

**IBM SPSS Analytic Server
3.0 版**

管理者手冊

IBM

附註

在使用本資訊及它支援的產品之前，請閱讀第 11 頁的『注意事項』中的資訊。

產品資訊

此版本適用於 IBM SPSS Analytic Server 3.0.0.0 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 承租人管理 | 1 | 版本資訊 | 7 |
| 命名規則 | 2 | 日誌收集器 | 7 |
| 第 2 章 使用者入門 | 3 | 常見問題 | 7 |
| 第 3 章 Analytic Server 工作名稱 | 5 | 效能調整 | 8 |
| 第 4 章 疑難排解 | 7 | 注意事項 | 11 |
| 記載 | 7 | 商標 | 12 |

第 1 章 承租人管理

承租人提供使用者、專案及資料來源的高層次分割，從而無法在承租人之間共用物件。每一個使用者都可存取其指派之承租人環境定義中的系統。

您可以在 **Analytic Server** 主控台中，管理承租人，以及將使用者指派給承租人。「承租人」頁面的視圖取決於登入至主控台之使用者的角色：

- 安裝期間設定的「超級使用者」管理者是承租人管理員。只有此使用者可以建立新的承租人，以及編輯任何承租人的內容。
- 具有「管理者」角色的使用者可以編輯他們登入之承租人的內容。
- 具有「使用者」角色的使用者無法編輯承租人內容。「承租人」頁面對他們隱藏。
- 具有讀者角色的使用者不能編輯資料來源，甚至無法登入 **Analytic Server** 主控台。

管理者可以存取「專案」及「資料來源」頁面，以及管理任何專案或資料來源以進行清理及管理。如需相關資訊，請參閱 *IBM® SPSS® Analytic Server 使用手冊*。

承租人清單

主要「承租人」頁面在表格中顯示現有承租人。只有「超級使用者」管理者能夠在此頁面上進行編輯。

- 按一下承租人名稱，以顯示其詳細資料，並編輯其內容。
- 按一下承租人 URL，以在該承租人的環境定義中開啓主控台。

註：您將登出主控台，並將需要使用承租人的有效認證來登入。

- 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承租人。
- 按一下**新建**，以使用您在**新增承租人**對話框中指定的名稱建立新承租人。請參閱第 2 頁的『命名規則』，以取得您可以為承租人提供之名稱的限制。
- 按一下**刪除**，以移除選取的承租人。
- 按一下**重新整理**，以更新清單。

個別承租人詳細資料

內容區劃分為數個可收合區段。

詳細資料

| | |
|------------|---|
| 名稱 | 一個可編輯的文字欄位，顯示承租人的名稱。 |
| 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 主控台登入承租人，以及用來配置 SPSS Modeler 伺服器。如需配置 SPSS Modeler 的詳細資料，請參閱 <i>IBM SPSS Analytic Server Installation and Configuration Guide</i> 。 |
| 狀態 | 作用中 承租人目前正在使用中。讓承租人處於 非作用中Inactive 會阻止使用者登入該承租人，但不會刪除任何基礎資訊。 |

主體

主體是從安裝期間設定之安全提供者處得來的使用者和群組。您可以主體新增至承租人以作為「管理者」或「使用者」。

- 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。從下拉清單中選取**管理者**、**使用者**或**讀者**，以指派其在租戶內的角色。按一下**新增參與者**，將他們新增至作者清單。
- 若要移除參與者，請在成員清單中選取使用者或群組，然後按一下**移除參與者**。

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。
- 您可以設定檔案大小上限 (MB)。上傳檔案時，會檢查該限制。

安全提供者配置

可讓您指定使用者鑑別提供者。**預設值**使用在安裝及配置期間設定的預設承租人提供者。**LDAP** 可讓您向外部 LDAP 伺服器（例如 Active Directory 或 OpenLDAP）鑑別使用者。指定提供者的設定，並選擇性地指定過濾器設定，以控制「主體」區段中可用的使用者和群組。

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。

- 在單個租戶內，名稱在相同類型的物件內必須是唯一的。例如，兩個資料來源無法同時命名為 insuranceClaims，但一個資料來源及一個專案可以分別命名為 insuranceClaims。
- 名稱區分大小寫。例如，insuranceClaims 與 InsuranceClaims 會視為唯一名稱。
- 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 2 章 使用者入門

告知使用者導覽至 `http://<host>:<port>/<context-root>/admin/<tenant>`，並輸入其使用者名稱與密碼，從而登入至 Analytic Server 主控台。

<host>

Analytic Server 主機的位址。

<port>

Analytic Server 接聽的埠。依預設，這是 9080。

<context-root>

Analytic Server 的環境定義根目錄。依預設，這是分析伺服器。

<tenant>

在多租戶環境中，表示您所屬的租戶。在單一租戶環境中，預設租戶為 **ibm**。

例如，如果主機的 IP 位址為 9.86.44.232，您已建立租戶 "mycompany" 並向其新增了使用者，且將其他設定保留為其預設值，則使用者應該導覽至 `http://9.86.44.232:9080/analyticserver/admin/mycompany`，以存取 Analytic Server 主控台。

第 3 章 Analytic Server 工作名稱

Analytic Server 產生對映化簡工作，可以透過 Hadoop 叢集的工作追蹤器進行監視。

對映化簡工作具有下列結構。

AS/{tenant name}/{user name}/{algorithm name}

{tenant name}

這是執行工作所使用的租戶名稱。

{user name}

這是要求工作的使用者。

{algorithm name}

這是工作中的主要演算法。請注意，單一串流可能產生多個對映化簡工作；類似地，串流內的數個作業可以包含在單一對映化簡工作內。

第 4 章 疑難排解

Analytic Server 提供數個有用的工具來協助判斷問題。

記載

Analytic Server 會建立客戶日誌檔和追蹤檔案，對問題診斷會有幫助。藉由預設的 Liberty 安裝，您可以在 {AS_ROOT}/ae_wlpserver/usr/servers/aeserver/logs 目錄中找到日誌檔。

預設記載配置會產生兩個日誌檔（每天輪換）。

as.log 此檔案包含參考資訊警告及錯誤訊息的高層次摘要。當無法利用使用者介面中所顯示錯誤訊息來解決發生的伺服器錯誤時，應該首先檢查該檔案。

as_trace.log

此檔案包含 as.log 中的所有項目，但會新增主要針對 IBM 支援中心及用於開發除錯的其他資訊。

Analytic Server 使用 Apache LOG4J 作為其基礎記載機能。藉由 LOG4J，記載可以透過編輯 {AS_SERVER_ROOT}/configuration/log4j.xml 配置檔進行動態調整。支援中心可能會要求您執行此動作來協助論斷問題，或要求您修改它來限制保留的日誌檔數目。系統會在數秒內自動偵測檔案變更，因此 Analytic Server 不需要重新啟動。

如需 log4j 及配置檔的相關資訊，請參閱官方 Apache 網站 <http://logging.apache.org/log4j/> 的說明文件。

版本資訊

您可以檢查 {AS_ROOT}/properties/version 資料夾，來確定安裝哪個版本的 Analytic Server。下列檔案包含版本資訊。

IBM_SPSS_Analytic_Server-*.swtag

包含詳細產品資訊。

version.txt

已安裝產品的版本和建置號碼。

日誌收集器

當無法直接檢閱日誌檔來解決問題時，您可以組合所有日誌並將其傳送至 IBM 支援中心。提供了公用程式，以便輕鬆收集所有必要的資料。

使用指令 Shell，來執行下列指令：

```
cd {AS_ROOT}/bin
run >sh ./logcollector.sh
```

這些指令會在 {AS_ROOT}/bin 下建立壓縮檔。壓縮檔包含所有日誌檔及產品版本資訊。

常見問題

本節說明了部分常見管理問題，以及如何對其進行修正。

安全

嘗試存取 HCatalog 資料來源時，Kerberos 鑑別失敗

如果您在日誌中看到類似下列錯誤：

```
cause:javax.security.sasl.SaslException: Failure to initialize security context  
com.spss.analyticframework.api.exceptions.ComponentException: Cannot access HCatalog
```

您必須確保快取 HDFS 使用者的 Kerberos TGT，並在 Analytic Server 伺服器主機上可用。若要執行此動作：

1. 停止 Analytic Server 程序。
2. 透過 Analytic Server 主機執行 `kinit -f $hdfs.user`，其中 `$hdfs.user` 如 `config.properties` 檔案中所定義的，且具有分析根目錄的寫入權
3. 啟動 Analytic Server。

Analytic Server 主控台

在 iOS 上透過 Safari 存取 Analytic Server 主控台

租戶狀態下拉清單不工作，且您無法更新檔案型資料來源的資料模型。執行這些動作時，請使用其他瀏覽器。

執行中串流

R 工作將非英文單字轉換為 Unicode

在 Cloudera 叢集上，如果 Hadoop 伺服器的系統編碼不是 UTF-8，則 R 會將非英文單字轉換為 Unicode。

1. 導覽至 Cloudera Manager 主控台內的 YARN 配置標籤。
2. 將下列設定新增至「NodeManager 環境進階配置 Snippet（安全閥）」欄位。

```
LC_ALL=""  
LANG=en_US.utf8
```

PySpark 工作無法執行

1. 在 Ambari 主控台中，將 `export SPARK_HOME=/usr/iop/current/spark-client`（適用於 BigInsights）或 `SPARK_HOME=/usr/hdp/current/spark-client`（適用於 Hortonworks）新增至 YARN 服務的 `yarn-env` 參數；在 Cloudera Management 中，將 `SPARK_HOME=/opt/cloudera/parcels/CDH/lib/spark` 新增至 YARN 服務 YARN（包括 MR2）服務環境進階配置 Snippet（安全閥）。
2. 確保 Spark 服務部署在所有叢集節點上。
3. 重新啟動 YARN 與 Analytic Server 服務。

效能調整

本節說明最佳化系統效能的方法。

Analytic Server 是 Ambari 架構中的一個元件，該元件會利用其他元件，例如 HDFS、Yarn 及 Spark。Hadoop、HDFS 及 Spark 的一般效能調整技術適用於 Analytic Server 工作量。每個 Analytic Server 工作量都有所不同，因此，需要根據您的特定部署工作量進行調整實驗。下列內容及調整提示是主要變更，已影響 Analytic Server 評比及調整大小測試的結果。

當第一個工作在 Analytic Server 上執行時，伺服器將啟動持續性 Spark 應用程式，該應用程式將處於作用中，直到 Analytic Server 關閉為止。在 Analytic Server 執行期間，持續性 Spark 應用程式將在所有配置給它的叢集資源上進行配置並予以保留，即使 Analytic Server 工作未在積極地執行中。應該小心計算配置給 Ana-

lytic Server Spark 應用程式的資源量。如果所有叢集資源都配置給 Analytic Server Spark 應用程式，則其他工作可能會延遲或未執行。這些工作可以排入佇列中等待足夠的可用資源，並且 Analytic Server Spark 應用程式將耗用那些資源。

如果配置並部署多個 Analytic Server 服務，則每一個服務實例都可以潛在配置其自己的持續性 Spark 應用程式。例如，如果部署兩個 Analytic Server 服務以支援高可用性失效接手，則您可以看到兩個持續性 Spark 應用程式在作用中，每一個都配置叢集資源。

其他複雜性是在某些狀況下，Analytic Server 可能啟動需要叢集資源的對映減少工作。這些對映減少工作將需要未配置給 Spark 應用程式的資源。需要對映減少工作的特定元件是 PSM 模型建置。

可以配置下列內容，以將資源配置給 Spark 應用程式。如果在 Spark 安裝的 `spark-defaults.conf` 中設定它們，則可以為環境中執行的所有 Spark 工作配置它們。如果在 Analytic Server 配置中將它們設為「自訂 `analytic.cfg`」區段下的自訂內容，則只會為 Analytic Server Spark 應用程式配置它們。

spark.executor.memory

每個執行程式處理程序使用的記憶體數量。

spark.executor.instances

要啟動的執行程式處理程序數目。

spark.executor.cores

每個執行程式處理程序的執行程式工作程式執行緒數目。此值應該介於 1 和 5 之間。

設定三個主要 Spark 內容的範例。HDFS 叢集中有 10 個資料節點，每一個資料節點有 24 個邏輯核心和 48 GB 記憶體，並且僅執行 HDFS 處理程序。這裡有一個配置此環境內容的方法，假設您僅在此環境上執行 Analytic Server 工作，並且想要對單一 Analytic Server Spark 應用程式配置上限。

- 設定 `spark.executor.instances=20`。這將嘗試在每個資料節點上執行 2 個 Spark 執行程式處理程序。
- 設定 `spark.executor.memory=22G`。這會將每一個 Spark 執行程式處理程序的資料堆大小上限設為 22 GB，並在每一個資料節點上配置 44 GB。其他 JVM 及作業系統需要額外記憶體。
- 設定 `spark.executor.cores=5`。這將為每一個 Spark 執行程式提供 5 個工作程式執行緒，並為每個資料節點提供總計 10 個工作程式執行緒。

監視用來執行工作的 Spark 使用者介面

如果您看到可能影響效能的磁碟「溢出」。部分可能的解決方案為：

- 透過 `spark.executor.memory` 提高記憶體並將它配置給 Spark 執行程式。
- 減少 `spark.executor.cores` 的數目。這將減少配置記憶體的並行工作執行緒的數目，但是它還會減少工作的平行化數量。
- 變更 Spark 記憶體內容。Spark 之 Spark 執行程式資料堆的 `spark.shuffle.memoryFraction` 與 `spark.storage.memoryFraction` 配置百分比。

確保相同的節點具有足夠的記憶體

如果 HDFS 中的區塊數目較大並在不斷增長，請確保您指定的節點資料堆增加以容納此增長。這是一般 HDFS 調整建議。

變更用於快取的記憶體數量

依預設，`spark.storage.memoryFraction` 的值為 0.6。當資料的 HDFS 區塊大小為 64MB 時，此可以增加達 0.8。如果輸入資料的 HDFS 區塊大小大於 64MB，則僅當每個作業配置的記憶體大於 2GB 時，才可以增加此值。

調整模型評分的效能

您可以透過使用下列步驟，在具有 Apache Spark 引擎的海量資料集上，改進模型評分工作的效能。請注意，這些步驟不得影響叢集上非 Analytic Server 服務的作業。

1. 檢查 `libtcmalloc_minimal.so{.version}` 是否已安裝在叢集中的每一個節點上。
`whereis libtcmalloc_minimal.so.*`
2. 如果未安裝 `libtcmalloc_minimal.so`，請在叢集中的每一個節點上安裝包含 `libtcmalloc_minimal` 程式庫的作業系統特定套件，或者手動建置並安裝 `libtcmalloc_minimal`。例如：

Ubuntu：

```
sudo apt-get install libgoogle-perftools-dev
```

Red Hat Enterprise Linux 6.x (x64)：

- a. 為 RedHat 安裝 EPEL 儲存庫（如果尚未安裝）

```
wget http://dl.fedoraproject.org/pub/epel/6/x86_64/epel-release-6-8.noarch.rpm
sudo rpm -Uvh epel-release-6*.rpm
```

- b. `sudo yum install gperftools-libs.x86_64`

手動建置：

- a. 從鏈結 <https://github.com/gperftools/gperftools/releases> 中下載 `gperftools-2.4.tar.gz`
- b. `tar zxvf gperftools-2.4.tar.gz`
- c. `cd gperftools-2.4`
- d. `./configure --disable-cpu-profiler --disable-heap-profiler --disable-heap-checker --disable-debugalloc --enable-minimal`
- e. `make`
- f. `sudo make install`

3. 從一個以上節點上執行的下列指令中返回時，請記錄所安裝程式庫檔案 `libtcmalloc_minimal.so{.version}` 的其中一個位置。

```
whereis libtcmalloc_minimal.so.*
```

如果叢集有節點正在執行混合作業系統，則此檔案可能有多個位置。

4. 在 Ambari 主控台中，前往 Analytic Server 配置，並在「自訂 `analytics.cfg`」區段下，將程式庫的位置用作值，來配置主要 `spark.executorEnv.LD_PRELOAD`。進行此變更後，重新啟動 Analytic Server 服務。例如，如果程式庫安裝至 `/usr/lib64/libtcmalloc_minimal.so.4`，則配置將是：

```
spark.executorEnv.LD_PRELOAD=/usr/lib64/libtcmalloc_minimal.so.4
```

如果需要多個位置，請使用空格來區隔它們，如下列範例中所示。

```
spark.executorEnv.LD_PRELOAD=/usr/lib64/libtcmalloc_minimal.so.4 /usr/lib/libtcmalloc_minimal.so
```

如果有任何節點未在其中一個所配置的位置上安裝 `libtcmalloc_minimal.so` 程式庫，這不會造成錯誤，但是這些節點上模型評分的效能可能更為緩慢。

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。為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。所有這些名稱全為虛構，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。為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。所有這些名稱全為虛構，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，網址為 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。

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