Apache Spark Programming with Databricks Eğitimi

Süre: **2 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Overview

Databricks üzerinde Apache Spark programlamayı öğrenmek ve büyük ölçekli veri işleme süreçlerine giriş yapmak isteyenler için tasarlanmış uygulamalı bir eğitim programıdır.

Eğitim, her biri yaklaşık **4 saat süren dört ana modülden** oluşur. Katılımcılar Apache Spark'ın dağıtık mimarisinden başlayarak DataFrame API ile veri işleme, ETL geliştirme, gerçek zamanlı veri akışlarını Structured Streaming ile analiz etme ve Databricks üzerindeki Spark iş yüklerini optimize etme konularına ilerler.

Program boyunca **Apache Spark, Python, Spark DataFrame API, Structured Streaming, Delta Lake ve Databricks** birlikte ele alınır. Teorik konular demo ve laboratuvar çalışmalarıyla desteklenerek katılımcıların öğrendiklerini gerçek veri işleme senaryolarına uygulamaları sağlanır.

Eğitimin son bölümünde Lakehouse yaklaşımı, Delta Lake, performans optimizasyonu ve Databricks üzerindeki Spark iş yüklerinin izlenmesi gibi üretim ortamları açısından önemli konular üzerinde durulur.

Prerequisites

Katılımcıların aşağıdaki konularda temel bilgi sahibi olması beklenir:

- Python programlama diline aşinalık
- Veri tipleri, listeler, dictionary yapıları ve değişkenler
- Fonksiyonlar, döngüler ve koşullu ifadeler
- Exception handling
- Class'lara erişim ve üçüncü taraf Python kütüphanelerinin kullanımı
- Temel SQL bilgisi
- SELECT, WHERE, GROUP BY, ORDER BY, LIMIT ve JOIN kullanarak sorgu yazabilme

Who Should Attend

Bu eğitim özellikle:

- Apache Spark programlama becerilerini geliştirmek isteyen Data Engineer ve Data Scientist'ler
- Databricks üzerinde Apache Spark ve Delta Lake kullanmak isteyen geliştiriciler
- Büyük ölçekli veri işleme ve gerçek zamanlı analitik projelerinde çalışan profesyoneller

için uygundur.

Apache Spark Programming with Databricks Eğitimi

Süre: **2 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

Apache Spark'a Giriş

Katılımcılar Apache Spark'ın temel çalışma prensiplerini ve dağıtık mimarisini öğrenir. Büyük ölçekli veri işleme açısından Spark'ın nasıl kullanıldığı incelenirken farklı programlama yapıları ve Spark DataFrame API ele alınır.

Python tabanlı Spark iş akışları kullanılarak verilerin okunması, yazılması ve dönüştürülmesine yönelik temel beceriler geliştirilir.

Apache Spark ile Uygulama Geliştirme

Katılımcılar ölçeklenebilir veri işleme uygulamalarının nasıl geliştirileceğini öğrenir. Spark DataFrame API kullanılarak ETL pipeline'ları oluşturulur ve dağıtık veri dönüşümleri üzerinde çalışılır.

Grouping, aggregation, join, set operations ve window functions gibi veri işleme tekniklerinin yanı sıra array, map ve struct gibi complex data type'lar ele alınır. Performans açısından verimli Spark uygulamaları geliştirmeye yönelik iyi uygulamalar da incelenir.

Apache Spark ile Stream Processing ve Analiz

Bu bölümde gerçek zamanlı verilerin işlenmesine yönelik temel kavramlar ele alınır.

Katılımcılar **Spark Structured Streaming API** ile streaming uygulamaları geliştirir; stream aggregation ve window analysis gibi yöntemlerle sürekli gelen verilerin nasıl işlenebileceğini öğrenir. Ölçeklenebilir ve hata toleranslı streaming uygulamalarının geliştirilmesine yönelik temel yaklaşımlar uygulamalı olarak incelenir.

Databricks Üzerinde Spark İş Yüklerini İzleme ve Optimize Etme

Programın bu bölümünde **Lakehouse Architecture** ve **Medallion Architecture** yaklaşımları üzerinden ölçeklenebilir veri iş akışları ele alınır.

Unity Catalog ile veri yönetişimi, erişim kontrolü ve lineage konuları incelenirken Delta Lake kullanılarak güvenilir ve ACID uyumlu veri pipeline'larının nasıl oluşturulabileceği değerlendirilir.

Ayrıca partitioning, caching ve query tuning gibi Spark optimizasyon yöntemleri; performans izleme, troubleshooting ve verimli data engineering uygulamalarına yönelik iyi pratiklerle birlikte ele alınır.

Apache Spark Programming with Databricks Eğitimi

Süre: **2 Gün**

BİLGİNÇ IT Academy
info@bilginc.com
bilginc.com

Outline

Modül 1: Apache Spark'a Giriş

- Spark Runtime Architecture
- Databricks üzerinde Apache Spark mimarisinin incelenmesi
- Spark DataFrames ve SQL'e giriş
- DataFrame'lerle veri okuma ve yazma
- Dağıtık sistem programlamanın temelleri
- DataFrame API ile temel ETL
- DataFrame API kullanarak uçuş verileri üzerinde ETL
- DataFrame'lerle transaction verilerinin analizi

Modül 2: Apache Spark ile Uygulama Geliştirme

- DataFrame API temelleri
- Demo (Opsiyonel): DataFrame API ile temel ETL
- Verileri gruplama ve aggregate etme
- Demo: Grouping ve Aggregation
- Lab: E-ticaret verilerinin gruplandırılması ve aggregate edilmesi
- Relational Operations
- Demo: Apache Spark ile ilişkisel veri işlemleri
- Complex Data ile çalışma
- Demo: Apache Spark'ta Complex Data Type'larla çalışma
- Lab: E-ticaret verilerindeki Complex Data Type'larla çalışma

Modül 3: Apache Spark ile Stream Processing ve Analiz

- Stream Processing'e giriş
- Spark Structured Streaming
- Demo: Spark Structured Streaming'e giriş
- Lab: Spark Structured Streaming'e giriş
- İleri seviye Stream Processing ve analiz
- Demo: Spark Structured Streaming ile Window Aggregation
- Lab: Spark Structured Streaming ile Window Aggregation

Modül 4: Databricks Üzerinde Apache Spark İş Yüklerini İzleme ve Optimize Etme

- Apache Spark ve Databricks
- Apache Spark'ın Delta Lake ile kullanılması
- Demo: Delta Lake'e giriş
- Lab: Delta Lake'e giriş
- Apache Spark optimizasyonu
- Demo: Apache Spark optimizasyonu
- Lab: Apache Spark optimizasyonu