

The background is a complex, abstract digital pattern. It features a dense arrangement of small, semi-transparent squares in various shades of blue and cyan, some of which are slightly offset, creating a 3D effect. Interspersed among these are thin, glowing orange and yellow lines that curve and swirl, resembling data streams or particle paths. The overall composition has a sense of depth and movement, with a bright, circular light source or focal point on the right side, from which some of the orange lines radiate outwards.

Apache Spark

Course overview
Spark
Concepts
Hands on



Intro to BigData

Overview of BigData
BigData Ecosystems
Setting up BigData
env.

BigData with Kafka

Real-time data
processing
Key components of
Kafka
Hands on

Apache Spark

Architecture
Spark streaming
and jobs
Hands on

Data Lake

Design &
Architecture
Pipeline
Hands on

What is Spark?

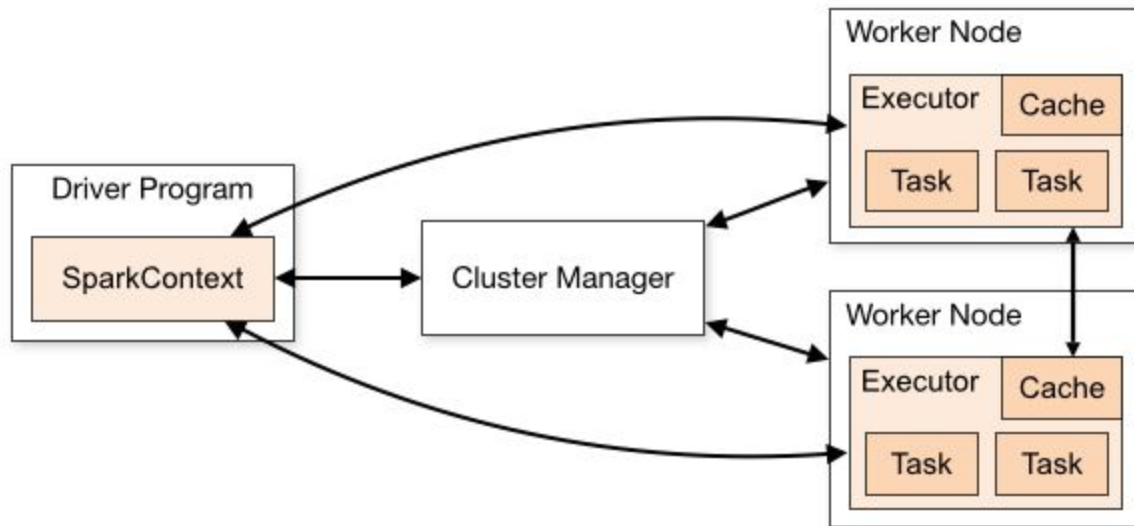
spark

What is Apache Spark?

IBM Cloud



In-memory computing
Distributed system
Multiple Language support
Unified engine
Scalability
Integration



Storage

RDD or DataFrame

Lazy Evaluation

Getting Started

Pandas API Spark

Creating & Loading datasets

Basic operations

Joining and Merging

Handling missing data

Conversion

Performance

Pandas vs Spark

Dataprocc



Run Spark and Hadoop
Faster with Cloud
Dataproc

Our 1st Cluster

Our 1st Job

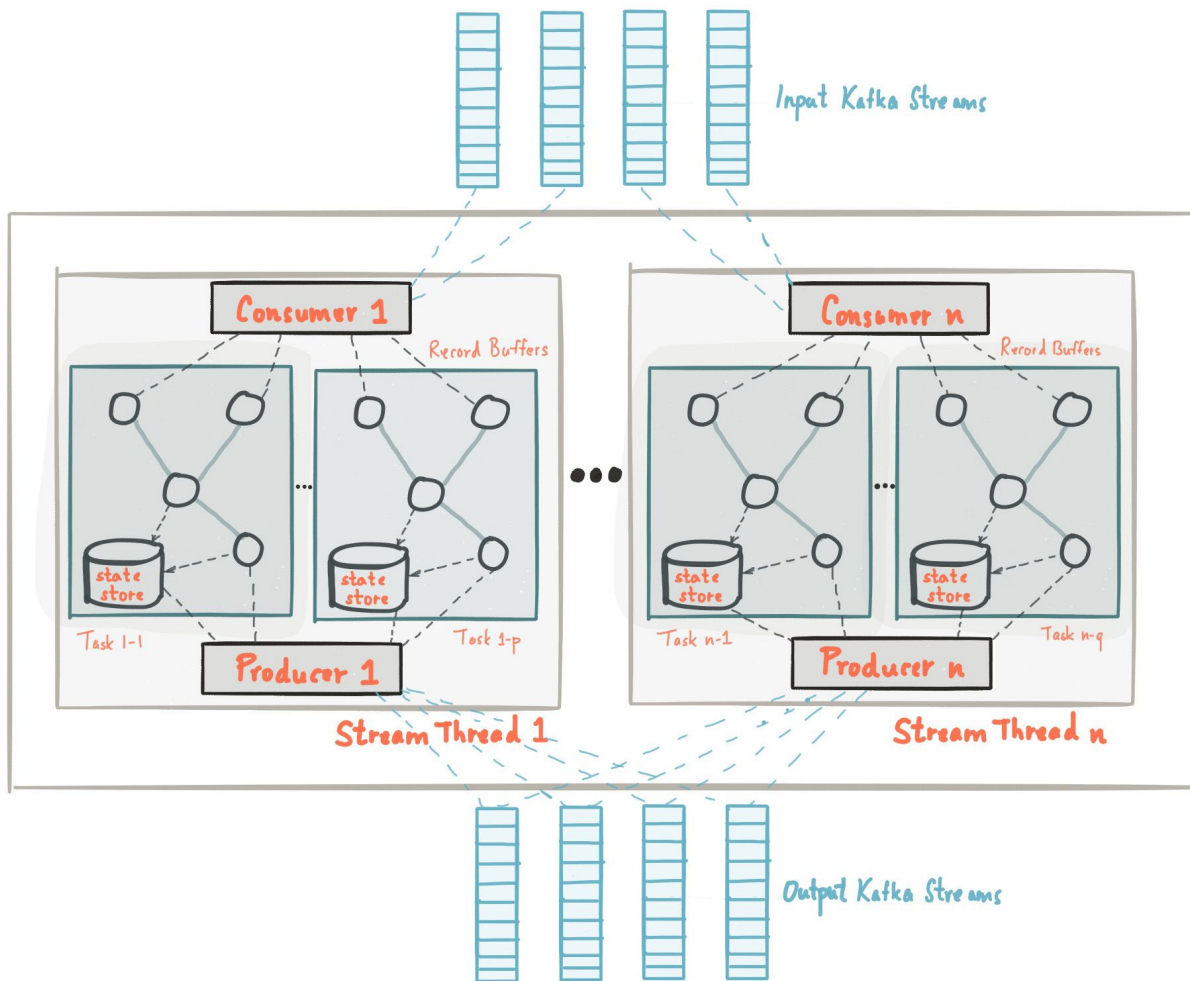
Cluster and Job

Simulate in Jupyter

Revisit Streaming



Kafka Streams



WHAT IS APACHE FLINK?





Flink

Core terminologies

Stream

Event

Task

Operator

Job

State

Checkpoint

Getting started with Flink

Have Questions?

**Working with DATA
is an ART**

