Curso de PySpark

O que é Apache Spark

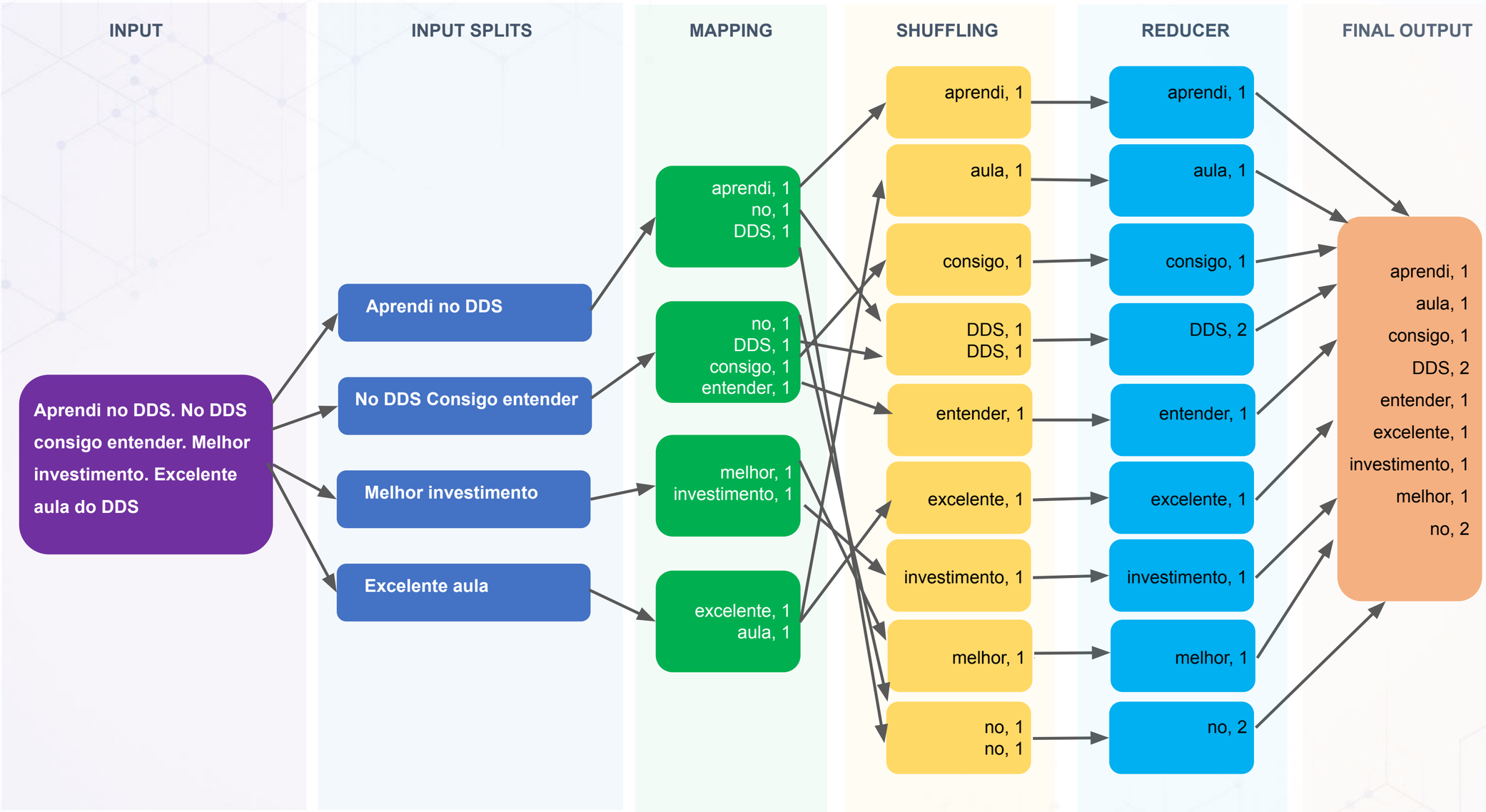
Programa flai

Dominando Data Science 2.0

Divisão e Conquista

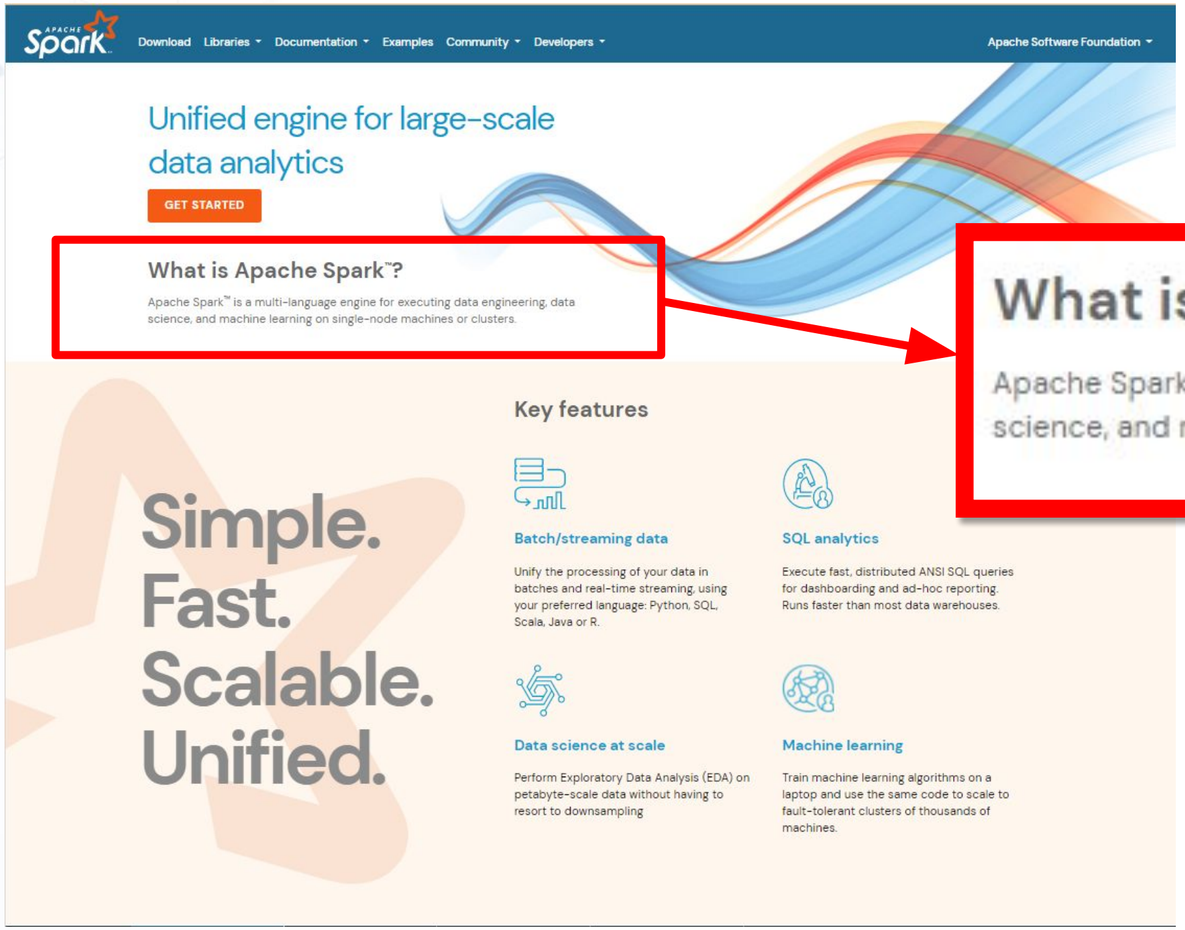
- Divide
- Conquista
- Combina

Divisão e Conquista


INPUT
INPUT SPLITS
MAPPING
SHUFFLING
REDUCER
FINAL OUTPUT
Aprendi no DDS. No DDS consigo entender. Melhor investimento. Excelente aula do DDS
Aprendi no DDS
No DDS Consigo entender
Melhor investimento
Excelente aula
aprendi, 1
no, 1
DDS, 1
no, 1
DDS, 1
consigo, 1
entender, 1
melhor, 1
investimento, 1
excelente, 1
aula, 1
aprendi, 1
aula, 1
consigo, 1
DDS, 1
DDS, 1
entender, 1
excelente, 1
investimento, 1
melhor, 1
no, 1
no, 1
aprendi, 1
aula, 1
consigo, 1
DDS, 2
entender, 1
excelente, 1
investimento, 1
melhor, 1
no, 2
aprendi, 1
aula, 1
consigo, 1
DDS, 2
entender, 1
excelente, 1
investimento, 1
melhor, 1
no, 2

Apache Spark

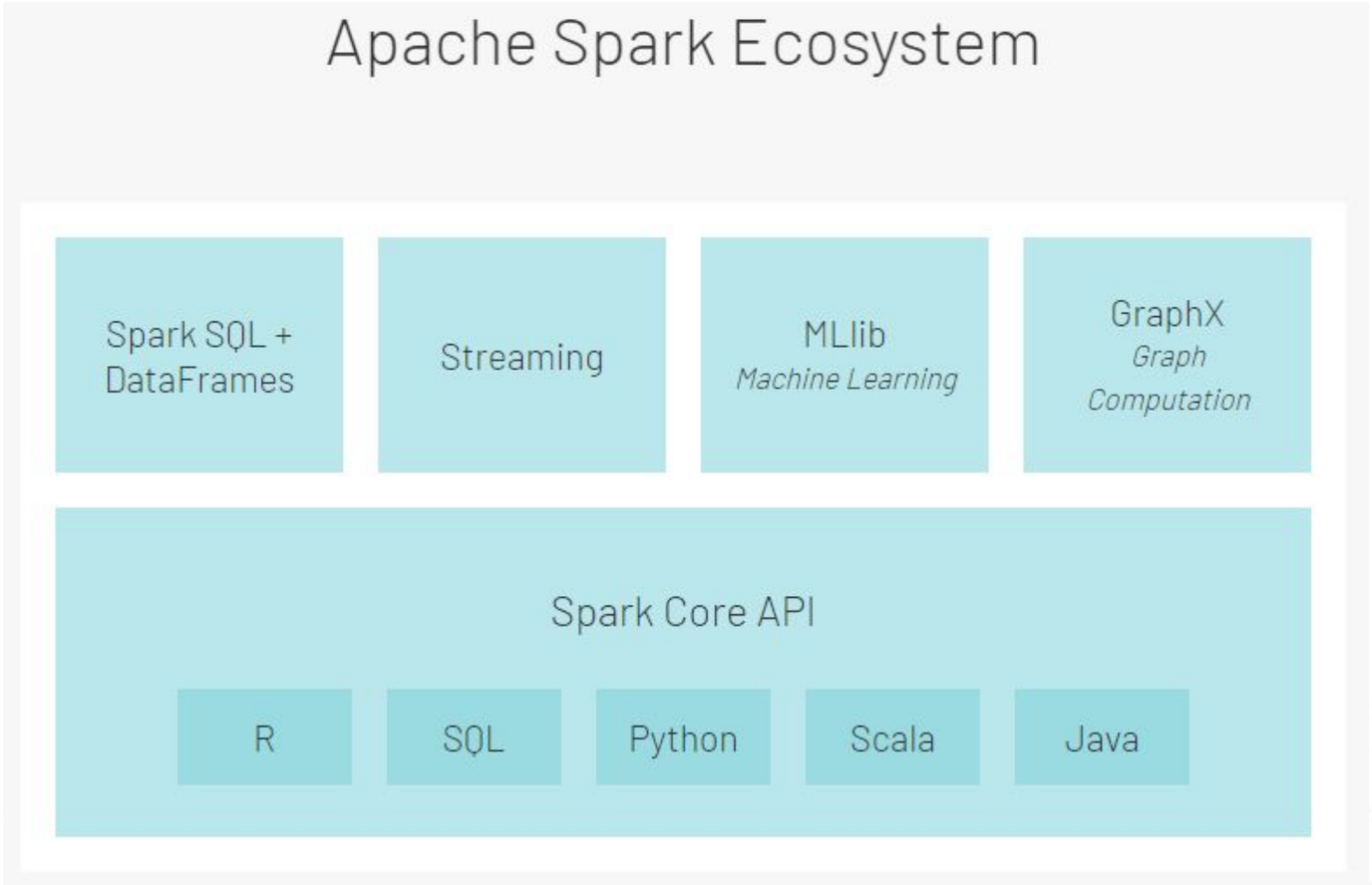
Apache Spark



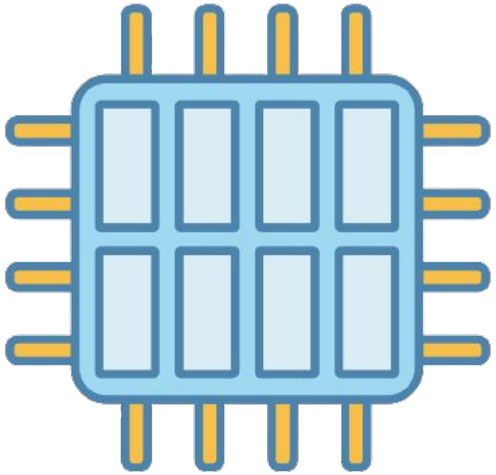
What is Apache Spark™?

Apache Spark™ is a multi-language engine for executing data engineering, data science, and machine learning on single-node machines or clusters.

Apache Spark

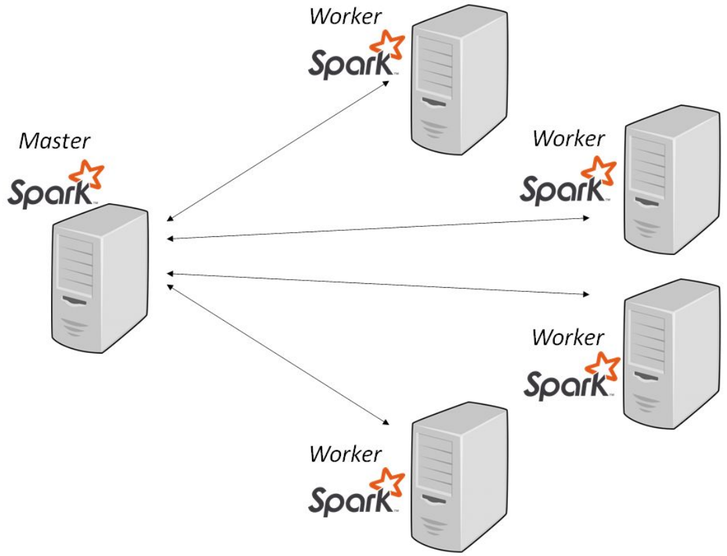


Infraestrutura



standalone

VS



Cluster

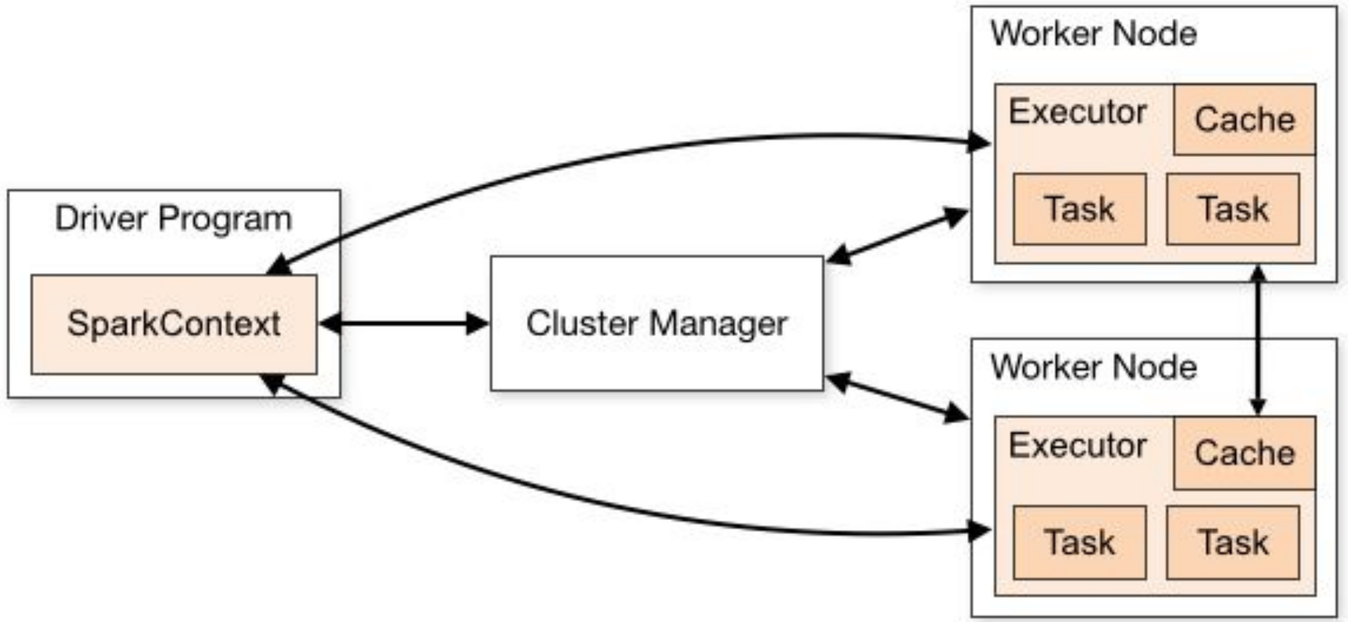
Infraestrutura

On Premise

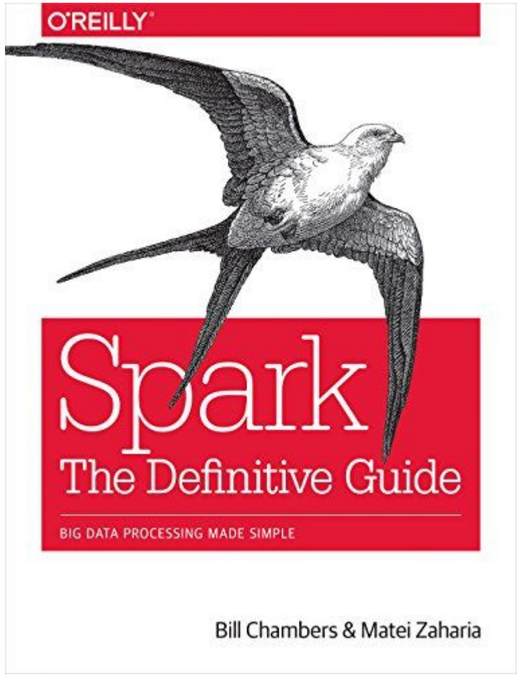
VS

Cloud

Funcionamento



Consultas importantes



- https://spark.apache.org/
- https://docs.microsoft.com/pt-br/dotnet/spark/what-is-spark
- https://cloud.google.com/learn/what-is-apache-spark
- https://aws.amazon.com/pt/big-data/what-is-spark/
- https://databricks.com/it/spark/about

PySpark

https://databricks.com/glossary/pyspark

#itstimetoflai 🚀