

Ödev konuları ve grupları aşağıda belirtildiği gibidir.

Aşağıda verilen başlıklar altında bulunan uygulamalar gerçekleştiriliip rapor haline getirilip sunulacaktır.

Grup 1:

1-) “Modeling structured data using Python” başlığı altındaki githubdaki chapter 6 da bulunan veri kümesi kullanılarak evlerin konumlarını harita üzerinde fiyatlarıyla birlikte görselleştirmek, ev fiyatlarını etkileyen faktörleri incelemek, parametre-fiyat ilişkilerini grafiklerle göstermek, Gradient Boosting Regression yöntemiyle fiyat tahmini yapmak ve fiyatı etkileyen parametreleri en etkili olandan en az etkili olana doğru sıralanması.

Temel olarak kullanılması gereken araçlar : Kafka,Spark,Hadoop

2-) “Modeling bitcoin data points with python” uygulaması gerçekleştirilecektir.

Grup 2:

1-) “Modelling with Unstructured Data” Başlığı altındaki “Unstructured text analysis using R” başlığı altında gösterilenlerin özetlenmesi uygulanması ve sunulması. (Data ingestion, Data cleaning and transformation, Data visualisation, improving the model)

2-) “Modelling Twitter feeds using python” uygulaması gerçekleştirilecektir.

Grup 3:

1-) “Modelling with streaming data” ve “Streaming sensor data” başlıkları altında kalan uygulamalar özetlenip uygulanması ve sunulması.

2-) “Modelling weather data points with python”

Grup 4:

1-) Apache Flink, Apache Pulsar teknolojileri ve Apache Ray Framework lerin detaylı bir şekilde incelenip uygulanması ve sunulması.

2-) “Modelling IMDB data points with python” başlığı altında gösterilenlerin uygulanması.

Grup 5:

1-) Apache dusk, Apache iceberg ve Apache Snowflake detaylı bir şekilde incelenip uygulanması, raporlanması ve sunulması.

2-) “Gerçek zamanlı satış ve stok takip sistemi” isimli ödev adımları takip edilip uygulanacak.

Grup 6:

1. “Web Sunucu Log Analizi ve Dashboard Geliştirme“ isimli ödev adımları takip edilip uygulanacak.

2. (Loki, Grafana, ELK, ElasticSearch,Kibana) araçlarının raporlanması ve sunulması

Gerçek Zamanlı Satış ve Stok Takip Sistemi

Bir e-ticaret firması, ürün satış ve stok verilerini analiz ederek:

- Gerçek zamanlı olarak stok seviyelerini izlemek,
- Bölgesel satış eğilimlerini analiz etmek,
- Geçmiş verilere göre talep tahminlemesi yapmak,
- Verileri hem **analitik** hem de **operasyonel** sistemlerde kullanılabilir hale getirmek istiyor.

Veriler:

1. Satış Verisi (CSV veya JSON)

- Ürün ID, Müşteri ID, Zaman, Bölge, Satış Adedi, Fiyat

2. Stok Verisi (CSV veya JSON)

- Ürün ID, Depo ID, Mevcut Stok, Güncelleme Tarihi

Görevler / Ödev Adımları:

1. Veriyi Simüle Et veya Gerçek Veriyi Temizle

- Python ile veri üret (veya Kaggle benzeri sitelerden çek)
- Farklı formatlarda veri oluştur (CSV, Parquet)

2. Dask ile Büyük Veriyi İşle

- Verileri Dask ile yükle
- Satışların bölgeye göre toplamı, zaman bazlı eğilimler
- Günlük stok düşüşlerini tespit et (günlük talep = günlük satış)
- Temel ETL işlemleri (tarih formatları, boş veri temizliği)

3. Iceberg'e Yaz ve Versiyonlama

- Iceberg tabloları oluştur (örneğin: sales_data, stock_data)
- Zaman damgası ve bölgeye göre böülümlendirme (partitioning)
- Veri ekleme, silme ve zaman içinde versiyon oluşturma
- Iceberg tablosunda "güncel stok durumu" gibi bir görünüm (view)

4. Snowflake'e Bağlan ve Verileri Aktar

- Iceberg → Snowflake aktarımı (PyArrow, Snowpipe, veya Dask + Connector)
- Alternatif: Snowflake External Table üzerinden doğrudan okuma
- Satış verisi üzerinde Snowflake SQL ile analizler:
 - En çok satılan ürünler
 - Bölge bazlı satış eğilimi
 - Stok yetersizliği uyarıları

5. Opsiyonel: Dashboard ile Görselleştirme

- Snowflake verisini Power BI, Tableau veya Apache Superset üzerinden bağlayarak görselleştir
- Satış ve stoklara ilişkin dinamik grafikler

Apache Dask, Apache Iceberg ve Snowflake detaylıca raporlanacak.

Proje Başlığı: Web Sunucu Log Analizi ve Dashboard Geliştirme

Proje Amacı

Bir web sunucusunun (örneğin Nginx veya Apache) ürettiği log dosyaları analiz edilerek:

- En yoğun trafik saatleri,
- Hangi IP'lerden, hangi endpoint'lere erişim sağlandığı,
- Hatalı istekler (404, 500 gibi),
- Bot veya anormal trafik davranışları,

tespit edilecek ve bu bilgiler **görselleştirilmiş dashboardlar** aracılığıyla izlenecektir.

Seçenek 1: ELK Stack

| Bileşen | Açıklama |
|---------------|--|
| Logstash | Logları alır, işler ve Elasticsearch'e gönderir |
| Elasticsearch | Verilerin indekslendiği ve sorgulandığı dağıtık arama/veritabanı |
| Kibana | Dashboard oluşturma ve veri görselleştirme aracı |

Seçenek 2: Grafana + Loki

| Bileşen | Açıklama |
|---------|--|
| Loki | Prometheus benzeri log toplayıcı sistem |
| Grafana | Loki'den logları çeker ve dashboard'larla görselleştirir |

Kullanılacak Veri: Web Sunucu Logları

Log alanları:

- IP adresi
- Tarih/saat
- HTTP metodu ve endpoint
- HTTP yanıt kodu (200, 404, 500)
- User agent

Yapılacak Adımlar

Log Üretime veya Toplanması

- **Gerçek loglar:** Bir Nginx/Apache sunucusu üzerinden alınabilir.
- **Yapay log üretimi (ödev için):**
Python ile 10.000 satırlık sahte log üretilebilir.

Log İşleme (Parsing)

ELK için:

- **Logstash config** dosyası yazılır:

Loki için:

- Loglar bir dizine yazılır ve Loki'nin agent'ı (Promtail) bu logları okur.
- Promtail config ile regex tanımlanır.

Logların Elasticsearch / Loki'ye Aktarılması

- Logstash → Elasticsearch
- Promtail → Loki

Dashboard Oluşturma

Kibana (ELK)

Örnek dashboard bileşenleri:

- Zaman serisine göre istek sayısı (time histogram)
- En çok erişilen endpoint'ler (bar chart)
- En sık gelen IP adresleri (table)
- HTTP status kod dağılımı (pie chart)
- Anlık log görünümü (log viewer paneli)

Grafana (Loki)

- Log stream viewer
- Status code heatmap
- Query: `{job="nginx"}` ile endpoint'e göre filtreleme
- Alert tanımlamaları (örneğin: 500 hataları artarsa uyarı)

Opsiyonel: Anomali Tespiti ve Alarm Mekanizması

- Kibana veya Grafana üzerinde:
 - 5xx hataları 1 dakikada 10'dan fazla olursa uyarı
 - Belirli IP'den aşırı istek varsa uyarı

Grup1:

Melih Eröz
Şevval Duygu Örklezmez
Rehyan Çolak
Muhammed Irfan Bakar
Semanur Enç
Egemen Bozca
Furkan Ay
Dilara Gülbabaş
Rabia Varol
HATİCE TURAN
Alperen İsmet ESEN

Grup 2:

Tarık Kalyoncu
MAHMOUD ALDAHER
Recep Salih Kılınç
Umut Kuzyaka
MUSTAFA SAIT KARADAĞ
ÜMRANAY ADIYAMAN
BURAK CAN EVCİ
Ali Kutay KILINÇ
SHHADA AL HARIRI
Hüdanur Duranoğlu
Şimal Ece Kazdal

Grup 3 :

MUHAMMED YUSUF YAĞCI
MEHMET CAN ERMURAT
İsmail Demir
Muhammed Taha Duran
AHMET SEVİLİ
Hasan Abdulhay
Buğra Bastaban
Yasemin Sude Yalçınkaya
YOUNES RAHEBI
Yusuf Taha Ezgin

Grup4:

Beyzanur Karaçam
YAREN CERRAH
Umut Direk
Reyhan Genç
Mustafa Alperen Akça
UĞUR KAYA
Berkay Göçer
YAVUZ SELİM SAY

İSMAİL UYGAR KESKİN

Samet Güzel

Grup5:

Ayşegül Bilici
Şevval Nur Kalaycı
Gülseher Başağa
MUHAMMED BAHA BAKAN
SIMA BANNOUD
Semih Özçaka
Emine Demirbaş
Berat Alpsar
İbrahim Talha Durna
Nourjan Ahmet

Grup6:

Özgür Akbay
Alper Zeybek
MUHAMMED NECİB TAVLAŞOĞLU
Abdulsamed Kurubal
Sude Dönmez
HÜSEYİN GÖBEKLİ
Najma Bakdach
Ayşe Verda Gülcemal
İbrahim ALİKILIÇ
Şeyma Orhan