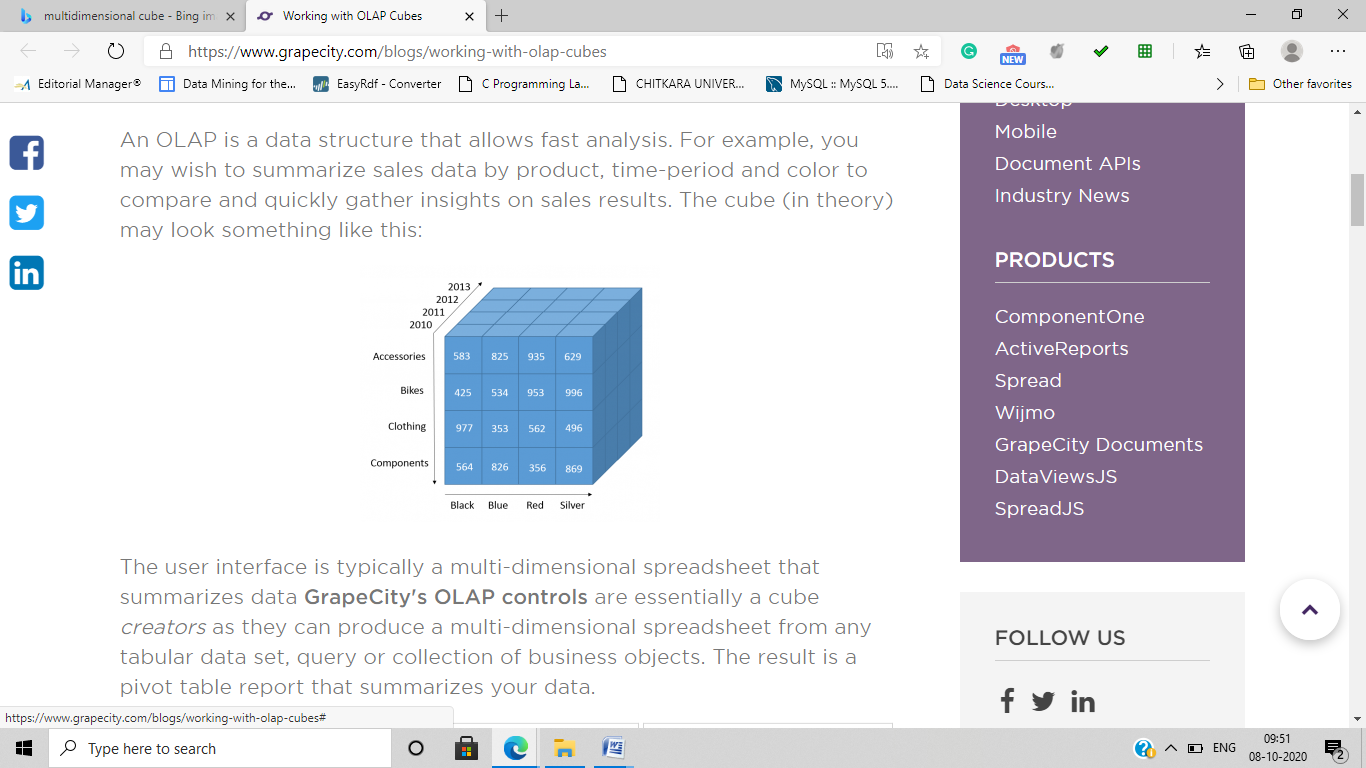
**Multidimensional Logical Modeling (OLAP)**

In a Logical relational data model (MDDM), data is stored in the form of a two dimensional table containing n number of rows and columns. But in MDDM data is visualized as a multidimensional cube. In a cube the dimensions are displayed along each axis and measures are stored in individual cells

A cube is a multidimensional structure that contains information for analytical purposes; the main constituents of a cube are dimensions and measures. Dimensions define the structure of the cube that you use to slice and dice over, and measures provide aggregated numerical values of interest to the end user. As a logical structure, a cube allows a client application to retrieve values, of measures, as if they were contained in cells in the cube; cells are defined for every possible summarized value. A cell, in the cube, is defined by the intersection of dimension members and contains the aggregated values of the measures at that specific intersection.

**SQL Server Analysis Services -> Azure Server Analysis Service**

****

Dimension Analysis

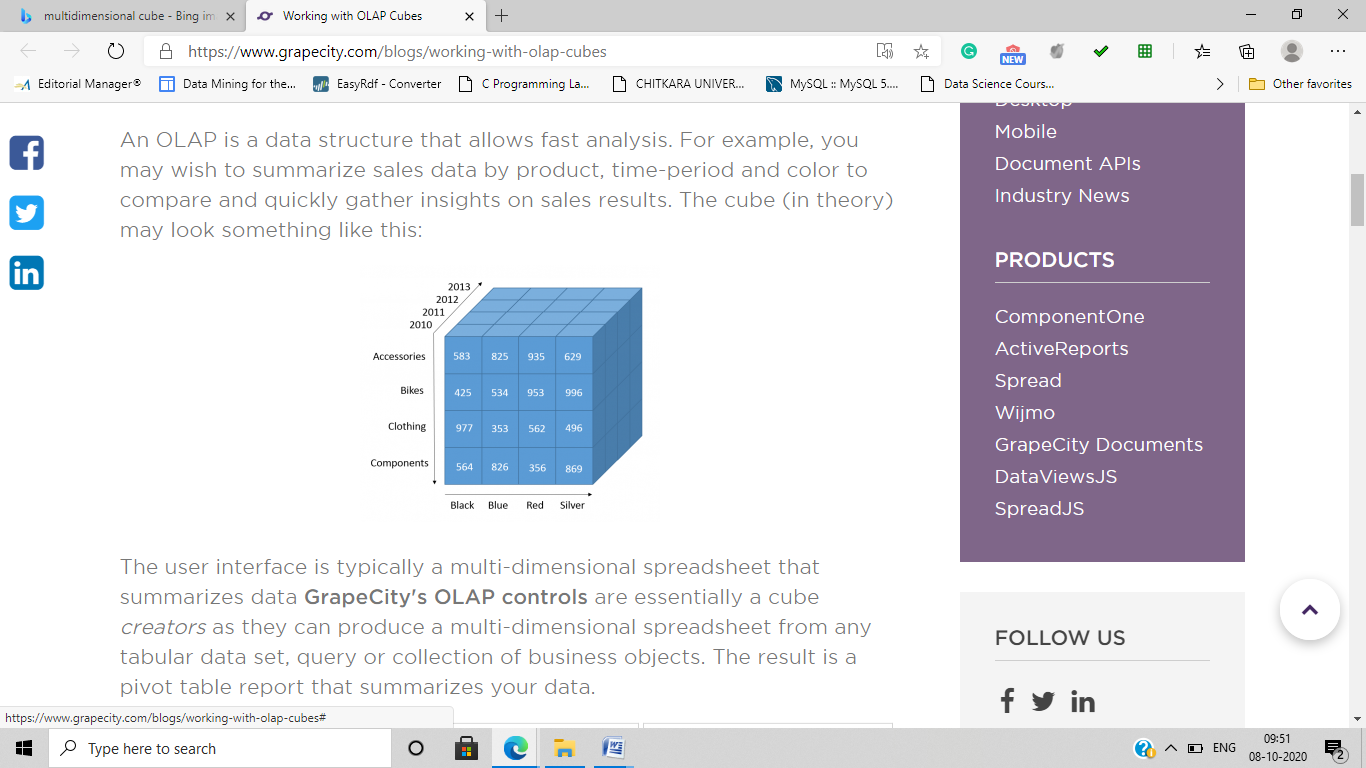
One Dimensional Data –

|  |  |
| --- | --- |
| Section | Sales Amount |
| Infant | 1000 |
| Kids | 2000 |
| Men | 3000 |

Two Dimensional Data

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Accessories | Clothing | Sales Amount |
|  |  |  |  |
|  |  |  |  |

Three Dimensional Data



To understand how data is stored in a three dimensional cube take an analogy of a notebook. The notebook is a cube which has several pages on each page data along two dimension is stored. Each page act as a attribute of the third dimension.

Hyper Cubes.

The hyper cube is a generalization of a 3 dimensional cube to n dimensional cube.

**OLAP operations in multidimensional**

A number of OLAP data cube operations exist to materialize these different views, allowing querying and analysis of the data at the end.

Rollup – The rollup operation also known as drill up performs aggregation on a data cube by climbing up a dimensional hierarchy. For Example week-Month-Quarter-Year.

Drill Down – It is the reverse of roll up. It navigates from less detailed data to more detailed data.

These two operations are also know as Aggregations

Slice

The slice operation performs a selection on only one dimension of a given cube.

Dice

The dice operation on the other hand defines a subcube by performing a selection on two or more dimensions.

These two operations are part of Restrictions (Selection/Projection)

PIVOT or Rotate

Pivot is a visualization operation that rotates the data axis in view in order to provide an alternative presentation of the data.

**EF Codd Rules to MDDM**

EF Codd has given twelve basic guidelines which are as follows

1) Multidimensional conceptual view – Data to be presented to the user in multidimensional paradigm.

2) Transparency – User need not know that they are using OLAP dbs

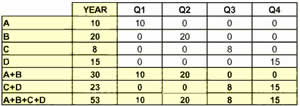
3) Accessibility – Reports, Dashboards, OLAP Queries

4) Consistent Reporting Performance – Performance should be same regardless the number of dimension in use

5) Client Server Architecture – Tools should be deployed in the client server architecture. The server component must enable clients to be attached with a minimum effort of integration programming

6) Generic Dimensionality – Dimensions are all equal

7) Dynamic Sparse Matrix Handling – Null values should be stored in an efficient way.



8) Multi user support – Tools should support more than one users.

9) Unrestricted Cross Dimensional Support – Aggregation rules should be applied consistently across all dimensions.

10) Intuitive data manipulation –

The user view of data should contain everything visible on the interface itself.

11) Flexible Reporting – User should be able to present data in any way they like.

12) Unlimited dimensions and aggregation levels – There should be no limit to the number of dimensions and levels in the model.

**Different OLAP architectures**

**DW Overview: Data Warehouse Architecture**

Below is a diagram showing a typical data warehouse architecture.

|  |
| --- |
| Data Warehouse Architecture Figure 1. Data Warehouse Architecture |

Data warehouse architecture should include tools for the following:

* extracting data from multiple operational databases and external sources
* cleaning, transforming, and integrating the data
* loading data into the data warehouse
* periodically refreshing the data warehouse to reflect updates at the source and to purge data from the data warehouse

In addition to the main data warehouse, there may be several departmental data marts. Data marts are departmental subsets focused on selected subjects. Data in data warehouses and data marts are stored and managed by one or more data warehouse servers. The data warehouse servers present multidimensional views of data to a variety of front end tools, including analysis tools, query/reporting tools, and data mining tools. And there is a repository for storing and managing metadata and tools which monitor and administer the data warehousing system.

What are some different architectural alternatives? Many organizations want to implement an integrated enterprise warehouse that collects information about all different aspects of the whole organization. This requires extensive business modeling which may take a long time and cost a lot in order to complete it successfully. As an alternative to this, some organizations choose to settle for data marts instead. This enables faster roll out in comparison to the enterprise wide warehouse, but it may lead to complex integration problems in the long run.

**Relational -ROLAP**

**Multidimensional- MOLAP**

**Hybrid-HOLAP**

1) ROLAP

**OLAP: Relational OLAP**

Relational On-Line Analytical Processing (ROLAP) performs dynamic multidimensional analysis of data stored in a relational database, rather than in a multidimensional database. The traditional OLAP's slice and dice functionality is equivalent to adding a WHERE clause in the SQL statement. The design may be structured in the form of a star or its variations. A typical use of ROLAP is for large data size that is infrequently queried, such as historical data.

ROLAP tools feature the ability to ask any question and one is not limited to the contents of a cube

Characteristics of this model

Data and meta data is stored as records in Relational Tables ROLAP

End users are supplied with a multidimensional viewing tool

There is high capacity connectivity to the powerful servers

There is no limitation on the size of the database

Complex SQL code is generated by ROLAP tools

The chart below highlights advantages and disadvantages of ROLAP.

|  |  |
| --- | --- |
| ***Advantages*** | Well known environments (relational database). |
| Can leverage functionality that comes with relational database with ROLAP technologies. |
| Can be used with data warehouse and OLTP systems. |
| No pre-aggregation is needed - avoid the data explosion effect that some MOLAP implementations incur with large scale models. |
| Can handle large amounts of data - the limitation is the data size of the underlying relational database. OLAP itself has no limitation on data amount. |
| Full security and administration is provided through RDBMSs. |
| Performs better than MOLAP when the data is sparse. |
| Performance is getting better by adding more OLAP functions and employing various storage and query optimization techniques. |
| ***Disadvantages*** | Performance can be slow, since each ROLAP report is a SQL query in the relational database. |
| Does not have complex functions that are provided by OLAP tools. |
| Limited by SQL functionality. |
| Hard to maintain aggregate tables in the data warehouse. |
| ***Major Players*** | Discover 3 from Oracle**, DSS Agent from MicroStrategy**, MetaCube from IBM Informix, Platinum Beacon from Platinum, Brio, Business Objects, Cognos Powerplay, Microsoft Analysis Services |

|  |
| --- |
| ROLAP Architecture Figure 1. ROLAP architecture |

As explained in the previous section, ROLAP accesses data stored in a data warehouse (relational database) to provide OLAP analyses. The premise of ROLAP is that OLAP capabilities are best provided directly against the *relational database*.

After defining the data model for the data warehouse, data from OLTP systems is loaded into the database. Database routines are run to aggregate the data, if required by the data model. Then indices are created to optimize query access time. Users submit multidimensional analyses to the ROLAP engine, which then dynamically transforms the requests into SQL execution plans. The SQL is submitted to the relational database for processing, the relational query results are cross-tabulated. A multidimensional result set is returned to the users. It is capable of utilizing pre-calculated results when they are available, or dynamically generating results from atomic information when necessary.

The ROLAP architecture was invented to directly access data from data warehouses, thereby supporting optimization techniques to meet batch window requirements and providing fast response time. These optimization techniques include application-level table partitioning, aggregate inferencing, denormalization support, and multiple fact table joins.

ROLAP is a three-tier, client/server architecture. The database layer utilizes relational databases for data storage, access, and retrieval processes. The application logic layer is the ROLAP engine which executes the multidimensional reports from multiple users. The ROLAP engine integrates with a variety of presentation layers, through which users perform OLAP analyses.

|  |
| --- |
| ROLAP Architecture Figure 1. ROLAP architecture |

2) MOLAP

**OLAP: Multidimensional OLAP**

Multidimensional On-Line Analytical Processing (MOLAP) is the more traditional way of OLAP analysis, in which, data is stored in a multidimensional cube. This allows users to view different aspects of data aggregates such as sales by time period, geography, or product. The storage is not in a relational database. If the data is stored in a relational database, it can be viewed multidimensionally, but only by successively accessing and processing a table for each dimension. MOLAP processes data that is already stored in a multidimensional array in which all possible combinations of data are reflected, each in a cell that can be accessed directly. MOLAP is more appropriate for cubes with frequent use and the necessity for rapid query response.

MOLAP Data Model uses Proprietary databases ( therefore is data that is owned and controlled by an individual, an organization or a group.)

Characteristics

Data is stored in pre-calculated tables

The data resides in proprietary databases.

The database is organized to allow rapid retrievals

MOLAP tools has the ability to process data which is there in the form of a cube.

The chart below highlights advantages and disadvantages of MOLAP.

|  |  |
| --- | --- |
| ***Advantages*** | Excellent performance since pre-aggregation provides quicker response time. |
| Availability of extensive libraries of complex functions for OLAP analyses. |
| Optimal for slice and dice operations. |
| Performs better than ROLAP when data is dense. |
| ***Disadvantages*** | Usually more than 90% of cells are empty - issue with sparsity. |
| Limited in the amount of data it can handle, since all calculations are performed when the cube is built. Therefore, it is not commonly used above 20-50 GB - scalability problem. |
| Difficult to change dimension without re-aggregation. |
| Data must be copied and moved into data stores. |
| Originated from query tools, thereby lacking the architecture. |
| Requires additional investment since cube technology is often proprietary and does not already exist in organizations. |
| Lacks security and administration features which RDBMSs can bring. |
| ***Major Players*** | Hyperion, Executive Viewer, CFO Vision, BI/Analyze, PowerPlay, Business Objects, Genita, Holos, MS OLAP Services, Pilot, ProCube |

|  |
| --- |
| MOLAP Architecture Figure 1. MOLAP architecture |

As explained in the previous section, MOLAP utilizes a proprietary multidimensional database to provide OLAP analyses. The main premise of this architecture is that data must be stored *multidimensionally* to be viewed multidimensionally.

Data from various operational systems is loaded into a multidimensional database through a series of batch routines. Once this atomic data has been loaded into the multidimensional database, the general approach is to perform a series of calculations in batch to aggregate along the dimensions and fill the multidimensional array structures. Then indices are created, and hashing algorithms are used to improve query access time.

When users request OLAP reports through the interface, the application logic layer of the multidimensional database retrieves the stored data. It reads the pre-compiled data and has limited capabilities to dynamically create aggregations or to calculate metrics that have not been pre-calculated and stored.

MOLAP is a two-tier, client/server architecture. The multidimensional database serves as both the database layer and the application logic layer. In the database layer, it is responsible for all data storage, access, and retrieval processes. In the application logic layer, it is responsible for the execution of all OLAP requests. The presentation layer integrates with the application logic layer and provides an interface through which the users view and request OLAP analyses. The client/server architecture allows multiple users to access the same multidimensional database.

|  |
| --- |
| MOLAP Architecture Figure 1. MOLAP architecture |

3) HOLAP

**HOLAP: Hybrid OLAP**

Hybrid On-Line Analytic Processing (HOLAP) is a mixture of MOLAP and ROLAP technologies. For summary type query, HOLAP leverages cube technology for faster performance. When detail information is needed, it can drill through from the cube into the underlying relational database. Cubes stored as HOLAP are smaller than equivalent MOLAP cubes and respond quicker than ROLAP cubes for queries involving summary data. HOLAP storage is generally suitable for cubes that require rapid query response for summaries based on a large amount of base data.

In order to deliver the combined strengths of MOLAP and ROLAP technologies, HOLAP systems must comply with the following rules:

* Fast access at all levels of aggregation (MOLAP requirement)
* Easy aggregate maintenance (MOLAP requirement)
* Compact aggregate storage (MOLAP requirement) - for high-level aggregates in order to economize disk space.
* Dynamically updated dimensions (ROLAP requirement) - real time access to the data itself and to rapidly changing structures.
* Multidimensional view based on RDBMS metadata (ROLAP requirement) - should point to the appropriate RDBMS tables and automatically generate required SQL statements when modifying the multidimensional view. It reduces development time and maintenance.

The chart below highlights advantages and disadvantages of HOLAP.

|  |  |
| --- | --- |
| ***Advantages*** | Combined advantages of both MOLAP and ROLAP (for a full list, look at the MOLAP and ROLAP sections). |
| Can combine the ROLAP technology for sparse regions and MOLAP for dense regions. Also ROLAP for storing the detailed data and MOLAP for higher-level summary data. |
| ***Disadvantages*** | Complex - HOLAP server must support both MOLAP and ROLAP engines and tools to combine both storage engines and operations. |
| Functionality overlap - between storage and optimization techniques in ROLAP and MOLAP engines. |
| ***Major Players*** | Express from Oracle, IBM DB 2 OLAP Server, Microsoft OLAP Services, Sagent Holos |

|  |
| --- |
| HOLAP architecture Figure 1. HOLAP architecture |

4) DOLAP

**DOLAP: Desktop OLAP**

Desktop On-Line Analytic Processing (DOLAP) is single-tier, desktop-based OLAP technology. It is able to download a relatively small hypercube from a central point, usually from data mart or data warehouse, and perform multidimensional analyses while disconnected from the source. Data sets are limited to the boundaries defined by the user with no access to granular data. In general, cubes contain summarized data, organized in a fixed structure of dimensions. Therefore, it is ideal for well-understood, recurring analytic questions and reporting.

The chart below highlights advantages and disadvantages of DOLAP.

|  |  |
| --- | --- |
| ***Advantages*** | User friendly - user can pivot and manipulate data locally from the returned result set stored on the desktop. |
| Excellent query performance - it collects, aggregates, and calculates data in advance of the analysis. |
| Low cost per seat and maintenance. |
| Useful for mobile users who cannot always connect to the data warehouse. |
| Easiest to deploy among all OLAP approaches. |
| ***Disadvantages*** | Limited functionality and data capacity. |
| ***Major Players*** | Cognos, Business Objects, Brio, Crystal Decisions, Hummingbird, Oracle |

ROLAP

Query Processing Mechanism

User uses his own language to fetch the data from the warehouse.

The ROLAP server into complex SQL queries

The query is executed on the data warehouse

The analytical server checks the form in which data is required by the user and pass it to the end user.

1) Detailed and lightly summarized data

2) Very large tables with textual data

3) Data is store in Relational

4) Complex function are not available it is limited to what is their in SQL

5) Slow Performace

Query processing mechanism in MOLAP

The OLAP servers creates and stored data in cubes.

warehouse.

The OLAP server checks whether summary is needed - Multidimensional cubes