OLAP QUERIES
Online Analytic Processing
OLAP
OLAP

- **OLAP**: Online Analytic Processing

- **OLAP queries are complex queries that**
  - Touch large amounts of data
  - Discover patterns and trends in the data
  - Typically expensive queries that take long time
  - Also called decision-support queries

- **In contrast to OLAP:**
  - **OLTP**: Online Transaction Processing
  - OLTP queries are simple queries, e.g., over banking or airline systems
  - OLTP queries touch small amount of data for fast transactions

```sql
Select salary
From Emp
Where ID = 100;
```
OLTP vs. OLAP

- **On-Line Transaction Processing (OLTP):**
  - technology used to perform updates on operational or transactional systems (e.g., point of sale systems)

- **On-Line Analytical Processing (OLAP):**
  - technology used to perform complex analysis of the data in a data warehouse

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 dimensionality of the enterprise as understood by the user.

[source: OLAP Council: www.olapcouncil.org]
OLAP AND DATA WAREHOUSE
Typically, OLAP queries are executed over a separate copy of the working data
- Over data warehouse

Data warehouse is periodically updated, e.g., overnight
- OLAP queries tolerate such out-of-date gaps

Why run OLAP queries over data warehouse??
- Warehouse collects and combines data from multiple sources
- Warehouse may organize the data in certain formats to support OLAP queries
- OLAP queries are complex and touch large amounts of data
  - They may lock the database for long periods of time
  - Negatively affects all other OLTP transactions
OLAP ARCHITECTURE
EXAMPLE OLAP APPLICATIONS

• **Market Analysis**
  • Find which items are frequently sold over the summer but not over winter?

• **Credit Card Companies**
  • Given a new applicant, does (s)he a credit-worthy?
  • Need to check other similar applicants (age, gender, income, etc…) and observe how they perform, then do prediction for new applicant

OLAP queries are also called “decision-support” queries
MULTI-DIMENSIONAL VIEW

- Data is typically viewed as points in multi-dimensional space

Raw data cubes (raw level without aggregation)

Typical OLAP applications have many dimensions
### ANOTHER EXAMPLE

<table>
<thead>
<tr>
<th>gender</th>
<th>age</th>
<th>accident</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>27</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>Male</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td>Male</td>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>Male</td>
<td>39</td>
<td>4</td>
</tr>
<tr>
<td>Male</td>
<td>43</td>
<td>0</td>
</tr>
<tr>
<td>Male</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Male</td>
<td>49</td>
<td>1</td>
</tr>
<tr>
<td>Male</td>
<td>44</td>
<td>2</td>
</tr>
<tr>
<td>Male</td>
<td>43</td>
<td>3</td>
</tr>
<tr>
<td>Male</td>
<td>53</td>
<td>4</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>Female</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>Female</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>Female</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>Female</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Female</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>Female</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>Female</td>
<td>43</td>
<td>1</td>
</tr>
</tbody>
</table>
APPROACHES FOR OLAP

• Relational OLAP (ROLAP)

• Multi-dimensional OLAP (MOLAP)

• Hybrid OLAP (HOLAP) = ROLAP + MOLAP
RELATIONAL OLAP: ROLAP

- Data are stored in relational model (tables)
- Special schema called Star Schema
- One relation is the fact table, all the others are dimension tables
CUBE vs. STAR SCHEMA

Dimension tables describe the dimensions

Data inside the cube are the fact records
ROLAP: EXTENSIONS TO DBMS

- Schema design
- Specialized scan, indexing and join techniques
- Handling of aggregate views (querying and materialization)
- Supporting query language extensions beyond SQL
- Complex query processing and optimization
- Data partitioning and parallelism
SLICING & DICING

**Dicing**
- how each dimension in the cube is divided
- Different granularities
- When building the data cube

**Slicing**
- Selecting slices of the data cube to answer the OLAP query
- When answering a query
SLICING & DICING: EXAMPLE 1

**Slicing operation in ROLAP is basically:**
-- Selection conditions on some attributes (WHERE clause) +
-- Group by and aggregation
SLICING & DICING: EXAMPLE 2

shop = Pinewood
product = 4
date = ‘7 Mar 2004’

- count = 5
- value = $215
- discount = $32
- cost = $75
SLICING & DICING: EXAMPLE 3

These subcubes correspond purchase of the same product at one store one all days.

These are some of the subcube in more detail.

The answer to the question is the number of subcubes where count is not equal to 0.
**DRILL-DOWN & ROLL-UP**

- **Drill-down** (Group by Nation)

<table>
<thead>
<tr>
<th>Nation</th>
<th>Sales variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>123%</td>
</tr>
<tr>
<td>Japan</td>
<td>52%</td>
</tr>
<tr>
<td>India</td>
<td>87%</td>
</tr>
<tr>
<td>Singapore</td>
<td>95%</td>
</tr>
</tbody>
</table>

- **Roll-up** (group by Region)

<table>
<thead>
<tr>
<th>Region</th>
<th>Sales variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>105%</td>
</tr>
<tr>
<td>Asia</td>
<td>57%</td>
</tr>
<tr>
<td>Europe</td>
<td>122%</td>
</tr>
<tr>
<td>North America</td>
<td>97%</td>
</tr>
<tr>
<td>Pacific</td>
<td>85%</td>
</tr>
<tr>
<td>South America</td>
<td>163%</td>
</tr>
</tbody>
</table>
SELECT dealer, year, SUM(price) 
FROM (Sales NATURAL JOIN Autos) JOIN Days ON date = day 
WHERE model = 'Gobi' AND 
    color = 'red' AND 
    (year = 2001 OR year = 2002) 
GROUP BY year, dealer; 

SELECT dealer, month, SUM(price) 
FROM (Sales NATURAL JOIN Autos) JOIN Days ON date = day 
WHERE model = 'Gobi' AND color = 'red' 
GROUP BY month, dealer;
MOLAP

- Unlike ROLAP, in MOLAP data are stored in special structures called “Data Cubes” (Array-bases storage)

- Data cubes pre-compute and aggregate the data
  - Possibly several data cubes with different granularities
  - Data cubes are aggregated materialized views over the data

- As long as the data does not change frequently, the overhead of data cubes is manageable

Every day, every item, every city

Every week, every item category, every city
MOLAP: CUBE OPERATOR

Aggregation over the Z axis

Aggregation over the Y axis

Aggregation over the X,Y

Aggregation over the X axis

Raw-data (fact table)
MOLAP & ROLAP

• Commercial offerings of both types are available

• In general, **MOLAP** is good for smaller warehouses and is optimized for canned queries

• In general, **ROLAP** is more flexible and leverages relational technology

• **ROLAP** May pay a performance penalty to realize flexibility
## OLTP vs. OLAP

<table>
<thead>
<tr>
<th>User</th>
<th>OLTP</th>
<th>OLAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>Clerk, IT Professional</td>
<td>Knowledge worker</td>
</tr>
<tr>
<td></td>
<td>Day to day operations</td>
<td>Decision support</td>
</tr>
<tr>
<td></td>
<td>Application-oriented (E-R based)</td>
<td>Subject-oriented (Star, snowflake)</td>
</tr>
<tr>
<td>Data</td>
<td>Current, Isolated</td>
<td>Historical, Consolidated</td>
</tr>
<tr>
<td>View</td>
<td>Detailed, Flat relational</td>
<td>Summarized, Multidimensional</td>
</tr>
<tr>
<td>Usage</td>
<td>Structured, Repetitive</td>
<td>Ad hoc</td>
</tr>
<tr>
<td>Unit of work</td>
<td>Short, Simple transaction</td>
<td>Complex query</td>
</tr>
<tr>
<td>Access</td>
<td>Read/write</td>
<td>Read Mostly</td>
</tr>
<tr>
<td>Operations</td>
<td>Index/hash on prim. Key</td>
<td>Lots of Scans</td>
</tr>
<tr>
<td># Records accessed</td>
<td>Tens</td>
<td>Millions</td>
</tr>
<tr>
<td>#Users</td>
<td>Thousands</td>
<td>Hundreds</td>
</tr>
<tr>
<td>Db size</td>
<td>100 MB-GB</td>
<td>100GB-TB</td>
</tr>
<tr>
<td>Metric</td>
<td>Trans. throughput</td>
<td>Query throughput, response</td>
</tr>
</tbody>
</table>

Source: Datta, GT
OLAP: SUMMARY

• OLAP stands for Online Analytic Processing and used in decision support systems
  • Usually runs on data warehouse

• In contrast to OLTP, OLAP queries are complex, touch large amounts of data, try to discover patterns or trends in the data

• **OLAP Models**
  • **Relational (ROLAP):** uses relational star schema
  • **Multidimensional (MOLAP):** uses data cubes