Course Name	Data Analytics and Machine Learning with Databricks
Course Duration	4 Days (32 hours)
Time Division	**Break:** 1 Hr. 15 Minutes /Day **Session:** 6 Hrs. 45 Minutes/Day
Pre-requisite	Should familiar with programming language like python/C++/S QL/Scala/R

Important Note:

- Courseware – Reference material/ppt along with lab files/exercises will be provided
- Azure subscription is needed to perform labs.
- All labs will be performed on Azure Databricks Platform.

Section	Sub-sections	Estimated Hours
1. Overview of Apache Spark and Databricks	- What is Apache Spark?	6 hours
	- 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	
Lab 1: Introduction to Apache Spark	**Title**: Getting Started with Apache Spark	
	Aim: To understand the basics of Apache Spark and its role in Big Data processing.	
2. Databricks Benefits	- Collaboration	2 hours
	- Scaling	
	- Integrating into Pipelines	
3. Getting Started with Databricks	- Creating a Databricks Workspace on Azure	4 hours

	- Creating and configuring your Cluster	
	- Creating and attaching your first Notebook	
	- Testing your Notebook	
Lab 2: Exploring Databricks Community Edition	**Title**: Setting Up Databricks Community Edition	
	Aim: To create and configure a Databricks workspace and cluster.	
4. Uploading Data	- Creating a Table	3 hours
	- Connecting to a Spark data source	
	- Previewing your Table	
	- Columns and Datatypes basics	
Lab 3: Data Ingestion and Table Creation	**Title**: Uploading and Managing Data in Databricks	
	Aim: To upload data, create tables, and understand columns and datatypes.	
5. Bringing Your Data into Your Notebook	- Writing the initial SQL query to import	3 hours
	- Viewing aggregates	
	- Performing Joins	
Lab 4: Data Analysis with SQL	**Title**: Performing Data Analysis Using SQL in Databricks	
	Aim: To write SQL queries to import data, view aggregates, and perform joins.	
6. Visualisations & DataFrames	- Datatypes	5 hours
	- DataFrames	
	- Images	
	- Structured Streaming DataFrames	
	- Plots	
	- Choosing Chart types	

	- Chart Toolbar	
	- Layout and styling considerations	
	- Machine Learning visualisations	
Lab 5: Data Visualisation Techniques	**Title**: Visualising Data with Databricks	
	Aim: To create visualisations using DataFrames and understand different chart types.	
7. Databricks Jobs	- Creating a Job	4 hours
	- Viewing Jobs and Job details	
	- Running your first Job	
	- Scheduling Jobs	
	- Setting Parameters	
	- Viewing completed jobs	
	- Managing Dependencies	
	- Setting up Alerts	
Lab 6: Running Jobs in Databricks	**Title**: Automating Tasks with Databricks Jobs	
	Aim: To create, schedule, and manage jobs in Databricks.	
8. Delta Lake and Delta Tables	- Getting data into Delta Lake	5 hours
	- Reads and Writes	
	- Batch	
	- Streaming	
	- Delete, update, merge	
	- Constraints	
	- Versioning	

	- Concurrency	
	- Integrations	
	- Overview of Delta Engine	
Lab 7: Introduction to Delta Lake	**Title**: Working with Delta Lake	
	Aim: To understand Delta Lake, perform batch and streaming reads/writes, and manage data versioning.	
Lab 8: Advanced Data Management	**Title**: Advanced Data Management with Delta Tables	
	Aim: To perform advanced data operations like delete, update, merge, and manage constraints and concurrency.	
Lab 9: Machine Learning with Databricks	**Title**: Implementing Machine Learning Models in Databricks	
	Aim: To build and visualise machine learning models using Databricks.	