# Data Engineering on AWS Training

Duration: **3 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

## Overview

**Data Engineering on AWS** is a **3-day intermediate course** designed for professionals who want to master data engineering solutions using AWS services.

Through **hands-on labs, theory, and real-world exercises**, participants learn how to **design, build, optimize, and secure** modern data architectures — including **data lakes, warehouses, and pipelines (batch and streaming)**.

The course prepares learners for the **AWS Certified Data Engineer – Associate** certification.

## Prerequisites

Participants should have:

- Basic understanding of **machine learning concepts**,
- Familiarity with **Python** and common data libraries (NumPy, Pandas, Scikit-learn),
- Knowledge of **cloud computing and AWS fundamentals**,
- Experience with **SQL** recommended,
- Familiarity with **Git** is beneficial.

## Who Should Attend

This course is ideal for:

- Data engineers and architects working with AWS,
- Professionals building or managing large-scale data platforms,
- Teams migrating on-premises data infrastructure to AWS,
- Candidates preparing for the **AWS Data Engineer Associate (DEA-C01)** exam.

# Data Engineering on AWS Training

Duration: **3 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

## What You Will Learn

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

- Understand data engineering roles and key AWS services,
- Design and implement **data lakes and warehouses** on AWS,
- Apply **performance tuning, orchestration, and security** best practices,
- Build and manage **batch and streaming data pipelines**,
- Optimize data architectures for scalability, cost, and compliance.

## Outline

**Day 1**

**Module 1: Data Engineering Roles and Key Concepts**

- Role of a Data Engineer
- Key functions of a Data Engineer
- Data Personas
- Data Discovery
- AWS Data Services

**Module 2: AWS Data Engineering Tools and Services**

- Orchestration and Automation
- Data Engineering Security
- Monitoring
- Continuous Integration and Continuous Delivery
- Infrastructure as Code
- AWS Serverless Application Model
- Networking Considerations
- Cost Optimization Tools

**Module 3: Designing and Implementing Data Lakes**

- Data lake introduction
- Data lake storage
- Ingest data into a data lake
- Catalog data
- Transform data
- Server data for consumption
- Hands-on lab: Setting up a Data Lake on AWS

**Module 4: Optimizing and Securing a Data Lake Solution**

- Open Table Formats
- Security using AWS Lake Formation

BILGINÇ IT Academy

info@bilginc.com

bilginc.com


- Setting permissions with Lake Formation
- Security and governance
- Troubleshooting
- Hand-on lab: Automating Data Lake Creation using AWS Lake Formation Blueprints

**Day 2**

**Module 5: Data Warehouse Architecture and Design Principles**

- Introduction to data warehouses
- Amazon Redshift Overview
- Ingesting data into Redshift
- Processing data
- Serving data for consumption
- Hands-on Lab: Setting up a Data Warehouse using Amazon Redshift Serverless

**Module 6: Performance Optimization Techniques for Data Warehouses**

- Monitoring and optimization options
- Data optimization in Amazon Redshift
- Query optimization in Amazon Redshift
- Orchestration options

**Module 7: Security and Access Control for Data Warehouses**

- Authentication and access control in Amazon Redshift
- Data security in Amazon Redshift
- Auditing and compliance in Amazon Redshift
- Hands-on lab: Managing Access Control in Redshift

**Module 8: Designing Batch Data Pipelines**

- Introduction to batch data pipelines
- Designing a batch data pipeline
- AWS services for batch data processing

**Module 9: Implementing Strategies for Batch Data Pipeline**

- Elements of a batch data pipeline
- Processing and transforming data
- Integrating and cataloging your data
- Serving data for consumption
- Hands-on lab: A Day in the Life of a Data Engineer

**Day 3**

**Module 10: Optimizing, Orchestrating, and Securing Batch Data Pipelines**

- Optimizing the batch data pipeline
- Orchestrating the batch data pipeline
- Securing the batch data pipeline
- Hands-on lab: Orchestrating Data Processing in Spark using AWS Step Functions

**Module 11: Streaming Data Architecture Patterns**

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com


- Introduction to streaming data pipelines
- Ingesting data from stream sources
- Streaming data ingestion services
- Storing streaming data
- Processing Streaming Data
- Analyzing Streaming Data with AWS Services
- Hands-on lab: Streaming Analytics with Amazon Managed Service for Apache Flink

**Module 12: Optimizing and Securing Streaming Solutions**

- Optimizing a streaming data solution
- Securing a streaming data pipeline
- Compliance considerations
- Hands-on lab: Access Control with Amazon Managed Streaming for Apache Kafka