

# Microsoft - Implement a data engineering solution with Azure Databricks

Download Whitepaper: Accelerate Your Modernization Efforts with a Cloud-Native Strategy  
Get Your Free Copy Now

**Course Number: DP-3027**

**Duration: 1 days**

## Overview

### Course Description

Important

This course will be available in Q2

Learn how to harness the power of Apache Spark and powerful clusters running on the Azure Databricks platform to run large data engineering workloads in the cloud.

### Audience Profile

Data Engineers, Data Scientists, ELT Developers learn how to harness the power of Apache Spark and powerful clusters running on the Azure Databricks platform to run large data engineering workloads in the cloud.

### Audience

## Course Details

### Outline

- Perform incremental processing with spark structured streaming
  - Set up real-time data sources for incremental processing
  - Optimize Delta Lake for incremental processing in Azure Databricks
  - Handle late data and out-of-order events in incremental processing
  - Monitoring and performance tuning strategies for incremental processing in Azure Databricks
  - Exercise - Real-time ingestion and processing with Delta Live Tables with Azure Databricks
  - Module assessment
- Implement streaming architecture patterns with Delta Live Tables
  - Event driven architectures with Delta Live tables
  - Ingest data with structured streaming
  - Maintain data consistency and reliability with structured streaming
  - Scale streaming workloads with Delta Live tables
  - Exercise - end-to-end streaming pipeline with Delta Live tables
  - Module assessment
- Optimize performance with Spark and Delta Live Tables
  - Optimize performance with Spark and Delta Live Tables
  - Perform cost-based optimization and query tuning
  - Use change data capture (CDC)
  - Use enhanced autoscaling
  - Implement observability and data quality metrics
  - Exercise - optimize data pipelines for better performance in Azure Databricks
  - Module assessment
- Implement CI/CD workflows in Azure Databricks
  - Implement version control and Git integration
  - Perform unit testing and integration testing
  - Manage and configure your environment
  - Implement rollback and roll-forward strategies
  - Exercise - Implement CI/CD workflows

- Module assessment
- Automate workloads with Azure Databricks Jobs
  - Implement job scheduling and automation
  - Optimize workflows with parameters
  - Handle dependency management
  - Implement error handling and retry mechanisms
  - Explore best practices and guidelines
  - Exercise - Automate data ingestion and processing
  - Module assessment
- Manage data privacy and governance with Azure Databricks
  - Implement data encryption techniques in Azure Databricks
  - Manage access controls in Azure Databricks
  - Implement data masking and anonymization in Azure Databricks
  - Use compliance frameworks and secure data sharing in Azure Databricks
  - Use data lineage and metadata management
  - Implement governance automation in Azure Databricks
  - Exercise - Practice the implementation of Unity Catalog
  - Module assessment
- Use SQL Warehouses in Azure Databricks
  - Get started with SQL Warehouses
  - Create databases and tables
  - Create queries and dashboards
  - Exercise - Use a SQL Warehouse in Azure Databricks
  - Module assessment
- Run Azure Databricks Notebooks with Azure Data Factory
  - Understand Azure Databricks notebooks and pipelines
  - Create a linked service for Azure Databricks
  - Use a Notebook activity in a pipeline
  - Use parameters in a notebook
  - Exercise - Run an Azure Databricks Notebook with Azure Data Factory
  - Module assessment