

Modern Lakehouse & Data Platform Architecture Eğitimi

Süre: 3 Gün

Açıklama

Modern Lakehouse & Data Platform Architecture Eğitimi, geleneksel Hadoop ekosisteminden modern veri platformlarına geçiş süreçlerini, yeni nesil veri mimarilerini ve açık tablo formatlarını ele alan ileri düzey bir teknik eğitim programıdır. Eğitim kapsamında Data Lake, Data Warehouse ve Lakehouse mimarileri karşılaştırılarak Apache Iceberg ve Delta Lake teknolojilerinin mimari bileşenleri, kullanım alanları ve kurumsal veri platformlarına sağladığı avantajlar incelenmektedir. ACID Transactions, Schema Evolution, Time Travel ve veri versiyonlama gibi özelliklerin büyük ölçekli veri yönetiminde nasıl kullanıldığı uygulamalı örneklerle ele alınmaktadır. Programda ayrıca mevcut Hadoop, HDFS ve Hive tabanlı altyapıların modern Lakehouse mimarilerine dönüştürülmesi, veri ve metadata geçiş stratejileri, performans optimizasyonu ve kurumsal mimari tasarımı konularına odaklanılmaktadır. Katılımcıların modern Lakehouse mimarilerini ve açık tablo formatlarını teknik açıdan değerlendirebilmelerini, Apache Iceberg ve Delta Lake teknolojilerini kullanabilmelerini ve mevcut Hadoop tabanlı veri platformlarından modern, ölçeklenebilir ve sürdürülebilir veri mimarilerine geçiş için uygun modernizasyon stratejileri geliştirebilmelerini sağlamaktır. **Bu eğitimde neler öğreneceksiniz?**

- Geleneksel Hadoop ekosistemi ile modern veri platformları arasındaki mimari farklılıkları açıklayabilecek
- Data Lake, Data Warehouse ve Lakehouse yaklaşımlarını karşılaştırabilecek
- Açık tablo formatlarının temel özelliklerini ve kullanım senaryolarını değerlendirebilecek
- Apache Iceberg ve Delta Lake mimarilerini anlayabilecek
- ACID Transactions, Schema Evolution ve Time Travel özelliklerini uygulayabilecek
- Veri tutarlılığı, versiyonlama ve performans optimizasyonu süreçlerini yönetebilecek
- Hadoop tabanlı veri platformlarından Lakehouse mimarilerine geçiş stratejileri geliştirebilecek
- Kurumsal gereksinimlere uygun modern veri platformu mimarileri tasarlayabileceklerdir

Kimler Katılmalı?

- Data Engineer / Senior Data Engineer
- Data Architect / Enterprise Data Architect

- Data Platform Engineer
- Big Data Engineer
- Solution Architect
- Veri platformlarının geliştirilmesi, yönetimi ve modernizasyonundan sorumlu teknik ekipler

Eğitim İçeriği

1. Modern Data Platform Architecture

- Geleneksel veri platformlarından modern veri mimarilerine geçiş
- Hadoop ekosisteminin temel bileşenleri
- HDFS, Hive ve Apache Spark mimarileri
- Hadoop tabanlı veri platformlarının mimari ve operasyonel sınırlamaları
- Modern Data Platform kavramı
- Storage ve Compute katmanlarının ayrıştırılması
- Ölçeklenebilirlik, esneklik ve sürdürülebilirlik
- On-Premise, Cloud ve Hybrid veri platformları

2. Data Lake, Data Warehouse & Lakehouse Architecture

- Data Warehouse mimarisi ve temel özellikleri
- Data Lake mimarisi ve kullanım alanları
- Data Lake ve Data Warehouse karşılaştırması
- Geleneksel Data Lake mimarilerinin sınırlamaları
- Lakehouse mimarisinin ortaya çıkışı
- Lakehouse mimarisinin temel bileşenleri
- Batch ve Streaming veri işleme yaklaşımları
- Lakehouse mimarisinde veri yönetimi
- Kurumsal veri platformları için mimari seçim kriterleri

3. Open Table Formats

- Open Table Format kavramı
- Parquet ve benzeri dosya formatlarının rolü
- Geleneksel dosya tabanlı veri depolamanın sınırlamaları
- Metadata ve Transaction Log yapıları
- Apache Iceberg, Delta Lake ve Apache Hudi'ye genel bakış
- Açık tablo formatlarının karşılaştırılması
- Storage Engine ve Query Engine ilişkisi
- Açık tablo formatlarında birlikte çalışabilirlik
- Vendor bağımsızlığı ve teknoloji seçim kriterleri

4. Apache Iceberg Architecture

- Apache Iceberg mimarisine giriş
- Iceberg Table yapısı
- Metadata Files ve Manifest Files
- Snapshot mimarisi
- Hidden Partitioning
- Partition Evolution
- Schema Evolution
- Snapshot Management
- Time Travel ve veri versiyonlama
- Iceberg ile veri okuma ve yazma işlemleri
- Apache Spark ve Spark SQL ile Iceberg kullanımı
- Uygulama: Iceberg tablolarının oluşturulması ve yönetilmesi

5. Delta Lake Architecture

- Delta Lake mimarisine giriş
- Delta Table yapısı
- Delta Transaction Log
- ACID Transactions
- Schema Enforcement
- Schema Evolution
- Time Travel ve Versioning
- MERGE, UPDATE ve DELETE işlemleri
- OPTIMIZE ve Compaction yaklaşımları
- Delta Lake performans özellikleri
- Apache Spark ile Delta Lake kullanımı
- Uygulama: Delta tabloları üzerinde veri yönetimi işlemleri

6. ACID Transactions & Data Consistency

- ACID prensipleri
- Dağıtık veri platformlarında transaction yönetimi
- Atomicity, Consistency, Isolation ve Durability
- Snapshot Isolation
- Optimistic Concurrency Control
- Eşzamanlı veri okuma ve yazma işlemleri
- Veri tutarlılığı ve güvenilirlik
- Apache Iceberg ve Delta Lake transaction mekanizmalarının karşılaştırılması
- Uygulama: Transaction ve veri tutarlılığı senaryoları

7. Schema Evolution & Time Travel

- Schema Evolution kavramı
- Kolon ekleme, silme ve değiştirme işlemleri
- Schema Compatibility
- Veri versiyonlama yaklaşımları
- Snapshot tabanlı veri yönetimi
- Time Travel sorguları
- Historical Data Analysis
- Rollback ve veri kurtarma senaryoları
- Veri saklama ve versiyon temizleme politikaları
- Uygulama: Schema Evolution ve Time Travel işlemleri

8. Hadoop to Modern Lakehouse Migration

- Mevcut Hadoop ekosisteminin değerlendirilmesi
- HDFS ve Hive tabanlı veri mimarilerinin analizi
- Hadoop modernizasyon stratejileri
- Mevcut veri kaynaklarının ve bağımlılıkların belirlenmesi
- Hive tablolarından Apache Iceberg veya Delta Lake'e geçiş
- Veri taşıma ve dönüştürme yaklaşımları
- Metadata ve katalog geçişi
- Mevcut ETL/ELT süreçlerinin modernizasyonu
- Veri doğrulama ve tutarlılık kontrolleri
- Kademeli geçiş ve paralel çalışma stratejileri
- Geçiş riskleri ve geri dönüş planlaması
- Hadoop'dan Lakehouse'a geçiş yol haritasının oluşturulması

9. Lakehouse Performance & Optimization

- Partitioning stratejileri
- File Size Optimization
- Small Files problemi
- Compaction
- Query Performance Optimization
- Metadata Optimization
- Storage ve Compute kaynaklarının yönetimi
- Performans ve maliyet optimizasyonu
- Apache Iceberg ve Delta Lake performans yaklaşımlarının karşılaştırılması

10. Enterprise Lakehouse Architecture Workshop

- Örnek Hadoop mimarisinin değerlendirilmesi

- Mevcut mimarideki teknik sınırlamaların belirlenmesi
- Hedef Lakehouse mimarisinin tasarlanması
- Apache Iceberg ve Delta Lake seçim kriterleri
- Veri depolama ve işleme katmanlarının belirlenmesi
- Metadata ve katalog mimarisinin oluşturulması
- ACID Transactions, Schema Evolution ve Time Travel senaryolarının değerlendirilmesi
- Veri geçişi ve modernizasyon stratejisinin belirlenmesi
- Hedef mimari için teknik yol haritasının hazırlanması

Ön Koşullar

- SQL ve veri işleme süreçleri hakkında bilgi sahibi olmak
- Hadoop ekosistemi veya dağıtık veri işleme teknolojilerine aşina olmak
- ETL/ELT süreçleri ve veri depolama mimarileri hakkında temel deneyime sahip olmak
- Apache Spark veya benzeri büyük veri işleme teknolojileri hakkında bilgi sahibi olunması önerilir