

Databricks Certified Data Engineer Associate

Build reliable lakehouse pipelines on Databricks with Spark, Delta Lake and Unity Catalog.

DURATION

2 days · 16 hours

LEVEL

Intermediate

CURRICULUM

8 modules · 5 labs

DELIVERY

Live online, Onsite,
Corporate

Course overview

The lakehouse approach combines the flexibility of a data lake with the reliability of a warehouse, and Databricks is one of the most widely adopted platforms for it. This two-day course teaches the core data engineering skills you need to build and run pipelines there.

Day one introduces the Databricks Data Intelligence Platform — workspaces, compute, notebooks and Git folders — and moves straight into transforming data with Spark SQL and PySpark. You will work with Delta Lake tables, understand transactions, time travel, schema enforcement and optimisation, and ingest data incrementally with Auto Loader and COPY INTO.

Day two focuses on production pipelines and governance. You will build a medallion (bronze, silver, gold) pipeline using declarative pipelines with data quality expectations, orchestrate tasks with jobs, and handle change data capture. You will then manage access and lineage with Unity Catalog, share data securely, and finish with an exam readiness session.

Labs run in a live Databricks workspace using realistic datasets, so you finish with working notebooks and pipelines you can adapt.

This course is aligned to the Databricks Certified Data Engineer Associate exam. Databricks updates its exam guide periodically; please check the current version on the Databricks certification site before booking.

Highlights

- 16 hours of live, trainer-led sessions
- Hands-on Databricks workspace for every lab
- Medallion architecture with Delta Lake
- Incremental ingestion, declarative pipelines and jobs
- Practice questions with review

What's included

- Live Databricks workspace labs
- Spark SQL and PySpark transformations
- Delta Lake operations and optimisation
- Auto Loader incremental ingestion
- Declarative pipelines with quality expectations
- Unity Catalog governance
- Course completion certificate from Ace Skills Consulting

Who this course is for

Who should attend

- Data engineers new to Databricks
- ETL and data warehouse developers
- Data analysts moving into engineering
- Cloud engineers supporting data platforms

Prerequisites

- Working knowledge of SQL
- Basic Python is helpful
- Familiarity with data warehousing concepts

Detailed curriculum — 2 days, 16 hours, 8 modules, 5 hands-on labs

Day 1 Platform, Spark and Delta Lake

8 hours

1.1 Databricks platform

- Lakehouse architecture
- Notebooks and Git folders
- Workspaces and compute
- Serverless and SQL warehouses

1.2 Transforming data

- Spark SQL
- Complex types and higher-order functions
- PySpark DataFrames
- User-defined functions

LAB Clean and transform raw JSON with Spark SQL and PySpark · 50 min

1.3 Delta Lake

- ACID transactions
- Schema enforcement and evolution
- Time travel
- MERGE, OPTIMIZE and liquid clustering

LAB Upsert and time travel on Delta tables · 45 min

1.4 Incremental ingestion

- Auto Loader
- Structured Streaming basics
- COPY INTO

LAB Ingest files incrementally with Auto Loader · 35 min

Day 2 Pipelines, Governance and Exam Readiness

8 hours

2.1 Declarative pipelines

- Medallion architecture
- Data quality expectations
- Streaming tables and materialised views
- Change data capture

LAB Build a bronze-silver-gold pipeline with expectations · 60 min

2.2 Orchestration

- Jobs and tasks
- Scheduling and alerts
- Dependencies and parameters
- Repair and rerun

2.3 Governance

- Unity Catalog object model
- Lineage
- Permissions and grants
- Delta Sharing overview

LAB Secure tables with Unity Catalog · 35 min

2.4 Exam readiness

- Exam guide review
- Practice questions with review
- Question styles

Skills and tools

Skills you will build

Databricks

Lakehouse architecture

Spark SQL

PySpark

Delta Lake

Auto Loader

Declarative pipelines

Medallion architecture

Databricks Jobs

Unity Catalog

Data quality

Tools used in class

Databricks

Apache Spark

Delta Lake

Unity Catalog

Databricks SQL

Git

Certification exam

This course is aligned to the exam below and prepares you for it. The exam itself is booked and taken with the vendor.

Exam	Databricks Certified Data Engineer Associate
Exam code	Data Engineer Associate
Questions	45
Duration	90 minutes
Format	Multiple choice
Delivery	Online proctored
Validity	2 years
Official page	https://www.databricks.com/learn/certification

Exam skill areas covered

Databricks Certified Data Engineer Associate

Exam details are pending verification. Confirm the current exam code, format and fee on the official vendor page before booking.

Frequently asked questions

Do I need a Databricks account?

No. We provide workspace access for all labs.

Is Python required?

SQL is the main language; basic Python helps for the PySpark sections and is taught as needed.

How current is the course?

We update content when Databricks revises the exam guide. Please confirm the current guide before booking.

Which cloud does Databricks run on in the labs?

Databricks behaves the same on AWS, Azure and Google Cloud; differences are noted where relevant.

Book a batch or talk to an advisor

Course page: aceskillsconsulting.com/course/databricks-certified-data-engineer-associate · Email: info@aceskillsconsulting.com

Chennai office

Guindy, Chennai, Tamil Nadu

Delhi office

B-95, Old Post Office Street, Shakarpur

New Delhi 110092

Office: 98180 12523

Databricks is a trademark of Databricks, Inc. All product names, logos, brands, trademarks and registered trademarks are the property of their respective owners. Ace Skills Consulting is an independent training provider; unless explicitly stated, it is not affiliated with, endorsed by or sponsored by any vendor mentioned on this site. Exam details are provided for guidance and may change — always confirm on the vendor's official website. Curriculum generated on 23 Oct 2026; topics and order may be adjusted by the trainer to suit each batch.