**SESSION 15: ADVANCED HIVE OPERATIONS**

**ASSIGNMENT 1**

Dataset: <https://github.com/hortonworks/tutorials/blob/hdp-2.5/driver_data.zip>

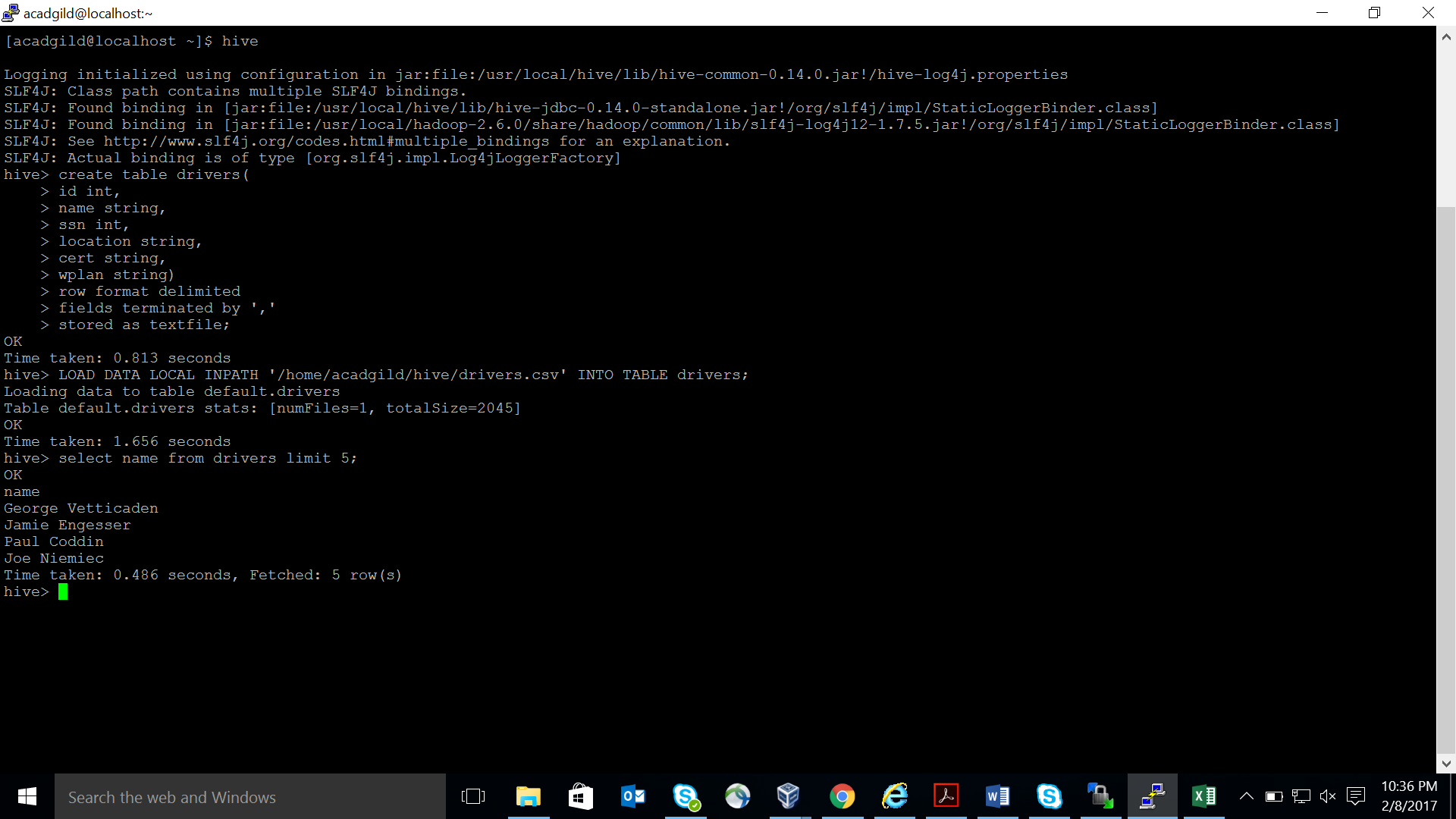
**TEXTFILE**

hive> create table drivers(id int, name string, ssn int, location string, cert string, wplan string)

row format delimited fields terminated by ‘,’ stored as textfile;

hive> LOAD DATA LOCAL INPATH ‘/home/acadgild/hive/drivers.cvs’ INTO TABLE drivers;

hive> select name from drivers limit 5;



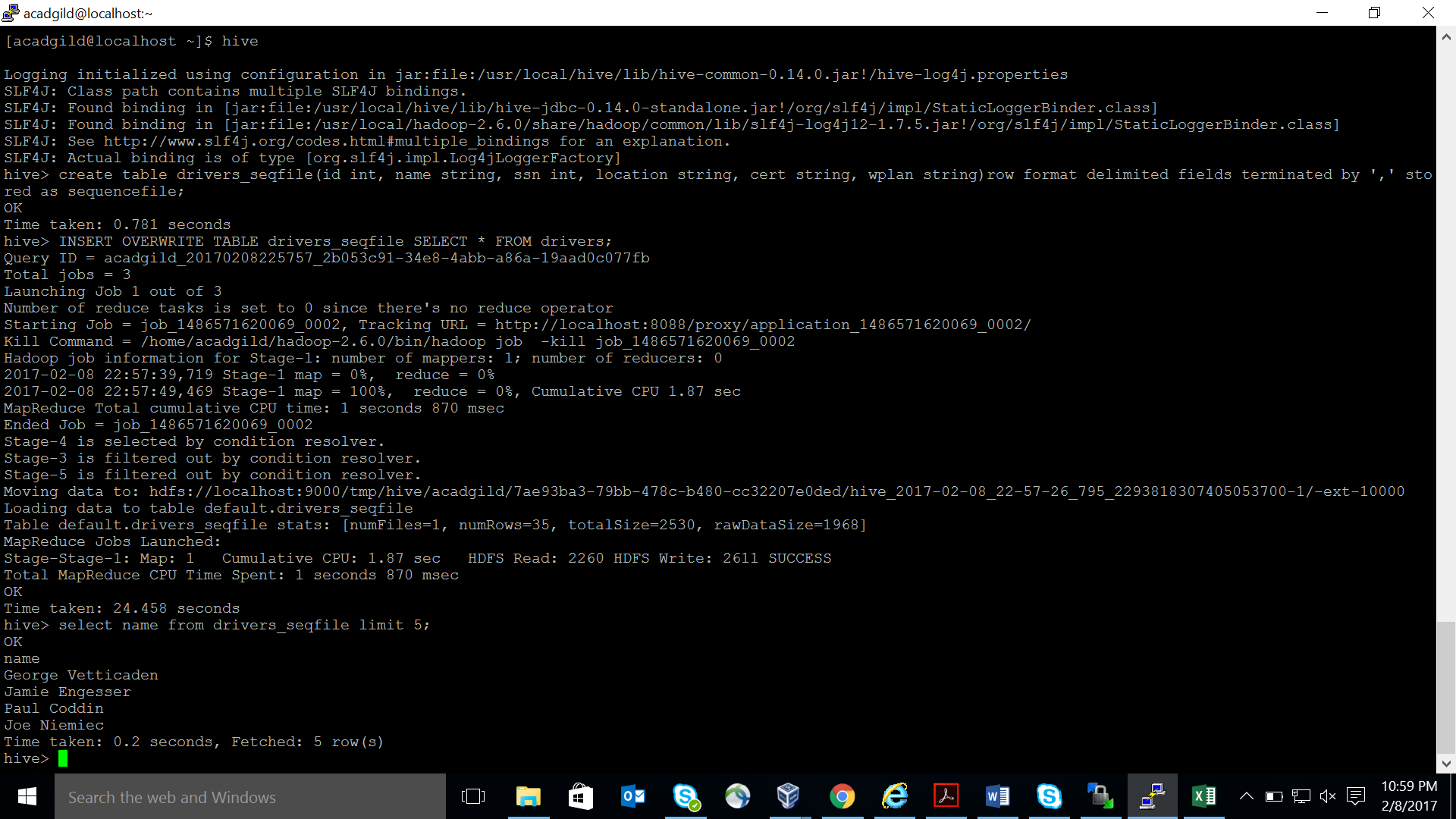
**SEQUENCEFILE**

hive> create table drivers\_seqfile(id int, name string, ssn int, location string, cert string, wplan

string) row format delimited fields terminated by ‘,’ stored as sequencefile;

hive> INSERT OVERWRITE TABLE drivers\_seqfile SELECT \* FROM drivers;

hive> select name from drivers\_seqfile limit 5;



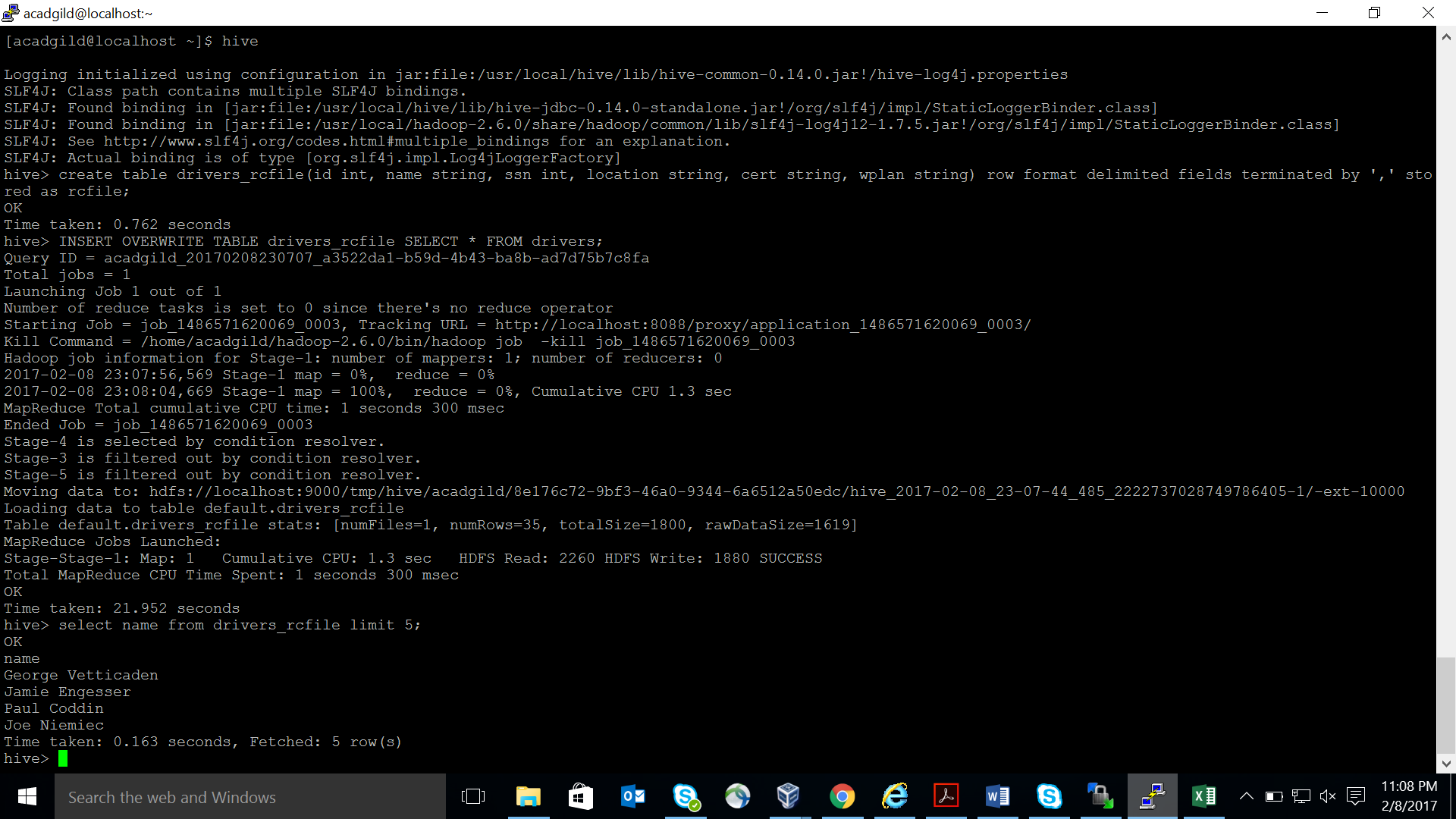
**RCFILE**

hive> create table drivers\_rcfile(id int, name string, ssn int, location string, cert string, wplan

string) row format delimited fields terminated by ‘,’ stored as rcfile;

hive> INSERT OVERWRITE TABLE drivers\_rcfile SELECT \* FROM drivers;

hive> select name from drivers\_rcfile limit 5;



**ORCFILE**

hive> create table drivers\_orcfile(id int, name string, ssn int, location string, cert string, wplan

string) row format delimited fields terminated by ‘,’ stored as orcfile;

hive> INSERT OVERWRITE TABLE drivers\_orcfile SELECT \* FROM drivers;

hive> select name from drivers\_orcfile limit 5;

