科技新知

Apache Spark：巨量資料分析軟體，由Scala開發。運算速度快、可在雲端平台執行、支援多種語言(python、java、R)、GraphY是Spark上的分散式圖形處理框架

物聯網：應用層、網路層、感知層

雲端

佈屬模式：公有雲、私有雲、混和雲、社群雲

服務模式：IaaS PaaS SaaS

特徵：

on-demand self-service(隨選自助服務)、

Broad Network access(隨時隨地用任何網路裝置存取)、

Resource Pooling(多人共享資源池)、

Rapid Elasticity(快速重新佈署靈活度)、

Measured Service(被監控與量測服務)

中華電信的HiCloud、amazon的EC2是Iaas ( 提供各等級的運算資源、儲存空間和網路流量，並以此計價 )

Google maill、salesforce.com的CRM(costermer relationship manarge)

Google docs

google engine

amazon SQS/SNS

Microsoft Azure

在多層式(Muti-tiers)平台裡，

客戶端電腦：存取的程式為呈現層(presentation)的部分

伺服端電腦：業務邏輯層(business logic layer)

伺服端電腦：資料層(data layer)

AI的兩大流派：法則學派(EX:專家系統)、機器學習學派

機器學習

**監督式學習**常用的演算法：二元分類、多元分類、回歸分析 ~~分群~~

**傳統統計模型：**

|  |  |
| --- | --- |
| 分類模型 | 決策樹、  區別分析、  貝斯統計分析、  支援向量機 |
| 預測 | 線性規劃  決策樹  隨機森林 |

**神經網路深度學習法：**

**非監督式學習**：分群(cluster analysis)、連結分析(asoociation rule)、維度縮減(dimensionality reduce)

**分群**

分群演算法：集群分析、k-means、馬可夫模型、類神經網路

分群演算法的績效衡量簡單明暸：**組間差異大，組內差異小**

所謂的**差異**指的就是觀測值之間的距離遠近作為衡量，最常見還是使用[歐氏距離（Euclidean distance）](https://en.wikipedia.org/wiki/Euclidean_distance)

以距離作為度量，在資料的預處理程序中，與 k-Nearest Neighbors 分類器一樣我們必須將所有的數值型變數標準化（Normalization），避免因為單位不同，在距離的計算上失真。

K-Means 演算法可以非常快速地完成分群任務，但是如果觀測值具有雜訊（Noise）或者極端值，其分群結果容易被這些雜訊與極端值影響，適合處理分布集中的大型樣本資料。

<https://ithelp.ithome.com.tw/articles/10187314>

與 K-Means 演算法不同的地方在於不需要事先設定 k 值，Hierarchical Clustering 演算法每一次只將兩個觀測值歸為一類，然後在演算過程中得到

k = 1 一直到 k = n（觀測值個數）群的結果。

Big data(巨量資料)

<https://zh.wikipedia.org/wiki/%E5%A4%A7%E6%95%B8%E6%93%9A>

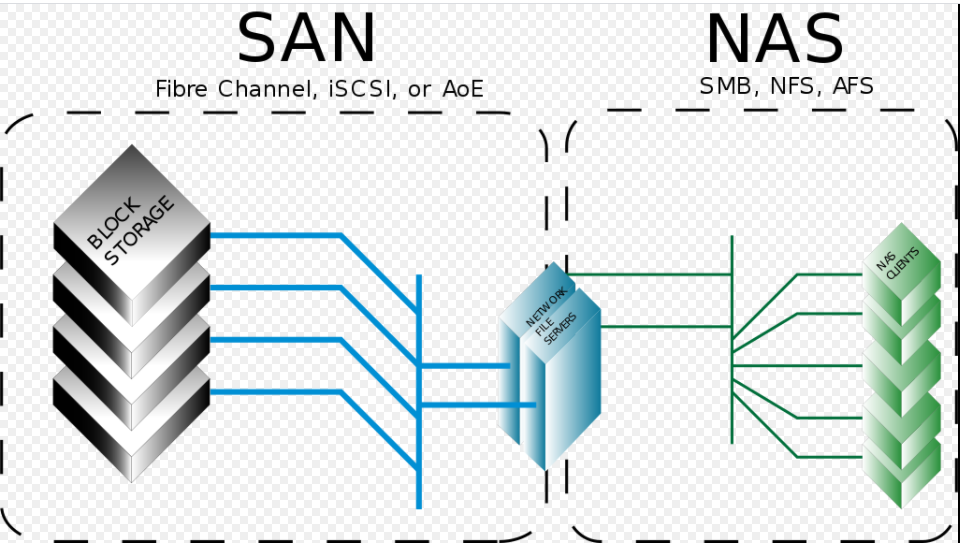
特性：資料量大(Volume,TB~PB)、資料輸入出速度(Velocity)、多樣性(Variety)、真實性（Veracity）、價值性(Value)

NAS(Network-attached storage)

<https://zh.wikipedia.org/wiki/%E7%B6%B2%E8%B7%AF%E9%99%84%E5%8A%A0%E5%84%B2%E5%AD%98>

採用NAS的優點：

1. 有標準協定，支援多種平台
2. 透過網路連接，配置距離彈性較大
3. 各式應用的伺服器不需額外處理資料存取管理作業
4. ~~不具備檔案系統~~



SAN（[儲存區域網路](https://zh.wikipedia.org/wiki/%E5%AD%98%E5%82%A8%E5%8C%BA%E5%9F%9F%E7%BD%91%E7%BB%9C)）和NAS

資料倉儲

**線上分析處理**（英語：Online analytical processing），簡稱**OLAP** ([/ˈoʊlæp/](https://zh.wikipedia.org/wiki/Help:%E8%8B%B1%E8%AA%9E%E5%9C%8B%E9%9A%9B%E9%9F%B3%E6%A8%99))，是[電腦技術](https://zh.wikipedia.org/wiki/%E8%AE%A1%E7%AE%97_(%E8%AE%A1%E7%AE%97%E6%9C%BA%E7%A7%91%E5%AD%A6))中快速解決[多維分析問題](https://zh.wikipedia.org/w/index.php?title=%E5%A4%9A%E7%BB%B4%E5%88%86%E6%9E%90%E9%97%AE%E9%A2%98&action=edit&redlink=1)（MDA）的一種方法。

OLAP工具讓使用者能夠從多個角度互動地分析多維資料。OLAP由三個基本的分析操作組成：上卷（roll-up）、鑽取（drill-down）、切片（slicing）和切塊（dicing）。[[6]](https://zh.wikipedia.org/wiki/%E7%B7%9A%E4%B8%8A%E5%88%86%E6%9E%90%E8%99%95%E7%90%86#cite_note-OBrien-6):402-403 上卷涉及可以在一個或多個維度中累積和計算的資料的聚合。