Data Engineering on Google Cloud Eğitimi

Süre: **4 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Overview

Kuruluşlar her geçen gün daha fazla veriyi farklı kaynaklardan topluyor ve bu verileri hızlı, güvenilir ve ölçeklenebilir sistemlerle işleyebilmek büyük önem taşıyor. Google Cloud'un modern veri platformu; veri toplama, depolama, dönüştürme ve analiz süreçlerini tek bir ekosistem içerisinde yönetebilme imkânı sunmaktadır.

Data Engineering on Google Cloud eğitimi, katılımcılara Google Cloud servislerini kullanarak uçtan uca veri işleme çözümleri geliştirme becerisi kazandırmayı amaçlayan uygulamalı bir programdır. Eğitim boyunca gerçek laboratuvar çalışmaları ve örnek senaryolar üzerinden veri gölleri, veri ambarları, ETL/ELT süreçleri, batch ve streaming veri işleme mimarileri ele alınmaktadır.

Katılımcılar ayrıca Google Cloud'un sunucusuz veri işleme servislerini kullanarak yüksek performanslı veri pipeline'ları oluşturmayı ve büyük veri kümelerinden anlamlı içgörüler elde etmeyi öğrenirler.

Eğitimde Kullanılan Google Cloud Servisleri

- BigQuery
- Bigtable
- Cloud Storage
- Cloud SQL
- Cloud Spanner
- Dataproc
- Dataflow
- Cloud Data Fusion
- Cloud Composer
- Pub/Sub

Prerequisites

Katılımcıların aşağıdaki konularda temel bilgi sahibi olmaları önerilir:

- Google Cloud Console ve Cloud Shell kullanımı
- SQL sorgulama bilgisi
- ETL veya veri entegrasyonu süreçleri
- Veri modelleme kavramları
- Python gibi yaygın programlama dillerinden en az biriyle geliştirme deneyimi

Data Engineering on Google Cloud Eğitimi

Süre: **4 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Who Should Attend

Bu eğitim aşağıdaki roller için uygundur:

- Data Engineer
- Database Administrator (DBA)
- System Administrator
- Cloud Engineer
- Analytics Engineer
- Veri Platformu Uzmanları
- Google Cloud üzerinde veri çözümleri geliştiren yazılım profesyonelleri

What You Will Learn

Eğitim sonunda katılımcılar;

- Google Cloud üzerinde ölçeklenebilir veri işleme mimarileri tasarlayabilir.
- Batch ve gerçek zamanlı veri akışlarını yönetebilen pipeline'lar oluşturabilir.
- Dataflow kullanarak otomatik ölçeklenebilir veri işleme süreçleri geliştirebilir.
- BigQuery ile büyük veri kümelerini analiz ederek iş kararlarını destekleyen çıktılar üretebilir.
- Dataproc üzerinde Apache Spark kullanarak yapılandırılmış ve yapılandırılmamış verileri işleyebilir.
- Gerçek zamanlı veri akışlarından anlık analizler elde edebilir.
- Veri gölü ve veri ambarı mimarilerini Google Cloud servisleriyle oluşturabilir.
- Veri yönetimi, güvenliği ve yönetişimi için uygun servisleri seçebilir.

Outline

Modül 1 – Google Cloud'da Veri Mühendisliğine Giriş

- Veri mühendisinin görev ve sorumlulukları
- Veri kaynakları ve veri senkronizasyonu
- Veri formatları
- Google Cloud depolama seçenekleri
- Metadata yönetimi
- Analytics Hub ile veri paylaşımı

Modül 2 – Veri Taşıma ve Replikasyon

- Veri replikasyon mimarileri
- gcloud komut satırı aracı
- Veri taşıma yöntemleri
- Datastream kullanımı

Modül 3 – Extract & Load Pipeline Tasarımı

- EL mimarisi
- bq komut satırı aracı
- BigQuery Data Transfer Service

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- BigLake

Modül 4 – ELT Yaklaşımı

- Extract, Load, Transform mimarisi
- BigQuery SQL Script
- İş planlama
- Dataform ile veri dönüşümü

Modül 5 – ETL Pipeline Geliştirme

- ETL mimarisi
- Google Cloud görsel ETL araçları
- Dataproc ile batch veri işleme
- Streaming veri işleme seçenekleri
- Bigtable entegrasyonu

Modül 6 – Pipeline Otomasyonu

- Pipeline otomasyon yaklaşımları
- Cloud Scheduler
- Workflows
- Cloud Composer
- Cloud Run Functions
- Eventarc

Modül 7 – Modern Veri Mühendisliği Yaklaşımları

- Veri mühendisliği süreçleri
- Karşılaşılan zorluklar
- BigQuery temelleri
- Data Lake ve Data Warehouse kavramları
- OLTP sistemleri ile veri ambarlarının karşılaştırılması
- Veri yönetişimi
- Erişim yönetimi
- Üretim ortamına uygun veri pipeline'ları
- Google Cloud müşteri senaryoları

Modül 8 – Data Lake Oluşturma

- Veri gölü mimarisi
- Cloud Storage kullanımı
- Güvenli veri depolama
- Farklı veri tiplerinin yönetimi
- Cloud SQL ile operasyonel verilerin kullanılması

Modül 9 – Data Warehouse Tasarımı

- Modern veri ambarı yaklaşımı
- BigQuery kullanımı
- Veri yükleme süreçleri
- Şema tasarımı
- Nested ve Repeated Fields

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- Partitioning
- Clustering optimizasyonu

Modül 10 – Batch Veri İşleme

- EL, ELT ve ETL yaklaşımları
- Veri kalitesi
- BigQuery işlem yöntemleri
- Performans değerlendirmeleri
- Veri kalitesini artırma teknikleri

Modül 11 – Apache Spark ile Dataproc

- Hadoop ekosistemi
- Dataproc kullanımı
- Cloud Storage entegrasyonu
- Dataproc optimizasyon teknikleri

Modül 12 – Serverless Veri İşleme

- Dataflow mimarisi
- Veri işleme pipeline'ları
- GroupByKey
- Combine
- Side Inputs
- Windowing
- Dataflow Templates

Modül 13 – Veri Pipeline Yönetimi

- Cloud Data Fusion ile görsel pipeline geliştirme
- Pipeline bileşenleri
- Wrangler kullanımı
- Cloud Composer
- Apache Airflow
- DAG yapıları
- İş akışı zamanlama
- İzleme ve log yönetimi

Modül 14 – Pub/Sub ile Mesajlaşma

- Pub/Sub mimarisi
- Push ve Pull modelleri
- Yayınlama ve abonelik yapıları

Modül 15 – Streaming Veri İşleme

- Streaming mimarileri
- Windowing teknikleri
- Gerçek zamanlı veri akışı yönetimi

Modül 16 – BigQuery ve Bigtable Performansı

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- Streaming veri yükleme
- Bigtable performans optimizasyonu
- Gerçek zamanlı analiz

Modül 17 – Gelişmiş BigQuery Özellikleri

- Analitik Window Functions
- GIS fonksiyonları
- Performans optimizasyon teknikleri