

# Spark con Python

Ricardo Jaume

Introducción a Apache Spark

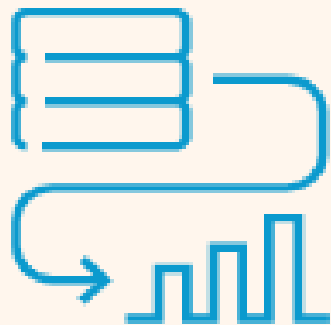


# Spark

## Características



**Simple.  
Fast.  
Scalable.  
Unified.**



# Batch/streaming data



# SQL analytics



**Data science at scale**

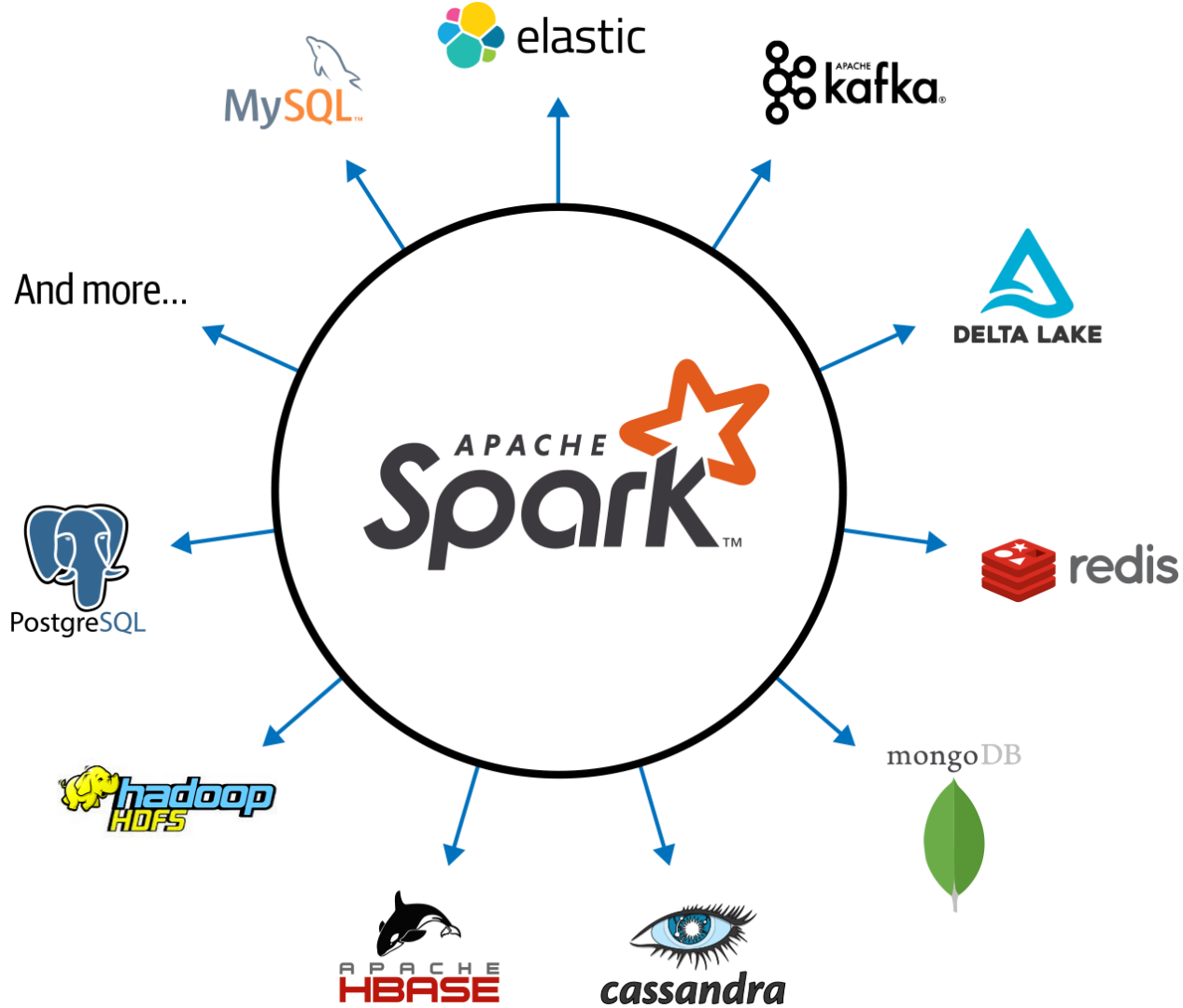


## Machine learning

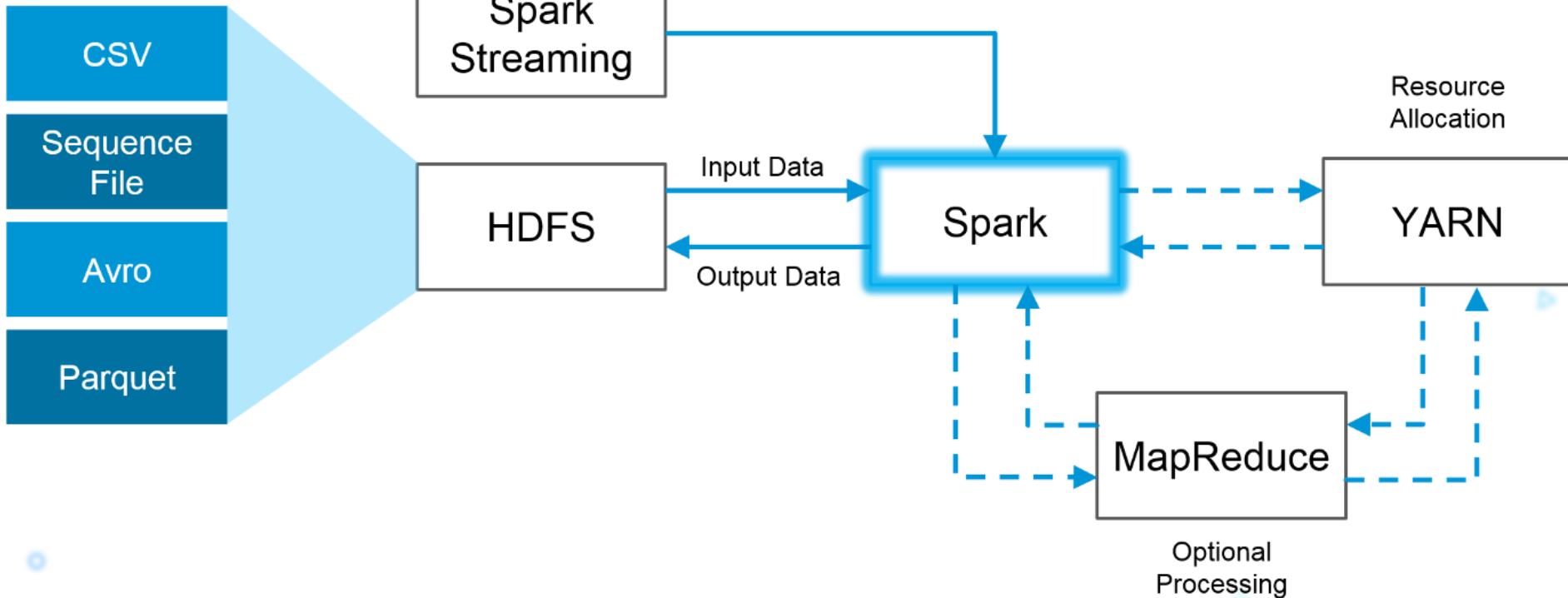


# Spark

## Arquitectura

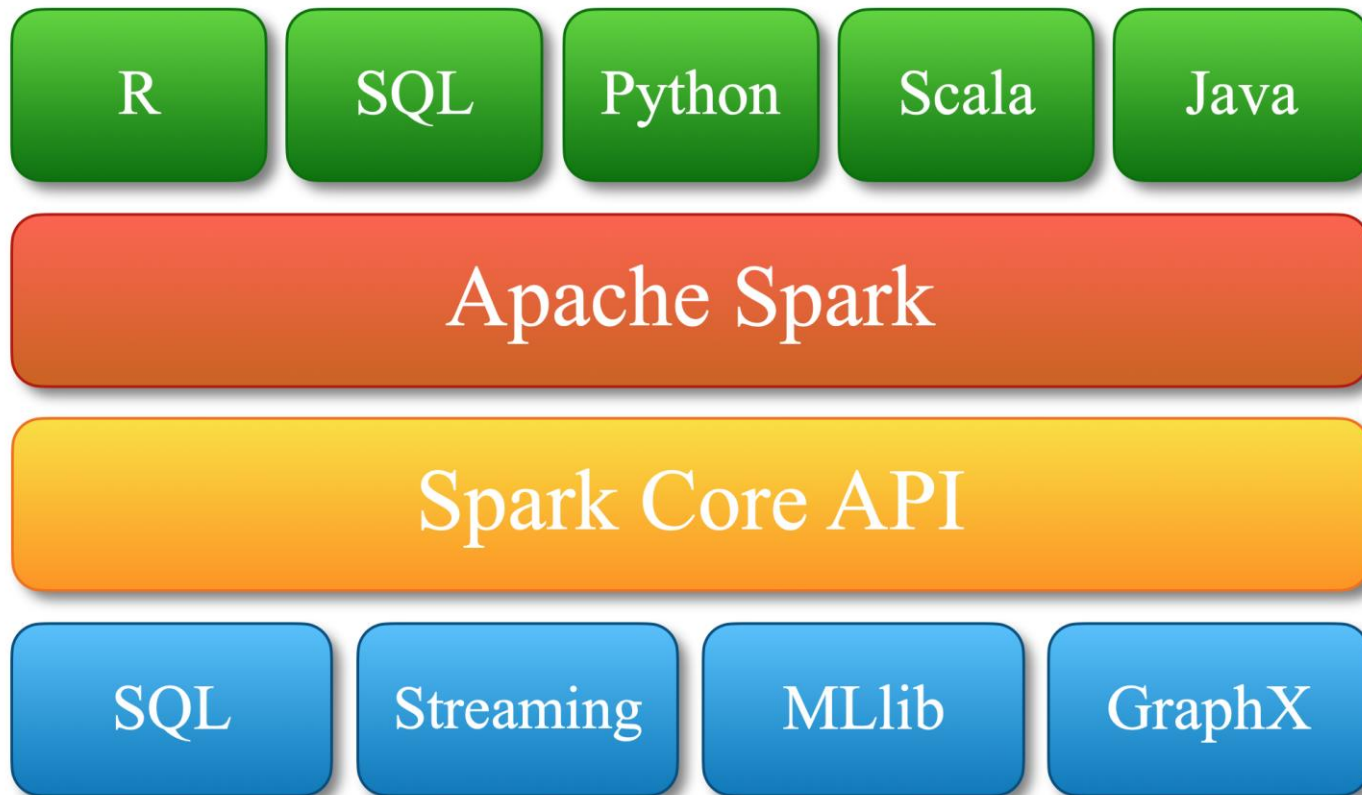


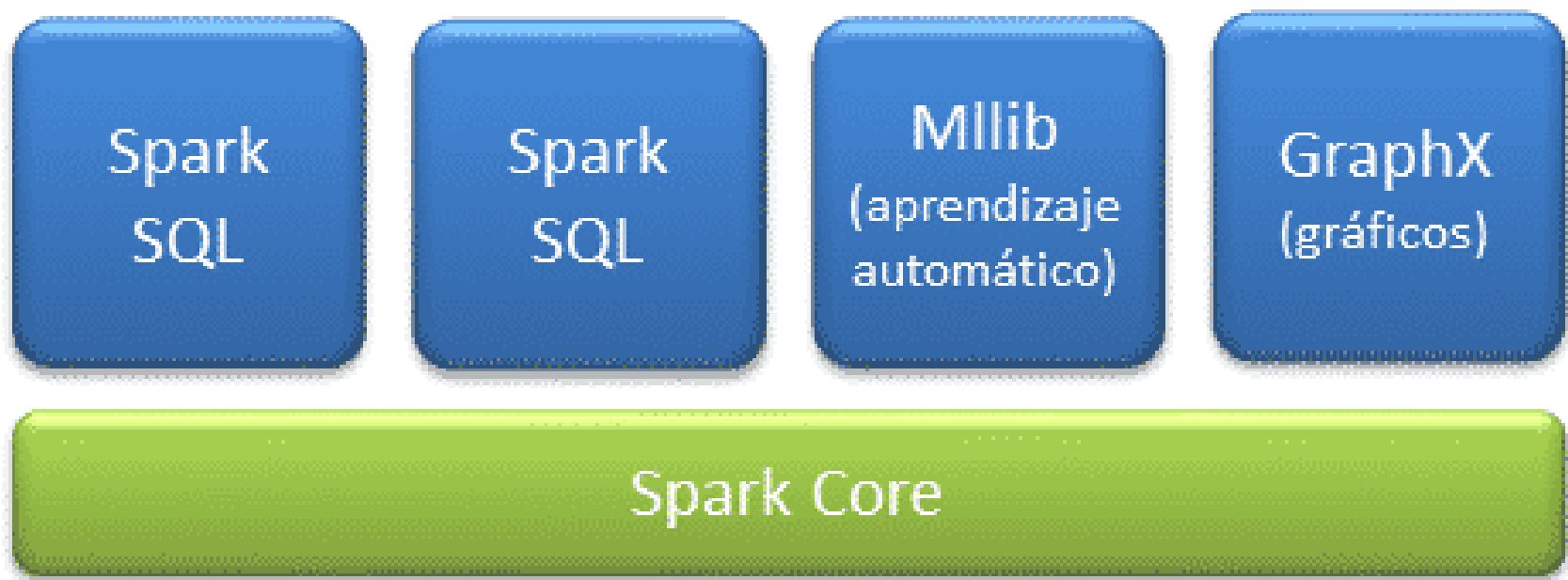
## Storage Sources

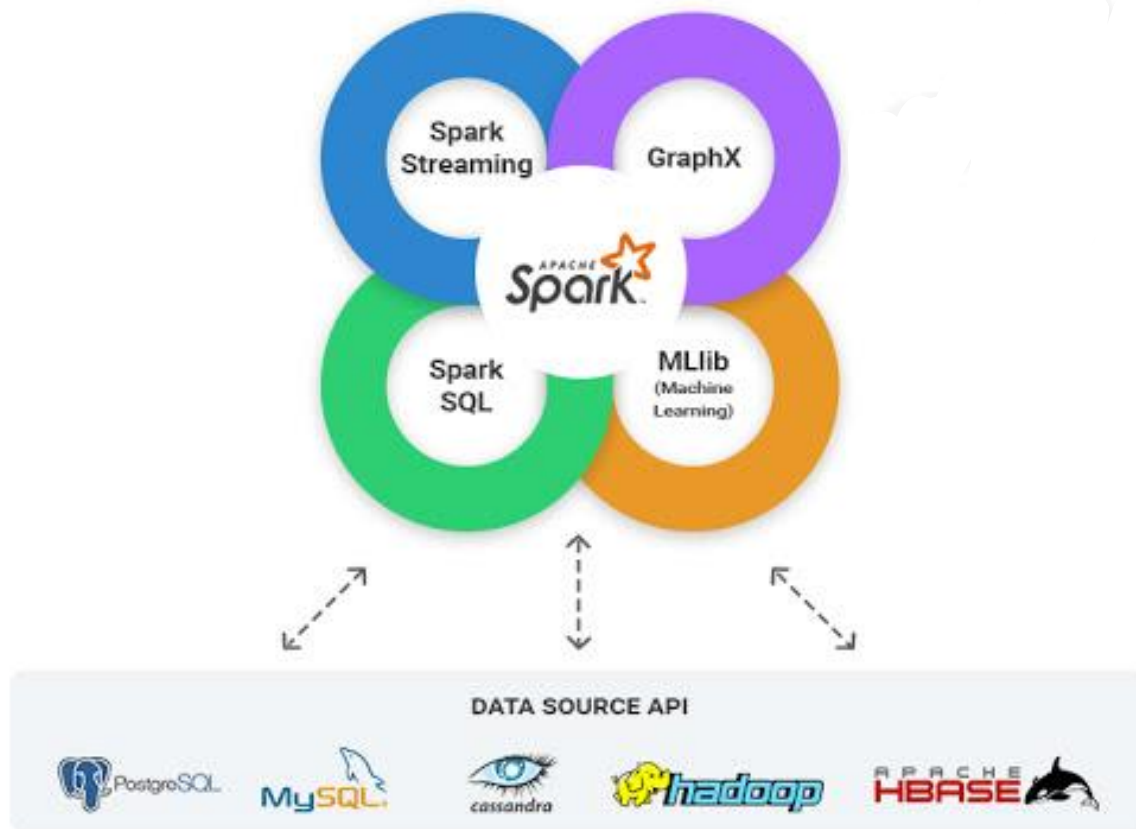


# Spark

## Componentes



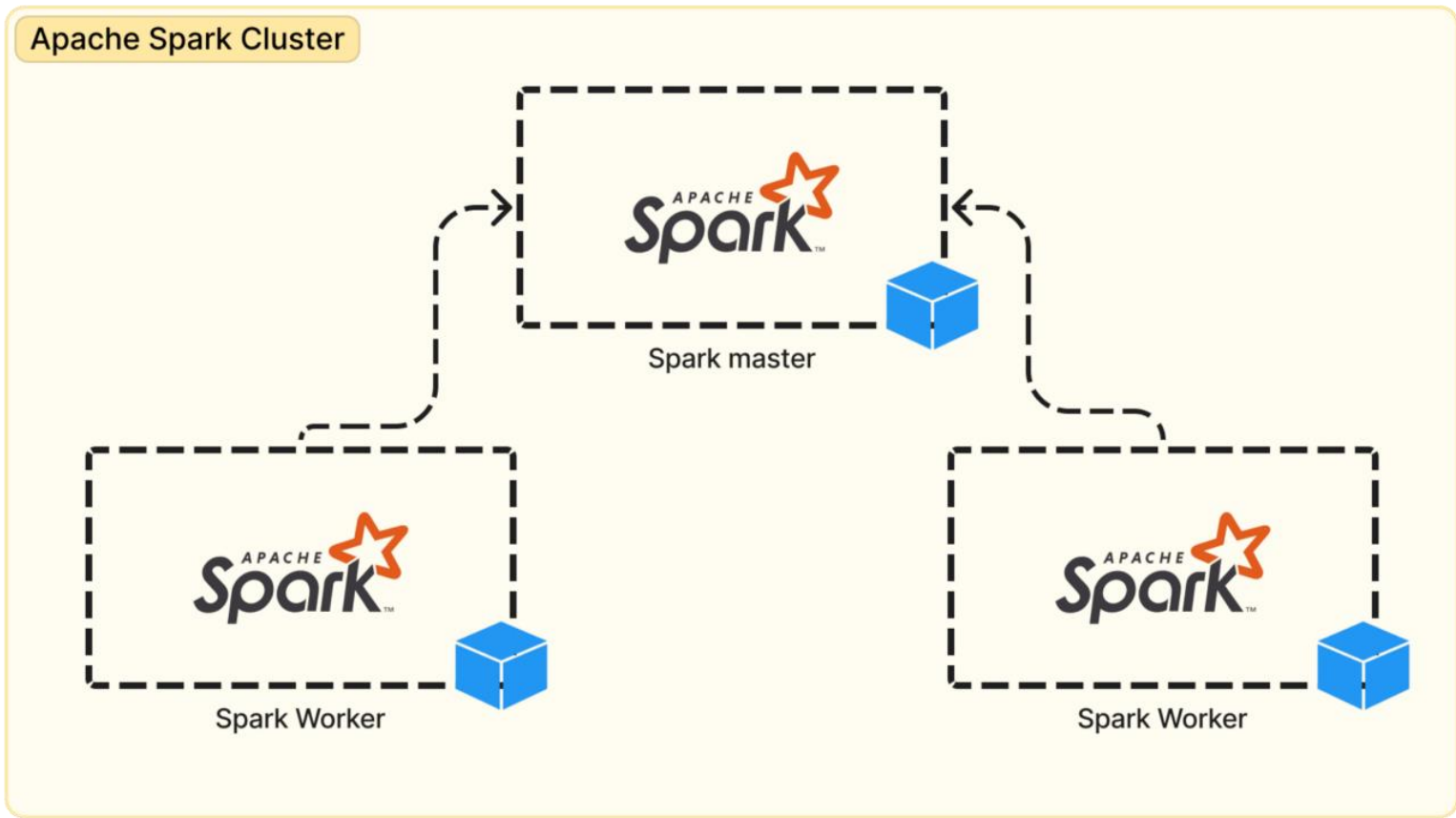


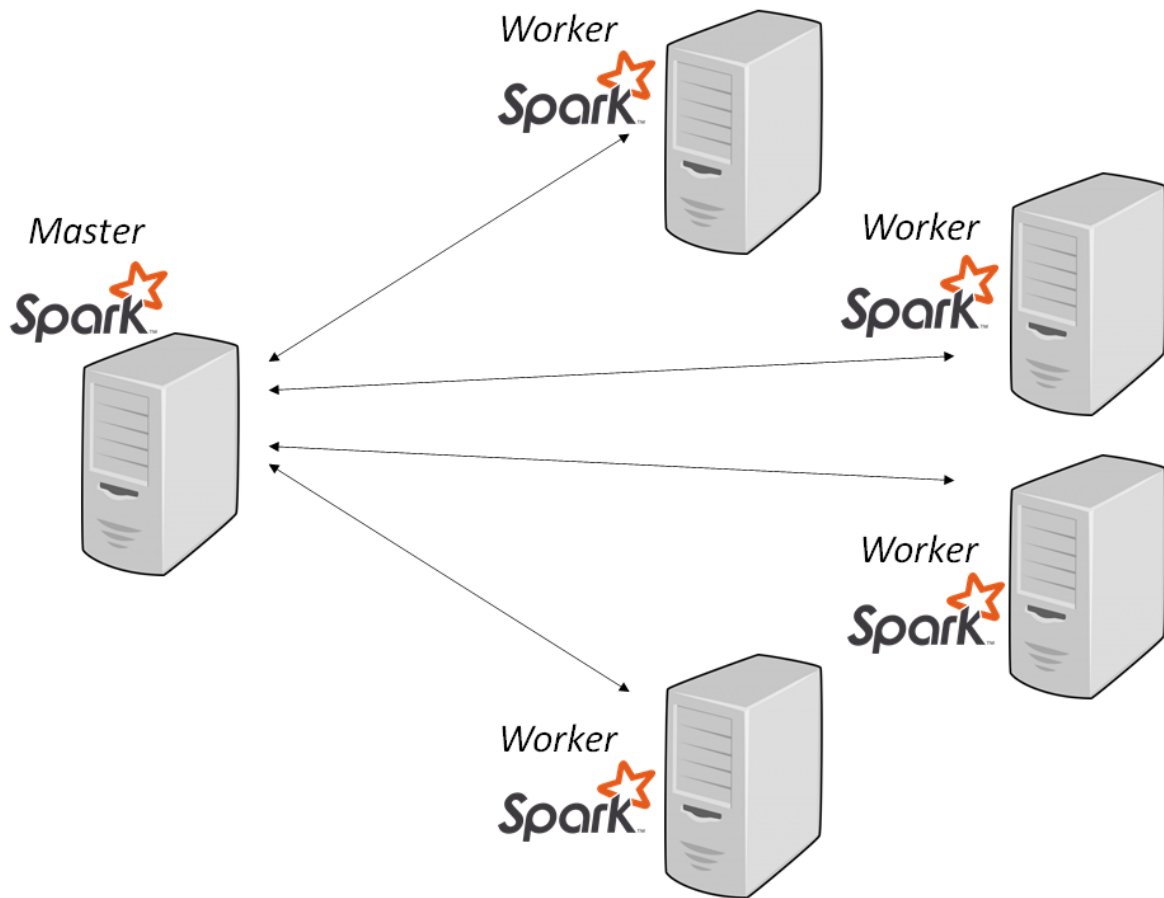


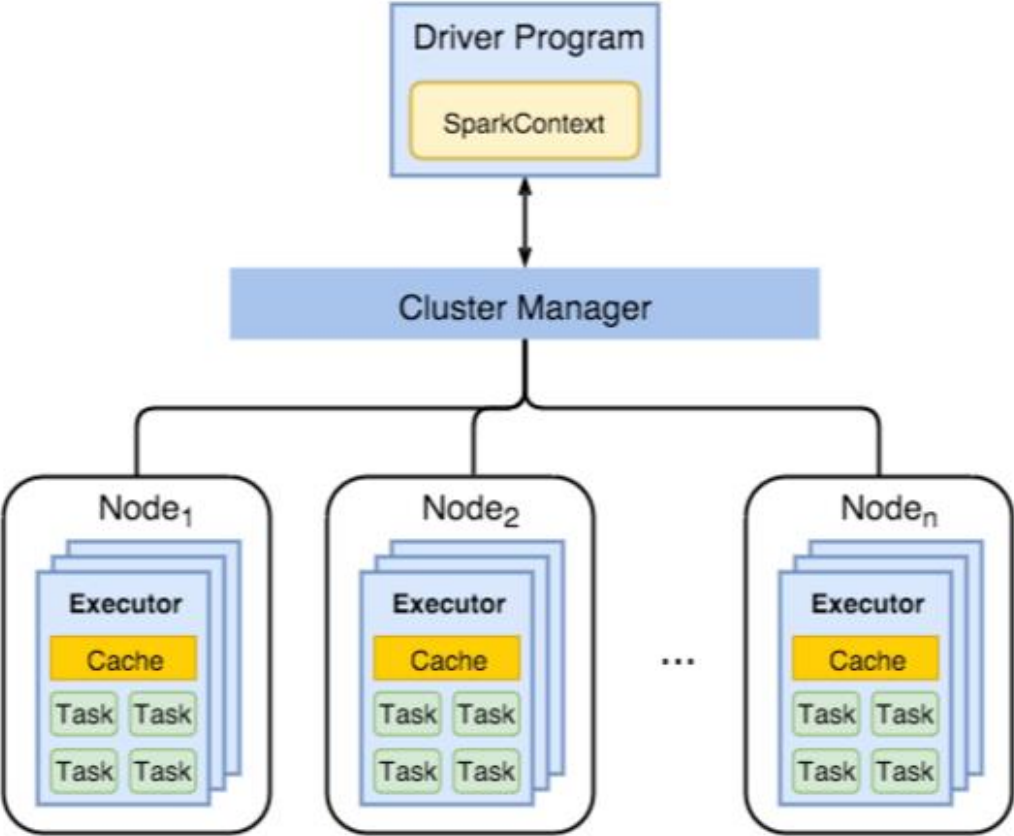
# Spark

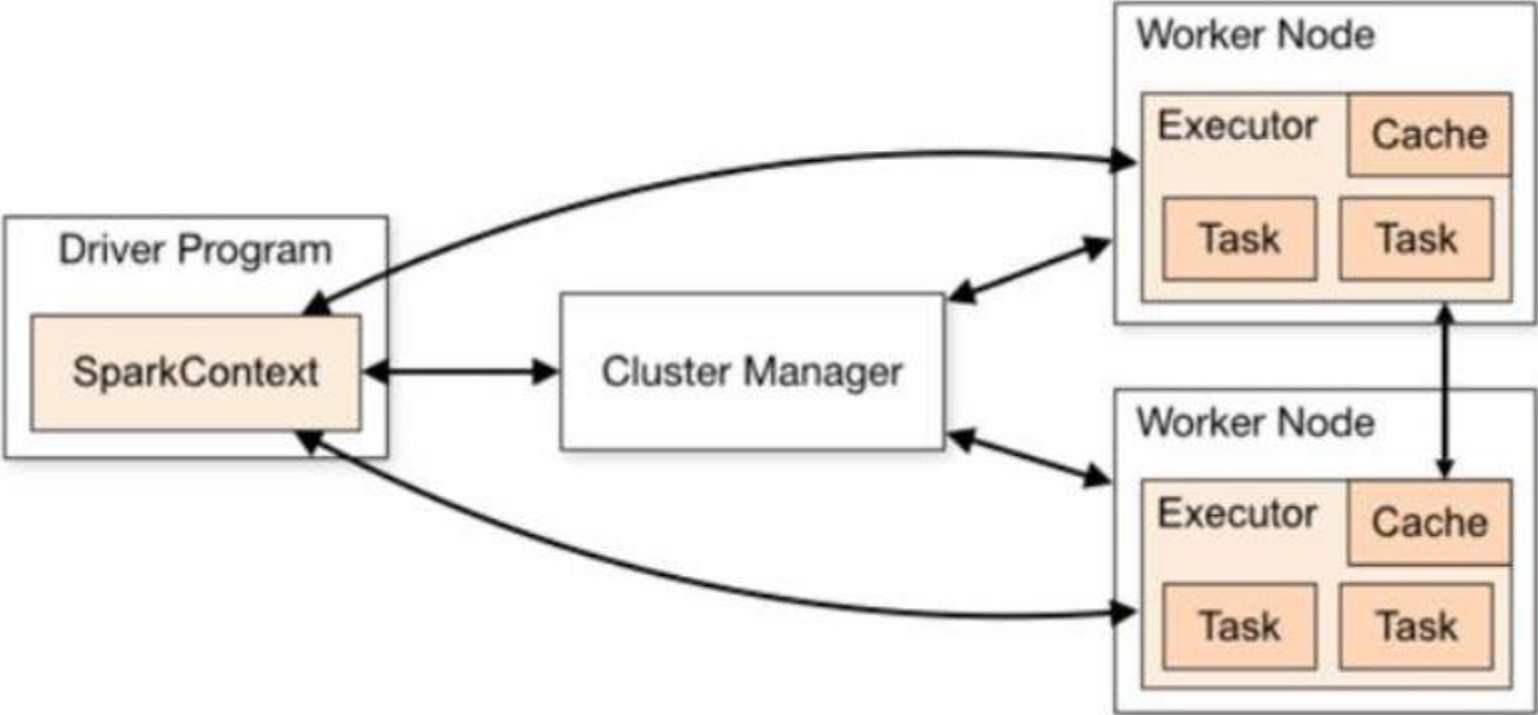
## Clusterización





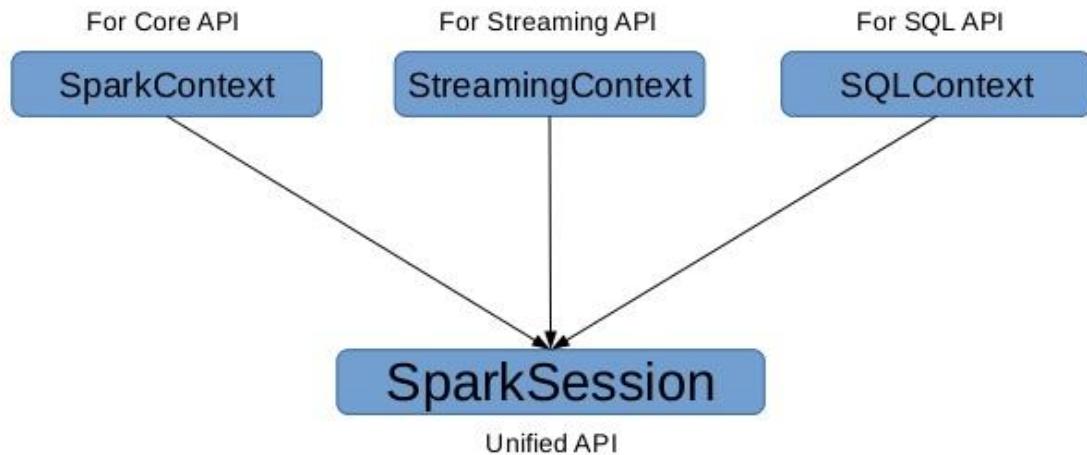






# Spark

## Desarrollo



## DataFrames & Datasets

Batch

Streaming



