**HIVE MINI PROJECT : 3**

**Description:**

This is a project which is created in hive. The project was to create and maintain a database for candidate enrollment where they provide training for various courses .

In this project we have created five tables. We have performed various operations in this project like table join , partitioning, Clustering, insertion etc. We have also performed queries in in tables where we extracted the data from it.

**🡺Create a Table Enquiry**

create table enquiry\_txts(id int,name string,mail STRING,mob BIGINT,course STRING,fee INT,discount INT,dte STRING,status STRING,status\_final STRING)

row format delimited

fields terminated by ','

stored as textfile;

**🡺Create ORC table with partitioning and clustering**

create table enquiry

(

id INT,

name STRING,

mail STRING,

mob BIGINT,

course STRING,

fee INT,

discount INT,

dte STRING,

status STRING

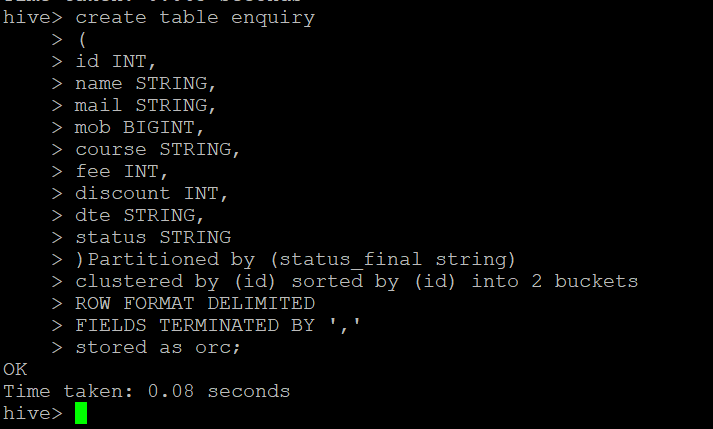
)Partitioned by (status\_final string)

clustered by (id) sorted by (id) into 2 buckets

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

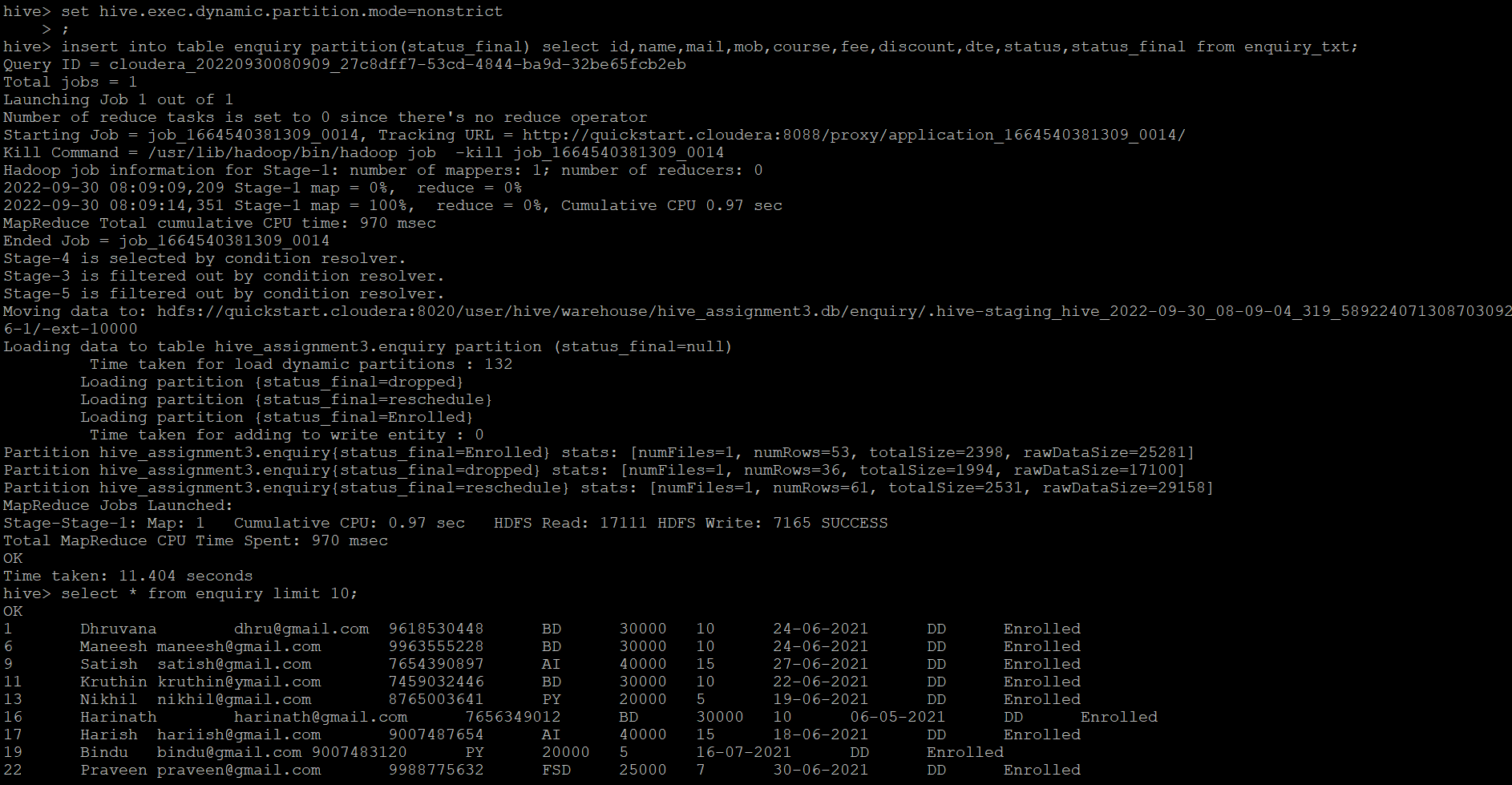
stored as orc;



**🡺Load tha Data into tables:**

load data local inpath '/home/cloudera/hive\_assignment3/Enquiry.csv' into table enquiry\_txt;

set hive.exec.dynamic.partition.mode=nonstrict;

insert into table enquiry partition(status\_final) select id,name,mail,mob,course,fee,discount,dte,status,status\_ﬁnal from enquiry\_txt;

**🡺DEMO\_SCHECULE Table**

create table demo\_schedule(id INT,name STRING,mail STRING,mob BIGINT,course STRING,demoday STRING,day STRING,dte string,time STRING)

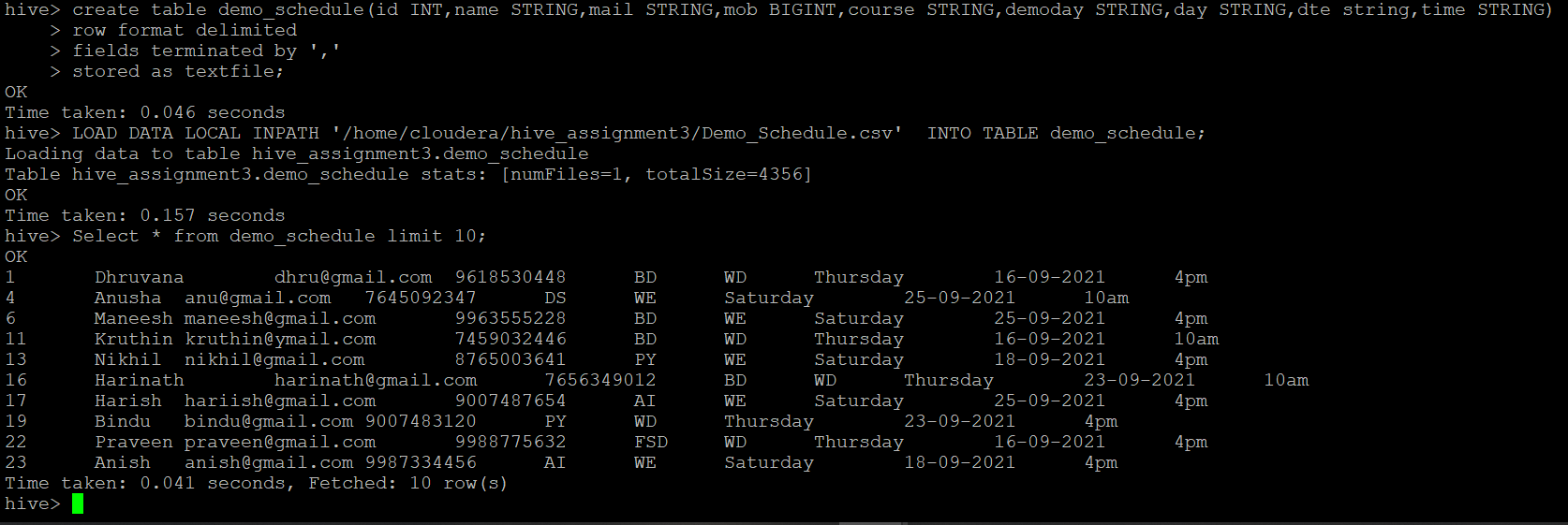
row format delimited

fields terminated by ','

stored as textfile;

**🡺Load the Data in table:**

LOAD DATA LOCAL INPATH '/home/cloudera/hive\_assignment3/Demo\_Schedule.csv' INTO TABLE demo\_schedule;



**🡺DEMO MISSED Table:**

create table demo\_missed(id int,name string,mail string,mob bigint,course string,reschedule\_date string,time string)

row format delimited

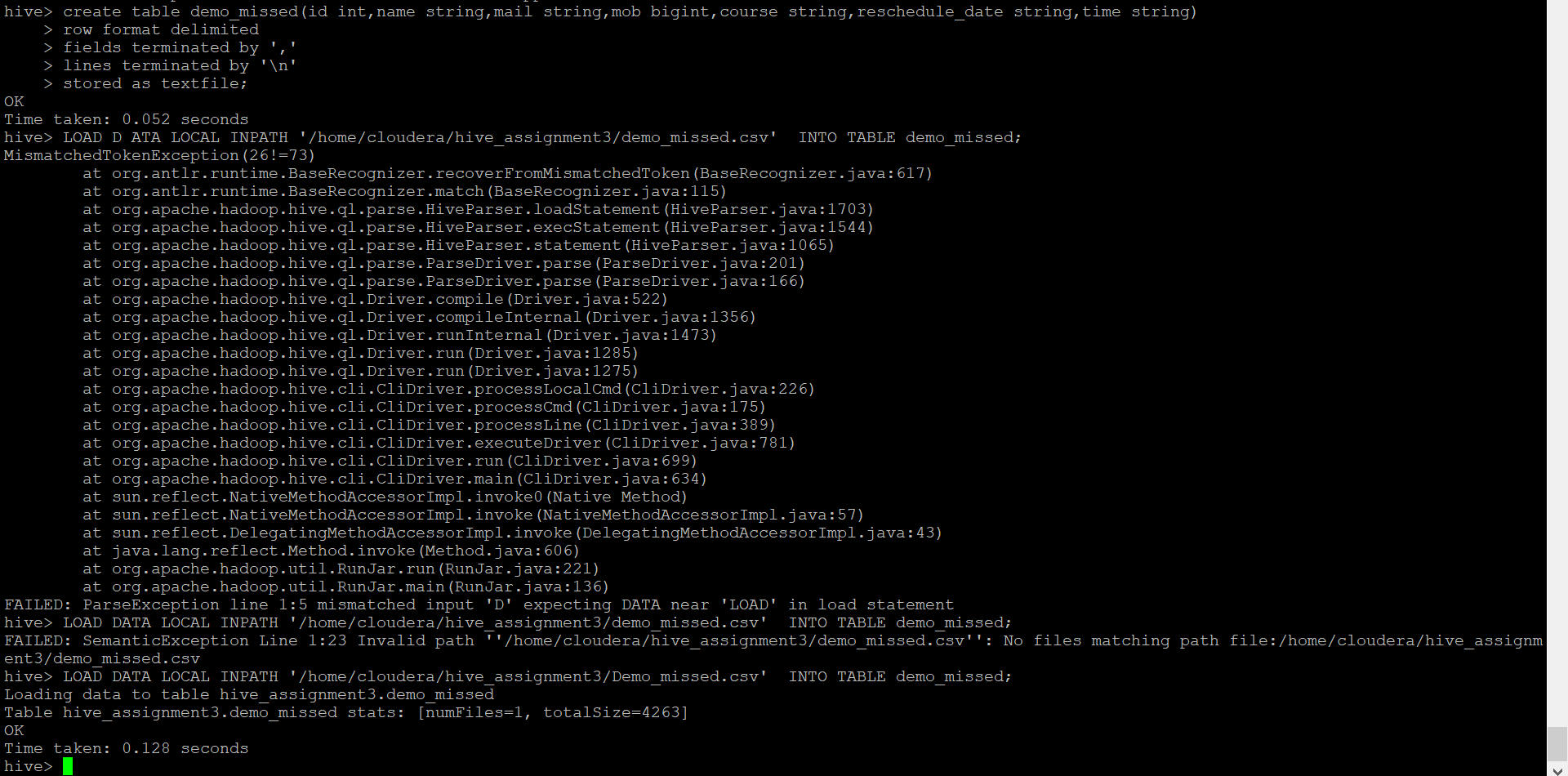
fields terminated by ','

lines terminated by '\n'

stored as textfile;

**🡺Load Data into Table**

LOAD DATA LOCAL INPATH '/home/cloudera/hive\_assignment3/demo\_missed.csv' INTO TABLE demo\_missed;



**🡺ENROLL Table:**

create table enroll (id int,name string,mail string,mob bigint,course string,status string,payment\_mode string,total\_installment int,installment\_paid int,Next\_installment\_date string,payment\_status string,installment\_left int,final\_fee int)

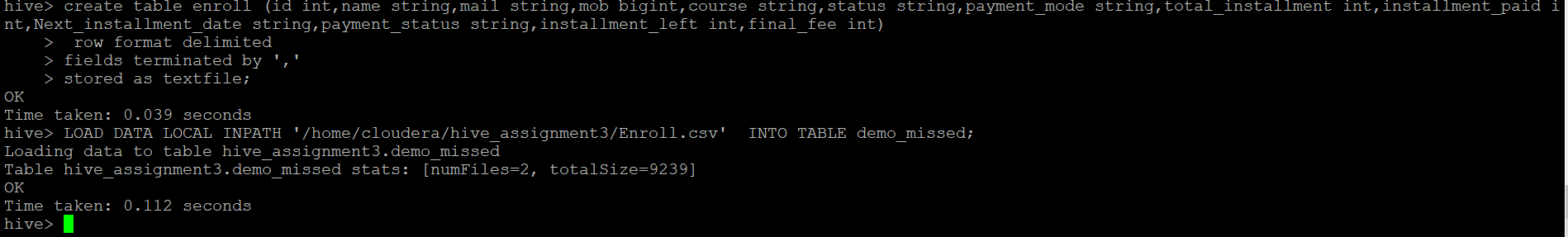
row format delimited

fields terminated by ','

stored as textfile;

**🡺Load data into table:**

LOAD DATA LOCAL INPATH ‘/home/cloudera/hive\_assignment3/Enroll.csv' INTO TABLE enroll;



**🡺BATCH DISTRIBUTION table:**

create table batch\_distribution(id int,name string,mail string,mob bigint,course string,joining\_date string,week\_day string,day string ,time string)

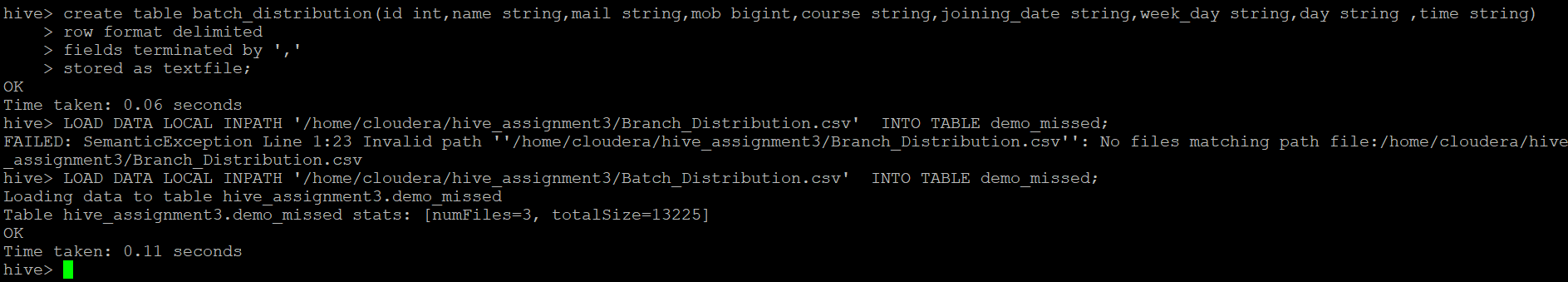
row format delimited

fields terminated by ','

stored as textfile;

**🡺Load data into table:**

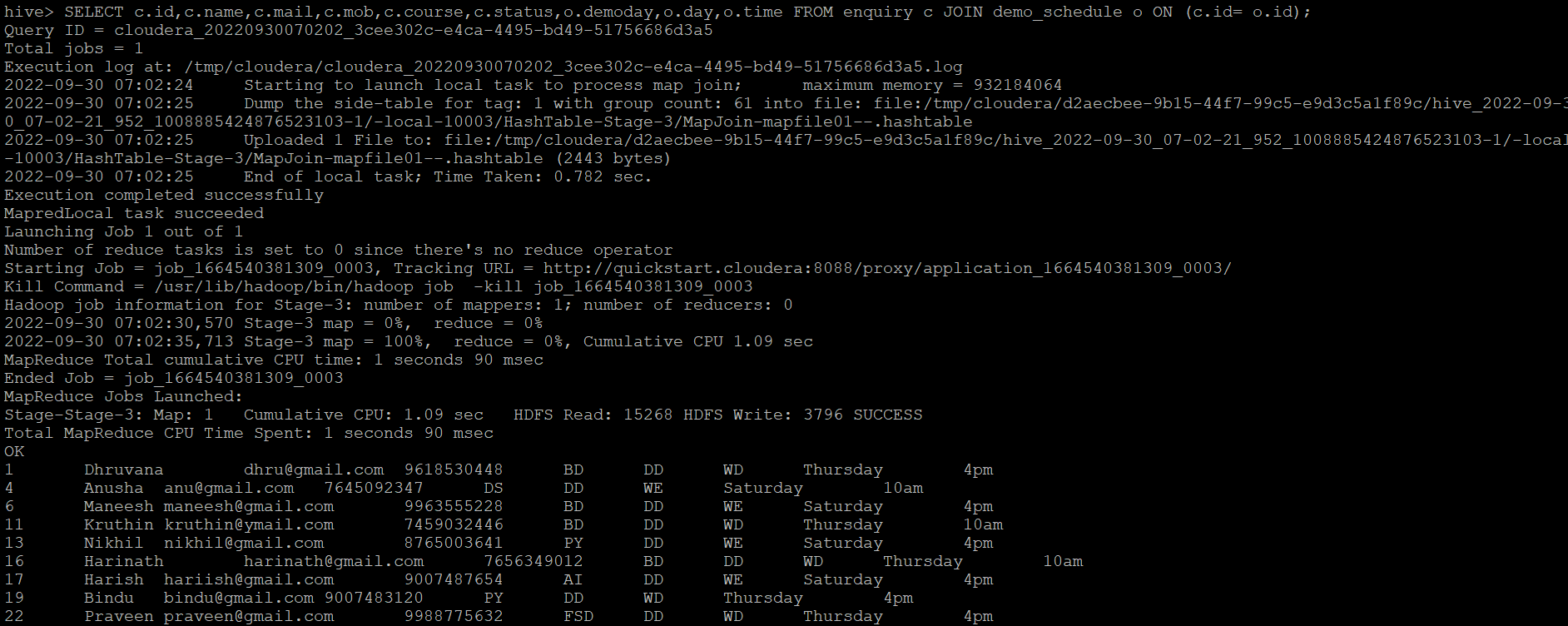
LOAD DATA LOCAL INPATH '/home/cloudera/hive\_assignment3/batch\_distribution.csv' INTO TABLEbatch\_distribution;



**Questions:**

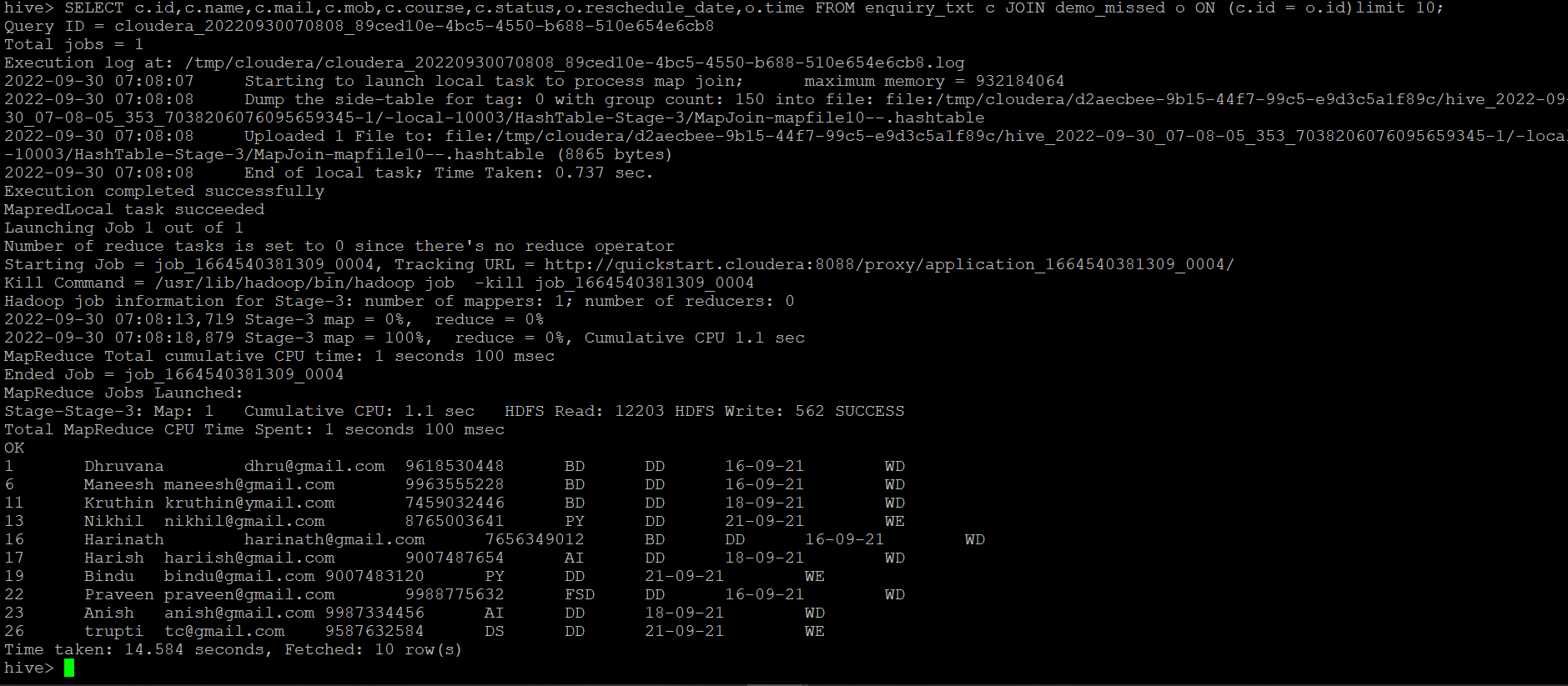
**🡺Join on enquiry and demo\_schedule table:**

SELECT c.id,c.name,c.mail,c.mob,c.course,c.status,o.demoday,o.day,o.time FROM enquiry c JOIN demo\_schedule o ON (c.id= o.id);



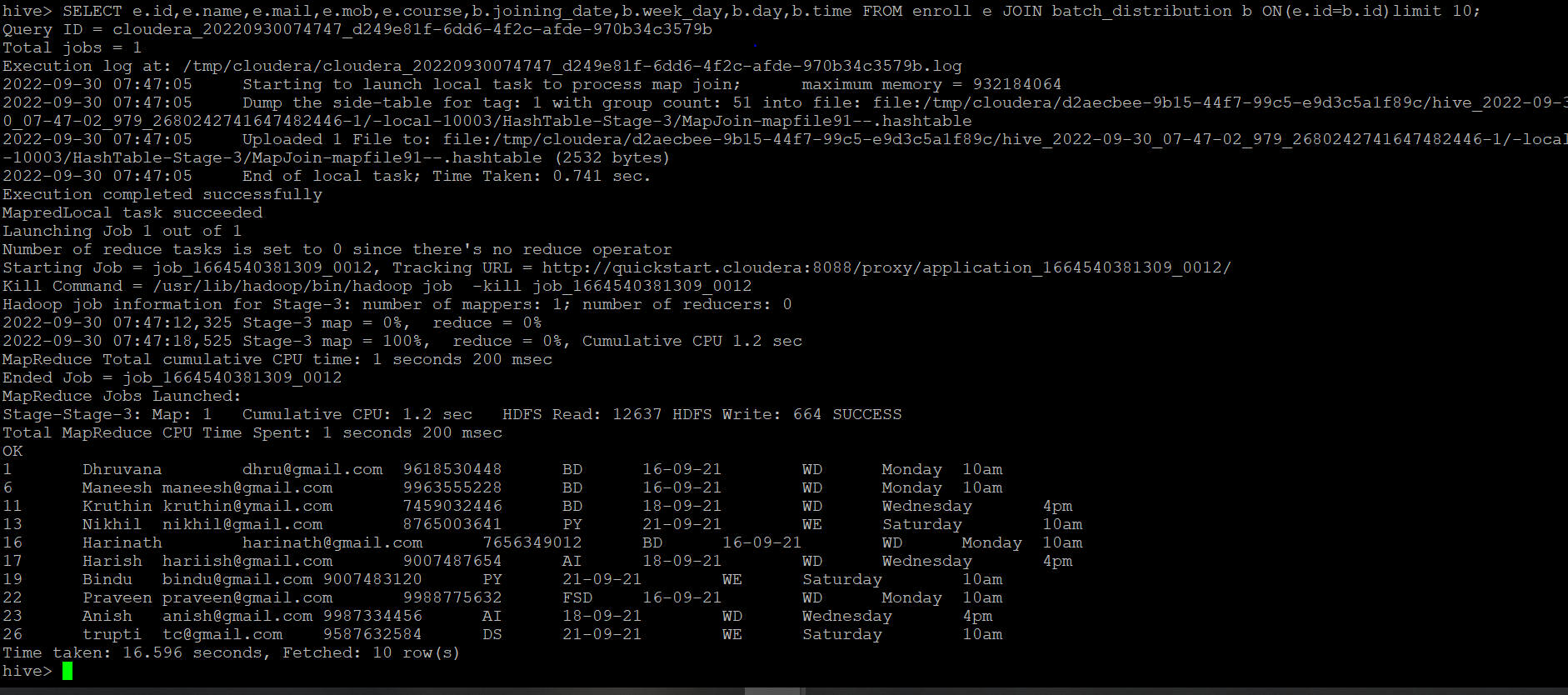
🡺 **Join on enquiry\_txt and demo\_missed:**

SELECT c.id,c.name,c.mail,c.mob,c.course,c.status,o.reschedule\_date,o.time FROM enquiry\_txt c JOIN demo\_missed o ON (c.id = o.id);



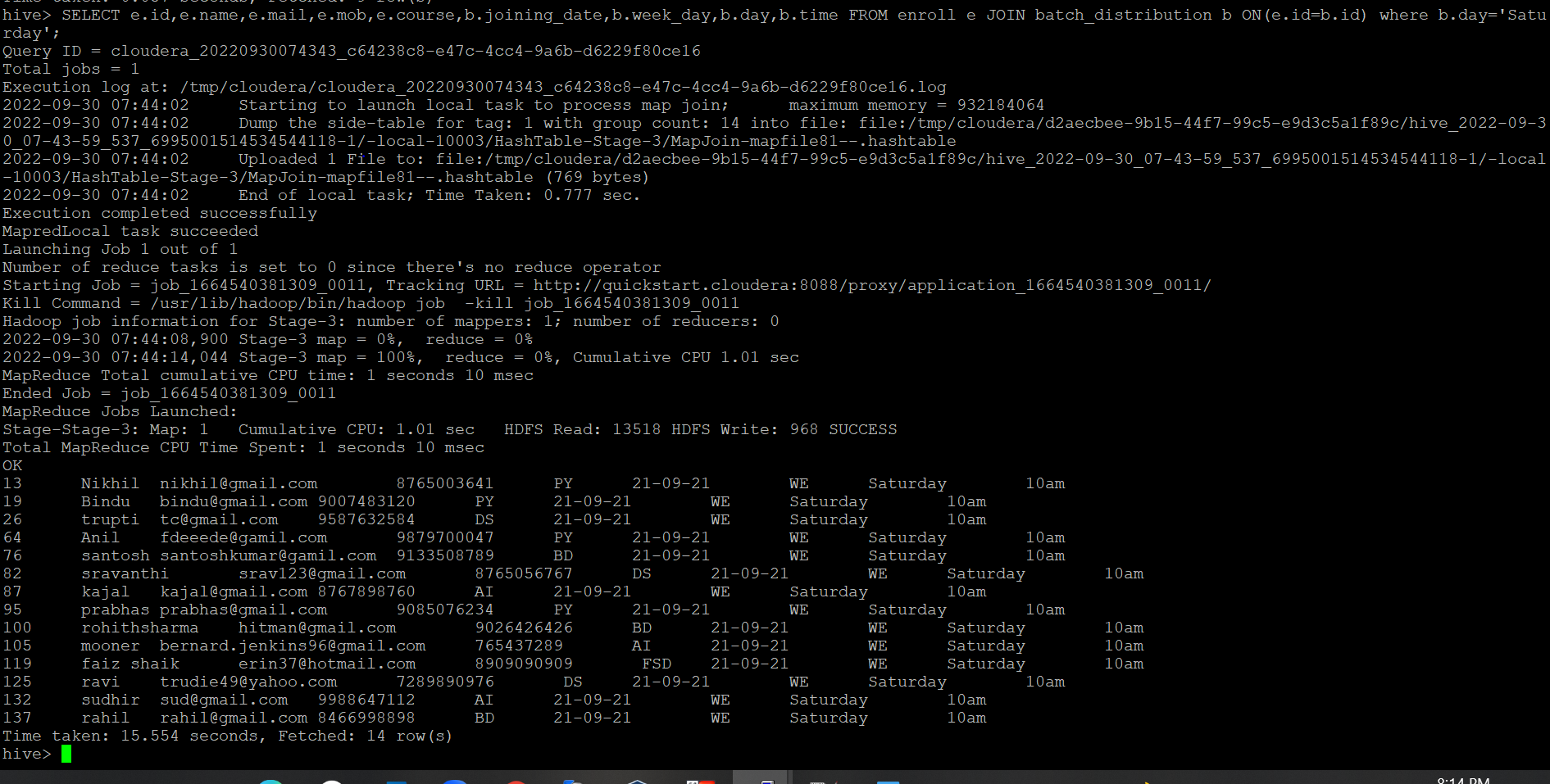
**🡺** **enroll and batch(which student enroll for which batch)**

SELECT e.id,e.name,e.mail,e.mob,e.course,b.joining\_date,b.week\_day,b.day,b.time FROM enroll e JOIN batch\_distribution b ON(e.id=b.id)limit 10;



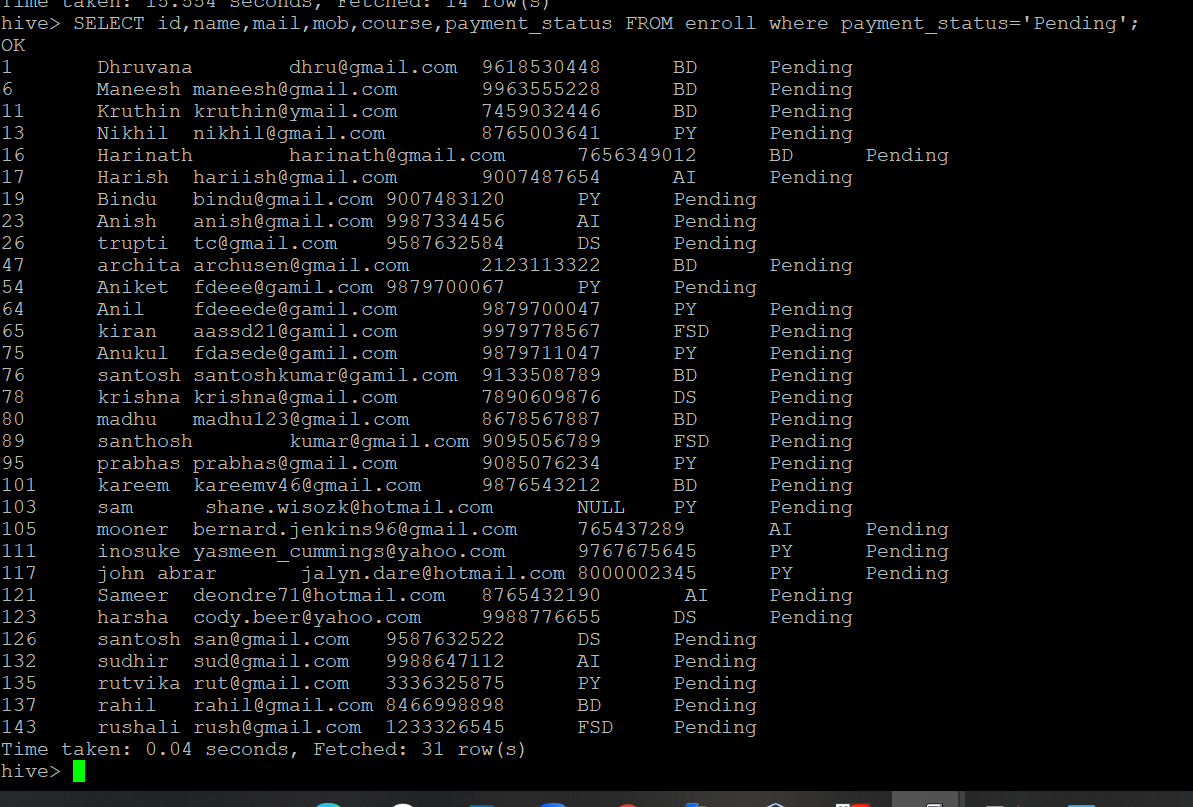
**🡺** **students who enrolled for weekend batch**

SELECT e.id,e.name,e.mail,e.mob,e.course,b.joining\_date,b.week\_day,b.day,b.time FROM enroll e JOIN batch\_distribution b ON(e.id=b.id) where b.day='Saturday';



**🡺Details of student whoes fees pending**

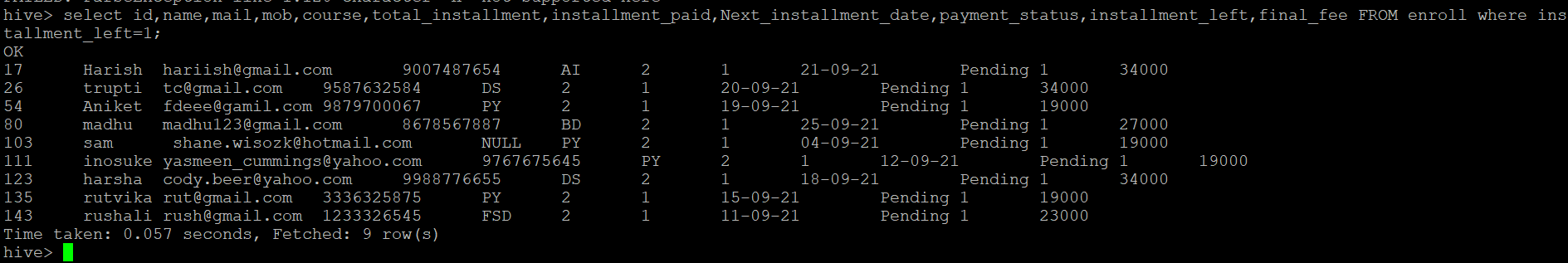
SELECT id,name,mail,mob,course,payment\_status FROM enroll where payment\_status='Pending';



**🡺Details of student whose 1 installment pending**

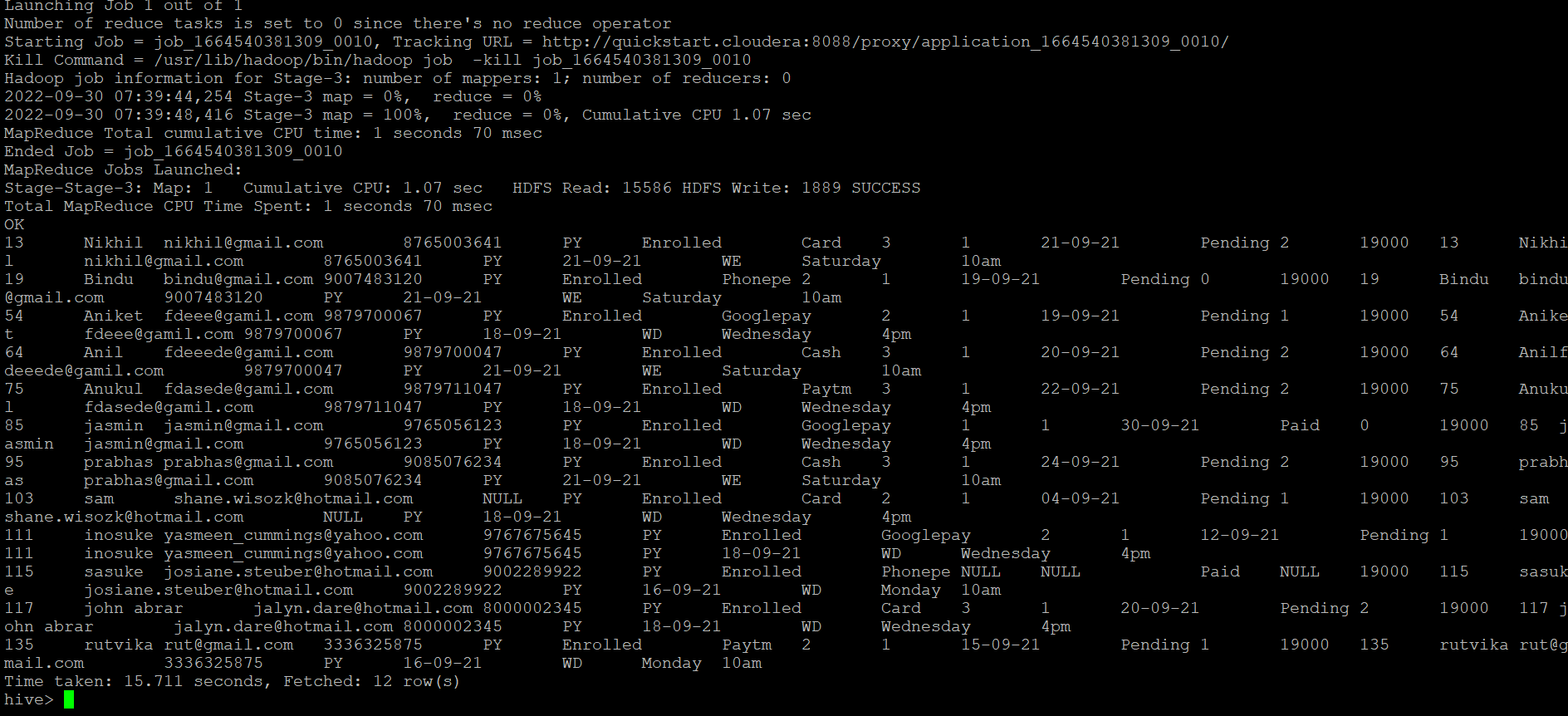
Select id,name,mail,mob,course,total\_installment,installment\_paid,Next\_installment\_date,

payment\_status,installment\_left,ﬁnal\_fee FROM enroll where installment\_left=1;



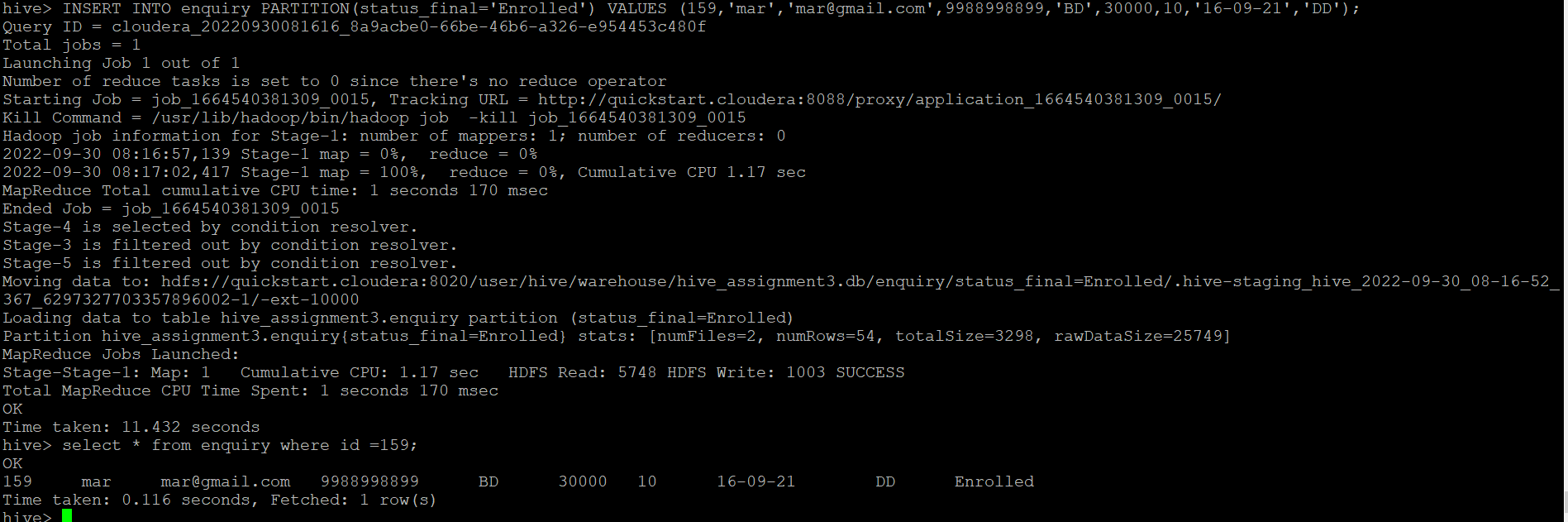
**🡺All details of student who enrolled for python batch**

SELECT \* FROM enroll e JOIN batch\_distribution b ON(e.id=b.id) where e.course=‘PY’



**🡺Insert into table Enquiry:**

INSERT INTO enquiry PARTITION(status\_final='Enrolled') VALUES (159,'mar','mar@gmail.com',9988998899,'BD',30000,10,'16-09-21','DD');



**🡺List the student whose demo scheduled at 4pm.**

Select \* from demo\_schedule where day= ‘25-09-2021’ and time = ‘4pm’;

