



DATABRICKS TRAINING

Databricks Data Engineer Associate

Become a Databricks Data Engineer Associate. Master data pipelines, Delta Lake, and Apache Spark for robust data solutions.

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

The Databricks Data Engineer Associate course is meticulously designed to equip aspiring and current data professionals with the essential skills to build and manage robust data pipelines using the Databricks Lakehouse Platform. This intermediate-level training dives deep into the core functionalities of Apache Spark and Delta Lake, empowering you to handle large-scale data ingestion, transformation, and storage efficiently. You will gain hands-on experience in developing scalable ETL/ELT processes, ensuring data quality, and leveraging the full potential of a unified data platform for analytics and machine learning workloads. Upon completion, you will be proficient in orchestrating data workflows and optimizing performance on Databricks. This comprehensive course goes beyond theoretical concepts, providing practical scenarios and exercises that simulate real-world data engineering challenges. You'll learn best practices for working with structured, semi-structured, and unstructured data, mastering techniques for data versioning, schema evolution, and ACID transactions with Delta Lake. Successfully completing this certification not only validates your expertise in modern data engineering practices but also significantly enhances your career prospects in an increasingly data-driven world, positioning you as a valuable asset capable of driving impactful data initiatives.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Introduction to the Databricks Lakehouse Platform

- Understanding the Lakehouse Architecture and its benefits
- Introduction to Databricks Workspace and key components
- Working with Databricks Notebooks and Clusters
- Overview of Apache Spark for big data processing
- Introduction to Delta Lake: The Foundation of the Lakehouse

02 Module 2: Core Apache Spark Programming with Python

- Spark DataFrames: Creation, transformations, and actions
- Working with different data sources (CSV, JSON, Parquet)
- Advanced DataFrame operations: Joins, aggregations, window functions
- Optimizing Spark jobs and understanding execution plans
- Introduction to User-Defined Functions (UDFs) and performance considerations

03 Module 3: Building Reliable Data Pipelines with Delta Lake

- Delta Lake fundamentals: ACID properties, schema enforcement, schema evolution
- Data ingestion strategies into Delta Lake (batch and streaming)
- Upserts, deletes, and merges with Delta Lake
- Time travel and data versioning with Delta Lake
- Optimizing Delta Lake tables for performance (Z-ordering, VACUUM, OPTIMIZE)

04 Module 4: Data Ingestion and Transformation Patterns

- Ingesting data from various cloud storage solutions (S3, ADLS Gen2, GCS)
- Implementing Medallion Architecture (Bronze, Silver, Gold layers)
- Structured Streaming with Spark for real-time data processing
- Handling late-arriving data and watermarking in streams
- Error handling and fault tolerance in data pipelines

Module 5: Workflow Orchestration and Data Governance

- 05
- Scheduling and orchestrating jobs using Databricks Jobs
 - Introduction to Databricks SQL for analytics workloads
 - Implementing access control and permissions within Databricks
 - Monitoring and logging data pipeline health and performance
 - Best practices for MLOps integration and data sharing

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Databricks Platform Data Engineer**
- > **Cloud ETL Developer (Spark/Delta Lake Specialist)**
- > **Lakehouse Solutions Architect**

FREQUENTLY ASKED QUESTIONS

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

While there are no strict official prerequisites, candidates will benefit greatly from having a foundational understanding of Python programming, basic SQL knowledge, and familiarity with cloud computing concepts (e.g., AWS, Azure, GCP). Prior experience with big data concepts or Apache Spark is advantageous but not mandatory, as core Spark concepts are covered.

Q: How does this Databricks Data Engineer Associate certification enhance my ability to build scalable data pipelines in a production environment?

This certification validates your proficiency in leveraging Databricks' unified platform to build, deploy, and manage production-grade data pipelines. You will master techniques for ingesting data efficiently, performing complex transformations with Apache Spark, ensuring data reliability and governance using Delta Lake, and orchestrating workflows. This directly translates to the ability to design and implement robust, high-performance data solutions capable of handling enterprise-scale data volumes and velocity.

Q: Beyond technical skills, what career outcomes can I expect from achieving the Databricks Data Engineer Associate certification?

Achieving this certification positions you as a recognized expert in modern data engineering on the Databricks Lakehouse Platform. This opens doors to specialized roles such as Databricks Platform Engineer, Cloud ETL Developer, or even contributing to Lakehouse architecture design. It demonstrates to employers your capability to work with cutting-edge data technologies, solve complex data challenges, and contribute significantly to data-driven decision-making within an organization, leading to increased job opportunities and career advancement.

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