



DATABRICKS TRAINING

Databricks Azure

Master data engineering on Azure with Databricks. Learn to build and manage scalable data solutions in the cloud.

DURATION

Flexible

LEVEL

Intermediate

FORMAT

Instructor-Led

- Globally Recognized Certification

- Expert-Led Hands-On Training

- Flexible Scheduling

Computer Pride | 1st Floor, ICEA Building, Kenyatta Avenue, Nairobi | www.computer-pride.co.ke

COURSE OVERVIEW

This intermediate-level "Databricks Azure" course provides comprehensive training on leveraging the Databricks Lakehouse Platform within the Microsoft Azure ecosystem. Participants will gain in-depth knowledge and hands-on experience in building scalable data engineering pipelines, developing advanced machine learning models, and performing robust analytics directly on Azure. The curriculum focuses on integrating Databricks with key Azure services, ensuring you can design, implement, and manage end-to-end data solutions in a cloud-native environment.

By the end of this course, you will be proficient in utilizing core Databricks technologies like Apache Spark for big data processing, Delta Lake for reliable data lakes, and MLflow for managing the machine learning lifecycle, all while operating within Azure's powerful cloud infrastructure. This expertise is crucial for organizations looking to modernize their data platforms, enable real-time data insights, and accelerate their AI initiatives by combining the power of Databricks with the extensive capabilities of Azure. It equips professionals with the skills to tackle complex data challenges, from ingestion and transformation to analysis and deployment of intelligent applications.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Introduction to Azure Databricks and the Lakehouse Architecture

- Understanding the Databricks Lakehouse Platform
- Overview of Azure Databricks Workspace and Components
- Integrating Databricks with Azure Services (ADLS Gen2, Azure Synapse, Azure DevOps)
- Introduction to Apache Spark on Databricks

02 Module 2: Data Engineering with Delta Lake on Azure

- Working with Delta Lake: ACID Transactions, Schema Enforcement, Time Travel
- Building ELT/ETL Pipelines with Spark SQL and PySpark
- Data Ingestion from various Azure sources (Event Hubs, Blob Storage, Cosmos DB)
- Optimizing Delta Lake Tables (Z-ordering, Compaction)

03 Module 3: Advanced Data Transformations and Performance Tuning

- Complex Spark Transformations and User-Defined Functions (UDFs)
- Structured Streaming for Real-time Data Processing
- Performance Monitoring and Tuning Spark Applications on Azure Databricks
- Debugging and Troubleshooting Spark jobs

04 Module 4: Machine Learning with MLflow on Azure Databricks

- Introduction to MLflow: Tracking, Projects, Models, Registry
- Data Preparation and Feature Engineering for ML models
- Training and Evaluating Machine Learning Models at Scale
- Deploying and Managing ML Models using Azure Machine Learning integration

05 Module 5: Security, Governance, and Operations on Azure Databricks

- Workspace Administration and Access Control (IAM, AAD integration)
- Implementing Data Governance with Unity Catalog
- Monitoring Azure Databricks Workloads

- Automating Workflows with Azure Data Factory and Databricks Jobs
- Best Practices for Production Deployments

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Azure Databricks Data Engineer**
- > **Cloud Machine Learning Engineer (Azure Focus)**
- > **Lakehouse Architect (Azure Platform)**

FREQUENTLY ASKED QUESTIONS

Q: What are the prerequisites for enrolling in the Databricks Azure course?

Participants should have a foundational understanding of cloud concepts, preferably with some familiarity with Azure services. Basic knowledge of Python or SQL is highly recommended for engaging with the hands-on coding exercises and understanding data manipulation concepts effectively. Prior experience with big data concepts or Apache Spark is beneficial but not strictly required, as the course will cover Spark fundamentals.

Q: How does this course prepare me for real-world data engineering or machine learning projects on the Azure platform?

This course is designed with a strong practical emphasis, featuring hands-on labs and real-world scenarios. You'll learn to build end-to-end data pipelines, manage data lakes with Delta Lake, develop and deploy scalable machine learning models using MLflow, and integrate Databricks with various Azure data services. This practical exposure ensures you gain the concrete skills necessary to design, implement, and optimize data and AI solutions within an Azure cloud environment.

Q: Will I gain practical experience with core Databricks components like Delta Lake and MLflow on Azure through this course?

Absolutely. A significant portion of this course is dedicated to hands-on exercises utilizing Delta Lake for building reliable and performant data lakes, including concepts like ACID transactions, schema enforcement, and time travel. You will also extensively use MLflow to manage the entire machine learning lifecycle, from experiment tracking and model versioning to model deployment within the Azure Databricks environment. Practical application of these components is central to the curriculum.

ABOUT COMPUTER PRIDE



East Africa's Premier IT Training Center

With over 15 years of excellence, Computer Pride delivers globally recognized certification programs from leading technology vendors. Located on Kenyatta Avenue in Nairobi's CBD, our state-of-the-art facilities and expert instructors

have empowered thousands of professionals across East Africa.

15+

Years of Excellence

10,000+

Professionals Trained

50+

Vendor Certifications

READY TO ENROL?

Take the next step in your technology career.

Location

1st Flr, ICEA Building
Kenyatta Ave, Nairobi

Email

info@computer-pride.co.ke

Phone

+254 705 900 900

www.computer-pride.co.ke