

IBM SPSS Analytic Server
第 1 版

使用手冊

IBM

附註

在使用本資訊及它支援的產品之前，請閱讀第 17 頁的『注意事項』中的資訊。

產品資訊

此版本適用於 IBM SPSS Analytic Server 1.0.0 版及所有後續版本和修訂版，直到在新版本中另有聲明為止。

目錄

第 1 章 概觀	1	資料來源	9
架構	2	設定 (檔案資料來源)	12
第 2 章 SPSS Modeler 整合	3	預覽與 meta 資料 (資料來源)	13
Analytic Server 來源	3	專案	14
Analytic Server 選取資料來源	3	注意事項	17
Analytic Server 匯出	3	商標	18
Analytic Server 串流內容	4	索引	21
支援的節點	4		
第 3 章 Analytic Server 主控台	9		

第 1 章 概觀

IBM® SPSS® Analytic Server 是用於大型資料分析的解決方案，它結合 IBM SPSS 技術與大型資料系統，可容許您使用熟悉的 IBM SPSS 使用者介面來解決規模空前的問題。

大型資料分析為何重要

組織收集的資料量以指數增長；例如，財務與零售公司具有一年（或二年，或十年）之內的所有客戶交易，telco 提供者具有呼叫資料記錄 (CDR) 及裝置感應器讀數，而網際網路公司具有 Web 搜索的結果。

存在下列項目時，需要大型資料分析：

- 大量資料 (TB、PB、EB)，特別當結構化與非結構化資料混合時
- 快速變更/累積資料

當發生下列狀況時，大型資料分析也能予以協助：

- 建置大量（數以千計）模型
- 頻繁地建置/重新整理模型

盤查

因為各種原因，收集大量資料的相同組織通常難以實際予以使用：

- 傳統分析產品的架構不適合分散式計算，以及
- 現有統計演算法未設計為處理大型資料（這些演算法預期資料送入其中，但是大型資料移動成本太高），因此
- 大型資料的圖形分析執行狀態需要大型資料系統的新技能，且對這些系統非常熟悉。極少數分析師具有這些技能。
- 記憶體內解決方案適用於中等大小問題，但是不適合真正的大型資料。

解決方案

Analytic Server 提供：

- 以資料為中心的架構，利用大型資料系統，例如 HDFS 中含資料的「Hadoop 對映」/「Hadoop 減少」。
- 已定義介面，併入設計以跳至資料的新統計演算法。
- 熟悉 IBM SPSS 使用者介面，隱藏大型資料環境的詳細資料，以便分析師可以聚焦於分析資料。
- 可調整以適合任何大小問題的解決方案。

架構

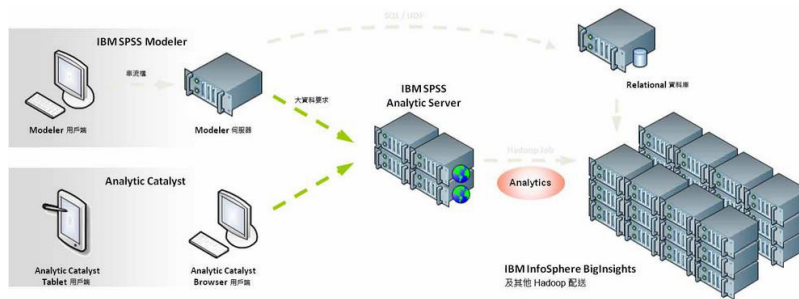


圖 1. 架構

Analytic Server 位於用戶端應用程式與 Hadoop 雲端之間。假設資料位於雲端中，使用 Analytic Server 的概要為：

1. 定義基於雲端中資料的 Analytic Server 資料來源。
2. 定義要在用戶端應用程式中執行的分析。對於現行版本，用戶端應用程式為 IBM SPSS Modeler 及 IBM SPSS Analytic Catalyst。
3. 當您執行分析時，用戶端應用程式提交 Analytic Server 執行要求。
4. Analytic Server 編排工作以在 Hadoop 雲端中執行，並將結果報告給用戶端應用程式。
5. 您可以使用結果來定義進一步分析，並重複該循環。

第 2 章 SPSS Modeler 整合

SPSS Modeler 是具有視覺化分析方法的資料採礦工作台。工作中每個獨一無二的動作，從存取資料來源、合併記錄到寫出新檔案或建置模型，在畫布上都用節點表示。我們將這些動作鏈結在一起來形成串流。

為了構建可在 HDFS 中執行的 SPSS Modeler 串流，可從『Analytic Server 來源』節點開始。SPSS Modeler 會將盡可能多的串流推送回 Analytic Server，然後從 HDFS 取回記錄的子集，以完成在 SPSS Modeler 伺服器中「以本端方式」執行串流（必要的話）。您可以在第 4 頁的『Analytic Server 串流內容』中設定 SPSS Modeler 將下載的記錄數目上限。

如果您的分析以將記錄寫回 HDFS 結束，請透過『Analytic Server 匯出』節點完成串流。

Analytic Server 來源

Analytic Server 來源可讓您在 Hadoop 分散式檔案系統 (HDFS) 上執行串流。Analytic Server 資料來源中的資訊可來自許多位置，其中包括：

- HDFS 上的文字檔
- 資料庫
- HCatalog

通常，具有 Analytic Server 來源的串流將在 HDFS 上執行；不過，如果串流包含不支援在 HDFS 上執行的節點，則會將盡可能多的串流「推送回」Analytic Server，然後 SPSS Modeler Server 將嘗試處理串流的其餘部分。您將需要對超大型資料集進行二次取樣；例如，透過將「取樣」節點放置在串流中。

資料來源。假設您的 SPSS Modeler Server 管理者已建立連線，您選取包含您要使用的資料的資料來源。資料來源包含與該來源相關聯的檔案及 meta 資料。按一下**選取**以顯示可用的資料來源。如需相關資訊，請參閱『Analytic Server 選取資料來源』。

如果您需要建立新的資料來源或編輯現有資料來源，請按一下**啟動資料來源編輯器...**。如需建立和編輯資料來源的詳細資料，請參閱第 9 頁的『資料來源』。

Analytic Server 選取資料來源

「資料來源」表格顯示可用的資料來源清單。選取您要使用的來源，並按一下**確定**。

按一下**顯示擁有者**以顯示資料來源擁有者。

過濾依據可讓您過濾在**關鍵字**上的資料來源清單；該功能會針對資料來源名稱和資料來源說明或**擁有者**檢查過濾準則。您可以輸入字串、數字或萬用字元 (%) 作為過濾準則。搜尋字串區分大小寫。按一下**重新整理**以更新「資料來源」表格。

Analytic Server 匯出

Analytic Server 匯出可讓您將資料從分析寫入現有的 Analytic Server 資料來源。例如，這可以是 Hadoop 分散式檔案系統 (HDFS) 或資料庫上的文字檔。

具有 Analytic Server 匯出節點的串流通常也以 Analytic Server 來源節點開頭，並提交給 Analytic Server 並在 HDFS 上執行。另外，具有「本端」資料來源的串流可以結束於 Analytic Server 匯出節點，以便上傳相對小的資料集（不超過 100,000 個記錄）供與 Analytic Server 搭配使用。

資料來源。 選取包含您要使用的資料的資料來源。資料來源包含與該來源相關聯的檔案及 meta 資料。按一下 **選取** 以顯示可用的資料來源。如需相關資訊，請參閱第 3 頁的『Analytic Server 選取資料來源』。

如果您需要建立新的資料來源或編輯現有資料來源，請按一下 **啟動資料來源編輯器...**。如需建立和編輯資料來源的詳細資料，請參閱第 9 頁的『資料來源』。

模式。 選取 **附加** 以新增至現有資料來源，或選取 **置換** 以取代資料來源的內容。

產生此資料的匯入節點。 選取以在將資料匯出到指定的資料來源時，產生資料的來源節點。此節點將新增至串流畫布。

Analytic Server 串流內容

這些設定提供使用 Analytic Server 的許多選項。

在 Analytic Server 外部可處理的記錄數目上限

指定要從 Analytic Server 資料來源匯入到 SPSS Modeler 伺服器中的記錄數目上限。

無法在 Analytic Server 中處理節點時的通知

此設定用來斷定當要確定至 Analytic Server 之串流包含的節點，無法在 Analytic Server 中處理時，會出現什麼情況。指定是發出警告並繼續處理串流，還是擲出錯誤並停止處理。

分割模型儲存設定

當模型大小 (MB) 超出時，透過參照將分割模型儲存在 Analytic Server

模型塊通常儲存為串流的一部分。具有許多分割塊的分割模型可產生許多塊，在串流與 Analytic Server 之間來回移動這些塊可能會影響效能。作為解決方案，當分割模型超過指定的大小時，系統會將它儲存在 Analytic Server 上，並且 SPSS Modeler 中的塊包含對該模型的參照。

執行完成後要透過參照儲存在 Analytic Server 上的預設資料夾

指定 Analytic Server 上要儲存分割模型的預設路徑。路徑應該以有效的 Analytic Server 專案名稱開頭。

用於儲存已提升模型的資料夾

指定要用於儲存「已提升」模型的預設路徑。當 SPSS Modeler 階段作業結束時，不會清除已提升模型。

支援的節點

對於在 HDFS 上執行，支援許多 SPSS Modeler 節點，但是在某些節點中的執行可能有一些差異，並且目前不支援有些節點。本主題詳細說明現行支援層次。

概要

- Analytic Server 不接受在帶引號的 Modeler 欄位中通常可接受的某些字元。
- 為了讓 Modeler 串流在 Analytic Server 中執行，它必須以一或多個 Analytic Server 來源節點開始，以一個建模節點或 Analytic Server 匯出節點結束；允許合併，但是不允許分出。
- 建議您將連續目標儲存設定為實數而不是整數。對於連續目標，評分模型一律將實際值寫入輸出資料檔案，而評分的輸出資料模型追蹤目標的儲存。因此，如果連續目標具有整數儲存，則寫入的值將與評分的資料模型不符，而此不符情況將導致在嘗試讀取已評分資料時發生錯誤。

來源

- 以 Analytic Server 來源節點以外的任何節點開頭的串流將在本端執行。

記錄作業

支援所有記錄作業。後面說明這些節點功能的進一步注意事項。

選取

- 支援「衍生」節點所支援的相同功能集（請參閱下面的內容）。

範例

- 不支援區塊層取樣。
- 不支援複式取樣方法。

聚集

- 不支援連續的索引鍵。
- 不支援順序統計資料（中位數、第一四分位數、第三四分位數）。

排序

- 不支援「最佳化」標籤。

合併

- 不支援依序合併。
- 不支援依條件合併。
- 不支援「最佳化」標籤。
- 目前不支援將「範例」節點或模型塊放置在 Analytic Server 「來源」節點與「合併」節點之間。通常可以指定「選取」節點來取代「範例」節點的功能。
- Analytic Server 不會結合空字串索引鍵；亦即，如果您要用來合併的索引鍵包含空字串，則會從合併輸出種刪除包含空字串的所有記錄。
- 合併作業相對較慢。如果您在 HDFS 中有可用空間，則與合併每個串流中的資料來源相比，合併資料來源並使用下列串流中的已合併來源的速度將比較快。

欄位作業

支援自動資料準備、輸入、過濾、衍生、整合、填充值、重新分類、進倉、RFM 分析、分割區、SetToFlag、重組和欄位重新排序節點。後面說明這些節點功能的進一步注意事項。

自動資料準備

- 不支援訓練節點。支援件將已訓練「自動資料準備」節點中的轉換套用至新資料。

輸入

- 不支援「檢查」直欄。
- 不支援「格式」標籤。

衍生

- 支援所有衍生功能，順序功能除外。
- 分割欄位不能在使用它們作為分割塊的相同串流種衍生；一個衍生分割欄位，另一個使用該欄位做為分割塊。
- 在進行比較時，旗標欄位不能使用它本身；即，`if (flagField) then ... endif` 將導致錯誤；暫行解決方法是使用 `if (flagField=trueValue) then ... endif endif` 將導致錯誤；暫行解決方法是使用 `(flagField=trueValue) then ... endif`

- 當使用 ** 運算子來指定指數做為實數（例如 $x^{2.0}$ 而不是 x^{**2} ）以與 Modeler 中的結果比對時，建議使用此項

填充值

- 支援「衍生」節點所支援的相同功能集（請參閱上面的內容）。

RFM 分析

- 不支援用於處理連結空間的「保持現行」選項。RFM 最近、頻率和貨幣評分一律不會 Modeler 與從相同資料計算所得值相符。評分範圍將相同，但是評分指派（進倉數目）可能相差一。

圖形 支援所有「圖形」節點。

建模 只有有限的「建模」節點數目，包括：線性、類神經網路、C&RT、Chaid 和搜尋。後面說明這些節點功能的進一步注意事項。

線性

- 不支援現有 PSM 模型的後續訓練。
- 僅當分割欄位的定義方式使得每個分割中僅含數目不太大的記錄時（其中「太大」的定義與 Hadoop 叢集中個別節點的能力相關），才建議使用標準模型建置目標。相比之下，您還需要小心謹慎，確保分割塊沒有定義的過細，否則記錄太少，無法建置模型。
- 不支援「提高」目標。
- 不支援「裝袋」目標。
- 當記錄較少時，建議不要使用超大資料集目標；它通常既不會建置模型，又不會建置欠佳模型。如果採用會違反所用組合模型建置演算法背後隨機假設的某種系統方法對輸入記錄進行排序，您可能也會遇到問題。
- 不支援自動資料準備。嘗試針對具有許多遺漏值的資料建置模型時，這可能會導致問題；這些問題通常部分歸咎於自動資料準備。暫行解決方法是透過「進階」設定使用樹狀結構模型或類神經網路來推導選取得遺漏值。
- 不會計算分割模型的精確度統計資料。

類神經網路

- 不支援現有標準或 PSM 模型的後續訓練。
- 僅當分割欄位的定義方式使得每個分割中僅含不多的記錄數目時（其中「太大」的定義與 Hadoop 叢集中個別節點的能力相關），才建議使用標準模型建置目標。相比之下，您還需要小心謹慎，確保分割塊沒有定義的過細，否則記錄太少，無法建置模型。
- 不支援「提高」目標。
- 不支援「裝袋」目標。
- 當記錄較少時，建議不要使用超大資料集目標；它通常既不會建置模型，又不會建置欠佳模型。如果採用會違反所用組合模型建置演算法背後隨機假設的某種系統方法對輸入記錄進行排序，您可能也會遇到問題。
- 當資料中有許多遺漏的值時，使用「進階」設定來推導遺漏的值。
- 不會計算分割模型的精確度統計資料。

C&R 樹狀結構、CHAID 和要求

- 不支援現有 PSM 模型的後續訓練。
- 僅當分割欄位的定義方式使得每個分割中僅含不多的記錄數目時（其中「太大」的定義與 Hadoop 叢集中個別節點的能力相關），才建議使用標準模型建置目標。相比之下，您還需要小心謹慎，確保分割塊沒有定義的過細，否則記錄太少，無法建置模型。

- 不支援「提高」目標。
- 不支援「裝袋」目標。
- 當記錄較少時，建議不要使用超大資料集目標；它通常既不會建置模型，又不會建置欠佳模型。如果採用會違反所用組合模型建置演算法背後隨機假設的某種系統方法對輸入記錄進行排序，您可能也會遇到問題。
- 不支援互動式階段作業。
- 不會計算分割模型的精確度統計資料。

模型評分

對於評分，支援下列模型塊：線性、類神經網絡、C&RT、Chaid、搜尋、邏輯、回歸、Genlin、TwoStep 叢集、C5、貝葉斯網路、SVM、R 和文字採礦。

- 不會對原始或調整的傾向評分。作為暫行解決方法，透過將「衍生節點」與下列表示式搭配使用來手動計算原始傾向，可以取得相同的效果：`if 'predicted-value' == 'value-of-interest' then 'prob-of-that-value' else 1-'prob-of-that-value' endif`
- 對模型進行評分時，Analytic Server 不會檢查以瞭解是否模型中的所有欄位都存在於資料集中，因此，在執行 Analytic Server 之前，請確保模型中的所有欄位都存在於資料集中

輸出 支援「矩陣」、「分析」、「資料審核」、「轉換」、「統計資料」和「方法」節點。

匯出 串流可以從 Analytic Server 來源節點開始，以 Analytic Server 匯出節點以外的匯出節點結束，但是資料將從 HDFS 移至 SPSS Modeler Server，最後移至匯出位置。

第 3 章 Analytic Server 主控台

Analytic Server 提供一個小型用戶端介面，用於管理資料來源和專案；存取該介面的方法有兩種：

- 透過 Analytic Server 來源或匯出節點，這會在瀏覽器中開啓主控台
- 直接透過瀏覽器
 1. 在瀏覽器的位址列中輸入 Analytic Server 的 URL。這可向伺服器管理者索要。
 2. 輸入用來登入伺服器的使用者名稱。
 3. 輸入與指定使用者名稱相關聯的密碼。

登入之後，預設 Accordion 是「資料來源 Accordion」。

導覽主控台

Analytic Server 主控台含有四個組成部分：

- 標頭顯示產品名稱和目前登入之使用者的名稱。
- 左側直欄顯示可用的 Accordion 或功能分組。所選取 Accordion 決定內容區中顯示的內容。
- 內容區。顯示與目前選定 Accordion 相關聯的控制項。每一個 Accordion 內容的詳細資料在下面各小節中顯示。
- 標底包含登出鏈結，並顯示 Analytic Server 的已安裝版本。

資料來源

資料來源是記錄集合加上資料模型（定義資料集以供分析）。記錄來源可以是 HDFS 上的檔案（定界文字、固定寬度文字、Excel）、資料庫或 HCatalog。資料模型會定義分析資料所需要的所有 meta 資料（欄位名稱、儲存體、測量層次等）。資料來源擁有人可以授與或限制對資料來源的存取權。

左側直欄

左側直欄在 Accordion 標題下顯示現有資料來源。

- 選取資料來源以在內容區中顯示其詳細資料，並編輯其內容。在搜尋區中輸入內容可過濾清單，以僅顯示名稱中含有搜尋字串的資料來源。
- 按一下**新建資料來源**按鈕會使用您在**新增資料來源**對話框中指定的名稱和內容類型建立新的資料來源。
 - 資料來源名稱必須區分大小寫。系統會忽略前導及尾端空格。某些名稱會遭到拒絕以免發生 SQL 注入。
 - 可用的內容類型為檔案、資料庫和 HCatalog。

註：選取內容類型之後，就無法編輯。

- 按一下**刪除資料來源**按鈕以移除資料來源。這個動作會讓與資料來源相關的所有檔案都保持完好。

內容區

內容區劃分為若干個區段，視資料來源的內容類型而定。指定資料來源的設定之後，按一下預覽和 meta 資料以終結資料來源。

資料來源內容

所有內容類型的共同設定。

- 名稱** 一個可編輯的文字欄位，顯示資料來源的名稱。資料來源名稱必須區分大小寫。系統會忽略前導及尾端空格。某些名稱可能會遭到拒絕以免發生 SQL 注入。
- 說明** 一個可編輯的文字欄位，用來提供關於資料來源的解釋性文字。
- 為公用** 一個勾選框，指出是任何人都可以看到專案（勾選）還是必須明確地將使用者和群組新增至所有者清單（清除）。
- 共用** 您可以透過將使用者和群組新增為作者來共用資料來源的所有權。
- 在文字框中輸入內容可過濾名稱中含有搜尋字串的使用者和群組。按一下**新增參與者**按鈕以將他們新增至作者清單。
 - 若要移除作者，請在「作者」清單中選取使用者或群組，然後按一下**移除參與者**按鈕。
- 註：** 管理者不管是否列出為作者，都對每個資料來源具有讀寫存取權。

檔案輸入

專用於使用檔案內容類型定義資料來源的設定。

檔案檢視器

顯示可併入資料來源中的檔案。選取**專案**模式以檢視 Analytic Server 專案結構中的檔案，或選取 **HDFS** 以檢視 Hadoop 分散式檔案系統的其餘部分。您可以瀏覽任一種資料夾結構，但是 HDFS 完全不可編輯，而 Analytic Server 資料夾僅在定義的專案中可編輯。亦即，您無法在**專案**模式的根層次新增檔案、建立資料夾或刪除項目。若要建立、編輯或刪除專案，請使用專案 Accordion。

- 按一下 **將檔案上傳到 HDFS** 按鈕可將檔案上傳至現行專案/子資料夾。
- 按一下**建立新資料夾**按鈕可在現行資料夾下，使用您在「新資料夾名稱」對話框中指定的名稱建立新的資料夾。
- 按一下**將檔案下載到本端檔案系統**按鈕可將所選檔案下載到本端檔案系統。
- 按一下**刪除所選檔案**按鈕可按鈕所選檔案/資料夾。

資料來源定義中包括的檔案

使用移動按鈕將所選檔案新增至資料來源或從資料來源中移除所選檔案。對於資料來源中每個選定的檔案，按一下設定以定義讀取檔案的規格。

檔案輸出

可對具有檔案內容類型的資料來源附加來自執行於 Analytic Server 上之串流的輸出。選取**設為可寫入**以啟用附加，然後選擇要將新檔案寫入其中的輸出資料夾。

資料庫選擇

指定包含記錄內容之資料庫的連線參數。

資料庫 選取您要連接的資料庫類型。從 DB2、Oracle、SQL Server、TeraData 或 Netezza 中進行選擇。

伺服器位址

輸入管理資料庫之伺服器的 URL。

伺服器埠

資料庫接聽所在的埠號。

資料庫名稱

您要連接的資料庫名稱。

使用者名稱

如果資料庫的密碼受到保護，則輸入使用者名稱。

密碼 如果資料庫的密碼受到保護，則輸入您的密碼。

表格名稱

輸入您要使用之資料庫的某個表格名稱。

並行讀取數上限

資料庫輸出

可對具有資料庫內容類型的資料來源附加來自執行於 Analytic Server 上之串流的輸出。選取**設為可寫入**以啓用附加，然後選擇要將輸出資料寫入其中的輸出資料庫表格。

HCatalog 選擇

指定用於存取資料且在 Apache HCatalog 下管理的參數。

資料庫 HCatalog 資料庫的名稱。

表格名稱

輸入您要使用之資料庫的某個表格名稱。

過濾器 表格的分割區過濾器（如果將表格建立成分割區表格的話）。

HCatalog 綱目

顯示所指定表格的結構。HCatalog 可支援高度結構化的資料集。若要對此類資料定義 Analytic Server 資料來源，則結構必須壓縮至簡式列和欄。選取綱目中的元素，然後按一下移動按鈕將它對映至欄位以進行分析。並非所有樹狀結構節點都可以對映。例如，複式類型的陣列或對映被視為「母項」，無法對映。這些節點可由樹狀結構中結尾為 `...:array:struct` 或 `...:map:struct` 的標籤進行識別。

HCatalog 欄位對映

顯示將 HCatalog 中的元素與資料來源中欄位的對映。按一下**預覽原始資料**可依記錄儲存在 HCatalog 中的現狀查看記錄；這可協助您決定如何將 HCatalog 綱目對映至欄位。

HCatalog 元素

按兩下資料格以進行編輯。當 HCatalog 元素為陣列或對映時，您必須編輯資料格。藉由陣列，指定整數，該整數對於要對映至欄位的陣列成員。藉由對映，指定引號內的字串，該字串對映於要對映至欄位的索引鍵。請參閱第 12 頁的圖 2，以取得有關如何使用原始資料預覽來判定對應於對映索引之字串的範例。

對映欄位

顯示在 Analytic Server 資料來源中的欄位。按兩下資料格以進行編輯。「對映欄位」直欄中不允許有重複的值，否則會導致錯誤。

儲存 欄位的儲存。儲存衍生自 HCatalog，無法編輯。

註：按一下預覽和 meta 資料以終結 HCatalog 資料來源時，不存在編輯選項。

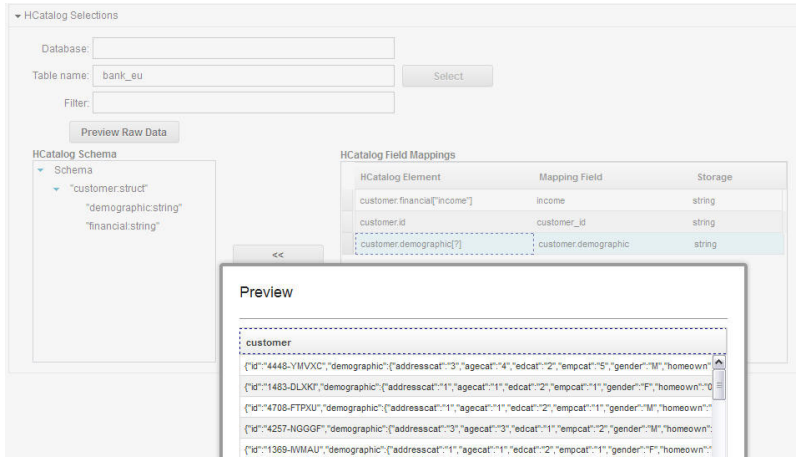


圖 2. 用來定義 HCatalog 資料來源的資料來源 accordion

設定 (檔案資料來源)

字集編碼

檔案的字元編碼。選取或指定 Java 字集名稱，例如 "UTF-8"、"ISO-8859-2" 和 "GB18030"。預設值是 "UTF-8"。

語言環境

定義語言環境。選用項目。預設為伺服器語言環境。語言環境字串應該採用以下格式指定：
 <language>[_country[_variant]]，其中：

language

ISO-639 定義的兩個小寫字母的有效代碼。必要項目。

country

ISO-3166 定義的兩個大寫字母的有效代碼。選用項目。

variant

供應商或瀏覽器專用代碼。選用項目。

修整空格

移除字串欄位開頭及/或末尾的空格字元。預設為「無」。支援下列值：

- 無** 不移除空格字元。
- 左側** 移除字串開頭的空格字元。
- 右側** 移除字串末尾的空格字元。
- 兩者** 移除字串欄位開頭和末尾的空格字元。

分組符號

設定是否將語言環境特有的語言環境用作千位分隔字元。

欄位定界字元

標示欄位界限的一或多個字元。每一個字元都視為獨立定界字元予以採用。例如，",\t" 表示逗點或 tab 標記欄位界限。如果控制字元定界欄位，則在這裡指定的字元視為除了控制字元以外的定界字元。如果控制字元不定界欄位，則預設值為 ","，如果控制字元定界欄位，則預設值為空字串。

控制字元定界欄位

設定視為欄位定界字元的 ASCII 控制字元，LF 和 CR 除外。預設值為否。

第一列包含欄位名稱

設定是否使用第一列來決定欄位名稱。預設值為否。

要跳過的起始字元數目

檔案開頭要跳過的字元數目。非負整數。預設值為零。

合併空格

設定是否將空格和/或 tab 的多個相鄰出現視為單個欄位定界字元。如果空格與 tab 都不是欄位定界字元，則無效。預設值是 true。

行尾註解字元

標示行尾註解的一或多個字元。將會忽略記錄上該字元及隨後的所有內容。每一個字元都視為獨立註解標記予以採用。例如，"/*" 表示斜線或星號開始註解。不能定義多字元註解標記，例如 "//"。空字串表示未定義任何註解字元。如果已定義，會先檢查註解字元，再處理引號或跳過要跳過的起始字元。預設值是空字串。

無效字元

判定如何處理無效字元（位元組順序未對應於編碼中的字元）。空字串表示捨棄無效字元。非空字串（通常是單個字元）表示用字串的內容取代無效字元。預設值是空字串。

單引號 指定單引號（所有格號）的處理。預設值為**保留**。

保留 單引號沒有特殊意義，與任何其他字元一樣處理。

捨棄 除非用引號括起，否則刪除單引號

對組 單引號視為引號字元，單引號配對之間的字元失去任何特殊意義（將它們視為用引號括起）。用單引號括起的字串內部出現的單引號本身由設定**透過重複輸入括起引號**來決定。

雙引號 指定雙引號的處理。預設值為**對組**。

保留 雙引號沒有特殊意義，與任何其他字元一樣處理。

捨棄 除非用引號括起，否則刪除雙引號

對組 雙引號視為引號字元，雙引號配對之間的字元失去任何特殊意義（將它們視為用引號括起）。用雙引號括起的字串內部出現的雙引號本身由設定**透過重複輸入括起引號**來決定。

透過重複輸入括起引號

指出雙引號是否可以出現在用雙引號括起的字串中以及當設定為**對組**時，單引號是否可以出現在單引號括起的字串中。如果為**是**，則跳出雙引號括起之字串中的雙引號的方法是重複輸入雙引號，跳出單引號括起之字串中的單引號的方法是重複輸入單引號。如果為**否**，則無法用引號括起已用雙引號括起的字串中的雙引號，或用引號括起已用單引號括起的字串中的雙引號。預設值為**是**。

預覽與 meta 資料（資料來源）

按一下**預覽與 meta 資料**可根據記錄的範例產生資料來源的資料模型。在這裡，您有機會檢查並編輯基本 meta 資料資訊。

預覽 「預覽」標籤顯示記錄的少量範例及其欄位值。

編輯

「編輯」標籤可容許您修正基本欄位 meta 資料。

欄位 按兩下欄位名稱可以進行編輯。

度量 這是度量層次，用於說明給定欄位中資料的性質。

角色 用來告訴建模節點，欄位將是機器學習處理程序的輸入（預測工具欄位）還是目標（預測欄位）。「兩者」和「無」也是可用角色，「分割區」也是，「分割區」表示用來將記錄分割成個別範例以進行訓練、測試和驗證的欄位。「分割」值指定將為欄位的每個可能值建置個別模型。

儲存 儲存說明將資料儲存在欄位中的方式。例如，值為 1 和 0 的欄位儲存整數資料。這與度量層次截然不同，後者說明資料使用情形，不影響儲存。例如，您可能要將值為 1 和 0 之整數欄位的度量層次設為「旗標」。這通常表示 1 = True，而 0 = False。

專案

專案是用來儲存工作輸入以及存取工作輸出的工作區。它們提供用於包含檔案和資料夾的最上層組織結構。可以與個別使用者和群組共用專案。

左側直欄

左側直欄在 **Accordion** 標題下顯示現有專案。

- 選取專案以在內容區中顯示其詳細資料，並編輯其內容。在搜尋區中輸入內容可過濾清單，以僅顯示名稱中含有搜尋字串的專案。
- 按一下**新建專案**以使用您在「新增專案」對話框中指定的名稱建立新專案。名稱區分大小寫，忽略前導和尾端空格，並防止發生 SQL 注入。
- 按一下**刪除專案**以移除專案。這個動作會讓與資料來源相關的所有檔案都保持完好。

內容區

內容區劃分為**設定**、**擁有者**和**版本**三個標籤。

設定

專案說明

一個可編輯的文字欄位，用來提供關於專案的解釋性文字。

為公用 一個勾選框，指出是任何人都可以看到專案（勾選）還是必須明確地將使用者和群組新增為擁有者（清除）。

按一下**儲存**，儲存設定的現行狀態。

專案資料來源

一個不可編輯的區域，列出與專案相關聯的所有資料來源。

專案結構窗格

右窗格顯示目前選取之專案的專案/資料夾結構。您可以瀏覽資料夾結構，但是除非透過按鈕，否則不可予以編輯。

- 按一下**將檔案下載到本端檔案系統**以將所選檔案下載到本端檔案系統。
- 按一下**刪除所選取檔案**以移除所選取的檔案/資料夾。

擁有者

擁有者是專案的完整成員，可以修改專案以及專案中的資料夾和檔案。

「可用的使用者和群組清單」會顯示目前與此專案不關聯的作用中租戶中的使用者和群組。

- 在搜尋窗格中予以鍵入，過濾以顯示其名稱中包含搜尋字串的使用者和群組。
- 選取清單上方的使用者圖示會顯示可用的使用者。清除該圖示會隱藏使用者。依預設會選取此圖示。

- 選取清單上方的群組圖示會顯示可用的群組。清除該圖示會隱藏群組。依預設會選取此圖示。

可以使用移動按鈕將使用者和群組移至「專案使用者和群組」。當透過 IBM SPSS Modeler 連接至 Analytic Server 時，這些群組的這些使用者及成員具有對這個專案的讀取（「Analytic Server 來源」節點）權及寫入（「Analytic Server 匯出」節點）權。

註：在「擁有者」標籤上進行的變更會立即自動套用。

註：管理者不管是否特地列出為擁有者，都對每個專案具有讀寫存取權。

版本

專案的版本取決於對檔案及資料夾內容的變更。對專案屬性的變更（例如說明，它是否是公用的，以及其共用人員）不需要新版本。新增、修改或刪除檔案或資料夾不需要新版本。

專案版本化表格

表格顯示現有專案版本、其建立及確定日期、負責每一個版本的使用者及母項版本。母項版本是所選取版本基於的版本。

- 按一下**鎖定**可變更所選專案版本內容。
- 按一下**確定**儲存對專案進行的所有變更，並讓此版本成為專案的目前可見狀態。
- 按一下**捨棄**以捨棄對已鎖定專案進行的所有變更，並將專案的可見狀態恢復為最近確定的版本。
- 按一下**刪除**以移除選取的版本。

當版本數超過時自動清除

當版本數超出指定的數目時，自動刪除最舊的已確定專案版本。

注意事項

本資訊係針對 IBM 在美國所提供之產品與服務所開發。

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
Armonk, NY 10504-1785 U.S.A.

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.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

下列段落不適用於英國，若與任何其他國家之法律條款抵觸，亦不適用於該國：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。有些地區在特定交易上，不允許排除明示或暗示的保證，因此，這項聲明不一定適合您。

本資訊中可能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。這些網站所提供的資料不是 IBM 本產品的資料內容，如果要使用這些網站的資料，您必須自行承擔風險。

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爲了 (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 Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

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。

本文件中所含的任何效能資料是在控制環境中得出。因此，在其他作業環境中獲得的結果可能有明顯的差異。在開發層次的系統上可能有做過一些測量，但不保證這些測量在市面上普遍發行的系統上有相同的結果。再者，有些測定可能是透過推測方式來評估。實際結果可能不同。本文件的使用者應驗證其特定環境適用的資料。

本文件所提及之非 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。有關非 IBM 產品的性能問題應直接洽詢該產品供應商。

所有關於 IBM 未來方針或目的之聲明，隨時可能更改或撤銷，不必另行通知，且僅代表目標與主旨。

所有 IBM 價格都是 IBM 建議的零售價格，可隨時變更而不另行通知。經銷商價格可不同。

本資訊僅作規劃目的。在產品可用前，此處的資訊可能變更。

本資訊含有日常商業運作所用之資料和報告範例。爲了盡可能地加以完整說明，範例中含有個人、公司、品牌及產品的名稱。所有這些名稱全爲虛構，任何與實際商場企業使用的名稱及地址類似之處，純屬巧合。

這些範例程式或任何衍生成果的每份複本或任何部分，都必須依照下列方式併入著作權聲明：

本資訊含有日常商業運作所用之資料和報告範例。爲了盡可能地加以完整說明，範例中含有個人、公司、品牌及產品的名稱。所有這些名稱全爲虛構，任何與實際商場企業使用的名稱及地址類似之處，純屬巧合。

這些範例程式或任何衍生成果的每份複本或任何部分，都必須依照下列方式併入著作權聲明：

©（您的公司名稱）（年份）。本程式之若干部分係衍生自 IBM 公司的範例程式。

© Copyright IBM Corp.（輸入年份）。All rights reserved.

若 貴客戶正在閱讀本項資訊的電子檔，可能不會有照片和彩色說明。

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 商標的最新清單可在 Web 的『Copyright and trademark information』中找到，網址爲 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。

IT Infrastructure Library 是 Central Computer and Telecommunications Agency（現在是 Office of Government Commerce 的一部分）的註冊商標。

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。

ITIL 是 Minister for the Cabinet Office 在美國 Patent and Trademark Office 註冊的註冊商標及註冊社群商標。

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。

Cell Broadband Engine 是 Sony Computer Entertainment, Inc. 在美國及/或其他國家或地區的商標並在當地軟體使用權下使用。

Linear Tape-Open、LTO、LTO 標誌、Ultrium 及 Ultrium 標誌是 HP、IBM Corp. 及 Quantum 在美國及其他國家的商標。

索引

索引順序以中文字，英文字，及特殊符號之次序排列。

〔七劃〕

串流內容

Analytic Server 4

〔八劃〕

來源節點

Analytic Server 來源 3

〔十一劃〕

專案 14

〔十三劃〕

匯出節點

Analytic Server 匯出 3

資料來源 9

A

Analytic Server 來源 3

Analytic Server 匯出 3



Printed in Taiwan