

資料工程人員職能基準

版本	職能基準代碼	職能基準名稱	狀態	更新說明	發展更新日期
V2	SMS2529-001v2	資料工程人員	最新版本	略	2021/12/31
V1	SMS2529-001v1	資料工程人員	歷史版本	已被《SMS2529-001v2》取代	2019/12/19

職能基準代碼		SMS2529-001v2			
職能基準名稱 (擇一填寫)		職類			
		職業	資料工程人員		
所屬 類別	職類別	科學、技術、工程、數學 / 數學及科學		職類別代碼	SMS
	職業別	其他資料庫及網路專業人員		職業別代碼	2529
	行業別	專業、科學及技術服務業 / 研究發展服務業		行業別代碼	M7230
工作描述		依客戶之需求確認資料蒐集內容、建立自動化之資料介接流程並進行資料處理、轉置，以及數據平台之維運管理。			
基準級別		4			

主要職責	工作任務	工作產出	行為指標	職能 級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
T1 確認資料內容	T1.1 確認客戶與資料分析師需求	O1.1.1 需求確認書	P1.1.1 與客戶或資料分析師面談，以決定資料分析或處理之目的或需求 P1.1.2 將文件、紀錄等相關之需求資訊，與客戶或分析師確認。	3	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K03 資料結構知識 K04 組織資料蒐集處理相關政策	S01 資料蒐集技能 S02 資料結構分析技能 S03 資料需求分析技能 S04 溝通技能
	T1.2 建立資料蒐集清單	O1.2.1 資料清冊 (含資料 workflow)	P1.2.1 以結構化的方式將文件分類，並記錄資料蒐集的範圍及項目。	3	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K03 資料結構知識 K04 組織資料處理蒐集相關政策	S01 資料蒐集技能 S02 資料結構分析技能 S03 資料需求分析技能 S04 溝通技能

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
		程)				
	T1.3 確認資料來源與格式	O1.3.1 資料清冊 (含資料工作流程)	P1.3.1 文件化方式記錄各項欲蒐集資料之來源、數量、資料原始型態、資料提供之形式及其他相關必要資訊。	3	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K03 資料結構知識 K04 組織資料處理蒐集相關政策	S01 資料蒐集技能 S02 資料結構分析技能 S03 資料需求分析技能 S04 溝通技能

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
T2 設計、建置與開發數據平台 (Data Warehouse) 自動化之資料介接機制	T2.1 確認技術方案	O2.1.1 技術設計文件	<p>P2.1.1 依據資料量與特性，評估各儲存方案之資料接收及讀取方式、效率、可儲存容量以及擴充機制等所需規格資訊。</p> <p>P2.1.2 依據上述資料進行實機概念驗證 (POC)，並記錄驗證過程中之相關數據，比對各儲存方案之規格資訊，以釐清正確性。</p> <p>P2.1.3 依據客戶需求，建議最適之資料儲存方案，將技術規格設計細節製作為文件，並取得客戶與相關人員之同意。</p>	4	K05 資料結構 K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K09 資訊網路概論 K10 現行伺服器應用程式、相容性問題與解決程序 K11 作業系統支援公用程式 K12 資料查詢設計 K13 伺服器架構	S05 流程分析 S06 <u>資料庫技術</u> 【註1】應用 S07 <u>程式語言</u> 【註2】應用 S08 讀寫技能 S09 計算技能 S10 <u>分散運算與儲存平台技術</u> 【註3】應用 S11 客戶需求整合能力 S12 <u>巨量資料儲存分析平台</u> 【註4】操作 S13 <u>分散運算與儲存技術</u> 【註5】應用 S14 資料串流分析與技術應用 S15 <u>低結構化資料組織技術</u> 【註6】應用 S16 <u>低結構化資料處理</u> 【註7】 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
	T2.2 進行數據平台建置	O2.2.1 數據平台安裝文件 O2.2.2 數據平台組態文件 O2.2.3 數據平台權限設定文件	P2.2.1 確認數據平台安裝之所需軟硬體清單。 P2.2.2 製作安裝程序與配置之相關文件 (如安裝 SOP)。 P2.2.3 安裝數據平台軟體與相關服務。 P2.2.4 配置數據平台環境組態與服務。 P2.2.5 驗證數據平台軟體與服務功能，確保可正確運作。 P2.2.6 設定數據平台權限，包含管理者與使用者。	4	K05 資料結構 K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K09 資訊網路概論 K10 現行伺服器應用程式、相容性問題與解決程序 K11 作業系統支援公用程式 K12 資料查詢設計 K13 伺服器架構	S05 流程分析 S06 資料庫技術應用 S07 程式語言應用 S08 讀寫技能 S09 計算技能 S10 分散運算與儲存平台技術應用 S12 巨量資料儲存分析平台操作 S13 分散運算與儲存技術應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理
	T2.3 進行資料存放	O2.3.1 資料配置文件	P2.3.1 依據欲儲存之資料類型、來源或需求，設計資料集或資料標籤。 P2.3.2 設定各資料集組態，例如資料存放位置、資料欄位名稱、資料長度、資料最大使用空間等必要資訊。 P2.3.3 進行資料匯入作業，並驗證資料存放位置與組態設定一致性。	4	K05 資料結構 K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K09 資訊網路概論 K10 現行伺服器應用程式、相容性問題與解決程序	S05 流程分析 S06 資料庫技術應用 S07 程式語言應用 S08 讀寫技能 S09 計算技能 S10 分散運算與儲存平台技術應用

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
					K11 作業系統支援公用程式 K12 資料查詢設計 K13 伺服器架構	S12 巨量資料儲存分析平台操作 S13 分散運算與儲存技術應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理
	T2.4 建立自動化程序	O2.4.1 自動化程式設定或開發文件	P2.4.1 確認資料來源與型式與資料接收後之儲存方式，包含儲存位置，切分方式，資料格式與所需內容。 P2.4.2 依資料需求，評估建立資料爬蟲技術進行資料蒐集，並建立資料儲存機制或自動化技術，存放至數據平台。 P2.4.3 設定或開發資料拋接方式與內容過濾機制。 P2.4.4 在數據平台中透過適當之資料查詢方式，驗證資料接入結果是否與預期相同。	4	K05 資料結構 K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K09 資訊網路概論 K10 現行伺服器應用程式、相容性問題與解決程序 K11 作業系統支援公用程式 K12 資料查詢設計 K13 伺服器架構	S05 流程分析 S06 資料庫技術應用 S07 程式語言應用 S08 讀寫技能 S09 計算技能 S10 分散運算與儲存平台技術應用 S12 巨量資料儲存分析平台操作 S13 分散運算與儲存技術應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
						S19 資料數值處理 S20 資料編碼處理 S21 整合測試能力 S22 網路及資料爬蟲技術應用
T3 數據處理與轉置	T3.1 進行資料萃取	O3.1.1 資料集或資料表	<p>P3.1.1 由數據平台中確定所需資訊，並確認資料分析平台或資料呈現平台之資料接取機制。</p> <p>P3.1.2 使用適當的資料查詢語法或開發所需程式，將所需資料接取至資料分析平台或資料呈現平台。</p> <p>P3.1.3 在目的端上，計算資料接取完成率與正確率。</p>	4	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K05 資料結構 K06 程式語言 K12 資料查詢設計 K14 資料平台操作 K15 應用統計學 K16 網頁資料擷取概念 K17 文字字詞處理 K18 資料處理 K19 ETL 機制	S07 程式語言應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理 S23 資料分析工具 ^{【註8】} 應用 S24 資料處理語法 ^{【註9】} 應用 S25 ETL 相關軟體操作
	T3.2 進行資料描述	O3.2.1 資料描述資訊	<p>P3.2.1 針對數值型資料執行所需之統計分析，取得值域資訊，如資料筆數、最大值、最小值、中位數、欄位定義等必要資訊。</p> <p>P3.2.2 針對非數值資訊，取得值域資訊，如統計關鍵字詞的個數、類別數、分佈狀態集中或分散等相關資訊。</p>	4	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K05 資料結構 K06 程式語言 K12 資料查詢設計 K14 資料平台操作 K15 應用統計學 K16 網頁資料擷取概念	S07 程式語言應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
					K17 文字字詞處理 K18 資料處理 K19 ETL 機制	S23 資料分析工具應用 S24 資料處理語法應用 S25 ETL 相關軟體操作
	T3.3 處理異常資料	O3.3.1 調整完成之資料集或資料表	<p>P3.3.1 以適當之資料查詢語法或開發所需程式，逐筆驗證資料完整性，並針對不完整之資料，以適當方式進行修正或補正，並將修正結果回存。</p> <p>P3.3.2 以適當之資料查詢語法或開發所需程式，過濾出重複資料，並向相關人員確認處理方式，進行修正後將修正結果回存。</p> <p>P3.3.3 以適當之資料查詢語法或開發所需程式，過濾會影響資訊分析正確性之錯誤資訊（如格式錯誤、數值異常、類別錯誤等），進行修正後將修正結果回存。</p>	4	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K05 資料結構 K06 程式語言 K12 資料查詢設計 K14 資料平台操作 K15 應用統計學 K16 網頁資料擷取概念 K17 文字字詞處理 K18 資料處理 K19 ETL 機制	S07 程式語言應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理 S23 資料分析工具應用 S24 資料處理語法應用 S25 ETL 相關軟體操作

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
	T3.4 轉置資料	O3.4.1 轉置完成之資料集或資料表	<p>P3.4.1 與相關人員討論，確認後續分析所需之資料轉置需求，並完成紀錄。</p> <p>P3.4.2 以適當之資料查詢語法或開發所需程式，將資料內容轉換至可分析之型態，包含對文詞資料進行分類、置換、字詞轉置，為數值資料進行分類、貼標等必要作業。</p> <p>P3.4.3 調整資料轉置作業，確保資料轉置效率與正確率之最佳化。</p> <p>P3.4.4 驗證資料轉置完成率與正確率能符合作業之需求。</p> <p>P3.4.5 將轉置完成之資料，依據資料模型存入數據平台</p>	4	K01 資料應用場域的基本知識 K02 系統日誌相關知識 K05 資料結構 K06 程式語言 K12 資料查詢設計 K14 資料平台操作 K15 應用統計學 K16 網頁資料擷取概念 K17 文字字詞處理 K18 資料處理 K19ETL 機制	S07 程式語言應用 S14 資料串流分析與技術應用 S15 低結構化資料組織技術應用 S16 低結構化資料處理 S17 資料過濾技術應用 S18 資料字串處理 S19 資料數值處理 S20 資料編碼處理 S23 資料分析工具應用 S24 資料處理語法應用 S25ETL 相關軟體操作
T4 數據平台維運與管理	T4.1 建立數據平台管理機制	O4.1.1 監控機制設定與組態文件 O4.1.2 監控日誌	<p>P4.1.1 建立數據平台的效能與容量標準。</p> <p>P4.1.2 針對數據平台或資料分析平台之系統資源進行管理 (CPU/GPU)。</p> <p>P4.1.3 建立監控機制，以確保數據平台符合效能與容量標準。</p> <p>P4.1.4 將相關監控機制留存於日誌，以追蹤未達標準之情況，並主動進行改善。</p>	3	K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K20 系統功能性 K21 資訊安全 K22 個資保護相關法規 K23 伺服器管理	S07 程式語言應用 S08 讀寫技能 S10 分散運算與儲存平台技術應用 S12 巨量資料儲存分析平台操作 S26 作業系統操作

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
	T4.2 確保數據平台安全與完整性	O4.2.1 數據平台權限申請程序	<p>P4.2.1 依資訊安全政策與規範，建立數據平台使用之權限申請程序。</p> <p>P4.2.2 建立數據平台資料使用之相關規範，如必要之去識別化與遮罩、資料攜出提供原則等，以確保符合現行相關法規與組織內之相關要求。</p> <p>P4.2.3 建立數據平台活動監控機制，包含使用操作、運作情況與網路活動監控等，以確保數據平台安全性。</p>	3	K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K20 系統功能性 K21 資訊安全 K22 個資保護相關法規 K23 伺服器管理	S07 程式語言應用 S08 讀寫技能 S10 分散運算與儲存平台技術應用 S12 巨量資料儲存分析平台操作 S26 作業系統操作
	T4.3 進行數據平台的備份與最佳化處理	O4.3.1 資料備份原則管理文件 O4.3.2 資料備份復原操作 SOP O4.3.3 資料還原演練紀錄	<p>P4.3.1 建立數據平台備份復原程序 (如 Business continuously Plan)。</p> <p>P4.3.2 設定或建立數據平台備份作業。</p> <p>P4.3.3 定期進行資料復原演練，並確認資料還原正確性。</p>	3	K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制 K20 系統功能性 K21 資訊安全 K22 個資保護相關法規 K23 伺服器管理	S07 程式語言應用 S08 讀寫技能 S10 分散運算與儲存平台技術應用 S12 巨量資料儲存分析平台操作 S26 作業系統操作
	T4.4 留存數據平台操作軌跡	O4.4.1 數據平台日誌管理文	<p>P4.4.1 建立數據平台日誌留存程序。</p> <p>P4.4.2 配置數據平台或開發所需程式，針對數據平台操作、運作之相關軌跡以日誌方式記</p>	3	K06 程式語言 K07 關聯式與非關聯式資料庫 K08 資料儲存機制	S07 程式語言應用 S08 讀寫技能 S10 分散運算與儲存平台技術應

主要職責	工作任務	工作產出	行為指標	職能級別	職能內涵 (K=knowledge 知識)	職能內涵 (S=skills 技能)
		件 O4.4.2 數據平台操作日誌	錄。		K20 系統功能性 K21 資訊安全 K22 個資保護相關法規 K23 伺服器管理	用 S12 巨量資料儲存分析平台操作 S26 作業系統操作

職能內涵 (A=attitude 態度)

- A01主動積極：不需他人指示或要求能自動自發做事，面臨問題立即採取行動加以解決，且為達目標願意主動承擔額外責任。
- A02自我管理：設立定義明確且實際可行的個人目標；對於及時完成任務展現高度進取、努力、承諾及負責任的行為。
- A03謹慎細心：對於任務的執行過程，能謹慎考量及處理所有細節，精確地檢視每個程序，並持續對其保持高度關注。
- A04持續學習：能夠展現自我提升的企圖心，利用且積極參與各種機會，學習任務所需的新知識與技能，並能有效應用在特定任務。
- A05好奇開放：容易受到複雜新穎的事物吸引，且易於接受新觀念的傾向。
- A06團隊意識：積極參與並支持團隊，能彼此鼓勵共同達成團隊目標。

說明與補充事項

- 建議擔任此職類/職業之學歷/經歷/或能力條件：
 - 大專以上資訊相關科系畢業，或具備2年以上巨量資料處理相關工作經驗。
 - 全民英檢初級或同等語言能力以上。
- 其他補充說明：
 - 【註1】資料庫技術：如 SQL、NoSQL 等。
 - 【註2】程式語言：如 Java、C/C++、Python、.NET 等。
 - 【註3】分散運算與儲存平台技術：如 Hadoop、Hive 等。
 - 【註4】巨量資料儲存分析平台：如 Hadoop、Splunk、Mongo DB、ELK 等。

說明與補充事項

- 【註5】分散運算與儲存技術：如 Hadoop、Hive 等。
- 【註6】低結構化資料組織技術：如 HTML、XML、JSON 等。
- 【註7】低結構化資料處理：如文字、影音、影像與機器資料處理等。
- 【註8】資料分析工具：如 SAS、SPSS、Splunk、R、Python 等。
- 【註9】資料處理語法：如 SQL、R、Splunk language 等。