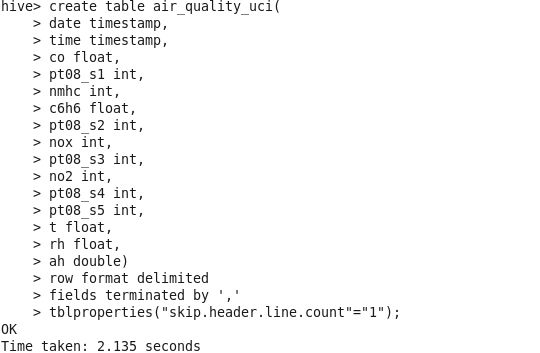
**BUILD A DATA PIPELINE WITH HIVE**

Download a **data** from the given location -

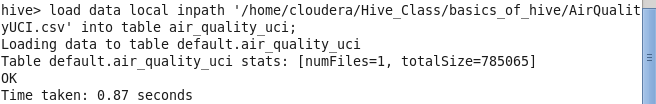
https://archive.ics.uci.edu/ml/machine-learning-databases/00360/

Note: the task screenshots are from the executions that I have done on cloudera environment inside the VM

1. Create a hive table as per given schema in your dataset



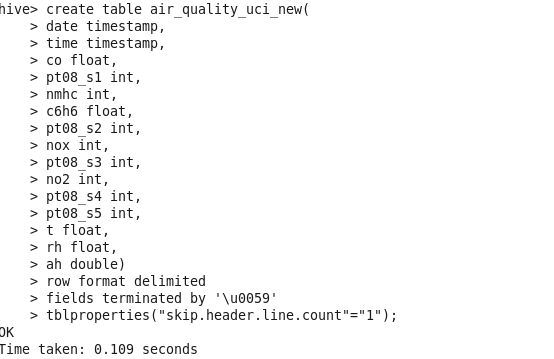
2. try to place a data into table location



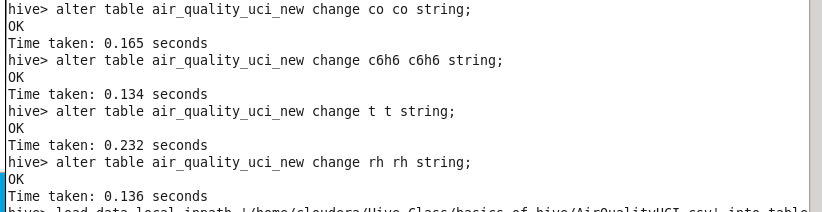
This approach doesn’t work as the field in the file are terminated by; and the decimal place are separated by,

So, we need transformation on this table

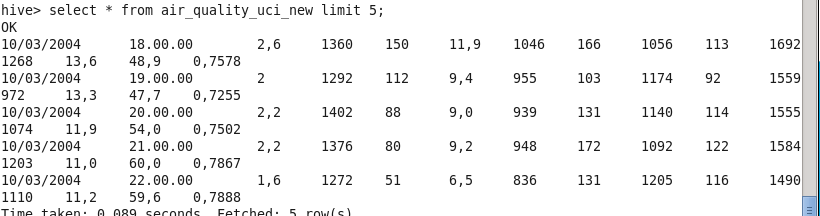
Step 1: take datatime as string in the table and decimal values as string in the table



Using alter table command to change the datatype of the columns

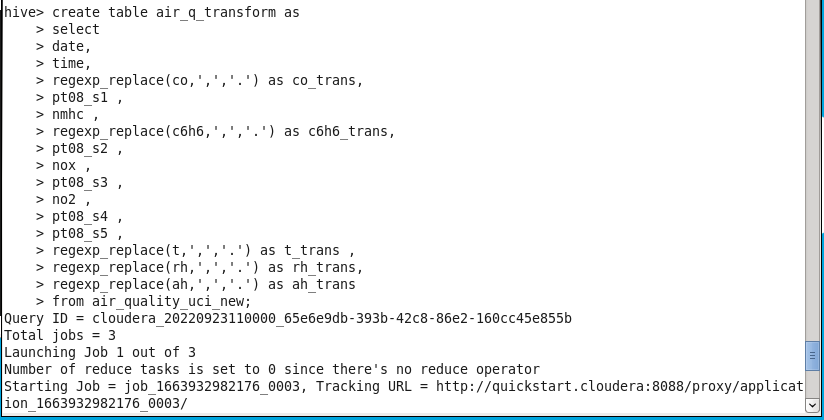


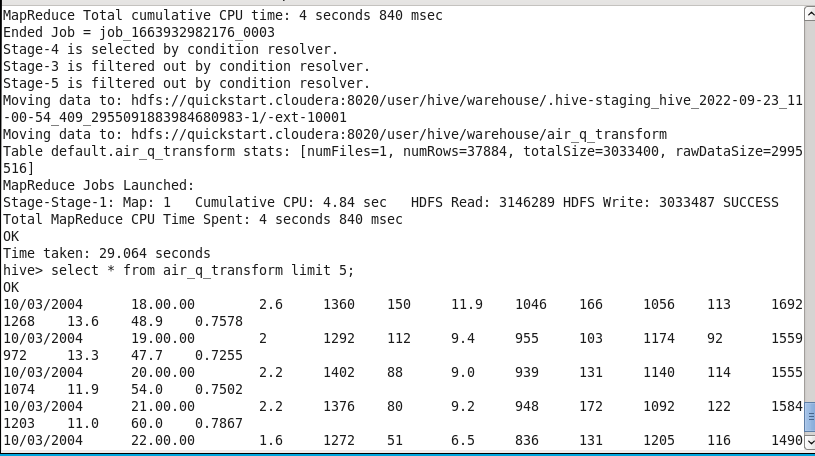
Loading Data in the table



As we can see the decimal values are stored as comma separated values ,so we need to transform this

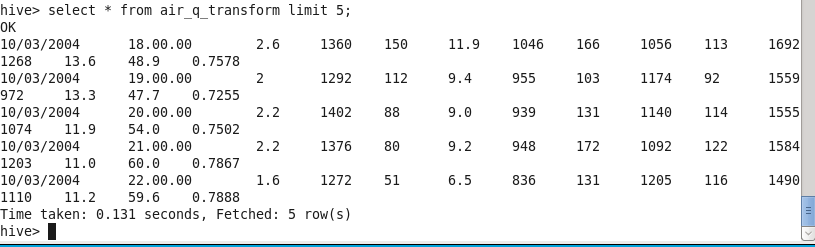
Step 2: transform the column of decimal values using regex\_replace and store data into other table





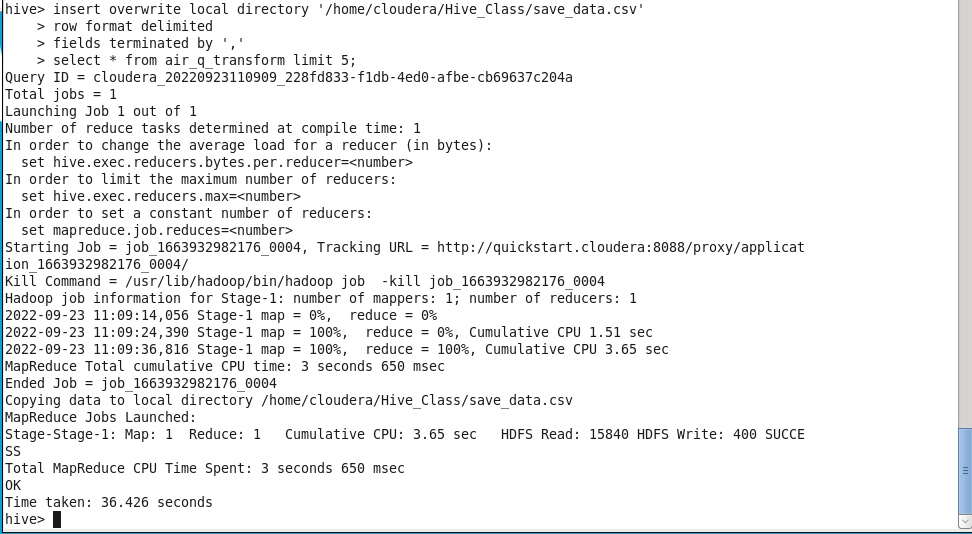
As u can see the comma separated decimal data got transformed into point separated data

3. Perform a select operation.

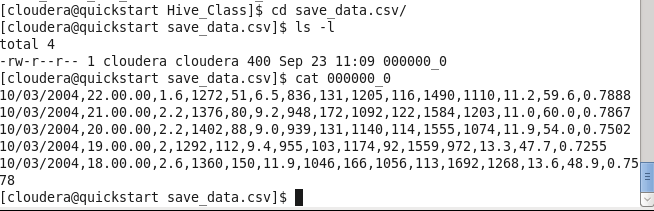


4. Fetch the result of the select operation in your local as a csv file.

Using insert overwrite local directory command I save the result of the query in the local system



Proof of data being stored in the local system in the form of partitioned file

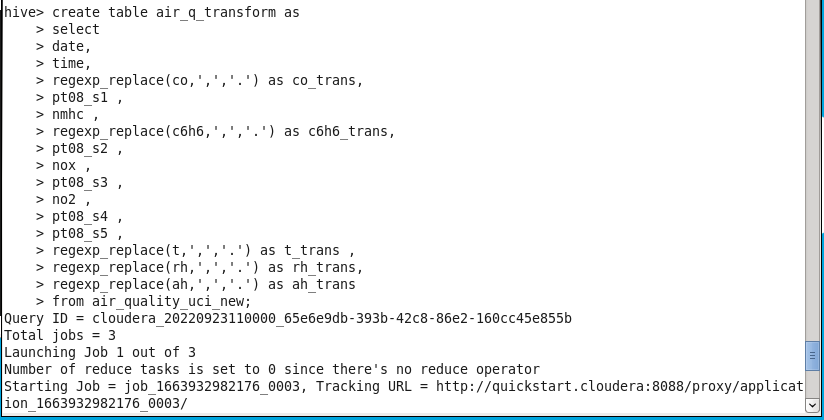


5. Perform group by operation.

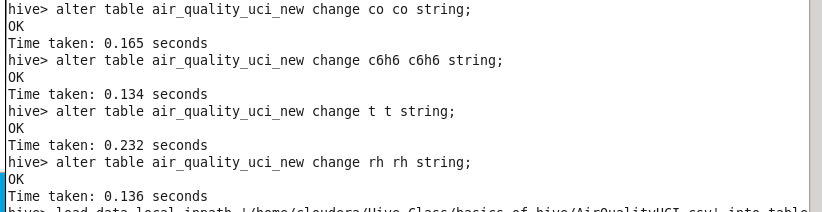


7. Perform filter operation at least 5 kinds of filter examples.

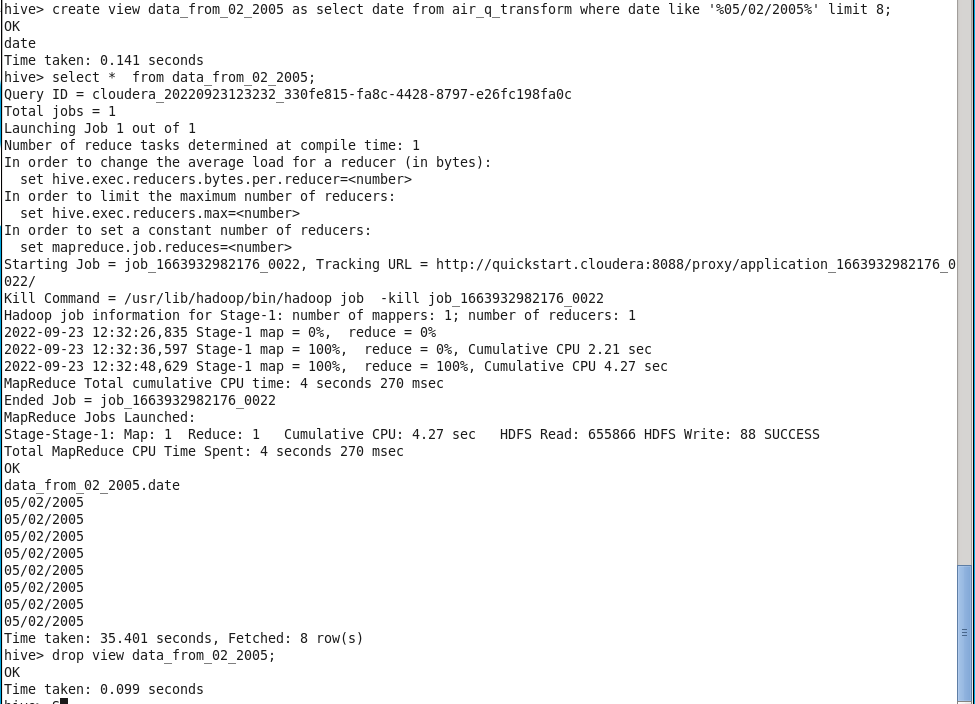
8. Show and example of regex operation



9. Alter table operation



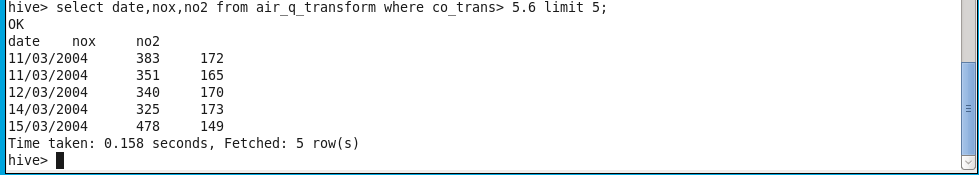
10. Drop table operation



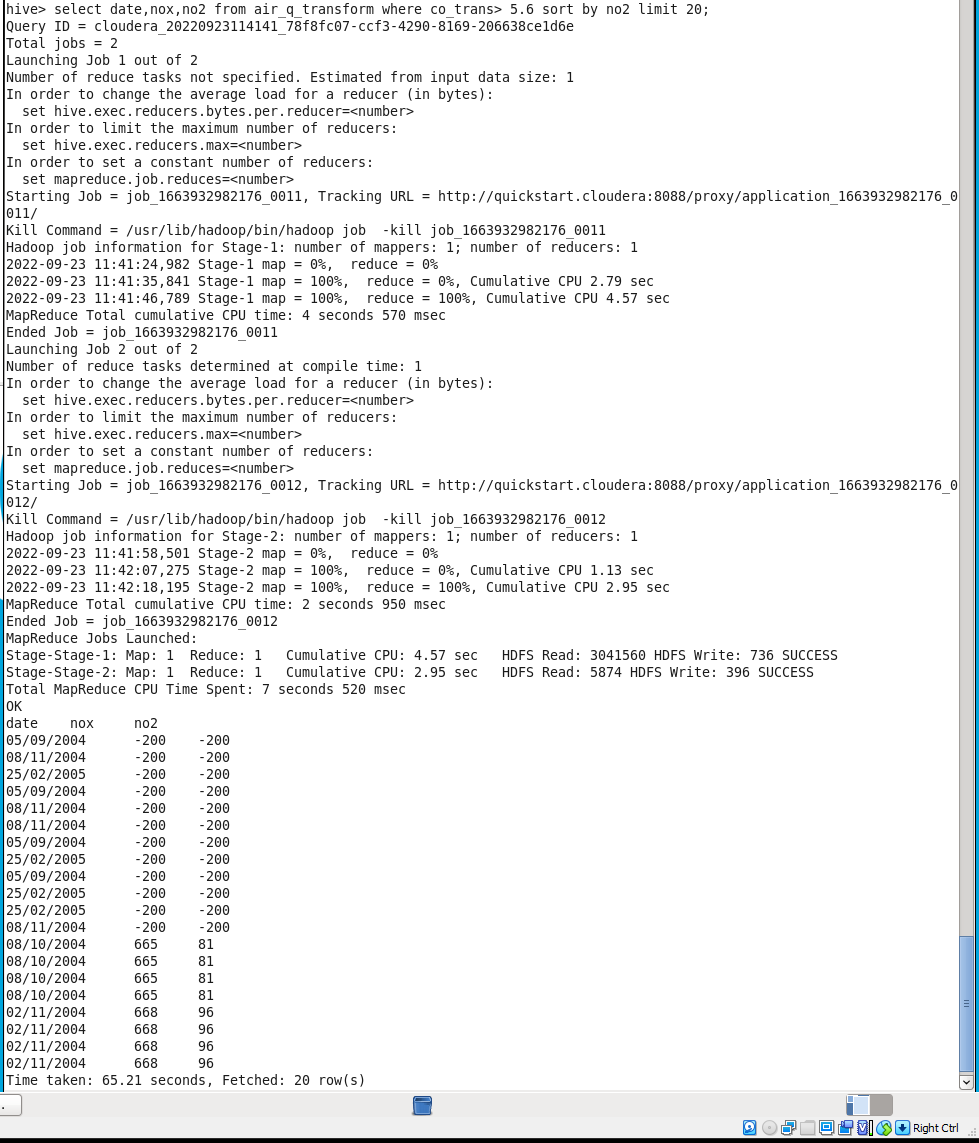
12. Order by operation.

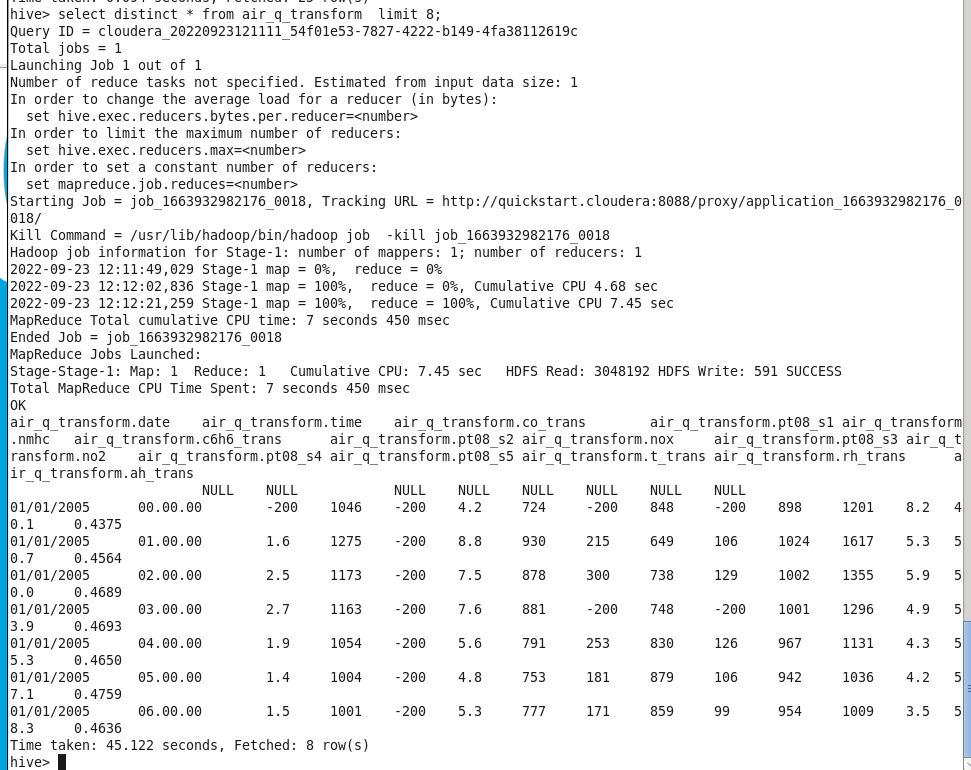


13. Where clause operations you have to perform.

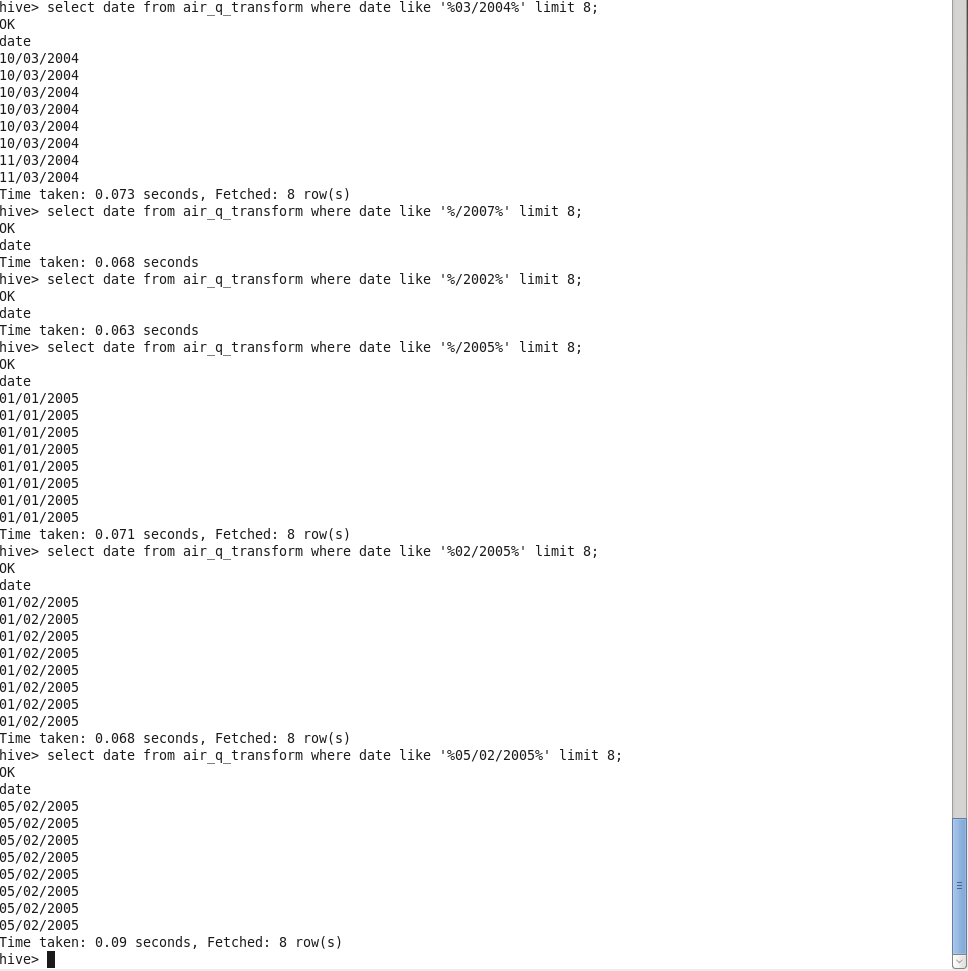


14. Sorting operation you have to perform.

15. Distinct operation you have to perform.



16. Like an operation you have to perform.



17. Union operation you have to perform.





18. Table view operation you have to perform.

