



DATABRICKS TRAINING

Databricks Data Engineering

Master Databricks Data Engineering to build robust, scalable data pipelines on the Lakehouse Platform.

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 from Computer Pride is meticulously designed to equip aspiring and current data professionals with the essential skills to build, manage, and optimize robust data pipelines on the Databricks Lakehouse Platform. Participants will gain deep expertise in leveraging Apache Spark for large-scale data processing, mastering Delta Lake for reliable and performant data lake architectures, and implementing efficient Extract, Transform, Load (ETL) operations. The curriculum emphasizes hands-on application, covering best practices in using SQL, Python, or Scala within the Databricks environment to solve complex data challenges.

Upon completion, learners will be proficient in designing scalable and fault-tolerant data solutions, ensuring data quality and governance, and optimizing performance for big data workloads. This certification not only validates your ability to work effectively with Databricks but also positions you as a critical asset in organizations looking to modernize their data infrastructure and harness the full potential of their data for analytics and machine learning initiatives. Elevate your data engineering capabilities and drive significant value in any data-driven enterprise.

Level	Intermediate
Vendor	Databricks

COURSE CURRICULUM

01 Module 1: Introduction to Databricks Lakehouse & Apache Spark Fundamentals

- Overview of the Databricks Platform and its components
- Understanding the Lakehouse Architecture
- Introduction to Apache Spark: Architecture and Core Concepts
- Working with Spark DataFrames and Spark SQL
- Setting up your Databricks Workspace and Notebooks

02 Module 2: Mastering Delta Lake for Reliable Data Lakes

- Introduction to Delta Lake: ACID Transactions, Schema Enforcement, Time Travel
- Reading and Writing Data with Delta Lake
- Optimizing Delta Lake Tables (Z-ordering, Compaction)
- Implementing Change Data Capture (CDC) with Delta Lake
- Data Versioning and Auditing

03 Module 3: Building ETL/ELT Pipelines with Databricks

- Designing Data Ingestion Strategies (Batch and Streaming)
- Developing ETL pipelines using Spark SQL, Python/Scala
- Data Cleansing, Transformation, and Enrichment techniques
- Orchestration of pipelines using Databricks Jobs
- Working with various data sources and sinks (ADLS, S3, Blob Storage, external databases)

04 Module 4: Advanced Databricks Features & Performance Optimization

- Structured Streaming for Real-time Data Processing
- User-Defined Functions (UDFs) and Advanced Spark APIs
- Performance Tuning of Spark Applications (Caching, Partitioning, Shuffling)
- Monitoring Spark Jobs and Databricks Workflows
- Best practices for production-grade data engineering

Module 5: Data Governance, Security, and Collaboration

- 05**
- Access Control and Permissions in Databricks (Unity Catalog)
 - Data Masking and Encryption techniques
 - Managing Secrets with Databricks Secrets
 - Collaboration features in Databricks (Repos, Git Integration)
 - Introduction to Delta Live Tables (DLT) for declarative pipelines

CAREER PATHWAYS

Graduates of this program are prepared for the following roles:

- > **Databricks Platform Engineer**
- > **Senior Cloud Data Engineer (specializing in Azure/AWS/GCP with Databricks)**
- > **Lakehouse Architect**

FREQUENTLY ASKED QUESTIONS

Q: What prior programming or data experience is required to succeed in this Databricks Data Engineering course?

This intermediate-level course assumes a foundational understanding of SQL and basic programming concepts in either Python or Scala. While prior experience with big data concepts or cloud platforms (AWS, Azure, GCP) is beneficial, it is not strictly required, as core concepts will be revisited in the context of Databricks.

Q: How does this specific Databricks Data Engineering course prepare me for real-world projects and potential Databricks certification exams?

This course is intensely hands-on, focusing on practical labs and real-world scenarios for building and optimizing data pipelines using Databricks, Apache Spark, and Delta Lake. While not a direct exam prep course, the comprehensive curriculum covers the core domains and skills tested in official Databricks Data Engineer Associate certification, significantly enhancing your readiness for such an exam and empowering you to immediately contribute to data engineering projects.

Q: Will I gain skills in optimizing large-scale data processing jobs and managing complex data lake architectures using Databricks?

Absolutely. A significant portion of this course is dedicated to advanced topics such as performance tuning of Spark applications, efficient use of Delta Lake features like Z-ordering and compaction, and understanding Databricks-specific optimizations. You will learn to design robust, scalable, and performant data lake architectures and troubleshoot common big data processing challenges.

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