

DP-3014: Build machine learning solutions using Azure Databricks

Course Description

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.

Course Duration

1 day

Prerequisites

This learning path assumes that you have experience of using Python to explore data and train machine learning models with common open source frameworks, like Scikit-Learn, PyTorch, and TensorFlow.

Course Outline

1 - Explore Azure Databricks

- Get started with Azure Databricks
- Identify Azure Databricks workloads
- Understand key concepts
- Data governance using Unity Catalog and Microsoft Purview
- Module assessment

2 - Use Apache Spark in Azure Databricks

- Get to know Spark
- Create a Spark cluster
- Use Spark in notebooks
- Use Spark to work with data files
- Visualize data
- Module assessment

3 - Train a machine learning model in Azure Databricks

- 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
- Module assessment

4 - Use MLflow in Azure Databricks

- Capabilities of MLflow
- Run experiments with MLflow
- Register and serve models with MLflow
- Module assessment

5 - Tune hyperparameters in Azure Databricks

- Optimize hyperparameters with Optuna
- Review trials
- Scale hyperparameter optimization
- Module assessment

6 - Use AutoML in Azure Databricks

- What is AutoML?
- Use AutoML in the Azure Databricks user interface
- Use code to run an AutoML experiment
- Module assessment

7 - Train deep learning models in Azure Databricks

- Understand deep learning concepts
- Train models with PyTorch
- Distribute PyTorch training with TorchDistributor
- Module assessment

8 - Manage machine learning in production with Azure Databricks

- Automate your data transformations
- Explore model development
- Explore model deployment strategies
- Explore model versioning and lifecycle management
- Module assessment

Who Should Attend

- This course is designed for aspiring data scientists and AI engineers who need to train and manage machine learning models by using Azure Databricks.