**SESSION 10-11: PROJECT I**

**USA Crime Analysis**

**Processing Data through PIG:**

REGISTER /usr/local/pig/lib/piggybank.jar;

data = LOAD '/user/acadgild/crimes.csv' USING

org.apache.pig.piggybank.storage.CSVExcelStorage() AS (caseid:chararray,

caseno:chararray, date:chararray, block:chararray, iucr:chararray, type:chararray,

desc:chararray, locdesc:chararray, arrest:chararray, domestic:chararray, beat:chararray,

district:chararray, ward:chararray, commarea:chararray, fbicode:chararray, x:chararray,

y:chararray, year:chararray, updatedon:chararray, lat:chararray, long:chararray,

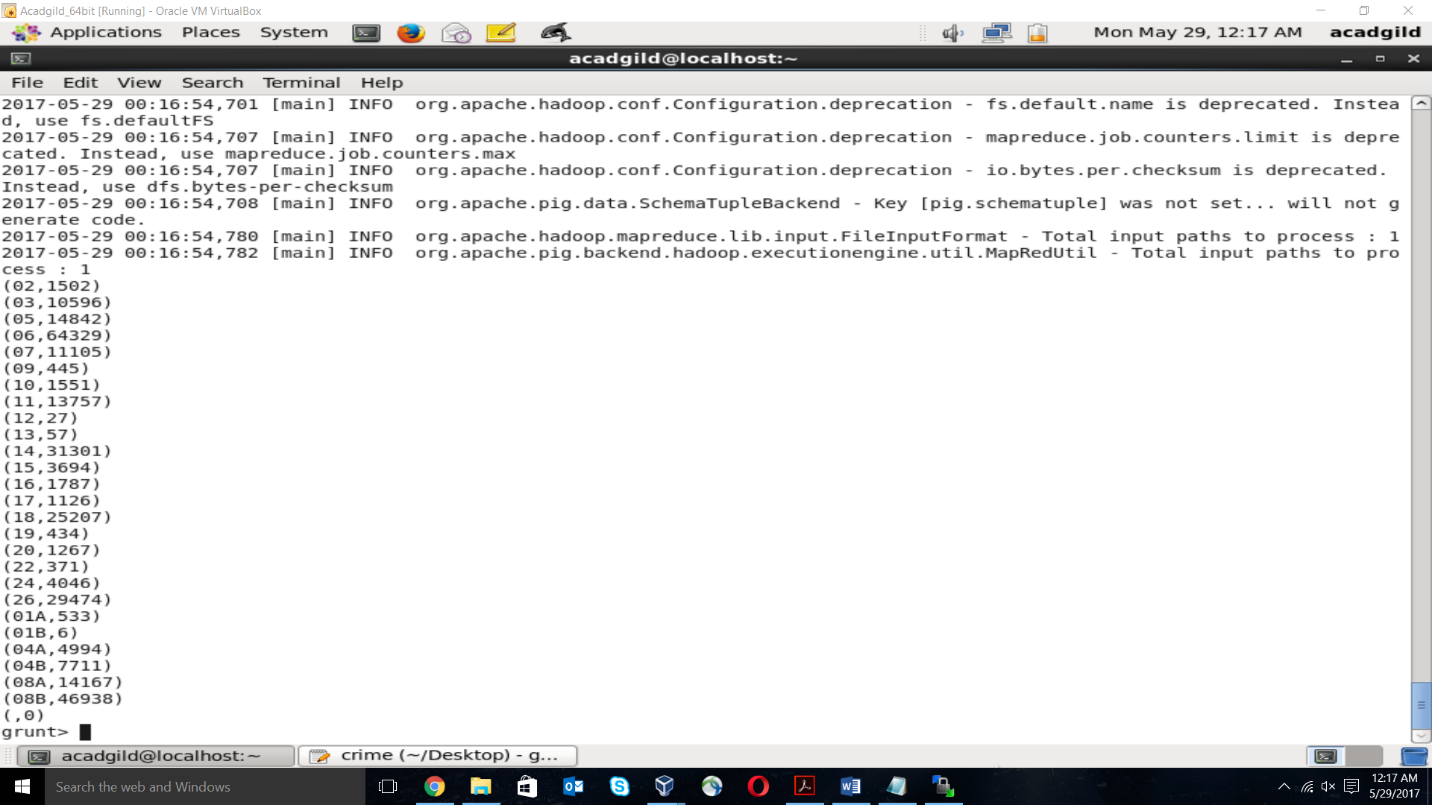
loc:chararray);

1. **Write a MR/Pig program to calculate the number of cases investigated under each FBI code.**

group\_data = GROUP data BY fbicode;

result = FOREACH group\_data GENERATE group as code, COUNT(data.caseno) as count;

dump result;



1. **Write a MR/Pig program to calculate the number of cases investigated under Ward number 32.**

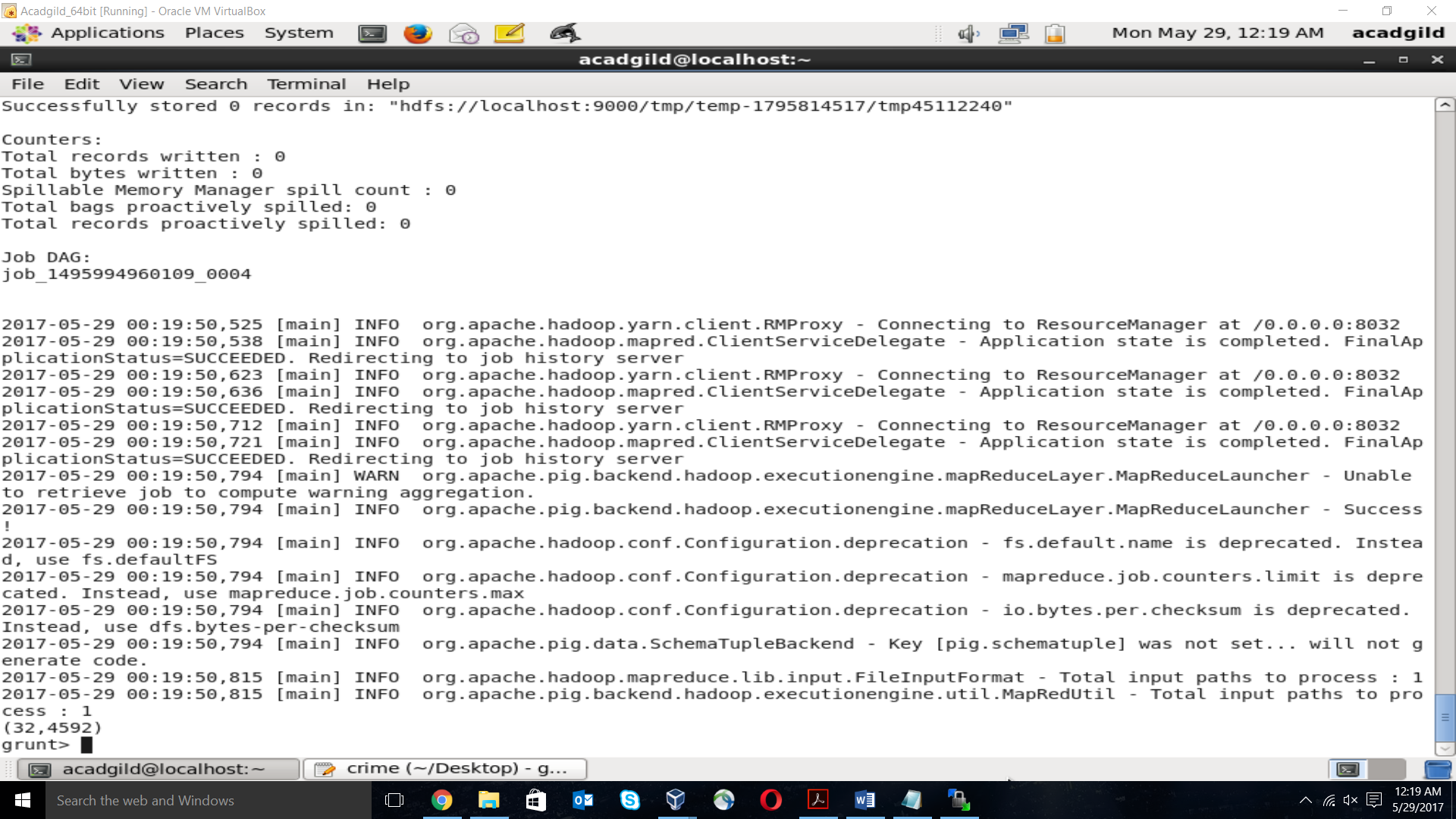
filter\_data = FILTER data by (ward == '32');

group\_data = GROUP filter\_data by ward;

result = FOREACH group\_data GENERATE group as ward, COUNT(filter\_data.caseno) as

count;

dump result;



1. **Write a MR/Pig program to calculate the number of arrests in theft district wise.**

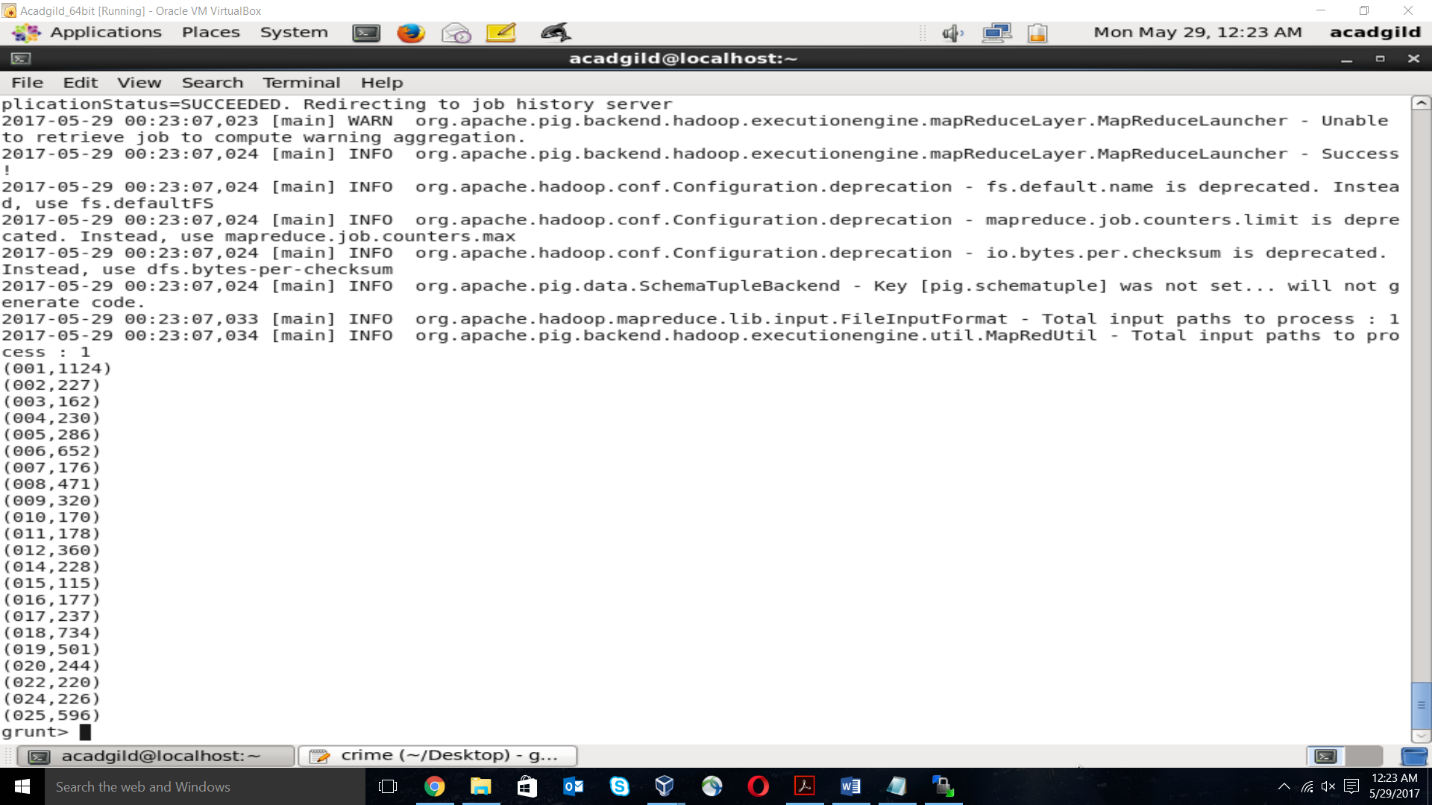
filter\_data = FILTER data BY ((type == 'THEFT') AND (arrest == 'true'));

group\_data = GROUP filter\_data BY district;

result = FOREACH group\_data GENERATE group as disctrict, COUNT(filter\_data.caseno)

as count;

dump result;



1. **Write a MR/Pig program to calculate the number of arrests done between October 2014 and October 2015.**

ts\_data = foreach data generate ToDate(date,'MM/dd/yyyy hh:mm:ss aa') as date,

arrest;

dt\_data = FOREACH ts\_data GENERATE GetYear(date) as year, GetMonth(date) as

month, arrest;

filter\_data = FILTER dt\_data BY ((year == 2014 AND month >= 10) OR (year == 2015 AND

month <= 10)) AND (arrest == 'true');

grp\_data = GROUP filter\_data ALL;

result = FOREACH grp\_data GENERATE COUNT(filter\_data);

dump result;

