

**IBM SPSS Analytic Server
3.2.1 版**

概觀

IBM

附註

在使用本資訊及它支援的產品之前，請閱讀第 5 頁的『注意事項』中的資訊。

產品資訊

此版本適用於 IBM SPSS Analytic Server 3.2.1 版及所有後續版本與修訂，但新版本另有說明者不在此限。

目錄

概觀	1	注意事項	5
架構	2	商標	6
Spark 與 Analytic Server	2		
3.2.1 版的新增功能	3		

概觀

IBM® SPSS® Analytic Server 是用於海量資料分析的解決方案，它結合 IBM SPSS 技術與海量資料系統，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。

海量資料分析為何重要

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，需要海量資料分析：

- 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，海量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海量資料 (這些演算法預期資料送入其中，但是海量資料移動成本太高)，因此
- 海量資料的圖形分析執行狀態需要海量資料系統的新技能，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，但是不適合真正的海量資料。

解決方案

Analytic Server 提供：

- 以資料為中心的架構，利用海量資料系統，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，隱藏海量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

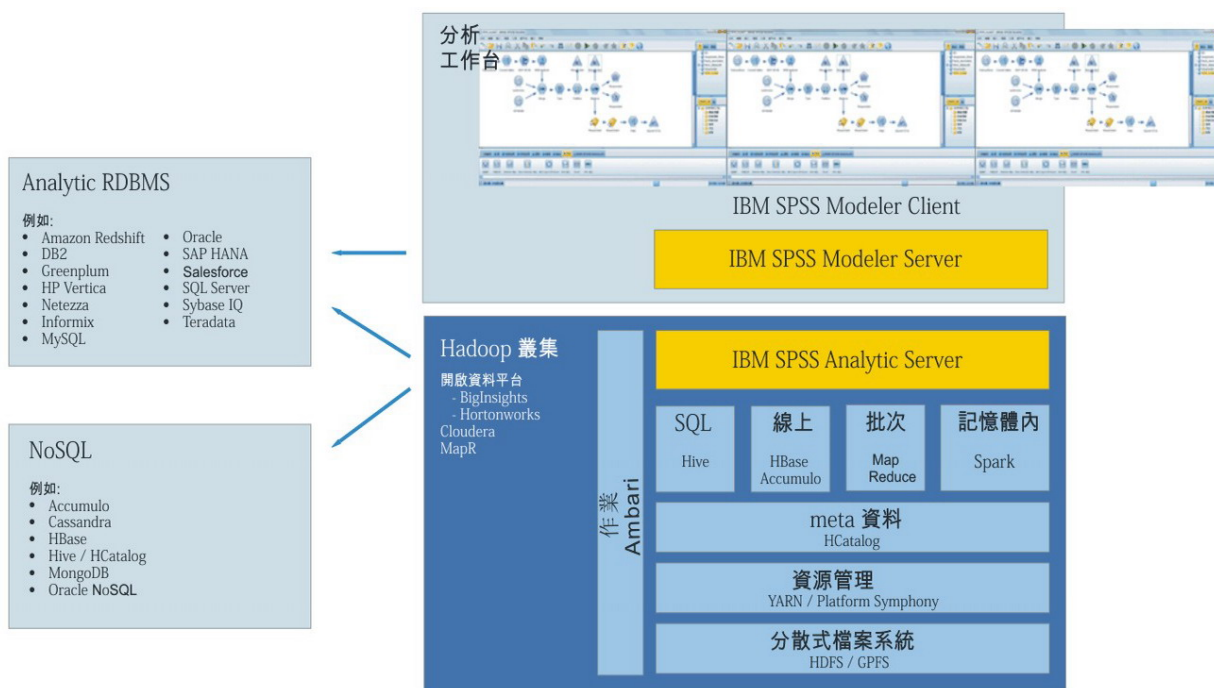


圖 1. 架構

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。
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。

Spark 與 Analytic Server

Analytic Server 與 Apache Spark 整合，以提高效能。

何時使用及不使用 Spark

如果 Spark 作為 Ambari 服務安裝在 Hadoop 叢集上，則 Analytic Server 使用它來處理海量資料工作。下列準則適用於判定何時不使用 Spark。

1. 如果資料集小於 128MB，則 Analytic Server 在 Analytic Server JVM 中使用內嵌的 MapReduce 功能，並且不會利用 Spark 或 Hadoop 叢集。
2. 如果 Spark 未安裝在叢集上，則 Analytic Server 會使用 MapReduce 第 2 版。
3. Analytic Server 使用 MapReduce 第 2 版來建置 PSM 模型。當工作以 PSM 模型建置結束時，Analytic Server 使用 Spark 來處理工作，完成用來建置模型的所有步驟，然後寫入至磁碟，並使用 MapReduce

來建置 PSM 模型。例如，如果工作包括一個結合，後面是 PSM 模型建置，則結合在 Spark 中執行，並且 PSM 在 MapReduce 中的已結合資料上執行。

如何使用 Spark

當 Analytic Server 服務啟動，並探索到 Spark 可用時，它會起始設定「Spark Hadoop 工作」，允許跨叢集在已配送的作業之間進行通訊。只要 Analytic Server 服務執行，此工作就會執行，並且用於所有 Analytic Server 執行。此方法會改進與編排多個 MapReduce Hadoop 工作相關的效能，原因是它會刪除重新載入每一個「Hadoop 工作」之所有 Analytic Server 元件的超載。

Spark 能夠執行 MapReduce 工作。這容許 Analytic Server 使用「原生」Spark 演算法，例如結合、排序及聯集（可用時）。同時，Analytic Server 可以在 Spark 中執行現有 SPSS 的「對映」及「減少」演算法，而無需直接使用 Hadoop API。

3.2.1 版的新增功能

3.2.1 版

平台 支援 Hadoop Data Platform (HDP) 3.0 及 3.1。

支援 Cloudera 6.0 及 6.1。

Rank 功能

rank 功能用來將輸入資料集分割成單獨的分割區並生成一個顯示每個分割區列分級的新欄位。該功能與 Hive 功能類似 **rank()**、**dense_rank()** 及 **row_number()**。

UDF Hive 撤回

已建立新的 Hive UDF 功能。將 Hive UDF 登錄至 HiveDB 之後，Analytic Server 可以使用新的 UDF 功能以執行撤回。

如需最新的系統需求資訊，請使用 IBM 技術支援網站上的詳細系統需求報告：<http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>。在此頁面上：

1. 輸入 SPSS Analytic Server 作為產品名稱，然後按一下搜尋。
2. 選取報告的所需版本及範圍，然後按一下提交。

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。IBM 可能以其他語言提供本資訊。不過，您可能需要擁有一份該語言的產品或產品版本的副本，才能存取該產品或產品版本。

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 10504-1785
US*

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 10504-1785
US

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。

所述的效能資料及客戶範例僅供示範之用。實際的效能結果可能會因特定的配置及作業狀況而異。

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、品牌及產品名稱。此等名稱皆屬虛構，凡有類似實際個人或企業所用之名稱及地址者，皆屬巧合。

著作權：

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、品牌及產品名稱。此等名稱皆屬虛構，凡有類似實際個人或企業所用之名稱及地址者，皆屬巧合。

這些範例程式或任何衍生著作的每份副本或任何部分，都必須依照下列方式併入著作權聲明：

© IBM 2019. 本程式之若干部分係衍生自 IBM 公司的範例程式。

© Copyright IBM Corp. 1989 - 20019. All rights reserved.

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