BUSINESS INTELLIGENCE

OLAP: On-Line Analytical Processing
BI Architecture

Figure 1. Typical business intelligence architecture.
WHICH DBMS FOR DATAWAREHOUSING?

CHALLENGERS

1. Amazon Web Services
2. HPE
3. Infobright
4. MarkLogic
5. 1010data

LEADERS

1. Oracle
2. Teradata
3. Microsoft
4. IBM
5. SAP

NICHES PLAYERS

1. Cloudera
2. MapR Technologies
3. Hortonworks
4. Transwarp
5. Pivotal
6. MemSQL

VISIONARIES

1. Kognitio
2. Exasol
3. Actian
4. MongoDB
5. Hitachi

ABILITY TO EXECUTE

COMPLETENESS OF VISION

As of February 2016
ON-LINE ANALYTICAL PROCESSING (OLAP)

- An OLAP server provides a *multidimensional view* starting from a datawarehouse.

- The multidimensional view can be navigated through pivot tables, reports, 2-D or 3-D plots, or it can be queried using a query language (e.g., MDX - MultiDimensional eXpressions).
The multidimensional model is useful to understand interactive data analysis, and how to improve the execution performance.
2-D CUBE

Fact Table

<table>
<thead>
<tr>
<th>StoreId</th>
<th>ProductId</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>P1</td>
<td>300</td>
</tr>
<tr>
<td>S2</td>
<td>P1</td>
<td>500</td>
</tr>
<tr>
<td>S3</td>
<td>P1</td>
<td>50</td>
</tr>
<tr>
<td>S1</td>
<td>P2</td>
<td>30</td>
</tr>
<tr>
<td>S2</td>
<td>P2</td>
<td>50</td>
</tr>
<tr>
<td>S3</td>
<td>P2</td>
<td>400</td>
</tr>
</tbody>
</table>

2-D Cube

CROSS TABULATION

<table>
<thead>
<tr>
<th>ProductId</th>
<th>StoreId</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>P1</td>
<td>300</td>
</tr>
<tr>
<td>P2</td>
<td>30</td>
</tr>
</tbody>
</table>

Cube, A. Albano
3-D CUBE

Sales

<table>
<thead>
<tr>
<th>StoreId</th>
<th>ProductId</th>
<th>DateId</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>P1</td>
<td>D1</td>
<td>300</td>
</tr>
<tr>
<td>S2</td>
<td>P1</td>
<td>D1</td>
<td>500</td>
</tr>
<tr>
<td>S3</td>
<td>P1</td>
<td>D1</td>
<td>50</td>
</tr>
<tr>
<td>S1</td>
<td>P2</td>
<td>D1</td>
<td>30</td>
</tr>
<tr>
<td>S2</td>
<td>P2</td>
<td>D1</td>
<td>50</td>
</tr>
<tr>
<td>S3</td>
<td>P2</td>
<td>D1</td>
<td>400</td>
</tr>
<tr>
<td>S2</td>
<td>P1</td>
<td>D2</td>
<td>200</td>
</tr>
<tr>
<td>S3</td>
<td>P1</td>
<td>D2</td>
<td>600</td>
</tr>
<tr>
<td>S1</td>
<td>P2</td>
<td>D2</td>
<td>900</td>
</tr>
<tr>
<td>S2</td>
<td>P2</td>
<td>D2</td>
<td>800</td>
</tr>
<tr>
<td>S3</td>
<td>P2</td>
<td>D2</td>
<td>70</td>
</tr>
</tbody>
</table>

Fact Table

3-D Cube
CUBE OPERATOR: SLICE

Sales **SLICE FOR** DateId = ‘D1’;

![Diagram showing the process of slicing a cube]
CUBE OPERATOR: DICE

Sales DICE FOR DateId = ‘D1’
StoreId IN (‘S1’, ‘S2’);
CUBE OPERATOR: PIVOT

PIVOT (Sales SLICE FOR DateId = ‘D1’);

**Rotate**: reorient the cube, visualization, 3D to series of 2D planes
CUBE OPERATORS: ROLL-UP and DRILL-DOWN

Roll-up aggregates data by dimension reduction or by navigating attribute hierarchy (Drill-down is the reverse of roll-up)

Hypothesis: one measure and aggregations by sum.

SALES ROLL-UP ON DateId
(total Qty by ProductId and by StoreId)
CUBE OPERATORS: ROLL-UP and DRILL-DOWN

0-D cube or Apex-cube
**CUBE OPERATORS: DRILL THROUGH**

**Drill-through** produces the facts that satisfy a cell coordinate.

**Sales**

<table>
<thead>
<tr>
<th>StoreId</th>
<th>ProductId</th>
<th>DatedId</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>P1</td>
<td>D1</td>
<td>300</td>
</tr>
<tr>
<td>S2</td>
<td>P1</td>
<td>D1</td>
<td>500</td>
</tr>
<tr>
<td>S3</td>
<td>P1</td>
<td>D1</td>
<td>50</td>
</tr>
<tr>
<td>S1</td>
<td>P2</td>
<td>D1</td>
<td>30</td>
</tr>
<tr>
<td>S2</td>
<td>P2</td>
<td>D1</td>
<td>50</td>
</tr>
<tr>
<td>S3</td>
<td>P2</td>
<td>D1</td>
<td>400</td>
</tr>
<tr>
<td>S2</td>
<td>P1</td>
<td>D2</td>
<td>200</td>
</tr>
<tr>
<td>S3</td>
<td>P1</td>
<td>D2</td>
<td>600</td>
</tr>
<tr>
<td>S1</td>
<td>P2</td>
<td>D2</td>
<td>900</td>
</tr>
<tr>
<td>S2</td>
<td>P2</td>
<td>D2</td>
<td>800</td>
</tr>
<tr>
<td>S3</td>
<td>P2</td>
<td>D2</td>
<td>70</td>
</tr>
</tbody>
</table>

- **Cube:**
  - P1: 1650
  - P2: 2250

- **Drill-down:**
  - S1 P2 D1 30
  - S2 P2 D1 50
  - S3 P2 D1 400
  - S1 P2 D2 900
  - S2 P2 D2 800
  - S3 P2 D2 70

*Cube, A. Albano*
CUBE NAVIGATION BY DIFFERENT USERS

- Branch manager look at sales of his/her stores for any product and any period.
- Product managers look at sales of some products in any period and in any market.
- Finance manager look at sales of a period compared to the previous period for any product and any market.
TEXTUAL NOTATION FOR CUBE OPERATORS

Hypothesis: one measure and aggregations by sum.

Sales(StoreId, ProductId, DateId)
is the cube with dimensions StoreId, ProdottoId, DataId, and measure M

A cube operation is denoted by substituting a dimension with a value
Sales(StoreId, ProductId, ‘D1’)  slice

Sales(‘S1’, ProductId, ‘D1’)  dice

Sales(‘S1’, ‘P1’, ‘D1’)  dice
Each dimension domain is extended with the value “*”, that means summarize data (sum) by all the dimension values.

Sales(StoreId, ProductId, *)

Sales by roll-up on DateId with sum(M)
CUBE OPERATORS: EXAMPLES

\[
\text{Sales}(\text{StoreId}, \text{ProductId}, \text{DateId}) =
\]

\[
\text{Sales}(\text{StoreId}, \text{ProductId}, *) =
\]

\[
\text{Sales}(\text{StoreId}, *, *) =
\]

\[
\text{Sales}(*, *, *) =
\]
What is

Sales(StoreId, 'P1', *) =

\begin{array}{c}
\text{P1} \\
\text{S1} \\
300 \\
\hline
\text{S2} \\
700 \\
\hline
\text{S3} \\
650
\end{array}
A data cube is extended with the value ‘*’ for each dimensions, and in the corresponding cells is stored the sum of the measure.
EXTENDED CUBE

With the ‘*’ values, the cube becomes a set of **cuboids**:

- **white** cells are the data cube
- **gray** cells are roll-up by a dimension,
- **dark gray** cells are roll-up by two dimensions
- **black** cells are roll-up by all dimensions.
## EXTENDED CROSS TABULATION

### CROSS TABULATION

<table>
<thead>
<tr>
<th>Storeld</th>
<th>ProductId</th>
<th>Qty</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>P1</td>
<td>300</td>
</tr>
<tr>
<td>S2</td>
<td>P1</td>
<td>500</td>
</tr>
<tr>
<td>S3</td>
<td>P1</td>
<td>50</td>
</tr>
<tr>
<td>S1</td>
<td>P2</td>
<td>30</td>
</tr>
<tr>
<td>S2</td>
<td>P2</td>
<td>50</td>
</tr>
<tr>
<td>S3</td>
<td>P2</td>
<td>400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ProductId</th>
<th>Storeld</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
</tr>
<tr>
<td>P1</td>
<td>300</td>
</tr>
<tr>
<td>P2</td>
<td>30</td>
</tr>
</tbody>
</table>

## EXTENDED CROSS TABULATION

<table>
<thead>
<tr>
<th>Storeld</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ProductId</td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
</tr>
<tr>
<td>P1</td>
<td>300</td>
<td>500</td>
<td>50</td>
</tr>
<tr>
<td>P2</td>
<td>30</td>
<td>50</td>
<td>400</td>
</tr>
<tr>
<td>Total</td>
<td>330</td>
<td>550</td>
<td>450</td>
</tr>
</tbody>
</table>
On the set of cuboids is defined the following partial order relation:

\[ C_1 \leq C_2 \quad \text{if } C_1 \text{ dimensions are included in } C_2 \text{ dimensions.} \]
• D = \{d_1, \ldots, d_N\} dimensions (degenerate or flat)
  • \(2^N\) cuboids

• Let \(#d_i\) = number of values for dimension \(d_i\)

• How many cells in total?

\[
\sum_{C \subseteq D} \prod_{d \in C} \#d = \prod_{i=1}^{N}(\#d_i + 1)
\]
CUBOIDS MATERIALIZATION

Complete

Partial
AGGREGATION FUNCTIONS TYPES

\[ V = V_1 \cup V_2 \quad V_1 \cap V_2 = \emptyset \]

Distributive

E.g., `sum()`, `min()`, `max()`, `count()`

\[
\begin{align*}
\text{sum}(V) &= \text{sum}(V_1) + \text{sum}(V_2) \\
\text{count}(V) &= \text{count}(V_1) + \text{count}(V_2)
\end{align*}
\]

\[ \text{sum}({v}) = v \quad \text{count}({v}) = 1 \]

Algebraic

E.g., `avg()`, `standard\_deviation()`

\[
\begin{align*}
\text{avg}(V) &= \frac{\text{sum}(V)}{\text{count}(V)} \\
\text{var}(V) &= \frac{\text{sum}(V^2) - \text{sum}(V)^2}{\text{count}(V)} \\
&= \frac{\text{sum}({v}^2)}{\text{count}(V) - 1}
\end{align*}
\]

\[ \text{sum}({v}^2) = v^2 \]

Holistic

E.g., `median()`, `mode()`, `rank()`.
CUBOIDS MATERIALIZATION

Complete

Partial

If the materialization is partial, which cuboids do we select?
**OLAP SYSTEMS SOLUTIONS**

**OLAP** refers to the technique of performing complex business analysis over the information stored in a data warehouse.

We will see how report developers use SQL to write queries, but there are business intelligence tools that allows a user or a developer to make data analysis and to build beautiful reports without any knowledge of SQL... which is generated automatically.
The DW is managed by a specialized RDBMS (Relational Data Server).

The **OLAP Client** provides presentation and reporting tools to deal with data analysis and visualization, and interacts with the Data Server.
The **OLAP Client** interacts with an **OLAP Server**, that supports multidimensional data and operations, and can be one of the following type:

- **MOLAP**, which stores in the local memory both the data cube, taken from the Data Server, and the aggregates of the extended cube, using a specialized data structure. *A MOLAP server does not support SQL, but MDX.*

- **ROLAP** which stores both the data and the aggregates of the extended cube in the Data Server. *ROLAP servers may also implement functionalities not supported in the SQL of the Data server.*

- **HOLAP** which stores the data in the Data Server, and the aggregates of the extended cube in the local memory.
The ROLAP case

Data storage modes

<table>
<thead>
<tr>
<th>OLAP Server</th>
<th>MOLAP</th>
<th>ROLAP</th>
<th>HOLAP</th>
<th>Offline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essbase</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>icCube</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Offline Cubes</td>
</tr>
<tr>
<td>Microsoft Analysis Services</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Local cubes, PowerPivot for Excel</td>
</tr>
<tr>
<td>MicroStrategy Intelligence Server</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>MicroStrategy Office, Dynamic Dashboards</td>
</tr>
<tr>
<td>Mondrian OLAP server</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Oracle Database OLAP Option</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Palo</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>SAS OLAP Server</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>TM1</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>SAP NetWeaver BW</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

Figure 20: ROLAP Meets Application Requirements
OLAP SYSTEMS: SOLUTION 3

The **OLAP client** interacts with a local **DOLAP system (Desktop OLAP)** which manages small amount of data extracted from the **OLAP server**, the **Data server** or an **operational DBMS**. It a good choice for those who travel and move extensively, by using portable computers.

E.g., Microsoft Power Pivot (Add-in of Excel)
MULTIDIMENSIONAL MODEL (CUBE)

The multidimensional model is useful to understand interactive data analysis, and how to improve the execution performance.
The OLAP client interacts with a local DOLAP system (Desktop OLAP) which manages small amount of data extracted from the OLAP server, the Data server or an operational DBMS. It a good choice for those who travel and move extensively, by using portable computers.

DEMO WITH Microsoft Power Pivot (Add-in of Excel)

Power BI - Overview and Learning

Microsoft Power BI is a collection of online services and features that enables you to find and visualize data, share discoveries, and collaborate in intuitive new ways. There are two experiences now available for Power BI: the new experience, generally referred to as Power BI, and the previous experience which is referred to as Power BI for Office 365.