BITS Pilani
Hyderabad Campus

Department Of Computer Science and Information Systems

A report on Advanced Database Management System Assignment

Project Title: Collaborative Sustainability: Leveraging Confidential Data for Shared Targets

Under the Supervision of

Dr. Subhrakanta Panada

Members

Kushal Chakraborty [2022H1030089H]

Mohammad Avesh Husain [2022H1030090H]

Aritra Kumar Dutta [2022H1030096H]

Utsav Seth [2022H1030074H]

Introduction: Data Lake

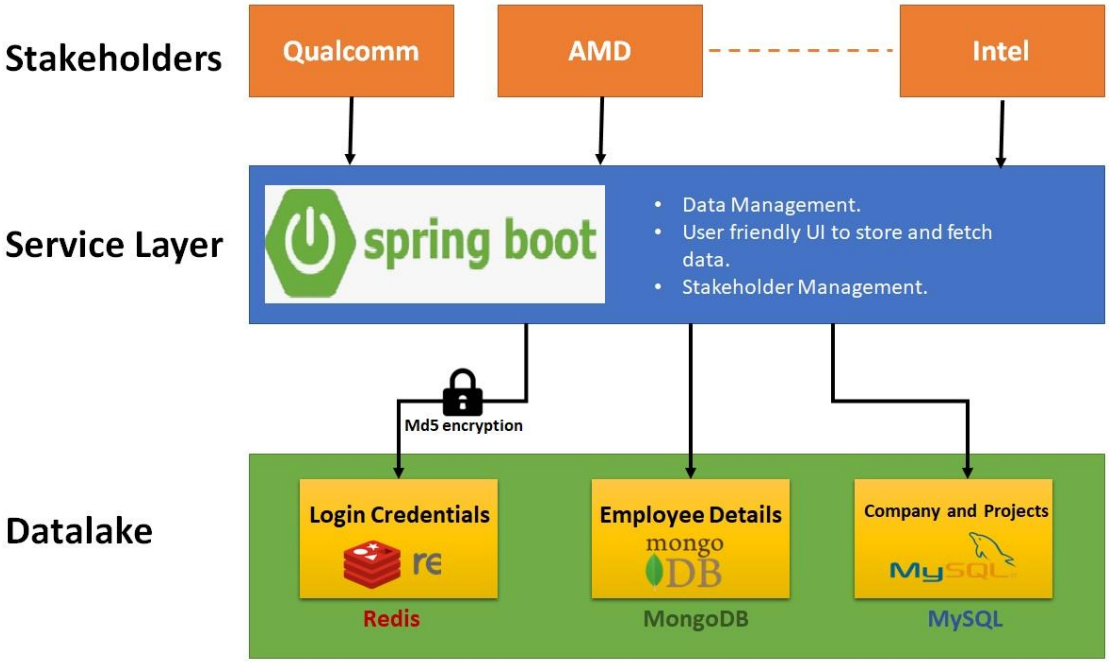
A data lake is a centralized repository that allows organizations to store, manage, and analyse large amounts of raw, unstructured, and structured data at scale. Unlike traditional data storage systems, which often require data to be pre-processed and structured before it can be stored and analysed, a data lake allows organizations to store all types of data in its native format. This means that data can be collected from a variety of sources, including social media, IoT devices, and other sources, without the need for extensive data transformation.

The goal of a data lake is to provide a cost-effective, scalable, and flexible solution for managing large amounts of data. Data lakes typically use distributed computing technologies such as Hadoop and Spark to store and process data across multiple servers or clusters. By storing data in its raw form, organizations can perform ad-hoc analysis on the data and gain valuable insights that might not have been possible using a more traditional data warehousing approach.

Technologies Used:

- Backend
    - Python
    - SpringBoot, SpringMVC
    - Encryption (MD5)
- Frontend
    - HTML
    - CSS
    - JavaScript
    - Ajax
    - jQuery
- Databases used
    - Reddis
    - MongoDB
    - Neo4J
    - SQL

System Architecture:

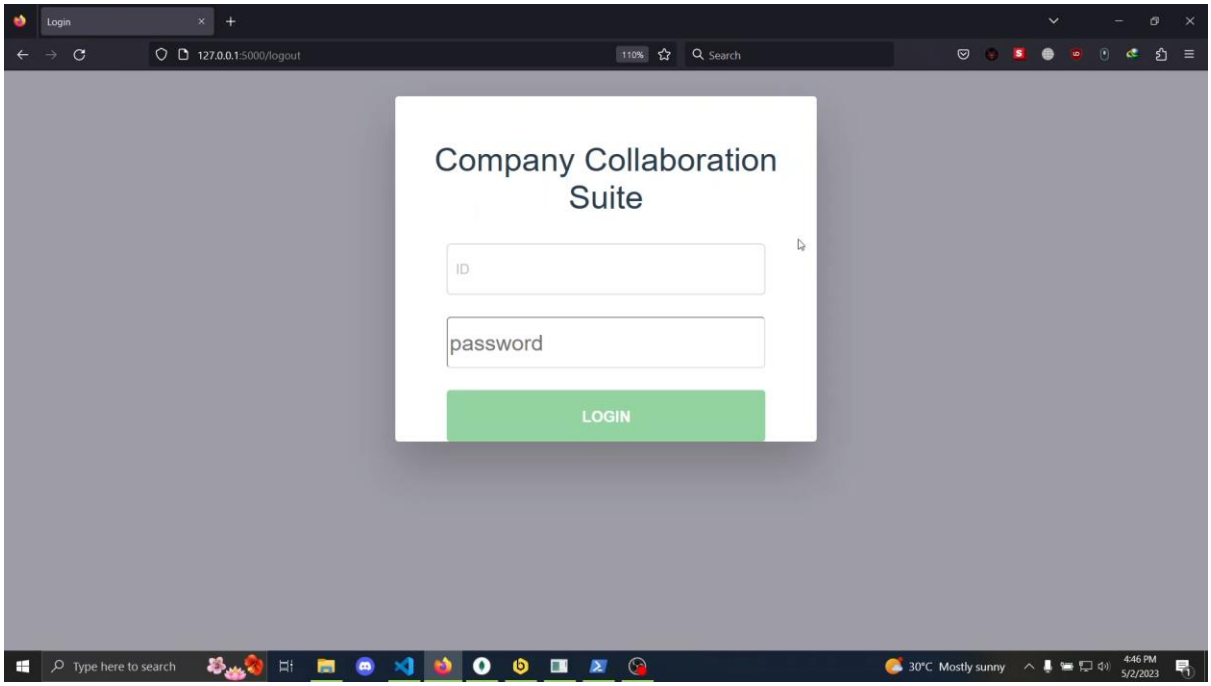


Architecture proposed in the above figure has been implemented to achieve Data Lake using multiple database(polyglot) architecture. The Scenario that has been assumed is that there are multiple companies who are working on several projects. Companies maintain their own projects and set of employee details, and it can also outsource a project to other companies and work in collaboration. Details regarding the employees and the projects are being stored in MongoDB and MySQL databases. Whereas the login credentials of each company have been hashed using MD5 Encryption and saved in Redis database.Neo4J is used for indexing the files for faster access.

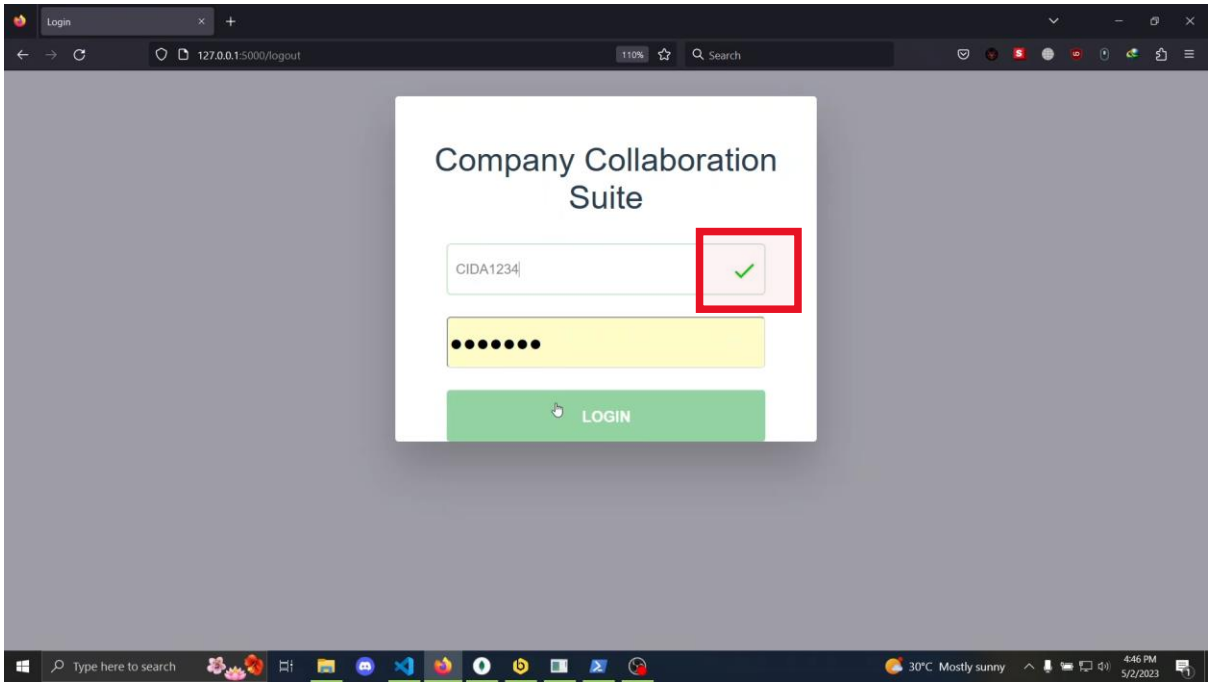
For backend Flask has been used which a light weight framework written in python programming language. For frontend technologies HTML, CSS, JavaScript and for request and response Ajax with jQuery has been used.

Project Demonstration:

Step 1: The web application is hosted on the local machine. And when accessed asks a user ID and password for a particular company.

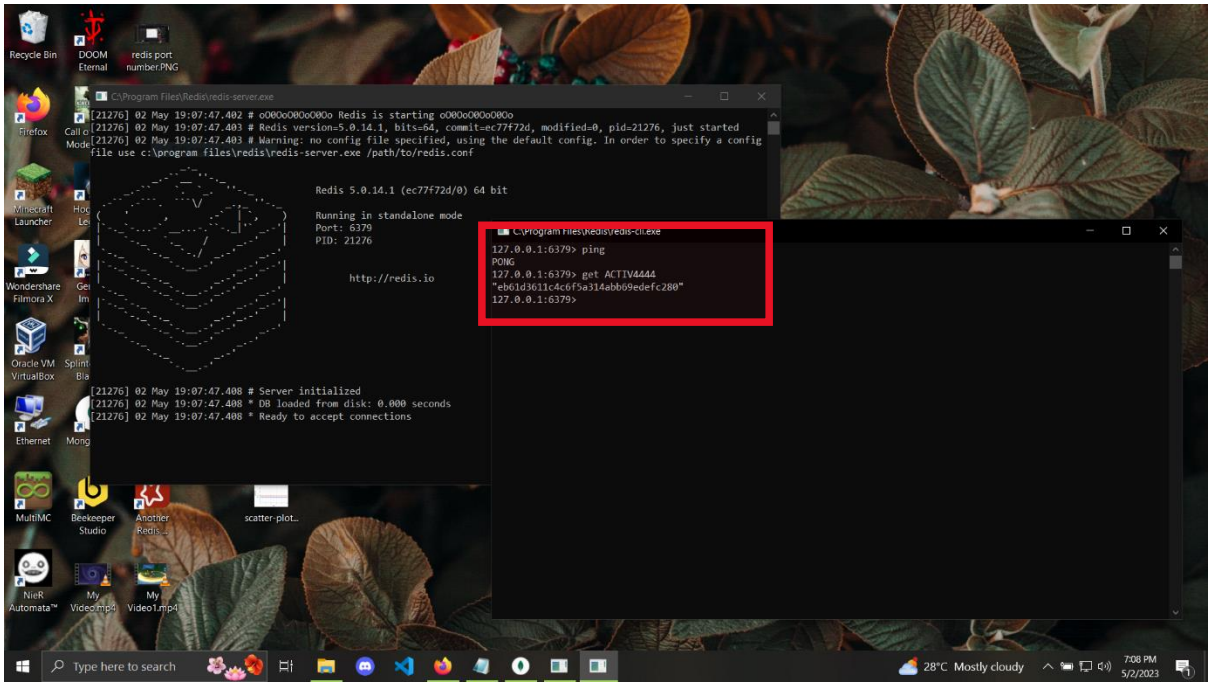


Step 2: Give a user name and Password and click on login

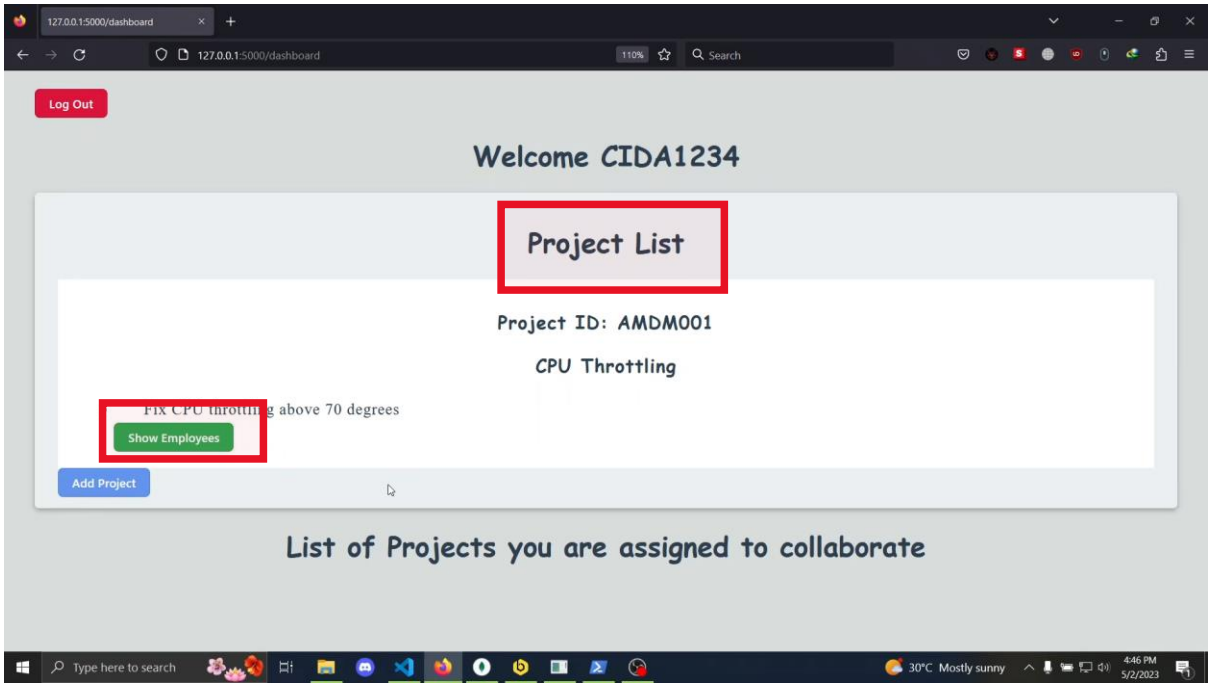


** when the username is valid a green check will appear

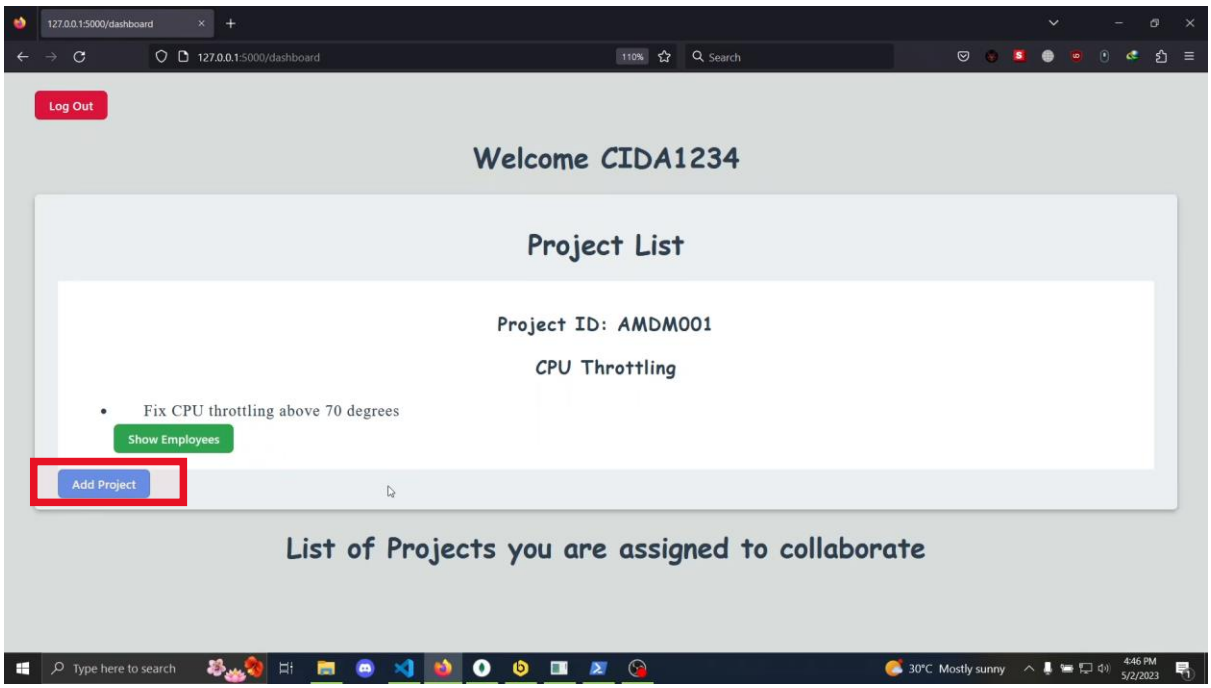
** Database related to Login Credential of the Companies has been Hashed using Md5 hashing function and stored in Redis database to provide security



** When we get inside companies profile we can see the Project list and Employee details of a particular project

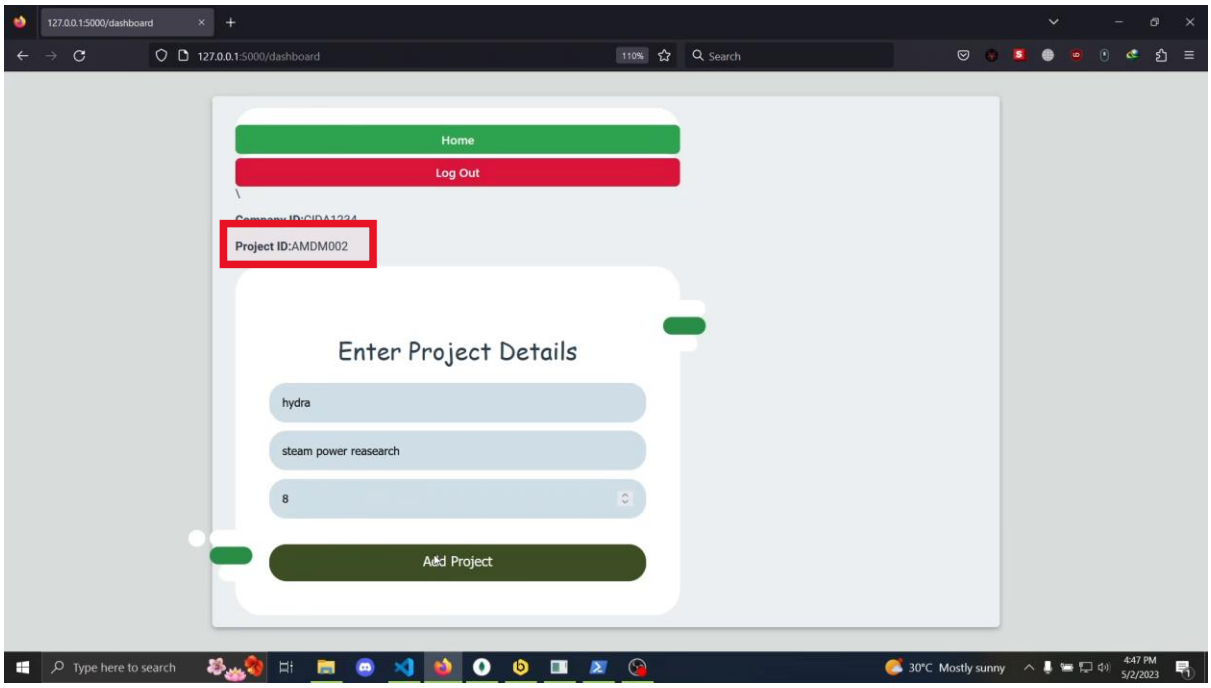


Step 3: Click on Add project to add a new project for a company

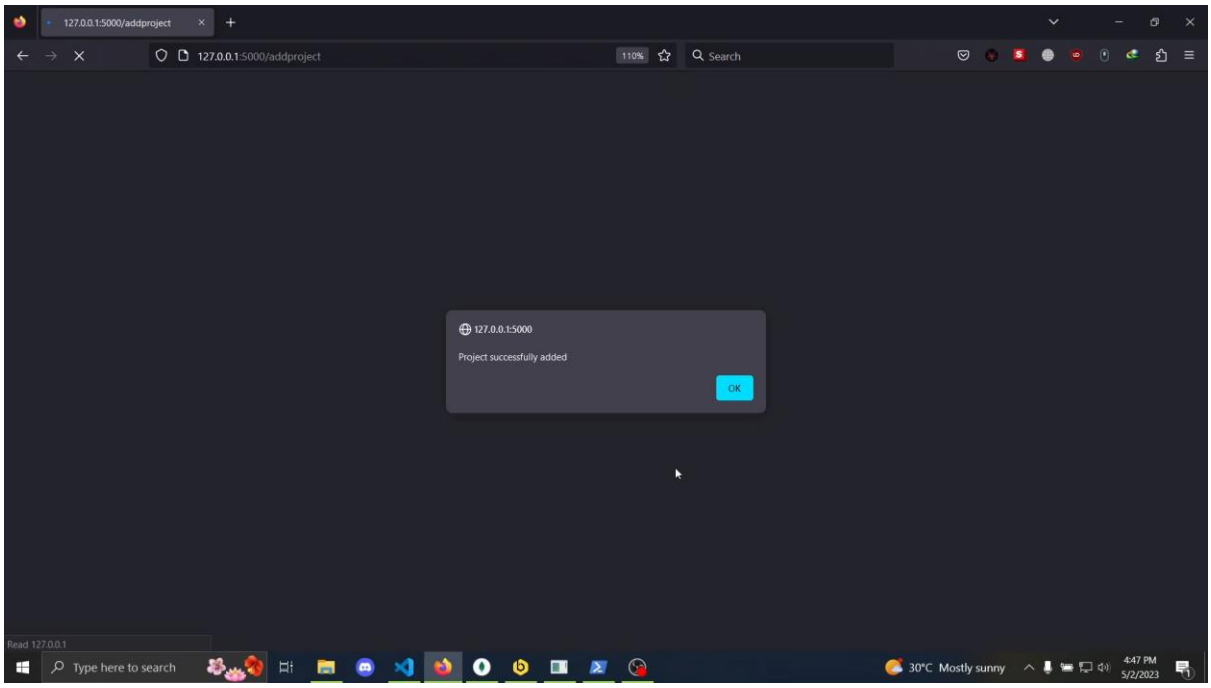


Step 4: A new window will pop up enter the relevant details for a project and click “Add project” button.

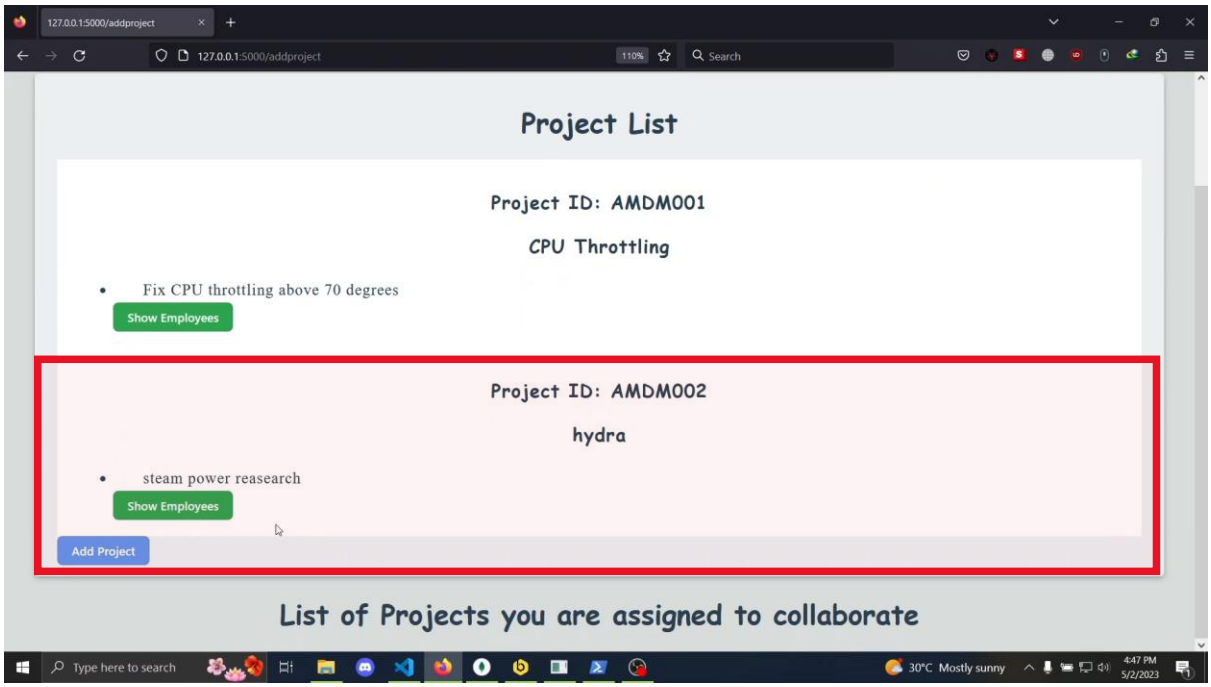
**A project id will automatically be generated



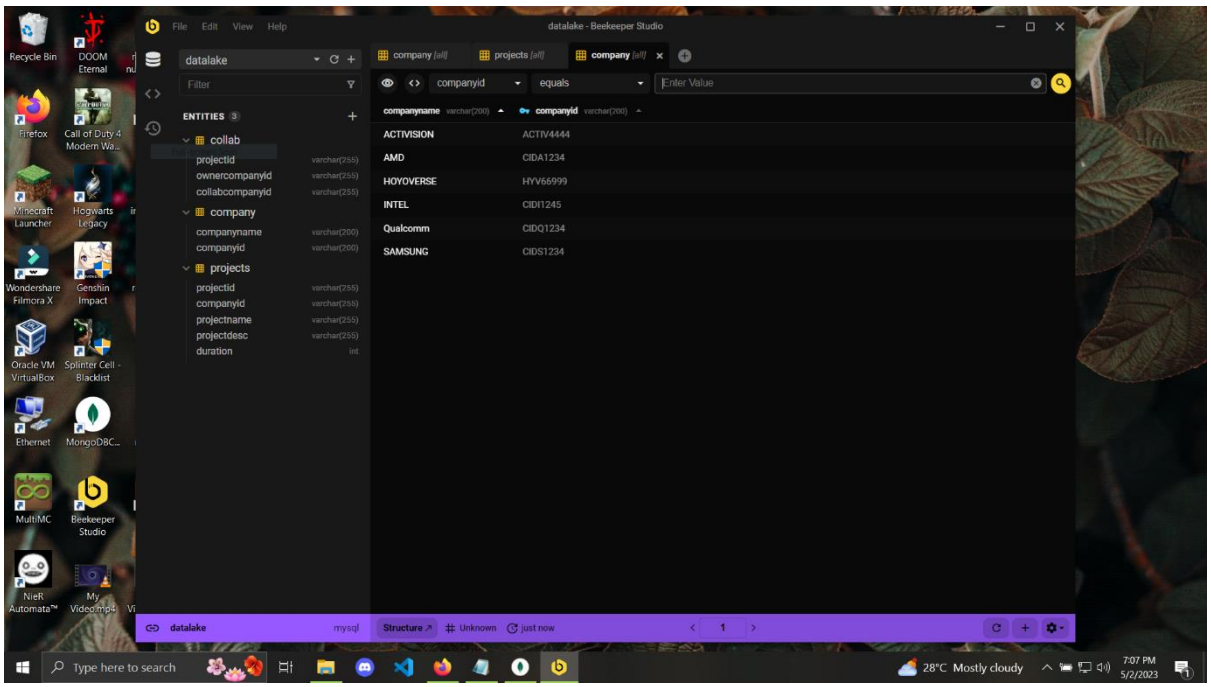
Step 5: A popup will alert about the successful addition of the project, Click "ok"



** Now we can see in the Companies Dashboard that the newly added project appears

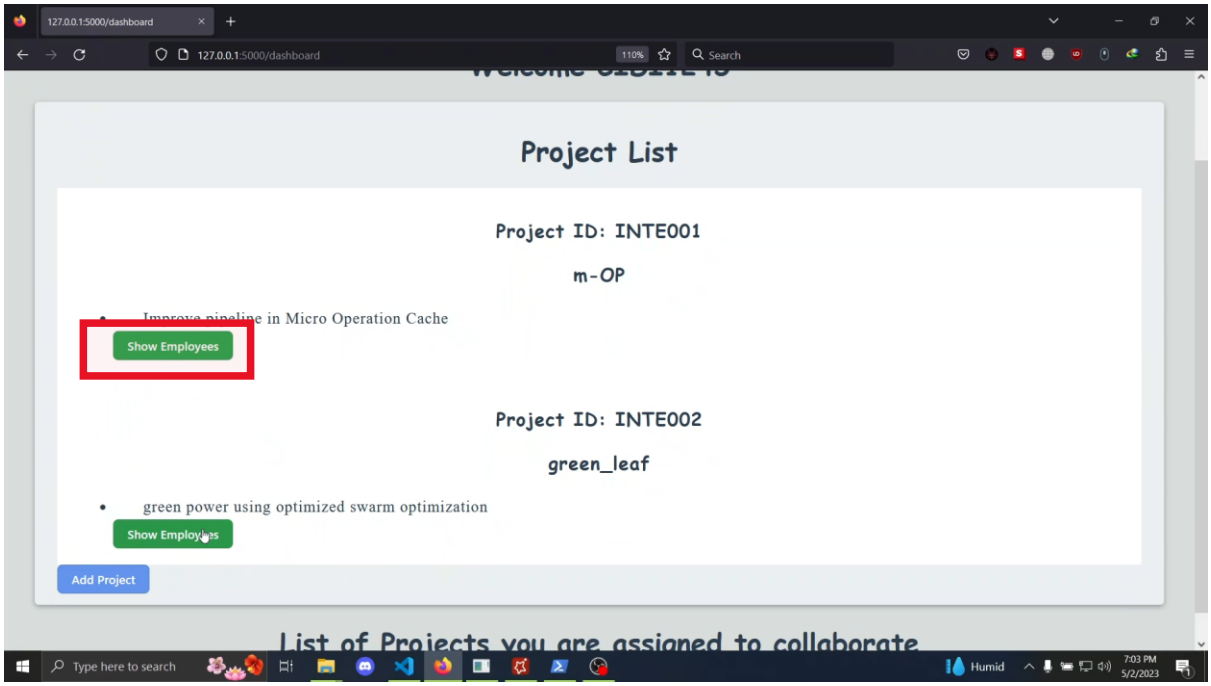


** All the details regarding the project of a Company is stored in SQL databses and hosted locally

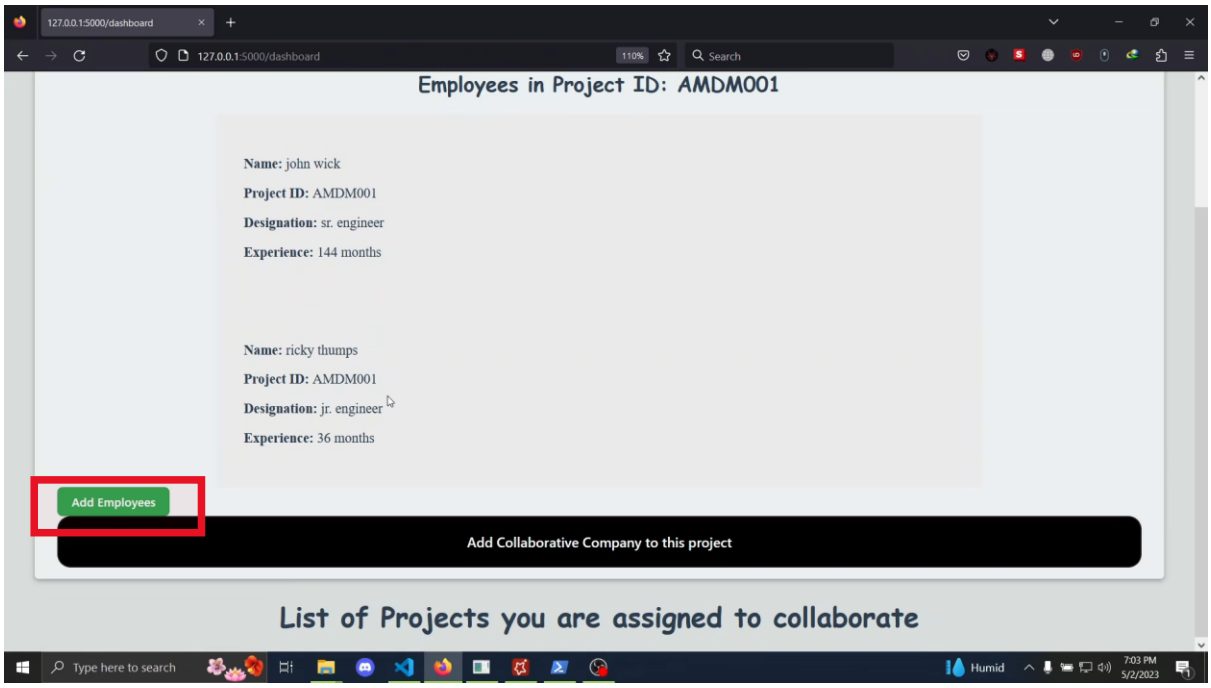


**We can also add Employee details for a particular project

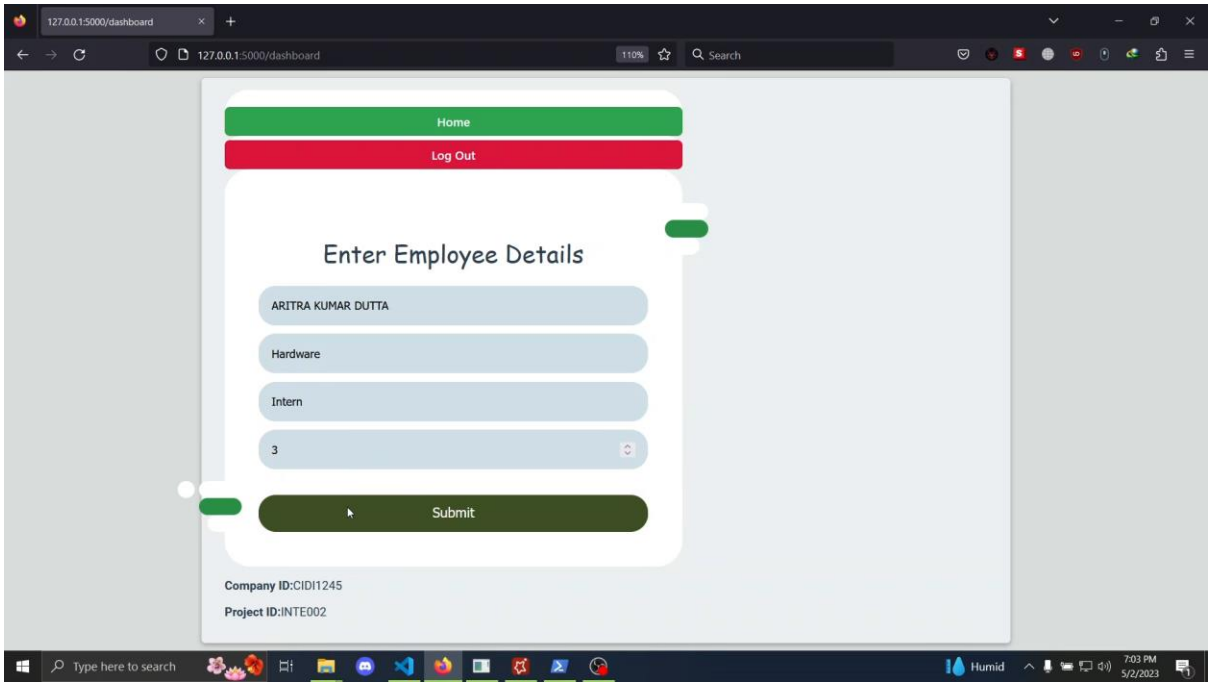
Step 6: login to a company and click on show employee of a project



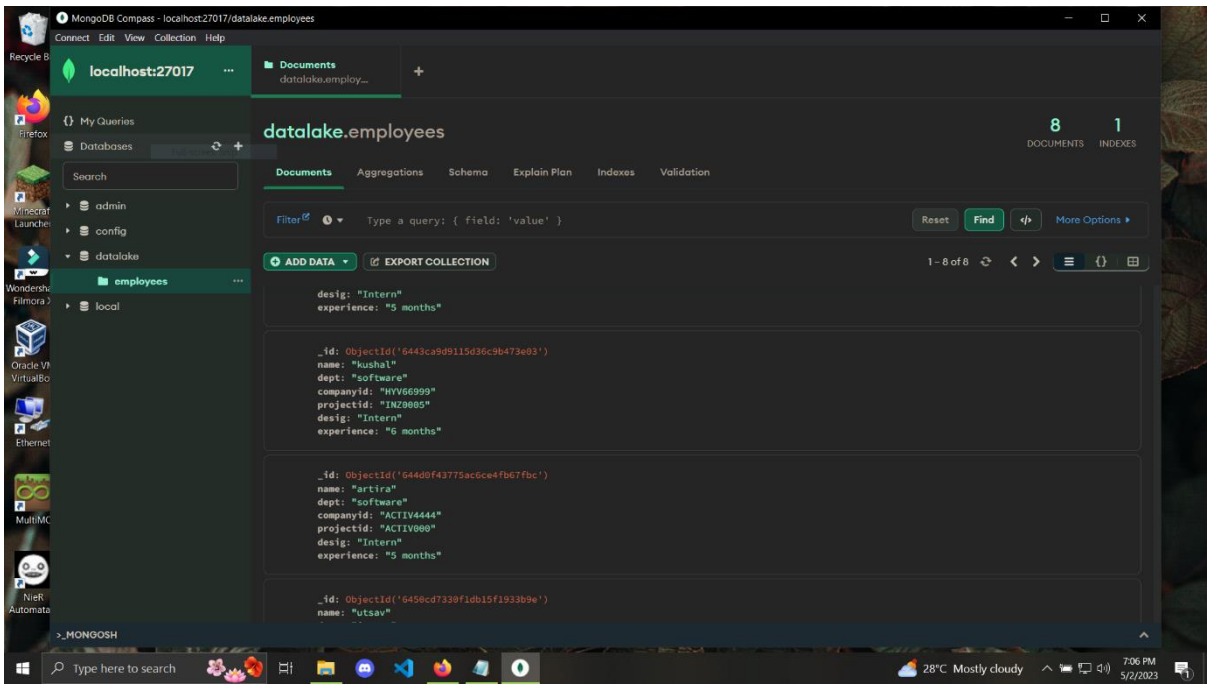
Step 7: List of employees and their details will be displayed click on Add Employees button



Step 8: Fill in the required details and Click on Add Employee

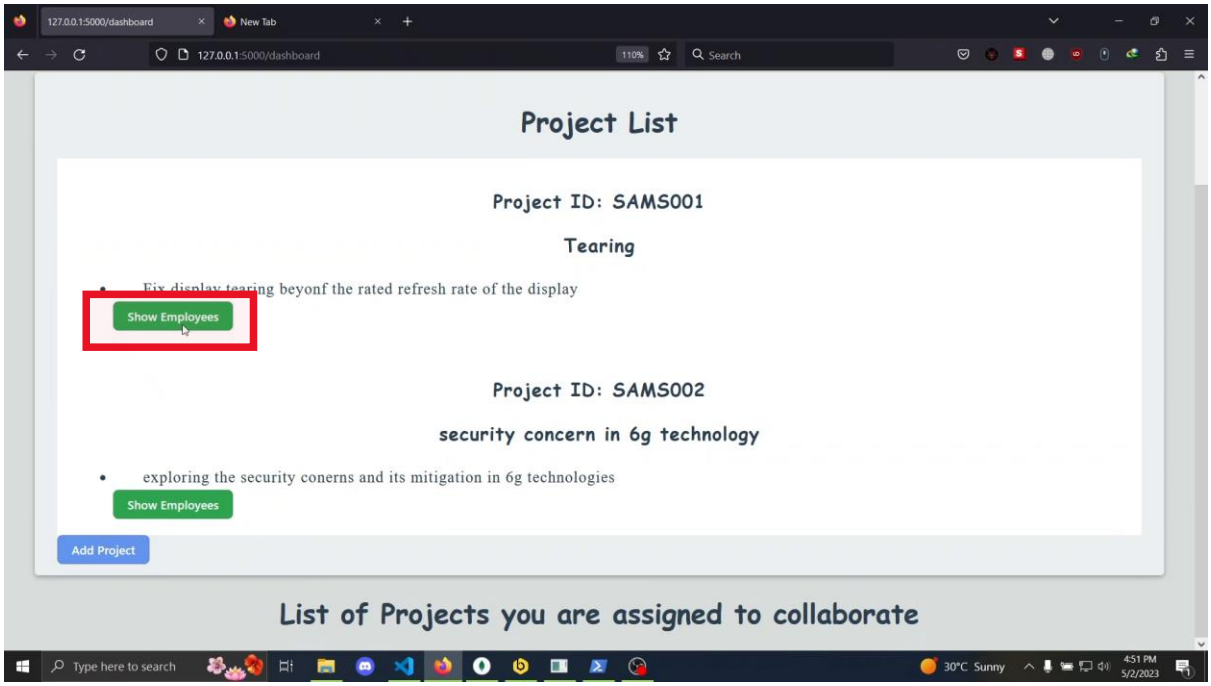


**** Employee details for Each coutry has been Stored using No-SQL MongoDB Database**

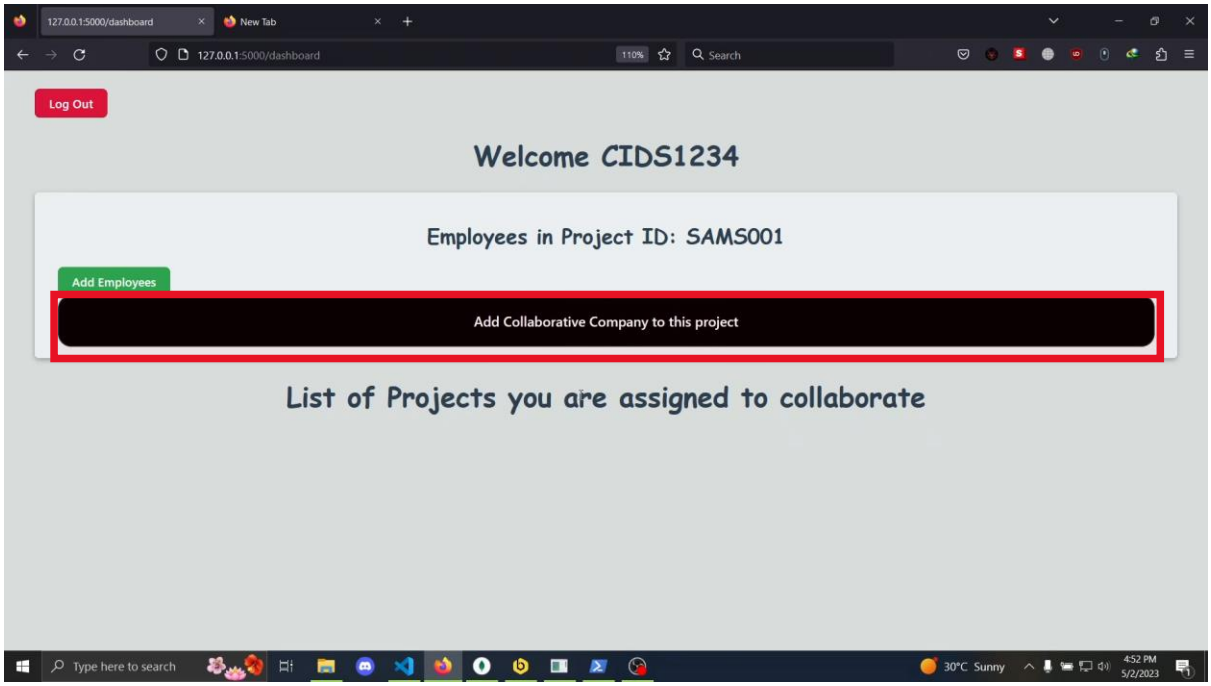


**** we can also add a collaborative company along with the parent company on a project**

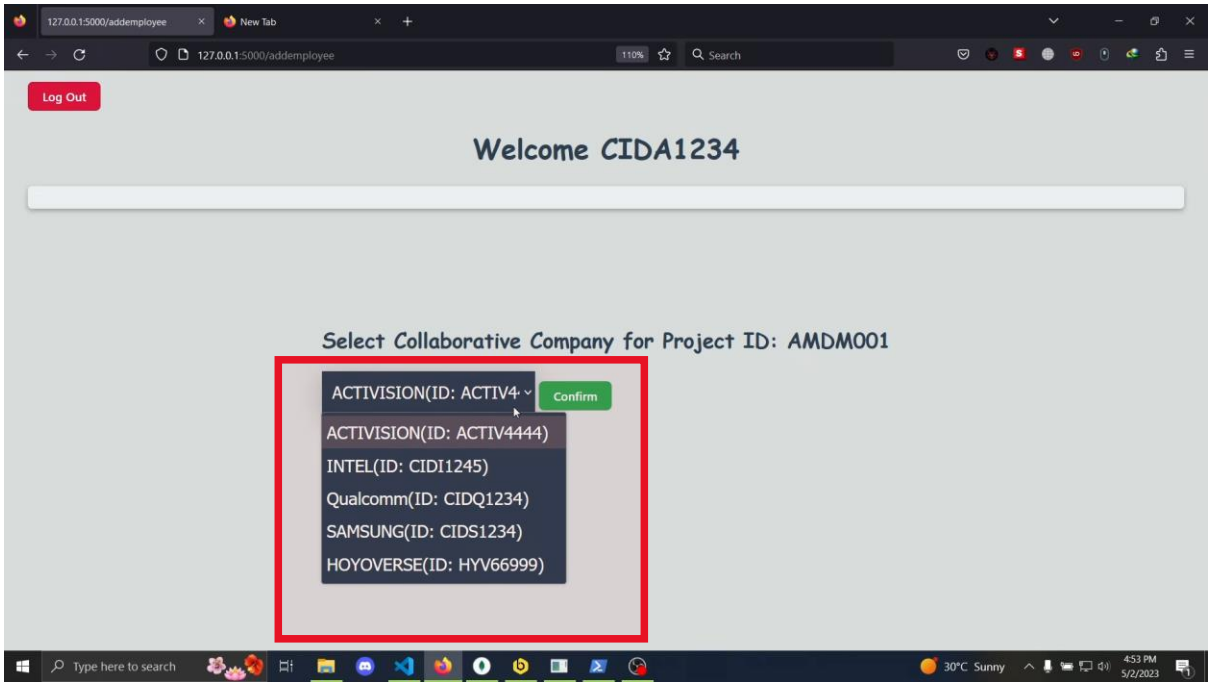
Step 9: Click on show Employee button



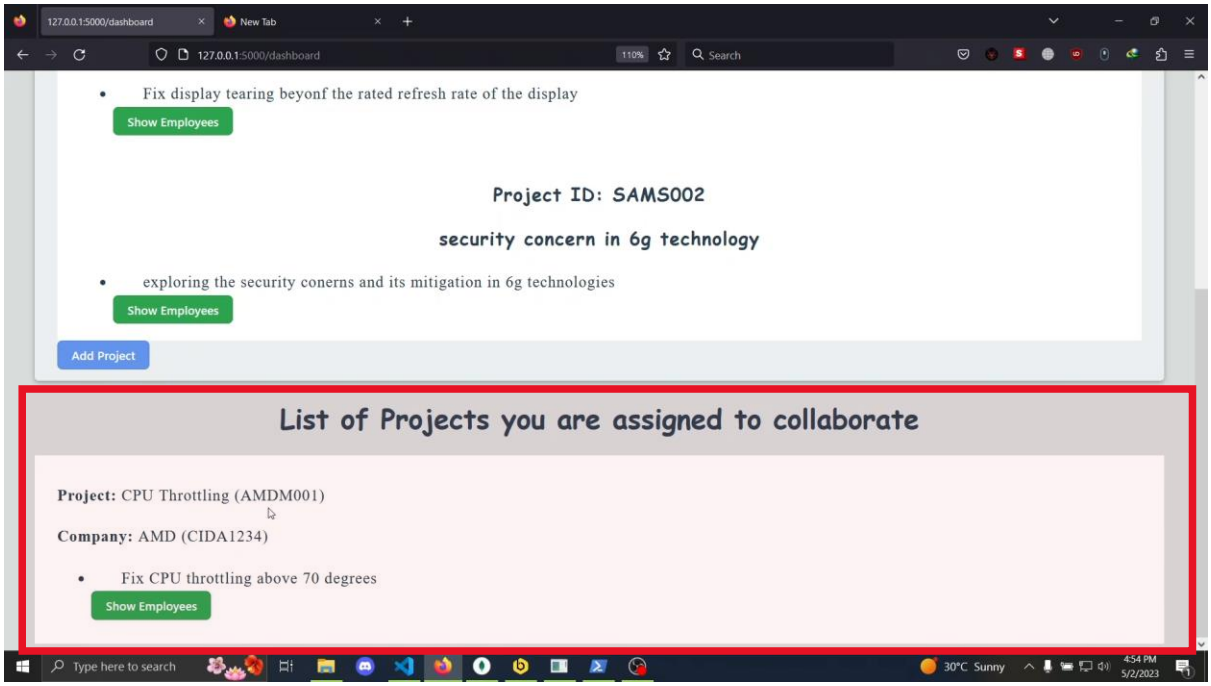
Step 10: Click on "Add Collaborative Company"



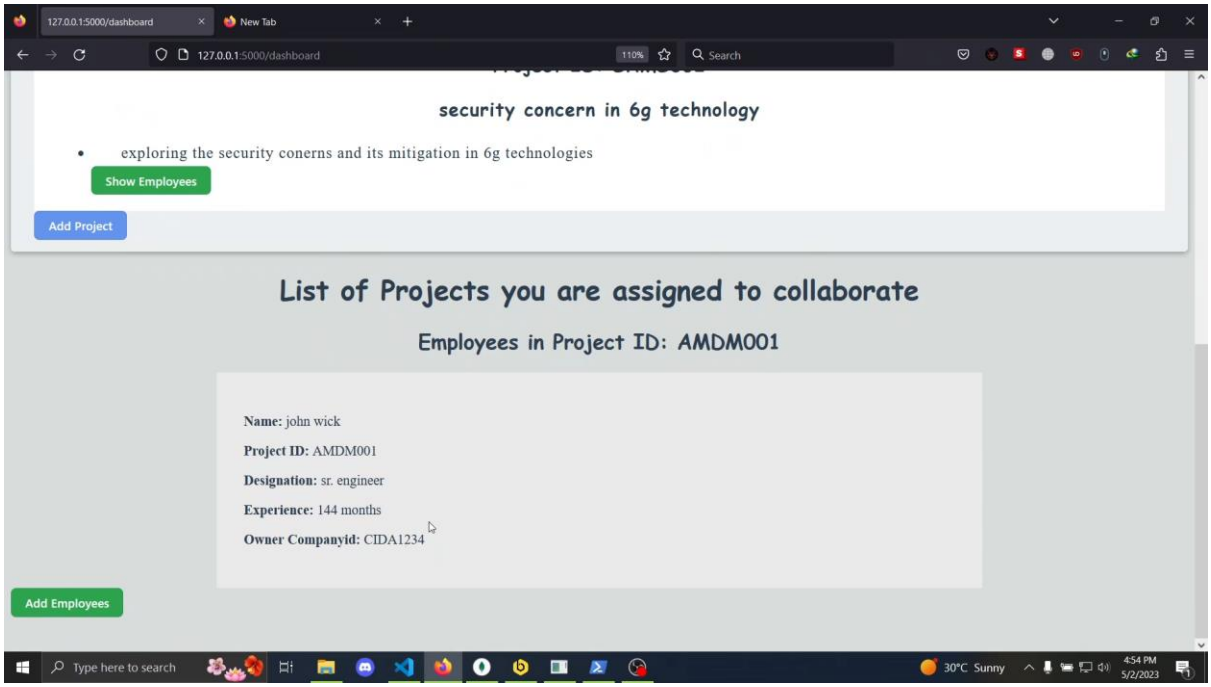
Step 11: From the drop down list select the company to collaborate with and click Confirm



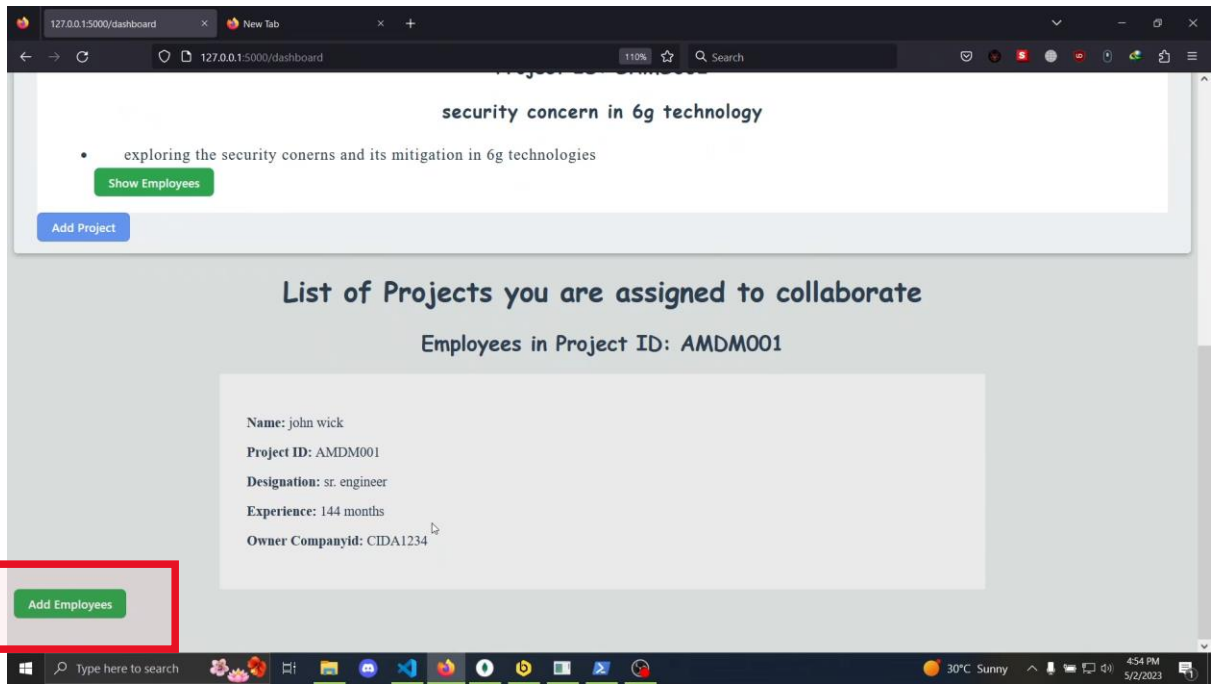
Step 12: Login to the collaborative company Dashboard. We can see the collaborated project appears in the collaborated projects

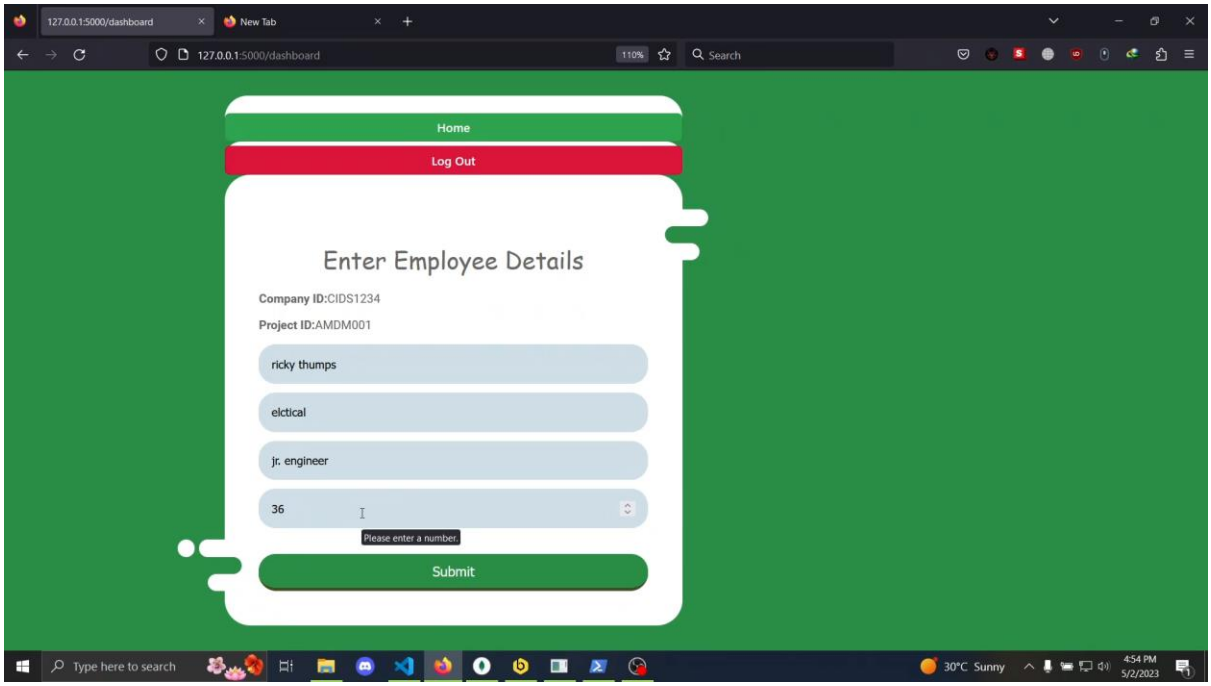


** form here we can also see the Employee of the parent company associated with the projet



Step 13: Collaborative company can also assigns their employee to the project by clicking on “Add Employee” button





** and the result will be reflected to the parent company