Program 6: Construct the Pig Latin Scripts to find a max temp for each and every year.

Login into Hadoop user you used while installing Hadoop , here we use hdoop user

```
hdoop@NuvobookV1:~$
```

$ su – hdoop

-start the Hadoop server

$ cd hadoop-3.4.0/sbin/

$ ./start-dfs.sh

$ ./start-yarn

$ jps

```
hdoop@hdoop-VirtualBox:~$ cd hadoop-3.2.1/sbin
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [hdoop-VirtualBox]
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$ jps
4817 DataNode
5298 ResourceManager
5000 SecondaryNameNode
5450 NodeManager
4683 NameNode
5982 Jps
hdoop@hdoop-VirtualBox:~/hadoop-3.2.1/sbin$
```

$ sudo mkdir p6/

$ cd p5/

#create the dataset:

$sudo nano temperature.txt

-Insert this dataset:

1938 5.0

1938 13.3

1941 33.9

1974 8.9

1974 8.9

1983 31.7

1983 30.6

1991 13.3

1993 32.8

1950 7.2

1951 8.3

1955 10.0

1960 19.4

1961 22.2

1965 17.2

1966 23.9

1970 25.6

1972 27.8

1975 27.2

1976 22.8

1980 18.3

1981 31.1

1984 28.3

1985 28.9

1990 21.1

1992 29.4

1994 32.2

1995 20.0

1996 15.6

2000 18.9

2001 26.7

2002 25.0

2003 26.1

2005 30.0

2006 23.9

2008 27.8

2010 27.2

2012 20.6

2013 31.1

2014 33.3

2015 30.6

2016 33.9

2017 22.2

2018 32.8

2019 31.7

2020 34.4

2021 31.1

2022 33.3

2023 35.0

2024 29.4

```
hdoop@NuvobookV1:~$ cd lab/p6/
hdoop@NuvobookV1:~/lab/p6$ cat temperature.txt
1938 5.0
1938 13.3
1941 33.9
1974 8.9
1974 8.9
1983 31.7
1983 30.6
1991 13.3
1993 32.8
1950 7.2
1951 8.3
1955 10.0
1960 19.4
1961 22.2
1965 17.2
1966 23.9
1970 25.6
1972 27.8
1975 27.2
1976 22.8
1980 18.3
1981 31.1
1984 28.3
1985 28.9
1990 21.1
1992 29.4
1994 32.2
1995 20.0
1996 15.6
2000 18.9
2001 26.7
2002 25.0
2003 26.1
2005 30.0
2006 23.9
2008 27.8
2010 27.2
```

#upload the file to Hadoop:

$ hdfs dfs -mkdir /p6/

$hadoop fs -put temperature.txt /p6/

```
hdoop@NuvobookV1:~/lab/p6$ hadoop fs -ls /p6
Found 1 items
-rw-r--r--   1 hdoop supergroup        495 2024-10-18 03:01 /p6/temper
ature.txt
hdoop@NuvobookV1:~/lab/p6$
```

#run pig :

$ pig

```
hdoop@NuvobookV1:~/lab/p5$ pig -x mapreduce
2024-10-16 19:59:03,530 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
2024-10-16 19:59:03,532 INFO pig.ExecTypeProvider: Trying ExecType : MAPREDU
CE
2024-10-16 19:59:03,532 INFO pig.ExecTypeProvider: Picked MAPREDUCE as the E
xecType
2024-10-16 19:59:03,588 WARN pig.Main: Cannot write to log file: /home/hdoop
/lab/p5/pig_1729108743588.log
2024-10-16 19:59:03,599 [main] INFO  org.apache.pig.Main - Apache Pig versio
n 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2024-10-16 19:59:03,635 [main] INFO  org.apache.pig.impl.util.Utils - Defaul
t bootup file /home/hdoop/.pigbootup not found
2024-10-16 19:59:03,918 [main] INFO  org.apache.hadoop.conf.Configuration.de
precation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtrac
ker.address
2024-10-16 19:59:03,918 [main] INFO  org.apache.pig.backend.hadoop.execution
engine.HExecutionEngine - Connecting to hadoop file system at: hdfs://172.18
.132.51:9000
2024-10-16 19:59:04,458 [main] INFO  org.apache.pig.PigServer - Pig Script I
D for the session: PIG-default-7825e024-6755-4975-aad7-14900747aff5
2024-10-16 19:59:04,458 [main] WARN  org.apache.pig.PigServer - ATS is disab
led since yarn.timeline-service.enabled set to false
grunt>
```

#write this query in pig:

-load the file temperature that we upload to Hadoop:

data = LOAD '/p6/temperature.txt' USING PigStorage(' ') AS (year:int, temp:int);

```
grunt> data = LOAD '/p6/temperature.txt' USING PigStorage(' ') AS (year:int,
 temp:int);
2024-10-18 03:15:00,345 [main] INFO  org.apache.hadoop.conf.Configuration.de
precation - yarn.resourcemanager.system-metrics-publisher.enabled is depreca
ted. Instead, use yarn.system-metrics-publisher.enabled
grunt>
```

- Group the data by year:

grouped_data = GROUP data BY year;

- Find the maximum temperature for each year

max_temp_by_year = FOREACH grouped_data GENERATE group AS year, MAX(data.temp) AS max_temp;

- Store the result in an output file

STORE max_temp_by_year INTO '/p6/max_temperature_per_year' USING PigStorage(',');

```
grunt> data = LOAD '/p6/temperature.txt' USING PigStorage(' ') AS (year:int,
 temp:int);
2024-10-18 03:15:00,345 [main] INFO  org.apache.hadoop.conf.Configuration.de
precation - yarn.resourcemanager.system-metrics-publisher.enabled is depreca
ted. Instead, use yarn.system-metrics-publisher.enabled
grunt> grouped_data = GROUP data BY year;
grunt> max_temp_by_year = FOREACH grouped_data GENERATE group AS year, MAX(d
ata.temp) AS max_temp;
grunt> STORE max_temp_by_year INTO '/p6/max_temperature_per_year' USING PigS
torage(',');
```

Note: it will take approx. 10 minutes to compile.

```
engine.mapReduceLayer.MapReduceLauncher - Success!
grunt>
```

Output:

-in new terminal log in to hdoop again and chechk the output by writing this command line (it print out the content of your out put file that is inside Hadoop file system)

$hadoop fs -cat /p6/max_temperature_per_year/part-r-00000

```
hdoop@NuvobookV1:~/lab/p6$ hadoop fs -cat /p6/max_temperature_per_year
/part-r-00000
1938,13
1941,33
1950,7
1951,8
1955,10
1960,19
1961,22
1965,17
1966,23
1970,25
1972,27
1974,8
1975,27
1976,22
1980,18
1981,31
1983,31
1984,28
1985,28
1990,21
1991,13
1992,29
1993,32
1994,32
1995,20
1996,15
2000,18
2001,26
2002,25
2003,26
2005,30
2006,23
2008,27
2010,27
```

Note:" part-r-00000" file name might change , to check the exact name of the file run this line :

$ hadoop fs -ls /p6/max_temperature_per_year

```
hdoop@NuvobookV1:~/lab/p6$ hadoop fs -ls /p6/max_temperature_per_year
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2024-10-18 03:19 /p6/max_te
mperature_per_year/_SUCCESS
-rw-r--r--   1 hdoop supergroup        373 2024-10-18 03:19 /p6/max_te
mperature_per_year/part-r-00000
```

-write quit to stop pig

```
grunt> quit
2024-10-18 03:27:31,573 [main] INFO  org.apache.pig.Main - Pig script comple
ted in 54 minutes, 8 seconds and 763 milliseconds (3248763 ms)
hdoop@NuvobookV1:~$
```

#stop Hadoop:

$stop-all.sh