



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

結語

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

Overview

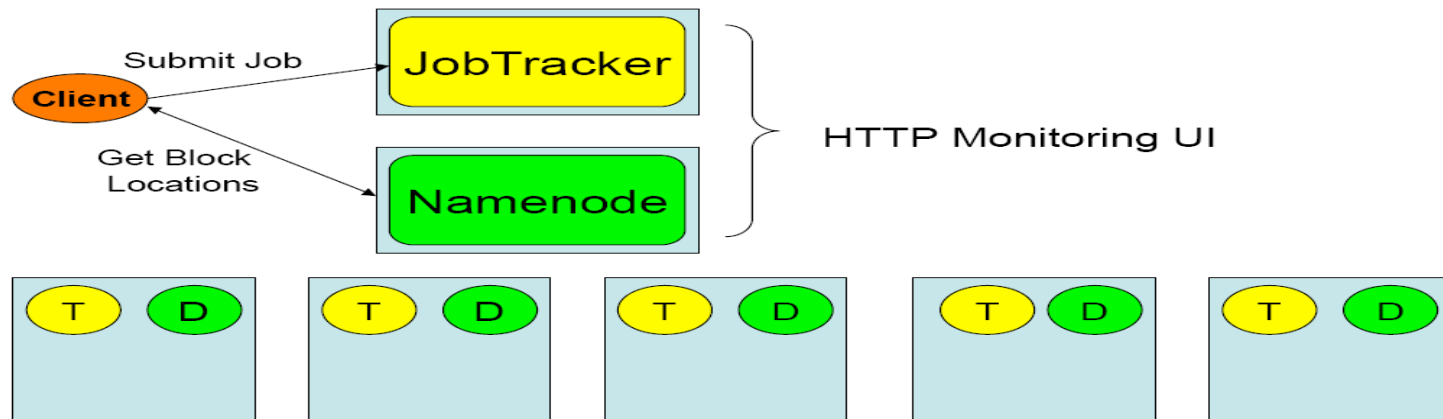
- 重點回顧
- 其他專案介紹
- 文獻參考

雲端運算

- 什麼是雲端運算？
 - 將使用者的命令透過介面，交由網路上眾多伺服器所組成的龐大系統運算之後，再把處理結果回傳給用戶
 - 雲＝～ 網路
- 網路運算的演化
 - 叢集 -> 平行分散 -> 格網 -> 雲端
- 服務型態
 - SaaS, Paas, Iaas
- 特色
 - 經濟、簡單、可擴充…

Hadoop

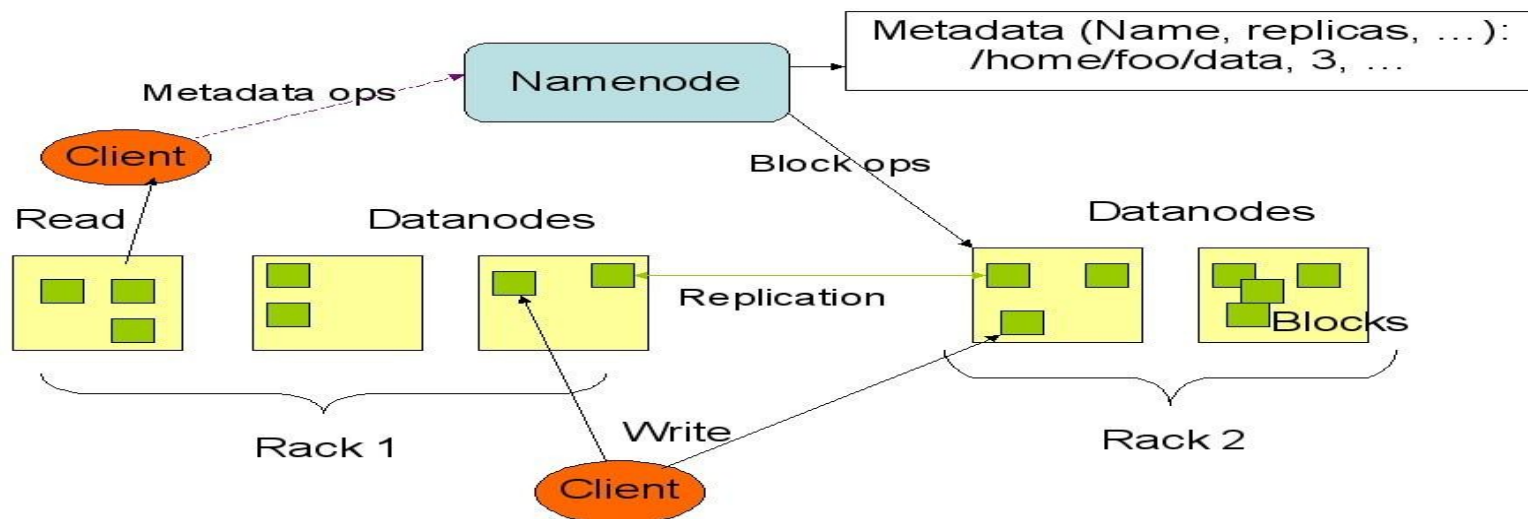
- 什麼是 Hadoop ?
 - 借鑑了 google 在分佈式系統上設計的分散式計算平台
 - 由 java 實做的自由軟體
 - 包含了 HDFS 、 MapReduce
- 特色
 - 巨量、經濟、高效率、可靠、持續更新



Hadoop Distributed File System

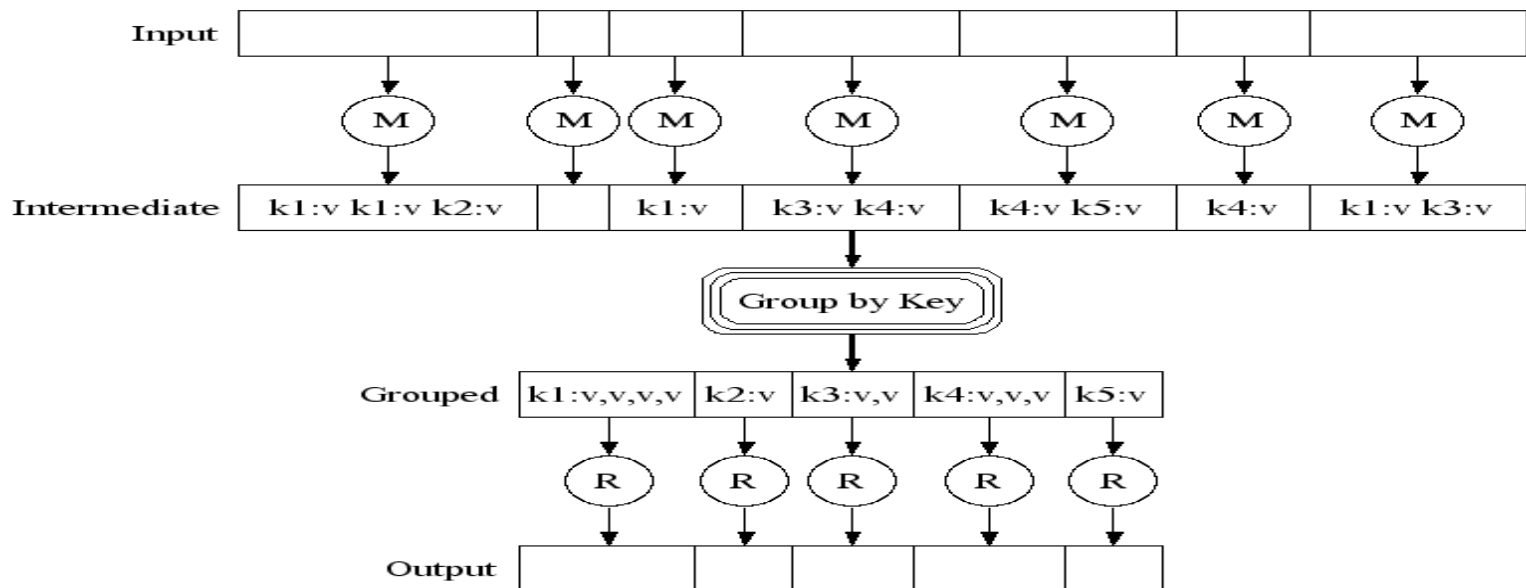
- 什麼是 HDFS ?
 - Hadoop 專案中的檔案系統
- 特色
 - 錯誤容忍、**高 Throughput**、大規模資料集、一次多次讀、在地運算、異質平台移植

HDFS Architecture



MapReduce

- 什麼是 MapReduce
 - Map 將每個資料視為一個 key，並作 $\langle \text{key}, \text{value} \rangle$ 的配對，Reduce 再統合所有的 Map 結果做出 $\langle \text{key}, \text{list}(\text{value}) \rangle$
- 運作方法



Hadoop 安裝設定

安裝

- step 1. 設定登入免密碼
- step 2. 安裝 java
- step 3. 下載安裝 Hadoop

Java-6-sun

Hadoop-core

設定

- 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

hadoop-env.sh

hadoop-site.xml

slaves

執行

- step 5.1 格式化 HDFS
 - bin/hadoop namenode -format
- step 5.2 啟動 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/>

程式基本寫法



編譯程式基本步驟

1. 編譯

- `javac -classpath hadoop-*-core.jar -d MyJava MyCode.java`

1. 封裝

- `jar -cvf MyJar.jar -C MyJava .`

1. 執行

- `bin/hadoop jar MyJar.jar MyCode HDFS_Input/
HDFS_Output/`

-
- 所在的執行目錄為 Hadoop_Home
 - `./MyJava` = 編譯後程式碼目錄
 - `MyJar.jar` = 封裝後的編譯檔

- 先放些文件檔到 HDFS 上的 input 目錄
- `./input`; `./output` = hdfs 的輸入、輸出目錄

其他相關專案

- HBase (<http://hadoop.apache.org/hbase/>)



- 用 Hadoop 為基礎的雲端資料庫

- Nutch (<http://lucene.apache.org/nutch/>)



- 以 Hadoop 為基礎的搜尋引擎

- Pig (<http://hadoop.apache.org/pig/>)

- 一個可用在 Hadoop 上的平台，提供一個全新語言 (Pig Latin) 以簡化撰寫分析的程式



- Disco (<http://discoproject.org/>)

- Nokia 所研發的 MapReduce 架構，用 erlang 實現 massive data - minimal code 者可以 python 驅動，類似 Hadoop 的自由軟體專案。

disco

massive data - minimal code

文獻參考

- Hadoop 官方網站
 - <http://hadoop.apache.org/core/>
- Hadoop API
 - <http://hadoop.apache.org/core/docs/r0.18.3/api/index.html>
- Hadoop Taiwan User Group
 - <http://www.hadoop.tw/>
- 中文 Hadoop 手冊
 - <http://cn.hadoop.org/doc/index.html>
- 維基百科
 - <http://en.wikipedia.org/wiki/Hadoop>
 - <http://zh.wikipedia.org/wiki/Hadoop>

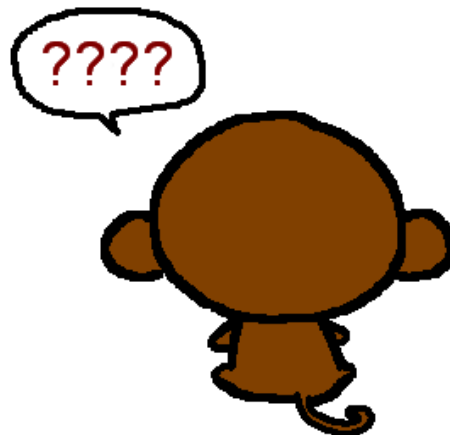


財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING

Question ?



http://chinese.storylands.org/1329magic_al/story1c15.php



自由軟體實驗室



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

Thank You !



<http://miumiu516.pixnet.net/album/photo/94262410>



自由軟體實驗室