***Experiments 9*: Use CDH (Cloudera Distribution for Hadoop) and HUE (Hadoop User Interface) to analyze data and generate reports for sample data sets**.

**AIM:** Use CDH (Cloudera Distribution for Hadoop) and HUE (Hadoop User Interface) to analyze the data and generate reports for sample data sets

Steps and Source Code:

**1.Start Cloudera services:**

sudo service cloudera-scm-server start

sudo service cloudera-scm-agent star

**Check the status:**

sudo service --status-all | grep cloudera

**2. Access HUE Web Interface**

Open your browser and navigate to:

http://localhost:8888

Login using HUE credentials:

Username: cloudera

Password: cloudera

**3. Upload Sample Dataset to HDFS**

Example Dataset: Employee Data (employees.csv)

id,name,department,salary

1,John,IT,70000

2,Alice,HR,60000

3,Bob,IT,75000

4,Charlie,Finance,80000

5,David,HR,62000

6,Eva,IT,72000

7,Frank,Finance,81000

8,Grace,HR,65000

**Upload employees.csv to HDFS using HUE**

**1. Navigate to HUE → File Browser → HDFS**

**2. Create a new directory** /user/cloudera/data

3. Click **Upload** → Select employees.csv → Upload it

Or through terminal

hdfs dfs -mkdir -p /user/cloudera/data

hdfs dfs -put employees.csv /user/cloudera/data/

**Create a Hive Table in HUE**

**Open HUE → Click Query Editors → Select Hive**

**4. Create a Hive table for the dataset:**

CREATE DATABASE IF NOT EXISTS company;

USE company;

CREATE TABLE employees (

id INT,

name STRING,

department STRING,

salary INT

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

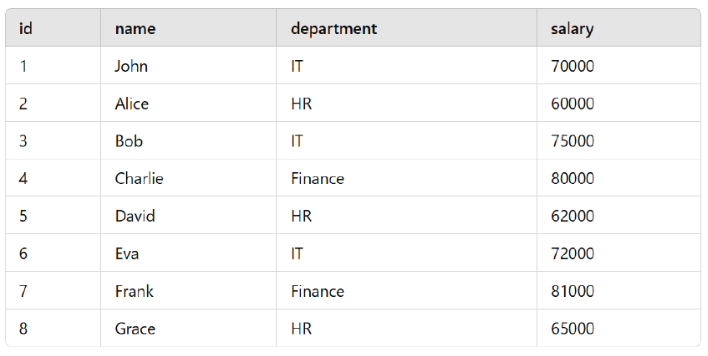
STORED AS TEXTFILE;

**Load data into the Hive table:**

LOAD DATA INPATH '/user/cloudera/data/employees.csv' INTO TABLE employees;

**Verify the data:**

SELECT \* FROM employees;



**Data Analysis Using Hive Queries in HUE**

**Find the Highest Salary in Each Department**

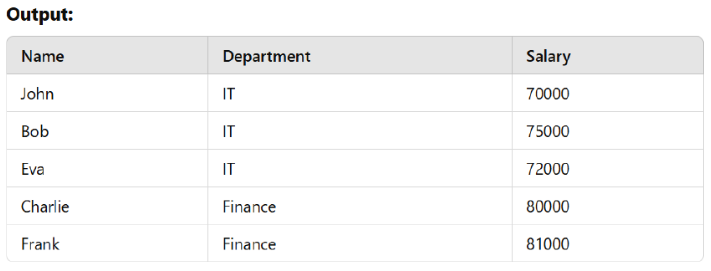
SELECT department, MAX(salary) AS highest\_salary

FROM employees

GROUP BY department;



**Get Employees with Salary Greater than 65000;**



**Generate Reports in HUE**

Step 1: Export Query Results

1. Run any of the above SQL queries in HUE Query Editor

2. Click Export → Choose format (CSV, Excel, JSON)

3. Download the report

Step 2: Create HUE Dashboard for Visualization

1. Open HUE → Click Dashboard

2. Click Create New Dashboard

3. Click Add Widget → Select Chart Type (Bar Chart, Pie Chart, etc.)

Enter Query → Example for Employee Count:

SELECT department, COUNT(\*) AS employee\_count

FROM employees

GROUP BY department;

**Click "Run Query" →** The visualization will be generated

