Hive Create table

Create external table with 6 columns:

CREATE EXTERNAL TABLE PERMA\_Receiver(status STRING, title STRING, skill STRING, wage INT, citizenship STRING, education STRING)ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ STORED AS TEXTFILE

LOCATION '/user/yt1579/PERMA\_Receiver'TBLPROPERTIES (“skip.herder.line.count”=”1”);

Copy the data:

hadoop fs -copyFromLocal /home/yt1579/perm\_output.csv /user/yt1579/PERMA\_Receiver;

图片包含 文本

描述已自动生成

Query

Count the number of certified applicant

SELECT count(status) from PERMA\_Receiver where status == ‘Certified’as success;

文本

描述已自动生成

Count the number of bad records(denied, withdrawn)

SELECT count(status) from PERMA\_Receiver where status == ‘Certified’as success;

文本

描述已自动生成

mean wage according to each country:

SELECT citizenship, avg(wage) as avg from PERMA\_Receiver WHERE status == ‘Certified’ GROUP BY citizenship ORDER BY avg DESC;

文本

描述已自动生成

mean wage according to each job title:

SELECT title, avg(wage) as avg from PERMA\_Receiver WHERE status == ‘Certified’ GROUP BY title ORDER BY avg DESC;

图形用户界面, 文本

描述已自动生成  
  
  
mean wage according to education level:

文本

描述已自动生成