Data AI:
Azure Databricks Essentials

WorkshopPLUS

Duration: 3 days

Focus Area: Business/IT Alignment

Difficulty: 300-Advanced

Overview

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform. This course will help you understand Azure Databricks, manage Azure Databricks cluster, develop in Azure Databricks and go through use cases like Streaming, Machine Learning and Data Exploration. Use your existing skills in Python, Scala, R, and SQL, as well as deep learning frameworks. Accelerate big data analytics and artificial intelligence (AI) solutions with Azure Databricks in your organization. This workshop is targeted at Database Administrators, Developers, Data scientists, Data Engineers, and Business Analysts who can collaborate on shared projects in an interactive workspace.

Objectives

After completing this training, students will be able to:

- Understand the features of Apache Spark
- Learn the basic definition of Databricks and its evolution
- Understand usage of notebooks
- Manage notebooks (create, delete, export, import, attach/detach)
- Use notebooks to run commands, create dashboards
- Create and schedule jobs
- Learn the basic mechanics of how Structured Streaming works
- Learn the difference between popular open-source streaming engines
- Learn an overview of some of the more advance features of Structured Streaming like fault tolerance and partitioning
- Understand what options are available for model export
- Explore the MLeap option

Key Takeaways

Course Material

- Learn how to use Azure Databricks to transform Big Data architectures and execute powerful Machine Learning models
- Learn how to manage your workspace and clusters to ensure you get the most performant and cost-effective experience

Hands-on Labs

- Most of the concepts covered above will be supported by hands-on labs and demos.
- Attendees have access to resources and labs for up to 6 months after workshop completion.

Agenda

Day 1

- Introduction to Azure Databricks
- Databricks Development

Day 2

- Data Exploration in Azure Databricks and Visualization in PowerBI
- Structured Streaming with Azure Databricks

Day 3

- Machine Learning with Azure Databricks

Plan for three full days. Early departure on any day is not recommended.

Course Details

Module 1: Introduction to Azure Databricks

- Introduction to Databricks
- Azure Databricks and Capabilities
- HDInsight Vs Azure Databricks
- Pricing in Azure Databricks
- Azure Databricks Artifacts
- Azure Databricks Clusters
- Azure Databricks Workspace

Module 2: Databricks Development

- Azure Databricks Notebooks and Jobs
- Working with Azure Databricks CLI
- Working with Storage Options
- Security with Azure Databricks

Module 3: Data Exploration in Azure Databricks and Visualization in PowerBI

- Data Exploration with Azure Databricks
- Visualization in PowerBI

Module 4: Structured Streaming with Azure Databricks

- Introduction to Structured Streaming
- Developing Structured Streaming Pipelines
- Advanced Pipeline Development
- Real-Time Structured Streaming Jobs

Module 5: Machine Learning with Azure Databricks

- Machine Learning (ML) Capabilities
- Azure Databricks ML Use Case
- Preprocessing Data for the Use Case
- Fit and Evaluate Models in ML Use Case
- Model Export in Azure Databricks

Recommended Qualifications

- Familiarity with Python, SQL, Machine Learning, Big Data on Azure
- Participants that have existing experience performing Python, SQL, and Big Data development will receive the most value from this course.

Hardware Requirements

- An Intel Core-i5-based PC
- USB port
- Microsoft/Windows Live ID to connect to the virtual environment
- 4 GB RAM
- 128 GB HDD
- Windows 7 SP1 or later
- Office 2013 Professional Plus
- Internet access with at least 1 Mbps bandwidth per student.

For more information

Contact your Henson Group representative for further details.