IS660J

Lecture 8

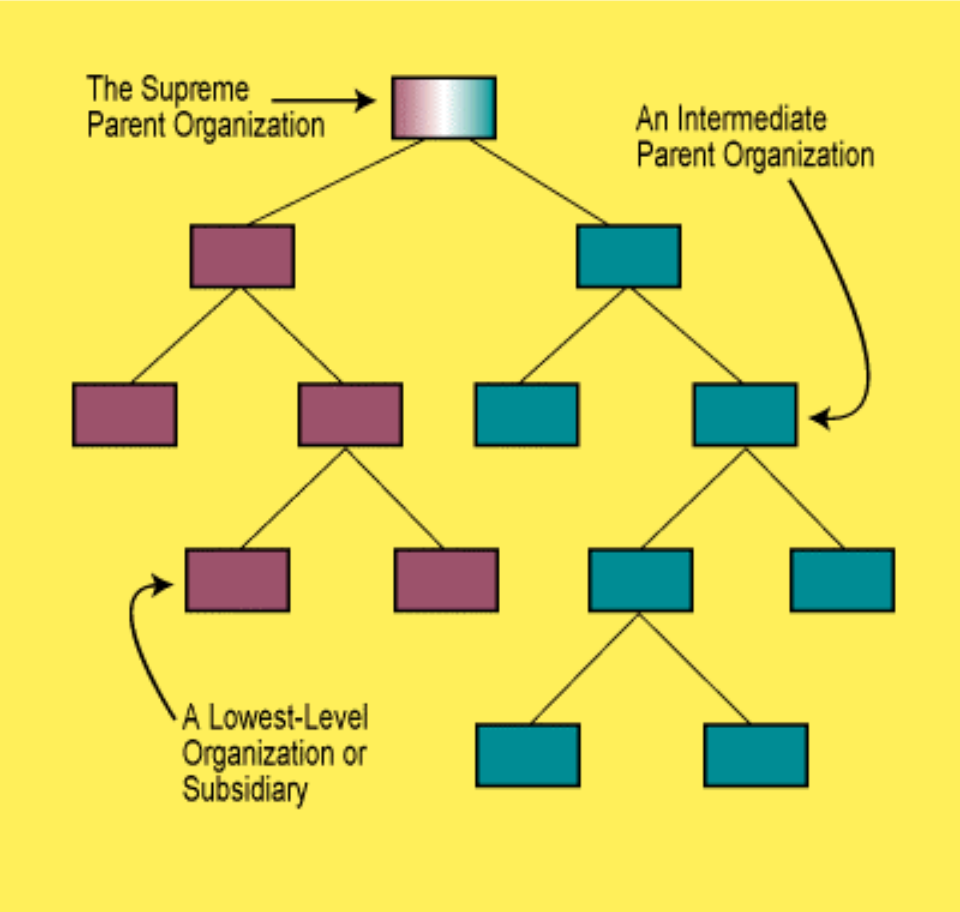
Professor K.M. Burns

Agenda

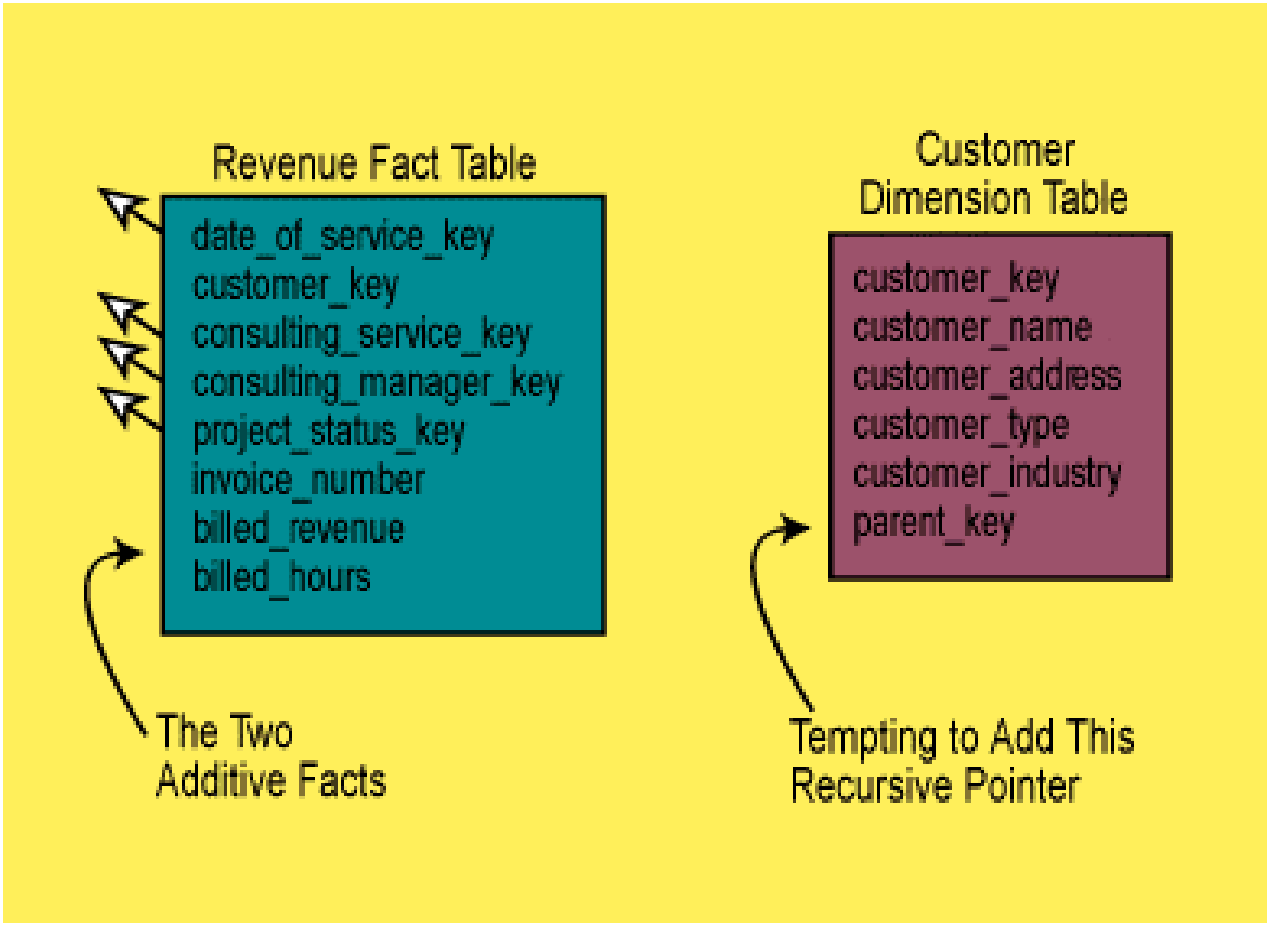
- Review Customer dimension and hierarchies (CRM) – text example
- Decision Support Systems
- OLAP
  - Why OLAP
  - Physical Architecture
  - Market
  - Demonstration
  - Potential Applications
- Wrap-up

Customer Hierarchies

Variable Depth Hierarchies



Example – Consulting Services

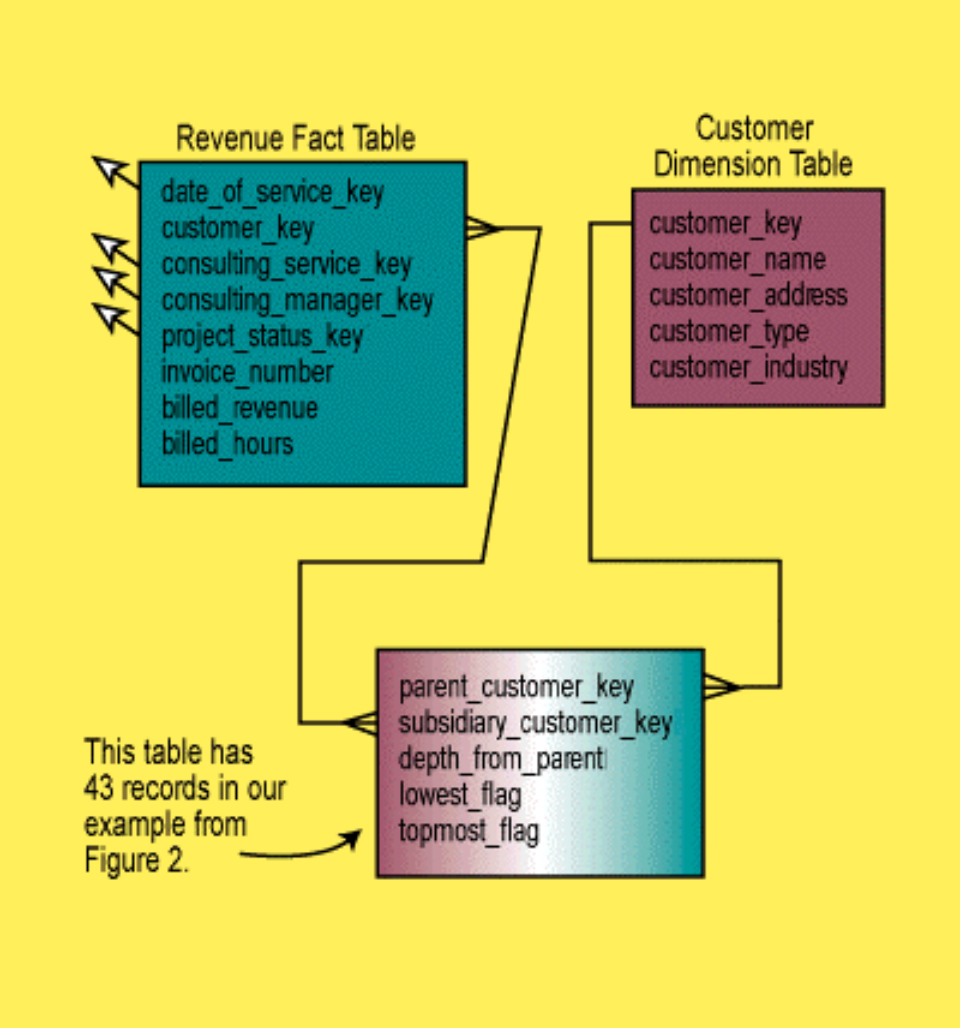


User would like to create reports that show total consulting sold not only to individual departments, but also to divisions, subsidiaries, and overall enterprises

Problem

Standard SQL makes no attempt to deal with recursive pointers, and even such facilities as Oracle's CONNECT BY do not let you use a join in the same SQL statement as CONNECT BY

Solution



The helper table contains one record for *each separate path from each node in the organization tree to itself and to every node below it*. There are, then, more records in the helper table than there are nodes in the tree.

Decision Support Systems

As the name implies, decision support systems are designed to empower the user with the ability to make effective decisions regarding both the current and future evolution of an organization. To do so, the DSS must not only encapsulate static information, but it must also allow for the abstraction of patterns and trends that would not be immediately obvious. Users must be able to visualize the relationships between such things as customers, vendors, products, inventory, geography, and sales. Moreover, they must understand these variables in a chronological context, since it is the time element that ultimately gives meaning to the observations that are formed.

Knowledge-based systems come in a number of distinct forms, each extending the capabilities of their simpler predecessors.

Evolution

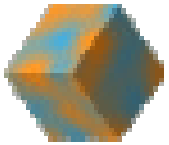
- Information Processing - fundamental querying and reporting functions
  - Query Driven Approaches to integration
  - Update Driven Approaches to integration (data warehouses)
- OLAP - extends the basic reporting capabilities of the first category by allowing a robust multidimensional understanding of the archived, enhancing interactivity, data from a variety of perspectives and hierarchies
- Data Mining - goal is to automate the discovery process so that trends and patterns can be retrieved with minimal user input

OLAP

OLAP

OLAP stands for On Line Analytical Processing. What it really means is multidimensional analysis—the ability to analyze corporate data the way you think about it, interactively.

Analysis



- ✓Slide-Dice
- ✓Drill down
- ✓Roll-up
- ✓Pivoting

OLAP (as Data Warehouses)

- Subject-oriented (business process)
- Integrated
- Time Variant
- Non-volatile

Handout Figure 2.4 and Table 2.3

- **Roll-up**. The roll-up operation collapses the dimension hierarchy along a particular dimension(s) so as to present the remaining dimensions at a coarser level of granularity.
- **Drill-down**. In contrast, the drill-down function allows users to obtain a more detailed view of a given dimension.
- **Slice**. Here, the objective is to extract a slice of the original cube corresponding to a single value of a given dimension. No aggregation is required with this option. Instead, we are allowing the user to focus in on values of interest.
- **Dice**. A related operation is the dice. In this case, we are defining a subcube of the original space. In other words, by specifying value ranges on two or more dimensions, the user can highlight meaningful blocks of aggregated data.
- **Pivot**. The pivot is a simple but effective operation that allows OLAP users to visualize cube values in more natural and intuitive ways.

Dice

Slice

Pivot

Rollup

Drill Down

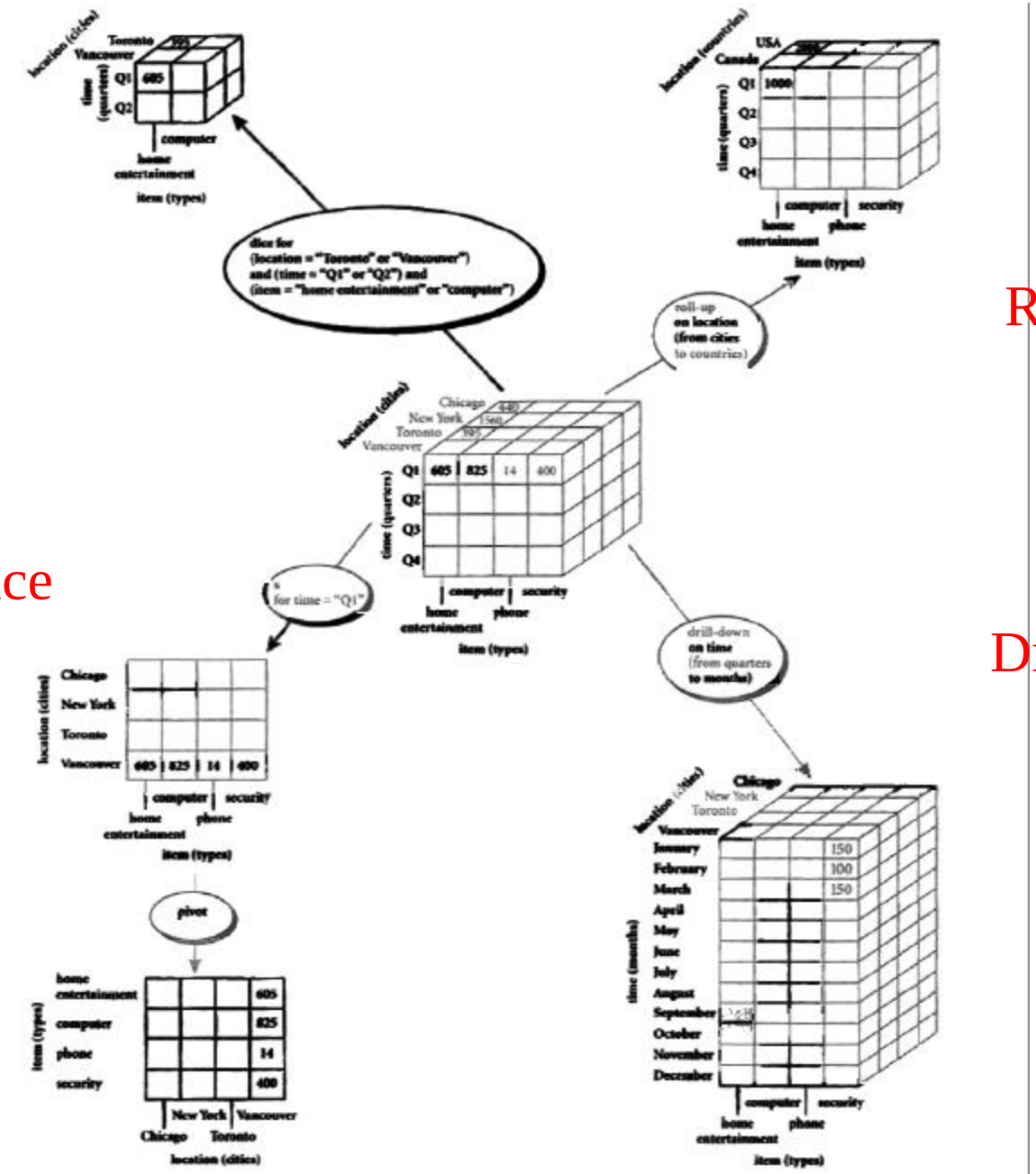
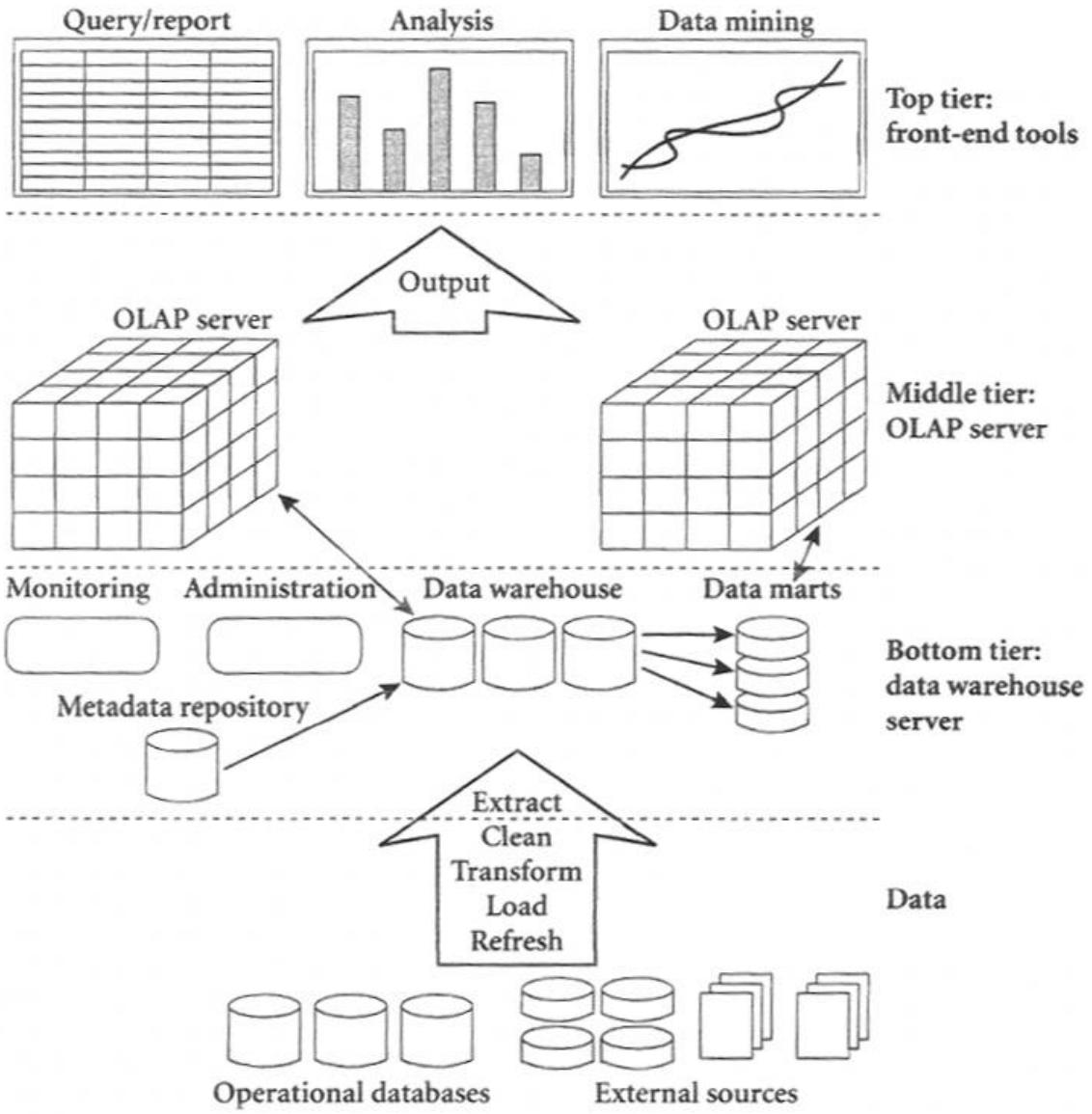


Figure 2.10 Examples of typical OLAP operations on multidimensional data.

Why OLAP?

- Interactivity; Ease of use
- SQL does not have the ability to perform multidimensional calculations in single statements, and complex multi-pass SQL is necessary to achieve more than the most trivial multidimensional functionality.
- In most cases, vendors do a limited range of suitable calculations in SQL, with the results then being used as input by a multidimensional engine, which does most of the work, either on the client or in a mid-tier server. There may also be a RAM resident cache which can hold data used in more than one query: this improves response dramatically.

Physical Architecture



igure 2.12 A three-tier data warehousing architecture.

Scalability & Specialization

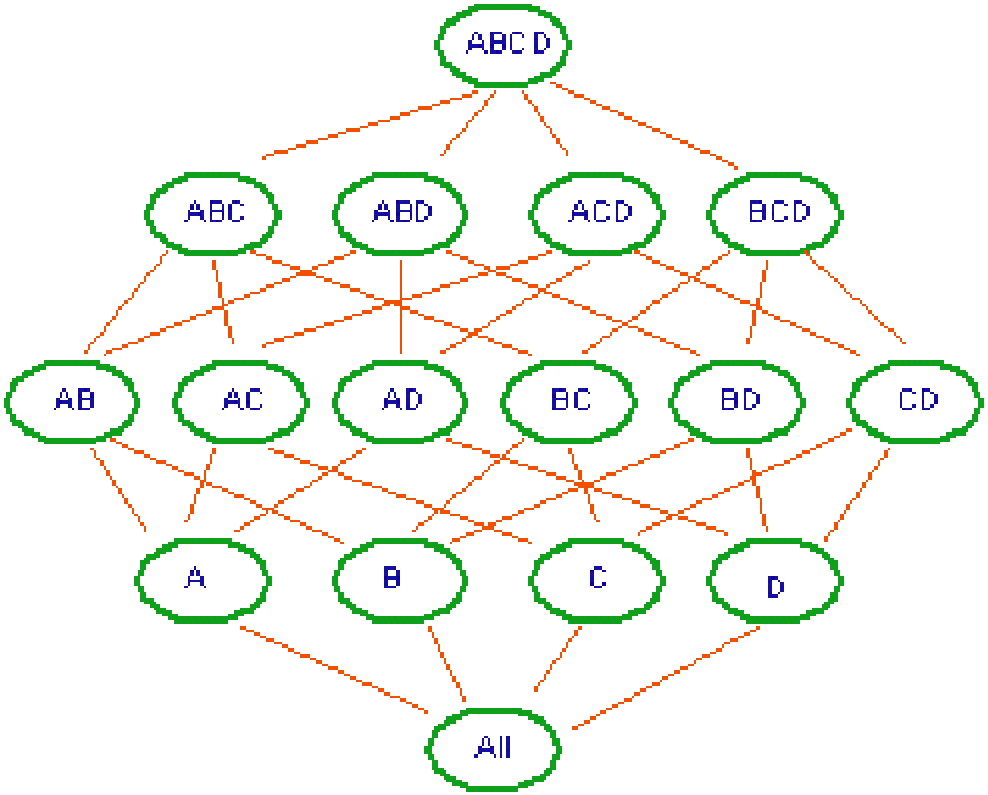
What is a cuboid?

- *n*-D (dimensional) data as a series of (n-1) -D "cubes"
- d-dimensional base cube is associated with 2^d cuboids
- Data at different levels of summarization (group by), or, summarized by a different subset of the dimensions
- Base cuboid holds the the lowest level of summarization
- Apex cubiod holds the highest level of summarization

What is a Data Cube?

- Lattice of cubiods forming a data cube for all the dimensions (ie. time, product, location and supplier)

Handout Figure 2.2 & 2.3



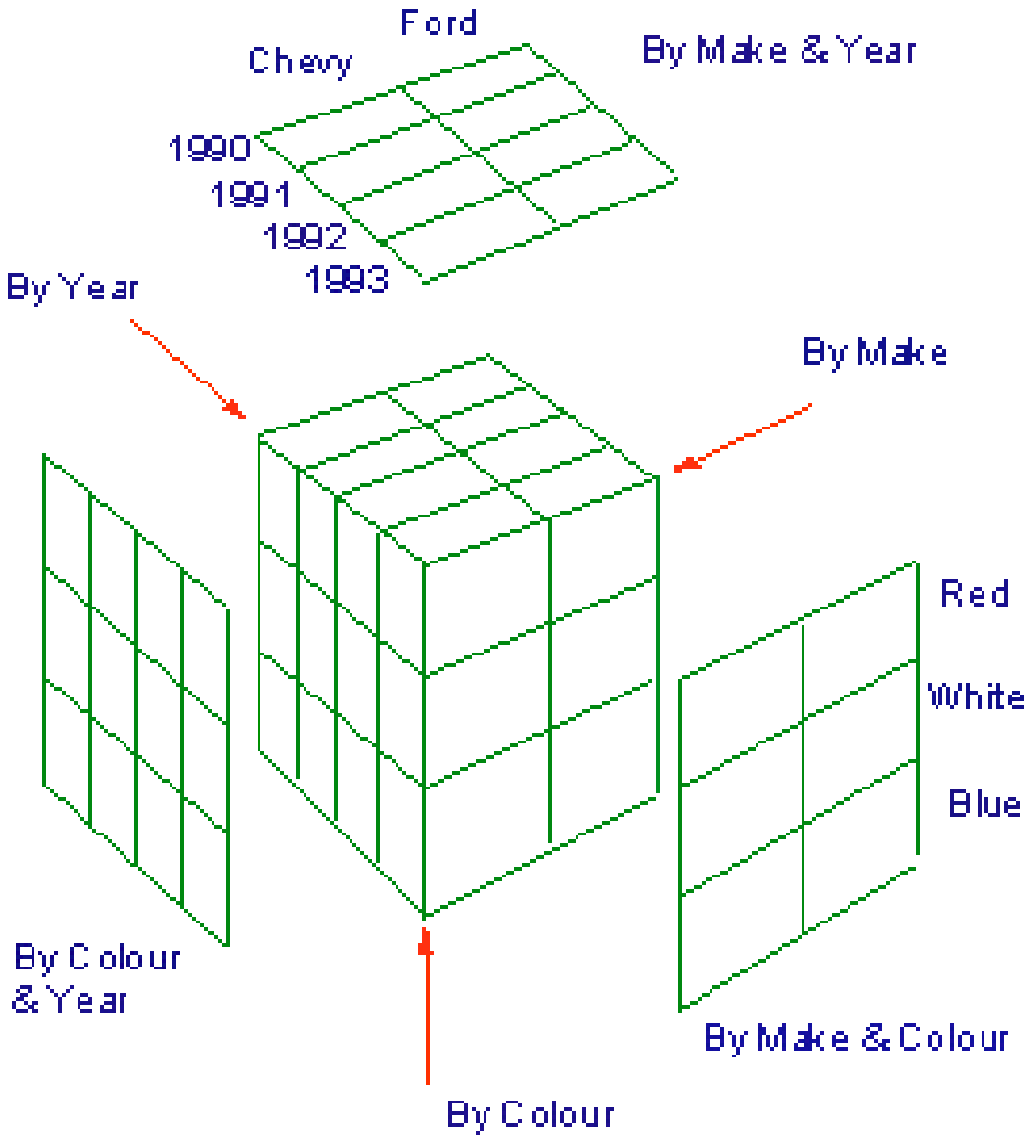


Figure 3: The Data Cube

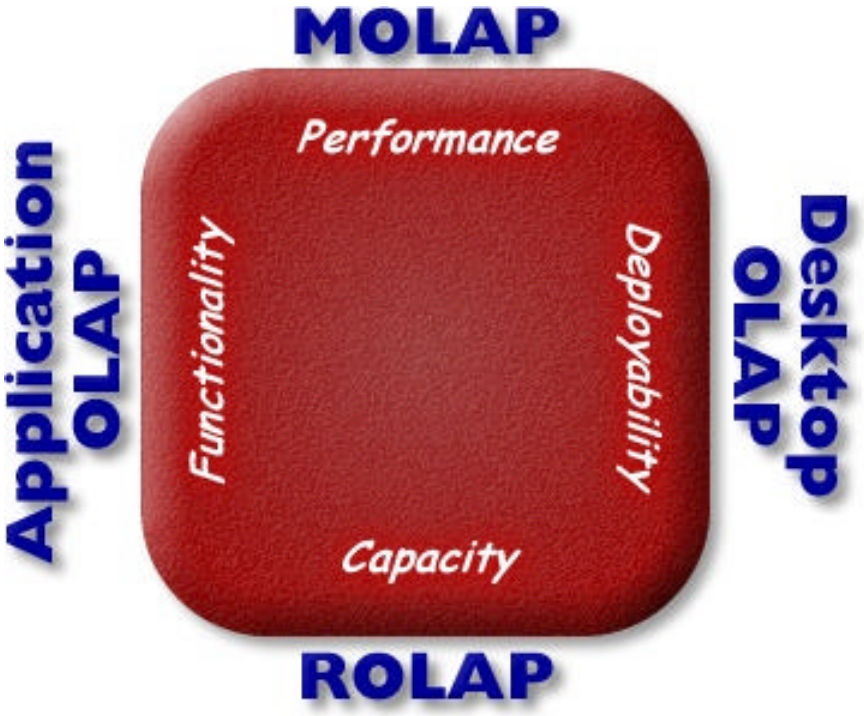
Aggregation Strategies

- Pre-calc
- Calc-on-the-fly
- Hybrid

We refer to the dimension to be aggregated as the *measure* attribute, while the remaining dimensions are known as the *feature* attributes.

Market

Market Segments



Leaders

At present, the list of major "players" includes companies such as Hyperion, Cognos, Microsoft, Oracle, MicroStrategy and Business Objects. Microsoft, in particular, with the introduction of its low cost SQL Server 7.0 and Analysis Services 2000, has become the fastest growing vendor and will likely challenge Hyperion as the market leader within the next two or three years. Conversely, many of the original powerhouses, such as Pilot, Gentia, and Informix, have succumbed to competitive pressure and have all but vanished.

Architectures & Vendors

Multidimensional processing options	Multidimensional data storage options: RDBMS	Multidimensional database server	Client files
Multi-pass SQL	*1* Cartesis Magnitude MicroStrategy		
Multidimensional server engine	*2* *Crystal Holos (ROLAP mode)* IBM DB2 OLAP Server CA EUREKA:Strategy Longview Khalix Informix MetaCube Speedware Media/MR Microsoft Analysis Services Oracle Express (ROLAP mode) *Pilot Analysis Server* Sagent Applix iTM1 WhiteLight	*4* SAS CFO Vision Crystal Holos Comshare Decision Hyperion Essbase Oracle Express Gentia Speedware Media/M Microsoft Analysis Services PowerPlay Enterprise Server Pilot Analysis Server Applix iTM1	
Client multidimensional engine	*3* Oracle Discoverer Informix MetaCube	*5* Comshare FDC Dimensional Insight Hyperion Enterprise Hyperion Pillar PwC CLIME	*6* Brio.Enterprise BusinessObjects Cognos PowerPlay Personal Express iTM1 Perspectives

- Relational OLAP (ROLAP) products are in squares ***1***, ***2*** and ***3***
- MDB (also known as MOLAP) products are in squares ***4*** and ***5***
- Desktop OLAP products are in square ***6***
- Hybrid OLAP products are those that are in **both** squares ***2*** and ***4*** (shown in *italics*)

Best Choices

Each of the options has its own strengths and weaknesses, and there is no single optimum choice. It is perfectly reasonable for sites to use products from more than one of the squares, and even more than one from a single square if they are specialized products used for different applications. As might be expected, the squares containing the most products are also the most widely used architectures, and vice versa. The choice of architecture does affect the performance, capacity, functionality and particularly the scalability of an OLAP solution.

Demonstration

- **Fast**. Vendors must be able to efficiently trade off pre-calculation costs and storage requirements with real-time query response. Studies have shown that users are likely to abort queries that take longer than thirty seconds to complete.
- **Analysis**. Tools should not only provide the five fundamental operations but extras such as times series analysis, currency translation, and data mining capabilities.
- **Shared**. Security and concurrency control should be available when required. It must be noted however that most OLAP systems assume that user-level updates will not be necessary.
- **Multidimensional**. This is the key FASMI requirement. Whether implemented with OLAP or MOLAP, the user must see the data in subject-oriented hierarchies.
- **Information**. Applications must be able to handle vast amounts of data. Again, regardless of the server model that is used, good OLAP applications may have to support data cubes that scale to the terabyte range.

Applications

Examples

- Global Planning for Sales & Marketing
- Portfolio Analysis
- Cost/Benefit Analysis for Infrastructure Investments
- Balanced Scorecard
- Performance Measurement

Sources

- OLAP Report - http://www.olapreport.com/index.htm
- Hyperion Solutions – http://www.hyperion.com
- Microsoft - http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnsql2k/html/olapunisys.asp?frame=true
- OLAP Solutions. Erik Thomsen.
- Panda Project - http://www.cs.dal.ca/~panda/overview.html

Differences between OLTP and DWs?

- Users and system orientation
- Data contents
- Database design
- View
- Access patterns