**Welcome to Advanced Data Engineering with Databricks!**

In this course, participants will build upon their existing knowledge of Apache Spark, Delta Lake, and Delta Live Tables to unlock the full potential of the data lakehouse by utilizing the suite of tools provided by Databricks. This course places a heavy emphasis on designs favoring incremental data processing, enabling systems optimized to continuously ingest and analyze ever-growing data. By designing workloads that leverage built-in platform optimizations, data engineers can reduce the burden of code maintenance and on-call emergencies, and quickly adapt production code to new demands with minimal refactoring or downtime. The topics in this course should be mastered prior to attempting the Databricks Certified Data Engineering Professional exam.

Note: This version of Advanced Data Engineering with Databricks was released in January 2024 and is an update to the course in the Databricks Academy by the title: Advanced Data Engineering with Databricks (2023). Although both courses can help you prepare for the Databricks Professional Data Engineering Certification Exam, we recommend preparing using this updated version of the course. For the latest about what is covered in the exam, view the exam guide accessible [from this page](https://www.databricks.com/learn/certification/data-engineer-professional).

**Course goals**

By the end of our time together, you’ll be able to:

* Design databases and pipelines optimized for the Databricks Data Intelligence Platform.
* Implement efficient incremental data processing to validate and enrich data driving business decisions and applications.
* Leverage Databricks-native features for managing access to sensitive data and fulfilling right-to-be-forgotten requests
* Manage code promotion, task orchestration, and production job monitoring using Databricks tools

**Prerequisites**

At a minimum, you should be familiar with the following before attempting to take this content:

* Ability to perform basic code development tasks using the Databricks Data Engineering & Data Science workspace (create clusters, run code in notebooks, use basic notebook operations, import repos from git, etc)
* Intermediate programming experience with PySpark
  + Extract data from a variety of file formats and data sources
  + Apply a number of common transformations to clean data
  + Reshape and manipulate complex data using advanced built-in functions
* Intermediate programming experience with Delta Lake (create tables, perform complete and incremental updates, compact files, restore previous versions etc.)
* Beginner experience configuring and scheduling data pipelines using the Delta Live Tables (DLT) UI
* Beginner experience defining Delta Live Tables pipelines using PySpark
  + Ingest and process data using Auto Loader and PySpark syntax
  + Process Change Data Capture feeds with APPLY CHANGES INTO syntax
  + Review pipeline event logs and results to troubleshoot DLT syntax

The data engineering skills required can be obtained by taking the Data Engineering with Databricks course in Databricks Academy.

**Technical considerations**

This course was designed to work with DBR 12.2 LTS.

Next, we'll discuss course logistics.