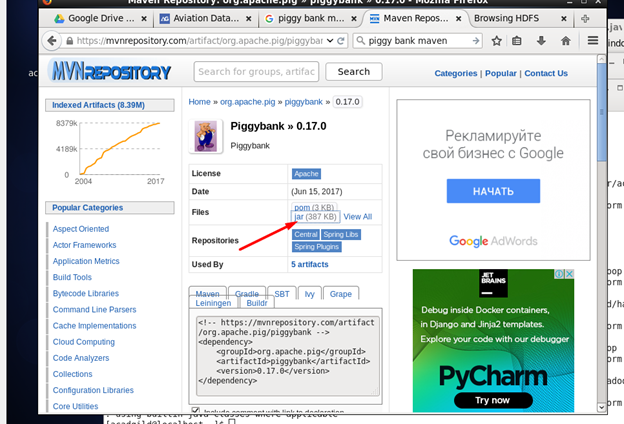
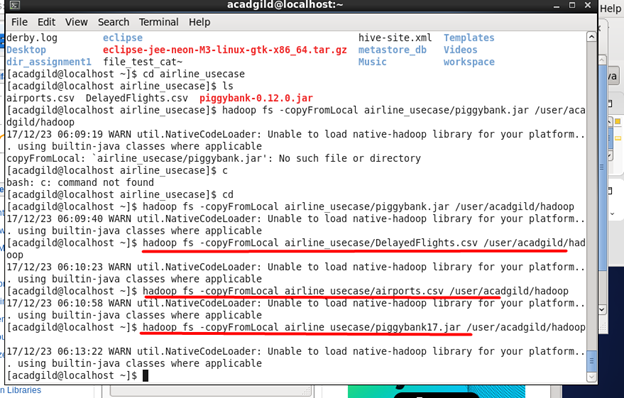
Session 5: Assignment 2

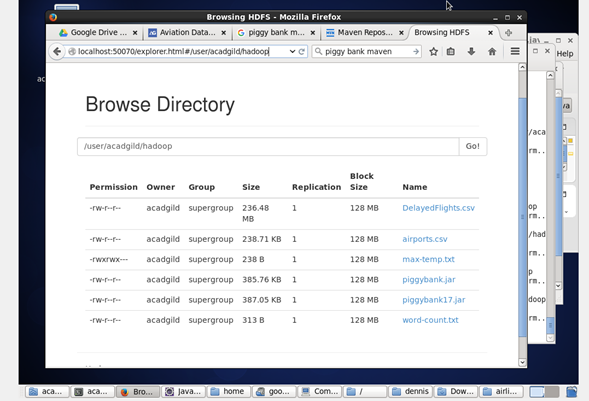
Download piggybank



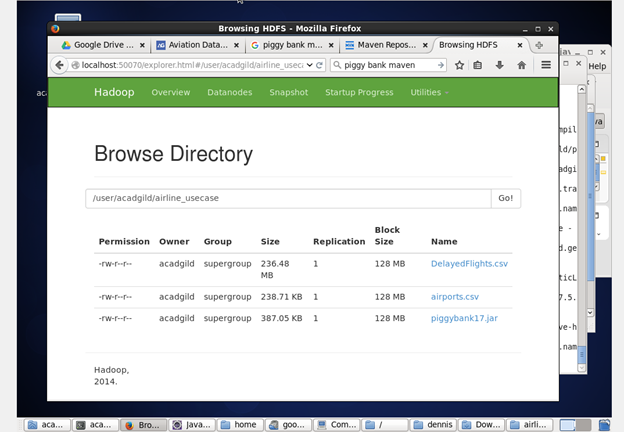
Move the jar to hdfs along with data for the assignment

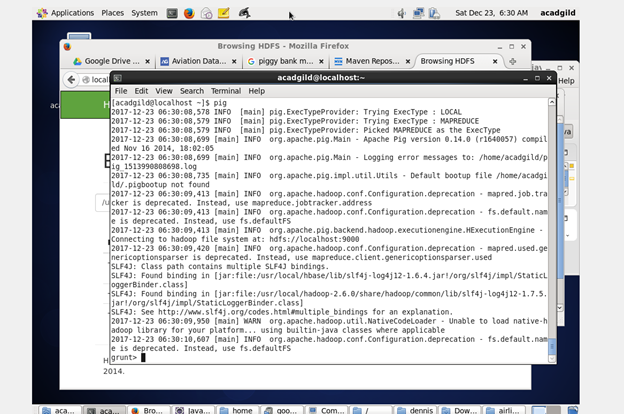


Check that data appears in the expected hdfs folder



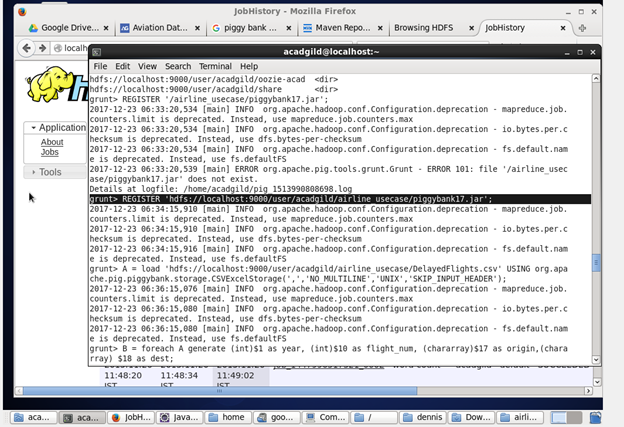
For more comfort create dir in the hdfs and move the files here :

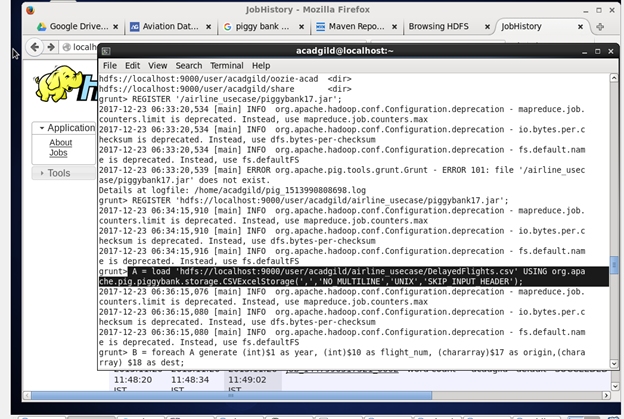


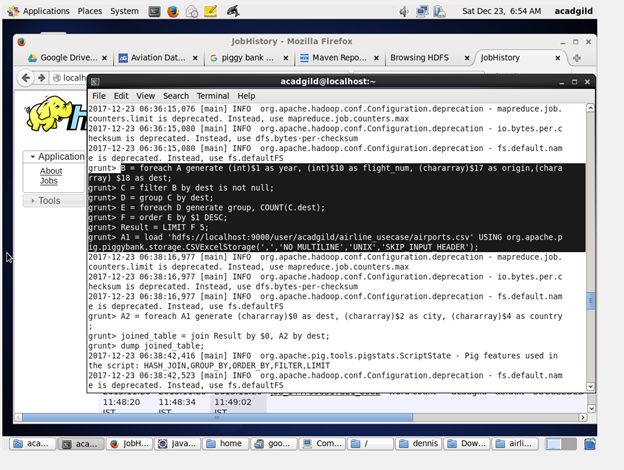


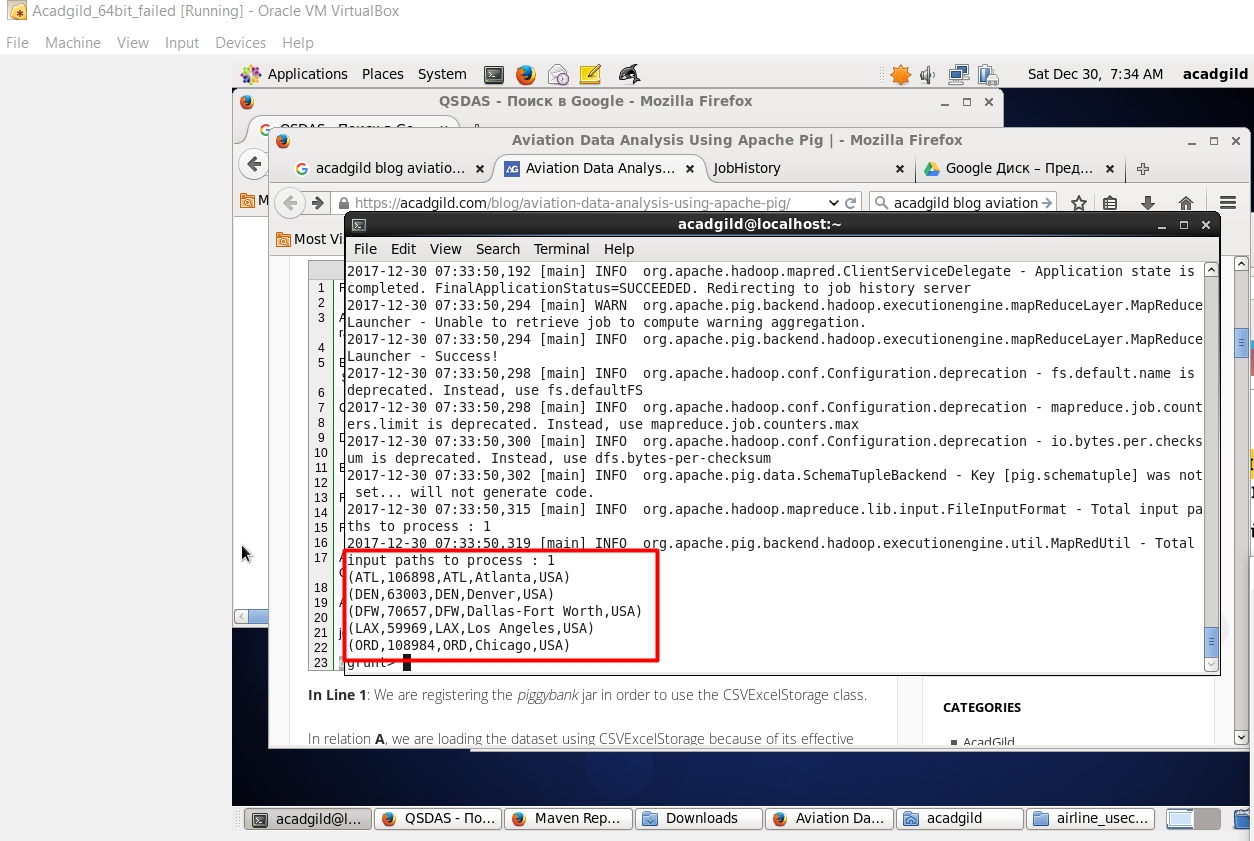
**Problem 1**

Run script from blog line by line





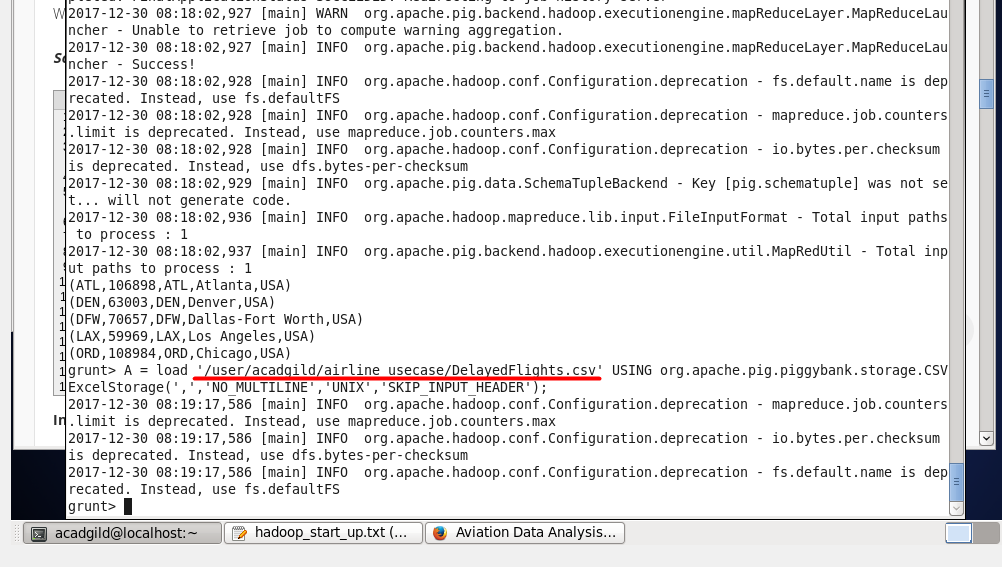




The image above is final result of problem statement 1

**Problem Statement 2**

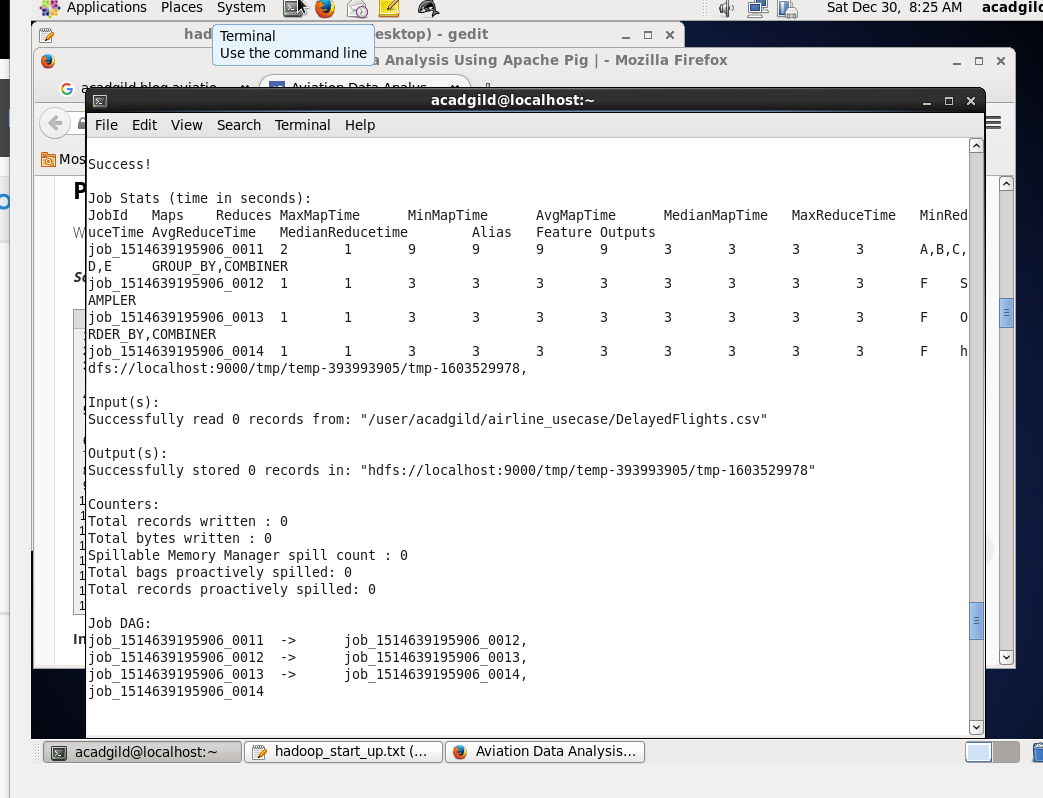
1. Load first piece of input data

****

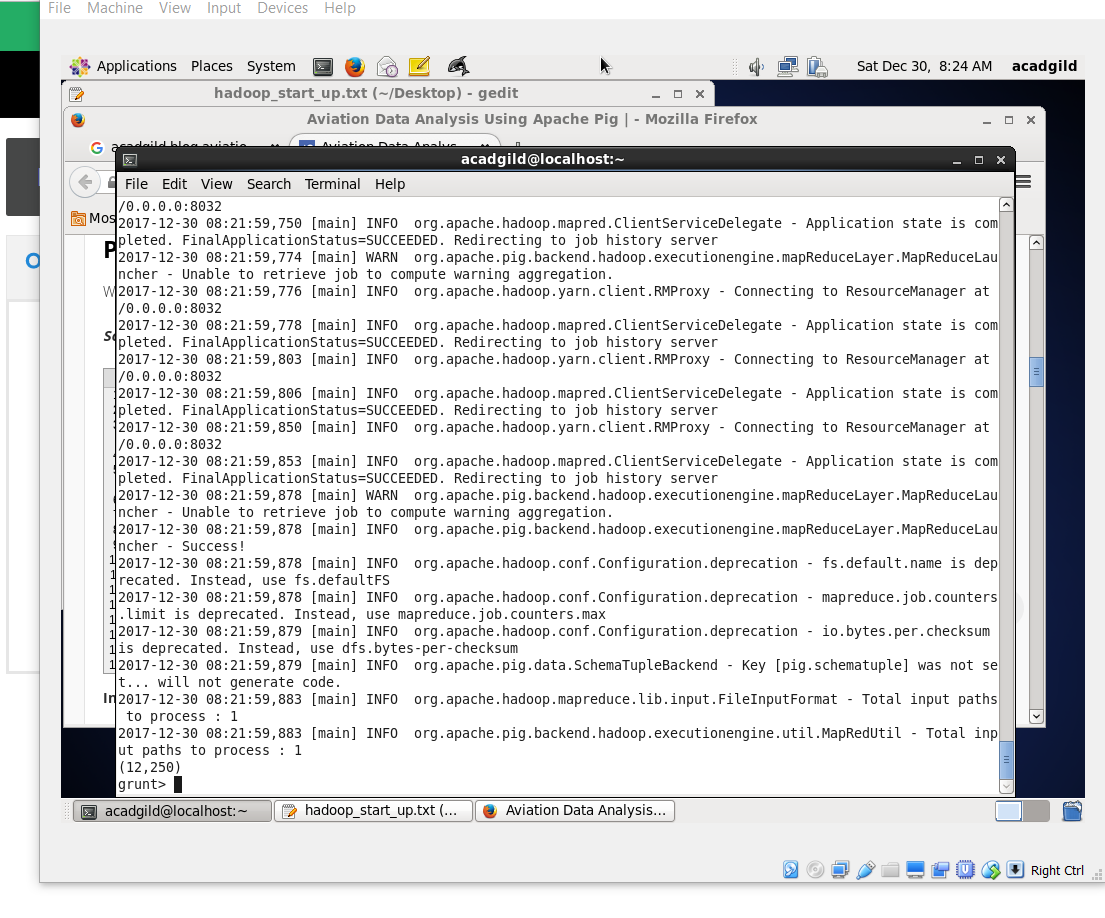
2. Creating local variables, job business logic for hadoop execution and load second piece of input data

3. Determine output format , determine details of joint operation around input data after the manipulation ( in details, the manipulations are described in the corresponding acadgild blog’s article ) shown above and run mapReduce job.

And receiving some intermediate results

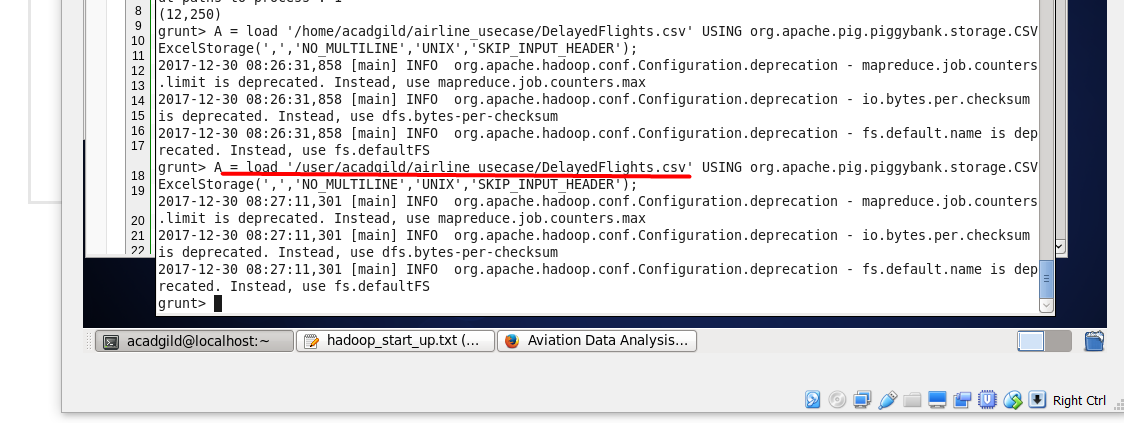
****

4. Receiving the final results

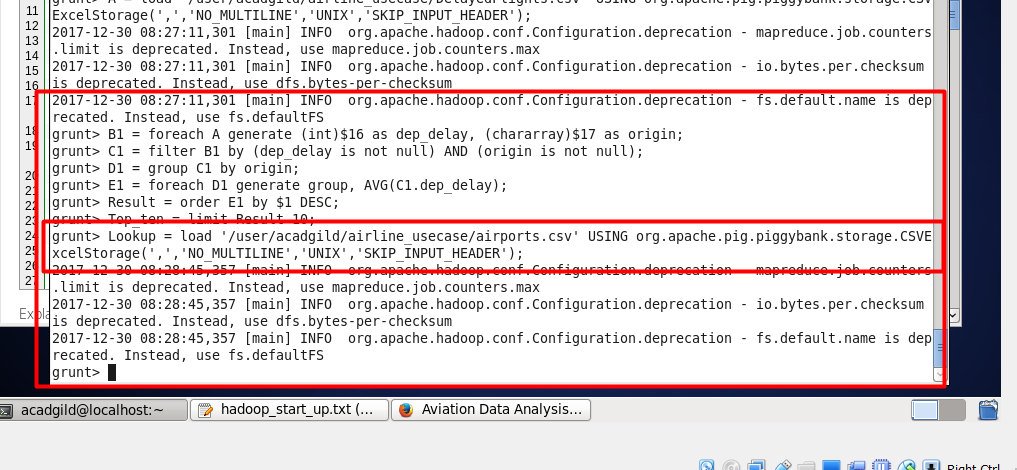
****

**Problem 3**

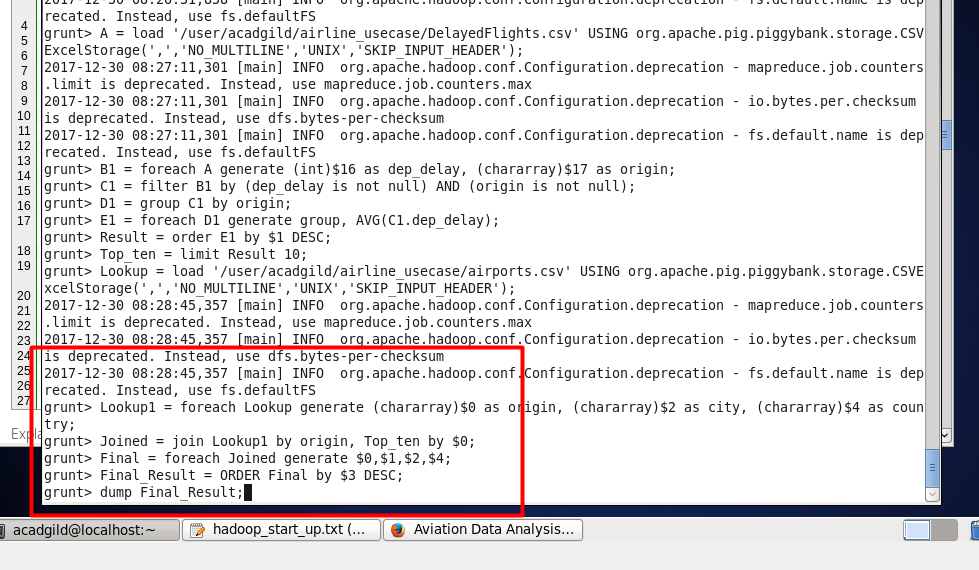
1. Load first piece of input data

****

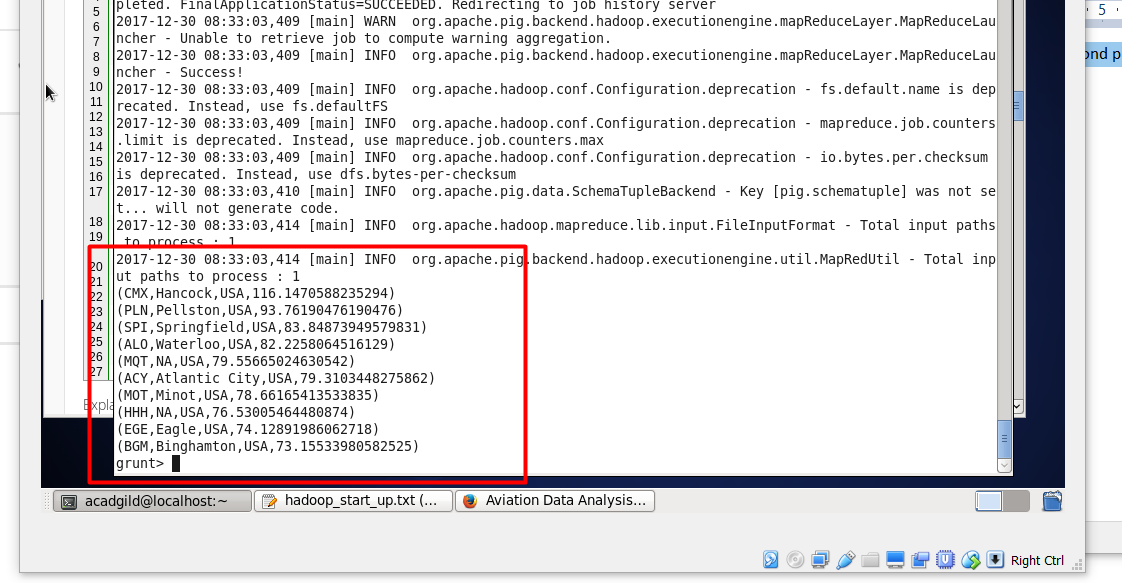
2. Creating local variables, job business logic for hadoop execution and load second piece of input data

****

3. Determine output format , determine details of joint operation around input data after the manipulation ( in details, the manipulations are described in the corresponding acadgild blog’s article ) shown above and run mapReduce job.

****

4. Receiving the final results

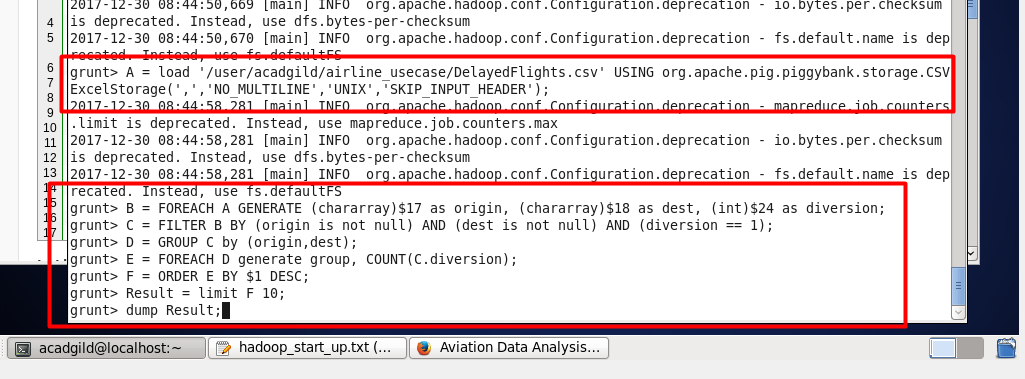
****

**Problem 4**

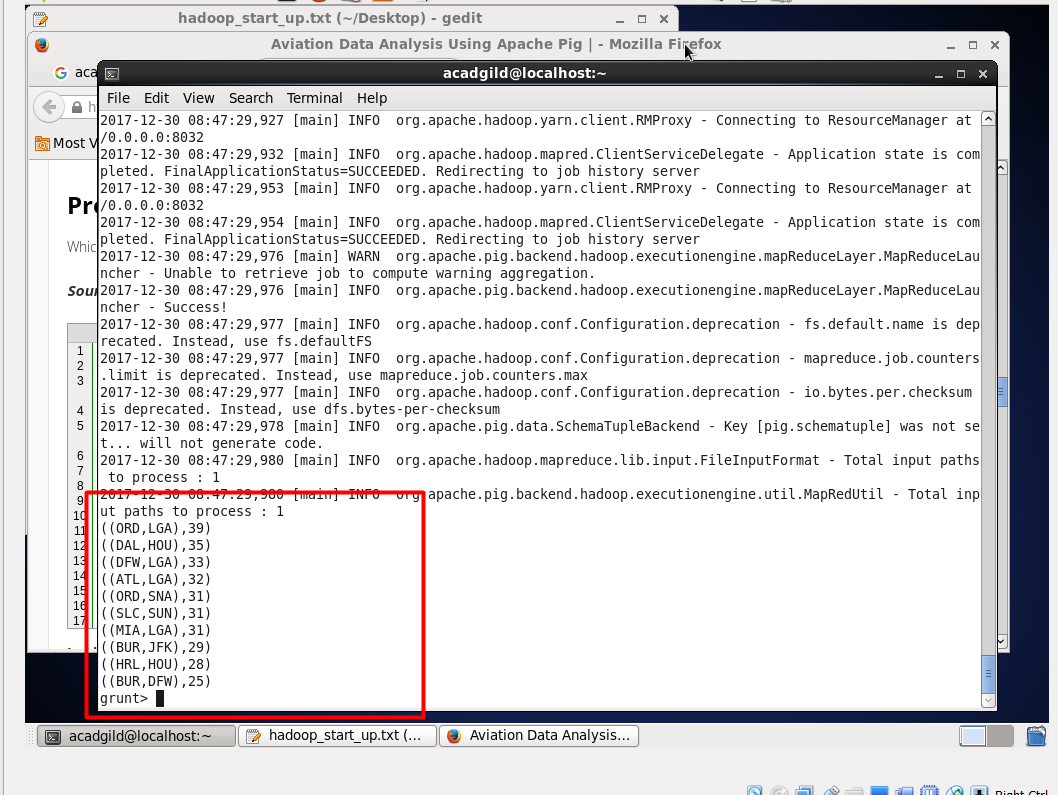
1. Load first piece of input data

2. Creating local variables, job business logic for hadoop execution and load second piece of input data

3. Determine output format , determine details of joint operation around input data after the manipulation ( in details, the manipulations are described in the corresponding acadgild blog’s article ) shown above and run mapReduce job.

****

4. Receiving the final results

****