
1 作者与邀请

作者：翁安栋

作者的网站：www.mathings.com

联系作者：wengad@foxmail.com

作者的介绍：架构师、大数据行者

邀请：有志于致力大数据等等基础学习、培训、应用的道友加入我的网站。

2 概述

本文描述了搭建 windows 下的 hadoop 开发环境的过程。以下为 windows 端的环境部署，默认 linux 端的 hadoop 伪分布式已经部署成功。

- 客户端开发环境：windows 10 64bit+Eclipse Mars.1 Release (4.5.1)
64bit+jdk1.7
- 服务器端环境：CENTOS 7 64BIT+HADOOP 2.7.3，伪分布式部署

3 软件下载

- www.eclipse.org 下在 eclipse 相应版本；
- 到百度云盘下载：hadoop-2.7.3.tar.gz，hadoop-eclipse-plugin-2.7.3.jar，hadoop-common-2.7.1-bin-master.zip、jdk-7u80-windows-x64.exe、YARNRunner.java
- 百度云盘下载地：<https://pan.baidu.com/s/1geU53WZ>

4 环境安装

以下都在 windows 下操作。

4.1 解压 eclipse

将 eclipse 解压到某个目录下。

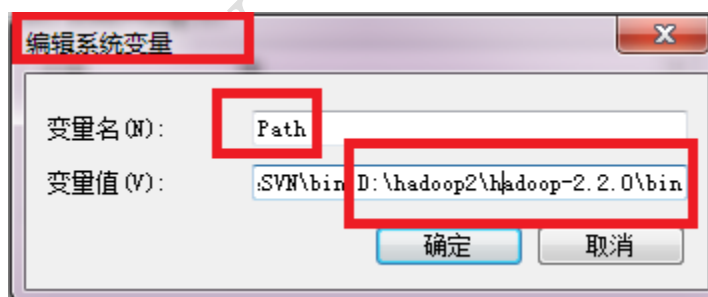
4.2 安装 jdk

安装 jdk-7u80-windows-x64.exe。

4.3 解压 hadoop 安装包

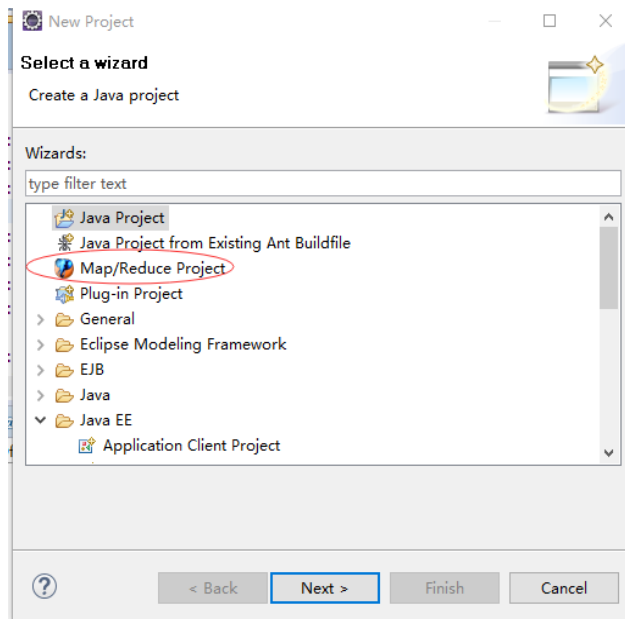
将 hadoop 的安装包 hadoop-2.7.3.tar.gz 解压到某个目录。比如 D:\software\apache\hadoop-2.7.3。

设置 windows 下的环境变量，HADOOP_HOME 和 PATH

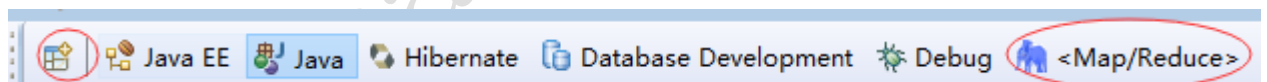


4.4 eclipse 的 hadoop 插件安装

将 eclipse 插件 `hadoop-eclipse-plugin-2.7.3.jar`，拷贝到 eclipse 的 `plugins` 目录，重启 eclipse，即可看到创建工程时有 Map/Reduce Project 类型的工程。

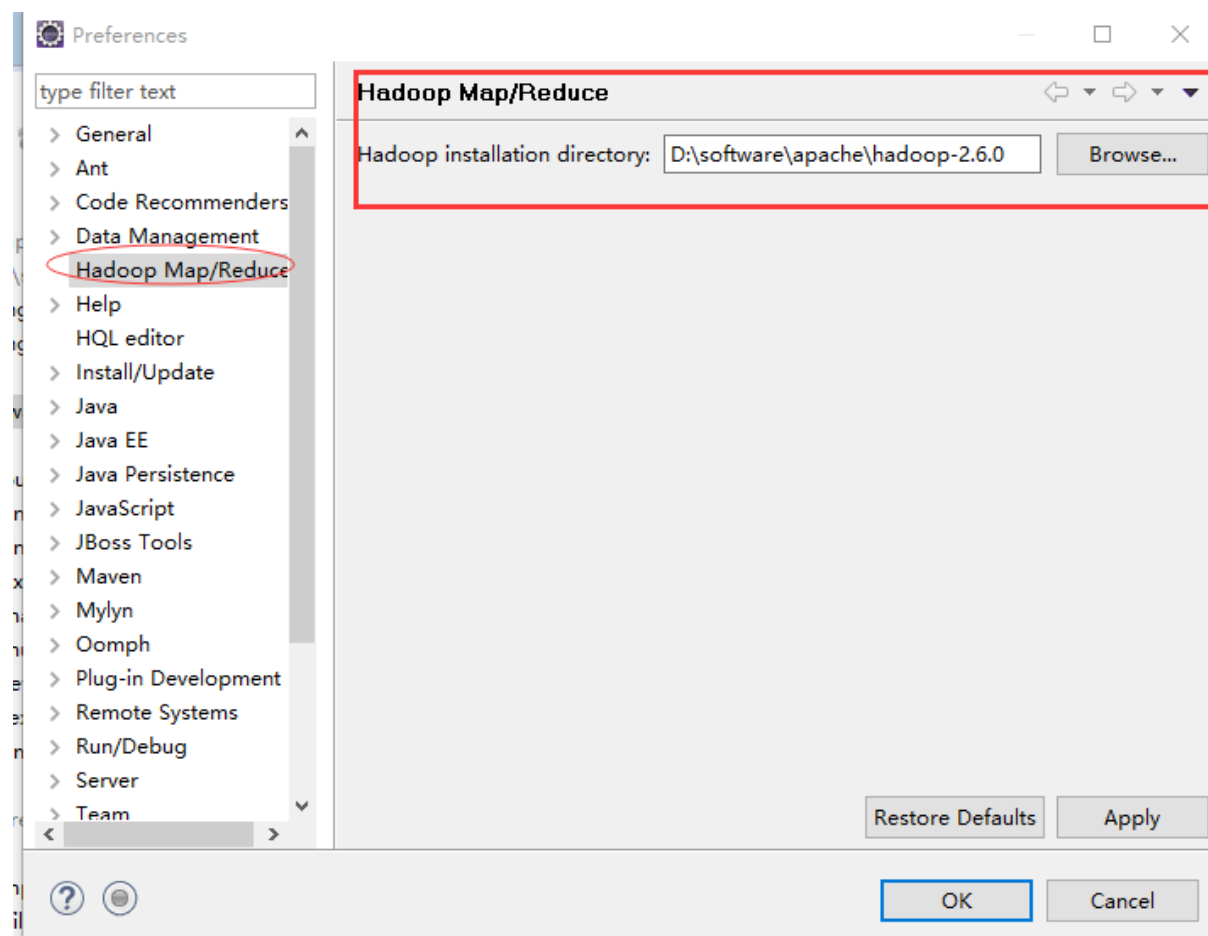


同时工程视图可以选择到 Map/Reduce



4.5 配置 eclipse 中的 hadoop 目录

配置 eclipse 环境的 hadoop 目录：在 eclipse 的 `Windows->Preferences` 的 `Hadoop Map/Reduce` 中设置安装目录，这个目录就是上面步骤在本机解压 hadoop 的主目录



4.6 部署一些 hadoop 的程序库等

将下载的 hadoop-common-2.7.1-bin-master.zip 包内的 hadoop.dll、libwinutils.lib、winutils.exe 拷贝到解压到本地的 hadoop 目录下的 bin 目录，比如 D:\software\apache\hadoop-2.7.3\bin

4.7 创建 mapreduce 工程

该步骤请创建为 MapReduce 类工程，否则若选择为普通的 java project，那么就需要自己手工引入 hadoop 下的一堆 jar 包，难免会有遗漏的情况。

如下图创建的 MapReduce 工程，默认的使用上面配置好的 hadoop 目录，创建工程时会将该版本下 hadoop 相关的包自动引入到 lib 环境中。

New MapReduce Project Wizard

MapReduce Project

Create a MapReduce project.

Project name:

☒ Use default location

Location: [Browse...](#)

Choose file system:

Hadoop MapReduce Library Installation Path

☒ Use default Hadoop [Configure Hadoop install directory...](#)

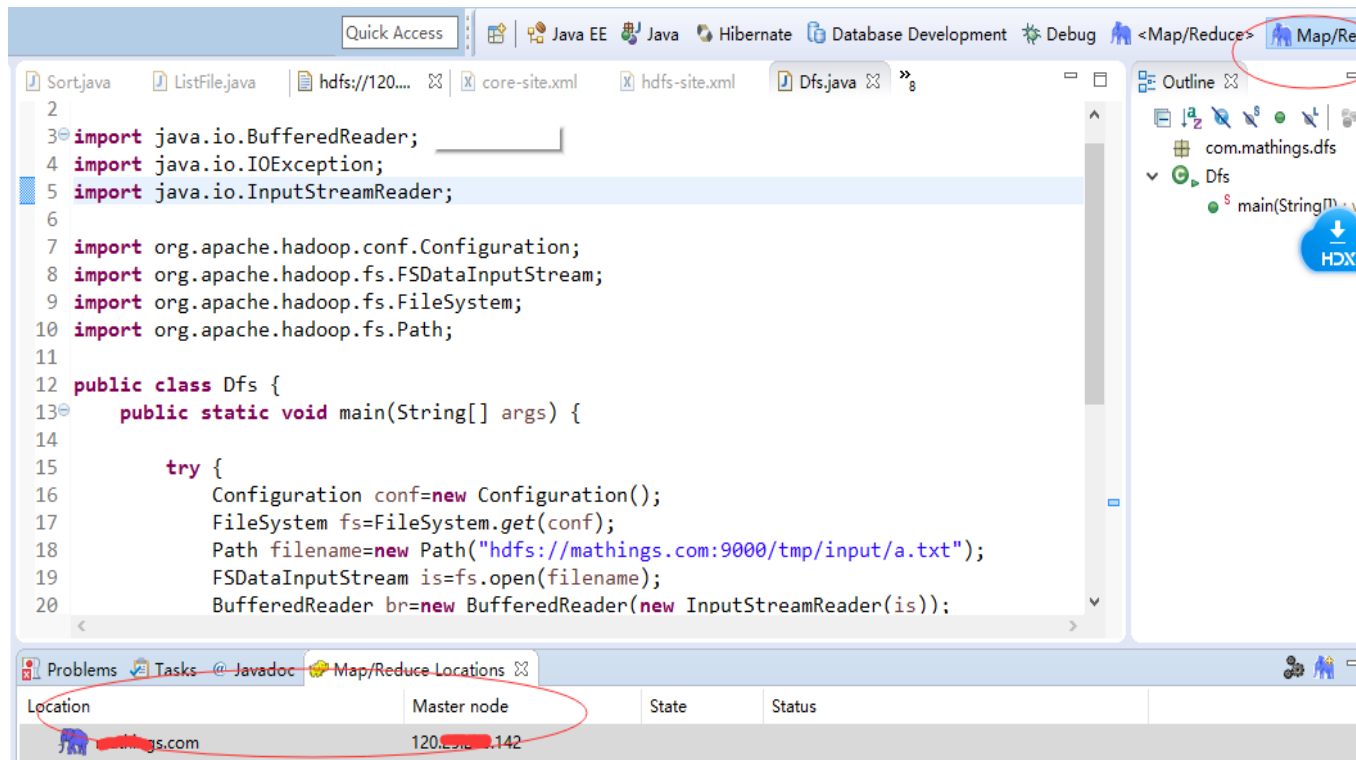
☐ Specify Hadoop library location [Browse...](#)

[?](#) [< Back](#) [Next >](#) [Finish](#) [Cancel](#)

- bigData
 - src
 - JRE System Library [jdk1.7.0_80]
 - zookeeper-3.4.6.jar - D:\software\apache\zookeeper-3.4.6\zookeeper-3.4.6.jar
 - xz-1.0.jar - D:\software\apache\xz-1.0\xz-1.0.jar
 - stax-api-1.0-2.jar - D:\software\apache\stax-api-1.0-2\stax-api-1.0-2.jar
 - javax.servlet-api-2.5.jar - D:\software\apache\javax.servlet-api-2.5\javax.servlet-api-2.5.jar
 - protobuf-java-2.5.0.jar - D:\software\apache\protobuf-java-2.5.0\protobuf-java-2.5.0.jar
 - netty-3.6.2.Final.jar - D:\software\apache\netty-3.6.2.Final\netty-3.6.2.Final.jar
 - log4j-1.2.17.jar - D:\software\apache\log4j-1.2.17\log4j-1.2.17.jar
 - leveldbjni-all-1.8.jar - D:\software\apache\leveldbjni-all-1.8\leveldbjni-all-1.8.jar
 - jsr305-1.3.9.jar - D:\software\apache\jsr305-1.3.9\jsr305-1.3.9.jar
 - jline-0.9.94.jar - D:\software\apache\jline-0.9.94\jline-0.9.94.jar
 - jetty-util-6.1.26.jar - D:\software\apache\jetty-util-6.1.26\jetty-util-6.1.26.jar
 - jetty-6.1.26.jar - D:\software\apache\jetty-6.1.26\jetty-6.1.26.jar
 - jettison-1.1.jar - D:\software\apache\jettison-1.1\jettison-1.1.jar
 - jersey-server-1.9.jar - D:\software\apache\jersey-server-1.9\jersey-server-1.9.jar
 - jersey-json-1.9.jar - D:\software\apache\jersey-json-1.9\jersey-json-1.9.jar
 - jersey-guice-1.9.jar - D:\software\apache\jersey-guice-1.9\jersey-guice-1.9.jar
 - jersey-core-1.9.jar - D:\software\apache\jersey-core-1.9\jersey-core-1.9.jar
 - jersey-client-1.9.jar - D:\software\apache\jersey-client-1.9\jersey-client-1.9.jar

4.8 创建 Mapreduce Location

若需要通过 eclipse 浏览 hadoop 下的文件，那么可以通过 MapReduce 工程视图创建 mapreduce location



Edit Hadoop location...

Define Hadoop location

Define the location of a Hadoop infrastructure for running MapReduce applications.

General Advanced parameters

Location name: [redacted].com

Map/Reduce(V2) Master

Host: [redacted]

Port: 50020

DFS Master

☐ Use M/R Master host

Host: [redacted]

Port: 9000

User name: hadoop

SOCKS proxy

☐ Enable SOCKS proxy

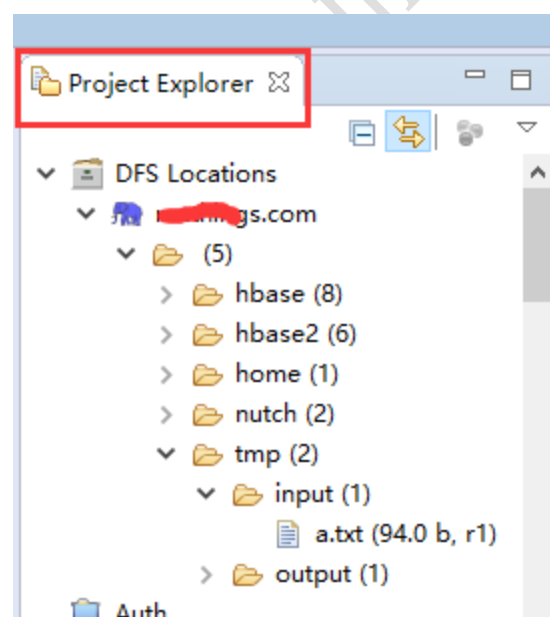
Host: host

Port: 1080

Load from file Validate location

Finish Cancel

配置完，若配置的正确，那么可以在工程浏览器下查看服务器端的 HDFS 目录与文件信息了。



4.9 部署 hadoop 配置文件

将服务器上的 `hdfs-site.xml`、`core-site.xml`、`mapred-site.xml`、`yarn-site.xml` 放到工程的 `bin` 下，将 `log4j.properties` 配置也放到工程的 `src` 和 `bin` 下。

4.10 修改 hadoop 源码

在工程中创建 `org.apache.hadoop.mapred.YARNRunner` 类，具体的代码使用云盘中的 `YARNRunner.java`。

4.11 开始 HDFS 编码

经过以上安装配置后就可以开始 hadoop 编码了，一个实例如下：

```
package com.mars.mr;

import java.io.BufferedReader;
import java.io.IOException;
import java.io.InputStreamReader;

import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.FSDataInputStream;
import org.apache.hadoop.fs.FileSystem;
import org.apache.hadoop.fs.Path;

public class ListFile {
    public static void main(String[] args) {

        try {
            Configuration conf=new Configuration();
            FileSystem fs=FileSystem.get(conf);
            Path filename=new
Path("hdfs://mars.com:9000/tmp/input/a.txt");
            FSDataInputStream is=fs.open(filename);
```

```
        BufferedReader br=new BufferedReader(new
InputStreamReader(is));
        String content=br.readLine();
        System.out.println(content);
        br.close();
        fs.close();

    } catch (IOException e) {
        // TODO Auto-generated catch block
        e.printStackTrace();
    }

}
```

再运行之，实际上运行成 JAVA Applicaiton 或者 Run on Hadoop 都可以正常执行。

4.12 开始 mapreduce 编码

参考云盘中的 WordCount. java