**Assignment 9.5**

**Problem Statement:**

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

You can refer the below blog for integrating tez with hadoop

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

You can refer the below blog to know how to run pig queries using tez <https://acadgild.com/blog/performance-analysis-of-tez/>

**Answer:**

**Running modes of pig:**

Pig has two modes for running scripts:

* Local mode
* MapReduce mode(Hadoop 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 you’re using a small set of data to developer or test your code, then local mode could be faster than going through the MapReduce infrastructure.

 To run Pig in local mode, you need access to a single machine; all files are installed and run using your local host and file system. Specify local mode using the -x flag (pig -x local). Note that local mode does not support parallel mapper execution with Hadoop 0.20.x and 1.0.0. This is because the LocalJobRunner of these Hadoop versions is not thread-safe.

**MapReduce mode (Hadoop mode):**

Pig is executed on the Hadoop cluster. In this case, the Pig Script gets converted into a series of MapReduce jobs that are then run on the Hadoop cluster.

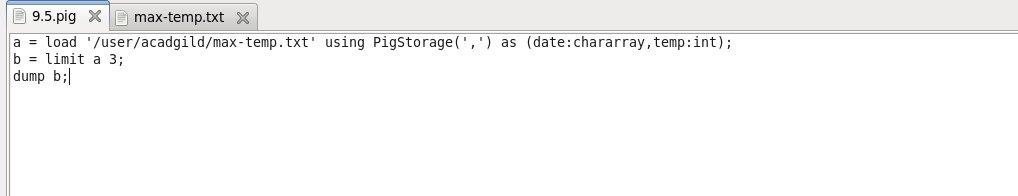
To run Pig in mapreduce mode, you need access to a Hadoop cluster and HDFS installation. Mapreduce mode is the default mode; you can, *but don't need to*, specify it using the -x flag (pig OR pig -x mapreduce).

You can run Pig in either mode using the "pig" command (the bin/pig Perl script) or the "java" command (java -cp pig.jar ...).

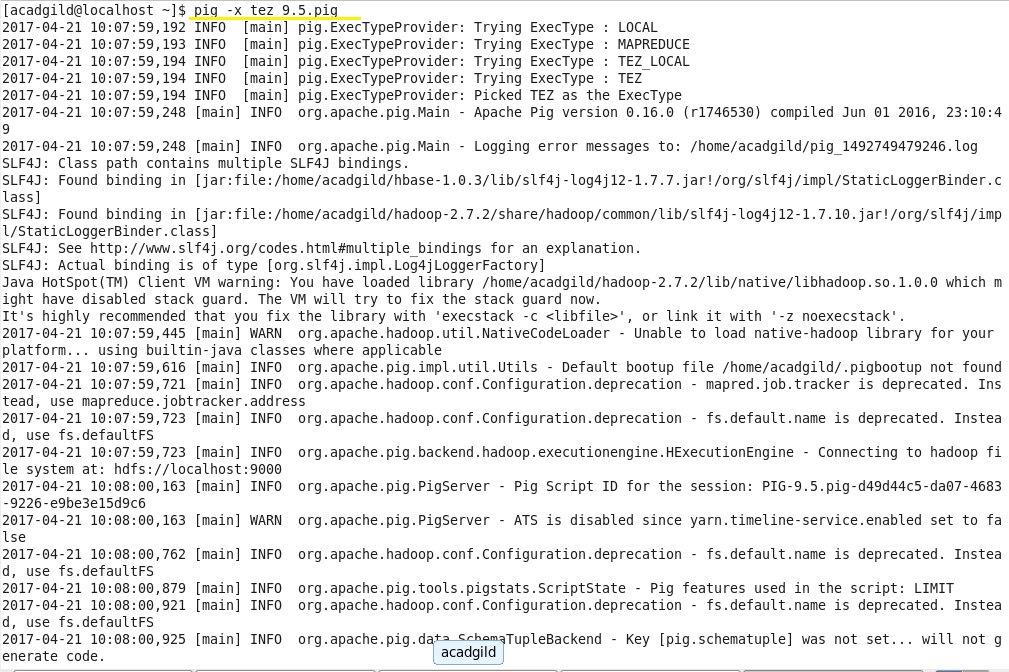
**Input:**

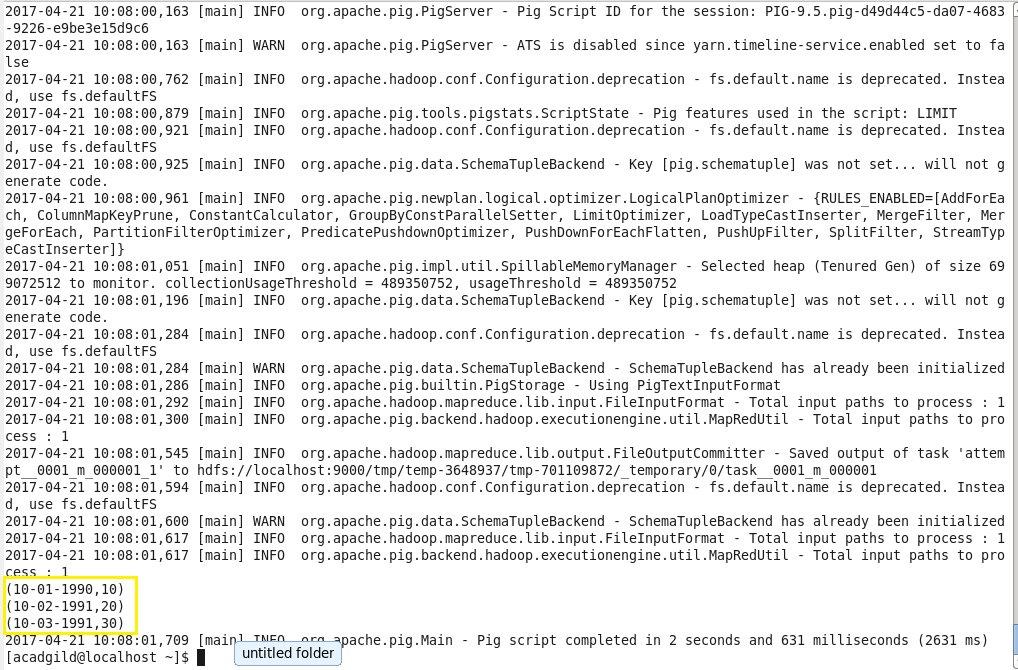
****

**Pig Script:**

****

**Output:**

****

****