**Setup**

**Create schema**

Access to folder containing *lab6\_create\_schema.sql*, then run hive command: *$ hive -f lab6\_create\_schema.sql*

*Text

Description automatically generated*

Copy ChicagoEmployeesDataset.csv file to HDFS directory: */hive/input/ChicagoEmployeesDataset*

*Graphical user interface, text, application, email

Description automatically generated*

**Queries**

From HUE, run these 3 queries against view *employees\_vw*:

--Find average annual salary of every department

SELECT department as Deparment, ceiling(AVG(annual\_salary)) as Average\_annual\_salary FROM employees\_vw WHERE salary\_or\_hour = 'Salary'

GROUP BY department;

**Graphical user interface, text, application

Description automatically generated**

--Find min income per week of job titles for each department

SELECT job\_title, department, min(hourly\_rate \* typical\_hours) as Min\_income\_per\_week FROM employees\_vw WHERE salary\_or\_hour = 'Hourly'

GROUP BY job\_title, department

ORDER BY job\_title ASC;

Graphical user interface, text, application

Description automatically generated

--Find the hihgest-income person (annual salary)

SELECT A.name as Name, B.Salary as Salary, A.job\_title as Job\_title, A.department as Department

FROM employees\_vw A ,(SELECT MAX(annual\_salary) as Salary FROM employees\_vw WHERE salary\_or\_hour = 'Salary') B

WHERE A.annual\_salary = B.Salary;

A screenshot of a computer

Description automatically generated