

**Instituto Tecnológico de Costa Rica
Carrera de Ingeniería en Computación
Carrera Gestión del Turismo Rural Sostenible
Campus Tecnológico Local San Carlos**

Proyecto:

*GEOTOROTUR: Geomática utilizando ToroDB aplicada al
turismo*

Informe Final

Investigadores:

Marlen Treviño Villalobos, Leonardo Víquez Acuña, Rocío
Quirós Oviedo, Lady Fernández Mora, Oscar Víquez Acuña,
Gaudy Esquivel Vega

Marzo, 2020

TABLA DE CONTENIDO

I.	Código y título del proyecto.....	2
II.	Autores.....	2
III.	Resumen	2
IV.	Metodología.....	3
V.	Resultados.....	7
VI.	Publicaciones.....	10
VII.	Referencias	11
VIII.	Anexos	12

I. CÓDIGO Y TÍTULO DEL PROYECTO

Código del Proyecto **0000-0218-8007**

Nombre del proyecto: **GEOTOROTUR: Geomática utilizando ToroDB aplicada al turismo**

II. AUTORES

TABLA 1: INVESTIGADORES DEL PROYECTO

Nombre y apellidos	Escuela	Jornada (h/semana)	Meses en el proyecto
Leonardo Viquez Acuña	Carrera de Ingeniería en Computación Campus Tecnológico Local San Carlos	8	Enero de 2017 a Diciembre de 2019
Oscar Viquez Acuña	Carrera de Ingeniería en Computación Campus Tecnológico Local San Carlos	4	Enero de 2017 a Diciembre de 2019
Marlen Treviño Villalobos	Carrera de Ingeniería en Computación Campus Tecnológico Local San Carlos	12	Enero de 2017 a Diciembre de 2019
Lady Fernández Mora	Escuela de Idiomas y Ciencias Sociales / Carrera de Gestión Turismo Rural Sostenible	8	Enero de 2018 a Diciembre de 2019
Rocío Quirós Oviedo	Carrera de Ingeniería en Computación Campus Tecnológico Local San Carlos	8	Enero de 2017 a Diciembre de 2019
Gaudy Esquivel Vega	Carrera de Ingeniería en Computación Campus Tecnológico Local San Carlos	4	Enero de 2017 a Junio de 2018

En el informe final de este proyecto se omite el Documento 1 - que constituye un informe técnico sobre los resultados obtenidos en el proyecto o actividad- ya que éste puede ser reemplazado por un artículo científico presentado para su publicación en una revista o conferencia indexada en ISI Web of Science (Thomson Reuters) o Scopus, según el acuerdo, tomado por el Consejo Institucional en la Sesión Ordinaria No. 2968, Artículo 8, del 28 de abril de 2016. Sin embargo, para ayudar a comprender mejor el cumplimiento de los objetivos del presente proyecto, en este documento se desarrollaron las secciones de resumen, metodología y resultados.

III. RESUMEN

La Carrera de Ingeniería en Computación poseen un geoportal denominado IDEHN (Infraestructura de Datos Espaciales de la Región Huetar Norte) y tiene una escala regional,

mientras que la Municipalidad de San Carlos también cuenta con un geoportal que se conoce como IDESCA (Infraestructura de Datos Espaciales de San Carlos) y su escala es local. Estos geoportales permiten a los usuarios compartir y editar datos geoespaciales mediante la utilización de geoservicios y tienen la capacidad de interactuar entre sí, generando un mayor consumo de los geoservicios de datos geográficos ofrecidos en la Región.

En ambos casos, los geoportales utilizan como servidor de mapas la herramienta Geoserver. Sin embargo, se identificó que este servidor de mapas está desaprovechando las ventajas ofrecidas por múltiples paradigmas de bases de datos (como el NoSQL) que brindan una simplicidad mayor en el almacenaje de los datos lo que implica una recuperación más eficiente de los mismos. Esto, se debe fundamentalmente a 4 razones: en los últimos años se ha presentado un aumento significativo en el número de usuarios de información geográfica, el volumen de datos geográficos crece diariamente de forma exponencial, de igual forma se incrementa la cantidad de información histórica almacenada y mucha de la información que se genera hoy en día corresponde a datos no estructurados.

El motivo anterior, llevó a presentar una investigación con el objetivo de contribuir con la mejora de la eficiencia de la herramienta Geoserver, mediante la implementación de una base de datos híbrida y distribuida para un efectivo aprovechamiento de las ventajas de los paradigmas SQL y NoSQL.

Los principales resultados obtenidos con la ejecución de este proyecto son: una base de datos distribuida con 45 capas de datos geográficas replicadas en MongoDB y PostgreSQL, un geoservicio OWS que trabaja con dicha base de datos, una aplicación web responsive para la gestión de consultas del Instituto Costarricense de Turismo (ICT) que tiene por objetivo evaluar si los resultados de esta propuesta mejoraron significativa la eficiencia de Geoserver y 5 artículos científicos escritos (ver sección de Publicaciones); de los que 3 ya fueron publicados y 2 se encuentran bajo la evaluación de las revistas donde fueron enviados. Además, en este momento se está confeccionando un artículo científico adicional que pretende evidenciar los resultados finales del proyecto en cuestión, que se espera a finales del mes de abril pueda ser enviado a una revista o conferencia.

IV. METODOLOGÍA

Objetivo Específico 1: Para alcanzar este objetivo se realizó un mapeo sistemático basado en la replicación de estudios empíricos (Silva y otros, 2012) de los gestores de bases de datos de código abierto que permitan una integración de los paradigmas NoSQL con el motor de bases de datos PostgreSQL permitiendo seleccionar la herramienta más óptima para la conformación de una base de datos híbrida.

- Se realizó una exploración de los sistemas gestores de bases de datos de paradigmas NoSQL analizando sus propiedades, con la finalidad de determinar si existe compatibilidad para la integración con el motor PostgreSQL. Los sistemas gestores de base de datos NoSQL identificados en este proceso fueron MongoDB, CouchDB y ArangoDB.

- Se instalaron y configuraron los servidores de base de datos: MongoDB, CouchDB y ArangoDB.
- Se desarrollaron pruebas de rendimiento a MongoDB, CouchDB y ArangoDB utilizando la herramienta JMeter.
- Se analizaron los resultados obtenidos de las pruebas de rendimiento realizadas y se determinó que la herramienta MongoDB tenía las mejores condiciones para implementar una base de datos híbrida y distribuida con el motor de bases de datos PostgreSQL.

Objetivo Específico 2: Mediante este objetivo se generó una base de datos con características híbridas y distribuidas. Esta base de datos está constituida por dos motores de bases de datos de paradigmas SQL y NoSQL sobre almacenes de datos NewSQL (Grolinger, Higashino, Tiwari, & Capretz, 2013). En este caso, se trabajó con PostgreSQL y MongoDB. Además, como mecanismo de replicación de los datos desde MongoDB hacia PostgreSQL se usó ToroDB de modo que se permite el manejo distribuido particularmente de la información geográfica para el consumo de la misma mediante geoservicios. Los pasos seguidos, son los siguientes:

- Se definieron los requisitos mínimos de hardware para un adecuado desempeño del servidor de bases de datos seleccionado.
- Se instalaron y configuraron los servidores de base de datos PostgreSQL y MongoDB para que trabajen con ToroDB.
- Se recibió una capacitación en temas básicos y avanzados de MongoDB.
- Se definió el formato GeoJSON para el manejo de los datos geográficos.
- Se analizó la base de datos geográfica implementada por Geoserver, pero se debió crear la base de datos desde cero. El mecanismo seguido fue importar las capas de datos geográficas en formato GeoJSON en MongoDB y la herramienta ToroDB se encarga de la replicación de los datos en PostgreSQL.
- Dado que ToroDB hace una descomposición de los objetos almacenados en MongoDB en varias tablas en PostgreSQL, fue necesaria la generación de una vista materializada por cada capa de datos geográfica, esto para unificar los datos desagrupados de dichas tablas y facilitar la consulta de datos.

Objetivo específico 3: Este objetivo buscó adaptar el geoservicio WMS a la modificación de una nueva base de datos distribuida con propiedades híbridas de sus paradigmas. Para lograr este objetivo se desarrollaron las siguientes actividades:

- Se capacitó a los estudiantes asistentes en la instalación, configuración y uso de Geoserver. Además, en los lenguajes de programación necesarios para trabajar con dicha herramienta.
- Se realizó un análisis de la implementación del protocolo WMS que implementa Geoserver para determinar la integración con el motor de bases de datos PostgreSQL.

- El análisis efectuado en el paso anterior, determinó que se debía desarrollar un nuevo geoservicio por lo que se decidió implementar un geoservicio OWS (Open Geospatial Consortium, 2016).
- También, fue necesario definir un mecanismo para realizar las solicitudes al servidor mediante pseudocódigo y este debió ser traducido para ejecutarlo en PostgreSQL o MongoDB. Este último punto, implicó el desarrollo de un analizador sintáctico (o parser).
- Además, se desarrollaron una serie de pruebas de rendimiento por tipo de consulta en PostgreSQL y MongoDB para que el analizador sintáctico tuviera un mecanismo para decidir en cuál de los dos sistemas gestores de bases de datos el geoservicio ejecutaría la consulta. De este modo, se determinó que cuando la consulta implique procesamiento geográfico el geoservicio utilizaría PostgreSQL y en cualquier otro caso ejecutaría la consulta en MongoDB.

Objetivo específico 4: Para alcanzar este objetivo se desarrolló una aplicación web responsive que consume los datos a través del geoservicio OWS implementado y que este a su vez utiliza la base de datos híbrida y distribuida con PostgreSQL y MongoDB. Pasos seguidos:

- Se definieron los requisitos mínimos que debe cumplir la herramienta a desarrollar para que trabaje con el geoservicio OWS.
- En conjunto con el ICT se identificaron un conjunto de posibles herramientas a desarrollar utilizando una lluvia de ideas, y se seleccionó una aplicación para la consulta de datos.
- Se definieron los requisitos para la aplicación.
- Se efectuó el diseño de la aplicación.
- Se desarrolló e implementó la aplicación coordinando con el equipo de investigación y asistentes contratados. Se utilizó una metodología basada en tareas considerando que el equipo de desarrollo estuvo conformado por asistentes especiales (Wurdel, Sinnig, & Forbrig, 2008). Dicha aplicación se encuentra disponible en: <http://geotorotur.ddns.net/> o a nivel interno del TEC se puede visualizar la aplicación en la siguiente dirección IP <http://172.24.164.174:8090/#!/>.
- Se recopilaron las capas de datos geográficos del ICT necesarias para la implementación de la aplicación y se verificó que cumplieran con los estándares definidos por el Instituto Geográfico Nacional (Instituto Geográfico Nacional, s.f.).
- Se realizaron pruebas de caja blanca y caja negra para la identificación de posibles defectos.
- Se capacitó a funcionarios del ICT en el uso de la herramienta desarrollada.

Objetivo específico 5: Con este objetivo se esperaba obtener un análisis del rendimiento del geoservicio WMS de Geoserver trabajando con una base de datos híbrida y distribuida. Pasos seguidos:

- Se instaló y configuró el ambiente de pruebas.

- Se definieron 75 consultas, 25 de estas implican procesamiento geográfico y las otras 50 consultas no requieren de procesamiento geográfico. Estas 75 consultas debieron ser escritas en el seudocódigo conocido como ECQL (Extensión del Lenguaje de Consulta Común) definido por Geoserver y también en el seudocódigo desarrollado para el geoservicio OWS implementado en este proyecto.
- Se definieron los mecanismos de pruebas haciendo uso de la herramienta JMeter.
- Dado que el proceso de configuración del ambiente de pruebas fue muy complejo y lento, así como la escritura de las consultas mencionadas anteriormente implicó más tiempo del que se tenía previsto; todavía se está en el proceso de ejecución de las pruebas.

V. RESULTADOS

El objetivo general de este proyecto fue contribuir con la mejora de la gestión de datos geográficos del nodo de la IDE Local y Regional, se considera que el grado de avance del proyecto al mes de marzo de 2020 fue de aproximadamente un 98%; pues, todavía se está trabajando en la redacción del artículo final del proyecto. Seguidamente, se describen los principales resultados del proyecto.

TABLA 2: RESULTADOS DEL PROYECTO

Producto	Descripción
<p>El diagrama muestra la estructura de la base de datos idehn.rios. Las tablas principales y sus relaciones son:</p> <ul style="list-style-type: none"> idehn.rios: Tabla principal que contiene referencias a otras tablas. idehn.rios_crs: Tabla de coordenadas de referencia espacial. idehn.rios_crs_properties: Tabla de propiedades de las coordenadas de referencia espacial. idehn.rios_features: Tabla de características geográficas. idehn.rios_features_geometry: Tabla de geometría de las características geográficas. idehn.rios_features_geometry_properties: Tabla de propiedades de la geometría de las características geográficas. idehn.rios_features_geometry_coordnates\$1, idehn.rios_features_geometry_coordnates\$2, idehn.rios_features_geometry_coordnates\$3: Tablas de coordenadas de las características geográficas. <p>Se incluyen también ejemplos de datos de cada una de estas tablas:</p> <pre> -- idehn.rios did rid seq type_s properties_e ----- ----- ----- ----- ----- 1 0 0 character varying f name ----- 1 -- idehn.rios_crs did rid pid seq name_s ----- ----- ----- ----- ----- 1 0 0 0 EPSG:4326 -- idehn.rios_crs_properties did rid pid seq coordinates_n type_s ----- ----- ----- ----- ----- 1 0 0 0 t MultilineString 2 0 1 1 t MultilineString 3 0 2 2 t MultilineString 4 0 3 3 t MultilineString 5 0 4 4 t MultilineString -- idehn.rios_features did rid seq geometry_name_s type_s properties_e ----- ----- ----- ----- ----- 1 0 0 rio_0 t Feature f 2 0 1 rio_1 t Feature f 3 0 2 rio_2 t Feature f 4 0 3 rio_3 t Feature f -- idehn.rios_features_geometry did rid pid seq nombre_s categoria_s ----- ----- ----- ----- ----- 1 0 0 0 SIN NOMBRE PERMANENTE 2 0 1 1 RIO SARDINALITO PERMANENTE 3 0 2 2 SIN NOMBRE PERMANENTE 4 0 3 3 RIO MORENO PERMANENTE 5 0 4 4 SIN NOMBRE PERMANENTE -- idehn.rios_features_geometry_coordnates\$1 did rid pid seq v_e ----- ----- ----- ----- ----- 1 0 0 0 0 t 2 0 1 1 1 0 t 3 0 2 2 2 0 t 4 0 3 3 3 0 t -- idehn.rios_features_geometry_coordnates\$2 did rid pid seq v_e ----- ----- ----- ----- ----- 1 0 0 0 0 t 2 0 1 0 1 t 3 0 2 0 2 t 4 0 3 0 3 t </pre>	<p>Se creó una base de datos distribuida, que funciona con MongoDB y PostgreSQL. Además, se replicaron un total de 45 capas de datos geográficas y se definieron 45 vistas materializadas en PostgreSQL para poder consultar dichos datos. Este último punto, debido a que ToroDB en el proceso de replicación descompone los objetos geográficos en varias tablas.</p>

Producto	Descripción
	<p>Se tiene funcionando un geoservicio OWS (Open Web Services) que consume los datos de la base de datos distribuida.</p>

Producto	Descripción
	<p>Se desarrolló una aplicación web responsive que se encuentra instalada y disponible en el URL http://geotorotur.ddns.net/. También, a nivel interno del TEC se puede visualizar la aplicación en la siguiente dirección IP http://172.24.164.174:8090/#!/. Dicha aplicación, utiliza las capas de hospedaje, gastronomía y atractivos turísticos que fueron proporcionadas por el ICT y están almacenados en la base de datos distribuida. El objetivo de esta aplicación es facilitar la consulta de los datos de hospedaje, gastronomía y atractivos turísticos del ICT a empresarios, inversionistas, turistas y público en general tanto a nivel nacional como internacional para una toma de decisiones oportuna. Adicionalmente, la aplicación hace uso del geoservicio desarrollado y de este modo se aprovecha la oportunidad de gestionar datos a través de aplicaciones web y móviles para el turista.</p>

VI. PUBLICACIONES

Como producto académico del proyecto, en la siguiente tabla se presentan las publicaciones realizadas durante el período de ejecución del proyecto.

TABLA 3: PUBLICACIONES DEL PROYECTO

Nombre de obra	Tipo de obra	Estado (aceptado por publicar y publicado)	Base de datos de indexación	Nombre de Evento	Contó con Comité científico (Si ó NO)	DOI
Análisis de rendimiento para bases de datos geográficas: PostgreSQL vs MongoDB	Artículo	Publicado		Congreso de Computación para el Desarrollo (COMPDES)	Sí	ANEXO #1 - LIBRO COMPDES 2018
Una comparación de rendimiento entre MongoDB, ArangoDB y CouchBase para la operación lectura sobre bases de datos geográficas	Artículo	Publicado	SCOPUS IEEE Xplore® Digital Library	CONCAPAN	Sí	https://doi.org/10.1109/CONCAPAN.2018.8596387 ANEXO #2 - Una comparación de rendimiento entre MongoDB, ArangoDB y CouchBase para la operación lectura sobre bases de datos geográficas
Una comparación de rendimiento entre bases de datos NoSQL: MongoDB y ArangoDB	Artículo	Publicado	Scielo	Tecnología en Marcha	Sí	https://doi.org/10.18845/tm.v32i6.4223 ANEXO #3 - Una comparación de rendimiento entre bases de datos NoSQL MongoDB y ArangoDB
Comparison of the response times of MongoDB and PostgreSQL according to type of query in geographical databases	Artículo	Presentado el 22 de octubre de 2019	SCOPUS	Revista Computación y Sistemas	Sí	ANEXO #4 - Comparison of the response times of MongoDB and PostgreSQL according to type of query in geographical databases
Hybrid Storage Engine for geospatial data using NoSQL and SQL paradigms	Artículo	Presentado el 12 de noviembre de 2019	Scielo	Tecnología en Marcha	Sí	ANEXO #5 - Hybrid Storage Engine for geospatial data using NoSQL and SQL paradigms

Estos artículos pueden ser consultados en la sección de Anexos, que se encuentra al final del documento.

VII. REFERENCIAS

- Alwan, A. A. (2009). Improved integrity constraints checking in distributed databases by exploiting local checking. *Journal of Computer Science and Technology. Journal of Computer Science and Technology*, 24(4), 665-674. doi:10.1007/s11390-009-9261-0
- Antoñana, J. V. (2008). *Pasado, presente y futuro de las Infraestructuras de Datos Espaciales*. Recuperado el 18 de abril de 2015, de Sistema de Información Territorial de Navarra: <http://ww2.pcypsitna.navarra.es/Aprende/Documents/PASADO-PRESENTE-Y-FUTURO-DE-LAS-INFRAESTRUCTURAS-DE-DATOS-ESPACIALES.pdf.pdf>
- Ballari, D., Vilches, L., Pérez, D., Pacheco, D., & Fernández, V. (17 de octubre de 2014). *Tendencias en infraestructuras de datos espaciales en el contexto*. Recuperado el 17 de abril de 2015, de Repositorio Digital de la Universidad de Cuenca: http://dspace.ucuenca.edu.ec/bitstream/123456789/21364/1/IC_17_Ballari%20et%20al.pdf
- Diagnóstico sobre temas relevantes de la gestión de información geoespacial y desarrollo de las Infraestructuras de Datos Espaciales – IDE en los países de las Américas*. (2012). Recuperado el 10 de abril de 2015, de Comité Permanente para la Infraestructura de datos Geoespaciales de las Américas (CP-IDEA): <http://www.cp-idea.org/index.php/component/jdownloads/finish/42-seminario-de-riesgos-y-desastres-naturales-cp-idea/245-diagnostico-sobre-temas-relevantes-de-la-gestion-de-informacion-geoespacial-y-desarrollo-de-las-infraestructuras-de-datos-espaciales--id>
- Gobierno de Costa Rica. (12 de julio de 2013). *La Gaceta*(134), págs. 7-8.
- Grolinger, K., Higashino, W. A., Tiwari, A., & Capretz, M. A. (2013). Data management in cloud environments: NoSQL and NewSQL data stores. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1), 1-24. doi:10.1186/2192-113X-2-22
- Instituto Geográfico Nacional. (s.f.). *Sistema Nacional de Información Territorial (SNIT)*. Recuperado el 24 de abril de 2016, de Documentos y Manuales: http://www.snitcr.go.cr/index.php?option=com_content&view=category&id=35&Itemid=67
- Instituto Municipal de Planeación del Ayuntamiento de Chihuahua. (s.f.). Recuperado el 2014 de junio de 18, de <http://www.implanchihuahua.gob.mx/PDU2040/>

- Open Geospatial Consortium. (2016). Open Geospatial Consortium. Recuperado el 18 de abril de 2016, de <http://www.opengeospatial.org/>
- Open Source Geospatial Content Management System. (2014). *GeoNode*. Recuperado el 2014 de junio de 18, de <http://geonode.org/>
- Silva, F. Q., Suassuna, M., Francca, A. C., Grubb, A. M., Gouveia, T. B., & Monteiro, C. V. (2012). Replication of empirical studies in software engineering research: A systematic mapping study. *Empirical Software Engineering*, 19(3), 501-557. doi:10.1007/s10664-012-9227-7
- Wurdel, M., Sinnig, D., & Forbrig, P. (2008). Engineering interactive systems: Second conference on human-centered software engineering. *HCSE 2008, and 7th international workshop on task models and diagrams, pisa, italy, september 25-26, 2008. proceedings.(Task-Based Development Methodology for Collaborative Environments)*, 118-125. doi:10.1007/978-3-540-85992-5_10
- Zhang, C., Zhao, T., & Li, W. (2015). Geospatial semantic web.(Geospatial Data Interoperability, Geography Markup Language (GML), Scalable Vector Graphics (SVG), and Geospatial Web Services). 1-33. doi:doi:10.1007/978-3-319-17801-1_1

VIII. ANEXOS

En la carpeta de Anexos, se adjunta con este documento, una carpeta digital con los archivos en formato PDF de las publicaciones realizadas como producto del desarrollo del proyecto. Asimismo, se encuentra en los anexos 6, 7 y 8, el código fuente de las aplicaciones implementadas como productos del proyecto. Por último, el anexo 9 muestra una evidencia de las sesiones de trabajo sostenidas con el ICT durante la ejecución del proyecto.

1. Libro COMPDES 2018
2. Una comparación de rendimiento entre MongoDB, ArangoDB y CouchBase para la operación lectura sobre bases de datos geográficas
3. Una comparación de rendimiento entre bases de datos NoSQL MongoDB y ArangoDB
4. Comparison of the response times of MongoDB and PostgreSQL according to type of query in geographical databases
5. Hybrid Storage Engine for geospatial data using NoSQL and SQL paradigms
6. App ICT
7. Geoservicio
8. Base de datos distribuida

9. Capacitaciones y reuniones