A Technical Review on On-Line Analytical Processing (OLAP)

K. Jayapriya1, E. Girija2,III-M.C.A., R.Uma.3,M.C.A.,M.Phil.,
Department of computer applications, Assit.Prof,Dept of M.C.A,
Dhanalakshmi Srinivasan College of Arts & Science,Perambalur 621 212. Tamil Nadu.
lallipriya1994@gmail.com,
girijashivani94@gmail.com,
Uky.uma@gmail.com
9791386538.

ABSTRACT:
Online Analytical Processing (OLAP) refers to the general activity of querying and presenting text and number data from data warehouses and/or data marts for analytical purposes. Database solved one part of the problem that is the data storage. To retrieve the appropriate information from the business data a variety of tools, using different techniques, exists that performs simple or complex tasks involving mathematical and statistical operation. Its main idea is providing navigation through data to non-expert users, so that they are able to interactively generate ad-hoc queries without the intervention of IT professionals. This name was introduced in contrast to On-Line Transactional Processing (OLTP), so that it reflected the different requirements and characteristics between both kinds of uses.

Keywords: Data Mining, Data Warehouse, Decision Support System(DSS), Desktop OLAP (DOLAP), Hybrid OLAP (HOLAP), Multidimensional Data, Multidimensional OLAP (MOLAP), On-Line Analytical Processing (OLAP), Relational OLAP (ROLAP).

1. INTRODUCTION:

OLAP (On-Line Analytical Processing) is different from both; it is primarily a software technology concerned with fast analysis of enterprise information. Often OLAP systems are data warehouse front-end software tools to make aggregate data available efficiently, for advanced analysis, to an enterprise’s managers.

It is essential that an OLAP system provides facilities for a manager to pose ad hoc complex queries to obtain the information that he/she requires. Another term that is being used increasing is business intelligence. It is sometimes used to mean both data warehousing and OLAP.

DATA MINING:

Data mining is a collection of techniques for efficient automated discovery of previously unknown, valid, novel, useful and understandable patterns must be actionable so that they may be
used in an enterprise’s decision making process.

**GROWTH IN OLTP DATA:**

The first database systems were implemented in the 1960s and 1970s. Many enterprises therefore have more than 30 years of experience in using database systems and they have accumulated large amounts of data during that time. Also as noted earlier, other technological developments also have resulted in growth, in some cases dramatic growth, in accumulated data.

**DATA MINING PROCESS:** Online analytical processing (OLAP) that is designed to assist decision makers in an enterprise by providing summary data based on information that is stored in a data warehouse.

### 2. GUIDELINES FOR OLAP IMPLEMENTATION:

The guidelines are, perhaps not surprisingly, somewhat similar to those presented for data warehouse implementation.

- **Vision:**
  The OLAP team must, in consultation with the users, develop a clear vision for the OLAP system.

- **Senior management support:**
  The OLAP project should be fully supported by the senior managers. Since a data warehouse may have been developed already, is should not be difficult.

- **Selecting an OLAP tool:**

  The OLAP team should familiarize themselves with the ROLAP and MOLAP tools available in the market.

- **Corporate strategy:**
  The OLAP strategy should fit with the enterprise strategy and business objectives.

- **Focus on the users:**
  The OLAP project should be focused on the users. A good GUI user interface should be provided to non-technical users.

- **Joint management:**
  The OLAP project must be managed by both the IT and business professionals.

- **Review and adapt:**
  Organizations evolve and so must the OLAP system.

### 3. Types of OLAP servers:

**ROLAP versus MOLAP versus HOLAP:**

OLAP servers present business users with multidimensional data from data warehouses or data marts, without concerns regarding how or where the data are stored. However, the physical architecture and implementation of OLAP servers must consider data storage issues.

**Relational OLAP (ROLAP) servers:**

These are the intermediate servers that stand in between a relational back-end server and client front-end tools. They use a relational or extended-relational DBMS to store and manage warehouse data, and OLAP middleware to support missing pieces. ROLAP
technology tends to have greater scalability than MOLAP technology.

The features of ROLAP are;

- Store and manage warehouse data in relational DBMS or specialized relational DBMS.
- Maps operations on multidimensional data to standard relational operator.
- Requires OLAP middleware to support missing pieces.

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.

Multidimensional OLAP (MOLAP) servers:

These servers support multidimensional views of data through array-based multidimensional storage engines. They map multidimensional views directly to data cube array structures.

The advantages of using a data cube is that allows fast indexing to precompiled summarized data. Many MOLAP servers adopt a two-level storage representation to handle dense and sparse data sets.

The features of MOLAP are;

- Store and manage warehouse data in multi-dimensional DBMS.
- Array based storage structure.
- Direct access to array data structure.

Advantages:

- Excellent performance: MOLAP cubes are built for fast data retrieval, and is 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.

Hybrid OLAP (HOLAP) servers:

The hybrid OLAP approach combines ROLAP and MOLAP technology, benefiting from the greater scalability of ROLAP and the faster computation of MOLAP. The Microsoft SQL server 2000 supports a hybrid OLAP server.

The features of HOLAP are;

- Storing detailed data in RDBMS
- Storing aggregated data in MDDMS
- User access via MOLAP tools.

4. Indexing OLAP data:

To facilitate efficient data accessing, most data warehouse systems support index structures and materialized views. General methods of select cuboids for materialization were discussed in the previous section. In this section, we examine how to index OLAP data by bitmap indexing and join indexing.
The bitmap indexing method is popular in OLAP products because it allows quick searching in data cubes. The bitmap indexing is an alternative representation of the record-ID (RID) list. If the attribute has the value \( v \) for a given row in the data to 1 in the corresponding row of the bitmap index. All other bits for that row are set to 0.

<table>
<thead>
<tr>
<th>RID</th>
<th>item</th>
<th>city</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>H</td>
<td>V</td>
</tr>
<tr>
<td>R2</td>
<td>C</td>
<td>V</td>
</tr>
<tr>
<td>R3</td>
<td>P</td>
<td>V</td>
</tr>
<tr>
<td>R4</td>
<td>S</td>
<td>V</td>
</tr>
<tr>
<td>R5</td>
<td>H</td>
<td>T</td>
</tr>
<tr>
<td>R6</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>R7</td>
<td>P</td>
<td>T</td>
</tr>
<tr>
<td>R8</td>
<td>S</td>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RID</th>
<th>H</th>
<th>C</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R6</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>R7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RID</th>
<th>V</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>R5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>R8</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

The join indexing method gained popularly from its use in relational database query processing. Traditional indexing maps the value in a given column to a list of rows having the value. Then the join index record contains the pair (RID, SID), Where RID and SID are record identifiers from the joinable relation.

Advantages:

I. Useful for low cardinality domains.

II. Reduction in space and processing time.

5. Efficient processing of OLAP Queries:

The purpose of materializing cuboids and constructing OLAP index structure is to speed up query processing in data cubes.

Determine which operations should be performed on the available cuboids: This involves transforming any selection, projection, roll-up, and drill-down operations specified in the query into corresponding SQL and/or OLAP operations.

Determine to which materialized cuboids the relevant operations should be applied: This involves identifying all of the materialized cuboids that may potentially be used to answer the query, pruning the above set using knowledge of “dominance” relationships among the cuboids, estimating the costs of using the remaining materialized cuboids, and selecting the cuboids with the least cost.
6. From On-Line Analytical Processing To On-Line Analytical Mining:

Online analytical mining (OLAM) (also called OLAP mining) integrates on-line analytical processing (OLAP) with data mining and mining knowledge in multidimensional databases. Among the many different paradigms and architectures of data mining systems, OLAM is particularly important for the following reasons.

- High quality of data in data warehouse.
- Available information processing infrastructure surrounding data warehouses.
- OLAP based exploratory data analysis.
- Online selection of data mining functions.

7. Architecture for On-Line Analytical Processing:

The OLAM and OLAP server platform analytical mining in data cubes in a similar manner as an OLAP server performs on-line analytical processing.

A metadata directory is used to guide the access of the data cube. The data cube can be constructed by accessing and/or integrating multiple databases via an MDDB API and/or by filtering a data warehouse via a database API that may support OLE DB or ODBC connections.

Since, such as concept description, association, classification, prediction, clustering, time-series analysis, and usually consists of multiple integrated data mining modules and is more sophisticated than an OLAP server.
The capability of OLAP to provide multiple and dynamic views of summarized data in a data warehouse sets a solid foundation for successful data mining.

8. OLAP Operations in the Multidimensional Data Model:

A number of OLAP data cube operations exists to materialize these different views, allowing interactive querying and analysis of the data at hand. Hence OLAP provides a user-friendly environment for interactive data analysis.

- **Roll-up**: The roll-up operation performs aggregate on data cube, either by **climbing-up a concept hierarchy for a dimension reduction**.

- **Drill-down**: Drill down is the reverse of roll-up. Drill-down can be realized by either **stepping down a concept hierarchy for a dimension or introducing additional dimensions**.

- **Slice and dice**: The slice operation performs a selection on one dimension of the given cube, resulting in a sub cube.

- **Pivot**: Pivot is a visualization operation that rotates the data axes in views in order to provide an alternative presentation of the data. Shows a pivot operation where the item and location axes in a 2-d slice are rotated.

9. Need for OLAP:

OLAP is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user.

- Calculation and modeling applied across dimensions, through hierarchies and/or across members
- Trend analysis over sequential time periods
- Slicing subsets for on-screen viewing
- Drill-down to deeper levels of consolidation
- Reach-through to underlying detail data
- Rotation to new dimensional comparison in the viewing area.

OLAP is implemented in a multi-user client/server mode and offers consistently rapid response to queries, regardless of database size and complexity.

10. OLAP Server:

An OLAP server is a high-capacity, multi-user data manipulation engine specifically designed to support and operate on multi-dimensional data structures. The OLAP server may either physically stage the processed multi-dimensional information to deliver consistent and rapid response times to end users, or it may populate its data structures in real-time relational or other databases, or offer a choice of both.

Given the current state of technology and the user requirement for consistent and rapid response times, staging the multi-dimensional data in the
OLAP server is often the preferred method.

11. OLAP Applications:

OLAP is widely used in several realms of data management. Some of these applications include:

1. Financial Applications
   - Activity based costing (resource allocation)
   - Budgeting
2. Marketing/Sales Applications
   - Market Research Analysis
   - Sales Forecasting
   - Promotions Analysis
   - Customer Analyses
   - Market/Customer Segmentation
3. Business Modeling
   - Simulating business behavior.
   - Extensive, real time decision support system for managers

All of the above applications need ability to provide managers with the information they need to make effective decision about an organisation’s strategic directions. The key indicator of a successful OLAP applications is its ability to provide information, as needed, that is, its ability to provide “just-in-time” information for effective decision making. This requires more than a base level of detailed data.

12. Benefits of using OLAP:

OLAP holds several benefits for business:

i. OLAP helps managers in decision-making through the multidimensional data views that is capable of providing, thus increasing their productivity.

ii. OLAP applications are self-sufficient owing to the inherent flexibility provided to the organized databases.

iii. It enables simulation of business models and problems, through extensive usage of analysis-capabilities.

iv. In conjunction with data warehousing, OLAP can be used to provide reduction in the application backlog, faster information retrieval and reduction in query drag.

13. Conclusion:

The applications and benefits of OLAP indicate clearly that is one of the revolutionary technologies that can be used extensively in data management. All business, big or small, can benefit from using OLAP, at some scale or other. The instant access to information and the apparent knowledge gained from this information is invaluable considering against the cost involved in establishing this setup.

References:


