

南昌大学实验报告

姓名：陈华豪

学号：6130116238

邮箱地址：6130116238@email.ncu.edu.cn

专业班级：网络工程161班

实验日期：2019.05.19

课程名称：云计算

实验项目名称

Lab 7 Introduction to Cloud Computing-- Hadoop and MapReduce

实验目的

• Understanding the concept of MapReduce Model • Building a file system in a distributed way that store large data trunks • Merging all things together into a docker image • Writing some sample demo on running MapReduce models

实验基础

• Read the Hadoop wiki page and learn it • <https://wiki.apache.org/hadoop> • Install a Hadoop Docker image, you know where to find it • Finish the report on your experience of realizing two classic demos • Word Count: https://hadoop.apache.org/docs/stable/hadoop-mapreduceclient/hadoop-mapreduce-clientcore/MapReduceTutorial.html#Example:_WordCount_v1.0 • Quantity Analysis: <https://docs.mongodb.com/manual/tutorial/map-reduceexamples/>

实验步骤

1. 拉镜像 `docker pull sequenceiq/hadoop-docker`

```
root@vps-sfo181022:~# docker pull sequenceiq/hadoop-docker
Using default tag: latest
latest: Pulling from sequenceiq/hadoop-docker
b253335dcf03: Pulling fs layer
a3ed95caeb02: Pulling fs layer
69623ef05416: Pulling fs layer
8d2023764774: Pulling fs layer
0c3c0ff61963: Pulling fs layer
ff0696749bf6: Pull complete
72accdc282f3: Pull complete
5298ddb3b339: Pull complete
f252bbba6bda: Pull complete
3984257f0553: Pull complete
26343a20fa29: Pull complete
faf713cb3a4f: Pull complete
b7fa02232da2: Pull complete
b6e914dfab57: Pull complete
ed0c4a7a604c: Pull complete
d5381e4d358c: Pull complete
d31888a1eaab: Pull complete
7f6c4dc6b5: Pull complete
ed6629089518: Pull complete
36a49c5cc049: Pull complete
e6a7899cc72b: Pull complete
Digest: sha256:5a971a61840d9d32752330a891d528d35a558adf31d99c46205f1180af8e1abd
Status: Downloaded newer image for sequenceiq/hadoop-docker:latest
root@vps-sfo181022:~#
```

2. 运行容器 `docker run -it sequenceiq/hadoop-docker /bin/bash`

```
root@vps-sfo181022:~# docker run -it sequenceiq/hadoop-docker /bin/bash
bash-4.1#
```

3. 加入java环境

```
export JAVA_HOME=/usr/java/default export PATH=${JAVA_HOME}/bin:${PATH} export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar
bash-4.1# export JAVA_HOME=/usr/java/default
bash-4.1# export PATH=${JAVA_HOME}/bin:${PATH}
bash-4.1# export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar
bash-4.1# bin/hadoop com.sun.tools.javac.Main WordCount.java
bash: bin/hadoop: No such file or directory
bash-4.1# hadoop
bash: hadoop: command not found
```

4. 创建input、output文件夹 `mkdir wordcount cd wordcount mkdir input mkdir output`

```
bash-4.1# export JAVA_HOME=/usr/java/default
bash-4.1# export PATH=${JAVA_HOME}/bin:${PATH}
bash-4.1# export HADOOP_CLASSPATH=${JAVA_HOME}/lib/tools.jar
bash-4.1# bin/hadoop com.sun.tools.javac.Main WordCount.java
bash: bin/hadoop: No such file or directory
bash-4.1# hadoop
bash: hadoop: command not found
```

5. 编写file01, file02 `vi file01 vi file02`

6. 运行程序 `bin/hadoop jar wc.jar WordCount /user/joe/wordcount/input /user/joe/wordcount/output`

```
bash-4.1# bin/hadoop jar wc.jar WordCount wordcount/input /user/joe/wordcount/output
19/05/19 11:20:29 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
Exception in thread "main" java.net.ConnectException: Call From 93ce86a981b2/172.17.0.3 to localhost:9000 failed on connection exception: java.net.ConnectExc
eption: Connection refused; For more details see: http://wiki.apache.org/hadoop/ConnectionRefused
    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
    at sun.reflect.NativeConstructorAccessorImpl.newInstance(NativeConstructorAccessorImpl.java:57)
    at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(DelegatingConstructorAccessorImpl.java:45)
    at java.lang.reflect.Constructor.newInstance(Constructor.java:526)
    at org.apache.hadoop.net.NetUtils.wrapWithMessage(NetUtils.java:792)
    at org.apache.hadoop.net.NetUtils.wrapException(NetUtils.java:732)
    at org.apache.hadoop.ipc.Client.call(Client.java:1480)
    at org.apache.hadoop.ipc.Client.call(Client.java:1407)
    at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEngine.java:229)
    at com.sun.proxy.$Proxy9.getFileInfo(Unknown Source)
    at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.getFileInfo(ClientNamenodeProtocolTranslatorPB.java:771)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:606)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryInvocationHandler.java:187)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocationHandler.java:102)
    at com.sun.proxy.$Proxy10.getFileInfo(Unknown Source)
    at org.apache.hadoop.hdfs.DFSClient.getFileInfo(DFSClient.java:2113)
    at org.apache.hadoop.hdfs.DistributedFileSystem$22.doCall(DistributedFileSystem.java:1305)
    at org.apache.hadoop.hdfs.DistributedFileSystem$22.doCall(DistributedFileSystem.java:1301)
    at org.apache.hadoop.fs.FileSystemLinkResolver.resolve(FileSystemLinkResolver.java:81)
    at org.apache.hadoop.hdfs.DistributedFileSystem.getFileStatus(DistributedFileSystem.java:1301)
    at org.apache.hadoop.fs.FileSystem.exists(FileSystem.java:1424)
    at org.apache.hadoop.mapreduce.lib.output.FileOutputFormat.checkOutputSpecs(FileOutputFormat.java:145)
    at org.apache.hadoop.mapreduce.JobSubmitter.checkSpecs(JobSubmitter.java:269)
    at org.apache.hadoop.mapreduce.JobSubmitter.submitJobInternal(JobSubmitter.java:142)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1290)
    at org.apache.hadoop.mapreduce.Job$10.run(Job.java:1287)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:415)
    at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1657)
    at org.apache.hadoop.mapreduce.Job.submit(Job.java:1287)
    at org.apache.hadoop.mapreduce.Job.waitForCompletion(Job.java:1308)
    at WordCount.main(WordCount.java:59)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
```

实验数据或结果

如上

实验思考

参考资料

- https://hadoop.apache.org/docs/stable/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html#Example:_WordCount_v1.0
- <https://docs.mongodb.com/manual/tutorial/map-reduce-examples/>