



# 線上資料分析處理

---

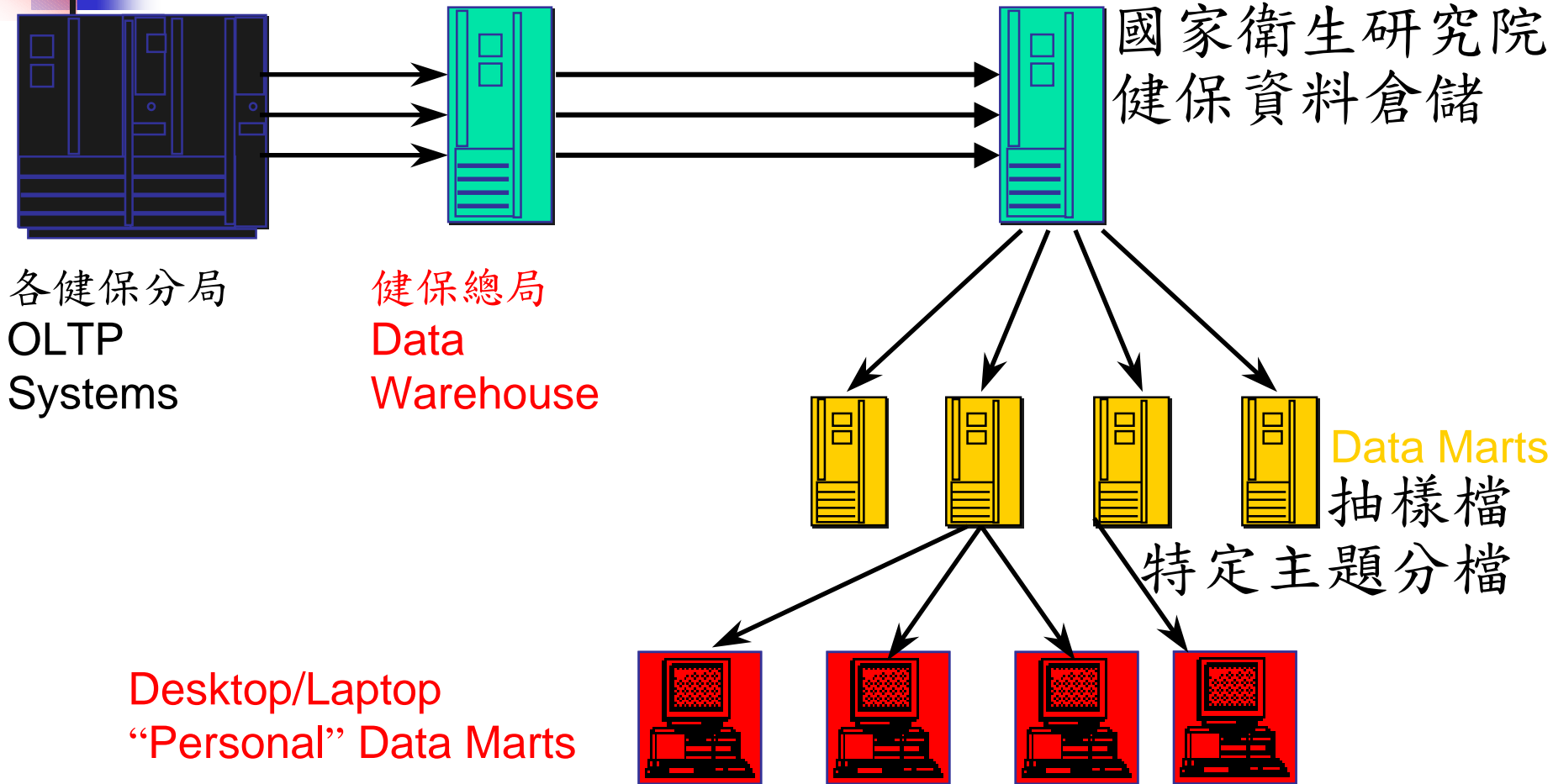
劉德明

國立陽明大學

衛生資訊與決策研究所副教授

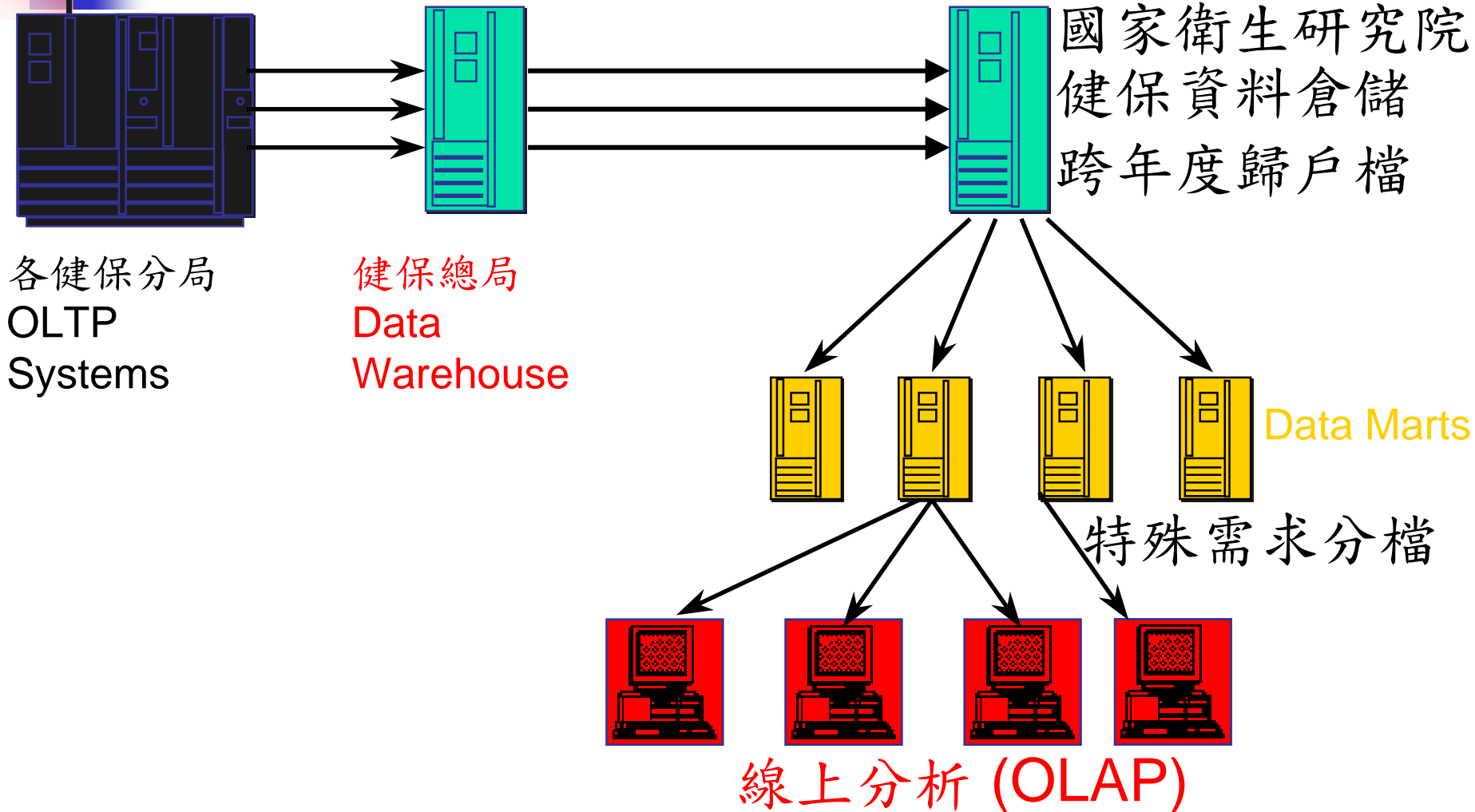


# 健保資料倉儲現況





# 健保資料倉儲未來





# 資料倉儲的用途

---

- 線上分析系統 (On-Line Analytical Processing)
- 決策支援系統 (Decision Support System)
- 資料探勘 (Data mining)

A decorative graphic consisting of overlapping yellow, blue, and red squares with a black crosshair.

# 資料探勘

---

- Knowledge discovery
- Technologies
  - Classification rules
  - Association rules
  - Sequence rules
  - Similar time series
  - Clustering rules



# 資料探勘的步驟

---

- Selection and sampling
- Preprocessing and cleaning
- Transformation and reduction
- Data mining
  - Pattern
- Visualization
- Evaluation



# 從資料表格建立資料方塊

- A data warehouse is based on a **multidimensional data model** which views data in the form of a data cube
- A data cube, such as 就醫人數, allows data to be modeled and viewed in multiple dimensions
  - Dimension tables, such as Hospital (item\_name, brand, type), or time(day, week, month, quarter, year) or Patient(age,sex,address)
  - Fact table contains measures (such as 門診處方及治療明細檔) and keys to each of the related dimension tables



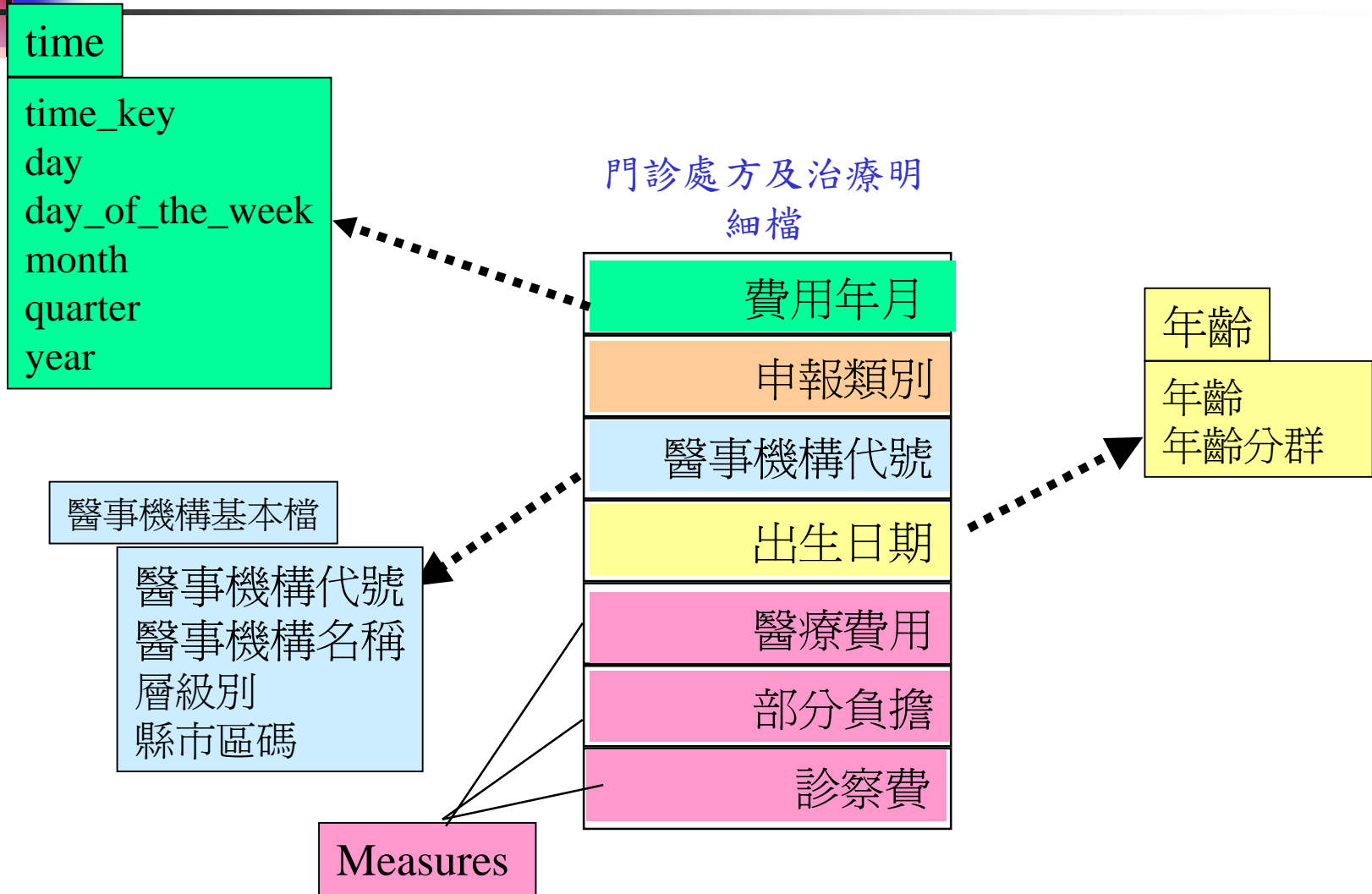
# 資料倉儲的結構模式

- Modeling data warehouses: dimensions & measures
  - Star schema: A fact table in the middle connected to a set of dimension tables
  - Snowflake schema: A refinement of star schema where some dimensional hierarchy is **normalized** into a set of smaller dimension tables, forming a shape similar to snowflake



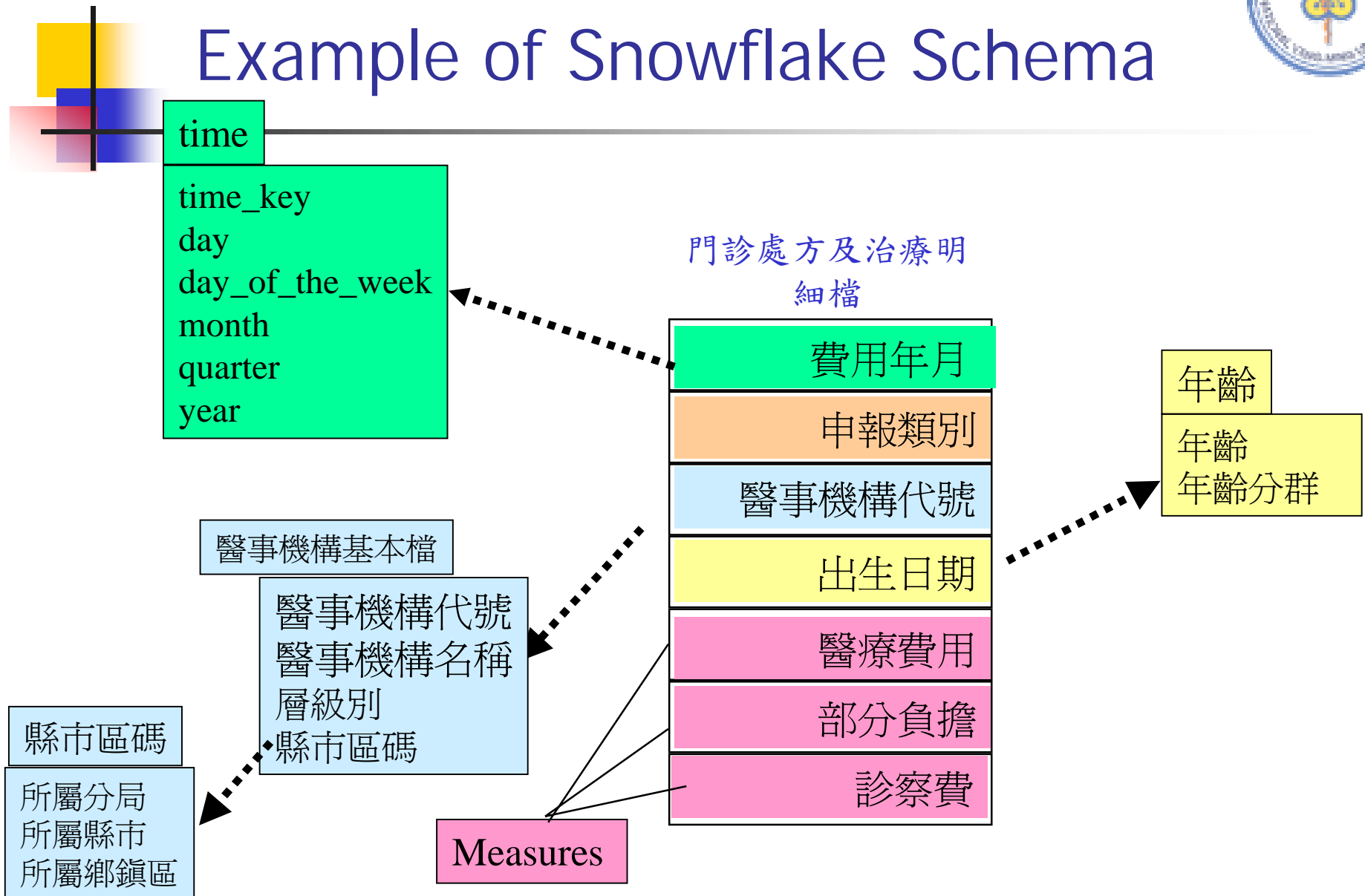


# Example of Star Schema

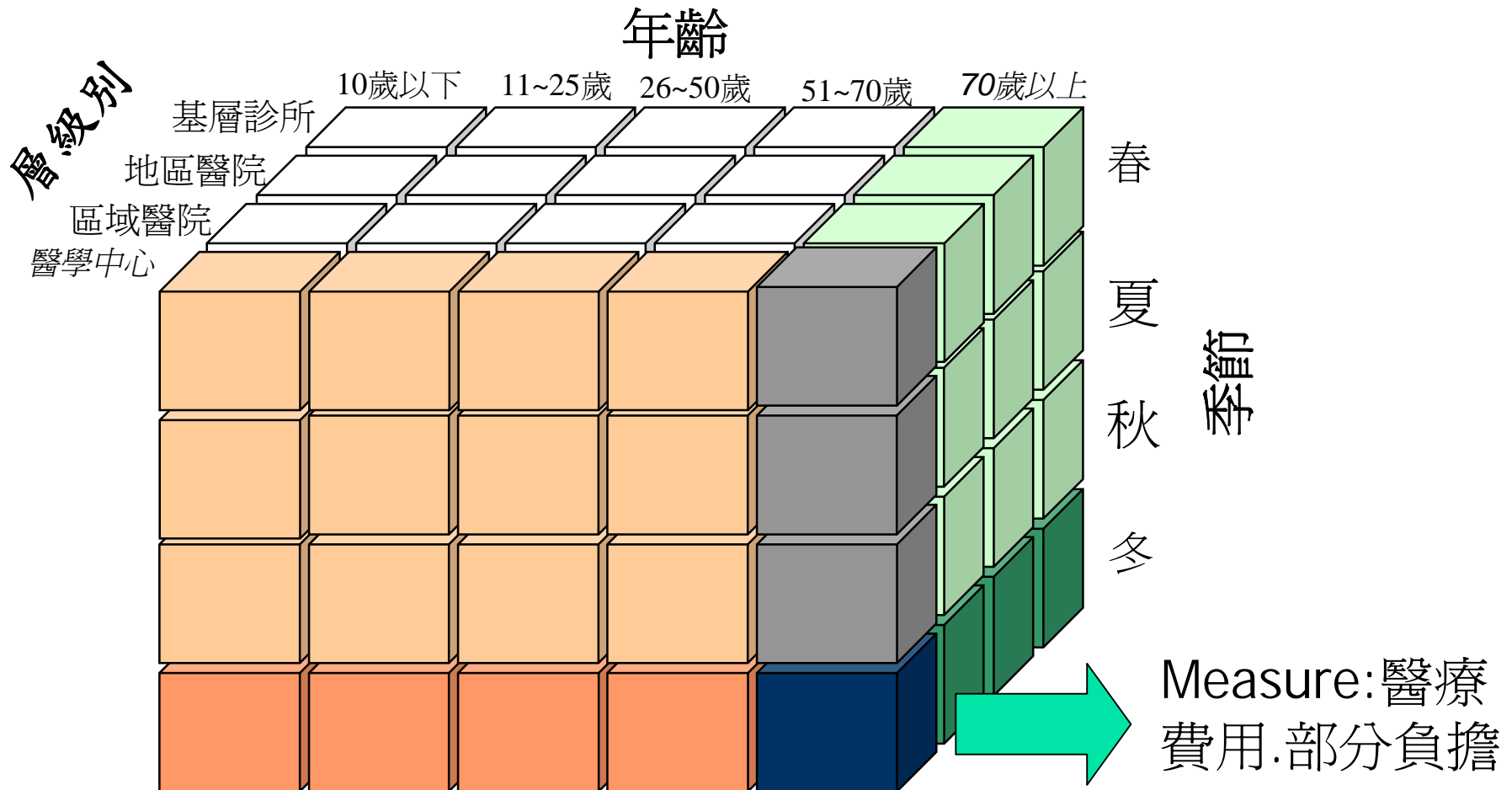




# Example of Snowflake Schema



# 門診醫療費用資料方塊架構圖





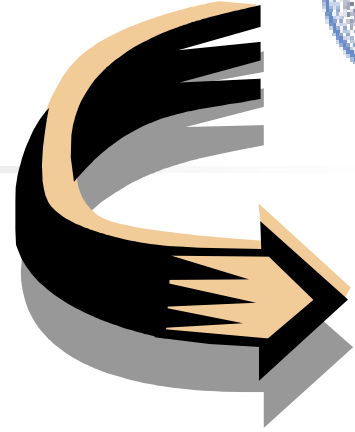
## 案例討論 – 專案規劃



- 健保資料倉儲專案定義
- 預期目標
  - 提供使用者一個簡易不需預設的使用環境
  - 提供彈性、易於開發不影響日常交易行為的資料倉儲架構
  - 提供網際網路系統架構
  - 提供線上分析(On Line Analytical Processing; OLAP)的使用者介面



# 案例討論 - 需求規劃



- 關鍵效能指標
  - 住院費用分析
  - 門診費用分析
- 加入疾病分類分析
- 醫事機構累積次數、醫事人員累積次數、就診或住院累積次數



# 案例討論 – 技術架構設計

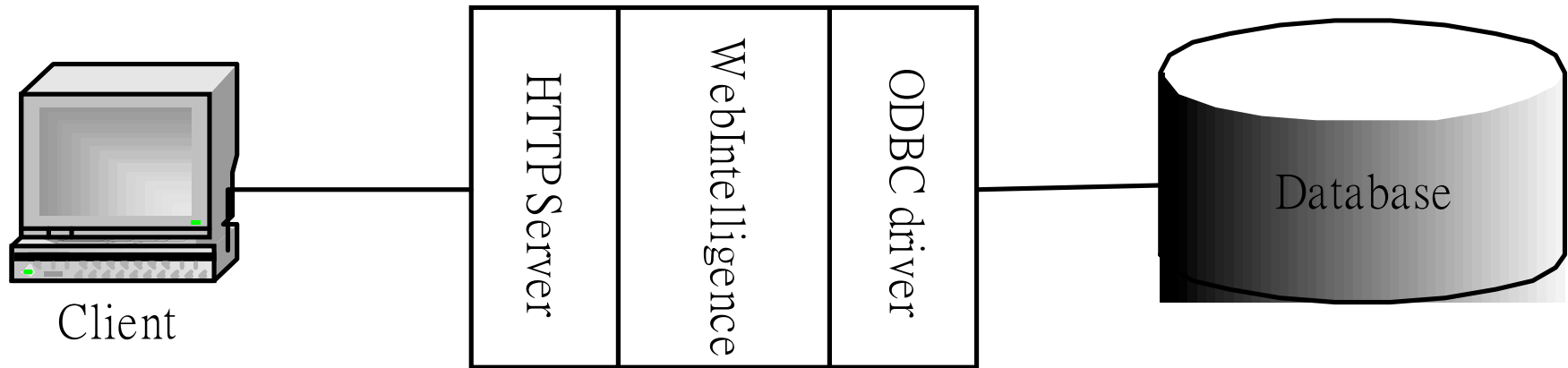
- 資料量評估
- 硬體設備評估
- 三層式設計架構
- 資料庫、資料集結使用工具
- 線上分析設計工具
- 網際網路報表工具
- 前端瀏覽器

# 案例討論 – 原始資料檔

檔案類別	來源資料中文檔名	英文檔名	資料筆數	檔案大小
基本資料檔	醫事機構基本資料檔	HOSB1997	27,441	4.9
	醫事機構副檔資料檔	HOSX1997	25,088	5.6
	醫事機構診療科別明細檔	DETA1997	27,233	0.8
	醫事機構病床主檔	BED1997	212,739	12
	醫事人員基本資料檔	PER1997	149,053	12.8
	專科醫師證書主檔	DOC1997	61,153	4.1
	重大傷病證明明細檔	HV1997	627,842	146.7
	門診費用申請總表主檔	CT1997	197,506	65.5
	住院費用申請總表主檔	DT1997	12,758	2.9
系統抽樣檔*	門診處方及治療明細檔	S_CD1997	585,789	117.3
	門診處方醫令明細檔	S_OO1997	2,063,586	188.9
	住院醫療費用清單明細檔	S_DD1997	117,647	47.7
	住院醫療費用醫令清單明細檔	S_DO1997	6,320,416	482.2
特定主題分檔	特約藥局處方及調劑明細檔	GD1997	2,068,195	258.4
	特約藥局處方及調劑醫令檔	GO1997	8,596,370	623.1



# 案例討論 – 三層式設計架構



PresentationLayer

BusinessLogicLayer

DataAccessLayer



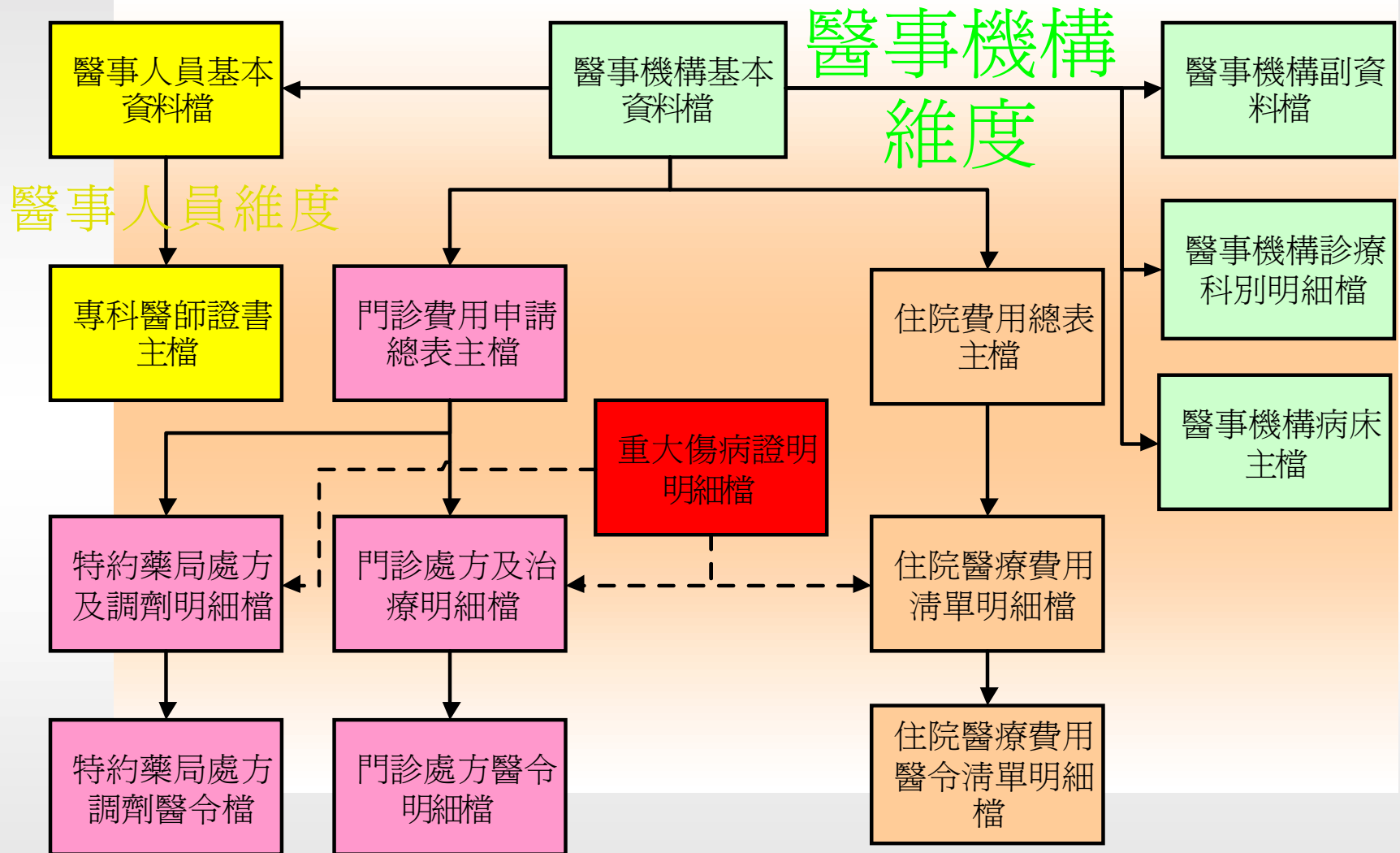


# 案例討論 – 工作平台的選擇及安裝

- 資料庫、資料集結使用工具
  - MS SQL Server 7.0 Sp2, Microsoft Data Transformation Services (DTS)。
- 線上分析設計工具
  - Business Objects 5.0
- 網際網路報表工具
  - WebIntelligent 2.5
- 網際網路瀏覽器



# 案例討論 – 原始資料結構



# 案例討論 – 使用之資料檔及欄位 (一)

檔名	使用欄位	分析說明
DD1997	ID(身分證統一編號) ID_BIRTHDAY(出生年月日) ID_SEX(性別) CARD_SEQ_NO(健保卡就醫序號) ICD9CM_CODE(主診斷代碼)	病患維度
	DIAG_AMT(診察費) ROOM_AMT(病房費) MEAL_AMT(管灌膳食費) AMIN_AMT(檢查費) RADO_AMT(放射線診療費) THRP_AMT(治療處置費) SGRY_AMT(手術費) PHSC_AMT(復健治療費) BLOD_AMT(血液透析費) HD_AMT(血液血漿費) ANE_AMT(麻醉費) METR_AMT(特殊材料費)	以各項費用為事實表格(fact table)的中心

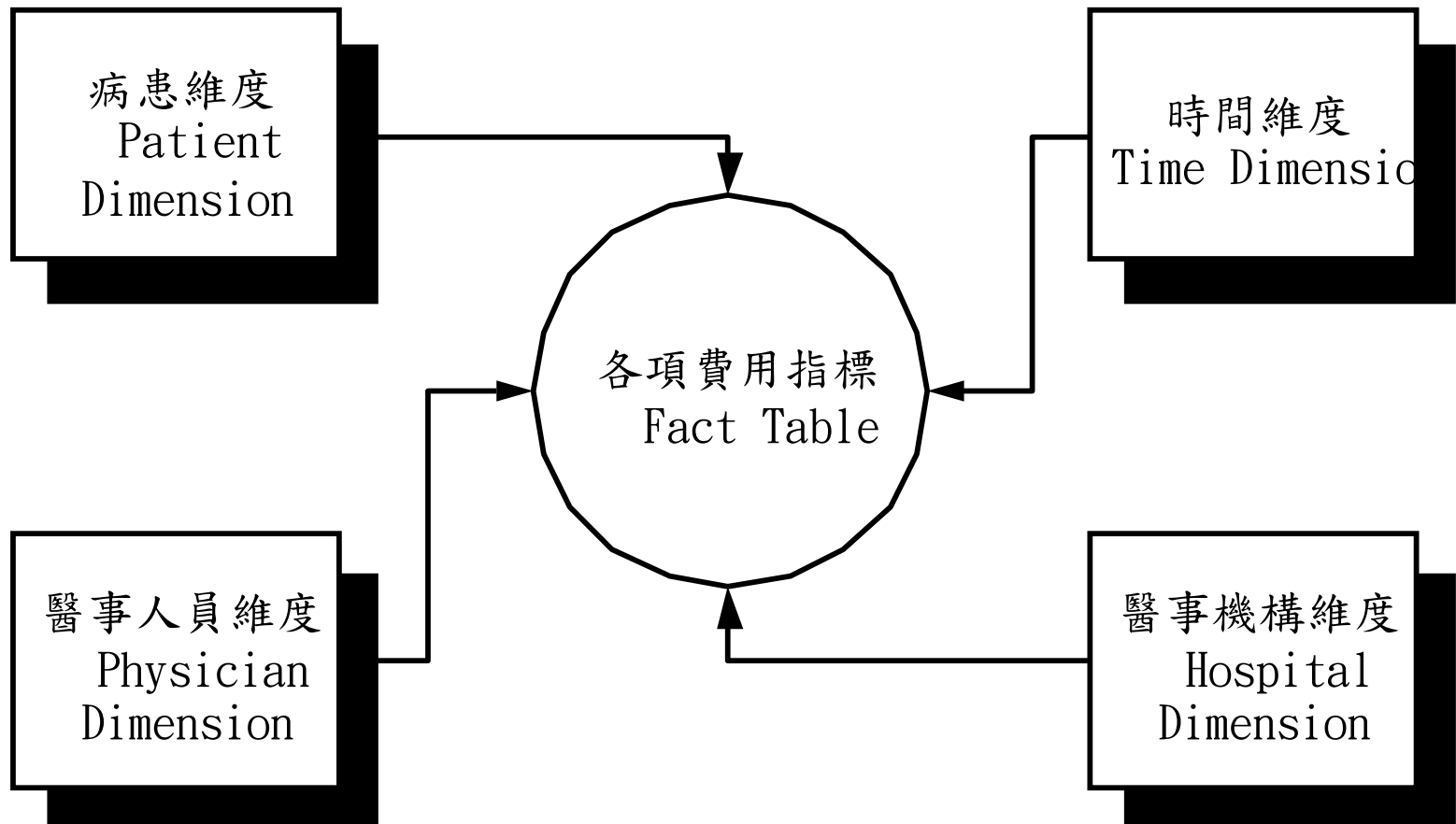
## 案例討論 – 使用之資料檔及欄位 (二)

DD1997	DRUG_AMT(藥費) DSVC_AMT(藥事服務費) NRTP_AMT(精神科治療費) INJT_AMT(注射技術費) BABY_AMT(嬰兒費) PART_AMT(部分負擔金額) APPL_AMT(申請費用金額)	
	IN_DATE(入院年月日)	時間維度
HOSB1997	HOSP_ID(醫事機構代號) HOSP_CONT_TYPE(特約類別) HOSP_TYPE_ID(型態別) HOSP_GRAD_ID(評鑑等級) AREA_NO_H(縣市區碼)	醫事機構維度
PER1997	PRSN_ID(醫事人員身分證號) BIRTHDAY(出生日期) PRSN_SEX(性別) PRSN_TYPE(醫事人員類別)	醫事人員維度



## 案例討論 – 維度設計

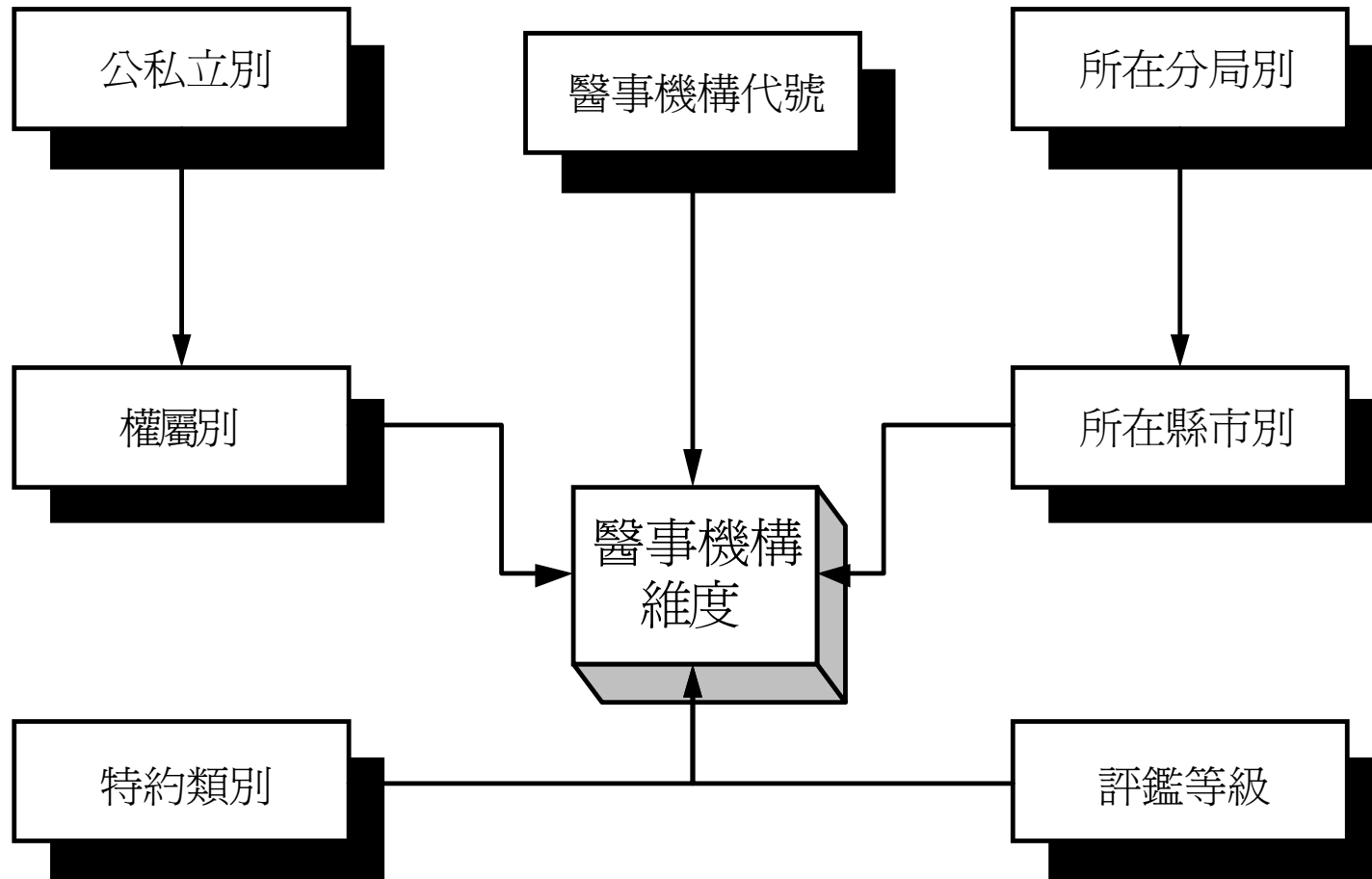
- 展開上表後，可以更清楚維度設計





# 案例討論 – 維度細分

## Information Object 分類成數個 Hierarchies





## 案例討論 — 資料庫實體設計

- 根據維度模式設計，採用星狀模式設計  
(Dimensional Modeling : Star Schema)

# 案例討論 — 資料結構

dbo.time\_dim

time\_key  
in\_date  
year  
season  
month

dbo.hosb\_dim

hosp\_key  
hosp\_id  
hosp\_PandP  
hosp\_belong  
hosp\_cont\_type  
hosp\_type\_id  
hosp\_grad\_id  
area\_branch  
area\_no\_h

dbo.inpatient\_fact

pati\_key  
time\_key  
hosp\_key  
prsn\_key  
em\_date  
lon\_date  
diag\_amt  
room\_amt  
meal\_amt  
amin\_amt  
rado\_amt  
thrp\_amt  
sgry\_amt  
phsc\_amt  
blod\_amt  
hd\_amt  
ane\_amt  
metr\_amt  
drug\_amt  
dsvc\_amt  
nrtp\_amt  
injt\_amt  
baby\_amt  
med\_amt  
part\_amt  
charg\_amt  
appl\_amt

dbo.pati\_dim

pati\_key  
pati\_id  
pati\_birthday  
pati\_age  
pati\_age\_range  
card\_seq\_no  
pati\_sex  
icd9cm\_code

dbo.prsn\_dim

prsn\_key  
prsn\_id  
prsn\_age  
prsn\_age\_range  
prsn\_sex  
prsn\_type

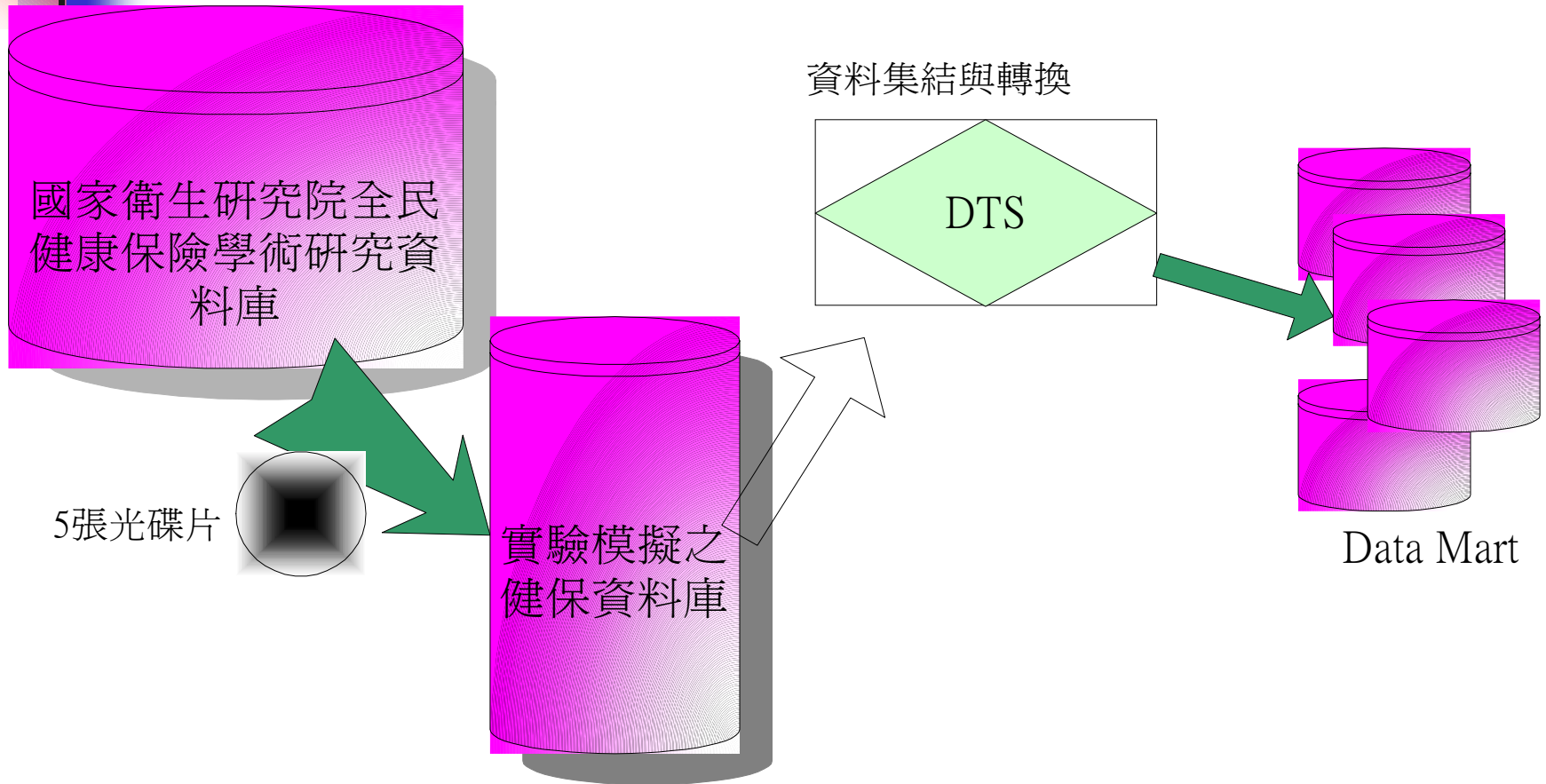




## 案例討論 – 建立資料表格

- ● Fact Table 特性
  - 1. 包含單一性的複合主鍵
  - 2. 包含數值欄位
  - 3. 資料列相當龐大
  
- ● Dimension Table 的特性
  - 1. 包含代理鍵的主鍵
  - 2. 跟事實表格有一對多的關係
  - 3. 至少包含一個決策因子
  - 4. 包含 Multi-Level 的維面階層欄位
  - 5. 包含隨時間變化的資料記錄欄

# 案例討論 – 資料集結設計及開發

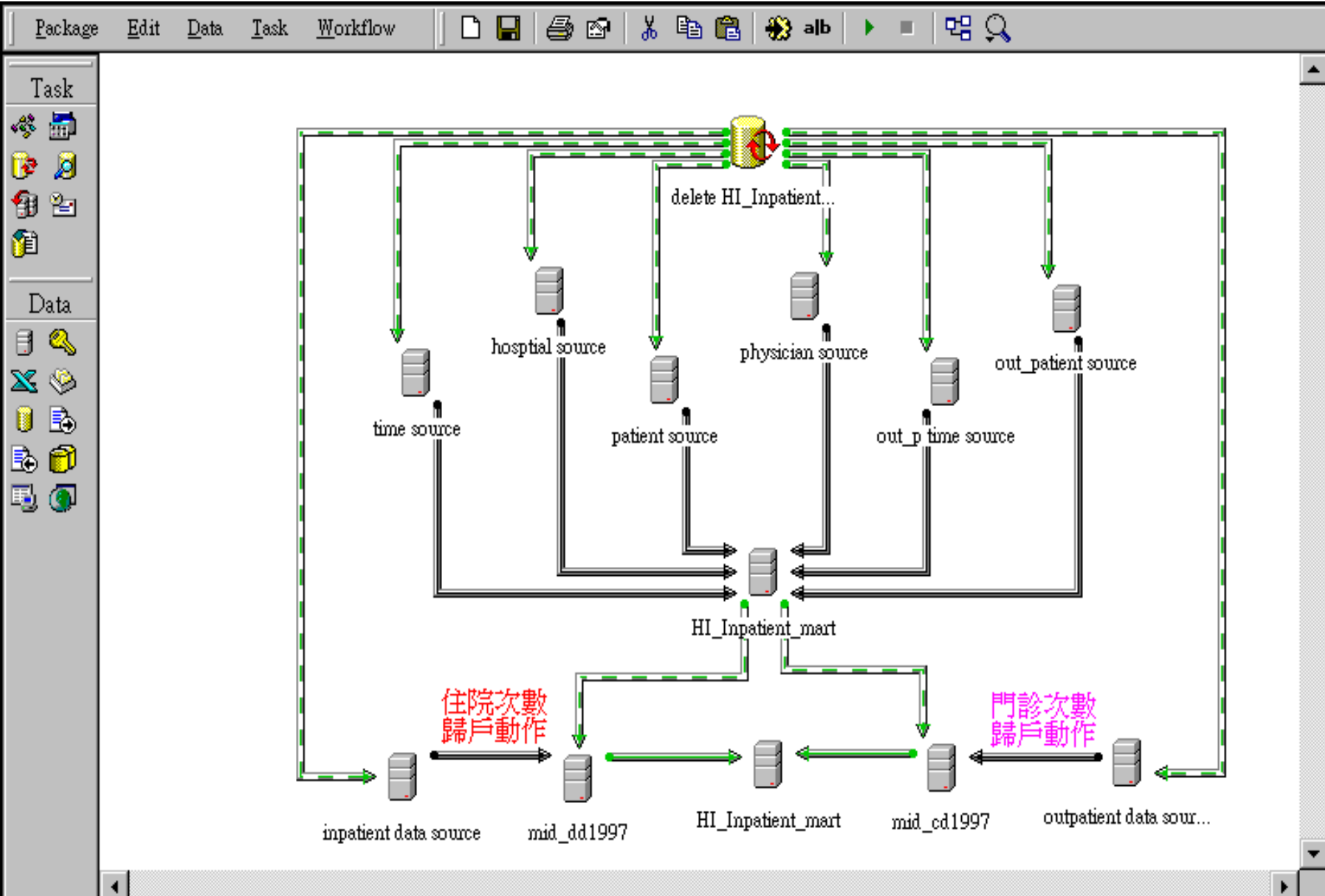


- 實體資料及交易資料檔總共有 4,372MB。

# 案例討論 – 清理過的資料

檔案類別	來源資料中文檔名	英文檔名	資料筆數	檔案大小
基本資料檔	醫事機構基本資料檔	HOSB1997	27,419	11
	醫事機構副檔資料檔	HOSX1997	25,039	6
	醫事機構診療科別明細檔	DETA1997	26,856	1
	醫事機構病床主檔	BED1997	210,555	14
	醫事人員基本資料檔	PER1997	148,771	14
	專科醫師證書主檔	DOC1997	59,224	5
	重大傷病證明明細檔	HV1997	627,858	164
	門診費用申請總表主檔	CT1997	197,430	64
	住院費用申請總表主檔	DT1997	12,758	3
系統抽樣檔	門診處方及治療明細檔	S_CD1997	585,625	117
	門診處方醫令明細檔	S_OO1997	2,063,502	197
	住院醫療費用清單明細檔	S_DD1997	117,647	42
	住院醫療費用醫令清單明細檔	S_DO1997	6,320,415	514
*特定主題分檔	特約藥局處方及調劑明細檔	GD1997	765,568	100
	特約藥局處方及調劑醫令檔	GO1997	3,233,513	232

# 案例討論 – 資料匯入資料庫



# 案例討論 – 資料匯入的時間

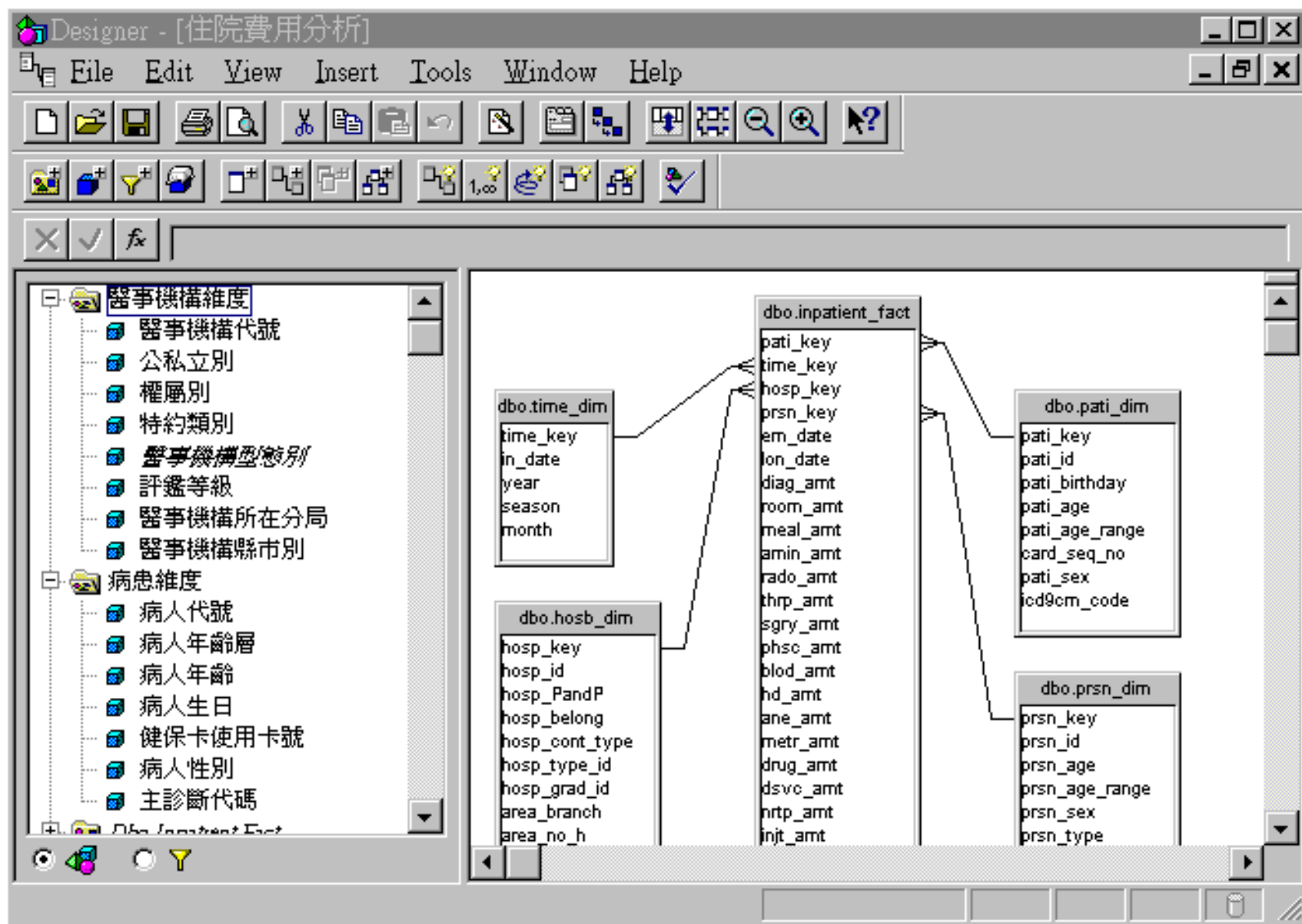
執行內容	處理資料筆數	所耗時間
刪除舊有的資料		8分50秒
轉換住院時間維度資料	1013	15秒
轉換醫事機構維度資料	27419	1分39秒
轉換住院病人維度資料	116980	17分29秒
轉換醫事人員維度資料	82444	18分23秒
住院資料歸戶	117176	3分38秒
轉換住院事實表格資料	116882	3分6秒
轉換門診時間維度資料	452	1分45秒
轉換門診病人維度資料	585522	11分42秒
門診資料歸戶	585597	13分41秒
轉換門診事實表格資料	580465	11分27秒
總和	2213950	29分41秒*



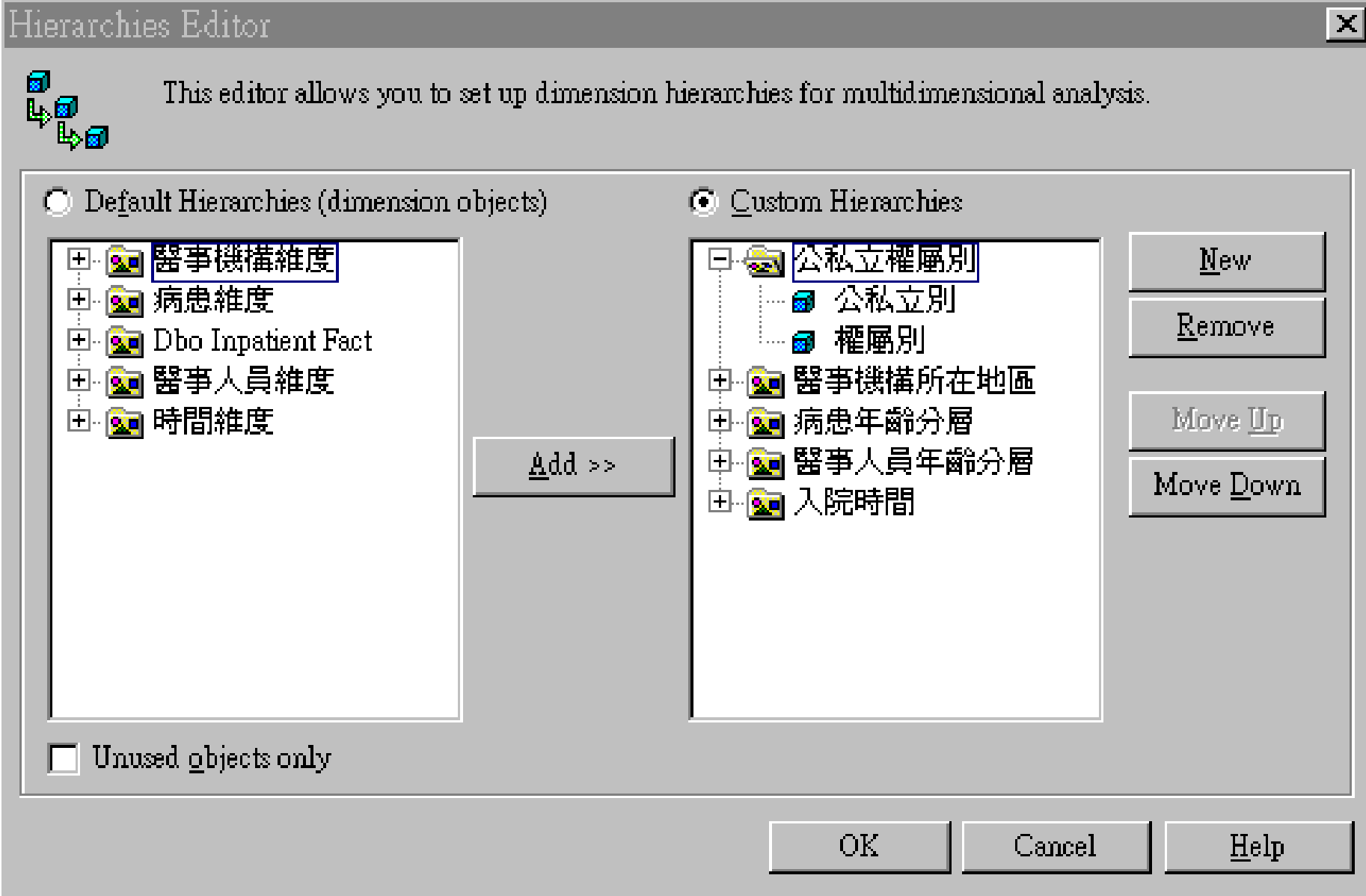
## 案例討論 – 使用者介面開發

- 定義中介資料 (matadata) 。
- 利用 WebIntelligence 2.5 為使用介面及報表產生器 。
- 線上分析工具的成功與否，最要緊的是前置作業中的維度分析及中介資料的定義 。

# 案例討論 – Metadata 設計



# 案例討論 – Dimension 設計






# 案例討論 - 使用者系統簽入

健保OLAP系統登入 - Microsoft Internet Explorer

檔案(E) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

網址(D) <http://ym79171/vi/> 移至

**全民健保 醫療費用 線上分析系統**




**Check Browser** 歡迎使用此系統，此系統需要使用Microsoft IE4.0或Netscape 4.x以上的瀏覽器，或其他支援JAVA或ActiveX的瀏覽器。若你不能確定使用的瀏覽器是否可以在本系統執行，請按下左方的"Check Browser"按鈕。

**User Help** 按下左方的"User Help"你可以取得Only help的資料，以利於了解如何使用本系統

**Log In**

系統製作者 [YAKIBUTA](#) 2000/07/08

建議使用IE4以上的瀏覽器以800\*600的方式取得最佳效果



近端 intranet

# 案例討論 – 分析維度選擇

住院費用分析

All Objects

- 醫事機構維度
- 病患維度
- 醫事人員維度
- 時間維度
- 各項統計值

Section

To create sections, drag objects from the list and drop them here (optional).

Details: Simple Table

To create a table, drag objects from the list and drop them here.

Scope of Analysis

☐ Drill mode

Results | Conditions | Settings

Apply Format Run Query

# 案例討論 – 報表格式設計

住院費用分析

All Objects

- 醫事機構維度
- 病患維度
- 醫事人員維度
- 時間維度
- 各項統計值

Section

To create sections, drag objects from the

Details: Simple Table

To create a table, drag objects from

Scope of Analysis

Tables

- Table
- Financial
- Form
- Crosstab

Bar Charts

Line Charts

Area Charts

Pie Charts

Radar Charts

☐ Drill mode

# 案例討論 – 維度定義

## 住院費用分析

All Objects

醫事機構維度

醫事機構代號

公私立別

權屬別

特約類別

評鑑等級

醫事機構所在分局

醫事機構縣市別

病患維度

醫事人員維度

時間維度

各項統計值

Section

To create sections, drag objects from the list and drop them here (optional).

Details: Crosstab Table

To create the horizontal crosstab header, drag objects from the list and drop them here.

醫事機構代號

To create the vertical crosstab header, drag objects from the list and drop them here.

To create a table, drag objects from the list and drop them here.

Scope of Analysis

Results

Conditions

Settings


Drill mode

Apply Format

Run Query

# 案例討論 - 報表製作

## 住院費用分析



### All Objects

- 血液透析費
- 血液血漿費
- 麻醉費
- 特殊材料費
- 藥費
- 藥事服務費
- 精神科治療費
- 注射技術費
- 嬰兒費
- 醫療費用
- 代辦費
- 部分負擔金額
- 申請費用金額
- 入院人次
- 醫事機構累計
- 醫事人員累計

### Section

To create sections, drag objects from the list and drop them here (optional).

### Details: Crosstab Table

☑ 公私立別

☑ 醫事機構縣市別

☑ 醫事機構代號

☑ 病人年齡層

● 申請費用金額

### Scope of Analysis

Results

Conditions

Settings

☐ Drill mode

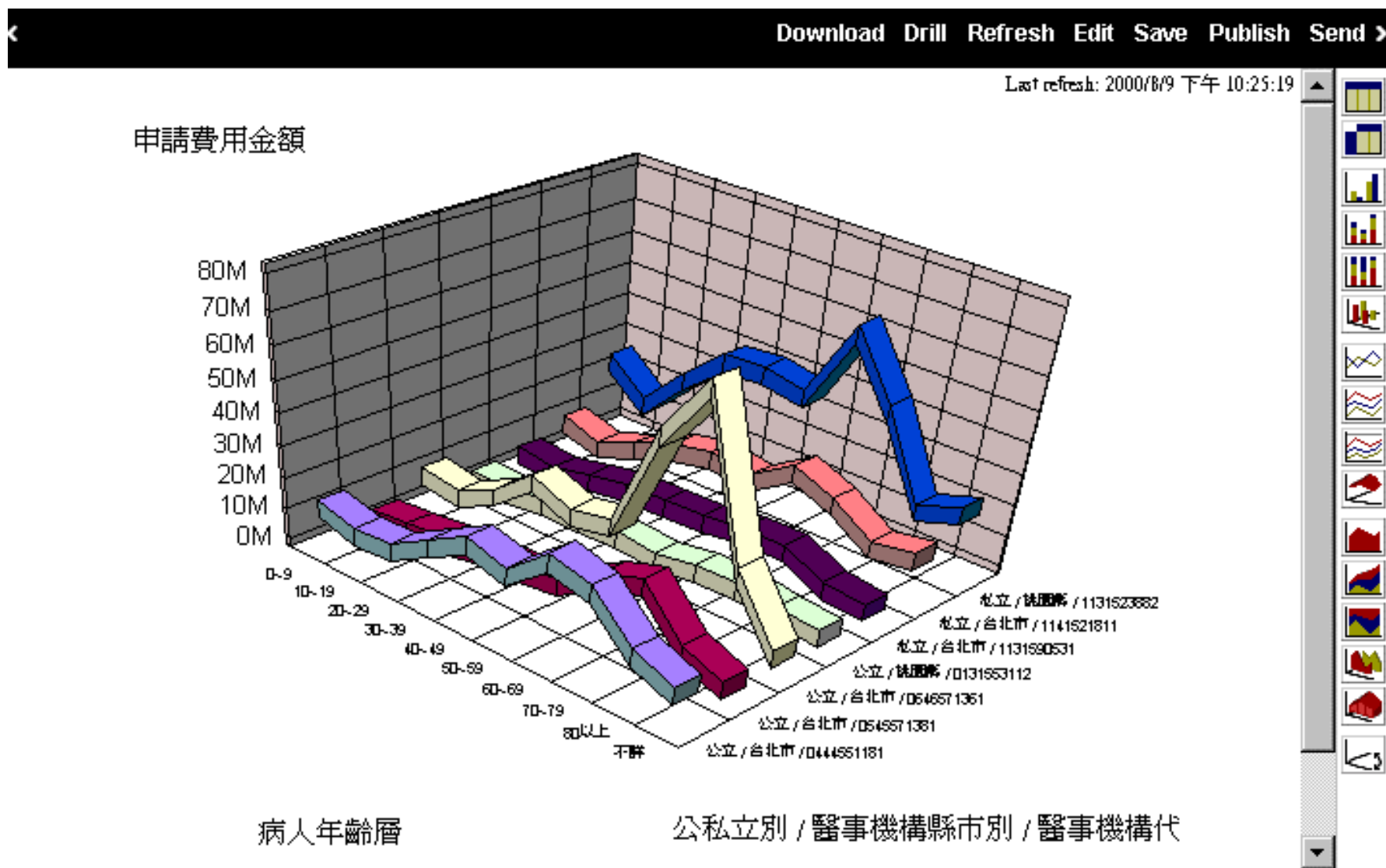
Apply Format

Run Query

# 案例討論 – 完成報表

Download Drill Refresh Edit Save Publish Send <							
Last refresh: 2000/8/9 下午 10:25:19 ▲							
	公立				私立		
	台北市			桃園縣	台北市		桃園縣
	0444551181	0545571381	0646571361	0131553112	1131590531	1141521811	1131523882
	申請費用金額	申請費用金額	申請費用金額	申請費用金額	申請費用金額	申請費用金額	申請費用金額
0~9	11225702	1155295	8517063	634197	2804974	7427134	22926678
10~19	7480751	2230357	6163318	630329	1239851	3393481	12725299
20~29	8513173	3781501	12373431	2610901	4044422	8477973	25854271
30~39	15657789	5759520	22671958	2300964	6265304	13539481	37001062
40~49	22749223	6071262	17468098	3200581	6623645	16396223	37987809
50~59	19106597	8361101	20246606	1751047	7060160	14093020	35439898
60~69	29828284	16441535	53161649	3772690	8939595	22565988	60298016
70~79	26585182	20740079	72013627	4322257	9121229	17541661	40878743
80以上	13363999	6028789	22680867	2449596	2858117	9181275	14238743
不詳	8055055	831227	2550828	1239446	1617027	11149137	18860242

# 案例討論 – 圖形模擬



# 案例討論 – 資料匯出

Microsoft Excel - Document5.csv

檔案(E) 編輯(E) 檢視(V) 插入(I) 格式(O) 工具(T) 資料(D) 視窗(W) 說明(H)

新細明體 12 B I U \$ % , % .00 % 100%

A1 = 醫事機構代號

	A	B	C	D	E	F	G	H
1	醫事機構代號	公私立別	醫事機構縣市別	病人年齡層	申請費用金額			
2	444551181	公立	台北市	10~19	7480751			
3	131553112	公立	桃園縣	20~29	2610901			
4	1131523882	私立	桃園縣	60~69	60298016			
5	444551181	公立	台北市	30~39	15657789			
6	1131523882	私立	桃園縣	50~59	35439898			
7	131553112	公立	桃園縣	不詳	1239446			
8	1141521811	私立	台北市	10~19	3393481			
9	444551181	公立	台北市	80以上	13363999			
10	131553112	公立	桃園縣	50~59	1751047			
11	545571381	公立	台北市	10~19	2230357			
12	545571381	公立	台北市	60~69	16441535			
13	545571381	公立	台北市	0~9	1155295			
14	1141521811	私立	台北市	40~49	16396223			
15	646571361	公立	台北市	30~39	22671958			
16	1131523882	私立	桃園縣	80以上	14238743			
17	444551181	公立	台北市	0~9	11225702			
18	1131590531	私立	台北市	10~19	1239851			

Document5

就緒

開始 | 檔案 | 第... | W... | M... | CH PM 10:53





# 典型的 OLAP 動作

- Roll up (drill-up): summarize data
  - *by climbing up hierarchy or by dimension reduction*
- Drill down (roll down): reverse of roll-up
  - *from higher level summary to lower level summary or detailed data, or introducing new dimensions*
- Slice and dice:
  - *project and select*
- Pivot (rotate):
  - *reorient the cube, visualization, 3D to series of 2D planes.*
- Other operations
  - *drill across: involving (across) more than one fact table*
  - *drill through: through the bottom level of the cube to its back-end relational tables (using SQL)*

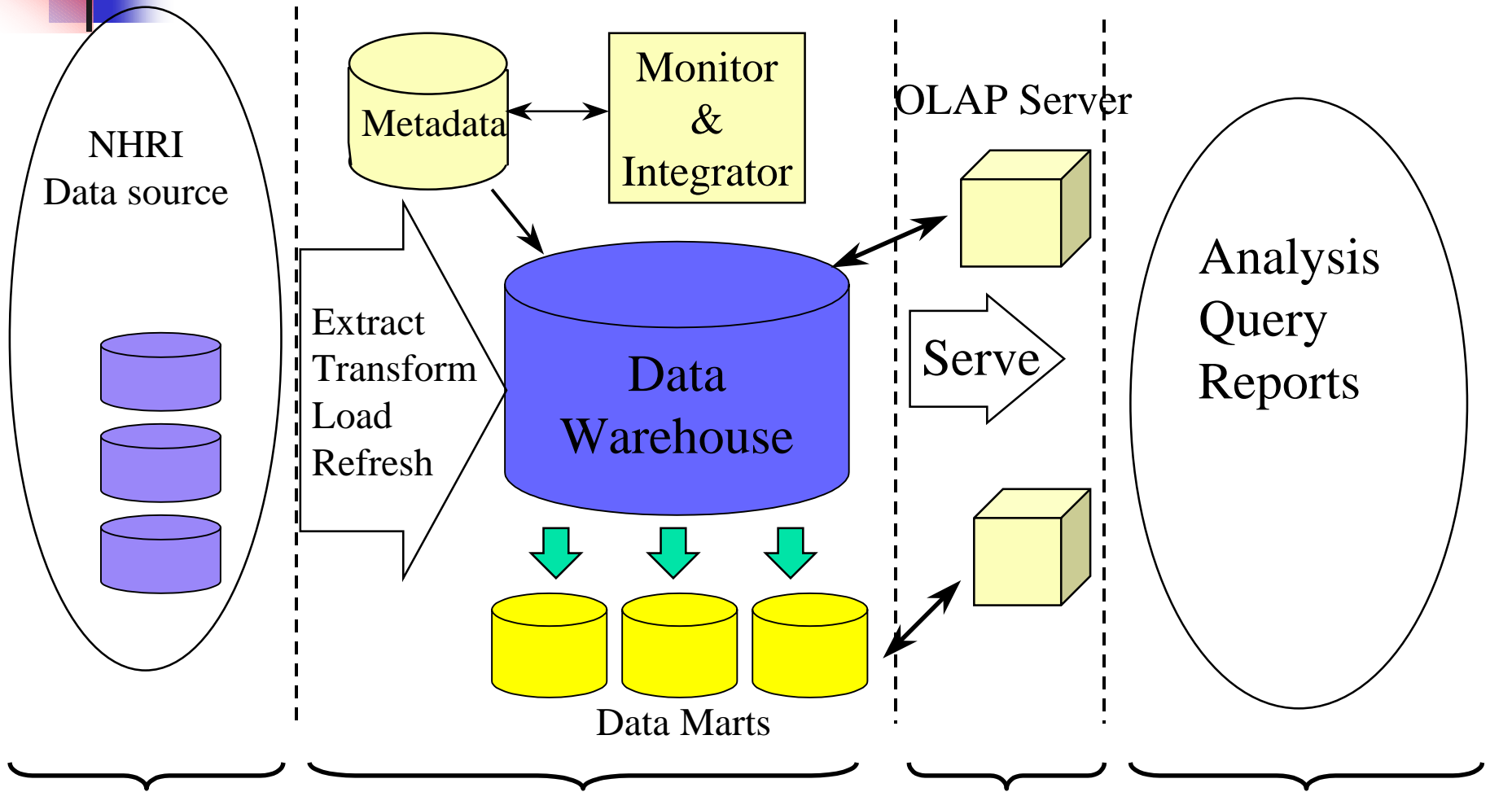


# 資料倉儲的設計觀點

- Four views regarding the design of a data warehouse
  - Top-down view
    - allows selection of the relevant information necessary for the data warehouse
  - Data source view
    - exposes the information being captured, stored, and managed by operational systems
  - Data warehouse view
    - consists of fact tables and dimension tables
  - Business query view
    - sees the perspectives of data in the warehouse from the view of end-user



# 多層次的系統架構



Data Sources

Data Storage

OLAP Engine

Front-End Tools



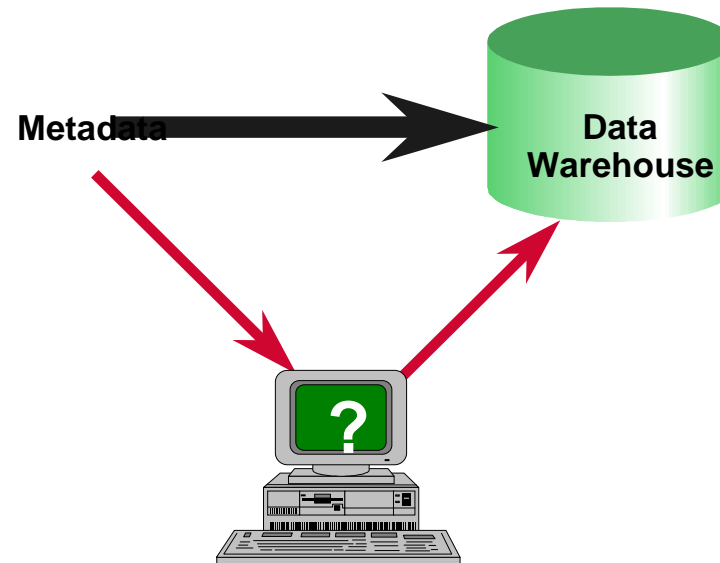
# OLAP 伺服器的架構

- Relational OLAP (ROLAP)
  - Use relational or extended-relational DBMS to store and manage warehouse data and OLAP middle ware to support missing pieces
  - Include optimization of DBMS backend, implementation of aggregation navigation logic, and additional tools and services
  - greater scalability
- Multidimensional OLAP (MOLAP)
  - Array-based multidimensional storage engine (sparse matrix techniques)
  - fast indexing to pre-computed summarized data
- Hybrid OLAP (HOLAP)
  - User flexibility, e.g., low level: relational, high-level: array



## 如何定義 Metadata?

**Metadata provides information about what data are available, where it came from, how it got there, and what happened to it along the way**



*Data? about data*



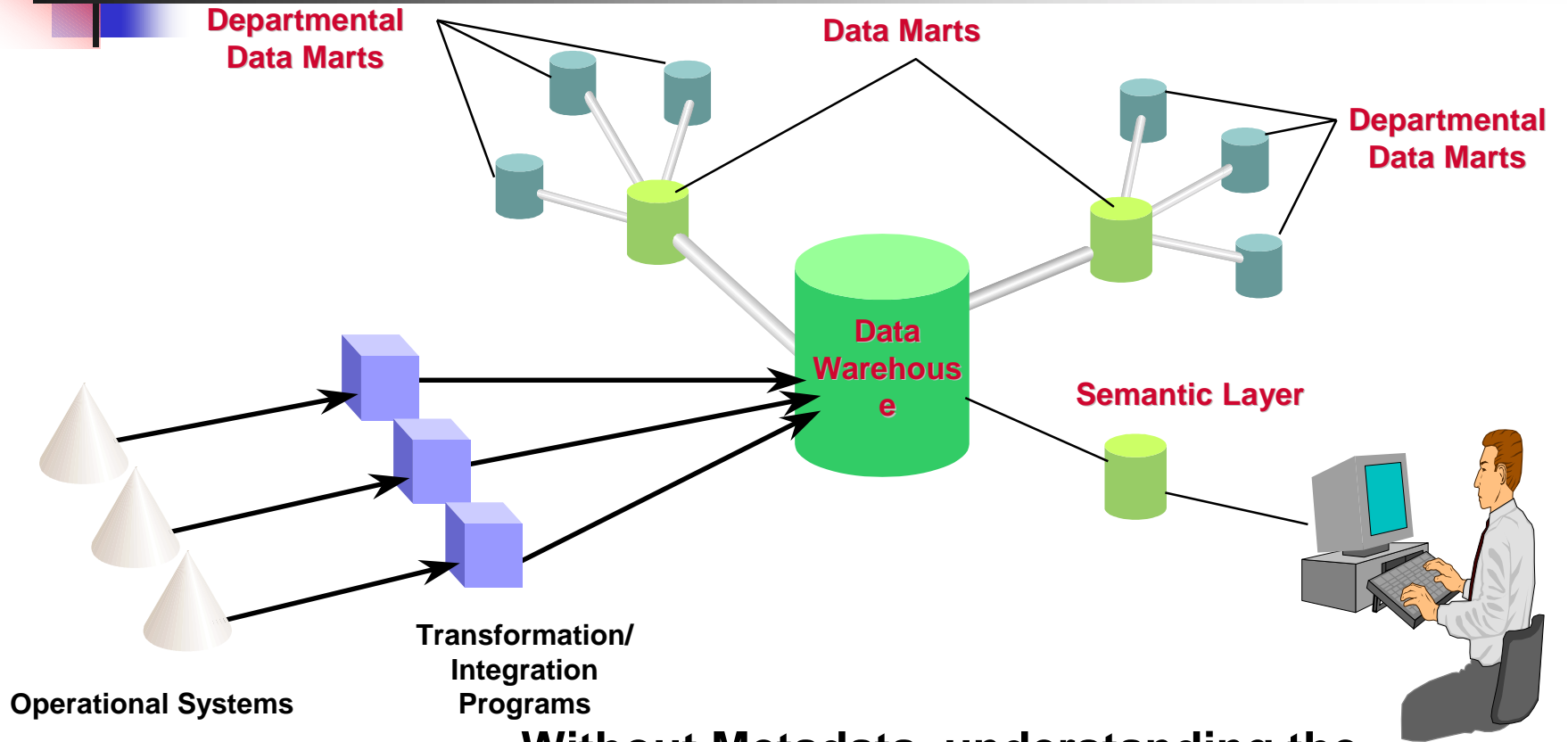
# 為何要有 Metadata?

**To answer questions such as**

- **What is in the data warehouse?**
- **Where did it come from?**
- **How was it transformed?**
- **How has it changed over time?**
- **When was it last refreshed?**
- **How is data logically related?**
- **What are the business terms that are applicable?**
- **What business assumptions have been made?**
- **What is the purge criteria of data?**
- **What data has been summarized?**
- **How was this summarization created?**



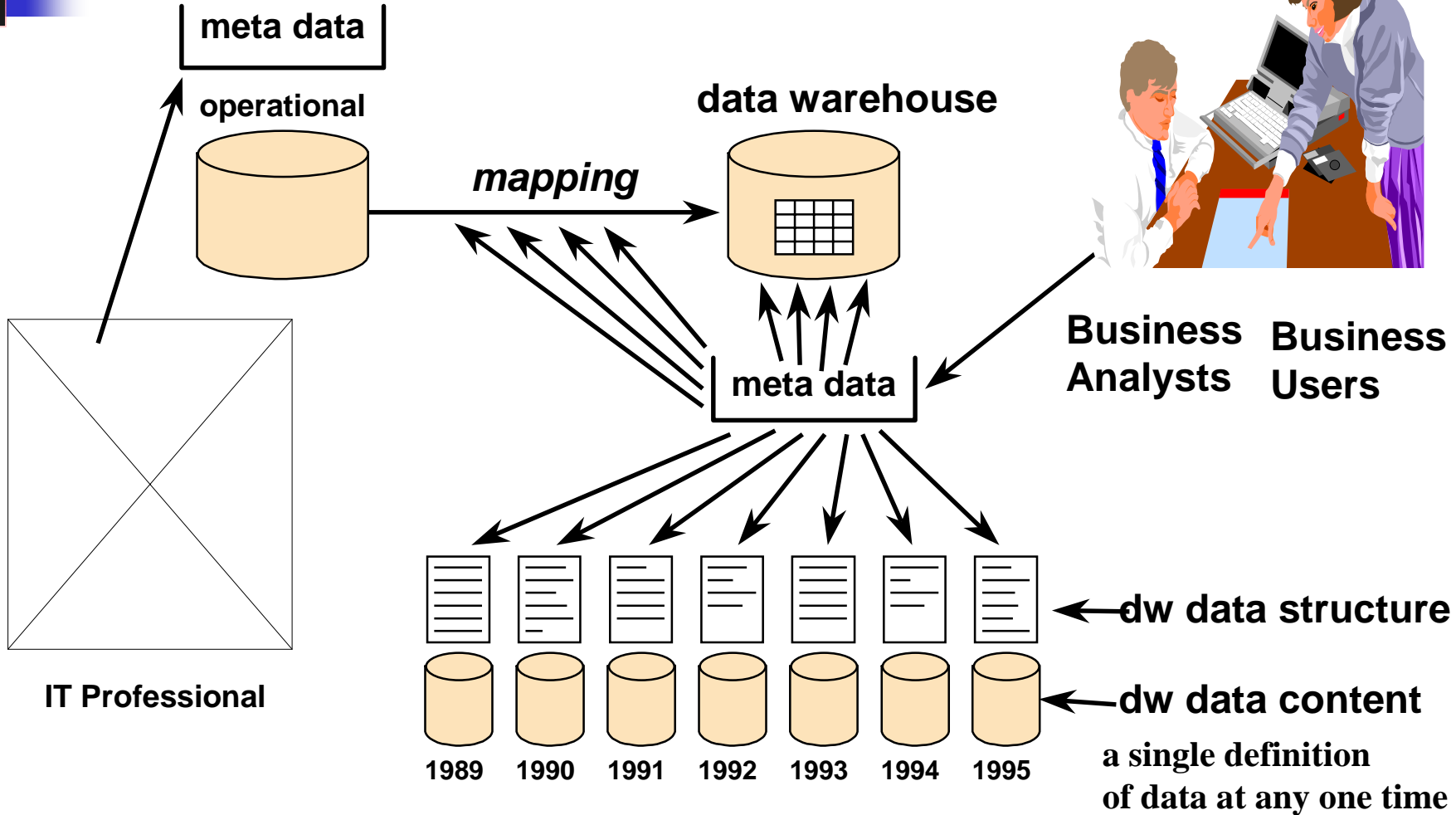
# 不使用 Metadata



**Without Metadata, understanding the structure, content and meaning of the data at each level becomes increasingly difficult.**



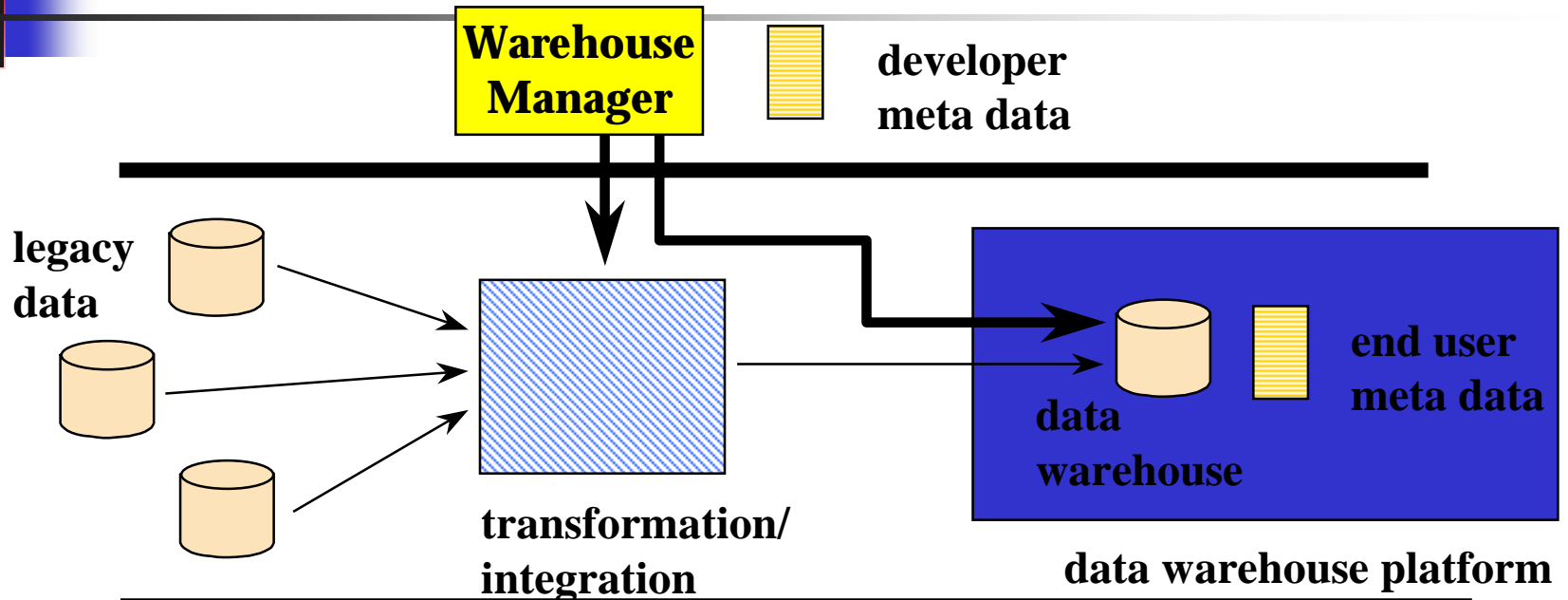
# 運用 Metadata







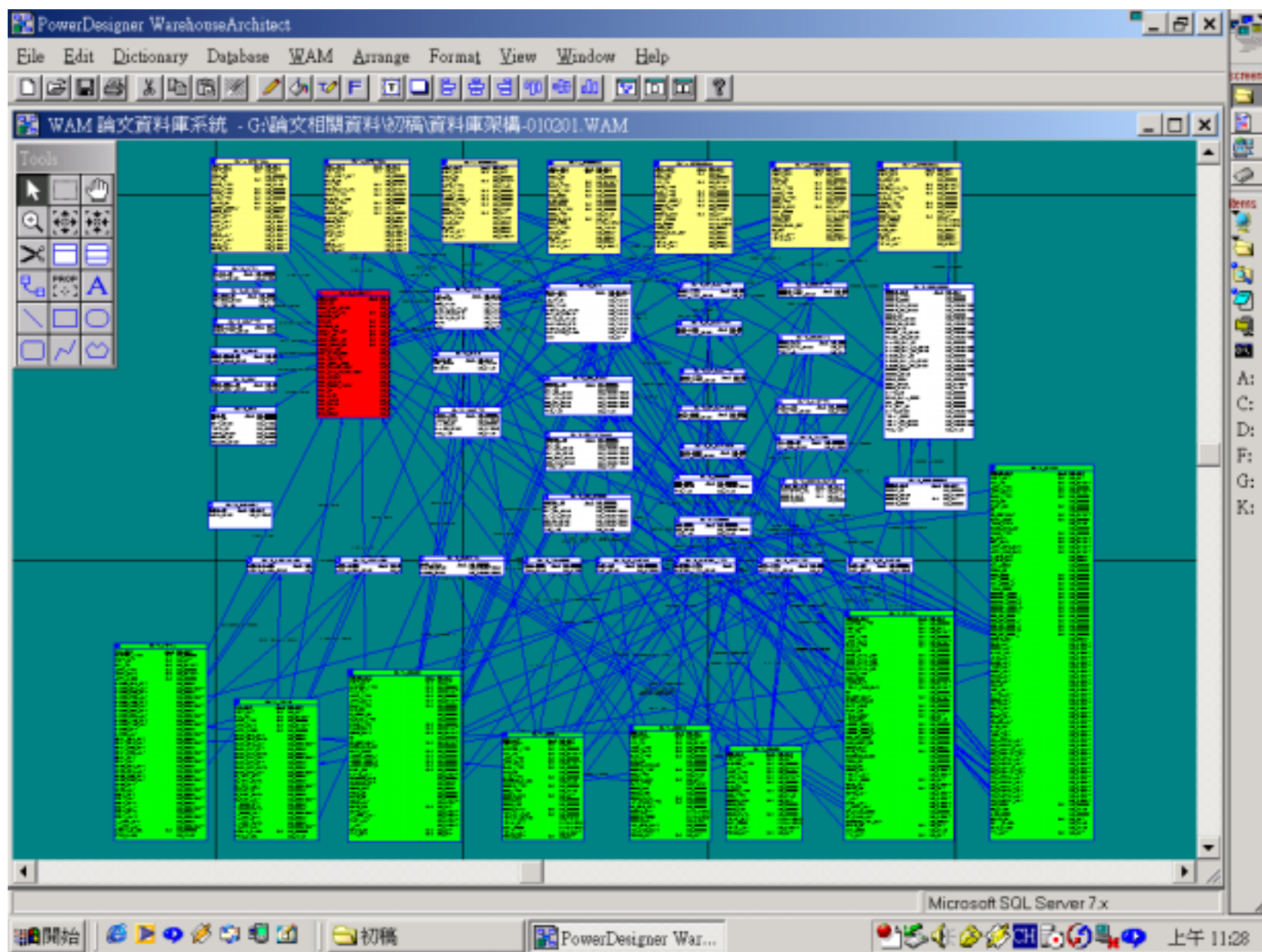
# 利用自動的工具管理 Metadata



**Metadata is developed as an interactive product of designing and developing the data warehouse environment:**

- ♦ it is complete
- ♦ it is done every time
- ♦ it is done automatically
- ♦ it is not an afterthought
- ♦ it is not optional
- ♦ it requires no extra effort ?manually or otherwise

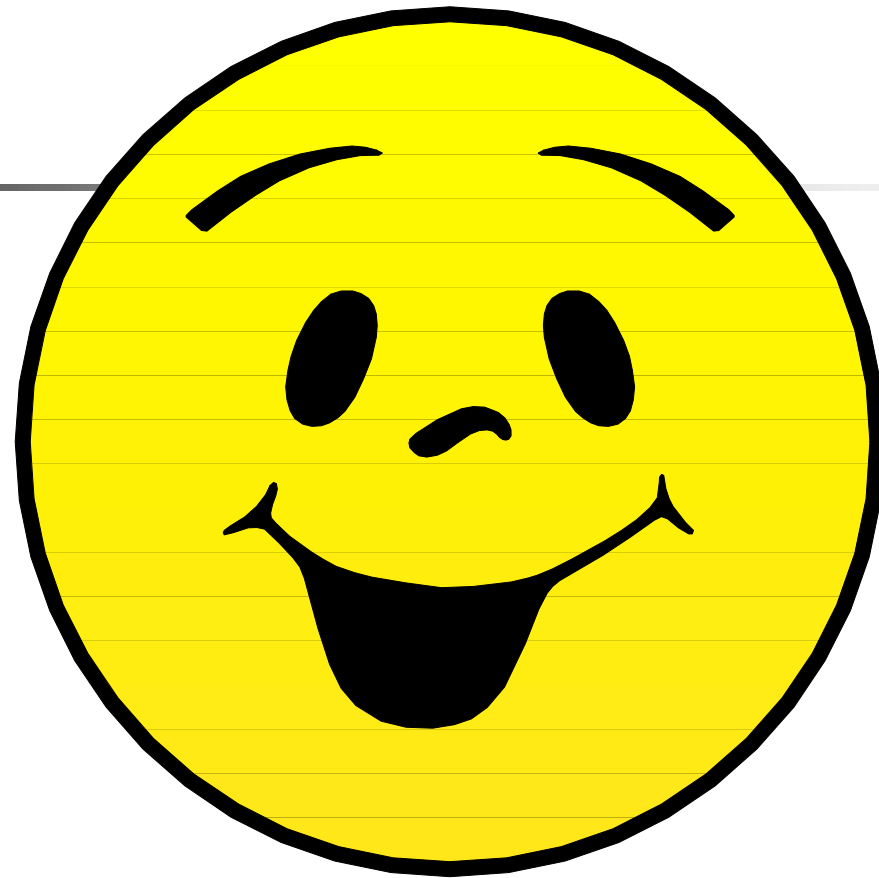
# 健保資料倉儲的維度結構





# 結論

- Data warehouse
  - A subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process
- A multi-dimensional model of a data warehouse
  - Star schema, snowflake schema, fact constellations
  - A data cube consists of dimensions & measures
- OLAP operations: drilling, rolling, slicing, dicing and pivoting
- OLAP servers: ROLAP, MOLAP, HOLAP



Thank you !!!