



DATABRICKS TRAINING

Databricks Aws

*Master Databricks on AWS to build scalable data analytics and machine learning solutions.
Unleash the power of big data 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 AWS course provides a comprehensive deep dive into leveraging the Databricks Lakehouse Platform within the Amazon Web Services (AWS) ecosystem. Participants will gain hands-on expertise in building, deploying, and managing scalable data engineering, data science, and machine learning workloads. You'll master core technologies such as Apache Spark for distributed data processing, Delta Lake for reliable data lakes, and MLflow for robust MLOps, all while integrating seamlessly with essential AWS services like S3, EC2, and IAM. This course focuses on practical application, enabling you to tackle real-world big data challenges.

Upon completion, you will be proficient in designing and implementing end-to-end data solutions, from raw data ingestion and transformation to advanced analytics and machine learning model deployment on a cloud-native, unified platform. This certification empowers data professionals to unlock the full potential of their data assets on AWS, driving faster insights, improved data quality, and more efficient operational workflows, making you an invaluable asset in any data-driven organization looking to modernize its data strategy.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Getting Started with Databricks on AWS

- Introduction to the Databricks Lakehouse Platform
- Understanding the AWS-Databricks architecture and deployment models
- Setting up and configuring Databricks workspaces on AWS
- Navigating the Databricks UI, notebooks, and core components
- Connecting Databricks to AWS services: S3, IAM roles, VPC peering

02 Module 2: Apache Spark Fundamentals on Databricks

- Deep dive into Spark architecture, clusters, and execution model on Databricks
- Working with DataFrames, Datasets, and RDDs in PySpark and Scala
- Mastering Spark SQL for data manipulation and analysis
- Performance optimization techniques for Spark jobs (caching, partitioning, shuffles)
- Introduction to structured streaming for real-time data processing

03 Module 3: Data Ingestion and Transformation with Delta Lake

- Introduction to Delta Lake: ACID transactions, schema enforcement, time travel
- Building reliable data lakes on AWS S3 using Delta Lake format
- Implementing batch and streaming data ingestion strategies (e.g., Auto Loader)
- Designing and executing ETL/ELT pipelines with Databricks notebooks and Delta Live Tables (DLT)
- Handling data quality issues, data versioning, and change data capture (CDC) with MERGE INTO

04 Module 4: Advanced Analytics and Machine Learning Workflows

- Data preparation and feature engineering for machine learning on Databricks
- Introduction to MLflow for experiment tracking, model packaging, and deployment
- Implementing distributed machine learning with Spark MLlib and popular ML frameworks (Scikit-learn, TensorFlow, PyTorch)
- Integrating Databricks ML workflows with AWS SageMaker and other ML services
- Building and deploying ML models using Databricks Model Serving and endpoint management

Module 5: Data Governance, Security, and Operations

- 05**
- Implementing data governance and fine-grained access control with Unity Catalog
 - Monitoring and logging Databricks workloads and clusters on AWS
 - Best practices for performance tuning, cost optimization, and resource management
 - Securing data, workspaces, and connections on Databricks AWS (encryption, network security)
 - Operationalizing data pipelines with Databricks Jobs, Workflows, and alerts

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Cloud Data Engineer (AWS Specialist)**
- > **Machine Learning Engineer (Databricks/AWS)**
- > **Lakehouse Architect**

FREQUENTLY ASKED QUESTIONS

Q: What specific data engineering tasks will I master on Databricks AWS to manage large datasets efficiently?

This course will equip you with the ability to design and implement robust ETL/ELT pipelines using Apache Spark and Delta Lake on AWS S3. You'll master techniques for schema evolution, ACID transactions, data quality enforcement, efficient data ingestion from various sources (batch and streaming via Auto Loader), and leveraging Databricks Workflows to orchestrate complex data flows. This expertise will make you adept at managing petabyte-scale data lakes with high reliability and performance.

Q: What prior experience with AWS or Apache Spark is recommended for an intermediate course in Databricks AWS?

While this is an intermediate course, a foundational understanding of AWS cloud concepts (such as S3, EC2, IAM, VPC) and basic familiarity with SQL or Python programming is highly recommended. Prior exposure to Apache Spark concepts or distributed computing would be beneficial but is not strictly required, as the course will cover Spark fundamentals within the Databricks context to ensure a solid understanding for platform-specific applications.

Q: How does proficiency in Databricks AWS enhance my ability to deploy and manage Machine Learning models in a production environment?

By mastering Databricks AWS, you will gain hands-on experience with MLflow for comprehensive MLOps, enabling you to track experiments, package models, and manage their lifecycle from development to production deployment. You'll learn to leverage Databricks Model Serving for scalable inference and integrate with AWS services to build observable, reproducible, and robust machine learning solutions. This significantly accelerates your ability to bring AI initiatives from concept to operational reality within a cloud-native ecosystem.

ABOUT COMPUTER PRIDE

**computerpride*****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