



DATABRICKS TRAINING

Databricks Bootcamp

Master Databricks and Apache Spark for scalable data engineering and machine learning. Build robust data pipelines and ML models.

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

Dive deep into the Databricks Lakehouse Platform with this intensive Databricks Bootcamp, designed for intermediate professionals looking to master unified data analytics and AI. This course provides comprehensive hands-on training on leveraging Apache Spark for large-scale data processing, understanding the power of Delta Lake for reliable data lakes, and implementing robust machine learning workflows with MLflow. Participants will gain practical experience in building scalable data pipelines, performing advanced analytics, and deploying machine learning models directly within the Databricks environment. By the end of this bootcamp, you'll be adept at transforming raw data into actionable insights, orchestrating complex ETL processes, and developing AI-powered solutions that drive business value. This expertise is crucial for modern data professionals seeking to bridge the gap between data engineering and data science, ensuring faster development cycles and more reliable data infrastructure. Elevate your career by becoming a proficient Databricks practitioner, capable of tackling real-world big data challenges and contributing to data-driven innovation.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Introduction to Databricks and the Lakehouse Platform

- Understanding the Databricks Ecosystem
- Introduction to Lakehouse Architecture and its benefits
- Databricks Workspace Navigation and Configuration
- Clusters, Notebooks, and Jobs
- Data Loading and Ingestion techniques on Databricks

02 Module 2: Apache Spark Fundamentals on Databricks

- Spark Architecture and Execution Model
- Working with Spark DataFrames and SQL
- Transformations and Actions
- Optimizing Spark Performance
- User-Defined Functions (UDFs)

03 Module 3: Data Engineering with Delta Lake

- Introduction to Delta Lake: ACID Transactions, Schema Enforcement, Time Travel
- Implementing ETL/ELT Pipelines with Delta Lake
- Data Versioning and Auditing
- Optimizing Delta Lake Tables (Z-Ordering, Vacuum)
- Stream Processing with Structured Streaming and Delta Lake

04 Module 4: Machine Learning with MLflow and Databricks

- Introduction to Machine Learning on Databricks
- MLflow for Experiment Tracking, Model Management, and Deployment
- Feature Engineering and Vectorizing Data
- Building and Training ML Models (Scikit-learn, TensorFlow, PyTorch)
- Model Deployment and Inference on Databricks

05 Module 5: Advanced Databricks Features and Optimization

- Databricks SQL Analytics and Dashboards
- Working with Unity Catalog for Governance

- Advanced Performance Tuning for Spark and Delta Lake
- Security and Access Control on Databricks
- Automating Workflows with Databricks Jobs and APIs

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Databricks Data Engineer**
- > **Machine Learning Operations (MLOps) Engineer specializing in Databricks**
- > **Lakehouse Architect**

FREQUENTLY ASKED QUESTIONS

Q: What prior knowledge or experience is recommended to get the most out of this Databricks Bootcamp?

This bootcamp is designed for intermediate professionals. A foundational understanding of Python or Scala, SQL, and basic concepts of big data processing (e.g., ETL principles) is highly recommended. Familiarity with cloud environments (AWS, Azure, or GCP) will also be beneficial, though not strictly required, as the course will focus on the Databricks platform itself.

Q: How will this bootcamp specifically prepare me to design and implement a Lakehouse architecture?

This bootcamp dedicates significant modules to the Lakehouse architecture, particularly focusing on Delta Lake. You will gain hands-on experience in building reliable, scalable data pipelines using Delta Lake's ACID properties, schema enforcement, and time travel capabilities. You'll also explore best practices for unifying data warehousing and data lake functionalities, enabling you to confidently design and implement robust Lakehouse solutions from data ingestion to consumption, optimizing for performance and data quality.

Q: Beyond general data roles, what specific types of projects will I be able to lead or contribute to after completing this Databricks Bootcamp?

Upon completion, you'll be equipped to lead projects involving the migration of legacy data warehouses to a modern Lakehouse paradigm, develop end-to-end machine learning operationalization (MLOps) pipelines on Databricks, and construct real-time analytics dashboards powered by Delta Lake and Databricks SQL. You will also be able to optimize large-scale data processing workloads and implement robust data governance strategies using Unity Catalog across diverse data sources and use cases.

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