

Spark en Crossdata



¿Qué es Crossdata?

Crossdata es un producto de Stratio que tiene como **finalidad abstraer al usuario de las particularidades del lenguaje al consultar bases de datos.**

Por ejemplo: tanto si los datos están en Parquet, MongoDB, Postgres etc., puedes consultar dichos datos con un lenguaje SQL, y Crossdata se encargará de traducir esa query al lenguaje correspondiente.

Crossdata ofrece un **driver JDBC/ODBC** para que ciertas herramientas se enganchen a la capa de datos a través de él.

Arquitectura - modos de ejecución

Crossdata puede usarse como:

1. En **modo servidor**
2. Como **librería** sobre un cluster de Spark propio ya existente.

Arquitectura - catálogo

- El catálogo almacenará la **estructura de las tablas definidas en Crossdata**, su origen de datos, su base de datos, etc asignándoles un identificador único.
- Del mismo modo almacenará también las bases de datos definidas en crossdata.
- Compartido por todo el banco. Pero cada Sandbox tiene acceso a “su propio” catálogo de ZK
 - Es una forma de **compartición de catálogo entre los usuarios de un Sandbox**

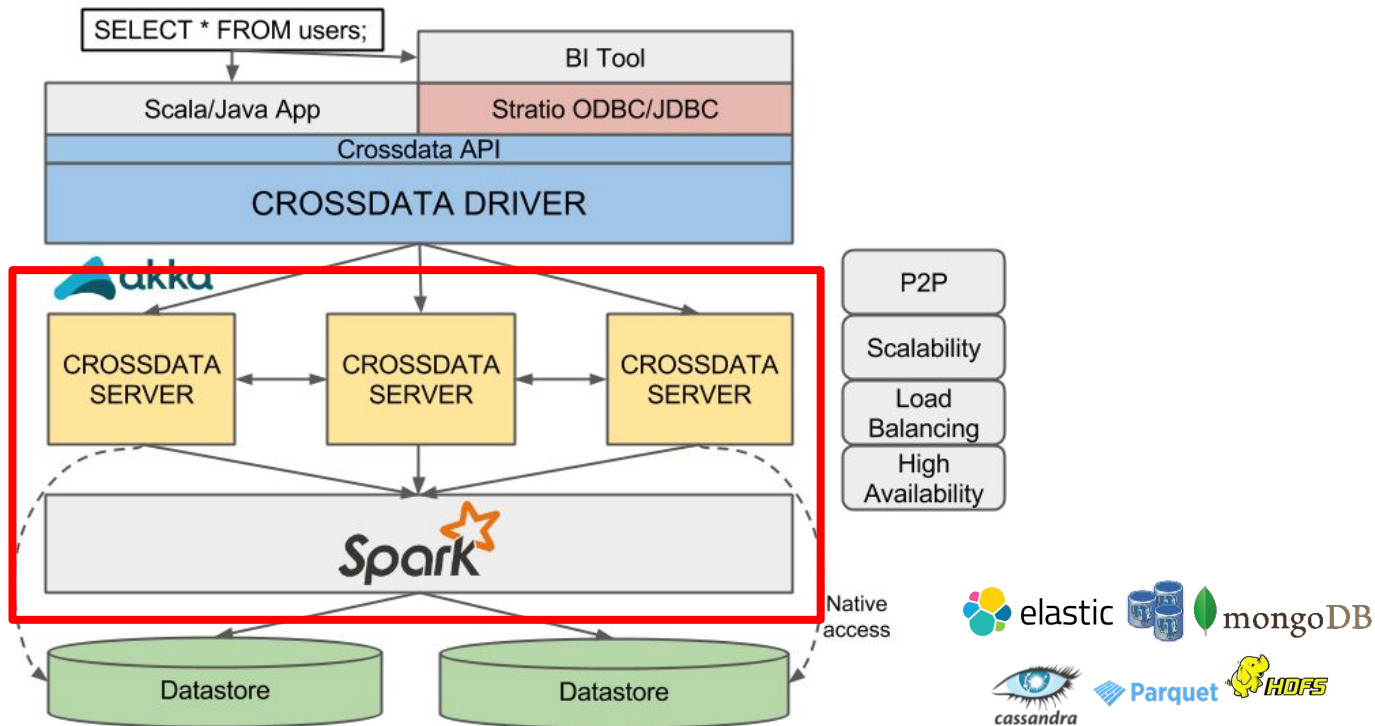
XDSession vs SparkSession

Diferencias

- **XDSession** envuelve la **SparkSession**.
- **Spark Session** tiene el catálogo en **memoria** de Spark por lo tanto, los demás usuarios del SandBox **no pueden acceder** a él. Se borra al finalizar la sesión.
- **XDSession** tiene el catálogo compartido para todos los usuarios de SandBox y por tanto todos ellos pueden acceder a él.

Arquitectura - modos de ejecución

Modo Servidor



NOTEBOOK

Conceptos básicos de cómo
utilizar Crossdata en Spark

Spark en Crossdata

Spark en Crossdata

Thanks