**Index**

1. [**Introduction**](#Intro)
2. [**Objectives**](#objectives)
3. [**Problem Statement**](#Synopsis)
4. [**Hardware/ Software Requirements**](#hswreq)

**Members:**

|  |  |
| --- | --- |
| **Student1367861** | **Muhammad Umer** |
| **Student1367863** | **Haris Khan** |
| **Student1356627** | **Muhammad Usman Abdul Waris** |
| **Student1367013** | **Talha** |

**Introduction**

The thirst for learning, upgrading technical skills and applying the concepts in real life environment at a fast pace is what the industry demands from IT professionals today. However busy work schedules, far-flung locations, unavailability of convenient time-slots pose as major barriers when it comes to applying the concepts into realism. And hence the need to look out for alternative means of implementation in the form of laddered approach.

The above truly pose as constraints especially for our students too! With their busy schedules, it is indeed difficult for our students to keep up with the genuine and constant need for integrated application which can be seen live especially so in the field of IT education where technology can change on the spur of a moment. *Well, technology does come to our rescue at such times!!*

Keeping the above in mind and in tune with our constant endeavour to use Technology in our training model, we at Aptech have thought of revolutionizing the way our students learn and implement the concepts using tools themselves by providing a *live and synchronous eProject learning environment!*

**So what is this eProject?**

eProject is a step by step learning environment that closely simulates the classroom and Lab based learning environment into actual implementation. It is a project implementation at your fingertips!! An electronic, live juncture on the machine that allows you to

* Practice step by step i.e. laddered approach.
* Build a larger more robust application.
* Usage of certain utilities in applications designed by user.
* Single program to unified code leading to a complete application.
* Learn implementation of concepts in a phased manner.
* Enhance skills and add value.
* Work on real life projects.
* Give a real life scenario and help to create applications more complicated and useful.
* Mentoring through email support.

The students at the centre are expected to complete this eProject and send complete documentation with source code to eProjects Team

Looking forward to a positive response from your end!!

**Objectives of the project**

The Objective of this program is to give a sample project to work on real life projects. These applications help you build a larger more robust application.

The objective is not to teach you the concepts but to provide you with a real life scenario and help you create applications using the tools.

You can revise them before you start with the project.

It is very essential that a student has a clear understanding of the subject.

Kindly get back to eProjects Team in case of any doubts regarding the application or its objectives.

**Background**

Climate change is a global challenge that demands comprehensive analysis and proactive measures to mitigate its impacts. The increase in greenhouse gas emissions, deforestation, and changes in land use have led to unprecedented shifts in climate patterns, resulting in more frequent and severe weather events. To address this complex issue, the integration of big data technologies, particularly Hadoop, becomes crucial for processing, analyzing, and interpreting the vast amount of climate-related data available.

The project is initiated for a EarthScape Climate Agency, dedicated to monitoring and addressing climate change issues. The agency is responsible for collecting vast amounts of climate-related data from various sources, including satellites, weather stations, and environmental sensors. They seek a robust solution to process, analyze, and visualize this data to gain actionable insights for informed decision-making.

**Functional Requirements**

**User Authentication and Authorization:**

Implement a secure authentication system for users with different roles (e.g., administrators, analysts).

Define access controls to restrict data access based on user roles and responsibilities.

**Data Ingestion:**

The system should support the ingestion of diverse climate-related datasets, including satellite imagery, weather station records, and environmental sensor data.

Implement mechanisms to handle both historical and real-time data sources.

Ensure compatibility with common data formats used in climate science.

**Data Storage:**

Utilize the Hadoop Distributed File System (HDFS) for scalable and fault-tolerant storage of large climate datasets.

Implement data partitioning and organization strategies to optimize retrieval and processing.

**Data Processing:**

Implement Hadoop MapReduce jobs for parallel processing of climate data across distributed nodes.

Develop algorithms for the identification of climate patterns, anomalies, and correlations.

Include mechanisms to handle missing or incomplete data gracefully.

**Real-time Data Processing:**

Integrate real-time data streaming capabilities.

Ensure seamless integration with batch processing for a comprehensive analysis.

**Machine Learning Models:**

Develop machine learning models for predictive analysis of climate trends and impacts.

Include algorithms for anomaly detection, trend prediction, and correlation analysis.

Regularly update and refine models based on the latest available data.

**Data Visualization:**

Create interactive dashboards.

Develop visual representations of climate patterns, anomalies, and predictions.

Provide customizable and user-friendly interfaces for stakeholders to explore data

**Notifications and Alerts:**

Set up automated notifications and alerts for stakeholders based on predefined thresholds for climate anomalies or significant events.

Enable configurable alerting mechanisms to notify users in real-time.

**Feedback and Support:**

A support system for users to contact for assistance, report issues, and provide feedback.

**Non-Functional Requirements**

**Performance:**

Implement monitoring tools to track system performance, resource utilization, and data processing times.

Include optimization strategies for enhancing the overall efficiency of the system.

**Data Security:**

Implement encryption mechanisms to secure sensitive climate data during storage and transmission.

Ensure compliance with relevant data protection regulations and standards.

**Reliability:**

**Uptime:** The application should aim for a minimum of 99% uptime, with scheduled maintenance communicated in advance.

**Data Backup:** Regular automated data backups must be performed to prevent data loss in case of system failures.

**Scalability:**

The architecture should support horizontal scaling to accommodate increased data volumes and processing demands.

Implement load balancing mechanisms for optimal resource utilization.

**Performance Monitoring:**

Implement monitoring tools to track system performance, resource utilization, and data processing times.

Include optimization strategies for enhancing the overall efficiency of the system.

**Compliance and Standards:**

Ensure adherence to relevant environmental data standards and protocols.

Comply with industry best practices for big data processing and analytics.

**Documentation:**

**User Documentation:** Provide user guides, FAQs, and tutorials to help users understand and navigate the application.

**Developer Documentation:** Provide thorough documentation for system architecture, data processing workflows, and machine learning models. Maintain developer documentation to assist in further development and maintenance.

**Video:** Provide video displaying complete working of the application.

**[Hardware/ Software Requirements](#hswreq)**

**Hardware Requirements**

* Minimum i5 with 4 Cores but i7 is recommended.
* 16 GB of RAM
* 500 GB of SSD storage
* Graphics Card
* 64Bit Windows 10 or higher

**Software:**

* Jupiter Anaconda Notebook 3
* RStudio
* Visual Code Studio / PyCharm
* Mongo Db Compass, and Shell
* Hadoop,HDFS and Apache server
* Tableau
* Impala server

**Project Deliverables**

You will design and build the project and submit it along with a complete project

report that includes:

• Problem Definition

• Design specifications

• Diagrams such as flowcharts for various activities, Data Flow Diagrams etc.

• Source Code

• Test Data Used in the Project

• Project Installation Instructions (if any)

Documentation is considered as a very important part of the project. Ensure that documentation is complete and comprehensive. The consolidated project will be submitted as a zip file with a ReadMe.doc file listing assumptions (if any) made.

**Submit a video clip demonstrating the working of the Website.** Optionally, a live hosted URL can be supplied for the site. Over and above the given specifications, you can apply your creativity and logic to improve the portal.