测试报告

2015 年 5 月

1. 编写说明

1.1. 标识

文档标题：Hadoop-MapReduce 测试报告

版本号：V2

1.2. 历史版本

编号	被修改版本	生成版本	修订人	修订章节	修改内容	修订日期
1	V0	V1	郑思文	全部	完成测试报告初稿	2015.5.19
2	V1	V2	郑思文	3.3 3.4 3.7	完善测试用例	2015.5.26

2. 实际测试用例对应表

本测试报告中已经完成的测试用例与计划的测试用例及需求用例的对应关系如表 2-1 所示。本报告仅完成了部分重要用例的测试，包括分配 Mapper 任务测试、分配 Reducer 任务测试、并行计算测试等，后续将会完善所有的测试用例。

表 2-1　实际测试用例与计划测试用例及需求用例的对应关系表

需求用例	测试用例	实际测试用例
配置作业信息	Config_test（配置测试）	√
提交作业	submit_test（作业提交测试）	√
杀死任务	killTask_test（杀死任务测试）	√
杀死作业	killJob_test（杀死作业测试）	√
处理任务	ExeTask_test（处理任务测试）	√
分配 Mapper 任务	AllocMap_test（分配 Mapper 任务测试）	√
分配 Reducer 任务	AllocRedu_test（分配 Reducer 任务测试）	√
无	parallel_test（并行计算测试）	√

3. 测试用例

3.1. 配置测试

3.1.1. 测试目标

测试目标：覆盖配置测试用例。

测试依据：需求规格说明书中配置作业信息规格说明、测试需求规格说明书中配置信息测试用例规格说明。

3.1.2. 测试用例分析

本测试用例主要实现的是配置测试，测试者实现作业相关接口和作业配置，并测试可能出现的中断错误，完成配置测试。

3.1.3. 测试内容 1 及结果

（1）目标

测试未对 MapReduce 作业分片信息进行配置时，系统的处理情况。

（2）测试脚本（详见 测试脚本/scripts/TestConfig1/）

TestConfig1.java，TestConfig1_Map.java，TestConfig1_Reduce.java

（3）操作过程

注释对分片信息的配置

```
//        TextInputFormat.setMinInputSplitSize(job, 1024L);
//        TextInputFormat.setMaxInputSplitSize(job, 1024*1024*40L);
```

（4）结果

```
2015-05-19 00:18:07,748 INFO [org.apache.hadoop.conf.Configuration.deprecation] - session.id is deprecated. Instead, use dfs.metrics.session-id
2015-05-19 00:18:07,752 INFO [org.apache.hadoop.metrics.jvm.JvmMetrics] - Initializing JVM Metrics with processName=JobTracker, sessionId=
2015-05-19 00:18:09,711 WARN [org.apache.hadoop.mapreduce.JobSubmitter] - No job jar file set.  User classes may not be found. See Job or Job#setJar(String).
2015-05-19 00:18:09,932 INFO [org.apache.hadoop.mapreduce.lib.input.FileInputFormat] - Total input paths to process : 1
2015-05-19 00:18:10,287 INFO [org.apache.hadoop.mapreduce.JobSubmitter] - number of splits:1
2015-05-19 00:18:10,766 INFO [org.apache.hadoop.mapreduce.JobSubmitter] - Submitting tokens for job: job_local1557306723_0001
```

（5）结果分析

结果显示，在未对 MapReduce 作业分片信息进行配置时，默认分片数为 1。

3.1.4. 测试内容 2 及结果

（1）目标

测试对 MapReduce 作业分片信息进行配置时，系统的处理情况。

（2）测试脚本（详见 测试脚本/scripts/TestConfig2/）

TestConfig2.java，TestConfig2_Map.java，TestConfig2_Reduce.java

（3）操作过程

配置信息如下：

```
TextInputFormat.setMinInputSplitSize(job, 1024L);
TextInputFormat.setMaxInputSplitSize(job, 1024*1024*40L);
```

（4）结果

```
2015-05-19 00:20:34,959 INFO [org.apache.hadoop.conf.Configuration.deprecation] - session.id is deprecated. Instead, use dfs.metrics.session-id
2015-05-19 00:20:34,961 INFO [org.apache.hadoop.metrics.jvm.JvmMetrics] - Initializing JVM Metrics with processName=JobTracker, sessionId=
2015-05-19 00:20:35,464 WARN [org.apache.hadoop.mapreduce.JobSubmitter] - No job jar file set.  User classes may not be found. See Job or Job#setJar(String).
2015-05-19 00:20:35,523 INFO [org.apache.hadoop.mapreduce.lib.input.FileInputFormat] - Total input paths to process : 1
2015-05-19 00:20:35,702 INFO [org.apache.hadoop.mapreduce.JobSubmitter] - number of splits:2
2015-05-19 00:20:35,974 INFO [org.apache.hadoop.mapreduce.JobSubmitter] - Submitting tokens for job: job_local50568606_0001
```

（5）结果分析

结果显示，设置最大分片大小为 40M，数据文件为 77.4M，进行分片之后，分片数为 2。

3.1.5. 测试内容 3 及结果

（1）目标

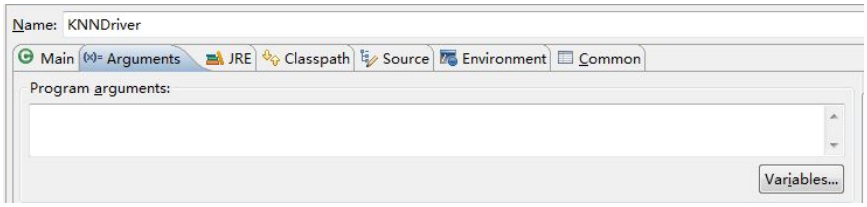
测试未对 MapReduce 作业输入输出路径进行配置时，系统的处理情况。

（2）测试脚本（详见 测试脚本/scripts/TestConfig3/）

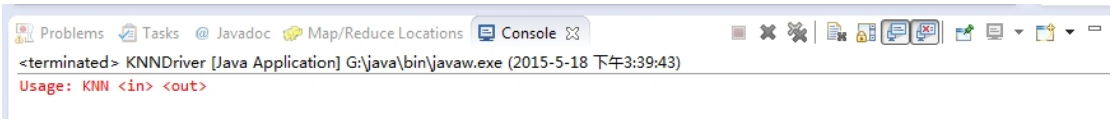
TestConfig3.java，TestConfig3_Map.java，TestConfig3_Reduce.java

（3）操作过程

这部分操作不能写进脚本，需要手动进行操作。如下图所示，没有配置输入输出路径。



（4）结果



（5）结果分析

结果显示错误信息。

3.1.6. 测试内容 4 及结果

（1）目标

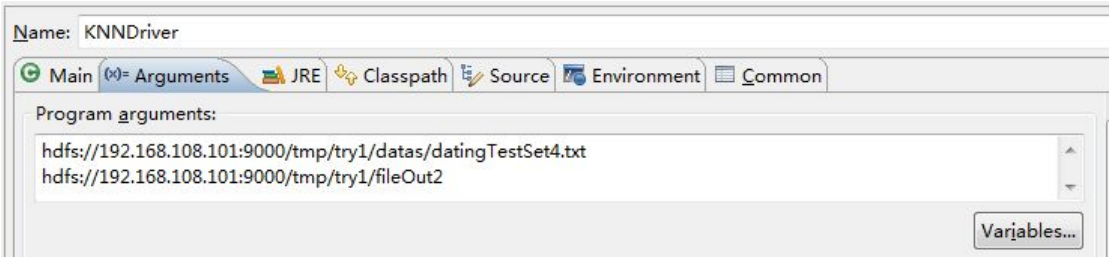
测试对 MapReduce 作业输入输出路径进行配置时，系统的处理情况。

（2）测试脚本（详见 测试脚本/scripts/TestConfig3/）

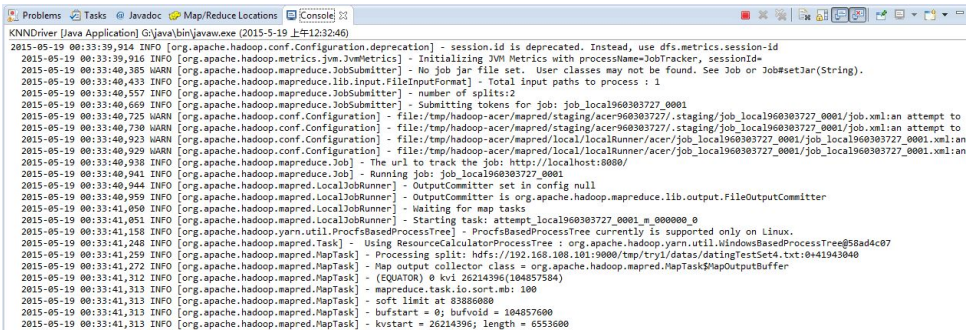
TestConfig3.java，TestConfig3_Map.java，TestConfig3_Reduce.java

（3）操作过程

这部分操作不能写进脚本，需要手动进行操作。如下图所示，进行配置。



（4）结果



（5）结果分析

结果显示，作业已被处理。

3.2. 提交作业测试

3.2.1. 测试目标

测试目标：覆盖提交作业测试用例。

测试依据：需求规格说明书中提交作业规格说明、测试需求规格说明书中的提交作业测试用例规格说明。

3.2.2. 测试用例分析

本测试用例主要实现的是提交作业测试，测试者编写 mapreduce 程序并打包成 jar 文件，通过 shell 命令提交作业，系统接收作业后存储到 HDFS，也需要测试可能出现的中断错误。

3.2.3. 测试内容及结果

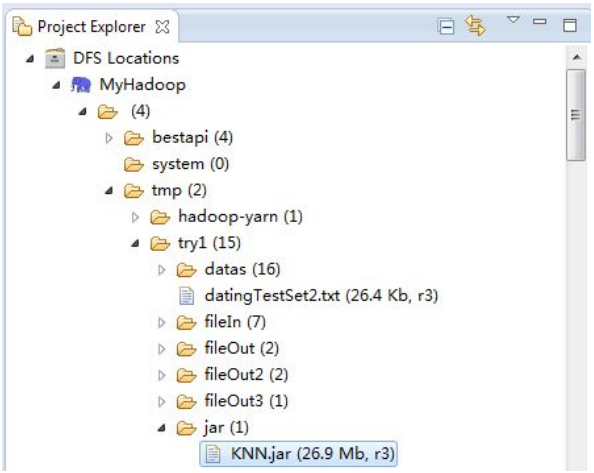
（1）目标

测试系统提交作业功能。

（2）测试脚本（详见 测试脚本/scripts/TestSubmit/）

TestSubmit.java，TestSubmit_Map.java，TestSubmit_Reduce.java

（3）结果



（4）结果分析

从截图可以看出，作业成功提交。

3.3. 杀死作业测试

3.3.1. 测试目标

测试目标：覆盖杀死作业测试用例。

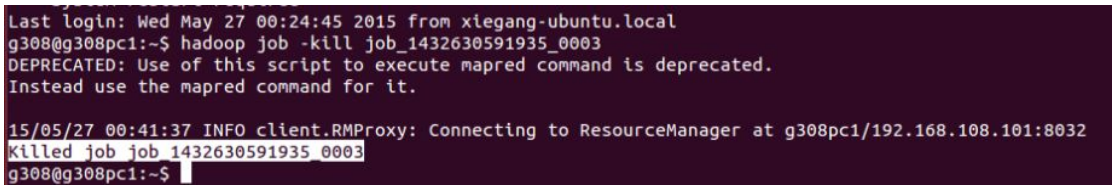
测试依据：需求规格说明书中杀死作业规格说明、测试需求规格说明书中的杀死作业测试用例规格说明。

3.3.2. 测试内容及结果

（1）目标

测试系统杀死作业的功能。

（2）结果



（3）结果分析

从截图可以看出，作业被成功杀死。

3.4. 杀死任务测试

3.4.1. 测试目标

测试目标：覆盖杀死任务测试用例。

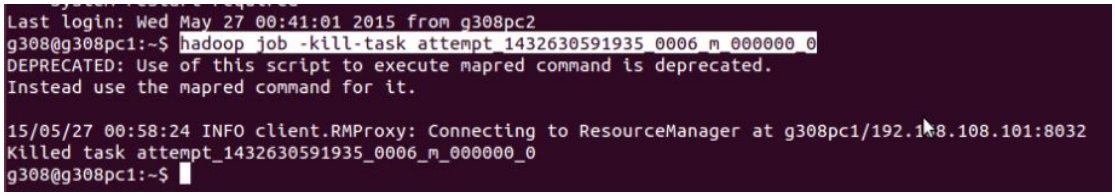
测试依据：需求规格说明书中杀死任务规格说明、测试需求规格说明书中的杀死任务测试用例规格说明。

3.4.2. 测试内容及结果

（1）目标

测试系统杀死任务的功能。

（2）结果



（3）结果分析

从截图可以看出，任务被成功杀死。

3.5. 系统执行 map 任务测试（包含分配 Mapper 任务测试）

3.5.1. 测试目标

测试目标：覆盖分配 Mapper 任务测试用例。

测试依据：需求规格说明书中分配 Mapper 任务规格说明、测试需求规格说明书中分配 Mapper 任务测试用例规格说明。

3.5.2. 测试用例分析

本测试用例主要实现的是分配 Mapper 任务测试，测试者实现作业提交后，系统经过一系列调用，将 Mapper 任务进行分配，并测试可能出现的中断错误。

3.5.3. 测试内容及结果

（1）目标

测试系统执行 map 任务功能。

（2）测试脚本（详见 测试脚本/scripts/TestMapper/）

TestMapper.java，TestMapper_Map.java，TestMapper_Reduce.java

（3）结果

分配 Mapper 任务结果：

```
2015-05-18 17:40:39,035 INFO [org.apache.hadoop.conf.Configuration.deprecation] - session.id is deprecated. Instead, use dfs.metrics.session-id
2015-05-18 17:40:39,039 INFO [org.apache.hadoop.metrics.jvm.JvmMetrics] - Initializing JVM Metrics with processName=JobTracker, sessionId=
2015-05-18 17:40:39,545 WARN [org.apache.hadoop.mapreduce.JobSubmitter] - No job jar file set.  User classes may not be found. See Job or Job#setJar(String).
2015-05-18 17:40:39,591 INFO [org.apache.hadoop.mapreduce.lib.input.FileInputFormat] - Total input paths to process : 1
2015-05-18 17:40:39,699 INFO [org.apache.hadoop.mapreduce.JobSubmitter] - number of splits:2
2015-05-18 17:40:39,810 INFO [org.apache.hadoop.mapreduce.JobSubmitter] - Submitting tokens for job: job_local2083213196_0001
2015-05-18 17:40:39,873 WARN [org.apache.hadoop.conf.Configuration] - file:/tmp/hadoop-acer/mapred/staging/acer2083213196/.staging/job_local2083213196_0001/job.xml:an attempt to
2015-05-18 17:40:39,878 WARN [org.apache.hadoop.conf.Configuration] - file:/tmp/hadoop-acer/mapred/staging/acer2083213196/.staging/job_local2083213196_0001/job.xml:an attempt to
2015-05-18 17:40:40,037 WARN [org.apache.hadoop.conf.Configuration] - file:/tmp/hadoop-acer/mapred/local/localRunner/acer/job_local2083213196_0001/job_local2083213196_0001.xml:ar
2015-05-18 17:40:40,041 WARN [org.apache.hadoop.conf.Configuration] - file:/tmp/hadoop-acer/mapred/local/localRunner/acer/job_local2083213196_0001/job_local2083213196_0001.xml:ar
2015-05-18 17:40:40,050 INFO [org.apache.hadoop.mapreduce.Job] - The url to track the job: http://localhost:8080/
2015-05-18 17:40:40,052 INFO [org.apache.hadoop.mapreduce.Job] - Running job: job_local2083213196_0001
2015-05-18 17:40:40,055 INFO [org.apache.hadoop.mapred.LocalJobRunner] - OutputCommitter set in config null
2015-05-18 17:40:40,066 INFO [org.apache.hadoop.mapred.LocalJobRunner] - OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2015-05-18 17:40:40,158 INFO [org.apache.hadoop.mapred.LocalJobRunner] - Waiting for map tasks
2015-05-18 17:40:40,159 INFO [org.apache.hadoop.mapred.LocalJobRunner] - Starting task: attempt_local2083213196_0001_m_000000_0
2015-05-18 17:40:40,253 INFO [org.apache.hadoop.yarn.util.ProcfsBasedProcessTree] - ProcfsBasedProcessTree currently is supported only on Linux.
2015-05-18 17:40:41,056 INFO [org.apache.hadoop.mapreduce.Job] - Job job_local2083213196_0001 running in uber mode : false
```

执行 Mapper 任务结果：

```
2015-05-18 17:40:50,064 INFO [org.apache.hadoop.mapreduce.Job] -  map 1% reduce 0%
2015-05-18 17:40:51,199 INFO [org.apache.hadoop.mapred.MapTask] - Finished spill 0
2015-05-18 17:40:51,200 INFO [org.apache.hadoop.mapred.MapTask] - (RESET) equator 69149575 kv 17287388(69149552) kvi 16059240(64236960)
2015-05-18 17:40:52,274 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map > map
2015-05-18 17:40:53,081 INFO [org.apache.hadoop.mapreduce.Job] -  map 2% reduce 0%
2015-05-18 17:40:53,249 INFO [org.apache.hadoop.mapred.MapTask] - Spilling map output
2015-05-18 17:40:53,249 INFO [org.apache.hadoop.mapred.MapTask] - bufstart = 69149575; bufend = 28507089; bufvoid = 104857600
2015-05-18 17:40:53,249 INFO [org.apache.hadoop.mapred.MapTask] - kvstart = 17287388(69149552); kvend = 12369640(49478560); length = 4917749/6553600
2015-05-18 17:40:53,250 INFO [org.apache.hadoop.mapred.MapTask] - (EQUATOR) 33441537 kvi 8360380(33441520)
2015-05-18 17:40:55,274 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map > map
2015-05-18 17:40:55,432 INFO [org.apache.hadoop.mapred.MapTask] - Finished spill 1
2015-05-18 17:40:55,433 INFO [org.apache.hadoop.mapred.MapTask] - (RESET) equator 33441537 kv 8360380(33441520) kvi 7132200(28528800)
2015-05-18 17:40:56,082 INFO [org.apache.hadoop.mapreduce.Job] -  map 3% reduce 0%
2015-05-18 17:40:57,440 INFO [org.apache.hadoop.mapred.MapTask] - Spilling map output
2015-05-18 17:40:57,440 INFO [org.apache.hadoop.mapred.MapTask] - bufstart = 33441537; bufend = 97656549; bufvoid = 104857600
2015-05-18 17:40:57,440 INFO [org.apache.hadoop.mapred.MapTask] - kvstart = 8360380(33441520); kvend = 3442612(13770448); length = 4917769/6553600
2015-05-18 17:40:57,440 INFO [org.apache.hadoop.mapred.MapTask] - (EQUATOR) 102591013 kvi 25647748(102590992)
2015-05-18 17:40:58,274 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map > map
2015-05-18 17:40:59,082 INFO [org.apache.hadoop.mapreduce.Job] -  map 4% reduce 0%
2015-05-18 17:40:59,530 INFO [org.apache.hadoop.mapred.MapTask] - Finished spill 2
2015-05-18 17:40:59,530 INFO [org.apache.hadoop.mapred.MapTask] - (RESET) equator 102591013 kv 25647748(102590992) kvi 24419604(97678416)
2015-05-18 17:41:01,275 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map > map
```

（4）结果分析

从截图可以看出，系统成功分配并执行 Mapper 任务，完成 map 环节。

3.6. 系统执行 Reduce 任务测试（包含分配 Reducer 任务测试）

3.6.1. 测试目标

测试目标：覆盖分配 Reducer 任务测试用例。

测试依据：需求规格说明书中分配 Reducer 任务规格说明、测试需求规格说明书中分配 Reducer 任务测试用例规格说明。

3.6.2. 测试用例分析

本测试用例主要实现的是分配 Reducer 任务测试，测试者实现作业提交后，系统经过一系列调用，将 Reducer 任务进行分配，并测试可能出现的中断错误。

3.6.3. 测试内容及结果

（1）目标

测试系统执行 reduce 任务功能。

（2）测试脚本（详见 测试脚本/scripts/TestReducer/）

TestReducer.java，TestReducer_Map.java，TestReducer_Reduce.java

（3）结果

分配 Reducer 任务结果：

```
2015-05-18 23:02:53,949 INFO [org.apache.hadoop.mapreduce.Job] -  map 100% reduce 0%
2015-05-18 23:02:56,730 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map > sort >
2015-05-18 23:02:57,012 INFO [org.apache.hadoop.mapred.Task] - Task:attempt_local1049132994_0001_m_000001_0 is done. And is in the process of committing
2015-05-18 23:02:57,024 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map > sort
2015-05-18 23:02:57,024 INFO [org.apache.hadoop.mapred.Task] - Task 'attempt_local1049132994_0001_m_000001_0' done.
2015-05-18 23:02:57,024 INFO [org.apache.hadoop.mapred.LocalJobRunner] - Finishing task: attempt_local1049132994_0001_m_000001_0
2015-05-18 23:02:57,024 INFO [org.apache.hadoop.mapred.LocalJobRunner] - map task executor complete.
2015-05-18 23:02:57,052 INFO [org.apache.hadoop.mapred.LocalJobRunner] - Waiting for reduce tasks
2015-05-18 23:02:57,052 INFO [org.apache.hadoop.mapred.LocalJobRunner] - Starting task: attempt_local1049132994_0001_r_000000_0
2015-05-18 23:02:57,089 INFO [org.apache.hadoop.yarn.util.ProcfsBasedProcessTree] - ProcfsBasedProcessTree currently is supported only on Linux.
2015-05-18 23:02:59,263 INFO [org.apache.hadoop.mapred.Task] -  Using ResourceCalculatorProcessTree : org.apache.hadoop.yarn.util.WindowsBasedProcessTree@7b1ae06d
2015-05-18 23:02:59,432 INFO [org.apache.hadoop.mapred.ReduceTask] - Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@968bafe
2015-05-18 23:02:59,499 INFO [org.apache.hadoop.mapreduce.task.reduce.MergeManagerImpl] - MergerManager: memoryLimit=614727680, maxSingleShuffleLimit=153681920, mergeThreshold=4
2015-05-18 23:02:59,506 INFO [org.apache.hadoop.mapreduce.task.reduce.EventFetcher] - attempt_local1049132994_0001_r_000000_0 Thread started: EventFetcher for fetching Map Comple
2015-05-18 23:02:59,595 INFO [org.apache.hadoop.mapreduce.task.reduce.MergeManagerImpl] - attempt_local1049132994_0001_m_000001_0: Shuffling to disk since 1573135873 is greater t
2015-05-18 23:02:59,627 INFO [org.apache.hadoop.mapreduce.task.reduce.LocalFetcher] - localfetcher#1 about to shuffle output of map attempt_local1049132994_0001_m_000001_0 decomp
2015-05-18 23:03:03,091 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > copy
2015-05-18 23:03:06,093 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > copy
2015-05-18 23:03:09,093 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > copy
2015-05-18 23:03:12,093 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > copy
```

执行 Reducer 任务结果：

```
2015-05-18 23:07:37,505 INFO [org.apache.hadoop.conf.Configuration.deprecation] - mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords
2015-05-18 23:07:39,280 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:07:40,009 INFO [org.apache.hadoop.mapreduce.Job] -  map 100% reduce 67%
2015-05-18 23:07:42,281 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:07:45,282 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:07:48,282 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:07:51,282 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:07:52,011 INFO [org.apache.hadoop.mapreduce.Job] -  map 100% reduce 68%
2015-05-18 23:07:54,282 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:07:55,016 INFO [org.apache.hadoop.mapreduce.Job] -  map 100% reduce 69%
2015-05-18 23:07:57,282 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:00,282 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:03,285 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:06,285 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:09,285 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:12,285 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:15,285 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:16,017 INFO [org.apache.hadoop.mapreduce.Job] -  map 100% reduce 70%
2015-05-18 23:08:18,285 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:21,286 INFO [org.apache.hadoop.mapred.LocalJobRunner] - reduce > reduce
2015-05-18 23:08:22,018 INFO [org.apache.hadoop.mapreduce.Job] -  map 100% reduce 71%
```

（4）结果分析

从截图可以看出，系统成功分配并执行 Reducer 任务，完成 reduce 环节。

3.7. 并行计算测试

3.7.1. 测试目标

测试目标：运用 KNN 算法，检验系统的并行计算能力。

测试依据：比较 MapReduce 的计算时间和单机上运行 KNN 算法的计算时间，来检验系统的并行计算能力。

3.7.2. 测试内容及结果

（1）目标

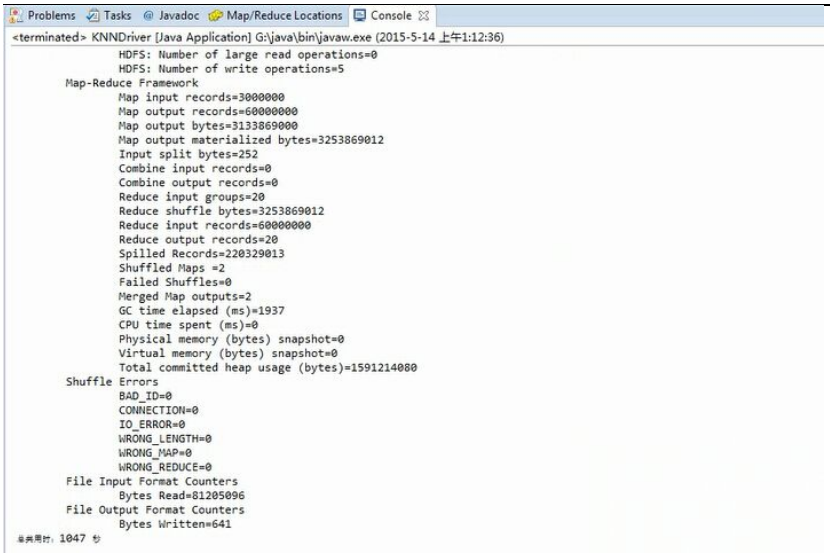
测试系统的并行计算能力（在两个从节点的情况下）。

（2）测试脚本（详见 测试脚本/scripts/TestParallel_MapReduce/和/scripts/TestParallel_local/）

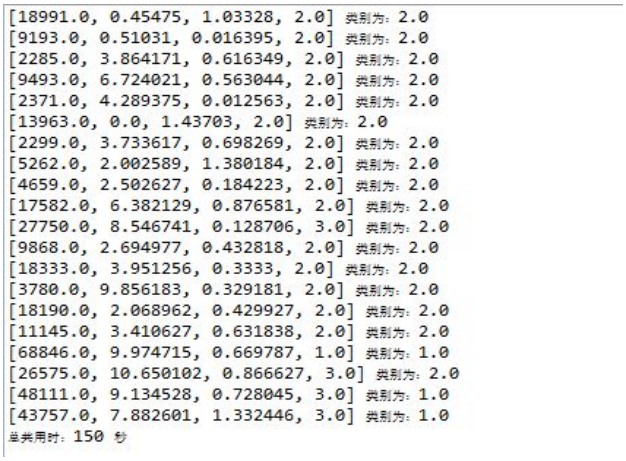
KNN_Map.java，KNN_Reduce.java，KNNDriver.java ；KNN.java，TestKNN.java

（3）结果

MapReduce 并行计算结果：



本地单机计算结果：



（4）结果分析

对比 MapReduce 和单机运行 KNN 算法给出的时间，可以发现在只有两个从节点的情况下，由于远程数据传输（实验室网速）等原因，并行结果很差，没有体现 MapReduce 的优势。但这并不意味着否定 MapReduce 的并行计算能力。下面通过改变数据集的大小，来进一步验证我们的猜测。

3.7.3. 改变测试集的数据大小

（1）数据集的大小为 242M 时

处理的结果中，耗时 401 秒。

```
<terminated> WordCountDriver [Java Application] G:\java\bin\javaw.exe (2015-5-26 下午10:59:04)
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=5
        Map-Reduce Framework
                Map input records=5975325
                Map output records=48280626
                Map output bytes=441457011
                Map output materialized bytes=538018275
                Input split bytes=250
                Combine input records=0
                Combine output records=0
                Reduce input groups=141
                Reduce shuffle bytes=538018275
                Reduce input records=48280626
                Reduce output records=141
                Spilled Records=144841878
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=1308
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=1069547520
        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=254313928
        File Output Format Counters
                Bytes Written=1746
单共用时：401 秒
```

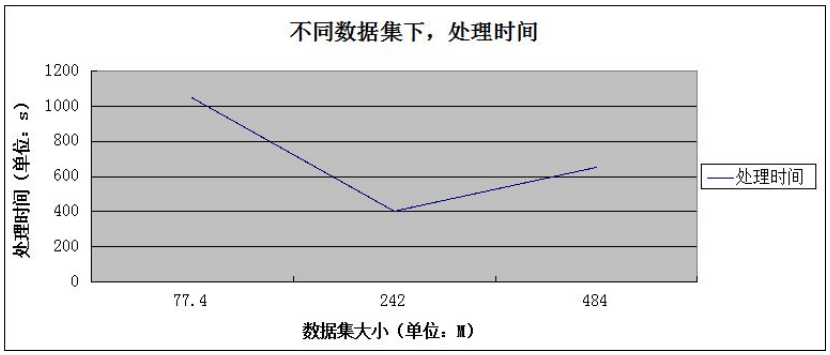
（2）数据集的大小为 484M 时

处理的结果中，耗时 653 秒。

```
<terminated> WordCountDriver [Java Application] G:\java\bin\javaw.exe (2015-5-26 下午11:21:47)
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=7
        Map-Reduce Framework
                Map input records=11950650
                Map output records=96561252
                Map output bytes=882914022
                Map output materialized bytes=1076036550
                Input split bytes=504
                Combine input records=0
                Combine output records=0
                Reduce input groups=141
                Reduce shuffle bytes=1076036550
                Reduce input records=96561252
                Reduce output records=141
                Spilled Records=289683756
                Shuffled Maps =4
                Failed Shuffles=0
                Merged Map outputs=4
                GC time elapsed (ms)=1992
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=1842348032
        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=508631952
        File Output Format Counters
                Bytes Written=1751
单共用时：653 秒
```

3.7.4. 结论

使用 KNN 算法，当数据集大小为 77.4M、242M、484M 时，所耗时间分别为 1047 秒、401 秒、653 秒。绘制折线图如下图所示。



从图可以看出，当数据集较小时，MapReduce 处理的时间反而更长，没有体现出其处理并发任务的能力。当增大数据集时，所需处理时间有所减短。可见，MapReduce 不适合处理较小的数据集。

以实验室现有条件难以从根本上改变 MapReduce 的并行计算能力。后期如果能够增加从节点的个数，提高网速，将有可能进一步提高 MapReduce 的并行计算能力。