

Büyük Veri Altyapısı Eğitimi

Eğitim Hakkında

Büyük Veri Altyapısı eğitim kursu, geliştiricilere, veri bilimcilere ve DevOps uzmanlarına Büyük Veri hizmetlerinde nasıl gezineceklerini ve bunları nasıl anlayacaklarını öğretir. Katılımcılar Kapsayıcılar (Docker), Hadoop Dağıtılmış Dosya Sistemi (HDFS), Apache Spark, Doğal Dil İşleme (NLP) uygulamaları, Cassandra, Kubernetes ve daha fazlasını nasıl kullanacaklarını öğrenirler.

Neler Öğreneceksiniz

“Büyük Veri” yi ve ondan içgörü elde etmeye çalışırken ortaya çıkabilecek zorlukları açıklayın: Hacim, hız, çeşitlilik, değişkenlik ve karmaşıklık.

Verilerle geniş ölçekte çalışmak için yaygın olarak kullanılan açık kaynaklı sistemleri tanıyın: Cassandra, Elasticsearch, Hadoop / HDFS, Spark ve Kafka.

Büyük veri ekosisteminin her bir üyesinin genel bakışını ve ele alınması amaçlanan sorunların örneklerini sağlayın.

Kafka'yı derinlemesine inceleyin ve veri akışlarını yönetmek için ardışık düzenler oluşturmak için nasıl kullanılabileceğini görün.

İlişkisel veritabanlarında, Cassandra, Elasticsearch ve Hadoop / HDFS'de depolanan farklı veri türlerini analiz etmek için nasıl kullanılabileceğini göstermek için Spark'a derinlemesine dalın.

Eğitim İçeriği

Başlangıç

Kursta kullanılacak araçları tanıyın
Python ve Anaconda
Jupyter

Veri ve Büyük Veri

Verilerin Zorlukları: 5VC
Büyük Veri nedir?
Kurtarılacak Teknoloji: Büyük veri kümeleriyle çalışmayı kolaylaştırma
Sanallaştırma
Everything as a Service (EaaS): Karmaşık hesaplama araçlarını yönetmek için stratejiler
Dağıtılmış Depolama ve Hesaplama
Bir Büyük Veri Çözümünün Özellikleri

Mesos ve DC / OS: İşletim Platformu Modern Veri Merkezi

Mimari

Yapılandırma ve yönetim

Uygulama ve hizmet dağıtımı

Ağ İletişimi, yük dengeleme ve uygulama izolasyonu

Kafka

Gerekçe ve rol: Kafka hangi sorunu çözer?

Mimari ve temel bileşenler

DC / OS içinde servis kurulumu

API: Tüketiciler ve Üreticiler

Python ve Java istemci kitaplıkları

Kafka Connect: Yapılandırılmış verileri taşımak ve bunlarla çalışmak için araçlar

Hadoop Dağıtılmış Dosya Depolama (HDFS)

HDFS nedir ve Hadoop dünyasına nasıl uymaktadır?

HDFS API ve daha sonra analiz için veri almaya yönelik araçlar

Python ve Java istemci kitaplıkları

Apache Spark: Büyük Ölçekli Veri İşleme için Genel Motor

Spark nedir?

Pratikte nasıl kullanılır?

Mimari ve bileşenler

Ekosistem: Çekirdek, SQL, Makine Öğrenimi (MLlib), Grafik

DC / OS içinde servis kurulumu

API ve ortam

Yapılandırılmış verilerin analizi

Spark, akış veri kümelerini analiz etmek için nasıl kullanılabilir?

ElasticSearch: Aramaya Karşı Depolama

Büyük Veri içinde aramanın rolü nedir? Bir arama motoru ne gibi fayda sağlar?

Mimari ve veri depolama

Elasticsearch'ün bir analiz platformu olarak kullanılmasını sağlayan nedir: toplama ve belirtme

DC / OS içinde servis kurulumu

Anahtar API'ler

Doğal Dil İşleme platformu olarak Rolü

ES verilerini keşfetmek için bir görselleştirme platformu olarak Kibana'nın kullanılması

Cassandra: Ölçekte Yapılandırılmamış Veri Depolama

Cassandra'nın Amacı: Hangi sorunları çözüyor?

Mimari ve bileşenler

API ve entegrasyon

DC / OS'de servis kurulumu

Python ve Java istemcisi

Kubernetes

DevOps nedir?

DevOps uygulamaları Büyük Veri ile nasıl ilişkilidir?

Kubernetes DC / OS içinde hangi rolü oynar ve özel yazılımın yönetimini nasıl kolaylaştırabilir?

Mimari ve bileşenler

DC / OS'de servis kurulumu

Veri hizmetlerini Kubernetes'e dağıtmak için iş akışlarını gösterin
Kafka bağlantı uygulamalarını bir pod olarak paketleyin ve dağıtın
Apache Spark Streaming uygulamalarını bir pod olarak paketleyin ve dağıtın