

Introduction to Apache Spark



After this video you will be able to..

- List the main motivations for the development of Spark
- Draw the Spark stack as a layer diagram
- Explain the functionality of the components in the Spark stack

Why Spark?

Hadoop MapReduce Shortcomings

Only for Map and Reduce based computations

Relies on reading data from HDFS

Native support for Java only

No interactive shell support

No support for streaming



Basics of Data Analysis with Spark

Expressive programming model

In-memory processing

Support for diverse workloads

Interactive shell

The Spark Stack

SparkSQL

**Spark
Streaming**

MLlib

GraphX

Spark Core

The Spark Stack

SparkSQL

**Spark
Streaming**

MLlib

GraphX

Spark Core

The Spark Stack

SparkSQL

**Spark
Streaming**

MLlib

GraphX

Spark Core

The Spark Stack

SparkSQL

**Spark
Streaming**

MLlib

GraphX

Spark Core

The Spark Stack

SparkSQL

**Spark
Streaming**

MLlib

GraphX

Spark Core



```
graph TD; subgraph TopLayer; direction LR; A[SparkSQL]; B[Spark Streaming]; C[MLlib]; D[GraphX]; end; B1[Spark Core]; TopLayer --- B1;
```

The Spark Stack

SparkSQL

**Spark
Streaming**

MLlib

GraphX

Spark Core



```
graph TD; subgraph Stack; direction TB; A[SparkSQL] --- B[Spark Streaming]; B --- C[MLlib]; C --- D[GraphX]; end; E[Spark Core] --- Stack;
```

