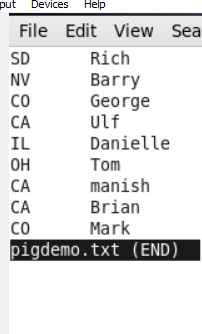


Viewing contents of file

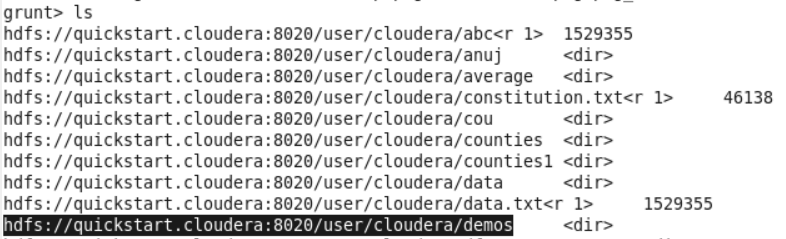
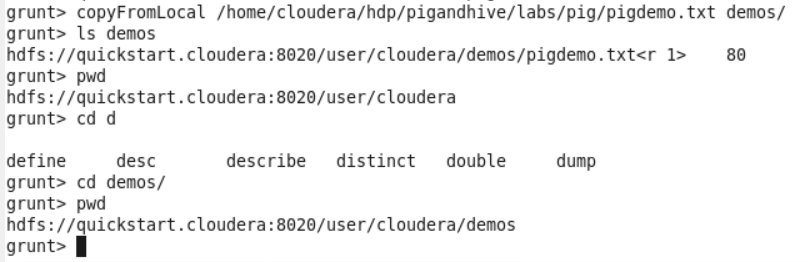
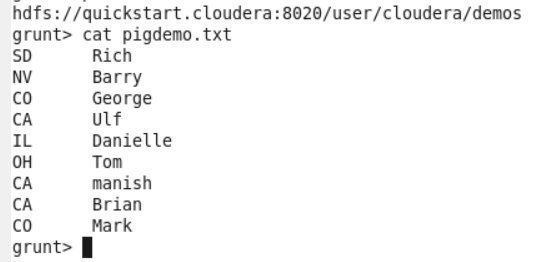


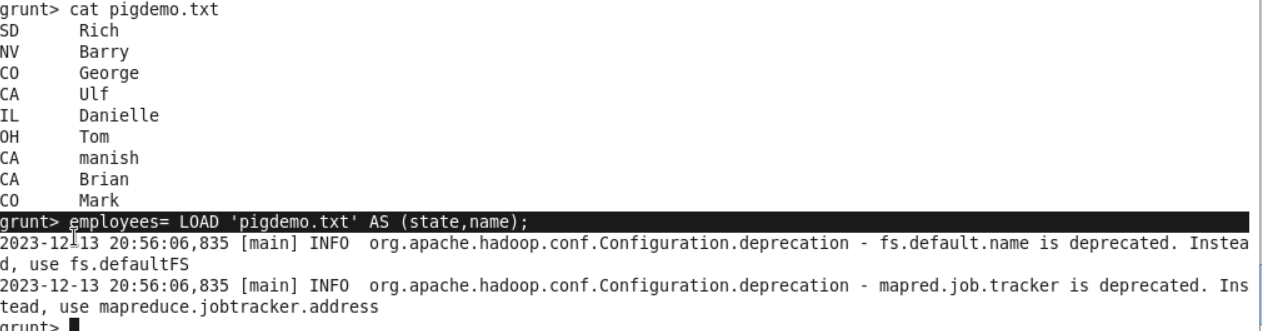
Pig

Grunt>

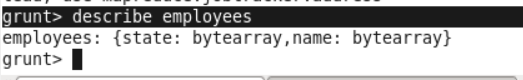
Ls

Make directory

* Mkdir demos
* 
* …
* 
* …
* 
* …

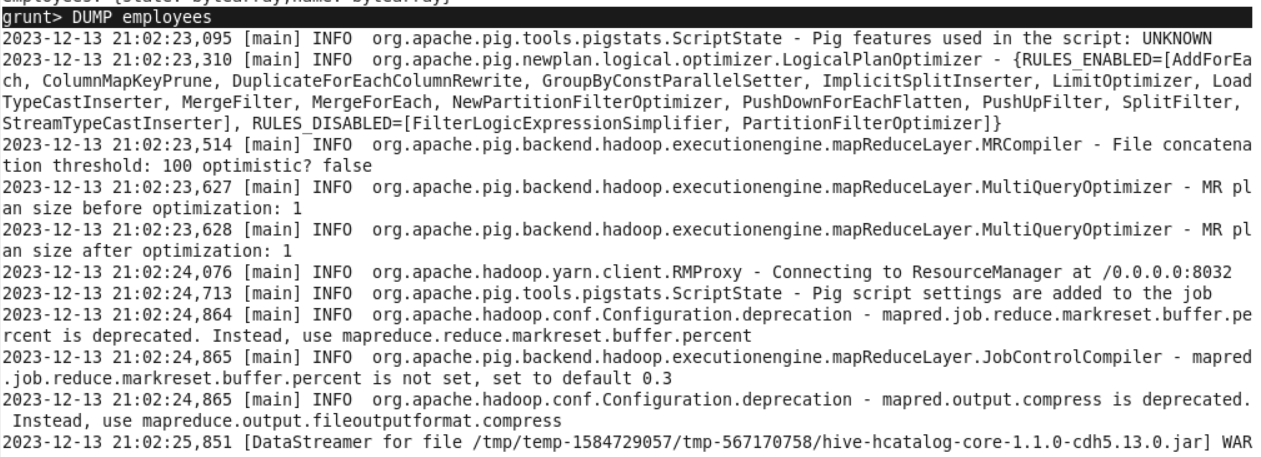


Just added not running

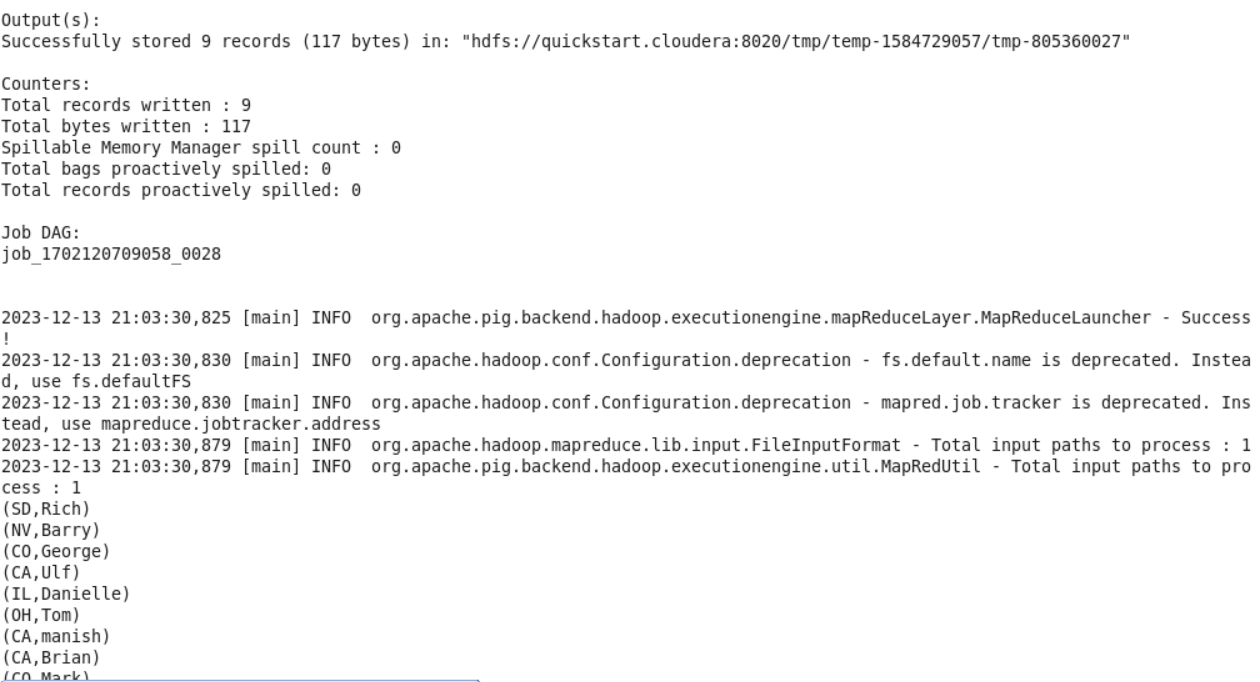


employees: {state: bytearray,name: bytearray}

if not specified, Default datatype is bytearray



Will dump data on console



When we give DUMP or STORE command, parser will understand to run the statement, then it will see data in logical plan and will then execute the statement and then DUMP data on console.



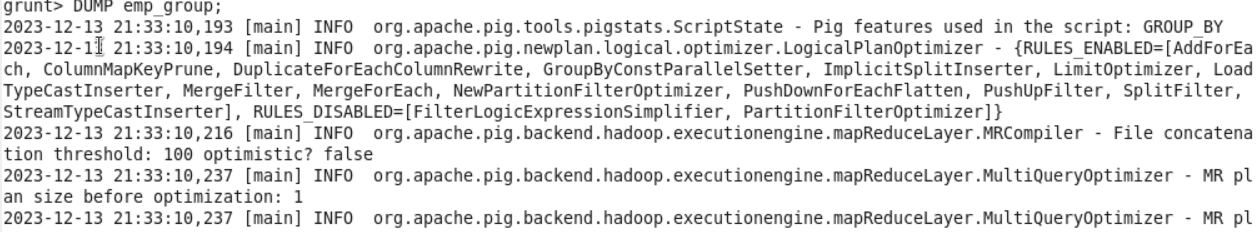


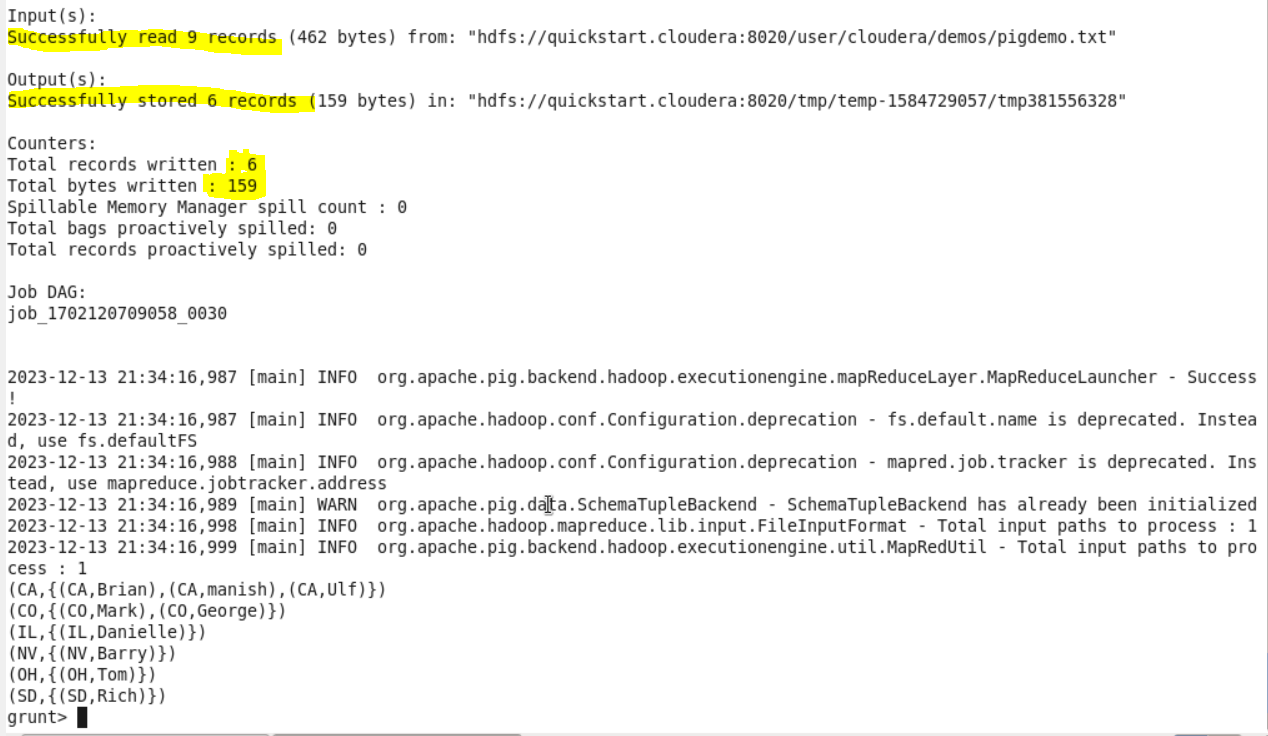
Lazy evaluation



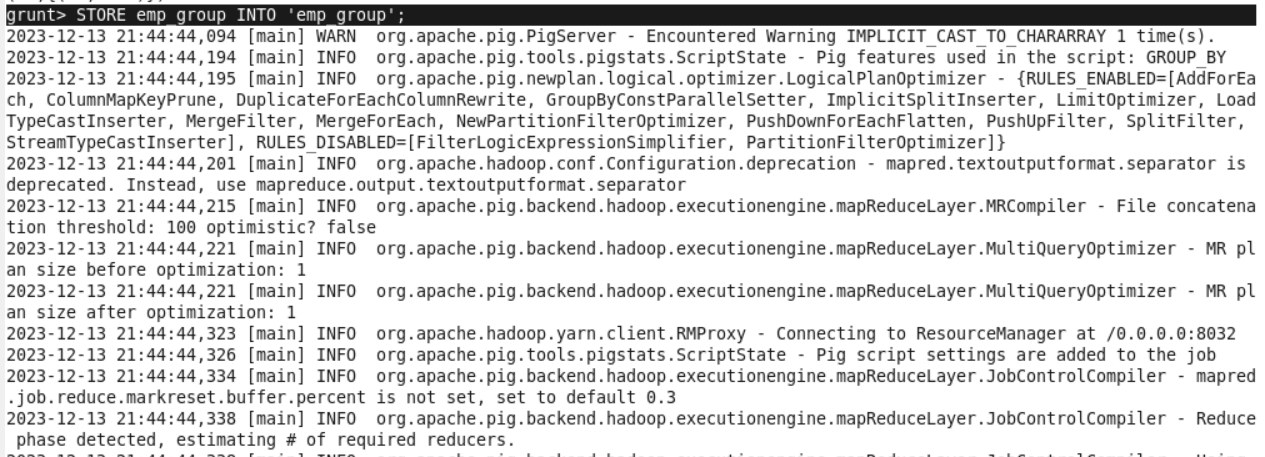


…

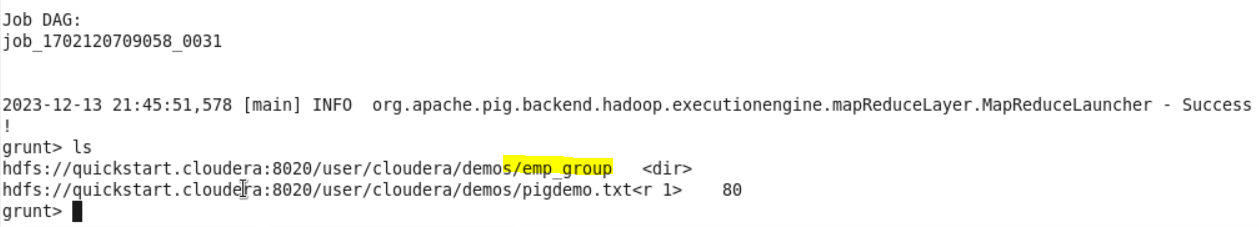


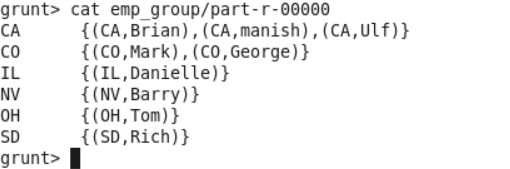


…

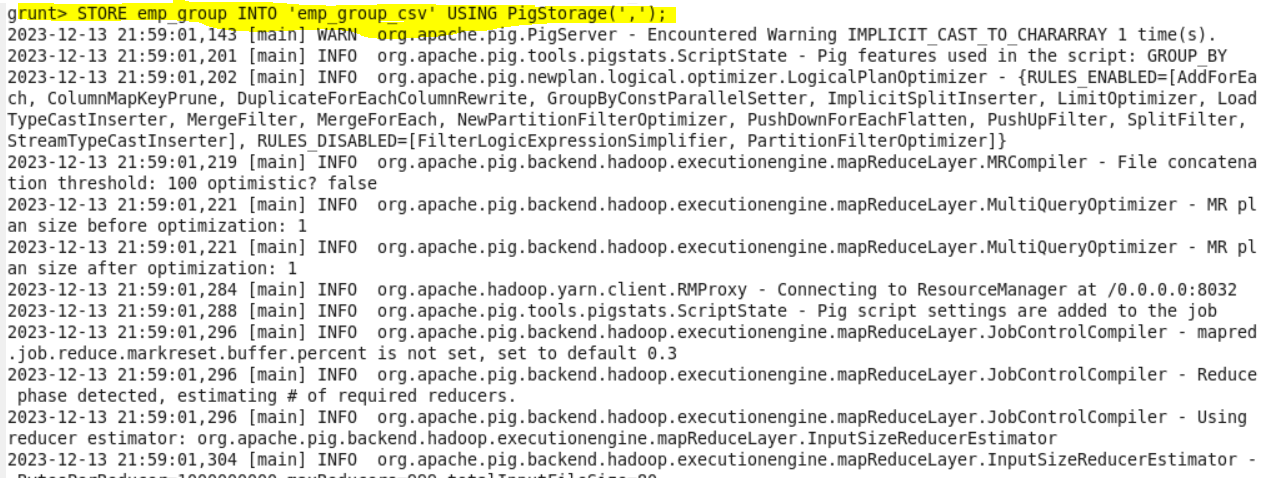


…

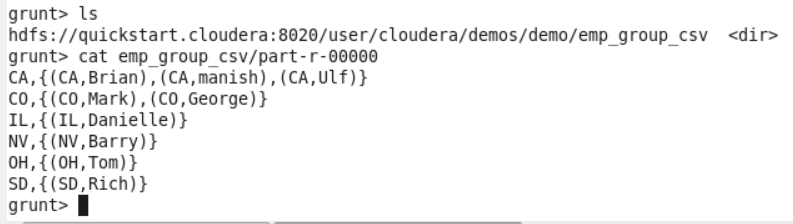




…



..



 Getting Started with Pig

1) Download the RAR, extract into **STAGING\_AREA** location

2) Pull the extracted folder in **LABS\_AREA**/Lab5.1

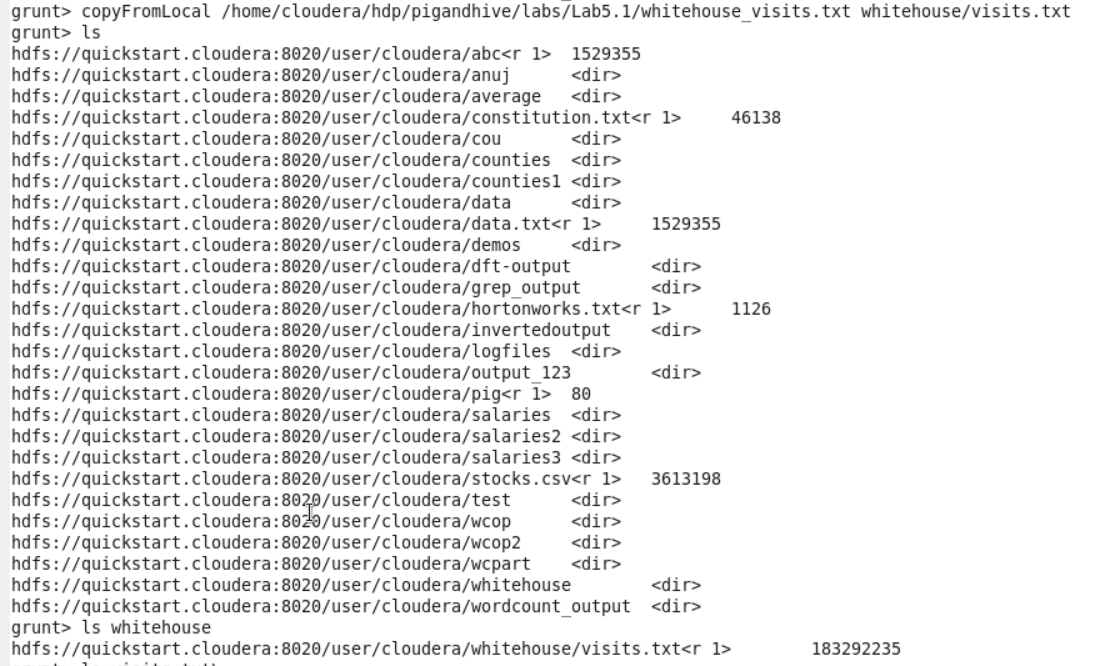
3) Follow the instructions from the lab manual

* Code/Dataset

data/Lab5.1.rar

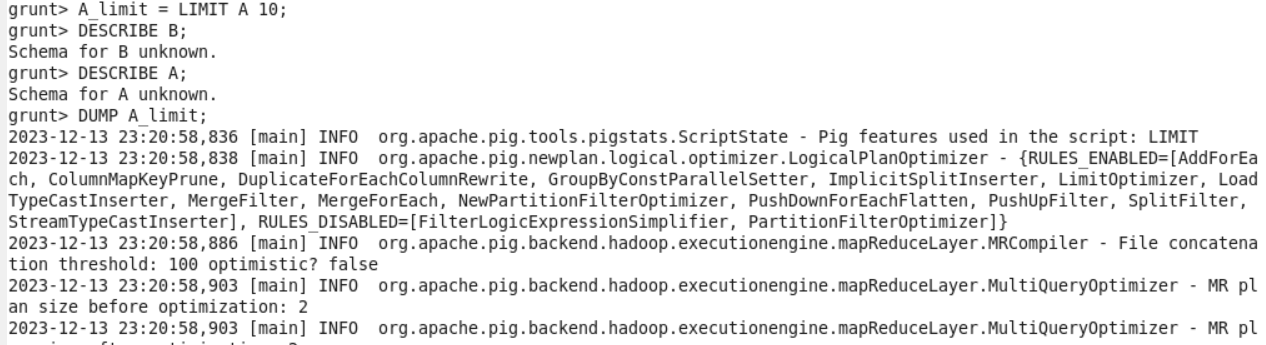


…

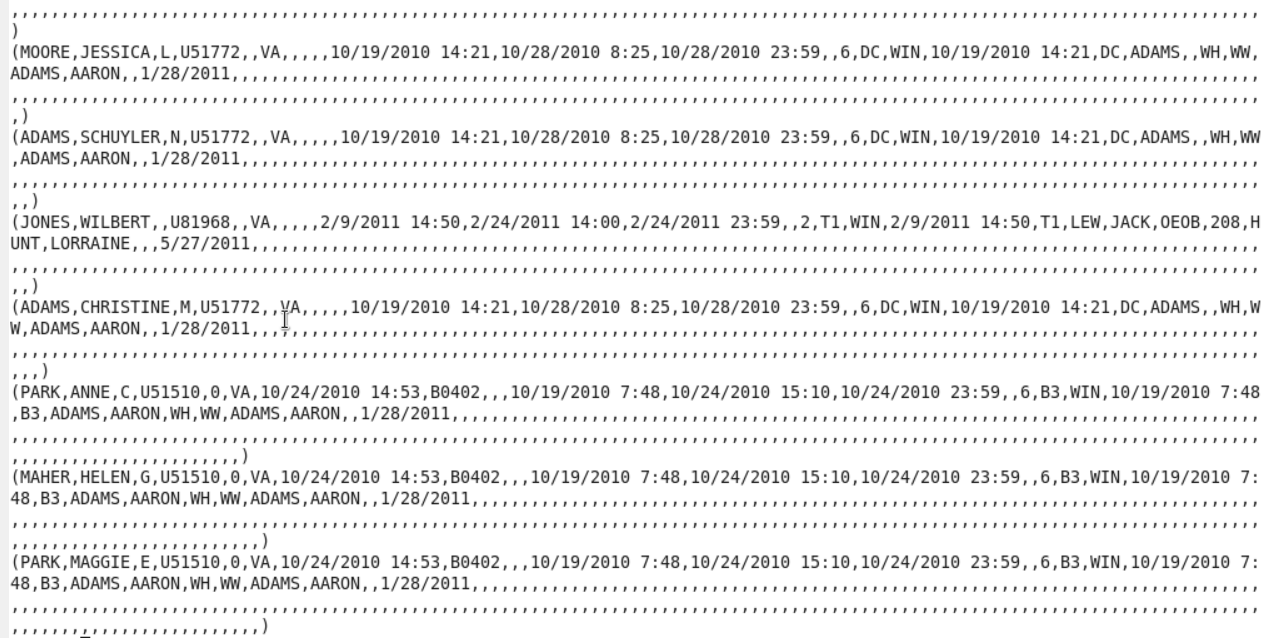


…

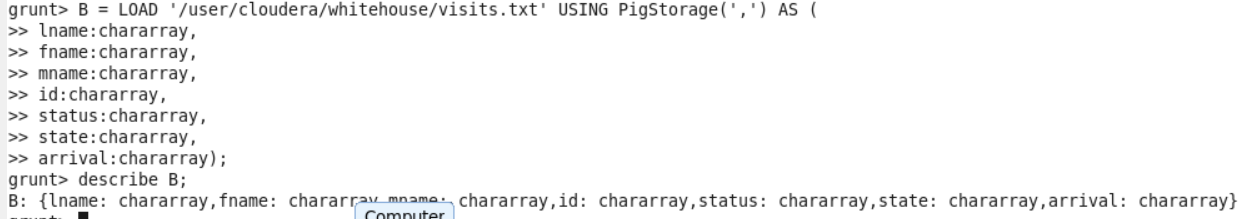




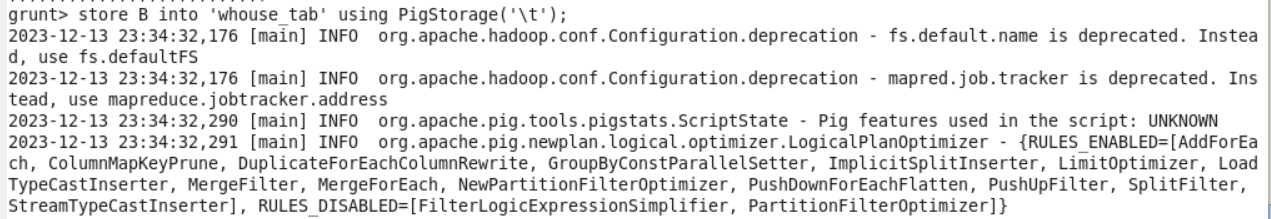
…



…

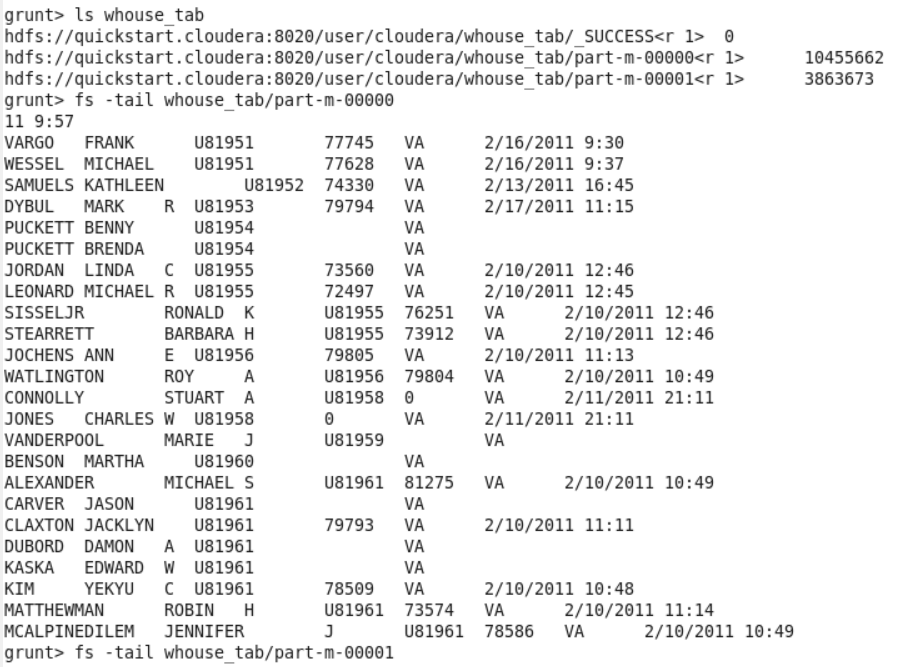


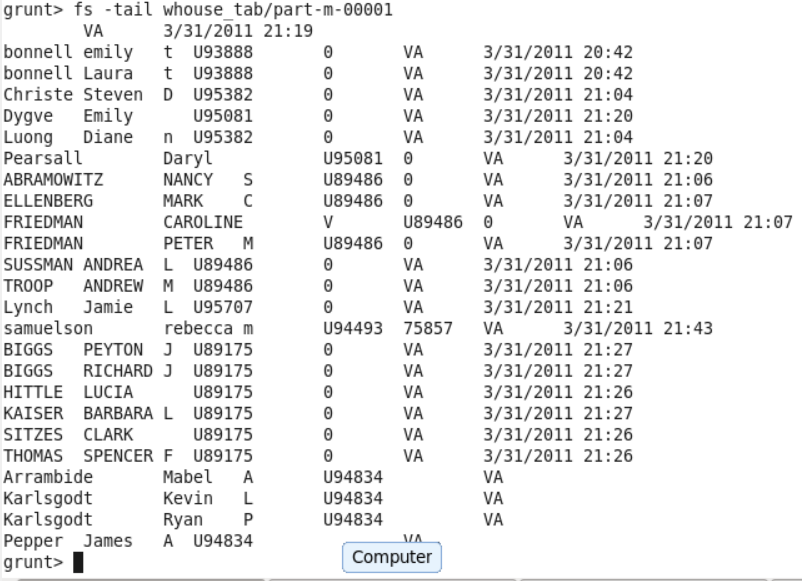
…





…

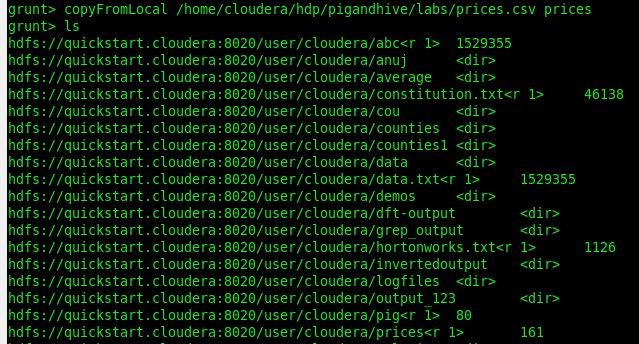




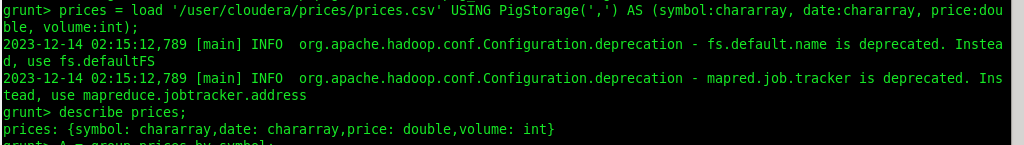
…



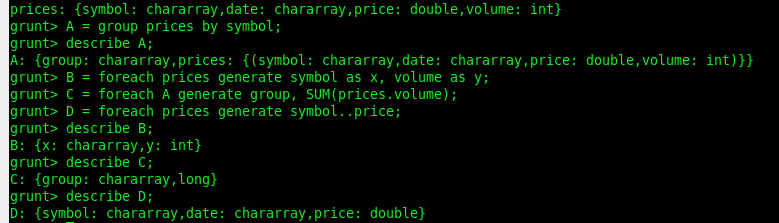




…



…



…

**Lab: Exploring Data with Pig**

**About this Lab**

**Objective:**

Use Pig to navigate through HDFS and explore a datase

**File locations:**

whitehouse/visits.txt in HDFS

**Successful outcome:**

You will have written several Pig scripts that analyze and query the White House visitors’ data, including a list of people who visited the President.

**Before you begin:**

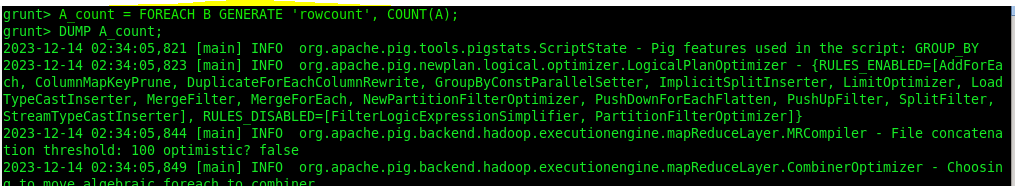
At a minimum, complete steps 1 and 2 of the Getting Started with Pig lab.



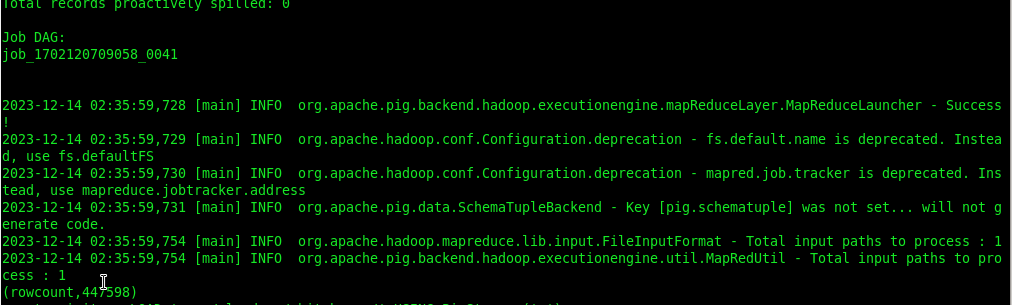
…



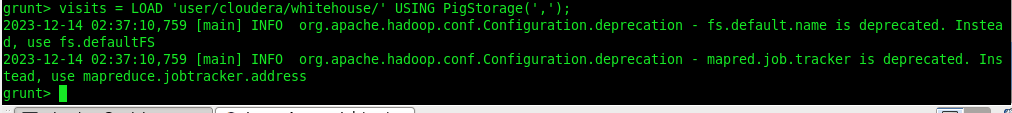
…



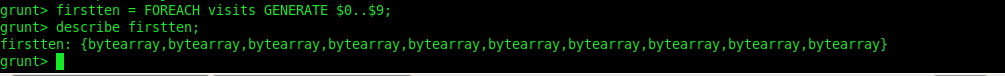
…

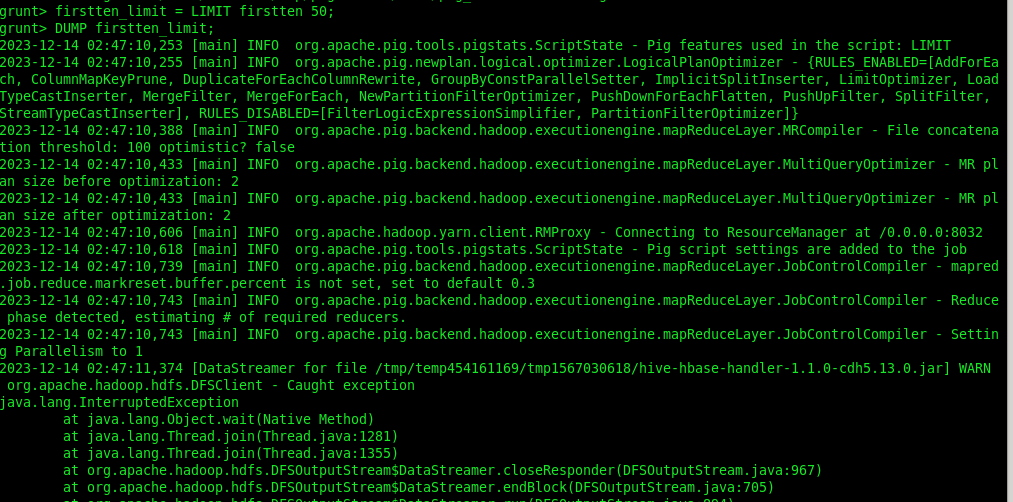


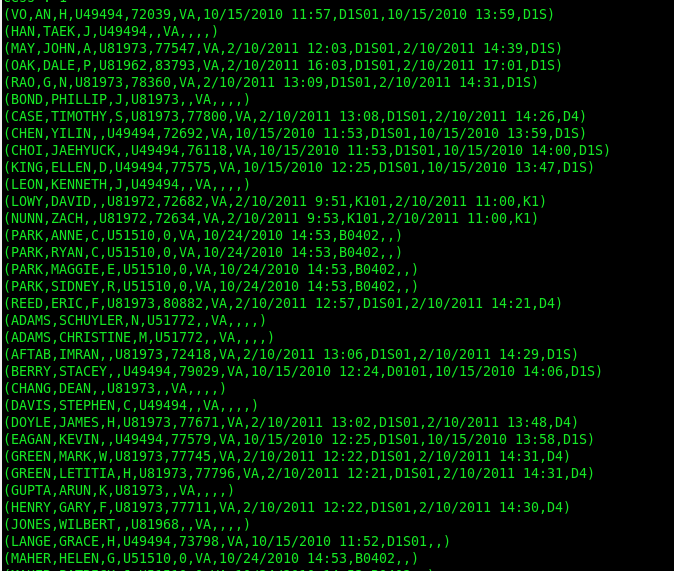
…



…

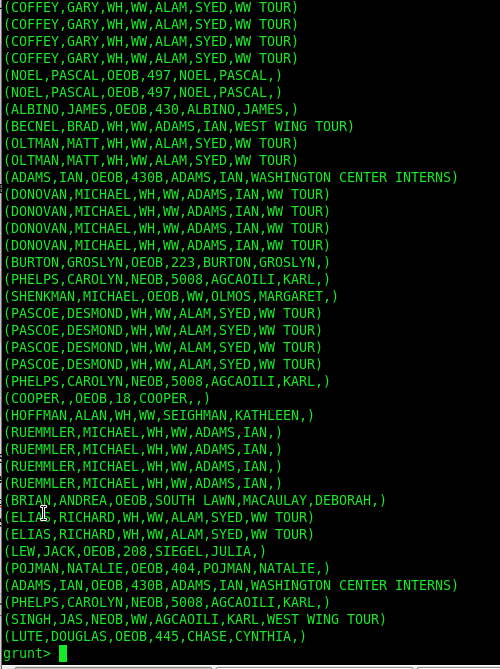




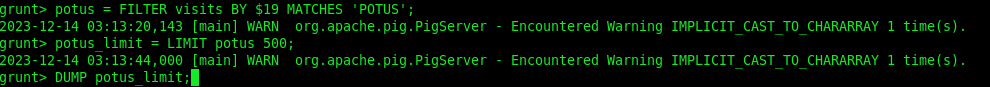




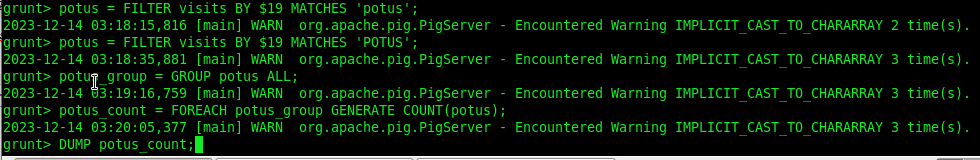
…



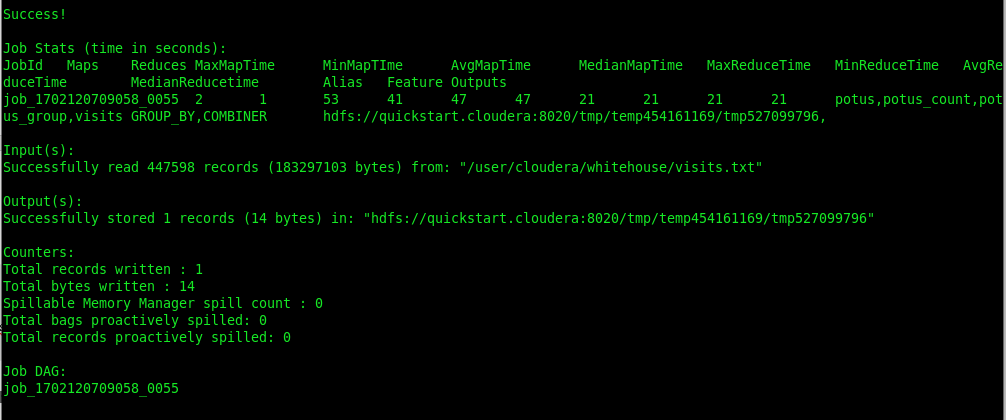
…



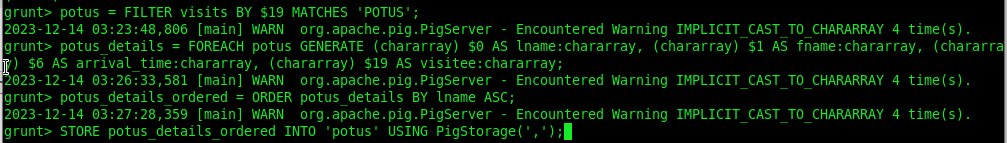
…



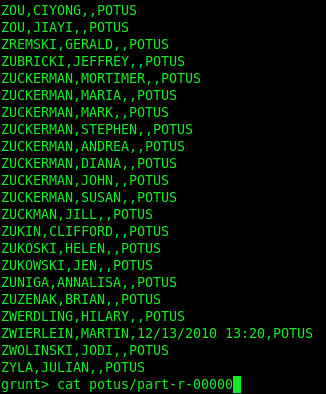
…



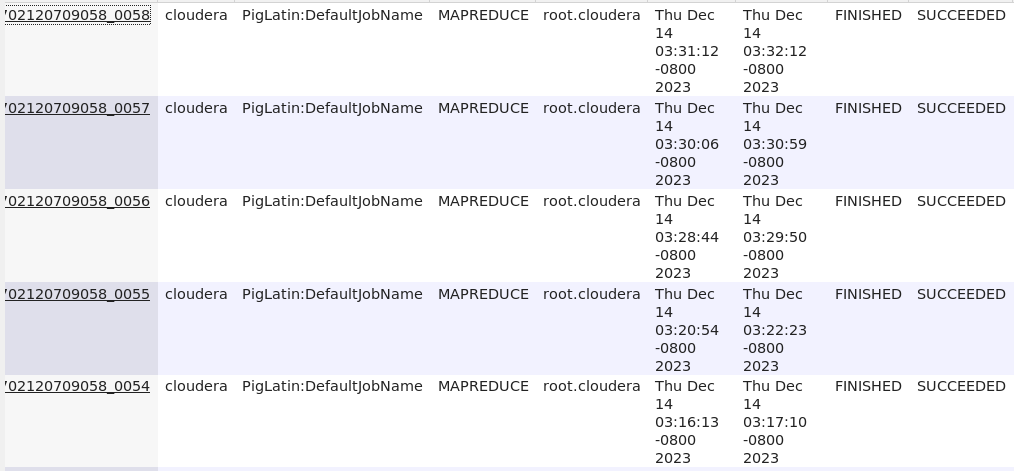
…



…



…



4] Preparing data for Hive

1) Download the attachment in **STAGING\_AREA**

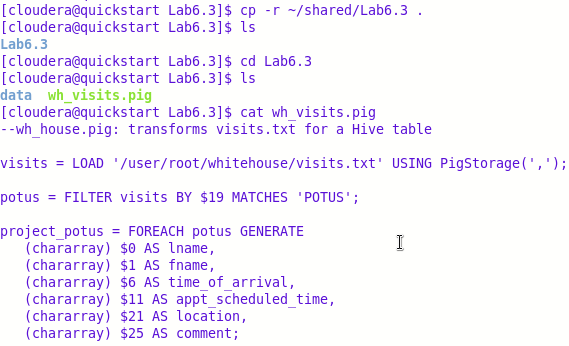
2) Extract it

3) Pull the extracted contents in **LABS\_AREA**/lab6.3

4) Refer the lab from the manual and wait for the instructions by an Instructor

* Code/Dataset

data/Lab6.3.rar



Pig