```
start-dfs.sh
start-yarn.sh
mapred --daemon start historyserver
```

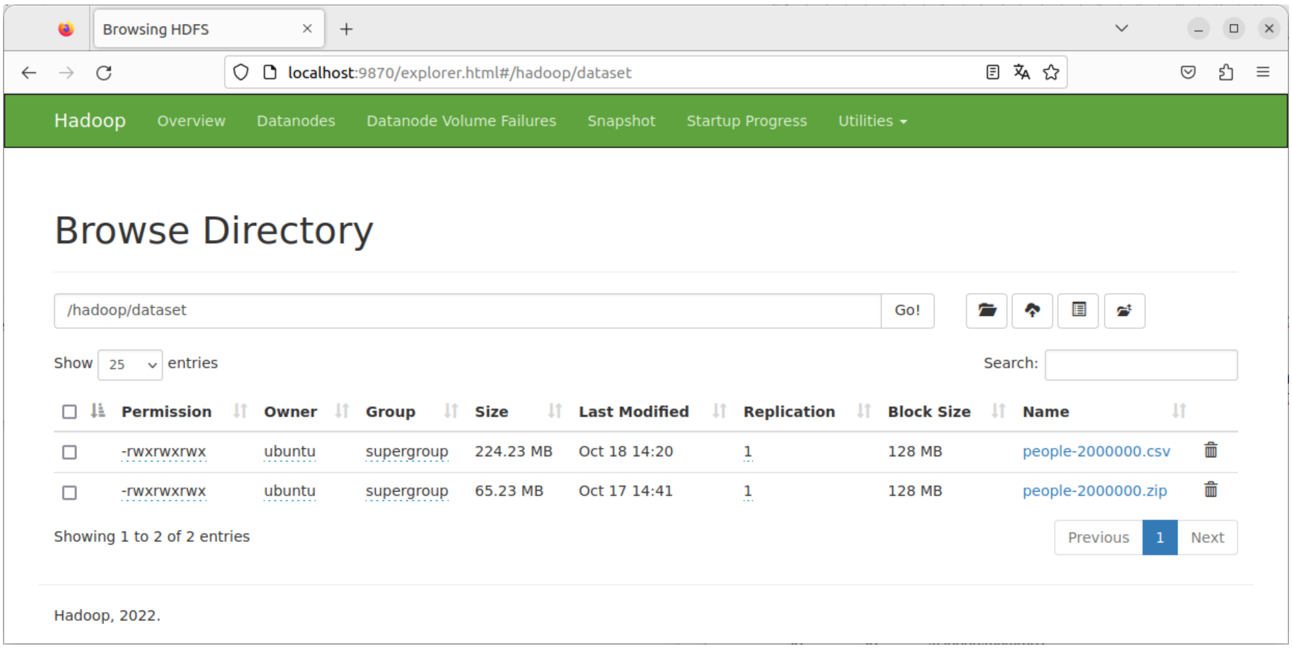
2.1 Carga de Dataset

1) # Crea una carpeta en HDFS en la ruta /hadoop/dataset.

```
$ hdfs dfs -mkdir -p /hadoop/dataset
```

2) # Descarga el fichero people-2000000.zip desde la plataforma virtual y cópialo en la ruta que acabas de crear.

```
$ hdfs dfs -copyFromLocal /home/ubuntu/Descargas/people-2000000.zip /hadoop/dataset
$ hdfs dfs -copyFromLocal /home/ubuntu/Descargas/people-2000000.csv /hadoop/dataset
```



Browse Directory

/hadoop/dataset

☐	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
☐	-rwxrwxrwx	ubuntu	supergroup	224.23 MB	Oct 18 14:20	1	128 MB	people-2000000.csv
☐	-rwxrwxrwx	ubuntu	supergroup	65.23 MB	Oct 17 14:41	1	128 MB	people-2000000.zip

Showing 1 to 2 of 2 entries

Hadoop, 2022.

a) # Lista el contenido de la carpeta, incluyendo el tamaño del achivo.

```
$ hdfs dfs -du -h /hadoop/dataset
```

```
224.2 M  224.2 M  /hadoop/dataset/people-2000000.csv
65.2 M   65.2 M   /hadoop/dataset/people-2000000.zip
```

```
$ hdfs dfs -du -h /hadoop
```

```
0         0         /hadoop/dataout
289.5 M   289.5 M   /hadoop/dataset
50        50        /hadoop/ejemplo1
```

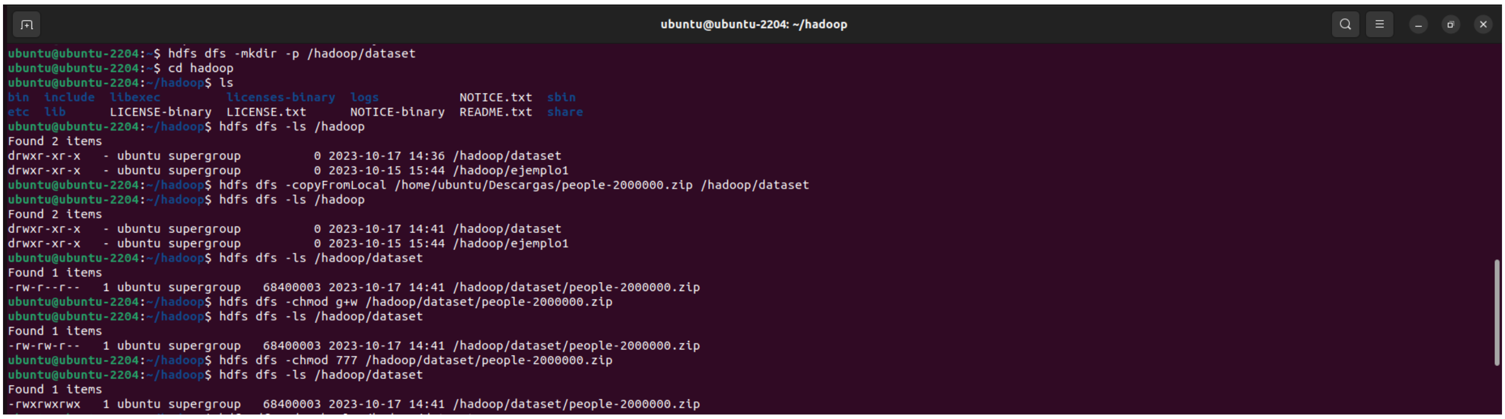
```
$ hdfs dfs -ls /hadoop/dataset

Found 2 items
-rwxrwxrwx   1 ubuntu supergroup  235121126 2023-10-18 14:20 /hadoop/dataset/people-2000000.csv
-rwxrwxrwx   1 ubuntu supergroup   68400003 2023-10-17 14:41 /hadoop/dataset/people-2000000.zip
```

```
$ hdfs dfs -ls /hadoop

Found 3 items
drwxr-xr-x   - ubuntu supergroup          0 2023-10-18 14:39 /hadoop/dataout
drwxr-xr-x   - ubuntu supergroup          0 2023-10-18 14:20 /hadoop/dataset
drwxr-xr-x   - ubuntu supergroup          0 2023-10-15 15:44 /hadoop/ejemplo1
```

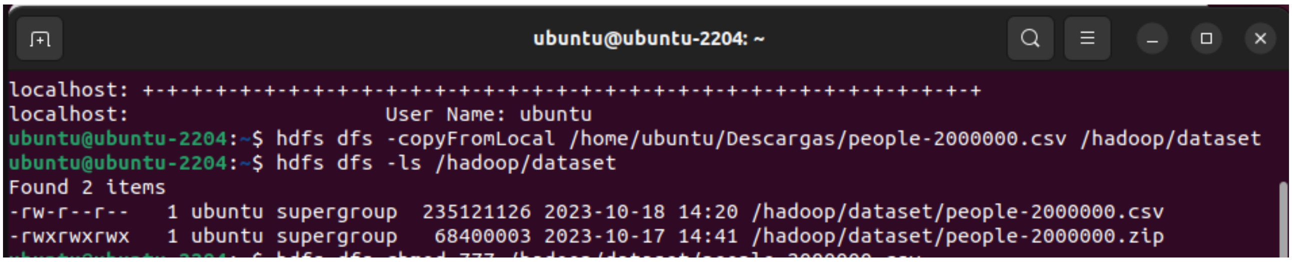
b) #Muestra las primeras líneas del fichero, leyendo directamente de HDFS (sin copiar al sistema de ficheros local) en línea de comandos.

```
$ hadoop fs -cat /hadoop/dataset/people-2000000.csv | head

Index,User Id,First Name,Last Name,Sex,Email,Phone,Date of birth,Job Title
1,4defE49671cF860,Sydney,Shannon,Male,tvang@example.net,574-440-1423x9799,2020-07-09,Technical brewer
2,F89B87bCf8f210b,Regina,Lin,Male,helen14@example.net,001-273-664-2268x90121,1909-06-20,"Teacher, adult education"
3,Cad6052BDd5DEaf,Pamela,Blake,Female,brent05@example.org,927-880-5785x85266,1964-08-19,Armed forces operational officer
4,e83E46f80f629CD,Dave,Hoffman,Female,munozcraig@example.org,001-147-429-8340x608,2009-02-19,Ship broker
5,60AAc4DcaBcE3b6,Ian,Campos,Female,brownevelyn@example.net,166-126-4390,1997-10-02,Media planner
6,7ACb92d81A42fdf,Valerie,Patel,Male,muellerjoel@example.net,001-379-612-1298x853,2021-04-07,"Engineer, materials"
7,A00bacC18101d37,Dan,Castillo,Female,billmoody@example.net,(448)494-0852x63243,1975-04-09,Historic buildings inspector/conservation officer
8,B012698Cf31cfec,Clinton,Cochran,Male,glenn94@example.org,4425100065,1966-07-19,"Engineer, mining"
9,a5bd11BD7dA1a4B,Gabriella,Richard,Female,blane@example.org,352.362.4148x8344,2021-09-02,Wellsite geologist
cat: Unable to write to output stream.
```

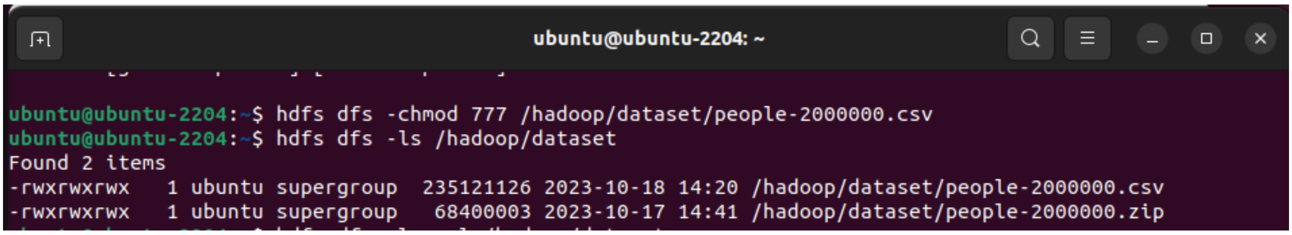
4)
```
$ hdfs dfs -mkdir -p /hadoop/dataout
```



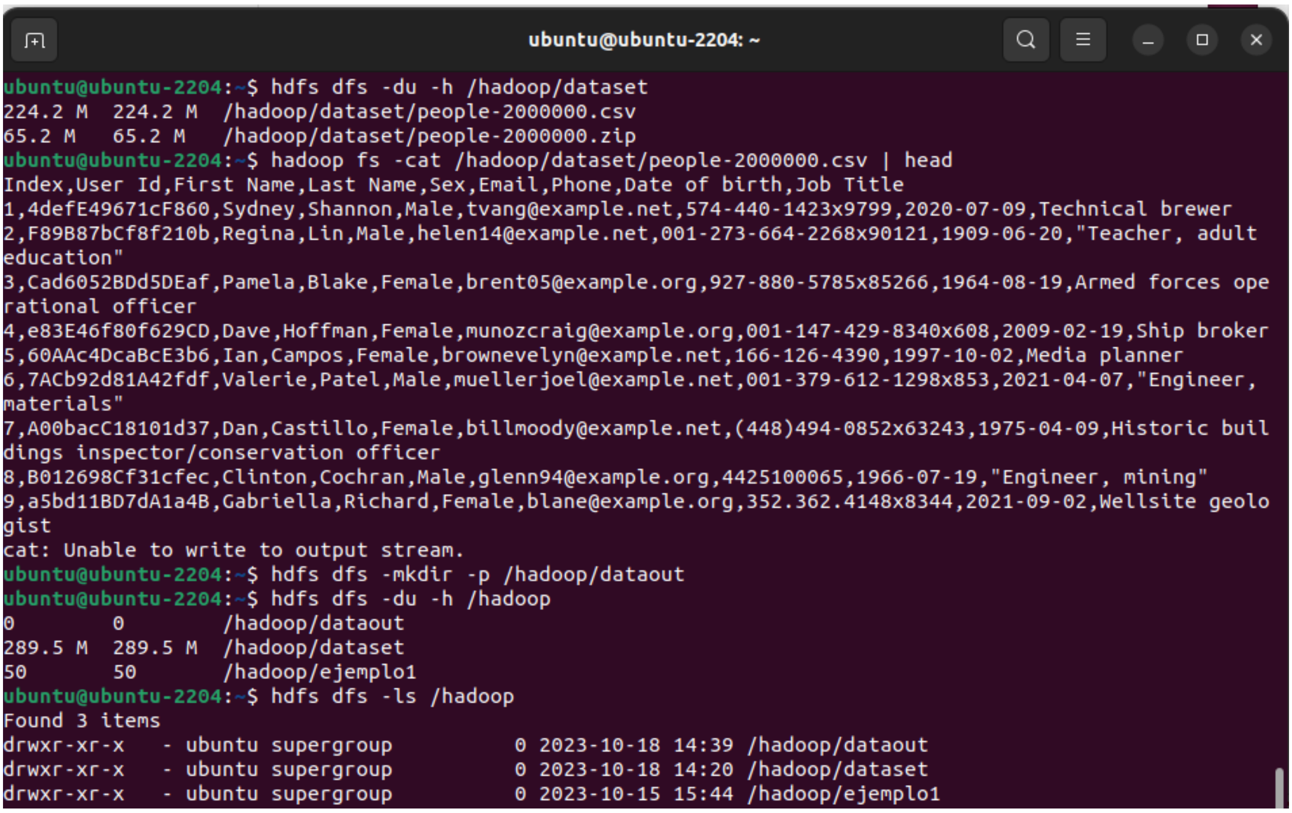
```
ubuntu@ubuntu-2204:~$ hdfs dfs -mkdir -p /hadoop/dataset
ubuntu@ubuntu-2204:~$ cd hadoop
ubuntu@ubuntu-2204:~/hadoop$ ls
bin  include  libexec         licenses-binary  logs           NOTICE.txt  sbin
etc  lib      LICENSE-binary  LICENSE.txt      NOTICE-binary  README.txt  share
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -ls /hadoop
Found 2 items
drwxr-xr-x   - ubuntu supergroup          0 2023-10-17 14:36 /hadoop/dataset
drwxr-xr-x   - ubuntu supergroup          0 2023-10-15 15:44 /hadoop/ejemplo1
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -copyFromLocal /home/ubuntu/Descargas/people-2000000.zip /hadoop/dataset
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -ls /hadoop
Found 2 items
drwxr-xr-x   - ubuntu supergroup          0 2023-10-17 14:41 /hadoop/dataset
drwxr-xr-x   - ubuntu supergroup          0 2023-10-15 15:44 /hadoop/ejemplo1
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -ls /hadoop/dataset
Found 1 items
-rw-r--r--   1 ubuntu supergroup   68400003 2023-10-17 14:41 /hadoop/dataset/people-2000000.zip
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -chmod g+w /hadoop/dataset/people-2000000.zip
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -ls /hadoop/dataset
Found 1 items
-rw-rw-r--   1 ubuntu supergroup   68400003 2023-10-17 14:41 /hadoop/dataset/people-2000000.zip
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -chmod 777 /hadoop/dataset/people-2000000.zip
ubuntu@ubuntu-2204:~/hadoop$ hdfs dfs -ls /hadoop/dataset
Found 1 items
-rwxrwxrwx   1 ubuntu supergroup   68400003 2023-10-17 14:41 /hadoop/dataset/people-2000000.zip
```



```
localhost: +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
localhost:                          User Name: ubuntu
ubuntu@ubuntu-2204:~$ hdfs dfs -copyFromLocal /home/ubuntu/Descargas/people-2000000.csv /hadoop/dataset
ubuntu@ubuntu-2204:~$ hdfs dfs -ls /hadoop/dataset
Found 2 items
-rw-r--r--   1 ubuntu supergroup  235121126 2023-10-18 14:20 /hadoop/dataset/people-2000000.csv
-rwxrwxrwx   1 ubuntu supergroup   68400003 2023-10-17 14:41 /hadoop/dataset/people-2000000.zip
```



```
ubuntu@ubuntu-2204:~$ hdfs dfs -chmod 777 /hadoop/dataset/people-2000000.csv
ubuntu@ubuntu-2204:~$ hdfs dfs -ls /hadoop/dataset
Found 2 items
-rwxrwxrwx   1 ubuntu supergroup  235121126 2023-10-18 14:20 /hadoop/dataset/people-2000000.csv
-rwxrwxrwx   1 ubuntu supergroup   68400003 2023-10-17 14:41 /hadoop/dataset/people-2000000.zip
```

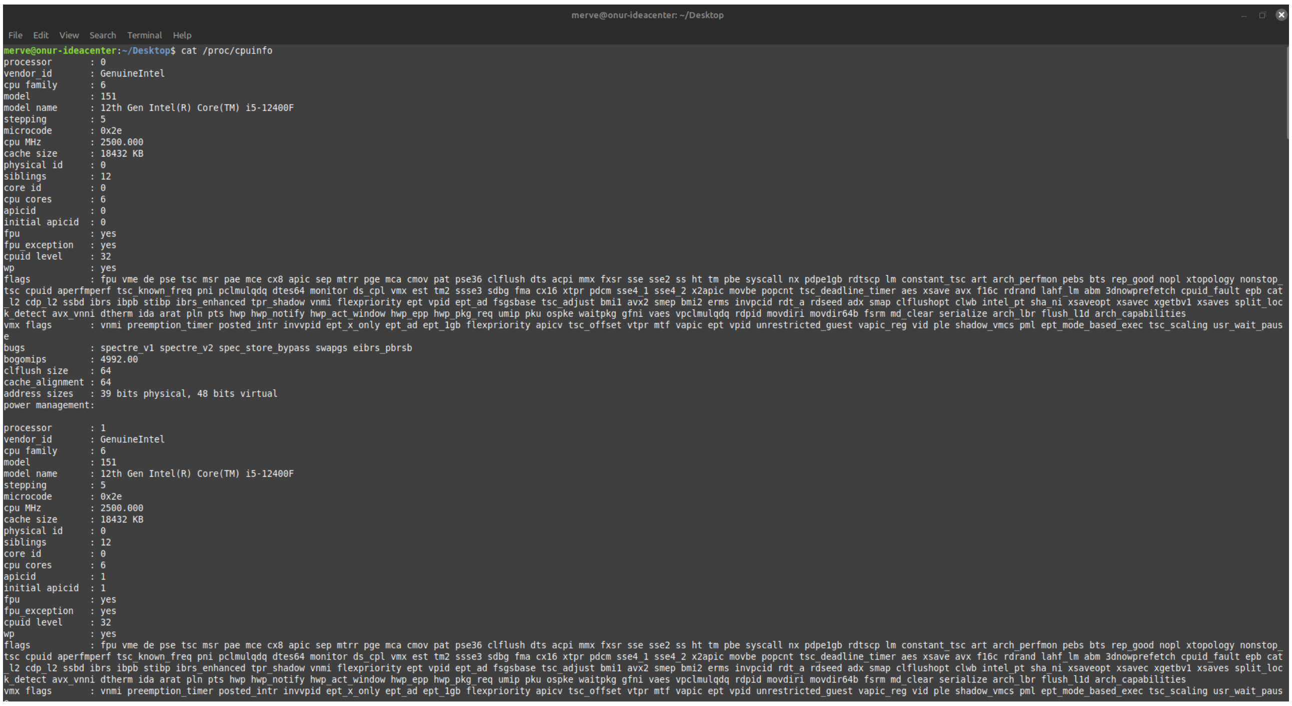


```
ubuntu@ubuntu-2204:~$ hdfs dfs -du -h /hadoop/dataset
224.2 M  224.2 M  /hadoop/dataset/people-2000000.csv
65.2 M   65.2 M   /hadoop/dataset/people-2000000.zip
ubuntu@ubuntu-2204:~$ hadoop fs -cat /hadoop/dataset/people-2000000.csv | head
Index,User Id,First Name,Last Name,Sex,Email,Phone,Date of birth,Job Title
1,4defE49671cF860,Sydney,Shannon,Male,tvang@example.net,574-440-1423x9799,2020-07-09,Technical brewer
2,F89B87bCf8f210b,Regina,Lin,Male,helen14@example.net,001-273-664-2268x90121,1909-06-20,"Teacher, adult
education"
3,Cad6052BDd5DEaf,Pamela,Blake,Female,brent05@example.org,927-880-5785x85266,1964-08-19,Armed forces ope
rational officer
4,e83E46f80f629CD,Dave,Hoffman,Female,munozcraig@example.org,001-147-429-8340x608,2009-02-19,Ship broker
5,60AAc4DcaBcE3b6,Ian,Campos,Female,brownevelyn@example.net,166-126-4390,1997-10-02,Media planner
6,7ACb92d81A42fdf,Valerie,Patel,Male,muellerjoel@example.net,001-379-612-1298x853,2021-04-07,"Engineer,
materials"
7,A00bacC18101d37,Dan,Castillo,Female,billmoody@example.net,(448)494-0852x63243,1975-04-09,Historic buil
dings inspector/conservation officer
8,B012698Cf31cfec,Clinton,Cochran,Male,glenn94@example.org,4425100065,1966-07-19,"Engineer, mining"
9,a5bd11BD7dA1a4B,Gabriella,Richard,Female,blane@example.org,352.362.4148x8344,2021-09-02,Wellsite geolo
gist
cat: Unable to write to output stream.
ubuntu@ubuntu-2204:~$ hdfs dfs -mkdir -p /hadoop/dataout
ubuntu@ubuntu-2204:~$ hdfs dfs -du -h /hadoop
0        0        /hadoop/dataout
289.5 M  289.5 M  /hadoop/dataset
50       50       /hadoop/ejemplo1
ubuntu@ubuntu-2204:~$ hdfs dfs -ls /hadoop
Found 3 items
drwxr-xr-x   - ubuntu supergroup          0 2023-10-18 14:39 /hadoop/dataout
drwxr-xr-x   - ubuntu supergroup          0 2023-10-18 14:20 /hadoop/dataset
drwxr-xr-x   - ubuntu supergroup          0 2023-10-15 15:44 /hadoop/ejemplo1
```

2.2 Consultas

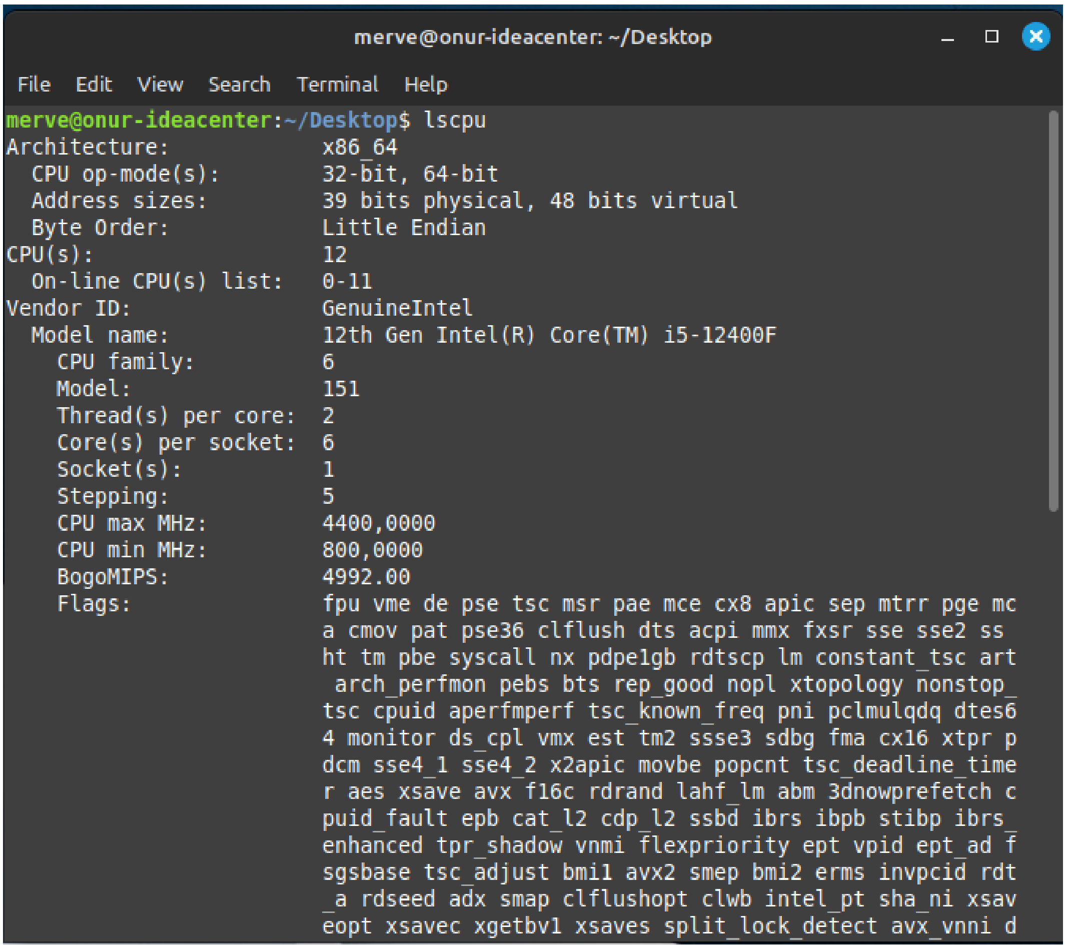
# Rellena la siguiente tabla con los datos solicitados:

$ cat /proc/cpuinfo


merve@onur-ideacenter: ~/Desktop
File Edit View Search Terminal Help

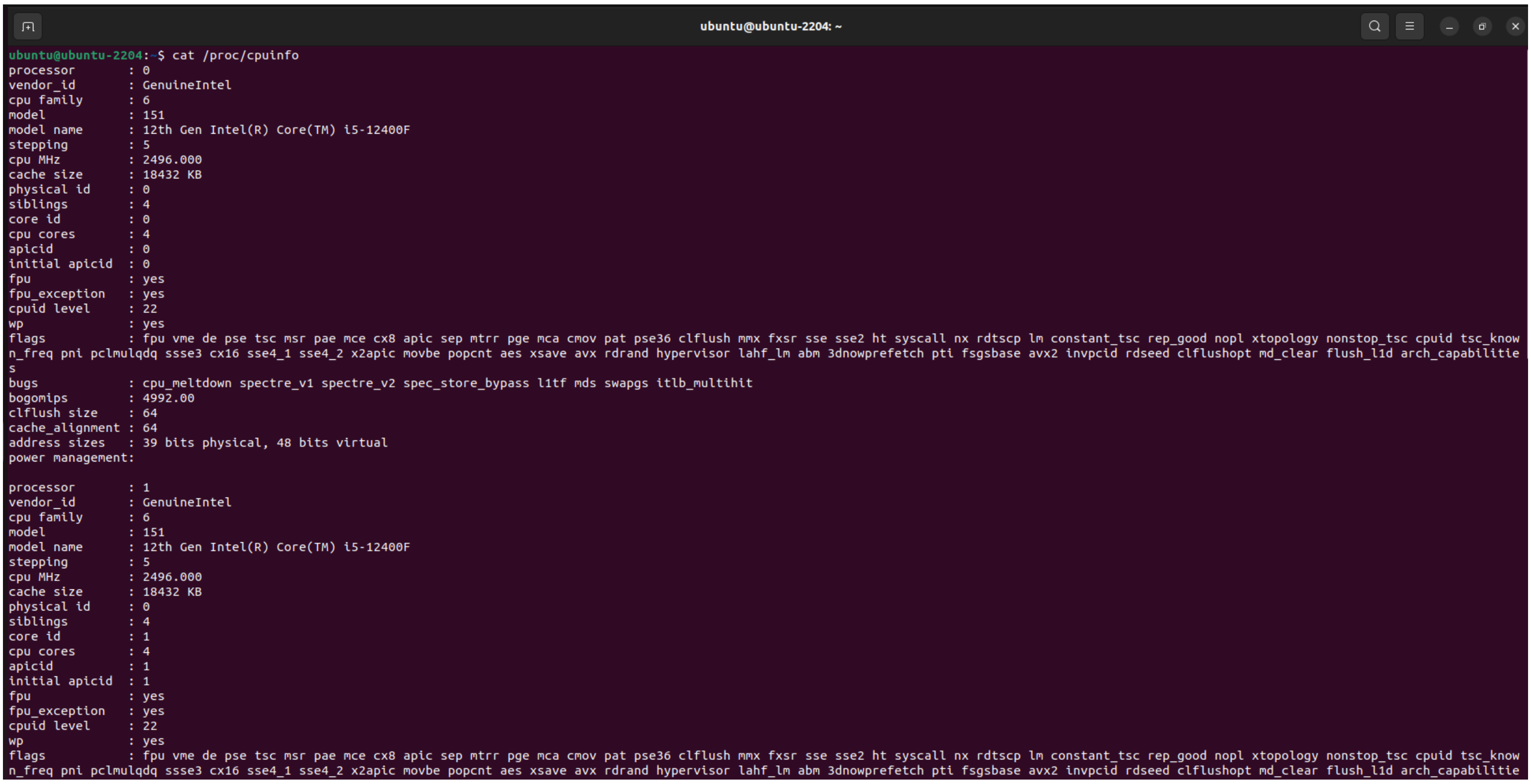
```
merve@onur-ideacenter:~/Desktop$ cat /proc/cpuinfo
processor       : 0
vendor_id       : GenuineIntel
cpu family      : 6
model           : 151
model name      : 12th Gen Intel(R) Core(TM) i5-12400F
stepping        : 5
microcode       : 0x2e
cpu MHz         : 2500.000
cache size      : 18432 KB
physical id     : 0
siblings        : 12
core id         : 0
cpu cores       : 6
apicid          : 0
initial apicid  : 0
fpu             : yes
fpu_exception   : yes
cpuid level     : 32
wp              : yes
flags           : fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art arch_perfmon pebs bts rep_good nopl xtopology nonstop_
tsc cpuid aperfmperf tsc_known_freq pni pclmulqdq dtes64 monitor ds_cpl vmx est tm2 ssse3 sdbg fma cx16 xtpr pdcm sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsave avx f16c rdrand lahf_lm abm 3dnowprefetch cpuid_fault epb cat
_l2 cdp_l2 ssbd ibrs ibpb stibp ibrs_enhanced tpr_shadow vnmi flexpriority ept vpid ept_ad fsgsbase tsc_adjust bmi1 avx2 smep bmi2 erms invpcid rdt_a rdseed adx smap clflushopt clwb intel_pt sha_ni xsaveopt xsavec xgetbv1 xsaves split_loc
k_detect avx_vnni dtherm ida arat pln pts hwp hwp_notify hwp_act_window hwp_epp hwp_pkg_req umip pku ospke waitpkg gfni vaes vpclmulqdq rdpid movdiri movdir64b fsrm md_clear serialize arch_lbr flush_l1d arch_capabilities
vmx flags       : vnmi preemption_timer posted_intr invvpid ept_x_only ept_ad ept_1gb flexpriority apicv tsc_offset vtpr mtf vapic ept vpid unrestricted_guest vapic_reg vid ple shadow_vmcs pml ept_mode_based_exec tsc_scaling usr_wait_paus
e
bugs            : spectre_v1 spectre_v2 spec_store_bypass swapgs eibrs_pbrsb
bogomips        : 4992.00
clflush size    : 64
cache_alignment : 64
address sizes   : 39 bits physical, 48 bits virtual
power management:

processor       : 1
vendor_id       : GenuineIntel
cpu family      : 6
model           : 151
model name      : 12th Gen Intel(R) Core(TM) i5-12400F
stepping        : 5
microcode       : 0x2e
cpu MHz         : 2500.000
cache size      : 18432 KB
physical id     : 0
siblings        : 12
core id         : 0
cpu cores       : 6
apicid          : 1
initial apicid  : 1
fpu             : yes
fpu_exception   : yes
cpuid level     : 32
wp              : yes
flags           : fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art arch_perfmon pebs bts rep_good nopl xtopology nonstop_
tsc cpuid aperfmperf tsc_known_freq pni pclmulqdq dtes64 monitor ds_cpl vmx est tm2 ssse3 sdbg fma cx16 xtpr pdcm sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_timer aes xsave avx f16c rdrand lahf_lm abm 3dnowprefetch cpuid_fault epb cat
_l2 cdp_l2 ssbd ibrs ibpb stibp ibrs_enhanced tpr_shadow vnmi flexpriority ept vpid ept_ad fsgsbase tsc_adjust bmi1 avx2 smep bmi2 erms invpcid rdt_a rdseed adx smap clflushopt clwb intel_pt sha_ni xsaveopt xsavec xgetbv1 xsaves split_loc
k_detect avx_vnni dtherm ida arat pln pts hwp hwp_notify hwp_act_window hwp_epp hwp_pkg_req umip pku ospke waitpkg gfni vaes vpclmulqdq rdpid movdiri movdir64b fsrm md_clear serialize arch_lbr flush_l1d arch_capabilities
vmx flags       : vnmi preemption_timer posted_intr invvpid ept_x_only ept_ad ept_1gb flexpriority apicv tsc_offset vtpr mtf vapic ept vpid unrestricted_guest vapic_reg vid ple shadow_vmcs pml ept_mode_based_exec tsc_scaling usr_wait_paus
```

$ lscpu


merve@onur-ideacenter: ~/Desktop
File Edit View Search Terminal Help

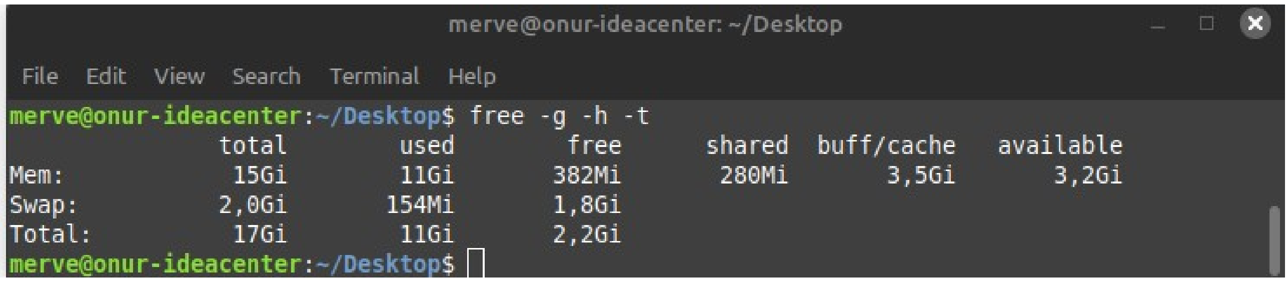
```
merve@onur-ideacenter:~/Desktop$ lscpu
Architecture:            x86_64
  CPU op-mode(s):        32-bit, 64-bit
  Address sizes:         39 bits physical, 48 bits virtual
  Byte Order:            Little Endian
CPU(s):                  12
  On-line CPU(s) list:   0-11
Vendor ID:               GenuineIntel
  Model name:            12th Gen Intel(R) Core(TM) i5-12400F
    CPU family:          6
    Model:               151
    Thread(s) per core:  2
    Core(s) per socket:  6
    Socket(s):           1
    Stepping:            5
    CPU max MHz:         4400,0000
    CPU min MHz:         800,0000
    BogoMIPS:            4992.00
    Flags:               fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mc
                         a cmov pat pse36 clflush dts acpi mmx fxsr sse sse2 ss
                         ht tm pbe syscall nx pdpe1gb rdtscp lm constant_tsc art
                          arch_perfmon pebs bts rep_good nopl xtopology nonstop_
                         tsc cpuid aperfmperf tsc_known_freq pni pclmulqdq dtes6
                         4 monitor ds_cpl vmx est tm2 ssse3 sdbg fma cx16 xtpr p
                         dcm sse4_1 sse4_2 x2apic movbe popcnt tsc_deadline_time
                         r aes xsave avx f16c rdrand lahf_lm abm 3dnowprefetch c
                         puid_fault epb cat_l2 cdp_l2 ssbd ibrs ibpb stibp ibrs_
                         enhanced tpr_shadow vnmi flexpriority ept vpid ept_ad f
                         sgsbase tsc_adjust bmi1 avx2 smep bmi2 erms invpcid rdt
                         _a rdseed adx smap clflushopt clwb intel_pt sha_ni xsav
                         eopt xsavec xgetbv1 xsaves split_lock_detect avx_vnni d
```

En Máquina Virtual



```
ubuntu@ubuntu-2204:~$ cat /proc/cpuinfo
processor       : 0
vendor_id       : GenuineIntel
cpu family      : 6
model           : 151
model name      : 12th Gen Intel(R) Core(TM) i5-12400F
stepping        : 5
cpu MHz         : 2496.000
cache size      : 18432 KB
physical id     : 0
siblings        : 4
core id         : 0
cpu cores       : 4
apicid          : 0
initial apicid  : 0
fpu             : yes
fpu_exception   : yes
cpuid level     : 22
wp              : yes
flags           : fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush mmx fxsr sse sse2 ht syscall nx rdtscp lm constant_tsc rep_good nopl xtopology nonstop_tsc cpuid tsc_know
n_freq pni pclmulqdq ssse3 cx16 sse4_1 sse4_2 x2apic movbe popcnt aes xsave avx rdrand hypervisor lahf_lm abm 3dnowprefetch pti fsgsbase avx2 invpcid rdseed clflushopt md_clear flush_l1d arch_capabilitie
s
bugs            : cpu_meltdown spectre_v1 spectre_v2 spec_store_bypass l1tf mds swapgs itlb_multihit
bogomips        : 4992.00
clflush size    : 64
cache_alignment : 64
address sizes   : 39 bits physical, 48 bits virtual
power management:

processor       : 1
vendor_id       : GenuineIntel
cpu family      : 6
model           : 151
model name      : 12th Gen Intel(R) Core(TM) i5-12400F
stepping        : 5
cpu MHz         : 2496.000
cache size      : 18432 KB
physical id     : 0
siblings        : 4
core id         : 1
cpu cores       : 4
apicid          : 1
initial apicid  : 1
fpu             : yes
fpu_exception   : yes
cpuid level     : 22
wp              : yes
flags           : fpu vme de pse tsc msr pae mce cx8 apic sep mtrr pge mca cmov pat pse36 clflush mmx fxsr sse sse2 ht syscall nx rdtscp lm constant_tsc rep_good nopl xtopology nonstop_tsc cpuid tsc_know
n_freq pni pclmulqdq ssse3 cx16 sse4_1 sse4_2 x2apic movbe popcnt aes xsave avx rdrand hypervisor lahf_lm abm 3dnowprefetch pti fsgsbase avx2 invpcid rdseed clflushopt md_clear flush_l1d arch_capabilitie
```

$ free -g -h -t

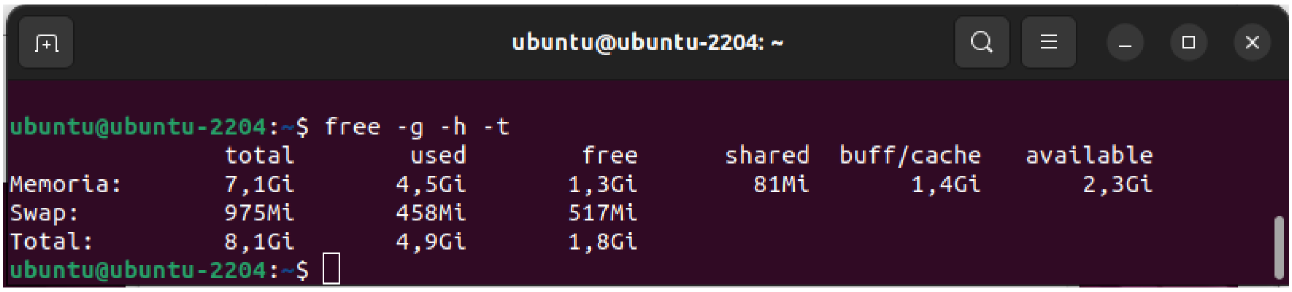


merve@onur-ideacenter: ~/Desktop
File Edit View Search Terminal Help

```
merve@onur-ideacenter:~/Desktop$ free -g -h -t
              total        used        free      shared  buff/cache   available
Mem:           15Gi        11Gi       382Mi       280Mi       3,5Gi       3,2Gi
Swap:         2,0Gi       154Mi       1,8Gi
Total:         17Gi        11Gi       2,2Gi
merve@onur-ideacenter:~/Desktop$
```

Memoria asignada a Máquina Virtual

$ free -g -h -t



```
ubuntu@ubuntu-2204:~$ free -g -h -t
               total        used        free      shared  buff/cache   available
Memoria:       7,1Gi       4,5Gi       1,3Gi        81Mi       1,4Gi       2,3Gi
Swap:          975Mi       458Mi       517Mi
Total:         8,1Gi       4,9Gi       1,8Gi
ubuntu@ubuntu-2204:~$
```

CPU	Memoria PC	Memoria asignada a Máquina Virtual
12th Gen Intel i5	16 GB	7

# Rellena esta tabla con el tiempo (segundos) que tarda en realizarse cada consulta.

Consultas	Pig Local	Pig MapReduce	Hive
C1	35 s	1 min 42 s	26 s
C2	10 s	1 min 21 s	36 s
C3	11 s	1 min 1 s	36 s
C4	11 s	1 min 18s	35 s

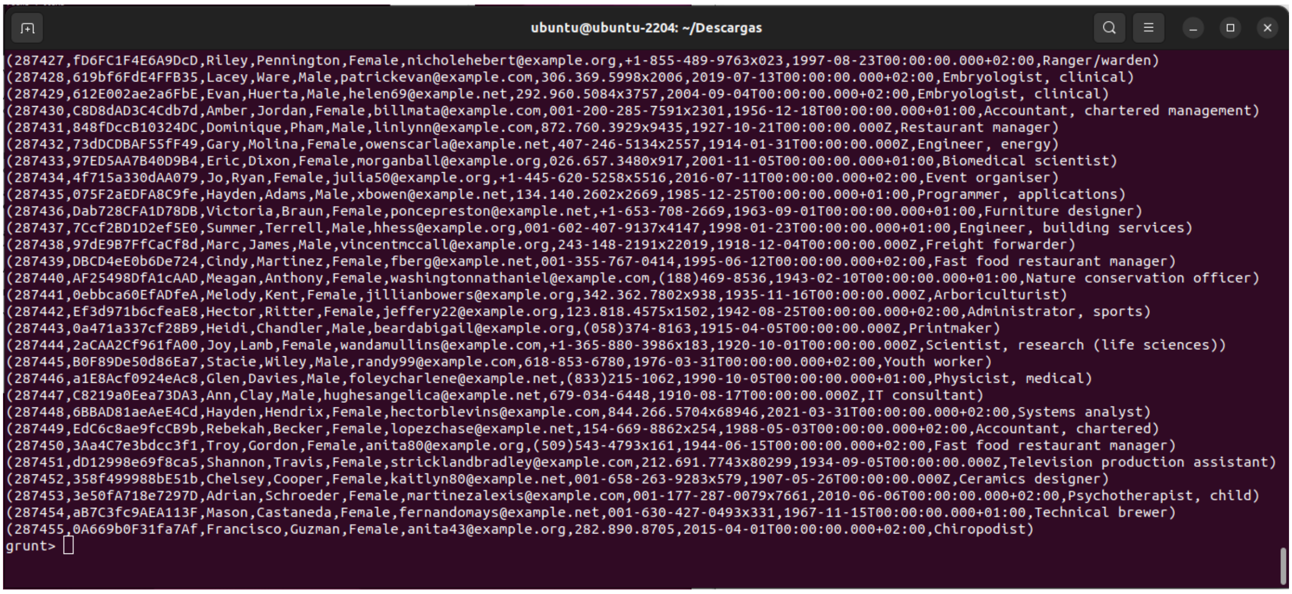
# Analiza los resultados obtenidos e intenta justificar porqué se obtienen esos resultados. ¿Cuál es más rápido Pig o Hive? ¿Por qué?

Según mis resultados, Pig en modo MapReduce es el más lento y Pig en modo local es el más rápido, mientras que Hive funciona ligeramente más lento que Pig en modo local.

2.2.1. Pig

grunt> people = LOAD '/home/ubuntu/Descargas/people-2000000.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE', 'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray, sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime, cargo:chararray);

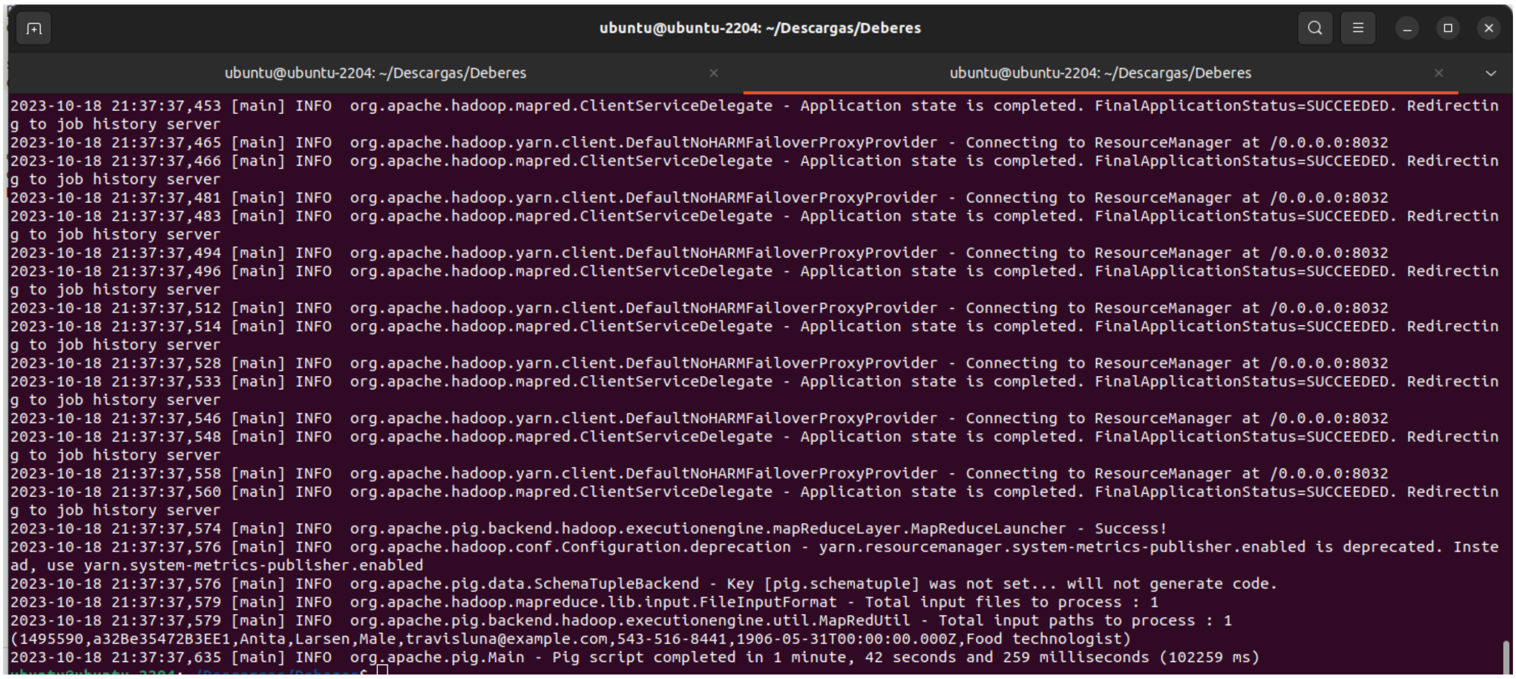
grunt> DUMP people;



```
(287427,fD6FC1F4E6A9DcD,Riley,Pennington,Female,nicholehebert@example.org,+1-855-489-9763x023,1997-08-23T00:00:00.000+02:00,Ranger/warden)
(287428,619bf6FdE4FFB35,Lacey,Ware,Male,patrickevan@example.com,306.369.5998x2006,2019-07-13T00:00:00.000+02:00,Embryologist, clinical)
(287429,612E002ae2a6FbE,Evan,Huerta,Male,helen69@example.net,292.960.5084x3757,2004-09-04T00:00:00.000+02:00,Embryologist, clinical)
(287430,C8D8dAD3C4Cdb7d,Amber,Jordan,Female,billmata@example.com,001-200-285-7591x2301,1956-12-18T00:00:00.000+01:00,Accountant, chartered management)
(287431,848fDccB10324DC,Dominique,Pham,Male,linlynn@example.com,872.760.3929x9435,1927-10-21T00:00:00.000Z,Restaurant manager)
(287432,73dDCDBAF55fF49,Gary,Molina,Female,owenscarla@example.net,407-246-5134x2557,1914-01-31T00:00:00.000Z,Engineer, energy)
(287433,97ED5AA7B40D9B4,Eric,Dixon,Female,morganball@example.org,026.657.3480x917,2001-11-05T00:00:00.000+01:00,Biomedical scientist)
(287434,4f715a330dAA079,Jo,Ryan,Female,julia50@example.org,+1-445-620-5258x5516,2016-07-11T00:00:00.000+02:00,Event organiser)
(287435,075F2aEDFA8C9fe,Hayden,Adams,Male,xbowen@example.net,134.140.2602x2669,1985-12-25T00:00:00.000+01:00,Programmer, applications)
(287436,Dab728CFA1D78DB,Victoria,Braun,Female,poncepreston@example.net,+1-653-708-2669,1963-09-01T00:00:00.000+01:00,Furniture designer)
(287437,7Ccf2BD1D2ef5E0,Summer,Terrell,Male,hhess@example.org,001-602-407-9137x4147,1998-01-23T00:00:00.000+01:00,Engineer, building services)
(287438,97dE9B7FfCaCf8d,Marc,James,Male,vincentmccall@example.org,243-148-2191x22019,1918-12-04T00:00:00.000Z,Freight forwarder)
(287439,DBCD4eE0b6De724,Cindy,Martinez,Female,fberg@example.net,001-355-767-0414,1995-06-12T00:00:00.000+02:00,Fast food restaurant manager)
(287440,AF25498DfA1cAAD,Meagan,Anthony,Female,washingtonnathaniel@example.com,(188)469-8536,1943-02-10T00:00:00.000+01:00,Nature conservation officer)
(287441,0ebbca60EfADfeA,Melody,Kent,Female,jillianbowers@example.org,342.362.7802x938,1935-11-16T00:00:00.000Z,Arboriculturist)
(287442,Ef3d971b6cfeaE8,Hector,Ritter,Female,jeffery22@example.org,123.818.4575x1502,1942-08-25T00:00:00.000+02:00,Administrator, sports)
(287443,0a471a337cf28B9,Heidi,Chandler,Male,beardabigail@example.org,(058)374-8163,1915-04-05T00:00:00.000Z,Printmaker)
(287444,2aCAA2Cf961fA00,Joy,Lamb,Female,wandamullins@example.com,+1-365-880-3986x183,1920-10-01T00:00:00.000Z,Scientist, research (life sciences))
(287445,B0F89De50d86Ea7,Stacie,Wiley,Male,randy99@example.com,618-853-6780,1976-03-31T00:00:00.000+02:00,Youth worker)
(287446,a1E8Acf0924eAc8,Glen,Davies,Male,foleycharlene@example.net,(833)215-1062,1990-10-05T00:00:00.000+01:00,Physicist, medical)
(287447,C8219a0Eea73DA3,Ann,Clay,Male,hughesangelica@example.net,679-034-6448,1910-08-17T00:00:00.000Z,IT consultant)
(287448,6BBAD81aeAeE4Cd,Hayden,Hendrix,Female,hectorblevins@example.com,844.266.5704x68946,2021-03-31T00:00:00.000+02:00,Systems analyst)
(287449,EdC6c8ae9fcCB9b,Rebekah,Becker,Female,lopezchase@example.net,154-669-8862x254,1988-05-03T00:00:00.000+02:00,Accountant, chartered)
(287450,3Aa4C7e3bdcc3f1,Troy,Gordon,Female,anita80@example.org,(509)543-4793x161,1944-06-15T00:00:00.000+02:00,Fast food restaurant manager)
(287451,dD12998e69f8ca5,Shannon,Travis,Female,stricklandbradley@example.com,212.691.7743x80299,1934-09-05T00:00:00.000Z,Television production assistant)
(287452,358f499988bE51b,Chelsey,Cooper,Female,kaitlyn80@example.net,001-658-263-9283x579,1907-05-26T00:00:00.000Z,Ceramics designer)
(287453,3e50fA718e7297D,Adrian,Schroeder,Female,martinezalexis@example.com,001-177-287-0079x7661,2010-06-06T00:00:00.000+02:00,Psychotherapist, child)
(287454,aB7C3fc9AEA113F,Mason,Castaneda,Female,fernandomays@example.net,001-630-427-0493x331,1967-11-15T00:00:00.000+01:00,Technical brewer)
(287455,0A669b0F31fa7Af,Francisco,Guzman,Female,anita43@example.org,282.890.8705,2015-04-01T00:00:00.000+02:00,Chiropodist)
grunt> 
```

En C1_mapreduce.pig script ($ pig C1_mapreduce.pig) :

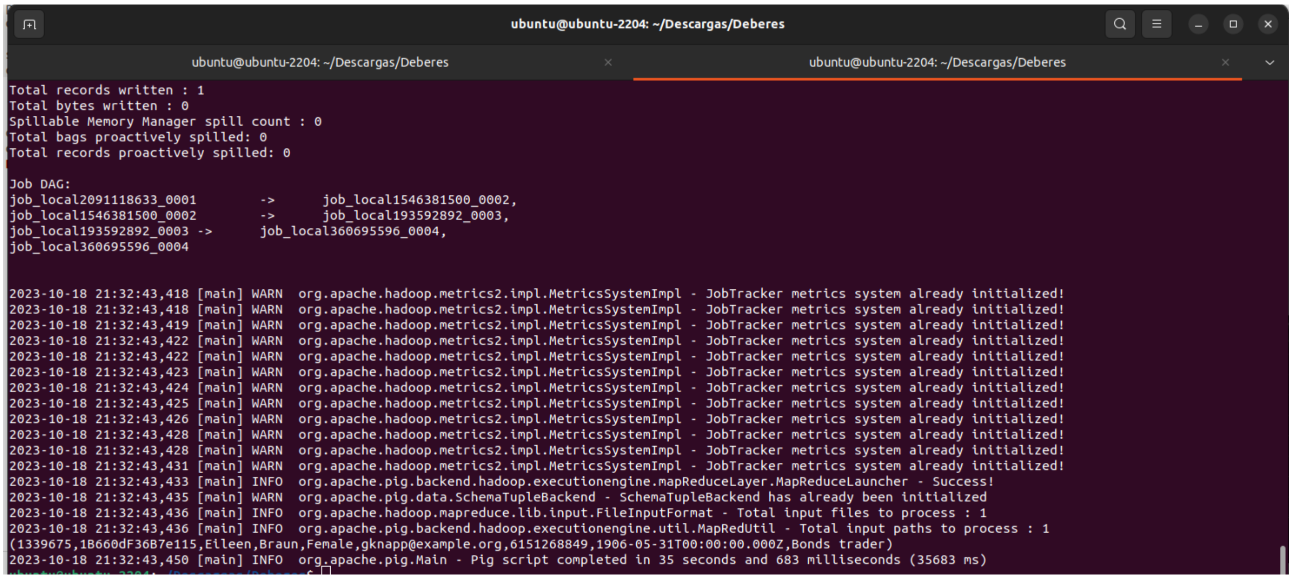
```
people = LOAD '/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- DUMP people;

oldest_person = ORDER people BY fecha_de_nacimiento ASC;
oldest_person = LIMIT oldest_person 1;
DUMP oldest_person;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
2023-10-18 21:37:37,453 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,465 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,466 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,481 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,483 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,494 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,496 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,512 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,514 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,528 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,533 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,546 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,548 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,558 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 21:37:37,560 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 21:37:37,574 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2023-10-18 21:37:37,576 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Inste
ad, use yarn.system-metrics-publisher.enabled
2023-10-18 21:37:37,576 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2023-10-18 21:37:37,579 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-18 21:37:37,579 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1495590,a32Be35472B3EE1,Anita,Larsen,Male,travisluna@example.com,543-516-8441,1906-05-31T00:00:00.000Z,Food technologist)
2023-10-18 21:37:37,635 [main] INFO  org.apache.pig.Main - Pig script completed in 1 minute, 42 seconds and 259 milliseconds (102259 ms)
```

En C1_pig.pig script ($ pig -x local C1_pig.pig) :

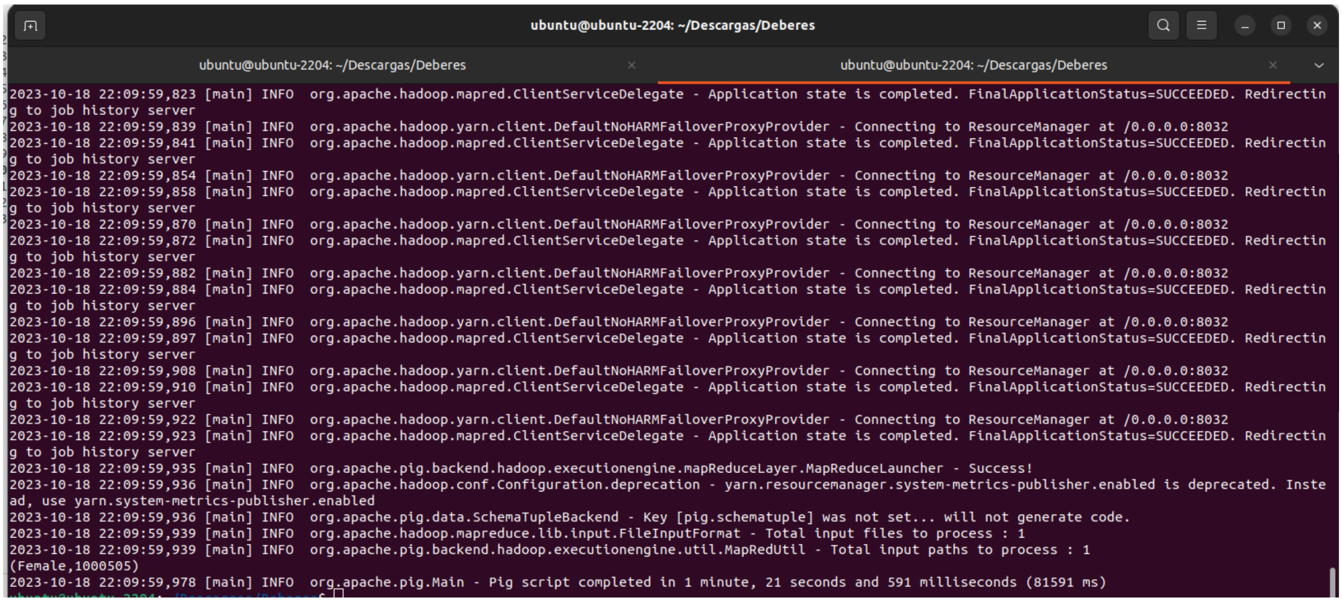
```
people = LOAD '/home/ubuntu/Descargas/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- DUMP people;

oldest_person = ORDER people BY fecha_de_nacimiento ASC;
oldest_person = LIMIT oldest_person 1;
DUMP oldest_person;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local2091118633_0001        ->      job_local1546381500_0002,
job_local1546381500_0002        ->      job_local193592892_0003,
job_local193592892_0003 ->      job_local360695596_0004,
job_local360695596_0004


2023-10-18 21:32:43,418 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,418 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,419 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,422 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,422 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,423 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,424 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,425 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,426 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,428 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,428 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,431 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 21:32:43,433 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2023-10-18 21:32:43,435 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-10-18 21:32:43,436 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-18 21:32:43,436 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1339675,1B660dF36B7e115,Eileen,Braun,Female,gknapp@example.org,6151268849,1906-05-31T00:00:00.000Z,Bonds trader)
2023-10-18 21:32:43,450 [main] INFO  org.apache.pig.Main - Pig script completed in 35 seconds and 683 milliseconds (35683 ms)
```

En C2_mapreduce.pig script ($ pig C2_mapreduce.pig) :

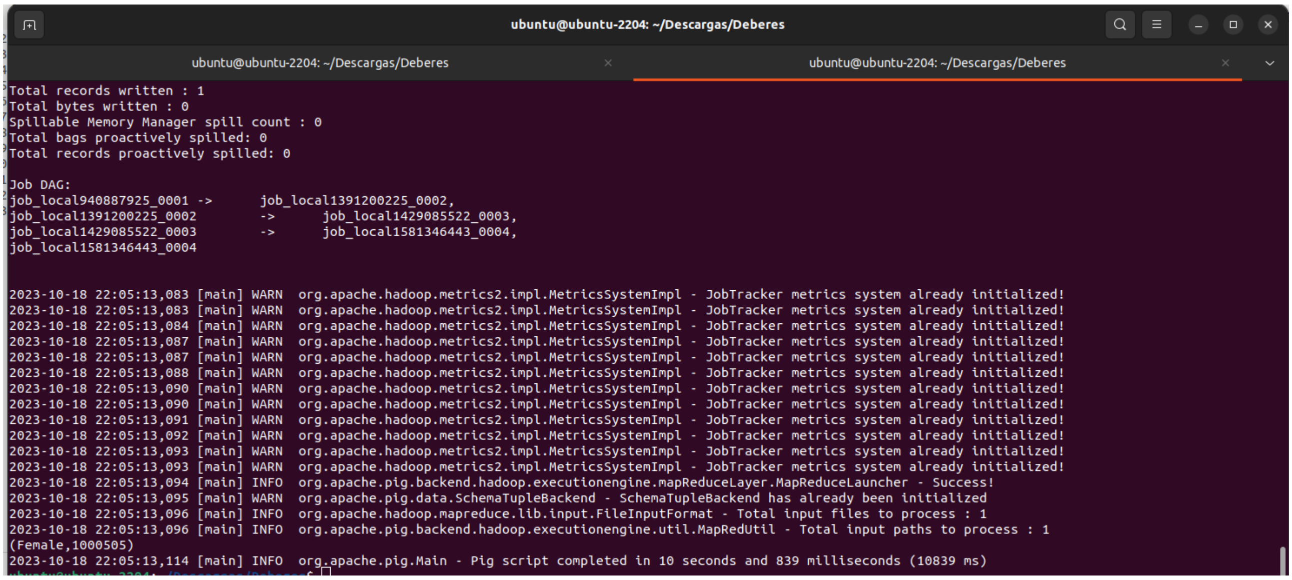
```
people = LOAD '/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- Count Each Gender
gender_el = GROUP people BY sexo;
gender_count = FOREACH gender_el GENERATE group AS sexo, COUNT(people) AS count;

-- DUMP gender_count;

-- Count Men & Women
sorted_data = ORDER gender_count BY count DESC;
max_gender = LIMIT sorted_data 1;
DUMP max_gender;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
2023-10-18 22:09:59,823 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,839 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,841 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,854 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,858 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,870 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,872 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,882 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,884 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,896 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,897 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,908 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,910 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,922 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:09:59,923 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:09:59,935 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2023-10-18 22:09:59,936 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Inste
ad, use yarn.system-metrics-publisher.enabled
2023-10-18 22:09:59,936 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2023-10-18 22:09:59,939 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-18 22:09:59,939 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Female,1000505)
2023-10-18 22:09:59,978 [main] INFO  org.apache.pig.Main - Pig script completed in 1 minute, 21 seconds and 591 milliseconds (81591 ms)
```

En C2_pig.pig script ($ pig -x local C2_pig.pig) :

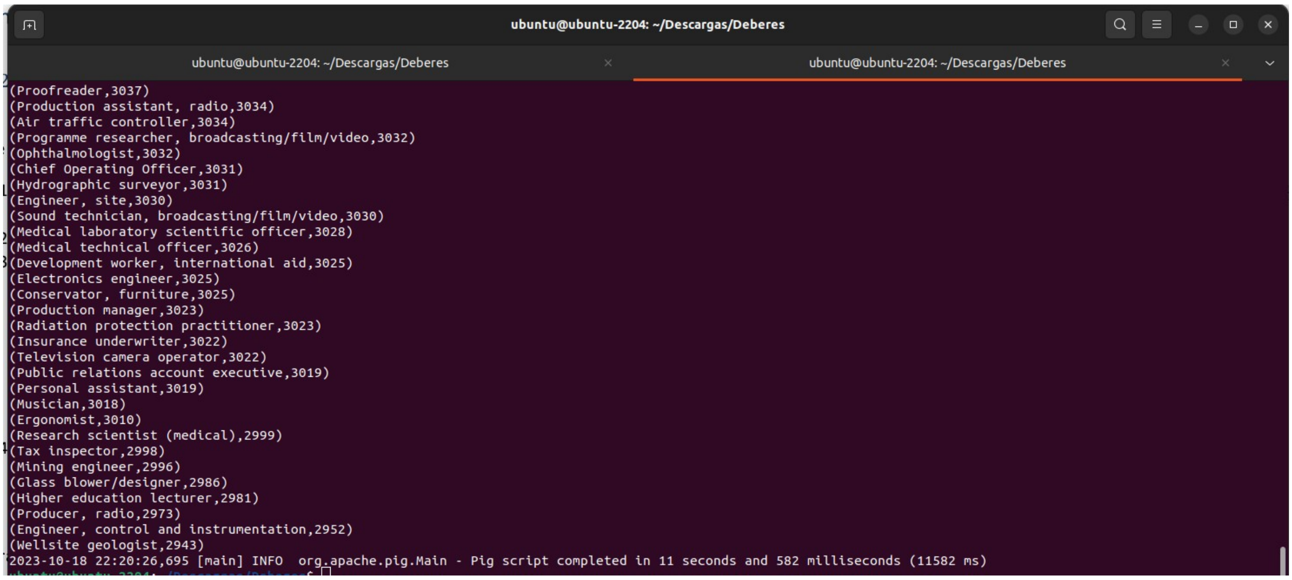
```
people = LOAD '/home/ubuntu/Descargas/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- Count Each Gender
gender_el = GROUP people BY sexo;
gender_count = FOREACH gender_el GENERATE group AS sexo, COUNT(people) AS count;

-- DUMP gender_count;

-- Count Men & Women
sorted_data = ORDER gender_count BY count DESC;
max_gender = LIMIT sorted_data 1;
DUMP max_gender;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local940887925_0001 ->      job_local1391200225_0002,
job_local1391200225_0002        ->      job_local1429085522_0003,
job_local1429085522_0003        ->      job_local1581346443_0004,
job_local1581346443_0004


2023-10-18 22:05:13,083 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,083 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,084 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,087 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,087 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,088 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,090 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,090 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,091 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,092 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,093 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,093 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:05:13,094 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2023-10-18 22:05:13,095 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-10-18 22:05:13,096 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-18 22:05:13,096 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Female,1000505)
2023-10-18 22:05:13,114 [main] INFO  org.apache.pig.Main - Pig script completed in 10 seconds and 839 milliseconds (10839 ms)
```

En C3_pig.pig script ($ pig -x local C3_pig.pig) :

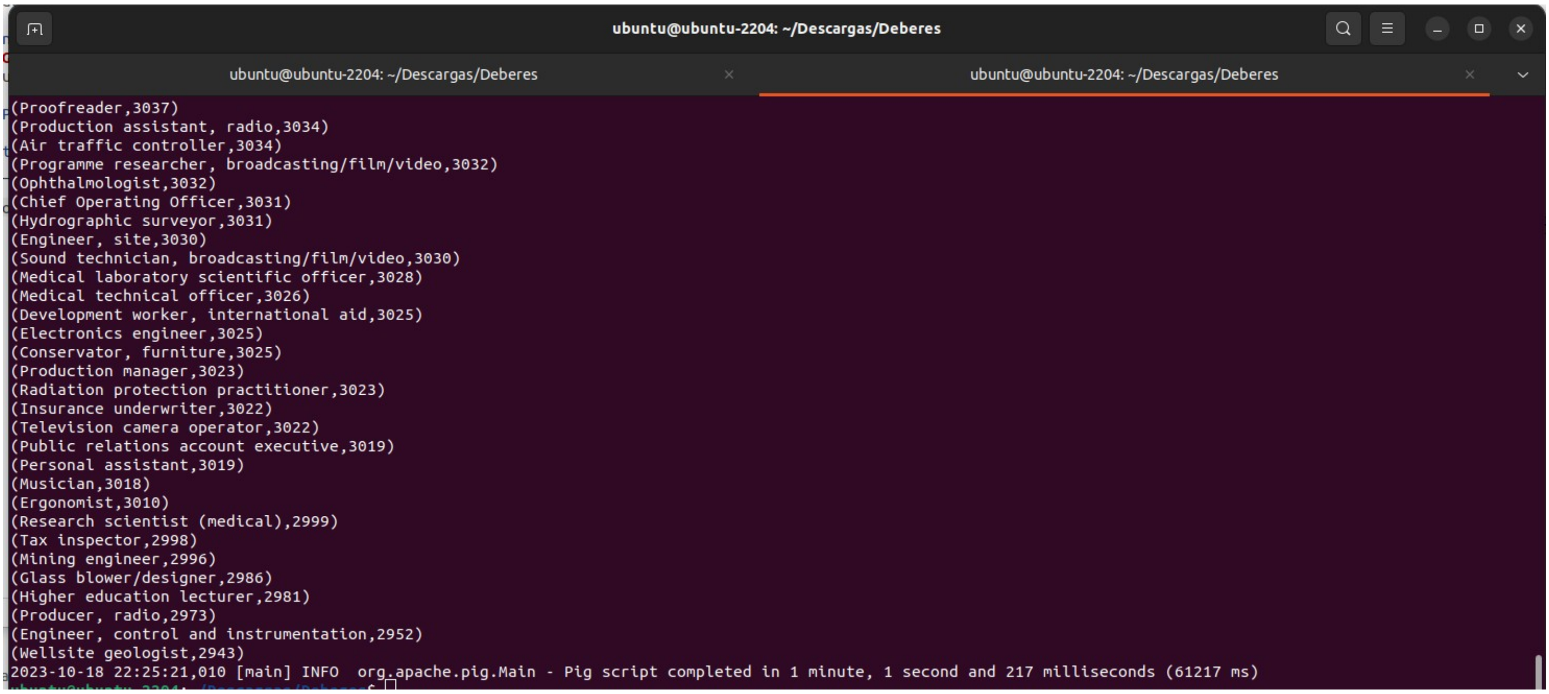
```
people = LOAD '/home/ubuntu/Descargas/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- Count Each Job
job = GROUP people BY cargo;
job_count = FOREACH job GENERATE group AS cargo, COUNT(people) AS count;

-- DUMP job_count;

-- Sort Job
sorted_data = ORDER job_count BY count DESC;

DUMP sorted_data;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
(Proofreader,3037)
(Production assistant, radio,3034)
(Air traffic controller,3034)
(Programme researcher, broadcasting/film/video,3032)
(Ophthalmologist,3032)
(Chief Operating Officer,3031)
(Hydrographic surveyor,3031)
(Engineer, site,3030)
(Sound technician, broadcasting/film/video,3030)
(Medical laboratory scientific officer,3028)
(Medical technical officer,3026)
(Development worker, international aid,3025)
(Electronics engineer,3025)
(Conservator, furniture,3025)
(Production manager,3023)
(Radiation protection practitioner,3023)
(Insurance underwriter,3022)
(Television camera operator,3022)
(Public relations account executive,3019)
(Personal assistant,3019)
(Musician,3018)
(Ergonomist,3010)
(Research scientist (medical),2999)
(Tax inspector,2998)
(Mining engineer,2996)
(Glass blower/designer,2986)
(Higher education lecturer,2981)
(Producer, radio,2973)
(Engineer, control and instrumentation,2952)
(Wellsite geologist,2943)
2023-10-18 22:20:26,695 [main] INFO  org.apache.pig.Main - Pig script completed in 11 seconds and 582 milliseconds (11582 ms)
```

En C3_mapreduce.pig script ($ pig C3_mapreduce.pig) :

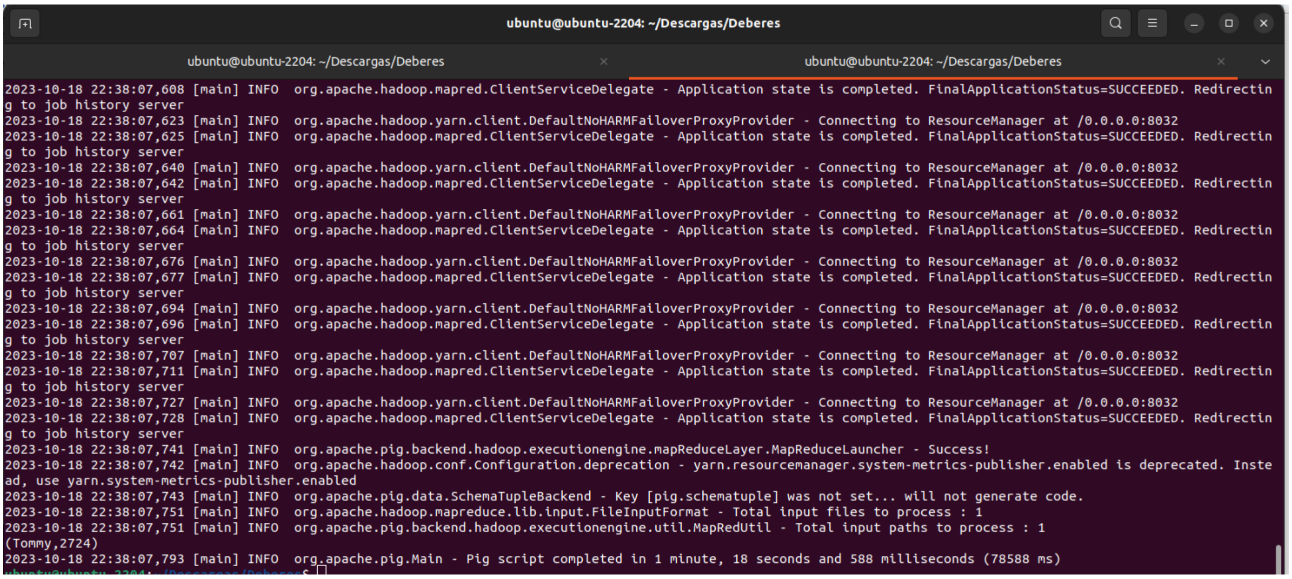
```
people = LOAD '/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- Count Each Job
job = GROUP people BY cargo;
job_count = FOREACH job GENERATE group AS cargo, COUNT(people) AS count;

-- DUMP job_count;

-- Sort Job
sorted_data = ORDER job_count BY count DESC;

DUMP sorted_data;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
(Proofreader,3037)
(Production assistant, radio,3034)
(Air traffic controller,3034)
(Programme researcher, broadcasting/film/video,3032)
(Ophthalmologist,3032)
(Chief Operating Officer,3031)
(Hydrographic surveyor,3031)
(Engineer, site,3030)
(Sound technician, broadcasting/film/video,3030)
(Medical laboratory scientific officer,3028)
(Medical technical officer,3026)
(Development worker, international aid,3025)
(Electronics engineer,3025)
(Conservator, furniture,3025)
(Production manager,3023)
(Radiation protection practitioner,3023)
(Insurance underwriter,3022)
(Television camera operator,3022)
(Public relations account executive,3019)
(Personal assistant,3019)
(Musician,3018)
(Ergonomist,3010)
(Research scientist (medical),2999)
(Tax inspector,2998)
(Mining engineer,2996)
(Glass blower/designer,2986)
(Higher education lecturer,2981)
(Producer, radio,2973)
(Engineer, control and instrumentation,2952)
(Wellsite geologist,2943)
2023-10-18 22:25:21,010 [main] INFO  org.apache.pig.Main - Pig script completed in 1 minute, 1 second and 217 milliseconds (61217 ms)
ubuntu@ubuntu-2204: ~/Descargas/Deberes$
```

En C4_mapreduce.pig script ($ pig C4_mapreduce.pig) :

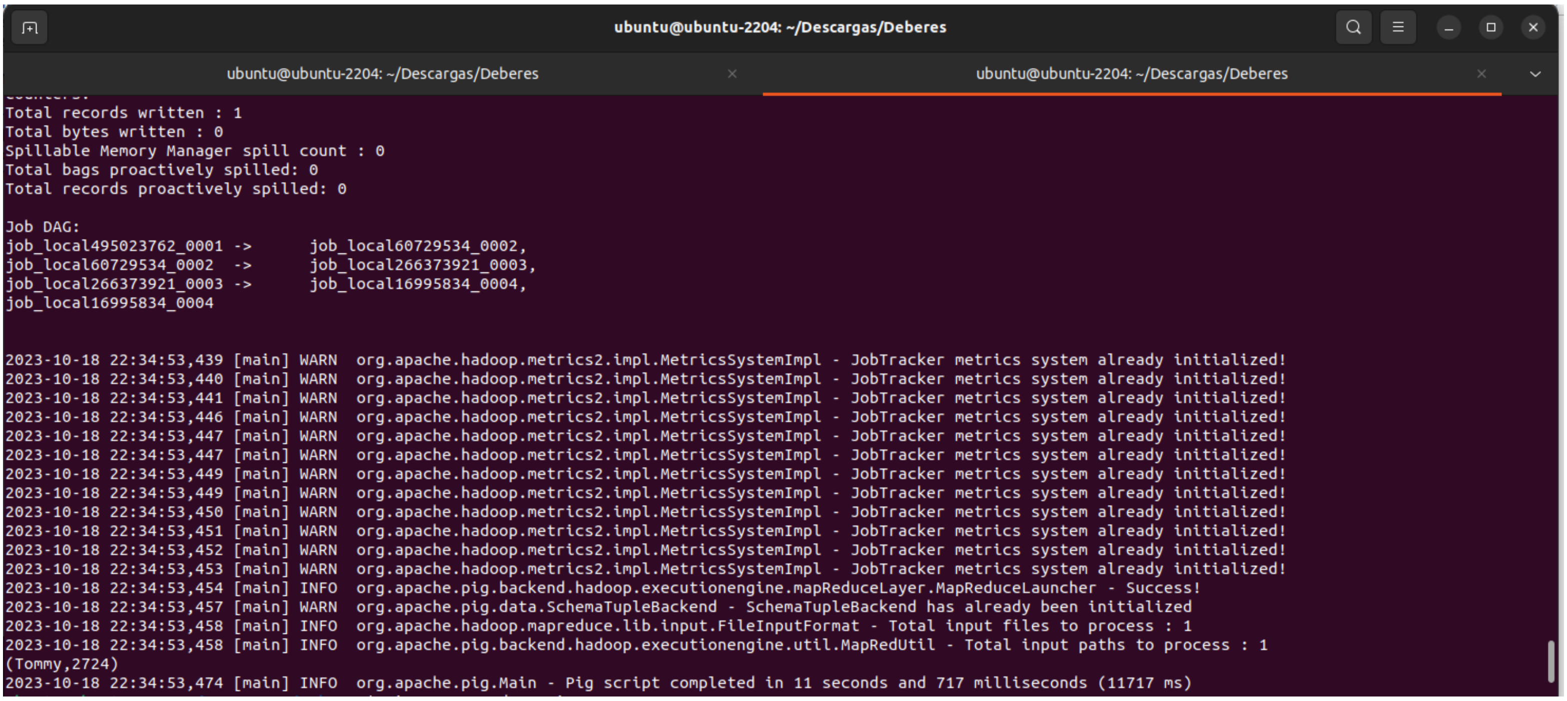
```
-- Indica el nombre menos repetido
people = LOAD '/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- Count Nombres
nombre = GROUP people BY nombre;
nombre_count = FOREACH nombre GENERATE group AS nombre, COUNT(people) AS count;

-- DUMP nombre_count;

-- Sort Nombres
sorted_data = ORDER nombre_count BY count ASC;
min_nombre = LIMIT sorted_data 1;
DUMP min_nombre;
```


ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
2023-10-18 22:38:07,608 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,623 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,625 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,640 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,642 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,661 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,664 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,676 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,677 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,694 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,696 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,707 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,711 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,727 [main] INFO  org.apache.hadoop.yarn.client.DefaultNoHARMFailoverProxyProvider - Connecting to ResourceManager at /0.0.0.0:8032
2023-10-18 22:38:07,728 [main] INFO  org.apache.hadoop.mapred.ClientServiceDelegate - Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirectin
g to job history server
2023-10-18 22:38:07,741 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2023-10-18 22:38:07,742 [main] INFO  org.apache.hadoop.conf.Configuration.deprecation - yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Inste
ad, use yarn.system-metrics-publisher.enabled
2023-10-18 22:38:07,743 [main] INFO  org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2023-10-18 22:38:07,751 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-18 22:38:07,751 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Tommy,2724)
2023-10-18 22:38:07,793 [main] INFO  org.apache.pig.Main - Pig script completed in 1 minute, 18 seconds and 588 milliseconds (78588 ms)
ubuntu@ubuntu-2204:~/Descargas/Deberes$
```

En C4_pig.pig script ($ pig -x local C4_pig.pig) :

```
-- Indica el nombre menos repetido

people = LOAD '/home/ubuntu/Descargas/people-2000000.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'YES_MULTILINE', 'NOCHANGE',
'SKIP_INPUT_HEADER') AS (indice:int, id:chararray, nombre:chararray, apellidos:chararray,
sexo:chararray, correo:chararray, telefono:chararray, fecha_de_nacimiento:datetime,
cargo:chararray);

-- Count Nombres
nombre = GROUP people BY nombre;
nombre_count = FOREACH nombre GENERATE group AS nombre, COUNT(people) AS count;

-- DUMP nombre_count;

-- Sort Nombres
sorted_data = ORDER nombre_count BY count ASC;
min_nombre = LIMIT sorted_data 1;
DUMP min_nombre;
```

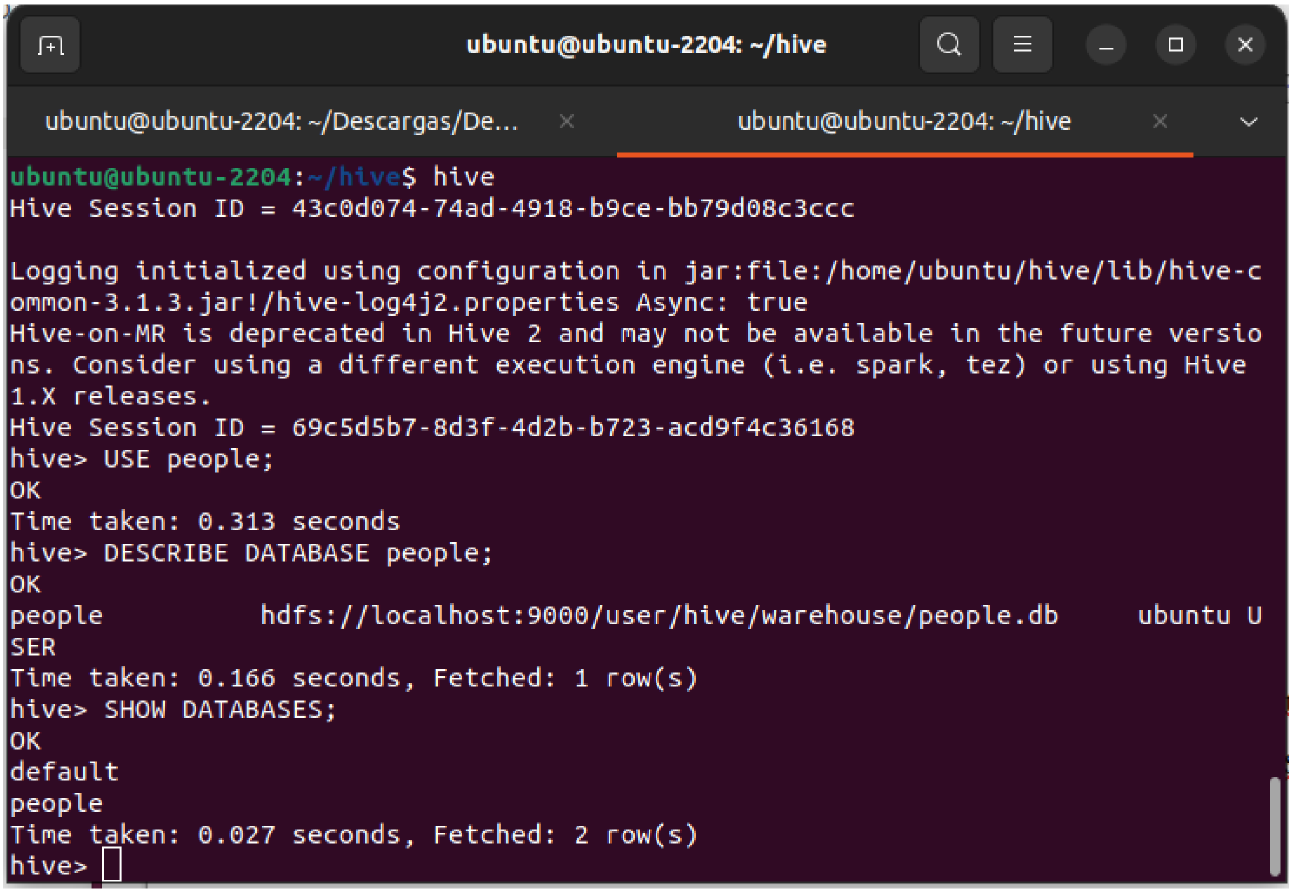

ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/Descargas/Deberes

```
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local495023762_0001 ->      job_local60729534_0002,
job_local60729534_0002  ->      job_local266373921_0003,
job_local266373921_0003 ->      job_local16995834_0004,
job_local16995834_0004


2023-10-18 22:34:53,439 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,440 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,441 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,446 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,447 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,447 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,449 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,449 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,450 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,451 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,452 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,453 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2023-10-18 22:34:53,454 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2023-10-18 22:34:53,457 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2023-10-18 22:34:53,458 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2023-10-18 22:34:53,458 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Tommy,2724)
2023-10-18 22:34:53,474 [main] INFO  org.apache.pig.Main - Pig script completed in 11 seconds and 717 milliseconds (11717 ms)
```

2.2.2. Hive

1) a) Crea una base de datos que se llame people y actívala.

CREATE DATABASE people;
USE people;
DESCRIBE DATABASE people;
SHOW DATABASES;


ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/Descargas/De...
ubuntu@ubuntu-2204: ~/hive

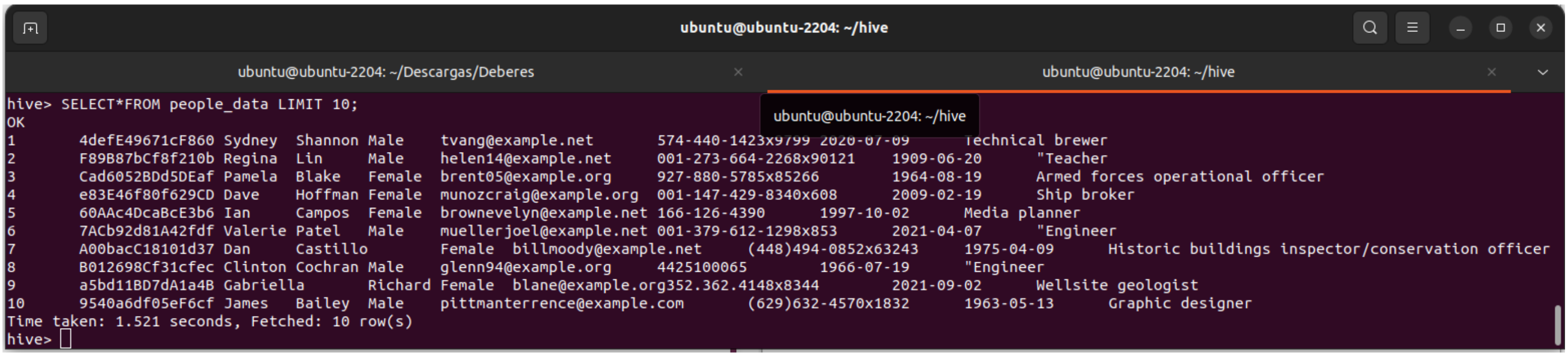
```
ubuntu@ubuntu-2204:~/hive$ hive
Hive Session ID = 43c0d074-74ad-4918-b9ce-bb79d08c3ccc

Logging initialized using configuration in jar:file:/home/ubuntu/hive/lib/hive-c
ommon-3.1.3.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versio
ns. Consider using a different execution engine (i.e. spark, tez) or using Hive
1.X releases.
Hive Session ID = 69c5d5b7-8d3f-4d2b-b723-acd9f4c36168
hive> USE people;
OK
Time taken: 0.313 seconds
hive> DESCRIBE DATABASE people;
OK
people          hdfs://localhost:9000/user/hive/warehouse/people.db     ubuntu U
SER
Time taken: 0.166 seconds, Fetched: 1 row(s)
hive> SHOW DATABASES;
OK
default
people
Time taken: 0.027 seconds, Fetched: 2 row(s)
hive> 
```

b) Crea una tabla para alojar los datos del dataset. Debes tener en cuenta el formato del fichero y los campos que tiene.

CREATE TABLE people_data (indice int, id string, nombre string, apellidos string, sexo string, correo string, telefono string, fecha_de_nacimiento date, cargo string) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE TBLPROPERTIES ("skip.header.line.count"="1");

c) Carga el fichero en la tabla y haz una primera selección de todos los campos con un máximo de 10 registros.

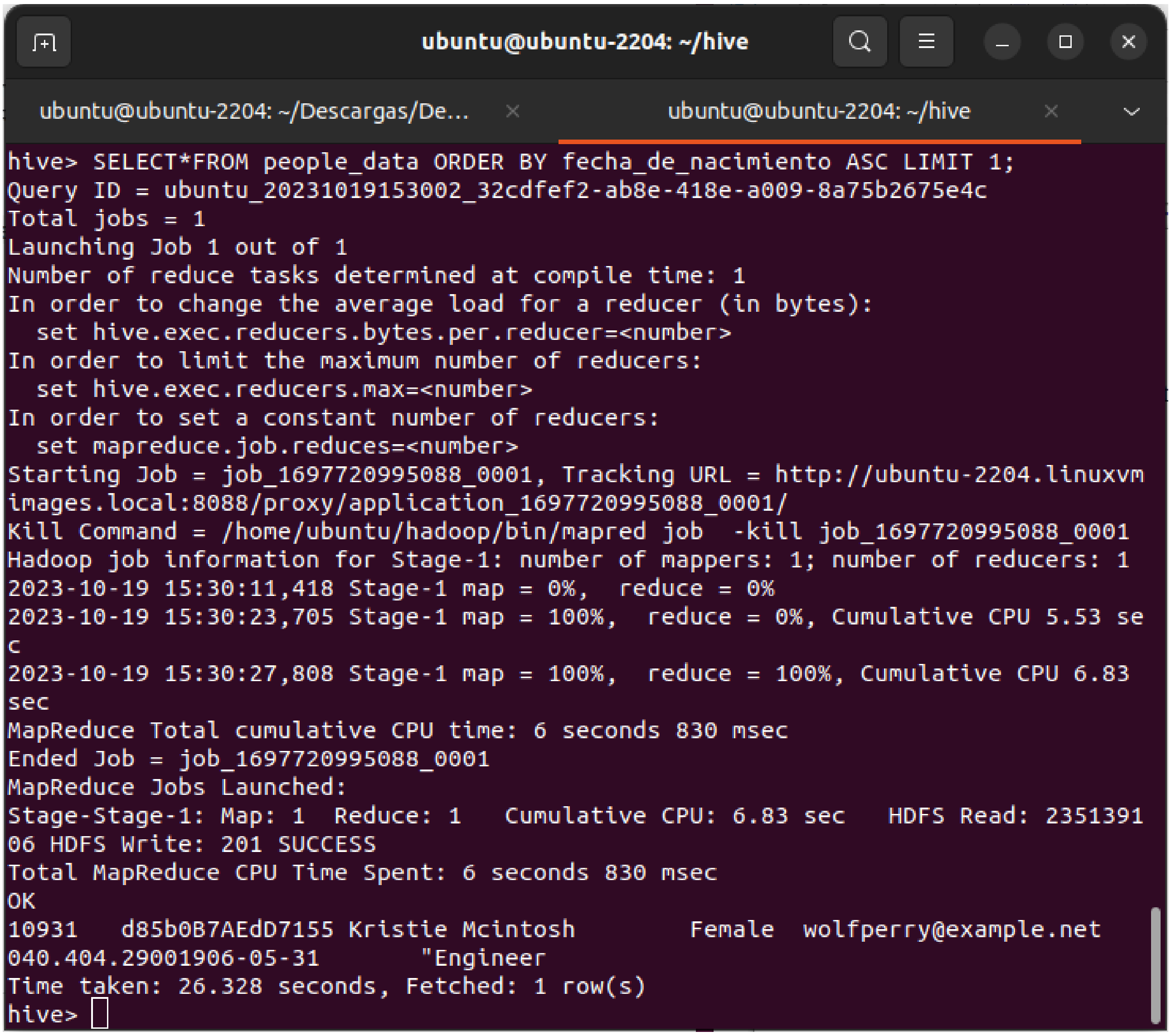
LOAD DATA INPATH '/people-2000000.csv' OVERWRITE INTO TABLE people_data;
SELECT*FROM people_data;
SELECT*FROM people_data LIMIT 10;


ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/hive

```
hive> SELECT*FROM people_data LIMIT 10;
OK
1       4defE49671cF860 Sydney  Shannon Male    tvang@example.net       574-440-1423x9799 2020-07-09      Technical brewer
2       F89B87bCf8f210b Regina  Lin     Male    helen14@example.net     001-273-664-2268x90121    1909-06-20      "Teacher
3       Cad6052BDd5DEaf Pamela  Blake   Female  brent05@example.org     927-880-5785x85266        1964-08-19      Armed forces operational officer
4       e83E46f80f629CD Dave    Hoffman Female  munozcraig@example.org  001-147-429-8340x608      2009-02-19      Ship broker
5       60AAc4DcaBcE3b6 Ian     Campos  Female  brownevelyn@example.net 166-126-4390      1997-10-02      Media planner
6       7ACb92d81A42fdf Valerie Patel   Male    muellerjoel@example.net 001-379-612-1298x853      2021-04-07      "Engineer
7       A00bacC18101d37 Dan     Castillo        Female  billmoody@example.net     (448)494-0852x63243     1975-04-09      Historic buildings inspector/conservation officer
8       B012698Cf31cfec Clinton Cochran Male    glenn94@example.org     4425100065        1966-07-19      "Engineer
9       a5bd11BD7dA1a4B Gabriella       Richard Female  blane@example.org352.362.4148x8344        2021-09-02      Wellsite geologist
10      9540a6df05eF6cf James   Bailey  Male    pittmanterrence@example.com       (629)632-4570x1832      1963-05-13      Graphic designer
Time taken: 1.521 seconds, Fetched: 10 row(s)
hive> 
```

En C1 hive

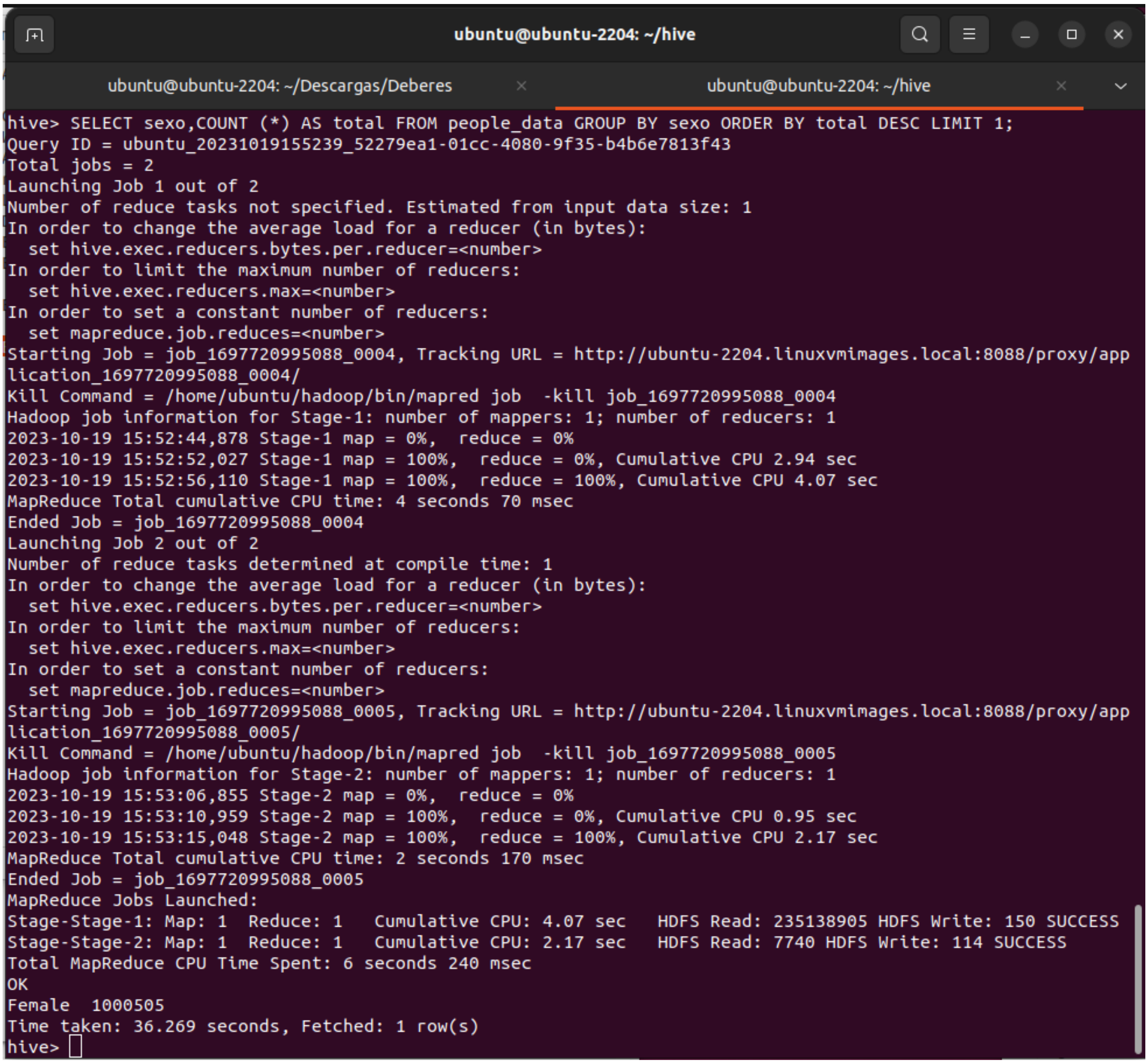
SELECT*FROM people_data ORDER BY fecha_de_nacimiento ASC LIMIT 1;


ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/Descargas/De...
ubuntu@ubuntu-2204: ~/hive

```
hive> SELECT*FROM people_data ORDER BY fecha_de_nacimiento ASC LIMIT 1;
Query ID = ubuntu_20231019153002_32cdfef2-ab8e-418e-a009-8a75b2675e4c
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1697720995088_0001, Tracking URL = http://ubuntu-2204.linuxvm
images.local:8088/proxy/application_1697720995088_0001/
Kill Command = /home/ubuntu/hadoop/bin/mapred job  -kill job_1697720995088_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-10-19 15:30:11,418 Stage-1 map = 0%,  reduce = 0%
2023-10-19 15:30:23,705 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 5.53 se
c
2023-10-19 15:30:27,808 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.83
sec
MapReduce Total cumulative CPU time: 6 seconds 830 msec
Ended Job = job_1697720995088_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.83 sec   HDFS Read: 2351391
06 HDFS Write: 201 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 830 msec
OK
10931   d85b0B7AEdD7155 Kristie Mcintosh        Female  wolfperry@example.net
040.404.29001906-05-31        "Engineer
Time taken: 26.328 seconds, Fetched: 1 row(s)
hive>
```

En C2 hive

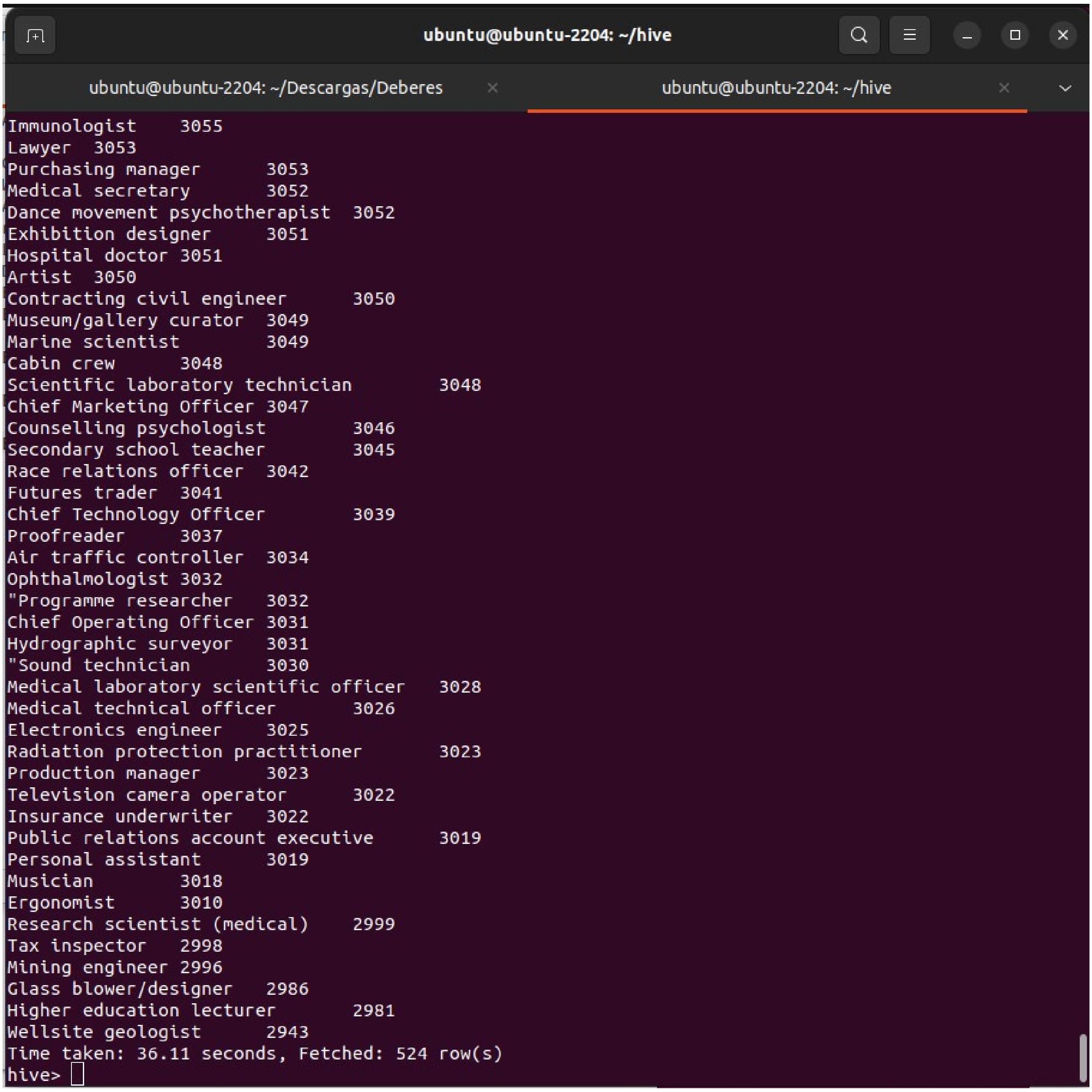
SELECT sexo, COUNT (*) AS total FROM people_data GROUP BY sexo ORDER BY total DESC LIMIT 1;


ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/hive

```
hive> SELECT sexo,COUNT (*) AS total FROM people_data GROUP BY sexo ORDER BY total DESC LIMIT 1;
Query ID = ubuntu_20231019155239_52279ea1-01cc-4080-9f35-b4b6e7813f43
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1697720995088_0004, Tracking URL = http://ubuntu-2204.linuxvmimages.local:8088/proxy/app
lication_1697720995088_0004/
Kill Command = /home/ubuntu/hadoop/bin/mapred job  -kill job_1697720995088_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-10-19 15:52:44,878 Stage-1 map = 0%,  reduce = 0%
2023-10-19 15:52:52,027 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.94 sec
2023-10-19 15:52:56,110 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.07 sec
MapReduce Total cumulative CPU time: 4 seconds 70 msec
Ended Job = job_1697720995088_0004
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1697720995088_0005, Tracking URL = http://ubuntu-2204.linuxvmimages.local:8088/proxy/app
lication_1697720995088_0005/
Kill Command = /home/ubuntu/hadoop/bin/mapred job  -kill job_1697720995088_0005
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-10-19 15:53:06,855 Stage-2 map = 0%,  reduce = 0%
2023-10-19 15:53:10,959 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.95 sec
2023-10-19 15:53:15,048 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.17 sec
MapReduce Total cumulative CPU time: 2 seconds 170 msec
Ended Job = job_1697720995088_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 4.07 sec   HDFS Read: 235138905 HDFS Write: 150 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.17 sec   HDFS Read: 7740 HDFS Write: 114 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 240 msec
OK
Female  1000505
Time taken: 36.269 seconds, Fetched: 1 row(s)
hive> 
```

En C3 hive

SELECT cargo, COUNT (*) AS total FROM people_data GROUP BY cargo ORDER BY total DESC;

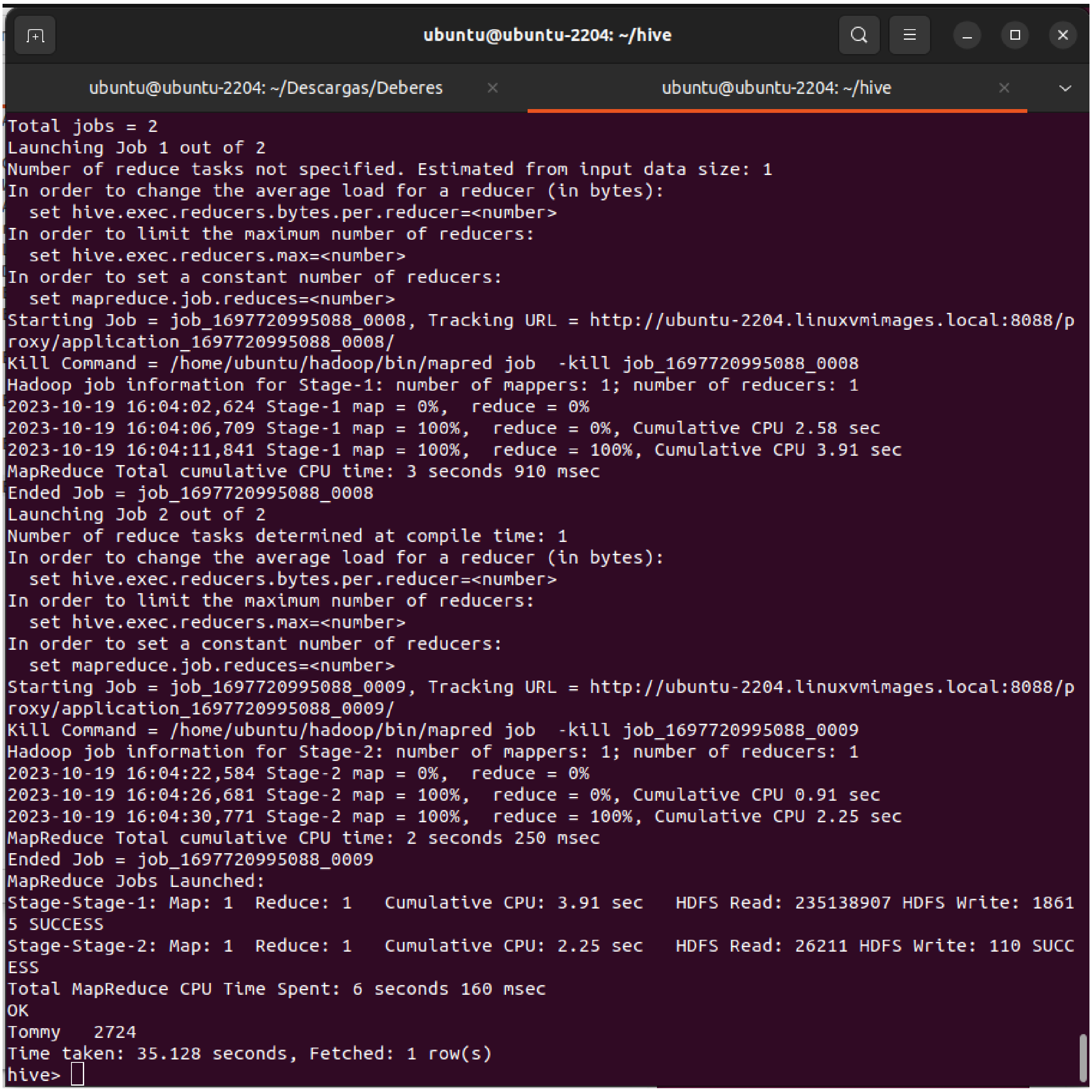

ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/hive

```
Immunologist    3055
Lawyer  3053
Purchasing manager      3053
Medical secretary       3052
Dance movement psychotherapist  3052
Exhibition designer     3051
Hospital doctor 3051
Artist  3050
Contracting civil engineer      3050
Museum/gallery curator  3049
Marine scientist        3049
Cabin crew      3048
Scientific laboratory technician        3048
Chief Marketing Officer 3047
Counselling psychologist        3046
Secondary school teacher        3045
Race relations officer  3042
Futures trader  3041
Chief Technology Officer        3039
Proofreader     3037
Air traffic controller  3034
Ophthalmologist 3032
"Programme researcher   3032
Chief Operating Officer 3031
Hydrographic surveyor   3031
"Sound technician       3030
Medical laboratory scientific officer   3028
Medical technical officer       3026
Electronics engineer    3025
Radiation protection practitioner       3023
Production manager      3023
Television camera operator      3022
Insurance underwriter   3022
Public relations account executive      3019
Personal assistant      3019
Musician        3018
Ergonomist      3010
Research scientist (medical)    2999
Tax inspector   2998
Mining engineer 2996
Glass blower/designer   2986
Higher education lecturer       2981
Wellsite geologist      2943
Time taken: 36.11 seconds, Fetched: 524 row(s)
hive> 
```

En C4 hive

-- Indica el nombre menos repetido

SELECT nombre, COUNT (*) AS total FROM people_data GROUP BY nombre ORDER BY total ASC LIMIT 1;


ubuntu@ubuntu-2204: ~/hive
ubuntu@ubuntu-2204: ~/Descargas/Deberes
ubuntu@ubuntu-2204: ~/hive

```
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1697720995088_0008, Tracking URL = http://ubuntu-2204.linuxvmimages.local:8088/p
roxy/application_1697720995088_0008/
Kill Command = /home/ubuntu/hadoop/bin/mapred job  -kill job_1697720995088_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-10-19 16:04:02,624 Stage-1 map = 0%,  reduce = 0%
2023-10-19 16:04:06,709 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.58 sec
2023-10-19 16:04:11,841 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.91 sec
MapReduce Total cumulative CPU time: 3 seconds 910 msec
Ended Job = job_1697720995088_0008
Launching Job 2 out of 2
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1697720995088_0009, Tracking URL = http://ubuntu-2204.linuxvmimages.local:8088/p
roxy/application_1697720995088_0009/
Kill Command = /home/ubuntu/hadoop/bin/mapred job  -kill job_1697720995088_0009
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-10-19 16:04:22,584 Stage-2 map = 0%,  reduce = 0%
2023-10-19 16:04:26,681 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 0.91 sec
2023-10-19 16:04:30,771 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 2.25 sec
MapReduce Total cumulative CPU time: 2 seconds 250 msec
Ended Job = job_1697720995088_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.91 sec   HDFS Read: 235138907 HDFS Write: 1861
5 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 2.25 sec   HDFS Read: 26211 HDFS Write: 110 SUCC
ESS
Total MapReduce CPU Time Spent: 6 seconds 160 msec
OK
Tommy   2724
Time taken: 35.128 seconds, Fetched: 1 row(s)
hive> 
```