

#### Spark con Python

Ricardo Jaume

Introducción a Apache Spark









0

Características



0

Simple. Fast. Scalable. Unified.





Batch/streaming data



#### **SQL** analytics



Data science at scale



Machine learning

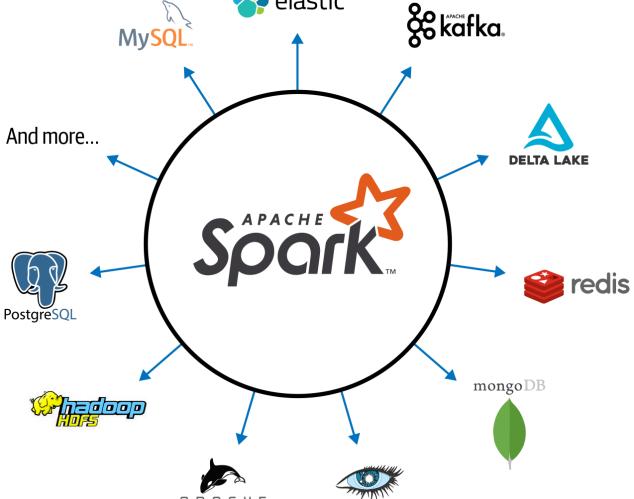
0

Arquitectura



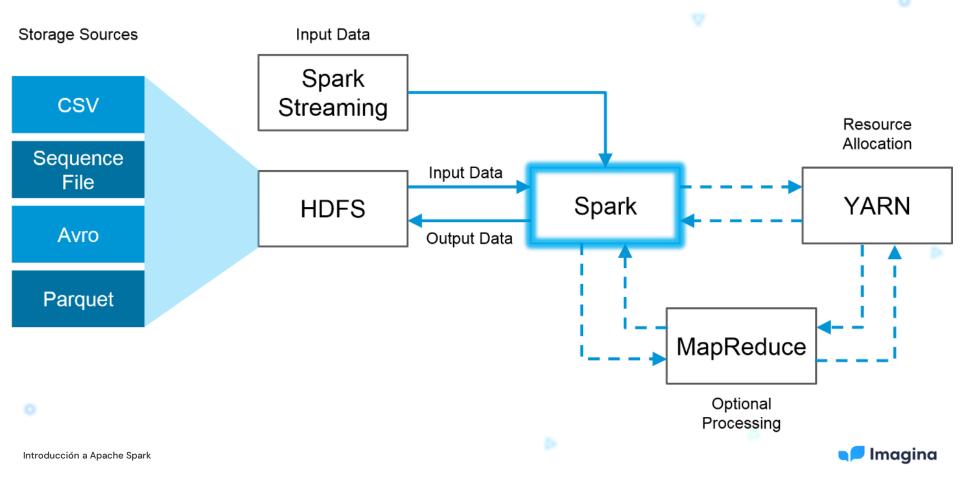
0

ARQUITECTURA Page elastic 2022



cassandra

ARQUITECTURA 2022



0

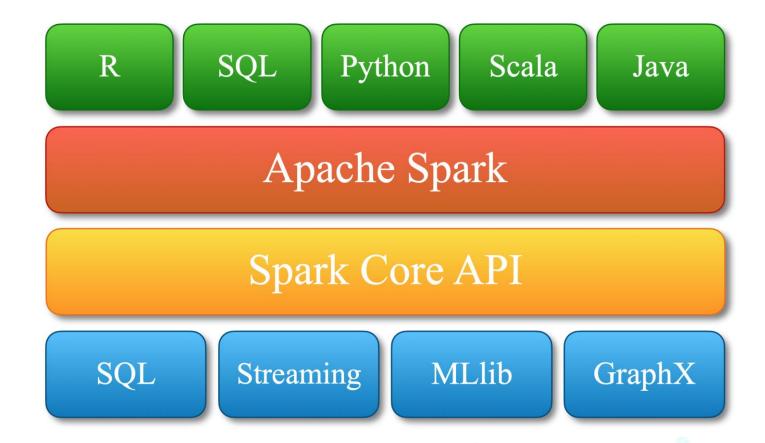
Componentes



D

0

COMPONENTES 2022





COMPONENTES 2022

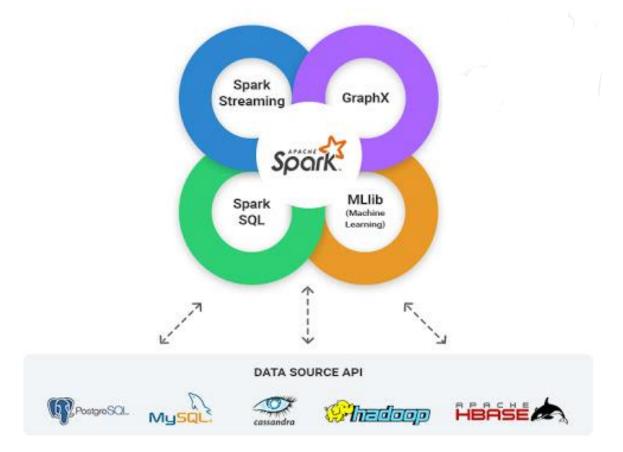
Spark SQL Spark SQL Mllib (aprendizaje automático)

GraphX (gráficos)

Spark Core



COMPONENTES 2022



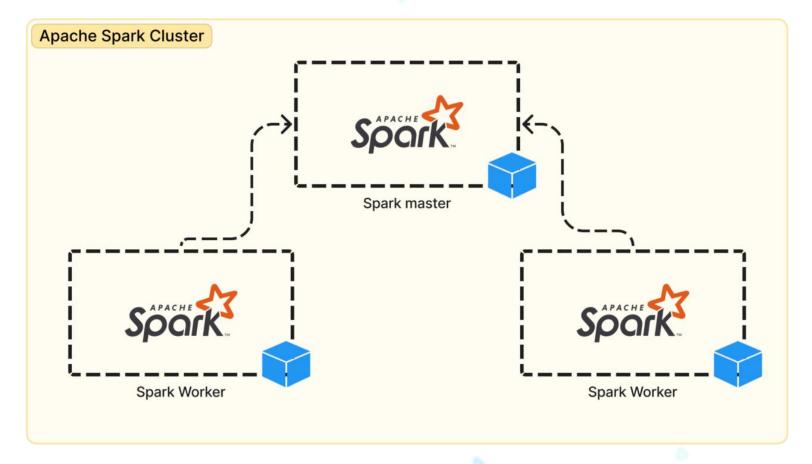


0

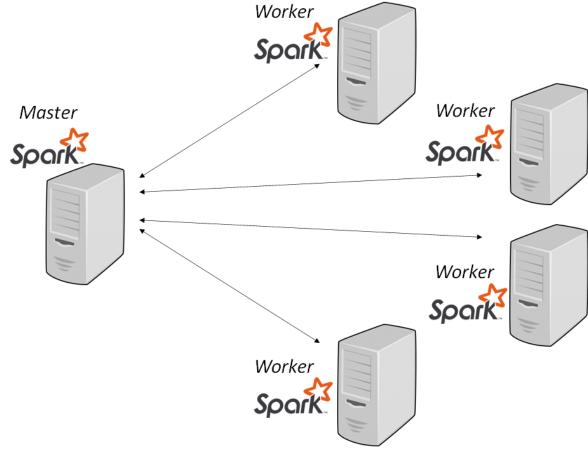
Clusterización

Δ

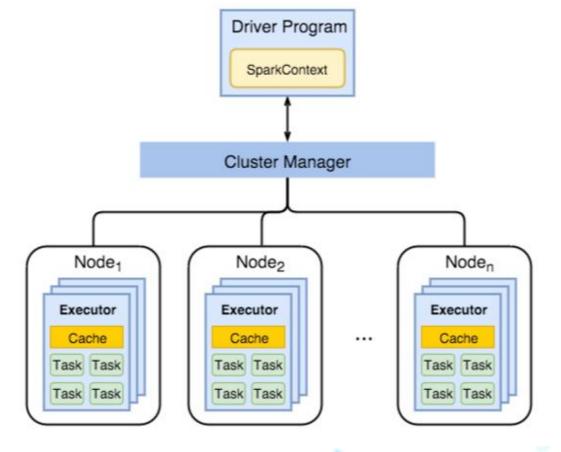
0

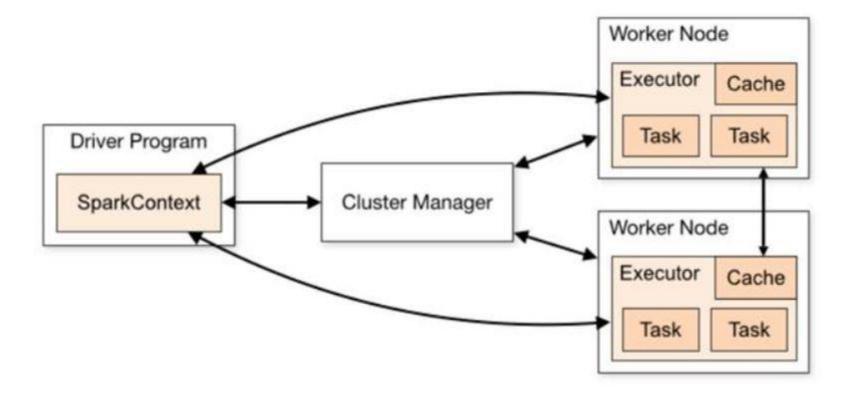












9

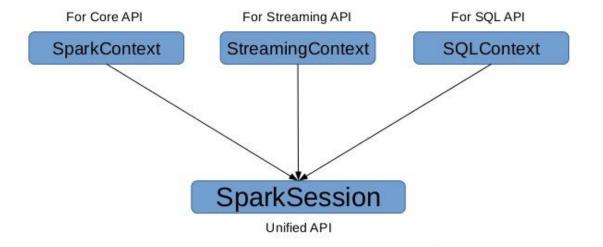
Desarrollo



D

0

DESARROLLO 2022



DESARROLLO 2022

**DataFrames & Datasets** 

**Batch** 

Streaming



