**MAJOR PROJECT**

* 1. **Project Overview : To develop the System to analyze the log data (In XML format) of government progress of various development activities.**

**Requirements :**

**• The FLUME job which will format the data and place the data to HDFS**

**• Pig/MapReduce job for parsing the XML data.**

**• Create Pig scripts/MapReduce jobs to analyze the data**

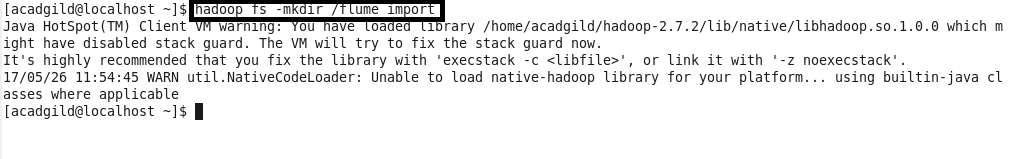
**• Create the Sqoop job to store the data in database**

**Problem statement 1.**

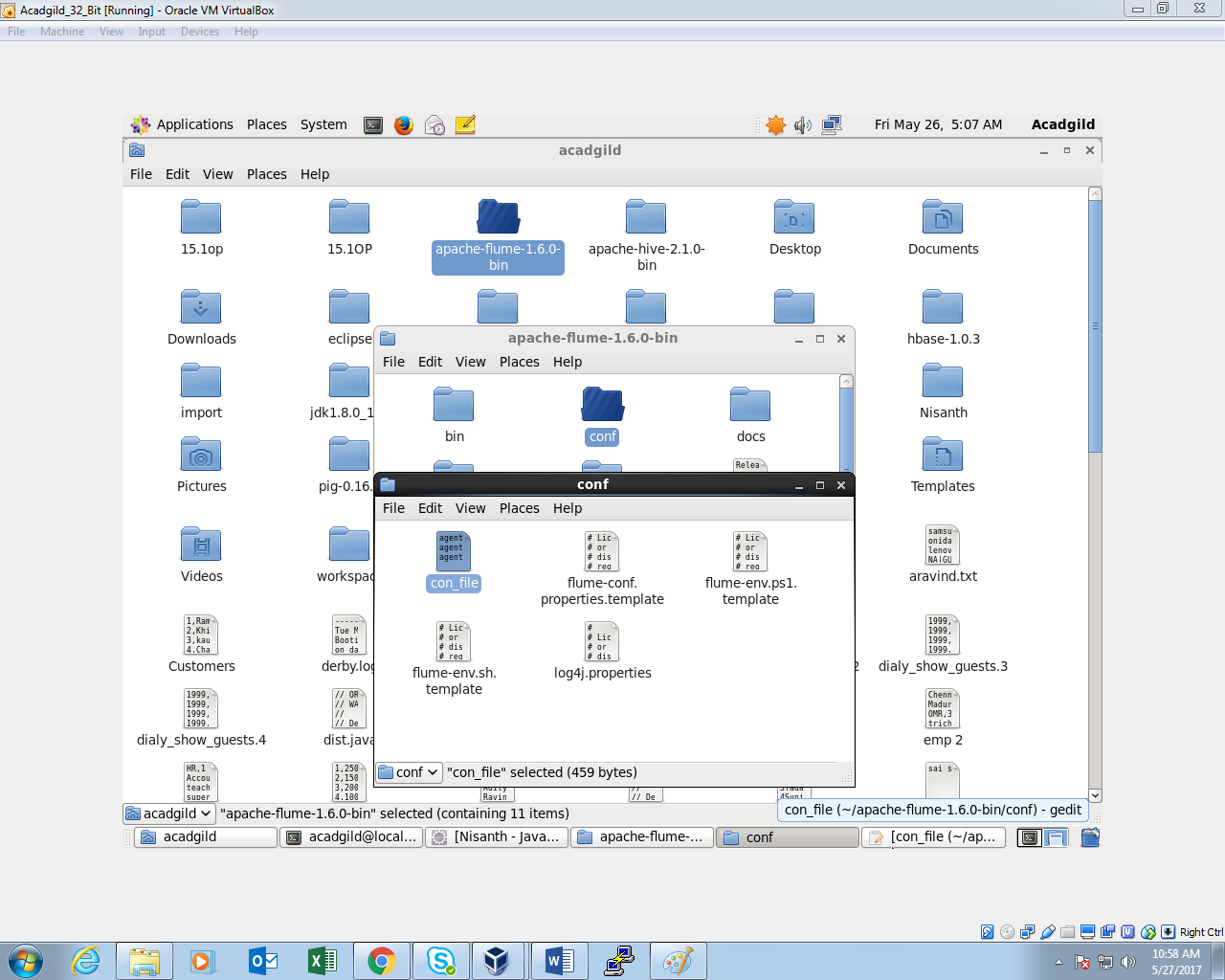
**Find out the districts who achieved 100 percent objective in BPL cards Export the results to MySql using Sqoop**

***Copy dataset from local file system to HDFS using flume.***

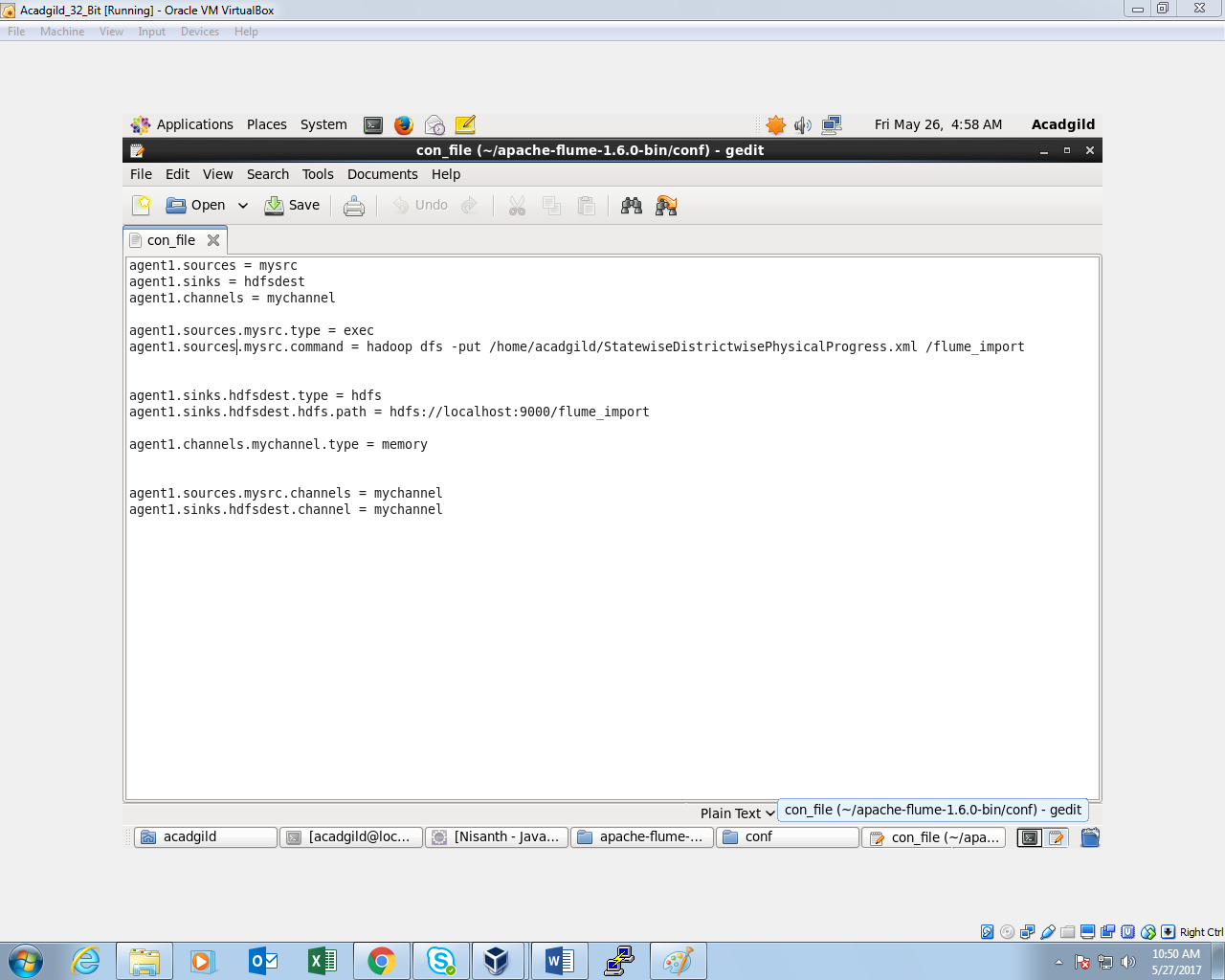
***Making directory in flume:***

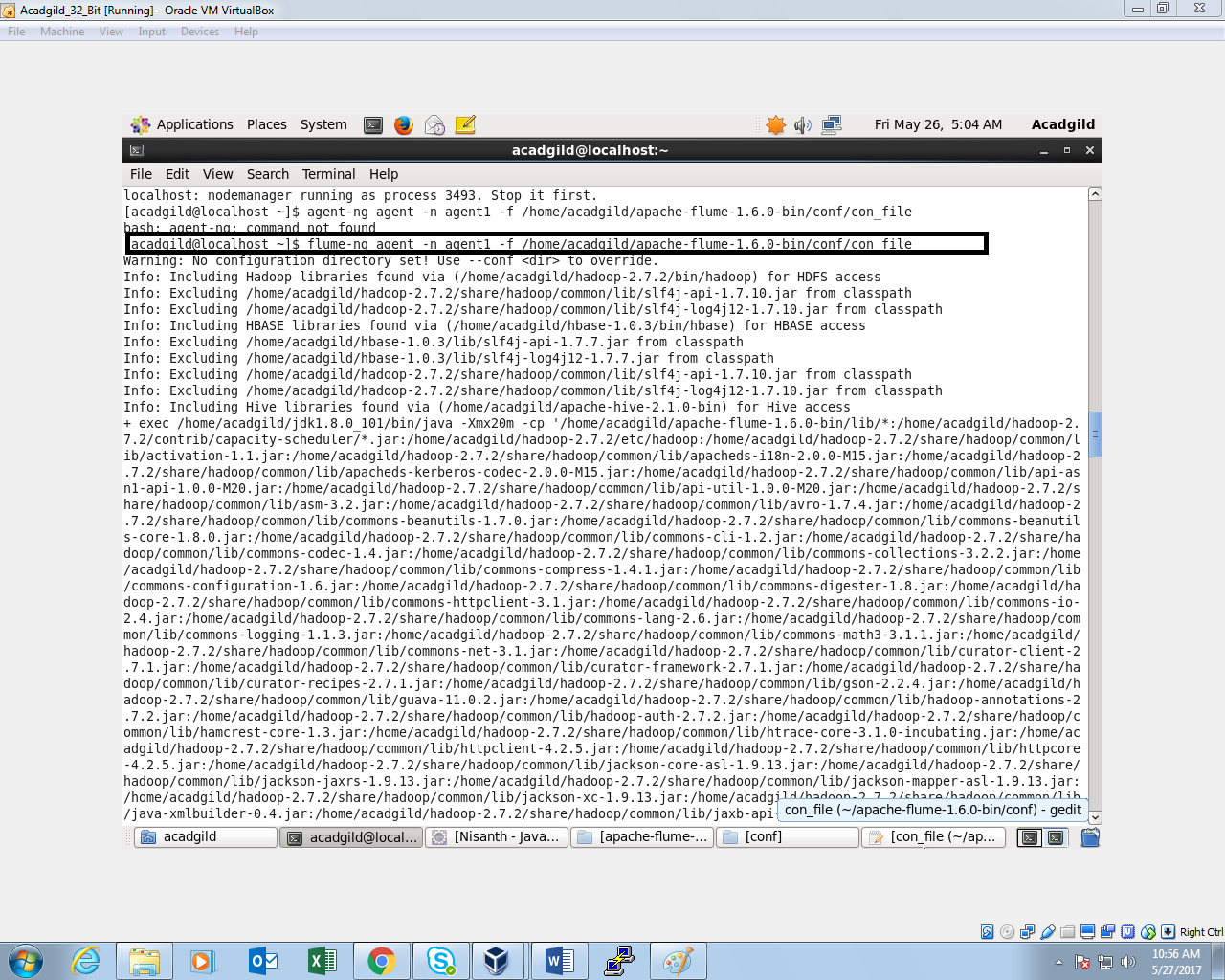
******

***Using the given configuration file:***

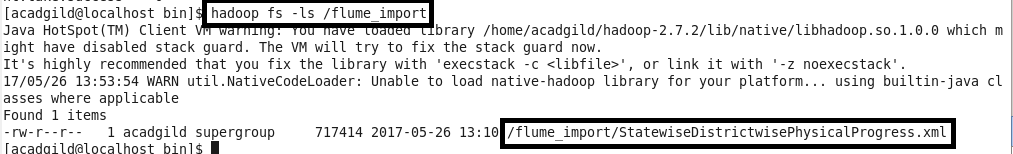


***Configuration file:***



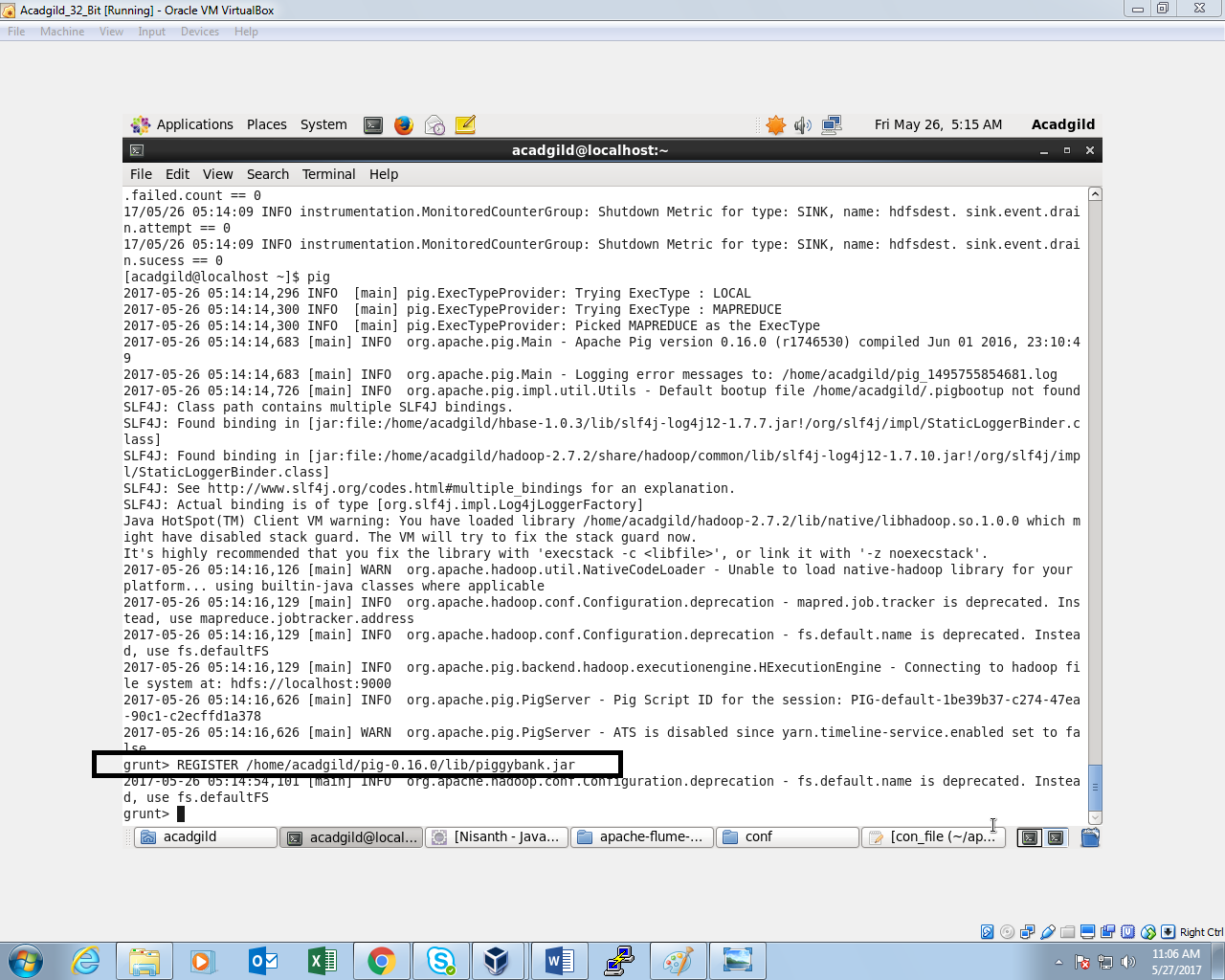


***Listing the files inside the directory /flume\_import***

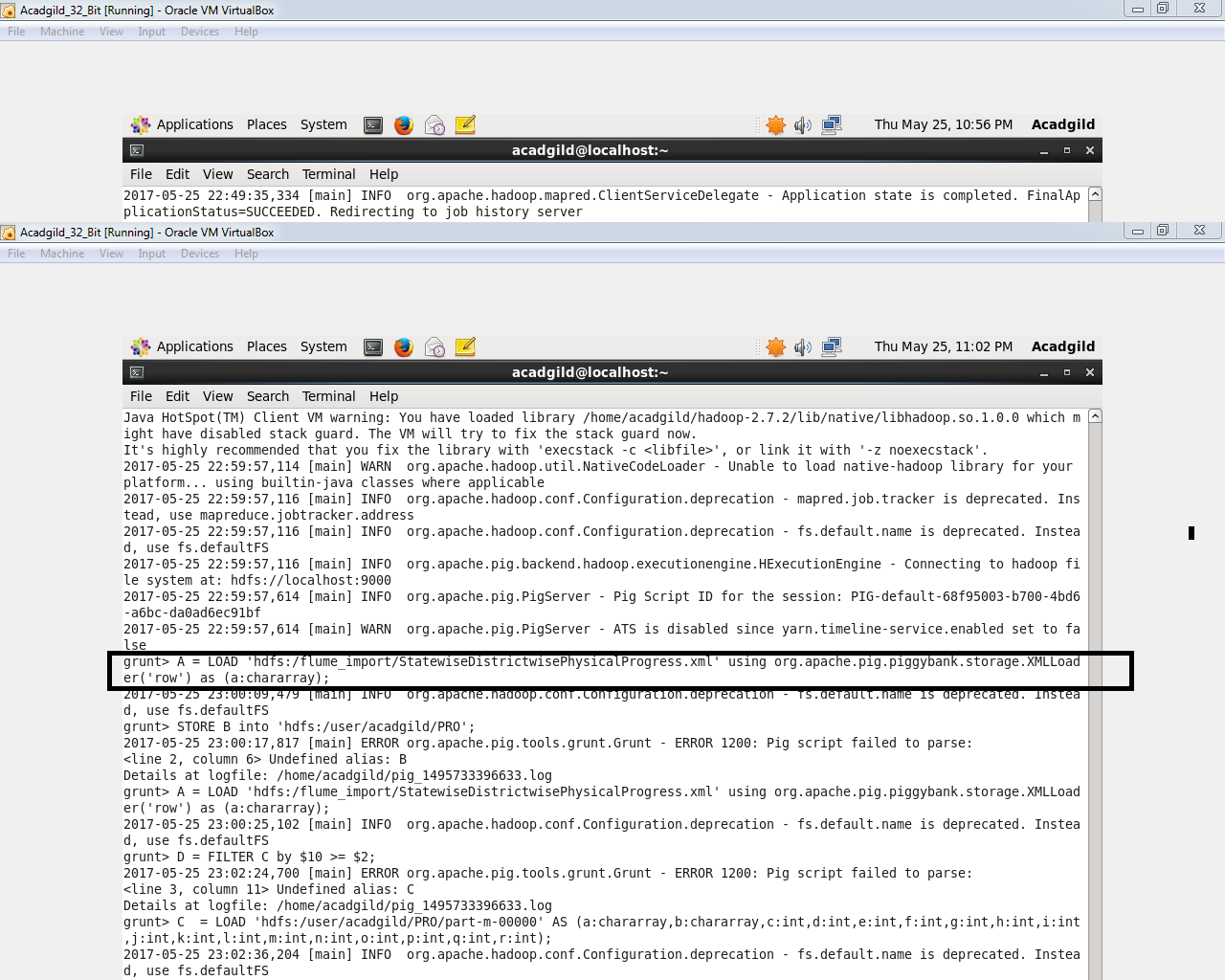


**Input file is in the XML format.so pig is used to parse the data and get the results**

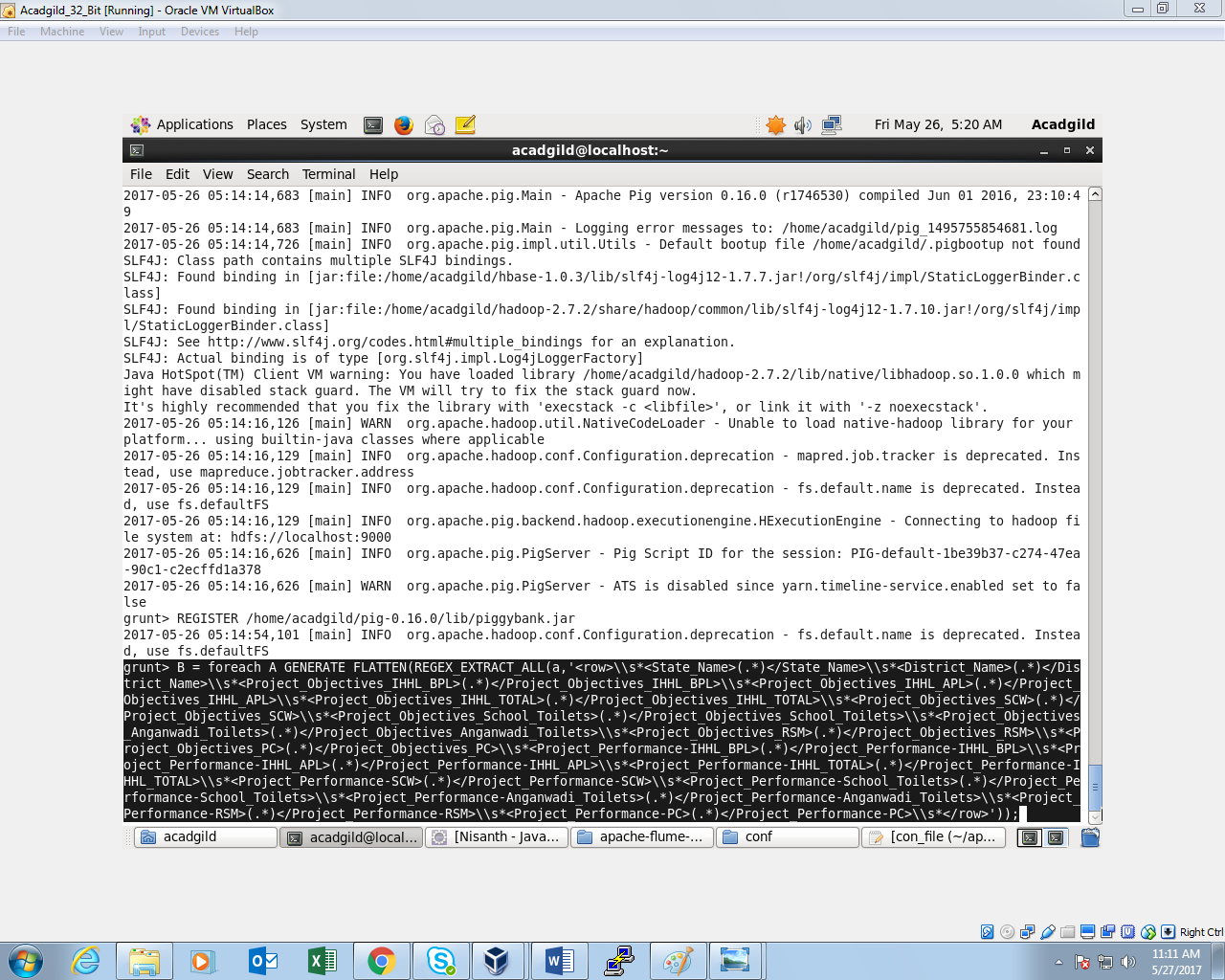
***REGISTERING PIG JAR:***

****

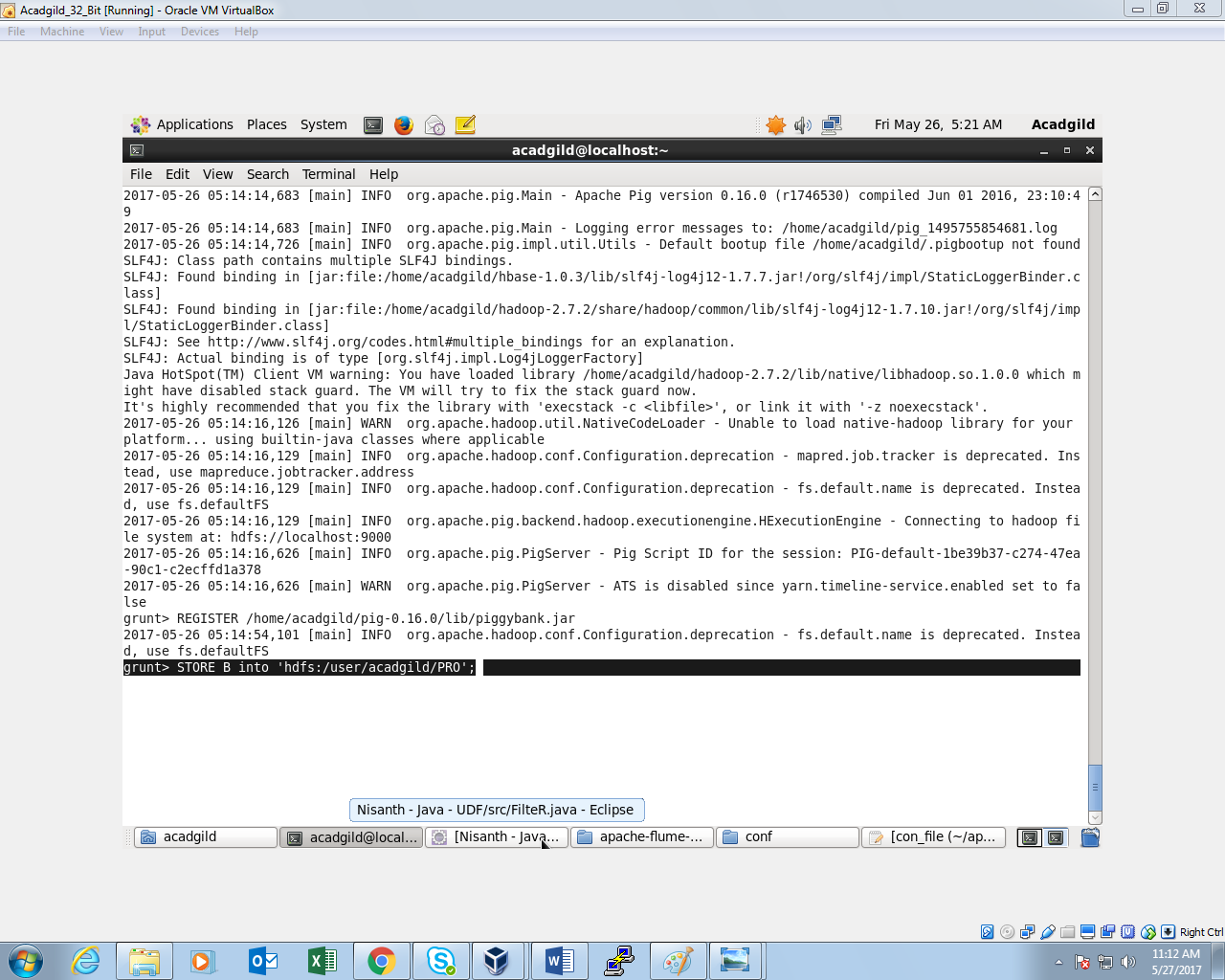
***Loading the file from HDFS into pig shell:***



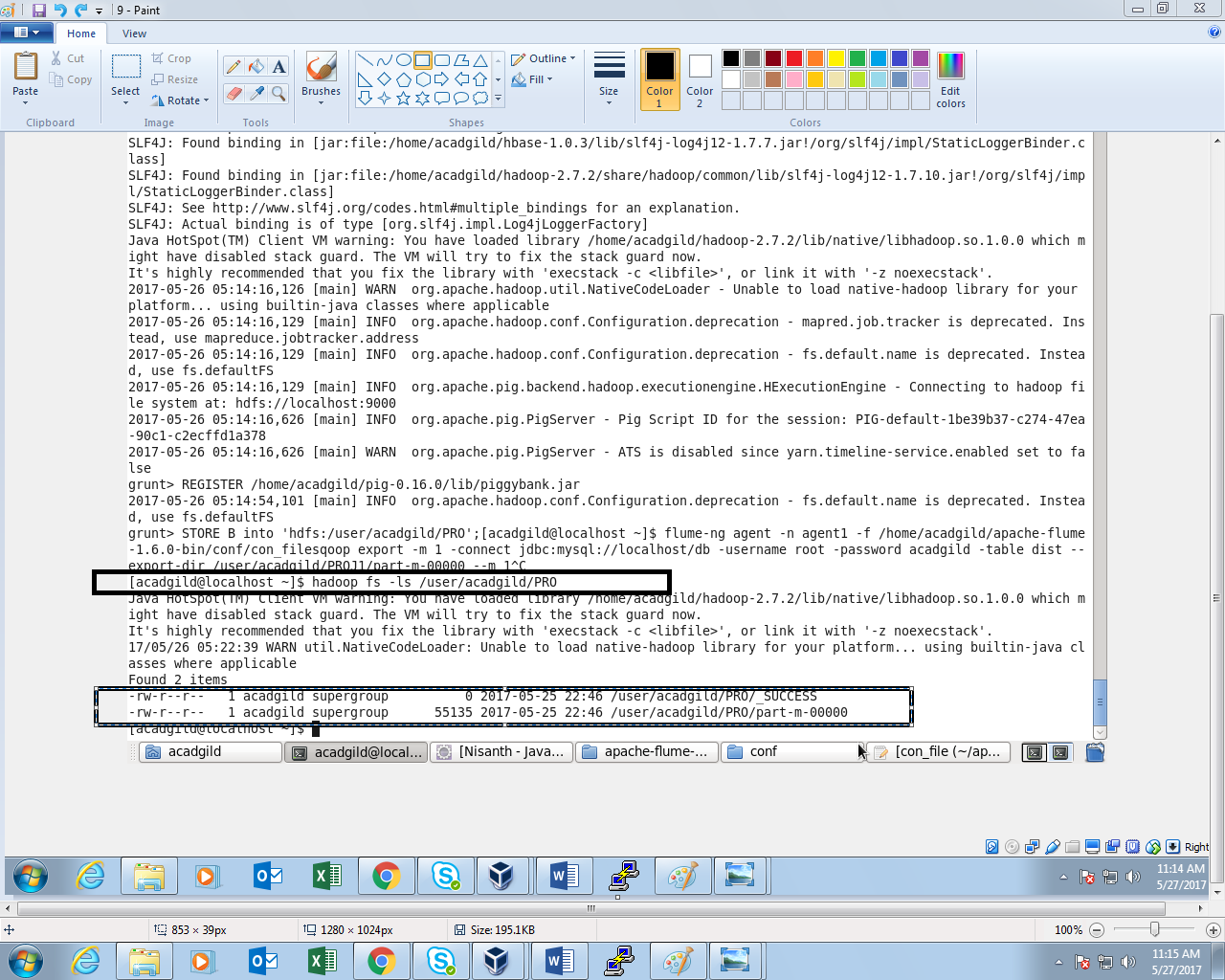
***Parsing the XML format file:***



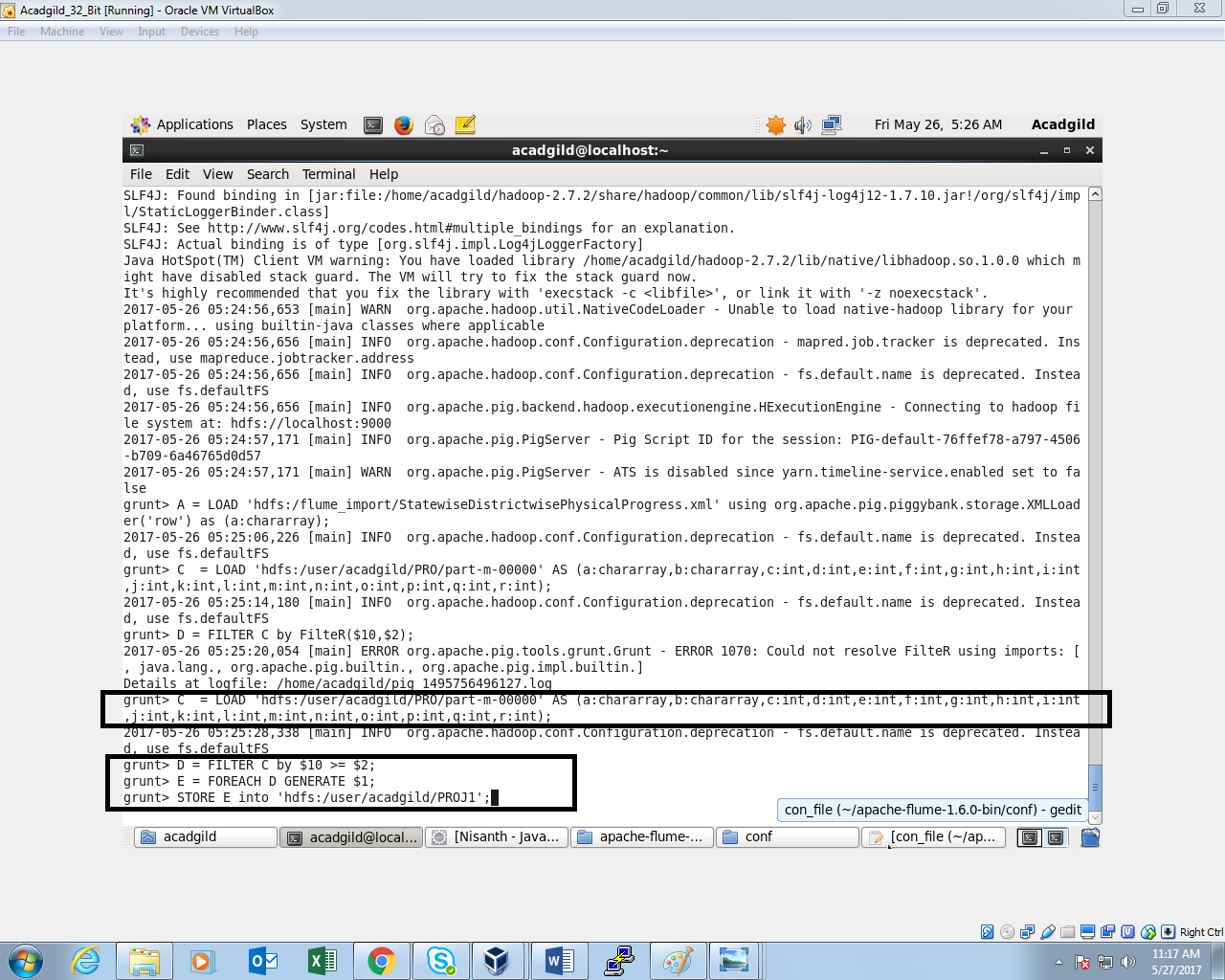
***Storing the parsed file into file “PRO”:***



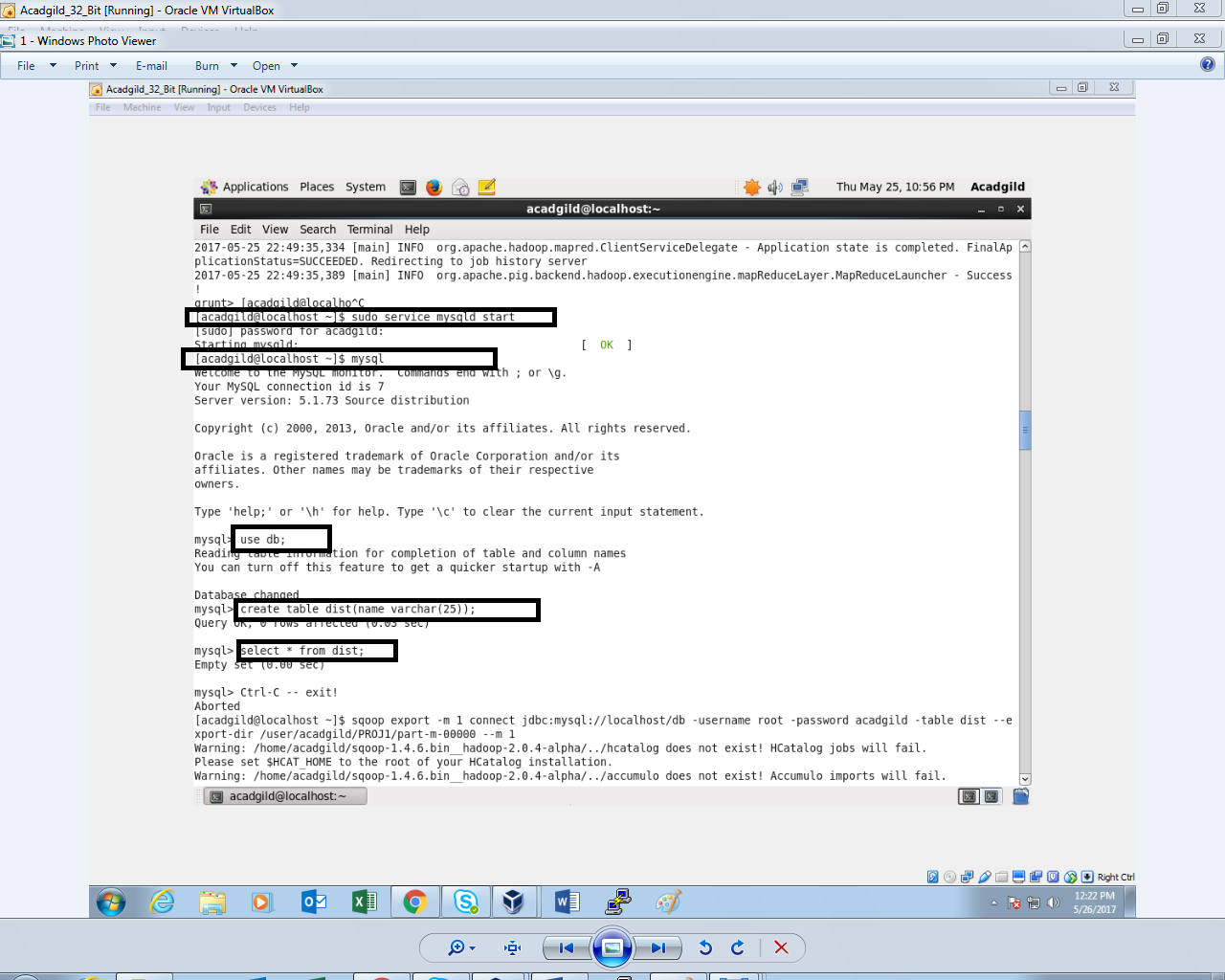
***Listing the file from the folder PRO:***



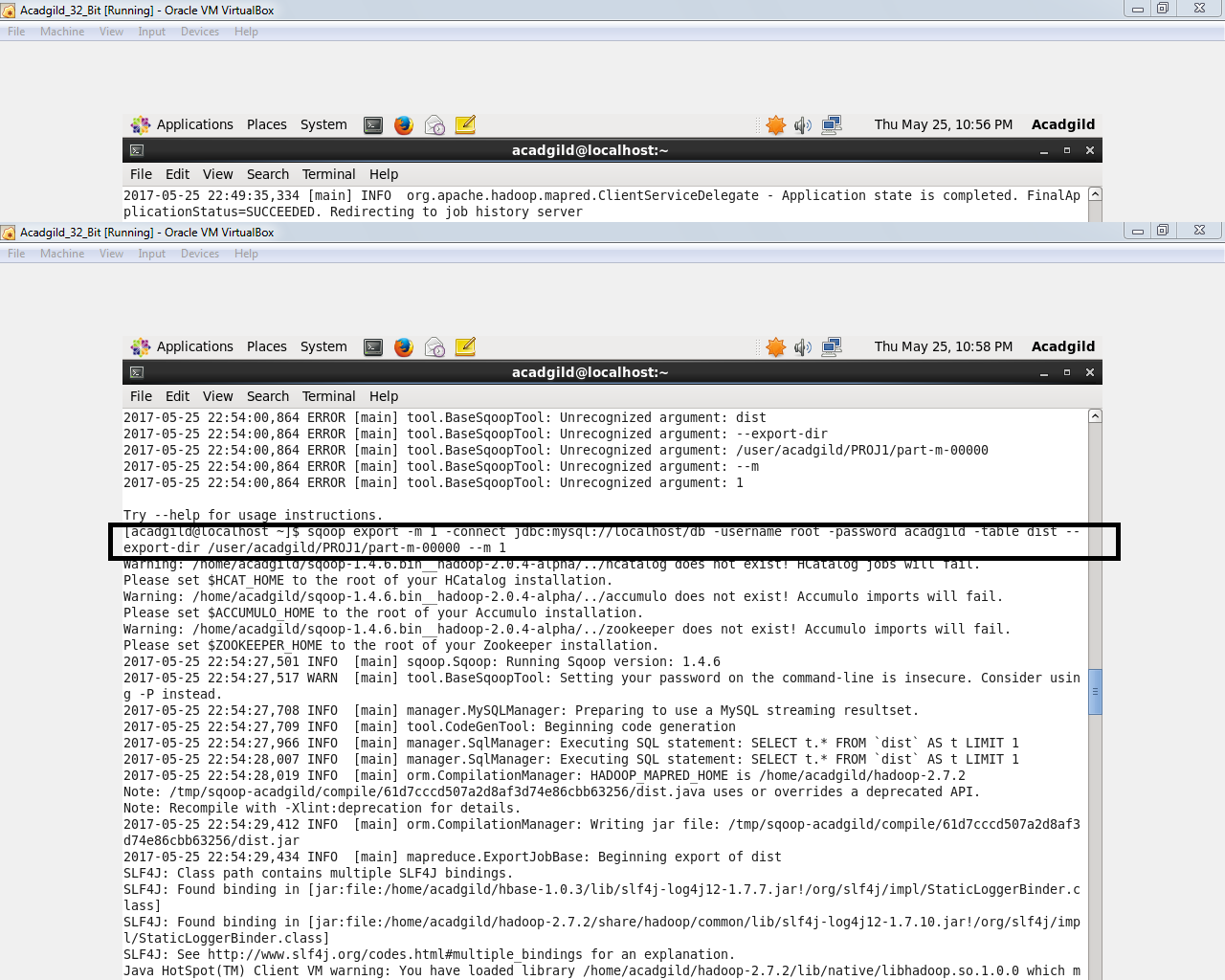
***Performing pig commands and Storing into file PROJ1:***



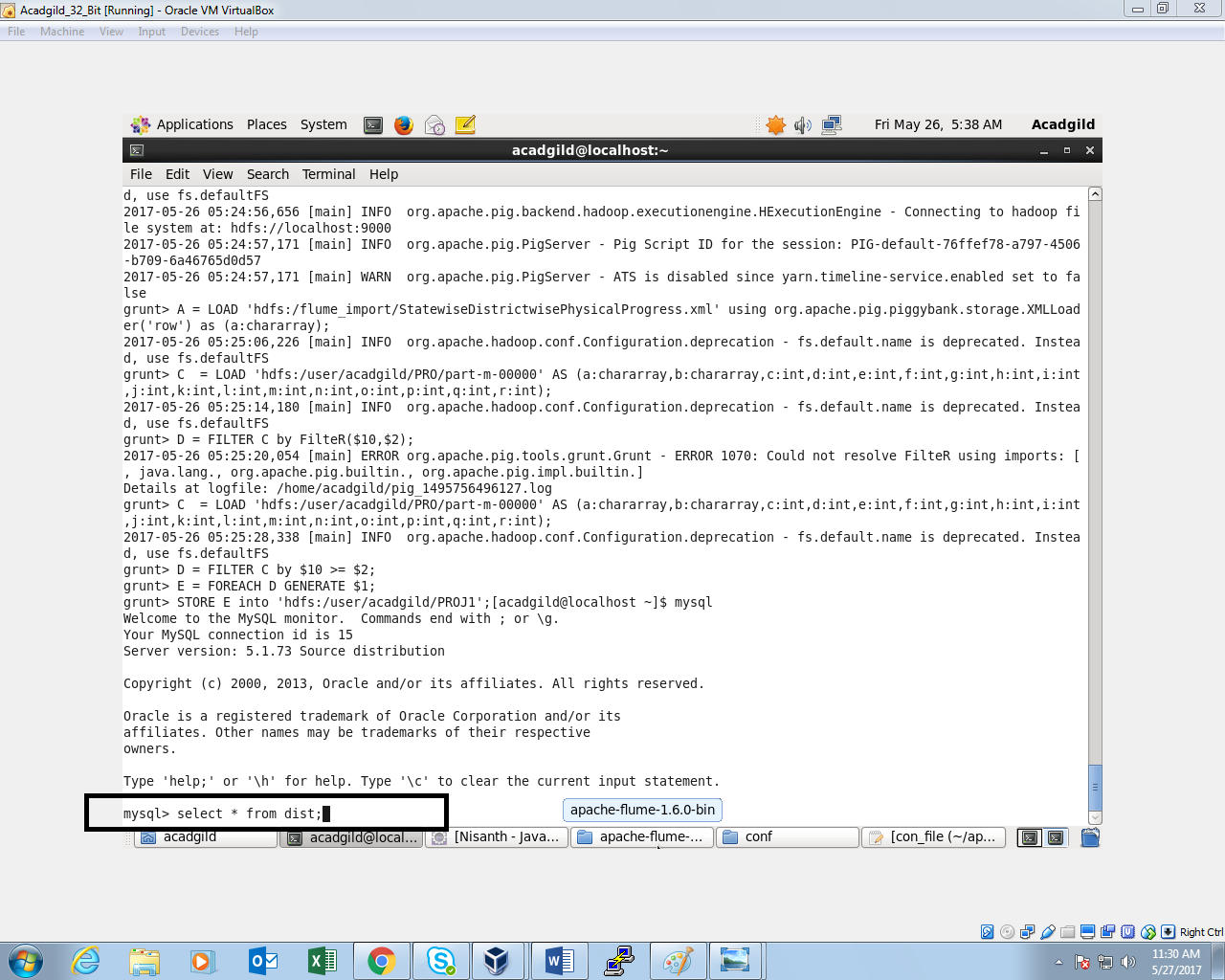
***Starting mysql and creating a table named dist:***



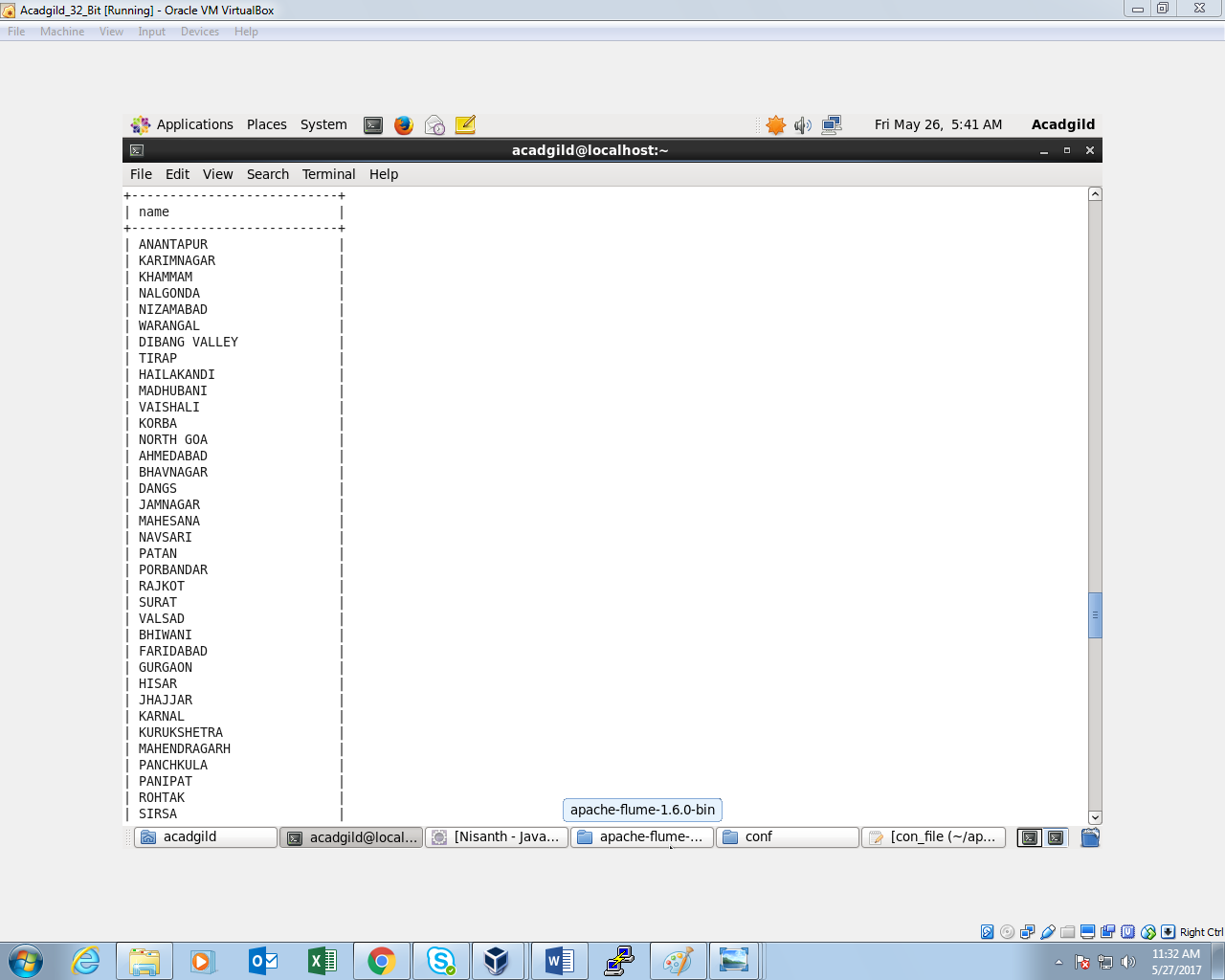
***Data from hdfs is transferred to the table using sqoop:***

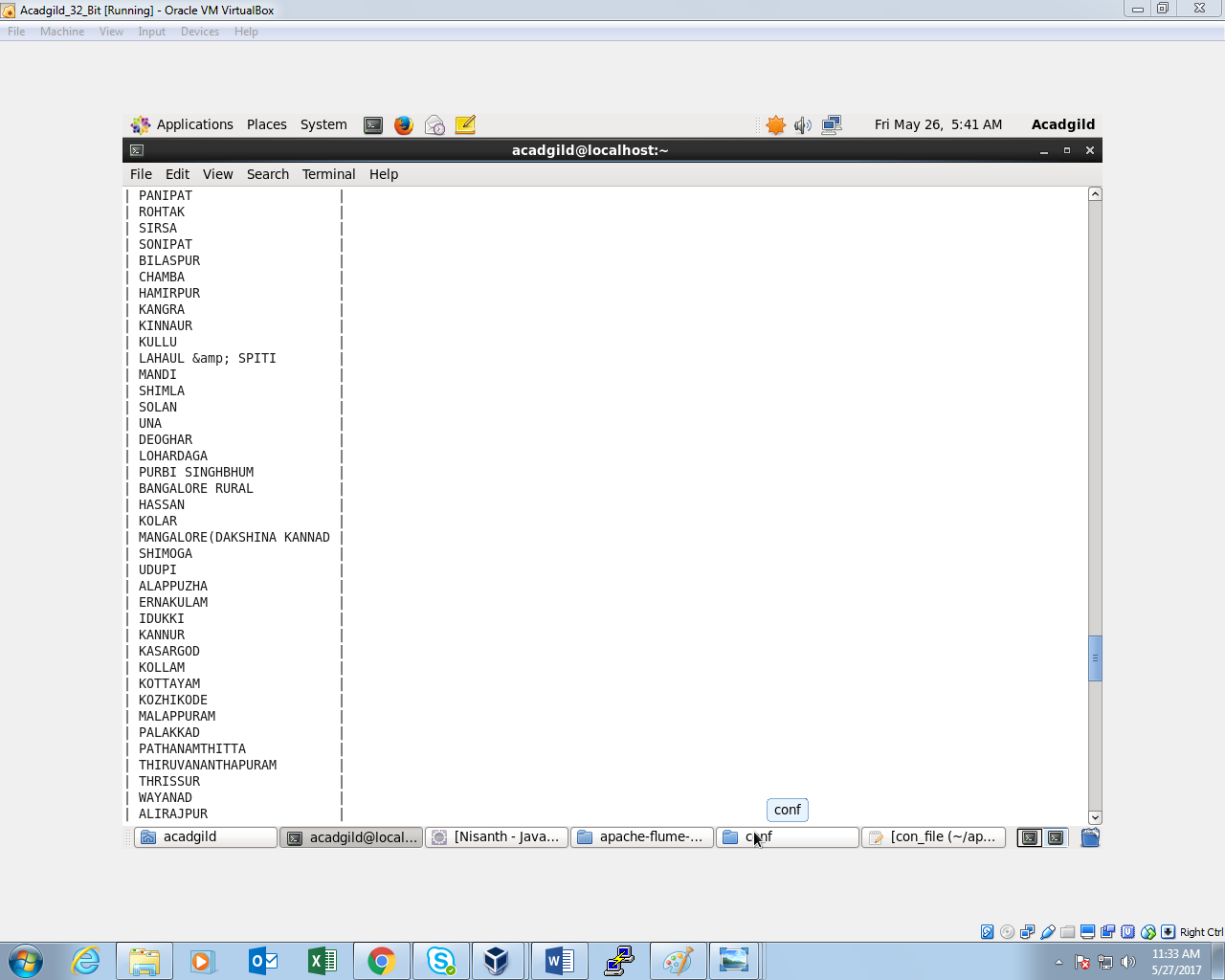


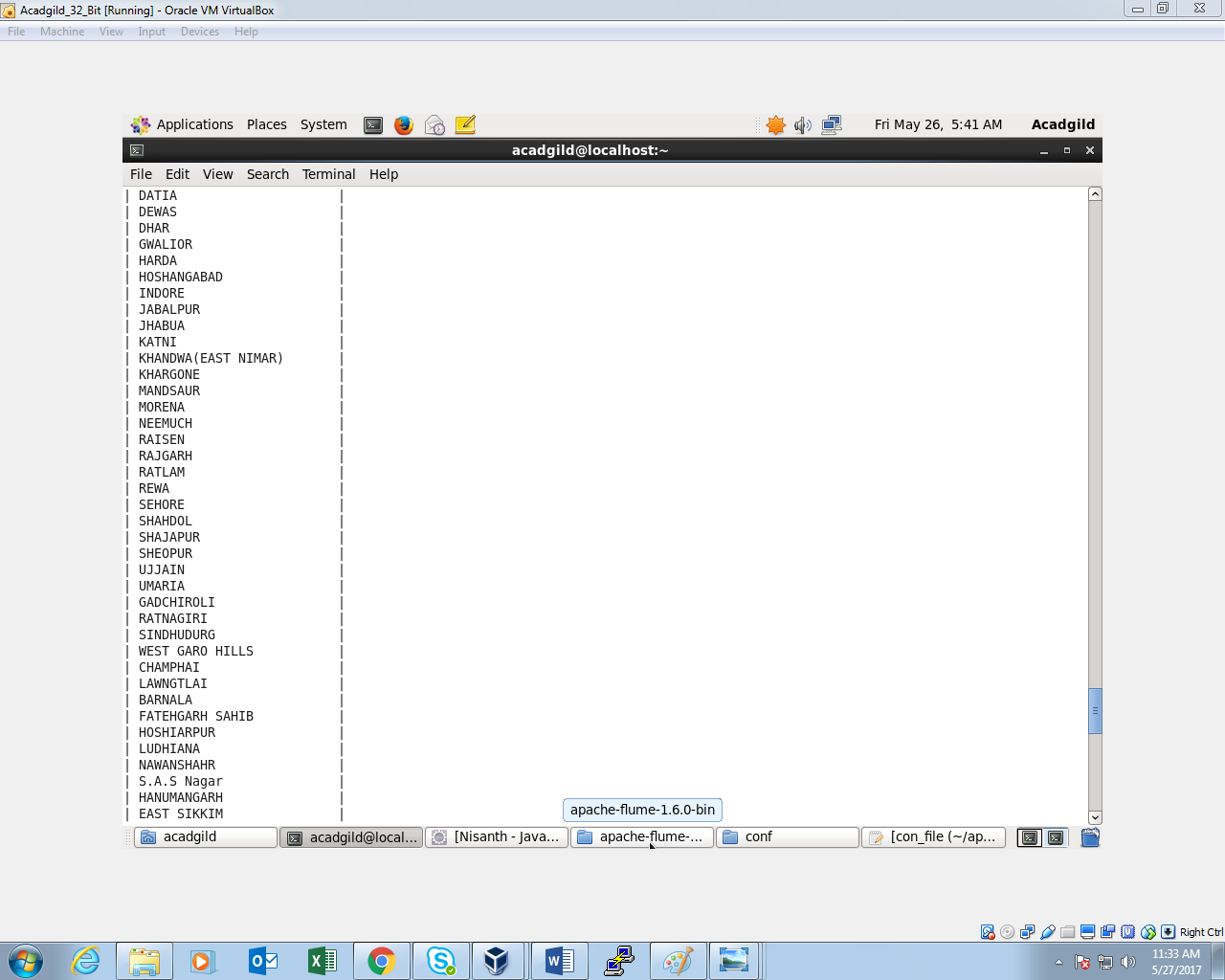
***Showing data from the table:***



***OUTPUT:***





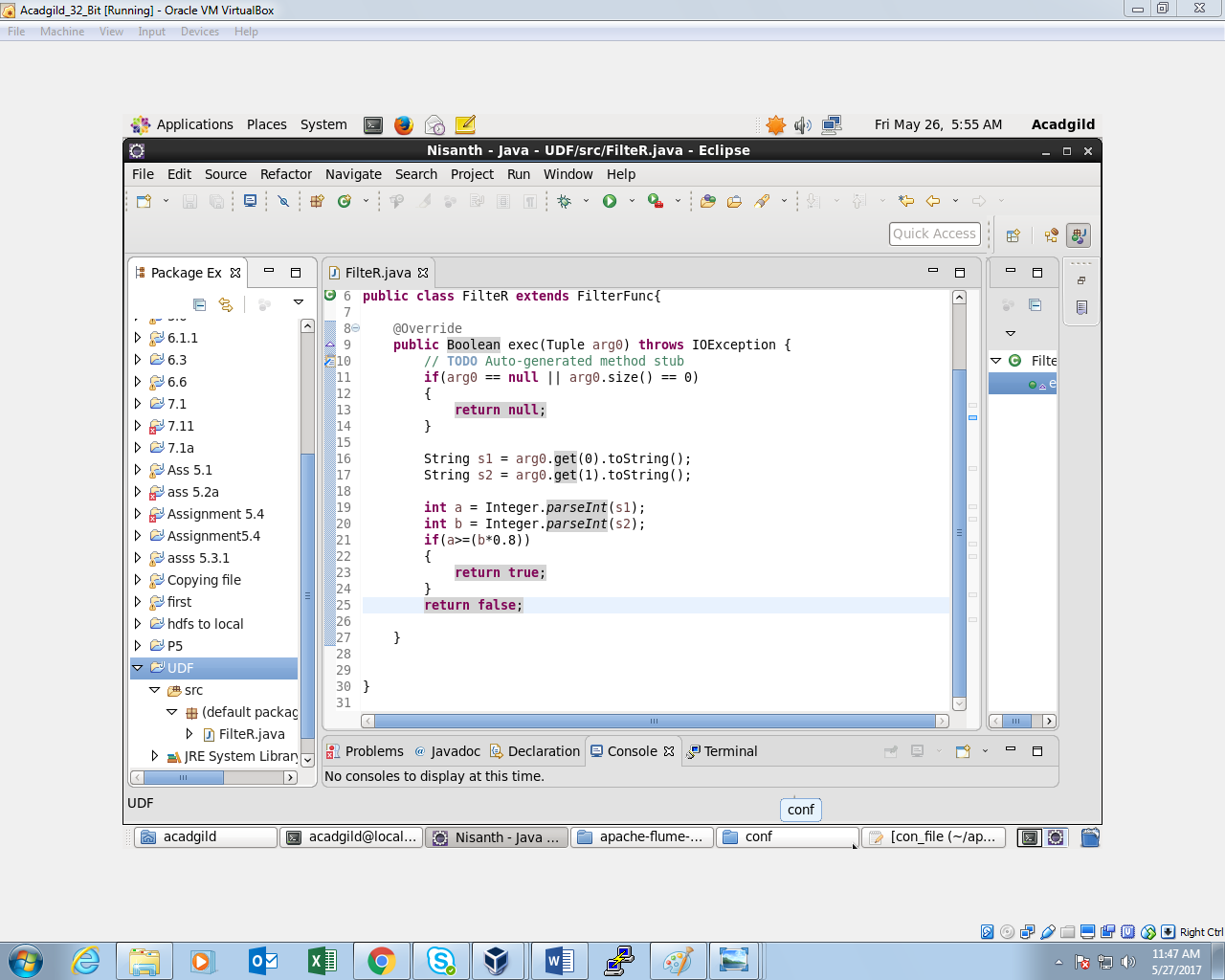


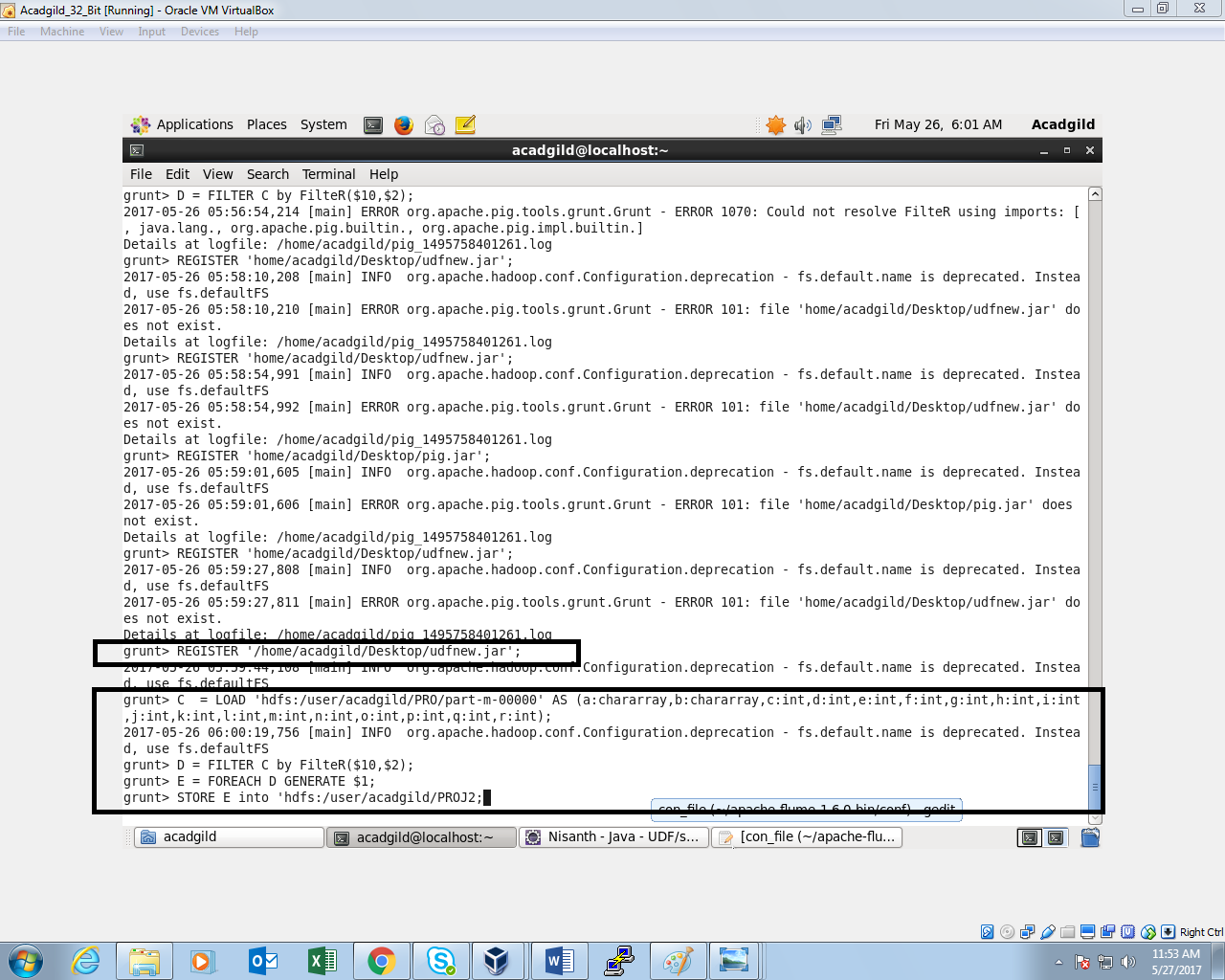


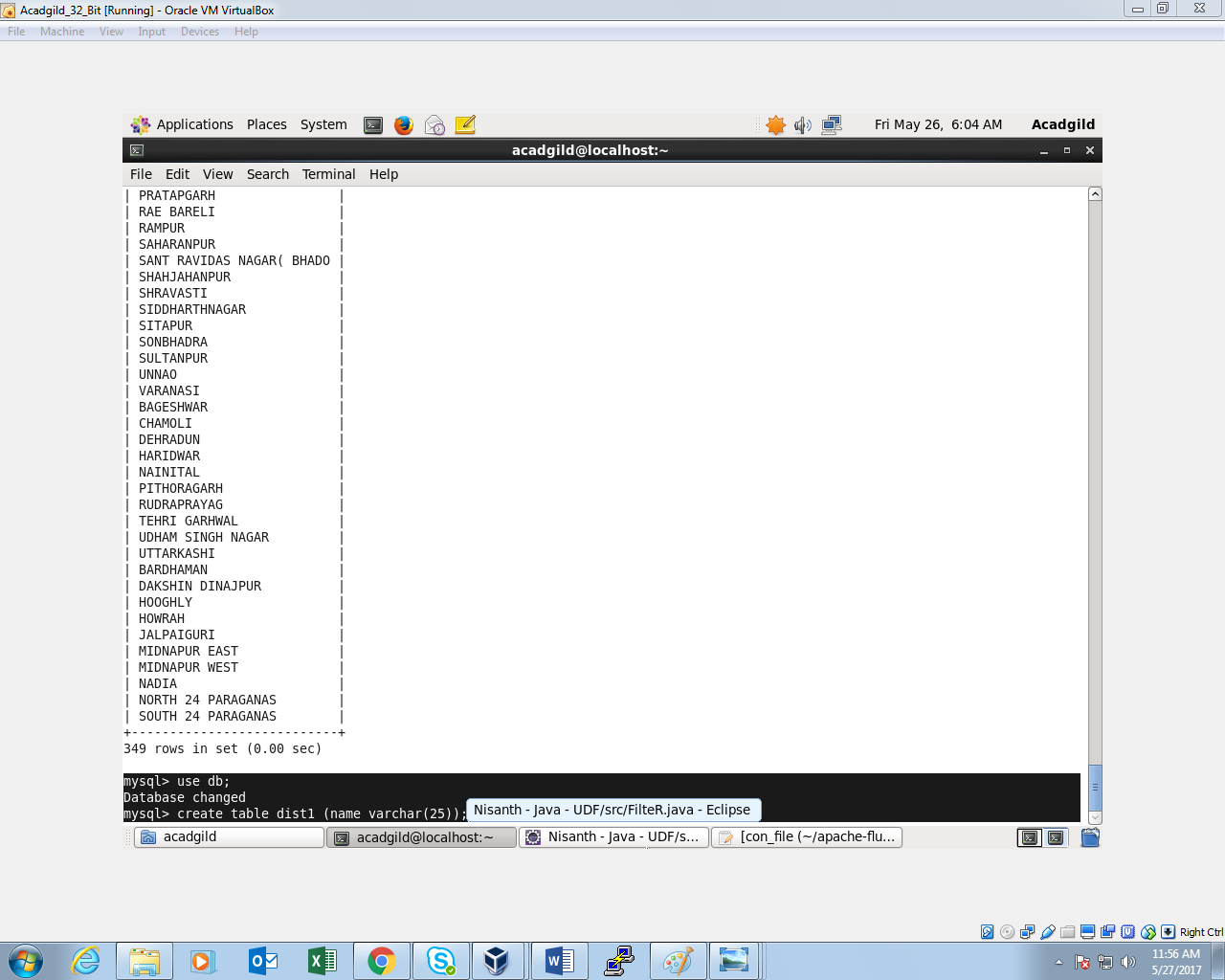
**Thus, 176 districts are found achieved 100 percent objective in BPL cards**

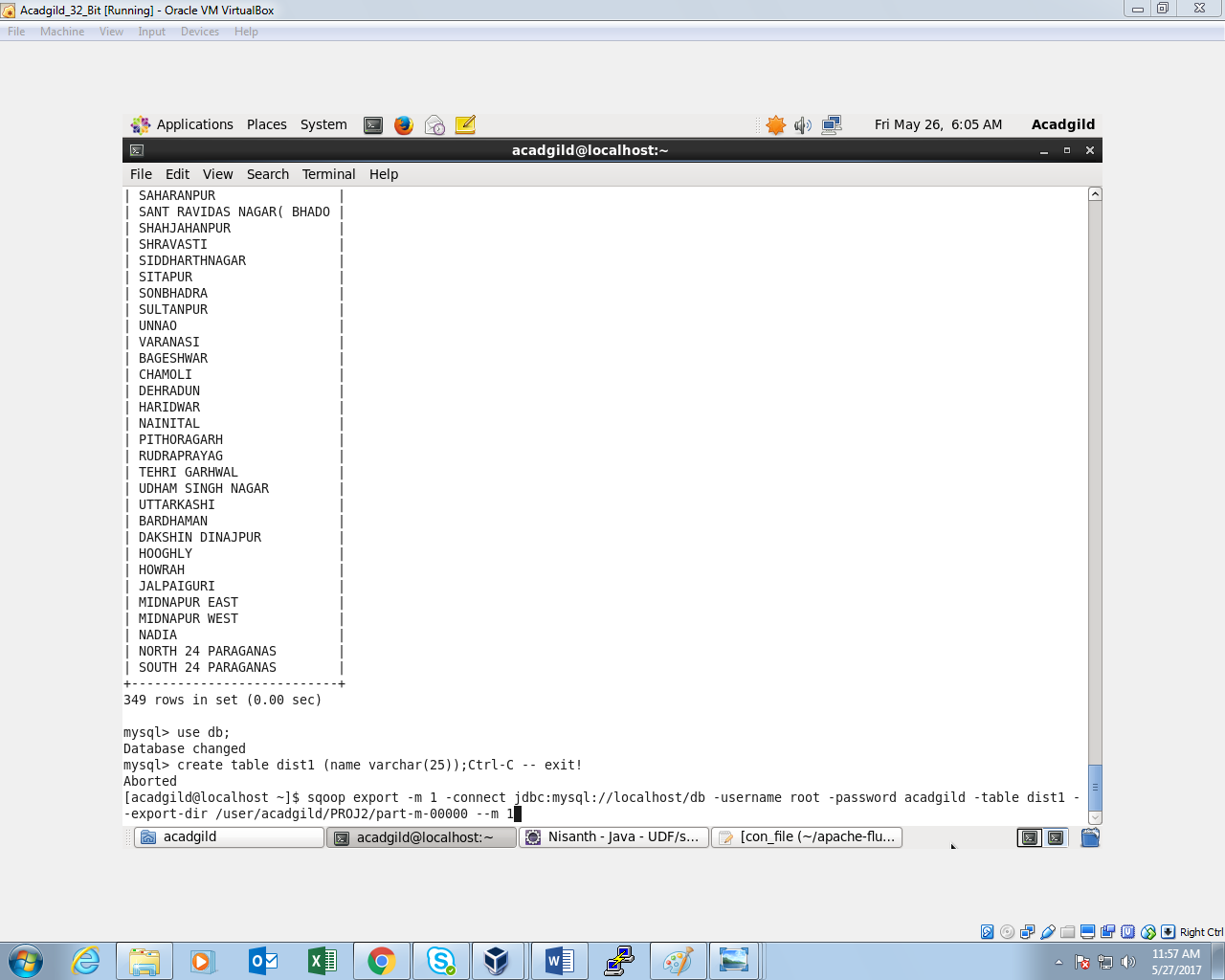
**Write a Pig UDF to filter the districts which have reached 80% of objectives of BPL cards. Export the results to MySQL using Sqoop.**

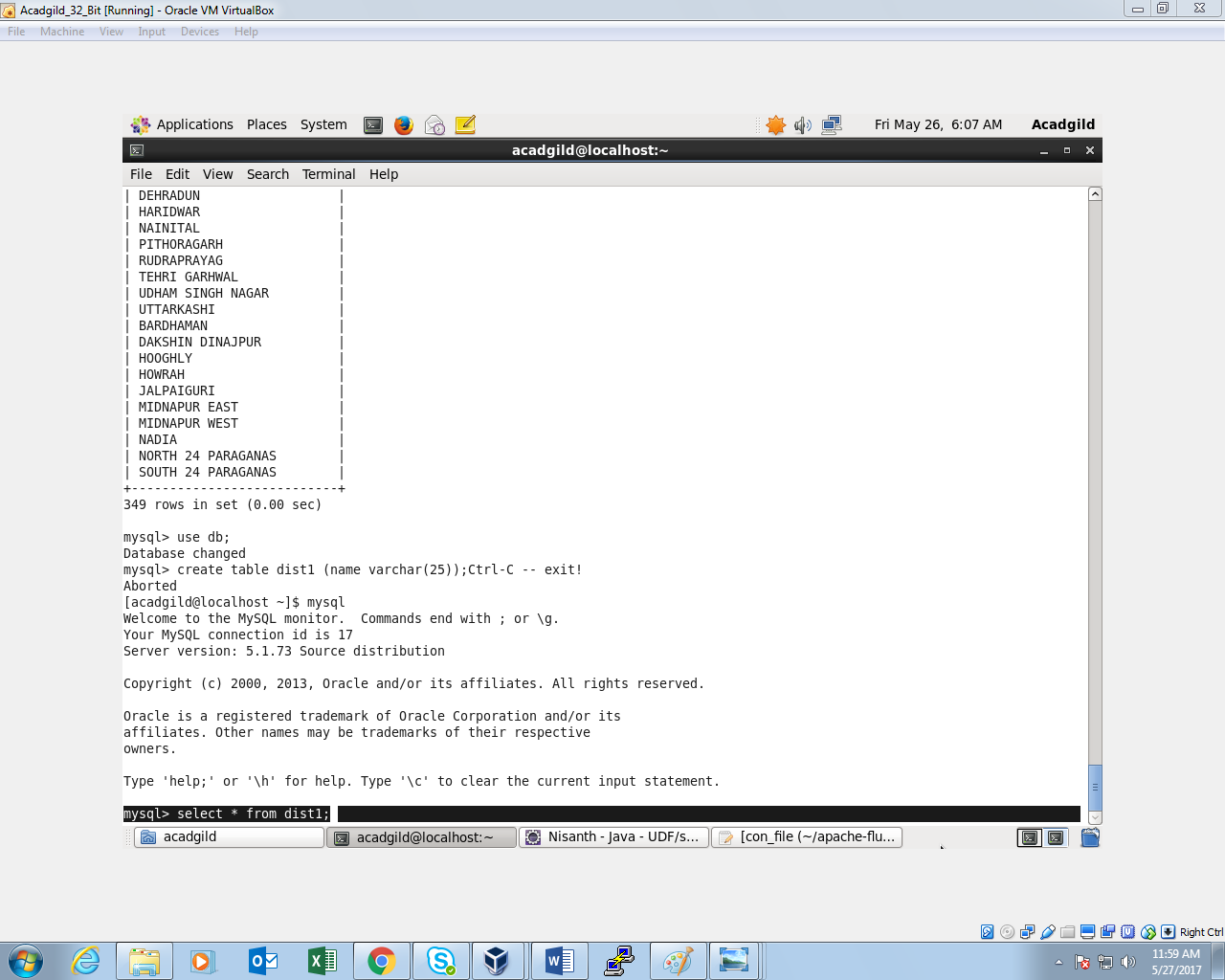
***UDF:***



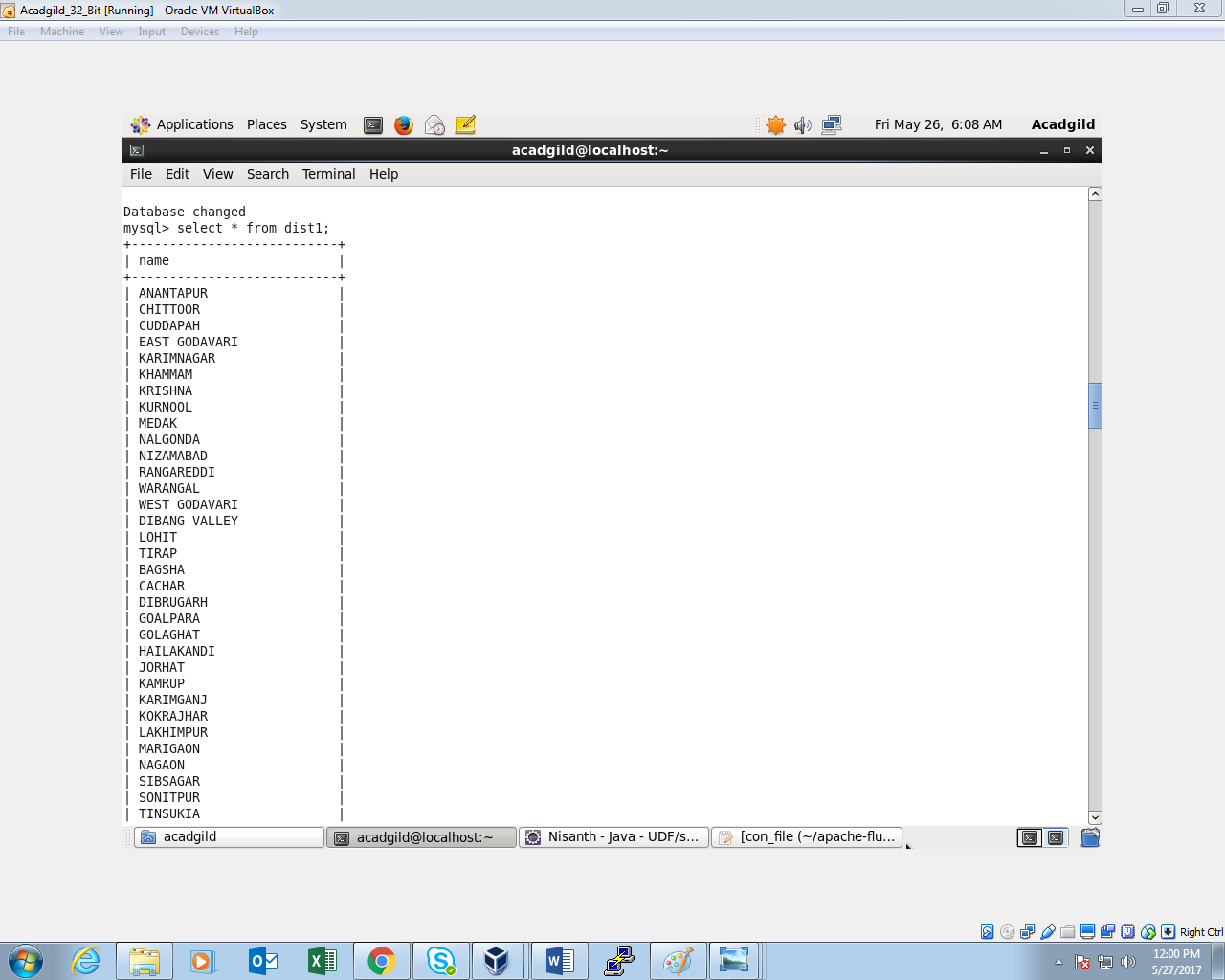


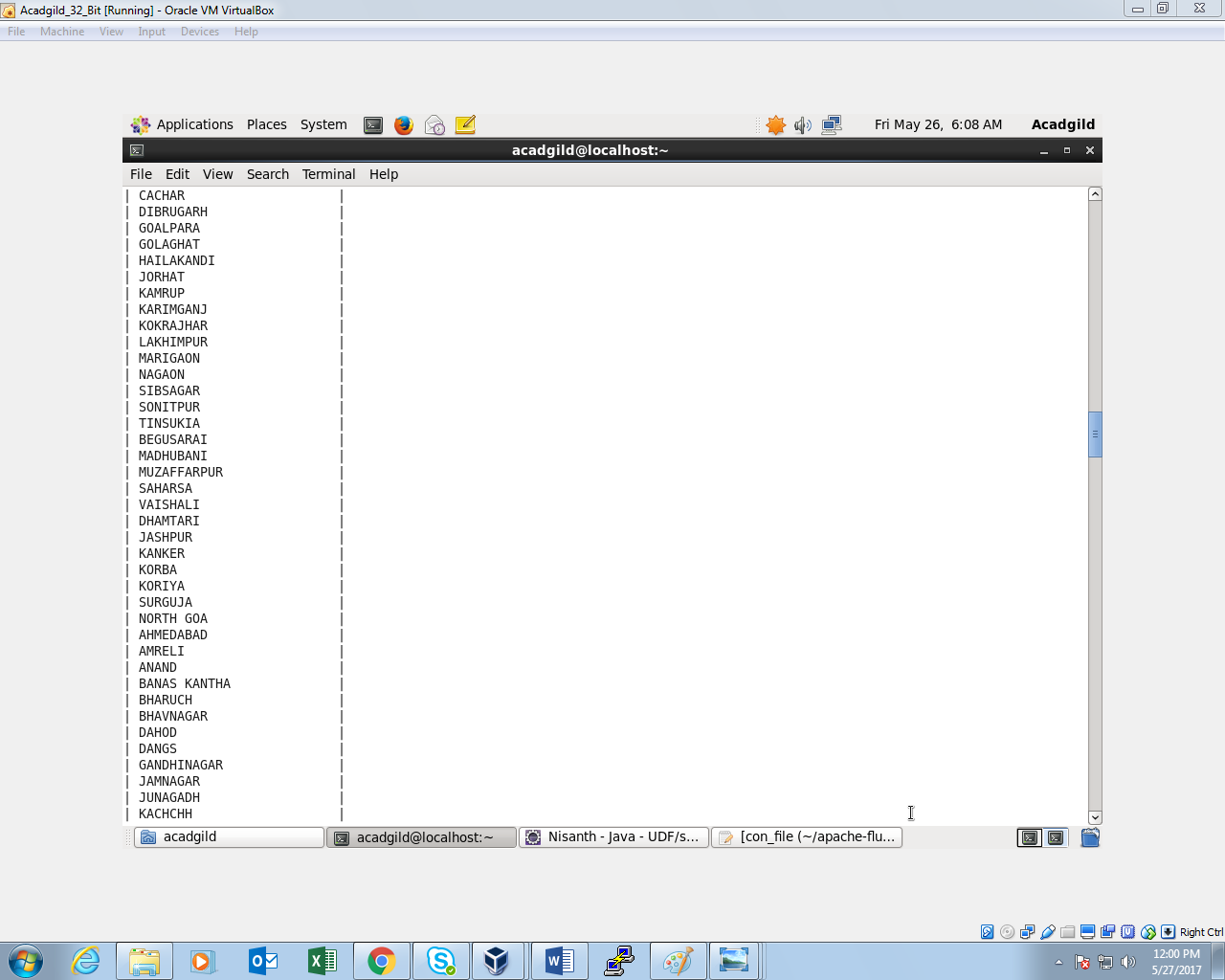


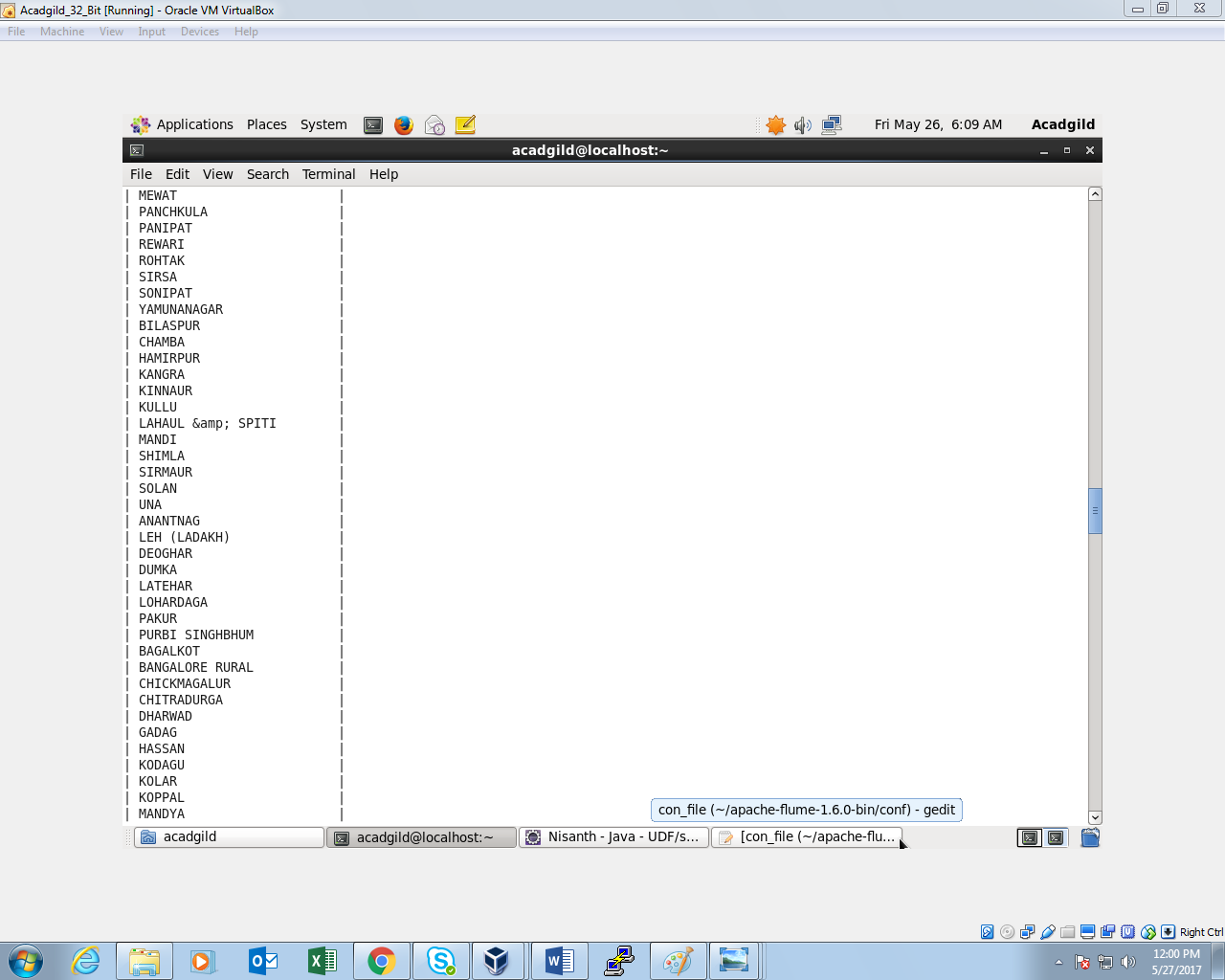


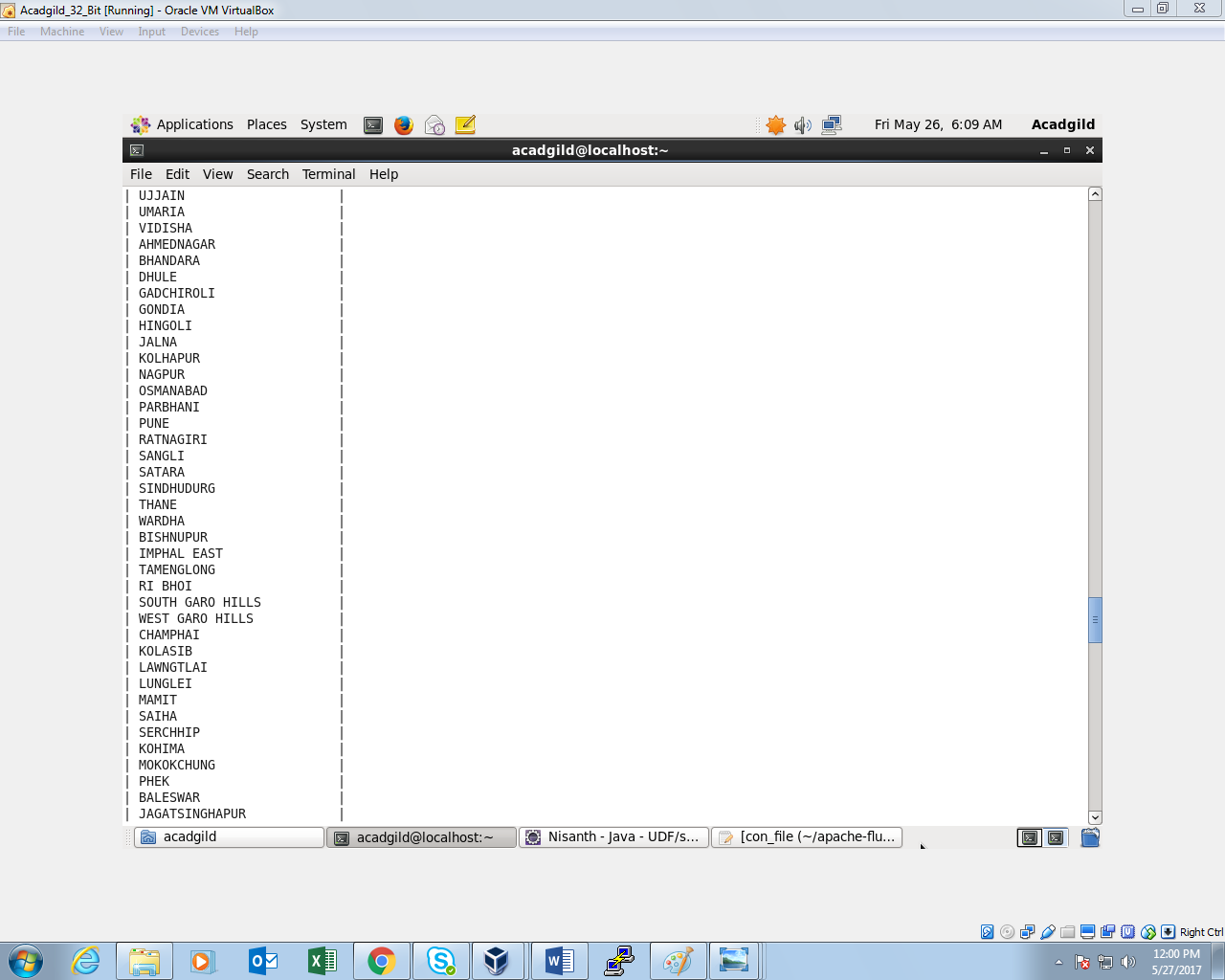


***OUTPUT:***











**Thu, 349 districts are shown which have reached 80% of objectives of BPL cards.**