Implement Data Engineering Solutions using Azure Databricks Eğitimi

Süre: **4 Gün**

BİLGİNÇ
IT Academy

info@bilginc.com
bilginc.com

Implement Data Engineering Solutions using Azure Databricks eğitimi, Azure Databricks ve lakehouse mimarisini kullanarak modern veri mühendisliği çözümleri geliştirmeye yönelik kapsamlı ve uygulama odaklı bir programdır. Eğitim, kurumsal analitik ve makine öğrenmesi iş yüklerini destekleyebilecek ölçeklenebilir veri pipeline'larının tasarlanması, geliştirilmesi, yönetilmesi ve optimize edilmesine odaklanır.

Program boyunca katılımcılar Azure Databricks çalışma ortamının yapılandırılmasından başlayarak veri yönetişimi, veri alımı, dönüşüm, pipeline orchestration ve üretim ortamı yönetimine kadar uzanan uçtan uca veri mühendisliği süreçlerini ele alır.

Eğitimin önemli başlıklarından biri **Unity Catalog** kullanılarak merkezi ve güvenli veri yönetişiminin oluşturulmasıdır. Katılımcılar metastore, catalog, schema ve table yapılarını inceleyerek farklı çalışma alanlarında veri erişiminin nasıl kontrol edilebileceğini, yetkilendirme mekanizmalarının nasıl uygulanacağını ve data lineage ile audit yeteneklerinin yönetişim süreçlerine nasıl dahil edilebileceğini öğrenir.

Veri işleme tarafında SQL ve Python kullanılarak batch ve streaming verilerin alınması, hazırlanması ve dönüştürülmesi ele alınır. **Delta Lake** ile güvenilir ve ölçeklenebilir veri depolama yaklaşımı incelenirken veri kalitesi, temizleme, doğrulama ve performans optimizasyonu gibi üretim ortamları açısından önemli konular da uygulamalı olarak değerlendirilir.

Programın ilerleyen bölümlerinde Databricks jobs ve workflows kullanılarak güvenilir veri pipeline'larının oluşturulması, CI/CD süreçlerinin uygulanması, üretim iş yüklerinin izlenmesi ve performans ile maliyet açısından optimize edilmesi üzerinde durulur.

Eğitmen liderliğindeki oturumlar ve Microsoft Learn ile uyumlu uygulamalı laboratuvarlar sayesinde katılımcıların öğrendikleri teknikleri gerçek kurumsal veri mühendisliği senaryolarına aktarabilecek yetkinliğe ulaşmaları hedeflenir.

Önkoşullar

Katılımcıların aşağıdaki konularda bilgi ve deneyime sahip olmaları önerilir:

- SQL ile sorgu yazma deneyimi
- Python ve notebook kullanımı
- Veri mühendisliği ve data warehouse kavramlarının anlaşılması
- Azure servislerine aşinalık
- Temel güvenlik prensipleri hakkında bilgi
- Git kullanarak temel version control bilgisi

Implement Data Engineering Solutions using Azure Databricks Eğitimi

Süre: **4 Gün**

BİLGİNÇ
IT Academy
info@bilginc.com
bilginc.com

Kimler Katılmalı

Bu eğitim özellikle:

- Veri pipeline'larının geliştirilmesi ve yönetilmesinden sorumlu data engineer'lar,
- Data engineering rollerine geçiş yapmak isteyen analitik profesyonelleri,
- Azure veri platformlarıyla çalışan cloud profesyonelleri,
- Databricks kullanarak ölçeklenebilir lakehouse çözümleri oluşturmak isteyen kurum ve teknik ekipler

için tasarlanmıştır.

Neler Öğreneceksiniz

Eğitimin sonunda katılımcılar:

- Azure Databricks workspace'lerini ve compute kaynaklarını yapılandırabilir ve yönetebilir.
- Unity Catalog kullanarak güvenli ve ölçeklenebilir veri yönetişimi uygulayabilir.
- SQL, Python ve Delta Lake ile verileri alabilir, hazırlayabilir ve dönüştürebilir.
- Batch ve streaming veri işleme gereksinimlerine uygun yaklaşımlar geliştirebilir.
- Üretim iş yükleri için güvenilir veri pipeline'ları tasarlayabilir ve orchestrate edebilir.
- Pipeline ve veri işleme performansını izleyebilir.
- Performansı ve maliyet verimliliğini geliştirmek için optimizasyon teknikleri uygulayabilir.
- Azure üzerinde kurumsal ölçekte veri mühendisliği çözümleri devreye alabilir ve sürdürebilir.

Eğitim İçeriği

1. Azure Databricks Ortamının Kurulması ve Yapılandırılması

Bu bölümde Azure Databricks'in temel mimarisi ve veri mühendisliği çalışmalarının gerçekleştirileceği ortamın hazırlanması ele alınır.

- Azure Databricks mimarisinin anlaşılması
- Temel Databricks bileşenleri
- Ortak veri mühendisliği çalışmaları için workspace yapılandırması
- Farklı iş yüklerine uygun cluster ve compute kaynaklarının oluşturulması
- Compute kaynaklarının yönetilmesi
- Git ile geliştirme ortamlarının entegre edilmesi
- Version control yaklaşımının uygulanması
- Microsoft Entra ID ile identity ve access management kontrolleri
- Workspace'ler genelinde temel governance uygulamalarının oluşturulması

2. Unity Catalog ile Veri Güvenliği ve Yönetişimi

Bu bölüm, kurumsal veri varlıklarının merkezi ve kontrollü biçimde yönetilmesine odaklanır.

- Unity Catalog mimarisine genel bakış
- Metastore yapısının anlaşılması

BİLGİNÇ IT Academy

info@bilginc.com

bilginc.com

- Catalog, schema ve table organizasyonu
- Workspace'ler arasında merkezi veri erişim kontrolü
- Role-Based Access Control (RBAC) yaklaşımı
- Güvenli veri operasyonları için erişim politikaları
- Data lineage takibi
- Audit yetenekleri
- Uyumluluk ve kurumsal standartları destekleyen governance uygulamaları

3. Azure Databricks ile Verilerin Hazırlanması ve İşlenmesi

Bu bölümde farklı kaynaklardan gelen verilerin analitik ve diğer kurumsal iş yüklarına hazırlanması incelenir.

- Batch veri için ingestion stratejileri
- Streaming veri ingestion yaklaşımları
- SQL kullanarak veri hazırlama ve dönüşüm
- Python ile veri işleme
- Delta Lake ile güvenilir ve ölçeklenebilir veri depolama
- Structured veri setlerinin hazırlanması
- Veri kalitesi kontrollerinin uygulanması
- Data cleansing işlemleri
- Veri doğrulama süreçleri
- Transformation işlemlerinin performans ve ölçeklenebilirlik açısından optimize edilmesi

4. Veri Pipeline'larının ve İş Yüklerinin Devreye Alınması

Bu bölümde geliştirilen veri süreçlerinin üretim ortamına taşınması ve otomatik olarak çalıştırılması ele alınır.

- Databricks jobs ile veri işlerinin oluşturulması
- Workflows ile pipeline orchestration
- Veri pipeline'larında bağımlılıkların yönetilmesi
- Databricks iş yükleri için Continuous Integration ve Continuous Deployment yaklaşımı
- Geliştirmeden production ortamına geçiş süreçleri
- Tekrarlanabilir deployment yapılarının oluşturulması

5. Production İş Yüklarının İzlenmesi ve Optimizasyonu

Üretim ortamındaki veri çözümlerinin sürdürülebilir biçimde çalışması için monitoring ve optimizasyon süreçleri incelenir.

- Pipeline performansının izlenmesi
- Üretim ortamındaki problemlerin belirlenmesi
- Troubleshooting yaklaşımı
- Veri işleme performansının geliştirilmesi
- Kaynak kullanımının optimize edilmesi
- Maliyetlerin azaltılmasına yönelik optimizasyon teknikleri
- Kurumsal ölçekte production workload yönetimi
- Veri mühendisliği çözümlerinin bakımının sürdürülmesi