**Lab 6**

**//Create external tbl**

create external table records (Name string, JobTitles string,

Department string, ContractType string,

WageType string, TypicalHours int,

AnnualSalary int, HourlyRate float)

row format serde 'org.apache.hadoop.hive.serde2.OpenCSVSerde' with serdeproperties

(

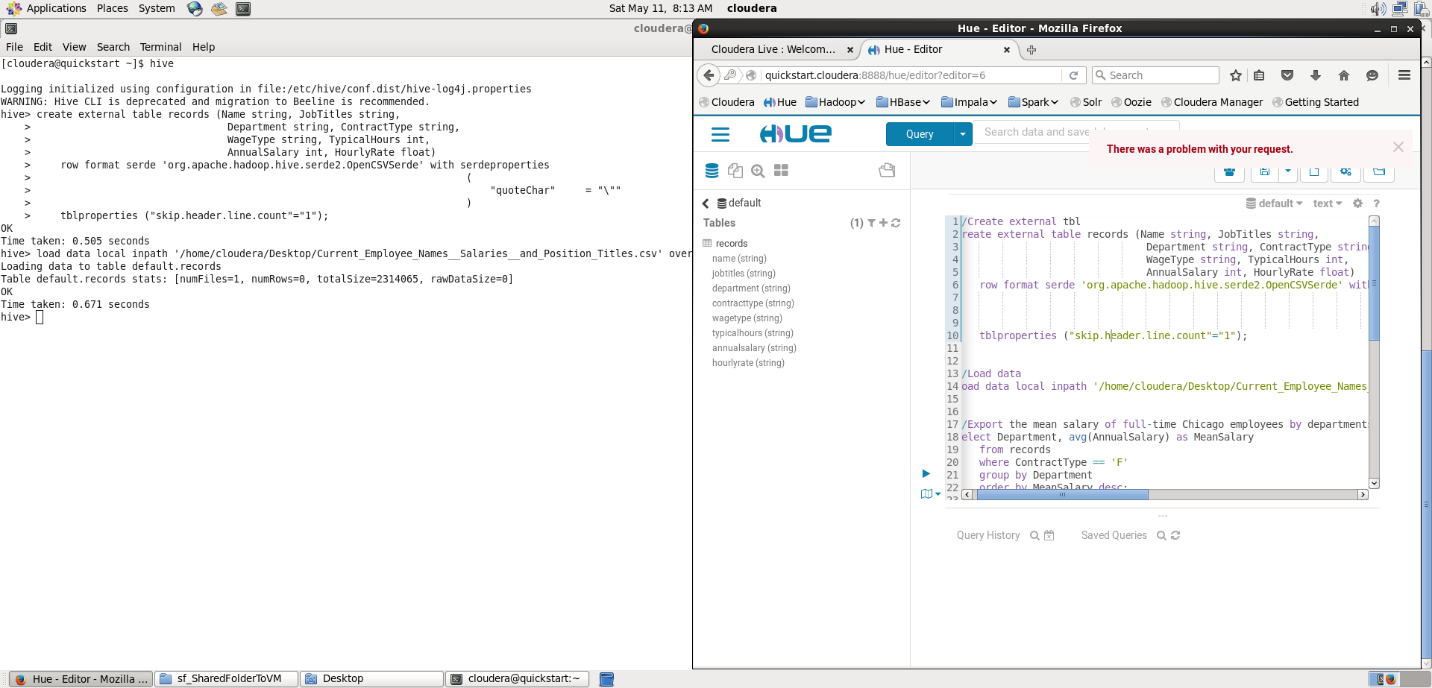
"quoteChar" = "\""

)

tblproperties ("skip.header.line.count"="1");

**//Load data**

load data local inpath '/home/cloudera/Desktop/Current\_Employee\_Names\_\_Salaries\_\_and\_Position\_Titles.csv' overwrite into table records;



**//Export the top 10 mean salary of full-time Chicago employees by departments**

select Department, avg(AnnualSalary) as MeanSalary

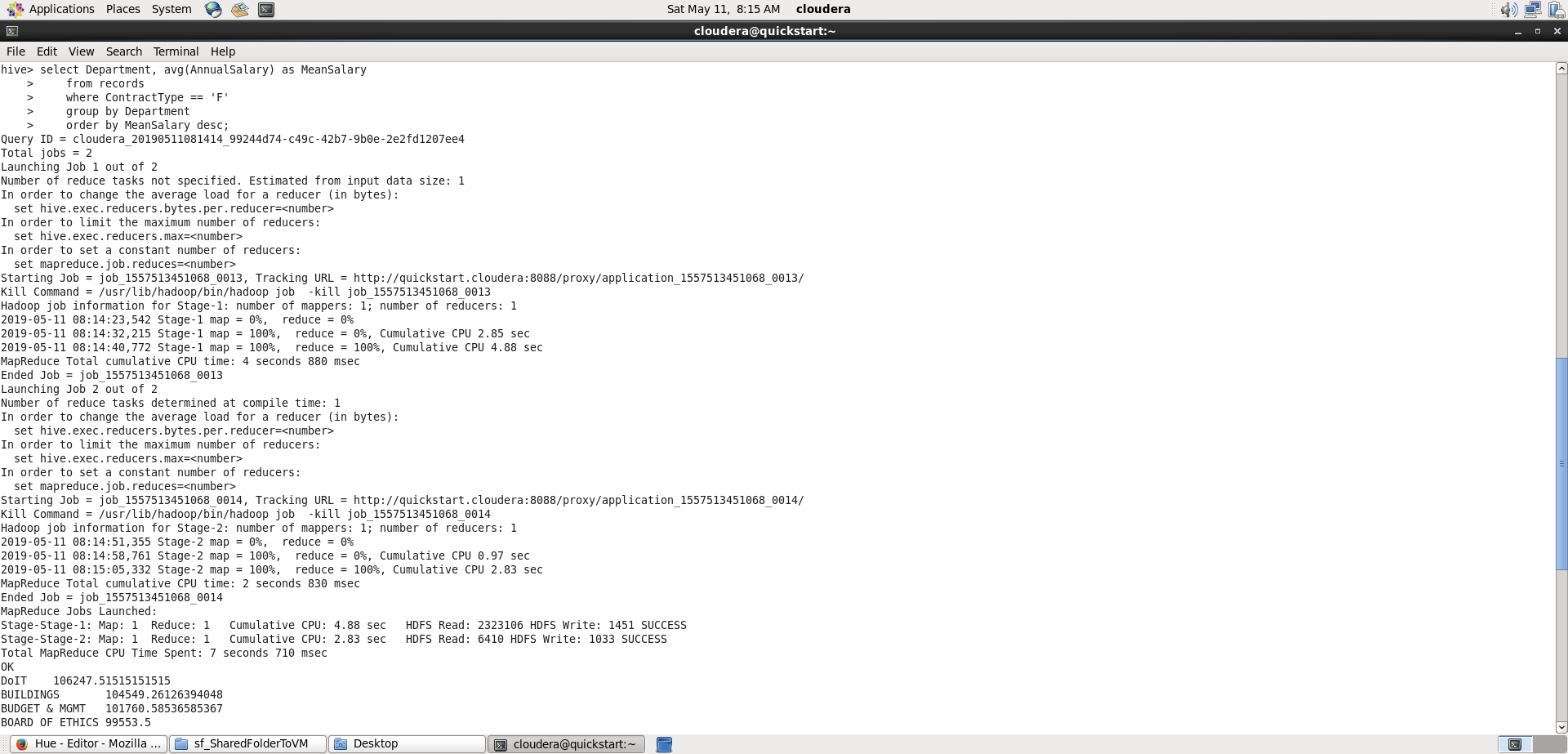
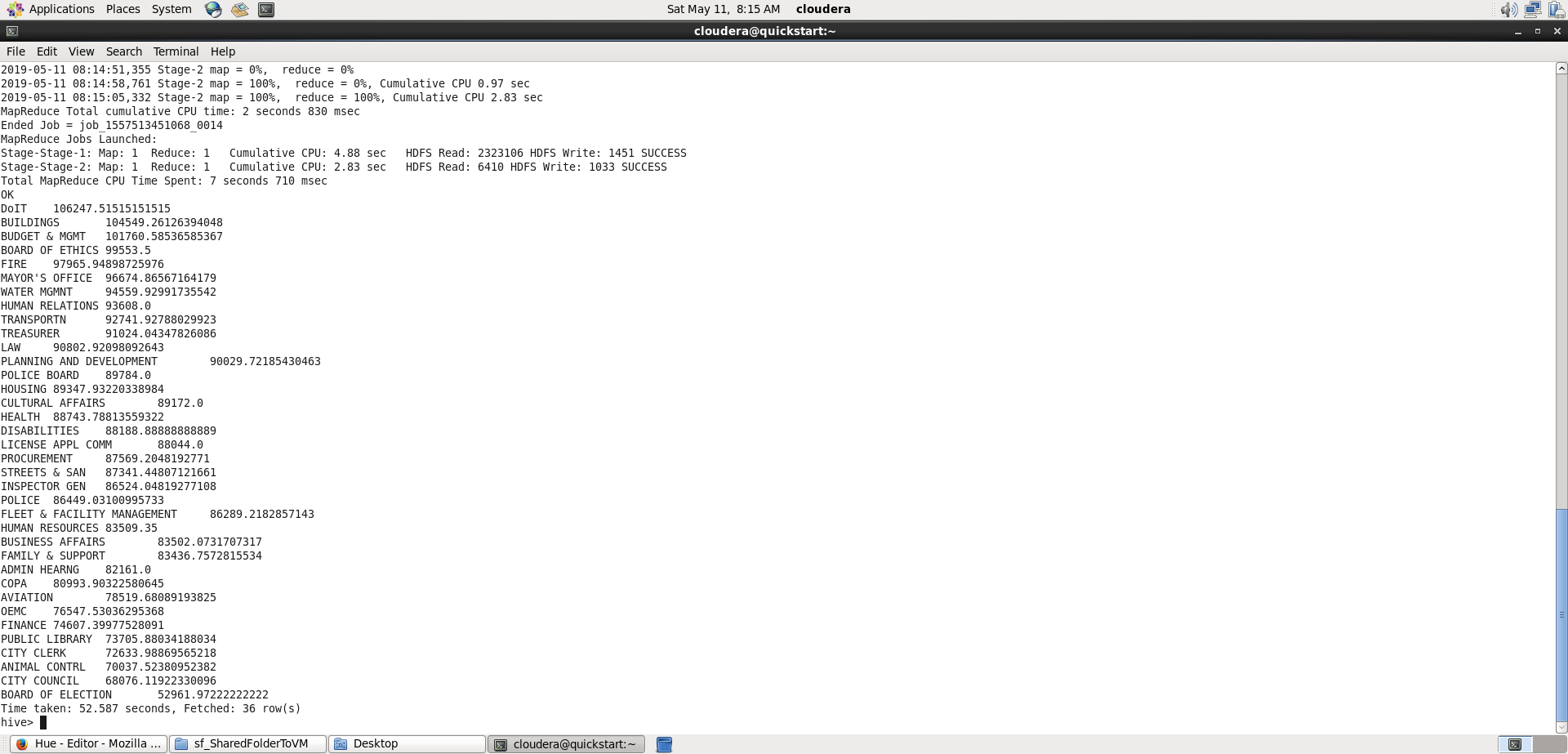
from records

where ContractType == 'F'

group by Department

order by MeanSalary desc

limit 10;



**//Export the top 10 job titles whith highest mean salary of full-time Chicago employees**

select JobTitles, avg(AnnualSalary) as MeanSalary

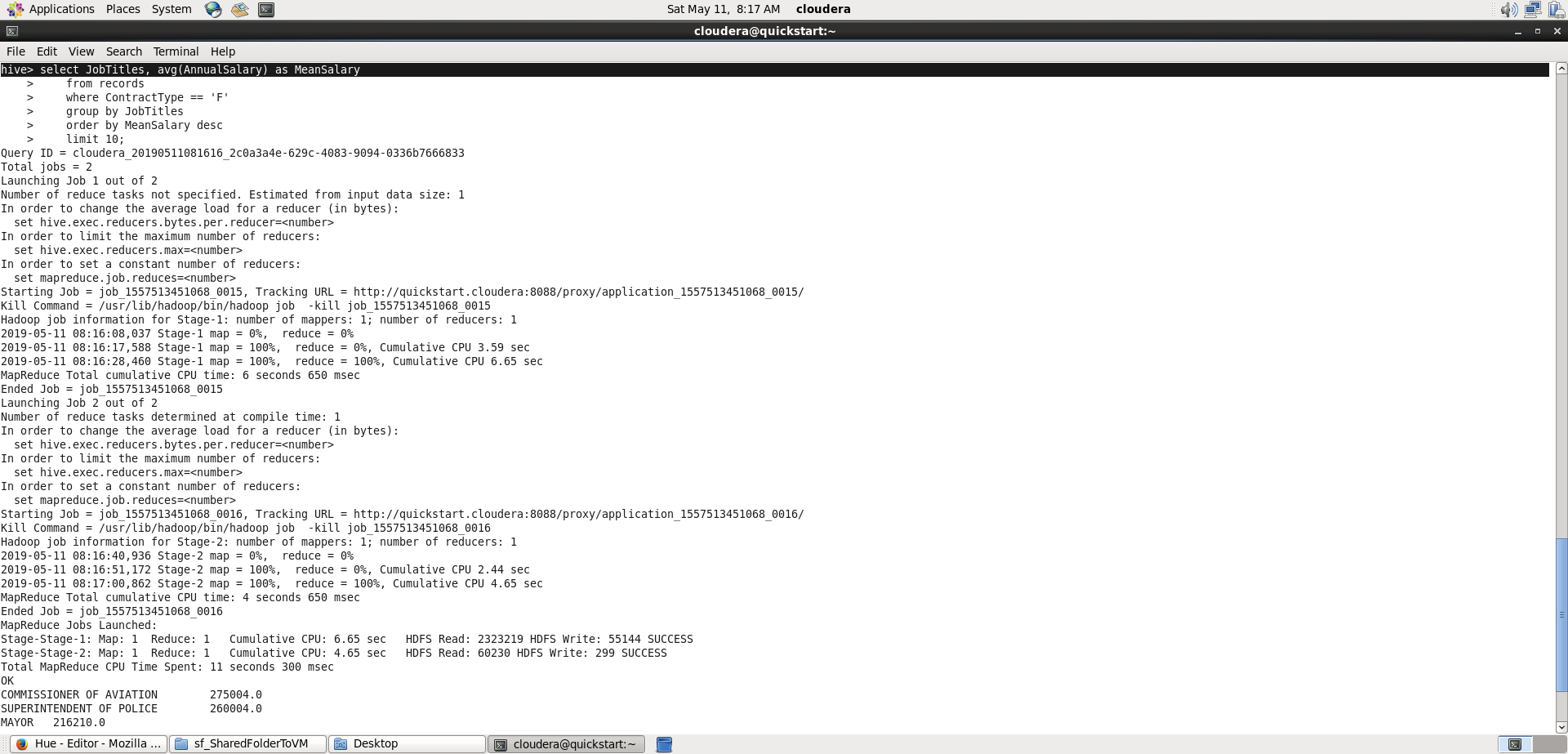
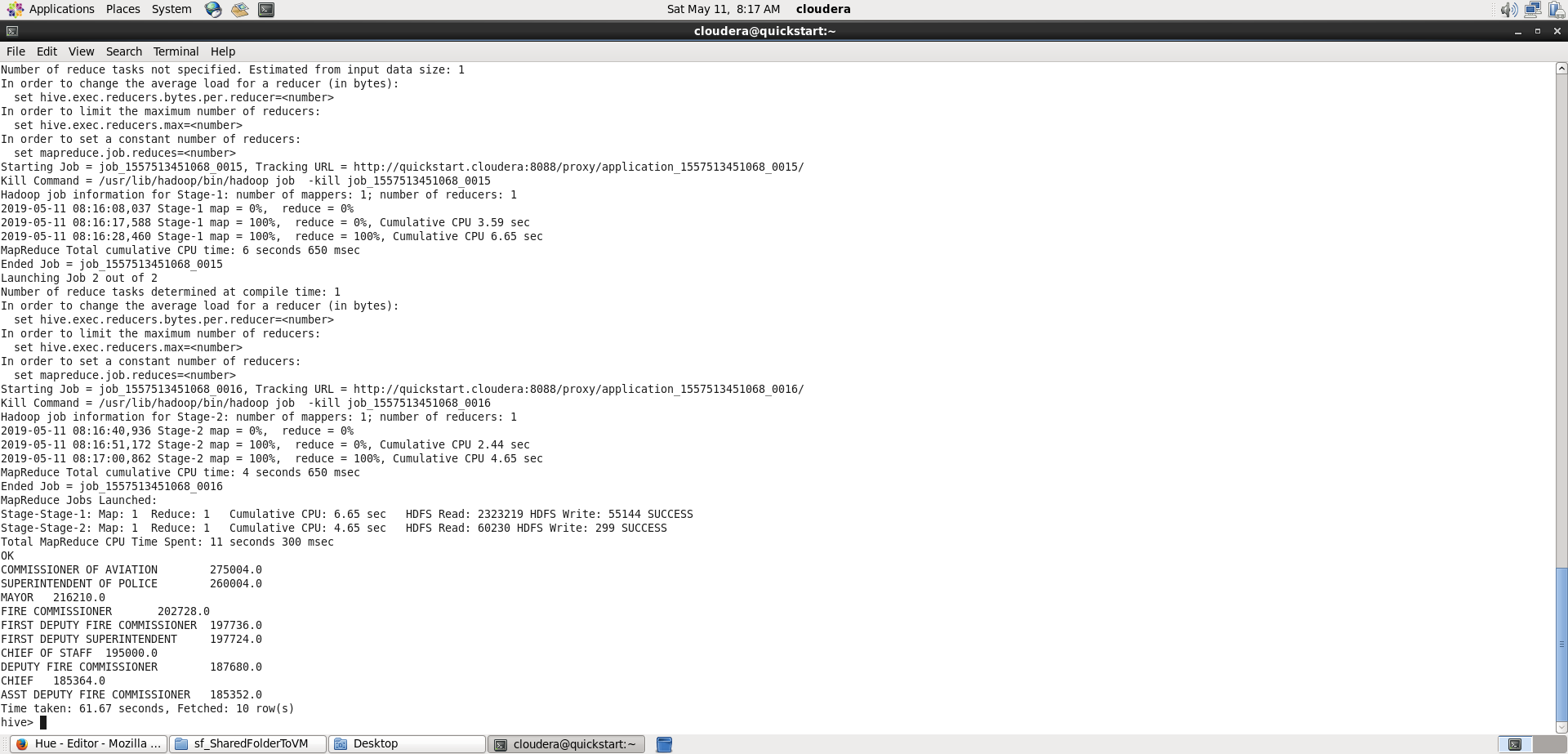
from records

where ContractType == 'F'

group by JobTitles

order by MeanSalary desc

limit 10;



**//Export the top 100 job titles which has largest workforce of full-time Chicago employees**

select JobTitles, count(\*) as count

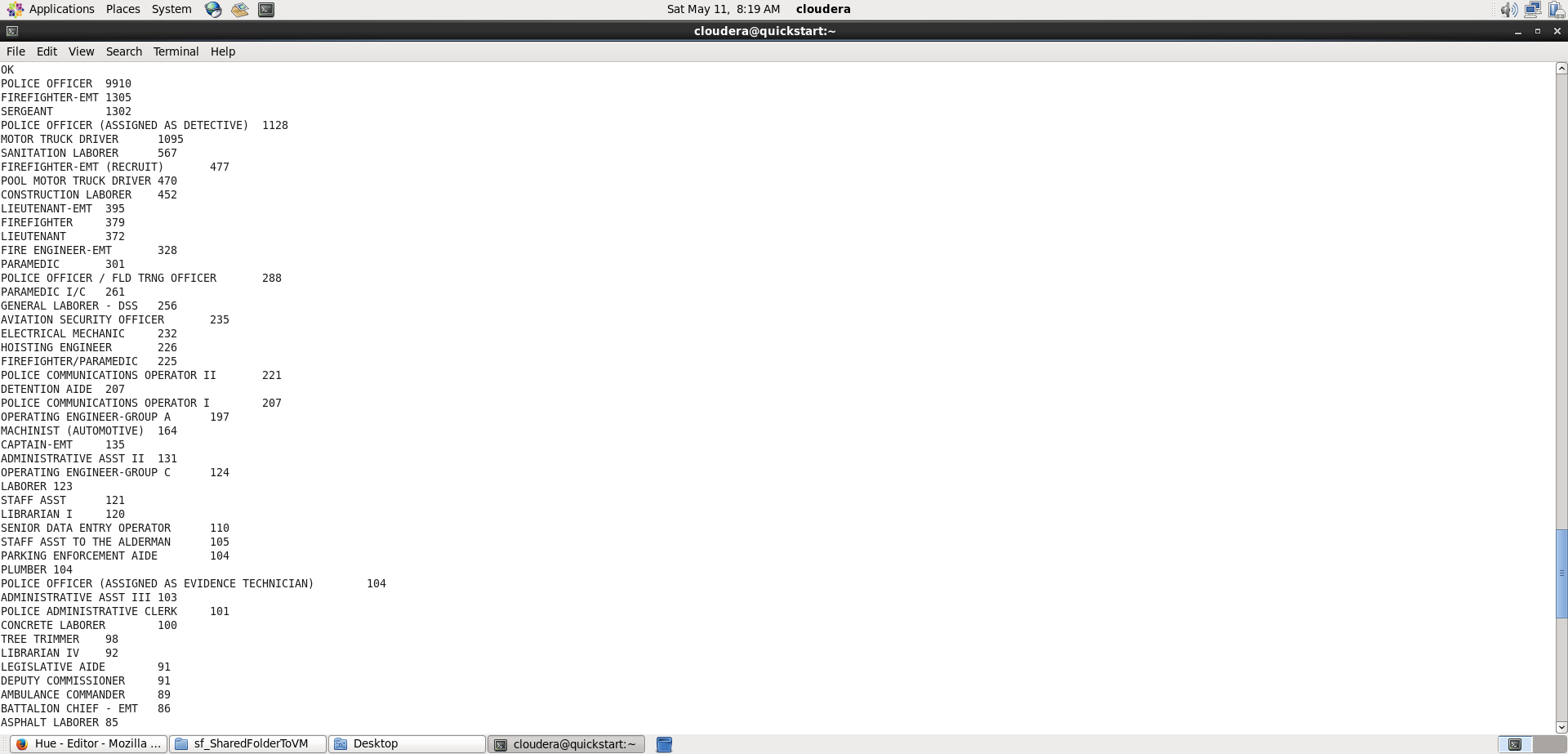
from records

where ContractType == 'F'

group by JobTitles

order by count desc

limit 100;



**//Delete tbl**

drop table records;