進階課程

Hadoop 進階程式設計與 HBase 資料庫整合實作

王耀聰　陳威宇

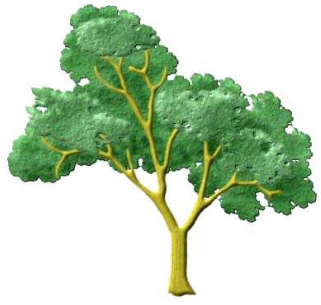
jazz@nchc.org.tw

waue@nchc.org.tw

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

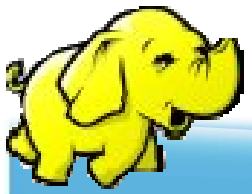


課程大綱 （1）

第一天	
09:30~10:20	介紹課程 與 Hadoop簡介
10:20~10:30	休息
10:30~12:00	Hadoop生態系簡介
	實作一：Hadoop Streaming 範例操作
12:00~13:00	午餐
13:00~15:00	開發輔助工具 Eclipse
	Map Reduce 程式架構
15:00~15:10	休息
15:10~16:30	程式設計I- HDFS 操作
	程式設計II-範例程式

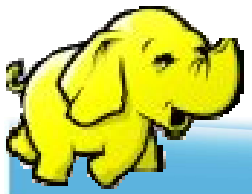
課程大綱 （2）

第二天	
09:30~10:20	HBase 簡介與架構
10:20~10:30	休息
10:30~12:00	HBase 安裝操作說明
12:00~13:00	午餐
13:00~15:00	HBase 程式架構與範例
15:00~15:10	休息
15:10~16:00	Hadoop + HBase + PHP 案例實務
16:00~16:30	hadoop ＋ 關聯式資料庫



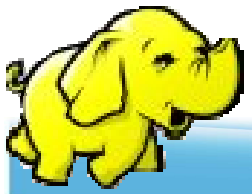
學員背景調查

- Java 語言 ??
- PHP 語言 ?? MySQL 資料庫 ??
- Linux 操作 ?? 電腦叢集維護 ??
- 安裝過 Hadoop ??
- 參加過 Hadoop 基礎課程 ??



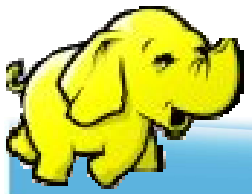
It’s Show Time

- 名稱
- 服務公司/就讀學校
- 報名原因
- 預期收穫



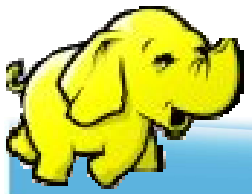
引言

雲端運算這個名詞雖然紅，
但我一定需要雲端運算嗎？
他用在什麼場合？又或非它不可嗎？



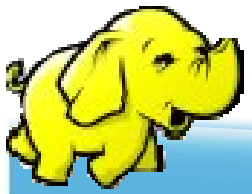
Computing with big datasets

is a fundamentally different challenge than doing “big compute” over a small dataset



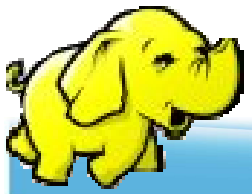
平行分散式運算

- 格網運算(網格運算, Grid computing)
  - MPI, PVM, Condor…
- 著重於: 分散工作量
- 目前的問題在於：如何分散資料量
  - Reading 100 GB off a single filer would leave nodes starved – just store data locally



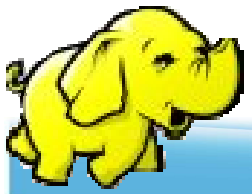
分散大量資料： Slow and Tricky

- 交換資料需同步處理
  - **Deadlock** becomes a problem
- 有限的頻寬
  - Failovers can cause **cascading failure**



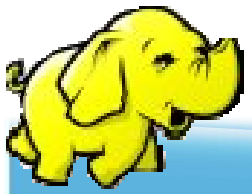
數字會說話

- Data processed by Google every month: 400 PB … in 2007
  - ◆Max data in memory: 32 GB
  - ◆Max data per computer: 12 TB
  - ◆Average job size: 180 GB
- 光一個device的讀取時間= 45 minutes



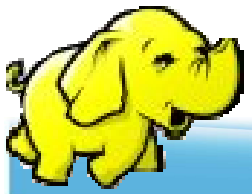
所以 ...

- 運算資料可以很快速，但瓶頸在於硬碟的 I/O
  - 1 HDD = 75 MB/sec
- 解法: parallel reads
  - 1000 HDDs = 75 GB/sec



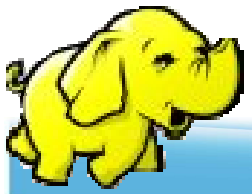
MapReduce 的動機

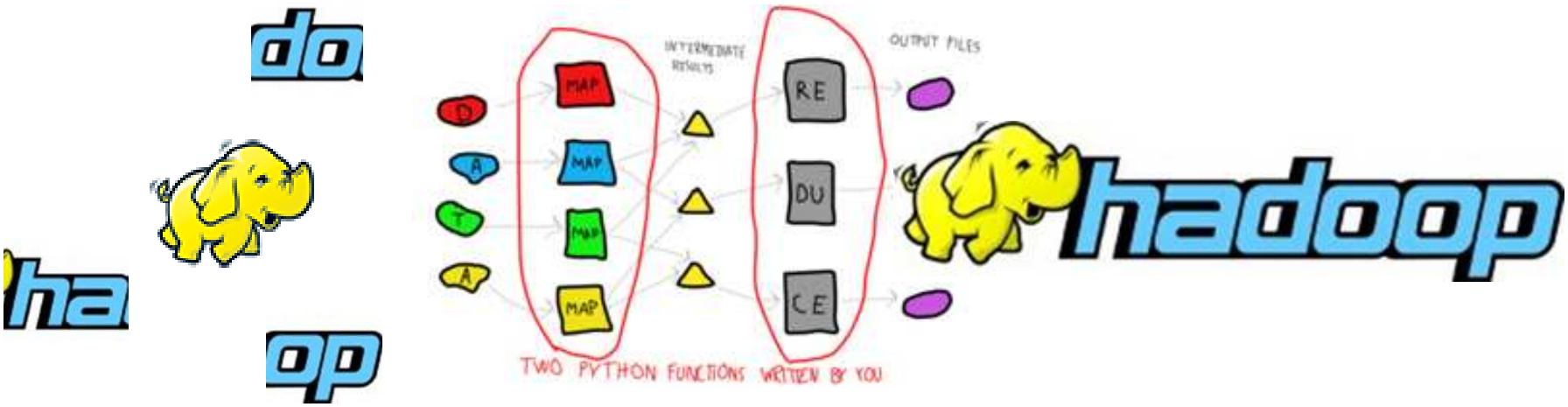
- Data > 1 TB
- 交互運算於大量的CPU
- 容易開發與使用
  - High-level applications written in MapReduce
  - Programmers don’t worry about socket(), etc.



Conclusions

- “大資料集的運算”與“小資料集的高速運算”，兩者是迥然不同的挑戰。
- 大資料量的解決方法將需要：
  - 提供囊括所有解決之道的新工具
  - MapReduce 與 HDFS 等工具是其中之一



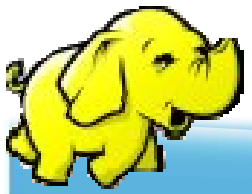


一、Hadoop 簡介

Hadoop 是一套儲存並處理
petabytes 等級資訊的
雲端運算技術

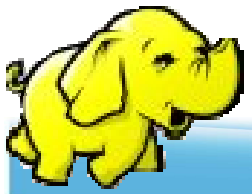
Hadoop

- 以Java開發
- 自由軟體
- 上千個節點
- Petabyte等級的資料量
- 創始者 Doug Cutting
- 為Apache 軟體基金會的 top level project



特色

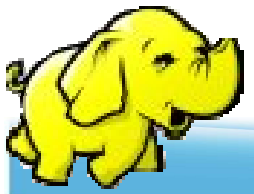
- 巨量
  - 擁有儲存與處理大量資料的能力
- 經濟
  - 由一般個人電腦所架設的叢集環境
- 效率
  - 藉由平行分散檔案以致得到快速的回應
- 可靠
  - 當某個節點發生錯誤，系統能即時自動的取得備份資料以及佈署運算資源



實用案例

Hadoop於Yahoo的運作資訊

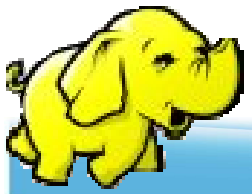
年份	日期	節點數	耗時（小時）
2006	四月	188	47.9
2006	五月	500	42
2006	十一月	20	1.8
2006	十一月	100	3.3
2006	十一月	500	5.2
2006	十一月	900	7.8
2007	七月	20	1.2
2007	七月	100	1.3
2007	七月	500	2
2007	七月	900	2.5



Sort benchmark, every nodes with terabytes data.

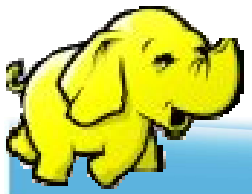
誰在用Hadoop ？(1)

- Facebook
  - 處理 internal log and dimension data sources
  - for reporting/analytics and machine learning.
- IBM
  - Blue Cloud Computing Clusters
- Journey Dynamics
  - 用 Hadoop MapReduce 分析 billions of lines of GPS data 並產生交通路線資訊.
- Krugle
  - 用 Hadoop and Nutch 建構 原始碼搜尋引擎



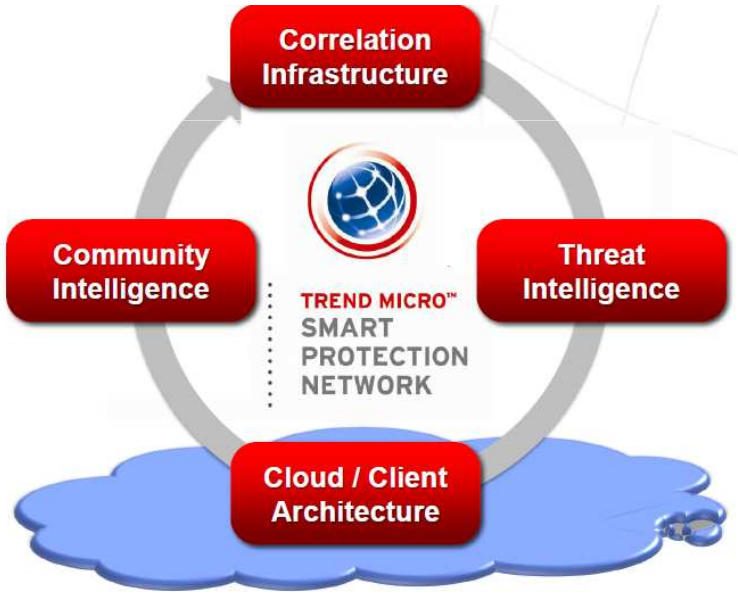
誰在用Hadoop？(2)

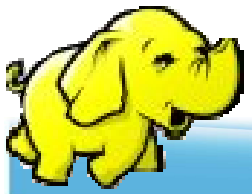
- SEDNS - Security Enhanced DNS Group
  - 收集全世界的 DNS 以探索網路分散式內容.
- Technical analysis and Stock Research
  - 分析股票資訊
- University of Nebraska Lincoln, Research Computing Facility
  - 用Hadoop跑約200TB的CMS經驗分析
  - **緊湊渺子線圈**（**CMS**，**C**ompact **M**uon **S**olenoid）為瑞士歐洲核子研究組織CERN的大型強子對撞器計劃的兩大通用型粒子偵測器中的一個。



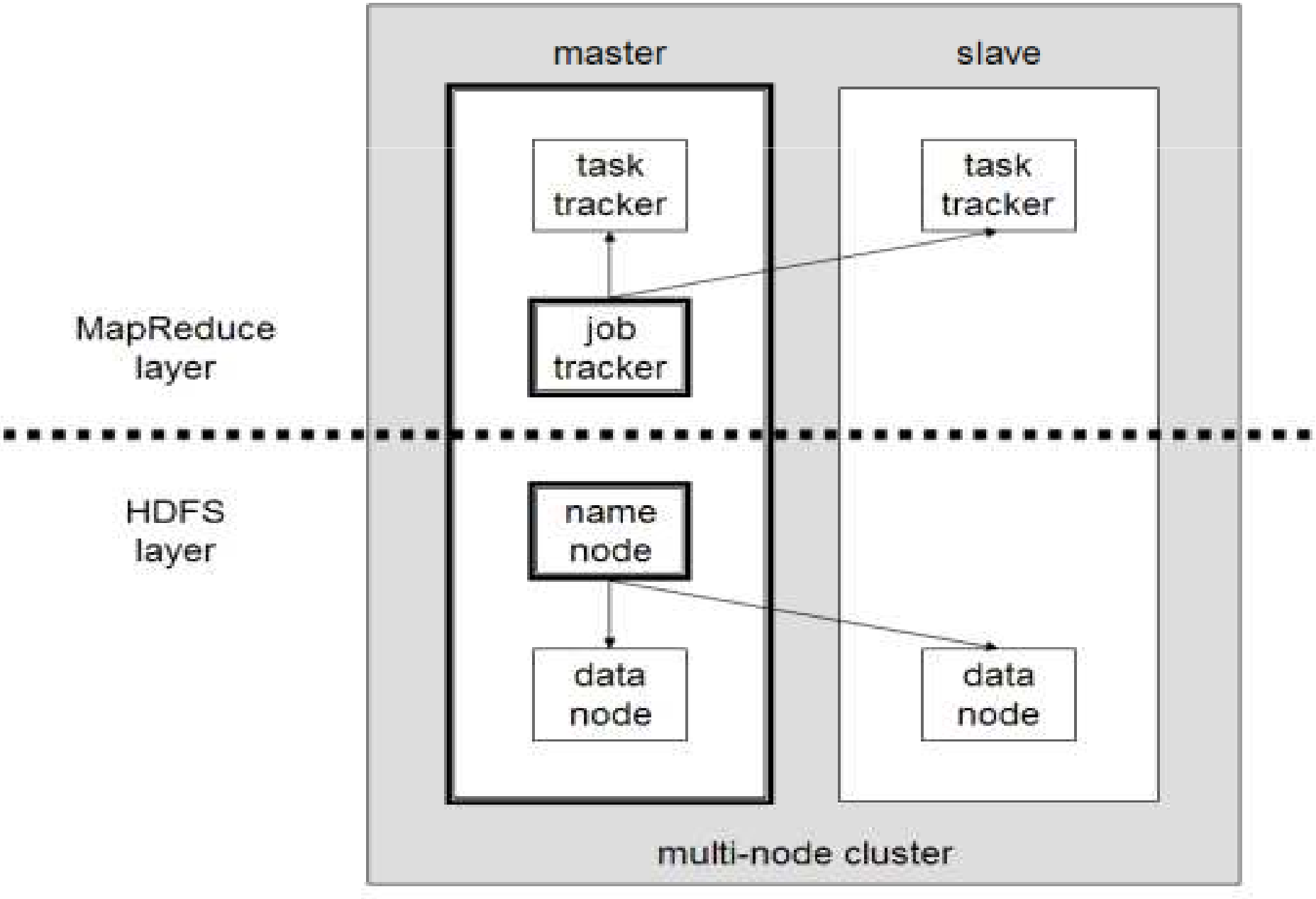
誰在用Hadoop？(3)

- Yahoo!
  - Used to support research for Ad Systems and Web Search
  - 使用Hadoop平台來發現發送垃圾郵件的殭屍網絡
- 趨勢科技
  - 過濾像是釣魚網站或惡意連結的網頁內容

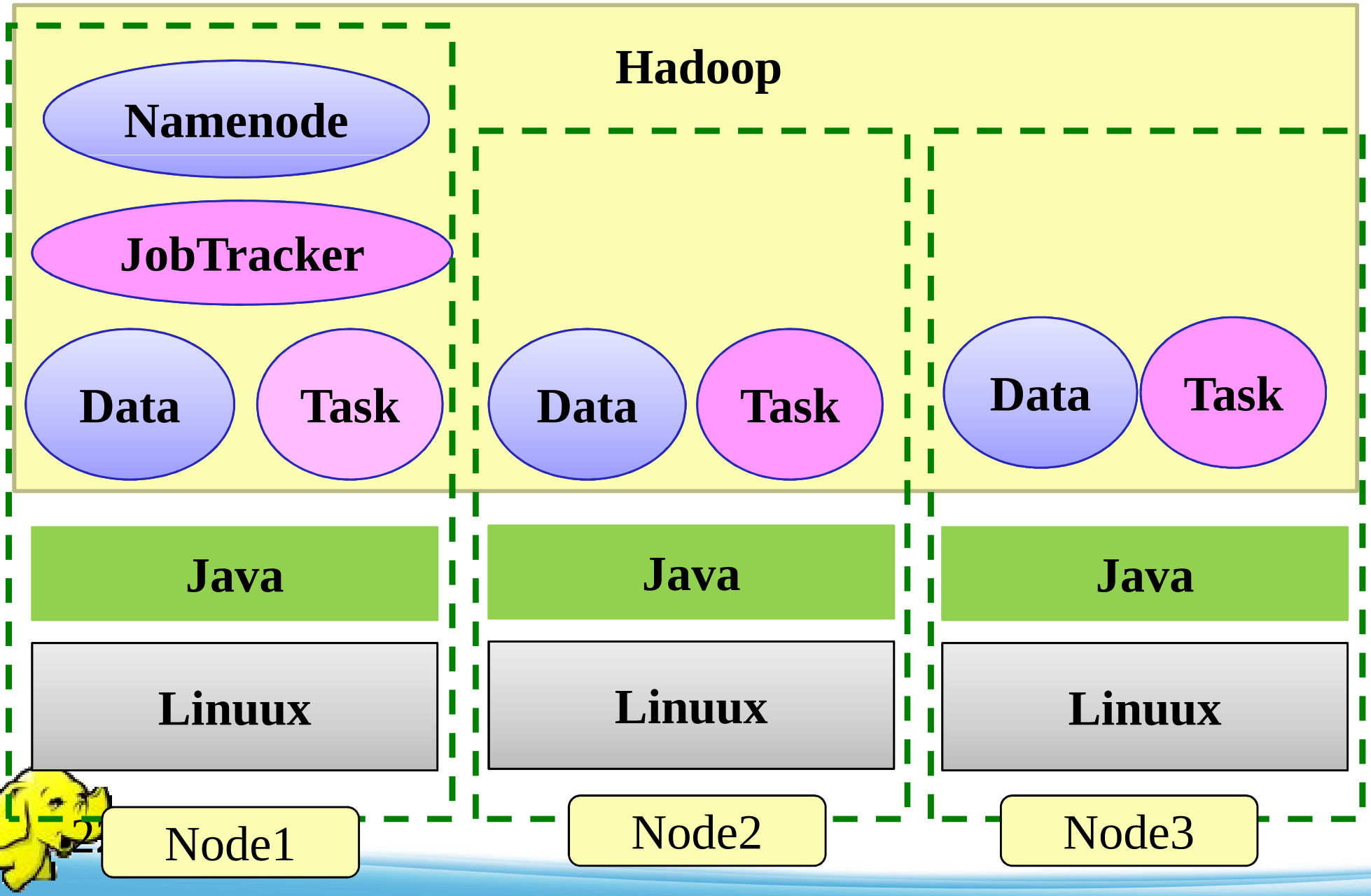




Hadoop 的主要架構



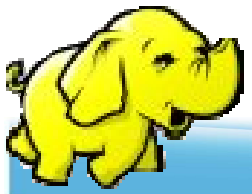
Building Hadoop



名詞

- Job
  - 任務
- Task
  - 小工作
- JobTracker
  - 任務分派者
- TaskTracker
  - 小工作的執行者
- Client
  - 發起任務的客戶端
- Map
  - 應對
- Reduce
  - 總和

- Namenode
  - 名稱節點
- Datanode
  - 資料節點
- Namespace
  - 名稱空間
- Replication
  - 副本
- Blocks
  - 檔案區塊 (64M)
- Metadata
  - 屬性資料



管理資料

Namenode

- Master
- 管理HDFS的名稱空間
- 控制對檔案的讀/寫
- 配置副本策略
- 對名稱空間作檢查及紀錄
- 只能有一個

Datanode

- Workers
- 執行讀/寫動作
- 執行Namenode的副本策略
- 可多個

分派程序

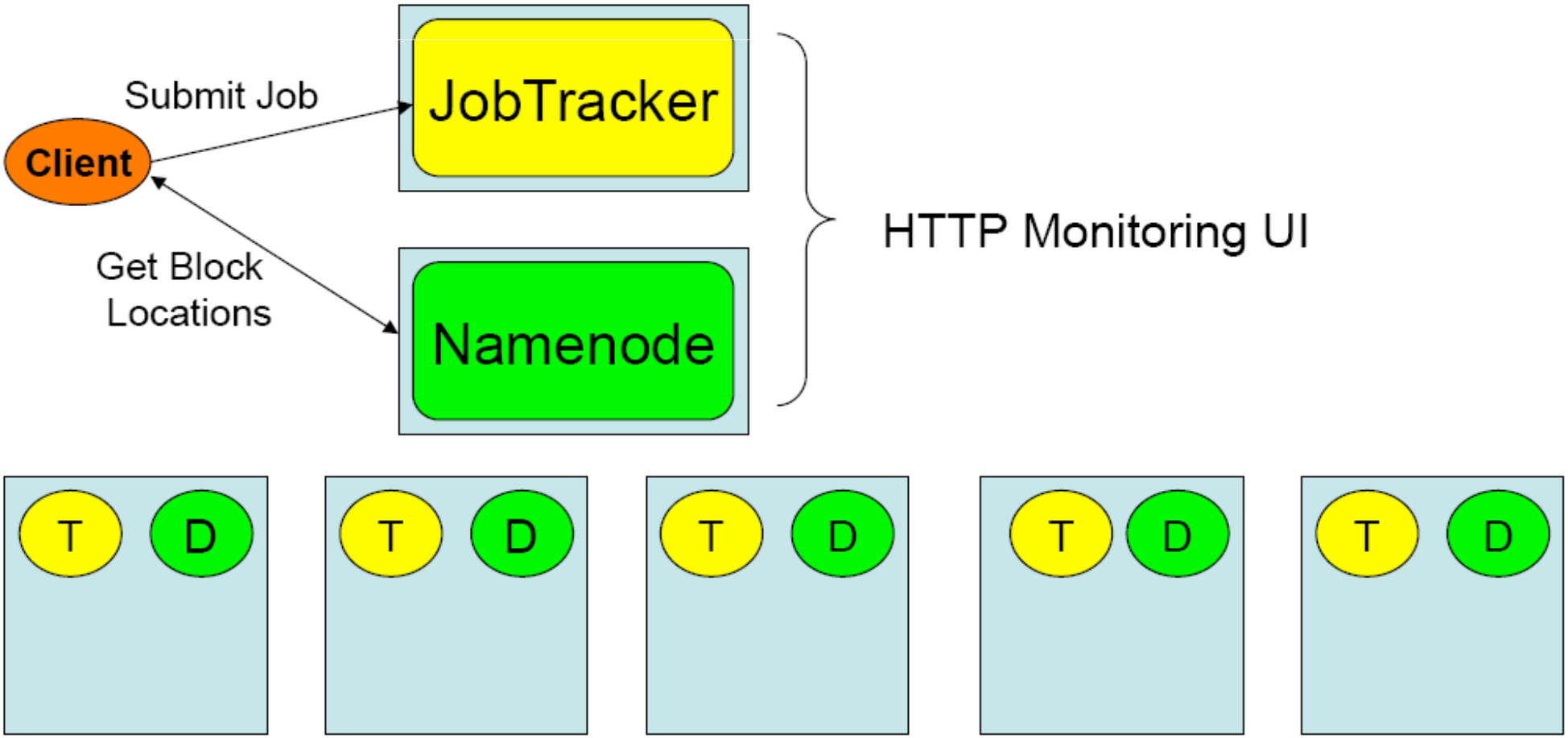
Jobtracker

- Master
- 使用者發起工作
- 指派工作給 Tasktrackers
- 排程決策、工作分配、錯誤處理
- 只能有一個

Tasktrackers

- Workers
- 運作Map 與 Reduce 的工作
- 管理儲存、回覆運算結果
- 可多個

不在雲裡的 Client

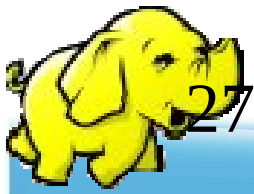


其他的 Open Source 專案:

Sector	The National Center for Data Mining (NCDM)	http://sector.sourceforge.net/

其他不同語言實作的 MapReduce 函式庫

http://trac.nchc.org.tw/grid/wiki/jazz/09-04-14#MapReduce



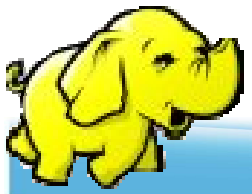
27

關於 Sector / Sphere

- http://sector.sourceforge.net/
- 由美國資料探勘中心(National Center for Data Mining)研發的自由軟體專案。
- 採用C/C++語言撰寫，因此效能較 Hadoop 更好。
- 提供「類似」Google File System與MapReduce的機制
- 基於UDT高效率網路協定來加速資料傳輸效率
- Open Cloud Consortium的Open Cloud Testbed，有提供測試環境，並開發了MalStone效能評比軟體。

National Center for Data Mining
University of Illinois at Chicago

Open Data Group
http://www.opendatagroup.com/



Conclusions

- 所有工作都由JobTracker統一分派，由眾多TaskTracker 執行，每個TaskTracker又可以執行多個Task threads
- 所有名稱空間與檔案的metadata都由一個Namenode統籌，檔案空間為所有Datanode的集合，hdfs的基本單位為block
- Client 只需要丟工作或存取在“雲”的資料