Implement Data Engineering Solutions using Azure Databricks Training

Duration: **4 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Implement Data Engineering Solutions using Azure Databricks is a comprehensive, hands-on course focused on developing modern data engineering solutions with Azure Databricks and the lakehouse architecture. The program explores how scalable data pipelines can be designed, developed, managed and optimised to support enterprise analytics and machine learning workloads.

Learners progress through an end-to-end data engineering journey, beginning with Azure Databricks environment configuration and continuing through data governance, ingestion, transformation, pipeline orchestration and production workload management.

A major area of the course is establishing centralised and secure data governance with**Unity Catalog**. Learners examine metastores, catalogs, schemas and tables while exploring how access can be managed across workspaces, how appropriate permissions can be applied, and how data lineage and auditing capabilities support wider governance requirements.

For data processing, the course covers the ingestion, preparation and transformation of batch and streaming data using SQL and Python. **Delta Lake** is introduced as a foundation for reliable and scalable data storage, alongside practical approaches to data quality, cleansing, validation and transformation optimisation.

Later sections focus on building dependable data pipelines using Databricks jobs and workflows, introducing CI/CD practices, monitoring production workloads, troubleshooting issues and improving both performance and cost efficiency.

Through instructor-led sessions and practical labs aligned with Microsoft Learn, learners develop the confidence to apply these data engineering techniques within realistic enterprise environments.

Prerequisites

Participants should have:

- Experience writing SQL queries
- Experience using Python and notebooks
- An understanding of data engineering and data warehousing concepts
- Familiarity with Azure services
- Knowledge of foundational security principles
- Basic version control experience using Git

Implement Data Engineering Solutions using Azure Databricks Training

Duration: **4 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

Who Should Attend

This course is designed for:

- Data engineers responsible for developing and maintaining data pipelines
- Analytics professionals transitioning towards data engineering roles
- Cloud professionals working with Azure data platforms
- Organisations and technical teams looking to develop scalable lakehouse solutions with Databricks

What You Will Learn

By the end of the course, learners will be able to:

- Configure and manage Azure Databricks workspaces and compute resources.
- Implement secure and scalable data governance with Unity Catalog.
- Ingest, prepare and transform data using SQL, Python and Delta Lake.
- Develop suitable approaches for batch and streaming data processing.
- Design and orchestrate reliable pipelines for production workloads.
- Monitor pipeline and data processing performance.
- Apply optimisation techniques to improve performance and cost efficiency.
- Deploy and maintain enterprise-grade data engineering solutions on Azure.

Training Outline

1. Setting Up and Configuring an Azure Databricks Environment

This section introduces the Azure Databricks architecture and establishes the development environment required for modern data engineering.

- Understanding Azure Databricks architecture
- Exploring core Databricks components
- Configuring workspaces for collaborative data engineering
- Creating clusters and compute resources for different workloads
- Managing compute resources
- Integrating development environments with Git
- Applying version control practices
- Implementing identity and access controls with Microsoft Entra ID
- Establishing foundational governance practices across workspaces

2. Securing and Governing Data with Unity Catalog

This section focuses on the centralised and controlled management of enterprise data assets.

- Understanding Unity Catalog architecture
- Working with metastores
- Organising catalogs, schemas and tables
- Implementing centralised data access across workspaces

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

- Applying Role-Based Access Control (RBAC)
- Managing permissions for secure data operations
- Tracking data lineage
- Using auditing capabilities
- Applying governance practices to support compliance and organisational standards

3. Preparing and Processing Data with Azure Databricks

Learners explore how data can be ingested, transformed and prepared for analytical and enterprise workloads.

- Designing batch data ingestion strategies
- Working with streaming ingestion
- Preparing and transforming data with SQL
- Processing data with Python
- Using Delta Lake for reliable and scalable data storage
- Preparing structured datasets
- Implementing data quality checks
- Applying data cleansing processes
- Validating data
- Optimising transformations for performance and scalability

4. Deploying Data Pipelines and Workloads

This section considers how developed data processes can be orchestrated and moved into production.

- Building data processes with Databricks jobs
- Orchestrating pipelines with workflows
- Managing pipeline dependencies
- Applying Continuous Integration and Continuous Deployment to Databricks workloads
- Moving workloads from development into production
- Establishing repeatable deployment practices

5. Monitoring and Optimising Production Workloads

The final section focuses on maintaining reliable and efficient data engineering solutions after deployment.

- Monitoring pipeline performance
- Identifying production issues
- Troubleshooting workloads
- Improving data processing performance
- Optimising resource utilisation
- Applying techniques to reduce operational costs
- Managing production workloads at enterprise scale
- Maintaining enterprise data engineering solutions