**AIM: To write a HIVE program to find average rainfall in INDIA (from given climate dataset)with Index.**

**PROGRAM:**

CREATE DATABASE climate\_one;

CREATE TABLE climate\_one.cli\_tab (station string,month string,span string,num int,max\_temp float,min\_temp float,rainfall float) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

LOAD DATA LOCAL INPATH '/home/hduser/hive/data.csv' OVERWRITE INTO TABLE climate\_one.cli\_tab;

CREATE INDEX climate\_index ON TABLE climate\_one.cli\_tab (rainfall) AS 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler' WITH DEFERRED REBUILD;

ALTER INDEX climate\_index on climate\_one.cli\_tab REBUILD;

SELECT AVG(rainfall) FROM climate\_one.cli\_tab;

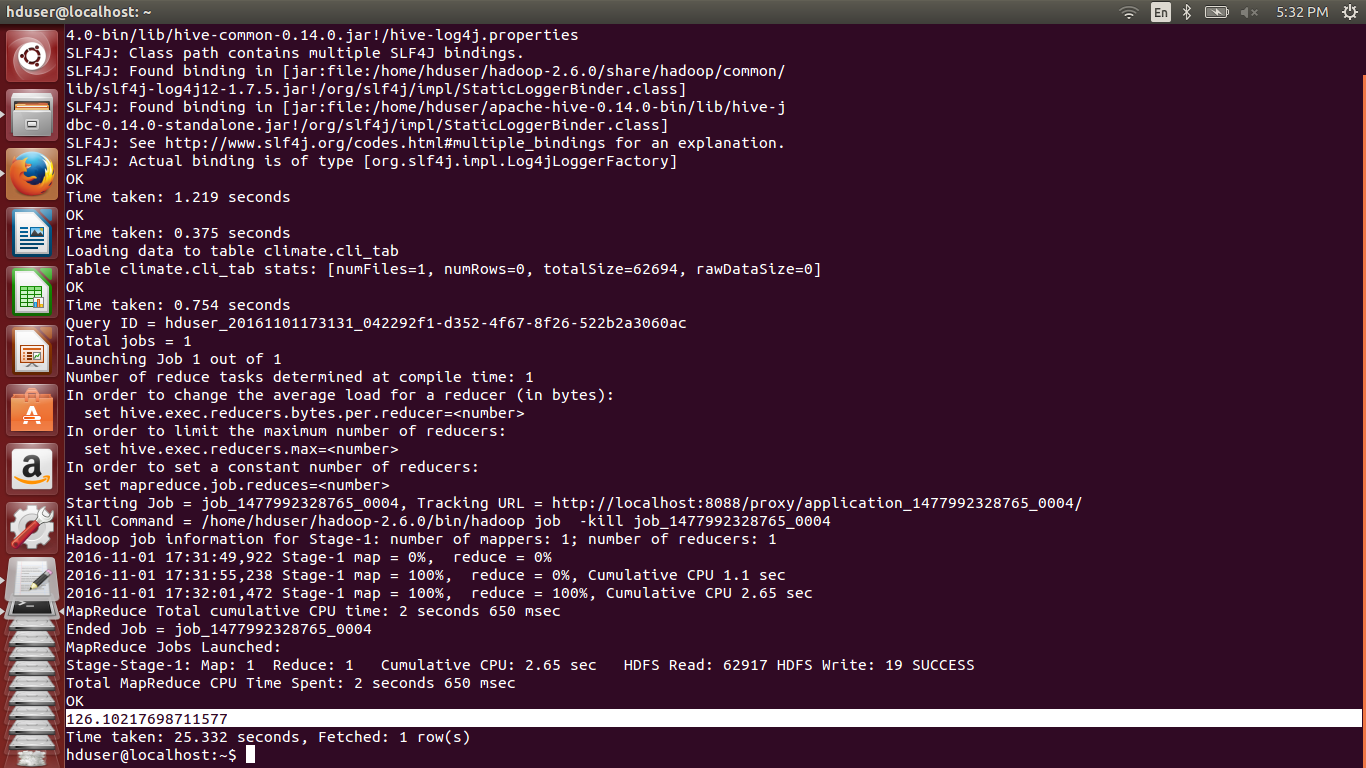
**INPUT:**

climate dataset

**OUTPUT:**

The time taken to caluclate average rainfall with indexes(21.815 sec) is less than the time taken to caluclate average without index i.e 25.332 sec

without index:



with index:

