

Data Engineering with Databricks

Code:	DEWD
Length:	2 days
URL:	View Online

This course begins with a review of programming with Spark APIs and an introduction to key terms and definitions of Databricks data engineering tools, followed by an overview of DB Connect, the Spark UI, and writing testable code. Participants will learn about the Cloud Data Platform in terms of data architecture concepts and will build an end-to-end OLAP data pipeline using Delta Lake with batch and streaming data, learning best practices throughout. Participants who wish to dive deeper into tuning and optimization can take the Advanced Data Engineering with Databricks course.

Skills Gained

- Build an end-to-end batch and streaming OLAP data pipeline
- Make data available for consumption by downstream stakeholders using specified design patterns
- Apply Databricks' recommended best practices in engineering a single source of truth Delta architecture

Who Can Benefit

- Data Engineers and Machine Learning Engineers

Prerequisites

- Intermediate to advanced programming skills in Python or Scala
 - Intermediate to advanced SQL skills
 - Beginning experience using the Spark DataFrames API
 - Beginning knowledge of general data engineering concepts
 - Beginning knowledge of the core features and use cases of Delta Lake
-