1)

REGISTER '/home/acadgild/pig/piggybank-0.15.0.jar';

delayed\_flights\_data = LOAD 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

flight\_details = FOREACH delayed\_flights\_data GENERATE (int)$1 as year, (int)$10 as flight\_num, (chararray)$17 as origin,(chararray) $18 as destination;

filtered\_flight\_details = FILTER flight\_details by destination is not null;

grp\_flight\_details = GROUP filtered\_flight\_details by destination;

destination\_count = FOREACH grp\_flight\_details GENERATE group, COUNT(filtered\_flight\_details.destination);

ordered\_destination\_count = ORDER destination\_count by $1 DESC;

top5\_destinations\_code = LIMIT ordered\_destination\_count 5;

airports\_data = LOAD 'airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

airports\_details = FOREACH airports\_data GENERATE (chararray)$0 as destination, (chararray)$2 as city, (chararray)$4 as country;

top5\_most\_visited\_destinations = JOIN top5\_destinations\_code by $0, airports\_details by destination;

DUMP top5\_most\_visited\_destinations;

2)

REGISTER '/home/acadgild/pig/piggybank-0.15.0.jar';

delayed\_flights\_data = LOAD 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

flight\_details = FOREACH delayed\_flights\_data GENERATE (int)$2 as month,(int)$10 as flight\_num,(int)$22 as cancelled,(chararray)$23 as cancel\_code;

cancelled\_flights = FILTER flight\_details by cancelled == 1 AND cancel\_code =='B';

grp\_cancelled\_flights = GROUP cancelled\_flights by month;

cancelled\_flights\_bymonth\_count = FOREACH grp\_cancelled\_flights GENERATE group, COUNT(cancelled\_flights.cancelled);

cancelled\_flights\_bymonth\_ordered = ORDER cancelled\_flights\_bymonth\_count by $1 DESC;

month\_max\_flights\_cancelled = LIMIT cancelled\_flights\_bymonth\_ordered 1;

DUMP month\_max\_flights\_cancelled;

3)

REGISTER '/home/acadgild/pig/piggybank-0.15.0.jar';

delayed\_flights\_data = LOAD 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

flight\_details = FOREACH delayed\_flights\_data GENERATE (int)$16 as dep\_delay, (chararray)$17 as origin;

filtered\_flight\_details = FILTER flight\_details by (dep\_delay is not null) AND (origin is not null);

grp\_flight\_details = GROUP filtered\_flight\_details by origin;

origins\_with\_avg\_dep\_delay = FOREACH grp\_flight\_details GENERATE group, AVG(filtered\_flight\_details.dep\_delay);

ordered\_origins\_with\_delay = ORDER origins\_with\_avg\_dep\_delay by $1 DESC;

top10\_origins = LIMIT ordered\_origins\_with\_delay 10;

airports\_data = LOAD 'airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

airport\_details = FOREACH airports\_data GENERATE (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;

top10\_origins\_details = JOIN airport\_details by origin, top10\_origins by $0;

DUMP top10\_origins\_details;

4)

REGISTER '/home/acadgild/pig/piggybank-0.15.0.jar';

delayed\_flights\_data = LOAD 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

flight\_details = FOREACH delayed\_flights\_data GENERATE (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;

filtered\_flight\_details = FILTER flight\_details BY (origin is not null) AND (dest is not null) AND (diversion == 1);

grp\_flight\_details = GROUP filtered\_flight\_details by (origin,dest);

route\_diversion = FOREACH grp\_flight\_details GENERATE group, COUNT(filtered\_flight\_details.diversion);

ordered\_route\_diversion = ORDER route\_diversion BY $1 DESC;

max\_route\_diversion = LIMIT ordered\_route\_diversion 10;

DUMP max\_route\_diversion;