*HIVE COMMANDS*

create database dna;

show databases;

use dna;

create table info(name varchar(50), AGATC int, AATG int, TATC int);

load data local inpath '/home/hadoop/Documents/small.csv' into table info;

insert into info values('Alice',2,8,3);

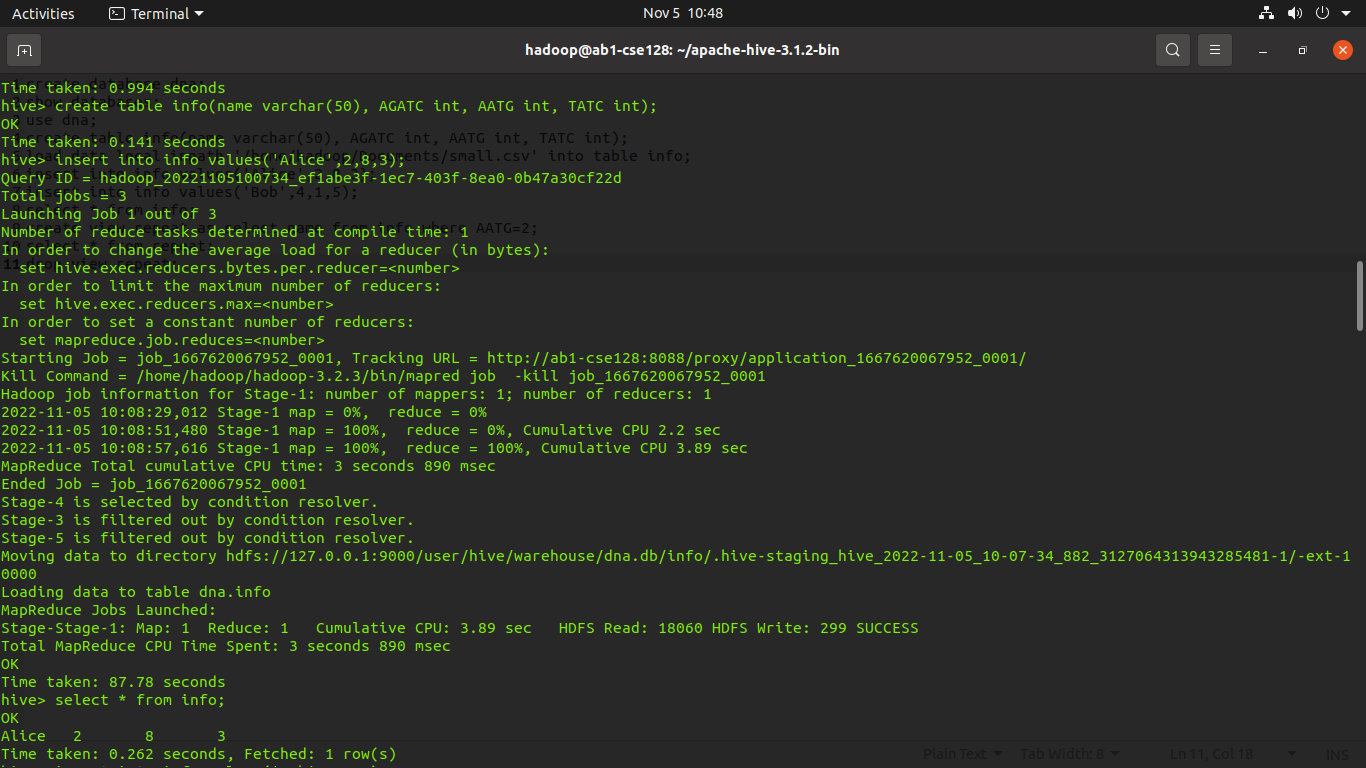
insert into info values('Bob',4,1,5);

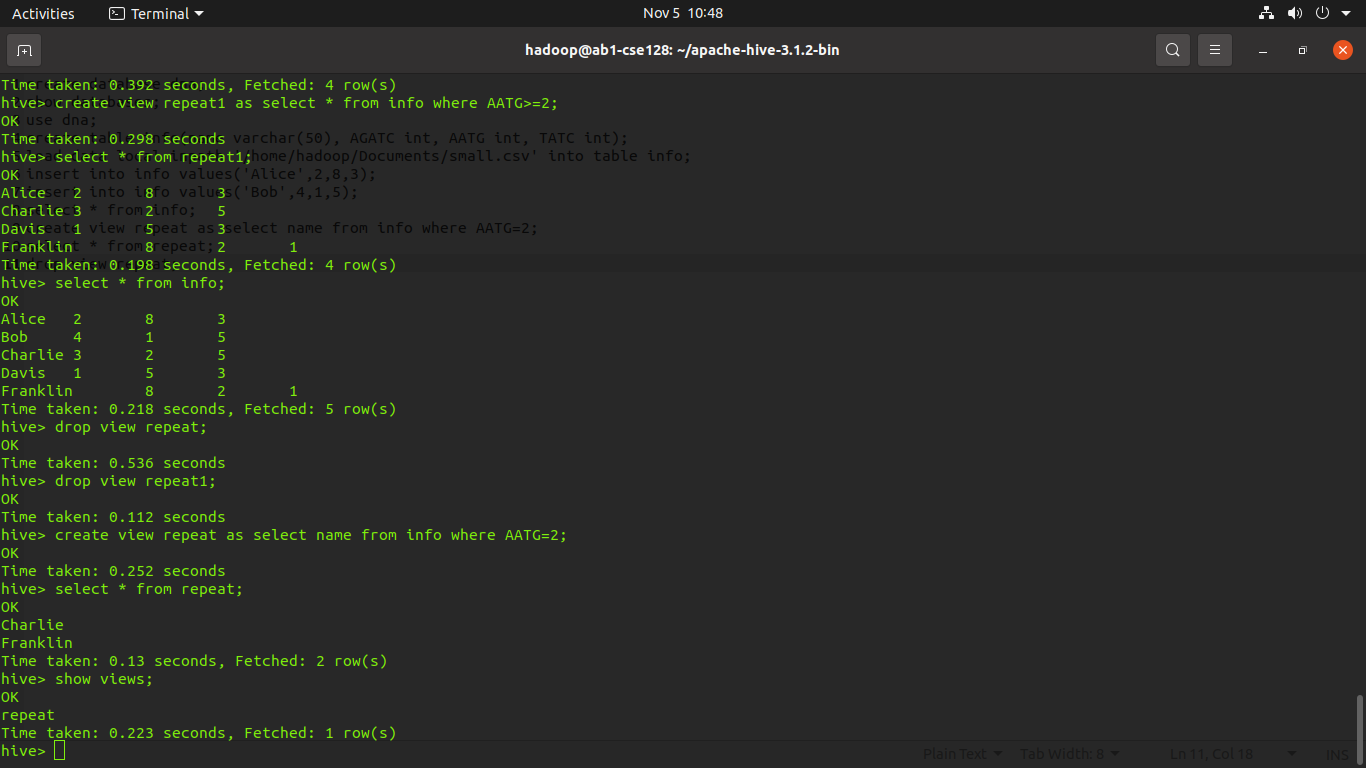
select \* from info;

create view repeat as select name from info where AATG=2;

select \* from repeat;

drop view repeat;





name AGATC AATG TATC  
Alice 2 8 3  
Bob 4 1 5  
Charlie 3 2 5

Sequencing DNA means determining the order of the four chemical building blocks - called "bases" - that make up the DNA molecule. The sequence tells scientists the kind of genetic information that is carried in a particular DNA segment.

*PIG COMMANDS*

hadoop fs -mkdir hdfs://localhost:9000/dna

hadoop fs -put '/home/hadoop/Downloads/small.csv' /dna

pig -x local

short\_dna = LOAD 'hdfs://localhost:9000/dna/small.csv' USING PigStorage(',') as (name:chararray,AGATC:int,AATG:int,TATC:int);

dump short\_dna;

explain short\_dna;

illustrate short\_dna;

filter\_data = FILTER short\_dna BY name=='Alice';

dump filter\_data;

order\_by\_data = ORDER short\_dna BY AATG DESC;

dump order\_by\_data;

