

» 巨量資料分析與應用 (4)

楊立偉教授

台灣科技大學資管系

2013 Fall

楊立偉教授

◆ 專長領域為資料庫及語意分析技術、知識管理、數位行銷

現任	台科大資管系兼任助理教授 2008~
	台大工管系暨商研所兼任助理教授 2006~
	資訊及通信國家標準技術委員
	意藍資訊 董事總經理（創辦人） 1999~ 國內規模最大的網路情報與社群口碑自動分析平台
	龍捲風科技 董事總經理 國內企業搜尋引擎市佔率最高；國際檢索競賽第一名
經歷	智威湯遜數位行銷首席顧問、尚藍互動行銷共同創辦人
	2009年獲選100 MVP最有價值經理人，擁有超過20項語意分析專利
	2012年榮獲國家雲端創新獎、數位時代「創業之星」首獎

課程大綱

◆ 第一部份

- 巨量資料導論
- 巨量資料分析與管理架構
- 巨量資料分析技術

◆ 第二部份

- 應用案例與研討 – 企業個案 (1)
- 應用案例與研討 – 企業個案 (2)
- 應用案例與研討 – Open Data

應用案例與研討 – 大數據資料倉儲

資料倉儲 Data warehouse

- ◆ A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions

企業為何需要 Data warehouse

- ◆ Increasing customer focus, which includes the analysis of customer buying patterns.
- ◆ Repositioning products and managing product portfolios by comparing the performance of sales by time or regions, to fine-tune production strategies
- ◆ Analyzing operations and looking for sources of profit
- ◆ Managing the customer relationship
- ◆ Managing the cost of corporate assets

Data warehouse 特性

- ◆ for analytical tasks, using data from multiple applications 分析用途為主
- ◆ relatively small number of users with relatively long interactions 使用者人數較少，但互動程度高
- ◆ read-intensive 讀取為主
- ◆ periodically updated 週期性更新資料
- ◆ contains current and historical data 新舊資料並存
- ◆ Each query frequently results in a large result set and full table scan and multi-table joins 運算量大



◆ Why not use Database directly ?

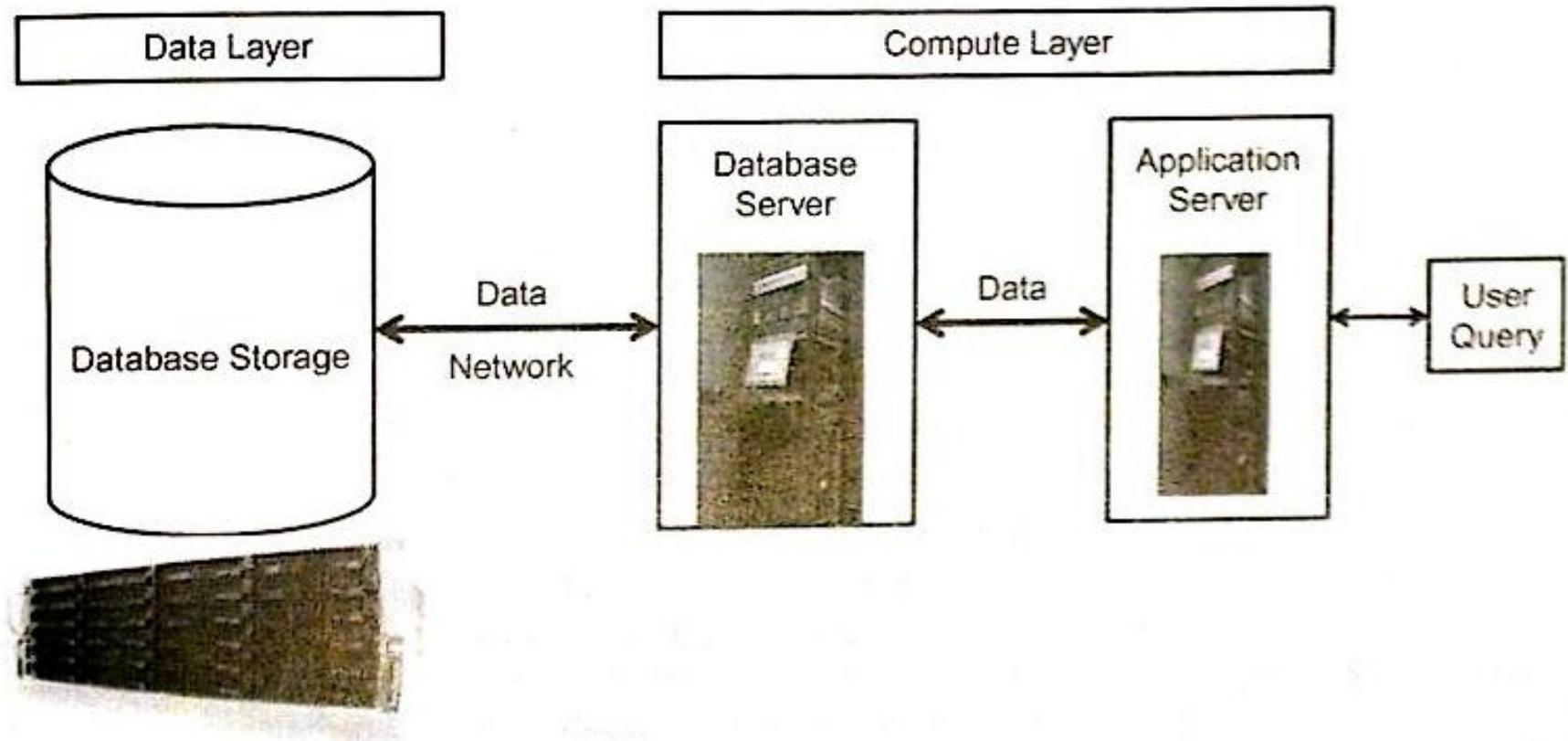
- update-driven approach is inefficient.
- Potentially expensive for frequent queries.

◆ Use Data warehouse instead

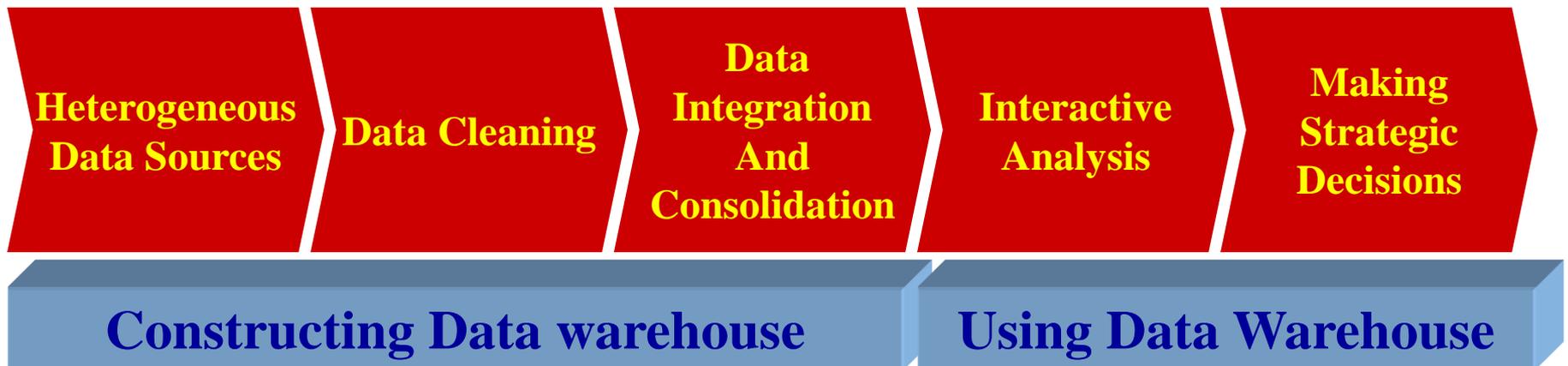
- query-driven approach is enough for making strategic decisions.
- Separate the operational DBMS for daily and critical operations.

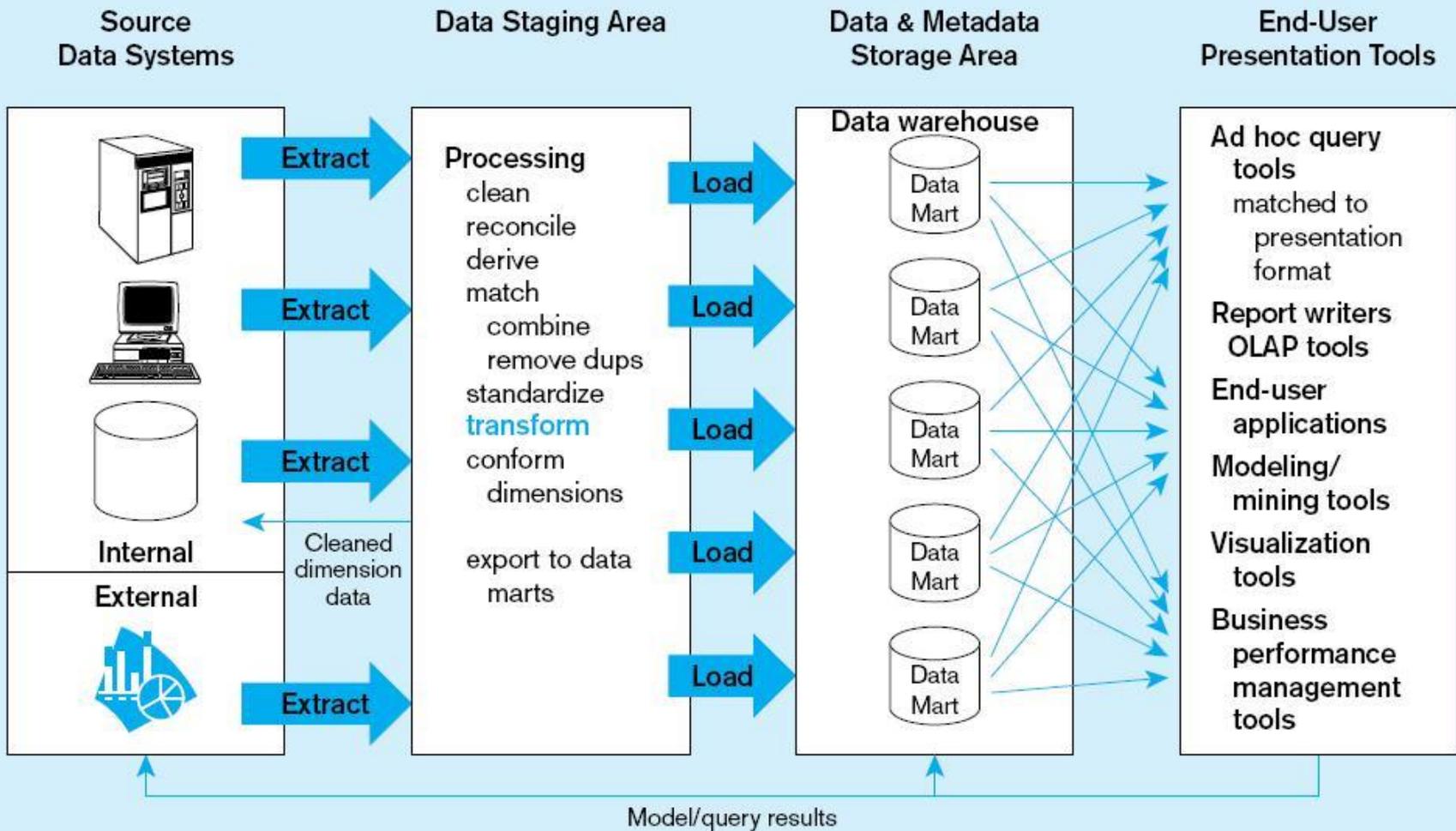
Data warehouse 處理流程

- ◆ data warehouse processing (original)



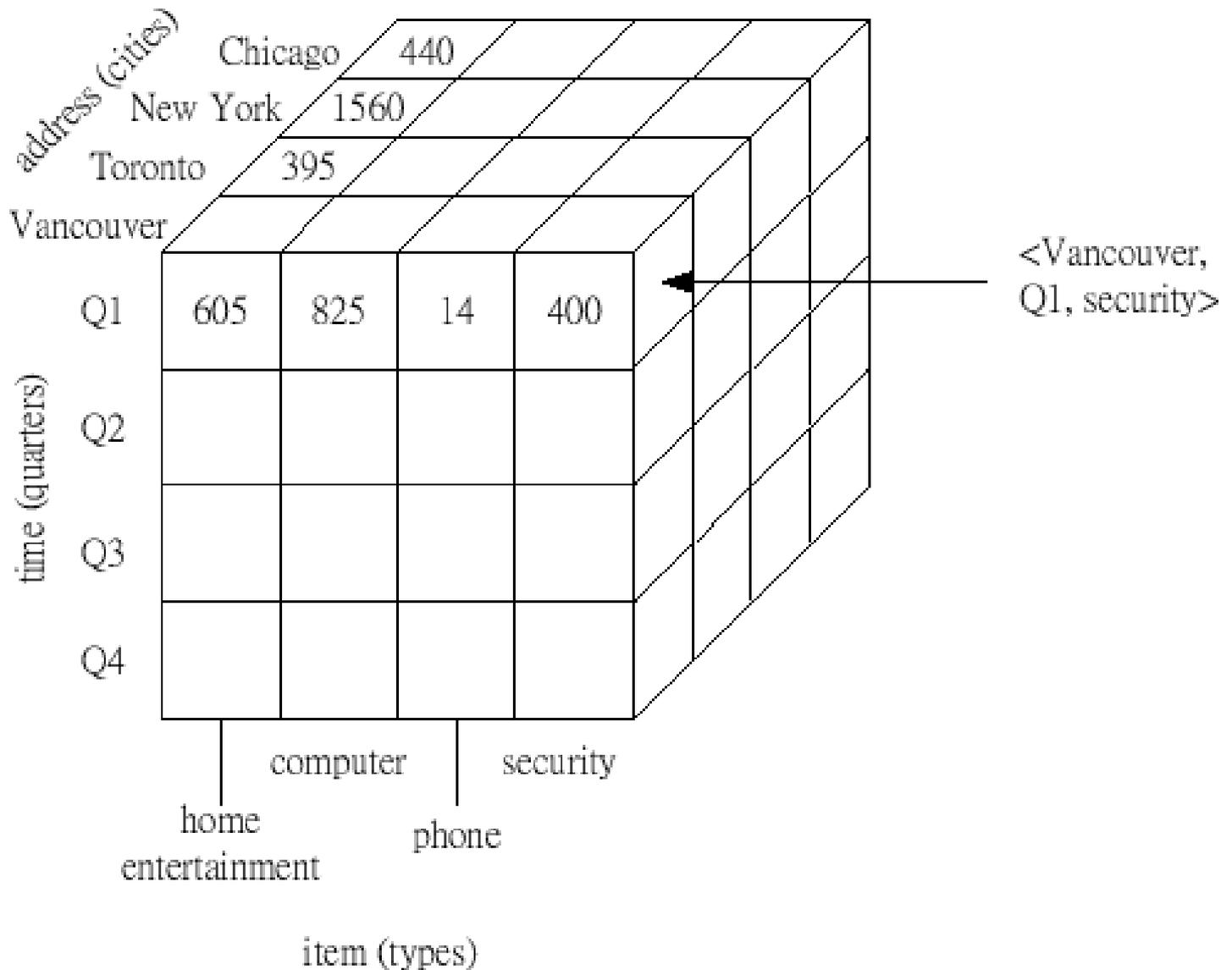
◆ data warehouse processing (original)



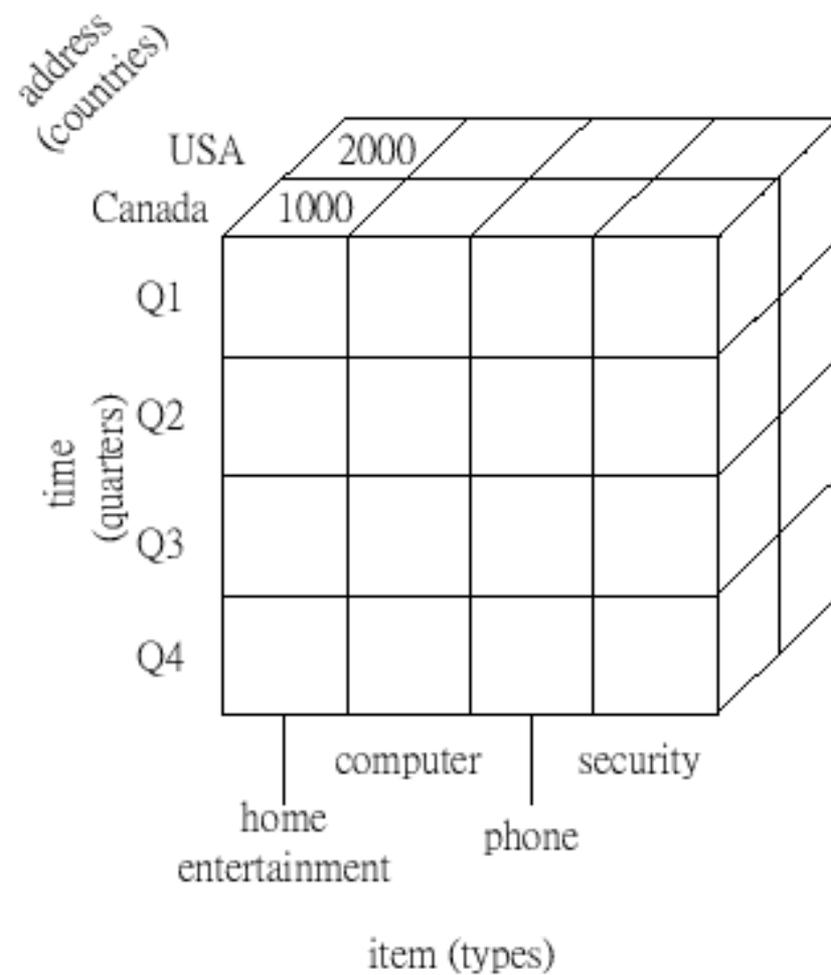
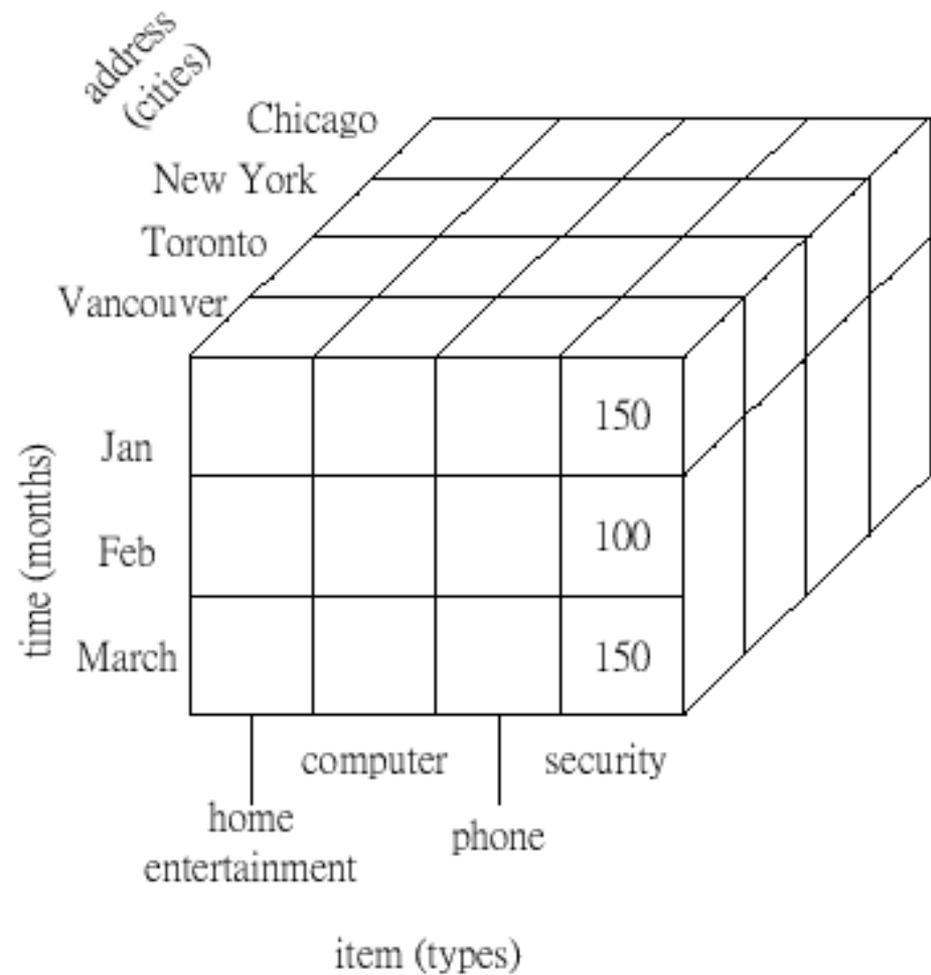


Data warehouse 原理 – Data cube

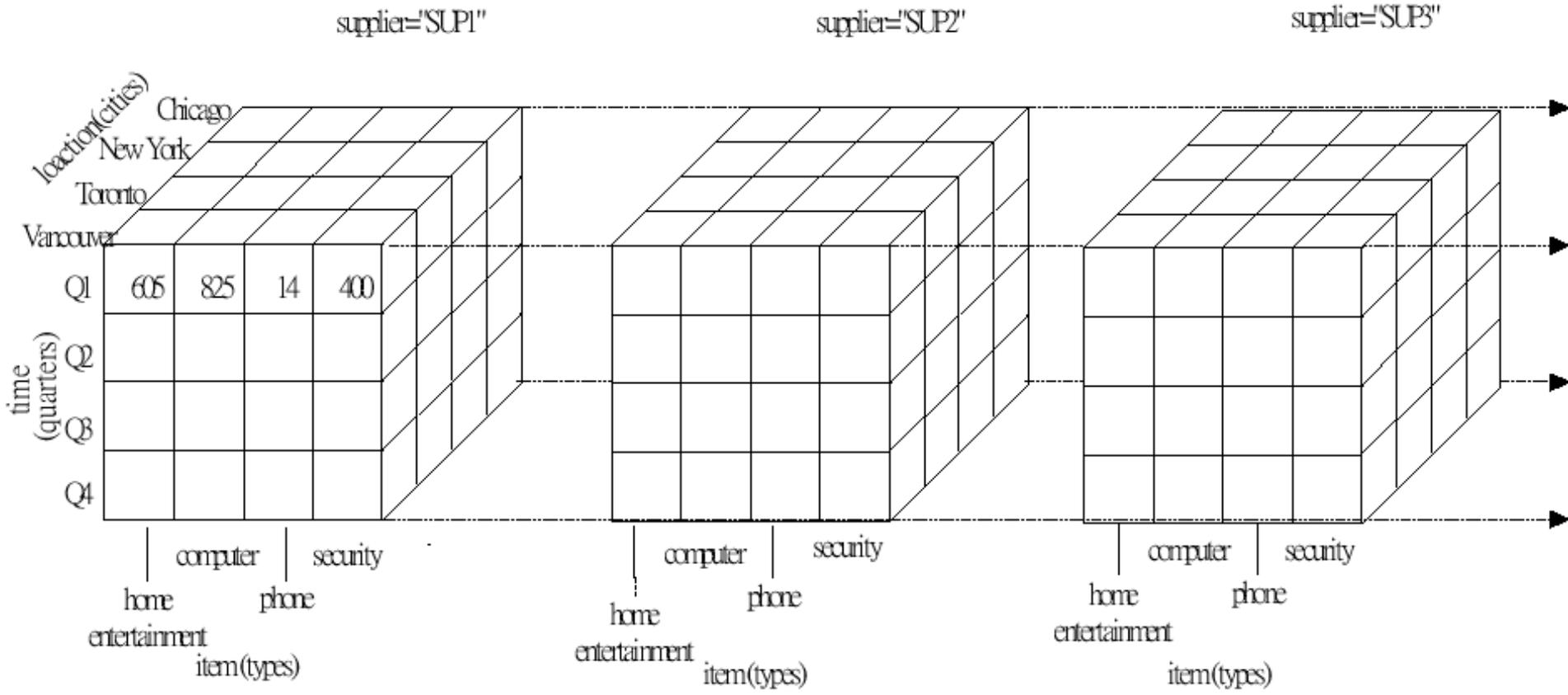
- ◆ A multidimensional, logical view of the data
- ◆ Concept hierarchy
 - Multiple data granularity 多重的資料顆粒度
 - Data summarization 資料加總
 - Data generalization 資料一般化



(a)



- Roll-up on address

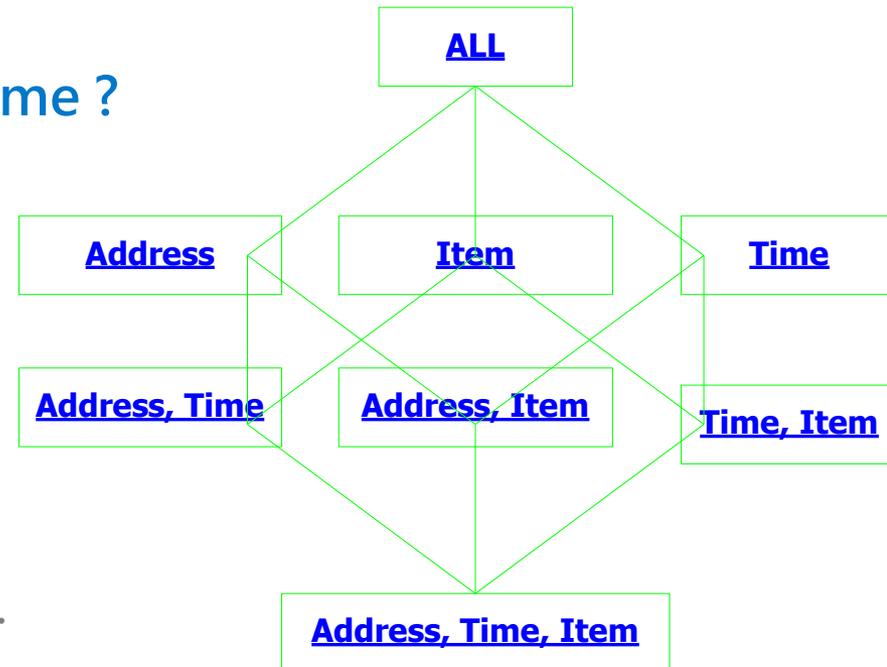


◆ The challenges

- There are 2^N combination for data cube computation.
- Concept hierarchy and Aggregations makes it more complicated

◆ Efficient data cube computation

- Materialize every, none, or some ?
- Algorithms for selection
 - Based on size
 - Based on sharing,
 - Based on access frequency.

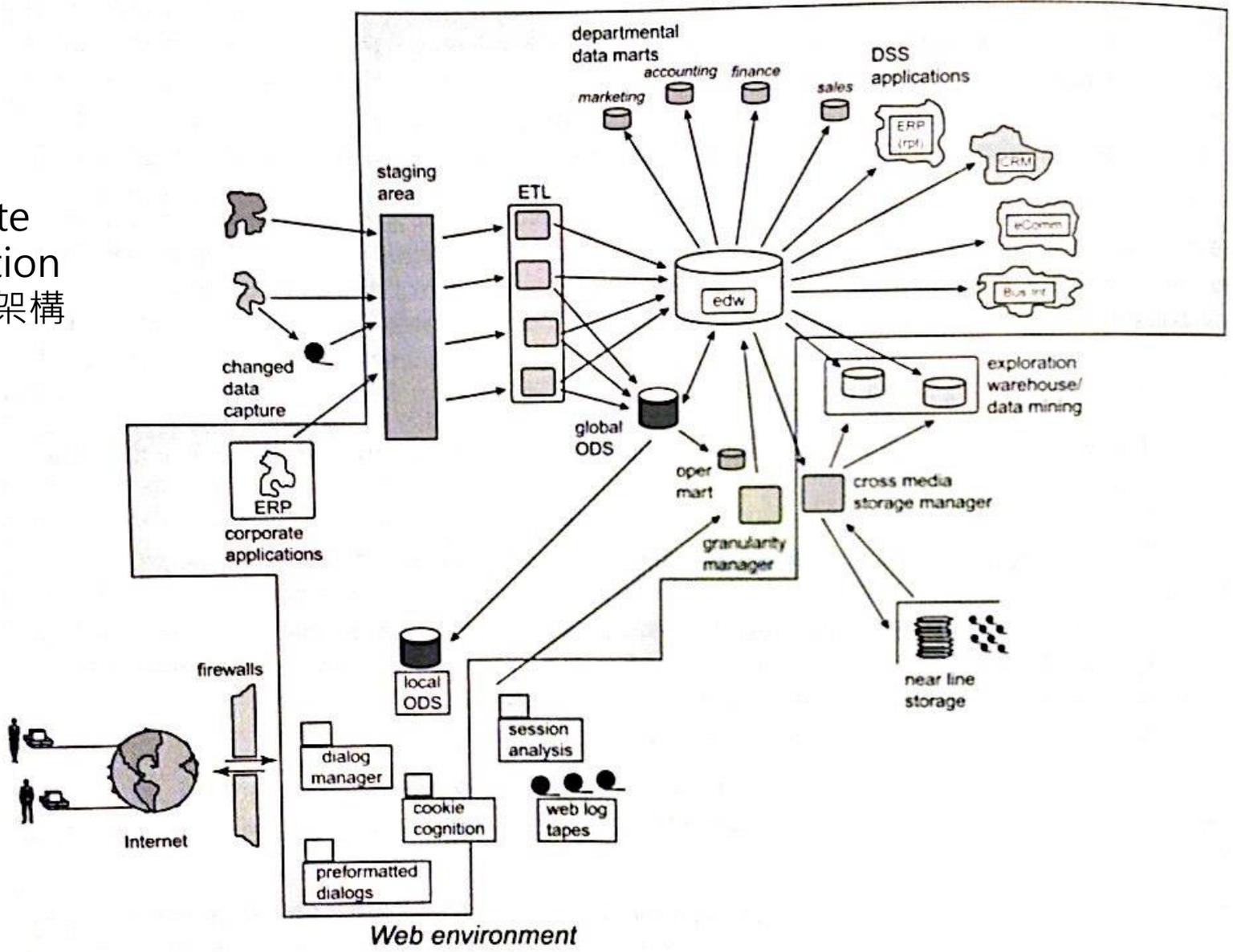


On-Line Analytical Processing (OLAP)

- ◆ Fast on-line processing of data cubes or multi-dimensional databases
 - Drill down / Roll up 、Pivoting 樞紐分析、Slicing / Dicing, etc.
- ◆ Provide interactive analysis and quick response to OLAP queries.
 - Summarization and aggregations at every dimension
 - Retrieval and display of data in 2-D or 3-D cross-tabs, charts, and graphs, with easy pivoting of the axes.
 - Analytical modeling : deriving ratios, variance, etc.
 - Use for forecasting, trend, statistical analysis.

Corporate information factory

Corporate Information Factory 架構

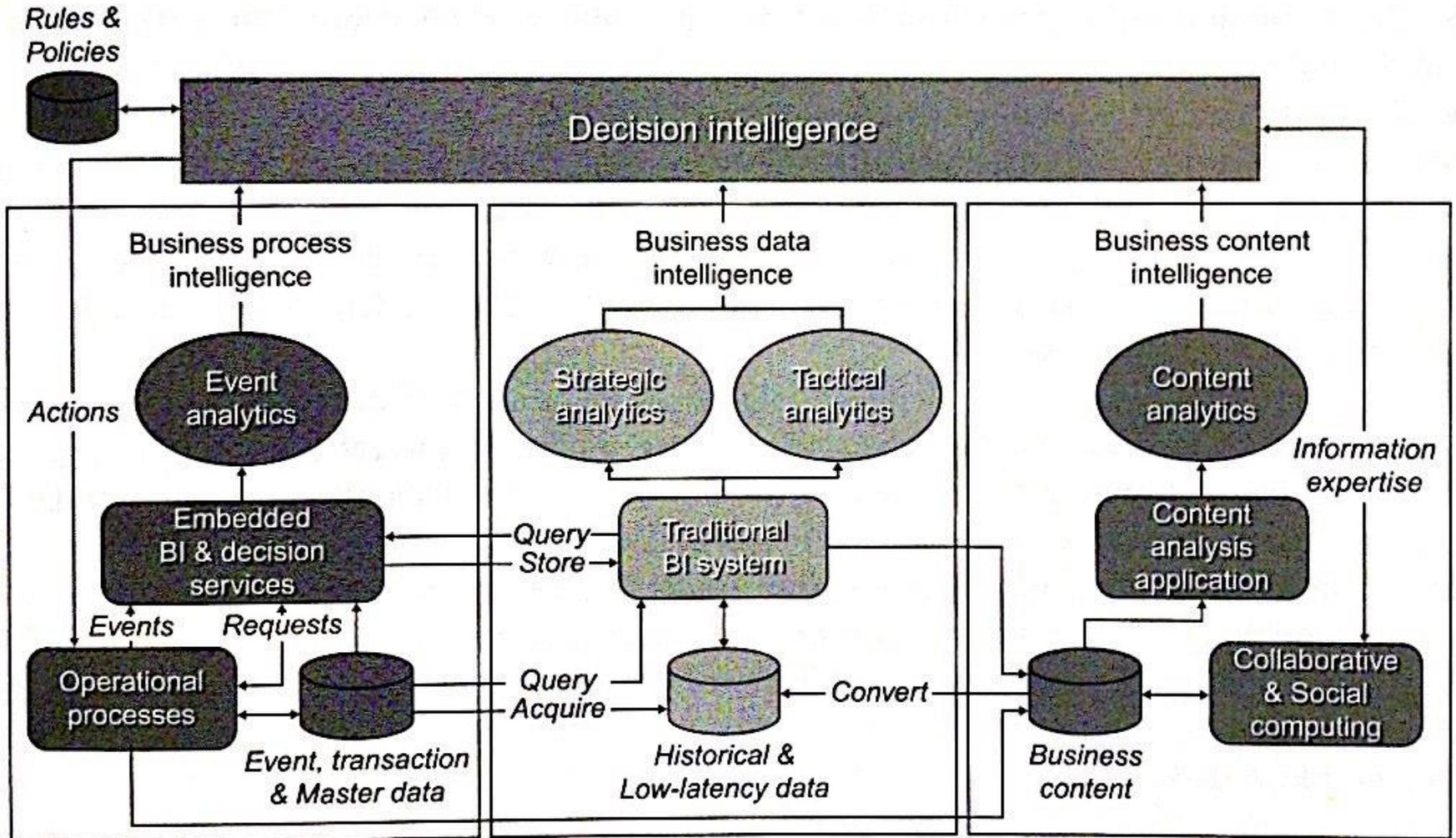


Data warehouse 2.0

◆ The BI matrix

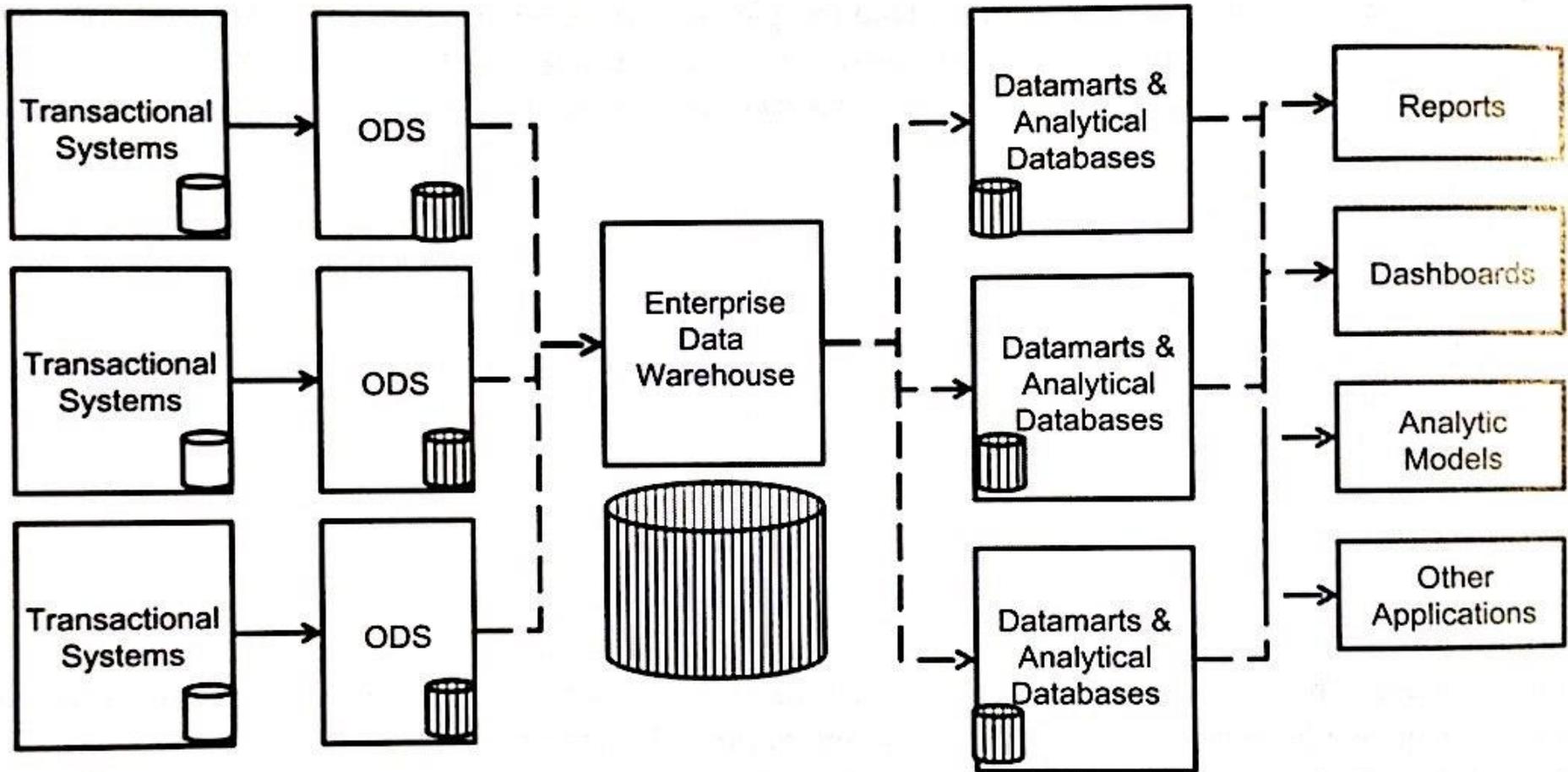
	Strategic BI	Tactical BI	Operational BI
Business focus	Achieve long-term business goals	Manage tactical initiatives to achieve strategic goals	Monitor & optimize operational business processes
Primary users	Executives & business analysts	Business analysts, & LOB managers	LOB managers, operational users & operational processes
Time-frame	Months to years	Days to weeks to months	Intra-day to daily
Data	Historical data	Historical data	Real-time, low-latency & historical data
Mode of operation	User driven Data centric	User driven Data centric	Event driven Process centric

◆ New architecture



◆ 目前企業常見的資料倉儲架構

data distribution in a data warehouse today



整合大數據與資料倉儲

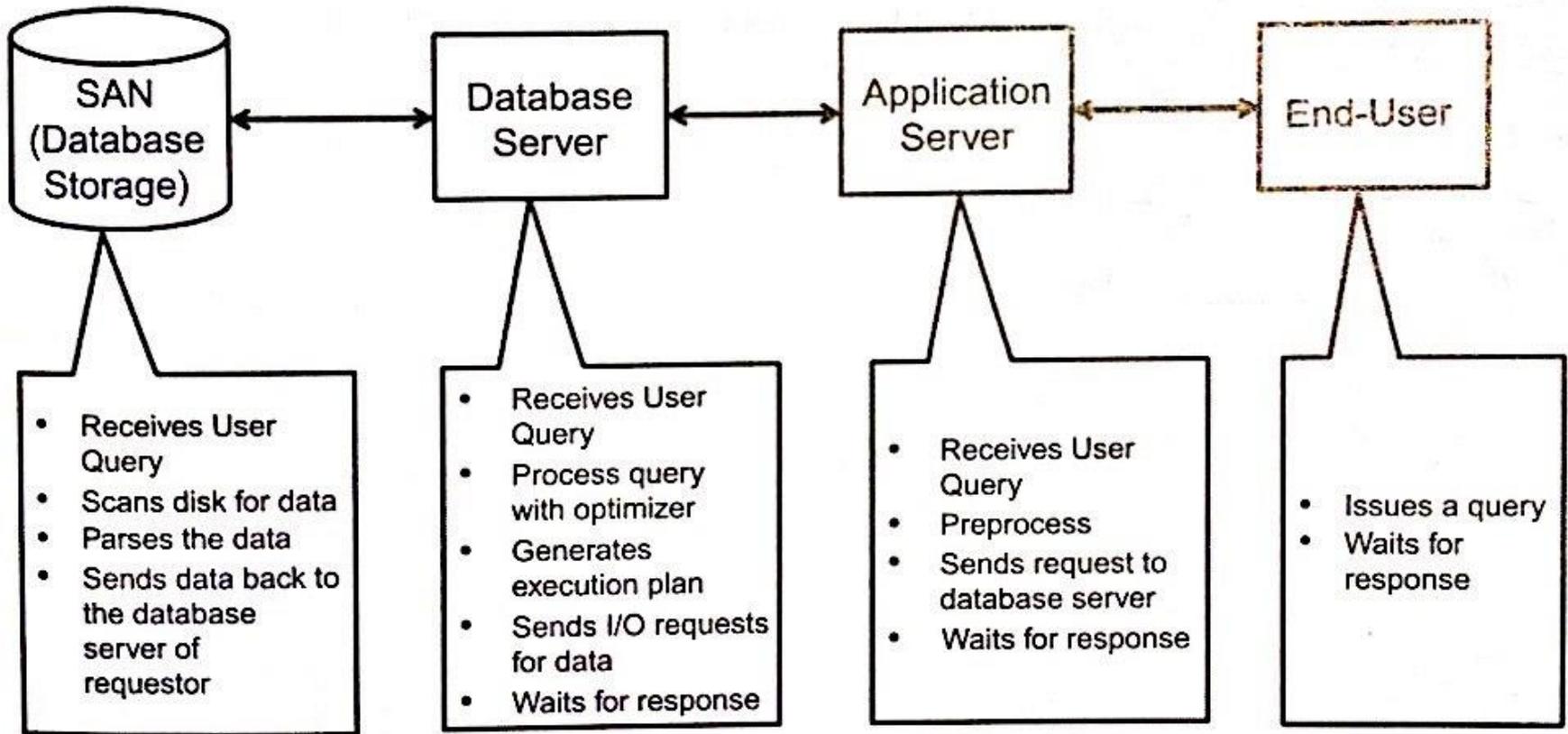
設計前題

- ◆ Identify the business objective
- ◆ Understand the workload
 - 每次查詢會需要處理的資料量, 以及回傳結果的資料量

Wide Wide	Narrow Wide
Wide Narrow	Narrow Narrow

- ◆ Understand the data
 - Unstructured
 - Semi-structured

◆ Understand the workload : 傳統的系統架構



可使用的新技术 (1) 雲端技術

◆ Cloud computing components

MANAGEMENT SERVICES:

- Security
- Usage
- Monitoring
- Service Level Agreements (SLAs)

BUSINESS APPLICATIONS & SERVICES (SaaS) :

- Web 2.0 Business Models
- Collaboration
- Social Networking
- Business Applications – HR, Sales, Supply Chain Management, ERP
- Services – Payment, Hospitality, Tourism
- Communication – Email, Skype, GoToMeeting, Join.Me
- Business Tools – Office, Notes, Calendar

PLATFORM SERVICES (PaaS):

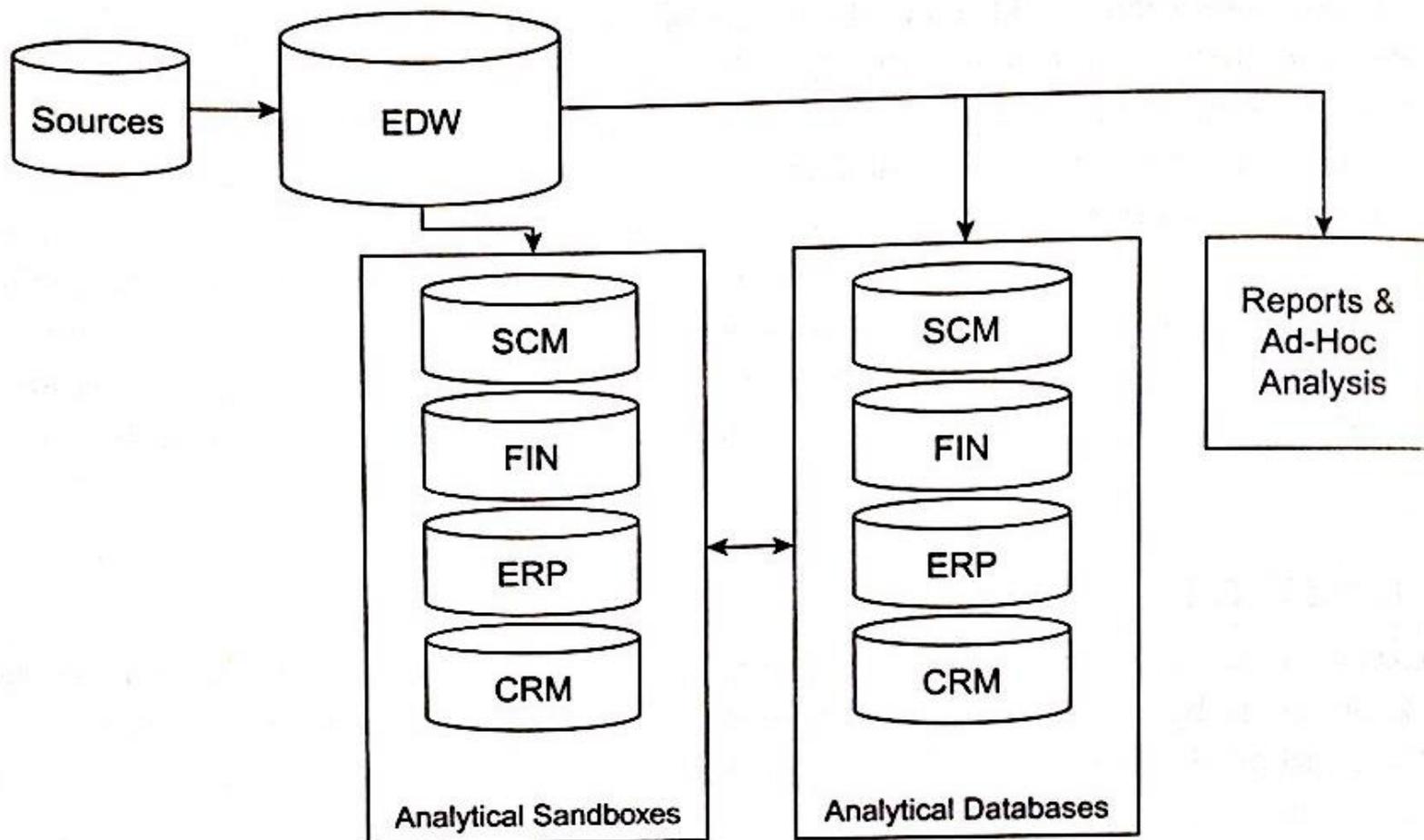
- Security – Authentication & Authorization, Single Sign-on
- Database
- Workflow Automation
- Application Development & Testing

INFRASTRUCTURE SERVICES (IaaS) :

- Server
- Storage
- Virtualization
- Management

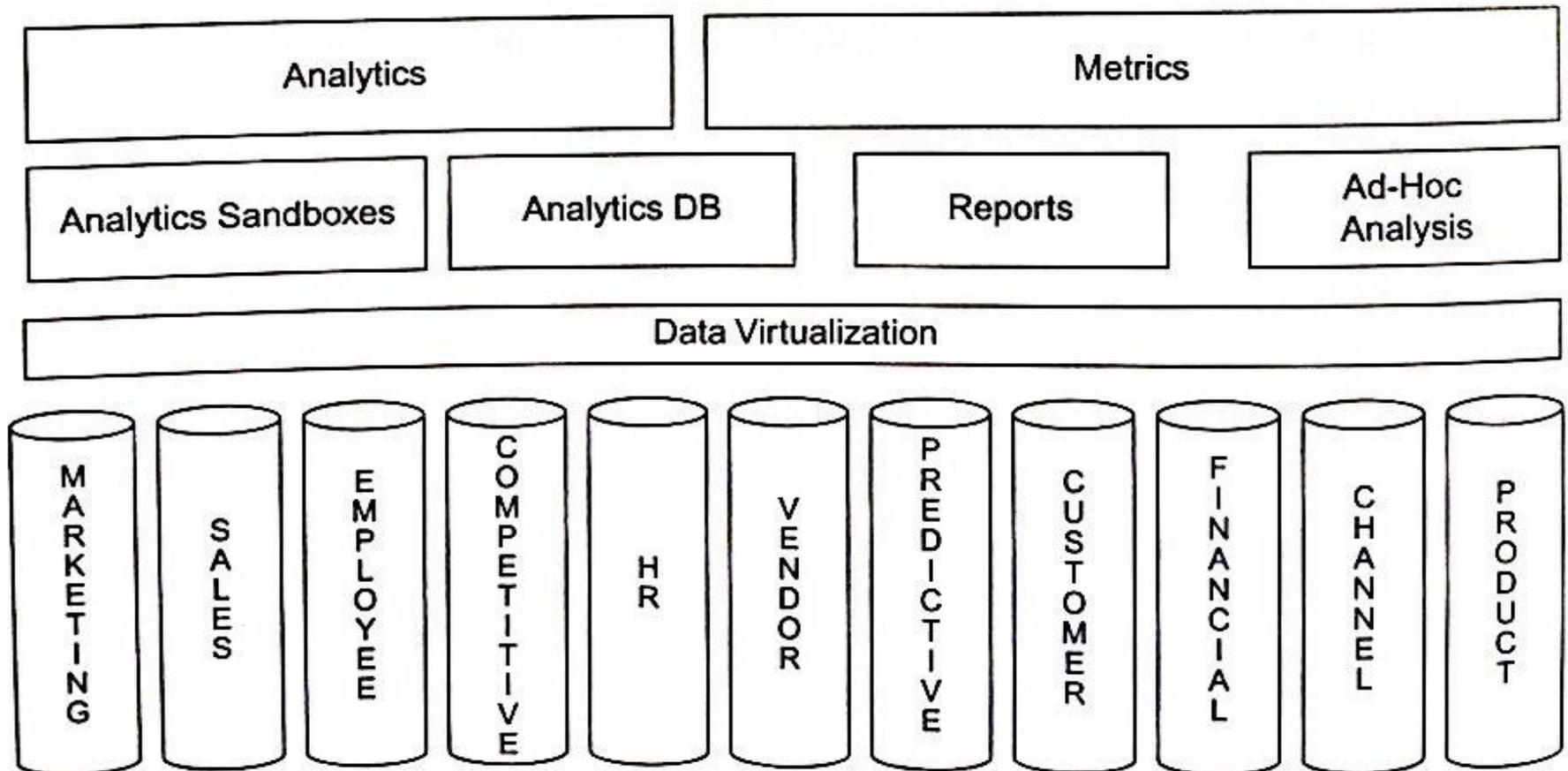
可使用的新技术 (2)

◆ Different state of analytics and reporting



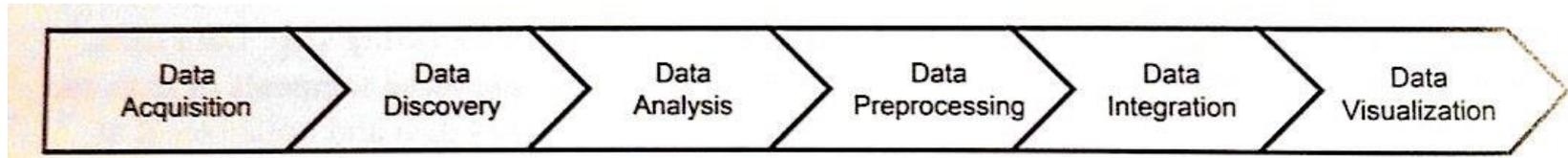
可使用的新技术 (3) 资料虚拟化

◆ Data virtualization-based architecture



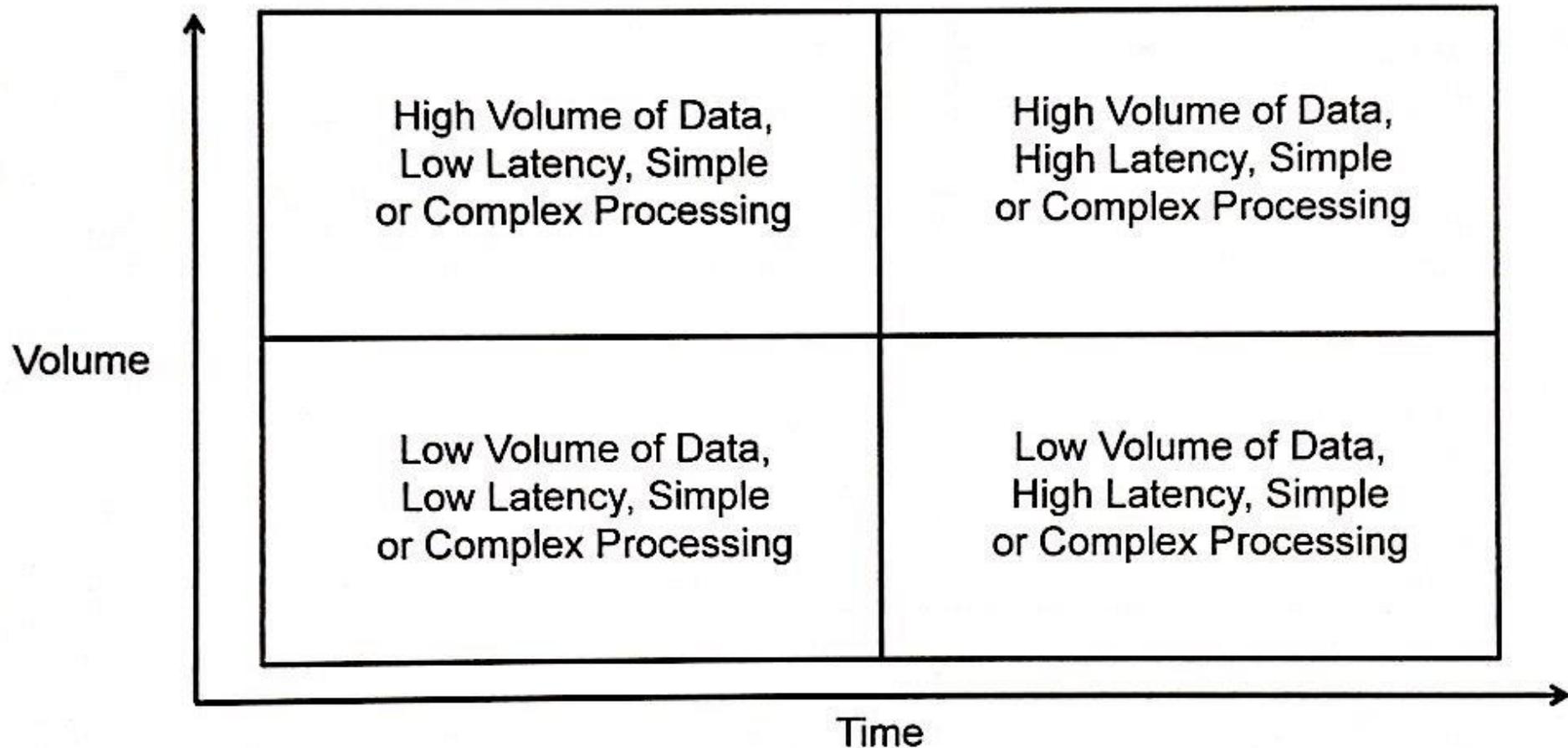
用Big data方式整合Data warehouse

- ◆ 依下列流程進行資料處理

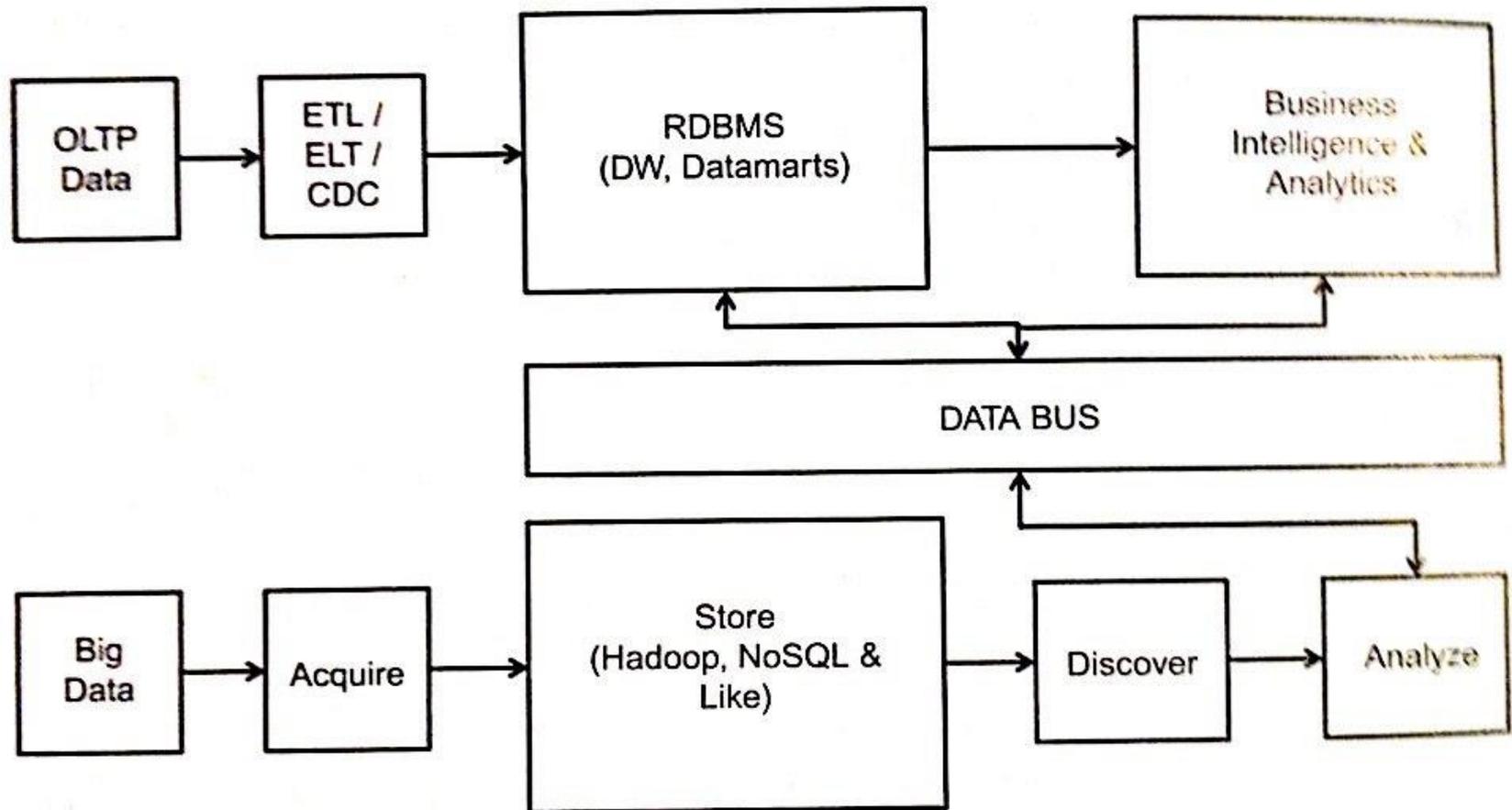


- ◆ 整合結構與非結構性資料
- ◆ 整合大數據架構與BI / Data warehouse架構

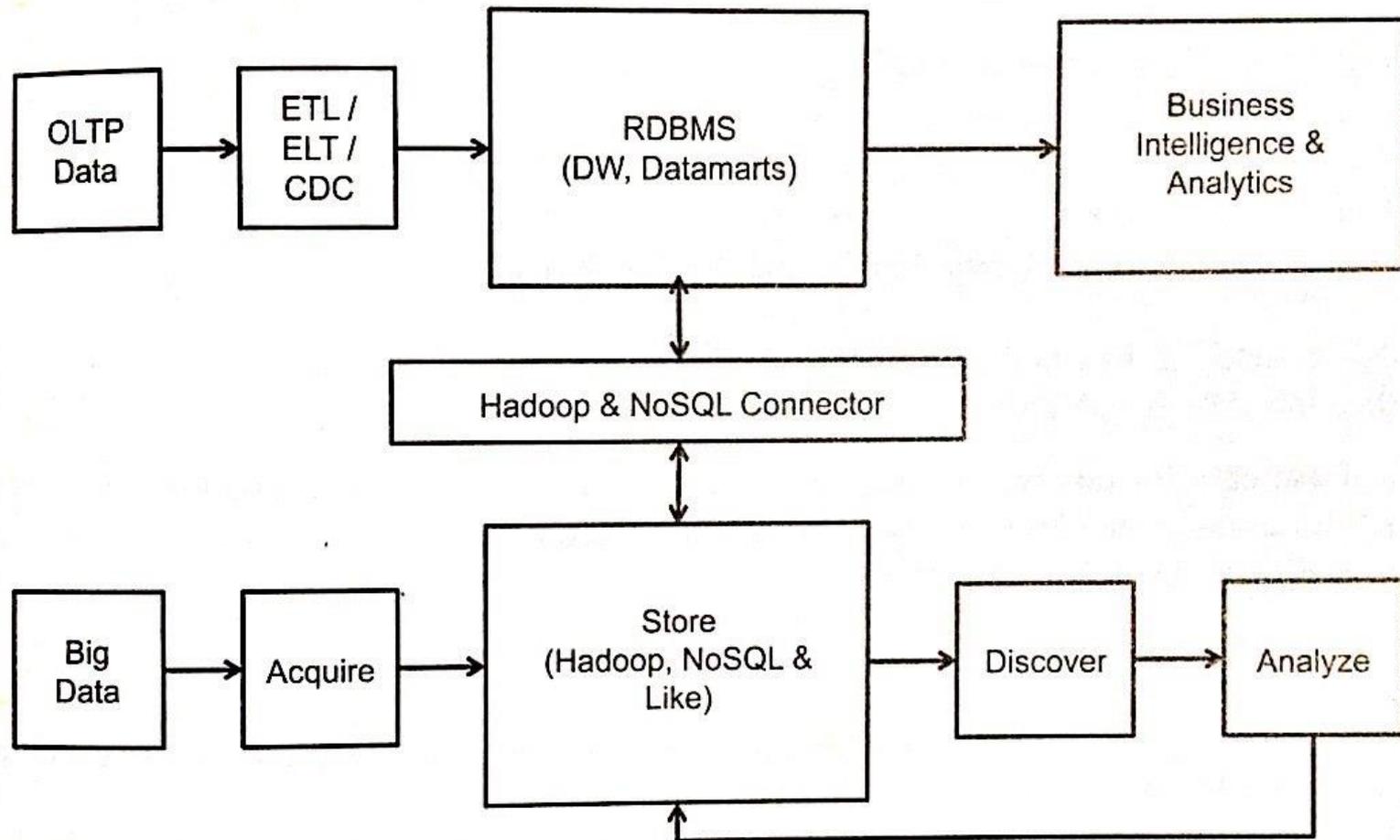
◆ Workload category



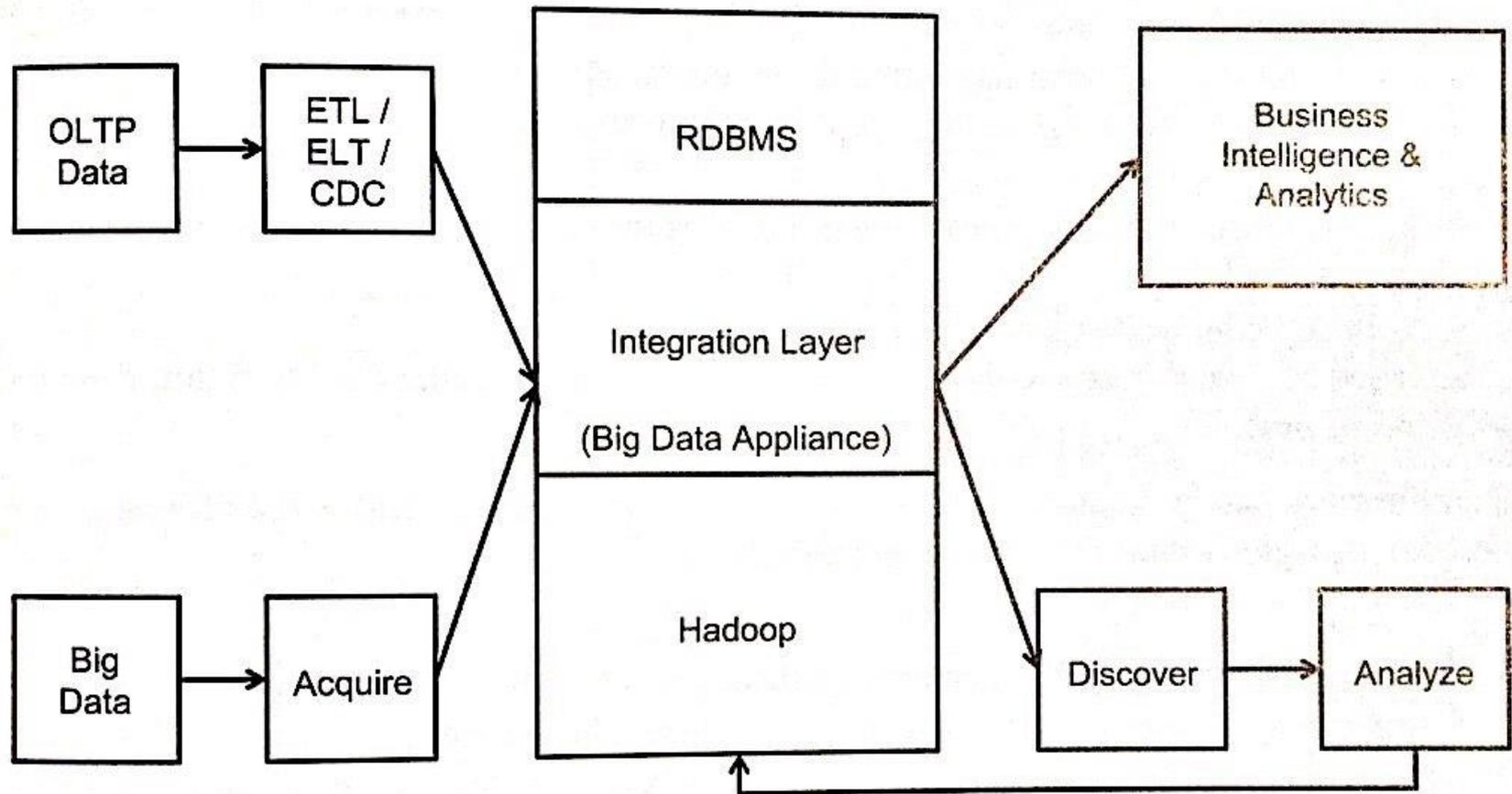
◆ External data integration



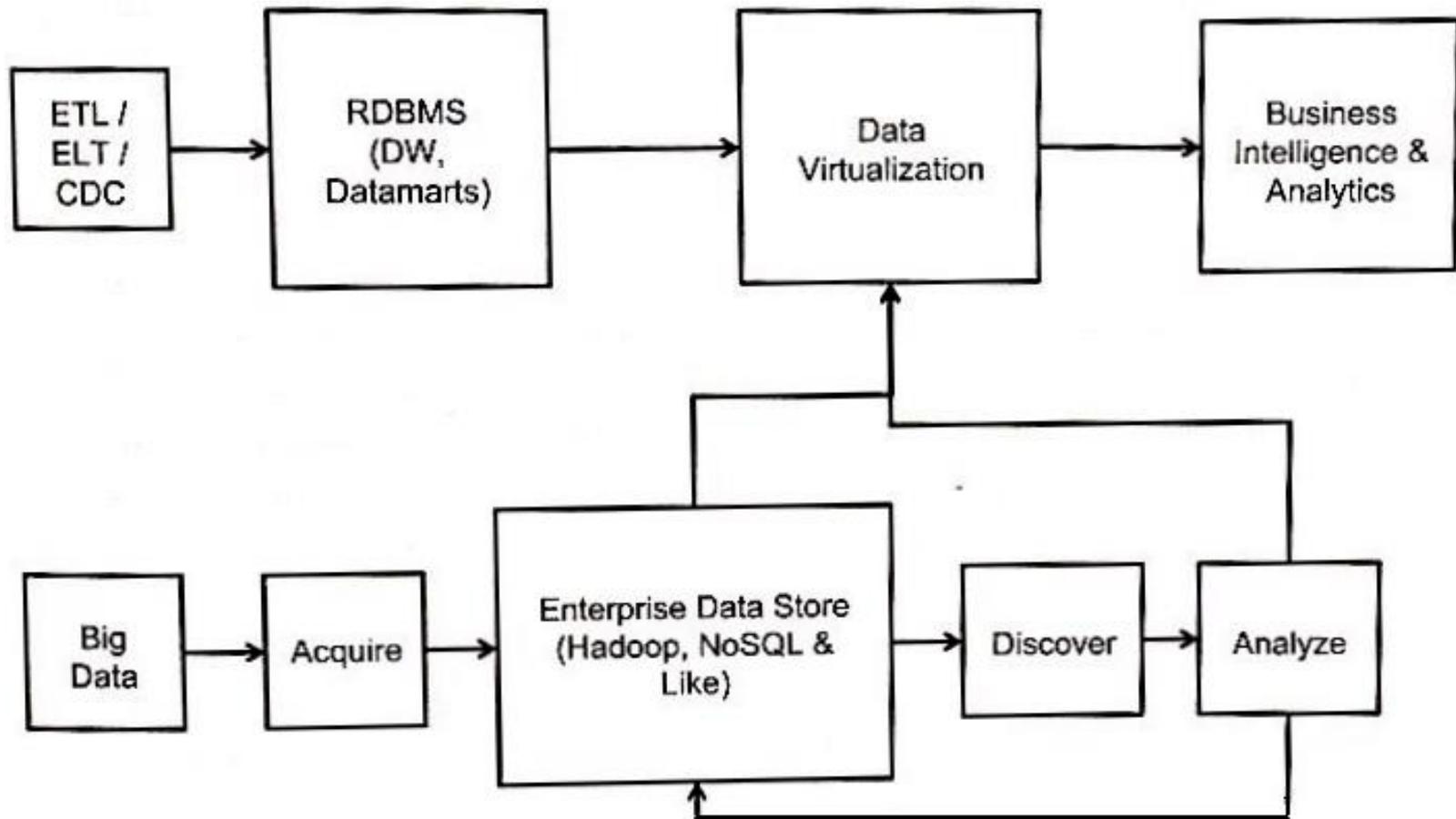
◆ Integration-driven approach



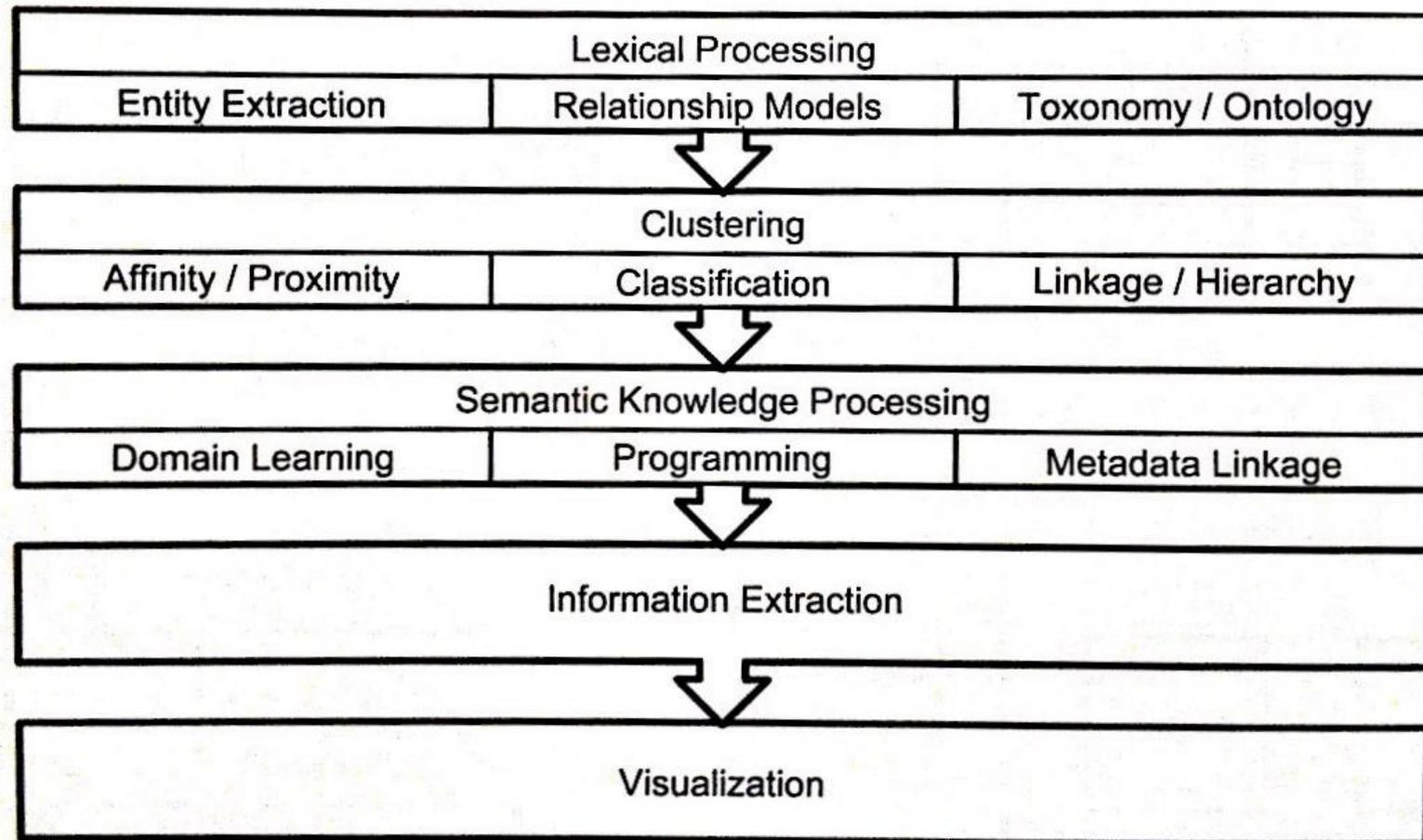
◆ Conceptual big data appliance



◆ Data virtualization-based big data integration

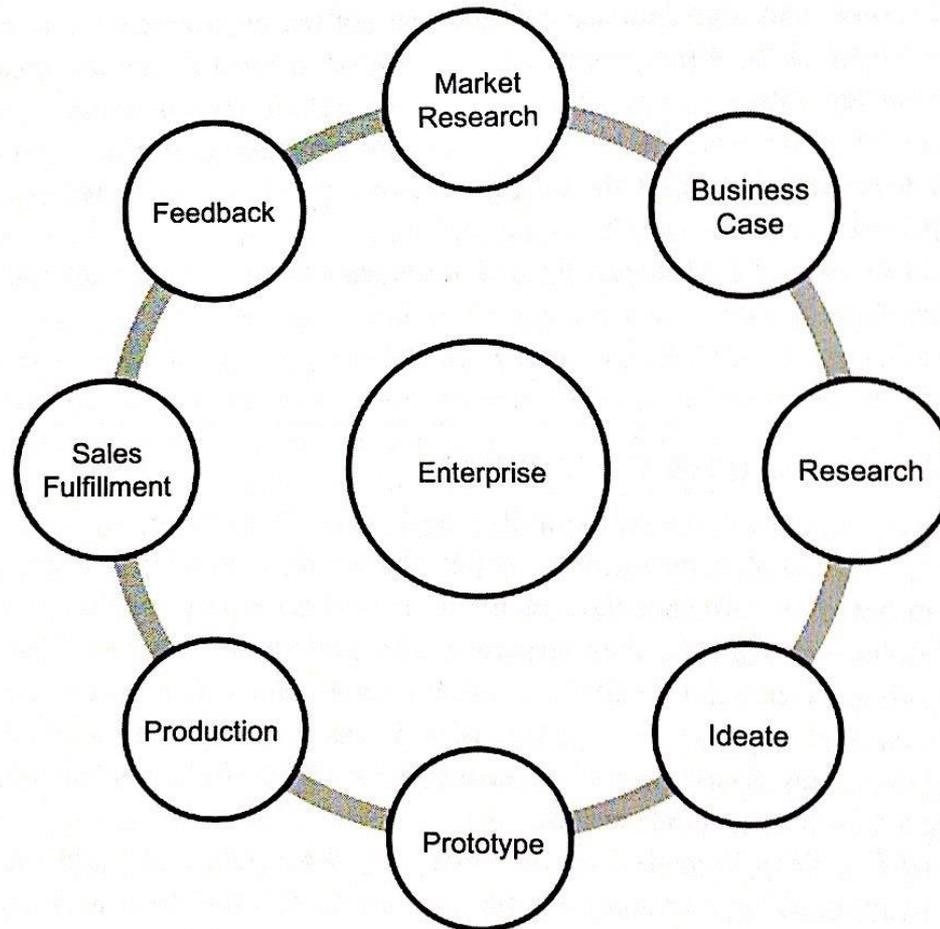


◆ Data virtualization-based : semantic framework

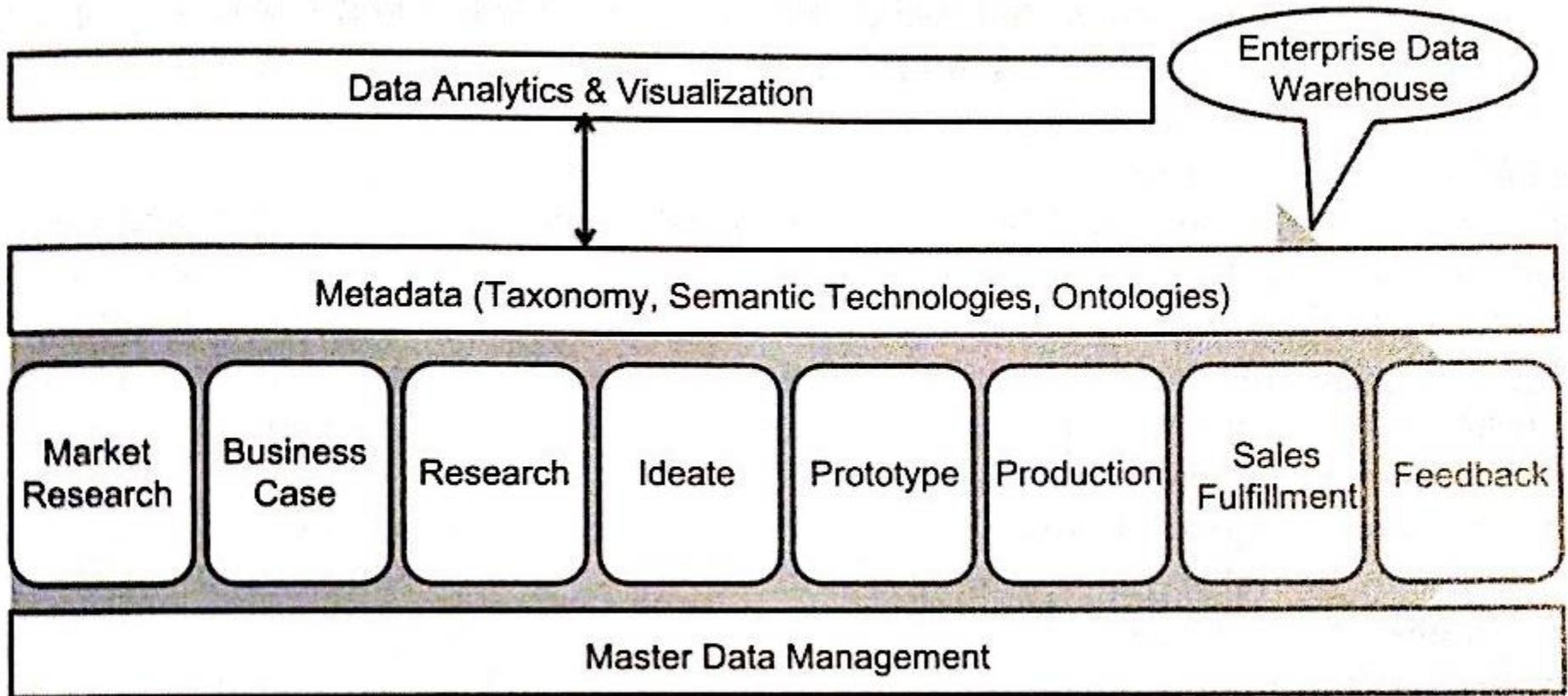


Data-driven architecture for big data

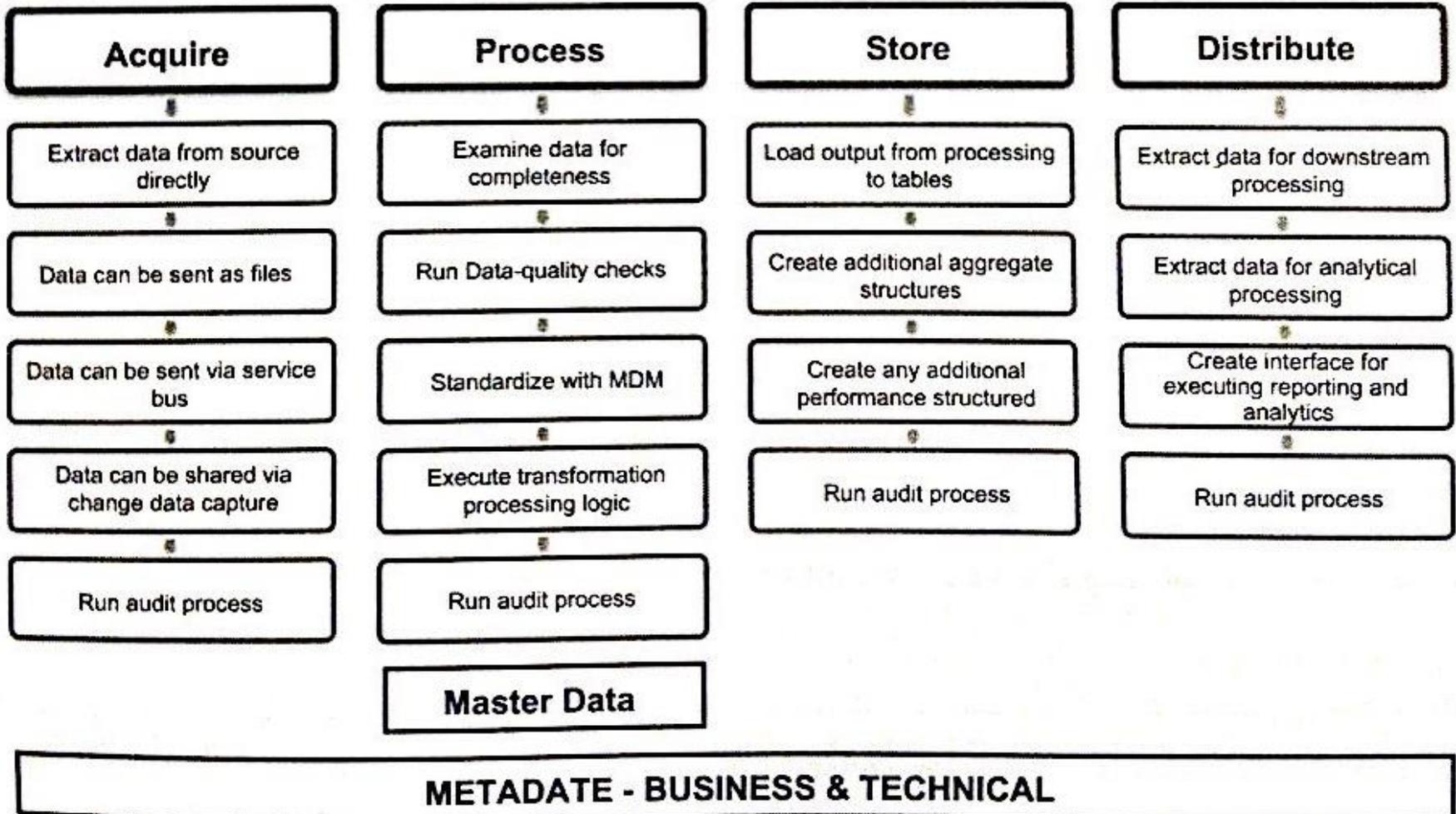
- ◆ Enterprise use of data before big data : 分散的子系統



◆ Enterprise data-driven architecture : 統一的使用方式

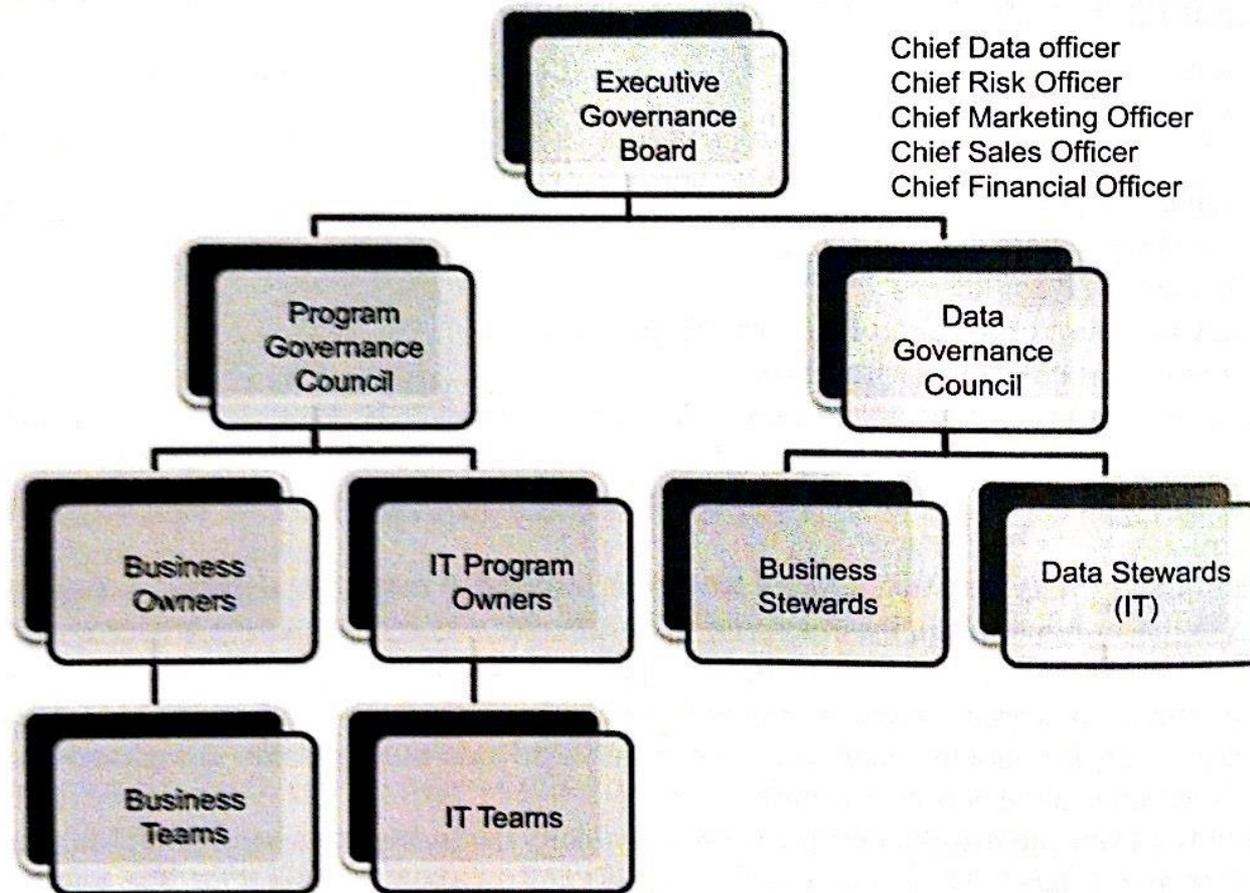


◆ Data processing cycles with integration of MDM and metadata



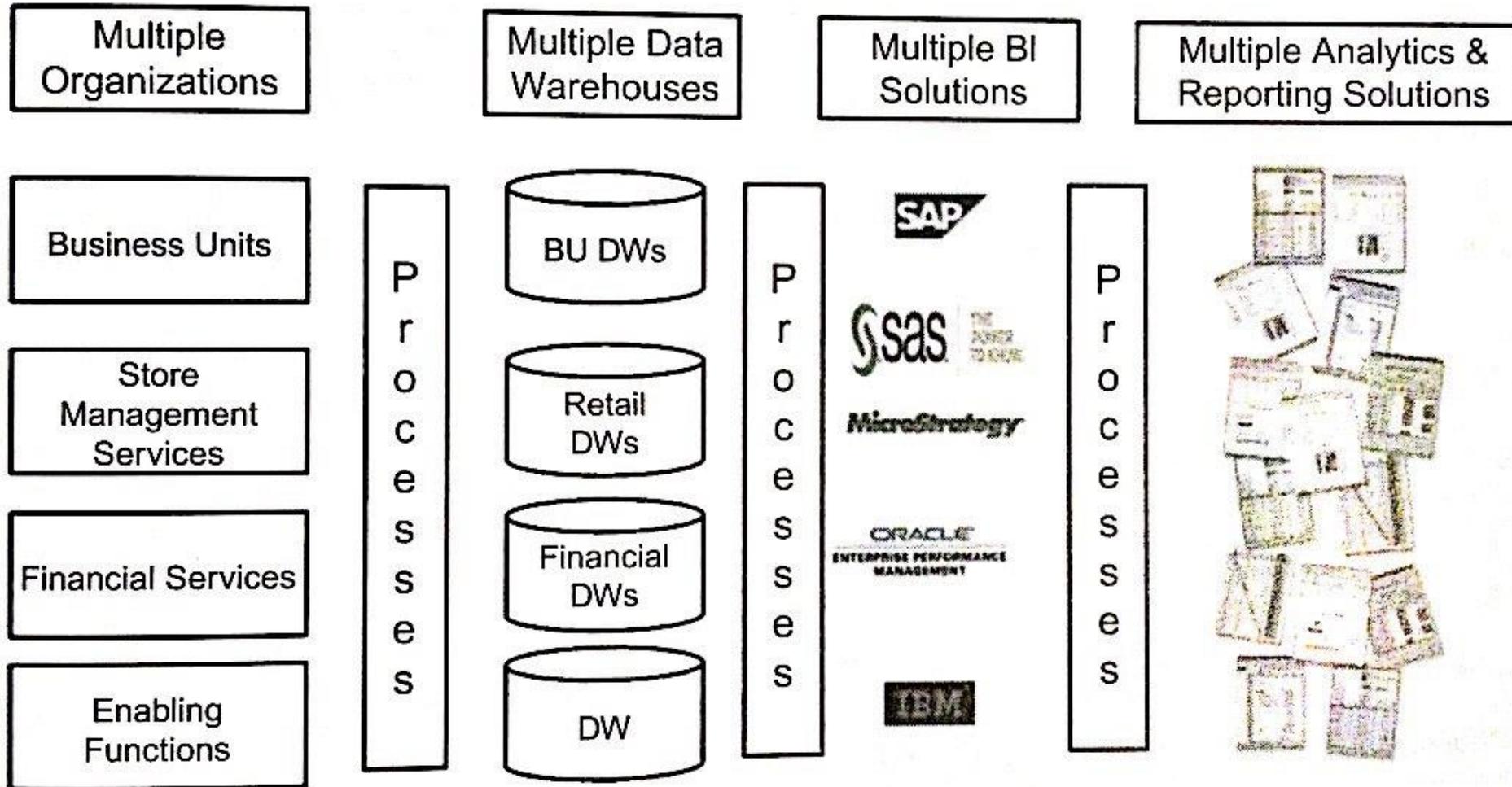
資訊管理與Life cycle of big data

◆ Data governance teams

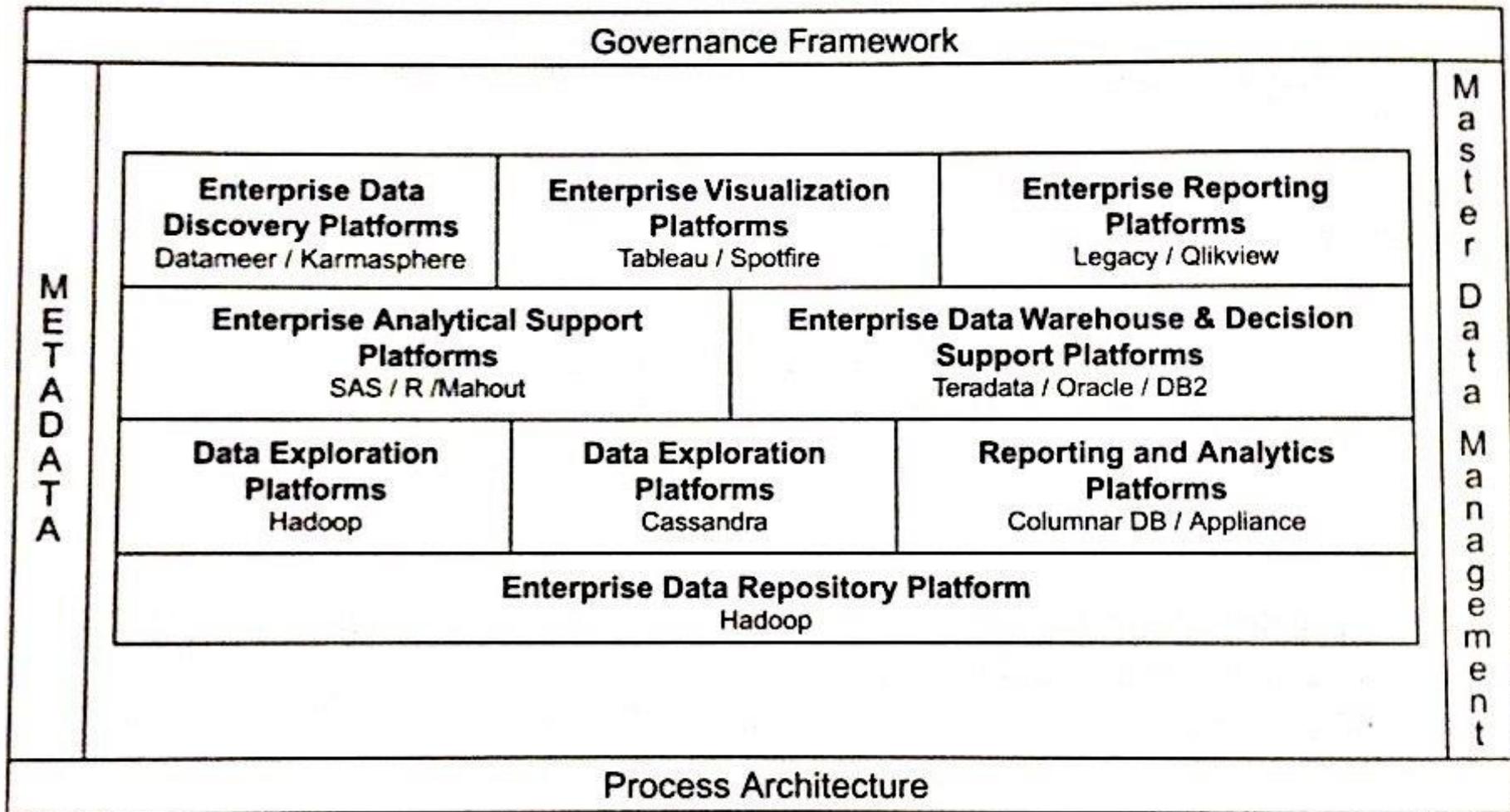


Implement the big data/DW in real life

◆ Current state architecture

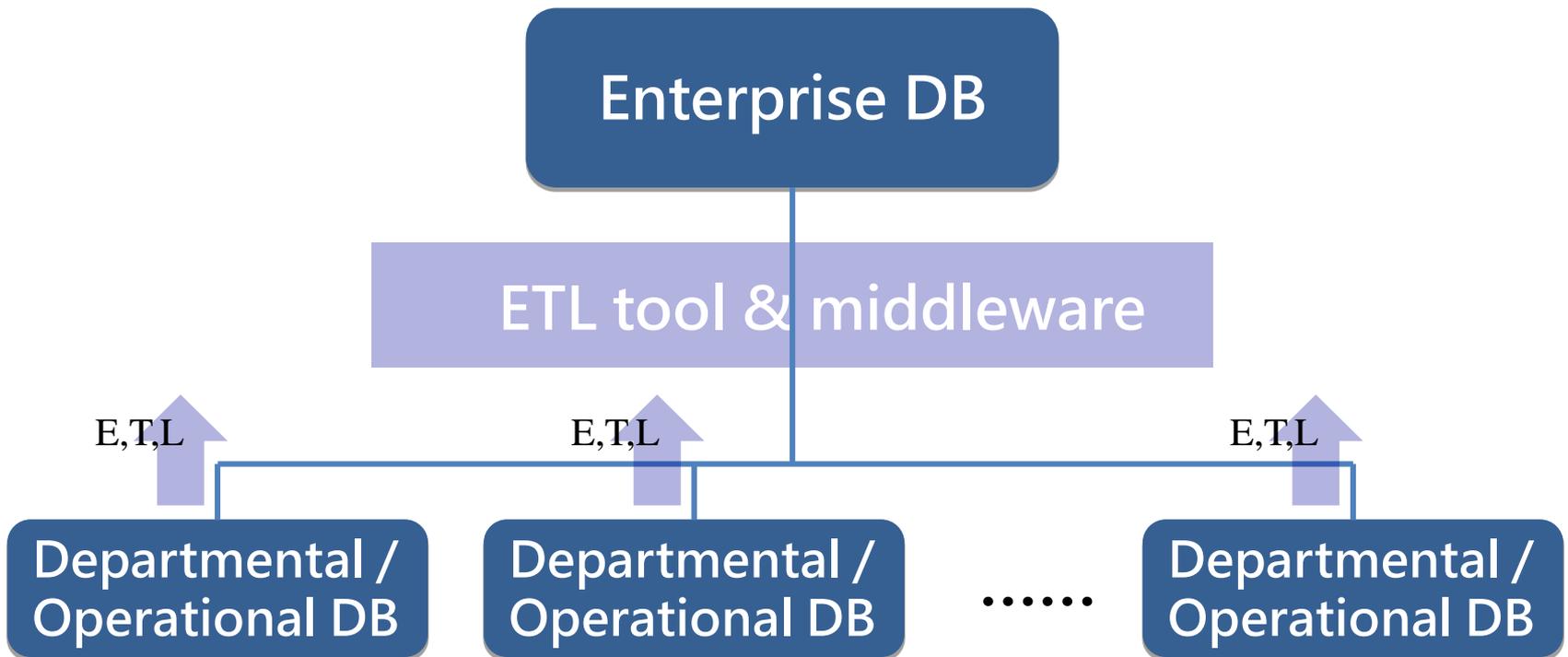


◆ Big data modular architecture conceptual model



整合大數據與資料倉儲 – 實務案例

知識整合 (1) Database approach

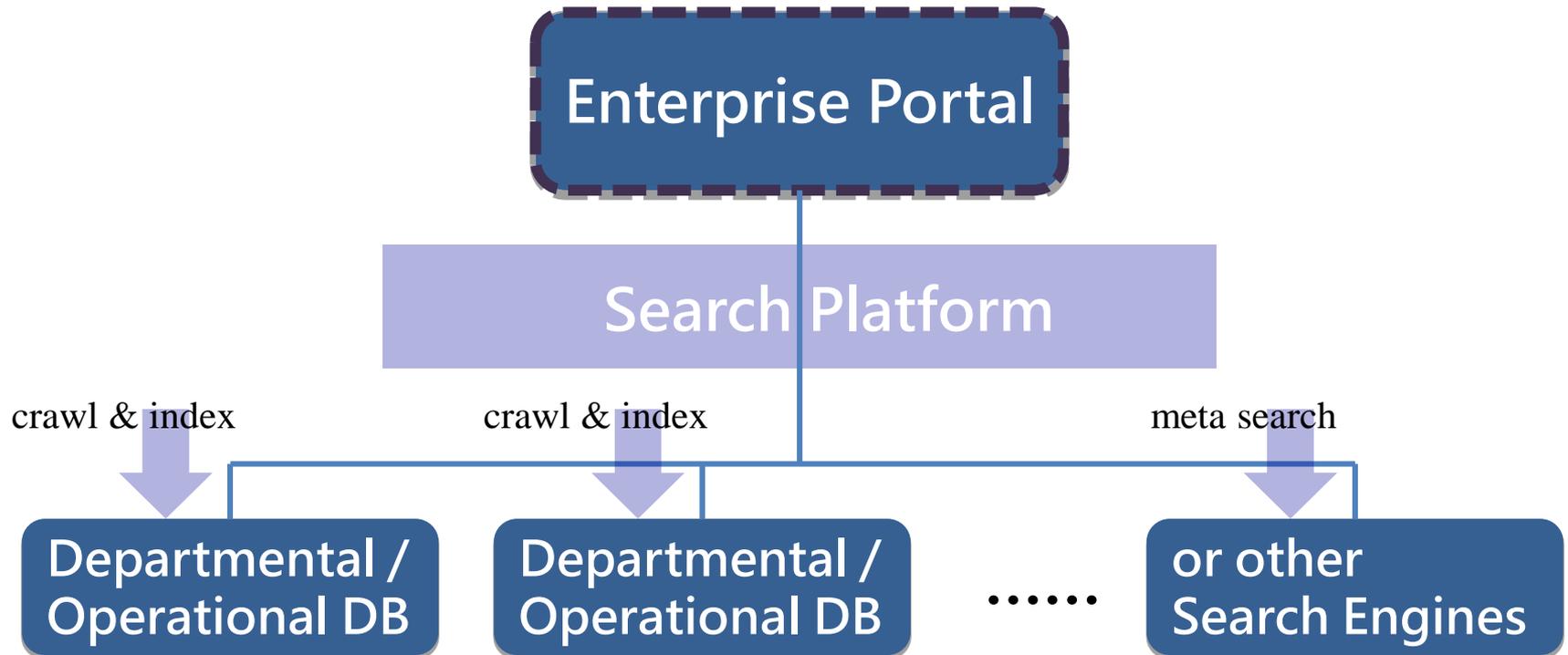


For structured data :

heterogeneous sources → ETL

data integration → middleware

知識整合 (2) Big data approach



For (un)structured data :

heterogeneous sources → crawl & index
data integration → query & result

資訊整合與重構 – 以旅遊產業為例

- ◆ Travel Vertical Portal founded in 2007
- ◆ 國內最大的旅遊搜尋入口



國內最大的旅遊商品
及旅遊情報資料庫

旅遊資訊入口

旅遊產品比價

搜尋技術匯集資訊



旅行，一定要先比價！

共89家航空公司、443,439筆資料，票價、税金一目瞭然！

◎ 依目的地搜尋 BY CITY

▶ 航班與機位查詢

出發地

目的地

出發日期 ~ 回程日期

艙等 不限 經濟艙 商務艙 頭等艙

航班 不限 來回 單程

比價來源網站：
吉蒂旅遊

約 313,736 筆資料

比價

揪團飛香港
湊團票含税\$6,379

我最低價

01 02 03

本週限時精選 ★★★★★

- 01 ▼ 比價省689 ▲ 首爾最便宜機票 含税只要\$8,036
- 02 【最低價只剩4位】直飛上海虹橋 \$9,192含税
- 03 【飛香港不到8千】全市場最低價 (還有位子哦!)

NO	航空公司	訂購網站	航段	轉機	艙等	票種	出發效期	未稅票價	税金	機票總價▲	查機位/訂購	說明	轉寄
1	菲律賓航空	玉山	來回	轉機	T經濟艙	外勞票 0天-1年	02-06 03-31	6,667	3,300	9,967			
2	國泰航空	易飛	來回	轉機	N經濟艙	3天-30天外勞票	02-05 06-30	6,634	3,489	10,123			
3	國泰航空	玉山	來回	轉機	N經濟艙	外勞票 3天-30天	02-05 06-30	6,526	3,599	10,125			
4	菲律賓航空	吉帝	來回	轉機	T經濟艙	1年/外勞票	02-06 03-31	6,465	3,670	10,135			
5	菲律賓航空	雄獅	來回	轉機	T經濟艙	1年外勞票 年票	02-06 03-31	6,567	3,647	10,214			
6	菲律賓航空	燦星	來回	轉機	T經濟艙	0天-12月 外勞票	02-06 03-31	6,567	3,647	10,214			
7	長榮航空	易遊	來回	直飛	U經濟艙	1個月外勞票	02-11 06-24	7,944	2,321	10,265			
8	長榮航空	雄獅	來回	直飛	U經濟艙	1月外勞票 旅遊票	02-11 06-24	7,944	2,321	10,265			
9	長榮航空	玉山	來回	直飛	U經濟艙	外勞票 3天-1月	02-11 06-24	7,979	2,321	10,300			
10	長榮航空	易飛	來回	直飛	U經濟艙	3天-1個月 外勞票	02-11 06-24	8,025	2,321	10,346			
11	國泰航空	雄獅	來回	轉機	N經濟艙	30天外勞票 旅遊票	02-05 06-30	6,567	3,795	10,362			

● 航班與機位查詢

精選!!

**淡季出遊
省最多
排行榜**

FunTime市集

陽明山天籟溫泉會館
午餐券

03/14開搶
\$550

◆ Data integration from various sources, both structured and unstructured data.

FUNTIME Beta
全國最大旅遊搜尋網站

免費

限定便宜 暑假出國減5000~小孩12900

加入最愛 | 首頁 | 登入/加入會員 | 忘記密碼

國內訂房比價 | 國際訂房比價 | 國際機票比價 | 今日最新折扣 | 團體/自由行比價 | 旅遊情報

預約暑假 一定要知的優惠

國內訂房		團體/自由行	國際機票		
北基桃竹苗	中彰投雲	嘉南高屏	宜花東離島		
NO.	名稱	售價	商品內容	比價	說明/訂購
1	萬里太平洋翡翠灣福華渡假飯店	雄雞:3190 定價:5800	【只要\$798/人】6月份限時專案 4人入住	✦	🛒
	台北晶華酒店	易達:4820 雄X:5799	【早餐只要10元】比價省\$979-台北知名五★	✦	🛒
1	台北八方美學商旅	易達:3552 定價:8800	【千元有找】4人同行 每人只要\$888	✦	🛒

熱門推薦

台北豪景大飯店【比價省600】
四方:2200 雄X:2800

台北歐華酒店【標準客房比價省\$512】
雄雞:2688 四星:3200

台北HOTEL QUOTE Taipei【商務客的最佳選

Innovation - reorganize all travel information

依主題或景點，將相關資訊進行重構

搜尋結果就像一本有圖文「活的旅遊雜誌」

TOP 10 一定要做的事 Must Do

NO.1 2010 盛夏京阪神紀行--強烈推薦到 gourmetCity 買飲料
Emily 於 2010-08-27 發表
在日本，要喝乾淨的水很方便，水龍頭一轉開就有了。還記得年輕時候，第一次跟團到北海道，導遊跟我們講說，結東喝開水，飯店浴室的水龍頭一轉開就有水喝了，而且...

NO.2 [遊記] 一個人的關西 DAY2-大阪(四天王寺、通天閣)...

NO.3 [遊記] 新手京阪行-清水寺、柳園、心齋橋

NO.4 [遊記] 新手京阪行-心齋橋、道頓堀、難波

NO.5 [大阪道頓堀] 本家章魚燒，用料毫不手軟的著名小吃

NO.6 2010 京阪奈之秋 - Day1 - 錦市場之大白天就喝大...

NO.7 PP MM BB 三人行-日本京都&大阪自由7天【之四...】

NO.8 《關西自助記行07》Day2-4：京蘇・清水寺

NO.9 [遊記] 2010 關西-Day6 京都金閣寺、龍安寺

NO.10 [遊記] 12/5 京都賞楓-平安神宮/南禪寺/永觀堂



FUNTIME Beta
全國最大旅遊搜尋網站

旅遊市集下殺3折起

加入最愛 | 首頁 | 登入/加入會員 | 忘記密碼

國內訂房比價 國際訂房比價 國際機票比價 今日最新折扣 團體/自由行比價 找行程安排

Facebook Plurk 部落格

首頁 > 找行程 > 大阪&京都

京都 & 大阪
KYOTO & OSAKA

現在均溫 4°C
跟我這樣穿就對了!

秒遊京阪
: 90天皆可免費讀
: 日本語
: 100V, 插頭形狀與台灣相同
: 1台幣=2.6日圓
: UTC +9:00, 較台灣快一小時

快速前往 >> 一定要做的事 / 老饕說 / 行程這樣安排 / 最火的話題 / 巴士地鐵這樣搭

比價最便宜 Bon Prix

機票						行程		
航空	票價	稅金	總價	說明	訂購	商品名稱	每人價格	說明
airline	price	tax	total	desc	buy	product	price	desc
	9,405	2,603	12,008			中華航空大阪新大阪車站自由5天	15,500	
	9,405	2,649	12,054			中華航空【呼朋引伴】大阪自由4天(3人...)	15,900	
	9,450	2,610	12,060			中華航空大阪梅田自由5天	15,900	

更多 More



Let's Have FunTime Together

老饕說 Cuisine



[食記] 大阪梅田車站・松葉食車炸初體驗 (地下店)-立

大家行程這樣安排 Itinerary

2010.12.06 大阪 北海道賞楓看雪之旅 ~ 北海道 札幌

結束我們5天4夜的北海道自駕之旅趁著還有半天的空檔我們安排了 "Asahihon 朝日啤酒工廠見習之旅" 這個行程一直...

共 0 則回文 huivuan 於 2011-01-06 發表

其他更多:

- 【行程】大阪、神戶、京都五天四夜行程規劃詳誌
- [遊記] HOTEL VISTA GRANDE 大阪市區住宿首選
- [遊記] 2010 DEC 京都賞楓六天五夜行程

京都金閣寺

金閣寺，基本款，很多介紹金閣寺的照片都是差不多這個角度金閣寺，算是側面吧，屋頂上那隻鸕可以看的比較清...

共 0 則回文 Pearlwen 於 2009-04-02 發表

其他更多:

- [旅人] 1/29-2/7期間 京都大阪自由行徵男旅伴
- [食記] 京洋 子可
- 京都住宿

京都自由行 2010/11/19: 漫步三条

也是行程的第七日瑪小姐早上拜訪過二條城後下午依計劃就是在京都小巷中漫步遊逛小店 簡單說~就是購物行程...

共 0 則回文 marissa-blog 於 2010-12-18 發表

其他更多:

- [嚴利品] - 京阪神家買速遞-公仔篇
- [京都] 親子井名店-西陣 烏岩樓 (おしじん どり)
- [實況] 京都市明年四月一日起開徵入湯稅

» 問題討論