

Databricks Certified Data Engineer Associate

Course Duration ~40 Hours · Exam-Aligned Curriculum

Intelligence Platform

Development & Ingestion

Processing & Transformations

Productionizing Pipelines

Governance & Quality

In short: This is an instructor-led certification program mapped 1:1 to the official Databricks Certified Data Engineer Associate exam guide (live as of July 25, 2025) — the entry-level credential in the Databricks data engineering track. You learn to complete introductory data engineering tasks on the Databricks Data Intelligence Platform: ingesting with Auto Loader, performing ETL with Spark SQL and PySpark, building Medallion-style pipelines with Lakeflow Declarative Pipelines, deploying workflows with Databricks Asset Bundles and serverless compute, and governing data with Unity Catalog, Delta Sharing, and Lakehouse Federation — finishing with full-length mock exams and a lakehouse capstone. On completion you are ready to pass the exam and take on Data Engineer and Analytics Engineer roles.

Certification	Databricks Certified Data Engineer Associate	Offered by	Databricks
Level	Associate — entry-level credential in the Databricks data engineering track	Platform	Databricks Data Intelligence Platform
Prerequisite	None required; course attendance plus ~6 months of hands-on Databricks experience is highly recommended		
Format	45 scored multiple-choice questions (may include unscored items)	Time limit	90 minutes
Delivery	Online proctored or test center; no test aids allowed	Validity	2 years (recertify via the current live exam)
Code samples	SQL where possible, PySpark (Python) otherwise	Exam guide	Live as of July 25, 2025

Section 1: Databricks Intelligence Platform

ESTIMATED TIME

4–5 Hours

TOPICS

- Enable features that simplify data layout decisions and optimize query performance
- Explain the value of the Data Intelligence Platform
- Identify the applicable compute to use for a specific use case

HANDS-ON

- Workspace tour: clusters, SQL warehouses, and platform features
- Compare compute options against workload scenarios

Assessment: Quiz

Section 2: Development and Ingestion

ESTIMATED TIME

7–8 Hours

TOPICS

- Use Databricks Connect in a data engineering workflow
- Determine the capabilities of Notebooks functionality
- Classify valid Auto Loader sources and use cases
- Demonstrate knowledge of Auto Loader syntax
- Use Databricks' built-in debugging tools to troubleshoot a given issue

HANDS-ON

- Build an Auto Loader ingestion stream from cloud storage
- Debug a failing notebook job with the built-in tools

Assessment: Quiz + Lab

Section 3: Data Processing & Transformations

ESTIMATED TIME

10–11 Hours

TOPICS

- Describe the three layers of the Medallion Architecture and the purpose of each layer in a data processing pipeline
- Classify the cluster type and configuration for optimal performance based on the scenario
- Explain the advantages of Lakeflow Declarative Pipelines (LDP) for ETL in Databricks
- Implement data pipelines using LDP
- Identify DDL (Data Definition Language) / DML features
- Compute complex aggregations and metrics with PySpark DataFrames

HANDS-ON

- Build a bronze → silver → gold pipeline with LDP
- Compute complex aggregations with PySpark DataFrames and UDFs

Assessment: Quiz + Lab

Section 4: Productionizing Data Pipelines

ESTIMATED TIME

8–9 Hours

TOPICS

- Identify the difference between Databricks Asset Bundles (DABs) and traditional deployment methods
- Identify the structure of Asset Bundles
- Deploy a workflow, and repair and rerun a task in case of failure
- Use serverless for hands-off, auto-optimized compute managed by Databricks
- Analyze the Spark UI to optimize a query

HANDS-ON

- Package and deploy a workflow with Databricks Asset Bundles
- Repair/rerun a failed task and tune a query via the Spark UI

Assessment: Quiz + Lab

Section 5: Data Governance & Quality

ESTIMATED TIME

7–8 Hours

TOPICS

- Explain the difference between managed and external tables
- Identify how to grant permissions to users and groups within Unity Catalog (UC)
- Identify key roles in Unity Catalog
- Identify how audit logs are stored
- Use lineage features in Unity Catalog
- Use the Delta Sharing feature available with Unity Catalog to share data
- Identify the advantages and limitations of Delta Sharing
- Identify types of Delta Sharing — Databricks-to-Databricks vs. external system
- Analyze the cost considerations of data sharing across clouds
- Identify use cases of Lakehouse Federation when connected to external sources

HANDS-ON

- Configure Unity Catalog permissions, roles, and lineage for a pipeline
- Share a dataset via Delta Sharing and query a federated external source

Assessment: Quiz + Lab

Exam Readiness & Mock Exams**ESTIMATED TIME**

3–4 Hours

TOPICS

- Section-by-section review and weak-area drills
- Question strategy and elimination techniques (no test aids allowed)
- Recertification path: 2-year validity via the current live exam

HANDS-ON

- Full-length timed mock exams with detailed review
- Capstone: end-to-end lakehouse pipeline from ingestion to governed sharing

Assessment: Capstone + Mock Exams