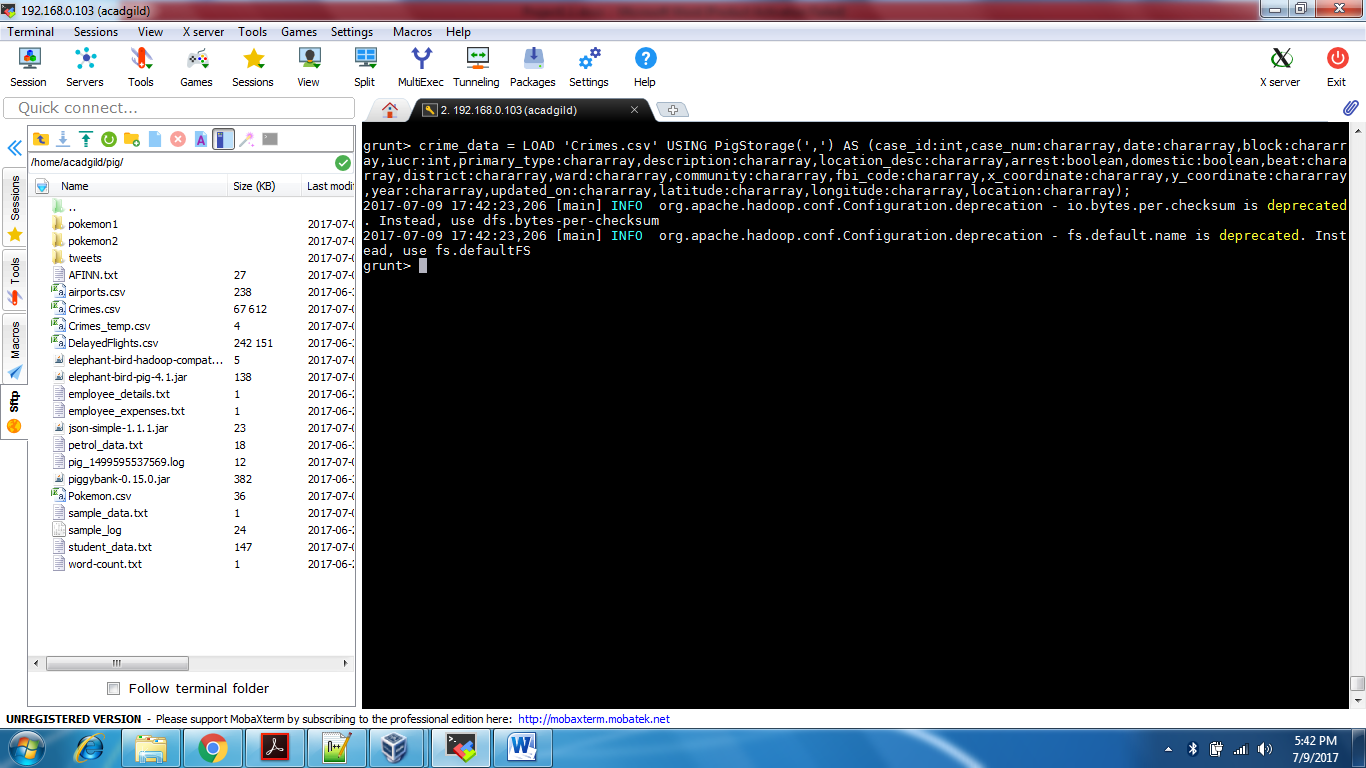
Loading crime data:

At first we are loading the csv file into crime\_data relation

crime\_data = LOAD 'Crimes.csv' USING PigStorage(',') AS (case\_id:int,case\_num:chararray,date:chararray,block:chararray,iucr:int,primary\_type:chararray,description:chararray,location\_desc:chararray,arrest:boolean,domestic:boolean,beat:chararray,district:chararray,ward:chararray,community:chararray,fbi\_code:chararray,x\_coordinate:chararray,y\_coordinate:chararray,year:chararray,updated\_on:chararray,latitude:chararray,longitude:chararray,location:chararray);



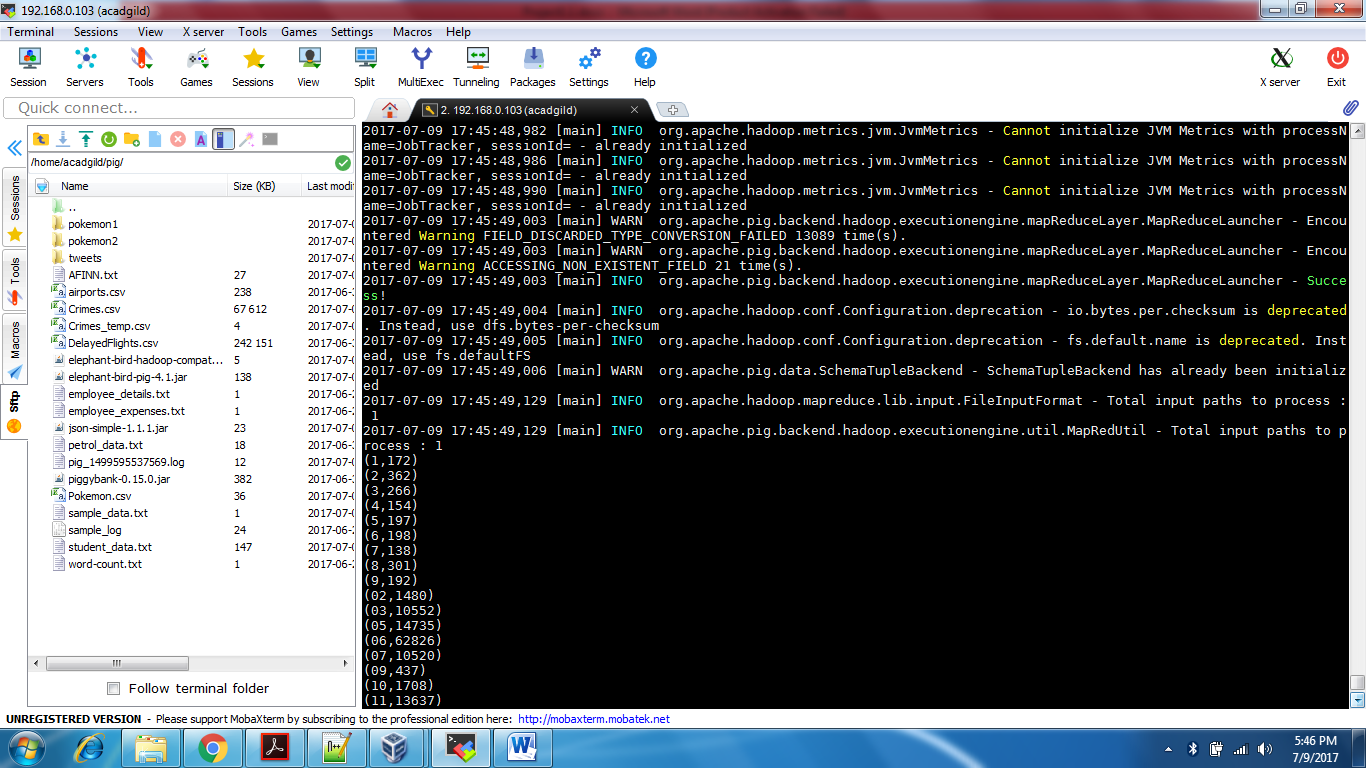
Number of cases investigated under each FBI code:

* Grouping data based on FBICode.
* Counting number of records for each FBICode.
* Getting the result by DUMP command.

grp\_crime\_data = GROUP crime\_data BY fbi\_code;

count\_cases = FOREACH grp\_crime\_data GENERATE group, COUNT(crime\_data);

DUMP count\_cases;



Number of cases investigated under FBI code 32:

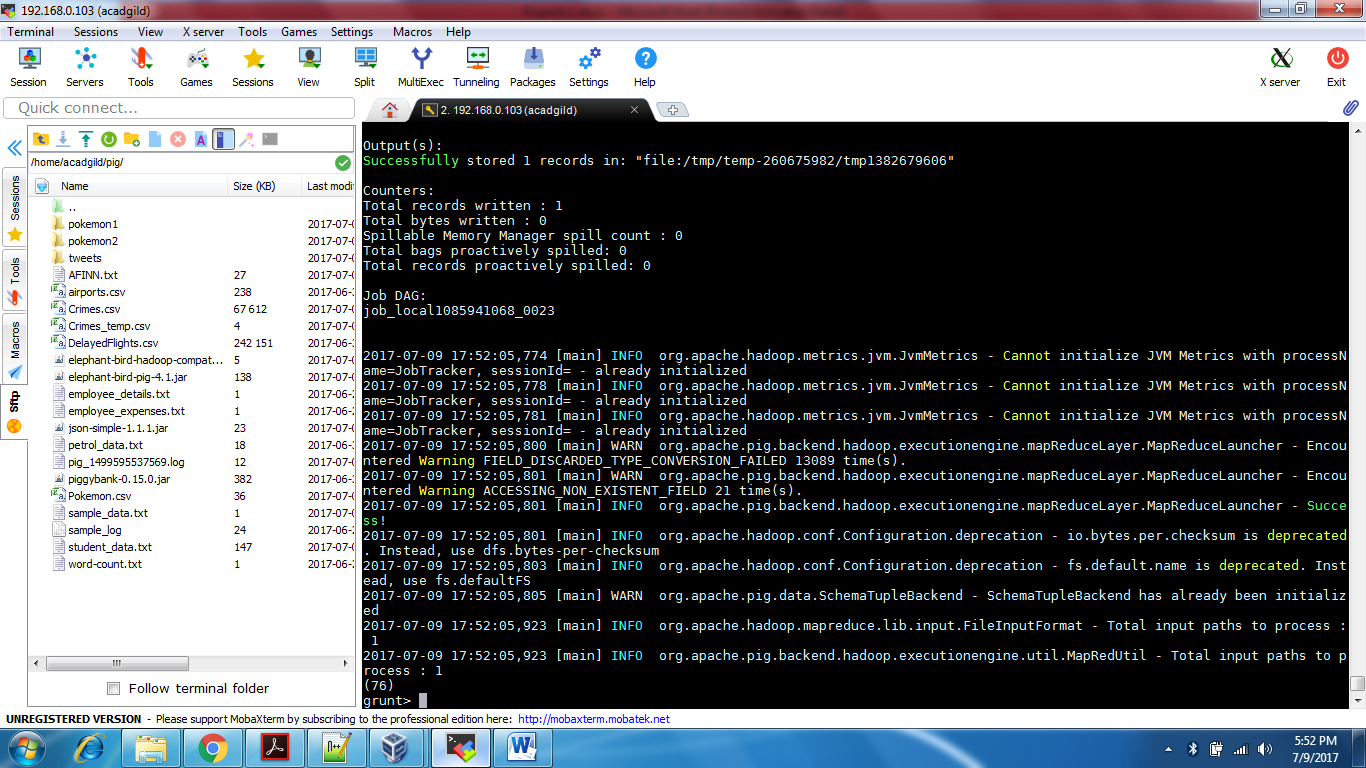
* Filtering data based on FBICode.
* Grouping filtered data by null to get all the records.
* Counting number of records for FBICode 32.
* Getting the result by DUMP command.

filtered\_crime\_data = FILTER crime\_data BY (fbi\_code matches '32');

grp\_crime\_data\_32 = GROUP filtered\_crime\_data BY NULL;

count\_case\_32 = FOREACH grp\_crime\_data\_32 GENERATE COUNT (filtered\_crime\_data);

DUMP count\_case\_32;



Number of arrests in theft district wise:

* Generating Primary Type, Arrest, and District for each record.
* Filtering data based on Primary type as ‘THEFT’ and arrest as true.
* Grouping data based on District.
* Counting number of records for arrests in ‘THEFT’.
* Getting the result by DUMP command.

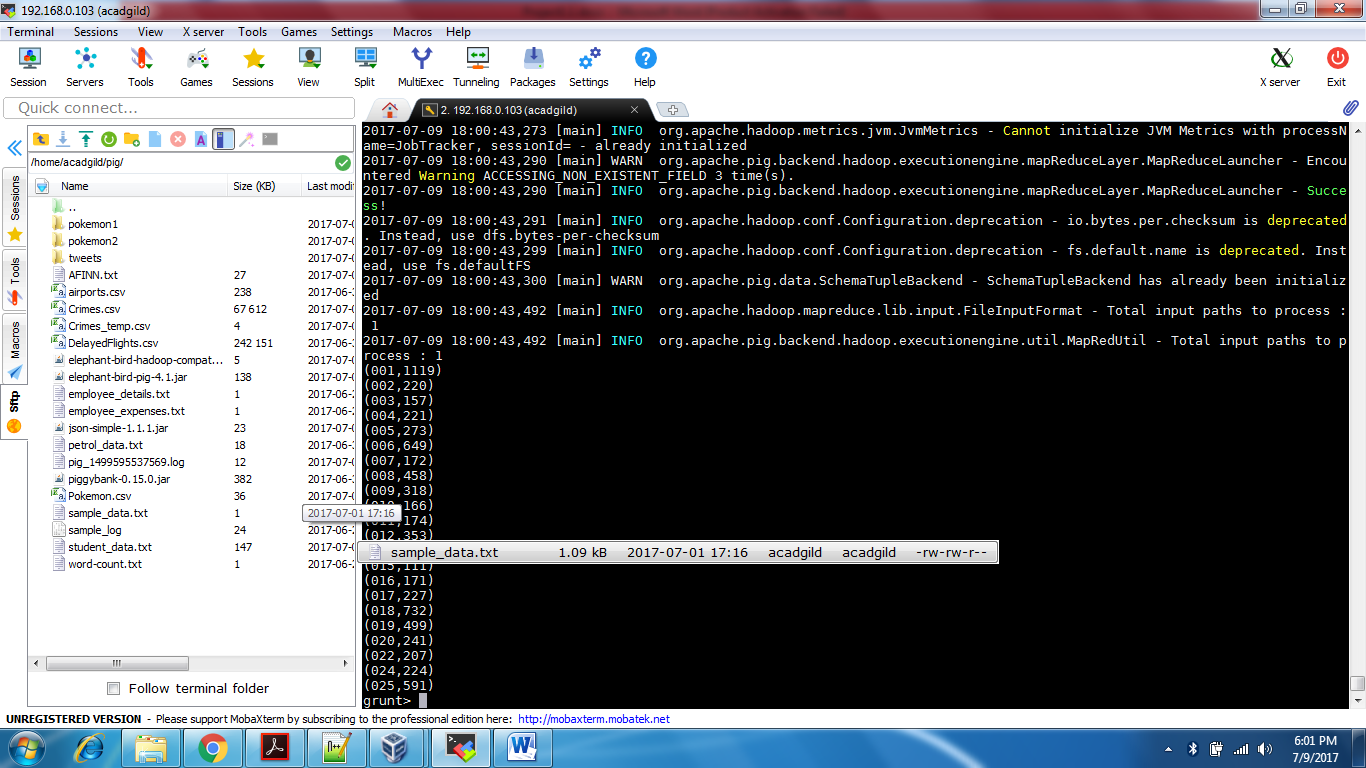
crime\_data\_arrest = FOREACH crime\_data GENERATE primary\_type, arrest, district;

filtered\_crime\_data\_arrest = FILTER crime\_data\_arrest BY ((primary\_type matches 'THEFT') AND (arrest == true));

grp\_crime\_data\_arrest = GROUP filtered\_crime\_data\_arrest BY district;

count\_theft\_arrest = FOREACH grp\_crime\_data\_arrest GENERATE group, COUNT (filtered\_crime\_data\_arrest);

DUMP count\_theft\_arrest;



Number of arrests done between October 2014 and October 2015:

* Generating Date, Arrest for each record.

1. Here first SUBSTRING function has been used to neglect (AM/PM).
2. Then we gave converted the date from chararray to datetime format.

* Filtering data based on the desired duration and arrest as true.
* Grouping filtered data by null to get all the records.
* Counting number of records for arrests between OCT 2014 and OCT 2015.
* Getting the result by DUMP command.

crime\_data\_date = FOREACH crime\_data GENERATE ToDate (SUBSTRING (date, 0, 19),'mm/dd/yyyy HH:mm:ss') AS fdate, arrest;

filtered\_crime\_data\_date = FILTER crime\_data\_date BY ((ToDate ('2014-09-30') < fdate) AND (fdate < ToDate ('2015-11-01') AND (arrest == true)));

grp\_crime\_data\_date = GROUP filtered\_crime\_data\_date BY NULL;

count\_arrest = FOREACH grp\_crime\_data\_date GENERATE COUNT (filtered\_crime\_data\_date);

Dump count\_arrest;

