

## LAGOS, SILOS Y CONSULTAS DE GRANDES VOLÚMENES DE DATOS

### ¿Sabes qué hacer con la gran cantidad de datos en tu empresa?

SANTIAGO RODRIGUEZ, IVÁN RIVERA.  
GERENTE DE BIG DATA, COORDINADOR DE BIG DATA.

En promedio las empresas llegan a duplicar su volumen de datos cada dos años, debido a la información generada por sistemas de gestión tradicionales como nómina, ERP y CRM, adicionalmente la data obtenida de las redes sociales y ni qué decir de aquellas compañías que se han dejado enamorar por los beneficios del Internet de las cosas (IoT), y la gran extracción de información que ésta genera al conectar diferentes dispositivos dentro de una compañía.

De hecho, la gobernabilidad de datos, es un problema que actualmente se está originando desde diferentes fuentes como Devops, sistemas de Recursos Humanos, incluso aplicaciones SaaS, Office 365 y Dropbox, contando con muchas infraestructuras informativas separadas.

#### ¿Cómo optimizar la búsqueda con información separada o en islas?

Una forma óptima de realizar búsquedas en estos silos o islas es usar herramientas como *Elasticsearch*, que maneja una arquitectura abierta, y escalable para desarrollar capacidades de búsqueda empresarial.

*Elasticsearch*, se basa en el motor de búsqueda llamado *Apache Lucene*, *Elasticsearch* tiene un esquema distribuido flexible, que permite escalar horizontalmente y manejar miles de millones de registros en múltiples índices, agregando nuevos campos a un registro dinámicamente sin tener la rigidez de las tablas de lenguaje de consulta estructuradas tradicionales.

Además, la búsqueda es casi instantánea con registros disponibles dentro de aproximadamente un segundo de indexación.

Dadas las capacidades de búsqueda empresarial ofrecidas por *Elasticsearch*, surge una nueva pila de búsqueda empresarial denominada “ELK” (*Elasticsearch*, *Logstash*, *Kibana*), que proporciona un marco común para indexar un conjunto de datos empresariales distribuidos en silos dentro del lago de datos a través de una sola vista de búsqueda usando una pequeña fracción del tamaño real de los datos en sí.



## ¿Cómo extraer inteligente y ordenadamente los datos?

Geoff Tudor, gerente de servicios de nube en Panzura, empresa situada en Campbell, California dedicada a desarrollar productos para la nube, indica que las empresas buscan aprovechar la infraestructura de la nube para reducir costos y mejorar la productividad, dando sentido a los grandes volúmenes de datos a través de estructuras denominadas : lagos de datos (*data lake*), arquitectura de Big Data enfocada en almacenar datos no estructurados o semiestructurados en su forma original, en un único repositorio para su análisis.

## ¿Qué otros beneficios recibes de los lagos de datos?

Los lagos de datos hacen que sea más fácil adoptar un enfoque de "almacenar ahora, analizar más tarde", ya que hay muy poco esfuerzo involucrado en la ingesta de datos en el lago.

En la Gerencia de Big Data de PRAXIS, sabemos la importancia y el impacto que tiene contar con herramientas que te permitan tener a la mano datos que llegan a determinar el rumbo de la organización. Es por ello que contamos con consultores expertos en *Elasticsearch* que pueden ayudar a la organización, a diseñar las consultas óptimas para explotar sus grandes volúmenes de datos independientemente de la forma en que se haya integrado el lago de datos a consultar.

Actualmente contamos con clientes de diversas industrias (banca, manufactura, retail y automotriz) a quienes hemos ayudado a agilizar sus operaciones diarias por medio de una arquitectura de limpieza de datos (*data governance*) y una búsqueda eficiente de los mismos (*Elasticsearch*), éstas tecnologías han causado un impacto de optimización en la procesos de los servicios de búsqueda de datos o estadísticas internas de las compañías.



**SANTIAGO RODRÍGUEZ MACÍAS**  
*Gerente de Big Data*

☎ Tel. 55 50800048 Ext. 2449  
✉ [bigdata@praxisglobe.com](mailto:bigdata@praxisglobe.com)  
🏠 Santiago Rodríguez Macías

## BIBLIOGRAFÍA:

Gutiérrez, D. (2020). The Power of Data Is Stuck In Silos - insideBIGDATA. Recuperado el 11 de enero de 2020, de <https://insidebigdata.com/2019/01/11/power-data-stuck-silos/>

