



Área Académica de Administración de Tecnologías de Información

Propuesta para la elaboración de reportes crediticios de una entidad
financiera

Trabajo final de graduación para optar al grado de Licenciatura en Administración de
Tecnología de Información

Elaborado por: Fernando Antonio Mata Orozco

Prof. Tutor: Lorena Zúñiga Segura

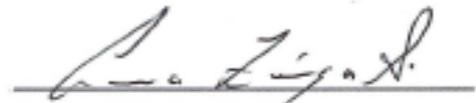
Cartago, Costa Rica

Noviembre, 2018



ÁREA ACADÉMICA DE ADMINISTRACIÓN DE TECNOLOGÍAS DE INFORMACIÓN
GRADO ACADÉMICO: LICENCIATURA

Los miembros del Tribunal Examinador del Área Académica de Administración de Tecnologías de Información, recomendamos que el siguiente Trabajo Final de Graduación del estudiante Fernando Antonio Mata Orozco sea aceptado como requisito parcial para optar al grado académico de Licenciatura en Administración de Tecnología de Información.



PhD. Lorena Zúñiga Segura
Profesor Tutor



María José Artavia
Profesor Lector



MBA. Luis Javier Chavarría Sánchez
Coordinador de la Carrera



MBA. Sonia Mora González
Coordinadora del Trabajo Final de Graduación

Nota aclaratoria

Género:

La actual tendencia al desdoblamiento indiscriminado del sustantivo en su forma masculina y femenina va contra el principio de economía del lenguaje y se funda en razones extralingüísticas. Por tanto, deben evitarse estas repeticiones, que generan dificultades sintácticas y de concordancia, que complican innecesariamente la redacción y la lectura de los textos.

Este documento se redacta de acuerdo con las disposiciones actuales de la Real Academia Española con relación al uso del “género inclusivo”. Al mismo tiempo se aclara que estamos a favor de la igualdad de derechos entre los géneros.

Dedicatoria

A mis padres por ser ejemplo de esfuerzo y dedicación durante todos estos años, por brindarme la oportunidad de estudiar y elegir el camino de vida, pese a las adversidades siempre han sido constantes con su apoyo incondicional.

Son y seguirán siendo mi más valioso tesoro y les dedico estas palabras.

Agradecimientos

Propuesta para la elaboración de reportes crediticios de una entidad financiera

A mis padres Fernando y Myriam por su apoyo incondicional durante las distintas etapas de mi vida, por su constancia y paciencia en los momentos difíciles.

A mi tía Roxana, Magda y abuela Doris por formar parte integral de mi formación ética, académica y profesional, por todas sus horas de esfuerzo y dedicación que invirtieron para convertirme en quien soy ahora.

A la profesora Lorena Zúñiga por su apoyo y guía durante la elaboración de este proyecto.

A Lorena Castillo por sus sabias palabras y el apoyo brindado para el desarrollo de este proyecto, todo esto no habría sido posible sin su colaboración.

A Cinthya por su apoyo, paciencia y motivación durante este proceso, por sus palabras de aliento y por estar a mi lado sin importar lo complicado que fuese.

A la familia Vargas Ramírez por su apoyo durante el desarrollo de su carrera y por abrirme las puertas de su hogar de forma tan atenta.

A todos mis amigos y compañeros que formaron parte de mi proceso de formación académico, por su simpatía, consejos, buenos deseos e increíbles momentos que compartimos a lo largo de mi educación universitaria.

ÍNDICE GENERAL

	Página
1. Introducción	1
1.1. Antecedentes.....	3
1.1.1. Descripción de la organización	3
1.1.2. Equipo de trabajo del proyecto.....	5
1.1.3. Trabajos similares realizados dentro y fuera de la organización	6
1.2. Planteamiento del problema	7
1.2.1. Situación problemática	7
1.2.2. Beneficios o aportes del proyecto a la organización	9
1.3. Objetivos del proyecto.....	11
1.4. Alcance	12
1.4.1. Fase I: Descubrimiento inicial.....	13
1.4.2. Fase II: Diseño de los componentes de inteligencia de negocio	14
1.4.3. Fase III: Implementación de los componentes de inteligencia de negocio	15
1.4.4. Fase IV: Comprobación de resultados	16
1.5. Aspectos excluidos del alcance del proyecto	16
1.6. Supuestos	17
1.7. Entregables	18
1.7.1. Entregables de Gestión del proyecto	18
1.7.2. Entregables de producto	18
1.7.3. Entregables académicos	22
1.8. Limitaciones	22
2. Marco teórico	23
2.1. Inteligencia de negocio (IN)	23
2.2. Arquitectura de inteligencia de negocio	24
2.3. Data warehouse y Data mart.....	26
2.4. Modelos multidimensionales	30

Propuesta para la elaboración de reportes crediticios de una entidad financiera

2.4.1.	Tablas de hechos	31
2.4.2.	Tablas dimensionales	33
2.4.3.	Granularidad.....	34
2.4.4.	Ventajas de los modelos multidimensionales.....	36
2.4.5.	Esquemas del modelo multidimensional.....	36
2.4.6.	Implementaciones físicas del modelo multidimensional	39
2.5.	Proceso de Extracción, Transformación y Carga	40
2.5.1.	Funciones de extracción	43
2.5.2.	Funciones de transformación	44
2.5.3.	Funciones de carga	45
2.5.4.	Enfoque ECL-TL.....	46
2.6.	Metadatos.....	48
2.7.	Visualización de datos en inteligencia de negocios.....	51
2.7.1.	Tipos de visualización	52
2.8.	Enfoques para el desarrollo de un <i>Data mart</i>	58
2.8.1.	Enfoque planteado por Paulraj Ponniah	59
2.8.2.	Enfoque planteado por Ralph Kimball y Margi Ross	69
2.9.	Prácticas recomendadas para implementar un <i>Data mart</i>	76
2.9.1.	Mejores prácticas para diseño del modelo multidimensional	76
2.9.2.	Mejores prácticas para implementar procesos de ETL	78
2.10.	Prácticas para evaluar alternativas de implementación	81
2.10.1.	Costos asociados a la evaluación de proyectos	82
2.10.2.	Cálculo del ROI.....	83
2.10.3.	Flujos de caja descontados	84
3.	Desarrollo Metodológico	86
3.1.	Tipo de investigación.....	87
3.1.1.	Investigación Cuantitativa.....	87

Propuesta para la elaboración de reportes crediticios de una entidad financiera

3.1.2.	Investigación Cualitativa.....	88
3.1.3.	Investigación Mixta.....	89
3.1.4.	Selección del tipo de investigación	90
3.2.	Diseño de la investigación.....	91
3.3.	Fuentes de información	92
3.3.1.	Fuentes primarias	92
3.3.2.	Fuentes secundarias.....	94
3.4.	Instrumentos y técnicas para recolección de datos	94
3.4.1.	Entrevista.....	94
3.4.2.	Revisión documental.....	96
3.5.	Instrumentos y técnicas para el análisis de datos	96
3.5.1.	Tabulación de datos.....	96
3.5.2.	Diagramación	96
3.5.3.	Comparativo para herramientas de integración y visualización.....	97
3.6.	Selección de la muestra para análisis de investigación	97
3.6.1.	Descripción de los sujetos de información.....	98
3.7.	Análisis de la información.....	101
3.8.	Fases de la metodología aplicada	102
3.8.1.	Fase I: Descubrimiento inicial.....	105
3.7.2.	Fase II: Diseño de los componentes de inteligencia de negocio	111
3.7.3.	Fase III: Implementación del <i>Data mart</i> , proceso de ETL y visualizaciones	116
3.7.4.	Fase IV: Comprobación de resultados	119
4.	Análisis de resultados	121
4.1.	Fase I: Descubrimiento inicial.....	121
4.1.1.	Entender el negocio	122
4.1.2.	Revisar las mejores prácticas y restricciones regulatorias	145
4.1.3.	Analizar las alternativas	152

Propuesta para la elaboración de reportes crediticios de una entidad financiera

4.2.	Fase II: Diseño de los componentes de inteligencia de negocio	162
4.2.1.	Diseñar el <i>Data mart</i> (DM).....	163
4.2.2.	Diseñar el proceso ETL.....	165
4.2.3.	Diseñar las visualizaciones.....	167
4.3.	Fase III: Implementación del <i>Data mart</i> , proceso de ETL y visualizaciones	168
4.3.1.	Implementar el <i>Data mart</i> (DM).....	168
4.3.2.	Implementar el proceso ETL.....	168
4.3.3.	Implementar las visualizaciones.....	169
4.4.	Fase IV: Comprobación de resultados.....	169
4.4.1.	Ejecutar las pruebas.....	169
4.4.2.	Realizar el comparativo de resultados.....	170
5.	Propuesta de solución	172
5.1.	Diseño del modelo multidimensional del <i>Data mart</i>	173
5.2.	Diseño del proceso de Extracción Transformación y Carga de datos (ETL)	180
5.2.1.	Diagrama lógico del proceso.....	180
5.2.2.	Detalle de las transformaciones.....	181
5.2.3.	Mecanismos de integración entre transformaciones y carga.....	182
5.2.4.	Periodicidad de ejecución.....	182
5.2.5.	Prerrequisitos.....	182
5.3.	Diseño de las visualizaciones de datos	182
5.3.1.	Mapeo de requerimientos contra las funcionalidades de las herramientas	183
5.4.	Implementación de las soluciones de IN	185
5.4.1.	Implementación del modelo multidimensional	185
5.4.2.	Implementación del proceso de ETL	186
5.4.3.	Implementación de las visualizaciones de datos	203
5.5.	Comprobación de resultados del proceso de ETL	206
5.5.1.	Validación de la cantidad de registros procesados.....	207

Propuesta para la elaboración de reportes crediticios de una entidad financiera

5.5.2. Validación del mapeo para datos categóricos	209
5.6. Verificación del proceso para creación de reportes post implementación de la propuesta.....	213
6. Conclusiones	218
7. Recomendaciones	222
8. Apéndices.....	224
Apéndice A: Tabla para Benchmarking comparativo de herramientas	224
Apéndice B: Guía de entrevista para identificación de necesidades	225
Apéndice C: Plantilla para recolección de requisitos	226
Apéndice D: Guía de entrevista para mapeo de fuentes transaccionales	227
Apéndice E: Guía de entrevista para mapeo de los procesos de elaboración de reportes	228
Apéndice F: Guía de entrevista para selección de criterios del <i>Benchmarking</i>	229
Apéndice G: Plantilla de paquetes de información	230
Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta	231
Apéndice J: Aprobación del modelo de datos para el Data mart.....	232
Apéndice K: Aprobación de visualizaciones.....	233
Apéndice L: Composición de las tablas utilitarias para proceso de ETL.....	234
Apéndice M: Entrevista sobre necesidades con analista de riesgo crediticio.....	236
Apéndice N: Entrevista sobre necesidades con analista de riesgo cambiario y liquidez	237
Apéndice Ñ: Entrevista sobre necesidades con el auxiliar de riesgos	238
Apéndice O: Entrevista sobre necesidades con analista operativo de crédito	239
Apéndice P: Entrevista sobre necesidades con analista de mora.....	240
Apéndice Q: Entrevista sobre necesidades con analista de mercado	241
Apéndice R: Entrevista sobre necesidades con analista financiero	242
Apéndice S: Recolección de requisitos con el área de riesgos	243
Apéndice T: Recolección de requisitos con el área de crédito	246
Apéndice V: Entrevista sobre fuentes transaccionales con analista de mora.....	249

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Apéndice W: Entrevista sobre fuentes transaccionales con analista de mercado.....251

Apéndice X: Entrevista sobre fuentes transaccionales con analista financiero.....253

Apéndice U: Entrevista sobre fuentes transaccionales con analista operativo de crédito 255

Apéndice Y: Entrevista sobre fuentes transaccionales con analista de riesgo crediticio .257

Apéndice Z: Entrevista sobre fuentes transaccionales con analista de riesgo cambiario y liquidez259

Apéndice AA: Entrevista sobre fuentes transaccionales con auxiliar de riesgo.....261

Apéndice BB: Entrevista para mapeo de procesos para elaboración de reportes analista de riesgo crediticio263

Apéndice CC: Entrevista para mapeo de procesos para elaboración de reportes analista de riesgo de liquidez y riesgo cambiario265

Apéndice DD: Entrevista para mapeo de procesos para elaboración de reportes auxiliar de riesgo267

Apéndice EE: Entrevista para mapeo de procesos para elaboración de reportes – analista operativo de crédito269

Apéndice FF: Entrevista para mapeo de procesos para elaboración de reportes analista de mora.....271

Apéndice GG: Entrevista para mapeo de procesos para elaboración de reportes analista de mercado273

Apéndice HH: Entrevista para mapeo de procesos para elaboración de reportes analista financiero.....275

Apéndice II: Entrevista para selección de criterios del Benchmarking.....277

Apéndice JJ: Ejecución de pruebas para comprobar la cantidad de registros procesados279

Apéndice KK: Ejecución de pruebas para comprobar las transformaciones de campos numéricos a rangos de valores.....281

Apéndice LL: Comprobación de resultados del proyecto283

9. Anexos285

Anexo 1: Notación estándar BPMN 2.0285

Anexo 2: Plantilla para la documentación de pruebas de la organización286

Anexo 3: Aval del profesor tutor para la defensa del TFG	287
Anexo 4: primera evaluación del estudiante.....	288
Anexo 5: segunda evaluación del estudiante	289
Anexo 6: tercera evaluación del estudiante	290
Anexo 7: Plantilla para minuta de reunión	291
Anexo 8: Matriz de trazabilidad	291
10. Bibliografía.....	297
11. Glosario	300
11.1. Siglas.....	301

ÍNDICE DE FIGURAS

Figura No.	Descripción	Página
Figura 1.1.	Organigrama del equipo del proyecto.....	5
Figura 1.2.	Fases del proyecto.....	12
Figura 1.3.	Alcance del proyecto	14
Figura 2.1.	Arquitectura estándar de inteligencia de negocio.	24
Figura 2.2.	Arquitectura para la integración con Data marts.	28
Figura 2.3.	Tabla de hechos agregada.	32
Figura 2.4.	Tabla dimensional.....	33
Figura 2.5.	Esquema dimensional.	37
Figura 2.6.	Esquema copo de nieve.....	38
Figura 2.7.	Cubo OLAP.	40
Figura 2.8.	Tipos de carga.	46
Figura 2.9.	Proceso ECL-TL.	47
Figura 2.10.	Flujo de metadatos en el proceso de ETL.....	49
Figura 2.11.	Ejemplo de gráfico lineal.	53
Figura 2.12.	Ejemplo de gráfico de columna.	54
Figura 2.13.	Ejemplo de gráfico de barras.	54
Figura 2.14.	Ejemplo de gráfico de columnas apiladas.....	55
Figura 2.15.	Ejemplo de gráfico de barras apiladas.	55

Figura 2.16. Ejemplo de gráfico de dispersión.....	55
Figura 2.17. Ejemplo de Slopegraph.	57
Figura 2.18. Gráfico de pastel sobre la atención de gestiones.	58
Figura 2.19. Propuesta de ciclo de vida para un DW según Paulraj Ponniah.	59
Figura 2.20. Ciclo de vida para proyectos de inteligencia de negocio.	69
Figura 3.1. Fases del proyecto.....	103
Figura 3.2. Fase I: Descubrimiento inicial.	106
Figura 3.3. Entendimiento del negocio, primera actividad de la fase I del proyecto.	106
Figura 3.4. Revisión de mejores prácticas, segunda actividad de la fase I del proyecto.	109
Figura 3.5. Análisis de alternativas, tercera actividad de la fase I del proyecto.	110
Figura 3.6. Fase II: Diseño de los componentes de inteligencia de negocio.....	112
Figura 3.7. Diseño del DM, primera actividad de la fase II del proyecto.	112
Figura 3.8. Diseño del proceso ETL, segunda actividad de la fase II del proyecto.	114
Figura 3.9. Diseño de visualizaciones, tercera actividad de la fase II del proyecto.....	115
Figura 3.10. Fase III: Implementación del Data mart, proceso de ETL y visualizaciones. .	116
Figura 3.11. Implementación del DM, primera actividad de la fase III del proyecto.	117
Figura 3.12. Implementación del proceso ETL, segunda actividad de la fase III del proyecto.	117
Figura 3.13. Implementación de visualizaciones, tercera actividad de la fase III del proyecto.	118
Figura 3.14. Fase IV: Comprobación de resultados.	119
Figura 4.1. Secuencia de actividades para la construcción de un reporte área de Riesgo....	138
Figura 4.2. Secuencia de actividades para la construcción de un reporte área de Crédito...	141
Figura 4.3. Proceso general para elaboración de reportes.....	145
Figura 5.1. Componentes de la propuesta de solución.....	172
Figura 5.2. Modelo lógico del Data mart	173
Figura 5.3. Flujo lógico del proceso de ETL.....	180
Figura 5.4. Implementación física del modelo multidimensional crediticio.....	185
Figura 5.5. Flujo del proceso ETL implementado en SQL Server Integration Services.....	186
Figura 5.6. Flujo de carga a base de datos intermedia	187
Figura 5.7. Composición de archivos en el directorio de carga intermedia	188
Figura 5.8. Codificación de datos para carga a base de datos intermedia.....	189
Figura 5.9. Mapeo de campos para carga a la base de datos intermedia de la solución.	190
Figura 5.10. Consulta sobre los datos de la base de datos intermedia de Crédito.....	190

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Figura 5.11. Flujo de carga a la dimensión ejecutivo.....	191
Figura 5.12. Flujo de carga a la dimensión cliente.....	192
Figura 5.13. Estructura condicional para el atributo generador de divisas de un cliente.....	192
Figura 5.14. Creación de nueva columna para el atributo Generador de divisa.	193
Figura 5.15. Bifurcación del flujo de procesamiento basado en categoría de riesgo.....	193
Figura 5.16. Creación de la columna para la nueva categoría de riesgo.....	193
Figura 5.17. Transformación del tipo de datos para los atributos Generador de divisa y Categoría de Riesgo.....	194
Figura 5.18. Mapeo de los datos procesados y la dimensión cliente.....	194
Figura 5.19. Flujo de carga a la dimensión operación.	195
Figura 5.20. Estructura condicional de días de atraso para división del flujo de procesamiento.....	196
Figura 5.21. Mapeo de los campos de la fuente transaccional contra la tabla de operaciones en el DM.....	196
Figura 5.22. Flujo de carga a la dimensión tiempo.....	197
Figura 5.23. Estructura condicional de la dimensión temporal.....	198
Figura 5.24. Creación de nuevos atributos para la primera ramificación del flujo en dimensión tiempo.....	198
Figura 5.25. Conversión de las columnas trimestre y semestre.....	198
Figura 5.26. Mapeo de información de la fuente transaccional y la dimensión tiempo.....	199
Figura 5.27. Flujo de carga a la tabla de hechos.....	200
Figura 5.28. Mapeo para la búsqueda de la llave primaria de la dimensión operación.....	200
Figura 5.29. Búsqueda del tipo de cambio para las transformaciones de saldos en la tabla de hechos.....	201
Figura 5.30. Transformación del monto total de la deuda a dólares.....	201
Figura 5.31. Mapeo de atributos para la tabla de hechos.....	202
Figura 5.32. Reporte gráfico de monto total de la deuda dolarizado.....	203
Figura 5.33. Reporte gráfico del monto principal directo por moneda.....	204
Figura 5.34. Reporte gráfico del monto de estimación genérica por año.....	205
Figura 5.35. Dashboard de la cartera crediticia.....	206
Figura 5.36. Consulta ejecutada sobre la dimensión tiempo.....	207
Figura 5.37. Consulta ejecutada sobre la dimensión cliente.....	208
Figura 5.38. Consulta ejecutada sobre la dimensión operación.....	208
Figura 5.39. Consulta ejecutada sobre la dimensión ejecutivo.....	209

Figura 5.40. Consulta ejecutada sobre la tabla de hechos	209
Figura 5.41. Consulta de clientes Generadores de divisas	210
Figura 5.42. Consultas sobre la categoría de riesgo de la dimensión cliente	211
Figura 5.43. Consulta a la tabla operación sobre días de atraso de 0-30	211
Figura 5.44. Consulta a la tabla operación sobre días de atraso de 31-60	212
Figura 5.45. Consulta a la tabla operación sobre días de atraso de 61-90	212
Figura 5.46. Consulta a la tabla operación sobre días de atraso de 91-120	212
Figura 5.47. Consulta a la tabla operación sobre días de atraso más de 120 días	212
Figura 5.48. Tiempo dedicado a la extracción de los datos en los archivos fuente	214
Figura 5.49. Tiempo dedicado al procesamiento y carga de datos al modelo multidimensional	215

ÍNDICE DE TABLAS

Tabla No.	Descripción	Página
Tabla 1.1.	Entregables por fase	19
Tabla 2.1.	Diferencias entre sistemas transaccionales y los DW.	27
Tabla 2.2.	Ejemplo de representación tabular.	56
Tabla 2.3.	Ejemplo de mapa de calor.	56
Tabla 2.4.	Métodos para el desarrollo de un DW o DM.	58
Tabla 2.5.	Contenido del plan de capacitación.....	68
Tabla 3.1.	Descripción de los sujetos de información.....	99
Tabla 3.2.	Objetivos por fase metodológica.....	104
Tabla 4.1.	Requisitos para estructura del repositorio multidimensional Riesgos.	124
Tabla 4.2.	Requisitos para procesamiento de datos Riesgos.....	125
Tabla 4.3.	Requisitos para visualizaciones de datos Riesgos.....	126
Tabla 4.4.	Requisitos para estructura del repositorio multidimensional Crédito.	127
Tabla 4.5.	Requisitos para procesamiento de datos Crédito.	128
Tabla 4.6.	Requisitos para visualizaciones de datos Crédito.	129
Tabla 4.7.	Descripción de las fuentes de datos	135
Tabla 4.8.	Tiempo promedio para la construcción de informes.....	139
Tabla 4.9.	Tiempo para elaboración de informes crédito.....	143

Tabla 4.10. Resumen de artículos tomados en cuenta para las restricciones regulatorias del proyecto.....	146
Tabla 4.11. Verificación de factibilidad de buenas prácticas.....	150
Tabla 4.12. Revisión de brechas con respecto a las mejores prácticas	151
Tabla 4.13. Comparativo de características para herramientas de integración	154
Tabla 4.14. Comparativo de características para herramientas de integración	156
Tabla 4.15. Tabla resumen de la cotización para el software de visualización de datos	157
Tabla 4.16. Tabla de resumen de las cotizaciones para el software de integración de datos.	157
Tabla 4.17. Mapeo del costo de licencias para herramientas de implementación.	158
Tabla 4.18. Desglose de costos para herramientas de integración y visualización de datos	160
Tabla 4.19. Cálculos del ROI para las alternativas de inversión.....	162
Tabla 4.20. Paquete de información para la cartera crediticia	163
Tabla 4.21. Descripción de las métricas de evaluación para el proceso de construcción de informes.....	170
Tabla 5.1. Descripción del modelo multidimensional.....	174
Tabla 5.2. Detalle de las transformaciones para el proceso de ETL	181
Tabla 5.3. Mapeo de requisitos contra funcionalidades de las herramientas	184
Tabla 5.4. Cantidad de registros en la fuente transaccional	207
Tabla 5.5. Beneficios esperados del proyecto al implementar la propuesta de solución	214
Tabla 5.6. Tiempos para la construcción de reportes en el departamento de riesgo post implementación de la propuesta	216
Tabla 5.7. Tiempos para la construcción de reportes en el departamento de riesgo post implementación de propuesta.....	216
Tabla 8.1. Simbología notación BPMN 2.0	285

Propuesta para la elaboración de reportes crediticios de una entidad financiera

El siguiente documento describe las necesidades de la entidad financiera que patrocinó el proyecto, esas necesidades ofrecieron la base para el diseño y desarrollo de una solución de inteligencia de negocio que elimina la ejecución de tareas manuales relacionadas a la extracción, transformación y carga de un conjunto diverso de información, disminuyendo la cantidad del tiempo invertida en tareas operativas de los departamentos de riesgo y crédito de la organización.

Actualmente los analistas de la entidad financiera ejecutan una variedad de tareas manuales para crear múltiples reportes crediticios, esto tiene un impacto sobre los tiempos de entrega acordados de la información que forma parte del proceso de toma de decisiones.

El objetivo del Proyecto se enfocó en el diseño y desarrollo de una solución de IN que disminuyera el tiempo que le toma a un analista desarrollar el reporte.

Desde una perspectiva metodológica la iniciativa sigue cuatro fases principales que incluyen el descubrimiento inicial del estado actual del proceso para la creación de reportes, evaluación de herramientas especializadas en integración y visualización de datos, el diseño del modelo multidimensional, proceso de ETL y los tableros de información basados en la información crediticia o de préstamos, finalmente, el proyecto concluye con la evaluación respectiva de los resultados de la iniciativa y los objetivos del proyecto.

Haciendo uso de la técnica que entrevista a través del desarrollo del proyecto fue posible determinar las causas que impactan la entrega oportuna de los reportes crediticios y señalar una solución concreta para los problemas organizacionales relacionados a la entrega tardía de información crucial.

Tomando en cuenta lo anterior la institución financiera necesita sacar ventaja de los beneficios que la propuesta del proyecto en el corto plazo para maximizar el tiempo invertido en tareas de analítica pura.

Palabras clave: finanzas, crédito, inteligencia de negocio, reportes, tableros de control

Abstract

The following document describes the necessities of the financial entity that sponsored the project, those necessities offered the basis for the design and development of a business intelligence solution that eliminates the execution of manual tasks regarding the extraction, transformation and load of a diverse sets of information, decreasing the amount of time spent in operative tasks on the departments of risk and credit of the organization.

Currently the analyst of the financial entity perform a variety of manual tasks in order to create several credit reports, this has an impact on the agreed delivery times of the information that forms part of the decision making process.

The project objective focused on the design and development of a BI solution that decreases the time that takes an analyst to develop the report.

From a methodological perspective the initiative follows four main phases that include an initial discovery of the current state of the report creation processes, evaluation of tools specialized on data integration and visualization, the design of the multidimensional model, ETL process, and dashboards based on credit or loan information, finally the project concludes with the respective evaluation the results of the initiative and the project objectives.

Making use of the interview technique throughout the project development it was possible to determine the underlying causes that impacted the timely delivery of the credit reports and pinpoint a concrete solution to the organizational issues regarding the late delivery of crucial data.

Taking into account the above the financial institution needs to leverage the benefits that the project proposal provided in the short term in order to maximize the time invested in pure analytical tasks.

Keywords: finance, credit, business intelligence, reports, dashboard

1

Introducción

La automatización en el procesamiento de datos para analítica de negocios se ha convertido en un pilar de competitividad en las organizaciones financieras a nivel mundial.

Una gestión oportuna y decisiones basadas en datos cuantificables brindan a las entidades bancarias flexibilidad y una visión clara sobre nuevos escenarios u oportunidades para reducir sus costos, incrementar su margen de ganancia o competitividad en el sector financiero (Michalewicz, Michalewicz, Chiriac y Schmidt, 2006).

Debido a esto, la necesidad de las organizaciones por mantener un flujo de datos oportuno para la toma de decisiones se ha incrementado durante la última década. Según mencionan Moro, Cortez y Rita (2015), la resolución de problemas complejos y el proceso para la toma oportuna de decisiones en la industria financiera demanda el uso de sistemas o mecanismos de soporte e inteligencia de negocio que apoyen los análisis.

La toma de decisiones informada y oportuna fundamenta los cursos de acción que siguen las organizaciones para mantener o incrementar su margen competitivo. En este trabajo se estudiarán las causas que afectan la entrega oportuna de los reportes crediticios destinados al proceso de toma de decisiones en una entidad financiera ubicada en Costa Rica.

El trabajo desarrollado, se enfoca en el diseño e implementación de pruebas de concepto para un modelo multidimensional, proceso de ETL y las visualizaciones de datos que requieren los analistas crediticios y gerentes de la organización durante el proceso de toma de decisiones. Además, la propuesta de solución planteada toma en cuenta las restricciones impuestas por los entes reguladores de instituciones financieras para el procesamiento de datos personales de clientes.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

El presente documento, se compone de seis capítulos cuyo contenido se detalla a continuación:

- ❏ Capítulo I: Introducción. En este capítulo se incluyen los antecedentes de la organización donde se desarrolla el TFG, el problema a resolver y los objetivos del proyecto. Adicionalmente, se delimita el alcance del trabajo realizado, sus limitaciones y supuestos.
- ❏ Capítulo II: Marco teórico. Este capítulo detalla el sustento teórico que respalda la elaboración de la propuesta de solución del proyecto y los elementos que lo componen.
- ❏ Capítulo III: Marco Metodológico. El marco metodológico incluido en este capítulo especifica los métodos y procedimientos seleccionados para la recolección y análisis de datos.
- ❏ Capítulo IV: Resultados. En este capítulo se detallan los resultados obtenidos al aplicar los instrumentos y técnicas seleccionadas en la metodología del proyecto. Los datos incluidos en el capítulo contemplan los resultados de las dos etapas que conforman el proyecto.
- ❏ Capítulo V: Propuesta. En este capítulo se especifica la solución planteada para resolver la situación problemática que afecta la organización.
- ❏ Capítulo VI: Conclusiones y recomendaciones. Finalmente, en este capítulo se exponen las conclusiones generadas producto de los resultados obtenidos del proyecto.

1.1. Antecedentes

En esta sección se especifican los antecedentes de la entidad financiera donde se realizó el TFG, el equipo del proyecto y los trabajos similares realizados anteriormente en la organización.

1.1.1. Descripción de la organización

El presente trabajo se desarrolla en una entidad bancaria con más de 10 años de experiencia en la oferta de servicios financieros y con posicionamiento en las distintas provincias del territorio costarricense. A continuación, se presenta la misión y visión organizacional.

Misión

Ofrecer a pequeñas y medianas empresas oportunidades de financiamiento, mediante servicios de un alto valor agregado con un enfoque personalizado y de respuesta rápida.

Visión

Aspiramos a distinguirnos por ser una entidad financiera centrada en los buenos hábitos del ahorro, trabajo arduo, constancia y fidelidad de la relación comercial con nuestros clientes.

La organización nace con el objetivo de promover el desarrollo comercial entre Costa Rica y países ubicados alrededor del mundo. Las estrechas relaciones que posee la entidad con sus bancos corresponsales favorecen la intermediación ágil y segura con entes financieros fuera del país.

La institución se encuentra enfocada en un nicho de índole empresarial, sus servicios están dirigidos a financiar los medianos y pequeños empresarios costarricenses.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Para facilitar la oferta y prestación de servicios, la organización cuenta con el apoyo de sistemas informáticos que brindan soporte a sus operaciones y funcionan como catalizadores para las propuestas de innovación en todos los departamentos.

La estructura departamental de la organización está conformada por las áreas de Riesgo, Finanzas, Recursos Humanos, Seguridad, Negocios, Tecnologías de Información, Crédito, Comercial, Mercadeo, Oficialía de Procesos, Operaciones Bancarias, Contabilidad, Análisis de Crédito, Banca de Personas, Coordinación de Sucursales y la Oficialía de Cumplimiento.

Actualmente la entidad financiera cuenta con 250 colaboradores y más de quince sucursales a nivel nacional, las cuales se enfocan en el servicio personalizado a sus clientes.

Manteniendo el lema de excelencia en el servicio, la entidad ha ganado reconocimiento y exposición a nivel nacional, incrementando su valor en activos y favoritismo entre las opciones de financiamiento de los costarricenses durante la última década.

En cuanto a los servicios que ofrece la institución a sus clientes se destaca principalmente:

-  Líneas de Crédito empresariales y personales.
-  Créditos específicos.
-  Intra y extra-financiamientos.
-  Crédito prendario.
-  Tarjeta de Crédito.
-  Cuentas de ahorro, corrientes y alto rendimiento.
-  Apertura de inversiones a plazo.
-  Transferencias al exterior y a nivel nacional.
-  Mesa de cambio (compra y venta de divisas).
-  Descuento de Facturas.
-  Arrendamiento de activos.

1.1.2. Equipo de trabajo del proyecto

Según se establece en la Figura 1.1 el desarrollo de la iniciativa se ve influenciado principalmente por tres colaboradores de la entidad financiera vinculados a las áreas de tecnología, riesgo y crédito.

Los aportes desde la perspectiva de negocio se obtienen de la Encargada de Riesgo de Crédito y de la jefatura de Licitaciones y Factoreo, mientras que la jefatura de sistemas provee la perspectiva técnica y experiencia de implementación adquirida durante proyectos similares ejecutados en la entidad.

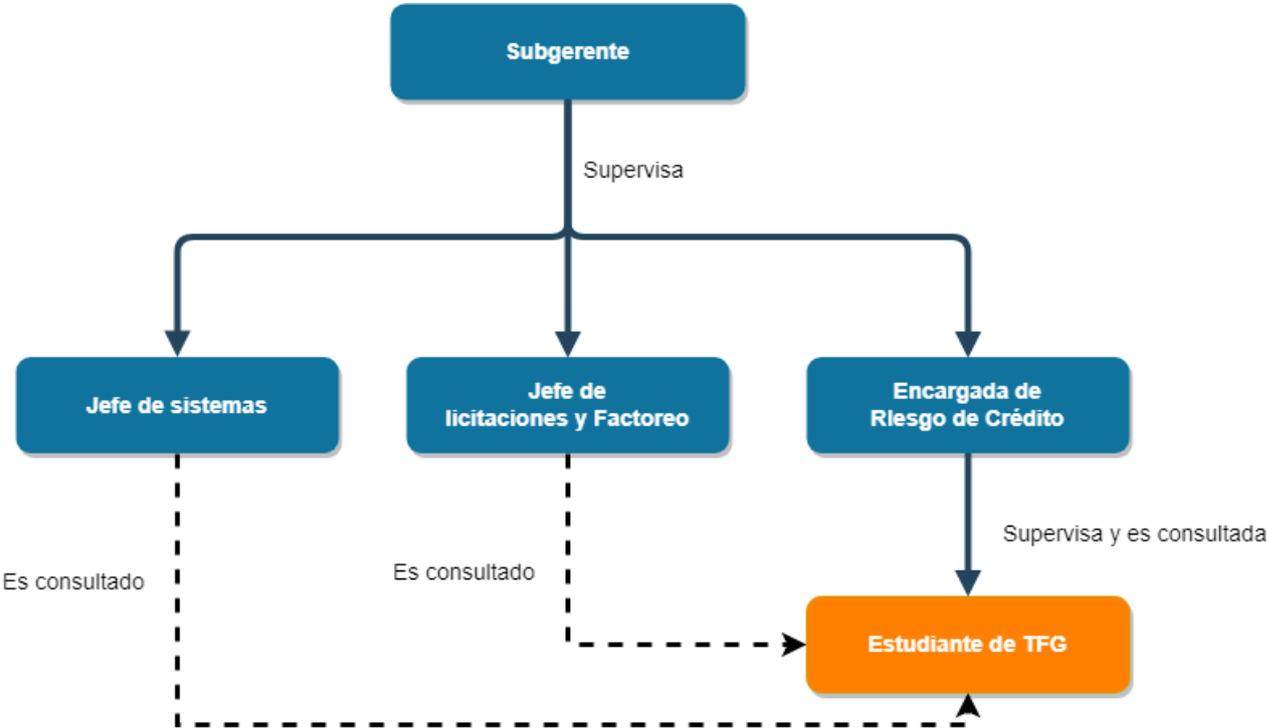


Figura 1.1. Organigrama del equipo del proyecto

Nota: elaboración propia

Adicionalmente, cabe destacar que el rol de supervisión que cumple la persona Encargada de riesgo crediticio favorece el desarrollo del proyecto, debido a su vasta experiencia en las áreas de Riesgo y Crédito del sector financiero.

1.1.3. Trabajos similares realizados dentro y fuera de la organización

Anteriormente la entidad implementó tres iniciativas relacionadas con la automatización de tareas operativas que se ejecutan durante la construcción de informes. El presente trabajo considera las lecciones aprendidas que se documentaron al finalizar cada iniciativa.

A continuación, se brinda una breve explicación de cada iniciativa desarrollada en la institución:

1.1.3.1. Automatización por medio de Macros en el área financiera

El área de finanzas de la entidad implementó un conjunto de macros en Excel para automatizar el procesamiento de grandes volúmenes de partidas contables, dicho proceso se realizaba manualmente y su duración se extendía en dos semanas.

Los macros implementados permitieron extraer los valores disponibles para las partidas de análisis y convertirlas en indicadores para el Gerente Financiero de la organización.

1.1.3.2. Creación de herramienta para análisis de industria

Otra iniciativa que se gestó en el departamento de finanzas de la organización es el desarrollo de una herramienta dedicada a la creación de reportes que extrae datos del sector financiero nacional y procesa indicadores tales como tasas activas, pasivas y los saldos totales en cartera crediticia.

Cabe destacar que toda la información disponible en esta solución es de conocimiento público ya que la SUGEF publica cada cierre de mes los datos mencionados para cada entidad financiera pública o privada.

1.1.3.3. Creación del repositorio institucional

Finalmente, la iniciativa que se encuentra relacionada directamente con el desarrollo del proyecto es la creación del repositorio de datos institucional. La construcción de esta fuente de datos se gestó debido al cambio en el sistema transaccional maestro de la entidad.

El cambio en el sistema principal brindó la posibilidad para que el departamento de TI de la organización junto a representantes de las áreas de negocio desarrollaran una herramienta de almacenamiento histórico transaccional a disposición de los usuarios, sin

Propuesta para la elaboración de reportes crediticios de una entidad financiera

embargo no se tomaron en cuenta áreas relevantes de análisis para la entidad financiera, tales como crédito y riesgo.

Aunque la información se encuentra disponible en un motor de base de datos, su mecanismo de entrega es por medio de archivos de Excel, esto incrementa la cantidad de procesamiento manual que deben realizar los analistas financieros para obtener indicadores relevantes que den soporte a la toma de decisiones.

Actualmente la gestión y producción de los datos se encuentra centralizada en el departamento de Tecnología de información, esto repercute en los usuarios finales quienes deben incorporarse en una lista de espera para recibir los datos solicitados.

1.2. Planteamiento del problema

En esta sección se describe la situación problemática que enfrenta la entidad financiera la cual motiva el desarrollo del proyecto como medio para solventar las necesidades presentes en los procesos dedicados a la construcción de reportes crediticios. Adicionalmente, en esta sección se mencionan los beneficios directos e indirectos que generará el proyecto en la institución.

1.2.1. Situación problemática

La situación problemática se origina en los procesos vinculados a la elaboración de informes en las áreas de Riesgo y Crédito de la organización. Específicamente los procesos que presentan oportunidades de mejora son los enfocados en extracción y transformación de datos financieros que se utilizan para determinar la concentración de cartera y exposición crediticia.

La construcción de informes integra fuentes internas tales como información de clientes y operaciones de crédito, brindando a la entidad una visión integral sobre la situación y salud de su cartera crediticia. El consolidar la información disponible en las distintas fuentes de datos requiere que los analistas concilien y consoliden la información antes de iniciar la preparación de sus reportes.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

La información utilizada para análisis generalmente se encuentra disponible en archivos de Excel, archivos separados por comas (CSV) o documentos XML, los analistas deben estandarizar la información y posteriormente realizar sus cálculos lo que conlleva a una disminución en el tiempo de entrega de sus reportes.

Debido a la descentralización de las fuentes de datos y a la diferencia en las tareas que realizan los analistas durante la preparación de sus informes se han presentado diferencias en los montos de las operaciones e indicadores crediticios contruidos a partir de una misma fuente de datos.

Finalmente, la cantidad de iteraciones sobre el procesamiento de datos que realizan los analistas durante la construcción de los informes repercute en las fechas de entrega de la información crediticia.

1.2.1.1. Condiciones que enfrenta la organización producto de la situación problemática

Actualmente la fuerza laboral de la organización realiza los siguientes reportes con base en la cartera de operaciones crediticias:

- ❏ Matrices de transición (análisis sobre el deterioro en categoría de riesgo en la cartera crediticia).
- ❏ Informes en concentración de cartera (indica cuánto representa el crédito que posee un cliente con respecto a la cartera total de la entidad financiera).
- ❏ Informe de excepciones (destaca las excepciones que se realizaron durante el proceso de aprobación crediticio).
- ❏ Estudio del comportamiento sobre el monto total de la deuda y estimación que generan las operaciones de crédito.
- ❏ Evaluación del comportamiento de pago con respecto a los días de atraso de las operaciones crediticias.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Durante el desarrollo de los informes se han determinado condiciones que afectan negativamente la función analítica de las áreas de Crédito y Riesgo, a continuación, se detallan cada una de ellas.

La primera condición que afecta el desarrollo y entrega ágil de los reportes es la subutilización de herramientas especializadas en visualización de datos como *Tableau*, *Power BI* y *Reporting Services*. La subutilización de este software se debe a la dificultad de acceso para los usuarios finales y la centralización del conocimiento requerido para implementar los *dashboards* en el área de finanzas. La subutilización de este tipo de software representa un costo para la organización y una oportunidad perdida para incrementar la agilidad en la entrega de datos.

La segunda condición que repercute en la entrega oportuna de la información analítica es que el repositorio de información no es de acceso libre para los colaboradores de la entidad. La centralización del conocimiento en las áreas técnicas provoca que los departamentos opten por extraer la información y procesar sus reportes de forma independiente.

La situación descrita contradice lo establecido por Kimball y Ross (2013) quienes afirman que: “El *Data Warehouse* debe hacer la información de fácil acceso y su contenido debe ser comprensible para los usuarios de negocio.”.

Otro factor que provoca la construcción independiente de reportes es la incertidumbre sobre los datos incluidos en el repositorio institucional y su calidad. De mantenerse la incertidumbre sobre los datos que gestiona esta fuente de información, se podría correr el riesgo de desuso de esta herramienta.

1.2.2. Beneficios o aportes del proyecto a la organización

Con la realización de este proyecto se busca solventar en la entidad financiera los problemas mencionados anteriormente, obteniendo a su vez los siguientes beneficios directos e indirectos.

1.2.2.1. Beneficios directos

Producto de la ejecución del proyecto se espera que la solución propuesta ofrezca los siguientes beneficios a la entidad financiera.

 Disminución en las cargas de trabajo del personal

Se espera que el personal de análisis disminuya la cantidad de tiempo que invierte en tareas de extracción, preparación y carga de datos manual en al menos un 50% del tiempo actual.

 Disminución de la cantidad de iteraciones que realizan los analistas para construir sus reportes.

Se busca reducir en un 80% la cantidad de iteraciones realizadas por los analistas debido a cambios en los requerimientos gerenciales o errores en el procesamiento de información.

1.2.2.2. Beneficios indirectos

Al finalizar la ejecución del proyecto se espera que la propuesta genere los siguientes beneficios indirectos a la entidad:

 Diagnóstico sobre problemas en limpieza o estructura de los repositorios de información

Como parte del proyecto se identificarán fuentes de datos transaccionales que generan información para los análisis crediticios tales como el Core Bancario, el sistema que gestiona las Tarjetas de Crédito y el crédito empresarial. Durante el mapeo realizado se busca determinar cuáles repositorios poseen datos que deben incluirse posteriormente en iniciativas de limpieza y reestructuración de información.

1.3. Objetivos del proyecto

En esta sección se delimitan los objetivos del proyecto a realizar en la entidad financiera. Se formula el objetivo general del proyecto y los objetivos específicos.

Objetivo general

Desarrollar el diseño de un modelo multidimensional y la automatización de los procesos de ETL en las áreas de riesgo y crédito de una entidad bancaria, mediante un análisis de las mejores prácticas de la industria, para la disminución de las brechas en tiempos de entrega de información a los departamentos comerciales de la entidad financiera.

Objetivos específicos

1. Diagnosticar las brechas en las tareas de extracción, transformación y carga en los procesos de inteligencia de negocio del área de crédito y riesgo con respecto a las mejores prácticas de la industria, para la identificación de las causas que imposibilitan la entrega de información oportuna.
2. Recomendar tres alternativas de inversión acordes a las necesidades actuales de la entidad financiera en términos cualitativos y cuantitativos estableciendo a la gerencia opciones que permitan la implementación de la propuesta para la automatización del procesamiento de datos.
3. Diseñar el modelo multidimensional, *dashboards* de información crediticia y los procesos de ETL asociados a la preparación de datos que se utilizan como insumo en los análisis de las áreas de crédito y riesgo, para la entrega de información a las áreas comerciales.
4. Construir una solución automatizada que facilite el procesamiento de datos en las tareas analíticas de Crédito y Riesgo para la disminución del reprocesamiento de informes en un 80% y la reducción del tiempo de elaboración en por lo menos un 50%.

1.4. Alcance

Esta sección establece el alcance del proyecto planteado en la entidad bancaria como parte del Trabajo Final de Graduación de la carrera de Administración de Tecnología de Información.

El proyecto se enfoca en el área de inteligencia de negocios, específicamente en la rama dedicada a procesos de extracción, transformación y carga de datos (procesos ETL) sobre repositorios multidimensionales.

El desarrollo del proyecto se plantea en cuatro fases, la primera fase comprende el descubrimiento inicial de información interna y externa para el desarrollo del proyecto, la segunda fase se enfoca en el diseño del modelo multidimensional, proceso de ETL y visualizaciones de datos. Las últimas dos fases del proyecto se relacionan directamente con la implementación y comprobación de resultados de la propuesta de solución. La Figura 1.2, muestra de manera gráfica la secuencia de las fases del proyecto.

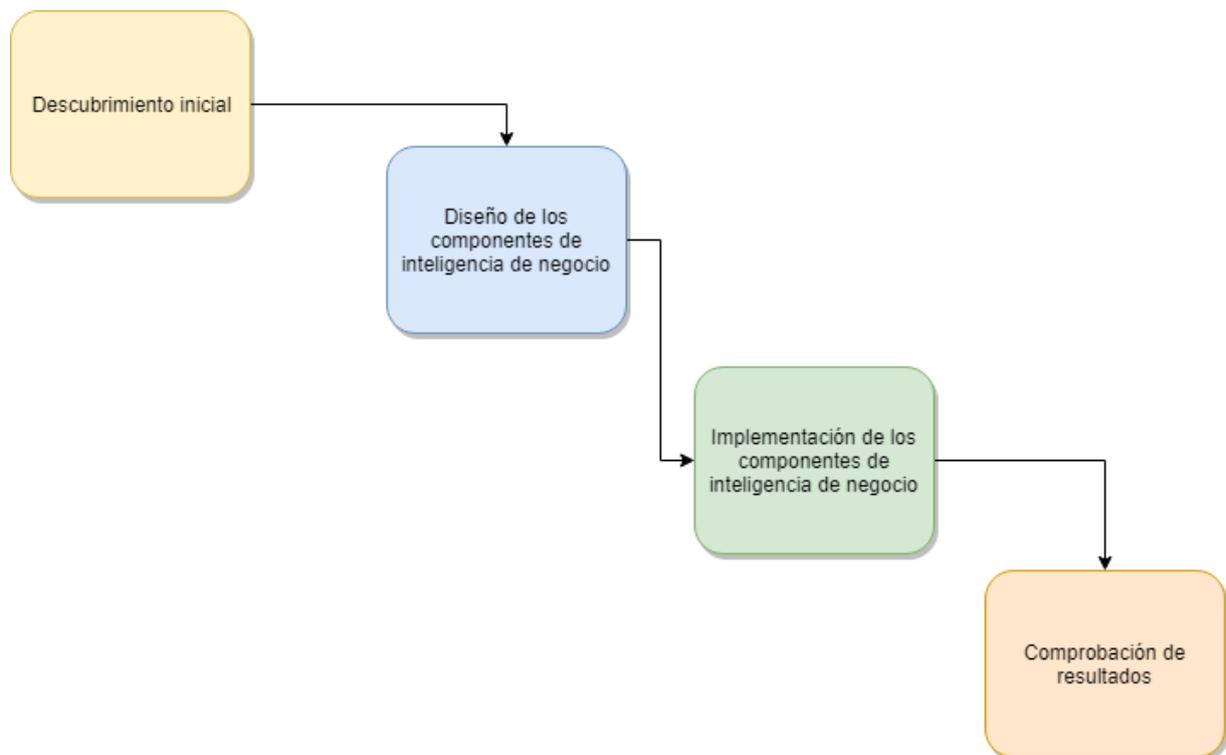


Figura 1.2. Fases del proyecto.

Nota: elaboración propia

1.4.1. Fase I: Descubrimiento inicial

La primera fase tiene como objetivo recolectar el respaldo teórico fundamental para el diseño e implementación de la propuesta que solventa la situación problemática en etapas posteriores de la iniciativa. Las actividades desarrolladas durante la primera etapa del proyecto son:

1.4.1.1. Desarrollar el compendio de información interna y externa

El compendio de información interna incorpora el conocimiento disponible en la entidad bancaria donde se desarrolla el TFG. La perspectiva interna incluye:

Conocimiento de negocio: se documentarán los informes a incluir en el proceso automatizado de ETL, las transformaciones que se deben realizar sobre los datos y las consideraciones o restricciones de la información a tomar en cuenta durante la creación de informes.

Fuentes de información a utilizar: otro componente a incluir en el compendio es el mapeo de las fuentes de datos transaccionales. Las fuentes de datos identificadas formarán parte del flujo de información que se integrará durante el diseño del proceso de ETL en la segunda etapa del proyecto.

Procesos actuales: se documentará el proceso para elaborar informes que realizan los analistas de los departamentos de Riesgo y Crédito. La documentación incluye mediciones con respecto al tiempo que toma crear cada informe.

La perspectiva externa del compendio incluye los siguientes elementos:

Restricciones regulatorias: este componente del compendio incorpora las restricciones regulatorias impuestas por la SUGEF y CONASSIF para el procesamiento de datos de los clientes.

Investigación de mejores prácticas: finalmente, se incluirán las mejores prácticas para el diseño de procesos ETL, esto establecerá el punto de partida para la fase de diseño e implementación de la propuesta.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

1.4.1.2. Diagnosticar los procesos de la entidad con respecto a las mejores prácticas

Se realizará el diagnóstico inicial de los procesos con el objetivo de medir el tiempo que requieren actualmente los analistas de Crédito y Riesgo para elaborar sus respectivos informes.

1.4.1.3. Benchmark sobre alternativas para implementar el proceso ETL

Se realizará un comparativo con criterios cualitativos para cuatro soluciones que ofrece el mercado para la integración y visualización de datos. Adicionalmente, se realizará el estudio cuantitativo de las alternativas óptimas de implementación. Al finalizar el comparativo se determinan las alternativas óptimas de inversión para implementar la propuesta del proyecto.

1.4.2. Fase II: Diseño de los componentes de inteligencia de negocio

Esta fase utiliza la información recolectada durante la primera fase del proyecto y la aplica en el diseño de los componentes de la solución de inteligencia de negocio. La Figura 1.3 muestra un resumen de las actividades principales que se realizan en la etapa II del proyecto.

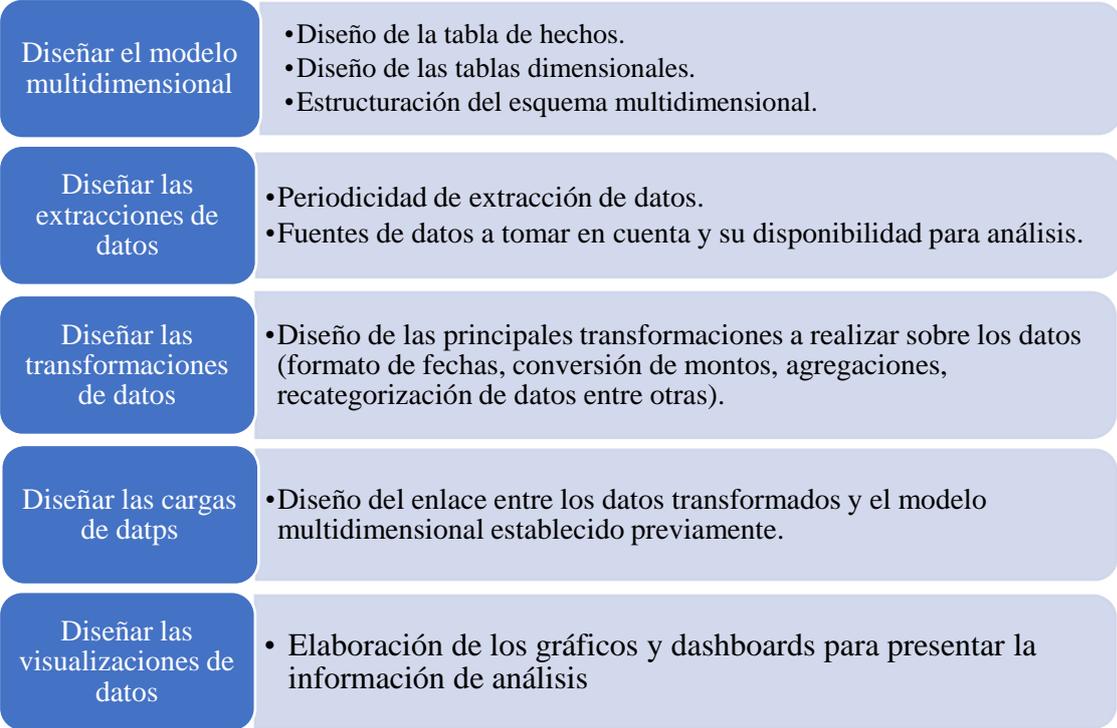


Figura 1.3. Alcance del proyecto

Nota: elaboración propia

A continuación, se describen las actividades de diseño correspondientes a la fase II del proyecto:

1.4.2.1. Diseñar el modelo multidimensional

El diseño del modelo multidimensional se realizará con base en los requerimientos establecidos por las áreas de crédito y riesgo, las características de la información almacenada en las fuentes de datos y la organización jerárquica de la información que posee la entidad.

1.4.2.2. Diseñar el proceso ETL

En esta etapa se diseñará el proceso de ETL que automatiza la extracción, transformación y carga de datos en el modelo multidimensional previamente establecido. El diseño del proceso considera la periodicidad de extracción de los datos, restricciones sobre la disponibilidad de la información, las transformaciones a realizar sobre los datos de análisis y las condiciones relacionadas a la carga al modelo multidimensional.

1.4.2.3. Diseñar las visualizaciones

Adicionalmente, se realizará el diseño de las visualizaciones asociadas a cada reporte relacionado a la información crediticia que se procesó por medio del ETL para su posterior presentación en dashboards.

1.4.3. Fase III: Implementación de los componentes de inteligencia de negocio

La fase tres tiene como objetivo implementar los diseños realizados en la etapa dos del proyecto. Las actividades por desarrollar durante la implementación de la propuesta son:

1.4.3.1. Desarrollar el modelo multidimensional

Utilizando los diseños realizados se procede a implementar el modelo de datos en la alternativa de software seleccionada. Una vez desarrollados los modelos multidimensionales se procede a la siguiente etapa del proceso de implementación.

1.4.3.2. Desarrollar el proceso de ETL

Se realizará la implementación del proceso de Extracción, Transformación y Carga en la herramienta seleccionada durante el comparativo de alternativas de implementación.

En este caso se programarán las extracciones periódicas de datos y la carga inicial. Adicionalmente, se incorporará a la solución el conjunto de transformaciones definidas por las

Propuesta para la elaboración de reportes crediticios de una entidad financiera

áreas de riesgo y crédito. Finalmente, se construirá el enlace entre el área de transformación y el repositorio multidimensional donde se almacenarán los datos crediticios.

1.4.3.2. Desarrollar las visualizaciones de datos

La actividad final de la fase III se encarga de la elaboración visual de los reportes en el software que dispone la entidad financiera. Se desarrollarán los gráficos requeridos y los tableros de control o dashboards.

1.4.4. Fase IV: Comprobación de resultados

La fase cuatro tiene como objetivo comparar las métricas obtenidas al ejecutar el proceso de ETL implementado contra los objetivos del proyecto. Las actividades por desarrollar durante la primera fase del proyecto son:

1.4.4.1. Evaluar los resultados con respecto a las métricas establecidas

El proyecto finaliza con la ejecución de las pruebas de conceptos sobre la solución implementada y el comparativo de los resultados correspondiente del proceso de ETL y los objetivos de mejora planteados. El comparativo se incluye en el informe de resultados del proyecto.

Para la documentación de las pruebas aplicadas sobre los componentes implementados se utilizará el formato establecido por la entidad financiera.

El detalle de las etapas del proyecto se encuentra especificado en el capítulo III (marco teórico) del presente documento.

1.5. Aspectos excluidos del alcance del proyecto

Los elementos que se describen a continuación se encuentran fuera del alcance del proyecto:

- ❏ Creación de estándares o planes de acción para el control de la calidad de datos en las fuentes transaccionales.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ Puesta en marcha de la propuesta en el ambiente productivo de la entidad financiera.

- ❏ Extracción, transformación y carga de datos no relacionados con las carteras crediticias de la organización.

1.6. Supuestos

Para el desarrollo de este proyecto se toma en consideración el siguiente listado de supuestos:

1. Se cuenta con la debida autorización de los clientes para almacenar y procesar su información.

2. Se cuenta con el apoyo y disposición del personal clave en los departamentos de TI, riesgo y crédito para el desarrollo del proyecto.

3. La versión de los datos disponibles para el procesamiento posibilita la elaboración del proceso de ETL.

4. Se pondrán a disposición del proyecto las herramientas de visualización necesarias para crear los dashboards de análisis.

1.7. Entregables

Esta sección describe los artefactos que se desarrollan como parte del proyecto planteado, los entregables se clasifican en tres tipos: entregables de gestión del proyecto, entregables de producto y los académicos.

1.7.1. Entregables de Gestión del proyecto

Los entregables de gestión se generan durante la ejecución del proyecto, ayudan a mantener un adecuado control sobre las actividades, alcance y cambios que se presentan en el desarrollo del trabajo.

Minutas de reunión

Las minutas de reunión son el mecanismo donde se registran los acuerdos tomados durante las reuniones del proyecto. Se generan dos tipos de minuta, una para las reuniones con la organización y otra para las reuniones con la contraparte académica.

Cronograma de proyecto

El cronograma de proyecto sirve de apoyo para dar seguimiento a las fechas de entrega para la organización y el área académica de Administración de Tecnología de Información.

Gestión de cambios

Permiten mantener un control sobre los cambios que se realizan a lo largo del proyecto, en dicho documento se detalla el motivo del cambio, impacto en términos de tiempo, alcance o costo.

Plantilla para requerimientos

Permite documentar los requerimientos organizacionales identificados en la fase inicial del proyecto y además guían el desarrollo de los entregables de producto que se obtendrán al finalizar el Trabajo Final de Graduación.

1.7.2. Entregables de producto

La Tabla 1.1, muestra la relación entre cada una de las etapas y los objetivos del proyecto, así como los entregables por cada una de las fases que componen el trabajo.

Tabla 1.1.
Entregables por fase

Etapa	Objetivos	Entregables
Fase I Descubrimiento inicial	Diagnosticar las brechas en las tareas de extracción, transformación y carga en los procesos de inteligencia de negocio del área de crédito y riesgo con respecto a las mejores prácticas de la industria, para la identificación de las causas que imposibilitan la entrega de información oportuna.	<ul style="list-style-type: none">  Compendio de información interna y mejores prácticas de la industria para soluciones de ETL y modelos multidimensionales.  Mapeo de fuentes transaccionales.
	Recomendar tres alternativas de inversión acordes a las necesidades actuales de la entidad financiera en términos cualitativos y cuantitativos estableciendo a la gerencia opciones que permitan la implementación de la propuesta para la automatización del procesamiento de datos.	<ul style="list-style-type: none">  Lista de requerimientos para proceso ETL.  Análisis de alternativas de inversión.
Fase II Diseño de los componentes de inteligencia de negocio	Diseñar el modelo multidimensional, dashboards de información crediticia y los procesos de ETL asociados a la preparación de datos que se utilizan como insumo en los análisis de las áreas de crédito y riesgo, para la entrega de información a las áreas comerciales.	<ul style="list-style-type: none">  Diseño del modelo multidimensional.  Diseño del proceso ETL.  Diseño de las visualizaciones de datos.
Fase III Implementación del Data mart, proceso de ETL y visualizaciones	Construir una solución automatizada que facilite el procesamiento de datos en las tareas analíticas de Crédito y Riesgo para la disminución del reprocesamiento de informes en un 80% y la reducción del tiempo de elaboración en por lo menos un 50%.	<ul style="list-style-type: none">  Implementación del modelo multidimensional, dashboards y el proceso de ETL.  Casos de prueba.
Fase IV Comprobación de resultados		<ul style="list-style-type: none">  Informe de resultados.

Nota: elaboración propia

Propuesta para la elaboración de reportes crediticios de una entidad financiera

A continuación, se detalla cada entregable de producto del proyecto:

Compendio de información interna y externa

Es un compilado de conocimiento interno y externo que debe ser incorporado durante el diseño del proceso de ETL. Los componentes asociados a cada tipo de conocimiento son:

- **Conocimiento interno**

Periodicidad de extracción, restricciones regulatorias, supuestos, cálculos o transformaciones en los datos, estandarizaciones en la información.

- **Conocimiento externo**

Mejores prácticas de la industria para el diseño de ETL, proyectos similares, restricciones regulatorias a tomar en cuenta.

Mapeo de fuentes transaccionales

Incluye la lista de las fuentes de información a utilizar como insumo durante el proceso de ETL, especificando mecanismos de comunicación con los repositorios o sistemas, datos disponibles en cada fuente y restricciones de extracción a tomar en cuenta durante el diseño del proceso automatizado de extracción.

Lista de requerimientos para el proceso ETL

Comprende el listado de requerimientos del área de crédito y riesgo de la organización que deben ser incluidos en el proceso de ETL. Este entregable funciona como mecanismo de control para verificar que el producto final cumple con lo solicitado inicialmente. La recolección de estos requerimientos es fundamental ya que los requisitos son la guía para construir la solución de inteligencia de negocios, siendo imprescindible que el “*Data Warehouse* incluya los principales elementos de información en los formatos requeridos por los usuarios” (Ponniah, 2001, P. 109).

Análisis de alternativas de implementación

Este entregable incluye el comparativo de alternativas de herramientas para la implementación del proceso de ETL y visualización de datos. Este entregable brinda los fundamentos para tomar la decisión sobre la solución óptima que debe adquirirse según las necesidades actuales de la organización.

Diseño del proceso ETL y modelo multidimensional

Este entregable incluye el mapeo de los requerimientos del negocio contra las funcionalidades de la herramienta de ETL, el modelo dimensional planteado, detalle de las transformaciones implementadas, mecanismos de integración entre la actividad de transformación y la carga de los datos resultantes en el modelo multidimensional, periodicidad de ejecución del proceso y prerequisites para el correcto funcionamiento del proceso automatizado.

Diseño de las visualizaciones

Este artefacto incluye el diseño de los *dashboards* solicitados por las áreas de Crédito y Riesgo, considerando el tipo de visualización adecuada para los datos de análisis. Adicionalmente, se incluyen los datos que se visualizaron y la periodicidad para generar los reportes gráficos. Este entregable es fundamental ya que según establece Eckerson “Lo que las organizaciones realmente necesitan es un tablero de desempeño que traslade la estrategia de la organización en objetivos, métricas e iniciativas y tareas dirigidas a cada individuo o grupo de la organización” (Eckerson, 2005, p. 4).

Casos de prueba

Este entregable incluye un conjunto de casos de prueba que fueron aplicados sobre el proceso automatizado de ETL, en cada caso de prueba se incluye las entradas que se brindaron al proceso, el objetivo de la prueba, el resultado esperado, resultado obtenido al ejecutar las pruebas.

Informe de resultados

Es el entregable final de producto que detalla el comparativo entre el proceso implementado y la situación previa de la entidad. Este informe incluye el comparativo de la situación previa del proceso y los resultados del proceso automatizado y especifica la mejora exacta generada por el proyecto.

1.7.3. Entregables académicos

A continuación, se describen los entregables académicos que se obtendrán como parte del desarrollo del proyecto:

Informes de avance

Los informes de avance son los entregables periódicos que el desarrollador del proyecto realiza al profesor tutor durante la realización del TFG. Dichos informes se entregan de manera periódica a lo largo de las 16 semanas de duración del semestre en el que se matricula el curso de TFG.

Informe final académico

El informe final académico es el resultado final de Trabajo Final de Graduación que se entrega a la Coordinación del área académica de Administración de Tecnología de Información. El informe final se compone del Capítulo I, marco teórico, marco metodológico, análisis de resultados, propuesta de solución, conclusiones y recomendaciones del proyecto.

1.8. Limitaciones

A continuación, se expone el listado de limitaciones que tendrá el proyecto durante su ejecución:

-  Ley 8968 – Protección de la persona frente al tratamiento de sus datos personales.
-  Definiciones o configuración faltantes o erróneas en las fuentes transaccionales.
-  Desviaciones o errores en los datos que se ingresan en los sistemas transaccionales.
-  Compatibilidad entre componentes legados que gestiona la organización y los componentes propuestos en el proyecto.

2

Marco teórico

El propósito del presente capítulo es definir el fundamento teórico del proyecto mediante la presentación ordenada de las principales teorías en el área de inteligencia de negocio.

El capítulo inicia explorando el concepto de inteligencia de negocio, seguidamente, se establece la terminología asociada al *Data warehouse* (DW) y el *Data mart* (DM), los modelos multidimensionales, procesos de ETL, metadatos y las visualizaciones de Información. Las secciones finales del capítulo se enfocan en la descripción de algunos métodos para la implementación de proyectos enfocados en inteligencia de negocio, los lineamientos recomendados para la evaluación de proyectos de este tipo y un resumen de mejores prácticas a tomar en cuenta para el diseño multidimensional y del proceso de ETL.

2.1. Inteligencia de negocio (IN)

La inteligencia de negocio es una “Disciplina que se enfoca en la aceleración de la toma de decisiones, combinando componentes de infraestructura tecnológica, procesos y aplicaciones, con el objetivo de identificar, calcular y distribuir información relevante a las áreas de negocio. “(Turban, Aronson y Liang, 2005, P. 130).

La definición anterior establece que la inteligencia de negocio integra elementos técnicos y de gestión organizacional con la premisa de generar activos de información destinados al personal de las áreas de negocio. La relevancia de estos activos nace del rol que desempeñan en la toma de decisiones organizacional y de su impacto en los procesos departamentales.

Los objetivos de la IN que complementan la definición anterior según Kimball y Ross (2013) son:

Accesibilidad de la información: la información que provee la IN debe ser comprensible y accesible para los usuarios de negocio, no solamente para los desarrolladores de sus aplicaciones o sistemas.

Ofrecer la información oportunamente: la IN debe convertir datos en información valiosa en cuestión de horas, minutos o segundos. Es imprescindible que los usuarios de negocio y los desarrolladores concilien sus expectativas sobre los tiempos de entrega de información.

Fundamentar la toma de decisiones basadas en información de confianza: las soluciones de inteligencia de negocio deben asegurar la entrega de información correcta y consistente que respalde los procesos de toma de decisiones.

Para cumplir los objetivos de la IN se ha desarrollado el concepto de arquitectura de inteligencia de negocio, a continuación, se describe este término y sus componentes.

2.2. Arquitectura de inteligencia de negocio

La arquitectura de IN se define como la organización y vínculo de componentes de gestión, información y herramientas tecnológicas, así como las relaciones que existen entre ellos (Shariat, Hightower, 2007). La integración de los componentes en una arquitectura de inteligencia de negocios fundamenta la ejecución de los procesos dedicados a la entrega de datos. La Figura 2.1, muestra los componentes generales en una arquitectura de inteligencia de negocios que brinda soporte a la toma de decisiones.

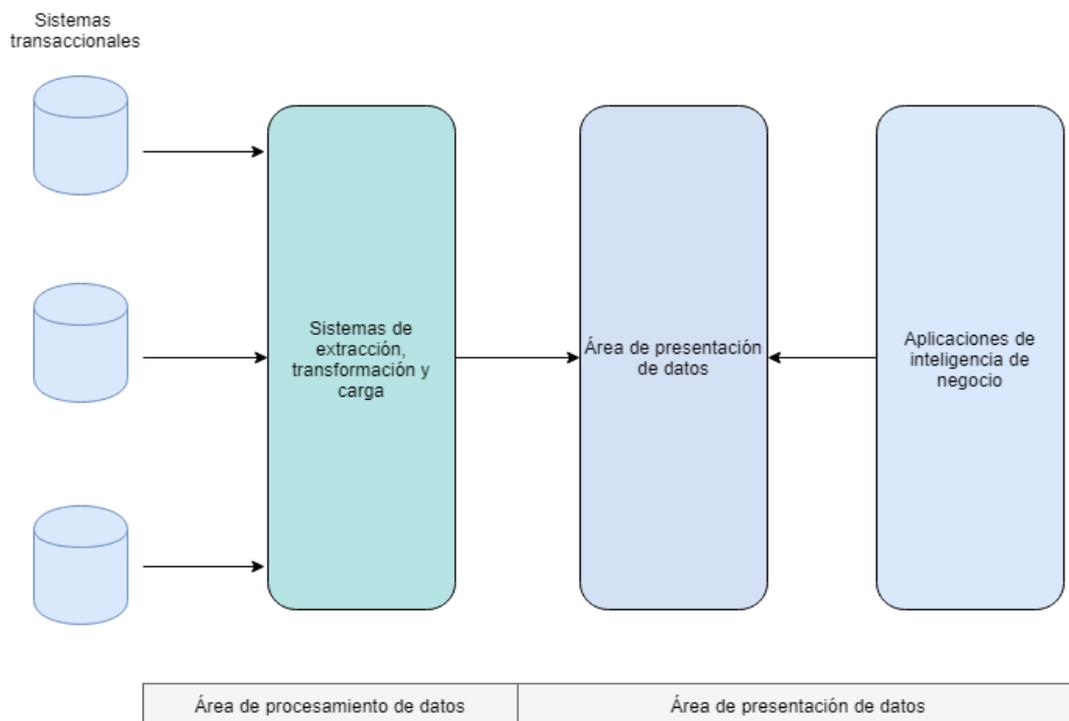


Figura 2.1. Arquitectura estándar de inteligencia de negocio.

Nota: Adaptado de Kimball y Ross (2013).

Las funciones principales que desempeñan los componentes de las arquitecturas de IN son:

Sistemas transaccionales o fuente

Son los sistemas encargados de registrar las operaciones resultantes del giro de negocio que posee la organización, por ende, estos sistemas representan el punto de inicio para el flujo de datos en una arquitectura de IN. Los sistemas transaccionales por lo general no mantienen un histórico de la información procesada.

Sistemas de extracción transformación y carga

El propósito de este componente es procesar la información que proviene de los sistemas transaccionales, mediante la estandarización y limpieza de los registros que se incorporarán en los repositorios históricos. Las funciones que realizan estos sistemas durante las tareas de extracción, transformación y carga de datos se mencionan en la sección 2.4.6 de este documento.

Área de presentación de datos

El área de presentación es el componente donde se organizan y almacenan los datos para su posterior consulta por parte de los usuarios finales. Aquí se almacenan datos con el nivel de detalle requerido para dar solución a las consultas que plantean los usuarios de forma dinámica.

Aplicaciones de inteligencia de negocios

Son las aplicaciones encargadas de consultar la información almacenada en el área de presentación, su objetivo es ofrecer a los usuarios información certera que apoye el proceso de toma de decisiones.

Las herramientas incluidas en este componente varían en cuanto a las funcionalidades que proveen al usuario, algunas aplicaciones integran capacidades para crear consultas personalizadas directamente en el modelo de datos, otras aplicaciones proveen consultas predefinidas ofreciendo una guía al usuario durante el proceso de descubrimiento de datos. Finalmente, existen aplicativos dedicados a la visualización de datos, que toman los registros almacenados en los repositorios de información y los presentan gráficamente a los usuarios finales.

Uno de los elementos principales que alimentan las aplicaciones de inteligencia de negocios son los repositorios de información histórica o *Data warehouse* ubicados en el área de presentación de datos. Estos almacenes de información proveen los insumos para los análisis vinculados a la toma de decisiones en todos los niveles de la organización.

Para comprender la función que desempeñan estos repositorios en el proceso de toma de decisiones es necesario establecer su definición, características y diferencias con respecto a los sistemas transaccionales.

2.3. *Data warehouse* y *Data mart*

Los DW se pueden definir como bases de datos dedicadas a la recolección, almacenamiento y consulta de datos históricos a nivel organizacional. Este tipo de repositorios integra información de múltiples fuentes o procesos de negocio, estableciendo una versión íntegra y consistente de los datos que procesan los sistemas transaccionales.

El-Sappagh, Hendawi, y Bastawissy (2011) establecen que el DW es una colección de tecnologías que habilita la toma de decisiones basadas en información histórica. Otros autores complementan esta definición al mencionar que los DW se caracterizan por su capacidad para mantener datos históricos, carencia de volatilidad en la información y la falta de normalización en sus estructuras (Turban et al., 2006, P. 236).

La carencia de volatilidad en los datos del DW juega un rol fundamental en el proceso de toma de decisiones, ya que permite la ejecución de análisis históricos al mantener un esquema de actualización nulo o casi nulo. Los análisis históricos no se podrían ejecutar si los datos almacenados en el DW se sobrescriben con información actualizada periódicamente. (Nagabhushana, 2006).

Los DW ofrecen capacidades que no se encuentran disponibles en sistemas transaccionales, particularmente se destaca su capacidad para almacenar grandes cantidades de información histórica. La Tabla 2.1, resume otras diferencias que existen entre los almacenes de datos históricos y sus contrapartes transaccionales.

Tabla 2.1.

Diferencias entre sistemas transaccionales y los DW.

DW	Sistemas transaccionales
Gestiona grandes cantidades de información no volátil.	Gestiona Información volátil.
Alcance organizacional.	Dedicados generalmente a uno o dos procesos de negocio.
Se enfocan en dar soporte a la toma de decisiones.	Dan soporte a la ejecución de los procesos de negocio diarios, tales como el proceso de venta.
Almacenan información procesada mediante reglas de negocio.	Registran información transaccional de manera diaria.

Nota: elaboración propia basado Kimball y Ross (2013).

En principio los DW representan un desafío para el equipo encargado de su diseño e implementación, ya que la premisa de integrar datos almacenados en fuentes heterogéneas puede presentar una complejidad mayor que las implementaciones de sistemas que poseen una fuente de datos única. Algunas organizaciones optan por implementar almacenes de información de menor escala enfocados en un proceso de negocio específico antes de aventurarse en el desarrollo de proyectos complejos de *data warehousing*. El resultado de este tipo de proyectos se conoce como *Data mart* (DM).

El DM se puede definir como una especialización del DW, cuyo objetivo es satisfacer las necesidades de información en un departamento o área de negocio en particular. Generalmente los repositorios de información categorizados como DM incluyen subconjuntos de datos cuya fuente son los sistemas transaccionales que utilizan las áreas de negocio diariamente (IBM, 2005).

Algunas características que poseen los *Data mart* según Ponniah (2001) son:

- ☐ Poseen alcance departamental.
- ☐ Abarcan un proceso de negocio únicamente.
- ☐ Optimizan el acceso a los datos y su análisis.
- ☐ Se encuentran estructurados para adaptarse a la forma en la que los departamentos de negocio visualizan la información.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Los DM facilitan las labores analíticas que realizan los departamentos organizacionales al proveer una versión íntegra y estructurada de la información.

Los DM son estructuras de almacenamiento que generalmente se integran a la arquitectura de otras soluciones de inteligencia de negocio de mayor escala como el DW. A continuación, se tipifican los DM según las relaciones que poseen con otros elementos en una arquitectura de inteligencia de negocio, tales como los sistemas de extracción, transformación y carga, áreas de presentación de datos y las aplicaciones de inteligencia de negocios.

Tipos de *data marts*

IBM (2005) establece que un DM puede clasificarse tomando en cuenta el enlace que posee con el DW organizacional a nivel de arquitectura. La Figura 2.2, muestra un diagrama que incluye la relación entre las fuentes transaccionales, el repositorio principal y los DM correspondientes.

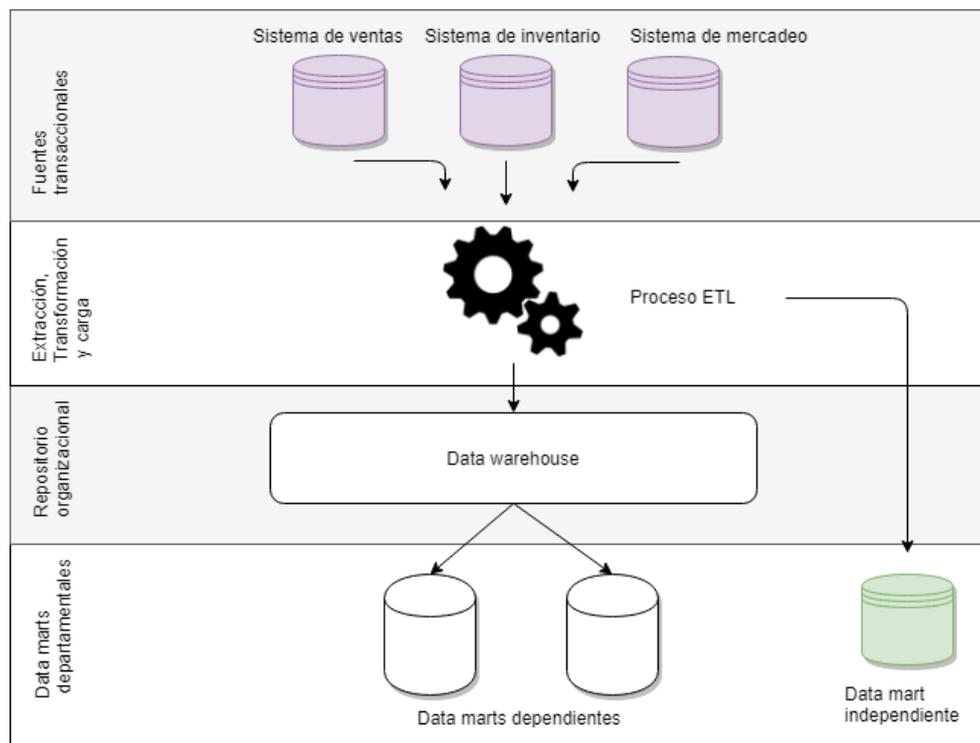


Figura 2.2. Arquitectura para la integración con Data marts.

Nota: Adaptado de IBM (2005)

Propuesta para la elaboración de reportes crediticios de una entidad financiera

En la clasificación que presenta IBM se mencionan dos tipos principales de DM: los repositorios dependientes y los independientes.

- **Data mart dependiente:** incluye información extraída directamente del DW, por lo tanto, sus datos son consistentes y se encuentran integrados con respecto al repositorio principal.
- **Data mart Independiente:** almacena información externa que no se encuentra disponible en el DW, esto implica que los datos incluidos en sus estructuras podrían no estar integrados o ser consistentes con el almacén de datos principal.

Según Kimball y Ross (2013), una arquitectura de inteligencia de negocio que posee DM independientes genera duplicidad en la información de análisis, ya que las áreas de negocio trabajan en la definición de sus requerimientos de manera separada, aplicando sus propias reglas de negocio y terminología sobre los datos.

Kimball y Ross (2013) establecen que al implementar soluciones de inteligencia de negocios con DM independientes se debe tomar en cuenta el siguiente conjunto de ventajas y desventajas.

Ventajas

- ❏ Rapidez de implementación con respecto al enfoque que integra los DM con el DW principal.
- ❏ Menor costo de implementación, al no tomar en cuenta tareas de integración con otros DM.
- ❏ Representa el camino de menor resistencia para implementaciones organizacionales, ya que se enfoca en las necesidades de un subgrupo de usuarios de negocio.

Desventajas

- ❏ No se comparte o integra información organizacional, cada DM funciona de manera aislada.
- ❏ Inconsistencia en la información, especialmente en casos donde se almacena información similar en repositorios separados.

- ❏ Genera esfuerzos para conciliar los datos cuando la información es inconsistente entre los repositorios.
- ❏ Incertidumbre sobre la calidad de datos debido a la duplicidad e inconsistencia de información.

Otra desventaja de implementar una solución bajo una arquitectura con *Data marts* independientes es que “contradice el principio básico de integración de datos en los proyectos de *Data warehousing*.” (Nagabhushana, 2006, P. 31), ya que algunos de esos repositorios podrían incluir información disonante con respecto a los datos almacenados en el DW.

Por otro lado, las soluciones que poseen *Data marts* dependientes, proveen gran valor al negocio ya que “sin importar cuantos repositorios de este tipo existan y lo heterogéneo de las tecnologías utilizadas durante su implementación, todos los usuarios tendrán acceso a una sola versión de los datos integrados” (Nagabhushana, 2006, P. 30). El mantener una sola versión consistente de los datos brinda a los usuarios de negocio claridad durante la toma de decisiones y provee una base sólida para futuros proyectos de inteligencia de negocio.

Los DM son repositorios de información que deben seguir un modelo de datos que permita organizar y relacionar los registros almacenados en sus estructuras. En el caso específico del DW o DM a este modelo de información se le conoce como modelo multidimensional. Los modelos de datos en un DW o DM poseen su propio conjunto de estructuras y técnicas de implementación. Por tanto, es necesario explicar en qué consisten estos elementos.

2.4. Modelos multidimensionales

En el núcleo del DM se encuentran los modelos multidimensionales, que son estructuras optimizadas para el procesamiento, almacenamiento y análisis de información integrales desde múltiples perspectivas (Reeves, 2009).

Las técnicas de modelado multidimensional implementan estructuras lógicas llamadas tablas hechos y tablas dimensionales, este tipo de estructuración permite organizar y representar los datos de manera sencilla y comprensible para los usuarios de negocio. Reeves (2009) establece que “el objetivo del modelado dimensional es hacer la creación de reportes y analítica de datos más intuitiva para los usuarios.” (P. 178).

Con el objetivo de comprender la función que desempeñan los modelos multidimensionales en los proyectos de *Data warehousing*, es necesario explorar con mayor detalle los conceptos de tablas de hechos y tablas dimensionales.

2.4.1. Tablas de hechos

Según Kimball y Ross (2013), los hechos son medidas resultantes de eventos en los procesos de negocio y generalmente representan variables numéricas. Nagabhushana (2006), establece que:

Una tabla de hechos se utiliza en los modelos multidimensionales para registrar las métricas o ítems numéricos que son de interés para el negocio. Ejemplos de hechos por industria son:

- Ventas al por menor: cantidad vendida y monto de ventas.
- Telecomunicaciones: duración de llamadas y cantidad promedio de llamadas.
- Banca: monto transaccional y balance diario promedio.

Kimball y Ross (2013) establecen que los hechos pueden agruparse en dos grandes categorías, hechos aditivos y semi-aditivos:

Hechos aditivos

Este tipo de medidas pueden ser agregadas o sumadas a través de cualquier dimensión asociada a la tabla de hechos. Hechos aditivos pueden ser cantidades vendidas, ingresos, costos y ganancias. Por ejemplo, si un usuario desea conocer las ganancias que han generado las ventas de resmas de papel en tres continentes distintos durante el último trimestre, es posible realizar la sumatoria de las ganancias a través de criterios temporales, de producto y localización, obteniendo resultados consistentes con respecto a las cantidades vendidas realmente.

Hechos semi-aditivos

Las medidas semi-aditivas pueden ser agregadas solamente en algunas dimensiones enlazadas a la tabla de hechos.

Algunas medidas semi-aditivas son los saldos de cuentas de ahorro y los niveles de inventario de una tienda, estos datos no se pueden sumar a través de la dimensión temporal ya que produciría resultados erróneos en los conteos de inventario o los balances de cuentas

bancarias. A manera de ejemplo, si un usuario desea conocer el saldo que mantuvo en su cuenta bancaria durante el transcurso de una semana, no se podrían agrupar los saldos registrados en su cuenta todos los días, ya que se estaría calculando un saldo mucho mayor al saldo real que posee esa persona en su cuenta.

Tablas de hechos agregadas

Ponniah (2001, P. 239), establece las tablas de hechos agregadas como tablas que contienen la suma de los hechos registrados con el mayor nivel de detalle. Esta variante de las tablas de hechos se utiliza para incrementar el desempeño de las consultas sobre el modelo dimensional al evitar el procesamiento de una gran cantidad de registros, especialmente sobre estructuras que poseen millones de datos. La Figura 2.3, muestra un ejemplo de una tabla de hechos agregada y una tabla de hechos normal. La tabla de hechos convencional almacena la información a un mayor nivel de detalle, mientras que la tabla agregada incluye la misma información, pero con un nivel de detalle mucho menor para optimizar la búsqueda de registros.

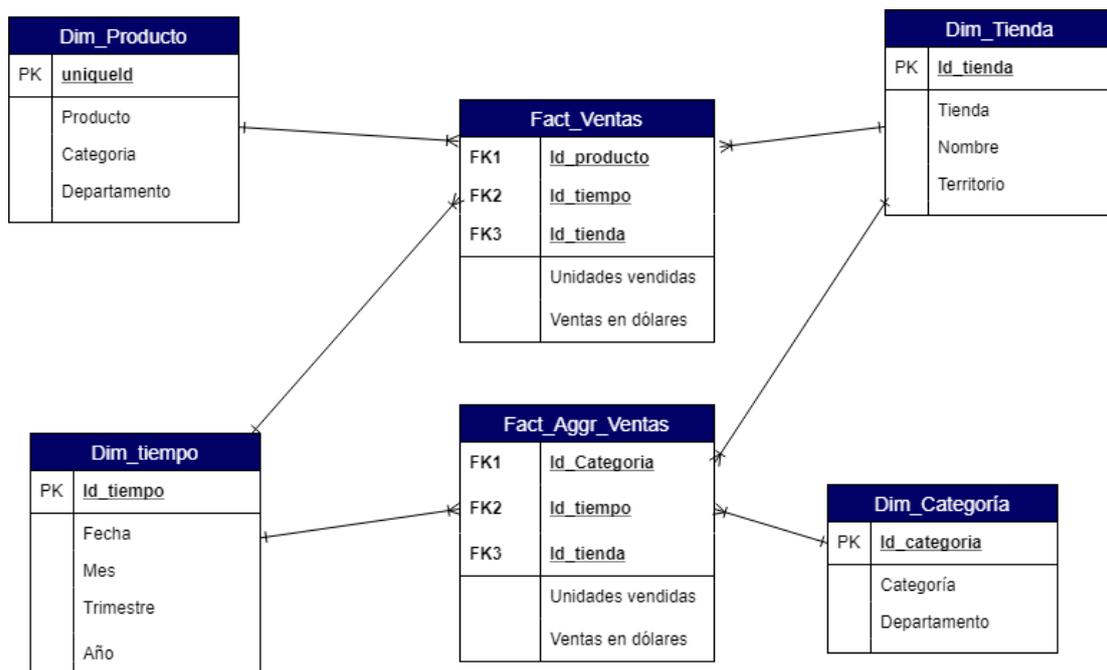


Figura 2.3. Tabla de hechos agregada.

Nota: Adaptado de Ponniah (2001).

Otro elemento fundamental en los modelos multidimensionales son las tablas de dimensión, la Figura 2.3, muestra cuatro tablas dimensionales enlazadas a las distintas tablas

de hechos. La relación que existe entre los dos tipos de tablas permite a los usuarios navegar hasta el nivel de detalle deseado en sus análisis, contextualizando la información cuantitativa de las tablas de hechos. A continuación, se explica brevemente el concepto de tabla dimensional.

2.4.2. Tablas dimensionales

Nagabhushana (2006), menciona que las dimensiones son aquellas tablas que establecen el contexto de las métricas en la tabla de hechos mediante datos categóricos o descriptivos, expresados generalmente por medio de atributos textuales. La Figura 2.4, muestra un ejemplo de una tabla dimensional y los elementos que la conforman.

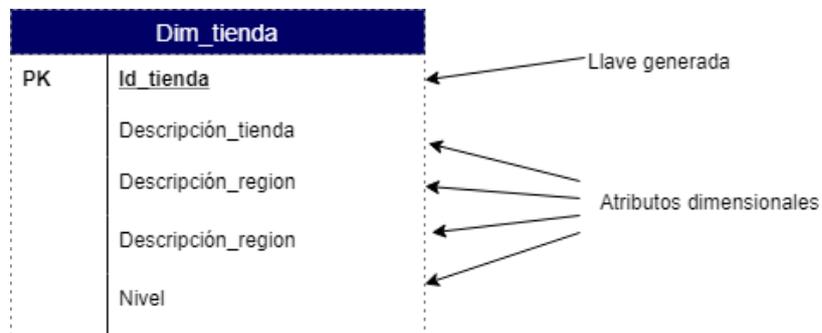


Figura 2.4. Tabla dimensional.

Nota: Adaptado de Ponniah (2001).

Las tablas de dimensión por lo general no se encuentran normalizadas, la carencia de normalización optimiza el desempeño de las consultas e incrementa la legibilidad de los modelos multidimensionales. Nagabhushana (2006) afirma que:

La falta de normalización en las tablas dimensiones implica que este tipo de tablas generalmente posee estructuras jerárquicas que agrupan los datos de manera lógica, emulando los agrupamientos que realizan los usuarios de negocio durante la ejecución de sus análisis (P.169).

Las jerarquías disponibles en las tablas dimensionales fortalecen las capacidades exploratorias de los usuarios de negocio, quienes por medio de estas estructuras pueden obtener

información específica para solventar sus interrogantes sobre los datos. Sobre las jerarquías en las tablas de dimensión Kimball y Ross (2013) establecen que:

Los usuarios de negocio utilizan las jerarquías disponibles en las tablas dimensionales para navegar a un mayor detalle en la información o bien cuando requieren visualizar los datos del modelo resumidos, mediante operaciones *Drill Down* o *Roll Up* respectivamente (P. 214).

Las tablas dimensionales poseen una llave primaria única que enlaza cada registro en la tabla dimensional con una fila o medida en la tabla de hechos.

Llaves subrogadas en dimensiones

La llave primaria en las tablas dimensionales no puede ser igual a la llave primaria que posee la información en los sistemas transaccionales, ya que con el paso del tiempo podrían existir múltiples filas en las tablas dimensionales que coincidan con la llave del sistema transaccional, provocando inconsistencias en los análisis finales. Para solucionar este problema los autores recomiendan crear llaves primarias numéricas para cada dimensión, utilizando números secuenciales (Kimball y Ross, 2013, P. 46).

Un concepto que se debe tomar en cuenta durante el diseño de las tablas de hechos o las tablas dimensionales es la granularidad de los datos. A continuación, se conceptualiza brevemente la granularidad de la información en los modelos multidimensionales.

2.4.3. Granularidad

La granularidad en los modelos multidimensionales representa el nivel de detalle utilizado para almacenar los datos en las estructuras del DW o DM.

Sobre la selección de la granularidad se destaca que “La granularidad debe establecerse antes de escoger las dimensiones o los hechos a incluir en el modelo, ya que todo candidato a tabla dimensional o tabla de hechos debe ser consistente con el nivel de granularidad escogido” (Kimball y Ross, 2013, P.39).

El identificar el nivel de granularidad es crucial para determinar la utilidad y costo de la solución.

Un nivel de granularidad demasiado alto limita la capacidad de los usuarios para obtener detalles sobre la información de análisis, en contraste la granularidad a niveles muy bajos puede provocar un incremento en los requisitos de almacenamiento en la solución (Nagabhushana, 2006, P.171).

Un ejemplo de granularidad de bajo nivel son las transacciones registradas por hora en los sistemas de venta de las cajas registradoras pertenecientes a una cadena de tiendas departamentales, ubicadas en la costa este de los Estados Unidos. Se podría determinar que la granularidad de los datos son las ventas por caja registradora, por tienda, por ubicación geográfica en periodos de una hora. Con este nivel de detalle es posible determinar cuáles cajas registradoras en tiendas específicas registran la mayor cantidad de ventas por región en el lapso de una hora.

Por otro lado, un nivel de granularidad alto para el mismo sistema se podría establecer como las transacciones de venta que registra una tienda específica por región geográfica en el transcurso de una semana.

Kimball y Ross (2013) recomiendan enfocarse inicialmente en el nivel más detallado de la información que generalmente se captura en los sistemas transaccionales, ya que mantener un grano fino en los datos permite responder con mayor flexibilidad a posibles requerimientos del usuario.

En resumen, un nivel de granularidad alto no permite que los usuarios naveguen hasta la información más detallada, en contraste, un nivel de granularidad bajo habilita las tareas exploratorias y el descubrimiento de patrones de negocio que podrían pasar desapercibidos en los análisis sobre información resumida.

El diseño de las tablas de dimensión, tablas de hechos y la definición de la granularidad de los datos, permite la construcción del modelo multidimensional que permite ejecutar las funciones del DW o *Data mart*. El uso de este tipo de modelos en las soluciones de *data warehousing* se propicia debido a las ventajas que ofrece al personal técnico y de negocio que interactúa con estos repositorios.

2.4.4. Ventajas de los modelos multidimensionales

Según Nagabhushana (2006), las ventajas de utilizar los modelos multidimensionales en el diseño de un DW o DM son:

Simplicidad del modelado: el modelo dimensional, permite que los diseñadores de la solución desarrollen esquemas de base de datos que los usuarios de negocio pueden comprender fácilmente. Debido a la simplicidad del modelo no se debe invertir en horas de capacitación para que los usuarios puedan leer y comprender los diagramas desarrollados, ya que no existen relaciones complejas entre los datos.

Promueve la calidad de datos: por medio de la integridad referencial entre las tablas de la base de datos multidimensional, los diseñadores del DW o DM pueden garantizar que no hay registros faltantes en la base de datos dimensional.

Optimización del desempeño en consultas: por medio de agregaciones los modelos multidimensionales pueden optimizar el desempeño de las consultas, mediante agrupaciones de información que reducen la carga del procesamiento sobre registros detallados.

Utiliza tecnología de bases de datos relacionales: al utilizar modelos multidimensionales implementados con tecnología relacional es posible sacar ventaja de la escalabilidad que poseen estos sistemas de bases de datos sin comprometer la usabilidad del DW o DM.

Quizás una de las mayores ventajas que poseen los modelos multidimensionales es su capacidad para solventar las necesidades de información de los usuarios mediante un esquema de datos simple. La representación de los modelos multidimensionales se da por medio de esquemas que organizan e interrelacionan las tablas dimensionales y las tablas de hechos. A continuación, se describen las variantes de esquematización para estas estructuras.

2.4.5. Esquemas del modelo multidimensional

En la teoría sobre bases de datos multidimensionales expuesta por los principales autores se pueden extraer dos variantes o tipos de esquema que facilitan la representación del diseño multidimensional de análisis. A continuación, se explican los dos tipos principales de esquema:

Esquema estrella

Este tipo de esquema se utiliza para representar modelos de información implementados en sistemas gestores de bases de datos relacionales, por la estructura que poseen se asemejan a la figura de una estrella (Kimball y Ross, 2013). Este tipo de esquemas utiliza las llaves primarias y foráneas características de los sistemas relacionales para establecer la conexión entre las tablas del modelo. La Figura 2.5, ejemplifica un esquema de tipo estrella donde se pueden apreciar los enlaces mencionados anteriormente.

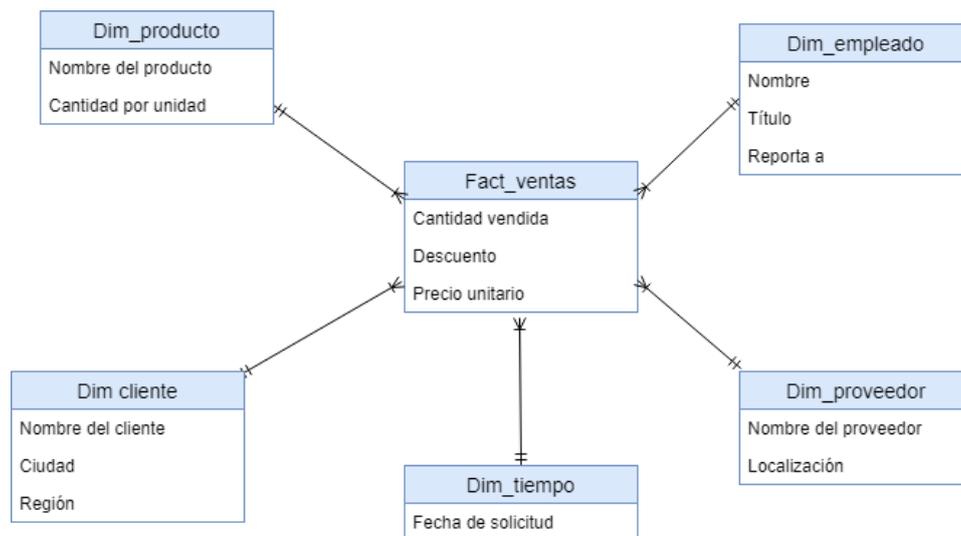


Figura 2.5. Esquema dimensional.

Nota: Adaptado de Zhang y Wu (2010).

Los esquemas de tipo estrella “se caracterizan por la ausencia de normalización en sus tablas dimensionales” (Ponniah, 2001, P. 220), almacenando grandes cantidades de información en cada tabla.

Ponniah (2001), expone además una serie de ventajas que se pueden obtener al implementar este esquema en un DW o DM, las aserciones principales que realiza el autor son:

- 📦 **Facilidad de comprensión:** por medio de la representación del esquema estrella, los usuarios pueden comprender la estructura que compone el DW o DM de manera rápida para formular sus consultas.
- 📦 **Optimiza la navegación:** debido a la simplicidad de los enlaces entre las tablas que componen el esquema estrella es posible navegar rápidamente por las estructuras de la solución, incrementando los tiempos de respuesta para los análisis.

- ❏ **Adecuado para el procesamiento de consultas:** debido a los mecanismos para ejecución de las consultas en el esquema estrella es posible obtener ventajas en los tiempos de procesamiento, al seleccionar primero los valores de filtro de las consultas en las tablas dimensionales y realizando búsquedas sobre la tabla de hechos donde únicamente se extraen los registros que coinciden con los conjuntos de datos filtrados en las tablas dimensionales.

Esquema copo de nieve

La segunda variante en esquemas dimensionales que predomina en los almacenes de datos analíticos es el copo de nieve. Los modelos basados en la estructura de copo nieve “se caracterizan por incluir al menos una tabla dimensional normalizada” (Ponniah, 2001, P. 235). La Figura 2.6, muestra un esquema parcialmente normalizado que subdivide la dimensión producto en una tabla para marcas y otra para categorías de producto.

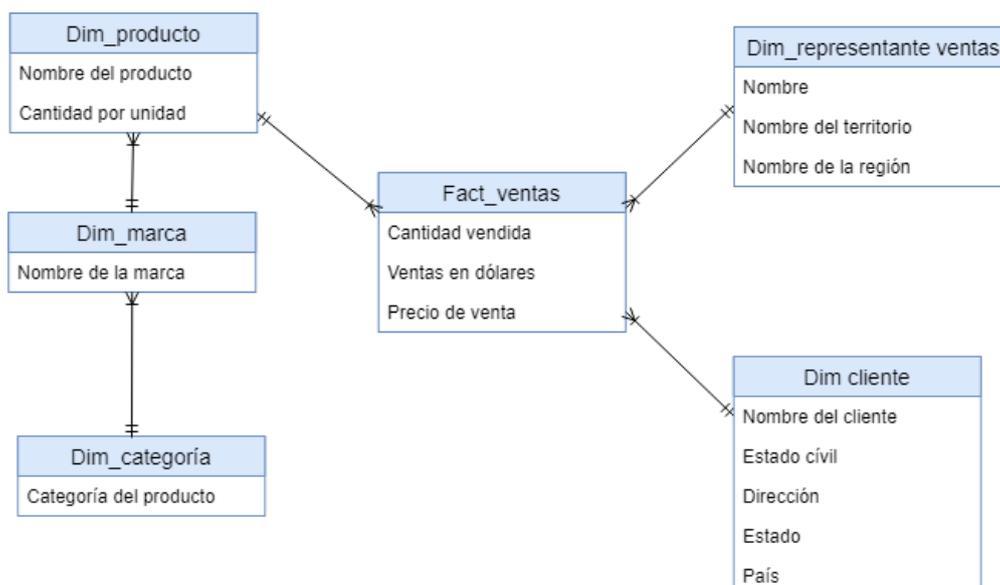


Figura 2.6. Esquema copo de nieve.

Nota: Adaptado de Ponniah (2001).

Ponniah (2001), expone algunas ventajas y desventajas de adaptar el esquema copo de nieve en el diseño de un DW o DM, en las afirmaciones realizadas por este autor se destacan las siguientes ventajas:

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ Ahorros en espacio de almacenamiento, al reducir atributos con baja cardinalidad de las tablas dimensionales.
- ❏ Facilidad para dar mantenimiento y actualizar las estructuras normalizadas.

Adicionalmente, Ponniah (2001) señala las siguientes desventajas que se generan al implementar un esquema de este tipo:

- ❏ El esquema es menos intuitivo para los usuarios.
- ❏ La navegación entre los ítems de información se vuelve compleja.
- ❏ Disminución en el desempeño de las consultas debido a las uniones adicionales entre tablas del modelo.

Kimball y Ross (2013) respaldan las aseveraciones descritas anteriormente al mencionar que “normalizar atributos con baja cardinalidad en tablas secundarias enlazas a la tabla dimensional contribuye a representar con precisión los datos, pero compromete el desempeño de las consultas” (P. 50).

La selección del esquema depende de los requisitos de la solución, por tanto, no es posible establecer cuál enfoque es el óptimo de forma general. Una vez seleccionado el esquema a implementar en el proyecto se procede a elegir el método de implementación correspondiente que permita obtener los resultados esperados por los usuarios de negocio.

2.4.6. Implementaciones físicas del modelo multidimensional

Las implementaciones físicas de los modelos multidimensionales pueden clasificarse según las tecnologías utilizadas para su desarrollo en los ambientes empresariales. Kimball y Ross (2013), mencionan que los modelos multidimensionales pueden implementarse mediante esquemas estrella en sistemas relacionales de base de datos o por medio de cubos OLAP (*Online Analytical Processing*).

Esquemas estrella

Los esquemas estrellas “son estructuras dimensionales implementadas en un sistema gestor de bases de datos relacional, principalmente consiste en tablas de hechos enlazadas con tablas dimensionales por medio de llave primarias y foráneas” (Kimball y Ross, 2013, P. 8).

Los esquemas estrella implementados en sistemas relacionales toman ventaja de la optimización de consultas y escalabilidad inherente en las bases relacionales de información.

Cubos OLAP

Son estructuras dimensionales implementadas en una base de datos multidimensional, y pueden ser equivalentes o derivadas de un esquema estrella. Un cubo OLAP contiene hechos y dimensiones, pero su información solamente se puede acceder mediante XMLA (*Extensible Markup Language for Analysis*) o MDX (*MultiDimensional Expressions*) (Kimball y Ross, 2013). La Figura 2.7, muestra la representación gráfica de un cubo OLAP que posee tres dimensiones de análisis.

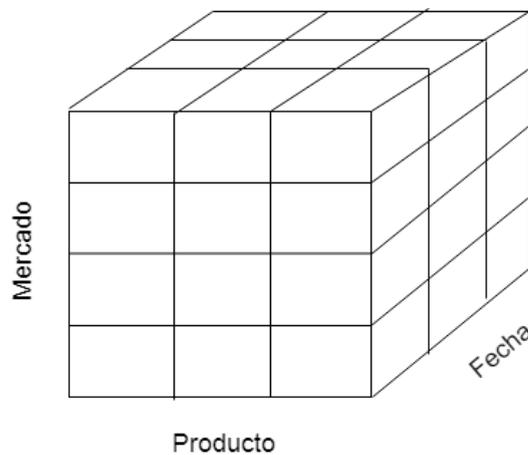


Figura 2.7. Cubo OLAP.

Nota: Adaptado de Kimball y Ross (2013)

Una vez establecidas las posibles representaciones físicas del modelo multidimensional, es necesario explorar los mecanismos que incorporarán la información en las estructuras del DM. El proceso encargado de realizar la transformación de datos y su posterior almacenamiento en el repositorio multidimensional se conoce como proceso de Extracción, Transformación y Carga (ETL por sus siglas en inglés).

2.5. Proceso de Extracción, Transformación y Carga

El objetivo del proceso de ETL es orquestar la ejecución de las tareas para el procesamiento de la información disponible en sistemas transaccionales para su posterior almacenamiento en las estructuras del DW o DM (Diouf, Boly y Ndiaye, 2018). Por medio de los procesos de ETL, también es posible “integrar de forma periódica información que no se

encuentra almacenada en sistemas de procesamiento transaccional, tales como archivos de texto, hojas de cálculo o sistemas legados “(Anand y Kumar, 2013. P.1).

Los procesos de ETL resultan particularmente útiles para reducir la carga de procesamiento manual sobre los datos, adicionalmente ofrecen la oportunidad de estandarizar los mecanismos de extracción, conversión y carga sobre la información.

Los procesos de ETL incrementan su complejidad y alcance cuando se deben integrar múltiples fuentes de datos al repositorio histórico de información, ya que durante la ejecución de este proceso se debe asegurar que se tomaron en cuenta todas las reglas de negocio y limitaciones que poseen las plataformas fuente. Sobre la complejidad que genera la integración de fuentes heterogéneas de conocimiento autores como Ponniah (2001) afirman que “incorporar datos de múltiples fuentes de información representa un desafío para el equipo que diseña el proceso, ya que los datos se deben estandarizar, convertir y formatear adecuadamente para su eventual consulta y análisis por parte de los usuarios “(P. 31).

Anand y Kumar (2013), mencionan otras funcionalidades a incluir en el proceso de ETL tales como:

- ❏ Identificar información relevante en las fuentes de datos.
- ❏ Transportar la información al área de preparación de datos o *staging area*.
- ❏ Cargar la información periódicamente a los repositorios de datos correspondientes (*Data Warehouse* o *Data Mart*).
- ❏ Limpiar e integrar los datos.
- ❏ Crear hechos o campos calculados.

Debido a la variedad de funciones que debe ejecutar el proceso de ETL, se debe establecer un nexo de comunicación consistente entre los involucrados durante el diseño e implementación de este componente. La comunicación estrecha entre miembros del equipo desarrollador y los usuarios de negocio minimiza la cantidad de iteraciones que se deben realizar sobre el diseño.

Uno de los primeros pasos en la construcción de un proceso de ETL es el diseño del modelo lógico del proceso. Kimball y Caserta (2004) mencionan que la información especificada en este modelo guiará el proceso de implementación de las tareas de extracción,

Propuesta para la elaboración de reportes crediticios de una entidad financiera

transformación y carga. Durante el diseño del proceso se deben tomar en cuenta todas las reglas de transformaciones por implementar y la documentación de los esquemas resultantes en el repositorio de metadatos (Anand y Kumar, 2013).

Kimball y Caserta (2004), establecen además que el modelo lógico del proceso ETL generalmente se representa mediante una tabla que incluye:

- ✓ Nombre de la tabla objetivo en el DW o DM.
- ✓ Nombre de la columna objetivo.
- ✓ Tipo de tabla (hechos, dimensión o subdimensión).
- ✓ Fuente de los datos: nombre de la base de datos donde reside la información, nombre del archivo o fichero.
- ✓ Nombre de la tabla fuente: cuando se tiene más de una tabla que se utiliza para llenar la tabla del DW se deben listar todas las tablas necesarias.
- ✓ Nombre de la columna fuente: listar todas las columnas requeridas para cargar la columna objetivo.
- ✓ La manipulación exacta requerida para que los datos en las fuentes de datos cumplan el formato adecuado en el almacén objetivo. Normalmente las transformaciones se documentan mediante sentencias de SQL o Pseudo-código. Cuando el campo de transformaciones aparece en blanco en el modelo lógico se interpreta que únicamente se realizará una carga directa de información desde la fuente.

Al establecer el modelo lógico del proceso se puede iniciar con el desarrollo o implementación del proceso en cada una de las actividades principales que componen un ETL.

Los procesos de ETL están compuestos por tres actividades fundamentales: la extracción, transformación y carga. Cada actividad se divide en funciones cuyo objetivo es llevar a cabo el procesamiento de datos (Ponniah, 2001). A continuación, se detallan las funciones que integran el proceso.

2.5.1. Funciones de extracción

Estas funciones son las encargadas de extraer los datos en las fuentes transaccionales, sistemas legados o archivos planos para su procesamiento en actividades subsecuentes del proceso ETL.

Ponniah (2001) establece que: “con frecuencia las soluciones de inteligencia de negocios utilizan un área intermedia dedicada a la preparación de datos, las funciones de extracción son las responsables de trasladar la información al repositorio intermedio de datos “(Ponniah, 2001, P. 31). El área de preparación de datos o *staging area* ofrece la posibilidad de mantener un respaldo de los datos procesados en un paso intermedio. Si el proceso falla o se detiene este respaldo permite reanudar la ejecución del ETL a partir de los datos transformados y evita la ejecución completa del proceso.

Según Ponniah (2001) entre las responsabilidades que deben cumplir las funciones de extracción se destaca el siguiente conjunto de tareas:

- **Identificación de fuentes:** se identifican las aplicaciones fuente y sus estructuras de datos.
- **Métodos de extracción:** para cada fuente se define si la extracción será manual o por medio de herramientas.
- **Frecuencias de extracción y ventanas de tiempo:** por cada fuente de información se debe establecer la frecuencia para la extracción de datos y la ventana de tiempo para realizar la extracción.
- **Secuencia de Jobs:** determinar si los *Jobs* o tareas programadas deben ejecutarse de manera secuencial o si algunos pueden ser de ejecución paralela.
- **Manejo de excepciones:** se debe establecer cómo se deben manejar los registros que no pueden extraerse de las fuentes.

Las funciones de extracción también son las encargadas de detectar anomalías en la información, por ejemplo, se realiza la identificación de valores nulos y se lleva a cabo el reemplazo de los registros con un valor por defecto cuando sea posible. Otras anomalías que se deben tomar en cuenta en las funciones de extracción son las fechas almacenadas en campos que no fueron diseñados para almacenar los tipos de datos (Kimball y Ross, 2013).

2.5.2. Funciones de transformación

Según Ponniah (2001), las funciones de extracción se encargan de preparar los datos para su posterior almacenamiento en el repositorio dimensional, algunas tareas que realizan las funciones de transformación son: limpieza de datos, estandarización de tipos y ajustes a longitudes de los campos que se incluirán en el modelo dimensional.

Algunas tareas que se ejecutan durante la transformación de datos son:

- ❏ Estandarización semántica, que se realiza cuando se tienen dos o más términos que significan lo mismo (sinónimos) o cuando un solo término posee distintos significados en sistemas independientes (homónimos).
- ❏ Asignación de llaves subrogadas que se derivan de llaves primarias del sistema transaccional de donde proceden.
- ❏ Agregaciones de datos numéricos también tienen lugar en las funciones de transformación.
- ❏ Separación o unión de campos: en esta tarea se dividen o unen los campos cuando sea necesario.
- ❏ Conversión de datos: las conversiones de datos cumplen dos objetivos principales, estandarizar valores recolectados de múltiples fuentes y contextualizar la información para incrementar su legibilidad de cara a los usuarios.
- ❏ Agrupación: estas funciones toman los valores que se encuentran en un mayor nivel de detalle en las fuentes de datos y los agrupan para crear información consolidada.
- ❏ Enriquecimiento: estas funciones reorganizan y simplifican los atributos individuales para incrementar su utilidad para los usuarios, otro ejemplo es la consolidación de campos de múltiples fuentes transaccionales.

Según Ponniah (2001), las funciones de transformación de datos pueden clasificarse en:

Revisión de formatos: las revisiones incluyen verificación de los tipos de datos y longitud de los atributos.

Decodificación de atributos: estas funciones se encargan de transformar los atributos de los sistemas transaccionales que poseen valores poco representativos en información comprensible

Propuesta para la elaboración de reportes crediticios de una entidad financiera

para los usuarios finales, por ejemplo, se transforman los valores 1 y 2 que representan el sexo femenino o masculino en un sistema transaccional a los datos categóricos Masculino y Femenino para su correcto almacenamiento en el DW.

Cálculo de campos derivados: estas funciones se encargan de calcular atributos que no se encuentran disponibles en las fuentes de datos. Los campos calculados son el resultado de tomar los registros de las fuentes transaccionales y aplicar reglas de negocio para obtener nuevos valores numéricos.

Conversión de unidades de medida: estas funciones se encargan de estandarizar las unidades de medida de las fuentes de datos cuando se requiere consolidar información de sistemas con diferentes estándares de medición.

Conversión de fechas/tiempo: estas funciones se encargan de estandarizar la representación de los atributos de tiempo y fecha en un formato común.

Reestructuración de llaves: esta función es la encargada de asignar las llaves primarias a los registros del DW.

Eliminación de registros duplicados: esta función consolida información duplicada en los sistemas transaccionales en un solo registro a nivel del DW.

2.5.3. Funciones de carga

Según Ponniah (2001), en las funciones de carga existen dos grupos distintos de tareas. Las tareas de carga inicial y los procesos de actualización incremental que se ejecutan sobre la información cuando se requiere actualizar el repositorio o incluir nuevos datos en sus estructuras. La Figura 2.8, ejemplifica los tipos de carga que se pueden realizar en un esquema de DW.

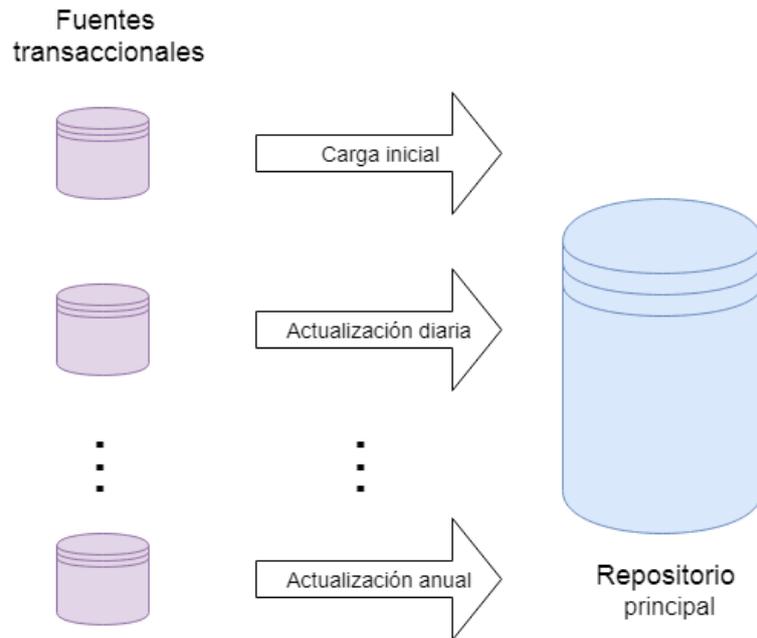


Figura 2.8. Tipos de carga.

Nota: Adaptado de Ponniah (2001).

Las cargas incrementales de información permiten agilizar el procesamiento de datos al tomar en cuenta únicamente los datos nuevos o modificados en las fuentes transaccionales durante las cargas al repositorio principal o DW.

2.5.4. Enfoque ECL-TL

Pan, Zhang y Qin (2018) proponen una variante para el proceso de ETL tradicional, en la propuesta se subdivide la actividad de extracción en tareas de extracción, limpieza y carga. La Figura 2.9, ejemplifica el flujo de actividades y los componentes del enfoque Extracción, Limpieza, Carga – Transformación y Carga (ECL – TL por sus siglas en inglés).

Pan et al. (2018) mencionan las siguientes ventajas de implementar el enfoque ECL-TL:

1. Eficiencia y estabilidad al separar los componentes de limpieza y extracción de datos.
2. Toda la información relacionada al sistema de negocio se encuentra integrada.
3. Provee una mayor flexibilidad a las tareas de extracción para gestionar requerimientos complejos de negocio.

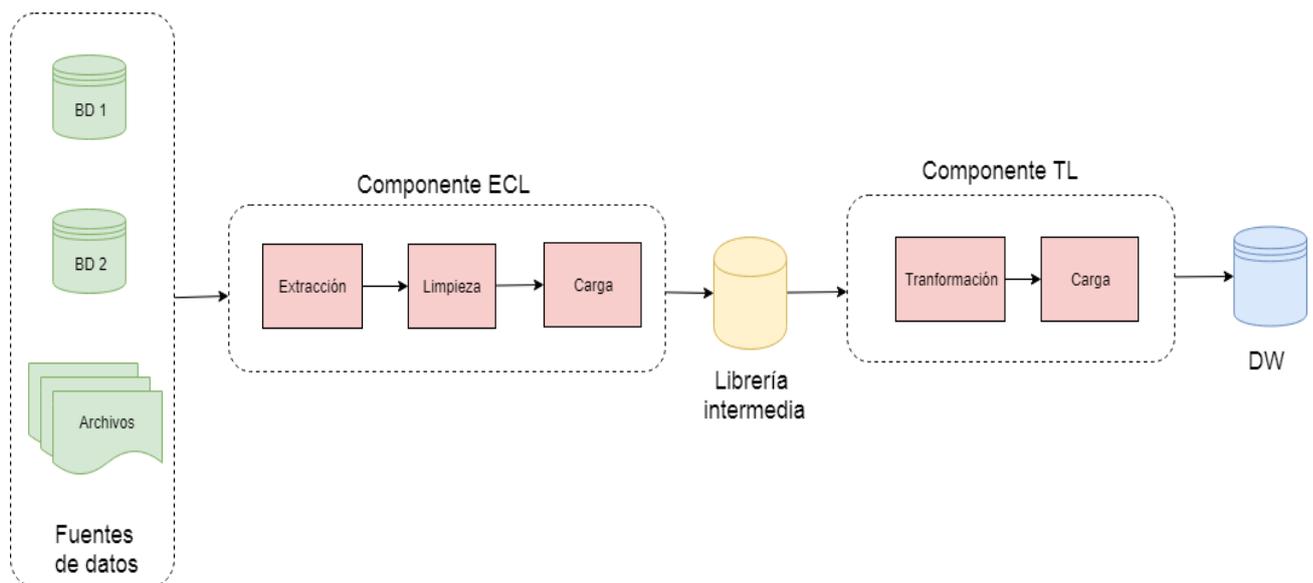


Figura 2.9. Proceso ECL-TL.

Nota: Adaptado de Pan, Zhang y Qin (2018)

A continuación, se explican los componentes del proceso de ECL-TL representados en la Figura 2.9.

Componente de ECL

Pan et al. (2018) establecen que el propósito principal de este componente es limpiar la información que proviene de las fuentes de datos. Las actividades que integran el componente de ECL son:

- ❏ Extracción: obtiene la información almacenada en las fuentes transaccionales tales como bases de datos o archivos.
- ❏ Limpieza: estandariza, limpia y ordena los datos de las fuentes transaccionales según las reglas de negocio de los usuarios.
- ❏ Carga: es la actividad encargada de construir la base de información que será trasladada al siguiente componente del proceso que es la librería intermedia.

Librería intermedia

Pan et al. (2018), establecen este componente como el núcleo del método ECL-TL. El propósito de la librería intermedia es proveer información consistente al componente de TL, de

esta forma los errores en la información de los sistemas transaccionales serán corregidos antes de iniciar las tareas de transformación.

La actualización de datos es la clave para la construcción de la librería intermedia, ya que los métodos de actualización son los encargados de refrescar los registros en este componente cuando se realizan cambios en las fuentes transaccionales. La actualización de la librería intermedia se realiza por medio de cargas completas o cargas incrementales según se requiera (Pan et al, 2018).

Componente de TL

Este componente recibe como insumo la información almacenada en la librería intermedia y la transforma en conocimiento que apoya la toma de decisiones por medio de algoritmos y reglas de transformación definidas por los usuarios de negocio (Pan et al, 2018).

Adicionalmente, el componente de TL se encarga del traslado de información a las estructuras del DW. Durante la ejecución de las actividades de este componente se realizan las agregaciones o separaciones de datos que permiten ajustar el nivel de granularidad requerido para almacenar la información en el DW.

Durante la ejecución de los procesos de extracción, transformación y carga se genera información relacionada a su desempeño y procesamiento. A este tipo de información se le conoce como metadatos y no se encuentran limitados únicamente a las tareas del ETL, por el contrario, en las soluciones de *Data warehousing* los metadatos juegan un rol fundamental en los procesos de diseño, implementación y puesta en marcha del proyecto.

2.6. Metadatos

En la literatura relacionada a los DW o DM se hace referencia al concepto de Metadatos como el componente del DW que describe los registros o datos integrados en los repositorios de información. Una definición formal puede extraerse de la teoría planteada por Ponniah (2001) quien menciona que:

Los metadatos proveen una descripción completa y precisa de los aspectos pertinentes a la información del DW, la información en los metadatos atañe a usuarios finales y a los

Propuesta para la elaboración de reportes crediticios de una entidad financiera

miembros del equipo del proyecto, ya que contextualizan los elementos incorporados en el repositorio de datos (P. 174).

De la afirmación anterior se puede establecer que los metadatos representan un conglomerado de información descriptiva del repositorio de información al que hacen referencia.

Kimball y Caserta (2004) realizan su aporte al concepto de metadatos al expresar que este componente del DW integra:

El linaje de datos, este elemento traza la información del DW desde la localización exacta en la que se extrajo la información del sistema fuente y documenta con precisión las transformaciones realizadas antes de su carga final en el repositorio de datos. (P. 351).

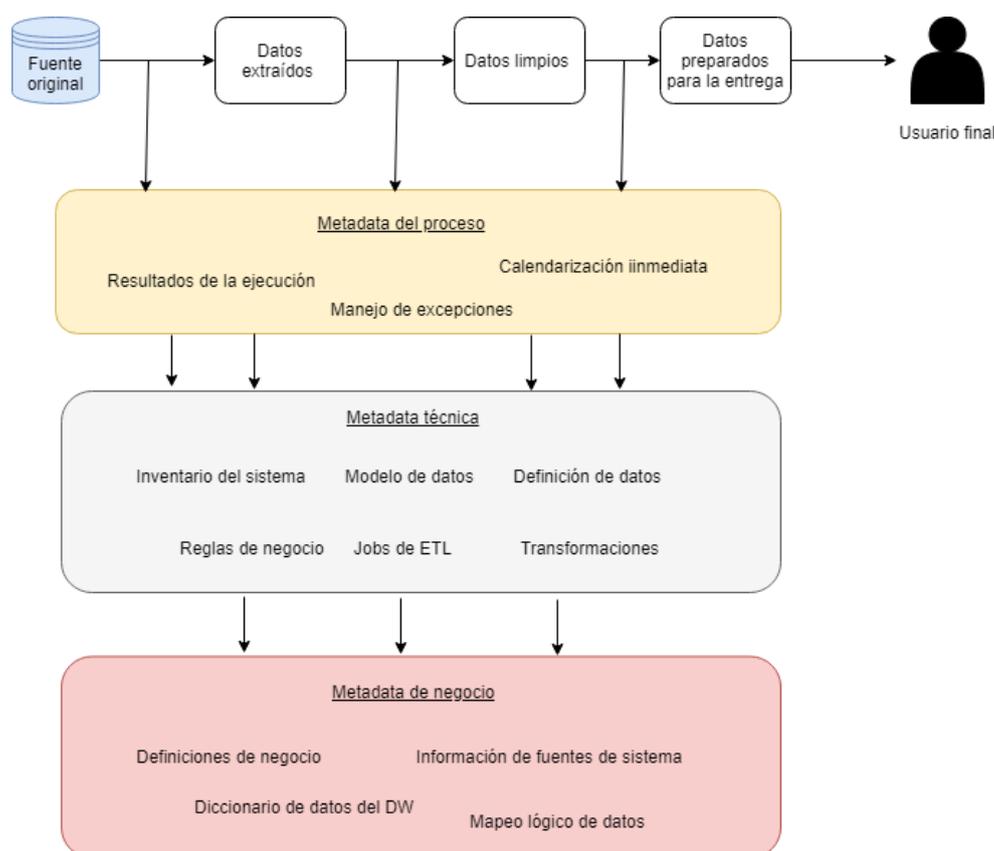


Figura 2.10. Flujo de metadatos en el proceso de ETL.

Nota: Adaptado de Kimball y Caserta (2004).

La Figura 2.10, muestra las relaciones entre los tipos de metadatos y el proceso de transformación en un proyecto de inteligencia de negocios enfocado en un DW.

La metadata técnica o de bajo nivel ayuda al gestor de base de datos a traer la información al DW, incluso ayuda a los usuarios de negocio a conocer de donde proviene la información, ejemplos de metadatos técnicos son los modelos de datos y las reglas de negocio asociadas a las transformaciones ejecutadas en los procesos de ETL. Los metadatos de alto nivel son de mayor provecho para los usuarios finales ya que contienen un diccionario del contenido de negocio que representa los elementos de datos (Kimball y Caserta, 2004, P. 353), un ejemplo de metadatos de alto nivel son los diccionarios de datos que se incluyen en el DW.

A continuación, se realiza una breve descripción de los metadatos técnicos y de negocio que integran este componente del DW.

Metadatos de negocio

Conectan a los usuarios de negocio con el DW, permitiéndoles conocer qué elementos se encuentran disponibles en el DW desde una perspectiva no técnica. Los metadatos de negocio representan una vista guiada para los ejecutivos de negocio y una hoja de ruta para los analistas de negocio (Kimball y Ross, 2013, P. 187).

Según Kimball y Ross (2013), los metadatos de negocio benefician a los analistas de negocio, usuarios regulares y casuales del DW, administradores experimentados y ejecutivos. Algunos ejemplos de metadatos de negocio son:

- ❖ Estructura general de los datos expresados en términos de negocio.
- ❖ Reglas de negocio para la transformación de datos.
- ❖ Nombres de tablas y definiciones de negocio.
- ❖ Reportes y consultas predefinidas.
- ❖ Cronograma de actualización del DW.

Metadatos técnicos

Los metadatos técnicos están dirigidos al personal de TI responsable de la gestión y desarrollo del DW, el personal técnico requiere información que guíe el diseño de cada proceso. Los metadatos representan una guía de soporte para los profesionales de TI para mantener, construir y gestionar el DW (Kimball y Ross, 2013, P. 190).

Según Kimball y Ross (2013), los metadatos técnicos benefician al administrador del DW, al administrador del proyecto, al analista de calidad de datos, al arquitecto de seguridad y al especialista de infraestructura entre otros. Algunos ejemplos de metadatos técnicos son:

- ❖ Modelos de datos para los sistemas fuente.
- ❖ Reglas de extracción de datos y su calendarización.
- ❖ Reglas de agregación y limpieza de datos.
- ❖ Modelo de datos del DW.
- ❖ Nombres de columnas y descripciones de los campos.
- ❖ Mapeos entre el modelo lógico y el modelo físico de implementación.

Los productos de información generados por el proceso de ETL se almacenan en los modelos multidimensionales del DW o DM, donde se encuentran a disposición de los usuarios de negocio. Al extraer y representar los datos del DW o DM mediante reportes gráficos o visualizaciones de información es posible identificar con un vistazo patrones relevantes a tomar en cuenta durante la toma de decisiones. A continuación, se describen los principales tipos de visualización de datos.

2.7. Visualización de datos en inteligencia de negocios

La Real Academia Española define el verbo visualizar como: representar mediante imágenes ópticas fenómenos de otro carácter. Teniendo en cuenta la definición suministrada, es posible contextualizar las visualizaciones de datos en los negocios como representaciones visuales de los fenómenos o eventos vinculados a sus procesos o eventos.

La meta de la visualización es “dar respuesta a las preguntas de los usuarios, resaltando patrones y valores disonantes” (Stacey, Salvatore y Jorgensen, 2013, P. 13).

A continuación, se mencionan metas complementarias para las visualizaciones:

- Ilustrar patrones no aparentes en la información.
- Comparar valores de forma gráfica.
- Mostrar cambios en la información durante periodos específicos de tiempo.
- Crear un mecanismo ágil para observar valores o datos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Las visualizaciones se pueden categorizar bajo perspectivas estratégicas, tácticas y operativas. La diferencia entre perspectivas radica en el nivel de detalle requerido en sus visualizaciones y la magnitud o importancia de la decisión debe tomarse basada en su información.

La perspectiva estratégica se caracteriza por ofrecer capacidades de monitoreo sobre la información. Mientras nuevas oportunidades de negocio están siendo evaluadas por las altas jerarquías organizacionales, el personal que tiene acceso a la información desde una perspectiva estratégica puede indagar si el negocio se comporta según los patrones esperados, o bien si es necesario tomar acciones para corregir el rumbo (Stacey et al, 2013).

La perspectiva operativa se representa generalmente con un mayor nivel de detalle y las decisiones que dependen de estas vistas son más frecuentes que las estratégicas, individualmente las decisiones operativas generan un impacto menor en la organización, pero de manera colectiva su importancia se incrementa (Stacey et al, 2013).

La perspectiva táctica generalmente posee un carácter exploratorio y no de monitoreo a diferencia de las perspectivas anteriores. El impacto de las decisiones tácticas basadas en visualizaciones es menor que el impacto de las decisiones estratégicas, pero su efecto es mayor que las decisiones operativas (Stacey et al, 2013).

2.7.1. Tipos de visualización

Las visualizaciones de información pueden clasificarse en distintos tipos según el propósito de su diseño. En esta sección se describen los tipos predominantes en las visualizaciones de datos en ambientes empresariales.

Las visualizaciones temporales cobran gran relevancia durante el proceso de toma de decisión ya que proveen un marco comparativo entre dos o más períodos de tiempo. Stacey et al (2013) señalan que:

Los análisis temporales ilustran el paso del tiempo y la tendencia en la información mediante elementos gráficos. Convencionalmente las representaciones temporales inician desde el margen izquierdo del gráfico hasta el extremo derecho de la visualización (P. 192).

Algunos gráficos utilizados comúnmente para representar los análisis temporales son:

Gráficos lineales

Este tipo de visualizaciones “son utilizadas para representar datos continuos al conectar los puntos representados mediante una línea” (Nusbaunmer, 2015, P. 52).

Sobre la información que se pueden representar mediante gráficos lineales Stacey et al (2013) establecen que:

Un ejemplo de un caso donde no se debe utilizar este gráfico es para representar ventas semanales de una tienda que opera de lunes a sábado con una línea continua, conectando los valores del sábado con los valores del lunes siguiente (P. 196).

La Figura 2.11, ejemplifica un gráfico lineal que posee solo una serie de tiempo.



Figura 2.11. Ejemplo de gráfico lineal.

Nota: Adaptado de Stacey et al. (2013)

Gráficos de barras y columnas

Estos gráficos son útiles para mostrar valores cuya adición individual representa información valiosa, por ejemplo, el valor de ventas para un periodo específico de venta tal como los meses financieros. Es importante tomar en cuenta que los ojos humanos comparan rápidamente los extremos de cada barra o columna para determinar la categoría predominante y la categoría más pequeña (Nusbaunmer, 2015).

La diferencia entre un gráfico de barras y uno de columnas es que un gráfico de barras “utiliza la longitud de cada barra para mostrar una fecha de inicio y una fecha de fin, mientras

Propuesta para la elaboración de reportes crediticios de una entidad financiera

que un gráfico de columnas utiliza la longitud de sus columnas individuales para mostrar un valor específico” (Stacey et al, 2013, P. 197).

La Figura 2.12 y la Figura 2.13 muestran ejemplos de un gráfico de columnas y de barras respectivamente.

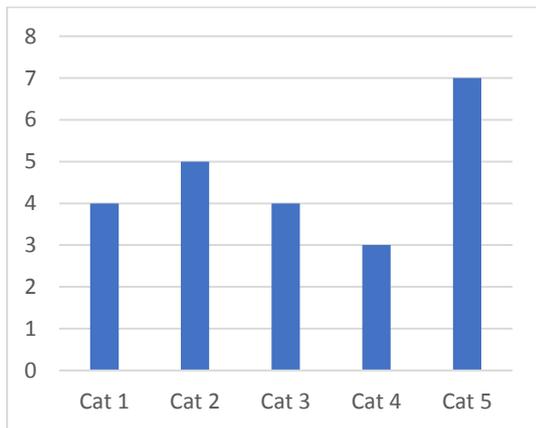


Figura 2.12. Ejemplo de gráfico de columna.

Nota: elaboración propia

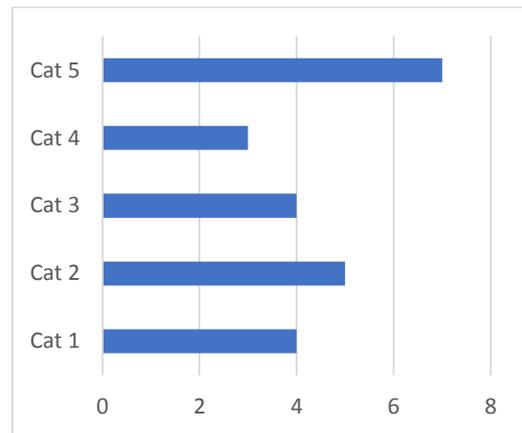


Figura 2.13. Ejemplo de gráfico de barras.

Nota: elaboración propia

En los gráficos de barras y columnas existe una subdivisión conocida como gráficos apilados. Nusbaunmer (2015) establece que:

Los casos de uso para este tipo de visualizaciones son mucho más reducidos, ya que fueron diseñados para comparar totales entre datos categóricos y visualizar los subcomponentes de cada categoría, pero resultan abrumadores visualmente y la comparación de valores resulta confusa ya que no se tiene una línea base de comparación consistente, a diferencia de la alternativa convencional de gráficos de barras o columnas (P. 59).

La Figura 2.14 y Figura 2.15, muestran ejemplos de gráficos de barras y columnas apiladas.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

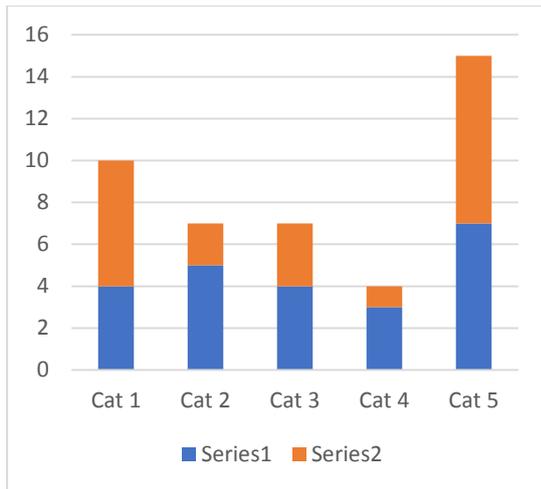


Figura 2.14. Ejemplo de gráfico de columnas apiladas.

Nota: elaboración propia.

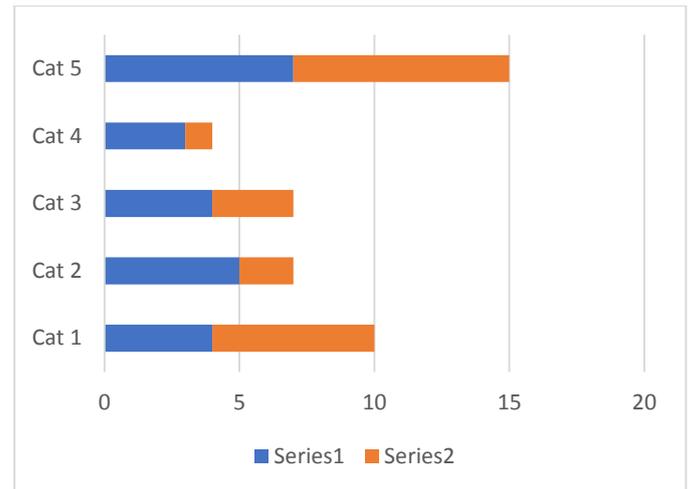


Figura 2.15. Ejemplo de gráfico de barras apiladas.

Nota: elaboración propia.

Gráficos de dispersión

Los gráficos de dispersión representan un conjunto de puntos graficados sobre los ejes X y Y de un plano, para efectos de análisis temporales, el eje horizontal siempre representará la dimensión de tiempo.

Estos tipos de gráficos poseen gran utilidad en campos científicos, ya que permiten representar los resultados de encuestas de opinión o la relación entre dos variables de forma sencilla (Nusbaunmer, 2015).

La Figura 2.16, muestra un ejemplo de un gráfico de dispersión convencional.

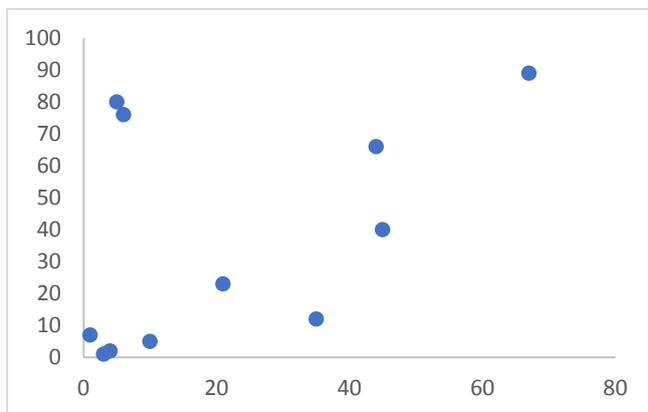


Figura 2.16. Ejemplo de gráfico de dispersión.

Nota: elaboración propia.

Nussbaumer (2015) complementa los tipos de visualización explicados anteriormente al incluir los siguientes elementos visuales:

Tablas

Las tablas proveen gran utilidad cuando se requiere “representar múltiples unidades de medida, las representaciones tabulares se utilizan para organizar información en columnas y filas para comparar valores” (Nusbaumer, 2015, P. 47). La Tabla 2.2, muestra un ejemplo de una representación tabular simple.

Tabla 2.2.

Ejemplo de representación tabular.

Grupo	Métrica A	Métrica B	Métrica C
Grupo 1	\$	%	Z
Grupo 2	\$	%	Z
Grupo 3	\$	\$	Z
Grupo 4	\$	\$	Z

Nota: adaptado de Stacey et al. (2013)

Mapa de calor

Los mapas de calor son representaciones tabulares que “toman ventaja de pistas visuales mediante la implementación del color, los colores en este tipo de visualizaciones ayudan a resaltar la magnitud de los valores representados” (Nusbaumer, 2015, P. 48). La Tabla 2.3, ejemplifica la representación de un mapa de calor para datos categóricos.

Tabla 2.3.

Ejemplo de mapa de calor.

Bajo - Alto	A	B	C
Categoría 1	15%	22%	42%
Categoría 2	40%	36%	20%
Categoría 3	35%	17%	34%
Categoría 4	30%	39%	26%
Categoría 5	55%	30%	58%
Categoría 6	11%	25%	49%

Nota: adaptado de Stacey et al. (2013)

Slopegraph

Estas representaciones son particularmente útiles cuando se requieren ejemplificar “incrementos o decrementos en categorías sobre dos periodos de tiempo independientes” (Nusbaunmer, 2015, P. 54). La Figura 2.17, muestra un ejemplo simple de un *Slopegraph* para la retroalimentación de empleados durante dos periodos de tiempo.



Figura 2.17. Ejemplo de Slopegraph.

Nota: adaptado de Stacey et al. (2013).

Pastel

Las representaciones de este tipo representan un desafío para el lector cuando los segmentos que componen el gráfico presentan tamaños similares. Es posible interpretar cual segmento posee un mayor tamaño, pero no la diferencia absoluta entre dos segmentos particulares, esto puede solventarse mediante el uso de etiquetas para la información del gráfico (Nusbaunmer, 2015).

Finalmente, es posible integrar las visualizaciones descritas anteriormente en un *Dashboard* o tablero de control. Un *Dashboard* según Rasmussen, Bansal y Chen (2009) es:

Una pantalla diseñada para mostrar información clave que ayuda a los administradores durante el monitoreo de las áreas que tienen a cargo. Los *Dashboards* brindan la posibilidad para descubrir problemas y tomar acciones correctivas que incrementen el desempeño organizacional mediante elementos visuales (P. 3).

La Figura 2.18, muestra un ejemplo de gráfico de pastel para la atención de gestiones de clientes.

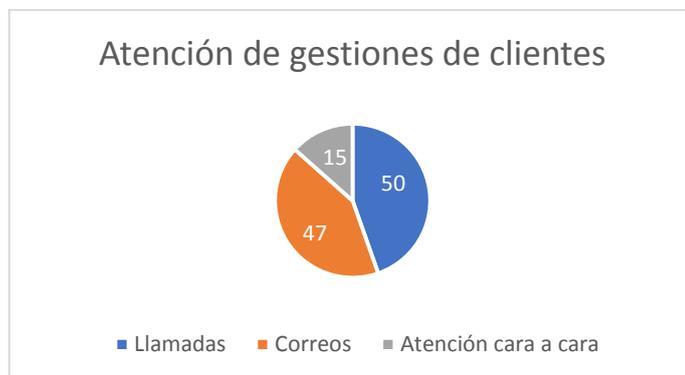


Figura 2.18. Gráfico de pastel sobre la atención de gestiones.

Nota: elaboración propia

Una vez establecidos los conceptos que integran una solución de DW o DM desde la construcción de sus modelos de datos hasta las visualizaciones de información, se procede a explorar los métodos disponibles que garantizan el desarrollo y entrega de este tipo de soluciones a los usuarios finales.

2.8. Enfoques para el desarrollo de un *Data mart*

En esta sección se describen dos enfoques de implementación para proyectos de inteligencia de negocio orientados al desarrollo de un *Data warehouse* o su especialización el *Data mart*. La Tabla 2.4, muestra un resumen de las etapas que componen los métodos establecidos por los autores Paulraj Ponniah, Ralph Kimball y Margy Ross.

Tabla 2.4.

Métodos para el desarrollo de un DW o DM.

Paulraj Ponniah	Ralph Kimball y Margy Ross
Planificación	Planificación
Requerimientos	Requerimientos
Diseño	Diseño de arquitectura
Construcción	Selección e instalación de productos
Liberación	Diseño modelo dimensional y modelo físico
Mantenimiento	Diseño e implementación del ETL

Nota: adaptado de Ponniah (2001), Kimball y Ross (2013).

A continuación, se detallan los enfoques planteados por los autores mencionados en la Tabla 2.4.

2.8.1. Enfoque planteado por Paulraj Ponniah

Ponniah (2001), establece seis grandes etapas en su enfoque del ciclo de vida para los proyectos de *Data warehousing*. La Figura 2.19, muestra las etapas planteadas por el autor, cada etapa se compone de tareas iterativas, el objetivo de realizar tareas iterativas es refinar los entregables finales de cada fase.

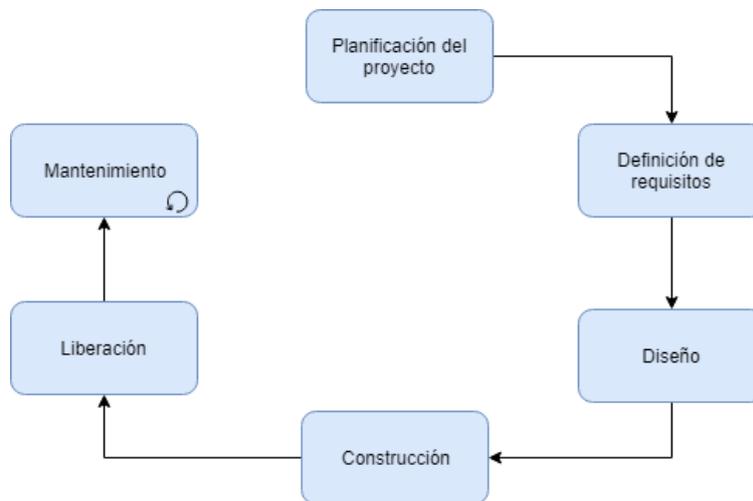


Figura 2.19. Propuesta de ciclo de vida para un DW según Paulraj Ponniah.

Nota: adaptado de Ponniah (2001).

A continuación, se describen la subdivisión de actividades para cada etapa del ciclo de vida de Ponniah.

2.8.1.1. Planificación

El ciclo de vida de Ponniah (2001) inicia por la planificación del proyecto, en esta etapa se elabora el plan de la iniciativa que delimita su alcance y las tareas por realizar. El plan del proyecto describe los objetivos específicos, menciona los factores críticos de éxito, enumera supuestos, resalta aspectos claves para la ejecución del proyecto, especifica el cronograma propuesto e incluye las tareas o asignaciones para cada miembro del equipo del proyecto.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Durante la etapa de planificación se procede a identificar los recursos tecnológicos y de información que se requieren para la correcta ejecución de la iniciativa, además, se establece el equipo que ejecutará las actividades de desarrollo y puesta en marcha de la solución.

2.8.1.2. Requerimientos

La fase de requerimientos del ciclo de vida fundamenta el diseño e implementación del *Data warehouse*, ya que documenta las necesidades de información de las áreas de negocio. Las tareas que se realizan durante la definición de requerimientos son:

Definición de la estrategia para recolección de requisitos

Ponniah (2001), menciona que antes de iniciar la recolección de requisitos se deben ejecutar las siguientes tareas que guiarán el proceso:

- ❏ Establecer los tipos de requerimientos a recolectar: se debe procurar el desarrollo de un plan para recolección de requerimientos de datos, extracción y reglas de negocio.
- ❏ Definir los usuarios o personal a entrevistar: el personal que será entrevistado incluye generalmente usuarios de negocio, personal técnico y de altos niveles gerenciales. Al seleccionar personal en distintos niveles de la organización se puede crear una visión integral de las necesidades del negocio.
- ❏ Investigación previa sobre el personal a entrevistar: una vez identificado el personal a entrevistar se debe investigar sobre la estructura de la unidad de negocio a la que pertenecen, sus roles, responsabilidades y la razón de ser para cada unidad de negocio.
- ❏ Definir las técnicas a utilizar durante la recolección de requerimientos: las técnicas a utilizar incluyen reuniones individuales o con grupos pequeños de personas (dos o tres personas) y las sesiones para desarrollo de aplicaciones conjuntas (conformados por grupos con no más de 20 integrantes).

Luego de ejecutar las tareas iniciales, Ponniah (2001), propone la recolección de requerimientos mediante la ejecución del siguiente conjunto de actividades:

Ejecución de entrevistas y sesiones grupales

Las técnicas de entrevista y las sesiones grupales son elementos básicos para la captura de requerimientos. Por medio de entrevistas individuales se pueden especificar los detalles más intrincados de las necesidades de negocio, mientras que durante las sesiones grupales es posible alcanzar consensos sobre los requerimientos detectados previamente (Ponniah, 2001, P. 99).

Revisión de documentos existentes

Las entrevistas y sesiones grupales ayudan a especificar gran parte de los requerimientos de los usuarios, complementando la recolección de requisitos con revisiones documentales es posible detectar información valiosa sobre elementos técnicos y de negocio que no habían sido mapeados. La revisión documental incluye investigación de los análisis o reportes realizados en los departamentos de negocio, documentación técnica sobre sistemas transaccionales y sus bases de datos (Ponniah, 2001, P. 104).

Preparación del documento formal de requerimientos

Una vez consolidados los requerimientos de negocio en un documento formal, es posible iniciar el proceso de validación con los usuarios. Generalmente, los documentos de requisitos mapean los requerimientos para fuentes de datos, transformaciones, almacenamiento de información y entrega de datos.

2.8.1.3. Diseño

La definición de requerimientos guía el proceso completo de diseño y desarrollo del DW. La fase de diseño del ciclo de vida es la encargada de crear modelos y mecanismos que reflejan la manera en la que se ejecutan los procesos de negocio de la organización para incorporarlos en el sistema (Ponniah, 2001, P.109).

Las actividades principales de la fase de diseño incluyen:

Diseño del modelo lógico

Durante la fase de diseño, se desarrolla el modelo de datos para el área de preparación de datos y el repositorio o DW. Los modelos se utilizarán para comunicar al equipo del proyecto la información que se incluirá en el entregable final de la iniciativa (Ponniah, 2001, P.111).

Los detalles técnicos del diseño o modelado lógico de un DW se explican en la sección 2.4 de este capítulo.

Diseño del proceso ETL

El diseño de este proceso incluye: diseño de las extracciones, transformaciones y cargas para tablas de hechos, dimensiones, estructuración del área de preparación, planificar la construcción de tablas agregadas, reglas de limpieza de datos y el determinar las fuentes de datos internas y externas. (Ponniah, 2001)

Los detalles técnicos del diseño para procesos de ETL se explican en la sección 2.4.6 de este capítulo.

Diseño del framework para la entrega de información

Tomando en cuenta las necesidades de extracción de datos documentadas durante la fase de requerimientos, se procede a desarrollar las funcionalidades del componente para entrega de información a incorporar en el DW (Ponniah, 2001).

Las características identificadas en la comunidad de usuarios dictarán los mecanismos a implementar en la estrategia de entrega de información, por ejemplo, si la comunidad se compone de usuarios de negocio o analistas, la estrategia de entrega de información estará orientada a herramientas analíticas, mientras que para usuarios con roles técnicos se debe proveer una plataforma con capacidades para desarrollo de consultas. (Ponniah, 2001).

Establecer requerimientos de almacenamiento

Los requerimientos de negocio determinan la robustez y apertura de los sistemas de base de datos que conforman el ambiente del DW. Los requerimientos de almacenamiento se ven dictados por la interacción que posee este componente con respecto a herramientas de procesamiento y visualización de datos en la arquitectura planteada (Ponniah, 2001).

Definición de arquitectura

La arquitectura del DW se refiere a la forma de estructurar componentes tecnológicos para maximizar el beneficio de la solución implementada. Durante el diseño de la arquitectura se ajustan los componentes para adaptarlos a las necesidades del ambiente del DW, esta labor se realiza para asegurar que todos los componentes trabajen bajo un esquema integrado.

Los componentes principales de la arquitectura de un proceso de ETL según Ponniah (2001) son:

<u>Fuentes de información</u>	<u>Preparación de datos</u>	<u>Carga de datos</u>
<ul style="list-style-type: none">• Producción de datos• Información interna• Información archivada• Información externa	<ul style="list-style-type: none">• Extracción de datos• Transformación de datos	<ul style="list-style-type: none">• Almacenamiento de datos• Almacenamiento de información• Metadatos• Entrega de datos• Gestión y control

2.8.1.4. Construcción

En esta fase del ciclo de vida se realiza la implementación de los modelos y estructuras creadas durante la fase de diseño.

Las actividades que componen la fase de construcción del DW según Ponniah (2001) son:

Selección e instalación del DBMS

Según Ponniah (2001), La selección del sistema gestor de base de datos se ve influenciada por los requerimientos del negocio, en este caso es importante determinar:

- ❏ Nivel de experiencia del usuario: es necesario conocer si los usuarios cuentan con experiencia en sistemas que dan soporte a la toma de decisiones, ya que esto determinará las características técnicas que deben integrarse en la solución.
- ❏ Tipos de consultas: conocer los tipos de consultas que realizarán los usuarios ayudará a determinar si el sistema gestor debe incorporar un optimizador de consultas robusto o si se requiere un sistema interno que balancee la ejecución de las consultas.
- ❏ Necesidad de apertura: se debe establecer las capacidades de integración requeridas para el sistema gestor de base de datos con respecto a los componentes de la arquitectura del DW.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ Cargas de datos: se debe estimar el volumen de datos y frecuencias de carga que debe realizarse.
- ❏ Localización del repositorio de datos: se tiene que determinar si el sistema debe seguir un esquema centralizado o distribuido.
- ❏ Crecimiento del DW: es necesario estimar el crecimiento en cantidad de usuarios y la complejidad de las consultas que se realizarán, ya que esto dicta la escalabilidad requerida en el sistema.

Selección e instalación de las herramientas de ETL

Para la selección de herramientas orientadas a la extracción, transformación y carga de datos Ponniah (2001) menciona los siguientes aspectos a tomar en consideración para cada actividad del proceso de ETL:

Extracción de datos

La herramienta debe tener la capacidad para realizar extracciones masivas para actualizaciones completas de las estructuras de datos y extracciones parciales para cargas incrementales.

Otros aspectos que se deben tomar en cuenta al seleccionar herramientas de extracción son: sistemas y bases de datos operacionales de los cuales se extraerá la información, disponibilidad de los datos para extracción y duplicación de datos en sistemas transaccionales (Ponniah, 2001).

Transformación de datos

En cuanto a la transformación de datos Ponniah (2001) establece que las herramientas deben contar con la capacidad para transformar los datos extraídos en formatos y estructuras apropiadas para el análisis de información, proveer valores por defecto cuando se requiera y realizar separaciones de campos y consolidación de datos.

Carga de datos

Según Ponniah (2001), las herramientas de ETL deben tener la capacidad para realizar la carga de información transformada y consolidada por medio de imágenes de carga al repositorio o DW, además de contar con mecanismos que permitan realizar las cargas de información periódicamente.

Diseño físico

El diseño físico es la última actividad de la fase de construcción descrita en el ciclo de vida de Ponniah (2001), las tareas que componen esta actividad son:

Desarrollo de estándares

El desarrollo de estándares asegura la consistencia de información entre las áreas de negocio, esta práctica procura reducir la ambigüedad en los modelos de implementación cuando personal de diversas áreas se involucran en el proceso de desarrollo (Ponniah, 2001).

Creación de un plan para agregaciones

Durante el diseño físico se debe corroborar la posibilidad de construir tablas con agregaciones para incrementar el desempeño de las consultas en datos consolidados. El plan debe especificar los tipos agregación para las tablas que constituyen el modelo físico (Ponniah, 2001).

Esquematización para las particiones de datos

Cuando las tablas de hechos o tablas dimensionales poseen una gran cantidad de registros o atributos se deben tomar en cuentas las opciones de particionamiento para facilitar los procesos de actualización y de recuperación de datos (Ponniah, 2001).

Ponniah (2001), establece que un plan de particionamiento debe incluir:

- ❏ Las tablas de hechos y tablas dimensionales seleccionadas para el proceso de particionamiento.
- ❏ Tipo de particionamiento a realizar (vertical u horizontal).
- ❏ Cantidad de particiones por cada tabla.
- ❏ El criterio utilizado para dividir cada tabla.
- ❏ Consideraciones para realizar consultas sobre tablas particionadas.

Establecer las opciones de agrupamiento para las tablas

Se deben explorar las opciones de agrupamientos para las tablas de base de datos antes de finalizar el diseño físico. Durante el proceso de examinación se debe identificar aquellas estructuras que se encuentren relacionadas, esto permite establecer cuales tablas son consultadas al mismo tiempo durante el procesamiento. Las agrupaciones de tablas incrementan

Propuesta para la elaboración de reportes crediticios de una entidad financiera

el rendimiento de las consultas ya que solamente se debe realizar un acceso al repositorio de información para extraer los datos (Ponniah, 2001, 432).

Preparar una estrategia de indexación

El desarrollo de la estrategia de indexación es fundamental para incrementar el rendimiento de las consultas realizadas sobre el modelo de solución planteado. Las estrategias de indexación incluyen planes para la creación de índices para cada tabla del modelo, en estos planes se indica cuáles campos son candidatos para la creación de índices (Ponniah, 2001, P.432).

Asignar estructuras de almacenamiento

Durante la ejecución de esta tarea se realiza la asignación de estructuras de almacenamiento, para las tablas del DW, además se incluyen los almacenamientos para archivos temporales, el área de preparación y las estructuras requeridas por las aplicaciones de usuario final (Ponniah, 2001, P. 432)

2.8.1.5. Liberación

Al finalizar la fase de construcción todos los componentes desarrollados ya han sido probados y se encuentran listos para la puesta en marcha de la solución. Las actividades que se deben ejecutar en esta fase del ciclo de vida según Ponniah (2001) son:

Aceptación de usuarios

La liberación de la iniciativa debe ser respaldada por los representantes claves de las áreas de negocio, la liberación no debe forzarse si los usuarios no han expresado su satisfacción con el DW. Opciones para obtener la aceptación de los usuarios incluyen procesos formales de entrega o pruebas de aceptación para cada función que realiza el sistema (Ponniah, 2001).

Análisis del desempeño para cargas iniciales

En esta actividad se realiza la carga inicial y se ejecuta la primera carga incremental, normalmente 24 horas después de la carga inicial (Ponniah, 2001).

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Para completar la carga inicial Ponniah (2001) recomienda:

- ❏ Eliminar los índices en las tablas relacionales del DW, ya que comprometen el rendimiento y los tiempos de carga.
- ❏ Eliminar la restricción de integridad referencial para acelerar el proceso de carga.
- ❏ Cargar primero las tablas dimensionales, comenzando por las más pequeñas primero.
- ❏ Cargar la tabla de hechos.
- ❏ Crear las tablas agregadas basándose en las tablas dimensionales y de hechos.
- ❏ Construir los índices del modelo para optimizar el rendimiento de consultas
- ❏ Si no se desactivó la regla de integridad referencial, se deben revisar las bitácoras para resolver excepciones en las cargas.

Entrenamiento y soporte inicial a usuarios

Ponniah (2001) establece que para entrenar a los usuarios se debe:

- ❏ Establecer conceptos básicos de bases de datos y almacenamiento de información.
- ❏ Establecer claramente las funcionalidades fundamentales del DW disponibles.
- ❏ Navegar sobre los contenidos del DW.
- ❏ Proveer instrucciones sobre el uso de las herramientas de acceso y recuperación de datos.
- ❏ Establecer reportes y consultas predefinidas para los usuarios.
- ❏ Establecer y comunicar el cronograma donde se especifican los tiempos de entrega y de creación de reportes.
- ❏ Comunicar el cronograma para carga de datos y corte de actualización de los datos.

Adicionalmente, se debe proveer soporte de primera mano al usuario durante las primeras fases de implementación del proyecto, ya que las preguntas del usuario pueden variar desde inicios de sesión hasta el desarrollo de análisis tipo *Drill Down* complejos (Ponniah, 2001)

2.8.1.6. Mantenimiento

La última fase del ciclo de vida es la fase de mantenimiento, las actividades que conforman esta fase son ejecutadas de forma iterativa posterior a la liberación del sistema. Según Ponniah (2001), las tareas a realizar en la fase de mantenimiento son:

Recolección de estadísticas

La recolección de indicadores permite obtener información sobre el uso de los recursos de software y hardware del sistema, estos indicadores brindan información sobre las tendencias de crecimiento, funcionalidad de servidores y la utilidad de las herramientas de usuario final. La recolección de estadísticas se puede realizar mediante intervalos de tiempo o por medio de eventos (Ponniah, 2001).

Entrenamiento y soporte

Durante la etapa de mantenimiento se debe establecer un programa de entrenamiento formal para los usuarios, considerando todas las áreas de conocimiento que los usuarios necesitarán para aprovechar las funciones que ofrece el DW.

Los contenidos del programa de entrenamiento pueden variar según los tipos de usuarios a los que se dirige la capacitación, Ponniah (2001) establece la división del entrenamiento en tres grandes áreas: contenido de datos, aplicaciones y herramientas. La Tabla 2.5, muestra el contenido del plan de capacitación para los distintos grupos de usuarios.

Tabla 2.5.

Contenido del plan de capacitación

Contenido de datos	Aplicaciones	Herramientas
Dimensiones y hechos del DW.	Consultas predefinidas.	Funcionalidades de las herramientas de usuario final.
Navegación del DW.	Reportes predefinidos.	Procedimientos para ingresar en las herramientas de software.
Granularidad de los datos y tablas agregadas.	Resúmenes OLAP y análisis multidimensionales.	Uso de herramientas para navegar sobre el contenido del DW.
Extracciones de sistemas transaccionales.	Flujos de datos a las aplicaciones de negocio.	Uso de herramientas para formularlas consultas y obtener resultados.

Nota: adaptado de Ponniah (2001)

Adicionalmente, durante el periodo inicial de la puesta en marcha del sistema y luego de establecer la primera versión del DW se requiere dar soporte continuo a los usuarios, especialmente si los conceptos del DW son nuevos para la organización (Ponniah, 2001).

Propuesta para la elaboración de reportes crediticios de una entidad financiera

2.8.2. Enfoque planteado por Ralph Kimball y Margi Ross

Kimball y Ross (2013), proponen el desarrollo de iniciativas enfocadas en *data warehousing* bajo un esquema de ciclo de vida. La Figura 2.20, muestra las fases principales que componen la propuesta de los autores.

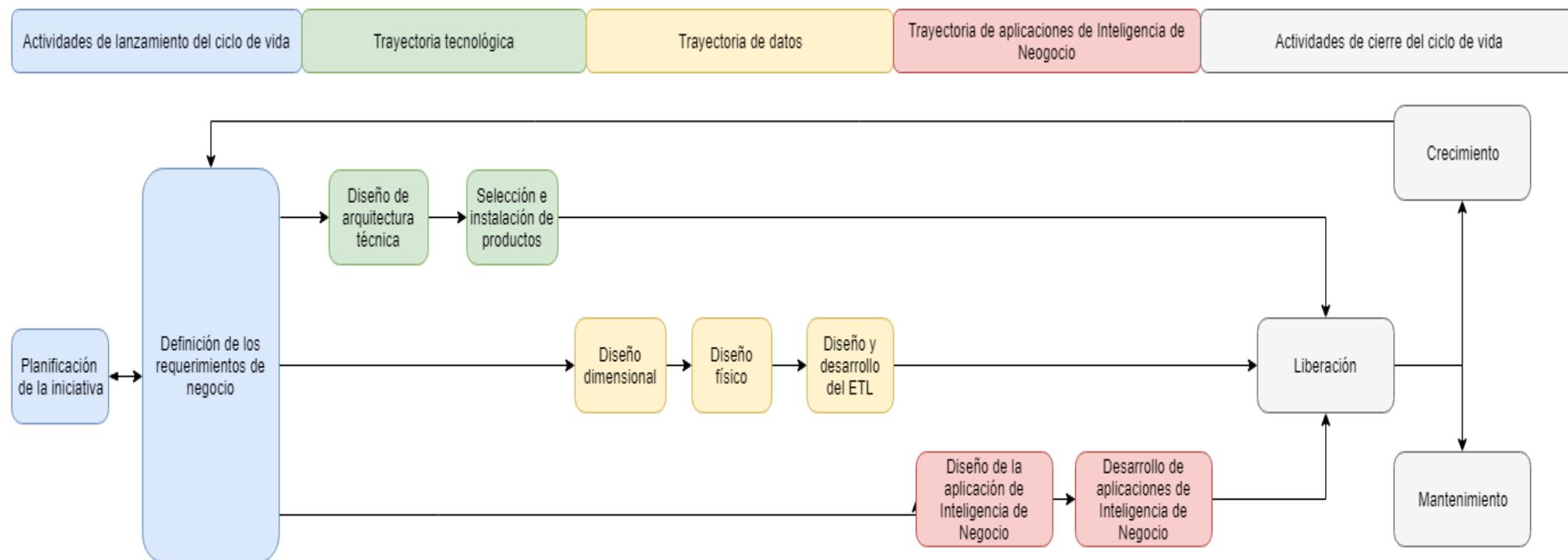


Figura 2.20. Ciclo de vida para proyectos de inteligencia de negocio.

Nota: adaptado de Kimball y Ross (2013).

Las fases que componen el enfoque de Kimball y Ross se describen a continuación.

2.8.2.1. Actividades de lanzamiento

Las tareas de planificación y definición de requerimientos ejecutadas durante la primera fase del ciclo de vida establecen la estructura que guiará el diseño e implementación de la iniciativa en fases posteriores. A continuación, se describen las actividades que conforman esta primera etapa.

Planificación de la iniciativa

Durante la planificación se ejecutan tareas cuyo enfoque es establecer la factibilidad del proyecto, delimitar su alcance y justificar su implementación. La planificación según Kimball y Ross incluye:

Evaluar la capacidad de la organización para implementar la iniciativa

Previo al desarrollo del proyecto, se recomienda evaluar si la organización está preparada para implementar la solución. Para guiar la evaluación, se mencionan tres indicadores principales:

- ❖ **Patrocinio o respaldo de la iniciativa:** los involucrados en el proyecto deben tener una visión clara de los beneficios e impacto del trabajo a realizar en la organización.
- ❖ **Motivación para el desarrollo de la iniciativa:** debe existir motivación de los actores principales que se ven involucrados en el desarrollo del proyecto, el aspecto motivacional debe estar acompañado de un sentido de urgencia dirigido a solventar las necesidades del negocio.
- ❖ **Factibilidad técnica, de datos y recursos organizacionales:** se debe tener claridad con respecto a la disponibilidad de los recursos técnicos, recursos de negocio y los datos que se utilizarán para la construcción de los entregables del proyecto.

Una vez establecida la capacidad organizacional para llevar a cabo el proyecto, se procede con la especificación de su alcance, acompañado de la justificación que respalda la ejecución de la iniciativa.

Delimitación de alcance y justificación de la iniciativa

La definición del alcance del proyecto es una tarea que requiere del consenso entre los usuarios de negocio y el equipo técnico, ambas partes deben estar de acuerdo sobre cuál es el objetivo del proyecto, las necesidades del negocio y las capacidades técnicas para solventar los requerimientos de los usuarios. Los autores mencionan que durante esta etapa las organizaciones que se encuentran iniciando procesos de inteligencia de negocio deben concentrar sus esfuerzos en proyectos que involucren un solo proceso de negocio.

Posterior a la definición del alcance es posible formar los equipos multifuncionales del proyecto que ejecutará la iniciativa. Los equipos del proyecto incluyen personal de negocio, técnico e incorpora colaboradores que cumplen un rol de facilitador para apoyar la iniciativa por medio de su visión estratégica.

Una vez establecido claramente el alcance del proyecto, se prepara la justificación de la iniciativa. Generalmente, la justificación para este tipo de implementaciones requiere una estimación de los beneficios y costos asociados al proyecto. Al finalizar la justificación del proyecto se debe tener claridad sobre los beneficios que generará la solución en términos monetarios.

Al finiquitar las tareas de planificación se procede con la definición de los requerimientos del proyecto.

Definición de requerimientos

La recolección de requerimientos es la actividad que culmina la primera etapa del ciclo de vida. Los autores recomiendan dos enfoques principales para recolectar requisitos: las entrevistas y las sesiones grupales, ambos enfoques brindan información valiosa que no podría recolectarse utilizando otras técnicas.

Kimball y Ross (2013), mencionan que iniciativas desarrolladas bajo un enfoque híbrido incorporan ambas técnicas para la recolección de requisitos alcanzando excelentes resultados, se inicia con entrevistas individuales donde los usuarios pueden expresar sus ideas libremente y finaliza con la sesión grupal donde se llega a un consenso sobre los requisitos a implementar.

Durante la recolección de requisitos es recomendable tomar en consideración los informes o reportes que realizan los usuarios actualmente, ya que brindan una guía para identificar las necesidades del negocio.

Finalmente, se procede a documentar los resultados obtenidos de las entrevistas en un documento que integre los requerimientos de negocio. El documento de requerimientos se utilizará en fases subsecuentes del ciclo de vida para desarrollar el diseño e implementar la solución de inteligencia de negocio.

2.8.2.2. Trayectoria tecnológica

La trayectoria tecnológica es la segunda fase del ciclo de vida según Kimball y Ross, durante esta fase se realiza el diseño de la arquitectura de solución y lleva a cabo la selección e instalación de posibles alternativas de implementación para las herramientas del ambiente de inteligencia de negocio. Las actividades ejecutadas durante esta etapa comprenden:

Diseño de arquitectura técnica

La arquitectura técnica especifica la manera de interactuar de la tecnología y aplicaciones en el ambiente de la solución de inteligencia de negocios. El objetivo de la arquitectura técnica es coordinar esfuerzos paralelos de desarrollo en los equipos multifuncionales, mientras se reutilizan componentes modulares. Durante el desarrollo de la arquitectura técnica se debe:

- ❏ **Recolectar requerimientos relacionadas a la arquitectura:** en este caso se busca identificar y recolectar las implicaciones de arquitectura inmersas en las necesidades de negocio y las lecciones aprendidas de iniciativas previas.
- ❏ **Documentar requerimientos de arquitectura:** se toman cada uno de los requerimientos del negocio y se documenta su impacto sobre la arquitectura y las implicaciones arquitecturales a tomar en cuenta para dar solución a los requerimientos.
- ❏ **Crear el modelo de arquitectura:** tomando como insumo los requerimientos y necesidades de negocio se deben diseñar los modelos para la solución de Inteligencia de Negocio. Los requerimientos se agrupan por categoría en ETL, inteligencia de negocio, metadatos e infraestructura, esta clasificación brinda una mayor estructura al plan de arquitectura.

- ❏ **Determinar las fases de implementación de la arquitectura:** se deben priorizar las necesidades del negocio, distinguiendo las funcionalidades que son obligatorias y las necesidades deseables en la solución, con el objetivo de priorizar o enfocar el desarrollo en las necesidades primordiales.
- ❏ **Revisar y finalizar la arquitectura técnica:** finalmente la arquitectura propuesta debe ser comunicada a los equipos de negocio y técnicos para su revisión. Una vez finalizada la verificación se ejecutan los cambios requeridos y se actualiza la documentación correspondiente.

Selección e instalación de soluciones

La selección e instalación de soluciones se divide en seis tareas principales.

- ❏ **Entendimiento el proceso de compra organizacional:** se debe comprender el proceso de compra para el software y hardware de la organización.
- ❏ **Desarrollo de una matriz para evaluación de productos:** utilizando el plan de arquitectura, se desarrolla una matriz de evaluación para las posibles alternativas o productos disponibles en el mercado para cumplir los requerimientos del negocio.
- ❏ **Realizar una investigación de mercado:** la investigación de mercado ayudará a identificar potenciales vendedores que ofrecen soluciones para solventar lo requerido por la organización.
- ❏ **Evaluar una lista reducida de opciones:** luego de analizar la matriz de evaluación de productos y seleccionar aquellos proveedores que cumplen las expectativas y necesidades de la organización, se procede con la reducción de la lista de posibles alternativas con el objetivo de realizar un análisis más detallado a las soluciones que obtuvieron la mejor calificación.
- ❏ **De ser necesario realizar una prueba:** en caso de que existan dudas sobre el proveedor ideal una vez finalizada la evaluación, se aconseja realizar un prototipo con no más de dos alternativas, el prototipo ayudará a solventar las dudas y facilitará la selección del proveedor idóneo.
- ❏ **Seleccionar el producto, instalar versión de prueba y negociar:** antes de comunicar la decisión al proveedor favorecido, es recomendable realizar un periodo de prueba con la alternativa seleccionada, esto con el objetivo de observar el comportamiento de la solución en el ambiente empresarial.

2.8.2.3. Trayectoria de datos

Durante la etapa de trayectoria de datos se realizan los diseños dimensionales que se traducen en el modelo físico de base de datos para la solución. Las actividades que componen esta fase del ciclo son:

Diseño del modelo dimensional

En este caso se debe crear el modelo dimensional que dará respuesta a los requerimientos del negocio, el diseño de los modelos dimensionales varía según las necesidades de cada organización. Al terminar esta actividad de diseño se puede proceder con la traducción del modelo dimensional a un modelo físico para la solución.

Diseño del modelo físico

Para una adecuada traducción del modelo dimensional a un modelo físico se aconseja:

Establecer convenciones de nombramiento y estándares para BD

Los nombres de los datos deben ser significativos para los usuarios de negocio, adicionalmente se deben establecer estándares relacionados con la declaración de llaves y la permisibilidad de valores nulos para los atributos del modelo.

Desarrollo del modelo físico de la BD

Se recomienda que este modelo sea construido inicialmente en el servidor de desarrollo donde el equipo del proceso de ETL podrá tener acceso a él. Adicionalmente, existen diversas tablas que deben crearse en el ambiente de desarrollo del DW, por ejemplo, tablas de preparación para soporte del proceso de ETL, y las tablas de auditoría para el procesamiento de ETL.

Diseño de agregaciones en tablas dimensionales y de hechos

Primero se debe pensar en los patrones de acceso de los usuarios y en el monitoreo de los accesos realizados a la información, seguidamente se debe analizar estadísticamente cuales son los puntos donde se deben implementar las agregaciones para mejorar el rendimiento de las consultas.

Diseño y desarrollo del proceso ETL

El diseño y desarrollo del proceso de ETL no se detalla en la propuesta de Kimball y Ross para su ciclo de vida, ya que los autores detallan que existe un libro dedicado únicamente al desarrollo de este proceso. La obra se titula *The Data Warehouse ETL Toolkit, Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data* y fue desarrollada por Kimball y Caserta (2004).

2.8.2.4. Trayectoria de aplicaciones de Inteligencia de Negocio

La fase que define la trayectoria de las aplicaciones de inteligencia de negocio se encarga de establecer y desarrollar las aplicaciones que integran la solución de IN. Las actividades de esta etapa del ciclo de vida incluyen:

Especificación de las aplicaciones de Inteligencia de Negocio

Siguiendo la definición de los requerimientos de negocio, se deben corroborar los aspectos identificados durante este proceso y recolectar reportes de ejemplo para identificar un concretamente el material generado por los usuarios del negocio.

Desarrollo de las aplicaciones de Inteligencia de Negocio

Cuando se inicia con las actividades de desarrollo en aplicaciones de IN se deben enfocar esfuerzos en mantener estándares, convenciones de nombramiento, cálculos y codificación para disminuir posibles iteraciones del desarrollo en el futuro.

2.8.2.5. Actividades de cierre del ciclo

La etapa final del ciclo de vida que proponen Kimball y Ross incluye las actividades de liberación o puesta en marcha y las actividades que mapean el mantenimiento y crecimiento potencial de la solución. A continuación, se describe cada una de las actividades mencionadas.

Liberación

Es la actividad donde convergen las etapas anteriores, por tanto, requiere de planificación previa, adicionalmente durante la liberación se busca evaluar si la solución está lista para ser utilizada por los usuarios.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

En esta actividad se realizan pruebas de sistema integrales, incluyendo el aseguramiento de la calidad de los datos, procesamiento de operaciones y pruebas de usabilidad.

Mantenimiento y crecimiento

Luego de finalizar la actividad de liberación se debe invertir tiempo y esfuerzo en:

- ❏ **Soporte:** para ganar credibilidad con los usuarios y asegurar la aceptación del entregable de la iniciativa se debe proveer soporte a los usuarios y corregir posibles errores generados durante el proceso de liberación.
- ❏ **Educación:** Se debe proveer un programa de educación continuo para usuarios de negocio y desarrolladores.
- ❏ **Soporte técnico:** soporte técnico para monitorear proactivamente el desempeño y tendencias en la capacidad de la solución.
- ❏ **Soporte al programa:** Posterior a la implementación de las soluciones se debe corroborar que se estén cumpliendo las necesidades del negocio.

Para seleccionar el enfoque óptimo de implementación para el DM o DW se deben tomar los objetivos del proyecto, las lecciones aprendidas y mejores prácticas disponibles para construir una solución eficiente que cumpla las necesidades del negocio.

2.9. Prácticas recomendadas para implementar un *Data mart*

Las mejores prácticas se pueden considerar como un conglomerado de consejos y lecciones aprendidas generadas durante la implementación de los repositorios de datos históricos. Esta sección establece un conjunto de prácticas recomendadas para el modelado dimensional y la implementación de procesos de ETL.

2.9.1. Mejores prácticas para diseño del modelo multidimensional

Durante la fase de diseño del modelo multidimensional se pueden presentar situaciones especiales en los requerimientos de negocio o características de la información que ameriten el desarrollo de estrategias o técnicas alternativas que permitan obtener un modelo de datos óptimo para el DW o DM. Prácticas implementadas comúnmente en el modelado multidimensional según Kimball y Ross (2013) incluyen:

Manejo de valores nulos en tablas de hechos

Para el manejo de los valores nulos en las llaves foráneas de la tabla de hechos, se aconseja que las tablas dimensionales asociadas cuenten con un registro por defecto que permita representar los valores desconocidos o faltantes. Al incluir valores nulos en las llaves foráneas de la tabla de hechos, se estaría incumpliendo con la regla de integridad referencial entre las tablas del modelo.

Atributos nulos en tablas dimensionales

Los valores nulos en las tablas dimensionales surgen cuando los registros de las tablas dimensionales no se han cargado por completo o existen atributos que no son aplicables a todos los registros de la dimensión. Para estos casos se recomienda sustituir los valores nulos por un valor textual descriptivo tal como “desconocido” o “No aplica”. La sustitución mencionada se realiza para evitar inconsistencias en el manejo de registros nulos a través de los distintos motores de bases de datos.

Dimensiones resumidas

Son dimensiones que representan un subconjunto de filas y columnas de una tabla dimensional base. El uso de este tipo de dimensiones se recomienda cuando se construyen tablas de hechos agregadas o bien cuando los procesos de negocio capturan de manera natural un nivel mayor de granularidad.

Hechos con diferentes monedas

Para tablas de hechos que registran transacciones financieras en múltiples monedas se aconseja manejar columnas separadas por cada moneda, una columna contiene el hecho expresado en la moneda real de la transacción, mientras que otra columna guarda el mismo valor pero expresado en una moneda estándar, por ejemplo el dólar. Los valores estandarizados para las monedas se crean durante la ejecución del proceso de ETL y deben incorporar las reglas de negocio que dictan la conversión de la divisa. Adicionalmente, para estos casos se recomienda el manejo de una tabla dimensional separada que identifique la moneda real de las transacciones en la tabla de hechos.

Hechos con múltiples unidades de medida

Cuando se requiere que los hechos sean almacenados en distintas unidades de medida, se recomienda establecer un acuerdo con los usuarios sobre el uso de unidades de medida estándar y las reglas de conversión para cada tipo de hecho según los requerimientos del negocio. A nivel de la tabla de hechos se debe almacenar el valor de la medida estándar y los factores de conversión. El mantener los factores de conversión almacenados junto a los hechos asegura que el despliegue de las vistas solicitadas por el usuario se realice de forma simple y correcta, minimizando al mismo tiempo la complejidad de las consultas sobre el modelo de datos.

2.9.2. Mejores prácticas para implementar procesos de ETL

Prácticas recomendadas para la implementación de procesos de ETL también se encuentran disponibles en la literatura relacionada a la disciplina de IN. Este conjunto de recomendaciones tiene como objetivo agilizar el procesamiento de datos, garantizando la entrega de información oportuna a los usuarios finales. Prácticas implementadas comúnmente en los procesos de ETL según Kimball y Caserta (2004) son:

Extraer solamente los datos requeridos

Las consultas óptimas devuelven solamente los datos necesarios para el procesamiento del ETL, no se recomienda extraer las tablas de base de datos completas y aplicar filtros sobre los valores no deseados en el proceso de ETL que se ejecutará posteriormente.

Evitar el uso de la cláusula NOT en las consultas a la base de datos

Se recomienda evitar el uso de restricciones o uniones que incluyen cláusulas de negación o los operadores ' \neq ', ya que algunos motores de bases de datos optan por realizar un escaneo completo de la tabla fuente y sin tomar en cuenta los índices creados en las columnas de búsqueda.

Detección de borrados y reemplazo de valores en los registros de los sistemas fuente

Con el objetivo de minimizar el impacto que tienen estos cambios sobre los datos que procesa el sistema de ETL, se recomienda negociar con los dueños de los sistemas fuente para que de ser posible notifiquen sobre los registros eliminados o que se sobrescribieron.

Adicionalmente, se aconseja revisar periódicamente los totales históricos de las medidas en los sistemas fuente para alertar al equipo del proceso ETL sobre los cambios realizados. Cuando se detecta un cambio se debe navegar al detalle de los registros para aislar e identificar el cambio.

Detección de anomalías en los datos

Para llevar a cabo la detección de anomalías, se recomienda seguir una técnica simple de muestreo de datos. En este proceso se realiza un análisis de distribución sobre los datos agrupando los valores disponibles en la columna bajo revisión, al agrupar estos valores se destacan los valores disonantes. Los registros identificados como incorrectos se deben presentar al responsable de negocio correspondiente para su valoración.

Chequeos sobre calidad de datos

Otras recomendaciones a tomar en cuenta durante la detección de anomalías son las reglas que capturan valores incorrectos en las tablas de la base de datos. Los chequeos adicionales incluyen:

Reforzamiento basado en propiedades de columnas

Asegura que los datos que ingresan al flujo de información contienen los valores esperados. En esta revisión se corrobora que no existan valores nulos en columnas requeridas, también se revisan los valores numéricos que se encuentran fuera de los rangos esperados, longitudes de valores que son inesperadamente cortos o largos y las columnas que poseen valores fuera de los valores discretos establecidos.

Reforzamiento basado en estructura

Esta restricción se enfoca en la relación que existe entre columnas, asegurando que las tablas poseen llaves primarias, foráneas y mantienen la integridad referencial.

Incrementando el rendimiento del proceso de ETL

Para incrementar el rendimiento de los procesos de extracción, transformación y carga se recomienda:

 Eliminar las lecturas y escrituras a la base de datos

Durante el procesamiento ejecutado en el área de preparación de datos, se recomienda utilizar archivos planos para almacenar la información procesada, ya que consultar la información en una base de datos puede incrementar los tiempos de ejecución. La desventaja que posee este enfoque es la pérdida de metadatos generados por los sistemas que gestionan los repositorios de información durante el procesamiento del ETL.

 Filtrar los datos antes de iniciar el procesamiento del ETL

Se recomienda reducir la cantidad de registros que se deben procesar, evitando las transformaciones sobre registros que no llegarán al almacenamiento en el DW. En este caso se deben utilizar filtros sobre campos o columnas indexadas en los sistemas o fuentes transaccionales para evitar la ejecución de escaneos completos sobre las tablas.

 Paralelizar el procesamiento de datos

Procesar la información en paralelo incrementa el desempeño del proceso completo y reduce potencialmente su tiempo de ejecución. Tareas que se pueden paralelizar incluyen las consultas a las bases de datos fuente, transformaciones y preparaciones de datos, partición y carga de tablas objetivo en el DW o DM.

 Actualizar datos agregados incrementalmente

Se debe evitar el procesamiento desde cero para los datos agregados, en su lugar se debe preferir la inclusión de valores nuevos a los hechos agregados ya existentes en la base de datos multidimensional.

 Descartar las restricciones e índices en las bases de datos

Las restricciones de llave foránea representan sobrecarga innecesaria sobre el procesamiento de datos, se recomienda que previo a la ejecución del proceso de ETL se deshabiliten estas reglas en la base de datos fuente y posterior a la ejecución del proceso se habiliten nuevamente.

 Permitir al sistema de ETL hacer el trabajo

Se recomienda evadir la ejecución de procedimientos almacenados, funciones y disparadores que identifican valores duplicados a nivel de la base de datos transaccional, ya que estas tareas las puede realizar el sistema de ETL.

Otras prácticas complementarias en el desarrollo de proyectos de *data warehousing*, son las enfocadas en la evaluación de alternativas de implementación. Estos lineamientos buscan asegurar la adquisición de soluciones óptimas que cumplan las expectativas funcionales, técnicas y de costo que posee la organización.

2.10. Prácticas para evaluar alternativas de implementación

El proceso de evaluación en proyectos de inteligencia de negocio tiene por objetivo seleccionar la alternativa óptima de solución que cumpla con los requerimientos de los usuarios y las expectativas del patrocinador del proyecto en términos del retorno de la inversión.

Esta sección explora prácticas utilizadas comúnmente durante la evaluación de alternativas de implementación en proyectos.

Ponniah (2001), menciona algunos lineamientos básicos para iniciar el proceso de valoración para las soluciones disponibles en la rama de inteligencia de negocio. Los lineamientos a tomar en cuenta son:

- 1) Determinar las funciones del DW que se requieren en las herramientas diseñadas por los proveedores.
- 2) Para cada herramienta identificada se dividen las funcionalidades esperadas en grupos de importancia (alto, medio y bajo).
- 3) Destinar tiempo suficiente para investigar las soluciones y los proveedores a detalle.
- 4) Mantener una lista de dos o tres proveedores cuyos productos se adecuen a las necesidades o requerimientos.

- 5) De ser requerido se debe asegurar que las soluciones del proveedor permitan gestionar metadatos.
- 6) Probar las herramientas y productos en el ambiente organizacional incorporando subconjuntos de datos propios.
- 7) Establecer un método definitivo para comparar y calificar las soluciones del proveedor, se debe colocar una calificación por cada funcionalidad.
- 8) Se deben considerar herramientas intuitivas para los usuarios, ya que posiblemente sea su primera experiencia utilizando las herramientas de un DW.

Aunque los aportes anteriores son cruciales durante la evaluación de alternativas, no contemplan aspectos relacionados al Retorno de Inversión (ROI por sus siglas en inglés). Con el objetivo de complementar los criterios de evaluación mencionados por Ponniah, a continuación, se integra el concepto e implicaciones del cálculo del ROI para proyectos de IN.

Uno de los principales componentes que se deben tomar en cuenta durante el cálculo del Retorno de Inversión son los costos.

2.10.1. Costos asociados a la evaluación de proyectos

Phillips, Brantley y Phillips (2011) mencionan que un enfoque conservador para estimar los costos del proyecto debe incluir los elementos que se explican a continuación:

- ❏ **Evaluación y análisis inicial:** este criterio incluye los costos de la recolección de datos, resolución de problemas, evaluación y análisis del proyecto.
- ❏ **Desarrollo de soluciones:** estos costos incluyen el diseño e implementación del proyecto, así como la adquisición de los materiales utilizados directamente en el proyecto.
- ❏ **Costos de adquisición:** este criterio toma en cuenta los costos de adquirir las soluciones que integran el proyecto, materiales de soporte y acuerdos de licenciamiento.

- ❏ **Costos de implementación y aplicación:** es el criterio de costo más grande, ya que incluye conceptos de salarios del personal involucrado.

- ❏ **Monitoreo y mantenimiento:** estos costos incluyen los gastos rutinarios para mantener la operativa del proyecto.

- ❏ **Evaluación y reportes:** este criterio incluye tiempo, materiales, instrumentos adquiridos y cargos por consultorías vinculadas a la preparación, diseño y comunicación de los resultados del proyecto.

Una vez identificados los costos que posee el proyecto se procede con el cálculo del Retorno de Inversión. A continuación, se describen dos métodos para calcular esta razón financiera.

2.10.2. Cálculo del ROI

El cálculo del Retorno sobre la Inversión debe tomar en cuenta los beneficios del proyecto ya que según mencionan Philips et al. (2011):

Presentar los costos sin identificar los beneficios correspondientes puede resultar problemático. Cuando la mayoría de los ejecutivos revisan los costos del proyecto, se genera una pregunta lógica: ¿Qué beneficios se recibieron de este proyecto? Esta es una reacción típica de los administradores cuando se perciben costos muy elevados en el proyecto (P. 203).

Los beneficios del proyecto pueden clasificarse en incrementos en utilidades o ahorros en costos. En la realidad existen mayores oportunidades para ahorrar costos que para incrementar utilidades, debido a esto la mayoría de los proyectos se enfocan en la reducción de costos (Philips et al, 2011).

Razón Beneficio/Costo

Es un indicador o “razón que compara de manera simple los beneficios del proyecto con respecto a sus costos” Phillips et al. (2011), para construir este indicador se utiliza la siguiente formula:

$$BCR = \frac{\text{Beneficios del proyecto}}{\text{Costos del proyecto}}$$

Los resultados de la Razón Beneficio/Costo indican los beneficios en términos monetarios de invertir en el proyecto, por ejemplo, un resultado de \$1 indica que los beneficios son iguales a los costos, un resultado de \$5 indica que por cada dólar invertido se retornan 5 dólares (Phillips et al, 2011, P. 210).

Formula del Retorno sobre la Inversión

Phillips et al. (2011), con respecto al Retorno sobre la Inversión establecen que:

La fórmula apropiada para evaluar las inversiones en proyectos se obtiene al dividir los beneficios netos del programa entre los costos. Los resultados de esta fórmula se expresan generalmente en términos de porcentaje donde los valores fraccionados se multiplican por 100, con la siguiente formula:

$$ROI(\%) = \frac{\text{Beneficios netos}}{\text{Costos del proyecto}} \times 100$$

Los resultados del ROI indican que por cada dólar invertido se obtendrá un beneficio de \$1, multiplicado por el resultado de la fórmula.

2.10.3. Flujos de caja descontados

Además del ROI existen otros indicadores para la evaluación de proyecto de inversión, a continuación, se presenta una breve descripción de cada uno.

Flujos de caja descontados

Es un método que parte del supuesto de que un dólar ganado hoy tiene un mayor valor que un dólar ganado dentro de un año. Este acercamiento es utilizado comúnmente para el cálculo del valor presente neto de la inversión (Philips et al., 2011, P. 215).

Philips et al (2011), establecen que el cálculo de los flujos de caja descontados se realiza tomando en cuenta que:

Los ahorros de cada año se comparan con las salidas de efectivo requeridas por el proyecto, seguidamente se descuenta los ahorros anuales esperados y las salidas de efectivo utilizando la tasa de interés seleccionada. Si el valor presente de los ahorros descontados supera las salidas de efectivo descontadas, se considera que el proyecto es aceptable (P. 215).

Al contrastar los costos totales del proyecto y sus beneficios, es posible discernir cuál debe ser la alternativa que genera los resultados esperados por los niveles gerenciales de la organización a un nivel de costo razonable, justificando a su vez la inversión realizada.

Una vez finalizada la recolección de aspectos teóricos que fundamentan los componentes del proyecto se procedió a establecer los lineamientos metodológicos que permitieron ejecutar la iniciativa plasmada en este documento.

3

Desarrollo Metodológico

Este capítulo describe la metodología utilizada para el desarrollo del proyecto. En las secciones iniciales del capítulo se especifica el tipo de investigación seleccionada, fuentes de información, instrumentos y las técnicas implementadas para la recolección de datos. Secciones posteriores detallan los procedimientos metodológicos y las fases que componen la metodología utilizada para el desarrollo del proyecto.

La investigación se describe como el “conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema” (Hernández, Fernández y Baptista, 2014, P.4).

Dada la definición anterior, se puede establecer que los trabajos de índole investigativo se componen de tareas definidas que al ser ejecutadas de forma ordenada y crítica permiten ampliar el conocimiento disponible en las áreas de conocimiento relacionadas al estudio.

Según Hernández et al. (2014), las investigaciones se pueden agrupar en tres tipos o enfoques principales:

- ❏ Cualitativas: que buscan conocer las percepciones de los sujetos de estudio para determinar las causas relacionadas a los fenómenos de investigación.
- ❏ Cuantitativas: que se enfocan en la búsqueda de generalizaciones para un conjunto de casos de estudio.
- ❏ Mixtas: que combinan aspectos cualitativos y cuantitativos para formar un espectro más amplio de análisis que incorpora criterios causales con criterios cuantitativos.

Para clasificar la presente investigación es necesario definir y establecer las características de los tres enfoques mencionados anteriormente.

3.1. Tipo de investigación

Debido a la variedad de casos de estudio que se presentan en las áreas de conocimiento, se han desarrollado un conjunto de enfoques que permiten indagar y comprender las causas de los fenómenos que originan los trabajos de investigación. Los enfoques comprenden la estructuración del estudio fenomenológico desde su gestación hasta la presentación de las conclusiones que se generan producto del análisis de resultados. A continuación, se describen los enfoques principales que puede seguir una investigación.

3.1.1. Investigación Cuantitativa

Este enfoque se compone de pasos secuenciales que parten de la formulación de una idea, que una vez delimitada, deriva los objetivos y preguntas de investigación. A partir de estas se establecen las hipótesis y se determinan las variables de estudio. Finalmente se traza un plan estructurado que permite medir las variables, analizarlas y extraer un conjunto de conclusiones relacionadas a las hipótesis de investigación.

Algunas características que se destacan en los métodos cuantitativos según Hernández et al. (2014) son:

- Buscan la comprobación de hipótesis con base en la medición numérica y el análisis estadístico.
- Procuran la objetividad del estudio al evitar que el proceso se vea afectado por el investigador o alterado por las tendencias de personas vinculadas a la investigación.
- Mantienen un patrón predecible y estructurado en el proceso, donde las decisiones sobre el método se realizan previo a la recolección de datos.
- Confirman y predicen los fenómenos de investigación, buscando regularidades y relaciones entre variables de investigación.

El acercamiento cuantitativo es particularmente útil cuando se busca generalizar los resultados de investigación de un fenómeno mediante procedimientos estadísticos de análisis sobre los datos.

Es por medio de la estructuración rigurosa del proceso de recolección y análisis estadístico que es posible repetir o reproducir el estudio aplicando la misma serie de pasos y procedimientos.

3.1.2. Investigación Cualitativa

Las investigaciones cualitativas buscan “examinar la forma en que los individuos perciben y experimentan los fenómenos que los rodean, priorizando en sus puntos de vista, interpretaciones y significados” (Hernández et al, 2014, P. 358).

Dada la definición anterior se puede determinar que las investigaciones cualitativas se enfocan en las experiencias de los sujetos de estudio que se ven afectados por un fenómeno o situación. Este aspecto de las investigaciones cualitativas es relevante cuando se requiere diagnosticar un fenómeno o problema que se presenta en una comunidad específica de personas.

Según Hernández et al. (2014), los estudios cualitativos pueden desarrollar preguntas e hipótesis antes, durante o incluso después de realizada la recolección y el análisis de datos a diferencia del enfoque cuantitativo donde las hipótesis se formulan previo a la ejecución del estudio. Las investigaciones cualitativas por lo general no presentan regularidad en la secuencia de sus pasos, ya que su ejecución varía según el estudio realizado.

Según Hernández et al. (2014), los métodos cualitativos se caracterizan por:

- No seguir un proceso definido claramente para su desarrollo, ya que su ejecución depende del fenómeno de estudio.
- Establecer como propósito la reconstrucción de la realidad, tal y como la observan los sujetos de investigación.
- Realizar una inmersión inicial en el campo de estudio para identificar informantes que aporten datos, guíen al investigador en el proceso y ayuden a verificar la factibilidad del estudio.
- Evaluar el desarrollo natural de los sucesos, sin propiciar la manipulación de la realidad del fenómeno de estudio.

El enfoque cualitativo resulta útil cuando se requiere comprender la causalidad de los fenómenos que afectan una comunidad particular de estudio, por medio de la inmersión inicial del investigador en el campo o área donde se presenta el problema.

3.1.3. Investigación Mixta

Los métodos mixtos representan un enfoque combinado de las investigaciones cualitativas y cuantitativas. Según Hernández et al (2014), las investigaciones mixtas son:

“...procesos sistemáticos, empíricos y críticos de investigación e implican la recolección y el análisis de datos cuantitativos y cualitativos, así como su integración y discusión conjunta, para realizar inferencias producto de toda la información recabada y lograr un mayor entendimiento del fenómeno bajo estudio” (Hernández et al, 2014, P. 534).

Los métodos mixtos permiten alcanzar una perspectiva más amplia y profunda del fenómeno al sustentar el estudio desde la perspectiva numérica e indagatoria. Los procesos mixtos incorporan técnicas propias de los enfoques cuantitativos y cualitativos, por ejemplo, encuestas, entrevistas y observación. Mediante el uso de las técnicas de recolección de datos es posible construir un rango amplio de evidencia que fundamenta el diagnóstico del fenómeno y favorece su comprensión.

Algunas características preponderantes de las metodologías mixtas según Hernández et al. (2014) son:

- ❏ Su orientación hacia el planteamiento del problema para definir los métodos a emplearse en un estudio determinado.
- ❏ Parte de un conjunto de diseños y procesos analíticos que se realizan de acuerdo a las circunstancias del estudio.
- ❏ Su proceso de investigación es iterativo y cíclico.

3.1.4. Selección del tipo de investigación

Para la selección del enfoque del proyecto, se tomaron en cuenta algunos lineamientos generales mencionados por Hernández et al. (2014) quienes establecen que:

- ❏ El enfoque a seleccionar se debe ajustar al que el investigador considere armoniza o se adapta mejor a su planteamiento del problema.
- ❏ Los problemas que buscan establecer tendencias se acomodan mejor a un diseño cuantitativo.
- ❏ Los problemas que requieren ser explorados para obtener un entendimiento profundo se adaptan mejor a un diseño cualitativo.
- ❏ Cuando el problema o fenómeno es complejo, los métodos mixtos ofrecen herramientas para solventar las interrogantes de estudio.

Tomando en cuenta los puntos anteriores y la descripción de los tres enfoques de investigación, se selecciona el enfoque cualitativo, ya que el proyecto desarrollado requiere comprender y explorar el fenómeno que afecta actualmente los procesos de creación de informes en la entidad financiera desde la perspectiva del personal que elabora el material de apoyo para la toma de decisiones.

Adicionalmente, el enfoque cualitativo seleccionado facilita la comprensión de las causas relacionadas al fenómeno de estudio por medio de la inmersión inicial en el ambiente donde se presenta el problema a resolver. El proceso de inmersión permite la formulación y desarrollo de la propuesta de solución del proyecto, al tomar en cuenta las experiencias y conocimiento del personal de las áreas de Riesgo y Crédito que ha experimentado las repercusiones del problema de estudio.

3.2. Diseño de la investigación

El método cualitativo posee un conjunto de diseños genéricos que guían el desarrollo de la investigación, Hernández et al. (2014) mencionan que dichas perspectivas pueden clasificarse en:

- a) Teoría fundamentada: se utiliza para problemas donde no se dispone de teorías para el contexto, caso, muestra o bien cuando las teorías existentes no son adecuadas.
- b) Diseños etnográficos: se utilizan cuando se busca describir, entender y explicar un sistema social.
- c) Diseños narrativos: busca comprender una sucesión de eventos, a través de la narrativa detallada de quienes los vivieron.
- d) Diseños fenomenológicos: se utiliza cuando se busca comprender las experiencias de varias personas sobre un fenómeno o perspectivas de este.
- e) Diseños de investigación-acción: se selecciona este diseño cuando el foco de la investigación es resolver la problemática de una comunidad y alcanzar un cambio con respecto a la situación actual.

Al tomar en cuenta la descripción brindada por Hernández et al. (2014) se decidió adoptar el enfoque investigación-acción para el proyecto, ya que este diseño resulta ideal cuando se requiere estudiar una problemática que afecta una comunidad, con el objetivo de plantear una solución que mitigue o elimine los efectos adversos del fenómeno investigado.

Dado que la finalidad de la investigación – acción es comprender y resolver una problemática, se busca que el investigador realice una inmersión inicial en el campo de estudio que le permita comprender a cabalidad las implicaciones del problema, con el objetivo de realizar aportes sobre los procesos que mitigan los efectos perjudiciales del fenómeno de estudio.

La investigación-acción puede clasificarse por medio de sus diseños básicos en participativa o práctica. Según Hernández et al. (2014), el diseño participativo resalta la colaboración equitativa de todo el grupo o comunidad de estudio y se enfoca en mejorar el nivel de vida y desarrollo humano de los individuos. Por otro lado, el diseño práctico implementa un

plan de acción para resolver el problema, introducir mejoras o generar el cambio por medio del estudio de las prácticas del grupo o comunidad.

Por las características que posee el problema de estudio se selecciona el diseño práctico de la investigación-acción, ya que se requiere plantear una propuesta de solución que atienda las necesidades de los sujetos de estudio.

Las fases principales del proceso investigación-acción son: “observar (construir un bosquejo del problema y recolectar datos), pensar (analizar e interpretar) y actuar (resolver problemáticas e implementar mejoras, las cuales se dan de manera cíclica” (Hernández et al., 2014, P. 497).

Durante la fase de observación se realizó el planteamiento del problema y la recolección de los datos que se analizaron posteriormente en la segunda fase de la investigación (pensar), una vez construido el panorama que representa la situación problemática, se procedió a ejecutar la fase final (actuar) donde se planteó e implementó la propuesta para la resolución del problema. El detalle de las fases requeridas para dar solución al problema de la entidad financiera se especifica en la sección 3.7 de este documento.

Una vez establecido el tipo de investigación a realizar, se procede con la especificación de las fuentes de información que se tomarán en cuenta durante el desarrollo del proyecto. A continuación, se detalla cada tipo de fuente utilizada durante el proyecto.

3.3. Fuentes de información

A continuación, se presentan las fuentes de información primarias y secundarias que se tomaron en cuenta para el desarrollo del proyecto, las cuales incluyen material bibliográfico de relevancia para la iniciativa y las fuentes organizacionales de conocimiento.

3.3.1. Fuentes primarias

En cuanto a las fuentes primarias de información que se tomaron en cuenta durante el desarrollo del proyecto se destacan:

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ La documentación interna de las bases de datos o repositorios de información, donde se incluyen metadatos sobre los campos de cada tabla que se tomó en cuenta para el procesamiento del ETL, la estructura de los repositorios y la interrelación entre tablas de la base de datos.
- ❏ Las políticas internas sobre el manejo de datos de los clientes y acuerdos de confidencialidad que resguardan a la institución financiera ante la divulgación no autorizada de su información financiera o de contacto de sus clientes.
- ❏ La documentación asociada a manuales de procedimiento disponibles para la creación de reportes en los departamentos de la organización.
- ❏ Las obras de Ralph Kimball que se enfocan en el diseño e implementación de soluciones multidimensionales del tipo *Data warehouse* o *Data mart* y procesos de ETL vinculados a repositorios históricos de información.
- ❏ La literatura de Paulraj Ponniah que facilitó un espectro más amplio de conocimiento para respaldar la formulación de propuestas. Los aportes de Ponniah cobran relevancia por su impacto en el tema de estudio y por su practicidad en el desarrollo de técnicas de implementación para los sistemas de toma de decisiones basados en DW o DM.
- ❏ Artículos relacionados a la creación de soluciones de inteligencia de negocio, donde se destaca el modelo para el ETL de un *Data Warehouse* que proponen El-Sappagh, Hendawi y Bastawissy y la propuesta de modelado y optimización de los procesos de ETL en el DW presentado por Anand y Kumar.
- ❏ En cuanto a visualización de datos se incorporó la guía establecida por Stacey, Salvatore y Jorgensen sobre inteligencia visual y representación de datos.
- ❏ Para el desarrollo del *benchmark* de opciones de implementación, específicamente en las secciones de evaluación de alternativas se optó por incorporar la literatura que propone Resch sobre la entrega del ROI máximo y valor del negocio sostenible.

- ❏ En el caso de los criterios del *Benchmarking* se tomaron en cuenta las recomendaciones realizadas por Oracle para herramientas de integración, adicionalmente se tomaron en cuenta los criterios establecidos por Poess, Rabl y Caufield en su trabajo de investigación.

- ❏ El conocimiento de negocio que posee el personal interno de la organización. Se destaca el conocimiento requerido para la construcción de informes que poseen los analistas en las áreas de crédito y riesgo. El conocimiento del personal interno organizacional incluye desde las excepciones a tomar en cuenta durante los procesos de elaboración de reportes hasta los sistemas o archivos utilizados como insumo en los procesos de toma de decisiones.

3.3.2. Fuentes secundarias

Para el desarrollo del proyecto, se tomó en cuenta el compendio resumen de literatura sobre inteligencia de negocio en la industria bancaria (período 2002 - 2013) que realizaron los autores Moro, Cortéz y Rita, por su relevancia con respecto al giro de negocio de la organización donde se llevó a cabo el proyecto.

Una vez establecidas las fuentes principales de información para el proyecto se procede a definir cuáles son los instrumentos y técnicas que se utilizaran para su recolección y análisis.

3.4. Instrumentos y técnicas para recolección de datos

Para recolectar el conocimiento que poseen los colaboradores de la institución y establecer una base para el análisis de la situación problemática que permita fundamentar las propuestas de solución, se establecen los siguientes instrumentos para la recolección de datos.

3.4.1. Entrevista

Esta técnica consiste en formular preguntas de forma verbal a las personas afectadas por el fenómeno de investigación, con el objetivo de obtener respuestas o afirmaciones que contribuyan a la verificación o comprobación de hipótesis del trabajo.

Ñaupas, Mejía, Novoa y Villagómez (2014) apoyados por Hernández et al. (2014) establecen que las entrevistas se pueden clasificar en:

- **Entrevistas Estructuradas:** llamadas también entrevistas dirigidas, controladas, o guiadas; son aquellas que siguen un plan y diseño preestablecido. Se realizan de acuerdo con una guía o formulario previamente preparado con preguntas enfocadas a responder las hipótesis de investigación formuladas. Este tipo de entrevistas permiten procesar mejor los datos proporcionados por el entrevistado, al aplicar las preguntas de forma ordenada y consistente.
- **Entrevistas Semiestructuradas.** al igual que las entrevistas estructuradas utiliza una guía de preguntas, pero no mantiene la formalidad y rigidez que caracteriza a la versión estructurada del instrumento, esto permite que el entrevistador introduzca algunas preguntas para esclarecer vacíos en la información; ya que no todas las preguntas están predeterminadas.
- **Entrevistas No Estructuradas.** en este tipo de entrevistas, el entrevistador tiene libertad para hacer las preguntas, siempre basándose en una guía, general de contenido y no una guía específica. Las entrevistas no estructuradas se ajustan a las necesidades de la investigación de tipo cualitativa. Su mayor debilidad radica en que son difíciles de procesar debido a la gran variedad de respuestas que pueden recolectarse.

Durante el desarrollo del proyecto se utilizaron las entrevistas semiestructuradas, lo que permitió mantener una guía sobre la entrevista y a la vez flexibilizó el proceso de obtención de datos al permitir la formulación de preguntas adicionales basadas en las respuestas del entrevistado. Estos instrumentos se aplicaron a tres analistas del área de riesgo y a cuatro analistas del área de crédito, en total se aplicaron cinco entrevistas a los siete colaboradores de la organización. La guía de las preguntas aplicadas se detalla en el Apéndice B: Guía de entrevista para identificación de necesidades y en el Apéndice C: Plantilla para recolección de requisitos.

3.4.2. Revisión documental

Esta técnica se concentra en la recolección de documentos, archivos y procedimientos relacionados al tema de investigación o estudio realizado.

La información recolectada durante la revisión documental complementa los datos obtenidos durante el proceso de entrevistas o grupos focales, lo que permite incrementar el espectro de conocimiento disponible para el investigador, esto resulta vital para comprender el fenómeno de estudio. Es importante destacar que durante la inmersión inicial del proceso de investigación cualitativa esta técnica cobra relevancia al contextualizar al investigador en aspectos relacionados con fenómeno de estudio o bien en el ambiente donde este se manifiesta.

3.5. Instrumentos y técnicas para el análisis de datos

Para analizar y estructurar la información que se recolectó por medio de entrevistas, grupos focales y revisión documental se utilizaron las siguientes técnicas e instrumentos:

 Tabulación de datos.

 Diagramación.

A continuación, se describe cada uno de los instrumentos o técnicas utilizadas.

3.5.1. Tabulación de datos

Esta técnica permite organizar la información obtenida durante los procesos de entrevista y revisión documental por medio de las estructuras de tipo tabla. La tabulación de datos permite separar, identificar y comparar resultados específicos. Un ejemplo de tabulación es el comparativo de alternativas de implementación o *benchmark* que se detalla en capítulos posteriores de este documento.

3.5.2. Diagramación

La diagramación provee un mecanismo para representar de forma gráfica conceptos, procesos, actividades y ciclos. Esta técnica permite expresar un flujo de eventos que para efectos del presente proyecto representan los procesos de negocio enfocados en la producción de reportes. En cuanto a los estándares de diagramación utilizados se destaca el uso de BPMN 2.0 que se especifica en el Anexo 1: Notación estándar BPMN 2.0 de este documento.

3.5.3. Comparativo para herramientas de integración y visualización

Esta técnica permite realizar un comparativo de dos o más alternativas que poseen características similares, para facilitar la comparación de datos este instrumento utiliza la presentación tabular la información donde cada columna representa una alternativa u opción de análisis.

Por medio de la estructura del comparativo se facilita la verificación de que las alternativas analizadas cumplen con los criterios o requerimientos de la organización donde se desarrolla el proyecto.

El formato tabular utilizado para el *Benchmark* que compara las alternativas de software para la implementación de la propuesta del presente proyecto se encuentra en el Apéndice A: Tabla para *Benchmarking* comparativo de herramientas.

3.6. Selección de la muestra para análisis de investigación

En esta sección se especifica la unidad de análisis, población, muestra y las características que describen los sujetos de información sobre los cuales se aplicaron los instrumentos de recolección de datos.

La unidad de análisis para la investigación la conformaron los colaboradores de los departamentos de riesgo y crédito de la entidad financiera encargados de construir los informes relacionados a la cartera crediticia. La experiencia acumulada por los analistas de ambos departamentos contribuyó a identificar el conjunto de actividades que ejecutaban manualmente estos colaboradores, lo que permitió determinar oportunidades de mejora para el proceso de elaboración de los reportes crediticios.

Las características de la unidad de análisis seleccionada para el desarrollo del proyecto son:

 Personal analista de las áreas de riesgo o crédito de la entidad financiera.

- 📦 Deben utilizar como insumo para la creación de sus reportes la cartera crediticia de la organización.
- 📦 Personal que ejecuta tareas manuales para construir sus reportes con la información de las carteras de crédito empresarial o personal.

Una vez establecida la unidad de análisis se procedió a determinar la población que conforma el universo de todos los posibles casos que cumplen las características mencionadas. En este caso la población estaba constituida por tres analistas del departamento de riesgo y cuatro analistas del departamento de crédito quienes ejecutan funciones de construcción y análisis de reportes manualmente.

Una vez identificada la población se procedió a delimitar la muestra sobre la cual se aplicaron los instrumentos de recolección de datos. Neuman (2009) menciona que en los estudios cualitativos el tamaño de muestra no se fija a priori (antes de la recolección de los datos), en su lugar se define el tipo de unidad de análisis y se perfila un número aproximado de casos para iniciar el proceso. La muestra final se establece cuando las unidades de análisis seleccionadas no aportan información o datos novedosos al proceso.

Para efectos del presente proyecto, la selección de la muestra se realizó mediante juicio por conveniencia, ya que según Hernández et al. (2014) “esta muestra está formada por los casos disponibles a los que tenemos acceso” (P. 390). Teniendo en cuenta la afirmación realizada por estos autores y para obtener una visión holística de las actividades que ejecutan los analistas durante la elaboración de los reportes crediticios, se estableció que la muestra debía abarcar la totalidad de la población de estudio conformada por siete analistas de la entidad.

3.6.1. Descripción de los sujetos de información

A continuación, se describen los sujetos de información que cumplen las características de la unidad de análisis y por tanto formaron parte de la población/muestra de la investigación. Adicionalmente, se incluye el perfil del jefe de sistemas de información quien tuvo una participación en el proceso. En la Tabla 3.1 se describe para cada departamento el puesto y perfil del personal entrevistado.

Tabla 3.1.

Descripción de los sujetos de información.

Departamento	Descripción	Funciones	Perfil
	Analista de riesgo crediticio	Debido a su vasto conocimiento en el proceso de análisis crediticio de índole personal y empresarial se dedica a construir informes de concentración de cartera que se dirigen a la Junta Directiva de la entidad. Los análisis que realiza este analista permiten identificar comportamientos anómalos en la cartera de clientes, así como alertas tempranas de deterioro en los deudores.	Las funciones que desarrolla este analista requieren el perfil de licenciatura en ciencias económicas o administrativas.
Riesgo	Analista de riesgo de cambiario y liquidez	Ejecuta análisis de liquidez sobre las carteras de clientes, su labor permite identificar riesgos potenciales en la solvencia de la institución y por tanto la capacidad que posee la entidad financiera para cumplir sus obligaciones ante entres reguladores. Adicionalmente, este analista evalúa el comportamiento del tipo de cambio para determinar el nivel de exposición crediticia que posee la organización con respecto a los deudores que solicitan préstamos en moneda extranjera, pero generan ingresos en moneda nacional.	Las funciones que desarrolla este analista requieren el perfil de licenciatura en ciencias económicas o administrativas.
	Auxiliar de riesgo	Brinda soporte a las tareas de los analistas del departamento por tanto posee conocimiento tanto en el área crediticia, de liquidez y riesgo cambiario.	Las funciones que desarrolla este analista requieren el perfil de licenciatura en ciencias económicas o administrativas.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Continuación de la tabla 3.1

Departamento	Descripción	Funciones	Perfil
Crédito	Analista operativo de crédito	Se encarga de identificar el estado de la cartera crediticia desde el punto de vista operativo, identificando operaciones formalizadas y construyendo los reportes de cartera que se dirigen a los entes reguladores, los cuales incluyen montos de estimación, garantías, saldos de operación y mora.	El puesto desempeñado por este analista requiere como mínimo el grado de bachillerato universitario en administración.
	Analista de mora	Se encarga de estudiar las operaciones que se encuentran en mora en la cartera crediticia y que potencialmente podrían afectar los indicadores de estimación de la entidad financiera.	El puesto desempeñado por este analista requiere como mínimo el grado de bachillerato universitario en administración.
	Analista de mercado	Construye informes que integran información de la cartera interna de clientes con datos externos del sector bancario para determinar la posición financiera de la entidad con respecto a sus pares.	El puesto desempeñado por este analista requiere como mínimo el grado de bachillerato universitario en administración.
	Analista financiero	Realiza análisis sobre indicadores financieros para recomendar cursos de acción en cuanto a colocación y recuperación de créditos.	El puesto desempeñado por este analista requiere como mínimo el grado de bachillerato universitario en administración.
Tecnología de información	Jefe de sistemas de información	Es el encargado de supervisar el desarrollo y mantenimiento de las soluciones tecnológicas de la organización.	Este puesto requiere una maestría en sistemas de información.

Nota: elaboración propia.

Una vez establecidos los lineamientos generales que guiaron el desarrollo del proyecto, se procedió a definir el conjunto de fases o etapas del proyecto que detallan cómo se llevará a cabo el desarrollo de la iniciativa.

3.7. Análisis de la información

En esta sección se especifican los métodos utilizados para estructurar y analizar la información recolectada al aplicar los instrumentos mencionados en el apartado 3.4 de este documento.

Por entrevista en la que participaron los colaboradores de la organización se ejecutaron las siguientes tareas de análisis de datos.

Entrevista para la identificación de necesidades

En este caso se esquematizaron las observaciones de los analistas de la entidad y se incorporó el resumen de las respuestas que brindaron los colaboradores de la entidad a la guía de preguntas que se especifica en el Apéndice B: Guía de entrevista para identificación de necesidades.

Sesión para captura de requisitos

Para la recolección de requisitos se llevaron a cabo sesiones grupales con los analistas de las áreas de crédito y riesgo de forma independiente, con el objetivo de consolidar en plantillas separadas los requerimientos de ambos departamentos. La plantilla para la recolección de requisitos se especifica en el Apéndice C: Plantilla para recolección de requisitos de este documento.

Entrevista para el mapeo de las fuentes transaccionales

Para identificar las fuentes de datos que se utilizan en los reportes crediticios de la entidad se procedió a entrevistas a los analistas del área de crédito y riesgo de la entidad utilizando la guía de preguntas que se detalla en el Apéndice D: Guía de entrevista para mapeo de fuentes transaccionales. Una vez recolectada la información se procedió a esquematizar, tabular y resumir las respuestas de los colaboradores para la presentación de los resultados.

Entrevista para el mapeo de los procesos de elaboración de reportes

En este caso el análisis de los datos realizado se llevó a cabo por medio de la tabulación de las respuestas de los analistas, la diagramación de las tareas en común que realizan los colaboradores en cada uno de sus procesos. La guía de preguntas utilizada durante las entrevistas se especifica en el Apéndice E: Guía de entrevista para mapeo de los procesos de elaboración de reportes

Entrevista para la selección de criterios del *Benchmarking*

Para la estructuración de los criterios del comparativo se utilizó el formato tabular que se describe en el Apéndice A: Tabla para Benchmarking comparativo de herramientas de este documento. Adicionalmente, para completar la estructura del *Benchmark* se requirió de una entrevista con la jefatura de sistemas del área de tecnología de información, la guía de preguntas realizada se encuentra en el Apéndice F: Guía de entrevista para selección de criterios del *Benchmarking*.

3.8. Fases de la metodología aplicada

En esta sección se detallan las fases que se ejecutaron para dar solución a la problemática de la entidad financiera. Cada etapa planteada contribuye al cumplimiento de los objetivos del proyecto y el desarrollo de sus entregables. La ejecución de la iniciativa se plantea en cuatro etapas secuenciales que incluyen:

-  **Descubrimiento inicial:** esta fase consolida el conocimiento interno organizacional y las prácticas utilizadas comúnmente en soluciones de Data warehousing para solventar las necesidades de información en las organizaciones.

-  **Diseño de los componentes de inteligencia de negocio:** en la segunda etapa se diseña el Data mart, proceso de ETL y visualizaciones de datos que integran la solución de IN.

- ❏ **Implementación de los componentes de inteligencia de negocio:** abarca el desarrollo de los diseños realizados durante la fase anterior en las herramientas provisionadas por la entidad financiera.
- ❏ **Comprobación de los resultados:** incluye el análisis comparativo de los resultados que se generaron al implementar los diseños de la solución contra los objetivos del proyecto.

La Figura 3.1, muestra la interacción entre las fases del proyecto mencionadas y su relación con los objetivos de la iniciativa.

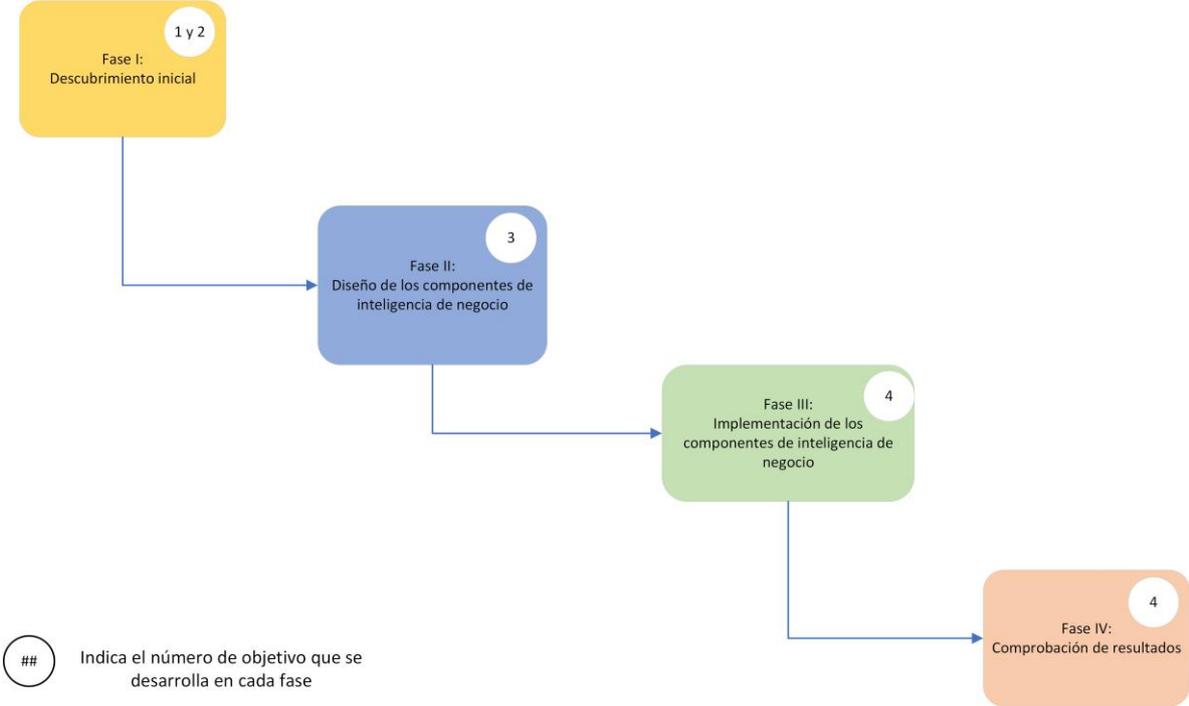


Figura 3.1. Fases del proyecto.

Nota: elaboración propia

La relación que existe entre las fases del proyecto y sus objetivos permite mapear de forma clara cuál es la finalidad que persigue cada etapa. De forma complementaria la Tabla 3.2 muestra la relación entre las fases del proyecto y las actividades que se ejecutaron en cada etapa para cumplir los objetivos correspondientes.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Tabla 3.2.

Objetivos por fase metodológica.

Objetivo	Fase	Actividades
Diagnosticar las brechas en las tareas de extracción, transformación y carga en los procesos de inteligencia de negocio del área de crédito y riesgo con respecto a las mejores prácticas de la industria, para la identificación de las causas que imposibilitan la entrega de información oportuna.	Fase I	<ul style="list-style-type: none">  Entendimiento del negocio  Revisión de mejores prácticas
Diseñar el modelo multidimensional, <i>dashboards</i> de información crediticia y los procesos de ETL asociados a la preparación de datos que se utilizan como insumo en los análisis de las áreas de crédito y riesgo.	Fase I	<ul style="list-style-type: none">  Análisis de alternativas de inversión
Diseñar el modelo multidimensional, <i>dashboards</i> de información crediticia y los procesos de ETL asociados a la preparación de datos que se utilizan como insumo en los análisis de las áreas de crédito y riesgo.	Fase II	<ul style="list-style-type: none">  Diseño del Data mart  Diseño del ETL  Diseño de las visualizaciones
Construir una solución automatizada que facilite el procesamiento de datos en las tareas analíticas de Crédito y Riesgo para la disminución del reprocesamiento de informes en un 80% y la reducción del tiempo de elaboración en al menos un 50%.	Fase III y IV	<ul style="list-style-type: none">  Implementación del Data mart  Implementación del ETL  Implementación de visualizaciones.

Continuación de tabla 3.2

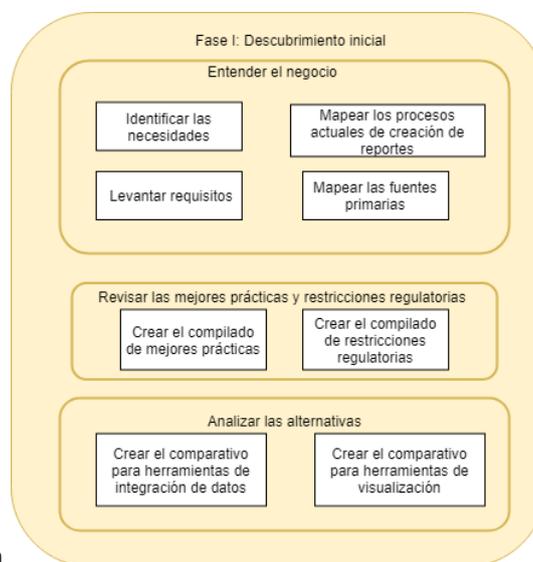
Objetivo	Fase	Actividades
Construir una solución automatizada que facilite el procesamiento de datos en las tareas analíticas de Crédito y Riesgo para la disminución del reprocesamiento de informes en un 80% y la reducción del tiempo de elaboración por lo menos en un 50%.	Fase III y IV	<ul style="list-style-type: none"> 📦 Ejecución de pruebas sobre el proceso de ETL 📦 Comprobación de resultados

Nota: elaboración propia

A continuación, se especifican las actividades que componen cada fase mencionada en la Figura 3.1.

3.8.1. Fase I: Descubrimiento inicial

Durante esta fase se recolectó la información interna de la entidad financiera y las prácticas recomendadas en la literatura para proyectos de *Data warehousing*. El objetivo de esta etapa era consolidar el fundamento teórico de las propuestas de solución que solventan



las necesidades de la organización. La

Figura 3.2 muestra gráficamente los componentes de la fase orientada al descubrimiento inicial.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

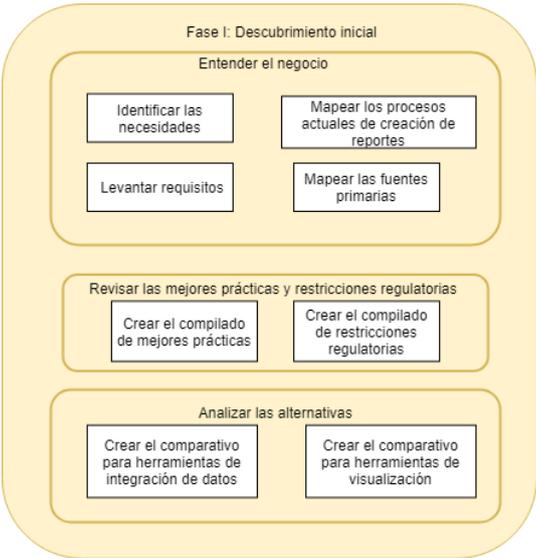


Figura 3.2. Fase I: Descubrimiento inicial.

Nota: elaboración propia.

La primera fase del proyecto se compone de tres actividades principales enfocadas en la recolección de datos de fuentes internas y externas de conocimiento. Cada actividad posee un conjunto de tareas o subactividades, cuya finalidad es contribuir la creación de los entregables del proyecto. A continuación, se detalla cada actividad y las subactividades que las componen.

3.7.1.1. Entender el negocio

Durante esta actividad se recopiló el conocimiento asociado a la inteligencia de negocio en las áreas de crédito y riesgo de la organización, la información recolectada fundamentó la creación de propuestas para la entidad financiera. La recopilación de datos demandó que se profundizara en la operativa de ambos departamentos y el proceso que realizan para elaborar sus informes, esto se llevó a cabo por medio de entrevistas enfocadas en comprender los reportes que realizan los analistas de ambas áreas.

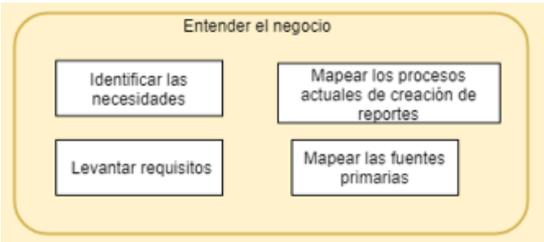


Figura 3.3. Entender el negocio, primera actividad de la fase I del proyecto.

Nota: elaboración propia

A continuación, se detallan las tareas específicas que se ejecutaron durante la primera actividad de la fase I del proyecto que se muestra en la Figura 3.3.

Identificar las necesidades

Para identificar las necesidades en los procesos de producción de informes que poseen las áreas de crédito y riesgo se realizaron entrevistas a los analistas de ambas áreas.

Las entrevistas realizadas se enfocaron en determinar requerimientos implícitos en las necesidades de los analistas de la institución. La guía de preguntas utilizada durante la ejecución de esta subactividad se encuentra en el Apéndice B: Guía de entrevista para identificación de necesidades de este documento. Este instrumento se aplicó a los siete analistas que integraron la muestra de análisis del proyecto.

Levantar requisitos

El enfoque de esta subactividad fue la documentación de los requerimientos que se debían incorporar en la solución de IN. El levantamiento de requisitos se realizó por medio de entrevistas a los analistas de las áreas de crédito y riesgo. Durante las entrevistas se consultó sobre las funcionalidades deseadas en la solución de inteligencia de negocio y los factores que debían tomarse en cuenta durante el diseño de los componentes de la propuesta.

Adicionalmente, el levantamiento de requisitos buscaba identificar la estructura para el repositorio multidimensional, transformaciones realizadas sobre los datos para el proceso de ETL y las métricas o datos que se deben representar en las visualizaciones de información.

En el Apéndice C: Plantilla para recolección de requisitos se muestra la plantilla utilizada para la documentar los requerimientos de las áreas de negocio, en este caso se realizó una sesión grupal donde se consolidaron los requerimientos de los analistas de riesgo y crédito.

Mapear las fuentes primarias

Luego de identificar las necesidades de las áreas de negocio se procedió a mapear las fuentes de información que utilizan los analistas en la construcción de sus informes. Este mapeo tomó en cuenta sistemas transaccionales que registran las operaciones de los clientes con la

Propuesta para la elaboración de reportes crediticios de una entidad financiera

entidad, archivos de Excel o archivos planos donde se almacenan datos relacionados al giro de negocio de la institución.

El mapeo de fuentes se realizó por medio de entrevistas aplicadas a los analistas de la entidad. Con el desarrollo de las entrevistas se buscaba identificar todos los datos que se utilizan como insumo para elaborar los reportes dirigidos a la toma de decisiones organizacional. Las preguntas utilizadas para esta entrevista se especifican en el Apéndice D: Guía de entrevista para mapeo de fuentes transaccionales, este instrumento fue aplicado a los siete analistas que integraron la muestra del proyecto.

Durante esta actividad se llevó a cabo la revisión documental de los metadatos vinculados a los repositorios de información, lo que permitió identificar las características de los datos que se debían incluir en el modelo multidimensional y el proceso de ETL que se desarrollaron en etapas posteriores del proyecto.

Mapear los procesos actuales de creación de reportes

Por medio de entrevistas a los analistas de la entidad financiera se mapearon y documentaron los procesos ejecutados durante la elaboración de informes dirigidos a otros departamentos. En el levantamiento de estos procesos se requería identificar los tiempos promedio que le toma a cada analista construir sus informes.

La serie de preguntas realizadas a los analistas de la organización se encuentra disponible en el Apéndice E: Guía de entrevista para mapeo de los procesos de elaboración de reportes. Este instrumento se aplicó dos sesiones grupales independientes para los analistas del área de riesgos y crédito respectivamente.

La documentación del proceso para creación de reportes se realizó por medio de la notación BPMN 2.0, estableciendo un estándar que facilitó la comprensión del flujo de tareas realizado durante la creación de reportes.

3.7.1.2.Revisar las mejores prácticas y restricciones regulatorias

Esta actividad comprende la recopilación de mejores prácticas y restricciones regulatorias a tomar en cuenta durante el diseño e implementación del modelo multidimensional y el proceso de ETL.

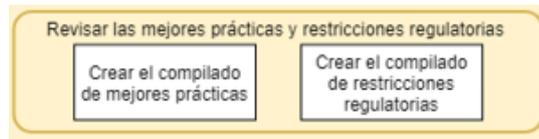


Figura 3.4. Revisar las mejores prácticas y restricciones regulatorias, segunda actividad de la fase I del proyecto.

Nota: elaboración propia.

Las subactividades que componen la revisión de mejores prácticas que se muestra en la Figura 3.4 son:

Crear el compilado de mejores prácticas

Se investigaron prácticas utilizadas comúnmente durante el diseño de modelos multidimensionales y procesos de ETL documentadas en textos bibliográficos, que permitieron establecer una base de conocimiento aplicable durante el desarrollo de estos componentes.

Una vez finalizada la recopilación de mejores prácticas se procedió a valorar su factibilidad de implementación y beneficios que ofrecían con respecto a la situación actual de la Organización. La viabilidad de las prácticas a implementar se valoró contra las necesidades y requerimientos de la organización.

Crear el compilado de restricciones regulatorias

En esta subactividad se estudiaron las restricciones regulatorias vinculadas al manejo y procesamiento de datos de los clientes. Adicionalmente, se identificaron las restricciones normativas que dispone la Superintendencia de Entidades Financieras (SUFEG) y del Consejo Nacional de Supervisión del Sistema Financiero (CONASSIF) para el desarrollo del proyecto.

Las normas que se tomaron en cuenta son:

- Ley orgánica del sistema bancario nacional.
- Ley orgánica del Banco Central de Costa Rica.
- Ley de protección de datos de la persona frente al tratamiento de sus datos.

3.7.1.3. Analizar las alternativas

En esta actividad se realizó la investigación de alternativas disponibles para llevar a cabo la automatización del proceso de ETL. Se indagó sobre las soluciones o herramientas tecnológicas de automatización que ofrece el mercado para implementar procesos de extracción, transformación y carga. Durante la selección se tomaron en cuenta soluciones ubicadas en el cuadrante mágico de Gartner por su relevancia en el mercado internacional.

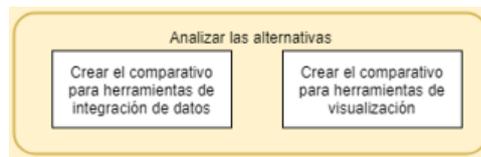


Figura 3.5. Analizar las alternativas, tercera actividad de la fase I del proyecto.

Nota: elaboración propia.

A continuación, se detallan las tareas de la Figura 3.5 correspondiente a la tercera actividad de la primera fase del proyecto.

Crear el comparativo para herramientas de integración de datos

La selección de herramientas para el comparativo se realizó tomando en cuenta la afinidad o experiencia que ha tenido la organización con los distintos proveedores del cuadrante mágico de Gartner. La selección de los criterios de comparación se realizó mediante la revisión documental de artículos científicos y prácticas recomendadas por los proveedores de herramientas de integración tales como Oracle y Talend.

Crear el comparativo para herramientas de visualización

En este comparativo se tomaron en cuenta cuatro proveedores de herramientas para visualización de datos. La selección de herramientas para el comparativo se realizó tomando en cuenta la afinidad o experiencia que ha tenido la organización con los distintos proveedores del cuadrante mágico de Gartner. Para la selección de los criterios de evaluación se realizó una revisión documental de los componentes principales que poseen las herramientas de

visualización en el mercado, durante la revisión se tomó en cuenta principalmente la propuesta de Tableau y Microsoft por su relevancia y trayectoria con la institución financiera. Adicionalmente, se llevó a cabo una entrevista con la jefatura de sistemas de la entidad para identificar criterios complementarios para el análisis. La guía de entrevista se detalla en el Apéndice F: Guía de entrevista para selección de criterios del *Benchmarking*. La información del comparativo se consolidó en la estructura establecida en el Apéndice A: Tabla para *Benchmarking* comparativo de herramientas.

Análisis de inversión

Para cada alternativa se realizó un análisis del ROI que otorgará la inversión, lo que permitió identificar la solución óptima que satisface los requerimientos del negocio y genera los beneficios esperados sin sobrepasar el costo requerido para su implementación.

El análisis de inversión expresa los beneficios de cada alternativa y los costos de su implementación en la fórmula del ROI.

$$ROI(\%) = \frac{\text{Beneficios netos}}{\text{Costos de cada alternativa}} \times 100$$

Al aplicar la fórmula se obtuvieron las soluciones que brindan el mayor retorno con respecto a los costos de implementación.

3.7.2. Fase II: Diseño de los componentes de inteligencia de negocio

La fase II del proyecto se enfoca en las actividades de diseño para el modelo multidimensional, proceso de ETL y las visualizaciones de datos. La Figura 3.6 muestra gráficamente los componentes de la fase orientada al diseño de los componentes de IN.

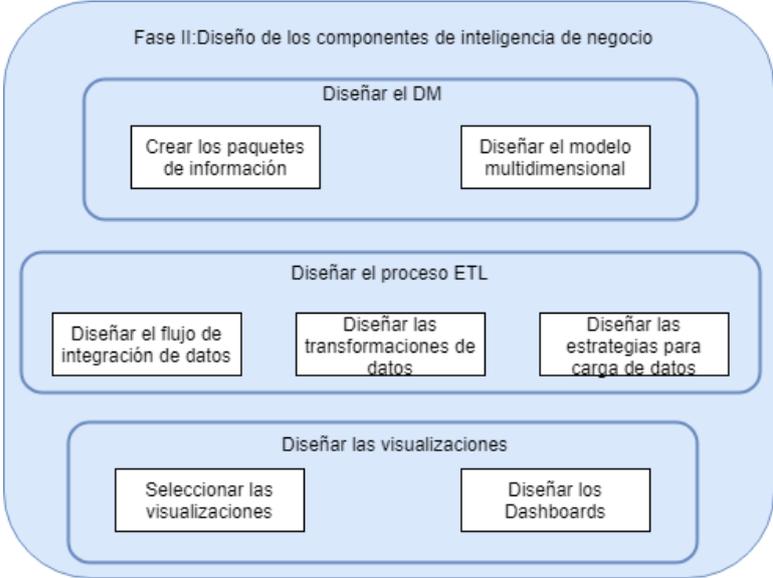


Figura 3.6. Fase II: Diseño de los componentes de inteligencia de negocio.

Nota: elaboración propia.

A continuación, se detalla cada una de las actividades de diseño que conforman la segunda fase del proyecto.

3.7.2.1. Diseñar el DM

En esta actividad se diseña el modelo de datos que rige la información del *Data mart*. La Figura 3.7 muestra las tareas que comprende el diseño del repositorio multidimensional de datos que alberga la información crediticia de la organización.

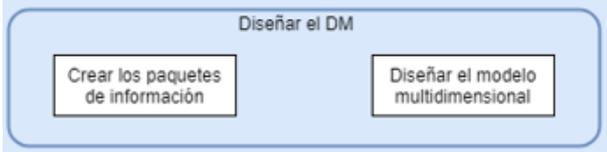


Figura 3.7. Diseñar el DM, primera actividad de la fase II del proyecto.

Nota: elaboración propia.

Las tareas que integran la primera actividad en la segunda etapa de desarrollo del proyecto son:

Crear los paquetes de información

Previo a la construcción del modelo multidimensional, se confeccionaron los paquetes de información que se originan de los requerimientos del negocio. El contenido de los paquetes se estructuró en un formato tabular donde se especifican las dimensiones, sus atributos y las medidas o hechos numéricos que se debían incluir en el modelo multidimensional. La estructura de los paquetes de información se muestra en el Apéndice G: Plantilla de paquetes de información.

Diseñar el modelo multidimensional

En esta subactividad se realizó el diseño del modelo multidimensional donde se almacenaron los datos de las áreas de crédito y riesgo al finalizar la ejecución del proceso ETL, en dicho modelo se tomaron en cuenta los datos que las áreas solicitaron para elaborar sus informes.

El diseño del modelo multidimensional incluyó la estructuración de la tabla de hechos (para datos cuantitativos o métricas), delimitación de las tablas dimensionales (almacenan datos categóricos relacionados con las métricas de la tabla de hechos) y los enlaces entre tablas.

Durante el diseño del modelo multidimensional se analizó el tipo de esquema requerido para dar solución a las necesidades del departamento de Crédito y Riesgo. Los esquemas considerados durante el diseño del modelo fueron estrella y copo de nieve.

Cabe destacar que el diseño del modelo multidimensional requería ser aprobado en una reunión presencial y documentado mediante la minuta respectiva. La aprobación se registrará utilizando la plantilla de minuta que se muestra en el Anexo 7: Plantilla para minuta de reunión.

3.7.2.2. Diseñar el proceso ETL

Esta actividad se enfocó en el diseño del proceso que toma los datos almacenados en sistemas transaccionales, los transforma y almacena en el modelo dimensional desarrollado en la actividad anterior de esta fase. La Figura 3.8 muestra las subactividades o tareas que componen la actividad de diseño para el proceso de ETL de la fase II del proyecto.

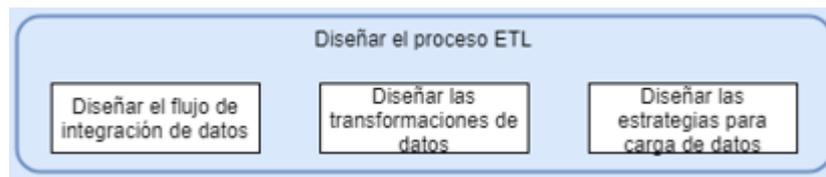


Figura 3.8. Diseñar el proceso ETL, segunda actividad de la fase II del proyecto.

Nota: elaboración propia.

A continuación, se detallan las tareas que se ejecutan durante la segunda fase del proyecto.

Diseñar el flujo de integración de datos

Se realizó la integración de los sistemas de procesamiento transaccional, donde se extrajo la información insumo para los análisis crediticios generada en las herramientas de negocio, principalmente el Core Bancario de la institución.

Como parte del mapeo de los datos que proveen estos sistemas se identificó el conjunto de información cuantitativa clave o primaria de análisis, los datos categóricos complementarios, la periodicidad de actualización, extracción y granularidad de los datos. La información mencionada anteriormente se utilizó durante el diseño de las transformaciones que forman parte del proceso de ETL.

Diseñar las transformaciones de datos

Tomando los datos primarios identificados anteriormente se automatizaron en la herramienta los cálculos o transformaciones que realizan actualmente los colaboradores de la organización. Las transformaciones programadas incluyen: dolarización de saldos o montos, conversiones a otros tipos de moneda extranjera, formato de la información (estructuración de fechas o identificadores) y agregaciones sobre datos cuantitativos. El resultado de estos cálculos o transformaciones se tomó en cuenta durante el diseño del último paso del ETL, relacionado a la carga final de datos.

Diseñar las estrategias para carga de datos

Luego de realizar el diseño de las transformaciones se procedió a establecer la estrategia de carga de la información, donde los datos primarios se enlazaron a la tabla de hechos, mientras que los datos complementarios se vincularon a las tablas dimensionales del modelo. La relación

entre tablas se realizó mediante llaves foráneas, lo que habilitó el desarrollo de análisis que contemplan distintas perspectivas de negocio.

3.7.2.3. Diseñar las visualizaciones

En esta actividad se realizó la planificación y diseño de las visualizaciones de datos basadas en la información del modelo multidimensional que requieren los analistas de las áreas de riesgo y crédito.

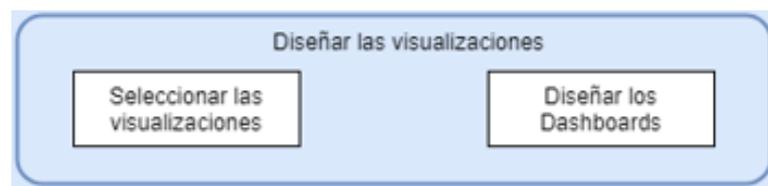


Figura 3.9. Diseñar las visualizaciones, tercera actividad de la fase II del proyecto.

Nota: elaboración propia.

La Figura 3.9 muestra las subactividades que se realizaron durante el diseño de las visualizaciones, cada tarea que compone esta actividad se describe a continuación.

Seleccionar las visualizaciones

Algunos análisis de información crediticia se benefician de las representaciones gráficas, lo que permite a los analistas la identificación de tendencias en los datos que podrían no ser intuitivas o evidentes. Por tanto, en esta actividad se seleccionaron las representaciones gráficas que mejor se ajustan al análisis requerido por el personal de la institución financiera.

La selección de visualizaciones se aprobó en una reunión con los analistas, quienes determinaron si la propuesta cumple con sus requerimientos. La aprobación de las visualizaciones se documentó mediante la plantilla de minuta ubicada en el Anexo 7: Plantilla para minuta de reunión.

Diseñar los dashboards

Finalmente, se crearon los diseños de un conjunto de *dashboards* utilizando las herramientas de visualización disponibles en la institución, para corroborar que el resultado obtenido concuerda con las expectativas de las áreas de negocio.

Se corroboró que los *dashboards* permitieran al usuario realizar operaciones de *Drill Down* sobre la información, al igual que lo hacen muchas herramientas para modelos multidimensionales; con el objetivo de presentar un tablero general de información que las gerencias puedan utilizar como medio para investigar a detalle los datos que les presentan los analistas.

3.7.3. Fase III: Implementación del *Data mart*, proceso de ETL y visualizaciones

La fase III del proyecto se enfoca en la implementación de los diseños realizados en la fase anterior los cuales incluyen el modelo multidimensional, proceso de ETL y las visualizaciones de datos. La Figura 3.10 muestra gráficamente los componentes de la fase a la implementación de los componentes de IN.

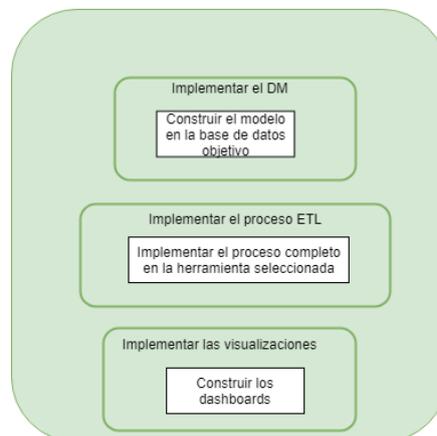


Figura 3.10. Fase III: Implementación del *Data mart*, proceso de ETL y visualizaciones.

Nota: elaboración propia.

Al igual que las fases anteriores la implementación de los diseños se compone de subactividades que permiten alcanzar el objetivo de la fase. A continuación, se describe cada una de ellas.

3.7.3.1. Implementar el *Data mart*

La implementación del *Data mart* constituye la tarea inicial de la fase IV y sobre esta tarea se establecen las bases para la ejecución de tareas posteriores. La Figura 3.11 muestra la subactividad que compone la implementación del DM del proyecto.

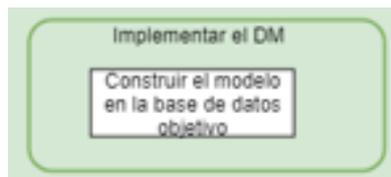


Figura 3.11. Implementar el DM, primera actividad de la fase III del proyecto.

Nota: elaboración propia.

Construir el modelo en la base de datos objetivo

En esta actividad se implementó el modelo lógico del *Data mart* en el repositorio que pone a disposición la entidad financiera. La implementación del modelo de datos es fundamental ya que este repositorio alberga la información procesada del ETL y es la fuente de datos principal que alimenta las visualizaciones y *Dashboards* finales que utilizan los analistas para sus análisis.

La creación del DM incluye la construcción de las tablas de hechos, tablas dimensionales y sus relaciones por medio de sentencias estándar de SQL en el motor de base de datos destino. El detalle de la implementación del modelo de datos se especifica en la sección 5.4.1 de este documento.

3.7.3.2. Implementar el ETL

En esta actividad se llevó a cabo la implementación del proceso de Extracción, Transformación y Carga en la herramienta seleccionada. La implementación del ETL incluyó la construcción de las conexiones con los sistemas transaccionales y el área de preparación de datos. La Figura 3.12 muestra la subactividad que compone la implementación del proceso de ETL que forma parte de la iniciativa.

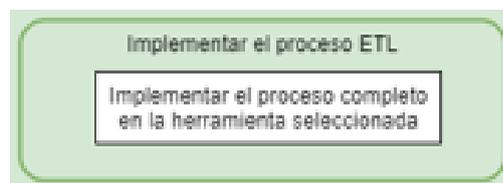


Figura 3.12. Implementar el proceso ETL, segunda actividad de la fase III del proyecto.

Nota: elaboración propia.

Implementar el proceso completo en la herramienta seleccionada

Durante la construcción del proceso de ETL se programaron las extracciones, transformaciones y la carga de información, tomando como insumo los datos ubicados en las fuentes transaccionales.

Seguidamente, se establecieron las conexiones entre el proceso de ETL y el modelo multidimensional del DM, donde se almacenaron los datos resultantes del proceso.

Finalmente, se verificó que todos los datos fueran consistentes con respecto a las bases transaccionales, por medio de un proceso de revisión donde se verificó la cantidad de registros almacenados en el DM con respecto a la fuente transaccional de información.

3.7.3.3. Implementar las visualizaciones

Esta actividad incluye la implementación de las representaciones gráficas de datos que se integraron posteriormente en los *Dashboards* solicitados por el personal de la entidad financiera. La Figura 3.13 muestra la subactividad que compone la implementación de las visualizaciones del proyecto.

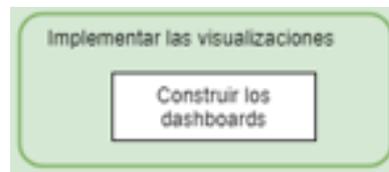


Figura 3.13. Implementar las visualizaciones, tercera actividad de la fase III del proyecto.

Nota: elaboración propia.

Implementar las visualizaciones

Una vez procesados y almacenados los datos crediticios en el *Data mart*, se procedió a establecer la representación gráfica de la información, habilitando al personal de la institución financiera la ejecución de análisis que identifican tendencias en los datos y patrones de comportamiento de la cartera crediticia.

3.7.4. Fase IV: Comprobación de resultados

La fase IV del proyecto se enfoca en la comprobación de resultados que se obtienen del proceso de ETL. La Figura 3.14 muestra gráficamente los componentes de la fase de comprobación de resultados.



Figura 3.14. Fase IV: Comprobación de resultados.

Nota: elaboración propia.

3.7.4.1. Ejecutar las pruebas

Las pruebas de concepto brindaron evidencia sobre el funcionamiento del proceso de ETL diseñado, lo que permitió determinar si la solución es candidata para realizar el pase a los ambientes de producción de la entidad financiera.

Los pasos realizados en esta actividad son el diseño, ejecución y documentación de las pruebas de concepto. Las pruebas se documentaron según el estándar definido por la organización, especificado en el Anexo 2: Plantilla para la documentación de pruebas de este documento. Los resultados de estas pruebas permitieron definir si la solución era apta para ser utilizada por los usuarios o si requería ajustes antes de ser liberada en el ambiente de producción.

3.7.4.2. Realizar el comparativo de resultados

Tomando los indicadores de tiempo que brinda la solución de ETL diseñada se realizó una comparación contra los indicadores actuales de los procesos que se realizan de forma manual actualmente, para determinar el impacto de las mejoras.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Finalmente se realizó una comprobación de la cantidad de tiempo que toma elaborar un informe de la cartera crediticia utilizando la solución desarrollada, contra el tiempo que toma elaborar el mismo informe con el proceso sin automatizar según las entrevistas realizadas a los analistas. La comprobación realizada permitió determinar si la solución propuesta cumple con la reducción de tiempo para el procesamiento de datos en al menos un 50% del tiempo actual (aproximadamente dos semanas).

La comprobación de resultados se documenta mediante la plantilla de minuta disponible en el Anexo 7: Plantilla para minuta de reunión.

Una vez definidos los aspectos metodológicos del proyecto, se procedió a realizar el análisis de los resultados obtenidos al ejecutar las fases o etapas de la iniciativa que se especificadas en este capítulo.

4

Análisis de resultados

En este capítulo se detallan los resultados obtenidos al aplicar los instrumentos y técnicas especificadas en las fases del desarrollo metodológico del proyecto.

Producto del análisis de resultados se obtuvo un conjunto de elementos que contribuyeron a la formulación de la propuesta. A continuación, se explica cada uno de ellos.

4.1. Fase I: Descubrimiento inicial

Durante esta fase se llevó a cabo la inmersión inicial del estudiante en el ambiente donde se presenta el problema de estudio. Al tratarse del primer acercamiento al problema se procedió con la recolección y documentación del material necesario para comprender la situación actual de la entidad financiera y las posibles alternativas de mejora en los procesos de creación de reportes. Los hallazgos encontrados en esta etapa se subdividen en:

- Entendimiento del negocio.
 - Identificación de necesidades.
 - Levantamiento de requisitos.
 - Mapeo de fuentes primarias
 - Mapeo de procesos actuales en creación de reportes.
- Revisión de mejores prácticas y restricciones regulatorias.
 - Compilado de mejores prácticas.
 - Compilado de restricciones regulatorias.
- Análisis de alternativas.
 - Creación del comparativo para herramientas de integración de datos.
 - Creación del comparativo para herramientas de visualización.

A continuación, se describen los resultados obtenidos para cada actividad que compone el descubrimiento inicial.

4.1.1. Entender el negocio

Esta actividad se enfocó en la recolección y estructuración del conocimiento organizacional interno por medio de entrevistas al personal de crédito y riesgo. Los principales hallazgos que se obtuvieron durante la ejecución de las actividades de entendimiento de negocio son:

4.1.1.1. Identificar las necesidades

El punto de partida de la inmersión inicial al ambiente donde se presenta el problema fue el diagnóstico de las necesidades que tienen los analistas en las áreas de crédito y riesgo con respecto al proceso de creación de reportes. El proceso de diagnóstico se llevó a cabo por medio de entrevistas al personal de ambos departamentos, producto de este proceso se identificaron los siguientes resultados. La guía de preguntas utilizada se especifica en el Apéndice B: Guía de entrevista para identificación de necesidades

Necesidades del departamento de Riesgo

El proceso de identificación de necesidades para el área de riesgos resulta fundamental debido al rol activo que desempeñan sus integrantes en el proceso que identifica comportamientos extraordinarios en la cartera de clientes. Los reportes que construye este departamento están dirigidos a identificar riesgos potenciales de liquidez, crédito, tipo de cambio y continuidad, por tanto, se requirió identificar qué necesidades poseen los analistas del área con respecto a la construcción de reportes que les permitan monitorear el estado de las operaciones de crédito en la entidad. Al finalizar el proceso de entrevistas se identificó que el departamento de riesgos busca:

- ❏ Disminuir la ejecución de tareas manuales que realizan los analistas y que comprenden la extracción, validación, integración y transformación de datos, así como la construcción y visualización de los reportes asociados al monitoreo de la cartera de clientes.
- ❏ Enfocar los esfuerzos del departamento en tareas analíticas y reducir el procesamiento manual de datos.
- ❏ Eliminar la actualización manual de los reportes que realizan mensualmente los analistas, ya que este proceso consume tiempo que podría enfocarse en el análisis de información ya procesada.

Necesidades del departamento de Crédito

El departamento de Crédito de la entidad es el responsable de mantener la cartera de clientes de la entidad en óptimas condiciones que permitan maximizar las ganancias obtenidas por concepto de comisiones y recuperación de recursos. Debido a las funciones que realiza el área de crédito es necesario mantener un flujo de información actualizado y consistente sobre la cartera de Crédito, bajo esta premisa se identifica el siguiente conjunto de necesidades para este departamento:

- ❏ Dado que la creación de reportes se ha concentrado en el personal con mayor experiencia del departamento se busca encontrar un mecanismo que permita la construcción ágil de reportes y cuyo funcionamiento permita que todo el personal del área se involucre en la creación de conocimiento, mitigando la dependencia que posee la entidad con el personal que construye los informes.
- ❏ Se busca que la construcción de reportes se realice solo una vez y que su actualización se realice automáticamente al registrarse nueva información de la cartera crediticia al cierre de mes.
- ❏ Eliminar el procesamiento manual de archivos que se requieren para elaborar un reporte, ya que esto podría inducir a error en el procesamiento de datos.

El detalle de las entrevistas para la identificación de necesidades se encuentra en el Apéndice M: Entrevista sobre necesidades con analista de riesgo crediticio, Apéndice N: Entrevista sobre necesidades con analista de riesgo cambiario y liquidez, Apéndice Ñ: Entrevista sobre necesidades con el auxiliar de riesgos, Apéndice O: Entrevista sobre necesidades con analista operativo de crédito, Apéndice P: Entrevista sobre necesidades con analista de mora, Apéndice Q: Entrevista sobre necesidades con analista de mercado y en el Apéndice R: Entrevista sobre necesidades con analista financiero

4.1.1.2. Levantar requisitos

Al finalizar el proceso de entrevistas con el personal de las áreas de negocio se identificaron una serie de requerimientos para la solución de inteligencia de negocio que procesa y visualiza los datos de la cartera crediticia de la entidad financiera.

Recolección de requerimientos para construcción de reportes de riesgos

La recolección de requerimientos realizada con los analistas del área de riesgos se puede observar en la Tabla 4.1, Tabla 4.2 y Tabla 4.3. Cabe destacar que para ambos departamentos se unificaron los resultados obtenidos en una de las plantillas que se muestran a continuación. La recolección de requisitos para el área de riesgos se puede encontrar en el Apéndice R: Entrevista sobre necesidades con analista financiero

Tabla 4.1.

Requisitos para estructura del repositorio multidimensional Riesgos.

N° de requerimiento	01
Versión del documento	01
Departamento	Área de Riesgos
Rol en el proyecto	Consultados
Enfoque del requerimiento	Estructura del repositorio multidimensional
Descripción del requerimiento	
Consolidar la información de las carteras crediticias en una base de datos única que permita la construcción de informes basados en los datos de clientes y operaciones de la entidad.	
Precondiciones del requerimiento	
Verificar que los archivos que se incluirán en la solución cuenten con toda la información solicitada.	
Detalle del requerimiento	
Incluir en la solución los siguientes datos:	
<input type="checkbox"/>	Descripción del producto y subproducto.
<input type="checkbox"/>	Días de atraso.
<input type="checkbox"/>	Estado de la operación.
<input type="checkbox"/>	Monto de estimación genérica.
<input type="checkbox"/>	Monto de estimación específica saldo con garantía.
<input type="checkbox"/>	Monto de estimación específica saldo sin garantía.
<input type="checkbox"/>	Monto total de la deuda colonizado y dolarizado.
<input type="checkbox"/>	Moneda de la operación.
<input type="checkbox"/>	Nombre del cliente.
<input type="checkbox"/>	Identificación.
<input type="checkbox"/>	Categoría de Riesgo.
<input type="checkbox"/>	Actividad.
<input type="checkbox"/>	Indicador para generadores de divisa.

Continuación de la tabla 4.1

Restricciones a tomar en cuenta

En los archivos originales los montos de las operaciones no se encuentran colonizados ni dolarizados, en su lugar se especifican el importe de la operación y una columna que indica la moneda en la que se registró la operación.

Nota: elaboración propia.

Tabla 4.2.

Requisitos para procesamiento de datos Riesgos.

N° de requerimiento	02
Versión del documento	01
Departamento	Área de Riesgos
Rol en el proyecto	Consultados
Enfoque del requerimiento	Procesamiento de datos

Descripción del requerimiento

Eliminar el procesamiento manual requerido para extraer, modificar y estructurar los datos que se utilizan en los reportes de concentración de la cartera crediticia.

Precondiciones del requerimiento

Se debe contar con la información correspondiente a la cartera crediticia de los meses bajo análisis.

Detalle del requerimiento

Se debe dolarizar el dato del monto total de la deuda colonizado, utilizando el tipo de cambio de referencia brindado por el Banco Central de Costa Rica al cierre del mes que se está procesando.

Unificar la información de las columnas de categoría de Riesgo que son alfanuméricas para que incluyan letras únicamente.

Convertir la columna de días de atraso a un rango de valores de 0 a 30, 31 a 60, 61 a 90, 91 a 120 y más de 120 días.

Continuación de la tabla 4.2

Detalle del requerimiento

Convertir la columna del indicador para generadores de divisa a texto bajo los siguientes criterios, el valor 1 se convierte a Generador, el 2 a No generador de bajo riesgo, el 3 a No generador de alto riesgo y el 4 a No aplica.

Restricciones a tomar en cuenta

La información de insumo debe procesarse al finalizar el cierre operativo mensual.

Nota: elaboración propia.

Tabla 4.3.

Requisitos para visualizaciones de datos Riesgos

N° de requerimiento	03
Versión del documento	01
Departamento	Área de Riesgos
Rol en el proyecto	Consultados
Enfoque del requerimiento	Visualizaciones de datos

Descripción del requerimiento

Visualizar la información histórica de la cartera crediticia para determinar comportamientos anómalos en las operaciones de crédito o los clientes de la entidad financiera.

Precondiciones del requerimiento

Contar con la base de datos integrada de operaciones históricas de la cartera de crédito que contemple el procesamiento requerido por el departamento.

Detalle del requerimiento

Se busca construir reportes gráficos que permitan observar la tendencia histórica de la cartera crediticia de la organización.

Adicionalmente, se debe permitir que los gráficos sean interactivos al ofrecer la posibilidad de navegar al detalle de la información que se presenta en las gráficas del reporte.

Continuación de la tabla 4.3

Restricciones a tomar en cuenta

Se debe evitar el uso de gráficos en tres dimensiones y de tipo pastel, ya que generalmente no permiten la interpretación correcta de los datos.

Nota: elaboración propia.

Seguidamente se procedió a entrevistar al personal del departamento de Crédito para determinar sus requerimientos con respecto a la solución que se plantea en la entidad financiera.

Recolección de requerimientos para construcción de reportes de crédito

El detalle de la recolección de requerimientos realizada a los analistas del área de crédito se puede observar en la Tabla 4.4, Tabla 4.5 y Tabla 4.6. La recolección de requisitos para el área de riesgos se puede encontrar en el Apéndice S: Recolección de requisitos con el área de riesgos de este documento.

Tabla 4.4.

Requisitos para estructura del repositorio multidimensional Crédito.

N° de requerimiento	01
Versión del documento	01
Departamento	Área de Crédito
Rol en el proyecto	Consultados
Enfoque del requerimiento	Estructura del repositorio multidimensional
Descripción del requerimiento	
Incorporar los datos de las carteras de crédito en una herramienta única que permita centralizar el acceso a los datos para el área de Crédito, evitando a su vez el uso de archivos independientes para la construcción de reportes en el área.	
Precondiciones del requerimiento	
Se debe tener en cuenta la extracción de los archivos utilizados para construir los reportes debe ejecutarse posterior al cierre mensual de operaciones de crédito.	

Continuación de la tabla 4.4

Detalle del requerimiento

Se debe incluir en la solución las siguientes columnas del archivo de cartera:

- Monto total de la deuda dolarizado y colonizado.
- Monto mitigador.
- Monto principal directo y contingente por moneda.
- Monto producto colonizado.
- Monto de cuenta.
- Nombre del ejecutivo.
- Identificación.
- Nombre del cliente.
- Actividad.
- Descripción del producto.
- Estado de la operación.
- Días de atraso.
- Moneda.

Restricciones a tomar en cuenta

Existen operaciones que no tienen un ejecutivo asignado en las carteras crediticias.

Nota: elaboración propia.

Tabla 4.5.

Requisitos para procesamiento de datos Crédito.

N° de requerimiento	02
Versión del documento	01
Departamento	Área de Crédito
Rol en el proyecto	Consultados
Enfoque del requerimiento	Procesamiento de datos

Descripción del requerimiento

Se busca una alternativa para la construcción de reportes que permita analizar tendencias históricas en el comportamiento de los clientes por producto, días de atraso, estado, actividad y moneda de la operación.

Precondiciones del requerimiento

Se requiere que previo a la construcción del reporte se organicen las mantas o archivos de datos crediticios de la entidad financiera.

Continuación de la tabla 4.5

Detalle del requerimiento

Se debe dolarizar el monto total de la deuda utilizando el tipo de cambio de referencia de compra del Banco Central reportado al cierre de mes.

Se busca segmentar los valores correspondientes a días de atraso para determinar cuántas operaciones se encuentran en los rangos de 0-30, 31-60 días y 61-90 días de atraso.

Restricciones a tomar en cuenta

La información que se utiliza como insumo debe procesarse siempre después del cierre operativo mensual.

Nota: elaboración propia.

Tabla 4.6.

Requisitos para visualizaciones de datos Crédito.

N° de requerimiento	03
Versión del documento	01
Departamento	Área de Crédito
Rol en el proyecto	Consultados
Enfoque del requerimiento	Visualizaciones de datos

Descripción del requerimiento

Visualizar la información correspondiente a la cartera de crédito de forma clara, adicionalmente se busca determinar tendencias en los datos que habiliten la toma de decisiones certera en el departamento.

Precondiciones del requerimiento

Información procesada según las reglas de negocio establecidas anteriormente en la entrevista.

Continuación de la tabla 4.6

Detalle del requerimiento

Se requiere que las visualizaciones permitan diferenciar claramente la evolución interanual de la cartera de crédito.

Se busca que las visualizaciones de datos mantengan una paleta de colores oscura para los fondos y de colores cálidos para la información analizada, de esta forma se pueden resaltar adecuadamente los datos principales del reporte.

Restricciones a tomar en cuenta

No registran restricciones para este requerimiento.

Nota: elaboración propia.

Una vez establecidos los requerimientos de las áreas de negocio se procedió a determinar las fuentes de datos que se debían incorporar en la propuesta de solución.

4.1.1.3. Mapear las fuentes primarias

Previo a realizar el diseño e implementación de la propuesta de solución se llevó a cabo la verificación de las fuentes de datos que se debían tomar en cuenta para la construcción del proceso automatizado de ETL. El mapeo de las fuentes de datos asociadas a la construcción de reportes de la entidad se realizó por medio de entrevistas al personal de las áreas de crédito y riesgo de la organización. El detalle de la información recolectada se encuentra especificada en el Apéndice Y: Entrevista sobre fuentes transaccionales con analista de riesgo crediticio, Apéndice Z: Entrevista sobre fuentes transaccionales con analista de riesgo cambiario y liquidez y Apéndice AA: Entrevista sobre fuentes transaccionales con auxiliar de riesgo para el área de riesgo y en el Apéndice U: Entrevista sobre fuentes transaccionales con analista operativo de crédito, Apéndice V: Entrevista sobre fuentes transaccionales con analista de mora, Apéndice W: Entrevista sobre fuentes transaccionales con analista de mercado y Apéndice X: Entrevista sobre fuentes transaccionales con analista financiero para el área de crédito.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Entrevista al área de Riesgo sobre las fuentes de datos para sus reportes

Este conjunto de entrevistas buscaba determinar cuáles son las fuentes de datos que utilizan los analistas del departamento de riesgos para construir sus reportes. Los hallazgos principales fueron:

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Para la elaboración de reportes se toman en cuenta fuentes internas tales como los archivos de la cartera crediticia donde se registran las operaciones de crédito tramitadas durante el mes.

En cuanto a fuentes externas se procesan los datos del comportamiento crediticio de los clientes en el sector financiero nacional.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

La información que se recibe de las áreas operativas de crédito se gestiona por medio de Excel o archivos de texto plano.

La información de fuentes externas generalmente se maneja en formato XML o texto plano que se procesan por medio de libros de Excel.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

Se utilizan principalmente los archivos de cartera que se reportan a los entes regulatorios al final de cada mes y algunos archivos auxiliares tales como los XML que se descargan del Centro de Información Crediticia (CIC) donde se almacena la información de todos los deudores a nivel nacional.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

En el caso de los archivos de la cartera crediticia se solicitan al área operativa correspondiente una vez realizado el cierre mensual.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Los tiempos de extracción para los datos que provienen de fuentes externas varían según la cantidad de registros que se deben descargar. En el caso de los datos del archivo CIC para la totalidad de la cartera crediticia de la organización el tiempo máximo de extracción ha sido de 12 horas.

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

La información que proviene de los archivos CIC no se depura ya que proviene de fuentes oficiales del sector financiero regulado a nivel nacional.

Por otro lado, la información de la cartera crediticia ha presentado algunas inconsistencias anteriormente donde las columnas de clasificación de producto y algunos saldos de operación no poseen datos en los archivos. Los errores identificados se notifican al área operativa correspondiente para su corrección inmediata y a nivel del archivo que se está procesando se realiza el ajuste requerido.

En cuanto a procesamiento de la información comúnmente se dolarizan saldos, montos de operación aprobados, se crean rangos para identificar de forma sencilla clientes generadores y no generadores de moneda extranjera y se agrupan operaciones que no poseen un ejecutivo bajo una clasificación separada que permite discernir los casos específicos en análisis posteriores.

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Los datos internos se actualizan bajo un corte mensual, una vez que el departamento operativo de crédito finaliza la construcción del informe de cartera, que se distribuye por lo general una semana después de finalizar el cierre operativo.

Entrevista al área de Crédito sobre la fuente de datos para sus reportes

Este conjunto de entrevistas buscaba determinar cuáles son las fuentes de datos que utilizan los analistas del departamento de crédito para construir sus reportes. Los principales hallazgos encontrados son:

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Los reportes que se construyen en el área de crédito se encuentran basados en información interna de las operaciones crediticias que se han registrado desde el cambio del Core bancario de la entidad.

En cuanto a fuentes externas el departamento de crédito utiliza la información disponible en el sistema que monitorea el comportamiento de las carteras de clientes en las instituciones competencia de la organización a nivel nacional.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

Los datos procesados se encuentran en archivos Excel ya que este es el formato designado por el área técnica para extraer la información del sistema maestro de datos que posee la entidad.

¿Cuál son las fuentes primarias de información para la construcción de sus informes?

La fuente principal para la construcción de reportes es la cartera crediticia que se genera posterior al cierre mensual.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

La extracción de datos se realiza en 30 minutos aproximadamente, sin embargo, la verificación y depuración de la cartera puede tomar alrededor de un día, debido a que se debe asegurar la consistencia de los datos para la construcción de los informes.

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

La información extraída se encuentra en su estado natural, esto implica que debido a las restricciones y diseño del sistema maestro de datos existe información que se maneja por medio de indicadores numéricos que no son representativos o explícitos para las áreas de negocio, por lo tanto, se debe realizar la conversión de estos códigos numéricos a valores textuales que representen la información de análisis.

En el caso de las operaciones que poseen valores nulos se procede a notificar al responsable de ingresar la información operativamente para que verifique el caso y realice las correcciones respectivas,

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Los datos que se utilizan para construir los reportes del área se actualizan de forma mensual ya que posterior al cierre es cuando se tiene certeza de que las operaciones poseen información correcta.

Resultados obtenidos para las fuentes de datos

Una vez identificadas las fuentes de datos que se debían tomar en cuenta para construir la propuesta del proyecto, se realizó el perfilado de las fuentes de datos. El proceso de perfilado inició con la identificación del conjunto de archivos que poseen los datos de las operaciones crediticias registradas durante cada mes operativo de la entidad. Cabe destacar que estos archivos poseen la información consolidada de todas las operaciones de Crédito que se realizan durante un mes natural.

Partiendo de la estructura que poseen los archivos de datos de la cartera crediticia se establece que la información que se debe extraer es:

- | | |
|--|--|
| <input type="checkbox"/> Monto total de la deuda en colones. | <input type="checkbox"/> Monto estimación específica saldo con garantía. |
| <input type="checkbox"/> Monto total de la deuda en dólares (campo calculado). | <input type="checkbox"/> Nombre y código del ejecutivo asociado al cliente que realiza la transacción. |
| <input type="checkbox"/> Monto principal directo moneda. | <input type="checkbox"/> Descripción producto. |
| <input type="checkbox"/> Monto principal contingente moneda. | <input type="checkbox"/> Descripción subproducto. |
| <input type="checkbox"/> Monto producto colonizado. | <input type="checkbox"/> Días de atraso. |
| <input type="checkbox"/> Monto de cuenta. | <input type="checkbox"/> Estado de operación. |
| <input type="checkbox"/> Monto mitigador. | <input type="checkbox"/> Moneda de la operación. |
| <input type="checkbox"/> Monto estimación genérica. | |
| <input type="checkbox"/> Monto estimación específica saldo sin garantía. | |

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Adicionalmente, se tomó en cuenta la información almacenada de los clientes que posee la entidad en su histórico de bases de datos. Las variables principales en este almacén son:

- ☐ Identificación.
- ☐ Identificador único a nivel de la entidad financiera.
- ☐ Nombre del cliente.
- ☐ Indicador para generadores de divisa.
- ☐ Actividad del cliente.
- ☐ Categoría de Riesgo.

Para cada dato mencionado anteriormente se determinó su tipo (numérico o textual) y los posibles rangos de valores para los casos donde un atributo específico de una operación de crédito o de un cliente posee un conjunto finito de valores. La Tabla 4.7 muestra un resumen de los atributos que integran cada fuente de datos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Tabla 4.7.

Descripción de las fuentes de datos

Nombre del campo en la fuente de datos	Tipo del dato	Fuente de datos	Rango de valores
Monto total de la deuda colones	Numérico	Archivos de cartera crediticia	No tiene
Monto total de la deuda dólares	Numérico	Archivos de cartera crediticia	No tiene
Monto principal directo moneda	Numérico	Archivos de cartera crediticia	No tiene
Monto principal contingente moneda	Numérico	Archivos de cartera crediticia	No tiene
Monto producto colonizado	Numérico	Archivos de cartera crediticia	No tiene
Monto de cuenta	Numérico	Archivos de cartera crediticia	No tiene
Monto mitigador	Numérico	Archivos de cartera crediticia	No tiene
Monto estimación genérica	Numérico	Archivos de cartera crediticia	No tiene
Monto estimación específica saldo sin garantía	Numérico	Archivos de cartera crediticia	No tiene
Monto estimación específica saldo con garantía	Numérico	Archivos de cartera crediticia	No tiene
Nombre del ejecutivo	Textual	Archivos de cartera crediticia	No tiene
Código del ejecutivo	Numérico	Archivos de cartera crediticia	No tiene
Descripción producto	Textual	Archivos de cartera crediticia	No tiene

Continuación de la tabla 4.7

Nombre del campo en la fuente de datos	Tipo del dato	Fuente de datos	Rango de valores
Descripción subproducto	Textual	Archivos de cartera crediticia	No tiene
Días de atraso	Numérico	Archivos de cartera crediticia	No tiene
Estado de la operación	Textual	Archivos de cartera crediticia	Activa, Cobro, Judicial, Vencida, Vigente
Moneda de la operación	Textual	Archivos de cartera crediticia	USD, CRC, EUR
Identificación	Textual	Base de datos de clientes	Puede incluir cédulas de residencia o pasaporte
BP	Numérico	Base de datos de clientes	No tiene
Nombre del cliente	Textual	Base de datos de clientes	No tiene
Indicador para generadores de divisa	Numérico	Base de datos de clientes	1,2,3,4 que corresponde a Generador, No generador de bajo riesgo, no generador de alto riesgo y no aplica
Actividad del cliente	Textual	Base de datos de clientes	No tiene
Categoría de Riesgo	Textual	Base de datos de clientes	A1,A2,B1,B2,C1,C2,D,E

Nota: elaboración propia.

Los atributos que poseen rangos finitos de valores son exclusivos es decir un cliente específico no puede tener dos categorías de riesgo asociadas al mismo tiempo.

4.1.1.4. Mapear los procesos actuales de creación de reportes

El diagnóstico de la situación actual de los procesos de construcción de reportes en la organización se llevó a cabo mediante entrevistas a los analistas de las áreas de Riesgo y Crédito que se dedican la construcción y análisis de informes. El primer grupo de entrevistas se

enfocaron en conocer y diagnosticar la operativa correspondiente a la creación de los reportes en las dos áreas, con el objetivo de identificar puntos de mejora en el procesamiento de los datos crediticios.

A continuación, se muestran los resultados obtenidos al aplicar la guía de preguntas listada en el Apéndice E: Guía de entrevista para mapeo de los procesos de elaboración de reportes.

Resultados de las entrevistas al personal de Riesgos

Las entrevistas se realizaron de manera individual para asegurar que la información brindada por los analistas del departamento fuera precisa y no se viera afectada por experiencias grupales. El detalle de las entrevistas se encuentra en el Apéndice BB: Entrevista para mapeo de procesos para elaboración de reportes analista de riesgo crediticio, Apéndice CC: Entrevista para mapeo de procesos para elaboración de reportes analista de riesgo de liquidez y riesgo cambiario y en el Apéndice DD: Entrevista para mapeo de procesos para elaboración de reportes auxiliar de riesgo. Las observaciones o resultados principales para cada pregunta son:

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes construidos se dirigen a diversos comités que se realizan semanal y mensualmente en la organización, adicionalmente, los resultados principales de los análisis de datos se presentan a la Junta Directiva de la organización con una periodicidad mensual.

¿Cuál es la periodicidad con la que elabora estos reportes?

La mayoría de los reportes se realizan de manera mensual debido a los procesos de cierre operativo, sin embargo, se llevan a cabo reportes crediticios bajo demanda a lo largo del mes.

¿Qué actividades lleva a cabo para construir sus reportes?

Pese a que la construcción de los reportes se realiza de distintas maneras en el caso de cada analista se lograron identificar una serie de pasos en común. La Figura 4.1 muestra las actividades que componen el proceso general para la creación de reportes en el área de Riesgo.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

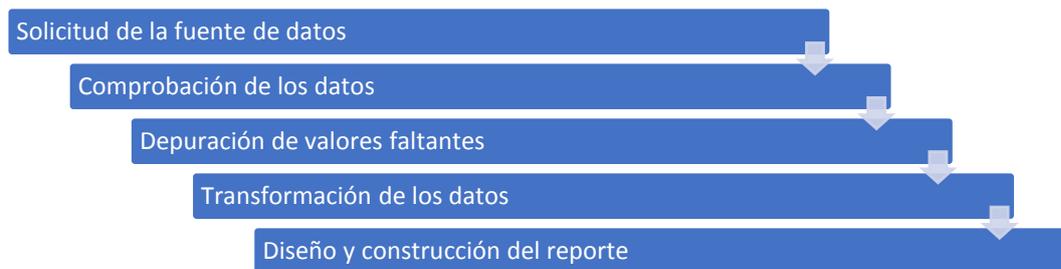


Figura 4.1. Secuencia de actividades para la construcción de un reporte área de Riesgo.

Nota: elaboración propia.

A continuación, se describen las observaciones realizadas por los analistas del área crediticia para cada una de las actividades.

- ❏ Solicitud de la manta o archivo de datos: en esta actividad se realiza la petición de los datos requeridos para construir el reporte (no incluye el tiempo que le toma al área gestora del archivo la construcción del mismo).
- ❏ Comprobación de los datos: en este paso se realiza la comprobación de los valores disponibles en la manta o archivo contra la información de los reportes contables organizacionales, en caso de encontrar diferencias se notifica al área correspondiente para que realice las verificaciones y correcciones del caso.
- ❏ Depuración de valores faltantes: en este caso los analistas no toman en cuenta valores nulos para sus análisis en su lugar estos se extraen de la manta principal y se registran en un archivo separado para reportarlos al área respectiva.
- ❏ Transformación de los datos (aplicación de las reglas de negocio): Los analistas ejecutan transformaciones sobre la información que incluye comúnmente dolarización y colonización de saldos, formateo de fechas y transformación de códigos numéricos a rangos descriptivos de texto, como por ejemplo los días de atraso y el indicador para generadores de divisa.
- ❏ Diseño y construcción del reporte en Excel: una vez finalizado el procesamiento y depuración de los datos, los analistas proceden a diseñar el informe requerido donde seleccionan las variables que desean representar y los valores a resaltar en el reporte.

Seguidamente toman la información procesada y aplican un diseño tabular en la herramienta Excel.

- Implementar representación visual: con los datos tabulados en Excel los analistas proceden a construir la representación visual de la información, generalmente por medio de gráficos lineales (de gran utilidad para demostrar el comportamiento de la cartera crediticia).

¿En promedio cuánto tiempo invierte en elaborar cada informe?

En este caso se tomaron en cuenta las actividades comunes identificadas, se solicitó a cada analista que especificara el tiempo que dedica a cada actividad con el objetivo de identificar tiempos promedio para la construcción de reportes. Los resultados obtenidos se muestran en la Tabla 4.8, los valores se expresaron en minutos para mantener consistencia entre las unidades de medida de cada valor.

Tabla 4.8.

Tiempo promedio para la construcción de informes – área de riesgos

Actividad	Analista uno	Analista dos	Analista tres
Solicitud de la manta o archivo de datos	2 minutos	2 minutos	2 minutos
Comprobación de los datos	30 minutos	45 minutos	20 minutos
Depuración de valores faltantes	120 minutos	30 minutos	20 minutos
Transformación de los datos	180 minutos	80 minutos	80 minutos
Diseño y construcción del reporte en Excel	60 minutos	90 minutos	80 minutos
Realizar representación visual de los datos	10 minutos	15 minutos	10 minutos
Total	402 minutos	262 minutos	212 minutos

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Los tiempos representados corresponden a la construcción del reporte únicamente, en este caso se excluyó el tiempo que toma al departamento crediticio la elaboración de la manta o archivo de datos.

Tomando los tiempos de elaboración indicados por los analistas se puede calcular el tiempo promedio que le toma al departamento construir sus reportes. El cálculo del tiempo promedio se muestra a continuación.

$$\frac{402 \text{ minutos} + 262 \text{ minutos} + 212 \text{ minutos}}{3}$$
$$\frac{876 \text{ minutos}}{3}$$

292 minutos, tiempo promedio para la elaboración de un informe

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

En cuanto a oportunidades de mejora se identificó que el área de Riesgo establece como oportunidad de mejora la reducción del procesamiento manual en la construcción de los informes, ya que esto implica una mayor carga operativa e incrementa la probabilidad de cometer errores durante la elaboración del reporte.

Otra oportunidad de mejora mencionada por los analistas fue la eliminación del paso de solicitud de datos a otras áreas de negocio al sustituir esta solicitud por la extracción automática de datos que favorezca la construcción oportuna de los informes.

¿De qué forma gestiona los informes de índole histórica?

En el caso de los reportes sobre datos históricos algunos analistas mencionan que representan una complejidad mayor al tener que procesar grandes cantidades de información en periodos prolongados de tiempo lo que previamente ha provocado inconsistencias en los informes.

Otros analistas mencionan que la manera en la que gestionan la construcción de informes sobre datos históricos es por medio de archivos formulados en Excel donde realizan la sustitución de la base de datos fuente y la herramienta se encarga de construir los indicadores requeridos. Mencionan que este enfoque resulta más práctico que procesar los datos desde cero ya que en ocasiones la transformación de datos consume parte del tiempo dedicado al análisis de los reportes.

Resultados de las entrevistas al personal de Crédito

Al igual que en el caso del departamento de riesgos, se realizaron un conjunto de entrevistas individuales para asegurar que la información brindada por los analistas fuera precisa y no se viera afectada por experiencias grupales. El detalle de las entrevista se encuentra en el Apéndice EE: Entrevista para mapeo de procesos para elaboración de reportes – analista operativo de crédito, Apéndice FF: Entrevista para mapeo de procesos para elaboración de reportes analista de mora, Apéndice GG: Entrevista para mapeo de procesos para elaboración de reportes analista de mercado y en el Apéndice HH: Entrevista para mapeo de procesos para elaboración de reportes analista financiero Las observaciones o resultados principales para cada pregunta son:

¿Qué actividades lleva a cabo para construir sus reportes?

Pese a que la construcción de los reportes se realiza de distintas maneras en el caso de cada analista se identificaron una serie de pasos en común que realizan todos los integrantes del equipo de análisis. Los pasos principales identificados se muestran en la Figura 4.2.



Figura 4.2. Secuencia de actividades para la construcción de un reporte área de Crédito.

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

A continuación, se describen las observaciones realizadas por los analistas del área crediticia para cada una de las actividades.

- ❏ Extracción de datos: en este paso se realiza la selección y extracción de los datos desde el core bancario de la organización que es la fuente principal de datos donde se registran todas las transacciones de la entidad.
- ❏ Depuración de datos: esta actividad comprende la revisión e identificación de valores disonantes en las fuentes de datos que deben corregirse para el procesamiento del cierre mensual.
- ❏ Transformación de datos: en cuanto a transformación de datos el área crediticia maneja saldos totalmente dolarizados o totalmente colonizados, la tarea de conversión de divisas es una de las transformaciones preponderantes en este proceso, otras transformaciones incluyen la categorización de códigos numéricos y la agregación de saldos.
- ❏ Creación del reporte: en esta actividad se tabula la información procesada bajo el formato ya establecido por el personal experimentado del departamento lo que propicia la estandarización y facilita la construcción visual del informe posteriormente.
- ❏ Creación del reporte visual: la elaboración del reporte se realiza por medio de la herramienta Excel donde se toman los datos tabulados para crear gráficos lineales que ejemplifican la tendencia de la cartera de clientes desde el punto de vista de recuperaciones y colocaciones de créditos.

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes que se generan en el área se dirigen a diversos comités crediticios, Junta Directiva, jefaturas y gerentes de la entidad.

¿Cuál es la periodicidad con la que elabora estos reportes?

Los reportes del área se realizan de forma mensual ya que posterior al cierre mensual es cuando se puede tener certeza sobre el estado de la cartera. Los analistas mencionan que de realizar su análisis previo al cierre operativo la información podría no reflejar correctamente la situación actual de la entidad financiera.

¿En promedio cuánto tiempo invierte en elaborar cada informe?

En este caso se tomaron en cuenta las actividades comunes identificadas, se solicitó a cada analista que mencionara el tiempo que dedica a cada actividad con el objetivo de identificar tiempos promedio para la construcción de reportes. Los resultados obtenidos se muestran en la Tabla 4.9. Los valores de la tabla se expresaron en minutos para mantener consistencia entre los resultados obtenidos.

Tabla 4.9.

Tiempo para elaboración de informes crédito

Actividad	Analista uno	Analista dos	Analista tres	Analista cuatro
Extracción de datos		30 minutos		
Depuración de datos	30 minutos	60 minutos	25 minutos	60 minutos
Transformación de datos	60 minutos	180 minutos	90 minutos	60 minutos
Creación del reporte	30 minutos	20 minutos	60 minutos	30 minutos
Creación del reporte visual	10 minutos	25 minutos	15 minutos	20 minutos
Total	160 minutos	315 minutos	190 minutos	170 minutos

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Tomando los tiempos de elaboración indicados por los analistas se puede calcular el tiempo promedio que le toma al departamento construir un reporte. El resultado del cálculo para determinar el tiempo promedio se muestra a continuación.

$$\frac{160 \text{ minutos} + 315 \text{ minutos} + 190 \text{ minutos} + 170 \text{ minutos}}{4}$$
$$\frac{835 \text{ minutos}}{4}$$

208,75 minutos, tiempo promedio para la elaboración de un informe

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Se identifica como oportunidad de mejora encontrar un mecanismo para integrar la información de los archivos o fuentes de datos, evitando el manejo de múltiples archivos de Excel independientes.

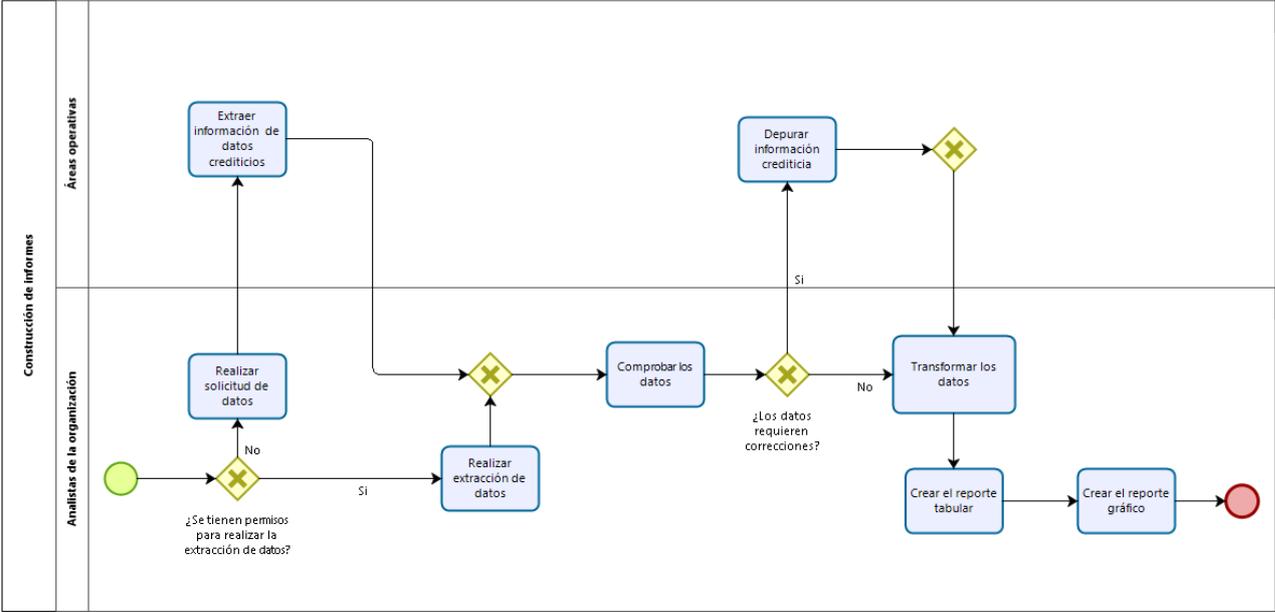
¿De qué forma gestiona los informes de índole histórica?

En el caso de los informes históricos se mantiene un archivo consolidado con la información procesada anteriormente, cuando se requiere agregar información del nuevo periodo se construye el reporte mensual y se integra al consolidado.

Utilizando la información recolectada durante las entrevistas se procedió a diagramar el proceso general para la elaboración de reportes de la entidad financiera.

Al unificar las actividades identificadas durante las entrevistas a las áreas se estableció el proceso que enmarca la secuencia de pasos para la elaboración de los reportes a nivel crediticio de las áreas de negocio. La documentación del proceso se realizó por medio de la notación BPMN 2.0 y se representa en la Figura 4.3. La explicación de los elementos que conforman esta notación se especifica en el Anexo 1: Notación estándar BPMN 2.0 de este documento.

Propuesta para la elaboración de reportes crediticios de una entidad financiera



Powered by bizagi Modeler

Figura 4.3. Proceso general para elaboración de reportes.

Nota: elaboración propia utilizando la herramienta Bizagi.

Como se puede observar en el proceso cuando las áreas de análisis no pueden extraer la información crediticia por su cuenta deben solicitarla a las áreas operativas lo que provoca un incremento en el tiempo dedicado a la creación del reporte lo que repercute en la entrega final.

4.1.2. Revisar las mejores prácticas y restricciones regulatorias

En esta actividad del proyecto se procedió a corroborar las mejores prácticas y restricciones regulatorias que impactan directamente al proyecto. A continuación se describen los resultados encontrados.

4.1.2.1. Crear el compilado de restricciones regulatorias

Previo al desarrollo del proyecto se destacó la importancia de estudiar la normativa vigente relacionada al procesamiento y gestión de datos por medio de herramientas tecnológicas. Para la construcción del consolidado se realizó una revisión documental donde se verificaron aquellas restricciones que delimitan el desarrollo de los componentes que integran la propuesta de solución. En la Tabla 4.10 se observa un resumen de los reglamentos y normativas vinculadas al objeto del proyecto.

Tabla 4.10.

Resumen de artículos tomados en cuenta para las restricciones regulatorias del proyecto.

Normativa	Ley Número	Artículos tomados en cuenta
Reglamento de Riesgo Operativo	Acuerdo SUGEF 18-16	13
Reglamento General de Gestión de la Tecnología de Información	Acuerdo SUGEF 14-17	19
Ley de protección de datos	Ley n° 8968	11

Nota: elaboración propia.

A continuación, se presenta un breve resumen del contenido correspondiente a cada artículo mencionado en la Tabla 4.10.

Reglamento de Riesgo Operativo – Artículo 13

El reglamento de Riesgo Operativo establece en su artículo 13 que la entidad financiera debe contar con políticas y procedimientos de gestión y seguridad de la información, que consideren entre otros aspectos:

- a) La autenticación para el acceso lógico a los sistemas y servicios informáticos internos y externos.
- b) La conservación ordenada, completa, íntegra, oportuna de la información y documentación (registros) que soporta las operaciones de la entidad.
- c) La divulgación y uso no autorizado de información confidencial o protegida por ley.

Este artículo específicamente en el ítem b se refiere al mantenimiento que se debe brindar a la información que brinda soporte a las operaciones de la entidad. Sobre este artículo se destaca la importancia de mantener un flujo de datos consistente y preciso en las soluciones que gestionan y procesan los datos de cliente u operaciones de la organización.

Reglamento General de Gestión de la Tecnología de Información – Artículo 19

En el artículo 19 de este reglamento se menciona que las bases de datos actualizadas, así como, las aplicaciones vigentes que procesan o dan acceso a estas bases, pueden mantenerse en servicios de computación en la nube, siempre y cuando se cumplan con los requisitos legales, de seguridad y de acceso del supervisor, de acuerdo a la normativa aplicable por cada superintendencia.

La respectiva superintendencia puede requerir un modelo de gestión de infraestructura tecnológica diferente al de los servicios de computación en la nube, cuando en estos: la entidad no cumpla los requisitos legales y de seguridad; no se brinde acceso al supervisor; la información que la entidad desea mantener sea sensible o crítica para la continuidad del negocio; la computación en la nube represente un riesgo para el sistema financiero; o cuando afecte los intereses de los clientes.

En este artículo del reglamento se menciona que no existe restricción sobre los servicios en la nube a los que puede suscribirse la organización, siempre y cuando se respeten los lineamientos legales, de seguridad y acceso de la información almacenada o procesada dentro de este tipo de esquemas de almacenamiento. Sobre este artículo se destaca la posibilidad de gestionar la información de los clientes mediante servicios en la nube, siempre y cuando esto no signifique un riesgo de divulgación para la entidad financiera o sus clientes.

Ley de protección de datos – Artículo 11

La ley de protección de datos en su artículo 11 establece que la persona responsable y quienes intervengan en cualquier fase del tratamiento de datos personales están obligadas al secreto profesional o funcional, aún después de finalizada su relación con la base de datos.

En este artículo de la ley se hace referencia a que las personas que intervengan en el tratamiento de datos personales (información de los clientes en el caso del presente proyecto) deben mantener la confidencialidad de los datos que se ponen a su disposición sin que esto signifique un riesgo para ellos o la entidad para la que laboran.

Dado que la información procesada en la solución propuesta es de índole personal y contiene además información financiera se requiere mantener un alto grado de confidencialidad en los datos.

Una vez establecidas las principales restricciones del proyecto se realizó la investigación de mejores prácticas para los proyectos de *Data warehousing*.

4.1.2.2. Crear el compilado de mejores prácticas

Con el objetivo de plantear una propuesta de solución óptima a la entidad financiera se realizó el estudio documental de mejores prácticas para el desarrollo de proyectos de *Data warehousing*. A continuación, se presentan prácticas recomendadas por los principales autores en esta rama de la inteligencia de negocio.

Ponniah (2001) menciona que los modelos multidimensionales se pueden implementar por medio de bases de datos relacionales o herramientas OLAP, favoreciendo la rapidez en la ejecución de las consultas sobre la base de datos objetivo.

Otro aspecto que favorece la implementación del modelo multidimensional en un RDBMs es la popularidad de esta tecnología en diversas industrias. La popularidad de las bases de datos relacionales propicia la existencia de una amplia variedad para opciones de soporte y corrección de errores bajo un esquema de tiempo menor con respecto a otras tecnologías.

En cuanto a los metadatos de la solución, Lwayer y Chowdhury (2004) aconsejan que “los metadatos deben mantener nombres y definiciones de negocio apropiadas, así como nombres técnicos estandarizados a nivel de la base de datos” (P.5). El mantener la estandarización sobre este componente de la solución propicia el uso de un lenguaje común para los elementos que constituyen el ambiente del *Data mart*, favoreciendo la comunicación efectiva entre los colaboradores técnicos y de negocio que interactúan con el DM.

Kimball y Ross (2014) recomiendan para iniciar con la implementación de soluciones de *data warehousing* se deben tomar en cuenta los siguientes puntos:

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ Los equipos involucrados en el desarrollo de la solución deben comprender por completo la estructura, contenido, relaciones y reglas de las fuentes de datos.
- ❏ Se debe corroborar que la información se encuentre en un estado adecuado para su consumo.
- ❏ Los nombres o etiquetas del modelo de datos deben ser descriptivos y consistentes con la perspectiva del negocio.

Ponniah (2001) recomienda que desde la fase de requerimientos se evalúe cuales fuentes de datos podrían presentar problemas en su información y los tipos de problemas en calidad de datos que podrían encontrarse en los repositorios. Adicionalmente este autor hace referencia dos prácticas adicionales para las soluciones de inteligencia de negocio relacionadas al procesamiento durante el ETL y a las cargas de datos al repositorio histórico de información.

- ❏ Las soluciones de *Data warehousing* comúnmente utilizan un área intermedia para el procesamiento de información llamada *staging area* que permite mantener un respaldo de la información transformada durante la ejecución del ETL.
- ❏ Las cargas de información que actualizan los datos del modelo multidimensional pueden seguir dos enfoques, uno de actualización para cargas incrementales y otro para refrescar la totalidad de los datos en intervalos específicos de tiempo. El autor menciona que la selección del enfoque debe tomar en consideración la cantidad de registros por carga incremental contra la cantidad de registros que se deben eliminar e insertar cuando se opta por un enfoque donde se refresca la totalidad de la información.

La Tabla 4.11 muestra la validación de las buenas prácticas descritas anteriormente para contra las necesidades o requerimientos de la organización.

Tabla 4.11.

Verificación de factibilidad de buenas prácticas

Buenas prácticas	Necesidad o requerimiento	Viabilidad	Mecanismos de implementación
Mantener los datos procesados del ETL en el área de preparación (staging area) como respaldo hasta su traslado al DM.	Mantener un histórico de la información procesada para evitar el procesamiento recurrente de archivos.	Si	Se planteó realizar respaldos de la información en un motor de base de datos relacional o bien en ficheros planos
Perfilado de datos para identificar potenciales problemas en las fuentes de información.	Identificar la fuente de datos que se deben incorporar en la creación de reportes.	Ya ejecutado	Producto de los proyectos de limpieza de datos se han identificado problemas recurrentes en la información. Para efectos de este proyecto la calidad de datos encuentra fuera de alcance.
Gestión de metadatos en la solución en el repositorio del <i>Data mart</i> .	Se requiere documentar el conocimiento relacionado a la creación de reportes en las áreas de Crédito y Riesgo de la entidad.	Si	Para la implementación se planteó gestionar los metadatos mediante el motor de base de datos relacional.
La actualización de datos en el repositorio multidimensional se realiza bajo dos esquemas principales que son las cargas incrementales y otro para la actualización total de la carga de datos.	Se debe gestionar un mecanismo que permita cargar la información de las carteras crediticias de los años 2014 y 2015, en cuanto se realice la actualización mensual de los datos se manejará una carga adicional que incluirá la información de la nueva cartera en el consolidado histórico.	Si	Dado que la información final de las carteras posee un corte mensual se plantea la ejecución de cargas incrementales para procesar los datos una vez finalizado el cierre operativo mensual.
Aprovechar las características de los sistemas relacionales de bases de datos para optimizar la velocidad de las consultas.	Reducir los tiempos de entrega de información a los distintos foros de toma de decisiones que integran la organización.	Si	Se plantea aprovechar las características que posee el motor de base de datos Microsoft que posee la organización para la optimización de las consultas sobre el modelo de datos multidimensional.

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Tomando en cuenta la información especificada en la sección 4.1.1.4 y de las mejores prácticas identificadas en la sección 2.9 y esta sección se determinaron las brechas que se detallan en la Tabla 4.12.

Tabla 4.12.

Revisión de brechas con respecto a las mejores prácticas

Mejor práctica	Situación del proceso	Brecha
Reemplazar valores nulos por valores textuales, que incluyen: “desconocido” o “no aplica”	Entre las correcciones de datos mencionadas por los analistas se establece que los valores nulos identificados se trasladan a un archivo independiente para ser reportados al área operativa, encargada de realizar su corrección.	Dado que la corrección de datos se realiza en el área operativa, no se tiene control sobre los cambios realizados en las fuentes de datos, por tanto, no se puede determinar si se incluyen valores textuales descriptivos cuando la información no se puede de clientes u operaciones no se puede localizar.
Extraer solamente los datos necesarios para el procesamiento del ETL	En las tareas mencionadas por los analistas no se menciona la extracción o filtrado de datos para el procesamiento de información, es hasta la construcción del informe donde se realiza la selección de variables para el reporte.	En este caso se están procesando datos que no llegarán a representarse en el informe final del proceso, generando un desperdicio en el tiempo de procesamiento.
Filtrar la información antes de iniciar el procesamiento del ETL	Los analistas entrevistados no realizan filtros sobre la información de las fuentes transaccionales, salvo en el caso de los valores faltantes.	Se recalca la importancia de procesar únicamente los registros requeridos para los reportes, evitando transformar o incluso extraer información que podría no ser necesaria.
Paralelizar el procesamiento de datos, por ejemplo transformaciones y cargas	Dado que el procesamiento de datos de realiza en archivos de Excel los analistas deben ejecutar sus tareas de forma secuencial.	La ejecución secuencial de tareas imposibilita la reducción de los tiempos de procesamiento de datos.
Manejo de un área intermedia para almacenar el procesamiento de datos	Durante el proceso de entrevistas al personal de la entidad no se identificó el manejo de un área o archivos intermedios para el respaldo de datos durante el procesamiento, sin embargo, si se destaca que el departamento de crédito mantiene un archivo que almacena la información consolidada de periodos anteriores.	En este caso se identifica que existe una brecha con respecto a la mejora práctica, dado que si el procesamiento de datos resulta incorrecto o se tiene una falla sobre el archivo que se está procesando no se tiene un respaldo que permita recuperar la información transformada previamente.

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Al identificar las brechas del proceso de construcción de informes con respecto a las mejores prácticas se determina que los factores que afectan principalmente los tiempos de entrega son:

- ❏ El procesamiento sobre los archivos completos (sin filtrar) de la cartera crediticia.
- ❏ La ejecución secuencial de las transformaciones de datos, debido a la falta de tareas paralelizadas.

4.1.3. Analizar las alternativas

Luego de ejecutar la revisión documental sobre las recomendaciones que brindan los proveedores Oracle y Microsoft para herramientas de integración de datos se procedió a consultar a la jefatura de sistemas de la organización criterios de valoración para las alternativas de software, el detalle de la entrevista se encuentra en el Apéndice II: Entrevista para selección de criterios del Benchmarking. Producto del análisis de criterios se obtuvo el siguiente listado de características deseables para las herramientas de procesamiento de datos:

- ❏ Capacidad para crear flujos de datos por medio de entornos gráficos o sentencias de SQL.
- ❏ Capacidad para incorporar librerías nativas de la herramienta que permitan agilizar las integraciones de datos.
- ❏ Capacidad para orquestar y combinar múltiples flujos de transformaciones de datos.
- ❏ Capacidad para integrarse con los sistemas gestores de bases de datos de la organización bajo conexiones *OLEDB*.
- ❏ Capacidad para incorporar extracciones de datos con fuentes de información basadas en archivos de datos planos.
- ❏ Capacidad para realizar perfilado sobre los datos de fuentes transaccionales.

- ❏ Documentación en múltiples idiomas incluido inglés y español.
- ❏ Ofrece paquetes de soporte para usuarios finales por medio de correo electrónico o llamada.

Por otro lado, los criterios para las herramientas de visualización seleccionados por medio de investigación sobre las alternativas disponibles en el mercado fueron:

- ❏ Capacidad para integrarse con sistemas gestores de bases de datos relacionales y archivos planos de datos.
- ❏ Diversidad de visualizaciones que permitan identificar tendencias en la información.
- ❏ Facilidad para construir los gráficos de información por medio de mecanismos *drag and drop*.
- ❏ Cuenta con alternativas para consolidación de visualizaciones en *dashboards*.
- ❏ Permite la creación de filtros y la visualización del detalle de la información mediante operaciones *Drill Down* o en su defecto mediante pistas desplegadas sobre los datos del gráfico.
- ❏ Capacidad para realizar despliegues de los reportes sobre dispositivos móviles.
- ❏ Cuenta con software complementario para visualizar los gráficos que no requieren de licencias adicionales para creación de datos.
- ❏ Permite la actualización periódica de la información sin intervención de personal técnico o de negocio.

Los resultados del proceso comparativo de *Benchmarking* para herramientas de integración y visualización se presentan a continuación.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

4.1.3.1. Crear el comparativo para herramientas de integración de datos

En el caso de las herramientas de integración se tomaron en cuenta cuatro proveedores reconocidos en el mercado por su trayectoria en este tipo de soluciones. La Tabla 4.13 muestra los resultados del comparativo para las soluciones que ofrecen los proveedores seleccionados.

Tabla 4.13.

Comparativo de características para herramientas de integración

Criterio de comparación	Microsoft Integration Services	Pentaho Data Integration	Talend Data Integration	Oracle Data integrator 12 c
Capacidad para crear flujos de datos por medio de entornos gráficos o sentencias de SQL.	Cumple	Cumple	Cumple	Cumple
Capacidad para incorporar librerías nativas de la herramienta que permitan agilizar las integraciones de datos.	No cumple	Cumple	Cumple	Cumple
Capacidad para orquestar y combinar múltiples flujos de transformaciones de datos.	Cumple	Cumple	Cumple	Cumple
Capacidad para integrarse con los sistemas gestores de bases de datos de la organización bajo conexiones OLEDB.	Cumple	Cumple	Cumple	Cumple
Capacidad para incorporar extracciones de datos con fuentes de información basadas en archivos de datos planos.	Cumple	Cumple	Cumple	Cumple
Capacidad para realizar perfilado sobre los datos de fuentes transaccionales.	No cumple	Cumple	No cumple	No cumple
Documentación en múltiples idiomas incluido inglés y español.	Cumple	Cumple	Cumple	Cumple
Ofrece paquetes de soporte para usuarios finales por medio de correo electrónico o llamada.	Cumple	Cumple	Cumple	Cumple

Nota: elaboración propia

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Producto del comparativo se determina que el listado de alternativas recomendadas para herramientas de integración está liderado por *Pentaho Data Integration* del proveedor *Hitachi Vantara*, seguido por *Talend Data Integration*, *Oracle Data Integrator 12c* y *SQL Integration Services*.

Aunque la solución de *Hitachi Vantara* cumple con la mayoría de los criterios deseables se opta por recomendar la herramienta de *SQL Server Integration Services*, debido a los acuerdos de licenciamiento que posee la organización y por, la cantidad de documentación que cuenta el proveedor Microsoft para sus herramientas. Además, se tomó en consideración la posibilidad que ofrece la herramienta de Microsoft para integrarse con las bases de datos actuales de la entidad financiera. La aprobación para utilizar *SQL Server Integration Services* como herramienta de implementación se documentó y aprobó en el Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

4.1.3.2. Crear el comparativo para herramientas de visualización

Para el comparativo de las herramientas de visualización se tomaron en cuenta cuatro proveedores reconocidos en el mercado por su trayectoria en el campo de visualización de datos. La Tabla 4.14 muestra los resultados del comparativo para los proveedores seleccionados.

Tabla 4.14.

Comparativo de características para herramientas de integración

Criterio de comparación	Power BI	Tableau	Qlik Technologies	Sisense
Capacidad para integrarse con sistemas gestores de bases de datos relacionales y archivos planos de datos.	Cumple	Cumple	Cumple	Cumple
Diversidad de visualizaciones que permitan identificar tendencias en la información.	Cumple	Cumple	Cumple	Cumple
Facilidad para construir los gráficos de información por medio de mecanismos <i>drag and drop</i> .	Cumple	Cumple	Cumple	Cumple
Cuenta con alternativas para consolidación de visualizaciones en <i>dashboards</i> .	Cumple	Cumple	Cumple	Cumple
Permite la creación de filtros y la visualización del detalle de la información mediante operaciones Drill Down o en su defecto mediante pistas despleables sobre los datos del gráfico.	Cumple	Cumple	Cumple	Cumple
Capacidad para realizar despliegues de los reportes sobre dispositivos móviles.	Cumple	Cumple	No cumple	No cumple
Cuenta con software complementario para visualizar los gráficos que no requieren de licencias adicionales para creación de datos.	No cumple	Cumple	No cumple	No cumple
Permite la actualización periódica de la información sin intervención de personal técnico.	Cumple	Cumple	Cumple	Cumple

Nota: elaboración propia.

Al analizar el resultado del cuadro comparativo se determinó que la solución del proveedor *Tableau Software* cumple a cabalidad los criterios de comparación establecidos. Cabe destacar que actualmente la organización cuenta con un acuerdo de licenciamiento con este proveedor. La aprobación de *Tableau* como herramienta de implementación para las visualizaciones de datos se documentó en el Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta de este documento.

4.1.3.3. Análisis de inversión

Para el análisis de inversión se tomaron en cuenta los costos de licenciamiento, así como los beneficios económicos para cada herramienta de integración y visualización de datos. Los resultados obtenidos se describen a continuación.

En primer lugar, se realizó la cotización individual para las herramientas de integración y visualización de datos que se muestra en la Tabla 4.15.

Tabla 4.15.

Tabla resumen de la cotización para el software de visualización de datos

Herramienta	Precio mensual por usuario	Precio mensual (siete usuarios)
Power BI	\$9,99	\$69,93
Tableau	\$70	\$490
Sisense	\$15	\$105
Qlik technologies	\$65	\$455

Nota: elaboración propia.

Adicionalmente, se realizó el mismo proceso con el software para la integración de datos, la Tabla 4.16 muestra las cotizaciones del software enfocado en la automatización del proceso ETL.

Tabla 4.16.

Tabla de resumen de las cotizaciones para el software de integración de datos.

Herramienta	Precio para siete usuarios mensuales
SQL Server Integration Services	\$1188
Pentaho Data Integration	\$1166,66
Talend Data Integration	\$1170
Oracle data integrator 12c	\$3050

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Tomando como insumo la información de las cotizaciones, se elaboró la Tabla 4.17 que muestra un resumen de las combinaciones posibles para la selección de herramientas de implementación.

Tabla 4.17.

Mapeo del costo de licencias para herramientas de implementación.

Herramienta de integración	Precio mensual	Herramienta de visualización	Precio mensual	Costo mensual ambas herramientas
SQL Server Integration Services	\$1188	Power BI	\$69,93	\$1257,93
		Tableau	\$490	\$1678
		Sisense	\$105	\$1293
		Qlik	\$455	\$1643
Pentaho Data Integration	\$1166,66	Power BI	\$69,93	\$1236,59
		Tableau	\$490	\$1656,66
		Sisense	\$105	\$1271,66
		Qlik	\$455	\$1621,66
Talend Data Integration	\$1170	Power BI	\$69,93	\$1239,93
		Tableau	\$490	\$1660
		Sisense	\$105	\$1275
		Qlik	\$455	\$1625
Oracle data integrator 12c	\$3050	Power BI	\$69,93	\$3119,63
		Tableau	\$490	\$3540
		Sisense	\$105	\$3155
		Qlik	\$455	\$3505

Nota: elaboración propia.

Teniendo en cuenta que las herramientas recomendadas para la implementación fueron Pentaho Data Integration y Tableau se determina que el costo mensual de licenciamiento para la solución es de \$1656,66, mientras que el costo diario es de \$82,833.

Seguidamente, tomando como base la cantidad de tiempo que invierten los analistas en promedio para la construcción especificada en la Tabla 4.8 y en la Tabla 4.9 realiza el siguiente cálculo para estimar el ahorro o beneficio monetario del proyecto.

Si se toma en cuenta que el grado académico mínimo requerido para la ejecución de las tareas de análisis es Licenciatura se establece que el salario mensual mínimo según el Ministerio de Trabajo de Costa Rica de un analista es de ₡644689,30. Al convertir el monto mencionado con el tipo de cambio de referencia del BCCR fijado en 627,38 colones por dólar, se tendría que el salario mensual dolarizado de un analista como mínimo es de \$1027,60.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Al calcular el salario por minuto del personal que construye los reportes se establece que por persona el costo por minuto es de \$0,0714.

Si se tiene en cuenta que el beneficio esperado del proyecto es la reducción del tiempo requerido para construir los reportes en al menos un 50%, se tendría que estimar una reducción de tiempo de 146 minutos en el departamento de riesgos y de 104,375 minutos en el área de crédito según lo establecido en la Tabla 4.8 y en la Tabla 4.9.

En este caso el ahorro estimado al multiplicar el costo por minuto que invierte cada analista en construir un reporte y los beneficios esperados del proyecto en términos de ahorro por reporte serían de \$10,4244 aproximadamente en el área de riesgo y de \$7,4523 aproximadamente por cada reporte que realizan los analistas del área de crédito.

Si un analista de riesgo invierte su tiempo en ejecutar tres informes al día se estaría generando un ahorro de \$31,27.

Por otro lado, para un analista de crédito que realiza la misma labor se calcula que el ahorro monetario generado es de \$22,35.

Calculando el retorno de la inversión diario se tiene que esta alternativa brinda los siguientes beneficios para tres analistas de riesgo y cuatro de crédito:

ROI = beneficios / costos

ROI = (beneficio por analista * cantidad de analistas) + (beneficio por analista * cantidad de analistas) / costo diario de licenciamiento

ROI = ((\$31,27*3) + (\$22,35*4)) / \$82, 83

ROI = (\$93,81 + \$89,4) / \$82,83 = \$2,21

Es decir, el ROI generado sería de \$2,21 dólares por cada dólar invertido.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Para las otras alternativas de inversión en cuanto a herramientas de integración y visualización de datos se realizaron los siguientes cálculos.

Costo de licenciamiento mensual para herramientas de integración y visualización

En primer lugar, se calculó el costo diario de licenciamiento para las herramientas de visualización e integración de datos que se muestra en la Tabla 4.18.

Tabla 4.18.

Desglose de costos para herramientas de integración y visualización de datos

Herramienta de integración	Precio mensual	Herramienta de visualización	Precio mensual	Costo mensual ambas herramientas	Costo diario
SQL Server Integration Services	\$1188	Power BI	\$69,93	\$1257,93	\$62,8965
		Tableau	\$490	\$1678	\$83,9
		Sisense	\$105	\$1293	\$64,65
		Qlik	\$455	\$1643	\$82,15
Pentaho Data Integration	\$1166,66	Power BI	\$69,93	\$1236,59	\$61,829
		Tableau	\$490	\$1656,66	\$82,833
		Sisense	\$105	\$1271,66	\$63,583
		Qlik	\$455	\$1621,66	\$81,083
Talend Data Integration	\$1170	Power BI	\$69,93	\$1239,93	\$61,996
		Tableau	\$490	\$1660	\$83
		Sisense	\$105	\$1275	\$63,75
		Qlik	\$455	\$1625	\$81,25
Oracle data integrator 12c	\$3050	Power BI	\$69,93	\$3119,63	\$155,981
		Tableau	\$490	\$3540	\$177
		Sisense	\$105	\$3155	\$157,75
		Qlik	\$455	\$3505	\$175,25

Nota: elaboración propia.

Una vez calculado el costo diario del licenciamiento para ambas herramientas se procedió a estimar los beneficios del proyecto para los analistas de riesgo y crédito de la organización.

Cálculo de los beneficios del proyecto

Para estimar los beneficios del proyecto se calculó el ahorro en términos monetarios que genera la solución, multiplicando el salario mínimo de cada analista por valores meta de reducción en tiempos de procesamiento de reportes fijados como objetivos del proyecto.

Tomando en cuenta que el grado académico mínimo requerido para la ejecución de las tareas de análisis es Licenciatura se establece que el salario mensual mínimo según el Ministerio de Trabajo de Costa Rica de un analista es de ₡644689,30. Al convertir el monto mencionado con el tipo de cambio de referencia del BCCR fijado en 627,38 colones por dólar, se tendría que el salario mensual dolarizado de un analista como mínimo es de \$1027,60.

Al calcular el salario por minuto del personal que construye los reportes se establece que por persona el costo por minuto es de \$0,0714.

Si se tiene en cuenta que el beneficio esperado del proyecto es la reducción del tiempo requerido para construir los reportes en al menos un 50%, se tendría que estimar una reducción de tiempo de 146 minutos en el departamento de riesgos y de 104,375 minutos en el área de crédito según lo establecido en la Tabla 4.8 y en la Tabla 4.9.

En este caso el ahorro estimado al multiplicar el costo por minuto que invierte cada analista en construir un reporte y los beneficios esperados del proyecto en términos de ahorro por reporte serían de \$10,4244 aproximadamente en el área de riesgo y de \$7,4523 aproximadamente por cada reporte que realizan los analistas del área de crédito.

Si un analista de riesgo invierte su tiempo en ejecutar tres informes al día se estaría generando un ahorro de \$31,27.

Por otro lado, para un analista de crédito que realiza la misma labor se calcula que el ahorro monetario generado es de \$22,35.

Cálculo del ROI

Para el cálculo del ROI se mantuvo la estructura de los beneficios para manejar un punto de comparación consistente entre alternativas. Sin embargo, para los cálculos de costo se especificaron los cotizados para cada combinación de herramientas de integración y visualización. La Tabla 4.19 muestra un resumen de los cálculos realizados.

Tabla 4.19.

Cálculos del ROI para las alternativas de inversión.

Integración	Visualización	Beneficios riesgos	Beneficios crédito	Costo	ROI
Integration Services	Tableau			\$83,9	2,184
	Power BI			\$62,896	2,913
	Sisense			\$64,65	2,834
	Qlik			\$82,15	2,230
Pentaho Data Integration	Tableau			\$82,833	2,212
	Power BI			\$61,829	2,963
	Sisense			\$63,583	2,881
	Qlik			\$81,083	2,260
		\$93,81	\$89,4		
Talend Data Integration	Tableau			\$83	2,207
	Power BI			\$61,996	2,955
	Sisense			\$63,75	2,874
	Qlik			\$81,25	2,255
Oracle data integrator 12c	Tableau			\$177	1,035
	Power BI			\$155,981	1,175
	Sisense			\$157,75	1,161
	Qlik			\$175,25	1,045

Nota: elaboración propia.

Al realizar el cálculo del ROI se determina que las dos combinaciones que brindan el mayor retorno de la inversión en primer lugar son Pentaho Data Integrarion y Power BI, mientras que Talend Data Integration y Power BI es la segunda combinación que brinda un mayor rendimiento con respecto a los costos del proyecto.

4.2. Fase II: Diseño de los componentes de inteligencia de negocio

Los resultados obtenidos durante la ejecución de esta fase propiciaron la elaboración de los diseños del repositorio multidimensional de datos, las visualizaciones de información y el flujo del proceso de Extracción, Transformación y Carga que alimenta el *Data mart* de la solución. A continuación, se describen los principales hallazgos de esta etapa del proyecto.

4.2.1. Diseñar el *Data mart* (DM)

El diseño de este componente requirió que se elaboraran los paquetes de información que fundamentaron el modelo multidimensional que almacena los datos para análisis crediticios en las áreas de crédito y riesgo.

4.2.1.1. Crear los paquetes de información

Producto de la recolección de requerimientos realizada con los analistas de crédito y riesgo se generaron los paquetes de información donde se especifican las dimensiones y los hechos que se incorporaron posteriormente en el modelo multidimensional de la propuesta. La Tabla 4.20, muestra los paquetes de información obtenidos al finalizar la recolección de requerimientos.

Tabla 4.20.

Paquete de información para la cartera crediticia

Paquete de información Cartera crediticia			
Dimensiones			
Tiempo	Operación	Ejecutivo	Cliente
Mes.	Descripción del producto.	Código Ejecutivo.	Actividad económica.
Trimestre.	Descripción subproducto.	Nombre Ejecutivo.	Identificador organizacional único.
Semestre.	Número de operación.		Identificación.
Año.	Días de atraso.		Nombre del cliente.
	Estado de operación.		Generador de divisa.
	Moneda.		Categoría de Riesgo.
	Año de operación.		
	Mes de operación		
Hechos			
Monto total deuda colonizado		Monto principal contingente moneda	
Monto total deuda dolarizado		Monto principal directo moneda	

Nota: elaboración propia.

El identificar los atributos que se debían incluir en las estructuras del modelo multidimensional resultó fundamental para elaborar el diseño lógico del DM que se especifica en la sección 5.1 del capítulo cinco del presente documento.

4.2.1.2. Diseñar el modelo multidimensional

Para iniciar el diseño del modelo lógico del *Data mart* fue necesario determinar el tipo de esquema que debía seguir la base de datos multidimensional donde se almacenó la información crediticia. El proceso de análisis tomó en cuenta lo establecido en la sección 2.4.5 de este documento, donde se menciona que la selección del esquema depende de los requerimientos de la solución.

A continuación, se establece cómo las variantes para representar el modelo de datos multidimensional satisfacen los requisitos de la entidad financiera.

Esquema copo de nieve

- Facilita el mantenimiento y actualización de las estructuras normalizadas, en este caso la información procesada no requiere actualización, ya que el comportamiento de la cartera se mantiene consistente una vez realizado el cierre operativo mensual de la organización.

Esquema estrella

- Su falta de normalización permite almacenar y consultar grandes cantidades de información de manera rápida, esto favorece la construcción de los reportes crediticios que incluyen en promedio 8000 registros por cada archivo.
- La representación visual del modelo es más intuitiva para los usuarios, esto permite que los reportes requeridos se puedan realizar de manera ágil.
- Las relaciones entre las tablas de la base de datos multidimensional facilitan la construcción de reportes en las herramientas de visualización que posee la entidad.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Partiendo de las ventajas que ofrece cada enfoque se determinó que el modelo de tipo estrella era la alternativa óptima para representar la información de crédito. Otros aspectos importantes para el diseño del modelo multidimensional incluyen:

Capacidad para manejar consultas relacionadas a la información de los clientes, ejecutivos, operaciones en un marco temporal definido.

El modelo debe mantener la granularidad de los datos que posee el archivo de la cartera crediticia, es decir debe mantener el mayor nivel de detalle disponible en esta fuente de datos que es la operación de un cliente.

4.2.2. Diseñar el proceso ETL

Para el diseño del proceso de Extracción, Transformación y Carga se identificaron algunos aspectos importantes en cada componente a continuación se describe cada uno de ellos.

4.2.2.1. Diseñar el flujo de integración de datos

Como se mencionó anteriormente en los resultados de las entrevistas al departamento de Riesgo y Crédito de la entidad, la información de la cartera crediticia se encuentra en su estado más consistente luego del cierre mensual, por tanto, se considera que la ejecución de las extracciones debe realizarse posterior a la fecha de cierre.

La información crediticia no se actualiza posterior al cierre mensual ya que estos datos muestran el estado de las operaciones al finalizar el mes natural. Cualquier operación modificada en los sistemas transaccionales después del cierre operativo mensual se considera como parte integral del siguiente mes en el que opera la entidad financiera.

4.2.2.2. Diseñar las transformaciones de datos

Producto de las entrevistas realizadas al personal de la entidad financiera se identificaron una serie de transformaciones sobre los datos para generar el resultado final del proceso de ETL. Las transformaciones principales identificadas en el procesamiento de los analistas incluyen:

- ❏ Creación de campos calculados especialmente sobre saldos consolidados de los productos financieros.

 Dolarización y colonización de saldos.

En este caso se realiza la conversión contra el tipo de cambio de compra al momento del cierre mensual de referencia brindado por el Banco Central de Costa Rica (BCCR).

 Conversión de valores numéricos a campos categóricos o rangos de valores.

La primera conversión de valores se realiza sobre el dato numérico de días de atraso que se debía convertir a un rango de valores con la siguiente estructura:

- 0-30.
- 31-60.
- 61-90.
- 91-120.
- Más de 120.

La segunda transformación se realiza sobre el dato numérico Generador de divisa donde se realiza el siguiente mapeo:

- 1 se convierte a Generador.
- 2 se convierte a No generador de bajo riesgo.
- 3 se convierte a No generador de alto riesgo.
- 4 se convierte a no aplica.

La tercera transformación se realiza sobre el dato textual de categoría de Riesgo donde se realiza el siguiente mapeo:

- A1 y A2 se convierte a A.
- B1 y B2 se convierte a B.
- C1 y C2 se convierte a C.
- D y E se mantiene en el estado correspondiente a los datos fuente.

 Consolidación de datos disponibles en múltiples archivos de Excel.

4.2.2.3. Diseñar las estrategias para carga de datos

Para la carga de datos se tomó en cuenta que la ventana de tiempo ideal para ejecutar los procesos de carga de datos en los repositorios organizacionales es posterior al cierre operativo diario, debido a la consolidación de saldos que realizan las áreas operativas.

4.2.3. Diseñar las visualizaciones

El diseño de las visualizaciones de datos crediticios se basó en la selección de las herramientas requeridas para construir las gráficas y en las consideraciones que mencionaron los analistas para la construcción visual de sus reportes durante la recolección de requisitos.

4.2.3.1. Consideraciones para la selección de herramientas de visualización

La selección de herramientas de visualización se basa en el grado de familiaridad que posee el personal de la institución financiera con software de visualización tales como Tableau y Power BI. Otro factor que se tomó en cuenta para la selección de las herramientas que permiten elaborar los gráficos es la capacidad que poseen para conectarse con repositorios de información del tipo SQL Server y Oracle, que son tecnologías predominantes en la entidad financiera.

4.2.3.2. Consideraciones para el diseño de visualizaciones

Para la selección de las visualizaciones de datos se tomó en cuenta que el objetivo principal es analizar tendencias sobre datos históricos de la cartera crediticia. Es imperativo que la herramienta seleccionada permita la creación de visualizaciones que reflejen el comportamiento crediticio de los clientes que posee la entidad financiera.

Para el diseño de visualizaciones es importante mantener una paleta de colores oscura para los fondos de los gráficos y de colores cálidos que contrasten con el fondo, por medio del contraste posible resaltar o hacer énfasis en la información que se desea enfatizar en el reporte.

La escala de los datos en todos los reportes se debe realizar en millones de colones o dólares. El resumir los datos en millones facilita la representación de montos elevados para las operaciones de crédito.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Se debe evitar el manejo de gráficos circulares o en tres dimensiones ya que no permiten quienes interpretan el reporte identificar claramente las diferencias entre los datos que se muestran en las gráficas.

4.3. Fase III: Implementación del *Data mart*, proceso de ETL y visualizaciones

Durante esta fase se realizó la creación física de los componentes del *Data mart*, proceso de ETL y visualizaciones diseñadas previamente. Algunos aspectos que se deben considerar durante la implementación de la propuesta son:

4.3.1. Implementar el *Data mart* (DM)

Para realizar la implementación del repositorio histórico de información crediticia es necesario traducir el diseño lógico en el modelo físico de la base de datos objetivo.

4.3.1.1. Construir el modelo en la base de datos objetivo

Dado que la selección de la base de datos objetivo para la implementación del modelo multidimensional es del tipo SQL Server es necesario tomar en cuenta los tipos de datos propietarios disponibles en esta tecnología y evitar su uso en la propuesta final. Al seguir este lineamiento se mantuvo la posibilidad de implementar el modelo en otro motor de base de datos de manera ágil.

4.3.2. Implementar el proceso ETL

Utilizando el diseño del proceso que se realizó en la etapa anterior se procedió a implementar la propuesta en la herramienta de integración seleccionada.

4.3.2.1. Implementar el proceso completo en la herramienta seleccionada

Debido al acuerdo realizado con la organización sobre el desarrollo de la propuesta en herramientas del proveedor Microsoft, se procedió a generar las representaciones físicas del repositorio de datos los repositorios en dicha tecnología. El acuerdo de reunión se encuentra documentado en el Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta.

4.3.3. Implementar las visualizaciones

La elaboración de las visualizaciones considera los aspectos mencionados para el diseño de los *dashboards* y los plasma en la herramienta que pone a disposición la entidad financiera para el proyecto.

4.3.3.1. Construir los *dashboards*

En cuanto a las herramientas de visualización se tomó la decisión de utilizar *Tableau* por su capacidad de entrega de información y esquema de licenciamiento adquirido por la entidad financiera para los próximos cinco años.

Los aspectos específicos de la implementación de cada componente corresponden a la propuesta de solución del proyecto y se describe en el capítulo dedicado a la Propuesta de solución de este documento.

4.4. Fase IV: Comprobación de resultados

Durante la última fase del proyecto se llevó a cabo la recolección de resultados del proceso de ETL, utilizados para contrastar la mejora obtenida con los objetivos del proyecto.

4.4.1. Ejecutar las pruebas

Para corroborar que la solución provee los resultados esperados se verificaron los siguientes componentes:

- Lógica de ETL
 - Se comprobó que la cantidad de registros procesados durante la primera ejecución del proceso de ETL concordara con la cantidad de registros existentes en los archivos de la cartera crediticia.
 - Se realizó el perfilado de los valores categóricos que en la fuente transaccional están representados con valores numéricos.
 - Se comprobó que la cantidad de registros extraídos para la primera ejecución del proceso coincidiera con la cantidad de registros cargados en el *Data mart*.

Para la ejecución de las pruebas sobre la solución se utilizó la plantilla ubicada en el Anexo 2: Plantilla para la documentación de pruebas.

4.4.2. Realizar el comparativo de resultados

A continuación, se describen las métricas que permitieron evaluar las mejoras sobre el proceso de creación de informes.

- **Tiempo promedio para creación de reportes**

Tiempo dedicado a la extracción de datos más el tiempo el asignado a las transformaciones de información y el tiempo dedicado a la construcción de reporte tabular y visual. Los tiempos promedio se producen al realizar la sumatoria de todos los tiempos para elaboración de reportes de cada analista y dividirlos por el número total de analistas.

- **Tasa de reprocesos o iteraciones**

Cantidad de reprocesamiento mensual/cantidad de reprocesamiento anual.

La Tabla 4.21 muestra el detalle de las métricas de mejora y los valores meta para cada una de ellas.

Tabla 4.21.

Descripción de las métricas de evaluación para el proceso de construcción de informes

Métrica	Cálculo	Meta	Responsable del cumplimiento
Tiempo promedio para elaboración del reporte	Sumatoria de todos los tiempos para elaboración de reportes dividido entre el total de reportes realizados	146 minutos para el área de riesgo 104,375 minutos para el área de crédito	Estudiante (desarrollador del proyecto)
Tasa de reprocesos o iteraciones	Cantidad de reprocesos mensuales divididos por la cantidad total de reprocesos	Cinco informes para el área de riesgos y diez informes para el departamento de crédito de un total de 74 informes anuales	Estudiante (desarrollador del proyecto)

Nota: elaboración propia

Propuesta para la elaboración de reportes crediticios de una entidad financiera

El cálculo de los valores meta para cada métrica se basa en los objetivos del proyecto y la información de tiempos promedio para la construcción de informes identificados en la sección 4.1.1.4 de este documento.

Tomando como referencia los resultados obtenidos en este capítulo se procedió a desarrollar la propuesta de solución para la problemática que afectaba a la entidad financiera.

5

Propuesta de solución

En este capítulo se detalla la propuesta que busca solventar la situación problemática identificada en la entidad financiera. El contenido de este capítulo se formuló con base en los resultados que se especifican en el capítulo cuatro de este documento.

Para solucionar la problemática de la entidad financiera se plantea el desarrollo de tres componentes de inteligencia de negocio que incluyen el almacenamiento, transformación y visualización de los datos relacionados a la cartera crediticia de clientes.

La Figura 5.1 muestra los elementos que integran la solución de inteligencia de negocio, iniciando por las Fuentes de datos transaccionales que almacenan la información de las operaciones crediticias. El segundo componente corresponde al proceso de ETL que extrae, transforma y almacena los datos en el modelo multidimensional correspondiente al tercer elemento de la solución. Finalmente, el último componente de la propuesta se encarga de visualizar los datos almacenados en el modelo multidimensional según los requisitos de las áreas de negocio.

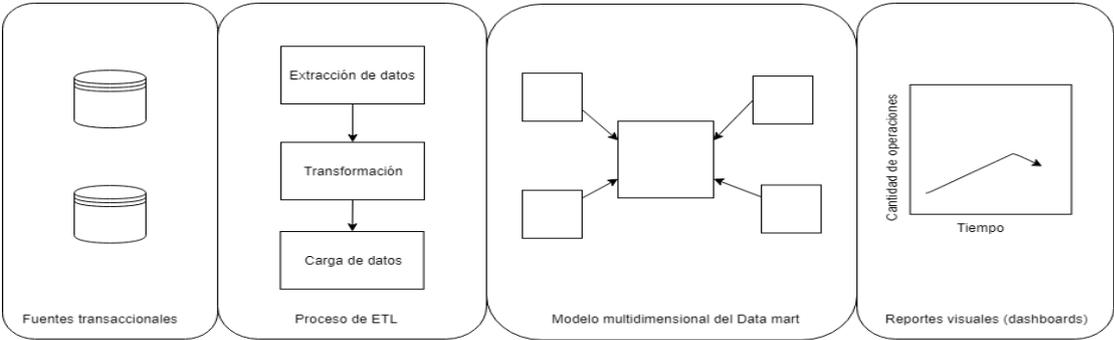


Figura 5.1. Componentes de la propuesta de solución

Nota: elaboración propia.

A continuación, se detallan los elementos diseñados durante la ejecución del proyecto que incluyen el proceso de ETL, modelo lógico del *Data mart* y las visualizaciones de datos.

5.1. Diseño del modelo multidimensional del *Data mart*

Como se mencionó en el capítulo IV, producto de las entrevistas se identificaron las métricas y datos categóricos que constituyen las tablas dimensionales y la tabla de hechos del modelo. Al estructurar las métricas y atributos identificados se procedió a diagramar modelo de datos del *Data mart* que se muestra en la Figura 5.2, el mismo fue aprobado según la minuta de reunión que se muestra en el Apéndice J: Aprobación del modelo de datos para el *Data mart* de este documento.



Figura 5.2. Modelo lógico del *Data mart*

Nota: Elaboración propia

Como se puede observar en el modelo de datos se cuenta con cuatro tablas dimensionales y una tabla de hechos que permiten la construcción de los reportes que realizan los analistas en las áreas de Crédito y Riesgo de la entidad financiera. El detalle de los elementos que integran el modelo multidimensional se especifica en la Tabla 5.1.

Tabla 5.1.

Descripción del modelo multidimensional

Nombre de la tabla	Descripción
	<p>Esta tabla dimensional incluye los datos del personal de ventas asociado a cada operación crediticia a lo largo de la historia de la entidad financiera. A continuación, se describen los atributos que componen esta tabla.</p>
	<p> Id_ejecutivo: es la llave primaria de la dimensión Ejecutivo y por tanto, representa el identificador único de los colaboradores a nivel del modelo multidimensional. Esta llave permite establecer la relación con la tabla de hechos de crédito y combinar los resultados de las consultas multidimensionales sobre el <i>Data mart</i>.</p>
Dim_ejecutivo	<p> Cod_ejecutivo: es el código que representa de forma única a cada ejecutivo que labora para la entidad financiera a nivel del sistema maestro transaccional.</p>
	<p> Nombre_ejecutivo: incluye el nombre completo de cada ejecutivo que ha laborado para la entidad financiera.</p>

Continuación de la tabla 5.1

Nombre de la tabla	Descripción
	<p>Esta dimensión incorpora los datos correspondientes a las operaciones de crédito de la entidad financiera. A continuación, se describen los atributos que componen esta dimensión.</p>
	<p> Id_operacion: Representa de forma única cada operación a nivel de la tabla dimensional. Esta llave permite establecer la relación con la tabla de hechos de crédito y combinar los resultados sobre la base de datos del Data mart.</p>
	<p> Desc_producto: este atributo describe el producto asociado a cada operación de crédito.</p>
	<p> Desc_subproducto: este atributo describe el subproducto de cada operación registrada en la tabla dimensional.</p>
Dim_operacion	<p> Num_operacion: representa el consecutivo alfanumérico que identifica cada operación en el sistema transaccional.</p>
	<p> Dias_atraso: representa la cantidad de días transcurridos desde la fecha de vencimiento de las operaciones crediticias para las cuales no se ha registrado un pago u amortización.</p>
	<p> Estado_operacion: indica el estado de las operaciones en el sistema transaccional de la institución. Los estados posibles de una operación incluyen: vigente, vencida o en cobro judicial.</p>
	<p> Moneda: es la moneda en la que registró o formalizó la operación en la fuente de datos transaccional.</p>
	<p> Año_op y Mes op: identifican el año y mes en el que se registran los atributos de la operación en el sistema transaccional.</p>

Continuación de la tabla 5.1

Nombre de la tabla	Descripción
	<p>La dimensión temporal desglosa por mes, semestre, trimestre y año las operaciones registradas en el sistema maestro transaccional de la entidad financiera.</p>
	<ul style="list-style-type: none"> <li data-bbox="477 607 1393 869">  Id_tiempo: Representa el identificador único de las unidades de tiempo a nivel del modelo multidimensional. Esta llave permite establecer la relación con la tabla de hechos de crédito y habilita la construcción de consultas sobre la tabla de hechos y los registros temporales de la cartera crediticia. <li data-bbox="477 936 1393 1032">  Mes: incluye el valor textual asociado al mes en el que fueron registradas las operaciones de crédito.
Dim_tiempo	<ul style="list-style-type: none"> <li data-bbox="477 1106 1393 1202">  Trimestre: incluye el valor textual asociado al trimestre en el que fueron registradas las operaciones de crédito. <li data-bbox="477 1270 1393 1366">  Semestre: incluye el valor textual asociado al semestre en el que fueron registradas las operaciones de crédito. <li data-bbox="477 1433 1393 1529">  Año: incluye el valor textual asociado al año en el que fueron registradas las operaciones de crédito. <p data-bbox="432 1615 1393 1809">Esta dimensión es fundamental para la construcción de los análisis enfocados en determinar el comportamiento de la cartera crediticia, al incorporar variables temporales a las métricas y atributos de las tablas dimensionales.</p>

Continuación de la tabla 5.1

Nombre de la tabla	Descripción
	<p>Esta dimensión incorpora la información que describe los clientes de la entidad financiera. Los atributos incluidos en la tabla histórica de clientes son:</p> <ul style="list-style-type: none"> <li data-bbox="478 526 1394 683">  Id_cliente: representa de forma única los registros de cada cliente de la entidad a nivel del modelo multidimensional. Por medio de esta llave se establece la relación con la tabla de hechos de crédito. <li data-bbox="478 728 1394 884">  Identificación: este atributo incluye el número de cédula nacional, cédula de residencia o pasaporte de los clientes de la entidad financiera. <li data-bbox="478 929 1394 1041">  Codigo_org: incluye el código único del cliente que se encuentra registrado a nivel del sistema transaccional.
Dim_cliente	<ul style="list-style-type: none"> <li data-bbox="478 1086 1394 1198">  Nombre_cliente: almacena el nombre de los clientes físicos y jurídicos registrados en las carteras crediticias. <li data-bbox="478 1243 1394 1400">  Generador_divisa: indica la categoría de generador de divisa que poseen los clientes, que puede ser generador, no generado o no aplica. <li data-bbox="478 1444 1394 1556">  Actividad: incluye la actividad relacionada al giro de negocio del cliente registrado en la cartera de crédito. <li data-bbox="478 1601 1394 1803">  Categoria_de_riesgo: incluye la categoría de riesgo asociada al cliente que puede tener un valor de A, B, C, D, E. Este atributo indica los clientes que representar un menor o mayor riesgo para la entidad.

Continuación de la tabla 5.1

Nombre de la tabla	Descripción
Fact_Saldos	<p>La tabla de hechos se compone de variables relacionadas con los saldos de las operaciones de los clientes en colones y moneda extranjera. Sus atributos incluyen:</p> <ul style="list-style-type: none"> <li data-bbox="480 521 1390 674">  Id_credito: es la llave primaria de la tabla de hechos y representa de forma única cada registro incluido en la base de datos como una operación de crédito independiente. <li data-bbox="480 745 1390 835">  Id_ejecutivo: es la llave foránea que relaciona la tabla de hechos de crédito con la tabla dimensional de ejecutivos. <li data-bbox="480 907 1390 996">  Id_operacion: es la llave foránea que relaciona la tabla de hechos de crédito con la tabla dimensional de operaciones. <li data-bbox="480 1068 1390 1158">  Id_tiempo: es la llave foránea que relaciona la tabla de hechos de crédito con la tabla dimensional de tiempo. <li data-bbox="480 1229 1390 1319">  Id_cliente: es la llave foránea que relaciona la tabla de hechos de crédito con la tabla dimensional de clientes. <li data-bbox="480 1391 1390 1543">  Monto_total_deuda_col: almacena la suma del monto principal directo y contingente colonizado al cierre mensual en la que se procesó la cartera crediticia. <li data-bbox="480 1615 1390 1767">  Monto_total_deuda_dol: es un campo calculado que toma el monto total de la deuda colonizado y lo dolariza al tipo de cambio de compra de referencia del BCCR al cierre mensual.

Continuación de la tabla 5.1

Nombre de la tabla	Descripción
	<p> Monto_principal_directo_moneda: Incluye los montos relacionados a todas las operaciones directas de la cartera de crédito.</p>
	<p> Monto_principal_contingente_moneda: Incluye los montos relacionados a todas las operaciones contingentes de la cartera de crédito.</p>
	<p> Monto_producto_colonizado: representa el interés al cobro relacionado a la operación crediticia.</p>
	<p> Monto_cta: incluye cargos varios relacionados a la operación de crédito tales como avalúos y morosidad.</p>
Fact_Saldos	<p> Monto_estimacion_generica: representa la estimación regulatoria que se aplica a todas las operaciones de crédito de las entidades financieras.</p>
	<p> Monto_estimacion_especifica_saldo_sin_garantia: Incluye el monto de estimación basada en la calificación del cliente con respecto a su categoría de riesgo.</p>
	<p> Monto_estimacion_especifica_saldo_con_garantia: Incluye la estimación de 0.5 % sobre el saldo cubierto con garantía de la operación de crédito.</p>
	<p> Monto_mitigador: monto cubierto de la operación crediticia que se utiliza en caso de incobrables.</p>

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Una vez diseñado el modelo lógico de datos que almacena y estructura la información del *Data mart* crediticio se procedió a diseñar el proceso encargado de poblar las tablas dimensionales y la tabla de hechos del modelo.

5.2. Diseño del proceso de Extracción Transformación y Carga de datos (ETL)

El proceso de ETL diseñado se encarga de extraer, procesar y almacenar la información de las fuentes transaccionales en el modelo de datos del *Data Mart* diseñado previamente. El objetivo del diseño fue facilitar la:

- ❏ Extracción de datos de 24 archivos de Excel que incluyen información de la cartera crediticia.
- ❏ Incorporación de los datos de la cartera crediticia en una base de datos relacional para facilitar el procesamiento de información.
- ❏ Transformación de los datos relacionados a la dimensión tiempo, cliente y operación.
- ❏ Inserción de datos a las tablas dimensionales y posteriormente a la tabla de hechos del *Data mart*.

El diseño del proceso de ETL inicia con la creación del flujo lógico de actividades que se debían ejecutar para el procesamiento de los datos crediticios.

5.2.1. Diagrama lógico del proceso

La Figura 5.3 muestra el flujo lógico de las tareas que componen el proceso de ETL propuesto como solución para la situación actual de la entidad financiera.

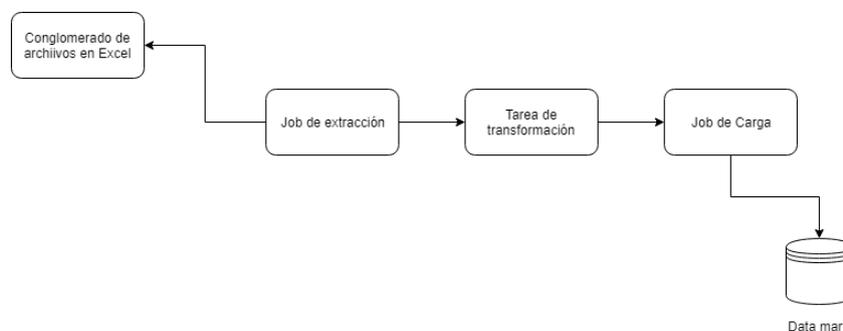


Figura 5.3. Flujo lógico del proceso de ETL

Nota: Elaboración propia

Sobre este diagrama lógico se destaca la integración de múltiples fuentes de datos disponibles en archivos de Excel.

5.2.2. Detalle de las transformaciones

El detalle de las transformaciones requeridas sobre la información de las fuentes transaccionales se especifica en la Tabla 5.2.

Tabla 5.2.

Detalle de las transformaciones para el proceso de ETL

Transformación	Datos originales	Datos transformados
Dolarización y colonización de saldos.	Monto total de deuda colonizado.	Monto total de deuda dolarizado al tipo de cambio de compra de referencia del BCCR al cierre mensual.
Conversión de datos numéricos a rangos o categorías de información para los días de atraso.	Días de atraso en el pago de las operaciones expresado como un número entero.	Creación de los rangos para los días de atraso con la siguiente estructura: 0-30 31-60 61-90 91-120 Más de 120
Conversión de códigos numéricos a rangos o categorías de información para indicar si un cliente es generador de divisas.	Los archivos de cartera manejan una columna que contiene los valores 1, 2, 3 y 4 para indicar la categoría del cliente en cuanto a generación de divisas.	Se transforma el código numérico bajo las siguientes reglas: <ul style="list-style-type: none">  1 se convierte a Generador.  2 se convierte a No generador de bajo riesgo.  3 se convierte a No generador de alto riesgo.  4 se convierte a no aplica.
Mapeo de categorías de riesgo alfanuméricas a datos textuales que indican la categoría de riesgo de un cliente.	Los archivos de cartera poseen una columna que indica la categoría de riesgo de cada cliente según sus características. Las categorías de clasificación principales son A1, A2, B1, B2, C1, C2, D y E.	En la transformación se convierten los datos alfanuméricos bajo las siguientes reglas: <ul style="list-style-type: none">  A1 y A2 se convierten al valor A.  B1 y B2 se convierten al valor B.  C1 y C2 se convierten al valor C.  D y E se mantiene igual que en las fuentes transaccionales.

Nota elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

El detalle de los datos disponibles en las fuentes de información para la carga de datos se encuentra especificado en la sección 4.1.1.3 de este documento.

5.2.3. Mecanismos de integración entre transformaciones y carga

Para el traslado de la información transformada al modelo multidimensional se planteó el uso del enlace OLEDB nativo que posee la herramienta SQL Server Integration Services del proveedor Microsoft.

5.2.4. Periodicidad de ejecución

El proceso de carga inicial se ejecuta únicamente la primera vez que se crea el modelo multidimensional, esto se debe a que la información de las carteras de crédito no se actualiza una vez realizado el cierre operativo mensual como se mencionó en la sección 4.2.2.1 de este documento.

5.2.5. Prerrequisitos

El requisito para la ejecución de la carga inicial de datos es que la información de las carteras a procesar se encuentre disponible en el directorio destinado para la solución, ya sea una carpeta local o remota

Adicionalmente, se requiere que los nombres de las columnas en los archivos de carteras crediticias sean iguales en todos los ficheros que se deben procesar.

5.3. Diseño de las visualizaciones de datos

Cada reporte que se deriva de la propuesta del proyecto se encuentra conectado al modelo de datos del Data mart. El enlace entre el modelo multidimensional y el software de visualización propicia la construcción de los reportes crediticios al ofrecer los insumos para el reporte solicitado de forma ágil.

La información que se incorpora en los reportes corresponde al consolidado de las carteras crediticias de años anteriores, lo que permitió identificar tendencias en el comportamiento de los clientes de crédito.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Para la construcción de reportes se solicitó a los analistas de crédito y riesgo seleccionar tres reportes que construyen regularmente con base en la cartera de operaciones crediticias. Los reportes seleccionados fueron:

-  Tendencia del monto total de la deuda colonizado por categoría de riesgo y días de atraso.
-  Distribución del monto directo moneda anual por días de atraso y moneda.
-  Distribución de la estimación genérica por año.

Los tipos de visualización aprobadas para los reportes según el Apéndice K: Aprobación de visualizaciones son de tipo lineal y barras. Los datos requeridos para cada reporte gráfico son:

Tendencia del monto total de la deuda por categoría de riesgo y días de atraso

Los datos utilizados para este reporte son el monto total de la deuda y la fecha en la que fueron registradas.

Distribución del monto directo moneda anual por días de atraso y moneda

Los datos utilizados para este reporte son el monto principal directo moneda, el año de registro de la operación, la moneda y la categoría de días de atraso.

Distribución de la estimación genérica por año

Los datos utilizados para este reporte son el monto de estimación genérica y el año en el que fue registrada la operación de crédito.

Cabe destacar que la periodicidad de actualización para los reportes visuales mencionados anteriormente es mensual.

5.3.1. Mapeo de requerimientos contra las funcionalidades de las herramientas

La Tabla 5.3 muestra el mapeo de las funcionalidades que poseen las herramientas seleccionadas para la implementación de la propuesta del proyecto y los requerimientos recolectados de los usuarios finales (analistas de riesgo y crédito de la entidad financiera).

Tabla 5.3.

Mapeo de requisitos contra funcionalidades de las herramientas

Requerimiento	Herramienta	Funcionalidad de las herramientas
Dolarización de montos correspondiente a los saldos de operación.	SQL Server Integration Services	Creación de campo calculado para saldos dolarizados
Conversión de códigos numéricos a categorías y rangos de valores.	SQL Server Integration Services	Creación de flujos condicionales y campos calculados para el mapeo de rangos y categorías
Agregación de datos y navegación al detalle de las operaciones	Tableau Data Visualization	Desglose de los datos al incorporar columnas disponibles sobre las tablas dimensionales que integran el cubo de datos crediticios.
Creación de reportes que incluyan las perspectivas de cliente, ejecutivo y operaciones en un contexto histórico	Tableau Data Visualization	Creación de consultas en un entorno gráfico donde se puede incorporar la información a un panel de datos arrastrando columnas.
Graficar la tendencia en el comportamiento de la cartera crediticia	Tableau Data Visualization	Elaboración de gráficos lineales.
Capacidad para construir gráficos interactivos que permitan navegar al detalle de la información representada.	Tableau Data Visualization	Creación de niveles y capas de datos, adicionalmente, se pueden utilizar los elementos de tipo hint para ofrecer más detalle en la gráfica.

Nota: elaboración propia

5.4. Implementación de las soluciones de IN

Este apartado detalla la implementación de los componentes diseñados como parte de la propuesta de solución del proyecto, a continuación, se detalla el desarrollo del modelo multidimensional, proceso de ETL y las visualizaciones de datos en las herramientas seleccionadas por la entidad financiera.

5.4.1. Implementación del modelo multidimensional

El modelo multidimensional fue implementado en el gestor de base de datos relacional Microsoft SQL Server. El diagrama físico del modelo implementado se muestra en la Figura 5.4.

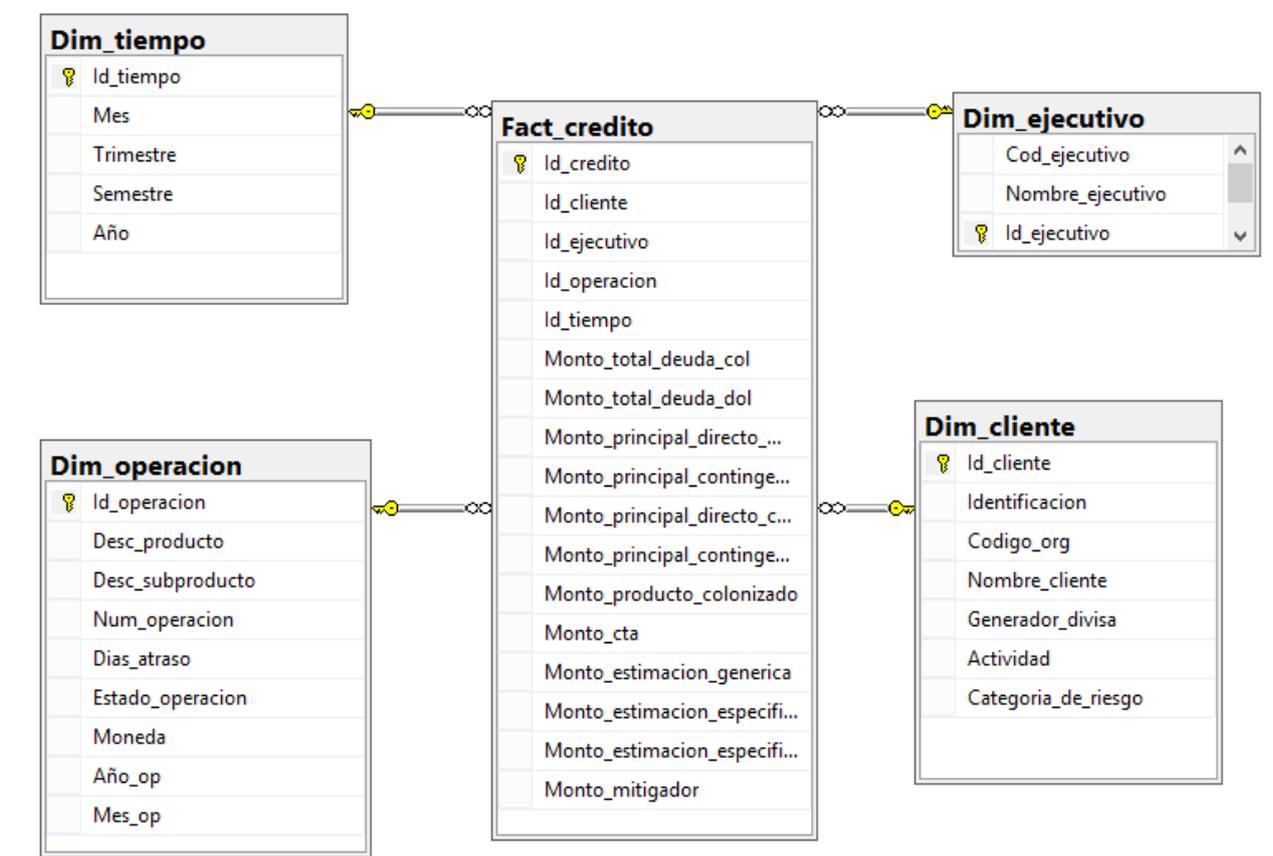


Figura 5.4. Implementación física del modelo multidimensional crediticio

Nota: elaboración propia utilizando la herramienta SQL Server

5.4.2. Implementación del proceso de ETL

La implementación del proceso se realizó por medio de la herramienta de integración SQL Server Integration Services y su interconexión con el entorno de desarrollo Visual Studio.

A continuación, se describe el flujo del proceso de ETL, las cargas a las tablas de dimensión y la tabla de hechos del modelo multidimensional que se implementaron como parte de la propuesta del proyecto.

5.4.2.1. Flujo del proceso implementado

El proceso se compone de seis tareas principales que gestionan la extracción, transformación y carga de los datos al modelo multidimensional del DM crediticio. La secuencia de las tareas en el flujo de datos se muestra en la Figura 5.5. Es importante destacar que en adelante se invirtieron los colores de las figuras en este capítulo para facilitar su comprensión y legibilidad.

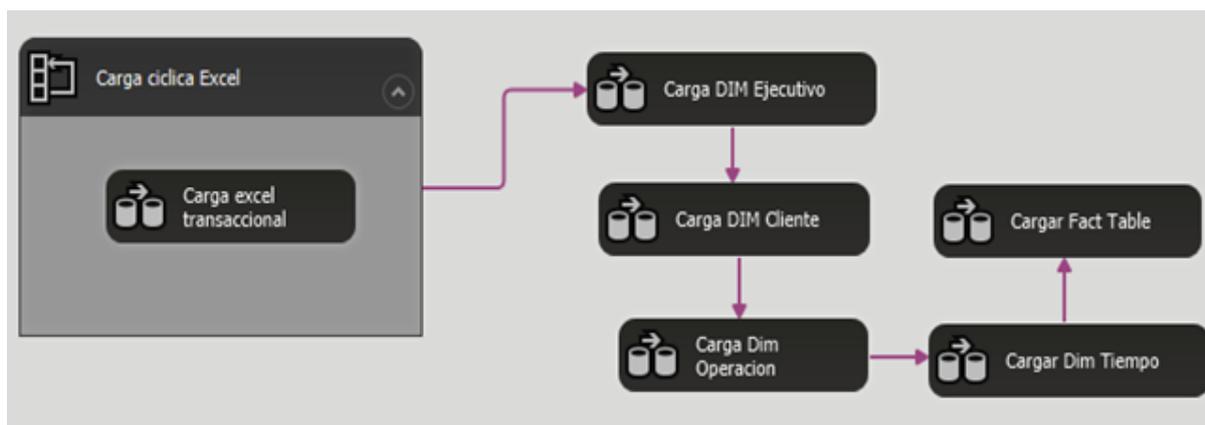


Figura 5.5. Flujo del proceso ETL implementado en SQL Server Integration Services.

Nota: elaboración propia.

Como se puede observar en la Figura 5.5 la primera actividad ejecutada es la extracción y carga de los datos disponibles en los archivos de la cartera crediticia de la entidad a una base de datos relacional intermedia, esta actividad es primordial ya que brinda el insumo para la construcción de las tablas dimensionales y de hechos posteriormente. Cabe destacar que la primera actividad del proceso es cíclica y se ejecuta por cada archivo disponible en el directorio asignado para la extracción de datos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Las tareas que se ejecutan luego de realizar la carga en la base intermedia corresponden a la transformación de datos y carga de información a las tablas dimensionales.

La última actividad del flujo corresponde a la carga de la tabla de hechos en el modelo multidimensional. Esta actividad se ejecuta al final del proceso ya que requiere que las tablas dimensionales se encuentren pobladas con la información de las carteras crediticias, permitiendo la creación de los enlaces entre las tablas dimensionales y la tabla de hechos.

A continuación, se describe cada componente ejecutado durante el proceso de extracción, transformación y carga de los datos al DM.

5.4.2.2. Carga cíclica Excel

Como se mencionó anteriormente la primera actividad del proceso corresponde a la extracción de los archivos de la cartera crediticia y su posterior carga en la base de datos intermedia. La extracción y carga de datos se ejecuta cíclicamente hasta completar la totalidad de los archivos disponibles en el directorio de extracción. La Figura 5.6 muestra gráficamente la secuencia de tareas ejecutadas durante la carga de archivos Excel a la base de datos.



Figura 5.6. Flujo de carga a base de datos intermedia

Nota: elaboración propia.

La primera tarea que se presenta en la Figura 5.6 corresponde a la extracción de los datos disponibles en las carteras crediticias de la entidad financiera. La Figura 5.7 muestra la composición del directorio de donde se realiza la extracción de datos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Name	Date modified	Type	Size
Cartera_credito_2014_01_31	17/10/2018 03:58 ...	Microsoft Excel W...	1.808 KB
Cartera_credito_2014_02_28	16/10/2018 04:27 ...	Microsoft Excel W...	3.085 KB
Cartera_credito_2014_03_31	16/10/2018 04:29 ...	Microsoft Excel W...	1.815 KB
Cartera_credito_2014_04_30	16/10/2018 04:31 ...	Microsoft Excel W...	1.843 KB
Cartera_credito_2014_05_31	16/10/2018 01:42 a...	Microsoft Excel W...	1.871 KB
Cartera_credito_2014_06_30	16/10/2018 01:43 a...	Microsoft Excel W...	1.875 KB
Cartera_credito_2014_07_31	16/10/2018 01:43 a...	Microsoft Excel W...	1.870 KB
Cartera_credito_2014_08_31	16/10/2018 01:44 a...	Microsoft Excel W...	1.869 KB
Cartera_credito_2014_09_30	16/10/2018 01:44 a...	Microsoft Excel W...	1.884 KB
Cartera_credito_2014_10_31	16/10/2018 01:45 a...	Microsoft Excel W...	1.882 KB
Cartera_credito_2014_11_30	16/10/2018 01:45 a...	Microsoft Excel W...	1.847 KB
Cartera_credito_2014_12_31	16/10/2018 01:46 a...	Microsoft Excel W...	1.925 KB
Cartera_credito_2015_01_31	17/10/2018 07:56 ...	Microsoft Excel W...	1.784 KB
Cartera_credito_2015_02_28	17/10/2018 08:48 ...	Microsoft Excel W...	1.711 KB
Cartera_credito_2015_03_31	17/10/2018 09:08 ...	Microsoft Excel W...	1.696 KB
Cartera_credito_2015_04_30	17/10/2018 08:50 ...	Microsoft Excel W...	1.632 KB
Cartera_credito_2015_05_31	17/10/2018 08:51 ...	Microsoft Excel W...	1.665 KB

Figura 5.7. Composición de archivos en el directorio de carga intermedia

Nota: elaboración propia.

La estructura que deben mantener los archivos de Excel para el procesamiento del ETL es:

- Año
- Mes
- Ejecutivo
- Nombre Ejecutivo
- Cod cliente
- ID
- Nombre
- Operación
- Descripción Producto
- Descripción Subproducto
- Estado
- Mto Principal Directo Moneda
- Mto Principal Contingente Moneda
- Mto Productos X Cobrar Moneda
- Moneda Oper
- Mto Principal Directo Colonizado
- Mto Principal Contingente
- Colonizado
- Mto Producto Colonizado
- Mto Cta X Cobrar Colonizado
- Mto Total Deuda Colonizado
- Capacidad de Pago
- Comportamiento Pago Histórico
- Categoría de Riesgo
- Dias atraso
- Mto Mitigador
- Mto Estimación Genérica
- Mto Estimación Especifica Saldo Cubierto
- Mto Estimación Especifica Saldo Descubierta
- Generador de Divisa
- Desc.Actividad Económica

Propuesta para la elaboración de reportes crediticios de una entidad financiera

La segunda tarea de la carga cíclica se encarga de la conversión de los datos extraídos al tipo string. Esta tarea es fundamental ya que se deben convertir los datos antes de realizar la carga a la base de datos intermedia. La Figura 5.8 muestra la conversión de los datos ejecutada sobre la información de los archivos de Excel.

Input Column	Output Alias	Data Type	Length
Desc# Actividad Econó...	DesActividad Económic...	string [DT_STR]	200
Generador de Divisa	Generador de Divisa_STR	string [DT_STR]	10
Categoría de Riesgo	Categoría de Riesgo_STR	string [DT_STR]	10
Moneda Oper	Moneda Oper_STR	string [DT_STR]	10
Estado	Estado_STR	string [DT_STR]	20
Nombre	Nombre_STR	string [DT_STR]	200
Operación	Operación_STR	string [DT_STR]	100
Descripción Subproducto	Descripción Subproduct...	string [DT_STR]	70
Descripción Producto	Descripción Producto_STR	string [DT_STR]	70
ID	ID_STR	string [DT_STR]	50
Nombre Ejecutivo	Nombre Ejecutivo_STR	string [DT_STR]	100
Capacidad de Pago	Capacidad de Pago_STR	string [DT_STR]	10
Comportamiento Pago ...	Comportamiento Pago ...	string [DT_STR]	10
Año	Año_STR	string [DT_STR]	10

Figura 5.8. Codificación de datos para carga a base de datos intermedia

Nota: elaboración propia.

La carga cíclica culmina con el almacenamiento de la información en la base de datos intermedia. Para ejecutar el proceso de almacenamiento se requiere mapear los atributos extraídos de los archivos fuente contra los atributos que integran las tablas de la base relacional intermedia. La Figura 5.9 muestra el mapeo de campos realizado entre la fuente transaccional y la base de datos destino.

Los atributos de las tablas que constituyen la base de datos intermedia se encuentran listados en el Apéndice L: Composición de las tablas utilitarias para proceso de ETL.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

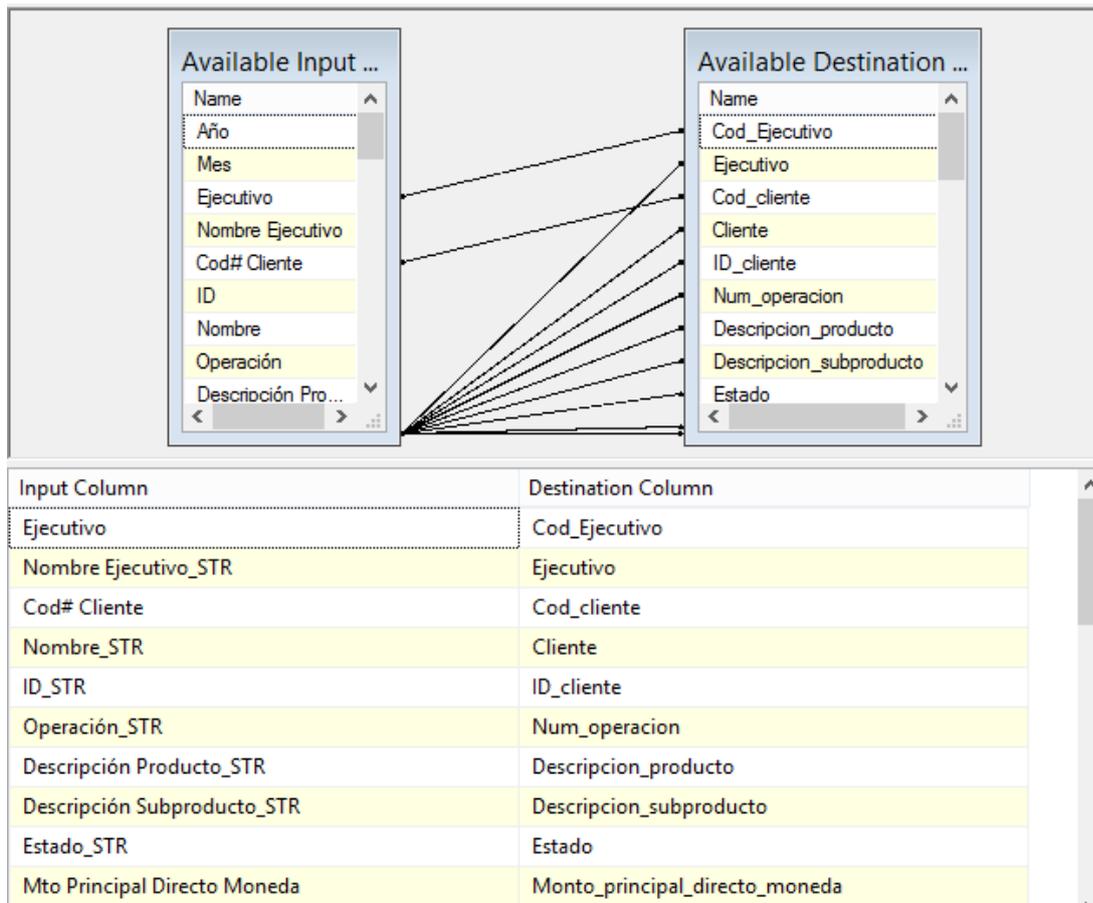


Figura 5.9. Mapeo de campos para carga a la base de datos intermedia de la solución.

Nota: elaboración propia.

El resultado final de la ejecución de la carga de datos se corrobora contra la consulta directa a la base de datos intermedia de la solución, la cual se muestra en la Figura 5.10.

	Mto_productosxCobrar_moneda	Moneda_operacion	Monto_principal_directo_colonizado	Monto_principal_contingente_colonizado	Mto_producto_colonizado
1	1083,07	USD	28556248,27	0	547372,75
2	806,14	USD	17760177,85	0	407415,09
3	290,4	USD	8723011,18	0	146765,26
4	1896,89	USD	15538806,86	0	958669,24
5	25,66	USD	3112929,49	0	12968,31
6	16,74	USD	1891710,15	0	8460,23
7	205,38	USD	10714283,16	0	103797
8	78,43	USD	4721085,52	0	39637,74
9	332,93	USD	21422622,94	0	168259,49
10	25,81	USD	5022484,96	0	13044,12

Figura 5.10. Consulta sobre los datos de la base de datos intermedia de Crédito.

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

La información almacenada en el repositorio intermedio se utiliza como insumo para iniciar el procesamiento de la información crediticia que se almacena en las tablas dimensionales.

5.4.2.3. Carga de Dimensión ejecutivo

La carga a la dimensión ejecutivo incluye la extracción y almacenamiento de los datos del personal de ventas en la tabla dimensional de Ejecutivo. La Figura 5.11 muestra las tareas ejecutadas durante esta actividad.



Figura 5.11. Flujo de carga a la dimensión ejecutivo

Nota: elaboración propia.

Para realizar la extracción de datos del personal de ventas se realizó la siguiente consulta utilizando el lenguaje Transact – SQL sobre la base de datos intermedia.

```
SELECT DISTINCT Cod_Ejecutivo, Ejecutivo  
FROM Tabla_operaciones
```

Esta consulta permitió seleccionar solamente los registros únicos de ejecutivos en la base de datos intermedia, disminuyendo la cantidad de registros que se debían insertar en la tabla dimensional de 241629 filas a únicamente 51 filas.

5.4.2.4. Carga de Dimensión Cliente

Seguidamente se realizó el procesamiento de la información del cliente, esta actividad se encargó principalmente de la transformación de códigos numéricos y alfanuméricos a información categórica basada en rangos descriptivos que facilitan el análisis de los datos. La Figura 5.12 muestra el flujo de las tareas que componen esta actividad.

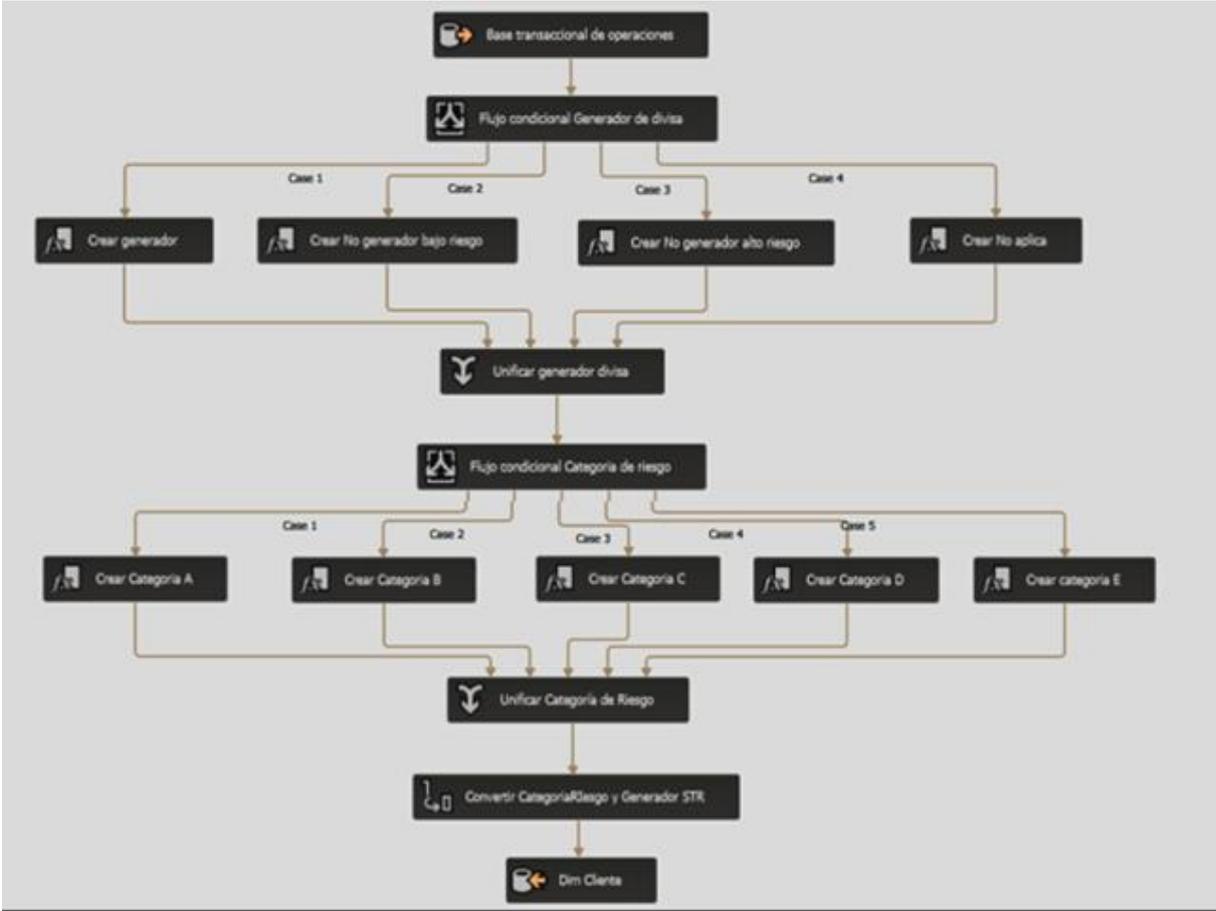


Figura 5.12. Flujo de carga a la dimensión cliente.

Nota: elaboración propia.

Como se puede observar en la Figura 5.12 el flujo de la actividad inicia por la extracción de los datos disponibles en la base de datos intermedia. Seguidamente se distribuye el procesamiento por medio de una estructura condicional que evalúa el código numérico que indica si un cliente es generador de divisas. La Figura 5.13 muestra las condiciones evaluadas para separar el flujo del proceso.

Order	Output Name	Condition
1	Case 1	Generador_divisa == "1"
2	Case 2	Generador_divisa == "2"
3	Case 3	Generador_divisa == "3"
4	Case 4	Generador_divisa == "4"

Figura 5.13. Estructura condicional para el atributo generador de divisas de un cliente.

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Por cada bifurcación del flujo se creó una columna nueva sobre el conjunto de datos que permitió transformar la codificación numérica del indicador Generador de divisa a un valor textual descriptivo según los requerimientos de los analistas de negocio. La Figura 5.14 ejemplifica una de las tareas utilizadas para crear la nueva columna que describe si un cliente es generador de divisas o no.

Derived Column Name	Derived Column	Expression	Data Type
NoGeneradorBajo	<add as new column>	"No generador bajo riesgo"	Unicode string [DT_WSTR]

Figura 5.14. Creación de nueva columna para el atributo Generador de divisa.

Nota: elaboración propia.

Una vez finalizada la creación de la nueva columna se realizó la unificación del flujo de procesamiento, esto permitió mantener solo una columna para representar el valor de los generadores de divisa en el conjunto de datos. Seguidamente se bifurca el flujo de procesamiento para evaluar la categoría de riesgo del cliente, la Figura 5.15 muestra la estructura condicional que divide el flujo del proceso.

Order	Output Name	Condition
1	Case 1	Categoria_riesgo == "A1" Categoria_riesgo == "A2"
2	Case 2	Categoria_riesgo == "B1" Categoria_riesgo == "B2"
3	Case 3	Categoria_riesgo == "C1" Categoria_riesgo == "C2"
4	Case 4	Categoria_riesgo == "D" Categoria_riesgo == "E"

Figura 5.15. Bifurcación del flujo de procesamiento basado en categoría de riesgo

Nota: elaboración propia.

Cada rama del flujo ejecuta la creación de una columna nueva para representar la categoría de riesgo, transformando el código alfanumérico de la fuente transaccional a un valor textual descriptivo según los requerimientos de las áreas de negocio. La Figura 5.16 ejemplifica una de las tareas de creación para la nueva categoría de riesgo.

Derived Column Name	Derived Column	Expression	Data Type
CategoriaA	<add as new column>	"A"	Unicode string [DT_WSTR]

Figura 5.16. Creación de la columna para la nueva categoría de riesgo

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Una vez finalizada la creación de las nuevas categorías de riesgo se unificó el flujo del proceso, permitiendo el manejo de una columna que almacena la transformación de la categoría de riesgo del cliente para su posterior carga a la tabla dimensional de clientes.

Luego de ejecutar la transformación de los datos correspondientes al atributo Generador de divisa y Categoría de riesgo se procedió a convertir los nuevos atributos al tipo de datos string requerido para su almacenamiento en el modelo multidimensional como se muestra en la Figura 5.17.

Input Column	Output Alias	Data Type	Length
CategoriaUnificado	CategoriaUnificadoSTR	string [DT_STR]	1
GeneradorDivisaSTR	GeneradorDivisaSTR	string [DT_STR]	24

Figura 5.17. Transformación del tipo de datos para los atributos Generador de divisa y Categoría de Riesgo

Nota: elaboración propia.

El procesamiento de la dimensión cliente culmina con la carga de los datos transformados a la tabla dimensional de clientes, esto se lleva a cabo por medio del mapeo de atributos entre la fuente de datos y el DM que se muestra en la Figura 5.18.

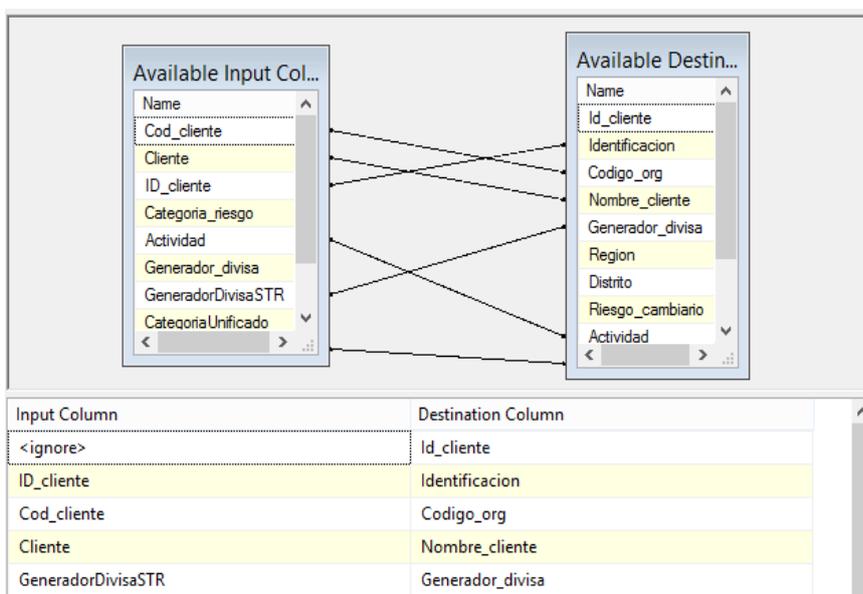


Figura 5.18. Mapeo de los datos procesados y la dimensión cliente

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Cabe destacar que se migró la totalidad de los registros de los clientes registrados en las carteras crediticias, ya que los atributos de cada cliente pueden variar mes a mes. Para efectos de control se requiere visualizar el cambio en los atributos del cliente, especialmente sobre su categoría de riesgo.

5.4.2.5. Carga de Dimensión operación

Esta actividad ejecuta la extracción de los datos transaccionales de cada operación, adicionalmente se encarga de aplicar las reglas de transformación solicitadas por las áreas de negocio para facilitar la construcción de los reportes crediticios. La Figura 5.19 muestra el flujo de las tareas que componen el procesamiento de las operaciones de crédito.

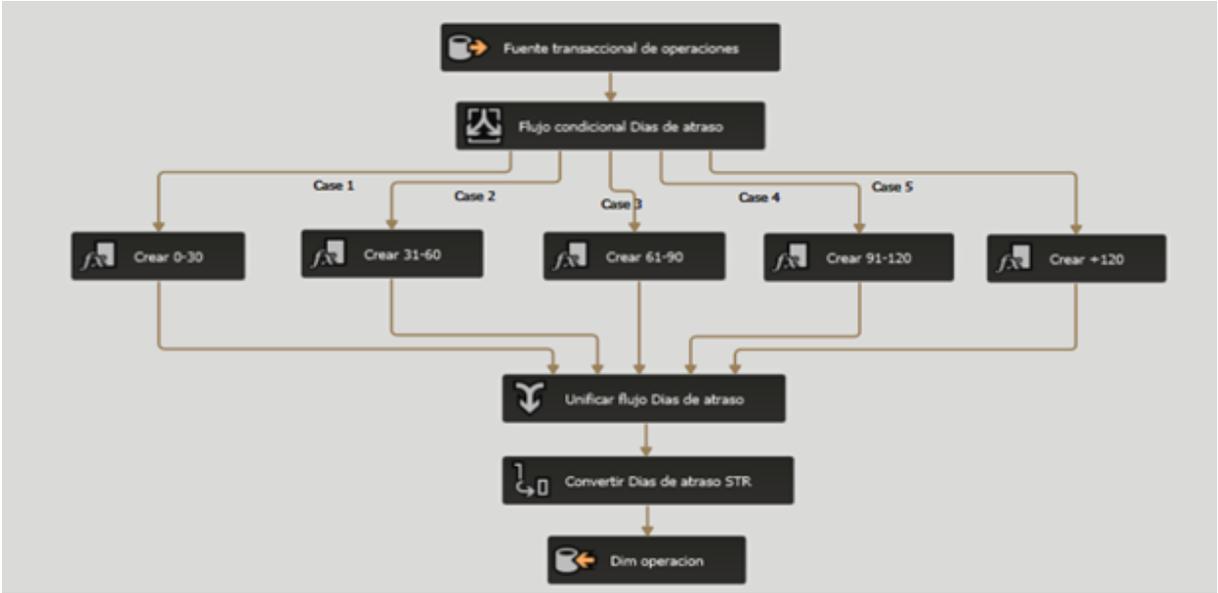


Figura 5.19. Flujo de carga a la dimensión operación.

Nota: elaboración propia.

Esta actividad inicia con la extracción de los datos almacenados en el repositorio intermedio de crédito que se mencionó en la sección 5.4.2.2 de este documento. Seguidamente se realizó la división del flujo de procesamiento basado en la cantidad de días de atraso que posee cada operación, la Figura 5.20 muestra las condiciones evaluadas para cada registro.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Order	Output Name	Condition
1	Case 1	0 <= Dias_atraso && Dias_atraso <= 30
2	Case 2	31 <= Dias_atraso && Dias_atraso <= 60
3	Case 3	61 <= Dias_atraso && Dias_atraso <= 90
4	Case 4	91 <= Dias_atraso && Dias_atraso <= 120
5	Case 5	Dias_atraso > 120

Figura 5.20. Estructura condicional de días de atraso para división del flujo de procesamiento

Nota: elaboración propia.

Cada rama del flujo ejecuta la creación de un atributo nuevo donde se almacena el rango de valores que describe la cantidad de días de atraso en el pago de una operación crediticia. Una vez creado el nuevo atributo en cada bifurcación del flujo se procede a unificar el procesamiento de datos para mantener solamente una columna que represente los días de atraso para cada operación específica.

Una vez integrado el flujo de información se procede a convertir el tipo de datos de la columna correspondiente a los días de atraso al tipo *string*, requerido para su inserción en la tabla dimensional de operaciones.

Esta actividad culmina con la carga de la información a la tabla dimensional de operaciones, esto requirió el mapeo de los campos transformados contra la tabla destino en el modelo multidimensional como se observa en la Figura 5.21.

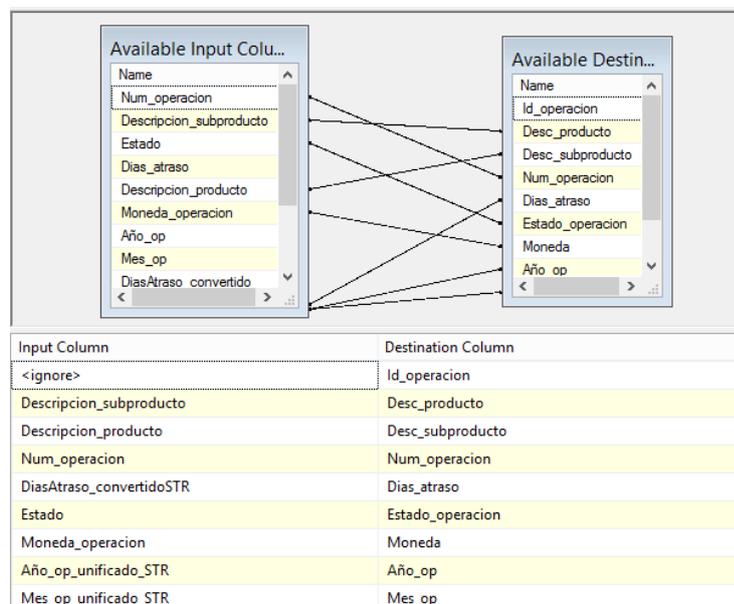


Figura 5.21. Mapeo de los campos de la fuente transaccional contra la tabla de operaciones en el DM

Nota: elaboración propia.

5.4.2.6. Carga de Dimensión Tiempo

La cuarta actividad que se ejecuta durante el proceso de ETL es la carga de la dimensión tiempo que toma los registros de la fuente transaccional, aplica la segmentación temporal de los registros para su carga posterior a la tabla dimensional destino. El flujo del proceso para dimensión tiempo se muestra Figura 5.22.

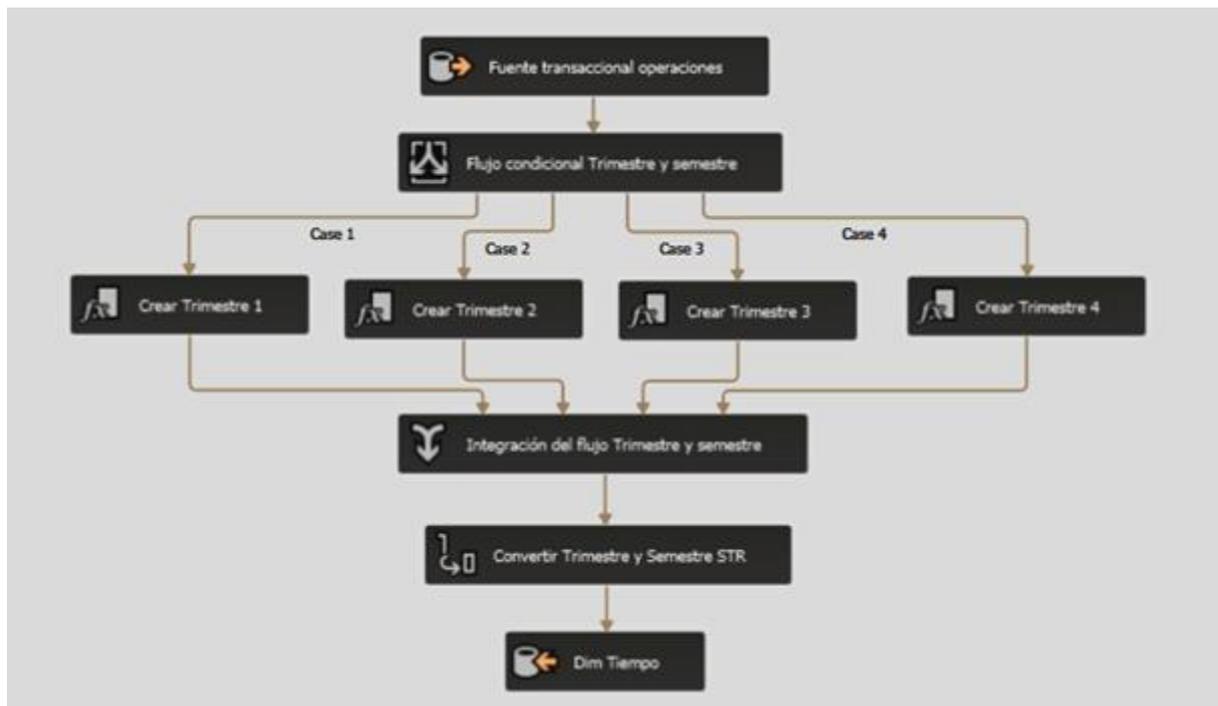


Figura 5.22. Flujo de carga a la dimensión tiempo

Nota: elaboración propia.

Esta actividad inicia con la extracción de los datos disponibles en la base de datos intermedia de la solución, para extraer los valores temporales únicos de la tabla de operaciones se ejecutó la siguiente consulta.

```
SELECT DISTINCT Año, Mes  
FROM Tabla_operaciones
```

Este filtro de extracción permitió reducir la cantidad de registros extraídos y procesados de 241629 a 24 registros únicos, uno por cada mes de la carta crediticia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Seguidamente se ejecutó una bifurcación condicional sobre el año y mes extraídos previamente, que permitió crear los atributos nuevos para trimestre y semestre la Figura 5.23 muestra las sentencias evaluadas en la información crediticia.

Order	Output Name	Condition
1	Case 1	Mes == "01" Mes == "02" Mes == "03"
2	Case 2	Mes == "04" Mes == "05" Mes == "06"
3	Case 3	Mes == "07" Mes == "08" Mes == "09"
4	Case 4	Mes == "10" Mes == "11" Mes == "12"

Figura 5.23. Estructura condicional de la dimensión temporal.

Nota: elaboración propia.

Cada ramificación del flujo se encarga de crear dos columnas nuevas que almacenan el número de trimestre y semestre asociado a cada mes como se muestra en la Figura 5.24.

Derived Column Name	Derived Column	Expression	Data Type
Trimestre1	<add as new column>	"Trimestre 1"	Unicode string [DT_WSTR]
Semestre 1	<add as new column>	"Semestre 1"	Unicode string [DT_WSTR]

Figura 5.24. Creación de nuevos atributos para la primera ramificación del flujo en dimensión tiempo

Nota: elaboración propia.

Al finalizar la creación de las nuevas columnas se unificó el flujo de datos nuevamente y se realizó la conversión del tipo de datos de las nuevas columnas a tipo string como se muestra en la Figura 5.25.

Input Column	Output Alias	Data Type	Length
Trimestre1	TrimestreSTR	string [DT_STR]	11
Semestre 1	SemestreSTR	string [DT_STR]	10

Figura 5.25. Conversión de las columnas trimestre y semestre

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Esta actividad culmina con el mapeo y posterior carga de la información a la tabla dimensional Tiempo. La Figura 5.26 muestra el mapeo realizado entre los datos de la fuente transaccional y la tabla de dimensión tiempo.

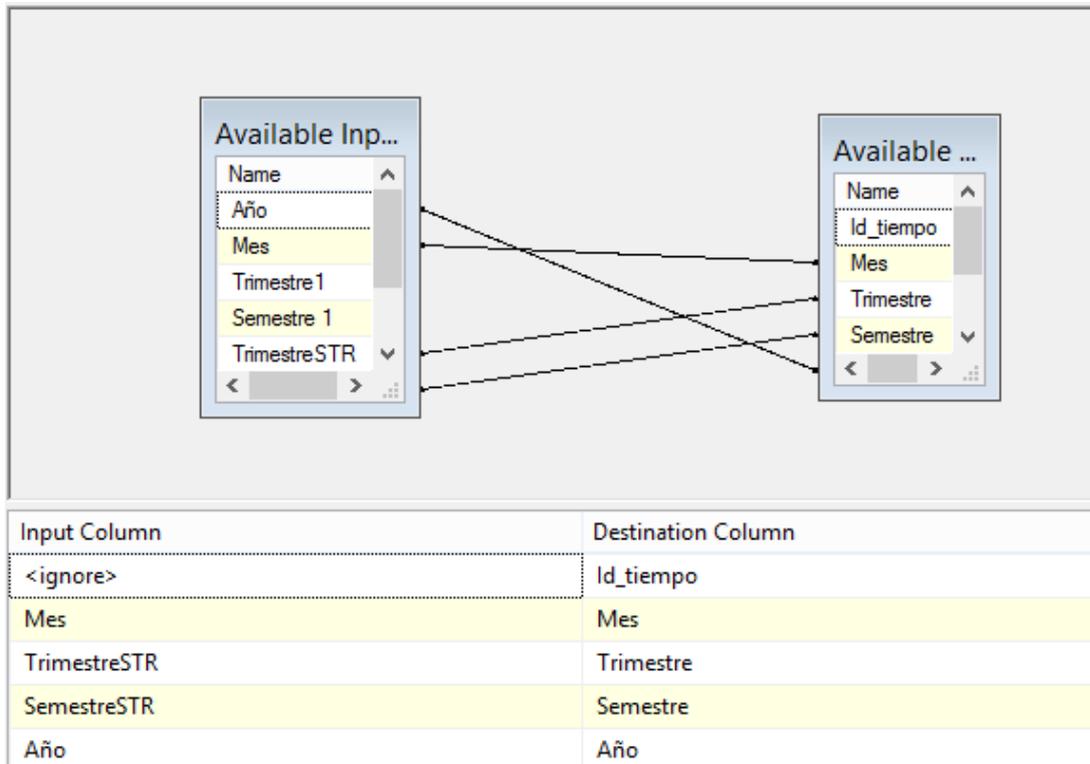


Figura 5.26. Mapeo de información de la fuente transaccional y la dimensión tiempo

Nota: elaboración propia.

Con la ejecución de la carga a la tabla dimensional de tiempo se finaliza el procesamiento de las tablas dimensionales. Con las estructuras de dimensión pobladas se procedió a ejecutar la actividad final del proceso de ETL que corresponde a la carga de la tabla de hechos del modelo multidimensional.

5.4.2.7. Carga de la Tabla de hechos de crédito

Esta actividad se ejecutó al final del proceso ya que para construir la relación entre los registros de tabla de hechos y las filas almacenadas en las tablas dimensionales se requiere que cada dimensión esté poblada con los datos requeridos. La Figura 5.27 muestra el flujo de procesamiento correspondiente a la carga de la tabla de hechos de la solución.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

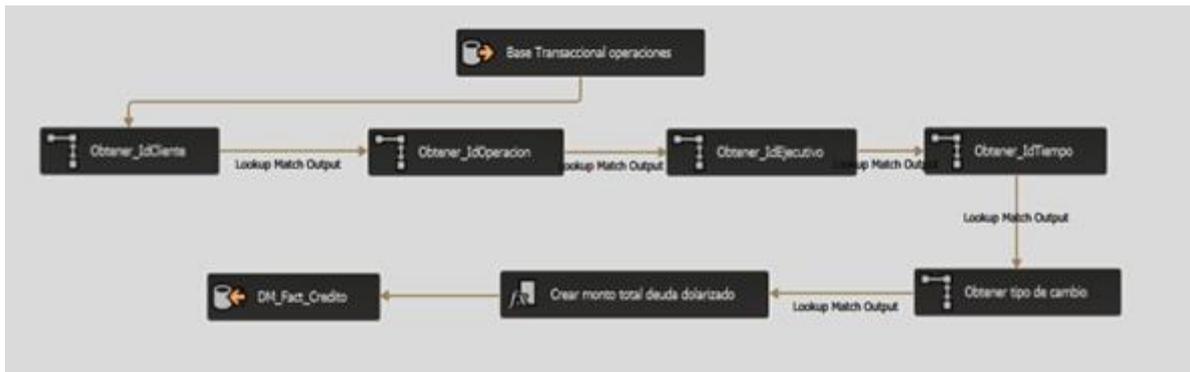


Figura 5.27. Flujo de carga a la tabla de hechos

Nota: elaboración propia.

La primera tarea de esta actividad se encarga de extraer los datos de la base intermedia de operaciones.

Seguidamente, se procede con el mapeo de las relaciones entre las tablas de dimensión y la base intermedia de operaciones, esto permite obtener las llaves foráneas que se insertaran en la última tarea de carga de esta actividad. La Figura 5.28 muestra un ejemplo del mapeo y realizado entre la base de datos intermedia de crédito y la tabla dimensional de operaciones, el resultado de esta tarea es la llave primaria de la dimensión operación que se utilizó como llave foránea para la tabla de hechos del *Data mart*.

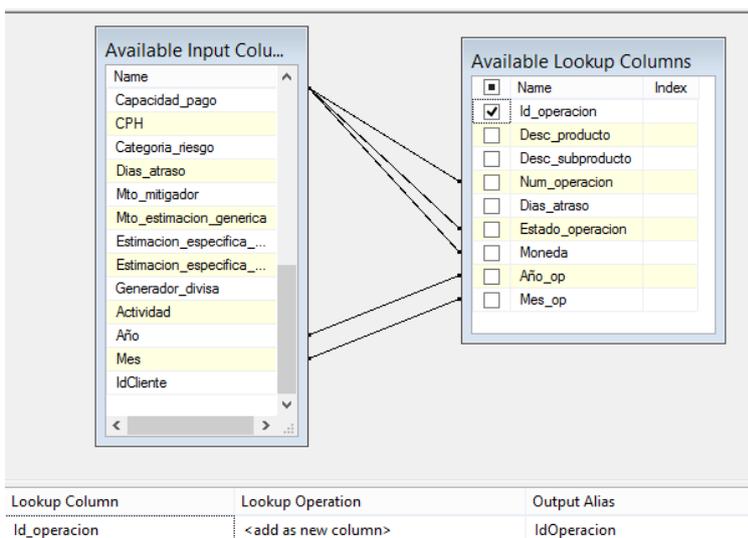


Figura 5.28. Mapeo para la búsqueda de la llave primaria de la dimensión operación

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Una vez extraídas las llaves primarias de las tablas dimensionales, se procede con las tareas de transformación y carga de las medidas que se incorporaron en la tabla de hechos. En primer lugar, se realiza una búsqueda del tipo de cambio almacenado en una tabla utilitaria de la base de datos intermedia. La Figura 5.29 muestra la extracción realizada sobre la tabla de tipo de cambio de la base de datos transaccional.

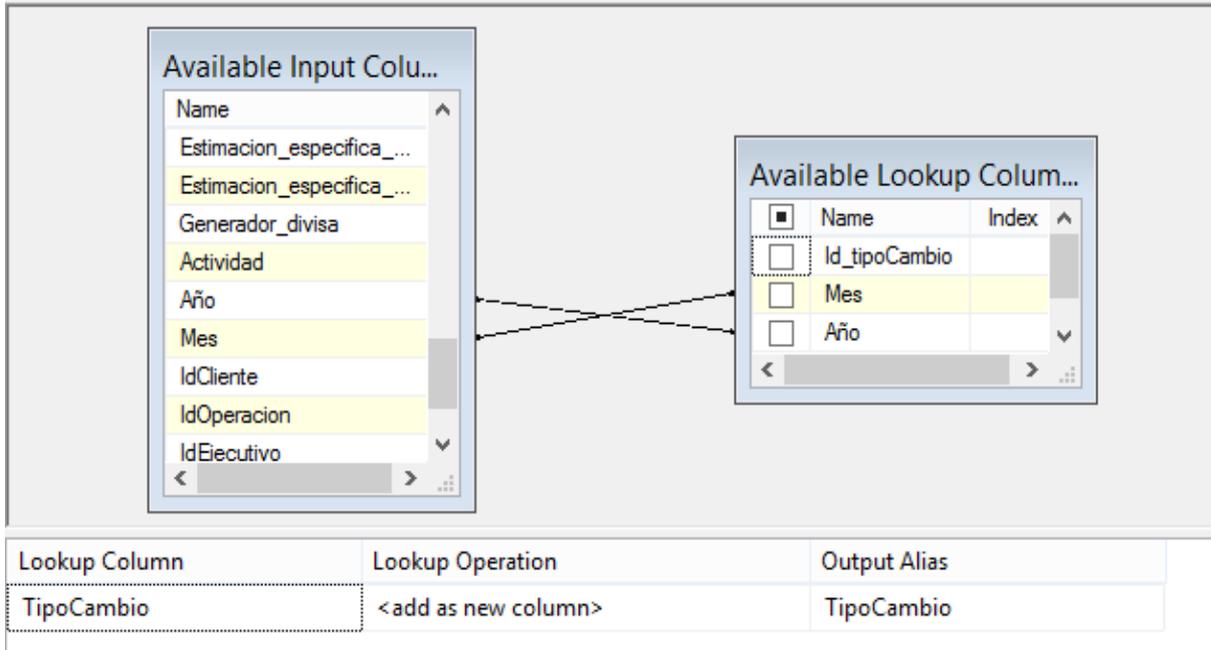


Figura 5.29. Búsqueda del tipo de cambio para las transformaciones de saldos en la tabla de hechos.

Nota: elaboración propia.

Seguidamente se creó un nuevo atributo para almacenar el monto total de la deuda dolarizado. La creación del atributo se realizó al multiplicar el monto total de la deuda colonizado disponible en la base de datos intermedia y el tipo de cambio extraído durante la tarea anterior. La Figura 5.30 muestra la implementación del campo calculado para la métrica dolarizada.

Derived Column Name	Derived Column	Expression	Data Type
Monto_total_deudaDolari...	<add as new column>	Mto_total_deuda_colonizado * TipoCambio	double-precision float [D...

Figura 5.30. Transformación del monto total de la deuda a dólares

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Una vez extraídas las llaves primarias de las dimensiones y transformados los saldos de la fuente transaccional se procedió con el mapeo de los atributos de la base intermedia y la tabla de hechos del modelo dimensional. La Figura 5.31 muestra el mapeo de campos para la carga de la información a la tabla de hechos del DM.

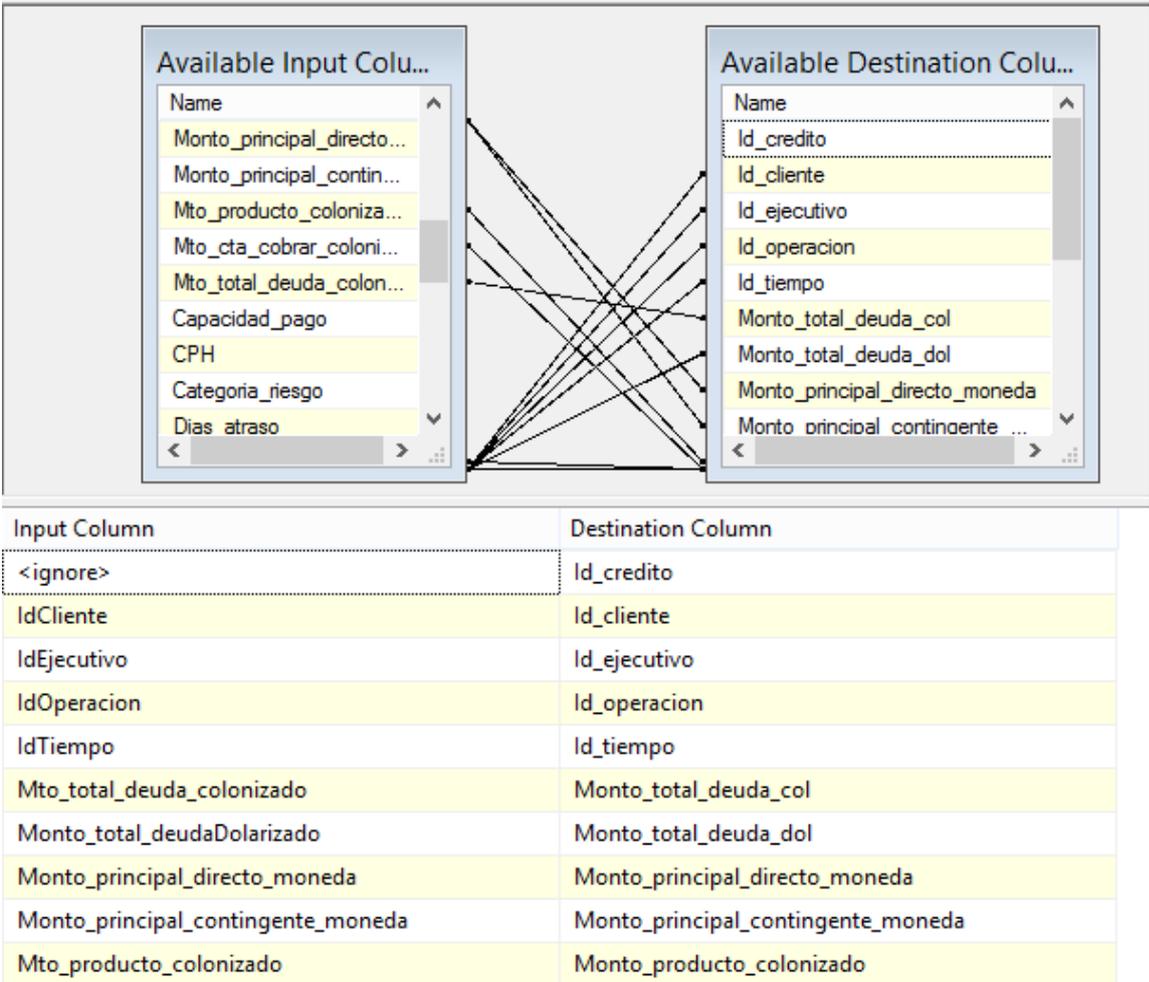


Figura 5.31. Mapeo de atributos para la tabla de hechos

Nota: elaboración propia.

Al finalizar la ejecución de esta actividad se cuenta con el repositorio multidimensional poblado con los datos de las operaciones de crédito de la entidad financiera que provee los insumos para los otros componentes desarrollados en la solución que incluyen las visualizaciones de datos.

5.4.3. Implementación de las visualizaciones de datos

Utilizando como insumo la información procesada por las tareas del ETL se construyeron las tres visualizaciones diseñadas en la sección 5.3 de este documento. El desarrollo de los reportes gráficos se realizó por medio de la herramienta Tableau y con el objetivo de identificar los siguientes aspectos en la cartera crediticia.

- 📦 Tendencia del monto total de la deuda por categoría de riesgo y días de atraso.
- 📦 Distribución del monto directo moneda anual por días de atraso y moneda.
- 📦 Distribución de la estimación genérica por año.

5.4.3.1. Implementación del reporte de tendencia del monto total de la deuda por categoría de riesgo y días de atraso

Por medio de la funcionalidad de *Drag and Drop* que provee la solución Tableau se construyó un gráfico donde se muestra la tendencia por mes del Monto total de la deuda dolarizado para los años 2014 y 2015. La Figura 5.32 muestra el informe gráfico construido en la herramienta.

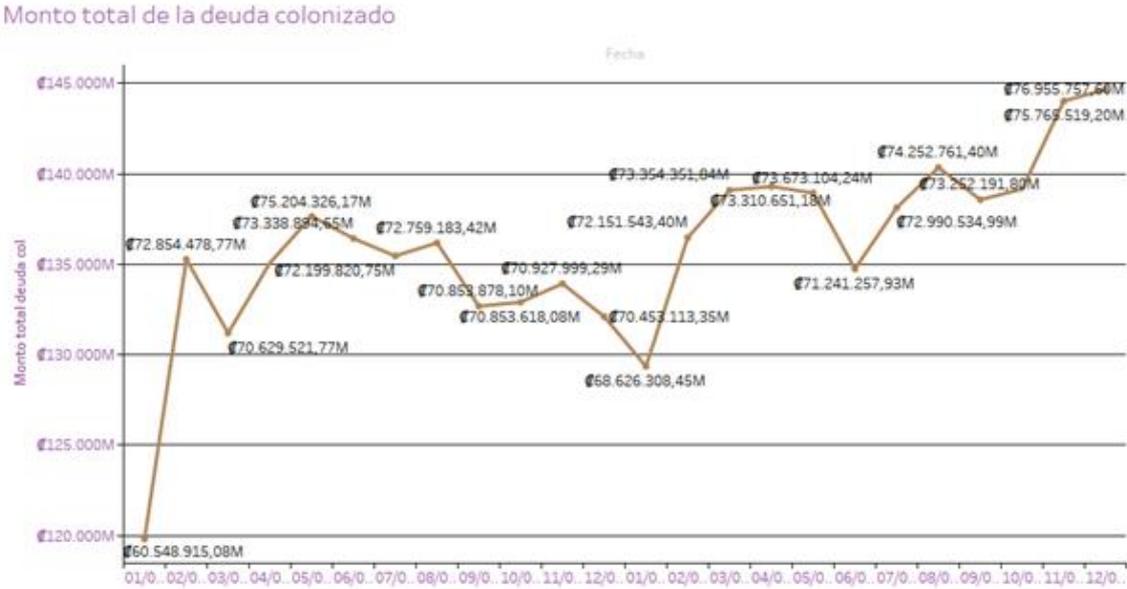


Figura 5.32. Reporte gráfico de monto total de la deuda dolarizado

Nota: elaboración propia.

5.4.3.2. Implementación del reporte de distribución del monto directo anual de las operaciones por días de atraso y moneda

Por medio de la funcionalidad de *Drag and Drop* que provee la solución Tableau se construyó un gráfico que muestra la distribución del monto principal directo de las operaciones por moneda y el rango de días de atraso para los periodos 2014 y 2015. La Figura 5.33 muestra el informe gráfico construido en la herramienta.

Estimación genérica por año

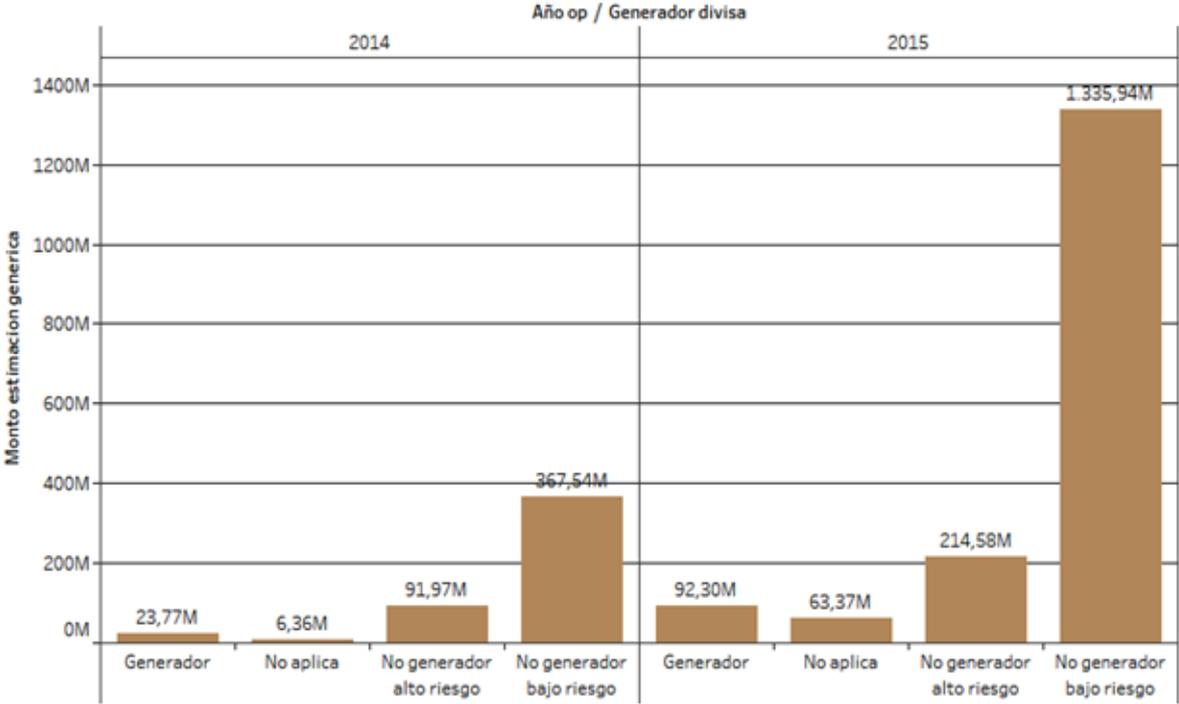


Figura 5.33. Reporte gráfico del monto principal directo por moneda

Nota: elaboración propia.

4.4.4.3. Implementación del reporte de distribución de la estimación genérica por año

Utilizando la herramienta Tableau se construyó el reporte gráfico que muestra la distribución de la estimación genérica de las operaciones para las categorías del atributo Generador de divisa en los periodos 2014 y 2015. La Figura 5.34 muestra el informe gráfico construido en la herramienta.

Principal directo diferencia interanual por moneda

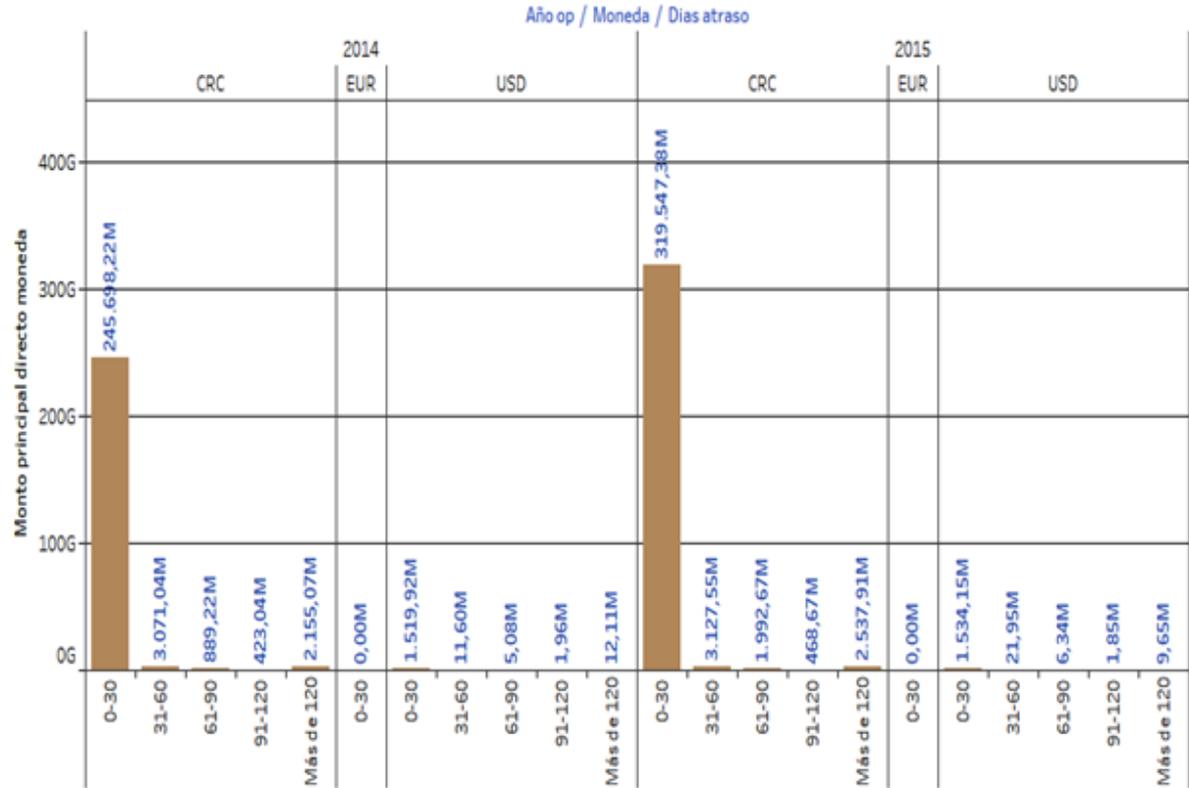


Figura 5.34. Reporte gráfico del monto de estimación genérica por año

Nota: elaboración propia.

Una vez implementados los reportes gráficos se procede a construir el *dashboard* de la cartera crediticia que incorpora las tres visualizaciones desarrolladas, el resultado de la unificación de reportes visuales se muestra en la Figura 5.35.

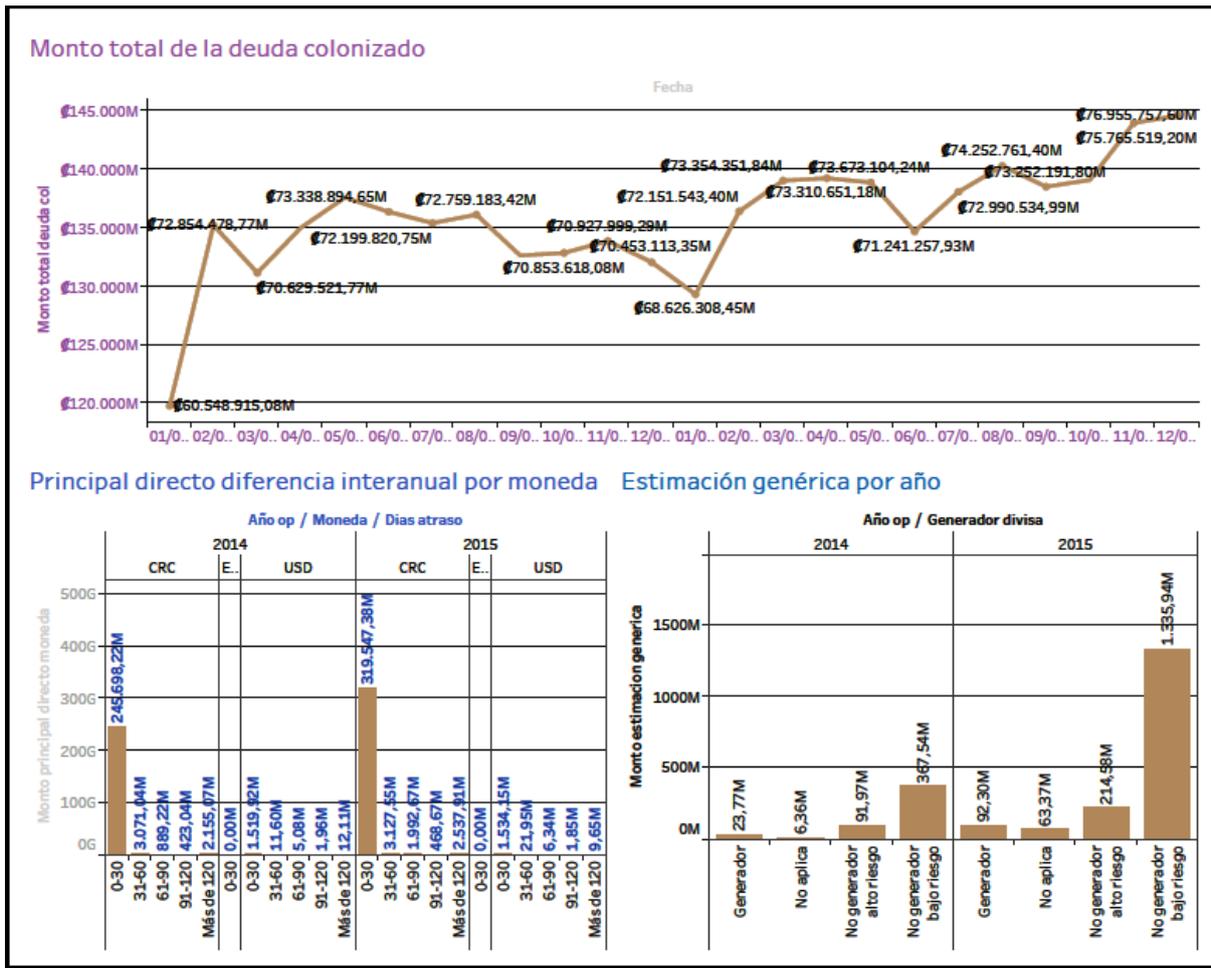


Figura 5.35. Dashboard de la cartera crediticia

Nota: elaboración propia.

Al finalizar la implementación de los componentes diseñados como parte de la propuesta de solución se procedió con la comprobación de resultados que buscaba verificar la efectividad de la solución implementada con respecto al planteamiento inicial del proyecto.

5.5. Comprobación de resultados del proceso de ETL

La comprobación de resultados del proceso de extracción, transformación y carga incluye la validación de la cantidad de registros procesados con respecto a la fuente original de información y la verificación de las transformaciones de datos ejecutadas. A continuación, se describen los resultados obtenidos.

5.5.1. Validación de la cantidad de registros procesados

Para verificar que la carga de información se ejecutó correctamente se llevaron a cabo una serie de validaciones sobre los datos migrados al modelo multidimensional del *Data mart*. La Tabla 5.4 muestra la cantidad original de operaciones crediticias para los años procesados como parte de la propuesta de solución.

Tabla 5.4.

Cantidad de registros en la fuente transaccional

Año de cartera	Cantidad total de registros en la fuente original
2014	126522
2015	115107
Total de registro procesados	241629

Nota: elaboración propia

Una vez establecida la cantidad de registros disponibles en las fuentes transaccionales se procedió a validar el resultado contra el número de registros del DM por medio de consultas sobre las tablas dimensionales y de hechos. A continuación, se muestran las verificaciones ejecutadas sobre el modelo implementado.

5.5.1.1. Verificación de la Dimensión tiempo

En la ejecución del proceso ETL se transformaron 24 archivos de cartera crediticia, uno por cada mes en un periodo de dos años, esto brinda un total de 24 combinaciones únicas de los atributos año y mes. La Figura 5.36 muestra el resultado de la consulta ejecutada sobre la dimensión tiempo del modelo multidimensional.

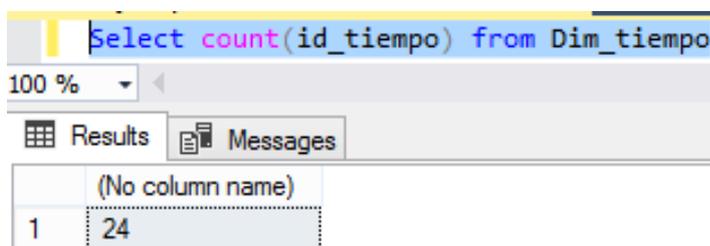
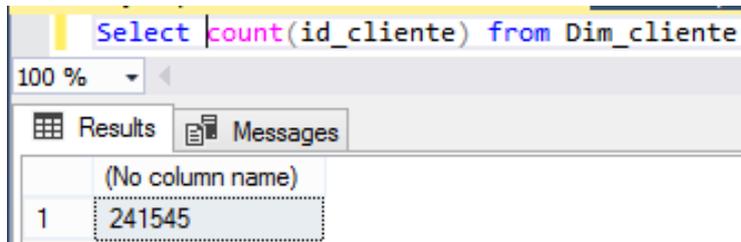


Figura 5.36. Consulta ejecutada sobre la dimensión tiempo

Nota: elaboración propia.

5.5.1.2. Verificación de la Dimensión Cliente

Para esta prueba se consultó la cantidad de registros insertados en la dimensión cliente y se comparó contra el total de operaciones registradas para los clientes durante el mes de cierre de la cartera crediticia. La Figura 5.37 muestra el resultado de la consulta realizada sobre la dimensión cliente del modelo multidimensional.



```
Select count(id_cliente) from Dim_cliente
```

(No column name)
1

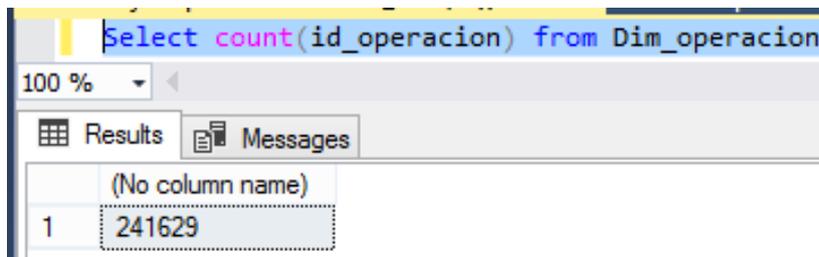
The screenshot shows a SQL query execution window. The query is 'Select count(id_cliente) from Dim_cliente'. The results pane shows a single row with the value 241545. The window title is '100 %' and it has tabs for 'Results' and 'Messages'.

Figura 5.37. Consulta ejecutada sobre la dimensión cliente

Nota: elaboración propia.

5.5.1.3. Verificación de la Dimensión operación

Para el desarrollo de esta prueba se consultó la cantidad de registros insertados en la tabla dimensional de operaciones y se corroboró contra la cantidad total de operaciones registradas en las carteras de crédito (241629 operaciones). La Figura 5.38 muestra el resultado de la consulta realizada sobre la dimensión operación del modelo multidimensional.



```
Select count(id_operacion) from Dim_operacion
```

(No column name)
1

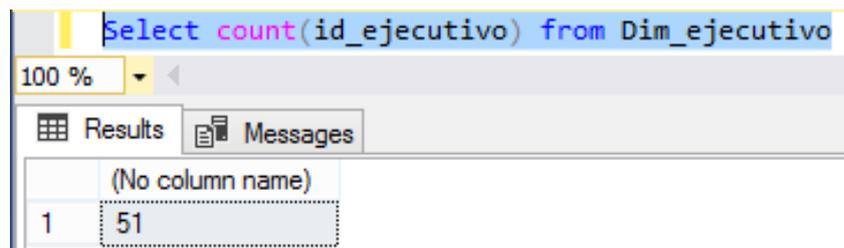
The screenshot shows a SQL query execution window. The query is 'Select count(id_operacion) from Dim_operacion'. The results pane shows a single row with the value 241629. The window title is '100 %' and it has tabs for 'Results' and 'Messages'.

Figura 5.38. Consulta ejecutada sobre la dimensión operación

Nota: elaboración propia.

5.5.1.4. Verificación de la Dimensión ejecutivo

Durante esta prueba se consultó la cantidad de registros únicos insertados en la tabla dimensional de ejecutivos y se corroboró contra la cantidad total de ejecutivos únicos incluidos en las carteras de crédito de la entidad. La Figura 5.39 muestra el resultado de la consulta ejecutada sobre la dimensión ejecutivo del modelo multidimensional.



```
Select count(id_ejecutivo) from Dim_ejecutivo
```

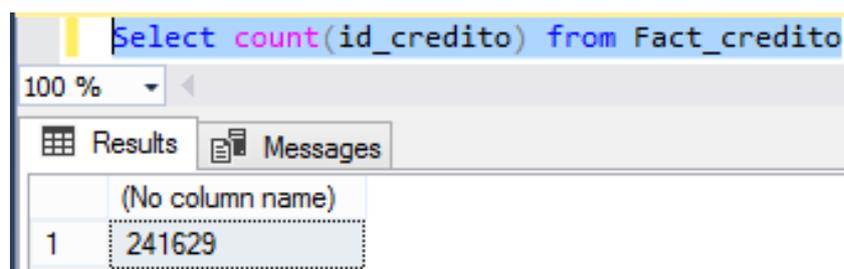
	(No column name)
1	51

Figura 5.39. Consulta ejecutada sobre la dimensión ejecutivo

Nota: elaboración propia.

5.5.1.5. Verificación de la tabla de hechos

Para el desarrollo de esta prueba se consultó la cantidad de identificadores únicos de operaciones de crédito insertados en la tabla de hechos de la solución y se corroboró contra la cantidad total de operaciones registradas en las carteras de crédito de la entidad. La Figura 5.40 muestra el resultado de la consulta ejecutada sobre la dimensión ejecutivo del modelo multidimensional.



```
Select count(id_credito) from Fact_credito
```

	(No column name)
1	241629

Figura 5.40. Consulta ejecutada sobre la tabla de hechos

Nota: elaboración propia.

5.5.2. Validación del mapeo para datos categóricos

La segunda validación sobre los datos procesados por el ETL se enfocó en la transformación de los atributos numéricos o alfanuméricos en rangos de valores o campos descriptivos. A continuación, se describen los resultados de las validaciones sobre el modelo multidimensional.

5.5.2.1. Verificación de la Dimensión cliente

En la dimensión cliente se realizaron dos transformaciones de códigos numéricos y alfanuméricos a información categórica. A continuación, se muestran los resultados obtenidos durante para cada atributo transformado.

Atributo Generador de divisa

En este caso se corroboró que en la tabla dimensional todos los registros incluyeran los valores Generador, No generador de alto riesgo, No generador de bajo riesgo y No aplica en el atributo Generador de divisa. La Figura 5.41 muestra la consulta realizada sobre la dimensión cliente.

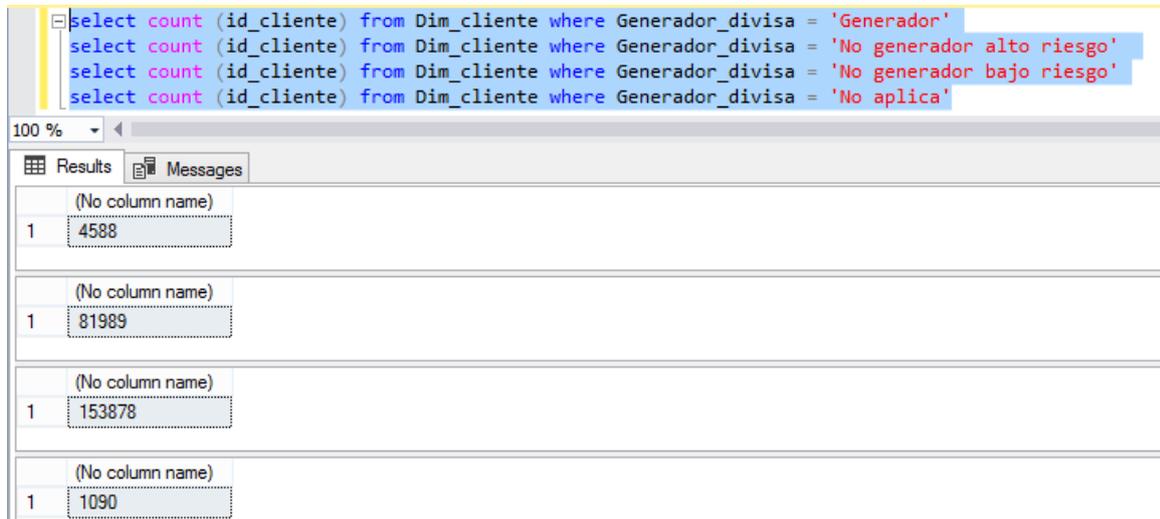


Figura 5.41. Consulta de clientes Generadores de divisas

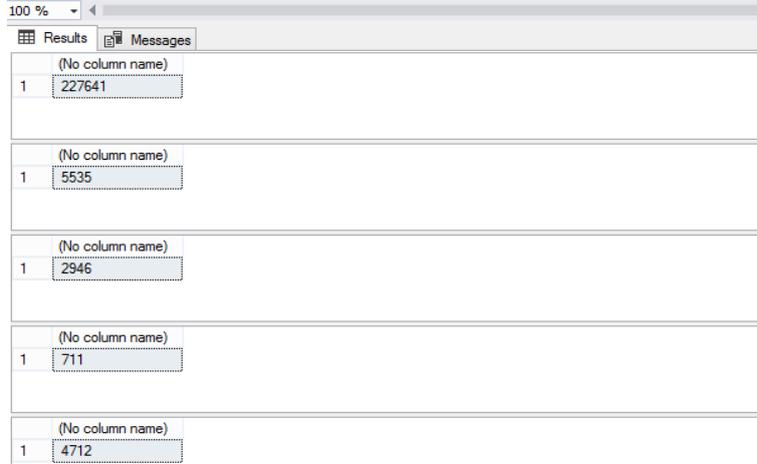
Nota: elaboración propia.

Al sumar el resultado de las consultas se obtiene que 241545 registros poseen una clasificación en el atributo generador de divisa, lo cual coincide con la cantidad de clientes registrados en las fuentes transaccionales.

Atributo Categoría de riesgo

En este caso se corroboró que en la tabla dimensional de clientes todos los registros incluyeran los valores A, B, C, D o E. La Figura 5.42 muestra la consulta realizada sobre la tabla dimensional.

```
select count(id_cliente) from Dim_cliente where Categoria_de_riesgo = 'A'  
select count(id_cliente) from Dim_cliente where Categoria_de_riesgo = 'B'  
select count(id_cliente) from Dim_cliente where Categoria_de_riesgo = 'C'  
select count(id_cliente) from Dim_cliente where Categoria_de_riesgo = 'D'  
select count(id_cliente) from Dim_cliente where Categoria_de_riesgo = 'E'
```



	(No column name)
1	227641
1	5535
1	2946
1	711
1	4712

Figura 5.42. Consultas sobre la categoría de riesgo de la dimensión cliente

Nota: elaboración propia.

Por medio de la consulta sobre la base de datos se determina que todos los clientes del modelo multidimensional poseen al menos un valor en el atributo categoría de riesgo.

5.5.2.2. Verificación de la tabla dimensional operación

En este caso se realizó una sola transformación de datos y esta corresponde al atributo numérico de días de atraso, durante el procesamiento del ETL se transformó este campo a un rango de valores textuales según lo especificado en la sección 5.2.2 de este capítulo.

En este caso particular se verificó que todas las operaciones registradas en la tabla dimensional tuvieran asignado un rango correspondiente a los días de atraso de la operación según la regla de negocio establecida por los analistas. Los resultados de las consultas ejecutadas sobre el modelo multidimensional se muestran a continuación.

```
select count(id_operacion) from Dim_operacion where Dias_atraso = '0-30'
```

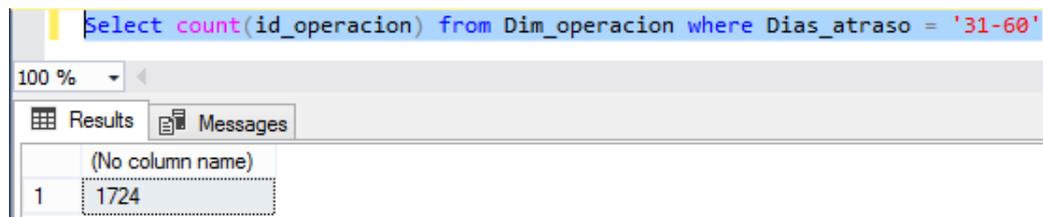


	(No column name)
1	237505

Figura 5.43. Consulta a la tabla operación sobre días de atraso de 0-30

Nota: elaboración propia.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

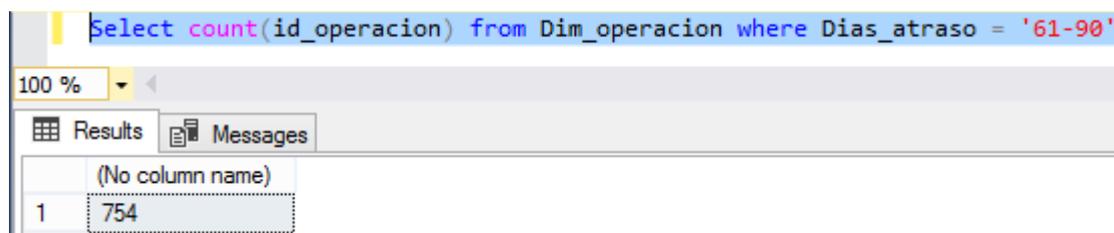


```
Select count(id_operacion) from Dim_operacion where Dias_atraso = '31-60'
```

	(No column name)
1	1724

Figura 5.44. Consulta a la tabla operación sobre días de atraso de 31-60

Nota: elaboración propia.

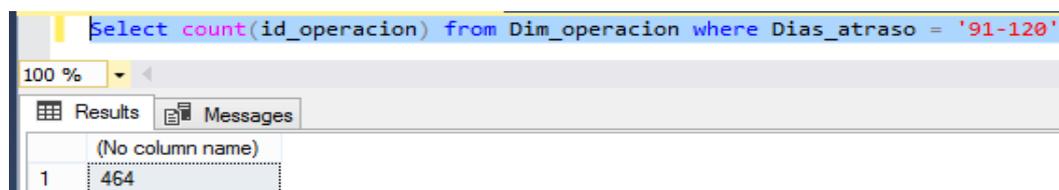


```
Select count(id_operacion) from Dim_operacion where Dias_atraso = '61-90'
```

	(No column name)
1	754

Figura 5.45. Consulta a la tabla operación sobre días de atraso de 61-90

Nota: elaboración propia.

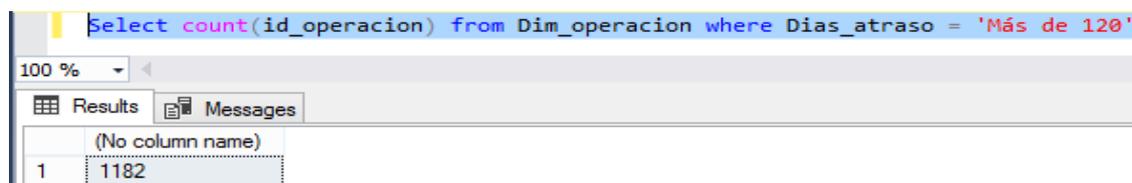


```
Select count(id_operacion) from Dim_operacion where Dias_atraso = '91-120'
```

	(No column name)
1	464

Figura 5.46. Consulta a la tabla operación sobre días de atraso de 91-120

Nota: elaboración propia.



```
Select count(id_operacion) from Dim_operacion where Dias_atraso = 'Más de 120'
```

	(No column name)
1	1182

Figura 5.47. Consulta a la tabla operación sobre días de atraso más de 120 días

Nota: elaboración propia.

Al sumar el resultado de las consultas de las figuras Figura 5.44, Figura 5.45, Figura 5.46, Figura 5.47 y Figura 5.47 se corrobora que la cantidad de registros coincide con el total de operaciones almacenadas en los archivos de las carteras crediticias.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Adicionalmente, se llevaron a cabo las pruebas que se detallan en el Apéndice JJ: Ejecución de pruebas para comprobar la cantidad de registros procesados y en el Apéndice KK: Ejecución de pruebas para comprobar las transformaciones de campos numéricos a rangos de valores de este documento.

5.6. Verificación del proceso para creación de reportes post implementación de la propuesta

Una vez implementada la propuesta de solución del proyecto se procedió a verificar si el diseño planteado cumple con la reducción en el tiempo de procesamiento y la cantidad de iteraciones (tasa de reprocesamiento) que debe realizar el personal de la entidad financiera para generar sus reportes.

En el capítulo I de este documento se plantearon como beneficios directos de la solución los siguientes enunciados:

- ❏ Disminución en la cantidad de tiempo que invierte el personal en tareas de extracción, preparación y carga de información manual en al menos un 50% del tiempo actual.
- ❏ Reducir en un 80% la cantidad de reprocesamiento que deben realizar los analistas debido a cambios en los requerimientos gerenciales.

Para estimar concretamente el beneficio esperado de la solución se llevaron a cabo una serie de entrevistas al personal de las áreas de riesgo y crédito de la organización, los resultados obtenidos se detallan en el capítulo IV de este documento.

Durante las entrevistas se determinaron los tiempos promedio para la construcción de informes y se calculó la reducción meta en los indicadores derivados de los beneficios del proyecto.

La mejora esperada para los indicadores de tiempo promedio y tasa de reprocesamiento se expresa en la Tabla 5.5.

Tabla 5.5.

Beneficios esperados del proyecto al implementar la propuesta de solución

Departamento	Indicador	Situación previa organizacional	Mejora relativa esperada	Mejora absoluta esperada
Riesgo	Tiempo promedio para la construcción de informes	292 minutos	50%	146 minutos
	Tasa de reprocesamiento	cinco informes	80%	cinco informes
Crédito	Tiempo promedio para la construcción de informes	208,75 minutos	50%	104,375 minutos
	Tasa de reprocesamiento	diez informes	80%	dos informes

Nota: elaboración propia basada en los resultados en la Tabla 4.8 y en la Tabla 4.9 del capítulo de análisis de resultados.

Luego de establecer la mejora esperada para los indicadores del proceso de creación de reportes se procedió a corroborar su cumplimiento contra la ejecución de la solución planteada del proceso de ETL. Los resultados de la ejecución del proceso de ETL ejecutado se muestran en la Figura 5.48 y en la Figura 5.49.

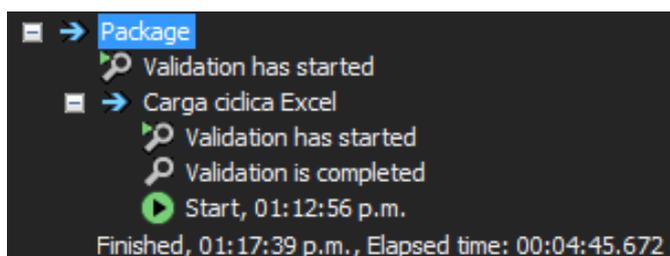


Figura 5.48. Tiempo dedicado a la extracción de los datos en los archivos fuente

Nota: elaboración propia.

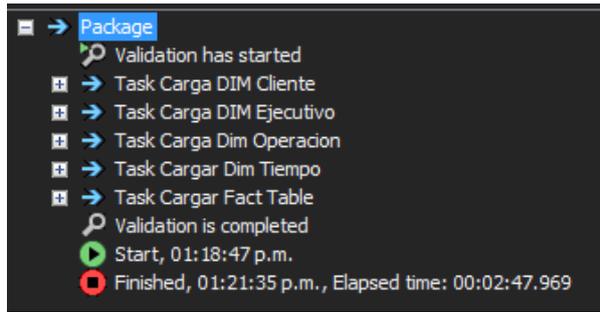


Figura 5.49. Tiempo dedicado al procesamiento y carga de datos al modelo multidimensional

Nota: elaboración propia.

Posterior a la implementación del proceso se determina que la extracción, transformación y carga de datos ejecutada por medio de la herramienta SSIS toma en total siete minutos y medio desde que inicia la extracción de los archivos fuente hasta que se realiza la carga al modelo multidimensional del DM.

En este caso se destaca una mejoría del tiempo de procesamiento de datos para los departamentos de Riesgo y Crédito quienes previamente dedicaban 203,66 y 171,25 minutos en promedio para procesar la información de sus reportes. El desglose de los tiempos que invertían los analistas en procesar la información se muestra en la Tabla 4.8 y en la Tabla 4.9 del capítulo de análisis de resultados.

Una vez finalizado el procesamiento de la información que se utiliza como insumo en la construcción de reportes se procede con la construcción del reporte en la herramienta Tableau.

Para la construcción del reporte se solicitó que los analistas de riesgos y crédito realizaran un reporte gráfico por medio de la herramienta Tableau para determinar la cantidad de tiempo que invierten en construir sus reportes gráficos una vez implementada la propuesta del proyecto. Los resultados obtenidos se muestran en la Tabla 5.6 y en la Tabla 5.7.

Tabla 5.6.

Tiempos para la construcción de reportes en el departamento de riesgo post implementación de la propuesta

Departamento de Riesgos después de integrar la solución a Tableau			
Actividad	Analista uno	Analista dos	Analista tres
Construcción del reporte visual	6 minutos	5 minutos	5 minutos
Total	6 minutos	5 minutos	5 minutos
Tiempo promedio	5,33 minutos		

Nota: prueba ejecutada con los analistas de riesgo de la entidad financiera

Tabla 5.7.

Tiempos para la construcción de reportes en el departamento de riesgo post implementación de propuesta

Departamento de Crédito después de integrar la solución a Tableau				
Actividad	Analista uno	Analista dos	Analista tres	Analista cuatro
Construcción del reporte visual	10 minutos	10 minutos	7 minutos	5 minutos
Total	10 minutos	10 minutos	7 minutos	5 minutos
Tiempo promedio	8 minutos			

Nota: prueba ejecutada con los analistas de crédito de la entidad financiera

Como se especifica en la Tabla 4.8 y en la Tabla 4.9 del capítulo de análisis de resultados antes de iniciar el desarrollo del proyecto el departamento de riesgos invertía 83,33 minutos en promedio para construir sus reportes tabulares y visuales, mientras que el área de crédito invertía 52,3 minutos en promedio en dichas tareas.

Al comparar los resultados obtenidos al implementar la propuesta de solución se denota una disminución en el tiempo que les toma a ambos departamentos elaborar sus informes de 78 minutos en promedio para el departamento de riesgos y de 46,3 minutos en promedio para el departamento de crédito.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

En cuanto a la tasa de reprocesamiento de reportes se redujo por completo, ya que el procesamiento de datos no se realiza manualmente y los usuarios finales únicamente tienen acceso a la herramienta de visualización de datos donde pueden ajustar sus reportes gráficos bajo los parámetros y atributos definidos en el DM.

6

Conclusiones

En este capítulo se establecen las conclusiones del proyecto desarrollado para solventar la situación problemática de la entidad financiera.

Como parte del desarrollo del proyecto se establecen las siguientes conclusiones para cada objetivo.

Objetivo general: Desarrollar el diseño de un modelo multidimensional y la automatización de los procesos de ETL en las áreas de riesgo y crédito de una entidad bancaria, mediante un análisis de las mejores prácticas de la industria, para la disminución de las brechas en tiempos de entrega de información a los departamentos comerciales de la entidad financiera.

- ▣ Producto del desarrollo del proyecto se establece que la propuesta de solución cumple a cabalidad con el objetivo del proyecto al brindar valor agregado a la organización en cuanto a la reducción de tiempo total de procesamiento de informes en más de un 50% del tiempo y en términos monetarios un ahorro de \$2,21 por cada dólar invertido como se evidencia en la sección 4.1.3 de este documento.

Objetivo 1: Diagnosticar las brechas en las tareas de extracción, transformación y carga en los procesos de inteligencia de negocio del área de crédito y riesgo con respecto a las mejores prácticas de la industria, para la identificación de las causas que imposibilitan la entrega de información oportuna.

- ▣ Según los resultados de las entrevistas especificadas en la sección 4.1.1.4 de este documento, la ejecución manual de las extracciones y procesamiento de datos consume 203,66 y 171,25 minutos en promedio a los departamentos de riesgo y crédito, indicando que uno de los factores que afectan los tiempos de entrega de los reportes crediticios es la ejecución de tareas manuales dedicadas a la extracción y procesamiento de datos.

- Se identificó una brecha con respecto a la práctica mencionada en la sección 2.9.2 de este documento donde se establece que únicamente se deben procesar los datos que se migrarán al *Data mart*, según el proceso para la construcción de informes detallado en la sección 4.1.1.4 del análisis de resultados, los analistas de la institución no aplican filtros sobre la información extraída de las fuentes transaccionales antes de iniciar las transformaciones de datos.

Producto de las conclusiones mencionadas se denota el cumplimiento del objetivo al identificar que las causas principales que afectan la entrega de informes son el procesamiento manual de información y la omisión del filtrado de datos durante la extracción de registros.

Objetivo 2: Recomendar tres alternativas de inversión acordes a las necesidades actuales de la entidad financiera en términos cualitativos y cuantitativos estableciendo a la gerencia opciones que permitan la implementación de la propuesta para la automatización del procesamiento de datos.

- Producto de los resultados del comparativo de herramientas, Pentaho Data Integration es la opción que cumple la mayoría de los criterios cualitativos para herramientas de integración con un 100% (ocho criterios de ocho) de cumplimiento, mientras que el software de visualización Tableau se establece como la alternativa óptima al cumplir con un 100% (ocho de ocho) de los criterios cualitativos. En este caso ROI obtenido al escoger estas herramientas para implementar la solución es de \$2,21. Cabe destacar que según lo indicado en el Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta
- Producto del comparativo de herramienta se establece que el segundo conjunto de herramientas que cumplen la mayoría de los criterios son Talend Data Integration con un 87,5% de cumplimiento (siete de ocho criterios) y Power BI que cumple en un 87,5% (siete de ocho características) los criterios. En este caso el ROI obtenido al implementar la solución con estas herramientas es de \$2,9.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ Al realizar el comparativo de herramientas se determinó que Oracle Data Integrator 12 c que cumple en un 87,5% (siete de ocho) de los criterios establecidos y mientras que Sisense cumple en un 75% (seis de ocho) los criterios. En este caso el ROI obtenido al implementar la solución con estas herramientas es de \$1,16.

Tomando en cuenta las conclusiones anteriores se determina el cumplimiento del objetivo del proyecto, al brindar a la gerencia de la entidad financiera tres alternativas que cumplen los requisitos de implementación de la propuesta en términos cualitativos y de inversión, resultado como alternativas óptimas Pentaho Data Integration y Tableau.

Objetivo 3: Diseñar el modelo multidimensional, dashboards de información crediticia y los procesos de ETL asociados a la preparación de datos que se utilizan como insumo en los análisis de las áreas de crédito y riesgo, para la entrega de información a las áreas comerciales.

- ❏ Se diseñó el modelo multidimensional que solventa los requerimientos de las áreas de negocio plasmados en la sección 4.1.1.2 del capítulo de análisis de resultados.
- ❏ Se elaboraron los diseños para tres reportes gráficos basados en la información del modelo multidimensional lo que facilitó la construcción de los informes crediticios.
- ❏ Se diseñó el proceso de Extracción, Transformación y Carga que se encarga de integrar la información de las carteras crediticias históricas de la entidad, reduciendo la intervención manual de los analistas en las tareas de procesamiento en un 100%.

Producto de las conclusiones mencionadas se establece el cumplimiento del tercer objetivo del proyecto al propiciarse la construcción del modelo multidimensional, proceso de ETL y visualizaciones de datos utilizando las herramientas designadas por la entidad financiera.

Objetivo 4: Construir una solución automatizada que facilite el procesamiento de datos en las tareas analíticas de Crédito y Riesgo para la disminución del reprocesamiento de informes en un 80% y la reducción del tiempo de elaboración en por lo menos un 50%.

- ❏ Al implementar la propuesta del proyecto se disminuyó el tiempo de procesamiento de datos del área de riesgo en 279 minutos con respecto al tiempo de construcción manual del reporte, mientras que para el área de crédito se logró disminuir el tiempo de procesamiento en 193,03 minutos con respecto al tiempo de elaboración manual.

- ❏ Por medio de la implementación de la propuesta se logró reducir al 100% la cantidad de iteraciones que realizaban los analistas en las tareas de procesamiento. Alcanzando el objetivo de disminuir en 80% el reprocesamiento de informes al brindar acceso a los usuarios únicamente sobre la capa de visualización de datos donde pueden elaborar sus propios informes sobre un conjunto de datos pre procesados.

Tomando en cuenta las conclusiones anteriores se evidencia el cumplimiento del cuarto objetivo del proyecto, al alcanzarse una disminución del tiempo para elaboración de informes en 95,6% para el área de riesgos y en 92,66% para el área de crédito. Así bien, se generó la disminución total de las iteraciones o reprocesamientos que ejecutaban los analistas durante el procesamiento de datos.

7

Recomendaciones

En este capítulo se establece un conjunto de acciones y pasos a seguir para complementar el desarrollo del proyecto en la entidad financiera. Adicionalmente, se incluyen algunas recomendaciones generales que se identificaron durante el desarrollo del proyecto.

Las recomendaciones del proyecto se plantean desde tres perspectivas distintas: técnica, de negocio y de información. A continuación, se detallan las recomendaciones para cada una de las perspectivas mencionadas.

Recomendaciones técnicas

- ✎ Incorporar la ejecución de la propuesta desarrollada en la ventana mensual de actualización que posee la entidad financiera, para los procesos que tramitan la información crediticia.
- ✎ Diseñar los roles y accesos a las bases de datos de producción a los usuarios finales que tendrán acceso a los datos del DM por medio de las herramientas de visualización de datos.

Recomendaciones de negocio

- ✎ Gestionar el acceso a las herramientas de visualización en cada área para garantizar la construcción y disponibilidad de la información para análisis.
- ✎ Establecer un procedimiento estándar para la construcción de reportes en las áreas de negocio de la entidad financiera.
- ✎ Producto del comparativo de alternativas existen opciones para herramientas de integración y visualización que son de interés para la entidad financiera tales como Power BI, Talend Data Integration y Sisense. Se recomienda retomar la discusión de estas soluciones en los foros de adquisición anual de licencias.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

- ❏ Realizar el mapeo y construcción de reportes adicionales que se podrían incluir en la solución propuesta.

Recomendaciones de Información

- ❏ Realizar la gestión de limpieza de datos y actualización de los datos de los clientes con el fin de mantener información actualizada a nivel de las carteras crediticias y la solución planteada.
- ❏ Mantener la estructura de los archivos de cartera crediticia según lo establecido en el capítulo V de este documento.
- ❏ Para asegurar la creación consistente de los reportes que realizan las áreas de Riesgo y Crédito de la entidad es necesario reforzar la estandarización y controlar las fuentes de ingreso de datos que alimentan las fuentes de datos analíticas.
- ❏ Construir una bitácora de los errores presentados en la información para determinar las causas y aplicar las acciones correctivas correspondientes.



Apéndices

En esta sección se adjuntan los elementos complementarios al trabajo de investigación, tales como plantillas, minutas y herramientas aplicadas para la recolección de datos de análisis.

Apéndice A: Tabla para Benchmarking comparativo de herramientas

Criterios analizados	Alternativa 1	Alternativa 2	Alternativa ...	Alternativa n
Criterio 1				
Criterio ...				
Criterio n				

Apéndice B: Guía de entrevista para identificación de necesidades

Datos generales

Puesto del entrevistado	
Rol del entrevistado en el proyecto	
Fecha de aplicación	

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento?
Mencione el impacto que tiene esta labor para la organización.
- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?
- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Apéndice C: Plantilla para recolección de requisitos

N° de requerimiento	Consecutivo del requisito
Versión del documento	Versión del documento
Departamento	Departamento involucrado en la recolección de requisitos.
Rol en el proyecto	Rol del entrevistado (a) en el proyecto.
Enfoque requerimiento	Área específica del requerimiento.
Descripción del requerimiento	
[Descripción breve del requerimiento que propone el entrevistado (a)]	
Precondiciones del requerimiento	
[Condiciones previas para el desarrollo del requerimiento]	
Detalle del requerimiento	
[Detalle del requerimiento que toma en cuenta los insumos y resultados esperados]	
Restricciones a tomar en cuenta	
[Restricciones a tomar en cuenta durante el desarrollo del requisito]	

Apéndice D: Guía de entrevista para mapeo de fuentes transaccionales

Datos generales

Puesto del entrevistado	
Rol del entrevistado en el proyecto	
Fecha de aplicación	

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

Preguntas

- ¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?
- ¿En qué formato se encuentra la información que utiliza para construir sus reportes?
- ¿Cuál son las fuentes primarias de información para la construcción de sus informes?
- ¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?
- ¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?
- ¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Apéndice E: Guía de entrevista para mapeo de los procesos de elaboración de reportes

Datos generales

Puesto del entrevistado	
Rol del entrevistado en el proyecto	
Fecha de aplicación	

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

Preguntas

- ¿Qué actividades lleva a cabo para construir sus reportes?
- ¿A quién van dirigidos los reportes que genera actualmente?
- ¿Cuál es la periodicidad con la que elabora estos reportes?
- ¿En promedio cuánto tiempo invierte en elaborar cada informe?
- ¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?
- ¿De qué forma gestiona los informes de índole histórica?

Apéndice F: Guía de entrevista para selección de criterios del *Benchmarking*

Datos generales

Puesto del entrevistado	
Rol del entrevistado en el proyecto	
Fecha de aplicación	

El objetivo de esta entrevista es indagar al personal técnico y de negocio de la organización sobre las características deseables en las herramientas de integración y visualización de datos.

Preguntas

- ¿Existen algunas herramientas que requiera incluir en el análisis comparativo?
- ¿Cuáles son las características básicas deseables de las soluciones de integración y visualización de datos?
- ¿Con cuáles herramientas he trabajado la organización anteriormente en cuanto a integración y visualización de datos?
- ¿Cuál es su expectativa con respecto a las soluciones de visualización e integración de datos?
- ¿El personal cuenta con experiencia utilizando herramientas de integración o visualización de datos?

Apéndice G: Plantilla de paquetes de información

Dimensiones			
Nombre de dimensión	Nombre de dimensión	Nombre de dimensión	Nombre de dimensión
Atributo dimensional 1	Atributo dimensional 1	Atributo dimensional 1	Atributo dimensional 1
Atributo dimensional 2	Atributo dimensional 2	Atributo dimensional 2	Atributo dimensional 2
...
Atributo dimensional n	Atributo dimensional n	Atributo dimensional n	Atributo dimensional n
Hechos			
Hecho o medida 1			
Hecho o medida 2			
Hecho o medida ...			
Hecho o medida n			

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta

Reunión No.	05	Fecha:	8/10/2018	
Lugar:	Oficinas centrales de la entidad financiera	Hora Inicio:	2:00 p.m	
		Hora Finalización	3:00 p.m	
Objetivo de la reunión:	Documentar acuerdo sobre las herramientas que se utilizarán para el desarrollo de la propuesta del proyecto.			
Participantes:	Presentes: Fernando Mata Orozco Lorena Castillo García Jefatura de sistema de la organización			
	Ausentes: -			
Temas Tratados				
No.	Asunto	Comentarios	Acuerdos	Responsable
I	Documentar el acuerdo sobre las herramientas de integración y visualización que se utilizarán durante el desarrollo de la propuesta del proyecto.	Se establece que debido al acuerdo de licenciamiento actual que posee la entidad financiera se hará uso de las herramientas SQL Server Integration Services y Tableau para implementar la propuesta de solución.	Se acuerda utilizar las herramientas del proveedor Microsoft y Tableau para ejecutar la implementación del proyecto.	Fernando Mata
Próxima reunión				
Temas a tratar		Fecha	Convocados	
Por definir		Por definir	Por definir	
Firmas de aprobación				
Nombre	Rol	Firma		
Lorena Castillo García	Contraparte organizacional			

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Apéndice J: Aprobación del modelo de datos para el Data mart

Reunión No.	03	Fecha:	4/9/2018	
Lugar:	Oficinas centrales de la entidad financiera	Hora Inicio:	2:00 p.m	
		Hora Finalización	3:00 p.m	
Objetivo de la reunión:	Aprobar el modelo de datos que se utilizará para la propuesta del proyecto.			
Participantes:	Presentes: Fernando Mata Orozco Lorena Castillo García			
	Ausentes: -			
Temas Tratados				
No.	Asunto	Comentarios	Acuerdos	Responsable
1	Aprobación del modelo de datos para la propuesta del proyecto.	Se presenta el modelo del Data mart que integra la información de la cartera crediticia para la aprobación correspondiente por parte de la entidad financiera.	Se acuerda que el modelo de datos debe incluir toda la información solicitada por los analistas durante la recolección de requerimientos.	Fernando Mata
Próxima reunión				
Temas a tratar	Fecha	Convocados		
Aprobación de visualizaciones de datos para la propuesta		Lorena Castillo García Fernando Mata Orozco		
Firmas de aprobación				
Nombre	Rol	Firma		
Lorena Castillo García	Contraparte organizacional			

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Apéndice K: Aprobación de visualizaciones

Reunión No.	04	Fecha:	1/10/2018	
Lugar:	Oficinas centrales de la entidad financiera	Hora Inicio:	9:00 a.m.	
		Hora Finalización	9:15 a.m.	
Objetivo de la reunión:	Aprobar las visualizaciones que se incorporarán en la propuesta del proyecto.			
Participantes:	Presentes: Fernando Mata Orozco			
	Lorena Castillo Garcia			
	Ausentes: -			
Temas Tratados				
No.	Asunto	Comentarios	Acuerdos	Responsable
I	Aprobación de las visualizaciones que se incorporarán en la propuesta del proyecto.	Se presenta la propuesta de los tipos de visualización enfocados en representar el comportamiento histórico de la cartera crediticia. Los tipos de visualización presentados incluyen gráficos lineales y de barras.	Se acuerda que los tipos de visualización seleccionados propician la correcta representación de los datos almacenados en las carteras crediticias.	Fernando Mata
Próxima reunión				
Temas a tratar		Fecha	Convocados	
Por definir		Por definir	Por definir	
Firmas de aprobación				
Nombre		Rol	Firma	
Lorena Castillo García		Contraparte organizacional		

Apéndice L: Composición de las tablas utilitarias para proceso de ETL

Las tablas intermedias para el manejo del respaldo durante el procesamiento del ETL poseen los siguientes atributos y su objetivo es proveer parámetros para campos calculados o bien almacenar información temporalmente.

Tabla de operaciones

📦 Cod_Ejecutivo	📦 Mto_cta_cobrar_colonizado
📦 Ejecutivo	📦 Mto_total_deuda_colonizado
📦 Cod_Ejecutivo	📦 Capacidad_pago
📦 Cliente	📦 CPH
📦 Id_cliente	📦 Categoria_riesgo
📦 Num_operacion	📦 Dias_atraso
📦 Descripcion_producto	📦 Mto_mitigador
📦 Descripcion_subproducto	📦 Mto_estimacion_generica
📦 Estado	📦 Estimacion_especifica_saldo_cubier to
📦 Monto_principal_directo	📦 Estimacion_especifica_saldo_descu bierto
📦 Monto_principal_contingente	📦 Generador_divisa
📦 Mto_productosxCobrar	📦 Actividad
📦 Moneda_operacion	📦 Año
📦 Monto_principal_directo	📦 Mes
📦 Monto_principal_contingente	
📦 Mto_producto_colonizado	

Tabla de tipo de cambio

📦 Mes
📦 Año
📦 TipoCambioCompra

Apéndice M: Entrevista sobre necesidades con analista de riesgo crediticio

Datos generales

Puesto del entrevistado	Analista de riesgo crediticio
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	20/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- **¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.**

Analizar el comportamiento de los clientes crediticios desde la perspectiva de riesgo, para identificar su comportamiento a lo largo de la relación que mantienen con la entidad.

- **¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?**

El procesamiento manual del reporte en archivos de Excel que en ocasiones ha propiciado errores durante los cálculos de indicadores.

- **¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?**

Eliminar la intervención manual en las tareas de procesamiento de datos.

Apéndice N: Entrevista sobre necesidades con analista de riesgo cambiario y liquidez

Datos generales

Puesto del entrevistado	Analista de riesgo cambiario y liquidez
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	20/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.

Nuestros reportes se realizan para brindar a las gerencias de la organización información que les permita conocer el estado de las operaciones y clientes de la institución desde distintas perspectivas, en mi caso con respecto al tema de liquidez y tipo de cambio.

- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?

La cantidad de tiempo que se tarda en construir los reportes, debido a la construcción manual y verificación posterior del informe que se desarrolló.

- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Utilizaría datos ya procesados de la cartera de clientes para elaborar los reportes, así reduciría la cantidad de intervenciones manuales sobre la información.

Apéndice Ñ: Entrevista sobre necesidades con el auxiliar de riesgos

Datos generales

Puesto del entrevistado	Departamento de riesgo
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	21/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.

Los reportes de riesgos permiten a la organización comprender los movimientos y el comportamiento de la cartera de clientes, proveedores, socios, accionistas y codeudores desde una perspectiva de exposición.

- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?

La cantidad de tareas manuales que se llevan a cabo con el tema de los reportes consume bastante tiempo que se podría dedicar al análisis de datos.

- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Evitaría que el personal del departamento intervenga en el procesamiento de datos y en su lugar se enfoque en analizar la información.

Apéndice O: Entrevista sobre necesidades con analista operativo de crédito

Datos generales

Puesto del entrevistado	Analista operativo de crédito
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	21/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.

Los reportes del departamento buscan estudiar el estado de la cartera crediticia en pro de identificar oportunidades de negocio que permitan maximizar las ganancias de la entidad.

- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?

En ocasiones la disponibilidad de los recursos con mayor experiencia del departamento impacta en la construcción y disponibilidad de los reportes.

- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Crearía un repositorio de información que permita construir los reportes sin depender de otras personas del departamento.

Apéndice P: Entrevista sobre necesidades con analista de mora

Datos generales

Puesto del entrevistado	Analista de mora
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	21/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.

Los reportes del departamento de crédito permiten estudiar la cartera de clientes y operaciones para determinar nuevas oportunidades de negocio o indicios de que una relación comercial ya no es prolífica.

- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?

La dependencia con los analistas más experimentados de la organización para la construcción de los informes de cartera.

- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Pondría a disposición del departamento los datos ya procesados de la cartera crediticia y capacitaría a los analistas en la construcción de los reportes principales o más importantes.

Apéndice Q: Entrevista sobre necesidades con analista de mercado

Datos generales

Puesto del entrevistado	Analista de mercado
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	21/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.

Los reportes del departamento permiten estudiar nuevos negocios y su factibilidad con respecto al estado actual de la organización.

- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?

Se han estado manejando muchos archivos independientes con la información crediticia y esto dificulta el procesamiento de información.

- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Unificaría los datos en una sola fuente que sea de acceso universal para el departamento.

Apéndice R: Entrevista sobre necesidades con analista financiero

Datos generales

Puesto del entrevistado	Analista financiero
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	21/8/2018

El objetivo de esta entrevista es determinar los requerimientos implícitos en las necesidades de los analistas de la organización donde se desarrolló el TFG.

Preguntas

- ¿Cuál es la finalidad que persiguen los reportes que construye su departamento? Mencione el impacto que tiene esta labor para la organización.

Los reportes del área crediticia permiten a la organización estudiar nuevas vertientes de colocación de crédito con clientes existentes o potenciales.

- ¿Cuáles son los problemas principales que ha identificado durante la creación de los reportes en su departamento?

Se manejan muchos archivos dispersos de datos y resulta tedioso el manejo de la información histórica.

- ¿De qué manera solventaría los problemas que se presentan actualmente durante la creación de los reportes?

Unificaría las fuentes de datos y eliminaría el procesamiento recurrente de datos. La información debería procesarse una vez únicamente.

Apéndice S: Recolección de requisitos con el área de riesgos

La recolección de requisitos se realizó de manera conjunta para el auxiliar de riesgos, analista de riesgo crediticio y el analista de riesgo de liquidez y riesgo cambiario. Los resultados se presentan en las siguientes plantillas.

Requerimientos del repositorio multidimensional en el área de riesgos

Fecha de aplicación: 23/08/2018	
N° de requerimiento	01
Versión del documento	01
Departamento	Área de Riesgos
Rol en el proyecto	Consultados
Enfoque del requerimiento	Estructura del repositorio multidimensional

Descripción del requerimiento

Consolidar la información de las carteras crediticias en una base de datos única que permita la construcción de informes basados en los datos de clientes y operaciones de la entidad.

Precondiciones del requerimiento

Verificar que los archivos que se incluirán en la solución cuenten con toda la información solicitada.

Detalle del requerimiento

Incluir en la solución los siguientes datos:

- Descripción del producto y subproducto.
- Días de atraso.
- Estado de la operación.
- Monto de estimación genérica.
- Monto de estimación específica saldo con garantía.
- Monto de estimación específica saldo sin garantía.
- Monto total de la deuda colonizado y dolarizado.
- Moneda de la operación.
- Nombre del cliente.
- Identificación.
- Categoría de Riesgo.
- Actividad.
- Indicador para generadores de divisa.

Restricciones a tomar en cuenta

En los archivos originales los montos de las operaciones no se encuentran colonizados ni dolarizados, en su lugar se especifican el importe de la operación y una columna que indica la moneda en la que se registró la operación.

Requerimientos para el procesamiento de datos en el área de riesgos

Fecha de aplicación: 23/08/2018

N° de requerimiento 02

Versión del documento 01

Departamento Área de Riesgos

Rol en el proyecto Consultados

Enfoque del requerimiento Procesamiento de datos

Descripción del requerimiento

Eliminar el procesamiento manual requerido para extraer, modificar y estructurar los datos que se utilizan en los reportes de concentración de la cartera crediticia.

Precondiciones del requerimiento

Se debe contar con la información correspondiente a la cartera crediticia de los meses bajo análisis.

Detalle del requerimiento

Se debe dolarizar el dato del monto total de la deuda colonizado, utilizando el tipo de cambio de referencia brindado por el Banco Central de Costa Rica al cierre del mes que se está procesando.

Unificar la información de las columnas de categoría de Riesgo que son alfanuméricas para que incluyan letras únicamente.

Convertir la columna de días de atraso a un rango de valores de 0 a 30, 31 a 60, 61 a 90, 91 a 120 y más de 120 días.

Convertir la columna del indicador para generadores de divisa a texto bajo los siguientes criterios, el valor 1 se convierte a Generador, el 2 a No generador de bajo riesgo, el 3 a No generador de alto riesgo y el 4 a No aplica.

Restricciones a tomar en cuenta

La información de insumo debe procesarse al finalizar el cierre operativo mensual.

Requerimientos para las visualizaciones de datos en riesgos

Fecha de aplicación: 23/08/2018

N° de requerimiento 03

Versión del documento 01

Departamento Área de Riesgos

Rol en el proyecto Consultados

Enfoque del requerimiento Visualizaciones de datos

Descripción del requerimiento

Visualizar la información histórica de la cartera crediticia para determinar comportamientos anómalos en las operaciones de crédito o los clientes de la entidad financiera.

Precondiciones del requerimiento

Contar con la base de datos integrada de operaciones históricas de la cartera de crédito que contemple el procesamiento requerido por el departamento.

Detalle del requerimiento

Se busca construir reportes gráficos que permitan observar la tendencia histórica de la cartera crediticia de la organización.

Adicionalmente, se debe permitir que los gráficos sean interactivos al ofrecer la posibilidad de navegar al detalle de la información que se presenta en las gráficas del reporte.

Restricciones a tomar en cuenta

Se debe evitar el uso de gráficos en tres dimensiones y de tipo pastel, ya que generalmente no permiten la interpretación correcta de los datos.

Apéndice T: Recolección de requisitos con el área de crédito

La recolección de requisitos se realizó de manera conjunta para el analista operativo de crédito, analista de mercado, analista de mora y el analista financiero. Los resultados se presentan en las siguientes plantillas.

Requerimientos para el modelo multidimensional en crédito

Fecha de aplicación 24/08/2018	
N° de requerimiento	01
Versión del documento	01
Departamento	Área de Crédito
Rol en el proyecto	Consultados
Enfoque del requerimiento	Estructura del repositorio multidimensional

Descripción del requerimiento

Incorporar los datos de las carteras de crédito en una herramienta única que permita centralizar el acceso a los datos para el área de Crédito, evitando a su vez el uso de archivos independientes para la construcción de reportes en el área.

Precondiciones del requerimiento

Se debe tener en cuenta la extracción de los archivos utilizados para construir los reportes debe ejecutarse posterior al cierre mensual de operaciones de crédito.

Detalle del requerimiento

Se debe incluir en la solución las siguientes columnas del archivo de cartera:

- Monto total de la deuda dolarizado y colonizado.
- Monto mitigador.
- Monto principal directo y contingente por moneda.
- Monto producto colonizado.
- Monto de cuenta.
- Nombre del ejecutivo.
- Identificación.
- Nombre del cliente.
- Actividad.
- Descripción del producto.
- Estado de la operación.
- Días de atraso.
- Moneda.

Restricciones a tomar en cuenta

Existen operaciones que no tienen un ejecutivo asignado en las carteras crediticias.

Requerimientos para el procesamiento de datos

Fecha de aplicación	24/08/2018
N° de requerimiento	02
Versión del documento	01
Departamento	Área de Crédito
Rol en el proyecto	Consultados
Enfoque del requerimiento	Procesamiento de datos

Descripción del requerimiento

Se busca una alternativa para la construcción de reportes que permita analizar tendencias históricas en el comportamiento de los clientes por producto, días de atraso, estado, actividad y moneda de la operación.

Precondiciones del requerimiento

Se requiere que previo a la construcción del reporte se organicen las mantas o archivos de datos crediticios de la entidad financiera.

Detalle del requerimiento

Se debe dolarizar el monto total de la deuda utilizando el tipo de cambio de referencia de compra del Banco Central reportado al cierre de mes.

Se busca segmentar los valores correspondientes a días de atraso para determinar cuántas operaciones se encuentran en los rangos de 0-30, 31-60 días y 61-90 días de atraso.

Restricciones a tomar en cuenta

La información que se utiliza como insumo debe procesarse siempre después del cierre operativo mensual.

Requerimientos para las visualizaciones de datos

Fecha de aplicación 24/08/2018	
N° de requerimiento	03
Versión del documento	01
Departamento	Área de Crédito
Rol en el proyecto	Consultados
Enfoque del requerimiento	Visualizaciones de datos

Descripción del requerimiento

Visualizar la información correspondiente a la cartera de crédito de forma clara, adicionalmente se busca determinar tendencias en los datos que habiliten la toma de decisiones certera en el departamento.

Precondiciones del requerimiento

Información procesada según las reglas de negocio establecidas anteriormente en la entrevista.

Detalle del requerimiento

Se requiere que las visualizaciones permitan diferenciar claramente la evolución interanual de la cartera de crédito.

Se busca que las visualizaciones de datos mantengan una paleta de colores oscura para los fondos y de colores cálidos para la información analizada, de esta forma se pueden resaltar adecuadamente los datos principales del reporte.

Restricciones a tomar en cuenta

No registran restricciones para este requerimiento.

Apéndice V: Entrevista sobre fuentes transaccionales con analista de mora

Datos generales

Puesto del entrevistado	Analista de mora
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	28/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Generalmente los reportes se construyen utilizando el archivo de la cartera crediticia del Banco.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

La cartera crediticia se exporta en archivos de Excel desde SAP.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

En mi caso sería la cartera de crédito que tiene la información de clientes y operaciones de crédito para construir los informes de mora.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

La extracción tarda alrededor de unos 30 minutos, pero siempre hay que revisar que la información concuerde con los cierres de conta.

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

En mi caso debo agrupar las operaciones por rangos de días de atraso, por medio de una formula en Excel creo los rangos y construyó el reporte.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

La cartera se actualiza diariamente, pero para tener claridad de las operaciones que se encuentra en mora y sus implicaciones lo ideal es realizar los reportes después del cierre de mes. Aunque siempre se realizan reportes diarios para ver el movimiento de la mora.

Apéndice W: Entrevista sobre fuentes transaccionales con analista de mercado

Datos generales

Puesto del entrevistado	Analista de mercado
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	28/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Para los reportes de cartera utilizó el archivo transaccional de SAP.

Cuando se requieren reportes del sector financiero también utilizó la información disponible en el sistema que monitorea el comportamiento de las carteras de clientes en las instituciones consideradas como pares de la organización a nivel nacional.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

La información interna siempre se ha trabajado en Excel y la externa por lo general viene en formato XML.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

La fuente principal es la cartera crediticia de clientes y el archivo del Centro de Información Crediticia.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

Alrededor de 30 minutos porque son bastantes transacciones que extrae el sistema.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

Tenemos indicadores numéricos que no intuitivos para alguien fuera de la operativa, es nuestro trabajo convertir esos datos a valores comprensibles por las gerencias y foros de la entidad.

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Los datos que utilizó se actualizan mensualmente en el caso de la cartera crediticia y cada dos meses en el caso de los archivos CIC.

Apéndice X: Entrevista sobre fuentes transaccionales con analista financiero

Datos generales

Puesto del entrevistado	Analista financiero
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	28/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Los reportes financieros requieren datos internos y externos debido a que se requiere tener un panorama claro de la situación interna de la entidad y en que posición nos encontramos con respecto a nuestros pares.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

Generalmente la información se encuentra en archivos de Excel o archivos planos en menor medida, son casos esporádicos donde se debe trabajar con archivos de texto.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

Principalmente se utiliza la cartera crediticia que se genera posterior al cierre mensual y la información del CIC, JANFOR también proporciona información complementaria.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

Media hora aproximadamente por la cantidad de operaciones que maneja la organización.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

Generalmente lo que se identifica son valores nulos, en ese caso se notifica al encargado de su corrección en la parte operativa para que verifique el caso y realice las correcciones respectivas,

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Los datos que se utilizan para construir los reportes del área se actualizan de forma mensual y en casos del Centro de Información Crediticia es cada dos meses.

Apéndice U: Entrevista sobre fuentes transaccionales con analista operativo de crédito

Datos generales

Puesto del entrevistado	Analista operativo de crédito
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	28/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Los reportes que se construyen en el área utilizan las operaciones de la cartera de crédito que tiene datos de saldos, estimación, días de atraso y clientes.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

En mi caso trabajo en su totalidad con archivos de Excel.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

En la perspectiva operativa que me corresponde utilizo la cartera de crédito más que todo.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

La extracción de datos tarda de 25 minutos en adelante, esto depende de la cantidad de datos que se requiera extraer de SAP.

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

La corrección de valores faltantes es lo que se realiza generalmente fuera de eso no realizo nada más para adecuar la información.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Los datos se actualizan diariamente por el movimiento de saldos, cancelación de operaciones y aperturas de líneas, sin embargo los reportes siempre es recomendable realizarlos al finalizar el cierre de mes.

Apéndice Y: Entrevista sobre fuentes transaccionales con analista de riesgo crediticio

Datos generales

Puesto del entrevistado	Analista de riesgo crediticio
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	29/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Los datos se extraen de la cartera mensual de crédito y del CIC, aunque en ocasiones los archivos de cartera incorporan los datos del CIC.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

La información se gestiona por lo general en Excel o archivos de texto plano para los datos internos. La información de fuentes externas generalmente se maneja en formato XML o texto plano, pero igual se procesan por medio de Excel.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

Se procesa la información que se reporta a los entes reguladores a final de mes.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

En el caso de los archivos de la cartera crediticia se solicitan al área operativa correspondiente una vez realizado el cierre mensual.

En el caso de los datos del archivo CIC para la totalidad de la cartera crediticia de la organización el tiempo máximo de extracción ha sido de 12 horas.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

Comúnmente se dolarizan saldos, montos de operación aprobados, se crean rangos para identificar de forma sencilla clientes generadores y no generadores de moneda extranjera y se agrupan operaciones que no poseen un ejecutivo bajo una clasificación separada que permite discernir los casos específicos en análisis posteriores.

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Los datos internos se actualizan mensualmente, una vez que el departamento operativo de crédito finaliza la construcción del informe de cartera, que se distribuye por lo general una semana después de finalizar el cierre operativo.

Apéndice Z: Entrevista sobre fuentes transaccionales con analista de riesgo cambiario y liquidez

Datos generales

Puesto del entrevistado	Analista de riesgo cambiario y liquidez
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	29/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Siempre se ha utilizado la cartera de crédito donde están las operaciones de crédito de la organización.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

Los datos por lo general vienen en archivos de Excel, es el formato en que hemos trabajado siempre.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

La cartera de operaciones crediticias es lo que se utiliza generalmente.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

Los archivos de cartera se solicitan al área operativa y ellos extraen la información según su disponibilidad, si ya lo tienen extraído tardan un par de minutos, en caso contrario es alrededor de una hora.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

La información de la cartera crediticia si ha presentado errores de clasificación en el pasado y otros errores en saldos de operación, esto se reporta para ser corregido.

En mi caso tengo que transformar el dato de generadores de divisa porque en el archivo de cartera solo viene un código numérico.

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Mensual para tener los datos consistentes de la cartera.

Apéndice AA: Entrevista sobre fuentes transaccionales con auxiliar de riesgo

Datos generales

Puesto del entrevistado	Auxiliar de riesgo
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	29/8/2018

El objetivo de esta entrevista es conocer las fuentes primarias que utilizan los analistas de la entidad financiera durante el proceso de elaboración de informes.

¿La información que utiliza para construir los informes proviene de fuentes internas o externas de datos?

Debido a mis funciones debo manejar información interna y externa principalmente los archivos de la cartera crediticias y los descargables del CIC.

¿En qué formato se encuentra la información que utiliza para construir sus reportes?

La información proviene principalmente en formato XML o Excel.

¿Cuáles son las fuentes primarias de información para la construcción de sus informes?

La cartera de operaciones crediticias y el comportamiento de pago histórico de los clientes.

¿Cuánto tiempo invierte en extraer la información que requiere de las fuentes de datos?

No realizo la extracción, en este caso se solicita al área de crédito que lo realice y ellos tardan alrededor de 30 – 40 minutos.

¿La información que extrae de los sistemas fuente requiere algún tipo de procesamiento para crear el informe final? De ser así ¿Cuáles transformaciones aplica sobre los datos de la fuente?

Lo que se hace es consolidar las categorías de riesgo en una sola, crear rangos para los días de atraso y cambiar los valores para generadores y no generadores de divisa a texto para que se puedan comprender en un gráfico.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Cuál es la frecuencia de actualización de los datos que procesa actualmente?

Mensual en el caso de la cartera y bimensual en el caso del CIC.

Apéndice BB: Entrevista para mapeo de procesos para elaboración de reportes analista de riesgo crediticio

Datos generales

Puesto del entrevistado	Analista de riesgo crediticio
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	30/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Solicito el archivo de cartera
- Reviso la información los datos
- Notifico errores en valores faltantes
- Aplico los cambios en la información
- Construyo el reporte

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes van destinados a Junta Directiva, foros de discusión internos y gerencia.

¿Cuál es la periodicidad con la que elabora estos reportes?

La mayoría de los reportes son mensuales por los procesos de cierre.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

En la solicitud del archivo un par de minutos, en comprobación de datos media hora, en depuración alrededor de dos horas debido a que la información debe cerrar perfectamente con respecto a conta, en los cambios sobre los datos de dos horas a tres, construyendo el informe una hora y el gráfico unos 10 minutos.

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Reducir las tareas manuales en procesamiento de datos, ya que incrementa la probabilidad de cometer errores.

¿De qué forma gestiona los informes de índole histórica?

En el caso de los reportes sobre datos históricos algunos analistas mencionan que representan una complejidad mayor al tener que procesar grandes cantidades de información en periodos prolongados de tiempo lo que previamente ha provocado inconsistencias en los informes.

Con datos históricos tenemos una complejidad mayor ya que se deben procesar bastantes archivos con una cantidad de filas considerable que por lo general excede las 8000.

Apéndice CC: Entrevista para mapeo de procesos para elaboración de reportes analista de riesgo de liquidez y riesgo cambiario

Datos generales

Puesto del entrevistado	Analista de riesgo de liquidez y riesgo cambiario
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	30/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Solicitud de la fuente de datos
- Comprobación de los datos contra reporte de conta
- Reporto errores en los datos
- Aplicó la transformación de los datos
- Diseño y construyo mi reporte

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes construidos en riesgo se dirigen a Junta por regulación y se comunican a las gerencias para que los tengan en cuenta.

¿Cuál es la periodicidad con la que elabora estos reportes?

Mensual, pero por lo general se llevan a cabo reportes bajo demanda a lo largo del mes.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

Solicitud de la fuente de datos unos dos minutos, la comprobación contra la conta unos 30-45 minutos, el reporte y depuración de errores 30 minutos, las transformaciones hora y veinte u hora y media, la construcción del reporte hora y media y el gráfico unos 15 minutos.

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Eliminar la solicitud de datos a otras áreas de negocio, si se pudieran extraer los datos sin tener que solicitarlos sería muy provechoso.

¿De qué forma gestiona los informes de índole histórica?

Manejo unos archivos formulados en Excel donde sustituyo la base de datos que estoy analizando y mis formulas se encargan de construir los indicadores que necesito. Esto es más práctico que procesar los datos desde cero.

Apéndice DD: Entrevista para mapeo de procesos para elaboración de reportes auxiliar de riesgo

Datos generales

Puesto del entrevistado	Auxiliar de riesgo
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	30/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Solicitud de la fuente de datos
- Compruebo los datos
- Reporto los errores y depuro mi base
- Creo los rangos de valores requeridos datos
- Construyo mi reporte

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes se envían a junta y gerencias principalmente.

¿Cuál es la periodicidad con la que elabora estos reportes?

La periodicidad es mensual, sin embargo, hemos realizado reportes semanales e incluso diarios bajo demanda.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

La solicitud de los datos tarda unos dos minutos, la comprobación uno veinte minutos, la depuración otros veinte, la creación de rangos una hora quince u hora y veinte minutos, la construcción del informe una hora y media y el gráfico unos 10 minutos.

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Eliminar los pasos manuales para construir el informe ya que consumen demasiado tiempo en tareas operativas.

¿De qué forma gestiona los informes de índole histórica?

Generalmente utilizo archivos formulados en Excel para agilizar el procesamiento, sin embargo, siempre se tiene que intervenir aunque sea un poco cuando se presentan excepciones en la información que no se pueden controlar como la clasificación incorrecta de operaciones.

Apéndice EE: Entrevista para mapeo de procesos para elaboración de reportes – analista operativo de crédito

Datos generales

Puesto del entrevistado	Analista operativo de crédito
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	31/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Extracción de datos
- Depuración de valores
- Transformaciones en la información fuente
- Creo una manta procesada en Excel
- Creo el reporte visual

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes que se generan van dirigidos por lo general a las jefaturas y gerentes de la entidad.

¿Cuál es la periodicidad con la que elabora estos reportes?

Mensual debido a los procesos de cierre y a que es posterior al cierre mensual cuando se tiene certeza sobre el estado de la cartera.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

En la extracción se tarda media hora, depurando otra media hora, transformando la información una hora, creando el reporte en Excel media hora y el gráfico 10 minutos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Integrar la información de las fuentes de datos para evitar el manejo de múltiples archivos de Excel.

¿De qué forma gestiona los informes de índole histórica?

Mantenemos un archivo consolidado con la información procesada anteriormente, solo esos construimos los reportes.

Apéndice FF: Entrevista para mapeo de procesos para elaboración de reportes analista de mora

Datos generales

Puesto del entrevistado	Analista de mora
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	31/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Extraigo los datos de SAP
- Depuro la información
- Aplico las transformaciones
- Hago el reporte tabular
- Construyo el reporte visual

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes que se generan de mora van para diversos comités crediticios, a Junta Directiva y gerentes de la entidad.

¿Cuál es la periodicidad con la que elabora estos reportes?

Diarios y otros mensuales según los requieran los foros de la organización.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

Extraer los datos de SAP 30 minutos, depurar la información una hora, aplicar las transformaciones dos horas o dos horas y media, creando el reporte tardo unos 20 minutos y el gráfico unos 25 minutos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Unificar la información principalmente, el manejo de muchos Exceles se vuelve complejo y lo hace propenso a errores.

¿De qué forma gestiona los informes de índole histórica?

Hay un archivo histórico que se actualiza cada cierre, cuando se requiere agregar información del nuevo periodo se construye el reporte mensual y se integra al consolidado.

Apéndice GG: Entrevista para mapeo de procesos para elaboración de reportes analista de mercado

Datos generales

Puesto del entrevistado	Analista de mercado
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	31/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Extraer de datos
- Depuro los datos
- Aplico las transformaciones
- Creo el reporte en tablas de Excel
- Creo el gráfico

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes que se generan van para Comité crediticio, Gerencia y a las jefaturas.

¿Cuál es la periodicidad con la que elabora estos reportes?

Mensuales ya que si se ejecuta el análisis previo al cierre operativo la información podría no reflejar correctamente la situación actual de la entidad financiera.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

Extrayendo los datos tarda unos 30 minutos, la corrección de datos unos 25 minutos, las transformaciones hora y media, la creación del reporte en Excel una hora y el gráfico unos 15 minutos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Eliminar el procesamiento en Excel ya que estamos muy propensos a cometer errores en el procesamiento.

¿De qué forma gestiona los informes de índole histórica?

Utilizó el archivo consolidado del área y sobre eso construyo los informes internos.

Apéndice HH: Entrevista para mapeo de procesos para elaboración de reportes analista financiero

Datos generales

Puesto del entrevistado	Analista financiero
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	31/8/2018

El objetivo de esta entrevista es conocer el conjunto de informes que realizan los analistas de Riesgo y Crédito de la entidad financiera donde se lleva a cabo el TFG.

¿Qué actividades lleva a cabo para construir sus reportes?

- Extracción de datos
- Notificación de correcciones y depuración
- Aplico las transformaciones
- Creo el reporte tabular
- Creo el reporte visual

¿A quién van dirigidos los reportes que genera actualmente?

Los reportes que se generan en el área se dirigen a diversos comités crediticios, Junta Directiva, jefaturas y gerentes de la entidad.

¿Cuál es la periodicidad con la que elabora estos reportes?

La periodicidad es mensual y se ejecutan después del cierre.

¿En promedio cuanto tiempo invierte en elaborar cada informe?

En la extracción duro unos 30 minutos, en la depuración una hora, transformando otra hora, creando el reporte en tablas 30 minutos y el gráfico unos 20 minutos.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

¿Qué oportunidades de mejora ha identificado en el proceso para elaborar informes?

Remover el procesamiento manual de archivos de Excel y unificar la información de crédito en una fuente de consulta.

¿De qué forma gestiona los informes de índole histórica?

Prefiero utilizar mi consolidado de operaciones ya que posee las transformaciones que aplico en mi caso particularmente.

Apéndice II: Entrevista para selección de criterios del Benchmarking

Datos generales

Puesto del entrevistado	Jefatura de sistema de la organización
Rol del entrevistado en el proyecto	Consultado
Fecha de aplicación	19/09/2018

El objetivo de esta entrevista es indagar al personal técnico y de negocio de la organización sobre las características deseables en las herramientas de integración y visualización de datos.

Preguntas

- **¿Existen algunas herramientas que requiera incluir en el análisis comparativo?**
Dado que la organización ha trabajado con las plataformas de Microsoft y Oracle es necesario tomarlos en cuenta para el análisis que se realizará.
- **¿Cuáles son las características básicas deseables de las soluciones de integración y visualización de datos?**
Características necesarias la conexión con las bases de datos internas por medio de OLE DB, combinación de flujos de procesamiento y la compatibilidad con archivos planos o de Excel.
- **¿Con cuáles herramientas he trabajado la organización anteriormente en cuanto a integración y visualización de datos?**
Se ha trabajado más que todo con SQL Server Integration Service, Tableau y Power BI.
- **¿Cuál es su expectativa con respecto a las soluciones de visualización e integración de datos?**
Contar con una perspectiva nueva con respecto a este tipo de herramientas ante el vencimiento del acuerdo actual de licencias se requiere corroborar herramientas alternativas ante modificaciones en el presupuesto.

- **¿El personal cuenta con experiencia utilizando herramientas de integración o visualización de datos?**

En el área técnica tenemos una persona especializada en ambientes de SQL Server, en temas de visualización el área técnica queda fuera de alcance.

Apéndice JJ: Ejecución de pruebas para comprobar la cantidad de registros procesados

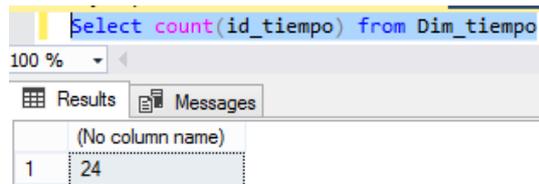
Caso de prueba			
N° de caso	01	Responsable de ejecución	Fernando Mata
Objetivo de la prueba	Corroborar que la cantidad de registros migrada al Data mart corresponda con la cantidad de datos almacenados en las fuentes transaccionales.		
Descripción de la prueba			
Componente o elemento bajo prueba	Componente de carga del proceso de ETL diseñado como parte de la propuesta del proyecto.		
Entradas/insumos	Base de datos multidimensional poblada con la información de la cartera crediticia (cargada por medio del proceso de ETL). Archivos de datos fuente (24 archivos de la cartera crediticia)		
Resultados esperados	Se espera que la cantidad de registros almacenados en la base de datos multidimensional coincida con la cantidad de datos registrados en los archivos fuente.		
Secuencia de pasos ejecutados			
Cálculo de la cantidad de registros almacenados en las carteras crediticias de la entidad financiera.			
Consulta sobre las tablas dimensionales para comprobar la cantidad de registros almacenados en sus estructuras.			
Comparativo de resultados			
Resultados obtenidos			

En primer lugar, la cantidad de registros almacenados en las carteras crediticias es:

126522 filas para 2014 y 115107 filas para 2015, lo que brinda un total de 241629 registros almacenados en 24 archivos de Excel correspondientes a las carteras de crédito.

Seguidamente se realizaron las consultas sobre las tablas dimensionales, obteniendo los siguientes resultados:

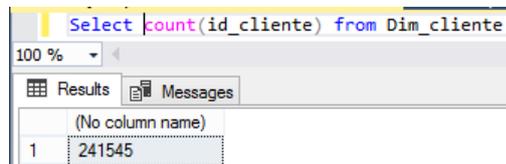
Dimensión tiempo



The screenshot shows a SQL query window with the following text: `Select count(id_tiempo) from Dim_tiempo`. Below the query, there are tabs for 'Results' and 'Messages'. The 'Results' tab is active, showing a table with one row and one column. The column header is '(No column name)' and the value is 24.

(No column name)
24

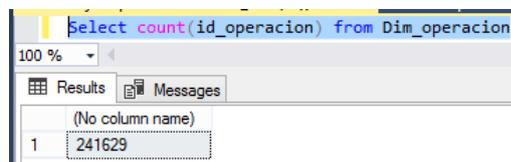
Dimensión cliente



The screenshot shows a SQL query window with the following text: `Select count(id_cliente) from Dim_cliente`. Below the query, there are tabs for 'Results' and 'Messages'. The 'Results' tab is active, showing a table with one row and one column. The column header is '(No column name)' and the value is 241545.

(No column name)
241545

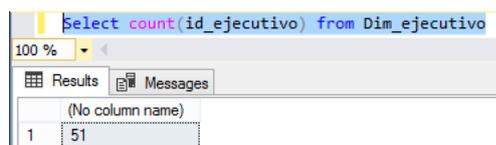
Dimensión operación



The screenshot shows a SQL query window with the following text: `Select count(id_operacion) from Dim_operacion`. Below the query, there are tabs for 'Results' and 'Messages'. The 'Results' tab is active, showing a table with one row and one column. The column header is '(No column name)' and the value is 241629.

(No column name)
241629

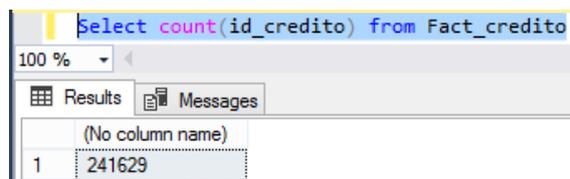
Dimensión ejecutiva



The screenshot shows a SQL query window with the following text: `Select count(id_ejecutivo) from Dim_ejecutivo`. Below the query, there are tabs for 'Results' and 'Messages'. The 'Results' tab is active, showing a table with one row and one column. The column header is '(No column name)' and the value is 51.

(No column name)
51

Tabla de hechos



The screenshot shows a SQL query window with the following text: `Select count(id_credito) from Fact_credito`. Below the query, there are tabs for 'Results' and 'Messages'. The 'Results' tab is active, showing a table with one row and one column. The column header is '(No column name)' and the value is 241629.

(No column name)
241629

Producto de la ejecución de las consultas se confirmó que todos los registros disponibles en las fuentes de datos se migraron satisfactoriamente durante el procesamiento del ETL.

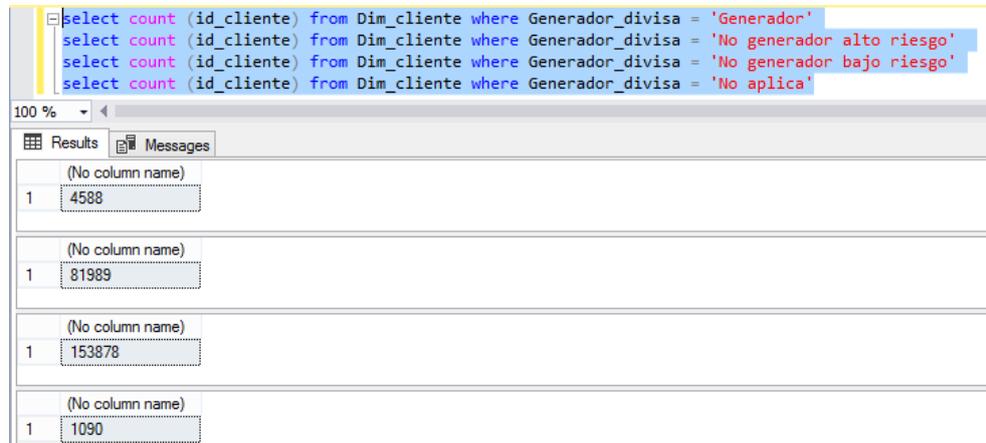
Apéndice KK: Ejecución de pruebas para comprobar las transformaciones de campos numéricos a rangos de valores

Caso de prueba			
N° de caso	02	Responsable de ejecución	Fernando Mata
Objetivo de la prueba	Verificación de los valores transformados para los atributos generador de divisa y categoría de riesgo		
Descripción de la prueba			
Componente o elemento bajo prueba	Transformación realizadas durante el procesamiento del ETL.		
Entradas/insumos	<p>Información almacenada en la base de datos multidimensional del Data mart.</p> <p>Archivos fuente de la cartera crediticia (Archivos de Excel)</p>		
Resultados esperados	<p>Se espera que para el atributo generador de divisa todos los registros tengan al menos un valor que corresponda a No generador, Generador de alto riesgo, Generador de bajo riesgo, No aplica.</p> <p>Para el atributo categoría de riesgo se espera que los registros del modelo multidimensional posean el valor A, B, C, D o E según lo especificado en las transformaciones requeridas por los analistas del área de crédito y riesgo.</p>		
Secuencia de pasos ejecutados			
Ejecución de consultas sobre el modelo dimensional para la dimensión cliente con el objetivo de verificar las transformaciones aplicadas.			
Resultados obtenidos			

Los resultados obtenidos para las consultas sobre el modelo multidimensional se muestran a continuación

Atributo Generador de divisa

```
select count (id_cliente) from Dim_cliente where Generador_divisa = 'Generador'  
select count (id_cliente) from Dim_cliente where Generador_divisa = 'No generador alto riesgo'  
select count (id_cliente) from Dim_cliente where Generador_divisa = 'No generador bajo riesgo'  
select count (id_cliente) from Dim_cliente where Generador_divisa = 'No aplica'
```

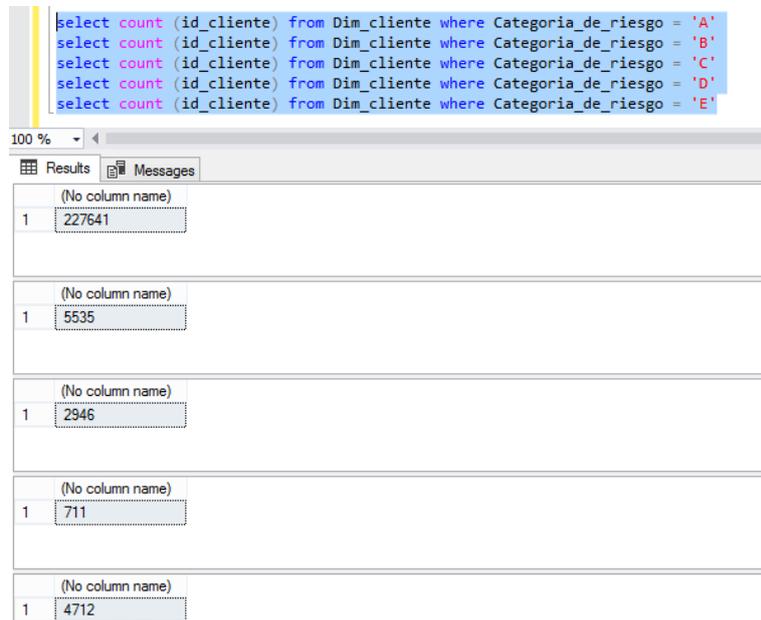


(No column name)
1 4588
1 81989
1 153878
1 1090

Al sumar el resultado de las consultas se obtiene que 241545 registros poseen una clasificación en el atributo generador de divisa, lo cual coincide con la cantidad de clientes registrados en las fuentes transaccionales.

Atributo Categoría de riesgo

```
select count (id_cliente) from Dim_cliente where Categoria_de_riesgo = 'A'  
select count (id_cliente) from Dim_cliente where Categoria_de_riesgo = 'B'  
select count (id_cliente) from Dim_cliente where Categoria_de_riesgo = 'C'  
select count (id_cliente) from Dim_cliente where Categoria_de_riesgo = 'D'  
select count (id_cliente) from Dim_cliente where Categoria_de_riesgo = 'E'
```



(No column name)
1 227641
1 5535
1 2946
1 711
1 4712

Por medio de la consulta sobre la base de datos se determina que todos los clientes del modelo multidimensional poseen al menos un valor en el atributo categoría de riesgo.

Apéndice LL: Comprobación de resultados del proyecto

Una vez implementada la propuesta de solución del proyecto se procedió a verificar si el diseño planteado cumple con la reducción en el tiempo de procesamiento y la cantidad de iteraciones (tasa de reprocesamiento) plasmados en los objetivos de la iniciativa.

Como beneficios del proyecto se plantearon los siguientes enunciados:

- ▣ Disminución en la cantidad de tiempo que invierte el personal en tareas de extracción, preparación y carga de información manual en al menos un 50% del tiempo actual.
- ▣ Reducir en un 80% la cantidad de reprocesamiento que deben realizar los analistas debido a cambios en los requerimientos gerenciales.

Concretamente la mejora esperada sobre los procesos dedicados a la construcción de reportes se menciona a continuación.

Departamento	Indicador	Situación previa organizacional	Mejora relativa esperada	Mejora absoluta esperada
Riesgo	Tiempo promedio para la construcción de informes	292 minutos	50%	146 minutos
	Tasa de reprocesamiento	cinco informes	80%	un informe
Crédito	Tiempo promedio para la construcción de informes	208,75 minutos	50%	104,375 minutos
	Tasa de reprocesamiento	diez informes	80%	dos informes

Una vez que se establece la mejora esperada para los indicadores del proceso se procedió a corroborar su cumplimiento contra la ejecución de la propuesta del proyecto.

Durante el desarrollo de la iniciativa se ejecutó la extracción, transformación y carga de datos correspondientes a 24 meses de la cartera crediticia de la entidad financiera, en una ventana de tiempo de siete minutos y medio aproximadamente.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Se destaca una mejoría del tiempo de procesamiento de datos para los departamentos de Riesgo y Crédito quienes previamente dedicaban 203,66 y 171,25 minutos en promedio para procesar la información de sus reportes.

Una vez finalizado el procesamiento de la información que se utiliza como insumo en la construcción de reportes se procede con la construcción del informe visual en la herramienta Tableau.

Para la construcción del reporte se solicitó que los analistas de las áreas de negocio construir un reporte gráfico por medio de la herramienta Tableau para determinar la cantidad de tiempo que invierten en la elaboración del informe utilizando la propuesta implementada. Producto del desarrollo del proyecto la construcción del reporte visual pasó de 83,33 minutos en promedio a 5,33 minutos para el área de riesgo, mientras que para el área de crédito se pasó de 52,3 minutos en promedio a 8 minutos.

Al comparar los resultados obtenidos al implementar la propuesta de solución se denota una disminución adicional en el tiempo que dedican ambos departamentos a elaborar sus informes, la diferencia fue de 78 minutos en promedio para el departamento de riesgos y de 46,3 minutos en promedio para el departamento de crédito.

En cuanto a la tasa de reprocesamiento de reportes se redujo por completo, ya que el procesamiento de datos no se realiza manualmente y a los usuarios finales únicamente tienen acceso a la herramienta de visualización de datos donde pueden ajustar sus reportes gráficos bajo los parámetros y atributos definidos en el DM únicamente.

Firma de aprobación

Nombre	Rol en el proyecto	Firma
Lorena Castillo García	Contraparte Organizacional	

9

Anexos

Anexo 1: Notación estándar BPMN 2.0

La notación BPMN 2.0 ofrece una serie de elementos gráficos estandarizados para representar los procesos de negocio de las organizaciones sin importar su giro de negocio. A continuación, se describen los elementos principales que componen esta herramienta y su significado en los procesos organizacionales. La Tabla 9.1 muestra la simbología correspondiente a la notación para procesos de negocio.

Tabla 9.1. Simbología notación BPMN 2.0

Componente	Descripción	Representación gráfica
Evento de inicio	Indica el inicio del proceso de negocio.	
Evento de fin	Indica la finalización del proceso de negocio.	
Actividad	Se utiliza para representar las tareas que integran el flujo del proceso de negocio.	
Compuerta exclusiva	Se utiliza para bifurcar el flujo del proceso por medio de una condición booleana o bien para unificar dos líneas de proceso separadas anteriormente.	
Compuerta paralela	Se utiliza bifurcar dos o más flujos paralelos en el proceso. Esta variante de la compuerta detiene la secuencia del flujo hasta que todos los caminos bifurcados finalicen su procesamiento.	
Pool	Enmarca los componentes que especifican el flujo del proceso	
Lane	Se utiliza para indicar roles en el proceso y las tareas que ejecuta cada rol dentro del flujo.	
Flujo de secuencia	Este componente representa el flujo de las actividades, eventos y compuertas en el proceso de negocio.	

Anexo 2: Plantilla para la documentación de pruebas de la organización

Caso de prueba			
N° de caso		Responsable de ejecución	
Objetivo de la prueba			
Descripción de la prueba			
Componente o elemento bajo prueba			
Entradas/insumos			
Resultados esperados			
Secuencia de pasos ejecutados			
Resultados obtenidos			

Anexo 3: Aval del profesor tutor para la defensa del TFG

Aval de Entrega del Documento de Trabajo Final de Graduación

Nota aclaratoria:

Este documento se redacta de acuerdo a las disposiciones actuales de la Real Academia Española con relación al uso del género inclusivo (<https://goo.gl/ITVYiN>).
Al mismo tiempo, se aclara que estamos a favor de la igual de derechos entre los géneros.

Responsabilidad del Profesor Tutor:

1. A solicitud del estudiante, completar el formulario de Aval de Entrega del Documento de Trabajo Final de Graduación.
2. Devolver una respuesta al estudiante que realizó la solicitud de Aval de Entrega del Documento de Trabajo Final de Graduación. La respuesta debe ser por correo (en formato pdf).

Formulario de Aval de Entrega del Documento de Trabajo Final de Graduación:

Yo Lorena Zúñiga Segura Profesor Tutor del Estudiante
Fernando Nafa Orozco carné 201235809,

hago constar que he revisado exhaustivamente el documento académico final del Trabajo Final de Graduación, realizado en el II semestre del 20 18. Asimismo, he verificado la atención de las correcciones realizadas en mi condición de Profesor Tutor. Por lo tanto, autorizo entregar este documento a la Coordinación de Trabajos Finales de Graduación para que se realicen las gestiones correspondientes para la programación de la defensa.

Responsabilidades del estudiante:

1. Solicitar al Profesor Tutor el Aval de Entrega del Documento de Trabajo Final de Graduación. Esta solicitud se debe realizar por correo al Profesor Tutor, después de haber enviado con al menos una semana hábil el documento académico completo para la respectiva revisión integral final.
2. Enviar a la Coordinación de Trabajos Finales de Graduación la respuesta otorgada por el Profesor Tutor según el formato indicado en este documento. Para esto, debe realizar un reenvío del correo a smora@itcr.ac.cr con copia:
 - a. El correo del Profesor Tutor y
 - b. Al correo soniamora0407@gmail.com

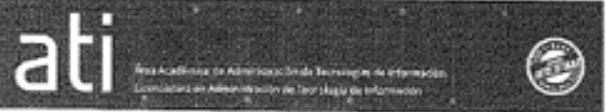
No se requiere la firma del Profesor Tutor, dado que el reenvío del correo del Profesor Tutor garantiza la identidad del Profesor.



Área Académica de Administración de Tecnologías de Información



Anexo 4: primera evaluación del estudiante

Rúbrica de evaluación por parte de la Organización

Criterios básicos de evaluación por parte de la Organización

Nota: Esta es una guía que contiene criterios básicos para la evaluación de los estudiantes de Proyecto Final de Graduación por parte de la Organización. Se compone de 3 secciones y cada contiene un valor porcentual máximo. El evaluador deberá asignar una calificación para cada sección, tomando en consideración los criterios que la componen. La suma de los porcentajes asignados en cada sección dará como resultado la nota total asignada al estudiante en esta evaluación.

NOMBRE DEL ESTUDIANTE: Fernanda Hata Orzco

ORGANIZACIÓN/DEPARTAMENTO/UNIDAD: _____ **FECHA:** Agosto 21, 2018

Por favor, rellene cada uno de los campos.

A. ASPECTOS GENERALES (nota TOTAL máxima: 20%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Nivel de responsabilidad (5% máximo)	5%
Puntualidad (5% máximo)	5%
Cumplimiento de las normas de la organización (5% máxima)	5%
Relaciones interpersonales y de comunicación (5% máxima)	5%
TOTAL	20%

B. ACERCA DEL TRABAJO REALIZADO A LA FECHA (nota TOTAL máxima: 40%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Cumplimiento del cronograma y de las metas establecidas (8% nota máxima)	8%
Calidad del conocimiento técnico (8% nota máxima)	8%
Calidad de los aportes (8% nota máxima)	8%
Resolución de problemas mediante metodologías adecuadas y un proceso ordenado (8% nota máxima)	8%
Disposición para aprender (8% nota máxima)	8%
TOTAL	40%

C. ASPECTOS ACERCA DEL ÚLTIMO INFORME/DOCUMENTO/ENTREGABLE REVISADO (nota TOTAL máxima: 40%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Presentación del informe (8% nota máxima)	8%
Complejidad del informe (8% nota máxima)	8%
Redacción del informe (8% nota máxima)	7%
Aporte a la organización del contenido del informe (8% nota máxima)	8%
Calidad de los productos generados (8% nota máxima)	8%
TOTAL	39%

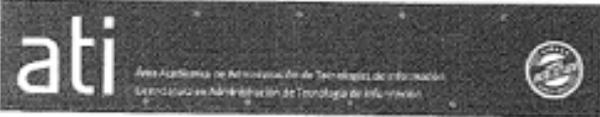
Observaciones Generales:
Fernanda muestra mucha dedicación e interés por cumplir su tiempo y firma con el trabajo por escrito. Se ha esforzado en entender tanto la operativa, origen de la información como el requerimiento y necesidad del trabajo.

Nombre del Evaluador/Contraparte de la Organización: Ana Lorena Castillo Garcia

Firma del Evaluador/Contraparte de la Organización: [Firma]

Anexo 5: segunda evaluación del estudiante

TEC | Tecnológico de Costa Rica




Rúbrica de evaluación por parte de la Organización

Criterios básicos de evaluación por parte de la Organización

Nota: Esta es una guía que contiene criterios básicos para la evaluación de los estudiantes de Proyecto Final de Graduación por parte de la Organización. Se compone de 3 secciones y cada contiene un valor porcentual máximo. El evaluador deberá asignar una calificación para cada sección, tomando en consideración los criterios que la componen. La suma de los porcentajes asignados en cada sección dará como resultado la nota total asignada al estudiante en esta evaluación.

NOMBRE DEL ESTUDIANTE: Fernando Mata Orozco

ORGANIZACIÓN/DEPARTAMENTO/UNIDAD: _____ FECHA: Septiembre 27, 2018.

Por favor, rellene cada uno de los campos.

A. ASPECTOS GENERALES (nota TOTAL máxima: 20%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Nivel de responsabilidad (5% máximo)	5%
Puntualidad (5% máximo)	5%
Cumplimiento de las normas de la organización (5% máxima)	5%
Relaciones interpersonales y de comunicación (5% máxima)	5%
TOTAL	20%

B. ACERCA DEL TRABAJO REALIZADO A LA FECHA (nota TOTAL máxima: 40%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Cumplimiento del cronograma y de las metas establecidas (8% nota máxima)	8%
Calidad del conocimiento técnico (8% nota máxima)	8%
Calidad de los aportes (8% nota máxima)	8%
Resolución de problemas mediante metodologías adecuadas y un proceso ordenado (8% nota máxima)	8%
Disposición para aprender (8% nota máxima)	8%
TOTAL	40%

C. ASPECTOS ACERCA DEL ÚLTIMO INFORME/DOCUMENTO/ENTREGABLE REVISADO (nota TOTAL máxima: 40%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Presentación del informe (8% nota máxima)	8%
Complejidad del informe (8% nota máxima)	8%
Redacción del informe (8% nota máxima)	7,5%
Aporte a la organización del contenido del informe (8% nota máxima)	8%
Calidad de los productos generados (8% nota máxima)	8%
TOTAL	39,5%

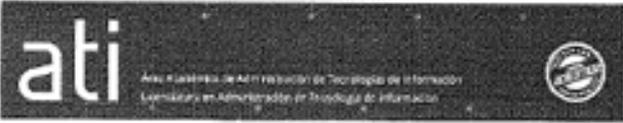
Observaciones Generales:

Nombre del Evaluador/Contraparte de la Organización: Ana Lorena Costillo Guerra

Firma del Evaluador/Contraparte de la Organización: [Firma]

Anexo 6: tercera evaluación del estudiante

TEC Escuela Superior de Tecnología




Rúbrica de evaluación por parte de la Organización

Criterios básicos de evaluación por parte de la Organización

Nota: Esta es una guía que contiene criterios básicos para la evaluación de los estudiantes de Proyecto Final de Graduación por parte de la Organización. Se compone de 3 secciones y cada contiene un valor porcentual máximo. El evaluador deberá asignar una calificación para cada sección, tomando en consideración los criterios que la componen. La suma de los porcentajes asignados en cada sección dará como resultado la nota total asignada al estudiante en esta evaluación.

NOMBRE DEL ESTUDIANTE: Fernando Mata Orzoco

ORGANIZACIÓN/DEPARTAMENTO/UNIDAD: _____ **FECHA:** Noviembre 2, 2018

Por favor, rellene cada uno de los campos.

A. ASPECTOS GENERALES (nota TOTAL máxima: 20%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Nivel de responsabilidad (5% máximo)	5%
Puntualidad (5% máximo)	5%
Cumplimiento de las normas de la organización (5% máxima)	5%
Relaciones interpersonales y de comunicación (5% máxima)	5%
TOTAL	20%

B. ACERCA DEL TRABAJO REALIZADO A LA FECHA (nota TOTAL máxima: 40%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Cumplimiento del cronograma y de las metas establecidas (8% nota máxima)	8%
Calidad del conocimiento técnico (8% nota máxima)	8%
Calidad de los aportes (8% nota máxima)	8%
Resolución de problemas mediante metodologías adecuadas y un proceso ordenado (8% nota máxima)	8%
Disposición para aprender (8% nota máxima)	8%
TOTAL	40%

C. ASPECTOS ACERCA DEL ÚLTIMO INFORME/DOCUMENTO/ENTREGABLE REVISADO (nota TOTAL máxima: 40%)

Rubro a evaluar (DISTRIBUCIÓN)	Nota porcentual a asignar
Presentación del Informe (8% nota máxima)	8%
Compleitud del informe (8% nota máxima)	8%
Redacción del informe (8% nota máxima)	7,5%
Aporte a la organización del contenido del informe (8% nota máxima)	8%
Calidad de los productos generados (8% nota máxima)	8%
TOTAL	39,5%

Observaciones Generales:

Nombre del Evaluador/Contraparte de la Organización: Ana Legena Castillo Garcia

Firma del Evaluador/Contraparte de la Organización: [Firma]

Anexo 7: Plantilla para minuta de reunión

Reunión No.	01	Fecha:		
Lugar:		Hora Inicio:		
		Hora Finalización		
Objetivo de la reunión:				
Participantes:	Presentes:			
	Ausentes: -			
Temas Tratados				
No.	Asunto	Comentarios	Acuerdos	Responsable
1				
Próxima reunión				
Temas a tratar	Fecha	Convocados		
Firmas de aprobación				
Nombre	Rol	Firma		

Anexo 8: Matriz de trazabilidad

Título del TFG	Propuesta para la elaboración de reportes crediticios de una entidad financiera		
Situación problemática	<p>Los analistas deben estandarizar la información y posteriormente realizar sus cálculos lo que conlleva a una disminución en el tiempo de entrega de sus reportes.</p> <p>Debido a la descentralización de las fuentes de datos y a la diferencia en las tareas que realizan los analistas durante la preparación de sus informes se han presentado diferencias en los montos de las operaciones e indicadores crediticios construidos a partir de una misma fuente de datos.</p> <p>La cantidad de iteraciones sobre el procesamiento de datos que realizan los analistas durante la construcción de los informes repercute en las fechas de entrega de la información crediticia.</p>		
Objetivo general	Desarrollar el diseño de un modelo multidimensional y la automatización de los procesos de ETL en las áreas de riesgo y crédito de una entidad bancaria, mediante un análisis de las mejores prácticas de la industria, para la disminución de las brechas en tiempos de entrega de información a los departamentos comerciales de la entidad financiera.		
Objetivos específicos	Entregables	Variables	Instrumentos y técnicas
Diagnosticar las brechas en las tareas de extracción, transformación y carga en los procesos de inteligencia de negocio del área de crédito y riesgo con respecto a las mejores prácticas de la industria, para la identificación de las causas que imposibilitan la entrega de información oportuna.	<ul style="list-style-type: none"> ☐ Compendio de información interna y mejores prácticas de la industria para soluciones de ETL y modelos multidimensionales. ☐ Mapeo de fuentes transaccionales. 	<ul style="list-style-type: none"> ☐ Brechas en las tareas de ETL. ☐ Mejores prácticas para ETL 	<ul style="list-style-type: none"> ☐ Revisión documental ☐ Entrevista ☐ Tabulación de datos ☐ Diagramación

Propuesta para la elaboración de reportes crediticios de una entidad financiera

<p>Recomendar tres alternativas de inversión acordes a las necesidades actuales de la entidad financiera en términos cualitativos y cuantitativos estableciendo a la gerencia opciones que permitan la implementación de la propuesta para la automatización del procesamiento de datos.</p>	<ul style="list-style-type: none"> ☐ Lista de requerimientos para proceso ETL. ☐ Análisis de alternativas de inversión. 	<ul style="list-style-type: none"> ☐ Alternativas de inversión para implementar la propuesta. 	<ul style="list-style-type: none"> ☐ Benchmark ☐ Tabulación de datos ☐ Revisión documental
<p>Diseñar el modelo multidimensional, <i>dashboards</i> de información crediticia y los procesos de ETL asociados a la preparación de datos que se utilizan como insumo en los análisis de las áreas de crédito y riesgo, para la entrega de información a las áreas comerciales.</p>	<ul style="list-style-type: none"> ☐ Diseño del modelo multidimensional. ☐ Diseño del proceso ETL. ☐ Diseño de las visualizaciones de datos. 	<ul style="list-style-type: none"> ☐ Modelos multidimensionales de datos. ☐ Procesos de ETL ☐ Visualizaciones de datos. 	<ul style="list-style-type: none"> ☐ Tabulación de datos ☐ Revisión documental
<p>Construir una solución automatizada que facilite el procesamiento de datos en las tareas analíticas de Crédito y Riesgo para la disminución del reprocesamiento de informes en un 80% y la reducción del tiempo de elaboración en por lo menos un 50%.</p>	<ul style="list-style-type: none"> ☐ Implementación del modelo multidimensional, <i>dashboards</i> y el proceso de ETL. ☐ Casos de prueba. ☐ Informe de resultados. 	<ul style="list-style-type: none"> ☐ Procesamiento de datos. ☐ Reprocesamiento de informes ☐ Tiempo para la elaboración de informes. 	<ul style="list-style-type: none"> ☐ Tabulación de datos

<p>Conclusiones</p>	<p>Conclusiones del objetivo específico 1</p> <ul style="list-style-type: none">❏ Según los resultados de las entrevistas especificadas en la sección 4.1.1.4 de este documento, la ejecución manual de las extracciones y procesamiento de datos consume 203,66 y 171,25 minutos en promedio a los departamentos de riesgo y crédito, indicando que uno de los factores que afectan los tiempos de entrega de los reportes crediticios es la ejecución de tareas manuales dedicadas a la extracción y procesamiento de datos.❏ Se identificó una brecha con respecto a la práctica mencionada en la sección 2.9.2 de este documento donde se establece que únicamente se deben procesar los datos que se migrarán al <i>Data mart</i>, según el proceso para la construcción de informes detallado en la sección 4.1.1.4 del análisis de resultados, los analistas de la institución no aplican filtros sobre la información extraída de las fuentes transaccionales antes de iniciar las transformaciones de datos. <p>Conclusiones del objetivo específico 2</p> <ul style="list-style-type: none">❏ Producto de los resultados del comparativo de herramientas, Pentaho Data Integration es la opción que cumple la mayoría de los criterios cualitativos para herramientas de integración con un 100% (ocho criterios de ocho) de cumplimiento, mientras que el software de visualización Tableau se establece como la alternativa óptima al cumplir con un 100% (ocho de ocho) de los criterios cualitativos. En este caso ROI obtenido al escoger estas herramientas para implementar la solución es de \$2,21. Cabe destacar que según lo indicado en el Apéndice I: Acuerdo sobre herramientas para desarrollo de la propuesta❏ Producto del comparativo de herramienta se establece que el segundo conjunto de herramientas que cumplen la mayoría de los criterios son Talend Data Integration con un 87,5% de cumplimiento (siete de ocho criterios) y Power BI que cumple en un 87,5% (siete de ocho características) los criterios. En este caso el ROI obtenido al implementar la solución con estas herramientas es de \$2,9.❏ Al realizar el comparativo de herramientas se determinó que Oracle Data Integrator 12 c que cumple en un 87,5% (siete de ocho) de los criterios establecidos y mientras que Sisense cumple
----------------------------	--

	<p>en un 75% (seis de ocho) los criterios. En este caso el ROI obtenido al implementar la solución con estas herramientas es de \$1,16.</p> <p>Conclusiones del objetivo específico 3</p> <ul style="list-style-type: none">❏ Se diseñó el modelo multidimensional que solventa los requerimientos de las áreas de negocio plasmados en la sección 4.1.1.2 del capítulo de análisis de resultados.❏ Se elaboraron los diseños para tres reportes gráficos basados en la información del modelo multidimensional lo que facilitó la construcción de los informes crediticios.❏ Se diseñó el proceso de Extracción, Transformación y Carga que se encarga de integrar la información de las carteras crediticias históricas de la entidad, reduciendo la intervención manual de los analistas en las tareas de procesamiento en un 100%. <p>Conclusiones del objetivo específico 4</p> <ul style="list-style-type: none">❏ Al implementar la propuesta del proyecto se disminuyó el tiempo de procesamiento de datos del área de riesgo en 279 minutos con respecto al tiempo de construcción manual del reporte, mientras que para el área de crédito se logró disminuir el tiempo de procesamiento en 193,03 minutos con respecto al tiempo de elaboración manual.❏ Por medio de la implementación de la propuesta se logró reducir al 100% la cantidad de iteraciones que realizaban los analistas en las tareas de procesamiento. Alcanzando el objetivo de disminuir en 80% el reprocesamiento de informes al brindar acceso a los usuarios únicamente sobre la capa de visualización de datos donde pueden elaborar sus propios informes sobre un conjunto de datos pre procesados.
--	--

Recomendaciones	<p>Recomendaciones para Objetivo 1</p> <ul style="list-style-type: none">❏ Establecer un procedimiento estándar para la construcción de reportes en las áreas de negocio de la entidad financiera.❏ Realizar la gestión de limpieza de datos y actualización de los datos de los clientes con el fin de mantener información actualizada a nivel de las carteras crediticias y la solución planteada.❏ Para asegurar la creación consistente de los reportes que realizan las áreas de Riesgo y Crédito de la entidad es necesario reforzar la estandarización y controlar las fuentes de ingreso de datos que alimentan las fuentes de datos analíticas.❏ Construir una bitácora de los errores presentados en la información para determinar las causas y aplicar las acciones correctivas correspondientes. <p>Recomendaciones para Objetivo 2</p> <ul style="list-style-type: none">❏ Producto del comparativo de alternativas existen opciones para herramientas de integración y visualización que son de interés para la entidad financiera tales como Power BI, Talend Data Integration y Sisense. Se recomienda retomar la discusión de estas soluciones en los foros de adquisición anual de licencias. <p>Recomendaciones para Objetivo 3</p> <ul style="list-style-type: none">❏ Realizar el mapeo y construcción de reportes adicionales que se podrían incluir en la solución propuesta. <p>Recomendaciones para Objetivo 4</p> <ul style="list-style-type: none">❏ Incorporar la ejecución de la propuesta desarrollada en la ventana mensual de actualización que posee la entidad financiera, para los procesos que tramitan la información crediticia.
------------------------	---

Propuesta para la elaboración de reportes crediticios de una entidad financiera

	<ul style="list-style-type: none"><li data-bbox="745 193 2047 268"> Diseñar los roles y accesos a las bases de datos de producción a los usuarios finales que tendrán acceso a los datos del DM por medio de las herramientas de visualización de datos.<li data-bbox="745 304 2047 379"> Gestionar el acceso a las herramientas de visualización en cada área para garantizar la construcción y disponibilidad de la información para análisis.<li data-bbox="745 416 2047 491"> Mantener la estructura de los archivos de cartera crediticia según lo establecido en el capítulo V de este documento.
--	--

10

Bibliografía

Anand, N. y Kumar, M. (2013). Modeling and optimization of extraction-transformation-loading (ETL) processes in data warehouse: An overview. *Fourth International Conference on Computing Communications and NETWORKING Technologies*, pp. 1-5.

Diouf, P.S., Boly, A., & Ndiaye, S. (2018). Variety of data in the ETL processes in the cloud: State of the art. *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, pp. 1-5.

Eckerson, W. (2005). Performance dashboards measuring monitoring and managing your business, John Wiley & Sons Inc

El-Sappagh, A., Hendawi, A. y Bastawissy, E (2011). A proposed model for data warehouse ETL process. *ScienceDirect*, 23 (2), 91-104

Hernández, R, Fernández, C y Baptista, M. (2014). Metodología de la investigación. McGraw Hill

IBM (2016). IBM Informix V 12.10 documentation. https://www.ibm.com/support/knowledgecenter/en/SSGU8G_12.1.0/com.ibm.sqlt.doc/ids_sqt_255.htm

IBM, R. (2005). *Data mart consolidation: getting control of your enterprise information*. Retrieved from <https://ebookcentral.proquest.com>

Kimball, R., Ross, M. (2013). The data warehouse toolkit. John Wiley & Sons Inc

Kimball, R y Caserta, J. (2004). The Data Warehouse ETL Toolkit, Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. John Wiley & Sons Inc

Michalewicz, M., Michalewicz, Z., Schmidt, M. y Chiriach, C. (2006). *Adaptative Business Intelligence*, Berlin, Springer

Microsoft (2017). Description of the database normalization basics. <https://support.microsoft.com/en-us/help/283878/description-of-the-database-normalization-basics>

Moro, S., Cortez, P. y Rita, P. (2014). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *ScienceDirect*, 42 (3), 1314-1324

Nagabhushana, S. (2006). *Data warehousing, OLAP and data mining*. New Age International Publishers

Nussbaumer, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons Inc.

Ñaupas, H, Mejía E, Novoa, E, Villagómez, A. (2014). *Metodología de la investigación. Cuantitativa – Cualitativa y redacción de la tesis*. Ediciones de la U

Pan, B., Zhang, G., Qin, X. (2018). Design and Realization of an ETL Method in Business Intelligence Project. 2018 the 3rd IEEE International Conference on Cloud Computing and Big Data Analysis, pp. 275 – 279

Phillips, J., Brantley, W., Phillips, P. (2011). *Project Management ROI: A Step-by-Step Guide for Measuring the Impact and ROI for Projects*. John Wiley & Sons Inc.

Ponniah, P. (2001). *Data warehousing fundamentals a comprehensive guide for it professionals*, John Wiley & Sons Inc

Rasmussen, N., Bansal, M., y Chen, C. (2009). *Business dashboards: a visual catalog for design and deployment*. John Wiley & Sons Inc.

Propuesta para la elaboración de reportes crediticios de una entidad financiera

Real Academia Española. (2017). Diccionario de la lengua española (23.a ed.1). Consultado en <http://dle.rae.es/srv/search?m=30&w=visualizar>

Reeves, L. (2009). A manager's guide to data warehousing.

Resch, M (2011). Strategic Project Management Transformation: Delivering Maximum Roi & Sustainable Business Value. J.Ross Publishing

Stacey, M., Salvatore, J., Jorgensen, A (2013). Visual Intelligence. Microsoft Tool and Techniques for Visualizing Data. John Wiley & Sons Inc.

Shariat, M., Hightower, R. (2007). Conceptualizing business intelligence architecture. Florida A & M University

Turban, E., Aronson, J. y Liang, T. (2005). Decision support systems and intelligent systems. Prentice Hall

Williams, N., Williams S. (2007). The profit impact of business intelligence, Morgan Kaufmann

Wijaya, R. y Pudjoatmodjo,B. (2015). An Overview and Implementation of Extraction-Transformation-Loading (ETL) Process in Data Warehouse (Case Study: Department of Agriculture). 2015 3rd International Conference on Information and Communication Technology (ICoICT), pp. 1-5

Wu, X. y Zhang, H. (2010). Design and Implementation of Data Warehouse of Minor Chain Supermarkets. IEEE

11

Glosario

Inteligencia de negocio: disciplina que engloba el uso de información y la aplicación de diversas técnicas, metodologías y herramientas para solventar las necesidades del negocio o sacar ventaja de sus fortalezas.

BPMN Business Process Management Notation: notación estándar que proporciona la capacidad para ejemplificar el flujo de un proceso de negocio por medio de elementos gráficos.

Data warehouse: es un repositorio histórico de información que almacena grandes cantidades de información para favorecer las técnicas de inteligencia de negocio.

Data mart: especialización del Data warehouse a menor escala, generalmente dedicado a un proceso de negocio único.

Modelo de datos: es la representación ordenada de los componentes que integran una estructura de almacenamiento o base de datos.

Sujeto de información: son las personas u objetos de estudio que forman parte de un proceso de investigación.

Dashboard: es la representación gráfica de los indicadores de negocio de interés para una organización.

Fuente transaccional: es el repositorio o sistema que almacena los datos de transacciones vinculadas al giro de negocio o razón social de una empresa u organización.

Sistema maestro: es el sistema principal de una organización, donde se procesa o almacena la información vital para la subsistencia del negocio.

Repositorio de información: es la estructura de almacenamiento que gestiona la información relacionada a una o más fuentes transaccionales.

11.1. Siglas

BCCR Banco Central de Costa Rica.

SUGEF Superintendencia General de Entidades Financieras.

CONASSIF Consejo Nacional de Supervisión del Sistema Financiero.

ETL Extraction, Transformation, Load