hadoop 集群初体验

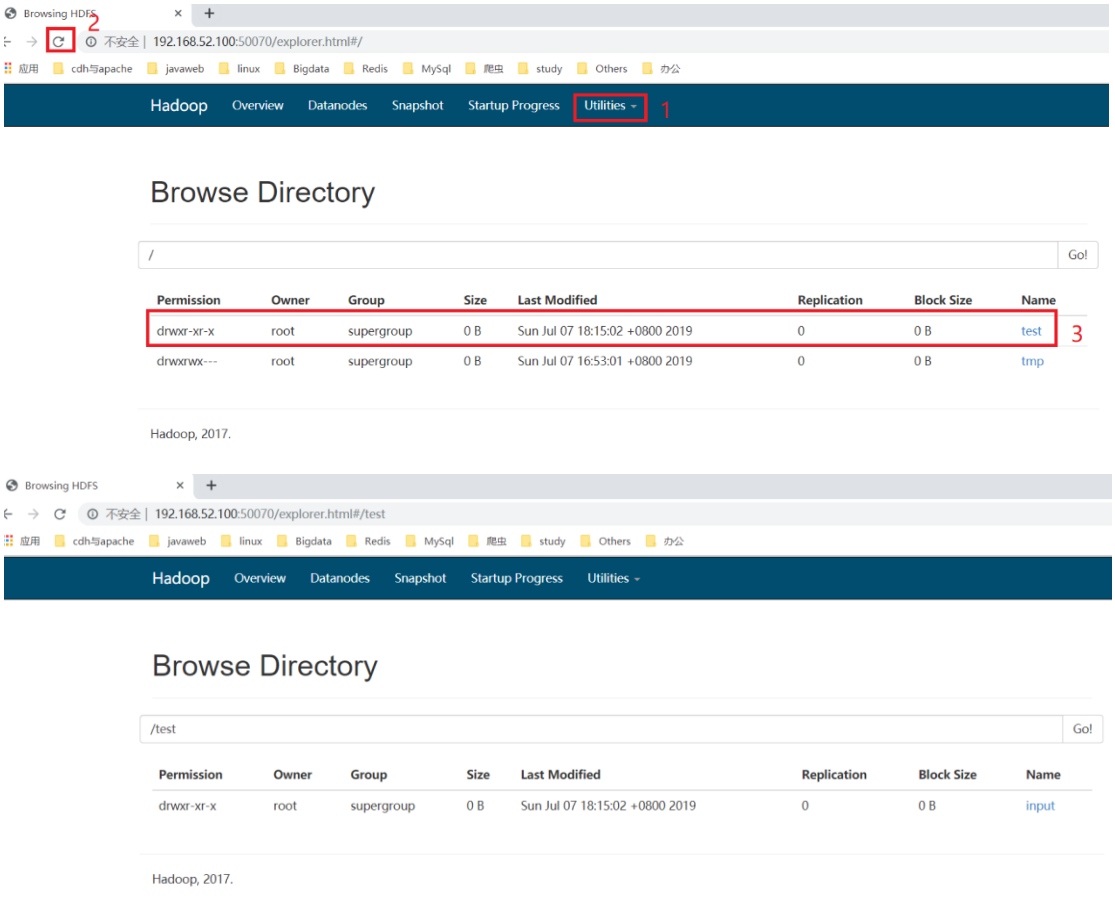
1、HDFS 使用体验

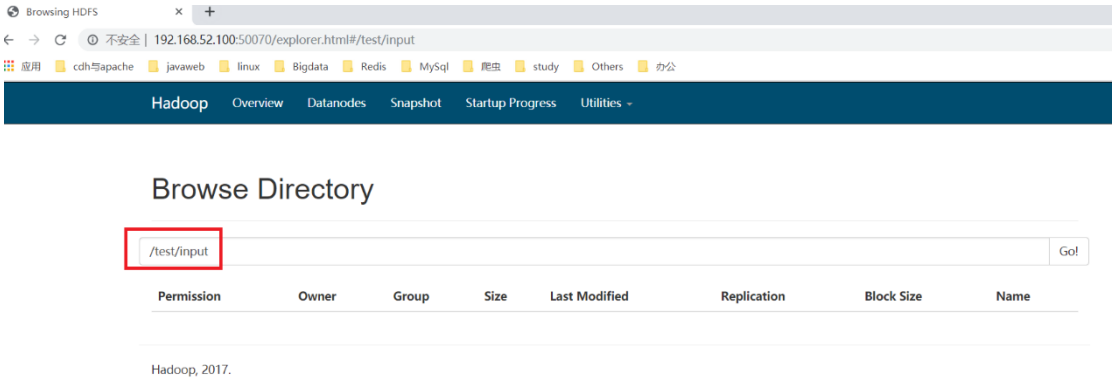
从 Linux 本地上传一个文本文件到 hdfs 的/test/input 目录下

递归的创建文件夹：

```
hdfs dfs -mkdir -p /test/input
```


```
hdfs dfs -ls /
```



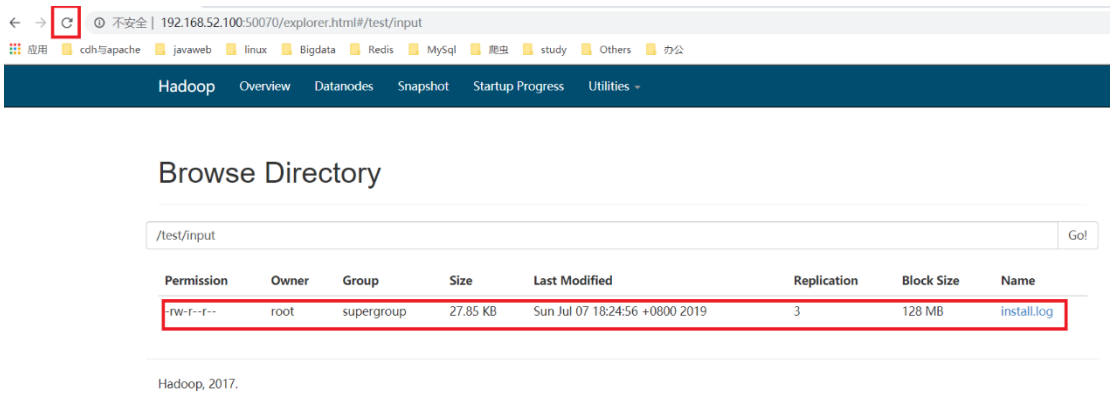



```
[root@node01 /]# cd ~
[root@node01 ~]# ll
total 44
-rw-------. 1 root root  1429 Jul 15 21:18 anaconda-ks.cfg
-rw-r--r--. 1 root root 28523 Jul 15 21:18 install.log
-rw-r--r--. 1 root root  7570 Jul 15 21:16 install.log.syslog
You have new mail in /var/spool/mail/root
[root@node01 ~]#
```

分布式文件系统来源于本地磁盘

```
hdfs dfs -put /root/install.log  /test/input
```

```
[root@node01 ~]# hdfs dfs -put /root/install.log  /test/input
You have new mail in /var/spool/mail/root
[root@node01 ~]# hdfs dfs -ls /
Found 2 items
drwxr-xr-x   - root supergroup          0 2019-07-07 18:15 /test
drwxrwx---   - root supergroup          0 2019-07-07 16:53 /tmp
[root@node01 ~]#
```



2、mapreduce 程序初体验

在 Hadoop 安装包的

hadoop-2.6.0-cdh5.14.0/share/hadoop/mapreduce 下有官方自带的

mapreduce 程序。我们可以使用如下的命令进行运行测试。

示例程序 jar:

hadoop-mapreduce-examples-2.6.0-cdh5.14.0.jar

计算圆周率:

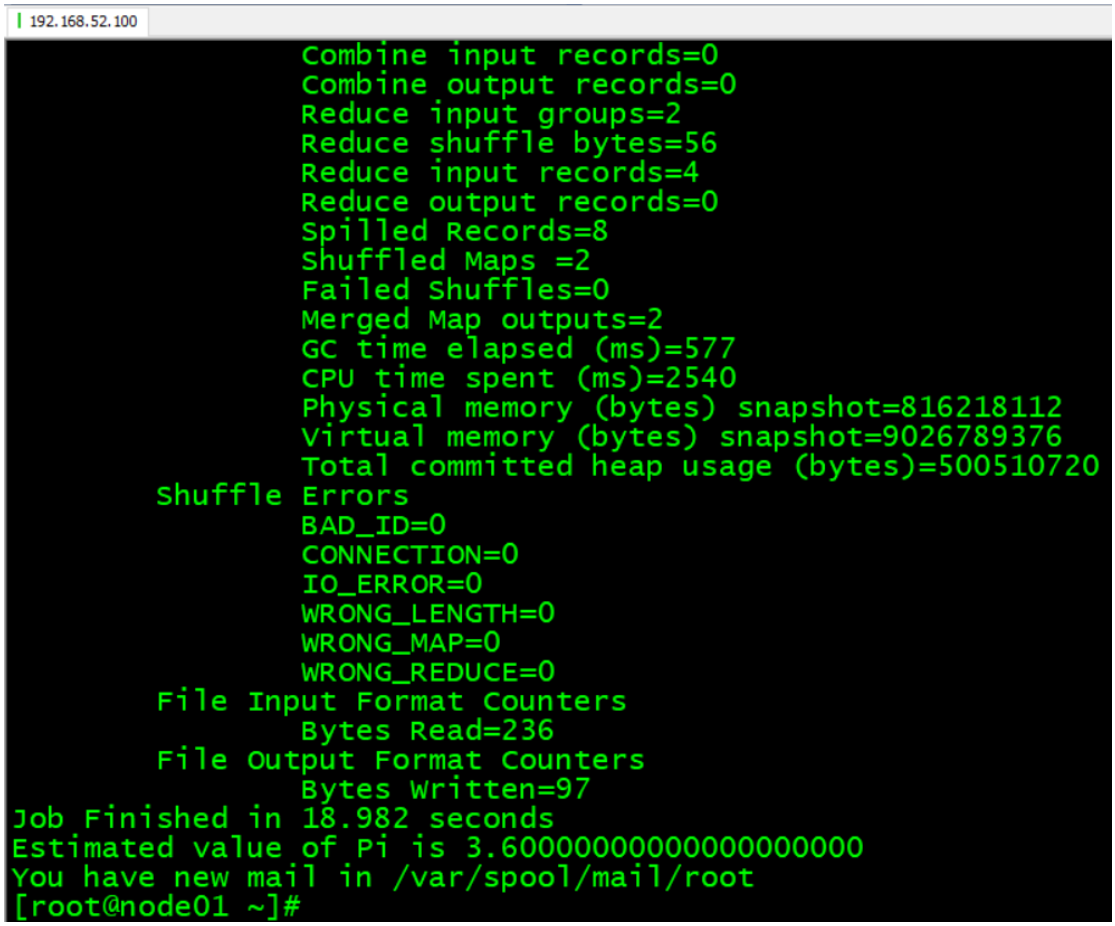
```
hadoop jar /export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.14.0.jar pi 2 5
```

关于圆周率的估算，感兴趣的可以查询资料 Monte Carlo 方法来计算 Pi 值。

```
[root@node01 ~]# clear

[root@node01 ~]# hadoop jar /export/servers/hadoop-2.6.0-cdh5.14.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0-cdh5.14.0.jar pi 2 5
Number of Maps  = 2             2启动2个maptask  5启动5个reducetask
Samples per Map = 5
Wrote input for Map #0
19/07/07 18:30:25 WARN hdfs.DFSClient: Caught exception
java.lang.InterruptedException
        at java.lang.Object.wait(Native Method)
        at java.lang.Thread.join(Thread.java:1252)
        at java.lang.Thread.join(Thread.java:1326)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.closeResponder(DFSOutputStream.java:967)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.endBlock(DFSOutputStream.java:705)
        at org.apache.hadoop.hdfs.DFSOutputStream$DataStreamer.run(DFSOutputStream.java:894)
Wrote input for Map #1
Starting Job
19/07/07 18:30:25 INFO client.RMProxy: Connecting to ResourceManager at node01/192.168.52.100:8032
19/07/07 18:30:26 INFO input.FileInputFormat: Total input paths to process : 2
19/07/07 18:30:26 INFO mapreduce.JobSubmitter: number of splits:2
19/07/07 18:30:26 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1562489571734_0001
19/07/07 18:30:27 INFO impl.YarnClientImpl: Submitted application application_1562489571734_0001
19/07/07 18:30:27 INFO mapreduce.Job: The url to track the job: http://node01:8088/proxy/application_1562489571734_0001/
19/07/07 18:30:27 INFO mapreduce.Job: Running job: job_1562489571734_0001
19/07/07 18:30:39 INFO mapreduce.Job: Job job_1562489571734_0001 running in uber mode : true
19/07/07 18:30:39 INFO mapreduce.Job:  map 0% reduce 0%
19/07/07 18:30:41 INFO mapreduce.Job:  map 100% reduce 0%
19/07/07 18:30:43 INFO mapreduce.Job:  map 100% reduce 100%
19/07/07 18:30:43 INFO mapreduce.Job: Job job_1562489571734_0001 completed successfully
```

底层日志线程异常，不用管!


192.168.52.100
Combine input records=0
Combine output records=0
Reduce input groups=2
Reduce shuffle bytes=56
Reduce input records=4
Reduce output records=0
Spilled Records=8
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=577
CPU time spent (ms)=2540
Physical memory (bytes) snapshot=816218112
Virtual memory (bytes) snapshot=9026789376
Total committed heap usage (bytes)=500510720
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=236
File Output Format Counters
Bytes Written=97
Job Finished in 18.982 seconds
Estimated value of Pi is 3.60000000000000000000
You have new mail in /var/spool/mail/root
[root@node01 ~]#