
Capstone Project: Smart Home Energy Usage Tracker

Objective:

Build a lightweight system to collect, process, and analyze smart home energy consumption data — helping users monitor usage, identify wastage, and track device-level power trends.

Week 1 - Database Foundations: MySQL & MongoDB

Tools: MySQL, MongoDB

Capstone Tasks:

- Create MySQL tables for devices, energy_logs, and rooms
- Perform CRUD operations to update device status and usage records
- Write a stored procedure to calculate total energy usage per room per day
- Use MongoDB to store raw sensor logs (JSON format)
- Add indexes for quick lookup by device ID or timestamp

Deliverables:

- SQL schema + CRUD operations + stored procedure
 - MongoDB script with sample sensor readings and indexing
-

Week 2 - Data Preprocessing in Python

Tools: Python (Pandas, NumPy)

Capstone Tasks:

- Load energy logs from CSV, API, or sensor simulator
- Clean missing or malformed readings
- Use numpy to calculate total and average energy per device
- Use pandas to generate room-level summaries

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

```
df = pd.readcsv("energyusage.csv") df['timestamp'] =  
pd.to_datetime(df['timestamp']) df['energykwh'] =  
df['energy_kwh'].astype(float)
```

```
roomsummary = df.groupby('roomid')['energykwh'].sum()  
print(roomsummary) ```
```

Deliverables:

- Cleaned dataset with usable energy stats
 - Python script generating summaries by room and device
-

Week 3 - Spark for Device-Level Aggregation

Tools: PySpark

Capstone Tasks:

- Load a large dataset of sensor logs using PySpark
- Group by device and calculate peak vs off-peak usage
- Identify top energy-consuming devices

Deliverables:

- PySpark script for aggregation and filtering
 - Output file of top devices by usage
-

Week 4 - Databricks ETL for Smart Energy Monitoring

Tools: Azure Databricks

Capstone Tasks:

- Upload cleaned logs to Databricks
- Build an ETL pipeline to calculate daily/weekly summaries
- Save final results in Delta format or CSV
- Optional: use SQL to query energy savings opportunities

Deliverables:

- Databricks notebook with full ETL pipeline
 - Output files for dashboards or alerts
-

Week 5 - CI/CD Automation with Azure DevOps

Tools: Azure DevOps

Capstone Tasks:

- Automate the energy report generation weekly
- Pipeline to fetch new data, clean, and summarize
- Alert or log if usage crosses a threshold (e.g., >10 kWh per device per day)

Deliverables:

- Azure DevOps pipeline YAML
 - Execution log and report output
-

Final Outcome by Week 5:

- Smart home device data stored and processed end-to-end
 - MySQL and MongoDB for structured/unstructured data
 - Python and PySpark for data cleaning and analysis
 - Databricks for ETL and metrics
 - Azure DevOps for automation and alerting
-