

Course Name	Data Processing & Machine Learning with Azure Databricks
Course Duration	3 Day (24 hours)
Target Audience	Data Scientists, ML Engineer, ML Enthusiasts
Course Outcomes	Gain proficiency in using Azure Databricks and Spark to process big data
	Develop skills to work with data from multiple sources in various formats
	Learn how to create and apply user-defined functions to process data
	Build machine learning workflows and pipelines using PySpark's machine learning package
	Acquire knowledge of MLflow and Spark's machine learning library for model selection and hyperparameter tuning
Module	Content
01	Introduction to Azure Databricks
1.1	Explore Azure Databricks capabilities for processing big data using Apache Spark notebooks
1.2	Identify tasks suitable for Apache Spark on Azure Databricks
1.3	Understand the architecture of an Azure Databricks Spark Cluster and Spark Jobs
02	Data Handling in Azure Databricks
2.1	Utilize Azure Databricks for day-to-day data-handling functions such as reads, writes, and queries
2.2	Work with large amounts of data from multiple sources in different raw formats
2.3	Learn to use DataFrame Column Class Azure Databricks to apply column-level transformations, such as sorts, filters, and aggregations
2.4	Use advanced DataFrame functions to manipulate data, apply aggregates, and perform date and time operations in Azure Databricks
03	Processing Data with Azure Databricks
3.1	Use Azure Databricks built-in SQL functions and learn to write custom functions with User-Defined Function (UDF)
3.2	Register and invoke UDFs
3.3	Create, append, and upsert data to Apache Spark tables using Delta Lake, taking advantage of built-in reliability and optimizations
04	Building Machine Learning Workflows with Databricks
4.1	Use PySpark's machine learning package to build exploratory data analysis, model training, and model evaluation components of the machine learning workflows
4.2	Build pipelines for common data featurization tasks
05	Managing Machine Learning Lifecycles and Fine-Tuning Models
5.1	Track machine learning experiments using MLflow
5.2	Use modules from the Spark's machine learning library for hyperparameter tuning and model selection
06	Training and Serving Distributed Neural Networks with Azure Machine Learning
5.4	Use of Uber's Horovod framework along with the Petastorm library to run distributed, deep learning training jobs on Spark

5.5	Use of MLflow and Azure Machine Learning service to register, package, and deploy trained models to Azure Container Instance and Azure Kubernetes Service as a scoring web service
-----	--