

Curso de PySpark

O que é Apache Spark

Programa  flai

Dominando Data Science

2.0

Divisão e Conquista

- Divide
- Conquista
- Combina



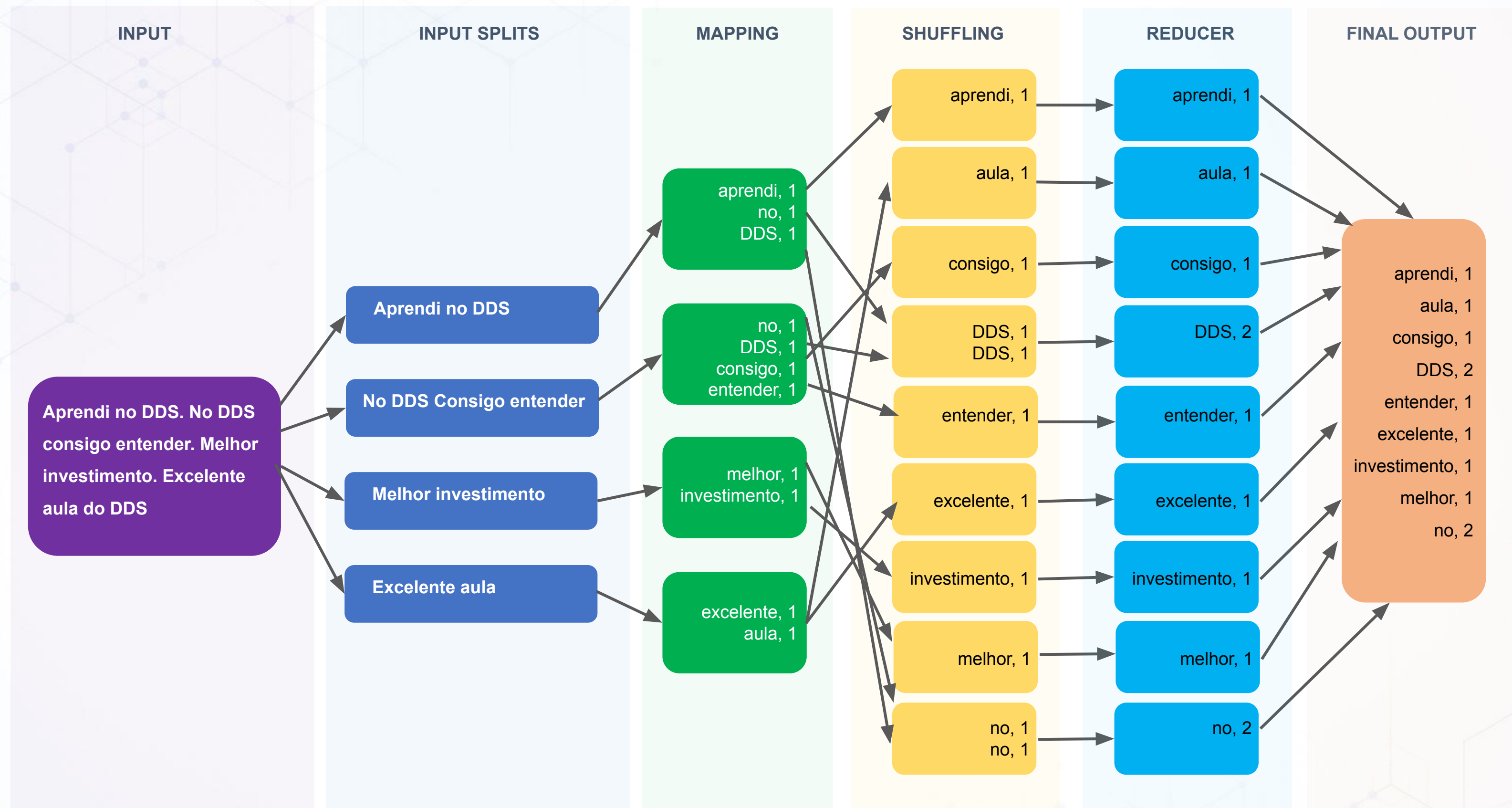
Divisão e Conquista

Programa flai

Dominando Data Science

2.0





Apache Spark

Programa  flai

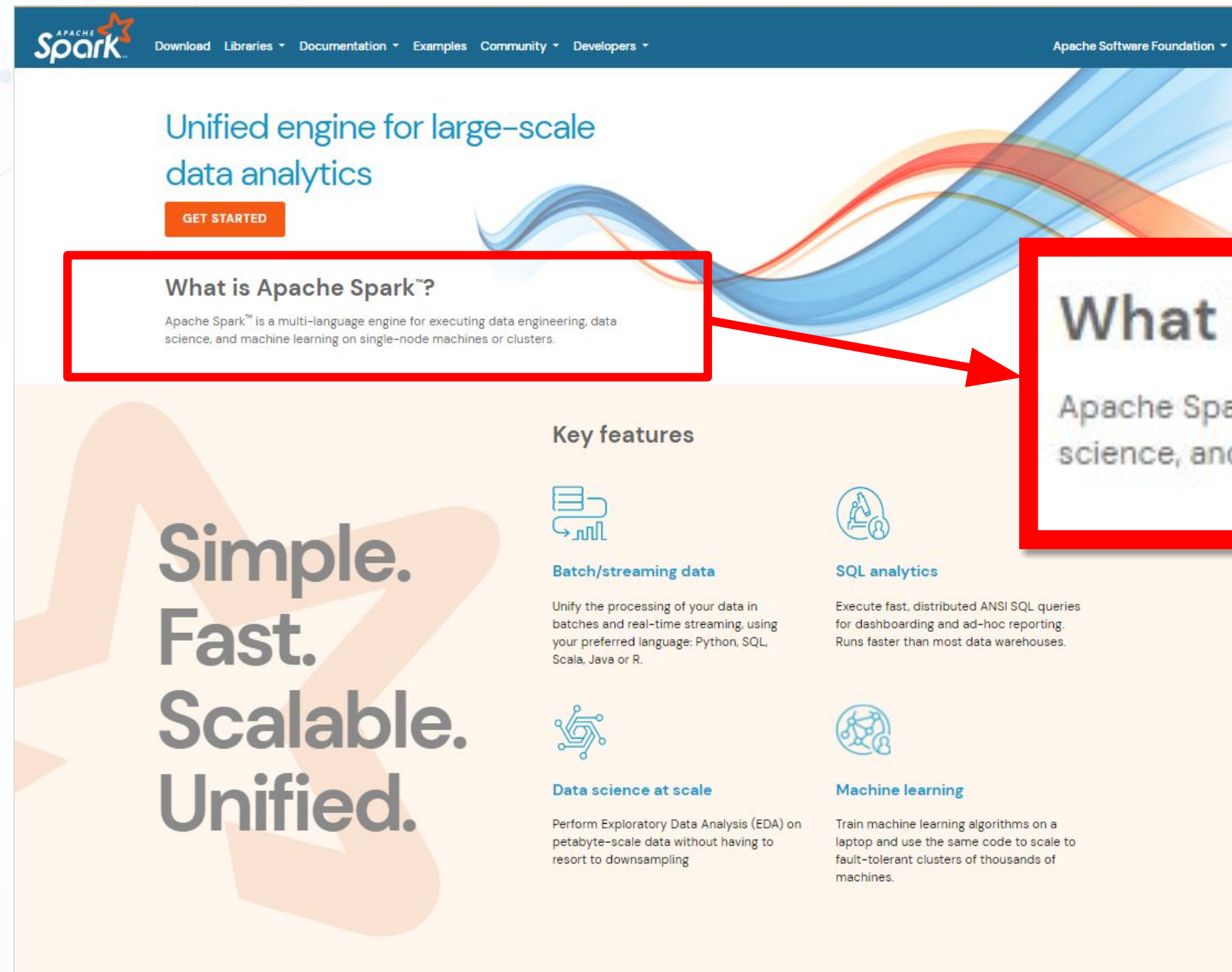
Dominando Data Science

2.0



Apache Spark

Programa  flai
Dominando Data Science
2.0



The screenshot shows the Apache Spark homepage. A red box highlights the 'What is Apache Spark™?' section, and a red arrow points from it to a larger, zoomed-in version of the same text on the right. The website header includes the Apache Spark logo and navigation links: Download, Libraries, Documentation, Examples, Community, and Developers. The main heading is 'Unified engine for large-scale data analytics' with a 'GET STARTED' button. Below this, the 'What is Apache Spark™?' section defines it as a multi-language engine for data engineering, data science, and machine learning. The 'Key features' section lists four capabilities: Batch/streaming data, SQL analytics, Data science at scale, and Machine learning, each with a brief description and an icon. On the left side of the page, the text 'Simple. Fast. Scalable. Unified.' is displayed over a large orange hand icon.

APACHE **Spark**

Download Libraries Documentation Examples Community Developers Apache Software Foundation

Unified engine for large-scale data analytics

GET STARTED

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.

Simple. Fast. Scalable. Unified.

Key features

- Batch/streaming data**
Unify the processing of your data in batches and real-time streaming, using your preferred language: Python, SQL, Scala, Java or R.
- SQL analytics**
Execute fast, distributed ANSI SQL queries for dashboarding and ad-hoc reporting. Runs faster than most data warehouses.
- Data science at scale**
Perform Exploratory Data Analysis (EDA) on petabyte-scale data without having to resort to downsampling
- Machine learning**
Train machine learning algorithms on a laptop and use the same code to scale to fault-tolerant clusters of thousands of machines.

What is Apache Spark™?

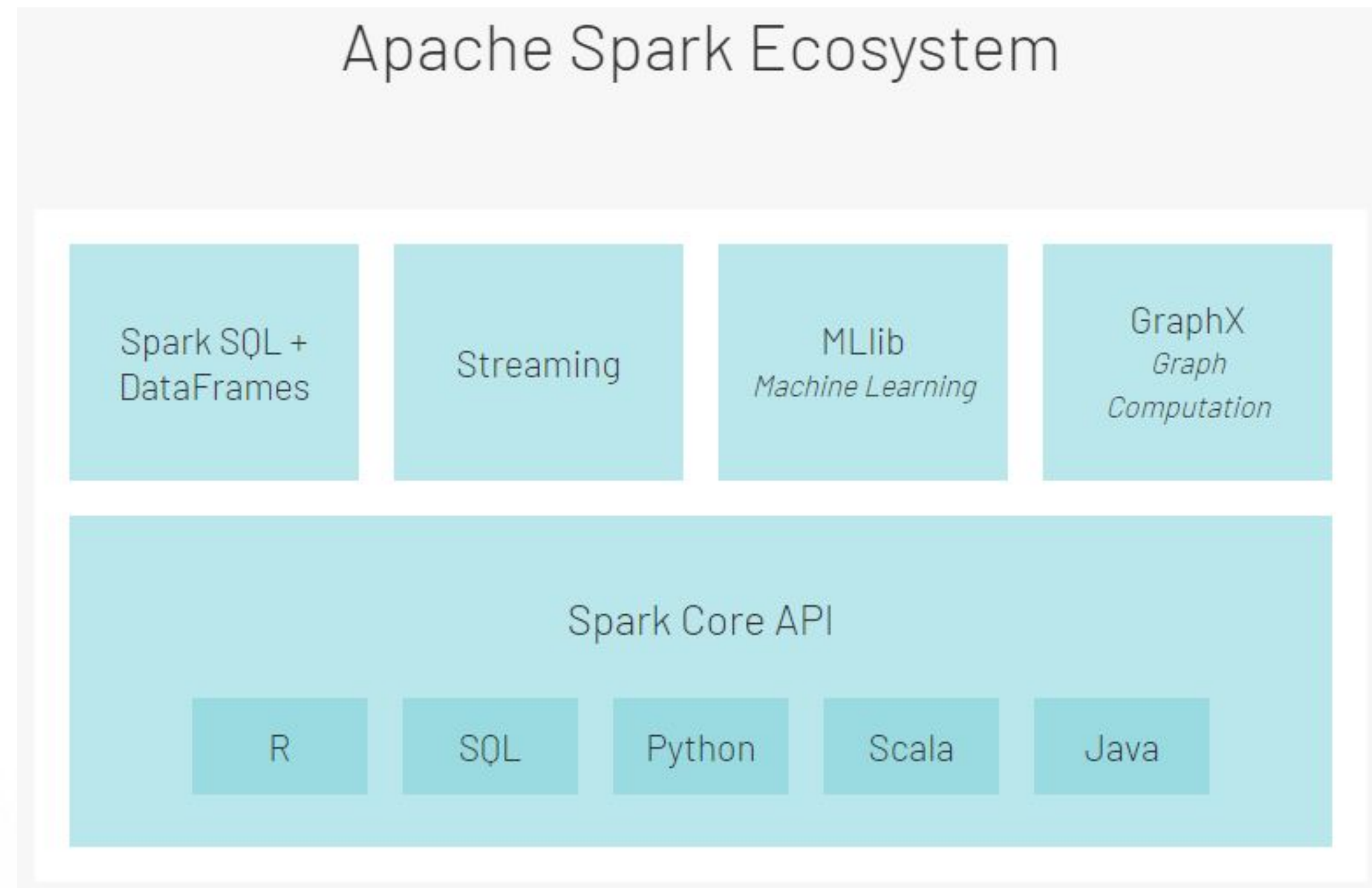
Apache Spark™ is a multi-language engine for executing data engineering, data science, and machine learning on single-node machines or clusters.

Apache Spark

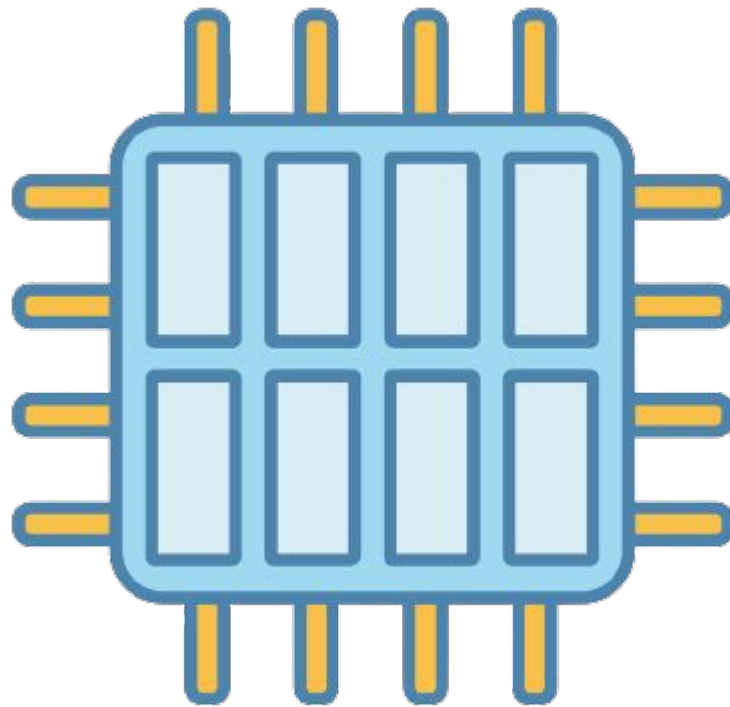
Programa flai

Dominando Data Science

2.0

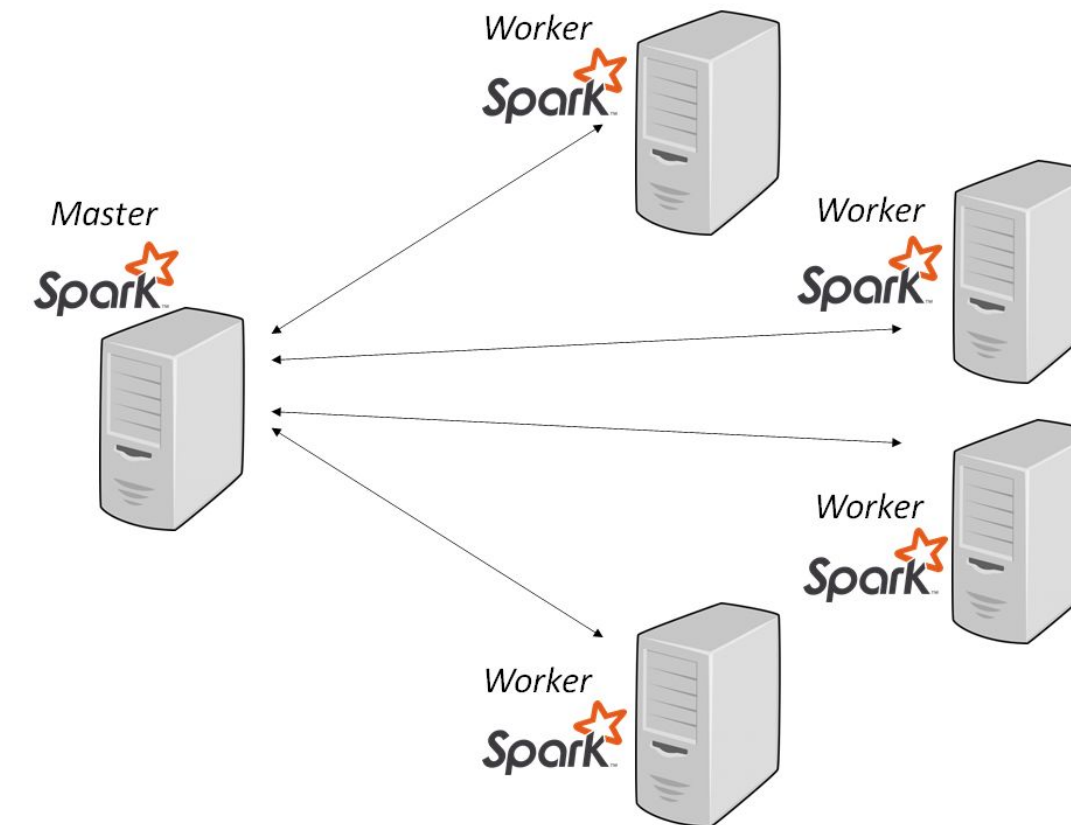


Infraestrutura



standalone

VS



Cluster

Infraestrutura



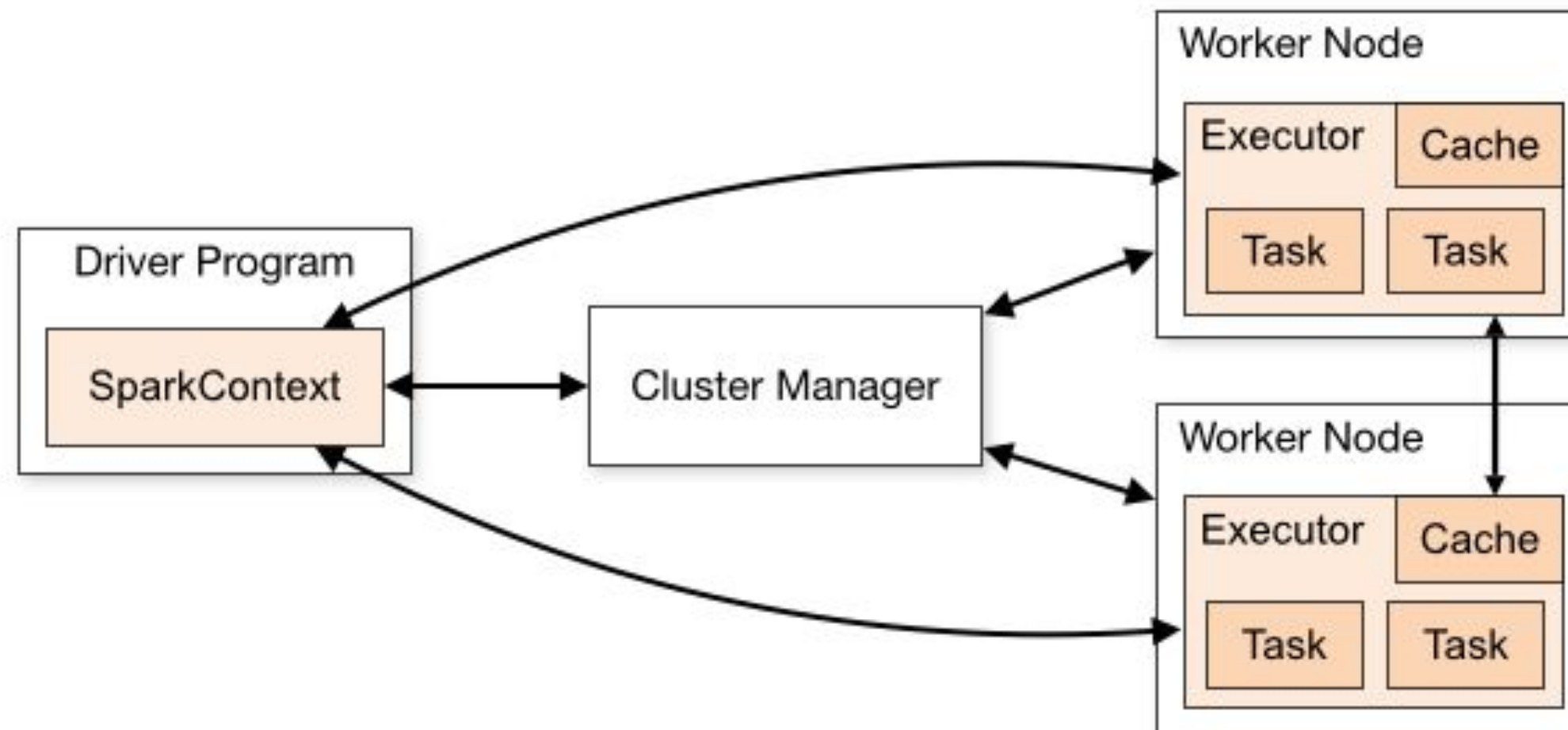
On Premise

VS

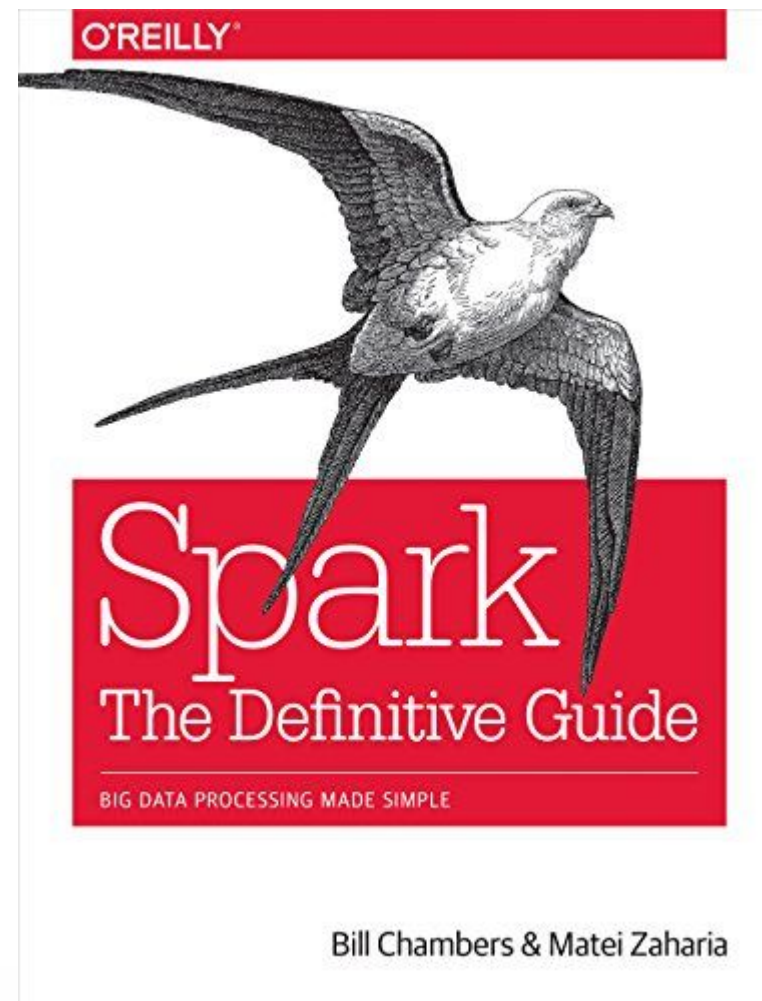


Cloud

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

Programa  flai
Dominando Data Science
2.0



#itstimetoflai 🚀

