Run an Example MapReduce program On Ubuntu 14.04 Linux

---Multi-Node Cluster

此文档默认已经配置好 Hadoop 集群并且已经正常开启集群。

1. Run an example

1)check the status

```
$ jps
$ bin/hdfs dfsadmin -report
```

2)Run a Map Reduce job

Create a directory in HDFS, and put some files to input, and list the files:

```
$ bin/hdfs dfs –mkdir –p /user/hadoop
$ bin/hdfs dfs –put etc/hadoop/hadoop-env.sh input
$ bin/hdfs dfs –ls /user/hadoop/input
```

Run WordCount:

```
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar
wordcount input output
```

Check output:

```
$ bin/hdfs dfs –cat output/*
```

2. Run your own procedure(手动编译运行 WordCount 程序)

Source code:

https://github.com/apache/hadoop-common/blob/trunk/hadoop-mapreduce-project/hadoop-mapreduce-examples/src/main/java/org/apache/hadoop/examples/WordCount.java

or use command wget to get: https://10.1.220.23:8888/test/WordCount.java

1) compile WordCount.java

   在/usr/local/hadoop 文件夹下创建文件夹 code/以存放 WordCount.java 代码。

   Hadoop 2.x 版本中 jar 不再集中在一个 hadoop-core*.jar 中，而是分成多个 jar，如运行 WordCount 实例需要如下三个 jar:

   $HADOOP_HOME/share/hadoop/common/hadoop-common-2.6.0.jar

   $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.6.0.jar

   $HADOOP_HOME/share/hadoop/common/lib/commons-cli-1.2.jar

```
$ cd code/
$ javac -cp ../share/hadoop/common/hadoop-common-
2.6.0.jar:../share/hadoop/mapreduce/hadoop-mapreduce-client-core-
2.6.0.jar:../share/hadoop/common/lib/commons-cli-1.2.jar WordCount.java -d ./
```

此时可以看到生产的 class 文件在目录/code/org/apache/hadoop/examples/下。

2) package to jar

把生成的 class 文件打包成 jar 文件：

```
$ jar -cvf WordCount.jar org/apache/hadoop/examples/WordCount*.class
```

3) run

由于之前已经运行过自带的例子，input 文件已经有了，而 hadoop 是不会自动覆盖 output 文件的，故每一次运行之前都应该将 output 文件夹下面的内容删除或者重新制定一个输出目录，如这里使用 output1.

```
$ /usr/local/hadoop/bin/hadoop jar WordCount.jar
org.apache.hadoop.examples.WordCount input output1
```

4) Check output:

```
$ bin/hdfs dfs –cat output1/*
```