Data Engineering with Databricks Eğitimi

Süre: **2 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

Overview

Data Engineering with Databricks eğitimi, farklı veri disiplinlerinde çalışan profesyonellere Databricks Lakehouse Platformunun üretim ortamlarında ETL pipeline'ları oluşturmayı destekleyen temel bileşenlerini kapsamlı biçimde tanıtır.

Katılımcılar eğitim boyunca SQL ve Python kullanarak farklı veri kaynaklarından gelen yeni verileri artımlı olarak işleyen pipeline'lar oluşturmayı ve zamanlamayı öğrenir. Bu pipeline'lar, lakehouse ortamındaki analitik uygulamaları ve dashboard'ları beslemek üzere tasarlanır.

Eğitim, Databricks Data Science and Engineering Workspace, Databricks SQL, Delta Live Tables, Databricks Repos, Databricks Task Orchestration ve Unity Catalog gibi temel Databricks bileşenleri üzerinde uygulamalı çalışma imkânı sunar.

Program aynı zamanda katılımcıların **Databricks Certified Data Engineer Associate** sertifika sınavına hazırlanmasına katkı sağlar.

Prerequisites

Katılımcıların aşağıdaki bilgi ve deneyimlere sahip olmaları önerilir:

- Sanal makineler, object storage ve identity management gibi temel bulut kavramlarına başlangıç seviyesinde aşinalık,
- Compute instance oluşturma, notebook içerisinde kod çalıştırma, temel notebook işlemlerini kullanma ve Git üzerinden repository içe aktarma gibi temel kod geliştirme görevlerini gerçekleştirebilme,
- `CREATE`, `SELECT`, `INSERT`, `UPDATE`, `DELETE`, `GROUP BY` ve `JOIN` gibi komutları içeren orta seviye SQL bilgisi,
- Aggregate functions, filters, sorting, indexes, tables ve views gibi SQL kavramlarında orta seviye deneyim,
- Python programlama, Jupyter Notebook arayüzü ve PySpark temelleri hakkında başlangıç seviyesinde bilgi.

Bu ön koşullardan bir veya birkaçına sahip olmayan katılımcılar için SQL sorgulama ve Python ile veri işleme konularında temel eğitimler faydalı olabilir.

Who Should Attend

Bu eğitim özellikle aşağıdaki katılımcılar için uygundur:

- Databricks ve Delta Lake bilgisini geliştirmek isteyen Data Engineer'lar,
- Veri pipeline'ları ve veri dönüşümü konularındaki uzmanlığını artırmak isteyen Data Analyst'ler,
- Büyük veri framework'leriyle çalışan Cloud Engineer ve Developer'lar,
- Databricks Associate Data Engineering sertifikasyonuna hazırlanan profesyoneller.

Data Engineering with Databricks Eğitimi

Süre: **2 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

What You Will Learn

Eğitimin sonunda katılımcılar:

- Veri pipeline geliştirme süreçlerindeki temel sorumlulukları yerine getirmek için Databricks Lakehouse Platformunu kullanabilir,
- SQL ve Python ile verileri extract, transform ve load ederek lakehouse içerisindeki table ve view'lara aktaran production-ready veri pipeline'ları yazabilir,
- Delta Live Tables dahil olmak üzere Databricks'e özgü özellik ve sözdizimlerini kullanarak veri alımını ve artımlı değişikliklerin aktarımını kolaylaştırabilir,
- Ad hoc analitik ve dashboard kullanımları için güncel sonuçlar üreten production pipeline'larını orkestre edebilir.

Outline

1. Gün

Delta Lake

- Delta Lake'in temel yapısı ve veri mühendisliği süreçlerindeki kullanımı

Databricks Üzerinde İlişkisel Varlıklar

- Databricks ortamındaki relational entities yapılarının incelenmesi

Spark SQL ile ETL

- Spark SQL kullanarak extract, transform ve load süreçlerinin gerçekleştirilmesi

Spark SQL için Gerekli Python Temelleri

- Spark SQL çalışmalarını destekleyecek temel Python bilgisinin kullanılması

Structured Streaming ve Auto Loader ile Artımlı Veri İşleme

- Yeni gelen verilerin artımlı biçimde işlenmesi
- Structured Streaming kullanımı
- Auto Loader ile veri alım süreçlerinin yönetilmesi

2. Gün

Data Lakehouse'ta Medallion Architecture

- Medallion Architecture yaklaşımının lakehouse ortamında uygulanması

Delta Live Tables

Data Engineering with Databricks Eğitimi

Süre: **2 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
+44 20 3807 0925
+90 212 282 7700

- Delta Live Tables kullanarak veri pipeline'larının oluşturulması ve yönetilmesi

Databricks Jobs ile Task Orchestration

- Databricks Jobs kullanarak veri işlerinin ve pipeline süreçlerinin orkestre edilmesi

Databricks SQL

- Databricks SQL kullanarak analitik sorgular ve veri işlemleri gerçekleştirme

Lakehouse Ortamında Yetki Yönetimi

- Veri ve kaynak erişimlerinin yönetilmesi

Databricks SQL Üzerinde Dashboard ve Sorguları Production Ortamına Alma

- Dashboard ve sorguların üretim kullanımına hazırlanması
- Analitik sonuçların düzenli ve sürdürülebilir şekilde sunulması