

## MS-DP3014: BUILD MACHINE LEARNING SOLUTIONS USING AZURE DATABRICKS



| DURATION | LEVEL        | TECHNOLOGY | DELIVERY METHOD | TRAINING CREDITS |
|----------|--------------|------------|-----------------|------------------|
| 1 Day    | Intermediate | Azure Data | Instructor-led  | NA               |

### INTRODUCTION

Azure Databricks is a fully managed, cloud-based data analytics platform, which empowers developers to accelerate AI and innovation by simplifying the process of building enterprise-grade data applications. Built as a joint effort by Microsoft and the team that started Apache Spark, Azure Databricks provides data science, engineering, and analytical teams with a single platform for big data processing and machine learning. In this course, you'll learn how to use Azure Databricks to train and deploy machine learning models.

### AUDIENCE PROFILE

The primary audience includes:

- Data Scientists: Individuals who build and deploy machine learning models using large datasets.
- Machine Learning Engineers: Professionals who focus on the engineering aspects of machine learning, including model deployment and monitoring.
- Data Engineers: Those responsible for preparing and managing data for machine learning projects.
- AI Engineers: Engineers who design and implement AI solutions using machine learning models.
- Developers: Software developers who integrate machine learning models into applications.
- IT Professionals: Individuals who support the infrastructure and deployment of machine learning models.

### PREREQUISITES

This learning path assumes that you have experience using Python to explore data and train machine learning models with common open-source frameworks, like Scikit-Learn, PyTorch, and TensorFlow. Consider completing the Create machine learning models learning path before starting this one.

### COURSE OBJECTIVES

After attending this course, delegates will be able to:

- Provision an Azure Databricks Workspace: Learn to set up and configure an Azure Databricks workspace and cluster.
- Use Apache Spark in Azure Databricks: Understand how to use Apache Spark for data processing and analysis.
- Prepare Data for Machine Learning: Learn techniques for data ingestion, preparation, and preprocessing.
- Train Machine Learning Models: Use Azure Databricks to train machine learning models with various frameworks like Scikit-Learn, PyTorch, and TensorFlow.
- Track Experiments with MLflow: Utilize MLflow to log parameters, metrics, and manage the lifecycle of machine learning models.
- Tune Hyperparameters: Optimize model performance by tuning hyperparameters using libraries like Hyperopt.
- Use AutoML: Simplify the process of building machine learning models with Azure Databricks' AutoML capabilities.
- Deploy and Monitor Models: Deploy trained models to production and monitor their performance.

### COURSE CONTENT

#### Module 1: Explore Azure Databricks

Azure Databricks is a cloud service that provides a scalable platform for data analytics using Apache Spark.

Lessons

- Introduction.
- Get started with Azure Databricks.
- Identify Azure Databricks workloads.
- Understand key concepts.
- Exercise - Explore Azure Databricks.
- Knowledge check.
- Summary.

By the end of this module, you'll be able to:

- Provision an Azure Databricks workspace.
- Identify core workloads and personas for Azure Databricks.
- Describe key concepts of an Azure Databricks solution

#### Module 2: Use Apache Spark in Azure Databricks

Azure Databricks is built on Apache Spark and enables data engineers and analysts to run Spark jobs to transform, analyse and visualize data at scale.

Lessons

- Introduction.
- Get to know Spark.
- Create a Spark cluster.
- Use Spark in notebooks.

- Use Spark to work with data files.
- Visualize data.
- Exercise - Use Spark in Azure Databricks.
- Knowledge check.
- Summary.

By the end of this module, you'll be able to:

- Describe key elements of the Apache Spark architecture.
- Create and configure a Spark cluster.
- Describe use cases for Spark.
- Use Spark to process and analyze data stored in files.
- Use Spark to visualize data.

## Module 3: Train a machine learning model in Azure Databricks

Machine learning involves using data to train a predictive model. Azure Databricks support multiple commonly used machine learning frameworks that you can use to train models.

### Lessons

- Introduction.
- Understand principles of machine learning.
- Machine learning in Azure Databricks.
- Prepare data for machine learning.
- Train a machine learning model.
- Evaluate a machine learning model.
- Exercise - Train a machine learning model in Azure Databricks.
- Knowledge check.
- Summary.

By the end of this module, you'll be able to:

- Prepare data for machine learning.
- Train a machine learning model.
- Evaluate a machine learning model.

## Module 4: Use MLflow in Azure Databricks

MLflow is an open-source platform for managing the machine learning lifecycle that is natively supported in Azure Databricks.

### Lessons

- Introduction.
- Capabilities of MLflow.
- Run experiments with MLflow.
- Register and serve models with MLflow.
- Exercise - Use MLflow in Azure Databricks.
- Knowledge check.
- Summary.

By the end of this module, you'll be able to:

- Use MLflow to log parameters, metrics, and other details from experiment runs.

- Use MLflow to manage and deploy trained models.

## Module 5: Tune hyperparameters in Azure Databricks

Tuning hyperparameters is an essential part of machine learning. In Azure Databricks, you can use the Hyperopt library to optimize hyperparameters automatically.

### Lessons

- Introduction
- Optimize hyperparameters with Hyperopt
- Review Hyperopt trials
- Scale Hyperopt trials
- Exercise - Optimize hyperparameters for machine learning in Azure Databricks
- Knowledge check
- Summary

By the end of this module, you'll be able to:

- Use the Hyperopt library to optimize hyperparameters.
- Distribute hyperparameter tuning across multiple worker nodes.

## Module 6: Use AutoML in Azure Databricks

AutoML in Azure Databricks simplifies the process of building an effective machine learning model for your data.

### Lessons

- Introduction.
- What is AutoML?
- Use AutoML in the Azure Databricks user interface.
- Use code to run an AutoML experiment.
- Exercise - Use AutoML in Azure Databricks.
- Knowledge check.
- Summary.

### Learning objectives

- Use the AutoML user interface in Azure Databricks.
- Use the AutoML API in Azure Databricks.

## Module 7: Train deep learning models in Azure Databricks

Deep learning uses neural networks to train highly effective machine learning models for complex forecasting, computer

vision, natural language processing, and other AI workloads.

### Lessons

- Introduction.
- Understand deep learning concepts.
- Train models with PyTorch.
- Distribute PyTorch training with Horovod.
- Exercise - Train deep learning models on Azure Databricks.
- Knowledge check.
- Summary.

### Learning Objectives

- Train a deep learning model in Azure Databricks
- Distribute deep learning training by using the Horovod library.

## Module 8: Manage machine learning in production with Azure Databricks

Machine learning enables data-driven decision-making and automation, but deploying models into production for real-time insights is challenging. Azure Databricks simplifies this process by providing a unified platform for building, training, and deploying machine learning models at scale, fostering collaboration between data scientists and engineers.

### Lessons

- Introduction
- Automate your data transformations.
- Explore model development.
- Explore model deployment strategies.
- Explore model versioning and lifecycle management.
- Exercise - Manage a machine learning model.
- Knowledge check.
- Summary.

### Learning Objectives

- Automate feature engineering and data pipelines.
- Model development and training.
- Model deployment strategies.
- Model versioning and lifecycle management.

## ASSOCIATED CERTIFICATIONS & EXAM

There is no Associated Certification or Exam for this course.