[](http://blog.fens.me/wp-content/uploads/2013/09/hadoop-maven.png)

**前言**

Hadoop的MapReduce环境是一个复杂的编程环境，所以我们要尽可能地简化构建MapReduce项目的过程。Maven是一个很不错的自动化项目构建工具，通过Maven来帮助我们从复杂的环境配置中解脱出来，从而标准化开发过程。所以，写MapReduce之前，让我们先花点时间把刀磨快！！当然，除了Maven还有其他的选择Gradle(推荐), Ivy….

后面将会有介绍几篇MapReduce开发的文章，都要依赖于本文中Maven的构建的MapReduce环境。

**目录**

1. Maven介绍
2. Maven安装(win)
3. Hadoop开发环境介绍
4. 用Maven构建Hadoop环境
5. MapReduce程序开发
6. 模板项目上传github

**1. Maven介绍**

Apache Maven，是一个Java的项目管理及自动构建工具，由Apache软件基金会所提供。基于项目对象模型（缩写：POM）概念，Maven利用一个中央信息片断能管理一个项目的构建、报告和文档等步骤。曾是Jakarta项目的子项目，现为独立Apache项目。

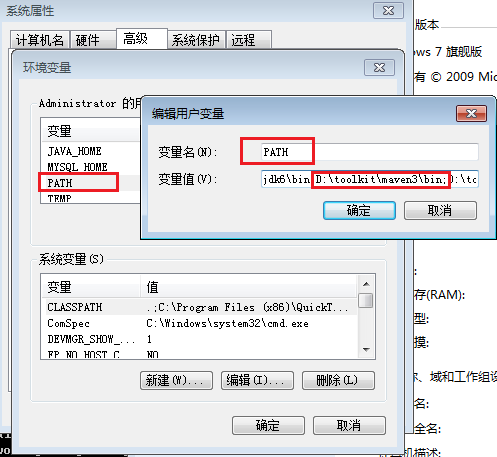
maven的开发者在他们开发网站上指出，maven的目标是要使得项目的构建更加容易，它把编译、打包、测试、发布等开发过程中的不同环节有机的串联了起来，并产生一致的、高质量的项目信息，使得项目成员能够及时地得到反馈。maven有效地支持了测试优先、持续集成，体现了鼓励沟通，及时反馈的软件开发理念。如果说Ant的复用是建立在”拷贝–粘贴”的基础上的，那么Maven通过插件的机制实现了项目构建逻辑的真正复用。

**2. Maven安装(win)**

下载Maven：<http://maven.apache.org/download.cgi>

下载最新的xxx-bin.zip文件，在win上解压到 D:\toolkit\maven3

并把maven/bin目录设置在环境变量PATH：

[](http://blog.fens.me/wp-content/uploads/2013/09/win7-maven.png)

然后，打开命令行输入mvn，我们会看到mvn命令的运行效果

~ C:\Users\Administrator>mvn

[INFO] Scanning for projects...

[INFO] ------------------------------------------------------------------------

[INFO] BUILD FAILURE

[INFO] ------------------------------------------------------------------------

[INFO] Total time: 0.086s

[INFO] Finished at: Mon Sep 30 18:26:58 CST 2013

[INFO] Final Memory: 2M/179M

[INFO] ------------------------------------------------------------------------

[ERROR] No goals have been specified for this build. You must specify a valid lifecycle phase or a goal in the format : or :[:]:. Available lifecycle phases are: validate, initialize, generate-sources, process-sources, generate-resources, process-resources, compile, process-class

es, generate-test-sources, process-test-sources, generate-test-resources, process-test-resources, test-compile, process-test-classes, test, prepare-package, package, pre-integration-test, integration-test, post-integration-test, verify, install, deploy, pre-clean, clean, post-clean, pre-site, site, post-site, site-deploy. -> [Help 1]

[ERROR]

[ERROR] To see the full stack trace of the errors, re-run Maven with the -e switch.

[ERROR] Re-run Maven using the -X switch to enable full debug logging.

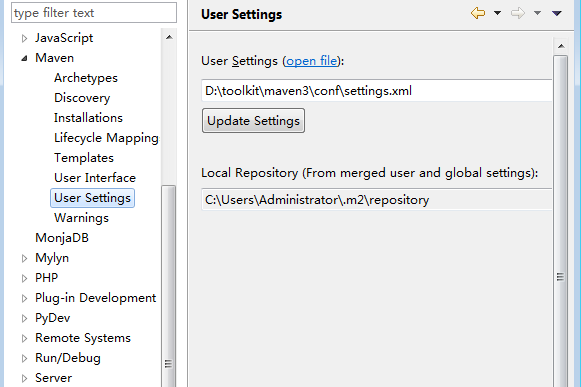
[ERROR]

[ERROR] For more information about the errors and possible solutions, please read the following articles:

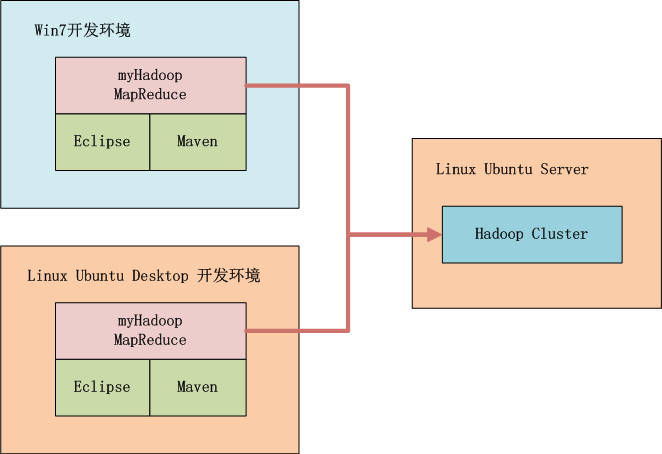
[ERROR] [Help 1] http://cwiki.apache.org/confluence/display/MAVEN/NoGoalSpecifiedException

安装Eclipse的Maven插件：[Maven Integration for Eclipse](http://www.eclipse.org/m2e/)

Maven的Eclipse插件配置

[](http://blog.fens.me/wp-content/uploads/2013/09/eclipse-maven.png)

**3. Hadoop开发环境介绍**

[](http://blog.fens.me/wp-content/uploads/2013/09/hadoop-dev.png)

如上图所示，我们可以选择在win中开发，也可以在linux中开发，本地启动Hadoop或者远程调用Hadoop，标配的工具都是Maven和Eclipse。

Hadoop集群系统环境：

* Linux: Ubuntu 12.04.2 LTS 64bit Server
* Java: 1.6.0\_29
* Hadoop: hadoop-1.0.3，单节点，IP:192.168.1.210

**4. 用Maven构建Hadoop环境**

* 1. 用Maven创建一个标准化的Java项目
* 2. 导入项目到eclipse
* 3. 增加hadoop依赖，修改pom.xml
* 4. 下载依赖
* 5. 从Hadoop集群环境下载hadoop配置文件
* 6. 配置本地host

**1). 用Maven创建一个标准化的Java项目**

~ D:\workspace\java>mvn archetype:generate -DarchetypeGroupId=org.apache.maven.archetypes -DgroupId=org.conan.myhadoop.mr

-DartifactId=myHadoop -DpackageName=org.conan.myhadoop.mr -Dversion=1.0-SNAPSHOT -DinteractiveMode=false

[INFO] Scanning for projects...

[INFO]

[INFO] ------------------------------------------------------------------------

[INFO] Building Maven Stub Project (No POM) 1

[INFO] ------------------------------------------------------------------------

[INFO]

[INFO] >>> maven-archetype-plugin:2.2:generate (default-cli) @ standalone-pom >>>

[INFO]

[INFO] <<< maven-archetype-plugin:2.2:generate (default-cli) @ standalone-pom <<<

[INFO]

[INFO] --- maven-archetype-plugin:2.2:generate (default-cli) @ standalone-pom ---

[INFO] Generating project in Batch mode

[INFO] No archetype defined. Using maven-archetype-quickstart (org.apache.maven.archetypes:maven-archetype-quickstart:1.

0)

Downloading: http://repo.maven.apache.org/maven2/org/apache/maven/archetypes/maven-archetype-quickstart/1.0/maven-archet

ype-quickstart-1.0.jar

Downloaded: http://repo.maven.apache.org/maven2/org/apache/maven/archetypes/maven-archetype-quickstart/1.0/maven-archety

pe-quickstart-1.0.jar (5 KB at 4.3 KB/sec)

Downloading: http://repo.maven.apache.org/maven2/org/apache/maven/archetypes/maven-archetype-quickstart/1.0/maven-archet

ype-quickstart-1.0.pom

Downloaded: http://repo.maven.apache.org/maven2/org/apache/maven/archetypes/maven-archetype-quickstart/1.0/maven-archety

pe-quickstart-1.0.pom (703 B at 1.6 KB/sec)

[INFO] ----------------------------------------------------------------------------

[INFO] Using following parameters for creating project from Old (1.x) Archetype: maven-archetype-quickstart:1.0

[INFO] ----------------------------------------------------------------------------

[INFO] Parameter: groupId, Value: org.conan.myhadoop.mr

[INFO] Parameter: packageName, Value: org.conan.myhadoop.mr

[INFO] Parameter: package, Value: org.conan.myhadoop.mr

[INFO] Parameter: artifactId, Value: myHadoop

[INFO] Parameter: basedir, Value: D:\workspace\java

[INFO] Parameter: version, Value: 1.0-SNAPSHOT

[INFO] project created from Old (1.x) Archetype in dir: D:\workspace\java\myHadoop

[INFO] ------------------------------------------------------------------------

[INFO] BUILD SUCCESS

[INFO] ------------------------------------------------------------------------

[INFO] Total time: 8.896s

[INFO] Finished at: Sun Sep 29 20:57:07 CST 2013

[INFO] Final Memory: 9M/179M

[INFO] ------------------------------------------------------------------------

进入项目，执行mvn命令

~ D:\workspace\java>cd myHadoop

~ D:\workspace\java\myHadoop>mvn clean install

[INFO]

[INFO] --- maven-jar-plugin:2.3.2:jar (default-jar) @ myHadoop ---

[INFO] Building jar: D:\workspace\java\myHadoop\target\myHadoop-1.0-SNAPSHOT.jar

[INFO]

[INFO] --- maven-install-plugin:2.3.1:install (default-install) @ myHadoop ---

[INFO] Installing D:\workspace\java\myHadoop\target\myHadoop-1.0-SNAPSHOT.jar to C:\Users\Administrator\.m2\repository\o

rg\conan\myhadoop\mr\myHadoop\1.0-SNAPSHOT\myHadoop-1.0-SNAPSHOT.jar

[INFO] Installing D:\workspace\java\myHadoop\pom.xml to C:\Users\Administrator\.m2\repository\org\conan\myhadoop\mr\myHa

doop\1.0-SNAPSHOT\myHadoop-1.0-SNAPSHOT.pom

[INFO] ------------------------------------------------------------------------

[INFO] BUILD SUCCESS

[INFO] ------------------------------------------------------------------------

[INFO] Total time: 4.348s

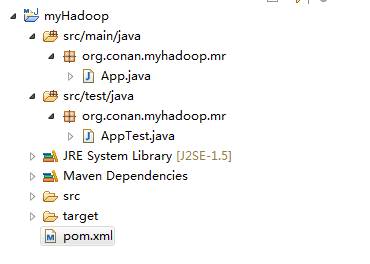
[INFO] Finished at: Sun Sep 29 20:58:43 CST 2013

[INFO] Final Memory: 11M/179M

[INFO] ------------------------------------------------------------------------

**2). 导入项目到eclipse**

我们创建好了一个基本的maven项目，然后导入到eclipse中。 这里我们最好已安装好了Maven的插件。

[](http://blog.fens.me/wp-content/uploads/2013/09/hadoop-eclipse.png)

**3). 增加hadoop依赖**

这里我使用hadoop-1.0.3版本，修改文件：pom.xml

~ vi pom.xml

<project xmlns="http://maven.apache.org/POM/4.0.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://maven.apache.org/POM/4.0.0 http://maven.apache.org/maven-v4\_0\_0.xsd">

<modelVersion>4.0.0</modelVersion>

<groupId>org.conan.myhadoop.mr</groupId>

<artifactId>myHadoop</artifactId>

<packaging>jar</packaging>

<version>1.0-SNAPSHOT</version>

<name>myHadoop</name>

<url>http://maven.apache.org</url>

<dependencies>

<dependency>

<groupId>org.apache.hadoop</groupId>

<artifactId>hadoop-core</artifactId>

<version>1.0.3</version>

</dependency>

<dependency>

<groupId>junit</groupId>

<artifactId>junit</artifactId>

<version>4.4</version>

<scope>test</scope>

</dependency>

</dependencies>

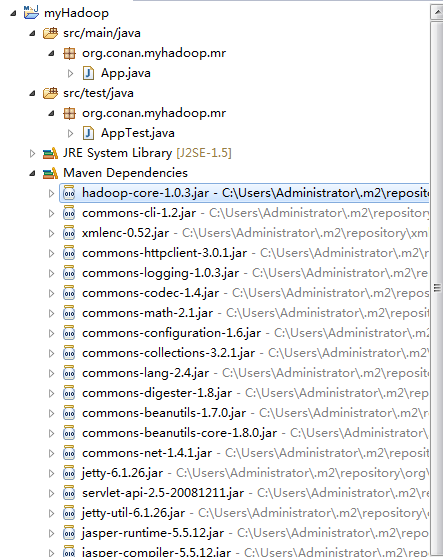
</project>

**4). 下载依赖**

下载依赖：

~ mvn clean install

在eclipse中刷新项目：

[](http://blog.fens.me/wp-content/uploads/2013/09/hadoop-eclipse-maven.png)

项目的依赖程序，被自动加载的库路径下面。

**5). 从Hadoop集群环境下载hadoop配置文件**

* + core-site.xml
  + hdfs-site.xml
  + mapred-site.xml

查看core-site.xml

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>fs.default.name</name>

<value>hdfs://master:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/home/conan/hadoop/tmp</value>

</property>

<property>

<name>io.sort.mb</name>

<value>256</value>

</property>

</configuration>

查看hdfs-site.xml

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

<name>dfs.data.dir</name>

<value>/home/conan/hadoop/data</value>

</property>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.permissions</name>

<value>false</value>

</property>

</configuration>

查看mapred-site.xml

<?xml version="1.0"?>

<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>

<property>

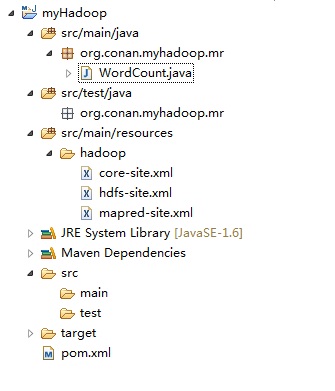
<name>mapred.job.tracker</name>

<value>hdfs://master:9001</value>

</property>

</configuration>

保存在src/main/resources/hadoop目录下面

[](http://blog.fens.me/wp-content/uploads/2013/09/hadoop-config.png)

删除原自动生成的文件：App.java和AppTest.java

**6).配置本地host，增加master的域名指向**

~ vi c:/Windows/System32/drivers/etc/hosts

192.168.1.210 master

**6. MapReduce程序开发**

编写一个简单的MapReduce程序，实现wordcount功能。

新一个Java文件：WordCount.java

package org.conan.myhadoop.mr;

import java.io.IOException;

import java.util.Iterator;

import java.util.StringTokenizer;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapred.FileInputFormat;

import org.apache.hadoop.mapred.FileOutputFormat;

import org.apache.hadoop.mapred.JobClient;

import org.apache.hadoop.mapred.JobConf;

import org.apache.hadoop.mapred.MapReduceBase;

import org.apache.hadoop.mapred.Mapper;

import org.apache.hadoop.mapred.OutputCollector;

import org.apache.hadoop.mapred.Reducer;

import org.apache.hadoop.mapred.Reporter;

import org.apache.hadoop.mapred.TextInputFormat;

import org.apache.hadoop.mapred.TextOutputFormat;

public class WordCount {

public static class WordCountMapper extends MapReduceBase implements Mapper<Object, Text, Text, IntWritable> {

private final static IntWritable one = new IntWritable(1);

private Text word = new Text();

@Override

public void map(Object key, Text value, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {

StringTokenizer itr = new StringTokenizer(value.toString());

while (itr.hasMoreTokens()) {

word.set(itr.nextToken());

output.collect(word, one);

}

}

}

public static class WordCountReducer extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {

private IntWritable result = new IntWritable();

@Override

public void reduce(Text key, Iterator values, OutputCollector<Text, IntWritable> output, Reporter reporter) throws IOException {

int sum = 0;

while (values.hasNext()) {

sum += values.next().get();

}

result.set(sum);

output.collect(key, result);

}

}

public static void main(String[] args) throws Exception {

String input = "hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account";

String output = "hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/result";

JobConf conf = new JobConf(WordCount.class);

conf.setJobName("WordCount");

conf.addResource("classpath:/hadoop/core-site.xml");

conf.addResource("classpath:/hadoop/hdfs-site.xml");

conf.addResource("classpath:/hadoop/mapred-site.xml");

conf.setOutputKeyClass(Text.class);

conf.setOutputValueClass(IntWritable.class);

conf.setMapperClass(WordCountMapper.class);

conf.setCombinerClass(WordCountReducer.class);

conf.setReducerClass(WordCountReducer.class);

conf.setInputFormat(TextInputFormat.class);

conf.setOutputFormat(TextOutputFormat.class);

FileInputFormat.setInputPaths(conf, new Path(input));

FileOutputFormat.setOutputPath(conf, new Path(output));

JobClient.runJob(conf);

System.exit(0);

}

}

**启动Java APP.**

控制台错误

2013-9-30 19:25:02 org.apache.hadoop.util.NativeCodeLoader

警告: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

2013-9-30 19:25:02 org.apache.hadoop.security.UserGroupInformation doAs

严重: PriviledgedActionException as:Administrator cause:java.io.IOException: Failed to set permissions of path: \tmp\hadoop-Administrator\mapred\staging\Administrator1702422322\.staging to 0700

Exception in thread "main" java.io.IOException: Failed to set permissions of path: \tmp\hadoop-Administrator\mapred\staging\Administrator1702422322\.staging to 0700

at org.apache.hadoop.fs.FileUtil.checkReturnValue(FileUtil.java:689)

at org.apache.hadoop.fs.FileUtil.setPermission(FileUtil.java:662)

at org.apache.hadoop.fs.RawLocalFileSystem.setPermission(RawLocalFileSystem.java:509)

at org.apache.hadoop.fs.RawLocalFileSystem.mkdirs(RawLocalFileSystem.java:344)

at org.apache.hadoop.fs.FilterFileSystem.mkdirs(FilterFileSystem.java:189)

at org.apache.hadoop.mapreduce.JobSubmissionFiles.getStagingDir(JobSubmissionFiles.java:116)

at org.apache.hadoop.mapred.JobClient$2.run(JobClient.java:856)

at org.apache.hadoop.mapred.JobClient$2.run(JobClient.java:850)

at java.security.AccessController.doPrivileged(Native Method)

at javax.security.auth.Subject.doAs(Subject.java:396)

at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1121)

at org.apache.hadoop.mapred.JobClient.submitJobInternal(JobClient.java:850)

at org.apache.hadoop.mapred.JobClient.submitJob(JobClient.java:824)

at org.apache.hadoop.mapred.JobClient.runJob(JobClient.java:1261)

at org.conan.myhadoop.mr.WordCount.main(WordCount.java:78)

这个错误是win中开发特有的错误，文件权限问题，在Linux下可以正常运行。

解决方法是，修改/hadoop-1.0.3/src/core/org/apache/hadoop/fs/FileUtil.java文件

688-692行注释，然后重新编译源代码，重新打一个hadoop.jar的包。

685 private static void checkReturnValue(boolean rv, File p,

686 FsPermission permission

687 ) throws IOException {

688 /\*if (!rv) {

689 throw new IOException("Failed to set permissions of path: " + p +

690 " to " +

691 String.format("%04o", permission.toShort()));

692 }\*/

693 }

我这里自己打了一个hadoop-core-1.0.3.jar包，放到了lib下面。

我们还要替换maven中的hadoop类库。

~ cp lib/hadoop-core-1.0.3.jar C:\Users\Administrator\.m2\repository\org\apache\hadoop\hadoop-core\1.0.3\hadoop-core-1.0.3.jar

再次启动Java APP，控制台输出：

2013-9-30 19:50:49 org.apache.hadoop.util.NativeCodeLoader

警告: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

2013-9-30 19:50:49 org.apache.hadoop.mapred.JobClient copyAndConfigureFiles

警告: Use GenericOptionsParser for parsing the arguments. Applications should implement Tool for the same.

2013-9-30 19:50:49 org.apache.hadoop.mapred.JobClient copyAndConfigureFiles

警告: No job jar file set. User classes may not be found. See JobConf(Class) or JobConf#setJar(String).

2013-9-30 19:50:49 org.apache.hadoop.io.compress.snappy.LoadSnappy

警告: Snappy native library not loaded

2013-9-30 19:50:49 org.apache.hadoop.mapred.FileInputFormat listStatus

信息: Total input paths to process : 4

2013-9-30 19:50:50 org.apache.hadoop.mapred.JobClient monitorAndPrintJob

信息: Running job: job\_local\_0001

2013-9-30 19:50:50 org.apache.hadoop.mapred.Task initialize

信息: Using ResourceCalculatorPlugin : null

2013-9-30 19:50:50 org.apache.hadoop.mapred.MapTask runOldMapper

信息: numReduceTasks: 1

2013-9-30 19:50:50 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: io.sort.mb = 100

2013-9-30 19:50:50 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: data buffer = 79691776/99614720

2013-9-30 19:50:50 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: record buffer = 262144/327680

2013-9-30 19:50:50 org.apache.hadoop.mapred.MapTask$MapOutputBuffer flush

信息: Starting flush of map output

2013-9-30 19:50:50 org.apache.hadoop.mapred.MapTask$MapOutputBuffer sortAndSpill

信息: Finished spill 0

2013-9-30 19:50:50 org.apache.hadoop.mapred.Task done

信息: Task:attempt\_local\_0001\_m\_000000\_0 is done. And is in the process of commiting

2013-9-30 19:50:51 org.apache.hadoop.mapred.JobClient monitorAndPrintJob

信息: map 0% reduce 0%

2013-9-30 19:50:53 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息: hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/part-m-00003:0+119

2013-9-30 19:50:53 org.apache.hadoop.mapred.Task sendDone

信息: Task 'attempt\_local\_0001\_m\_000000\_0' done.

2013-9-30 19:50:53 org.apache.hadoop.mapred.Task initialize

信息: Using ResourceCalculatorPlugin : null

2013-9-30 19:50:53 org.apache.hadoop.mapred.MapTask runOldMapper

信息: numReduceTasks: 1

2013-9-30 19:50:53 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: io.sort.mb = 100

2013-9-30 19:50:53 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: data buffer = 79691776/99614720

2013-9-30 19:50:53 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: record buffer = 262144/327680

2013-9-30 19:50:53 org.apache.hadoop.mapred.MapTask$MapOutputBuffer flush

信息: Starting flush of map output

2013-9-30 19:50:53 org.apache.hadoop.mapred.MapTask$MapOutputBuffer sortAndSpill

信息: Finished spill 0

2013-9-30 19:50:53 org.apache.hadoop.mapred.Task done

信息: Task:attempt\_local\_0001\_m\_000001\_0 is done. And is in the process of commiting

2013-9-30 19:50:54 org.apache.hadoop.mapred.JobClient monitorAndPrintJob

信息: map 100% reduce 0%

2013-9-30 19:50:56 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息: hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/part-m-00000:0+113

2013-9-30 19:50:56 org.apache.hadoop.mapred.Task sendDone

信息: Task 'attempt\_local\_0001\_m\_000001\_0' done.

2013-9-30 19:50:56 org.apache.hadoop.mapred.Task initialize

信息: Using ResourceCalculatorPlugin : null

2013-9-30 19:50:56 org.apache.hadoop.mapred.MapTask runOldMapper

信息: numReduceTasks: 1

2013-9-30 19:50:56 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: io.sort.mb = 100

2013-9-30 19:50:56 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: data buffer = 79691776/99614720

2013-9-30 19:50:56 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: record buffer = 262144/327680

2013-9-30 19:50:56 org.apache.hadoop.mapred.MapTask$MapOutputBuffer flush

信息: Starting flush of map output

2013-9-30 19:50:56 org.apache.hadoop.mapred.MapTask$MapOutputBuffer sortAndSpill

信息: Finished spill 0

2013-9-30 19:50:56 org.apache.hadoop.mapred.Task done

信息: Task:attempt\_local\_0001\_m\_000002\_0 is done. And is in the process of commiting

2013-9-30 19:50:59 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息: hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/part-m-00001:0+110

2013-9-30 19:50:59 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息: hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/part-m-00001:0+110

2013-9-30 19:50:59 org.apache.hadoop.mapred.Task sendDone

信息: Task 'attempt\_local\_0001\_m\_000002\_0' done.

2013-9-30 19:50:59 org.apache.hadoop.mapred.Task initialize

信息: Using ResourceCalculatorPlugin : null

2013-9-30 19:50:59 org.apache.hadoop.mapred.MapTask runOldMapper

信息: numReduceTasks: 1

2013-9-30 19:50:59 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: io.sort.mb = 100

2013-9-30 19:50:59 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: data buffer = 79691776/99614720

2013-9-30 19:50:59 org.apache.hadoop.mapred.MapTask$MapOutputBuffer

信息: record buffer = 262144/327680

2013-9-30 19:50:59 org.apache.hadoop.mapred.MapTask$MapOutputBuffer flush

信息: Starting flush of map output

2013-9-30 19:50:59 org.apache.hadoop.mapred.MapTask$MapOutputBuffer sortAndSpill

信息: Finished spill 0

2013-9-30 19:50:59 org.apache.hadoop.mapred.Task done

信息: Task:attempt\_local\_0001\_m\_000003\_0 is done. And is in the process of commiting

2013-9-30 19:51:02 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息: hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/part-m-00002:0+79

2013-9-30 19:51:02 org.apache.hadoop.mapred.Task sendDone

信息: Task 'attempt\_local\_0001\_m\_000003\_0' done.

2013-9-30 19:51:02 org.apache.hadoop.mapred.Task initialize

信息: Using ResourceCalculatorPlugin : null

2013-9-30 19:51:02 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息:

2013-9-30 19:51:02 org.apache.hadoop.mapred.Merger$MergeQueue merge

信息: Merging 4 sorted segments

2013-9-30 19:51:02 org.apache.hadoop.mapred.Merger$MergeQueue merge

信息: Down to the last merge-pass, with 4 segments left of total size: 442 bytes

2013-9-30 19:51:02 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息:

2013-9-30 19:51:02 org.apache.hadoop.mapred.Task done

信息: Task:attempt\_local\_0001\_r\_000000\_0 is done. And is in the process of commiting

2013-9-30 19:51:02 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息:

2013-9-30 19:51:02 org.apache.hadoop.mapred.Task commit

信息: Task attempt\_local\_0001\_r\_000000\_0 is allowed to commit now

2013-9-30 19:51:02 org.apache.hadoop.mapred.FileOutputCommitter commitTask

信息: Saved output of task 'attempt\_local\_0001\_r\_000000\_0' to hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/result

2013-9-30 19:51:05 org.apache.hadoop.mapred.LocalJobRunner$Job statusUpdate

信息: reduce > reduce

2013-9-30 19:51:05 org.apache.hadoop.mapred.Task sendDone

信息: Task 'attempt\_local\_0001\_r\_000000\_0' done.

2013-9-30 19:51:06 org.apache.hadoop.mapred.JobClient monitorAndPrintJob

信息: map 100% reduce 100%

2013-9-30 19:51:06 org.apache.hadoop.mapred.JobClient monitorAndPrintJob

信息: Job complete: job\_local\_0001

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Counters: 20

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: File Input Format Counters

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Bytes Read=421

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: File Output Format Counters

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Bytes Written=348

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: FileSystemCounters

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: FILE\_BYTES\_READ=7377

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: HDFS\_BYTES\_READ=1535

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: FILE\_BYTES\_WRITTEN=209510

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: HDFS\_BYTES\_WRITTEN=348

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Map-Reduce Framework

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Map output materialized bytes=458

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Map input records=11

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Reduce shuffle bytes=0

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Spilled Records=30

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Map output bytes=509

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Total committed heap usage (bytes)=1838546944

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Map input bytes=421

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: SPLIT\_RAW\_BYTES=452

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Combine input records=22

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Reduce input records=15

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Reduce input groups=13

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Combine output records=15

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Reduce output records=13

2013-9-30 19:51:06 org.apache.hadoop.mapred.Counters log

信息: Map output records=22

成功运行了wordcount程序，通过命令我们查看输出结果

~ hadoop fs -ls hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/result

Found 2 items

-rw-r--r-- 3 Administrator supergroup 0 2013-09-30 19:51 /user/hdfs/o\_t\_account/result/\_SUCCESS

-rw-r--r-- 3 Administrator supergroup 348 2013-09-30 19:51 /user/hdfs/o\_t\_account/result/part-00000

~ hadoop fs -cat hdfs://192.168.1.210:9000/user/hdfs/o\_t\_account/result/part-00000

1,abc@163.com,2013-04-22 1

10,ade121@sohu.com,2013-04-23 1

11,addde@sohu.com,2013-04-23 1

17:21:24.0 5

2,dedac@163.com,2013-04-22 1

20:21:39.0 6

3,qq8fed@163.com,2013-04-22 1

4,qw1@163.com,2013-04-22 1

5,af3d@163.com,2013-04-22 1

6,ab34@163.com,2013-04-22 1

7,q8d1@gmail.com,2013-04-23 1

8,conan@gmail.com,2013-04-23 1

9,adeg@sohu.com,2013-04-23 1

这样，我们就实现了在win7中的开发，通过Maven构建Hadoop依赖环境，在Eclipse中开发MapReduce的程序，然后运行JavaAPP。Hadoop应用会自动把我们的MR程序打成jar包，再上传的远程的hadoop环境中运行，返回日志在Eclipse控制台输出。

**7. 模板项目上传github**

<https://github.com/bsspirit/maven_hadoop_template>

大家可以下载这个项目，做为开发的起点。

~ git clone https://github.com/bsspirit/maven\_hadoop\_template.git

我们完成第一步，下面就将正式进入MapReduce开发实践。