Armazéns de Dados

Departamento de Engenharia Informática (DEI/ISEP)
Paulo Oliveira
pjo@isep.ipp.pt

Adaptado do Original de:
Fátima Rodrigues (DEI/ISEP)

1

Bibliography

- OLAP Solutions
  Building Multidimensional Information Systems (2nd Ed.)
  Erik Thomsen
  John Wiley and Sons, 2002

2

On-Line Analytical Processing (OLAP)

3

On-Line Analytical Processing

- Defined by **E.F. Codd**
  - White paper "Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate", 1993
  - OLAP is "the dynamic synthesis, analysis, and consolidation of large volumes of **multidimensional** data."
  - Codd proposed an arrangement of data in arrays to allow fast analysis – **cubes**
  - **OLAP cubes** can produce an answer in around 0.1% of the time for the same query on a OLTP relational database
- **Complements** OLTP and DW systems

4

4

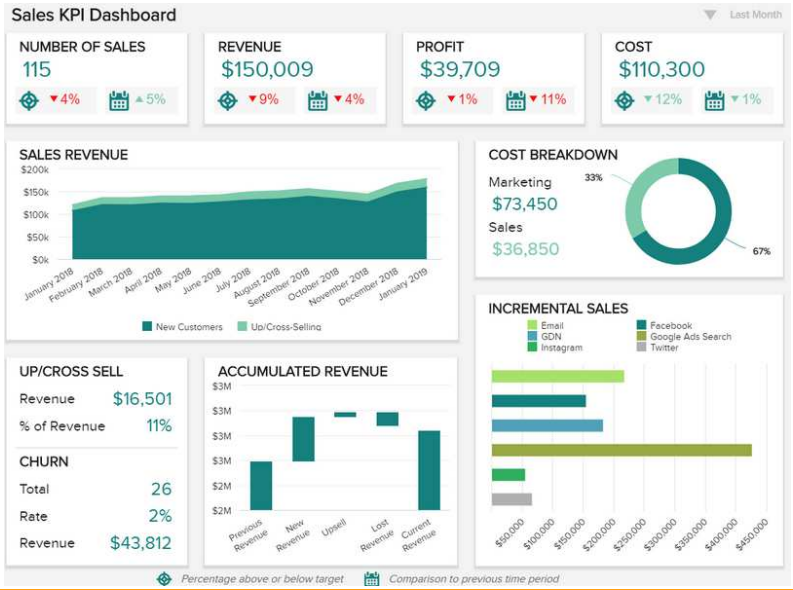
OLAP Analytical Goals

- Before developing OLAP applications it is crucial to identify the business analytical goals – **Key Performance Indicators (KPIs)**
- **KPIs are financial and non-financial metrics** used to provide business users with an **indication of the current and historical performance** of different aspects of the business
- **KPIs** examples:
  - Profit
  - Loss
  - Gross margin
  - Customer attrition/churn
  - …

5

5

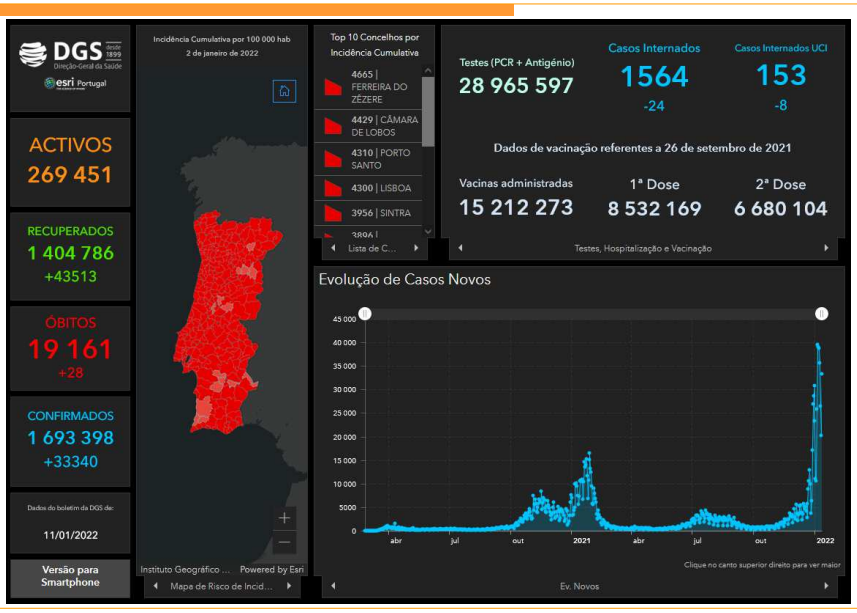
KPI Dashboard Example



6

6

Another KPI Dashboard Example

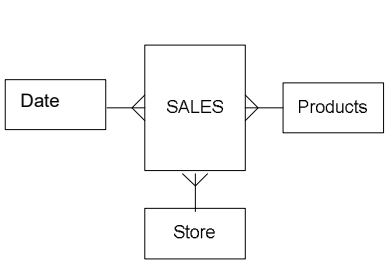


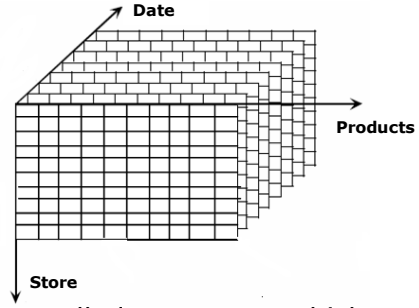
On-Line Analytical Processing

- Approach to quickly provide the answer to **analytical queries** that are multidimensional in nature
  - Usually the data comes from the DW
- **Analytical queries** involve inter-row calculations, time series analysis, and access to aggregated historical and atomic/ aggregated current data
- Databases configured for OLAP employ **a multidimensional data model**, allowing analytical ad-hoc queries with fast execution time
- **Output of an OLAP query** is typically displayed in a **matrix** (or pivot) format:
  - **Dimensions** – rows and columns of the matrix
  - **Measures** (facts) – values of each cell

OLAP Functionality

- In the core of any OLAP system is the concept of an **OLAP cube** also called a ***multidimensional cube*** or a ***hypercube***





- ***Hypercube*** consists of numeric facts called ***measures*** which are categorized by ***dimensions***
- Cube is typically created from a star schema of tables in a relational database
  - **Measures** are derived from the records in the **fact table**
  - **Dimensions** are derived from the records in the **dimension tables**

9

9

OLAP Applications

- **Multidimensional** data structure
- Automatic **roll-ups** and **drill-downs**
  - No programming required
- Support **complex calculations**
  - Sales forecasting, moving averages, growth percentage, etc.
- **Time intelligence**
  - Comparison of performance between different time periods

10

10

Kinds of On-Line Analytical Processing

- **MOLAP (Multidimensional OLAP)**
  - "Classic" form of OLAP, sometimes referred to as just OLAP
  - Data is stored in a special, usually proprietary, data structure
- **ROLAP (Relational OLAP)**
  - Works directly with relational databases
  - Tools access the DW directly
- **HOLAP (Hybrid OLAP)**
  - Takes synergies from ROLAP and MOLAP
  - Divides data between relational and specialized storage
    - Uses relational tables to hold larger quantities of detailed data
    - Uses specialized storage for smaller quantities of more-aggregate or less-detailed data

Multidimensional OLAP

- Stores data into an **optimized multidimensional array storage**, usually a three or higher dimensional array (**hipercube**)
- **Requires the pre-computation and storage of data in the hipercube** for fast end-user response
  - Operation known as ***processing***
    - Expensive
    - Makes loading slower
- Examples: Cognos Powerplay; Hyperion ESSbase; SQL Server Analysis Services

MOLAP Advantages

- Delivers **better performance** due to specialized indexing and storage optimizations
  - Array model provides **natural indexing** – easy and quick to access an array's position
- **Automated computation of higher-level aggregates** of data
- Does not need to have a **permanent connection** to the underlying DW

13

13

MOLAP Disadvantages

- **Data loading** into these structures can be **lengthy**, especially of large data volumes
  - Must be **periodically updated** to remain current
  - ↪ **Solution**: **incremental processing** – only the data which is new or has been updated, instead of reprocessing the entire hypercube
- Stores a copy of the data at OLAP server and so **requires additional storage**

14

14

Relational OLAP

- **Does not require the pre-computation of data** (as in MOLAP)
  - Creates multidimensional views from a relational database
- ROLAP tools **access the data in a relational database** and generate SQL queries to retrieve data when an end-user requests it
- **Relational database** must be carefully **designed for ROLAP use** – a database which was designed for OLTP will not function well as a ROLAP database
  - Data must be stored in a **multidimensional data model**
  - **It is possible to create additional database tables** (*aggregations*) which summarize the data at any desired combination of dimensions
- Examples: SAP BW; MicroStrategy; SQL Server Analysis Services

15

ROLAP Advantages

- Uses an **underlying RDBMS**, rather than a specialized multidimensional data structure
  - **Better scalability** since is able to handle very high volumes of data
  - **Maturity** of the RDBMS technology
- **Frequent updates** are not a problem
- It has **less storage requirements**, since it does not make a copy of the data (as in MOLAP)

16

ROLAP Disadvantages

- As the data structure is relational, **SQL must be used to access the detailed records**
  - ROLAP engine must perform additional work
  - SQL generation may be costly
- **Slower data retrieval** than with MOLAP tools

17

17

Hybrid OLAP

- **Combination** of MOLAP and ROLAP that mixes the **best** of both
  - Can pre-process quickly and scale well
- Combines the **performance and functionality of the multidimensional database** with the **ability to access directly detailed data on relational databases**
  - Stores the **aggregations** that exist within a **multidimensional** schema, leaving the **most detailed data** in **relational** form
- Most commercial OLAP tools **now support** a hybrid OLAP approach
- Examples: Holos; Oracle Advanced Analytic Services; SQL Server Analysis Services

18

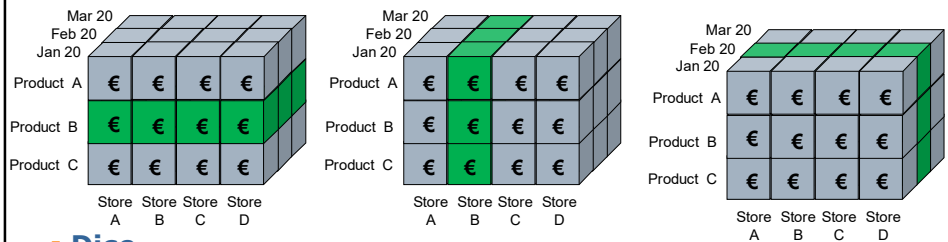
18

OLAP Basic Operations

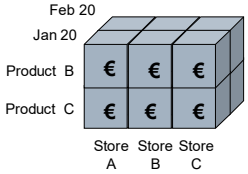
19

Slice & Dice

- **Slice**
  - Subset of multidimensional data
  - Defined through the selection of specific dimension attribute values



- **Dice**
  - Extracts a sub-cube from the original cube
  - Can be combined with slicing

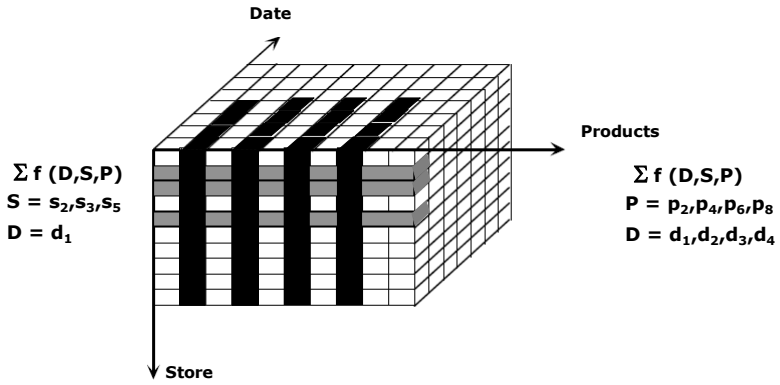


20

20

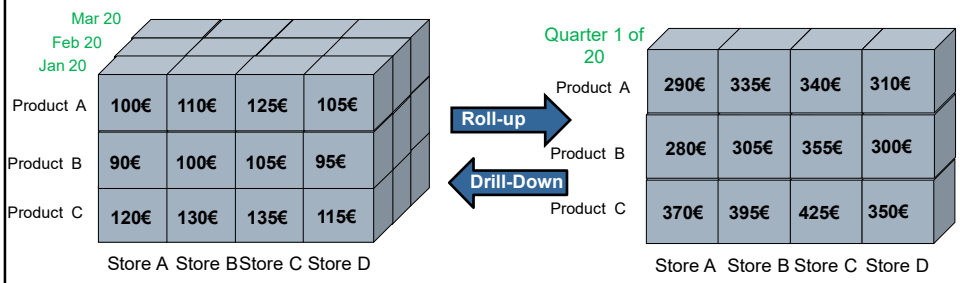
Aggregations

- Built by **restricting the values of certain dimension attributes** and **aggregating data along them**



Drill-Down & Roll-Up

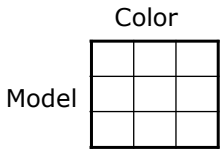
- **Drill-down** means to move **from summary information to more detailed data** by focusing on something
- **Roll-up** means to move **from more detailed data to summary information** by abstracting of something



Pivot Tables

Data Table		
Model	**Color**	**Sales**
Van	Blue	6
Van	Red	5
Van	White	4
Coupe	Blue	3
Coupe	Red	5
Coupe	White	5
Sedan	Blue	4
Sedan	Red	3
Sedan	White	2

Pivot Tables are especially well-suited for **taking enormous amounts of data and summarizing that data into useful reports**



Pivot Table 3×3

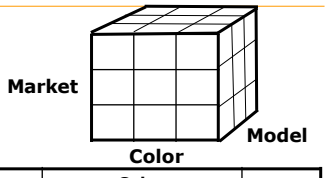
Pivot Table				
Model/Color	**Blue**	**Red**	**White**	**Total**
Van	6	5	4	**15**
Coupe	3	5	5	**13**
Sedan	4	3	2	**9**
Total	**13**	**13**	**11**	**37**

23

23

One More Dimension ...

Data Table			
Market	**Model**	**Color**	**Sales**
Europe	Van	Blue	6
America	Van	Red	5
Europe	Van	White	4
Europe	Coupe	Blue	3
America	Coupe	Red	5
Asia	Van	Red	3
America	Coupe	White	5
America	Van	Blue	3
Europe	Coupe	Red	5
Asia	Coupe	White	5
Asia	Sedan	Blue	4
America	Sedan	Red	3
Europe	Sedan	White	2



Pivot Table 3×3×3

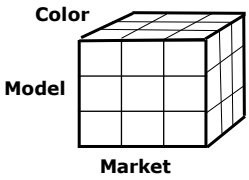
		Color			
Market	**Model**	**Blue**	**Red**	**white**	**Total**
America	Coupe	0	5	5	**10**
	Sedan	0	3	0	**3**
	Van	3	5	0	**8**
	Total	**3**	**13**	**5**	**21**
Europe	Coupe	3	5	0	**8**
	Sedan	0	0	2	**2**
	Van	6	0	4	**10**
	Total	**9**	**5**	**6**	**20**
Asia	Coupe	0	0	5	**5**
	Sedan	4	0	0	**4**
	Van	0	3	0	**3**
	Total	**4**	**3**	**5**	**12**
Total		**16**	**21**	**16**	**53**

24

24

Rotate Data Cube

- Allow the user to **view the same data from a different perspective**



		Market			
Model	**Color**	**America**	**Europe**	**Asia**	**Total**
Coupe	Blue	0	3	0	**3**
	Red	5	5	0	**10**
	White	5	0	5	**10**
	Total	**10**	**8**	**5**	**23**
Sedan	Blue	0	0	4	**4**
	Red	3	0	0	**3**
	White	0	2	0	**2**
	Total	**3**	**2**	**4**	**9**
Van	Blue	3	6	0	**9**
	Red	5	0	3	**8**
	White	0	4	0	**4**
	Total	**8**	**10**	**3**	**21**
Total		**21**	**20**	**12**	**53**

Pivot Table 3×3×3