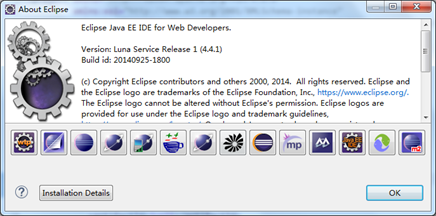
# 1. 环境

Eclipse版本Luna 4.4.1

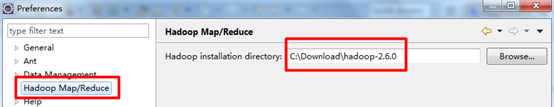


安装插件hadoop-eclipse-plugin-2.6.0.jar，下载后放到eclipse/plugins目录即可。

# 2. 配置插件

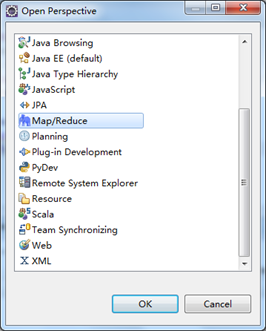
## 2.1 配置hadoop主目录

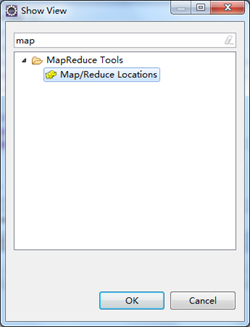
解压缩hadoop-2.6.0.tar.gz到C:\Downloads\hadoop-2.6.0，在eclipse的Windows->Preferences的Hadoop Map/Reduce中设置安装目录。



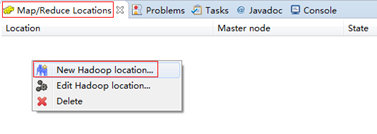
## 2.2 配置插件

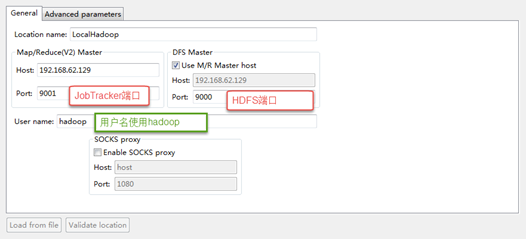
打开Windows->Open Perspective中的Map/Reduce，在此perspective下进行hadoop程序开发。





打开Windows->Show View中的Map/Reduce Locations，如下图右键选择New Hadoop location…新建hadoop连接。

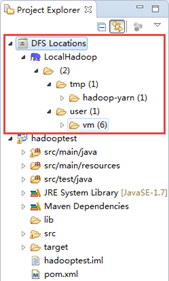




确认完成以后如下，eclipse会连接hadoop集群。

802162916118.png

如果连接成功，在project explorer的DFS Locations下会展现hdfs集群中的文件。



# 3. 开发hadoop程序

## 3.1 程序开发

开发一个Sort示例，对输入整数进行排序。输入文件格式是每行一个整数。

1 package com.ccb;

2

3 /\*\*

4 \* Created by hp on 2015-7-20.

5 \*/

6

7import java.io.IOException;

8

9 import org.apache.hadoop.conf.Configuration;

10 import org.apache.hadoop.fs.FileSystem;

11 import org.apache.hadoop.fs.Path;

12 import org.apache.hadoop.io.IntWritable;

13 import org.apache.hadoop.io.Text;

14 import org.apache.hadoop.mapreduce.Job;

15 import org.apache.hadoop.mapreduce.Mapper;

16 import org.apache.hadoop.mapreduce.Reducer;

17 import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;

18 import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

19

20 public class Sort {

21

22 // 每行记录是一个整数。将Text文本转换为IntWritable类型，作为map的key

23 public static class Map extends Mapper<Object, Text, IntWritable, IntWritable> {

24 private static IntWritable data = new IntWritable();

25

26 // 实现map函数

27 public void map(Object key, Text value, Context context) throws IOException, InterruptedException {

28 String line = value.toString();

29 data.set(Integer.parseInt(line));

30 context.write(data, new IntWritable(1));

31 }

32 }

33

34 // reduce之前hadoop框架会进行shuffle和排序，因此直接输出key即可。

35 public static class Reduce extends Reducer<IntWritable, IntWritable, IntWritable, Text> {

36

37 //实现reduce函数

38 public void reduce(IntWritable key, Iterable<IntWritable> values, Context context) throws IOException, InterruptedException {

39 for (IntWritable v : values) {

40 context.write(key, new Text(""));

41 }

42 }

43 }

44

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

46 Configuration conf = new Configuration();

47

48 // 指定JobTracker地址

49 conf.set("mapred.job.tracker", "192.168.62.129:9001");

50 if (args.length != 2) {

51 System.err.println("Usage: Data Sort <in> <out>");

52 System.exit(2);

53 }

54 System.out.println(args[0]);

55 System.out.println(args[1]);

56

57 Job job = Job.getInstance(conf, "Data Sort");

58 job.setJarByClass(Sort.class);

59

60 //设置Map和Reduce处理类

61 job.setMapperClass(Map.class);

62 job.setReducerClass(Reduce.class);

63

64 //设置输出类型

65 job.setOutputKeyClass(IntWritable.class);

66 job.setOutputValueClass(IntWritable.class);

67

68 //设置输入和输出目录

69 FileInputFormat.addInputPath(job, new Path(args[0]));

70 FileOutputFormat.setOutputPath(job, new Path(args[1]));

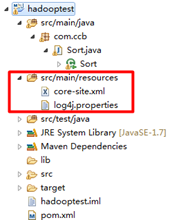
71 System.exit(job.waitForCompletion(true) ? 0 : 1);

72 }

73 }

## **3.2 配置文件**

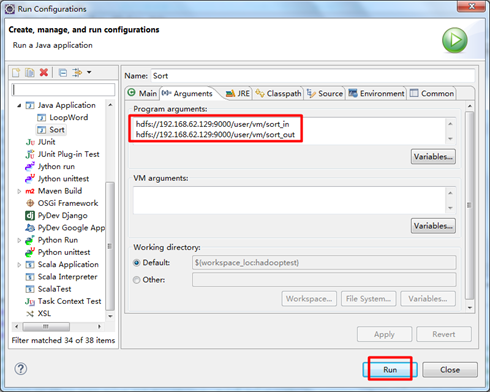
把log4j.properties和hadoop集群中的core-site.xml加入到classpath中。我的示例工程是maven组织，因此放到src/main/resources目录。



程序执行时会从core-site.xml中获取hdfs地址。

## 3.3 程序执行

右键选择Run As -> Run Configurations…，在参数中填好输入输出目录，执行Run即可。



# 4.windows环境

windows环境执行程序可能出现：Could not locate executable null\bin\winutils.exe in the Hadoop binaries

在源码中：

public static final String getQualifiedBinPath(String executable) throws IOException {

    // construct hadoop bin path to the specified executable

    String fullExeName = HADOOP\_HOME\_DIR + File.separator + "bin" + File.separator + executable;

    File exeFile = new File(fullExeName);

    if (!exeFile.exists()) {

      throw new IOException("Could not locate executable " + fullExeName

        + " in the Hadoop binaries.");

    }

    return exeFile.getCanonicalPath();

  }

private static String HADOOP\_HOME\_DIR = checkHadoopHome();

private static String checkHadoopHome() {

    // first check the Dflag hadoop.home.dir with JVM scope

    String home = System.getProperty("hadoop.home.dir");

    // fall back to the system/user-global env variable

    if (home == null) {

      home = System.getenv("HADOOP\_HOME");

    }

     ...

}

很明显应该是HADOOP\_HOME的问题。如果HADOOP\_HOME为空，必然fullExeName为null\bin\winutils.exe。

解决方法很简单，配置环境变量，不想重启电脑可以在程序里加上：System.setProperty("hadoop.home.dir", "E:\\Program Files\\hadoop-2.7.0");

之后继续执行程序，发现依旧不行，因为hadoop-x.x.x/bin目录下，压根就没有winutils.exe。

下载地址是：https://github.com/srccodes/hadoop-common-2.2.0-bin

不要顾虑它的版本，不用怕，因为我用的最新的hadoop-2.7.0都没问题！下载好后，把winutils.exe加入你的hadoop-x.x.x/bin下。

# 5. 可能出现的问题

## 5.1 权限问题，无法访问HDFS

修改集群hdfs-site.xml配置，关闭hadoop集群的权限校验。

|  |
| --- |
| <property> <name>dfs.permissions</name> <value>false</value> </property> |

## 5.2 出现NullPointerException异常

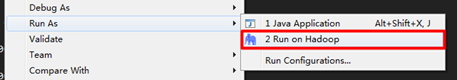
在环境变量中配置%HADOOP\_HOME%为C:\Download\hadoop-2.6.0\

下载winutils.exe和hadoop.dll到C:\Download\hadoop-2.6.0\bin

注意：网上很多资料说的是下载hadoop-common-2.2.0-bin-master.zip，但很多不支持hadoop2.6.0版本。需要下载支持hadoop2.6.0版本的程序。

## 5.3 程序执行失败

需要执行Run on Hadoop，而不是Java Application。（其实都是可以的没有问题）



# 6.参考链接

[Ubuntu](http://www.linuxidc.com/topicnews.aspx?tid=2)<http://www.linuxidc.com/Linux/2015-02/113487.htm>14.04下Hadoop2.4.1单机/伪分布式安装配置教程

**[CentOS](http://www.linuxidc.com/topicnews.aspx?tid=14)**<http://www.linuxidc.com/Linux/2014-01/94685.htm>**安装和配置Hadoop2.2.0**

<http://www.linuxidc.com/Linux/2013-06/86106.htm>Ubuntu 13.04上搭建Hadoop环境

<http://www.linuxidc.com/Linux/2013-09/90600.htm>Ubuntu 12.10 +Hadoop 1.2.1版本集群配置

<http://www.linuxidc.com/Linux/2013-01/77681.htm>Ubuntu上搭建Hadoop环境（单机模式+伪分布模式）

<http://www.linuxidc.com/Linux/2012-11/74539.htm>Ubuntu下Hadoop环境的配置

<http://www.linuxidc.com/Linux/2012-02/53927.htm>单机版搭建Hadoop环境图文教程详解