

# Apache Spark 2 Programlamaya Giriş Eğitimi

## Eğitim Hakkında

Apache Spark 2'ye Giriş eğitimi, katılımcılar Spark mimarisine ve Spark'ın nasıl çalıştığına dair sağlam bir teknik giriş sağlar. Katılımcılar, RDD'ler ve dağıtılmış hesaplama motoru dahil olmak üzere Spark'ın temel yapı taşlarını ve ayrıca Spark SQL ve DataFrames dahil olmak üzere daha basit ve daha yetenekli bir arabirim sağlayan üst düzey yapıları öğrenirler.

## Neler Öğreneceksiniz

Veri işlemede Spark ihtiyacını anlamayı,

Spark mimarisini ve hesaplamaları küme düğümlerine nasıl dağıttığını anlamayı,

Spark'ın temel kurulum / kurulum / düzenine aşina olmayı,

Spark'ı etkileşimli ve anlık işlemler için kullanmayı,

Yapılandırılmış verileri verimli bir şekilde işlemek için Dataset / DataFrame / Spark SQL kullanmayı.

## Eğitim İçeriği

### Scala Ramp Up (İsteğe Bağlı)

Scala Tanıtımı, Değişkenler, Veri Tipleri, Kontrol Akışı  
Scala Tercümanı  
Koleksiyonlar ve Standart Yöntemleri (ör. Harita ())  
İşlevler, Yöntemler, İşlev Değişmezleri  
Sınıf, Nesne, Özellik

### Spark'a Giriş

Genel Bakış, Motivasyonlar, Kıvılcım Sistemleri  
Kıvılcım Ekosistemi  
Spark ve Hadoop  
Tipik Kıvılcım Dağıtımı ve Kullanım Ortamları

### RDD'ler ve Spark Mimarisi

RDD Kavramları, Bölümler, Yaşam Döngüsü, Tembel Değerlendirme  
RDD'lerle Çalışma - Oluşturma ve Dönüştürme (harita, filtre vb.)  
Önbelleğe Alma - Kavramlar, Depolama Türü, Yönergeler

## **DataSets / DataFrames ve Spark SQL**

Giriş ve Kullanım  
Veri Kümesi Oluşturma ve Kullanma  
JSON ile çalışmak  
DataSet DSL'i kullanma  
Spark ile SQL kullanma  
Veri Biçimleri  
Optimizasyonlar: Katalizör ve Tungsten  
Veri Kümeleri, Veri Çerçevesi ve RDD'ler

## **Spark Uygulamaları Oluşturma**

Genel Bakış, Temel Sürücü Kodu, SparkConf  
SparkContext / SparkSession Oluşturma ve Kullanma  
Uygulamalar Oluşturma ve Çalıştırma  
Uygulama Yaşam Döngüsü  
Küme Yöneticileri  
Günlük Kaydı ve Hata Ayıklama

## **Kıvılcım Akışı**

Genel Bakış ve Akışla İlgili Temel Bilgiler  
Yapılandırılmış Akış  
DStreams (Discretized Steams),  
Mimari, Durum Bilgisi Olmayan, Durum Bilgisi Olan ve Pencereli Dönüşümler  
Spark Streaming API  
Programlama ve Dönüşümler

## **Performans Özellikleri ve Ayarlama**

Spark UI  
Dar ve Geniş Bağımlılıklar  
Veri İşleme ve Karıştırmayı En Aza İndirmek  
Önbelleğe Alma - Kavramlar, Depolama Türü, Yönergeler  
Önbelleğe Almayı Kullanma  
Yayın Değişkenlerini ve Akümülatörleri Kullanma

## **(İsteğe bağlı): Spark GraphX'e Genel Bakış**

Giriş  
Basit Grafikler Oluşturmak  
GraphX API  
En Kısa Yol Örneği

## **(İsteğe bağlı): MLLib'e Genel Bakış**

Giriş  
Özellik Vektörleri  
Kümeleme / Gruplama, K-Ortalamalar  
Öneriler  
Sınıflandırmalar