Decision Support Systems
aka
Analytical Systems
Decision Support Systems

• Systems that are used to transform data into information, to manage the organization: OLAP vs OLTP
## OLTP vs OLAP

<table>
<thead>
<tr>
<th></th>
<th>Transactions</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objectives</strong></td>
<td>Operations</td>
<td>Decisions</td>
</tr>
<tr>
<td><strong>Users</strong></td>
<td>Operatives</td>
<td>Managers and analysts</td>
</tr>
<tr>
<td><strong>Detail</strong></td>
<td>Analytic</td>
<td>May be aggregated</td>
</tr>
<tr>
<td><strong>Data origin</strong></td>
<td>Internal</td>
<td>Internal and external</td>
</tr>
<tr>
<td><strong>Applications</strong></td>
<td>Known a priori</td>
<td>Ad hoc</td>
</tr>
<tr>
<td><strong>Data items</strong></td>
<td>Few (tens)</td>
<td>Many (millions)</td>
</tr>
<tr>
<td><strong>Per operation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Updates</strong></td>
<td>Frequent, Small</td>
<td>Rare, Mostly massive</td>
</tr>
<tr>
<td><strong>State of data</strong></td>
<td>Current data</td>
<td>Historic</td>
</tr>
<tr>
<td><strong>Stored in</strong></td>
<td>Data Base</td>
<td>Data Warehouse / DSS</td>
</tr>
</tbody>
</table>
Operational systems vs DSS

• Operational systems:
  – Data are organized in a DB.
  – Data are managed by a traditional DBMS.
  – Applications used to perform structured operational activities

• Decision Support Systems:
  – Data are organized in a separate specialized DB (Data Warehouse (DW)).
  – Data are managed by a specialized DBMS.
  – Business Intelligence applications used to analyze data
Data Warehouse

- The first definition of data warehouse was provided by William Inmon in 1990.
- A DW is a specialized database
  - static (non volatile),
  - with integrated data from different data sources
  - organized to analyze subjects of interest,
  - with historical data,
  - used to produce summarized data to support decision-making processes
Building the data warehouse
DW design

• The creation of a data warehouse takes place gradually, at different levels of abstraction: a **conceptual model**, a **logical model** and **physical model**.
Data models for DW

• The Dimensional Fact Model (DFM) is a graphical conceptual model.
• The Relational Data Model, as a logical model
• The Multidimensional Model (called Cube), useful to understand OLAP operations
Design of a Data Mart

**DATA BASE**

- **Customer**
  - SSN
  - Name
  - Address
  - City
  - Phone

- **Payment Method**
  - Description

- **Invoice**
  - Date

- **Order**
  - Number
  - Date

- **OrderLine**
  - LineNo
  - QuantityOrdered

- **Employee**
  - SSN
  - Name
  - Qualification
  - OfficePhone

- **Shipper**
  - Name

- **Product**
  - Name
  - Category
  - UnitPrice
  - Quantity

**BUSINESS QUESTIONS**

- Number of items ordered, **by product**, **by customer**, **by month**

- Total revenue **by product category**, **by customer**, **by year**

- Total revenue from Italian customers **by customer city**, **by year**, **by quarter**
A data model for conceptual design

• Object data model:
  – Classes with
    • Attributes

• Data warehouse data model
  – Facts tables with
    • Measures
    • Dimensions
The facts

• Facts are typically transactions / events
  – Sales, clicks, complains, visits
• Periodic facts
  – One fact for a group of transactions made over a period of time.
  – Example: the monthly balance of all monthly transactions
• Accumulating fact
  – One fact for the entire lifetime of an evolving event that has a duration and change over time
  – Example: the life cycle of a mortgage application
The measures

• Measures: quantitative information whose aggregation is of interest
• Aggregation is often Sum, but not always (count, average...)

![Diagram of database model]

DW Design
Dimensions

• Dimensions relate to who, what, why, when, and where:
  – Who is involved? What is about? When happens? Where occurs?

• Dimensions: the variables that influence the measures and indicate possible intervention levers

• How does a measure depend on a dimension?
Dimensions have attributes and hierarchies

Without hierarchies

With hierarchies
Case study

An hospital needs a DM to extract information from their operational database with information about inpatients treatments.

1. Total billed amount for hospitalizations, by diagnosis code and description, by month (year).
2. Total number of hospitalizations and billed amount, by ward, by patient gender (age at date of admission, city, region).
3. Total billed amount, average length of stay and average waiting time, by diagnosis code and description, by name (specialization) of the physician who has admitted the patient.
4. Total billed amount, and average waiting time of admission, by patient age (region), by treatment code (description).
## Requirements specification

<table>
<thead>
<tr>
<th>Requirements analysis</th>
<th>Dimensions</th>
<th>Measures</th>
<th>Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total billed amount for hospitalizations, by diagnosis code and description, by month (year).</td>
<td>Diagnosis (ICD code, Description), Date (Month, Year)</td>
<td>Amount</td>
<td>Total Amount</td>
</tr>
<tr>
<td>Total number of hospitalizations and billed amount, by ward, by patient gender (age at date of admission, city, region)</td>
<td>Ward, Patient (Gender, Age, City, Region)</td>
<td>Amount</td>
<td>Total number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total Amount</td>
<td></td>
</tr>
<tr>
<td>Total billed amount, average length of stay and average waiting time by diagnosis code and description, by name (specialization) of the physician who admitted the patient.</td>
<td>Diagnosis (ICD code, Description), Physician (Name, Specialization)</td>
<td>Amount, Duration, WaitingTime</td>
<td>Total Amount Average Duration Average WaitingTime</td>
</tr>
<tr>
<td>Total billed amount, and average waiting time for admission by patient age (region), by treatment code (description).</td>
<td>Patient (Age, Region), Treatment (Code, Description)</td>
<td>Amount, Duration, WaitingTime</td>
<td>Total Amount Average WaitingTime</td>
</tr>
</tbody>
</table>
# Requirements specification

<table>
<thead>
<tr>
<th>Description</th>
<th>A fact is a hospitalization of a patient, assuming that they may require one treatment only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminary dimensions</td>
<td>Patient, Date, Ward, Diagnosis, Treatment, Physician</td>
</tr>
<tr>
<td>Preliminary measures</td>
<td>Duration, WaitingTime, Amount</td>
</tr>
</tbody>
</table>
Multidimensional model (cube)

- The cube model helps understanding interactive data analysis.
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

<table>
<thead>
<tr>
<th>ProductId</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P1</td>
<td>P2</td>
<td></td>
</tr>
<tr>
<td>StoreId</td>
<td>S1</td>
<td>S2</td>
<td>S3</td>
</tr>
<tr>
<td>-----------</td>
<td>---</td>
<td>---</td>
<td>---</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>
</tbody>
</table>

Cross tabulation
### 3-D Cube

#### Fact Table

<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>

#### 3-D Cube
Cube operator: Slice

• Sales **SLICE FOR** Dateld = ‘D1’;
Cube operator: Dice

- Sales DICE FOR Dateld = ‘D1’
  Storeld IN (‘S1’, ‘S2’);
Cube operator: Pivot

- **PIVOT** *(Sales SLICE FOR DateId = 'D1');*
Cube operators: Roll-up

Data cube
Cube operators: Roll-up and Drill-down

- Roll-up aggregates data by dimension reduction or by navigating dimension hierarchy
- Drill-down: the reverse
- SALES ROLL-UP ON DateId

Data cube
Extended cube

- Extend each dimension with *, mapped to the sum
Extended cube

- The cube is now a set of cuboids
- **white** cells are the data cube
- **gray** cells are rolled-up by a dimension
- **dark gray** cells are rolled-up by two dimensions
- **black** cells are rolled-up by all dimensions.
DW Lattice: a lattice of cuboids

• On the set of cuboids is defined the following partial order relation:
  – \( C_1 \leq C_2 \) if \( C_1 \) can be computed by \( C_2 \)
Cuboid materialization

• If the materialization is partial, which cuboids do we store?
Relational representation

- Relational OLAP systems are relational DBMS extended with specific features to support business intelligence analysis.
- A DW is represented with a special kind of relational schema:
  - star schema
  - snowflake schema
  - constellation schema
Star schema

• A fact table with a foreign key for each dimension
• Each dimension has one table with an artificial key
Snowflake schema

- Dimension tables are ‘normalized’
- Not very popular
Data Warehouse Management System

• Interface: what to change?
• Engine: what to change?