CLOUD APPLICATION DEVELOPMENT

BIG DATA ANALYIS ITH IBM CLOUD DATABASES

Phase 1: Problem Definition and Design Thinking

TEAM DETAILS:

TEAM NAME:

TEAM MEMBER NAMES	COLLEGE CODE	REGISTER NO	EMAIL ID
KOTIPATRUNI TIRUMALARAO	3128	312821205302	22le-it02@act.edu.in
DHARSHAN.K	3128	312821205004	22it07@act.edu.in
RAHUL.N	3128	312821205032	21it39@act.edu.in
DINAKARAN.S.A	3128	312821205005	21it08@act.edu.in
SANTHSOH.G	3128	312821205037	21it55@act.edu.in
JAGADEESH.D	3128	312821205010	21it15@act.edu.in

Phase 1: Problem Definition and Design Thinking

Data Selection	Database Setup	Data Exploration	Analysis Techniques	Visualization	Business Insights

Problem Definition:

Project Objective: The project aims to perform comprehensive big data analysis using IBM Cloud Databases. The primary objectives include extracting valuable insights from extensive datasets that cover a wide spectrum of domains, such as climate trends and social patterns. This analysis process involves:

1. Designing the Analysis Process

2. Setting up IBM Cloud Databases

3. Performing Data Analysis

4. Visualizing Results for Business Intelligence

Design Thinking:

To achieve our project objectives, we will follow a structured design thinking approach, as outlined below:

1. Data Selection:

Objective: Identify and select the datasets to be analysed, ensuring alignment with project objectives.

Approach:
- Define Specific Goals: Establish the key objectives for the dataset selection process.
- Identify Data Sources: Explore various sources, including online repositories, APIs, and proprietary providers.
- Data Acquisition Strategy: Develop a strategy for gathering and preparing the data for analysis.

2. Database Setup:

Objective: Configure IBM Cloud Databases to effectively store and manage large volumes of data.

Approach:

- Select Database Type: Choose IBM Db2 as the database technology within IBM Cloud.
- Database Configuration: Configure IBM Db2 to ensure scalability, reliability, and data security.
- Data Ingestion: Implement methods for ingesting data into IBM Db2.

3.Data Exploration:

Objective: Develop queries and scripts to explore the datasets, extract relevant information, and identify patterns.

Approach:

- Develop Data Exploration Scripts: Create Python scripts and queries to navigate and understand the dataset's structure and content.

4.Analysis Techniques:

Objective: Apply suitable analysis techniques, including statistical methods and machine learning, to unearth valuable insights.

Approach:

- **Select Analysis Methods:** Choose appropriate statistical analysis and machine learning techniques based on the dataset's characteristics.

5. Front-End and Back-End Design:

Objective: Develop a user-friendly web-based front-end (HTML/CSS) and a Flask-based back end for data visualization and interaction.

Front-End Design:

Objective: Design a user-friendly web-based front-end for data presentation and interaction**.**

Approach:

- User Requirement Analysis: Engage with stakeholders to gather specific requirements for data presentation and user interaction.
- HTML/CSS Selection: Utilize HTML/CSS to create responsive and accessible web-based data presentation components.
- UI Development: Create user interface elements and layouts for displaying data visualizations.
- Integration: Seamlessly integrate data visualizations, charts, and graphics into the front end.

Technologies: HTML/CSS for front-end development.

Back-End Design:

- **Objective:** Build a Flask-based back end for data handling, processing, and interaction**.**
- **Approach:**
- **Flask Selection:** Utilize Flask, a Python web framework, for robust and scalable back-end development.
- **API Design:** Define RESTful APIs to facilitate communication between the front end and back end.
- **API Development:** Build APIs for data retrieval and interaction**.**
- **Data Processing:** Implement data processing logic for real-time or on-demand analysis results.
- **Security Measures:** Incorporate security measures to protect data and user interactions.
- **Integration:** Ensure seamless communication between the back end and front end.

Technologies: Flask for back-end development**.**

Resource Requirements:

Hardware:

1. A high-performance server with at least GB RAM and multi-core processors for database hosting and data processing.
2. Backup and redundancy hardware for data storage.
3. A workstation with a system meeting the minimum requirements for development and analysis tools.

Software:

1. The operating system of choice is Windows for both the database server and development workstations.
2. IBM Db2 will serve as our chosen Database Management System (DBMS) for data storage.
3. Development tools include Python IDE (PyCharm), text editors (Visual Studio Code), and Jupyter Notebook for data analysis.

4. Web development tools such as HTML/CSS development tools (Visual Studio Code).
5. Flask, a Python framework, will be employed for back-end development.
6. For data analysis, we will leverage Python libraries (pandas, NumPy, scikit-learn).
7. Version control will be managed through Git.
8. Project management and collaboration tools like Jira.

Cloud Services:

IBM Cloud Services will be our primary cloud provider for hosting databases, web applications, and deployment.

Project Flow:

The overall flow of the project is as follows:
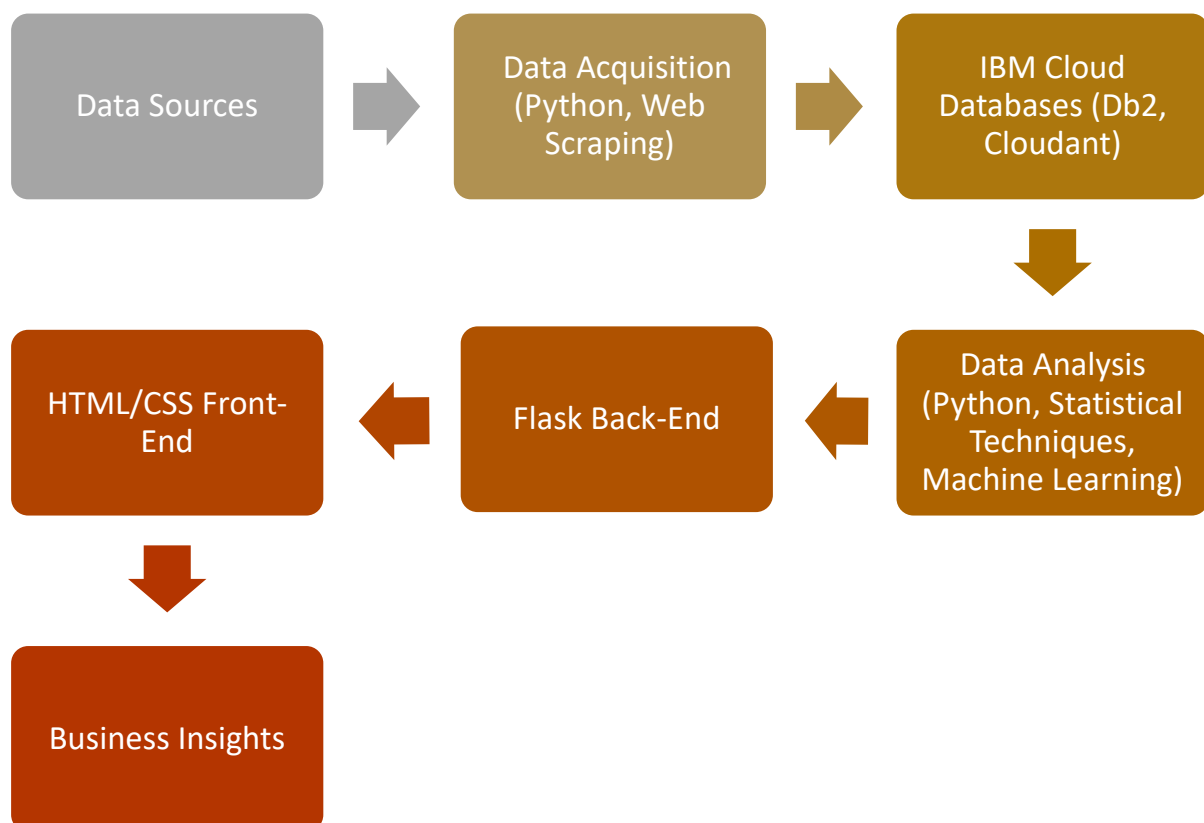
Components:

1.Data Sources: This component represents various data sources, including online repositories, APIs, and proprietary data providers. Data is collected from these sources for analysis.

2.Data Acquisition: In this part, Python scripts and web scraping tools are used to acquire data from the sources. Data is then prepared for storage and analysis.

3.IBM Cloud Databases: IBM Cloud Databases, including Db2 and Cloudant, serve as the central data storage and management systems. They ensure scalability, reliability, and security.

4.Data Analysis: Data analysis is performed using Python libraries, statistical techniques, and machine learning algorithms. This phase uncovers valuable insights from the datasets.

5.Flask Back-End: The Flask-based back-end component handles data processing and interaction with the front-end.

6. HTML/CSS Front-End: The HTML/CSS front-end enables stakeholders to visualize and interact with analysis results, providing a user-friendly web interface.

7. Business Insights: The final component involves interpreting analysis findings and provides actionable recommendations

Project Execution Plan:

1.Initiation Phase:

- Create Project Charter

2.Data Selection and Acquisition Phase:

- Identify Data Sources
- Develop Data Acquisition Scripts
- Implement Data Cleaning and Transformation

3.Database Setup Phase:

- IBM Db2 Configuration
- Configure IBM Db2
- Implement Data Ingestion

4.Data Exploration Phase:

- Develop Data Exploration Scripts

5.Analysis Techniques Phase:

- Select Analysis Methods
- Perform Data Analysis

6.Front-End and Back-End Development Phase:

1.Front-End Design:

- User Requirement Analysis
- HTML/CSS Development
- UI Development
- Integration

2.Back-End Design

- Flask Selection
- API Design
- API Development
- Data Processing
- Security Measures
- Integration

7.Visualization Phase

- Design Visualizations

8.Business Insights Phase

- Interpret Analysis Findings
- Deliver Actionable Recommendations

9.Testing and Quality Assurance Phase:

- Testing (Unit, Integration, User Acceptance)
- Quality Assurance (Data Validation, Error Handling)

10.Deployment and User Training Phase:

- Deploy Front-End and Back-End
- Provide User Training (if applicable)

11.Hold Project Closure Meeting:

- Archive Documentation
- Handover Deliverables

Key Elements to Include:

- Dependencies between tasks
- Start and end dates for each task.
- Responsible team members
- Milestones
- Timeline and task durations

Project timeline:

- **Start Date:** September 22,2023
- **End Data:** October 30,2023