

資料工程人員職能基準

職能基準代碼		SMS2529-001v1			
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、轉製，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 ^{【註1】} 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系統日誌本身之意涵 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
	T1.2建立資料蒐集清單	O1.2.1欲蒐集之資料清冊(含資料工作流程)	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。			
	T1.3確認資料來源與格式	O1.3.1欲蒐集之資料清冊(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。			

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(含網際網路與物聯網) 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(協助與支援公用程式) K12資料查詢設計 K13伺服器架構	S05流程分析 S06資料庫技術 ^{【註2】} 應用 S07程式語言 ^{【註3】} 應用 S08讀寫技能 ^{【註4】} S09計算技能 ^{【註5】} S10分散運算與儲存平台技術 ^{【註6】} 應用 S11整合客戶需求 ^{【註7】} S12巨量資料儲存分析平台 ^{【註8】} 操作 S13分散運算與儲存技術 ^{【註9】} 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 ^{【註10】} 應用 S16低結構化資料處理 ^{【註11】} 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 S21整合測試能力 S22網路及資料爬蟲技術應用
	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	<p>P2.2.1確認數據平台安裝之所需軟硬體清單。</p> <p>P2.2.2製作安裝程序與配置之相關文件(如安裝 SOP)。</p> <p>P2.2.3安裝數據平台軟體與相關服務。</p> <p>P2.2.4配置數據平台環境組態與服務。</p> <p>P2.2.5驗證數據平台軟體與服務功能，確保可正確運作。</p> <p>P2.2.6設定數據平台權限，包含管理者與使用者。</p>			
	T2.3進行資	O2.3.1資料配	P2.3.1依據欲儲存之資料類型、來源或			

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)
	料存放	置文件	<p>需求，設計資料集或資料標籤。</p> <p>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。</p> <p>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。</p>			
	T2.4建立自動化程序	O2.4.1自動化程式設定或開發文件	<p>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。</p> <p>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。</p> <p>P2.4.3設定或開發資料拋接方式與內容過濾機制。</p> <p>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。</p>			
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	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。	4	K01 資料應用場域的基本知識 K02 系統日誌本身之意涵	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

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)
			<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>		<p>K05資料結構</p> <p>K06程式語言</p> <p>K12資料查詢設計</p> <p>K14資料平台操作</p> <p>K15應用統計學</p> <p>K16網頁資料擷取概念</p> <p>K17文字字詞處理</p> <p>K18資料處理</p> <p>K19 ETL 機制</p>	<p>S16低結構化資料處理</p> <p>S17資料過濾技術應用</p> <p>S18資料字串處理</p> <p>S19資料數值處理</p> <p>S20資料編碼處理</p> <p>S23資料分析工具^{【註12】}應用</p> <p>S24資料處理語法^{【註13】}應用</p> <p>S25 ETL 相關軟體操作</p>
	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>			
	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</p>			

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)
			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。			
	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、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>			
T4數據平台維運與管理	T4.1建立數據平台管理機制	<p>O4.1.1監控機制設定與配置文件</p> <p>O4.1.2監控日</p>	<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>	3	<p>K06程式語言</p> <p>K07關聯式與非關聯式資料庫</p> <p>K08資料儲存機制</p>	<p>S07程式語言應用</p> <p>S08讀寫技能【註14】</p> <p>S10分散運算與儲存平台技術應用</p> <p>S12巨量資料儲存分析平台操作</p>

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)
		誌	<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>		<p>K20系統功能性</p> <p>K21資訊安全</p> <p>K22個資保護相關法規</p> <p>K23伺服器管理</p>	S26作業系統操作
	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>			
	T4.3進行數據平台的備份與最佳化處理	<p>O4.3.1資料備份原則管理文件</p> <p>O4.3.2資料備份復原操作SOP</p> <p>O4.3.3資料還</p>	<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>			

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)
		原演練紀錄				
	T4.4留存數據平台操作軌跡	O4.4.1數據平台日誌管理文件 O4.4.2數據平台操作日誌	P4.4.1建立數據平台日誌留存程序。 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錄。			

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自我提升：能夠展現持續學習的企圖心，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年以上資訊相關工作經驗。
 - 全民英檢初級或同等語言能力以上。
- 其他補充說明：
 - 【註1】客戶：係指接受資料工程師服務之組織。
 - 【註2】資料庫技術：如 SQL, NoSQL 等。

說明與補充事項

- 【註 3】程式語言：如 Java, C/C++, Python, .NET 等。
- 【註 4】讀寫技能：包括解讀技術文件、閱讀與解讀企業程序/手冊/規範、評估與呈現資訊、依照規定格式撰寫報告等。
- 【註 5】計算技能：包括進行測試測量、解讀結果等。
- 【註 6】分散運算與儲存平台技術：如 Hadoop, Hive 等。
- 【註 7】整合客戶需求：在進行設計、安裝數據平台伺服器與相關網路服務等工作時，須能發掘與評估技術方案，考量相關限制與可管理性等問題。
- 【註 8】巨量資料儲存分析平台：如 Hadoop, Splunk, Mongo DB, ETK 等。
- 【註 9】分散運算與儲存技術：如 Hadoop, Hive 等。
- 【註 10】低結構化資料組織技術：如 HTML/XML/JSON 等。
- 【註 11】低結構化資料處理：如文字、影音、影像與機器資料處理等。
- 【註 12】資料分析工具：如 SAS, SPSS, Splunk, R, Python 等。
- 【註 13】資料處理語法：如 SQL, R, Splunk language 等。
- 【註 14】讀寫技能：此處所指之讀寫技能，包括閱讀與解讀企業程序/手冊/規範、遵守製造商建議、依照組織標準修改文件等。