
CARTO Big Data connectors

Versión 1.0.0

Alberto Romeu Carrasco

01 de octubre de 2017

Contents:

1. Resumen ejecutivo	1
2. Glosario	3
3. Indices and tables	5

Resumen ejecutivo

CARTO¹ es una plataforma de *Location Intelligence*² que permite transformar datos geoespaciales en resultados de negocio.

Actualmente, las organizaciones que utilizan **CARTO** como herramienta de análisis geoespacial tienen multitud de fuentes de información y aplicaciones ya instaladas que generan continuamente nuevos datos.

El principal valor de **CARTO** para estas organizaciones es el de conectarse con esas fuentes de información (*datos de CRM, ERPs, hojas de cálculo, archivos con contenido geoespacial, bases de datos relacionales, etc.*) a través de una interfaz sencilla e intuitiva, para generar nueva información de valor añadido para su negocio mediante análisis geoespaciales y visualizaciones.

En determinadas organizaciones, especialmente de tamaño medio o grande, ocurre que diversos equipos gestionan sus datos con sistemas de información heterogéneos que utilizan repositorios de datos tales como:

- Hojas de cálculo y archivos CSV
- Documentos de Google Drive
- CRMs tales como Salesforce
- Herramientas de marketing digital como Mailchimp
- Servidores de datos geoespaciales como ArcGIS
- Bases de datos operacionales (relacionales o NoSQL)
- Sistemas de ficheros distribuidos HDFS
- Otros sistemas (Twitter, Dropbox, Instagram, etc.)

Estas organizaciones utilizan **CARTO** para importar sus datos y analizar la información que generan los distintos departamentos de manera conjunta, respondiendo a sus preguntas de negocio y encontrando además respuesta a otras preguntas que no se habían planteado en un principio.

¹ <https://www.carto.com>

² *Location Intelligence* Añadir directamente aquí la definición

CARTO cuenta con la posibilidad de importar datos desde diversas fuentes de datos, pero carece de soporte nativo para conectar a sistemas de almacenamiento masivo de datos usados generalmente para almacenar datos operacionales y resultados agregados obtenidos por los departamentos de *data science*.

El objetivo de este trabajo final de máster consiste en el desarrollo de conectores para **CARTO** que permitan incluir en los cuadros de mandos (*dashboards*), información proveniente de sistemas de almacenamiento masivo tales como: HDFS, Hive, Impala, MongoDB, Cassandra, BigQuery u otros.

Palabras clave: *BASH, Docker, Vagrant, Location Intelligence, AWS, HDFS, Hadoop, BigQuery, Hive, Impala, Spark, NoSQL, Cassandra, MongoDB, CARTO, dashboards, análisis geoespacial*

2.1 CRM

2.2 CSV

2.3 ERP

2.4 HDFS

2.5 Location Intelligence

2.6 NoSQL

CAPÍTULO 3

Indices and tables

- `genindex`
- `modindex`
- `search`