**Index:**

|  |  |  |
| --- | --- | --- |
| S.No | Operator | Page No. |
| 1 | Load datasets | 2 |
| 2 | Filter(basic) | 5 |
| 3 | Filter & Foreach | 6 |
| 4 | Bagging | 7 |
| 5 | Group & Group all | 8 |
| 6 | Split | 9 |
| 7 | Store | 10 |
| 8 | Order | 11 |
| 9 | Join | 12 |
| 10 | Bag | 13 |

**PIG**

**Ravella Abhinav**

**CB.EN.U4CSE19453**

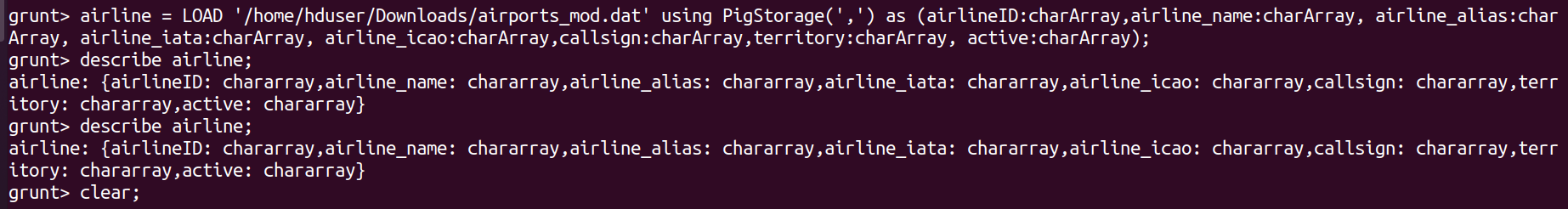
**Load datasets:**

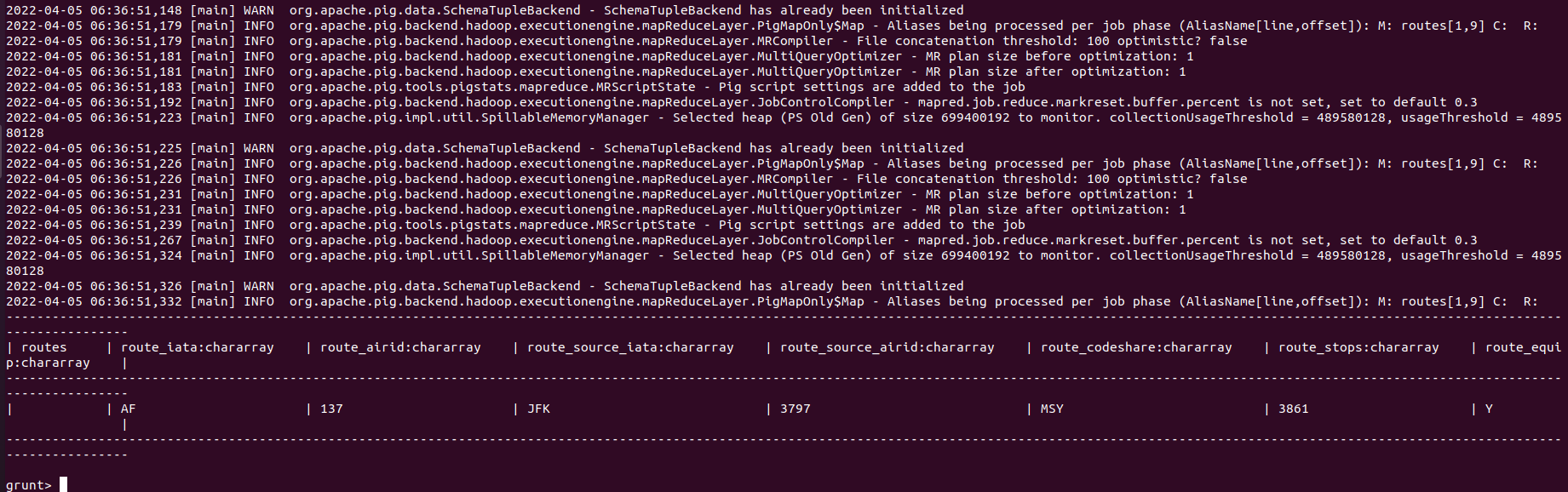
**Airlines:**

**Query:**

airline = LOAD '/home/hduser/Downloads/airports\_mod.dat' using PigStorage(',') as (airlineID:charArray,airline\_name:charArray, airline\_alias:charArray, airline\_iata:charArray, airline\_icao:charArray,callsign:charArray,territory:charArray, active:charArray);

**Output:**



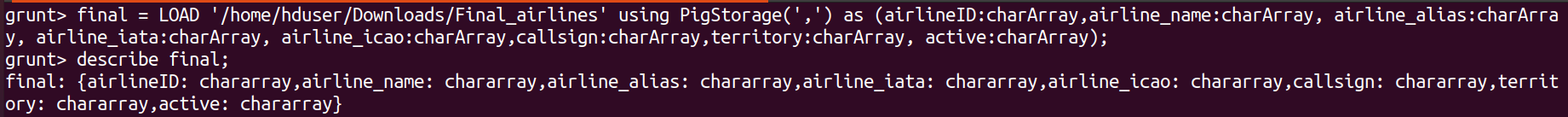


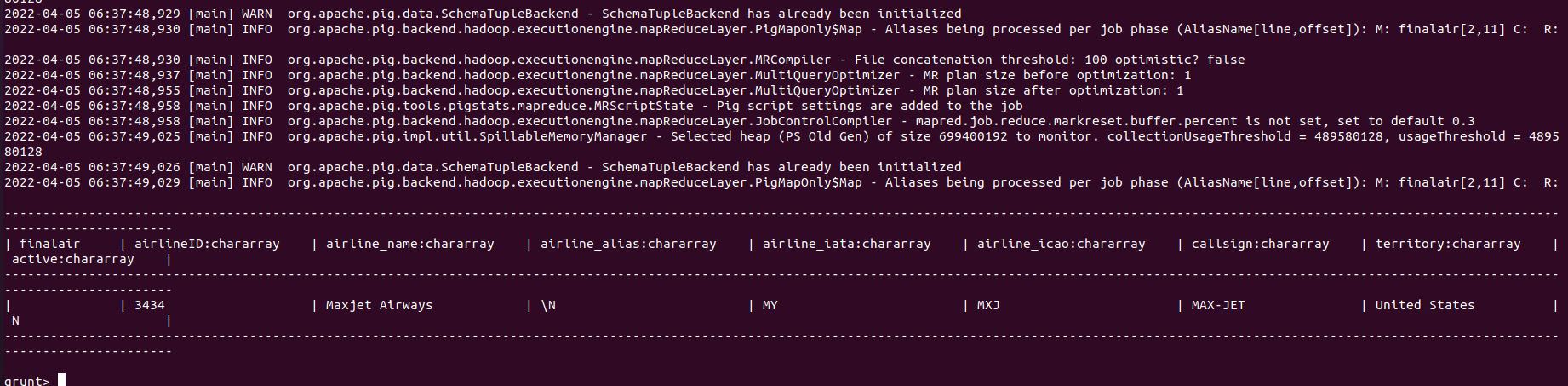
**Final Airlines:**

**Query:**

final = LOAD '/home/hduser/Downloads/Final\_airlines' using PigStorage(',') as (airlineID:charArray,airline\_name:charArray, airline\_alias:charArray, airline\_iata:charArray, airline\_icao:charArray,callsign:charArray,territory:charArray, active:charArray);

**Output:**

****

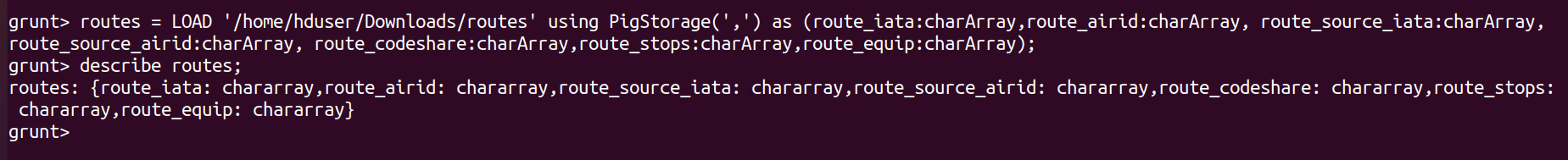
****

**Routes:**

**Query:**

routes = LOAD '/home/hduser/Downloads/routes' using PigStorage(',') as (route\_iata:charArray, route\_airid:charArray, route\_source\_iata:charArray, route\_source\_airid:charArray, route\_codeshare:charArray, route\_stops:charArray, route\_equip:charArray);

**Output:**

****

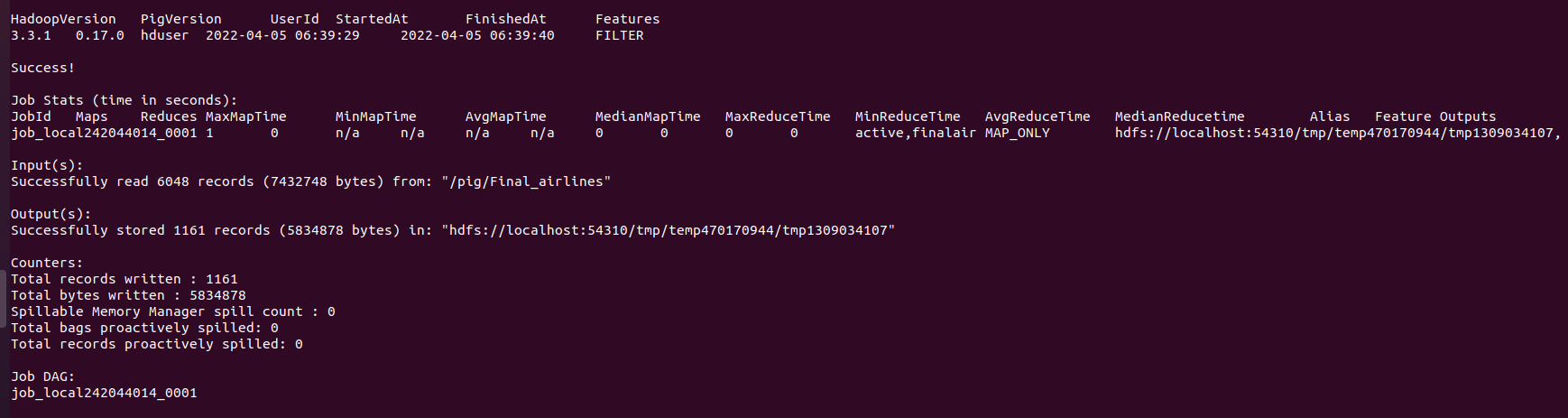
****

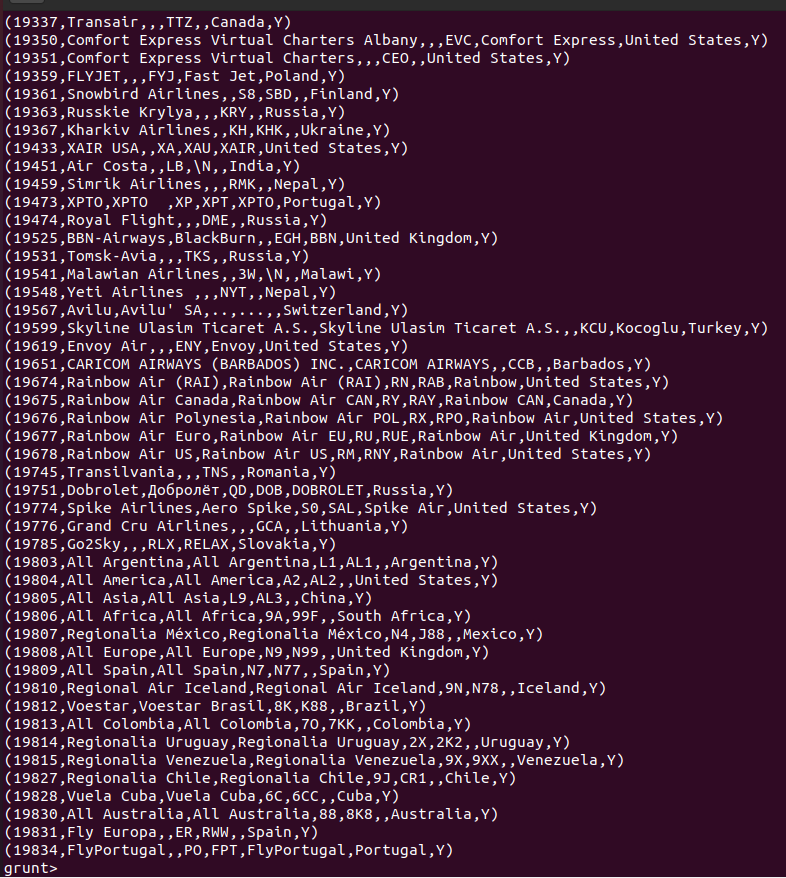
**Queries:**

1. **Find the airlines that are active:**

* active = filter final by Active == 'Y';
* limit = LIMIT active 50;
* dump active;
* dump limit;

**Output:**

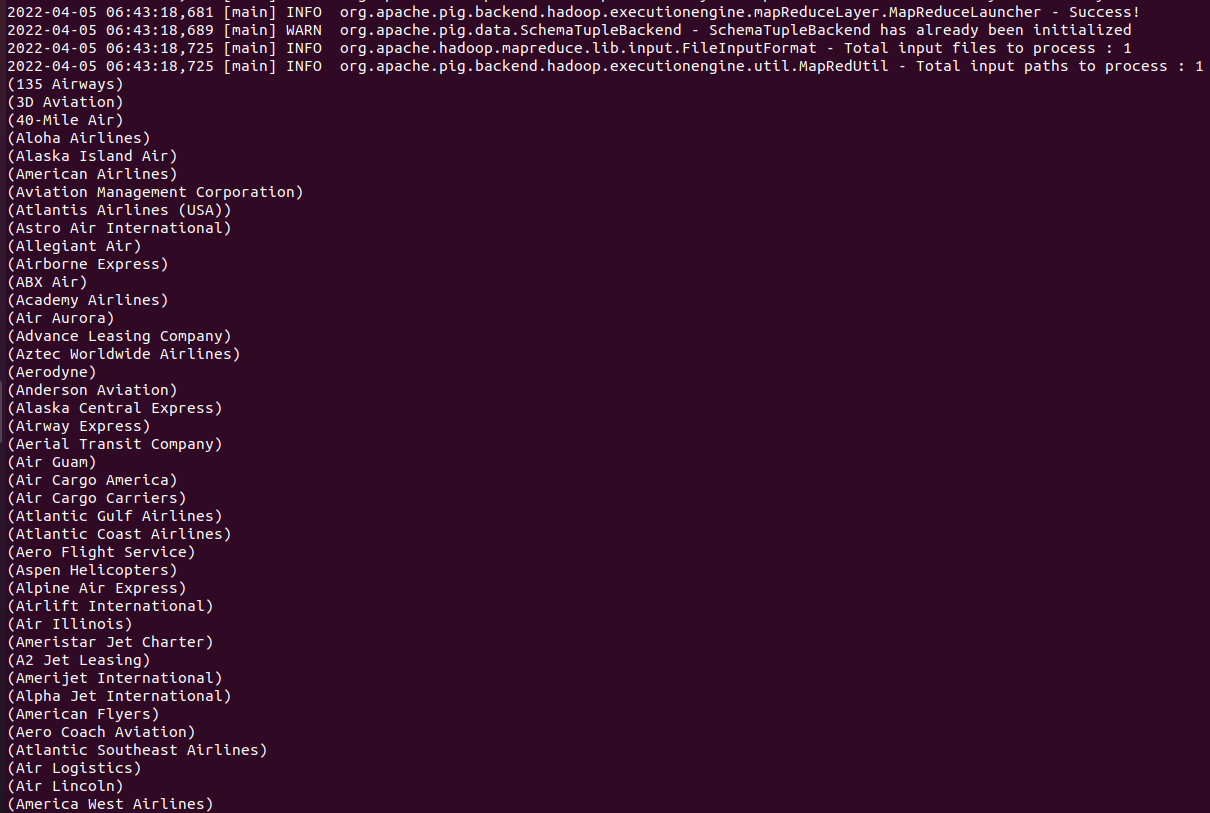


****

1. **Find the names of active airlines in territory ‘Unites States’?**

Query:

* active\_airlines = filter finalair by active == 'Y';
* active\_airlines\_usa = filter active\_airlines by territory == 'United States';
* find\_active\_airline\_names\_in\_usa = foreach active\_airlines\_usa generate airline\_name;
* dump find\_active\_airline\_names\_in\_usa;

****

**Explanation:**

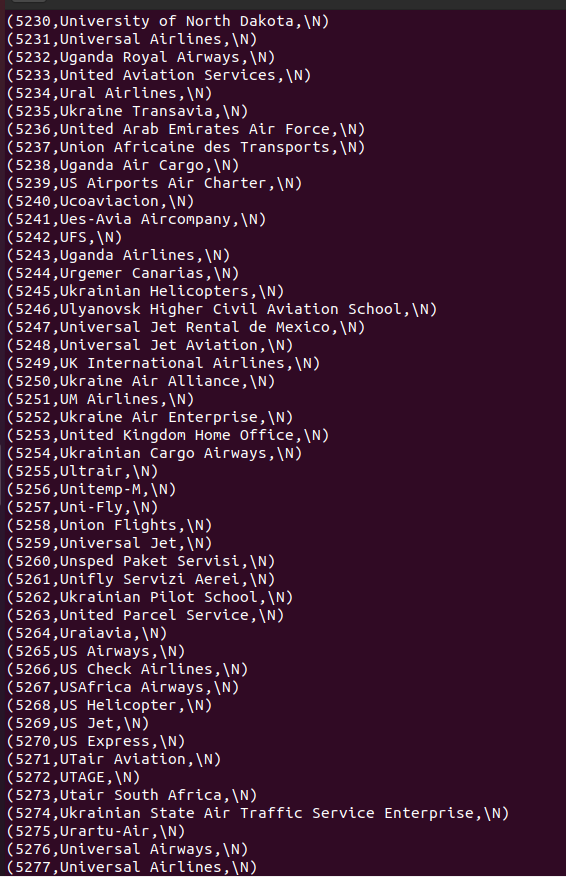
Job is to find all the airlines that are active in the United States which are currently active. So first filtering all the airlines that are active using filter and by keywords and saving them in a table and then querying that table to check for the United States territory using the same keywords and printing the airline names only using foreach and generate keywords.

1. **Get all the alias names of all airlines with airlineID**

**Query:**

* foreach\_data = FOREACH finalair GENERATE airlineID,airline\_name,airline\_alias;
* dump foreach\_data;

**Output:**

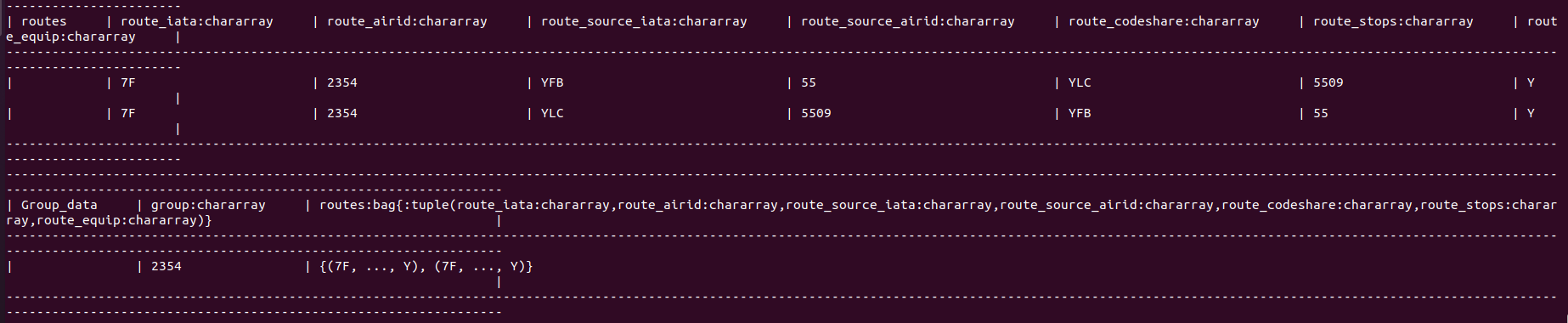
****

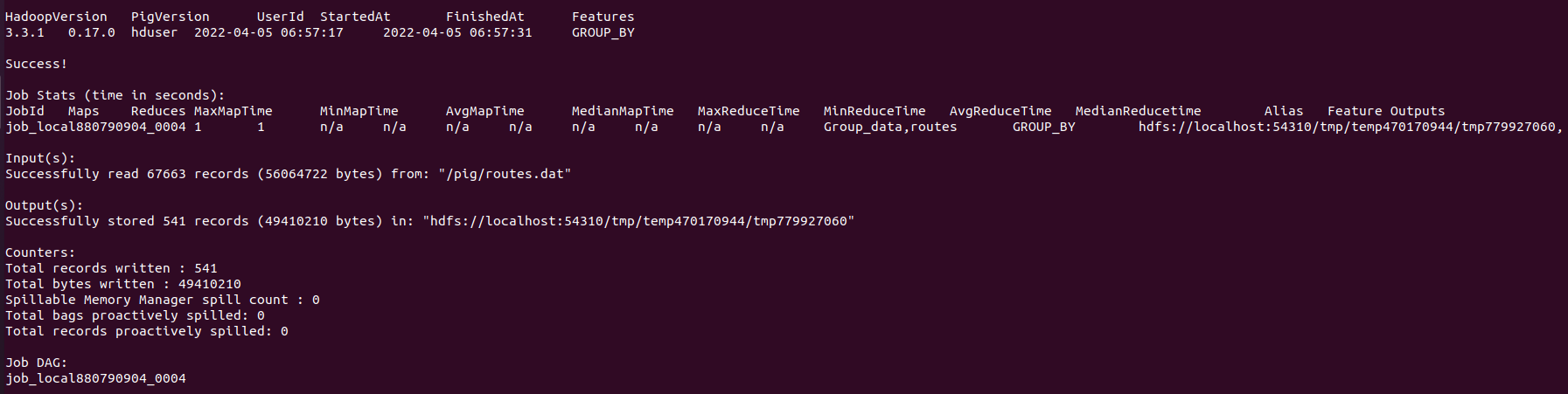
1. **Exploring group command:**

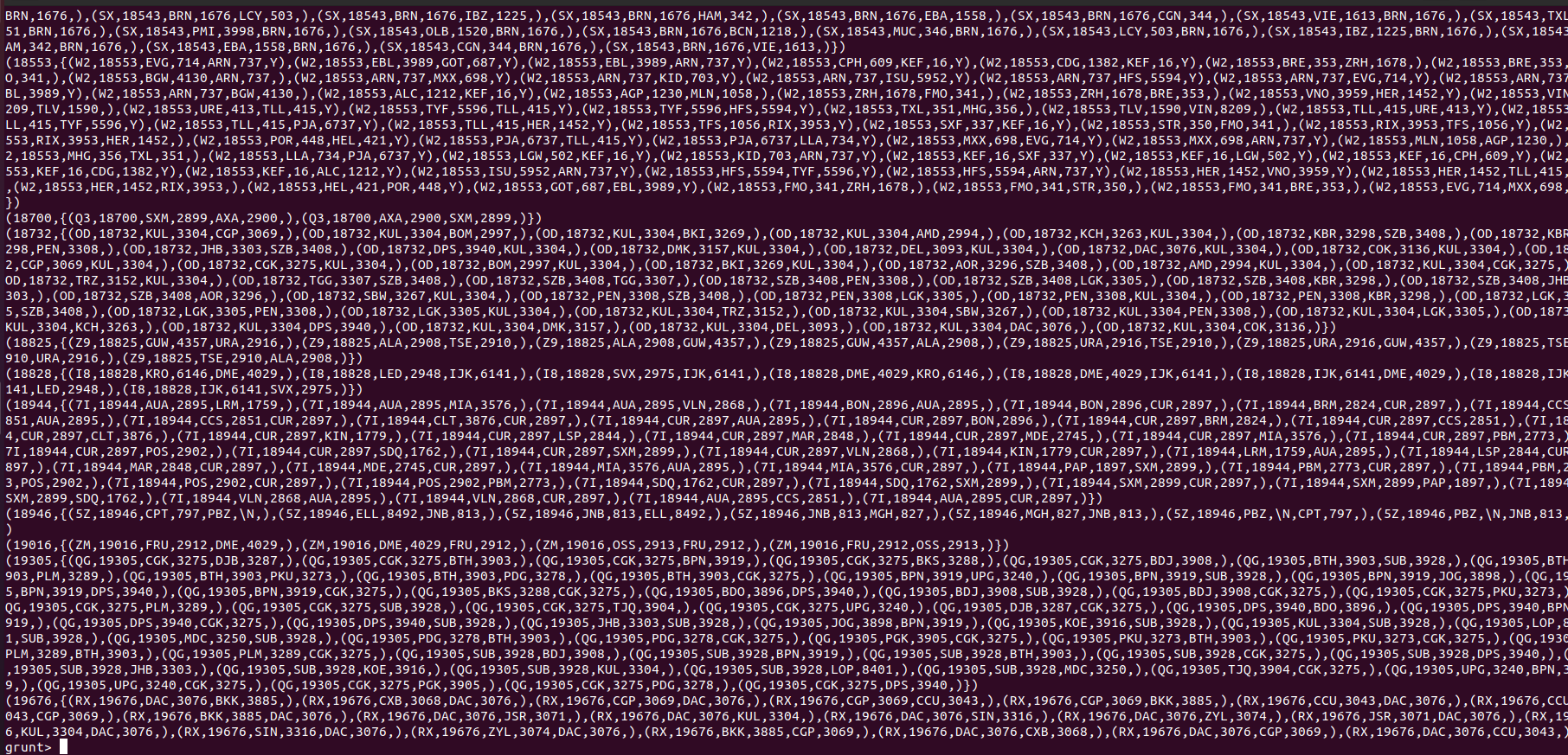
**Query:**

* Group\_data = GROUP routes BY (route\_iata,route\_airid);
* Limit\_group = LIMIT Group\_data 10;
* Dump Limit\_group;
* group\_all = GROUP routes All;

**Output:**

****

****

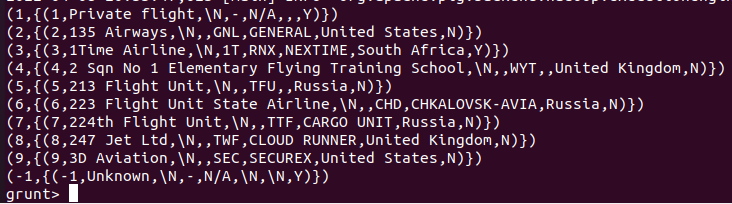
****

1. **Group airline id by active status;**

**Query:**

* ae = group finalair by airlineID;
* ans = limit ae 10;
* illustrate as;
* dump as;

**Output:**

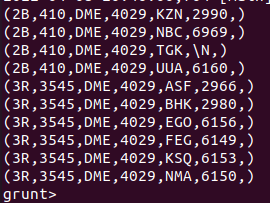
****

1. **Split relation based on src airport id i.e, equal to DME to get exclusive info.**

**Query:**

* split routes into ru\_dme if route\_source\_iata == 'DME' , ru\_not\_dme if route\_source\_iata != 'DME';
* dump ru\_dme;
* dump ru\_not\_dme;

**Output:**

****

**Explaination:**

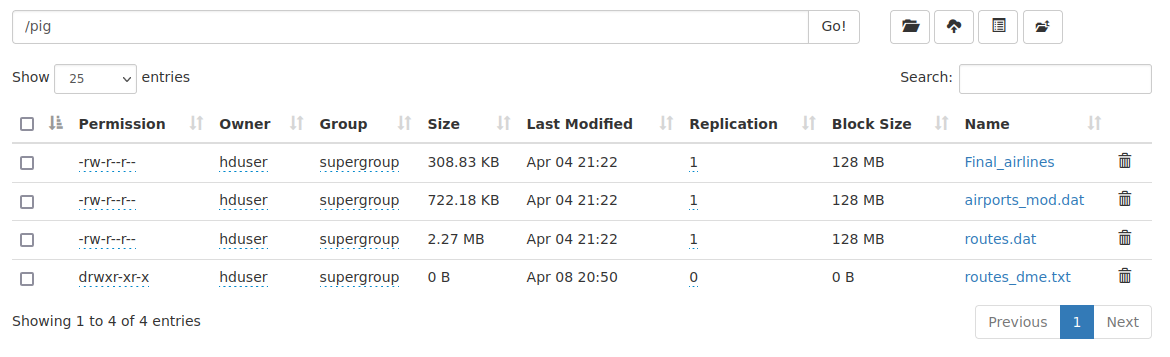
* We are trying to split our relation routes in source airport as DME and not equal to DME using split operator.

1. **Store ru\_dme relation into text file separated by delimiter “,”.**

**Query:**

* store ru\_dme into '/pig/routes\_dme.txt' using PigStorage(',');

**Output:**

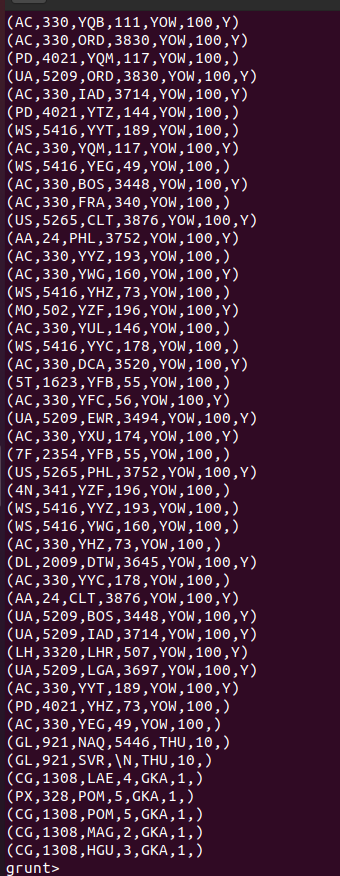
****

1. **Find the flights that has most number of stops?**

**Query:**

* stops = ORDER routes by route\_stops DESC;
* dump stops;

**Output:**

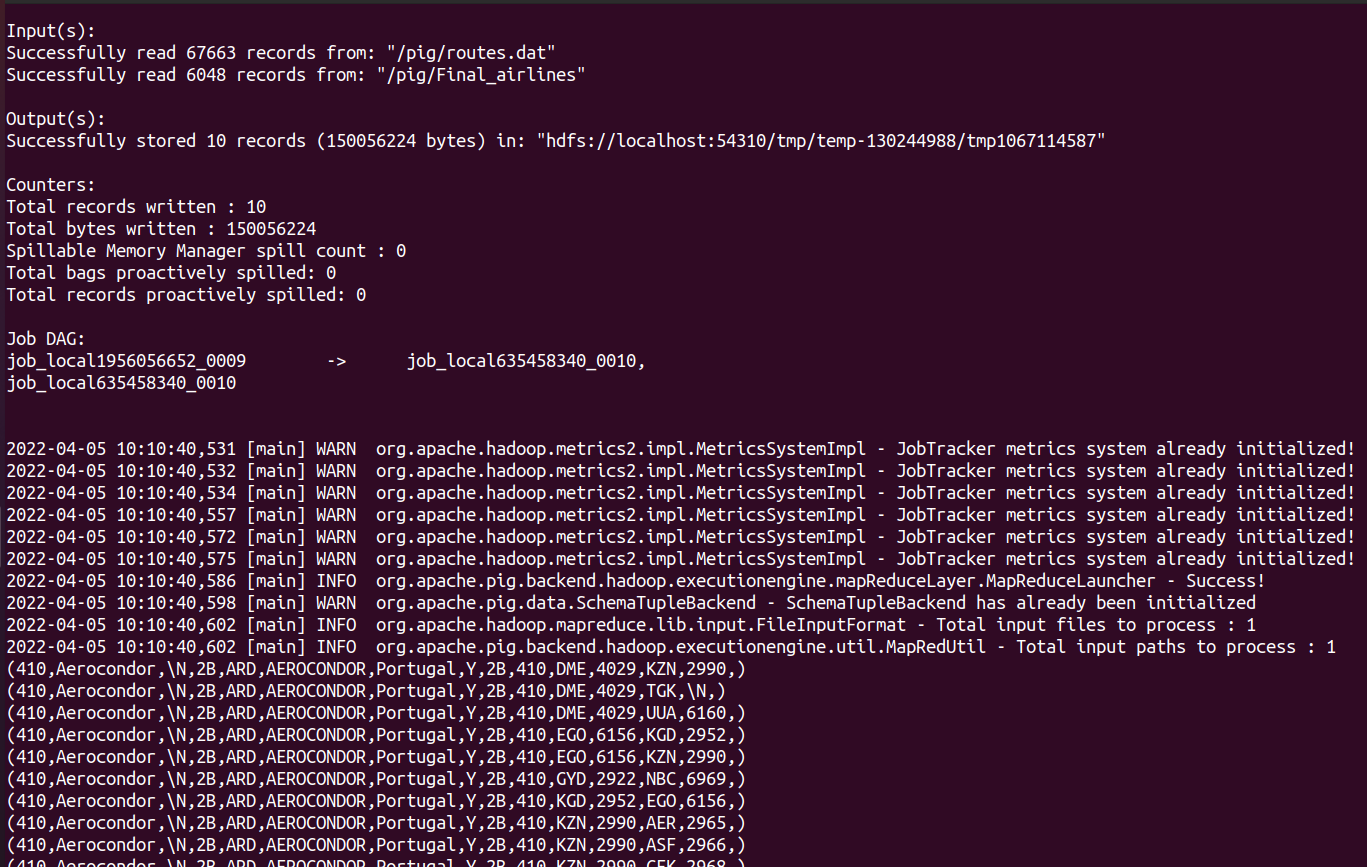
****

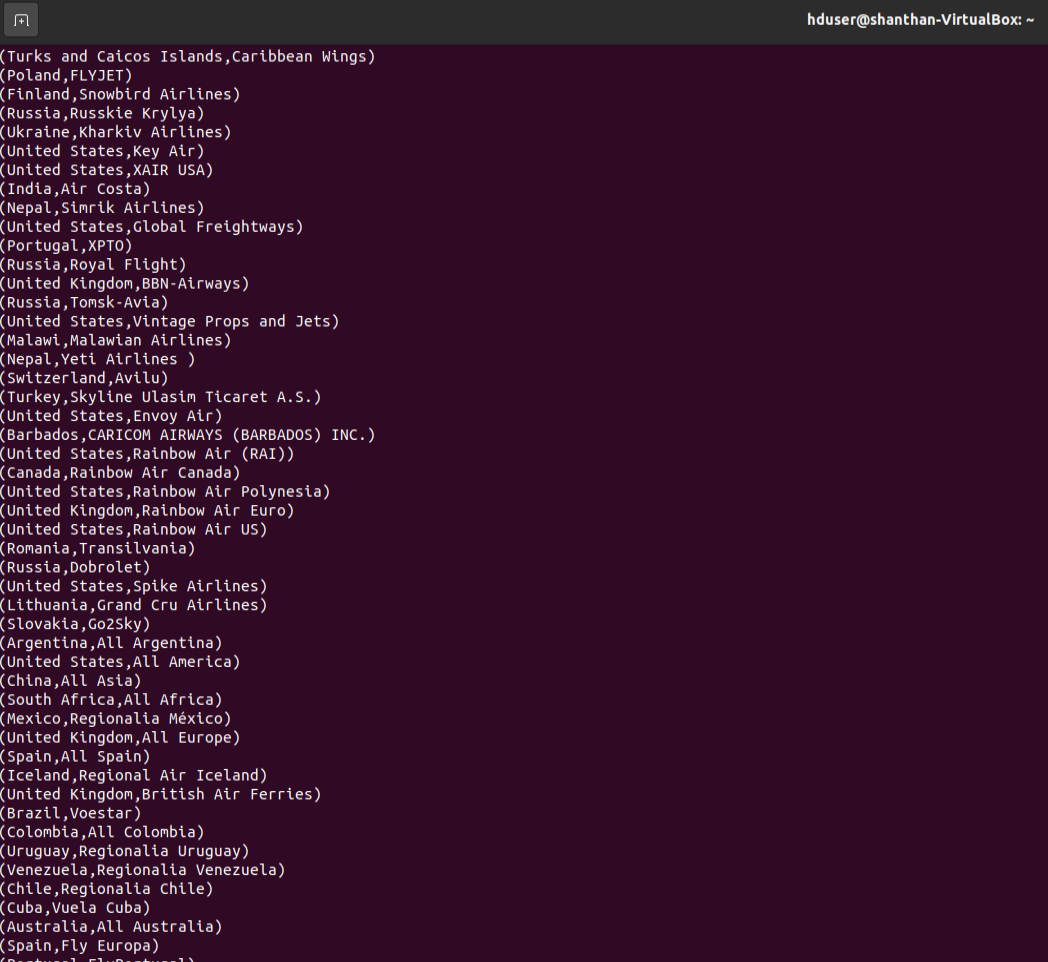
1. **JOIN Command:**

**Query:**

* right\_join = JOIN finalair BY airline\_iata RIGHT OUTER, routes BY route\_iata;
* LIM = LIMIT right\_join 10;
* dump LIM;

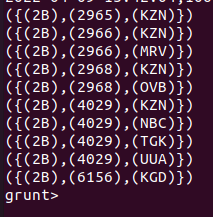
**Output:**

****



1. **Bag source,dest airport and airline.**

* baged\_airline = foreach routes generate TOBAG(route\_iata,route\_source\_airid,route\_codeshare);
* bg = limit baged\_airline 10;
* dump bg;

****