

ESTUDIO DE CASO

# Tecnologías de Big Data geoespacial en el Centro de Información Geoespacial de Geocuba

## *Geospacial Big Data Technologies in the Geocuba Geospace Information Center*

*José Luis Capote Fernández*

*josel.capote@gmail.com* • <https://orcid.org/0000-0002-3733-0893>

*Rafael Cruz Iglesias*

*rcruz@geomix.geocuba.cu* • <https://orcid.org/0000-0002-7564-7779>

GEOCUBA INVESTIGACIÓN Y CONSULTORÍA, CUBA

Recibido: 2019-12-20 • Aceptado: 2020-04-20

## RESUMEN

Desde 2006 la Agencia GeoMIX, de Geocuba, ha estado lidiando con los problemas relacionados con *Big Data* en el marco del proyecto de gestión y control de flotas *MovilWeb*. El gran volumen de información, la variedad de la misma y la velocidad con que se requieren los análisis han sido una constante en este trabajo. Por otro lado, Geocuba ha potenciado la utilización de los Vehículos Aéreos no Tripulados (VANT) en el levantamiento de diferente índole, acumulando un gran volumen de datos que crece continuamente. Hasta hace poco las soluciones a estos problemas estuvieron siempre relacionadas con cambios de arquitectura, optimización de índices, incremento de recursos de cómputo, etc. Aunque las tecnologías relacionadas con *Big Data* ya se utilizaban en el contexto global, no es hasta hace poco que, gracias a las bondades de *Docker*, se habilitaron para una más amplia utilización. En este documento se exponen los resultados, desde el punto de vista de Gestión de Información Geoespacial, de evaluar tres tecnologías de *Big Data* geoespacial que son *GeoMesa*, *GeoTrellis* y *MrGeo* en un experimento cuyo objetivo es completar un flujo de gestión que comprende desde la ingestión en el clúster de *Big Data* hasta su publicación como servicio web, tanto de datos vectoriales como ráster. El completamiento de este flujo garantiza la generación de una cadena de valor agregado sobre estos datos geoespaciales primarios. El presente trabajo

expone las experiencias obtenidas en el trabajo con *Big Data* geoespacial, su vinculación a algunos proyectos actuales y las perspectivas futuras de su utilización en el Centro de Información Geoespacial de Geocuba.

**PALABRAS CLAVE:** big data geoespacial; centro de información geoespacial; infraestructuras de datos espaciales.

## ABSTRACT

*Since 2006, Geocuba's GeoMIX Agency has been dealing with Big Data related issues under the Fleet Management and Control project, MovilWeb. The large volume of information, the variety of it, and the speed with which the analysis is required have been a constant in this work. On the other hand, Geocuba has promoted the use of UAVs in the survey of different kinds, accumulating a large volume of data that is continually growing. Until recently, solutions to these problems were always related to changes in architecture, index optimization, increased computing resources, etc. Although Big Data related technologies were already used in the global context, it was not until recently that, thanks to the benefits of Docker, they were enabled for wider use. This document presents the results, for Geospatial Information Management, of evaluating three Big Data Geospatial technologies that are GeoMesa, GeoTrellis, and MrGeo in an experiment whose objective is to complete a management flow that includes from ingestion in the Big Data cluster until publication as a Web Service, both vector and raster data. Completion of this flow ensures the generation of an added value chain on this primary geospatial data. The present work exposes the experiences obtained in the work with Big Data Geospatial, its connection to some current projects, and the future perspectives of its use in the Geospatial Information Center of Geocuba.*

**KEYWORDS:** *geospatial big data; geospatial information center; spatial data infrastructures.*

## INTRODUCCIÓN

Los datos vectoriales espaciales con alta precisión y amplia cobertura se han disparado a nivel mundial, como la cobertura terrestre, las redes sociales y otros conjuntos de datos, lo que brinda una buena oportunidad para mejorar la toma de decisiones nacionales macroscópicas, la supervisión social, los servicios públicos y capacidades de emergencia. Este crecimiento de

datos a volúmenes nunca antes alcanzados, a una alta velocidad y con una variedad de información que requiere ser analizada es actualmente tratado como *Big Data* (Cugler, *et al.*, 2013; McCoy, 2017; Shekhar, 2012). En el entorno espacial tradicional los datos suelen ser de dos tipos, vectoriales o ráster, y en el entorno de *Big Data* encontramos las tecnologías correspondientes para soportar estos tipos de datos.

## ESCENARIOS IDENTIFICADOS DE BIG DATA EN GEOCUBA

### Big Data en *MovilWeb*

*MovilWeb* es el grupo de aplicaciones que sirven de plataforma de *software* para el Sistema de Gestión y Control de Flotas (González Suárez, *et al.*, 2018). Esta plataforma utiliza dos variantes de operación, modo diferido, donde los datos son almacenados en un computador de a bordo y modo tiempo real, donde se transmiten vía GPRS a los servidores de forma continua.

En el marco del proyecto *MovilWeb*, se fue produciendo un escalado en vehículos que implica hoy enfrentar un problema de *Big Data* geoespacial. Durante mucho tiempo se han utilizado diferentes estrategias para mitigar el efecto de este crecimiento acelerado de los datos en la celeridad de los análisis y el procesamiento de los datos (ver tabla 1).

El problema inicial estuvo relacionado con el almacenamiento de los datos de las trazas de los vehículos. En un inicio estos datos se comenzaron a almacenar en las bases de datos, separados por campos, pero se comenzó a enfrentar el problema del volumen, el crecimiento de las tablas era excesivamente grande, y no se quería eliminar datos porque se sabía que en algún momento serían necesarios. Se desarrollaron varios talleres a los cuales se invitaron algunas instituciones y como conclusión se adoptó la variante de ubicar las trazas en archivos en la Nube y desarrollar los procedimientos para su procesamiento, indexación y utilización bajo demanda. Ya en la base de datos sólo quedaban los índices de las trazas que eran subidas al sistema y, aun así, su crecimiento era sin precedentes.

Los archivos con las trazas en el disco duro fue también una buena solución durante determinado tiempo. El crecimiento de estos, cada uno de aproximadamente 200 kilobytes, implicaban la distribución en varios servidores y el desarrollo de sistemas de gestión de la información en múltiples volúmenes. También la carga y descarga de una trayectoria en el sistema para la visualización y análisis, un proceso que fue más engorroso en la medida que fueron creciendo los usuarios.

Una de las primeras estrategias utilizadas fue el particionamiento horizontal de los datos en las bases de datos espaciales de *PostgreSQL/Postgis* utilizando diferentes índices (Abreus,

Tabla 1. Cantidad de medios instalados por año (González Suárez, *et al.*, 2018).

Año	Bases	Total de medios
2006	27	1241
2007	37	1320
2008	85	3771
2009	175	6878
2010	193	7796
2011	206	9070
2012	233	11234
2013	262	13497
2014	372	16 797
2015	413	17013
2016	500	20000
2017	506	22077

2006). En su momento esta solución permitió mejorar considerablemente el rendimiento de las consultas. Pero los datos continuaron creciendo a mayor velocidad que antes y el particionado se volvió también complejo porque se generaban muchas tablas y el mantenimiento de las estadísticas del sistema se hacía lento.

Al inicio de la tarea, en mayo de 2006, con una sola base de transporte la base de datos crecía a un ritmo de 2 000 000 (300 MB) de artículos mensuales y el promedio de visitas diarias a la aplicación era de 27 (hoy es de 50 500 visitas diarias). Cuatro meses más tarde, cuando había 20 bases de transporte y 1000 vehículos, tenía 42 millones artículos y ocupaba 10 GB al procesar un total de 10 000 trayectorias. Hoy hay 500 bases de transporte, 20 000 vehículos se procesan 6 000 trayectorias diarias y se reciben 50 500 visitas diarias.

Actualmente se han establecido configuraciones avanzadas de clúster de *PostgreSQL* y arquitecturas de Balanceo de carga en las aplicaciones que han garantizado la operatividad del sistema que continúa creciendo. El escenario de *MovilWeb* posee las tres “V” descritas para la *Big Data* (Furht & Villanustre, 2016; Shekhar, 2012) y es ideal para comenzar a utilizar un grupo de soluciones ya evaluadas con muy buenos resultados.

### Las imágenes de los VANT

Desde hace varios años el grupo empresarial Geocuba trabaja con el uso de los VANT para el levantamiento en escenarios donde esta tecnología es factible. En los inicios los resultados de estos trabajos, que generalmente tienen gran tamaño, quedaban en los discos duros de los especialistas, acumulándose una gran cantidad de información que muchas veces terminaba siendo borrada por necesidad de espacio. Poco a poco se fue organizando la información en los archivos técnicos y creándose metadatos para la misma, pero estos quedaban almacenados de forma inoperante para su reutilización de forma fácil.

Actualmente la tecnología VANT se ha difundido a las empresas de Geocuba en todas las provincias y se ha identificado una gran cantidad de modelos de negocios relacionados con ella. Solamente es necesario un grupo de tecnologías de apoyo que garanticen la gestión de los resultados de los análisis de una forma eficiente para que estos modelos de negocios sean habilitados. Muchas de estas tecnologías habilitadoras de apoyo las encontraremos en *Big Data* geoespacial donde es posible almacenar, recuperar, analizar y visualizar estos datos ráster con un alto desempeño.

### Generación de mapas isoyéticos en el proyecto HidroSIG

*HidroSIG* es una herramienta computacional que combina funciones propias de los sistemas de información geográfica con herramientas que le permiten al usuario georeferenciar los datos asociados a los recursos hídricos para la visualización, análisis y toma de decisiones. En el marco de este proyecto se implementaron procesos WPS (*Web Processing Service*) (Schut, 2007), para generar diariamente de forma automática los isoyéticos con la información de las lluvias, y para obtener las lluvias por diferentes niveles de agregación espacial (cuenca y municipio), estos procesos WPS se incorporaron como extensiones de *Geoserver*. El análisis de

la información de los pluviómetros reportada diariamente y la generación de los mapas isotéticos fue resuelta utilizando métodos tradicionales, enfrentando los costos que implica la generación y almacenamiento de los datos ráster una vez obtenidos, para ahorrar recursos de procesamiento. Este proyecto constituye un escenario adecuado para la utilización de tecnologías de *Big Data* geoespacial para el procesamiento y gestión de los mismos.

### **DevOps y Docker: Tecnologías habilitadoras**

El término *DevOps*, corresponde a las siglas de *Development & Operations*. El término existe desde hace algunos años, y si bien su popularidad sigue aumentando, aún no es fácil encontrar una definición clara y consensuada que dé identidad al concepto de *DevOps* (Willis, 2012). A pesar de esta diversidad de definiciones, es claro que lo principal de *DevOps* son sus aspectos culturales como el mantener un buen flujo de información, colaboración inter-funcional, responsabilidades compartidas, y fomentar nuevas ideas. En otras palabras, *DevOps* busca establecer un enfoque en la colaboración entre equipos con un objetivo común, el de desarrollar *software* de alta calidad. (Kim, *et al.*, 2015)

*Docker* es una plataforma que permite la creación, ejecución y gestión de ambientes aislados para las aplicaciones de *software* (conocido generalmente como “containerización”) mediante un conjunto de herramientas integradas que brindan solución a algunas de las problemáticas principales que se presentan cuando se debe implementar la gestión de ambientes de ejecución.

La principal funcionalidad de *Docker* consiste en la generación de ambientes de ejecución que permite el empaquetado en imágenes livianas de todas las dependencias requeridas por las aplicaciones que vivirán en él. Estas imágenes luego pueden ser distribuidas, y, al contener su entorno de ejecución completo, pueden ser ejecutadas en contenedores, lo que permite asegurar que las aplicaciones se comportarán de la misma manera en todas las ejecuciones. *Docker* surge como alternativa frente a la clásica técnica de virtualización de servidores con un enfoque arquitectónico distinto que permite la reutilización de ambientes de ejecución con un mejor uso de los recursos de *hardware*, principalmente de espacio en disco (Moreno, *et al.*, 2016).

Como suele ser el caso de las aplicaciones de *Big Data* geoespacial (múltiples sistemas interdependientes que, en producción, viven en diferentes máquinas), se presupone un clúster completamente funcional incluso para las operaciones básicas. Por ejemplo, *Apache Accumulo* requiere que tanto *Hadoop* (HDFS) como *ZooKeeper* estén configurados y en ejecución incluso para inicializar. El mantenimiento de una instalación local de estos recursos ya es en sí mismo un proceso difícil, aumenta la complejidad, ralentiza el flujo de trabajo de desarrollo e introduce oportunidades para las incompatibilidades del entorno.

En este caso *Docker* permitió acceder a un grupo de contenedores relacionados con proyectos exitosos de *Big Data* geoespacial no solo para evaluar sus capacidades sino para utilizarlos como componentes en proyectos. Muchos de estos proyectos estaban implementados en diferentes lenguajes utilizando una gran cantidad de dependencia y su construcción desde el código fuente y configuración para el despliegue constituía una barrera a su utilización.

Este trabajo se propone compartir las experiencias del Centro de Información Geoespacial de Geocuba en la gestión de datos geoespaciales, a partir de la evaluación y experimentación de varias tecnologías de *Big Data* geoespacial.

## METODOLOGÍA

Para enfrentar los retos de los escenarios identificados se realizó una evaluación consistente en un levantamiento de diferentes tecnologías cuya asimilación fuera factible teniendo en cuenta el lenguaje de implementación y las facilidades de despliegue. En cada uno de los casos el proceso de evaluación incluye la preparación del escenario de pruebas para la ejecución de experimentos y la compilación del código fuente para garantizar la posibilidad de suplir deficiencias.

En este contexto, y como segundo paso de la metodología, se desarrollaron experimentos orientados a ofrecer soporte a los tipos de datos que se gestionan actualmente con características de *Big Data*, las trazas de los vehículos y las imágenes de satélites de los drones. El objetivo de los experimentos desarrollado es completar el ciclo de gestión de información geoespacial, que comprende desde su captura hasta su publicación en web por medio de algún servicio estándar. No se hace referencia a las características del *hardware* en los experimentos desarrollados, ni se establecen comparaciones, porque no fue posible en las condiciones de Cuba lograr escenarios comparables con los que aparecen en publicaciones similares. El principal requerimiento es la completitud del proceso.

## RESULTADOS Y DISCUSIÓN

### EVALUACIÓN DE TECNOLOGÍAS

Hasta el momento se han evaluado tres tecnologías *GeoMesa* (Hughes, *et al.*, 2015), *GeoTrellis* (OSGeo, 2018) y *MrGeo* (NGA, 2015) buscando resolver diferentes problemas en escenarios identificados.

#### **GeoMesa**

*GeoMesa* es un conjunto de herramientas de código abierto con licencia *Apache* que permite análisis geoespaciales a gran escala en la nube y sistemas informáticos distribuidos, permitiéndole administrar y analizar los enormes conjuntos de datos espacio-temporales que las aplicaciones de IoT (*Internet of Things*), redes sociales, seguimiento y teléfonos móviles buscan ventaja de hoy.

*GeoMesa* proporciona persistencia de datos espacio-temporales sobre bases de datos distribuidas orientadas a columnas, como *Accumulo*, *HBase* y *Cassandra*, para el almacenamiento masivo de datos de puntos, líneas y polígonos. Permite un acceso rápido a estos datos a través de consultas que aprovechan al máximo las propiedades geográficas para especificar la distancia y el área. *GeoMesa* también proporciona soporte para el procesamiento de flujo

en tiempo casi real de datos espacio-temporales mediante la estratificación de la semántica espacial en la parte superior del sistema de mensajería *Apache Kafka*.

A través de un servidor de información geográfica como *GeoServer*, *GeoMesa* facilita la integración con una amplia gama de clientes de mapeo existentes al permitir el acceso a sus bases de datos y capacidades de transmisión sobre API y protocolos OGC (*Open Geospatial Consortium*) estándar como WFS (Vretanos, 2005) y WMS (de la Beaujardiere, 2003). Estas interfaces también permiten que *GeoMesa* administre interfaces de usuario de mapas y sirva datos para análisis, como consultas, histogramas, mapas de calor y análisis de series de tiempo (Hughes, *et al.*, 2015).

### **GeoTrellis**

*GeoTrellis* es un proyecto desarrollado por la empresa *LocationTech* que da una respuesta a la creciente necesidad de procesamiento escalable de datos ráster. Al utilizar *Apache Spark* como tecnología base para distribuir el procesamiento, los límites de datos a manejar están dados por los de esta tecnología, logrando procesar volúmenes del orden de terabytes.

*GeoTrellis* también proporciona herramientas para representar datos ráster en PNG o para almacenar metadatos sobre archivos ráster como JSON. Su objetivo es proporcionar procesamiento ráster a velocidades web (por debajo de un segundo o menos) con servicios *Web RESTful*, así como proporcionar procesamiento rápido por lotes de grandes conjuntos de datos ráster. De esta forma proporciona un escenario adecuado para el desarrollo de aplicaciones geoespaciales en web donde se requiere tecnologías de *Big Data*.

Sus funcionalidades están divididas en aquellas orientadas al trabajo con datos geoespaciales con *Apache Spark*, las orientadas al trabajo con datos ráster y las destinadas a trabajar con datos vectoriales. Usa el concepto de RDD (conjuntos de datos resilientes distribuidos) (Yao, *et al.*, 2017) tanto para almacenar como para proporcionar el procesamiento escalable de los datos permitiendo operaciones de recortar, filtrar, unir, enmascarar, fusionar, crear pirámides, remuestrear, dividir, crear mosaicos, etc. sobre capas ráster en este formato. Al utilizar datos vectoriales implementa múltiples métodos para su procesamiento, sobre todo métodos para la generación de datos ráster por medio de interpolación (Kriging, Distancia Inversa Simple y Pesada, etc.), para la generación de diferentes variogramas, triangulación de Delaunay y diagramas de Voronoy, etc. (OSGeo, 2018).

*GeoTrellis* está disponible para ser utilizado desde el lenguaje *Python* por medio del proyecto *GeoPySpark* (Locationtech Labs, 2016/2018). Esta característica le habilita para ser utilizado en aplicaciones que ejecutan guiones en la Nube (por ejemplo *GeoNotebook*). Además, permite crear un servicio de mapas en internet basado en mosaicos (TMS) que le permite cerrar el ciclo desde la gestión hasta la publicación de los datos desde la Nube.

### **MrGeo**

*MrGeo* es un conjunto de herramientas geoespaciales diseñado para realizar un procesamiento geoespacial basado en ráster a escala regional o global. Al distribuir el cómputo entre varias computadoras puede reducir considerablemente el procesamiento en comparación a los

sistemas tradicionales. *MrGeo* proporciona dos formas de visualizar estos productos a través de los estándares *Open Geospatial Consortium* (OGC), utilizando *GeoServer* por medio de una extensión o por una implementación de servicio WMS propia del proyecto.

Dentro de las funcionalidades que presenta está el almacenamiento y procesamiento escalable de datos ráster, propone almacenar los datos en un formato listo para el cálculo, eliminando varios pasos de pre-procesamiento del flujo de trabajo de producción. Además, proporciona un conjunto de métodos de análisis robustos para *Apache Spark* que incluyen operaciones matemáticas algebraicas, operaciones focales (por ejemplo, pendiente y aspecto). *MrGeo* presenta una interfaz de álgebra de mapas que permite el desarrollo de algoritmos personalizados en una simple API de *scripting*.

Una de las ventajas de *MrGeo* es que garantiza un flujo de trabajo desde el almacenamiento del ráster hasta su publicación como servicio OGC. *MrGeo* se ha utilizado para almacenar, indexar, crear mosaicos y bases de datos de imágenes a escala de múltiples terabytes en pirámide. Una vez almacenados, estos datos están disponibles a través de un servicio de mosaicos de mapas (TMS) y servicio de mapa en web (WMS).

Al igual que *GeoTellis*, *MrGeo* incluye un subproyecto que permite su utilización desde el lenguaje *Python* llamado *pymrgeo*. Este proyecto hace simple la ejecución de las funcionalidades de álgebras de mapas dentro de un guion de ejecución.

En el caso de *MrGeo* se realizaron modificaciones a la versión del código fuente disponible en *Github* para resolver algunos problemas detectados en la ejecución de las pruebas con los juegos de datos de nuestra agencia y para la creación de un producto más afín con las tecnologías que se utilizan en la actualidad.

## EXPERIMENTACIÓN

En este contexto se desarrollaron experimentos orientados a ofrecer soporte a los tipos de datos que se gestionan actualmente con características de *Big Data*, las trazas de los vehículos y las imágenes de satélites de los drones. El objetivo de los experimentos desarrollado es completar el ciclo de gestión de información geoespacial, que comprende desde su captura hasta su publicación en web por medio de algún servicio estándar. No se hace referencia a las características del *hardware* en los experimentos desarrollados, ni se establecen comparaciones, porque no fue posible en las condiciones de Cuba lograr escenarios comparables con los que aparecen en publicaciones similares. El principal requerimiento es la completitud del proceso.

### Almacenamiento de Trazas GPS en Big Data

Las trazas GPS están almacenadas en la plataforma como archivos compactados. En las primeras versiones se utilizaba un archivo XML compactado para almacenar la información y en la actualidad se almacena un archivo compactado que contiene varios archivos en formato SHP de ESRI. Es necesario señalar que aún coexisten dispositivos que trabajan utilizando los dos formatos. Estos archivos compactados son accesibles mediante un sitio FTP seguro y su ubicación se obtiene desde un índice en la base de datos de la plataforma.

Como parte del proyecto *Geoloc*, para la creación de la *Plataforma Nacional de soporte espacial para servicios de geolocalización con actualización continua* se estableció un caso de estudio dedicado a la migración y análisis de trazas acumuladas durante los últimos dos años en la plataforma *MovilWeb* a un entorno de *Big Data*. Para desarrollar este experimento se utilizó el proyecto *GeoDocker* y se creó un clúster de *GeoMesa* en un servidor de nuestra Intranet. A partir de tutoriales existentes en Internet (Graser, 2017) y de la experiencia anterior en el trabajo con el paquete *GeoTools* se desarrolló una aplicación en *Java* que lee las trayectorias desde el servicio FTP de la plataforma, en cualquiera de los formatos existentes y las inserta en una tabla geoespacial en el clúster.

Una vez insertadas las trazas GPS en el clúster de *Big Data* se configuró el *Geoserver* para publicar las posiciones de las trazas en forma de mapas. Los resultados obtenidos fueron satisfactorios tanto para los tiempos de inserción de los datos como para los tiempos de respuesta a solicitudes de mapas y consultas utilizando dicha interfaz para seleccionar los datos de una trayectoria en la tabla.

Además de mejorarse los tiempos de respuesta con estas tecnologías respecto a la solución existente se habilitan un grupo de análisis que hoy no es posible poner a disposición de los usuarios por la complejidad de su implementación. Se habilita también la utilización de técnicas de minería de datos para buscar patrones y extraer información de los datos primarios.

### **Almacenamiento de resultados de levantamientos realizados con VANT**

Los VANT tienen una amplia gama de aplicaciones y modelos. Se clasifican en tres categorías: Control de seguridad, Investigación científica y Aplicaciones comerciales. Sin embargo, para lograr una aplicación de VANT bien diseñada, debe haber un soporte de información preciso, que es necesario para un sistema exitoso. Es bien sabido que las aplicaciones de VANT se han involucrado en muchas industrias que van desde la agricultura hasta la producción y el transporte de petróleo y gas (Mohammed, *et al.*, 2014).

Geocuba ha optado esta tecnología para realizar levantamientos de diferente índole. Actualmente varias empresas del Grupo Empresarial poseen VANT y ofrecen el servicio (Fernandez, *et al.*, s. f.). Como resultado del procesamiento del material primario se obtienen varios componentes, una ortoimagen georreferenciada, una nube de puntos, un modelo digital de elevación, etc. Todos estos resultados poseen gran tamaño y en la actualidad no existe una gestión adecuada de los mismos que favorezca su reutilización en otras aplicaciones. El proceso de producción de los sistemas basados en VANT, según recoge la bibliografía, está compuesto por cuatro momentos, adquisición, procesamiento, administración y disseminación (Percivall, *et al.*, 2015). En el momento de madurez en que se encuentra nuestra empresa los dos últimos procesos no están eficientemente implementados.

A partir de haber asimilado tres proyectos orientados a los datos ráster y de disponer en nuestra infraestructura de los escenarios para desarrollar las pruebas se evaluó la ingestión de ortofotos en formato *GeoTiff* utilizando los diferentes escenarios, *GeoMesa*, *GeoTrellis* y *MrGeo*.

Para apoyar la implementación de estos dos momentos se desarrolló un experimento utilizando el clúster de *GeoDocker* mencionado anteriormente. Se tomó una imagen en formato

*GeoTiff*, resultado de un vuelo con VANT y se implementó en *Java* una aplicación para introducirla en el clúster. En este caso no contamos con ejemplos previos públicos en línea que pudiéramos reutilizar o de los cuales tomar experiencias, solo las pruebas unitarias de algunas clases del proyecto *GeoMesa*.

El proceso de ingestión de imágenes ráster en *GeoMesa* parte de realizar un re-teselado para generar múltiples imágenes más pequeñas que sean más fáciles de gestionar por el clúster. En este caso se generaron cuatro niveles de visualización de mosaicos de 256 por 256 píxeles en formato *GeoTiff*. Este proceso de generación de los mosaicos se realiza utilizando la herramienta *gdal\_retile* asociada al proyecto *gdal* (GDAL, 2012).

En el caso de *GeoTrellis* se utilizó el proyecto *GeoPySpark* para realizar las pruebas, para lo que se preparó un escenario en *Python 3.6*. El proceso de ingestión de las imágenes con *GeoTrellis* es más simple ya que no es necesario el proceso de crear los mosaicos antes de la ingestión. Además, *GeoTrellis* facilita el proceso de creación de las pirámides de mosaicos para los diferentes niveles de visualización.

Para el caso del servicio de visualización TMS que proporciona *GeoTrellis* se logró configurar para acceder a capas de una sola banda, el proceso no fue exitoso para ortofotos de tres o más bandas. No encontramos en la documentación información sobre este problema.

Para desarrollar los experimentos con *MrGeo* se utilizó tanto las facilidades del proyecto para ser utilizado desde la línea de comandos como desde el proyecto *PyMrGeo*. Se desarro-



Figura 1: Imagen resultado de visualizar una ortofoto

llaron pruebas para diferentes tipos de imágenes que gestiona Geocuba, en este caso imágenes multiespectrales *Landsat*, modelos de elevación *ASTER* y ortofotos producto de levantamientos con VANT.

Al igual que *GeoTrellis*, *MrGeo* facilita la ingestión de las imágenes y la creación de las pirámides de visualización. En su caso se pudo completar el ciclo completo desde la ingestión hasta la publicación de las imágenes.

Un elemento importante de *MrGeo* es que el servicio WMS que ofrece no está limitado a un archivo de configuración y por medio del mismo es posible acceder a todas las imágenes del clúster.

## CONCLUSIONES

Durante el proceso de evaluación fueron identificadas fortalezas y debilidades que nos permiten definir estrategias para utilización de estas soluciones en distintos escenarios. Para el caso de los datos vectoriales el producto *GeoMesa* puede ser una solución adecuada ya que garantiza un ciclo completo de gestión de los datos, desde su almacenamiento hasta su publicación en web desde *GeoServer*. Esta publicación en web desde *GeoServer* garantiza la utilización de estándares de estilos de mapas garantizando una mejor explotación de las potencialidades del servicio.

En cuanto a los datos ráster hay dos escenarios donde utilizar las tecnologías evaluadas. Para el caso de gestión de imágenes satelitales u ortofotos resultantes de levantamientos con VANT la plataforma que ofrece *MrGeo* es más adecuada teniendo en cuenta las facilidades de servicio WMS para la visualización de imágenes de múltiples bandas. En aquellos escenarios donde se generan datos ráster a partir de sensores se puede utilizar *GeoTrellis* teniendo en cuenta los múltiples métodos de interpolación y visualización que ofrece.

## REFERENCIAS

- Abreus, Y. A. (2006). Tipos de fragmentación en bases de datos distribuidas. *Ingeniería Industrial*, 27(1), 3–6.
- Cugler, D. C., Oliver, D., Evans, M. R., Shekhar, S., & Medeiros, C. B. (2013). Spatial big data: Platforms, analytics, and science. *GeoJournal*. <https://pdfs.semanticscholar.org/c64e/b7f733cf78573e962c6b5df24860eed3aabe.pdf>
- de la Beaujardiere, J. (2003). *OGC Implementation Specification OGC 01-068r3: OpenGIS Web Map Service*. Open Geospatial Consortium. <http://www.opengeospatial.org/standards/wms>
- Fernandez, P. M., Martínez, D. O., García, E. P., & Lorenzo, F. B. (s. f.). *Vehículos aéreos no tripulados (vant) en Cuba, aplicados a la Geomática. Estado actual, perspectivas y desarrollo*.
- Furht, B., & Villanustre, F. (2016). Introduction to big data. En *Big Data technologies and applications* (pp. 3–11). Springer.
- GDAL (2012). *GDAL - Geospatial Data Abstraction Library*. <http://www.gdal.org/>
- González Suarez, G., Capote Fernández, J. L., Cruz Iglesias, R., Batule Dominguez, M., Farre Rosales, D., & Fernández Águila, L. (2018, marzo 15). Movilweb sistema de control de flotas,

- estado actual y perspectivas. *Informática Habana 2018*. X Congreso Internacional de Geomática 2018, La Habana, Cuba. <http://www.informaticahabana.cu/es/node/4495>
- Graser, A. (2017, agosto 27). *Getting started with GeoMesa using Geodocker | Free and Open Source GIS Ramblings*. <https://anitagraser.com/2017/08/27/getting-started-with-geomesa-using-geodocker/>
- Hughes, J. N., Annex, A., Eichelberger, C. N., Fox, A., Hulbert, A., & Ronquest, M. (2015). Geomesa: A distributed architecture for spatio-temporal fusion. *Geospatial Informatics, Fusion, and Motion Video Analytics V*, 9473, 94730F.
- Kim, G., Humble, J., Debois, P., & Willis, J. (2015). *The DevOps Handbook*. IT Revolution.
- Locationtech Labs. (2018). *GeoTrellis for PySpark*. Contribute to locationtech-labs/geopyspark development by creating an account on GitHub [Python]. LocationTech Labs. <https://github.com/locationtech-labs/geopyspark> (Original work published 2016)
- McCoy, M. D. (2017). Geospatial Big Data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science*, 84, 74–94.
- Mohammed, F., Idries, A., Mohamed, N., Al-Jaroodi, J., & Jawhar, I. (2014). UAVs for smart cities: Opportunities and challenges. *Unmanned Aircraft Systems (ICUAS)*, 2014 International Conference on, 267–273.
- Moreno, C. S., Valenti, H. R., Rubio, D. M., & Gonzalez, C. J. (2016). *Experiencias de Implementación de Despliegue Continuo con una Infraestructura “Conteinerizada”*.
- NGA. (2015). *Home • ngageoint/mrgeo Wiki • GitHub*. <https://github.com/ngageoint/mrgeo/wiki>
- OSGeo. (2018). *Geotrellis*. Geotrellis. <https://www.osgeo.org/projects/geotrellis/>
- Percivall, G., Reichardt, M., & Taylor, T. (2015). Common Approach To Geoprocessing of Uav Data Across Application Domains. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(1), 275.
- Schut, P. (2007). *OGC Implementation Specification 05-007r7: OpenGIS Web Processing Service*.
- Shekhar, S. (2012). Spatial big data challenges. *Keynote at ARO/NSF Workshop on Big Data at Large: Applications and Algorithms, Durham, NC*. <http://www-users.cs.umn.edu/~shekhar/talk/2012/12.11.sbd.acmgis.pdf>
- Vretanos, A. (2005). *OGC Implementation Specification 04-094: Web Feature Service Implementation Specification*. Technical report. Open Geospatial Consortium. <http://www.opengeospatial.org/standards/wfs>
- Willis, J. (2012). The convergence of DevOps. *DevOps Blog*.
- Yao, X., Mokbel, M. F., Alarabi, L., Eldawy, A., Yang, J., Yun, W., Li, L., Ye, S., & Zhu, D. (2017). Spatial coding-based approach for partitioning big spatial data in Hadoop. *Computers & Geosciences*, 106, 60–67.

