Online Analytical Processing (OLAP) – Codd, 1993.

Definition (The OLAP Council): 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.

OLAP

- Support (almost) ad-hoc querying for business analyst
- Think in terms of spreadsheets
  - View sales data by geography, time, or product
- Extend spreadsheet analysis model to work with warehouse data
  - Large data sets
  - Combine interactive queries with reporting functions
- Multidimensional view of data is the foundation of OLAP
  - Data model, operations, etc.

OLAP Operations

- **rotate-pivoting**: pivoting the cube
- **roll up / drill down**: going up/down in the hierarchy
day → week → month → quarter → year
city → province → state → country → continent
- **slice and dice**: projecting (fixing one or more dimension)
  select, Time, Day, StoreName, SalesVolume
  from Sales, Product, Store, Time
  where ProductType = "The item PC" is

Data Warehousing and OLAP

- Although sometimes used interchangeably, the terms **data warehousing** and **online analytical processing** (OLAP) apply to different components of systems often referred to as decision support systems or business intelligence systems. Components of these types of systems include databases and applications that provide the tools analysts need to support organizational decision-making.
Data Warehousing and OLAP

- A data warehouse is a database containing data that usually represents the business history of an organization.
- This historical data is used for analysis that supports business decisions at many levels, from strategic planning to performance evaluation of a discrete organizational unit.
- Data in a data warehouse is organized to support analysis rather than to process real-time transactions as in online transaction processing systems (OLTP).

Note for OLAP we need:
Database (Data Warehouse Storage)
OLAP Server
OLAP Client (Like Data Analyzer or Excel)

OLAP stages

- picking the data from the data warehouse (DB)
- building the “cube” (OLAP SERVER)
  - cube = pre-computation of “all” possible combinations
  - The high performance comes from the fact that all the “queries” (combinations) have been anticipated.
- playing with the “cube” (OLAP CLIENT)
  - slice-dice, roll-up/drill-down, etc.

Three-Tier Decision Support Systems

- Warehouse database server
  - Almost always a relational DBMS, rarely flat files
- OLAP servers
- Clients
  - Query and reporting tools
  - Analysis tools
  - Data mining tools
**The Complete Decision Support System**

- Information Sources
- Data Warehouse Server (Tier 1)
- OLAP Servers (Tier 2)
- Clients (Tier 3)

**OLAP Server: Query Engine Requirements**

- Aggregates (maintenance and querying)
  - Decide what to precompute and when
- Query language to support multidimensional operations
  - Standard SQL falls short
- Scalable query processing
  - Data intensive and data selective queries

**OLAP Server Architectures**

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)
- Desktop OLAP (DOLAP)

**ROLAP**

- Relational OLAP (ROLAP)
  - Use relational or extended-relational DBMS to store and manage warehouse data and OLAP middle ware to support missing pieces
  - Include optimization of DBMS backend, implementation of aggregation navigation logic, and additional tools and services
  - greater scalability
ROLAP As-Is not Appropriate For OLAP

- n-Dimensional Structure
  - Summarization may not be possible
- End-User Confusion about SQL Query Tools
- Poor Performance

Relational Table

<table>
<thead>
<tr>
<th>Product</th>
<th>Region</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuts</td>
<td>East</td>
<td>50</td>
</tr>
<tr>
<td>Nuts</td>
<td>West</td>
<td>60</td>
</tr>
<tr>
<td>Nuts</td>
<td>Central</td>
<td>100</td>
</tr>
<tr>
<td>Screws</td>
<td>East</td>
<td>40</td>
</tr>
<tr>
<td>Screws</td>
<td>West</td>
<td>70</td>
</tr>
<tr>
<td>Screws</td>
<td>Central</td>
<td>80</td>
</tr>
<tr>
<td>Bolts</td>
<td>East</td>
<td>90</td>
</tr>
<tr>
<td>Bolts</td>
<td>West</td>
<td>120</td>
</tr>
<tr>
<td>Bolts</td>
<td>Central</td>
<td>140</td>
</tr>
<tr>
<td>Washers</td>
<td>East</td>
<td>20</td>
</tr>
<tr>
<td>Washers</td>
<td>West</td>
<td>10</td>
</tr>
<tr>
<td>Washers</td>
<td>Central</td>
<td>30</td>
</tr>
</tbody>
</table>

2 Dimensional Matrix

Sales Dimensioned By Products & Region

<table>
<thead>
<tr>
<th></th>
<th>East</th>
<th>West</th>
<th>Central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuts</td>
<td>50</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>Screws</td>
<td>40</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Bolts</td>
<td>90</td>
<td>120</td>
<td>140</td>
</tr>
<tr>
<td>Washers</td>
<td>20</td>
<td>10</td>
<td>30</td>
</tr>
</tbody>
</table>

No Standard SQL statement can return this structure of report
SQL Extensions for OLAP

• Extended family of aggregate functions
  – rank (top 10), N_Tile (percentile, e.g., 30% for each product)
• Reporting features
  – Matrix results
• Results of multiple group by’s
  – Running totals, cumulative totals
• Research on computing the cube efficiently
  – Share work among multiple group-by operations
  – Rollup requires single sort
• Papers by J. Gray (cube & roll-up, supported in SQL Server 6.5) and Red Brick (RISQL)

GROUPING SETS

```sql
SELECT COURSE, YEAR, AVG(grade) as GRADES_AVG
FROM GRADES
GROUP BY GROUPING SETS ((COURSE,YEAR),(COURSE),())
```

<table>
<thead>
<tr>
<th>COURSE</th>
<th>YEAR</th>
<th>GRADES_AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB</td>
<td>1998</td>
<td>95.03</td>
</tr>
<tr>
<td>DB</td>
<td>1999</td>
<td>93.00</td>
</tr>
<tr>
<td>DB</td>
<td></td>
<td>94.01</td>
</tr>
<tr>
<td>DW</td>
<td>1998</td>
<td>90.00</td>
</tr>
<tr>
<td>DW</td>
<td>1999</td>
<td>95.00</td>
</tr>
<tr>
<td>DW</td>
<td></td>
<td>92.50</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>93.25</td>
</tr>
</tbody>
</table>

MDX = “SQL” for Cubes instead for tables

```sql
SELECT [Route.nonground.Members] ON COLUMNS,
[Time.[1st half].Members] ON ROWS
FROM TestCube
WHERE ([Measures].[Packages])
```

MOLAP

• Multidimensional OLAP (MOLAP)
  – Array-based storage structures
  – Direct access to array data structures
• The First Generation OLAP.
• Good query performance
• Not Flexible – Require Multi Dimensional Design
• Require Expensive IT Resources.
• very poor storage utilization when the data is sparse
**DOLAP**

- Desktop OLAP.
- For small databases.
- Easy to integrate.
Codd’s Criteria for OLAP

- Multi-Dimensional Conceptual View
- Transparency
- Accessibility – Different sources.
- Consistent Reporting Performance
- Dynamic Sparse Matrix Handling
- Multi-User Support
- Intuitive Data Manipulation
- Unlimited Dimensions and Aggregation Levels

*Codd's Paper*