

Spark Course

Chapter 01

- Course Introduction
- Course References
- What is Apache Spark?
- Python Vs. Scala
- History of Apache Spark
- About Databricks
- Apache Spark in Data Platforms
- Running Spark
 - Demo: Databricks Community Edition
 - Demo: Install Spark Locally
 - Demo: Spark on Docker
 - Demo: Spark Shell
 - Demo: Spark SQL
- From MapReduce To Apache Spark
- What is Apache Spark
 - Apache Spark
 - Ease of Use
 - Speed
 - Modularity
 - Extensibility
- Apache Spark as a unified stack
 - Languages
 - Low level APIs
 - Use Case: GeoData
 - Structured APIs
 - RDD vs Dataframe vs Dataset ↗
 - Structured Streaming
 - Advanced Analytics
 - Libraries & EcoSystems
- Who Uses Spark, and for What?
- Lab: Run simple analysis using SparkSQL (Provide sample code to create the table)

Chapter 02

- Spark Applications Concept
 - Application
 - SparkSession
 - Job
 - Stage
 - Task
- Spark Architecture
 - Spark Driver
 - Spark Session
 - Cluster Manager
 - Spark Executor
 - Deployment Mode
 - Distributed data and partitions
 - JVM
- Spark Operations
 - Transformations
 - Narrow vs Wide
 - Lazy Evalution
 - Actions
- Lab: Running spark etl for counting workds and m&M

Chapter 03

Chapter 04

Chapter 05

Chapter 06

Chapter 07

Chapter 08