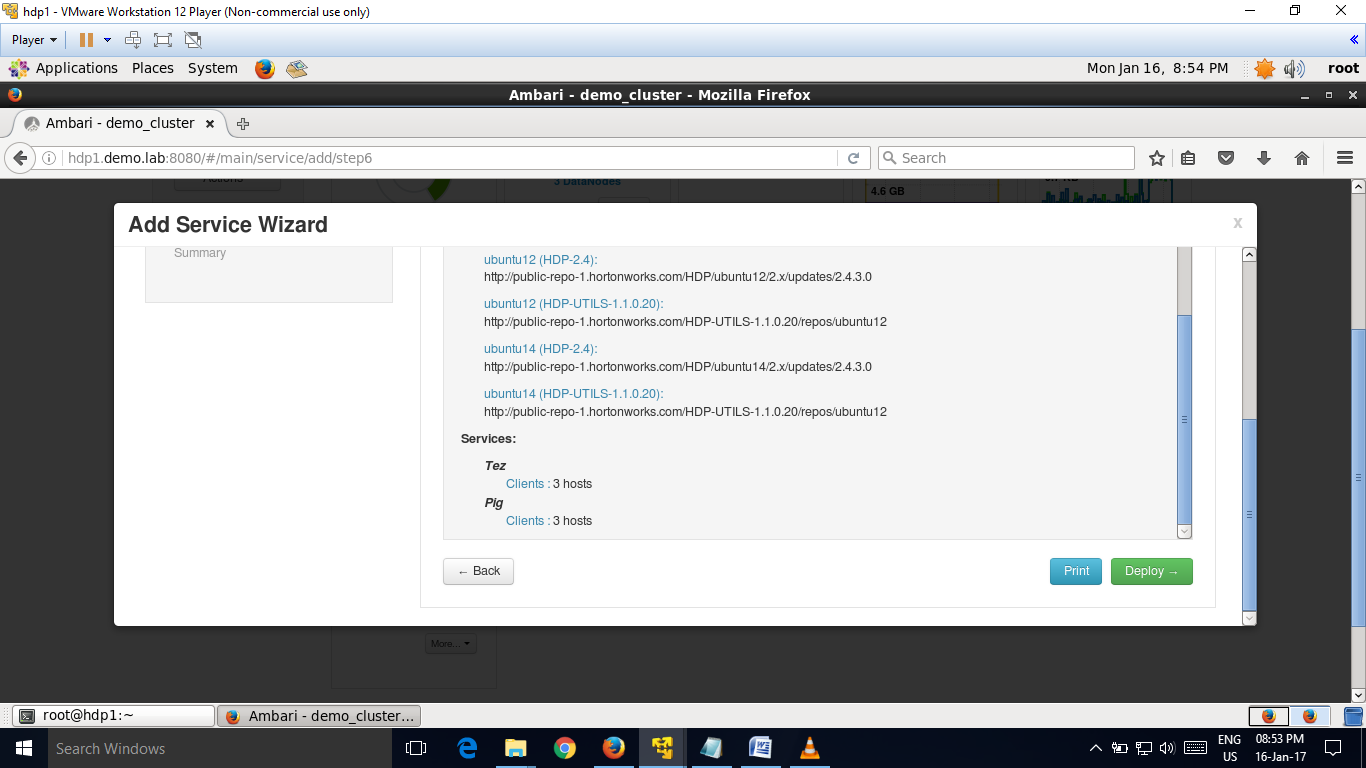
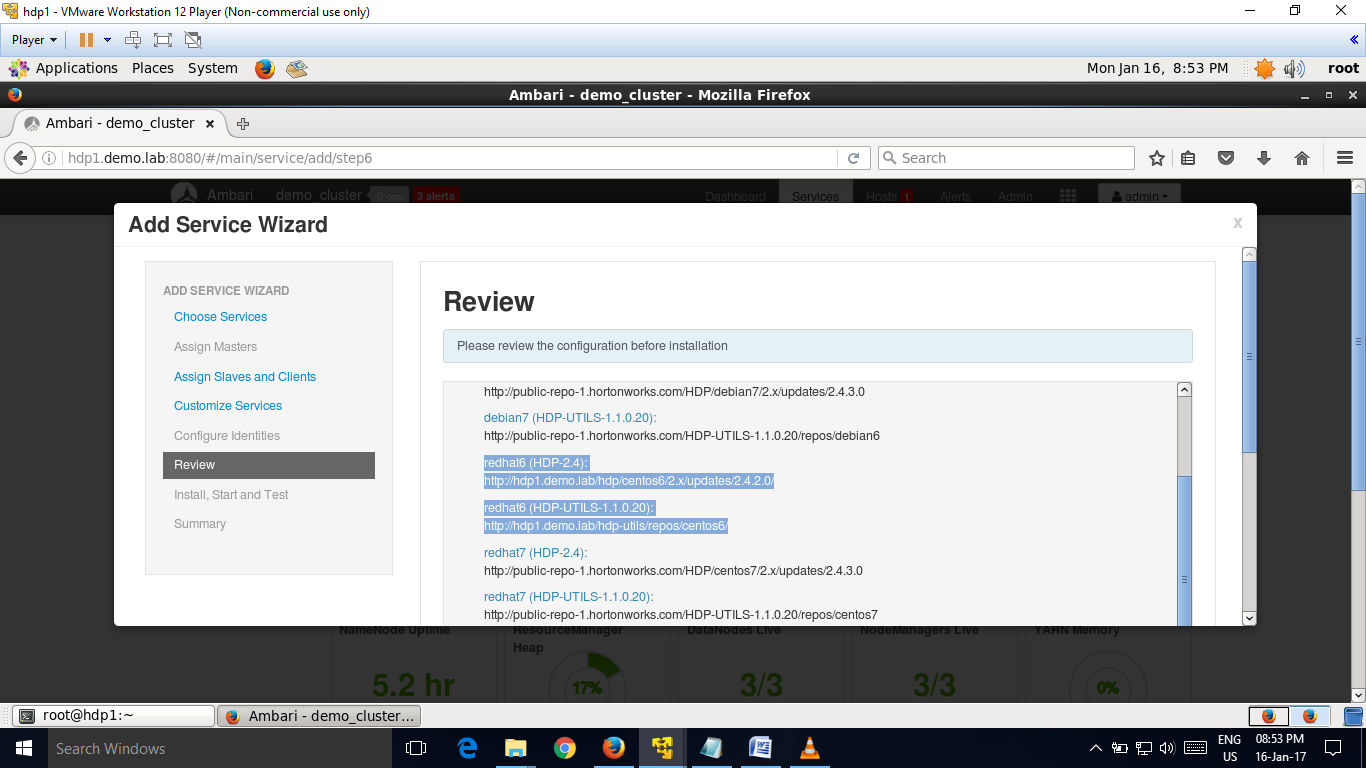
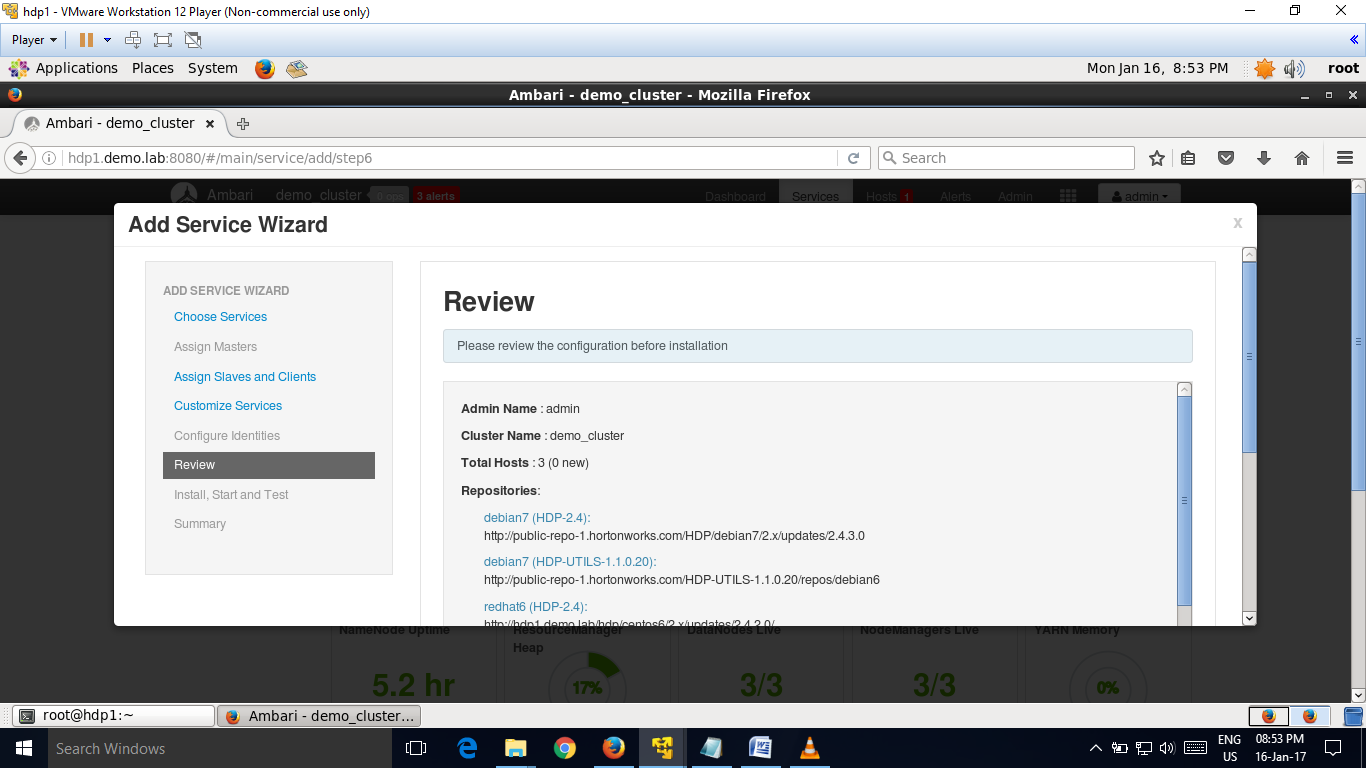
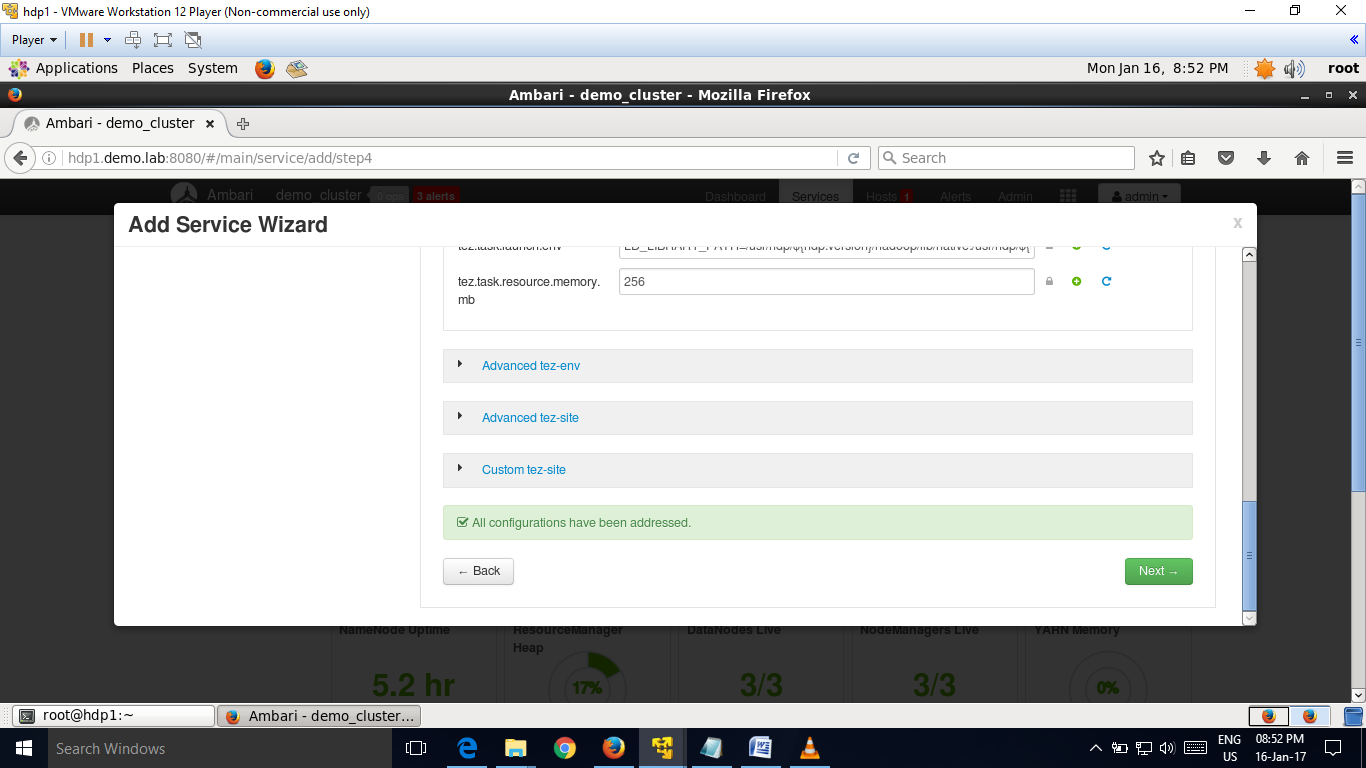
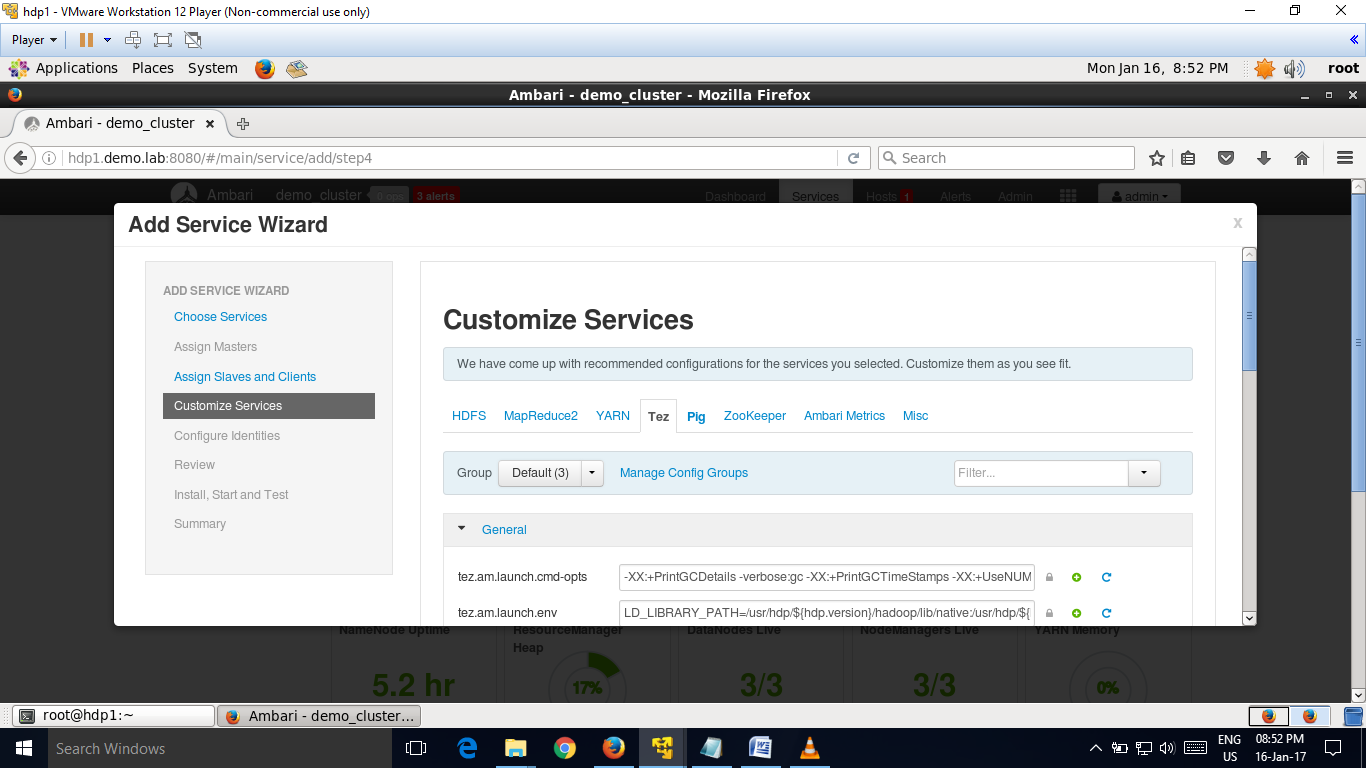
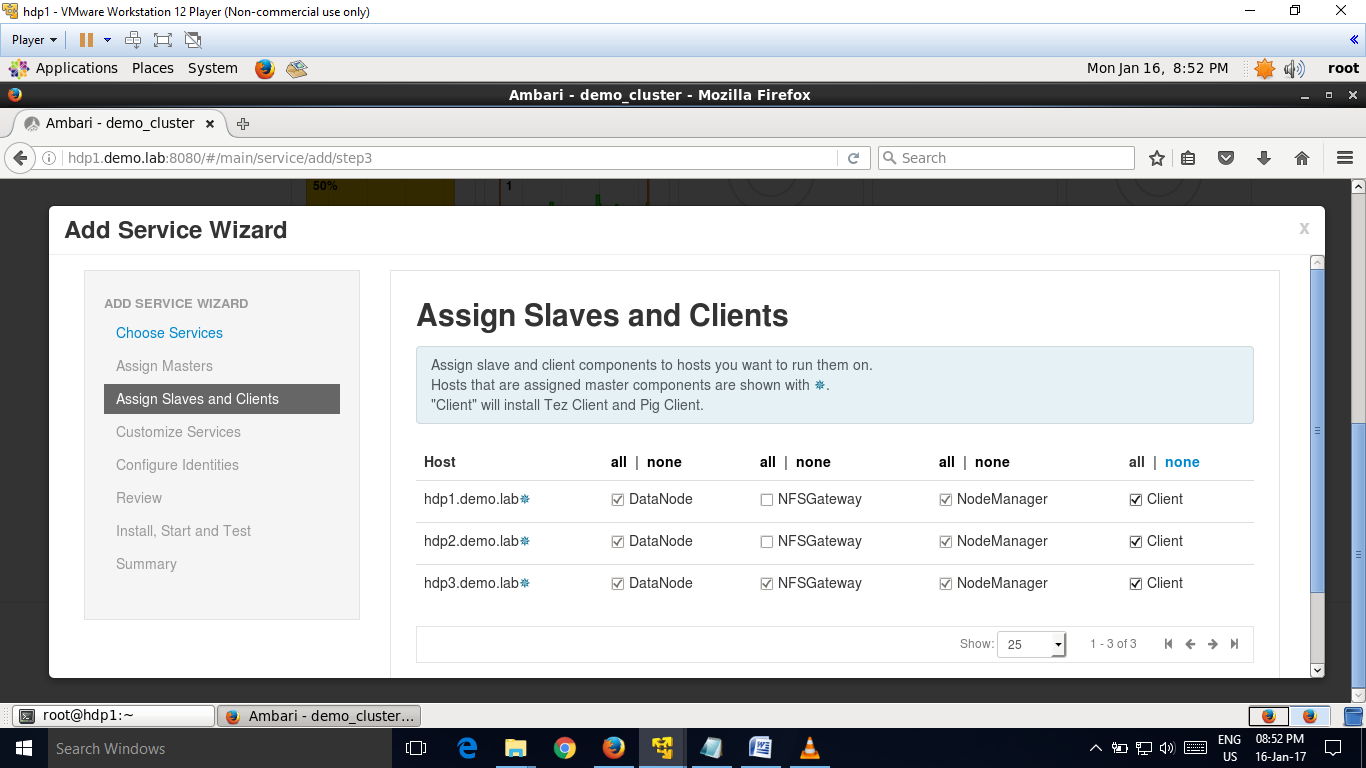
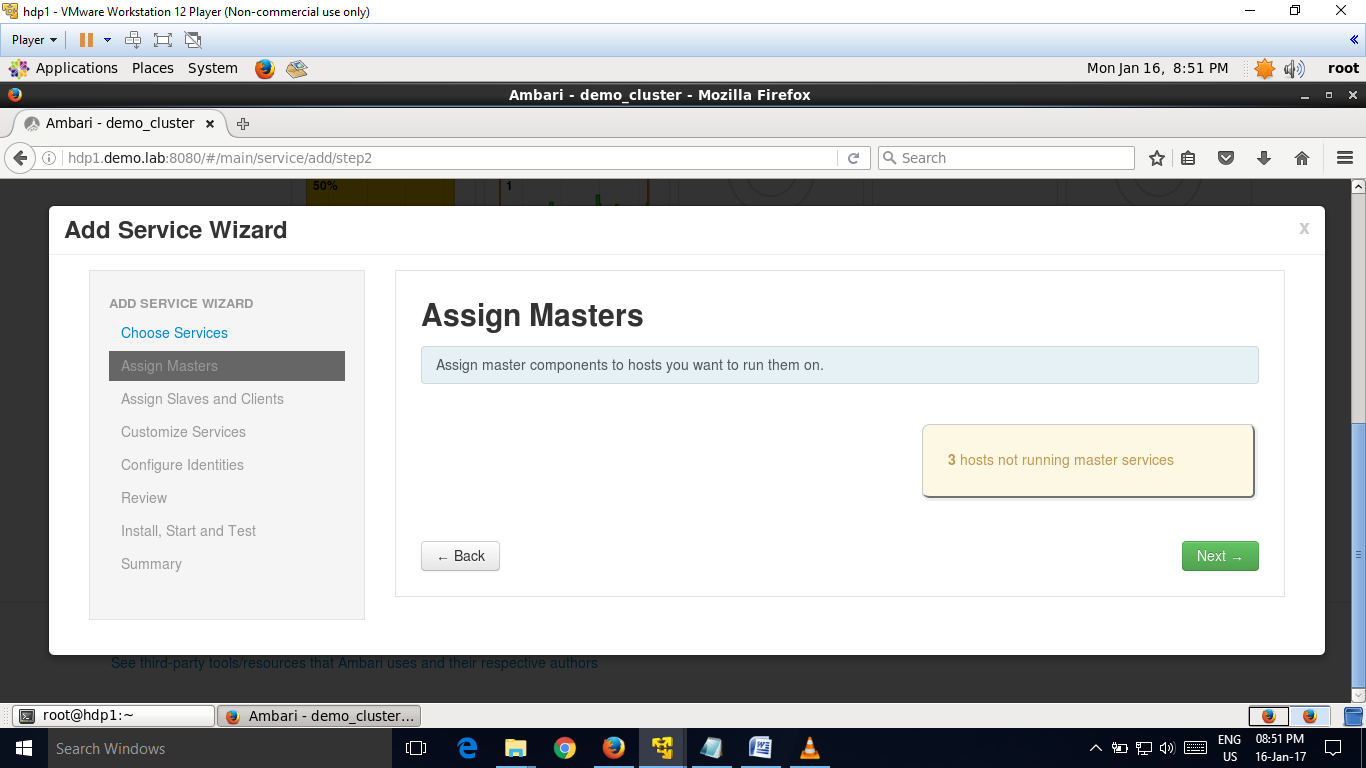
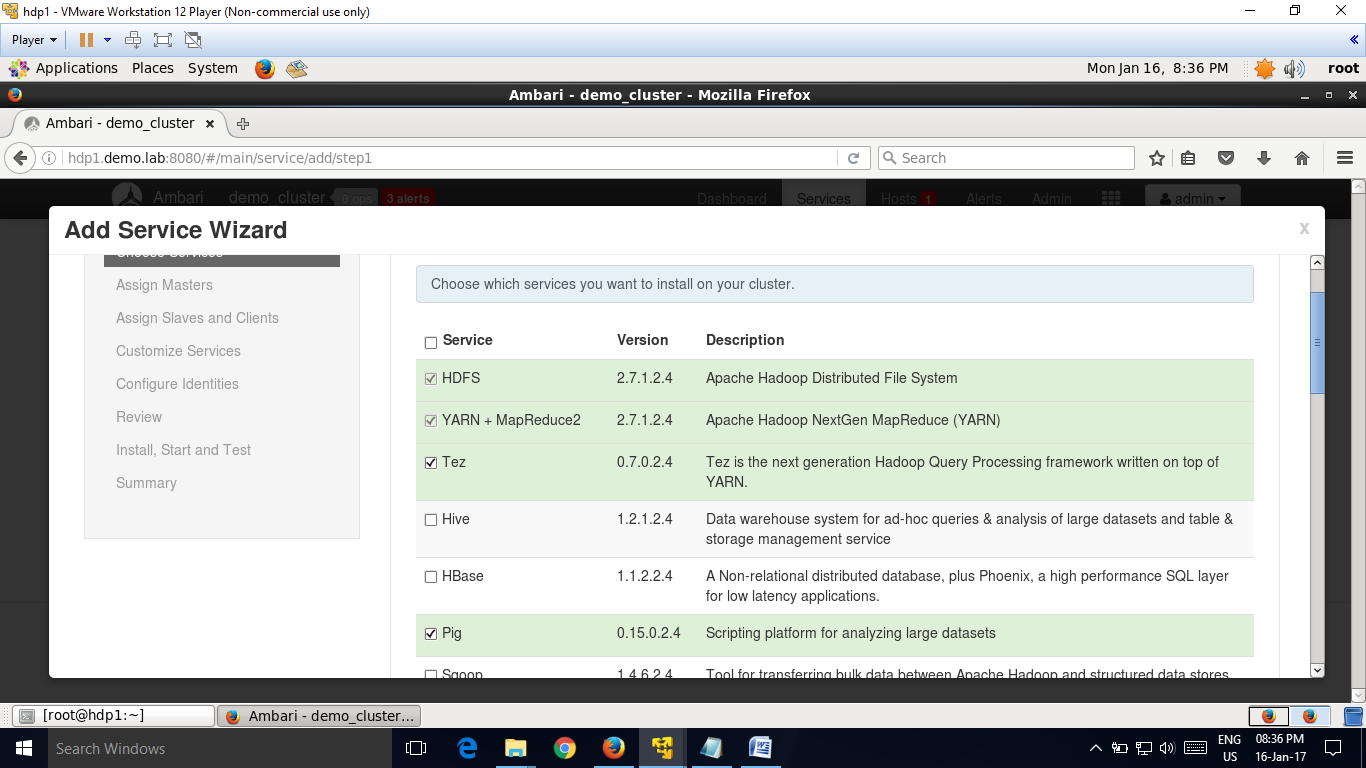
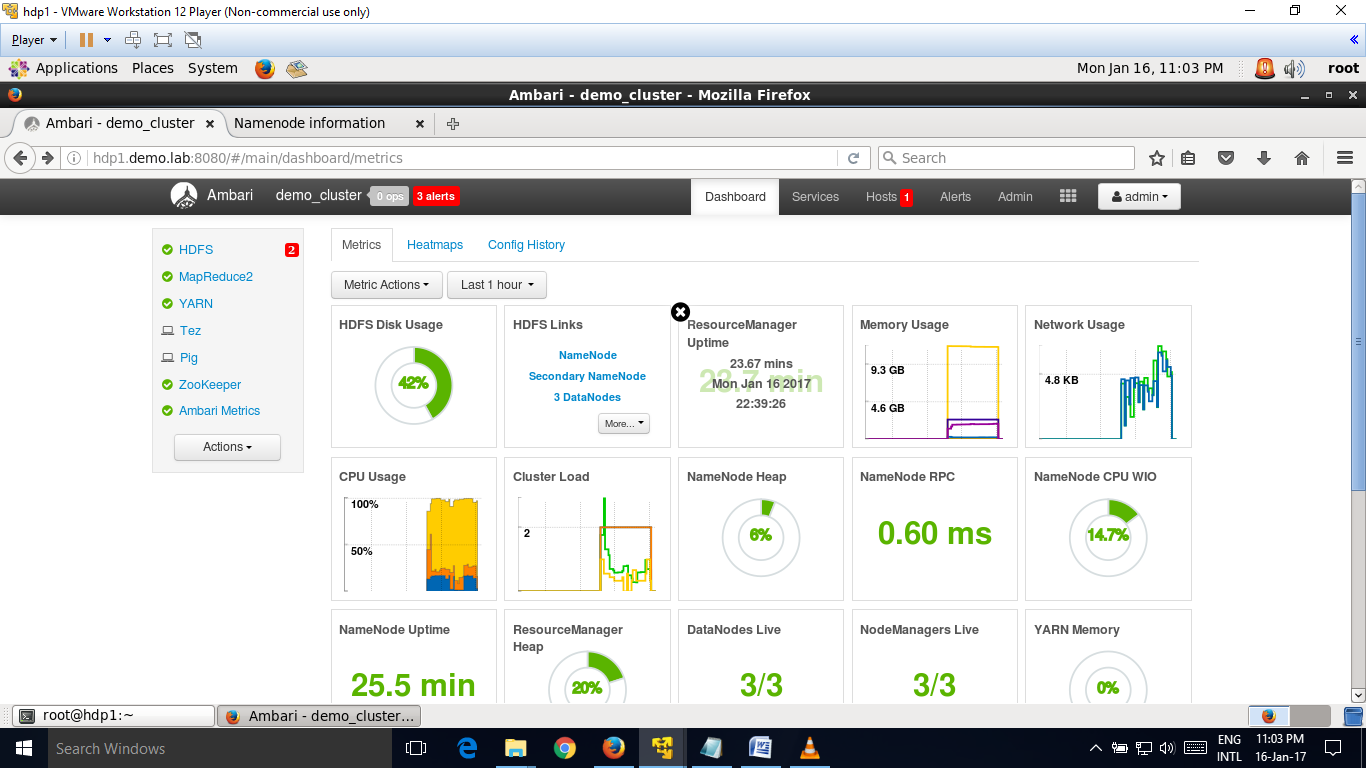
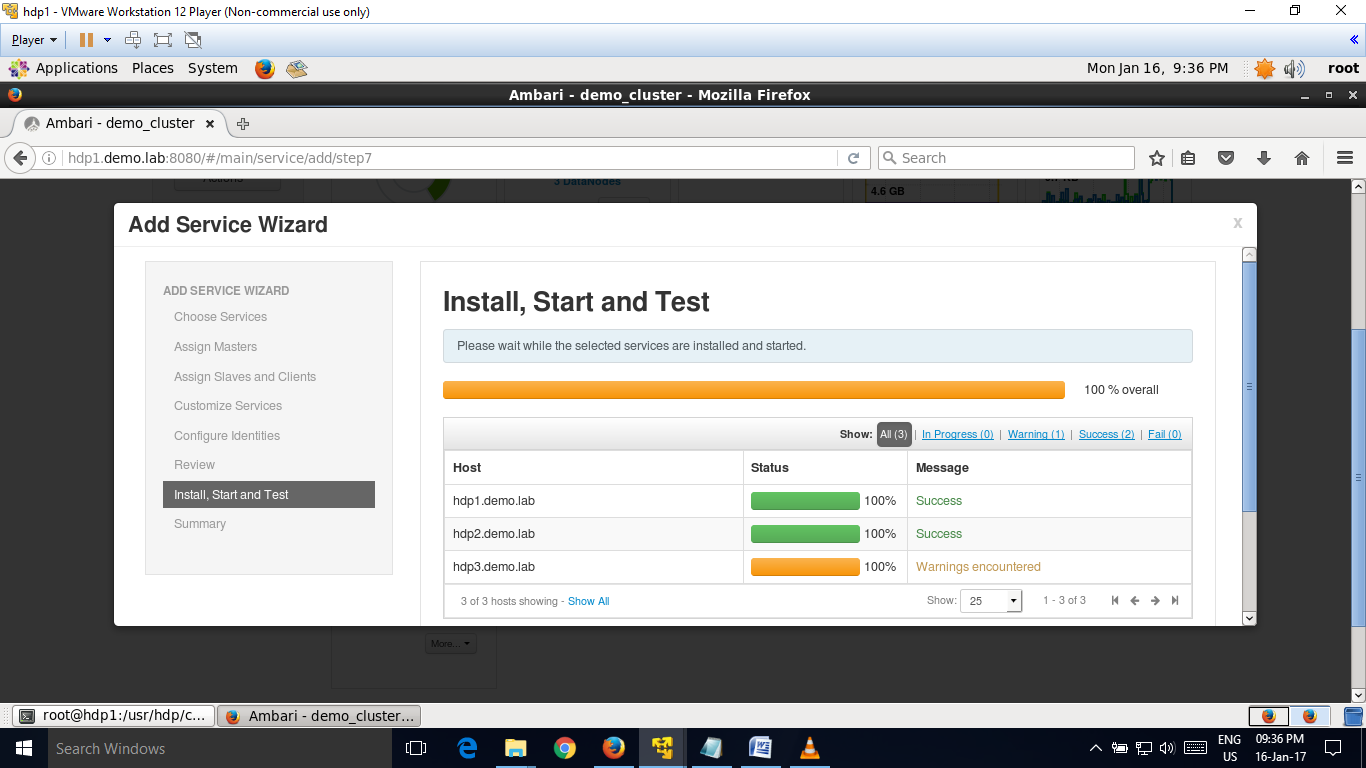
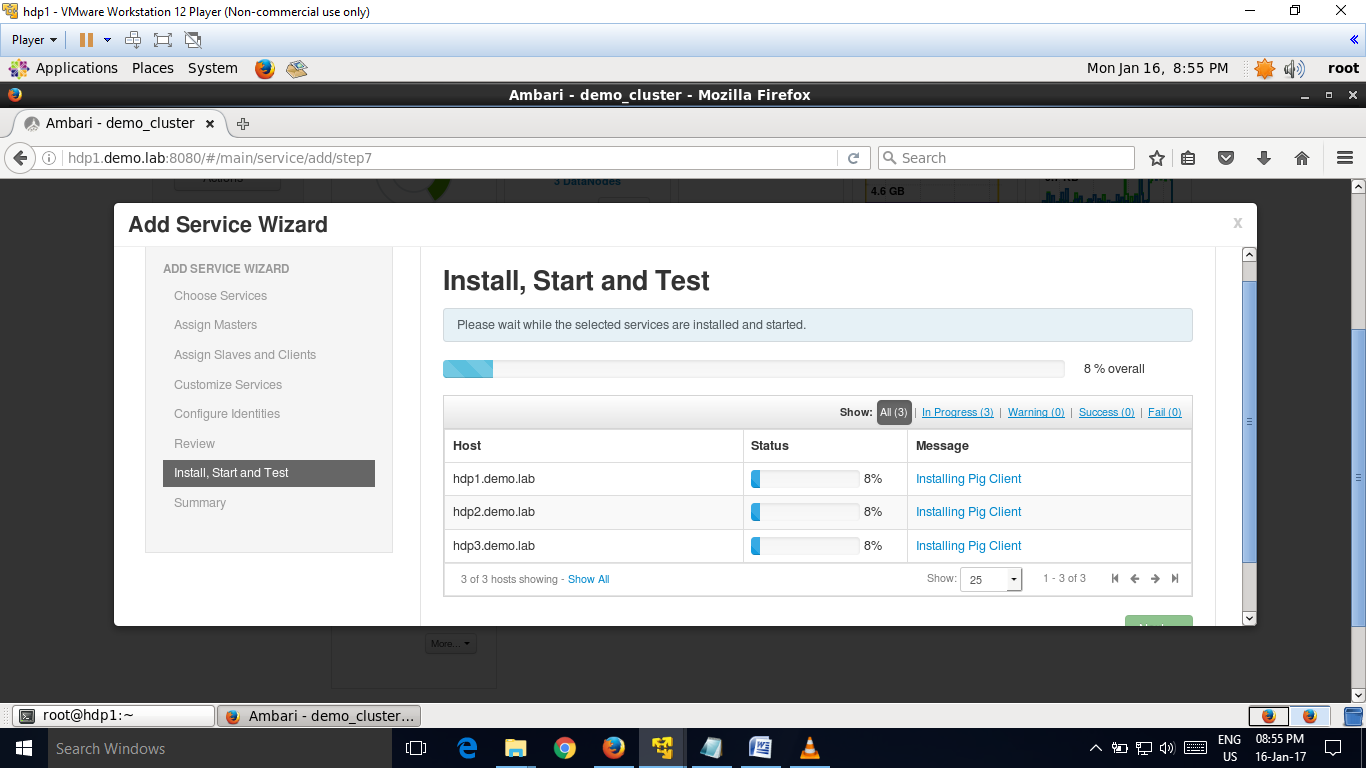
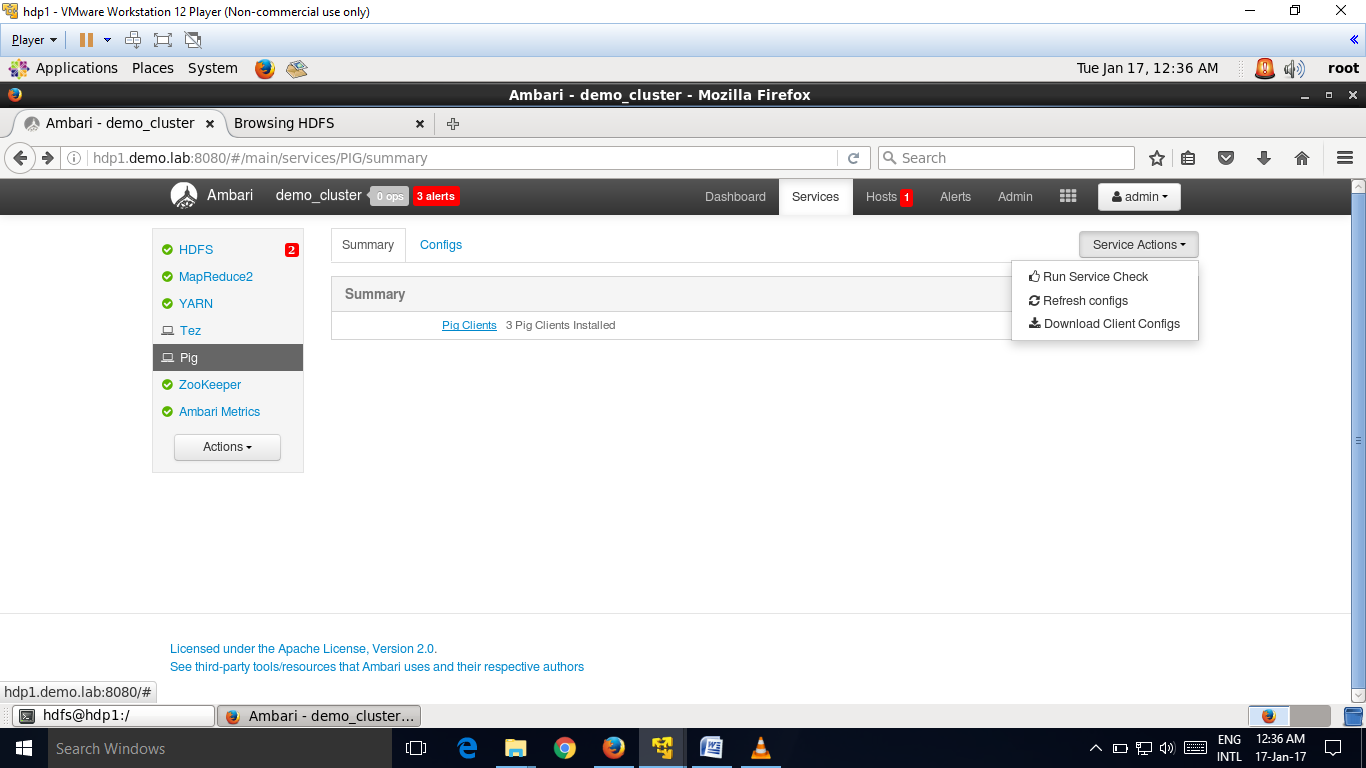
**Screenshots to configure pig in hortonworks:**







Pig is a scripting language to analyze large amount of data.

Pig is a high level scripting language that is used with apache hadoop.

A good example of pig application is the extract,transfer and load transaction model.

Pig has two run modes, local and mapreduce:

a)localmode- to run pig in local mode, we need access to a single machine.

b)mapreduce mode= to run pig in mapreduce mode, we need access to hadoop cluster and hdfs installation.

We can run pig three ways – using either local mode or hadoop map reduce mode :

i)grunt shell- enter pig commands manually using interactive shell,grunt.

ii)script file- place pig commands in a script file and run the script.

iii)embedded system- embedded pig commands in a host language and run the program.

**One example of pig commands**-

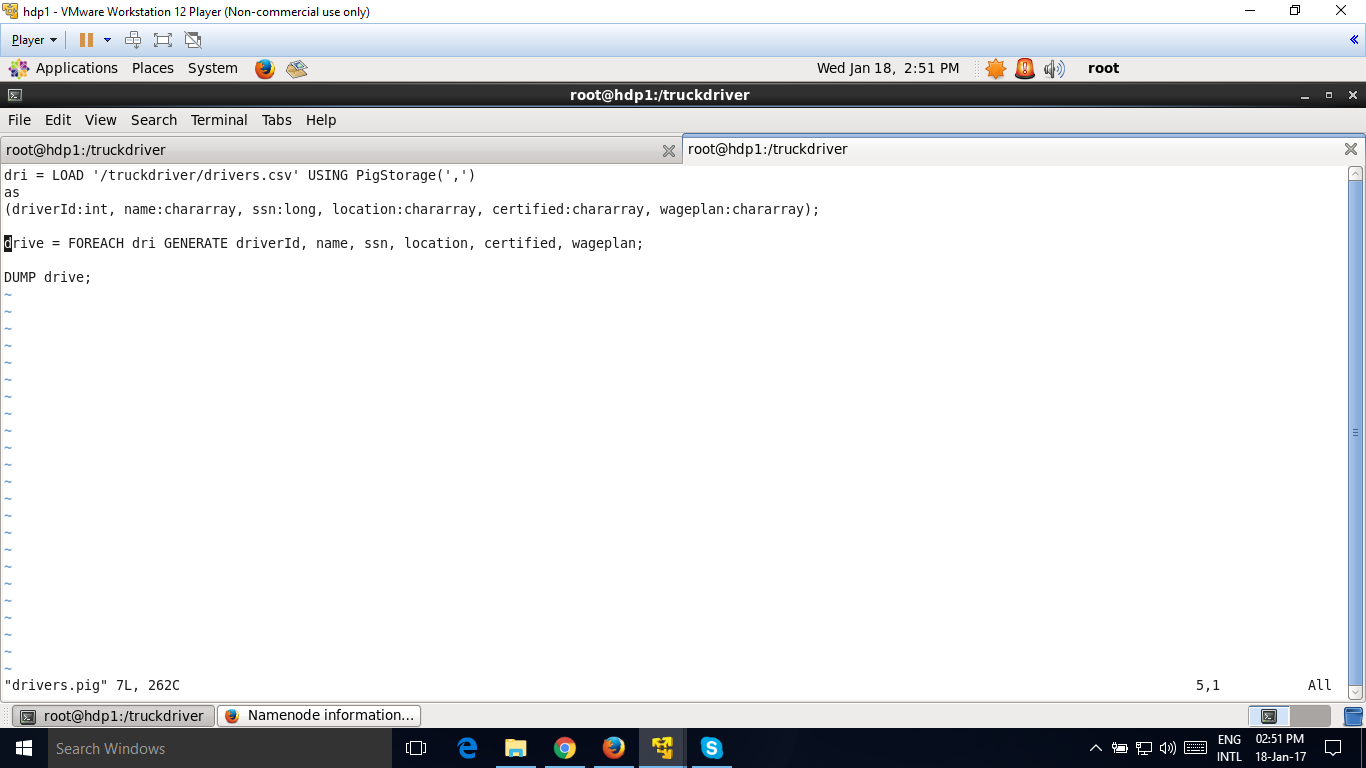
dri = LOAD '/truckdriver/drivers.csv' USING PigStorage(',')

as

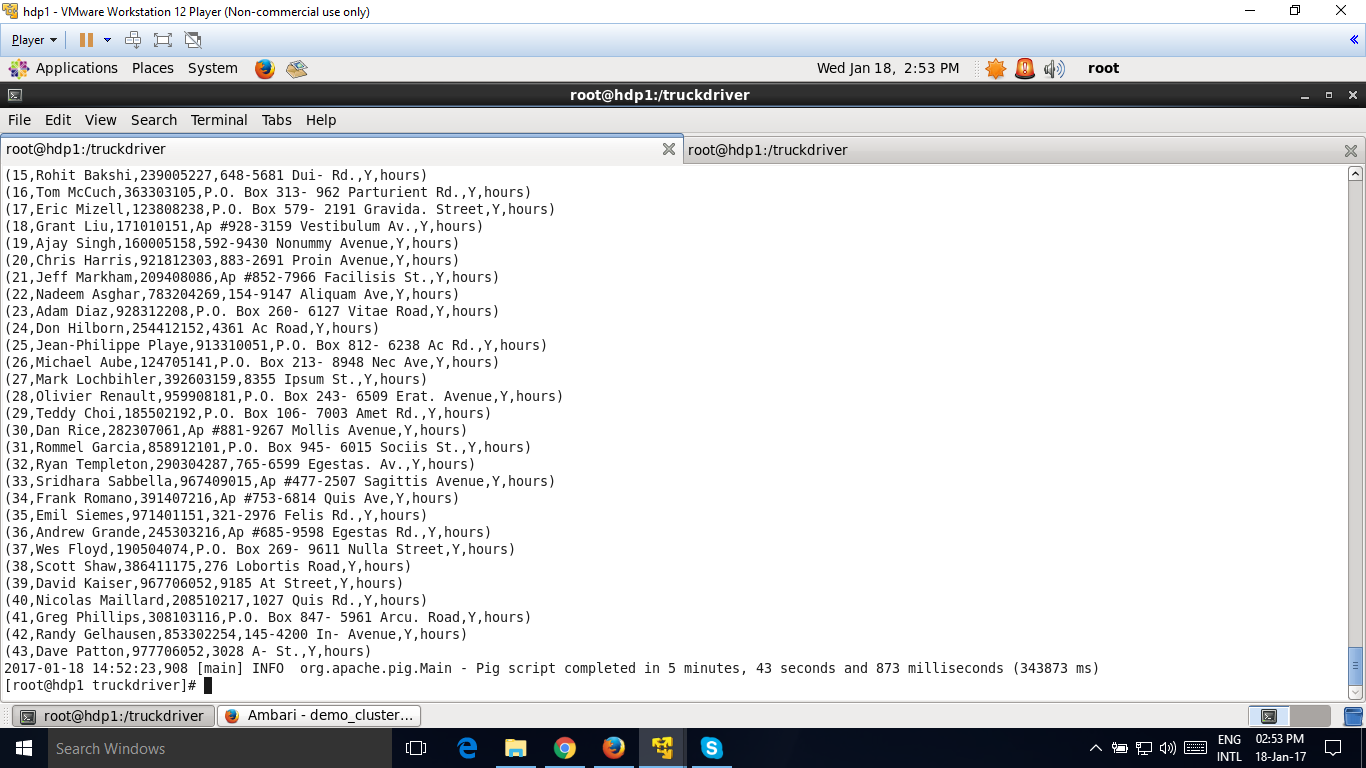
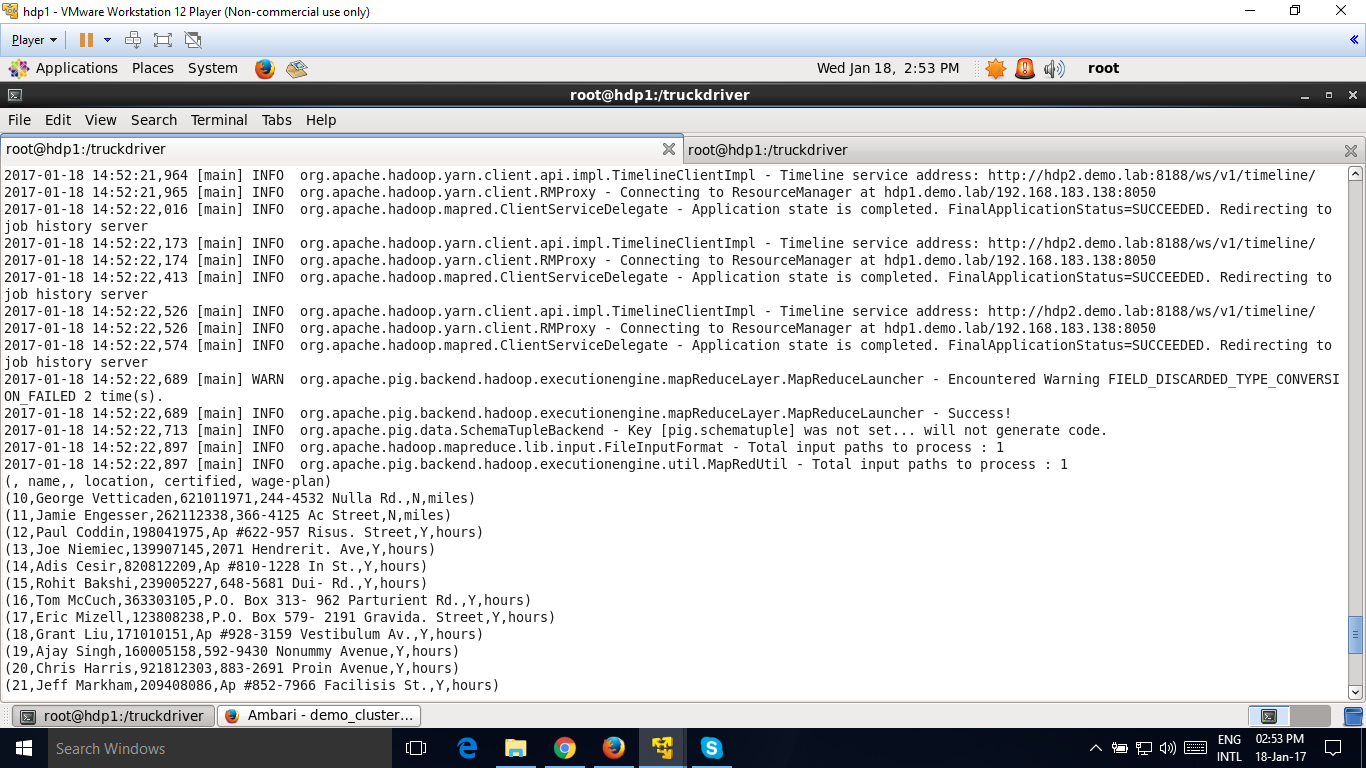
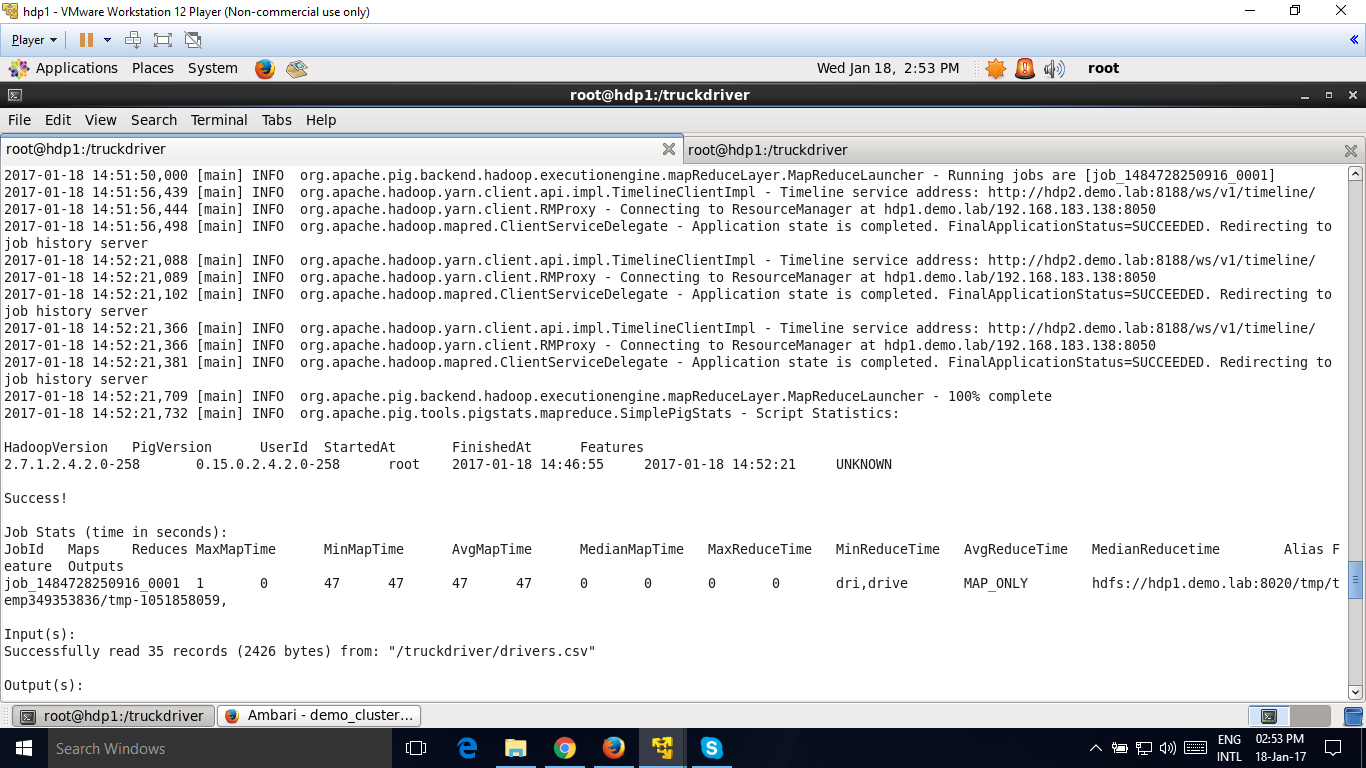
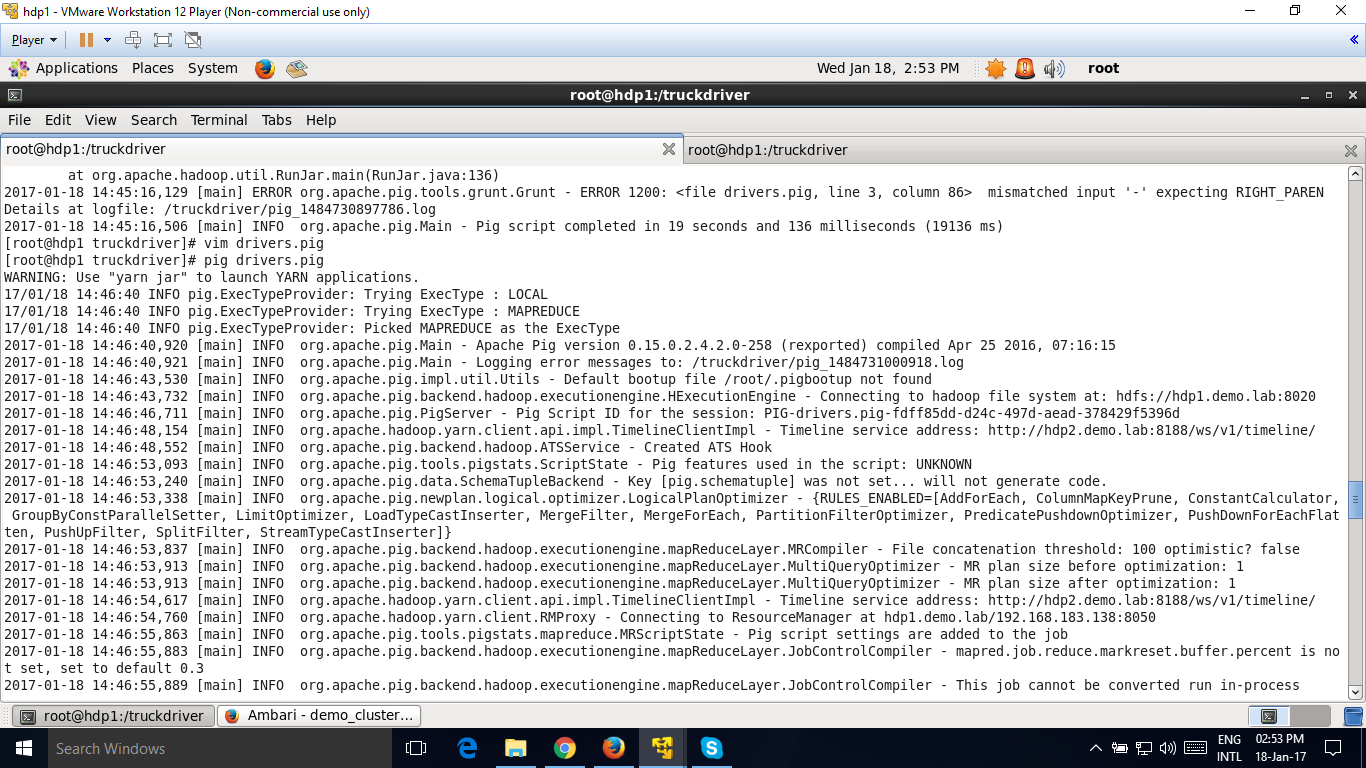
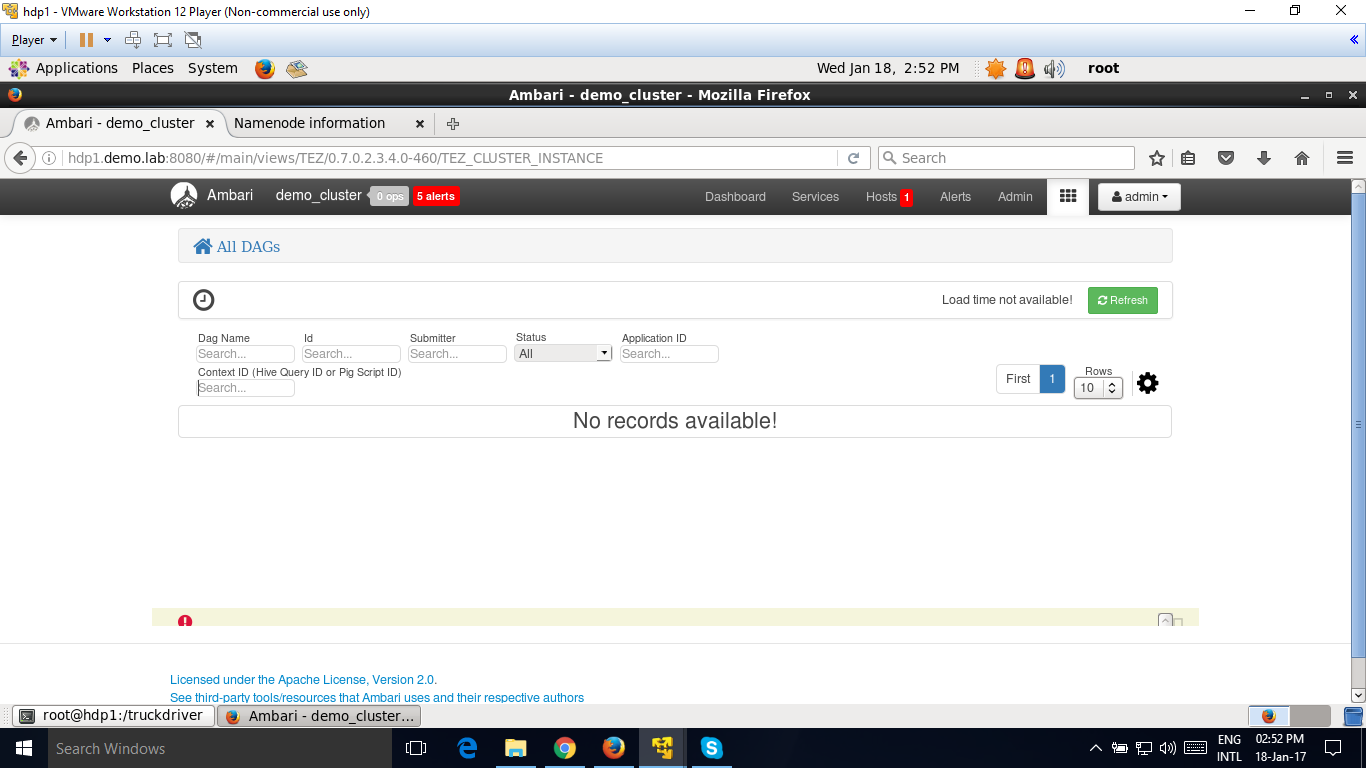
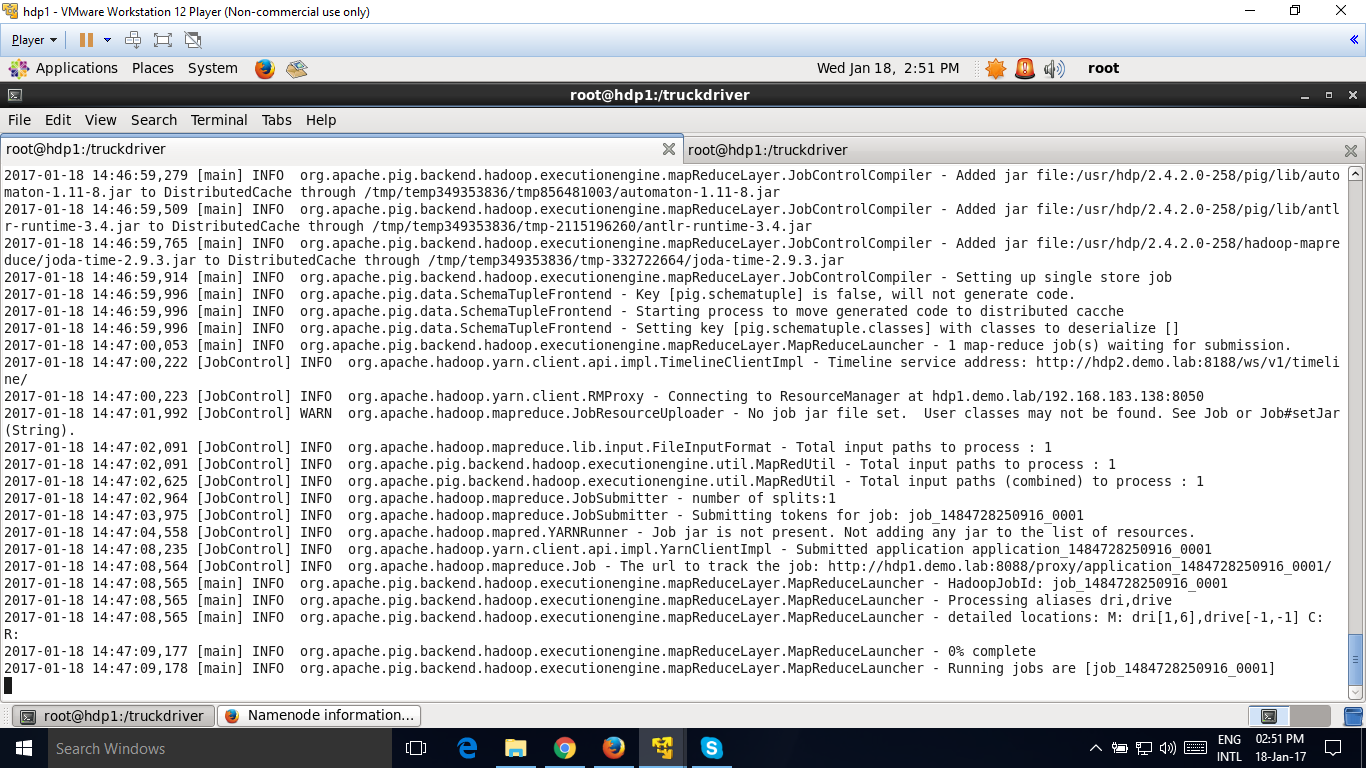
(driverId:int, name:chararray, ssn:long, location:chararray, certified:chararray, wageplan:chararray);

drive = FOREACH dri GENERATE driverId, name, ssn, location, certified, wageplan;

DUMP drive;



**Output**

****

Running job with tez:

Pig drivers.pig –x tez

