RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-11

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

OUTLINE

- WHAT IS OLAP
- OLAP
- OLAP TOOLS DELIVER POWERFUL ANALYSIS
- THE TYPICAL ARCHITECTURE OF AN OLAP
- ORACLE OLAP TOOL
- OLAP GUIDELINES (DR. E.F. CODD RULE)
- MCQ
- REFERENCES

What is OLAP

OLAP is an acronym for On Line Analytical Processing. OLAP performs multidimensional analysis of business data and provides the capability for complex calculations, trend analysis, and sophisticated data modeling. It is quickly becoming the fundamental foundation for Intelligent Solutions including Business Performance Management, Planning, Budgeting, Forecasting, Financial Reporting, Analysis, Simulation Models, Knowledge Discovery, and Data Warehouse Reporting. OLAP enables end-users to perform ad hoc analysis of data in multiple dimensions, thereby providing the insight and understanding they need for better decision making.

Typical OLAP Applications:

- ❑ Financial modeling (budgeting, planning)
- ❑ Sales Forecasting
- ❑ Customer & Product Profitability
- ❑ Exception Reporting
- ❑ Resource allocation & Capacity Planning
- ❑ Variance Analysis
- ❑ Promotion Planning
- ❑ Market Share Analysis

OLAP

- Data is integrated by loading it into a centralized database called a Data Warehouse. The data warehouse holds data from all sources and is usually clean and ready for business use.
- However, most business users do not have direct access to the data warehouse but rather to OLAP cubes that are built over it.

• Why OLAP

- OLAP cubes have two main purposes.
- The first is to provide business users with a data model more intuitive to them than a tabular model. This model is called a Dimensional Model.
- The second purpose is to enable fast query response that is usually difficult to achieve using tabular models.

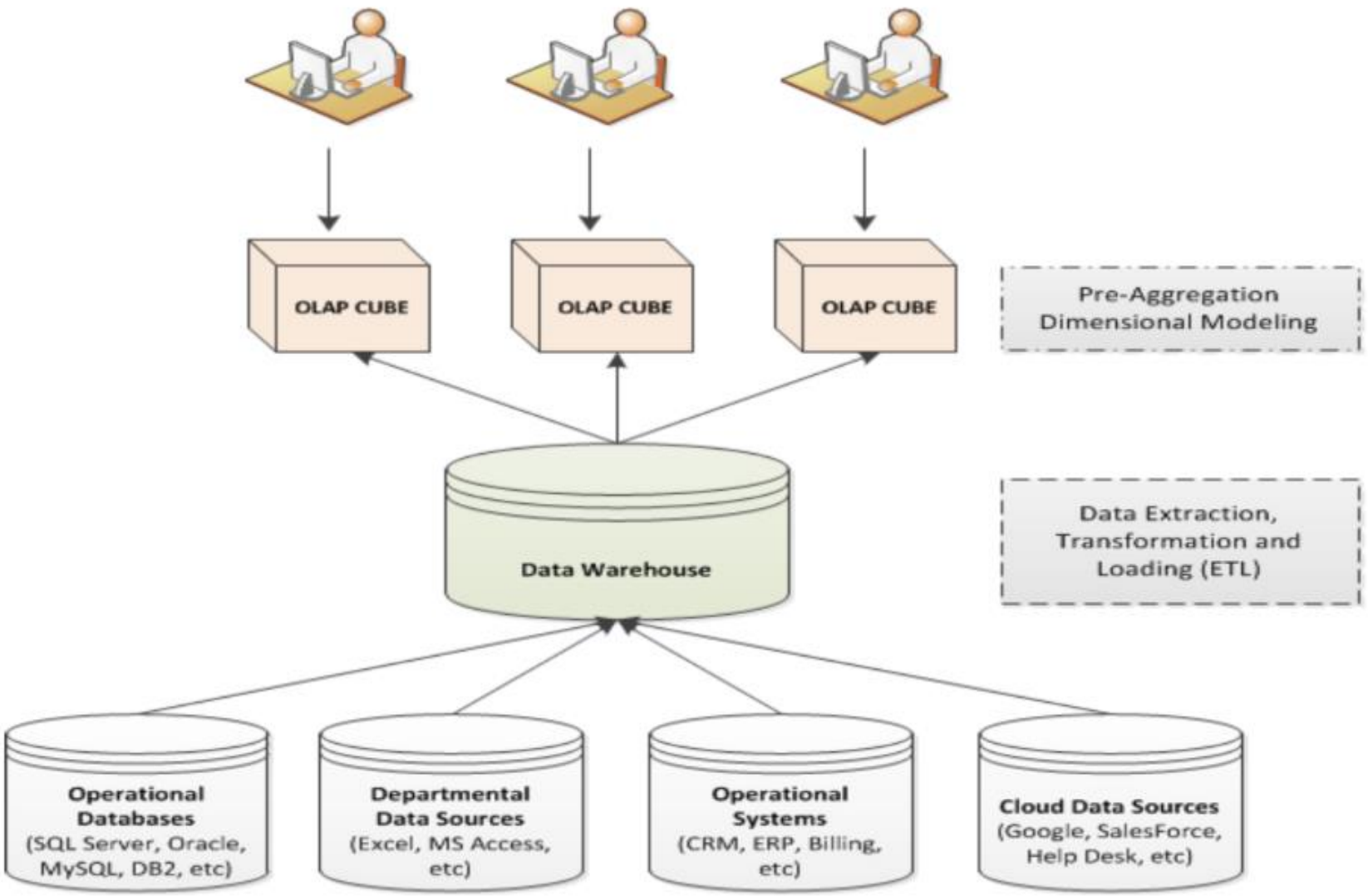
• How OLAP Works

- Fundamentally, OLAP has a very simple concept. It pre-calculates most of the queries that are typically very hard to execute over tabular databases, namely – aggregation, joining and grouping. These queries are calculated during a process that is usually called 'building' or 'processing' of the OLAP cube. This process happens overnight, and by the time end users get to work - data will have been updated.

OLAP tools deliver powerful analysis

- Companies today have come to rely on state-of-the-art OLAP tools for high-level, multidimensional analysis of corporate performance. OLAP tools allow companies to get to the bottom of complicated queries through quick consolidation of information in data-rich matrices. This kind of reporting functionality is important for all businesses, but the ability to link reporting to performance strategy is the key to enterprise success. Integrating OLAP tools with a comprehensive performance management program results in better business intelligence and more informed planning about processes, budgeting, and other critical pieces of business infrastructure.

THE TYPICAL ARCHITECTURE OF AN OLAP



Oracle OLAP Tool

- Oracle OLAP offers the following types of numeric functions:
- General numeric functions for typical mathematical processing (for example, ranking and finding logs and tangets
- Financial functions.
- Statistical functions.
- Time-series functions such as LAG and MOVINGMIN.
- Aggregation functions, such as COUNT and TOTAL.

Oracle OLAP Tool

❑OLAP Functions

❑ Oracle OLAP makes it easy to produce analytic measures, including time-series calculations, financial models, forecasts, allocations, regressions, and more. Hundreds of analytic functions can be easily combined in custom functions to solve nearly any analytic calculation requirement. Oracle OLAP cubes are represented using a star schema design: dimension views form a constellation around the cube (or fact) view. This standard representation of OLAP data makes it easy for any reporting and analysis tool or application - including sophisticated business intelligence solutions, SQL-based development tools and Microsoft Excel - to leverage the power of Oracle OLAP in a simple and productive way.

❑ With Oracle OLAP you can:

- Easily define a multidimensional model with advanced analytic calculations
- Productively deliver rich analytics to any reporting and analysis tool using simple SQL
- Transparently improve summary queries against tables using cube-based materialized views
- Combine OLAP data with any other data in your Oracle Database – including spatial, data mining, XML, documents and more
- Leverage your existing Oracle Database expertise and software investment

OLAP Guidelines (Dr. E.F. Codd Rule)

- Dr. E.F. Codd the “father” of the relational model has formulated a list of 12 guidelines and requirements as the basis for selecting OLAP systems:
- **Multidimensional conceptual view :** User-analysts would view an enterprise as being multidimensional in nature – for example, profits could be viewed by region, product, time period, or scenario (such as actual, budget, or forecast). Multi-dimensional data models enable more straightforward and intuitive manipulation of data by users, including “slicing and dicing“.
- **Transparency:** The user should not know what concrete resources are used for storage and data processing and how the data are organized. Without dependence from that, the OLAP-product a part of resources of the user is whether or not, this fact should be transparent for the user.
- **Accessibility:** The OLAP tool should be capable of applying its own logical structure to access heterogeneous sources of data and perform any conversions necessary to present a coherent view to the user. The tool (and not the user) should be concerned with where the physical data comes from. Business analyst should have a possibility to analyze within the framework of the common conceptual scheme, thus the data may remain under the control of old, "inherited" DBMS,
- **Consistent reporting performance:** Performance should not degrade as the number of dimensions in the model increases. With increasing of numbers of measures and database size analysts should not face with any decrease of productivity.

OLAP Guidelines (Dr. E.F. Codd Rule)

- **Client / Server Architecture:** The server component of OLAP tools should be sufficiently intelligent that the various clients can be attached with minimum effort. The server should be capable of mapping and consolidating data between disparate databases. Requires open, modular systems. Not only the product should be client/server but the server component of an OLAP product should allow that various clients could be attached with minimum effort and programming for integration.
- **Generic dimensionality:** Not limited to 3-D and not biased toward any particular dimension. A function applied to one dimension should also be able to be applied to another. Every data dimension should be equivalent in its structure and operational capabilities.
- **Dynamic sparse matrix handling:** Related both to the idea of nulls in relational databases and to the notion of compressing large files, a sparse matrix is one in which not every cell contains data. OLAP systems should accommodate varying storage and data-handling options. OLAP tool should guarantee optimal processing of the sparse matrixes. Access speed should be saved without dependence from data cells layout and to be a constant for the models having different number of measures and different data sparse.
- **Multiuser support:** Supports multiple concurrent users, including their individual views or slices of a common database. OLAP tools must provide concurrent retrieval and update access, integrity and security.

OLAP Guidelines (Dr. E.F. Codd Rule)

- **Unrestricted cross-dimensional operations:** Computational facilities must allow calculation and data manipulation across any number of data dimensions, and must not restrict any relationship between data cells. All dimensions are created equal, so all forms of calculation must be allowed across all dimensions, not just the measures dimension.
- **Intuitive data manipulation:** Users shouldn't have to use menus or perform complex multiple step operations when an intuitive drag and drop action will do. Data manipulation inherent in the consolidation path, such as drilling down or zooming out, should be accomplished via direct action on the analytical model's cells, and not require use of a menu or multiple trips across the user interface.
- **Flexible reporting:** Users should be able to print just what they need, and any changes to the underlying model should be automatically reflected in reports. Reporting facilities should present information in any way the user wants to view it.
- **Unlimited dimensions & aggregation levels:** Supports at least 15, and preferably 20, dimensions. The number of data dimensions supported should, to all intents and purposes, be unlimited. Each generic dimensions should enable an essentially unlimited number of user-defined aggregation levels within any given consolidation path.

Multiple Choice Question

1. _____ is a method of incremental conceptual clustering.
a) CORBA.
b) OLAP.
c) COBWEB.
d) STING.

2.. Effect of one attribute value on a given class is independent of values of other attribute is called _____.
a) value independence.
b) class conditional independence.
c) conditional independence.
d) unconditional independence.

3. The main organizational justification for implementing a data warehouse is to provide _____.
a) cheaper ways of handling transportation.
b) decision support.
c) storing large volume of data.
d) access to data.

4. Multidimensional database is otherwise known as_____.
a) RDBMS
b) DBMS
c) EXTENDED RDBMS
d) EXTENDED DBMS

5. Data warehouse architecture is based on _____.
a) DBMS
b) RDBMS
c) Sybase
d) SQL Server.

REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf DATA MINING BOOK WRITTEN BY Micheline Kamber
- https://www.javatpoint.com/three-tier-data-warehouse-architecture
- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education
- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier
- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education
- Mallach," Data Warehousing System", TMH