Capstone Project: Employee Attendance and Productivity Tracker

Objective:
Build a lightweight system to track employee attendance and productivity across departments. The goal is to collect data, identify patterns (like frequent absenteeism or underperformance), and generate reports for HR.

Week 1 – Database Foundations: MySQL & MongoDB

Tools: MySQL, MongoDB

Capstone Tasks:
- Create MySQL tables for `employees`, `attendance`, and `tasks`
- Perform CRUD operations (e.g., clock-in, clock-out)
- Write a stored procedure to calculate total working hours per employee
- Use MongoDB to store unstructured task feedback and notes
- Create indexes for fast querying by `employee_id` or `department`

Deliverables:
- SQL script with schema, CRUD, and stored proc
- MongoDB script with sample notes and index

Week 2 – Attendance & Task Data in Python

Tools: Python (Pandas, NumPy)

Capstone Tasks:
- Read attendance logs and task tracking data from CSV or mock API
- Clean missing or invalid entries
- Use `numpy` to calculate work hours, break times, and productivity scores
- Use `pandas` to find top performers and frequent absentees

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

df = pd.readcsv("attendance.csv") df['workhours'] = (pd.todatetime(df['clockout']) - pd.todatetime(df['clockin'])).dt.totalseconds() / 3600 df['productivityscore'] = df['taskscompleted'] / df['workhours']

```
summary = df.groupby('employeeid')['workhours',
'productivity_score'].mean() print(summary) ```
```

Deliverables:
- Cleaned attendance and task dataset
- Python report of top and bottom performers

Week 3 – PySpark for Attendance Analysis

Tools: PySpark

Capstone Tasks:
- Load large attendance logs in PySpark
- Filter for late logins and absences
- Group by department to get average work hours and productivity

Deliverables:
- PySpark script with filtering and group aggregations
- Output showing attendance issues by department

Week 4 – ETL in Azure Databricks

Tools: Azure Databricks

Capstone Tasks:
- Load employee attendance and task records into Databricks
- Clean and combine data to create department-level metrics
- Save output in Delta or CSV format for visualization/dashboard use

Deliverables:
- Databricks notebook with full ETL process
- Exported file showing attendance and task KPIs

Week 5 – Automation via Azure DevOps

Tools: Azure DevOps

Capstone Tasks:
- Set up a DevOps pipeline to automate weekly processing
- Schedule the pipeline to run every Monday
- Output a report with top 5 absentees or lowest performing departments

Deliverables:
- YAML pipeline file
- Report/log file of latest attendance metrics

Final Outcome by Week 5:

- A fully functional system that tracks employee hours, tasks, and productivity
- Automated reporting pipeline with actionable HR insights
- Uses MySQL + MongoDB + Python + PySpark + Databricks + Azure DevOps in a real scenario