

Microsoft Data Platform Strategy

微軟商業智慧開啟新視界

Isaac Song 宋明遠

isaacso@Microsoft.com

Technology Manager

Microsoft Taiwan

Microsoft®

新技術帶來新契機亦或新挑戰！？

該怎麼做？



做個有效可運用的分析？



商業智慧 or Big Data?

Microsoft Data Platform Strategy

微軟巨量資料運算開啟新視界

Isaac Song 宋明遠
isaacso@Microsoft.com
Technology Manager
Microsoft Taiwan

Big Data 海量資料的層面



資料量日益增加
(Large)

紙張文件轉換成數位形式，更多的資料需要被收集，並提高現有被採集資料的準確性。



資料種類多元
(Unstructured)

特點是基於文字文件，圖像，音樂和視訊文件，新的數據來源和非標準的資料模型。



資料應用變動快速
(Complex)

透過即時資料收集，快速變化的信息，以及社群網路的發展推波助瀾。

企業面對資料的想法正在改變

以前要不是技術做不到，就是成本太高

巨量資料的
儲存與使用



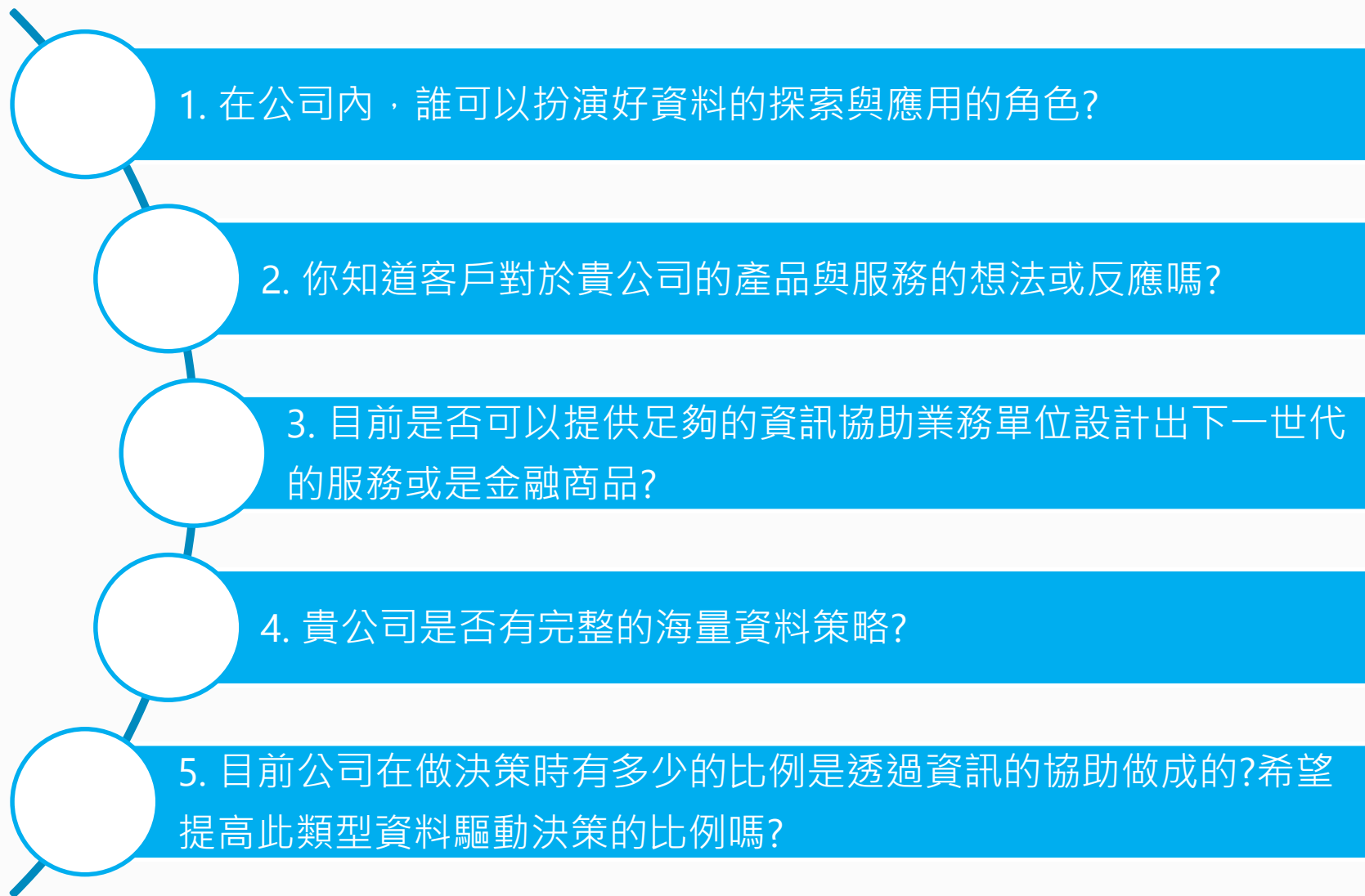
因應新興技術
Hadoop



持續增長的資料
與分析複雜度



海量資料如何開始第一步



資料價值公式 – 用手中的資料賺錢



- 結構化資料
- 非結構化資料
- 持續產生的資料

- **Power BI**
- **Excel Power Pivot, Power View, Power Query 及 Power Map**

- 任何人

- 即時

- 4 年內預估 1.6 兆美金
由資料創造的新營收

微軟海量資料情境體驗

海量資料處理平台



Azure Cloud ML
HDInsight



Microsoft®
SQL Server®

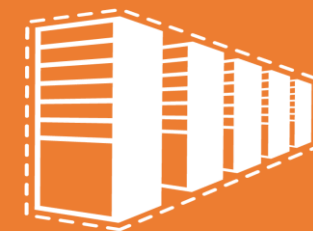
社群媒體
與情分析



銀行轉型
議題：
服務



Machine Learning
產品主動推薦



海量分析視覺化饗宴

微軟海量資料情境體驗

海量資料處理平台



Azure Cloud ML
HDInsight



Microsoft®
SQL Server®

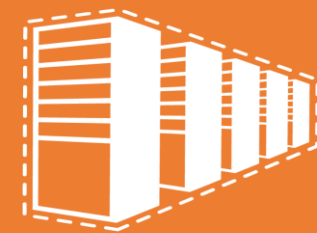
社群媒體
與情分析



政府轉型
議題：
服務



Machine Learning
產品主動推薦



海量分析視覺化饗宴

教育人口地理資訊分析

DEMO



微軟海量資料情境體驗

海量資料處理平台



Azure Cloud ML
HDInsight



Microsoft®
SQL Server®

社群媒體
與情分析



政府轉型
議題：
服務



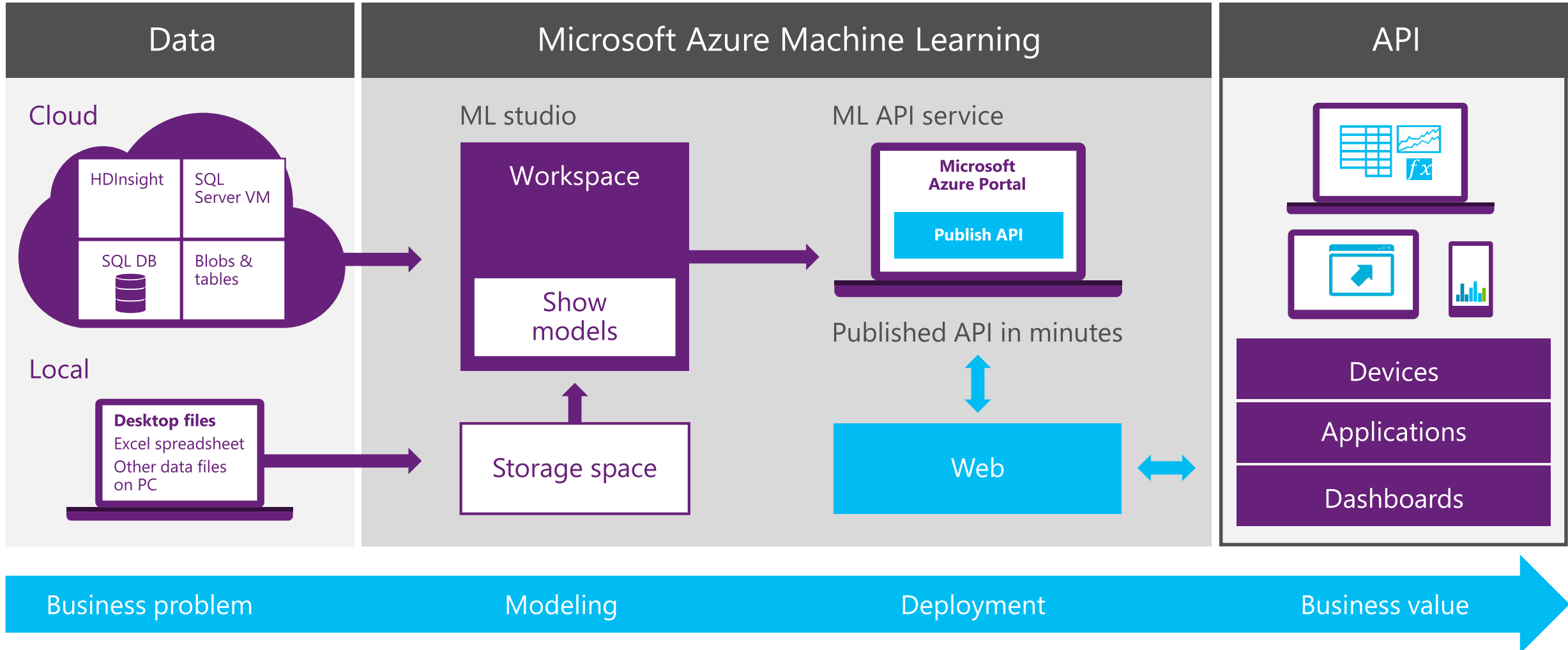
Machine Learning
產品主動推薦



海量分析視覺化饗宴

Cloud - Machine Learning

Microsoft Azure Machine Learning



Azure Cloud ML

基金購買主動推薦

DEMO



微軟金控 歡迎您!

使用者代號

使用者密碼

☐ 記住我的使用者代號

登入

微軟海量資料情境體驗

海量資料處理平台



Azure Cloud ML
HDInsight



Microsoft®
SQL Server®

社群媒體
與情分析



銀行轉型
議題：
服務



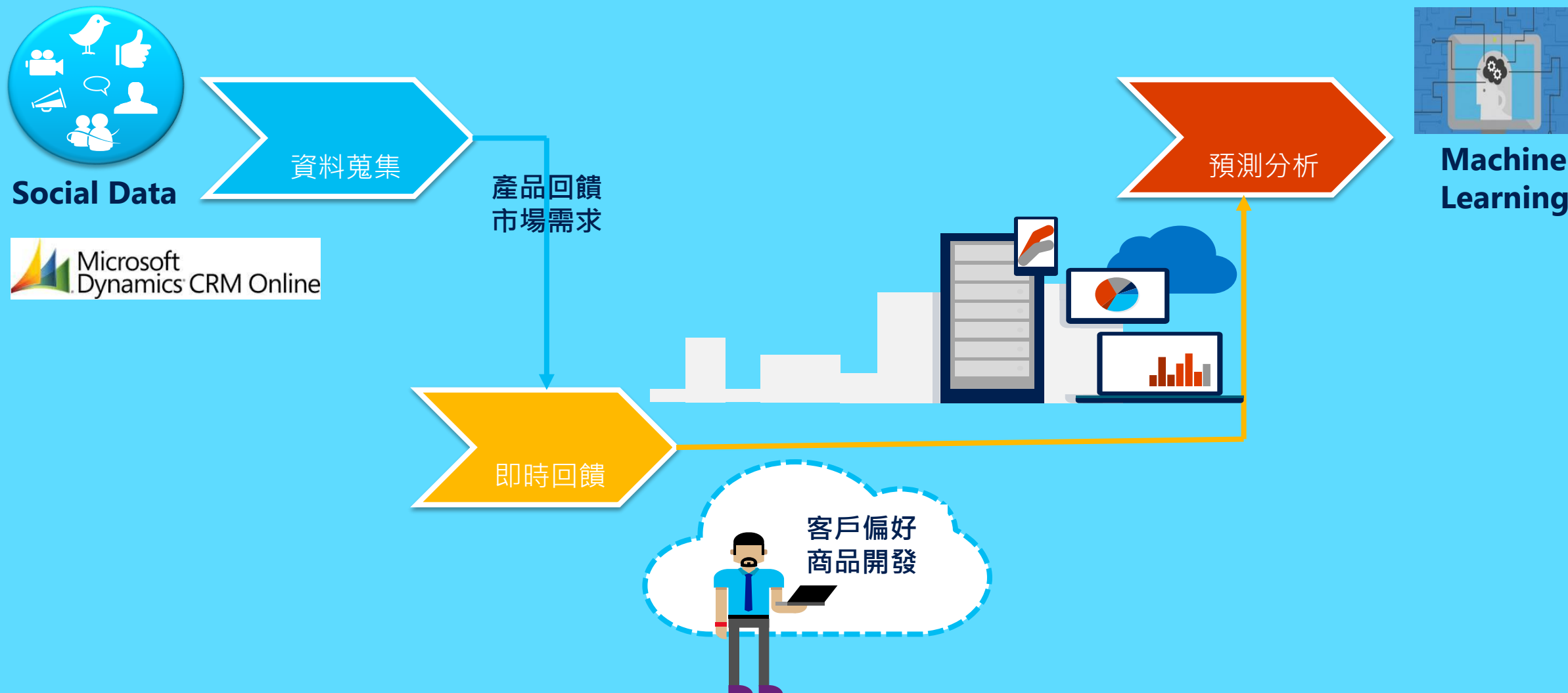
Machine Learning
產品主動推薦



海量分析視覺化饗宴

一呼百應-善用與觀察社群網站的力量

產品生命週期



自然語言處理

後來打給客服要求剪卡，可是他的態度很差

中文斷字

後來(副詞)+打給(動詞)+客服(名詞)+要求(動詞)+剪卡(動詞)+
可是(連結詞)+他的(代名詞)+態度(名詞)+很差(形容詞)

排停用字

打給(動詞)+客服(名詞)+要求(動詞)+剪卡(動詞)+態度(名詞)+
很差(形容詞)

語意解析

客服(名詞)+態度(名詞)+差(形容詞)→負面體驗

分詞邏輯架構

客戶需求

挖掘客戶金融相關
需求

提及產品

擷取客服發言所提
及之產品與品類

正負面體驗

評估與CUBC往來
的正負面使用體驗

正負面情感

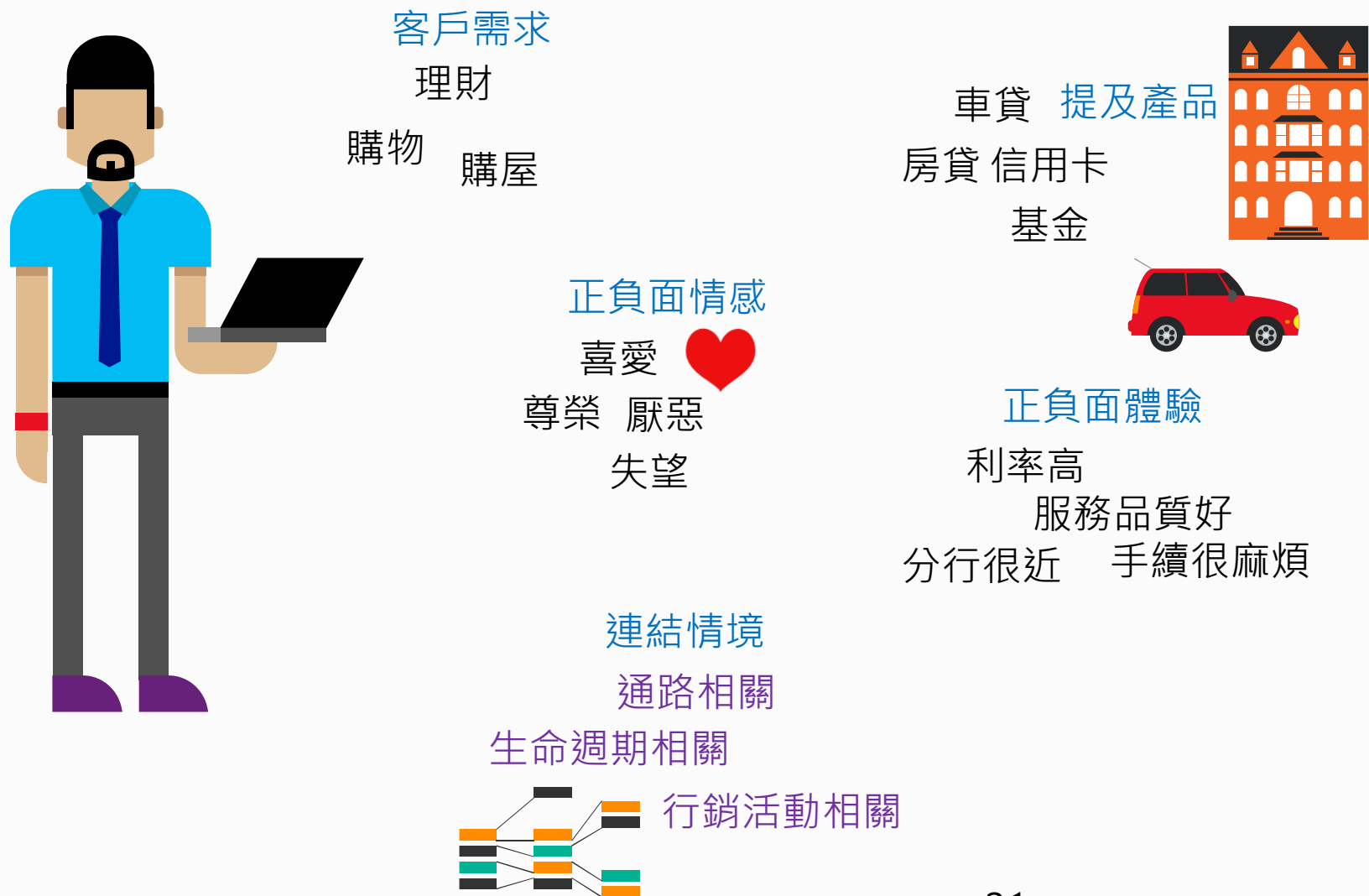
對品牌與產品的正
負面情感

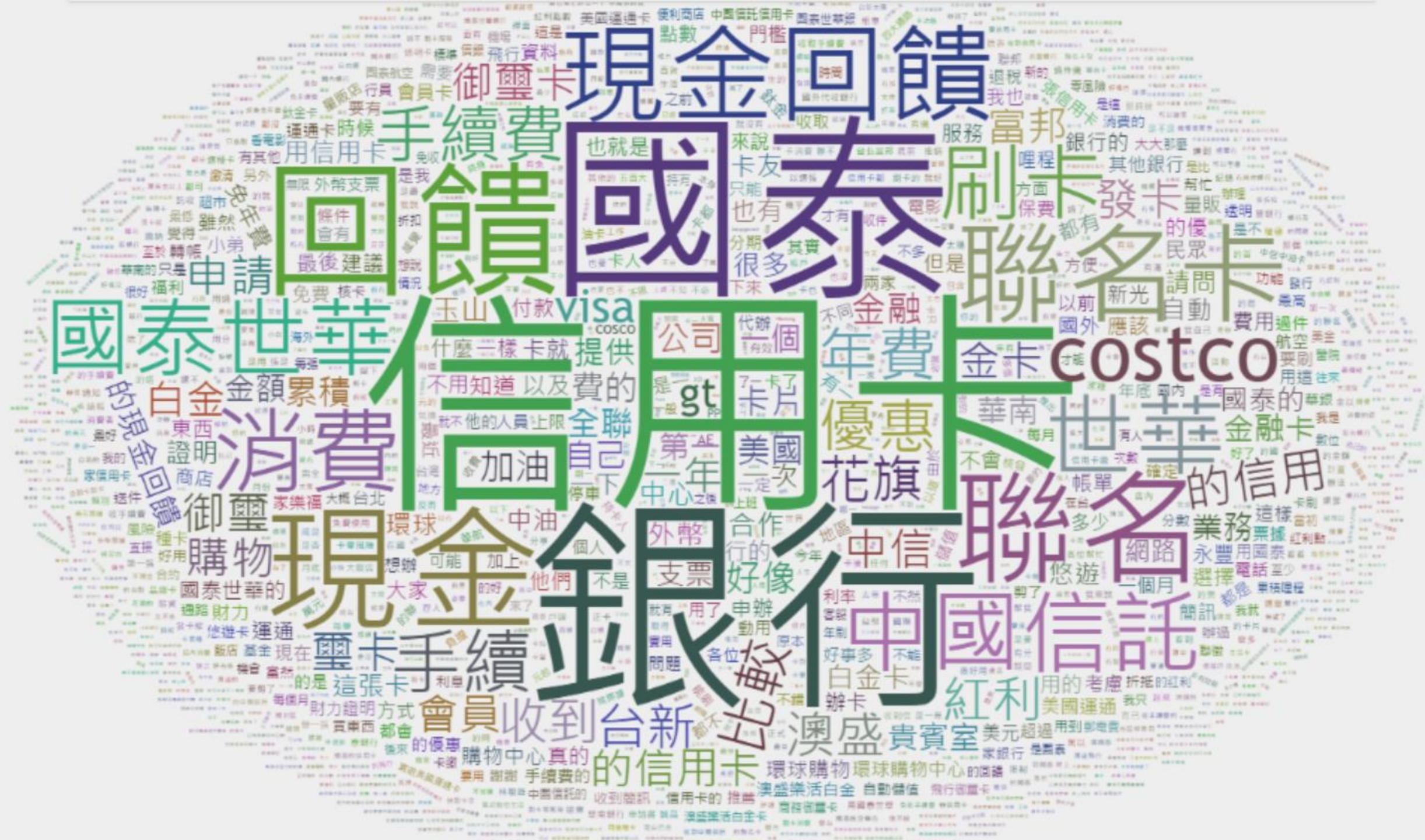
連結情境

挖掘客服來電內容
之提及情境

啟發式分詞

分詞邏輯架構





企業面對資料的想法正在改變

以前要不是技術做不到，就是成本太高

巨量資料的
儲存與使用



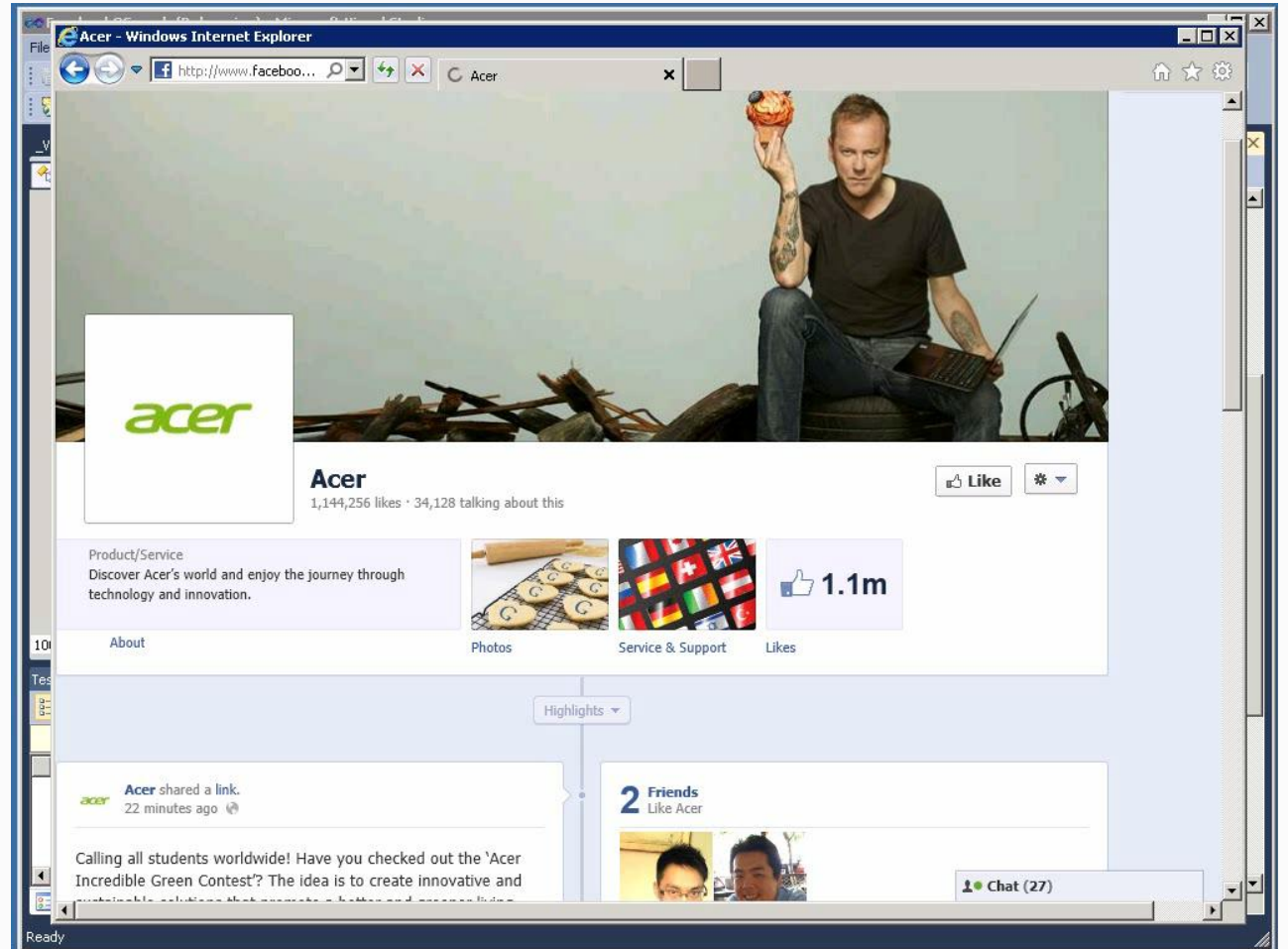
因應新興技術
Hadoop



持續增長的資料
與分析複雜度



Big Data 在社群網路上的應用



Windows (TW) 粉絲各國統計

國家，次數 修改



內政部參考情境

- 透過Facebook可以反映民眾目前在意事項
- 所以可以透過{讚}的數量，來警示或是提醒是否應該針對此議題做出反應機制



全台實價登錄網站 動態時報 最新



全台實價登錄網站分享省地大帝的影片。

7月23日

影片是今天徐世榮教授在路邊喊口號，被警方以「公共危險罪」逮捕的畫面。影片中有人要求警方唸出法條，並說明「公共危險罪」的成立要件。

這段畫面讓我非常的憤怒，因為我想起了與內政部地政司游科長的來往郵件。所以我決定公佈郵件的內容。信件內容未作修改。

謹以此文表達我身為納稅人，對中華民國政府官員的憤怒。本文謹代表我本人發言，不包括其他三位成員。

----- 更多

2013年7月23日 13:49

推擠後警察正要以@'%.,@.#%;%的理由抓走徐世榮教授，1949進行式！

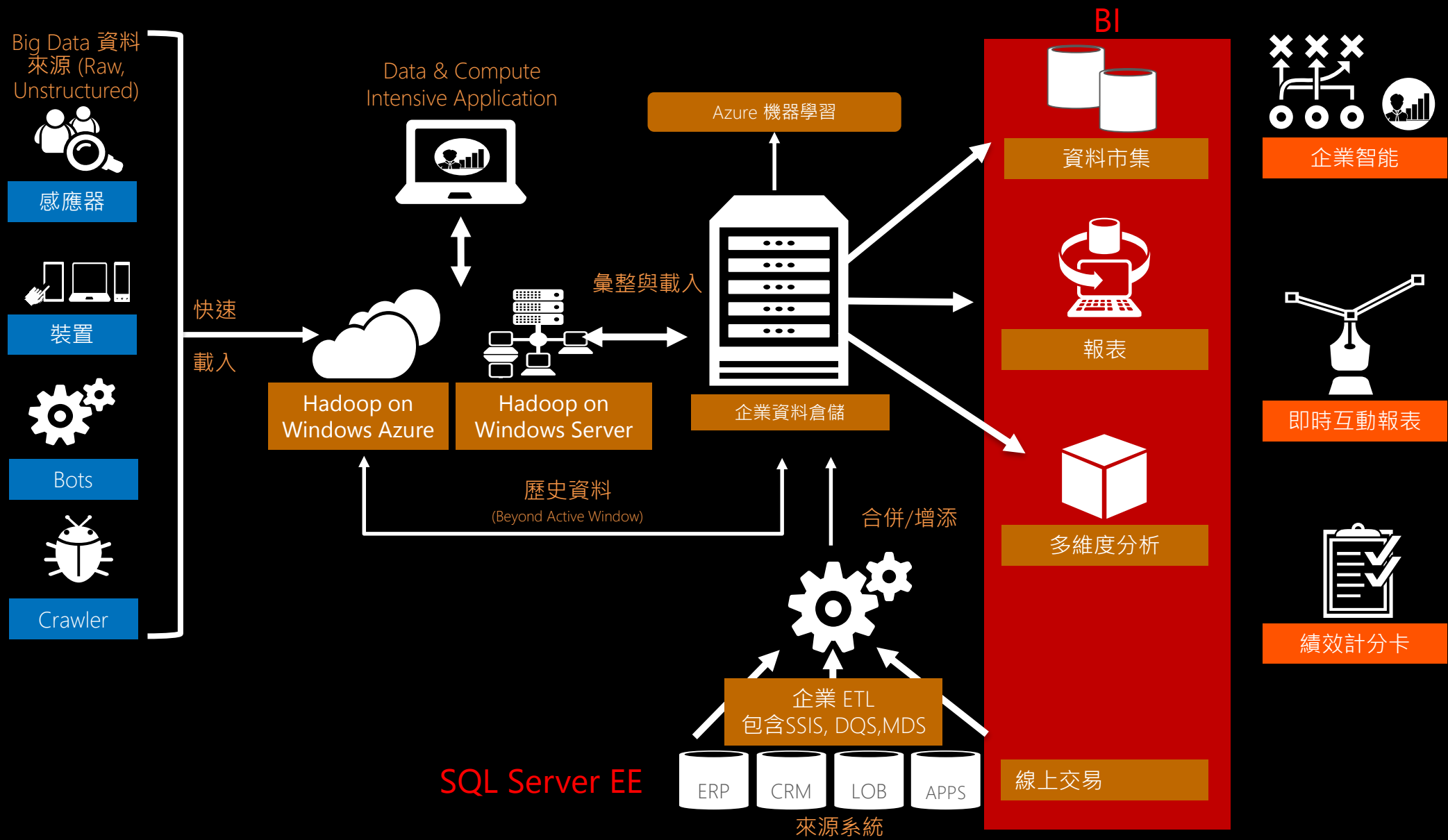


讚 留言 分享

112

561 個人都說讚。

Microsoft Business Intelligence in Big Data



微軟 Big Data 的效益



Breakthrough Insights

- 使用熟悉的工具分析 Big Data 背後的意義 (Excel, PowerPivot, Power View)
- JavaScript 為基礎的簡易程式開發



Broader Access

- 建構在 Windows 或 Azure 上的 Hadoop
- Excel Hive 增益集
- Fast Track 與 PDW



Enterprise Ready

- 整合 Active Directory
- 整合 System Center
- 企業關鍵資料庫
SQL Server 2012



補充資料

微軟海量資料 解決方案的組成元件

Hadoop on Windows Server & Azure

DW: SQL Server Fast Track & PDW

Business Intelligence : Self Service BI

企業級的 Hadoop 建構在 Windows 或 Cloud

支援非結構化的資料

效益



利用 Big Data 延伸
Data Warehouse



企業級 Big Data
平台 on-premises



雲端應用高彈性
Big Data 平台

關鍵特色



Microsoft®
SQL Server®

Hadoop
connectors for
SQL Server



Windows Server®

Hadoop-based
distribution on
Windows Server



Windows Azure™

Hadoop-based
Service on
Windows Azure

Hadoop on Azure- JavaScript Console

The screenshot shows a web browser window with the URL <https://www.hadooponazure.com/Account>. The page header includes the Windows Azure logo and the text "Apache™ Hadoop™-based Services for Windows Azure". The user is logged in as "HoRich" and can click "Sign Out".

The main heading asks, "Would a cluster by any other name compute as fast?". Below this, the cluster name "richho.cloudapp.net" is displayed, along with the expiration time "Expires in 2 hours and 17 minutes.". Two buttons, "Extend now" and "Release cluster", are available.

The "Your Tasks" section contains a grid of task cards:

- Create Job**: Status "New" (indicated by a plus icon).
- C# Streaming Example**: Status "Complete" (indicated by a checkmark icon).
- 10GB_Terasort Step 1**: Status "Complete" (indicated by a checkmark icon).
- 10GB_Terasort Step 2**: Status "Complete" (indicated by a checkmark icon).
- WordCount Example**: Status "Complete" (indicated by a checkmark icon).

Below the grid, there is an additional task card labeled "Complete" with a checkmark icon.

將 Windows 和 SQL Server 方便管理的特性 帶到 Hadoop

效益



提供Hadoop方便
管理 on Windows



企業級的安全性



同時容易管理設定
on-premises, Cloud

關鍵特色



整合 Microsoft
System Center



整合 Windows Active
DirectoryServer®

Smart packaging of
Hadoop on Windows

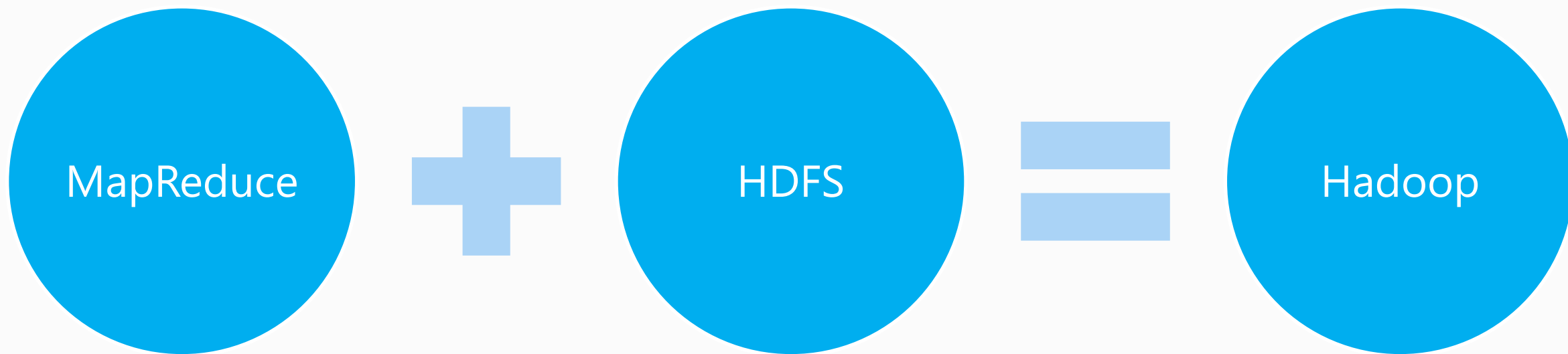
Hadoop-based
service on Windows
Azure

現行 80% 企業面對 Big Data
首要解決的都是大量的
累積資料，集中起來往往都
超過 Terabyte

可是**資料倉儲**的造價
都很昂貴，硬體絕大部分
都是特定獨規，很難彰顯
效益



何謂Hadoop?



What is our Hadoop Solution?

Three Hadoop distribution on Windows

HD Insight Service

- Integrated into Windows Azure platform.
- Lower cost , elastic scale



Microsoft



Windows Azure

HD Insight Server (preview)

- Can be managed by System Center
- Offer Rapid insights with Microsoft BI Tools



Microsoft



Windows Server 2012

Hortonworks Data Platform (HDP) for Windows (beta)

- 100% compatible with Apache Hadoop
- 100% open source

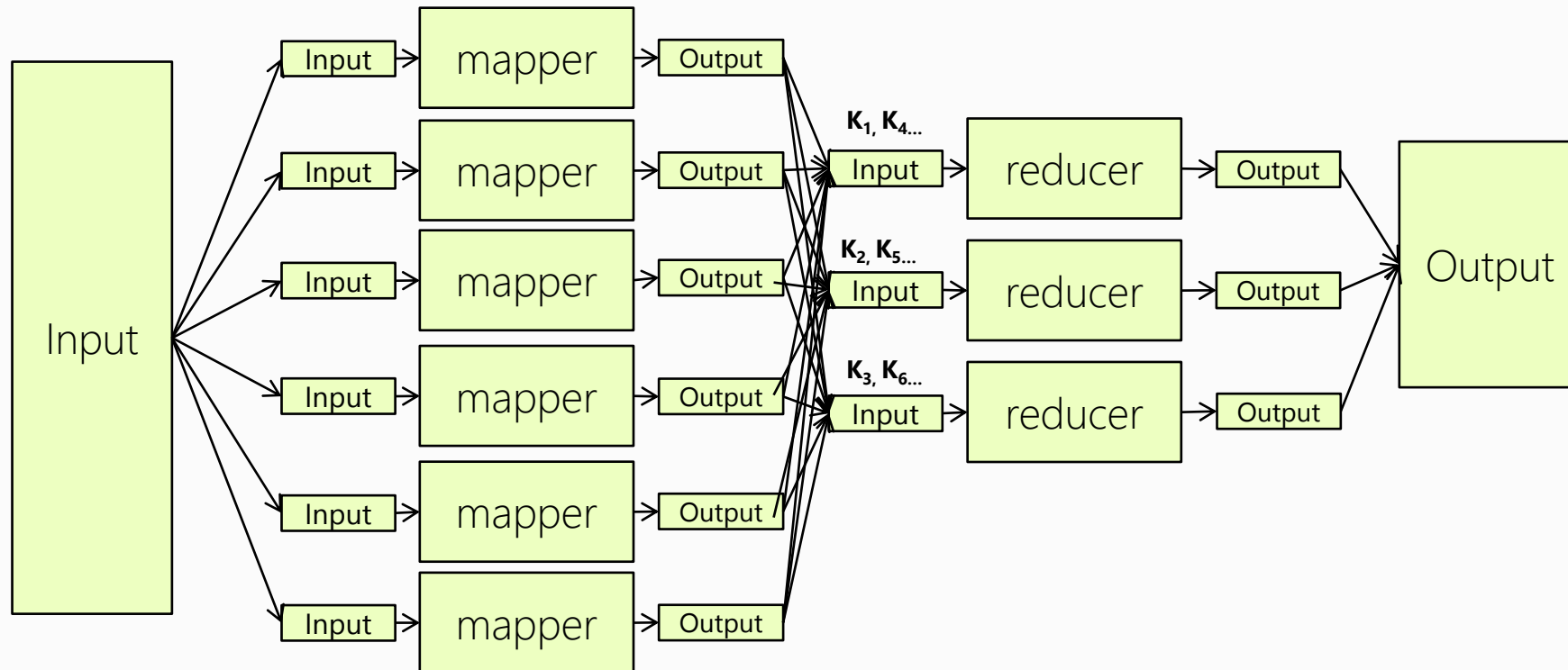


Hortonworks

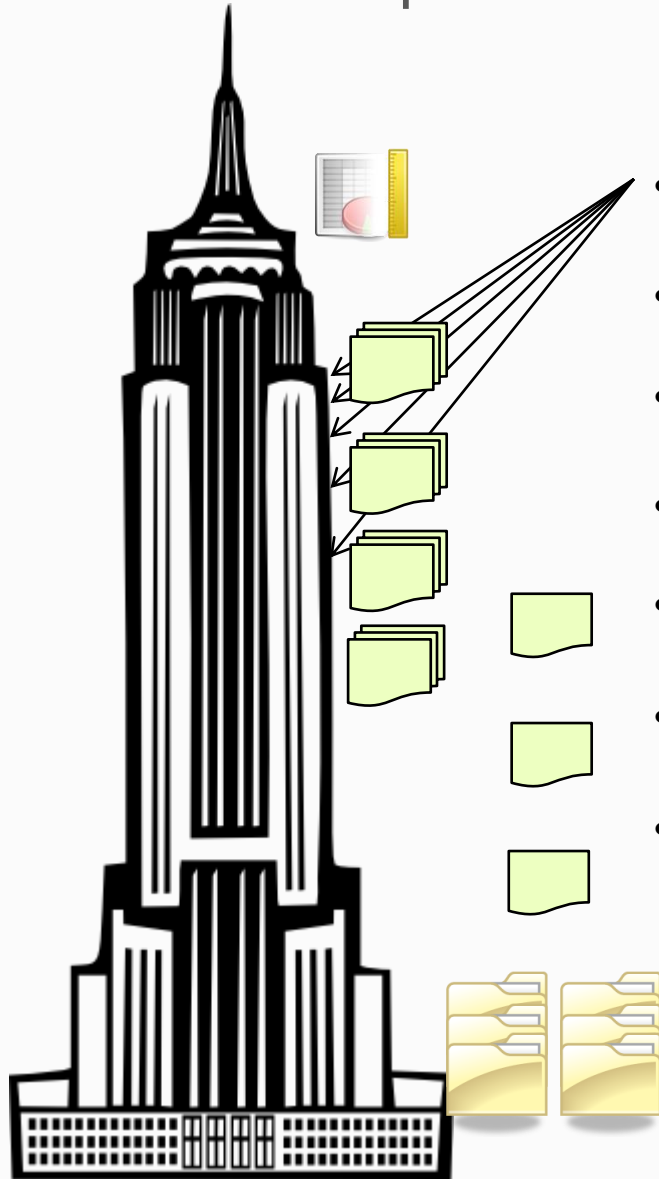


Windows Server 2012

MapReduce, in a Diagram



A MapReduce Example



- Count by suite, on each floor



to lobby

- ... photos to 10th, 20th, 30th floor
- Tally up each
- Collect the tallies
- Merge tallies into one spreadsheet



面對 55T 的 Big Data，趨勢科技如何使用微軟 Solution

利用 SQL Server 2012 資料倉儲資料行存放區索引技術提升應用效能

Microsoft®

Case Studies

Search Microsoft.com



擁有高達55T的資料量
查詢速度平均比原來效能快三倍
原本要跑一兩天的報表，現在幾小時就能跑出

Trend Micro

Internet Security Firm Fights Malware with Faster Data Processing and Contains Costs

Trend Micro needed faster data processing to keep pace with the growing number of potential threats on the Internet. The company's data warehouse had grown to more than 55 terabytes and data queries were getting slower. To keep up with threats—and to stay ahead of the competition—the company decided to upgrade to a Microsoft solution. Throughput test measurements indicate that the new solution will triple average query speeds and help block threats much more quickly.

Business Needs

Tokyo-based Trend Micro develops Internet content security and threat management solutions for businesses and consumers worldwide. It is fast-paced work. Tens of thousands of new pieces of malware invade the Internet every day, and trying to find and block these threats requires searching large volumes of data and responding quickly. "The environment changes very rapidly, so we have to act quickly in response," says Charlene Lee, Project Manager, Trend Micro.



“ With the old

To keep track of emerging threats and prevent them from infecting its customers' computers and mobile devices, Trend Micro collects suspicious samples and evidence of threats in a massive data warehouse. The company's analysts query that data to find the suspicious behaviors, or "signatures," that indicate that a



Securing Your Journey
to the Cloud

EMAIL LINK



DOWNLOAD



Organization Size: 4400
employees

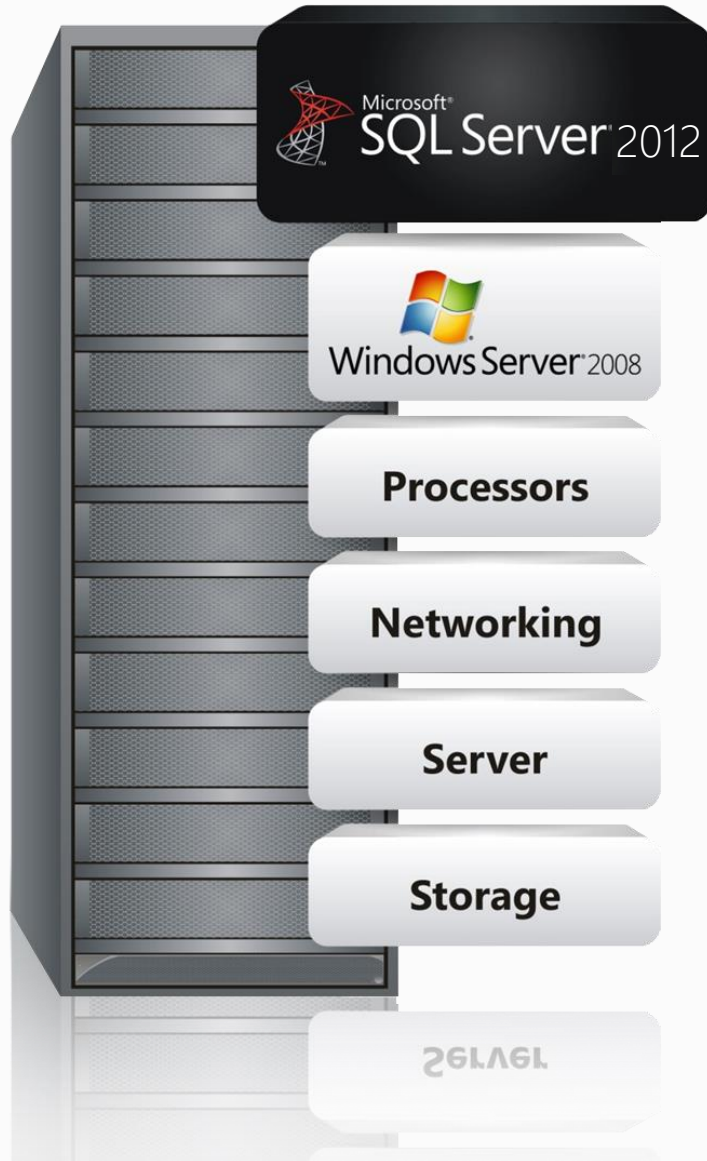
Organization Profile
Headquartered in Tokyo, Trend Micro provides Internet security software and services to businesses and consumers around the world.

Hardware
Dell PowerEdge R810 server
computers

Software and Services

- Microsoft SQL Server 2012 Enterprise
- Microsoft Services

SMP 資料倉儲首選：SQL Server Fast Track



- 資料庫伺服器加上硬體調校
- 適用 5T 至 80 T 資料量
- 專屬硬體資源的資料倉儲
- 支援多維度分析運算能力
- 鉅細靡遺的手冊指引安裝
- 多家硬體廠商與 SI 範本
 - HP
 - Dell
 - Bull
 - IBM
 - EMC

SQL Server Fast Track 優異表現

完整的產品提供快速的
建置及最佳效益

台灣現行廠商參考型號

2 Processor Configurations (5 – 20 TB, 2-3.7 GB/s)

HP ProLiant DL380 G7
HP ProLiant DL385 G7
IBM System x3650 M3
Bull Novascale R460 E2

4 Processor Configurations (20 – 40 TB, 3.5-7.5 GB/s)

HP ProLiant DL 580 G7
HP ProLiant DL 585 G7
IBM System x3850 x5
Bull Novascale R480 E1

8 processor Configurations (40 – 80 TB, 7.5-14 GB/s)

HP ProLiant DL 980 G7

80%

加快報表產生速度

95%

系統查詢較原有DW系統快

\$600K

每年節省硬體投資

50%

節省建置時間

10x

節省儲存空間



Represents storage array fully populated with 300 GB, 15,000 SAS, and use of 2.5:1 compression ratio. This includes the addition of storage expansion tray per enclosure.

30 percent of this storage should be reserved for DBA operations.

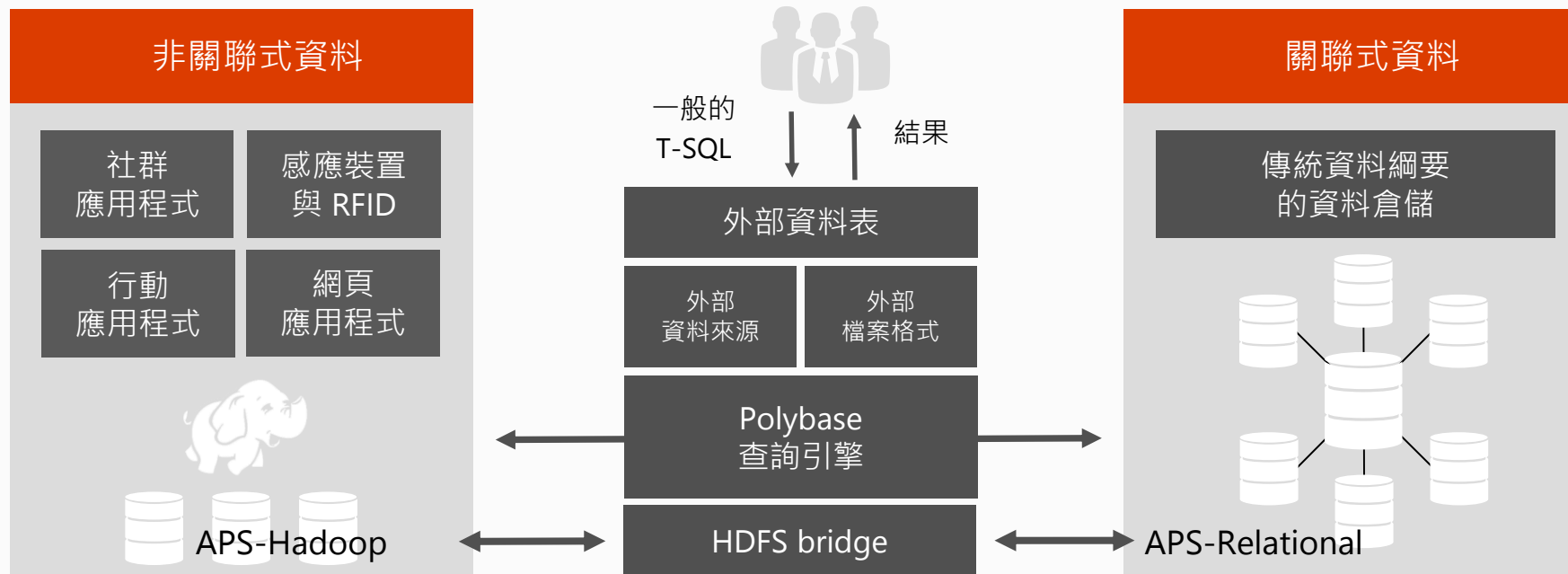
單獨的 Hadoop 不能解決巨量資料所有的挑戰

艱深的學習曲線、速度慢、效率低

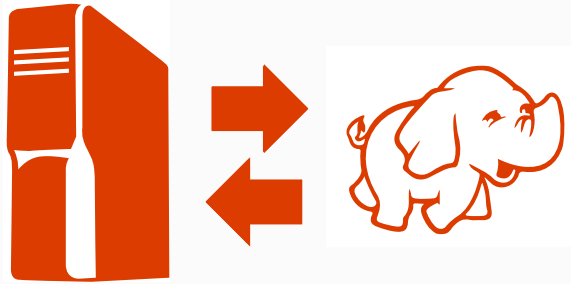


Analytical Platform System (APS)

- 直接且平行存取 HDFS
- 透過 外部資料表、Polybase查詢引擎與 HDFS Bridge

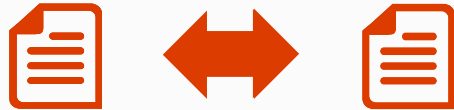


PolyBase 簡化 Hadoop 資料應用



進行 Hadoop 查詢時獲得更高的效能

不須特別技術，即可將資料倉儲的資料
封存至 Hadoop 中 (搬移)



匯出關聯式資料至 Hadoop (複製)



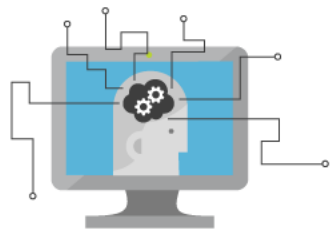
匯出 Hadoop 資料至資料倉儲 (複製)

以 HDFS 上的歷史事實資料表聯結
資料庫中的客戶資料、產品、銷售
金額，來建立新的外部資料表

```
PDW SQLQuery4.dsql * PDW SQLQuery3.dsql * PDW SQLQuery2.dsql *
PDW_POC_921

CREATE EXTERNAL TABLE FactSalesSummary_H
WITH ( LOCATION = 'hdfs://172.16.252.98:8020/user/pdw_user/factsalessummary_h.tbl'
      FORMAT_OPTIONS( FIELD_TERMINATOR = '|' )
)
as
select  fsh.CustomerKey, fs.ProductKey, sum( fs.SalesAmount ) SalesAmount
from    FactSalesHeader fsh join FactSales_HX fs on ( fsh.SalesOrderNumber =
group by fsh.CustomerKey, fs.ProductKey;
```

Data Mining Platform



認識 Azure 機器學習

有疑問嗎？機器學習以往皆需要複雜的軟體、高階的電腦以及熟悉機器的資深科學家才能進行。對於許多初學者，甚至是大型企業來說，這些都過於艱難且昂貴。加入 Azure 機器學習，即可使用受到完整管理的雲端預測分析服務。藉由雲端，Azure ML 能讓更多人接觸到機器學習。預測未來成果如今已可實現。

拖、放、預測

Azure ML 讓人們可以在沒有深入的資料科學背景下開始資料採礦與預測。ML Studio 是一個整合的開發環境，它利用拖放手勢與簡單的資料流程圖來設定實驗。您無需撰寫程式碼即可執行多項工作。ML Studio 也提供省時的範例實驗程式庫以及由微軟研究院開發的複雜演算法，包括經過相同嚴格測試的 Bing 與 Xbox 演算法。



可在此使用 R 語言

Azure ML 同時也為經驗豐富的資料科學家所設計。它支援 R 語言，是一種可供統計和資料採礦使用的開放原始碼程式環境。您可以現有的 R 程式碼直接拖放到工作環境，或在 ML Studio 中自行撰寫 R 程式碼，這兩種方式均支援超過 350 個 R 套件。可立即混搭匹配已驗證的演算法，並測試最多十種不同的模型。您可以輕鬆地共用 Azure ML 實驗，讓其他人可從您中斷的地方接續下去。

雲端所提供的內容

使用 Azure ML，不需要安裝任何軟體，不需要設定硬體，也沒有難懂的開發環境需要傷腦筋。您只需要一個瀏覽器登入 Azure，就可以隨時隨地開始開發預測模型，並以前所未有的速度部署新的分析模型。Azure ML 也讓您可以在 Azure 儲存體中保留無數個檔案，而且順暢地連結其他 Azure 資料相關服務，包括 HDInsight、Hadoop 巨量資料解決方案、SQL Database 和虛擬機器。



如何有效的運用 BIG DATA



Simplicity



Analysis with
familiar tools



Collaboration on
insights

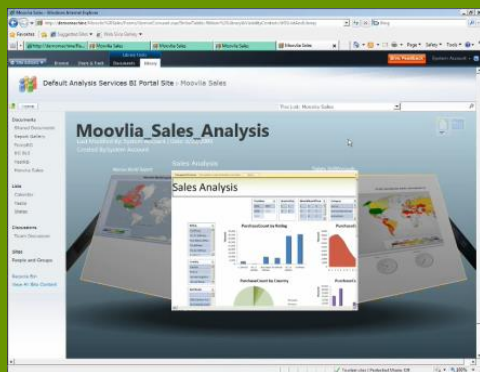
敏捷與熟悉的商業智慧分析能力



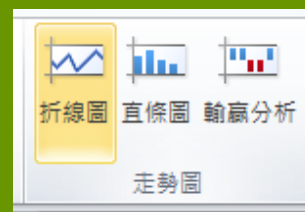
能處理巨量資料

	A	D	E	F	G	H
5						
6						
7	產品	第三季的加總	第四季的加總			產品
8	羊肉	\$ 4,836.00	\$ 6,087.90			主廚秋葵蛋糕
9	洋蔥香腸漿	\$ 140.00	\$ 440.00			主廚原住民族風味醬
10	波士頓蟹肉	\$ 4,412.32	\$ 1,656.00			巧克力餅乾
11	法國軟質乳酪	\$ 9,579.50	\$ 3,060.00			多層蛋糕
12	主廚原住民族風味醬	\$ 1,337.60	\$ 682.00			有機水餃餃
13	主廚秋葵蛋糕	\$ 288.22	\$ 85.40			羊肉
14	多層蛋糕	\$ 299.80	\$ 904.75			果醬 (素和)
15	義大利乾酪	\$ 515.00	\$ 2,681.87			法國軟質乳酪
16	野莓莓醬	\$ 1,750.00	\$ 750.00			
17	馬來西亞咖啡	\$ 1,196.00	\$ 3,979.00			
18	蛤蜊海鮮濃湯	\$ 1,582.60	\$ 1,664.62			
19	淡啤酒	\$ 350.00	\$ 42.00			
20	長壽豆腐	\$ -	\$ 512.50			
21	辣雞麵	\$ 1,375.62	\$ 3,899.51			
22	麻辣秋葵	\$ 68.00	\$ 850.00			
23	義大利白乾酪	\$ 3,027.60	\$ 2,697.00			
24	蕃茄醬	\$ -	\$ 2,960.00			
25	義大利餃	\$ 390.00	\$ 984.75			
26	麥芽酒	\$ -	\$ 890.40			
27	橘子果醬	\$ 3,061.80	\$ -			
28	辣雞	\$ 1,733.00	\$ 1,434.00			

具備互動多維度分析



有效分享能自動化

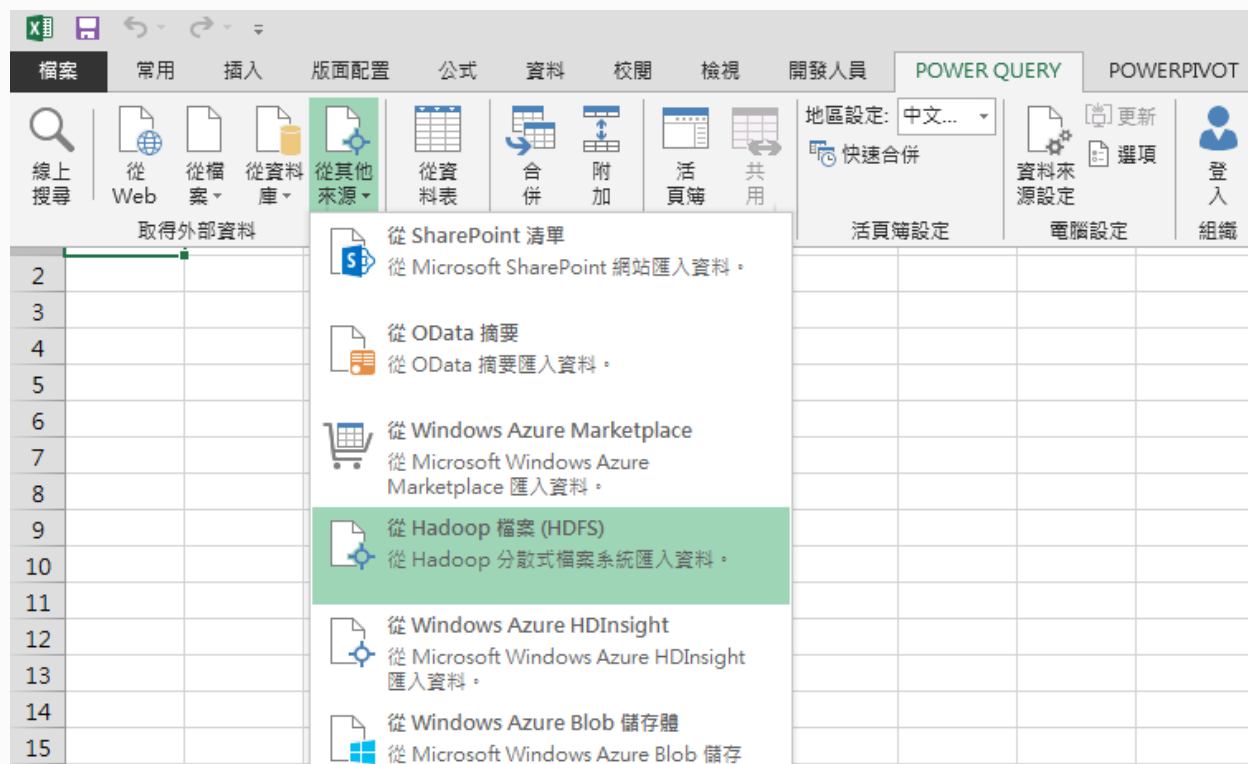


	數值	第一季的加總	第二季的加總	第三季的加總	第四季的加總
6					
7	產品				
8	羊肉	\$ 2,667.60	\$ 4,013.10	\$ 4,836.00	\$ 6,087.90
9	洋蔥香腸漿	\$ 544.00	\$ 600.00	\$ 140.00	\$ 440.00
10	波士頓蟹肉	\$ 1,768.41	\$ 1,978.00	\$ 4,412.32	\$ 1,656.00
11	法國軟質乳酪	\$ 3,182.40	\$ 4,683.50	\$ 9,579.50	\$ 3,060.00
12	主廚原住民族風味醬	\$ 225.28	\$ 2,970.00	\$ 1,337.60	\$ 682.00
13	主廚秋葵蛋糕	\$ -	\$ -	\$ 288.22	\$ 85.40

有意義的統計資訊

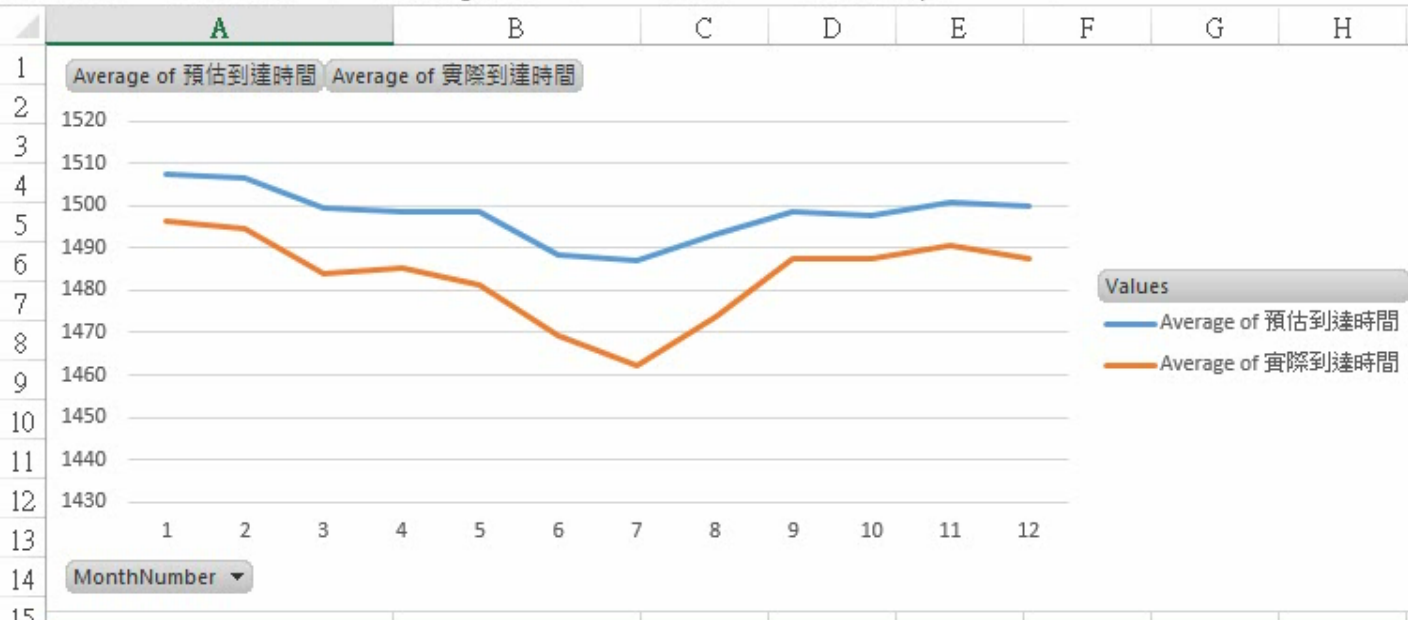
Excel 2013

- Excel 2013也能夠直接讀取APS-Hadoop之資料
- 使用資料無須再從Hadoop端匯出



Data Model **Calculations** **Slicer Alignment** **Tables** **Relationships**

Manage Calculated Fields KPIs Align Vertically Align Horizontally Add to Data Model Update All Detect Settings



- 國際機場**
- Abilene, TX: Abilene Municipal
 - Adak Island, AK: Adak Is
 - Aguadilla, PR: Borinquen
 - Akron, OH: Akron-Canton Regi...
 - Albany, GA: Albany Dougherty...
 - Albany, NY: Albany County
 - Albuquerque, NM: Albuquerque...
 - Alexandria, LA: Alexandria Int...
 - Allentown/Bethlehem/Easton, P...
 - Amarillo, TX: Amarillo Airport
 - Anchorage, AK: Anchorage Inte...

各因素總誤點次數總和	Column Labels										
Row Labels	1	2	3	4	5	6	7	8	9	10	11
AirTran Airways	103946	62841	79893	48250	113403	133882	273441	201463	110677	82312	96906
Alaska Airlines	113143	48138	84283	48219	46324	71597	78427	53502	41590	44933	79129
American Airline	318669	269351	481994	397049	540143	497485	602311	568416	405992	357684	355581
American Eagle	314625	249194	359615	285065	358712	372528	562615	406249	263760	257715	282652
Atlantic Southeast	712419	656331	994535	575481	804955	939683	1468469	416187	278892	241334	258353
Delta	363115	257009	418828	268215	429687	523190	831658	602931	344383	291711	294550
Froniter Airlines	68598	79059	62455	46795	49009	84650	76601	61284	43598	44314	38934
Hawaiian Airlines	22333	22871	21853	17012	15909	16031	31375	14177	15631	14956	17667
JetBlue Airways	174658	129182	200764	130620	175346	300397	363052	469785	218190	232487	124062

PivotTable Fields

ACTIVE ALL

Choose fields to add to report:

Month

- ☒ MonthNumber
- ☐ MonthOrderBy
- ☐ 月份

On-Performance

- ☐ Carrier
- ☐ DayOfMonth
- ☐ DayOfWeek
- ☐ DestState

Drag fields between areas below:

FILTERS

COLUMNS

MonthNumber

ROWS

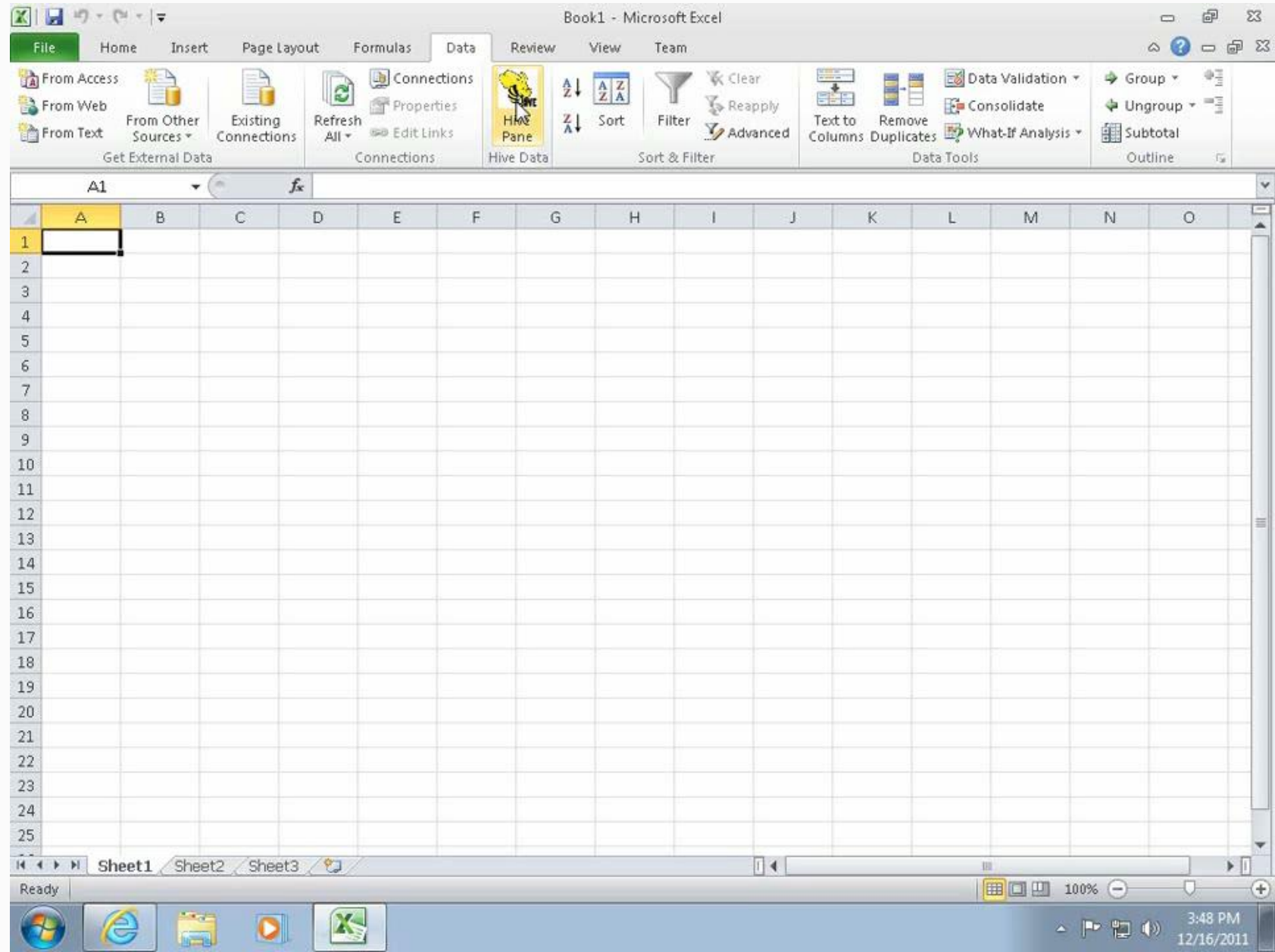
航空公司

VALUES

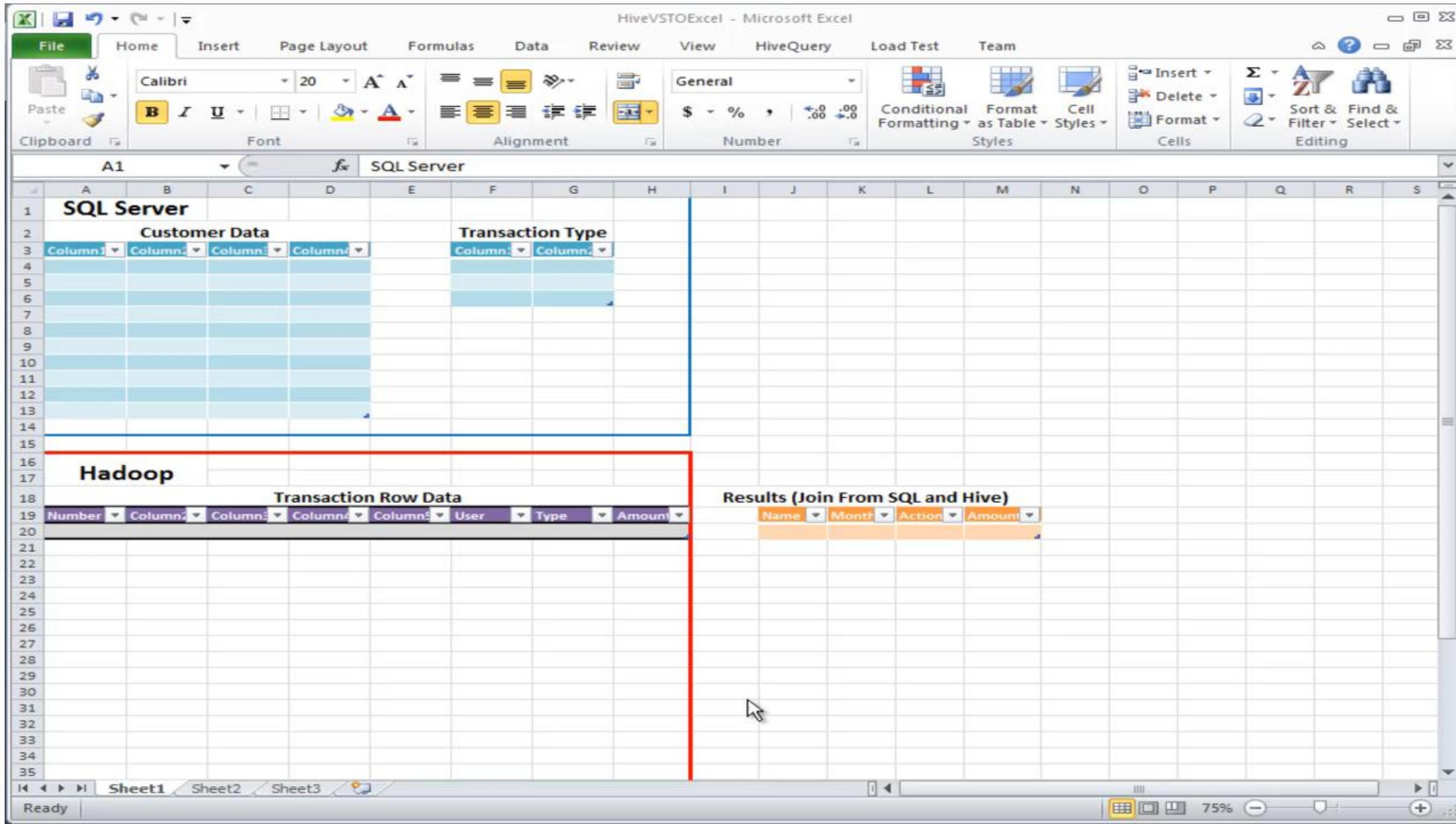
各因素總誤點...

☐ Defer Layout Update **UPDATE**

Hive Add-in for Excel



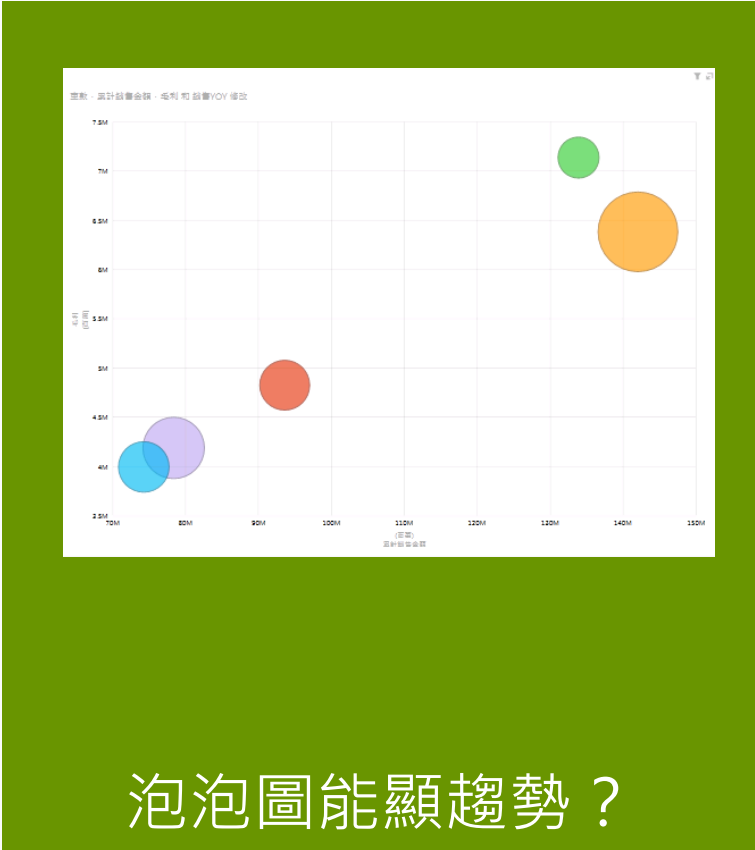
使用 Excel 連接 Hadoop 進行 BI 分析



讓數字、圖表與管理更具體的方法



分析呈現侷限在二維



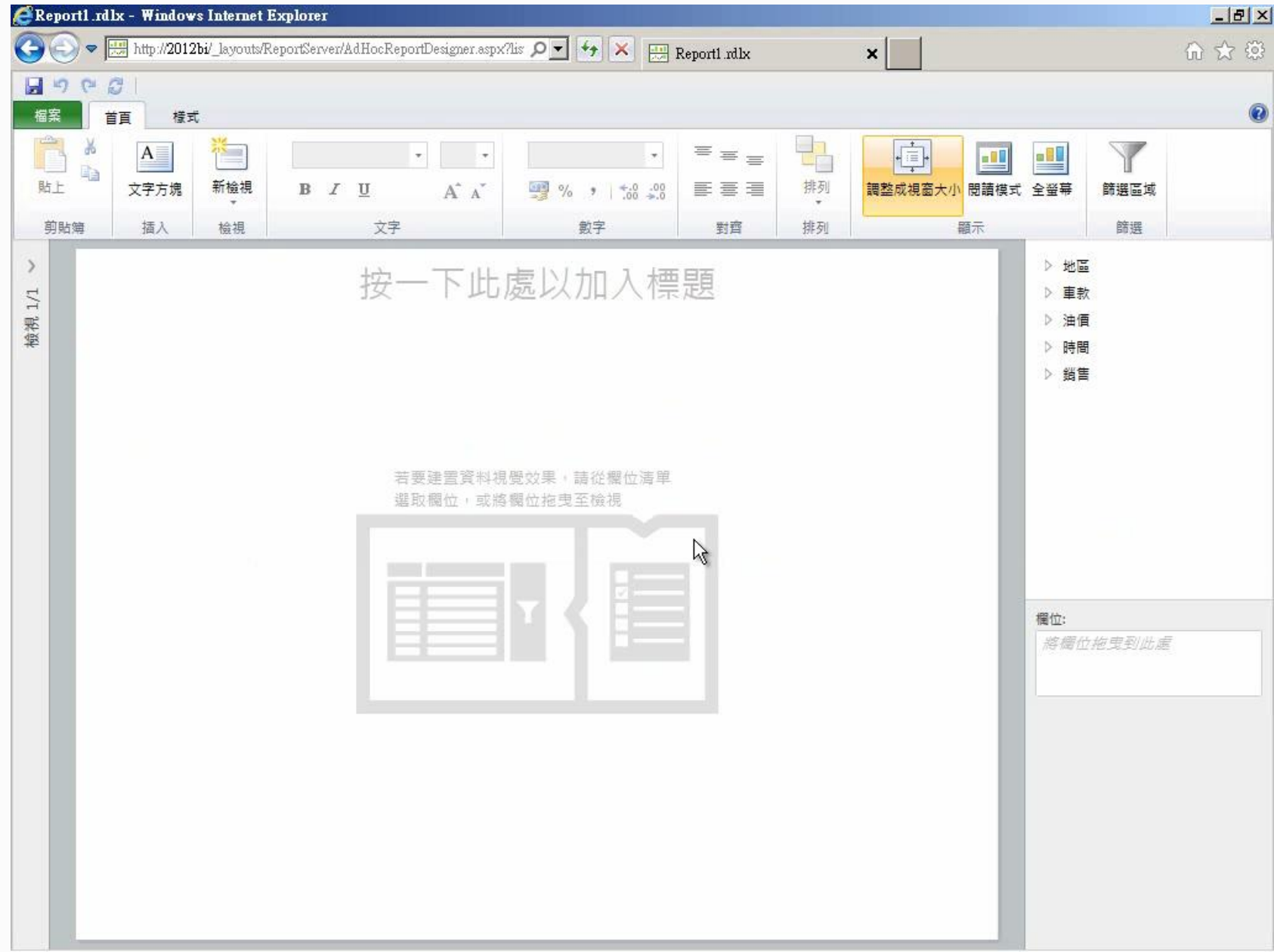
泡泡圖能顯趨勢？

台灣微軟股份有限公司
台北市11073信義區松仁路7號8樓
總機：(02) 3725-3888 傳真：(02) 3725-3686
專線：(02) 3725-3783 統一編號：23525730
客服中心：0800-008833 (市話專用)
(02) 2999-8833
行動電話：0921-053582
電子信箱：angiwang@microsoft.com
網址：http://www.microsoft.com/taiwan

王恩琦
專案技術經理
大型企業業務暨經銷事業群

善用名片圖像式管理

即時互動分析 Power View

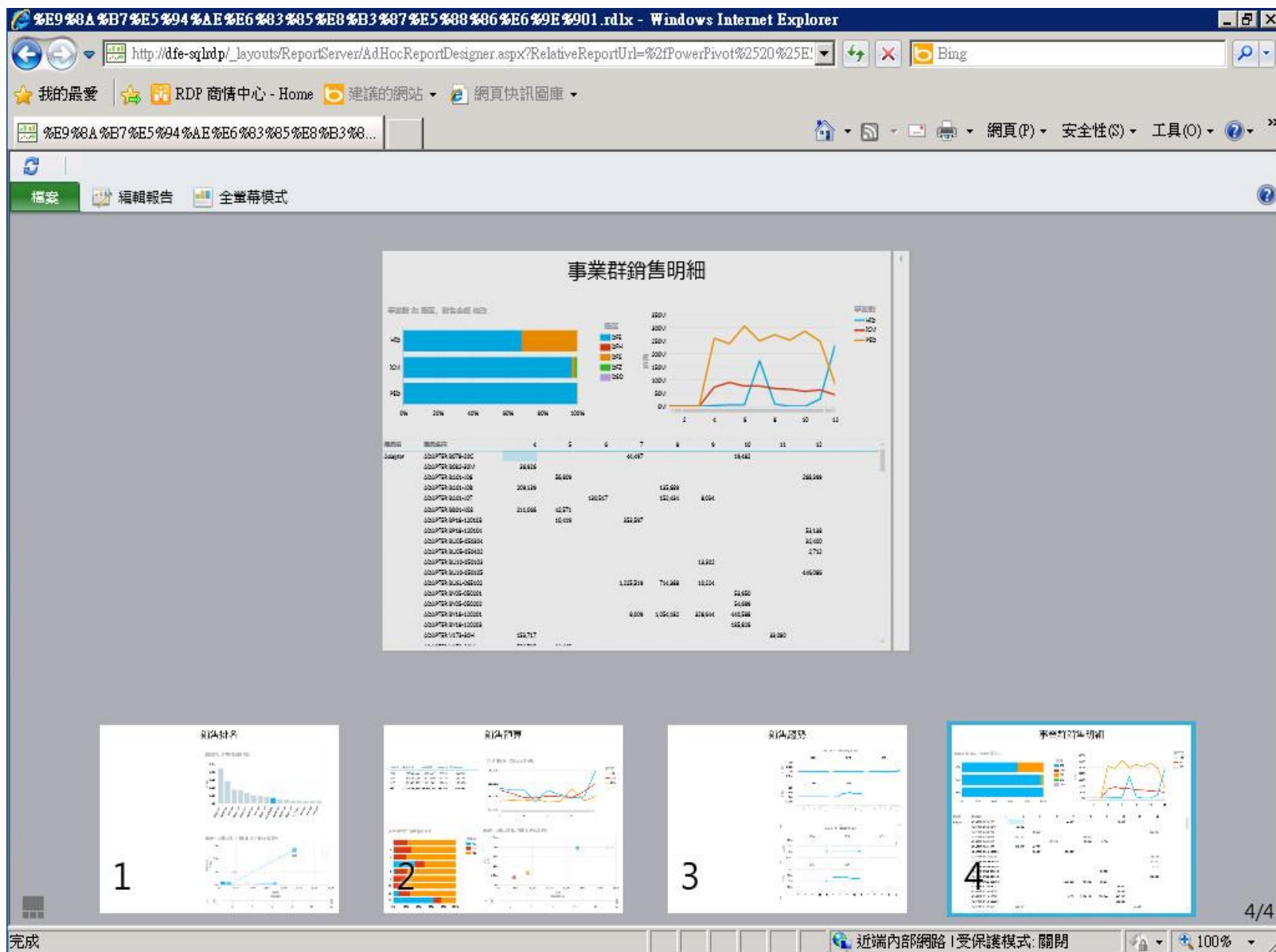


MLB Analytics with Microsoft Business Intelligence

HOME



全新互動Power View 主題式報表

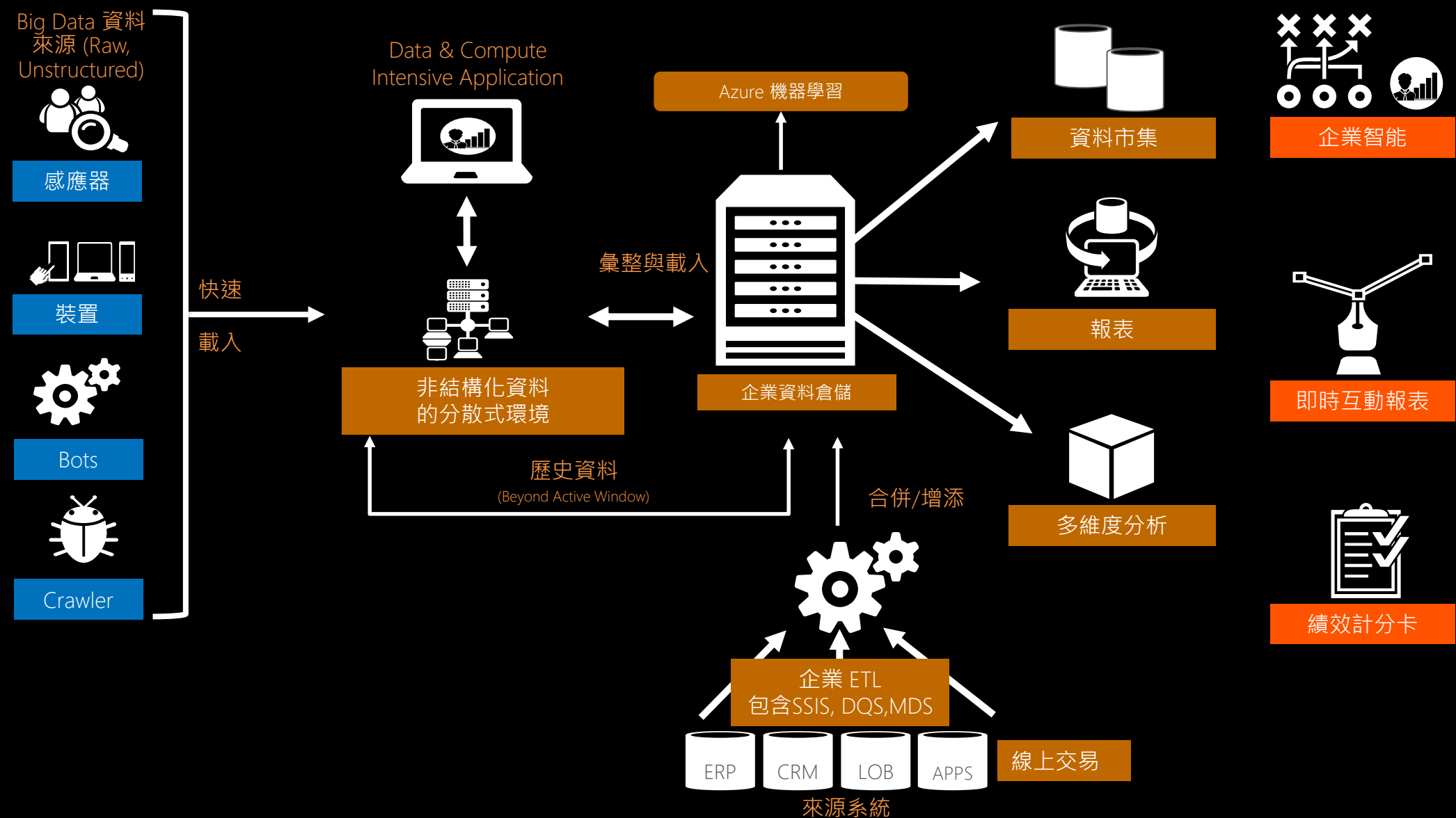


銷售排名與產品明細

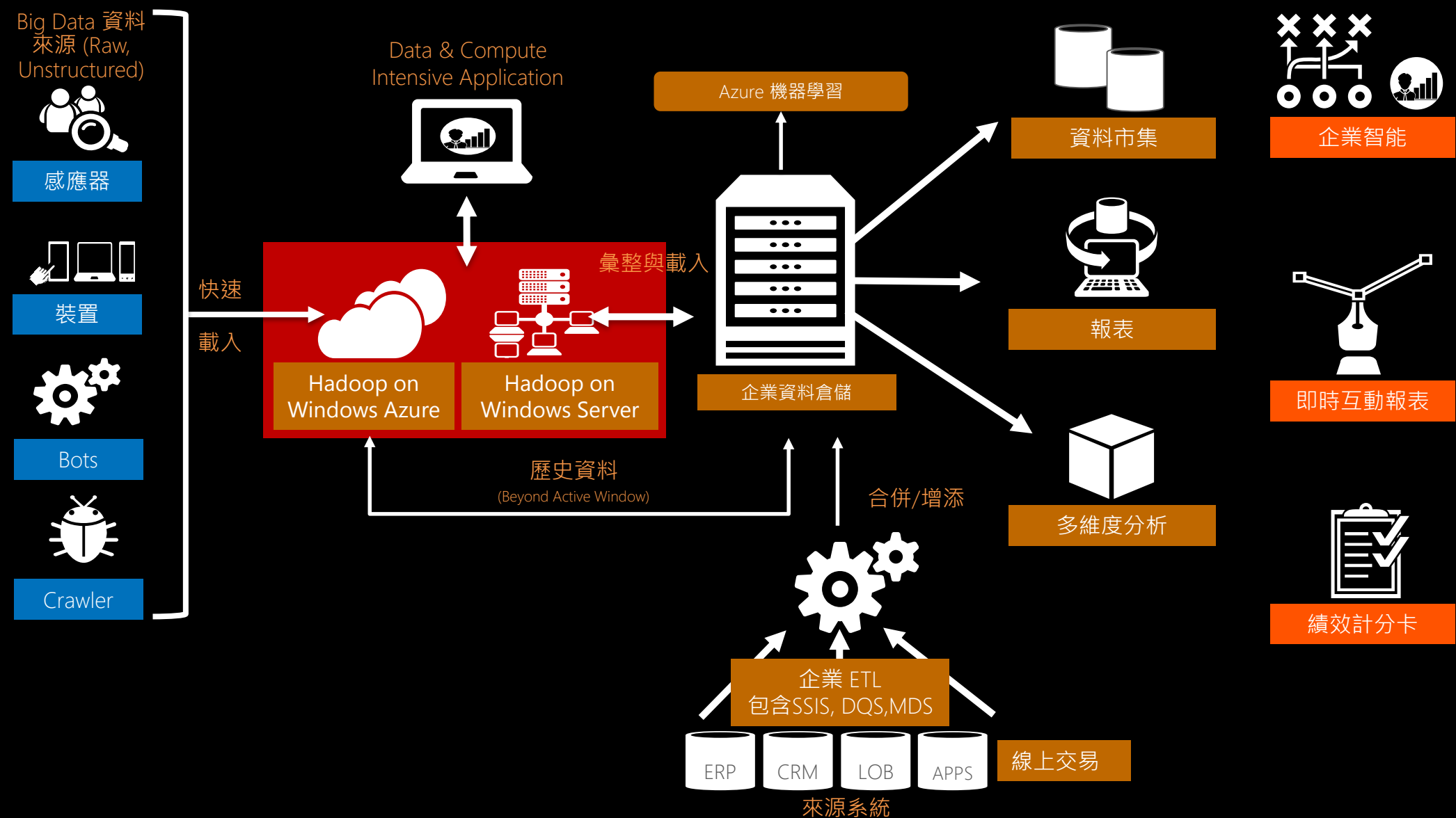


建構微軟的 海量資料及商業智慧

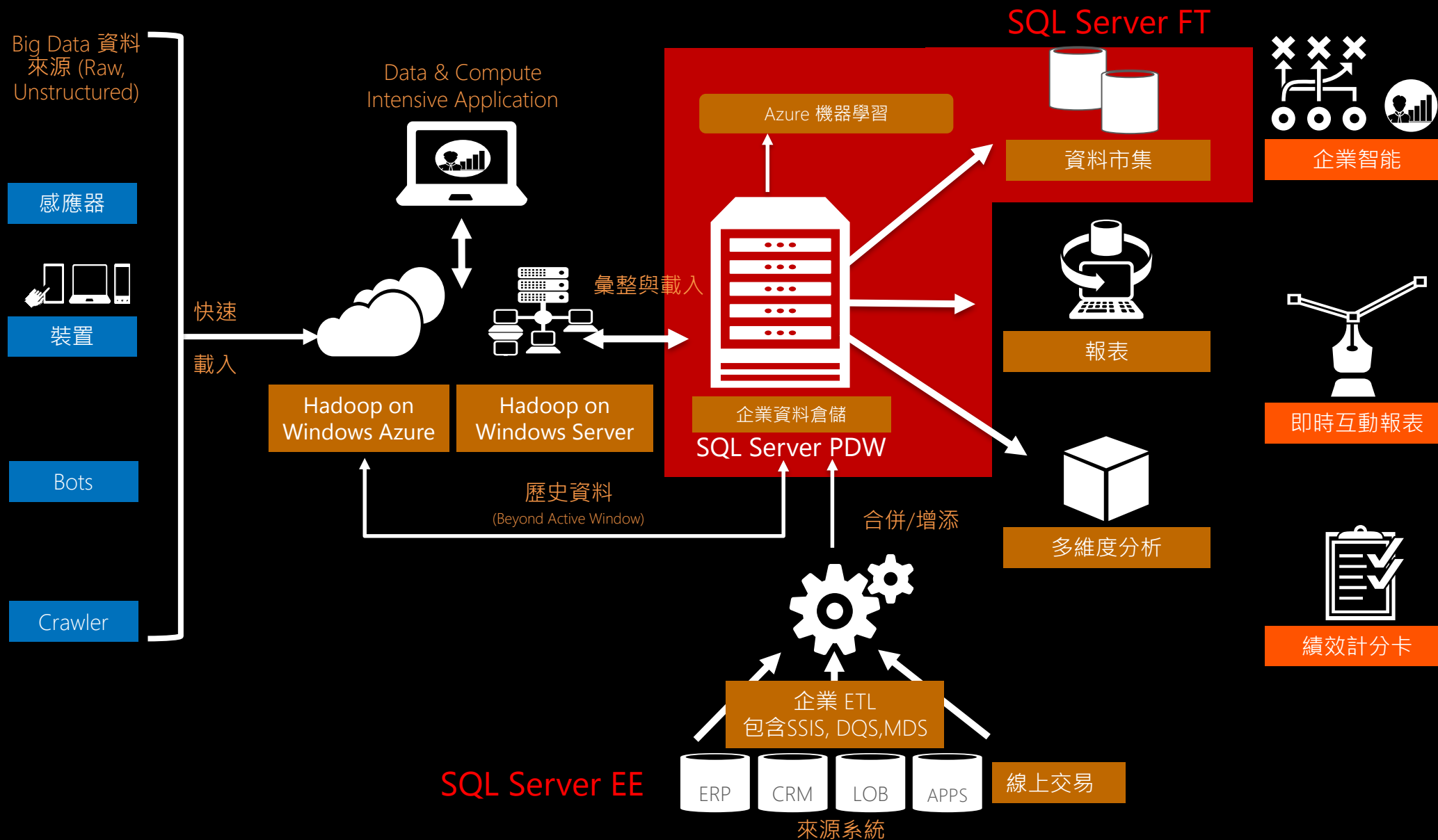
Enterprise Big Data Solution



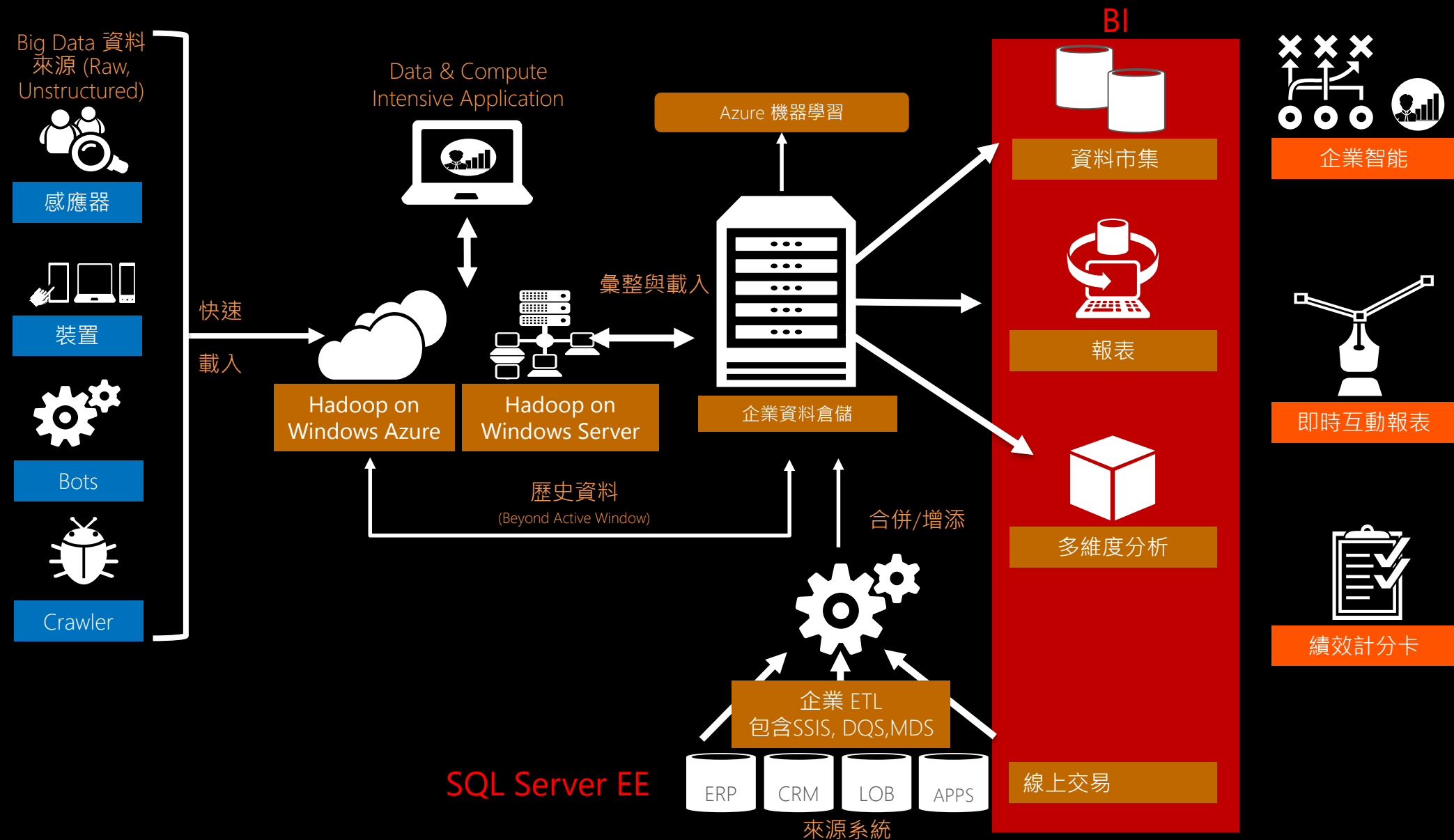
Microsoft Hadoop Support in Big Data



Microsoft Data Platform in Big Data



Microsoft Business Intelligence in Big Data



結論

The Microsoft logo is displayed in white on a solid blue background. It features the word "Microsoft" in a bold, italicized sans-serif font, followed by a registered trademark symbol (®).