



釋放雲端潛能 駕馭海量資料

Hadoop在地理資訊系統上的應用 以福衛二號衛星影像管理及應用為例

辜文元 / 逢甲大學GIS中心
cool@gis.tw

地理資訊系統面臨什麼問題？



累積資料量，愈來愈龐大，資料儲存問題

需要分析的資料量變大，分析速度變慢

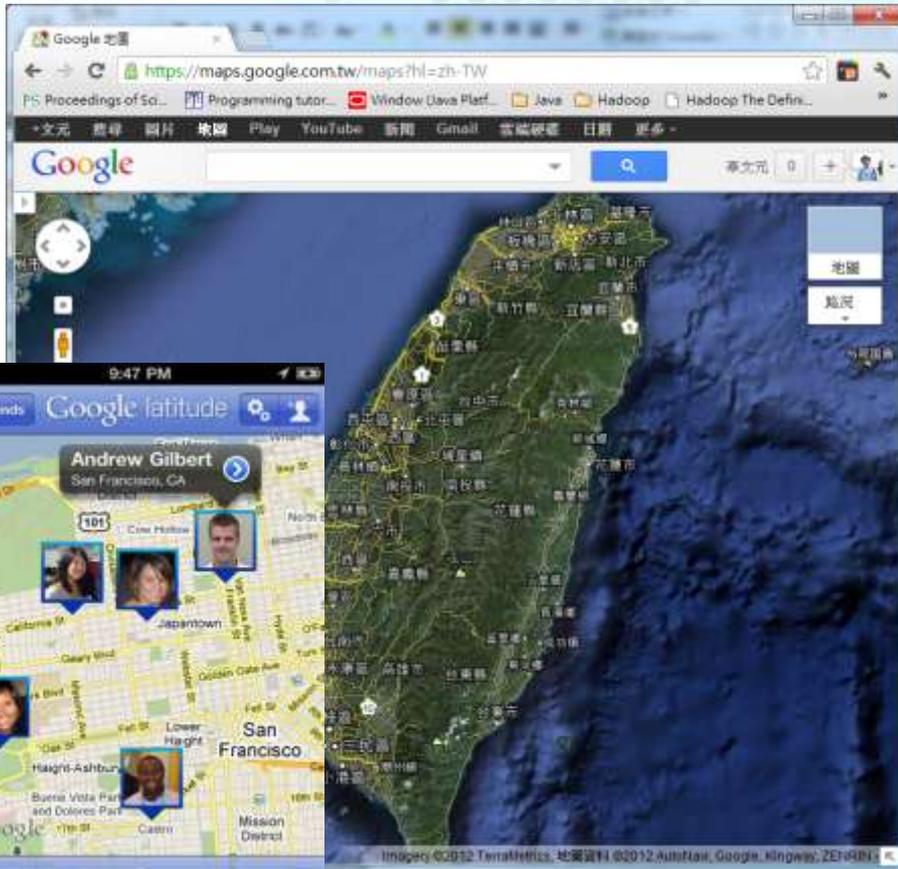
使用者愈多，圖台效率愈來愈差

套裝商用GIS軟體昂貴，擴充性受到阻礙

雲端運算技術的出現，對於地理資訊系統發展及應用面的影響？

從專業到大眾化

- GIS不再專屬於專業人士
- Google讓GIS更親民了



思維的改變

地圖回應速度

開放式架構

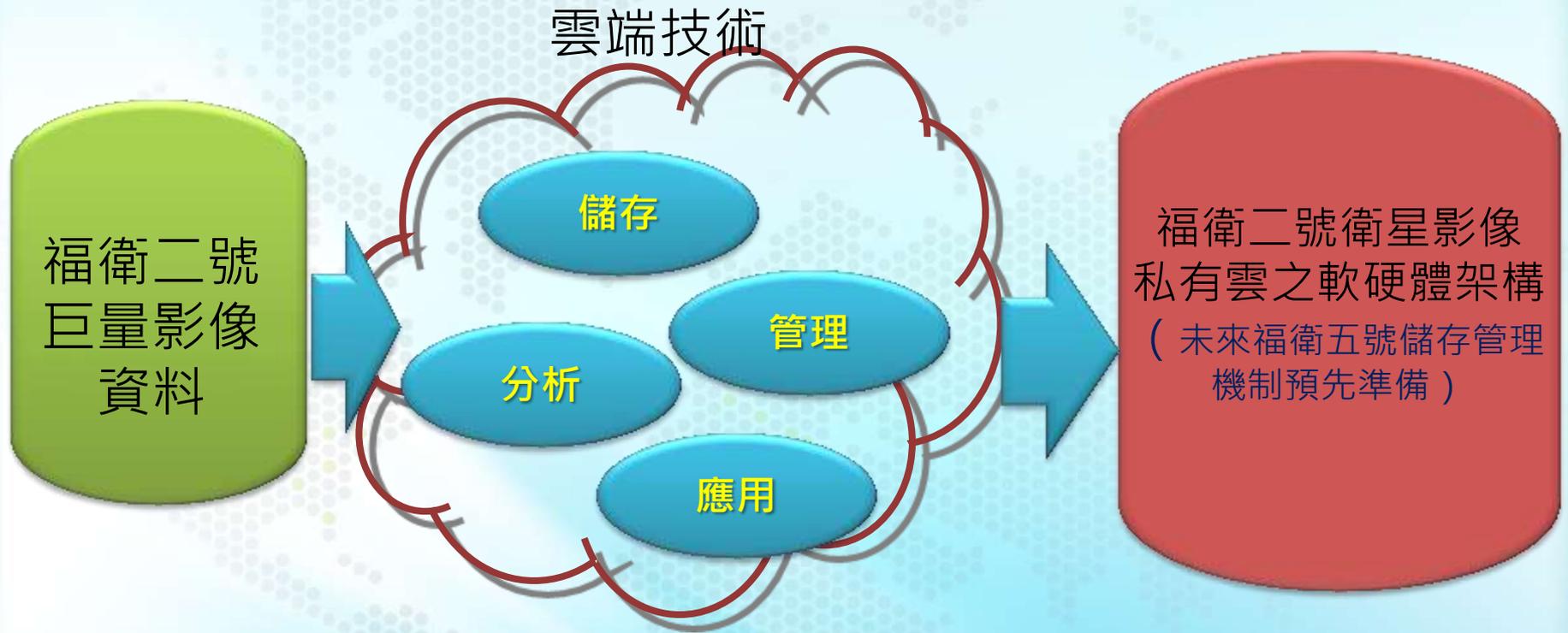
跨平台

以福衛二號為例

- 龐大的磁帶衛星影像資料
- 運用雲端技術對巨量影像資料進行有效的儲存與管理
- 磁帶媒體資料轉換



基本的構想



福爾摩沙衛星二號



我國自主擁有的第一枚遙測衛星

2004年5月21日成功發射

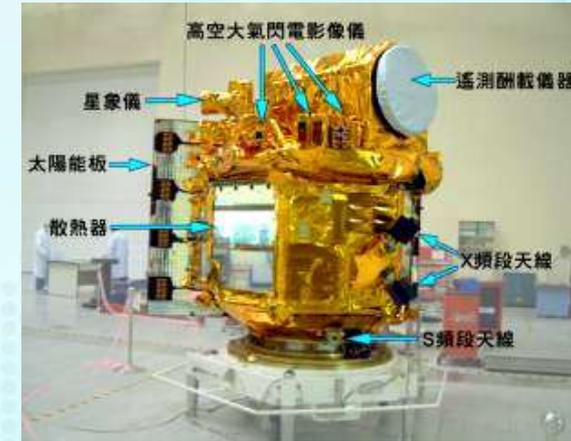
任務：對台灣及全球陸地及海域進行近實時之遙測作業

白晝地區：拍攝的影像資料可應用於國土規劃、資源探勘、環境保護、防災救災 ...等

黑夜地區：進行對高層大氣象上閃電之自然現象科學觀測，可做為科學實驗研究

福爾摩沙衛星二號基本資料

- 地球遙測應用衛星
- 重量：760 公斤左右
- 尺寸：高 2.4 米，外徑1.6 米
- 軌道：高度 891 公里，太陽同步軌道，每日通過台灣上空二次
- 繞行地球一週時間：約 103 分鐘。
- 遙測對地解析度：黑白影像 2 公尺，彩色影像 8 公尺
- 任務壽命：5 年



酬載儀器

- 遙測照相儀(白天)



- 高空大氣閃電影像儀(夜間)



福衛二號於救災上的應用#1



- 2011/03/11 日本 9.0 大地震引發海嘯
- 2011/03/12 早上，福衛二號拍到宮城影像
- 由衛星空拍圖可以看出海岸被海嘯吞噬約 5 公里。

福衛二號於救災上的應用#2

● 日本仙台地區影像

海嘯前衛星影像



海嘯後衛星影像



為什麼要導入雲端？

- 福衛二號巨量影像資料管理

既有資料現況

- 資料量：自發射升空到2012年1月30日
 - ▶ 國內站：
 - 儲存磁帶格式：DLT-IV(40GB) / AIT-3(100GB)
 - 共958捲，約42.29TB
 - ▶ 國外站：
 - 儲存磁帶格式：AIT-3(100GB)
 - 共7764捲，約543TB



面臨潛在的問題

AIT-3磁帶機已停產

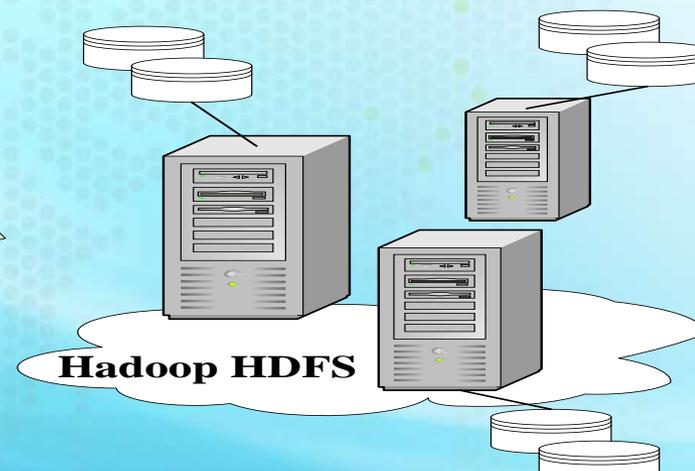
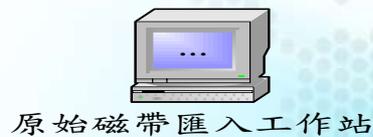
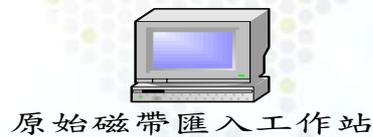
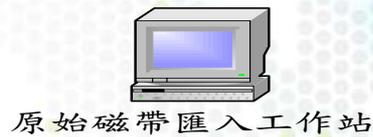
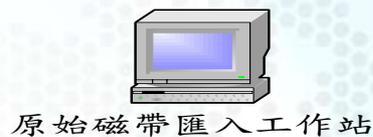
取得衛星影像資料程序

磁帶保存問題

衛星影像增值應用

原始磁帶資料匯入作業

- 將原始AIT-3磁帶，轉儲至LTO-4磁帶
- 轉換過程同時寫一份至HADOOP HDFS



福衛磁帶資料備份查詢介面

- 透過查詢介面，檢索及下載原始影像資料

利用日期與軌道編號進行查詢



利用日期與軌道編號查詢結果進行資料下載





解決了原始衛星影像磁帶保存問題，
接下來呢？

原始衛星影像資料轉換

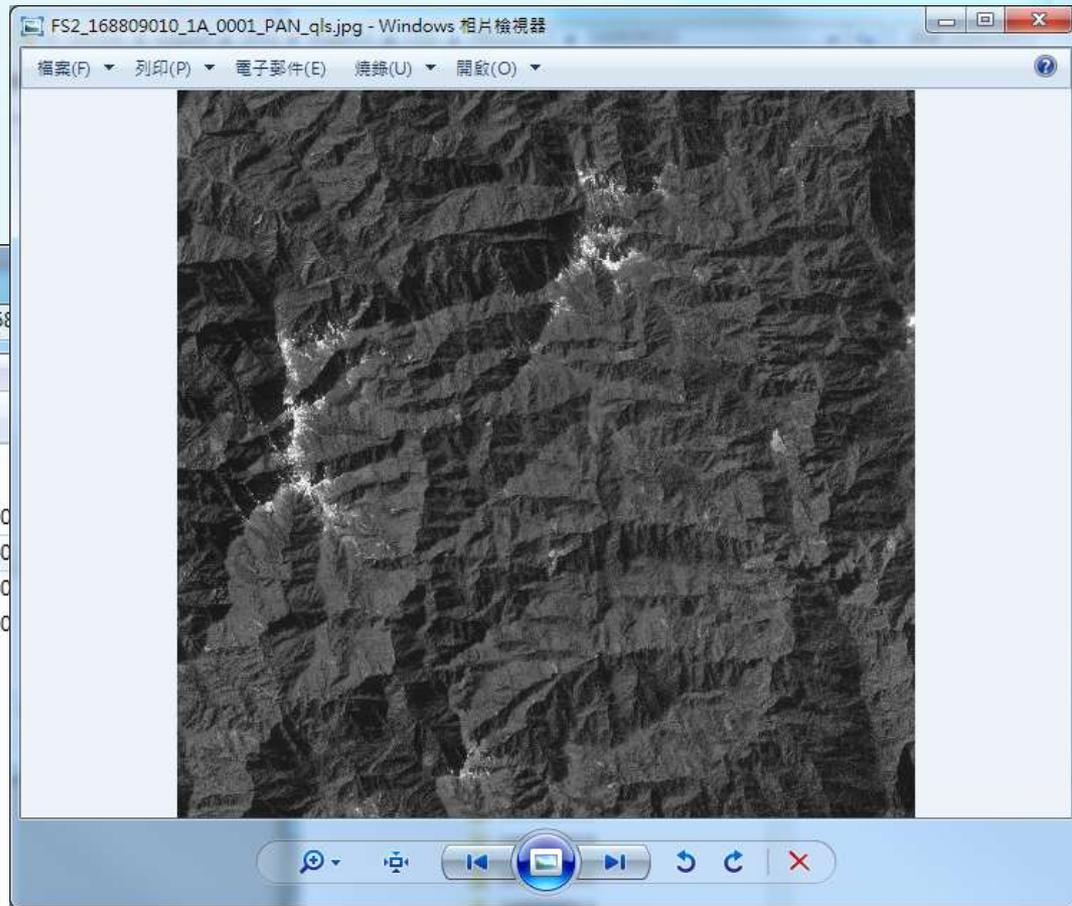
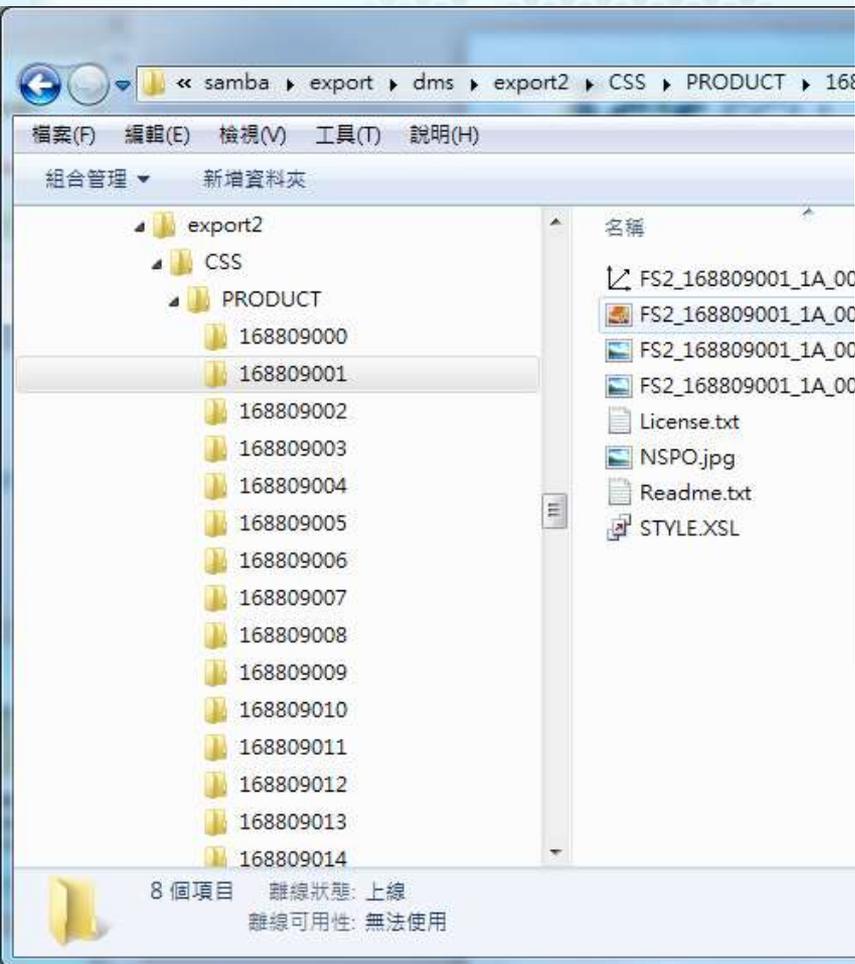


- 原始衛星資料記錄的是sensor回傳的數據
- 必需經過解算及處理後，才能變成衛星照片
- 受限於衛星影像處理模式，目前轉換方式僅能以單機方式運作
- 平均處理一個軌道資料需耗時約50分鐘

轉換程式的執行程序

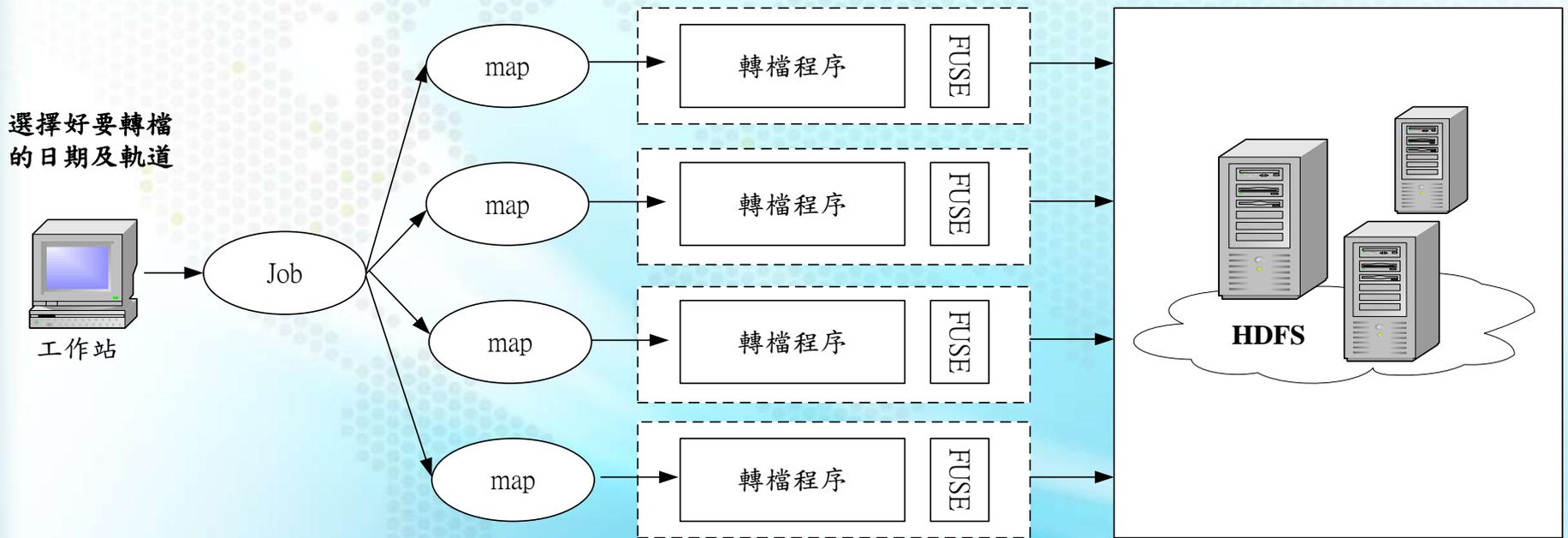
- # ./IDIBatchPan <Filename> <Product_parameter>
- # ./IDIBatchMS <Filename> <Product_parameter>
- # ./DPSmainrun <Product_parameter>
- # ./CSSBatchrun <MS_SegmentID> <PAN_SegmentID>
<Product_parameter>

轉換後的結果



結合Hadoop MapReduce平行轉檔

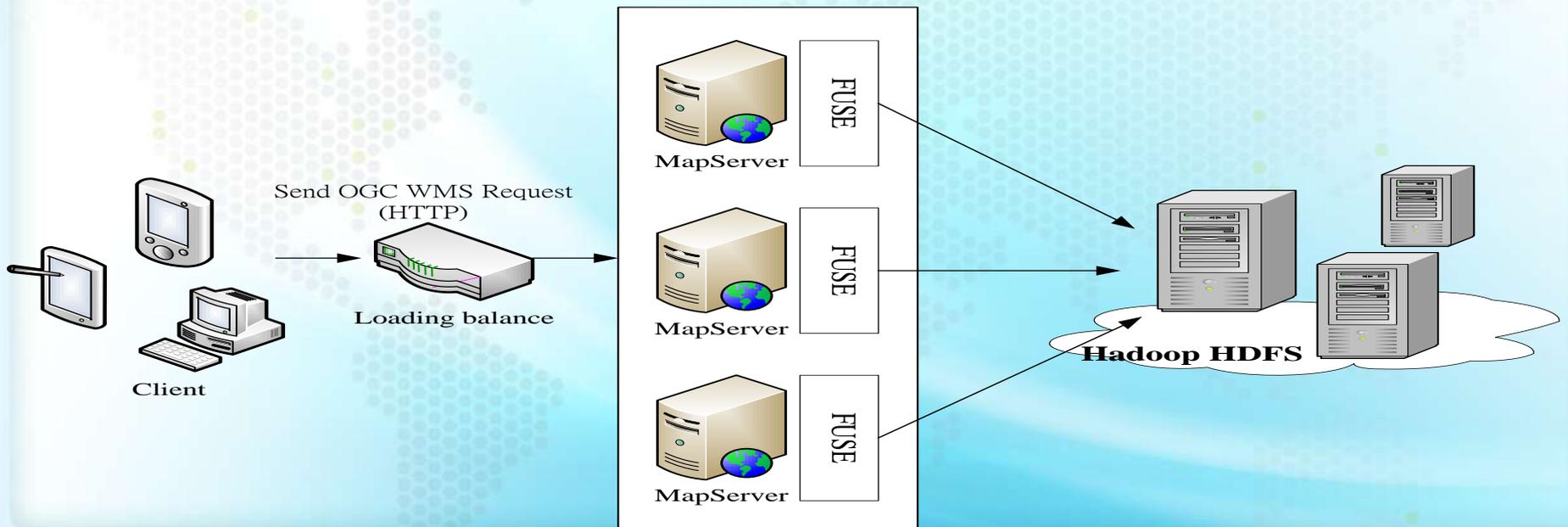
- 使用者透過介面選擇轉檔的日期及軌道
- 經由MapReduce分散處理，加速轉換時間



衛星影像有什麼用途呢？又該如何提供給使用者？

分散式動態WMS服務

- WMS(網路地圖服務)：提供即時網路地圖的服務
- 優點：伺服器接收到服務請求後，會立即讀取原始資料，並產生地圖影像



- Open Source Geospatial Foundation

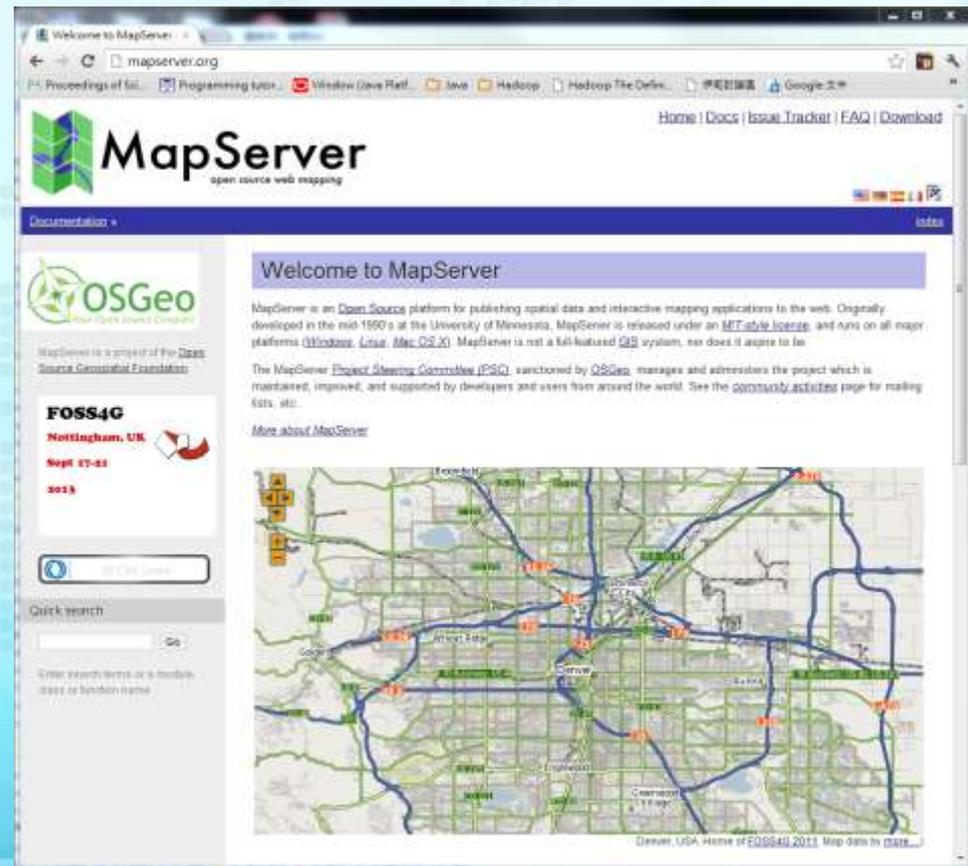
- ▶ Open Source GIS platform
- ▶ <http://mapserver.org/>

- 支援平台

- ▶ Microsoft/Linux/...

- 發佈地圖服務

- ▶ WMS
- ▶ WCS
- ▶ WFS



MapServer WMS MapFile/Sample

- <http://hadoop.gis.tw/cgi-bin/mapserv?map=/var/www/Taiwan.map&LAYERS=taiwan&TRANSPARENT=true&FORMAT=image/png&SERVICE=WMS&VERSION=1.1.1&REQUEST=GetMap&STYLES=&SRS=EPSG:4326&BBOX=120,20,123,25.3&WIDTH=768&HEIGHT=1024>

MAP

```
NAME "taiwan"
STATUS ON
#PROJ_LIB "/usr/share/proj"
```

...(略)...

```
### HDFS-map #####
```

LAYER

```
NAME "Taiwan"
```

```
METADATA
```

```
"wms_title" "taiwan" ##required
```

```
END
```

```
TYPE RASTER
```

```
STATUS ON
```

```
DATA "/mnt/hdfs/map/taiwan.tif"
```

```
END # Layer
```

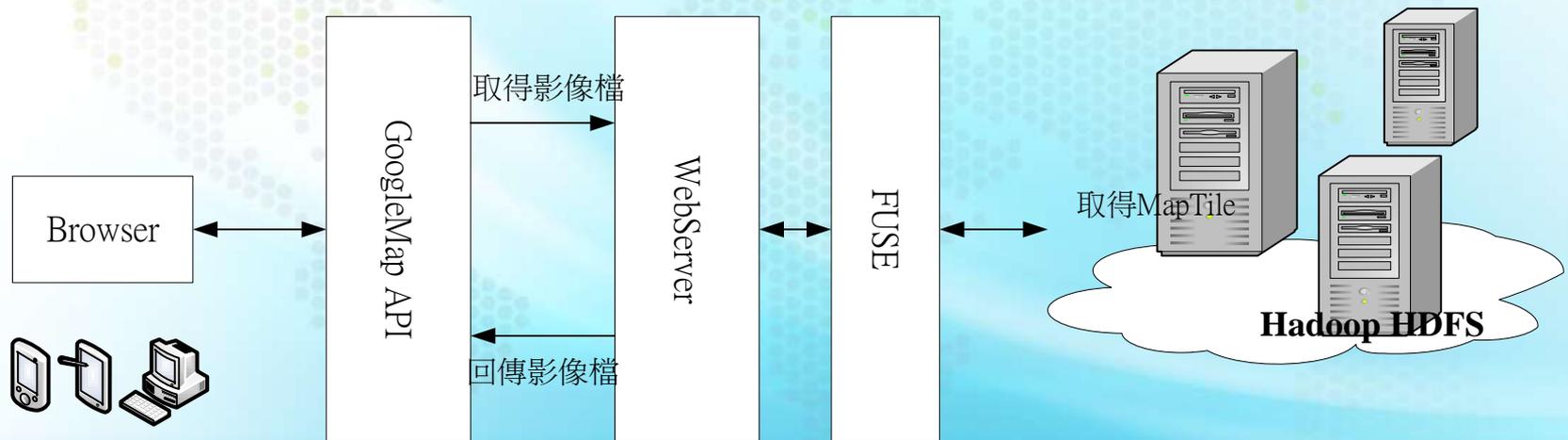
```
END # MAP
```



WMS固然好用，可以獲得取即時的空間資訊，但當面臨用戶端激增時，繁重的即時圖形產製工作，仍會導致回應速度不佳的問題

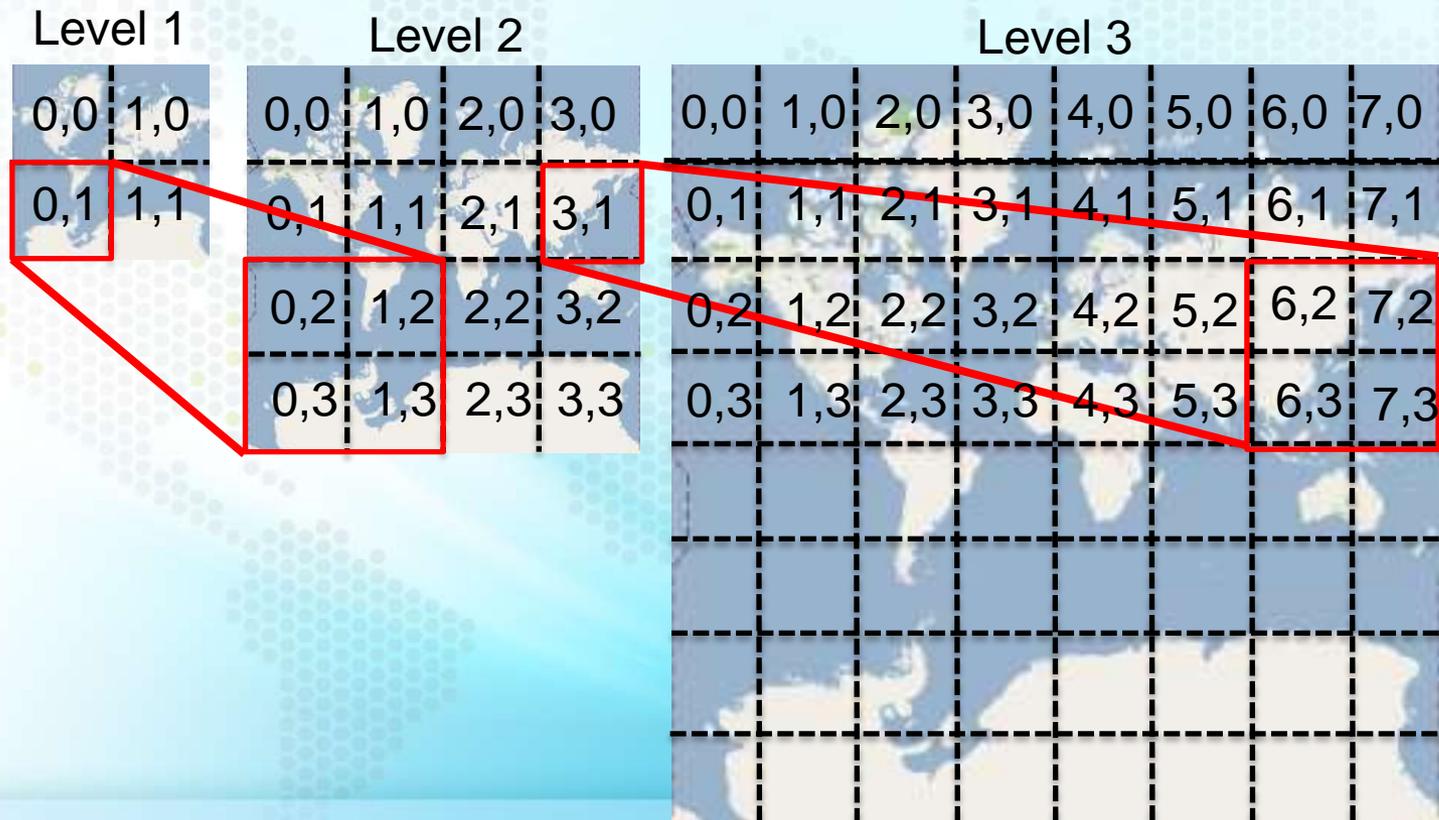
分散式地圖圖磚服務機制

- 針對異動率低的基礎底圖，採用事先處理機制，將地圖事前切割成不同比例尺的圖磚(MapTile)
- Google Map為最成功的案例
- 結合GoogleMap API，提供另一種圖層類型的選擇



全世界圖磚數量大概有多少

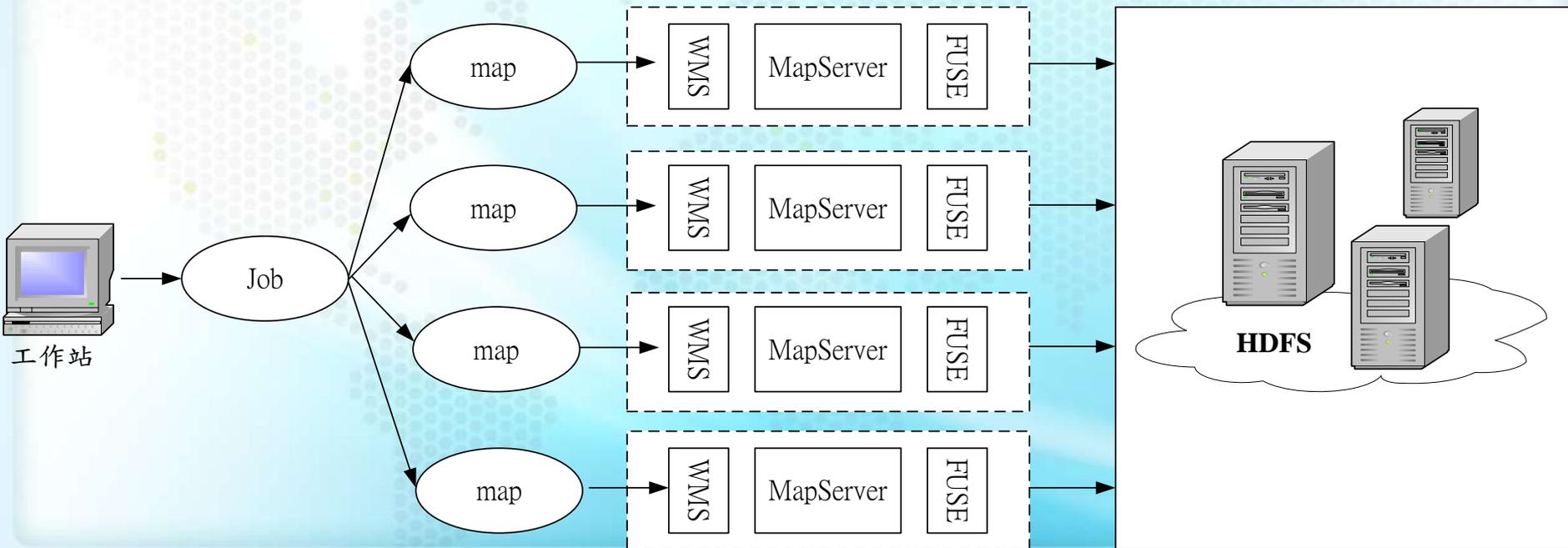
- 依據Google Map Tile的計算方式
- 總共19層，依四分樹方式進行分割
- 每一層數量公式： $2^n * 2^n$ ， n 為階層數
 - ▶ $n=16$ ，tiles=4,294,967,296



但要如何快速產生圖磚呢？

MapTile 平行產製

- 運用MapReduce平行化切圖程序
- 每一個節點安裝MapServer套件
- 透WMS取得每個Tile影像
- 切好的Tile儲存於HDFS



Map數量控制

- Map的數量取決於block的數量
- 依據使用者指定的地圖參數，產生若干的處理作業檔
- **bin/hadoop jar MapTile.jar zoom wmsUrl minx miny maxx maxy**



Zoom, x, miny, maxy
11,1700,870,901

MapTile產製結果

目錄結構

▶ Level

■ X軸

┆ 0.png (Y軸)

▶ 1

■ 0

┆ 0.png

┆ 1.png

■ 1

┆ 1.png

┆ 2.png

▶ 2

■ 0

┆ 0.png

┆ 1.png

┆ 2.png

┆ 3.png

■ 1

■ 2

■ 3



Contents of directory [/map/12/3420](#)

Goto :

[Go to parent directory](#)

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
1741.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1742.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1743.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1744.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1745.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1746.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1747.png	file	0.11 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1748.png	file	0.8 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1749.png	file	0.96 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup
1750.png	file	0.96 KB	3	64 MB	2012-09-26 18:51	rw-r--r--	hadoop	supergroup



1761.png



1762.png



1763.png



1764.png



1765.png



1766.png



1767.png



1768.png



1769.png



1770.png



1771.png



1772.png



1773.png



1774.png



1775.png



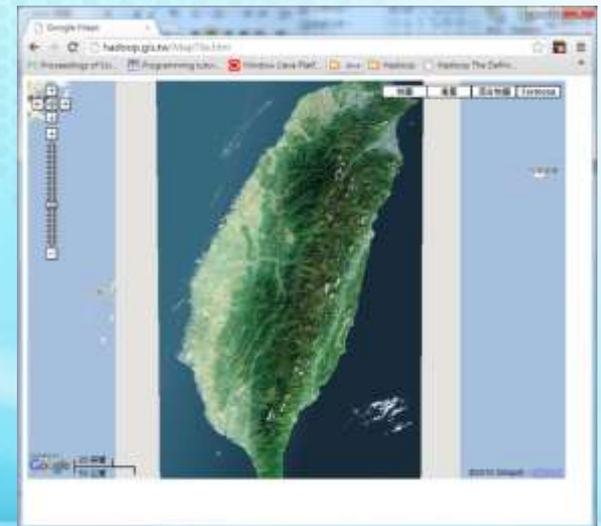
1776.png

與Google Map整合

```

var tileLayer = new google.maps.TileLayer(copyrightCollection, 1, 19);
tileLayer .getTileUrl = function(coord, zoom){
    var long =tile2long(coord.x, zoom) ;
    var lat = tile2lat(coord.y, zoom);
    if (zoom>=8 && zoom<=11 && lat>=21 && lat<=26 && long>=119 && long<=122) {
        return "http://hadoop.gis.tw/map/"+zoom+ "/" + coord.x+"/"+coord.y+".png"
    } else {
        return G_NORMAL_MAP.getTileLayers()[0].getTileUrl(coord, zoom);
    }
};

```



MapTile平行產製-優點

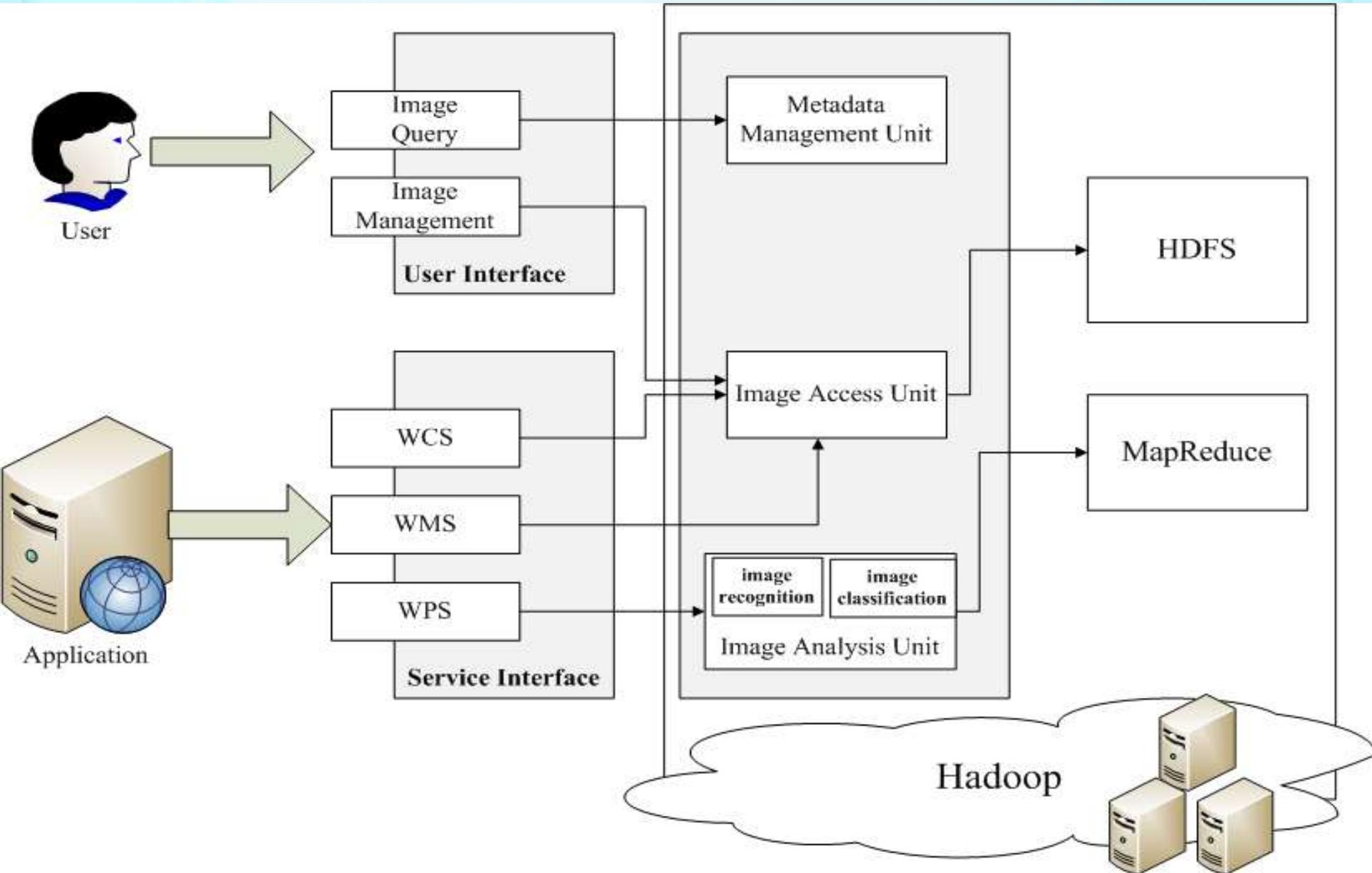
- MapServer免費，沒有額外的軟體採購費用
- MapServer採用CGI運作模式
 - ▶ 每次呼叫都是一個獨立process
 - ▶ 適合平行化使用
- 透過FUSE可以直掛載位於HDFS地圖資料
 - ▶ 資料分散，避免單一磁碟負載過重

未來的GIS雲端應用 -空間資訊雲

潛在的GIS運算應用



空間資訊雲架構



空間內插運算

常用的GIS分析功能，利用已知點資訊推估未知點資料。

例如：內插演算法

▶ 距離反比權重演算法(IDW)

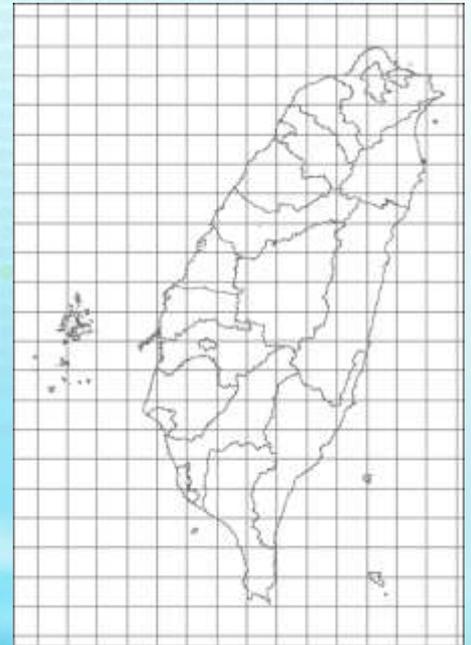
$$z(x, y) = \left[\sum_{i=1}^N w(d_i) Z_i \right] / \left[\sum_{i=1}^N w(d_i) \right]$$

解析度

1200 * 1600 (網格數量:192萬)

2400 * 3200 (網格數量:768萬)

3600 * 4800 (網格數量:1,728萬)

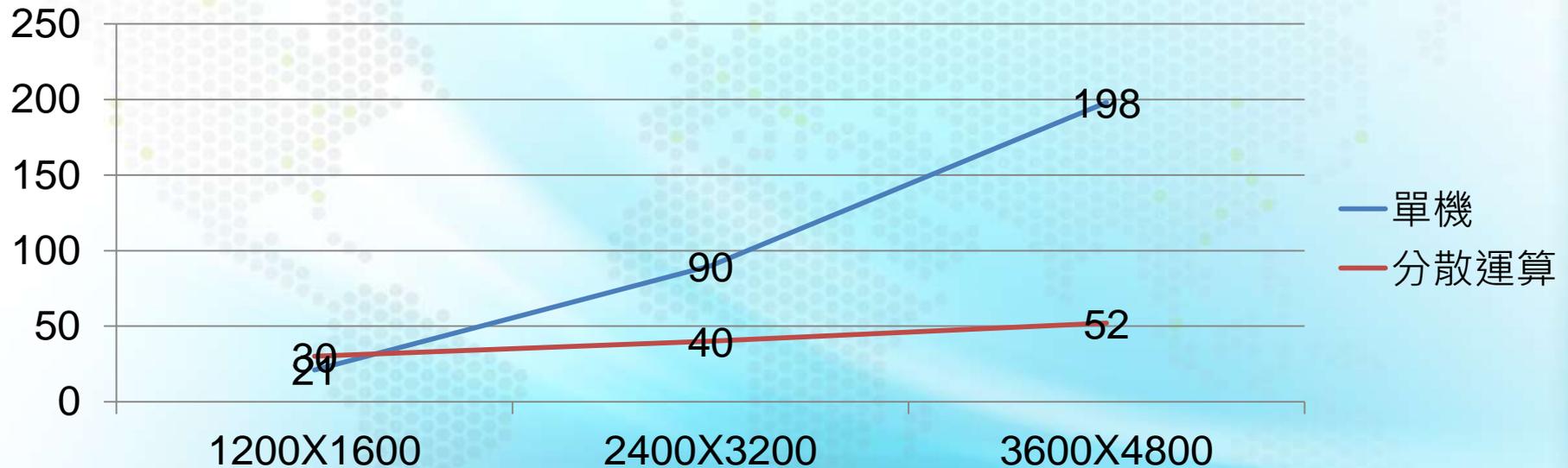


效能測試

- 單機及分散運算效能比較
- 測試硬體環境
 - ▶ 作業系統: Ubuntu Linux 11.10
 - ▶ CPU: i5-2400
 - ▶ RAM: 12MB
- Hadoop Cluster
 - ▶ 3 Nodes (硬體規格如上)
 - ▶ 每個Node最大同時運行工作數量: 10

測試結果

- 由於Hadoop分散運算Job初始化時間較長
- 對於少量運算，分散運算不見得佔有優勢
- 隨運算量增加，分散運算與單機運算明顯出現差異



結語

- 雲端運算，激發了GIS應用的發展
- 原本需要耗費長時間的空間分析，藉由雲端運算可以大幅縮短運算時間
- Hadoop + MapServer結合，是一個節省成本的GIS雲解決方案之一
- 未來Hadoop + GIS的整合還有很多努力的空間，但其應用卻也有無限可能

簡報結束
敬請指教