IBM SPSS Modeler CRISP-DM 手冊

IBM

附註

使用本資訊及其所支援的產品之前，請先詳閱第 39 頁的『注意事項』中的資訊。

產品資訊

本版適用於 IBM SPSS Modeler 18.2.0 版及所有後續版本和修正，直到新版中另有指示。

目錄

前言

IBM® SPSS® Modeler 是 IBM Corp. 企業強度的資料採礦工作台。SPSS Modeler 可透過更深入地理解資料來協助組織改進客戶與居民的關係。組織使用從 SPSS Modeler 獲取的見解來保持可盈利的客戶、識別交叉銷售機會、吸引新客戶、偵測欺詐、降低風險以及提升政府服務交付。

SPSS Modeler 的視覺化介面可邀請使用者套用其特定的業務專門知識，從而得到更為強大的預測模型並縮短問題解決時間。SPSS Modeler 提供許多建模技術，例如預測、分類、分段以及關聯偵測演算法。建立了模型之後，IBM SPSS Modeler Solution Publisher 可讓交付從企業層面延伸到決策制訂者或資料庫。

關於 IBM Business Analytics

IBM Business Analytics 軟體提供完整、一致且準確的資訊，決策者可信任此資訊，並藉以改善營運績效。商業智慧、預測分析、財務績效及策略管理，以及分析應用程式的綜合性產品組合會對現行績效提供清晰、即時而可行的洞察，且能夠預測未來結果。結合了豐富的業界解決方案、有效實證和專業服務，每種規模的組織都能引爆最高效能，確實自動化執行決策，並且交付更棒的成果。

作為此產品組合的一部分，IBM SPSS Predictive Analytics 軟體可協助組織預測未來事件，並根據促進較佳業務結果的洞察，主動採取行動。世界各地的商業、政府和學術客戶都依賴 IBM SPSS 技術，它在吸引、保留以及增長客戶，同時減少欺詐和降低風險方面存在競爭優勢。透過將 IBM SPSS 軟體引入其每天的作業，組織成為具有預測能力的企業，能夠直接或自動進行決策，以符合企業目標，並達成可測量的競爭優勢。如需更多資訊，或是聯絡代表人員，請造訪 http://www.ibm.com/spss。

技術支援人員

技術支援可用於維護客戶。客戶可能會聯絡技術支援，以取得使用 IBM Corp. 產品的協助，或其中一個受支援硬體環境的安裝協助。若要聯絡技術支援人員，請參閱 IBM Corp. 網站，網址為 http://www.ibm.com/support。要求協助時，請準備好識別您自己、您的組織及您的支援合約。

第 1 章 CRISP-DM 簡介

CRISP-DM 說明概觀

CRISP-DM 代表適用於資料採礦的跨業界標準程序，它是一種經業務證明可指引您進行資料採礦的方法。

- 作為**方法**，它包括專案的一般階段說明、每個階段涉及的作業以及這些作業之間的關係說明。
- 作為**程序模型**，CRISP-DM 提供資料採礦生命週期的概觀。

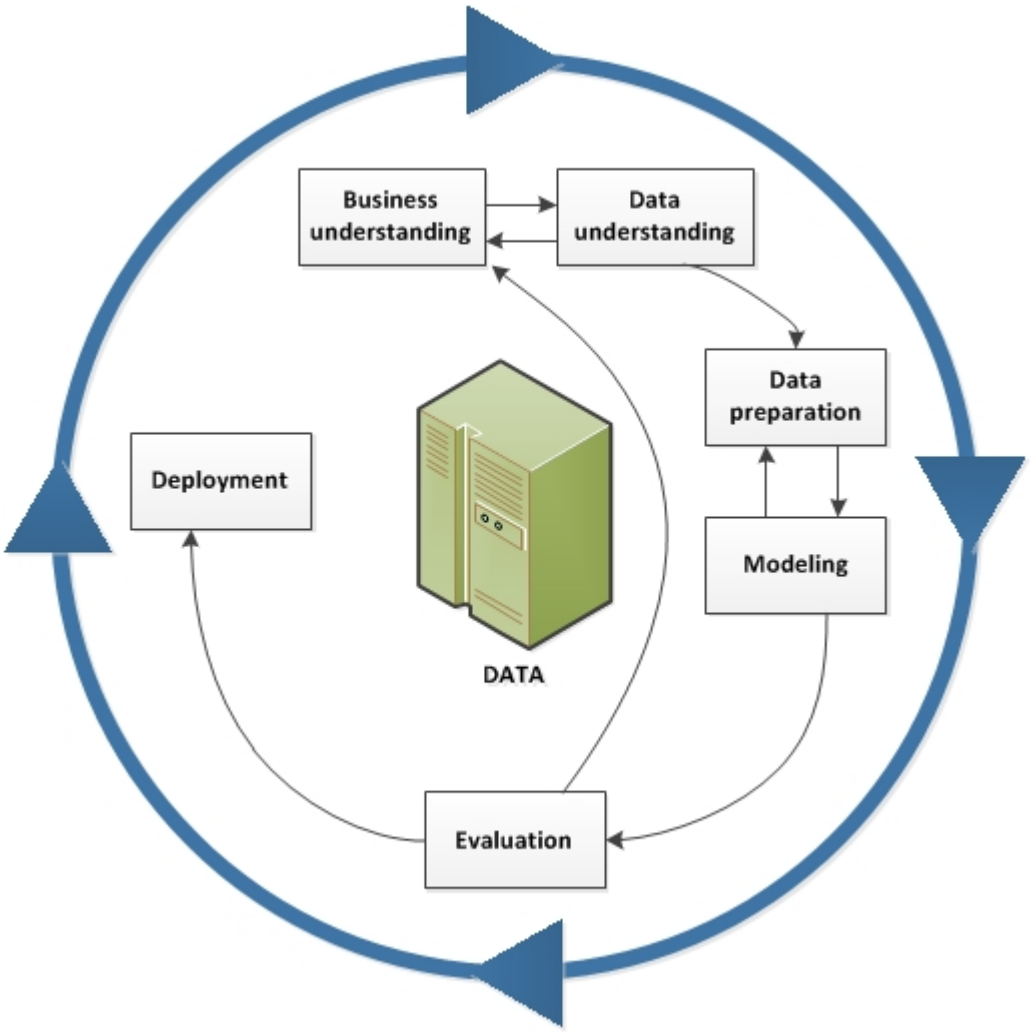

圖 *1.* 資料採礦生命週期

包含六個階段的生命週期模型，使用箭頭指出階段之間最重要以及最頻繁的相依關係。階段的順序並不嚴格。事實上，大部分專案會在階段之間來回移動（必要的話）。

CRISP-DM 模型是靈活的，可以輕鬆自訂。例如，如果您的組織旨在偵測洗錢，則很有可能您將在沒有特定建模目標的情況下篩選大量資料。您的工作將著重於資料探索和視覺化以找出財務資料中的可疑型樣，而不是建模。CRISP-DM 可讓您建立滿足您特定需要的資料採礦模型。

在這種狀況下，相較於資料理解和準備階段，建模、評估和部署階段的相關程度可能更小。但是，在長期規劃與未來資料採礦目標中將稍後的那些階段期間提出的部分問題納入考量，仍然很重要。

IBM SPSS Modeler 中的 CRISP-DM

IBM SPSS Modeler 可以兩種方式來合併 CRISP-DM 方法，以針對有效資料採礦提供唯一的支援。

- CRISP-DM 專案工具可協助您根據一般資料採礦專案的各個階段來組織專案串流、輸出以及註釋。您可在專案期間根據串流和 CRISP-DM 階段的附註隨時產生報告。
- CRISP-DM 的說明可引導您執行處理資料採礦專案的程序。說明系統包括每個步驟的作業清單以及 CRISP-DM 在真實世界中如何運作的範例。您可以從主視窗的說明功能表中選擇 **CRISP-DM 說明**，來存取 CRISP-DM 說明。

CRISP-DM 專案工具

CRISP-DM 專案工具針對資料採礦提供結構化方法，可協助確保專案成功。它在本質上是標準 IBM SPSS Modeler 專案工具的延伸。事實上，您可以在 CRISP-DM 視圖與標準的「典型」視圖之間切換，以查看依 CRISP-DM 的類型或階段組織的串流和輸出。

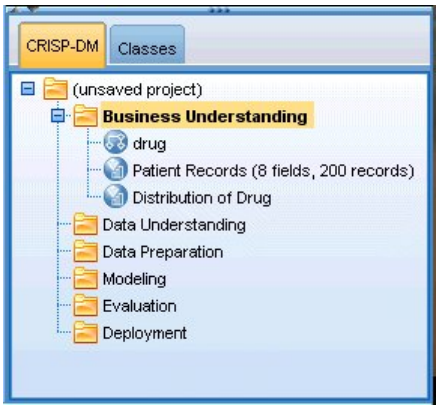

圖 2. *CRISP-DM* 專案工具

使用專案工具的 CRISP-DM 視圖，您可以執行下列作業：

- 根據資料採礦階段來組織專案的串流和輸出。
- 注意每一個階段的組織目標。
- 為每一個階段建立自訂工具提示。
- 注意根據特定圖形或模型得出的結論。
- 產生 HTML 報告或更新以配送給專案小組。

CRISP-DM 的說明

IBM SPSS Modeler 針對非專利 CRISP-DM 程序模型提供線上指引。該指引由專案階段進行組織，提供下列支援：

- CRISP-DM 的每個階段的概觀和作業清單
- 關於針對各種里程碑產生報告的說明
- 現實世界範例說明專案小組可如何使用 CRISP-DM 來照亮資料採礦的道路
- 指向 CRISP-DM 上的其他資源的鏈結

可從主視窗說明功能表中選擇 **CRISP-DM 說明**來存取 CRISP-DM 說明。

其他資源

除了對 CRISP-DM 的 IBM SPSS Modeler 支援以外，還有數種方式可擴充您對資料採礦程序的理解。

- 請閱讀 CRISP-DM 手冊，它由 CRISP-DM 聯盟建立並隨附於本版次。
- 請閱讀滿懷信心進行資料採礦，copyright 2002 by SPSS Inc., ISBN 1-56827-287-1。

第 2 章 商業理解

商業理解概觀

即使是在 IBM SPSS Modeler 中工作以前，您也應該花時間探索您的組織想要從資料採礦預期獲得哪些效益。請盡可能在這些討論中涉及足夠多的關鍵人員並記錄結果。此 CRISP-DM 階段的最終步驟討論如何使用這裡收集的資訊來產生專案計劃。

雖然此研究看起來可有可無，但事實並非如此。在擴充有價值的資源之前，對您的資料採礦工作的商業理由加以瞭解有助於確保所有人的目標一致。

確定企業目標

您的首要作業是嘗試盡可能獲取更多對企業目標的見解以用於資料採礦。這可能不像看起來這麼簡單，但您可以透過釐清問題、目標和資源來最小化稍後的風險。

CRISP-DM 方法為您提供一種結構化方式來完成這項作業。

作業清單

- 開始收集關於現行業務狀況的b背景資訊。
- 記錄特定的商業目標由關鍵決策制訂者決定。
- 對用來確定企業角度的資料採礦成功的準則達成協議。

電子零售範例--尋找企業目標

使用 CRISP-DM 的 Web 採礦實務範例

由於越來越多的公司轉換成透過網路銷售，因此建立的電腦/電子零售商面臨來自新網站的競爭也日益增加。面臨的現實是網路商店出現的速度越來越快，甚至超過客戶移轉至網路的速度，盡管收獲客戶的成本越來越高，但公司必須尋找出路來保持盈利。提出的一個解決方案是培養現有的客戶關係以最大化公司目前的每個客戶的價值。

因此，賦予研究的目標如下所示：

- 進行更佳的推薦來提升交叉銷售。
- 提供更加個人化的服務來增加客戶的忠誠度。

如果達到下列成效，則暫且認為研究成功：

- 交叉銷售量增加 10%。
- 客戶每次造訪網站花費更多的時間並且查看更多的網頁。
- 研究準時並且在預算之內完成。

編譯商業背景

瞭解您組織的業務狀況可協助您瞭解您處理事項時的依據：

- 可用資源（人員和物件）
- 問題
- 目標

您將需要對現行的業務狀況進行一些研究，以尋找可能會影響資料採礦專案結果的問題的實際答案。

作業 1--判定組織結構

- 開發組織圖表以說明公司部門、部門以及專案群組。請務必包括經理名稱和責任。
- 識別組織中的關鍵人員。
- 識別將提供財務支援及/或領域專門知識的內部贊助者。
- 判定是否有成員指導委員會並聘雇成員清單。
- 識別將受到資料採礦專案影響的業務單位。

作業 2--說明問題區域

- 識別問題區域，例如市場行銷、客戶關懷或業務開發。
- 用一般術語說明問題。
- 釐清專案的必備項目。專案背後的動機是什麼？業務已經在使用資料採礦嗎？
- 檢查商業群組中的資料採礦專案的狀態。工作已被核准嗎，或需要將資料採礦通告為商業群組的關鍵技術嗎？
- 必要的話，請準備資料採礦相關的參考簡報以呈現給您的組織。

作業 3--說明現行解決方案

- 說明目前用來解決商業問題的任何解決方案。
- 說明現行解決方案的優點與缺點。此外，解決此解決方案在組織內具有的接受層次。

定義企業目標

可在這裡具體化事項。您進行研究和會議之後，應該建構一個由專案贊助者同意的具體主要目標，以及受到結果影響的其他業務單位。此目標最終將會從「減少客戶流失」這種模糊概念轉換成將引導您的分析的特定資料採礦目標。

作業清單

請務必注意下列幾點，以便稍後併入專案計劃中。請記得保持目標切合實際。

- 說明您要使用資料採礦解決的問題。
- 盡可能詳細地指定所有商業問題。
- 確定任何其他業務需求（例如增加交叉銷售機會的同時不流失任何現有的客戶）。
- 用商業術語來指定預期的效益（例如將高價值客戶流失率減少 10%）。

企業成功準則

目標之前可能是明確的，但當您身臨其境時，您會知道嗎？先對資料採礦專案定義企業成功的本質，然後再進一步處理，這一點很重要。成功準則分成兩個種類：

- **目標。**這些準則可能很簡單，例如審核精確度有特定增加或客戶流失達到協議的降低水準。
- **主觀。**主觀準則（例如發現有效治療的叢集）更難以約束，但是您可以協議做出最終決策的人員。

作業清單

- 盡可能準確地記錄此專案的成功準則。
- 確保每個商業目標都有相關的成功準則。
- 讓仲裁者遵循主觀的成功度量。可能的話，請注意其預期目標。

評量狀況

現在，您已明確地定義目標，是時候評量您現在進行到哪裡了。這個步驟包括詢問類似以下的問題：

- 哪種類型的資料可用於分析？
- 您是否具備完成此專案所需的人員？
- 涉及的最大風險因素是什麼？
- 對於每個風險您有應急計劃嗎？

電子零售範例--評量狀況

使用 CRISP-DM 的 Web 採礦實務範例

這是電子零售商第一次嘗試進行 Web 採礦，且公司已決定諮詢資料採礦專家來協助入門。顧問所面臨的其中一個作業是存取公司用於資料採礦的資源。

人員。顯然，在管理伺服器日誌和產品與購買資料庫方面有內部的專門知識，但是在資料倉儲與資料清除以供分析方面知之甚少。因此，可能還需要諮詢資料庫專家。由於公司希冀研究結果將作為繼續進行的 Web 採礦程序的一部分，因此管理還必須考量在現行工作期間建立的任何位置是否將會是永久位置。

資料。由於這是一個知名公司，因此需要從中擷取大量網路日誌和購買資料。事實上，對於起始研究，公司會將分析限於已註冊網站的客戶。成功之後，即可擴充方案。

風險。除了聘請顧問的貨幣支出以及員工花費在研究上的時間以外，在此商業冒險中沒有大量的直接風險。但是，時間是非常重要的，因此將這個起始專案安排在單一財務季節進行。

此外，此時並沒有太多額外的現金流，因此將研究經費控制在預算之內極其重要。如果這些目標之一或有危險，則公司經理建議應該縮小專案範圍。

資源庫存

對您的資源進行精確的盤點這一作業必不可少。仔細查看硬體、資料來源以及人員問題可以節省大量時間和精力。

作業 1--研究硬體資源

- 您需要支援哪種硬體？

作業 2--識別資料來源和知識庫

- 哪些資料來源可用於資料採礦？記下資料類型和格式。
- 如何儲存資料？您有權即時存取資料倉儲或作業資料庫嗎？
- 您計劃購買外部資料（例如人口資訊）嗎？
- 是否存在任何安全問題防止您存取所需的資料？

作業 3--識別作業資源

- 您能夠聯絡業務和資料專家嗎？
- 您識別了資料庫管理者以及可能需要的其他支援人員嗎？

提出這些問題之後，請包括階段報告的聯絡人和資源的清單。

需求、假設與限制

如果您誠實評價專案的債務情況，則您的努力更有可能得到回報。盡可能明確地釐清這些問題將有助於避免將來發生問題。

作業 1--確定需求

基本需求是之前討論的企業目標，但請考量下列事項：

- 資料或專案結果存在安全和合法的限制嗎？
- 所有人都遵循專案排程需求嗎？
- 結果部署有相關需求嗎（例如，將 Web 或讀取評分發佈至資料庫）？

作業 2--釐清假設

- 存在可能會影響專案的經濟因素嗎（例如，諮詢費或競爭產品）？
- 有資料品質假設嗎？
- 專案贊助者/管理小組預期以何種方式來檢視結果？換言之，他們是想要瞭解模型本身還是只需檢視結果？

作業 3--驗證限制

- 您擁有存取資料所需的所有密碼嗎？
- 您驗證過對資料使用情況的所有合法限制項嗎？
- 專案預算中涵蓋了所有財務限制嗎？

風險與意外事故

考量專案過程中可能存在的所有風險，同樣也是一項明智之舉。風險類型包括：

- 排程（如果專案花費的時間大於預期怎麼辦？）
- 財務（如果專案贊助者遇到預算問題怎麼辦？）
- 資料（如果資料品質不佳或涵蓋面不佳怎麼辦？）
- 結果（如果起始結果比預期遜色怎麼辦？）

您考慮過各自風險之後，提出一個應急計劃來協助避免發生災難。

作業清單

- 記錄每一個可能的風險。
- 記錄每一個風險的應急計劃。

專有名詞

若要確保企業和資料採礦小組「使用相同的語言」，您應該考量編譯名詞解釋，其中包括技術術語以及需要釐清的專業術語。例如，如果企業的「流失」有特殊的唯一意義，則值得明確指出以利於整個團隊。同樣地，釐清增益圖表的使用情形也可讓小組受益。

作業清單

- 保留術語或團隊成員混淆的專門術語的清單。同時包括企業和資料採礦術語。
- 考量將清單發佈在企業內部網路或其他專案文件中。

成本/效益分析

此步驟會在最終評量期間回答**您的底線是什麼？**這個問題，將專案成本與成功的潛在效益相比較很重要。

作業清單

併入下列項目的分析預估成本：

- 使用的資料收集和任何外部資料
- 結果部署
- 營運成本

然後，考量下列項目的效益：

- 符合的主要目標
- 資料探索所產生的其他見解
- 較佳的資料理解可能帶來的效益

確定資料採礦目標

現在，企業目標清晰，是時候將其轉換成現實的資料採礦了。例如，「減少流失」的企業目標可轉換成包括以下項目的資料採礦目標：

- 根據最近購買資料來識別高價值客戶
- 使用可用的客戶來建置模型，以預測每個客戶的流失可能性
- 根據流失傾向和客戶價值向每個客戶指派一個等級。

如果符合這些資料採礦目標，則商業可利用這些目標來減少最有價值的客戶流失。

如您所見，商業和技術必須攜手運作才能達到有效的資料採礦。閱讀特定的提示以瞭解如何確定資料採礦目標。

資料採礦目標

當您處理業務以及資料分析師定義解決商業問題的技術解決方案時，請記得保持具體的事項。

作業清單

- 說明資料採礦問題的**類型**，例如叢集作業、預測或分類。
- 使用特定的時間單位來記載技術目標，例如使用三個月有效性進行預測。
- 可能的話，提供所需結果的實際數目，例如針對 80% 的現有客戶產生流失分數。

電子零售範例--資料採礦目標

使用 CRISP-DM 的 Web 採礦實務範例

在資料採礦顧問的協助下，電子零售商已經能夠將公司的企業目標轉換成資料採購項目。本季度要完成的起始研究目標是：

- 使用以前的購買的歷程相關資訊來產生鏈結「相關」項目的模型。當使用者查看某個項目說明時，會提供指向相關群組中的其他項目的鏈結（**購物籃分析**）。
- 使用網路日誌來確定不同的客戶嘗試尋找哪些項目，然後重新設計網站以強調顯示這些項目。每個不同的客戶「類型」將看到不同的網站主頁面（**性能分析**）。

- 在給定人員來自哪裡以及已在您的網站上的資訊後，使用網路日誌來嘗試預測人員接下來會去哪裡（**序列分析**）。

資料採礦成功準則

成功也必須以技術術語進行定義，以讓您的資料採礦工作步入正軌。使用之前確定的資料採礦目標來制訂成功基準。IBM SPSS Modeler 提供工具（例如「評估」節點和「分析」節點來協助您分析結果的正確性和有效性）。

作業清單

- 說明模型評量（例如正確性和效能等）的方法。
- 定義評估成功的基準。提供特定的數字。
- 盡您最大的努力定義主觀度量，並決定成功仲裁者。
- 考量成功部署模型結果是否為資料採礦成功的一部分。立即開始著手規劃部署。

產生專案計劃

此時，您已準備好產生因應資料採礦專案的計劃。您截止目前為止提出的問題以及您表述的商業和資料採礦目標將形成這個導覽圖的基準。

撰寫專案計劃

專案計劃是用於記錄所有資料採礦工作的主要文件。如果執行良好，則它可以通知與目標、資源以及風險的專案相關聯的每個人，並排定資料採礦的全部階段。您可能想要將計劃以及在這個階段過程中收集的說明文件發佈至您的企業內部網路。

作業清單

建立計劃時，請確保您已回答下列問題：

- 您討論過專案作業並提出了涉及每個人的計劃嗎？
- 所有階段或作業都包括時間估計值嗎？
- 您併入了部署結果或商業解決方案所需的工作和資源嗎？
- 計劃中強調顯示了決策點和審查要求嗎？
- 您標示了通常進行多個疊代的階段嗎（例如建模）？

專案計劃範例

研究的概觀計劃如下表所示。

表 *1.* 專案計劃概觀範例

階段	時間	資源	風險
商業理解	1 週	所有分析師	經濟變更
資料理解	3 週	所有分析師	資料問題, 技術問題
資料準備	5 週	資料採礦顧問, 部分資料庫分析師時間	資料問題, 技術問題
建模	2 週	資料採礦顧問, 部分資料庫分析師時間	技術問題, 找不到適當的模型
評估	1 週	所有分析師	經濟變更, 無法實作結果

表 *1.* 專案計劃概觀範例 *(*繼續*)*

階段	時間	資源	風險
部署	1 週	資料採礦顧問, 部分資料庫分析師時間	經濟變更, 無法實作結果

評量工具與技術

由於您已選擇使用 IBM SPSS Modeler 作為資料採礦成功的工具，您可以使用此步驟來研究哪些資料採礦技術最適合您的企業需求。IBM SPSS Modeler 為資料採礦的每一個階段提供全範圍的工具。若要決定何時使用各種技術，請諮詢線上說明的建模區段。

準備好執行下一步嗎？

在 IBM SPSS Modeler 中探索資料並開始工作之前，請務必先回答下列問題。

從公司的觀點來看：

- 您的公司希望從這個專案獲取什麼樣的收益？
- 您將如何定義我們工作是否成功完成？
- 您具備達成目標所需的預算和資源嗎？
- 您有權存取此專案所需的所有資料嗎？
- 您和您的團隊討論過與此專案相關聯的風險和意外事故嗎？
- 您的成本/收益分析結果顯示此專案是否有價值？

您回答完上述問題之後，您有將這些回答都轉換成資料採礦目標嗎？

從資料採礦的觀點來看：

- 資料採礦在協助您達成企業目標方面有何特別之處？
- 您瞭解哪些資料採礦技術可能產生最佳結果嗎？
- 您將如何得知您的結果正確或足夠有效？（我們設定了資料採礦成功的測量方法嗎？）
- 將如何部署建模結果？您考量過在專案計劃中部署嗎？
- 專案計劃包括 CRISP-DM 的所有階段嗎？
- 計劃中納入了風險和相依關係嗎？

如果您對上述問題均回答「是」，則表示您已準備好進一步查看資料。

第 3 章 資料理解

資料理解概觀

CRISP-DM 的資料理解階段包括進一步查看可用於採礦的資料。此步驟對於避免在下一個階段（即資料準備）期間發生非預期的問題很重要，通常這是專案中最長的一部分。

資料理解涉及存取資料以及探索資料，探索方式是透過可在 IBM SPSS Modeler 中使用 CRISP-DM 專案工具組織的表格和圖形進行。這可讓您確定資料的品質並說明專案說明文件中的這些步驟的結果。

收集起始資料

此時在 CRISP-DM 中，您已準備好存取資料並將資料帶入 IBM SPSS Modeler。資料來自各種來源，例如：

- **現有資料。**這包括各種資料，例如交易式資料、意見調查資料、網路日誌等。請考量現有資料是否足以符合您的需求。
- **購買的資料。**您的組織使用補充資料（例如個人背景資訊）嗎？如果未使用，請考量是否需要補充資料。
- **其他資料。**如果上述來源都不符合您的需求，您可能需要進行意見調查或開始其他追蹤以補充現有的資料儲存。

作業清單

查看 IBM SPSS Modeler 中的資料並考量下列問題：請務必注意您的發現項目。如需相關資訊，請參閱第 14 頁的『撰寫資料收集報告』主題。

- 資料庫中的哪些屬性（直欄）看上去最有價值？
- 哪些屬性看上去不相關並且可以排除？
- 資料是否足夠得出概括性結論或進行準確預測？
- 您所選的建模方法存在太多屬性嗎？
- 您正在合併資料來源嗎？如果是，在合併時是否有區域可能產生問題？
- 您有考量過在每一個資料來源中如何處理遺漏值嗎？

電子零售範例--起始資料收集

使用 CRISP-DM 的 Web 採礦實務範例

本範例中的電子零售商使用數個重要的資料來源，其中包括：

網路日誌。原始存取日誌包含客戶如何導覽網站的所有相關資訊。在資料準備過程中，將會需要移除對網路日誌中的影像檔和其他無意義項目的參照。

購買資料。當客戶提交訂單時，將儲存與該訂單相關的所有資訊。購買資料庫中的訂單必須對映至網路日誌中對應的階段作業。

產品資料庫。判定「相關」產品時，產品屬性可能很有用。該產品資訊必須對映到對應的訂單。

客戶資料庫。此資料庫包含從已註冊客戶收集的額外資訊。記錄並沒有完成，因為很多客戶未填寫問卷調查。客戶資訊必須對映至網路日誌中對應的購買和階段作業。

此時，公司沒有採購外部資料庫或花錢處理意見調查的計劃，因為其分析師正忙於管理目前具備的資料。但是，在某種情況下，分析師可能想要考量延伸部署資料採礦結果，在這種情況下，針對已取消註冊的客戶購買其他人口統計資料可能非常有用。此外，擁有人口統計資訊對於查看電子零售商的客戶基礎與平均網路購物者之間的差異程度也很有用。

撰寫資料收集報告

使用上一步收集的資料，您可以開始撰寫資料收集報告。完成之後，報告可機關報增至專案網站或配送至團隊。它也可以與後續步驟（資料說明、探索以及品質驗證）中準備的報告結合。這些報告將指引您完成資料準備階段。

說明資料

說明資料的方法有多種，但大部分說明著重於資料的數量與品質，即可用的資料有多少以及資料條件。下列內容是說明資料時要解決的部分關鍵性質。

- **資料量。**對於大部分建模技術，都存在與資料大小相關聯的權衡。大型資料集可產生更準確的模型，但它們也可能會延長處理時間。請考量是否可以使用一部分資料。對定案的報告進行備註時，請務必包括所有資料集的大小統計資料，並記得在說明資料時同時考量記錄數目與欄位（屬性）。
- **值類型。**資料可以採用各種格式，例如**數值**、**種類**（字串）或**布林** (true/false)。關注值類型可以避免稍後建模期間發生問題。
- **編碼方法。**通常，資料庫中的值代表性別或產品類型之類的性質。例如，一個資料集可能使用 *M* 和 *F* 來代表 *male* 和 *female*，而另一個可能使用數值 *1* 和 *2*。請注意資料報告中任何衝突的的方法。

掌握了這項知識，您現在便可以撰寫資料說明報告並與更多讀者分享您的發現項目。

電子零售範例--說明資料

使用 CRISP-DM 的 Web 採礦實務範例

在 Web 採礦應用程式中，需要處理很多記錄和屬性。雖然處理此資料採礦專案的電子零售商已將起始研究限制用於已在網站上註冊的客戶（大約 30,000 個），但網路日誌中仍有上百萬筆記錄。

這些資料來源中的大部分值類型都是符號值，即日期和時間、存取的網頁或來自登錄問卷調查中的多選問題的答案。這些變數中的一部分將用來建立新的數值變數，例如存取的網頁數目以及在網站上花費的時間。資料來源中一些現有的數值變數包括每個訂購產品的數目、在購買期間花費的金額，以及來自產品資料庫的產品加權與維度規格。

由於各種資料來源包含非常不同的屬性，因此資料來源的編碼方法中可能存在一些重疊情況。重疊的唯一變數是「鍵」，例如客戶 ID 和產品型號。資料來源之間的變數必須具有相同的編碼方法；否則，無法合併資料來源。重新為這些索引鍵欄位編碼以進行合併，必須準備一些額外的資料。

撰寫資料說明報告

若要繼續有效地處理資料採礦專案，請考使用下列度量值產生準確資料說明報告的價值：

資料數量

- 資料格式為何？
- 識別用來擷取資料的方法，例如，ODBC。
- 資料庫有多大（以列數和欄數表示）？

資料品質

- 資料包括與商業問題相關的性質嗎？
- 存在哪些資料類型（符號、數值等）？
- 您有比較過主要屬性的基本統計資料嗎？它對商業問題提供了什麼樣的見解？
- 您能夠設定相關屬性的優先順序嗎？如果不能，業務分析師能夠提供進一步的見解嗎？

探索資料

使用 CRISP-DM 的這個階段，搭配 IBM SPSS Modeler 中提供的表格、圖表和其他視覺化工具來探索資料。此類分析可協助解決在業務理解階段建構的資料採礦目標。它們也可以協助制定假設以及對資料準備期間進行的資料轉換作業設定形狀。

電子零售範例--探索資料

使用 CRISP-DM 的 Web 採礦實務範例

雖然 CRISP-DM 建議在這個點進行起始探索，但資料探索很困難，如果不可行，請探索我們的電子零售商已發現的原始網路日誌。通常，網路日誌資料必須先在資料準備階段進行處理，以產生讓探索有意義的資料。這會違背可以並且應該自訂程序以符合特定的資料採礦需求這一事實，這是 CRISP-DM 的底線。CRISP-DM 是週期性的，資料採礦者通常在階段之間來回移動。

雖然必須在探索之前先處理網路日誌，但可用於電子零售商的其他資料來源更適合用於探索。使用購買資料庫進行探索會顯示客戶相關摘要，例如他們花費的金額、他們每次購買了多少商品，以及他們來自哪裡。客戶資料庫的摘要將顯示對登錄問卷調查上項目回應的分佈。

探索也可用於尋找資料中的錯誤。大部分資料來源是自動產生的，而產品資料庫中的資訊是手動輸入的。所列出，產品尺寸的部分快速摘要將協助探索排版印刷上的訛誤，例如 "119-inch"（而不是 "19-inch"）顯示器。

撰寫資料探索報告

當您針對可用資料建立圖形並執行統計時，開始進行關於資料能夠如何回答技術和企業目標的假設。

作業清單

請注意併入資料探索報告中的發現項目。請務必回答下列問題：

- 您對資料進行了哪些類型的假設？
- 哪些屬性看起來有進一步分析的價值？
- 您的探索顯示了有關資料的新性質嗎？
- 這些探索變更了起始假設嗎？
- 您能夠識別特定的資料子集以供稍後使用嗎？
- 請再次查看您的資料採礦目標。此探索變更了目標嗎？

驗證資料品質

完美資料很少見。事實上，大部分資料都包含編碼錯誤、遺漏值或其他類型的不一致，從而導致有時候分析很棘手。在建模之前，先對可用的資料進行徹底的品質分析，這是避免潛在誤區的一個方法。

IBM SPSS Modeler 中的報告工具（例如資料審核、表格和其他輸出節點）可協助您尋找下列類型的問題：

- **遺漏資料**包括空白值或撰寫為非回應的值（例如 *$null$*、*?* 或 *999*）。
- **資料錯誤**通常是在輸入資料時出現的排版印刷上的訛誤。

- **測量錯誤**包括輸入正確但根據的是不正確的測量方法的資料。
- **編碼不一致內容**通常包括非標準度量單位或值不一致內容，例如，同時使用 *M* 和 *male* 表示性別。
- **meta 資料不正確**包括欄位的表面意義以及欄位名稱或定義中所指出意義之間的不符。

請務必注意此類品質問題。如需相關資訊，請參閱『撰寫資料品質報告』主題。

電子零售範例--驗證資料品質

使用 CRISP-DM 的 Web 採礦實務範例

通常是在說明和探索處理程序過程期間完成對資料品質的驗證。電子零售商遇到的部分問題包括：

遺漏資料。已知遺漏資料包括部分已註冊使用者未回答的問卷調查。如果沒有問卷調查提供的額外資訊，則這些客戶必須被後續模型排除。

資料錯誤。系統會自動產生大部分資料來源，因此不必太過於擔心這個問題。在探索處理程序期間可發現產品資料庫中有排版印刷上的訛誤。

測量錯誤。最有可能造成測量錯誤的來源是問卷調查。如果有任何項目考慮不周或用詞不當，則可能不會提供電子零售商希望取得的資訊。再者，在探索處理程序期間，請務必特別關注具備不尋常的答案分佈的項目。

撰寫資料品質報告

根據您對資料品質的探索與驗證，現在，您已備妥可以準備報告來引導您進行 CRISP-DM 的下一個階段了。如需相關資訊，請參閱第 15 頁的『驗證資料品質』主題。

作業清單

如前面所述，存在數種類型的資料品質問題。在移至下一步之前，請先考量下列品質問題並規則解決方案。在資料品質報告中記錄所有回應。

- 您識別了遺漏屬性和空白欄位嗎？如果已識別，那麼這類遺漏值背後是否有意義？
- 是否存在拼寫不一致而可能造成稍後的合併或轉換作業發生問題？
- 您探索過偏差來判定它們是否為值得進一步分析的雜訊或現象嗎？
- 您對值進行了合理性檢查嗎？注意任何明顯的衝突（例如有高收入等級的青少年）。
- 您考量過排除對您的假設毫無影響的資料嗎？
- 資料是以純文字檔儲存嗎？如果是，那麼檔案間的定界字元是否一致？每筆記錄均包含同樣數目的欄位嗎？

準備好執行下一步嗎？

準備資料以在 IBM SPSS Modeler 中建模之前，請考量下列幾點：

您對資料的理解程度為何？

- 所有資料來源都已經過明確識別和存取嗎？您知道任何問題或限制嗎？
- 您是否根據可用的資料識別出關鍵屬性？
- 這些屬性有助於您表述設想嗎？
- 您記錄了所有資料來源的大小嗎？
- 您能夠在適當時使用一部分資料嗎？
- 您為每個相關屬性計算了基本統計資料嗎？出現了有意義的資訊嗎？

- 您是否使用了探索圖形來獲取對關鍵屬性的進一步洞察？此洞察改變了您的任何設想嗎？
- 此專案的資料品質問題為何？您有解決這些問題的計劃嗎？
- 資料準備步驟是否清晰？例如，您知道要合並哪些資料來源以及要過濾或選取哪些屬性嗎？

現在，您已具備商業和資料理解，是時候使用 IBM SPSS Modeler 來準備資料進行建模了。

第 4 章 資料準備

資料準備概觀

資料準備是資料採礦最重要且通常很耗時的方面之一。事實上，預估資料準備花費的時間和精力通常佔到專案的 50-70%。致力於向先前的商業理解和資料理解階段注入足夠的能量，可以最小化開銷，但是您仍然需要花費大量精力來準備和包裝採礦資料。

視您的組織及其目標而定，資料準備一般涉及下列作業：

- 合併資料集及/或記錄
- 選取一部分範例資料
- 彙整記錄
- 衍生新屬性
- 排序建模資料
- 移除或取代空白或遺漏值
- 分割成訓練資料集和測試資料集

選取資料

根據之前的 CRISP-DM 階段中進行的起始資料收集，您已準備好開始選取與資料採礦目標相關的資料。通常，有兩種方法來選取資料：

- **選取項目（列）**包括制訂決策，例如要包括哪些帳戶、產品或客戶。
- **選取屬性或性質（直欄）**包括制訂如何使用性質的相關決策，例如交易金額或家庭收入。

電子零售範例--選取資料

使用 CRISP-DM 的 Web 採礦實務範例

在資料採礦程序的早期階段，很多電子零售商已決定選取哪些資料。

選取項目。起始研究將限制用於已在網站上註冊的客戶（大約 30,000 個），因此必須設定過濾器來排除已取消註冊之客戶的購買與網路日誌。應該建立其他過濾器來移除針對網路日誌中的影像檔和其他無意義項目的呼叫。

選取屬性。購買資料庫將包含電子零售商客戶的機密性資訊，因此請務必過濾客戶名稱、地址、電話號碼以及信用卡號碼之類的屬性。

併入或排除資料

當您決定要併入或排除的資料子集時，請務必記錄您制訂決策依據的基本原理。

考量的問題

- 給定的屬性與您的資料採礦目標相關嗎？
- 特定資料集或屬性的品質會妨礙結果有效性嗎？
- 您能夠援救此類資料嗎？
- 使用特定欄位（例如性別或種族）有任何限制嗎？

您在這裡制訂的決策與資料理解階段中表述的設想有不同嗎？如果有，請務必在專案報告中記錄您的理由。

清除資料

清除資料涉及進一步查看您選擇包括在分析中的資料中的問題。您可以使用數種方法並搭配 IBM SPSS Modeler 中的「記錄作業」和「欄位作業」節點來清除資料。

表 2. 清除資料

資料問題	可能的解決方案
遺漏資料	排除列或性質。或者，使用預估值填入空白。
資料錯誤	使用邏輯來手動探索錯誤並取代。或者，排除性質。
對不一致狀況進行編碼	根據單一編碼方法來決定，然後轉換並取代值。
遺漏 meta 資料或 meta 資料不正確	手動檢查可疑欄位並查出正確的意義。

在資料理解階段期間準備的資料品質報告包含特定於您資料的問題類型的相關詳細資料。您可以使用它作為在 IBM SPSS Modeler 中進行資料操作的起始點。

電子零售範例--清除資料

使用 CRISP-DM 的 Web 採礦實務範例

電子零售商使用資料清除程序來解決資料品質報告中記錄的問題。

遺漏資料。未完成線上問卷調查的客戶稍後可能會被某些模型排除。可能會再次詢問這些客戶填寫問卷調查，但這樣做花費的時間和資金讓電子零售商無力承擔。電子零售商能夠做的是針對回答了問卷調查的客戶與未回答問卷調查的客戶之間的購買差異進行建模。如果這兩組客戶有類似的購買習慣，則遺漏的問卷調查就不再這麼令人不安了。

資料錯誤。在這裡可以更正在探索處理程序期間發現的錯誤。雖然在大多數情況下，在客戶提交頁面給後端資料庫之前，會強制在網站上正確輸入資料。

測量錯誤。問卷調查上措辭不當的術語可能對資料品質有很大的影響。正如遺漏問卷調查一樣，這是一個很困難的問題，因為可能沒有足夠的時間和資金來收集新的取代問題的答案。對於有問題的術語，最佳解決方案可能是返回選取程序並根據進一步的分析來過濾這些術語。

撰寫資料清除報告

報告資料清除工作是追蹤資料疊代的必要工作。未來的資料採礦專案將受益於已備妥可用的工作詳細資料。

作業清單

在撰寫報告時，最好是考量下列問題：

- 資料中發生哪些類型的雜訊？
- 您使用哪些方法來移除雜訊？已順利完成哪些技術？
- 是否存在無法援救的任何案例或屬性？請務必記下由於雜訊而排除的資料。

建構新資料

您經常需要建構新資料。例如，它可能有助於建立新的直欄來標記每筆交易的延伸保固採購。這個新欄位 *purchased_warranty* 可使用 IBM SPSS Modeler 中的「設為旗標」節點輕鬆產生。

有兩種方法來建構新資料：

- 衍生屬性（直欄或性質）
- 產生記錄（列）

IBM SPSS Modeler 提供數種方法，供您使用其「記錄作業」和「欄位作業」節點來建構資料。

電子零售範例--建構資料

使用 CRISP-DM 的 Web 採礦實務範例

處理網路日誌可建立許多新的屬性。對於日誌中記錄的事件，電子零售商會想要建立時間戳記、識別訪客和階段作業，以及記錄存取頁面和事件代表的活動類型。這些變數中的一部分將用來建立更多屬性，例如某個階段作業中事件之間的時間。

在進行合併或其他資料重組之後可建立進一步屬性。例如，當每列事件網路日誌為「累積」以讓每列為一個階段作業時，將會建立記錄動作總數的新屬性、花費的總時間以及在階段作業期間進行的購買總數。當將網路日誌與客戶資料庫合併以讓每列為一個客戶時，將會建立記錄階段作業數目的新屬性、動作總數、花費的總時間以及每個客戶進行的購買總數。

建構新資料之後，電子零售商會經歷一個探索處理程序，以確保已正確執行資料建立。

衍生屬性

在 IBM SPSS Modeler 中，您可以使用下列「欄位作業」節點來衍生新屬性：

- 使用**衍生節點**來建立衍生自現有欄位的新欄位。
- 使用**設為旗標節點**來建立旗標欄位。

作業清單

- 衍生屬性時請考量建模的資料需求。建模演算法預期特定類型的資料（例如數值資料）嗎？如果是，請執行必要的轉換。
- 在建模之前需要正規化資料嗎？
- 可使用聚集、求平均值或歸納來建構遺漏屬性嗎？
- 根據您的背景知識，可從現有欄位衍生重要的事實（例如在網站花費的時間長度）嗎？

整合資料

同一組業務問題具有多個資料來源並不尋常。例如，您可能有權存取同一組客戶的抵押貸款資料以及購買的人口統計資料。如果這些資料集包含相同的唯一 ID（例如，社會保險號碼），則您可以使用這個索引鍵欄位將它們合併在 IBM SPSS Modeler 中。

有兩種基本方法來整合資料：

- **合併**資料包括合併具有類似記錄但屬性不同的兩個資料集。針對每筆記錄，使用相同的索引鍵 ID（例如客戶 ID）來合併資料。產生的資料的直欄或性質會增加。
- **附加**資料包括合併具有類似屬性但記錄不同的兩個以上資料集。資料是根據類似的欄位（例如產品名稱或合約長度）來整合。

電子零售範例--整合資料

使用 CRISP-DM 的 Web 採礦實務範例

使用多個資料來源時，電子零售商可使用數種不同的方法來整合資料：

- **將客戶與產品屬性新增至事件資料。**為了使用其他資料庫中的屬性來對網路日誌事件建模，必須正確地識別與每個事件相關聯的任何客戶 ID、產品編號以及採購單號碼，並且對應的屬性合併至已處理的網路日誌。請注意，合併檔案會在客戶或產品每一次與事件相關聯時抄寫客戶和產品資訊。
- **將採購和網路日誌資訊新增至客戶資料。**為了對客戶價值進行建模，必須從適當的資料庫挑選客戶的購買與階段作業資訊，然後進行加總並與客戶資料庫合併。這包括建立建構資料程序中所討論的新屬性。

整合資料庫之後，電子零售商會執行探索處理程序，以確保資料合併正確地執行。

整合作業

如果您未花費足夠的時間來瞭解您的資料，則整合資料會變得複雜。需要先對與資料採礦目標最相關的項目和屬性加以考量，然後再開始整合資料。

作業清單

- 使用 IBM SPSS Modeler 中的「合併」或「附加」節點來整合資料集有利於建模。
- 請考量在繼續建模之前先儲存產生的輸出。
- 合併之後，可透過**聚集**值來簡化資料。聚集表示透過彙總多筆記錄及/或表格中的資訊，來計算新值。
- 您可能還需要產生新記錄（例如數年的結合退稅的平均扣除）。

格式化資料

作為建模之前的最後一步，它有助於檢查某些技術是否需要對資料設定特殊格式或進行排序。例如，順序演算法要求預先排序資料然後再執行模型，這一情況並不罕見。即使模型可以為您執行排序，在建模之前使用「排序」節點也可以節省處理時間。

作業清單

格式化資料時，請考量下列問題：

- 您計劃使用哪些模型？
- 這些模型要求資料有特殊的格式或進行排序嗎？

如果建立變更，則 IBM SPSS Modeler 中的處理工具可協助您套用必要的資料操作。

準備好建模了嗎？

在 IBM SPSS Modeler 中建置模型之前，請務必先回答下列問題。

- 可從 IBM SPSS Modeler 存取所有資料嗎？
- 根據您的起始探索與理解，您能夠選取相關的資料子集嗎？
- 您有效清除了資料或移除了無法挽救的項目嗎？將任何決策都記錄在定案的報告中。
- 是否適當地整合多個資料集？是否存在任何應該予以記錄的合併問題？
- 您重新研究過計劃使用之建模工具的需求嗎？
- 是否存在您可以在建模之前解決的任何格式化問題？這同時包括必要的格式化問題以及可能會減少建模時間的作業。

如果您可以回答上述的問題，則表示您已準備好進行資料採礦的關鍵部分，即建模。

第 5 章 建模

建模概觀

這是您的辛苦工作開始取得成效的點。您費時準備的資料將進入 IBM SPSS Modeler 中的分析工具，且結果開始對商業理解期間公佈的業務問題帶來一些曙光。

建模通常以多次疊代來處理。通常，資料採礦者使用預設參數執行數個模型，然後精簡參數或回復至資料準備階段，進行其所選模型所需的操作。僅透過一個模型和一次執行就對組織的資料採礦問題提供滿意的回答，這種情況很少見。這一點讓資料採礦變得相當有趣 -- 您可以使用多種方式來查看給定的問題，且 IBM SPSS Modeler 提供大量工具來協助您這樣做。

選取建模技術

雖然您對於哪些類型的建模最適合您組織的需求可能略知一二，但現在應該決定確切要使用的類型。通常根據下列考量來確定最適當的模型：

- **可用於採礦的資料類型。**例如，是感興趣的種類（符號）領域嗎？
- **您的資料採礦目標。**您只是想要瞭解交易式資料儲存並發掘有趣的購買型樣嗎？還是需要產生評分，例如，指出學生貸款違約的傾向？
- **特定的建模需求。**模型需要特定資料大小或類型嗎？您是否需要具備易於呈現結果的模型？

如需 IBM SPSS Modeler 中的模型類型及其需求的相關資訊，請參閱 IBM SPSS Modeler 說明文件或線上說明。

電子零售範例--建模技術

電子零售商採用的建模技術由公司的資料採礦目標驅動：

改良的建議。簡單來說，這涉及形成採購單叢集來確定哪些是經常一起採購的產品。客戶資料甚至是造訪記錄都可以新增以豐富結果。Two-step 或 Kohonen 網路叢集作業技術適用於此類型的建模。之後，可使用 C5.0 規則集來設定叢集，以確定在客戶造訪期間的任意時刻，最適合推薦哪些項目。

改良的網站導覽。現在，電子零售商將著重於識別常用但需要使用者數次點擊滑鼠才能找到的頁面。這會導致排序演算法套用到網路日誌以便產生客戶透過網站採用的「唯一路徑」，然後在不執行動作的情況下（或在執行動作之前），特別尋找頁面造訪次數很多的階段作業。之後，在更深層次的分析中，叢集作業技術可用來識別不同類型的造訪和訪客，並且可以組織網站內容並根據類型來顯示。

選擇正確的建模技術

在 IBM SPSS Modeler 中提供了多個建模技術。通常，資料採礦者使用多種技術從數個方向來解決問題。

作業清單

決定要使用的模型時，請考量下列問題對您的選擇是否有影響：

- 模型需要將資料分割成測試集和訓練集嗎？
- 您是否具備足夠的資料可針對給定模型產生可靠的結果？
- 模型需要特定級別的資料品質嗎？您可以使用現行資料滿足此級別嗎？
- 您的資料類型適用於特定模型嗎？如果不適用，則您可以使用資料操作節點來進行必要的轉換嗎？

如需 IBM SPSS Modeler 中的模型類型及其需求的相關資訊，請參閱 IBM SPSS Modeler 說明文件或線上說明。

假設建模

當您開始縮小所選建模工具的範圍時，請注意決策制定程序。記錄任何資料假設及執行的任何資料操作以符合模型的需求。

例如，在執行之前，「邏輯迴歸」和「神經網路」節點兩者的資料類型（資料類型已知）都必須完全**實例化**。這表示您將需要將「類型」節點新增至串流並執行它來執行資料，然後再建置和執行模型。同樣，預測型模型（例如 C5.0）可在預測稀有事件的規則時透過重新平衡資料受益。進行這種類型的預測時，您通常可以透過將「平衡」節點插入串流並將更平衡的子集注入模型中，從而取得更佳的結果。

請務必記錄這些類型的決策。

建立測試設計

在實際建模的最後一步，您應該花費一點時間來重新考量將如何測試模型結果。用來產生綜合性測試設計的部分有兩個：

- 說明模型「良好度」的準則
- 定義將對其測試這些準則的資料

可以數種方式測量模型的**良好度**。對於受監督的模型（例如，C5.0 和 C&R 樹狀結構），測量良好度一般是估計特定模型的錯誤率不。對於不受監督的模型（例如 Kohonen 叢集網路），測量方式可能包括易於解譯、部署或必要的處理時間之類的準則。

請記住，建模是一個疊代程序。意思是您通常需要測試數個模型的結果，才能決定要使用和部署的模型。

撰寫測試設計

測試設計是您測試所產生模型時將採取的步驟的說明。由於建模是一個疊代過程，瞭解何時停止調整參數並嘗試其他方法或模型，這一點很重要。

作業清單

建立測試設計時，請考量下列問題：

- 將使用哪些資料來測試模型？您已將資料分割成訓練/測試集嗎？（這是建模中常用的方法。）
- 您要如何測量受監督模型（例如 C5.0）的成功？
- 您要如何測量不受監督模型（例如 Kohonen 叢集網路）的成功？
- 在嘗試另一種類型的模型之前，您希望對已調整設定的模型重新執行多少次？

電子零售範例--測試設計

使用 CRISP-DM 的 Web 採礦實務範例

評量模型時所依據的準則視考量的模型以及資料採礦目標而定：

改良的建議。在將改良的推薦項目呈現給即時客戶之前，不存在純目標方法來評量這些推薦項目。但是，電子零售商可能需要用來產生推薦項目的規則，對於企業而言足夠簡單易懂。同樣，規則應該足夠複雜，能夠為不同的客戶和階段作業產生不同的推薦項目。

改良的網站導覽。鑒于客戶可在網站上存取的頁面，電子零售商可以按照是否能夠輕鬆存取重要的頁面，來客觀地評量更新的網站設計。但是，正如推薦項目一樣，難以提前評量客戶對重組網站的適應程度。如果時間和資金足夠，進行一些可用性測試可能會更好。

建置模型

此時，您應該充分準備好建置您花費如此長時間考慮的模型。給自己時間和空間來試驗各種不同的模型，然後再做出最終結論。大部分資料採購者一般都會建置數個模型並比較結果，然後再部署或整合。

為了追蹤各種模型的進度，請務必注意用於每個模型的設定和資料。這將協助您與其他人討論結果並重新追蹤您的步驟（必要的話）。在建模程序結束時，您有三項資訊可用於資料採礦決策：

- **參數設定**包括您對產生最佳結果的參數所做的附註。
- 產生的實際**模型**。
- **模型結果的說明**，其中包括在執行模型以及探索模型結果期間發生的效能與資料問題。

電子零售範例--模型評量

使用 CRISP-DM 的 Web 採礦實務範例

改良的建議。針對各種不同層次的資料整合產生叢集作業，從只有購買資料庫開始，然後包括相關的客戶和階段作業資訊。對於每種層次的整合，會使用 Two-step 和 Kohonen 網路演算法的不同參數設定來產生叢集作業。對於這些叢集作業中的每一種，會使用不同的參數設定來產生數個 C5.0 規則集。

改良的網站導覽。「序列」建模節點用來產生客戶路徑。該演算法可讓您指定基本的支援準則，當著重於最常見客戶路徑時，這將十分有用。會嘗試使用參數的各種設定。

參數設定

大部分建模技術都有大量參數或設定，可以進行調整來控制建模程序。例如，可以透過調整樹狀結構深度、分割以及數個其他設定來控制決策樹狀結構。通常，大部分人員會先使用預設選項建置一個模型，然後在後續的階段作業期間精簡參數。

一旦您確定了產生最準確結果的參數之後，請務必儲存串流與產生的模型節點。此外，注意最佳設定可協助您決定自動化模型或使用新資料重建模型。

執行模型

在 IBM SPSS Modeler 中，執行模型是一項直接明確的作業。您將模型節點插入串流並編輯了任何參數之後，只需執行模型便能產生可檢視的結果。結果顯示在工作區右端的「產生的模型」導覽器中。您可以在模型上按一下滑鼠右鍵來瀏覽結果。對於大部分模型而言，您可以將產生的模型插入串流中以進一步評估及部署結果。模型還可以儲存在 IBM SPSS Modeler 中，方便重複使用。

模型說明

檢查模型結果時，請務必注意您的建模體驗。您可以使用節點註釋對話框或專案工具將記錄與模型本身一起儲存。

作業清單

對於每一個模型，記錄如下資訊：

- 您能夠根據此模型得出有意義的結論嗎？
- 模型顯示了新的見解或不同尋常的型樣嗎？

- 模型存在執行問題嗎？處理時間的合理程度如何？
- 模型的在處理資料品質問題方面（例如遺漏值太多）有困難嗎？
- 存在任何應該引起注意的計算不一致嗎？

評量模型

現在，您擁有一組起始模型，請進一步查看以確定哪些模型正確或足夠有效，可以成為最終模型。「最終」可能表示多種意思，例如「備妥可供部署」或「說明有趣的型樣」。參閱您之前建立的測試計劃可協助從組織的觀點進行此評量。

綜合性的模型評量

對於考量的每個模型，最好是根據測試計劃中產生的準則來建立方法評量。您可以在這裡將產生的模型新增至串流並使用評估圖表或分析節點來分析結果的有效性。您還應該考量結果是否有邏輯意義或結果是否對您的企業目標而言太過於簡單（例如，顯示購買的順序，例如，葡萄酒 > 葡萄酒 > 葡萄酒）。

建立評量之後，請根據目標（模型準確性）和主觀（易於使用或解譯結果）準則依序對模型評級。

作業清單

- 使用 IBM SPSS Modeler 中的資料採礦工具（例如評估圖表、分析節點或交叉驗證圖表），來評估模型的結果。
- 根據您對業務問題的理解來檢閱結果。請諮詢資料分析師或者對特定結果的相關性可能有深入瞭解的其他專家。
- 考量模型結果是否容易部署。您的組織要求透過 Web 部署結果或將結果傳回至資料倉儲嗎？
- 分析結果對成功準則的影響。它們在執行商業理解階段期間符合建立的目標嗎？

如果您能夠順利解決上述問題，並認為現行模型符合您的目標，則可以繼續進行更徹底地模型評估及最終部署。否則，使用您學到的知識來重新執行已調整參數設定的模型。

電子零售範例--模型評量

使用 CRISP-DM 的 Web 採礦實務範例

改良的建議。其中一個 Kohonen 網路以及一個 Two-step 叢集作業各提供合理的結果，電子零售商發現很難在兩者之間做出選擇。最後，公司希望同時使用兩者，接受兩種技術協議的推薦項目並更詳細地研究兩者的差異狀況。只需下一點功夫並套用商業知識，電子零售商即可開發出進一步的規則，來解決兩種技術之間存在的差異。

電子零售商還發現包括階段作業資訊的結果出奇地好。有證據表明推薦項目可以關聯於網站導覽。可以即時使用用來定義客戶下一步動向的規則集，以便在客戶瀏覽時直接影響網站內容。

改良的網站導覽。「序列」模型向電子零售商提供可預測特定客戶路徑的高信賴水準，產生結果建議對網站設計進行可管理的變更次數。

持續追蹤已修訂的參數

根據您在模型評量期間學習到的內容，是時候重新查看這些模型了。在這裡，您有兩個選項：

- 調整現有模型的參數。
- 選擇另一個模型來解決資料採礦問題。

在這兩種情況下，您都會回到建置模型作業並進行疊代，直到傳回成功的結果為止。重複此步驟時不必擔心。對於資料採礦者而言，數次評估以及重新執行模型直至找到符合其需求的模型，這一點極其普遍。在調整每個模型的參數之前，這是一個用於立即建置數個模型並比較結果的良好引數。

準備好執行下一步嗎？

在進入最終的模型評估之前，請考量您的起始評量是否徹底。

作業清單

- 您能夠理解模型的結果嗎？
- 從純邏輯角度而言，模型結果對您有意義嗎？是否存在明顯的不一致需要進一步探索？
- 初步看來，結果是否能夠解決組織的商業問題？
- 您使用了分析節點和提升或增益圖來比較和評估模型正確性嗎？
- 您探索了多種類型的模型並比較了結果嗎？
- 您的模型結果可以部署嗎？

如果資料建模的結果看起來正確且相關，則現在可以在進入最終部署之前進行更為徹底的評估。

第 6 章 評估

評估概觀

此時，您已完成大部分資料採礦專案。您也已在「建模」階段中根據早期定義的**資料採礦成功準則**，來確定所建置的模型在技術上是正確的並且有效。

但是，在繼續進行之前，您應該使用專案開頭所建立的**企業成功準則**來評估您的工作結果。這是確保您的組織能夠使用所取得結果的關鍵。使用資料採礦產生了兩種類型的結果：

- 在 CRISP-DM 的前一階段中選取的最終**模型**。
- 從模型本身以及資料採礦程序得出的任何結論或推斷。將它們稱之為**發現項目**。

評估結果

在這個階段中，您可以正式評量專案結果是否滿足您的企業成功準則。此步驟要求您明確理解指出的企業目標，因此，請務必在專案評量中包括關鍵決策制訂者。

作業清單

首先，您必須記錄資料採礦結果是否符合企業成功準則的評量。請在您的報告中考量下列問題：

- 您的結果是否明確闡述並且採用易於顯示的格式？
- 是否存在應該強調顯示的特定小說或唯一的發現項目？
- 您是否能夠按照模型和發現項目對企業目標的應用順序對其進行排名？
- 一般而言，這些結果在多大程度上對您組織的企業目標進行了回答？
- 您的結果提出了哪些其他問題？您如何以商業術語來定義這些問題？

評估結果之後，請編譯已核准的模型清單以併入定案的報告中。此清單應該包括同時滿足組織的資料採礦和企業目標的模型。

電子零售範例--評估結果

使用 CRISP-DM 的 Web 採礦實務範例

從企業的角度來看，電子零售商的第一次資料採礦經驗的整體結果很容易理解：研究產生的結果符合預期，即改良了產品推薦以及改進了網站設計。改良的網站基於客戶瀏覽順序，其顯示客戶所需的網站功能，但需要數個步驟才能達成。產品推薦變得更好這一跡像難以傳達，因為決策規則可能變得很複雜。若要產生定案的報告，分析師將嘗試識別規則集中一些可能更易於說明的一般趨勢。

對模型分級。由於數個起始模型似乎都對業務有意義，在該群組內的分級基於統計準則，輕鬆解譯以及多樣性。因此，模型針對不同的狀況提供不同的推薦。

新問題。研究提出的最重要的問題是，電子零售商如何找出其客戶的更多相關資訊？客戶資料庫中的資訊在構成推薦項目叢集時擔任重要的角色。特殊規則可用於向遺漏資訊的客戶推薦項目，而事實上，這些推薦相較於針對已註冊客戶進行的推薦更為普遍。

審查程序

有效方法通常包括用來反映剛完成之程序的成功與弱點的時間。資料採礦沒有什麼不同。CRISP-DM 的一部分是學習經驗，讓將來的資料採礦專案更有效率。

作業清單

首先，您應該彙總每個階段的活動和決策，其中包括資料準備步驟、模型建置等。然後，針對每個階段，考量下列問題並提出改進建議：

- 這個階段有助於最終結果值嗎？
- 是否有方法可以簡化或改進這個特定的階段或作業？
- 此階段的失敗或錯誤有哪些？下次可以如何避免這些失敗或錯誤？
- 遇到了困境嗎，例如經證明無用的特定模型？是否有方法可以預測此類困境，以讓這些努力能夠更有成效？
- 在此階段期間是否有任何驚喜（好和不好這兩方面）？回想一下，是否有明顯的方法可以預測此類狀況？
- 在給定階段中是否可能使用了替代的決策或策略？請記下此類替代方案，以供將來的資料採礦專案使用。

電子零售範例--審查報告

使用 CRISP-DM 的 Web 採礦實務範例

審查起始資料採礦專案的程序之後，電子零售商對程序中步驟之間的相互關係有了更為深刻的認識。最初，電子零售商在 CRISP-DM 程序中不願回溯，但是現在，他們發現程序的週期性可提升其功能。程序審查還導致電子零售商瞭解到：

- 當在 CRISP-DM 程序的另一階段出現不尋常事件時，有必要返回到探索處理程序。
- 資料準備（尤其是網路日誌）需要有耐心，因為它可能花費很長的時間。
- 保持關注即有的業務問題相當重要，因為一旦資料準備好進行分析，如果不考慮大局，會很容易就開始建構模型。
- 建模階段完成之後，在決定如何實作結果以及決定將來必須進行哪些研究方面，業務理解尤其重要。

決定後續步驟

到目前為止，您已產生結果並評估了資料採礦體驗，然後可能會考慮**接下來做什麼？**本階段可協助您回答根據企業目標提出的資料採礦相關問題。基本上，此時您有兩種選擇：

- **繼續前往部署階段。**下一個階段將協助您將模型結果合併到商業程序並產生定案的報告。即使您的資料採礦工作不成功，您也應該使用 CRISP-DM 的部署階段來建立定案的報告，以配送給專案贊助者。
- **回上一步並精簡或取代您的模型。**如果您發現您的結果幾乎（但並非完全）最佳，可考量進行另一輪建模。您可以利用在這個階段學到的知識來精簡模型並產生更佳的結果。

您在這個點所做的決策包括建模結果的正確性和相關性。如果結果符合您的資料採礦和企業目標，則表示您已準備好進入部署階段。無論做出什麼樣的決策，請務必徹底記錄評估程序。

電子零售範例--後續步驟

使用 CRISP-DM 的 Web 採礦實務範例

電子零售商對於專案結果的正確性與相關性相當有信心，因此繼續進入到部署階段。

與此同時，專案小組也已準備好返回並擴增部分模型以包括預測技術。此時，他們正在等待交付定案的報告以及決策制訂者的許可。

第 7 章 部署

部署概觀

部署是使用新的見解來改善組織的過程。這表示正式整合（例如實作 IBM SPSS Modeler 模型）產生的流失分數隨後會讀入資料倉儲中。或者，部署可表示您使用取自資料採礦的見解推導出組織中的變更。例如，或許您在資料中探索到警示型樣，指出年齡在 30 歲以上的客戶的行為轉變。這些結果可能未正確地整合至您的資訊系統，但在規劃和制訂市場行銷決策時無疑是有用的。

一般而言，CRISP-DM 的部署階段包括兩種類型的活動：

- 規劃和監視結果的部署
- 完成總結作業，例如產生定案的報告以及進行專案訪談

根據您的組織需求，您可能需要完成以下步驟中的一兩個。

規劃部署

雖然您可能急於想要共用資料採礦工作的成果，但請花些時間來規劃對結果進行流暢且綜合性的部署。

作業清單

- 首要步驟是彙總您的結果（包括模型和發現項目）。這有助於您確定哪些模型可整合到您的資料庫系統，以及哪些發現項目應該顯示給同事。
- 針對每個可部署的模型，建立部署及整合系統的逐步計劃。請記下任何技術詳細資料，例如適用於模型輸出的資料庫需求。例如，您的系統可能需要以 Tab 定界的格式來部署建模輸出。
- 針對每一個決定性的發現項目，建立一個計劃將此資訊散佈給策略制訂者。
- 對於這兩種類型的結果，有值得一提的替代部署計劃嗎？
- 考量將如何監視部署。例如，如何更新使用 IBM SPSS Modeler Solution Publisher 部署的模型？您如何決定模型不再適用？
- 識別任何部署問題並計劃因應意外事故。例如，決策制訂者可能需要建模結果的相關資訊，並且可能需要您提供進一步的技術詳細資料。

電子零售範例--部署規劃

使用 CRISP-DM 的 Web 採礦實務範例

順利部署電子零售商的資料採礦結果要求將正確的資訊送達正確的人員。

決策制訂者。需要將推薦項目與提議的網站變更通知決策制訂者，並向其提供這些變更有何幫助的簡短說明。假設他們接受研究結果，需要通知將實作變更的人員。

Web 開發人員。維護網站的人員必須合併網站內容的新推薦項目和組織。通知他們將來進行的研究可能會變更的內容，以便他們現在就可以著手準備。根據即時序列分析為小組準備好備用網站建構可能對以後會很有幫助。

資料庫專家。應該將如何使用資料庫的資訊以及哪些屬性可能會新增至未來專案中的資料庫，持續通知給維護客戶、購買和產品資料庫的人員。

最重要的是，專案小組需要保持與這些群組中的每個人員聯絡，以便協調結果的部署以及未來專案的規劃。

規劃監視及維護

在對建模結果進行完全部署和整合過程中，您的資料採礦工作可能在進行中。例如，如果已部署模型來預測個人資料夾購買項目的順序，則可能需要定期評估此模型，以確保它的有效性並進行連續改進。同樣地，在達到特定的保留層次之後，有可能需要調整用來增加高價值客戶的客戶保留的已部署模型。隨後可修改模型並重複使用，以保留較低層次（但在價值金字塔上仍保有盈利水平）的客戶。

作業清單

請注意下列問題並務必將其併入定案的報告中。

- 對於每一個模型或發現項目，需要追蹤哪些因素或影響（例如市場價值或週期性變動）？
- 如何測量及監視每個模型的有效性和正確性？
- 您將如何判定模型已過期？提供精確度臨界值相關的具體內容或在資料中提供預期變更等等。
- 模型到期時將發生什麼事？您可以只使用較新的資料重建模型或只是稍微進行調整嗎？或者，從普遍意義上講，變更是否足以要求新的資料採礦專案？
- 此模型過期之後能否用於類似的業務問題？在此，良好的說明文件變成評量每個資料採礦專案之商業用途的重要因素。

電子零售--監視與維護

使用 CRISP-DM 的 Web 採礦實務範例

監視的立即作業用來確定新的網路組織及改良的推薦項目實際上是否運作。亦即，使用者是否能夠採取更為直接的路徑到達他們尋找的頁面？所推薦項目的交叉銷售量增加了嗎？進行數週的監視之後，電子零售商將能夠確定研究的成功與否。

系統能夠自動處理的是併入新的已註冊使用者。當客戶註冊網站時，可將現行規則集套用至其資訊，來決定應該向其推薦哪些項目。

決定何時更新用來決定推薦項目的規則集這項作業更為棘手。更新規則集不是自動程序，原因是無論是否給定適當的叢集解決方案，建立叢集都要求人員輸入。

隨著將來的專案產生的模型日趨複雜，監視需求和監視量勢必會增加。可能的話，應該定期排定報告以供檢閱來自動化大量監視。或者，建立模型所提供的預測可能是公司想要的方向。這樣一來，需要團隊提供比第一個資料採礦專案更為複雜的專案。

產生定案的報告

撰寫定案的報告不僅能夠牽制住早期文件的尚未完結部分，它還可用於傳送您的結果。這看起來似乎直接明確，但將您的結果呈現給結果的各種利害關係人，這一點很重要。這可能同時包括技術管理者和市場行銷與管理贊助者，前者負責實作建模結果，後者將根據您的結果來制訂決策。

作業清單

首先，考量您的報告的讀者。他們是技術開發人員或專注市場的經理嗎？如果每個讀者的需求都不同，您可能需要為每個讀者建立個別的報告。無論如何，您的報告應該包括下列幾點中的大部分：

- 原始商業問題的完整說明
- 用來處理資料採礦的程序

- 專案成本
- 關於與原始專案計劃的任何偏差的附註
- 資料採礦結果（模型和發現項目）的摘要
- 提出的部署計劃概觀
- 用於進一步處理資料採礦的推薦項目，其中包括在探索和建模期間探索到的相關商機

準備最終呈現

除了專案報告以外，您也可能需要向贊助者或相關部門的團隊呈現專案發現項目。在這種情況下，您可以在報告中使用很多相同的資訊，但呈現的角度會更廣泛。可以輕鬆匯出 IBM SPSS Modeler 中的圖表和圖形，以實現這種類型的呈現。

電子零售範例--定案的報告

使用 CRISP-DM 的 Web 採礦實務範例

與原始專案計劃相差甚遠也是用於進一步處理資料採礦的相關商機。原始計劃旨在找出如何讓客戶花費更多的時間並在每次造訪時在網站上看到更多的頁面。

事實證明，擁有一個開心的客戶不僅僅是將他們保留在線上這麼簡單。每個階段作業所花費時間的次數分佈（視階段作業是否導致購買進行分割）發現，導致購買的大部分階段作業的階段作業時間介於非購買階段作業的兩個叢集的階段作業時間之間。

現在已瞭解，問題在於找出在網站上花費很長時間而不購買的這些客戶是僅僅瀏覽網站，還是因為找不到他們所需的項目。下一步是找出如何交付他們尋找的項目來鼓勵購買。

進行最終專案審查

這是 CRISP-DM 方法的最終步驟，它可讓您闡述您的最終印象，並整理資料採礦程序期間學習的課程。

作業清單

您應該對資料採礦程序中涉及的那些重要人員進行簡短的訪談。在這些訪談期間需要考量的問題包括以下內容：

- 您對專案的整體印象為何？
- 您在程序期間學習到了什麼（關於大體上的資料採礦以及可用的資料）？
- 專案的哪些部分進展順利？哪個地方出現困難？是否有資訊可協助您釐清困惑？

部署資料採礦結果之後，您還可以採訪那些受到結果影響的人員，例如客戶或業務合作夥伴。您在這裡的目標應該是確定專案是否值得做以及是否提供了開始所建立的好處。

這些訪談的結果可以與您對專案的印象一起彙總到定案的報告中，該報告應該著重於從資料儲存庫的採礦學習到的經驗。

電子零售範例--最終檢查

使用 CRISP-DM 的 Web 採礦實務範例

專案成員訪談。電子零售商發現，從研究開始到完成過程中最密切關聯的大部分專案成員，即是最熱衷於結果以及期望未來專案的人員。資料庫群組看上去謹慎樂觀；雖然他們得益於該研究的有用性，但會指出對資料庫資源的額外負擔。在研究期間隨附了顧問，但接下來隨著專案範圍的擴充，需要其他員工致力於資料庫維護。

客戶訪談。到目前為止，客戶意見回饋大多是積極正面的。網站設計變更對固定客戶產生影響，這一問題沒有充分考慮到。在數年之後，已註冊客戶對如何組織網站有著特定的預期。已註冊使用者提出的意見回饋不像未註冊客戶那樣積極正面，有一部分甚至很不喜歡這些變更。電子零售商必須持續瞭解此問題，並謹慎考量變更是否能夠帶來足夠多的客戶，以因應流失現有客戶的風險。

注意事項

本資訊係針對 IBM 在美國提供之產品與服務所開發。對於本資料，IBM 可能提供其他語言。但是，您可能需要具有該語言的產品或產品版本，才能存取該產品。

在其他國家，IBM 不見得有提供本文件所提及之各項產品、服務或功能。請洽詢當地的 IBM 業務代表，以取得當地目前提供的產品和服務之相關資訊。本文件在提及 IBM 的產品、程式或服務時，不表示或暗示只能使用 IBM 的產品、程式或服務。只要未侵犯 IBM 之智慧財產權，任何功能相當之產品、程式或服務皆可取代 IBM 之產品、程式或服務。不過，任何非 IBM 之產品、程式或服務，使用者必須自行負責作業之評估和驗證責任。

本文件所說明之主題內容，IBM 可能擁有其專利或專利申請案。提供本文件不代表提供這些專利的授權。您可以書面提出授權查詢，來函請寄到：

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785US

如果是有關雙位元組 (DBCS) 資訊的授權查詢，請洽詢所在國的 IBM 智慧財產部門，或書面提出授權查詢，來函請寄到：

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, Japan

International Business Machines Corporation 只依「現況」提供本出版品，不提供任何明示或默示之保證，其中包括且不限於不侵權、可商用性或特定目的之適用性的隱含保證。有些轄區在某些交易上並不接受明示或默示保證的免責聲明，因此此項聲明不見得適用於您。

本資訊中可能會有技術上或排版印刷上的訛誤。因此，IBM 會定期修訂；並將修訂後的內容納入新版中。IBM 隨時會改進及/或變更本出版品所提及的產品及/或程式，不另行通知。

本資訊中任何對非 IBM 網站的敘述僅供參考，IBM 對該網站並不提供任何保證。那些網站上的內容並非本 IBM 產品內容的一部分，使用那些網站時應自行承擔風險。

IBM 得以各種 IBM 認為適當的方式使用或散布 貴客戶提供的任何資訊，而無需對 貴客戶負責。

如果本程式之獲授權人為了 (i) 在個別建立的程式和其他程式（包括本程式）之間交換資訊，以及 (ii) 相互使用所交換的資訊，因而需要相關的資訊，請洽詢：

IBM Director of Licensing
IBM Corporation
North Castle Drive, MD-NC119
Armonk, NY 10504-1785US

上述資料之取得有其特殊要件，在某些情況下必須付費方得使用。

IBM 基於 IBM 客戶合約、IBM 國際程式授權合約或雙方之任何同等合約的條款，提供本文件所提及的授權程式與其所有適用的授權資料。

本文件中引用的效能資料及用戶範例僅供敘述之目的。實際效能結果可能會因特定配置及作業條件而異。

本書所提及之非 IBM 產品資訊，取自產品的供應商，或其發佈的聲明或其他公開管道。IBM 並未測試過這些產品，也無法確認這些非 IBM 產品的執行效能、相容性或任何對產品的其他主張是否完全無誤。如果您對非 IBM 產品的性能有任何的疑問，請逕向該產品的供應商查詢。

關於 IBM 未來方針或意圖的所有聲明僅代表目標或目的，得依規定未另行通知即變更或撤銷。

本資訊含有日常商業運作所用之資料和報告範例。為了盡可能描述完整，範例中涵蓋了個人、公司、品牌及產品的名稱。此等名稱皆屬虛構，凡有類似實際個人或企業者，皆屬巧合。

商標

IBM、IBM 標誌及 ibm.com 是 International Business Machines Corp. 在世界許多管轄區註冊的商標或註冊商標。其他產品及服務名稱可能是 IBM 或其他公司的商標。IBM 商標最新清單可於下列網站之「著作權與商標資訊」("Copyright and trademark information") 網頁上取得，網址如下：www.ibm.com/legal/copytrade.shtml。

Adobe、Adobe 標誌、PostScript 及 PostScript 標誌是 Adobe Systems Incorporated 在美國及（或）其他國家或地區的註冊商標或商標。

Intel、Intel 標誌、Intel Inside、Intel Inside 標誌、Intel Centrino、Intel Centrino 標誌、Celeron、Intel Xeon、Intel SpeedStep、Itanium 及 Pentium 是 Intel Corporation 或其子公司在美國及其他國家或地區的商標或註冊商標。

Linux 是 Linus Torvalds 在美國及（或）其他國家或地區的註冊商標。

Microsoft、Windows、Windows NT 及 Windows 標誌是 Microsoft Corporation 在美國及/或其他國家或地區的商標。

UNIX 是 The Open Group 在美國及其他國家或地區的註冊商標。

Java 和所有以 Java 為基礎的商標及標誌是 Oracle 及（或）其子公司的商標或註冊商標。

產品說明文件條款

這些出版品的使用許可權，係遵循下列條款而授與。

適用性

除了 IBM 網站的所有使用條款以外，還適用這些條款。

個人用途

貴客戶可以為了非商務性的私人用途而複製這些出版品，但必須保留全部的所有權聲明。如果沒有 IBM 的明文同意， 貴客戶不能散布、顯示或衍生這些出版品或其中的任何部分。

商業使用

貴客戶可以在企業內複製、散布和顯示這些出版品，但必須保留所有專利注意事項。如果沒有 IBM 的明文同意， 貴客戶不得在企業外衍生這些出版品，或複製、散布或顯示這些出版品或其中的任何部分。

權限

除了本項許可權所明確授與者之外，並未明示或暗示授與這些出版品或任何資訊、資料、軟體或其中的其他智慧財產的任何其他許可權、授權或權利。

IBM 保留在自行判定出版品的使用將損害其利益或由 IBM 判定未適當遵守上述指示時，撤銷此處所授與之許可權的權利。

除非完全符合所有適當的法律和規章，其中包括所有美國輸出法律和規章，否則， 貴客戶不能下載、輸出或再輸出本項資訊。

IBM 不提供這些出版品內容的任何保證。這些出版品是依「現狀」提供，不含任何明示或默示之保證（包括但不限於可售性、未涉侵權及符合特定效用的保證）。

索引

索引順序以中文字，英文字，及特殊符號之次序排列。

〔十三劃〕

〔十四劃〕

〔十五劃〕

〔十六劃〕

〔十八劃〕

〔二十一劃〕

C

H

M

W

Printed in Taiwan