Assignment 9.5:

1. Explain the running modes of pig and integrate Tez with Hadoop and run a simple pig query using tez

Running Modes of Pig:

Pig has two modes for running scripts: a.)Local Mode and b.)MapReduce Mode:

Local Mode:

All scripts are run on a single machine without requiring Hadoop MapReduce and HDFS. This can be useful for developing and testing Pig logic. If we are using a small set of data to develop or test our code, then local mode could be much quicker than going through the MapReduce infrastructure. Local mode doesn’t need Hadoop.

Mapreduce Mode:

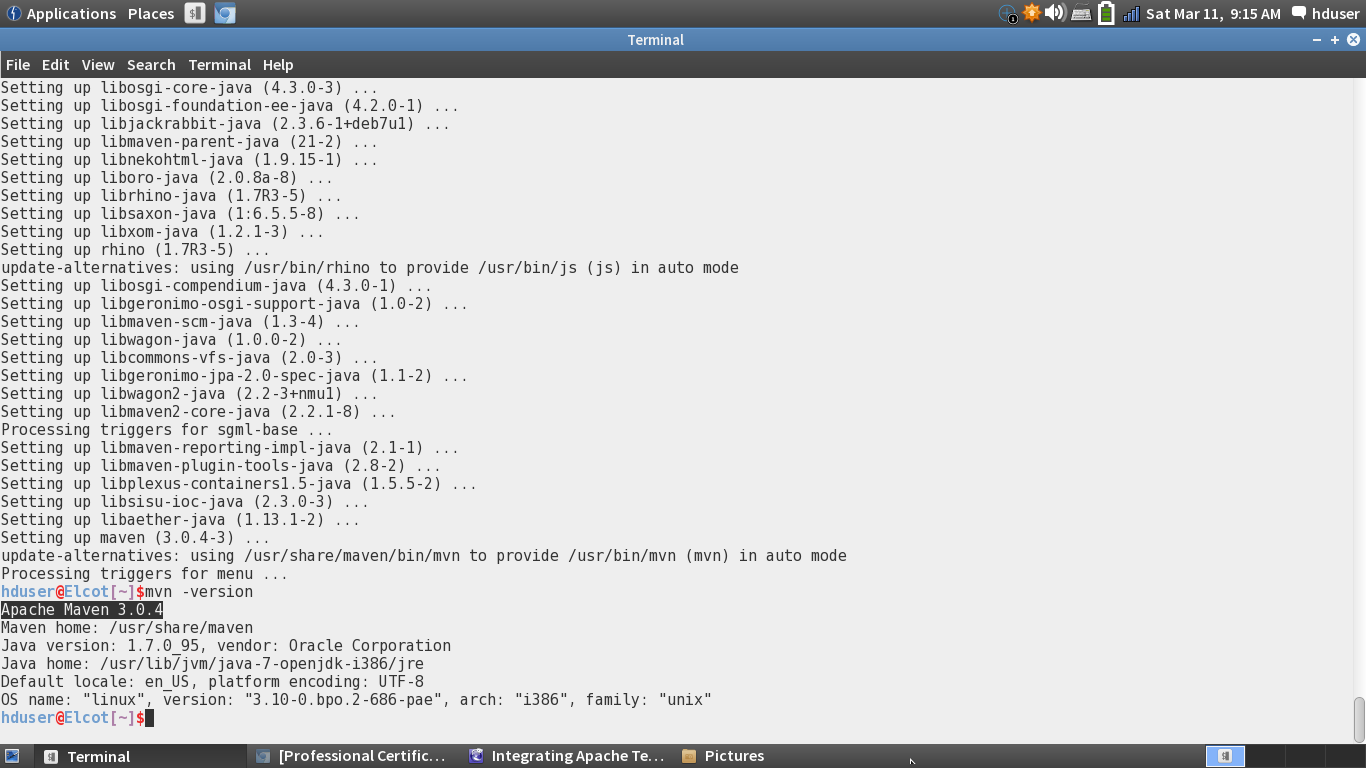
Pig always runs on the Hadoop cluster. In this mode, the Pig script gets translated into a series of MapReduce tasks that are then execute on the Hadoop cluster.

Integrating Tez with Hadoop:

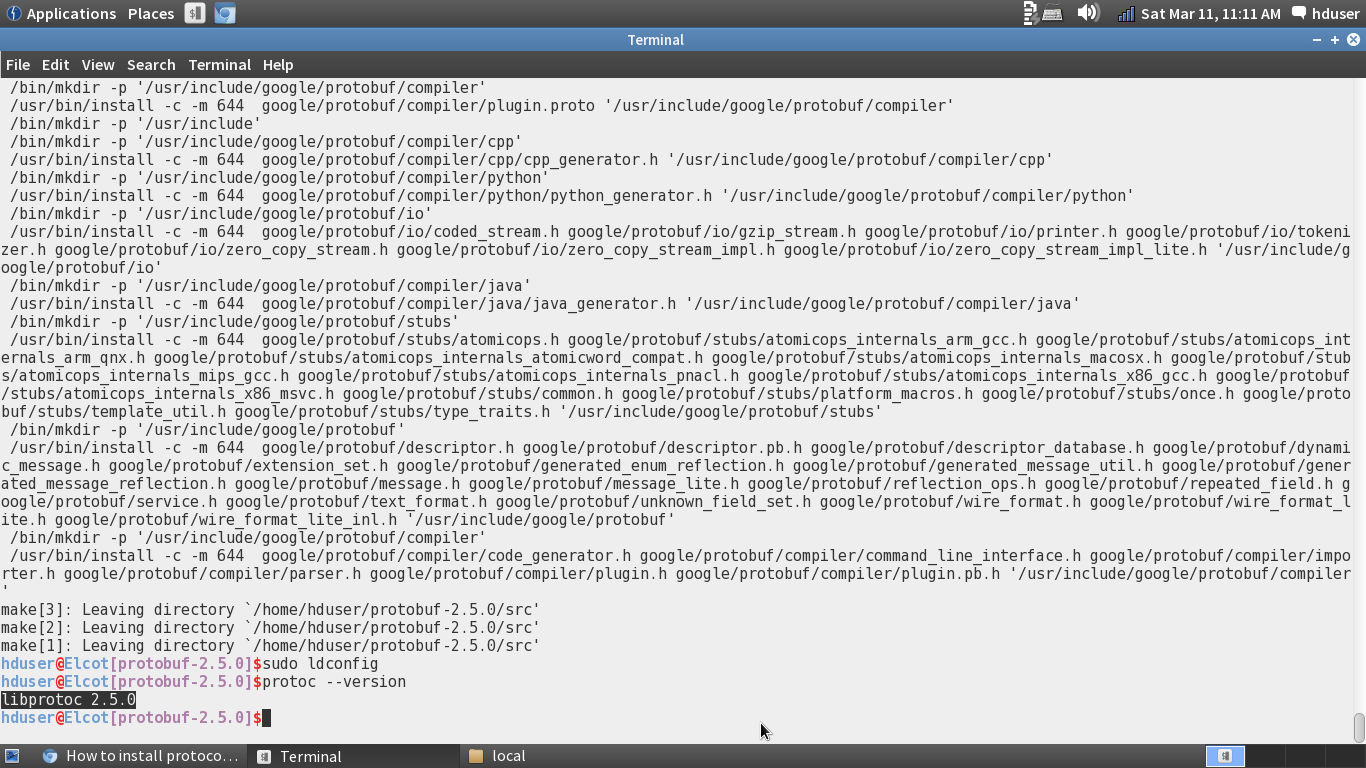
I referred this blog for the for the Integration of Tez With Hadoop:

https://acadgild.com/blog/integrating-apache-tez-with-hadoop/

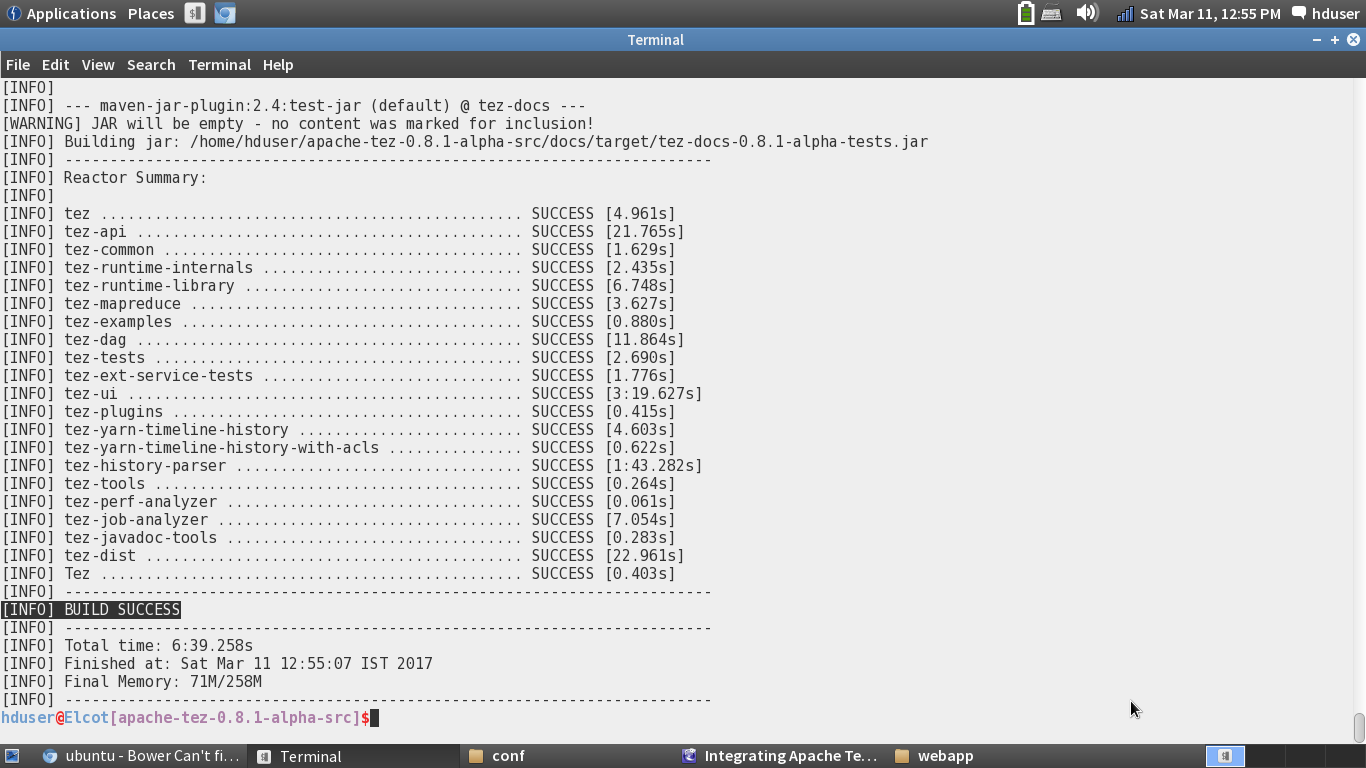
Step 1: Installing Maven



Step 2: Installing Protocol Buffer:



Step 3: Tez successfully build:



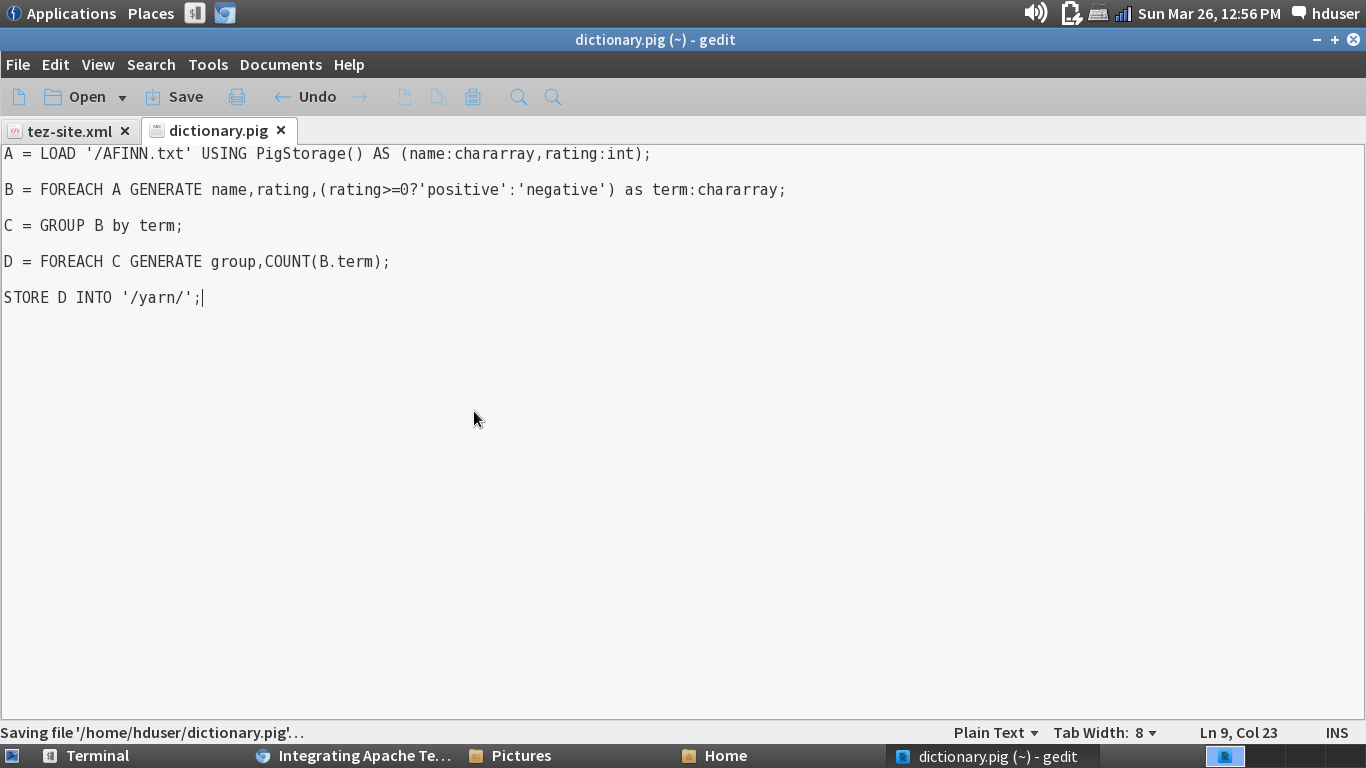
Running a pig query using Tez and performance analysis:

I referred to this blog for running the pig query <https://acadgild.com/blog/performance-analysis-of-tez/>.

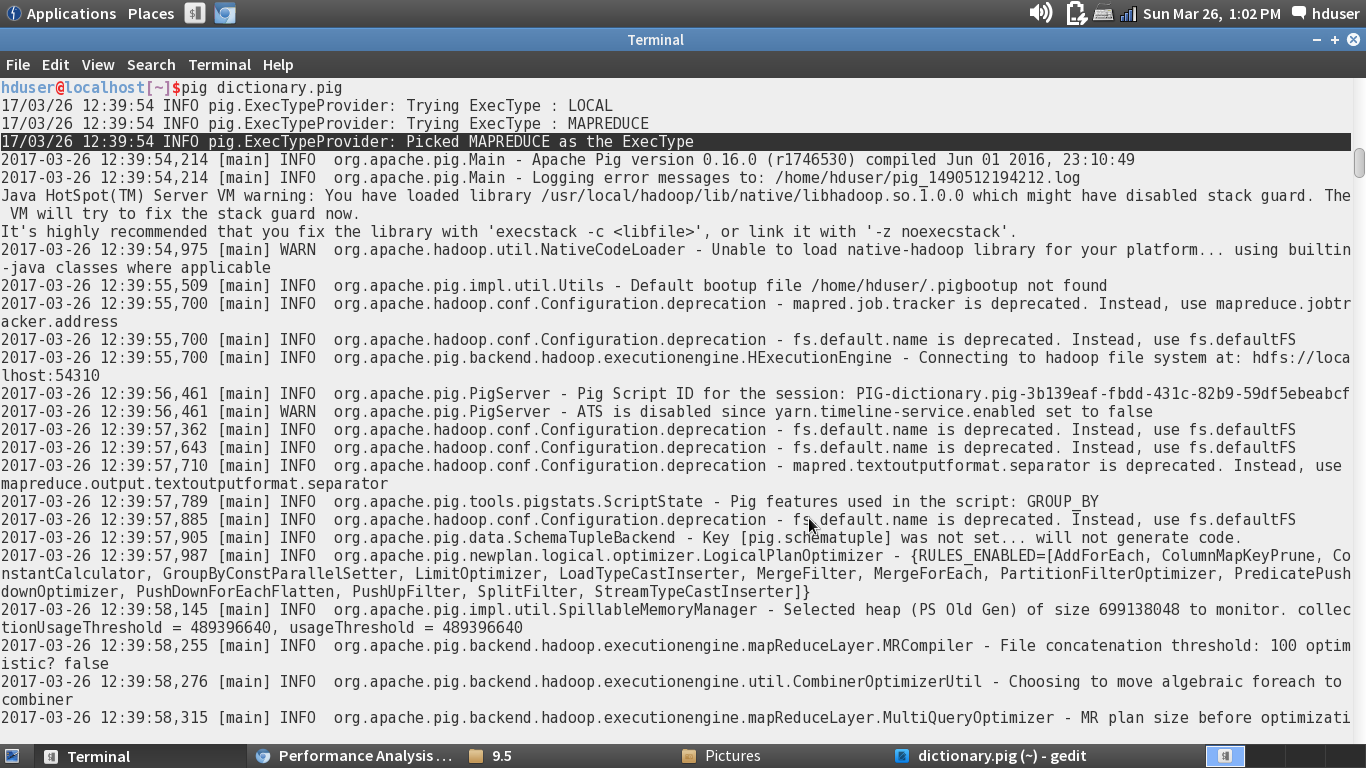
I wrote a Pig Script for a dictionary called AFINN, in which 2477 words are rated from -5 to +5 based on the words meaning. In the Pig script I counted how many positive words (0-5) and negative words (-5 to -1) are there.

First, I ran the script on the Mapreduce mode and after that I ran it on the Tez mode. And I compared the Time took for each mode to complete.

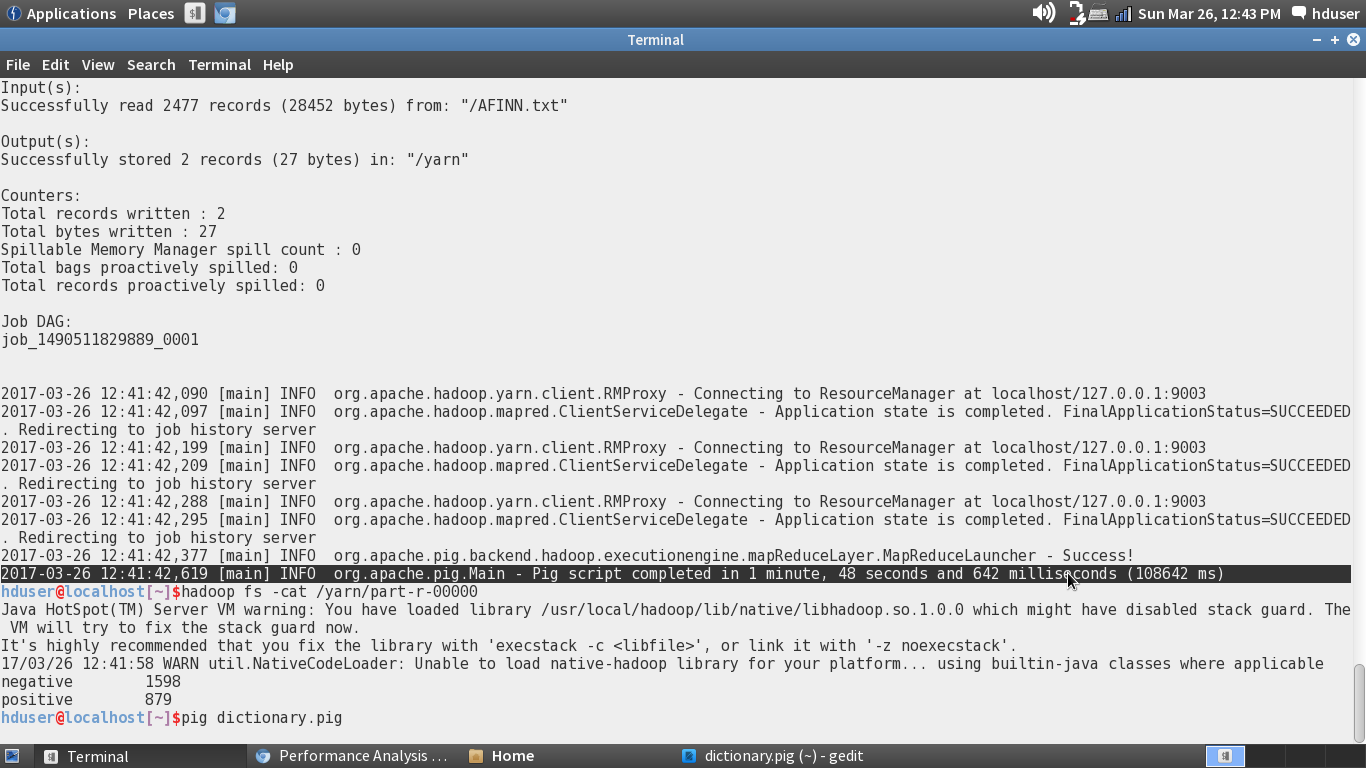
Yarn Pig script:



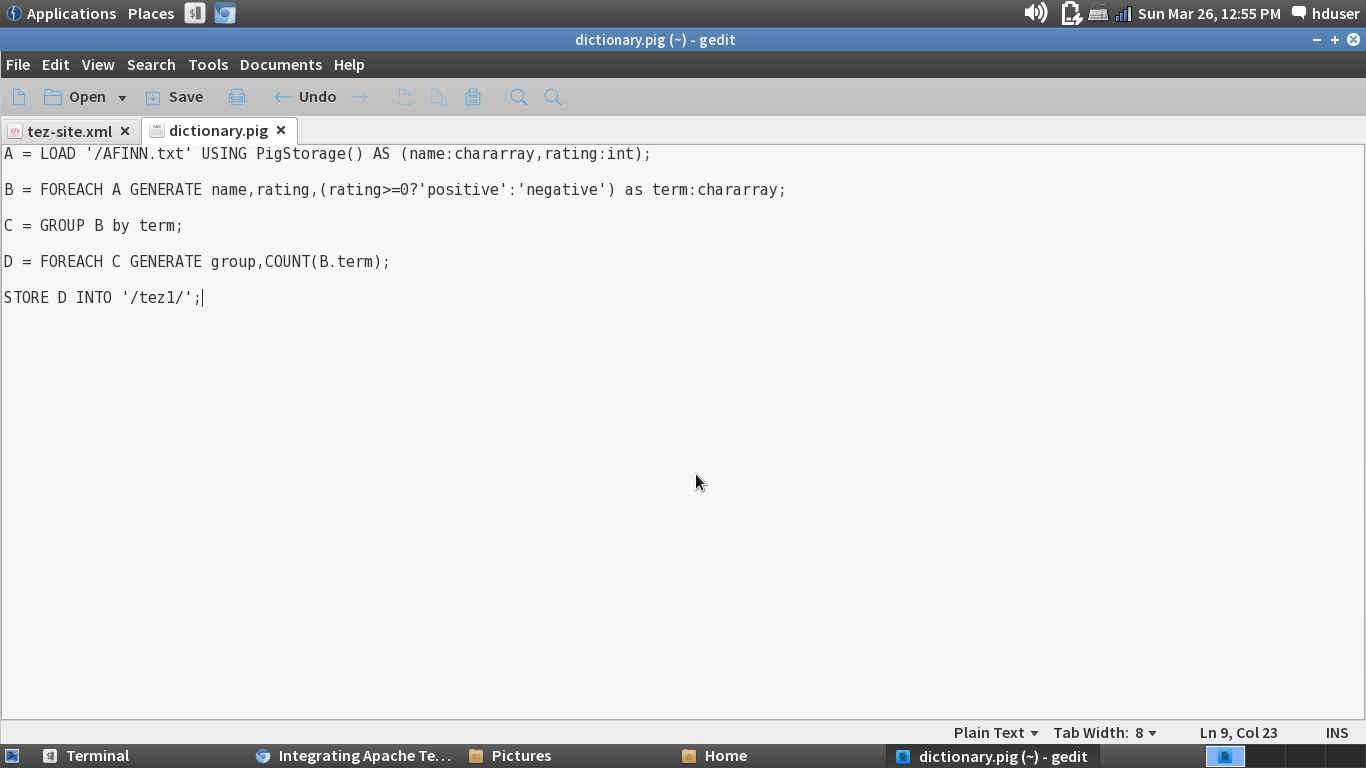
Running the pig script:



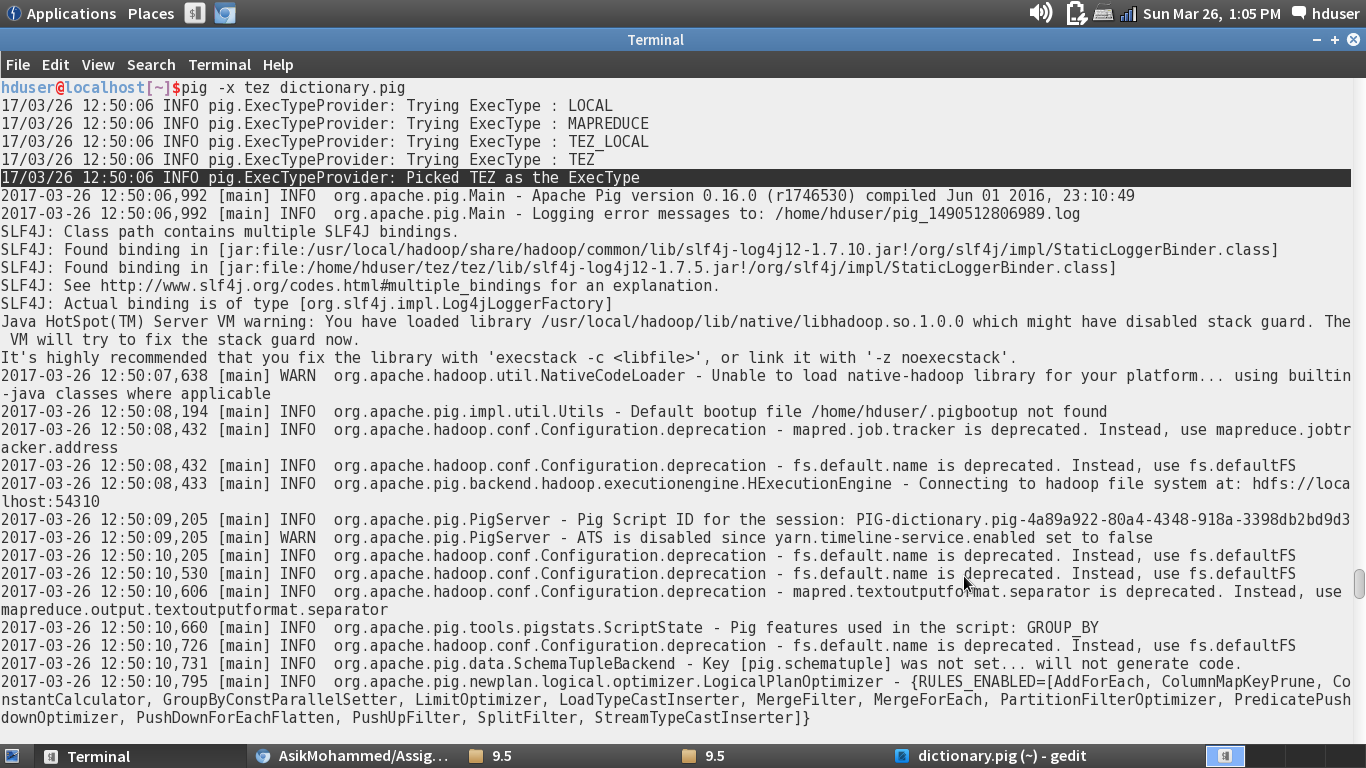
Time took for Yarn to complete the pig script and the Output of the script:



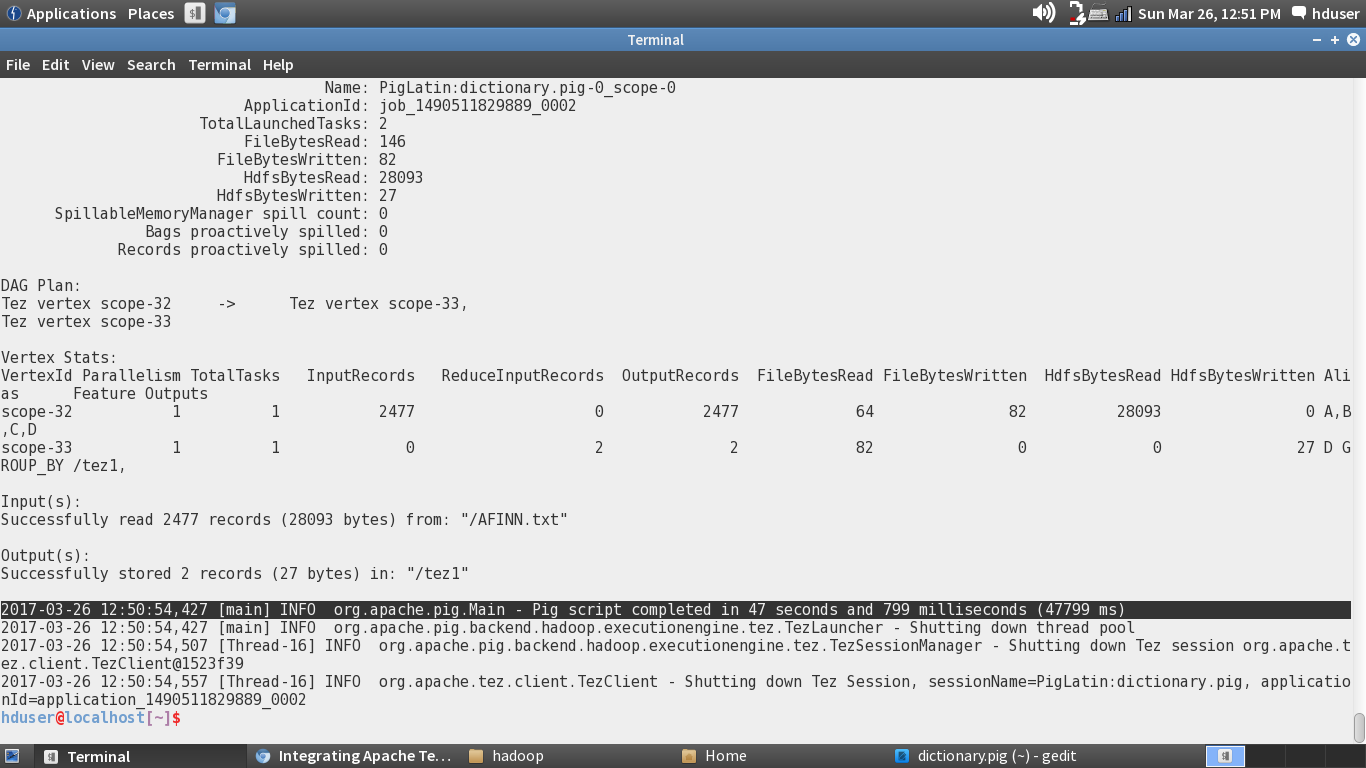
Tez Pig script:



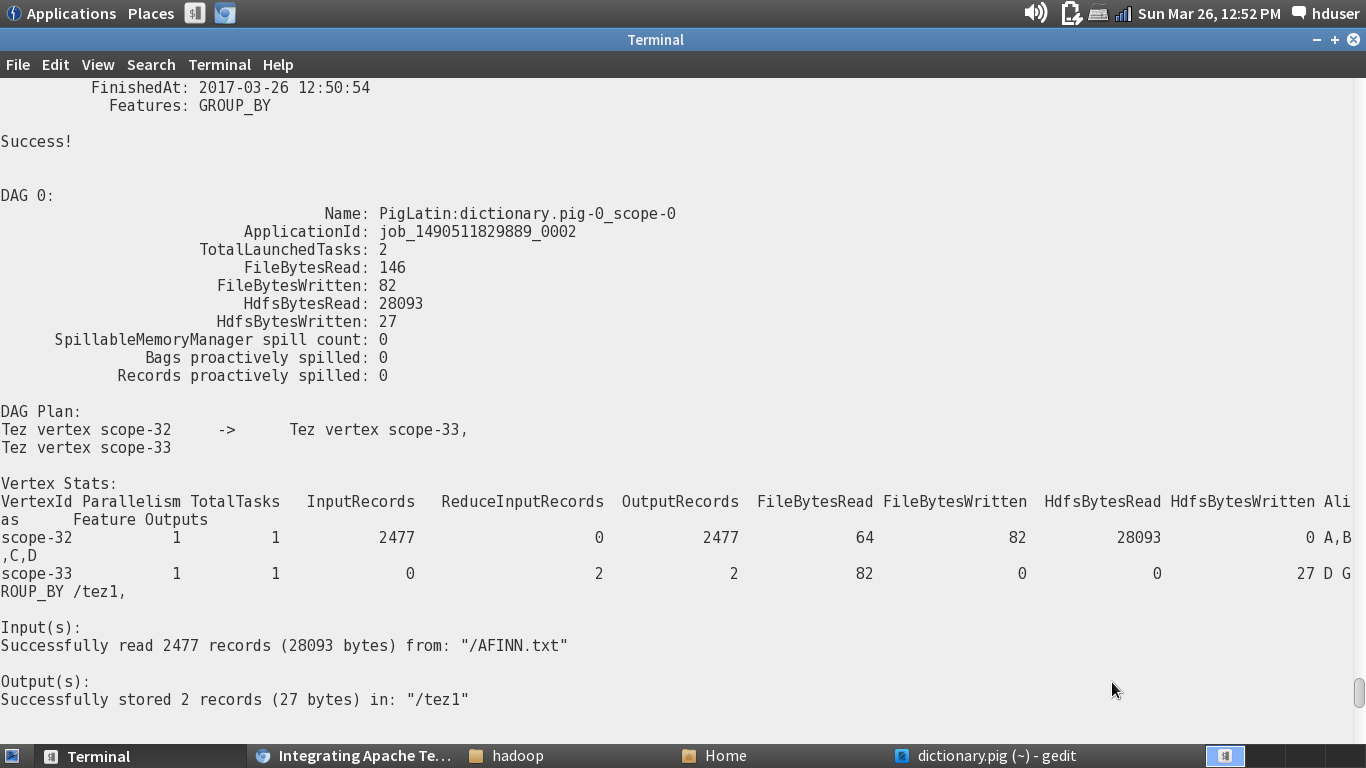
Executing the Tez Pig script:

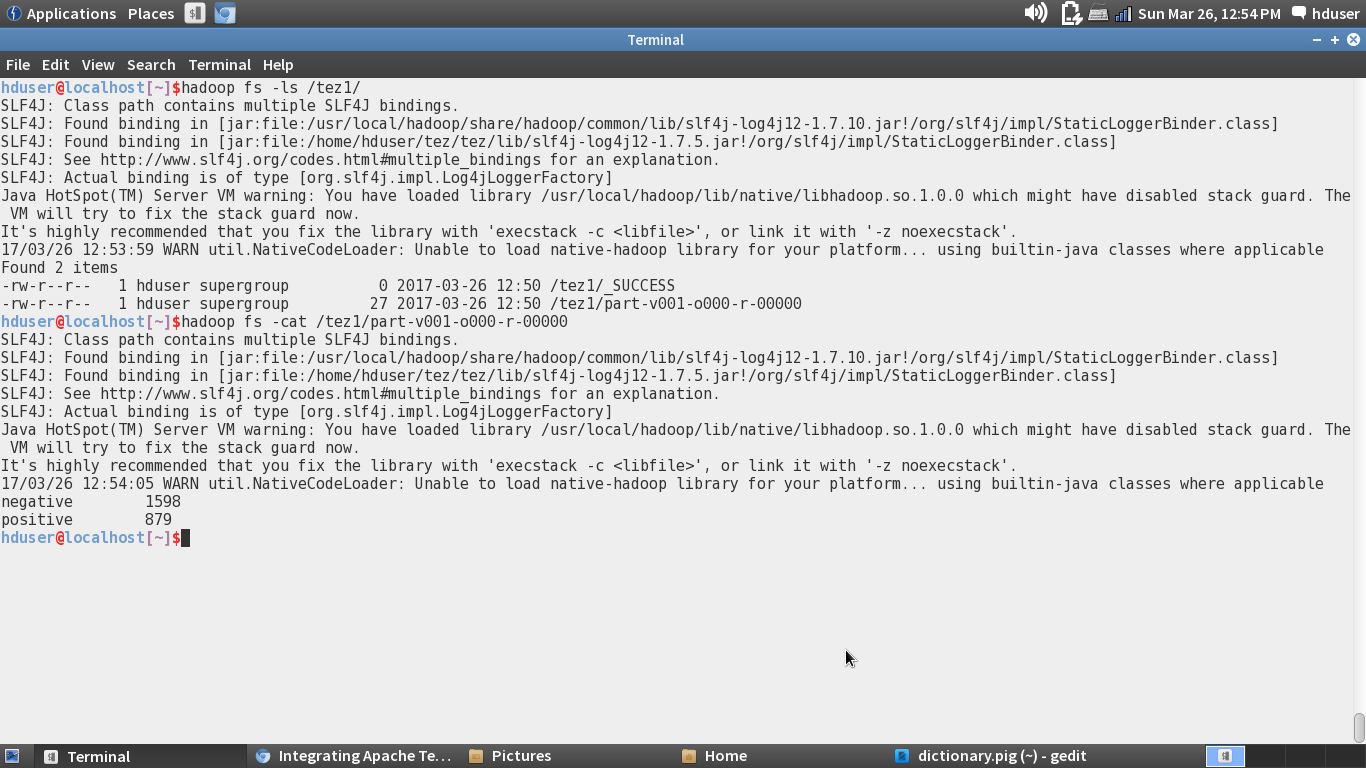


Time took for the Tez Pig script to complete:



Output of Tez:





Conclusion:

As we can see that the Time took for yarn to complete is around 1 minute and 48 seconds and for tez its only 47 seconds.