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

Hadoop Overview

王耀聰 陳威宇 楊順發

jazz@nchc.org.tw

waue@nchc.org.tw

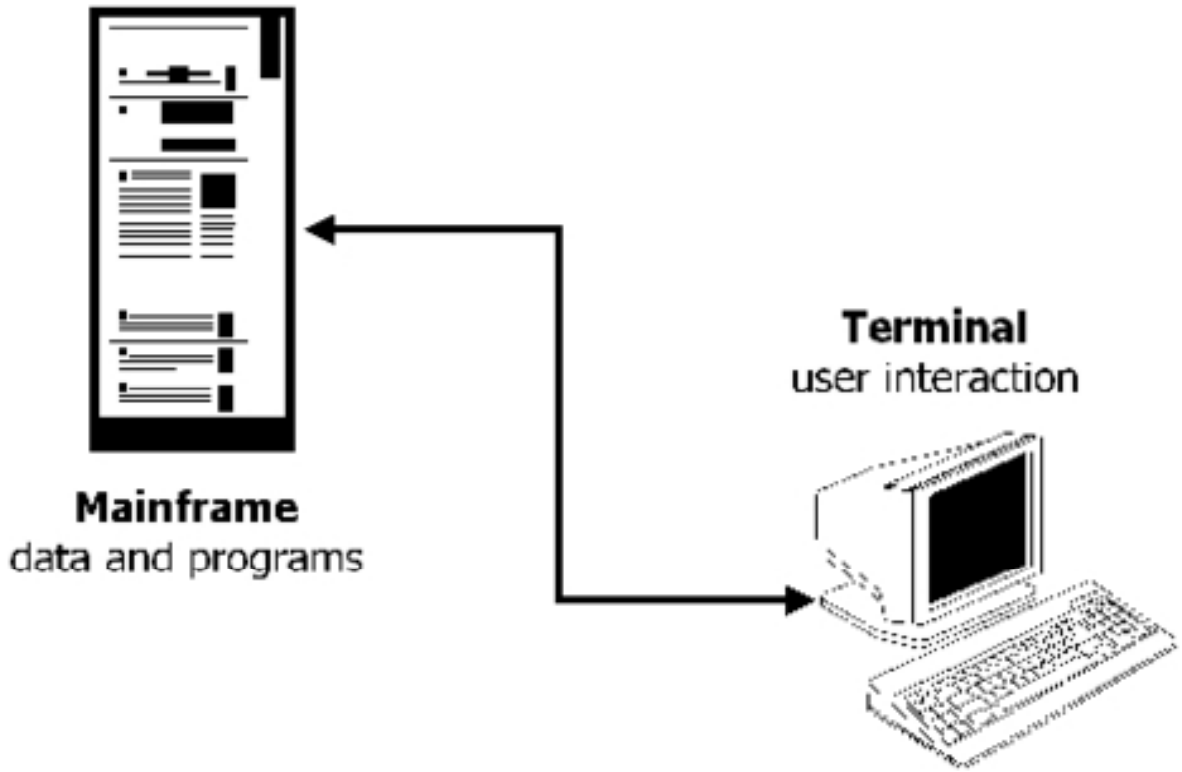
shunfa@nchc.org.tw

國家高速網路與計算中心(NCHC)

自由軟體實驗室

介紹Hadoop之前

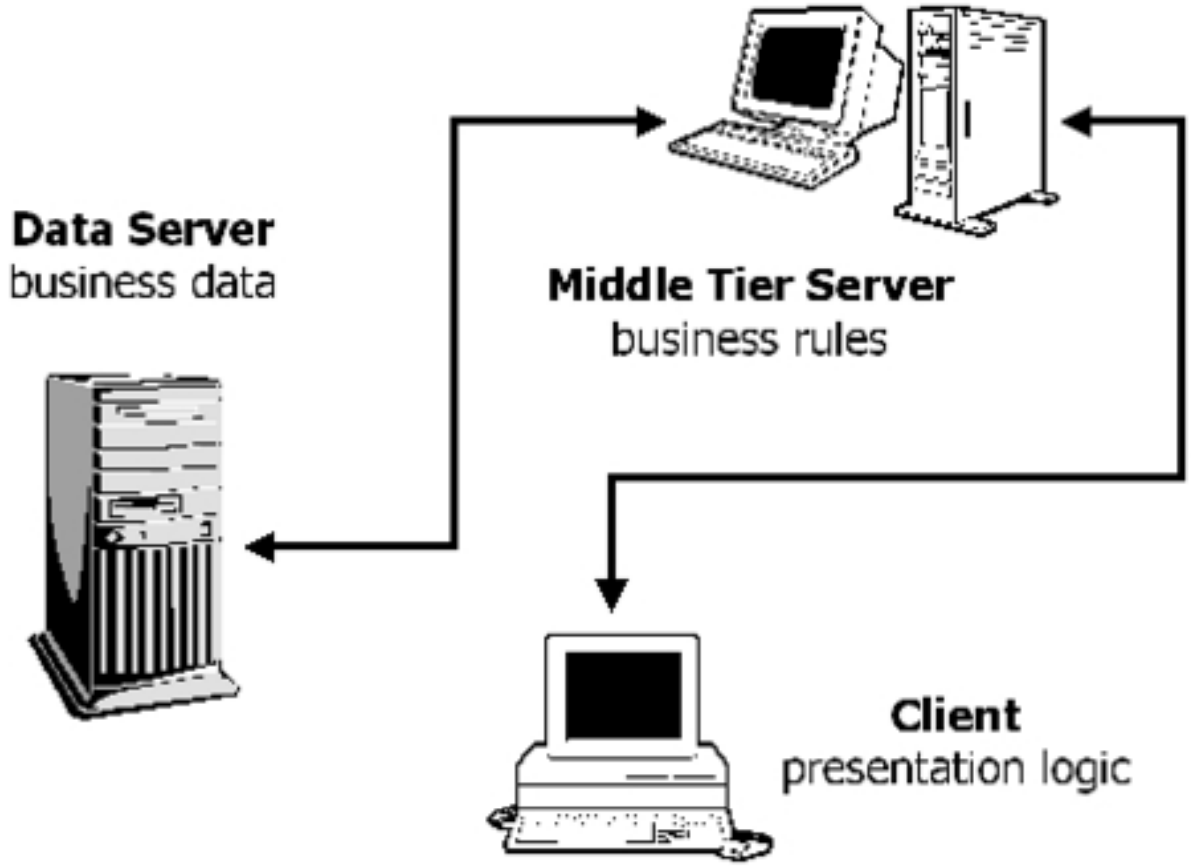
- 運算模式演進史- Mainframe Architecture



ref: http://www.linuxjournal.com/article/3508?page=0,0

介紹Hadoop之前

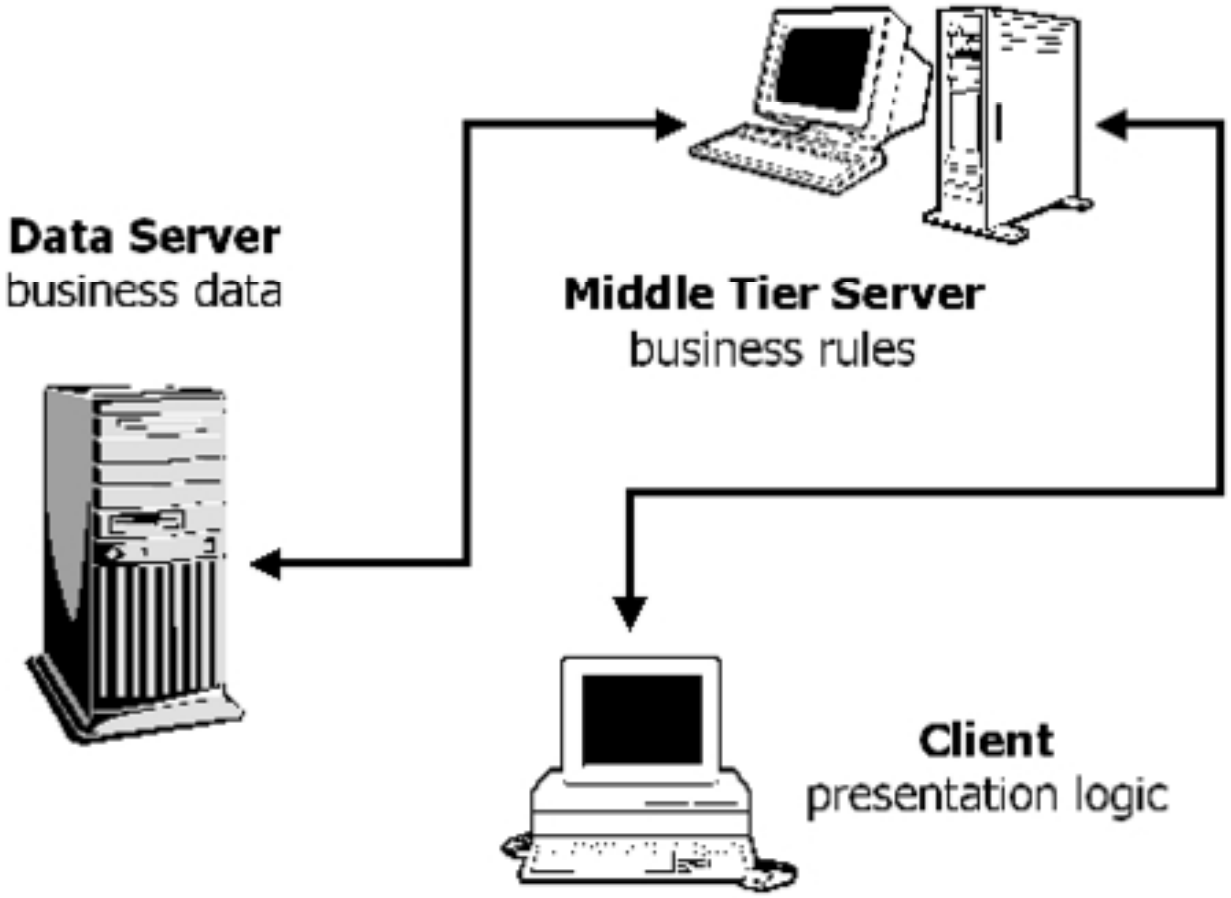
- 運算模式演進史- Two-Tier Client-Server Architecture



ref: http://www.linuxjournal.com/article/3508?page=0,0

介紹Hadoop之前

- 運算模式演進史- Three-Tier Architecture



ref: http://www.linuxjournal.com/article/3508?page=0,0

作業系統的最核心！

儲存空間的資源管理

記憶體空間與
行程分配

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

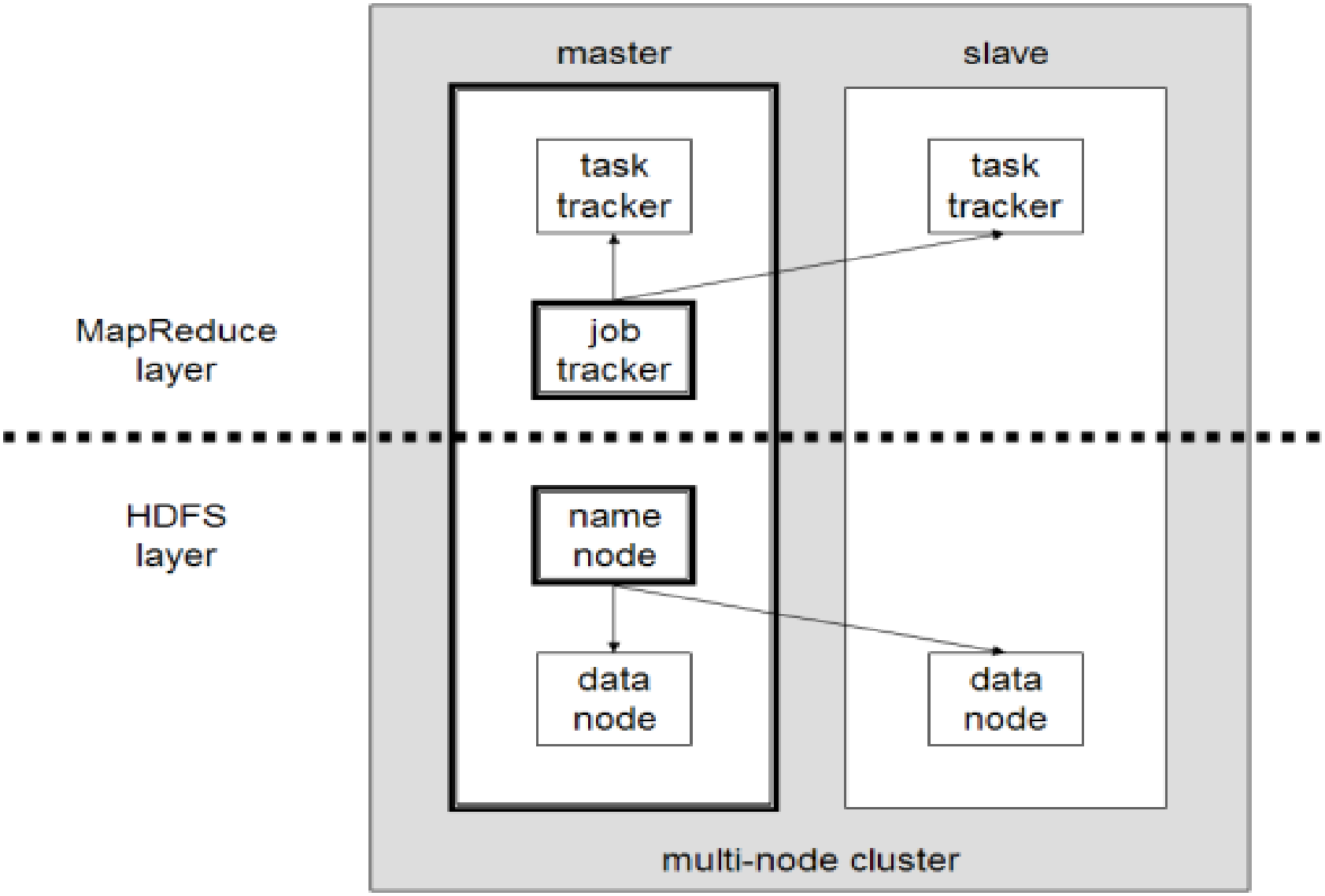
Jobtracker

- 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

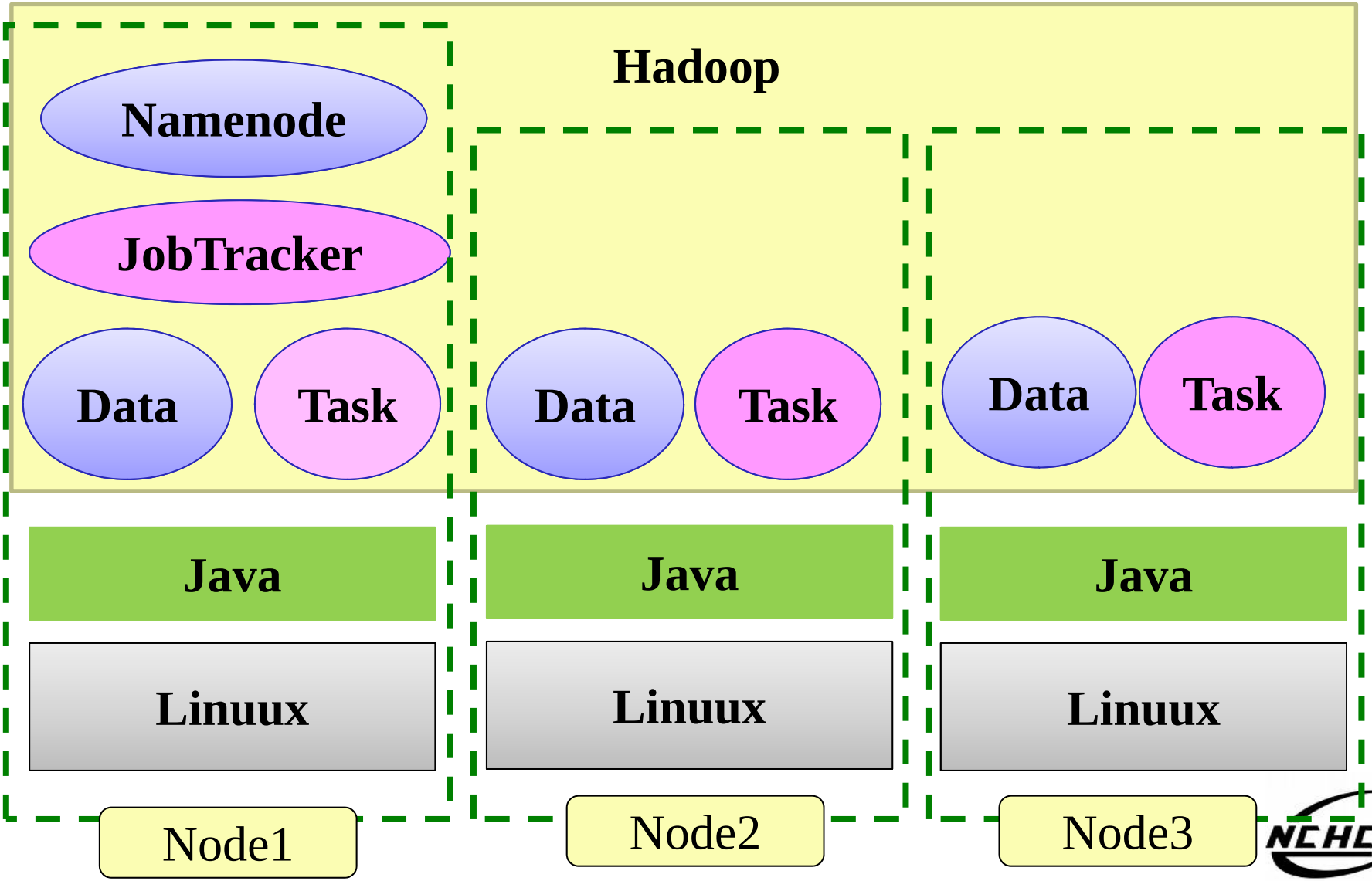
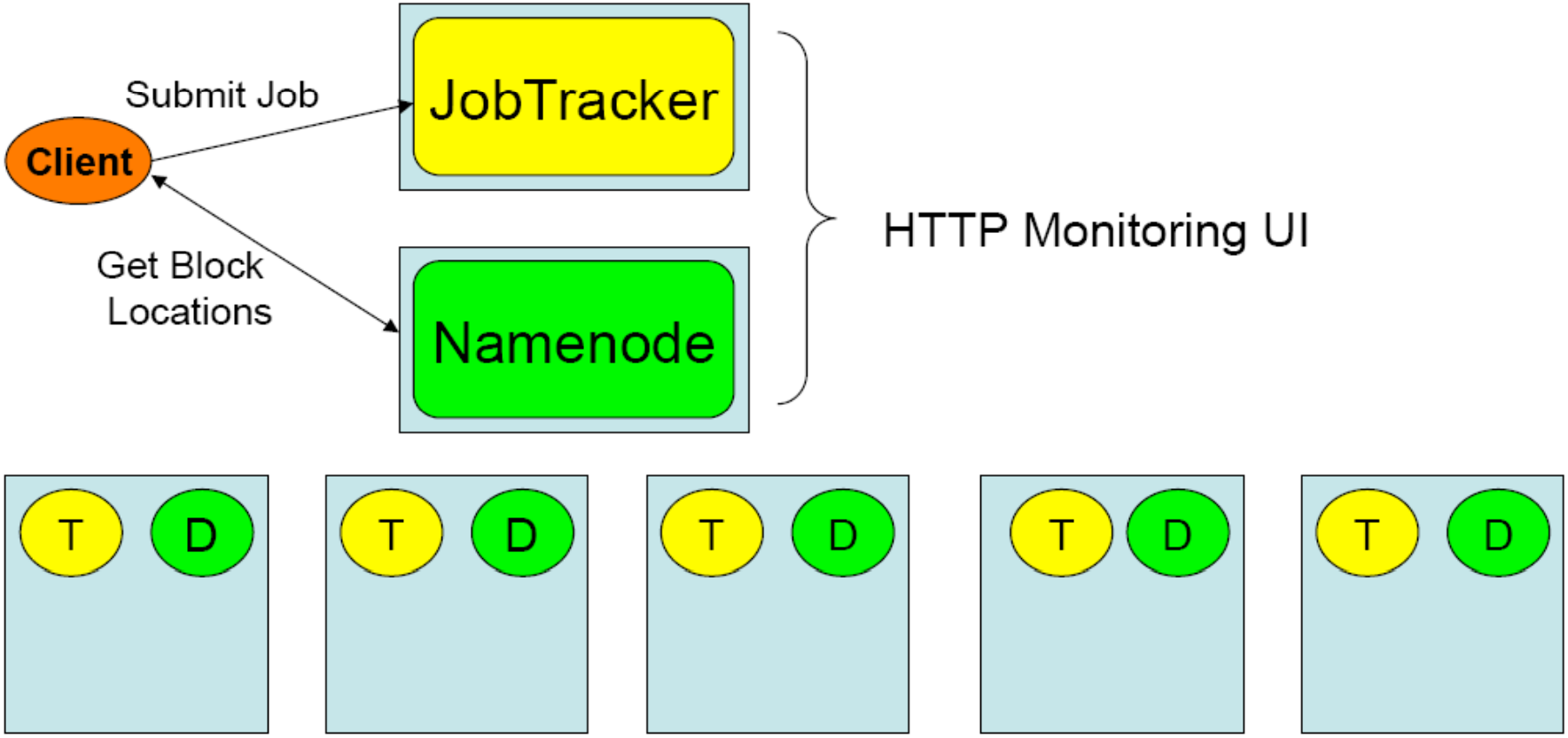
Tasktrackers

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

Hadoop的各種身份



Building Hadoop



不在雲裡的 Client