RHadoop running with Yarn

Spark and SparkR running cluster mode with Yarn

Streaming Python jobs running with Yarn

Downloaded and started configuring HiBench, need to set more environmental variables.

Datasets located at hdfs:///user/dotcz12/

Spark Jobs failed due to Java\_Heap space

15/03/04 13:49:33 WARN scheduler.TaskSetManager: Loss was due to java.lang.OutOfMemoryError

java.lang.OutOfMemoryError: Java heap space

at java.util.Arrays.copyOf(Arrays.java:2271)

at java.io.ByteArrayOutputStream.grow(ByteArrayOutputStream.java:113)

at java.io.ByteArrayOutputStream.ensureCapacity(ByteArrayOutputStream.java:93)

at java.io.ByteArrayOutputStream.write(ByteArrayOutputStream.java:140)

at java.io.ObjectOutputStream$BlockDataOutputStream.drain(ObjectOutputStream.java:1876)

at java.io.ObjectOutputStream$BlockDataOutputStream.setBlockDataMode(ObjectOutputStream.java:1785)

at java.io.ObjectOutputStream.writeObject0(ObjectOutputStream.java:1188)

at java.io.ObjectOutputStream.writeObject(ObjectOutputStream.java:347)

at org.apache.spark.serializer.JavaSerializationStream.writeObject(JavaSerializer.scala:42)

at org.apache.spark.serializer.JavaSerializerInstance.serialize(JavaSerializer.scala:71)

at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:197)

at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecutor.java:1145)

at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecutor.java:615)

at java.lang.Thread.run(Thread.java:745)

15/03/04 13:49:33 INFO scheduler.TaskSetManager: Starting task 0.12:1 as TID 152 on executor

Hadoop Streaming jobs are failing…..

15/03/04 13:30:27 INFO mapreduce.Job: map 100% reduce 100%

15/03/04 13:30:35 INFO mapreduce.Job: Task Id : attempt\_1422482982071\_0471\_r\_000000\_1, Status : FAILED

Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 139

at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:320)

at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:533)

at org.apache.hadoop.streaming.PipeReducer.close(PipeReducer.java:134)

at org.apache.hadoop.io.IOUtils.cleanup(IOUtils.java:237)

at org.apache.hadoop.mapred.ReduceTask.runOldReducer(ReduceTask.java:459)

at org.apache.hadoop.mapred.ReduceTask.run(ReduceTask.java:392)

at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:167)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:415)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1554)

at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:162)

15/03/04 13:30:36 INFO mapreduce.Job: map 100% reduce 0%

15/03/04 13:30:48 INFO mapreduce.Job: map 100% reduce 67%

15/03/04 13:30:51 INFO mapreduce.Job: map 100% reduce 68%

15/03/04 13:30:57 INFO mapreduce.Job: map 100% reduce 6

13:33:06 INFO mapreduce.Job: map 100% reduce 98%

15/03/04 13:33:15 INFO mapreduce.Job: map 100% reduce 99%

15/03/04 13:33:24 INFO mapreduce.Job: map 100% reduce 100%

15/03/04 13:33:35 INFO mapreduce.Job: Task Id : attempt\_1422482982071\_0471\_r\_000000\_2, Status : FAILED

Error: java.lang.RuntimeException: PipeMapRed.waitOutputThreads(): subprocess failed with code 139

at org.apache.hadoop.streaming.PipeMapRed.waitOutputThreads(PipeMapRed.java:320)

at org.apache.hadoop.streaming.PipeMapRed.mapRedFinished(PipeMapRed.java:533)

at org.apache.hadoop.streaming.PipeReducer.close(PipeReducer.java:134)

at org.apache.hadoop.io.IOUtils.cleanup(IOUtils.java:237)

at org.apache.hadoop.mapred.ReduceTask.runOldReducer(ReduceTask.java:459)

at org.apache.hadoop.mapred.ReduceTask.run(ReduceTask.java:392)

at org.apache.hadoop.mapred.YarnChild$2.run(YarnChild.java:167)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:415)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1554)

at org.apache.hadoop.mapred.YarnChild.main(YarnChild.java:162)

15/03/04 13:36:00 INFO mapreduce.Job: map 100% reduce 97%

15/03/04 13:36:06 INFO mapreduce.Job: map 100% reduce 98%

15/03/04 13:36:15 INFO mapreduce.Job: map 100% reduce 99%

15/03/04 13:36:25 INFO mapreduce.Job: map 100% reduce 100%

15/03/04 13:36:37 INFO mapreduce.Job: Job job\_1422482982071\_0471 failed with state FAILED due to: Task failed task\_1422482982071\_0471\_r\_000000

Job failed as tasks failed. failedMaps:0 failedReduces:1

15/03/04 13:36:37 INFO mapreduce.Job: Counters: 37

File System Counters

FILE: Number of bytes read=0

FILE: Number of bytes written=68657043

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=66139604

HDFS: Number of bytes written=0

HDFS: Number of read operations=6

HDFS: Number of large read operations=0

HDFS: Number of write operations=0

Job Counters

Failed reduce tasks=4

Launched map tasks=2

Launched reduce tasks=4

Rack-local map tasks=2

Total time spent by all maps in occupied slots (ms)=213840

Total time spent by all reduces in occupied slots (ms)=1374096

Total time spent by all map tasks (ms)=106920

Total time spent by all reduce tasks (ms)=687048

Total vcore-seconds taken by all map tasks=106920

Total vcore-seconds taken by all reduce tasks=687048

Total megabyte-seconds taken by all map tasks=865624320

Total megabyte-seconds taken by all reduce tasks=8244576000

Map-Reduce Framework

Map input records=2205657

Map output records=3464806

Map output bytes=61537573

Map output materialized bytes=68467209

Input split bytes=276

Combine input records=0

Spilled Records=3464806

Failed Shuffles=0

Merged Map outputs=0

GC time elapsed (ms)=775

CPU time spent (ms)=125790

Physical memory (bytes) snapshot=3893424128

Virtual memory (bytes) snapshot=18483359744

Total committed heap usage (bytes)=5539454976

File Input Format Counters

Bytes Read=66139328

15/03/04 13:36:37 ERROR streaming.StreamJob: Job not Successful!

15/03/04 13:55:16 ERROR streaming.StreamJob: Job not Successful!

Streaming Command Failed!

FAILED

-rw-r--r-- 3 dotcz12 dotcz12 66136448 2015-02-25 12:36 enwiki-latest-stub-articles8.xml

hadoop jar $HADOOP\_STREAMING\_JAR -file ./mapper.R -mapper ./mapper.R

-file ./reducer.R -reducer ./reducer.R -input ./enwiki-latest-stub-articles8.xml -output ./output/streaming\_test

PASSED

hadoop jar $HADOOP\_STREAMING\_JAR -file ./mapper.R -mapper ./mapper.R -file ./reducer.R -reducer ./reducer.R -input book.txt -output ./output/streaming-book