

巨量資料處理與分析 DPA

知識與技能規範	子項	細項
1. 資料科學與巨量資料	1.1 巨量資料的概念	1.1.1 巨量資料的定義與生活應用 1.1.2 處理巨量資料的常見支援工具(hadoop、mahout...等) 1.1.3 處理巨量資料的常見實現工具(R、Python、SAS、SPSS...等)
	1.2 資料科學的涵意	1.2.1 資料科學家的任務 1.2.2 統計學、電腦科學與商業解析 1.2.3 跨產業資料探勘通用流程
2. 資料處理的基礎	2.1 結構化資料與 SQL	2.1.1 屬性與尺度 2.1.2 資料表與資料庫 2.1.3 簡易 SQL 指令
	2.2 非結構化資料處理	2.2.1 非結構化資料種類 2.2.2 常用 NoSQL 資料庫
3. 統計分析的基礎	3.1 敘述統計	3.1.1 常用的統計值 3.1.2 常用的統計圖表
	3.2 推論統計	3.2.1 抽樣調查 3.2.2 估計與檢定
	3.3 相關分析與迴歸	3.3.1 相關分析 3.3.2 迴歸分析 3.3.3 變異數分析
4. 資料探勘與視覺化基礎	4.1 資料前處理	4.1.1 訓練集與測試集 4.1.2 資料清理 4.1.3 資料轉換 4.1.4 資料縮減
	4.2 資料探勘常用方法	4.2.1 決策樹的應用 4.2.2 購物籃分析與應用 4.2.3 集群分析的應用
	4.3 資料視覺化概念	4.3.1 儀表板 4.3.2 常見的資料視覺化圖型