Q1.Will the reducer work or not if you use “Limit 1” in any HiveQL query?

Ans: No reducers are not used in query where we use limit

Text

Description automatically generated

Q2.Suppose I have installed Apache Hive on top of my Hadoop cluster using default metastore configuration. Then, what will happen if we have multiple clients trying to access Hive at the same time?

Ans: Multiple users will not be able to connect at the same time and will get error because the default metastore configuration allows only one Hive session to be opened at a time for accessing the metastore

Q3. Suppose, I create a table that contains details of all the transactions done by the customers: CREATE TABLE transaction\_details (cust\_id INT, amount FLOAT, month STRING, country STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ ;Now, after inserting 50,000 records in this table, I want to know the total revenue generated for each month. But, Hive is taking too much time in processing this query. How will you solve this problem and list the steps that I will be taking in order to do so?

Ans:The following are the steps to follow

1.Apply partition on column month

2.For each partition apply sum(sales) on column sales in query hence we get total sales of that month from the partition of that month.

3.hence from 12 parttion we get sales of 12 month easily in less time as we need to scan only records of that month to get sales of that month.

Q.How can you add a new partition for the month December in the above partitioned table?

Ans :

transaction\_details is normal table . transaction\_details\_part is partioned table on column month.we can add record of month December in partition of December month

by using query:

insert overwrite table transaction\_details\_part partition(country = 'USA') select \* from transaction\_details where month=’december’;

Q. I am inserting data into a table based on partitions dynamically. But, I received an error – FAILED ERROR IN SEMANTIC ANALYSIS: Dynamic partition strict mode requires at least one static partition column. How will you remove this error?

Ans: By default hive.exec.dynamic.partition.mode value is strict .this means that while inserting records in partition table we must specify at least one static partition column.hence we get the error in dynamic partitioning.we can resolve this by set hive.exec.dynamic.partition.mode=nonstrict.

Q. Suppose, I have a CSV file – ‘sample.csv’ present in ‘/temp’ directory with the following entries:id first\_name last\_name email gender ip\_address.How will you consume this CSV file into the Hive warehouse using built-in SerDe?

Ans:we can create table and load data in hive using below query

CREATE EXTERNAL TABLE sample

(id int, first\_name string,

last\_name string, email string,

gender string, ip\_address string)

ROW FORMAT SERDE ‘org.apache.hadoop.hive.serde2.OpenCSVSerde’

STORED AS TEXTFILE

LOCATION ‘/temp’

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

Q.Suppose, I have a lot of small CSV files present in the input directory in HDFS and I want to create a single Hive table corresponding to these files. The data in these files are in the format: {id, name, e-mail, country}. Now, as we know, Hadoop performance degrades when we use lots of small files.

So, how will you solve this problem where we want to create a single Hive table for lots of small files without degrading the performance of the system?

Create a temporary table:

CREATE TABLE temporary\_table (id INT, name STRING, e-mail STRING, country STRING)

ROW FORMAT FIELDS DELIMITED TERMINATED BY ‘,’ STORED AS TEXTFILE;

Load the data into tempporary\_table:

LOAD DATA INPATH ‘/input’ INTO TABLE temporary\_table;

Create a table that will store data in SequenceFile format:

CREATE TABLE sample\_seqfile (id INT, name STRING, e-mail STRING, country STRING)

ROW FORMAT FIELDS DELIMITED TERMINATED BY ‘,’ STORED AS SEQUENCEFILE;

Transfer the data from the temporary table into the sample\_seqfile table:

INSERT OVERWRITE TABLE sample\_seqfile SELECT \* FROM temporary\_table;

Q. LOAD DATA LOCAL INPATH ‘Home/country/state/’

OVERWRITE INTO TABLE address;

The following statement failed to execute. What can be the cause?

Ans: The local inpath should contain a file and not a directory.here it contains a directory hence we are getting errors.

Q. Is it possible to add 100 nodes when we already have 100 nodes in Hive? If yes, how?

Ans: Yes, we can add the nodes by following the below steps:

Step 1: Take a new system; create a new username and password

Step 2: Install SSH and with the master node setup SSH connections

Step 3: Add ssh public\_rsa id key to the authorized keys file

Step 4: Add the new DataNode hostname, IP address, and other details in /etc/hosts slaves file:

192.168.1.102 slave3.in slave3

Step 5: Start the DataNode on a new node

Step 6: Login to the new node like suhadoop or:

ssh -X hadoop@192.168.1.103

Step 7: Start HDFS of the newly added slave node by using the following command:

./bin/hadoop-daemon.sh start data node

Step 8: Check the output of the jps command on the new node

**Hive Practical questions:**

**Hive Join operations**

Create a table named CUSTOMERS(ID | NAME | AGE | ADDRESS | SALARY)

Create a Second table ORDER(OID | DATE | CUSTOMER\_ID | AMOUNT

)

Now perform different joins operations on top of these tables

(Inner JOIN, LEFT OUTER JOIN ,RIGHT OUTER JOIN ,FULL OUTER JOIN)

Text

Description automatically generated

Text

Description automatically generated

select \* from customers inner join orders on customers.id=orders.customerid;

Text

Description automatically generated

select \* from customers left outer join orders on customers.id=orders.customerid;

Text

Description automatically generated

select \* from customers right outer join orders on customers.id=orders.customerid;

Text

Description automatically generated

select \* from customers full outer join orders on customers.id=orders.customerid;

Graphical user interface, text

Description automatically generated with medium confidence

**BUILD A DATA PIPELINE WITH HIVE**

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

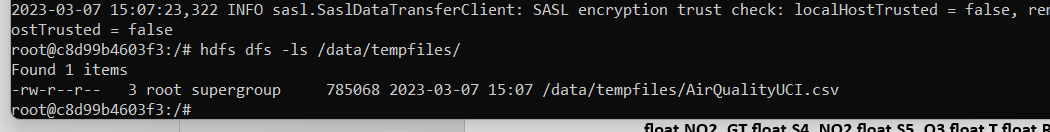
[**https://archive.ics.uci.edu/ml/machine-learning-databases/00360/**](https://archive.ics.uci.edu/ml/machine-learning-databases/00360/)

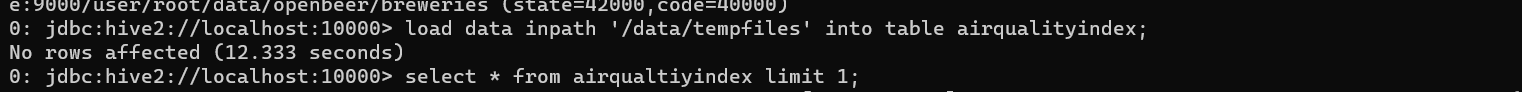
1. Create a hive table as per given schema in your dataset

**create table airqualityindex(dateofrecord string,timeofrecord string,Carbonoxide float,S1\_CO float,NMHC\_GT float ,C6H6\_GT float,S2\_NMHC float,NOx\_GT float,S3\_NOx float,NO2\_GT float,S4\_NO2 float,S5\_O3 float,T float,RH float,AH float)row format delimited fields terminated by ',';**

1. try to place a data into table location

**load data inpath '/data/tempfiles' into table airqualityindex;**

****

****

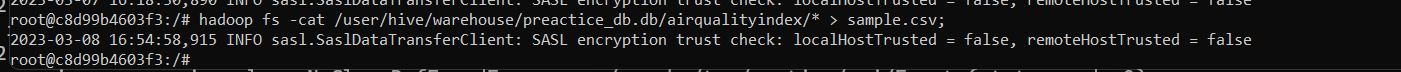
3. Perform a select operation

**Text

Description automatically generated with low confidence**

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

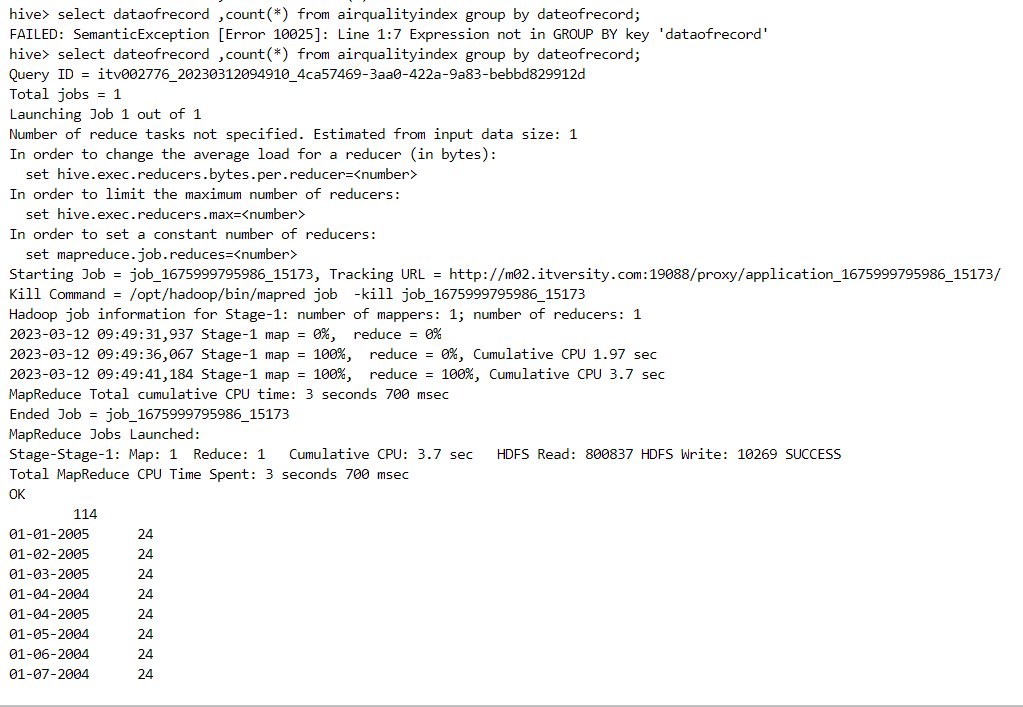
**Ans:**



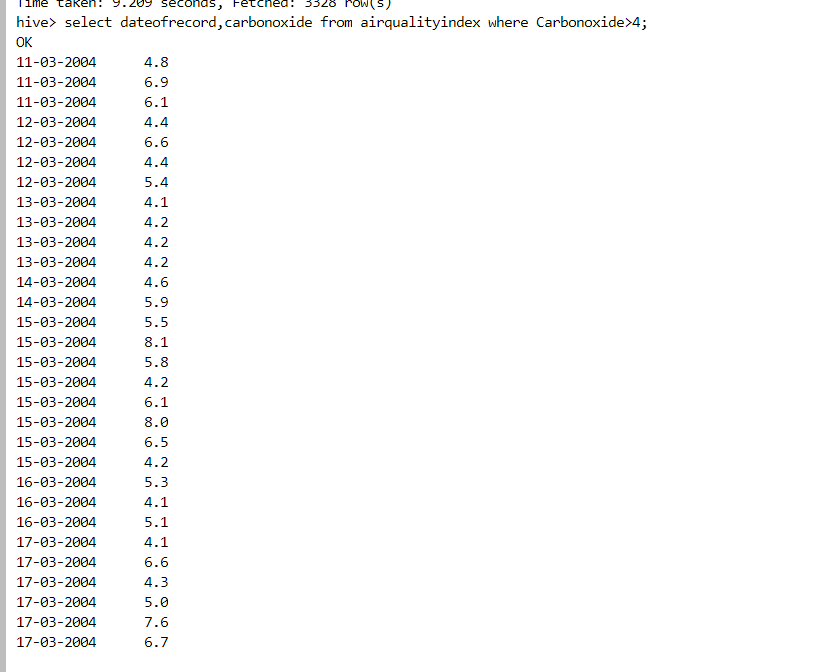


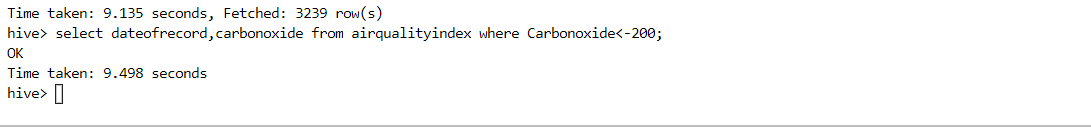
1. Perform group by operation

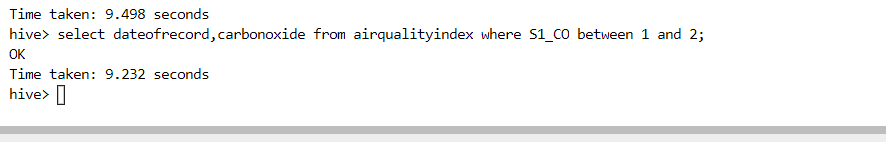
Ans:

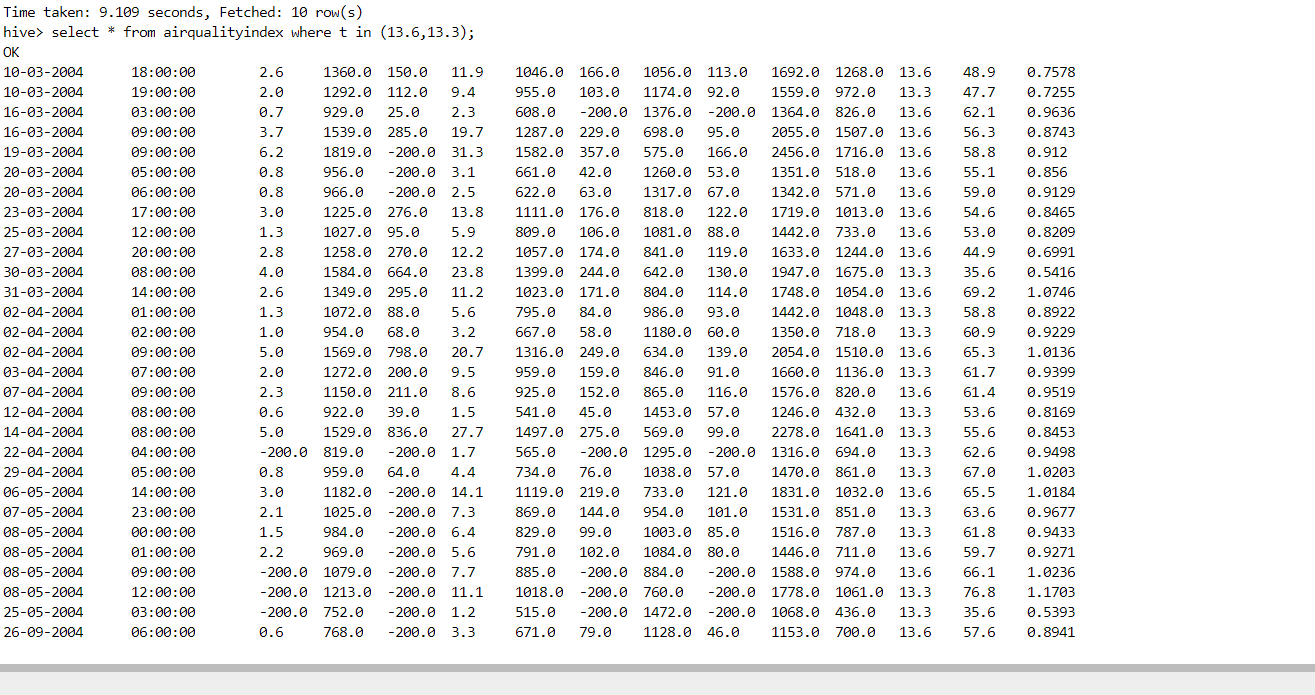


7. Perform filter operation at least 5 kinds of filter examples .

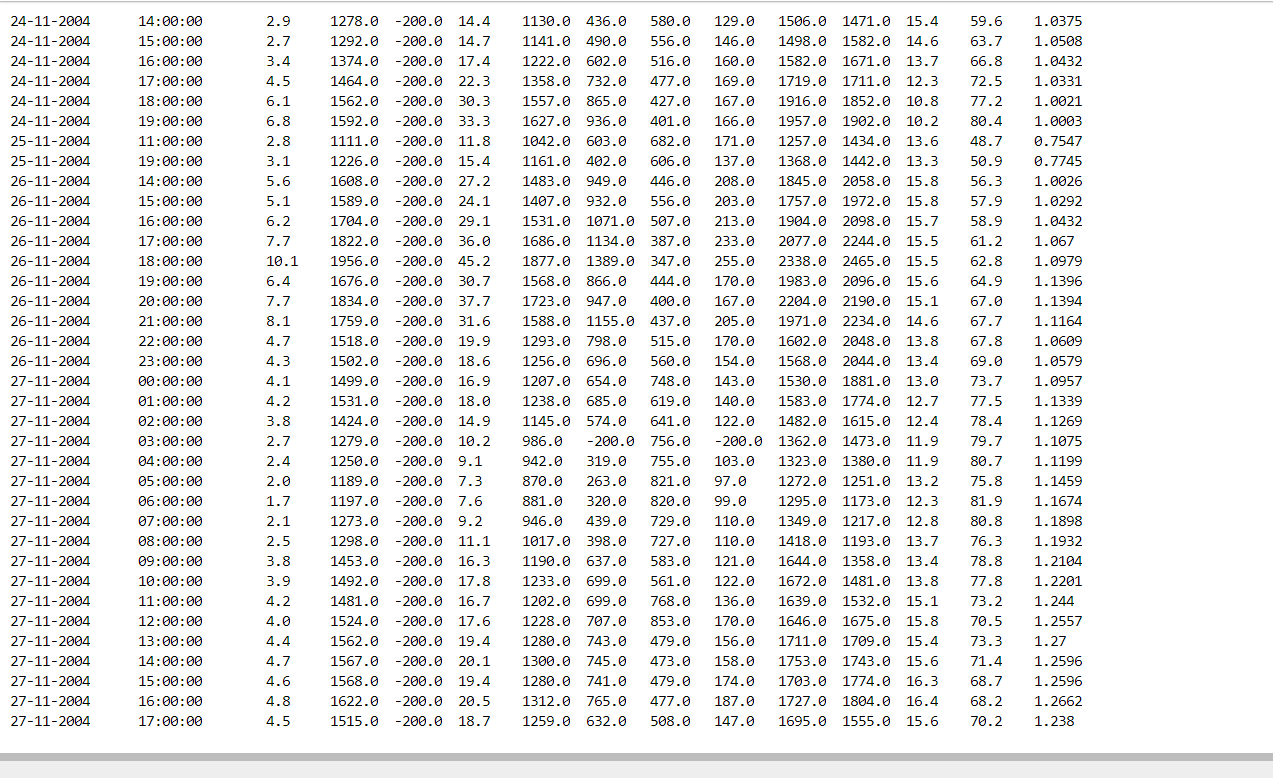




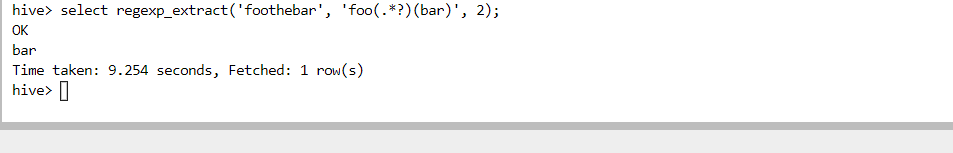




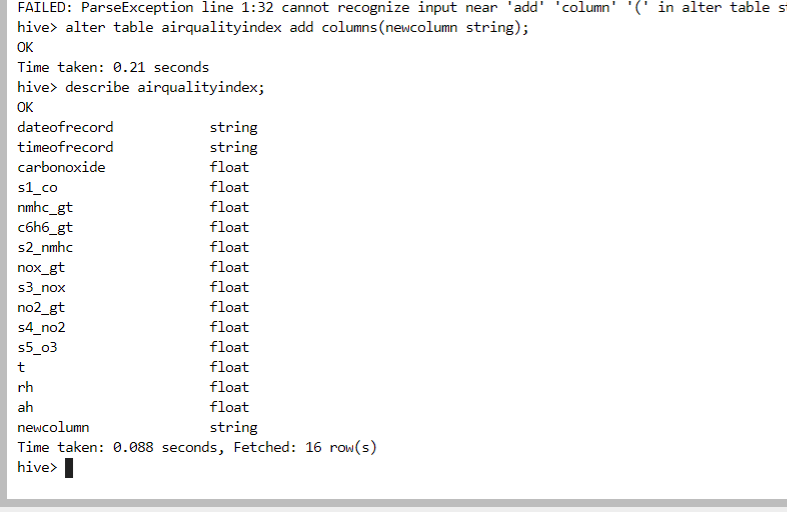
****

****

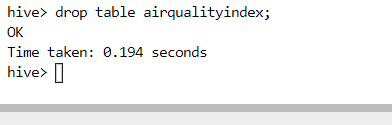
8. show and example of regex operation

****

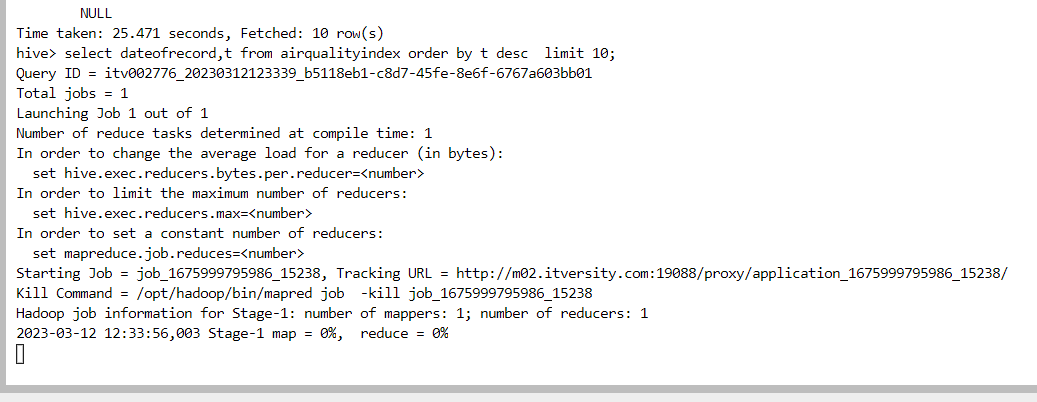
9. alter table operation

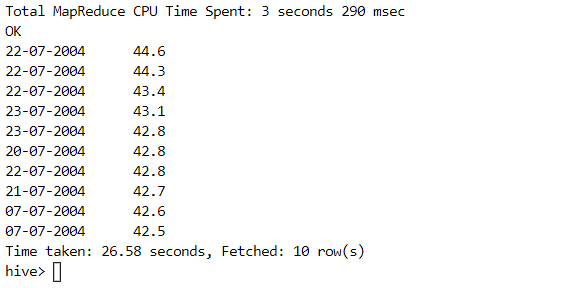
****

10 . drop table operation

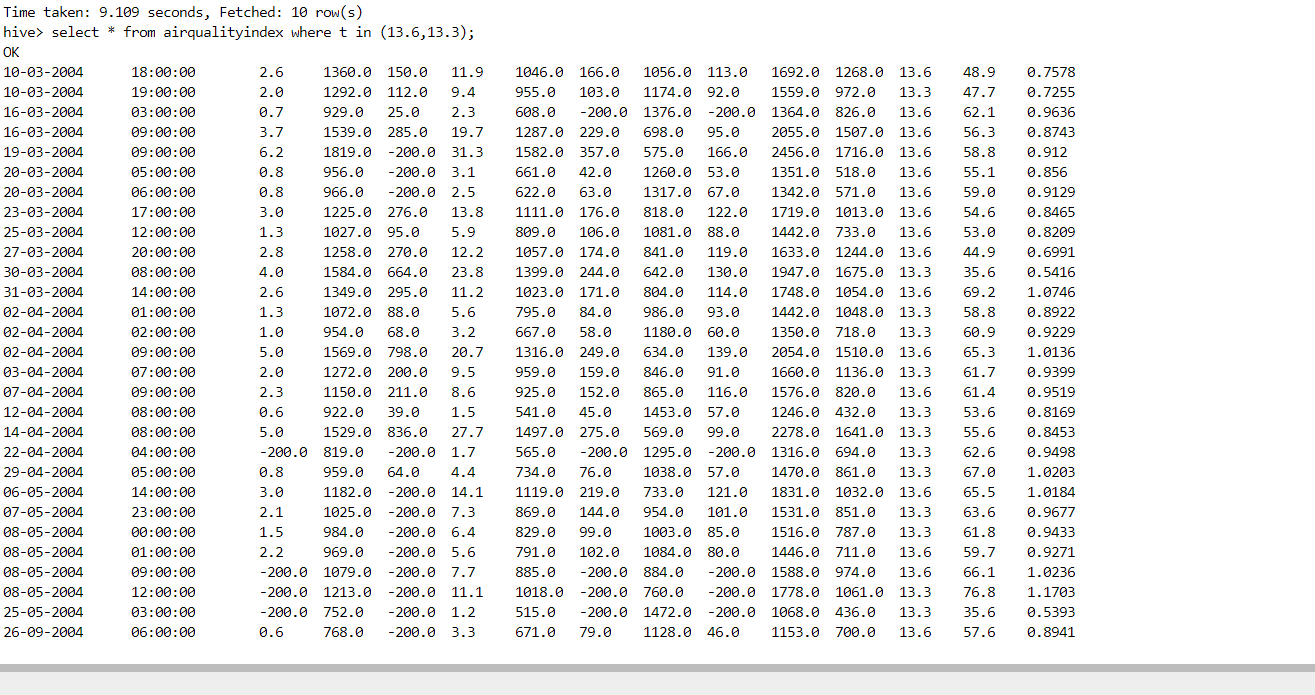


12 . order by operation .

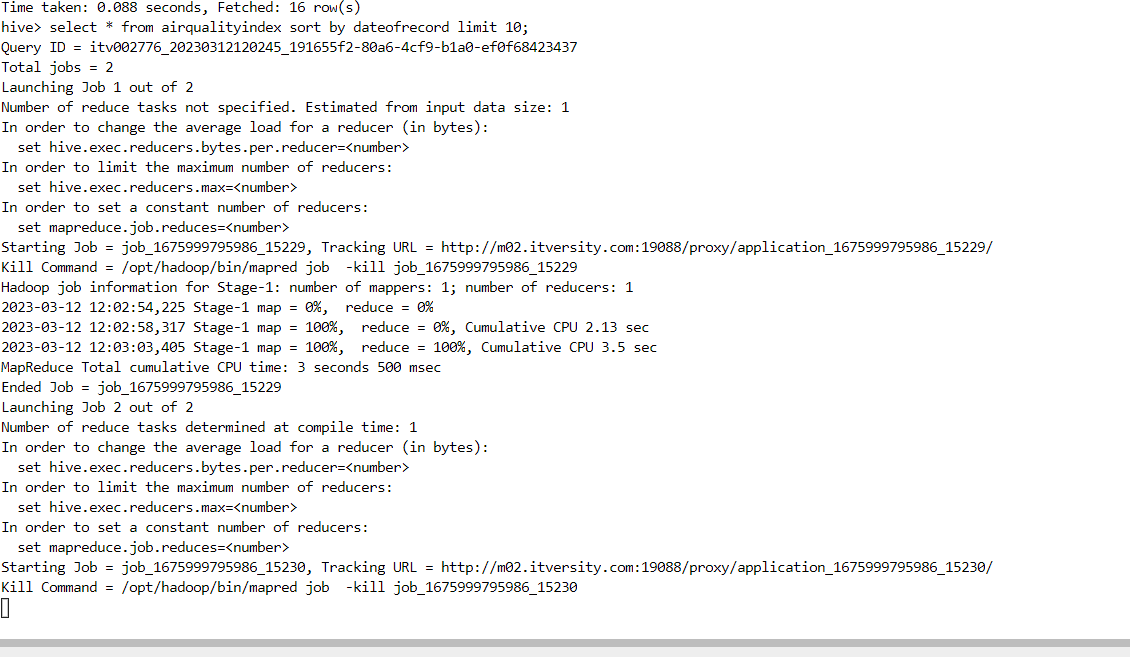
****

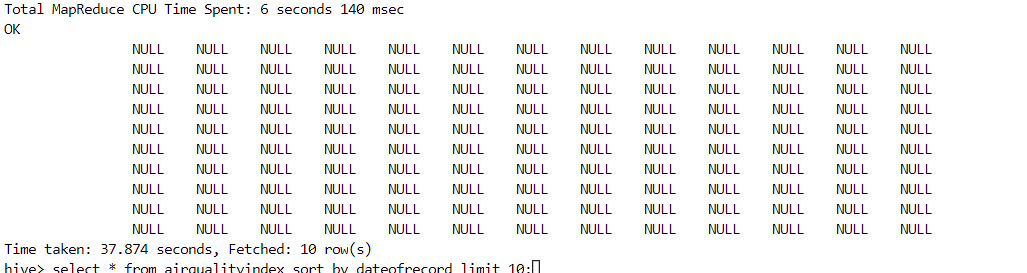
****

13 . where clause operations you have to perform .

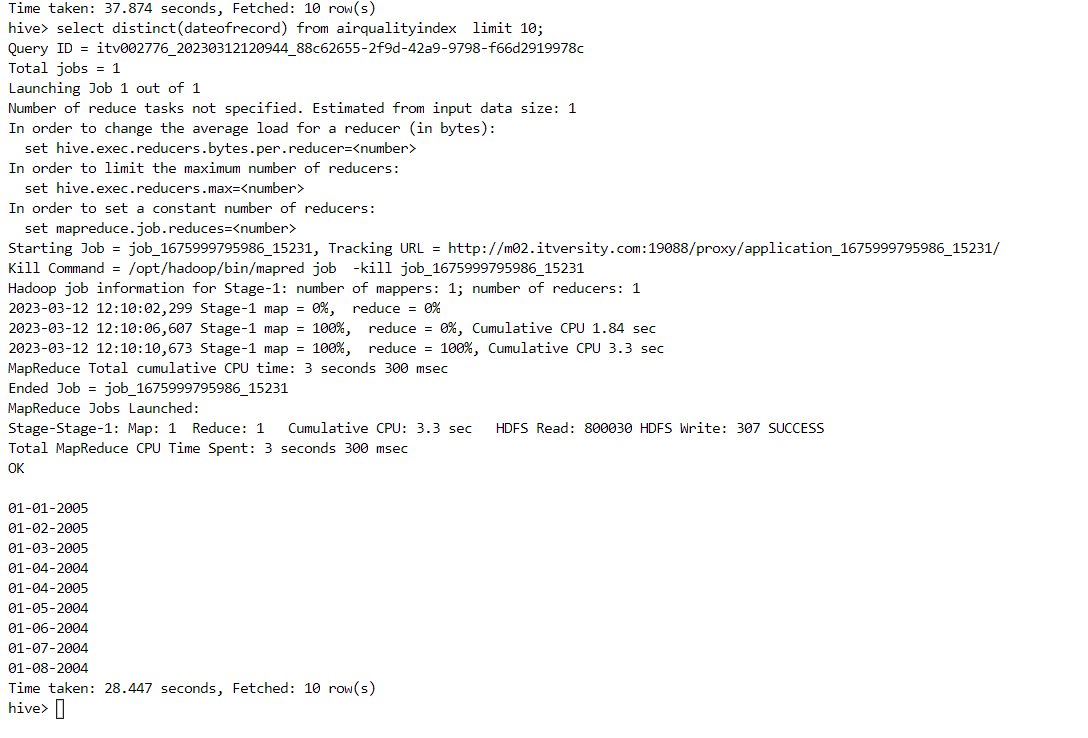


14 . sorting operation you have to perform .

****

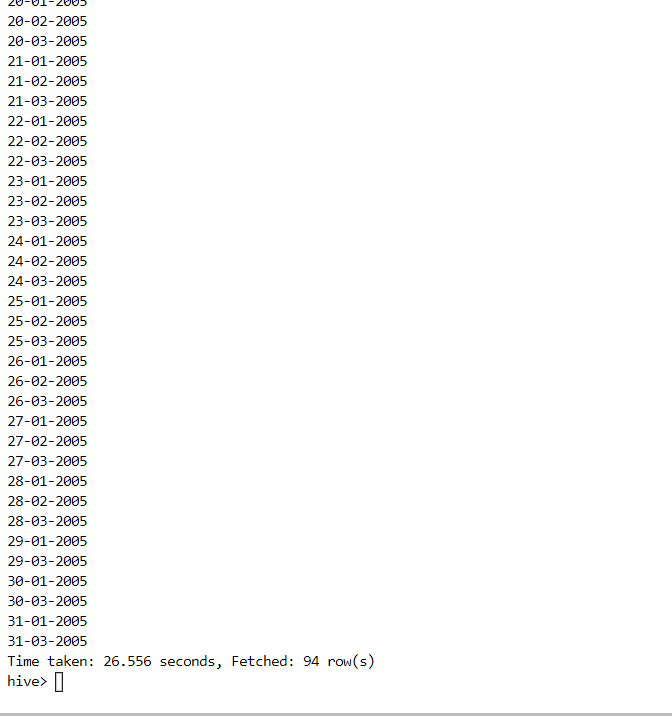
****

15 . distinct operation you have to perform .

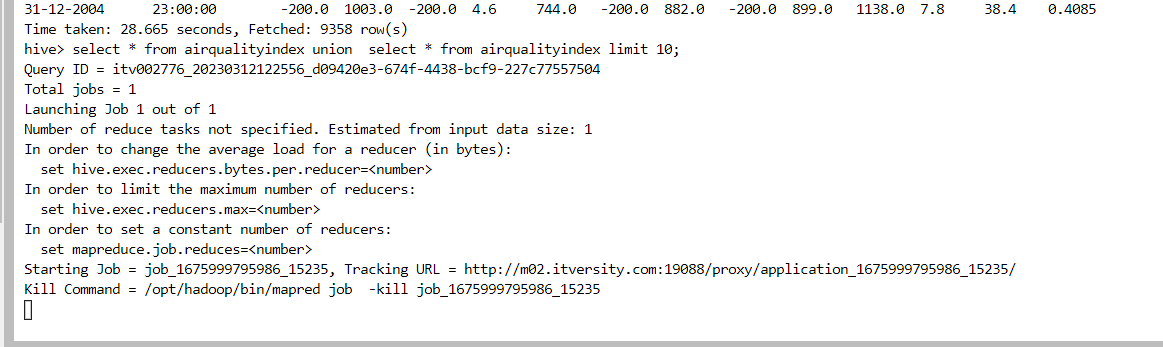
****

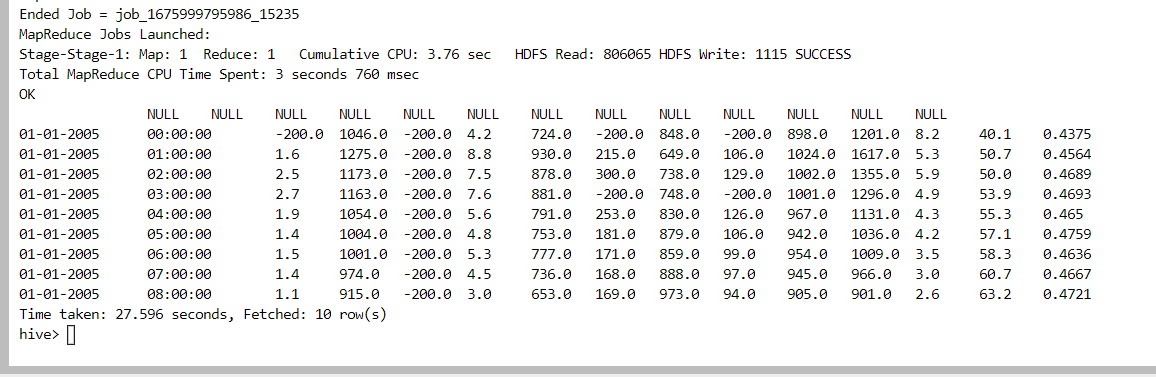
16 . like an operation you have to perform .

****

****

17 . union operation you have to perform .





18 . table view operation you have to perform .

