



DATABRICKS TRAINING

Aws Databricks

Master AWS Databricks to build scalable data engineering and machine learning solutions. Unlock powerful analytics capabilities.

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 AWS Databricks course offers a comprehensive deep dive into leveraging the powerful Databricks Lakehouse Platform within the Amazon Web Services (AWS) ecosystem. Participants will gain practical expertise in unifying data engineering, data science, and machine learning workloads, transforming raw data into actionable insights at scale. The curriculum focuses on building robust, scalable data pipelines, performing advanced analytics, and deploying machine learning models using core Databricks features like Apache Spark, Delta Lake, and MLflow, all seamlessly integrated with essential AWS services.

Throughout the course, you will master the art of data ingestion from various AWS sources, manage data with Delta Lake's ACID properties, and optimize Spark workloads for performance and cost-efficiency on AWS. You'll also explore best practices for MLOps, including experiment tracking, model versioning, and deployment using MLflow. By the end of this program, you'll be equipped with the skills to design, implement, and manage a unified data platform on AWS Databricks, significantly enhancing your ability to deliver high-impact data solutions for any organization.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Getting Started with Databricks on AWS

- Introduction to the Databricks Lakehouse Platform
- Databricks Architecture and Components on AWS
- Setting up a Databricks Workspace in AWS
- Navigating the Databricks UI and Notebooks
- Introduction to Apache Spark on Databricks
- Connecting Databricks to AWS S3 for Storage

02 Module 2: Data Engineering with Delta Lake on AWS

- Ingesting Data from AWS S3, RDS, and Kinesis into Databricks
- Introduction to Delta Lake: ACID Transactions and Schema Enforcement
- Building ETL/ELT Pipelines with PySpark/Scala
- Optimizing Delta Lake Tables (Z-Ordering, Compaction)
- Implementing Data Quality and Governance with Delta Lake
- Streaming Data with Structured Streaming and Delta Lake

03 Module 3: Advanced Spark Programming and Performance on AWS

- Advanced Spark Transformations and Actions
- Working with Different Data Formats (Parquet, ORC, JSON)
- Performance Tuning Spark Jobs on Databricks Clusters
- Understanding Cluster Configurations and Autoscaling on AWS
- Using Databricks Jobs for Orchestration and Automation
- Integrating with AWS Glue Data Catalog

04 Module 4: Machine Learning Workflows with MLflow on AWS

- Introduction to Machine Learning on Databricks
- Feature Engineering and Selection with Spark MLlib
- Model Training and Evaluation on Distributed Datasets
- MLflow for Experiment Tracking and Model Management

- Model Registering and Versioning with MLflow Model Registry
- Deploying and Serving Models (e.g., via SageMaker or Databricks Model Serving)

05 Module 5: Data Governance, Security, and Monitoring on AWS Databricks

- Implementing Unity Catalog for Data Governance and Access Control
- Managing Permissions and User/Group Access in Databricks
- Integrating with AWS IAM for Authentication and Authorization
- Monitoring Databricks Workloads and Clusters using AWS CloudWatch
- Security Best Practices for Databricks Workspaces on AWS
- Cost Optimization Strategies for Databricks on AWS

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **AWS Databricks Data Engineer**
- > **Cloud-Native Machine Learning Engineer (AWS & Databricks)**
- > **Senior Data Platform Architect (AWS Databricks Specialist)**

FREQUENTLY ASKED QUESTIONS

Q: What specific AWS services will I learn to integrate with Databricks during this course, and for what purposes?

This course extensively covers integration with AWS S3 for scalable data storage, AWS IAM for robust authentication and access control within your Databricks workspace, and touches upon services like AWS Glue Data Catalog for metadata management and AWS CloudWatch for monitoring. You will learn to leverage these services to build secure, efficient, and scalable data solutions on the Databricks Lakehouse Platform hosted on AWS.

Q: What prior experience with cloud platforms or data analytics is recommended to excel in this AWS Databricks course?

To get the most out of this intermediate-level course, we recommend participants have a foundational understanding of AWS cloud concepts, including basic navigation of the AWS console and familiarity with services like S3. Prior experience with SQL, Python (for PySpark), and a conceptual understanding of data warehousing or data engineering principles will be highly beneficial, though a deep background in Apache Spark is not strictly required as Spark fundamentals are covered.

Q: How does this course prepare me to build production-ready data pipelines and machine learning workflows leveraging Databricks on AWS?

This course is designed with a strong emphasis on practical application. You will gain hands-on experience in building end-to-end data pipelines using Delta Lake, optimizing Spark performance for large datasets, and implementing MLOps practices with MLflow. Through real-world scenarios and exercises, you'll learn to design scalable architectures, ensure data quality, manage model lifecycles, and monitor your Databricks workloads, all within the context of robust AWS integrations, preparing you for immediate impact in production environments.

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