UNIT II BUSINESS ANALYSIS

Reporting and Query tools and Applications – Tool Categories – The Need for Applications – Cognos Impromptu – Online Analytical Processing (OLAP) – Need – Multidimensional Data Model – OLAP Guidelines – Multidimensional versus Multirelational OLAP – Categories of Tools – OLAP Tools and the Internet.

Reporting Query tools and Applications

The data warehouse is accessed using an end-user query and reporting tool from Business Objects. Business Objects provides several tools to securely access the data warehouse or personal data files with a point-and-click interface including the following:

- **BusinessObjects** (Reporter and Explorer) ? a Microsoft Windows based query and reporting tool.
- **InfoView** - a web based tool, that allows reports to be refreshed on demand (but cannot create new reports).
- **InfoBurst** - a web based server tool that allows reports to be refreshed, scheduled and distributed. It can be used to distribute reports and data to users or servers in various formats (e.g. Text, Excel, PDF, HTML, etc.). For more information, see the documentation below:
    - InfoBurst Usage Notes (PDF)
    - InfoBurst User Guide (PDF)
- **Data Warehouse List Upload** - a web based tool that allows lists of data to be uploaded into the data warehouse for use as input to queries. For more information, see the documentation below:

    o Data Warehouse List Upload Instructions (PDF)

WSU has negotiated a contract with Business Objects for purchasing these tools at a discount. View BusObj Rates.

Selecting your Query Tool:

a. The query tools discussed in the next several slides represent the most commonly used query tools at Penn State.
b. A Data Warehouse user is free to select any query tool, and is not limited to the ones mentioned.
c. What is a ―Query Tool‖?
d. A query tool is a software package that allows you to connect to the data warehouse from your PC and develop queries and reports.

There are many query tools to choose from. Below is a listing of what is currently being used on the PC:

1. Microsoft Access

2. Microsoft Excel

3. Cognos Impromptu

Data Warehousing Tools and Technologies

a) Building a data warehouse is a complex task because there is no vendor that provides an end-to-end_ set of tools.

b) Necessitates that a data warehouse is built using multiple products from different vendors.

c) Ensuring that these products work well together and are fully integrated is a major challenge.

Cognos impromptu Query Tabs: Data

- Identify what to query
- Click and drag

Sort

- Hierarchy presentation
- Ascending or descending order

Group

- Summarized data by group order

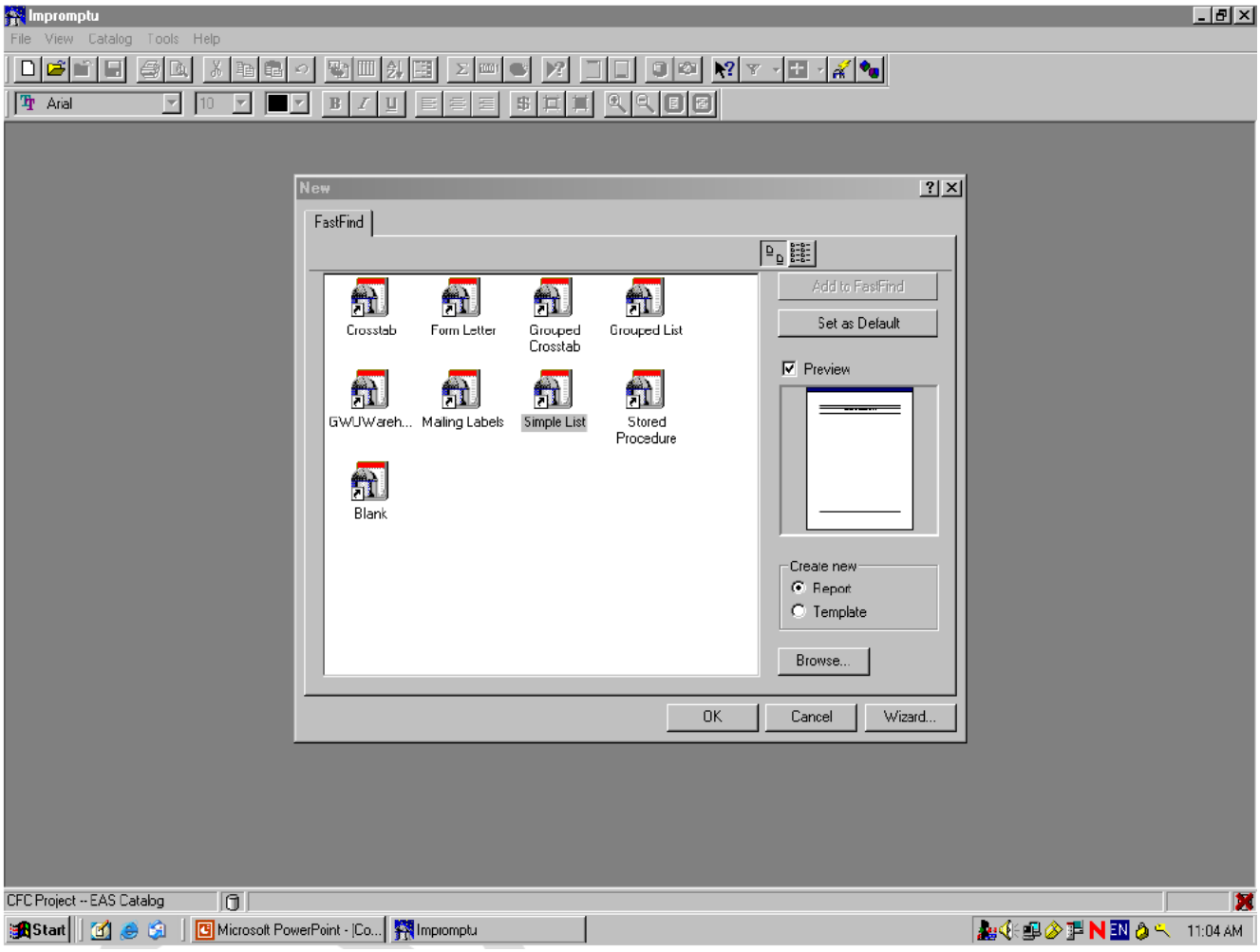
Filter

- Defines criteria
- Specifies query range

Administrative

- Access
- Profile
- Client/Server

Generating reports:



Edit features on the toolbar allowed changes to report data after the query has been completed

- Modify existing data
- Format numerical and date fields
- Perform calculations
- Group data
- Sort columns.

General Ledger System Data:

- **Data Elements**

  Table Format

  - Balances – Summary information
  - Lines – Journal entry detail

  Numeric

  - Detail and summary
  - Include calculations

  Descriptive

  - Accounting string segment values

Cognous Impromptu What is impromptu? Impromptu is an interactive database reporting tool. It allows Power Users to query data without programming knowledge. When using the Impromptu tool, no data is written or changed in the database. It is only capable of reading the data. Impromptu's main features includes,

- Interactive reporting capability
- Enterprise-wide scalability
- Superior user interface
- Fastest time to result
- Lowest cost of ownership

Catalogs Impromptu stores metadata in subject related folders. This metadata is what will be used to develop a query for a report. The metadata set is stored in a file called a _catalog_. The catalog does not contain any data. It just contains information about connecting to the database and the fields that will be accessible for reports. A catalog contains:

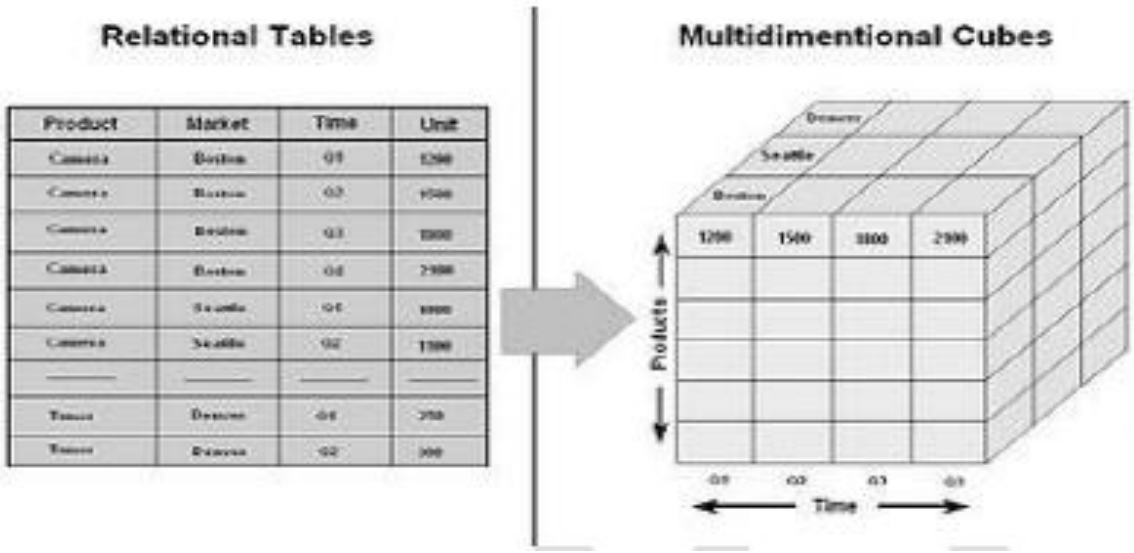
- Folders—meaningful groups of information representing columns from one or more tables
- Columns—individual data elements that can appear in one or more folders
- Calculations—expressions used to compute required values from existing data
- Conditions—used to filter information so that only a certain type of information is displayed
- Prompts—pre-defined selection criteria prompts that users can include in reports they create
- Other components, such as metadata, a logical database name, join information, and user classes

You can use catalogs to

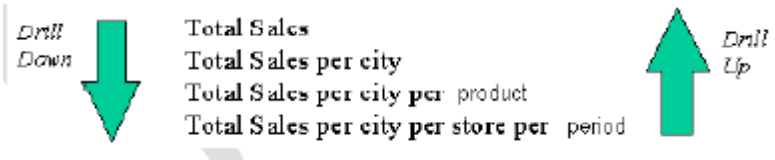
- view, run, and print reports
- export reports to other applications
- disconnect from and connect to the database
- create reports
- change the contents of the catalog
- add user classes

Online Analytical Processing (OLAP), OLAP Need, Multidimensional Data Model: The Multidimensional data Model

The multidimensional data model is an integral part of On-Line Analytical Processing, or OLAP. Because OLAP is on-line, it must provide answers quickly; analysts pose iterative queries during interactive sessions, not in batch jobs that run overnight. And because OLAP is also analytic, the queries are complex. The multidimensional data model is designed to solve complex queries in real time. Multidimensional data model is to view it as a cube. The cable at the left contains detailed sales data by product, market and time. The cube on the right associates sales number (unit sold) with dimensions-product type, market and time with the unit variables organized as cell in an array. This cube can be expended to include another array-price-which can be associates with all or only some dimensions. As number of dimensions increases number of cubes cell increase exponentially. Dimensions are hierarchical in nature i.e. time dimension may contain hierarchies for years, quarters, months, weak and day. GEOGRAPHY may contain country, state, city etc.



In this cube we can observe, that each side of the cube represents one of the elements of the question. The x-axis represents the time, the y-axis represents the products and the z-axis represents different centers. The cells of in the cube represents the number of product sold or can represent the price of the items.



This Figure also gives a different understanding to the drilling down operations. The relations defined must not be directly related, they related directly. The size of the dimension increase, the size of the cube will also increase exponentially. The time response of the cube depends on the size of the cube.

Operations in Multidimensional Data Model:

• Aggregation (roll-up)

– dimension reduction: e.g., total sales by city

– summarization over aggregate hierarchy:

e.g., total sales by city and year -> total sales by region and by year

• Selection (slice) defines a subcube – e.g., sales where city = Palo Alto and date = 1/15/96

• Navigation to detailed data (drill-down) – e.g., (sales - expense) by city, top 3% of cities by average income

• Visualization Operations (e.g., Pivot or dice)

OLAP

OLAP stands for Online Analytical Processing. It uses database tables (fact and dimension tables) to enable multidimensional viewing, analysis and querying of large amounts of data. E.g. OLAP technology could provide management with fast answers to complex queries on their operational data or enable them to analyze their company's historical data for trends and patterns. Online Analytical Processing (OLAP) applications and tools are those that are designed to ask ―complex queries of large multidimensional collections of data‖. Due to that OLAP is accompanied with data warehousing.

Need

The key driver of OLAP is the multidimensional nature of the business problem. These problems are characterized by retrieving a very large number of records that can reach gigabytes and terabytes and summarizing this data into a form information that can by used by business analysts. One of the limitations that SQL has, it cannot represent these complex problems. A query will be translated in to several SQL statements. These SQL statements will involve multiple joins, intermediate tables, sorting, aggregations and a huge temporary memory to store these tables. These procedures required a lot of computation which will require a long time in computing. The second limitation of SQL is its inability to use mathematical models in these SQL statements. If an analyst, could create these complex statements using SQL statements, still there will be a large number of computation and huge memory needed. Therefore the use of OLAP is preferable to solve this kind of problem.

CATEGORIES OF OLAP TOOLS

MOLAP This is the more traditional way of OLAP analysis. In MOLAP, data is stored in a multidimensional cube. The storage is not in the relational database, but in proprietary formats. That is, data stored in array-based structures.

Advantages:

- Excellent performance: MOLAP cubes are built for fast data retrieval, and are optimal for slicing and dicing operations.
- Can perform complex calculations: All calculations have been pre-generated when the cube is created. Hence, complex calculations are not only doable, but they return quickly.

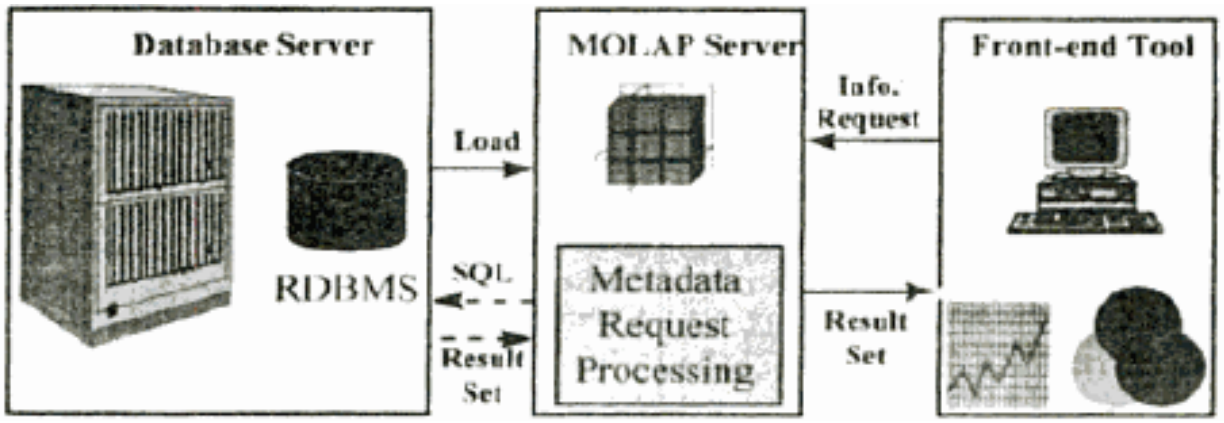
Disadvantages:

- Limited in the amount of data it can handle: Because all calculations are performed when the cube is built, it is not possible to include a large amount of data in the cube itself. This is not to say that the data in the cube cannot be derived from a

large amount of data. Indeed, this is possible. But in this case, only summary-level information will be included in the cube itself.

- Requires additional investment: Cube technology are often proprietary and do not already exist in the organization. Therefore, to adopt MOLAP technology, chances are additional investments in human and capital resources are needed.

Examples: Hyperion Essbase, Fusion (Information Builders)



ROLP

This methodology relies on manipulating the data stored in the relational database to give the appearance of traditional OLAP's slicing and dicing functionality. In essence, each action of slicing and dicing is equivalent to adding a "WHERE" clause in the SQL statement. Data stored in relational tables

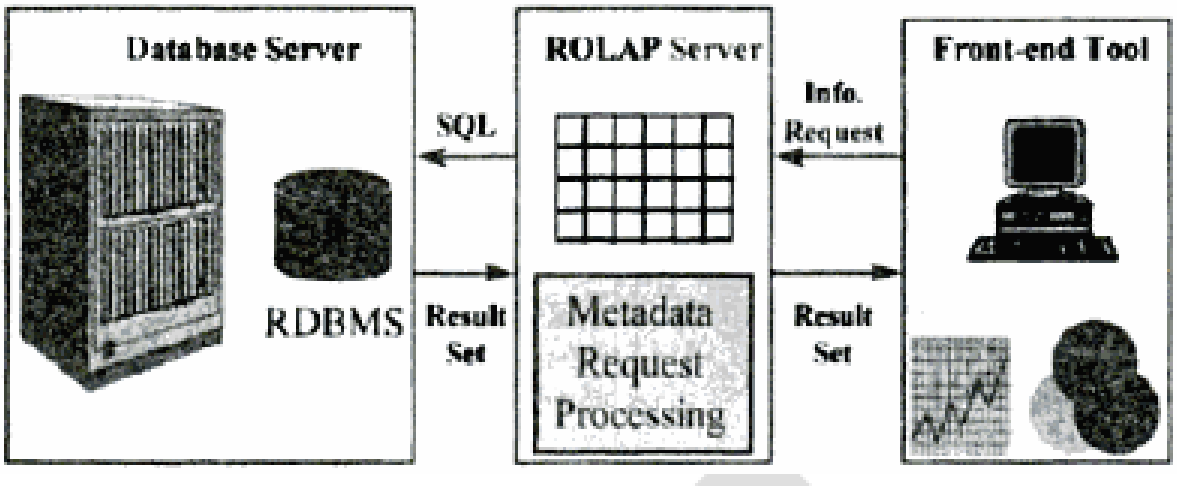
Advantages:

- Can handle large amounts of data: The data size limitation of ROLAP technology is the limitation on data size of the underlying relational database. In other words, ROLAP itself places no limitation on data amount.
- Can leverage functionalities inherent in the relational database: Often, relational database already comes with a host of functionalities. ROLAP technologies, since they sit on top of the relational database, can therefore leverage these functionalities.

Disadvantages:

- Performance can be slow: Because each ROLAP report is essentially a SQL query (or multiple SQL queries) in the relational database, the query time can be long if the underlying data size is large.
- Limited by SQL functionalities: Because ROLAP technology mainly relies on generating SQL statements to query the relational database, and SQL statements do not fit all needs (for example, it is difficult to perform complex calculations using SQL), ROLAP technologies are therefore traditionally limited by what SQL can do. ROLAP vendors have mitigated this risk by building into the tool out-of-the-box complex functions as well as the ability to allow users to define their own functions.

Examples: Microstrategy Intelligence Server, MetaCube (Informix/IBM)



HOLAP

(MQE: Managed Query Environment)

HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP. For summary-type information, HOLAP leverages cube technology for faster performance. It stores only the indexes and aggregations in the multidimensional form while the rest of the data is stored in the relational database.

Examples: PowerPlay (Cognos), Brio, Microsoft Analysis Services, Oracle Advanced Analytic Services.

OLAP Guidelines:

Dr. E.F. Codd the ―father‖ of the relational model, created a list of rules to deal with the OLAP systems. Users should priorities these rules according to their needs to match their business requirements (reference 3).

These rules are:

1) Multidimensional conceptual view: The OLAP should provide an appropriate multidimensional Business model that suits the Business problems and Requirements.

2) Transparency: The OLAP tool should provide transparency to the input data for the users.

3) Accessibility: The OLAP tool should only access the data required only to the analysis needed.

4) Consistent reporting performance: The Size of the database should not affect in any way the performance.

5) Client/server architecture: The OLAP tool should use the client server architecture to ensure better performance and flexibility.

6) Generic dimensionality: Data entered should be equivalent to the structure and operation requirements.

7) Dynamic sparse matrix handling: The OLAP too should be able to manage the sparse matrix and so maintain the level of performance.

8) Multi-user support: The OLAP should allow several users working concurrently to work together.

9) Unrestricted cross-dimensional operations: The OLAP tool should be able to perform operations across the dimensions of the cube.

10) Intuitive data manipulation. —Consolidation path re-orientation, drilling down across columns or rows, zooming out, and other manipulation inherent in the consolidation path outlines should be accomplished via direct action upon the cells of the analytical model, and should neither require the use of a menu nor multiple trips across the user interface.(Reference 4)

11) Flexible reporting: It is the ability of the tool to present the rows and column in a manner suitable to be analyzed.

12) Unlimited dimensions and aggregation levels: This depends on the kind of Business, where multiple dimensions and defining hierarchies can be made.

Features of OLTP and OLAP

The major distinguishing features between OLTP and OLAP are summarized as follows.

1. Users and system orientation: An OLTP system is customer-oriented and is used for transaction and query processing by clerks, clients, and information technology professionals. An OLAP system is market-oriented and is used for data analysis by knowledge workers, including managers, executives, and analysts.

2. Data contents: An OLTP system manages current data that, typically, are too detailed to be easily used for decision making. An OLAP system manages large amounts of historical data, provides facilities for summarization and aggregation, and stores and manages information at different levels of granularity. These features make the data easier for use in informed decision making.

3. Database design: An OLTP system usually adopts an entity-relationship (ER) data model and an application oriented database design. An OLAP system typically adopts either a star or snowflake model and a subject-oriented database design.

4. View: An OLTP system focuses mainly on the current data within an enterprise or department, without referring to historical data or data in different organizations. In contrast, an OLAP system often spans multiple versions of a database schema. OLAP

systems also deal with information that originates from different organizations, integrating information from many data stores. Because of their huge volume, OLAP data are stored on multiple storage media.

5. Access patterns: The access patterns of an OLTP system consist mainly of short, atomic transactions. Such a system requires concurrency control and recovery mechanisms. However, accesses to OLAP systems are mostly read-only operations although many could be complex queries.

Comparison between OLTP and OLAP systems.

Feature	OLTP	OLAP
Characteristic	operational processing	informational processing
Orientation	transaction	analysis
User	clerk, DBA, database professional	knowledge worker (e.g., manager, executive, analyst)
Function	day-to-day operations	long term informational requirements, decision support
DB design	E-R based, application-oriented	star/snowflake, subject-oriented
Data	current; guaranteed up-to-date	historical; accuracy maintained over time
Summarization	primitive, highly detailed	summarized, consolidated
View	detailed, flat relational	summarized, multidimensional
Unit of work	short, simple transaction	complex query
Access	read/write	mostly read
Focus	data in	information out
Operations	index/hash on primary key	lots of scans
# of records accessed	tens	millions
# of users	thousands	hundreds
DB size	100 MB to GB	100 GB to TB
Priority	high performance, high availability	high flexibility, end-user autonomy
Metric	transaction throughput	query throughput, response time

Multidimensional versus Multi relational OLAP, Categories of Tools

Representation of Multi-Dimensional Data:

- OLAP database servers use multi-dimensional structures to store data and relationships between data.
- Multi-dimensional structures are best-visualized as cubes of data, and cubes within cubes of data. Each side of a cube is a dimension.

Multi-dimensional OLAP supports common analytical operations, such as:

- Consolidation: involves the aggregation of data such as _roll-ups_or complex expressions involving interrelated data. Foe example, branch offices can be rolled up to cities and rolled up to countries.
- Drill-Down: is the reverse of consolidation and involves displaying the detailed data that comprises the consolidated data.
- Slicing and dicing: refers to the ability to look at the data from different viewpoints. Slicing and dicing is often performed along a time axis in order to analyze trends and find patterns.

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
.....	.....	.....
.....	.....	.....

(a)

City

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

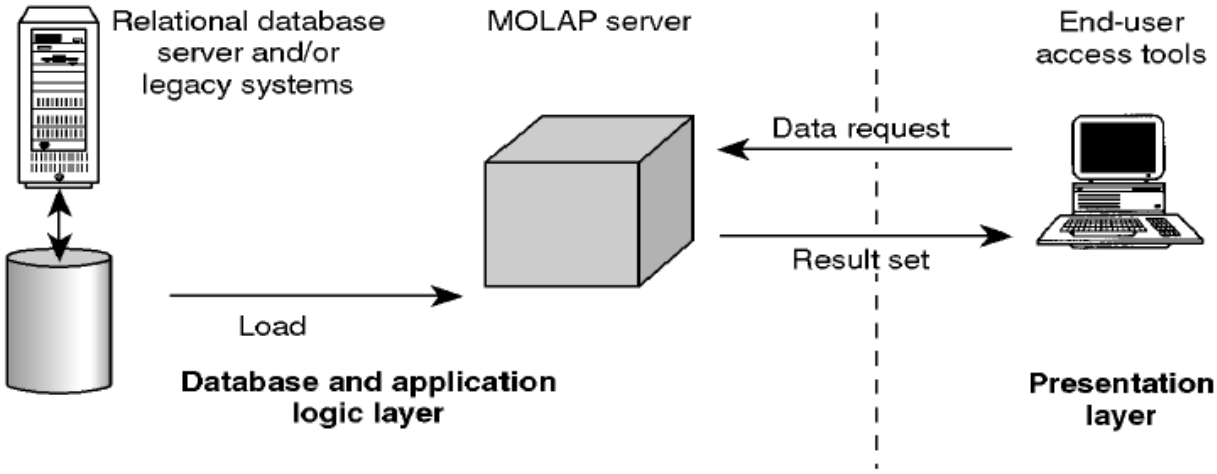
(b)

Relational OLAP (ROLAP)

- ROLAP is the fastest-growing type of OLAP tools.
- ROLAP supports RDBMS products through the use of a metadata layer, thus avoiding the requirement to create a static multi-dimensional data structure.
- This facilitates the creation of multiple multi-dimensional views of the two-dimensional relation.
- To improve performance, some ROLAP products have enhanced SQL engines to support the complexity of multi-dimensional analysis, while others recommend, or require, the use of highly denormalized database designs such as the star schema.
- The development issues associated with ROLAP technology:
- Performance problems associated with the processing of complex queries that require multiple passes through the relational data.
- Development of middleware to facilitate the development of multi-dimensional applications. Development of an option to create persistent multi-dimensional structures, together with facilities o assist in the administration of these structures.

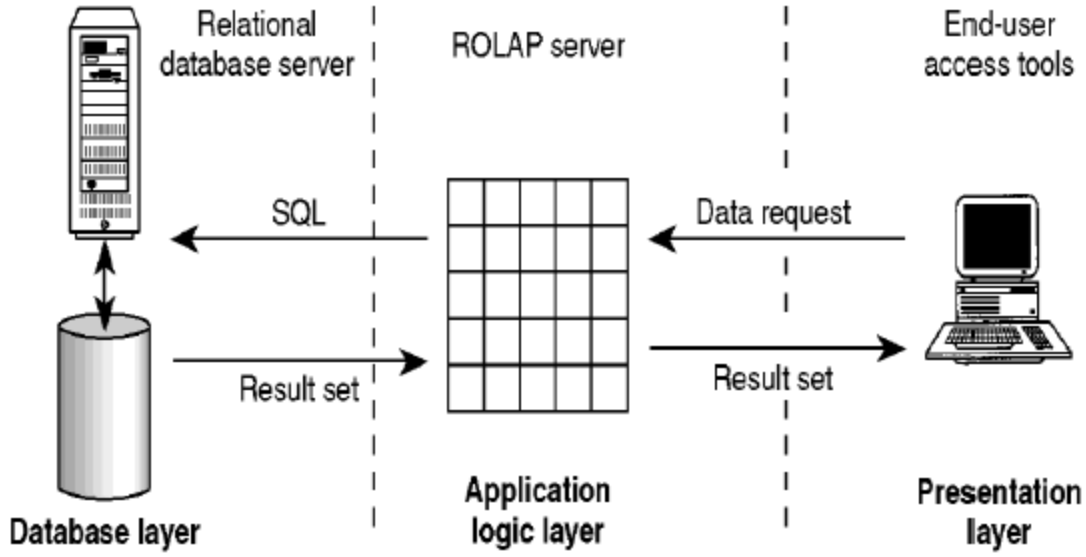
OLAP Tools and the Internet

Categorization of OLAP Tools OLAP tools are designed to manipulate and control multi-dimensional databases and help the sophisticated user to analyze the data using clear multidimensional complex views. Their typical applications include product performance and profitability, effectiveness of a sales program or a marketing campaign, sales forecasting, and capacity planning.



Describes The Relation of the MOLAP with the server and end user.

ROLAP



The mainly comprehensive premises in computing have been the internet and data warehousing thus the integration of these two giant technologies is a necessity. The advantages of using the Web for access are inevitable.These advantages are:

1. The internet provides connectivity between countries acting as a free resource.

2. The web eases administrative tasks of managing scattered locations.

3. The Web allows users to store and manage data and applications on servers that can be managed, maintained and updated centrally.

These reasons indicate the importance of the Web in data storage and manipulation. The Web-enabled data access has many significant features, such as:

- **The first**
- **The second**
- **The emerging third**
- **HTML publishing**
- **Helper applications**
- **Plug-ins**
- **Server-centric components**
- **Java and active-x applications**

Products for OLAP

Microsoft Analysis Services (previously called OLAP Services, part of SQL Server), IBM's DB2 OLAP Server, SAP BW and products from Brio, Business Objects, Cognos, Micro Strategy and others.

Companies using OLAP

MIS AG Overview

MIS AG is the leading European provider of business intelligence solutions and services, providing development, implementation, and service of systems for budgeting, reporting, consolidation, and analysis.

Poet Overview

With FastObjects™, German Poet Software GmbH (Poet) provides developers with a flexible Object-oriented Database Management System (ODBMS) solution optimized for managing complexity in high-performance applications using Java technology, C++ and .NET.