

Databricks Training

Overview of Apache Spark and Databricks

- What is Apache Spark?
- How do we define Big Data?
- Spark languages - Scala, Python, R, Java, SQL
- Databricks Community Edition
- Databricks Architecture
- Defining Data Analytics
- Defining Machine Learning
- Azure implementation
- AWS implementation

Databricks benefits

- Collaboration
- Scaling
- Integrating into Pipelines

Getting started with Databricks

- Creating a Databricks Workspace on Azure
- Creating and configuring your Cluster
- Creating and attaching your first Notebook
- Testing your Notebook

Uploading data

- Creating a Table
- Connecting to a Spark data source
- Previewing your Table
- Columns and Datatypes basics

Bringing your data into your Notebook

- Writing the initial SQL query to import
- View aggregates
- Perform Joins

Visualisations & DataFrames

- Datatypes
- DataFrames
- Images
- Structured Streaming DataFrames
- Plots
- Choosing Chart types
- Chart Toolbar
- Layout and styling considerations
- Machine Learning visualisations

Databricks Jobs

- Creating a Job
- View Jobs and Job details
- Running your first Job
- Scheduling Jobs
- Setting Parameters
- Viewing completed jobs
- Managing Dependencies
- Setting up Alerts

Delta Lake and Delta Tables

- Getting data into Delta Lake
- Reads and Writes
 - Batch
 - Streaming
- Delete, update, merge
- Constraints
- Versioning
- Concurrency
- Integrations
- Overview of Delta Engine