Eclipse Hadoop环境配置

机群配置基本信息

Ubuntu 10.04
Hadoop 0.20.2
1 个 namenode 162.105.146.30, 4 个 datanode 162.105.146.30~33

客户端配置

该文档测试时使用：
Eclipse 3.51
Hadoop 0.20.2

配置步骤

1. Eclipse plugin

a) 如果使用 Eclipse 3.5

下载
http://code.google.com/p/hadoop-eclipse-plugin/downloads/detail?name=hadoop-0.20.3-dev-eclipse-plugin.jar&can=2&q=
重命名为: hadoop-0.20.2-eclipse-plugin.jar
放在 Eclipse plugins 目录下面.（我的 ubuntu 该目录在/usr/lib/eclipse/plugins）

b) 如果使用 Eclipse 3.3 一下，可以选择使用 hadoop/contrib/eclipse-plugin 目录下面的 jar 文件

（这个我没有测试，可以自己试一下）

2. 编辑/etc/hosts (windows xp 目录是：C:\WINDOWS\system32\drivers\etc\hosts)

添加一行：162.105.146.30 sewm sewm

4. 下载 hadoop,解压到本地目录

5. 启动 Eclipse

6. 设置 hadoop 路径

Windows→Preferences→hadoop directory
指向解压的 hadoop 目录，编译的时候需要里面的 jar 文件

7. 切换到 Mapreduce Perspective

Windows→Open Perspective→Other



Linux公社（LinuxIDC.com）于2006年9月25日注册并开通网站，Linux现在已经成为一种广受关注和支持的一种操作系统，IDC是互联网数据中心，LinuxIDC就是关于Linux的数据中心。

LinuxIDC.com提供包括Ubuntu，Fedora，SUSE技术，以及最新IT资讯等Linux专业类网站。

并被收录到Google 网页目录-计算机 > 软件 > 操作系统 > Linux 目录下。

Linux公社（LinuxIDC.com）设置了有一定影响力的Linux专题栏目。

包括:

Ubuntu专题

Fedora专题

RedHat专题

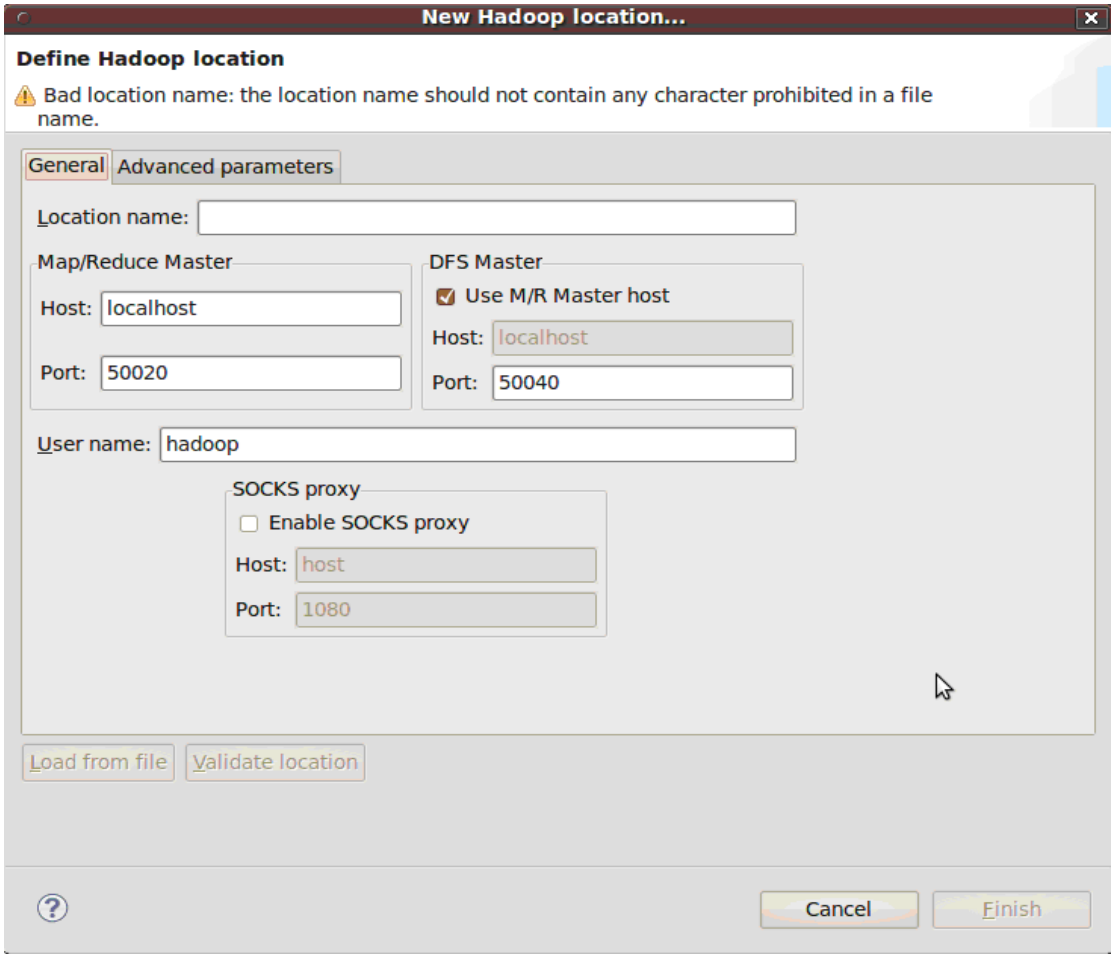
SUSE专题

红旗Linux专题

Android专题

本站（LinuxIDC）所刊载文章不代表同意其说法或描述，仅为提供更多信息，也不构成任何建议。
本站带宽由[6688.CC]友情提供
Copyright © 2006-2011 Linux公社 All rights reserved

8. 添加 Hadoop location



Location name: 随意指定

Map/Reduce Master:

Host: 162.105.146.30

Port: 9001

DFS Master:

Host: 162.105.146.30 （选中 User M/R Master host 即可）

Port: 9000

User name: hadoop

需要更改 Advance parameters 中的 hadoop.job.ugi，默认的好像是 DrWho,Tardis，将逗号前面的部分更改为 hadoop，最后：hadoop,Tardis

（如果运行 eclipse 的用户名是 hadoop，即与机群的账户名一致，不用更改这一项）

(在 New Location 的时候，可能看不到这个选项，一般情况下，新建完成后，选择新建的项，Edit Location，这个时候可以看到这一项，如果还是解决不了，重新启动 Eclipse，选择 Edit Location)

Finish 即可，在 Eclipse 中可以浏览到 hadoop dfs 目录

上传和输出的文件都在：

user/hadoop/ 目录下面

9. 添加 Project

添加 MapReduce Project，添加相应的 Maper, Reducer, Driver,

修改 Driver 中的输出文件夹名称，

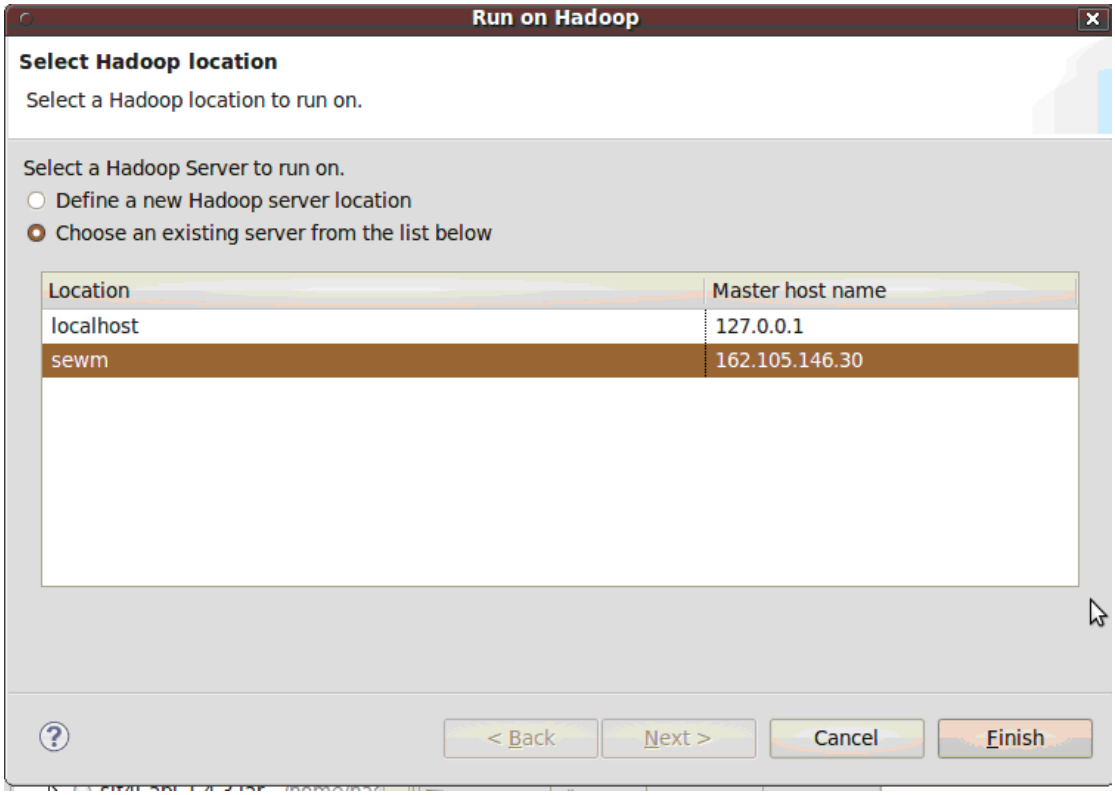
输出文件夹的命名：

学号_lab#_output，例如：10948244_lab1_output

如果多次试验，可以在 output 后面加一些序号之类的。

10. 运行

Run As—> Run on Hadoop→



选择 hadoop location, finish.

11. 查看输出结果

下载文件夹 /user/hadoop/{your output dir} 下面的文件

第一次实验

输入文件已经上传到 DFS，文件夹为 input（不要再上传文件）

如果希望上传自己的文件：

命名： 学号_lab#_input

输出文件夹，修改 MapReduce driver 程序中的输出文件夹