**Put the given dataset (ChicagoEmployeesDataset.csv) in HDFS.**

Graphical user interface, text, email

Description automatically generated

**Create EXTERNAL table in Hive for this given sample dataset and find out some interesting facts from this data.**

CREATE EXTERNAL TABLE employees (  
 name STRING,  
 job\_title STRING,  
 department STRING,  
 full\_or\_part\_time STRING,  
 salary\_or\_hour STRING,  
 typical\_hours STRING,  
 annual\_salary STRING,  
 hourly\_rate STRING  
)  
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'  
WITH SERDEPROPERTIES("quoteChar" = "\"")  
LOCATION '/user/cloudera/cs523/input'  
TBLPROPERTIES("skip.header.line.count" = "1");

**Execute script:**

Text

Description automatically generated

**Hive editor:**

Graphical user interface, table

Description automatically generated

**Case data type Employee**

CREATE VIEW employees\_vw  
AS SELECT  
 name,  
 job\_title,  
 department,  
 CAST(full\_or\_part\_time as char(1)),  
 CAST(salary\_or\_hour as varchar(10)),  
 CAST(typical\_hours as smallint),  
 CAST(annual\_salary as double),  
 CAST(hourly\_rate as float)  
FROM employees;

Describe tables:

A picture containing chart

Description automatically generated

Text

Description automatically generated

**Show top 5 maximum salaries of employees in each department:**

SELECT department, MAX(annual\_salary) max\_salary FROM employees\_vw GROUP BY department ORDER BY max\_salary DESC LIMIT 5;

Graphical user interface, table

Description automatically generated with medium confidence

**Show the average annual\_salary of ‘POLICE’ department:**

SELECT department, AVG(annual\_salary) avg\_salary FROM employees\_vw WHERE department = 'POLICE' GROUP BY department;

Graphical user interface, application

Description automatically generated

**Show top 20 employees part-time highest hourly\_rate:**

SELECT name, job\_title, department, hourly\_rate FROM employees\_vw WHERE hourly\_rate IS NOT NULL ORDER BY hourly\_rate DESC LIMIT 20;

Graphical user interface, application

Description automatically generated

**HIVE WITH AVRO**

**Put the given dataset (SampleWeatherForHive.txt) in HDFS.**

**Graphical user interface, table

Description automatically generated**

**Create table records:**

CREATE EXTERNAL TABLE records (year STRING, temperature INT, quality INT)  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n'  
LOCATION '/user/cloudera/cs523/input/avro';

**Create table as\_avro:**

CREATE TABLE as\_avro (year STRING, temperature INT, quality INT)  
STORED AS AVRO;

**Describe tables:**

**Text

Description automatically generated**

**Insert data into as\_avro:**

**INSERT INTO as\_avro SELECT \* FROM records;**

**Graphical user interface, application

Description automatically generated**