

OLYMPIC DATA ANALYSIS PROJECT

REPORT

In this project , I first loaded Olympic dataset on cloudera machine by creating shared folder both on windows and linux OS.

Then I create directory to load the csv files into it using hadoop and log in to hive shell. In hive I created table of data from csv file ,run queries and do analysis of data.

1) Creating Directory and Loading csv files into it

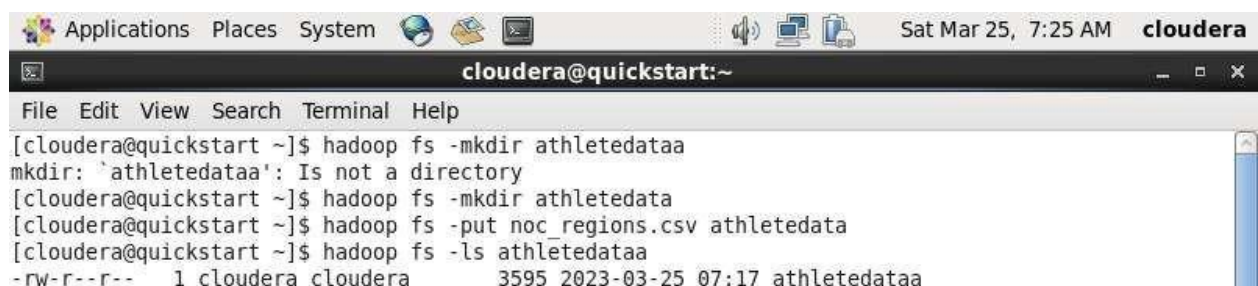
Command :

```
$ hadoop fs -mkdir athletedata
```

```
$ hadoop fs -put noc_regions.csv athletedata
```

```
$ hadoop fs -put athlete_events.csv athletedata
```

```
$ hadoop fs -ls athletedata
```



```
Applications Places System Sat Mar 25, 7:25 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -mkdir athletedataa
mkdir: 'athletedataa': Is not a directory
[cloudera@quickstart ~]$ hadoop fs -mkdir athletedata
[cloudera@quickstart ~]$ hadoop fs -put noc_regions.csv athletedata
[cloudera@quickstart ~]$ hadoop fs -ls athletedataa
-rw-r--r-- 1 cloudera cloudera 3595 2023-03-25 07:17 athletedataa
```

```
Applications Places System Mon Mar 27, 1:20 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -mkdir athletedata
[cloudera@quickstart ~]$ hadoop fs -put athlete_events.csv athletedata
[cloudera@quickstart ~]$ hadoop fs -ls athletedataa
ls: 'athletedataa': No such file or directory
[cloudera@quickstart ~]$ hadoop fs -ls athletedata
Found 1 items
-rw-r--r-- 1 cloudera cloudera 41500688 2023-03-27 01:03 athletedata/athlete
events.csv
```

2) Creating Tables

```
Applications Places System Mon Mar 27, 1:21 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
FAILED: ParseException line 1:180 cannot recognize input near 'tring' ',' 'medal
' in column type
hive> create table ath(id int, name string, sex string, age int, height int, wei
ght int, team string, NOC string, games string, year int, season string, city st
ring, sport string, event string, medal string) row format delimited fields term
inated by ',' stored as textfile location '/user/cloudera/athletedata/';
OK
Time taken: 0.93 seconds
hive> select * from athlete limit 5;
OK
Time taken: 0.772 seconds
hive> select * from ath limit 5;
OK
NULL "Name" "Sex" NULL NULL NULL "Team" "NOC" "Games" NULL "
Season" "City" "Sport" "Event" "Medal"
NULL "A Dijiang" "M" 24 180 80 "China" "CHN" "1992 Su
mmer" 1992 "Summer" "Barcelona" "Basketball" "Basketball Men'
s Basketball" NA
NULL "A Lamusi" "M" 23 170 60 "China" "CHN" "2012 Su
mmer" 2012 "Summer" "London" "Judo" "Judo Men's Extra-Lightw
eight" NA
NULL "Gunnar Nielsen Aaby" "M" 24 NULL NULL "Denmark" "
DEN" "1920 Summer" 1920 "Summer" "Antwerpen" "Football" "
Football Men's Football" NA
NULL "Edgar Lindenau Aabye" "M" 34 NULL NULL "Denmark/Sweden"
"DEN" "1900 Summer" 1900 "Summer" "Paris" "Tug-Of-War" "Tug-Of-
War Men's Tug-Of-War" "Gold"
Time taken: 0.104 seconds, Fetched: 5 row(s)
hive>
```

```
Applications Places System Sat Mar 25, 7:25 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hadoop fs -mkdir athletedataa
mkdir: 'athletedataa': Is not a directory
[cloudera@quickstart ~]$ hadoop fs -mkdir athletedata
[cloudera@quickstart ~]$ hadoop fs -put noc_regions.csv athletedata
[cloudera@quickstart ~]$ hadoop fs -ls athletedataa
-rw-r--r-- 1 cloudera cloudera 3595 2023-03-25 07:17 athletedataa
[cloudera@quickstart ~]$ hive

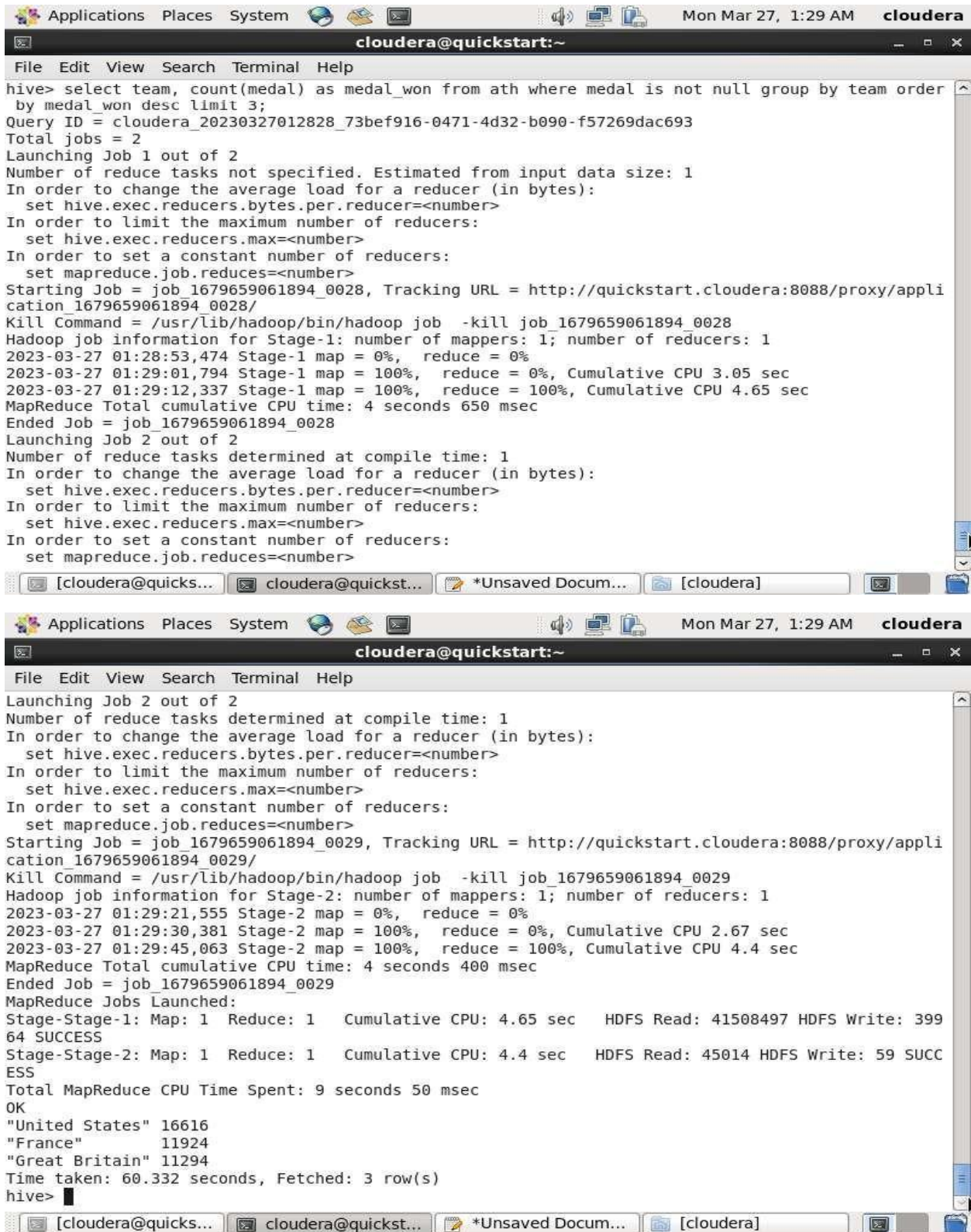
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.p
roperties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> use trial;
OK
Time taken: 0.28 seconds
hive> create table country(noc string, region string, notes string) row format d
eliminated fields terminated by ',' stored as textfile location '/user/cloudera/at
hletedata/';
OK
Time taken: 0.233 seconds
hive> select * from country limit 2;
OK
NOC      region notes
AFG      Afghanistan
Time taken: 0.758 seconds, Fetched: 2 row(s)
hive> select count(*) from country;
Query ID = cloudera_20230325072424_14492332-1dcc-4e51-a8ca-dfa29e274c02
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
```

3) Queries

Counting number of countries participated in Olympics

```
Applications Places System Sat Mar 25, 7:25 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
Time taken: 0.758 seconds, Fetched: 2 row(s)
hive> select count(*) from country;
Query ID = cloudera_20230325072424_14492332-1dcc-4e51-a8ca-dfa29e274c02
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_1679659061894_0007, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1679659061894_0007/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-25 07:24:20,579 Stage-1 map = 0%, reduce = 0%
2023-03-25 07:24:50,025 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.55 sec
2023-03-25 07:25:04,021 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.16 sec
MapReduce Total cumulative CPU time: 3 seconds 160 msec
Ended Job = job_1679659061894_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.16 sec HDFS Read: 10967 HDFS Write: 4 SUCC
ESS
Total MapReduce CPU Time Spent: 3 seconds 160 msec
OK
231
Time taken: 63.47 seconds, Fetched: 1 row(s)
hive>
```


Finding top 3 countries in terms of total number of medals



The image displays two screenshots of a terminal window on a Cloudera system, showing the execution of Hive queries and Hadoop jobs to find the top 3 countries by total number of medals.

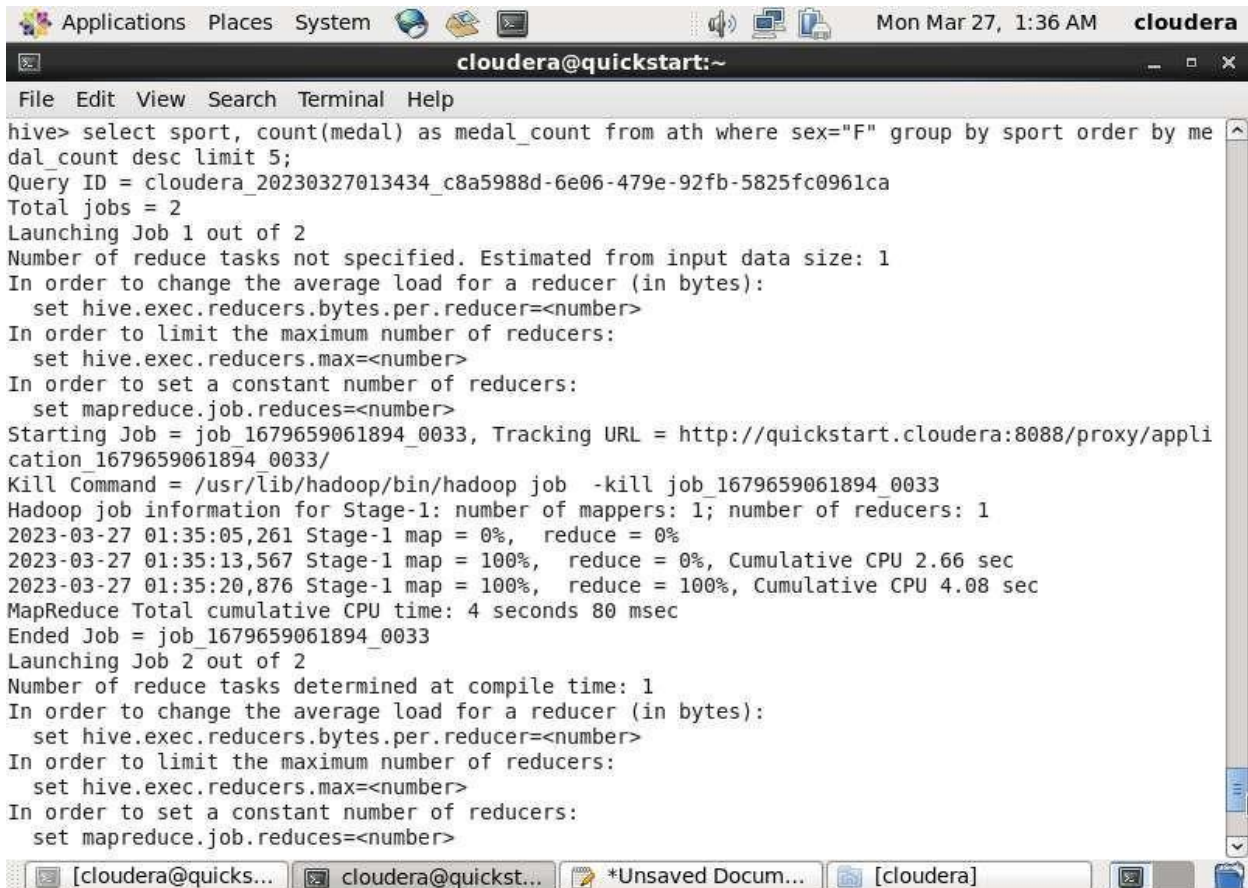
Top Screenshot:

```
File Edit View Search Terminal Help
hive> select team, count(medal) as medal_won from ath where medal is not null group by team order
  by medal_won desc limit 3;
Query ID = cloudera_20230327012828_73bef916-0471-4d32-b090-f57269dac693
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_1679659061894_0028, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation 1679659061894_0028/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0028
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:28:53,474 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:29:01,794 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.05 sec
2023-03-27 01:29:12,337 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.65 sec
MapReduce Total cumulative CPU time: 4 seconds 650 msec
Ended Job = job_1679659061894_0028
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>
```

Bottom Screenshot:

```
File Edit View Search Terminal Help
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_1679659061894_0029, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation 1679659061894_0029/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0029
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-03-27 01:29:21,555 Stage-2 map = 0%, reduce = 0%
2023-03-27 01:29:30,381 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.67 sec
2023-03-27 01:29:45,063 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 4.4 sec
MapReduce Total cumulative CPU time: 4 seconds 400 msec
Ended Job = job_1679659061894_0029
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.65 sec HDFS Read: 41508497 HDFS Write: 399
64 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 4.4 sec HDFS Read: 45014 HDFS Write: 59 SUCC
ESS
Total MapReduce CPU Time Spent: 9 seconds 50 msec
OK
"United States" 16616
"France" 11924
"Great Britain" 11294
Time taken: 60.332 seconds, Fetched: 3 row(s)
hive>
```

Count of female athletes in olympics



```
Applications Places System Mon Mar 27, 1:36 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select sport, count(medal) as medal_count from ath where sex="F" group by sport order by me
dal count desc limit 5;
Query ID = cloudera_20230327013434_c8a5988d-6e06-479e-92fb-5825fc0961ca
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_1679659061894_0033, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0033/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0033
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:35:05,261 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:35:13,567 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.66 sec
2023-03-27 01:35:20,876 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.08 sec
MapReduce Total cumulative CPU time: 4 seconds 80 msec
Ended Job = job_1679659061894_0033
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>
```

```
"F"      23.74107240661628
"M"      26.27707921005077
"Sex"    NULL
-Riddell)"  NULL
Time taken: 27.259 seconds, Fetched: 455 row(s)
hive> █
```

Top 5 sports in which athlete have won medals

```
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select sport, count(medal) as medal_count from ath group by sport order by medal_count des
c limit 5;
Query ID = cloudera_20230327013737_82432728-14ea-4be3-a38b-8a468b9a165d
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_1679659061894_0035, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0035/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0035
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:37:35,639 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:37:44,234 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.16 sec
2023-03-27 01:37:51,607 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.66 sec
MapReduce Total cumulative CPU time: 3 seconds 660 msec
Ended Job = job_1679659061894_0035
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>
```

```
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
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_1679659061894_0036, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0036/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0036
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2023-03-27 01:38:00,310 Stage-2 map = 0%, reduce = 0%
2023-03-27 01:38:07,841 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 1.44 sec
2023-03-27 01:38:20,410 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.79 sec
MapReduce Total cumulative CPU time: 2 seconds 790 msec
Ended Job = job_1679659061894_0036
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.66 sec HDFS Read: 41508769 HDFS Write: 365
2 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 2.79 sec HDFS Read: 8709 HDFS Write: 87 SUCC
ESS
Total MapReduce CPU Time Spent: 6 seconds 450 msec
OK
"Athletics"      38105
"Gymnastics"     26339
"Swimming"       22818
"Shooting"       11316
"Cycling"        10819
Time taken: 55.345 seconds, Fetched: 5 row(s)
```


No of events held for each sport in year after 1980

```
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select year, sport, count(*) from ath where sport!='None' and year>1980 group by year, spor
t;
Query ID = cloudera_20230327014444_7ca086c6-88b7-4f8d-8870-94d2b31c5cf3
Total jobs = 1
Launching Job 1 out of 1
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_1679659061894_0038, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0038/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0038
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:44:55,321 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:45:06,012 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.22 sec
2023-03-27 01:45:19,567 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.42 sec
MapReduce Total cumulative CPU time: 6 seconds 420 msec
Ended Job = job_1679659061894_0038
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.42 sec HDFS Read: 41510153 HDFS Write: 888
8 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 420 msec
OK
1984 "Alpine Skiing" 397
1984 "Archery" 109
1984 "Athletics" 1650
```

```
Applications Places System cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
1984 "Judo" 211
1984 "Luge" 88
1984 "Modern Pentathlon" 101
1984 "Nordic Combined" 28
1984 "Rhythmic Gymnastics" 33
1984 "Rowing" 461
1984 "Sailing" 292
1984 "Shooting" 530
1984 "Ski Jumping" 111
1984 "Speed Skating" 324
1984 "Swimming" 1221
1984 "Synchronized Swimming" 84
1984 "Volleyball" 208
1984 "Water Polo" 152
1984 "Weightlifting" 185
1984 "Wrestling" 286
1988 "Alpine Skiing" 660
1988 "Archery" 251
1988 "Athletics" 2018
1988 "Basketball" 230
1988 "Biathlon" 207
1988 "Bobsleigh" 184
1988 "Boxing" 431
1988 "Canoeing" 386
1988 "Cross Country Skiing" 526
1988 "Cycling" 518
1988 "Diving" 108
1988 "Equestrianism" 318
1988 "Fencing" 505
```

Applications Places System Mon Mar 27, 1:46 AM cloudera

cloudera@quickstart:~

File	Edit	View	Search	Terminal	Help
2004	"Volleyball"	283			
2004	"Water Polo"	255			
2004	"Weightlifting"	248			
2004	"Wrestling"	341			
2006	"Alpine Skiing"	619			
2006	"Biathlon"	651			
2006	"Bobsleigh"	190			
2006	"Cross Country Skiing"	812			
2006	"Curling"	91			
2006	"Figure Skating"	146			
2006	"Freestyle Skiing"	116			
2006	"Ice Hockey"	441			
2006	"Luge"	107			
2006	"Nordic Combined"	138			
2006	"Short Track Speed Skating"	238			
2006	"Skeleton"	42			
2006	"Ski Jumping"	202			
2006	"Snowboarding"	198			
2006	"Speed Skating"	379			
2008	"Archery"	194			
2008	"Athletics"	2234			
2008	"Badminton"	184			
2008	"Baseball"	191			
2008	"Basketball"	284			
2008	"Beach Volleyball"	96			
2008	"Boxing"	283			
2008	"Canoeing"	435			
2008	"Cycling"	651			
2008	"Diving"	182			

[cloudera@quicks... cloudera@quickst... *Unsaved Docum... [cloudera]

Applications Places System Mon Mar 27, 1:46 AM cloudera

cloudera@quickstart:~

File	Edit	View	Search	Terminal	Help
2016	"Cycling"	666			
2016	"Diving"	178			
2016	"Equestrianism"	355			
2016	"Fencing"	346			
2016	"Football"	473			
2016	"Golf"	119			
2016	"Gymnastics"	861			
2016	"Handball"	353			
2016	"Hockey"	390			
2016	"Judo"	389			
2016	"Modern Pentathlon"	72			
2016	"Rhythmic Gymnastics"	96			
2016	"Rowing"	549			
2016	"Rugby Sevens"	299			
2016	"Sailing"	380			
2016	"Shooting"	552			
2016	"Swimming"	1559			
2016	"Synchronized Swimming"	118			
2016	"Table Tennis"	236			
2016	"Taekwondo"	126			
2016	"Tennis"	286			
2016	"Trampolining"	32			
2016	"Triathlon"	110			
2016	"Volleyball"	283			
2016	"Water Polo"	258			
2016	"Weightlifting"	255			
2016	"Wrestling"	346			

Time taken: 34.516 seconds, Fetched: 398 row(s)
hive>

[cloudera@quicks... cloudera@quickst... *Unsaved Docum... [cloudera]

3 Items in Trash

Count of athletes in Olympics in year



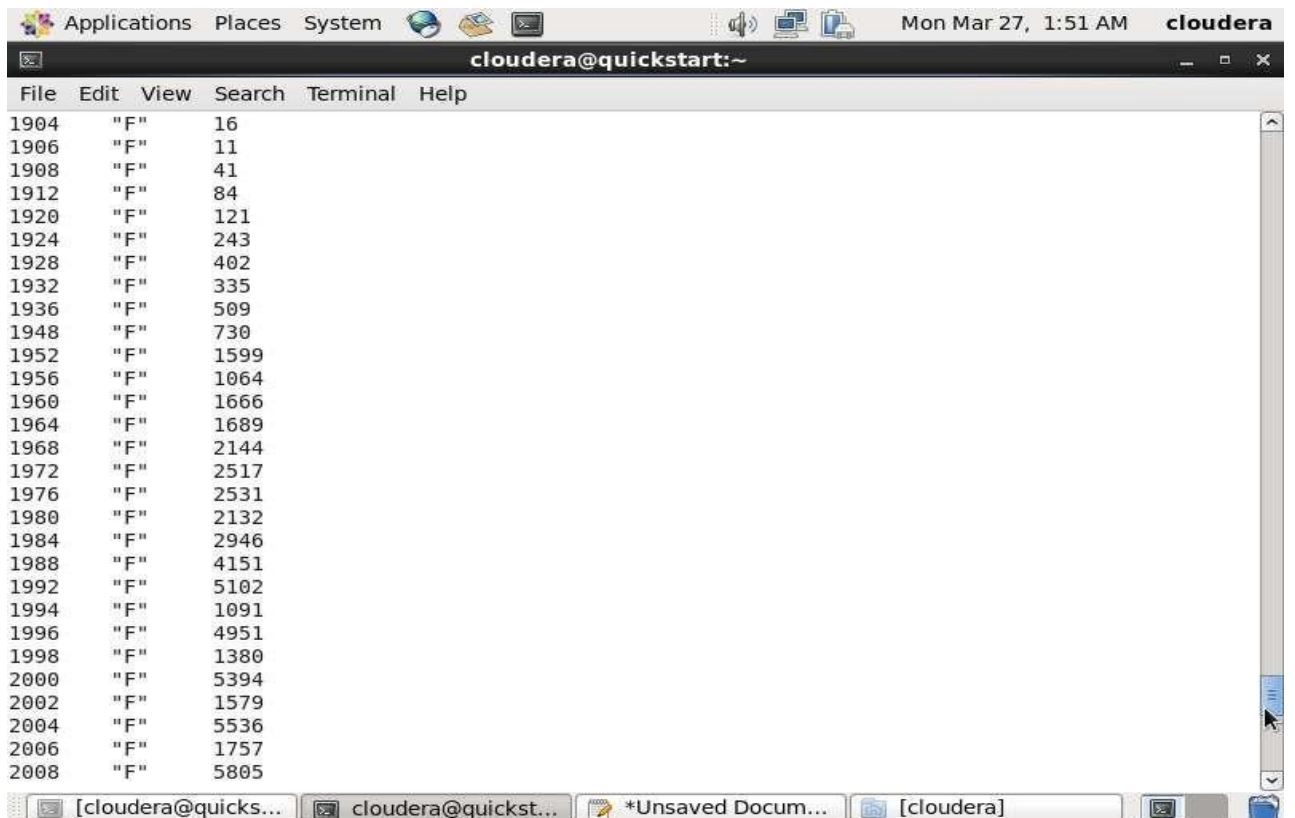
Applications Places System Mon Mar 27, 1:52 AM cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

```
2000 "M" 8357
2002 "M" 2520
2004 "M" 7860
2006 "M" 2613
2008 "M" 7754
2010 "M" 2540
2012 "M" 7071
2014 "M" 2858
2016 "M" 7445
NULL "Sex" 1
NULL "-Riddell)" 1
Time taken: 36.201 seconds, Fetched: 523 row(s)
hive> select year, sex, count(sex) as count from ath group by sex, year;
Query ID = cloudera_20230327015252_a4087334-6f9f-4fe5-b74e-f66666b87b73
Total jobs = 1
Launching Job 1 out of 1
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_1679659061894_0040, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0040/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0040
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:52:17,269 Stage-1 map = 0%, reduce = 0%
```

[cloudera@quicks... cloudera@quickst... *Unsaved Docum... [cloudera]



Applications Places System Mon Mar 27, 1:51 AM cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

```
1904 "F" 16
1906 "F" 11
1908 "F" 41
1912 "F" 84
1920 "F" 121
1924 "F" 243
1928 "F" 402
1932 "F" 335
1936 "F" 509
1948 "F" 730
1952 "F" 1599
1956 "F" 1064
1960 "F" 1666
1964 "F" 1689
1968 "F" 2144
1972 "F" 2517
1976 "F" 2531
1980 "F" 2132
1984 "F" 2946
1988 "F" 4151
1992 "F" 5102
1994 "F" 1091
1996 "F" 4951
1998 "F" 1380
2000 "F" 5394
2002 "F" 1579
2004 "F" 5536
2006 "F" 1757
2008 "F" 5805
```

[cloudera@quicks... cloudera@quickst... *Unsaved Docum... [cloudera]

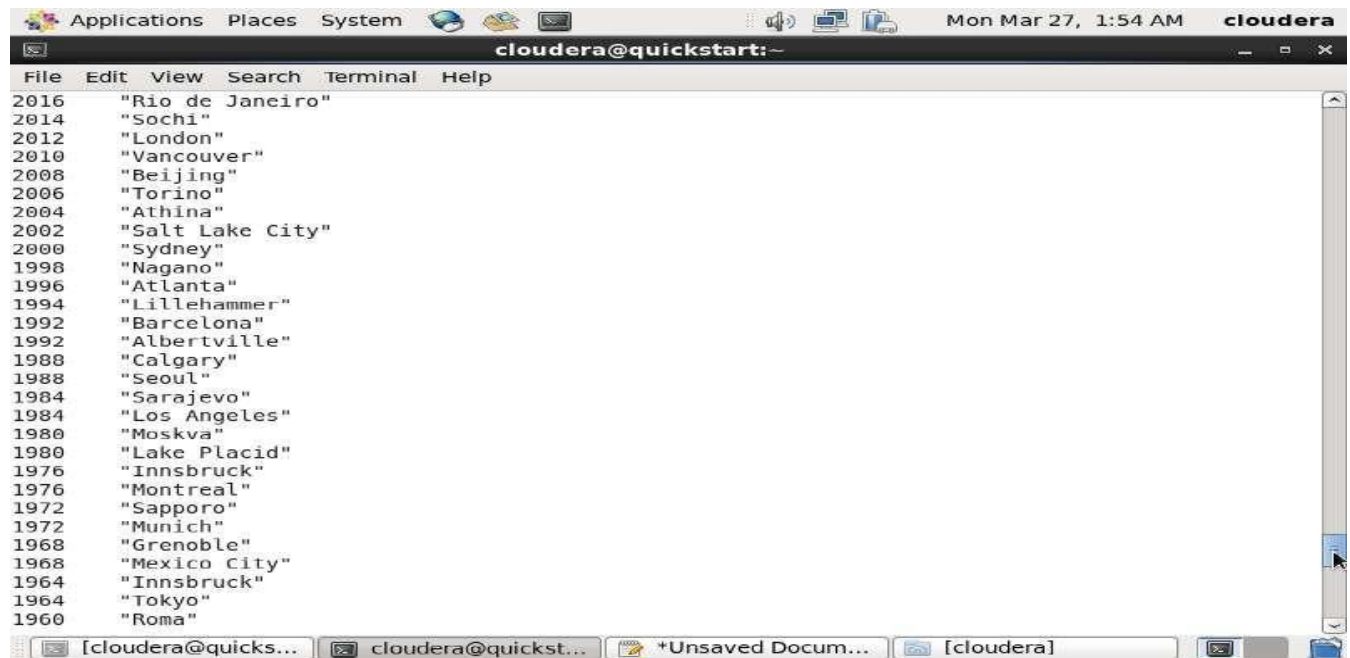
A screenshot of a Linux terminal window. The title bar at the top shows 'Applications Places System' on the left, a clock and date 'Mon Mar 27, 1:52 AM' in the center, and 'cloudera' on the right. Below the title bar, the terminal prompt is 'cloudera@quickstart:~'. The terminal content displays a list of years and values, organized into three columns: Year, a string in quotes, and a number. The years range from 1900 to 2006 in increments of 4, with the last entry being 2006. The values are: 1857, 1150, 1691, 3013, 3896, 4104, 5380, 5050, 2894, 6788, 6620, 7571, 5215, 7436, 7662, 8203, 9245, 7830, 6725, 8545, 10387, 11173, 2047, 8735, 2209, 8357, 2520, 7860, and 2613. The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The bottom of the image shows a taskbar with several open applications, including a terminal window with the prompt 'cloudera@quickst...', a document titled '*Unsaved Docum...', and a window titled '[cloudera]'. The system tray on the right shows icons for network, volume, and power.

Different cities in which Olympics has held

```

Applications Places System Mon Mar 27, 1:53 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select year, city from ath group by city, year order by year desc;
Query ID = cloudera_20230327015353_5b730973-aefe-4d81-af2f-84d0345a4b22
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_1679659061894_0041, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0041/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0041
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:53:21,262 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:53:28,588 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.42 sec
2023-03-27 01:53:35,903 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.04 sec
MapReduce Total cumulative CPU time: 4 seconds 40 msec
Ended Job = job_1679659061894_0041
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_1679659061894_0042, Tracking URL = http://quickstart.cloudera:8088/proxy/appli

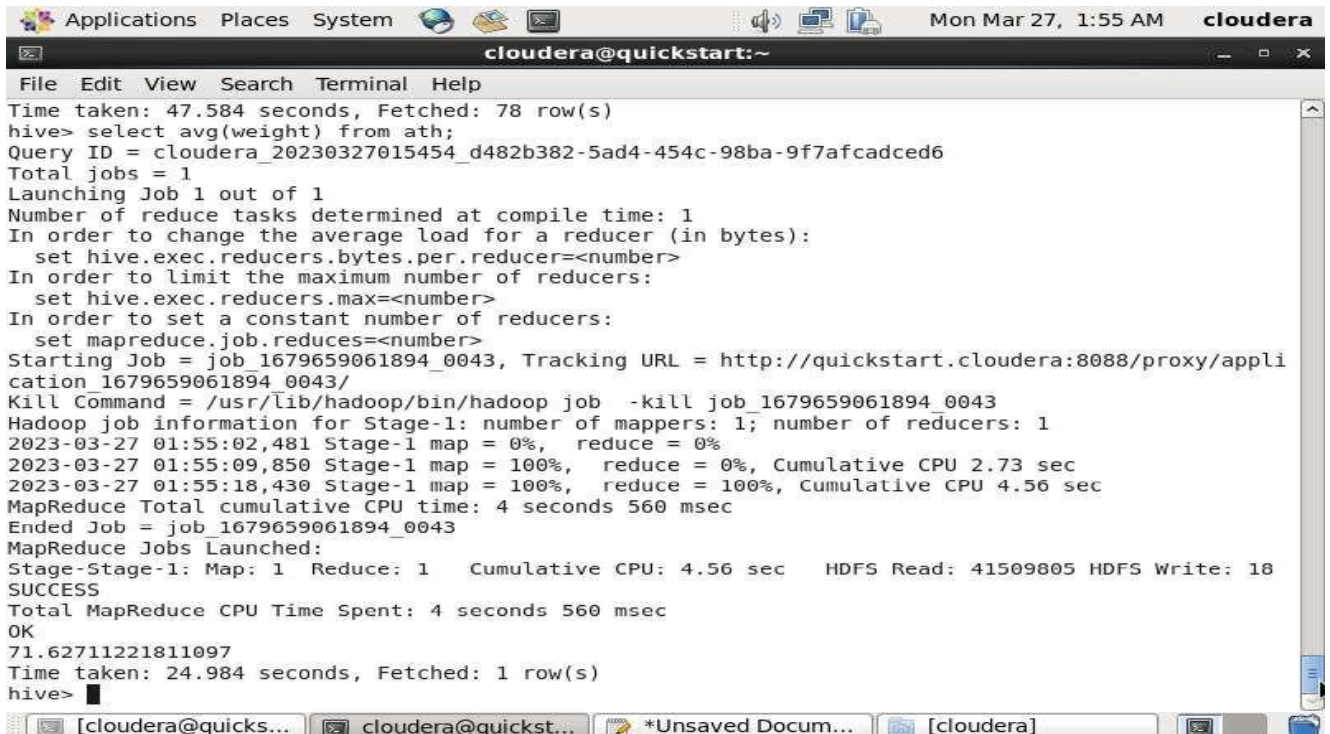
```



The screenshot shows a terminal window titled "cloudera@quickstart:~". The window contains a list of Olympic cities and the year they were hosted. The list is as follows:

Year	City
2016	"Rio de Janeiro"
2014	"Sochi"
2012	"London"
2010	"Vancouver"
2008	"Beijing"
2006	"Torino"
2004	"Athina"
2002	"Salt Lake City"
2000	"Sydney"
1998	"Nagano"
1996	"Atlanta"
1994	"Lillehammer"
1992	"Barcelona"
1992	"Albertville"
1988	"Calgary"
1988	"Seoul"
1984	"Sarajevo"
1984	"Los Angeles"
1980	"Moskva"
1980	"Lake Placid"
1976	"Innsbruck"
1976	"Montreal"
1972	"Sapporo"
1972	"Munich"
1968	"Grenoble"
1968	"Mexico City"
1964	"Innsbruck"
1964	"Tokyo"
1960	"Roma"

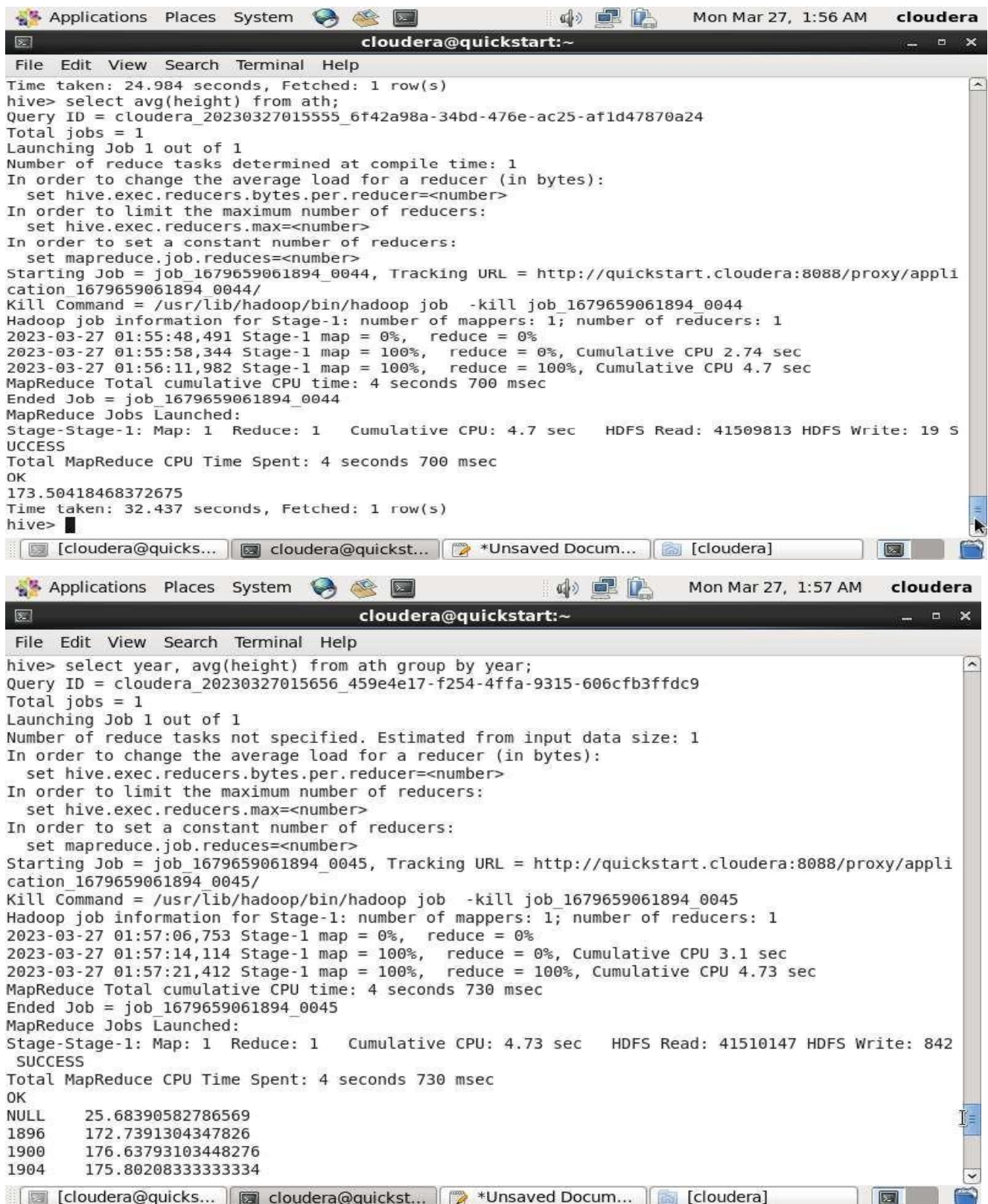
Average weights of all athletes



The screenshot shows a terminal window titled "cloudera@quickstart:~". The window contains the output of a Hive query and the execution details of a MapReduce job. The output is as follows:

```
Time taken: 47.584 seconds, Fetched: 78 row(s)
hive> select avg(weight) from ath;
Query ID = cloudera_20230327015454_d482b382-5ad4-454c-98ba-9f7afcadcde6
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_1679659061894_0043, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1679659061894_0043/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0043
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:55:02,481 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:55:09,850 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.73 sec
2023-03-27 01:55:18,430 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.56 sec
MapReduce Total cumulative CPU time: 4 seconds 560 msec
Ended Job = job_1679659061894_0043
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.56 sec HDFS Read: 41509805 HDFS Write: 18
SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 560 msec
OK
71.62711221811097
Time taken: 24.984 seconds, Fetched: 1 row(s)
hive>
```


Average heights of all athletes



The image shows two screenshots of a Cloudera terminal window. The top screenshot shows a Hive query to calculate the average height of athletes. The bottom screenshot shows a Hive query to calculate the average height of athletes grouped by year.

Top Screenshot:

```
Time taken: 24.984 seconds, Fetched: 1 row(s)
hive> select avg(height) from ath;
Query ID = cloudera_20230327015555_6f42a98a-34bd-476e-ac25-af1d47870a24
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_1679659061894_0044, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0044/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0044
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:55:48,491 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:55:58,344 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.74 sec
2023-03-27 01:56:11,982 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.7 sec
MapReduce Total cumulative CPU time: 4 seconds 700 msec
Ended Job = job_1679659061894_0044
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.7 sec HDFS Read: 41509813 HDFS Write: 19 S
UCCESS
Total MapReduce CPU Time Spent: 4 seconds 700 msec
OK
173.50418468372675
Time taken: 32.437 seconds, Fetched: 1 row(s)
hive>
```

Bottom Screenshot:

```
hive> select year, avg(height) from ath group by year;
Query ID = cloudera_20230327015656_459e4e17-f254-4ffa-9315-606cfb3ffdc9
Total jobs = 1
Launching Job 1 out of 1
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_1679659061894_0045, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0045/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0045
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 01:57:06,753 Stage-1 map = 0%, reduce = 0%
2023-03-27 01:57:14,114 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.1 sec
2023-03-27 01:57:21,412 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.73 sec
MapReduce Total cumulative CPU time: 4 seconds 730 msec
Ended Job = job_1679659061894_0045
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.73 sec HDFS Read: 41510147 HDFS Write: 842
SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 730 msec
OK
NULL      25.68390582786569
1896      172.7391304347826
1900      176.63793103448276
1904      175.80208333333334
```

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
1928 175.2515923566879
1932 174.28125
1936 175.76962457337885
1948 176.19823008849556
1952 174.07724719101122
1956 173.83745173745174
1960 173.14364573110893
1964 173.46286900369003
1968 173.93418342324668
1972 174.5901238876287
1976 174.9340288755401
1980 175.55449492508458
1984 175.52413729809103
1988 175.7497795414462
1992 176.1978936121499
1994 175.18970189701898
1996 175.90445590597784
1998 174.5700427960057
2000 176.08137657763427
2002 174.6995085995086
2004 175.9450598802395
2006 174.60013748854263
2008 176.18189275859498
2010 174.89459274469542
2012 176.23023676551563
2014 174.80189261468834
2016 176.01994217510565
Time taken: 25.839 seconds, Fetched: 36 row(s)
hive>
```

Number of distinct teams participated in every olympics

```
Applications Places System cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select year, count(distinct team) from ath group by year;
Query ID = cloudera_20230327053838_7733e87b-1f14-46d1-a634-47ee70c7c1e7
Total jobs = 1
Launching Job 1 out of 1
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 1679659061894 0051, Tracking URL = http://quickstart.cloudera:8088/proxy/appli
cation_1679659061894_0051/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1679659061894_0051
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2023-03-27 05:38:36,681 Stage-1 map = 0%, reduce = 0%
2023-03-27 05:38:47,342 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.36 sec
2023-03-27 05:38:54,704 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.7 sec
MapReduce Total cumulative CPU time: 5 seconds 700 msec
Ended Job = job 1679659061894_0051
MapReduce Jobs Launched:
Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.7 sec HDFS Read: 41509892 HDFS Write: 314
SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 700 msec
OK
NULL 115
1896 18
1900 185
1904 66
```

