

云计算实验三报告

--10215501435 杨茜雅

作业 1:

```
113.31.124.47:22
-bash: hadoop: command not found
[root@10-23-13-225 ~]#
Socket error Event: 32 Error: 10053.
Connection closing...Socket close.
Connection closed by foreign host.

Disconnected from remote host(113.31.103.133:22) at 22:32:51.

Type 'help' to learn how to use Xshell prompt.
[(:~)~]$ ssh root@113.31.124.47

Connecting to 113.31.124.47:22...
Connection established.
Warning: Permanently added host(113.31.124.47) to the list of known hosts.
root@113.31.124.47:~#
[WARNING] The remote SSH server rejected X11 forwarding request.
root@10-23-183-245 ~# hdfs dfsadmin -report
Configured Capacity: 633648111616 (590.13 GB)
Present Capacity: 601278963712 (555.90 GB)
DFS Remaining: 601278963712 (555.90 GB)
DFS Used: 86816 (84 KB)
DFS Used%: 0.00%
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0
Pending deletion blocks: 0

report: Access denied for user root. Superuser privilege is required
[root@10-23-183-245 ~]# ls
hadoop-2.6.0  hive  java
```

作业 2:

创建文件夹 test, 上传 info.txt, 使用 mkdir 和 put, 查看内容, 删除

```
113.31.124.47:22
Under replicated blocks: 0
Blocks with corrupt replicas: 0
Missing blocks: 0
Missing blocks (with replication factor 1): 0
Pending deletion blocks: 0

report: Access denied for user root. Superuser privilege is required
[root@10-23-183-245 ~]# ls
hadoop-2.6.0  hive  java
[root@10-23-183-245 ~]# hadoop fs -mkdir /test
[root@10-23-183-245 ~]# hadoop fs -put info.txt /test
put: 'info.txt': No such file or directory
[root@10-23-183-245 ~]# ls
hadoop-2.6.0  hive  java
[root@10-23-183-245 ~]# ls
hadoop-2.6.0  hive  info.txt  java
[root@10-23-183-245 ~]# hadoop fs -put /root/info.txt /test

-jt <local>[resourceManager:port] specify a ResourceManager.
-files <comma separated list of files> specify comma separated files to be copied to the map reduce cluster
-libjars <comma separated list of jars> specify comma separated jar files to include in the classpath.
-archives <comma separated list of archives> specify comma separated archive s to be unarchived on the compute machines.

The general command line syntax is
command [genericOptions] [commandOptions]

[root@10-23-183-245 ~]# hadoop fs -put info.txt /test
[root@10-23-183-245 ~]# hadoop fs -ls /test
Found 1 items
-rw-r--r-- 3 root supergroup 41 2023-10-08 23:00 /test/info.txt
[root@10-23-183-245 ~]# hadoop fs -cat /test/info.txt
云计算实验三:10215501435-杨茜雅 [root@10-23-183-245 ~]# hadoop fs -test -e info.txt
[root@10-23-183-245 ~]# hadoop fs -tail /test/info.txt
云计算实验三:10215501435-杨茜雅 [root@10-23-183-245 ~]# ls
hadoop-2.6.0  hive  info.txt  java
[root@10-23-183-245 ~]# hadoop fs -rm /test/info.txt
23/10/08 23:07:49 INFO fs.TrashPolicyDefault: Moved: 'hdfs://Ucluster/test/info.txt' to trash at: hdfs
~/Octastar/user/root/.Trash/Current/csc/info.txt
[root@10-23-183-245 ~]# ls
hadoop-2.6.0  hive  info.txt  java
[root@10-23-183-245 ~]# hadoop fs -ls
Found 1 items
drwx----- root root 0 2023-10-08 23:07 .Trash
[root@10-23-183-245 ~]# hadoop fs -test
-test: Not enough arguments: expected 1 but got 0
Usage: hadoop fs [generic options]
[-appendToFile <localSrc> ... <dst>]
[-cat [-ignoreCrc] <src> ...]
[-checksum <src> ...]
[-chgrp [-R] GROUP PATH...]
[-chmod [-R] <MODE[,MODE]...> [OCTALMODE] PATH...]
[-chown [-R] [OWNER][:(GROUP)] PATH...]
[-copyFromLocal [-f] [-p] [-l] [-d] <localSrc> ... <dst>]
[-copyToLocal [-f] [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
[-count [-q] [-h] [-v] [-t [<storage type>]] [-u] [-x] <path> ...]
```

作业 3:

```
1 113.31.124.47:22 +
The general command line syntax is
command [genericOptions] [commandOptions]

Usage: hadoop fs [-generic options] [-test [-defsz] <path>]
[root@10-23-183-245 ~]# hadoop fs -ls
Found 1 items
drwx----- root root 0 2023-10-08 23:07 .Trash
[root@10-23-183-245 ~]# cd /root/hadoop-2.6.0/share/hadoop/mapreduce
[root@10-23-183-245 mapreduce]# ls
hadoop-mapreduce-client-app-2.8.5.jar      hadoop-mapreduce-client-shuffle-2.8.5.jar
hadoop-mapreduce-client-common-2.8.5.jar  hadoop-mapreduce-examples-2.8.5.jar
hadoop-mapreduce-client-core-2.8.5.jar    idiff
hadoop-mapreduce-client-hs-2.8.5.jar      lib
hadoop-mapreduce-client-hs-plugins-2.8.5.jar lib-examples
hadoop-mapreduce-client-jobclient-2.8.5.jar sources
hadoop-mapreduce-client-jobclient-2.8.5-tests.jar
[root@10-23-183-245 mapreduce]# hadoop jar hadoop-mapreduce-examples-2.8.5.jar
An example program must be given as the first argument.
Valid program names are:
aggregatewordcount: An Aggregate based map/reduce program that counts the words in the input files.
aggregatewordhist: An Aggregate based map/reduce program that computes the histogram of the words in the input files.
bbp: A map/reduce program that uses Bailey-Borwein-Plouffe to compute exact digits of Pi.
dbcount: An example job that count the pageview counts from a database.
distbbp: A map/reduce program that uses a BBP-type formula to compute exact bits of Pi.
grep: A map/reduce program that counts the matches of a regex in the input.
join: A job that effects a join over sorted, equally partitioned datasets
multifilewc: A job that counts words from several files.
pentomino: A map/reduce tile laying program to find solutions to pentomino problems.
pi: A map/reduce program that estimates Pi using a quasi-Monte Carlo method.
randomtextwriter: A map/reduce program that writes 10GB of random textual data per node.
randomwriter: A map/reduce program that writes 10GB of random data per node.
secondarysort: An example defining a secondary sort to the reduce.
sort: A map/reduce program that sorts the data written by the random writer.
sudoku: A sudoku solver.
teragen: Generate data for the terasort
terasort: Run the terasort
teravalidate: Checking results of terasort
wordcount: A map/reduce program that counts the words in the input files.
wordmean: A map/reduce program that counts the average length of the words in the input files.
wordmedian: A map/reduce program that counts the median length of the words in the input files.
wordsstandarddeviation: A map/reduce program that counts the standard deviation of the length of the words in the input files.
[root@10-23-183-245 mapreduce]# hadoop fs -mkdir /input
[root@10-23-183-245 mapreduce]# hadoop fs -put /root/hadoop-2.6.0/etc/hadoop/input
put: /root/hadoop-2.6.0/etc/hadoop/input/: No such file or directory
[root@10-23-183-245 mapreduce]# hadoop fs -put /root/hadoop-2.6.0/etc/hadoop/* /input
[root@10-23-183-245 mapreduce]# hadoop jar /root/hadoop-2.6.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.5.jar wordcount /input /output
23/10/08 23:24:21 INFO client.AHSProxy: Connecting to Application History server at uhadop-os7h5mk1m6-master1:231
23/10/08 23:24:22 INFO InputFileInputFormat: Total input files to process : 33
23/10/08 23:24:22 INFO Lzo.GPLNativeCodeLoader: Loaded native gpl library from the embedded binaries
23/10/08 23:24:22 INFO Lzo.LzoCodec: Successfully loaded & initialized native-lzo library [hadoop-lzo
rw 52dec77982b58949896770d2720a91ad8c83f]
23/10/08 23:24:22 INFO mapreduce.JobSubmitter: number of splits:33
23/10/08 23:24:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1696774967540_0001
23/10/08 23:24:23 INFO Impl.YarnClientImpl: Submitted application application_1696774967540_0001
23/10/08 23:24:23 INFO mapreduce.Job: The url to track the job: http://uhadoop-os7h5mk1m6-master1:231
88/proxy/application_1696774967540_0001/
23/10/08 23:24:23 INFO mapreduce.Job: Running job: job_1696774967540_0001
23/10/08 23:24:28 INFO mapreduce.Job: Job job_1696774967540_0001 running in uber mode : false
23/10/08 23:24:28 INFO mapreduce.Job: map 0% reduce 0%
23/10/08 23:24:32 INFO mapreduce.Job: map 3% reduce 0%
23/10/08 23:24:34 INFO mapreduce.Job: map 15% reduce 0%
23/10/08 23:24:35 INFO mapreduce.Job: map 18% reduce 0%
23/10/08 23:24:38 INFO mapreduce.Job: map 33% reduce 0%
23/10/08 23:24:40 INFO mapreduce.Job: map 42% reduce 0%
23/10/08 23:24:41 INFO mapreduce.Job: map 48% reduce 0%
23/10/08 23:24:43 INFO mapreduce.Job: map 52% reduce 0%
23/10/08 23:24:44 INFO mapreduce.Job: map 64% reduce 0%
23/10/08 23:24:46 INFO mapreduce.Job: map 73% reduce 0%
23/10/08 23:24:47 INFO mapreduce.Job: map 76% reduce 0%
23/10/08 23:24:48 INFO mapreduce.Job: map 79% reduce 0%
23/10/08 23:24:49 INFO mapreduce.Job: map 88% reduce 0%
23/10/08 23:24:50 INFO mapreduce.Job: map 94% reduce 0%
23/10/08 23:24:52 INFO mapreduce.Job: map 100% reduce 0%
23/10/08 23:24:57 INFO mapreduce.Job: map 100% reduce 100%
23/10/08 23:24:57 INFO mapreduce.Job: Job job_1696774967540_0001 completed successfully
23/10/08 23:24:57 INFO mapreduce.Job: Counters: 50
File System Counters
FILE: Number of bytes read=91144
FILE: Number of bytes written=5941810
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=102489
HDFS: Number of bytes written=5391
HDFS: Number of read operations=102
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
Launched map tasks=33
Launched reduce tasks=1
Other local map tasks=2
Data-local map tasks=31
Total time spent by all maps in occupied slots (ms)=147050
Total time spent by all reduces in occupied slots (ms)=3670
Total time spent by all map tasks (ms)=73525
```

以下是部分输出截图，因为太长了

```
1 113.31.124.47:22 +
23/10/08 23:24:23 INFO mapreduce.Job: The url to track the job: http://uhadoop-os7h5mk1m6-master1:231
88/proxy/application_1696774967540_0001/
23/10/08 23:24:23 INFO mapreduce.Job: Running job: job_1696774967540_0001
23/10/08 23:24:28 INFO mapreduce.Job: Job job_1696774967540_0001 running in uber mode : false
23/10/08 23:24:32 INFO mapreduce.Job: map 0% reduce 0%
23/10/08 23:24:34 INFO mapreduce.Job: map 15% reduce 0%
23/10/08 23:24:35 INFO mapreduce.Job: map 18% reduce 0%
23/10/08 23:24:38 INFO mapreduce.Job: map 33% reduce 0%
23/10/08 23:24:40 INFO mapreduce.Job: map 42% reduce 0%
23/10/08 23:24:41 INFO mapreduce.Job: map 48% reduce 0%
23/10/08 23:24:43 INFO mapreduce.Job: map 52% reduce 0%
23/10/08 23:24:44 INFO mapreduce.Job: map 64% reduce 0%
23/10/08 23:24:46 INFO mapreduce.Job: map 73% reduce 0%
23/10/08 23:24:47 INFO mapreduce.Job: map 76% reduce 0%
23/10/08 23:24:48 INFO mapreduce.Job: map 79% reduce 0%
23/10/08 23:24:49 INFO mapreduce.Job: map 88% reduce 0%
23/10/08 23:24:50 INFO mapreduce.Job: map 94% reduce 0%
23/10/08 23:24:52 INFO mapreduce.Job: map 100% reduce 0%
23/10/08 23:24:57 INFO mapreduce.Job: map 100% reduce 100%
23/10/08 23:24:57 INFO mapreduce.Job: Job job_1696774967540_0001 completed successfully
23/10/08 23:24:57 INFO mapreduce.Job: Counters: 50
File System Counters
FILE: Number of bytes read=91144
FILE: Number of bytes written=5941810
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=102489
HDFS: Number of bytes written=5391
HDFS: Number of read operations=102
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
Job Counters
Launched map tasks=33
Launched reduce tasks=1
Other local map tasks=2
Data-local map tasks=31
Total time spent by all maps in occupied slots (ms)=147050
Total time spent by all reduces in occupied slots (ms)=3670
Total time spent by all map tasks (ms)=73525
```

```
113.31.124.47:22
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=9978
File Output Format Counters
  Bytes Written=53491
[root@10-23-183-245 mapreduce]# hadoop fs -cat /output/part-r-00000
--
  3
  4
  "SHADOOP_CLASSPATH" 1
  "SHADOOP_HEAPSIZE" 1
  "JAVA_HOME" 2
  "YARN_HEAPSIZE" 1
  "YARN_LOGFILE" 1
  "YARN_LOG_DIR" 1
  "YARN_POLICYFILE" 1
  "AS" 21
  "Error:" 1
  "License"); 21
  "alice,bob" 18
  "clustering" 1
  "console" 1
  "dfs" 3
  "hadoop.root.logger". 1
  "jks" 4
  "jvm" 3
  "mapred" 3
  "rpc" 3
  "run" 1
  "ugi" 3
  "x" 1
  "XJAVA_LIBRARY_PATH" 1
  # 394
  # //bin/bash 2
  ## 4
  #.sink.ganglia.dmax=jvm.metrics.threadsBlocked=70,jvm.metrics.memHeapUsedM=40 1
  #*.sink.ganglia.slope=jvm.metrics.gcCount=zero,jvm.metrics.memHeapUsedM=both 1
```

```
113.31.124.47:22
variable 4
variables 4
version 1
version="1.0" 7
version="1.0"> 1
version="1.0"> 6
via 3
view, 1
viewing 2
w/ 1
want 2
warnings 1
warnings, 1
weak 1
when 10
where 4
which 7
while 1
who 6
will 23
window 1
window, 1
with 57
within 4
without 1
work 11
writing, 21
written 2
xmlns:xsl="http://www.w3.org/1999/XSL/Transform" 1
yarn.emma.cleanupInterval=300 1
yarn.emma.maxUniqueness=250 1
yarn.emma.messageAgeLimitSeconds=86400 1
yarn.nodemanager.linux-container-executor.group 1
yarn.nodemanager.linux-container-executor.group=configured 1
yarn.server.resourcemanager.appsummary.log.file 1
yarn.server.resourcemanager.appsummary.log.file=rm-appsummary.log 1
yarn.server.resourcemanager.appsummary.logger 2
yarn.server.resourcemanager.appsummary.logger=${hadoop.root.logger} 1
you 26
zookeeper 1
[root@10-23-183-245 mapreduce]#
```

作业 4：实现单线程的 WordCount，记录运行时间

```
package wc;

import java.io.BufferedReader;
import java.io.File;
import java.io.FileReader;
import java.io.IOException;
import java.util.Iterator;
import java.util.Map;
import java.util.Set;
import java.util.TreeMap;

public class WordCountFolder {
    public static void main(String[] args) {
        long start = System.currentTimeMillis();
        Map<String, Integer> wordMap = new TreeMap<>();
        File folder = new File(pathname: "E:\\桌面\\mapreduce\\hadoop"); // 设置文件路径
        File[] files = folder.listFiles();
        if (files != null) {
            for (File file : files) {
                if (file.isFile()) {
                    processFile(file, wordMap);
                }
            }
        }
        long end = System.currentTimeMillis();
        Set<String> words = wordMap.keySet();
        Iterator<String> iterator = words.iterator();
        while (iterator.hasNext()) {
            String word = iterator.next();
            Integer count = wordMap.get(word);
            System.out.println("单词 " + word + ": " + count);
        }
        System.out.println("程序总用时: " + (end - start) + " 毫秒");
    }

    private static void processFile(File file, Map<String, Integer> wordMap) {
        try (BufferedReader br = new BufferedReader(new FileReader(file))) {
            String tempStr;
            while ((tempStr = br.readLine()) != null) {
                String[] split = tempStr.trim().split("\\s+");
                for (int i = 0; i < split.length; i++) {
                    String word = split[i];
                    wordMap.put(word, wordMap.getOrDefault(word, 0) + 1);
                }
            }
        } catch (IOException e) {
            e.printStackTrace();
        }
    }
}
```

结果输出（截取部分，因为太长了）：

```
WordCountFolder X
单词 <name>yarn.scheduler.capacity.root.default.state</name>: 1
单词 <name>yarn.scheduler.capacity.root.default.user-limit-factor</name>: 1
单词 <name>yarn.scheduler.capacity.root.queues</name>: 1
单词 <name>yarn.scheduler.fair.allow-undeclared-pools</name>: 1
单词 <name>yarn.scheduler.fair.user-as-default-queue</name>: 1
单词 <name>yarn.scheduler.maximum-allocation-mb</name>: 2
单词 <name>yarn.scheduler.maximum-allocation-vcores</name>: 1
单词 <name>yarn.timeline-service.address</name>: 1
单词 <name>yarn.timeline-service.enabled</name>: 1
单词 <name>yarn.timeline-service.generic-application-history.enabled</name>: 1
单词 <name>yarn.timeline-service.handler-thread-count</name>: 1
单词 <name>yarn.timeline-service.hostname</name>: 1
单词 <name>yarn.timeline-service.http-cross-origin.enabled</name>: 1
单词 <name>yarn.timeline-service.leveldb-state-store.path</name>: 1
单词 <name>yarn.timeline-service.leveldb-timeline-store.path</name>: 1
单词 <name>yarn.timeline-service.webapp.address</name>: 1
单词 <name>yarn.timeline-service.webapp.https.address</name>: 1
单词 <properties>: 2
单词 <property>: 2
单词 <property>: 201
单词 <queue>: 3
单词 <queues>: 1
单词 <state>running</state>: 1
单词 <table>: 1
单词 <td><a>: 1
单词 <td><xsl:value-of>: 2
单词 <td>description</td>: 1
```

```
WordCountFolder X
单词 <xsl:for-each>: 1
单词 <xsl:output>: 1
单词 <xsl:stylesheet>: 1
单词 <xsl:template>: 1
单词 =: 5
单词 @echo: 3
单词 @rem: 73
单词 A: 22
单词 ACL: 36
单词 ACL,: 2
单词 ACLs: 4
单词 ANY: 21
单词 ASF: 11
单词 AWS: 1
单词 AbhcYYwSdTgI7u+94zjilAtN3Gvj+dgvmUsqbDo4kGaR0FLD6oXwXg3zTgSAy+: 1
单词 Add: 1
单词 Admin: 2
单词 AdminOperationsProtocol.: 1
单词 Advanced: 1
单词 All: 3
单词 Apache: 33
单词 AppSummaryLogging: 1
单词 Appender: 8
单词 Application: 2
单词 ApplicationClientProtocol,: 1
单词 ApplicationHistoryProtocol,: 1
单词 ApplicationMaster: 1
```



```
1 113.31.124.47:22 +
variable 4
variables 4
version 1
version="1.0" 7
version="1.0"> 1
version="1.0"> 6
via 3
view, 1
viewing 2
w/ 1
want 2
warnings 1
warnings, 1
weak 1
when 10
where 4
which 7
while 1
who 6
will 23
window 1
window, 1
with 57
within 4
without 1
work 11
writing, 21
written 2
xmlns:xsl="http://www.w3.org/1999/XSL/Transform" 1
yarn.emma.cleanupInterval=300 1
yarn.emma.maxUniqueMessages=250 1
yarn.emma.messageAgeLimitSeconds=86400 1
yarn.nodemanager.linux-container-executor.group 1
yarn.nodemanager.linux-container-executor.group#configured 1
yarn.server.resourcemanager.appsummary.log.file 1
yarn.server.resourcemanager.appsummary.log.file=rm-appsummary.log 1
yarn.server.resourcemanager.appsummary.logger 2
yarn.server.resourcemanager.appsummary.logger=${hadoop.root.logger} 1
you 26
zookeeper 1
[root@10-23-183-245 mapreduce]#
```