IBM SPSS Analytic Server
第 2 版

使用手冊

附註

在使用本資訊及它支援的產品之前，請閱讀第 31 頁的『注意事項』中的資訊。

產品資訊

此版本適用於 IBM SPSS Analytic Server 2.0.0 版及後續之所有版本與修訂，但新版本另有說明者不在此限。

目錄

第 1 章 第 2 版中對使用者新增的功能

Analytic Server 主控台

新佈置 佈置已變更，因此現在透過首頁而不是 Accordion 來存取頁面。

資料來源

- 您可以定義資料來源的自訂屬性，並檢視其他應用程式建立的自訂屬性。
- 建立資料來源的 meta 資料時，您可以起始所有資料值的掃描，以判定種類值及範圍限制。 掃描所有資料值可確保 meta 資料正確，但是如果資料來源具有許多欄位及記錄，則可能會花費一些時間。
- 支援更多類型的資料來源。

  檔案內容類型

  多種檔案內容類型的支援包括其他設定及剖析器格式。 您還可以針對資料來源中的每一個檔案，定義欄位的剖析順序。 當向資料來源新增目錄時，您可以指定用於在該目錄或其子目錄中選取檔案的規則。

  半結構化檔案

  這些檔案（如 Web 日誌）雖然沒有定界文字檔的結構化程度高，但其包含的資料，可透過正規表示式，擷取至記錄及欄位。

  壓縮檔案

  受支援的壓縮格式包括 Gzip、Deflate、Bz2、Snappy 及 IBM CMX。 此外，還支援具有先前提到之任何壓縮格式的順序檔案。

  不同格式的文字型檔案

  對於 Text Analytics，單一文字型資料來源現在可以包含不同格式的文件（PDF、Microsoft Word 等）。

  SPSS Statistics 檔案

  SPSS Statistics 檔案（*.sav、*.zsav）是包含資料模型的二進位檔。

  可分割二進位格式檔案 (*.asbf)

  此檔案類型有時由 Analytic Server 輸出；例如，當分析需要使用具有清單值的欄位時。

  順序檔案

  順序檔案 (*.seq) 是結構化爲鍵值組的文字檔。 它們通常用作 MapReduce 工作中的中介格式。

  資料庫內容類型

  如果 Analytic Server 已配置爲能夠使用 Greenplum、MySQL 及 Sybase IQ 的資料來源，您可以定義那些資料來源。

  HCatalog 內容類型

  如果 Analytic Server 已配置爲能夠使用 Apache Cassandra、MongoDB 及 Oracle NoSQL 的資料來源，您可以定義那些資料來源。

  地理內容類型

  您可以使用 Shape 檔或線上對映服務，來定義地理的資料來源。

Analytics

新 SPSS Modeler 功能

合併 新增依據排名條件進行合併的支援。

時間序列

新增處理時間序列支援，以及對時間原因模型 (TCM) 的分散式建置及評分支援。 請參閱 SPSS Modeler 中的 AS 時間間隔、串流 TCM 及 TCM 節點。

空間資料

新增處理地理座標系統的支援，以及對地理關聯規則 (GSAR) 及時空點處理 (STP) 模型的分散式建置及評分支援。 請參閱 SPSS Modeler 中的重新投射、關聯規則及 STP 節點。

叢集作業

新增對兩步叢集模型的分散式建置及評分支援。 請參與 SPSS Modeler 中的 TwoStep-AS 節點。

現有 SPSS Modeler 功能的改進支援

聚集 字串欄位可以使用最小值、最大值及非空值計數進行聚集。 在「最佳化」標籤上，支援數值欄位的近似順序統計資料（中位數、四分位數）。

合併 新增依據條件合併及依據無鍵值的鍵合併的支援；例如，產生廣域平均值。

組合建模

對用於建置樹狀結構、線性及神經網絡模型之組合模型的演算法進行了改進，以更好地處理非隨機跨一致大小區塊分佈的資料。

第 2 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。

登入

1. 在瀏覽器的位址列中輸入 Analytic Server 的 URL。 該 URL 可從伺服器管理者處取得。
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，會顯示主控台首頁。

導覽主控台

- 標頭會顯示產品名稱、目前登入之使用者的名稱，以及說明系統的鏈結。 目前登入之使用者的名稱是包含登出鏈結之下拉清單的標頭。
- 內容區會顯示您可以從主控台首頁採取的動作。

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）。 資料來源擁有者可以授與或限制對資料來源的存取權。

資料來源清單

主要「資料來源」頁面提供現行使用者屬於其成員的資料來源清單。

- 按一下資料來源名稱，以顯示其詳細資料，並編輯其內容。
- 在搜尋區中鍵入內容可過濾清單，以僅顯示名稱中含有搜尋字串的資料來源。
- 按一下**新建**，以使用您在**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。
  - 請參閱第 25 頁的『命名規則』，以取得您可以爲資料來源提供之名稱的限制。
  - 可用的內容類型爲「檔案」、「資料庫」、HCatalog 及「地理」。

    註：僅當已配置 Analytic Server 以處理那些資料來源時，HCatalog 選項才可用。

    註：選取內容類型之後，就無法編輯。
- 按一下**刪除**，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。
- 按一下**重新整理**，以更新清單。
- 「動作」下拉清單會執行所選取的動作。
  1. 選取**匯出**，以建立資料來源的保存檔，並將其儲存至本端檔案系統。
  2. 選取**匯入**，以匯入「匯出」動作所建立的保存檔。
  3. 選取**複製**，以建立資料來源的副本。

個別資料來源詳細資料

內容區劃分爲若干個區段，視資料來源的內容類型而定。

詳細資料

這些設定對所有內容類型都是通用的。

名稱 一個可編輯的文字欄位，顯示資料來源的名稱。

顯示名稱

一個可編輯的文字欄位，如在其他應用程式中顯示的那樣，顯示資料來源的名稱。 如果該欄位爲空白，則「名稱」會用作顯示名稱。

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資料來源（勾選）還是必須明確地將使用者和群組新增爲成員（清除）。

自訂屬性

應用程式可以透過使用自訂屬性，將內容附加至資料來源，例如資料來源是否是暫存資料來源。 這些屬性在 Analytic Server 主控台中顯示，以提供對應用程式如何使用資料來源的進一步深入瞭解。

按一下**儲存**，以儲存設定的現行狀態。

共用

這些設定對所有內容類型都是通用的。

您可以透過將使用者和群組新增爲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**新增成員**，以將他們新增至作者清單。
- 若要移除作者，請在成員清單中選取使用者或群組，然後按一下**移除成員**。

註：管理者不管是否明確列出爲成員，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，或者選取**資料來源**，以檢視儲存在資料來源中的檔案，或者選取**檔案系統**，以檢視檔案系統（一般爲 HDFS）。 您可以瀏覽資料夾結構，但是 HDFS 根本無法編輯，且在**專案**模式下，您無法於根層次，而只能在已定義的專案內新增檔案、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。

- 按一下**上傳**，以將檔案上傳至現行資料來源或專案/子資料夾。 您可以瀏覽，以在單一目錄中尋找並選取多個檔案。
- 按一下**新增資料夾**，以使用您在「新資料夾名稱」對話框中指定的名稱，在現行資料夾下建立新的資料夾。
- 按一下**下載**，以將所選取的檔案下載至本端檔案系統。
- 按一下**刪除**，以移除所選取的檔案/資料夾。

資料來源定義中包括的檔案

使用移動按鈕，以將所選檔案及資料夾新增至資料來源或從資料來源中移除所選檔案及資料夾。 對於資料來源中每個所選取檔案或資料夾，按一下設定，以定義用於讀取檔案的規格。

當資料來源中包含多個檔案時，它們必須共用一般 meta 資料；即每一個檔案必須具有相同的欄位數目，欄位必須在每一個檔案中以相同的順序剖析，且每一個檔案必須跨所有檔案具有

相同的儲存。 檔案之間的不符可能會導致主控台無法建立預覽與 meta 資料，或者在 Analytic Server 讀取檔案時，有效值被剖析爲無效（空值）。

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 選取要連接的資料庫類型。 從 DB2、Greenplum、MySQL、Netezza、Oracle、SQL Server、Sybase IQ 或 TeraData 中進行選擇。 如果您正在尋找的類型未列出，請要求伺服器管理者，使用適當的 JDBC 驅動程式，對 Analytic Server 進行配置。

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

請輸入可從 Analytic Server 傳送至要讀取資料來源中指定表格之資料庫的平行查詢數目限制。

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）。 僅字串類型的 Hive 分割區索引鍵支援 HCatalog 過濾。

註：!=、<> 及 LIKE 運算子不會在特定 Hadoop 發行套件中運作。 這是 HCatalog 與那些發行套件之間的相容性問題。

HCatalog 欄位對映

顯示將 HCatalog 中的元素與資料來源中欄位的對映。 按一下編輯，以修改欄位對映。

註：在建立從 Hive 表格顯示資料的 HCatalog 型資料來源之後，您可能會發現，如果 Hive 表格由大量資料檔案構成，則每次 Analytic Server 開始從資料來源讀取資料時，都會發生長時間的延遲。 如果您注意到這種延遲，請使用較少數目的較大型資料檔案來重建 Hive 表格，並將檔案數目減少至 400 或更少。

地理選擇

指定用以存取地理資料的參數。

地理類型

地理資料可以來自線上對映服務或 Shape 檔。

如果您正在使用對映服務，請指定服務的 URL，並選取您要使用的對映層。

如果您正在使用 Shape 檔，請上傳 Shape 檔。

預覽與 meta 資料

指定資料來源的設定之後，按一下預覽與 meta 資料，以檢查並確認資料來源規格。

輸出 可對具有檔案或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：

- 對於具有資料庫內容類型的資料來源，選擇要將輸出資料寫入其中的輸出資料庫表格。
- 對於具有檔案內容類型的資料來源：
  1. 選擇要將新檔案寫入其中的輸出資料夾。

     提示： 對每一個資料來源使用不同的資料夾，以便可以更輕鬆地追蹤檔案與資料來源之間的關聯。
  2. 選取檔案格式；**CSV**（逗點區隔變數）或**可分割二進位格式**。
  3. 選擇性地選取**建立順序檔案**。 如果您要建立可在下游 MapReduce 工作中使用的可分割壓縮檔案，則此選項非常有用。
  4. 選取**可以跳出換行字元**，以讓資料中的換行字元在輸出檔中寫爲字串 ″\n″，而字串 ″\n″ 在輸出檔中寫爲 ″\\n″。 如果未選取，則字串 ″\n″ 會在輸出檔中寫爲 ″\n″，換行字元的存在會導致錯誤。
  5. 選取壓縮格式。 該清單包含已配置用於安裝 Analytic Server 的所有格式。

     註：壓縮格式與檔案格式的部分組合會導致輸出無法分割，因此不適用於進一步 MapReduce 處理。 當您進行此類選取時，Analytic Server 會在「輸出」區段產生警告。

設定（檔案資料來源）

「設定」對話框可讓您定義用於讀取檔案型資料的規格。 在**選取檔案**標籤上，這些設定會套用至所選取的所有檔案，以及所選取資料夾中符合準則的所有檔案。

指定檔案的不正確剖析器設定可能會導致主控台無法建立預覽與 meta 資料，或者在 Analytic Server 讀取檔案時，有效值被剖析爲無效（空值）。

設定標籤

「設定」標籤可讓您指定檔案類型，以及該檔案類型特定的剖析器設定。

您可以使用任何受支援檔案格式的壓縮檔案，來定義資料來源。 受支援的壓縮格式包括 Gzip、Deflate、Bz2、Snappy 及 IBM CMX。

定界檔案類型

定界檔案是自由欄位文字檔，其記錄包含固定數目的欄位，但每一個欄位的字元數是變化的。 定界檔案通常具有 *.csv 或 *.tab 副檔名。 如需相關資訊，請參閱第 7 頁的『定界檔案類型設定』。

固定檔案類型

固定欄位文字檔其欄位沒有定界，但開始於相同的位置，且長度固定。 固定欄位文字檔一般具有 *.dat 副檔名。 如需相關資訊，請參閱第 8 頁的『固定檔案類型設定』。

半結構化檔案類型

半結構化檔案（如 *.log）是文字檔，其具有可預測的結構，可透過正規表示式對映至欄位，但結構化程度沒有定界檔案那麼高。 如需相關資訊，請參閱第 9 頁的『半結構化檔案類型設定』。

Text Analytics 檔案類型

Text Analytics 檔案是可使用 SPSS Text Analytics 進行分析的文件（如 *.doc、*.pdf 或 *.txt）。

跳過空行
指定是否忽略所擷取文字內容中的空行。 預設值爲**否**。

行分隔字元
指定用於定義換行的字串。 預設值爲換行字元 "\n"。

SPSS Statistics 檔案類型

SPSS Statistics 檔案（*.sav、*.zsav）是包含資料模型的二進位檔。 對於此檔案類型，不需要在「設定」標籤上進行進一步設定。

可分割二進位格式檔案類型

指定檔案類型爲可分割二進位格式檔案 (*.asbf)。 此檔案類型有時由 Analytic Server 輸出；例如，當分析需要使用具有清單值的欄位時。 對於此檔案類型，不需要在「設定」標籤上進行進一步設定。

順序檔案類型

順序檔案 (*.seq) 是結構化爲鍵值組的文字檔。 它們通常用作 MapReduce 工作中的中介格式。

Excel 檔案類型

指定檔案類型爲 Microsoft Excel 檔案（*.xls、*.xlsx）。 如需相關資訊，請參閱第 10 頁的『Excel 檔案類型設定』。

定界檔案類型設定：

您可以針對定界檔案類型，指定下列設定。

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"。 預設值爲 **UTF-8**。

欄位定界字元
標示欄位界限的一或多個字元。 每一個字元都視爲獨立定界字元予以採用。 例如，如果您選取**逗點**及 **Tab**（或選取**其他**並鍵入 **,\t**），則表示以逗點或 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爲除了控制字元以外的定界字元。 如果控制字元不定界欄位，則預設值爲 ","；否則預設值爲空字串。

控制字元定界欄位
設定視爲欄位定界字元的 ASCII 控制字元，LF 和 CR 除外。 預設值爲**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。 預設值爲**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。 預設值爲 0。

合併空格

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

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。 每一個字元都視爲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。 預設值爲**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 單引號視爲引號字元，單引號配對之間的字元失去任何特殊意義（將它們視爲用引號括起）。 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。 預設值爲**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 雙引號視爲引號字元，雙引號配對之間的字元失去任何特殊意義（將它們視爲用引號括起）。 用雙引號括起的字串內部是否可以出現雙引號本身由設定**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，以及當設定爲**對組**時，單引號是否可以出現在單引號括起的字串中。 如果爲**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。 如果爲**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。 預設值爲**是**。

可以跳出換行字元

指出在讀取檔案時，剖析器是否將字串 ″\n″ 解譯爲換行字元。 如果未跳出換行字元，則僅會將 ″\n″ 讀取爲字串。 如果跳出換行字元，則 ″\n″ 會讀取爲 ASCII 換行字元，而 ″\\n″ 會讀取爲字串 ″\n″。 預設值爲**否**。

固定檔案類型設定：

您可以針對固定檔案類型，指定下列設定。

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″。 預設值爲 **UTF-8**。

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。

記錄長度

指出記錄是如何定義的。 如果是**換行字元定界**，則記錄由換行字元、檔案開頭或檔案結尾定義（定界）。 如果是**特定長度**，則記錄由記錄長度（以位元組爲單位）定義。 請指定正值。

起始要跳過的記錄

檔案開頭處要跳過的記錄數。 請指定非負整數。 預設值爲 0。

欄位 此區段會定義檔案中的欄位。 按一下**新增欄位**，並指定欄位名稱、欄位值開始所在的直欄，以及欄位值的長度。 檔案中的直欄編號起始於 0。

半結構化檔案類型設定：

半結構化檔案的設定包含用於將檔案內容對映至欄位的規則。

規則表格

個別規則會從記錄中擷取資訊，以建立欄位；這些規則一起在規則表格中，可定義可從資料來源中每一筆記錄擷取的所有欄位。

表格中的規則會按順序套用至每一筆記錄；如果表格中的所有規則都符合記錄，則任何其他規則表格都無需再處理該記錄，而繼續處理下一筆記錄。 如果表格中存在不相符的任何規則，則會捨棄表格中先前規則所擷取的所有欄位值；如果存在其他規則表格，則該表格中的規則會套用至該記錄。 如果沒有表格符合該記錄，則會套用「不符」規則。

不符 您可以選擇**跳過**不符合任何規則表格的記錄，或者將記錄中所有欄位的值設爲**遺漏**（空值）。

匯出規則

您可以儲存目前可見的規則表格，以供重複使用。 匯出的表格儲存在伺服器上。

匯入規則

您可以將已儲存的規則表格匯入至目前可見的的規則表格中。 這會改寫您對該表格定義的任何規則，因此最好是建立新表格，然後匯入規則表格。

規則編輯器

規則編輯器可讓您針對單一欄位，建立擷取規則。

匿名擷取群組

欄位擷取規則一般會從前一個規則停止位置處的記錄開始擷取資料。 當半結構化資料來源中的兩個欄位之間存在無關資訊時，定義匿名擷取群組，以將剖析器定位在下一個欄位開始位置會非常有用。 當您選取**匿名擷取群組**時，會停用對擷取群組進行命名及加標籤的控制項，但對話框的其餘部分功能正常。

欄位名稱

輸入欄位的名稱。 這用於定義資料來源 meta 資料。 欄位名稱在規則表格中必須唯一。

規則名稱

選擇性地輸入規則的敘述性標籤。

說明 選擇性地輸入規則的較詳細說明。

定義規則

有兩種方法可定義規則。

對擷取規則使用控制項

這會簡化擷取規則的建立。

1. 指定開始擷取欄位資料的點；**現行位置**會從前一個規則停止的位置開始，而**跳過直到**會從記錄的開頭開始，忽略所有字元，直到其到達文字框中所指定的位置爲止。 如果您想讓欄位資料包含開始位置處的字元，請選取**包括**。
2. 從**擷取**下拉清單中選取欄位擷取群組。
3. 選擇性地選取要停止擷取欄位資料的點；**空格**會在遇到任何空格字元（如空格或 Tab）時停止，而**特定字元**會在指定的字串處停止。 如果您想讓欄位資料包含停止位置處的字元，請選取**包括**。

手動定義正規表示式規則

如果您善於撰寫正規表示式語法，請選取此選項。 請在**正規表示式**文字框中，輸入正規表示式。

新增欄位擷取群組

這可讓您儲存正規表示式以供以後使用。 已儲存的擷取群組會顯示在**擷取**下拉清單中。

在已套用規則表格中的所有先前規則之後，「規則編輯器」會顯示依據此規則，從第一筆記錄擷取的資料預覽。

Excel 檔案類型設定：

您可以針對 Excel 檔案，指定下列設定。

選取工作表

選取要用作資料來源的 Excel 工作表。 指定數值索引（第一個工作表的索引爲 0），或者工作表的名稱。 預設爲使用第一個工作表。

選取要匯入的資料範圍。

您可以匯入以第一個非空白列或以明確儲存格範圍開頭的資料。

- **範圍起始於第一個非空白列。** 尋找第一個非空白儲存格，並將其用作資料範圍的左上角。
- 此外，也可以依列和欄，指定明確的儲存格範圍。 例如，若要指定 Excel 範圍 A1:D5，您可以在第一個欄位中輸入 A1，在第二個欄位中輸入 D5（或者 R1C1 及 R5C4）。 指定範圍內的所有列都會傳回，包括空白列。

第一列包含欄位名稱

指定所選取儲存格範圍的第一列是否包含欄位名稱。 預設值爲**否**。

遇到空白列之後停止讀取

指定在遇到多個空白列之後，是停止讀取記錄，還是繼續讀取所有資料，直到工作表的結尾，包括空白列。 預設值爲**否**。

格式

「格式」標籤可讓您定義剖析欄位的格式化資訊。

欄位轉換設定

修整空格

移除字串欄位開頭及/或末尾的空格字元。 預設值爲**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。

語言環境

定義語言環境。 預設爲伺服器語言環境。 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。

country

ISO-3166 定義的兩個大寫字母的有效代碼。

variant

供應商或瀏覽器專用代碼。

小數點 設定用作小數符號的字元。 預設爲語言環境特定設定。

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。

預設日期格式

定義預設日期格式。 支援 Unicode 語言環境資料標記語言 (LDML) 規格所定義的所有格式型樣。

預設時間格式

定義預設時間格式。

預設時間戳記

定義預設時間戳記格式。

預設時區

設定時區。 預設爲世界標準時間。 該設定會套用至沒有明確指定時區的時間及時間戳記欄位。

欄位置換

此區段可讓您將格式化指令指派給個別欄位。 從資料模型選取欄位，或者鍵入欄位名稱，並按一下**新增**，以將其新增至具有個別指令的欄位清單。 按一下**移除**，以將其從清單中移除。 對於清單中所選取的欄位，您可以設定欄位的下列內容。

儲存 設定欄位的儲存。

小數點 對於具有「實數」儲存的欄位，設定用作小數符號的字元。 預設爲語言環境特定設定。

分組符號

對於具有「整數」或「實數」儲存的欄位，設定是否應該使用用於千位分隔字元的語言環境特定字元。

格式 對於具有「日期」、「時間」或「時間戳記」儲存的欄位，設定格式。 從下拉清單中選擇格式。

欄位順序標籤

對於定界及 Excel 檔案類型，「欄位順序」標籤可讓您定義檔案的欄位剖析順序。 當資料來源中存在多個檔案時，這非常重要，因爲欄位的實際順序在檔案之間可能不同，但欄位的剖析順序必須相同，才能建立一致的資料模型。

對於固定及半結構化檔案類型，該順序會在「設定」標籤上定義。

當資料來源中只有一個檔案，或者所有檔案具有相同的欄位順序時，您可以使用預設的**欄位順序符合資料模型**。 如果資料來源中有多個檔案，且檔案中的欄位順序不相符，請定義**特定欄位順序**，以用於剖析檔案。

1. 若要將欄位新增至有序清單，請鍵入欄位名稱，或者從資料模型所提供的清單中進行選取。 您可以透過按一下**全部新增**，一次新增資料模型中的所有欄位。 欄位名稱僅會向有序清單中新增一次。

2. 使用箭頭按鈕，可根據需要對欄位進行排序。

當使用**特定欄位順序**時，未新增至清單的任何欄位都不會成爲此檔案結果集的一部分。 如果資料模型中有欄位未在此對話框中列出，則值在結果集中爲空值。

資料夾標籤

當指定資料夾的剖析器設定時，「資料夾」標籤可讓您選擇資料夾中的哪些檔案包含在資料來源中。

符合所選取資料夾中的所有檔案

資料來源包含最上層資料夾中的所有檔案；不包含子資料夾中的檔案。

符合使用正規表示式的檔案

資料來源包含最上層資料夾中符合指定正規表示式的所有檔案；不包含子資料夾中的檔案。

符合使用 Unix 檔名展開表示式（可能遞迴）的檔案

資料來源包含符合指定 Unix 檔名展開表示式的所有檔案；表示式可以包含所選取資料夾之子資料夾中的檔案。

HCatalog 欄位對映

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爲「母項」，無法直接對映；HCatalog 陣列或對映中的每一個簡式元素必須分別新增。 這些節點可由樹狀結構中結尾爲 ...:array:struct 或 ...:map:struct 的標籤進行識別。

例如：

- 對於整數陣列，您可以將欄位指派給陣列中的值：bigintarray[45]，但不能指派給陣列本身：bigintarray
- 對於對映，您可以將欄位指派給對映中的值：datamap[″key″]，但不能指派給對映本身：datamap
- 對於整數陣列的陣列，您可以將欄位指派給值 bigintarrayarray[45][2]，但不能指派給陣列本身 bigintarrayarray[45]。

因此，當您將欄位指派給陣列或對映元素時，元素的定義必須包括索引或鍵：bigintarray[index] 或 bigintmap[″key″]。

欄位對映

HCatalog 元素

按兩下資料格以進行編輯。 當 HCatalog 元素爲陣列或對映時，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。

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。

原始資料

以記錄在 HCatalog 中儲存的現狀顯示記錄；這可協助您決定如何將 HCatalog 綱目對映至欄位。

註： 「HCatalog 選項」中指定的任何過濾都會套用至原始資料的視圖。

啓用 HCatalog 資料來源

Analytic Server 提供 HCatalog 資料來源的支援。本節說明如何啓用各種基礎 NoSQL 資料庫。

Apache Accumulo

Analytic Server 提供對在 Apache Accumulo 中具有基礎內容之 HCatalog 資料來源的支援。

Apache Accumulo 分散式鍵/值儲存庫是基於 Google BigTable 設計的資料儲存及擷取系統，其以 Apache Hadoop、Zookeeper 及 Thrift 爲基礎進行建置。 Apache Accumulo 以儲存格型存取控制的形式，以及伺服器端程式設計的機制，可在資料管理處理程序的各個點上，對鍵值組進行修改，從而對 BigTable 設計進行了諸多創新改進。

若要在 Hive 中建立外部 Apache Accumulo 表格，請使用下列語法：

```
set accumulo.instance.id=<instance_name>;
set accumulo.user.name=<user_name>;
set accumulo.user.pass=<user_password>;
set accumulo.zookeepers=<zookeeper_host_port>;

CREATE EXTERNAL TABLE <hive_table_name>(<table_column_specifications>)
STORED BY 'com.ibm.spss.hcatalog.AccumuloStorageHandler'
WITH SERDEPROPERTIES (
'accumulo.columns.mapping' = '<family_and_qualifier_mappings>',
'accumulo.table.name' = '<Accumulo_table_name>')
TBLPROPERTIES (
    "accumulo.instance.id"="<instance_name>",
    "accumulo.zookeepers"="<zookeeper_host_port>"
);
```

例如：

```
set accumulo.instance.id=<id>;
set accumulo.user.name=admin;
set accumulo.user.pass=test;
set accumulo.zookeepers=<host>:<port>;

CREATE EXTERNAL TABLE acc_drug1n(rowid STRING,age STRING,sex STRING,bp STRING,
    cholesterol STRING,na STRING,k STRING,drug STRING)
STORED BY 'com.ibm.spss.hcatalog.AccumuloStorageHandler'
WITH SERDEPROPERTIES (
'accumulo.columns.mapping' = 'rowID,drug|age,drug|sex,drug|bp,drug|cholesterol,
    drug|na,drug|k,drug|drug',
'accumulo.table.name' = 'drug1n')
TBLPROPERTIES (
    "accumulo.instance.id"="<id>",
    "accumulo.zookeepers"="<host>:<port>"
);
```

註：給定 Accumulo 表格的 Accumulo 使用者名稱及密碼應該符合已鑑別 Analytic Server 使用者的使用者名稱及密碼。

Apache Cassandra

Analytic Server 提供對在 Apache Cassandra 中具有基礎內容之 HCatalog 資料來源的支援。

Cassandra 提供結構化的鍵值儲存庫。 鍵與多個值對映，這些值會分組爲直欄系列。 建立資料庫時，直欄系列是固定的，但可隨時新增直欄至系列。 此外，直欄只能新增至指定的鍵，因此不同的鍵可以在任何給定系列中，具有不同數量的直欄。 針對每一個鍵，來自直欄系列的值會儲存在一起。

有兩種方法來定義 Cassandra 表格：使用舊式 Cassandra 指令行介面 (cassandra-cli) 及新 CQL Shell (csqlsh)。

如果表格是使用舊式 CLI 建立的，請使用下列語法，在 Hive 中建立外部 Apache Cassandra 表格。

```
CREATE EXTERNAL TABLE <hive_table_name>  (<column specifications>)
STORED BY 'com.ibm.spss.hcatalog.CassandraStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "<cassandra_column_family>",
"cassandra.host"="<cassandra_host>","cassandra.port" = "<cassandra_port>")
TBLPROPERTIES ("cassandra.ks.name" = "<cassandra_keyspace>");
```

例如，對於下列 CLI 表格定義：

```
create keyspace test
with placement_strategy = 'org.apache.cassandra.locator.SimpleStrategy'
and strategy_options = [{replication_factor:1}];

create column family users with comparator = UTF8Type;

update column family users with
        column_metadata =
        [
        {column_name: first, validation_class: UTF8Type},
        {column_name: last, validation_class: UTF8Type},
        {column_name: age, validation_class: UTF8Type, index_type: KEYS}
        ];

assume users keys as utf8;

set users['jsmith']['first'] = 'John';
set users['jsmith']['last'] = 'Smith';
set users['jsmith']['age'] = '38';
set users['jdoe']['first'] = 'John';
set users['jdoe']['last'] = 'Dow';
set users['jdoe']['age'] = '42';

get users['jdoe'];
```

... Hive 表格 DDL 看起來如下：

```
CREATE EXTERNAL TABLE cassandra_users  (key string, first string, last string, age string)
STORED BY 'com.ibm.spss.hcatalog.CassandraStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "users",
"cassandra.host"="<cassandra_host>","cassandra.port" = "9160")
TBLPROPERTIES ("cassandra.ks.name" = "test");
```

如果表格是使用 CQL 建立的，請使用下列語法，在 Hive 中建立外部 Apache Cassandra 表格。

```
CREATE EXTERNAL TABLE <hive_table_name>  (<column specifications>)
STORED BY 'com.ibm.spss.hcatalog.CassandraCqlStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "<cassandra_column_family>",
"cassandra.host"="<cassandra_host>","cassandra.port" = "<cassandra_port>")
TBLPROPERTIES ("cassandra.ks.name" = "<cassandra_keyspace>");
```

例如，對於下列 CQL3 表格定義：

```
CREATE KEYSPACE TEST WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 2 };
USE TEST;

CREATE TABLE bankloan_10(
  row int,
  age int,
  ed int,
```

```
  employ int,
  address int,
  income int,
  debtinc double,
  creddebt double,
  othdebt double,
  default int,
  PRIMARY KEY(row)
);

INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (1,41,3,17,12,176,9.3,11.359392,5.008608,1);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (2,27,1,10,6,31,17.3,1.362202,4.000798,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (3,40,1,15,14,55,5.5,0.856075,2.168925,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (4,41,1,15,14,120,2.9,2.65872,0.82128,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (5,24,2,2,0,28,17.3,1.787436,3.056564,1);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (6,41,2,5,5,25,10.2,0.3927,2.1573,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (7,39,1,20,9,67,30.6,3.833874,16.668126,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (8,43,1,12,11,38,3.6,0.128592,1.239408,0);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (9,24,1,3,4,19,24.4,1.358348,3.277652,1);
INSERT INTO bankloan_10 (row, age,ed,employ,address,income,debtinc,creddebt,othdebt,default)
   VALUES (10,36,1,0,13,25,19.7,2.7777,2.1473,0);
```

... Hive 表格 DDL 如下所示：

```
CREATE EXTERNAL TABLE cassandra_bankloan_10 (row int, age int,ed int,employ int,address int,
   income int,debtinc double,creddebt double,othdebt double,default int)
STORED BY 'com.ibm.spss.hcatalog.CassandraCqlStorageHandler'
WITH SERDEPROPERTIES("cassandra.cf.name" = "bankloan_10","cassandra.host"="<cassandra_host>",
   "cassandra.port" = "9160")
TBLPROPERTIES ("cassandra.ks.name" = "test");
```

Apache HBase

Analytic Server 提供對在 Apache HBase 中具有基礎內容之 HCatalog 資料來源的支援。

Apache HBase 是以 Hadoop 及 HDFS 爲基礎的開放程式碼、分散式、已版本化、面向直欄的儲存庫。

若要在 Hive 中建立外部 HBase 表格，請使用下列語法：

```
CREATE EXTERNAL TABLE <tablename>(<table_column_specifications>)
STORED BY 'com.ibm.spss.hcatalog.HBaseStorageHandler'
WITH SERDEPROPERTIES ("hbase.columns.mapping" = "<column_mapping_spec>")
TBLPROPERTIES("hbase.table.name" = "<hbase_table_name>")
```

例如：

```
CREATE EXTERNAL TABLE hbase_drug1n(rowid STRING,age STRING,sex STRING,bp STRING,
    cholesterol STRING,na STRING,k STRING,drug STRING)
STORED BY 'com.ibm.spss.hcatalog.HBaseStorageHandler'
WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,drug:age,drug:sex,drug:bp,
    drug:cholesterol,drug:na,drug:k,drug:drug")
TBLPROPERTIES("hbase.table.name" = "drug1n");
```

註：如需如何建立 HBase 表格的相關資訊，請參閱 Apache HBase Reference Guide (http://hbase.apache.org/book.html)。

註：將資料庫名稱作爲前言，以指示資料庫的類型是很好的做法。 例如，將資料庫命名爲 HB_drug1n，以指示 HBase 資料庫，或者命名爲 ACC_drug1n，以指示 Accumulo 資料庫。 這將有助於在 Analytic Server 主控台中，選取 HCatalog 檔案。

MongoDB

Analytic Server 提供對在 MongoDB 中具有基礎內容之 HCatalog 資料來源的支援。

MongoDB 是開放程式碼文件資料庫，其爲以 C++ 撰寫的先進 NoSQL 資料庫。 該資料庫會以動態綱目，儲存 JSON 樣式的文件。

若要在 Hive 中建立外部 MongoDB 表格，請使用下列語法：

```
create external table <hive_table_name>(<column specifications>)
stored by "com.ibm.spss.hcatalog.MongoDBStorageHandler"
with serdeproperties ( "mongo.column.mapping" = "<MongoDB to Hive mapping>" )
tblproperties ( "mongo.uri" = "'mongodb://<host>:<port>/<database>.<collection>" );
```

例如：

```
create external table mongo_bankloan(age bigint,ed bigint,employ bigint, address bigint,income bigint,
    debtinc double, creddebt double,othdebt double,default bigint)
STORED BY 'com.ibm.spss.hcatalog.MongoDBStorageHandler'
with serdeproperties ( 'mongo.column.mapping' = '{"age":"age","ed":"ed","employ":"employ","address":"address",
    "income":"income","debtinc":"debtinc","creddebt":"creddebt","othdebt":"othdebt","default":"default"}' )
tblproperties ('mongo.uri'='mongodb://9.48.11.162:27017/test.bankloan');
```

Oracle NoSQL

Analytic Server 提供對在 Oracle NoSQL 中具有基礎內容之 HCatalog 資料來源的支援。

Oracle NoSQL Database 是分散式鍵值資料庫。 資料會儲存爲鍵值組，其會基於主要鍵的雜湊值，寫入特定儲存節點。 會對儲存節點進行抄寫，以確保高可用性。 客戶應用程式使用 Java/C API 撰寫，以讀取及寫入資料。

SerDe 及表格參數

Oracle NoSQL 儲存處理程式支援下列參數。

SERDEPROPERTIES 參數

kv.major.keys.mapping
逗點區隔的主要鍵清單。 必要項目

kv.minor.keys.mapping
逗點區隔的次要鍵清單。 選用項目

kv.parent.key
指定要由查詢傳回「子項」鍵值組的母項鍵。 主要鍵路徑必須是局部路徑，而次要鍵路徑必須是空的。 選用項目。

kv.avro.json.key
用於保留以 Avro 綱目定義之值的次要鍵名稱。 如果未定義次要鍵（通常會這樣），則預設爲 ″value″。 如果未定義該參數，則值會以 JSON 字串傳回。 選用項目。

kv.avro.json.keys.mapping.column
定義主要/次要鍵值組的 Hive 直欄名稱。 Hive 直欄應該具有對映 <string,string> 類型。 選用項目。

TABLEPROPERTIES 參數

kv.host.port

Oracle NoSQL 資料庫的 IP 位址及埠號。 必要項目

kv.name

Oracle NoSQL 鍵值儲存庫的名稱。 必要項目。

範例：簡式 Avro 綱目

資料佈置使用 Apache Avro 序列化架構進行建模。 若要遵循此方法，您要建立 Avro 綱目；例如：

```
{"type":"record",
 "name":"DrugSchema",
 "namespace":"avro",
 "fields":[
        {"name":"id","type":"string","default":""},
        {"name":"age","type":"string","default":""},
        {"name":"sex","type":"string","default":""},
        {"name":"bp","type":"string","default":""},
        {"name":"drug","type":"string","default":""}
]}
```

此綱目應該向 Oracle NoSQL 資料庫登錄，且已移入的資料應該包括該綱目的參照，如下所示。

```
put -key /drugstore_avro/1 -value
   "{\"id\":\"1\",\"age\":\"23\",\"sex\":\"F\",\"bp\":\"HIGH\",\"drug\":\"drugY\"}"
   -json avro.DrugSchema
put -key /drugstore_avro/2 -value
   "{\"id\":\"2\",\"age\":\"47\",\"sex\":\"M\",\"bp\":\"LOW\",\"drug\":\"drugC\"}"
   -json avro.DrugSchema
put -key /drugstore_avro/3 -value
   "{\"id\":\"3\",\"age\":\"47\",\"sex\":\"M\",\"bp\":\"LOW\",\"drug\":\"drugC\"}"
   -json avro.DrugSchema
put -key /drugstore_avro/4 -value
   "{\"id\":\"4\",\"age\":\"28\",\"sex\":\"F\",\"bp\":\"NORMAL\",\"drug\":\"drugX\"}"
   -json avro.DrugSchema
put -key /drugstore_avro/5 -value
   "{\"id\":\"5\",\"age\":\"61\",\"sex\":\"F\",\"bp\":\"LOW\",\"drug\":\"drugY\"}"
   -json avro.DrugSchema
```

若要顯示 Hive 中的資料，請建立外部表格，並在 SERDEPROPERTIES 區段中指定其他內容 **kv.avro.json.key**。 內容的值應該是次要鍵的名稱或 **value** 的預先定義名稱（如果未定義次要鍵）。

```
CREATE EXTERNAL TABLE oracle_json(id string, age string, sex string, bp string, drug string)
  STORED BY 'com.ibm.spss.hcatalog.OracleKVStorageHandler'
  WITH SERDEPROPERTIES ("kv.major.keys.mapping" = "drugstore_avro,keyid",
        "kv.parent.key"="/drugstore_avro","kv.avro.json.key" = "value")
  TBLPROPERTIES ("kv.host.port" = "<hostname>:5000", "kv.name" = "kvstore");
```

執行 select * from oracle_json 會產生下列結果。

```
select * from oracle_json;

1 23 F HIGH drugY
5 61 F LOW drugY
3 47 M LOW drugC
2 47 M LOW drugC
4 28 F NORMAL drugX
```

表格 oracle_json 可以在 Analytic Server 主控台中使用，以建立 Oracle NoSQL 資料來源。

範例：複式鍵

現在，假設下列 Avro 綱目。

```
{ "type": "record",
  "name": "DrugSchema",
  "namespace":"avro",
  "fields": [
      {"name": "age", "type": "string", "default": ""}, // age
      {"name": "bp", "type":"string","default":""},     // blood pressure
      {"name": "drug", "type": "int", "default": ""},    // drug administered
]}
```

同時假設如下所示對鍵進行建模：

```
/u/<sex (M/F)>/<patient ID>
```

並使用下列指令，對資料儲存庫進行移入作業：

```
put -key /u/F/1 -value
  "{\"age\":\"23\",\"bp\":\"HIGH\",\"drug\":\"drugY\"}" -json avro.DrugSchema
put -key /u/M/2 -value
  "{\"age\":\"47\",\"bp\":\"LOW\",\"drug\":\"drugC\"}" -json avro.DrugSchema
put -key /u/M/3 -value
  "{\"age\":\"47\",\"bp\":\"LOW\",\"drug\":\"drugC\"}" -json avro.DrugSchema
put -key /u/F/4 -value
  "{\"age\":\"28\",\"bp\":\"NORMAL\",\"drug\":\"drugX\"}" -json avro.DrugSchema
put -key /u/F/5 -value
  "{\"age\":\"61\",\"bp\":\"LOW\",\"drug\":\"drugY\"}" -json avro.DrugSchema
```

若要保留主要來自鍵的性別及使用者 ID 相關資訊，應該使用其他 SERDEPROPERTIES 參數 **kv.avro.json.keys.mapping.column**，來建立表格。 參數的值應該是對映 <string,string> 類型的 Hive 直欄名稱。 對映中的鍵會是 **kv.*.keys.mapping** 內容中指定的記錄鍵名稱，而值會是實際的鍵值。 表格建立 DDL 如下所示：

```
CREATE EXTERNAL TABLE oracle_user(keys map<string,string>, age string, bp string, drug string)
  STORED BY 'com.ibm.spss.hcatalog.OracleKVStorageHandler'
  WITH SERDEPROPERTIES ("kv.major.keys.mapping" = "DrugSchema,sex,patientid",
      "kv.parent.key" = "/u",
      "kv.avro.json.key" = "value",
      "kv.avro.json.keys.mapping.column" = "keys")
  TBLPROPERTIES ("kv.host.port" = "<hostname>:5000", "kv.name" = "kvstore");
```

執行 select * from oracle_user 會產生下列結果：

```
select * from
        oracle_user; {"user":"u","gender":"m"," userid":"125"} joe smith 77 13
      {"user":"u","gender":"m"," userid":"129"} jeff smith 67 27
      {"user":"u","gender":"m"," userid":"127"} jim smith 78 11
      {"user":"u","gender":"f"," userid":"131"} jen schmitt 70 20
      {"user":"u","gender":"m"," userid":"130"} jed schmidt 60 31
      {"user":"u","gender":"f"," userid":"128"} jan smythe 79 10
      {"user":"u","gender":"f"," userid":"126"} jess smith 76 12
```

oracle_user 表格可以在 Analytic Server 主控台中使用，以建立 Oracle NoSQL 資料來源。 來自 Avro 綱目的性別及 patientid 鍵以及直欄名稱可用於定義資料來源的對應欄位。

範圍掃描

Analytic Server 支援基於主要鍵及子範圍母項字首的範圍掃描，以進一步將範圍限制在母項鍵下。

母項鍵會指定要傳回之「子項」鍵值組的字首。 空字首會導致提取儲存庫中的所有鍵。 如果字首不是空的，則主要鍵路徑必須是局部路徑，而次要鍵路徑必須是空的。 母項鍵會儲存爲 **com.ibm.spss.ae.hcatalog.range.parent** 資料來源屬性。

子範圍會進一步將母項鍵下的範圍限制爲子範圍中的主要鍵路徑元件。 子範圍啓動鍵會儲存爲 **com.ibm.spss.ae.hcatalog.range.start**，而子範圍結束鍵會儲存爲 **com.ibm.spss.ae.hcatalog.range.end**。啓動鍵在字典上應該小於或等於結束鍵。 子範圍參數爲選用參數。

XML 資料來源

Analytic Server 透過 HCatalog，提供對 XML 資料的支援。

範例

1. 根據下列規則，透過 Hive 資料定義語言 (DDL)，將 XML 綱目對映至 Hive 資料類型。

```
CREATE [EXTERNAL] TABLE <table_name> (<column_specifications>)
ROW FORMAT SERDE "com.ibm.spss.hive.serde2.xml.XmlSerDe"
WITH SERDEPROPERTIES (
    ["xml.processor.class"="<xml_processor_class_name>",]
    "column.xpath.<column_name>"="<xpath_query>",
     ...
    ["xml.map.specification.<element_name>"="<map_specification>"
     ...
    ]
)
STORED AS
    INPUTFORMAT "com.ibm.spss.hive.serde2.xml.XmlInputFormat"
    OUTPUTFORMAT "org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat"
[LOCATION "<data_location>"]
TBLPROPERTIES (
    "xmlinput.start"="<start_tag ",
    "xmlinput.end"="<end_tag>"
);
```

註：如果您的 XML 檔案使用 Bz2 壓縮進行壓縮，則 INPUTFORMAT 應該設爲 com.ibm.spss.hive.serde2.xml.SplittableXmlInputFormat。 如果它們使用 CMX 壓縮進行壓縮，則該值應該設爲 com.ibm.spss.hive.serde2.xml.CmxXmlInputFormat。

例如，下列 XML...

```
<records>
  <record customer_id="0000-JTALA">
    <demographics>
      <gender>F</gender>
      <agecat>1</agecat>
      <edcat>1</edcat>
      <jobcat>2</jobcat>
      <empcat>2</empcat>
      <retire>0</retire>
      <jobsat>1</jobsat>
      <marital>1</marital>
      <spousedcat>1</spousedcat>
      <residecat>4</residecat>
      <homeown>0</homeown>
      <hometype>2</hometype>
      <addresscat>2</addresscat>
    </demographics>
    <financial>
      <income>18</income>
      <creddebt>1.003392</creddebt>
      <othdebt>2.740608</othdebt>
      <default>0</default>
    </financial>
  </record>
</records>
```

...會由下列 Hive DDL 表示。

```
CREATE TABLE xml_bank(customer_id STRING, demographics map<string,string>, financial map<string,string>)
ROW FORMAT SERDE 'com.ibm.spss.hive.serde2.xml.XmlSerDe'
WITH SERDEPROPERTIES (
    "column.xpath.customer_id"="/record/@customer_id",
    "column.xpath.demographics"="/record/demographics/*",
    "column.xpath.financial"="/record/financial/*"
)
STORED AS
    INPUTFORMAT 'com.ibm.spss.hive.serde2.xml.XmlInputFormat'
    OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.IgnoreKeyTextOutputFormat'
TBLPROPERTIES (
    "xmlinput.start"="<record customer",
    "xmlinput.end"="</record>"
);
```

如需相關資訊，請參閱『XML 與 Hive 資料類型的對映』。

2. 在 Analytic Server 主控台中，建立具有 HCatalog 內容類型的 Analytic Server 資料來源。

限制

- 目前僅支援 XPath 1.0 規格。
- 處理 Hive 欄位名稱時，會使用元素及屬性完整名稱的本端部分。 會忽略名稱空間字首。

XML 與 Hive 資料類型的對映： 使用下面記載的使用慣例，可以將 XML 中建模的資料轉換爲 Hive 資料類型。

結構

XML 元素可以直接與 Hive 結構類型相對映，因此所有屬性都會成爲資料成員。 元素的內容會成爲初始或複式類型的其他成員。

XML 資料

```
<result name="ID_DATUM">03.06.2009</result>
```

Hive DDL 及原始資料

```
struct<name:string,result:string>
```

```
{"name":"ID_DATUM", "result":"0.3.06.2009"}
```

陣列

元素的 XML 順序可以表示爲初始或複式類型的 Hive 陣列。 下列範例顯示使用者可以如何使用 XML <result> 元素的內容，來定義字串的陣列。

XML 資料

```
<result>03.06.2009</result>
<result>03.06.2010</result>
<result>03.06.2011</result>
```

Hive DDL 及原始資料

```
result array<string>
```

```
{"result":["03.06.2009","03.06.2010",...]}
```

對映

XML 綱目不提供對映的原生支援。 有三種通用的方法，可以在 XML 中建立對映模型。 爲了容納不同的方法，我們使用下列語法：

```
"xml.map.specification.<element_name>"="<key>-><value>"
```

其中

element_name
要視爲對映登錄的 XML 元素名稱

key 對映登錄鍵 XML 節點

value 對映登錄值 XML 節點

給定 XML 元素的對映規格應該在 Hive 表格建立 DDL 的 SERDEPROPERTIES 區段下定義。 使用下列語法，可以定義鍵及值：

@attribute
@attribute 規格可讓使用者將屬性的值用作對映的鍵或值。

element
元素名稱可用作鍵或值。

#content
元素的內容可用作鍵或值。 由於對映鍵只能是初始類型，因此複式內容會轉換爲字串。

在 XML 中代表對映的方法及其對應的 Hive DDL 及原始資料如下所示。

元素名稱與內容
元素的名稱會用作鍵，而內容會用作值。 這是其中一個常見的技術，在將 XML 對映至 Hive 對映類型時，依預設會使用該技術。 使用此方法的明顯限制是對映鍵只能是字串類型。

XML 資料

```
<entry1>value1</entry1>
<entry2>value2</entry2>
<entry3>value3</entry3>
```

對映、Hive DDL 及原始資料

在此情況下，您無需指定對映，因爲依預設，元素的名稱會用作鍵，而內容會用作值。

```
result map<string,string>
```

```
{"result":{"entry1": "value1", "entry2": "value2", "entry3": "value3"}}
```

屬性與元素內容
將屬性值用作鍵，將元素內容用作值。

XML 資料

```
<entry name="key1">value1</entry>
<entry name="key2">value2</entry>
<entry name="key3">value3</entry>
```

對映、Hive DDL 及原始資料

```
"xml.map.specification.entry"="@name->#content"
```

```
result map<string,string>
```

```
{"result":{"key1": "value1", "key2": "value2", "key3": "value3"}}
```

屬性與屬性

XML 資料

```
<entry name="key1" value="value1"/>
<entry name="key2" value="value2"/>
<entry name="key3" value="value3"/>
```

對映、Hive DDL 及原始資料

```
"xml.map.specification.entry"="@name->@value"
```

```
result map<string,string>
{"result":{"key1": "value1", "key2": "value2", "key3": "value3"}}
```

複式內容

透過新增稱爲 <string> 的根元素，用作初始類型的複式內容會轉換爲有效的 XML 字串。 假設有下列 XML：

```
<dataset>
 <value>10</value>
 <value>20</value>
 <value>30</value>
</dataset>
```

XPath 表示式 /dataset/* 會導致傳回大量 <value> XML 節點。 如果目標欄位爲初始類型，則實作會透過新增 <string> 根節點，將查詢的結果轉換爲有效的 XML。

```
<string>
 <value>10</value>
 <value>20</value>
 <value>30</value>
</string>
```

註：如果查詢的結果是單一 XML 元素，則實作不會新增根元素 <string>。

文字內容

會忽略 XML 元素的僅含空格文字內容。

預覽與 meta 資料（資料來源）

按一下**預覽與 meta 資料**可顯示記錄的樣本及資料來源的資料模型。 在這裡，您有機會檢閱基本 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」標籤會顯示基本欄位 meta 資料。 對於具有檔案內容類型的資料來源，資料模型從一小段記錄樣本產生，您可以在此標籤上，手動編輯欄位 meta 資料。 對於具有 HCatalog 內容類型的資料來源，資料模型基於 HCatalog 欄位對映產生，您無法在此標籤上編輯欄位儲存。

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」值指定將爲欄位的每個可能值建置個別模型。頻率會指定欄位值應該用作每一筆記錄的頻率加權。 記錄 ID 用於識別輸出中的記錄。

儲存 儲存說明將資料儲存在欄位中的方式。 例如，值爲 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，您可能要將值爲 1 和 0 之整數欄位的度量層次設爲「旗標」。這通常表示 1 = True，而 0 = False。

值 顯示具有分類度量之欄位的個別值，或者具有連續度量之欄位的值範圍。

結構 指出欄位中的記錄是包含單一值（初始值）還是值清單。

深度 指出清單的深度；0 是初始值的清單，1 是清單的清單，以此類推。

掃描所有資料值

這可讓您起始及取消資料來源資料值的掃描，以判定種類值及範圍限制。 如果掃描正在進行，請按一下按鈕，以**取消資料掃描**。 掃描所有資料值可確保 meta 資料正確，但是如果資料來源具有許多欄位及記錄，則可能會花費一些時間。

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。

專案清單

主要「專案」頁面提供現行使用者屬於其成員的專案清單。

- 按一下專案名稱，以顯示其詳細資料，並編輯其內容。
- 在搜尋區中鍵入內容可過濾清單，以僅顯示名稱中含有搜尋字串的專案。
- 按一下**新建**，以使用您在**新增專案**對話框中指定的名稱建立新專案。 請參閱第 25 頁的『命名規則』，以取得您可以爲專案提供之名稱的限制。
- 按一下**刪除**，以移除選取的專案。 此動作會從 HDFS 移除專案，並刪除與該專案相關聯的所有資料。
- 按一下**重新整理**，以更新清單。

個別專案詳細資料

內容區劃分爲**詳細資料**、**共用**、**檔案**和**版本**可收合區段。

詳細資料

名稱 一個可編輯的文字欄位，顯示專案的名稱。

顯示名稱

一個可編輯的文字欄位，如在其他應用程式中顯示的那樣，顯示專案的名稱。 如果該欄位爲空白，則「名稱」會用作顯示名稱。

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。

要保留的版本數

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。 預設值爲 25。

註：清除處理程序不會立即執行，但會每隔 20 分鐘，在背景執行一次。

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爲成員（清除）。

按一下**儲存**，以儲存設定的現行狀態。

共用 您可以透過將使用者和群組新增爲作者或檢視者，來共用專案。

- 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。 選取共用的層次，並按一下**新增成員**，以新增至成員清單。
  - 作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。 當透過 IBM® SPSS® Modeler 連接至 Analytic Server 時，這些使用者及這些群組的成員具有對這個專案的寫入（Analytic Server「匯出」節點）權。
  - 檢視者可以查看專案中的資料夾和檔案，並在專案的物件之上定義資料來源，但無法修改專案。
- 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**移除成員**。

註：管理者不管是否明確地列出爲成員，都對每個專案具有讀寫存取權。

註：對「共用」進行的變更會立即自動套用。

檔案

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。

檔案檢視器

顯示現行專案的資料夾結構。 資料夾結構僅在已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。 若要建立或刪除專案，請回到「專案」清單。

- 按一下**將檔案上傳到 HDFS**，以將檔案上傳至現行專案/子資料夾。
- 按一下**建立新資料夾**，以在現行資料夾下，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**，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。

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，並讓此版本成爲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爲最近確定的版本。
- 按一下**刪除**以移除選取的版本。

使用者管理

管理者可以透過「使用者」頁面，管理使用者和群組的角色。

內容區劃分爲**詳細資料**及**主體**可收合區段。

詳細資料

名稱 一個不可編輯的文字欄位，顯示租戶的名稱。

說明 一個可編輯的文字欄位，容許您提供關於租戶的解釋性文字。

URL 此 URL 提供給使用者，以透過 Analytic Server 主控台登入至租戶。

主體

主體是從配置期間設定之安全提供者處得來的使用者和群組。 您可以將主體的角色變更爲「管理者」或「使用者」。

度量值 可讓您配置租戶的資源限制。 報告租戶目前使用的磁碟空間。

- 您可以設定租戶的磁碟空間配額上限；當達到此限制時，無法將更多資料寫入至此租戶上的磁碟，除非清除足夠的磁碟空間，讓租戶磁碟空間用量低於配額。
- 您可以設定租戶的磁碟空間警告層次；當超出該配額時，此租戶上的主體無法提交任何分析工作，除非清除足夠的磁碟空間，讓租戶磁碟空間用量低於配額。
- 您可以設定此租戶上於單一時間可以執行的平行工作數目上限；當超出該配額時，此租戶上的主體無法提交任何分析工作，除非目前執行中的工作完成。
- 您可以設定資料來源可以具有的欄位數目上限。 每當建立或更新資料來源時，會檢查該限制。
- 您可以設定資料來源可以具有的記錄數目上限。 每當建立或更新資料來源時，會檢查該限制；例如，當您新增檔案或變更檔案設定時。
- 您可以設定檔案大小上限 (MB)。 上傳檔案時，會檢查該限制。

命名規則

對於 Analytic Server 中可以給定唯一名稱的任何項目（如資料來源及專案），下列規則適用於那些名稱。

- 相同類型的物件中，名稱必須唯一。 例如，兩個資料來源無法同時命名爲 insuranceClaims，但一個資料來源及一個專案可以分別命名爲 insuranceClaims。
- 名稱區分大小寫。 例如，insuranceClaims 與 InsuranceClaims 會視爲唯一名稱。
- 名稱忽略前導及尾端空格。
- 下列字元在名稱中無效。

  ~, #, %, &, *, {, }, \\, :, <, >, ?, /, |, ", \t, \r, \n

第 3 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。

爲了構建可依據 Analytic Server 資料來源執行的 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 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。

請參閱 SPSS Modeler 說明文件，以取得這些節點的詳細資料。

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。
- 爲了讓 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。
- 建議您將連續目標的儲存體設定爲實數，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，串流 TS 及 Space-Time-Boxes 節點除外。 後面說明受支援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。

取樣

- 不支援區塊層取樣。
- 不支援複式取樣方法。

聚集

- 不支援連續金鑰。 如果您正在重複使用的現有串流設定以排序資料，然後在「聚集」節點中使用此設定，請變更串流以移除「排序」節點。
- 順序統計資料（中位數、第一四分位數、第三四分位數）會近似計算，並透過「最佳化」標籤進行支援。

排序

- 不支援「最佳化」標籤。

在分散式環境中，存在有限數目的作業，可保留「排序」節點所建立的記錄順序。

- 後接「匯出」節點的「排序」會產生已排序的資料來源。
- 後接具有**第一筆**記錄取樣之「樣本」節點的「排序」會傳回第 *N* 筆記錄。
- 後接具有**針對非常大型資料集最佳化**目標（神經網絡、線性、C&R 樹狀結構、搜尋或 CHAID）之建模節點的「排序」對於使用以下方法，隨機調整記錄而言，是非常有用的型樣：在衍生的亂數鍵上排序，以免在對原始記錄進行排序時，產生可能會引入模型建置演算法的偏差。

一般而言，您應該盡可能近地，將「排序」節點放置在需要排序記錄之作業的旁邊。

合併

- 不支援「依順序合併」。
- 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。 一般可以指定「選取」節點來取代「取樣」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。
- 合併作業相對緩慢。 如果您在 HDFS 中有可用空間，則可以更快地一次合併資料來源，並在下列串流中使用合併的來源，而不是在每一個串流中合併資料來源。

R 轉換

節點中的 R 語法應該包含一次一筆記錄作業。

欄位作業

支援所有欄位作業，移轉、時間間隔及歷程節點除外。 後面說明受支援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。 支援將已訓練的「自動資料準備」節點中的轉換套用至新資料。

類型

- 不支援「檢查」直欄。
- 不支援「格式化」標籤。

衍生

- 支援所有「衍生」功能，順序功能除外。
- 不能在使用分割欄位作爲分割項的相同串流中衍生分割欄位；您將需要建立兩個串流；一個用於衍生分割欄位，另一個使用欄位作爲分割項。
- 在進行比較時，旗標欄位不能由其本身使用；即，if (flagField) then ... endif 將導致錯誤；暫行解決方法是使用 if (flagField=trueValue) then ... endif
- 在使用 ** 運算子指定指數爲實數（例如，x**2.0，而不是 x**2），以符合 Modeler 中的結果時建議這麼做

填充值

- 支援衍生節點所支援的相同功能集。

進倉 不支援下列功能。

- 最佳進倉
- 等級
- 並排 -> 並排：值的總和
- 並排 -> 連結空間：保持現行並隨機指派

- 並排 -> 自訂 N：超過 100 的值，以及 100 % N 不等於零的任何 N 值。

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 支援下列「建模」節點：線性、神經網絡、C&RT、Chaid、搜尋、TCM、TwoStep-AS、STP 及關聯規則。 後面說明這些節點功能的進一步注意事項。

線性 當在大資料上建置模型時，您通常想要將目標變更爲非常大型的資料集，或者指定分割。

- 不支援繼續訓練現有的 PSM 模型。
- 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。
- 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于自動資料準備。 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絡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。

類神經網絡

當在大資料上建置模型時，您通常想要將目標變更爲非常大型的資料集，或者指定分割。

- 不支援繼續訓練現有的標準或 PSM 模型。
- 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。
- 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

當在大資料上建置模型時，您通常想要將目標變更爲非常大型的資料集，或者指定分割。

- 不支援繼續訓練現有的 PSM 模型。
- 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。
- 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

支援用於建模的所有模型還支援用於評分。 此外，針對下列節點本端建置的模型塊也支援用於評分：C&RT、搜尋、CHAID、線性及神經網絡（無論模型是標準的、Boosted Bagged，還是用於非常大型資料集）、回歸、C5.0、Logistic、Genlin、GLMM、Cox、SVM、Bayes Net、TwoStep、KNN、決策清單、區別元件、自我學習、異常偵測、Apriori、Carma、K-Means、Kohonen、R 與文字採礦。

- 將不對原始或已調整的習性評分。 作爲暫行解決方法，您可以手動使用下列表示式透過「衍生」節點來計算原始習性，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

R 塊中的 R 語法應該包含一次一筆記錄作業。

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

透過撰寫包含上游作業結果的暫用 Analytic Server 資料來源，支援「表格」節點。 「表格」節點隨後會翻看該資料來源的內容。

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。

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。 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 認爲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。

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 爲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。

Printed in Taiwan