

## DATA ENGINEERING AND ANALYTICS

# Modern Analytics with Azure Databricks

Level: Practitioner • 2 days (expandable to 3) • Virtual, In-person

## Overview

Plenty of organizations have data in a lakehouse; far fewer reliably turn it into answers people trust. Azure Databricks gives analysts and analytics engineers a single platform for the whole path from raw data to insight, but that breadth is also the trap: it is easy to get lost in notebooks, clusters, and configuration before delivering a single useful dashboard. The skill this course teaches is the analytical workflow itself, using the platform in service of the question.

This is a hands-on, practitioner course. It is aligned with the ground covered by Microsoft's DP-3011 and follows the natural gradient of analytical work: get oriented on the platform, explore and understand the data, shape it into clean, queryable Delta tables, then deliver insight through Databricks SQL, dashboards, and the BI tools downstream. In keeping with a less-but-deeper philosophy, it stays focused on the analytics path; building production ingestion pipelines is the subject of *Data Engineering with Databricks*. Every module ends with a lab against a realistic dataset, and each module builds on the one before.

## Who Should Attend

- Data analysts and BI developers moving onto Azure Databricks
- Analytics engineers who shape lakehouse data for reporting and self-service use
- Data scientists and engineers who need the analytical side of the platform, not just the pipelines

## Prerequisites

- Solid SQL: joins, aggregation, and filtering (see *SQL Querying and T-SQL Fundamentals* for a foundation)
- Basic Python familiarity helps for the notebook modules but is not assumed
- General familiarity with Azure; no prior Databricks experience required

## What You Will Learn

- Navigate the Azure Databricks workspace: notebooks, clusters, SQL warehouses, and the lakehouse
- Explore and profile data with Spark SQL and PySpark in notebooks
- Clean and shape data into analysis-ready Delta tables
- Write analytical queries in Databricks SQL, including window functions and summary tables
- Build dashboards and alerts, and connect Power BI to the lakehouse
- Share governed, trustworthy results using Unity Catalog and sensible refresh and cost practices

## Course Outline

### Day one: the platform and the data

- Getting Oriented on Azure Databricks
  - The lakehouse in one picture: where analytics fits alongside engineering and ML

- Workspaces, clusters, and SQL warehouses: what to use for which kind of work
- Notebooks as the analyst's workbench
- Lab: connect to a workspace, attach compute, and run first queries against sample data
- Exploring and Profiling Data
  - Spark SQL and PySpark for exploration: the handful of operations that do most of the work
  - Profiling a dataset: distributions, nulls, duplicates, and outliers
  - Documenting what you find so the analysis is reproducible
  - Lab: profile a realistic raw dataset and record its quality problems
- Shaping Analysis-Ready Data
  - Cleaning and transforming with SQL and DataFrames
  - Delta tables for analytics: why they are more trustworthy than files
  - Organizing gold tables that answer business questions directly
  - Lab: transform the profiled data into clean, documented Delta tables

### **Day two: from queries to insight**

- Analytical SQL on Databricks
  - Databricks SQL warehouses and the SQL editor
  - Window functions, rollups, and the query patterns behind real business questions
  - Saving and organizing queries a team can reuse
  - Lab: answer a set of business questions with analytical SQL over the gold tables
- Dashboards and Delivery
  - Building dashboards in Databricks: visualizations, parameters, and refresh schedules
  - Alerts: letting the data tell people when something changed
  - Connecting Power BI to the lakehouse, and choosing between the two front ends
  - Lab: build a dashboard with a scheduled refresh and one meaningful alert
- Trustworthy, Governed Analytics
  - Unity Catalog for analysts: permissions, lineage, and discovering certified data
  - Performance and cost habits: warehouse sizing, caching, and query hygiene
  - Handing off: making your analysis something others can maintain
  - Lab: apply governance to the course assets and review a peer's dashboard for trust and clarity

### **Extended Version**

The three-day version keeps the same gradient and adds depth and a complete delivery cycle:

- Deeper analytical SQL: cohort, funnel, and time-series patterns
- A fuller Power BI integration workflow, including semantic model considerations
- An introduction to notebooks for lightweight statistical analysis and forecasting
- A capstone that takes a raw dataset and a business brief through to a governed, refreshing dashboard, presented and defended