Instituto Tecnológico de Costa Rica
Departamento de Computación
Programa de Maestría

Una valoración de las amenazas y propuesta para mejorar la seguridad de los depósitos de datos

Tesis para optar al grado de
Magister Scientiae en Computación

Sustentante: Marco V. Gámez Acuña

Profesor Asesor: Dr. Carlos A. González A.

Cartago, Costa Rica
Agosto, 2002

Resumen

Cada día las empresas dependen más de sus sistemas de información para apoyar los procesos de toma de decisiones. Una de las alternativas que más auge ha estado tomando en los últimos tiempos son las soluciones basadas en depósitos de datos.

A pesar de que ello constituye una alternativa viable, por el carácter estratégico que se le ha dado a la información en estos tiempos, el tema de la seguridad de la información contenida en los depósitos de datos se ha tornado sumamente importante, máxime si se toma en cuenta que existe la posibilidad de que los usuarios puedan hacer un uso inadecuado de esa información, al punto de que puede poner en riesgo hasta las vidas de los clientes o usuarios de los sistemas.

La presente investigación tuvo como propósito identificar y valorar las causas más comunes que amenazan los depósitos de datos, que pudiera servir de base para formular una propuesta para mejorar la seguridad en depósitos de datos. Dicha propuesta intenta ser útil para aplicaciones ya desarrolladas y para aquellos casos en que se planee desarrollar un nuevo depósito de datos.

La propuesta fue revisada mediante una encuesta que se aplicó entre algunas de las empresas costarricenses que tienen un depósito de datos, o que estén en proceso de desarrollar uno de éstos. También esta propuesta fue aplicada en algunas empresas que desarrollan soluciones de depósitos de datos. A pesar de que la validación no se pudo realizar con una representación estadística, por cuanto se carece de un registro en el que se pueda conocer qué empresas disponen o se aprestan a disponer de un depósito de datos, que se constituya en la población, y a partir de ella, seleccionar una muestra significativa, entre las empresas consultadas hubo consenso en cuanto a que las fases que componen la propuesta son necesarias todas dentro de una estrategia que procure mejorar la seguridad en los depósitos de datos.

La propuesta considera las siguientes actividades:

- Identificación y clasificación de los datos
- Cuantificación del valor de los datos
- Definición de expectativas de seguridad y sus métricas
- Identificación de amenazas
- Análisis de las amenazas
- Selección y planificación de las medidas mitigadoras
- Seguimiento y control de medidas mitigadoras adoptadas

Palabras claves: ***seguridad, depósitos de datos, propuesta metodológica***

Abstract

Companies are increasingly relying on their information systems to support their decision-making processes. One of the most important alternatives used in recent years is the data warehousing solution.

Although this is a practicable alternative due to the strategic value currently given to the information, the issue of security of the data warehouses information has become extremely important, especially if users are likely to use this information inappropriately, to the extent that the lives of the clients or users of the systems may be at risk.

This research is intended to identify and value the most common causes that threaten data warehouses, in order to formulate a proposal to improve their security. Such proposal could be used both with applications already developed and in those cases in which a new data warehouse is being planned.

This proposal has been reviewed through a survey applied among some Costa Rican companies having a data warehouse or in the process of developing one. It was also applied in some companies currently developing data warehousing solutions. Although it was not possible to make the validation using a statistical representation to be used as the population to choose a significant sample from it – there was no record available showing what companies have or are about to have a data warehouse – the companies surveyed agreed that all the stages making up the proposal are necessary as part of a strategy intended to enhance security in the deposits of information.

The proposal considers the following activities:

- Identification and classification of the information
- Quantification of the value of the information
- Definition of security expectations and their metrics
- Identification of threats
- Analysis of threats
- Selection and planning of the mitigating measures
- Follow-up and control of mitigating measures adopted

Keywords: ***security, data warehouses, methodological proposal***

Aprobación del Borrador de Tesis

Una valoración de las amenazas y propuesta para mejorar la seguridad de los depósitos de datos

_____

Dr. Carlos A. González A.

Profesor Asesor (Visto Bueno)

_____

Ing. Kirstein Gätjens S.

Director de la Maestría (Refrendo)

Dedicatoria

Al Todo Poderoso, de quien todo procede.

A mi esposa Ileana,
y a mis hijos Ariel Humberto y Laura Victoria,
por los incontables ratos que los he privado de mi compañía y atención,
en aras de alcanzar esta meta.

A mis sobrinos:
Ana Catalina, Albán Antonio,
Stephanie Alexandra, Keylin Melissa, Jeffrey Roberto,
Steven José (Q.d.D.G), Katherine Ariadna, Kristel Tatiana, Rugen Steven,
David Alberto,
María Isabel, Ivania María y Mario Alonso
(en representación de las nuevas generaciones),
para que valoren la importancia de superarse cada día,
pues los tiempos tienden a ser más difíciles y más competitivos.

Al "Guayacán", Tío Uladislao (Lalo) Gámez,
quien ha sido para mí un ejemplo a seguir.

A la memoria de mi abuelo Antonio Acuña Vargas.

A mis padres,
por haberme dado la vida,
y lo necesario para poderme valer por mí mismo.

Agradecimientos

Deseo dejar patente mi agradecimiento al Dr. Carlos González A., quien no sólo fungió como tutor de esta investigación, sino también por haberme introducido años atrás en el estudio del tema de los depósitos de datos.

A las altas autoridades de la Superintendencia General de Entidades Financieras y del Banco Central de Costa Rica, por el apoyo y la confianza que depositaron en mí al darme facilidades para la elaboración de la presente investigación.

A las diferentes personas (familiares, compañeros de trabajo y amigos en general) quienes con su pregunta ¿cómo va la tesis?, siempre me dieron el impulso y me comprometieron para seguir adelante hasta alcanzar esta meta.

Quiero hacer especial mención de la MBA Ana Isabel Méndez, ya que con su ayuda fue posible contactar a algunas de las empresas participantes en esta investigación. Asimismo, deseo agradecer la colaboración brindada por la MSc. Marlene V. Villanueva Sánchez, y la Licda. Lilliam Ma. Monge Bermúdez, en aspectos de interpretación estadística y traducción al inglés, respectivamente.

A los funcionarios de las empresas e instituciones que colaboraron respondiendo la encuesta, pues sus apreciaciones contribuyeron al mejoramiento de la propuesta que aquí se presenta.

A mis diferentes estudiantes con los que tuve oportunidad de comentar y discutir algunos tópicos de esta investigación, pues aunque tal vez no se hayan percatado de ello, sus observaciones me ayudaron a moldear mejor este trabajo.

A todos mi más sincero agradecimiento por el apoyo y la fortaleza que siempre me supieron dar.

Tabla de contenidos

Índice de figuras

Índice de tablas

Índice de gráficos

Capítulo 1. Introducción

1. Identificación del problema

1.1. Antecedentes

Día con día se ve que es mayor la necesidad de procesar datos a través de medios automatizados. Inmersos como se está en la Era de la Información, resulta difícil pensar en lo que sería el acontecer diario sin sistemas de información que ofrezcan ésta en forma rápida, confiable y oportuna.

Puesto que los volúmenes de datos que se deben procesar ya alcanzan tamaños considerables, y dado que en la mayoría de los casos no necesariamente éstos son manipulados por un mismo tipo o familia de herramientas, resulta muy común que para ese procesamiento se deba trabajar con ambientes heterogéneos: Dado este panorama es que desde hace ya algunos años se vienen estudiando nuevas alternativas para el procesamiento de datos altamente voluminosos y en donde se vean involucrados ambientes heterogéneos.

Es así como se ha llegado a un concepto que muchos concuerdan en denominar depósitos de datos (*data warehouses*), el cual representa a la alternativa viable que más se escucha en el medio. Relacionado con este concepto también se escuchan términos como mercado de datos (*data mart*), minería de datos (*data mining*) y consignación de datos (*data warehousing*).

No obstante, es poco lo que se reporta en relación con la seguridad (entiéndase por ello lo referente a la privacidad e integridad) que se obtiene de trabajar con este tipo de ambientes.

Para entender el concepto de seguridad de la información, se hace necesario comprender primero cuáles son las amenazas a que se ve expuesta la información almacenada en una computadora. A partir de ello se puede explicar mejor cómo es que otras formas de procesamiento o de almacenamiento son vulnerables a estas amenazas básicas. En la Figura 1 se esquematizan tales vulnerabilidades.

Las amenazas se pueden agrupar en dos categorías:

- las que atentan contra la confidencialidad o privacidad (el secreto) de los datos
- las que atentan contra la autenticidad (la integridad) de los datos

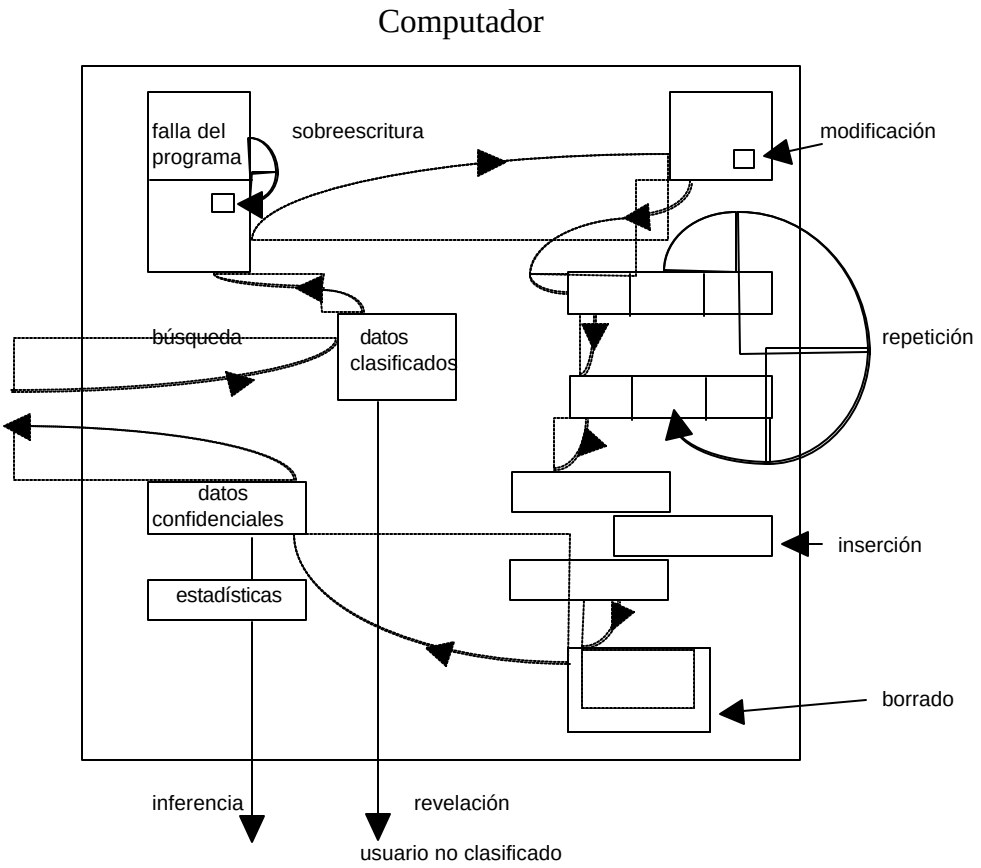


Figura 1. Tipos de amenazas que atentan contra los datos almacenados en una computadora.

En términos muy generales, las amenazas que atentan contra el secreto de los datos incluyen: *búsqueda*, *exposición* (o *revelación*) e *inferencia*.

La *búsqueda* se refiere a la acción de escudriñar cierta información a través de la memoria principal o de los dispositivos de almacenamiento secundario.

La *revelación* se refiere a la transmisión o cesión de datos a usuarios no autorizados, a través de procesos que tienen un acceso legítimo a ellos.

La *inferencia* se refiere a la deducción de datos confidenciales acerca de un individuo u objeto en particular, a partir de la correlación de datos estadísticos presentados en forma de grupos de individuos o de casos.

Por su parte, las amenazas que atentan contra la autenticidad incluyen la *alteración* y *destrucción* de los datos, así como la *suplantación* del autor de la información.

La *alteración* se refiere a la posibilidad de que alguien modifique, repita o inserte cierta información.

La *destrucción* (accidental o intencional) se refiere a la posibilidad de que alguien sobreescriba o borre información.

La *suplantación* se refiere a la posibilidad que adquieren ciertas personas para hacerle creer a otras que ellos son o no los autores de la información, según su conveniencia, pues en algunas ocasiones se puede pretender eludir la responsabilidad de la autoría de un dato, ya sea para no afrontar las consecuencias que ello conlleva, o bien, para engañar a otras personas. En cualquiera de los casos de que se trate, esta amenaza es un tanto seria, pues resulta difícil de determinar a ciencia cierta si una persona es o no el que generó cierta información, sobre todo si se toma en cuenta que un proceso de este tipo puede consumir cantidades de tiempo considerables, y a la postre, para cuando se logre dilucidar el asunto ya no sea oportuno el resultado.

Algunos autores indican que la problemática de la privacidad en los depósitos de datos no es ajena de aquélla que se debe afrontar con las mismas bases de datos, en el sentido de que una alternativa para ofrecer privacidad es acudir a la criptografía; sin embargo, ello lleva implícito todo lo referente a qué es lo que se debe encriptar, qué efectos tiene mantener toda la información encriptada a la hora de atender una interrogación, y quién debe asumir el proceso, si el cliente o el servidor, en caso de tratarse de un ambiente cliente/servidor, entre otras muchas más implicaciones.

Por su parte, en materia de integridad, preocupa asegurar que el dato que se recupere de un depósito sea confiable, es decir, que en vista de que la información que se obtiene, de alguna manera va a ser utilizada para apoyar a la toma de decisiones, se hace necesario conocer hasta qué punto se puede confiar en la información obtenida, pues como ya se indicó anteriormente, puede existir la posibilidad de que la información provenga de ambientes heterogéneos, y posiblemente dispersos geográficamente.

De este modo, la presente investigación procura establecer cuál es la problemática real en torno al procesamiento de información basado en depósitos de datos, intentando identificar las principales amenazas que atentan contra la seguridad de éstos, de forma tal que se pueda formular una propuesta viable que conduzca a minimizar tales amenazas.

Con el objeto de validar tal propuesta, se va a realizar una verificación de ella, al aplicar ésta en un depósito de datos que esté funcionando, de modo tal que la propuesta misma sea probada, y que no quede en un plano meramente teórico o académico, más bien, que se aproxime a la realidad que impera en los ambientes de producción del mundo real.

1.2. Formulación de objetivos

1.2.1. Objetivo General

Formular una guía para identificar y valorar las principales amenazas que atentan contra la seguridad de los depósitos de datos, así como para mejorar dicha seguridad.

1.2.2. Objetivos específicos

a. Compilar un marco teórico con la suficiencia necesaria para identificar y valorar las principales amenazas que atentan contra la seguridad de los depósitos de datos, lo que implica el estudio de algunas propuestas de seguridad para bases de datos.

b. Formular una guía para identificar y valorar las principales amenazas que atentan contra la seguridad de los depósitos de datos.

c. Formular una propuesta para mejorar la seguridad de los depósitos de datos.

d. Validar la propuesta a través de una consulta en algunas de las empresas que ya cuenten con un depósito de datos.

1.3. Alcance y limitaciones

1.3.1. Alcance

La presente investigación abarca los siguientes aspectos:

- un marco teórico en que se abordan los conceptos mínimos necesarios para facilitar el entendimiento de temas como seguridad en general (lo que considera entre otros el tema de la criptografía), depósitos de datos y seguridad en los depósitos de datos (en donde se compila la apreciación del tema por parte de diversos autores, y con variados enfoques)
- una clasificación de amenazas que atentan contra la seguridad de los depósitos de datos, la cual es resumida por el autor de acuerdo con criterios de afinidad entre las posibles causas de amenazas. En esta clasificación se consideran tanto factores de tipo técnico en torno a la seguridad de la información, como los de tipo operativo relacionados con la referida

seguridad, como podría ser el caso de situaciones derivadas de los procedimientos que se apliquen para lograr un ambiente seguro de procesamiento, o bien, de aspectos relacionados con las personas que participan en el procesamiento

- una estrategia para valorar el impacto de las amenazas identificadas, valoración que se basa en criterios tales como riesgo inherente y riesgo de control
- una propuesta para mejorar la seguridad en ambientes de depósitos de datos, la cual considera dos posibles situaciones: que el depósito de datos esté en proceso de desarrollo y que el depósito ya haya sido desarrollado
- una determinación del grado de adecuación de la propuesta, a través de una consulta a usuarios y técnicos involucrados con el proceso de desarrollo y operación de aplicaciones basadas en depósitos de datos

1.3.2. Limitaciones

Para esta investigación se apuntan las siguientes limitaciones:

- resulta difícil probar lo adecuado de la propuesta llevándola a la práctica, por cuanto para ello se requiere una necesidad real de desarrollar un depósito de datos, aspecto que escapa de las posibilidades del autor y de los objetivos de la investigación, además de que requiere una cantidad de tiempo y recursos económicos considerables
- en Costa Rica no existe un registro al cual acudir en procura de conocer qué empresas disponen de una solución basada en depósitos de datos, o que estén interesadas, o planeen en un futuro cercano, disponer de una aplicación de este tipo. Por lo tanto, no es posible realizar un muestreo, puesto que no se tiene identificada la población
- por tratarse en esta investigación de un tema delicado como lo es el de la seguridad, aún cuando fuera posible identificar todas y cada una de las empresas poseedoras o interesadas en poseer una solución basada en depósitos de datos, no existe garantía de que ellas accedan a participar en la parte práctica de la investigación, puesto que pueden existir disposiciones superiores en las empresas que prohíban suministrar información de esta naturaleza, o bien, que exista cierto temor por parte de las eventuales personas involucradas a exponer los intereses de la empresa para la que laboran, aún y cuando se les advierta que la información que suministren será tratada con la mayor confidencialidad del caso, y que los resultados sólo serán publicados en forma consolidada, de manera que se dificulte la identificación particular de las empresas participantes.

Capítulo 2. Marco Teórico

2. Marco Teórico.

2.1. Seguridad.

La seguridad es una serie de costos de oportunidad en donde se debe balancear la seguridad con la utilización y el costo. La seguridad es inversamente proporcional a la utilización y el costo. Conforme un sistema se vuelve más seguro, éste se vuelve más restrictivo y difícil de utilizar. Por lo tanto, el costo de administrar y mantener un sistema puede subir significativamente. Alcanzar un nivel apropiado de seguridad para un sistema implica un balanceo delicado entre protección y costo[1].

Las amenazas a los sistemas de información y la batalla para protegerlos son como la guerra de guerrillas. La guerrilla usa tácticas ocultas para minar y sabotear su sistema. El enemigo es usualmente desconocido y se preocupa de cubrir su rastro para poder atacar de nuevo sin problemas. En algunos casos, la guerrilla obtiene información necesaria para acceder a sus sistemas de sus propios empleados de confianza. En algunos casos, la guerrilla es el empleado de confianza[2].

2.1.1. Seguridad de la información.

La seguridad en la información es la práctica de proteger los recursos y los datos de un sistema de computadoras y redes, incluyendo la información guardada en dispositivos de almacenamiento y en su transmisión.

Las amenazas de seguridad son tanto naturales como intencionales. Los datos de los sistemas son vulnerables tanto a desastres naturales como a la corrupción creada por gente maliciosa. Inundaciones, fuegos, competidores, gobiernos extranjeros, personas vengativas –o sólo curiosas–, y muchos otros pueden ser la fuente. Un ataque perpetrado por humanos puede ser tanto activo como pasivo:

- los ataques activos tratan realmente de cambios en los datos almacenados o transmitidos. Puede consistir incluso en el borrado, la corrupción, el retraso o el atasco de transmisiones. Los ataques activos pueden aparentar ser accidentes.
- los ataques pasivos tratan de la recolección de información sin que nadie sepa que se está produciendo. Las escuchas y los pinchazos electrónicos son

[1] [HADF1997] página 289.
[2] [SHEL1997] página 16.

ejemplos de ello. La información con frecuencia se usa para atacar un sistema más importante[3].

El NIST (*National Institute for Standards and Technology*, Instituto Nacional para Estándares y Tecnología, de los Estados Unidos) ha resumido los siguientes estándares de seguridad que se refieren como los requerimientos funcionales mínimos de seguridad para sistemas operacionales multiusuario:

- identificación y autenticación: la identificación y la verificación de usuarios se realizan a través de un procedimiento de inicio de sesión y la autorización de uso de otros sistemas se basa en este tema de seguridad
- control de accesos: los derechos y permisos que controlan cómo los usuarios pueden acceder a los recursos y archivos de la red
- control de cuenta y auditoría: un sistema de registro y control de los inicios de sesión de las actividades en los sistemas de red y los enlaces entre ellos y las cuentas de usuarios específicos
- reutilización de objetos: métodos para suministrar a múltiples usuarios la posibilidad de acceder a recursos individuales
- precisión: métodos para proteger los recursos frente a errores, corrupción y accesos autorizados
- fiabilidad: métodos para asegurarse de que los sistemas y los recursos estén disponibles y para protegerlos frente a fallos o pérdidas
- intercambio de datos: métodos para asegurar las transmisiones de datos con canales de comunicaciones externas e internas[4].

En relación con los datos y sus características Chapman y Zwicky[5] mencionan que éstos tienen tres características que deben protegerse:

- confidencialidad: haciendo alusión a que quizá una persona no quiere que otras personas los conozcan
- integridad: es muy probable que el dueño de los datos no quiera que otras personas los cambien
- disponibilidad: con seguridad el usuario dueño de los datos quiere utilizarlos él mismo

Las personas tienden a enfocarse en los riesgos asociados con mantenerlos en secreto, y es cierto que por lo general son grandes riesgos. Muchas organizaciones

[3] [SHEL1997] página 17.
[4] [SHEL1997] páginas 36-37.
[5] [CHAPM1997] página 4.

tienen algunos de sus secretos más importantes (los diseños de sus productos, sus registros financieros, o registros de alumnos) en sus computadoras. Por otro lado, quizá encuentre que en un sitio es relativamente fácil desconectar las máquinas que contienen este tipo de datos altamente secretos en máquinas enlazadas a Internet.

Supóngase que sí puede separar sus datos de esta manera, y que ninguna de la información accesible a través de Internet es secreta. En ese caso, ¿por qué se ha preocupar por la seguridad?. Porque mantener los datos en secreto no es lo único que debe proteger. Aún debe preocuparse por la integridad y la disponibilidad. Después de todo, si sus datos no son secretos y si no importa que los cambien y si no le preocupa que alguien pueda llegar a ellos o no, ¿por qué desperdiciar espacio en disco?.

Aunque sus datos no sean particularmente secretos, sufrirá consecuencias si son destruidos o modificados. Algunas de estas consecuencias tienen costos que puede calcular fácilmente: si pierde datos, tendrá que pagar para reconstruirlos; si pensaba vender esos datos de alguna forma, habrá perdido ventas sin importar si los podía vender de manera directa, si contenían los diseños que utiliza para construir cosas o el código para desarrollar un programa. También hay costos intangibles asociados con cualquier incidente de seguridad. El más serio es la pérdida de la confianza (confianza del usuario, del cliente, del inversionista, de los empleados, de los alumnos, del público) en sus sistemas y datos y, en consecuencia, una pérdida de confianza en su organización.

2.1.2. Niveles de seguridad.

Hare y Siyan[6] hacen una breve exposición acerca de los diferentes niveles de seguridad definidos por el gobierno de los Estados Unidos, los cuales son tomados como punto de referencia por la comunidad profesional informática.

De acuerdo con los estándares para la seguridad de computadoras, desarrollados por el Departamento de Defensa (en inglés *DoD, Department of Defense*) de los Estados Unidos, los famosos Criterios de Evaluación de Estándares Confiables para Computación (en inglés, *Trusted Computer Standards Evaluations Criteria*), o mejor conocidos en el medio como el Libro Naranja (*Orange Book*), se usan algunos niveles de seguridad para proteger el hardware, el software y la información almacenada de ataques. Estos niveles describen diferentes tipos de seguridad física, autenticación de usuarios, confiabilidad del software del sistema operativo, y

[6] [HARE1996] páginas 60-62.

aplicaciones de los usuarios. Estos estándares también imponen límites sobre que tipo de sistemas se pueden conectar a una computadora.

- Nivel D1: esta es la forma más baja de seguridad disponible. Este estándar establece que el sistema completo es no confiable. No se dispone de ninguna protección para el hardware, el sistema operativo es fácil de violentar, no existen consideraciones para la autenticación de los usuarios ni de sus derechos para accesar la información almacenada en la computadora. Este nivel de seguridad típicamente se refiere a sistemas operativos como MS-DOS, MS-Windows y Apple Macintosh System 7.x.
- Nivel C: este nivel se subdivide en dos: el C1 y el C2.
  - El nivel C1, o Sistema de Seguridad Discrecional, describe la seguridad disponible en sistema Unix típico. Existen algunos niveles de protección para el hardware, por lo que éste no es fácil de violentar, a pesar de que ello sea posible. Los usuarios se deben identificar ellos mismos ante el sistema a través de una combinación de código de usuario y contraseña. Esta combinación es utilizada para determinar que derechos de acceso tiene cada usuario a programas e información. Estos derechos de acceso corresponden a los permisos a archivos y directorios. Estos Controles de Acceso Discrecional habilitan al dueño de un archivo o directorio, o al administrador del sistema, a prevenir que ciertas personas, o grupos de personas, accesen esos programas o información. Sin embargo, no previene que el administrador del sistema pueda realizar cualquier actividad. Consecuentemente, un administrador inescrupuloso puede fácilmente violentar la seguridad del sistema sin el conocimiento de los demás. Además, muchas de las tareas diarias de administración del sistema sólo pueden ser realizadas por un usuario conocido como *root*. Con la descentralización de los centros de cómputo imperante hoy día, no es de extrañarse que en una organización sean varias las personas que conocen y administran la contraseña de *root*. Ello conlleva a que sea difícil distinguir entre las acciones llevadas a cabo por cualesquiera dos usuarios que hayan tenido acceso a dicho código.
  - El Nivel C2, fue pensado para ayudar a combatir tales problemas. Partiendo de las características del nivel C1, este nivel agrega características adicionales de seguridad para crear un ambiente de acceso controlado. Este ambiente tiene la capacidad de restringir posteriormente a los usuarios de ejecutar ciertos comandos o accesar ciertos archivos basado no sólo en los permisos sino también en niveles de autorización. Además, este nivel de seguridad requiere que el sistema sea auditable. Ello implica escribir un registro de auditoría (también conocido como bitácora o pista de auditoría) por cada uno de los eventos que ocurren en el sistema. La auditoría se utiliza para mantener un registro de todos los eventos relacionados con la seguridad, tal como aquellas actividades realizadas por el administrador del sistema. La auditoría requiere una autenticación adicional puesto que sin

ella, como se puede estar seguro de que la persona que ejecutó un comando es realmente esa persona. La desventaja de auditar es que requiere tiempo de procesador adicional así como recursos del subsistema de disco. Con el uso de autorizaciones adicionales es posible para usuarios en un nivel C2 tener la autoridad para realizar tareas de administración del sistema sin tener acceso a *root*.

- Nivel B: este nivel se subdivide en tres: el B1, el B2 y el B3.
  - El nivel B1, o Protección de Seguridad Etiquetada, es el primer nivel que soporta seguridad multinivel, tal como secreto y ultra secreto. Este nivel establece que un objeto bajo un control de acceso mandatorio no puede tener sus premisos cambiados por el dueño del archivo.
  - El nivel B2, conocido como Protección Estructurada, requiere que cada uno de los objetos esté etiquetado. Dispositivos tales como discos, cintas o terminales deben tener asignado un nivel de seguridad simple o múltiple. Este es el primer nivel que empieza a direccionar el problema de un objeto del más alto nivel de seguridad comunicándose con otro objeto en el más bajo nivel de seguridad.
  - El nivel B3, o nivel de Dominios de Seguridad, refuerza el dominio con la instalación de hardware. Por ejemplo, el hardware de administración de memoria es utilizado para proteger la seguridad del dominio contra accesos no autorizados o modificaciones desde objetos en diferentes dominios de seguridad. Este nivel también requiere que las terminales de los usuarios sean conectadas al sistema a través de rutas confiables.
- Nivel A: o nivel de Diseño Verificado, es actualmente el más alto nivel de seguridad validado en el Libro Naranja. Este incluye un estricto proceso de diseño, control y verificación. Para que este nivel de seguridad sea alcanzado todos los componentes de los más bajos niveles deben ser incluidos; el diseño debe ser verificado matemáticamente; y se debe realizar un análisis de la cobertura de canales y la distribución confiable. Distribución confiable significa que el hardware y el software han sido protegidos durante el envío para prevenir que hayan sido interferidos con los sistemas de seguridad.

Puede consultarse la obra de Hadfield et al.[7], en especial su V Parte, si se desea mayor información sobre la implementación de los criterios de seguridad definidos por el Departamento de Defensa, en términos generales, y en el caso específico del nivel C2 para Windows NT.

[7] [HADF1997] páginas 413-446.

2.1.3. Mecanismo de control de acceso.

El mecanismo de control de acceso forma parte del subsistema de frontera que establece la interfaz entre un eventual usuario y un sistema[8]. El mecanismo de control de acceso se compone de tres etapas o procesos, por medio de los cuales un usuario debidamente autorizado obtiene los permisos que le han sido concedidos de previo para ejecutar acciones específicas sobre los diferentes objetos del sistema. A saber, dichos procesos son: identificación, autenticación y autorización[9].

2.1.3.1. Proceso de identificación.

Es el proceso por el cual el mecanismo de control de acceso le solicita al usuario algún tipo de información que sirva para determinar su identidad, ello con el objeto de conocer quién es el usuario que está intentando ingresar al sistema.

Con esta información el mecanismo de control de acceso debe determinar si en sus registros existe algún usuario que esté habilitado y activo, que corresponda con la información suministrada.

Dicha información se conoce con diferentes nombres, los más comunes son código de usuario (o *username*), número de cuenta (o *account*), nombre de conexión (o *login name*), por citar algunos, o bien, en el caso de tarjetas con banda magnética, en donde se ha grabado de previo la información de identificación. En la Figura 2 se aprecia el proceso de identificación.

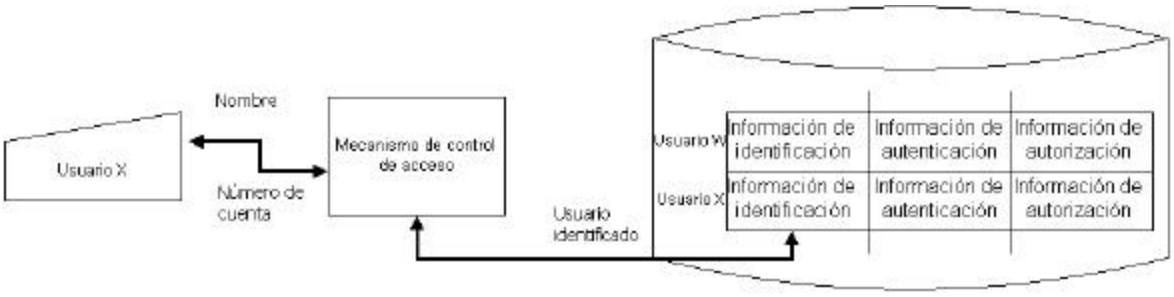


Figura 2. Proceso de identificación.
Fuente: [WEBE1988] página 312.

[8] [WEBE1988] página 310.
[9] [WEBE1988] página 313.

Una vez que el mecanismo de control de acceso ha capturado la información de identificación se procede con la segunda etapa o proceso.

2.1.3.2. Proceso de autenticación.

Puesto que en muchos casos la información que se utiliza para identificar un usuario puede ser pública, el mecanismo de control de acceso necesita autenticar a los usuarios.

El proceso de autenticación consiste en demostrar al mecanismo de control de acceso que el usuario que se está identificando es legítimo, por cuanto la información de identificación no es suficiente para asegurar que el usuario que está intentando ingresar al sistema sea quien dice ser. El proceso de autenticación se puede apreciar en la Figura 3.

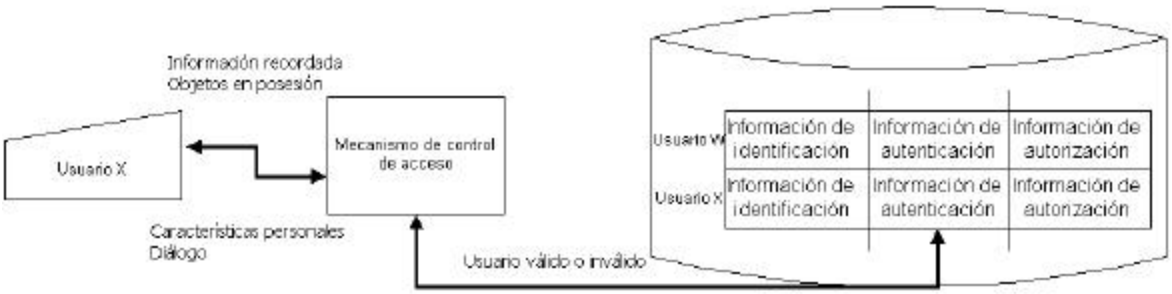


Figura 3. Proceso de autenticación.
Fuente: [WEBE1988] página 312.

Para ello, el usuario debe ingresar cierta información que le resulte privada o exclusiva, es decir, que no sea del dominio público del resto de posibles usuarios. En este sentido, existen diferentes técnicas de autenticación:

- por posesión: el usuario debe presentar ante el mecanismo de control de acceso algún objeto, el que con el sólo hecho de su presentación haga suponer que únicamente el usuario correspondiente es quién lo posee. Ejemplo de ello son llaves, tarjetas con banda magnética, tarjetas con microchip, por citar los que se usan más comúnmente.
- por conocimiento: en este caso, el usuario debe suministrar cierta información que él retiene en su mente, y que es de suponer que solamente el usuario involucrado en el proceso de ingreso la conoce. Ejemplos de este tipo de técnica son las contraseñas (palabras de paso o *passwords*) y los PIN (*Personal Identification Number*).

- características biométricas: otro método para autenticar usuarios es valerse de ciertos atributos físicos que en teoría no son tan fáciles de imitar o suplantar. En este caso constituyen ejemplos utilizados, la impresión de las huellas dactilares, el contorno de la mano, el patrón de la voz, reconocimiento de la retina o el iris, y hasta la firma manuscrita.
- diálogo: esta técnica consiste en que el mecanismo de control de acceso sostiene un diálogo con el usuario, en el cual el sistema le hace algunas preguntas y el usuario las debe responder correctamente, en el entendido de que el verdadero usuario suministró las respuestas correctas a la hora de registrarse inicialmente en el sistema. En estos casos, al usuario que está ingresando se le plantean sólo algunas de las preguntas, las cuales deben ser respondidas adecuadamente; para ello se parte del supuesto de que solamente el verdadero usuario conoce todas las respuestas correctas.

A pesar de todo, por sí sola ninguna de estas técnicas es suficiente para asegurar totalmente la autenticidad de un eventual usuario, razón por la cual se recomienda que para aumentar la efectividad del proceso se combinen las técnicas.

En caso de que la combinación (información de identificación/información de autenticación) suministrada no corresponda con la de alguno de los usuarios registrados en el mecanismo de control de acceso, éste produce un mensaje advirtiéndole al usuario que el intento de ingresar al sistema fue fallido, y que trate de nuevo.

En algunas implementaciones de mecanismos de control de acceso, luego de un número preestablecido de intentos fallidos se pueden generar diversas acciones, tales como el reporte en una bitácora acerca de la situación fallida, o bien, desactivando por un cierto tiempo la computadora desde donde se estuvieron haciendo los intentos de ingreso. Esto último corresponde a una medida extrema, la cual posiblemente puede obedecer a que existieran sospechas en cuanto a que se estuviera intentando ganar acceso, por parte de una persona extraña, a la cuenta de un usuario, sin su consentimiento.

2.1.3.3. Proceso de autorización.

En caso de que resulten satisfactorias las acciones de identificación y autenticación emprendidas por un usuario, éste será reconocido con válido por el mecanismo de control de acceso, y sólo le permitirá ejecutar aquellas acciones para las cuales fue previamente autorizado. Al conjunto de acciones o privilegios que un usuario tiene derecho sobre los diferentes

objetos o recursos del sistema se le conoce con el nombre de perfil (o *profile*) o roles. En la Figura 4 se muestra en qué consiste el proceso de autorización.

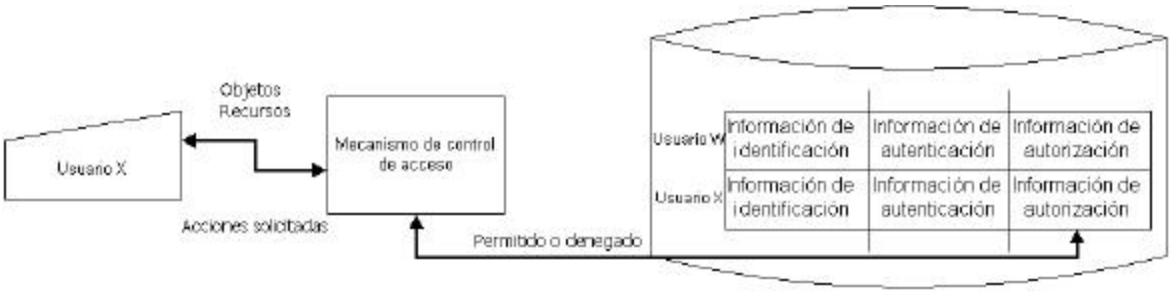


Figura 4. Proceso de autorización.
Fuente: [WEBE1988] página 312.

En consecuencia, no basta el sólo hecho de ingresar satisfactoriamente a un sistema, pues si no se cuenta con los permisos o privilegios suficientes es poco lo que un intruso puede hacer con los objetos definidos dentro de un sistema.

De hecho, antes de ejecutarse cada una de las acciones que emprenda un usuario auténtico de un sistema, se debe validar contra su correspondiente perfil si se cuenta con los derechos suficientes para poderla llevar a feliz término.

En caso de que los permisos establecidos no lo permitan, el mecanismo de control de acceso debe denegar la realización de la acción solicitada.

En caso contrario, posiblemente el usuario ni se percate de esta validación de privilegios.

La definición de permisos sobre objetos puede ser tan granular como se desee y el mecanismo de control de acceso lo permita. Se puede dar el caso en que una definición muy gruesa, es decir, que la definición se haga en términos muy generales, como por ejemplo al nivel de sistemas o de tabla, puede permitir que se lleven a cabo acciones no deseadas por parte de los usuarios. En otros casos, una definición muy granular (como por ejemplo al nivel de tupla o de atributo en un modelo relaciona) puede resultar muy engorrosa, y a la postre no sea justificable desde un punto de vista costo/beneficio, o sea, que se deben relacionar el costo de control y el costo

del objeto que se está controlando, de manera que sólo resultará razonable aplicar un control cuando esta relación sea conveniente (o sea, que el costo del control no exceda el costo del objeto).

2.1.4. Seguridad en bases de datos.

Según Date[10] el término seguridad se usa para hacer referencia a una protección necesaria, la cual se refiere a la necesidad de proteger la base de datos contra consultas y actualización ilícitas. En otras palabras, seguridad se usa para indicar la protección de la base de datos contra accesos no autorizados, alteración o destrucción de la información. El componente del sistema responsable de suministrar seguridad en este sentido se llama mecanismo de autorización (o subsistema de autorización).

En una base de datos esa seguridad se puede materializar mediante diferentes procesos:

- concesión de autoridad: un usuario no puede hacer nada sin autorización explícita, la cual es dada por intermedio del estatuto GRANT
- vistas y autorizaciones: las vistas proveen una medida sencilla pero efectiva, de hecho las vistas pueden usarse para ocultar datos a los usuarios que carecen de autorización para tener acceso a ellos. En este sentido, no basta que se tenga acceso a la vista, las cuales se crean a través del estatuto CREATE VIEW, sino que se requiere que se le dé la autorización correspondiente por medio del estatuto GRANT
- cancelación de la autoridad: puesto que mediante el estatuto GRANT se conceden derechos, debe existir la forma de poder anularlos. Ello se logra por intermedio del estatuto REVOKE, cuyo efecto es precisamente eliminar derechos previamente conferidos.

Adicionalmente Date apunta hacia la necesidad de que se tengan en cuenta otras consideraciones de este tema, como lo son:

- identificación: la necesidad de que los usuarios se identifiquen ante el sistema, ya sea a través de un código de usuario, código, número de cuenta o similar, que facilite al sistema el reconocimiento de los usuarios.
- autenticación: la necesidad imperiosa de que los usuarios se autentiquen ante el sistema, valiéndose para ello de alguna contraseña, palabra de paso, u otro medio que asegure al sistema que el usuario que está intentando solicitar sus servicios sea quien dice ser.

[10] [DATE1987] páginas 191-200.

- registros de auditoría: ante todo hay que partir del supuesto de que los sistemas no son perfectos, y por ende, que es necesario que se dejen las evidencias del caso, a través de registros de auditoría, bitácoras o trazas de auditoría, que permiten a posteriori reconstruir las acciones efectuadas sobre la bases de datos.
- utilización de técnicas de ciframiento: finalmente, se apunta la necesidad de recurrir a la criptografía como medio para preservar la confidencialidad que pueden requerir algunos de los datos almacenados en la bases de datos

Por su parte, Atre[11] define la seguridad en el medio de base de datos como la protección de los datos que se encuentran dentro de la base de datos contra revelaciones, alteraciones o destrucciones no autorizadas o accidentales.

El acceso a los datos sensibles debe controlarse. Esto significa que la cuestión de la seguridad se extiende más allá de los datos almacenados en la base de datos y del departamento de procesamiento de datos. Se extiende a todos los niveles de la administración y por lo tanto éstos deben interesarse en el establecimiento y mantenimiento de la seguridad de los datos.

Teniendo en cuenta que la seguridad perfecta es inalcanzable, el objetivo de un programa de seguridad de datos es minimizar el riesgo y probabilidad de pérdida y revelación al nivel más bajo permisible, así como implantar un programa de recuperación total si ocurre una pérdida.

En relación con la privacidad, Atre apunta que es el derecho que tiene un individuo para determinar el tipo de información, concerniente a él, que se debe almacenar, así como cuándo y cómo; además, es la elección del tipo de información que se puede transferir para otros propósitos distintos de aquellos para los cuales la información se recolectó originalmente. El individuo acerca del cual se recolectó la información es el único "propietario" de esa información.

La privacidad es más una cuestión ética y social que de procesamiento de información. Los sistemas de información se establecen para procesar y proteger los datos, no las fuentes de ellos.

El término privacidad frecuentemente se confunde con el término seguridad, pero las dos palabras tienen distintos significados. La privacidad de la información incluye el derecho de los individuos de saber que la información personal registrada acerca de ellos es exacta, pertinente, completa, actualizada y razonablemente segura de accesos no autorizados, accidentales o intencionales. Una vez que se ha

[11] [ATRE1988] páginas 330-343.

recolectado información relativa a la persona, la distinción entre la información que debe ser pública y la que debe permanecer confidencial, no siempre se establece claramente. Esta decisión requiere de juicios éticos, morales y legales, en los cuales los administradores pueden estar mal preparados o que simplemente no desean hacer. El problema de la privacidad de la información puede involucrar:

- información errónea o engañosa
- revelación o modificación accidental de la información
- infiltración accidental a un sistema de información
- pérdida de datos.

2.2. Depósito de datos (*data warehouse*).

Los depósitos de datos constituyen el centro de la arquitectura de los sistemas de información de los años 90 y dan el soporte al procesamiento de información en aras de proveer una plataforma sólida de datos históricos e integrados, a partir de los cuales se pueda hacer análisis. Además, éstos proveen la facilidad de integración en un mundo de aplicaciones no integradas. Los depósitos de datos se obtienen evolutivamente, al irse dando un paso a la vez. Los depósitos organizan y almacenan los datos necesarios para el procesamiento informacional y analítico sobre una perspectiva histórica de largo plazo. Por lo tanto, existe un mundo de promesa en la construcción y mantenimiento de depósitos de datos.

2.2.1. Definición.

William Inmon, considerado por muchos el padre de los depósitos de datos, define este concepto en función de las características de éstos, tal y como se aprecia en la Figura 5.

"Un depósito de datos es una colección de datos orientada a temas, integrada, variante en el tiempo y no volátil, para dar soporte al proceso de toma de decisiones"[12].

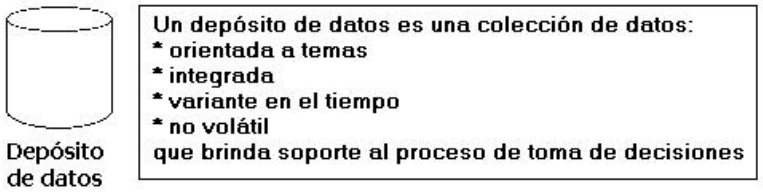


Figura 5. ¿Qué es un depósito de datos?.
Fuente: [INMO1995].

Los datos que entran al depósito provienen en la mayoría de los casos del ambiente operacional. El depósito de datos es siempre un almacenamiento físicamente separado de datos transformados, que provienen de los datos de las aplicaciones que se encuentran en el ambiente operacional.

[12] [INMO1996] página 33.

Puesto que la definición de depósito de datos antes dada resulta un tanto superficial, puesto que existen algunos tópicos importantes detrás de las características de un depósito, de seguido se dará una explicación más amplia de tales tópicos.

2.2.1.1. Orientación a temas.

La primera de las características de un depósito de datos es que está orientado en torno a los principales temas de la empresa.

El esquema dirigido por los datos (*data-driven*), es la más clásica orientación proceso/funcional de las aplicaciones, en el cual los más viejos sistemas operacionales están organizados en torno a las aplicaciones de las compañías. Un contraste de estas dos orientaciones se puede apreciar en la Figura 6.

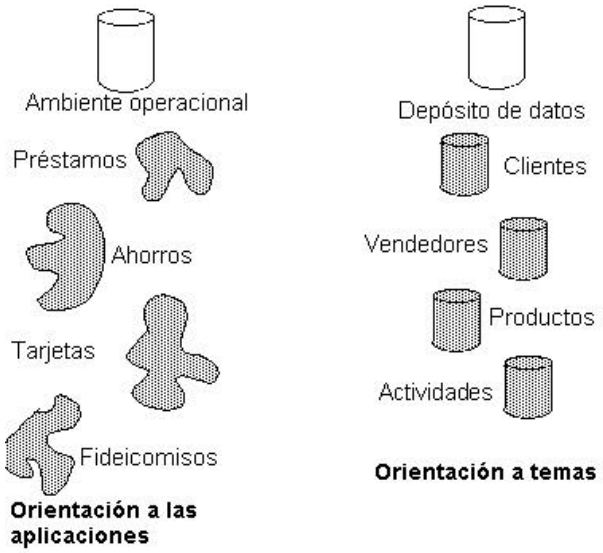


Figura 6. Contraste entre la orientación a las aplicaciones y la orientación a temas.
Fuente: [INMO1995].

El mundo operacional se diseña en torno a las aplicaciones y funciones tales como préstamos, ahorros, tarjetas y fideicomisos, en el caso de una institución financiera. El mundo de los depósitos de datos se organiza en torno a los principales temas, tales como clientes, vendedores, productos y actividades.

El mundo de las aplicaciones se interesa tanto en el diseño de la base de datos como en el diseño del proceso. El mundo de los depósitos de datos se

focaliza exclusivamente en el modelaje de los datos y en el diseño de la base de datos. El diseño del proceso (en su forma clásica) no es parte del ambiente de los depósitos de datos.

La diferencia entre la orientación a las aplicaciones (proceso/función) y la orientación a temas se muestra como una diferencia en el contenido de los datos, así como en el nivel de detalle. El depósito de datos excluye aquellos datos que no serán utilizados en el procesamiento de los sistemas de soporte a la toma de decisiones (DSS), mientras que los datos operacionales orientados a las aplicaciones contienen datos para satisfacer requerimientos inmediatos funcionales o de procesamiento que puede que sean utilizados o no por los analistas de los sistemas de soporte a la toma de decisiones.

Además, existe otro importante aspecto por el que difieren los datos operacionales orientados a las aplicaciones y los datos de un depósito de datos, y es en cuanto a la relación de los datos. Los datos operacionales mantienen una relación actual entre dos o más tablas basado en las reglas del negocio que hay definidas al efecto. Los datos de un depósito de datos atraviesan un espectro de tiempo y las relaciones que se encuentran en un depósito son muchas. Muchas reglas del negocio (y correspondientemente, muchas relaciones de datos) se representan en el depósito de datos entre dos o más tablas.

2.2.1.2. Integración.

Fácilmente se puede entender que el aspecto más importante de un ambiente de depósito de datos es que los datos que se encuentran dentro de él están integrados. De hecho ello debe ser siempre, y sin que quepan excepciones. La verdadera esencia de un ambiente de depósito de datos es que los datos contenidos dentro de sus límites estén efectivamente integrados.

La integración se puede mostrar de muchas formas: en convenciones que dirijan a la utilización de nombres consistentes, en la medición de las variables en forma consistente, en una codificación de estructuras consistente, en la definición consistente de atributos físicos de los datos, entre otros.

En la Figura 7 se muestra un contraste entre la integración encontrada dentro de un depósito de datos y la falta de integración que se encuentra en el ambiente de aplicaciones.

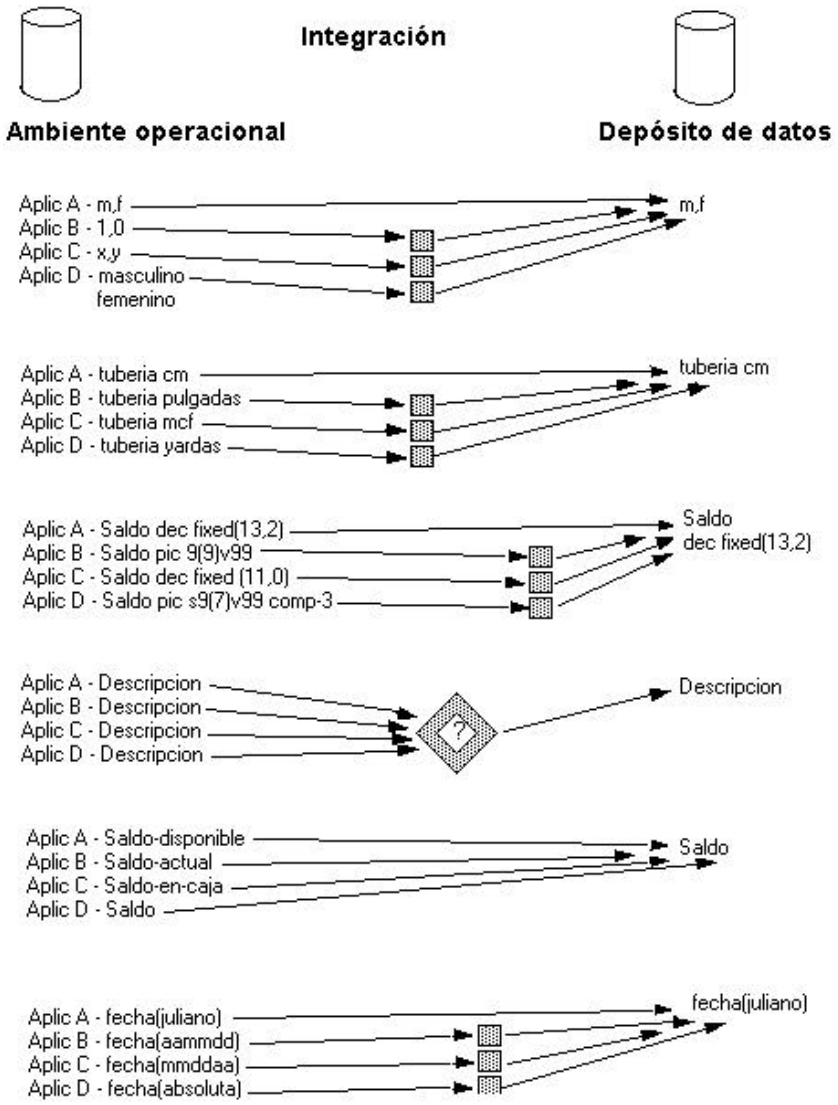


Figura 7. Contraste de integración entre el ambiente operacional y el ambiente de depósito de datos.
Fuente: [INMO1995].

A lo largo de los años los diseñadores de las diferentes aplicaciones han tomado numerosas decisiones individuales en torno a como deberían ser construidas éstas. Las decisiones de estilo y diseño individualizado que tomaron los analistas se muestran en cientos de formas diferentes: diferencias en la codificación, diferencias en las estructuras clave de los sistemas, diferencias en las características físicas, diferencias en las convenciones de nombres, entre otras muchas más. La habilidad colectiva de muchos diseñadores para crear aplicaciones inconsistentes es legendaria. A modo de ejemplo, se le puede dar una mirada nuevamente a la Figura 7 y observar la forma cómo se maneja el valor de fecha en las diversas aplicaciones.

Cualquiera que sea el tema de diseño, el resultado es el mismo, los datos necesitan ser almacenados en el depósito de datos en una forma singular y globalmente aceptable, aún cuando los sistemas operacionales involucrados almacenen los datos diferentemente. Cuando un analista de los sistemas de soporte a la toma de decisiones trabaja con el depósito de datos su enfoque debería ser utilizar los datos que están en el depósito, y no pensar tanto acerca de la credibilidad o consistencia de los datos, pues para ello están los técnicos responsables del depósito, los llamados a atender este tipo de aspectos.

2.2.1.3. Variante en el tiempo.

Todos los datos en el depósito de datos son exactos en algún momento en el tiempo. Esta característica básica de los datos en el depósito es muy diferente de los datos que se encuentran en el ambiente operacional. En el ambiente operacional los datos son exactos en el momento en que se accesan. En otras palabras, en el ambiente operacional, cuando se accesa una unidad de datos, se espera que ello reflejará sus valores exactos al momento del acceso.

Puesto que los datos en el depósito de datos son exactos en cualquier momento en el tiempo (es decir, no ahora mismo), los datos que se encuentran en éste se dice que son "variantes en el tiempo". La Figura 8 muestra precisamente el contraste en relación con esta característica.

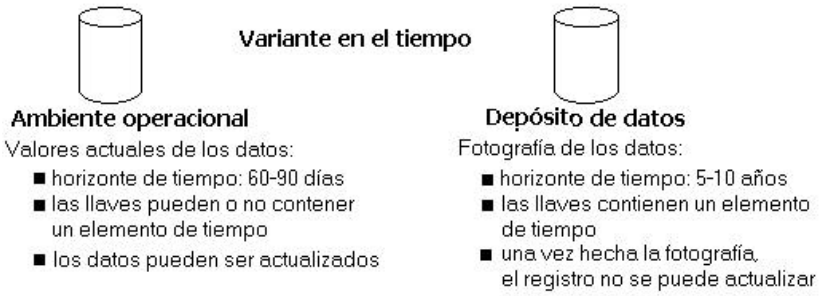


Figura 8. Contraste de la característica de variación en el tiempo entre el ambiente operacional y el ambiente de depósito de datos.
Fuente: [INMO1995].

La variación en el tiempo de los datos del depósito de datos se muestra de diferentes formas. La primera y más simple de ellas es que los datos en el depósito de datos representan datos sobre un horizonte de largo plazo (de entre cinco y diez años). El horizonte de tiempo que representa el ambiente

operacional es mucho más corto (desde el valor actual hasta sesenta o noventa días). Las aplicaciones deben exhibir un buen desempeño y deben ser capaces de procesar transacciones que requieran una mínima cantidad de datos, si es que tienen algún grado de flexibilidad para ello. Por lo tanto, las aplicaciones operacionales tienen un corto horizonte de tiempo, como una cuestión propia de su etapa de diseño.

La segunda forma en que se muestra la variación en el tiempo en el depósito de datos es en la estructura clave. Cualquier estructura clave en el depósito de datos contiene –implícita o explícitamente– un elemento de tiempo, tal como día, semana, mes, entre otros. El elemento de tiempo está casi siempre al inicio de la llave concatenada que se encuentra en el depósito de datos. En ocasiones, el elemento de tiempo podría existir implícitamente, tal es el caso en que un archivo completo se duplique al final del mes o del trimestre.

La tercera forma como aparece la variación en el tiempo es que los datos de un depósito de datos, una vez que se graban correctamente, no pueden ser actualizados. Por lo tanto, los datos de un depósito de datos son, para todos los efectos, una gran serie de fotografías. De hecho, si la fotografía de los datos se ha tomado incorrectamente, entonces la fotografía puede ser cambiada. Pero se parte de que las fotografías fueron hechas apropiadamente, entonces éstas no pueden ser cambiadas una vez que fueron tomadas. En algunos casos puede ser falto de ética y hasta ilegal que las fotografías en el depósito de datos se alteren. Los datos operacionales, en procura de ser exactos al momento del acceso pueden ser actualizados conforme surjan las necesidades.

2.2.1.4. No volátil.

La cuarta característica que define un depósito de datos es que no es volátil. En la Figura 9 se ilustra el contraste de este aspecto entre los depósitos de datos y el ambiente operacional.

Dicha figura muestra que las actualizaciones (sean éstas inserciones, borrados o cambios) son hechas regularmente en el ambiente operacional sobre una base de registro por registro. Pero la manipulación básica de los datos que ocurre en el depósito de datos es mucho más simple. Existen tan sólo dos clases de operaciones que pueden ocurrir en el depósito de datos: la carga inicial de los datos y el acceso a ellos. No existe actualización de los datos (en el sentido general de actualización) almacenados en el depósito de datos como parte normal del procesamiento.

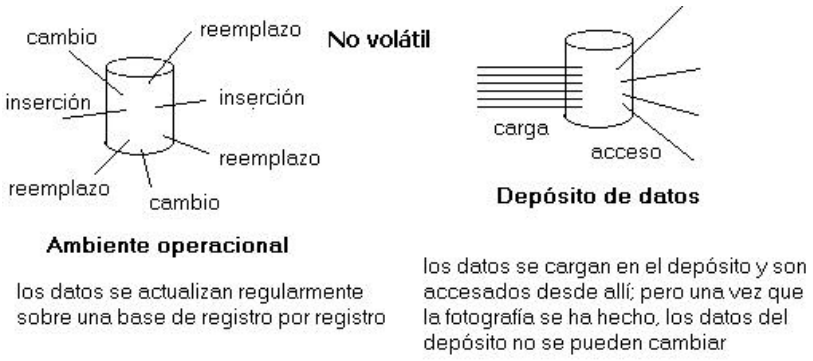


Figura 9. Contraste de la característica de no volátil entre el ambiente operacional y el ambiente de depósito de datos.
Fuente: [INMO1995].

Existen algunas importantes consecuencias de las diferencias básicas entre el procesamiento operacional y el procesamiento de los depósitos de datos. Por ejemplo, en la etapa de diseño no es un factor preocupante la necesidad de tomar precauciones para que la anomalía de las actualizaciones afecte al depósito de datos, pues no se hace actualización de ningún tipo. Esto significa que a nivel del diseño físico se pueden tener suficientes libertades como para optimizar el acceso a los datos, particularmente en el tratamiento de temas como la normalización y desnormalización física.

Otra consecuencia de la simplicidad de la operación de un depósito de datos está en la tecnología fundamental utilizada para correr el ambiente de depósitos de datos. Tener que soportar la actualización registro por registro en un modo de procesamiento en línea (que es el caso más frecuente del procesamiento operacional) hace que la tecnología deba tener un fundamento complejo por debajo de una fachada de simplicidad. La tecnología que ha de soportar respaldos y recuperaciones, transacciones e integridad de datos, así como la detección y resolución de entrabamientos (*deadlock)* es bastante compleja e innecesaria para el procesamiento de un depósito de datos.

Las características de un depósito de datos: orientación a temas del diseño, integración de los datos dentro del depósito, variación en el tiempo y simplicidad de la administración de datos, todas ellas apuntan a un ambiente que es muy, pero muy diferente de un clásico ambiente operacional.

La fuente de datos de casi todos los depósitos de datos es el ambiente operacional. Es una tentación pensar que existe una masiva redundancia de datos entre los dos ambientes. En efecto, la primera impresión que muchas personas tienen es que efectivamente existe una gran redundancia entre el ambiente operacional y el

ambiente de depósitos de datos. Entender eso así es ser superficial y demostrar una falta de entendimiento de lo que está ocurriendo en el depósito de datos.

De hecho, existe un mínimo de redundancia de datos entre el ambiente operacional y el ambiente de depósito de datos. Para ello, téngase en cuenta lo siguiente:

- los datos son filtrados conforme se pasan del ambiente operacional al ambiente de depósito de datos. Por lo tanto, nunca salen muchos datos del ambiente operacional. Sólo aquellos datos que se requieren en el procesamiento de los sistemas de soporte a la toma de decisiones encuentran cabida en el ambiente de depósito de datos.
- el horizonte de tiempo de los datos es bastante diferente entre un ambiente y el otro. Los datos en el ambiente operacional son mucho más frescos. Los datos en el depósito de datos son mucho más viejos. Desde tan sólo la perspectiva de los horizontes de tiempo, existe un muy pequeño traslape entre los ambientes operacional y de depósito de datos.
- el depósito de datos contiene información resumida que se encuentra en el ambiente operacional.
- los datos sufren una transformación fundamental conforme son pasados al depósito de datos. De hecho, la Figura 9 ilustra que la mayoría de los datos son significativamente alterados al ser seleccionados y puestos dentro del depósito de datos. Dicho de otra forma, la mayoría de los datos son física y radicalmente alterados conforme se introducen al depósito. Desde el punto de vista de integración, esto no es lo mismo que le sucede a los datos que residen en un ambiente operacional.

A la luz de estos factores, la redundancia entre los dos ambientes es una rara ocurrencia, resultando en menos de un 1% entre ambas.

2.2.2. Estructura de un depósito de datos.

El depósito de datos tiene una estructura diferente. Existen diferentes niveles de resumen y detalle que demarca el mismo depósito de datos. En la Figura 10 se muestra la estructura de un depósito de datos.

Dicha figura muestra que los componentes de un depósito de datos son:

- metadatos
- datos actuales detallados
- datos detallados más viejos

- datos ligeramente resumidos
- datos altamente resumidos

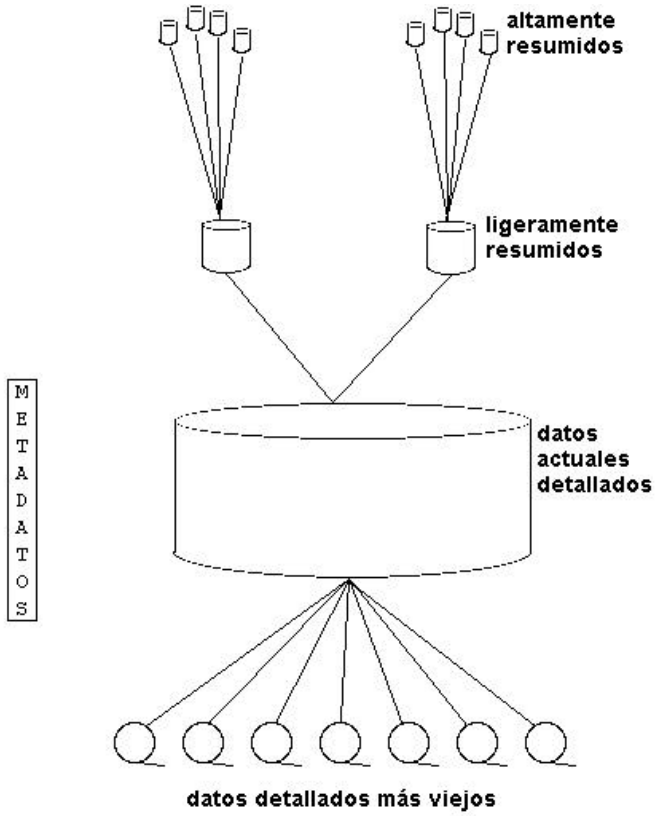


Figura 10. Estructura interna de los datos de un depósito de datos.
Fuente: [INMO1995].

Con mucho el principal interés está en los actuales datos detallados. Este principal interés se debe a que:

- los datos actuales detallados reflejan los acontecimientos más recientes, los cuales son siempre de mayor interés
- los datos actuales detallados son voluminosos debido a que se almacenan en el más bajo nivel de granularidad
- los datos actuales detallados son casi siempre almacenados en disco, lo cual es de rápido de acceso, pero costoso y complejo de administrar.

Los datos detallados más viejos son datos que se guardan en alguna forma de almacenamiento masivo. Son poco frecuentemente accesados y se almacenan en un nivel de detalle consistente con el de los datos actuales detallados. Salvo que medie una disposición mandatoria, los datos son almacenados en un medio de almacenamiento alternativo, dado que el gran volumen anticipado de los datos copa

con el acceso poco frecuente de ellos, y por ello, el medio de almacenamiento para los datos detallados más viejos es usualmente otro que el almacenamiento en disco.

Los datos ligeramente resumidos son datos que se destilan desde el bajo nivel de detalle encontrado en el actual nivel de detalle. Este nivel del depósito de datos es casi siempre almacenado en disco. Las preocupaciones que debe enfrentar el arquitecto de datos al construir este nivel del depósito de datos son:

- cuál es unidad de tiempo sobre la que se hace el resumen
- qué contenido –atributos– deberían tener los datos ligeramente resumidos

El próximo nivel de datos encontrado en un depósito de datos es el de los datos altamente resumidos. Los datos altamente resumidos son compactos y fácilmente accesibles. Algunas veces los datos altamente resumidos se encuentran en el ambiente de los depósitos de datos, y en otros casos dichos datos se encuentran fuera del depósito, en las paredes inmediatas de la tecnología que aloja al depósito de datos. (En algunos casos, los datos altamente resumidos son parte del depósito de datos, sin importar en dónde es que están alojados físicamente los datos).

El componente final del depósito de datos son los metadatos. En muchas formas los metadatos sientan en una dimensión diferente los datos de otros depósitos de datos, debido a que los depósitos de datos no contienen datos tomados directamente del ambiente operacional. Los metadatos juegan un papel especial y muy importante en los depósitos de datos. Los metadatos son utilizados como:

- un directorio que ayuda al analista del sistema de soporte a la toma de decisiones (DSS) a localizar el contenido del depósito de datos
- una guía para trazar un mapa de los datos conforme éstos se van transformando del ambiente operacional al ambiente de depósito de datos
- una guía para los algoritmos utilizados para la preparación de resúmenes, entre los datos actuales detallados y los datos ligeramente detallados, entre los datos ligeramente detallados y los datos altamente detallados, etc.

Los metadatos juegan un papel mucho más importante en el ambiente de depósitos de datos que el que han hecho alguna vez en el clásico ambiente operacional.

En procura de darle vida a los diferentes niveles de datos que se encuentran en un depósito de datos, considérese el ejemplo que se muestra en la Figura 11.

En dicha figura las viejas ventas detalladas constituyen el detalle de aquellas ventas que son anteriores a 1992. Todos los detalles de ventas desde 1982 (o desde cuando

el arquitecto de datos sea capaz de iniciar la colección histórica de detalles) son almacenados en el viejo nivel detallado de datos.

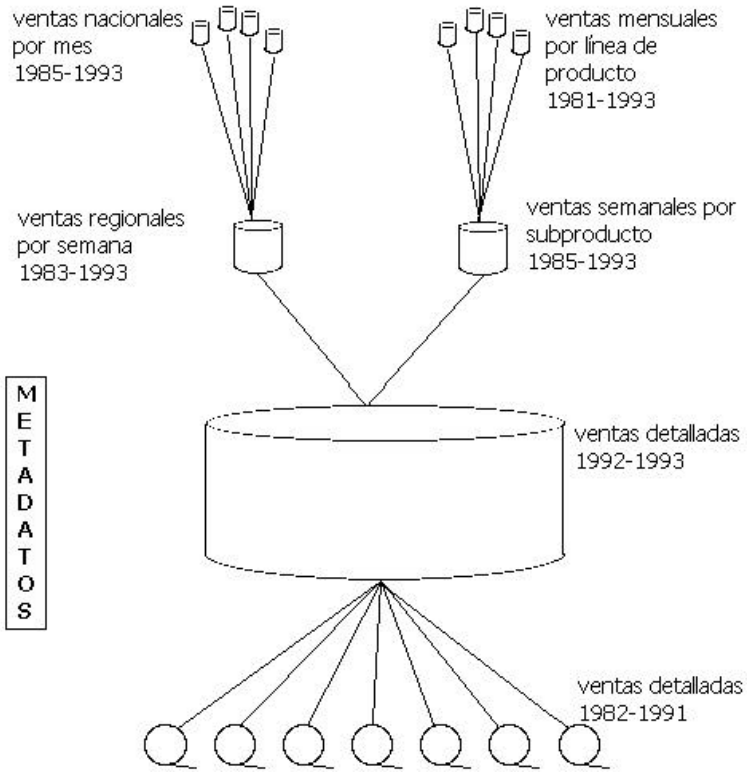


Figura 11. Ejemplo de los niveles de resumen que se pueden encontrar en un depósito de datos.
Fuente: [INMO1995].

El detalle de valores actuales contiene datos desde 1992 hasta 1993 (suponiendo que 1993 es el año actual). Como una regla de ventas el detalle no encuentra su forma en el nivel de detalle actual hasta que no hayan transcurrido por lo menos 24 horas desde que la información de las ventas se tuvo disponible en el ambiente operacional. En otras palabras, existió al menos un lapso de 24 horas entre el momento en que el ambiente operacional se enteró de las nuevas ventas y el momento en que los datos de las ventas fueron introducidos en el depósito de datos.

Las ventas detalladas se resumen semanalmente por línea de subproducto y por región para producir almacenes de datos ligeramente resumidos.

A su vez, las ventas semanales son posteriormente resumidas en forma mensual, a lo largo incluso de líneas más gruesas de productos para generar los datos altamente resumidos.

Los metadatos contienen al menos:

- la estructura de los datos
- el algoritmo utilizado para resumirlos
- el mapa de los datos desde el ambiente operacional hasta el depósito de datos

Nótese que no cualquier resumen que se haga se almacena en el depósito de datos. Podrían existir algunas ocasiones donde al hacer el análisis se generen diferentes tipos de resúmenes. El único tipo de resumen que se almacena permanentemente en el depósito de datos es aquél que corresponde al dato que es frecuentemente utilizado. En otras palabras, si un analista del sistema de soporte a la toma de decisiones (DSS) produce un resultado resumido que tiene una baja probabilidad de ser utilizada nuevamente, entonces dicho resumen no se almacena en el depósito de datos.

Medios de almacenamiento para los viejos datos detallados.

Los símbolos que se mostraron en la Figura 11 como medio de almacenamiento de los viejos datos detallados corresponden a cintas magnéticas. En efecto, las cintas magnéticas pueden ser utilizadas para este tipo de datos. Sin embargo, existe una variedad de otros medios de almacenamiento que podrían ser considerados para contener los viejos datos detallados. A manera de ejemplo, en la Figura 12 se muestran algunos de estos medios de almacenamiento.

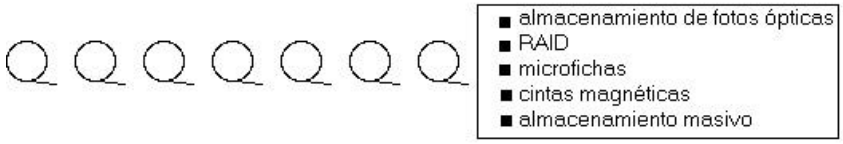


Figura 12. Tipos de medios de almacenamiento utilizados para contener la porción más voluminosa de un depósito de datos.
Fuente: [INMO1995].

Dependiendo del volumen de los datos, la frecuencia de acceso, el costo del medio y el tipo de acceso, es que se puede escoger el medio de almacenamiento que cumpla con las necesidades del viejo nivel de detalle en el depósito de datos.

Flujo de los datos.

Existe un flujo normal y predecible de los datos dentro del depósito de datos. En la Figura 13 se muestra ese flujo.

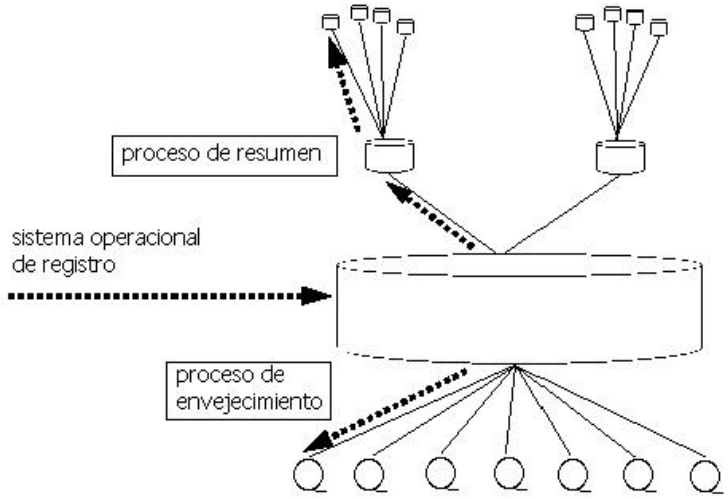


Figura 13. Flujo de los datos dentro de un depósito de datos.
Fuente: [INMO1995].

Los datos entran al depósito de datos, procedentes del ambiente operacional. (Nota: Existen unas pocas interesantes excepciones a esta regla. Sin embargo, casi todos los datos que entran al depósito de datos provienen del ambiente operacional). Conforme los datos van ingresando al depósito desde el ambiente operacional, éstos se van transformado, de acuerdo con lo descrito anteriormente.

Al ir entrando al depósito de datos, éstos vienen en el nivel de detalle actual, tal y como se ha mostrado. Estos residen aquí y se utilizan de esta forma hasta que uno de tres eventos ocurre:

- son purgados
- son resumidos
- son archivados

El proceso de añejamiento dentro del depósito de datos mueve datos actuales detallados hacia los viejos datos detallados, basándose para ello en la edad de los datos. El proceso de resumen utiliza el detalle de los datos para calcular los datos ligeramente resumidos y los datos altamente resumidos.

Existen una pocas excepciones a los flujos antes mostrados (los cuales serán analizados más adelante). Pero sin embargo, para la mayoría de los datos presentes en el interior de un depósito de datos el flujo de ellos es tal y como se mostró anteriormente.

Uso del depósito de datos.

Los diferentes niveles de datos dentro de un depósito de datos reciben diferentes niveles de uso. Se tiene como una regla que a más alto nivel de resumen, mayormente son utilizados los datos, tal y como se puede apreciar en la Figura 14.

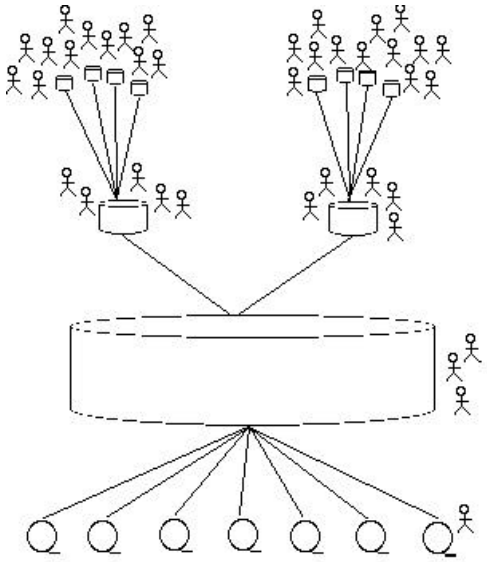


Figura 14. A mayor nivel de resumen, mayor utilización de los datos.
Fuente: [INMO1995].

Esta figura muestra que la mayor utilización de los datos sucede en los datos altamente resumidos, mientras que se recurre muy escasamente a los viejos datos detallados.

Existe una buena razón para que las organizaciones se dirijan al paradigma sugerido en la Figura 14, cual es la utilización de los recursos. A mayor resumen de los datos, es más rápido y eficiente su obtención. Si una organización encuentra que está haciendo mucho procesamiento en los niveles detallados del depósito de datos, entonces necesariamente está incurriendo en un mayor consumo del recurso de procesador. Por lo tanto, el principal interés de todos es procesar al más alto nivel de resumen que sea posible.

En muchas organizaciones, el analista del sistema de soporte a la toma de decisiones (DSS) ha utilizado datos en el nivel detallado en un ambiente de pre-depósito de datos. De alguna forma, acudir a los datos detallados es como una capa de seguridad, aún cuando estén disponibles otros niveles de resúmenes. Una de las tareas que debe cubrir el arquitecto de datos es apartar a los usuarios del sistema de soporte a la toma de decisiones (DSS) de utilizar constantemente datos procedentes

del más bajo nivel de detalle. Para ello existen dos elementos que están a disposición del arquitecto de datos:

- instalar un sistema de cobro, por medio del cual el usuario pague por los recursos consumidos
- señalar que los buenos tiempos de respuesta se pueden alcanzar cuando se trabaja con datos al más alto nivel de resumen, mientras que tiempos de respuesta pobres son producto de trabajar con datos al más bajo nivel de detalle

Otras consideraciones.

Existen algunas otras consideraciones en relación con la construcción y administración de un depósito de datos. La Figura 15 muestra algunas de esas consideraciones.

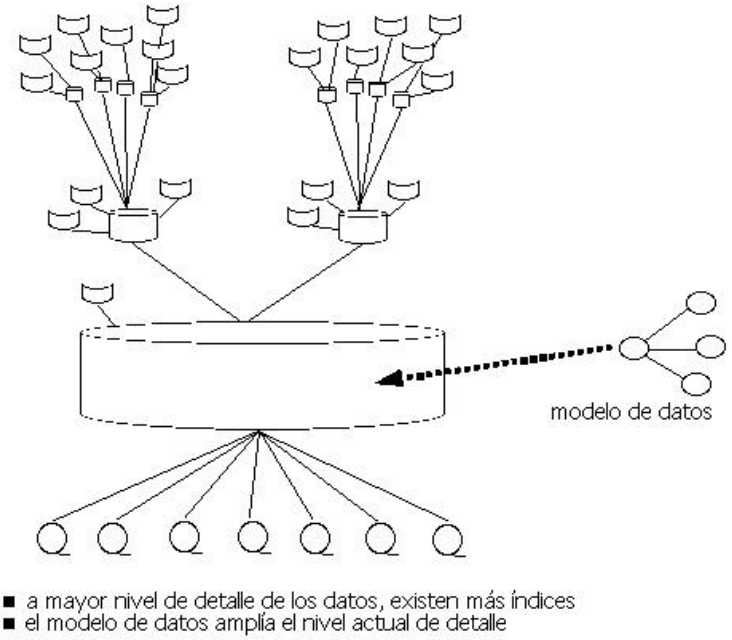


Figura 15. Otras consideraciones a tener en cuenta en el diseño de un depósito de datos.
Fuente: [INMO1995].

La primera de esas consideraciones trata acerca de los índices. Los datos en los más altos niveles de resumen puede que estén libres de índices, mientras que los datos en los más bajos niveles de detalle puede que estén de alguna forma indexados. Pero al mismo tiempo, los datos en los más altos niveles de detalle pueden ser reestructurados relativamente fácil, mientras que el volumen de los datos en los más bajos niveles de detalle es muy grande, de modo que una reestructuración de éstos no puede ser hecha fácilmente.

Consecuentemente, el modelo de datos y el trabajo de diseño formal hecho que se basa en los fundamentos para el depósito de datos aplica casi que exclusivamente para el nivel de detalle actual. En otras palabras, las actividades de modelaje no aplican para los niveles de resumen, en casi ningún caso.

La otra consideración estructural trata acerca del particionamiento de los datos del depósito de datos. La Figura 16 muestra que el nivel de detalle actual está casi siempre particionado.

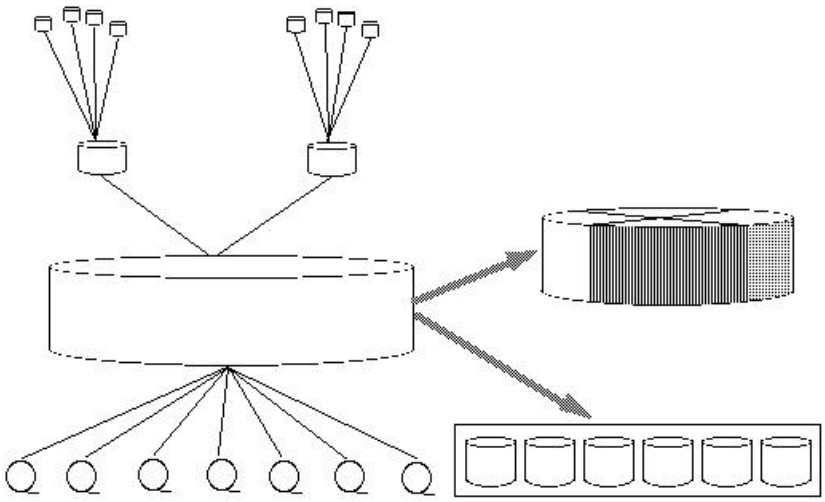


Figura 16. Los datos del nivel detallado casi siempre están particionados.
Fuente: [INMO1995].

Dicha figura muestra como es que el particionamiento se hace de dos formas: a nivel del sistema administrador de bases de datos (SABD), y al nivel de la aplicación. En el particionamiento a nivel del SABD, éste se percata de las particiones y las administra correspondientemente. En el caso del particionamiento al nivel de la aplicación, sólo el programador de la aplicación se percata de ello, y es su responsabilidad administrarlas.

Bajo el particionamiento a nivel del SABD, mucho del trabajo de infraestructura es hecho automáticamente, pero existe un tremendo grado de inflexibilidad asociado con la administración automática de las particiones. En el caso del particionamiento de los datos del depósito de datos al nivel de la aplicación, mucho del trabajo recae en el programador, pero el resultado final es mucho más flexible en materia de administración de los datos del depósito.

Un ejemplo de un depósito de datos.

La Figura 17 muestra un ejemplo hipotético de un depósito de datos estructurado para un ambiente de manufactura.

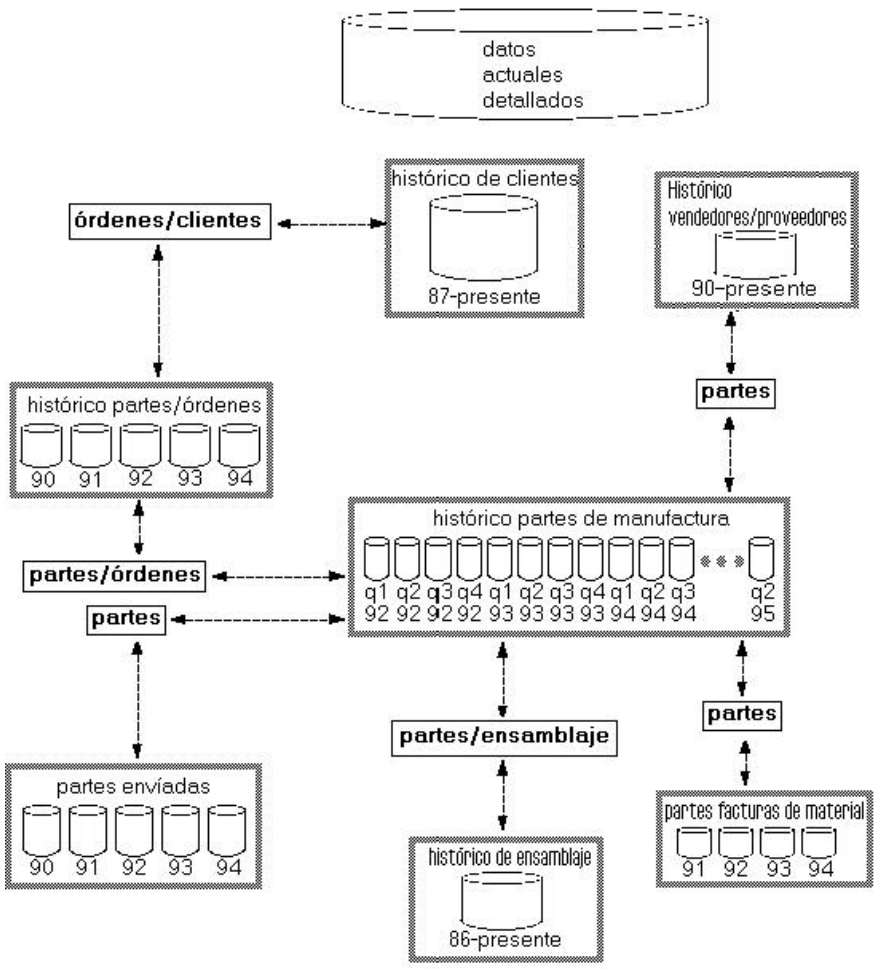


Figura 17. Estructura interna de los datos de un depósito de datos.
Fuente: [INMO1995].

La figura anterior sólo muestra los actuales datos detallados. Los niveles de resumen, así como los viejos datos detallados no se muestran.

En dicha figura se puede apreciar que las tablas del mismo tipo se dividen en el tiempo. Por ejemplo, para el historial de partes manufacturadas existen muchas tablas físicamente separadas, cada una de ellas representando un trimestre diferente. La estructura de los datos es consistente dentro de la tabla del historial de partes manufacturadas, aún cuando físicamente existan muchas tablas que lógicamente comprende el historial.

Nótese que para los diferentes tipos de tablas existen diferentes unidades de tiempo divididas en unidades de datos. El historial de manufactura se divide en trimestres, el historial de partes/órdenes se divide en años y el historial de clientes es un sólo archivo, no se divide de ninguna forma por el tiempo.

Nótese también que las diferentes tablas se ligan por medio de un identificador común, ya sea *partes*, *partes/órdenes*, etc. (Nota: la representación de una relación en el ambiente de depósito de datos toma una forma muy diferente a la de las relaciones representadas en otros ambientes, tal como el ambiente operacional).

Otras anomalías.

Mientras que los componentes de un depósito de datos trabajan de la forma descrita anteriormente para todos los datos, existen una pocas excepciones que vale la pena discutir.

Una de esas excepciones trata acerca del resumen de datos públicos. Un resumen de datos públicos es un resumen de datos que se calcula fuera de los límites del depósito de datos, pero que se utiliza dentro de la organización. El resumen de datos públicos se almacena y administra en el depósito, aún cuando su cálculo se haga fuera de él. Un ejemplo clásico de un resumen de datos públicos es el archivo trimestral que preparan algunas compañías públicas para el SEC (la Comisión de Valores de los Estados Unidos de América, según sus iniciales en inglés). Los contadores trabajan para producir diferentes cálculos como rentas trimestrales, gastos trimestrales, utilidades trimestrales, entre otros. El trabajo realizado por los contadores es externo al depósito de datos. Sin embargo, estos números producidos por los contadores son utilizados ampliamente en la empresa, ya sea por mercadeo, ventas, etc. Una vez que los archivos del SEC se preparan, los datos se almacenan en el depósito de datos.

Otro excepcional tipo de datos que algunas veces se encuentra en un depósito de datos es el de los datos detallados permanentes. Los datos detallados permanentes resultan de la necesidad que tienen las organizaciones de almacenar datos a un nivel de detalle en forma permanente, por razones éticas o legales. Si una empresa expone a sus empleados a sustancias peligrosas, existe una necesidad de mantener datos detallados permanentes. Si una empresa produce bienes que involucran la seguridad pública, tal como partes de avión, existe una necesidad por tener datos detallados permanentes. Si una empresa firma un contrato peligroso, también existe la necesidad de disponer de datos detallados permanentes.

La empresa simplemente no puede perder el detalle puesto que en el futuro (tal vez dentro de algunos años), en el caso de un juicio, un recordatorio, o una disputa por una falla en la construcción, la exposición de la empresa puede ser muy grande. Por lo tanto, existe un único tipo de datos en el depósito de datos que se conoce como datos detallados permanentes.

Los datos detallados permanentes comparten muchas de las mismas consideraciones de los otros datos del depósito, excepto que:

- El medio en que se almacenan los datos debe ser lo más seguro posible
- Los datos deben ser capaces de ser restaurados
- Los datos requieren un tratamiento especial de la indexación de ellos, pues de lo contrario pueden no ser accesibles, a pesar de que están almacenados en una forma segura.

2.2.3. Metadatos.

A continuación se presentan algunas de las definiciones que arroja la literatura consultada en relación con el término metadatos.

Primeramente, se tiene la definición que ofrecen Moriarty y Mandracchia[13], quienes definen los metadatos como: "datos codificados, reglas de derivación, reglas de resumen de datos, reglas de dependencia de datos y reglas para la integración transformación de datos".

También se pudo localizar la definición que da Devlin para este concepto. En efecto, Devlin define los metadatos como: "datos acerca de los datos"[14]. En un sentido más amplio, los metadatos definen y describen completamente el ambiente de sistemas de información, desde su relación con el negocio hasta su estructura técnica. Los metadatos responden preguntas tales como:

- ¿Qué significado tiene este campo en términos del negocio?
- ¿Cuáles procesos empresariales soportan este conjunto de consultas?
- ¿Cuándo se corrió por última vez el trabajo de actualización de los datos de clientes en nuestro mercado de datos?
- ¿Cuál de los archivos contiene los datos de productos?, ¿dónde reside?, ¿cuál es su estructura de detalle?

[13] [MORI1996b] página 70.
[14] [DEVL1998] página 8.

Debido a la variedad de metadatos, es necesario poder categorizarlos en tipos. Es ampliamente aceptado que los datos del negocio se clasifican mejor como operacionales o informacionales, dependiendo de su uso. Los metadatos se pueden clasificar similarmente basándose en cómo se utilicen ellos. Existen tres grandes categorías de metadatos:

a. metadatos de tiempo de construcción: cuando se diseña y construye un depósito se generan este tipo de metadatos. Estos metadatos ligan la terminología del negocio con la de sistemas de información, y describen la estructura técnica de los datos. Este es el tipo más detallado y exacto de metadatos y es utilizado intensamente por diseñadores, desarrolladores y administradores del depósito. Es la fuente primaria de la mayoría de los metadatos utilizados en el depósito.

b. metadatos de utilización: cuando el depósito es puesto en producción, los metadatos de utilización, los cuales se derivan de los metadatos de tiempo de construcción, son una herramienta importante para los usuarios y administradores de los datos. Estos metadatos son utilizados diferentemente a partir de los metadatos de tiempo de construcción y su estructura debe ser acomodada para ello.

c. metadatos de control: el tercer tipo de metadatos es usado, de hecho, por el motor de bases de datos y otras herramientas para administrar sus propias operaciones. Por ejemplo, un SABD construye una representación interna del catálogo de la base de datos para usarla como una copia de trabajo a partir del catálogo creado al momento de construirlo. Esta representación funciona como un metadatos de control. La mayoría de metadatos de control son sólo de interés de los programadores de sistemas. Sin embargo, un subconjunto, el cual se genera y utiliza por las herramientas que pueblan el depósito, es de un interés considerable para los usuarios y administradores del depósito de datos. Éstos ofrecen información vital acerca de la oportunidad del depósito y le permite a los usuarios rastrear la secuencia y momento de los eventos en el depósito.

Por su parte, Gill y Rao[15] manifiestan que por lo general, los metadatos se definen como “datos acerca de los datos”. En una base de datos, los metadatos son la representación de los diversos objetos que definen una base de datos. En una base de datos relacional, está representación consistiría en las definiciones de tablas, columnas, bases de datos, visualización y otros objetos. En un sentido más amplio, se puede usar el término metadatos para hacer referencia a todo lo que defina un objeto del depósito de datos –ya sea una tabla, una columna, un reporte, una consulta, una regla empresarial o una transformación dentro del depósito–. Esta amplia definición de metadatos permite abarcar las definiciones de todos los objetos significativos dentro del depósito.

[15] [GILL1996] páginas 147-149.

Los metadatos constan de los siguientes tipos de datos:

- ubicación y descripción de servidores, bases de datos, tablas, nombres y resúmenes del depósito de datos
- reglas para la profundización automática al detalle o al resumen y a través de jerarquías de dimensión empresarial, tales como productos, mercados y cuadros contables
- nombres elegidos o alias definidos por el usuario final para los encabezados y hechos de datos con nombres más técnicos
- reglas para cálculos personalizados definidos por el usuario final
- seguridad al nivel personal, de grupo de trabajo y de empresa, para visualizar, cambiar y distribuir resúmenes adaptados, cálculos y otros análisis de usuario final
- descripción de fuentes originales y transformaciones
- definiciones lógicas de tablas y atributos del depósito de datos
- definiciones físicas de tablas y columnas, así como de sus características
- ubicación integrada de las tablas del depósito de datos
- antecedentes de extracción
- información de alias
- algoritmos de resumen
- ubicación de área tema
- antecedentes de relaciones
- propiedad/gerencia
- patrones de acceso
- tablas de referencia y datos codificados
- criterios de envejecimiento y purga
- indicador de calidad de datos
- seguridad
- unidades de medida

En forma adicional, los metadatos pudieran también contener componentes de atributos para auxiliar en las siguientes tareas:

- identificación de fuentes operacionales
- ubicación sencilla de atributo a atributo
- conversiones de atributos

- conversiones de características físicas
- conversiones de codificación y tabla de referencia
- cambios de nombre
- cambios de llave
- valores predeterminados que se utilizan
- razón predeterminada
- lógica para elegir entre varias fuentes operacionales
- formulación algorítmica empleada

2.2.4. Clasificación de los depósitos de datos.

Alan Simon[16] clasifica los depósitos de datos en tres categorías. Para ello se ha permitido hacer la semejanza con las pizzas:

- depósito de datos liviano (o *lite*): corresponde a una implementación relativamente modesta, en la cual no se está yendo a los límites tecnológicos.
- depósito de datos de lujo (o *deluxe*): corresponde a una implementación estándar de depósito de datos que hace uso de la tecnología avanzada para resolver las necesidades analíticas y de información complejas del negocio.
- depósito de datos supremo (o *supreme*): corresponde al depósito de datos del futuro, con el uso de tecnología de punta, distribución de datos a gran escala y otras características no convencionales.

En la siguiente tabla se resumen las características más importantes que este autor señala como puntos distintivos de las categorías enunciadas por él.

[16] [SIMO1997b] páginas 34-49.

	liviano	**de lujo**	**supremo**
áreas de interés	se enfoca al análisis de una o dos áreas	un amplio rango de áreas de interés relacionadas	ilimitado
Fuentes de datos	pocas	es común encontrar en promedio de 8 a 10 aplicaciones y bases de datos externas	numerosas fuentes de datos
base de datos	SABDR (sistema administrador de bases de datos relacional), a lo sumo un BDM (base de datos multidimensional)	base de datos relacional del orden de los gigabytes, tendiendo a los terabytes, y en incremento.	una que cumpla con las siguientes características: • distribuida a lo largo de muchas plataformas • debe operar de una forma transparente a la localización • capacidad de orientación a objetos • debe permitir el acceso directo desde las bases de datos transaccionales sin tener que copiar la información en un depósito separado
datos contenidos	contenido limitado	necesita tener capacidades de acceso adicionales a los simples resultados de reporte	todos los diferentes tipos de datos que se necesiten para soportar el mejor proceso de toma de decisiones
herramientas	por lo general, poco complejas y fáciles de usar	reportes y consultas simples, OLAP, EIS, minería de datos	además de las mencionadas en el caso de la deluxe, los agentes inteligentes tendrán un papel importante
extracción, movimiento y carga de datos	lo más simple posible	sumamente complejo	habrá un mayor tráfico de mensajes entre las fuentes de datos y la base de datos del depósito. La fuente de datos determinará que datos deben ser movidos al depósito en vez de que sea el depósito quien tenga la responsabilidad por solicitar las actualizaciones y adiciones.
arquitectura	un SABD utilizado para almacenar los datos, herramientas de usuario final (*front-end*) para accesar los datos y el medio para mover los datos	además de los componentes mencionados en la lite, se deben considerar mercados de datos, estaciones de transformación y estaciones de aseguramiento de calidad de los datos	depósito de datos físico, vistas lógicas de los contenidos, herramientas de usuario final, servicios de acceso a los datos, entre otros componentes.

Tabla 1. Características más relevantes de las categorías de depósitos de datos propuestas por Alan Simon.

2.2.5. Mutaciones del concepto de depósitos de datos.

William Inmon[17] mantiene la tesis de que existen algunas notables formas mutantes de un depósito de datos. Una de ellas son los almacenes de datos operacionales (en inglés *Operational Data Store* – ODS). La otra lo son los mercados de datos (del inglés *Data Mart* – DM).

2.2.5.1. Almacén de datos operacionales (*Operational Data Store*, ODS).

Inmon indica que "un almacén de datos comparte muchas de las características de un depósito de datos, no obstante existe una que le distingue: un ODS puede ser actualizado y proveer un rápido tiempo de respuesta transaccional cosa que un depósito de datos clásico no puede hacer"[18].

Según Alan Simon: "Un ODS es un ambiente informacional y analítico que refleja en cualquier momento el estado operacional actual de sus principales temas, aún cuando los datos que hacen ese estado operacional se administren en diferentes aplicaciones dondequiera que sea dentro de la empresa"[19]. Ello significa:

- ambiente informacional y analítico: la interfaz de usuario y el comportamiento de un ODS se ve y se siente como un depósito de datos. Esto es, un usuario de un ODS tiene una herramienta para la generación de consultas y reportes, una herramienta OLAP, o posiblemente una herramienta EIS a través de la cual se solicita y envía la información y el análisis. (La minería de datos no es aplicable en un ODS).
- refleja en cualquier momento el estado operacional actual: al tomar una herramienta de interrogación y hacerle una consulta al ODS la respuesta que se recibe refleja los datos tal y como están actualmente almacenados en cualquiera de los sistemas operacionales de donde provengan éstos. Si ocurriera una actualización en uno de los sistemas operacionales, el ODS debe hacer el mismo cambio en tiempo real, o casi en tiempo real (o sea, muy rápidamente). En la mayoría de las ocasiones, sin embargo, no funciona extraer datos orientados a lotes para incluirlos en un ODS.
- principales temas: tal y como sucede con un depósito de datos, un ODS se crea con una misión de negocio específica en la mente para manejar un conjunto de áreas de interés.

[17] [INMO1998] página 88NA 1.
[18] [INMO1998] página 88NA 6.
[19] [SIMO1997b] página 242-243.

- datos administrados en diferentes aplicaciones dondequiera que sea dentro de la empresa: un ODS no es una simple base de datos unificada que utilizan un número de aplicaciones. Por el contrario, es una base de datos separada que recibe información de varias fuentes con las transformaciones apropiadas, aseguramiento de calidad y otros procesos.

José Paz[20], en su tesis de graduación cita dos definiciones de almacenes de datos operacionales que vale la pena transcribir:

a. la que da la revista Datamation: "... una base de datos integrada de los datos operacionales, cuyas fuentes incluyen viejos sistemas transaccionales y que contiene datos actuales o de muy corto plazo. Un almacén de datos operacional puede contener de 30 a 60 días de información".

b. la que ofrece Inmon: "... una edificación arquitectónica: orientada a temas, integrada (colectivamente integrada), volátil, con valores actuales, y que contiene sólo datos detallados de la organización".

Hammergren define un ODS como: "la base de datos asociada a un sistema operacional. Estas bases de datos se requieren típicamente para soportar el gran número de transacciones sobre una base diaria con la mínima latencia para el procesamiento de las transacciones"[21].

Finalmente, Singh ofrece la siguiente definición para ODS: "Es una base de datos integrada de datos operacionales, conteniendo información integrada, actual y oportunamente. Los datos son típicamente muy granulares. Estos sistemas son orientados a temas, y no a las aplicaciones, y son optimizados para buscar uno o dos registros a la vez para tomar decisiones. Sus fuentes incluyen sistemas heredados y contienen datos actuales o muy cercanos. Un ODS puede contener información de 30 a 60 días, mientras que un depósito de datos típicamente contiene años de datos"[22].

2.2.5.2. Mercado de datos (*Data Mart*, DM).

Inmon[23] menciona que la otra forma mutante de un depósito de datos es un mercado de datos. Un mercado de datos es una forma departamentalizada de un depósito de datos. Éstos son semejantes a los depósitos de datos, excepto porque los mercados de datos:

[20] [PAZ1999] página 19.
[21] [HAMM1996] página 452.
[22] [SING1999] página 447.
[23] [INMO1998] página 88NA 6.

- son configurados para un único departamento
- contienen menos información histórica que un depósito de datos
- operan con una tecnología que se ubica en un volumen de datos mucho menor que el de un depósito de datos
- tiene mucho más índices que un depósito, de modo que la estructura de datos óptima para un mercado de datos es la join * (o *star join* en inglés), puesto que para un depósito de datos más voluminoso es apropiada una estructura más normalizada
- tiene un patrón de utilización moderadamente predecible
- contiene mucho más datos resumidos
- luce bastante diferente de un departamento a otro
- retiene muy pocos, si acaso, datos detallados.

Según Alan Simon: "Un mercado de datos es un depósito de datos de menor escala"[24].

En este mismo sentido, Singh define un mercado de datos como: "Un subconjunto de una fuente de datos, usualmente orientada a un propósito específico o al tema principal de los datos, que puede estar distribuido para apoyar las necesidades del negocio. El concepto de mercado de datos puede aplicar a cualquier dato sea éste un dato operacional, un dato de evaluación, un dato espacial o un metadato"[25].

Por su parte, Gill y Rao definen un mercado de datos como: "una implementación de depósito de datos con un ámbito de datos y funciones de depósito más pequeño y reducido, que sirve a un departamento único o una parte de una organización. Una organización generalmente tiene varios mercados de datos"[26].

A su vez, Hammergren define un mercado de datos como: "un componente de un depósito de datos que se enfoca típicamente a un paquete de información o proceso del negocio específico. Un mercado de datos sigue los mismos principios de un depósito de datos completo, sin embargo, no son tan amplios"[27].

[24] [SIMO1997b] página 52.
[25] [SING1999] página 424.
[26] [GILL1996] página 358.
[27] [HAMM1996] página 445.

2.2.6. Ciclo de desarrollo de un depósito de datos.

Para la creación de un depósito de datos se deben seguir una serie de pasos o fases, todas las cuales conducen al desarrollo e implantación exitosa del producto final.

En este sentido, existen diferentes propuestas de cuáles deben ser esas fases. A continuación se presentan algunas de esas propuestas.

2.2.6.1. Propuesta metodológica de Gil y Rao.

Gill y Rao[28] proponen una estrategia de desarrollo de un depósito de datos que se componen de etapas, cada de una de las cuales tiene pasos a seguir, de la forma como se detalla de seguido:

Fase de planeación

- Selección de la estrategia de implementación
- Selección de la metodología de desarrollo
- Selección de ámbito de implementación
- Selección del enfoque arquitectónico
- Desarrollo de un programa y del presupuesto de proyecto
- Desarrollo de escenarios de uso empresarial
- Recopilación de metadatos

Fase de requerimientos

- Definir los requerimientos del propietario
- Definir los requerimientos del arquitecto
- Definir los requerimientos del desarrollador
- Definir los requerimientos del usuario final

Fase de análisis

- Requerimientos de enfoque empresarial que delinean las fronteras de la información que debe comprender el depósito de datos
- Especificaciones de requerimientos de fuentes de datos que delinean las fronteras de información disponibles en las fuentes de datos actuales
- Especificaciones de requerimientos de usuario final y acceso, las cuales definen cómo se utilizará la información del depósito de datos

[28] [GILL1996] páginas 77-109.

Fase de diseño

- Diseño detallado de la arquitectura de datos
- Diseño detallado de la arquitectura de aplicación

Fase de construcción

- Programas que creen y modifiquen las bases de datos para el depósito de datos y los mercados de datos
- Programas que extraigan datos de fuentes relacionales y no relacionales
- Programas que realicen transformaciones de datos, tales como integración, resumen y adición
- Programas que realicen actualizaciones de bases de datos relacionales
- Programas que efectúen búsquedas en bases de datos muy grandes

Fase de despliegue

- Proporcionar la instalación inicial, incluyendo facilidades para las conexiones básicas de datos con las fuentes y para la actualización y sincronización de datos
- Planeación y entrega de una implementación por etapas
- Proporcionar capacitación y orientación a todo tipo de usuarios
- Planeación e implementación de la actualización de plataformas y el mantenimiento necesario por el depósito de datos cuando se requiere
- Proporcionar la administración de usuarios y sistemas
- Proporcionar la capacidad de generar archivos permanentes y respaldos
- Proporcionar la capacidad de recuperación
- Asegurar la integración dentro de la infraestructura existente
- Proporcionar controles de acceso y seguridad
- Asegurar la completa disponibilidad y los procesos para manejar caídas de los sistemas y sus componentes de infraestructura

Fase de expansión

- Mejoras en las consultas empresariales que no pudieran formularse o satisfacerse debido a las limitaciones impuestas por el metamodelo del depósito de datos
- Mejoras en las consultas empresariales que comprendieran fuentes de datos externos que no formaron parte de la implementación inicial
- Mejoras en un desempeño no satisfactorio de componentes clave del depósito de datos
- Atender lo concerniente a que otros departamentos no querían configurar sus propios mercados de datos, haciendo necesario incrementar el ámbito del metamodelo del depósito de datos.

2.2.6.2. Propuesta metodológica de Jiménez y Ávalos.

Por su parte, Jiménez y Ávalos[29] proponen como estrategia de desarrollo de un depósito de datos la siguiente:

Etapa 1 Fundamentación

1. Definir las fronteras del sistema.
2. Identificar los objetivos empresariales del depósito de datos.
3. Identificar las consultas potenciales.
4. Identificar las fuentes de datos potenciales.
5. Pronosticar los volúmenes iniciales.
6. Analizar costo beneficio.
7. Obtener la aprobación del presupuesto.
8. Desarrollar un plan de contingencias.
9. Suministrar capacitación.

Etapa 2 Modelaje

1. Estudio detallado de las fuentes de información.
2. Realizar el análisis empresarial.
3. Analizar la arquitectura tecnológica de la organización.
4. Desarrollo de un modelo conceptual.
   - Identificar los tipos de entidades.
   - Identificar las asociaciones.
   - Identificar los atributos.
   - Identificar el dominio.
   - Estimar el tamaño de las fuentes de datos.
5. Diseño del depósito de datos.
   - Identificar la tabla de hechos.
   - Identificar las dimensiones requeridas.
   - Determinar el esquema de agregación.
   - Identificar las dimensiones cambiantes en el tiempo.
   - Identificar las llaves.
6. Diseño físico del depósito de datos.
   - Evaluar las técnicas de indexado.
   - Evaluar el sistema de consultas.
   - Definir y evaluar el sistema de acceso.
   - Elaborar el esquema físico.
7. Definir los componentes del depósito de datos.
   - Identificar los componentes de extracción.
   - Identificar los componentes de limpieza.

[29] [JIME1998b] página 55.

- Identificar los componentes de agregación.
- Identificar el intérprete de consultas.
- Identificar los metadatos.
- Análisis de componentes.

Etapa 3 Preparación e Implementación

1. Preparación
   - Establecer el equipo de apoyo al sistema
   - Planificar el proyecto de implementación
   - Establecer el plan de capacitación
   - Establecer el entorno de implementación
   - Ejecutar la carga preliminar del depósito de datos
2. Implementación
   - Ajustar hardware y software
   - Suspender el procesamiento actual
   - Ejecutar rutinas de conversión
3. Evaluación
   - Dirigir el proceso de implementación
   - Retroalimentación con los resultados obtenidos

Etapa 4 Operación

1. Suministrar asistencia a los usuarios para el funcionamiento del sistema
2. Suministrar asistencia operacional
3. Análisis del rendimiento del sistema
4. Afinamiento del depósito de datos
5. Evaluación post-implementación

2.2.6.3. Propuesta metodológica de Gamboa y Sánchez.

Gamboa y Sánchez[30] proponen una estrategia para el desarrollo de un depósito de datos de la siguiente forma:

Fase 1. Estudio de factibilidad

1. Evaluar factibilidad operativa
2. Evaluar factibilidad técnica
3. Evaluar factibilidad financiera y económica

Fase 2. Análisis de requerimientos

[30] [GAMB1998] páginas 123-146.

1. Entrevistar a usuarios que harán uso del depósito
2. Convertir los requerimientos recopilados en el paso anterior en un conjunto de especificaciones que puedan apoyar al diseño
3. Valorar las fuentes datos (sistemas operacionales corporativos, sistemas operacionales departamentales, fuentes externas)

Fase 3. Diseño de la arquitectura técnica

1. Entrevistar proveedores de hardware y software
2. Identificar la mejor de las tecnologías existentes en el mercado
3. Considerar que la infraestructura tecnológica que va a soportar el depósito de datos debe estar separada y ser diferente de la que soporta los sistemas operacionales
4. Retroalimentar las conclusiones iniciales, esperando mayores revisiones como resultado de este paso, porque aquí es donde el tamaño y costo verdadero del depósito de datos son claramente definidos
5. Refinar y volver a trabajar con las conclusiones iniciales
6. Documentar los pasos anteriores

Fase 4. Diseño y construcción de la estructura de la base de datos

1. Diseño físico de la base de datos. Convertir el modelo lógico a un diseño detallado de la base de datos
2. Creación de la base de datos física. Crear físicamente la base de datos, las particiones de las tablas y discos, entre otros
3. Diseño y construcción de las guías para generar los índices
4. Diseño y construcción de las guías para generar las tablas de resumen. Estas guías pueden ser diseñadas para generar las agregaciones de las definiciones de los metadatos en el depósito de datos
5. Diseño y construcción de las guías para crear las vistas de la organización

Fase 5. Diseño y construcción del proceso de administración de carga

1. Diseño y construcción del administrador de carga en la plataforma del depósito de datos
2. Diseño y construcción del controlador del proceso en el sistema fuente

Fase 6. Extracción y carga de los datos

1. Capturar información de datos externos para refrescar la información del depósito de dato, así como permitir la conexión a las estructuras de datos de dichas fuentes para mayor facilidad
2. Limpiar y completar los datos capturados: reestructurar campos, quitar ciertos datos no significativos, añadir y traducir campos, comprobar consistencia e integridad de contenidos y detectar errores, calcular

campos derivados y sumarizaciones, mezclar y operar sobre datos de distintas fuentes
3. Transportar los datos desde los sistemas origen al depósito de datos, utilizando la herramientas necesarias
4. Diseñar y construir la transferencia al nuevo hardware
5. Diseñar y construir la carga en el área temporal

Fase 7. Validación y limpieza de los datos

1. Diseño y construcción de las funciones de validación de los datos

Fase 8. Diseño y construcción de la infraestructura operacional

1. Diseño y construcción de los mecanismos de respaldo y recuperación
2. Diseño y construcción de los mecanismos de archivo
3. Diseño y construcción de los mecanismos de control de acceso
4. Instalación del software para respaldo y recuperación

Fase 9. Diseño y construcción del administrador del depósito

1. Diseño y construcción del administrador del depósito de datos
2. Diseño y construcción de las tablas de los metadatos e interfaz

Fase 10. Implantación

1. Ajuste del sistema operativo
2. Ajuste de la base de datos
3. Ajuste de las consultas
4. Ajuste de la herramienta de acceso del usuario
5. Ejecutar la carga de los datos iniciales
6. Soporte a la entrada en producción

2.2.6.4. Propuesta metodológica de Kimball et al.

Kimball y otros autores[31] proponen un ciclo de vida de los depósitos de datos en los siguiente términos.

Fase 1: Administración del proyecto y requerimientos.

Definición del proyecto

1. Evaluar la disposición para el depósito de datos
2. Desarrollar el alcance preliminar del proyecto
3. Construir la justificación del negocio

[31] [KIMB1998] páginas 737-746.

Planeación y administración del proyecto

1. Establecer la identidad del proyecto
2. Identificar los recursos del proyecto
3. Preparar el borrador del plan del proyecto
4. Conducir el arranque del equipo del proyecto y de la planeación
5. Revisar el plan del proyecto
6. Desarrollar el plan de comunicación del proyecto
7. Desarrollar el programa para la medición de los logros
8. Desarrollar el programa para administrar el alcance
9. Dar seguimiento a la administración del proyecto

Definición de los requerimientos del usuario

1. Identificar y preparar al equipo de entrevista
2. Seleccionar a los entrevistados
3. Calendarizar las entrevistas
4. Manejar el inicio de los usuarios y preparar a los entrevistados
5. Conducir las entrevistas de los usuarios del negocio
6. Conducir las entrevistas de los descubridores de datos
7. Analizar los hallazgos de las entrevistas
8. Documentar los hallazgos y revisarlos
9. Publicar los entregables de los requerimientos
10. Priorizar y revisar el alcance del proyecto
11. Aceptación del usuario/Revisión del proyecto

Fase 2: Diseño de datos

Modelamiento dimensional

1. Construir la matriz
2. Escoger el mercado de datos
3. Declarar la granularidad
4. Escoger las dimensiones
5. Desarrollar el diagrama de la tabla de hechos
6. Documentar el detalle de la tabla de hechos
7. Diseñar el detalle de dimensión
8. Desarrollar la hoja de trabajo de hechos derivados
9. Revisión y aceptación del usuario
10. Revisar las recomendaciones de diseño de bases de datos para herramientas de usuario final
11. Revisar las recomendaciones de diseño de bases de datos para el SABD
12. Completar el diseño lógico de la base de datos
13. Identificar candidatos de agregación prealmacenados
14. Desarrollar una estrategia para el diseño de las tablas de agregación

15. Revisar el diseño lógico de la base de datos con el equipo de trabajo
16. Certificar el diseño de la base de datos con el vendedor de las herramientas para el soporte a toma de decisiones (DSS)
17. Aceptación del usuario/Revisión del proyecto

Análisis de las fuentes de datos

1. Identificar fuentes de datos candidatas
2. Navegar por el contenido de los datos
3. Desarrollar el mapa de datos desde la fuente hasta el destino
4. Estimar el número de filas
5. Aceptación del usuario/Revisión del proyecto

Fase 3: Arquitectura

Diseño de la arquitectura técnica

1. Crear la fuerza de trabajo de arquitectura
2. Recolectar y documentar los requerimientos técnicos
3. Revisar el ambiente técnico actual
4. Crear el plan de arquitectura
5. Determinar la tentativa de implementación por fases
6. Crear el plan de infraestructura
7. Desarrollar la recomendaciones de configuración
8. Aceptación del usuario/Revisión del proyecto

Implementar medidas de seguridad tácticas

1. Desarrollar plan de seguridad táctico
2. Asegurar el ambiente físico
3. Instalar software para el chequeo de virus
4. Asegurar el acceso dentro el ambiente
5. Asegurar el acceso fuera del ambiente
6. Implementar un esquema riguroso de contraseñas
7. Implementar controles para la instalación de software
8. Auditar las violaciones de seguridad
9. Monitorear los privilegios de seguridad en forma individual
10. Aceptación del usuario/Revisión del proyecto

Desarrollar el plan de seguridad estratégico

1. Diseñar la arquitectura de seguridad
2. Implementar los *token* de acceso
3. Implementar esquema de llaves pública/privada para autenticación
4. Implementar túneles seguros para acceso remoto
5. Centralizar la autenticación y el control de acceso
6. Implementar certificados firmados para la descarga de software

7. Aceptación del usuario/Revisión del proyecto

Selección de productos

1. Desarrollar matriz de evaluación
2. Investigar productos candidatos
3. Desarrollar lista corta de productos
4. Evaluar opciones de los productos
5. Opcional para prototipos
   - Seleccionar proceso del negocio
   - Definir criterios de aceptación
   - Adquirir recursos (internos y del vendedor)
   - Determinar la configuración de prueba
   - Instalar prerrequisitos y componentes de evaluación
   - Entrenar al equipo de evaluación
   - Desarrollar y afinar el prototipo
   - Conducir las pruebas
   - Analizar y documentar los resultados
6. Determinar la recomendación del producto
7. Presentar los hallazgos/resultados a la administración
8. Negociar el contrato
9. Aceptación del usuario/Revisión del proyecto

Instalación del producto

1. Planeación de la instalación
2. Satisfacer los prerrequisitos
3. Instalar el hardware/software
4. Probar el hardware/software
5. Aceptación del usuario/Revisión del proyecto

Fase 4: Implementación

Diseño físico de la base de datos

1. Definir estándares
2. Diseñar las tablas físicas y columnas
3. Estimar el tamaño de la base de datos
4. Desarrollar el plan inicial de índices
5. Desarrollar el plan inicial de agregación
6. Desarrollar el plan inicial de particionamiento
7. Aceptación del usuario/Revisión del proyecto

Implementación de la base de datos física

1. Determinar los parámetros fijos del SABD
2. Instalar el SABD
3. Optimizar los parámetros ajustables del SABD
4. Construir la estructura de almacenamiento físico
5. Configurar el RAID

6. Completar el dimensionamiento de tablas e índices
7. Crear las tablas e índices
8. Aceptación del usuario/Revisión del proyecto

Diseñar y desarrollar el escenario de datos

1. Diseñar los procesos de escenificación de alto nivel
2. Desarrollar el plan detallado de escenificación por tabla
3. Establecer el ambiente de desarrollo
4. Definir e implementar los metadatos de escenificación
5. Desarrollar el proceso estático para la primera dimensión de tablas (extracción, transformación y carga)
6. Desarrollar el proceso de mantenimiento para la primera dimensión
7. Desarrollar los procesos para las tablas de las restantes dimensiones
8. Desarrollar los procesos para las tablas de hechos (extracción, transformación y carga)
9. Desarrollar el proceso incremental para la tabla de hechos
10. Desarrollar e implementar la limpieza de los datos
11. Desarrollar el proceso de agregación
12. Automatizar todo el proceso
13. Desarrollar procesos para el aseguramiento de la calidad de los datos
14. Implementar la administración de la base de datos (archivo, respaldo y recuperación)
15. Aceptación del usuario/Revisión del proyecto

Poblar y validar la base de datos

1. Establecer el ambiente de producción
2. Carga inicial de los datos de prueba
3. Validación de datos/Aseguramiento de la calidad inicial
4. Carga de datos históricos
5. Realización de la validación de datos/aseguramiento de la calidad
6. Aceptación del usuario/Revisión

Realizar afinamiento

1. Establecer consultas para *benchmark*
2. Revisar indización y agregación
3. Revisar la herramienta para el afinamiento específico
4. Dirigir el inicio del monitoreo de la base de datos
5. Aceptación del usuario/Revisión del proyecto

Especificación de aplicación de usuario final

1. Identificar y priorizar reportes candidatos
2. Diseñar la plantilla de navegación tentativa
3. Desarrollar los estándares de aplicación de usuario final

4. Documentar las especificaciones detalladas de la plantilla
5. Revisar las especificaciones de la aplicación de usuario final con los usuarios
6. Revisar las especificaciones de la aplicación de usuario final
7. Revisar el alcance del proyecto
8. Aceptación del usuario/Revisión del proyecto

Desarrollo de la aplicación de usuario final

1. Seleccionar la implementación tentativa
2. Revisar las especificaciones de la aplicación
3. Revisar los estándares de la aplicación
4. Poblar los metadatos de la herramienta de usuario final
5. Desarrollar las aplicaciones de usuario final
6. Proveer precisión en los datos y nitidez
7. Desarrollar la navegación de usuario final
8. Revisar con los usuarios
9. Documentar las aplicaciones de usuario final
10. Desarrollar procedimientos para el mantenimiento de las aplicaciones de usuario final
11. Desarrollar procedimientos para la liberación de aplicaciones de usuario final
12. Aceptación del usuario/Revisión del proyecto

Fase 5: Despliegue y crecimiento

Planeamiento del despliegue

1. Desarrollo de la lista de chequeo de la infraestructura de escritorio
2. Desarrollo de la estrategia inicial de entrenamiento para el usuario
3. Definir la estrategia de soporte al usuario
4. Definir el plan de liberación
5. Revisar la estrategia de despliegue y el plan de liberación
6. Desarrollar el material para el curso a los usuarios
7. Desarrollo de los procedimientos de soporte
8. Aceptación del usuario/Revisión del proyecto

Prueba completa del sistema

1. Correr el proceso para la escenificación completa de los datos
2. Ejecutar los procedimientos de aseguramiento de la calidad
3. Correr el núcleo de las aplicaciones de usuario final
4. Revisar la totalidad de los procesos
5. Aceptación del usuario/Revisión del proyecto

Despliegue (versiones alfa, beta y producción)

1. Valorar la preparación para el despliegue

2. Configurar y probar la infraestructura de escritorio
3. Establecer los privilegios de seguridad
4. Entrenar a los usuarios
5. Aceptación del usuario/Revisión del proyecto

Mantenimiento del depósito

1. Proveer soporte continuo a los usuarios
2. Proveer entrenamiento continuo a los usuarios
3. Mantener la infraestructura técnica
4. Monitorear el desempeño de las consultas de usuario final
5. Monitorear el desempeño de la escenificación de los datos
6. Comunicarle continuamente al mercado los éxitos logrados
7. Aceptación del usuario/Revisión del proyecto

Crecimiento del depósito de datos

1. Establecer un comité gerencial para el depósito de datos
2. Establecer una estrategia para la priorización del mejoramiento
3. Utilizar iterativamente el ciclo de vida dimensional de los negocios

2.2.6.5. Análisis de las propuestas.

Dada la naturaleza de la presente investigación resultan de interés aquellas actividades que estén relacionados con el tema de definición de aspectos de seguridad dentro del proceso de desarrollo de los depósitos de datos.

Sobre el particular, en la propuesta de Gil y Rao se aprecia que en la Fase de Análisis hay una actividad denominada "*Especificaciones de requerimientos de usuario final y acceso, las cuales definen cómo se utilizará la información del depósito de datos*". Asimismo, en la Fase de Despliegue se tienen dos actividades más relacionadas con el tema en cuestión, "*Proporcionar la administración de usuarios y sistemas*" y "*Proporcionar controles de acceso y seguridad*".

A pesar del poco detalle que ofrece la metodología sugerida por estos autores, se tiene que es en la Fase de Análisis donde primero se trata el tema de la seguridad y el acceso a los datos, sin que se aprecie que tales consideraciones sean abordadas, al menos de forma explícita, en la Fase de Construcción, y no es sino hasta la Fase de Despliegue que se retoma el tema de la seguridad, situación que permite concluir que desde la perspectiva de ellos la seguridad no es un tema de importancia tal que amerite un tratamiento especial.

En lo que respecta a la metodología sugerida por Jiménez y Ávalos, tan sólo se aprecia una sola actividad en la que se trate el tema de la seguridad, en concreto, la actividad "*Definir y evaluar el sistema de acceso*" que forma parte del paso 5. Diseño físico del depósito de datos, de la Etapa 2. Modelaje.

Esta situación lo que conduce a concluir es que el tema de la seguridad tampoco es relevante en esta metodología, en la que se le presta una mayor atención a aspectos tales como la fundamentación del proyecto, el modelaje del depósito desde la perspectiva de los requerimientos de la empresa (sin preocuparse por temas de seguridad y acceso a los datos), la preparación e implementación del depósito y finalmente, la operación de éste. Además, al igual que se comentara anteriormente, no parece tener sentido llegar a abordar el tema de la seguridad en una etapa de modelaje sin que previamente se hayan definido criterios al respecto.

En cuanto a la metodología de Gamboa y Sánchez, solamente en la Fase 8. Diseño y construcción de la infraestructura operacional, es que se define una actividad "*3. Diseño y construcción de los mecanismos de control de acceso*" en la cual se trata el tema de la seguridad del depósito de datos. Salvo que en la Fase 2. Análisis de requerimientos la actividad denominada "*2. Convertir los requerimientos recopilados en el paso anterior en un conjunto de especificaciones que puedan apoyar al diseño*" haya considerado lo referente al tema de la seguridad, como para que sirva de soporte a la actividad de la Fase 8 ya mencionada, nuevamente se puede concluir que el tema de la seguridad es tratado por esta metodología de una forma tangencial.

Es quizás la metodología propuesta por Kimball y otros la que mejor trata el tema de la seguridad en los depósitos de datos, de las metodologías aquí analizadas. Sin embargo, su enfoque es muy puntual, y sólo ve el tema como parte de la fase de arquitectura del depósito. De hecho, a pesar de que se dedica una porción considerable de las tareas que componen esta fase, no se aprecia que ello esté sustentado, por lo menos de una forma expresa y clara, en el trabajo o productos de las fases previas. Es loable reconocer que esta metodología incorpora actividades interesantes, como lo son "Implementar medidas de seguridad tácticas" o "Desarrollar un plan de seguridad estratégico"; sin embargo, no resulta claro cuál es el sustento para la realización de esas labores, sobre todo porque en la fase de "Administración del proyecto y requerimientos" no es explícito que entre éstos se consideren aspectos de seguridad del depósito.

Dado lo anteriormente expuesto, se puede concluir que, al menos, con las metodologías que se tuvo a la mano, y que se mostraron en este apartado, no se percibe la existencia de alguna que dé un tratamiento integral y metódico al tema de la seguridad en los depósitos de datos.

2.2.7. Limpieza y calidad de los datos del depósito.

Los temas limpieza y calidad de los datos están sumamente entrelazados como para poder referirse a uno de ellos sin que necesariamente se haga mención del otro.

A pesar de ello, a continuación se presentan dos secciones en donde se ha procurado abordar estos temas lo más independientemente posible uno del otro.

2.2.7.1. Limpieza de los datos.

El problema de la calidad en los datos no es reciente. Desde hace décadas se viene hablando de los "datos sucios" (en inglés, *dirty data*), y de hecho se ha tenido que vivir con este problema.

Sin embargo, tal y como lo hace ver Larissa Moss, esta situación retoma importancia en estos tiempos pues para poner a funcionar un depósito de datos se ha prometido que "los datos estarán limpios, integrados e históricos en un corto tiempo y a un bajo costo"[32], a pesar de que ha resultado difícil tratar con ellos. Algunos depósitos han fracasado debido a que las promesas de "limpios", "integrados" e "históricos" no se han podido cumplir. Por su parte, otros depósitos han fallado pues las promesas de "en corto tiempo" y a "bajo costo" se han excedido al intentar depurar los datos. En pocas palabras esos depósitos han fallado a causa de la dicotomía de la promesa hecha.

Entre las razones que se apuntan para que se produzcan los datos sucios se tienen las siguientes:

- Valores *dummy*: se refiere a la práctica de suministrar como un valor válido una repetición de números todos iguales, ya sea por desconocimiento del dato real, o bien, por no dejar el valor en blanco dado que la aplicación lo solicita; y los cuales en ese momento no tenían sentido, pero que con el tiempo ese valor llega a tener sentido.

[32] [MOSS1998].

Ejemplo de ello es el caso de indicar como un año 88 ó 99 antes de que esas fechas llegaren a ser reales.

- Ausencia de datos: se refiere a que algunas veces, a pesar de que las aplicaciones solicitaban ciertos datos, éstos no eran suministrados y por tanto no se almacenaba ningún valor en la variable correspondiente.
- Campos de propósitos múltiples: se refiere al caso en que un mismo campo se utilizó para diferentes propósitos, algunas veces ello se lograba a través de la redefinición del campo en cuestión. Este tipo de situaciones creaba trastornos sobre todo porque al realizar redefiniciones sucesivas, en función de los diferentes propósitos que se querían satisfacer, muchas veces ello se hacía sin la documentación debida, y por lo general se olvidaba el detalle necesario para interpretar los eventuales valores que se generaban en cada una de estas redefiniciones.
- Datos codificados: se refiere a la práctica de ahorrar espacio al codificar los posibles valores en unas pocas letras, lo que con el pasar del tiempo hacía que al requerirse nuevos valores se crearan nuevos códigos sin analizar la relación que pudiera existir entre los códigos ya definidos y los nuevos.
- Datos contradictorios: se refiere a la posibilidad que existió de que los datos fueran ingresados sin ninguna validación que hiciera evidente una inconsistencia en cuanto a los datos suministrados.
- Violación de las reglas del negocio: básicamente se refiere a aquellos datos que fueron ingresados incumpliendo las reglas del negocio imperantes al momento de tal acción. Ejemplo de ello puede ser encontrar tasas de interés sin un sentido real, de acuerdo con las políticas de la empresa, en un sistema de préstamos, o de inversiones a plazo.
- Llaves primarias reutilizadas: se refiere a que generalmente los sistemas transaccionales no conservan información histórica más allá de los 90 ó 180 días, existe la posibilidad de que se reutilicen las llaves primarias. El problema viene cuando se intenta recopilar la información histórica para el depósito de datos y se determina que una misma llave primaria, en el tiempo, ha servido para designar entidades u objetos totalmente diferentes.
- Identificadores no únicos: se refiere a que una misma entidad u objeto tenga diferentes identificadores en sistemas diferentes. Ejemplo de ello puede ser que un mismo cliente puede ser identificado por diferentes números, ya sea por su número de cuenta corriente, por número de tarjeta de crédito, por número de tarjeta de débito, etc.

- Problemas de integración de datos: se refiere principalmente a los problemas que se generan al integrar los datos. Ello se puede dar porque existen datos que deben ser relacionados y no se puede, o bien, que existen datos que inadvertidamente se les relacionó y no debiera ser así.

Ante el panorama ya descrito cabe entonces preguntarse si se lleva a cabo la limpieza o no. Para responder a esta interrogante, es necesario considerar cuatro aspectos:

- se puede hacer la limpieza de los datos
- se debieran limpiar los datos
- en qué lugar se va a realizar la limpieza de los datos
- se puede realizar una limpieza razonable de los datos

En torno al primero de estos aspectos, el que trata acerca de si la limpieza se puede hacer, la respuesta más frecuente es no, pues pueden haber situaciones en las que los datos no existan y no se pueden regenerar éstos a no ser que se tenga en acepte la considerable cantidad de esfuerzo manual y automatizado que ello requiere. Así también podrían existir casos en donde los valores estén tan confusos o dispersos en diferentes lugares con significados diferentes y opuestos, que pueden producir resultados más erróneos que los que se tienen, por lo que tal vez lo mejor sea dejar las cosas como están.

En lo referente al segundo aspecto, el que trata acerca de si se debiera limpiar los datos, de nuevo, la respuesta más frecuente es no, pues generalmente se cuestiona cuál sería el propósito de sacar datos sucios de los sistemas transaccionales e introducirlos así como vienen en el depósito de datos. Obviamente ninguno. Es claro que alguna clase de limpieza se debe dar. Sin embargo, se debe enfrentar la realidad actual de los negocios y sus expectativas de obtener información con valor agregado en el corto tiempo y a un bajo costo. Estos argumentos son suficientes como para que no haya ambiente para emprender grandes esfuerzos y gastar tiempo para crear y probar una amplia y compleja lógica que procure corregir algunos de los casos de datos sucios.

El tercer aspecto trata acerca de la selección del lugar en que se piensa hacer la limpieza de los datos. Acaso se estaría pensando en limpiar los datos operacionales en los sistemas operacionales, o se escogería hacer las transformaciones de limpieza como parte de la extracción y carga del depósito de datos. En torno a ello la primera reacción es limpiar los sistemas operacionales, sin embargo, quienes utilizan los sistemas operacionales con

propósitos operacionales no necesitan un limpiador de datos, lo que los conduce a resistirse a que se cambien sus sistemas. Además, en muchos casos ello puede tratarse de una labor muy intensiva en cuanto a trabajo, que no ofrezca una efectividad en lo que respecta a costos, o simplemente que sea imposible de hacer, lo que conduce a decidir que la limpieza se haga sobre los procesos de extracción y carga.

Finalmente, el cuarto aspecto a considerar se refiere a si es posible llevar a cabo un proceso de limpieza que resulte razonable. Dicho de otro forma, con los productos existentes en el mercado para realizar la limpieza de los datos se puede manejar la gran variedad de problemas de calidad de datos que comparten muchas de las organizaciones. La respuesta parece que sí es posible. Sin embargo, cabe cuestionarse entonces si esos productos son capaces de resolver todas esas situaciones de datos sucios que son muy complicadas y muy particulares de las empresas. Para esta pregunta la respuesta es no. Finalmente, si se tiene realmente la preocupación de crear información con valor agregado y conocimiento antes y después acerca de la condición de los datos operacionales, cabe entonces preguntarse si se está dispuesto a enfrentar la situación y escribir el código requerido. Para ésta, la respuesta es definitivamente que sí, si se quiere ser consecuente con el propósito que se persigue.

Algo importante a destacar es que la respuesta a estas preguntas debe provenir de los usuarios y no del área de sistema de información, a pesar que esta última área puede ayudar en la identificación, cuantificación, documentación y análisis de las necesidades del negocio.

2.2.7.2. Calidad de los datos.

La calidad de los datos obtenidos de un depósito representa uno de los muchos riesgos que se deben considerar a la hora de diseñar, desarrollar y operar un depósito de datos.

La calidad de los datos se puede medir en función de lo completos, válidos, consistentes, oportunos y precisos que resulten los datos ante un uso específico de éstos.

Se puede desarrollar una infraestructura de depósito de datos acorde con lo último de la tecnología, pero si los datos en el depósito no cumplen con las características de calidad requeridas para el soporte a la toma de decisiones todos los esfuerzos que se hayan hechos prontamente serán objetados aduciendo negligencia. Si se considera que este riesgo es significativo, se

debe escoger un método que evite introducir un conjunto indiscriminado de procedimientos que podrían resultar costosos tanto en términos de recursos como en soporto del negocio. Un método prudente deberá reconocer que algunos de los problemas de calidad de los datos presentan mayores riesgos que otros. Al adaptar el método que se escoja tomando en consideración lo antes mencionado permite que éste se enfoque en aplicar los recursos limitados para resolver los problemas de mayor riesgo en lugar de intentar resolver todos los posibles problemas que atentan contra los datos.

Hufford[33] sugiere un método mediante el cual se trata el tema del aseguramiento de la calidad de los datos, desde una perspectiva de administración del riesgo. Este método se adhiere a la filosofía del mejoramiento continuo propuesta por la administración de la calidad total (TQM, Total Quality Management). Sin embargo, en este método se utiliza el riesgo como un criterio para priorizar y enfocar los esfuerzos de mejoramiento de la calidad.

El método se compone de cinco actividades:

- Definir las expectativas de calidad de los datos y sus métricas: describir la calidad de los datos que se requieren para soportar las principales aplicaciones utilizadas para el apoyo a la toma de decisiones (DSS).
- Identificar los riesgos que atentan contra la calidad de los datos: pronosticar cómo los datos que se han puesto disponibles a través del depósito pueden fallar para alcanzar las expectativas.
- Valorar los riesgos que atentan contra la calidad de los datos: implementar mecanismos apropiados para la detección de los datos defectuosos y su respectivo reporte, lo cual ayude a clarificar los problemas de calidad de los datos y a decidir si se actúa y en qué forma para mitigar los riesgos observados.
- Mitigar los riesgos: tomar acciones para minimizar cada uno de los principales riesgos.
- Monitorear y evaluar los resultados: proveer visibilidad a los resultados obtenidos de los esfuerzos de mitigación de riesgos.

Dentro del conjunto de características que se refieren comúnmente en torno a la calidad de los datos se apuntan las siguientes:

- Exactitud: se refiere al grado de concordancia entre un conjunto de valores de datos y su correspondiente conjunto de valores correctos.

[33] [HUFF1996].

Un ejemplo de métrica que se puede utilizar para esta característica es el porcentaje de valores que son correctos cuando se les compara con las verdaderas características del objeto descrito por el dato.

- Completitud: se refiere al grado hasta el cual los valores están presentes en los atributos que los requieren. Una métrica para esta característica podría ser el porcentaje de datos que tienen valores almacenados.
- Consistencia: se refiere a la concordancia o coherencia lógica entre los datos que los liberan de variación o contradicción. Para este característica una métrica puede ser el porcentaje de condiciones de comparación de valores o de valores derivados que se satisfacen.
- Relatividad: se refiere a la concordancia o coherencia lógica que permite la correlación racional al comparar los datos con otros similares o parecidos. En este caso una métrica a utilizar puede ser el porcentaje de condiciones de integridad referencial que se satisfacen.
- Oportunidad: se refiere a la cantidad de ítemes o ítemes múltiples que se proveen en el tiempo requerido o especificado. Un ejemplo de métrica que se puede utilizar para esta característica es el porcentaje de datos disponibles dentro de un período de tiempo especificado.
- Unicidad: se refiere a los valores que son restringidos a un conjunto de entradas distintas, o sea, cada valor debe ser el único de su clase. La métrica que se puede utilizar para valorar esta característica puede ser el porcentaje de registros que violan la condición de unicidad (valores de llave primaria repetidos).
- Validez: se refiere a la conformación de los valores de los datos que se editan para aceptarlos. En este caso una métrica a utilizar puede ser el porcentaje de datos que tienen valores que caen dentro de su respectivo dominio de valores permitidos.

2.3. Seguridad en los depósitos de datos.

En las siguientes secciones se presentan las opiniones de diversos autores acerca del tema, la problemática, y las eventuales soluciones que se le pueden dar a la seguridad en los depósitos de datos.

2.3.1. Seguridad y administración del acceso a los datos contenidos en un depósito de datos.

En torno a ello, Gill y Rao[34] consideran que es muy importante la seguridad y administración del acceso a los datos dentro de un depósito de datos. Al incrementarse el grado de resumen, el valor de los datos se hace cada vez mayor. La información resumida que puede ayudar a una empresa a tomar decisiones es igual de valiosa para la competencia. El control del acceso a los depósitos de datos es todavía una tecnología en evolución. Esta tarea se complica por varios factores:

- el depósito de datos se construye como un conjunto abierto de datos de la empresa: Puede ayudar en la toma de decisiones y lo pueden utilizar analistas y personal operacional para mejorar sus operaciones y derivar una ventaja estratégica competitiva y sostenida. La incorporación de controles de seguridad va en contra de la necesidad de que sea un sistema abierto.
- los usuarios acceden datos dentro del depósito de datos a diferentes niveles de resumen: El mismo usuario puede comenzar con datos muy resumidos y "profundizar" en forma progresiva a datos cada vez más detallados. Otros usuarios pueden operar a un sólo nivel de resumen. Resulta difícil manejar el acceso a nivel de tabla e hilera de datos para cada uno de estos usuarios.
- la naturaleza de las herramientas OLAP y de acceso de datos en el medio del depósito de datos ha sido exploratoria: La mayoría de los usuarios utiliza el depósito de datos empleando un "proceso de descubrimiento". La incorporación de intrincados controles de seguridad puede hacer esto muy frustrante, al impedir que los usuarios avancen en su exploración.
- como el depósito de datos no maneja datos operacionales de misión crítica, la naturaleza de la amenaza a la seguridad no es la que causa daño a los datos, sino la que revela secretos y estrategias corporativas: Contrarrestar esta amenaza implica tener acceso sobre una base de "necesito saber". La política de seguridad impone restricciones a las capacidades de penetración a los detalles y el control de acceso a tablas específicas de datos resumidos y detalles operacionales. Se debe también manejar permisos para las restricciones de uso

[34] [GILL1996] páginas 136-137.

de recursos, tales como la capacidad de crear tablas temporales y consultas *ah doc*.

Otro escenario de amenaza es cuando usuarios hostiles acaparan grandes cantidades de recursos (prácticamente imposibilitando el acceso al depósito de datos). El manejo de consultas de escape, la creación de tablas temporales y la aplicación de límites de recursos a los perfiles de usuario son una forma inicial de abordar estos retos. Por lo tanto, el diseño de un plan de control de acceso y seguridad es una actividad esencial en el proceso de despliegue. Debido a la naturaleza de las aplicaciones cliente/servidor, resulta difícil manejar la seguridad desde un sólo punto de control. Los usuarios tienen claves de identificación y de acceso que, a menudo, son diferentes para cada estación de trabajo, acceso a la red, conexión remota al servidor y conexión remota a una o más bases de datos. Cuando un usuario sale de la organización, un equipo de depuración tiene que eliminar los controles de acceso en varios sistemas. Planear aplicaciones que manejen la depuración y eliminación de estos múltiples controles de acceso ayuda en la tarea de administrar la seguridad dentro del depósito de datos.

2.3.2. Administración de la seguridad en los depósitos de datos.

Inmon[35] se refiere al tema de la administración de la seguridad en los depósitos de datos de los siguientes términos.

Existen numerosas actividades que deben ser atendidas por el administrador de un depósito de datos y/o el administrador de la seguridad, en la labor de administración de un depósito de datos:

a. decidir el nivel de seguridad que se necesita
b. decidir el algoritmo de encripción que se utilizará
c. decidir en qué forma se encriptarán los datos
    - encripción por columna
    - encripción por fila
    - encripción por columna y fila
d. decidir qué usuarios tendrán acceso a qué datos
e. decidir qué tipo de respuesta recibirán los usuarios finales
    - el usuario final verá los datos encriptados
    - el usuario final verá los datos encriptados y una bandera que se lo indique

[35] [INMO1997a] páginas 62-63.

- el usuario final no verá ningún dato encriptado

f. dónde se ejecutará el algoritmo de desencripción
   - en el servidor del depósito de datos
   - en la estación de trabajo del usuario

g. qué tan frecuentemente se cambiarán las claves

h. quién tendrá acceso a las claves de seguridad

i. cómo se harán las auditorías

j. cómo serán manejados los algoritmos de encripción y desencripción en países extranjeros

k. cómo se limpiarán las estaciones de trabajo al final de una sesión

l. cuál será la base de autorización
   - el identificador de la estación de trabajo
   - la identificación del usuario final

m. resolver en qué librerías residirán los algoritmos de encripción y desencripción

n. determinar en qué casos es apropiada la seguridad basada en el estatuto VIEW (o sea, la seguridad basada en vistas)

o. determinar en qué casos es apropiada la seguridad basada en la aplicación

p. determinar en qué condiciones el cambio en la seguridad puede ser interpretado como una violación de seguridad

2.3.3. Seguridad en los depósitos de datos e Internet.

Inmon[36] aborda el tema de la seguridad en los depósitos de datos e Internet manifestando que los depósitos de datos representan un objetivo tentador para usuarios no autorizados, por cuanto los datos contenidos en éstos están:

- muy bien organizados
- integrados
- reunidos en una localización central
- optimizados para su acceso

Con la conjunción de los depósitos de datos e Internet, la tentación por hacerles travesuras y dañarlos seriamente es tan fuerte que sólo basta un poco de tiempo para

[36] [INMO1997b] páginas 8-11.

que ocurran, o más bien, se materialicen serias exposiciones en la seguridad de éstos.

En general, el efecto neto de los tipos de controles de seguridad que se pueden establecer para estos ambientes es alrededor del sistema y las bases de datos que se protegen. La idea detrás de este tipo de seguridad (o sea, la del tipo perimetral, o alrededor) es no hacer nada fundamental a lo interno del ambiente, sino más bien, crear barreras que impidan entrar a éste. Desgraciadamente, la seguridad perimetral tiene algunas fallas fundamentales:

- una vez que la barrera ha sido penetrada, no existe una protección posterior
- dentro de la barrera, los datos pueden ser descargados directamente desde uno de los discos e interpretados sin ninguna consideración de la base de datos o del sistema de seguridad
- la actividad de la red puede ser husmeada, tanto antes de que la información entre al área perimetral como después que ésta sale

Por lo tanto, la seguridad perimetral, en sus varias formas, no es segura del todo. Además, la granularidad de la seguridad ajustada a la seguridad perimetral es grosera. No existe forma de proveer protección a un nivel granular detallado dentro del ámbito del perímetro.

Otra falla de la seguridad basada en el perímetro es que, en muchos casos, ésta puede ser fácilmente penetrada. Algunas de las técnicas utilizadas para quebrantar y atravesar las medidas perimetrales clásicas son:

- fuerza bruta: tratar muchas combinaciones de contraseñas hasta descubrir la correcta
- ingeniería social: descubrir secretamente las contraseñas
- programas de contraseñas: algunos son muy buenos para adivinar las contraseñas comúnmente utilizadas
- Caballos de Troya o "engañadores": programas por medio de los cuales se gana acceso a datos que de otra forma estarían fuera de su alcance

Debido a la facilidad con que se puede entrar a un sistema de seguridad perimetral, la falta de granularidad de protección, proporcionada por la seguridad perimetral y las exposiciones de la seguridad que existen aún cuando el ambiente perimetral se haya implementado y se mantenga apropiadamente, este tipo de seguridad no es adecuado para la solidez industrial de seguridad que requiere un ambiente de depósito de datos mezclado con Internet.

De allí entonces, que se requieran mayores niveles de seguridad para garantizar una operación más confiable en tales ambientes. Para ello se sugiere utilizar la criptografía.

Existen diferentes clases de encripción. Las dos clases básicas son: la encripción de mensajes y la encripción de bases de datos/depósitos de datos.

Encripción de mensajes.

Es la técnica utilizada para asegurar el trasiego de mensajes entre dos partes. El mensaje necesita estar asegurado desde que deja el emisor, mientras es transportado, así como cuando llega a su destino. Los mensajes son enviados en un formato no estructurado. Tienen una longitud variable y carecen de estructura interna. La encripción de mensajes típicamente agrega algunos bytes al mensaje mismo como parte del proceso de encripción. Dado que no existe una estructura interna para los datos, y que el mensaje es de longitud variable, esta práctica de encripción es satisfactoria.

Encripción de datos en un ambiente de bases de datos/depósito de datos.

La encripción de los datos en una base de datos o depósito de datos es del todo diferente. En cualquiera de estos ambientes la estructura interna de los datos debe preservarse a través del proceso de encripción/desencripción. Por ejemplo, si un campo tiene una longitud de 10 bytes, cuando estos 10 bytes sean encriptados, el producto resultante debe ser un campo de 10 bytes exactos (ni uno más, ni uno menos). Además, si un campo A está antes de un campo B, el cual a su vez está antes de un campo C, en un estado no encriptado, entonces la estructura debe ser mantenida en el estado encriptado.

Otra consideración en torno a este tipo encripción lo constituye que los dominios de los datos en estos ambientes deben ser preservado a lo largo del proceso de encripción/desencripción. Si un campo ASCII va a ser encriptado, entonces el valor resultante también debe ser ASCII. O si un campo numérico se va a encriptar, el valor encriptado debe ser numérico. Mantener el dominio a través del proceso de encripción es una consideración complicada, pero es absolutamente necesaria si no se quiere que el sistema colapse ante un acceso inocente (o incidental) de un usuario no autorizado. Si el dominio no se preserva, ante un acceso de un usuario no autorizado se pueden producir excepciones de datos e incluso errores del sistema, conforme ese usuario intente recorrer a través de los datos encriptados.

En síntesis, las técnicas de encripción para los datos que residen en un depósito o base de datos son mucho más exigentes que aquellas técnicas que se encuentran en la encripción de mensajes.

También es necesario tener en cuenta que en el caso de encripción de los datos residentes en depósitos o bases de datos existe la posibilidad de encriptar algunos de los elementos de un atributo (o columna) con una llave, y otros elementos con otra llave, de manera tal que se pueda implementar un mayor nivel de granularidad de seguridad. Es más, también existe la posibilidad que no todos los atributos se encripten.

La habilidad de encriptar al nivel de columnas y de ocurrencias dentro de una columna permite un más bajo nivel de granularidad de la seguridad. Si se deseara, la seguridad se puede especificar al nivel de tabla/columna/ocurrencia. Esto significa que dentro del ambiente protegido existe una tremenda amplitud y protección de los datos.

A pesar de todo lo ya mencionado en relación con la encripción de datos residentes en un ambiente de depósitos o bases de datos, existen algunas restricciones que se asocian con la encripción de datos en un ambiente de depósitos. Los datos residen dentro de un depósito en un estado encriptado debido a que los datos se almacenan encriptados y retienen su estructura, sin embargo:

- los campos llave normalmente no se encriptan
- las columnas utilizadas en el procesamiento de las cláusulas WHERE tampoco se encriptan
- las relaciones de llave foránea tampoco pueden ser encriptadas
- los datos utilizados en el procesamiento estándar del SABD, tales como operaciones de comparación, adición y multiplicación, tampoco se encriptan

Estas consideraciones pueden parecer muy severas. En efecto, para algunas clases de procesamiento estas restricciones excluyen el uso de la encripción. Sin embargo, para muchas clases de datos estas restricciones no son problemáticas y la encripción en estos ambientes puede ser hecha sin la pérdida de funcionalidad para el usuario final.

2.3.4. Seguridad de los depósitos de datos desde la óptica de la auditoría informática.

En torno al tema de los depósitos de datos y los controles y estudios de auditoría que se pueden llevar a cabo acerca de ellos, Fryman[37] escribió lo siguiente.

[37] [FRYM1997] páginas 46-48.

Las auditorías y controles a los depósitos de datos han sido un tópico de discusión y frustración dentro de las áreas de tecnología de la información de las organizaciones desde hace algunos años. Inmon alguna vez sentenció que "los estudios de auditoría no se pueden llevar a cabo en los depósitos de datos". Esta afirmación es verdadera desde una perspectiva meramente contable. Sin embargo, muchas organizaciones han tomado al pie de la letra las palabras de Inmon, ignorando el hecho de que los datos en un depósito deben tener un nivel seguro y probado de exactitud, consistencia e integridad. Los ejecutivos y analistas empresariales no harán uso del depósito si ellos no están totalmente convencidos que pueden confiar en la información derivada de los depósitos.

Mientras los problemas empresariales y técnicos no son nuevos, pocas organizaciones han puesto funciones generales en orden para capturar, medir y rastrear los procesos de adquisición de datos y la calidad de los datos que se cargan dentro de sus depósitos y mercados de datos.

Un programa general de auditoría y controles puede ayudar a resolver tanto los problemas técnicos como empresariales, relacionados con la calidad de los datos. A continuación se presentan algunas de las preguntas típicas que se formulan luego de que un depósito de datos se ha implementado:

- ¿se corrieron todos los procesos de adquisición de datos?
- ¿dichos procesos se corrieron en la secuencia correcta?
- ¿se obtuvo el extracto correcto del libro mayor o fue este el mismo que se proceso el mes anterior?
- ¿se cargaron todas las transacciones de ventas que debían ser cargadas?
- ¿cuántos registros se borraron producto de las inconsistencias en las validaciones?
- ¿concuerdan la cantidad de registro y el total financiero?
- ¿cuándo se pueden accesar los nuevos datos en el depósito?
- ¿por qué existen diferencias entre los volúmenes de ventas en el depósito de datos y el sistema de captura de ventas?
- ¿se puede verificar que todos los datos en el depósito sean completos y exactos?

El reto de la consignación de datos es capturar, validar, integrar y transformar técnicamente los datos en información significativa y entonces almacenar esta información en el depósito o mercado de datos. El esfuerzo que ello implica es mucho más complejo que mover datos de una estructura de base de datos a otra. Por lo tanto, los auditores deben ser capaces de probar que la información en el depósito es completa y exacta. Mediante un programa de auditoría se puede proveer un paso significativo para resolver este problema.

Los objetivos de un programa de auditoría en este línea serían los siguientes:

- asegurar la exactitud y completitud de los datos fuentes que se cargan en el depósito
- mantener una actitud proactiva para identificar problemas de controles en las fuentes tan tempranamente como sea posible
- establecer el repositorio de metadatos como un componente activo del sistema que ayudará en el control de los procesos de adquisición de los datos
- proveer un proceso arquitectónico integrado para el auditoraje y control de la adquisición de los datos
- minimizar la necesidad de rastreos totales de las tablas y de recursos humanos involucrados en el proceso de balanceo
- proveer los mecanismos para registrar y monitorear los procesos de carga del depósito de datos y la calidad de los datos conforme estos se van ejecutando

Se debe hacer notar que estos objetivos no son necesarios para todos los datos de la organización. Es apropiado monitorear sólo los datos que son críticos para el negocio y funciones de éste soportados por el depósito de datos. Los datos que son críticos para el negocio deben estar definidos dentro de las metas y objetivos de cada organización individual. Se podría anticipar que los datos críticos de la empresa pueden cambiar con el tiempo, de la misma forma en que pueden cambiar las metas y objetivos, y la empresa misma.

2.3.5. Seguridad en un ambiente de depósitos de datos basado en Oracle.

2.3.5.1. Oracle8.

Según Corey et al.[38] como parte de una política de seguridad de depósitos de datos existen ciertas consideraciones a tomar en cuenta, algunas de ellas de sentido común, tales como:

- seguridad en las estaciones de trabajo: se refiere principalmente al uso de protectores de pantalla (*screen savers*) con contraseña, para evitar que en ausencia del usuario, otra persona utilice su estación y haga operaciones que luego le sean imputadas.
- posibilidad de curiosear (*snooping*): ello involucra examinar o capturar información de otra computadora, información a la que no

[38] [CORE1998] páginas 389-409.

se tiene acceso directamente desde su propio equipo. Ello puede tomar la forma de inspeccionar correos o vistas que se obtienen como producto del análisis hecho de otra persona, a la cual, un usuario no tendría normalmente acceso utilizando sus propios medios.

Este tipo de consideraciones tienen que ver directamente con los hábitos conforme a los cuales los usuarios interactúan con la información electrónica con la que se tiene contacto en un día normal de trabajo.

Todas las empresas atraviesan por una serie de ejercicios durante el desarrollo de los sistemas operacionales. Estos ejercicios definen quiénes estarían en capacidad de hacer qué cosa con cuáles datos. A algunos usuarios, basados en perfiles de seguridad predefinidos, se les permitirá ejecutar acciones tales como SELECT, CREATE, UPDATE y DELETE de la información en la base de datos, mientras que otros tendrán un puñado de estos privilegios, de acuerdo con sus necesidades. Independientemente de la forma cómo se accesen los datos, existen dos niveles de seguridad:

- la cesión de privilegios dados a través del mecanismo GRANT, en donde a los usuarios o grupos de usuarios se les da explícitamente permiso para accesar los datos
- el derecho para ejecutar ciertas operaciones sobre los datos, basado en quién es el usuario y dónde se ubica éste dentro de la estructura administrativa de la empresa

El primero de los niveles de seguridad antes mencionado, basta para que para la mayoría de usuarios de un depósito de datos puedan cumplir con sus labores. El segundo nivel sólo toma lugar para aquellos usuarios de Oracle quienes son dueños de los datos que están almacenados en el repositorio de los sistemas de soporte a la toma de decisiones (DSS).

El corazón de un sistema de seguridad en la mayoría, si no en todas las bases de datos relacionales, lo constituye la posibilidad de que los usuarios puedan ver los datos contenidos en éstas. Debido a que los datos en el depósito son de sólo lectura por naturaleza, el privilegio SELECT de Oracle se le da a los usuarios del depósito de datos. Este privilegio permite a los usuarios que se especifiquen ver los datos una vez que se ha establecido un *nombre universal* para ellos. Un nombre universal significa un nombre de tipos para la base de datos de objetos dentro de la cual se almacenan los datos. Oracle le llama a este nombre universal un *sinónimo público*, el cual entonces puede ser utilizado por todos los usuarios que se conecten a la base de datos para referirse a ese objeto en particular.

Ante un escenario como el que se ha venido presentando en torno al privilegio SELECT, surge la siguiente interrogante: ¿cómo se deben administrar los privilegios en un depósito de datos?.

Para responderla, Oracle8 tiene una característica llamada seguridad basada en roles que es ideal tanto para ambientes operacionales como para ambientes DSS.

Seguridad basada en roles.

Un rol es una agrupación lógica de uno o más usuarios de una base de datos Oracle, a la cual se le dan privilegios, basándose para ello en las responsabilidades funcionales de las personas registradas en el rol.

Derivado de esta idea surge entonces el *enrolamiento*, que consiste en el proceso de dar membresía a uno o más usuarios en un rol. Debe recordarse que la gente necesita una cuenta Oracle para poder conectarse a la base de datos. Por lo tanto, una vez que un usuario tiene la habilidad de conectarse a la base de datos (vía una cuenta), es el momento propicio para hacerlo miembro del rol que le corresponde en función de sus responsabilidades, y así poder desarrollar sus labores.

Entonces, una vez que los usuarios son incorporados en los roles apropiados, el propietario de los objetos de la base de datos en el depósito procede a ceder privilegios a los roles.

El poderío de la seguridad basada en roles es que una vez que éstos se han creado y han recibidos los privilegios apropiados, los nuevos usuarios simplemente son hechos miembros de uno o más roles, los que necesite para interactuar con el depósito de datos. Aún mejor, cuando a un nuevo usuario se le da la membresía en un rol, este nuevo usuario automáticamente hereda los privilegios que vienen desde antes con el rol.

Por otra parte, las vistas son el punto central de muchos mecanismos de seguridad, no sólo en los depósitos de datos sino también en muchos sistemas operacionales.

Seguridad basada en vistas.

Las vistas constituyen otra forma de restringir la información que los usuarios pueden ver. Una vista es una tabla lógica construida como un subconjunto de los datos contenidos en una o más tablas físicas. Aquí la

palabra *lógica* significa que la vista en sí misma no contiene datos; los datos se ensamblan a partir de las tablas involucradas al mencionar la vista en un estatuto SQL.

Puesto que las vistas generan tablas lógicas, las cuales restringen la información que pueden ver los usuarios, es factible combinar los conceptos de vistas y roles, de manera que se logre un mejor control de la información que se puede visualizar por parte de los usuarios de una base de datos o depósito de datos.

Existen algunas historias acerca de los depósitos de datos, en donde la seguridad se implementó al nivel de otras herramientas distintas de Oracle. Aquí el *nivel de herramientas* representa situaciones en donde los implementadores del depósito utilizaron características en las consultas y herramientas de análisis para controlar los datos que se les permite ver a los usuarios. A pesar de que estos mecanismos de seguridad pueden resolver problemas en el corto plazo, lo más juicioso es utilizar las facilidades de seguridad que ofrece la base de datos Oracle. El problema de la seguridad basada en herramientas tiene dos caras:

1. las características de seguridad que incorporan los productos de ciertos vendedores sólo trabajan para esas herramientas. Si se intenta implementar una seguridad basada en herramientas es muy probable que se deba hacer mucho trabajo manual para cada herramienta. Ello podría provocar que sea muy fácil de permitir que en forma inadvertida algunos usuarios puedan ver datos en una región del depósito utilizando una herramienta X, mientras que otra herramienta Y les restrinja de ver los mismos datos.
2. las características de seguridad propias de cada vendedor deben ser aprendidas por todos los miembros del personal de seguridad del depósito de datos. Conforme el personal cambia (como siempre sucede), la curva de aprendizaje toma el sentido del proceso de transferencia de responsabilidades.

Seguridad basada en herramientas.

La seguridad basada en herramientas, aunque no tan fluida y portable como la seguridad basada en el motor de base de datos, es práctica. Con muchas herramientas, entre las que se incluye Oracle Discoverer, no hay elección. La herramienta no se puede voltear a la comunidad de usuarios antes de que sus características de seguridad hayan sido implementadas. Oracle Discoverer utiliza el concepto de áreas de negocio para coleccionar un conjunto de información relacionada con un propósito común del negocio, organizados en carpetas (o *folders*), las carpetas contienen elementos (*items*), los cuales son una categoría particular de información dentro de la carpeta.

El acceso a las áreas de trabajo es donde se implementa el punto de partida de la seguridad en Discoverer.

Existen dos versiones del producto Discoverer 3, la edición Administrador y la edición Usuario. La separación del producto en dos versiones controla quién tiene acceso a poder establecer y cambiar las áreas de negocio, y quién puede simplemente trabajar con las áreas de negocio a las cuales se le ha dado acceso.

El acceso a las tablas se concede de la siguiente forma. Después de que se ha creado una nueva área de negocio, el administrador elige y selecciona los esquemas y objetos dentro de estos esquemas, que el usuario podrá ver utilizando esta área de negocio.

El acceso a las áreas de negocio se da de la forma que a continuación se explica. Después de definir el área de negocio y escoger los objetos para incluir en ésta, el administrador debe hacer uso de la caja de diálogo Seguridad. Existen dos carpetas en la aludida caja de diálogo.

1. En la primera es en donde se elige una de las áreas de negocio de las disponibles en la lista de escogencias y entonces se transfiere un usuario o rol, de la lista de disponibles a la lista de seleccionados (en este caso sería la carpeta etiquetada como **Business Area -> User**)
2. En la segunda es en donde se elige un usuario o rol de los disponibles en la lista de escogencias y entonces se transfiere el nombre del área de negocio, de la lista de disponibles a la listas de seleccionados (en este caso sería la carpeta etiquetada como **Users -> Business Area**)

Seguridad en las contraseñas con Oracle8.

Oracle8 ofrece el envejecimiento de contraseñas, así como un sofisticado conjunto de características para el control de contraseñas. Muchas personas, distintas de las que trabajan con seguridad en bases de datos, encuentran la administración de contraseñas algo molesto. Pero nadie quiere un acceso no autorizado a sus datos sensitivos.

Las facilidades que ofrece Oracle8 se basan en los perfiles (o *profiles*) de Oracle. Los perfiles se usan entonces para administrar las contraseñas en Oracle8. Un perfil viene a ser un conjunto de limitaciones a los recursos, que se distribuyen a los usuarios de la base de datos. Estas limitaciones son en áreas tales como consumo de tiempo de CPU por sesión o el tiempo inactivo de las terminales permisible antes de ser desconectada ésta de la

base de datos. Entre las facilidades que se ofrecen en Oracle8 están el bloqueo de cuentas y el envejecimiento de las contraseñas.

En relación con el bloqueo de cuentas, es aquí donde del DBA (administrador de la base de datos) especifica el número de intentos fallidos de ingreso consecutivos después del cual la cuenta se bloquea. Esta medida sirve para dos propósitos: a) para asistir en la prevención deliberada o intencional de que una persona no autorizada ingrese a la base de datos y b) detectar cuando alguien está tratando de violentar una cuenta.

Puesto que el envejecimiento de las contraseñas limita la vida de la contraseña de un usuario, Oracle8 usa un parámetro para controlar el tiempo de expiración de éstas.

Administración de los usuarios de la base de datos.

Existen diversas escuelas de pensamiento acerca de este tema. Hay quienes piensan que esto debería ser responsabilidad del DBA; otros piensan que ello merece ser administrado por un grupo de usuarios de confianza.

Privilegios del sistema.

Oracle8 permite el despliegue de cerca de 80 privilegios del sistema. Los privilegios del sistema le permiten a los usuarios realizar acciones que previamente sólo eran posibles de realizar con la cuenta del DBA. Esta característica va dirigida al deseo de separar algunas actividades hacia cuentas distintas de la del DBA, sin comprometer la seguridad total de la base de datos. Algunos de estos privilegios podrían ser el **Create User**, que permite crear cuenta en un servidor Oracle8; **Create Session**, que le permite a los usuarios conectarse a la base de datos; o el **Alter User**, que le permite a un usuario, entre otras cosas, cambiar la contraseña de otros usuarios de la base de datos.

2.3.5.2. Oracle8*i*.

En la sección precedente se expuso lo tratado en [CORE1999] acerca de la seguridad en los depósitos de datos basados en Oracle8. Los mismos autores[39] actualizaron el tema, adecuándolo a las particularidades de Oracle8*i*.

[39] [CORE2001] páginas 482-490.

Oracle8*i* permite desplegar más de 110 privilegios del sistema. Los privilegios del sistema le permiten a los usuarios realizar acciones que previamente sólo estaban disponibles para las cuentas de DBA. Esta característica va dirigida al deseo de separar algunas actividades hacia cuentas distintas de la del DBA, sin comprometer la seguridad total de la base de datos. Algunos de estos privilegios podrían ser el **Create User**, que permite crear cuenta en un servidor Oracle8i; **Create Session**, que le permite a los usuarios conectarse a la base de datos; o el **Alter User**, que le permite a un usuario, entre otras cosas, cambiar la contraseña de otros usuarios de la base de datos.

Al parecer son muchos los que prefieren utilizar la interfaz GUI del DBA Studio para realizar sus tareas relacionadas con la seguridad en depósitos de datos en Oracle8*i*. Los estatutos SQL que se construyen cuando se trabaja con el Studio son exactamente los mismos que se estructuran manualmente utilizando la interfaz del Server Manager, en línea de comando, o incluso con SQL*PLUS. Es desde el árbol Security (seguridad) en el DBA Studio donde se administran los roles, usuarios y perfiles.

Los estatutos SQL que se pueden construir y pasar a los usuarios en el DBA Studio son muy poderosos y pueden debilitar los esfuerzos del equipo de depósito de datos, si se abusa de ellos. Por ejemplo, en un ambiente cliente/servidor, cuando se da mantenimiento a usuarios, si se ha seleccionado EXTERNAL FOR AUTHENTICATION en una de las cajas de diálogo de CREATE o EDIT USER, se puede mantener en espera de conexión a la base de datos a un usuario hasta que la opción sea cambiada a PASSWORD.

En este sentido los autores recomiendan ser muy conservadores cuando se le da acceso al DBA Studio a usuario poderoso de una comunidad de usuarios. Mientras que es poco probable que un usuario poderoso pueda causar un daño de forma deliberada, usuarios inexpertos pueden hacer mucho daño inadvertido al depósito de datos.

2.3.5.3. Oracle9*i*.

Desde el punto de vista de Oracle9*i* como uno de los eventuales motores de base de datos para un depósito, Greenwald[40] apunta algunas consideraciones en torno a la forma como que se implementa la seguridad.

[40] [GREE2001] páginas 119-128.

Sobre el particular se indica que uno de los aspectos más importantes de la administración de una base de datos Oracle en un ambiente multiusuario es la creación de un esquema de seguridad para controlar el acceso y la modificación de la base de datos.

La administración de la seguridad de la base de datos se realiza en tres niveles diferentes:

- A nivel de la base de datos
- A nivel del sistema operativo
- A nivel de la seguridad de la red

Por ejemplo, en el ámbito del sistema operativo el DBA podrían tener la capacidad de crear y borrar archivos relacionados con la base de datos, en donde usuarios típicos de la base de datos no deberían tener esa capacidad. En grandes organizaciones, los DBA o los administradores de la seguridad de la base de datos trabajan estrechamente con los administradores del sistema computacional para coordinar especificaciones y prácticas de seguridad.

Las especificaciones de seguridad de la base de datos controlan quién tiene acceso a la base de datos y establecen límites a las capacidades de los usuarios a través del par código_de_usuario/contraseña (*username/password*). Tales especificaciones pueden limitar la asignación de recursos (espacio en disco y tiempo de procesador –CPU–) a los usuarios y autorizar la realización de auditorías a los usuarios. La seguridad de la base de datos en el nivel de la base de datos también permite controlar el acceso y el uso de esquemas de objetos específicos en la base de datos.

Debe recordarse que el DBA o el administrador de la seguridad de la base de datos debe crear un código_de_usuario (*username*) para cada usuario con el objeto de proveerle de un identificador único válido el cual debe utilizar para conectarse a la base de datos. Asociado a este código está la respectiva contraseña que permite confirmar la identidad del usuario que intenta conectarse a la base de datos.

Una vez que un usuario ha logrado conectarse exitosamente a una base de datos, su acceso está restringido por los privilegios, que no es otra cosa más que los derechos que se le otorgan para ejecutar ciertos estatutos del SQL. Algunos de esos privilegios se otorgan con una cobertura de todo el sistema (como la capacidad de borrar filas en cualquier lugar de la base de datos) o

pueden aplicar solamente a un esquema de objeto específico en la base de datos.

Se le llama rol a un grupo de privilegios, y éstos pueden ser creados, modificados o eliminados. En la mayoría de implementaciones, el DBA o el administrador de la seguridad de la base de datos crea los códigos para los usuarios y le asigna roles a usuarios específicos, y por este medio otorgándoles a éstos un conjunto de privilegios. Hoy día ello ha tendido a incrementarse por medio del uso del Oracle Enterprise Manager.

Existen cuatro privilegios básicos de seguridad que se aplican a los datos en una base de datos Oracle:

- SELECT: para realizar consultas
- INSERT: para agregar filas en las tablas o vistas
- UPDATE: para actualizar filas en las tablas o vistas
- DELETE: para eliminar filas de las tablas, particiones de tablas o vistas

Adicional a estos privilegios existen otros que aplican a los objetos dentro de un esquema de base de datos:

- CREATE: para crear una tabla en un esquema
- DROP: para eliminar una tabla de un esquema
- ALTER: para variar una tabla o vista

Todos estos privilegios son manejados a través de dos estatutos simples de SQL. El estatuto GRANT otorga un privilegio particular a un usuario o rol, mientras que el estatuto REVOKE cancela un privilegio específico. Por lo tanto, se pueden usar estatutos GRANT y REVOKE para modificar los privilegios de un individuo o rol.

En Oracle 9*i* existe una serie de roles y privilegios por omisión. Ellos son:

- CONNECT: permite conectarse a la base de datos, crear objetos y llevar a cabo exportaciones
- RESOURCE: permite crear procedimientos, *triggers* y tipos dentro del área de esquema del usuario
- DBA: permite privilegios virtualmente ilimitados

- SYSOPER: conjunto de privilegios. Permite conectarse a la base de datos de forma remota y realizar un conjunto limitado de acciones privilegiadas, lo que incluye iniciar (*start up*) y cerrar (*shut down*) la base de datos
- SYSDBA: conjunto de privilegios. Es muy similar al rol de DBA descrito anteriormente. Incluye el conjunto de privilegios de SYSOPER y todos aquellos privilegios con la especificación ADMIN OPTION. Permite conectarse a la base de datos y remotamente realizar acciones privilegiadas, tales como iniciar (*start up*) y cerrar (*shut down*) la base de datos
- EXP_FULL_DATABASE: permite realizar acciones de exportación sobre cualquier objeto de la base de datos y registrar tales actividades en el diccionario de datos
- IMP_FULL_DATABASE: permite convertirse en un usuario de modo que sus objetos pueden ser importados en el área apropiada del esquema
- DELETE_CATALOG_ROLE: permite eliminar filas de la tabla de auditoría SYS.AUD$
- EXECUTE_CATALOG_ROLE: permite ejecutar cualquier paquete exportado que aparezca listado en el catálogo de recuperación
- SELECT_CATALOG_ROLE: permite seleccionar roles a partir de todas las vistas o tablas exportadas en el catálogo de recuperación
- RECOVERY_CATALOG_OWNER: permite crear al propietario de un catálogo de recuperación
- SNMPAGENT: utilizado por el Oracle Enterprise Intelligent Agent

Por su parte, el rol de DBA es uno de los más importantes roles por omisión que viene con Oracle. Este rol le da a los usuarios que sean sus miembros todos los privilegios del sistema. Este rol le permite a sus miembros realizar cualquiera de las acciones, ya sea que lo hagan desde la línea de comandos del SQL*PLUS, desde el Oracle Server Manager (*svrmgrl*) o a través de la interfaz del Oracle Enterprise Manager:

- STARTUP: inicia una instancia de la base de datos
- SHUTDOWN: cierra una instancia de la base de datos
- ALTER DATABASE OPEN: abre una base de datos montada pero que está cerrada
- ALTER DATABASE MOUT: monta una base de datos utilizando una instancia previamente iniciada

- ALTER DATABASE BACKUP: por ejemplo, inicia un respaldo de un archivo de control. Sin embargo, en la actualidad los respaldos se hacen más frecuentemente por medio del RMAN.
- ALTER DATABASE ARCHIVELOG: especifica que los contenidos de un grupo de archivos de redo log deben ser archivados antes de que el grupo sea reutilizado
- ALTER DATABASE RECOVER: aplica los *logs* individualmente o inicia la aplicación automática de los redo *logs*.
- CREATE DATABASE: crea y nombra una base de datos, especifica los *datafiles* y sus tamaños, especifica los *logfiles* y sus tamaños, y establece los parámetros límites tales como MAXLOGFILES, MAXDATAFILES, entre otros
- RESTRICTED SESSION: permite conexiones a la base de datos iniciadas en un modo restringido (*Restricted mode*). El modo restringido se diseñó para mantener fuera de la base de datos a los usuarios durante actividades tales como solución de problemas (*troubleshooting*) y algunos tipos de mantenimiento

Las vistas pueden ser consideradas tablas virtuales por cuanto se definen a partir de consultas que extraen o se derivan datos de tablas físicas utilizadas como base. Debido a que ello puede ser utilizado para crear diferentes presentaciones de los datos para diferentes grupos de usuarios, se pueden utilizar vistas para presentarle a los usuarios solo las filas o columnas que un cierto grupo de usuarios debiera ser capaz de accesar.

En línea con lo anterior, implementar la seguridad puede ser un proceso que consume mucho tiempo, especialmente si se quiere basar la seguridad sobre un atributo con un amplio rango de valores. Por ejemplo, si se requiere ofrecer el acceso de escritura a un cierto grupo de usuario y el acceso de lectura a otro grupo, la situación se torna más compleja. A más pequeño alcance, o granular, en el control de acceso que se desee, mayor es el trabajo involucrado para crear y mantener los privilegios de seguridad.

Oracle ofrece un tipo de seguridad que se puede utilizar para brindar este tipo de control de acceso de fina granularidad (*fine-grained access control* –FGAC–). Las políticas de seguridad (implementadas como funciones PL/SQL) se pueden asociar con las tablas o vistas. Una política de seguridad retorna una condición que está asociada dinámicamente con un estatuto particular de SQL, el cual limita transparentemente el dato que se retorna. La política de seguridad podría retornar una cláusula WHERE, basada en una responsabilidad particular de los usuarios, que limita las filas que se retornan. De esta forma, se puede mantener el rango de cada usuario en una

tabla separada, la cual es dinámicamente consultada como parte de la función de la política de seguridad.

Se puede asociar una política de seguridad con una vista o tabla particular, utilizando para ello el paquete incorporado en PL/SQL DBMS_RLS, el cual permite que se refresque, habilite o deshabilite una política de seguridad.

Con la seguridad de fina granularidad también se puede implementar seguridad basada en el tipo de estatuto SQL que se utilice. La política de seguridad previamente descrita podría ser utilizada para limitar las operaciones de UPDATE, INSERT y DELETE a un conjunto de datos, pero permitir operaciones de SELECT sobre un grupo diferente de datos.

Una nueva solución de control de acceso de fina granularidad para Oracle9i es Oracle Label Security. Este producto agrega marcas y etiquetas especiales para las filas de datos y está construido sobre la tecnología de Oracle9i Virtual Private Database (VPB). Oracle Label Security incluye una herramienta para administrar políticas, etiquetas y etiquetas de autorización de usuarios.

Otra novedad en Oracle9i es la auditoría de fina granularidad, la cual permite auditorías selectivas de estatutos SELECT con variables vinculadas, basadas en el acceso de columnas específicas.

2.3.6. Seguridad en un ambiente de depósitos de datos basado en SQL Server.

2.3.6.1. SQL Server 7.

Corey et al.[41] se permiten hacer una presentación bastante similar a la de [CORE1998], con las diferencias del caso entre Oracle8 y SQL Server 7.

Se menciona que como parte de una política de seguridad de depósitos de datos existen ciertas consideraciones a tomar en cuenta, algunas de ellas de sentido común, tales como:

- seguridad en las estaciones de trabajo: se refiere principalmente al uso de protectores de pantalla (*screen savers*) con contraseña, para

[41] [CORE1999] páginas 245-257.

evitar que en ausencia del usuario, otra persona utilice su estación y haga operaciones que luego le sean imputadas.

- posibilidad de curiosear (*snooping*): ello involucra examinar o capturar información de otra computadora, información a la que no se tiene acceso directamente desde su propio equipo. Ello puede tomar la forma de inspeccionar correos o vistas que se obtienen como producto del análisis hecho de otra persona, a la cual, un usuario no tendría normalmente acceso utilizando sus propios medios.

Este tipo de consideraciones tienen que ver directamente con los hábitos conforme a los cuales los usuarios interactúan con la información electrónica con la que se tiene contacto en un día normal de trabajo.

Todas las empresas atraviesan por una serie de ejercicios durante el desarrollo de los sistemas operacionales. Estos ejercicios definen quiénes estarían en capacidad de hacer qué cosa con cuáles datos. A algunos usuarios, basados en perfiles de seguridad predefinidos, se les permitirá ejecutar acciones tales como SELECT, CREATE, UPDATE y DELETE de la información en la base de datos, mientras que otros tendrán un puñado de estos privilegios, de acuerdo con sus necesidades. Independientemente de cómo se accesen los datos, existen dos niveles de seguridad:

- la cesión de privilegios dados a través del mecanismo GRANT, en donde a los usuarios o grupos de usuarios se les da explícitamente permiso para accesar los datos
- el derecho para ejecutar ciertas operaciones sobre los datos, basado en quién es el usuario y dónde se ubica éste dentro de la estructura administrativa de la empresa

El primero de los niveles de seguridad antes mencionado, basta para que para la mayoría de usuarios de un depósito de datos puedan cumplir con sus labores. El segundo nivel sólo toma lugar para aquellos usuarios de SQL Server 7 quienes son dueños de los datos que están almacenados en el repositorio de los sistemas de soporte a la toma de decisiones (DSS).

El corazón de un sistema de seguridad en la mayoría, si no en todas las bases de datos relacionales, lo constituye la posibilidad de que los usuarios puedan ver los datos contenidos en éstas. Debido a que los datos en el depósito son de sólo lectura por naturaleza, el privilegio SELECT de SQL Server 7 se le da a los usuarios del depósito de datos. Este privilegio permite a los usuarios que se especifiquen ver cierta información dentro de la base de datos. Dependiendo de la herramienta que se utilice para accesar los datos, el primer nivel de seguridad tomar lugar.

Ante un escenario como el que se ha venido presentando en torno al privilegio SELECT, surge la siguiente interrogante: ¿cómo se deben administrar los privilegios en un depósito de datos?.

Para responderla, SQL Server 7 también tiene una característica llamada seguridad basada en roles que es ideal tanto para ambientes operacionales como para ambientes DSS.

Seguridad basada en roles.

Un rol es una agrupación lógica de uno o más usuarios de una base de datos SQL Server 7, a la cual se le dan privilegios, basándose para ello en las responsabilidades funcionales de las personas registradas en el rol.

Derivado de esta idea surge entonces el *enrolamiento*, que consiste en el proceso de dar membresía a uno o más usuarios en un rol. Debe recordarse que la gente necesita una cuenta SQL Server 7 para poder conectarse a la base de datos. Por lo tanto, una vez que un usuario tiene la habilidad de conectarse a la base de datos (vía una cuenta), es el momento propicio para hacerlo miembro del rol que le corresponde en función de sus responsabilidades, y así poder desarrollar sus labores.

Entonces, una vez que los usuarios son incorporados en los roles apropiados, el propietario de los objetos de la base de datos en el depósito procede a ceder privilegios a los roles.

El poderío de la seguridad basada en roles es que una vez que éstos se han creado y han recibidos los privilegios apropiados, los nuevos usuarios simplemente son hechos miembros de uno o más roles, los que necesite para interactuar con el depósito de datos. Aún mejor, cuando a un nuevo usuario se le da la membresía en un rol, este nuevo usuario automáticamente hereda los privilegios que vienen desde antes con el rol.

Un consejo útil que se da en torno a la concesión de privilegios es aquél que dice que no se le debe permitir a un usuario de un depósito de datos ver información en el depósito que tampoco le fuera permitida en un sistema operacional.

Por otra parte, las vistas son el punto central de muchos mecanismos de seguridad, no sólo en los depósitos de datos sino también en muchos sistemas operacionales.

Seguridad basada en vistas.

Las vistas constituyen otra forma de restringir la información que los usuarios pueden ver. Una vista es una tabla lógica construida como un subconjunto de los datos contenidos en una o más tablas físicas. Aquí la palabra *lógica* significa que la vista en sí misma no contiene datos; los datos se ensamblan a partir de las tablas involucradas al mencionar la vista en un estatuto SQL.

Puesto que las vistas generan tablas lógicas, las cuales restringen la información que pueden ver los usuarios, es factible combinar los conceptos de vistas y roles, de manera que se logre un mejor control de la información que se puede visualizar por parte de los usuarios de una base de datos o depósito de datos.

Existen algunas historias acerca de los depósitos de datos, en donde la seguridad se implementó al nivel de otras herramientas distintas de SQL Server 7. Aquí el *nivel de herramientas* representa situaciones en donde los implementadores del depósito utilizaron características en las consultas y herramientas de análisis para controlar los datos que se les permite ver a los usuarios. A pesar de que estos mecanismos de seguridad pueden resolver problemas en el corto plazo, lo más juicioso es utilizar las facilidades de seguridad que ofrece la base de datos SQL Server 7. El problema de la seguridad basada en herramientas tiene dos caras:

1. las características de seguridad que incorporan los productos de ciertos vendedores sólo trabajan para esas herramientas. Si se intenta implementar una seguridad basada en herramientas es muy probable que se deba hacer mucho trabajo manual para cada herramienta. Ello podría provocar que sea muy fácil de permitir que en forma inadvertida algunos usuarios puedan ver datos en una región del depósito utilizando una herramienta X, mientras que otra herramienta Y les restrinja de ver los mismos datos.
2. las características de seguridad propias de cada vendedor deben ser aprendidas por todos los miembros del personal de seguridad del depósito de datos. Conforme el personal cambia (como siempre sucede), la curva de aprendizaje toma el sentido del proceso de transferencia de responsabilidades.

Seguridad basada en herramientas.

La seguridad basada en herramientas, aunque no tan fluida y portable como la seguridad basada en el motor de base de datos, es práctica y podría formar parte de la solución de seguridad. Se debe tener presente que se podría implementar tanta seguridad como sea posible dentro de la base de datos, y así evitar los potenciales problemas de una seguridad basada en herramientas. Con muchas herramientas, entre las que se incluye Impromptu de Cognos, no hay elección. La herramienta no se puede voltear a la comunidad de usuarios antes de que sus características de seguridad hayan sido implementadas. Impromptu utiliza el concepto de catálogos para coleccionar un conjunto de información relacionada con un propósito común del negocio, organizados en carpetas (o *folders*), las carpetas contienen elementos (*items*), los cuales son una categoría particular de información dentro de la carpeta. El acceso a los catálogos es donde se implementa el punto de partida de la seguridad en Impromptu.

Existen dos versiones del producto Impromptu, la edición Administrador y la edición Usuario. La separación del producto en dos versiones controla quién tiene acceso a poder establecer y cambiar los catálogos, y quién puede simplemente trabajar con las áreas a las cuales se le ha dado acceso.

Dentro de la versión Administrador de Impromptu el usuario puede crear catálogos y usuarios, y entonces concederles o negarles acceso a los usuarios a carpetas (agrupamientos lógicos) y tablas (bases de datos fundamentales).

Una vez que se tiene control de quiénes tienen acceso a cada una de las tablas, se pueden crear carpetas que contengan columnas de muchas tablas diferentes, así como cálculos basados en los valores de las columnas. Normalmente las carpetas se definen para agrupaciones lógicas de tablas. Se puede imponer una restricción posterior a través de los permisos que le permiten a los usuarios utilizar los valores como parte de una cláusula SELECT, pero previniendo que se agreguen columnas.

Seguridad en las contraseñas con SQL Server 7.

La mejor forma de administrar la seguridad con SQL Server 7 es utilizar la seguridad integrada de Windows NT. La seguridad integrada ofrece una única clave de acceso a SQL Server 7, mientras le permite al administrador controlar quién tiene acceso a qué bases de datos, tablas y funciones.

Windows NT ofrece el envejecimiento de contraseñas, así como un sofisticado conjunto de características para el control de contraseñas, de modo tal que los usuarios ya no tienen por qué preocuparse por otra clave de acceso o contraseña.

Las políticas sobre las cuentas ayudan a reforzar la seguridad en la red, lo que incluye aspectos tales como longitud de la contraseña y el tiempo de expiración. También se puede controlar cuántas contraseñas se pueden recordar, previniendo que algún usuario reutilice una vieja contraseña hasta tanto ésta se haya cambiado un determinado número de veces. Es importante también reforzar la longitud mínima de la contraseña. Para ello se debe definir un valor que fuerce a los usuarios a utilizar una contraseña que sea suficientemente larga para prevenir que la adivinen fácilmente, pero lo suficientemente corta como para que ellos no tengan que acudir a escribirla de ninguna forma para recordarla. Igualmente se puede tener en cuenta forzar a que regularmente los usuarios estén cambiando las contraseñas sin llegar a extremos.

El bloqueo de cuentas sirve para dos propósitos: a) para asistir en la prevención deliberada o intencional de que una persona no autorizada ingrese a la base de datos y b) detectar cuando alguien está tratando de violentar una cuenta.

Administración de los usuarios de la base de datos.

Existen diversas escuelas de pensamiento acerca de este tema. Hay quienes piensan que esto debería ser responsabilidad del DBA; otros piensan que ello merece ser administrado por un grupo de usuarios de confianza.

2.3.6.2. SQL Server 2000.

En lo que respecta al tema de la seguridad en SQL Server 2000, Shapiro[42] inicia haciendo una diferenciación entre los conceptos de integridad y seguridad de los datos, valiéndose para ello de una cita a lo que C.J. Date opina sobre el tema. Indica Date que seguridad significa proteger los datos contra los usuarios no autorizados, mientras que integridad significa proteger los datos contra los usuarios autorizados.

El NT File System (NTFS, sistema de archivo NT), protege a los objetos verificando que un encargado de seguridad tiene acceso a un objeto. El sistema de archivo también verifica que los niveles de acceso han sido otorgados al encargado de seguridad, tales como *read*, *write*, *delete* y *execute*. *Deny* también es un nivel de acceso, a pesar de que previene el acceso al objeto. Entonces cabe preguntarse cómo es que el sistema de

[42] [SHAP2001] páginas. 206-261.

archivo conoce de quién se trata, sin tener que reautenticar al usuario, cada vez éste necesita accesar un objeto. Además, cómo es que el sistema conoce el nivel de acceso que conlleva la autenticación de un usuario. En un ambiente de Windows ello se logra utilizando *token* de acceso, identificadores de seguridad (SID, Security Identifier) y listas de control de acceso (ACL, por sus siglas en inglés de Access Control List).

Un *token* de acceso es una "insignia de confianza", la cual actúa como un poder (*proxy*), asignado al usuario para que se autentique satisfactoriamente en un dominio Windows NT o Windows 2000. El *token* es el medio por el cual el sistema de seguridad le otorga a un usuario el derecho de paso por la red, o de atravesar un dominio, a pesar de que el paso y el acceso a los recursos lo realiza el sistema en su representación, por medio del *token*. El derecho de paso no es abierto o ilimitado. Este puede ser cancelado por el sistema de seguridad o por el administrador, y la cancelación se da a partir del momento en que produce la desconexión (*log off*) de la base de datos.

Una vez que los usuarios están dentro del SQL Server, su objetivo es accesar la base de datos, mientras que el objetivo del motor es asegurarse de que los usuarios sólo accesen los datos que se les ha permitido accesar. También se debe asegurar que los usuarios no destruyan los datos a través de acciones intencionales o accidentales. Los mecanismos que se tienen para asegurar los datos, en el ámbito de los objetos, son muy similares a los mecanismos de seguridad que se disponen en el sistema de archivo o en el sistema operativo.

Una base datos se puede ver como un objeto que se compone de objetos. Ello constituye una jerarquía de objetos, la que se puede visualizar como una cadena de objetos. Todos los objetos en la cadena heredan propiedades del objeto padre cuando éstos son creados.

Microsoft introdujo un nuevo sistema de permisos para SQL Server empezando con la versión 7.0, su revisión monumental del siglo de este producto y precursor del SQL Server 2000. Dicho sistema ha sido extendido y mejorado precisamente con el SQL Server 2000. Se basa en el modelo de seguridad y en la arquitectura de permisos NTFS de Windows. Una lista de control de acceso para un objeto de la base de datos contiene el nombre de los usuarios o grupos de éstos a los que se les ha otorgado acceso a un objeto, y el nivel de este acceso. Un permiso especial (DENY) impide que un usuario o grupo de usuarios puedan accesar al objeto de una cierta forma.

Como sucede en su contraparte de los niveles de sistema operativo y NTFS, el usuario de SQL Server 2000 obtiene la sumatoria de los permisos asignados a nivel de usuario individual o de rol. Esto significa que si un usuario adquiere permiso para seleccionar registros a partir de un rol y de otro rol obtiene el permiso para agregar y seleccionar registros, los permisos combinados son de INSERT y SELECT. Sin embargo, si otro grupo en el mismo objeto deniega el SELECT, al usuario se le deniega el permiso de SELECT, sin importar que de alguna otra forma se le haya otorgado. En la seguridad de Windows NT y Windows 2000, los permisos sobre los objetos operan de la misma forma. DENY es el permiso más restrictivo y prevalece sobre cualquier otro.

Dentro del sistema administrador de SQL Server, no todos los objetos tienen permisos. Algunos objetos, por su naturaleza o propósito no pueden ser compartidos o accesados, a pesar de que algunos de ellos tienen algunas funciones para operar los datos, las cuales usualmente solo los ejecutadas por su dueño o por procesos internos.

El acceso a los objetos de la base de datos se provee a través de un sistema de permisos. En la parte más alta de la cadena de objetos, la base de datos no se pone inmediatamente disponible para los usuarios. Antes de que los usuarios puedan accesar la base de datos, el dueño de ésta debe primero otorgar explícitamente acceso a ella. Esto se logra permitiéndole a los grupos de usuarios tener acceso a la base de datos, y permisos que se pueden configurar interactivamente a través del Enterprise Manager o vía algunas acciones programables. Existen tres tipos de permisos que se pueden aplicar a los objetos de SQL Server:

- Permisos de objetos: si un usuario puede accesar un objeto esto significa que para accesar al objeto debe existir un sistema de permisos que gobierna su acceso
- Permisos de estatutos: controla el acceso a los estatutos T-SQL que crean objetos de la base de datos o que causan que el SQL Server ejecute actividades específicas centradas en la base de datos. Mientras una base de datos es un objeto, SQL Server 2000 clasifica los permisos que aplica a ésta como permisos de estatuto. Los permisos de estatuto también aplican a los procedimientos almacenados, a las funciones, tablas, vistas y reglas. Entre estos se tienen: BACKUP DATABASE, BACKUP LOG, CREATE DATABASE, CREATE TABLE, CREATE VIEW, CREATE DEFAULT, CREATE FUNCTION, CREATE PROCEDURE y CREATE RULE
- Permisos implicados: aplican a un amplio rango de actividades y sólo se pueden poner disponibles en forma implícita a los miembros de roles predefinidos en el sistema administrador de la base de datos, o

por los dueños de la base de datos. Tales permisos habilitan el uso de un permiso implícito que se hereda del dueño de la base de datos en virtud de su propiedad.

2.3.7. Consideraciones de seguridad y control en los depósitos de datos, según el planteamiento de Slemo Warigon.

Los requerimientos de seguridad de un ambiente de depósito de datos no son diferentes de los otros sistemas distribuidos. Por lo que, tener un mecanismo de control interno para asegurar la confidencialidad, integridad y disponibilidad de los datos en un ambiente distribuido, es de suma importancia. Desgraciadamente, la mayoría de los depósitos de datos se construyen como poca o ninguna consideración de seguridad durante las fases del desarrollo.

Slemo Warigon[43] plantea que lograr requerimientos proactivos de seguridad para un depósito de datos es un proceso que consta de siete fases:

1. identificar los datos
2. clasificar los datos
3. cuantificar el valor de los datos
4. identificar las vulnerabilidades de seguridad de los datos
5. identificar las medidas de protección de los datos y su costo
6. seleccionar las medidas de seguridad que resulten favorables en la relación costo-beneficio
7. evaluar la efectividad de las medidas de seguridad

A continuación una descripción de cada una de esas fases.

2.3.7.1. Fase 1. Identificación de los datos.

Se deben identificar todos los datos almacenados digitalmente dentro de la empresa que están puestos en el depósito de datos. Esta es una situación que frecuentemente se ignora, pero que se vuelve crítica para reunir los requerimientos de seguridad de un ambiente de depósito de datos desde sus inicios y hacia las etapas subsecuentes. Ello implica tener un inventario completo de los datos que están disponibles para los usuarios finales del depósito. El software de monitoreo de datos que se tenga instalado (que se

[43] [WARI1997] páginas. 3-7.

considera un importante componente del depósito), puede proveer una información precisa acerca de todas las bases de datos, tablas, columnas, filas de datos y perfiles de datos residentes en el ambiente de depósito tanto como quién está utilizándolos y qué tan frecuente hace uso de ellos.

Un procedimiento manual requeriría la preparación de una lista de chequeo de la misma información descrita anteriormente. Sin importar la forma como se colecte la información, sea por un medio automatizado o bien por un método manual, dicha información necesita ser organizada, documentada y conservada para la próxima fase.

2.3.7.2. Fase 2. Clasificación de los datos.

La clasificación de todos los datos contenidos en un ambiente de depósito de datos es necesaria para satisfacer, de una manera prudente, los requerimientos de confidencialidad, integridad y disponibilidad de los datos. En algunos casos, la clasificación de los datos puede obedecer a un requerimiento legal. Para realizar esta tarea se requiere el involucramiento de los dueños, custodios y usuarios finales de los datos. Los datos se clasifican generalmente sobre la base de la criticidad o sensitividad ante la exposición indebida, modificación o destrucción. La sensitividad de los datos de una empresa se puede clasificar en:

- públicos (los datos menos sensitivos): los datos de esta categoría usualmente no se clasifican y están propensos a la exposición pública por leyes, prácticas comerciales comunes, o políticas de la compañía. Todos los niveles de usuarios finales del depósito de datos pueden tener acceso a ellos
- confidenciales (datos moderadamente sensitivos): los datos de esta categoría no están sujetos a la exposición pública. Para la clasificación de los datos de esta categoría se aplica el principio del mínimo privilegio, y el acceso a ellos se limita sobre la base de la necesidad de conocer (*need-to-know*). Los usuarios sólo pueden tener acceso a estos datos si éstos son necesarios para realizar exitosamente su trabajo.
- muy confidenciales (los datos más sensitivos): los datos de esta categoría son altamente sensitivos y de misión crítica. El principio del mínimo privilegio también aplica a esta categoría (con requerimientos de acceso mucho más exigentes que para los datos confidenciales). Sólo los usuarios del más alto nivel del depósito de datos (es decir, con acceso ilimitado), con los permisos de seguridad apropiados pueden accesarlos. Los usuarios pueden accesar sólo los datos necesarios para cumplir con sus deberes críticos.

Sin importar cuáles sean las categorías que se utilicen para clasificar los datos sobre la base de sensitividad, la meta universal de clasificación de datos es establecer rangos para las categorías por incremento de los grados de criticidad, de modo que diferentes medidas de protección se puedan utilizar para diferentes categorías. Clasificar los datos en categorías diferentes no es tan fácil como parece. Ciertos datos representan una mezcla de dos o más categorías dependiendo del contexto utilizado (como por ejemplo, tiempo, localización geográfica, o las leyes que los afectan). Por lo tanto, determinar cómo clasificar estas clases de datos no sólo es interesante, sino también un reto.

2.3.7.3. Fase 3. Cuantificación del valor de los datos.

En la mayoría de las organizaciones la administración superior demanda ver el humo en las pistolas (es decir, gráficas de costos versus beneficios, o evidencia tangible de la comisión de fraudes), antes de girar fondos de la empresa para apoyar iniciativas de seguridad. Algunos administradores cínicos se aprestan a indicar que ellos tratan con realidades tangibles y no con variables etéreas concebidas hipotéticamente. Cuantificar el valor de las medidas protectoras que garanticen datos sensitivos es tan próximo al humo de las pistolas como obtener el apoyo y compromiso de la administración superior para iniciativas de seguridad en ambientes de depósito de datos.

El proceso de cuantificación trata principalmente de asignar un valor real (o valor en la calle) de los datos agrupados en las diferentes categorías de sensitividad. Por sí mismo, los datos no tienen un valor intrínseco. Sin embargo, el valor definitivo de los datos es frecuentemente medido por el costo de:

a) reconstruir datos perdidos,
b) restaurar la integridad de datos corruptos, interceptados o fabricados,
c) no tomar a tiempo una decisión a causa de la falta de servicio, o
d) pagar obligaciones financieras por la exposición pública de datos confidenciales.

El valor de los datos puede incluir ingresos dejados de recibir por causa de la fuga de secretos empresariales hacia sus competidores, y el uso anticipado de datos financieros secretos por parte de empleados deshonestos, antes de que éstos se hagan públicos en el mercado.

Medir el valor de sensitividad de los datos es frecuentemente una tarea titánica. Algunas organizaciones utilizan procedimientos simples para la medición del valor de los datos. Ellos preparan una hoja electrónica en la

que utilizan factores cualitativos y cuantitativos para estimar confiablemente la esperanza de pérdida anual de los datos en riesgo. Por ejemplo, si el costo anual es de $10,000 (basado en el número de horas) para reconstruir los datos clasificados como muy confidenciales y con un factor de riesgo asignado de 4, entonces la compañía podría esperar perder al menos $40,000 en un año si los datos con esta categoría no se protegen adecuadamente. Similarmente, si un empleado espera demandar exitosamente a la compañía y recuperar $250,000 en daños punitivos por la exposición pública de información personal protegida por privacidad, entonces los costos por responsabilidad más las costas legales pagadas a los abogados pueden ser utilizados para calcular el valor de los datos. El factor de riesgo (es decir, la probabilidad de ocurrencia) se puede determinar arbitrariamente o cuantitativamente. A mayor probabilidad de ataque de una unidad de datos particular, mayor es el factor de riesgo asignado a este conjunto de datos.

La medición del valor de la información estratégica basada en categorías de clasificación aceptadas puede ser utilizada para mostrar cómo una organización puede recuperar (en el caso de la tasa interna de retorno, TIR) si los activos son protegidos apropiadamente, o perder (en el caso de pérdida anual de dinero) si no se actúa para proteger los activos valiosos.

2.3.7.4. Fase 4. Identificación de las vulnerabilidades de los datos.

Esta fase requiere la identificación y documentación de las vulnerabilidades asociadas al ambiente de depósito de datos. Algunas de las vulnerabilidades de los depósitos de datos más comunes incluyen las siguientes:

- Seguridad incluida en el sistema administrador de bases de datos: la mayoría de los depósitos de datos confían plenamente en la seguridad que éstos proveen, principalmente basada en el estatuto VIEW. La seguridad basada en este estatuto es inadecuada para los depósitos de datos debido a que ésta puede ser burlada fácilmente si se hace un vaciado (*dump*) directamente de los datos. Tampoco protege a los datos durante la transmisión de estos entre el servidor y los clientes (facilitando una exposición de los datos a accesos no autorizados). La seguridad es igualmente inefectiva para ambientes de depósito de datos en donde las actividades de los usuarios finales no son fácilmente predecibles.
- Limitaciones del sistema administrador de bases de datos: no todos los sistemas de bases de datos que sustentan los datos de un depósito tienen la capacidad de manejar concurrentemente los datos a diferentes niveles de sensitividad. La mayoría de organizaciones, por ejemplo, utilizan un servidor de depósito de datos para procesar los datos confidenciales y muy confidenciales a la vez. Sin embargo, los programas que manejan datos de la más alta seguridad puede que no prevengan la falta de datos

de los programas que manejan información confidencial, así como que usuarios del depósito a los que se les ha limitado el acceso a sólo datos confidenciales no se les prevenga de accesar datos muy confidenciales.

- Bloque de seguridad dual: algunos depósitos combinan las características de seguridad incorporadas en el sistema administrador de bases de datos con el control de acceso de los sistemas operativos para satisfacer sus requerimientos de seguridad. Utilizar un bloque de seguridad dual tiende a presentar la oportunidad de anular la seguridad e incrementar la complejidad de la administración de la seguridad en un ambiente de depósito de datos.

- Ataques por inferencia: diferentes privilegios de acceso se conceden a diferentes usuarios del depósito de datos. Todos los usuarios pueden accesar la información pública, pero se presume que sólo unos pocos pueden accesar información confidencial o ultra secreta. Desdichadamente, usuarios generales pueden accesar información protegida por medio de la inferencia sin que para ello tengan un acceso directo a los datos protegidos. Los datos sensitivos son típicamente inferidos de información que parece ser no sensitiva. Llevar a cabo ataques de inferencia directa o indirectamente es una vulnerabilidad muy común en el ambiente de depósito de datos.

- Factor de disponibilidad: la disponibilidad es un requerimiento crítico sobre el que se construye la filosofía de acceso compartido de la arquitectura de un depósito de datos. Sin embargo, la disponibilidad puede entrar en conflicto o comprometer la confidencialidad y la integridad de los datos en el depósito si no se considera ésta cuidadosamente.

- Factores humanos: actos accidentales o intencionales, tales como errores, omisiones, modificaciones, destrucción, mal uso, divulgación, sabotaje, fraudes y negligencia, se cuentan entre las mayores causas de pérdidas cuantiosas en que incurren las organizaciones. Este tipo de actos afectan adversamente la integridad, confidencialidad y disponibilidad de los datos de un depósito.

- Amenazas internas: los usuarios (empleados) representan la más grande amenaza contra los datos valiosos. Empleados disgustados con acceso legítimo podrían revelar datos secretos a las empresas competidoras y divulgar públicamente ciertos datos confidenciales del recurso humano de la empresa. Empleados deshonestos podrían también obtener un beneficio personal al utilizar datos estratégicos de la empresa antes de que ésta se haga pública en el mercado. Estas actividades causan:

  a) tensas relaciones con los demás miembros del sector, así como con entidades gubernamentales,

  b) pérdidas económicas por asumir responsabilidades financieras,

  c) pérdida de confianza en la organización, y

d) pérdida de liderazgo.

- Amenazas externas: los competidores y terceros tienen en el ambiente de depósito de datos amenazas similares a las que pueden representar empleados que actúan en forma no ética. Estos foráneos utilizan espionaje electrónico y otras técnicas similares para robar, comprar o reunir datos estratégicos en un ambiente de depósito de datos. Entre los riesgos de este tipo de actividades se incluyen:

  a) publicidad negativa la cual destruye la capacidad de la compañía de atraer y retener clientes o mercados compartidos, y

  b) pérdida de continuidad de los recursos del depósito, lo cual niega la productividad de los usuarios.

  La pérdida resultante tiende a ser mucho más alta que la provocada por ataques internos.

- Factores naturales: daños provocados por el fuego, el agua y el aire pueden hacer que tanto los servidores como los clientes de un ambiente de depósito de datos se vuelvan inutilizables. Los riesgos y pérdidas varían de organización en organización, dependiendo mayormente de la ubicación y los factores de contingencia.

- Factores de utilidad: la interrupción del fluido eléctrico y del servicio de comunicaciones pueden causar costos disturbios en un ambiente de depósito de datos. Estos factores tienen una muy baja probabilidad de ocurrencia, pero tienden a causar pérdidas cuantiosas.

Un inventario completo de las vulnerabilidades inherentes en el ambiente de depósito de datos es necesario que se documente y organice (o sea, de mayor a menor) para la próxima fase.

2.3.7.5. Fase 5. Identificación de medidas de protección y su costo.

Las vulnerabilidades identificadas en la fase anterior deben ser consideradas en la medida de determinar las protecciones que resulten favorables en la relación costo-beneficio para los datos del depósito de datos a diferentes niveles de sensitividad. Algunas medidas de protección para los datos en el depósito incluyen:

- La barrera humana: los empleados representan el frente de defensa contra las vulnerabilidades de seguridad en cualquier ambiente de procesamiento centralizado, incluidos los depósitos de datos. Enfocarse en la contratación y entrenamiento del personal (sobre todo en lo que respecta a la conciencia de la seguridad), revisión de antecedentes periódicamente, traslado y despidos, como parte de los requerimientos de seguridad es útil para crear un ambiente de depósito de datos consciente

de la seguridad. Este método efectivamente ataca los orígenes, más que los síntomas de los problemas de seguridad. Los costos de la administración del recurso humano son fácilmente mensurables.

- Clasificación de acceso de los usuarios: se deben clasificar los usuarios del depósito de datos como:

  a) usuarios de acceso general,

  b) usuarios de acceso limitado, y

  c) usuarios de acceso ilimitado,

  para poder decidir en cuanto al control de acceso.

- Controles de acceso: se deben utilizar políticas de control de acceso basadas en los principios del mínimo privilegio y la protección adecuada de los datos. Se deben establecer restricciones de los controles de acceso efectivos y eficientes de modo que los usuarios finales puedan accesar sólo los datos y programas para los cuales ellos tienen los legítimos privilegios. Los datos corporativos deben ser protegidos al grado de consistencia que tiene su valor. Los usuarios necesitan obtener un permiso de seguridad granular antes de que se les conceda el acceso a datos sensitivos. Además, el acceso de los datos sensitivos debería de depender de más de un mecanismo de autenticación. Estos controles de acceso minimizan los daños provocados por ataques accidentales o maliciosos.

- Controles de integridad: se debe utilizar un mecanismo de control que:

  a) prevenga a todos los usuarios de actualizar y borrar datos históricos en el depósito de datos,

  b) restringir el acceso de mezclar datos a tan sólo actividades autorizadas,

  c) inmunizar los datos del depósito contra fallos en el suministro de energía eléctrica, caídas del sistema y corrupción de éstos,

  d) habilitar la rápida recuperación de los datos y las operaciones ante desastres, y

  e) asegurar la disponibilidad consistente, confiable y a tiempo de los datos para los usuarios.

  Todo ello se logra a través de controles de integridad del sistema operativo, así como con procedimientos de recuperación de desastres bien probados.

- Encripción de datos: se deben cifrar los datos sensitivos del depósito para asegurar que éstos sólo sean accesados por las personas autorizadas. Ello anula el valor potencial de los datos interceptados, así como la fabricación o modificación de éstos. También inhibe el vaciado (*dumping*) no autorizado e interpretación de los datos, habilitando una

autenticación segura de los usuarios. En breve, el ciframiento de los datos asegura la confidencialidad, integridad y disponibilidad de los datos en un ambiente de depósito de datos.

- Particionamiento: se debe utilizar un mecanismo de particionamiento de los datos sensitivos en tablas separadas, de modo que sólo los usuarios autorizados tengan acceso a estas tablas, basándose para ello en las legítimas necesidades de los usuarios. El esquema de particionamiento puede descansar en la simple característica de seguridad incorporada en los sistemas administradores de bases de datos para prevenir el acceso no autorizado a los datos sensitivos de un ambiente de depósito de datos. Sin embargo, el uso de este método presenta algunos problemas de redundancia.
- Controles en el desarrollo: se deben utilizar algunos controles estándares de calidad para guiar el desarrollo, prueba y mantenimiento de la arquitectura del depósito de datos. Este método asegura que los requerimientos de seguridad están suficientemente dirigidos durante y después de la fase de desarrollo de éste. También asegura que el sistema es altamente elástico (es decir, adaptable o que responde a cambios en las necesidades de seguridad).

El costo estimado de cada una de estas medidas de seguridad se debe determinar y documentar para la próxima fase. Algunos paquetes comerciales y aplicaciones internas pueden ayudar en la identificación de las medidas proactivas apropiadas para vulnerabilidades conocidas, y a cuantificar los costos asociados o el impacto fiscal. La medición de los costos usualmente involucra determinar los costos del desarrollo, implementación y mantenimiento de cada una de las medidas de seguridad.

2.3.7.6. Fase 6. Selección de medidas de seguridad con una relación costo-beneficio favorable.

Todas las medidas de seguridad implican gastos, y los gastos en seguridad requieren justificación. Esta fase descansa en los resultados de las fases anteriores para valorar el impacto fiscal de los datos corporativos en riesgo, y seleccionar las medidas que resulten favorables en su relación costo-beneficio para salvaguardar los datos contra vulnerabilidades conocidas. Seleccionar medidas de seguridad que resulten favorables en su relación costo-beneficio es congruente con una práctica del negocio prudente la cual asegure que los costos de protección de los datos en riesgo no excedan la pérdida esperada máxima para los datos. La administración superior, por ejemplo, juzgar si es prudente asignar $500.000 anualmente en salvaguardar datos cuya pérdida esperada máxima anual sea de $250.000.

Sin embargo, el factor costo podría ser el único criterio para seleccionar las medidas de seguridad apropiadas en un ambiente de depósito de datos. La compatibilidad, adaptabilidad e impacto potencial en el desempeño del depósito de datos podrían también ser tomadas en consideración. Adicionalmente, existen otros dos factores importantes. Primero, el principio de la economía del mecanismo dicta que una medida de protección simple y bien probada puede ser utilizada para controlar múltiples vulnerabilidades en un ambiente de depósito de datos. Segundo, los datos, al igual que el hardware y el software, son un elemento en el escenario de seguridad de los sistemas de información, y que tienen un período de vida muy corta. Entonces, el principio de la protección adecuada de los datos dicta que los datos de un depósito pueden ser protegidos con medidas de seguridad que sean suficientemente efectivas y eficientes para la corta vida de los datos.

2.3.7.7. Fase 7. Evaluación de la efectividad de las medidas de seguridad.

Una estrategia de seguridad acertada supone que todas las medidas de seguridad son quebrantables y no permanentemente efectivas. Cada vez que se identifica y selecciona una medida de seguridad que resulta favorable a su relación costo-beneficio para asegurar la información estratégica contra ciertos ataques, el atacante tiene a doblar sus esfuerzos en identificar métodos para doblegar la medida de seguridad implementada. La mejor que se puede hacer es prevenir que esto suceda, haciendo que el próximo ataque sea más difícil, o prepararse rápidamente por si la información es atacada. No se puede estar bien posicionado para llevar a cabo estas acciones si no se ha evaluado la efectividad de las medidas de seguridad sobre una base actualizada.

Para evaluar la efectividad de las medidas de seguridad se deben dirigir los esfuerzos para determinar continuamente si las medidas son:

1) pequeñas, simples y directas,
2) analizadas, probadas y verificadas cuidadosamente,
3) utilizadas y seleccionadas apropiadamente de modo que no excluyan los accesos legítimos,
4) elásticas de modo que puedan responder efectivamente a cambios en los requerimientos de seguridad, y
5) razonablemente eficientes en términos de tiempo, espacio de memoria, y actividades centradas en el usuario, de modo que no afecten adversamente a los recursos computacionales protegidos.

Es igualmente importante asegurar que los usuarios finales del depósito de datos entiendan y adopten la responsabilidad de las medidas de seguridad a

través de un programa de concienciación de seguridad efectivo. El administrador del depósito de datos con la autoridad delegada por la administración superior es el responsable por asegurar la efectividad de las medidas de seguridad.

Capítulo 3. Identificación de las principales amenazas que atentan contra un depósito de datos y su correspondiente análisis del impacto

3. Identificación de las principales amenazas que atentan contra un depósito de datos y su correspondiente análisis del impacto.

3.1. Introducción al análisis de amenazas.

Al igual que sucede con cualquier otra forma de procesamiento de datos, los depósitos de datos también se ven expuestos a amenazas. Existen diferentes clasificaciones de ellas, las cuales dependen del punto desde donde se le observe, sean éstas internas o externas, intencionales o accidentales, provocadas por el hombre o por la Naturaleza.

En una de las clasificaciones, las amenazas se pueden considerar provocadas por agentes externos o internos a la organización que sufre su materialización.

Se dice que una amenaza es externa cuando el factor que la provoca es ajeno a la empresa que sufre las consecuencias de su materialización. Por su parte, una amenaza es interna cuando el elemento que la provoca yace dentro de la misma empresa.

Otra de las clasificaciones agrupa las amenazas según la intención o premeditación de su materialización, agrupándolas en intencionales o accidentales.

Se dice que una amenaza es intencional cuando existe premeditación por parte de la persona o agente que provoca la materialización de la amenaza; mientras que se dice que una amenaza es accidental cuando la ocurrencia del evento se debe a una casualidad o a un hecho fortuito.

Una más de las clasificaciones conocidas agrupa las amenazas según sean originadas por la Naturaleza, o bien por el hombre.

Dentro de las amenazas provocadas por la Naturaleza se incluyen los terremotos, inundaciones, tormentas eléctricas, entre otras. Por su parte, dentro de las amenazas provocadas por el hombre se tienen sabotaje, robo, secuestro, suplantación, entre otras.

A manera de ejemplo, a continuación se presentan dos tablas. En la primera de ellas se muestran las posibles consecuencias que se pueden tener en un ambiente computacional, según sea la intencionalidad de la amenaza (intencionales o no intencionales), y de acuerdo con las consecuencias típicas de una materialización de éstas (daño, alternación y diseminación o revelación).

Acciones	Consecuencias		
	Daños	**Alteraciones**	**Diseminación**
Intencionales	fuego; robo de equipos, cintas, discos, casetes y disquetes; destrucción física de los datos y programas almacenados en cintas y discos; interrupción en la prestación de las facilidades que provee la computadora; motín, sabotaje, etc.	agregado, cambio o borrado de alguna porción de los archivos que contienen los datos o programas con propósitos personales, por diversión o por venganza	venta de información con propósitos personales, por diversión o por venganza
No intencionales	catástrofes naturales: tales como incendio, inundación, terremoto; daño accidental de cintas, discos, casetes, disquetes, pérdida de las facilidades proveídas por la computadora, etc.	agregado, cambio o borrado accidental de alguna porción de los archivos que contienen los datos o programas; alteración de los datos ocasionada por el mal funcionamiento de algún componente de hardware o software	mostrar alguna información a colegas; despliegue accidental de contraseñas o salidas que producen los sistemas de información

Tabla 2. Tipos de riesgos en un ambiente computacional, causas y consecuencias.
Fuente: [VALL1989] página 5.

En la segunda tabla se muestran las consecuencias de la materialización de una amenaza según la intencionalidad y en función del agente (principalmente las personas) u objeto (componentes físicos, telecomunicaciones, datos, aplicaciones de software, software del sistema y operaciones de la computadora) de riesgo propio de un ambiente computacional.

Acciones	Agentes de riesgo: personas	Objetos de riesgo					
		Físicos	**Telecomu-nicaciones**	**Datos**	**Aplicaciones de software**	**Software del sistema**	**Operaciones de la computadora**
Intencional	Daño Alteración Diseminación	Daño Alteración	Daño Alteración	Daño Alteración Diseminación	Daño Alteración Diseminación	Daño Alteración Diseminación	Daño Alteración
No intencional	Daño Alteración Diseminación	Daño Alteración	Daño Alteración	Daño Alteración Diseminación	Daño Alteración Diseminación	Daño Alteración Diseminación	Daño Alteración

Tabla 3. Agentes y objetos de riesgo en un ambiente comp utacional.
Fuente: [VALL1989] página 6.

En lo concerniente a la identificación de las amenazas a que se ven expuestos los depósitos de datos, se utilizará el esquema que plantean Kimball et al. en su capítulo 12, *A Graduate Course on the Internet and Security.*[44], el cual ofrece un análisis particular de las posibles consecuencias que se pueden derivar de tales amenazas.

Para lo que es el análisis de las amenazas identificadas, en la última sección de este capítulo se hará el respectivo análisis de éstas, agrupándolas según el área en la que corresponde actuar para minimizar su impacto.

Dentro de la identificación de las amenazas se considerarán aquéllas que afectan a los **activos físicos**, tales como sustracción o robo de éstos, destrucción intencional, incendio, humedecimiento, inmersión en agua, suciedad o acumulación de mugre, envejecimiento, problemas derivados de descargas o interferencias eléctricas, consecuencias de disturbios magnéticos, pérdida o extravío por descuido, pérdida por obsolescencia tecnológica, y consecuencias derivados del secuestro o pirateo (*hijack*) de tales activos.

Posteriormente se considerarán las amenazas que afectan directamente a los **datos** o **información** contenida en los depósitos de datos. Para ello, se considerarán situaciones tales como divulgación (en diversos aspectos, planes confidenciales, códigos, información confiada a la empresa, información sensitiva), pérdida de oportunidad, robo o sustracción de activos financieros, robo o sustracción de servicios, robo o sustracción de información con el fin de promover violencia o terrorismo, robo de la identidad de los clientes o usuarios del depósito, robo o pérdida de privacidad.

Puesto que el robo o sustracción se puede perpetrar de diferentes formas y con diversos propósitos, se hará una presentación de las **modalidades** más conocidas **de robo** y su repercusión en el uso de los depósitos de datos. Dentro de éstas se pueden citar el arrebato oportunista (*opportunistic snatching*), la divulgación inadvertida (*inadvertent broadcasting*), la escucha disimulada (*eavesdropping*), el robo físico de componentes como medio de robar información contenida en ellos, el pirateo de sesiones (*hijacked session*), la suplantación (*impersonation*), el uso de puertas traseras (*trapdoor*) y el soborno (*bribery*), asalto (*robbery*) y extorsión (*extortion*).

Como parte de las amenazas que procuran la **modificación de la información**, se considerarán aspectos tales como el desvío de la información que fluye a través de las redes de comunicación, el mal uso de la información enviada con el propósito de cometer fraudes o para desprestigiar al emisor de ésta, y el rechazo de la autoría de cierta información. Asimismo, se considerarán las implicaciones que se derivan del **robo del software**, entre las cuales se pueden citar el robo del código objeto, el robo del código fuente, la modificación del software con el objeto de propiciar un control "pirateado" (*hijacked control*), inducir a que se comprometa el proceso de certificación del software y la implantación o propagación de virus.

[44] [KIMB1998] páginas 449-500.

Finalmente se hará una identificación de aquellas amenazas que atentan contra el **buen funcionamiento de la empresa**, tales como los ataques de negación del, la inhabilidad de reconstruir puntos consistentes y los intentos de terrorismo.

3.2. Identificación de las amenazas.

3.2.1 Vulnerabilidades de los activos físicos.

3.2.1.1. Robo.

Cuando se habla de robo se está haciendo referencia a la pérdida o sustracción de un activo, en donde algún individuo lo ha tomado para llevárselo fuera de las instalaciones de la empresa, y no pretende devolverlo.

Para un depósito de datos tal vez el activo que más repercusión tiene que se vea afectado por esta amenaza es la pérdida o sustracción de una computadora personal o estación de trabajo, por cuanto en ella se puede tener almacenada información de muy diferente índole relacionada con el depósito. Por ejemplo, ésta puede contener información resultante de utilizar el depósito, o sea, que contenga información que ayude a apoyar el proceso de toma de decisiones, en cuyo caso, ello expondría a que personas ajenas a la empresa cuenten con información vital para la empresa y su proceso de toma de decisiones. También se podría suceder que la computadora contenga información que trate acerca de privilegios de ingreso, o incluso contraseñas para ingresar a un depósito de datos.

Asimismo, se podría dar el caso de que la computadora sustraída contenga contraseñas de encripción de un depósito de datos, las cuales no se encuentran almacenadas en ningún otro lugar, como podría ser el caso de la estación de trabajo de administrador de depósito, o bien, que por razones de seguridad no se quiera que de tales contraseñas se mantengan copias.

Otra consecuencia podría ser que la computadora contenga llaves privadas utilizadas para la autenticación de usuarios de un depósito de datos o certificaciones de los resultados de un depósito de datos que le han sido confiados a la empresa. Ya en términos generales, ciertos activos pueden contener metadatos necesarios para la operación del depósito de datos. En general, la pérdida por robo acarrea la posibilidad de que el contenido de la información caiga en manos hostiles, llegando a requerir que se deban

cambiar métodos de acceso, procedimientos y hasta personal para minimizar la ocurrencia de este tipo de amenaza.

3.2.1.2. Destrucción intencional.

Esta amenaza corresponde a un ataque deliberado en contra de un activo físico. Existe un sinfín de formas por medio de las cuales se puede destruir un activo físico por parte de un determinado individuo, lo que incluye chocarlo o golpearlo con un instrumento pesado, prenderle fuego, sumergirlo en un líquido, palanquearlo con desatornillador, lanzarlo desde una ventana, por citar algunas formas de intentar este cometido. Entre las consecuencias que se pueden derivar de este tipo de amenaza se puede citar la pérdida de datos de un depósito, la pérdida de contraseñas, la pérdida de metadatos que soportan las operaciones del depósito de datos, la pérdida de respaldos, la inhabilidad para recuperar información, lo que a la postre puede ocasionar una interrupción significativa en la prestación de los servicios, o bien, que se utilice para disfrazar la comisión de otros delitos, en cuyo caso lo que se estaría haciendo es distrayendo la atención o destruyendo la evidencia de tales acciones.

3.2.1.3. Incendio.

Esta amenaza se puede ver como el producto de un incendio accidental o provocado, que afecte la forma específica a los sistemas de información o que sea parte de un incendio mayor que afecte a toda la empresa. En este caso se deben considerar los daños ocasionados no sólo por el humo producto del incendio, sino también aquellos que se deriven de la utilización del agua como medio para combatir al incendio mismo. Aunque en términos generales las consecuencias de esta amenaza podrían ser las mismas que se enunciaron para el caso de una destrucción intencional, se debe considerar la posibilidad de que la afectación pueda ser más extensiva, lo cual pueda afectar tanto a los sistemas primarios como aquéllos de respaldo.

3.2.1.4. Humedecimiento.

Esta amenaza se refiere a la degradación que sufren los activos físicos a causa de un almacenamiento inapropiado, por medio del cual se facilita la acumulación de humedad. Entre las consecuencias que se derivan de esta amenaza se pueden citar la inhabilidad para recuperar viejas versiones de los datos; pérdida que probablemente se vuelva inadvertida hasta que es demasiado tarde para proveer la protección requerida.

3.2.1.5. Agua.

Al mencionar el agua como una amenaza que atenta contra los activos físicos se está haciendo alusión a la destrucción de un activo físico producto de la inundación, ya sea porque se sumerja al activo en el agua, o bien, porque el agua gotee sobre el activo.

Para comprender las consecuencias que se derivan de esta amenaza se debe tener en cuenta que muchos dispositivos portátiles de alta tecnología, tales como las computadoras personales portátiles, se vuelven inmediatamente inservibles cuando se mojan con agua. Si se comparan las consecuencias de esta amenaza con las del robo, se puede decir que son menos dañinas que las de este último debido a que la información tan sólo se pierde y no se compromete (es decir, que no se expone ante otros innecesariamente). En este sentido sus consecuencias se asemejan a las de una destrucción intencional o a las del incendio.

3.2.1.6. Suciedad.

Esta amenaza se refiere a la destrucción de un activo a causa de un almacenamiento y manejo inapropiado. La suciedad (mugre o impurezas) se filtra en los computadores, lo que puede provocar que los circuitos entren en corto o que las cabezas de los disquetes se estrellen contra éstos. Las consecuencias de esta amenaza generalmente resultan ser las mismas del caso de humedecimiento, con la única diferencia de que esta amenaza puede incluir la probabilidad de gastar tiempo y dinero tratando de usar y reparar componentes dañados.

3.2.1.7. Envejecimiento.

Esta amenaza se puede ver como la destrucción de un activo a causa de que éste se gaste, o sea, que alcance el fin de su vida útil, volviéndose frágil, o que gradualmente se altere químicamente. Algunas veces ello puede suceder como resultado que se le haya dejado al sol, o porque se haya manejado inadecuadamente. Sus consecuencias resultan ser las mismas que las que se producen por efecto del humedecimiento.

3.2.1.8. Descargas o interferencias eléctricas.

Al tratar este tipo de amenaza se debe tener en mente que ésta se refiere a la destrucción de un activo a causa de que no se le protegió contra sobretensión en el circuito, picos o apagones. También incluye los efectos de descargas causadas por rayos o relámpagos, así como por equipos tales como sistemas de seguridad en los aeropuertos. Entre las consecuencias de la materialización de una amenaza como esta se pueden mencionar la pérdida de un activo que esté en línea con el depósito de datos tal como una máquina cliente o un servidor, o bien, algún activo que esté en línea o que posea datos volátiles y que no se haya respaldado.

3.2.1.9. Disturbios magnéticos.

Esta amenaza considera la destrucción de un activo al ser expuesto a un campo magnético, tal como un potente parlante o un sistema de seguridad. Los discos duros, disquetes y otros medios magnéticos se pueden ver afectados por un campo magnético destructivo, lo que puede provocar que se pierdan completamente los datos contenidos en ellos y que no se pueden recuperar. Esta amenaza atenta en contra del almacenamiento de la información y el eventual traslado de ésta.

3.2.1.10. Pérdida por obsolescencia tecnológica.

La obsolescencia tecnológica puede tomar lugar en cualquier parte de un sistema: hardware, software o procedimientos. Por lo general la obsolescencia se da porque no existen más repuestos para reparar el activo, a pesar de que éste aún esté en condiciones de ser utilizado se si le repara. Aún los medios de almacenamiento que han sido cuidados e inventariados apropiadamente se pueden perder si el hardware, software y procedimientos alrededor de su uso no se han mantenido actualizados. Esta clase de pérdida constituye un riesgo alto para los propietarios de un depósito de datos. Los viejos datos que se supusieron estar disponibles no lo pueden estar por esta causa, o bien porque los datos actuales tienen un período de vida mucho más corto que lo planeado.

3.2.1.11. Secuestro o pirateo (*hijack*) de activos.

El secuestro o pirateo de activos se da cuando un intruso toma control total o parcial de una máquina física. Ello se puede ver como una pérdida de

control o de confianza comprometida. Quizás el intruso utilizó el acceso físico para reemplazar un componente de software confiable por uno deshonesto o con fines criminales. Por ejemplo, un *browser* para sitios Web de confianza puede ser reemplazado por otro que revele información que el usuario no desee que se divulgue. O quizás el intruso puede haber puesto una conexión en la intranet local. Un depósito de datos puede ser atacado o comprometido a través del "secuestro" de activos de muchas formas. Un control "pirateado" puede revelar códigos de seguridad que le den al intruso accesos no intencionados. Una conexión física a una red puede lograr el mismo propósito. Estos dos modos de ataque pueden de hecho enviar datos reales producidos por el intruso. Un activo "secuestrado" destruye la confianza en el sistema de depósito de datos.

3.2.2. Vulnerabilidades de la información.

3.2.2.1. Divulgación de planes confidenciales.

Puesto que el propósito de un depósito de datos es soportar el proceso de toma de decisiones, una de las principales vulnerabilidades que presenta es la posibilidad de que se divulguen los planes confidenciales o las decisiones tentativas de la empresa derivadas de la interacción o uso del depósito de datos. El depósito de datos es un lugar conocido para almacenar planes confidenciales tales como presupuestos, proyecciones, y otro tipo de análisis (sobre todo aquellos del tipo *qué pasa si* ...), y por lo tanto, será el objetivo de cualquiera que trate de encontrar estos planes. El depósito de datos puede señalar la presencia de información confidencial en muchas formas. La información detallada es obvia debido a que se le suele etiquetar explícitamente, pero los datos resumidos tales como proyecciones anuales o estadísticas atribuidas a un grupo de personas pueden ser reveladas justo cuando apenas se han obtenido.

3.2.2.2. Divulgación de códigos.

Esta amenaza se refiere a la posibilidad que existe de divulgar los códigos, números de cuenta o contraseñas de los usuarios, que pueden conducir a posteriores robos o acciones de obstrucción. Existen muchos lugares dentro de un depósito de datos donde se almacenan o introducen códigos de seguridad o de acceso. Muchas formas de metadatos se almacenan en el depósito de datos y pueden contener códigos explícitos. Muchas tablas de datos primarios de un depósito de datos contienen información sensitiva

acerca de los individuos tales como números de tarjetas de banco o datos personales acerca del usuario, los cuales constituyen alguna clase de códigos.

3.2.2.3. Divulgación de información dada en confianza a la empresa.

Esta amenaza se refiere a la situación en que se pueda dar una divulgación de información confidencial que se mantiene en confianza en el depósito de datos por encargo de un tercero, tal como números de tarjetas de crédito, números de cédula, información médica, o datos privados del negocio que le pertenecen a un tercero. Esta clase de exposición de la información es especialmente nociva no sólo porque se pueden ver afectados directamente los propietarios de ésta, sino también la propia empresa que actúa como custodio de confianza de la información, pues probablemente sea responsable por la pérdida y quizás por los daños adicionales. También se puede perder la relación que se tiene con el tercero. Si esto no fuera suficiente, también cabe la responsabilidad que le cabe al funcionario que directamente trataba los datos, lo que lo podría conducir a perder su trabajo si es por su causa que los datos fueron robados.

3.2.2.4. Divulgación de información sensitiva.

Con esta amenaza se desea cubrir la divulgación de aquella información que se considera política, ética o legalmente sensitiva, y que la empresa tiene almacenada, aún cuando esté o no verdaderamente facultada para tenerla. También considera la posibilidad de que esta información sea usada para realizar chantajes. El depósito de datos puede estar en una delicada posición de almacenar datos que realmente no se debieran almacenar o mantener por mucho tiempo. Si el depósito de datos contiene información sensitiva desde el punto de vista político, ético o legal, éste podría ser el blanco de acciones legales tales como citaciones. Tales datos también podrían ser el blanco principal de intrusos. Rastrear el comportamiento de los individuos a través de un depósito de datos tiene un conjunto de serias exposiciones que van desde simples objeciones de privacidad hasta el chantaje, pasando por requerimientos de privacidad legalmente protegidos.

3.2.2.5. Eliminación de protección o de oportunidad.

Esta amenaza trata acerca de la eliminación, en un sentido financiero, de una oportunidad que se pueda tener con los planes de producción, planes de inversión o conocimiento de otros eventos en el futuro. También en el mismo sentido, la divulgación de un secreto industrial que no está protegido

por una patente. Muchos análisis del tipo *qué pasa si* ... toman la forma de pronósticos de prueba o presupuestos de prueba, o cualquier otro tipo de análisis que involucren series de tiempo. Estas series de tiempo pueden ser almacenadas de forma natural en un ambiente de depósito de datos, y por lo tanto, ser el blanco de ataque de intrusos que anden buscando esta información.

3.2.2.6. Robo de activos financieros.

Se trata del robo directo de activos financieros por medio de la transferencia de fondos a otras cuentas o por estar en capacidad de efectuar transacciones financieras significativas utilizando para ello su autoridad. El robo directo de activos financieros podría ser el resultado de comprometer números de cuentas, códigos de acceso u otros elementos, tales como información personal de los usuarios, los cuales se encuentran disponibles en el depósito de datos.

3.2.2.7. Robo de servicios.

Esta amenaza se refiere al robo directo de servicio, como por ejemplo, ante la revelación de códigos de autorización para servicios telefónicos. Los números telefónicos y sus códigos de acceso constituyen una exposición muy obvia si es que el depósito de datos contiene tal información. Los números telefónicos que no aparecen en listados pueden provocar problemas, si ellos dan indicios a un intruso de que el sistema puede ser fácilmente burlado.

3.2.2.8. Robo de información para promover violencia o terrorismo.

Esta amenaza trata acerca del robo de información tal como planes de viaje, que ponen a los individuos en peligro de secuestro o de acciones terroristas. A mayor nivel de detalle con que se grabe cierta información en el depósito de datos acerca del comportamiento de los individuos, mayor es el riesgo de que esa información pueda ser utilizada para promover acciones de violencia o terrorismo. Por ejemplo, las transacciones hechas con tarjetas de crédito pueden ser muy reveladoras en cuanto a los gustos, preferencias y hábitos de los individuos, lo cual puede ser utilizado para determinar su comportamiento.

3.2.2.9. Robo de identidad.

Esta amenaza trata acerca del robo de identidad de una persona en el sentido de que su autoridad, su crédito o su reputación sea utilizado por alguien en contra de su deseo. Esta clase de robo frecuentemente resulta en una pérdida real para los individuos, materializada como la pérdida de activos físicos o financieros debido a la sorprendente dificultad para convencer a los demás de que el uso que se hizo de sus referencias fue fraudulento y de que el individuo es realmente quien dice ser y no el que malintencionadamente otro les hizo hacer creer. El robo de identidad puede requerir sólo unos pocos datos contenidos en el depósito de datos. Entre los principales están los números de cuenta, y elementos de seguridad tales como información personal.

3.2.2.10. Robo de privacidad.

Con esta amenaza se está considerando la posibilidad de que se pierda la privacidad (intimidad) de la persona, de modo que detalles de su vida personal, creencias y actividades se revelen en contra de su deseo. Un ejemplo de este tipo de caso se presenta en la industria, y en general, en cualquier actividad comercial, en donde se tiene un gran reto al intentar balancear el deseo de los departamentos de mercadeo y ventas de querer almacenar datos relacionados con el comportamiento de los clientes en el depósito de datos, contra el deseo opuesto de preservar el anonimato, la dignidad y la seguridad de esos clientes.

3.2.3. Modalidades de robo.

3.2.3.1. Arrebato oportunista (*Opportunistic Snatching*).

En este caso se trata de que la información queda desprotegida del sistema de seguridad por alguna circunstancia. Puede ser que la información esté en una localización oculta y que el atacante se la encuentre sin un intención previa de robársela, o bien, que sabiendo que existe esa información desprotegida la busque hasta dar con ella.

Muchos aspectos relacionados con los depósitos de datos no tienen o no se les percibe una verdadera necesidad de estar del todo protegidos. Por

ejemplo, las tablas temporales o las tablas de agregación que se deriven no tienen tal protección. Algunas veces lo referente a la pertenencia a los grupos de acceso de usuarios primarios no se mantiene de una forma efectiva, lo que permite que se puedan infiltrar usuarios deshonestos en tales grupos. Otras veces, por insuficiencia en los procedimientos internos se puede dar que las empresas contratadas, personas ajenas a la organización y los exempleados se mantienen como miembros de esos grupos. Finalmente, no siempre se da que los medios utilizados para respaldar información tengan la debida seguridad y que sean almacenados en ambientes seguros. Cualquiera de estas circunstancias puede favorecer a que se propicie la materialización de este tipo de amenaza.

3.2.3.2. Difusión inadvertida (*Inadvertent Broadcasting*).

En este caso la información es dejada expuesta accidentalmente, ello como producto de que alguien haya olvidado encriptar el archivo, o no le haya puesto los permisos de acceso; o bien, la información se envió inadvertidamente en un correo electrónico. A pesar de que las consecuencias de esta amenaza son similares a la anterior, la causa más probable es que se haya dado un incumplimiento en el procedimiento. Ejemplo de ello podría ser que el empleado es descuidado o que no entendió cómo usar el sistema. También podría darse el caso de que nadie haya revisado si los archivos compartidos contienen información sensitiva, o bien, que no se haya revisado quién ha estado accesando los archivos sensitivos. Estos problemas son típicamente simples errores de procedimientos causados por funcionarios y administradores que no piensan acerca de la seguridad y las consecuencias que se pueden derivar de una mala implementación de ésta.

3.2.3.3. Escucha disimulada (*Eavesdropping*).

Este tipo de amenaza se caracteriza porque el intruso monta un ataque deliberado al sistema de información y es capaz de interceptar la comunicación que se realiza con otros usuarios. La escucha disimulada incluye la intervención en la Internet o en un intranet local, así como la escucha electromagnética utilizando para ello sofisticado equipo capaz de capturar emanaciones desde cables, pantallas y teclados. También incluye escuchar conversaciones verbales y observar por encima del hombro de las personas cuando éstas usan la computadora. La escucha en una intranet local se puede llevar a cabo fácilmente al equipar una computadora personal con un programa que actúe como un "olfateador" (*sniffer*) de paquetes, el cual analiza todos los paquetes conforme ellos fluyen por el medio de transmisión. Lamentablemente muchas de las herramientas disponibles que comúnmente se utilizan para labores administrativas, que en condiciones

normales tienen un uso legítimo, pueden, por el contrario, ser utilizadas para rastrear correos electrónicos, ingreso de sesiones, y sesiones de búsqueda en Internet, en todos los usuarios remotos conectados a una intranet particular.

Este tipo de ataque puede involucrar a los depósitos de datos en muchas formas obvias, incluyendo el compromiso de contraseñas e información resultante. Puesto que la escucha puede tomar muchas formas, no existe una solución para ello. El primer paso que se puede tomar en este sentido es la concientización de los empleados y administradores acerca de ello. El uso de técnicas de encripción puede reducir muchos de los riesgos de este tipo de ataque.

3.2.3.4. Robo físico como medio para robar información.

En este tipo de amenaza el atacante toma físicamente la computadora, la *laptop*, o las cintas de respaldo en procura de usarlos como forma de obtener la información. Dicha información puede residir directamente en el activo físico que se tomó, o bien, el activo físico puede incluir los medios de autenticación para ganar acceso a la información. El administrador de un depósito de datos debe preocuparse acerca de la seguridad física de las *laptop* que poseen los empleados que tienen un acceso significativo al depósito de datos. Cada vez más, las *laptop* se están convirtiendo en el blanco de robo por lo que ellas contienen (en información) y no tanto por el hardware que la conforman.

3.2.3.5. "Pirateo" de sesión (*Hijacked Session*).

Con esta amenaza el atacante escucha disimuladamente la sesión del usuario hasta después de que se establece la autenticación, y entonces interfiere la conexión entre el usuario y el servidor, tomando control del resto de la sesión, de forma tal que el intruso se vuelve un usuario totalmente autenticado y autorizado.

El administrador del depósito de datos puede protegerse contra algunas formas de pirateo de sesiones al controlar la escucha disimulada, pero podría considerar sistemas como aquéllos protegidos por el esquema Kerberos para ofrecer una protección más efectiva.

3.2.3.6. Imitación o suplantación (*Impersonation*).

Esta amenaza se da cuando el atacante pretende ser alguien y que el sistema lo autentique como tal. Para ello el intruso lo puede llevar a cabo si obtiene los elementos de identificación. Si dichos elementos de seguridad son simplemente el nombre y la contraseña del usuario en su forma de texto en claro (es decir, no cifrada), la suplantación es sumamente fácil. Cuando las contraseñas son notorias no ofrecen una seguridad efectiva en el mundo real.

Muy frecuentemente las contraseñas son divulgadas. Las contraseñas pueden ser adivinadas en muchos casos si se trata de palabras del idioma, simples números o son muy cortas. La suplantación, también conocida como engaño (*spoofing*), en un ambiente de red puede que no requiera de un nombre de usuario y contraseña, pues puede requerir tan sólo de suplantar una dirección de una máquina, tal como una dirección IP en el protocolo TCP/IP utilizado entre máquina en Internet.

El administrador de un depósito de datos no debe ignorar los serios problemas de control de acceso y autenticación. Aún los sistemas estándares basados en contraseñas pueden ser bien seguros si la administración y los usuarios utilizan correctamente sus sistemas de contraseñas. Pero dado que la mayoría de las organizaciones carecen de voluntad para fortalecer la administración y uso de contraseñas robustas, el administrador del depósito de datos debería pensar en reemplazar las contraseñas convencionales por elementos de acceso tales como tarjetas inteligentes (*smart cards*) o sistemas de identificación o autenticación biométrica.

3.2.3.7. Puerta trasera (*Trapdoor*).

En el caso de esta amenaza el atacante prepara o implanta un software que deja disponible una puerta trasera. La puerta trasera es un medio no documentado de ingresar al sistema. El administrador del depósito de datos debe asumir la responsabilidad por la seguridad de la secuencia completa de acceso, desde la estación de trabajo del usuario hasta el disco duro del servidor en que reside el motor de la base de datos relacional. Típicamente nadie, además del administrador del depósito de datos, tiene esta perspectiva. Para cuidarse de las puertas traseras todo el software debería ser certificado como confiable por una autoridad de confianza y reconocida como tal.

3.2.3.8. Soborno, asalto y extorsión (*Bribery, Robbery, and Extortion*).

Estas amenazas tratan acerca de una manipulación directa de una persona o empleado de confianza quien entonces le cede acceso al atacante, y de esta forma le pasa la información o la sustrae. Puesto que este tipo de ataques por lo general se manifiesta a través de acciones lícitas en el uso del depósito, se debe ser capaz de diagnosticar o interceptar modos de acceso inusuales en el depósito. Principalmente lo que se necesita es un sistema de monitoreo que reporte patrones inusuales y que contabilice todos los accesos a ciertos datos.

3.2.4. Clases de modificación de la información.

3.2.4.1. Desvío de la información.

Una forma de modificar la información es copiándola y enviársela a un tercero, acción que se estaría dando en el momento en que ésta es transmitida. El objetivo de esta acción puede ser revelar información sensitiva, o bien, mantener la información que se recibe de primera mano. Los resultados que se obtienen de un depósito de datos frecuentemente se envían por medio del correo electrónico. Los usuarios con este tipo de acceso a los datos necesitan un protocolo de correo electrónico seguro y confiable que evite que se pueda estar copiando y desviando información producida a partir de la interacción con el depósito a personas ajenas a la organización.

3.2.4.2. Mal uso de la información transmitida.

Algunas veces se envía información con intenciones de cometer un fraude o con una dirección de origen falsa. Esta clase de debilidad en la seguridad es muy dañina debido a que es muy difícil localizar a todos aquéllos que hayan recibido el mensaje original fraudulento, y más difícil aún borrar en ellos la impresión original que este tipo de mensaje haya creado. Una forma más seria de este tipo de ataque incluye la destrucción de la reputación de una persona o empresa. Este tipo de problema es fácil que se dé en un ambiente abierto de correo electrónico, independientemente que sea dentro de una intranet o a través de la Web. El administrador del depósito de datos debe trabajar con los funcionarios del área de sistemas de información en procura de implementar la criptografía de llave pública en todo lo que es correo

electrónico de la empresa, de modo que el receptor pueda estar seguro que el emisor sea quien dice ser (o sea, logrando una verdadera autenticación de los emisores de mensajes).

3.2.4.3. Falso rechazo.

Este es un típico problema de seguridad, en donde una persona o empresa afirma que nunca se involucró en una comunicación determinada o generó una transacción, cuando de hecho lo hizo. El falso rechazo es un tema complicado en el ambiente de sistemas de procesamiento de transacciones, y es un tema relativamente mucho menor en un ambiente de depósito de datos. La preocupación principal es asegurarse que quién haya solicitado una modificación en el depósito de datos se le pueda atribuir responsabilidad por el cambio. Si los usuarios pueden actualizar pronósticos, presupuestos o planes entonces el tema del rechazo es importante. Aquellos usuarios que tengan privilegios o responsabilidades de administrador de base de datos necesitan pasar exitosamente la prueba del falso rechazo si es que ellos están tratando con información muy sensitiva

3.2.5. Vulnerabilidades en el robo del software.

3.2.5.1. Robo del código objeto.

En este caso el atacante copia el software ya sea para usarlo él o para vendérselo a otro. El uso que haga el ladrón, si es descubierto, puede crear cierta responsabilidad para con el desarrollador original del software. El ambiente de depósitos de datos frecuentemente presenta una serie de responsabilidades por el uso de diferentes licencias de software. Por lo tanto, se deben controlar las licencias de software de herramientas para consultas, reportes y modelaje, entre otras.

3.2.5.2. Robo del código fuente.

En este tipo de amenaza el atacante utiliza el código fuente para crear productos de software derivados o para divulgar secretos industriales en contra de los deseos del desarrollador. El depósito de datos puede tener una cierta cantidad de productos de software de terceros, desde código de sistemas específicos hasta fórmulas de cálculo en hojas electrónicas. El

código fuente puede revelar las técnicas de análisis empleadas por otros, o bien, errores en el código fuente, que en el caso de que se revele ello, puede conducir a demandas y publicidad dañina.

3.2.5.3. Control "pirateado".

Cuando se da este tipo de amenaza el atacante inserta código no deseado dentro de los componentes de software que antes de ello eran confiable. Si ello sucede, muchas otras formas de ruptura en la seguridad de un depósito de datos pueden ser posibles, dependiendo de cuánto control se tenga en el software en la máquina que se ve amenazada. Un caballo de Troya es una variante del control pirateado, en el cual una pieza de software que en apariencia es inofensiva o normal tiene oculta una función que aparece tiempo después o puede ser activada por el atacante en forma remota. La función oculta puede activarse y divulgar información del depósito de datos yendo en contra de los deseos de la organización, o bien, causar daños a la información almacenada

3.2.5.4. Certificación comprometida.

Cuando se habla de la posibilidad de que una certificación se vea comprometida se está haciendo referencia a la situación en que el atacante logra ganar acceso a un sistema de criptografía de llave privada que está siendo utilizado para certificar el software o para certificar la identidad de un usuario o de un servidor. Hasta tanto ello no sea descubierto, el atacante puede implantar controles de pirateo en todas esas máquinas, que de otra forma tendrían software confiable. La certificación comprometida es una pesadilla no sólo por el daño que causa directamente el intruso, sino también por el costo que involucra restaura la confianza y la emisión de nuevos certificados.

3.2.5.5. Virus.

Se trata de una pieza no deseada de software que entra al sistema y se reproduce a sí misma, llamando la atención por sí misma o causando daños a la información. Los virus se pueden adjuntar a los datos o programas. Las amenazas de virus son bastante bien conocidas por los administradores de los depósitos de datos en comparación con muchas otras amenazas a la seguridad. La mayoría de virus que invaden los ambientes de depósitos de datos causan una interrupción en la prestación de servicios mas que un compromiso directo de la información almacenada en el depósito de datos.

3.2.6. Vulnerabilidades que atentan contra el buen funcionamiento de la empresa.

3.2.6.1. Ataque de negación del servicio (*denial of service attack*).

Los sistemas de información se comprometen específicamente con el propósito de detener o convertir en cuellos de botella a uno o más componentes. Por ejemplo, los servidores pueden ser inundados por miles o millones de solicitudes falsas. Un ataque de negación de servicio bien conocido (el denominado torrente o inundación de SYN –*SYN flood*) involucra el envío de solicitudes de conexión de TCP/IP a un servidor sin completar todos los reconocimientos de máquina necesarios. Otro ataque de negación de servicio similar es el ataque FRAG, en el cual los paquetes TCP/IP son enviados con un número de secuencia confuso o inexistente.

La negación de servicio es muy posible en casi cualquier punto del conducto de envío de paquetes en un ambiente de depósito de datos. Los ataques desde afuera de la Web tienen la intención de afectar al muro de fuego o al servidor Web. Si el muro de fuego es derribado, los legítimos usuarios externos no estarían en capacidad de accesar el depósito de datos. Si el servidor Web es derribado, a todos los usuarios del depósito, tanto los que están dentro como fuera del muro de fuego, se les podría negar el servicio debido al incremento de confianza en las interfaces de usuario habilitadas para utilizar el *browser*. Un SABD podría ser el objetivo directo de un ataque de negación de servicio tomando la forma de muchas consultas enormes o de ciertas consultas que se conoce que hacen que el SABD se caiga. Aquellos usuarios que tienen acceso directo al SABD mediante la manipulación del SQL pueden hacer que éste se caiga fácilmente si este fuera su propósito.

3.2.6.2. Inhabilidad para reconstruir puntos consistentes.

Esta situación se puede dar por intermedio de sabotajes deliberados, pérdidas accidentales o simples descuidos, se puede propiciar un ambiente en el que no sea posible reconstruir un punto consistente de los sistemas. La pérdida de puntos consistentes de los datos almacenados en un depósito de datos tiene muchas consecuencias serias en el campo financiero y legal. La pérdida de puntos consistentes en los metadatos puede significar que no se puede restaura una trayectoria de datos o que se han perdido las reglas del negocio que se han aplicado en las transformaciones de datos.

3.2.6.3. Terrorismo.

Una de las más serias clases de ataque a la negación de servicio es aquélla que intenta que falle la principal infraestructura pública. Por ejemplo, el servicio de llamadas de emergencia 911 podría ser convertido en un cuello de botella o demorado. Se podría imaginar que un servicio público tal como la compañía telefónica o la de distribución de energía eléctrica sea objeto de un ataque de negación de servicio a través de una ruptura en la seguridad de sus sistemas de información. No está lejos que ello se dé. ¿Sería posible que tales servicios se puedan diagnosticar en forma remota?. Normalmente un verdadero depósito de datos no se pone para que esté en la línea crítica de un sistema operacional. Sin embargo, algunos tipos de depósitos de datos están muy cercanos de la línea de los sistemas operacionales. Por ejemplo, si un depósito de datos ofrece información para identificar a la persona que llama, y esta información está ligada a un sistema 911 ó a un sistema de respuesta de emergencias, entonces el sistema ha dejado de ser un depósito de datos y necesita ser considerado un sistema operacional del tipo que no deje de funcionar en ningún momento (o sea, tolerante a fallas). El administrador del depósito de datos necesita juzgar si el sistema está tendiendo hacia esa dirección.

3.3. Análisis del impacto derivado de las amenazas.

Por análisis del impacto derivado de las amenazas se ha de entender la valoración que se hace de las amenazas, tanto desde la óptica de sus consecuencias como desde el grado de preparación de la empresa para hacerle frente a la amenaza.

Sobre este tema la literatura atinente apunta a que existen diferentes factores a considerar a la hora de medir el impacto derivado de las amenazas. Son varios los factores que se pueden utilizan para intentar medir el referido impacto. De éstos existen dos que son los que interesan para la presente investigación; ellos son el riesgo inherente y el riesgo de control.

Con el primero se trata de establecer el nivel de afectación o repercusión que puede experimentar la organización en caso de que la amenaza se llegara a materializar, mientras que el segundo trata de valorar la suficiencia, pertinencia y validez de las medidas de control que la organización haya definido en aras de minimizar la amenaza. Más adelante se retomará este tema para desarrollarlo con mayor amplitud.

En este punto resulta necesario hacer un alto en la exposición del tema para recapitular lo tratado en las sesiones 2.1 (Seguridad), 2.3 (Seguridad en los depósitos de datos) y 3.2

(Identificación de las amenazas) del presente documento, en relación con las posibles amenazas que atentan contra los depósitos de datos, en procura de establecer una categorización más concisa de ellas. En virtud de ello se tiene lo siguiente:

a) según [SHEL1997] las amenazas pueden provenir de:
   - ataques activos que implican cambios en los datos (almacenados o transmitidos), y se puede manifestar en términos de borrado, corrupción, retraso o atasco de transmisiones
   - ataques pasivos tratan de la recolección de información sin que nadie sepa que se está produciendo, del tipo escucha (*eavesdropping*) o pinchazos electrónicos (*sniffing*, *spoofing*, etc.)

b) según [DENN1982] las amenazas se centran en dos objetivos:
   - el secreto o privacidad, que incluye búsqueda, exposición (o revelación) e inferencia
   - la autenticidad, que incluye la alteración y la destrucción accidental

c) según [DATE1987] las amenazas que atentan contra los datos almacenados en las bases de datos intentan realizar consultas y actualizaciones ilícitas. Asimismo, se considera lo referente a los accesos no autorizados y la alteración o destrucción de la información

d) según [ATRE1988] las amenazas contra una base de datos van en la línea de las revelaciones, alteraciones o destrucciones no autorizadas o accidentales. Más puntualmente se apunta que la privacidad involucra información errónea o engañosa, revelación o modificación accidental de la información, infiltración accidental a un sistema y pérdida de datos

e) según [MOSS1998] la pureza (o limpieza) de los datos es una amenaza particular de los depósitos de datos

f) según [HUFF1996] la calidad de los datos es otra amenaza que atenta contra los depósitos de datos

g) según [GILL1996] son varios los factores a considerar dentro de las amenazas que atentan contra un depósito de datos, a saber:
   - el depósito de datos es un conjunto abierto de datos de la empresa
   - los usuarios accesan los datos a diferentes niveles de resumen
   - las herramientas OLAP y el tema del acceso a los depósitos de datos muestra estar todavía en una etapa exploratoria (no se ha llegado a un estado de madurez)
   - la preocupación no es tanto causar daños en los datos, sino más bien revelar secretos o estrategias

h) según [WARI1997] las amenazas o vulnerabilidades que atentan contra los depósitos de datos provienen de factores tales:
   - la seguridad está incluida en el SABD
   - limitaciones del SABD

- seguridad sustentada en una estrategia de bloque dual (características de seguridad proveídas por el SABD y el sistema operativo)
- ataques por inferencia
- disponibilidad (al reñir con la confidencialidad y la integridad de los datos)
- humanos (actos accidentales o intencionales, tales como errores, omisiones, modificaciones, destrucción, mal uso, divulgación, sabotaje, fraude y negligencia)
- internas (empleados disgustados o deshonestos)
- externas (competidores y terceros que pueden utilizar espionaje electrónico y técnicas similares para robar, comprar o reunir datos estratégicos contenidos en el depósito de datos)
- naturales (fuego, agua y aire, así como terremotos)
- utilidad (interrupción del fluido eléctrico y del servicio de comunicaciones)

i) según [KIMB1998] entre las amenazas o vulnerabilidades de los depósitos de datos se pueden citar:

- de los activos físicos (robo, destrucción intencional, incendio, humedecimiento, inmersión, suciedad, envejecimiento, descargas o interferencias eléctricas, disturbios magnéticos, pérdida por obsolescencia tecnológica y secuestro o pirateo)
- de la información (divulgación de planes, divulgación de códigos, divulgación de información confiada a la empresa, eliminación de la protección, robo de activos financieros, robo de servicios, robo de información para promover violencia o terrorismo, robo de identidad y robo de privacidad)
- que atentan contra el funcionamiento de la empresa (negación del servicio –*denial of service attack*-, inhabilidad de reconstruir puntos consistentes y terrorismo)

Dado lo anterior, las amenazas se pueden agrupar de la siguiente forma:

a) Destrucción de la información

Dentro de esta categoría se pueden considerar acciones tales como: destrucción del dispositivo de almacenamiento por medio de la acción del fuego, agua, suciedad, acciones negligentes, sobreescritura intencional, borrado de la información, y acciones de sabotaje.

b) Alteración de la información

En esta categoría se consideran aspectos tales como: inserción, duplicación, sobreescritura, modificación, corrupción y actualización de los registros almacenados en el depósito, así como interferencia eléctrica, disturbios magnéticos y acciones de sabotaje.

c) Divulgación de la información

Esta categoría cubre situaciones tales como: infiltración y suplantación de usuarios, revelación, exposición, inferencia y mal uso de la información, así como acciones de sabotaje.

d) Sustracción de la información

En este grupo se han de considerar aspectos tales como: robo de códigos de acceso, o de activos físicos que contengan códigos de acceso o datos, infiltración y suplantación de usuarios, y el secuestro o pirateo de activos.

e) Pérdida de confidencialidad

En esta categoría se considera la posibilidad de que usuarios autorizados tengan acceso a información confidencial, como puede ser el caso de un ataque por inferencia, lo que a la postre conlleve a una pérdida de ésta.

f) Falta de disponibilidad del depósito de dato

Esta clase considera la posibilidad de que se pierda la continuidad en la prestación del servicio por ataques del tipo negación del servicio(*denial of service attack*).

g) Pérdida de integridad

Esta categoría contempla la amenaza de perder la integridad del depósito, que como cualquier base de datos que es en última instancia, la integridad es un requisito básico que se debe preservar.

h) Calidad de la información inaceptable

Esta categoría considera la amenaza de que la información contenida en el depósito de datos no sea de óptima calidad.

Una vez hecha la categorización de las amenazas, de seguido se presenta una exposición de algunas de las consideraciones que es importante tener en cuenta antes de proceder propiamente al desarrollo de los elementos necesarios para llevar a cabo el análisis del impacto de tales amenazas.

Es necesario considerar que las amenazas existen, y que los controles se definen con el objeto de minimizar la probabilidad de ocurrencia de la materialización de éstas, o bien, de minimizar la magnitud de las consecuencias que se pueden derivar en caso de una materialización de ellas. Algo que es importante tener claro es que la amenaza nunca

desaparece, por más (y mejores) controles que se definan la amenaza nunca desaparece, lo más que se logra con los controles es minimizarla llevándola a niveles aceptables o tolerables, pero la amenaza no va a desaparecer.

En este sentido, existen amenazas para las cuales es posible definir controles que se orienten solamente a atacar lo referente a la probabilidad de ocurrencia, existen otras para las cuales los posibles controles sólo estarán encaminados a minimizar la magnitud de las consecuencias, y finalmente existen unas pocas amenazas para las cuales los posibles controles permitirán hacerle frente a ambas situaciones (probabilidad de ocurrencia y magnitud de las consecuencias).

Puesto que los controles tienen un costo asociado, otro aspecto a considerar a la hora de intentar entender el por qué de la definición éstos, se refiere a la necesidad de hacer un análisis del costo que tiene el control y compararlo con el costo del bien o recurso sujeto al control. Dicha comparación tiene el propósito de determinar si el establecimiento del control está justificado, ello por cuanto debe existir una relación razonable entre el costo del control y el del bien controlado. Si el costo del control no está justificado, ello puede provocar que la empresa incurra en costos innecesarios. En otras palabras, en tanto el costo del control sea inferior al costo del bien controlado se justifica la implementación del control, si por el contrario el costo del control excede significativamente al del bien controlado, lo razonable es no optar por éste, y aceptar la situación tal cual, ya que ello implica la menor pérdida posible, pues en el peor de los casos (o sea, en caso de que la amenaza se llegare a materializar) el bien de por sí se iba a perder.

De lo anterior se tiene entonces, que si el costo del bien controlado es superior al de los controles requeridos, la organización debe seleccionar e implantar tales controles, no sólo porque así va a darle el mejor tratamiento a la amenaza, sino también porque con ello va a demostrar un estilo de administración diligente y responsable ante las exposiciones de riesgo que la amenazan.

De allí entonces que es necesario tener en cuenta que la definición e implantación de controles para hacerle frente a las amenazas requiere de una dotación de recursos económicos, lo cual muchas veces es restringido por las posibilidades de la organización y por la voluntad que tengan las instancias decisorias de ésta para facilitar tales recursos. Por lo tanto, es normal que no todas las amenazas sean cubiertas suficientemente, dadas las limitaciones antes mencionadas.

En vista de ello, es común que las organizaciones destinen sus recursos para atender las amenazas que presenten una mayor repercusión en caso de que se lleguen a materializar, y que aquéllas de menor trascendencia sean atendidas parcialmente en función de los recursos que les pueda destinar luego de atender las más preocupantes, y por ende, las más prioritarias.

3.3.1. Valoración del riesgo inherente.

El riesgo inherente representa la valoración que se le asigna al impacto que puede tener la materialización de una amenaza, medido éste en función de la trascendencia que ello puede tener para la continuidad de la organización que la llegara a sufrir. Para su análisis se definen tres categorías:

- Alta: una calificación de riesgo inherente alto se asigna a aquella amenaza cuya materialización compromete sustancialmente la continuidad de la prestación de los servicios que ofrece la organización. Por ejemplo, para una organización con cierto grado de automatización, una amenaza a la que le corresponde un riesgo inherente alto es la pérdida de su información, ya sea que ésta se produzca en forma accidental o premeditada, o que afecte parcial o totalmente la información, el sólo hecho de perder información puede provocar que la organización vea comprometida su capacidad de prestar los servicios en una forma continua.
- Media: una calificación de riesgo inherente medio se le otorga a aquella amenaza de la que se derivan consecuencias que no son tan críticas como las que se dan en el caso del riesgo inherente alto, pero que merecen una cierta consideración, por cuanto sus eventuales consecuencias no son del todo despreciables. Por ejemplo, en el ámbito del desarrollo de sistemas se puede considerar que el no disponer de un expediente actualizado para cada uno de los sistemas constituya un riesgo inherente medio, pues la sana práctica señala que para todo sistema se debe contar con un expediente en el que aparezcan documentados todos los aspectos relevantes de éste, desde que se inició su desarrollo (lo que implica el estudio preliminar y demás detalles correspondientes a las fases que componen el proceso de desarrollo), hasta las diferentes modificaciones que se le han practicado una vez que éste se puso en funcionamiento (o sea, las solicitudes de mantenimiento con su correspondiente documentación que respalde tales labores). Si bien es cierto disponer de tal expediente es importante para la empresa, en sí mismo esta situación no necesariamente representa una situación crítica que eventualmente pueda comprometer el continuo funcionamiento de la empresa. Sin embargo, la situación en comentario tampoco es del todo despreciable como para que la empresa la descuide del todo, pues ello repercute de alguna manera en la capacidad de darle mantenimiento a éste.
- Baja: una calificación de riesgo inherente bajo se le da a aquella amenaza cuyas consecuencias son prácticamente intrascendentes para la organización. Ejemplo de ello puede ser la rotación que puede enfrentar una empresa dentro de su personal informático. Es normal que las personas aspiren a superarse, razón por la cual es muy común que frecuentemente se presenten movimiento de personal dentro de las áreas informáticas. Puesto que nadie es indispensable, se espera

que el efecto de la renuncia, ascenso, despido o muerte de un cierto funcionario, independientemente del puesto que ocupe, la rotación no se debiera catalogar como algo muy preocupa, pues como ya se mencionó, el que la gente deje su puesto es algo normal y de esperar, es por ello entonces que asignarle una calificación de riesgo inherente bajo pareciera ser la más apropiada.

De seguido se procederá a valorar el riesgo inherente de las amenazas, según la agrupación presentada previamente.

3.3.1.1 Destrucción de la información.

Puesto que la información es el principal insumo de un depósito de datos, la destrucción de la información representa una amenaza a la que necesariamente se le debe asignar una valoración de riesgo inherente alto. Sin embargo, tomando en consideración lo reciente de la información y la utilidad que se desprenda de la información que se destruya, existe la posibilidad de que se le pueda dar la valoración de riesgo inherente medio, como sería el caso de la información sumamente antigua, la cual sea utilizada con muy poca frecuencia, o que haya perdido actualidad, y por ende, valor.

3.3.1.2 Alteración de la información.

Si se toma en consideración que un depósito de datos no se considera un sistema transaccional en el que las actualizaciones de la información y sus consecuencias derivadas (problemática de actualización y solución al problema de la exclusión mutua, así como la pérdida de integridad) son sumamente importantes, sino más bien un sistema que apoya el proceso de toma de decisiones, esta amenaza no representa una amenaza sumamente significativa desde el punto de vista meramente de la acción de alteración de la información. Además, los usuarios normalmente accesan la información del depósito con fines de consulta, pues por definición, un depósito de datos tiene ese propósito. En razón de ello, esta amenaza a lo más que llegaría es a obtener una valoración de riesgo inherente medio.

3.3.1.3 Divulgación de la información.

El depósito de datos es por excelencia un sitio en el que se encuentran los datos muy bien organizados, integrados, ubicados generalmente en una localización central, y optimizados para su acceso; todo ello con miras a

favorecer los procesos de toma de decisiones. Por lo tanto, es de esperar que una falla de seguridad que permita que se divulgue en forma inadecuada la información contenida en el depósito represente una amenaza seria para este tipo de sistemas, no sólo por las implicaciones que tiene en el sentido de que se podrían estar divulgando los elementos de tipo estratégico de la organización, sino que esta deficiencia atenta contra la privacidad y confidencialidad de la información contenida en el depósito. En razón de lo anterior, la valoración de riesgo inherente para esta amenaza necesariamente tiene que ser alta.

3.3.1.4 Sustracción de la información.

Dado que la sustracción implica premeditación, y de alguna manera obedece a un motivo, el hecho de que se sustraiga información contenida en un depósito de datos ello representa una acción grave. Sin embargo, el riesgo inherente de esta amenaza se debe establecer en función de la naturaleza del objeto sustraído. Procurando una definición más amplia de información en torno al concepto de depósito de datos, se puede considerar que los posibles objetos a sustraerse pueden ser propiamente los datos contenidos en el depósito, o bien, las contraseñas para accesar al depósito. En el caso de que se considere que el objeto capturado sea información extraída del depósito su valoración de riesgo inherente debiera ser alta, por tratarse de información ya procesada y dispuesta para el proceso de toma de decisiones. En lo tocante a la posibilidad de que las contraseñas de acceso sean obtenidas indebidamente, la repercusión de este hecho no necesariamente debiera ser tan grave, toda vez que todavía quedaría como obstáculo tener acceso a los medios para extraer la información, y también dependería del nivel de acceso que se tenga con las contraseñas sustraídas, por lo que tal vez una valoración de riesgo inherente medio sea adecuada. En todo caso, si esta valoración no fuera suficiente, se puede reconsiderar y ubicarla en el nivel de riesgo inherente alto, en el tanto se tengan indicios de que ese hecho realmente representa una amenaza sumamente seria para la empresa.

3.3.1.5 Pérdida de confidencialidad.

Tomando en consideración que la pérdida de confidencialidad en este punto se refiere a la eventual pérdida de privacidad de la información por parte de usuarios legítimos del depósito de datos, la valoración del riesgo inherente debe considerar las eventuales consecuencias que se pueden derivar de una acción que a todas luces puede resultar imperceptible. Por ejemplo, éste sería el caso de un usuario que teniendo un acceso válido al depósito, recolecte a lo largo de cierto tiempo parte de la información a la que ha tenido acceso, que bien podría tratarse como una modalidad muy solapada de

inferencia. La valoración del riesgo inherente de esta amenaza podría ser de riesgo medio, aunque dependiendo de la cantidad y sensitividad de la información a la que tuvo acceso se podría considerar la posibilidad de asignársele una valoración alta.

3.3.1.6 Falta de disponibilidad del depósito de dato.

La falta de disponibilidad del depósito de datos, al igual que sucede con la falta de disponibilidad de un servidor de bases de datos, no puede recibir una valoración de riesgo inherente baja. Sin embargo, si se toma en consideración que los depósitos de datos no son un sistema transaccional, del cual se depende para el trabajo diario de la empresa, tal vez la valoración de riesgo inherente se puede establecer en un nivel medio. En todo caso, la valoración que al final se le asigne a esta amenaza debe considerar también el eventual tiempo durante el cual no esté disponible el depósito, pues a pesar de lo dicho anteriormente, si un depósito de datos permanece fuera de servicio por un tiempo considerable, más allá del permisible por la empresa antes de que se empiecen a manifestar secuelas de la interrupción, definitivamente la valoración del riesgo inherente debe ser alta.

3.3.1.7 Pérdida de integridad.

Los depósitos de datos guardan una gran similitud con las bases de datos, pues en esencia se sustentan en un sistema administrador de bases de datos para todo lo referente a sus labores de administración del cúmulo de información que contienen. Perder la integridad implica que los datos contenidos no son coherentes, o bien, que no guardan una relación de armonía entre los diferentes elementos que lo componen. En razón de ello, si bien los procesos de carga (o poblamiento, del inglés *population*) del depósito de datos contienen una serie de rutinas que intentan asegurar la coherencia de los datos con que se está cargando el depósito, al punto de considerar no sólo los datos resumidos (en sus diferentes niveles de agregación), sino también los datos detallados, una falta de integridad, o una pérdida de ella, producto de un fallo en alguno de los componentes del depósito, puede comprometer las decisiones que se tomen a partir de la información contenida en el depósito. De allí entonces que la valoración del riesgo inherente para esta amenaza debe ser alta.

3.3.1.8 Calidad de la información inaceptable.

En el ámbito de los sistemas de información se tiene como máxima que la pureza de los datos que ingresan a éstos definen de forma muy significativa la integridad y utilidad de la información que se obtiene de ellos, a tal punto que generalmente se dice que "*si entra basura a un sistema de información, sale basura*". Si a ello se le agregan las consideraciones expuestas por Larissa Moss en [MOSS1998], en el sentido de que existe una serie de situaciones que impiden asegurar la pureza o limpieza de los datos que se incorporan a un depósito de datos, se tiene entonces que la calidad de los datos contenidos en un depósito es de por sí uno de sus factores críticos de éxito. Por lo tanto, la valoración del riesgo inherente que se le debe asignar a esta amenaza necesariamente debe ser alta.

3.3.2. Valoración del riesgo de control.

El riesgo de control se refiere a la valoración que se hace en relación con el grado de cobertura que ofrecen las medidas (generalmente referidas como controles) emprendidas por la organización, con el afán de minimizar una amenaza. Por lo tanto, es común escuchar que la valoración del riesgo de control mide la suficiencia, pertinencia y validez de los controles definidos en torno a una amenaza.

Al igual que con el riesgo inherente, el riesgo de control se puede valorar utilizando tres categorías:

- Alta: una calificación de riesgo de control alto se le asigna a aquella amenaza para la cual los controles que se han definido para hacerle frente no son suficientes, válidos o pertinentes, o bien, la situación imperante es tal que no se definieron controles para atender la amenaza. Generalmente dentro de un proceso de auditoría esta calificación está asociada a actuaciones negligentes o descuidadas por parte de la administración de la empresa.
- Media: una calificación de riesgo de control medio se otorga a aquella amenaza para la cual los controles que la administración ha definido para hacerle frente no son del todo suficientes, válidos o pertinentes. Ello significa entonces que sí existen medidas de control, pero éstas no son del todo satisfactorias o las mejores dentro de las posibilidades de la empresa.
- Baja: una calificación de riesgo de control bajo se le asigna a aquella amenaza para la cual los controles definidos por la administración resultan los más apropiados, de conformidad con las posibilidades de la organización y el tipo de amenaza de que se trata.

Tomando como base lo anterior, a continuación se hará una exposición de los elementos más importantes a tomar en cuenta para valorar cada una de las categorías de riesgo de control para las amenazas previamente analizadas en la sección anterior.

3.3.2.1 Destrucción de la información.

Entre las medidas de control que se pueden adoptar para hacerle frente a esta amenaza se tiene como la principal disponer de copias de respaldo. En la medida en que las copias que se hayan preparado permitan recuperar la información destruida, en el menor tiempo, y de la forma más completa posible, mejor será la calificación de riesgo de control que se le debe asignar.

Por lo tanto, si la empresa no dispone del todo de copias de respaldo, o si para la recuperación, además de los datos respaldos se requiere de trabajo adicional para volver al punto normal de procesamiento, la valoración de riesgo de control necesariamente debe ser alta, en señal de que la amenaza no está siendo atendida como se debe.

Si se dispone de copias de respaldo, pero el tiempo de recuperación no es inmediato, sin llegar al extremo de que se requiera un lapso mayor al aceptable por la empresa, la valoración de riesgo de control se puede ubicar en un rango medio, lo que evidenciaría que la amenaza no está adecuadamente cubierta, pero que tampoco está del todo desatendida.

En el caso de que las copias de respaldos permitan que prácticamente la recuperación sea inmediata, al punto de que resulte imperceptible para los usuarios la situación, la calificación del riesgo de control debe ser baja, como muestra de que la situación está razonablemente controlada.

3.3.2.2 Alteración de la información.

Por definición, un depósito de datos es sólo de consulta, lo cual implica que no es común que los usuarios estén actualizando la información, como sí sucede con los sistemas transaccionales u operacionales. En este sentido, según la literatura consultada, en un ambiente de depósitos de datos, los aspectos relativos a las anomalías de actualización, lo mismo que aspectos tales como detección y resolución de entrabamientos (*deadlocks*) no deben constituir una fuente de amenaza para los depósitos de datos.

A pesar de ello, existe la posibilidad de que la información que se extrae de un depósito de datos pueda ser alterada, ya sea durante la transferencia de los datos desde el repositorio hasta el usuario final, o durante el tiempo en que se mantienen almacenados los datos en el repositorio, y no porque corresponda a una acción de actualización típica de los sistemas operacionales, sino más bien a un intento deliberado de cambiar o sustituir la información. Dado que estas dos situaciones obedecen a naturalezas diferentes, de seguido se analizan cada una de ellas por separado.

En el caso de la interceptación con el propósito de modificar los datos que viajan por algún medio de transmisión, las posibilidades de éxito de este tipo de acciones dependen de la propensión a que ello sea permitido por el sistema de comunicaciones que se utilice. Generalmente los protocolos de comunicación utilizan esquemas de control tales que pueden detectar la alteración de los mensajes, al punto de que con códigos de paridad o de redundancia cíclica (del tipo CRC), es factible detectar la presencia de una modificación (accidental o intencional) del contenido del mensaje. Por lo tanto, en la medida en que los protocolos de comunicación exhiban características tales que detecten la ocurrencia de una variación en el mensaje, así será la valoración del riesgo de control que reciba este factor.

En el caso de que el protocolo detecte en un porcentaje bastante alto, tendiendo al 100% de las posibles ocurrencias, su valoración de riesgo de control sería baja.

Si la situación fuera que a pesar de utilizarse técnicas de detección de alteraciones en los mensajes la efectividad de éstas no fuera suficientemente alta, la valoración del riesgo de control se puede ubicar en un nivel medio.

Finalmente, si no utilizan técnicas para la detección de modificaciones en los mensajes la valoración del riesgo de control debe ser alta.

En lo que respecta a la posibilidad de modificar o alterar los datos contenidos en el depósito, el análisis se puede orientar de la siguiente forma. Puesto que en principio ningún usuario debiera tener privilegios de escritura o modificación sobre los datos del depósito, el hecho de preservar que ello sea así constituye de por sí en un control a implementar.

Por otro lado, dado que los únicos que deben poder hacer cambios en los datos contenidos en el depósito son los funcionarios a quienes se les haya asignado la labor de carga y mantenimiento del depósito, labores que generalmente se realizan en forma automática a través de programas o rutinas preparadas para esos efectos, y sólo en situaciones extraordinarias se justifica que dichos funcionarios interactúen directamente con el repositorio de datos, se podría pensar en el establecimiento de un control que restrinja los accesos directos al repositorio a tan sólo unas pocas personas, y que en el caso en que efectivamente requieran hacerlo tal acción se genere una entrada en una bitácora. En dicha bitácora se deberá consignar la identidad del usuario que llevó a cabo la acción, algunos datos que identifiquen el momento y desde dónde se hizo la acción, y una explicación o detalle de la naturaleza de la acción realizada. Puesto que este tipo de acciones sería poco frecuente, no debieran de existir criterios contrarios a la utilización de una medida de control de este tipo, inspirados éstos en argumentos como el espacio en disco que puede requerir la bitácora, o bien, el *overhead* que pueda implicar la generación de la entrada.

En razón de lo anterior, una valoración de riesgo de control baja se le puede asignar a una empresa en donde se tomen previsiones como las expuestas anteriormente, o muy similares.

En el caso de que las medidas que se definan no sean suficientes para identificar y establecer la naturaleza de los cambios aplicados al repositorio de datos necesariamente la valoración del riesgo de control se puede ubicar en un rango medio.

De no existir medidas que identifiquen al autor y la naturaleza de los cambios hechos al contenido de un depósito de datos, la valoración del riesgo de control debe ser alta.

Dado que se identificaron dos factores a tomar en cuenta a la hora de realizar la valoración de esta amenaza, la idea es que se asigne un único valor que los represente a ambos. Para ello se puede considerar que en caso de que las valoraciones individuales sean diferentes se asigne aquella que represente el mayor grado de riesgo.

3.3.2.3 Divulgación de la información.

En materia de divulgación de la información, las eventuales medidas de control que se pueden definir se concentran en torno a la concesión de

derechos de acceso. En este sentido, es común utilizar términos tales como códigos de usuario y perfiles o roles.

Tal y como se expusieran en la sección referente al mecanismo de control de acceso y los procesos que lo componen, es mediante este mecanismo que se intenta brindar un acceso controlado a la información. De allí entonces que las medidas de control tendientes a minimizar la divulgación de la información contenida en el depósito deban basarse e un esquema de este tipo.

Según esa teoría, un acceso controlado a la información debe pasar por tres etapas, a saber: identificación, autenticación y autorización.

Es mediante la primera de ellas que se asegura que sólo los usuarios previamente registrados pueden tener acceso a la información, lo que implica que un eventual usuario sin una identificación válida no tendrá acceso a la información.

En relación con el segundo proceso, puesto que la mayoría de las veces el elemento identificador no es secreto, sino más bien es común que se trate de un elemento público, el mecanismo necesita tener algún grado de certeza de que el usuario que solicita acceso sea quien dice ser. Para ello existen diferentes técnicas que intentan dar esa certeza de la autenticidad del solicitante de la información. Aunque no son infalibles tales técnicas, existe la posibilidad de llevarlas a niveles aceptables.

El tercer proceso es quizás el más delicado, pues es el que trata acerca de la concesión a los usuarios de derechos de acceso a la información. Para ello se trabaja bajo el esquema de lista de control de acceso, que no son otra cosa que una lista por usuario de los eventuales objetos a los que tiene acceso, asociado con el respectivo permiso o derecho para accionar con tales objetos de información. En su forma más simple, esta lista de control de acceso sería algo así como una colección de todos los pares (*objeto*, *acción*) a los que expresamente se le ha dado acceso al usuario, en donde *objeto* representa a los posibles objetos o elementos que componen el depósito de datos, y *acción* representa a las posibles operaciones o comandos que el usuario puede ejecutar sobre el objeto. Es claro que la lista de control de acceso se debe definir a partir de una base válida para el usuario, como lo puede ser el puesto o función que éste ejerce dentro de la organización. Es en este sentido que se indica que el proceso es tal vez el más delicado, por cuanto un exceso en la concesión de derechos puede permitir que un usuario tenga acceso a información que no le compete.

Puesto que la concesión de derechos se convierte en un trabajo muy delicado, y a la vez muy laborioso, típicamente las organizaciones realizan esta etapa a partir de la figura de los perfiles o roles, los cuales consisten en una definición impersonal de los derechos y objetos a los que deben tener acceso un grupo o categoría de funcionarios dentro de la organización. Dicha definición generalmente se basa en los principios de mínimo privilegio y el de la necesidad de conocer. El primero se refiere a que se le debe conceder a un usuario el mínimo de los privilegios (acciones que puede ejecutar sobre un objeto) requeridos para ejecutar sus funciones. Por ejemplo, si una persona requiere hacer sólo consultas en una tabla, solamente se le debe permitir ejecutar el comando SELECT, y no todas los estatutos del SQL. En caso de que un usuario que solamente debe tener acceso de consulta logre cambiar datos, no hay argumentos para culparlo de las consecuencias de su acción, dado que fue el sistema el que se lo permitió, y que el funcionario no hizo ninguna acción adicional para llevar a cabo tal cometido. En lo tocante al segundo principio, éste se refiere a que a un usuario solamente se le debe permitir acceso a los objetos que necesariamente requiere para cumplir con sus funciones laborales. Algunas veces en las organizaciones por simplificar el trabajo no toman en cuenta esta consideración, lo cual es la fuente más común de fugas de información, pero no es sino hasta que ello ocurre que se percatan de la debilidad de seguridad que tienen.

Luego de que se han definido los perfiles o roles, se procede a vincular (o afiliar) a los funcionarios con los roles o perfiles, de acuerdo con la naturaleza de su puesto, valiéndose para ello de un concepto similar al de herencia en el paradigma de orientación a objetos, en donde se "heredan" los derechos del grupo (o grupos) en el (o los) que está inscrito. De esta forma, si un funcionario tuviera alguna especialización dentro del grupo basta con agregar lo pertinente en la lista de control de acceso correspondiente a este funcionario para lograr tal efecto.

Puesto que este proceso de concesión no debiera ser ejecutado por cualquier usuario, sino más bien sólo por aquéllos con privilegios especiales, otro elemento a tomar en cuenta trata acerca de los controles que se puedan definir en torno a la función de administración de la seguridad. Ello implica no sólo el registro de las acciones que este funcionario en materia de concesión, remoción y variación de derechos de acceso, sino también en lo referente a la formalidad con que se realiza la función. Lo primero se refiere a la necesidad de utilizar una bitácora en la que se registren las diferentes operaciones que se realicen en relación con los derechos de acceso, mientras que lo segundo trata acerca de la necesidad de definir formalmente los aspectos operativos de la función, tales como procedimientos para solicitar

cambios en la definición de los derechos, procedimientos para el acceso y uso de la bitácora, entre otros.

Dado el panorama anterior, la valoración del riesgo de control de esta amenaza se debe realizar a la luz de los dos principales elementos antes expuestos, o sea, concesión de derechos y administración de la función de seguridad.

En lo referente a la concesión de derechos, en el caso de que la organización ejecute el proceso de autorización de una forma rudimentaria, en donde la definición no se hace por medio de perfiles, sino más bien por definición directa en la lista de control de acceso, ello representa una situación riesgosa, por cuanto incrementa la posibilidad de que a un usuario se le puedan dar accesos más allá de lo que realmente requiere, incumpliendo con los principios de mínimo privilegio y necesidad de conocer ya mencionados anteriormente. Por lo tanto, esta situación debiera de valorarse como de riesgo de control alto.

De tratarse una situación en la que se utilicen los perfiles, pero su administración y actualización sean poco efectivas, al punto de que más bien la concesión de derechos se haga más por la vía de las excepciones y no tanto por la generalidad que representa el perfil, la valoración del riesgo de control debiera de ser media.

Por el contrario, si en la organización se tiene claramente establecido el uso de perfiles, éstos se mantienen constantemente en revisión, procurando que los perfiles definidos reflejen a la mayor perfección posible las necesidades de los grupos a que representan, y se tiene comprobado que la concesión de derechos de los diferentes usuario se hace partiendo de los perfiles definidos, esta situación merece una valoración de riesgo de control baja, como indicio de que se han tomado las mejores medidas de control posibles.

En lo que respecta a la administración de la seguridad, en el tanto la organización cuente con una adecuada definición de los elementos mínimos de control requeridos para la ejecución de la función de la seguridad del depósito de datos, y se tenga comprobado que existe un compromiso porque tales elementos estén en constante revisión y actualización, así como que sean acatados por todo el personal involucrado, se puede considerar que la valoración de riesgo de control para este factor debe ser baja.

Si por el contrario, se notan algunas deficiencias en cuanto a la existencia de algunos de los elementos de control interno requeridos, o en lo que respecta a la revisión, actualización o cumplimiento de tales factores de control interno, la valoración de riesgo de control debe ubicarse en un nivel medio.

Finalmente, si el panorama prevaleciente en la organización es tal que no hay evidencia de que existan los elementos mínimos de control interno para la ejecución de la función de administración de la seguridad, o que se definieron en algún momento, pero en la práctica las labores en torno a esta función se realizan de una forma totalmente diferente a lo que se establece en la formulación, lo más apropiado es asignarle una calificación de riesgo de control alto.

Dado que en esta oportunidad también se consideraron dos factores para valorar esta amenaza, y puesto que la calificación debe ser un único valor, en caso de que ambas calificaciones sean distintas se debe tomar la que represente el mayor nivel de riesgo como la calificación que mejor representa la situación de la amenaza.

3.3.2.4 Sustracción de la información.

De conformidad con lo manifestado en la numeral precedente en que se analizó el riesgo inherente para esta amenaza, la sustracción implica intencionalidad a la hora de perpetrarla. También se indicó en esa oportunidad que en el ámbito de un depósito de datos existen claramente dos objetivos de sustracción, los datos extraídos del depósito y las claves de acceso.

En lo referente a los datos extraídos del depósito, para ayudar en la valoración de su riesgo de control se debe analizar primero en qué forma o momento es que se puede llevar a cabo la sustracción de los datos. En este sentido se puede considerar diferentes situaciones: primero, que los datos sean sustraídos directamente del depósito de datos. Otra situación sería que los datos sean sustraídos en el momento en que los datos están siendo recuperados por el usuario como parte de una sesión de trabajo, por medio de la interceptación de la comunicación que se tiene entre el servidor del depósito y la estación de trabajo del usuario. Otro escenario correspondería al caso en que los datos sean sustraídos una vez que éstos ya se encuentren almacenados en la estación de trabajo del usuario, o bien en poder de éste. Un último panorama sería el que corresponde a la sustracción de la información contenida en un informe impreso.

De ser el caso en que la información sea sustraída directamente del depósito de datos, esta situación corresponde a una sustracción de la información, aspecto que fue analizado como parte de la amenaza anterior, razón por la cual no será considerada nuevamente en este numeral.

De tratarse de una sustracción al momento de la transmisión de los datos entre el servidor y la estación de trabajo, las posibles medidas de control que se pueden utilizar para minimizar esta amenaza van en la dirección de buscar un medio de transmisión confiable. Dado que por definición los medios de transmisión son inseguros, se requiere de técnicas que provean confidencialidad en la transmisión, aspecto que se logra mediante la encriptación de la información que trasiega por el canal. En algunos casos, ello se obtiene utilizando redes privadas virtuales (VPN, *Virtual Private Network*) o túneles, sobre todo para lo que es conexiones entre redes que se valen de Internet como red de enlace entre ellas, o bien, en el caso de las comunicaciones internas dentro de las organizaciones, en donde se pueden emplear algoritmos criptográficos para cifrar la información que se transmite por su red interna.

Para el caso de que la sustracción se lleve a cabo una vez que los datos están almacenados en la estación de trabajo del usuario, las posibles medidas de control que se pueden implementar van en la dirección de restringir el acceso a la estación misma, ya sea por el uso de protectores de pantalla con contraseña; lo cual de paso no es muy efectivo, por cuanto si la información reside en el disco duro local siempre existe la posibilidad de apagar el equipo y accesar directamente la información en ese disco; o bien, por encriptación de la información valiosa contenida en el disco duro local, precisamente para minimizar lo comentado anteriormente en este mismo párrafo.

También se puede considerar el caso en que la información sea extraída del depósito de datos y almacenada en el servidor de archivos de la red. Si bien es cierto que ello resuelve un poco la posibilidad de que la información se extraiga fácilmente del disco duro local, esta medida aumenta la posibilidad de que la información pueda ser sustraída al viajar por el medio de transmisión.

Finalmente se tiene el caso de que la información sea sustraída una vez que fue consignada en un documento impreso. En esta situación las medidas de control aplicables estarían dirigidas a regular el uso y destrucción de los reportes impresos. En cuanto al uso, ello se refiere a la forma cómo son impresos y distribuidos los reportes, en procura que estos procesos se

realicen de una forma confiable y segura. En lo referente a la a la destrucción de los reportes impresos, las medidas de control estarían orientadas a definir y poner en práctica métodos confiables de destrucción de la información sensitiva, que por lo general es el caso de aquélla obtenida de un depósito de datos.

En general, la valoración de los anteriores cuatro formas de sustracción se deberá realizar individualmente de la siguiente forma. En el caso de que las medidas de control adoptadas para atender la forma de sustracción de que se trate resulten insuficientes según lo antes comentado, o si se aplican otras medidas de control que no tienden a alcanzar los objetivos que se destacaron en la anterior exposición, la valoración de riesgo de control deberá ser alta.

Si la situación fuera que las medidas de control están implementadas de manera deficiente, o existe una ligera discrepancia, pero aún así se trata de atender la amenaza de que se trata, se le puede asignar una valoración de riesgo de control media.

La valoración de riesgo de control baja se asignará en aquellos casos en que se tenga certeza de que las medidas adoptadas efectivamente ofrecen una confianza razonable de que la amenaza se ve sustancialmente disminuida con tales acciones.

Por tratarse de un factor que puede tener cuatro formas diferentes de materialización, y dado que existe la posibilidad de se puedan dar más de una a la vez, el análisis de este factor se debe hacer en forma individual para cada alternativa, para luego consolidarlo en una sola valoración, en donde debe imperar la que muestre el mayor nivel de riesgo.

Considerando ahora que el objeto sustraído sea las contraseñas y demás información de acceso al depósito de datos, si bien es cierto que se trata de información hasta cierto punto sensitiva para la organización, ello no es del todo cierto, pues como ya se indicara anteriormente, las claves por sí mismas no divulgan la información contenida en el depósito, toda vez que para ello se requiere la facilidad y la oportunidad de poder interactuar con el depósito. Es precisamente en esta línea que se pueden dirigir las medidas de control para hacerle frente a esta amenaza.

Por ejemplo, se puede considerar en primera instancia que para el acceso de cierta información de alto valor para la empresa (sobre todo información altamente sensitiva, la que típicamente se clasifica como secreta o

estratégica) se pueda considerar utilizar un esquema de autenticación más robusto, de manera que al intentar accesar ese tipo de información se requiera una segunda fase de autenticación por parte del usuario. También se puede pensar en utilizar un esquema mediante el cual se pueda restringir el acceso de algunos usuarios desde unas pocas estaciones de trabajo, o bien, que se restrinja el horario de acceso, todo como miras a que en caso de que sean sustraídas tales claves, el acceso efectivo deba hacerse desde ciertas estaciones de trabajo y en horarios laborales, propios por lo general del usuario dueño de la clave. Por decir algo, si la contraseña que se sustrajo fue la del gerente, quien tiene acceso irrestricto a la información estratégica contenida en el depósito, que exista el medio de restringir que ese usuario sólo pueda establecer conexiones desde unas pocas estaciones de trabajo (posiblemente desde su oficina y desde su casa de habitación, para lo cual basta con conocer las direcciones IP de tales equipos), y en horarios en que habitualmente esa persona está en tales direcciones.

Si bien es cierto que las medidas de control para atender esta amenaza pueden caer en extremos un tanto fantasiosos, la verdad es que no está de más que la organización sopese el eventual valor de la información que puede perder contra el costo y esfuerzo que ello pueda requerir.

La valoración del riesgo de control para este factor va estar entonces determinado por la suficiencia, pertinencia y validez de las medidas de control que la organización quiera establecer en este sentido. De nuevo la valoración de riesgo de control alta se le deberá asignar a la situación en que la amenaza no sea atendida del todo, la valoración de riesgo de control media a la situación en que se hayan tomado medidas, pero éstas no sean suficientes, y la valoración de riesgo de control bajo corresponderá a la situación en que la medidas de control resulten razonables y suficientes según las circunstancias.

Puesto que esta amenaza consideró dos factores, al final se debe dar una valoración única, por lo que se debe seleccionar aquélla que represente el mayor riesgo como la valoración de la amenaza.

3.3.2.5 Pérdida de confidencialidad.

Tal y como se definiera anteriormente, esta amenaza se refiere al caso en que un usuario debidamente autenticado haga un mal uso de la información a la que tuvo acceso. Entiéndase que al hablar de un usuario debidamente autenticado se está haciendo mención a que se trata del caso en que un usuario que pasó exitosamente las fases de identificación y autenticación,

propias de un mecanismo de control de acceso, hace uso de la información a la que tuvo acceso de una forma indebida. Sobre el particular también es importante señalar que este tipo de situaciones no es común que se lleve a cabo por parte de los usuarios, pero que a pesar de ello es posible que se den. Se considera que esta amenaza produce una pérdida de confidencialidad por cuanto el atacante es un miembro de la misma organización, a quien generalmente se le tiene un cierto grado de confianza, la que le permite tener acceso a información confidencial de ésta.

Ante este panorama es de comprender que los controles previos resulten sumamente difíciles de utilizar, por cuanto no se tiene certeza del tipo de ataque que se pueda perpetrar, por cuanto las alternativas pueden ser muchas y la probabilidad de que se lleguen a materializar pueden ser muy remotas. De allí entonces que muchas veces resulte prácticamente imposible prever todas las posibilidades y definir una forma de prevenir la materialización, por cuanto ello puede requerir tiempo y recursos considerables, los cuales se pueden destinar a atender situaciones con una probabilidad de ocurrencia mayor.

En todo caso, existen algunos elementos que ayudan a darle un tratamiento preventivo a esta amenaza. Por un lado están los principios éticos y morales que tengan los usuarios, pues para algunos usuarios su honorabilidad y principios no les permiten involucrarse en tales situaciones.

Por otro lado están los esfuerzos que emprenda la organización en procura de crear una conciencia entre sus empleados en materia de principios éticos y morales, así como a la definición y difusión de una política institucional en lo que respecta a la seguridad informática, que dé pie a que los eventuales usuarios tengan un conocimiento claro de los recursos disponibles, y de las implicaciones que se pueden derivar de su uso.

Como se puede ver, estos esfuerzos no son exclusivos de los depósitos de datos, y más bien caben dentro del tema de seguridad computacional en general, razón por la cual resulta conveniente analizar ahora la situación que compete a los depósitos y la pérdida de confidencialidad.

Para hacerle frente a esta amenaza, según lo explicado anteriormente, los controles que se pueden implementar estarían orientados hacia el uso de bitácoras en las que se registre información relacionada con los accesos que realizan los diferentes usuarios del depósito. Sobre este respecto es importante señalar que las bitácoras gozan de una impopularidad en la mayoría de organizaciones, por razones tales como el eventual "*overhead*"

que conlleva su generación, el espacio en disco, y en general, en almacenamiento secundario que se requiere para soportarlas, y la cuestionable utilidad que se obtiene de su uso, sobre todo si le ve desde la óptica del costo-beneficio. Tal impopularidad llega al extremo de que en muchas organizaciones el uso de las bitácoras se realiza más por acatamiento de disposiciones reglamentarias o estatutarias que por voluntad propia.

A pesar de ello, si se desea hacer frente a esta amenaza, y ante la dificultad (que raya en la imposibilidad) de definir controles previos que no afecten de manera significativa la utilidad del depósito de datos, no queda otro recurso que acudir a las bitácoras. Tal vez lo que se puede hacer para mitigar, no sólo los inconvenientes mencionados en el párrafo precedente, sino también otros argumentos negativos que surgen en torno al uso de las bitácoras, sería conveniente que la organización hiciera un esfuerzo por definir e implementar un esquema de bitácoras que le resulte eficiente.

Por lo tanto, para valorar esta amenaza se puede utilizar como parámetro el grado de preparación de la empresa en el tema de bitácoras para el depósito de datos, de manera tal que si no se tienen definidas éstas, la valoración de riesgo de control debe ser alta.

En el caso en que la organización tenga definido el uso de bitácoras en que se registre el acceso al depósito de datos, pero éstas no responden efectivamente a la utilidad que se espera recibir de tal herramienta, o bien, éstas no permiten obtener un resultado satisfactorio en el tiempo requerido, la valoración del riesgo de control se puede ubicar en la categoría media.

Finalmente, si la organización ha hecho los esfuerzos necesarios para llegar a una definición e implementación adecuada del uso de bitácoras, lo que incluye entre otras un procedimiento para la revisión periódica de la información contenida en las bitácoras que permita de manera proactiva detectar patrones sospechosos de utilización del depósito, así como que considere aspectos tales como la búsqueda, consulta, impresión y retención de la información contenida en la bitácora, la valoración del riesgo de control se puede ubicar en la categoría baja.

3.3.2.6 Falta de disponibilidad del depósito de datos.

El tema de la falta de disponibilidad del depósito de datos, tal y como se enunciara previamente, no se considera una amenaza grave, en el tanto el tiempo de indisponibilidad del depósito no sea excesivo. Sin embargo, luego

del atentado del 11 de setiembre del 2001 en los Estados Unidos, algunos autores han retomado el tema, y se han permitido hacer sus advertencias en torno a ello.

Tal es el caso de Ralph Kimball, que en [KIMB2001b][45], se refiere al tema en relación con los depósitos de datos. En este artículo Kimball menciona entre tales fallas la destrucción de la facilidad (en este caso del depósito de datos) producto de un ataque terrorista, el sabotaje deliberado por parte de personal interno, la "guerra cibernética" y las fallas de punto único ("*single point failures*") deliberadas o no.

Sobre el particular, Kimball sugiere considerar como medidas las arquitecturas distribuidas, las rutas paralelas de comunicación, las redes de área de almacenamiento extendido (SAN, del inglés *Storage Area Network*), los respaldos diarios en medios de almacenamiento removibles que se puedan almacenar en sitios seguros, el filtrado de paquetes a través de compuertas ("*gateways*") ubicadas estratégicamente y el acceso y autenticación basado en roles controlado centralizadamente.

Se debe recordar que los controles que se definan deben tener una relación adecuada del costo que implica éste en contraposición con el costo del bien en sí. Es por ello que, para algunas organizaciones, las medidas que sugiere Kimball en su artículo no les resulten del todo aplicable, por cuanto algunas de ellas les pueden resultar onerosas. Sin embargo, es conveniente que la organización haga un análisis de sus vulnerabilidades en este respecto, considere los eventuales efectos que se pueden derivar y decida cuáles de las sugerencias son viables.

La valoración de esta amenaza va a estar determinada por la identificación que haga la empresa de las amenazas que le pueden afectar y el grado de preparación que se haga para minimizar tales amenazas.

En este sentido, si la organización no ha hecho ni una identificación de sus amenazas más probables, ni la correspondiente valoración de ellas, la calificación de riesgo de control que se le debe asignar será alta, ello por cuanto cualquier esfuerzo por definir medidas de control en este sentido serán nulas o mal orientadas.

[45] [KIMB2001b] págs. 20, 22 y 48.

En el caso en que la organización haya hecho la identificación de las posibles amenazas que le pueden afectar, así como el respectivo análisis del impacto, pero las medidas de control que puede haber definido no resultan las más adecuadas dentro de sus posibilidades, la valoración que se le debe asignar es la de categoría media.

Finalmente, en el caso en que la organización no sólo haya hecho la identificación de las amenazas más probables, sino que también las haya valorado, y haya definido medidas de control que dentro de sus posibilidades sean las más adecuadas, así como que se tenga establecido un mecanismo de monitoreo de tales medidas de control, la valoración del riesgo de control a asignársele debe ser baja.

3.3.2.7 Pérdida de integridad.

La pérdida de integridad se refiere a que aún cuando la carga de los datos hacia el depósito se realiza por medio de procedimientos que han sido revisados y aprobados previamente, siempre existe la posibilidad de que se presenten errores que pueden provocar que la información contenida en el depósito resulte incoherente. A pesar de que esta amenaza puede tener una probabilidad de ocurrencia muy pequeña, el no contemplarla resulta una omisión de cuidado.

Ante esta situación, no está de más que la organización realice algunas pruebas luego de haberse ejecutado el proceso de carga de los datos al depósito, de manera que antes de que ellos se pongan a disposición de los usuarios se tenga un mayor grado de certeza de que no se vayan a presentar problemas derivados de dicho proceso y que puedan comprometer la integridad del depósito.

Es de esta forma que la valoración del riesgo de control se puede realizar tomando como aspecto de valoración el tratamiento que haga la organización al proceso de carga de datos y la posterior habilitación del depósito de datos. Con ello se pretende valorar el nivel de previsión que tiene la organización ante eventuales problemas derivados de fallas fortuitas ocurridas al momento de cargar el depósito.

Si la organización no tiene prevista ninguna prueba a la consistencia de los datos contenidos en el depósito luego de ejecutado el proceso de carga, la valoración del riesgo de control debe ubicarse en alta.

Si la organización tiene definidas pruebas que no son suficientes para detectar eventuales problemas derivados del proceso de carga, problemas que pueden comprometer en algún grado la integridad de los datos del depósito, la valoración de riesgo de control se debe ubicar en la categoría media.

Si por el contrario la organización dispone no solo de un conjunto de pruebas a ser aplicadas luego de concluido el proceso de carga del depósito, sino que ello forma parte de un procedimiento claramente establecido, en el que se ha considerado el reporte al administrador del depósito para su inmediata atención, la valoración del riesgo de control se debe ubicar en la categoría baja.

3.3.2.8 Calidad de la información inaceptable.

En primera instancia, no se debe obviar los planteado por Larissa Moss en [MOSS1998] en el sentido de que existe una serie de situaciones que dificultan la pureza o calidad de los datos recopilados y posteriormente incorporados en un depósito de datos. Es en este sentido que resulta útil de considerar el artículo de Duane Hufford [HUFF1996], en el cual trata el tema de la calidad de los datos de un depósito, a tal punto que se permite plantear un método compuesto de cinco actividades, por medio del cual, aplicando el concepto de administración de la calidad total a los datos, se puede llegar a tener una calidad de los datos de acuerdo con la definición que haga la organización al respecto. Sólo como recordatorio vale la pena anotar aquí dichas actividades:

- Definir las expectativas de calidad de los datos y sus métricas
- Identificar los riesgos que atentan contra la calidad de los datos
- Valorar los riesgos que atentan contra la calidad de los datos
- Mitigar los riesgos
- Monitorear y evaluar los resultados

En ese mismo artículo Hufford también hizo mención a una serie de características que influyen en la determinación de la calidad de los datos, a saber:

- Exactitud
- Completitud

- Consistencia
- Relatividad
- Oportunidad
- Unicidad
- Validez

Si desea un mayor detalle de tales características o actividades, se sugiere revisar la sección 2.2.7.2. Calidad de los datos.

Dado lo anterior, la valoración del riesgo de control de esta amenaza se puede realizar tomando como parámetro el grado de preparación que tenga la organización en cuanto a las actividades sugeridas por Hufford, u otras orientadas en esa misma dirección. De esta forma, si la organización no dispone de un método que le permita determinar la calidad de los datos que se agregan al depósito, la valoración del riesgo de control debe ser alta.

En el caso en que la organización disponga de un método similar al sugerido por Hufford, el cual considere la mayoría de las características definidas por este, la valoración del riesgo de control se puede ubicar en la categoría media.

Si la organización dispone de una metodología orientada en la misma dirección que sugiere Hufford, se consideran todas las características planteadas por éste, y además, la metodología se aplica en todos sus extremos, al punto de que la organización busca un mejoramiento continuo de la calidad de los datos del depósito, la valoración del riesgo de control se puede ubicar en la categoría baja sin reservas.

Los dos factores antes analizados están estrechamente relacionados, al punto de que para hacer una valoración más adecuada del riesgo se deben tomar ambos en forma conjunta. Considerar tan sólo el riesgo inherente puede conducir a resultados incorrectos, pues una amenaza catalogada con un riesgo inherente alto no necesariamente es una fuente de preocupación para la empresa, dado que será la valoración del riesgo de control correspondiente a esta amenaza la que determine si se trata de una actividad riesgosa (riesgo de control alto) o una razonablemente controlada (riesgo de control bajo).

De igual forma, considerar tan sólo la valoración del riesgo de control en forma aislada puede conducir a resultados erróneos o imprecisos. Por ejemplo, el hecho de que una amenaza presente una valoración de riesgo de control alto (o sea, que la amenaza no está

siendo debidamente controlada o atendida) no necesariamente debe ser un indicativo de que se trata de una situación preocupante, pues para ello se requiere considerar la valoración de riesgo inherente asociada a la amenaza en cuestión. No es lo mismo que una amenaza con riesgo inherente alto (de suma importancia para la empresa) presente un riesgo de control alto, que una amenaza con riesgo inherente bajo (situación poco significativa para la empresa) tenga asociado un riesgo de control alto, pues por lo comentado anteriormente, es común que las amenazas con un riesgo inherente bajo sean atendidas sólo después de que se ha hecho lo propio con las amenazas más prioritarias (las de riesgo inherente alto o medio).

Adicionalmente a los dos criterios antes analizados, el análisis del riesgo puede utilizar factores tales como la probabilidad de ocurrencia de la amenaza, o bien, el monto esperado de pérdida, los cuales a veces aparecen dentro del análisis de riesgo. Sin embargo, dada la poca información que se pudo localizar en torno a estos para el caso de los depósitos de datos, y dada la imposibilidad material de realizar evaluaciones en este respecto, para efectos de la presente investigación resultó más apropiado desarrollar los criterios de riesgos inherente y riesgo de control, los cuales de alguna forma resultan más intuitivos para los propósitos del presente trabajo.

Capítulo 4. Propuesta para el mejoramiento de la seguridad de un depósito de datos y su respectiva validación.

4. Propuesta para el mejoramiento de la seguridad de un depósito de datos y su respectiva validación.

4.1. Propuesta para el mejoramiento de la seguridad de un depósito de datos.

4.1.1. Proceso para el mejoramiento de la seguridad de un depósito de datos.

Para tener un poco más claro el concepto de que se trata, se consultó el Diccionario de la Real Academia Española en procura de conocer qué significa el verbo "mejorar". Una de las acepciones de este verbo hace mención a "adelantar, acrecentar una cosa, haciéndola pasar a un mejor estado"[46]. De lo anterior se puede concluir que la acción de mejorar implica necesariamente un cambio de estado hacia uno mejor.

Resulta lógico pensar que para cambiar es necesario saber en qué estado se encuentran las cosas en la actualidad, de allí entonces que el primer paso del proceso que aquí se propone trata acerca de ello; con esto se sienta la base para poder entonces establecer el estado que se desea alcanzar con del mejoramiento deseado.

En el proceso propuesto, el establecimiento de la situación actual se sitúa primero en lo que son los datos, para luego abocarse al tema de las amenazas que atentan contra éstos. De allí entonces el proceso prosigue con la mecánica necesaria para identificar y valorar tales amenazas, planificar las acciones tendentes a administrar el riesgo asociado a las amenazas de una manera adecuada, y dar seguimiento y controlar las acciones que se adopten para tales fines (refiérase a la Figura 18).

Dado que el entorno es cambiante, ello puede provocar que las acciones adoptadas no sean suficientes para controlar adecuadamente las amenazas identificadas, o bien, que surjan nuevas amenazas derivadas de los cambios en el entorno. Es por ello entonces que el proceso para el tratamiento de las amenazas se vuelve cíclico.

[46] [RAE1992] página 954.

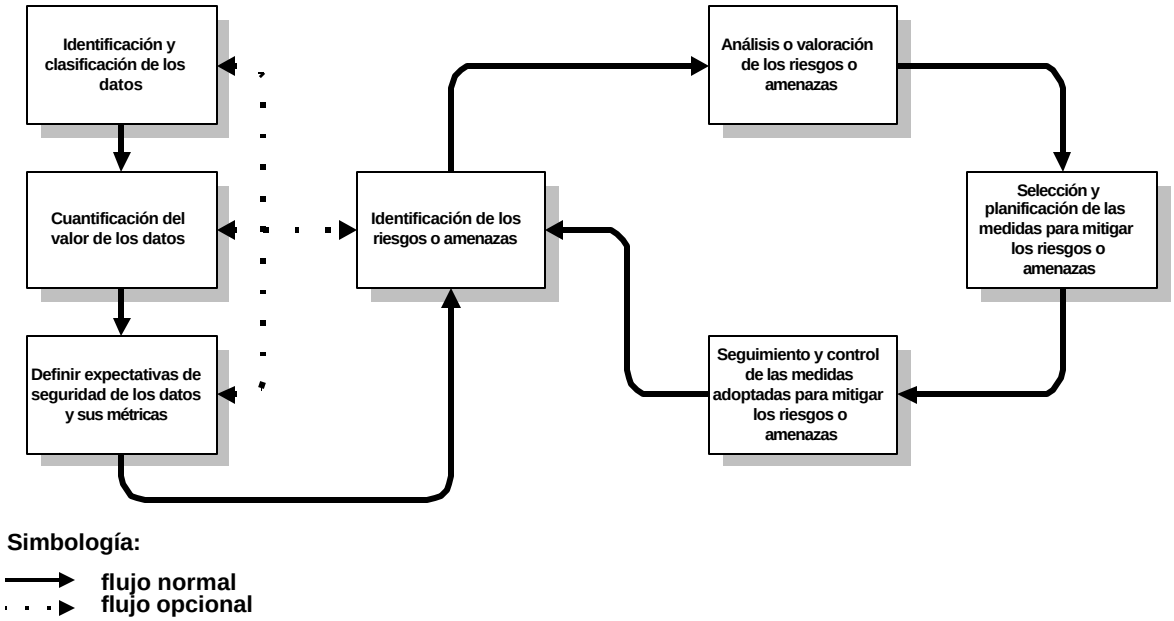


Figura 18. Proceso para el mejoramiento de la seguridad de los depósitos de datos.

En los siguientes numerales se detalla cada de una de las etapas que componen el proceso.

4.1.1.1. Identificación y clasificación de los datos.

Como un requisito básico del proceso, es necesario identificar todos los datos almacenados digitalmente dentro de la organización, y que serán puesto a disposición de los usuarios por medio del depósito. Tal identificación generalmente está asociada con la preparación de un inventario de los datos, en que se provee información precisa acerca de todas las bases de datos, tablas, columnas, filas de datos y perfiles de datos, información que servirá de base para la definición de los requerimientos de seguridad del ambiente de depósito de datos.

Existen un par de advertencias que es necesario tener presente. Primero, la definición de los requerimientos de seguridad se debe realizar desde el inicio del proceso de diseño del depósito, pues omitir la aludida identificación vuelve crítica esta etapa y tiene repercusiones en las etapas subsecuentes del proceso de diseño. Además, a efecto de que la información compilada en este punto sea de utilidad para las siguientes etapas, es necesario que ésta sea organizada, documentada y conservada adecuadamente.

Una vez que el inventario se ha preparado, se debe llevar a cabo una clasificación de los datos en función de su grado de sensitividad, confidencialidad o seguridad, según se le quiera ver a esta clasificación. Sobre este tema se debe apuntar que Slemo Warigon [WARI1997] sugiere que tal clasificación se puede hacer con base en tres categorías: datos públicos, confidenciales o muy confidenciales.

No obstante, se podría utilizar cualquiera otra categorización que haya definido la organización en este sentido, en el tanto las categorías reflejen el tipo de criticidad o sensitividad de los datos ante la exposición indebida, o la modificación o destrucción de ellos. Generalmente esas categorías se derivan de las políticas que sobre el particular disponga la organización, y en lo que necesariamente se involucra a dueños, custodios y usuarios finales de esos datos.

A final de cuentas el propósito que se persigue con la clasificación de los datos es establecer rangos para las categorías por incremento de los grados de criticidad, de modo que diferentes medidas de protección se puedan utilizar para diferentes categorías.

4.1.1.2. Cuantificación del valor de los datos.

Con anterioridad se expuso acerca de las implicaciones que se tienen al tratar de controlar los riesgos o amenazas. En esa oportunidad se indicó que las medidas de control tienen un costo asociado, y que éstas sólo son viables, o aceptadas por la administración, cuando el costo del control no supera el del bien controlado. Para ello es necesario realizar un análisis costo/beneficio de las posibles medidas de control para determinar si alguno de los controles se puede llevar a la práctica, o bien, si se asume el riesgo en cuestión.

Los datos, como cualquier otro bien que poseen las organizaciones, se les debe asignar un valor. No obstante, esta es una labor bastante difícil, pues los datos, en sí mismos, no tienen un valor intrínseco.

Por lo tanto, el propósito que persigue esta etapa es intentar de asignarle un valor real (o de mercado) a los datos ubicados anteriormente en las diferentes categorías de sensitividad.

Entre los elementos, que se utilizan frecuentemente para establecer el valor de los datos, se pueden mencionar los costos: a) de reconstruir los datos perdidos, b) de restaurar la integridad de datos corruptos o interceptados, c) de no tomar a tiempo una decisión a causa de la falta de servicio de la facilidad que pone a disposición los datos, o d) de pagar obligaciones financieras por la exposición pública de datos confidenciales. En la determinación del valor de los datos también se pueden incluir ingresos dejados de percibir por causa de la fuga de secretos empresariales hacia sus competidores, y el uso anticipado de datos financieros secretos por parte de empleados deshonestos (antes de que éstos se hagan públicos en el mercado o en el sector de que se trate).

4.1.1.3. Definición de expectativas de seguridad de los datos y sus métricas.

Una vez que se ha cuantificado el valor de los datos, otro paso importante a seguir trata acerca de la definición de las expectativas de seguridad de éstos, que no es otra cosa más que describir la seguridad (confidencialidad o privacidad) que requieren tener los datos para soportar las principales aplicaciones que se van a ofrecer a través del depósito de datos. Esta definición de expectativas implica necesariamente la definición de las métricas por medio de las cuales se va a poder establecer si se alcanzaron las expectativas, métricas que debieran ser más cuantitativas que cualitativas, para evitar ambigüedades y lograr un mayor grado de precisión.

La aludida definición de expectativas se debe hacer a partir de los requerimientos que tiene la organización, derivadas de sus propias reglas del negocio, y constituye un anhelo que puede ser modificado por la misma empresa, en función de los cambios que experimente ésta en respuesta a la variación de su entorno.

4.1.1.4. Identificación de las amenazas.

La identificación de las amenazas trata de pronosticar cómo o de qué forma se puede fallar al alcanzar las expectativas definidas en la etapa anterior.

Para la identificación de las amenazas se debe considerar la mayor diversidad posible de causas probables de problemas o afectaciones del funcionamiento del depósito, sin importar el grado en que eventualmente no se alcancen las expectativas de seguridad. Para tal efecto, en el capítulo precedente de enlistaron y analizaron algunas de las causas más comunes que atentan contra el funcionamiento de los depósitos de datos, y que

podrían considerarse como un punto de partida para la identificación de amenazas, no está de más considerar otras causas, o bien, prescindir de algunas de ellas, dependiendo de las particularidades del depósito de que se trate.

Las amenazas que se identificaron y analizaron en el capítulo anterior son las siguientes:

- Vulnerabilidades de los activos físicos
  - ❖ Robo destrucción intencional
  - ❖ Incendio
  - ❖ Humedecimiento
  - ❖ Agua
  - ❖ Suciedad
  - ❖ Envejecimiento
  - ❖ Descargas o interferencias eléctricas
  - ❖ Disturbios magnéticos
  - ❖ Pérdida por obsolescencia tecnológica
  - ❖ Secuestro o pirateo (*hijack*) de activos
- Vulnerabilidades de la información
  - ❖ Divulgación de planes confidenciales
  - ❖ Divulgación de códigos
  - ❖ Divulgación de información dada en confianza a la empresa
  - ❖ Divulgación de información sensitiva
  - ❖ Eliminación de protección o de oportunidad
  - ❖ Robo de activos financieros
  - ❖ Robo de servicios
  - ❖ Robo de información para promover violencia o terrorismo
  - ❖ Robo de identidad
  - ❖ Robo de privacidad
- Modalidades de robo
  - ❖ Arrebato oportunista (*Opportunistic Snatching*)
  - ❖ Difusión inadvertida (*Inadvertent Broadcasting*)
  - ❖ Escucha disimulada (*Eavesdropping*)

- ❖ Robo físico como medio para robar información
- ❖ "Pirateo" de sesión (*Hijacked Session*)
- ❖ Imitación o suplantación (*Impersonation*)
- ❖ Puerta trasera (*Trapdoor*)
- ❖ Soborno, asalto y extorsión (*Bribery, Robbery, and Extortion*)

- Clases de modificación de la información
  - ❖ Desvío de la información
  - ❖ Mal uso de la información transmitida
  - ❖ Falso rechazo
- Vulnerabilidades en el robo del software
  - ❖ Robo del código objeto
  - ❖ Robo del código fuente
  - ❖ Control "pirateado"
  - ❖ Certificación comprometida
  - ❖ Virus
- Vulnerabilidades que atentan contra el buen funcionamiento de la empresa
  - ❖ Ataque de negación del servicio (*denial of service attack*)
  - ❖ Inhabilidad para reconstruir puntos consistentes
  - ❖ Terrorismo

Un mayor detalle de estas posibles amenazas se puede encontrar en el Capítulo 3, Sección 3.2 y subsiguientes.

4.1.1.5. Análisis de las amenazas.

Al realizar el análisis de las amenazas se trata de evaluar las posibles consecuencias que se pueden derivar de ellas, los elementos que podrían estar propiciándolas y los mecanismos que se podrían implementar para su apropiada detección y mitigación.

Por análisis del impacto derivado de las amenazas se ha de entender la valoración que se hace de las amenazas, tanto desde la óptica de sus consecuencias como desde el grado de preparación de la empresa para hacerle frente a la amenaza.

Tal y como se discutiera anteriormente, existen diferentes factores a considerar a la hora de medir el impacto derivado de las amenazas. Son varios los factores que se pueden utilizan para intentar medir el referido impacto. En esta investigación se le puso énfasis a dos de ellos: el riesgo inherente y el riesgo de control.

Con el primero se trata de establecer el nivel de afectación o repercusión que puede experimentar la organización en caso de que la amenaza se llegara a materializar, mientras que el segundo trata de valorar la suficiencia, pertinencia y validez de las medidas de control que la organización haya definido en aras de minimizar la amenaza.

Es necesario considerar que las amenazas existen, y que los controles se definen con el objeto de minimizar la probabilidad de ocurrencia de la materialización de éstas, o bien, de minimizar la magnitud de las consecuencias que se pueden derivar en caso de una materialización de ellas. Algo que es importante tener claro es que la amenaza nunca desaparece, por más (y mejores) controles que se definan la amenaza nunca desaparece, lo más que se logra con los controles es minimizarla llevándola a niveles aceptables o tolerables, pero la amenaza no va a desaparecer.

En este sentido, existen amenazas para las cuales es posible definir controles que se orienten solamente a atacar lo referente a la probabilidad de ocurrencia, existen otras para las cuales los posibles controles sólo estarán encaminados a minimizar la magnitud de las consecuencias, y finalmente existen unas pocas amenazas para las cuales los posibles controles permitirán hacerle frente a ambas situaciones (probabilidad de ocurrencia y magnitud de las consecuencias).

Puesto que los controles tienen un costo asociado, otro aspecto a considerar a la hora de intentar entender el por qué de la definición éstos, se refiere a la necesidad de hacer un análisis del costo que tiene el control y compararlo con el costo del bien o recurso sujeto al control. Dicha comparación tiene el propósito de determinar si el establecimiento del control está justificado, ello por cuanto debe existir una relación razonable entre el costo del control y el del bien controlado. Si el costo del control no está justificado, ello puede provocar que la empresa incurra en costos innecesarios. En otras palabras, en tanto el costo del control sea inferior al costo del bien controlado se justifica la implementación del control, si por el contrario el costo del control excede significativamente al del bien controlado, lo razonable es no optar por éste, y aceptar la situación tal cual, ya que ello implica la menor pérdida

posible, pues en el peor de los casos (o sea, en caso de que la amenaza se llegare a materializar) el bien de por sí se iba a perder.

De lo anterior se tiene entonces, que si el costo del bien controlado es superior al de los controles requeridos, la organización debe seleccionar e implantar tales controles, no sólo porque así va a darle el mejor tratamiento a la amenaza, sino también porque con ello va a demostrar un estilo de administración diligente y responsable ante las exposiciones de riesgo que la amenazan.

De allí entonces que es necesario tener en cuenta que la definición e implantación de controles para hacerle frente a las amenazas requiere de una dotación de recursos económicos, lo cual muchas veces es restringido por las posibilidades de la organización y por la voluntad que tengan las instancias decisorias de ésta para facilitar tales recursos. Por lo tanto, es normal que no todas las amenazas sean cubiertas suficientemente, dadas las limitaciones antes mencionadas.

En vista de ello, es común que las organizaciones destinen sus recursos para atender las amenazas que presenten una mayor repercusión en caso de que se lleguen a materializar, y que aquéllas de menor trascendencia sean atendidas parcialmente en función de los recursos que les pueda destinar luego de atender las más preocupantes, y por ende, las más prioritarias.

El riesgo inherente representa la valoración que se le asigna al impacto que puede tener la materialización de una amenaza, medido éste en función de la trascendencia que ello puede tener para la continuidad de la organización que la llegara a sufrir. Para su análisis se definen tres categorías:

- Alta: una calificación de riesgo inherente alto se asigna a aquella amenaza cuya materialización compromete sustancialmente la continuidad de la prestación de los servicios que ofrece la organización. Por ejemplo, para una organización con cierto grado de automatización, una amenaza a la que le corresponde un riesgo inherente alto es la pérdida de su información, ya sea que ésta se produzca en forma accidental o premeditada, o que afecte parcial o totalmente la información, el sólo hecho de perder información puede provocar que la organización vea comprometida su capacidad de prestar los servicios en una forma continua.
- Media: una calificación de riesgo inherente medio se le otorga a aquella amenaza de la que se derivan consecuencias que no son tan críticas como las que se dan en el caso del riesgo inherente alto, pero que merecen una

cierta consideración, por cuanto sus eventuales consecuencias no son del todo despreciables. Por ejemplo, en el ámbito del desarrollo de sistemas se puede considerar que el no disponer de un expediente actualizado para cada uno de los sistemas constituya un riesgo inherente medio, pues la sana práctica señala que para todo sistema se debe contar con un expediente en el que aparezcan documentados todos los aspectos relevantes de éste, desde que se inició su desarrollo (lo que implica el estudio preliminar y demás detalles correspondientes a las fases que componen el proceso de desarrollo), hasta las diferentes modificaciones que se le han practicado una vez que éste se puso en funcionamiento (o sea, las solicitudes de mantenimiento con su correspondiente documentación que respalde tales labores). Si bien es cierto disponer de tal expediente es importante para la empresa, en sí mismo esta situación no necesariamente representa una situación crítica que eventualmente pueda comprometer el continuo funcionamiento de la empresa. Sin embargo, la situación en comentario tampoco es del todo despreciable como para que la empresa la descuide del todo, pues ello repercute de alguna manera en la capacidad de darle mantenimiento a éste.

- Baja: una calificación de riesgo inherente bajo se le da a aquella amenaza cuyas consecuencias son prácticamente intrascendentes para la organización. Ejemplo de ello puede ser la rotación que puede enfrentar una empresa dentro de su personal informático. Es normal que las personas aspiren a superarse, razón por la cual es muy común que frecuentemente se presenten movimiento de personal dentro de las áreas informáticas. Puesto que nadie es indispensable, se espera que el efecto de la renuncia, ascenso, despido o muerte de un cierto funcionario, independientemente del puesto que ocupe, la rotación no se debiera catalogar como algo muy preocupa, pues como ya se mencionó, el que la gente deje su puesto es algo normal y de esperar, es por ello entonces que asignarle una calificación de riesgo inherente bajo pareciera ser la más apropiada.

El riesgo de control se refiere a la valoración que se hace en relación con el grado de cobertura que ofrecen las medidas (generalmente referidas como controles) emprendidas por la organización, con el afán de minimizar una amenaza. Por lo tanto, es común escuchar que la valoración del riesgo de control mide la suficiencia, pertinencia y validez de los controles definidos en torno a una amenaza.

Al igual que con el riesgo inherente, el riesgo de control se puede valorar utilizando tres categorías:

- Alta: una calificación de riesgo de control alto se le asigna a aquella amenaza para la cual los controles que se han definido para hacerle

frente no son suficientes, válidos o pertinentes, o bien, la situación imperante es tal que no se definieron controles para atender la amenaza. Generalmente dentro de un proceso de auditoría esta calificación está asociada a actuaciones negligentes o descuidadas por parte de la administración de la empresa.

- Media: una calificación de riesgo de control medio se otorga a aquella amenaza para la cual los controles que la administración ha definido para hacerle frente no son del todo suficientes, válidos o pertinentes. Ello significa entonces que sí existen medidas de control, pero éstas no son del todo satisfactorias o las mejores dentro de las posibilidades de la empresa.
- Baja: una calificación de riesgo de control bajo se le asigna a aquella amenaza para la cual los controles definidos por la administración resultan los más apropiados, de conformidad con las posibilidades de la organización y el tipo de amenaza de que se trata.

Para mayor detalle, se puede consultar la Sección 3.3 y subsiguientes.

4.1.1.6. Selección y planificación de las medidas para mitigar las amenazas.

Una vez que las posibles amenazas han sido identificadas y analizadas en lo que respecta a las eventuales consecuencias que se podrían derivar de su materialización, y tomando en cuenta que anteriormente se consideraron las posibles acciones a emprender, en esta etapa lo que procede es actuar en procura de mitigar tales amenazas.

Para ello, es necesario seleccionar de entre las acciones alternativas para mitigar la amenaza, la que resulte más conveniente para la organización, conveniencia que debe ser analizada a la luz de las expectativas de seguridad definidas previamente y considerando para ello las posibilidades económicas de la organización, lo cual generalmente se satisface a través de un análisis costo/beneficio de las acciones.

No es de sorprender que puedan existir situaciones para las cuales la organización opte por asumir el riesgo de la materialización de una cierta amenaza, dado que la aludida relación costo/beneficio no resultó favorable, y por tanto, no permitió que se llevaran a la práctica las posibles acciones tendientes a su mitigación por presentar un costo de las acciones (o controles) superior al costo del bien o recurso controlado. Sin embargo, es de esperar que este sea el caso por excepción y no tan el común de todas las acciones.

4.1.1.7. Seguimiento y control de las medidas adoptadas.

Como regla general, una cosa es lo que se planea y otra muy distinta lo que en la práctica resulta. En el caso que nos ocupa, una cosa es la acción para mitigar la amenaza, vista como una acción aislada (o teórica), y otra la acción interactuando con las demás acciones implementadas (llevada a la práctica), además de que también puede producirse cambios motivados por el entorno en que se desarrolla la actividad del depósito de datos.

Todo ello hace necesario que se les deba dar un seguimiento y control a las acciones que se adoptaron en la etapa anterior, con lo cual se puedan valorar los resultados obtenidos de la implementación de tales acciones y determinar si los esfuerzos de mitigación de las amenazas surtieron sus frutos, o por el contrario, habrá que realizar ajustes en procura de llevar las amenazas a un nivel aceptable.

El proceso que se ha presentado en la Figura 18, considera dos escenarios diferentes: el caso de un depósito que se va a desarrollar y el de un depósito ya existente.

En el caso de tratarse de un nuevo desarrollo, el proceso empezaría por la etapa de Identificación y clasificación de los datos, continuando con la Cuantificación del valor de éstos, la Definición de las expectativas de seguridad y sus métricas, y prosiguiendo con el ciclo compuesto por las etapas de Identificación de amenazas, Análisis o valoración de los riesgos, Selección y planificación de las medidas de mitigación y Seguimiento y control de tales medidas. Producto de tal seguimiento y control es factible que se deba realizar una nueva identificación de amenazas, con lo cual se iniciaría nuevamente el aludido ciclo.

Por el contrario, si de lo que se trata es de considerar el segundo escenario, esto es, mejorar la seguridad de un depósito existente, las primeras tres etapas del proceso, existe la posibilidad de que no se realicen del todo o tan sólo de forma parcial, ello va a depender de si con anterioridad se hizo lo pertinente para contar con los elementos o productos esperados de ellas. En caso de que efectivamente se hayan realizado los esfuerzos necesarios por identificar los datos y clasificarlos según su confidencialidad o privacidad, por asignarles un valor a éstos y por definir expectativas de seguridad y sus respectivas métricas, no tiene sentido volver a ejecutar tales etapas; tal vez lo más que se puede hacer es valorar el grado de validez que tienen tales resultados.

Si la situación es que el depósito de datos se desarrolló sin contar con tales elementos, lo más apropiado sería ejecutar las mencionadas etapas, en procura de contar con una base sólida sobre la cual sustentar el mejoramiento de la seguridad del depósito. El ciclo que involucra las etapas de Identificación de amenazas, Análisis o valoración de los riesgos, Selección y planificación de las medidas de mitigación, y Seguimiento y control de tales medidas definitivamente debe ser parte del proceso de mejoramiento de la seguridad del depósito.

Resta por comentar el flujo que se muestra entre la etapa de Identificación de las amenazas y las tres primeras del proceso (o sea, Identificación y clasificación de los datos, Cuantificación del valor de éstos, y Definición de las expectativas de seguridad y sus métricas). La intención de este flujo es dejar abierta la posibilidad de que, en caso de que al realizar la identificación de amenazas aparezca una de ellas que presente características diferentes a las consideradas en lo que se estableció previamente, (como producto de las mencionadas etapas) motivado por un cambio significativo en el entorno que se escape de lo analizado en esa oportunidad, se pueda evaluar y actualizar, en lo que corresponda, lo ya establecido.

4.1.2. Estrategia para la valoración de la propuesta de mejoramiento de la seguridad de un depósito de datos.

La propuesta anterior, tal y como se presenta, a efecto de establecer si las actividades que la componen realmente son necesarias o no dentro de un proceso de mejoramiento de la seguridad de un depósito de datos, requiere ser valorada. Ahora bien, en vista de que una valoración efectiva de la propuesta resulta difícil, por cuanto se requiere para ello de una empresa que esté próxima a iniciar el desarrollo de un depósito de datos, en la presente investigación se optó por realizar la valoración acudiendo al juicio experto de personas que estuvieran involucradas con el proceso de desarrollo y administración de una solución basada en depósitos de datos. Por lo tanto, para alcanzar con dicho propósito se utilizó la técnica de encuesta dirigida. En el Anexo 1 se puede apreciar el formulario de la encuesta que se aplicó.

La encuesta utilizada en esta investigación se dividió en cinco secciones, a saber:

- Acerca del informante
- Acerca de la empresa
- Acerca de la solución de depósito de datos

- Acerca de la metodología utilizada para crear la solución de depósito de datos
- Acerca de la propuesta formulada en la presente investigación

Con la primera sección se buscó establecer el grupo al cual pertenece la persona que respondió la encuesta, ya sea personal técnico involucrado con la solución, personal usuario de la solución, o bien, empresa que se dedica a desarrollar soluciones de este tipo.

La segunda sección buscó establecer el sector económico en que se circunscribe la empresa, así como los años desde que fue fundada, ello para establecer si se trata de empresas novicias, o por el contrario, de empresas ya consolidadas en su operación.

La tercera sección de la encuesta se centró en la solución de depósitos de datos que se desarrolló o estaba próxima a desarrollarse. En ese sentido se formularon preguntas relacionadas con los siguientes tópicos:

- Áreas que cubre la solución: si se trata de un solo departamento o hasta la totalidad de la empresa
- Estrategia de desarrollo: lo que cubre tan sólo con personal interno de la empresa, hasta llegar a punto de utilizar sólo personal externo, pasando por alternativas como una integración mixta, o el caso de contratar un consultor externo que asesore al grupo de trabajo
- Modalidad con que se concibió originalmente la solución: pudiendo ser OLAP, Data Warehouse, Data Mart, u otro
- Tiempo desde que se desarrolló la solución
- Tiempo desde que se inició el desarrollo de la solución
- Plataforma utilizada para la solución: pudiendo ser Microsoft SQL Server, Oracle, Sybase u otra
- Tamaño del repositorio de datos: de manera que se tuviera una idea de la magnitud de la solución
- Número de personas involucradas directamente en las labores de administración y mantenimiento de la solución
- Sector de usuarios al que está orientado principalmente la solución: pudiendo ser nivel estratégico, nivel táctico, nivel operativo o todos los niveles
- Responsable por la administración de la solución
- Responsable por el monitoreo de la solución

La cuarta sección abordó el tema de la metodología utilizada por la empresa para desarrollar la solución con preguntas como:

- ¿Se dispuso de una metodología específica para desarrollar la solución?
- ¿La metodología utilizada dio un tratamiento adecuado a los temas de seguridad?
- Al momento del desarrollo de la solución, ¿existieron preocupaciones en torno al tema de la seguridad y de qué forma las solventaron?

Finalmente, la quinta sección de la encuesta se abocó a valorar cuán adecuada resultó la propuesta planteada en esta investigación para cumplir con el cometido, cual es, mejorar la seguridad de las soluciones basadas en depósitos de datos. Para ello se formularon preguntas tendientes a valorar aspectos tales como:

- Pertinencia de identificar todos los datos que se van a considerar como parte de la solución a desarrollar
- Utilidad de clasificar el grado de confidencialidad o privacidad de los datos previamente identificados en categorías tales como: público, confidencial o muy confidencial
- Importancia de cuantificar el valor de los datos
- Utilidad de definir expectativas de seguridad y sus respectivas métricas al momento de desarrollar la solución
- Utilidad de identificar las amenazas que atentan contra la seguridad de las soluciones basadas en depósitos de datos
- Valorar agregado que se genera al analizar las amenazas
- Validez de seleccionar y planificar medidas que tiendan a mitigar las amenazas
- Importancia de dar seguimiento y controlar las medidas que adopte la organización para mitigar las amenazas
- Criterio en cuanto a repetir pasos previos como identificación de los datos, clasificación de estos según su grado de confidencialidad, cuantificación del valor de los datos y definición de expectativas de seguridad y sus métricas
- Opinión en cuanto a considerar como posibles amenazas que afectan a los activos físicos que componen el depósito de datos acciones tales como el robo, la destrucción accidental o el secuestro de éstos
- Criterio en cuanto a considerar un uso restrictivo de la información contenida en el depósito de datos

- Conveniencia de acudir a la criptografía como opción para ofrecer un uso seguro de la información
- Opinión en cuanto a la buena definición de permisos de acceso y la identificación clara de necesidades de información de los diferentes usuarios del depósito

Con el propósito de contrastar la información suministrada, se remitió la encuesta del Anexo 1 en dos tantos a las empresas participantes en la investigación, de modo que una de las copias fuera llenada por el personal técnico a cargo del depósito de datos (principalmente área informática) y la otra copia por el funcionario que juega el papel de usuario responsable del depósito. Con ello se esperaba poder recabar no sólo el criterio de los técnicos en materia de desarrollo y operación del depósito, sino también el criterio práctico o de la vivencia diaria de los usuarios de éste.

Una vez que los cuestionarios fueron respondidos éstos se sometieron a un proceso de crítica, en procura de solventar cualquier inconsistencia o ambigüedad que se pudiera presentar en torno a las respuestas aportadas. Posterior a ello, los formularios fueron procesados por medio del paquete estadístico SPSS. Se procuró no sólo conocer la distribución de frecuencia de las respuestas dadas, sino también el cruce de algunas de las variables, de modo que se pudieran detectar interrelaciones entre las respuestas que pudieran aportar un mejor criterio en torno a cuán adecuada resulta la propuesta.

4.2. Resultados de la validación de la propuesta para el mejoramiento de la seguridad de un depósito de datos.

Con el objetivo de validar la presente propuesta se cursaron correos electrónicos a diversas empresas, lo que incluyó tanto empresas desarrolladoras como empresas poseedoras de (o muy próximas a poseer) una solución basada en depósitos de datos.

En ambos casos se planteó el interés de que la encuesta fuera contestada tanto por personas pertenecientes al área técnica como por usuarios, de manera tal que se tuvieran elementos para analizar si la percepción de seguridad era igual por parte de esos dos grupos. Más aún, en el caso de las empresas desarrolladoras se les solicitó que la encuesta fuera contestada por funcionarios suyos, y de ser posible, por alguna de las empresas que constituyen su grupo de clientes.

En este punto vale la pena retomar lo indicado en la sección de Limitaciones de este documento, en lo referente a la conformación de la muestra del estudio. Sobre el particular, se debe recordar que en la actualidad no existe un registro de las empresas en Costa Rica que cuenten, o tengan planeado contar a corto plazo, con una solución de la naturaleza de los depósitos de datos. Ello impidió realizar un muestreo que tuviera validez estadística.

Por otro lado, y tal como ocurrió en el desarrollo del presente trabajo, por tratarse de un tema delicado como lo es el de la seguridad, existió el temor de las empresas de revelar información sensitiva acerca de ellas, lo que provocó que algunas de ellas no respondieran el cuestionario, a pesar de que se les advirtió acerca del tratamiento confidencial que se le daría a la información que ellos suministraran.

Por lo tanto, el único recurso que quedó fue sondear el medio para identificar qué empresas poseían o estaban muy próximas a iniciar un proyecto de depósito de datos, y que además estaban anuentes a responder el cuestionario.

Es de esta forma como se recibió un total 13 cuestionarios contestados, distribuidos de la forma que se muestra en el Gráfico 1.

En el Anexo 2 se muestra la lista completa de empresas que respondieron el cuestionario preparado al efecto; en el Anexo 1 se puede apreciar el aludido cuestionario.

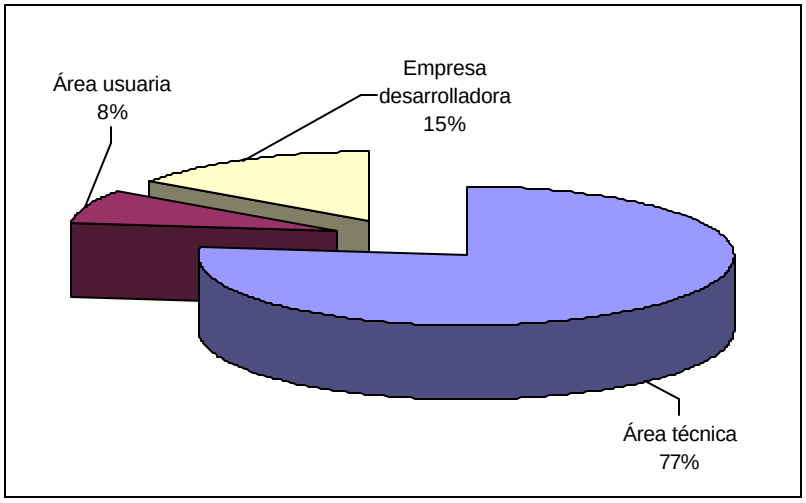


Gráfico 1. Composición del grupo de informantes.

De dicho gráfico se desprende que el 76.9% corresponde a informantes que pertenecen al área técnica (informática), un 15.4% corresponde a respuestas suministradas por empresas desarrolladoras y tan sólo un 7.7% (un informante) pertenece al área usuaria.

Aún cuando se esperaba que el cuestionario fuera respondido por un número muy similar de usuarios y técnicos, lo cierto es que tan sólo se recibió una encuesta respondida por la parte usuaria, lo cual podría estar influenciado por una falta de interés de parte de la persona con la que se hizo el contacto original de facilitar la copia del cuestionario a los usuarios, o bien, porque a pesar de que se les entregara, éstos optaran por no responderla.

Vale acotar que se invitó a varias empresas desarrolladoras a participar en este estudio; sin embargo, tan sólo dos de ellas aceptaron colaborar con la investigación contestando el cuestionario.

4.2.1. De las características de las empresas participantes en el estudio.

A efecto de conocer un poco acerca de las empresas participantes en la investigación, se formularon dos preguntas, a saber, sector económico en que se ubica su actividad principal y tiempo transcurrido desde su fundación.

En el Gráfico 2 se muestra la composición del grupo de estudio en lo referente a la actividad principal de la empresa.

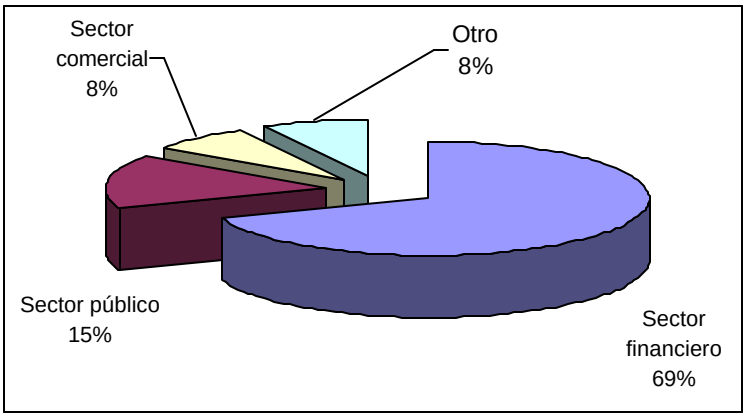


Gráfico 2. Sector económico en que se ubica la actividad principal de la empresa.

Del gráfico anterior se desprende que la mayoría de las empresas participantes en la investigación, y que poseen (o están próximas a disponer de) una solución basada en depósitos de datos pertenece al sector financiero, que bien pueden ser bancos, puestos de bolsa, o sus entes supervisores, situación que no está alejada de la

realidad, por cuanto es el sector financiero uno de los que mayormente demanda soluciones de este tipo.

En el Gráfico 3 se muestra la distribución de las empresas según su tiempo de fundada.

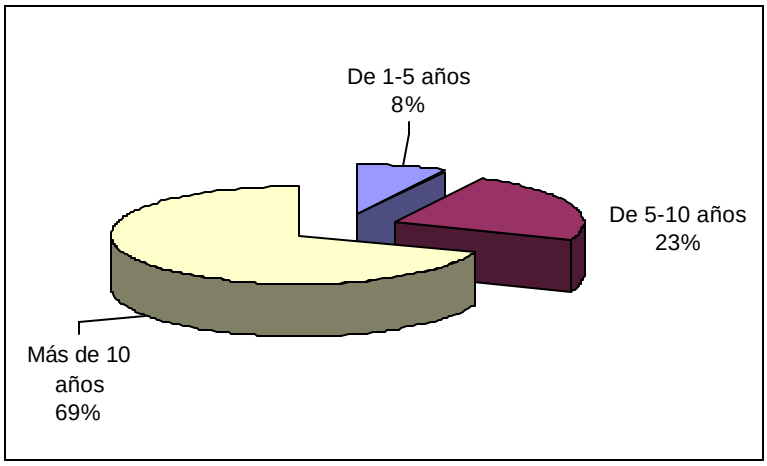


Gráfico 3. Tiempo de fundada la empresa.

Según este gráfico, cerca del 70% de las empresas participantes en la investigación tiene más de 10 años de haberse fundado, lo que es un indicador de que son empresas ya consolidadas.

4.2.2. De las características de la solución de depósito de datos.

Para conocer acerca de la solución de depósito de datos que poseen (o están próximas a poseer) las empresas participantes, se formularon preguntas orientadas hacia los siguientes tópicos:

- Áreas que cubre la solución
- Estrategia de desarrollo utilizada
- Concepción original de la solución
- Tiempo desde que se desarrolló la solución
- Tiempo desde que se inició el proyecto
- Plataforma en la que corre la solución
- Tamaño del repositorio de datos

- Número de personas dedicadas a labores de administración y mantenimiento de la solución
- Sector de la empresa al que se dedica la solución
- Responsable por la administración de la solución
- Responsable por monitorear la solución

En lo que respecta a las áreas que cubre la solución de depósito de datos, el Gráfico 4 muestra la distribución de respuestas. Según ese gráfico, en la mayoría de las empresas participantes (un 54%), este tipo de soluciones se ha desarrollado con el objeto que cubrir dos o más de sus áreas, sin llegar a cubrirla en su totalidad. En tan sólo en un 15% de los casos en que se indicó que la solución se diseñó con el propósito de atender las necesidades de información de toda la empresa.

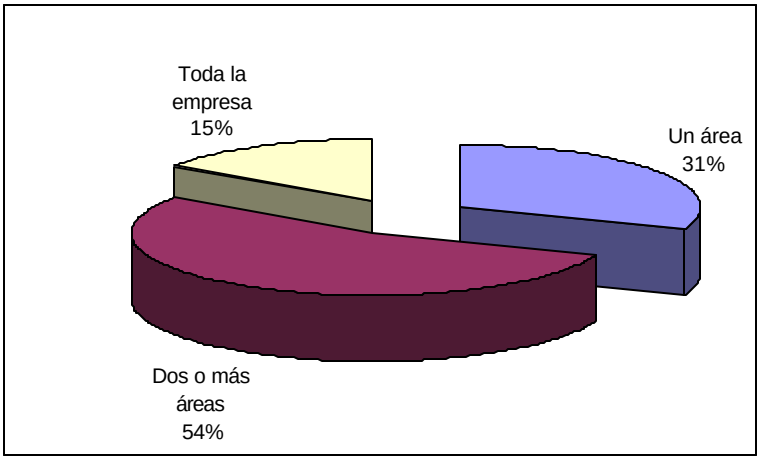


Gráfico 4. Áreas que cubre la solución de depósito de datos.

Según el Gráfico 5, la estrategia de desarrollo imperante entre las empresas consultadas fue la combinación de recurso humano interno y externo con un 69%. Llama la atención el caso en que la estrategia utilizada fue la contratación de un consultor que asesora al grupo de desarrollo, por cuanto con su 8% iguala a la alternativa de desarrollo externo y es tan sólo la mitad de la opción de desarrollar la solución empleando para ello tan sólo el recurso humano interno; ello por cuanto se trata de una alternativa interesante, y poco frecuente en este tipo de desarrollos.

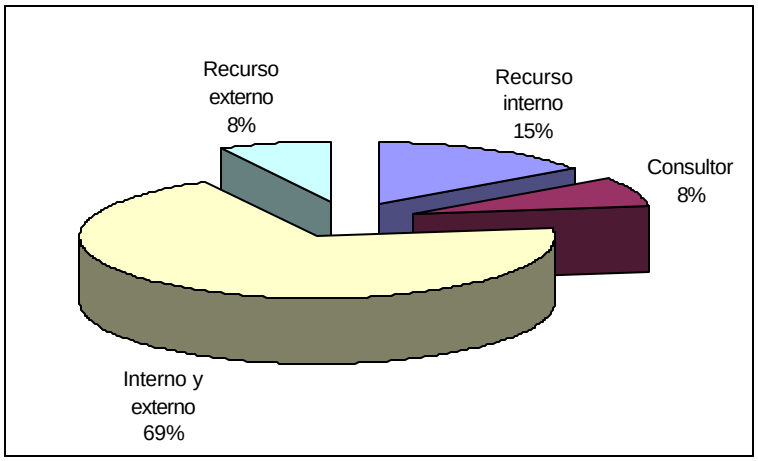


Gráfico 5. Estrategia de desarrollo utilizada.

El Gráfico 6 muestra la distribución en cuanto a la concepción original de la solución. Se aclara que la idea era conocer cuál fue la intención original del proyecto, lo que no necesariamente implica la naturaleza del producto que al final se generó.

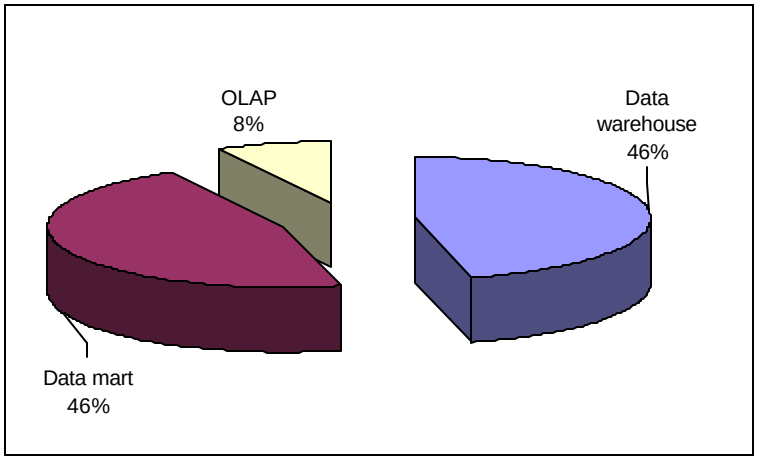


Gráfico 6. Concepción original de la solución de depósito de datos.

En torno a las respuestas obtenidas, se nota una paridad entre lo que son los *data warehouse* (o depósitos de datos) y los *data mart* (o mercados de datos), con un 46% para cada uno de ellos. Partiendo del hecho que la principal diferencia entre ellos es la cobertura que tienen, en donde por un lado se trata de toda la empresa, mientras que en el otro se refiere a un departamento o área de la empresa, es que se puede entender que algunas de las empresas hayan visto como estrategia desarrollar primero un mercado de datos para luego ir creciendo con la construcción de otros, hasta formar un depósito de datos, en el tanto se tomen las previsiones del caso.

Otros por el contrario puede haber concebido su solución directamente como un depósito de datos.

Al relacionar las respuestas obtenidas para la pregunta "Áreas que cubre la solución" (Gráfico 4) con las de esta pregunta (Concepción original de la solución, Gráfico 6) se obtuvo el siguiente resultado.

	Concepción original de la solución			
Área de cobertura	**Data warehouse**	**Data mart**	**OLAP**	**Total**
Un área	15,40%	15,40%	0,00%	30,80%
Dos o más áreas	23,10%	23,10%	7,70%	53,80%
Toda la empresa	7,70%	7,70%	0,00%	15,40%
Total	46,20%	46,20%	7,70%	100,00%

Independientemente de la cobertura de que se tratara, las concepciones de *data warehouse* y *data mart* fueron iguales. La única excepción se tuvo cuando la cobertura se trató de más de una área, pues aún cuando hubo una paridad entre ellas, también se indicó que la concepción fue de un OLAP.

En lo referente al tiempo desde que se desarrolló la solución, el Gráfico 7 muestra que una cuarta parte de las empresas consultadas la están desarrollando, y que de las restantes tres cuartas partes, en la mayoría de los casos la solución se puso a funcionar hace más de un año. Vale mencionar que una de cada tres empresas consultadas indicó que su solución de depósito de datos la habían puesto en funcionamiento entre los últimos seis meses y un año, lo que muestra que sus soluciones son muy recientes.

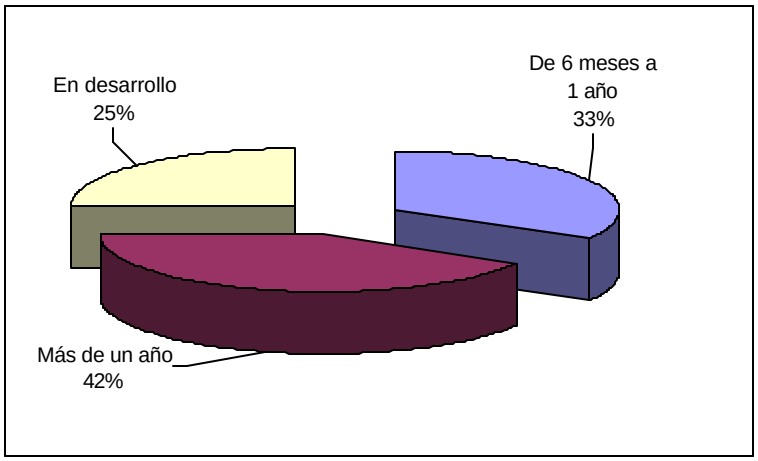


Gráfico 7. Tiempo transcurrido desde que se desarrolló la solución.

En lo tocante al tiempo transcurrido desde que inició el proyecto, en el Gráfico 8 se puede apreciar que dos de cada tres empresas consultadas indicó que éste inició desde hace uno a tres años. Destaca el caso en que un 17% de las empresas manifestó que el proyecto está por iniciarse, proporción que es igual a la suma conjunta de las empresas que indicaron que su proyecto había iniciado hace más de tres años y las que respondieron que el proyecto inició entre 6 meses y año. Agregando un poco los datos anteriores se tiene que en un 25% de los casos el proyecto tiene menos de un año de haberse iniciado o está por iniciar, mientras que el 75% de las empresas consultadas indicó que el proyecto tiene más de un año de haber comenzado, lo cual denota que la necesidad de este tipo de soluciones no es reciente, sino que por el contrario desde hace tal vez un par de años que las empresas muestran interés por esta alternativa.

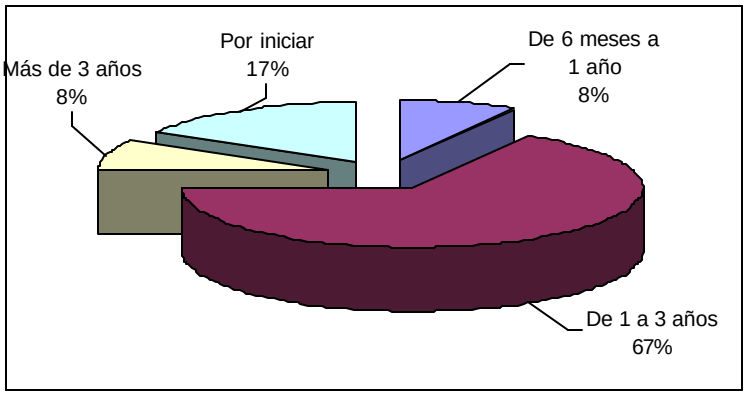


Gráfico 8. Tiempo transcurrido desde que inició el proyecto de depósito de datos.

El Gráfico 9 muestra el predominio de MS SQL Server como la plataforma de implementación para este tipo de soluciones, con un 54%.

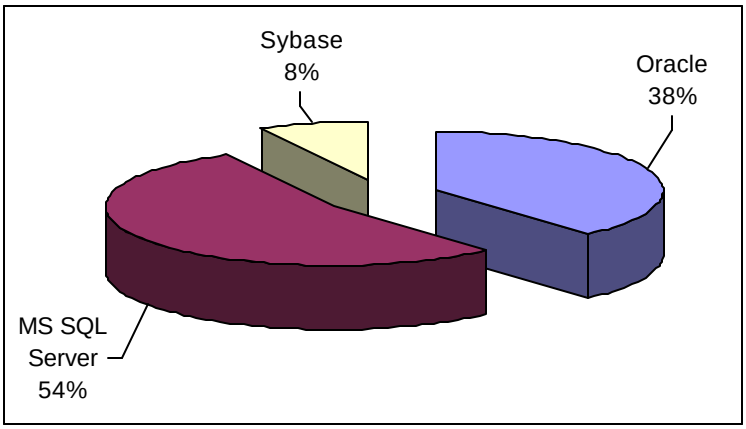


Gráfico 9. Plataforma en que corre la solución.

Sin embargo, la plataforma ofrecida por Oracle no resulta despreciable, pues en un 38% de las empresas consultadas se mencionó como la que ellos utilizan.

En lo referente al tamaño estimado del repositorio de datos, el Gráfico 10 muestra las respuestas obtenidas. Sobre el particular destaca el hecho de que en su mayoría (un 62%) tales repositorios no alcanzan los 200 GB, lo cual es un indicio de que se trata de soluciones relativamente pequeñas. De ese mismo gráfico se desprende que tan sólo en un 23% de las empresas consultadas se ha estimado que el repositorio puede alcanzar un tamaño superior a los 500 GB, en cuyo caso se advierte que para esas empresas sí se trata de una solución de depósito de datos de mediana cuantía.

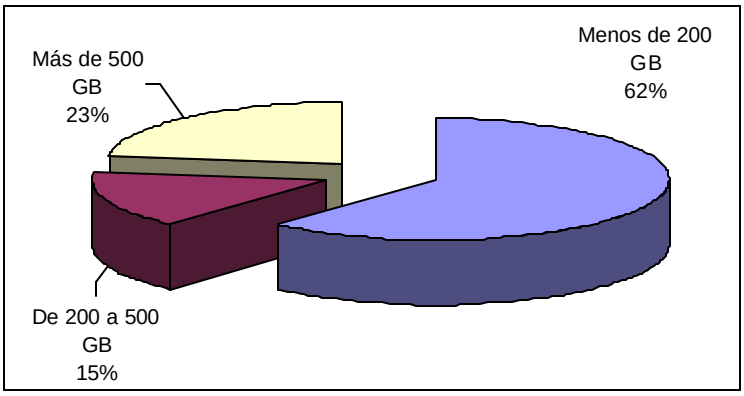


Gráfico 10. Tamaño del repositorio de datos.

Otra pregunta que se formuló fue aquélla relativa al número de personas involucradas directamente con la solución, principalmente en lo que atañe a las labores de administración y mantenimiento de ésta.

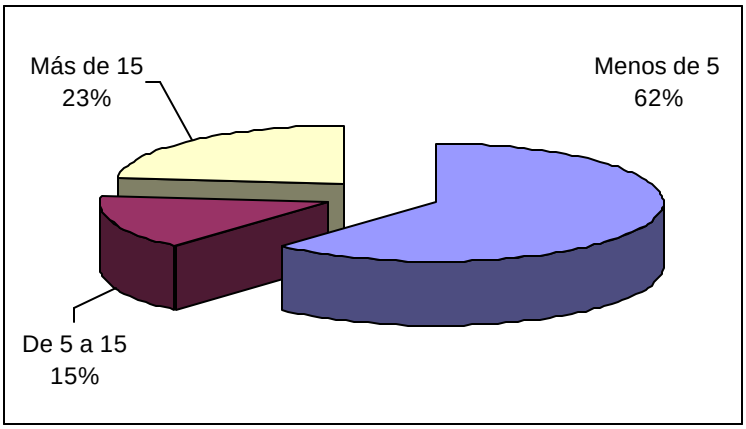


Gráfico 11. Número de personas involucradas en las labores de administración y mantenimiento.

Sobre el particular, el Gráfico 11 muestra que predomina en las empresas consultadas el criterio de que el grupo de personas debe ser reducido, en este caso no mayor a las 5 personas, que representó un 62% de las empresas, lo cual contrasta con un 23% de ellas que manifestó que ese grupo era de más de 15 personas. La experiencia indica que la tendencia es a que para atender adecuadamente una solución del tipo de depósitos de datos ronde las 10 personas, dependiendo del tamaño real de éste.

Por su parte, el Gráfico 12 muestra el nivel de la empresa al que está dirigida la solución, o sector meta de la solución. Sobre este aspecto, el 92% de las empresas consideró que el sector meta era al menos el nivel estratégico (pues se excluye de la totalidad solamente a las que indicaron que el sector meta era el nivel táctico, dado que la alternativa "Todo nivel" también incluye al nivel estratégico). Este resultado es coherente con el propósito que tienen las soluciones del tipo depósito de datos, cual es apoyar el proceso de toma de decisiones.

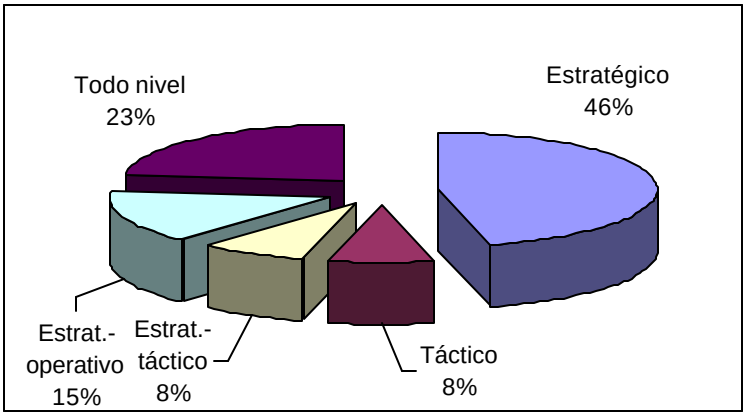


Gráfico 12. Nivel de la empresa al que se dedica la solución.

El Gráfico 13 muestra la forma como está definida la responsabilidad de administrar la solución de depósito de datos, por parte de las empresas consultadas en esta investigación. En la mayoría de ellas (un 54%), se consideró que tal responsabilidad debe recaer en el administrador de la base de datos (DBA, por sus siglas en inglés), y no tanto en un grupo del tipo OLTP (*On Line Transaction Processing*) que es la segunda forma mencionada, con un 38%. Este resultado lo que viene a representar es una situación en donde se entiende que este tipo de soluciones sigue siendo un asunto de una persona, y no tanto de un grupo como es la tendencia actual.

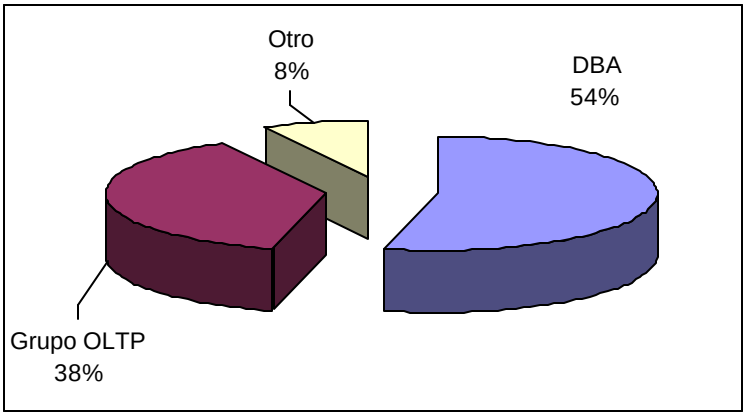


Gráfico 13. Responsable de administrar la solución.

El Gráfico 14 muestra el criterio de los encuestados en torno a la responsabilidad por el monitoreo de la solución. En este sentido, el 76% de las empresas manifestó que tal responsabilidad recaía en alguno de los funcionarios del área informática, mientras que las restantes alternativas (área usuaria, funcionario del nivel estratégico, o una combinación entre usuarios e informática) gozaron de una proporción idéntica del 8%. Si se analiza las respuestas desde otra perspectiva, se tiene que un 84% de las empresas consultadas consideró que la labor de monitoreo puede recaer al menos en un funcionario del área informática.

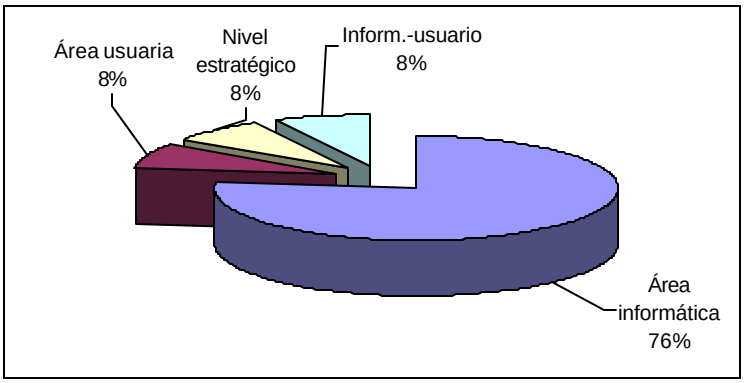


Gráfico 14. Responsable de monitorear la solución.

Con esta serie de preguntas se tiene una mejor idea de la forma cómo fueron concebidas y desarrolladas, y cómo están siendo administradas este tipo de soluciones, así como otras características generales de ellas.

4.2.3. De la metodología utilizada para crear la solución de depósito de datos.

Para conocer acerca de la metodología utilizada por parte de la empresa en la labor de desarrollar la solución de depósito de datos se plantearon tres preguntas concretas:

- ¿Dispuso la empresa de una metodología específica para el desarrollo de la solución?
- ¿Trató adecuadamente la metodología utilizada el tema de la seguridad en la solución a desarrollar?
- ¿Hubo alguna preocupación en torno al tema de la seguridad?

Ante la pregunta de si la empresa dispuso de una metodología específica para el desarrollo de la solución, el Gráfico 15 muestra la distribución obtenida de las respuestas. Según ese gráfico, la mayoría de las empresas (un 69%), dispuso de una metodología específica para tal efecto.

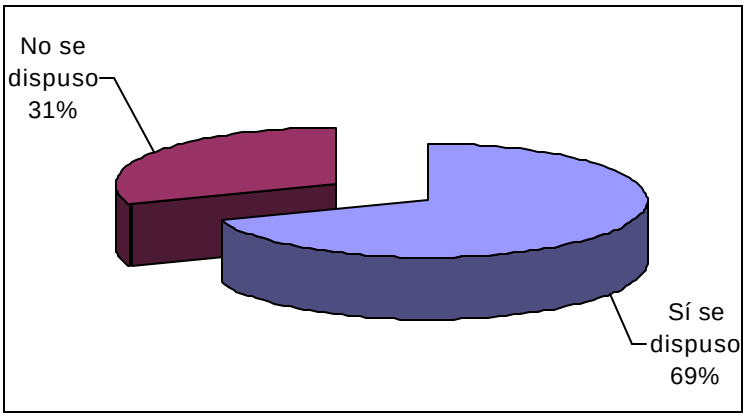


Gráfico 15. ¿Dispuso la empresa de una metodología específica para el desarrollo de la solución?.

Al consultárseles entonces cuál fue dicha metodología, las respuestas fueron las siguientes:

- Metodología Cognos Road Map
- Metodología Framework de Oracle
- Metodología de Ralph Kimball
- Metodología de la herramienta Business Object
- Metodología propuesta por la empresa contratada para el desarrollo (sin especificar cuál en particular)
- Combinación de la metodología de desarrollo en cascada con la que propuso la empresa contratada
- Metodología MSF de Microsoft y Metodología Cognos Road Map, enriquecidas con la experiencia de los clientes.

El Gráfico 16 muestra la situación resultante de tabular las respuestas a la pregunta referente a si se trató adecuadamente el tema de la seguridad en la metodología utilizada.

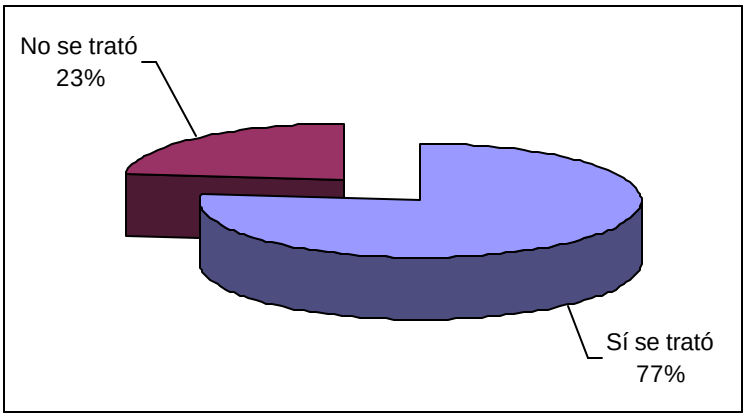


Gráfico 16. En la metodología utilizada, ¿se trató adecuadamente el tema de la seguridad?.

En este respecto, el gráfico muestra que en su mayoría (un 77%), las empresas manifestaron que la metodología dio un tratamiento adecuado al tema. Sin embargo, entre las empresas que indicaron que el tratamiento no fue el mejor mencionaron aspectos tales como que:

- Se le puso énfasis a aspectos de diseño tales como definición de requerimientos y diseño de estructuras y definición de roles del grupo de trabajo, pero no tanto al tema de la seguridad en sí
- No se le puso atención a la calidad de los datos de acceso a las consultas
- El cliente muchas veces no presta la atención adecuada al tema de la seguridad, de allí entonces que en la fase de visión se deban definir los accesos de la información, los perfiles de usuario, los riesgos de la liberación de la información, etc.

Para finalizar el tema de la metodología utilizada para el desarrollo de la solución, se le consultó a las empresas participantes que, independientemente de lo que establece esa metodología, si al desarrollar la solución se tuvo alguna preocupación en torno al tema de la seguridad. Para ello, el Gráfico 17 muestra la distribución de las respuestas obtenidas.

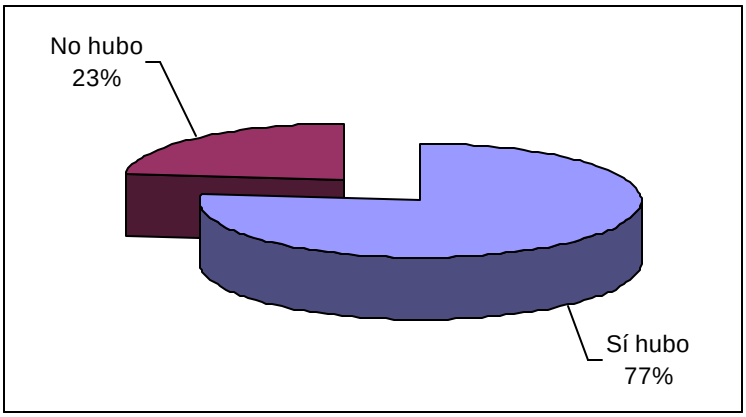


Gráfico 17. Al momento del desarrollo de la solución ¿hubo alguna preocupación en torno al tema de la seguridad?.

Se aprecia en este gráfico que en un 77% de los casos, sí hubo preocupaciones, entre las cuales se pueden mencionar las siguientes:

- Dependiendo de las necesidades de los usuarios y de las recomendaciones de la empresa desarrolladora, las soluciones de depósito de datos tienen sus niveles de seguridad de acuerdo con las políticas de seguridad de la empresa
- Se definen riesgos, pero no se les da seguimiento
- Permitir el acceso sólo a los usuarios que requieran de esa información
- Que la administración de la seguridad se manejara sólo por la herramienta
- Seguridad a nivel de base de datos, perfiles de acceso, perfiles para acceder a diferentes consultas, tanto a nivel de sistema operativo, como de SQL y software
- Que se consideraran aspectos tales como categorización o niveles de usuario, accesos de solo lectura aún en el *data warehouse*, seguridad a nivel físico (hardware), seguridad en los procesos de transformación
- Problemas de integridad al compartir datos entre empresas de la corporación y que los datos son utilizados para sustentar la toma de decisiones

4.2.4. De la propuesta formulada en esta investigación.

El propósito principal del cuestionario era valorar en qué grado la propuesta formulada en esta investigación cubría con suficiencia los aspectos más importantes a considerar dentro del tratamiento del tema de la seguridad en soluciones basadas

en depósitos de datos, razón por la cual se formuló una serie de preguntas que versaron sobre temas tales como:

- Identificar si resulta necesario hacer una identificación de todos los datos que se van a considerar como parte de la solución a desarrollar
- Realizar una clasificación de los datos identificados según su grado de confidencialidad va a resultar inútil en el proceso de definir la seguridad del depósito de datos
- Se considera que intentar cuantificar el valor de los datos resulta infructuoso
- Realizar una definición de las expectativas de seguridad y sus respectivas métricas resulta inútil
- Realizar una actividad tendente a identificar las amenazas puede resultar superflua
- Analizar las amenazas puede generar valor agregado
- Resulta ser el camino más apropiado seleccionar y planificar medidas que tiendan a mitigar las amenazas
- Dar seguimiento y controlar las medidas que adopte la organización para mitigar las amenazas puede resultar inútil
- Podría resultar valioso repetir pasos previos como identificación de los datos, clasificación de estos según su grado de confidencialidad, cuantificación del valor de los datos y definición de expectativas de seguridad y sus métricas
- Se deben considerar como posibles amenazas que afectan a los activos físicos que componen el depósito de datos acciones tales como el robo, la destrucción accidental o el secuestro de éstos
- Se debe considerar un uso restrictivo de la información contenida en el depósito de datos
- Se debe acudir a la criptografía como una opción para ofrecer un uso seguro de la información
- Se puede ofrecer un uso seguro y controlado de la información contenida en el depósito de datos a partir de una buena definición de permisos de acceso

Para aquellas preguntas en donde tuviera sentido hacerlo, se cruzaron las respuestas dadas a las preguntas de esta sección con la de la pregunta que se refirió al tratamiento que se le dio al tema de la seguridad en la metodología utilizada para el desarrollo de este tipo de soluciones.

La razón de hacer estos cruces era confirmar la validez de la respuesta dada a las preguntas referentes a la propuesta formulada en esta investigación, a partir de un

hecho concreto, como lo es el criterio que pudiera tener el informante acerca del tema de la seguridad en los depósitos de datos, que es precisamente lo que se les consultó cuando se les inquirió acerca de la forma de tratar este tema en la metodología empleada para el desarrollo de la solución.

4.2.4.1. Resulta necesario identificar todos los datos que se van a considerar.

Ante la consulta de si resulta necesario o no identificar todos los datos que estarían contenidos en una solución de depósito de datos, la mayoría de las empresas (un 85%) consideró que sí resulta necesario realizar esta tarea, tal y como se aprecia en el Gráfico 18.

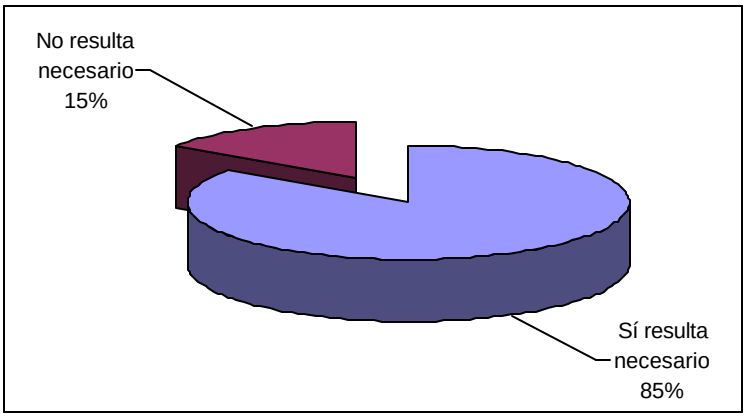


Gráfico 18. Resulta necesario identificar todos los datos que se van a considerar.

Las empresas que consideraron que tal labor no era necesaria adujeron razones tales como:

- Que el desarrollo de la solución se puede hacer en forma incremental, lo que no justifica entonces que se haga una identificación total al inicio del proyecto
- Que ello puede depender de la priorización de necesidades que hagan los usuarios

Al cruzar esta pregunta con la referida al tratamiento del tema de seguridad en la metodología utilizada se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	61.5%	23.1%	84.6%
No	15.4%	0.0%	15.4%
Total	76.9%	23.1%	100.0%

De aquí se desprende que del total de las empresas consultadas un 61.5% manifestó que era necesaria la tarea de identificar todos los datos a considerar en la solución a desarrollar, y que además el tratamiento del tema de seguridad en la metodología utilizada fue adecuado. Si se parte del hecho que fue un 84.6% de las empresas participantes las que indicaron que la tarea es necesaria, se tiene entonces aceptado por estas empresas que la tarea en comentario debe ser considerada dentro del conjunto de pasos a seguir para mejorar la seguridad de los depósitos de datos.

4.2.4.2. Resulta inútil clasificar los datos según grado de confidencialidad.

A las empresas participantes se les consultó acerca de si resulta inútil realizar una clasificación de los datos según su grado de confidencialidad. En el Gráfico 19 se aprecia la distribución de las respuestas.

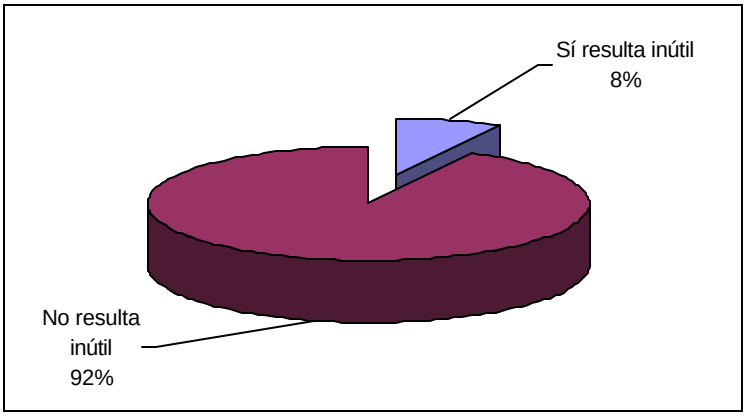


Gráfico 19. Resulta inútil clasificar los datos según grado de confidencialidad.

Según este gráfico, la mayoría de las empresas consultadas, un 92%, consideró que realizar esa labor no es inútil, es decir, que su realización produce algún beneficio.

Para la minoría de las empresas que manifestaron que realizar la clasificación sí resultaba inútil, se dieron respuestas tales como que:

- Existe información "sensible" para la toma de decisiones mezclada con información operativa, lo que complica llevar a cabo la aludida clasificación

Al cruzar esta pregunta con la referida al tratamiento del tema de seguridad en la metodología utilizada se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	7.7%	0.0%	7.7%
No	69.2%	23.1%	92.3%
Total	76.9%	23.1%	100.0%

Del total de las empresas consultadas un 69.2% manifestó que no consideraban inútil la realización de la mencionada clasificación, y que además en la metodología utilizada se dio un tratamiento adecuado al tema de la seguridad. Si se toma en consideración que del total de empresas participantes un 92.3% de ellas manifestó que realizar la clasificación no resultaba inútil, sin hacer distinción de la forma en que la metodología utilizada trató el tema de la seguridad, se puede dar por aceptado que esta labor también debe formar parte de la estrategia para mejorar la seguridad de los depósitos de datos.

4.2.4.3. Resulta infructuoso cuantificar el valor de los datos.

Ante la consulta que se le hiciera a las empresas participantes de si resultaba infructuoso cuantificar el valor de los datos, en su mayoría, el 85% de ellas respondió que la realización de una actividad no resultaba infructuosa, según se puede apreciar en el Gráfico 20.

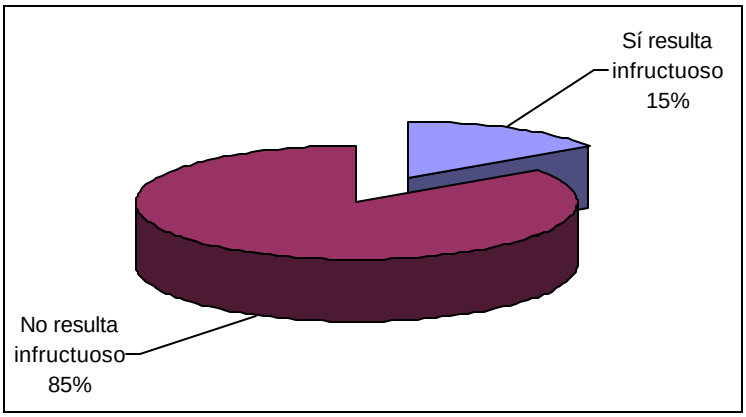


Gráfico 20. Resulta infructuoso intentar cuantificar el valor de los datos.

En relación con las empresas que respondieron que sí resultaba infructuoso realizar una actividad de esta naturaleza, se dieron opiniones como:

- Que se debía a la sensibilidad que presentaban ciertos datos, pues ello dificultaba realizar la cuantificación

Al cruzar esta pregunta con la referida al tratamiento del tema de seguridad en la metodología utilizada se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	7.7%	7.7%	15.4%
No	69.2%	15.4%	84.6%
Total	76.9%	23.1%	100.0%

En este caso, del total de las empresas consultadas un 69.2% opinó que realizar una actividad de este tipo no es infructuosa, máxime si se toma en cuenta que estas mismas empresas además indicaron que en metodología utilizada se dio un tratamiento adecuado al tema de la seguridad. Al considerar que del total de las empresas participantes un 84.6% de ellas manifestó que la realización de una actividad tendente a cuantificar el valor de los datos, independientemente del tratamiento que dio la metodología utilizada al tema de la seguridad, no es infructuosa da argumentos para concluir que esta actividad también es necesaria de considerar dentro de la estrategia para mejorar la seguridad de los depósitos de datos.

4.2.4.4. Resulta de poca utilidad definir expectativas de seguridad y sus métricas.

Otra de las preguntas que se le formuló a los encuestados trató acerca de si resultaba de poca utilidad definir algunas expectativas de seguridad y sus correspondientes métricas que le permitieran a las organizaciones determinar en qué medida las estaban alcanzando. Las respuestas a esta pregunta se muestran en el Gráfico 21.

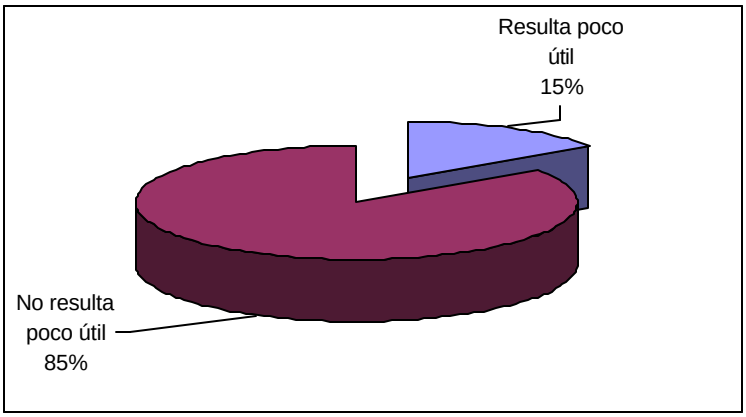


Gráfico 21. Resulta de poca utilidad definir expectativas de seguridad y sus métricas.

Como se puede apreciar en dicho gráfico, el 85% de los informantes manifestó que la realización de una actividad orientada en la línea antes reseñada no resulta de poca utilidad.

Entre las razones que se dieron para considerar tal situación se pueden mencionar las siguientes:

- La información es valiosa y si cae en malas manos puede ser peligroso para la empresa
- Desde el inicio del proyecto se deben tener presentes los aspectos de seguridad, sobre todo si se maneja información confidencial
- Los datos poseen un valor estratégico y la definición de un nivel adecuado de seguridad es vital
- Las expectativas de seguridad son de utilidad, más aún si se definen las métricas, porque permiten determinar si se alcanzan los niveles de seguridad aceptables para la empresa
- Ello depende del *data mart*, pues en su caso, se tienen *data mart* para las diferentes categorías de seguridad

- La seguridad es importante, más cuando se trata de información estratégica
- No todas las personas deben tener acceso a datos sensibles (o sensitivos)
- Es necesario que se definan tales expectativas de seguridad
- Es importante medir aspectos de seguridad, pues la información es la materia prima para la toma de decisiones y ello ayuda a tener control sobre el recurso informático que se presta

Al cruzar esta pregunta con aquélla destinada a determinar la forma cómo se trató el tema de la seguridad en la metodología utilizada, se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	0.0%	15.4%	15.4%
No	76.9%	7.7%	84.6%
Total	76.9%	23.1%	100.0%

Se observa que del total de las empresas consultadas un 76.9% manifestó que la definición de las expectativas sí es de utilidad, y que además en la metodología utilizada para el desarrollo del depósito de datos se le dio un tratamiento adecuado. Ello contrastado con el 84.6% de las empresas que indicaron que tal definición sí es de utilidad, sin importar el tipo de tratamiento que se le dio al tema de la seguridad en la metodología utilizada permite afirmar que esta actividad también debe formar parte de la estrategia que se debe utilizar para mejorar la seguridad en los depósitos de datos.

4.2.4.5. Resulta superflua realizar una actividad tendiente a identificar amenazas.

Ante la consulta que se hiciera de si resultaba superflua la realización de una actividad tendiente a identificar las amenazas que atentan contra los depósitos de datos, la totalidad de las empresas consultadas manifestó que de ninguna manera la actividad resultaba superflua, que por el contrario, una actividad de esta naturaleza debe formar parte del conjunto de pasos a seguir al tratar de mejorar la seguridad de los depósitos de datos. En el Gráfico 22 se muestra la situación en torno las respuestas obtenidas.

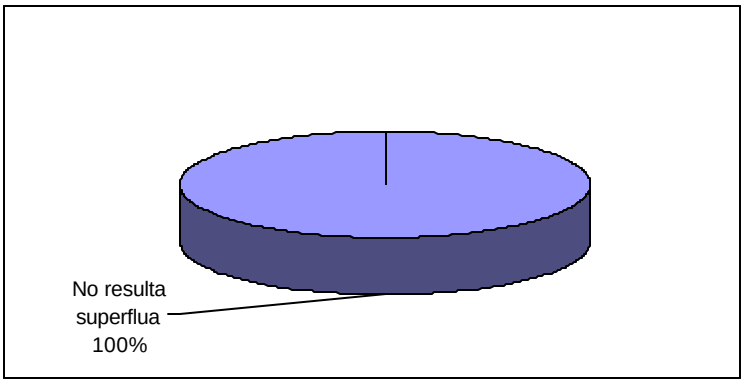


Gráfico 22. Resulta superflua realizar una actividad tendiente a identificar las amenazas.

4.2.4.6. Genera valor agregado analizar las amenazas.

Al ser consultadas las empresas participantes acerca de si al realizar una actividad que tuviera como objetivo analizar las amenazas ello genera algún valor agregado, la totalidad de los informantes indicó que efectivamente esa actividad producto un valor agregado tal que justifica su consideración dentro los pasos a seguir en procura de mejorar la seguridad de los depósitos de datos, tal y como se aprecia en el Gráfico 23.

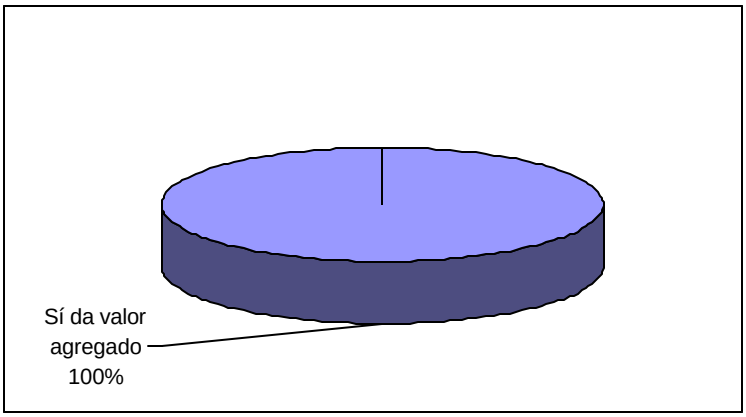


Gráfico 23. Genera valor agregado analizar las amenazas.

4.2.4.7. Seleccionar y planificar medidas mitigadoras resulta ser el camino más apropiado.

A las empresas participantes se les consultó acerca de si realizar una actividades dirigida a seleccionar y planificar las medidas mitigadoras de las amenazas resultaba ser el camino más apropiado para actuar. Una gran

mayoría de ellas, un 92%, consideró que efectivamente ese era el camino más apropiado de actuar, tal y como se muestra en el Gráfico 24.

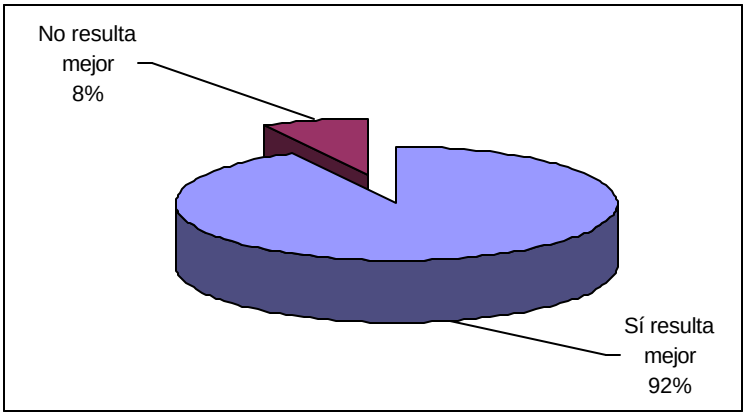


Gráfico 24. Seleccionar y planificar medidas mitigadoras resulta ser el camino más apropiado.

Entre los argumentos que dieron las empresas que consideraron que ese no era el mejor camino se dieron los siguientes:

- Tal vez la medida adoptada reduzca el riesgo, pero no necesariamente se alcanza el mínimo posible, debido a la cantidad de recursos que se pusieron a disposición para ello

Del cruce de esta pregunta con la referida al tratamiento del tema de la seguridad se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	69.2%	23.1%	92.3%
No	7.7%	0.0%	7.7%
Total	76.9%	23.1%	100.0%

De aquí se desprende que del total de empresas consultadas un 69.2% manifestó que actuar de la forma indicada anteriormente resulta el mejor camino a seguir, y que además indicó que el tratamiento del tema de la seguridad en la metodología utilizada para el desarrollo de soluciones del tipo depósito de datos fue apropiado. Si se parte del hecho que el 92.3% de las empresas participantes indicaron que esta forma de actuar es la correcta, se tiene entonces que una actividad dirigida en esta línea debe formar parte de los pasos a seguir para mejorar la seguridad de los depósitos de datos.

4.2.4.8. Resulta inútil dar seguimiento y controlar las medidas adoptadas.

Ante la consulta que se hiciera a las empresas participantes de si resultaba inútil dar seguimiento y controlar las medidas que se hayan adoptado para mitigar las amenazas que atentan contra los depósitos de datos, la totalidad de las empresas consultadas manifestó que de ninguna manera la actividad resultaba inútil, que por el contrario, una actividad de esta naturaleza debe formar parte del conjunto de pasos a seguir al tratar de mejorar la seguridad de los depósitos de datos. En el Gráfico 25 se muestra la situación en torno las respuestas obtenidas.

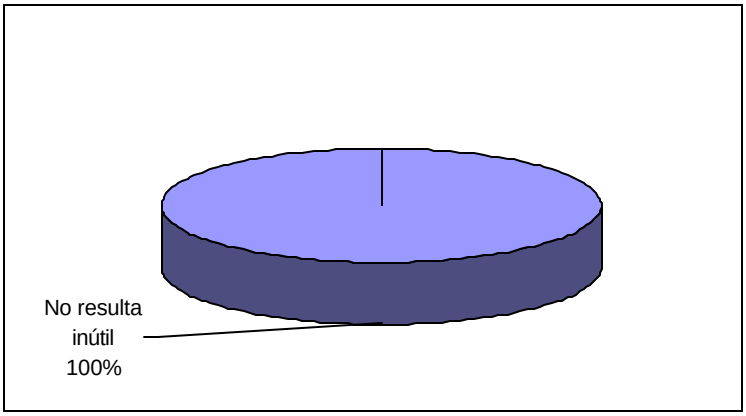


Gráfico 25. Resulta inútil dar seguimiento y controlar las medidas adoptadas.

4.2.4.9. Resulta valioso repetir pasos previos.

Ante la consulta que se le hizo a las empresas participantes acerca de si resulta valioso repetir los pasos previos de esta propuesta (identificación de todos los datos, clasificación de éstos según su grado sensitividad, cuantificar su valor y definir expectativas de seguridad y sus respectivas métricas), la mayoría de ellas, un 85%, consideró que realizar una actividad en esa línea verdaderamente resultaba valioso, tal y como se muestra en el Gráfico 26.

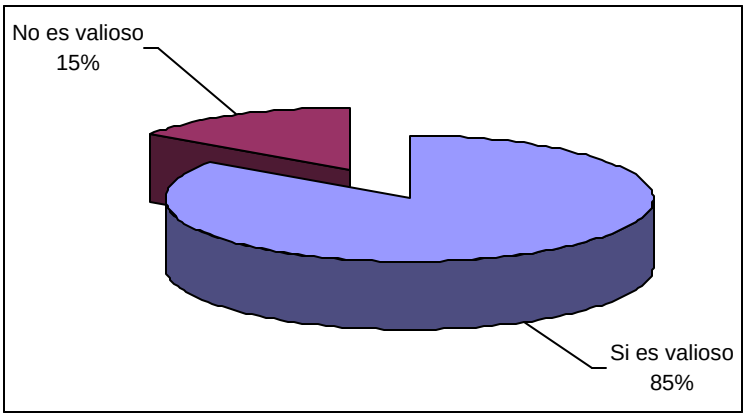


Gráfico 26. Resulta valioso repetir pasos previos.

Entre las empresas que respondieron negativamente a esta pregunta se dieron argumentos tales como:

- Que se pone en riesgo el proyecto. El proyecto debe seguir la línea de administración de proyectos, en donde deben existir fases concretas con productos claros, y el hecho de repetir pasos previos puede causar que los usuarios no le presten atención a la definición inicial. Por otra parte, cualquier metodología en este campo debe asegurar resultados en plazos muy cortos

Al cruzar esta pregunta con aquélla que trata acerca del tratamiento del tema de la seguridad en la metodología se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	61.5%	23.1%	84.6%
No	15.4%	0.0%	15.4%
Total	76.9%	23.1%	100.0%

Del total de las empresas consultadas un 61.5% manifestó que resulta valioso repetir los pasos previos de la propuesta, y que además el tratamiento del tema de la seguridad en la metodología empleada fue adecuado. Si al resultado anterior se le suma que el 84.6% del total de las empresas participantes en esta investigación respondió que sí resulta valioso repetir los aludidos pasos previos, si hacer distinción en cuanto a la forma de tratamiento del tema de la seguridad en la metodología utilizada para el desarrollo de depósitos de datos, se tiene entonces que existe fundamento

para concluir que una actividad como la comentada aquí debe formar parte de la estrategia que se formule para mejorar la seguridad de los depósitos de datos.

4.2.4.10. Se deben considerar acciones que mitiguen amenazas tales como robo, destrucción accidental y secuestro de los activos físicos.

Al ser consultadas las empresas participantes acerca de si se deben considerar acciones que tiendan a mitigar amenazas tales como el robo, la destrucción accidental y el secuestro de los activos físicos que conforman los depósitos de datos, la totalidad de los informantes indicó que efectivamente se debían de considerar acciones que minimicen los efectos de tales amenazas, las cuales debieran formar parte de los pasos a seguir en procura de mejorar la seguridad de los depósitos de datos, tal y como se aprecia en el Gráfico 27.

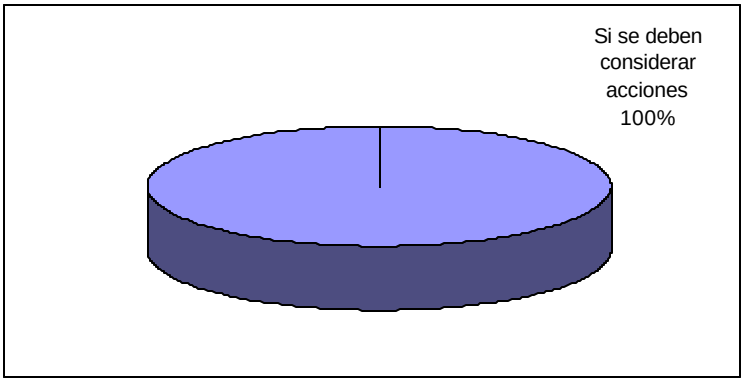


Gráfico 27. Se deben considerar acciones que mitiguen amenazas tales como robo, destrucción accidental y secuestro de los activos físicos.

4.2.4.11. Se debe considerar un uso restrictivo de la información del depósito de datos.

Cuando se le consultó a las empresas participantes su opinión respecto a considerar un uso restrictivo de la información contenida en el depósito de datos sustentado en el valor estratégico que tiene ésta, un 62% de las empresas manifestó que sí se debe considerar un uso restrictivo, tal y como se aprecia en el Gráfico 28.

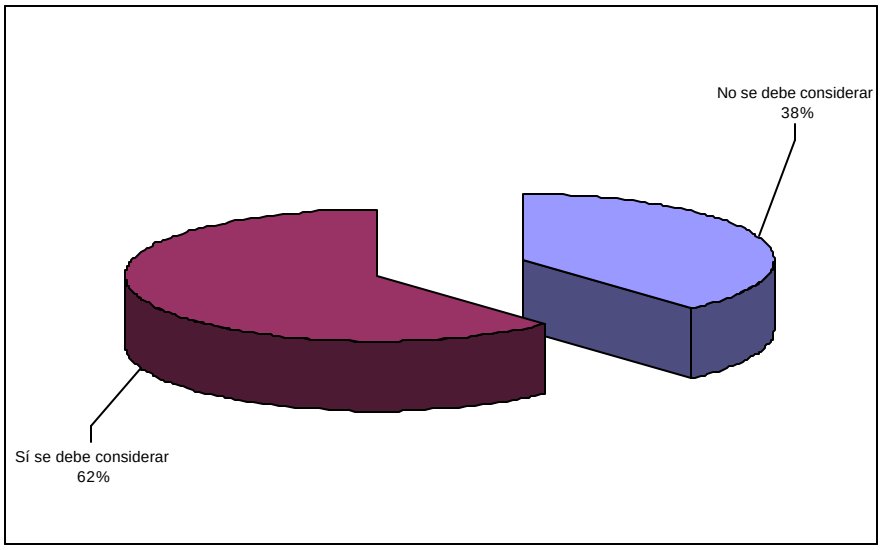


Gráfico 28. Se debe considerar un uso restrictivo de la información del depósito de datos.

Entre los argumentos vertidos para justificar que no se requiere un uso restrictivo de la información se dieron los siguientes:

- Que sólo las personas autorizadas tienen acceso a ese tipo de información
- Que no todos los depósitos de datos deben ser vigilados con la misma rigurosidad
- Que dada su orientación estratégica, no es conveniente que otras personas tengan acceso a toda la información importante de la empresa
- Que por tratarse de la materia prima para la toma de decisiones, es necesario tener control sobre los recursos informáticos
- Que se trata de información para la toma de decisiones gerenciales, y no debe ser de dominio público

Luego de cruzar esta pregunta con aquella otra que trata acerca de la forma en que se trató el tema de la seguridad en la metodología empleada, se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	46.2%	15.4%	61.6%
No	30.8%	7.7%	38.5%
Total	77.0%	23.1%	100.1%

Como se puede apreciar, de la totalidad de respuestas obtenidas, el 61.6% manifestó que sí se debe tener un uso restrictivo de la información contenida en el depósito. Una observación más minuciosa de la situación permite ver que ese resultado con tiene un 46.2% de respuestas de empresas que manifestaron que la metodología utilizada dio un tratamiento adecuado al tema de la seguridad, contra tan sólo un 15.4% de las empresas que manifestó que el tratamiento no fue adecuado. Aún cuando la combinación de las respuestas afirmativas de las dos preguntas da una mayoría, ésta no es del todo convincente.

4.2.4.12. Se debe acudir a la criptografía para cifrar toda la información, como una opción para ofrecer un uso seguro de ella.

Ante la consulta que se hiciera de si se debe acudir a la criptografía para cifrar toda la información contenida en el depósito de datos, como una opción para ofrecer un uso seguro de ésta, la totalidad de las empresas consultadas manifestó que no se debe acudir a tal extremo. En el Gráfico 29 se muestra la situación en torno las respuestas obtenidas.

Valga la oportunidad para advertir que el resultado de esta pregunta llama la atención, por cuanto hoy día la criptografía es la principal fuente a la que se acude en busca de seguridad de la información. Muestra de ello son los continuos esfuerzos por producir algoritmos criptográficos cada vez más robusto y seguro. A la postre, lo que puedo haber motivado este comportamiento es que la pregunta no estuviera claramente planteada, o bien, que exista desconocimiento o aversión de parte de las personas participante a este tipo de instrumentos.

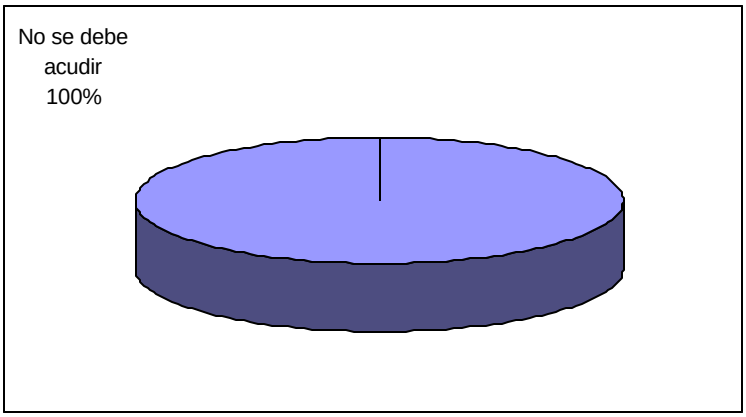


Gráfico 29. Se debe acudir a la criptografía para cifrar toda la información, como una opción para ofrecer un uso seguro de ella.

4.2.4.13. Se puede dar un uso seguro y controlado a partir de una buena definición de permisos.

A las empresas participantes se les consultó acerca de si se puede dar un uso seguro y controlado de la información contenida en el depósito de datos a partir de una buena definición de permisos de acceso. Una gran mayoría de ellas, un 92%, consideró que efectivamente sí se puede dar de esa forma, lo cual se muestra en el Gráfico 30.

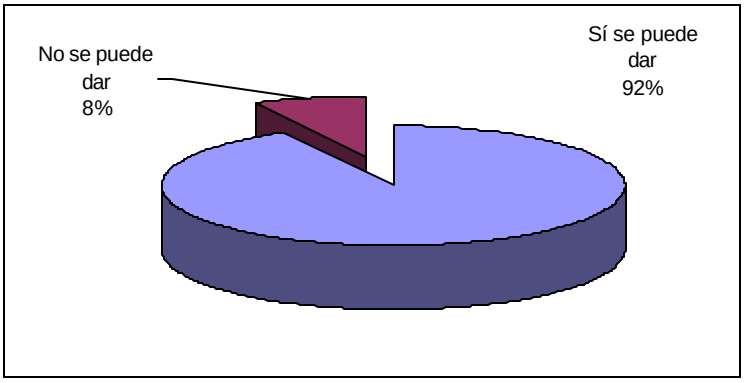


Gráfico 30. Se puede dar un uso seguro y controlado a partir de una buena definición de permisos.

Entre los argumentos que dieron las empresas que consideraron que ese no era el mejor camino se dieron los siguientes:

- Aún se puede correr el riesgo de que algún usuario pueda obtener información que no le competa

Al cruzar esta pregunta con la que se refiere a la forma de tratar el tema de la seguridad en la metodología utilizada se obtuvo el siguiente resultado.

	Trató adecuadamente tema seguridad		
	Sí	No	Total
Sí	69.2%	23.1%	92.3%
No	7.7%	0.0%	7.7%
Total	76.9%	23.1%	100.0%

De aquí se desprende que del total de empresas consultadas un 69.2% manifestó que sí se puede dar un uso seguro y controlado de la forma antes reseñada, y que además la metodología utilizada para el desarrollo de soluciones del tipo depósito de datos dio un tratamiento adecuado al tema de la seguridad. Si se parte del hecho que el 92.3% de las empresas participantes indicaron que esta forma de actuar es la correcta, se tiene entonces que una buena definición de permisos de acceso debe ser un elemento a considerar dentro las pautas a contemplar dentro de la estrategia que se defina para mejorar la seguridad de los depósitos de datos.

4.2.5. De la aceptación de la propuesta formulada en esta investigación.

Retomando la Figura 18, se tiene que la propuesta formulada en esta investigación consta de siete fases o etapas, a saber:

- Identificación y clasificación de los datos
- Cuantificación del valor de los datos
- Definición de expectativas de seguridad y sus métricas
- Identificación de las amenazas
- Análisis de las amenazas
- Selección y planificación de las medidas mitigadoras
- Seguimiento y control de las medidas mitigadoras adoptadas
- Repetir pasos o fases previas

De las secciones anteriores se tienen los resultados que se muestran en la Tabla 4.

Sección del documento	Número de la fase	Nombre de la fase	% de aceptación absoluto	% de aceptación relativo
4.2.4.1	1	Identificación de los datos	84.6	61.5
4.2.4.2	1	Clasificación de los datos	92.3	69.2
4.2.4.3	2	Cuantificación del valor de los datos	84.6	69.2
4.2.4.4	3	Definición de expectativas de seguridad y sus métricas	84.6	76.9
4.2.4.5	4	Identificación de amenazas	100.0	--
4.2.4.6	5	Análisis de las amenazas	100.0	--
4.2.4.7	6	Selección y planificación de las medidas mitigadoras	92.3	69.2
4.2.4.8	7	Seguimiento y control de medidas mitigadoras adoptadas	100.0	--
4.2.4.9		Repetir pasos previos	84.6	61.5

Tabla 4. Porcentaje de aceptación para cada una de las fases que componen la propuesta formulada.

En dicha tabla las columnas tienen el siguiente significado:

- Sección: Numeral de la sección del documento en que se analizó la actividad o fase
- Número y Nombre de la fase: Número y nombre de la fase que forma parte de la propuesta planteada
- % de aceptación absoluto: Porcentaje de aceptación de la fase o actividad, sin considerar el cruce de preguntas
- % de aceptación relativo: Porcentaje de aceptación de la fase o actividad, considerando que además el informante manifestó que la metodología utilizada para el desarrollo de la solución dio un tratamiento adecuado al tema de la seguridad.

Como se puede apreciar en la citada tabla, todas las actividades que conforman la propuesta analizada obtuvieron un porcentaje de aceptación absoluto significativo, al punto de que el menor valor observado fue de un 84.6%.

Si se evalúa la situación de las fases integrantes de la propuesta desde la perspectiva de la aceptación relativa, y partiendo del supuesto que se diera en su oportunidad en relación con la validez de la respuesta emitida por el informante a sabiendas de su conocimiento acerca del tema de la seguridad, se tiene entonces que todas las fases recibieron el beneplácito de los consultados, toda vez que los valores que se muestran están en relación con el total de las empresas consultadas, y que en ninguno de los casos alcanza un valor inferior al 50%.

Por lo tanto, del análisis presentado no se encuentran elementos que den pie a pensar que la propuesta es inadecuada.

Capítulo 5. Conclusiones y Recomendaciones

5. Conclusiones y recomendaciones.

5.1. Conclusiones.

5.1.1. Conclusiones generales.

- Esta investigación puso de manifiesto que no existe, dentro de las metodologías utilizadas para desarrollar soluciones basadas en depósitos de datos, un apartado que se dedique expresamente a tratar el tema de la seguridad en éstos. De hecho, el tratamiento que se le da al tema, al menos en las metodologías revisadas en este estudio, es tangencial o implícito a otras actividades, como la de determinación de requerimientos.
- En Costa Rica se carece de los elementos necesarios que permitan conocer qué empresas disponen de una solución basada en depósitos de datos. Más aún, tampoco es posible establecer qué empresas tienen contemplado en sus planes futuros incursionar en este tipo de sistemas de información.
- Luego de aplicar la encuesta que se preparó, con el propósito de establecer si la propuesta formulada en la investigación era adecuada o no para tratar el tema de la seguridad de los depósitos de datos, no se obtuvo un resultado tal que desaprobara o que permitiera concluir que la consideración de cada una de las actividades que componen la propuesta resulte innecesaria o superflua para mejorar la seguridad en soluciones del tipo de depósitos de datos, ya fuera que la solución existiera o que estuviera próxima a desarrollarse.

5.1.2. Conclusiones derivadas de la encuesta.

- De acuerdo con la encuesta aplicada en esta investigación (y teniendo en cuenta las limitaciones que se presentaron en torno a ella), la mayoría de las empresas que disponen de una solución del tipo depósito de datos se ubica en el sector financiero, lo que incluye a las empresas financieras en general y a sus órganos supervisores (las tres superintendencias: SUGEF, SUGEVAL y SUPEN).
- De los resultados de la encuesta utilizada se desprende que casi un 70% de las empresas participantes son organizaciones consolidadas en su accionar, con más de 10 años de haber sido fundadas o creadas.
- La encuesta puso en evidencia que tan sólo en un 31% de las empresas participantes en la investigación se pensó que la solución a desarrollar cubriría una sola área o departamento, en contraposición con el 69% de ellas que concibió que la solución abarcaría a más de un área o toda la empresa.

Este resultado es consecuente con la filosofía de los depósitos de datos al apoyar el proceso de toma de decisiones, lo que cubrir a toda la empresa.

- En materia de estrategia de desarrollo, la encuesta sacó a la luz que la combinación de recurso humano interno con personal externo es el predominante entre las empresas participantes en la investigación. Ello hace pensar que las empresas no desean que sus sistemas para la toma de decisiones sean desarrollados exclusivamente por personal externo, posiblemente por tratarse de sistemas específicos (que de alguna forma constituyen su *know how* del negocio), y que tampoco optan por el desarrollo con solo recurso interno, motivado a la postre por el grado de experticia que se requiere para un desarrollo de este tipo, que en muchas organizaciones se torna difícil de adquirirlo por cuanto implica incurrir en grandes inversiones de tiempo y dinero.

- De los resultados que arroja la encuesta se tiene que se existe un balance entre las empresas que concibieron la solución como un *data warehouse* y las que la concibieron como un *data mart*; constituyen una minoría poco significativa las empresas que manifestaron que la solución se concibió como un herramienta OLAP. Teniendo en cuenta que la principal diferencia que existe entre un *data mart* y un *data warehouse* es la cobertura de la solución, no es de extrañar el resultado obtenido dado que a la postre se refieren a lo mismo, por cuanto ello podría estar influenciado por la estrategia que al respecto haya establecido la organización (o sea, iniciar con un *data mart* y más adelante evolucionar a un *data warehouse*, al ir agregando otros *data mart* departamentales, hasta cubrir la totalidad de la organización).

- Un poco más del 40% de las empresas consultadas manifestó que disponen de una solución de depósito de datos desde hace más de un año, mientras que tan sólo un 25% del total de ellas se encuentra en proceso de desarrollo. Este resultado muestra que el uso de este tipo de soluciones no data mucho de haber incursiona en el medio costarricense, y que hay una tendencia importante entre las empresas por desarrollar sus propias soluciones de este tipo. Este hecho se ve reforzado al contrastar ello con las respuestas dadas por las empresas a la pregunta de desde hace cuánto habían iniciado sus proyectos de desarrollo, pues casi un 75% del total de las empresas consultadas respondió que sus proyectos habían iniciado al menos desde hace un año.

- La encuesta puso de manifiesto que la plataforma que predomina en Costa Rica para sustentar soluciones de depósitos de datos es MS SQL Server, seguido de la plataforma Oracle.

- En términos generales, se tiene que las soluciones que se han desarrollado, según la encuesta aplicada, se ubican en un rango de menos de los 200 GB, en lo que respecta al tamaño de la base de datos que sustenta la solución de depósito de datos. Este volumen de la base de datos corresponde a

soluciones pequeñas, posiblemente aún en estado de prueba o no consolidadas.

- En lo que respecta al número de personas directamente involucradas en las labores de mantenimiento y operación de la solución de depósito de datos, la encuesta arroja como resultado que la mayoría de empresas dedican un grupo no mayor a las 5 personas para ello. Este resultado muestra una actitud conservadora de parte de las empresas, en el sentido de que se estila que los grupos a cargo de este tipo de soluciones sean mayores en virtud de la complejidad de las tareas que se deben realizar.
- Partiendo del hecho de que los depósitos de datos están orientados a soportar los procesos de toma de decisiones, la encuesta aplicada mostró que entre las empresas consultadas prevalece significativamente el criterio de que el sector meta de la solución es en su mayoría al menos el nivel estratégico, pues tan sólo una proporción muy baja de las empresas consultadas manifestó en forma exclusiva que el sector fuera únicamente el nivel táctico. Al igual que en una conclusión anterior, este resultado viene a confirmar que la intención de las soluciones basadas en depósitos de datos es apoyar el proceso de toma de decisiones.
- Tratándose del tema de la definición de la responsabilidad de administrar la solución de depósito de datos, la mayoría de las empresas consultadas consideró que tal responsabilidad debe recaer en el administrador de la base de datos (DBA), y no tanto en un grupo del tipo OLTP (*On-Line Transaction Processing*). Este resultado lo que viene a representar es una situación en donde se entiende que este tipo de soluciones sigue siendo un asunto de una persona, y no tanto de un grupo como es la tendencia actual.
- Por su parte, en lo que respecta a la definición de la responsabilidad por el monitoreo de la solución, según los resultados obtenidos a partir de la encuesta, se tiene que en su mayoría las empresas manifestaron que tal responsabilidad debía recaer en alguno de los funcionarios del área informática, lo cual muestra una creencia, tal vez errónea entre las empresas, de que el tema de monitoreo es netamente técnico, cuando a la postre éste sea realmente un tema que atañe más a los usuarios, por ser ellos los "dueños" de la aplicación, y quienes tienen un mejor criterio acerca de los datos y de la mecánica de este tipo de aplicaciones.
- Una mayoría de las empresas consultadas manifestó disponer de una metodología específica para el desarrollo de soluciones del tipo de depósito de datos. Sin embargo, al pasar lista de las metodologías mencionadas, se aprecia que éstas corresponden a la metodología sugerida por la empresa contratada para desarrollar la solución, o bien, por un híbrido entre la metodología utilizada por la organización para su desarrollo de aplicaciones y la recomendada por la empresa, lo que permite afirmar que la empresa en sí misma no dispone de una metodología propia para este tipo de desarrollos.
- En su mayoría, las empresas consultadas manifestaron que el tema de la seguridad en los depósitos de datos se trató adecuadamente a la hora de

desarrollar la solución. Es precisamente esta valoración del tratamiento del tema de la seguridad, la que se utilizó como elemento para confrontar si la propuesta planteada en la presente investigación fue avalada o no por los informantes, al amparo de que para responder esta pregunta se debe tener el criterio suficiente.

- Según lo expuesto en el Capítulo 4 de este documento, la propuesta formulada consta de siete fases o etapas, las cuales fueron sometidas a consulta en forma separada, obteniéndose los resultados de aceptación que se detallan a continuación:

Número de la fase	Nombre de la fase	% de aceptación
1	Identificación de los datos	84.6
1	Clasificación de los datos	92.3
2	Cuantificación del valor de los datos	84.6
3	Definición de expectativas de seguridad y sus métricas	84.6
4	Identificación de amenazas	100.0
5	Análisis de las amenazas	100.0
6	Selección y planificación de las medidas mitigadoras	92.3
7	Seguimiento y control de medidas mitigadoras adoptadas	100.0
	Repetir pasos previos	84.6

De allí se desprende que todas y cada una de las fases que conforman la propuesta recibieron, un grado de aceptación de un 84.6%, o mayor, de parte de las empresas que fueron consultadas. A pesar de la limitación ya apuntada, en cuanto a que no se tuvo una población debidamente identificada, que permitiera seleccionar una muestra que tuviera una representación significativa estadística, los resultados de la encuesta aplicada al menos confirman que entre las empresas consultadas hay consenso en que las fases que componen la propuesta formulada en esta investigación son de utilidad para mejorar la seguridad en soluciones del tipo de depósitos de datos.

5.2. Recomendaciones.

- Someter a una prueba práctica la propuesta formulada. Para ello se sugiere, identificar alguna empresa que esté próxima a desarrollar una solución basada en depósitos de datos, y que esté anuente a realizar la prueba. Una vez que se tenga seleccionada la empresa, proceder a integrar la propuesta aquí formulada con su metodología para el desarrollo de depósito de datos. Posteriormente aplicar la metodología que resulte de esta integración y desarrollar la aplicación. Una vez que la aplicación esté desarrollada seleccionar algunas de las amenazas identificadas y determinar en qué grado la solución de depósitos de datos que se desarrolló la minimiza.
- Promover la utilización de la presente propuesta a través de la divulgación entre empresas desarrolladoras y profesionales (ya sean graduados o próximos a graduarse). Para lo primero se puede emplear una divulgación masiva de la propuesta, mientras que para lo segundo se puede incorporar el fondo de la investigación en los cursos que se dicten en las diferentes casas de enseñanza superior.
- Replantearse el alcance de la investigación en procura de conocer un poco más en torno a la situación real de los depósitos de datos en Costa Rica, tanto en lo que respecta a las preguntas formuladas en el cuestionario empleado, como en lo que respecta a la cobertura de la investigación, de modo que se pueda realizar un estudio con un grado de validez estadística aceptable, y en que se tenga una representación más realista de los sectores de la economía costarricense en los que ha permeado esta tecnología.

Bibliografía

Bibliografía.

A. Bibliografía referenciada.

[ATRE1988] Shakuntala Atre, *Técnicas de Bases de Datos Estructuración en diseño y administración*, Editorial Trillas, México, 1988.

[CHAPM1997] D. Brent Chapman, Elizabeth D. Zwicky, *Construya Firewalls para Internet*, McGraw-Hill, México, 1997.

[CORE1998] Michael Corey, Michael Abbey, Ian Abramson, Ben Taub, *Oracle8 Data Warehousing*, Osborne/McGraw-Hill, USA, 1998.

[CORE1999] Michael Corey, Michael Abbey, Ian Abramson, Larry Barnes, Benjamin Taub, Rajan Venkitachalam, *SQL Server 7 Data Warehousing*, Osborne/McGraw-Hill, USA, 1999.

[CORE2001] Michael Corey, Michael Abbey, Ian Abramson, Ben Taub, *Oracle8i Data Warehousing*, Osborne/McGraw-Hill, USA, 2001.

[DATE1987] C. J. Date, *Bases de Datos Una guía práctica*, Addison-Wesley Iberoamericana, México, 1987.

[DEVL1998] Barry Devlin, Metadata: The Warehouse Atlas, *DB2 Magazine*, Vol. 3, N° 1, Spring 1998, págs. 8-9.

[FRYM1997] Lowell W. Fryman, Audits and Controls for the Data Warehouse, *IS Audit & Control Journal*, Vol. IV, 1997, págs. 46-48.

[GAMB1998] Rocío Gamboa Gamboa, Erick Sánchez Del Valle, *Propuesta de una estrategia para el desarrollo de un depósito de datos para el Consejo de Seguridad Vial*, Tesis, Licenciatura en Ingeniería de Sistemas Informáticos, Universidad Latina de Costa Rica, Costa Rica, Set. 1998.

[GILL1996] Harjinder S. Gill, Prakash C. Rao, *La integración de información para la mejor toma de decisiones. Data Warehousing*, Prentice Hall Hispanoamericana, México, 1996.

[GREE2001] Rick Greenwald, Robert Stackowiak, Jonathan Stern, *Oracle Essentials Oracle9i, Oracle8i & Oracle8*, O'Reilly & Associates, Inc., USA, 2001.

[HADF1997] Lee Hadfield, Dave Hatter, Dave Bixler, *Windows NT Server 4 Security Handbook*, Que Corporation, USA, 1997.

[HAMM1996] Tom Hammergren, *Data Warehousing. Building the Corporate Knowledgebase*, International Thomson Computer Press, USA, 1996.

[HARE1996] Chris Hare, Karanjit Siyan, *Internet Firewalls and Network Security*, 2nd Ed., New Riders Publishing, USA, 1996.

[HUFF1996] Duane Hufford, Data Warehouse Quality: Special Feature from January 1996, *DM Review*, Jan. 1996, URL: **http://www.dmreview.com/master.cfm?NavId=55&EdId=1311**.

[INMO1995] William H. Inmon, Tech Topic What is a Data Warehouse?, *Prism*, Vol. 1, N° 1, USA, 1995, URL: **http://www.cait.wustl.edu/cait/papers/prism/vol1_no1**.

[INMO1996] William H. Inmon, *Building the Data Warehouse*, 2nd Ed., John Wiley & Sons, USA, 1996.

[INMO1997a] William Inmon, J. D. Welch, Katherine L. Glassey, *Managing the Data Warehouse*, John Wiley & Sons, USA, 1997.

[INMO1997b] William Inmon, Security in the Data Warehouse/Internet Environment, *IS Audit & Control Journal*, Vol. IV, 1997, págs. 8-11.

[INMO1998] William Inmon, Wherefore Warehouse?, *Byte*, Vol. 23, N° 1, Jan. 1998, págs. 88NA 1- 88NA 10.

[JIME1998b] Beatriz Jiménez S., Rafael Ávalos C., *Depósitos de datos*, Club de Investigación Tecnológica, Costa Rica, Nov. 1998.

[KIMB1998] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, *The Data Warehouse Lifecycle Toolkit*, John Wiley & Sons, USA, 1998.

[KIMB2001b] Ralph Kimball, Catastrophic Failure, *Intelligent Enterprise*, Vol. 4, N° 17, 12 Nov. 2001, págs. 20,22,48. URL: **http://wwww.iemagazine.com/011112/417warehouse1_1.shtml**.

[MORI1996b] Terry Moriarty, Christine Mandracchia, Heart of the Warehouse, *Database Programming & Design*, Vol. 9, N° 11, Nov. 1996, págs. 70-74.

[MOSS1998] Larissa Moss, Data Cleansing: A Dichotomy of Data Warehousing?, *DM Review*, Feb. 1998, URL: **http://www.dmreview.com/master.cfm?NavId=55&EdId=828**.

[PAZ1999] José Rafael Paz Barahona, *Depósito de datos vs. Almacén de datos operacional: Mejorando el entendimiento de ambos procesos*, Tesis, Magister Scientiae en Computación, Instituto Tecnológico de Costa Rica, Costa Rica, Mar. 1999.

[RAE1992] Real Academia Española, *Diccionario de la Lengua Española*, 21° Edición, Madrid, España, 1992.

[SHAP2001] Jeffrey R. Shapiro, *SQL Server 2000: The Complete Reference*, Osborne McGraw-Hill, USA, 2001.

[SHEL1997] Tom Sheldon, *Windows NT Security Handbook*, Osborne McGraw-Hill, USA, 1997.

[SIMO1997b] Alan Simon, *Data Warehousing for Dummies*, IDG Books Worldwide, USA, 1997.

[SING1999] Harry Singh, *Interactive Data Warehousing*, Prentice Hall PTR, USA, 1999.

[VALL1989] S. Rao Vallabhaneni, *Auditing Computer Security. A Manual with Case Studies*, John Wiley & Sons, USA, 1989.

[WARI1997] Slemo Warigon, *Data Warehouse Control and Security*, Association of College and University Auditors Ledger, Vol. 41, N° 2, Apr. 1997, págs. 3-7, URL: **http://all.net/books/audit/kits/dw.html**.

[WEBE1988] Ron Weber, *EDP Auditing Foundations and Practice*, 2nd Ed., McGraw-Hill, USA, 1988.

B. Bibliografía consultada.

[ALUR1995] Nagraj Alur, Missing Links in Data Warehousing, *Database Programming & Design*, Vol. 8, N° 9, Sept. 1995, págs. 21-23.

[AMO2000] William C. Amo, *Microsoft SQL Server OLAP Developer's Guide*, M&T Books, USA, 2000.

[ATKI1997] Derek Atkins, Paul Buis, Chris Hare, Robert Kelley, Carey Nachenberg, Anthony B. Nelson, Paul Phillips, Tim Ritchey, Tom Sheldon, Joel Snyder, *Internet Security Professional Reference*, 2nd Ed., New Riders Publishing, USA, 1997.

[AWAL2000] Don Awalt, Brian Lawton, Data Warehousing: Back to Basics, *SQL Server Magazine*, Feb. 2000, URL: **http://www.sqlmag.com/Articles/ArticleID=7833**.

[BARA1996a] Ahmad Baraani-Dastjerdi, Josef Pieprzyk, Reihaneh Safavi-Naini, *A Security Model for Multi-level Object-Oriented Databases Based on Views*, Department of Computer Science, University of Wollongong, Australia, Feb. 5, 1996.

[BARA1996b] Ahmad Baraani-Dastjerdi, Josef Pieprzyk, Reihaneh Safavi-Naini, *Security in Databases: A Survey Study*, Department of Computer Science, University of Wollongong, Australia, Feb. 7, 1996.

[BAUM1996a] David Baum, Building Blocks of Data Warehouses, *DataBased Advisor*, Vol. 14, N° 4, Apr. 1996, págs. 48-51.

[BAUM1996b] David Baum, Getting Data to Those Who Need It, *DataBased Advisor*, Vol. 14, N° 10, Oct. 1996, págs. 58-59.

[BAUM1997a] David Baum, Recipe for Success: Mixing Your Warehouse with the Web, *DataBased Advisor*, Vol. 15, N° 2, Feb. 1997, págs. 55-57.

[BAUM1997b] David Baum, Planning and Implementing a Data Warehouse, *Byte*, Vol. 22, N° 6, June 1997, págs. 120C-120H.

[BERG2000] Stacy Bergert, Power to the People, *Intelligent Enterprise*, Vol. 3, N° 18, Dec. 5, 2000, URL: **http://www.intelligententerprise.com/001205/feat4.shtml**.

[BERL2001] David Berlind, How ready is your business for worst-case scenario?, *Enterprise*, Oct. 11, 2001, URL: **http://techupdate.zdnet.com/techupdate/stories/main/0,14179,2817380,00.html**.

[BERN1996] Philip A. Bernstein, The Repository. A Modern Vision, *Database Programming & Design*, Vol. 9, N° 12, Dec. 1996, págs. 28-35.

[BERR1999] Michael J. A. Berry, The Privacy Backlash, *Decision Support*, Vol. 2, N° 15, Oct. 26, 1999, URL: **http://www.intelligententerprise.com/db_area/archives/1999/992610/decision.shtml**.

[BERS1997] Alex Berson, Stephen J. Smith, *Data Warehousing, Data Mining & OLAP*, McGraw-Hill, USA, 1997.

[BIGU1996] Joseph P. Bigus, *Data Mining with Neural Networks*, McGraw-Hill, USA, 1996.

[BISC1994] Joyce Bischoff, Achieving Warehouse Success, *Database Programming & Design*, Vol. 7, N° 7, July 1994, págs. 27-33.

[BISC1996] Joyce Bischoff, Richard Yevich, Building More than a Data Warehouse, *Database Programming & Design*, Vol. 9, N° 9, Sept. 1996, págs. 27-31,49.

[BROO1997] Peter Brooks, March of the Data Marts, *DBMS online*, March 1997, URL: **http://www.dbmsmag.com/9703d14.html**.

[BROW1997] Bradley D. Brown, Why Warehouse on the Web?, *Oracle Magazine*, Vol. XI, N° 5, Sept.-Oct. 1997, págs. 97-106.

[BRUS1998] Joe Bruscato, Tim Vokes, Data Warehousing for Health at Anthem, *Teradatareview*, Vol. 1, N° 1, Spring 1998, págs. 42-45.

[BURL1997] Donald Burleson, *High Performance Oracle Data Warehousing*, Coriolis Group Books, USA, 1997.

[CARL1996] Brad Carlile, Seeking the Balance: Large SMP Warehouses, *Database Programming & Design*, Vol. 9, N° 8, Aug. 1996, págs. 44-48.

[CHAP1996] Philip Chapnick, Data Mining: AI Redux?, *Database Programming & Design*, Vol. 9, N° 9, Sept. 1996, pág. S5.

[CHES1998] Diane Cheshire, Mining for Allies, *Teradatareview*, Vol. 1, N° 1, Spring 1998, pág. 7.

[CRAI1997] Robert Craig, Data Warehousing and the Web, *Oracle Magazine*, Vol. XI, N° 5, Sept.-Oct. 1997, págs. 11-12.

[CURT1997] Mary Curtis, Kailash Joshi, Internal Control Issues for Data Warehousing, *IS Audit & Control Journal*, Vol. IV, 1997, págs. 40-45.

[DAVY1997] Mark M. Davydov, Bridging the Data Archipelago, *Database Programming & Design*, Vol. 10, N° 8, Aug. 1997, págs. 54-63.

[DE VIL2001] Larry de Ville, Data Mining in SQL Server 2000, *SQL Server Magazine*, Jan. 2001, URL: **http://www.sqlmag.com/Articles/ArticleID=16175**.

[DENN1982] Dorothy E. R. Denning, *Cryptography and Data Security*, Addison-Wesley, USA, 1982.

[DODG1998] Gary Dodge, Tim Gorman, *Oracle8 Data Warehousing*, John Wiley & Sons, USA, 1998.

[DURA1998] William Gerardo Durán Ortiz, Luis Enrique Chacón Abarca, *Almacenes de datos (Data Warehouses)*, Tesis, Licenciatura en Ingeniería de Sistemas Informáticos, Universidad Latina de Costa Rica, Costa Rica, Oct. 1998.

[EDEL1997] Herb Edelstein, Data Mining. Exploiting the Hidden Trends in Your Data, *DB2 Magazine*, Vol. 2, N° 1, Spring 1997, págs. 18-23.

[ENGL1996] Larry P. English, Help for Data Quality Problems – A Number of Automated Tools can Ease Data Cleansing and Help Improve Data Quality, *InformationWeek*, Issue: 600,Oct. 7, 1996.

[EVAN1997] Robert B. Evans, Driving Down Costs: A Case Study in Data Mining, *Database Programming & Design*, Vol. 10, N° 4, Apr. 1997, págs. 42-49.

[FLOH1997] Udo Flohr, OLAP by Web, *Byte*, Vol. 22, N° 9, Sept. 1997, págs. 81-84.

[FORD1994] Warwick Ford, *Computer Communications Security. Principles, Standard Protocols and Techniques*, PTR Prentice Hall, USA, 1994.

[FORI2000] Ronald Forino, Data e.Quality: The Data Quality Assessment, Part 1, *DM Review OnLine*, Aug. 2000.

[FREE20002] Robert G. Freeman, *Oracle9i New Features*, Oracle Press, McGraw-Hill/Osborne, USA, 2002.

[GAME1995] Marco V. Gámez Acuña, *Implementación de mecanismos criptográficos para la seguridad de la información utilizando firmas digitales*, Tesis, Licenciatura en Ingeniería de Sistemas, Universidad Internacional de las Américas, Costa Rica, Jul. 1995.

[GARD1997a] Dana Marie Gardner, Editor, Cashing in With Data Warehouses and the Web, *DataBased Advisor*, Vol. 15, N° 2, Feb. 1997, págs. 60-63.

[GARD1997b] Stephen R. Gardner, The Quest to Standarize Metadata, *Byte*, Vol. 22. N° 11, 1997, págs. 47-48.

[GEIG1997] Jonathan G. Geiger, What is Housed in the Data Warehouse that the Auditor Can Use?, *IS Audit & Control Journal*, Vol. IV, 1997, págs. 31-34.

[GIOV2000] William A. Giovinazzo, *Object-Oriented Data Warehousing Design Building a Star Schema*, Prentice Hall PTR, USA, 2000.

[GONÇ1997] Markus Gonçalves, Arthur Donkers, Jon-Paul Harkin, Anne Hart, Kitty Niles, Kathryn Toyer, Matthews Willis, *Internet Privacy Kit*, Que Corporation, USA, 1997.

[GONZ1996] Carlos A. González Alvarado, *Sistemas de Bases de Datos*, Editorial Tecnológica de Costa Rica, Costa Rica, 1996.

[GUDE1998] Dave Guderian, Doug Leer, Steve Molini, Keeping the Data Warehouse On Track, *Database Programming & Design*, Vol. 11, N° 1, Jan. 1998, págs. 39-47.

[HACK1997] Richard Hackathorn, Data Warehousing's Credibility Crisis, *Byte*, Vol. 22, N° 8, Aug. 1997, págs. 43-44.

[HILD1998] Eyk Hildebrandt, Gunter Saake, *User Authentication in Multidatabase Systems*, ITI, Otto-von-Guericke-Universität Magdeburg (Germany), Proceed. Ninth International Workshop on Database and Expert Systems Applications, Vienna, Austria, Aug. 26-28, 1998.

[HILL1997a] Robert L. Hill, The Right Stuff, *Database Programming & Design*, Vol. 10, N° 8, Aug. 1997, págs. 46-52.

[HILL1997b] Robert L. Hill, Purple Parallel Database Haze, *Database Programming & Design*, Vol. 10, N° 9, Sept. 1997, págs. 40-47.

[HUBE1996] Martin G. Hubel, Who Cares About Database Security?, *Database Programming & Design*, Vol. 9, N° 2, Feb. 1996, págs. 43-45.

[HUGH1995] Larry J. Hughes, Jr., *Actually Useful Internet Security Techniques*, New Riders Publishing, USA, 1995.

[HURW1997] Mike Hurwicz, Take Your Data to the Cleaners, *Byte*, Vol. 22, N° 1, Jan. 1997, págs. 97-102.

[INMO1997c] William H. Inmon, Tech Topic 15. Data Warehouse and Internet Security, *Pine Cone Systems, Inc.*, 1997, URL: **http://www.billinmon.com/library/whiteprs/techtopic/tt15.pdf**.

[INMO2000] William H. Inmon, Data Warehouse and Security, *Billinmon.com*, 2000, URL: **http://www.billinmon.com/library/whiteprs/earlywp/ttsecur.pdf**.

[JIME1998a] Beatriz Jiménez Sobrino, Rafael Ávalos Cardoso, *Depósitos y minería de datos: Una visión estratégica y metodológica*, Tesis, Magister Scientiae en Computación, Instituto Tecnológico de Costa Rica, Costa Rica, Ago. 1998.

[JUDA2000] Jeffrey Juday, Readers' Tips and Tricks, *SQL Server Magazine*, Dec. 2000, URL: **http://www.sqlmag.com/Articles/ArticleID=15906**.

[KATIC1998] N. Katic, G. Quirchmayr, J. Schiefer, M. Stolba, A M. Tjoa, *A Prototype Model for Data Warehouse Security Based on Metadata*, Institute of Software Technology, Vienna University of Technology; Institute of Applied Computer Science and Information Systems, University of Vienna, Austria, DEXA Workshop 1998 Vienna, Austria, 1998, URL: **http://www.computer.org/conferen/dexa/8353/pdf/83530300.pdf**.

[KAY1997] Emily Kay, Interested In... ..DataWarehouse?. Security, *Data Warehousing Management*, July 1997, URL: **http://www.sentrytech.com/dw07sec.htm**.

[KEST1997a] Justin Kestelyn, Visualizing the Data Warehouse, *DB2 Magazine*, Vol. 2, N° 1, Spring 1997, págs. 44-45.

[KEST1997b] Justin Kestelyn, Extracting Fact from Fiction. Does Data Mining Fit in Your Enterprise?, *DB2 Magazine*, Vol. 2, N° 1, Spring 1997, págs. 26-29,47.

[KEST1999] Justin Kestelyn, What's Fair is Fair, *Intelligent Enterprise Magazine*, Vol. 2, N° 18, Dec. 21, 1999, URL: **http://www.intelligententerprise.com/db_area/archives/1999/992112/editpage.shtml**.

[KIMB1997] Ralph Kimball, Hackers, Crackers, and Spooks, *DBMS online*, Apr. 1997, URL: **http://www.dbmsmag.com/9704d05.html**.

[KIMB1999] Ralph Kimball, Remove Security from Your Database Tables, *Intelligent Enterprise Magazine*, Vol. 2, N° 14, Oct. 5, 1999, URL: **http://www.intelligententerprise.com/db_area/archives/1999/990510/warehouse.shtml**.

[KIMB2001a] Ralph Kimball, Adjust Your Thinking for SANs, *Intelligent Enterprise*, 8 Mar. 2001, URL: **http://www.intelligententerprise.com/010308/webhouse_1.shtml**.

[KOSA1997] Denis Kosar, A Proactive Approach to Data Warehousing, *DB2 Magazine*, Vol. 2, N° 1, Spring 1997, págs. 10-11.

[KRIV1996] Cheryl D. Krivda, Unearthing Underground Data, *LAN The Network Solutions Magazine*, Vol. 11, N° 5, May 1996, págs. 42-48.

[LAMB1997] Nevin Lambert, Manish Patel, *PCWEEK Windows NT Security. System Administrator's Guide*, Ziff-Davis Press, USA, 1997.

[LAMB1996] Bob Lambert, Data Warehousing Fundamentals: What you Need to Know to Successed: Special Feature from March 1996, *DM Review*, March 1996, URL: **http://www.dmreview.com/master.cfm?NavId=55&EdId=1313**.

[LAZA1996] Bill Lazar, The Data Breakthrough, *LAN The Network Solutions Magazine*, Vol. 11, N° 5, May 1996, págs. 50-55.

[LILL1997] R. S. Stephen Lilly, Client/Server Database Security Starts with Strong Passwords and Robust Auditing, *IS Audit & Control Journal*, Vol. VI, 1997, págs. 50-51.

[LINS1999] Mark A. Linsenbardt, M. Shane Stigler, *SQL Server 7 Administration*, Osborne/McGraw-Hill, USA, 1999.

[MADS1997] Mark Madsen, Warehousing Meets the Web, *Database Programming & Design*, Vol. 10, N° 8, Aug. 1997, págs. 37-45.

[MALE2000] Jonathan I. Maletic, Andrian Marcus, *Data Cleansing: Beyond Integrity Analysis*, Division of Computer Science, The Department of Mathematical Sciences, The University of Memphis, USA, Jun. 23, 2000.

[MARC2000a] Andrian Marcus, Jonathan I. Maletic, *Utilizing Association Rules for Identification of Possible Errors in Data Sets*, Technical Report CS-00-03, Division of Computer Science, The Department of Mathematical Sciences, The University of Memphis, USA, Feb. 28, 2000.

[MARC2000b] Andrian Marcus, Jonathan I. Maletic, *Utilizing Association Rules for Identification of Errors in Data*, Technical Report CS-00-04, Division of Computer Science, The Department of Mathematical Sciences, The University of Memphis, USA, May 8, 2000.

[MART1997] Fernando Martínez-Campos, Bigger Than a Database, *Database Programming & Design*, Vol. 10, N° 4, Apr. 1997, págs. 27-35.

[MASS1997] Paul Massiglia, The Dawn of Enterprise Storage, *Database Programming & Design*, Vol. 10, N° 6, June 1997, págs. 43-49.

[MATT1996] Rob Mattison, *Data Warehousing. Strategies, Technologies, and Techniques*, McGraw-Hill, USA, 1996.

[MCCL1996] David McClanahan, Making Sense of Enterprise Data, *DataBased Advisor*, Vol. 14, N° 11, Nov. 1996, págs. 76-79.

[MCKI1996] Stewart McKie, Data Warehouse 101, *Controller Magazine*, Oct. 1996, URL: **http://www.controllermag.com/issues/Oct96/Dataware.html**.

[MENA1997] Jesús Mena, Data Mining from the Ground Up, *Database Programming & Design*, Vol. 10, N° 2, Feb. 1997, págs. 63-67.

[MENN1997] Dave Menninger, Editor, Optimizing Your Data Warehouse for the Web, *DataBased Advisor*, Vol. 15, N° 2, Feb. 1997, págs. 40-47.

[MICR2001] Microsoft Corporation, *Administración de Riesgos*, 17 Ago. 2001, URL: **http://www.microsoft.com/latam/technet/admon/estrategia/art10/.**

[MILE1997a] Michael Miley, Bringing the World to Your Warehouse, *Oracle Magazine*, Vol. XI, N° 5, Sept.-Oct. 1997, págs. 51-75.

[MILE1997b] Michael Miley, First-Class Delivery, *Oracle Magazine*, Vol. XI, N° 5, Sept.-Oct. 1997, págs. 76-84.

[MOLI2000] Lucio A. Molina Focazzio, *Utilización de tecnología "Data Warehouse/Mining" para la detección de fraudes*, Sesión 322, 5ª Conferencia Anual Latin CACS, San José, Costa Rica, 16-19 oct. 2000.

[MORA2000] Brian Moran, Data Warehousing Step by Step, *SQL Server Magazine*, Feb. 2000, URL: **http://www.sqlmag.com/Articles/ArticleID=7832**.

[MORI1995a] Terry Moriarty, Modeling Data Warehouses, *Database Programming & Design*, Vol. 8, N° 8, Aug. 1995, págs. 61-65.

[MORI1995b] Terry Moriarty, Part of the Whole, *Database Programming & Design*, Vol. 8, N° 10, Oct. 1995, págs. 61-64.

[MORI1996a] Terry Moriarty, Richard P. Greenwood, Data's Quest—From Source to Query, *Database Programming & Design*, Vol. 9, N° 10, Oct. 1996, págs. 78-81.

[MORI1996c] Terry Moriarty, Art Moore, Data Warehouse Must-Reads, *Database Programming & Design*, Vol. 9, N° 12, Dec. 1996, págs. 61-63.

[MORI1997] Terry Moriarty, A Metadata Management How-To, *Database Programming & Design*, Vol. 10, N° 2, Feb. 1997, págs. 57-60.

[MOXO1996] Bruce Moxon, Defining Data Mining, *DBMS online*, Aug. 1996, URL: **http://www.dbmsmag.com/9608d53.html**.

[NEWQ1996] H. P. Newquist, Data Mining. The AI Metamorphosis, *Database Programming & Design*, Vol. 9, N° 9, Sept. 1996, págs. S12-S14.

[NORM1996] Mike Norman, The Broker of Networked Services, *Database Programming & Design*, Vol. 9, N° 13, Dec. 1996, págs. 31-37.

[OLSO1995] Jack Olson, Building a Database Topology Strategy, *Database Programming & Design*, Vol. 8, N° 6, June 1995, págs. 52-61.

[PARS1995] Kamran Parsaye, The Sandwich Paradigm, *Database Programming & Design*, Vol. 8, N° 4, Apr. 1995, págs. 50-55.

[PARS1996] Kamran Parsaye, Data Mines for Data Warehouses, *Database Programming & Design*, Vol. 9, N° 9, Sept. 1996, págs. S6-S11.

[PARS1997] Kamran Parsaye, OLAP & Data Mining. Bridging the Gap, *Database Programming & Design*, Vol. 10, N° 2, Feb. 1997, págs. 31-37.

[PEND1998] Nigel Pendse, OLAP Omnipresent, *Byte*, Vol. 23, N° 2, Feb. 1998, págs. 88NA1-88NA8.

[QUIN1996] Tim Quinlan, Report from the Trenches, *Database Programming & Design*, Vol. 9, N° 12, Dec. 1996, págs. 36-45.

[RAHM1997] Dan Rahmel, Database Security, *Internet Systems*, Apr. 1997, URL: **http://www.dbmsmag.com/9704i03.html**.

[REES1997] Joseph Reese, Making the Data Warehouse Work Overtime, *Database Programming & Design*, Vol. 10, N° 9, Sept. 1997, págs. S10-S15.

[REIN1998] David Reiner, Sold on Database Marketing, *Teradatareview*, Vol. 1, N° 1, Spring 1998, págs. 27-33,40.

[RENN1997] Martin Rennhackkamp, Database Security, *DBMS*, Feb. 1997, URL: **http://www.dbmsmag.com/9702d00.html**.

[RODR1995] Luis Ángel Rodríguez, *Seguridad de la información en sistemas de cómputo*, Ventura Ediciones, México D. F., 1995.

[ROSE1998] Michele Rosen, Internet Security Standards, *PC Magazine*, Vol. 17, N° 2, Jan. 20, 1998, 241-242.

[ROSEN2000] Arnon Rosenthal, Edward Sciore, View Security as the Basic for Data Warehouse Security, *Proceedings for the International Workshop on Design and Management of Data Warehouses*, Estocolmo, Suecia, Jun. 5-6, 2000, URL: **http://www.mitre.org/support/papers/tech_papers99_00/rosenthal_security.pdf**.

[SCHN1995a] Bruce Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, 2nd Ed., John Wiley & Sons, USA, 1994.

[SCHN1995b] Bruce Schneier, The GOST Encryption Algorithm, *Dr. Dobb's Journal*. N° 226, Jan. 1995, págs. 123-124, 143-144.

[SCOT2000a] Mark D. Scott, David Walls, Step-by-Step to Data Warehousing, *SQL Server Magazine*, Jan. 26,. 2000, URL: **http://www.sqlmag.com/Articles/ArticleID=8047**.

[SCOT2000b] Mark D. Scott, David Walls, 7 Steps to Data Warehousing, *SQL Server Magazine*, Feb. 2000, URL: **http://www.sqlmag.com/Articles/ArticleID=7834**.

[SHER1997] Richard Sherman, Intranet Data Access: The Fire Inside the Firewall, *Database Programming & Design*, Vol. 10, N° 4, Apr. 1997, págs. 36-41.

[SILB1999] Chris Silbernagel, Data Security: Protecting the Warehouse from Within, *DM Review*, Jun. 1999, URL: **http://www.dmreview.com/master.cfm?NavId=198&EdId=780**.

[SILVER] Len Silverston, Harry Smith, Mind if I Have a Look at Your Sensitive Warehouse Data?: Data Warehouse Security, *Decision Support & Data Warehousing*, Paper # 42, págs. 1-4, URL: **http://www.yale.edu/adminsys/datawhse/IDECI042.PDF**.

[SIMO1995] Alan Simon, I Want a Data Warehouse So, What Is It Again?, *Database Programming & Design*, Vol. 8, N° 12, Dec. 1995, págs. 26-31.

[SIMO1996] Alan Simon, Beyond the Warehouse, *Database Programming & Design*, Vol. 9, N° 13, Dec. 1996, págs. 43-46.

[SIMO1997a] Alan Simon, Better Clients, Better Decisions, *Byte*, Vol. 22, N° 1, Jan. 1997, págs. 91-94.

[SIMO1998a] Alan Simon, *90 Days to the Data Mart*, John Wiley & Sons, USA, 1998.

[SIMO1998b] Alan Simon, What Are Data Marts?, *Database Programming & Design*, Vol. 11, N° 2, Feb. 1998, págs. 23-25.

[STOD1997] David B. Stodder, Ten Years Alive in the Tenth Anniversary Database Industry, *Database Programming & Design*, Vol. 10, N° 3, March 1997, págs. 22-27,61.

[TANL1997] Rick Tanler, *The Intranet Data Warehouse*, John Wiley & Sons, USA, 1997.

[TAYL1997] Art Taylor, Java-Charging the Data Warehouse, *Database Programming & Design*, Vol. 10, N° 10, Oct. 1997, págs. 58-65.

[THOM1997a] Erik Thomsen, Dimensional Modeling: An Analytical Approach, *Database Programming & Design*, Vol. 10, N° 3, March 1997, págs. 29-35.

[THOM1997b] Erik Thomsen, Music of the Cubes, *Database Programming & Design*, Vol. 10, N° 4, Apr. 1997, págs. 66-69.

[THOM1997c] Erik Thomsen, Beyond the MOLAP/ROLAP Wars, *Database Programming & Design*, Vol. 10, N° 8, Aug. 1997, págs. 78-79.

[THOM1997d] Erik Thomsen, Mining Your Way to OLAP, *Database Programming & Design*, Vol. 10, N° 9, Sept. 1997, págs. 101-103.

[THOM1997e] Erik Thomsen, *OLAP Solutions. Building Multidimensional Information Systems*, John Wiley & Sons, USA, 1997.

[UDEL1998] Jon Udell, In Search of SSL Spidering, *Byte*, Vol. 23, N° 2, Feb. 1998, págs. 97-100.

[WATT1997] Karen Watterson, Attention, Data Mart Shoppers, *Byte*, Vol. 22, N° 7, July 1997, págs. 73-78.

[WELD1995] Jay-Louise Weldon, Managing Multimensional Data: Harnessing the Power, *Database Programming & Design*, Vol. 8, N° 8, Aug. 1995, págs. 24-33.

[WELD1997a] Jay-Louise Weldon, Alan Joch, Data Warehouse Building Blocks, *Byte*,Vol. 22, N° 1, Jan. 1997, págs. 82-83.

[WELD1997b] Jay-Louise Weldon, Warehouse Cornerstones, *Byte*, Vol. 22, N° 1, Jan. 1997, págs. 85-88.

[WENI1997] Joseph E. Wening, Gregory J. Sannik, 5 Critical Factors to Successful Data Warehouses, *DataBased Advisor*, Vol. 15, N° 1, Jan. 1997, págs. 74-77.

[WHIT1995] Colin White, The Key to a Data Warehouse, *Database Programming & Design*, Vol. 8, N° 2, Feb. 1995, págs. 23-25.

[WILL1997] Joseph Williams, Tools for Traveling Data, *DBMS online*, June 1997, URL: **http://www.dbmsmag.com/9706d16.html**.

[WINT1997a] Richard Winter, A Decade of VLDBs, *Database Programming & Design*, Vol. 10, N° 3, March 1997, págs. 19-21.

[WINT1997b] Richard Winter, Act, Don´t React, *Database Programming & Design*, Vol. 10, N° 4, Apr. 1997, págs. 21-23.

[WINT1997c] Richard Winter, Kathy Auerbach, Giants Walk the Earth, *Database Programming & Design*, Vol. 10, N° 9, Sept. 1997, págs. S2-S16.

[WINT1998] Richard Winter, Introducing the Data Warehouse Challenge, *Database Programming & Design*, Vol. 11, N° 2, Feb. 1998, págs. 19-21.

[ZIBI1998] Boris Zibitsker, Gdaliy Sigalov, Building the Perfect Data Warehouse, *Teradatareview*, Vol. 1, N° 1, Spring 1998, págs. 34-40.

[ZORN1997] Aaron Zornes, Serving Up a Web-Enabled Warehouse, *Oracle Magazine*, Vol. XI, N° 5, Sept.-Oct. 1997, pág. 160.

Anexos

Anexo 1

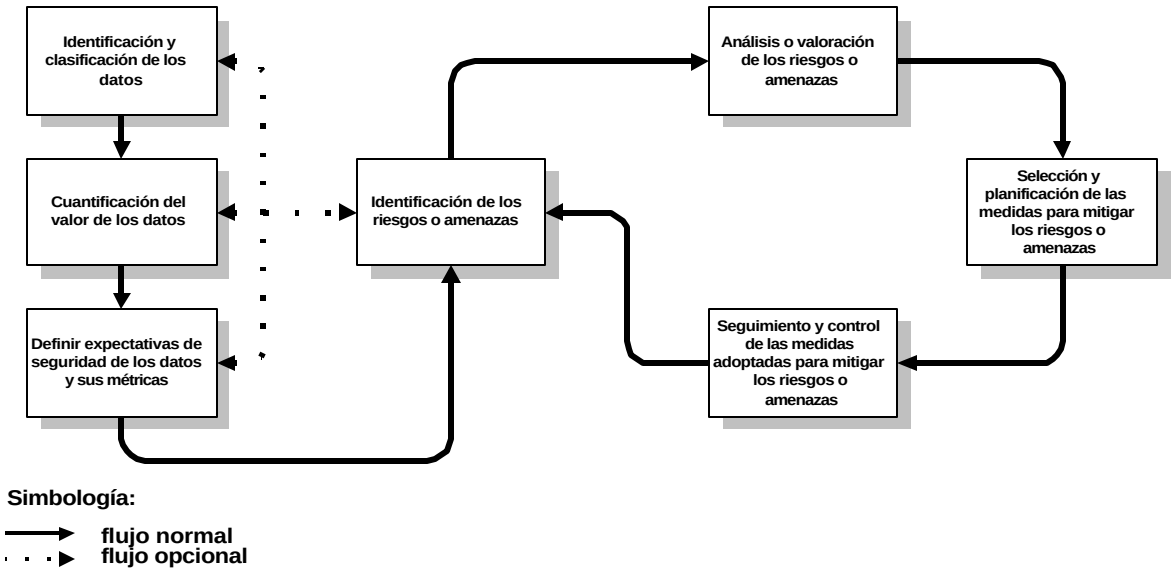
Cuestionario para valorar la propuesta para el mejoramiento de la seguridad de un depósito de datos

Instituto Tecnológico de Costa Rica
Departamento de Computación
Programa de Maestría en Computación

Cuestionario para valorar la propuesta para el mejoramiento de la seguridad de un depósito de datos.

El presente cuestionario se ha preparado como parte del proyecto de graduación del Programa de Maestría en Computación del Instituto Tecnológico de Costa Rica.

El propósito del cuestionario es valorar cuán adecuada resulta la propuesta para el mejoramiento de la seguridad en los depósitos de datos (data warehouses) que se sugiere en esta investigación.



La información que suministre será tratada con la más estricta confidencialidad del caso, y su divulgación no se hará sino por medio de cuadros estadísticos condensados, lo cual se permite identificar de manera particular a ninguna de las empresas participantes en el estudio.

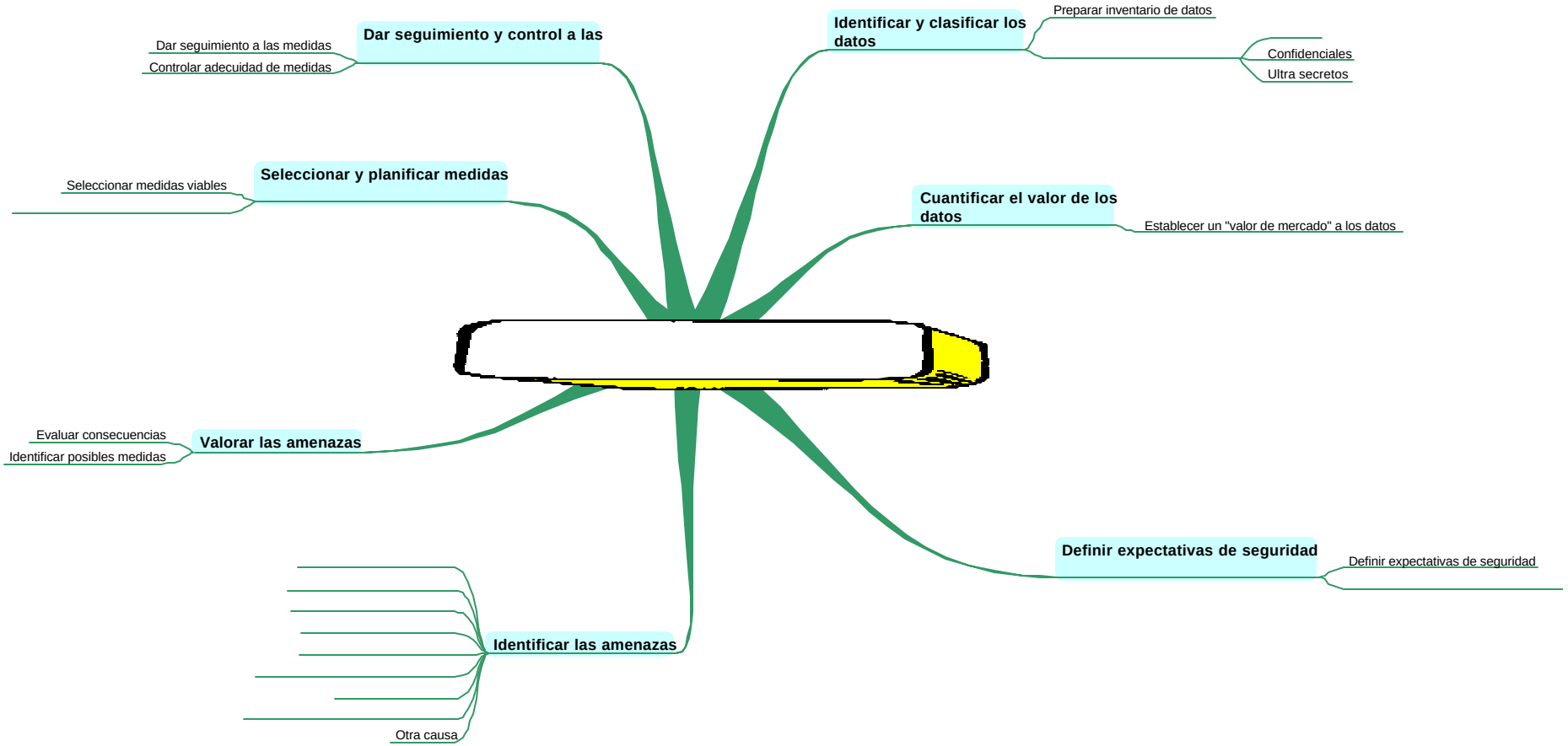
De igual forma, dado que la anterior figura corresponde a la propuesta de esta investigación, la cual aún no se ha publicado, se le solicita muy respetuosamente no divulgarla o utilizarla hasta tanto no culmine el proceso investigativo. En todo caso, si desea mayor información al respecto favor comunicarse con el autor.

Finalmente, aprovecho la oportunidad para agradecerle de antemano el tiempo y la atención que le sirva prestar a esta encuesta.

Muy atentamente,

Lic. Marco V. Gámez Acuña
mgameza@yahoo.com

Dr. Carlos A. González A.
Director de la Investigación


Dar seguimiento y control a las
Dar seguimiento a las medidas
Controlar adecuidad de medidas
Identificar y clasificar los datos
Preparar inventario de datos
Confidenciales
Ultra secretos
Seleccionar y planificar medidas
Seleccionar medidas viables
Cuantificar el valor de los datos
Establecer un "valor de mercado" a los datos
Valorar las amenazas
Evaluar consecuencias
Identificar posibles medidas
Definir expectativas de seguridad
Definir expectativas de seguridad
Identificar las amenazas
Otra causa

Indicaciones generales:
A continuación se le presenta una serie un conjunto de 31 preguntas, las cuales en su mayoría son de selección única. Por esta razón, mucho le agradeceríamos marcar con una equis (**X**) aquélla que mejor corresponda a la situación planteada.

A. De la persona que suministra los datos.

1. ¿En que área se ubica Usted en relación con el depósito de datos?
- ❑ Área técnica (informática)
- ❑ Área usuaria
- ❑ Empresa desarrolladora

B. De la descripción general de la empresa.

2. Indique cuál es la actividad principal de la empresa
- ❑ Sector financiero
- ❑ Sector público
- ❑ Sector comercial
- ❑ Sector industrial
- ❑ Otro

3. ¿Desde hace cuánto fue fundada u opera la empresa?
- ❑ Menos de un año
- ❑ De 1 a menos de 5 años
- ❑ De 5 a menos de 10 años
- ❑ Más de 10 años

C. De la descripción de la solución de depósito de datos

4. ¿Cuántas áreas de la empresa cubre la solución desarrollada o en proceso de desarrollo?
- ❑ Un departamento o área
- ❑ Dos o más áreas (pero no toda la empresa)
- ❑ Toda la empresa

5. Para el desarrollo de la solución de depósito de datos ¿qué estrategia de desarrollo se ha utilizado o se está utilizando?
- ❑ Sólo recurso humano interno
- ❑ Sólo recurso humano externo
- ❑ Tanto recurso humano interno como externo
- ❑ Se contrató un consultor externo para que asesorara al personal interno

6. La solución de depósito de datos se concibió originalmente como:
- ❑ Un *data warehouse*
- ❑ Un *data mart*
- ❑ Un *OLAP*
- ❑ Otro

7. ¿Hace cuánto se desarrolló la solución de depósito de datos?
- ❑ Menos de tres meses
- ❑ De tres a menos de seis meses
- ❑ De seis meses a menos de 1 año
- ❑ Más de un año
- ❑ Está en proceso de desarrollo

8. ¿Hace cuánto inició el proyecto de desarrollo?
- ❑ Hace seis meses
- ❑ De 6 meses a menos de un año
- ❑ De uno a tres años
- ❑ Más de tres años
- ❑ Está por iniciar

9. ¿En qué plataforma corre o va a correr la solución de depósitos de datos?
- ❑ Oracle
- ❑ Microsoft SQL Server
- ❑ Otra, especifique: _____

10. ¿En qué rango de tamaño ubica Usted al repositorio de datos?
- ❑ Menos de 200 Gigabytes
- ❑ De 200 a 500 Gigabytes
- ❑ De 500 Gigabytes a menos de 1 Terabytes
- ❑ Más de un Terabytes

11. ¿Cuántas personas están dedicadas a la administración y mantenimiento del depósito de datos?
- ❑ Menos de 5 personas
- ❑ De 5 a menos de 15 personas
- ❑ 15 ó más personas

12. ¿A qué sector de usuarios está orientada principalmente la solución de depósito de datos?
- ❑ Al nivel estratégico
- ❑ Al nivel táctico
- ❑ Al nivel operativo
- ❑ A todo nivel

13. ¿Sobre quién recae la responsabilidad de la administración de la solución de depósitos de datos que se desarrolló o va a desarrollar?
- ❑ Un administrador (del tipo DBA)
- ❑ Un grupo de personas (grupo OLTP)
- ❑ Otro, especifique: _____

14. ¿A quién le corresponde ejercer la labor de monitoreo de la solución de depósito de datos?
- ❑ Un funcionario del área informática
- ❑ Un funcionario del área usuaria
- ❑ Un funcionario del nivel estratégico de la empresa

D. De la metodología para el desarrollo de la solución de depósito de datos

15. Para el desarrollo de este tipo de solución ¿dispuso o dispone la empresa de una metodología específica?
- ❑ No
- ❑ Sí. Favor indique cuál:

_____

_____

_____

16. En su criterio, ¿la metodología utilizada para desarrollar la solución de depósito de datos le da o dio un tratamiento adecuado al tema de la seguridad?
- ❑ Sí
- ❑ No. Por favor explique brevemente en qué aspectos considera que la metodología no trata adecuadamente el tema de seguridad.

_____

_____

_____

17. Independientemente de lo que establece la metodología utilizada, al desarrollar la solución de depósito de datos ¿se tuvo alguna preocupación en cuanto al tema de la seguridad?
- ❑ No
- ❑ Sí. Por favor explique brevemente cuáles fueron esas preocupaciones y de qué forma se solventaron.

_____

_____

_____

E. Validación de la propuesta

18. En su criterio, a la hora de desarrollar una solución de depósito de datos ¿resulta realmente necesario hacer una identificación de todos los datos que se van a considerar como parte de la solución a desarrollar?

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

19. ¿Cree Usted que una vez identificados los datos que va a contener el depósito, realizar una clasificación de éstos según su grado de confidencialidad o privacidad, de manera que se establezca para cada uno de ellos una categoría (que puede ser público, confidencial o muy confidencial), va a resultar inútil en el proceso de definir la seguridad del depósito?

❑ No ❑ Sí ¿Por qué?

_____

_____

_____

20. Como parte del proceso de desarrollo de la solución, ¿considera Usted que intentar cuantificar el valor de los datos resulta infructuoso, es decir, que no es significativo el aporte a la hora de definir la seguridad del depósito?

❑ No ❑ Sí ¿Por qué?

_____

_____

_____

21. En la metodología que se propone en la presente investigación se sugiere que se deben definir expectativas de seguridad, las cuales deben ser consideradas a la hora de desarrollar la solución de depósito de datos y sus respectivas métricas, que permitan valorar en qué grado se alcanzan éstas. En su criterio, ¿realizar esta actividad resulta despreciable y de poca utilidad para la definición de un nivel adecuado de seguridad para la solución a desarrollar?

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

22. Las amenazas que atentan contra la solución de depósitos de datos que se va a desarrollar constituyen un tema que debe ser considerado a la hora de desarrollar los depósitos. ¿Considera Usted que realizar una actividad tendente a identificar tales amenazas resulta superflua o de poca importancia al definir una seguridad adecuada para la solución?

❑ No ❑ Sí ¿Por qué?

_____

_____

_____

23. Partiendo del hecho de que las amenazas que pueden atentar contra los depósitos de datos fueron identificadas de alguna forma, sea por la realización de la actividad anteriormente mencionada, o por otro medio, ¿cree Usted que analizar tales amenazas puede generar un valor agregado tal que justifique su realización?.

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

24. Para mitigar las amenazas que se detectaron y analizaron previamente, resulta lógico pensar que se deben establecer algunas medidas que las minimicen. De acuerdo con su perspectiva de la situación, y basado en la experiencia que Usted ha adquirido acerca del tema, ¿considera Usted que seleccionar y planificar las medidas que resulten más convenientes para la organización, circunscritas en las posibilidades económicas y tecnológicas de la organización resulta ser el camino más apropiado para llevar a su mínima expresión los riesgos involucrados?

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

25. Teniendo en cuenta que las medidas que se puedan haber adoptado para mitigar las amenazas están afectadas a no ser las mejores, ya sea por limitaciones propias de la organización, falta de conocimiento suficiente acerca de las amenazas, o bien, por cambios abruptos en el entorno, ¿cree Usted que dar seguimiento y controlar las medidas que se adoptaron puede resultar una actividad inútil o estéril en procura de dotar de la mejorar seguridad posible al depósito de datos?

❑ No ❑ Sí ¿Por qué?

_____

_____

_____

26. Puesto que el entorno de los depósitos de datos es por sí mismo continuamente cambiante, al punto de que los cambios pueden ser incluso abruptos e impredecibles, ¿considera Usted que a pesar de que en un inicio se pudo haber hecho una primera identificación de amenazas, tal vez resulte valioso y útil realizar una nueva revisión de los primeros elementos mencionados en la metodología propuesta en esta investigación, de manera tal que se reconsidere la identificación de los datos, la clasificación de estos según su grado de confidencialidad, la valoración de los datos y las expectativas de seguridad de éstos y sus respectivas métricas, y de ser el caso, que se actualicen tales productos?

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

27. Entre las posibles amenazas que pueden atentar contra el funcionamiento de un depósito de datos se tienen aquéllas que consideran a los activos físicos que lo componen, como por

ejemplo, el robo, la destrucción accidental de ellos o su secuestro. En virtud de ello, ¿considera Usted que estos aspectos deben formar parte de las acciones a establecer en la mitigación de las amenazas, por cuanto estos aspectos caben dentro de lo que sería un plan de contingencias para el área de procesamiento de datos?

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

28. ¿Cree Usted que por lo estratégica que resulta la información contenida en un depósito de datos, se deba considerar un uso sumamente restrictivo de ella?

❑ No ❑ Sí ¿Por qué?

_____

_____

_____

29. En su opinión, ¿para ofrecer un uso seguro de la información contenida en el depósito de datos se debe acudir a la criptografía para cifrar toda esa información?

❑ No ❑ Sí ¿Por qué?

_____

_____

_____

30. ¿Acaso un uso seguro y controlado de la información contenida en el depósito de datos se puede dar utilizando para ello una buena definición de permisos de acceso, los cuales se definan a partir de una clara identificación de las necesidades de información que tienen los diferentes usuarios del depósito?

❑ Sí ❑ No ¿Por qué?

_____

_____

_____

31. ¿Tiene Usted algún comentario en relación con la forma como en la empresa en que labora se trata el tema de la seguridad en el depósito de datos?

❑ No ❑ Sí Por favor desarrolle sus apreciaciones al respecto.

_____

_____

_____

_____

_____

Muchas gracias por su colaboración.

Anexo 2

Lista de empresas que participaron respondiendo el cuestionario

A. Empresas desarrolladoras de soluciones del tipo "*data warehouse*"

- Grupo Asesor
- Grupo MAS

B. Empresas o instituciones que cuentan (o están próximas a contar) con una solución del tipo "*data warehouse*"

- Superintendencia General de Entidades Financieras (SUGEF)
- Superintendencia General de Pensiones (SUPEN)
- Superintendencia General de Valores (SUGEVAL)
- Banco de Costa Rica
- Banco Nacional de Costa Rica
- Banco Popular y de Desarrollo Comunal
- Popular Valores
- Bolsa Nacional de Valores (BNV)
- Instituto Costarricense de Electricidad – Área de Telecomunicaciones (ICETEL)
- La Nación

Anexo 3

Programa SPSS para procesar la encuesta

```
SET DISK=ON EJECT=OFF LOG OFF MORE=off ECHO=OFF
    listing='tesis.lis' box '-|+' length=60.
TITLE 'Seguridad en depósitos de datos'.

TRANSLATE From 'tesis.dbf'.

VAR LABELS
    VAR01 'Area a que pertenece quien responde'/
    VAR02 'Actividad principal de la empresa'/
    VAR03 'Tiempo de fundada la empresa'/
    VAR04 'Areas que cubre data warehouse'/
    VAR05 'Estrategia de desarrollo utilizada'/
    VAR06 'Solución originalmente concibida como:'/
    VAR07 'Tiempo desde que se desarrolló solución'/
    VAR08 'Tiempo desde que inició el proyecto'/
    VAR09 'Plataforma en que corre la solución'/
    VAR10 'Tamaño del repositorio de datos'/
    VAR11 'Número personas en administr. y manten.'/
    VAR12 'Sector al que se dedica la solución'/
    VAR13 'Responsable de administrar la solución'/
    VAR14 'Responsable de monitorear la solución'/
    VAR15A 'Dispuso metodol. específ. desarrollo?'/
    VAR15B 'Metodologías disponibles'/
    VAR16A 'Se trató adecuadamente tema seguridad?'/
    VAR16B 'Aspectos de seguridad no tratados'/
    VAR17A 'Hubo preocupación temas seguridad?'/
    VAR17B 'Preocupaciones y forma solucionarlas'/
    VAR18A 'Resulta necesario identificar datos?'/
    VAR18B 'Razones por las cuales no es necesario'/
    VAR19A 'Resulta inútil clasificar confidenc.?'/
    VAR19B 'Razones por las cuales resulta inútil'/
    VAR20A 'Infructuoso cuantificar valor datos?'/
    VAR20B 'Razones porque es infructuoso'/
    VAR21A 'Poco util definir expectativas seguridad?'/
    VAR21B 'Razones de poca utilidad'/
    VAR22A 'Es superflua identificación de amenazas?'/
    VAR22B 'Razones por las cuales resulta superflua'/
    VAR23A 'Da valor agregado analizar las amenazas?'/
    VAR23B 'Razones porque no genera valor agregado'/
    VAR24A 'Es mejor seleccionar y planif. medidas?'/
    VAR24B 'Razones porque no es mejor seleccionar'/
    VAR25A 'Inútil dar seguim. y controlar medidas?'/
    VAR25B 'Razones por las cuales resulta inútil'/
    VAR26A 'Es valioso reejecutar etapas iniciales?'/
    VAR26B 'Razones porque no resulta valioso'/
    VAR27A 'Considerar robo, secuestro, destrucción?'/
    VAR27B 'Razones no considerarlas tales medidas'/
    VAR28A 'Considerar uso restrictivo información?'/
    VAR28B 'Razones para considerar uso restrictivo'/
    VAR29A 'Se debe acudir a la criptografía?'/
    VAR29B 'Razones para considerar uso criptografía'/
    VAR30A 'Se da buena definición permisos acceso?'/
    VAR30B 'Razones para buena definición permisos'/
    VAR31A 'Tiene comentarios acerca tema seguridad?'/
    VAR31B 'Comentarios acerca tema de seguridad'.
MISSING VALUES
    .
VALUE LABELS
    VAR01 1 'Area técnica'    2 'Area usuaria'   3 'Empresa desarr.'/
    VAR02 1 'Sector Financ.'  2  'Sector públ.'  3 'Sector comerc.'
        4 'Sector industr.' 5 'Otro'/
    VAR03 1 'Menos 1 año'     2 'De 1 a -5 años' 3 'De 5 a -10 años'
        4 'Más de 10 años'/
    VAR04 1 'Una área'        2 'Dos o más áreas' 3 'Toda la empresa'/
```

```
    VAR05 1 'Rec. interno'    2 'Rec. externo'   3 'Ambos'
        4 'Consultor'/
    VAR06 1 'Data warehouse'  2 'Data mart'      3 'OLAP'
        4 'Otro'/
    VAR07 1 'Menos 3 meses'   2 'De 3 a -6 meses' 3 'De 6m a -1 año'
        4 'Más de un año'   5 'En desarrollo'/
    VAR08 1 'Hace 6 meses'    2 'De 6m a -1 año'  3 'De 1 a 3 años'
        4 'Más de 3 años'   5 'Por iniciar'/
    VAR09 1 'Oracle'          2 'MS SQL Server'   3 'Sybase'
        4 'Otro'/
    VAR10 1 'Menos 200 GB'    2 'De 200 a 500 GB' 3 'De 500 GB a 1 TB'
        4 'Más de 1 TB'/
    VAR11 1 'Menos de 5'      2 'De 5 a -15'       3 '15 o más'/
    VAR12 1 'Estratégico'     2 'Táctico'         3 'Operativo'
          4 'Estratégico-táctico' 5 'Estratégico-operativo'
          6 'Todo nivel'/
    VAR13 1 'DBA'             2 'Grupo OLTP'      3 'Otro'/
    VAR14 1 'Area informát.'  2 'Area usuaria'    3 'Nivel estratégico'
          4 'Informát.-usuaria'/
    VAR16A,VAR18A,VAR21A,VAR23A,VAR24A,VAR26A,VAR27A,VAR30A
        1 'Si'              2 'No'/
    VAR15A,VAR17A,VAR19A,VAR20A,VAR22A,VAR25A,VAR28A,VAR29A,VAR31A
        1 'No'              2 'Si'.

FREQUENCIES  VARIABLES=VAR01 to VAR14,VAR15A,VAR16A,VAR17A,VAR18A,VAR19A,
             VAR20A,VAR21A,VAR22A,VAR23A,VAR24A,VAR25A,VAR26A,VAR27A,VAR28A,
             VAR29A,VAR30A,VAR31A.

CROSSTABS VAR18A,VAR19A,VAR20A,VAR21A,VAR22A,VAR24A,VAR26A,VAR28A,VAR30A by
          VAR15A,VAR16A,VAR17A/
  CELLS= /format=nobox.

FINISH.
```