Data Warehouse

Asst.Prof.Dr. Pattarachai Lalitrojwong

Faculty of Information Technology
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520

pattarachai@it.kmitl.ac.th

The Evolution of Data Warehousing

- Since 1970s, organizations gained competitive advantage through systems that automate business processes to offer more efficient and cost-effective services to the customer.
- This resulted in accumulation of growing amounts of data in operational databases.

The Evolution of Data Warehousing

- Organizations now focus on ways to use operational data to support decision-making, as a means of gaining competitive advantage.
- However, operational systems were never designed to support such business activities.
- Businesses typically have numerous operational systems with overlapping and sometimes contradictory definitions.

The Evolution of Data Warehousing

- Organizations need to turn their archives of data into a source of knowledge, so that a single integrated / consolidated view of the organization’s data is presented to the user.
- A data warehouse was deemed the solution to meet the requirements of a system capable of supporting decision-making, receiving data from multiple operational data sources.

The Need for Data Analysis

- Managers must be able to track daily transactions to evaluate how the business is performing
- By tapping into the operational database, management can develop strategies to meet organizational goals
- Data analysis can provide information about short-term tactical evaluations and strategies

Solving Business Problems and Adding Value with Data Warehouse-Based Solutions

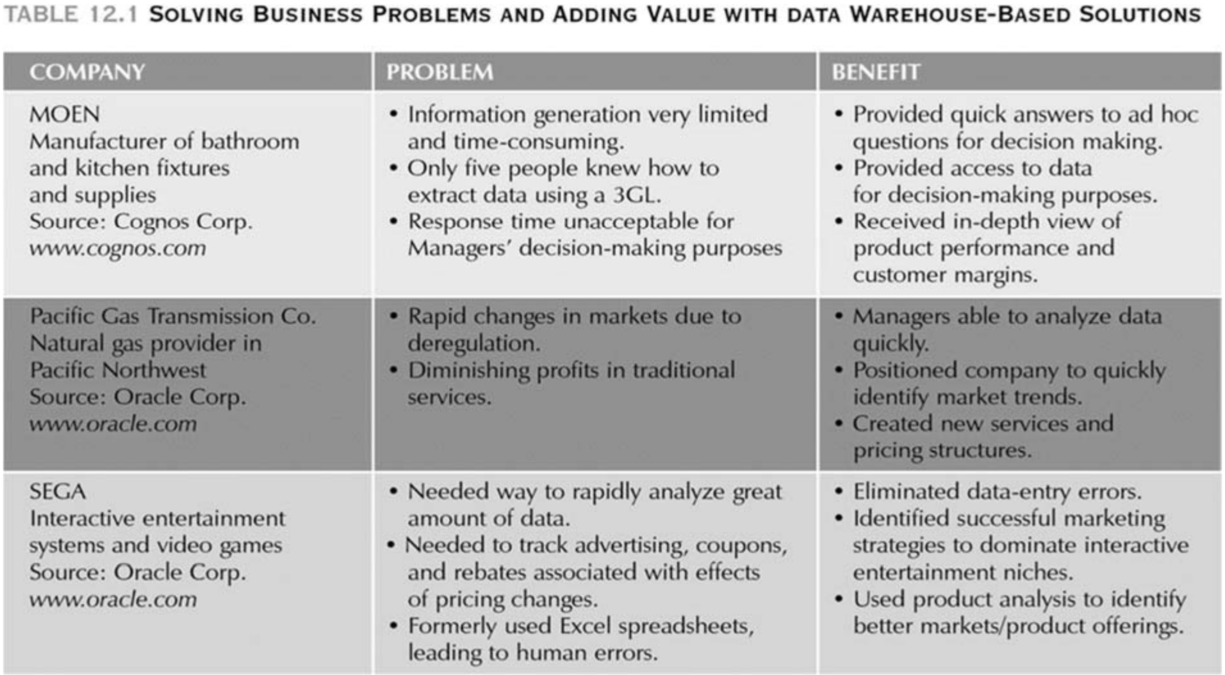


TABLE 12.1 **SOLVING BUSINESS PROBLEMS AND ADDING VALUE WITH DATA WAREHOUSE-BASED SOLUTIONS**

COMPANY	PROBLEM	BENEFIT
MOEN Manufacturer of bathroom and kitchen fixtures and supplies Source: Cognos Corp. *www.cognos.com*	• Information generation very limited and time-consuming. • Only five people knew how to extract data using a 3GL. • Response time unacceptable for Managers' decision-making purposes	• Provided quick answers to ad hoc questions for decision making. • Provided access to data for decision-making purposes. • Received in-depth view of product performance and customer margins.
Pacific Gas Transmission Co. Natural gas provider in Pacific Northwest Source: Oracle Corp. *www.oracle.com*	• Rapid changes in markets due to deregulation. • Diminishing profits in traditional services.	• Managers able to analyze data quickly. • Positioned company to quickly identify market trends. • Created new services and pricing structures.
SEGA Interactive entertainment systems and video games Source: Oracle Corp. *www.oracle.com*	• Needed way to rapidly analyze great amount of data. • Needed to track advertising, coupons, and rebates associated with effects of pricing changes. • Formerly used Excel spreadsheets, leading to human errors.	• Eliminated data-entry errors. • Identified successful marketing strategies to dominate interactive entertainment niches. • Used product analysis to identify better markets/product offerings.

Solving Business Problems and Adding Value with Data Warehouse-Based Solutions (continued)

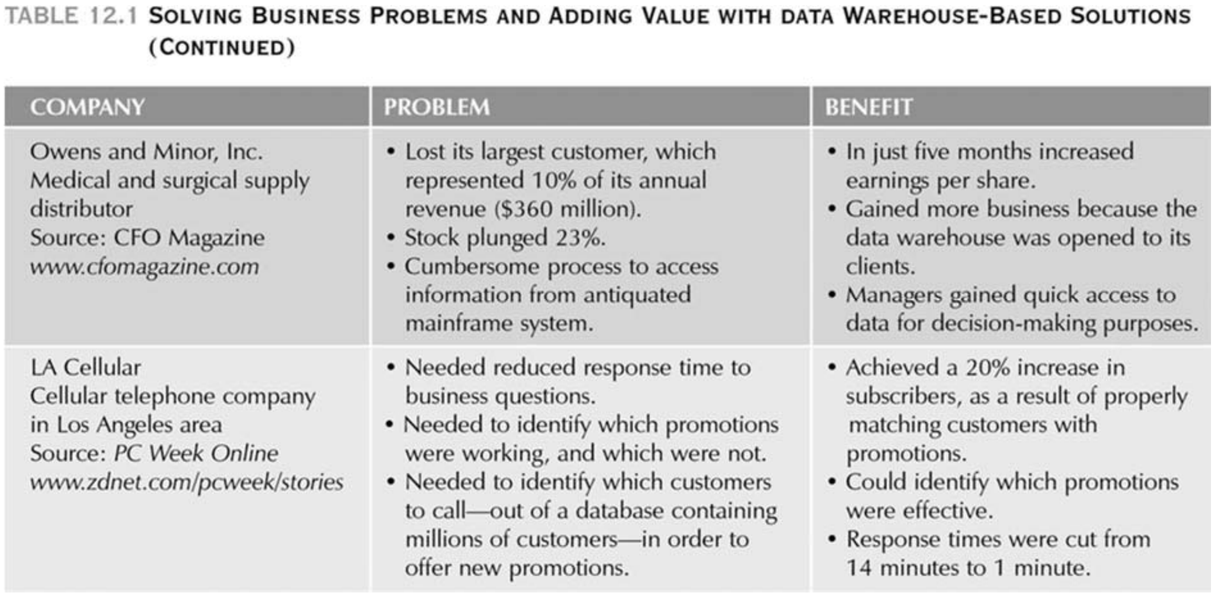


TABLE 12.1 **SOLVING BUSINESS PROBLEMS AND ADDING VALUE WITH DATA WAREHOUSE-BASED SOLUTIONS (CONTINUED)**

COMPANY	PROBLEM	BENEFIT
Owens and Minor, Inc. Medical and surgical supply distributor Source: CFO Magazine *www.cfomagazine.com*	• Lost its largest customer, which represented 10% of its annual revenue ($360 million). • Stock plunged 23%. • Cumbersome process to access information from antiquated mainframe system.	• In just five months increased earnings per share. • Gained more business because the data warehouse was opened to its clients. • Managers gained quick access to data for decision-making purposes.
LA Cellular Cellular telephone company in Los Angeles area Source: *PC Week Online* *www.zdnet.com/pcweek/stories*	• Needed reduced response time to business questions. • Needed to identify which promotions were working, and which were not. • Needed to identify which customers to call—out of a database containing millions of customers—in order to offer new promotions.	• Achieved a 20% increase in subscribers, as a result of properly matching customers with promotions. • Could identify which promotions were effective. • Response times were cut from 14 minutes to 1 minute.

Decision Support Systems

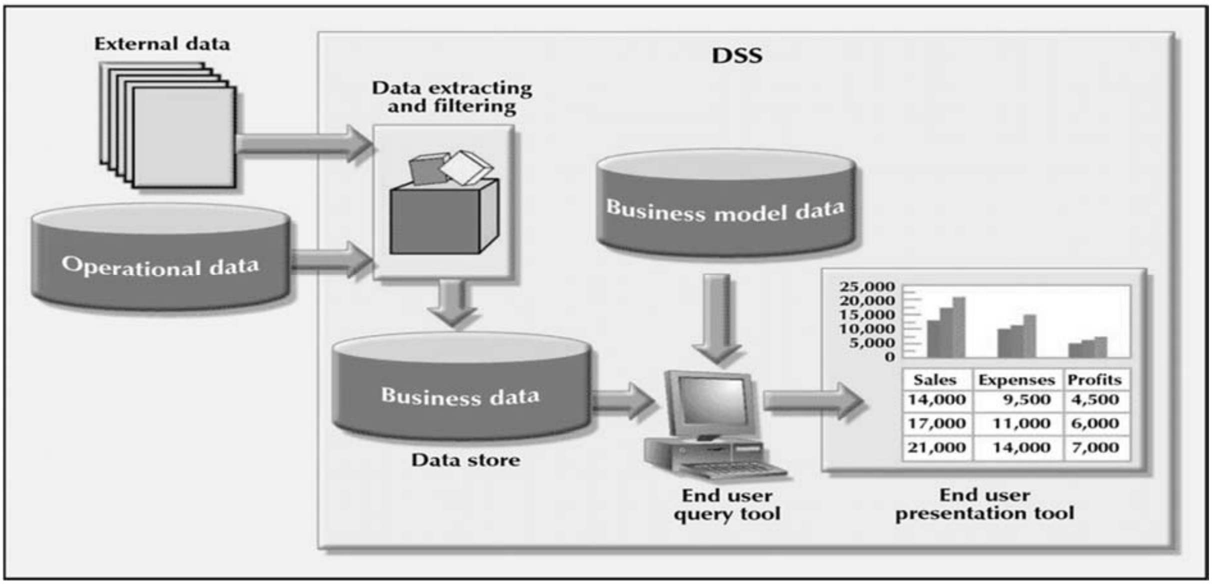
- Methodology (or series of methodologies) designed to extract information from data and to use such information as a basis for decision making
- Decision support system (DSS):
  - Arrangement of computerized tools used to assist managerial decision making within a business
  - Usually requires extensive data “massaging” to produce information
  - Used at all levels within an organization
  - Often tailored to focus on specific business areas
  - Provides ad hoc query tools to retrieve data and to display data in different formats

Decision Support Systems (continued)

- Composed of four main components:
  - Data store component
    - Basically a DSS database
  - Data extraction and filtering component
    - Used to extract and validate data taken from operational database and external data sources
  - End-user query tool
    - Used to create queries that access database
  - End-user presentation tool
    - Used to organize and present data

Main Components of a Decision Support System (DSS)

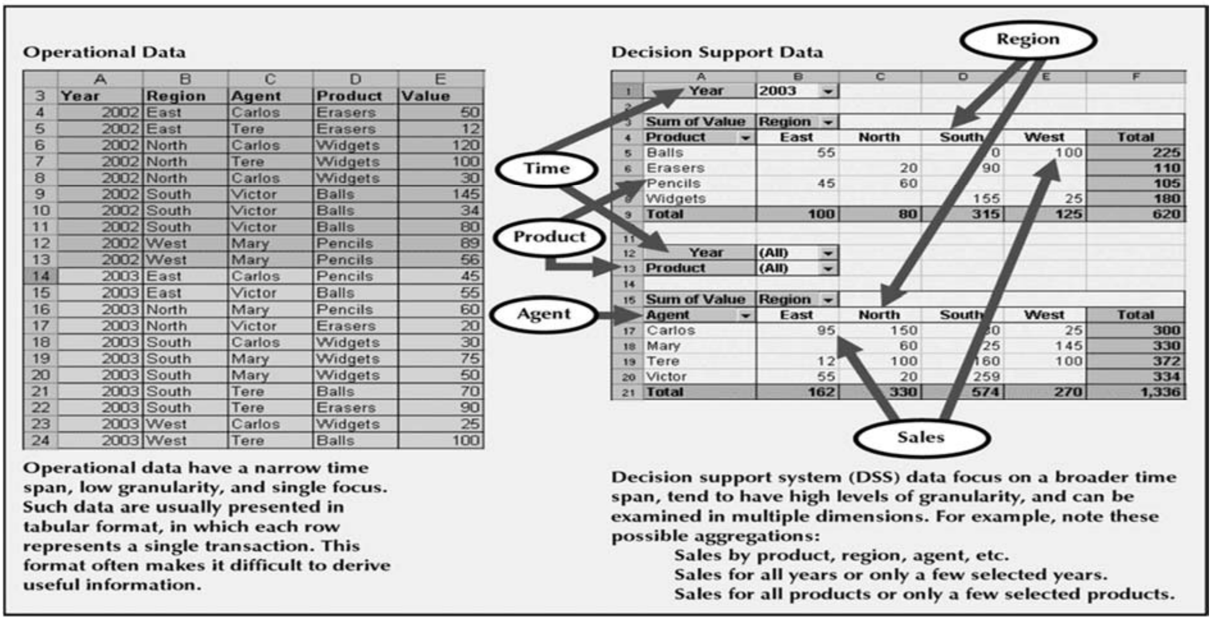
FIGURE 12.1 MAIN COMPONENTS OF A DECISION SUPPORT SYSTEM (DSS)



Sales	Expenses	Profits
14,000	9,500	4,500
17,000	11,000	6,000
21,000	14,000	7,000

Transforming Operational Data Into Decision Support Data

FIGURE 12.2 TRANSFORMING OPERATIONAL DATA INTO DECISION SUPPORT DATA



Operational Data

	A	B	C	D	E
3	Year	Region	Agent	Product	Value
4	2002	East	Carlos	Erasers	50
5	2002	East	Tere	Erasers	12
6	2002	North	Carlos	Widgets	120
7	2002	North	Tere	Widgets	100
8	2002	North	Carlos	Widgets	30
9	2002	South	Victor	Balls	145
10	2002	South	Victor	Balls	34
11	2002	South	Victor	Balls	80
12	2002	West	Mary	Pencils	89
13	2002	West	Mary	Pencils	56
14	2003	East	Carlos	Pencils	45
15	2003	East	Victor	Balls	55
16	2003	North	Mary	Pencils	60
17	2003	North	Victor	Erasers	20
18	2003	South	Carlos	Widgets	30
19	2003	South	Mary	Widgets	75
20	2003	South	Mary	Widgets	50
21	2003	South	Tere	Balls	70
22	2003	South	Tere	Erasers	90
23	2003	West	Carlos	Widgets	25
24	2003	West	Tere	Balls	100

Operational data have a narrow time span, low granularity, and single focus. Such data are usually presented in tabular format, in which each row represents a single transaction. This format often makes it difficult to derive useful information.

Decision Support Data

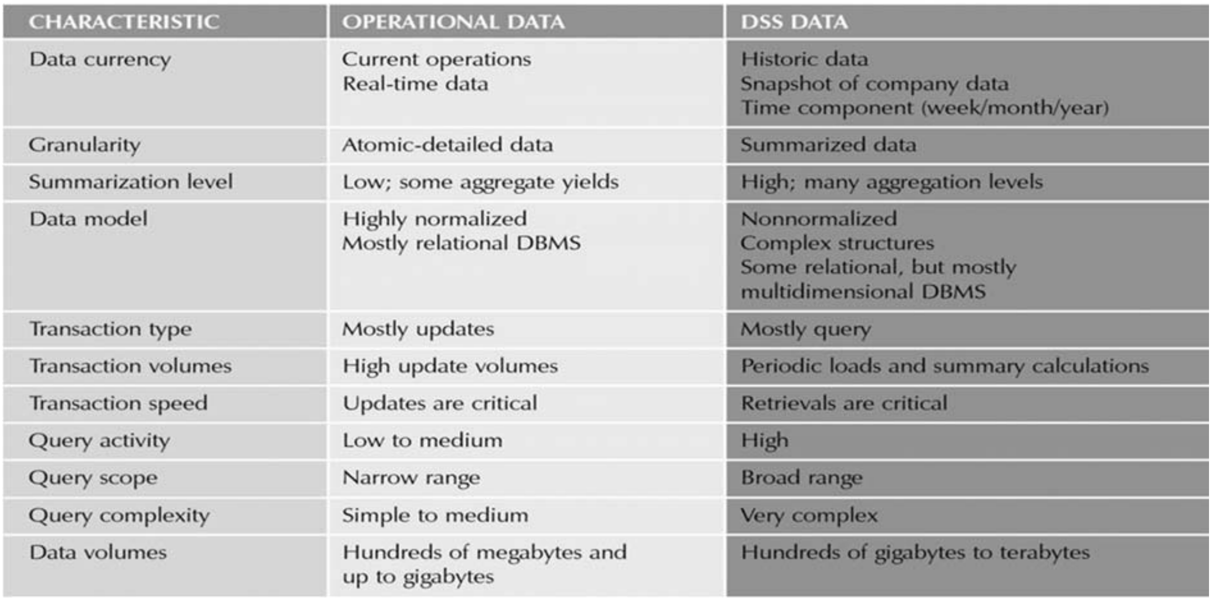
	A	B	C	D	E	F
1	Year	2003				
2						
3	Sum of Value	Region				
4	Product	East	North	South	West	Total
5	Balls	55		0	100	225
6	Erasers		20	90		110
7	Pencils	45	60			105
8	Widgets			155	25	180
9	Total	100	80	315	125	620
10						
11						
12	Year	(All)				
13	Product	(All)				
14						
15	Sum of Value	Region				
16	Agent	East	North	South	West	Total
17	Carlos	95	150	30	25	300
18	Mary		60	25	145	330
19	Tere	12	100	160	100	372
20	Victor	55	20	259		334
21	Total	162	330	574	270	1,336

Decision support system (DSS) data focus on a broader time span, tend to have high levels of granularity, and can be examined in multiple dimensions. For example, note these possible aggregations:

Sales by product, region, agent, etc.
Sales for all years or only a few selected years.
Sales for all products or only a few selected products.

Contrasting Operational and DSS Data Characteristics

TABLE 12.2 CONTRASTING OPERATIONAL AND DSS DATA CHARACTERISTICS



CHARACTERISTIC	OPERATIONAL DATA	DSS DATA
Data currency	Current operations Real-time data	Historic data Snapshot of company data Time component (week/month/year)
Granularity	Atomic-detailed data	Summarized data
Summarization level	Low; some aggregate yields	High; many aggregation levels
Data model	Highly normalized Mostly relational DBMS	Nonnormalized Complex structures Some relational, but mostly multidimensional DBMS
Transaction type	Mostly updates	Mostly query
Transaction volumes	High update volumes	Periodic loads and summary calculations
Transaction speed	Updates are critical	Retrievals are critical
Query activity	Low to medium	High
Query scope	Narrow range	Broad range
Query complexity	Simple to medium	Very complex
Data volumes	Hundreds of megabytes and up to gigabytes	Hundreds of gigabytes to terabytes

Ten-Year Sales History for a Single Department, in Millions of Dollars

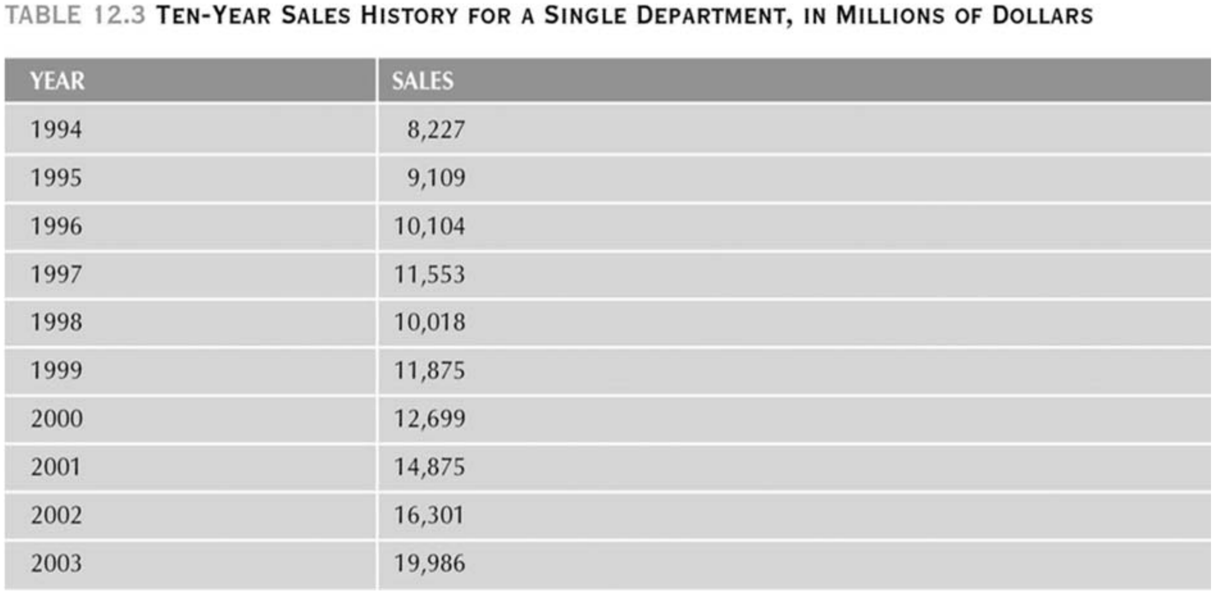


TABLE 12.3 **TEN-YEAR SALES HISTORY FOR A SINGLE DEPARTMENT, IN MILLIONS OF DOLLARS**

YEAR	SALES
1994	8,227
1995	9,109
1996	10,104
1997	11,553
1998	10,018
1999	11,875
2000	12,699
2001	14,875
2002	16,301
2003	19,986

Yearly Sales Summaries, Two Stores and Two Departments per Store, in Millions of Dollars

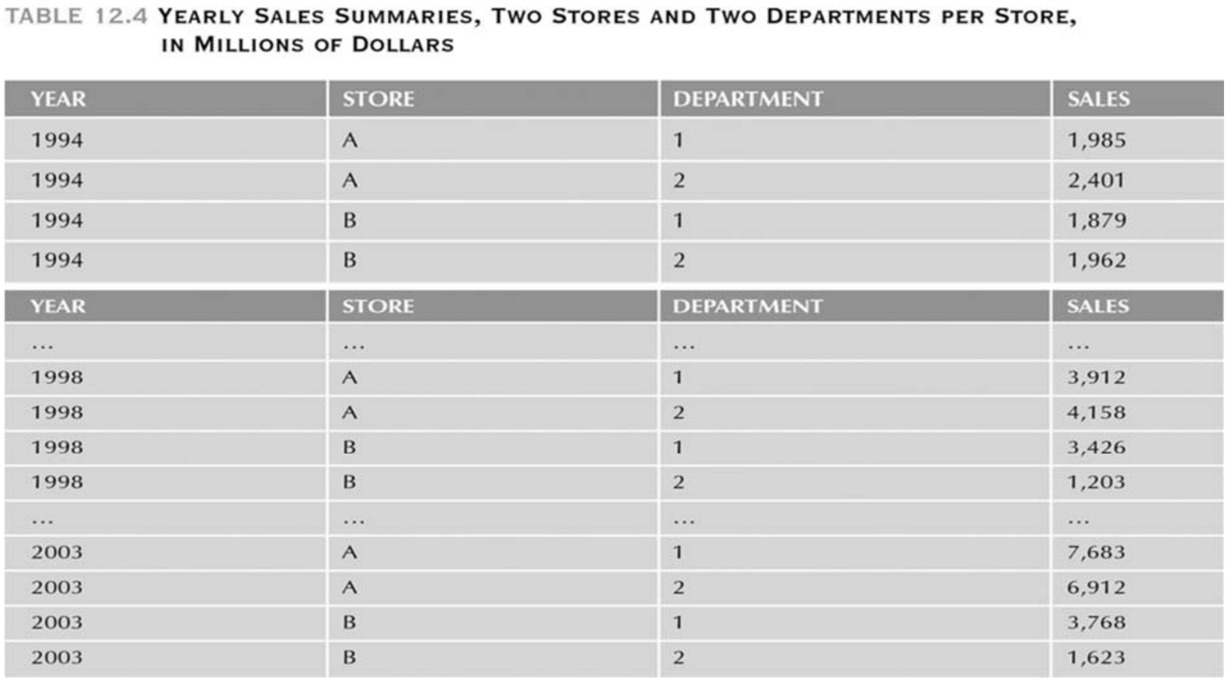


TABLE 12.4 **YEARLY SALES SUMMARIES, TWO STORES AND TWO DEPARTMENTS PER STORE, IN MILLIONS OF DOLLARS**

YEAR	STORE	DEPARTMENT	SALES
1994	A	1	1,985
1994	A	2	2,401
1994	B	1	1,879
1994	B	2	1,962
YEAR	STORE	DEPARTMENT	SALES
...	...	...	...
1998	A	1	3,912
1998	A	2	4,158
1998	B	1	3,426
1998	B	2	1,203
...	...	...	...
2003	A	1	7,683
2003	A	2	6,912
2003	B	1	3,768
2003	B	2	1,623

The Data Warehouse

- Integrated, subject-oriented, time-variant, nonvolatile database that provides support for decision making

Subject-Oriented Data

- Warehouse is organized around major subjects of the enterprise (e.g. customers, products, sales) rather than major application areas (e.g. customer invoicing, stock control, product sales).
- This is reflected in the need to store decision-support data rather than application-oriented data.

Integrated Data

- The data warehouse integrates corporate application-oriented data from different source systems, which often includes data that is inconsistent.
- The integrated data source must be made consistent to present a unified view of the data to the users.

Time-Variant Data

- Data in the warehouse is only accurate and valid at some point in time or over some time interval.
- Time-variance is also shown in the extended time that data is held, the implicit or explicit association of time with all data, and the fact that the data represents a series of snapshots.

Non-Volatile Data

- Data in the warehouse is not updated in real-time but is refreshed from operational systems on a regular basis.
- New data is always added as a supplement to the database, rather than a replacement.

A Comparison of Data Warehouse and Operational Database Characteristics

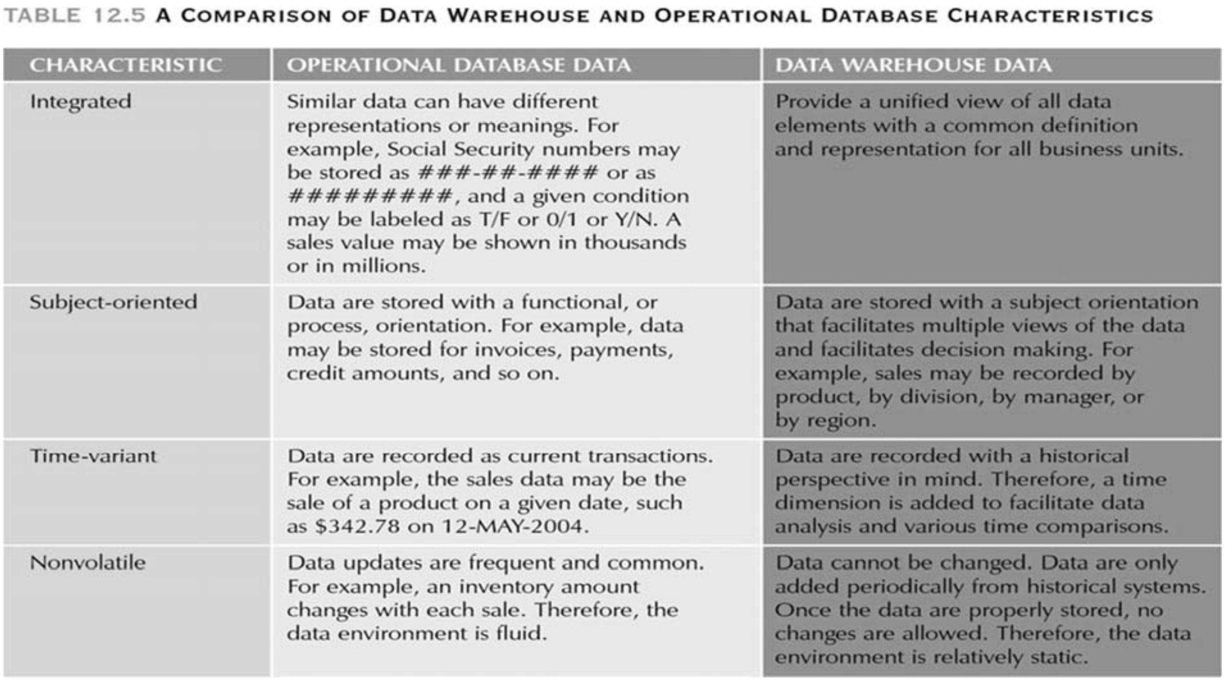


TABLE 12.5 **A COMPARISON OF DATA WAREHOUSE AND OPERATIONAL DATABASE CHARACTERISTICS**

CHARACTERISTIC	OPERATIONAL DATABASE DATA	DATA WAREHOUSE DATA
Integrated	Similar data can have different representations or meanings. For example, Social Security numbers may be stored as ###-##-#### or as #########, and a given condition may be labeled as T/F or 0/1 or Y/N. A sales value may be shown in thousands or in millions.	Provide a unified view of all data elements with a common definition and representation for all business units.
Subject-oriented	Data are stored with a functional, or process, orientation. For example, data may be stored for invoices, payments, credit amounts, and so on.	Data are stored with a subject orientation that facilitates multiple views of the data and facilitates decision making. For example, sales may be recorded by product, by division, by manager, or by region.
Time-variant	Data are recorded as current transactions. For example, the sales data may be the sale of a product on a given date, such as $342.78 on 12-MAY-2004.	Data are recorded with a historical perspective in mind. Therefore, a time dimension is added to facilitate data analysis and various time comparisons.
Nonvolatile	Data updates are frequent and common. For example, an inventory amount changes with each sale. Therefore, the data environment is fluid.	Data cannot be changed. Data are only added periodically from historical systems. Once the data are properly stored, no changes are allowed. Therefore, the data environment is relatively static.

Benefits of Data Warehousing

- Potential high returns on investment
- Competitive advantage
- Increased productivity of corporate decision-makers

Problems of Data Warehousing

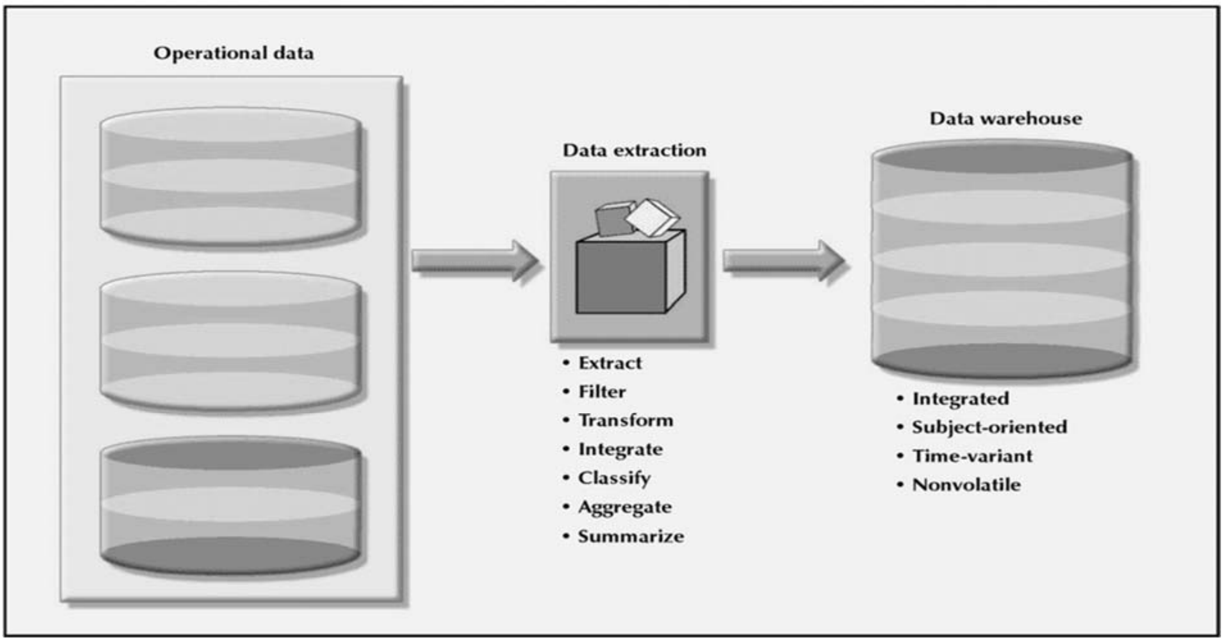
- Underestimation of resources for data loading
- Hidden problems with source systems
- Required data not captured
- Increased end-user demands
- Data homogenization

Problems of Data Warehousing

- High demand for resources
- Data ownership
- High maintenance
- Long duration projects
- Complexity of integration

Creating a Data Warehouse

FIGURE 12.3 CREATING A DATA WAREHOUSE



DSS Architectural Styles

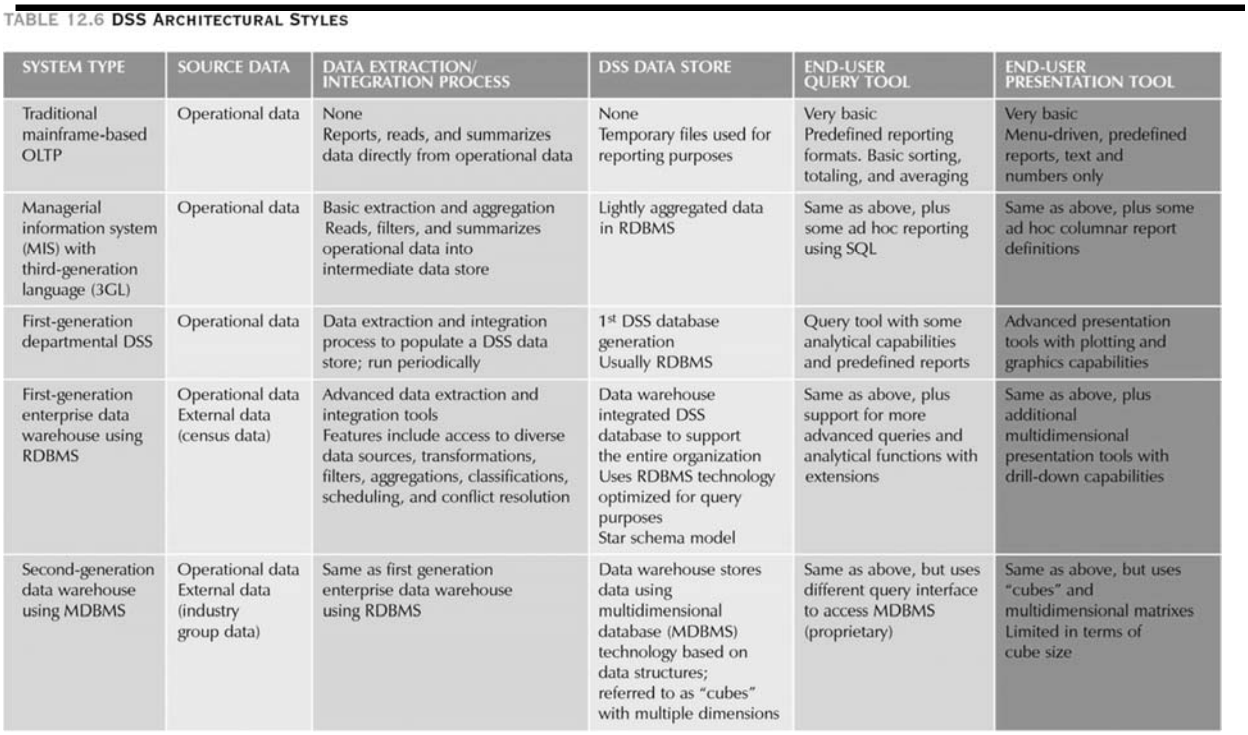


TABLE 12.6 DSS ARCHITECTURAL STYLES

SYSTEM TYPE	SOURCE DATA	DATA EXTRACTION/ INTEGRATION PROCESS	DSS DATA STORE	END-USER QUERY TOOL	END-USER PRESENTATION TOOL
Traditional mainframe-based OLTP	Operational data	None Reports, reads, and summarizes data directly from operational data	None Temporary files used for reporting purposes	Very basic Predefined reporting formats. Basic sorting, totaling, and averaging	Very basic Menu-driven, predefined reports, text and numbers only
Managerial information system (MIS) with third-generation language (3GL)	Operational data	Basic extraction and aggregation Reads, filters, and summarizes operational data into intermediate data store	Lightly aggregated data in RDBMS	Same as above, plus some ad hoc reporting using SQL	Same as above, plus some ad hoc columnar report definitions
First-generation departmental DSS	Operational data	Data extraction and integration process to populate a DSS data store; run periodically	1st DSS database generation Usually RDBMS	Query tool with some analytical capabilities and predefined reports	Advanced presentation tools with plotting and graphics capabilities
First-generation enterprise data warehouse using RDBMS	Operational data External data (census data)	Advanced data extraction and integration tools Features include access to diverse data sources, transformations, filters, aggregations, classifications, scheduling, and conflict resolution	Data warehouse integrated DSS database to support the entire organization Uses RDBMS technology optimized for query purposes Star schema model	Same as above, plus support for more advanced queries and analytical functions with extensions	Same as above, plus additional multidimensional presentation tools with drill-down capabilities
Second-generation data warehouse using MDBMS	Operational data External data (industry group data)	Same as first generation enterprise data warehouse using RDBMS	Data warehouse stores data using multidimensional database (MDBMS) technology based on data structures; referred to as "cubes" with multiple dimensions	Same as above, but uses different query interface to access MDBMS (proprietary)	Same as above, but uses "cubes" and multidimensional matrixes Limited in terms of cube size

Online Analytical Processing

- Advanced data analysis environment that supports decision making, business modeling, and operations research
- OLAP systems share four main characteristics:
  - Use multidimensional data analysis techniques
  - Provide advanced database support
  - Provide easy-to-use end-user interfaces
  - Support client/server architecture

Examples of OLAP Applications in Various Functional Areas

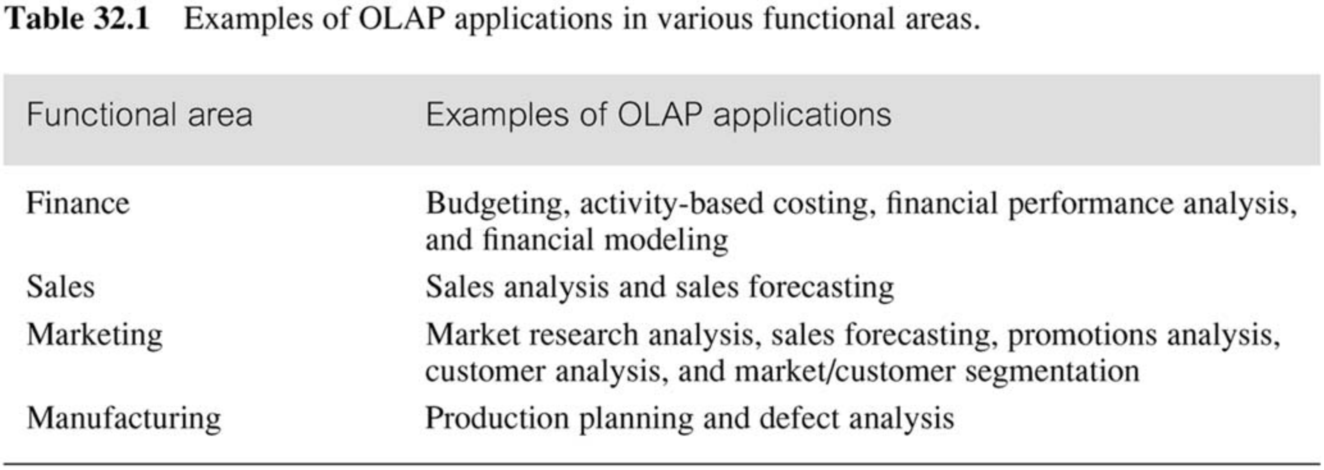
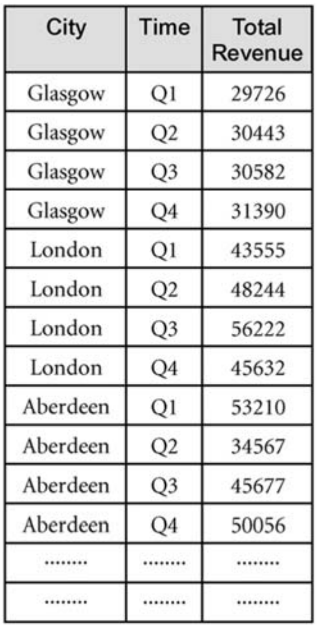


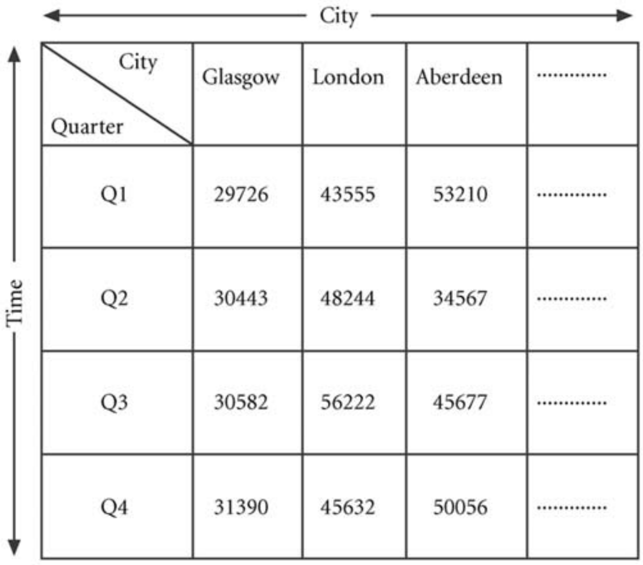
Table 32.1 Examples of OLAP applications in various functional areas.

Functional area	Examples of OLAP applications
Finance	Budgeting, activity-based costing, financial performance analysis, and financial modeling
Sales	Sales analysis and sales forecasting
Marketing	Market research analysis, sales forecasting, promotions analysis, customer analysis, and market/customer segmentation
Manufacturing	Production planning and defect analysis

Multi-Dimensional Data as Three-Field Table versus Two-Dimensional Matrix



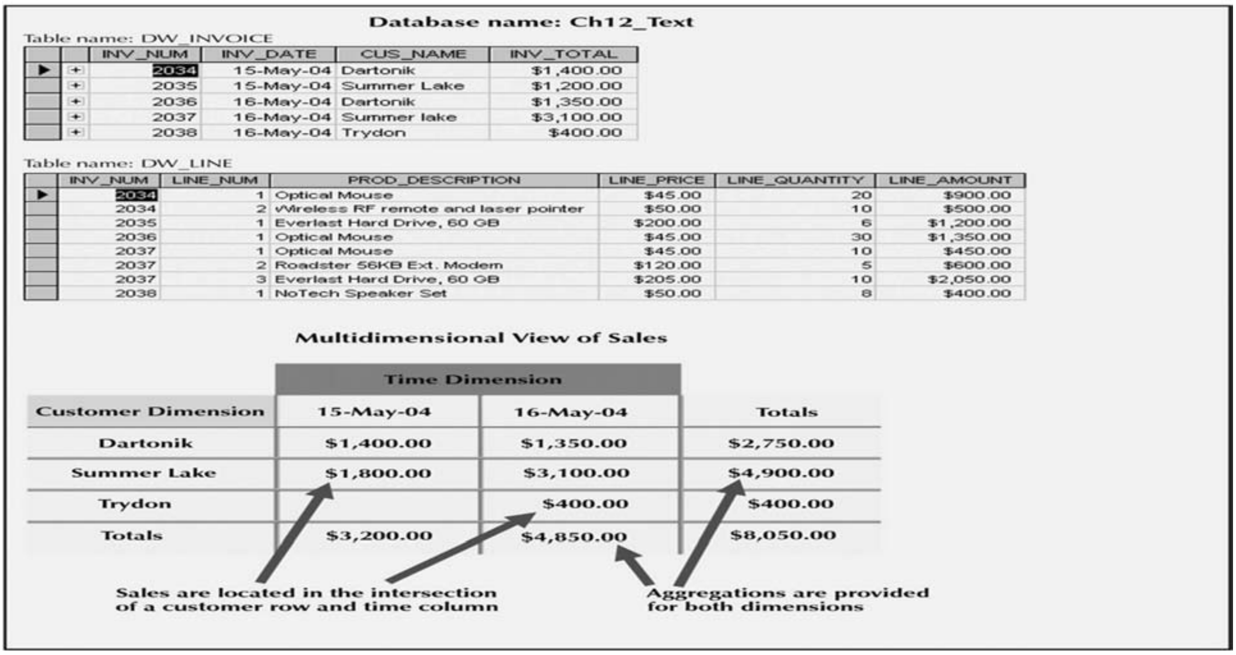
City	Time	Total Revenue
Glasgow	Q1	29726
Glasgow	Q2	30443
Glasgow	Q3	30582
Glasgow	Q4	31390
London	Q1	43555
London	Q2	48244
London	Q3	56222
London	Q4	45632
Aberdeen	Q1	53210
Aberdeen	Q2	34567
Aberdeen	Q3	45677
Aberdeen	Q4	50056
....	....	....
....	....	....



City / Quarter	Glasgow	London	Aberdeen	....
Q1	29726	43555	53210	....
Q2	30443	48244	34567	....
Q3	30582	56222	45677	....
Q4	31390	45632	50056	....

Operational vs. Multidimensional View of Sales

FIGURE 12.4 OPERATIONAL VS. MULTIDIMENSIONAL VIEW OF SALES



Database name: Ch12_Text

Table name: DW_INVOICE

INV_NUM	INV_DATE	CUS_NAME	INV_TOTAL
2034	15-May-04	Dartonik	$1,400.00
2035	15-May-04	Summer Lake	$1,200.00
2036	16-May-04	Dartonik	$1,350.00
2037	16-May-04	Summer lake	$3,100.00
2038	16-May-04	Trydon	$400.00

Table name: DW_LINE

INV_NUM	LINE_NUM	PROD_DESCRIPTION	LINE_PRICE	LINE_QUANTITY	LINE_AMOUNT
2034	1	Optical Mouse	$45.00	20	$900.00
2034	2	Wireless RF remote and laser pointer	$50.00	10	$500.00
2035	1	Everlast Hard Drive, 60 GB	$200.00	6	$1,200.00
2036	1	Optical Mouse	$45.00	30	$1,350.00
2037	1	Optical Mouse	$45.00	10	$450.00
2037	2	Roadster 56KB Ext. Modem	$120.00	5	$600.00
2037	3	Everlast Hard Drive, 60 GB	$205.00	10	$2,050.00
2038	1	NoTech Speaker Set	$50.00	8	$400.00

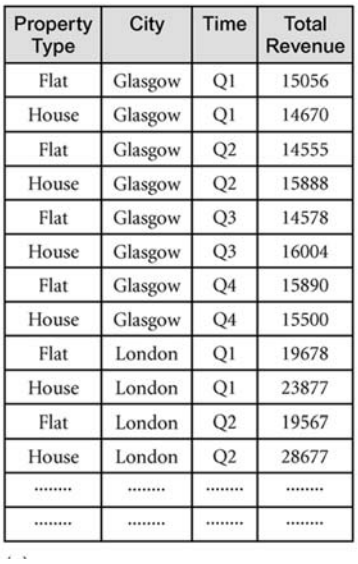
Multidimensional View of Sales

	Time Dimension		
Customer Dimension	15-May-04	16-May-04	Totals
Dartonik	$1,400.00	$1,350.00	$2,750.00
Summer Lake	$1,800.00	$3,100.00	$4,900.00
Trydon		$400.00	$400.00
Totals	$3,200.00	$4,850.00	$8,050.00

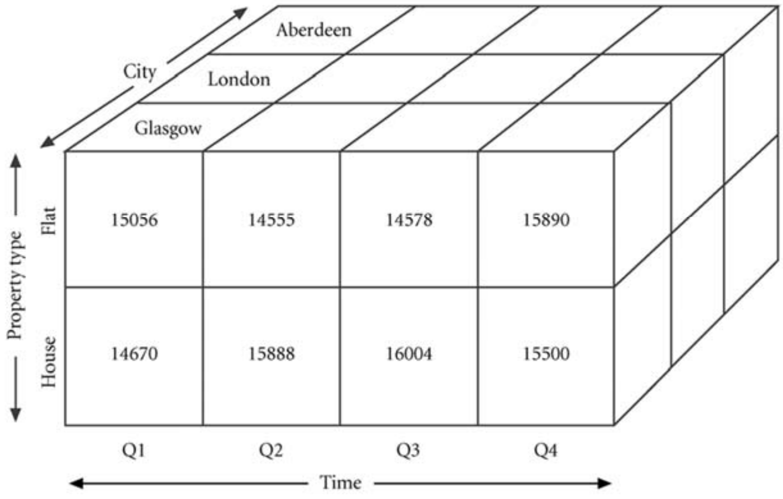
Sales are located in the intersection of a customer row and time column

Aggregations are provided for both dimensions

Multi-Dimensional Data as Four-Field Table versus Three-Dimensional Cube

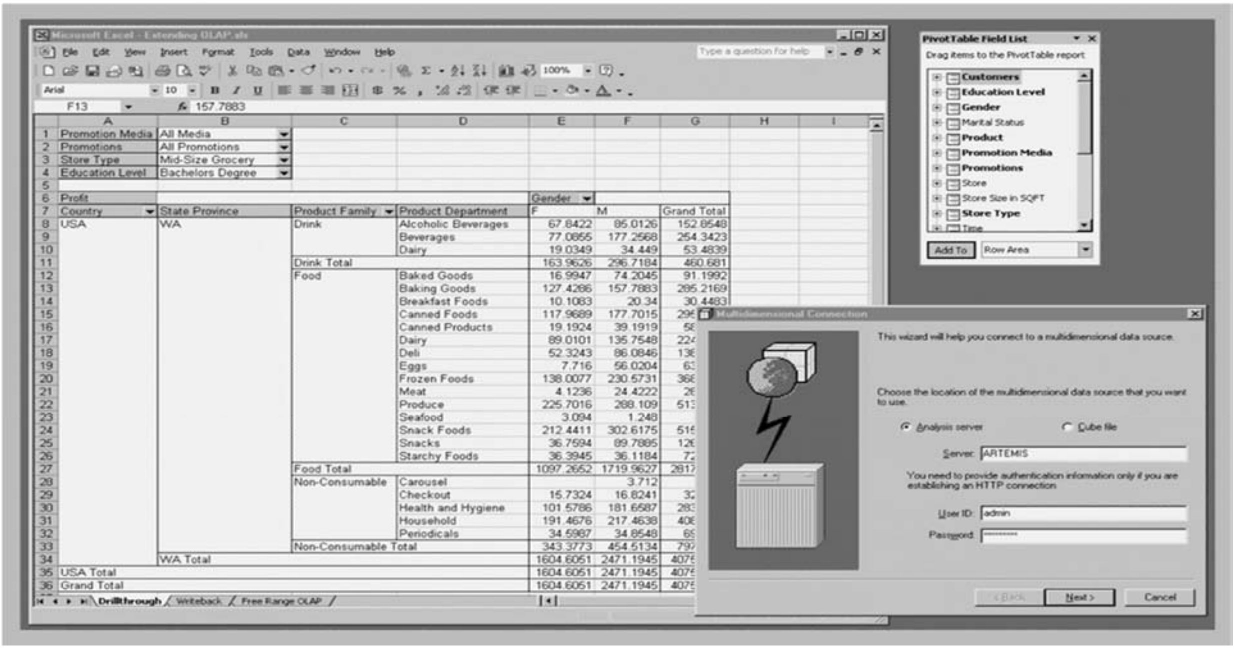


Property Type	City	Time	Total Revenue
Flat	Glasgow	Q1	15056
House	Glasgow	Q1	14670
Flat	Glasgow	Q2	14555
House	Glasgow	Q2	15888
Flat	Glasgow	Q3	14578
House	Glasgow	Q3	16004
Flat	Glasgow	Q4	15890
House	Glasgow	Q4	15500
Flat	London	Q1	19678
House	London	Q1	23877
Flat	London	Q2	19567
House	London	Q2	28677
....	....	....	....
....	....	....	....



Integration of OLAP with a Spreadsheet Program

FIGURE 12.5 INTEGRATION OF OLAP WITH A SPREADSHEET PROGRAM



Microsoft Excel - Extending OLAP.xls
File Edit View Insert Format Tools Data Window Help
Type a question for help
Arial
10
100%

F13 157.7883

	A	B	C	D	E	F	G	H	I
1	Promotion Media	All Media							
2	Promotions	All Promotions							
3	Store Type	Mid-Size Grocery							
4	Education Level	Bachelors Degree							
5									
6	Profit				Gender				
7	Country	State Province	Product Family	Product Department	F	M	Grand Total		
8	USA	WA	Drink	Alcoholic Beverages	67.8422	85.0126	152.8548		
9				Beverages	77.0855	177.2568	254.3423		
10				Dairy	19.0349	34.449	53.4839		
11			Drink Total		163.9626	296.7184	460.681		
12			Food	Baked Goods	16.9947	74.2045	91.1992		
13				Baking Goods	127.4286	157.7883	285.2169		
14				Breakfast Foods	10.1083	20.34	30.4483		
15				Canned Foods	117.9689	177.7015	295		
16				Canned Products	19.1924	39.1919	58		
17				Dairy	89.0101	135.7548	224		
18				Deli	52.3243	86.0846	138		
19				Eggs	7.716	56.0204	63		
20				Frozen Foods	138.0077	230.5731	368		
21				Meat	4.1236	24.4222	28		
22				Produce	225.7016	288.109	513		
23				Seafood	3.094	1.248			
24				Snack Foods	212.4411	302.6175	515		
25				Snacks	36.7594	89.7885	126		
26				Starchy Foods	36.3945	36.1184	72		
27			Food Total		1097.2652	1719.9627	2817		
28			Non-Consumable	Carousel		3.712			
29				Checkout	15.7324	16.8241	32		
30				Health and Hygiene	101.5786	181.6587	283		
31				Household	191.4676	217.4638	408		
32				Periodicals	34.5987	34.8548	69		
33			Non-Consumable Total		343.3773	454.5134	797		
34		WA Total			1604.6051	2471.1945	4075		
35	USA Total				1604.6051	2471.1945	4075		
36	Grand Total				1604.6051	2471.1945	4075		

PivotTable Field List

Drag items to the PivotTable report

- Customers
- Education Level
- Gender
- Marital Status
- Product
- Promotion Media
- Promotions
- Store
- Store Size in SQFT
- Store Type
- Time

Multidimensional Connection

This wizard will help you connect to a multidimensional data source.

Choose the location of the multidimensional data source that you want to use.

● Analysis server ○ Cube file

Server: ARTEMIS

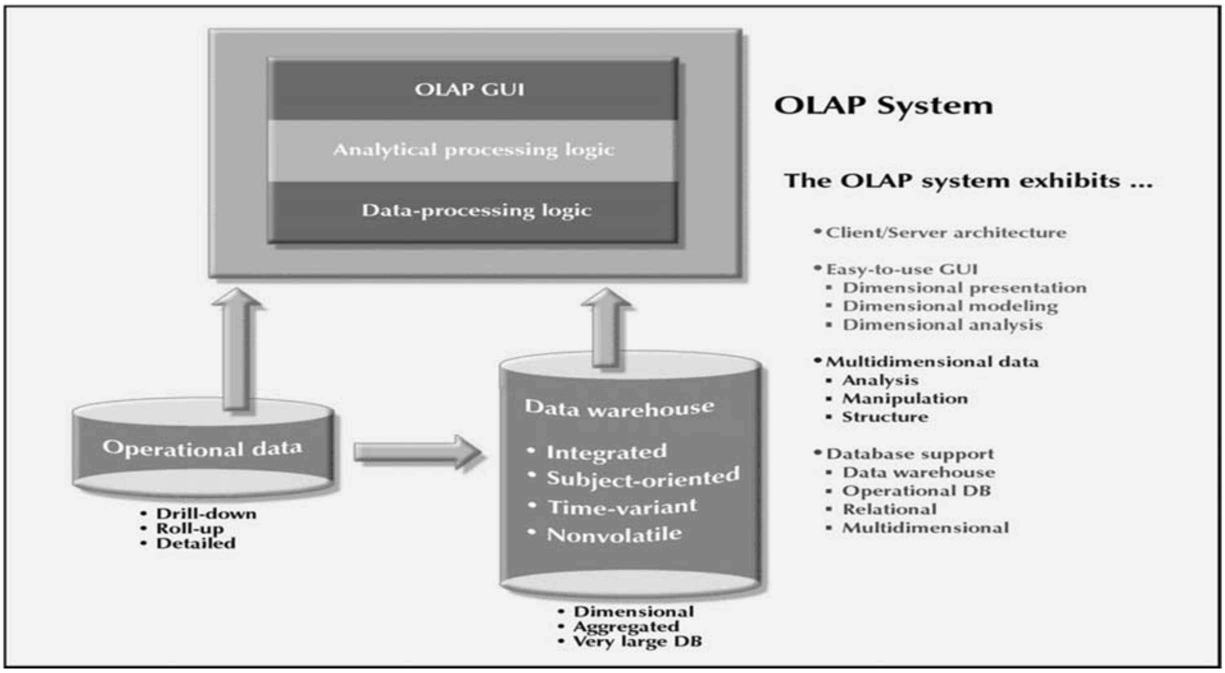
You need to provide authentication information only if you are establishing an HTTP connection.

User ID: admin

Password: **********

OLAP Client/Server Architecture

FIGURE 12.6 OLAP CLIENT/SERVER ARCHITECTURE



OLAP GUI

Analytical processing logic

Data-processing logic

OLAP System

The OLAP system exhibits ...

- Client/Server architecture
- Easy-to-use GUI
  - Dimensional presentation
  - Dimensional modeling
  - Dimensional analysis
- Multidimensional data
  - Analysis
  - Manipulation
  - Structure
- Database support
  - Data warehouse
  - Operational DB
  - Relational
  - Multidimensional

Operational data

- Drill-down
- Roll-up
- Detailed

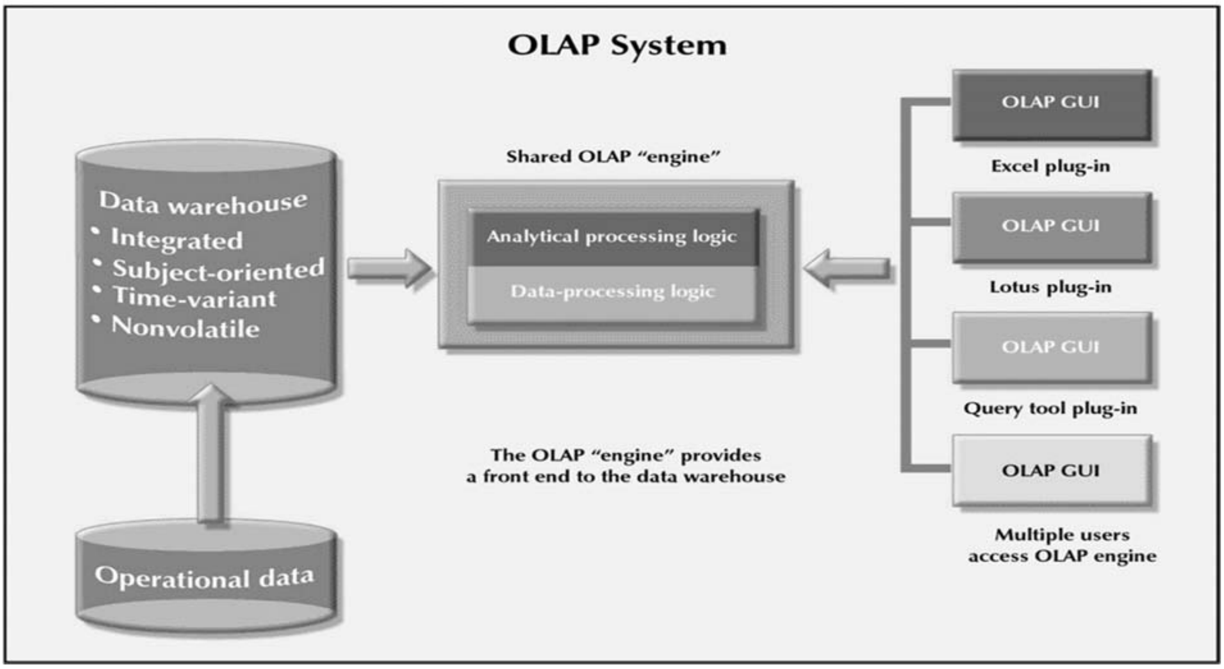
Data warehouse

- Integrated
- Subject-oriented
- Time-variant
- Nonvolatile

- Dimensional
- Aggregated
- Very large DB

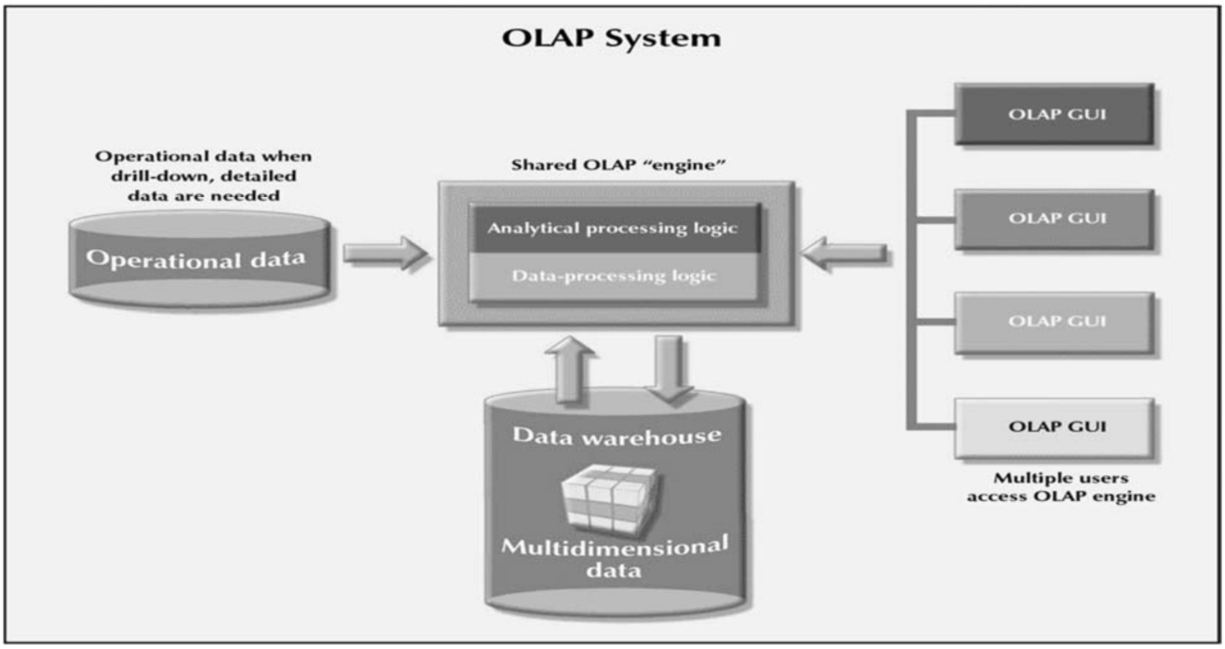
OLAP Server Arrangement

FIGURE 12.7 OLAP SERVER ARRANGEMENT



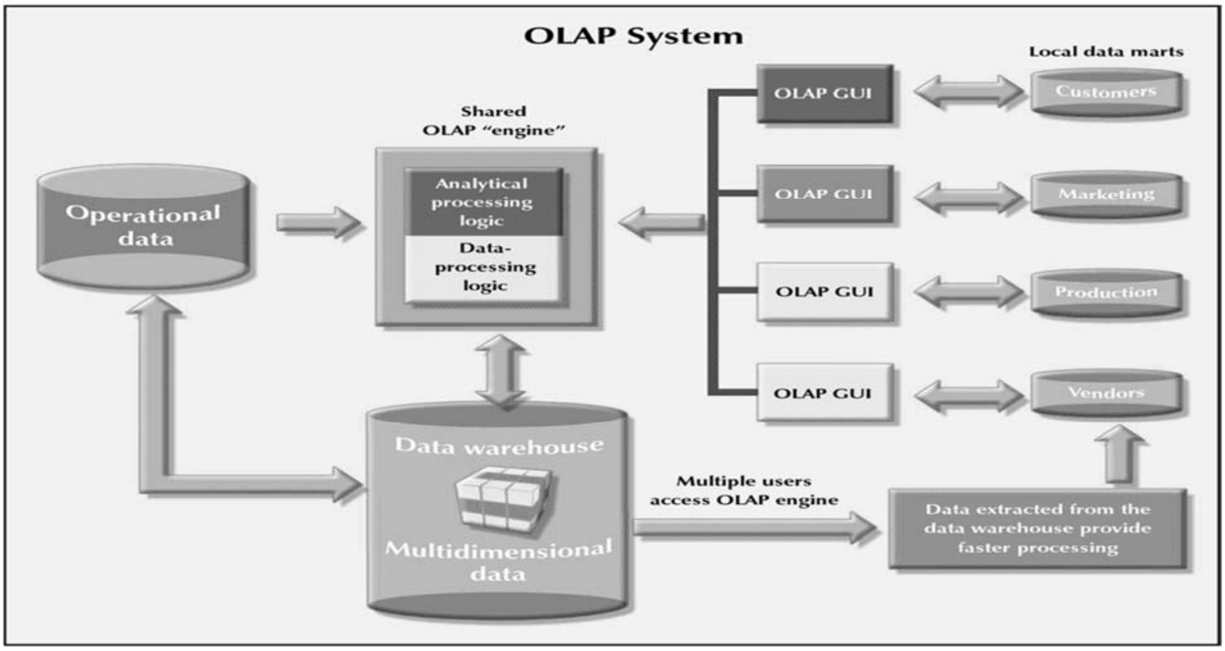
OLAP Server with Multidimensional Data Store Arrangement

FIGURE 12.8 OLAP SERVER WITH MULTIDIMENSIONAL DATA STORE ARRANGEMENT



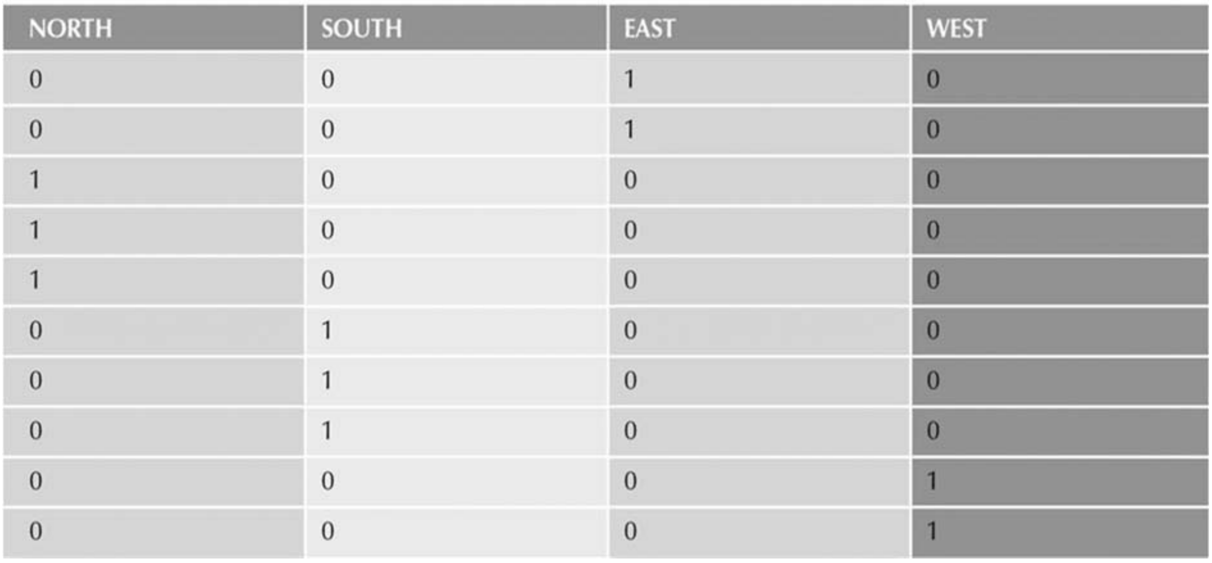
OLAP Server With Local Mini Data Marts

FIGURE 12.9 OLAP SERVER WITH LOCAL MINI DATA MARTS



Bitmap Representation of REGION Values

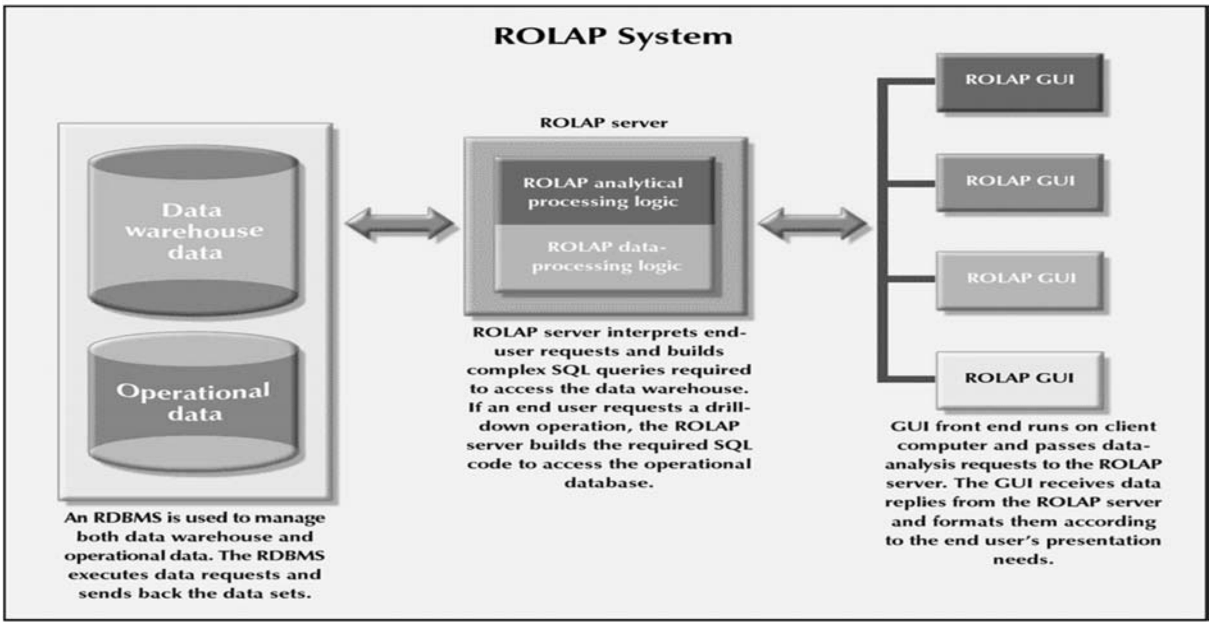
TABLE 12.7 BITMAP REPRESENTATION OF REGION VALUES



NORTH	SOUTH	EAST	WEST
0	0	1	0
0	0	1	0
1	0	0	0
1	0	0	0
1	0	0	0
0	1	0	0
0	1	0	0
0	1	0	0
0	0	0	1
0	0	0	1

Typical ROLAP Client/Server Architecture

FIGURE 12.10 TYPICAL ROLAP CLIENT/SERVER ARCHITECTURE



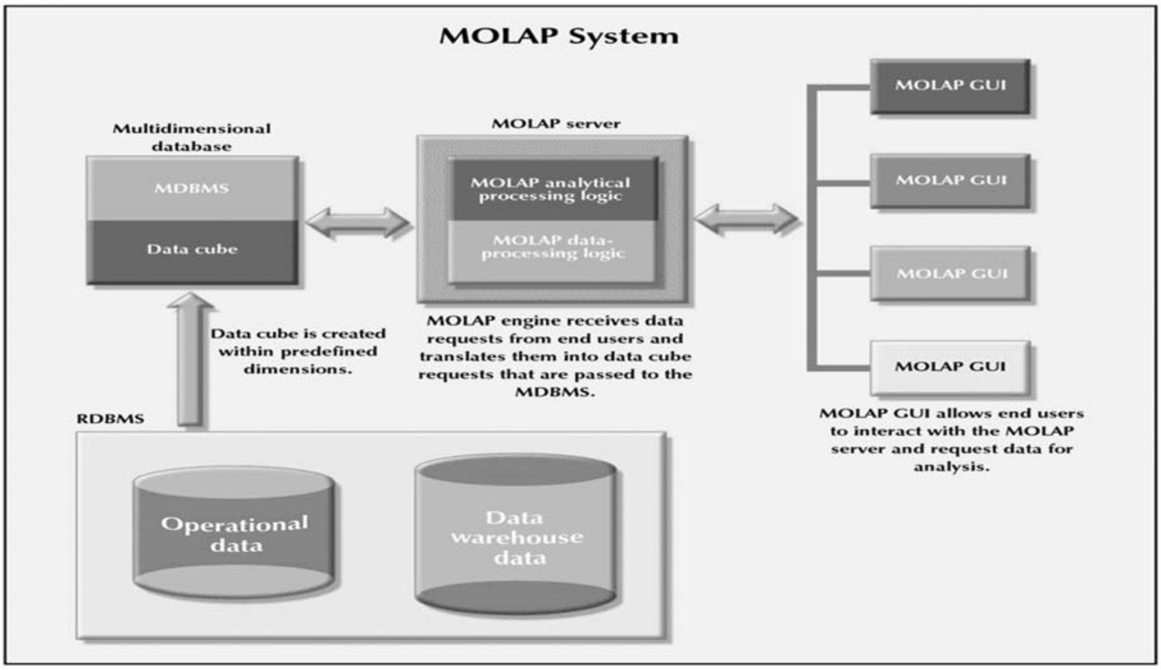
An RDBMS is used to manage both data warehouse and operational data. The RDBMS executes data requests and sends back the data sets.

ROLAP server interprets end-user requests and builds complex SQL queries required to access the data warehouse. If an end user requests a drill-down operation, the ROLAP server builds the required SQL code to access the operational database.

GUI front end runs on client computer and passes data-analysis requests to the ROLAP server. The GUI receives data replies from the ROLAP server and formats them according to the end user's presentation needs.

MOLAP Client/Server Architecture

FIGURE 12.11 MOLAP CLIENT/SERVER ARCHITECTURE



Data cube is created within predefined dimensions.

MOLAP engine receives data requests from end users and translates them into data cube requests that are passed to the MDBMS.

MOLAP GUI allows end users to interact with the MOLAP server and request data for analysis.

Multi-Dimensional OLAP Servers

- In summary, pre-aggregation, dimensional hierarchy, and sparse data management can significantly reduce the size of the cube and the need to calculate values ‘on-the-fly’.
- Removes need for multi-table joins and provides quick and direct access to arrays of data, thus significantly speeding up execution of multi-dimensional queries.

Relational vs. Multidimensional OLAP

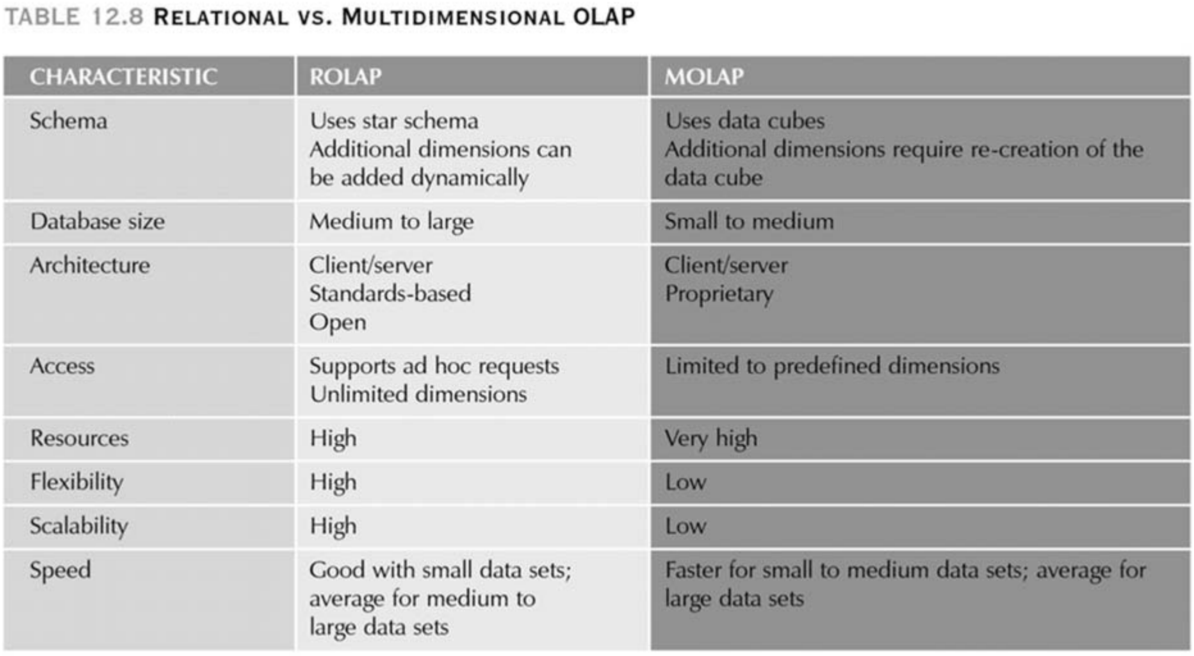


TABLE 12.8 **RELATIONAL VS. MULTIDIMENSIONAL OLAP**

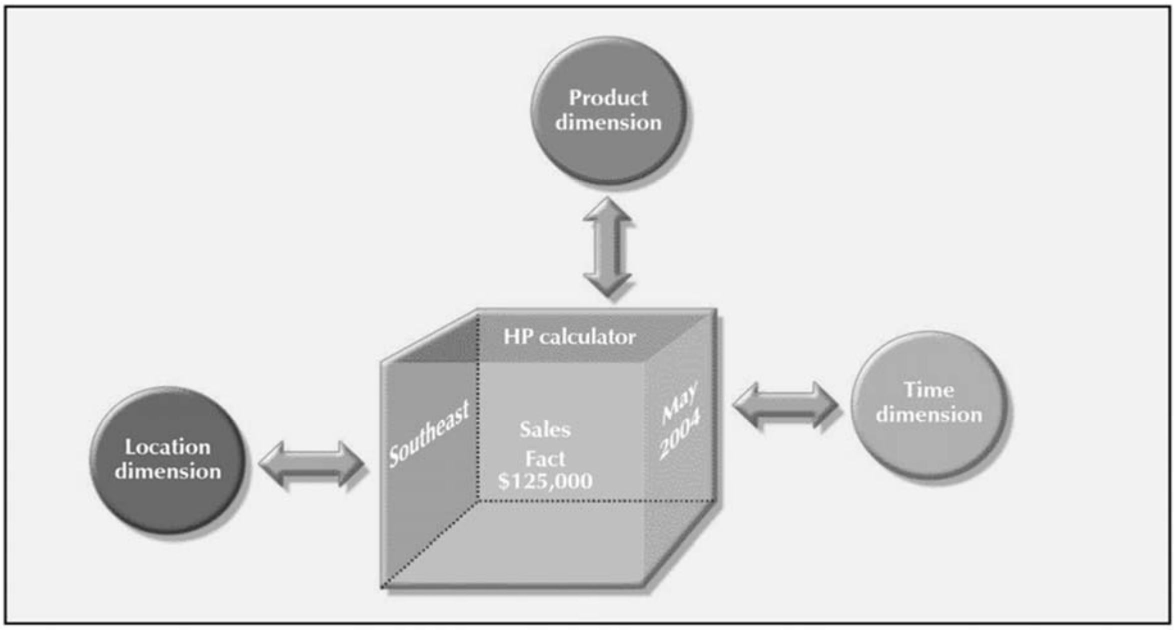
CHARACTERISTIC	ROLAP	MOLAP
Schema	Uses star schema Additional dimensions can be added dynamically	Uses data cubes Additional dimensions require re-creation of the data cube
Database size	Medium to large	Small to medium
Architecture	Client/server Standards-based Open	Client/server Proprietary
Access	Supports ad hoc requests Unlimited dimensions	Limited to predefined dimensions
Resources	High	Very high
Flexibility	High	Low
Scalability	High	Low
Speed	Good with small data sets; average for medium to large data sets	Faster for small to medium data sets; average for large data sets

Star Schemas

- Data modeling technique used to map multidimensional decision support data into a relational database
- Creates the near equivalent of a multidimensional database schema from the existing relational database
- Yield an easily implemented model for multidimensional data analysis, while still preserving the relational structures on which the operational database is built
- Has four components: facts, dimensions, attributes, and attribute hierarchies

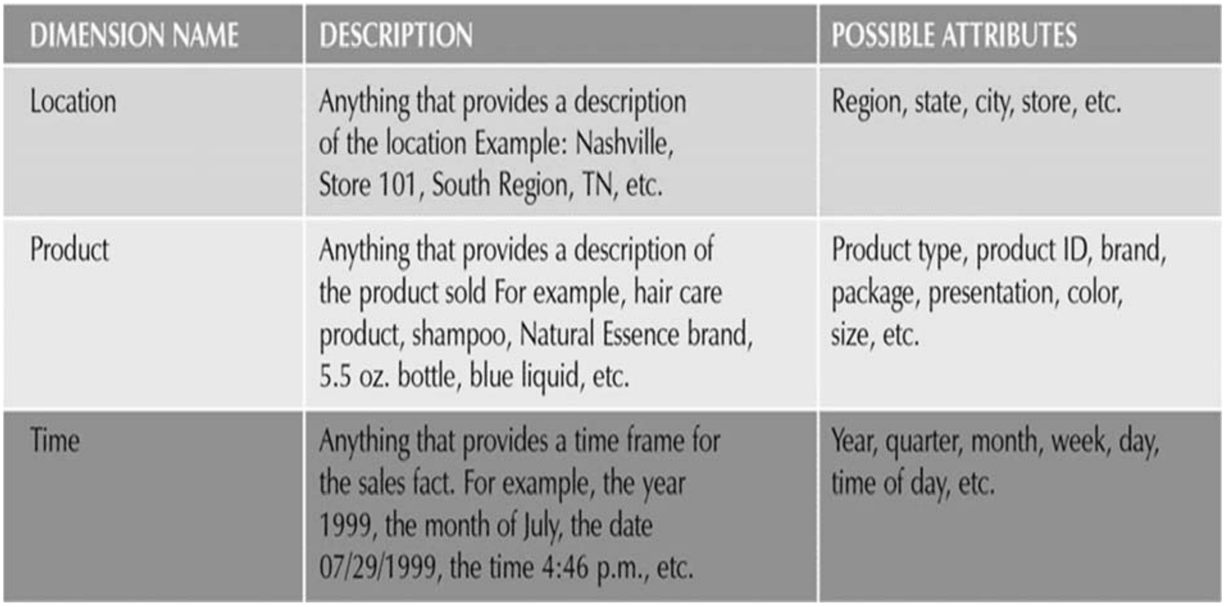
Simple Star Schema

FIGURE 12.12 SIMPLE STAR SCHEMA



Possible Attributes for Sales Dimensions

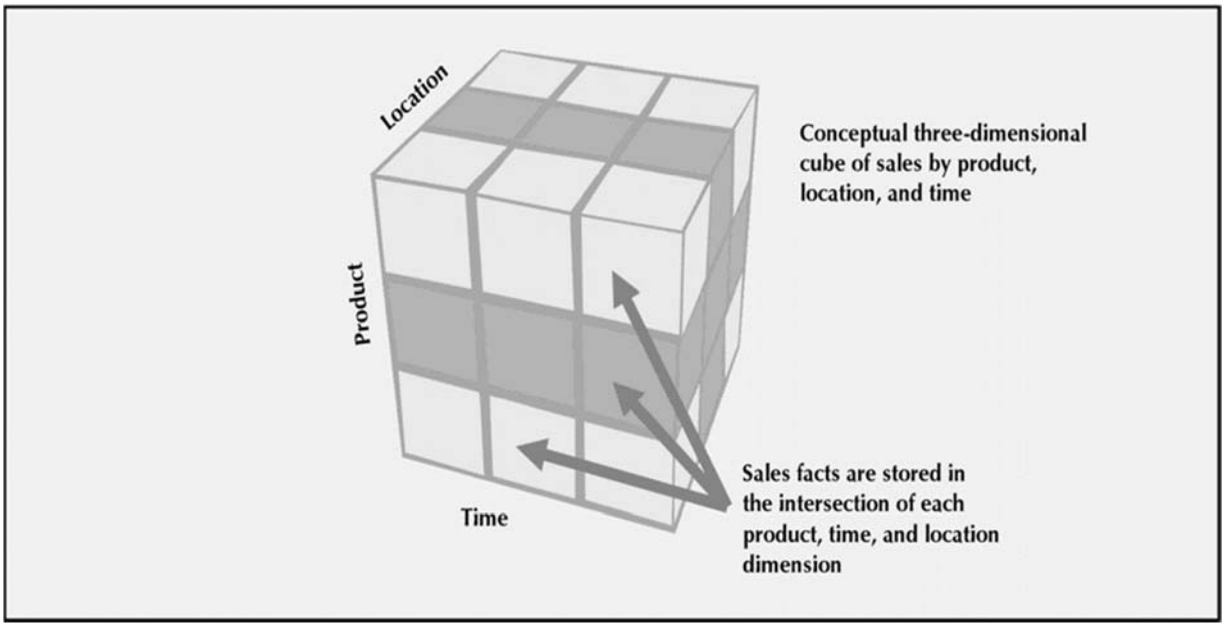
TABLE 12.9 POSSIBLE ATTRIBUTES FOR SALES DIMENSIONS



DIMENSION NAME	DESCRIPTION	POSSIBLE ATTRIBUTES
Location	Anything that provides a description of the location Example: Nashville, Store 101, South Region, TN, etc.	Region, state, city, store, etc.
Product	Anything that provides a description of the product sold For example, hair care product, shampoo, Natural Essence brand, 5.5 oz. bottle, blue liquid, etc.	Product type, product ID, brand, package, presentation, color, size, etc.
Time	Anything that provides a time frame for the sales fact. For example, the year 1999, the month of July, the date 07/29/1999, the time 4:46 p.m., etc.	Year, quarter, month, week, day, time of day, etc.

Three-Dimensional View of Sales

FIGURE 12.13 THREE-DIMENSIONAL VIEW OF SALES

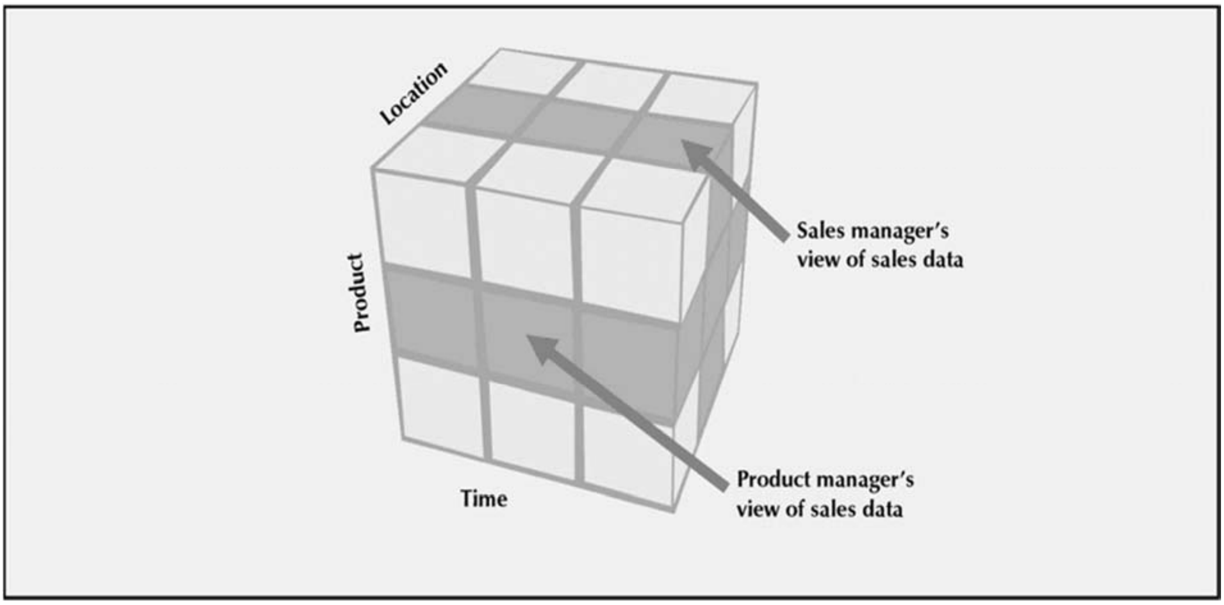


Conceptual three-dimensional cube of sales by product, location, and time

Sales facts are stored in the intersection of each product, time, and location dimension

Slice and Dice View of Sales

FIGURE 12.14 **Slice and Dice View of Sales**

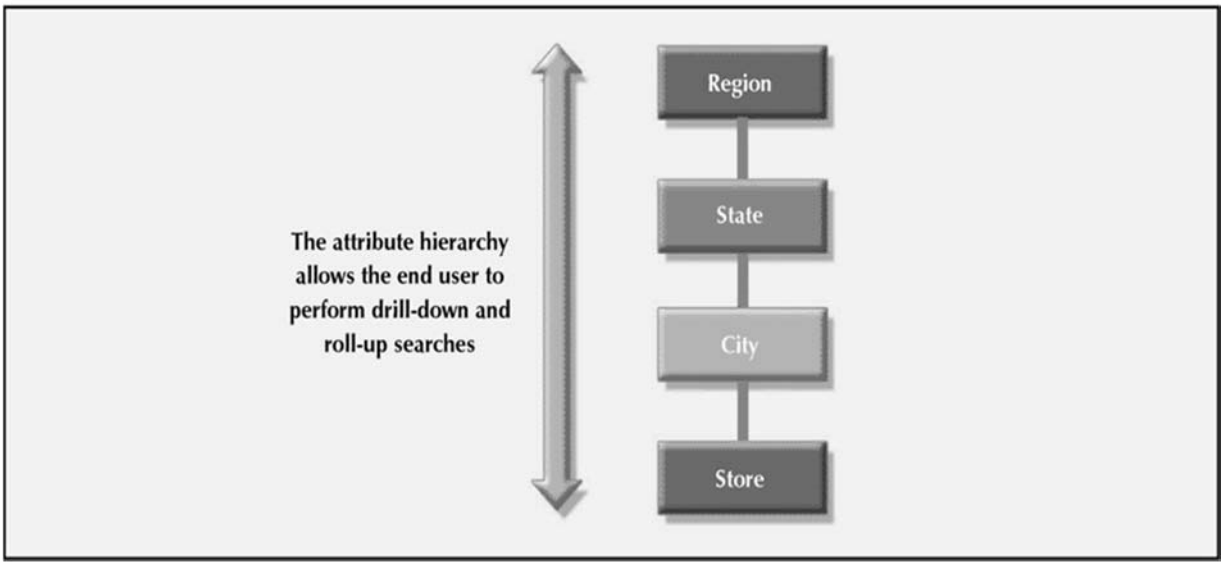


Sales manager's view of sales data

Product manager's view of sales data

Location Attribute Hierarchy

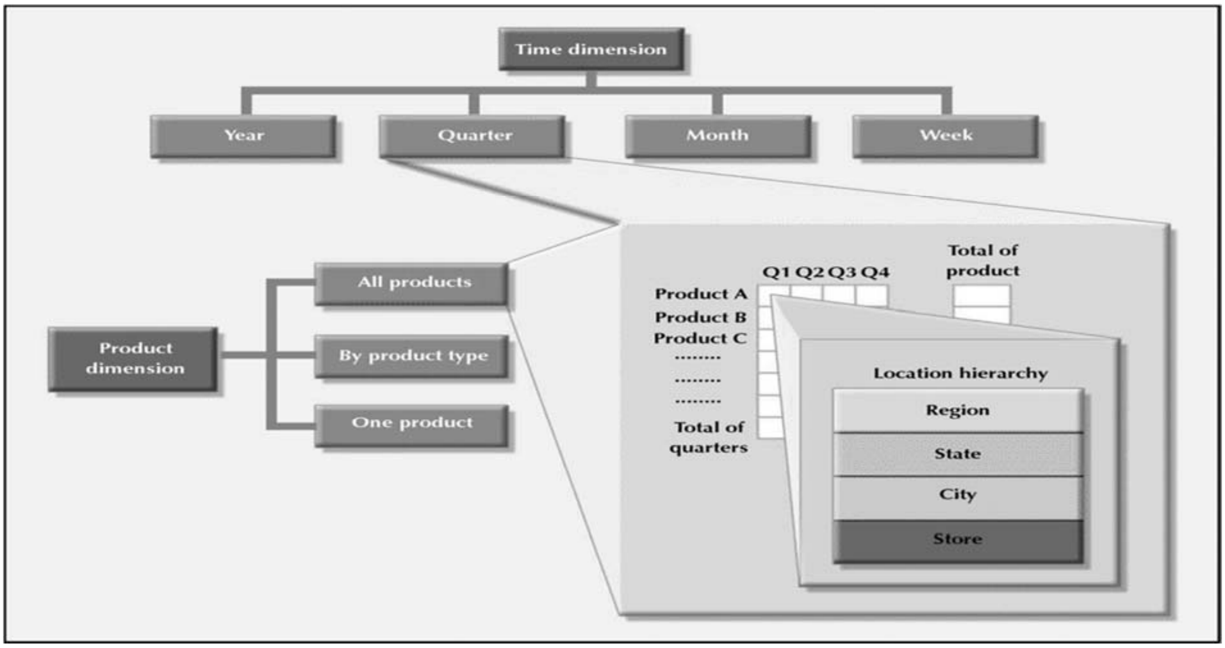
FIGURE 12.15 **Location Attribute Hierarchy**



The attribute hierarchy allows the end user to perform drill-down and roll-up searches

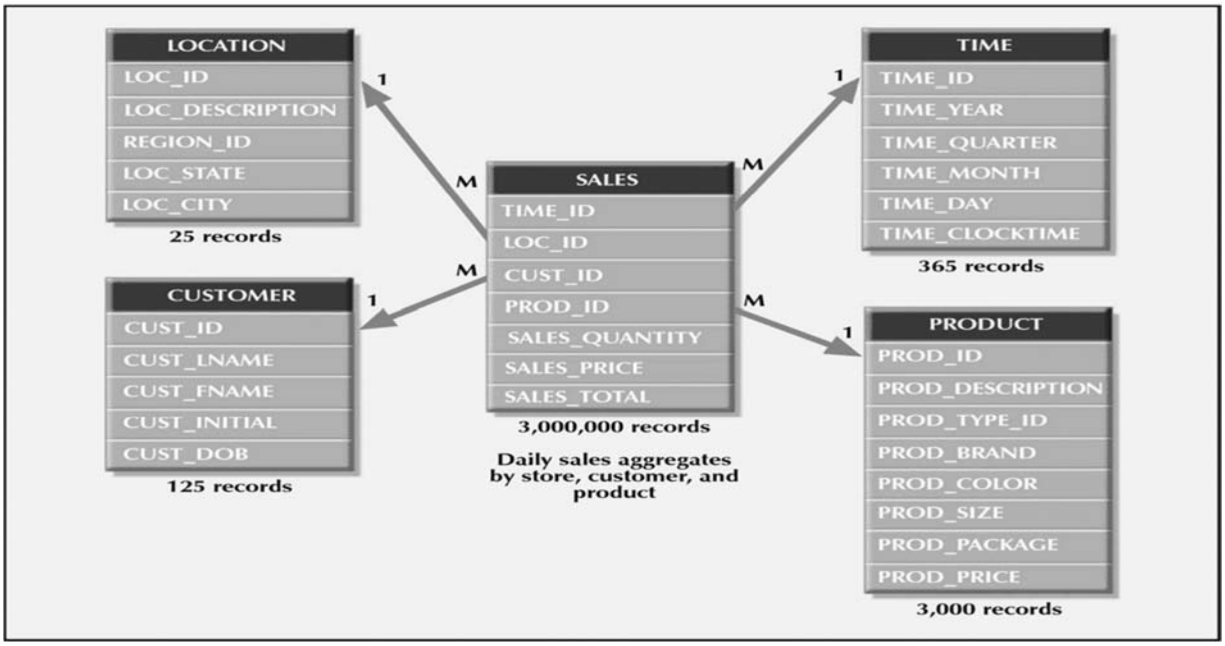
Attribute Hierarchies In Multidimensional Analysis

FIGURE 12.16 ATTRIBUTE HIERARCHIES IN MULTIDIMENSIONAL ANALYSIS



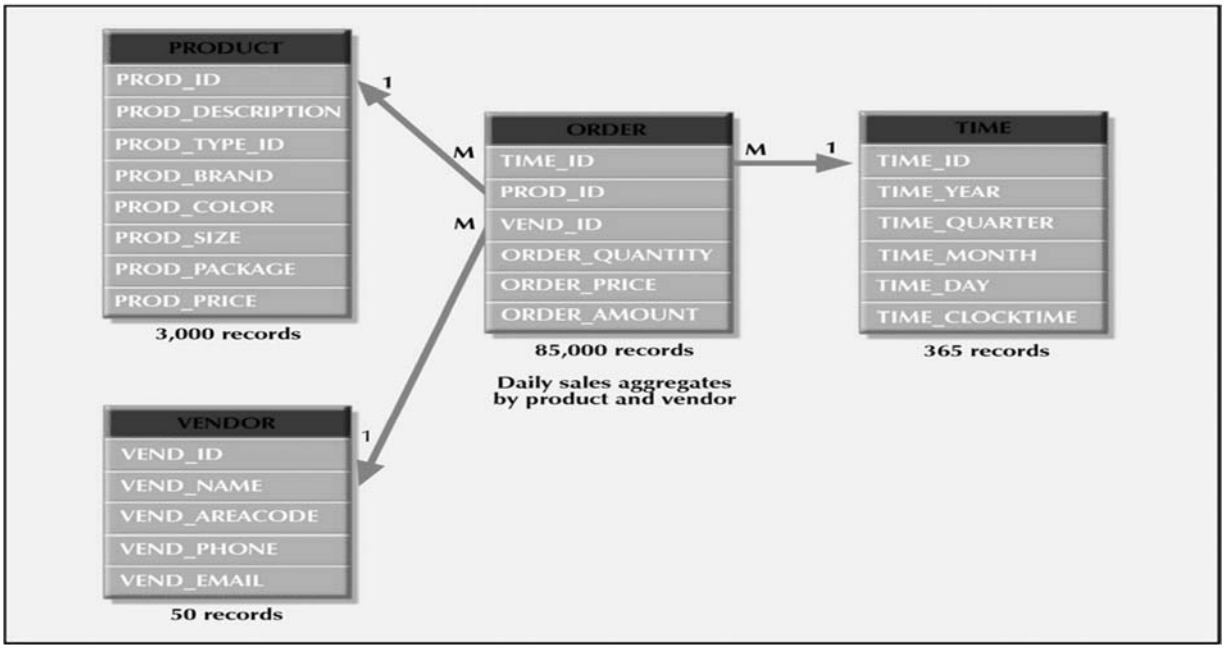
Star Schema for Sales

FIGURE 12.17 STAR SCHEMA FOR SALES



Orders Star Schema

FIGURE 12.18 ORDERS STAR SCHEMA



PRODUCT
PROD_ID
PROD_DESCRIPTION
PROD_TYPE_ID
PROD_BRAND
PROD_COLOR
PROD_SIZE
PROD_PACKAGE
PROD_PRICE
3,000 records

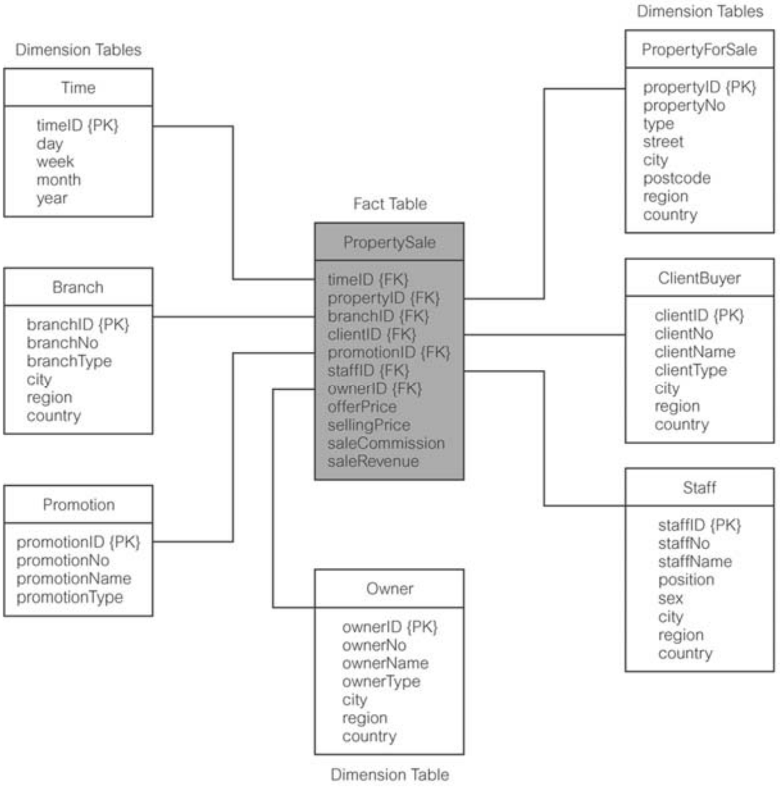
ORDER
TIME_ID
PROD_ID
VEND_ID
ORDER_QUANTITY
ORDER_PRICE
ORDER_AMOUNT
85,000 records

Daily sales aggregates
by product and vendor

TIME
TIME_ID
TIME_YEAR
TIME_QUARTER
TIME_MONTH
TIME_DAY
TIME_CLOCKTIME
365 records

VENDOR
VEND_ID
VEND_NAME
VEND_AREACODE
VEND_PHONE
VEND_EMAIL
50 records

Star Schema for Property Sales of DreamHome



Dimension Tables

Time
timeID {PK}
day
week
month
year

Branch
branchID {PK}
branchNo
branchType
city
region
country

Promotion
promotionID {PK}
promotionNo
promotionName
promotionType

Fact Table

PropertySale
timeID {FK}
propertyID {FK}
branchID {FK}
clientID {FK}
promotionID {FK}
staffID {FK}
ownerID {FK}
offerPrice
sellingPrice
saleCommission
saleRevenue

Owner
ownerID {PK}
ownerNo
ownerName
ownerType
city
region
country

Dimension Table

Dimension Tables

PropertyForSale
propertyID {PK}
propertyNo
type
street
city
postcode
region
country

ClientBuyer
clientID {PK}
clientNo
clientName
clientType
city
region
country

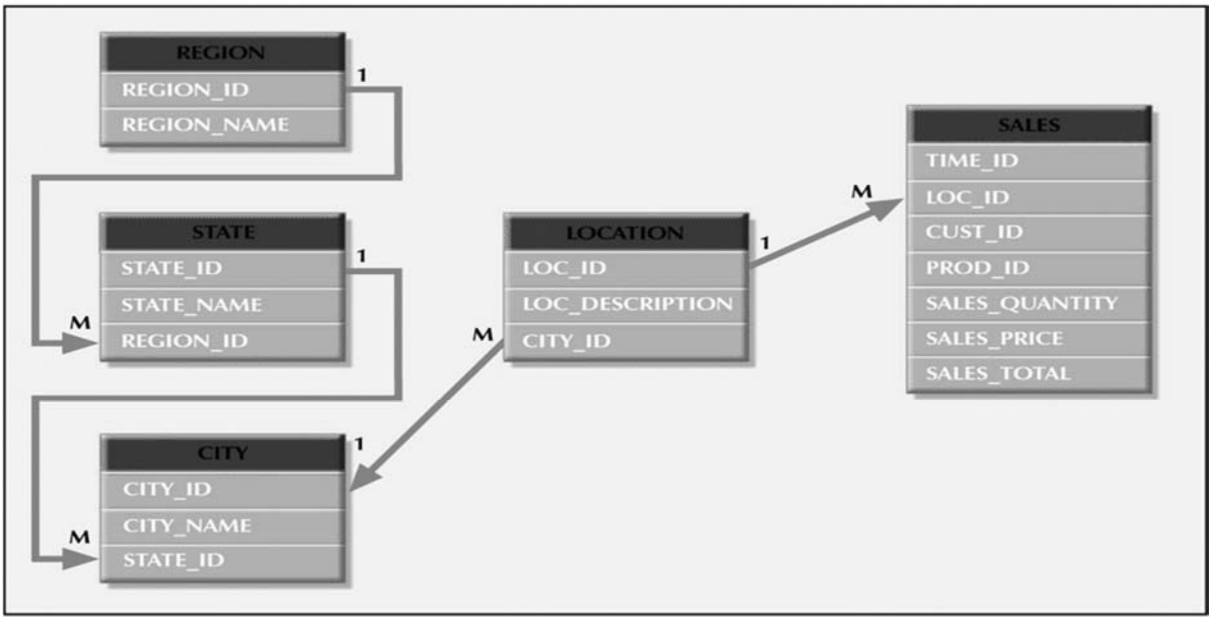
Staff
staffID {PK}
staffNo
staffName
position
sex
city
region
country

Snowflake & Starflake Schema

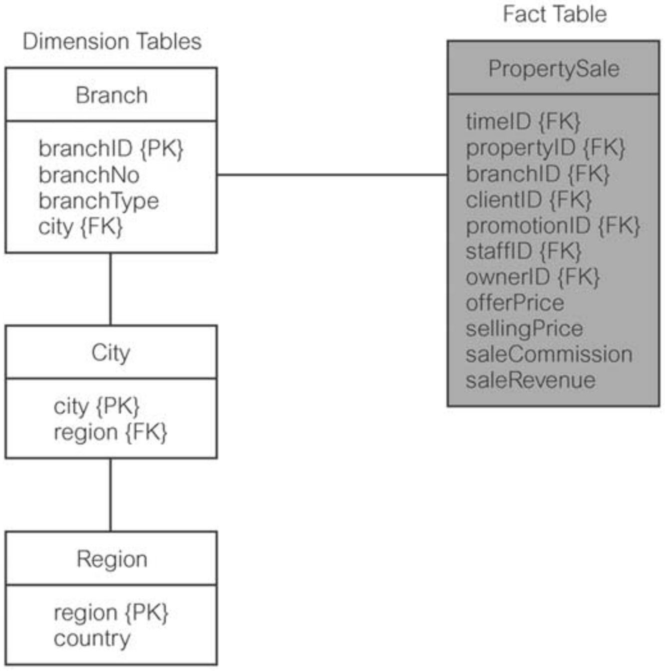
- Snowflake schema is a variant of the star schema where dimension tables do not contain denormalized data.
- Starflake schema is a hybrid structure that contains a mixture of star (denormalized) and snowflake (normalized) schemas. Allows dimensions to be present in both forms to cater for different query requirements.

Normalized Dimension Tables

FIGURE 12.19 NORMALIZED DIMENSION TABLES



Property Sales with Normalized Version of Branch Dimension Table

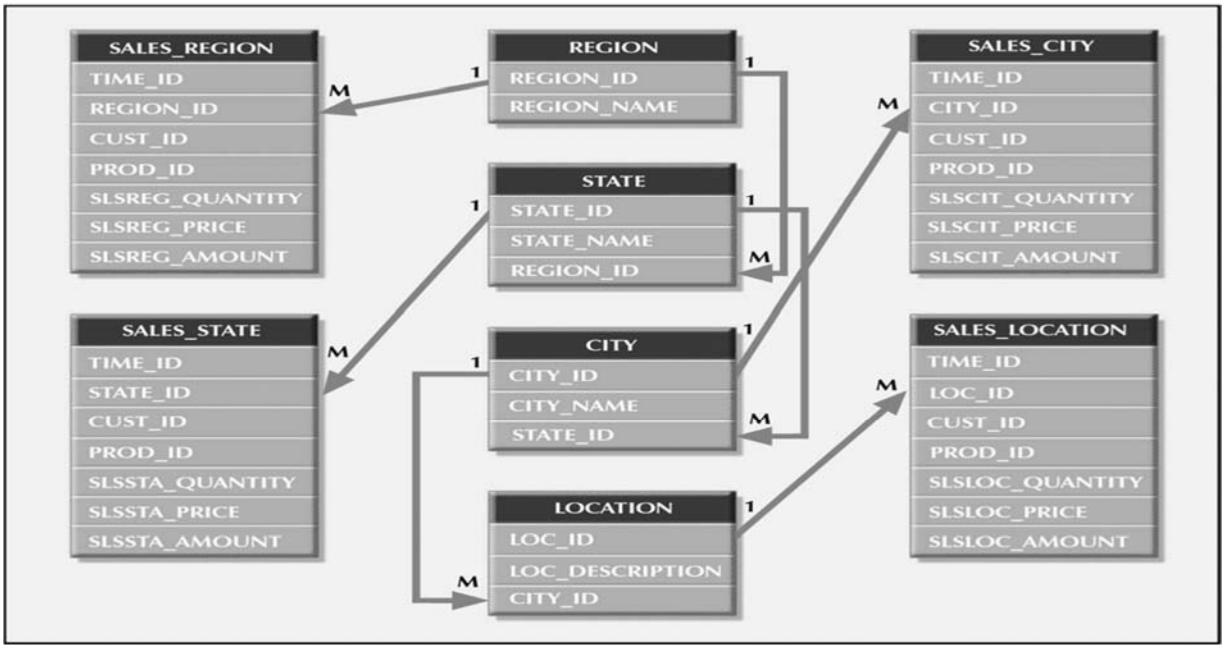


Fact Constellation

- A dimensional model, which contains more than one fact table sharing one or more conformed dimension tables, is referred to as a fact constellation.

Multiple Fact Tables

FIGURE 12.20 MULTIPLE FACT TABLES



SALES_REGION
TIME_ID
REGION_ID
CUST_ID
PROD_ID
SLSREG_QUANTITY
SLSREG_PRICE
SLSREG_AMOUNT

REGION
REGION_ID
REGION_NAME

SALES_CITY
TIME_ID
CITY_ID
CUST_ID
PROD_ID
SLSCIT_QUANTITY
SLSCIT_PRICE
SLSCIT_AMOUNT

STATE
STATE_ID
STATE_NAME
REGION_ID

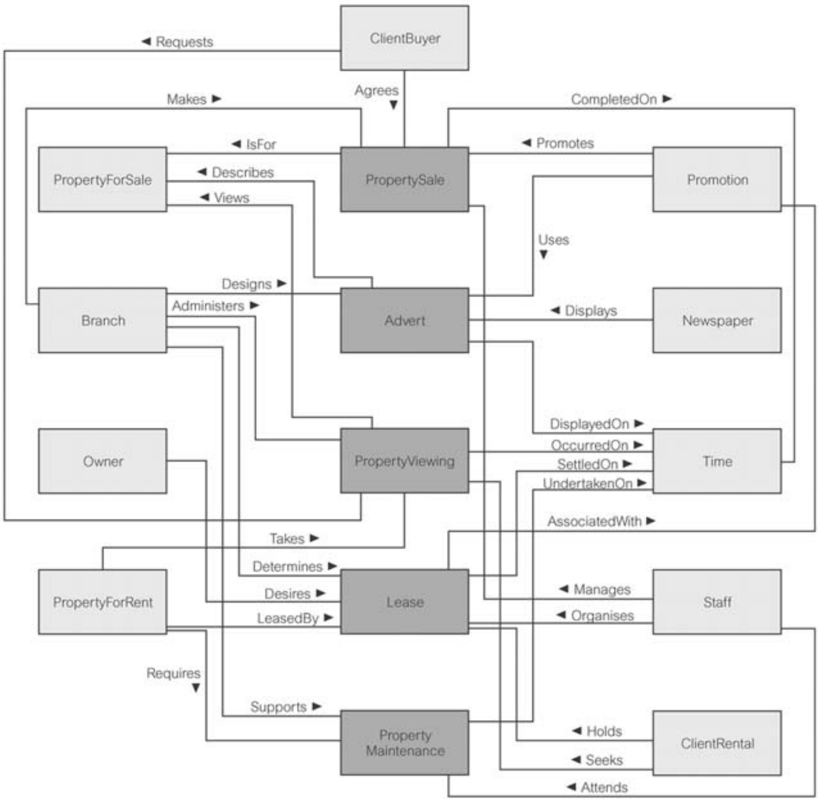
SALES_STATE
TIME_ID
STATE_ID
CUST_ID
PROD_ID
SLSSTA_QUANTITY
SLSSTA_PRICE
SLSSTA_AMOUNT

CITY
CITY_ID
CITY_NAME
STATE_ID

SALES_LOCATION
TIME_ID
LOC_ID
CUST_ID
PROD_ID
SLSLOC_QUANTITY
SLSLOC_PRICE
SLSLOC_AMOUNT

LOCATION
LOC_ID
LOC_DESCRIPTION
CITY_ID

Dimensional Model (Fact Constellation) for the DreamHome Data Warehouse



ClientBuyer
PropertyForSale
PropertySale
Promotion
Branch
Advert
Newspaper
Owner
PropertyViewing
Time
PropertyForRent
Lease
Staff
Property Maintenance
ClientRental

Implementing a Data Warehouse

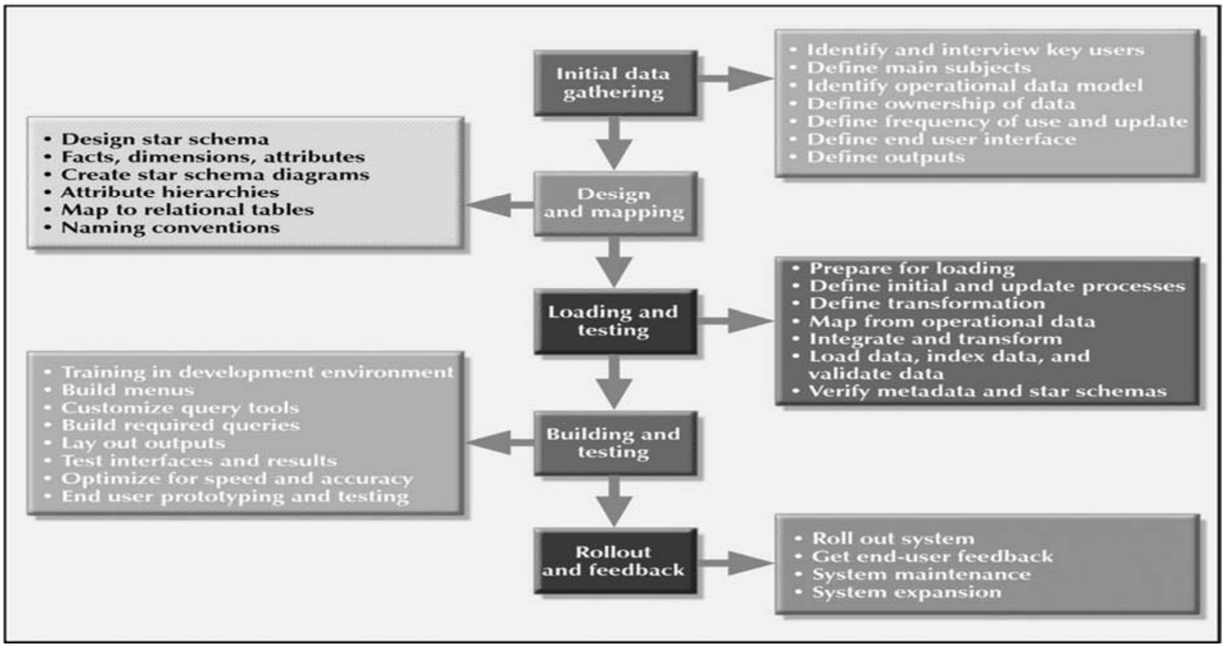
- Numerous constraints:
  - Available funding
  - Management’s view of the role played by an IS department and of the extent and depth of the information requirements
  - Corporate culture
- No single formula can describe perfect data warehouse development

Factors Common to Data Warehousing

- Data warehouse is not a static database
- Dynamic framework for decision support that is always a work in progress
- Data warehouse data cross departmental lines and geographical boundaries
- Must satisfy:
  - Data integration and loading criteria
  - Data analysis capabilities with acceptable query performance
  - End-user data analysis needs
- Apply database design procedures

Data Warehouse Implementation Road Map

FIGURE 12.21 DATA WAREHOUSE IMPLEMENTATION ROAD MAP



Initial data gathering

- Identify and interview key users
- Define main subjects
- Identify operational data model
- Define ownership of data
- Define frequency of use and update
- Define end user interface
- Define outputs

Design and mapping

- Design star schema
- Facts, dimensions, attributes
- Create star schema diagrams
- Attribute hierarchies
- Map to relational tables
- Naming conventions

Loading and testing

- Prepare for loading
- Define initial and update processes
- Define transformation
- Map from operational data
- Integrate and transform
- Load data, index data, and validate data
- Verify metadata and star schemas

Building and testing

- Training in development environment
- Build menus
- Customize query tools
- Build required queries
- Lay out outputs
- Test interfaces and results
- Optimize for speed and accuracy
- End user prototyping and testing

Rollout and feedback

- Roll out system
- Get end-user feedback
- System maintenance
- System expansion

Summary

- Data analysis is used to derive and interpret information from data
- Decision support is a methodology designed to extract information from data and to use such information as a basis for decision making
- Decision support system is an arrangement of computerized tools used to assist managerial decision making within a business
- Data warehouse is an integrated, subject-oriented, time-variant, nonvolatile database that provides support for decision making

Summary (continued)

- Online analytical processing is an advanced data analysis environment that supports decision making, business modeling, and operations research
- Star schema is a data-modeling technique used to map multidimensional decision support data into a relational database
- The implementation of any company-wide information system is subject to conflicting organizational and behavioral factors