



DATABRICKS TRAINING

Databricks Data Engineering

*Master Databricks Data Engineering to build robust, scalable data pipelines using Apache Spark.
Learn to transform and manage data for powerful analytics.*

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 Data Engineering course is meticulously designed for professionals aspiring to master the art of building and maintaining robust, scalable data pipelines on the Databricks Lakehouse Platform. You will delve deep into the core technologies that power modern data architectures, including Apache Spark for distributed processing, Delta Lake for ensuring data reliability and transactional capabilities, and Databricks SQL for analytical workloads. The course emphasizes hands-on experience, guiding you through practical scenarios to ingest, transform, and manage vast datasets efficiently and reliably.

Upon completion, you will possess the expertise to design and implement end-to-end data solutions, leverage Databricks notebooks for collaborative development, and optimize performance for various data engineering tasks. This certification will significantly enhance your ability to contribute to data-driven initiatives, enabling organizations to unlock critical insights from their data. It's an invaluable asset for anyone looking to solidify their position as a leading data professional in today's rapidly evolving technological landscape, providing a clear pathway to constructing the foundational data layers essential for analytics, machine learning, and business intelligence.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Introduction to the Databricks Lakehouse Platform

- Overview of the Databricks Lakehouse Architecture
- Introduction to Apache Spark and its ecosystem
- Understanding Delta Lake and its advantages
- Navigating the Databricks Workspace and UI
- Setting up clusters and environments

02 Module 2: Data Ingestion and Transformation with Apache Spark

- Reading and writing various data formats (CSV, JSON, Parquet, ORC)
- Working with DataFrames and Spark SQL for data manipulation
- Advanced DataFrame transformations (joins, aggregations, window functions)
- Optimizing Spark jobs for performance
- Error handling and fault tolerance in Spark applications

03 Module 3: Building Reliable Data Pipelines with Delta Lake

- Introduction to Delta Lake features (ACID transactions, schema enforcement, time travel)
- Implementing Change Data Capture (CDC) patterns with Delta Lake
- Building Medallion Architecture (Bronze, Silver, Gold layers)
- Managing Delta Lake tables (vacuum, optimize, describe history)
- Best practices for Delta Lake performance and cost optimization

04 Module 4: Stream Processing with Structured Streaming

- Fundamentals of Apache Spark Structured Streaming
- Ingesting real-time data from various sources (Kafka, Azure Event Hubs, AWS Kinesis)
- Performing continuous transformations on streaming data
- Managing stateful operations and checkpoints in Structured Streaming
- Monitoring and troubleshooting streaming pipelines

05 Module 5: Advanced Databricks Features and Data Governance

- Introduction to Databricks SQL for analytics and reporting
- Implementing data governance with Unity Catalog
- Monitoring and alerting for Databricks pipelines
- CI/CD practices for Databricks notebooks and jobs
- Integrating Databricks with other cloud services (ADLS, S3, Snowflake)

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Databricks Lakehouse Platform Engineer**
- > **Senior Spark ETL Developer**
- > **Cloud Data Pipeline Architect**

FREQUENTLY ASKED QUESTIONS

Q: What are the recommended prerequisites for enrolling in the Databricks Data Engineering course?

To get the most out of this course, participants should have foundational knowledge of Python or Scala programming, basic SQL querying skills, and an understanding of core data warehousing or ETL concepts. Familiarity with cloud platforms (Azure, AWS, or GCP) is beneficial but not strictly required, as the course focuses on Databricks-specific implementations.

Q: How will this Databricks Data Engineering certification enhance my ability to build real-world data solutions?

This certification provides hands-on expertise in critical areas like building scalable ETL/ELT pipelines using Apache Spark, ensuring data quality and reliability with Delta Lake's transactional capabilities, and implementing real-time data processing with Structured Streaming. You'll gain practical skills to design, deploy, and manage robust data architectures on the Databricks Lakehouse Platform, directly transferable to roles focused on large-scale data ingestion, transformation, and analytics projects.

Q: What specific skills will I acquire that are highly sought after by employers in the data engineering field?

You will master skills in distributed data processing using Apache Spark, specifically with Python or Scala, and become proficient in Delta Lake for creating performant and reliable data lakes. Key skills include implementing the Medallion architecture, optimizing Spark workloads, designing streaming data solutions, and leveraging Databricks' unified platform features for data governance via Unity Catalog. These capabilities are crucial for modern data engineering roles in cloud-native environments.

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