elite Analytical Solutions®
Business Analytics + Data Consultants

Intro to Databricks

databricks

Length: 2 Full Days / 4 Half Days

Overview: This course is tailored for individuals aiming to establish, deploy, and oversee data analytics solutions through Databricks. Participants will gain proficiency in setting up, deploying, and managing data analytics solutions using the Databricks platform. The course will cover topics such as configuring Databricks environments, deploying data analytics solutions, and effectively managing data workflows within Databricks. By the end of the course, students will have the skills to leverage Databricks for efficient data analytics operations.

Objectives:
- Set up and configure Databricks.
- Learn how Databricks and Apache Spark work in harmony.
- Load and transform data in Databricks.

Introduction

- Overview of Databricks and Apache Spark
- Understanding the Databricks architecture

Getting Started

- Setting up the Environment
- Setting up and configuring Databricks
- Navigating the Databricks user interface
- Creating a Databricks workspace

Working with Data in Databricks

- Connecting to an Apache Spark data source
- Understanding the basics columns and datatypes
- Managing file system into Notebookss

Managing Jobs and Clusters

- Creating and configuring clusters
- Creating jobs using Notebook
- Running jobs
- Viewing jobs and job details

Using Delta Lake in Databricks

- Loading data into Delta Lake
- Managing data in Delta Lake

Securing Databricks

- Managing Databricks security
- Managing backup and recovery

Troubleshooting

Summary and Plan Moving forward

Prerequisites

No prerequisites for this course.

Materials

- Software Needed on Each Student PC
  - Internet access
  - Web Browser
  - Access to a Databricks account
  - Optional – Python IDE such as Google Colab
  - Optional – SQL IDE such as SSMS
- All students will receive lecture material and data and labs.
- Related data and lab files will be provided