**Cloud Computing** ssh -i "ClassKeyPair.pem" hadoop@ec2-52-14-38-219.us-east-2.compute.amazonaws.com

The authenticity of host 'ec2-52-14-38-219.us-east-2.compute.amazonaws.com (52.14.38.219)' can't be established.

ECDSA key fingerprint is SHA256:o0PRcvIkuAkZO0WqG5yR/MjDnD+rzNy7IXfVvKvrQcI.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'ec2-52-14-38-219.us-east-2.compute.amazonaws.com,52.14.38.209' (ECDSA) to the list of known hosts.

Last login: Tue Mar 2 15:09:32 2021

\_\_| \_\_|\_ )

\_| ( / Amazon Linux 2 AMI

\_\_\_|\\_\_\_|\_\_\_|

https://aws.amazon.com/amazon-linux-2/

38 package(s) needed for security, out of 76 available

Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRR

E::::::::::::::::::E M:::::::M M:::::::M R::::::::::::::R

EE:::::EEEEEEEEE:::E M::::::::M M::::::::M R:::::RRRRRR:::::R

E::::E EEEEE M:::::::::M M:::::::::M RR::::R R::::R

E::::E M::::::M:::M M:::M::::::M R:::R R::::R

E:::::EEEEEEEEEE M:::::M M:::M M:::M M:::::M R:::RRRRRR:::::R

E::::::::::::::E M:::::M M:::M:::M M:::::M R:::::::::::RR

E:::::EEEEEEEEEE M:::::M M:::::M M:::::M R:::RRRRRR::::R

E::::E M:::::M M:::M M:::::M R:::R R::::R

E::::E EEEEE M:::::M MMM M:::::M R:::R R::::R

EE:::::EEEEEEEE::::E M:::::M M:::::M R:::R R::::R

E::::::::::::::::::E M:::::M M:::::M RR::::R R::::R

EEEEEEEEEEEEEEEEEEEE MMMMMMM MMMMMMM RRRRRRR RRRRRR

[hadoop@ip-172-31-20-52 ~]$ aws s3 cp s3://bucket801168390/PageRank.jar .

download: s3://bucket801168390/PageRank.jar to ./PageRank.jar

[hadoop@ip-172-31-20-52 ~]$ hadoop jar ./PageRank.jar Driver s3://bucket801168390/input-pages.txt /user/PageRankOutput 6

21/03/02 15:12:22 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:12:22 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:12:22 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:12:24 INFO input.FileInputFormat: Total input files to process : 1

21/03/02 15:12:24 INFO lzo.GPLNativeCodeLoader: Loaded native gpl library

21/03/02 15:12:24 INFO lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo rev 3fb854bbfdabadafad1fa2cca072658fa097fd67]

21/03/02 15:12:24 INFO mapreduce.JobSubmitter: number of splits:1

21/03/02 15:12:25 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0002

21/03/02 15:12:25 INFO conf.Configuration: resource-types.xml not found

21/03/02 15:12:25 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

21/03/02 15:12:25 INFO resource.ResourceUtils: Adding resource type - name = memory-mb, units = Mi, type = COUNTABLE

21/03/02 15:12:25 INFO resource.ResourceUtils: Adding resource type - name = vcores, units = , type = COUNTABLE

21/03/02 15:12:26 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0002

21/03/02 15:12:27 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0002/

21/03/02 15:12:27 INFO mapreduce.Job: Running job: job\_1614697513587\_0002

21/03/02 15:12:38 INFO mapreduce.Job: Job job\_1614697513587\_0002 running in uber mode : false

21/03/02 15:12:38 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:12:49 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:12:57 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:12:57 INFO mapreduce.Job: Job job\_1614697513587\_0002 completed successfully

21/03/02 15:12:57 INFO mapreduce.Job: Counters: 54

File System Counters

FILE: Number of bytes read=659512

FILE: Number of bytes written=1758605

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=101

HDFS: Number of bytes written=2093854

HDFS: Number of read operations=5

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

S3: Number of bytes read=5644081

S3: Number of bytes written=0

S3: Number of read operations=0

S3: Number of large read operations=0

S3: Number of write operations=0

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)=399504

Total time spent by all reduces in occupied slots (ms)=538560

Total time spent by all map tasks (ms)=8323

Total time spent by all reduce tasks (ms)=5610

Total vcore-milliseconds taken by all map tasks=8323

Total vcore-milliseconds taken by all reduce tasks=5610

Total megabyte-milliseconds taken by all map tasks=12784128

Total megabyte-milliseconds taken by all reduce tasks=17233920

Map-Reduce Framework

Map input records=2427

Map output records=2427

Map output bytes=2047882

Map output materialized bytes=659508

Input split bytes=101

Combine input records=0

Combine output records=0

Reduce input groups=1

Reduce shuffle bytes=659508

Reduce input records=2427

Reduce output records=2427

Spilled Records=4854

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=351

CPU time spent (ms)=9430

Physical memory (bytes) snapshot=959668224

Virtual memory (bytes) snapshot=8004730880

Total committed heap usage (bytes)=737673216

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=5644081

File Output Format Counters

Bytes Written=2093854

21/03/02 15:12:57 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:12:57 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:12:57 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:12:58 INFO input.FileInputFormat: Total input files to process : 1

21/03/02 15:12:58 INFO mapreduce.JobSubmitter: number of splits:1

21/03/02 15:12:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0003

21/03/02 15:12:58 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0003

21/03/02 15:12:58 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0003/

21/03/02 15:12:58 INFO mapreduce.Job: Running job: job\_1614697513587\_0003

21/03/02 15:13:07 INFO mapreduce.Job: Job job\_1614697513587\_0003 running in uber mode : false

21/03/02 15:13:07 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:13:14 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:13:23 INFO mapreduce.Job: map 100% reduce 33%

21/03/02 15:13:27 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:13:28 INFO mapreduce.Job: Job job\_1614697513587\_0003 completed successfully

21/03/02 15:13:28 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=1527609

FILE: Number of bytes written=3934429

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2094015

HDFS: Number of bytes written=2054983

HDFS: Number of read operations=12

HDFS: Number of large read operations=0

HDFS: Number of write operations=6

Job Counters

Killed reduce tasks=1

Launched map tasks=1

Launched reduce tasks=3

Data-local map tasks=1

Total time spent by all maps in occupied slots (ms)=243216

Total time spent by all reduces in occupied slots (ms)=2304768

Total time spent by all map tasks (ms)=5067

Total time spent by all reduce tasks (ms)=24008

Total vcore-milliseconds taken by all map tasks=5067

Total vcore-milliseconds taken by all reduce tasks=24008

Total megabyte-milliseconds taken by all map tasks=7782912

Total megabyte-milliseconds taken by all reduce tasks=73752576

Map-Reduce Framework

Map input records=2427

Map output records=48418

Map output bytes=4095001

Map output materialized bytes=1527597

Input split bytes=161

Combine input records=0

Combine output records=0

Reduce input groups=34374

Reduce shuffle bytes=1527597

Reduce input records=48418

Reduce output records=2427

Spilled Records=96836

Shuffled Maps =3

Failed Shuffles=0

Merged Map outputs=3

GC time elapsed (ms)=696

CPU time spent (ms)=8370

Physical memory (bytes) snapshot=1456947200

Virtual memory (bytes) snapshot=17329631232

Total committed heap usage (bytes)=1042284544

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=2093854

File Output Format Counters

Bytes Written=2054983

21/03/02 15:13:28 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:13:28 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:13:28 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:13:28 INFO input.FileInputFormat: Total input files to process : 3

21/03/02 15:13:28 INFO mapreduce.JobSubmitter: number of splits:3

21/03/02 15:13:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0004

21/03/02 15:13:28 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0004

21/03/02 15:13:28 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0004/

21/03/02 15:13:28 INFO mapreduce.Job: Running job: job\_1614697513587\_0004

21/03/02 15:13:36 INFO mapreduce.Job: Job job\_1614697513587\_0004 running in uber mode : false

21/03/02 15:13:36 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:13:43 INFO mapreduce.Job: map 33% reduce 0%

21/03/02 15:13:48 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:13:54 INFO mapreduce.Job: map 100% reduce 33%

21/03/02 15:13:55 INFO mapreduce.Job: map 100% reduce 67%

21/03/02 15:13:56 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:13:56 INFO mapreduce.Job: Job job\_1614697513587\_0004 completed successfully

21/03/02 15:13:56 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=1502000

FILE: Number of bytes written=4325104

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2055445

HDFS: Number of bytes written=2054914

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=6

Job Counters

Killed map tasks=1

Launched map tasks=3

Launched reduce tasks=3

Data-local map tasks=3

Total time spent by all maps in occupied slots (ms)=1106640

Total time spent by all reduces in occupied slots (ms)=2211744

Total time spent by all map tasks (ms)=23055

Total time spent by all reduce tasks (ms)=23039

Total vcore-milliseconds taken by all map tasks=23055

Total vcore-milliseconds taken by all reduce tasks=23039

Total megabyte-milliseconds taken by all map tasks=35412480

Total megabyte-milliseconds taken by all reduce tasks=70775808

Map-Reduce Framework

Map input records=2427

Map output records=48418

Map output bytes=3986384

Map output materialized bytes=1504157

Input split bytes=462

Combine input records=0

Combine output records=0

Reduce input groups=34374

Reduce shuffle bytes=1504157

Reduce input records=48418

Reduce output records=2427

Spilled Records=96836

Shuffled Maps =9

Failed Shuffles=0

Merged Map outputs=9

GC time elapsed (ms)=1137

CPU time spent (ms)=10970

Physical memory (bytes) snapshot=2446159872

Virtual memory (bytes) snapshot=23950721024

Total committed heap usage (bytes)=1808269312

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=2054983

File Output Format Counters

Bytes Written=2054914

21/03/02 15:13:56 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:13:56 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:13:56 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:13:56 INFO input.FileInputFormat: Total input files to process : 3

21/03/02 15:13:56 INFO mapreduce.JobSubmitter: number of splits:3

21/03/02 15:13:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0005

21/03/02 15:13:56 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0005

21/03/02 15:13:56 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0005/

21/03/02 15:13:56 INFO mapreduce.Job: Running job: job\_1614697513587\_0005

21/03/02 15:14:06 INFO mapreduce.Job: Job job\_1614697513587\_0005 running in uber mode : false

21/03/02 15:14:06 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:14:13 INFO mapreduce.Job: map 33% reduce 0%

21/03/02 15:14:17 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:14:23 INFO mapreduce.Job: map 100% reduce 33%

21/03/02 15:14:25 INFO mapreduce.Job: map 100% reduce 67%

21/03/02 15:14:26 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:14:26 INFO mapreduce.Job: Job job\_1614697513587\_0005 completed successfully

21/03/02 15:14:26 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=1500220

FILE: Number of bytes written=4321733

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2055376

HDFS: Number of bytes written=2054812

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=6

Job Counters

Killed reduce tasks=1

Launched map tasks=3

Launched reduce tasks=3

Data-local map tasks=3

Total time spent by all maps in occupied slots (ms)=1128048

Total time spent by all reduces in occupied slots (ms)=2281344

Total time spent by all map tasks (ms)=23501

Total time spent by all reduce tasks (ms)=23764

Total vcore-milliseconds taken by all map tasks=23501

Total vcore-milliseconds taken by all reduce tasks=23764

Total megabyte-milliseconds taken by all map tasks=36097536

Total megabyte-milliseconds taken by all reduce tasks=73003008

Map-Reduce Framework

Map input records=2427

Map output records=48418

Map output bytes=3985664

Map output materialized bytes=1502566

Input split bytes=462

Combine input records=0

Combine output records=0

Reduce input groups=34374

Reduce shuffle bytes=1502566

Reduce input records=48418

Reduce output records=2427

Spilled Records=96836

Shuffled Maps =9

Failed Shuffles=0

Merged Map outputs=9

GC time elapsed (ms)=1103

CPU time spent (ms)=10120

Physical memory (bytes) snapshot=2434736128

Virtual memory (bytes) snapshot=23946330112

Total committed heap usage (bytes)=1823473664

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=2054914

File Output Format Counters

Bytes Written=2054812

21/03/02 15:14:26 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:14:26 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:14:26 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:14:26 INFO input.FileInputFormat: Total input files to process : 3

21/03/02 15:14:26 INFO mapreduce.JobSubmitter: number of splits:3

21/03/02 15:14:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0006

21/03/02 15:14:26 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0006

21/03/02 15:14:26 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0006/

21/03/02 15:14:26 INFO mapreduce.Job: Running job: job\_1614697513587\_0006

21/03/02 15:14:35 INFO mapreduce.Job: Job job\_1614697513587\_0006 running in uber mode : false

21/03/02 15:14:35 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:14:42 INFO mapreduce.Job: map 33% reduce 0%

21/03/02 15:14:47 INFO mapreduce.Job: map 67% reduce 0%

21/03/02 15:14:48 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:14:50 INFO mapreduce.Job: map 100% reduce 33%

21/03/02 15:14:53 INFO mapreduce.Job: map 100% reduce 67%

21/03/02 15:14:56 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:14:56 INFO mapreduce.Job: Job job\_1614697513587\_0006 completed successfully

21/03/02 15:14:56 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=1500389

FILE: Number of bytes written=4321984

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2055274

HDFS: Number of bytes written=2054912

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=6

Job Counters

Killed map tasks=1

Launched map tasks=3

Launched reduce tasks=3

Data-local map tasks=3

Total time spent by all maps in occupied slots (ms)=1209360

Total time spent by all reduces in occupied slots (ms)=2129472

Total time spent by all map tasks (ms)=25195

Total time spent by all reduce tasks (ms)=22182

Total vcore-milliseconds taken by all map tasks=25195

Total vcore-milliseconds taken by all reduce tasks=22182

Total megabyte-milliseconds taken by all map tasks=38699520

Total megabyte-milliseconds taken by all reduce tasks=68143104

Map-Reduce Framework

Map input records=2427

Map output records=48418

Map output bytes=3985138

Map output materialized bytes=1502648

Input split bytes=462

Combine input records=0

Combine output records=0

Reduce input groups=34374

Reduce shuffle bytes=1502648

Reduce input records=48418

Reduce output records=2427

Spilled Records=96836

Shuffled Maps =9

Failed Shuffles=0

Merged Map outputs=9

GC time elapsed (ms)=916

CPU time spent (ms)=10530

Physical memory (bytes) snapshot=2427568128

Virtual memory (bytes) snapshot=23944540160

Total committed heap usage (bytes)=1782054912

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=2054812

File Output Format Counters

Bytes Written=2054912

21/03/02 15:14:56 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:14:56 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:14:56 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:14:57 INFO input.FileInputFormat: Total input files to process : 3

21/03/02 15:14:57 INFO mapreduce.JobSubmitter: number of splits:3

21/03/02 15:14:57 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0007

21/03/02 15:14:57 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0007

21/03/02 15:14:57 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0007/

21/03/02 15:14:57 INFO mapreduce.Job: Running job: job\_1614697513587\_0007

21/03/02 15:15:05 INFO mapreduce.Job: Job job\_1614697513587\_0007 running in uber mode : false

21/03/02 15:15:05 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:15:12 INFO mapreduce.Job: map 33% reduce 0%

21/03/02 15:15:17 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:15:19 INFO mapreduce.Job: map 100% reduce 33%

21/03/02 15:15:23 INFO mapreduce.Job: map 100% reduce 67%

21/03/02 15:15:26 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:15:26 INFO mapreduce.Job: Job job\_1614697513587\_0007 completed successfully

21/03/02 15:15:26 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=1500419

FILE: Number of bytes written=4322059

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2055374

HDFS: Number of bytes written=2054887

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=6

Job Counters

Killed map tasks=1

Launched map tasks=3

Launched reduce tasks=3

Data-local map tasks=3

Total time spent by all maps in occupied slots (ms)=1162560

Total time spent by all reduces in occupied slots (ms)=2016288

Total time spent by all map tasks (ms)=24220

Total time spent by all reduce tasks (ms)=21003

Total vcore-milliseconds taken by all map tasks=24220

Total vcore-milliseconds taken by all reduce tasks=21003

Total megabyte-milliseconds taken by all map tasks=37201920

Total megabyte-milliseconds taken by all reduce tasks=64521216

Map-Reduce Framework

Map input records=2427

Map output records=48418

Map output bytes=3985476

Map output materialized bytes=1502693

Input split bytes=462

Combine input records=0

Combine output records=0

Reduce input groups=34374

Reduce shuffle bytes=1502693

Reduce input records=48418

Reduce output records=2427

Spilled Records=96836

Shuffled Maps =9

Failed Shuffles=0

Merged Map outputs=9

GC time elapsed (ms)=925

CPU time spent (ms)=10110

Physical memory (bytes) snapshot=2429280256

Virtual memory (bytes) snapshot=23921434624

Total committed heap usage (bytes)=1787822080

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=2054912

File Output Format Counters

Bytes Written=2054887

21/03/02 15:15:26 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:15:26 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:15:26 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:15:27 INFO input.FileInputFormat: Total input files to process : 3

21/03/02 15:15:27 INFO mapreduce.JobSubmitter: number of splits:3

21/03/02 15:15:28 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0008

21/03/02 15:15:28 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0008

21/03/02 15:15:28 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0008/

21/03/02 15:15:28 INFO mapreduce.Job: Running job: job\_1614697513587\_0008

21/03/02 15:15:36 INFO mapreduce.Job: Job job\_1614697513587\_0008 running in uber mode : false

21/03/02 15:15:36 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:15:43 INFO mapreduce.Job: map 33% reduce 0%

21/03/02 15:15:48 INFO mapreduce.Job: map 67% reduce 0%

21/03/02 15:15:49 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:15:51 INFO mapreduce.Job: map 100% reduce 33%

21/03/02 15:15:55 INFO mapreduce.Job: map 100% reduce 67%

21/03/02 15:15:57 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:15:58 INFO mapreduce.Job: Job job\_1614697513587\_0008 completed successfully

21/03/02 15:15:58 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=1500033

FILE: Number of bytes written=4321434

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2055349

HDFS: Number of bytes written=2054906

HDFS: Number of read operations=18

HDFS: Number of large read operations=0

HDFS: Number of write operations=6

Job Counters

Killed map tasks=1

Launched map tasks=3

Launched reduce tasks=3

Data-local map tasks=3

Total time spent by all maps in occupied slots (ms)=1202304

Total time spent by all reduces in occupied slots (ms)=2073312

Total time spent by all map tasks (ms)=25048

Total time spent by all reduce tasks (ms)=21597

Total vcore-milliseconds taken by all map tasks=25048

Total vcore-milliseconds taken by all reduce tasks=21597

Total megabyte-milliseconds taken by all map tasks=38473728

Total megabyte-milliseconds taken by all reduce tasks=66345984

Map-Reduce Framework

Map input records=2427

Map output records=48418

Map output bytes=3985186

Map output materialized bytes=1502454

Input split bytes=462

Combine input records=0

Combine output records=0

Reduce input groups=34374

Reduce shuffle bytes=1502454

Reduce input records=48418

Reduce output records=2427

Spilled Records=96836

Shuffled Maps =9

Failed Shuffles=0

Merged Map outputs=9

GC time elapsed (ms)=856

CPU time spent (ms)=10210

Physical memory (bytes) snapshot=2411745280

Virtual memory (bytes) snapshot=23959244800

Total committed heap usage (bytes)=1753219072

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=2054887

File Output Format Counters

Bytes Written=2054906

21/03/02 15:15:58 INFO client.RMProxy: Connecting to ResourceManager at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:8032

21/03/02 15:15:58 INFO client.AHSProxy: Connecting to Application History server at ip-172-31-20-52.us-east-2.compute.internal/172.31.20.52:10200

21/03/02 15:15:58 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

21/03/02 15:15:58 INFO input.FileInputFormat: Total input files to process : 3

21/03/02 15:15:58 INFO mapreduce.JobSubmitter: number of splits:3

21/03/02 15:15:58 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1614697513587\_0009

21/03/02 15:15:58 INFO impl.YarnClientImpl: Submitted application application\_1614697513587\_0009

21/03/02 15:15:58 INFO mapreduce.Job: The url to track the job: http://ip-172-31-20-52.us-east-2.compute.internal:20888/proxy/application\_1614697513587\_0009/

21/03/02 15:15:58 INFO mapreduce.Job: Running job: job\_1614697513587\_0009

21/03/02 15:16:07 INFO mapreduce.Job: Job job\_1614697513587\_0009 running in uber mode : false

21/03/02 15:16:07 INFO mapreduce.Job: map 0% reduce 0%

21/03/02 15:16:14 INFO mapreduce.Job: map 33% reduce 0%

21/03/02 15:16:18 INFO mapreduce.Job: map 67% reduce 0%

21/03/02 15:16:19 INFO mapreduce.Job: map 100% reduce 0%

21/03/02 15:16:22 INFO mapreduce.Job: map 100% reduce 100%

21/03/02 15:16:22 INFO mapreduce.Job: Job job\_1614697513587\_0009 completed successfully

21/03/02 15:16:22 INFO mapreduce.Job: Counters: 50

File System Counters

FILE: Number of bytes read=47007

FILE: Number of bytes written=977424

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=2055368

HDFS: Number of bytes written=69801

HDFS: Number of read operations=12

HDFS: Number of large read operations=0

HDFS: Number of write operations=2

Job Counters

Killed map tasks=1

Launched map tasks=3

Launched reduce tasks=1

Data-local map tasks=3

Total time spent by all maps in occupied slots (ms)=1086048

Total time spent by all reduces in occupied slots (ms)=594336

Total time spent by all map tasks (ms)=22626

Total time spent by all reduce tasks (ms)=6191

Total vcore-milliseconds taken by all map tasks=22626

Total vcore-milliseconds taken by all reduce tasks=6191

Total megabyte-milliseconds taken by all map tasks=34753536

Total megabyte-milliseconds taken by all reduce tasks=19018752

Map-Reduce Framework

Map input records=2427

Map output records=2427

Map output bytes=74771

Map output materialized bytes=48868

Input split bytes=462

Combine input records=0

Combine output records=0

Reduce input groups=37

Reduce shuffle bytes=48868

Reduce input records=2427

Reduce output records=2427

Spilled Records=4854

Shuffled Maps =3

Failed Shuffles=0

Merged Map outputs=3

GC time elapsed (ms)=643

CPU time spent (ms)=4020

Physical memory (bytes) snapshot=1813372928

Virtual memory (bytes) snapshot=14599475200

Total committed heap usage (bytes)=1400373248

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=2054906

File Output Format Counters

Bytes Written=69801

[hadoop@ip-172-31-20-52 ~]$ hadoop fs -get /user/PageRankOutput/part-r-00000 .

[hadoop@ip-172-31-20-52 ~]$ hadoop fs -ls /user/PageRankOutput

Found 2 items

-rw-r--r-- 1 hadoop hadoop 0 2021-03-02 15:16 /user/PageRankOutput/\_SUCCESS

-rw-r--r-- 1 hadoop hadoop 69801 2021-03-02 15:16 /user/PageRankOutput/part-r-00000

[hadoop@ip-172-31-20-52 ~]$ aws s3 cp ./part-r-00000 s3://bucket801168390/PageRankOutput.txt

upload: ./part-r-00000 to s3://bucket801168390/PageRankOutput.txt

[hadoop@ip-172-31-20-52 ~]$