

DAY-4(HIVE)

1. Write a Python program that uses the HiveQL language to create a table named "Employees" with columns for "id," "name," and "salary."
2. Create a Python program that retrieves records from a Hive table named "Customers" where the age is greater than 30.
3. Write a Python script that sorts records in descending order based on the "timestamp" column in a Hive table named "Logs."
4. Write a Python program that connects to a Hive server using PyHive library and retrieves all records from a table named "Products".
5. Write a Python script that calculates the average salary of employees from a Hive table named "Employees".
6. Implement a Python program that uses Hive partitioning to create a partitioned table named "Sales_Data" based on the "year" and "month" columns.
7. Develop a Python script that adds a new column named "email" of type string to an existing Hive table named "Employees."
8. Create a Python program that performs an inner join between two Hive tables, "Orders" and "Customers," based on a common column.
9. Implement a Python program that uses the Hive SerDe library to process JSON data stored in a Hive table named "User_Activity_Logs."

Submission Guidelines:

- Answer all the questions in a single Jupyter Notebook file (.ipynb).
- Include necessary code, comments, and explanations to support your answers and implementation.
- Ensure the notebook runs without errors and is well-organized.
- Create a GitHub repository to host your assignment files.
- Rename the Jupyter Notebook file using the format "date_month_topic.ipynb" (e.g., "12_July_Hive.ipynb").

- Place the Jupyter Notebook file in the repository.
- Commit and push any additional files or resources required to run your code (if applicable) to the repository.
- Ensure the repository is publicly accessible.
- Submit the link to your GitHub repository as the assignment submission.

Grading Criteria:

1. Understanding and completeness of answers: 40%
2. Clarity and depth of explanations: 25%
3. Correct implementation and evaluation of optimizer techniques: 15%
4. Analysis and comparison of different optimizers: 10%
5. Proper code implementation and organization: 10%

Note:- Create your assignment in Jupyter notebook and upload it to GitHub & share that uploaded assignment file link through your dashboard. Make sure the repository is public.