**Course Project**

**Surabhi Asati June 3rd, 2018**

**ITM – 6273**

**Part 1:**

To develop a Mapper and Reducer application to retrieve Year and Temperature from original NCDC records and then write the Year and Temperature data into a text file

**Commands:**

1. Copy folder - ProjectData from local to hdfs:

hadoop dfs -copyFromLocal ProjectData /home/student2

1. Listing the folder on hdfs:

hadoop dfs -ls /home/student2/ProjectData

1. Compile Java files:

javac -cp src/:hadoop-common-2.6.1.jar:hadoop-mapreduce-client-core-2.6.1.jar:commons-cli-2.0.jar -d . JobBuilder.java NcdcRecordParser.java YearTemperatureReducer.java YearTemperature.java YearTemperatureMapper.java

1. Creating final-project.jar file

jar -cvf final-project.jar YearTemperatureReducer.class YearTemperature.class JobBuilder.class NcdcRecordParser.class YearTemperatureMapper.class

1. Running the program to generate file with Year and Temperatures

hadoop jar final-project.jar YearTemperature /home/student2/ProjectData /home/student2/finalproject

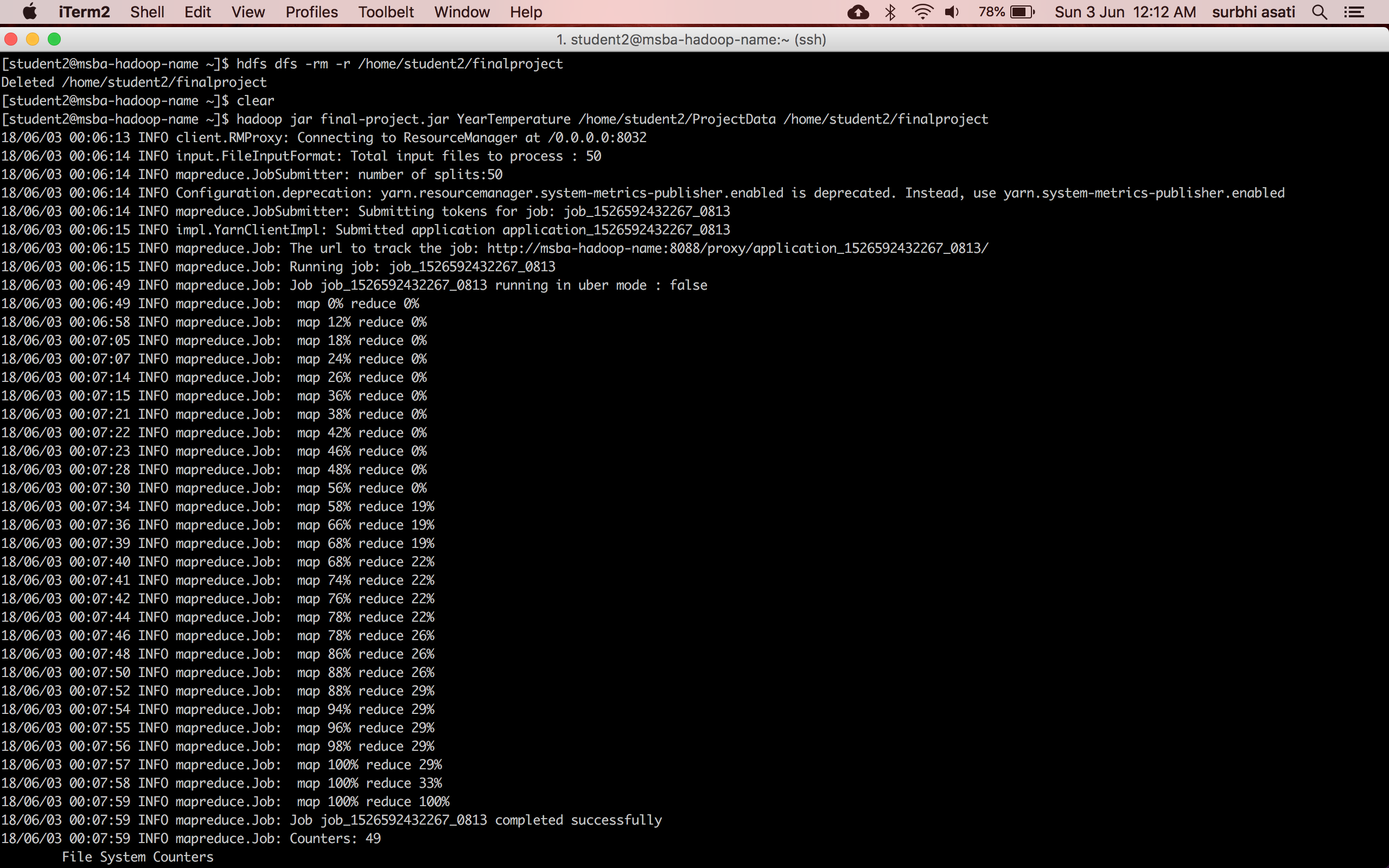
1. List the output of program

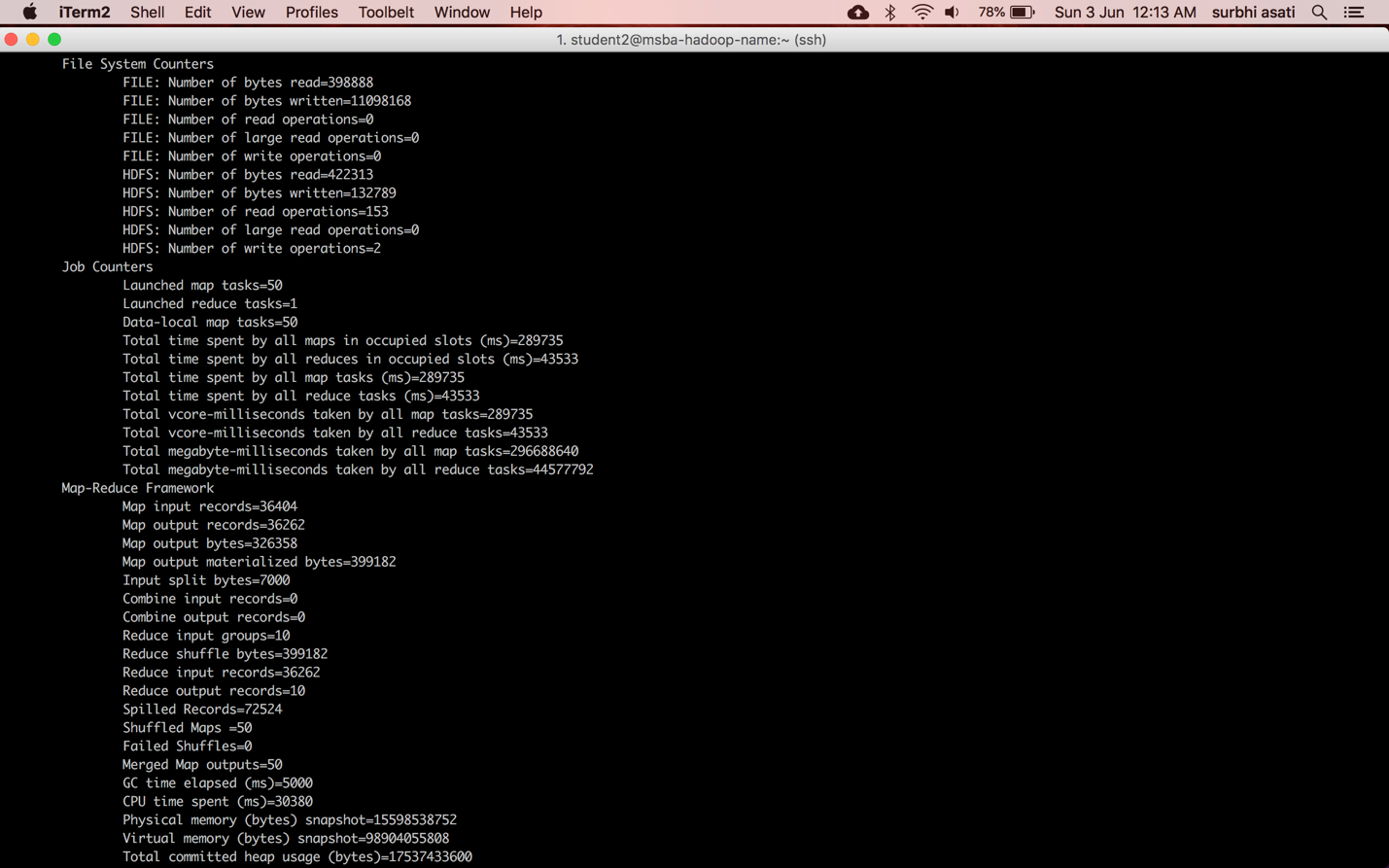
hadoop dfs -ls /home/student2/finalproject

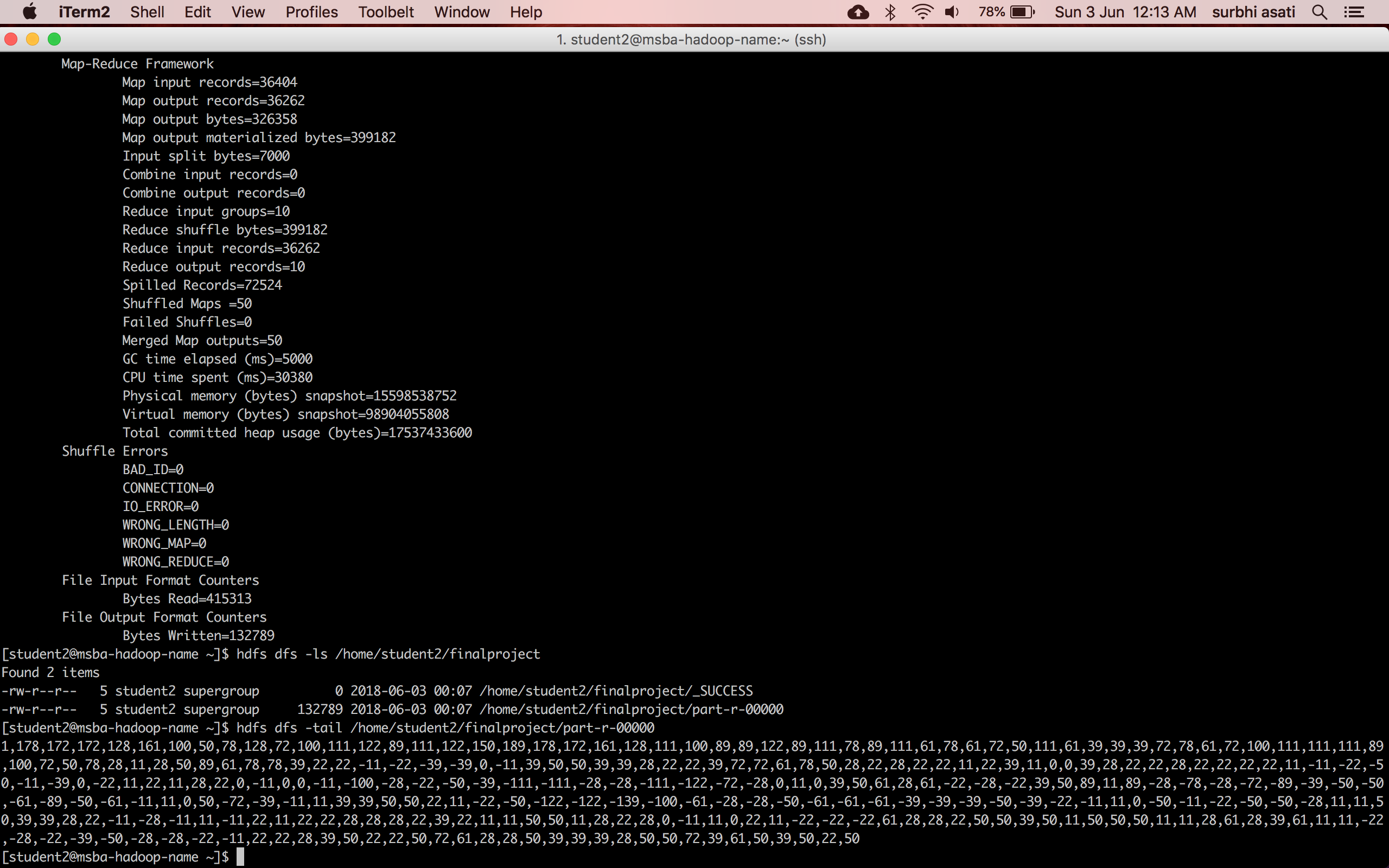
1. Display tail of the file (part-r-00000) with final result

hadoop dfs -tail /home/student2/finalproject/part-r-00000

**Final Screenshot:**







**Part 2:**

To load the text file into Pig and get the highest and lowest temperatures for each year.

**Commands:**

1. Loading data into pig

A = LOAD '/home/student2/finalproject/part-r-00000' USING PigStorage('\t') AS (year:chararray, temperatures:chararray);

1. Creating Bag data type containing year and temperature

B = FOREACH A GENERATE year, FLATTEN(TOKENIZE(temperatures)) as temperature;

1. Grouping by year

C = GROUP B BY year;

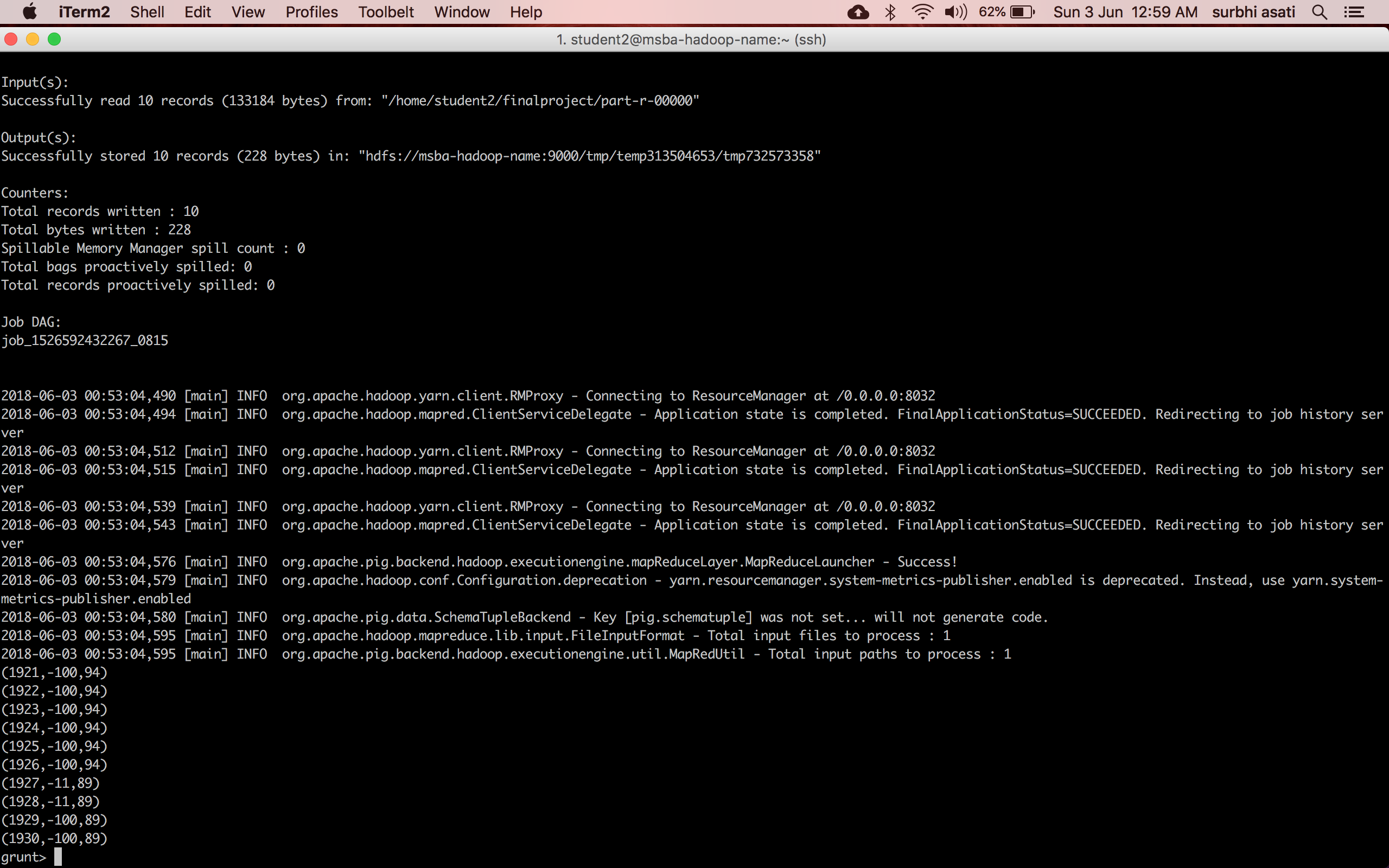
1. Calculating highest and lowest temperature for each year

D = FOREACH C GENERATE group, MIN(B.temperature), MAX(B.temperature);

1. Display Final results

DUMP D;

**The Final screenshot:**



**Part 3:**

To load the text file into Hive and get the average temperature for each year.

**Commands:**

1. Creating Table - FinalYearTemperature

create table FinalYearTemperature(year string, temp array<double>) row format delimited fields terminated by '\t' collection items terminated by ',';

1. Loading the text file into the table FinalYearTemperature

load data inpath '/home/student2/finalproject/part-r-00000' into table FinalYearTemperature;

1. Calculating the the average temperature for each year.

select

year,

avg(temperature)

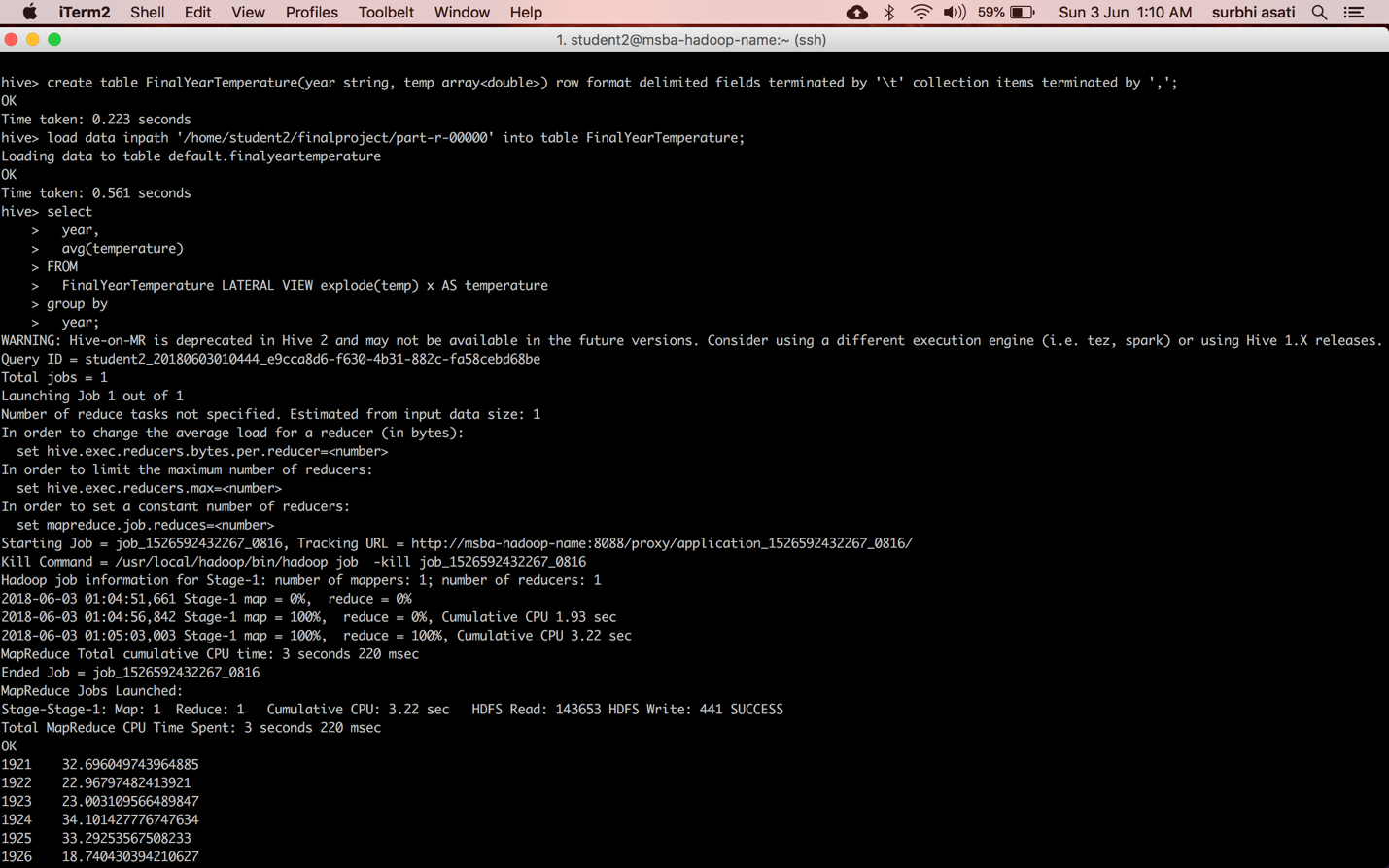
FROM

FinalYearTemperature LATERAL VIEW explode(temp) x AS temperature

group by

year;

**Screenshot – Final Results:**

****

