

《并行计算》上机报告

姓名:	李子旻	学号:	PB16110959	日期:	2019.6.2
上机题目:	MapReduce 实验				
实验环境: CPU: Intel core i5 8250 ; 内存: 8GB ;操作系统: Ubuntu 18.04 ;软件平台: vim ;					
一、算法设计与分析:					
题目:					
<p>1. 按照 Hadoop 安装运行说明文档中的指导自己搭建伪分布式 Hadoop 环境,熟悉 HDFS 的常用操作(参考 Hadoop 实战 第 31-36 页),运行 WordCount 程序,得到统计结果。请详细写出你每一步的操作,最好有截图,最后的结果部分必须有截图。</p> <p>2. 实现一个统计输入文件中各个长度的单词出现频次的程序</p>					
二、过程截图:					
jdk 安装成功					
 <pre>\$ java -version java version "1.7.0_09" Java(TM) SE Runtime Environment (build 1.7.0_09-b05) Java HotSpot(TM) Server VM (build 23.5-b02, mixed mode)</pre>					
ssh 配置完成					
 <pre>ssh localhost Welcome to Ubuntu 18.04.2 LTS (GNU/Linux 4.15.0-51-generic x86_64) * Documentation: https://help.ubuntu.com * Management: https://landscape.canonical.com * Support: https://ubuntu.com/advantage 1.并行域并行化 4.使用并行规约 均匀划分->局部排序->选取样本->样本排序->选择主元->主元划分->全局交换- snap install microk8s --classic * Canonical Livepatch is available for installation. - Reduce system reboots and improve kernel security. Activate at: https://ubuntu.com/livepatch 13 packages can be updated. 1 update is a security update. *** System restart required *** Last login: Fri Jun 7 19:39:34 2019 from 127.0.0.1</pre>					
启动 hadoop					

```

$ sh start-all.sh
Warning: $HADOOP_HOME is deprecated.

starting namenode, logging to /home/foenix/temp/hadoop/hadoop-1.0.4/libexec/../logs/hadoop-foenix-namenode-foenix.out
localhost: starting datanode, logging to /home/foenix/temp/hadoop/hadoop-1.0.4/libexec/../logs/hadoop-foenix-datanode-foenix.out
localhost: datanode running as process 1680. Stop it first.
localhost: starting secondarynamenode, logging to /home/foenix/temp/hadoop/hadoop-1.0.4/libexec/../logs/hadoop-foenix-secondarynamenode-foenix.out
localhost: starting secondarynamenode, logging to /home/foenix/temp/hadoop/hadoop-1.0.4/libexec/../logs/hadoop-foenix-secondarynamenode-foenix.out
localhost: Exception in thread "main" java.net.BindException: 地址已在使用
localhost: at sun.nio.ch.Net.bind0(Native Method)
localhost: at sun.nio.ch.Net.bind(Net.java:344)
localhost: at sun.nio.ch.Net.bind(Net.java:336)
localhost: at sun.nio.ch.ServerSocketChannelImpl.bind(ServerSocketChannelImpl.java:199)
localhost: at sun.nio.ch.ServerSocketAdaptor.bind(ServerSocketAdaptor.java:74)
localhost: at org.mortbay.jetty.nio.SelectChannelConnector.open(SelectChannelConnector.java:216)
localhost: at org.apache.hadoop.http.HttpServer.start(HttpServer.java:581)
localhost: at org.apache.hadoop.hdfs.server.namenode.SecondaryNameNode$1.run(SecondaryNameNode.java:216)
localhost: at org.apache.hadoop.hdfs.server.namenode.SecondaryNameNode$1.run(SecondaryNameNode.java:190)
localhost: Exception in thread "main" java.net.BindException: 地址已在使用
localhost: at sun.nio.ch.Net.bind0(Native Method)
localhost: at sun.nio.ch.Net.bind(Net.java:344)
localhost: at sun.nio.ch.Net.bind(Net.java:336)
localhost: at sun.nio.ch.ServerSocketChannelImpl.bind(ServerSocketChannelImpl.java:199)
localhost: at sun.nio.ch.ServerSocketAdaptor.bind(ServerSocketAdaptor.java:74)
localhost: at org.mortbay.jetty.nio.SelectChannelConnector.open(SelectChannelConnector.java:216)
localhost: at org.apache.hadoop.http.HttpServer.start(HttpServer.java:581)
localhost: at org.apache.hadoop.hdfs.server.namenode.SecondaryNameNode$1.run(SecondaryNameNode.java:216)
localhost: at org.apache.hadoop.hdfs.server.namenode.SecondaryNameNode$1.run(SecondaryNameNode.java:190)
localhost: starting jobtracker, logging to /home/foenix/temp/hadoop/hadoop-1.0.4/libexec/../logs/hadoop-foenix-jobtracker-foenix.out
localhost: starting tasktracker, logging to /home/foenix/temp/hadoop/hadoop-1.0.4/libexec/../logs/hadoop-foenix-tasktracker-foenix.out
localhost: tasktracker running as process 2721. Stop it first.

```

成功启动

```

$ jps
2442 JobTracker
1238 NameNode
1680 DataNode
19340 SecondaryNameNode
2721 TaskTracker
2804 Jps

```

添加输入文件

```
$ hadoop fs -lsr /
Warning: $HADOOP_HOME is deprecated.
drwxr-xr-x - foenix supergroup 0 2019-06-07 15:24 /tmp
drwxr-xr-x - foenix supergroup 0 2019-06-07 15:24 /tmp/hadoop-foenix
drwxr-xr-x - foenix supergroup 0 2019-06-07 19:52 /tmp/hadoop-foenix/mapred
drwxr-xr-x - foenix supergroup 0 2019-06-07 19:33 /tmp/hadoop-foenix/mapred/stag
ing
drwxr-xr-x - foenix supergroup 0 2019-06-07 19:33 /tmp/hadoop-foenix/mapred/stag
ing/foenix
drwx----- - foenix supergroup 0 2019-06-07 19:34 /tmp/hadoop-foenix/mapred/stag
ing/foenix/.staging
drwx----- - foenix supergroup 0 2019-06-07 19:52 /tmp/hadoop-foenix/mapred/syst
em
-rw----- 1 foenix supergroup 4 2019-06-07 19:52 /tmp/hadoop-foenix/mapred/syst
em/jobtracker.info
drwxr-xr-x - foenix supergroup 0 2019-06-07 19:01 /user
drwxr-xr-x - foenix supergroup 0 2019-06-07 19:54 /user/foenix
drwxr-xr-x - foenix supergroup 0 2019-06-07 20:04 /user/foenix/wordcount
-rw-r--r-- 1 foenix supergroup 3303 2019-06-07 20:02 /user/foenix/wordcount/WordCou
nt.jar
drwxr-xr-x - foenix supergroup 0 2019-06-07 20:04 /user/foenix/wordcount/input
-rw-r--r-- 1 foenix supergroup 29 2019-06-07 20:04 /user/foenix/wordcount/input/i
nput1.txt
-rw-r--r-- 1 foenix supergroup 29 2019-06-07 20:04 /user/foenix/wordcount/input/i
nput2.txt
```

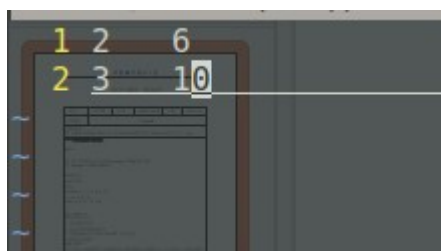
打包 wordcount.jar

```
$ ls
classes WordCount.jar WordCount.java
```

运行 wordcount.jar

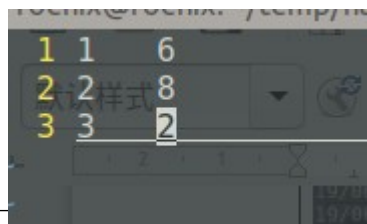

```
Warning: $HADOOP_HOME is deprecated.
19/06/07 20:06:06 INFO input.FileInputFormat: Total input paths to process : 2
19/06/07 20:06:06 INFO util.NativeCodeLoader: Loaded the native-hadoop library
19/06/07 20:06:06 WARN snappy.LoadSnappy: Snappy native library not loaded
19/06/07 20:06:06 INFO mapred.JobClient: Running job:/job_201906071951_0001
19/06/07 20:06:07 INFO mapred.JobClient: map 0% reduce 0%
19/06/07 20:06:20 INFO mapred.JobClient: map 100% reduce 0%
19/06/07 20:06:32 INFO mapred.JobClient: map 100% reduce 100%
19/06/07 20:06:37 INFO mapred.JobClient: Job complete: /job_201906071951_0001
19/06/07 20:06:37 INFO mapred.JobClient: Counters: 29
19/06/07 20:06:37 INFO mapred.JobClient: Job CountersOpenMP
19/06/07 20:06:37 INFO mapred.JobClient: Launched reduce tasks=1
19/06/07 20:06:37 INFO mapred.JobClient: SLOTS_MILLIS MAPS=16090
19/06/07 20:06:37 INFO mapred.JobClient: Total time spent by all reduces waiting after
reserving slots (ms)=0
19/06/07 20:06:37 INFO mapred.JobClient: Total time spent by all maps waiting after res
erving slots (ms)=0
19/06/07 20:06:37 INFO mapred.JobClient: Launched map tasks=2
19/06/07 20:06:37 INFO mapred.JobClient: Data-local map tasks=2
19/06/07 20:06:37 INFO mapred.JobClient: SLOTS_MILLIS REDUCES=10003
19/06/07 20:06:37 INFO mapred.JobClient: File Output Format Counters
19/06/07 20:06:37 INFO mapred.JobClient: Bytes Written=9
19/06/07 20:06:37 INFO mapred.JobClient: FileSystemCounters
19/06/07 20:06:37 INFO mapred.JobClient: FILE_BYTES_READ=38
19/06/07 20:06:37 INFO mapred.JobClient: HDFS_BYTES_READ=308
19/06/07 20:06:37 INFO mapred.JobClient: FILE_BYTES_WRITTEN=64689
19/06/07 20:06:37 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=9
19/06/07 20:06:37 INFO mapred.JobClient: File Input Format Counters
19/06/07 20:06:37 INFO mapred.JobClient: Bytes Read=58
19/06/07 20:06:37 INFO mapred.JobClient: Map-Reduce Framework
19/06/07 20:06:37 INFO mapred.JobClient: Map output materialized bytes=44
19/06/07 20:06:37 INFO mapred.JobClient: Map input records=4
19/06/07 20:06:37 INFO mapred.JobClient: Reduce shuffle bytes=22
19/06/07 20:06:37 INFO mapred.JobClient: Spilled Records=8
19/06/07 20:06:37 INFO mapred.JobClient: Map output bytes=96
19/06/07 20:06:37 INFO mapred.JobClient: Total committed heap usage (bytes)=416940032
19/06/07 20:06:37 INFO mapred.JobClient: CPU time spent (ms)=3800
19/06/07 20:06:37 INFO mapred.JobClient: Combine input records=16
19/06/07 20:06:37 INFO mapred.JobClient: SPLIT_RAW_BYTES=250
19/06/07 20:06:37 INFO mapred.JobClient: Reduce input records=4
19/06/07 20:06:37 INFO mapred.JobClient: Reduce input groups=2
19/06/07 20:06:37 INFO mapred.JobClient: Combine output records=4
19/06/07 20:06:37 INFO mapred.JobClient: Physical memory (bytes) snapshot=430219264
19/06/07 20:06:37 INFO mapred.JobClient: Reduce output records=2
19/06/07 20:06:37 INFO mapred.JobClient: Virtual memory (bytes) snapshot=1294180352
19/06/07 20:06:37 INFO mapred.JobClient: Map output records=16
```

获得运行结果



2.

对新的输入数据
结果是:



总结:	
附录 (源代码)	算法源代码 (C/C++/JAVA 描述)

--	--