



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING

設置 Hadoop 環境

王耀聰 陳威宇

jazz@nchc.org.tw

waue@nchc.org.tw

國家高速網路與計算中心 (NCHC)

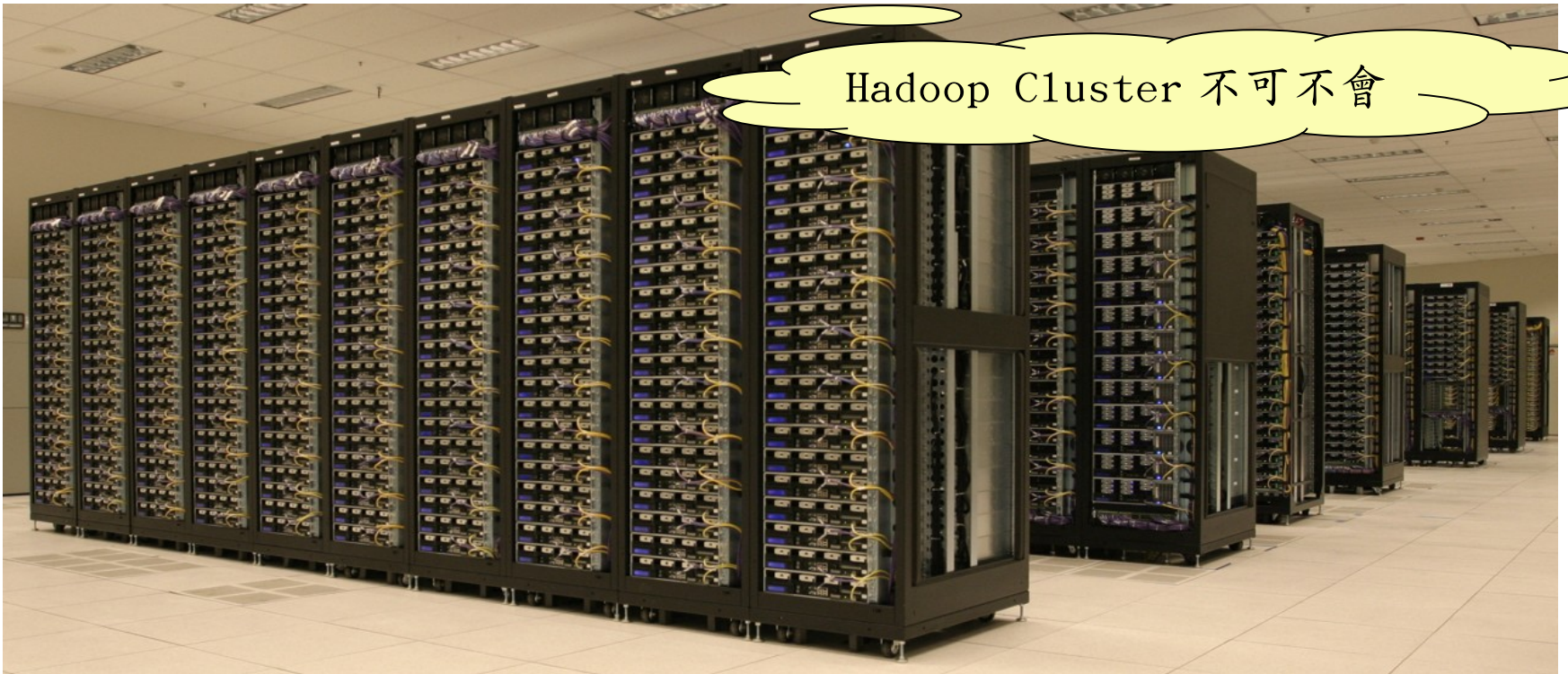


自由軟體實驗室

Yahoo's Hadoop Cluster

- ~10,000 machines running Hadoop in US
- The largest cluster is currently 2000 nodes
- Nearly 1 petabyte of user data (compressed, unreplicated)
- Running roughly 10,000 research jobs / week

Hadoop Cluster 不可不會



Hadoop 單機設定與啓動

- step 1. 設定登入免密碼
- step 2. 安裝 java
- step 3. 下載安裝 Hadoop
- step 4.1 設定 `hadoop-env.sh`
 - `export JAVA_HOME=/usr/lib/jvm/java-6-sun`
- step 4.2 設定 `hadoop-site.xml`
 - 設定 Namenode -> `hdfs://localhost:9000`
 - 設定 Jobtracker -> `localhost:9001`
- step 5.1 格式化 HDFS
 - `bin/hadoop namenode -format`
- step 5.2 啓動 Hadoop
 - `bin/start-all.sh`
- step 6. 完成！檢查運作狀態
 - Job admin `http://localhost:50030/` HDFS
 - `http://localhost:50070/`



Hadoop 單機環境示意圖

Node 1

conf /
hadoop-site.xml:

```
fs.default.name ->  
hdfs://localhost:9000  
mapred.job.tracker ->  
localhost:9001
```

conf/slaves:

```
localhost
```

Localhost

Namenode

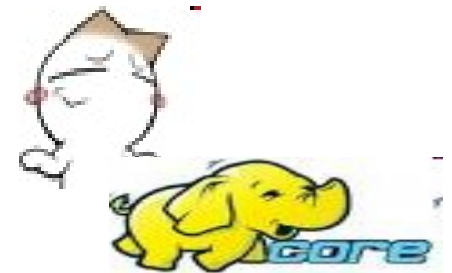
JobTracker

Datanode

Tasktracker

localhost:50070

localhost:50030



Hadoop 叢集設定與啓動

- step 1. 設定登入免密碼
- step 2. 安裝 java
- step 3. 下載安裝 Hadoop
- step 4.1 設定 `hadoop-env.sh`
 - `export JAVA_HOME=/usr/lib/jvm/java-6-sun`
- step 4.2 設定 `hadoop-site.xml`
 - 設定 Namenode -> `hdfs://x.x.x.1:9000`
 - 設定 Jobtracker -> `x.x.x.2:9001`
- step 4.3 設定 `slaves` 檔
- step 4.4 將叢集內的電腦 Hadoop 都做一樣的配置
- step 5.1 格式化 HDFS
 - `bin/hadoop namenode -format`
- step 5.2 啓動 Hadoop
 - `nodeN` 執行: `bin/start-dfs.sh` ; `nodeJ` 執行: `bin/start-mapred.sh`
- step 6. 完成! 檢查運作狀態
 - Job admin `http://x.x.x.2:50030/` HDFS `http://x.x.x.1:50070/`

情況一

Node 1

conf /
hadoop-site.xml:

```
fs.default.name ->  
hdfs://x.x.x.1:9000  
mapred.job.tracker ->  
x.x.x.1:9001
```

x.x.x.1

Namenode

http://x.x.x.1:50070

http://x.x.x.1:50030

JobTracker

Node 2

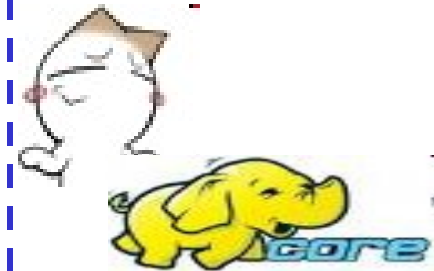
x.x.x.2

Datanode

Datanode

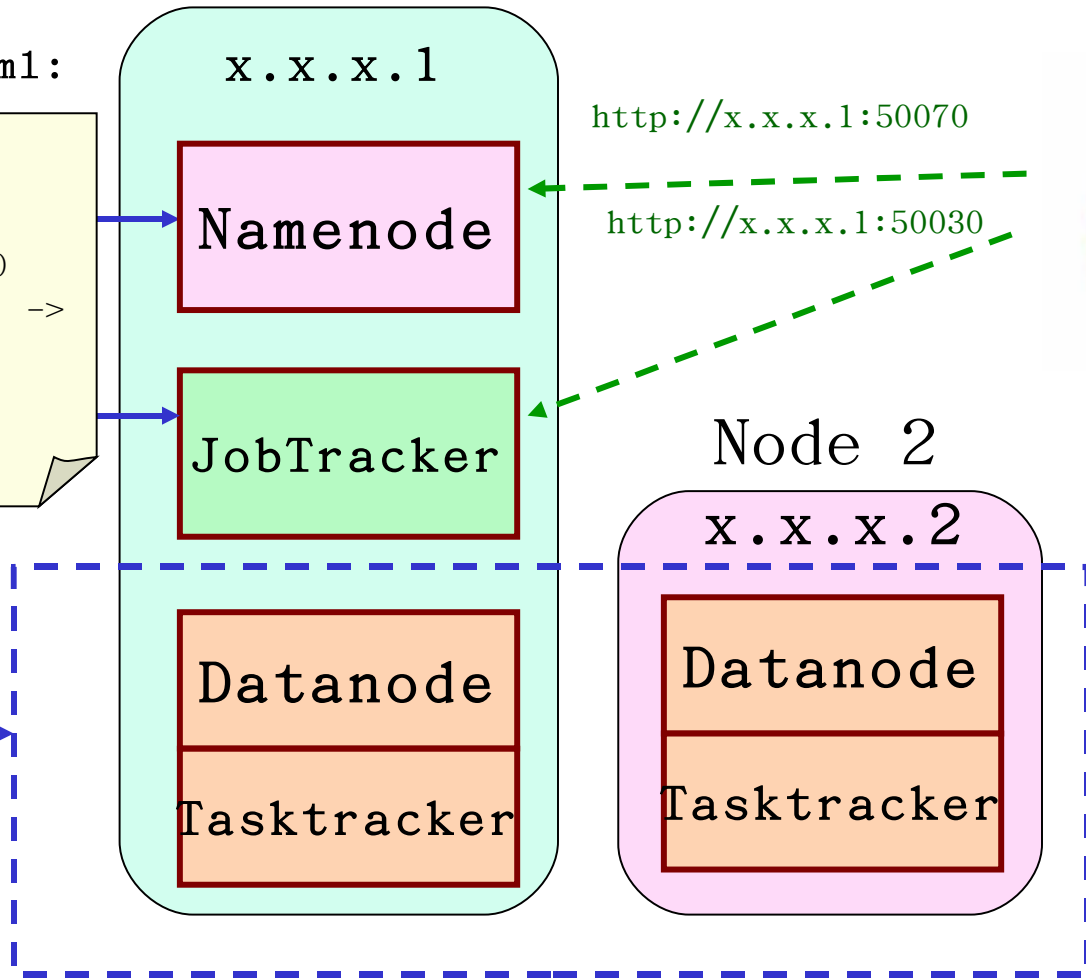
Tasktracker

Tasktracker



conf/slaves:

```
x.x.x.1  
x.x.x.2
```



情況二

conf /
hadoop-site.xml:

```
fs.default.name ->  
hdfs://x.x.x.1:9000  
mapred.job.tracker ->  
x.x.x.2:9001
```

Node 1

x.x.x.1

Namenode

Datanode

Tasktracker

Node 2

x.x.x.2

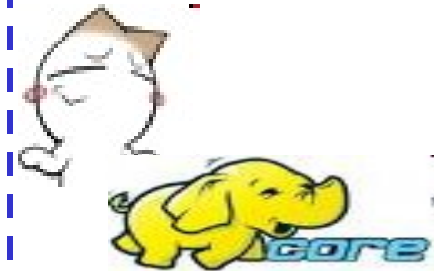
JobTracker

Datanode

Tasktracker

<http://x.x.x.1:50070>

<http://x.x.x.2:50030>



conf/slaves:

```
x.x.x.1  
x.x.x.2
```



情況三

conf /
hadoop.site.xml:

```
fs-default.name ->  
hdfs://x.x.x.1:9000  
mapred.job.tracker ->  
x.x.x.1:9001
```

Node 1

x.x.x.1

Namenode

JobTracker

http://x.x.x.1:50070

http://x.x.x.1:50030



conf/slaves:

```
x.x.x.2  
.....  
x.x.x.n
```

Node 2

x.x.x.2

Datanode

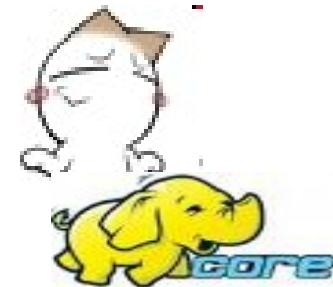
Tasktracker

Node N

x.x.x.n

Datanode

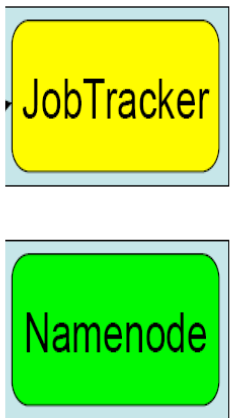
Tasktracker



情況四

conf /
hadoop-site.xml:

```
mapred.job.tracker->  
x.x.x.2:9001  
fs.default.name ->  
hdfs://x.x.x.1:9000
```



http://x.x.x.2:50030

HTTP Monitoring UI

http://x.x.x.1:50070



conf/slaves:

```
x.x.x.3  
.....  
x.x.x.n
```

