
Capstone Project: Personal Expense Monitoring System

Objective:

Build a basic system to track and analyze personal or household expenses. The system should help users categorize spending, identify unusual patterns, and automate monthly summaries.

Week 1 - Database Foundations: MySQL & MongoDB

Tools: MySQL, MongoDB

Capstone Tasks:

- Create MySQL tables for users, expenses, categories
- Perform basic CRUD operations (add/edit/delete expenses)
- Write a stored procedure to calculate monthly total expenses per category
- Store user notes or scanned receipts (as JSON) in MongoDB
- Add indexing in MongoDB for quick user/receipt lookup

Deliverables:

- SQL schema + CRUD + stored procedure
 - MongoDB script with sample receipt data and indexes
-

Week 2 - Data Processing with Python

Tools: Python (Pandas, NumPy)

Capstone Tasks:

- Load expense data from CSV or API
- Clean and standardize formats (e.g., dates, amounts)
- Use numpy to calculate monthly totals and averages
- Use pandas to create a breakdown of expenses by category

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

```
df = pd.readcsv('expenses.csv') df['amount'] = df['amount'].replace(['\$','], '', regex=True).astype(float) df['date'] = pd.to_datetime(df['date']) df['month'] = df['date'].dt.to_period('M')
```

```
monthlyexpense = df.groupby(['month', 'category'])['amount'].sum().unstack().fillna(0) print(monthlyexpense) ```
```

Deliverables:

- Cleaned dataset with monthly summaries
 - Python script for category-wise breakdown
-

Week 3 - PySpark for Transaction Volume Analysis

Tools: PySpark

Capstone Tasks:

- Load a large set of expense transactions
- Group by user to calculate total monthly spend
- Detect unusual spikes or large one-time expenses

Deliverables:

- PySpark script with grouping and simple anomaly detection logic
 - Output showing users with potential unusual spending
-

Week 4 - ETL in Azure Databricks

Tools: Azure Databricks

Capstone Tasks:

- Upload cleaned data into Databricks
- Combine user and expense data
- Create a summary table with monthly spend, savings, and alerts
- Save as Delta or CSV for dashboards

Deliverables:

- Databricks notebook with ETL steps
 - Final report stored in Delta/CSV
-

Week 5 - CI/CD Pipeline with Azure DevOps

Tools: Azure DevOps

Capstone Tasks:

- Build a pipeline that runs expense analysis weekly or monthly
- Output summary report as CSV
- Log or print a savings alert if expenses exceed threshold

Deliverables:

- Azure DevOps YAML pipeline
 - Output file of monthly analysis
-

Final Outcome by Week 5:

- A functional personal finance tracker
 - Stores data in MySQL and MongoDB
 - Uses Python for processing and summaries
 - Analyzes large data using PySpark
 - ETL runs in Databricks
 - Fully automated using Azure DevOps
-