Online Analytical Processing OLAP

- OLAP OVERVIEW
- What is OLAP?
- Designing OLAP Solutions

OLAP OVERVIEW,

- The data warehouse offloads data from a multitude of sources.
- The cleaned, validated and loaded data is voluminous and daunting.
- This data needs to be organized, categorized and arranged in meaningful order for analytical purposes.
- OLAP solutions are specifically designed to cater to this need.

OLAP OVERVIEW

- OLAP, are being used aggressively by organizations to discover valuable business trends from data marts and data warehouses.
- OLAP provides a historical view of data,
- although useful when used by itself,
- OLAP analysis becomes truly powerful when combined with predictive analysis like Data Mining.

OLAP OVERVIEW

- OLAP (or *Online Analytical Processing*) has been growing in popularity due to the increase in data volumes and the recognition of the business value of analytics.
- Until the mid-nineties, performing OLAP analysis was an extremely costly process mainly restricted to larger organizations.

OLAP OVERVIEW

- The major OLAP vendor are Hyperion, Cognos, Business Objects, MicroStrategy.
- The cost per seat were in the range of $1500 to $5000 per annum.
- The setting up of the environment to perform OLAP analysis would also require substantial investments in time and monetary resources.

OLAP OVERVIEW

- This has changed as the major database vendor have started to incorporate OLAP modules within their database offering –
- Microsoft SQL Server 2000 with Analysis Services,
- Oracle with Express and Darwin, and
- IBM with DB2

What is OLAP?

- OLAP allows business users to slice and dice data at will.
- Normally data in an organization is distributed in multiple data sources and are incompatible with each other.
- A retail example: Point-of-sales data and sales made via call-center or the Web are stored in different location and formats.

What is OLAP?

- It would a time consuming process for an executive to obtain OLAP reports such as –
- What are the most popular products purchased by customers between the ages 15 to 30?
- Part of the OLAP implementation process involves extracting data from the various data repositories and making them compatible.

What is OLAP?

- Making data compatible involves ensuring that the meaning of the data in one repository matches all other repositories.
- An example of incompatible data: Customer ages can be stored as birth date for purchases made over the web and stored as age categories (i.e. between 15 and 30) for in store sales.
- It is not always necessary to create a data warehouse for OLAP analysis.
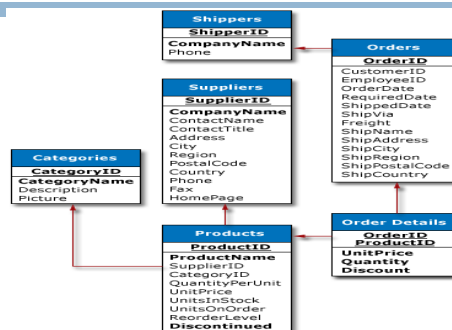
What is OLAP?

- Data stored by operational systems, such as point-of-sales, are in types of databases called OLTPs.
- OLTP, *Online Transaction Process*, databases do not have any difference from a structural perspective from any other databases.
- The main difference, and only, difference is the way in which data is stored.

What is OLAP?

- Examples of OLTPs can include ERP, CRM, SCM, Point-of-Sale applications, Call Center.
- OLTPs are designed for optimal transaction speed.
- When a consumer makes a purchase online, they expect the transactions to occur instantaneously.
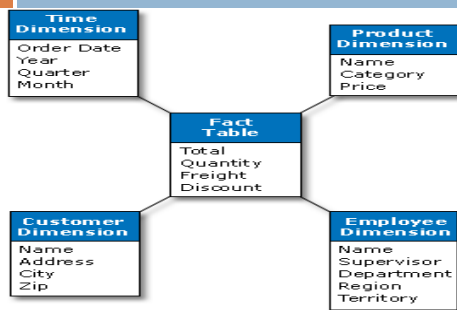
Data Model for OLTP

What is OLAP?

- Data are not typically stored for an extended period on OLTPs for storage cost and transaction speed reasons.
- OLAPs have a different mandate from OLTPs.

What is OLAP?

- OLAPs are designed to give an overview analysis of what happened.
- Hence the data storage (i.e. data modeling) has to be set up differently.
- The most common method is called the star design.

Star Data Model for OLAP

Time Dimension
Order Date
Year
Quarter
Month

Product Dimension
Name
Category
Price

Fact Table
Total
Quantity
Freight
Discount

Customer Dimension
Name
Address
City
Zip

Employee Dimension
Name
Supervisor
Department
Region
Territory

What is OLAP?

- The central table in an OLAP start data model is called the fact table.
- The surrounding tables are called the dimensions.
- Using the above data model, it is possible to build reports that answer questions such as:
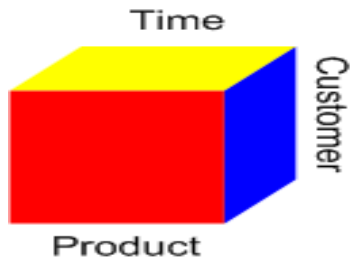
What is OLAP?

- The supervisor that gave the most discounts.
- The quantity shipped on a particular date, month, year or quarter.
- In which zip code did product A sell the most.
- To obtain answers, such as the ones above, from a data model OLAP *cubes* are created.

What is OLAP?

- OLAP cubes are not strictly cuboids - it is the name given to the process of linking data from the different dimensions.
- The cubes can be developed along business units such as sales or marketing.
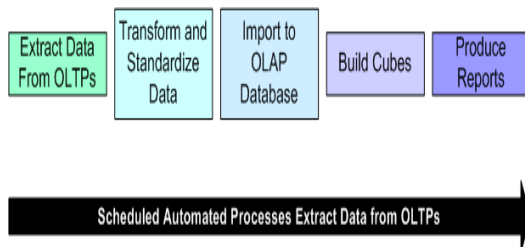- Or a giant cube can be formed with all the dimensions.

OLAP Cube with Time, Customer and Product Dimensions

What is OLAP?

- OLAP can be a valuable and rewarding business tool.
- Aside from producing reports, OLAP analysis can aid an organization evaluate balanced scorecard targets.

Steps in the OLAP Creation Process

Selecting an OLAP Application

- While selecting an OLAP tool that can fit an organization's need requires a closer inspection of its requirements.
- We will give an overview of some of the generic requirements that most organizations will have to consider in their selection process

Platform Choice

- The first step in narrowing the list of vendors is the strength of the OLAP tool on the platform that your organization will be using.
- Some vendors tend to be only Microsoft centric or have products that are better supported on a particular platform.

Vendor History

- There are many new OLAP vendors that have cropped up since the mid-nineties.
- Some of the newer companies have rewarding and rich OLAP tools that can be cost effective.
- Investigate how many installs the vendor has made and, if possible, interview some of the vendor's clients.

OLAP Cube Size and Transaction Speed

- There are products that are optimized to your organization's needs.
- Fine tune your vendor list to your operational needs.
- How large is your cube?

OLAP Cube Size and Transaction Speed

- What would be a tolerable usage of storage space and transaction speed?
- How will the OLAP cube be created - i.e. HOLAP, MOLAP, ROLAP?

Graphical User Interface (GUI)

- Will you prefer a thin client or desktop interface to interact with the OLAP cube?
- What are the trade-offs?
- List potential security concerns.

Consulting

- How much consulting would be require to install the product?
- If changes are needed after installation, will this require the need of external consultants?

Integration to Database

- How tightly integrated is the OLAP tool to your database?
- Can you perform SQL queries within the OLAP tool?

Integration to other Analysis Tools

- Does your organization have or plan to have other analysis tools such as data mining, reporting and data visualization?
- How will the OLAP application integrate with the other analysis tools?

Price

- It is often challenging to gauge the additional hidden cost from the actual purchase price.
- Some of the points above can help make a good financial forecast.
- Discuss with other users in your industry vertical or hire an independent consultant who is familiar with the process.

Designing OLAP Solutions

- MOLAP, ROLAP, HOLAP and other acronyms!
- After working out the business information of interest to extract from a data warehouse, and designing the OLAP cube with the relevant dimensions and measures,
- the next step is to choose the type of storage for the OLAP cube.

Designing OLAP Solutions

- The correct design decisions made at the inception of the data warehousing project will lower implementation and ongoing maintenance costs.
- And provide quick and easy to access analytics information to business users.

Designing OLAP Solutions

- The designer can store the OLAP data in the data warehouse itself or in a separate repository optimized for cube storage.
- Depending on the size of your data warehousing solution and the expected connection load, the ideal storage method can be chosen.

Designing OLAP Solutions

- For larger implementations with frequent queries, a separate optimized OLAP repository will be better suited.
- Queries made against an OLAP cube use large amounts of computer resources.

Designing OLAP Solutions

- In a perfect world, we can store all measures against each level of every combination of dimensions of the OLAP cube.
- Hence on each occasion a query is made, it can obtain the value from the pre-calculated aggregations.
- This can save time as the source data will not have to be queried.

Designing OLAP Solutions

- However this is not practical in most real-world applications.
- The number of aggregations increases exponentially with the number of OLAP levels and dimensions:

Number of OLAP Aggregations Increases Exponentially

- $a = l^d$
- a - Aggregations, l - Levels, d - Dimensions

Dimensions	Levels			
	2	3	4	5
2	4	9	16	25
3	8	27	64	125
4	16	81	256	625
4	32	243	1 024	3 125
5	64	729	4 096	15 625
6	128	2 187	16 384	78 125

Designing OLAP Solutions

- Many cubes used in practice have over ten dimensions, and can hold on average of four or more levels.
- These would require over 1 048 576 aggregations.
- To pre-calculate a million aggregations or more will require an inordinate amount of time and computer resources.
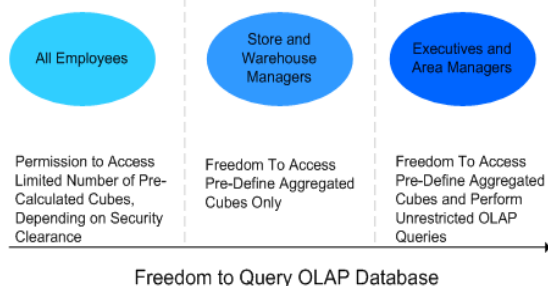
Designing OLAP Solutions

- Data Warehouse administrators often pre-calculate the more popular aggregations to reduce demand on the database.
- Companies also typically will restrict the use of OLAP to a select set of users.

Designing OLAP Solutions

- In an implementation DWreview recently made with a major retail chain, OLAP aggregations were pre-calculated for popular queries.
- These were accessible by a broad section of users.
- Users were given freedom to conducting OLAP queries depending on their levels of permission

Freedom to Query OLAP Database

All Employees	Store and Warehouse Managers	Executives and Area Managers
Permission to Access Limited Number of Pre-Calculated Cubes, Depending on Security Clearance	Freedom To Access Pre-Define Aggregated Cubes Only	Freedom To Access Pre-Define Aggregated Cubes and Perform Unrestricted OLAP Queries

Freedom to Query OLAP Database

Freedom to Query OLAP Database

- Using this design strategy saved considerable amount of hardware and software cost.
- Together with sound data management policies, the company saved substantially in the data warehousing implementation using this design strategy, without loosing functionality and reliability.

Designing OLAP Solutions

- OLAP storage is one of the critical choices to be made when designing the solution.
- OLAP storage comes in three forms:

MOLAP -

- Multidimensional OLAP. In MOLAP, both the source data and the aggregations are stored in a multidimensional format.
- MOLAP is the fastest option for data retrieval, but requires the most disk space.
- Disk space is less of a concern these days with lowering storage and processing cost.

ROLAP - Relational OLAP

- .All data, including the aggregations are stored within the source relational database.
- This will be a concern for larger data warehousing implementations which have higher usage needs.
- ROLAP is the slowest for data retrieval.
- Whether an aggregation exists or not, a ROLAP database must access the data warehouse itself.
- ROLAP is best suited for smaller data warehousing implementations.

HOLAP - Hybrid OLAP.

- HOLAP is a combination of both the above storage methodologies.
- HOLAP databases store the aggregations that exist within a multidimensional structure, leaving the cell-level data itself in a relational form.
- Where the data is pre aggregated, HOLAP offers the performance of MOLAP, where the data must be fetched from the tables.

HOLAP

- HOLAP is as slow as ROLAP.
- Due to shrinking hardware and processing cost, MOLAP are generally most often used.
- HOLAP is a better solution if the solution is accessing a stand-alone database.
- ROLAP are more convenient to set up when the query demands are relatively low and also on a stand-alone database.