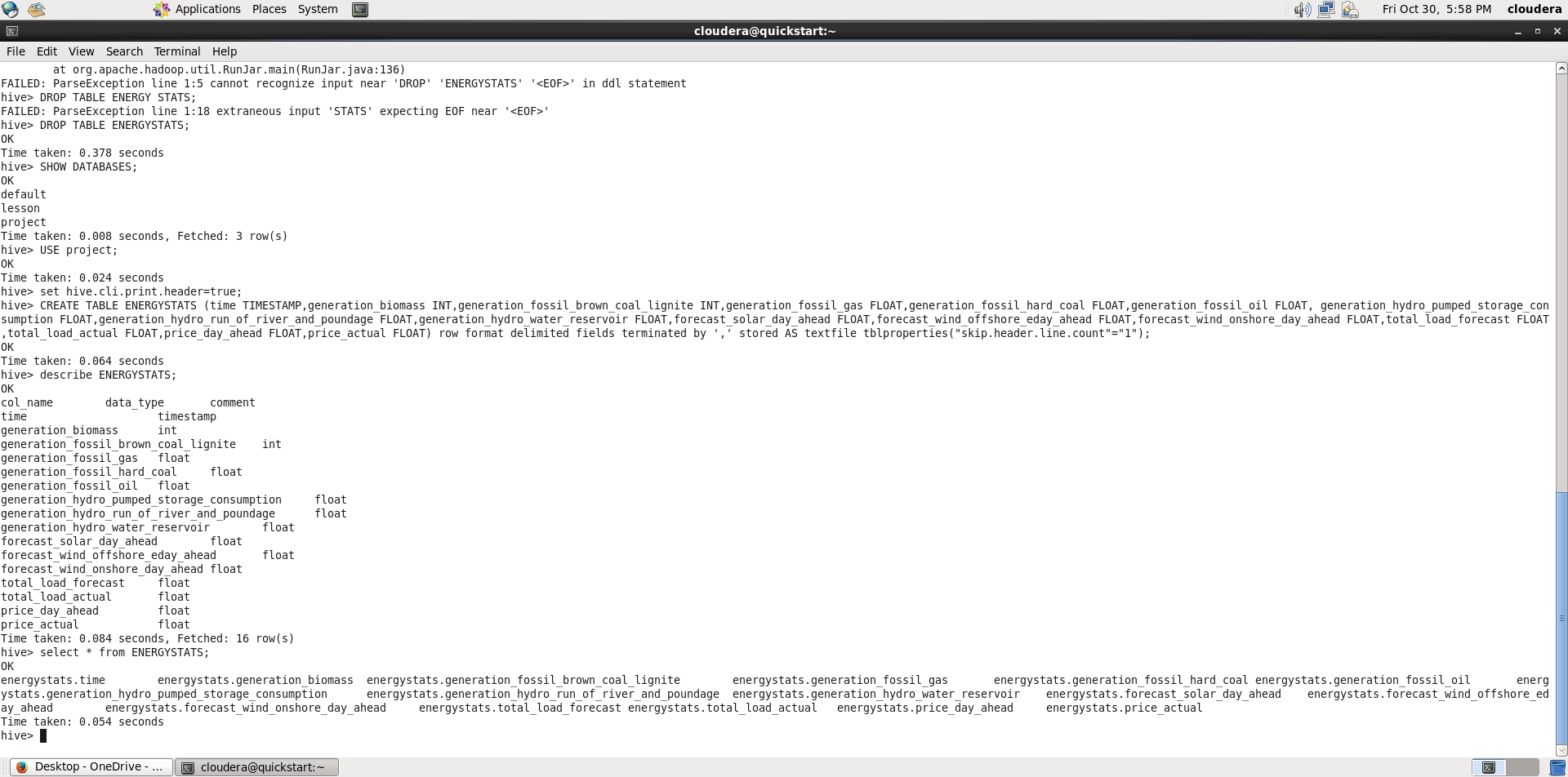
**Project increment by CLAIRE NDOFOR**

So Basically, my aim was to import our dataset into mysql and see what queries I can run to get information from this dataset. While working on this, I did realize most of the columns were hard for me to interpret what their values mean and how they were related to each other. That is one of the aspects I will have to focus on so it’s easier for me to decide on what to get out of this data.

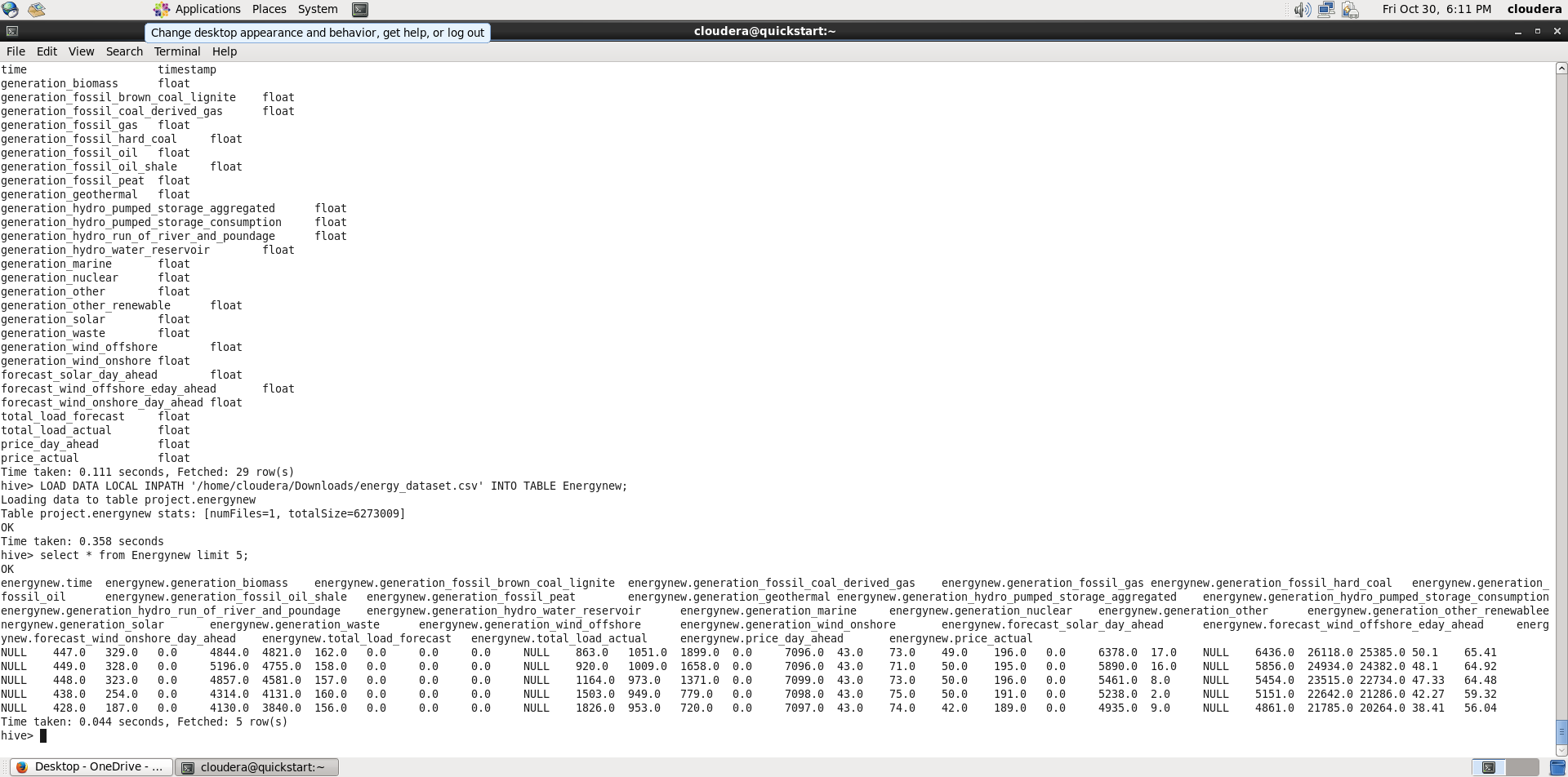
Importing dataset to mysql

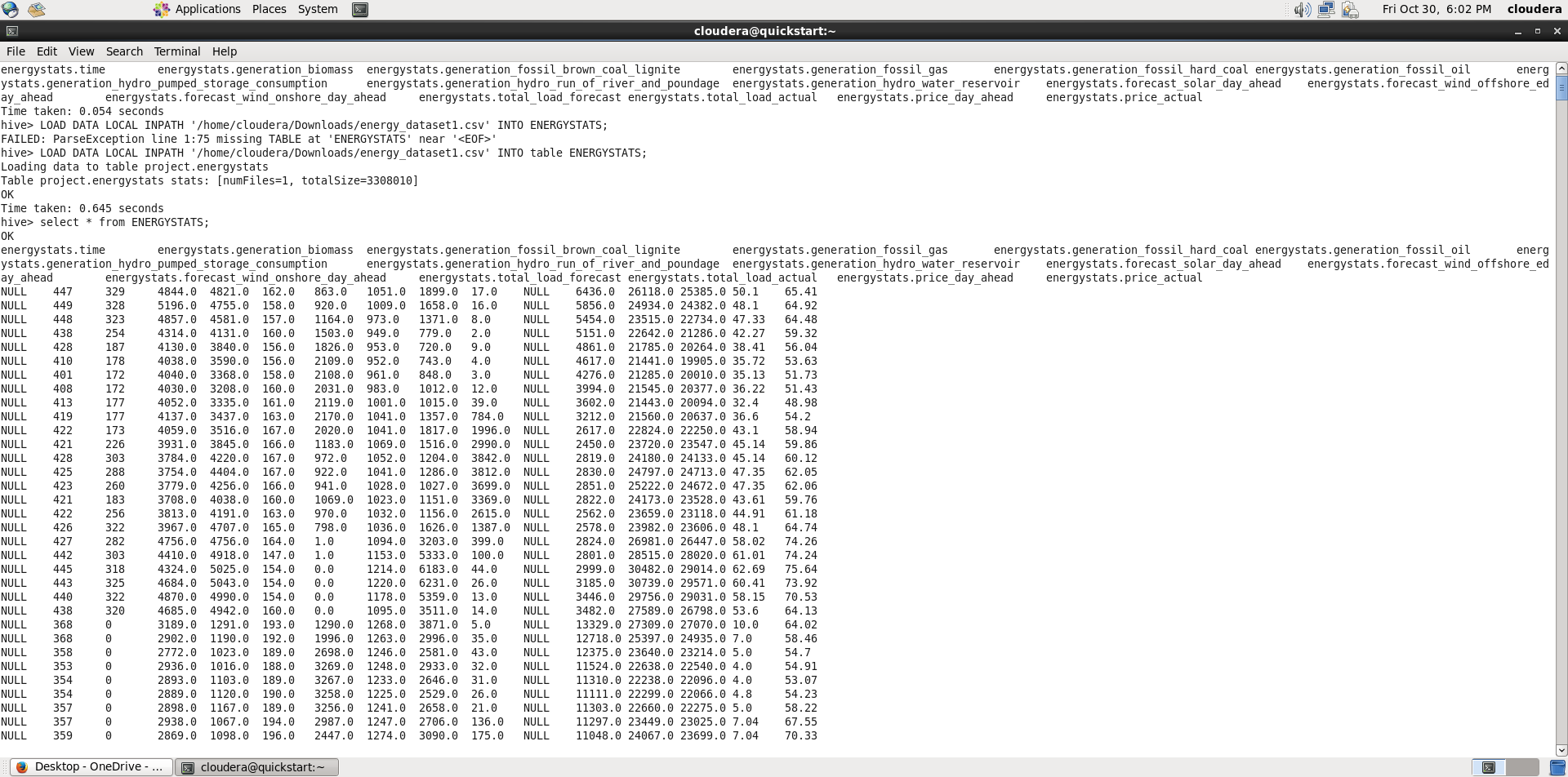
First I had to create a table in HIVE , import the dataset from hdfs to hive

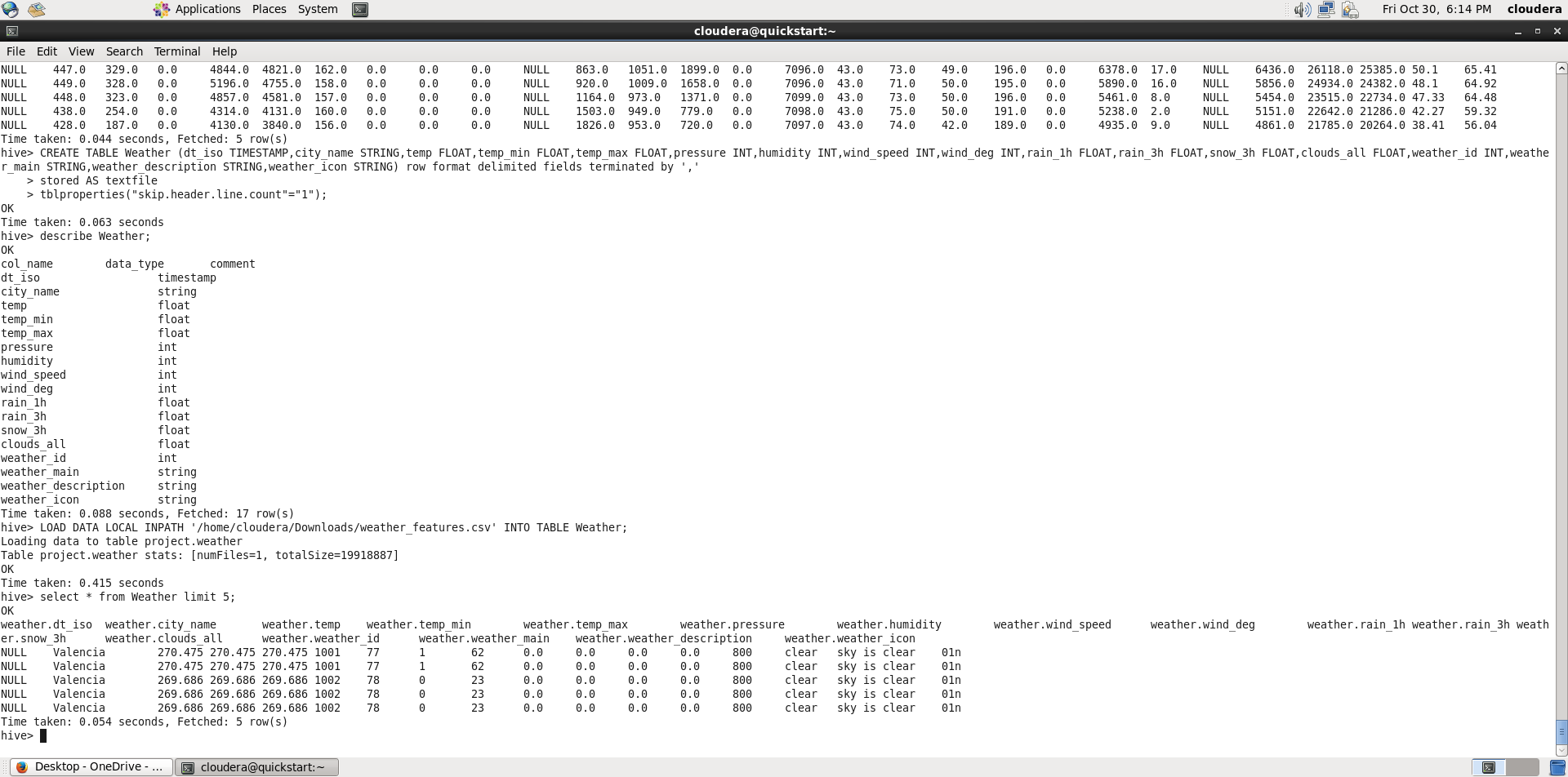
Screenshot

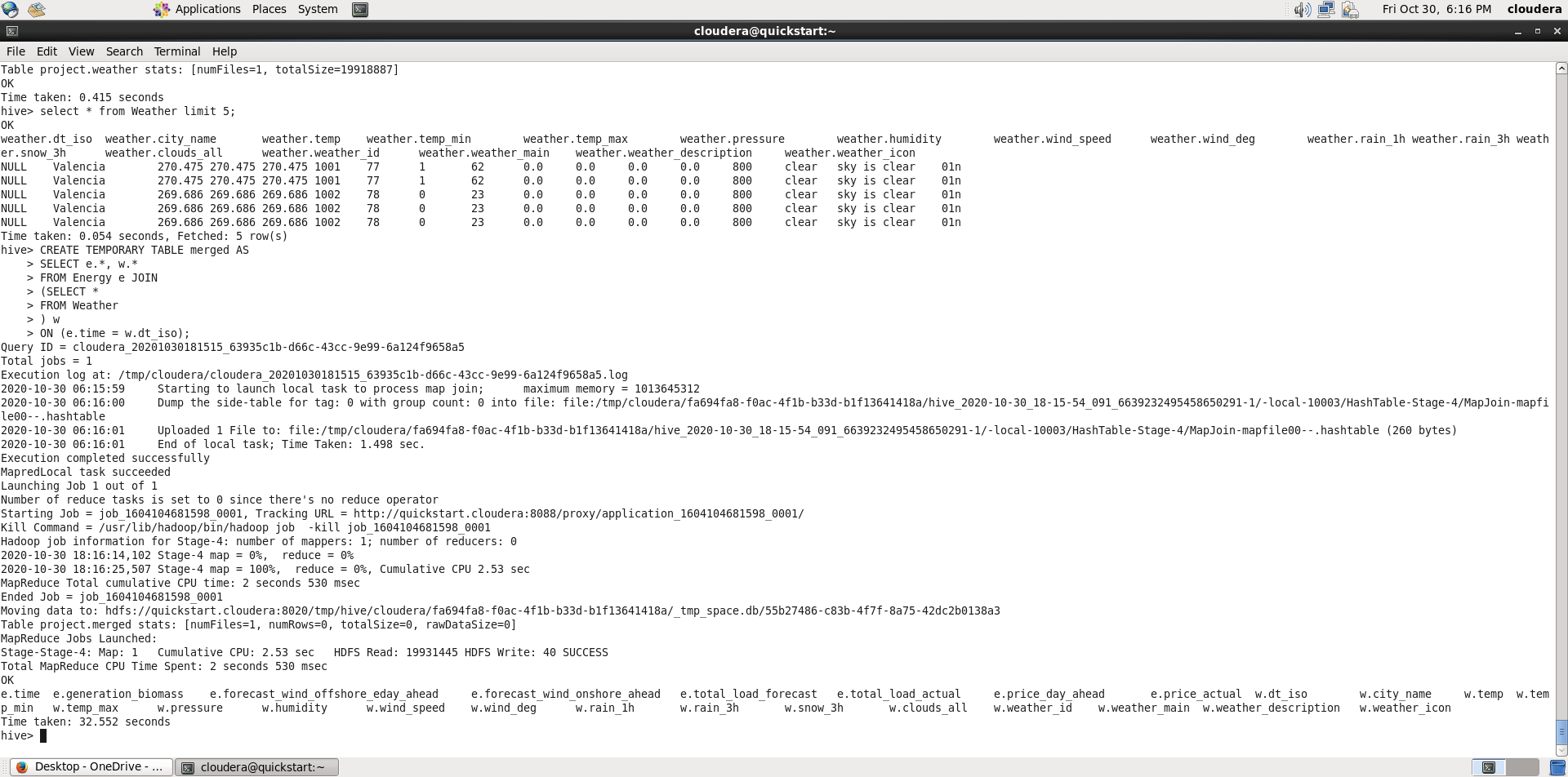






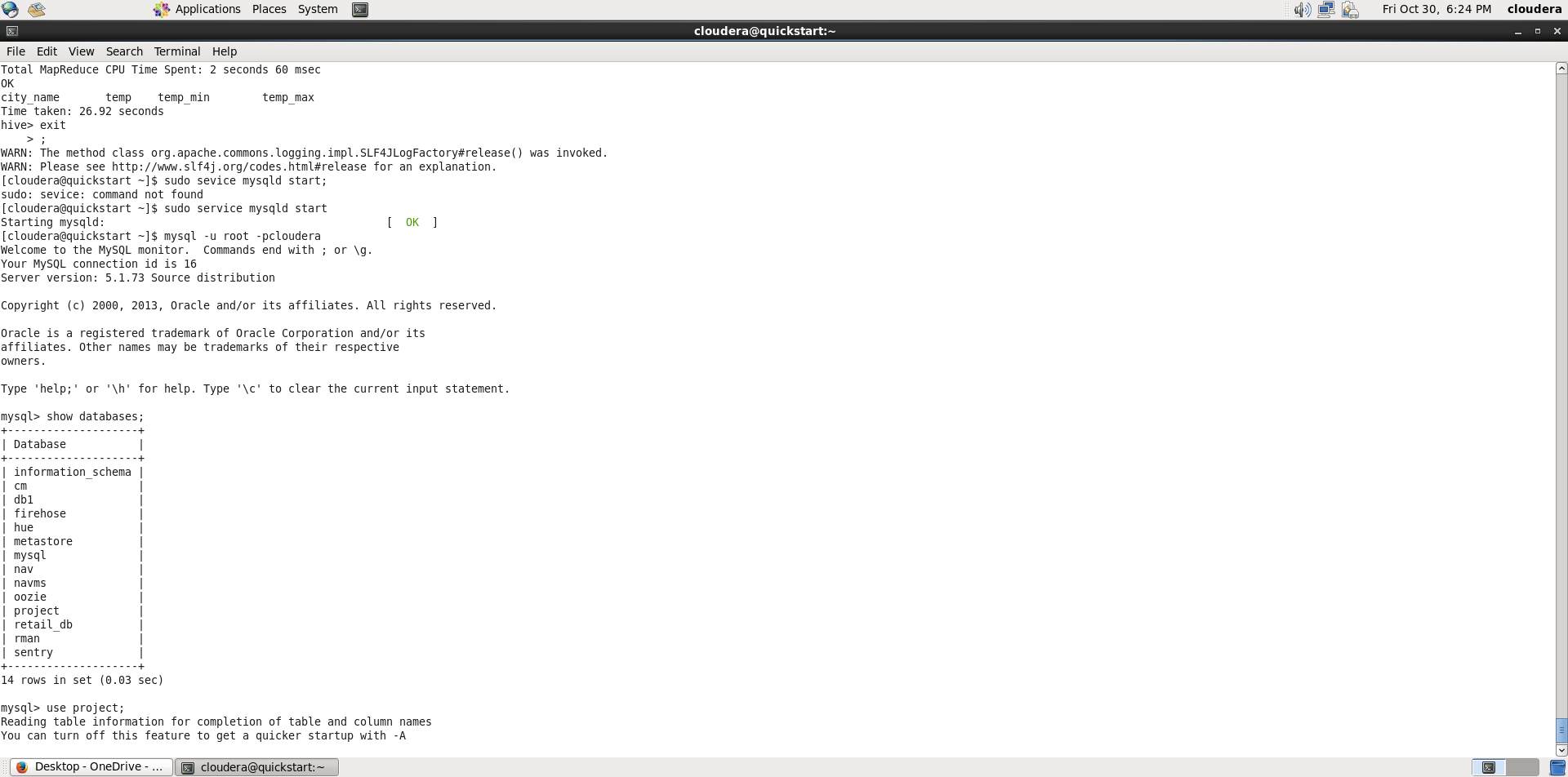


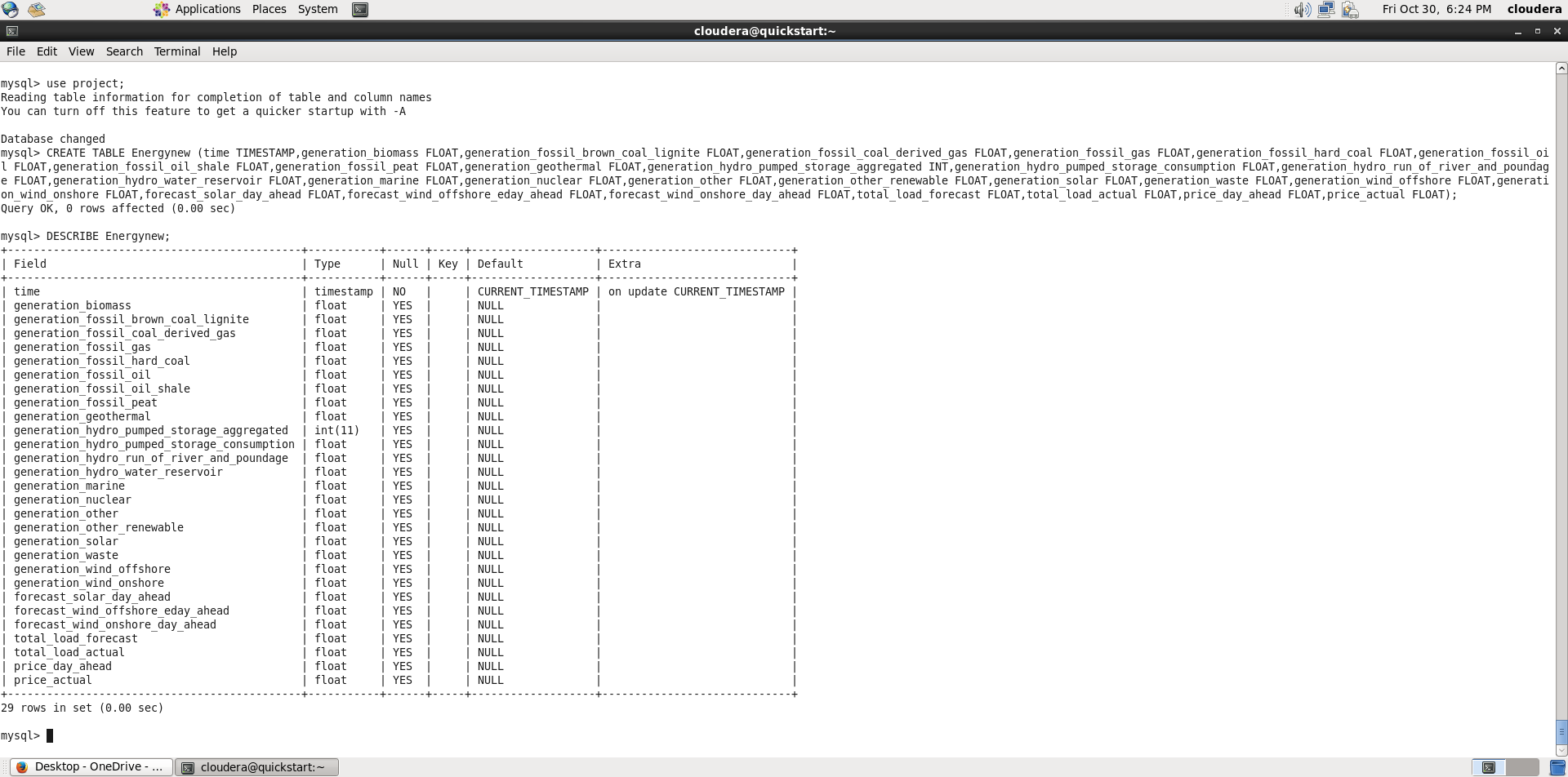


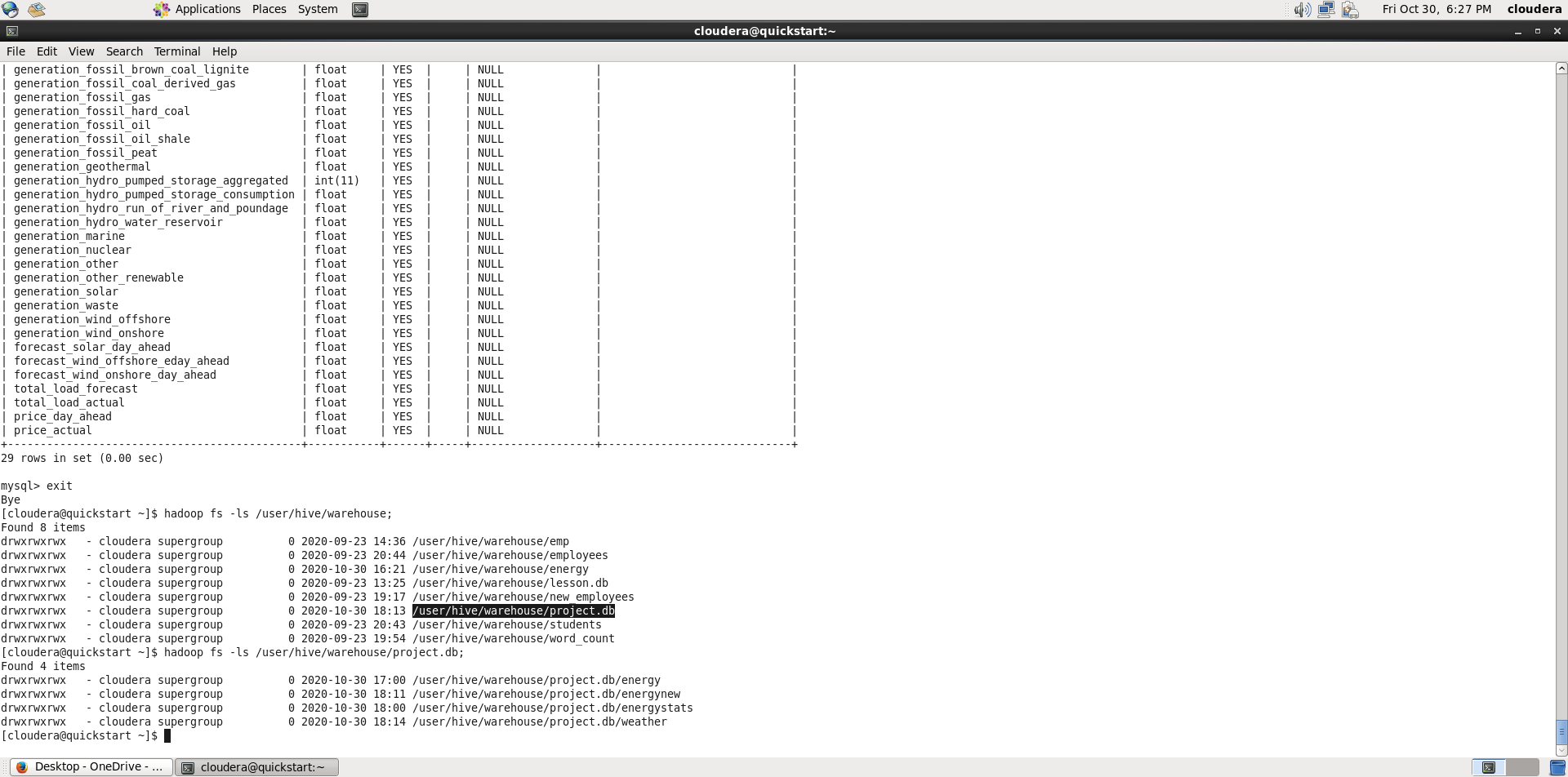


One thing I noticed was some of my columns returned NULL values and I tried fixing this by making sure ROW DELIMITERS was set but this didn’t help at all.

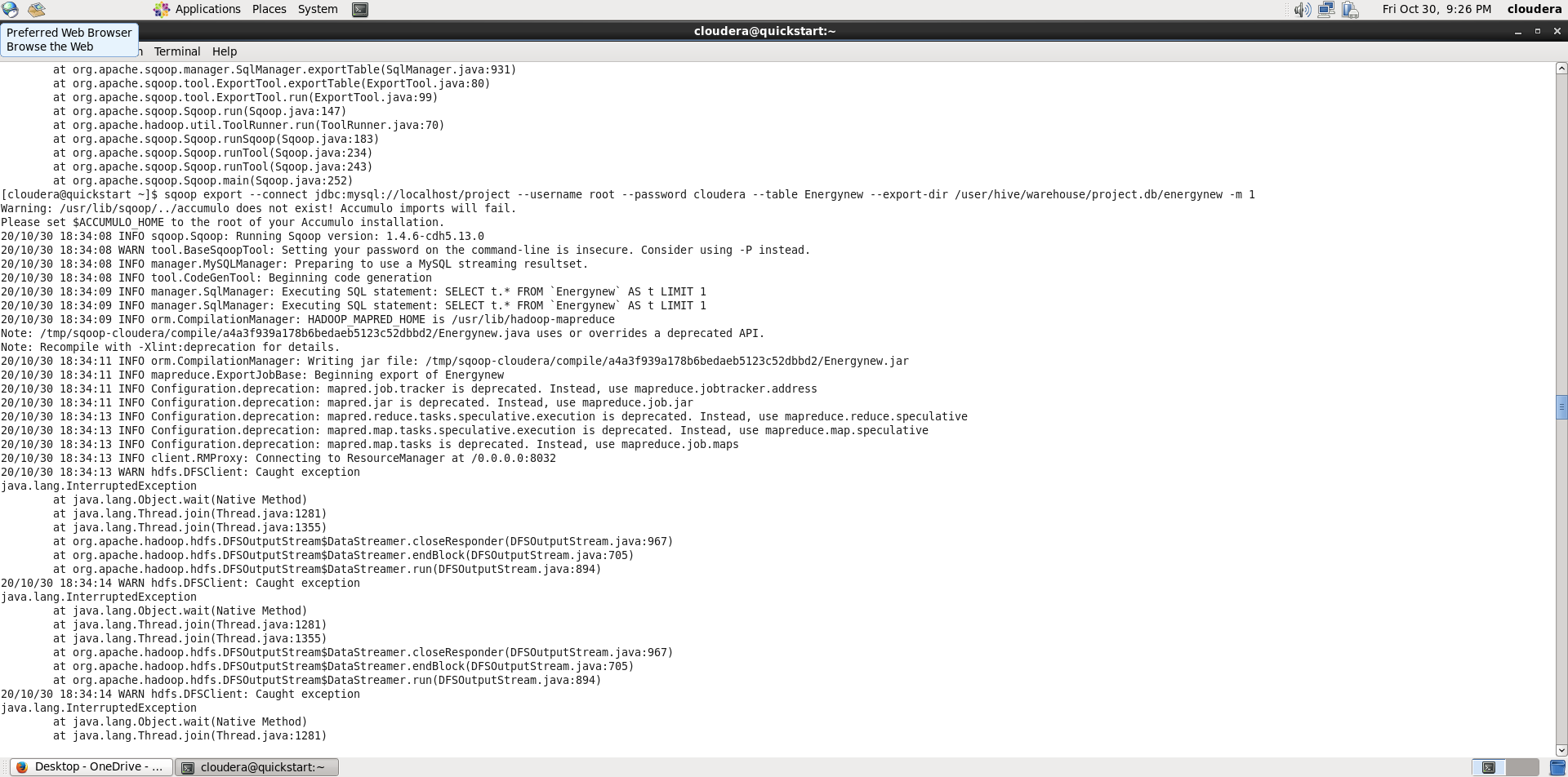
Secondly I had to create a table with similar columns in mysql so I can export data from hive to mysql via sqoop



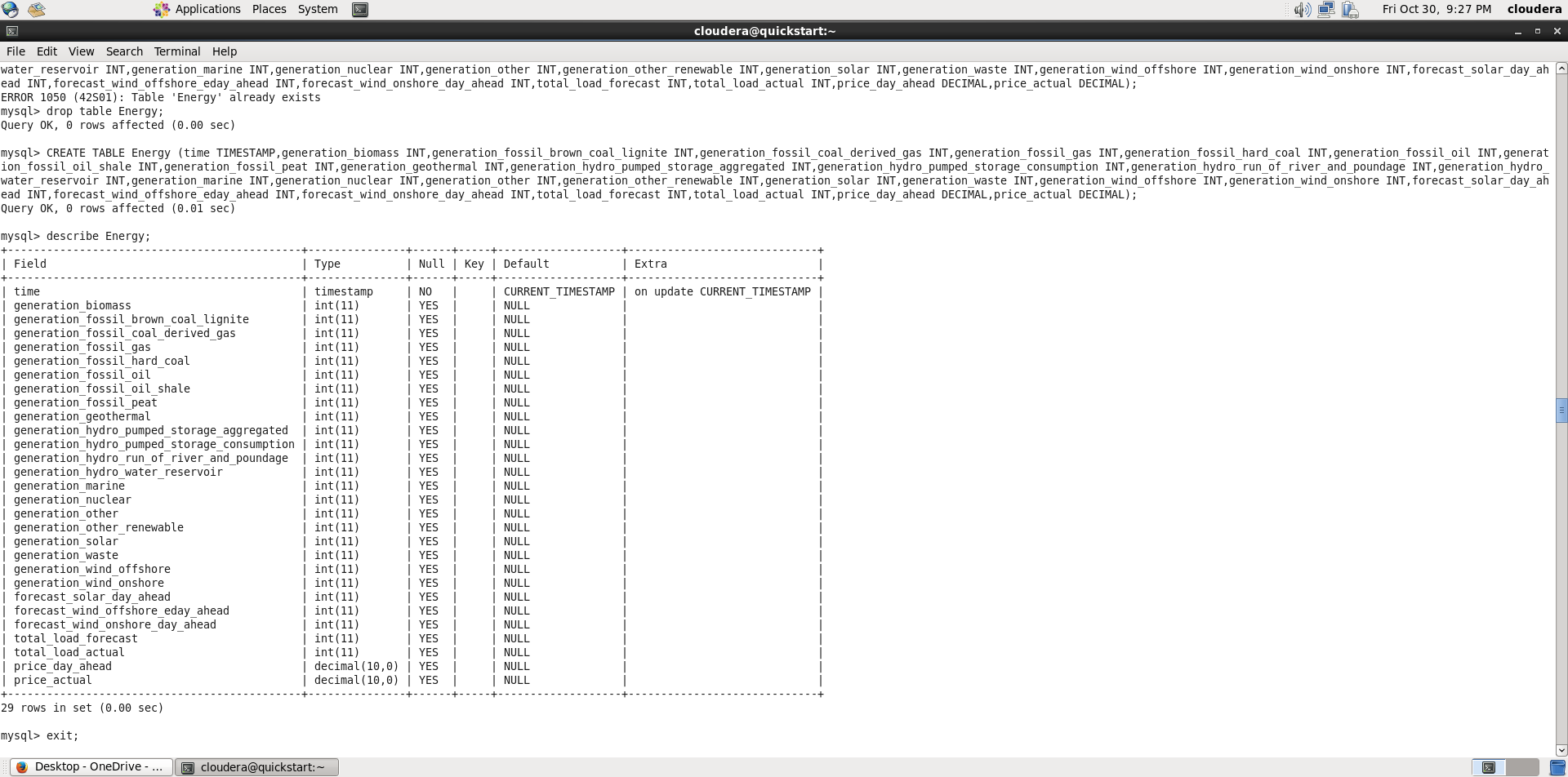




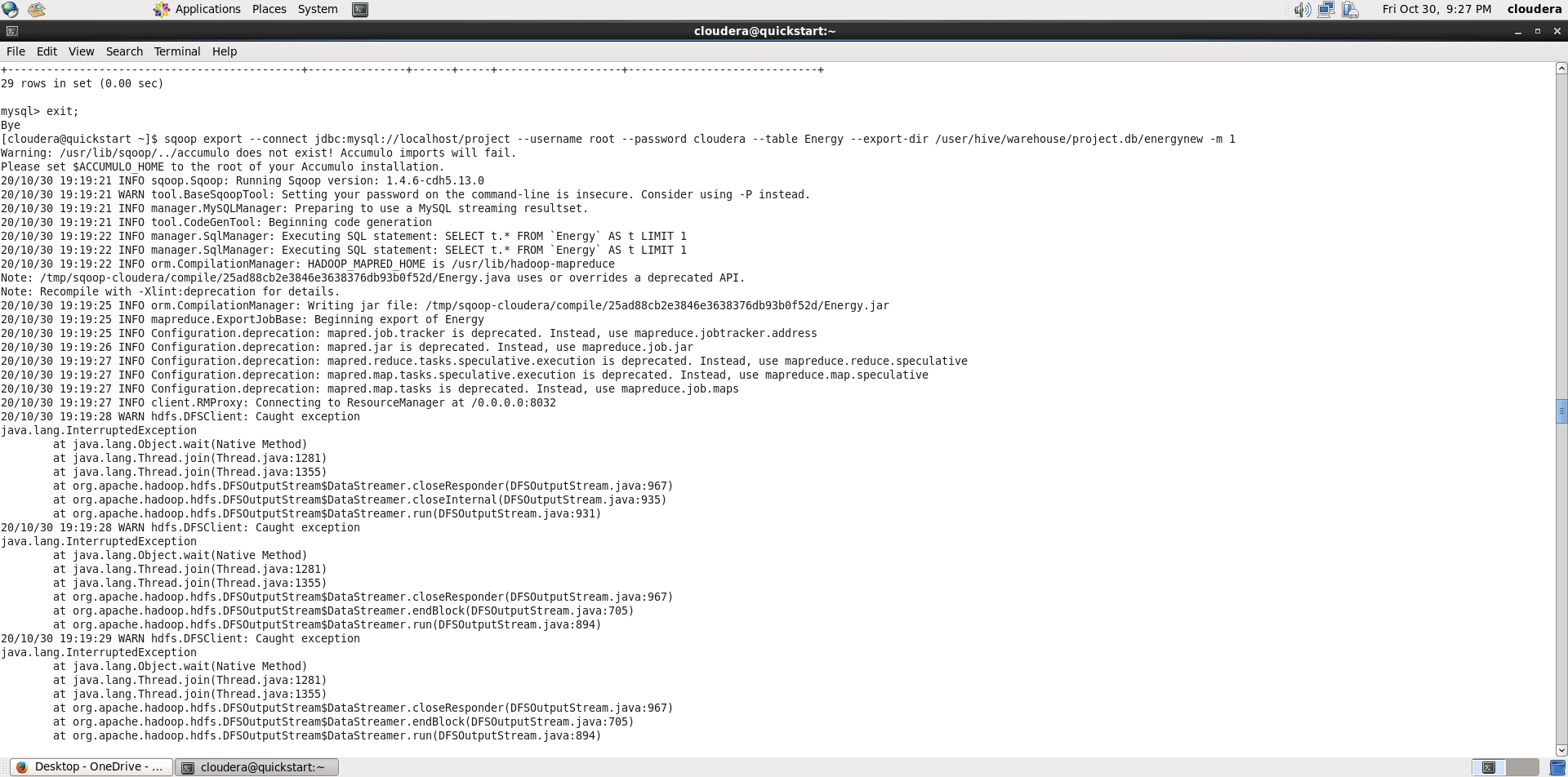
I succeeded to create a table in mysql which matched that in hive, Now the next step is to import the data from hive into mysql so I can run some queries



Unfortunately this process kept failing, so I had to recreate my tables and made sure I used the correct datatypes



That didn’t work either



FUTURE WORK

* Find out why imported data in hive has NULL columns and fix
* Successfully move data into mysql and run queries to see relationships that exists amongst these columns