

整合 InfoSphere Warehouse 資料採礦與 IBM Cognos 報告，第一部分： InfoSphere Warehouse 與 Cognos 整合架構概觀

有效散佈資料採礦結果

層級：中級

[Benjamin G. Leonhardi \(bleon@de.ibm.com\)](mailto:bleon@de.ibm.com), IBM 軟體工程師

[Christoph Sieb \(christoph.sieb@de.ibm.com\)](mailto:christoph.sieb@de.ibm.com), IBM 資深軟體工程師

[Michael J. Wurst 博士 \(mwurst@de.ibm.com\)](mailto:mwurst@de.ibm.com), IBM 資深軟體工程師

[IBM 技術銷售專員 John Rollins 專業工程師博士 \(rollins@us.ibm.com\)](mailto:rollins@us.ibm.com)

[Jo Ramos \(jaramos@us.ibm.com\)](mailto:jaramos@us.ibm.com), IBM 技術銷售專員

2008 年 10 月 30 日

資料採礦提供進階分析技術，可從大量資料中擷取有用的資訊。一直以來，資料採礦都是只有統計與資料分析專家才能完成的工作。另一方面，資料採礦的結果通常會與公司中的各種使用者相關。這一系列的文章涵蓋一般架構及商機，將 **IBM® InfoSphere™ Warehouse** 資料採礦與 **IBM Cognos®** 報表合併，此整合可以讓全公司的人使用資料採礦結果。這是第一篇文章，除了介紹基本整合架構，還包含簡短的技术個案研討，讓您對於如何完成整合有基本的瞭解。

序言

資料採礦可以讓專家、分析師及使用者，深入瞭解大量資料集中的型樣，並將其納入日常的商業程序中。一直以來，資料採礦都是只有統計與資料分析專家才能完成的工作。另一方面，資料採礦的結果通常會與公司中不同的使用者相關。

請參考下列案例。您收集客戶的相關資料，其中包括人口統計方面（例如年齡、職業及住處），以及過去的交易（例如銷售的商品及契約）。您的行銷部門想要針對具有相似性質的特定客戶群，開發新的商品。如何尋找這種典型的使用者群

組？資料叢集可以提供這個問題的解答，會自動根據屬性或特性，將資料集分組。然後分析師就可以檢閱這些群組，以互動的方式來加以修正，直到這些群組符合其需求。例如，行銷專家會發現有一個小型但在經濟上很重要的客戶群，他們的年紀高於平均年齡，且不使用網際網路銀行服務。根據此資訊，可以為這些客戶提供特殊服務。分析程序中的決定性步驟，是要讓使用者瞭解資料叢集步驟的結果。專業分析師通常不擅於低階資料庫程式設計。

如何將資料採礦的結果傳遞給有需要的分析師及員工？要以什麼樣的方式呈現結果，反映使用者所參與的商業程序？如何能夠達到安全性需求，讓各個使用者只看到他們能看到的內容？若要為那些問題提供令人滿意的答案，則需要改變一下觀點，從資料統計分析，改成實際使用者及其所參與的商業程序。**InfoSphere Warehouse** 是全公司資料倉儲的堅固底層，直接在資料所在的基礎 **DB2** 資料庫中，提供資料採礦功能。透過簡便的 **Eclipse-based** 工具平台，就可以使用完整的功能，且可以直接在工具中觸發部署作業。

IBM Cognos 是報表告解決方案的市場領導者之一，報表在許多公司中扮演重要的角色，因為其有助於針對不同的目標群眾，以不同的方式將相關資訊合併及視覺化。通常是將基本演算法套用至儲存在資料倉儲（例如，每個月的平均銷售量）中的資訊，而取得所呈現的結果。像上述範例中的進階分析，即不包括在報表架構的範圍與功能內。最後會使用 **InfoSphere Warehouse** 的可調式進階分析功能，以及 **IBM Cognos** 已建置的精密報告功能，將二者結合。以下各節將告訴您，**IBM InfoSphere Warehouse Data Mining** 與 **IBM Cognos** 如何能夠以彈性的方式合併，以達成這個目標。此整合不需要任何繁複的程式設計或設定；只要單純使用 **DB2 SQL**，即可完成此作業。

以下各節會簡介 **InfoSphere Warehouse** 及 **IBM Cognos** 的基本架構，接著，會提供如何將這二者整合的高階概念。最後，以保健部門為例，以簡單的步驟示範如何整合。本系列後續的文章會說明有關此整合更進階的技術及概念層面。

產品元件

IBM InfoSphere Warehouse

InfoSphere Warehouse 是以 **DB2** 為資料儲存基礎，具有資料庫分割功能 (DPF)，適用於資料倉儲資料的可調式、安全及高效儲存，並結合線上交易處理 (OLTP) 資料庫的優點與大型資料倉儲的儲存需求。**InfoSphere Warehouse** 具有各種工具，可用於倉儲管理及分析。分析元件如下：

- 資料方塊服務
- 資料採礦


- 文本分析

InfoSphere Warehouse Design Studio 是 Eclipse 型工具平台，可用來設計工作量規則、資料轉換流程，以及資料採礦和文本分析的分析流程。之後，這些流程可以部署至 InfoSphere Warehouse 管理主控台，依需要排程及執行。

InfoSphere Warehouse 資料採礦與 DB2 預儲程序與使用者定義的功能建置在一起，以提高在資料庫中執行的效能，利用 DB2 作為執行環境，您可以透過其 SQL 介面或 InfoSphere Warehouse 的採礦流程使用。

資料採礦是在大型資料集中尋找有用資訊的作業，而本文中用來作為執行範例的典型案例是保健資料。現在要收集大量保健資料，說明數年來許多不同患者的狀況是可行的。這類資料的重要用途之一，就是及早找出潛在疾病的指標。為此目的，您可以收集有關心臟患者的資料，並分析哪些因素（例如血壓或體重）與心臟病有關。必須收集並轉換此資料，才能進行資料採礦。更具體地說，就是您需要取得一個表格，其中每個病人各一列，並有數個直欄呈現該患者的相關資訊。此外，應該要有一個特殊的直欄，指出這個患者是否真的患有該疾病。可能的結構看起來會像「圖 1」：

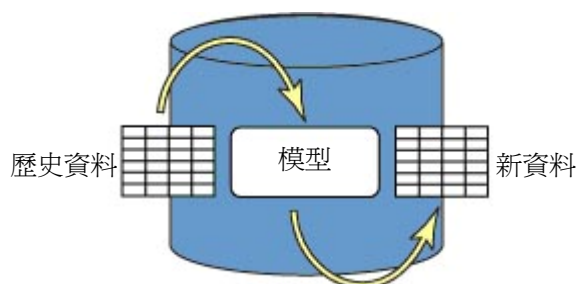
圖 1. 心臟病資料庫表格的結構（InfoSphere Warehouse 範例的一部分）

Key	Name	Data type	Length	Nullable
	RECORD_ID	INTEGER	4	No
	AGE	DOUBLE	8	Yes
	SEX	CHARACTER	1	Yes
	PAIN_TYPE	CHARACTER	1	Yes
	BLOOD_PRESSURE	DOUBLE	8	Yes
	CHOLESTEROL	DOUBLE	8	Yes
	ECG	CHARACTER	1	Yes
	HEART_RATE	DOUBLE	8	Yes
	ANGINA	CHARACTER	1	Yes
	OLD_PEAK	DOUBLE	8	Yes
	SLOPE	CHARACTER	1	Yes
	NUM_VESSELS	CHARACTER	1	Yes
	THAL	CHARACTER	1	Yes
	DISEASED	CHARACTER	1	Yes
	MEDICAL_HISTORY	VARCHAR	256	Yes
	KEYWORDS	VARCHAR	160	Yes

InfoSphere Warehouse 中的預儲程序可以依據這類資料，為任何新患者建置模型，預測他／她是否可能患有潛在的心臟病。此程序稱為「塑模」，所產生的 XML 資料採礦模型會儲存在資料庫中，並可透過 SQL/XQuery 瀏覽。其有二種用途：第一，您可以從此模型擷取資訊，以瞭解心臟病的重要指標為何，並取得其他統計資訊。第二，此模型可以自動套用至新的資料記錄。因此，您可以針對心

臟病直欄尚未包含任何值的任何記錄，自動新增此值，並以一個值來表示對此預測正確度的信心。此程序稱為「計分」，「圖 2」即顯示此結構：

圖 2. 根據歷史資料建置出的預測模型，之後可以套用至新資料



InfoSphere Warehouse 包含幾乎所有現行資料採礦演算法的高效實作，若要呼叫資料採礦，要先將資料寫入表格。再建立模型（之後可以套用至還不知道目標值為何的紀錄），產生預測及對此預測的信心值。所有資料採礦功能的呼叫方法，就如同一般 SQL 指令，如下所示，可以讓您在幾乎所有設定中輕鬆整合，例如 Web 服務。

IBM Cognos 8 Business Intelligence

IBM Cognos 8 Business Intelligence 提供一套完整的商業智慧 (BI) 功能，並以彈性的服務導向架構 (SOA) 為基礎。主要功能有報表、分析、儀表板及計分卡。

報表功能可用來收集商業程序的相關資料，將其分送給最需要該特定資訊的人員。在「資料採礦」環境中，從商業見解中受益最大的人員，會收到所產生的結果。

建立及發佈商業報告所需的 Cognos 8 元件如下：

- **Cognos Connection**：Cognos 8 功能的 Web 入口網站及進入點。使用者可以從這裡來管理、組織及檢視現有內容，例如報告或儀表板。您可以從 Cognos Connection 啟動 Web 型編寫工作區（例如 Report Studio），建立新的資源或編輯現有的資源。您也可以用它來管理 Cognos 伺服器來變更瀏覽限制、匯入內容或變更報告收件人的名單。
- **Cognos Framework Manager**：在 Cognos 8 中使用之資料來源的塑型工具。舉例來說，若要從 Cognos 報告中的大部分資料庫或其他來源（例如 Web 服務）瀏覽資料，則可在 Framework Manager 中描述 meta 資料層。資料庫表格、視圖及預儲程序都可以新增至 Cognos 套件，以作為查詢主旨。Framework Manager 可用來匯入及合併公司中不同資料來源的資訊，因此可以在像 Report Studio 這類的 Cognos 8 BI 工具中，

以一致的方式來使用資訊。有一點很重要，資料本身是保存在資料來源中，在建立報告時才加以使用。

- **Cognos Report Studio**：Cognos 8 BI 的其中一個 Web 型編寫工作區。可用來針對 Framework Manager 中描述的資料，建立及編輯精密的報告。Cognos Report Studio 提供強大的功能，例如逐層分析、提示及綜合的圖表庫。

若要在報表中使用此資料，需要執行幾個步驟：

- 在 Framework Manager 中，資料塑型器會建立 Cognos meta 資料，從商業的角度來描述資料庫中的資料（包括表格間的關係、值的商業名稱等）。
- 將 meta 資料塑型後，會將其以套件方式部署至 Cognos 8 內容儲存庫中，您可以透過 Cognos Connection 及編寫工作區使用。
- 報表設計者進入 Cognos Connection，並在所部署的套件上建立新報表。建立報表後，即可管理目標群組及配送格式（例如電子郵件或 Web 入口網站）。

從關聯式資料來源建立報表的能力，是 InfoSphere Warehouse 採礦與 IBM Cognos 整合的關鍵。

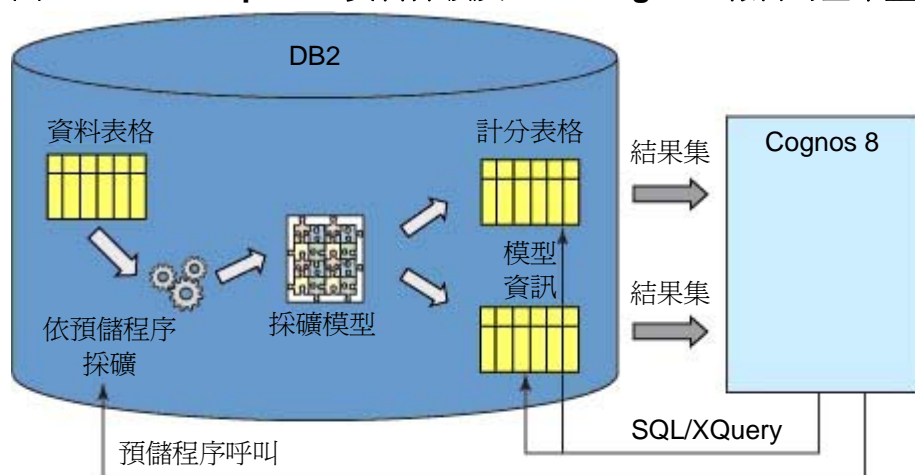
整合架構

Cognos 報表的內容包含關聯式資料來源所傳遞的結果集，如上述。特定報表的內容，則是取決於對一或多個資料來源的（動態）SQL 查詢。您可以利用這個基本通訊型樣，以下列方式來整合採礦及 Cognos 報表作業：

- Cognos 可用來顯示計分表格，可能與信心資訊一起顯示。
- Cognos 可用來顯示模型資訊，此資訊是由表格擷取程式功能或 XQuery 查詢，擷取自實際的 XML 模型。
- Cognos 可以藉由呼叫 SQL 預儲程序來動態呼叫資料採礦及計分作業，如此可進行下列作業：
 - 使用者在報表介面中提供的不同設定，呼叫採礦作業。
 - 在不同的資料子集上呼叫採礦作業（例如，建立遞迴式逐層分析報告）。
 - 依據使用者輸入，以動態方式為紀錄計分

圖 3. 彙總所有這些案例中使用的呼叫結構。

圖 3. IBM InfoSphere 資料採礦及 IBM Cognos 報告的基本整合架構



採礦作業是由預儲程序所呼叫，且會在資料庫中建立 XML 採礦模型。您可以在新資料上為此模型計分，或是將模型資訊擷取至表格中，Cognos 可以使用這些表格。使用者可以從 Cognos 報表，呼叫資料庫上相對應的預儲程序，以互動的方式來呼叫採礦作業。

此整合提供許多好處：

- 非常簡單，只需要 SQL 知識，不需要其他程式設計
- 採礦模型儲存在資料庫中，可以從 Cognos 以安全、有效率的方式使用
- 使用預儲程序，即可從 Cognos 觸發及控制整個採礦程序

下一節是這種整合的逐步式範例，也是最簡單的一種：將資料庫中的紀錄計分，並將結果顯示在 Cognos 報表中。本系列後續的文章中，會討論模型資訊的擷取，以及從 Cognos 報表進行採礦的動態呼叫。

使用 Cognos 報表中的資料採礦結果：取自保健部門的範例

此範例分析來自醫院的患者資料，醫院的心臟科有患者的主要紀錄，以及數項測量值，例如心跳速度、血壓、膽固醇等，患者會接受四種不同的心臟病檢查。患者紀錄中包含一個直欄，指出他們是否有這四種心臟病之一（y = 是，n = 否）。相對應的心臟病資料庫表格如圖 1。此表格位在 InfoSphere Warehouse 隨附的範例中。

分析目標是為新患者預測其患有這四種心臟病的風險，。如果風險很高，則應立即接受檢查。

此概念是要啓用風險管理，即使沒有針對其中一種心臟病進行專門的檢查，也可以使用其他領域前的檢查測量值。

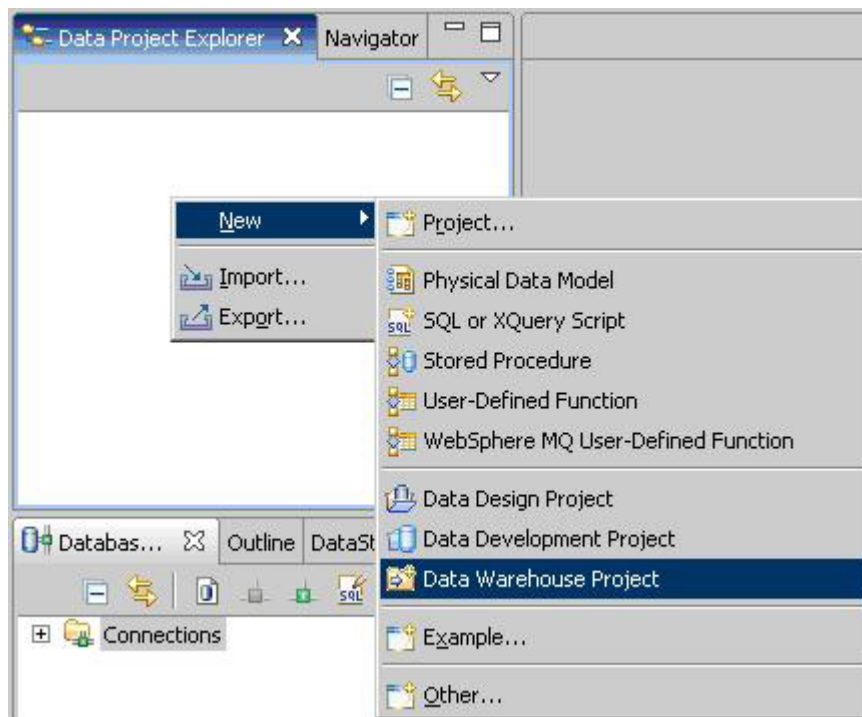
建立預測模型

首先，以 HEART 表格為基礎，建立預測模型，讓您能夠為患者預測罹患心臟病的風險。

建立「資料倉儲專案」：

用滑鼠右鍵按一下專案瀏覽器，再選取新增 > 資料倉儲專案（如下面的「圖 4」所示）。

圖 4. 建立「資料倉儲」專案



在下面的精靈中，輸入專案名稱，例如：`AdvancedAnalytics`。再按一下完成。

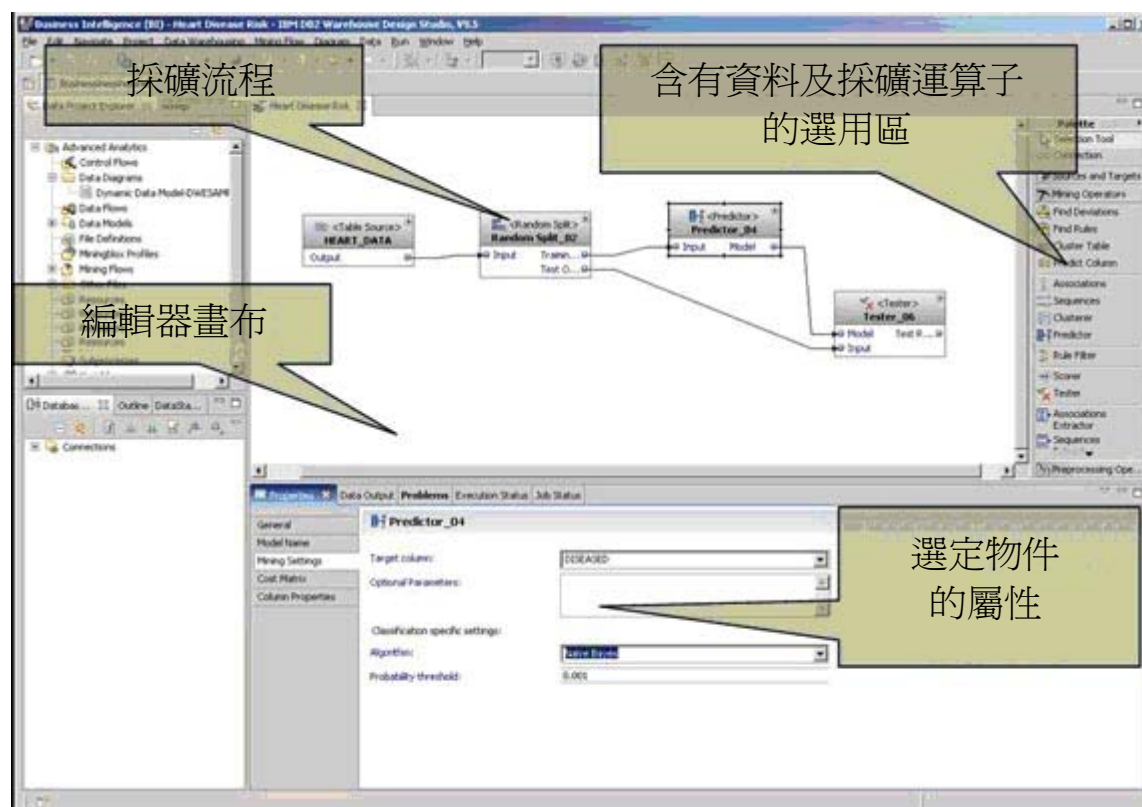
建立空的採礦流程：

1. 展開新建立的專案。
2. 用滑鼠右鍵按一下「採礦流程」資料夾，再選取新增 > 採礦流程。
3. 在即將出現的精靈中，輸入採礦流程名稱心臟病風險。
4. 在此範例中，您要對照資料庫來處理。因此，請保留預設值，並按下一步。
5. 選取資料庫，再按一下完成。

建立採礦流程：

「採礦流程」編輯器開啓。您可以在採礦編輯器右邊，看到含有運算子的選用區（請參閱「圖 5」）。您可以將這些運算子拖放至編輯器畫布上，以建置採礦流程。

圖 5. Design Studio 中的採礦流程



若要建立採礦模型，用來預測患者罹患疾病的風險，請遵循下列步驟：

1. 在選用區的「來源及目標」區段下，選取**表格來源**運算子，並將其拖放在編輯器上。
2. 在表格選項對話中，展開**綱目**，並選取 **HEART** 表格，再按一下**完成**。
3. 在選用區的「前處理運算子」區段下，選取**隨機分割**運算子，並將其拖放在編輯器上。
4. 利用簡單的拖曳動作，將第一個運算子的**輸出埠**連接至第二個運算子的**輸入埠**。
5. 選取**隨機分割**運算子。
6. 在採礦編輯器下方的「屬性」標籤上，將測試資料屬性的「百分比」設為**20**。表示我們稍後會使用 **20%** 的資料來驗證我們的模型。因此，您必須在建置預測模型前，先分割資料。
7. 在選用區的「採礦運算子」區段下，選取**預測器**運算子，並將其拖放在編輯器上。

8. 將分割運算子的**訓練輸出埠**與**預測器輸入埠**連接。
9. 選取**預測器**運算子。
10. 在採礦編輯器下方的「屬性」標籤上，選取左邊的**採礦設定**標籤。
11. 在「目標」直欄選項清單中，選取 **DISEASED** 作為您要預測的直欄。
12. Design Studio 會自動辨識出您想要預測名稱值直欄，並自動提供您可用於此目的的那些設定（在相同標籤中）。在「演算法」選項清單中，選取 **Naïve Bayes**。
13. 在「採礦設定」屬性標籤上方，選取**模型名稱**標籤。保留字首原狀，但將模型名稱變為 **HeartDiseasePrediction**。
14. 此外，也請在選用區的「採礦」區段下，選取**測試器**運算子，並將其拖放在編輯器上。
15. 將**預測器**的「模型」輸出埠與**測試器**的「模型」輸入埠連接，並將**隨機分割**的「測試」輸出埠與**測試器**的輸入埠連接。
16. 儲存採礦流程，例如，按一下編輯器區域，再按 **Ctrl+S** 鍵。

現在採礦流程已備妥，可供執行。

執行採礦流程：

您可以執行整個採礦流程編輯器，如果只要執行流程的專用路徑，請用滑鼠右鍵按一下運算子，再選取**執行此步驟...**。在此案例中，您要用滑鼠右鍵按一下**測試器**運算子，選取**執行此步驟...**，再按一下**完成**。此流程會產生可用來預測心臟病風險的模型，並將其儲存在資料庫中。其中 **80%** 的資料會受到訓練，並在剩餘的資料點上測試，如此便可讓您評估該模型對新資料的適用程度。直接用滑鼠右鍵按一下「測試器」運算子的**測試結果埠**。您也可以看一下模型本身，若要這麼做，請用滑鼠右鍵按一下「預測器」運算子的「模型」埠。

使用採礦模型，將新資料計分

計分的意思是將先前學習的模型套用至新資料。新資料沒有分類（在此案例中，沒有對心臟病進行檢查），且計分程序會根據採礦模型，為各個新紀錄指派預測結果。

建立新採礦流程，進行計分：

執行**建立新採礦流程**中的相同步驟，但您要另外命名，例如，**將新患者分類**。

建立計分流程：

若要建立計分流程，用來將新患者分類，請執行下列作業：

1. 在選用區的「來源及目標」區段下，選取**模型來源**運算子，並將其拖放在編輯器上。
2. 在採礦模型選項對話中，展開分類模型，並選取您的 **AdvancedAnalytics.HeartDiseasePrediction** 模型。
3. 在選用區的「來源及目標」區段下，將**表格來源**運算子拖放至編輯器上。
4. 在表格選項對話中，展開 **HEALTHCARE** 綱目，並選取 **HEART** 表格。此表格可以在每天晚上載入至倉儲中，或是有新患者來醫院時，以及一有必要測量值可用時，可即時載入至倉儲中。
5. 從選用區的「採礦運算子」區段下，將**計分器**運算子拖放至編輯器上。
6. 將模型及表格來源運算子連接至計分器。
7. 計分器會附加二個含有分類直欄（**是**，**否**），以及另一個指定此分類確實性的直欄。若要儲存此計分表格，您必須建立含有這些直欄的適當表格，以作為**表格目標**運算子。
若要這麼做，請用滑鼠右鍵按一下計分器的輸出埠，並選取**建立適當的表格...**。
8. 在即將出現的對話中，直接指定新表格的名稱：**NEW_PATIENT_CLASSIFICATION**，以及要在其中建立表格的綱目：**HEALTHCARE**。按一下**完成**。
9. 儲存計分流程。
10. 最後，用滑鼠右鍵按一下表格目標運算子，以執行流程，並執行至這個步驟。

執行此流程後，下面的視圖會出現產生自計分表格的範例（請參閱「圖 6」）。如果您捲動至右邊，您會看到二個已附計分器的直欄。

圖 6. 含範例結果的計分流程

計分流程：預測新患者的心臟病風險

含有二個附加直欄的範例表格：
預測的類別及預測的信賴度

Type	Name	Status	Action	Power	Columns
Run Flow	Classify New Patients	Succeeded	Run Flow	73	
Run Flow	Heart Disease Risk	Succeeded	Run Flow	73	
Run Flow	Heart Disease Risk	Failed	Run Flow	73	
Run Flow	Heart Disease Risk	Failed	Run Flow	73	

第一個附加直欄 **PREDICTED_CLASS**，包含預測結果，會說明採礦模型是否將現行患者分類為將會罹患該疾病的人（y 或 n）。第一個附加直欄 **CLASS_CONFIDENCE**，包含 0 到 1 之間的數字，根據模型說明計分器對該預測的信賴度。值 1 對應至「100 % 確定分類正確」。

在 Cognos Framework Manager 中建立 meta 資料

您必須先定義應使用的資源（資料庫、表格或視圖），才能使用採礦結果來建立報表。Framework Manager 也可讓您定義連線及新直欄（含表示式），以增加來源內容。

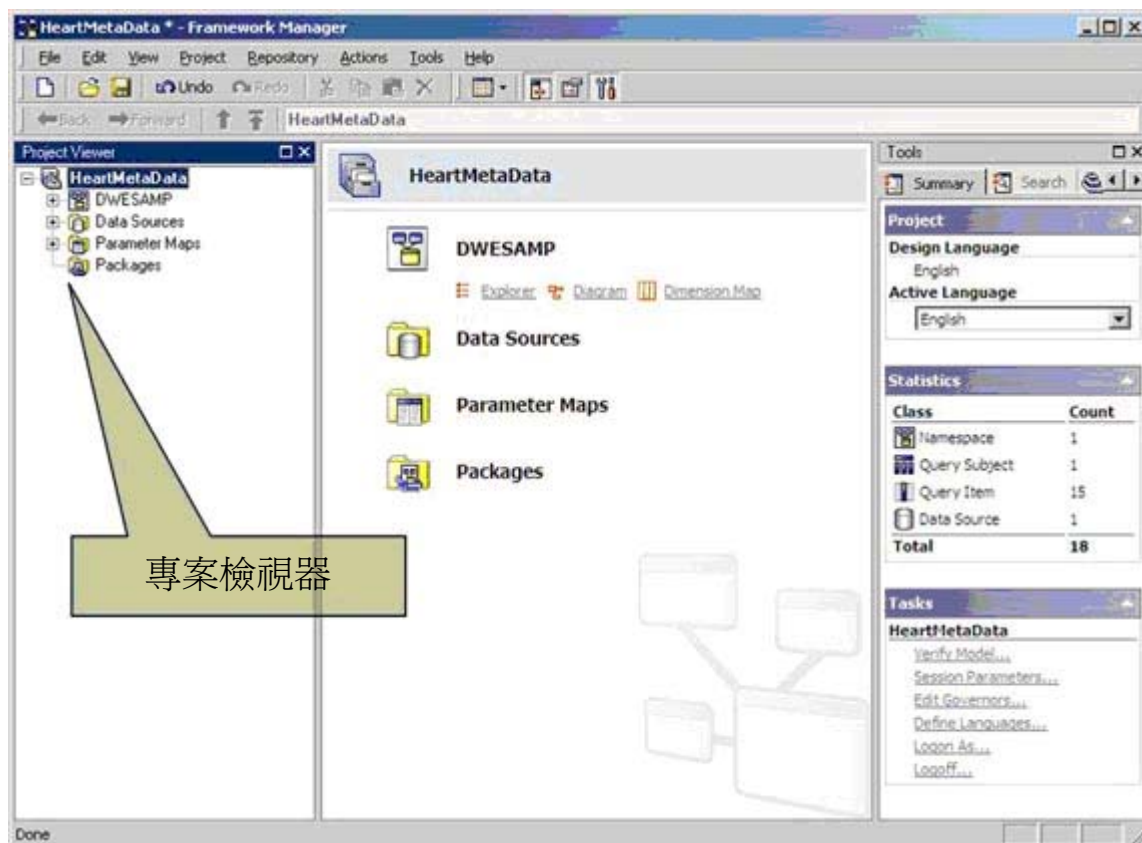
這個簡單的範例會建立 Cognos Framework Manager 專案，定義計分結果表格，並發佈 meta 資料。

開啓 Framework Manager，並建立新專案：

1. 您可以在桌面上啓動 Cognos Framework Manager。
2. 啓動完成後，按一下**建立新專案...**
3. 在即將出現的對話中，將名稱指定為 **HeartMetaData**，再按一下**確定**。
4. 語言請選取 **English**，再按一下**確定**。
5. 若您想要從 DB2 表格建立 meta 資料，請選取一般「資料來源」，再按一下**下一步**。

6. 選取 **DWESAMP** 資料庫，再按下一步。
7. 現在系統會要求您選取要匯入成 meta 資料的資料物件。展開 **HEALTHCARE** 綱目及表格資料夾，檢查您在執行計分流程期間建立的 **NEW_PATIENT_CLASSIFICATION** 表格。按 下一步。
8. 此頁面上的設定已經好了，所以請按一下匯入，再按一下完成。

圖 7. Cognos 8 Framework Manager

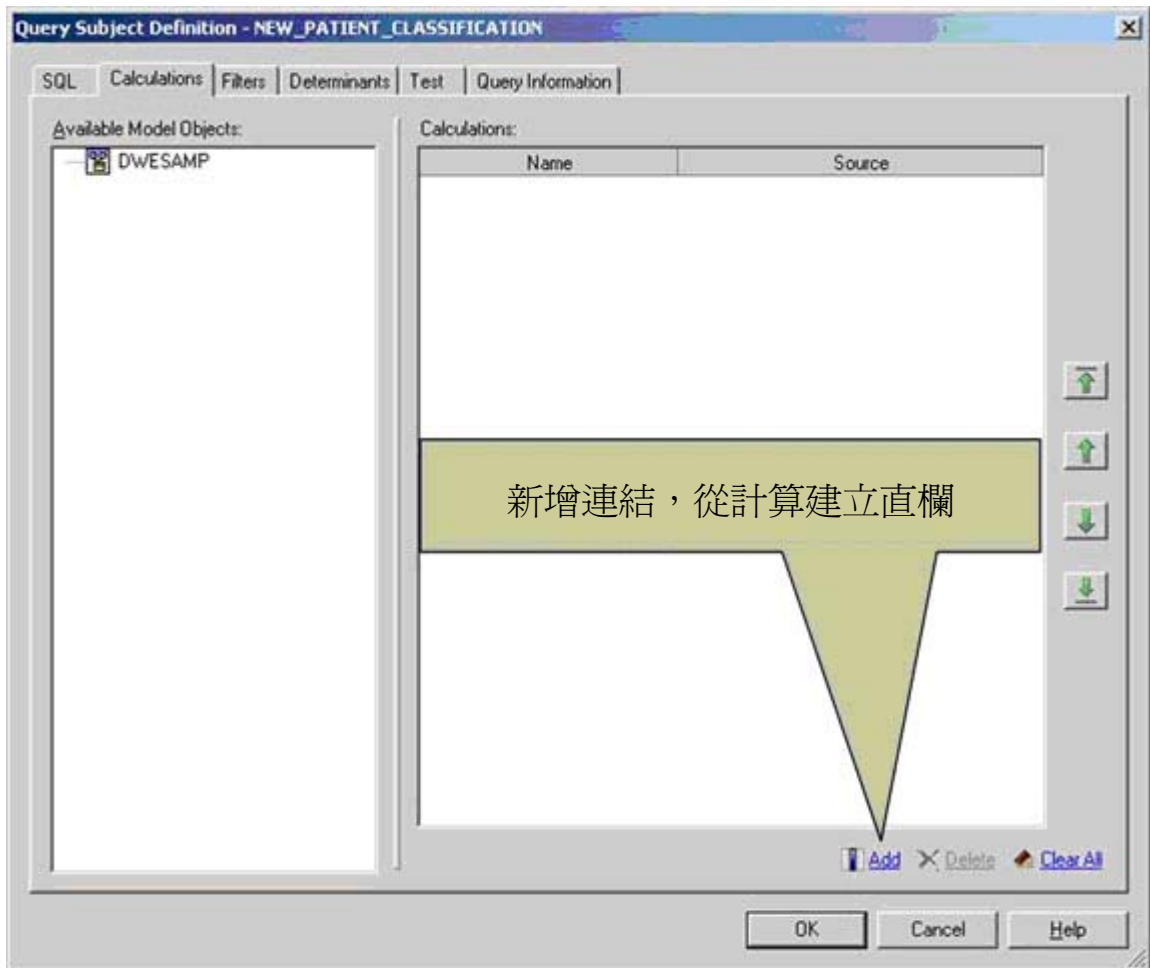


在「專案檢視器」中的左邊，您可以瀏覽新建立的專案。現在，請定義附加的直欄，將計分流程建立的二個預測直欄觀點合併。此概念是要以可理解的方式，將採礦觀點提供給醫生。

若要這麼做，請執行下列步驟：

1. 展開 **DWESAMP**。可看到計分流程中建立的表格，名為 **NEW_PATIENT_CLASSIFICATION**。
2. 按二下該表格。
3. 「查詢主旨定義」對話隨即出現，跳至「計算」標籤（請參閱圖 8）。

圖 8. 「查詢主旨定義」對話



4. 按一下**新增**，以建立從計算定義的新直欄。「計算定義」對話隨即出現。
5. 在名稱欄位中，輸入**檢查**，並在「表示式」定義文字欄位中，輸入下列表示式（您也可以桌面的 **Calculation Definition.txt** 檔案中找到這個表示式）：

```

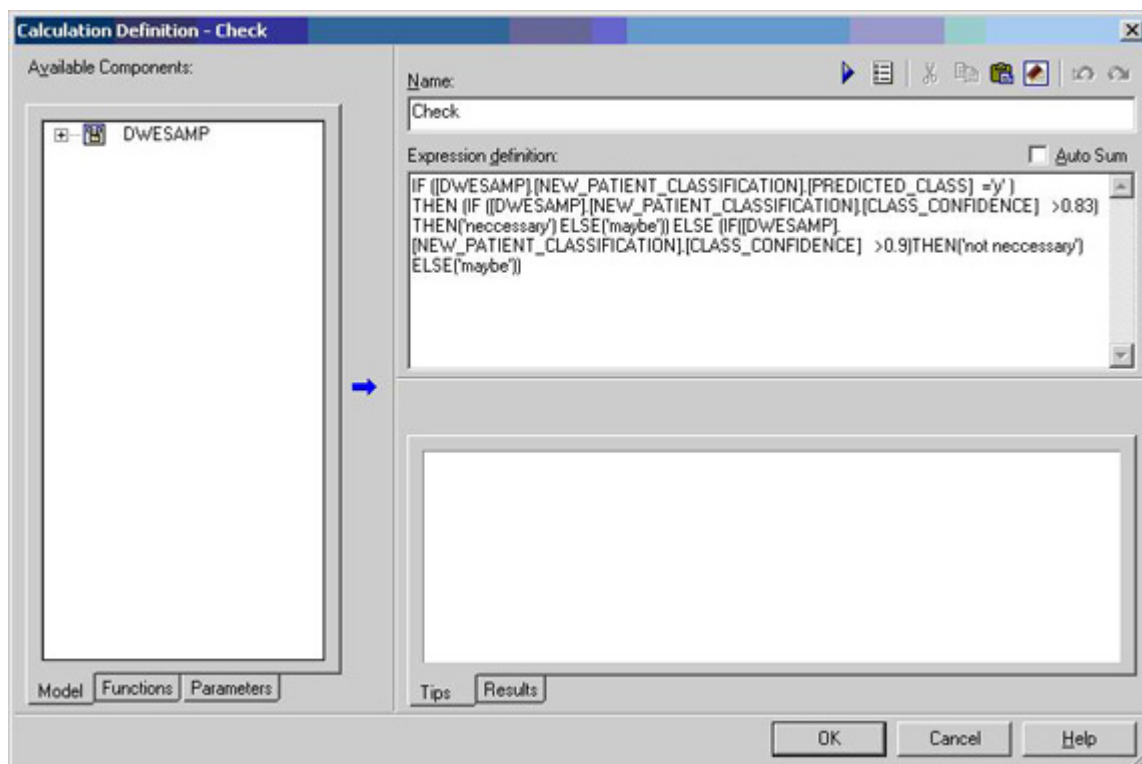
IF ([DWESAMP].[NEW_PATIENT_CLASSIFICATION].[PREDICTED_CLASS] = 'y')
THEN (IF ([DWESAMP].[NEW_PATIENT_CLASSIFICATION].[CLASS_CONFIDENCE]>0.83)
      THEN('necessary') ELSE('maybe'))
ELSE (IF([DWESAMP].[NEW_PATIENT_CLASSIFICATION].[CLASS_CONFIDENCE] > 0.9)
      THEN('not necessary')
      ELSE('maybe'))

```

上述表示式會從 **PREDICTED_CLASS** 標籤（y 及 n）及 **CLASS_CONFIDENCE**，建立三個新類別。此表示式會將預測類別為 y 的患者，且信賴度高於 83%，歸類為高風險患者（意即需要檢查）。類別為 n 的患者，且信賴度高於 90%，則會放在低風險患者群組中（意即不需要檢查）。所有其他患者，則歸類為中級風險（可能需要檢查）。圖 9 是

「計算定義」對話的範例。

圖 9. 「計算定義」對話



6. 按一下**確定**，以關閉「計算定義」對話。
7. 若要測試表示式，您可以跳至「測試」標籤，並按一下**測試範例**。
8. 按一下**確定**，以關閉「查詢主旨定義」對話。

建立套件，並加以發佈：

1. 若要讓最近建立的 **meta** 資料可供使用，請跳至「專案檢視器」，用滑鼠右鍵按一下**套件資料夾**，再選取**建立 > 套件**。
2. 在名稱欄位中，輸入**心臟**，再按下一步。
3. 按下一步，因為不需任何動作。
4. 在接下來的對話中，將 **DB2** 新增至「可用功能集清單」中，再按一下**完成**。
5. >在接下來的對話中，按一下**是**，繼續進行發佈程序。
6. 在「發佈精靈」的「選取發佈位置」部分中，按下一步，因為不需改變任何項目。
7. 不要在此範例中指定安全性設定。因此，請按下一步。
8. 再按「發佈」。
9. 在指出您已成功發佈套件的對話中，按一下**完成**。

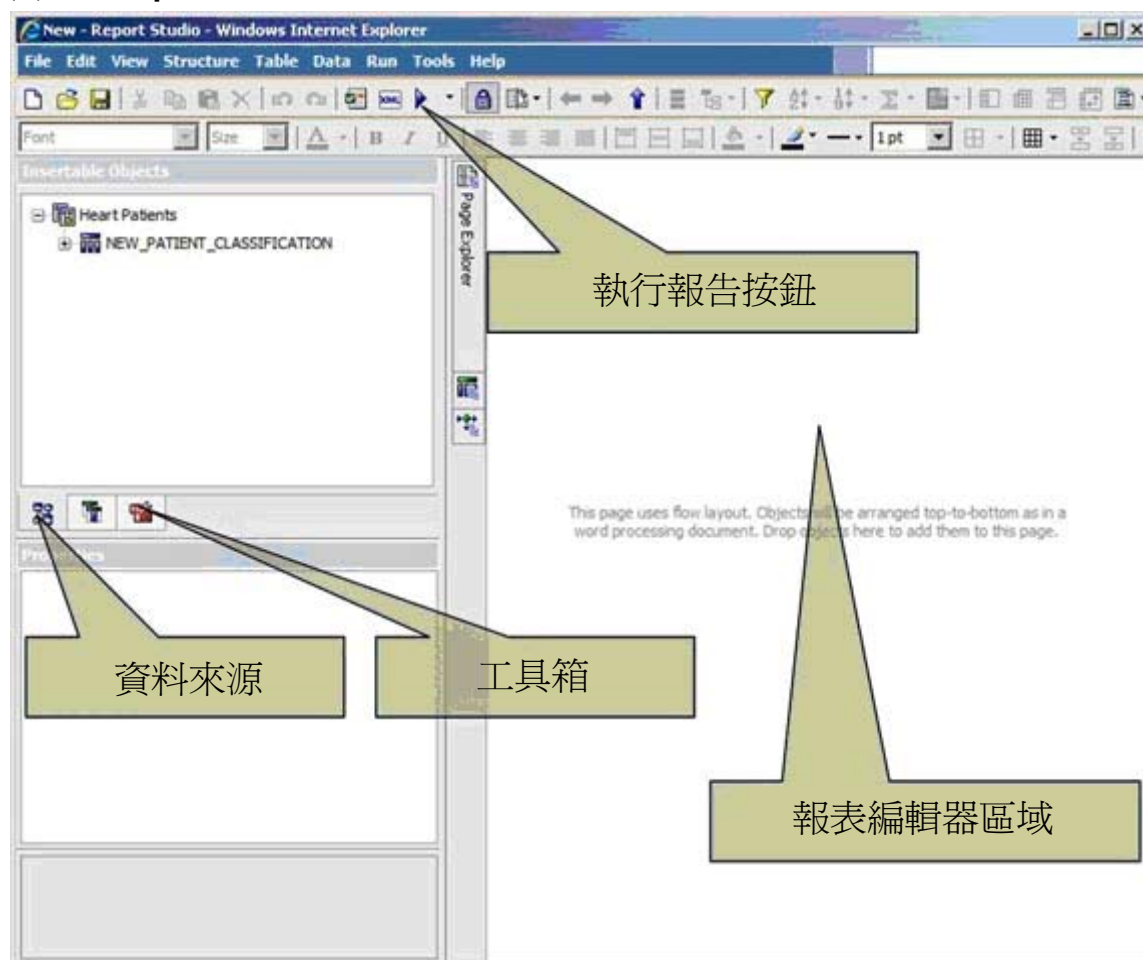
建立簡單的 Cognos 採礦報表

Cognos Report Studio 是完整的 Web 型應用程式，若要啟動 Report Studio，請按一下桌面上的 Internet Explorer 圖示 (左上角)。

您必須執行下列步驟，才能建立簡單的採礦報告：

1. 在「歡迎」畫面上，按一下**建立專業報表**。
2. 在「選取套件瀏覽器」中，按一下您在 Framework Manager 中發佈的**心臟**套件。
3. 在「歡迎」對話中，按一下**建立新報表或範本**。
4. 選擇空白範本，再按一下**確定**。
5. 在「可插入的物件」窗格中，您可以選擇不同的標籤。已選取來源（資料來源）標籤，您需要有佈置結構，才能新增資料至報表。您可以從「工具箱」標籤取得佈置結構，所以現在到「工具箱」吧。

圖 10. Report Studio



6. 從所顯示的物件清單中，將**文字項目**拖放至報表編輯器區域上(請參閱「圖 10」)。
7. 在文字方塊中，輸入下列名稱：**患者檢查建議報表**。
8. 在工具箱中尋找「清單項目」，並將其拖放至編輯器區域上。
9. 切換回「來源」標籤(資料來源)，並將 **NEW_PATIENT_CLASSIFICATION** 表格，拖曳至編輯器區域的清單中。
10. 您會看到已分類的患者表格，如果捲動至右邊，您也會看附加的直欄。由於您只需要在 Framework Manager 中，建立的附加直欄「檢查」，所以請移除 <PREDICTED_CLASS> 及 <CLASS_CONFIDENCE>。選取這二個直欄標頭(按住 **Ctrl** 鍵)，用滑鼠右鍵按一下所選取的標頭，再從快顯功能表選擇**刪除**。就目前的範例而言，您也要移除下列直欄：**<DISEASED>**、**<KEYWORDS>** 及 **<MEDICAL_HISTORY>**。
11. 現在，您想要用綠色強調顯示不需要檢查的患者(不需要)、用紅色強調顯示需要檢查的患者(需要)，其他則不強調顯示(可能)。若要這麼做，請用滑鼠右鍵按一下**檢查**直欄(附註：按一下直欄主體，而不是標頭)，再選取**樣式 > 條件樣式...**
12. 在「條件樣式」對話上，按一下左下角的圖示，以建立新的條件樣式。
13. 在下一個對話上，會選取「檢查」直欄。從條件樣式下拉清單的「類型」中，選取**字串**，再按一下**確定**。
14. 在下一個對話上，按一下左下角的圖示，並選取**選取多個值...**。Report Studio 會直接從資料庫，擷取可能的值(附註：Report Studio 可能會要求 DB2 使用者 ID 及密碼)。
15. 先將**需要**這個值新增至所選取的清單，再按一下**確定**。
16. 重複步驟 14 及 15，以新增**不需要**這個值。
17. 針對不需要這個值，從「樣式」下拉清單選取**優等**。
18. 針對不需要這個值，從「樣式」下拉清單選取**劣等**。
19. 按二次**確定**。
20. 變更標題樣式，請選取標題，再按一下「屬性」窗格上的**字型**快顯功能表。
21. 按一下**執行報表**按鈕，以檢視結果報表。

圖 11 顯示最後的報告，最右邊的直欄顯示風險分類。

圖 11. 保健心臟病風險報表

RECORD_ID	AGE	SEX	PAIN_TYPE	BLOOD_PRESSURE	CHOLESTEROL	ECG	HEART_RATE	ANGINA	OLD_PEAK	SLOPE	NUM_VESSELS	THAL	Check
120	41	f	3	112	268	2	172	y	0	1	0	3	not necessary
106	46	f	3	142	177	2	160	y	1.4	3	0	3	maybe
237	97	f	4	262	392	0	345	n	1.4	1	0	3	not necessary
75	67	f	4	106	223	0	142	n	0.3	1	2	3	not necessary
16	71	f	4	112	149	0	125	n	1.6	2	0	3	maybe
124	64	f	4	130	303	0	122	n	2	2	2	3	maybe
146	62	f	4	138	294	0	106	n	1.9	2	3	3	maybe
208	121	f	4	300	679	0	317	y	0.6	1	0	3	maybe
191	121	f	4	324	493	0	297	y	1.4	2	0	3	maybe
62	51	f	4	130	305	0	142	y	1.2	2	0	7	necessary
101	63	f	4	108	269	0	169	y	1.8	2	2	3	maybe
130	66	f	4	178	228	0	165	y	1	2	2	7	necessary
344	103	f	4	248	488	2	319	n	0	1	0	3	not necessary
15	57	f	4	128	303	2	159	n	0	1	1	3	not necessary
693	215	f	4	472	1,171	2	544	n	3.2	2	0	3	maybe
60	128	f	4	300	632	2	268	n	5	2	3	7	necessary
214	62	f	4	140	268	2	160	n	3.6	3	2	3	maybe
206	62	f	4	160	164	2	145	n	6.2	3	3	7	maybe
166	91	f	4	276	479	2	304	y	0.2	2	0	3	maybe
399	104	f	4	277	648	2	282	y	4	2	0	7	necessary

結論及展望

本文說明可輕易將 InfoSphere 資料採礦，整合至 Cognos 報表中的基礎架構。報表指出，如此的整合對於接受資料採礦有很大的影響，因為結果的使用者，絕不知道採礦程序的任何細節。取自保健部門的執行中範例顯示，如何讓部門花最少的精力，完成簡易整合。

同樣地，還是有很多可能性。後續的文章會討論更進階的主題，例如鑽探、在 Framework Manager 使用測量值，以及從 Cognos 報表動態呼叫資料採礦。

資源

學習

- [整合 InfoSphere Warehouse 資料採礦與 IBM Cognos 報表系列](#)：在此系列中尋找更多資訊。

- 「[動態倉儲：輕鬆的資料採礦](#)」：深入探討資料採礦的說明，以及 InfoSphere Warehouse 中的資料採礦演算法。
- [IBM InfoSphere Warehouse Information Center](#)：瞭解 InfoSphere Warehouse 的技術詳細資訊及參數。
- [Cognos 網站](#)：進一步瞭解 IBM Cognos。
- 逐步式範例中所使用的 HEART 資料集來源，是 [UCI 儲存庫](#)。
- [developerWorks Information Management 區域](#)：進一步瞭解 DB2。尋找技術文件、作法文章、教育訓練、下載、產品資訊等。
- 隨時掌握最新的 [developerWorks 技術活動及網路廣播](#)。

取得產品與技術

- [IBM InfoSphere Warehouse 網頁](#)：進一步瞭解 InfoSphere Warehouse。
- 利用 [IBM 試用軟體](#)（可直接從 developerWorks 下載）建置您的下一個開發專案。

討論

- 查看 [developerWorks 部落格](#)，以及參與 [developerWorks 社群](#)。

關於作者

Benjamin Leonhardi 是 InfoSphere Warehouse 資料採礦的軟體工程師，任職於德國 Boeblingen 的 IBM 研發實驗室。他專職於採礦視覺化、文本採礦及採礦報告解決方案。

Christoph Sieb 是資深軟體工程師，任職於德國 Boeblingen 的 IBM 研發實驗室。Christoph 負責資料採礦演算法，以及 InfoSphere Warehouse 資料採礦的相關架構議題。他在商業智慧及資料採礦領域，已有七年以上的開發人員及顧問經歷。

Michael Wurst 博士 是資深軟體工程師，任職於德國 Boeblingen 的 IBM 研發實驗室。他擁有電腦科學博士學位，負責資料採礦演算法及 InfoSphere Warehouse 的工具開發。Michael 在加入 IBM 前，曾擔任過 RapidMiner 資料採礦軟體的共同開發人員、設計師及顧問。

John Rollins 專業工程師博士，是 **IBM Americas Software Group** 的資料採礦技術領導者，過去 10 年來，他在 **IBM Americas Software Group** 從事的工作，包括技術銷售支援、廠區啓用、智慧資本開發及顧問。在 1996 年加入 **IBM** 前，他曾擔任過工程顧問、教授及研究員。

Jo Ramos 是商業智慧 (BI) 專家，任職於美國德州 **Dallas** 的 **BI Best Practices** 團隊。他在客戶 BI 專案方面有 8 年的經驗，負責支援資料倉儲及分析解決方案的諮詢、架構設計、資料塑模與實作。**Jo** 曾處理過大部分的 **IBM** 及 **Information Management** 產品組合，專攻分析解決方案。他擁有巴西聖塔卡林納 **FURB** 大學的經濟學士學位。