

İçerik

Ders Kodu	Dersin Adı	Yarıyıl	Teori	Uygulama	Lab	Kredisi	AKTS
INF 517	Veri Mühendisliği	2	3	0	0	3	6

Ön Koşul	
Derse Kabul Koşulları	

Dersin Dili	İngilizce
Türü	Seçmeli
Dersin Düzeyi	Yüksek Lisans

Dersin Amacı	<p>Veri mühendisliği, veri toplama, depolama, yönetim, güvenlik ve işleme için sistemlerin tasarımı ve analiz yöntemlerinin kullanımıyla ilgilenen bir disiplindir. İşlenebilir durumdaki büyük miktardaki "Büyük Veri"nin yönetimi için zengin veri yönetimi şemalarına ihtiyaç duyulmaktadır. Bu ders, Veri Mühendisliği prensipleri ve uygulamalarıyla ilgili temel bir kurs olacak ve aşağıdaki başlıklardan oluşacaktır:</p> <ol style="list-style-type: none">I. Veri yaşam döngüsüII. Veriyi düzenlemek ve yönetmek için veri modelleme teknikleriIII. Çoklu kaynak sistemlerinden veri toplamak, dönüştürmek, analiz etmek ve görselleştirmek için veri boruları oluşturmaIV. Farklı sorgu dilleriyle veriyi işlemeV. Veri analitiği uygulamaları ve algoritmalarıVI. Geleneksel olmayan veri türlerini yönetmeVII. Veri standartları ve veri kalitesi
--------------	--

İçerik	<ol style="list-style-type: none">1. Veri Mühendisliğine Giriş: Genel Kavramlar2. Veri Depolama Teknolojileri3. Bulut Veri Platformları (AWS/Azure/GCP)4. Veri Entegrasyon Yöntemleri ve Veri Hattı Mimarileri5. Apache Airflow ile İş Akışı Orkestrasyonu6. dbt (veri oluşturma aracı) ile Veri Dönüştürme7. Spark ile Toplu İşleme8. Akış İşleme Temelleri ve Apache Kafka9. Arama ve Bilgi Erişimi: Elastic Search10. Veri Gölü Evi: Mimari ve Prensipler11. Veri Ağ: Mimari ve Prensipler12. Veri Yönetimi - 1: Meta Veri Yönetimi13. Veri Yönetimi - 2: Veri Kalitesi ve Testi14. Veri Yönetimi - 2: Veri Soy Ağacı ve Gözlemlenebilirlik
--------	---

Kaynaklar	<p>Reis, J, Housley M, Fundamentals of Data Engineering: Plan and Build Robust Data Systems, 1st Edition, 2022, O'Reilly, 978-1098108304</p> <p>Warren, J., & Marz, N. (2015). Big Data: Principles and best practices of scalable realtime data systems. Simon and Schuster.</p> <p>Learning Spark: Lightning-Fast Big Data Analysis, by by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia. O'Reilly Media. Feb 2015</p> <p>Hadoop: The Definitive Guide, by Tom White. O'Reilly Media. April 2015. (Fourth edition of the book at Amazon.com)</p> <p>Gorelik, A. (2019). The enterprise big data lake: Delivering the promise of big data and data science. O'Reilly Media.</p>
-----------	---

Teori Konu Başlıkları

Hafta	Konu Başlıkları
1	Veri Mühendisliğine Giriş: Genel Kavramlar
2	Veri Depolama Yöntemleri - 1: Veritabanları, Veri Ambarları ve Veri Gölü
3	Veri Depolama Yöntemleri - 2 Hadoop Mimarisi ve Ekosistem + NoSQL veri tabanı
4	Veri Mühendisliği ardışık düzenleri: ETL - ELT - Data Ingestion
5	İş Akışı Orkestrasyonu
6	Büyük Veri ekosisteminde veri aktarımında kullanılan veri modelleri
7	Veri mühendisliğinde yeni mimari paradigmlar 1 - Data Lakehouse
8	Veri mühendisliğinde yeni mimari paradigmlar 2 - Data Mesh
9	İş Akışı Oİş Akışı Orkestrasyonu
10	Akış İşleme Temelleri ve Apache Kafka
11	Arama ve Bilgi Erişimi: Elastic Search
12	Veri Yönetimi - 1: Meta Veri Yönetimi
13	Veri Yönetimi - 2: Veri Kalitesi ve Testi
14	Veri Yönetimi - 2: Veri Soy Ağacı ve Gözlemlenebilirlik