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-java6bin 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:~/ (多合的話) \$ 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-gtkx86_64.tar.gz

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



\$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 &

e	N.	4ap/Redu	ce - ica	as/src	/Sample/	(mapper.java - E	clipse SDK
<u>F</u> ile <u>E</u> dit <u>S</u> ource Refac <u>t</u> or	<u>N</u> avigate	Se <u>a</u> rch <u>P</u>	roject	<u>R</u> an	<u>W</u> indow	<u>H</u> elp	
] ◘ × 🗟 👜] 🏇 × O × 💁 v	∕] 🖄 🕸	©~] 🥭	% ~]	P	<u>N</u> ew W New <u>E</u>	/indow ditor	3
Project Explorer 🛚 🗧 🖕				2	 Open F	erspective	> <u>O</u> ther

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





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

E	Map/Reduce - icas/src/Sample/mapper.java - Eclipse SDK
<u>File</u> <u>E</u> dit <u>S</u> ource Refac <u>t</u> or	<u>N</u> avigate Se <u>a</u> rch <u>P</u> roject <u>R</u> un <u>W</u> indow <u>H</u> elp
<u>N</u> ew	Shift+Alt+N > T Project
Open File	MapBeduce Driver
Close	Ctrl+W @ Mapper
Close All	Shift+Ctrl+W @ Reducer

E New Project	×
Select a wizard	\$
<u>W</u> izards:	
type filter text]
 Java Project Java Project from Existing Ant Buildfile Map/Reduce Project Plug-in Project General CVS CVS Java Map/Reduce Plug-in Development 	
(?) < <u>Back</u> <u>Next</u> > <u>Finish</u> Cance	el

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

🧲 New Ma	apReduce Project Wizard	×
MapReduce Project		
Create a MapReduce project.		
Project name: icas		
✓ Use <u>d</u> efault location		
Location: /home/waue/workspace/icas		Browse
Hadoop MapReduce Library Installation Path		
Use default Hadoop	Configure Hadoop install	directory
Specify Hadoop library location	Browse	
	Biombern	
\bigcirc	< Back Next > Finish	Cancel

							Maj	/Red	uce - Ecli	ipse SDK					_		×
Ei	le	<u>E</u> dit	<u>S</u> ource	Refac <u>t</u> or	<u>N</u> avigate	Se <u>a</u> rch	<u>P</u> roject	<u>R</u> un	<u>W</u> indow	<u>H</u> elp							
]	<u></u> *~		۵	~ () ~ Q	·] 🖄 🖶	@ ~]	🗢 🔗	_ <u>}</u> - ∠	₩~ *>					EP 月	Map/F	edu	»
P) Pr	oject	Explor 🕅										- 0	🗄 Outli	ne 🛛	-	
	J	icas	E	\$ ₹				_						An outlir available	ne is not e.		
	1	DFS	<u>N</u> ew		and the second		>										
花	I)	键	Sho <u>w</u>	In		Shift+A	lt+W >										
			📄 <u>С</u> ору				Ctrl+C										
			🗎 Сору	Qualified N	lame												
			💼 <u>P</u> aste				Ctrl+V										
			💢 <u>D</u> elet	e			Delete										
			<u>B</u> uild	Path			>										
			Refac	<u>t</u> or		Shift+	Alt+T >										
			占 Impor	t													
			🛃 Expor	t													
			8 Refre	sh			F5										
			Clo <u>s</u> e	Project													
			Close	<u>U</u> nrelated	Projects			idoc 🕯	🥟 Map/Re	educe Locations	8				an 🕴	† –	
			<u>R</u> un A	S			>			Master node		State	Status				
			<u>D</u> ebug	g As			>										
			T <u>e</u> am				>										
			Comp	are With			>										
			Resto	re from Lo	cal History												
			PDE T	ools			>			Ш							>
	•	\leq	P <u>r</u> ope	rties		A	lt+Enter		>								

	Properties for icas	×	
type filter text	Java Build Path	⇔ → →	
Resource Builders	Bource Projects Libraries Order and Export		
Java Build Path	JARS and class folders on the build path:		
Java Code Style	▷ 👼 core-3.1.1.jar - /opt/hadoop/lib	Add JARs	
Java Compiler	🗢 👼 hadoop=0:20:0-ant.jar - /opt/hadoop	Add External JARs	
Java Editor	Source attachment: (None)	Add Variabla	
Javadoc Location	🧖 Javadoc location: (None)		
Project References	📓 Native library location: (None)	Add Libr <u>a</u> ry	
Run/Debug Settings	Access rules: (No restrictions)	Add Class Folder	
	Image: mail the second seco		
	Source attachment: (None)	Add External Class <u>F</u> older	
	Javadoc location: (None)		
	📽 Native library location: (None)	<u>E</u> dit	
	े Access rules: (No restrictions)	Remove	
	hadoop-0.20.0-tools.jar - /opt/hadoop		
	🕼 Source attachment: (None)	Migrate JAR File	
	📓 Native library location: (None)		
	Access rules: (No restrictions)	source \rightarrow 脑入:/o	nt/hadoon/src/
	Isqldb-1.8.0.10.jar - /opt/hadoop/lib	Source> 前八・/0	punauoopisici
	jasper-compiler-5.5.12.jar - /opt/hadoop/lib	iavadoc -> 輸入:/	ont/hadoon/docs/ani/
	jasper-runtime-5.5.12.jar - /opt/hadoop/lib		
0		OK Cancel	

		Properties for icas	×
typ	pe filter text	Javadoc Location	(
	Resource Builders Java Build Path	Specify the location of the project's Javadoc documentation. This locatio Javadoc export wizard as the default value and by the 'Open External Jav example: 'file:/c:/myworkspace/myproject/doc'.	n is used by the vadoc' action. For
⊳	Java Code Style	Javadoc location path: file:/usr/lib/jvm/java-6-sun/docs/api/	<u>B</u> rowse
⊳	Java Editor		Validate
	Javadoc Location		•
	Run/Debug Settings		
		Restore Default	s Apply
?)	ОК	Cancel

找大象-連接HDFS

Ele Edit Source Refactor Navigate Search Project Bun Window Help	6	Map/Reduce - Eclipse SDK	- • × •
Project Explorer S Project Explorer S <td><u>F</u>ile <u>E</u>dit <u>S</u>ource Refac<u>t</u>or <u>N</u></td> <td>avigate Se<u>a</u>rch <u>P</u>roject <u>R</u>un <u>W</u>indow <u>H</u>elp</td> <td></td>	<u>F</u> ile <u>E</u> dit <u>S</u> ource Refac <u>t</u> or <u>N</u>	avigate Se <u>a</u> rch <u>P</u> roject <u>R</u> un <u>W</u> indow <u>H</u> elp	
Project Explorer 23 Contine 23 An outline is not available. An outline is not available. Problems Tasks @ Javadoc @ Map/Reduce Locations 23 Location Master node State Status Problems An available. Problems An available. Problems An avail] ◘ - 🔒 📄 🏇 - O - 🂁	28 📽 🞯 -] 🥭 🖋 -] 知、 樹、 🏷 ゆ、 ウィ	🗈 🧥 Map/Redu »
Image: Second State	🕒 Project Explorer 🛚 🗖 🗖		🗄 Outline 🕱 🗖 🗖
▶ Si icas DPS Locations In the selected	⊑ 🔄 ▽		An outline is not available.
DFS Locations Problems @ Tasks @ Javadoc @ Map/Reduce Locations 23 Docation Master node State Status New Ha Docation Master node State Status	🕨 😂 icas		
Problems @ Tasks @ Javadoc ♥ Map/Reduce Locations 33 Location Master node State Status	DFS Locations		
Problems Tasks @ Javadoc & Map/Reduce Locations X Location Master node State Status New Ha			
Problems Tasks @ Javadoc > Map/Reduce Locations 23 Location Master node State Status New Ha			
Problems Tasks @ Javadoc > Map/Reduce Locations S Location Master node State Status New He			
Problems Tasks @ Javadoc Map/Reduce Locations 3 Location Master node State Status			
Problems @ Tasks @ Javadoc @ Map/Reduce Locations 33 Location Master node State Status New Ha T T T T T T T T T T T T T T T T T T T			
Problems Tasks @ Javadoc @ Map/Reduce Locations X Location Master node State Status New Ha			
Problems Tasks @ Javadoc > Map/Reduce Locations X Location Master node State Status			
Problems Tasks @ Javadoc @ Map/Reduce Locations S3			
Problems Tasks @ Javadoc * Map/Reduce Locations X Location Master node State Status			
Problems Z Tasks @ Javadoc & Map/Reduce Locations X Location Master node State Status New Ha I items selected			
Problems Tasks @ Javadoc P Map/Reduce Locations X Location Master node State Status New Ha New Ha New Ha I items selected			
Location Master node State Status		🖹 Problems 🖉 Tasks @ Javadoc 🧽 Map/Reduce Locations 🕱	ೄ 🥂 🗆
< ···· → < ··· → < ··· → < ··· → > < ··· → > < ··· → > > > > > > > > > > > > > > > > >		Location Master node State Status	New Ha
<			
 ✓ ···· · · · · · · · · · · · · · · · ·			
< · · · · · · · · · · · · · · · · · · ·			
< ···			
↑ I items selected	< III >		>
	[↑] ¹ items selected		

找大象-連接HDFS

	New Hadoop location 🗙
	Define Hadoop location Define the location of a Hadoop infrastructure for running MapReduce applications. General Advanced parameters
	Location name: hadoop
打你namenode的IP	Map/Reduce Master DFS Master
11 Million of 11	Host: localhost Host: localhost Port: 9001 Port: 9001
	User name: waue
	SOCKS proxy
	Host: host
	Port: 1080
	Load from file
	⑦ <u>Finish</u> Cancel

找大象-連接HDFS

E	Ma	p/Reduce - Eclips	se SDK	
<u>File Edit Source Refactor Name</u>	vigate Se <u>a</u> rch <u>P</u> roject	<u>R</u> un <u>W</u> indow <u>I</u>	<u>H</u> elp	
] ǰ 🔒 📄 ∲× O× 🥵]	≝ ₩ @~] 😕 🔗~] & ~ 친~ ~ ~ <		
Project Explorer 🛿 🗖 🗖				
🕨 🔁 icas				
	Problems A Tasks Location R hadoop	@ Javadoc 😥 Map	/Reduce Locations ☎ Master node localhost	S
				1



獵大象

