

馴服紫色大象以及牠的好朋友們

Hadoop & MapReduce

資管碩一 戴睿宸



Frist Step: 挑好抓的大象

- Hadoop 目前已經改到 0.20.205.0
- 我們使用 Hadoop 0.20.2

要裝多台Hadoop的話(單台跳過)

- `sudo vi/etc/hostname`

改成你要的名字，區網內不要重複

EX:

第一台 handsomeboy11

第二台 handsomeboy22

- `Sudo vi /etc/hostname`

加你要連的機器ip和名字進入

EX:

192.168.0.100 handsomeboy11

192.168.0.101 handsomeboy22

192.168.0.102 handsomeboy33

要裝多台Hadoop的話(單台跳過)

- 上面設定完之後ping看看有沒有成功

```
$ping handsomeboy22
```

```
$ping handsomeboy11
```

改ssh設定

```
$ sudo vi /etc/ssh/ssh_config
```

找到

```
#StrictHostKeyChecking ask
```

改成

```
StrictHostKeyChecking no
```

Hadoop 安裝指令

- 先裝個java

指令(多台的話兩台都要裝)

-
- `$sudo apt-get install java-common sun-java6-bin sun-java6-jdk sun-java6-jre`

Hadoop 安裝指令

- 設定SSH登入免密碼

```
$ apt-get install ssh
$ ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
$ cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
$ scp -r ~/.ssh handsomeboy22:~/ (多台的話)
$ ssh localhost
```

沒出現要密碼就對了

```
$exit
```

Hadoop 安裝指令(紅色地方要改)

```
$wget http://apache.ntu.edu.tw/hadoop/core/  
hadoop-0.20.2/hadoop-0.20.2.tar.gz
```

```
$ tar zxvf hadoop-0.20.2.tar.gz
```

```
$ sudo mv hadoop-0.20.2 /opt/
```

```
$ sudo chown -R user:group /opt/hadoop-0.20.2
```

```
$ sudo ln -sf /opt/hadoop-0.20.2 /opt/hadoop
```

Hadoop 安裝指令 (vi可以換成gedit)

- `$sudo vi /opt/hadoop/conf/hadoop-env.sh`
- 然後貼上

```
export JAVA_HOME=/usr/lib/jvm/java-6-sun
export HADOOP_HOME=/opt/hadoop
export PATH=$PATH:/opt/hadoop/bin
```


Hadoop 安裝指令 (vi可以換成gedit)

```
$sudo vi /opt/hadoop/conf/core-site.xml
```

```
<configuration>  
  <property>  
    <name>fs.default.name</name>  
    <value>hdfs://localhost:9000</value>  
  </property>  
  <property>  
    <name>hadoop.tmp.dir</name>  
    <value>/tmp/hadoop/hadoop-`${user.name}</value>  
  </property>  
</configuration>
```

Hadoop 安裝指令 (vi可以換成gedit)

```
$sudo vi /opt/hadoop/conf/mapred-site.xml
```

```
<configuration>
```

```
  <property>
```

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

```
    <value>localhost:9001</value>
```

```
  </property>
```

```
</configuration>
```

Hadoop 安裝指令(如果單台 跳過這張)

```
$cd /opt/hadoop
```

```
$vi conf/slaves
```

- 加入你想要當作Datanode的hostname

EX:

```
handsomeboy11
```

```
handsomeboy22
```

Hadoop 安裝指令 (每個node2都要改)

```
$ ssh handsomeboy22
```

```
$sudo mkdir /opt/hadoop
```

```
$sudo chown -R User:Group /opt/hadoop
```

```
$sudo mkdir /var/hadoop
```

```
$sudo chown -R User:Group /var/hadoop
```

```
$exit
```

請確認已回到本機

```
$cd /opt/hadoop
```

```
$ scp -r /opt/hadoop/* handsomeboy22:/opt/hadoop/
```



Start Hadoop

```
$cd /opt/hadoop
```

```
$ bin/hadoop namenode -format
```

```
$ bin/start-all.sh
```

連50030跟50070看有沒有成功

Step2 環境設定

- 由於大象很難抓，所以要用好一點的工具來抓大象，在這邊我們選擇 **eclipse** 跟 **hadoop Plugin**

下載該載的東西

- Eclipse (64bit OS)

```
$cd ~
```

```
$wget http://archive.eclipse.org/eclipse/downloads/drops/  
R-3.3.2-200802211800/eclipse-SDK-3.3.2-linux-gtk-  
x86_64.tar.gz
```

- Jave Doc(沒辦法wget)
- <http://www.oracle.com/technetwork/java/javase/downloads/jdk-6u25-doc-download-355137.html>

安裝eclipse

```
$cd ~
```

```
$tar -zxvf eclipse-SDK-3.3.2-linux-gtk-x86_64.tar.gz
```

```
$sudo mv eclipse /opt
```

```
$sudo ln -sf /opt/eclipse/eclipse /usr/local/bin/
```


安裝hadoop Plugin

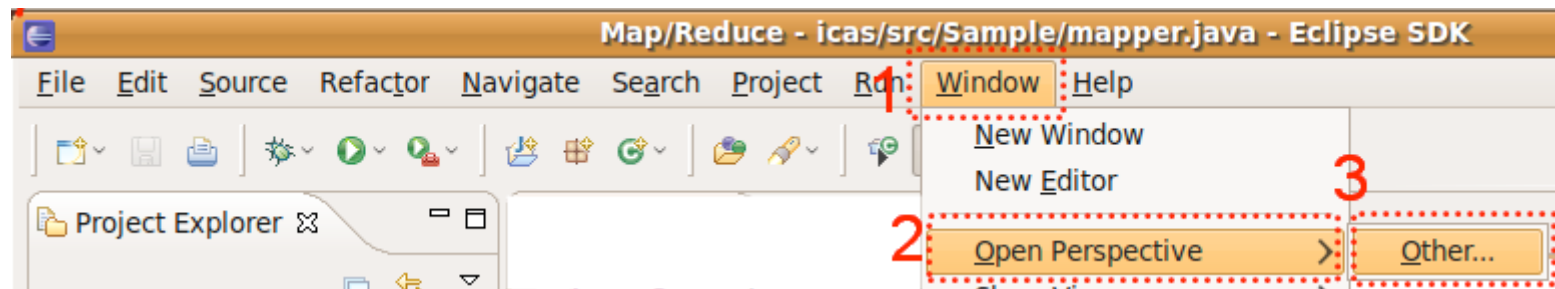
```
$ cd /opt/hadoop
```

```
$ sudo cp /opt/hadoop/contrib/eclipse-plugin/  
hadoop-0.20.2-eclipse-plugin.jar /opt/eclipse/plugins
```

打開eclipse

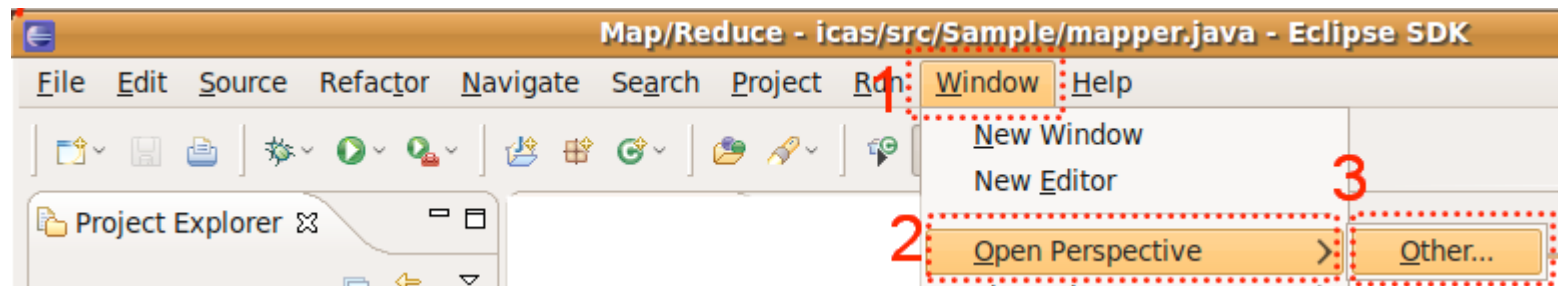
```
$eclipse &
```

Eclipse setting

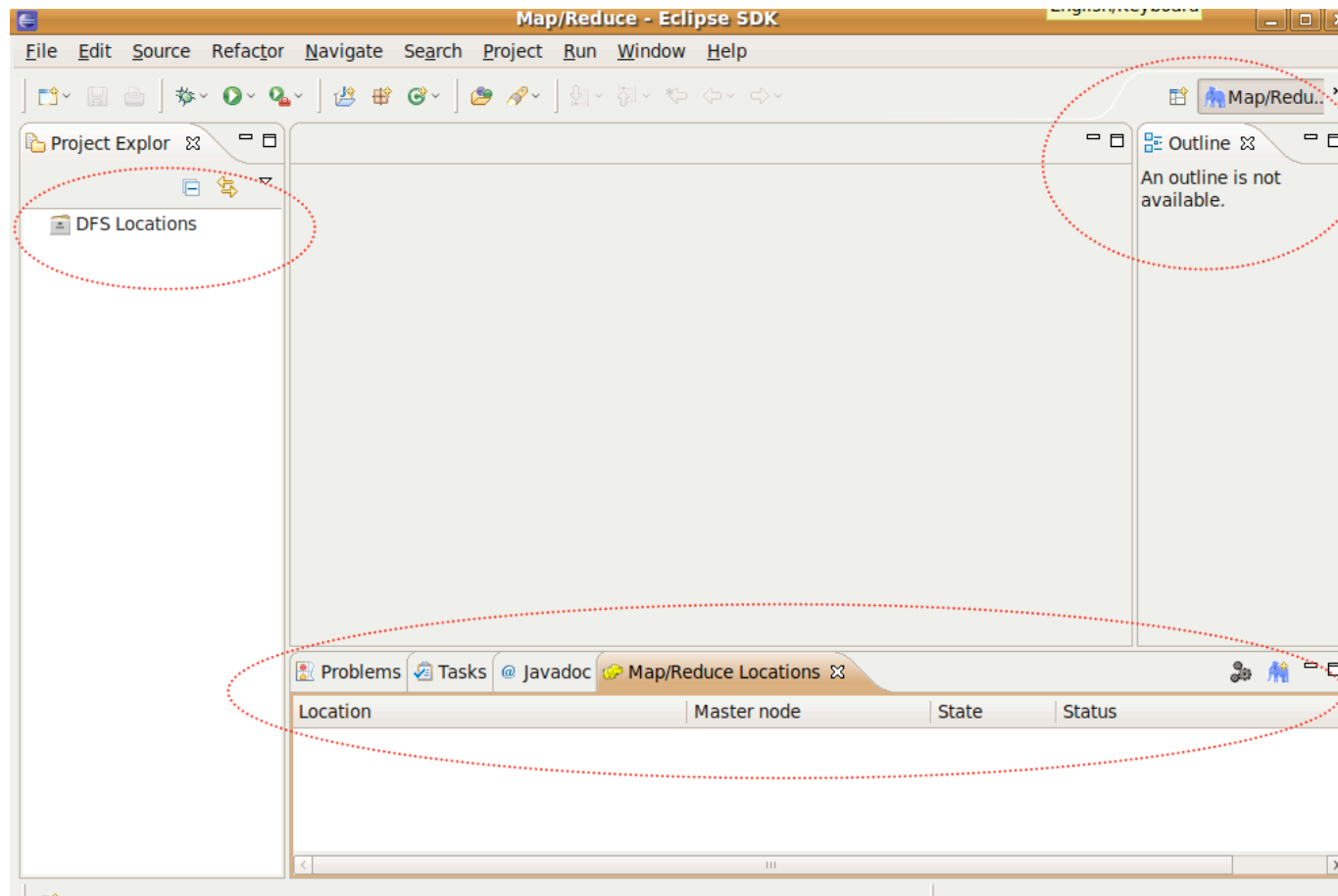


Eclipse setting

- Window->Open Perspective->Other->MapReduce

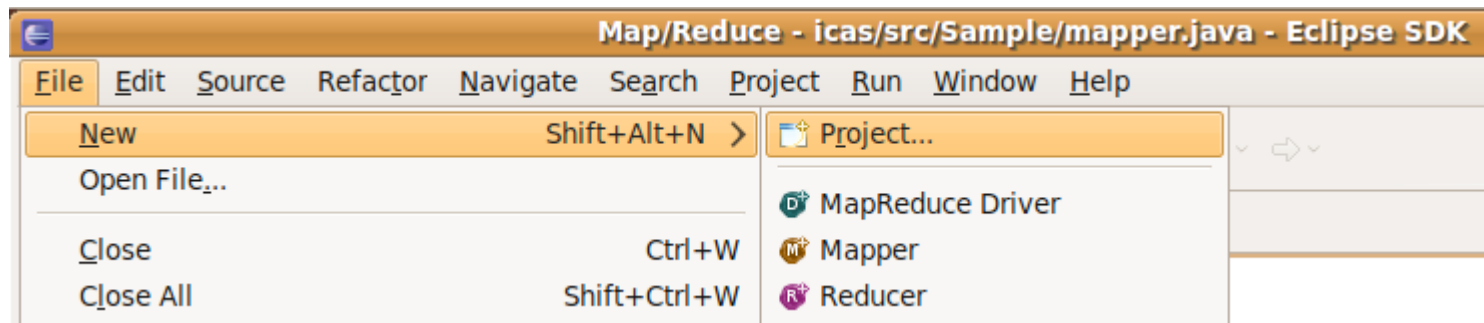


Eclipse setting

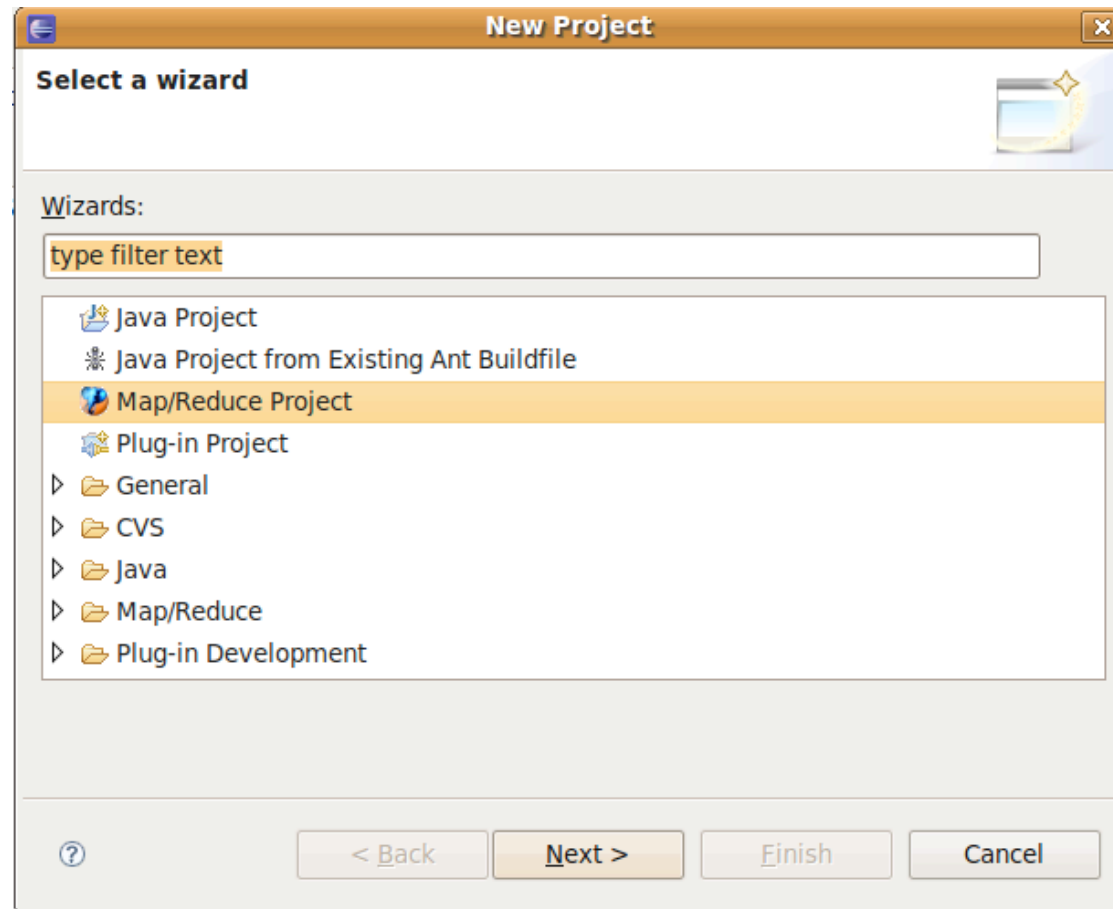


Eclipse setting

- 建立專案
- File->New->Project->MapReduce Project



Eclipse setting



Eclipse setting

- project name-> 隨意
- use default hadoop -> Configur Hadoop install...
-> 輸入："/opt/hadoop" -> ok Finish

Eclipse setting

New MapReduce Project Wizard

MapReduce Project
Create a MapReduce project.

Project name:

Use default location

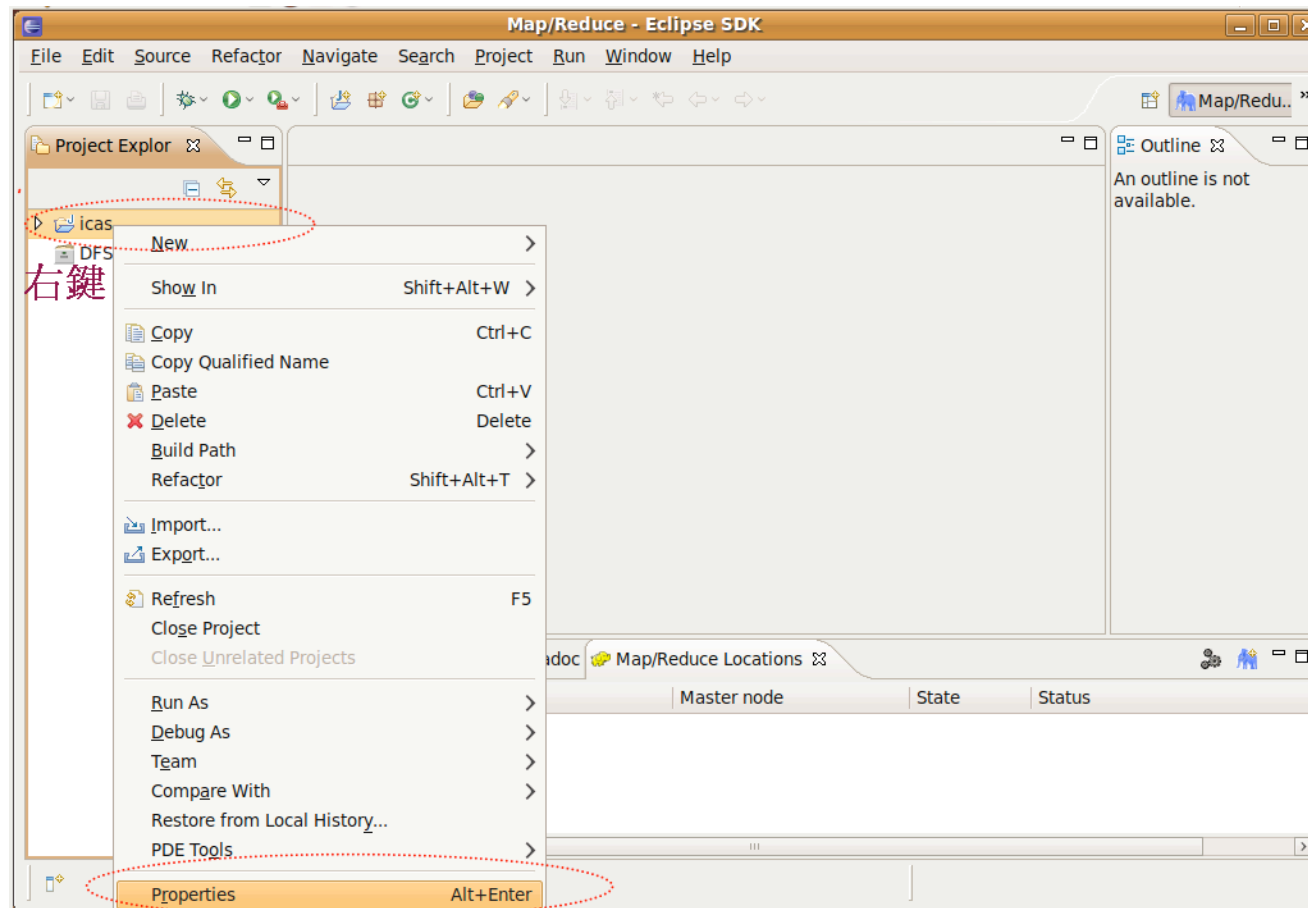
Location:

Hadoop MapReduce Library Installation Path

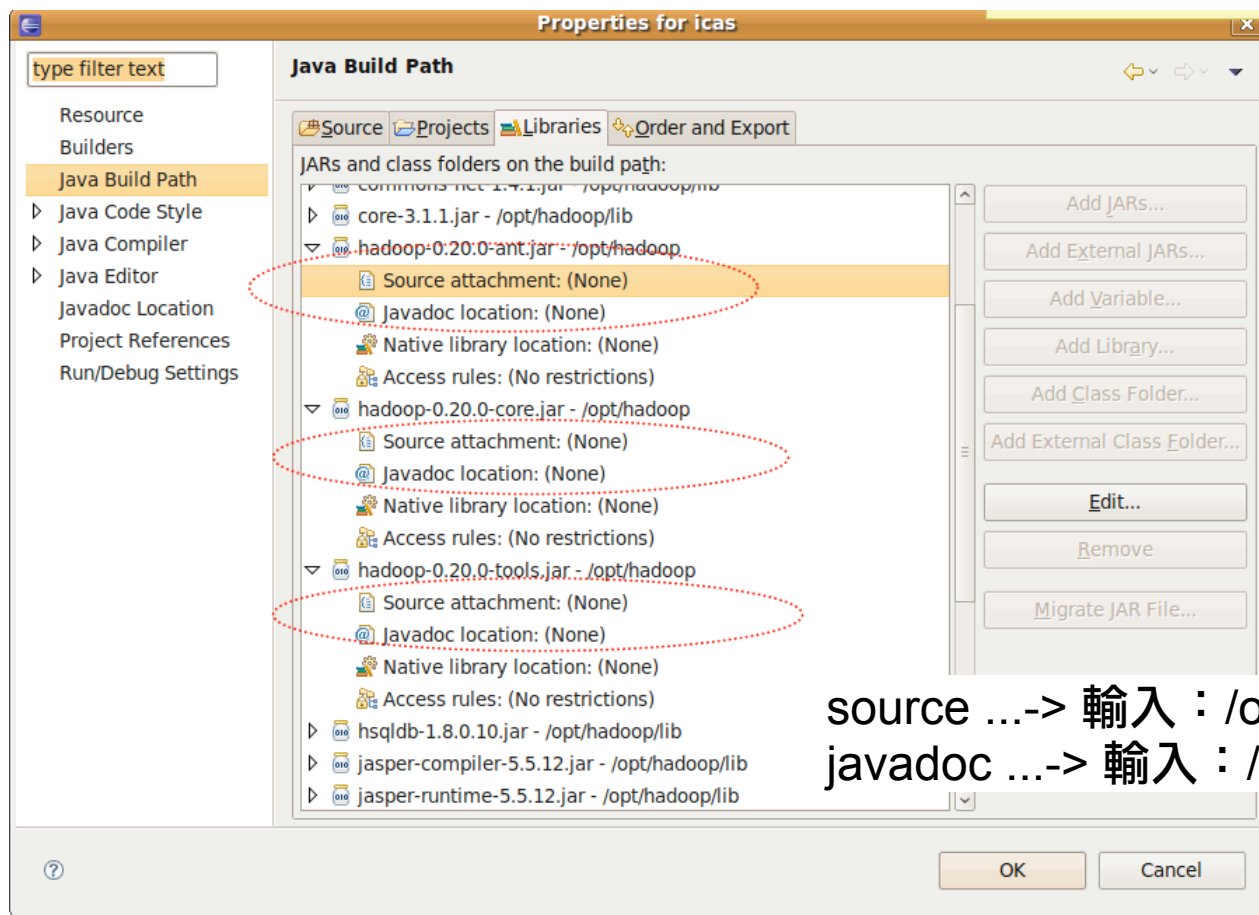
Use default Hadoop [Configure Hadoop install directory...](#)

Specify Hadoop library location

Eclipse setting

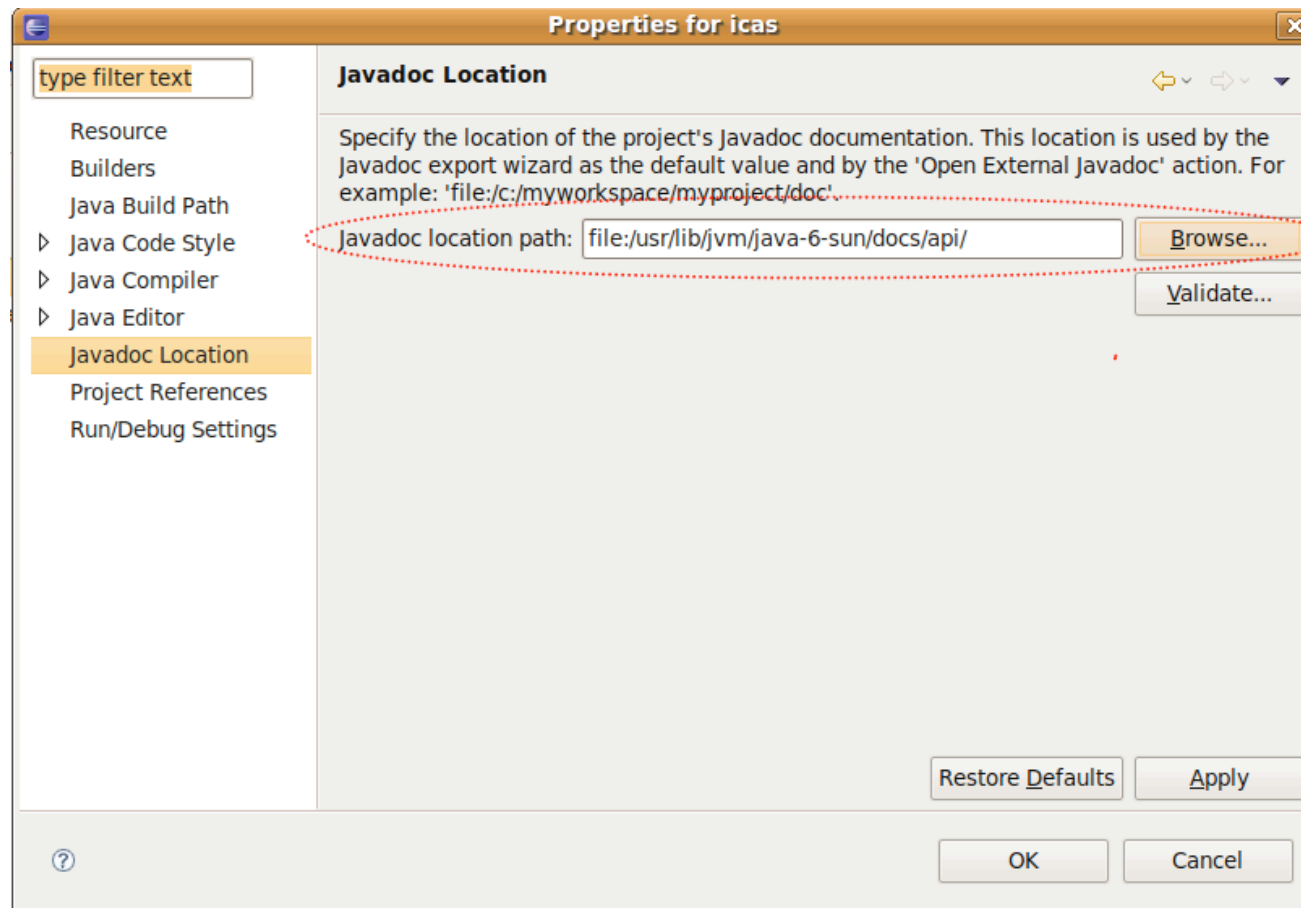


Eclipse setting

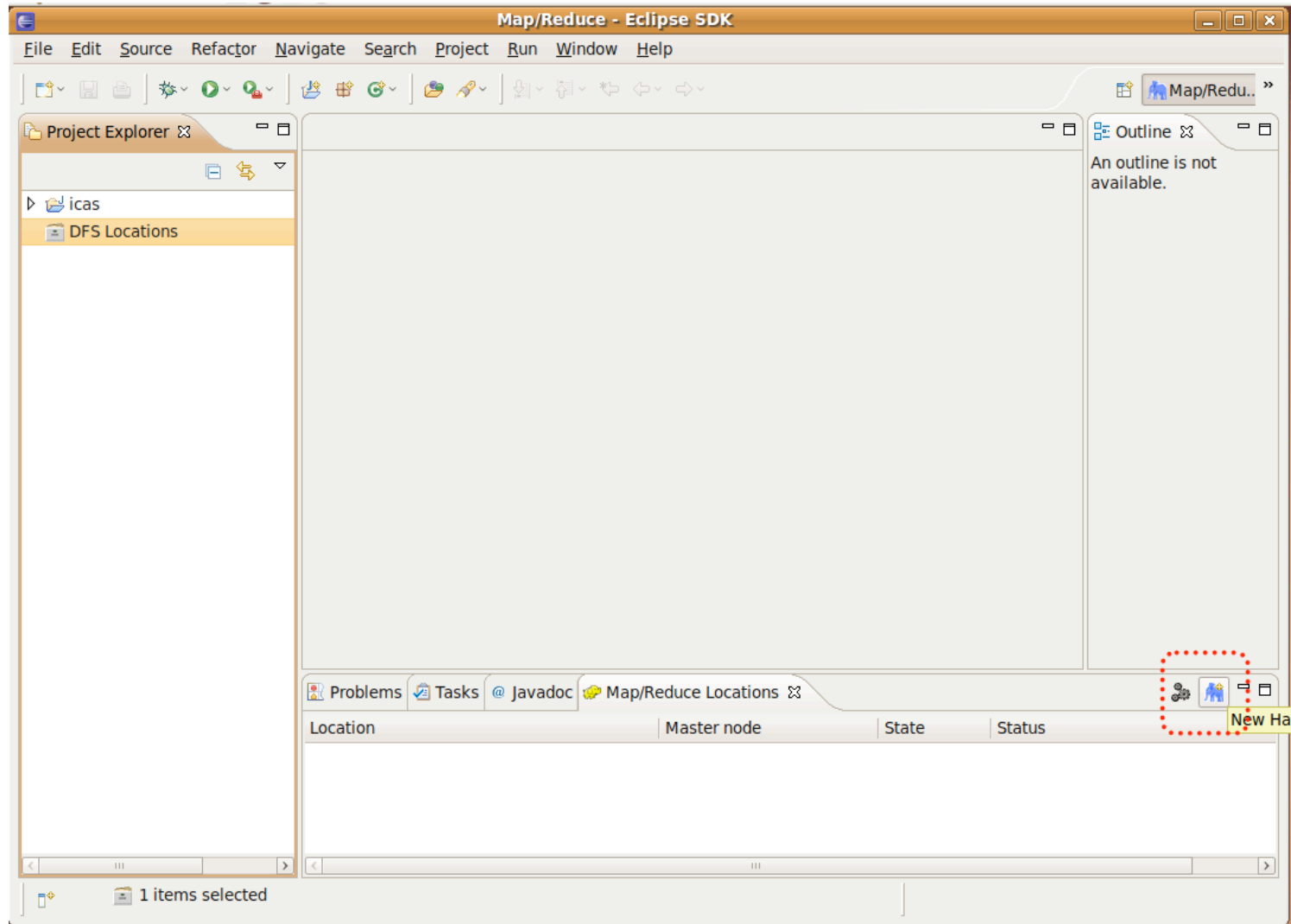


source ...-> 輸入 : /opt/hadoop/src/
javadoc ...-> 輸入 : /opt/hadoop/docs/api/

Eclipse setting



找大象-連接HDFS



找大象-連接HDFS

打你namenode的IP

New Hadoop location...

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

General | Advanced parameters

Location name: hadoop

Map/Reduce Master
Host: localhost
Port: 9001

DFS Master
 Use M/R Master host
Host: localhost
Port: 9000

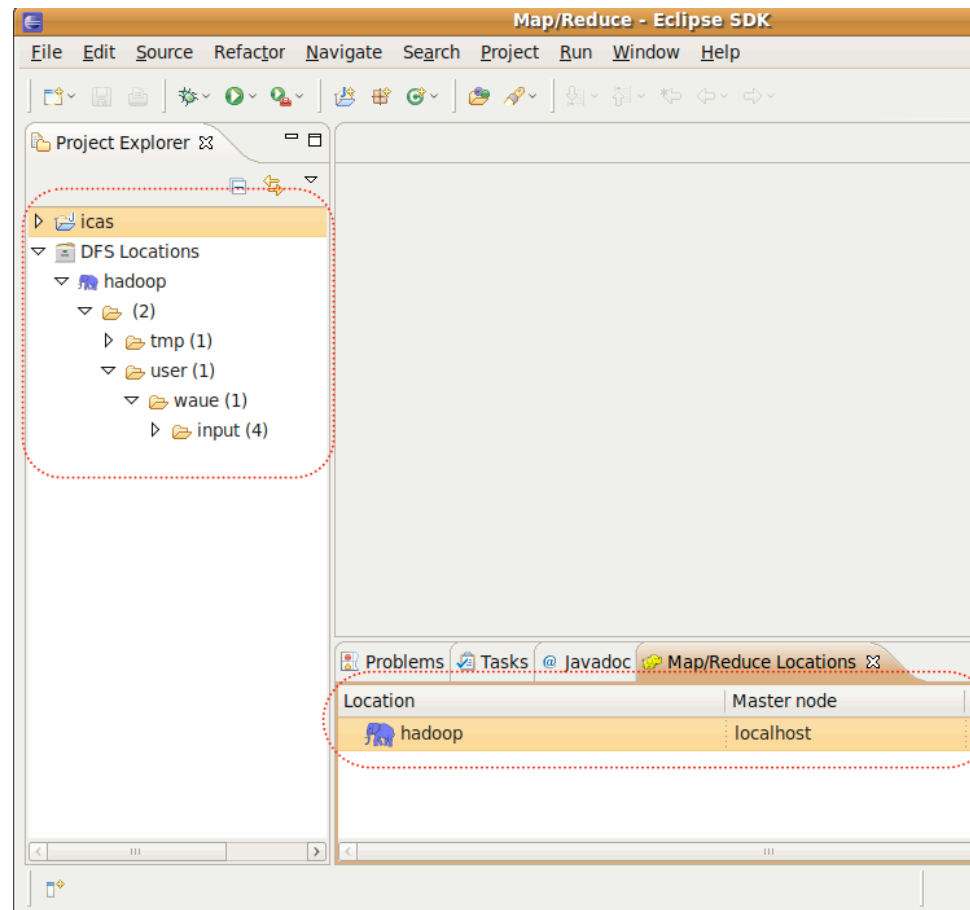
User name: waue

SOCKS proxy
 Enable SOCKS proxy
Host: host
Port: 1080

Load from file | Validate location

Finish | Cancel

找大象-連接HDFS





獵大象



END