
Capstone Project: Retail Inventory & Sales Performance Tracker

Objective:

Build a data-driven system that monitors product inventory levels, analyzes sales performance, and provides actionable retail insights — using **MySQL, MongoDB, Python, PySpark, Azure Databricks (executed in Google Colab)**, and **Azure DevOps**.

✅ Task 1 – Database Foundations: MySQL & MongoDB

Tools: MySQL, MongoDB

Capstone Tasks:

- Design MySQL tables for `products`, `sales`, and `inventory`.
- Perform basic CRUD operations on products and sales data.
- Write a stored procedure to identify low-stock items.
- Store customer or supplier feedback (unstructured) in MongoDB.
- Create a MongoDB index for quick search by `product_id`.

Deliverables:

- SQL script with tables, CRUD operations, and stored procedure.
- MongoDB script with sample feedback data and indexing.

✅ Task 2 – Data Processing with Python

Tools: Python (Pandas, NumPy, Requests)

Capstone Tasks:

- Load sales and inventory data from an API or CSV file.
- Clean missing values and format date/time columns.
- Use NumPy to calculate monthly sales and inventory turnover.
- Generate reports on top-selling and underperforming products.

Deliverables:

- Python script for data cleaning and sales analysis.
- Processed dataset summarizing sales performance.

✅ Task 3 – Retail Insights with PySpark

Tools: PySpark

Capstone Tasks:

- Load sales and inventory datasets into PySpark DataFrames.
- Join product, sales, and inventory data.
- Group data by region or category to analyze sales trends.
- Export aggregated results as CSV or Parquet files.

Deliverables:

- PySpark script performing joins and aggregations.

- Output file showing category-wise sales by region.

☁ Task 4 – ETL Pipeline in Azure Databricks (Executed in Google Colab)

Tools: Google Colab (used for Azure Databricks simulation)

Capstone Tasks:

- Load cleaned sales and inventory data into Colab notebook.
- Simulate an ETL pipeline to update latest sales and stock data.
- Save results as Delta-like or CSV outputs.
- Run SQL-style queries in Colab to identify top 5 best-selling products.

Deliverables:

- Google Colab notebook mimicking Databricks ETL steps.
- Output data stored in CSV/Parquet format.

☁ Task 5 – Project Automation and Tracking in Azure DevOps

Tools: Azure DevOps

Capstone Tasks:

- Create a single **Epic** in Azure DevOps named “Retail Inventory & Sales Performance Tracker.”
- Students should then break it down later into Features, Tasks, and User Stories inside the DevOps portal.
- Capture and submit screenshots showing the Epic creation and initial setup.

Deliverables:

- Screenshot of the Epic created in Azure DevOps.
- Evidence of portal setup and naming consistency.
