Report for MapReduce2 Java

HAN Duqing & GAO Xin

1.8 Remarkable trees of Paris

- List all jobs.

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar listjob
Unknown program 'listjob' chosen.
Valid program names are:
  districts: A map/reduce program that lists the districts in the trees.csv.
  mosttree: A map/reduce program that displays the district that contains the most trees in the trees.csv.
  oldesttree: A map/reduce program that displays the district where the oldest tree is in the trees.csv.
  sortheight: A map/reduce program that sorts the trees height from smallest to largest in the trees.csv.
  speciecount: A map/reduce program that counts the species in the trees.csv.
  speciemaxheight: A map/reduce program that calculates the height of the tallest tree of each
specie in the trees.csv.
  species: A map/reduce program that lists the species in the trees.csv.
  wordcount: A map/reduce program that counts the words in the input files.
```

1.8.1 Districts containing trees

A map/reduce program that lists the districts in the trees.csv.

	Mapper	**Reducer**
ClassName	DistrictsMapper	DistrictsReducer
InputKeyFormat	LongWritable	Text
InputValueFormat	Text	NullWritable
OutputKeyFormat	Text	Text
OutputValueFormat	NullWritable	NullWritable

- Run the job "districts".

```
yarn jar q7.jar districts trees.csv q1
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar districts trees.csv q1
20/11/07 12:28:27 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.o
nline/163.172.100.24:10200
20/11/07 12:28:27 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLIN
E, renewer=yarn, realUser=, issueDate=1604748507917, maxDate=1605353307917, sequenceNumber=4672, masterKeyI
d=42 on ha-hdfs:efrei
20/11/07 12:28:27 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service:
ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604748507917, maxDate=1605353307917, sequenceNumber=4672, masterKeyId=42)
20/11/07 12:28:28 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.stagin
g/job_1603290159664_2480
20/11/07 12:28:28 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 12:28:28 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 12:28:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2480
20/11/07 12:28:29 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service
: ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realU
ser=, issueDate=1604748507917, maxDate=1605353307917, sequenceNumber=4672, masterKeyId=42)]
20/11/07 12:28:29 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-15
2/0/resource-types.xml
20/11/07 12:28:29 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 12:28:29 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2480
20/11/07 12:28:29 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/p
roxy/application_1603290159664_2480/
20/11/07 12:28:29 INFO mapreduce.Job: Running job: job_1603290159664_2480
20/11/07 12:28:40 INFO mapreduce.Job: Job job_1603290159664_2480 running in uber mode : false
20/11/07 12:28:40 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 12:28:49 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 12:28:58 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 12:28:58 INFO mapreduce.Job: Job job_1603290159664_2480 completed successfully
20/11/07 12:28:58 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=84
                FILE: Number of bytes written=493223
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=44
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=20127
                Total time spent by all reduces in occupied slots (ms)=27272
                Total time spent by all map tasks (ms)=6709
                Total time spent by all reduce tasks (ms)=6818
                Total vcore-milliseconds taken by all map tasks=6709
                Total vcore-milliseconds taken by all reduce tasks=6818
                Total megabyte-milliseconds taken by all map tasks=10305024
                Total megabyte-milliseconds taken by all reduce tasks=13963264
        Map-Reduce Framework
                Map input records=98
                Map output records=97
                Map output bytes=277
                Map output materialized bytes=84
                Input split bytes=97
                Combine input records=97
                Combine output records=17
                Reduce input groups=17
                Reduce shuffle bytes=84
                Reduce input records=17
                Reduce output records=17
                Spilled Records=34
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=197
                CPU time spent (ms)=3330
                Physical memory (bytes) snapshot=1459490816
                Virtual memory (bytes) snapshot=7272185856
                Total committed heap usage (bytes)=1556086784
                Peak Map Physical memory (bytes)=1159405568
                Peak Map Virtual memory (bytes)=3394650112
                Peak Reduce Physical memory (bytes)=300085248
                Peak Reduce Virtual memory (bytes)=3877535744
```

```
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=16778
        File Output Format Counters
                Bytes Written=44
[dhan@hadoop-edge01 ~]$
```

- Display the result of job "districts".

```
hdfs dfs -cat q1/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q1/part-r-00000
11
12
13
14
15
16
17
18
19
20
3
4
5
6
7
8
9
```

1.8.2 Show all existing species

A map/reduce program that lists the species in the trees.csv.

	Mapper	**Reducer**
ClassName	SpeciesMapper	SpeciesReducer
InputKeyFormat	LongWritable	Text
InputValueFormat	Text	NullWritable
OutputKeyFormat	Text	Text
OutputValueFormat	NullWritable	NullWritable

- Run the job "species".

```
yarn jar q7.jar species trees.csv q2
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar species trees.csv q2
20/11/07 12:24:26 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.o
nline/163.172.100.24:10200
20/11/07 12:24:26 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLIN
E, renewer=yarn, realUser=, issueDate=1604748266975, maxDate=1605353066975, sequenceNumber=4671, masterKeyI
d=42 on ha-hdfs:efrei
20/11/07 12:24:27 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service:
ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604748266975, maxDate=1605353066975, sequenceNumber=4671, masterKeyId=42)
20/11/07 12:24:27 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.stagin
g/job_1603290159664_2479
20/11/07 12:24:27 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 12:24:27 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 12:24:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2479
20/11/07 12:24:28 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service
: ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realU
ser=, issueDate=1604748266975, maxDate=1605353066975, sequenceNumber=4671, masterKeyId=42)]
20/11/07 12:24:28 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-15
2/0/resource-types.xml
20/11/07 12:24:28 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 12:24:28 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2479
20/11/07 12:24:28 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/p
roxy/application_1603290159664_2479/
20/11/07 12:24:28 INFO mapreduce.Job: Running job: job_1603290159664_2479
20/11/07 12:24:39 INFO mapreduce.Job: Job job_1603290159664_2479 running in uber mode : false
20/11/07 12:24:39 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 12:24:48 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 12:24:58 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 12:24:59 INFO mapreduce.Job: Job job_1603290159664_2479 completed successfully
20/11/07 12:24:59 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=395
                FILE: Number of bytes written=493829
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=317
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=21540
                Total time spent by all reduces in occupied slots (ms)=28380
                Total time spent by all map tasks (ms)=7180
                Total time spent by all reduce tasks (ms)=7095
                Total vcore-milliseconds taken by all map tasks=7180
                Total vcore-milliseconds taken by all reduce tasks=7095
                Total megabyte-milliseconds taken by all map tasks=11028480
                Total megabyte-milliseconds taken by all reduce tasks=14530560
        Map-Reduce Framework
                Map input records=98
                Map output records=97
                Map output bytes=835
                Map output materialized bytes=395
                Input split bytes=97
                Combine input records=97
                Combine output records=36
                Reduce input groups=36
                Reduce shuffle bytes=395
                Reduce input records=36
                Reduce output records=36
                Spilled Records=72
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=201
                CPU time spent (ms)=3550
                Physical memory (bytes) snapshot=1456304128
                Virtual memory (bytes) snapshot=7266349056
                Total committed heap usage (bytes)=156378240
                Peak Map Physical memory (bytes)=1157955584
                Peak Map Virtual memory (bytes)=3393851392
                Peak Reduce Physical memory (bytes)=298348544
                Peak Reduce Virtual memory (bytes)=3872497664
```

```
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=16778
        File Output Format Counters
                Bytes Written=317
[dhan@hadoop-edge01 ~]$
```

- Display the result of job "species".

```
hdfs dfs -cat q2/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q2/part-r-00000
Acer
Aesculus
Ailanthus
Alnus
Araucaria
Broussonetia
Calocedrus
Catalpa
Cedrus
Celtis
Corylus
Davidia
Diospyros
Eucommia
Fagus
Fraxinus
Ginkgo
Gymnocladus
Juglans
Liriodendron
Maclura
Magnolia
Paulownia
Pinus
Platanus
Pterocarya
Quercus
Robinia
Sequoia
Sequoiadendron
Styphnolobium
Taxodium
Taxus
Tilia
Ulmus
Zelkova
[dhan@hadoop-edge01 ~]$
```

1.8.3 Number of trees by species

A map/reduce program that counts the species in the trees.csv.

	Mapper	Reducer
ClassName	SpecieCountMapper	SpecieCountReducer
InputKeyFormat	LongWritable	Text
InputValueFormat	Text	IntWritable
OutputKeyFormat	Text	Text
OutputValueFormat	IntWritable	IntWritable

- Run the job "speciecount".

```
yarn jar q7.jar speciecount trees.csv q3
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar speciecount trees.csv q3
20/11/07 12:30:42 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.o
nline/163.172.100.24:10200
20/11/07 12:30:42 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLIN
E, renewer=yarn, realUser=, issueDate=1604748642253, maxDate=1605353442253, sequenceNumber=4673, masterKeyI
d=42 on ha-hdfs:efrei
20/11/07 12:30:42 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service:
ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604748642253, maxDate=1605353442253, sequenceNumber=4673, masterKeyId=42)
20/11/07 12:30:42 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.stagin
g/job_1603290159664_2481
20/11/07 12:30:43 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 12:30:43 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 12:30:43 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2481
20/11/07 12:30:43 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service
: ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realU
ser=, issueDate=1604748642253, maxDate=1605353442253, sequenceNumber=4673, masterKeyId=42)]
20/11/07 12:30:43 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-15
2/0/resource-types.xml
20/11/07 12:30:43 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 12:30:44 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2481
20/11/07 12:30:44 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/p
roxy/application_1603290159664_2481/
20/11/07 12:30:44 INFO mapreduce.Job: Running job: job_1603290159664_2481
20/11/07 12:30:54 INFO mapreduce.Job: Job job_1603290159664_2481 running in uber mode : false
20/11/07 12:30:54 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 12:31:03 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 12:31:13 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 12:31:13 INFO mapreduce.Job: Job job_1603290159664_2481 completed successfully
20/11/07 12:31:13 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=539
                FILE: Number of bytes written=494147
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=390
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=21216
                Total time spent by all reduces in occupied slots (ms)=27632
                Total time spent by all map tasks (ms)=7072
                Total time spent by all reduce tasks (ms)=6908
                Total vcore-milliseconds taken by all map tasks=7072
                Total vcore-milliseconds taken by all reduce tasks=6908
                Total megabyte-milliseconds taken by all map tasks=10862592
                Total megabyte-milliseconds taken by all reduce tasks=14147584
        Map-Reduce Framework
                Map input records=98
                Map output records=97
                Map output bytes=1223
                Map output materialized bytes=539
                Input split bytes=97
                Combine input records=97
                Combine output records=36
                Reduce input groups=36
                Reduce shuffle bytes=539
                Reduce input records=36
                Reduce output records=36
                Spilled Records=72
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=205
                CPU time spent (ms)=3290
                Physical memory (bytes) snapshot=1457258496
                Virtual memory (bytes) snapshot=7273926656
                Total committed heap usage (bytes)=1560805376
                Peak Map Physical memory (bytes)=1156091904
                Peak Map Virtual memory (bytes)=3393040384
                Peak Reduce Physical memory (bytes)=301166592
                Peak Reduce Virtual memory (bytes)=3880886272
```

```
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=16778
        File Output Format Counters
                Bytes Written=390
[dhan@hadoop-edge01 ~]$
```

- Display the result of job "speciecount" (key-value: specie-count).

```
hdfs dfs -cat q3/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q3/part-r-00000
Acer    3
Aesculus        3
Ailanthus       1
Alnus   1
Araucaria       1
Broussonetia    1
Calocedrus      1
Catalpa 1
Cedrus  4
Celtis  1
Corylus 3
Davidia 1
Diospyros       4
Eucommia        1
Fagus   8
Fraxinus        1
Ginkgo  5
Gymnocladus     1
Juglans 1
Liriodendron    2
Maclura 1
Magnolia        1
Paulownia       1
Pinus   5
Platanus        19
Pterocarya      3
Quercus 4
Robinia 1
Sequoia 1
Sequoiadendron  5
Styphnolobium   1
Taxodium        3
Taxus   2
Tilia   1
Ulmus   1
Zelkova 4
[dhan@hadoop-edge01 ~]$
```

1.8.4 Maximum height per specie of tree

A map/reduce program that calculates the height of the tallest tree of each specie in the trees.csv.

	Mapper	Reducer
ClassName	SpecieMaxHeightMapper	SpecieMaxHeightReducer
InputKeyFormat	LongWritable	Text
InputValueFormat	Text	FloatWritable
OutputKeyFormat	Text	Text
OutputValueFormat	FloatWritable	FloatWritable

- Run the job "speciemaxheight".

```
yarn jar q7.jar speciemaxheight trees.csv q4
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar speciemaxheight trees.csv q4
20/11/07 12:33:55 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.o
nline/163.172.100.24:10200
20/11/07 12:33:55 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLIN
E, renewer=yarn, realUser=, issueDate=1604748835945, maxDate=1605353635945, sequenceNumber=4674, masterKeyI
d=42 on ha-hdfs:efrei
20/11/07 12:33:55 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service:
ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604748835945, maxDate=1605353635945, sequenceNumber=4674, masterKeyId=42)
20/11/07 12:33:56 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.stagin
g/job_1603290159664_2482
20/11/07 12:33:56 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 12:33:57 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 12:33:57 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2482
20/11/07 12:33:57 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service
: ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realU
ser=, issueDate=1604748835945, maxDate=1605353635945, sequenceNumber=4674, masterKeyId=42)]
20/11/07 12:33:57 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-15
2/0/resource-types.xml
20/11/07 12:33:57 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 12:33:57 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2482
20/11/07 12:33:57 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/p
roxy/application_1603290159664_2482/
20/11/07 12:33:57 INFO mapreduce.Job: Running job: job_1603290159664_2482
20/11/07 12:34:08 INFO mapreduce.Job: Job job_1603290159664_2482 running in uber mode : false
20/11/07 12:34:08 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 12:34:17 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 12:34:22 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 12:34:22 INFO mapreduce.Job: Job job_1603290159664_2482 completed successfully
20/11/07 12:34:22 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=539
                FILE: Number of bytes written=494183
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=496
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=19113
                Total time spent by all reduces in occupied slots (ms)=11792
                Total time spent by all map tasks (ms)=6371
                Total time spent by all reduce tasks (ms)=2948
                Total vcore-milliseconds taken by all map tasks=6371
                Total vcore-milliseconds taken by all reduce tasks=2948
                Total megabyte-milliseconds taken by all map tasks=9785856
                Total megabyte-milliseconds taken by all reduce tasks=6037504
        Map-Reduce Framework
                Map input records=98
                Map output records=96
                Map output bytes=1204
                Map output materialized bytes=539
                Input split bytes=97
                Combine input records=96
                Combine output records=36
                Reduce input groups=36
                Reduce shuffle bytes=539
                Reduce input records=36
                Reduce output records=36
                Spilled Records=72
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=183
                CPU time spent (ms)=3140
                Physical memory (bytes) snapshot=1452351488
                Virtual memory (bytes) snapshot=7267012608
                Total committed heap usage (bytes)=1570766848
                Peak Map Physical memory (bytes)=1158815744
                Peak Map Virtual memory (bytes)=3393449984
                Peak Reduce Physical memory (bytes)=293535744
                Peak Reduce Virtual memory (bytes)=3873562624
```

```
       Shuffle Errors
               BAD_ID=0
               CONNECTION=0
               IO_ERROR=0
               WRONG_LENGTH=0
               WRONG_MAP=0
               WRONG_REDUCE=0
       File Input Format Counters
               Bytes Read=16778
       File Output Format Counters
               Bytes Written=496
[dhan@hadoop-edge01 ~]$
```

- Display the result of job "speciemaxheight" (key-value: specie-maxheight).

```
hdfs dfs -cat q4/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q4/part-r-00000
Acer      16.0
Aesculus          30.0
Ailanthus         35.0
Alnus     16.0
Araucaria         9.0
Broussonetia      12.0
Calocedrus        20.0
Catalpa 15.0
Cedrus   30.0
Celtis   16.0
Corylus 20.0
Davidia 12.0
Diospyros         14.0
Eucommia          12.0
Fagus     30.0
Fraxinus          30.0
Ginkgo   33.0
Gymnocladus       10.0
Juglans 28.0
Liriodendron      35.0
Maclura 13.0
Magnolia          12.0
Paulownia         20.0
Pinus     30.0
Platanus          45.0
Pterocarya        30.0
Quercus 31.0
Robinia 11.0
Sequoia 30.0
Sequoiadendron    35.0
Styphnolobium     10.0
Taxodium          35.0
Taxus     13.0
Tilia     20.0
Ulmus     15.0
Zelkova 30.0
[dhan@hadoop-edge01 ~]$
```

1.8.5 Sort the trees height from smallest to largest

A map/reduce program that sorts the trees height from smallest to largest in the trees.csv.

	Mapper	**Reducer**
ClassName	SortHeightMapper	SortHeightReducer
InputKeyFormat	LongWritable	FloatWritable
InputValueFormat	Text	Text
OutputKeyFormat	FloatWritable	FloatWritable
OutputValueFormat	Text	Text

- Run the job "sortheight".

```
yarn jar q7.jar sortheight trees.csv q5
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar sortheight trees.csv q5
20/11/07 12:37:02 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.o
nline/163.172.100.24:10200
20/11/07 12:37:02 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLIN
E, renewer=yarn, realUser=, issueDate=1604749022353, maxDate=1605353822353, sequenceNumber=4675, masterKeyI
d=42 on ha-hdfs:efrei
20/11/07 12:37:02 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service:
ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604749022353, maxDate=1605353822353, sequenceNumber=4675, masterKeyId=42)
20/11/07 12:37:02 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.stagin
g/job_1603290159664_2483
20/11/07 12:37:03 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 12:37:03 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 12:37:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2483
20/11/07 12:37:03 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service
: ha-hdfs:efrei, Ident: (token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realU
ser=, issueDate=1604749022353, maxDate=1605353822353, sequenceNumber=4675, masterKeyId=42)]
20/11/07 12:37:03 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-15
2/0/resource-types.xml
20/11/07 12:37:03 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 12:37:04 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2483
20/11/07 12:37:04 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/p
roxy/application_1603290159664_2483/
20/11/07 12:37:04 INFO mapreduce.Job: Running job: job_1603290159664_2483
20/11/07 12:37:14 INFO mapreduce.Job: Job job_1603290159664_2483 running in uber mode : false
20/11/07 12:37:14 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 12:37:23 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 12:37:31 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 12:37:32 INFO mapreduce.Job: Job job_1603290159664_2483 completed successfully
20/11/07 12:37:32 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=861
                FILE: Number of bytes written=494787
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=755
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=20157
                Total time spent by all reduces in occupied slots (ms)=26596
                Total time spent by all map tasks (ms)=6719
                Total time spent by all reduce tasks (ms)=6649
                Total vcore-milliseconds taken by all map tasks=6719
                Total vcore-milliseconds taken by all reduce tasks=6649
                Total megabyte-milliseconds taken by all map tasks=10320384
                Total megabyte-milliseconds taken by all reduce tasks=13617152
        Map-Reduce Framework
                Map input records=98
                Map output records=96
                Map output bytes=663
                Map output materialized bytes=861
                Input split bytes=97
                Combine input records=96
                Combine output records=96
                Reduce input groups=28
                Reduce shuffle bytes=861
                Reduce input records=96
                Reduce output records=96
                Spilled Records=192
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=197
                CPU time spent (ms)=3220
                Physical memory (bytes) snapshot=1457455104
                Virtual memory (bytes) snapshot=7271501824
                Total committed heap usage (bytes)=1552416768
                Peak Map Physical memory (bytes)=1156358144
                Peak Map Virtual memory (bytes)=3394916352
                Peak Reduce Physical memory (bytes)=301096960
                Peak Reduce Virtual memory (bytes)=3876585472
```

```
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=16778
        File Output Format Counters
                Bytes Written=755
[dhan@hadoop-edge01 ~]$
```

- Display the result of job "sortheight" (key-value: height-objectid).

```
hdfs dfs -cat q5/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q5/part-r-00000
2.0      3
5.0      89
6.0      62
9.0      39
10.0     44
10.0     32
10.0     95
10.0     61
10.0     63
11.0     4
12.0     93
12.0     66
12.0     50
12.0     7
12.0     48
12.0     58
12.0     33
12.0     71
13.0     36
13.0     6
14.0     68
14.0     96
14.0     94
15.0     91
15.0     5
15.0     70
15.0     2
15.0     98
16.0     28
16.0     78
16.0     75
16.0     16
18.0     64
18.0     83
18.0     23
18.0     60
20.0     34
20.0     51
20.0     43
20.0     15
20.0     11
20.0     1
20.0     8
20.0     20
20.0     35
20.0     12
20.0     87
20.0     13
22.0     10
22.0     47
22.0     86
22.0     14
22.0     88
23.0     18
25.0     24
25.0     31
25.0     92
25.0     49
```

```
25.0    97
25.0    84
26.0    73
27.0    65
27.0    42
28.0    85
30.0    76
30.0    19
30.0    72
30.0    54
30.0    29
30.0    27
30.0    25
30.0    41
30.0    77
30.0    55
30.0    69
30.0    38
30.0    59
30.0    52
30.0    37
30.0    22
30.0    30
31.0    80
31.0    9
32.0    82
33.0    46
34.0    45
35.0    56
35.0    81
35.0    17
35.0    53
35.0    57
40.0    26
40.0    74
40.0    40
42.0    90
45.0    21
[dhan@hadoop-edge01 ~]$
```

1.8.6 District containing the oldest tree

A map/reduce program that displays the district where the oldest tree is in the trees.csv.

We define a Writable class called MyWritable which is the pair of (plantyear , district).

	Mapper	Reducer
ClassName	OldestTreeMapper	OldestTreeReducer
InputKeyFormat	LongWritable	Text
InputValueFormat	Text	MyWritable
OutputKeyFormat	Text	Text
OutputValueFormat	MyWritable	NullWritable

- Run the job "oldesttree" and display the result.

```
yarn jar q7.jar oldesttree trees.csv q6
hdfs dfs -cat q6/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q6.jar oldesttree trees.csv q6
20/11/07 14:29:57 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.online/163.172.100.24:10200
20/11/07 14:29:57 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, i
ssueDate=1604755797815, maxDate=1605360597815, sequenceNumber=4709, masterKeyId=42 on ha-hdfs:efrei
20/11/07 14:29:57 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:efrei, Ident: (token
for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, issueDate=1604755797815, maxDate=1605360597815, sequen
ceNumber=4709, masterKeyId=42)
20/11/07 14:29:57 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.staging/job_1603290159664_2515
20/11/07 14:29:58 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 14:29:58 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 14:29:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2515
20/11/07 14:29:58 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:efrei, Ident: (toke
n for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, issueDate=1604755797815, maxDate=1605360597815, sequ
enceNumber=4709, masterKeyId=42)]
20/11/07 14:29:59 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-152/0/resource-types.xml
20/11/07 14:29:59 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 14:29:59 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2515
20/11/07 14:29:59 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/proxy/application_160329015966
4_2515/
20/11/07 14:29:59 INFO mapreduce.Job: Running job: job_1603290159664_2515
20/11/07 14:30:09 INFO mapreduce.Job: Job job_1603290159664_2515 running in uber mode : false
20/11/07 14:30:09 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 14:30:19 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 14:30:29 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 14:30:30 INFO mapreduce.Job: Job job_1603290159664_2515 completed successfully
20/11/07 14:30:30 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=1072
                FILE: Number of bytes written=495575
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=2
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=21747
                Total time spent by all reduces in occupied slots (ms)=28528
                Total time spent by all map tasks (ms)=7249
                Total time spent by all reduce tasks (ms)=7132
                Total vcore-milliseconds taken by all map tasks=7249
                Total vcore-milliseconds taken by all reduce tasks=7132
                Total megabyte-milliseconds taken by all map tasks=11134464
                Total megabyte-milliseconds taken by all reduce tasks=14606336
        Map-Reduce Framework
                Map input records=98
                Map output records=77
                Map output bytes=912
                Map output materialized bytes=1072
                Input split bytes=97
                Combine input records=0
                Combine output records=0
                Reduce input groups=1
                Reduce shuffle bytes=1072
                Reduce input records=77
                Reduce output records=1
                Spilled Records=154
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=196
                CPU time spent (ms)=3450
                Physical memory (bytes) snapshot=1453199360
                Virtual memory (bytes) snapshot=7272005632
                Total committed heap usage (bytes)=1551368192
                Peak Map Physical memory (bytes)=1159045120
                Peak Map Virtual memory (bytes)=3395592192
                Peak Reduce Physical memory (bytes)=294154240
                Peak Reduce Virtual memory (bytes)=3876413440
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=16778
        File Output Format Counters
                Bytes Written=2
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q6/part-r-00000
5
[dhan@hadoop-edge01 ~]$
```

- Verify the result of job "oldesttree".

```
df[df['ANNEE PLANTATION']==min(df['ANNEE PLANTATION'])]
```

	GEOPOINT	ARRONDISSEMENT	GENRE	ESPECE	FAMILLE	ANNEE PLANTATION	HAUTEUR	CIRCONFERENCE	ADRESSE	NOM COMMUN	VARIETE	OBJEC
26	(48.8520958913, 2.34740754195)	5	Robinia	pseudoacacia	Fabaceae	1601.0	11.0	385.0	Quai de Montebello, rue Lagrange, rue Saint-Ju...	Robinier faux-acacia	NaN	

We can see that District 5 have the oldest tree is correct.

1.8.7 District containing the most trees

A map/reduce program that displays the district that contains the most trees in the trees.csv.

We define a Writable class called MyWritable which is the pair of (district , count).

We also create 2 MapReduce jobs to handle this question.

	Mapper1	**Reducer1**	**Mapper2**	**Reducer2**
ClassName	MostTreeMapper1	MostTreeReducer1	MostTreeMapper2	MostTreeReducer2
InputKeyFormat	LongWritable	Text	Oject	NullWritable
InputValueFormat	Text	IntWritable	Text	MyWritable
OutputKeyFormat	Text	Text	NullWritable	Text
OutputValueFormat	IntWritable	IntWritable	MyWritable	NullWritable

- Run the job "mosttree" and display the result.

```
yarn jar q7.jar mosttree trees.csv q7
hdfs dfs -cat q7/part-r-00000
```

```
[dhan@hadoop-edge01 ~]$ yarn jar q7.jar mosttree trees.csv q7
20/11/07 14:20:50 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.online/163.172.100.24:102
00
20/11/07 14:20:50 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604755250568, maxDate=1605360050568, sequenceNumber=4705, masterKeyId=42 on ha-hdfs:efrei
20/11/07 14:20:50 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:efrei, Ident: (t
oken for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, issueDate=1604755250568, maxDate=16053600505
68, sequenceNumber=4705, masterKeyId=42)
20/11/07 14:20:50 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.staging/job_1603290159664_2511
20/11/07 14:20:51 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 14:20:51 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 14:20:51 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2511
20/11/07 14:20:51 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:efrei, Ident:
(token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, issueDate=1604755250568, maxDate=160536005
0568, sequenceNumber=4705, masterKeyId=42)]
20/11/07 14:20:51 INFO conf.Configuration: found resource resource-types.xml at file:/etc/hadoop/3.1.5.0-152/0/resource-types.xml
20/11/07 14:20:51 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 14:20:52 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2511
20/11/07 14:20:52 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/proxy/application_1603290
159664_2511/
20/11/07 14:20:52 INFO mapreduce.Job: Running job: job_1603290159664_2511
20/11/07 14:21:02 INFO mapreduce.Job: Job job_1603290159664_2511 running in uber mode : false
20/11/07 14:21:02 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 14:21:11 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 14:21:20 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 14:21:21 INFO mapreduce.Job: Job job_1603290159664_2511 completed successfully
20/11/07 14:21:22 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=865
                FILE: Number of bytes written=495181
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=16875
                HDFS: Number of bytes written=80
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=20148
                Total time spent by all reduces in occupied slots (ms)=28080
                Total time spent by all map tasks (ms)=6716
                Total time spent by all reduce tasks (ms)=7020
                Total vcore-milliseconds taken by all map tasks=6716
                Total vcore-milliseconds taken by all reduce tasks=7020
                Total megabyte-milliseconds taken by all map tasks=10315776
                Total megabyte-milliseconds taken by all reduce tasks=14376960
        Map-Reduce Framework
                Map input records=98
                Map output records=97
                Map output bytes=665
                Map output materialized bytes=865
                Input split bytes=97
                Combine input records=0
                Combine output records=0
                Reduce input groups=17
                Reduce shuffle bytes=865
                Reduce input records=97
                Reduce output records=17
                Spilled Records=194
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=185
                CPU time spent (ms)=3560
                Physical memory (bytes) snapshot=1453039616
                Virtual memory (bytes) snapshot=7273865216
                Total committed heap usage (bytes)=1571815424
                Peak Map Physical memory (bytes)=1158545408
                Peak Map Virtual memory (bytes)=3397316608
                Peak Reduce Physical memory (bytes)=294494208
                Peak Reduce Virtual memory (bytes)=3876548608
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=16778
        File Output Format Counters
                Bytes Written=80
20/11/07 14:21:22 INFO client.AHSProxy: Connecting to Application History server at hadoop-master03.efrei.online/163.172.100.24:102
00
20/11/07 14:21:22 INFO hdfs.DFSClient: Created token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUse
r=, issueDate=1604755282102, maxDate=1605360082102, sequenceNumber=4706, masterKeyId=42 on ha-hdfs:efrei
20/11/07 14:21:22 INFO security.TokenCache: Got dt for hdfs://efrei; Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:efrei, Ident: (t
oken for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, issueDate=1604755282102, maxDate=16053600821
02, sequenceNumber=4706, masterKeyId=42)
20/11/07 14:21:22 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interfac
e and execute your application with ToolRunner to remedy this.
20/11/07 14:21:22 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /user/dhan/.staging/job_1603290159664_2512
```

```
20/11/07 14:21:22 INFO input.FileInputFormat: Total input files to process : 1
20/11/07 14:21:22 INFO mapreduce.JobSubmitter: number of splits:1
20/11/07 14:21:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1603290159664_2512
20/11/07 14:21:22 INFO mapreduce.JobSubmitter: Executing with tokens: [Kind: HDFS_DELEGATION_TOKEN, Service: ha-hdfs:efrei, Ident:
(token for dhan: HDFS_DELEGATION_TOKEN owner=dhan@EFREI.ONLINE, renewer=yarn, realUser=, issueDate=1604755282102, maxDate=160536008
2102, sequenceNumber=4706, masterKeyId=42)]
20/11/07 14:21:22 INFO impl.TimelineClientImpl: Timeline service address: hadoop-master03.efrei.online:8190
20/11/07 14:21:23 INFO impl.YarnClientImpl: Submitted application application_1603290159664_2512
20/11/07 14:21:23 INFO mapreduce.Job: The url to track the job: https://hadoop-master01.efrei.online:8090/proxy/application_1603290
159664_2512/
20/11/07 14:21:23 INFO mapreduce.Job: Running job: job_1603290159664_2512
20/11/07 14:21:33 INFO mapreduce.Job: Job job_1603290159664_2512 running in uber mode : false
20/11/07 14:21:33 INFO mapreduce.Job:  map 0% reduce 0%
20/11/07 14:21:42 INFO mapreduce.Job:  map 100% reduce 0%
20/11/07 14:21:52 INFO mapreduce.Job:  map 100% reduce 100%
20/11/07 14:21:52 INFO mapreduce.Job: Job job_1603290159664_2512 completed successfully
20/11/07 14:21:52 INFO mapreduce.Job: Counters: 53
        File System Counters
                FILE: Number of bytes read=169
                FILE: Number of bytes written=493427
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=192
                HDFS: Number of bytes written=3
                HDFS: Number of read operations=8
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=21090
                Total time spent by all reduces in occupied slots (ms)=26812
                Total time spent by all map tasks (ms)=7030
                Total time spent by all reduce tasks (ms)=6703
                Total vcore-milliseconds taken by all map tasks=7030
                Total vcore-milliseconds taken by all reduce tasks=6703
                Total megabyte-milliseconds taken by all map tasks=10798080
                Total megabyte-milliseconds taken by all reduce tasks=13727744
        Map-Reduce Framework
                Map input records=17
                Map output records=17
                Map output bytes=129
                Map output materialized bytes=169
                Input split bytes=112
                Combine input records=0
                Combine output records=0
                Reduce input groups=1
                Reduce shuffle bytes=169
                Reduce input records=17
                Reduce output records=1
                Spilled Records=34
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=171
                CPU time spent (ms)=3140
                Physical memory (bytes) snapshot=1452314624
                Virtual memory (bytes) snapshot=7268995072
                Total committed heap usage (bytes)=1556611072
                Peak Map Physical memory (bytes)=1163288576
                Peak Map Virtual memory (bytes)=3393609728
                Peak Reduce Physical memory (bytes)=289026048
                Peak Reduce Virtual memory (bytes)=3875385344
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=80
        File Output Format Counters
                Bytes Written=3
[dhan@hadoop-edge01 ~]$ hdfs dfs -cat q7/part-r-00000
16
[dhan@hadoop-edge01 ~]$
```

- Verify the result of job "oldesttree".

```
df.groupby('ARRONDISSEMENT').count()['OBJECTID'].sort_values(ascending=False)
```

```
ARRONDISSEMENT
16      36
12      29
19       6
8        5
14       3
7        3
20       3
13       2
5        2
15       1
11       1
9        1
17       1
18       1
6        1
4        1
3        1
Name: OBJECTID, dtype: int64
```

We can see that District 16 have most trees is correct.