

İçerik

Ders Kodu	Dersin Adı	Yarıyıl	Teori	Uygulama	Lab	Kredisi	AKTS
VM 514	Veri Bilimi Uygulamaları I	2	0	2	0	2	4

Ön Koşul	
Derse Kabul Koşulları	

Dersin Dili	Türkçe
Türü	Zorunlu
Dersin Düzeyi	Yüksek Lisans
Dersin Amacı	Veri Mühendisleri, veri akış kanallarının (data pipelines) ve veri işleme sistemlerinin (data processing systems) tasarımını, uygulamasını ve bakımını destekler. Bu sistemler, bilgilerin ölçeklenebilir, tekrarlanabilir ve güvenli bir şekilde toplanmasını, depolanmasını, toplu ve gerçek zamanlı işlenmesini ve analizini destekler. Temel olarak veri toplama, işleme ve depolamaya yönelik en uygun çözümleri tanımlamak/tasarlamakla yükümlüdürler. Sürecin sonunda, Veri Analistleri ve Veri Bilimcilerle çalışmalarında kullanabilmeleri için güvenilir ve temiz veri sağlanacaktır. Bu ders, katılımcılara veri ardışık düzenleri, farklı veri türleri ve bu verileri işlemek için kullanılan çeşitli veri platformları ile çalışma konusundaki temel kavramları tanıtmayı amaçlamaktadır. Katılımcılar, verileri farklı türdeki veritabanı sistemlerine aktarmanın, temizlemenin, işlemenin ve saklamanın yollarını öğreneceklerdir. Ayrıca öğrenciler, geleneksel sistemlerin idare edemediği 'büyük verileri' yönetmek ve manipüle etmek için kullanılan modern teknolojilerle de tanışacaklardır. Farklı veri ekosistemlerinin ihtiyaçları doğrultusunda verinin saklanması için farklı çözümleri oluşturma ve yönetme tekniklerini öğreneceklerdir.
İçerik	1. Genel Kavramlar, Büyük Veri ve Veri Mühendisliğine Giriş 2. Veri Saklama Yöntemleri – 1 Data Warehouses vs Data Lake 3. Veri Saklama Yöntemleri - 2 Hadoop Mimarisi ve Ekosistemi + NoSQL veritabanları 4. Veri aktarımı (ETL, ELT, Data Ingestion) 5. Basit veri toplama yöntemleri - Web Scraping 6. Büyük Veri ekosisteminde veri aktarımında kullanılan veri modelleri 7. Veri aktarımı araçları -- Apache Flume, Kafka -- Toplu veri işleme 8. Veri aktarımı araçları -- Kafka & Spark -- Akan veri işleme 9. Veri Saklama Yöntemleri - 3 Lakehouse Architecture 10. Bulut üzerinde veri aktarımı mimarileri : Lambda & Kappa Architecture 11. Bulut üzerinde Büyük Veri Analizi , Google Big Query
Kaynaklar	Warren, J., & Marz, N. (2015). Big Data: Principles and best practices of scalable realtime data systems. Simon and Schuster. Learning Spark: Lightning-Fast Big Data Analysis, by by Holden Karau, Andy Konwinski, Patrick Wendell, and Matei Zaharia. O'Reilly Media. Feb 2015 Hadoop: The Definitive Guide, by Tom White. O'Reilly Media. April 2015. (Fourth edition of the book at Amazon.com) Gorelik, A. (2019). The enterprise big data lake: Delivering the promise of big data and data science. O'Reilly Media. Reis, J, Housley M, Fundamentals of Data Engineering: Plan and Build Robust Data Systems, 1st Edition, 2022, O'Reilly, 978-1098108304

Teori Konu Başlıkları

Hafta	Konu Başlıkları
1	Genel Kavramlar, Büyük Veri ve Veri Mühendisliğine Giriş
2	Veri Saklama Yöntemleri – 1 Data Warehouses vs Data Lake
3	Veri Saklama Yöntemleri - 2 Hadoop Mimarisi ve Ekosistemi + NoSQL veritabanları
4	Veri aktarımı (ETL, ELT,Data Ingestion)
5	Basit veri toplama yöntemleri - Web Scraping
6	Büyük Veri ekosisteminde veri aktarımında kullanılan veri modelleri
7	Veri aktarımı araçları -- Apache Flume, Kafka -- Toplu veri işleme
8	Veri aktarımı araçları -- Kafka & Spark -- Akan veri işleme
9	Veri Saklama Yöntemleri - 3 Lakehouse Architecture
10	Veri Yönetiminde yeni paradigmalar : Data Mesh
11	Bulut üzerinde Büyük Veri Analizi , Google Big Query