

## GENEL TANIM / GENERAL DESCRIPTION

Ders Adı / Course Name	HIGH PERFORMANCE DATA PROCESSING / HIGH PERFORMANCE DATA PROCESSING	
Ders Kodu / Course Code	501004992025	
Ders Türü / Course Type		
Ders Seviyesi / Course Level	First Cycle / First Cycle	
Ders Akts Kredi / ECTS	5.00	
Haftalık Ders Saati (Kuramsal) / Course Hours For Week (Theoretical)	3.00	
Haftalık Uygulama Saati / Course Hours For Week (Objected)	0.00	
Haftalık Laboratuar Saati / Course Hours For Week (Laboratory)	0.00	
Dersin Verildiği Yıl / Year	4	
Öğretim Sistemi / Teaching System	Face to Face / Face to Face	
Eğitim Dili / Education Language	English / English	
Ön Koşulu Olan Ders(ler) / Precondition Courses	Yok	None
Amacı / Purpose	Ders kapsamında toplu işleme, akış işleme ve bunların birleşik mimarilerine odaklanılarak yüksek performanslı veri işleme mimarilerinin kapsamlı bir şekilde incelenmesi hedeflenmektedir. Hadoop, Spark, Kafka ve Flink gibi endüstri standardı araçları kullanılarak teorik temellerin pratik uygulamalarla bütünleştirilmesi amaçlanmaktadır. Ders sonunda öğrencilerin dağıtık ortamlarda hata toleransı, tutarlılık ve ölçeklenebilirlik gibi ölçütlere dikkat ederek büyük ölçekteki verileri verimli bir şekilde işleyen yüksek performanslı veri hatları tasarlaması, geliştirmesi ve bunları eniyileştirmesi beklenmektedir.	This course provides a comprehensive exploration of high-performance data processing architectures, with a focus on batch processing, stream processing, and their combined architectures. It integrates theoretical foundations with practical applications using industry-standard tools such as Hadoop, Spark, Kafka, and Flink. By the end of this course, students will design, build, and optimize high-performance data pipelines that process massive-scale data efficiently, ensuring fault tolerance, consistency, and scalability in distributed environments.

İçeriği / Content	<p>1. Veri Yoğun Uygulamaların Temelleri</p> <ul style="list-style-type: none"> <li>• Dağıtık veri sistemlerinde güvenilirlik, ölçeklenebilirlik ve sürdürülebilirlik.</li> <li>• Veri modellerinin kıyaslanması (ilişkisel, NoSQL, belge tabanlı, grafik tabanlı).</li> </ul> <p>2. Toplu İşleme Mimarileri</p> <ul style="list-style-type: none"> <li>• Hadoop ve MapReduce'a Giriş.</li> <li>• Spark ile toplu iş akışlarının optimizasyonu (RDD'ler, DataFrames, MLlib)</li> <li>• Apache Airflow ve Oozie ile veri hattı orkestrasyonu.</li> </ul> <p>3. Akış İşleme Mimarileri</p> <ul style="list-style-type: none"> <li>• Apache Kafka ile gerçek zamanlı olay odaklı işleme.</li> <li>• Apache Flink kullanarak durum bilgisi (stateful) akışı işleme (olay zamanı, pencereleme, hata toleransı).</li> <li>• Geciken (Late-arriving) verileri işleme, tutarlılık ve hata kurtarma.</li> </ul> <p>4. Birleşik Toplu ve Akış Mimarileri</p> <ul style="list-style-type: none"> <li>• Hibrit veri işleme için Lambda ve Kappa mimarileri.</li> <li>• Kafka, Spark ve Flink'i entegre eden pratik uygulamalar.</li> </ul> <p>5. Performans Optimizasyonu ve Gelişmiş Analizler</p> <ul style="list-style-type: none"> <li>• Sorgu optimizasyonu, indeksleme, bölümlendirme ve depolama verimliliği.</li> <li>• Spark MLlib ve GraphX ile ölçeklenebilir makine öğrenimi ve grafik işleme.</li> <li>• Netflix, LinkedIn ve Uber'den gerçek dünya vaka çalışmaları.</li> </ul>	<p>1. Foundations of Data-Intensive Applications</p> <ul style="list-style-type: none"> <li>• Reliability, scalability, and maintainability in distributed data systems.</li> <li>• Comparison of data models (relational, NoSQL, document-based, graph-based).</li> </ul> <p>2. Batch Processing Architectures</p> <ul style="list-style-type: none"> <li>• Introduction to Hadoop and MapReduce.</li> <li>• Optimizing batch workflows with Spark (RDDs, DataFrames, MLlib).</li> <li>• Data pipeline orchestration with Apache Airflow &amp; Oozie.</li> </ul> <p>3. Stream Processing Architectures</p> <ul style="list-style-type: none"> <li>• Real-time event-driven processing with Apache Kafka.</li> <li>• Stateful stream processing using Apache Flink (event-time, windowing, fault tolerance).</li> <li>• Handling late-arriving data, consistency, and failure recovery.</li> </ul> <p>4. Combined Batch and Streaming Architectures</p> <ul style="list-style-type: none"> <li>• Lambda &amp; Kappa architectures for hybrid data processing.</li> <li>• Practical implementations integrating Kafka, Spark, and Flink.</li> </ul> <p>5. Performance Optimization &amp; Advanced Analytics</p> <ul style="list-style-type: none"> <li>• Query optimization, indexing, partitioning, and storage efficiency.</li> <li>• Machine learning and graph processing at scale with Spark MLlib &amp; GraphX.</li> <li>• Real-world case studies from Netflix, LinkedIn, and Uber.</li> </ul>
Önerilen Diğer Hususlar / Recommended Other Considerations	Yok	None
Staj Durumu / Internship Status	Yok	None
Kitap / Malzemesi / Önerilen Kaynaklar / Books / Materials / Recommended Reading	<p>Ana Kaynak:</p> <p>1. Kleppmann, Martin. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. " O'Reilly Media, Inc.", 2017.</p> <p>Yardımcı Kaynaklar:</p> <p>2. White, Tom. "Hadoop: The Definitive Guide, Storage and Analysis at Internet Scale." White-London: O'Reilly Media(2015).</p> <p>3. Ryza, Sandy, et al. Advanced analytics with spark: patterns for learning from data at scale. " O'Reilly Media, Inc.", 2017.</p> <p>4. Narkhede, Neha, Gwen Shapira, and Todd Palino. Kafka: the definitive guide: real-time data and stream processing at scale. " O'Reilly Media, Inc.", 2017.</p> <p>5. Hueske, Fabian, and Vasiliki Kalavri. Stream processing with Apache Flink: fundamentals, implementation, and operation of streaming applications. O'Reilly Media, 2019.</p>	<p>Main Resource:</p> <p>1. Kleppmann, Martin. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. " O'Reilly Media, Inc.", 2017.</p> <p>Supplementary Resources:</p> <p>2. White, Tom. "Hadoop: The Definitive Guide, Storage and Analysis at Internet Scale." White-London: O'Reilly Media(2015).</p> <p>3. Ryza, Sandy, et al. Advanced analytics with spark: patterns for learning from data at scale. " O'Reilly Media, Inc.", 2017.</p> <p>4. Narkhede, Neha, Gwen Shapira, and Todd Palino. Kafka: the definitive guide: real-time data and stream processing at scale. " O'Reilly Media, Inc.", 2017.</p> <p>5. Hueske, Fabian, and Vasiliki Kalavri. Stream processing with Apache Flink: fundamentals, implementation, and operation of streaming applications. O'Reilly Media, 2019.</p>
Öğretim Üyesi (Üyeleri) / Faculty Member (Members)	Öğr. Gör. Dr. Burak YÖNYÜL	Teaching Assistant Dr. Burak YÖNYÜL

### ÖĞRENME ÇIKTILARI / LEARNING OUTCOMES

1	Veri-yoğun uygulamaların temel ilkelerini kavramak	Understand the principles of data intensive applications
2	Toplu ve akış işleme paradigmlarını gerçek dünya senaryoları üzerinde uygulamak	Apply batch and stream processing paradigms in real-world scenarios
3	Yüksek performanslı ve dağıtık veri mimarileri tasarlamak ve eniyileştirmek	Design and optimize high-performance distributed data architectures
4	Gelişmiş çözümler yapabilmek amacıyla uçtan uca veri hatları tasarlayabilmek ve gerçekleştirebilmek	Build and deploy end-to-end data pipelines for advanced analytics

## HAFTALIK DERS İÇERİĞİ / DETAILED COURSE OUTLINE

Hafta / Week					
1	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Yüksek Performanslı Veri İşlemeye Giriş			Kaynak-1	
	Introduction to High-Performance Data Processing			Reading-1	
2	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Veri Modelleri ve Sorgu Dilleri			Kaynak-1	
	Data Models and Query Languages			Reading-1	
3	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Dağıtık Sistemlerde Veri Depolama ve Veri Alımı			Kaynak-1	
	Storage and Retrieval in Distributed Systems			Reading-1	
4	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Hadoop ile Toplu İşlemeye Giriş			Kaynak-2	
	Introduction to Batch Processing with Hadoop			Reading-2	
5	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Apache Spark ile Gelişmiş Toplu İşleme			Kaynak-3	
	Advanced Batch Processing with Apache Spark			Reading-3	

	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
6	Veri Boru Hatları ve İş Akışı Orkestrasyonu			Kaynak-3	
	Data Pipelines and Workflow Orchestration			Reading-3	
7	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Apache Kafka ile Akış İşlemeye Giriş			Kaynak-4	
	Introduction to Stream Processing with Apache Kafka			Reading-4	
8	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Ara Sınav				
	Midterm Exam				
9	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Apache Flink ile Akış İşleme			Kaynak-5	
	Stream Processing with Apache Flink			Reading-5	
10	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Akış İşlemede Hata Toleransı ve Tutarlılık			Kaynak-4, 5	
	Fault Tolerance and Consistency in Stream Processing			Reading-4, 5	
11	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
	Lambda ve Kappa Mimarileri			Kaynak-1	
	Lambda and Kappa Architectures			Reading-1	

	Teorik Dersler / Theoretical	Uygulama	Lab	Öğretim Yöntem ve Teknikleri/Teaching Methods Techniques	Ön Hazırlık / Preliminary
12	Veri İşlemede Performansı Optimize Etme			Kaynak-3, 5	
	Optimizing Performance in Data Processing			Reading-3, 5	
13	Yüksek Performanslı Veri İşlemede Durum Çalışmaları			Kaynak-3, 4, 5	
	Case Studies in High-Performance Data Processing			Reading-3, 4, 5	
14	Veri İşlemede İleri Konular			Kaynak-3	
	Advanced Topics in Data Processing			Reading-3	
15	Ders Özeti ve Proje Sunumları / Kontrolü			Sunum ve Tartışma	
	Course Recap and Project Presentations & Control			Presentation and Discussion	
16	Final Sınavı				
	Final Exam				

DEĞERLENDİRME / EVALUATION

Yarıyıl (Yıl) İçi Etkinlikleri / Term (or Year) Learning Activities	Sayı / Number	Katkı Yüzdesi / Percentage of Contribution (%)
Ara Sınav / Midterm Examination	1	50
Proje Hazırlama / Project Preparation	1	50
Toplam / Total:	2	100
Başarı Notuna Katkı Yüzdesi / Contribution to Success Grade(%):		60

  

Yarıyıl (Yıl) Sonu Etkinlikleri / End Of Term (or Year) Learning Activities	Sayı / Number	Katkı Yüzdesi / Percentage of Contribution (%)
Final Sınavı / Final Examination	1	100
Toplam / Total:	1	100
Başarı Notuna Katkı Yüzdesi / Contribution to Success Grade(%):		40

  

Etkinliklerinin Başarı Notuna Katkı Yüzdesi(%) Toplamı / Total Percentage of Contribution (%) to Success Grade:	100
Değerlendirme Tipi / Evaluation Type:	DDS

İŞ YÜKÜ / WORKLOADS

Etkinlikler / Workloads	Sayı / Number	Süresi (Saat) / Duration (Hours)	Toplam İş Yüğü (Saat) / Total Work Load (Hour)
Ara Sınav / Midterm Examination	1	20.00	20.00
Derse Katılım / Attending Lectures	14	3.00	42.00
Final Sınavı / Final Examination	1	20.00	20.00
Okuma / Reading	14	2.00	28.00
Proje Hazırlama / Project Preparation	1	40.00	40.00
Toplam / Total:	31	85.00	150.00

Dersin AKTS Kredisi = Toplam İş Yüğü (Saat) / 30.00 (Saat/AKTS) = 150.00/30.00 = 5.00 ~ 5.00 / Course ECTS Credit = Total Workload (Hour) / 30.00 (Hour / ECTS) = 150.00 / 30.00 = 5.00 ~ 5.00

PROGRAM VE ÖĞRENME ÇIKTISI / PROGRAM LEARNING OUTCOMES

Öğrenme Çıktıları / Learning Outcomes	Program Çıktıları / Program Outcomes										
	1.1.1	1.1.2	1.1.3	1.1.4	1.1.5	1.1.6	1.1.7	1.1.8	1.1.9	1.1.10	1.1.11
1. Veri-yoğun uygulamaların temel ilkelerini kavramak / Understand the principles of data intensive applications	5						2				
2. Toplu ve akış işleme paradigmasını gerçek dünya senaryoları üzerinde uygulamak / Apply batch and stream processing paradigms in real-world scenarios	5		3		4	4	3	3			
3. Yüksek performanslı ve dağıtık veri mimarileri tasarlamak ve eniyileştirmek / Design and optimize high-performance distributed data architectures	5	4	5	5	5	5	3	3		3	
4. Gelişmiş çözümler yapabilmek amacıyla uçtan uca veri hatları tasarlayabilmek ve gerçekleştirebilmek / Build and deploy end-to-end data pipelines for advanced analytics	5	4	5	5	5	5	3			3	

Katkı Düzeyi / Contribution Level : 1-Çok Düşük / Very low, 2-Düşük / Low, 3-Orta / Moderate, 4-Yüksek / High, 5-Çok Yüksek / Very high