Capstone Project: Supply Chain Monitoring and Optimization Platform

Objective:
Build a practical system to track orders, inventory, and shipments using real-world tools like **MySQL**, **MongoDB**, **Python**, **PySpark**, **Azure Databricks**, and **Azure DevOps**.

Week 1 – Database Foundations: MySQL & MongoDB

Tools: MySQL, MongoDB

Capstone Tasks:

- Create MySQL tables for orders, suppliers, inventory
- Perform basic CRUD operations
- Write stored procedures (e.g., auto reorder trigger)
- Store shipment logs in MongoDB
- Create indexes for efficient querying

Deliverables:

- MySQL script with schema + CRUD + stored procedure
- MongoDB script with sample data and indexing

Week 2 – Data Collection & Preprocessing in Python

Tools: Python (Pandas, NumPy, Requests)

Capstone Tasks:

- Use `requests` to fetch data from a sample API or local JSON/CSV
- Clean data using `pandas` (drop nulls, format dates, etc.)
- Perform basic calculations using `numpy` (e.g., delays, stock levels)
- Display the cleaned and processed data

Sample Code Snippet:

```
import pandas as pd
import numpy as np
import requests

# Load sample data
data = requests.get("https://api.sampledata.com/orders").json()
df = pd.DataFrame(data)

# Clean and transform
df['delivery_date'] = pd.to_datetime(df['delivery_date'])
df['delay_days'] = (pd.Timestamp.today() - df['delivery_date']).dt.days
df['is_delayed'] = np.where(df['delay_days'] > 0, 1, 0)

print(df[['order_id', 'supplier_id', 'delay_days', 'is_delayed']].head())
```

Deliverables:

- Python script for collecting and processing supply chain data
- Cleaned `pandas` DataFrame with processed outputs

Week 3 – Intro to PySpark: Processing Big Data

Tools: PySpark

Capstone Tasks:

- Load order data from CSV using PySpark
- Filter delayed shipments
- Group by supplier and count delayed orders
- Save processed data to CSV or Parquet

Deliverables:

- PySpark script to load, process, and save supply chain data
- Output file showing grouped results

Week 4 – Simple ETL in Azure Databricks

Tools: Azure Databricks

Capstone Tasks:

- Upload CSV data into Databricks
- Run a notebook to clean and filter the data
- Save cleaned output as Delta or CSV
- Run basic analysis queries using SQL or PySpark

Deliverables:

- Azure Databricks notebook
- Cleaned output stored in Delta/CSV format

Week 5 – Deploy & Automate with Azure DevOps

Tools: Azure DevOps

Capstone Tasks:

- Create a simple Azure DevOps pipeline to run Python scripts
- Install dependencies and execute the project
- Log the results and mark completion

Sample YAML:

```
trigger:
  - main

pool:
  vmImage: 'ubuntu-latest'

steps:
  - script: |
      pip install -r requirements.txt
      python run_pipeline.py
    displayName: 'Run Supply Chain Script'
```

Deliverables:

- Azure DevOps YAML pipeline file
- Execution log or console output