

**Implementing a Machine Learning Solution with Azure Databricks (DP-3014)**

**Course Number:** MOC-DP-3014
**Duration:** 1 day

**Overview**

Azure Databricks is a cloud-scale platform for data analytics and machine learning (ML). This official Microsoft course, DP-3014: Implementing a Machine Learning Solution with Azure Databricks training, teaches data scientists and ML engineers how to use Azure Databricks to implement machine learning solutions at scale.

**Prerequisites**

All attendees must have experience in using Python to explore data and train machine learning models with common open-source frameworks, like Scikit-Learn, PyTorch, and TensorFlow.

**Materials**

Attendees will not need to install any software on their computers for this class. The class will be conducted in a remote environment that Accelebrate will provide; students will only need a local computer with a web browser and a stable Internet connection. Any recent version of Microsoft Edge, Mozilla Firefox, or Google Chrome will work well.

For all Microsoft Official Courses taught in their entirety that have a corresponding certification exam, an exam voucher is included for each participant.

**Software Needed on Each Student PC**

All Microsoft Azure Databricks training students receive Microsoft official courseware.

For all Microsoft Official Courses taught in their entirety that have a corresponding certification exam, an exam voucher is included for each participant.

**Objectives**

* Understand the core functionalities and applications of Azure Databricks
* Leverage Apache Spark for data processing and analysis within the Databricks environment
* Build, train, and evaluate machine learning models using various frameworks
* Implement MLflow to track experiments, manage models, and streamline deployment
* Apply advanced techniques like AutoML, hyperparameter tuning, and deep learning
* Deploy and manage machine learning models in a production setting

**Outline**

* Exploring Azure Databricks
	+ Introduction to Azure Databricks and its capabilities
	+ Key concepts and workloads
	+ Data governance with Unity Catalog and Microsoft Purview
	+ Hands-on exercise: Exploring Azure Databricks
* Using Apache Spark in Azure Databricks
	+ Introduction to Apache Spark
	+ Creating and managing Spark clusters
	+ Working with data using Spark in notebooks
	+ Data visualization techniques
	+ Hands-on exercise: Using Spark in Azure Databricks
* Training Machine Learning Models in Azure Databricks
	+ Machine learning principles and concepts
	+ Machine learning frameworks supported in Azure Databricks
	+ Data preparation for machine learning
	+ Model training and evaluation
	+ Hands-on exercise: Training a machine learning model
* Managing the Machine Learning Lifecycle
	+ MLflow for experiment tracking, model registry, and deployment
	+ Hands-on exercise: Using MLflow
	+ Hyperparameter tuning with Hyperopt
	+ Hands-on exercise: Optimizing hyperparameters
* Advanced Machine Learning Techniques
	+ AutoML for automated machine learning
	+ Hands-on exercise: Using AutoML
	+ Deep learning concepts and model training with PyTorch
	+ Distributed training with TorchDistributor
	+ Hands-on exercise: Training deep learning models
* Productionizing Machine Learning
	+ Automating data transformations
	+ Model development and deployment strategies
	+ Model versioning and lifecycle management
	+ Hands-on exercise: Managing a machine learning model in production