



Procesamiento de datos con Apache Spark

www.datapath.ai

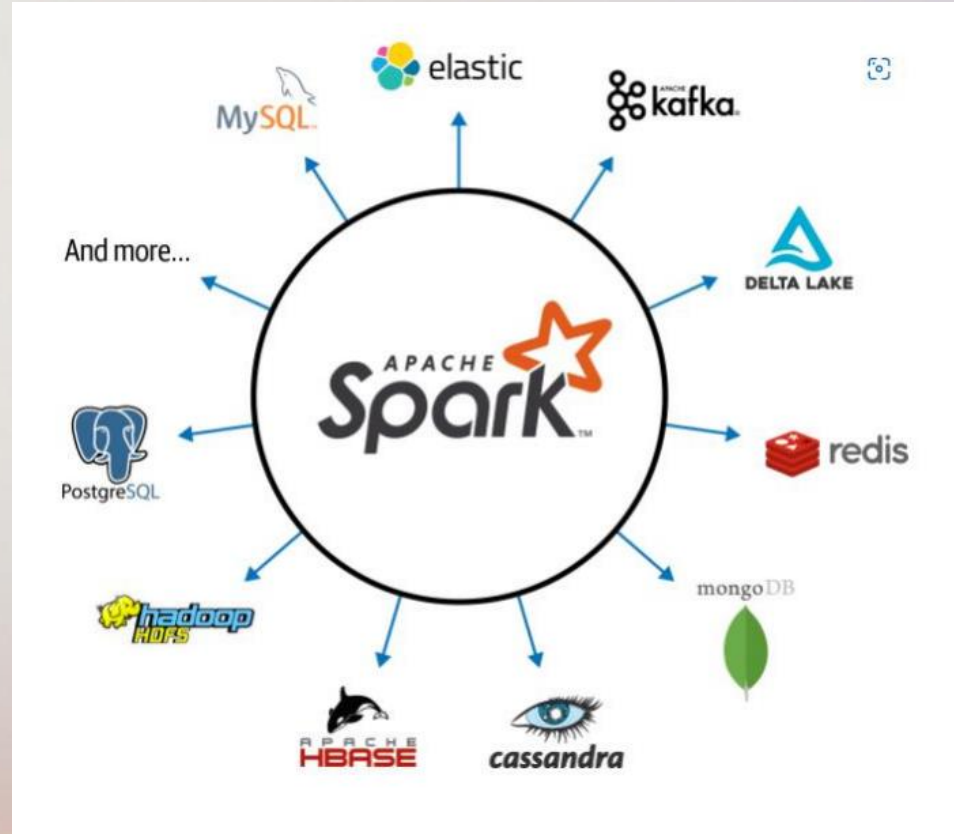
Objetivos de la clase

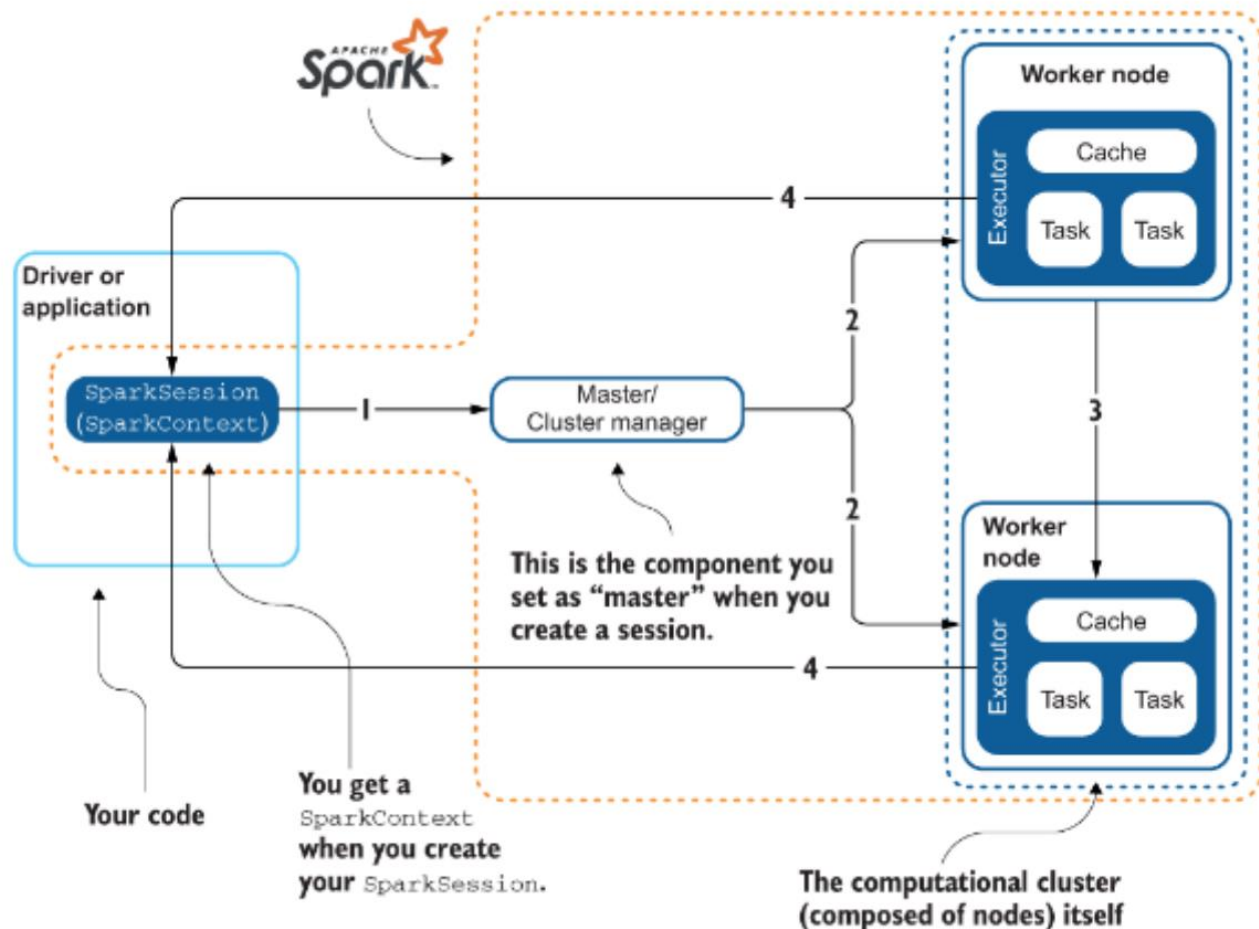
1. Comprender como trabajar con Apache Spark
2. Conocer como ver los logs de Spark en Databricks
3. Revisar los tipo de cluster para Databricks
4. Trabajar con Python/Scala en Spark
5. Trabajar con Apache Spark
6. Test Unit con Spark

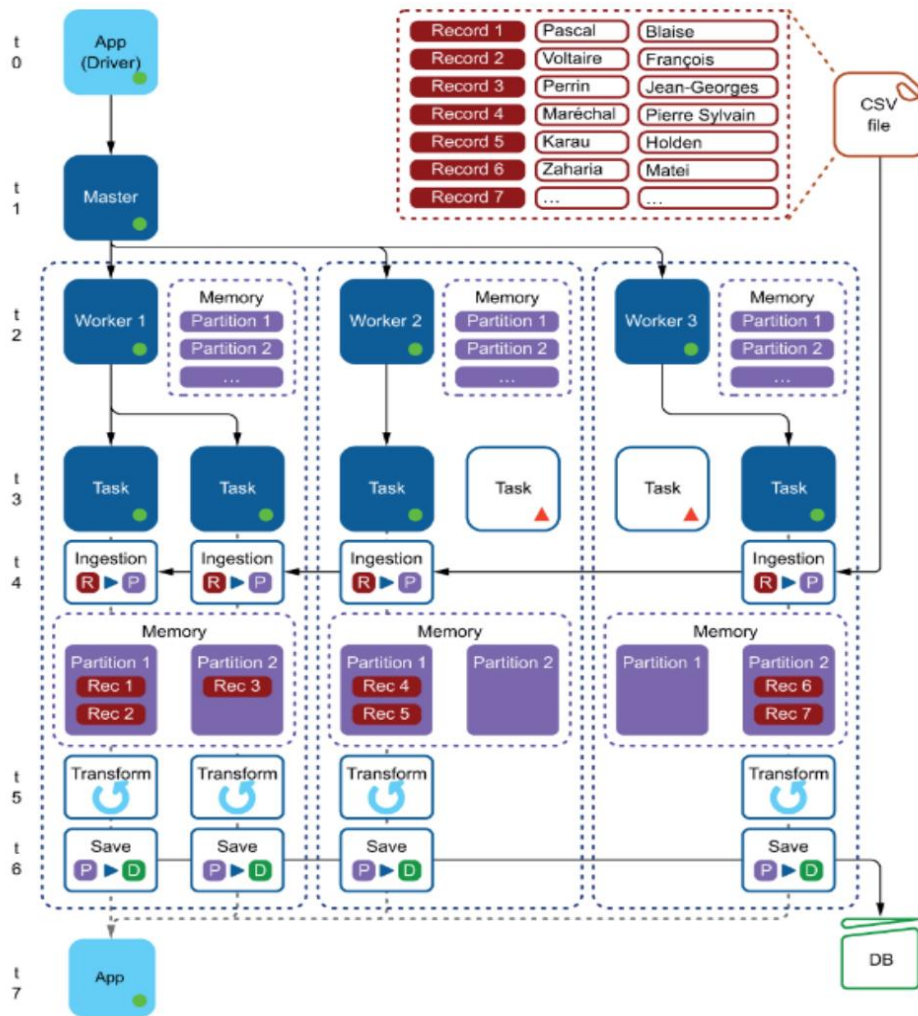
Let's gooo !!



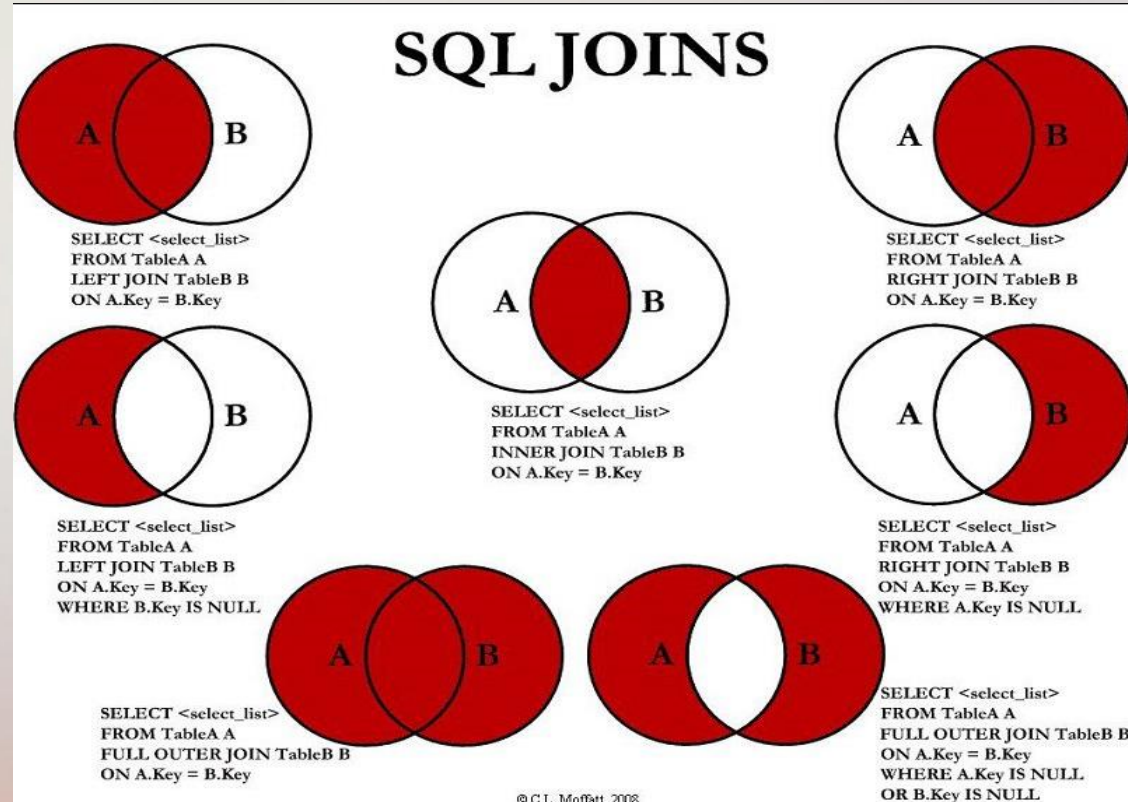
Spark connectors







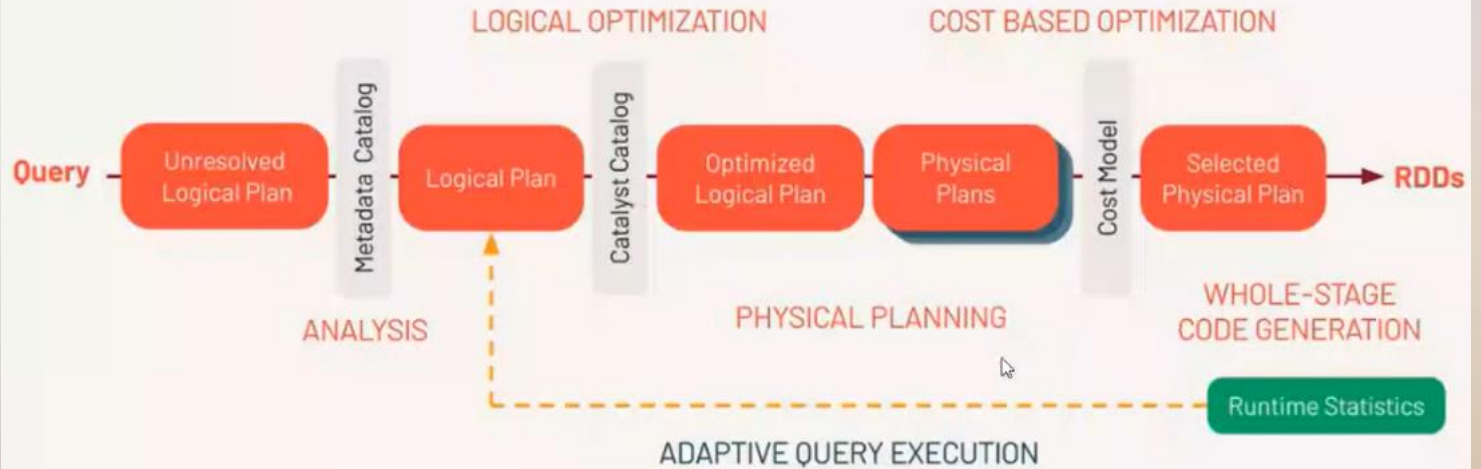
Type of Joins



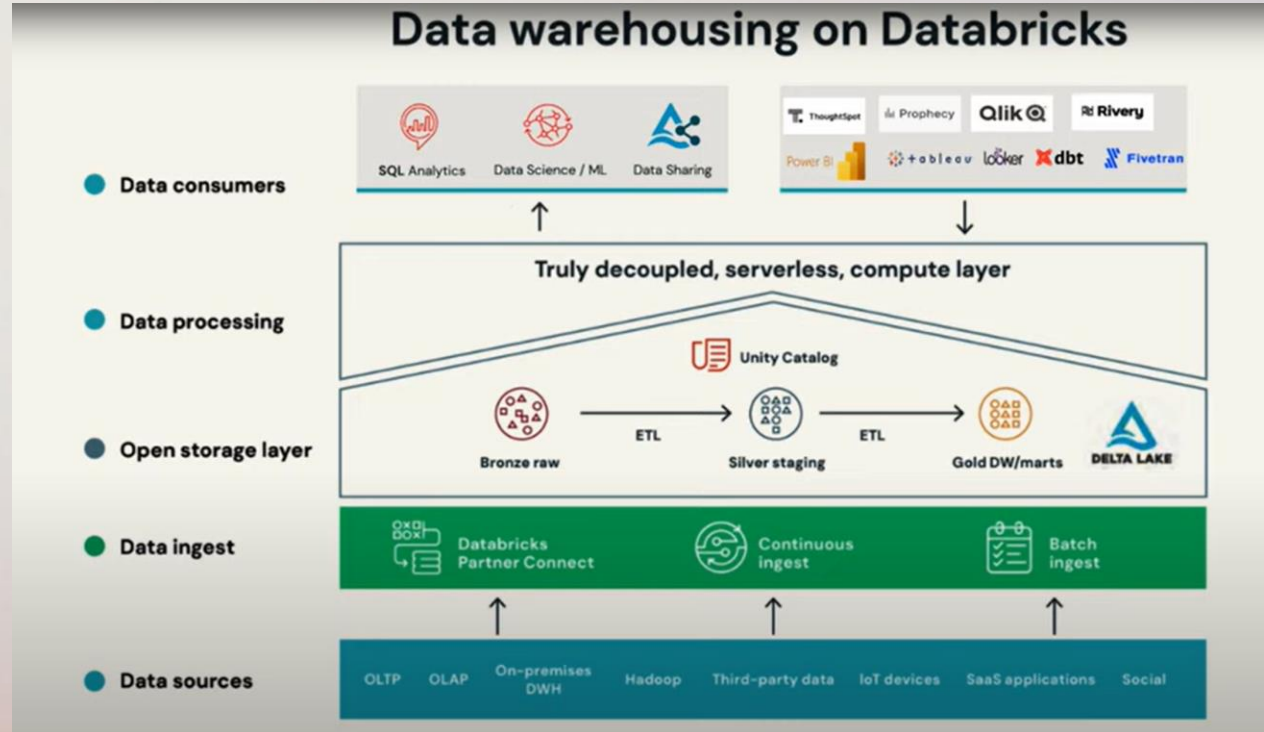
Deploy

Query Optimization with AQE

New in Spark 3.0, **disabled** by default - recommended to enable



Review Databricks



Cluster en Databricks

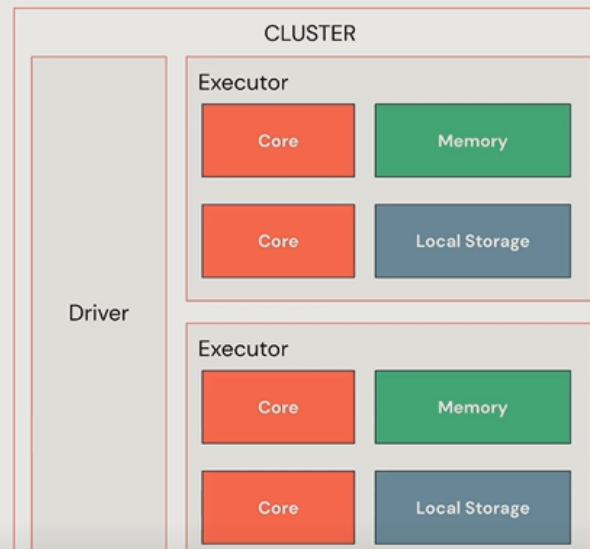
Clusters

Overview

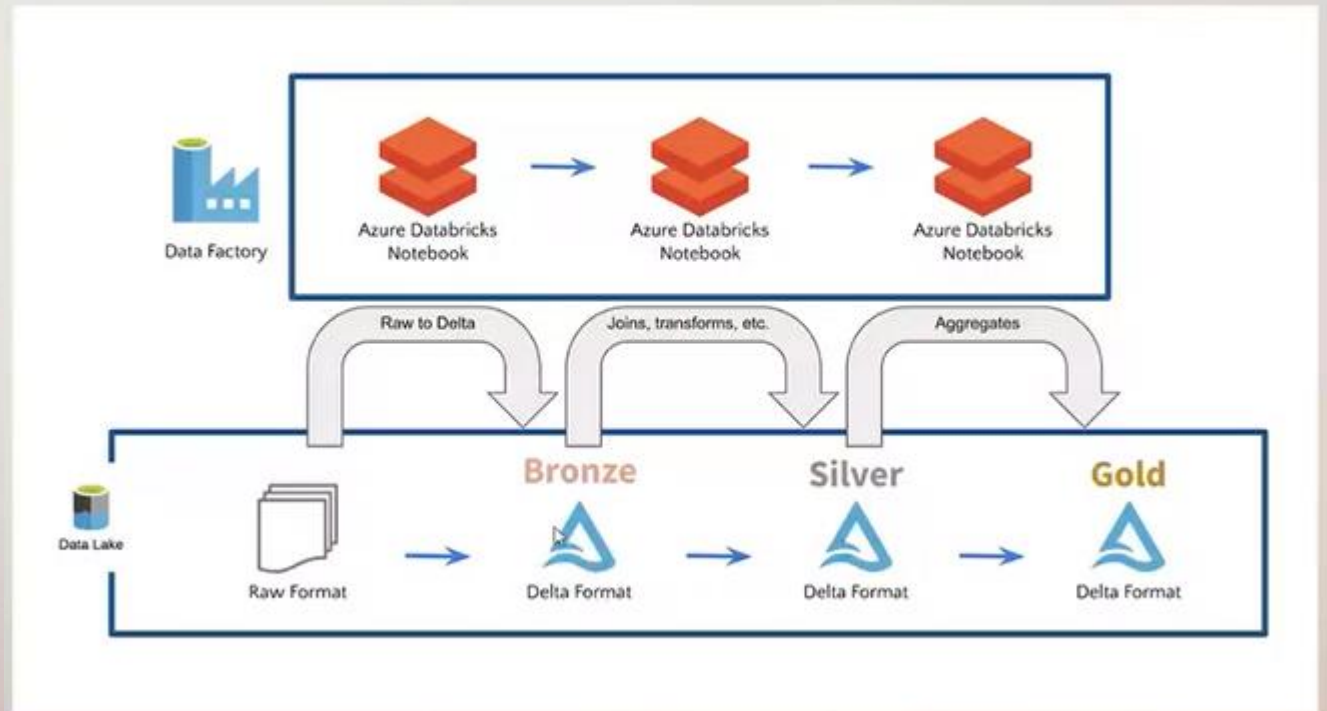
Clusters are made up of one or more virtual machine (VM) instances

Driver coordinates activities of executors

Executors run tasks composing a Spark job



Caso de uso



¿PREGUNTAS?

Aprende, aplica y crece

www.datapath.ai