

Azure Data bricks

Program Duration: 5 days

Introduction

- Overview of Big Data Architectures
- Top-down vs bottom-up
- What is Azure Databricks?

Scala Basics

- Data types List
- Dictionary, Tuples
- Loops, Functions
- Object oriented programming
- Object, Classes

Databricks concepts

- Workspace
- Interface
- Data Management
- Computation Management
- Model Management
- Authentication and Authorization

Apache Spark

- What is Apache Spark?
- Spark Architecture
- What is Ecosystem of Apache Spark?
- DataFrames and Datasets

Databricks development and Deployment

- Collaborative Workspace
- Perform ETL Operations
- Deploy production jobs and workflows
- Optimized data bricks runtime engine

Scala Spark

- What is Spark, anyway?
- Using Spark in Scala
- Examining The Spark Context
- Using DataFrames
- Creating a Spark Session Viewing tables Are you query-ious?
- Dropping the middleman

Manipulating Data

- Creating columns
- SQL in a nutshell
- SQL in a nutshell (2)
- Filtering Data
- Selecting
- Selecting II

- Aggregating Aggregating II
- Grouping and Aggregating I
- Grouping and Aggregating II
- Joining

Databricks Jobs & Cluster

- Introduction to Jobs and Cluster
- General Spark Cluster Architecture
- How to Submit Jobs using Job Cluster?
- Pool in Databricks
- Azure Databricks Integration with AAD
- Clusters: Auto Scaling and auto termination

Databricks Data Lake

- Data lake defined
- Hadoop as the data lake

Modern data warehouse

- Federated querying
- Solution in the cloud
- SMP Vs MPP

Spark Streaming

- Introduction to Spark Streaming
- Architecture of Spark Streaming
- Initializing Streaming Context
- Discretized Streams (DStreams)
- DStreams APIs
- Transformations on DStreams
- Input Operations on DStreams
- Output Operations on DStreams