Building Batch Data Pipelines on Google Cloud Training

Duration: **1 Day**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Overview

In this intermediate course, you will learn to design, build, and optimize robust batch data pipelines on Google Cloud. Moving beyond fundamental data handling, you will explore large-scale data transformations and efficient workflow orchestration, essential for timely business intelligence and critical reporting.

Get hands-on practice using Dataflow for Apache Beam and Serverless for Apache Spark (Dataproc Serverless) for implementation, and tackle crucial considerations for data quality, monitoring, and alerting to ensure pipeline

reliability and operational excellence. A basic knowledge of data warehousing, ETL/ELT, SQL, Python, and Google Cloud concepts is recommended.

Prerequisites

Participants should have:

- Basic proficiency with Data Warehousing and ETL/ELT concepts
- Basic proficiency in SQL
- Basic programming knowledge (Python recommended)
- Familiarity with gcloud CLI and the Google Cloud console
- Familiarity with core Google Cloud concepts and services

Target audience

This course is designed for:

- Data Engineers
- Data Analysts

What You Will Learn

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

- Determine whether batch data pipelines are the correct choice for your business use case.
- Design and build scalable batch data pipelines for high-volume ingestion and transformation.
- Implement data quality controls within batch pipelines to ensure data integrity.
- Orchestrate, manage, and monitor batch data pipeline workflows, implementing error handling and observability using logging and monitoring tools.