
Capstone Project: Customer Order Insights & Delivery Tracker

Objective:

Create a simple system that helps track customer orders, delivery status, and generate useful insights — using MySQL, MongoDB, Python, PySpark, Azure Databricks, and Azure DevOps.

Week 1 - Database Foundations: MySQL & MongoDB

Tools: MySQL, MongoDB

Capstone Tasks:

- Design MySQL tables for customers, orders, and delivery_status
- Perform basic CRUD operations on orders
- Write a stored procedure to fetch all delayed deliveries for a customer
- Store customer feedback (unstructured) in MongoDB
- Index MongoDB collection to search by customer ID

Deliverables:

- SQL script with tables + CRUD + stored procedure
 - MongoDB script with sample feedback data and indexing
-

Week 2 - Data Processing with Python

Tools: Python (Pandas, NumPy, Requests)

Capstone Tasks:

- Use Python to load customer order data from an API or CSV
- Clean missing fields and convert timestamps
- Use NumPy to calculate delivery delays
- Show top delayed customers and most common delivery issues

Sample Code Snippet: ```python import pandas as pd import numpy as np

```
df = pd.readcsv('orders.csv') df['deliverydate'] =  
pd.to_datetime(df['deliverydate']) df['delaydays'] = (pd.Timestamp.today() -  
df['deliverydate']).dt.days df['delayed'] = np.where(df['delay_days'] > 0, 1, 0)
```

```
print(df.groupby('customerid')  
['delayed'].sum().sortvalues(ascending=False)) ```
```

Deliverables:

- Cleaned and processed order dataset
 - Python script that prints delay summary by customer
-

Week 3 - Order Analysis with PySpark

Tools: PySpark**Capstone Tasks:**

- Load order and delivery data into PySpark
- Join orders and customer tables
- Group by region to count delays
- Save results to a file

Deliverables:

- PySpark script with join, group, and aggregation
 - Output file showing delayed orders by region
-

Week 4 - Simple ETL Pipeline in Azure Databricks

Tools: Azure Databricks**Capstone Tasks:**

- Load cleaned order data into Databricks
- Create a pipeline to update latest delivery status
- Save the results as Delta or CSV
- Optional: run a SQL query to show top 5 delayed customers

Deliverables:

- Databricks notebook with ETL steps
 - Output stored in Delta/CSV
-

Week 5 - Automation using Azure DevOps

Tools: Azure DevOps**Capstone Tasks:**

- Build an Azure DevOps pipeline to run the full Python analysis
- Log delay summary to a file
- Send a basic success notification (print/log)

Deliverables:

- Azure DevOps YAML pipeline
 - Output log file
-

Final Outcome by Week 5:

By the end of this project, participants will have built a simple system that: -
Tracks orders and deliveries across MySQL & MongoDB

- Analyzes customer delays with Python
 - Processes large datasets using PySpark
 - Runs lightweight ETL in Azure Databricks
 - Automates execution with Azure DevOps
-