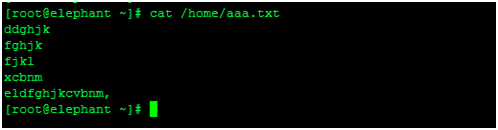
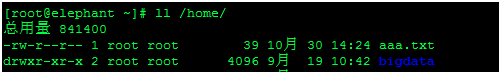
# MapReduce job实验

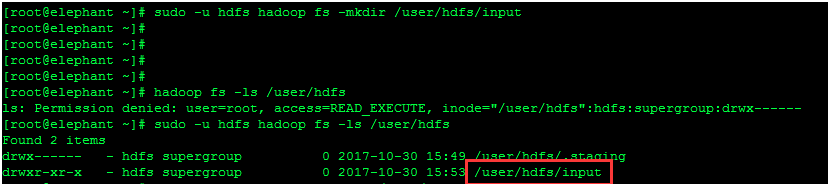
用MapReduce运行个简单的WordCount job

1. vi  /home/aaa.txt

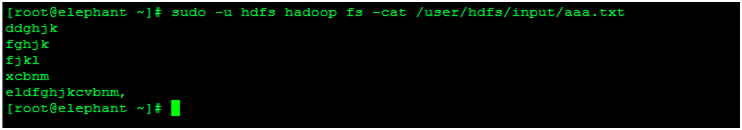
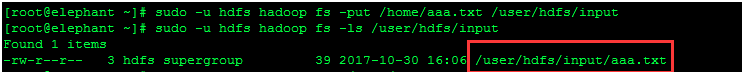
  cat /home/aaa.txt



2. sudo -u hdfs hadoop fs -mkdir /user/hdfs/input (在hdfs上创建input目录,用于存放需要统计的文件)



3. sudo -u hdfs hadoop fs -put /home/aaa.txt /user/hdfs/input（把home下的aaa文件上传到hdfs的input目录）



4. sudo -u hdfs hadoop jar /home/hadoop-mapreduce-examples.jar wordcount input counts  运行jar包

（hadoop-mapreduce-examples.jar包可以到hadoop官网下载，是一个单词统计包，感兴趣的小伙伴也可以自己写Python scala java脚本实现不同的功能，wordcount为作业名称，input为存放统计文件的目录，counts为mapreduce输入目录）

17/10/30 16:07:10 INFO client.RMProxy: Connecting to ResourceManager at elephant/192.168.10.128:8032

17/10/30 16:07:11 INFO input.FileInputFormat: Total input paths to process : 1

17/10/30 16:07:12 INFO mapreduce.JobSubmitter: number of splits:1

17/10/30 16:07:12 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1509333705246\_0002

17/10/30 16:07:14 INFO impl.YarnClientImpl: Submitted application application\_1509333705246\_0002

17/10/30 16:07:14 INFO mapreduce.Job: The url to track the job: http://elephant:8088/proxy/application\_1509333705246\_0002/

17/10/30 16:07:14 INFO mapreduce.Job: Running job: job\_1509333705246\_0002

17/10/30 16:09:05 INFO mapreduce.Job: Job job\_1509333705246\_0002 running in uber mode : false

17/10/30 16:09:05 INFO mapreduce.Job:  map 0% reduce 0%

17/10/30 16:13:57 INFO mapreduce.Job:  map 100% reduce 0%

17/10/30 16:15:33 INFO mapreduce.Job:  map 100% reduce 22%

17/10/30 16:15:47 INFO mapreduce.Job:  map 100% reduce 33%

17/10/30 16:15:59 INFO mapreduce.Job:  map 100% reduce 67%

17/10/30 16:17:13 INFO mapreduce.Job:  map 100% reduce 100%

17/10/30 16:17:17 INFO mapreduce.Job: Job job\_1509333705246\_0002 completed successfully（作业已经成功跑完）

17/10/30 16:17:22 INFO mapreduce.Job: Counters: 49

         File System Counters

                   FILE: Number of bytes read=124

                   FILE: Number of bytes written=588775

                   FILE: Number of read operations=0

                   FILE: Number of large read operations=0

                   FILE: Number of write operations=0

                   HDFS: Number of bytes read=147

                   HDFS: Number of bytes written=49

                   HDFS: Number of read operations=12

                   HDFS: Number of large read operations=0

                   HDFS: Number of write operations=6

         Job Counters

                   Launched map tasks=1

                   Launched reduce tasks=3

                   Data-local map tasks=1

                   Total time spent by all maps in occupied slots (ms)=274644

                   Total time spent by all reduces in occupied slots (ms)=243951

                   Total time spent by all map tasks (ms)=274644

                   Total time spent by all reduce tasks (ms)=243951

                   Total vcore-milliseconds taken by all map tasks=274644

                   Total vcore-milliseconds taken by all reduce tasks=243951

                   Total megabyte-milliseconds taken by all map tasks=281235456

                   Total megabyte-milliseconds taken by all reduce tasks=249805824

         Map-Reduce Framework

                   Map input records=5

                   Map output records=5

                   Map output bytes=59

                   Map output materialized bytes=112

                   Input split bytes=108

                   Combine input records=5

                   Combine output records=5

                   Reduce input groups=5

                   Reduce shuffle bytes=112

                   Reduce input records=5

                   Reduce output records=5

                   Spilled Records=10

                   Shuffled Maps =3

                   Failed Shuffles=0

                   Merged Map outputs=3

                   GC time elapsed (ms)=13313

                   CPU time spent (ms)=15470

                   Physical memory (bytes) snapshot=982937600

                   Virtual memory (bytes) snapshot=6274256896

                   Total committed heap usage (bytes)=765460480

         Shuffle Errors

                   BAD\_ID=0

                   CONNECTION=0

                   IO\_ERROR=0

                   WRONG\_LENGTH=0

                   WRONG\_MAP=0

                   WRONG\_REDUCE=0

         File Input Format Counters

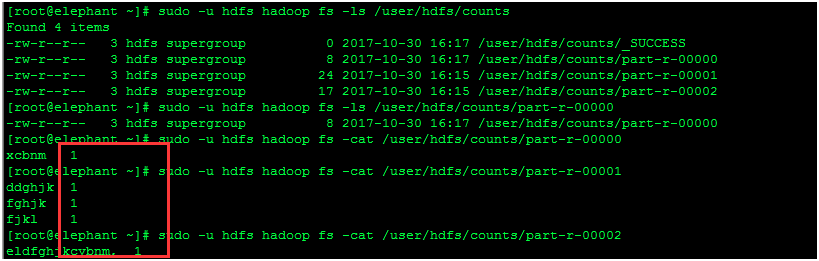
                   Bytes Read=39

         File Output Format Counters

                   Bytes Written=49

17/10/30 16:17:22 INFO mapred.ClientServiceDelegate: Application state is completed. FinalApplicationStatus=SUCCEEDED. Redirecting to job history server

5. sudo -u hdfs hadoop fs -ls /user/hdfs/counts（counts目录为上面跑mapreduce定义的输出目录，注意红框里面的1就是mapreduce统计出来的每行记录）



6.http://elephant:8088（通过ResourceManager Web UI 查看具体job）

