INTRODUCTION TO BUSINESS INTELLIGENCE

Lecture 3

Agenda

OLAP technology as an example of an exploratory and analytical system
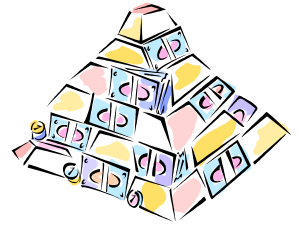
Data mining systems
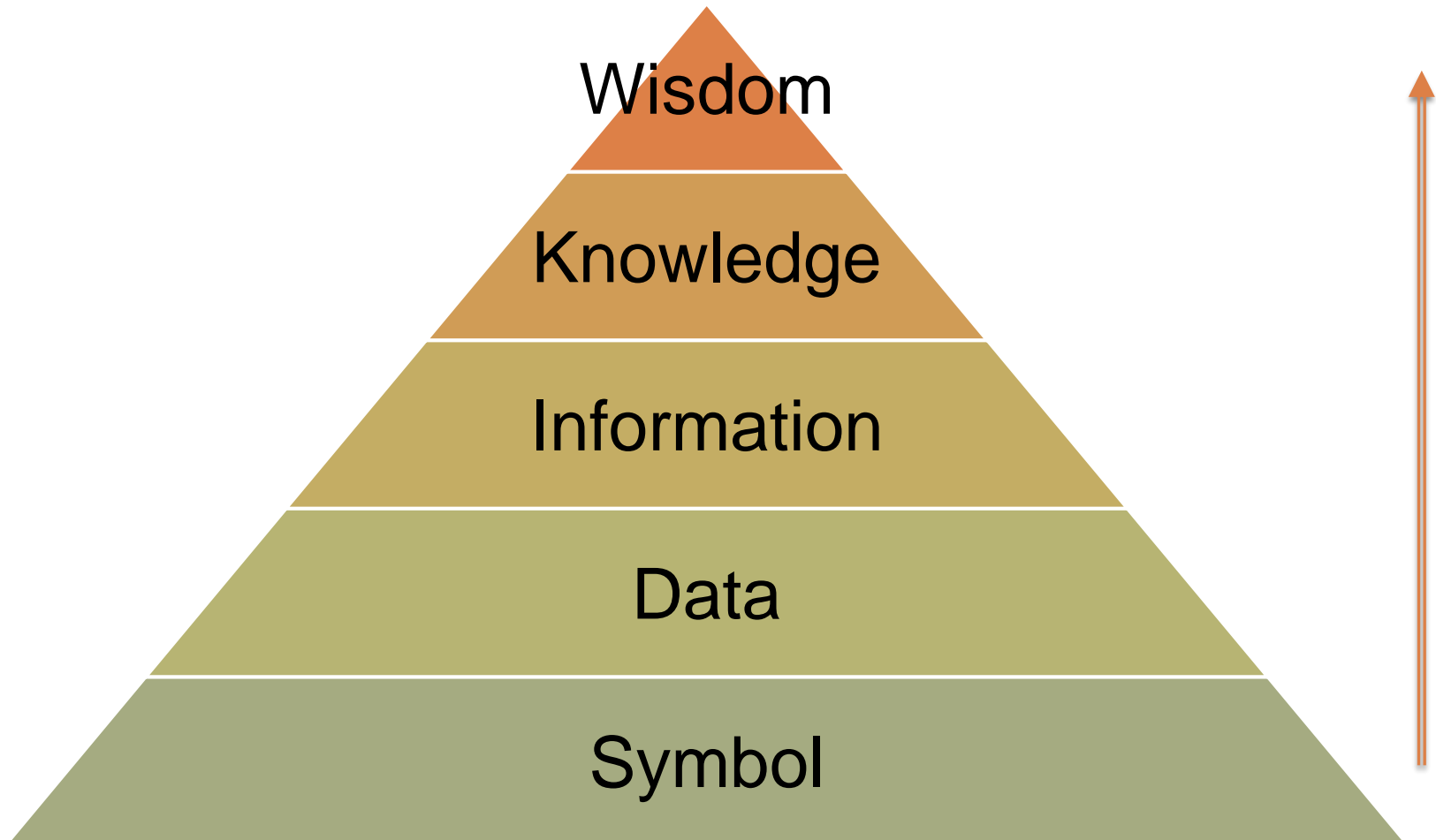
New data collection technologies

University of Gdańsk

3 OLAP

Information pyramid

Wisdom

Knowledge

Information

Data

Symbol

University of Gdańsk

Decision pyramid

Decision

DSS/BI/OLAP

Transaction

OLTP

Range of input and output variables

University of Gdańsk
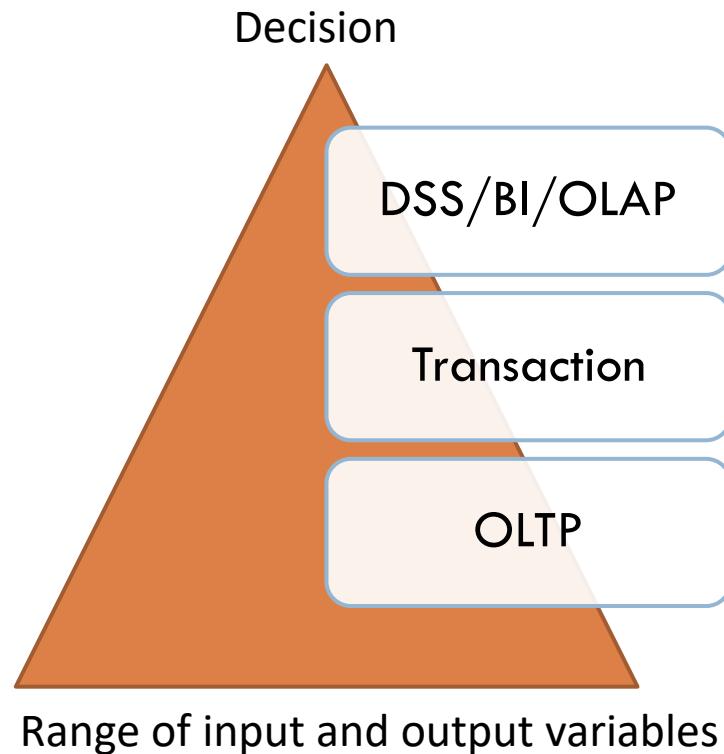
OLAP (On-Line Analytical Processing)

The functionality of performing analyzes on an ongoing basis, based on data contained in multidimensional cubes of a data warehouse

A critical tool of the organization, helping to identify past successes and failures, and based on them, predict future achievements and possible failures

Object-oriented user interfaces where users manipulate objects representing organized groups of data

University of Gdańsk

OLAP systems

These are the tools used to analyze economic information

They can be used by analysts, managers or executives to gain an insight into the functioning of the organization at any time in the past

They are a fast, consistent, interactive tool that provides a wide range of data views

University of Gdańsk

OLAP - Sales cube example

University of Gdańsk
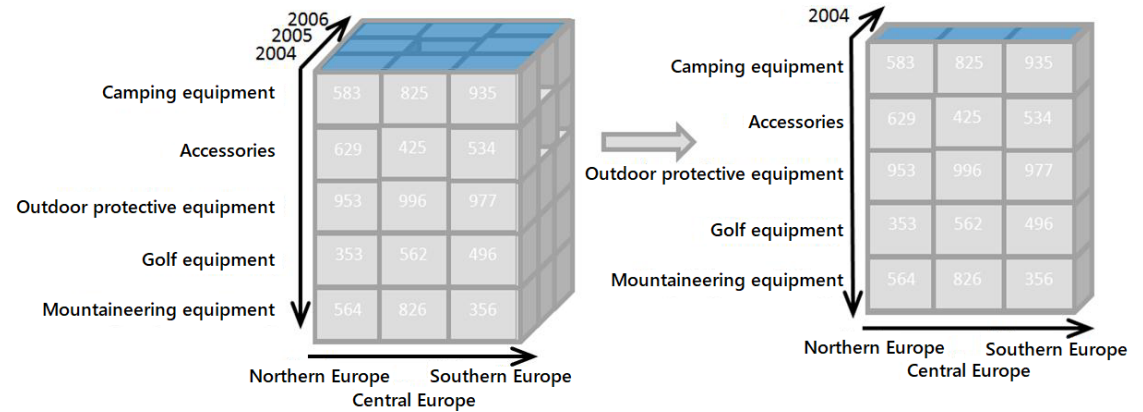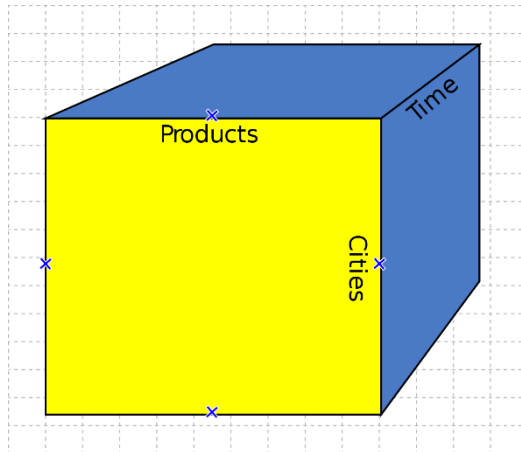
OLAP and OLTP - differences

System	OLTP	OLAP
Goal	Real-time data input	Reading and analysis of historical data
Updates	Yes	No – read only
Data edit	Yes	No
Data sources	Current data	Historical data
Associated database	Operational	Analytical, possible operational

University of Gdańsk

Components of OLAP systems

OLAP systems usually contain two components:

calculate component - used to perform operations such as sums, ratios, time series analysis, descriptive statistics, as well as customized formulas and algorithms, and modeling and forecasts,

multidimensional view component - for viewing data in defined dimensions.

University of Gdańsk
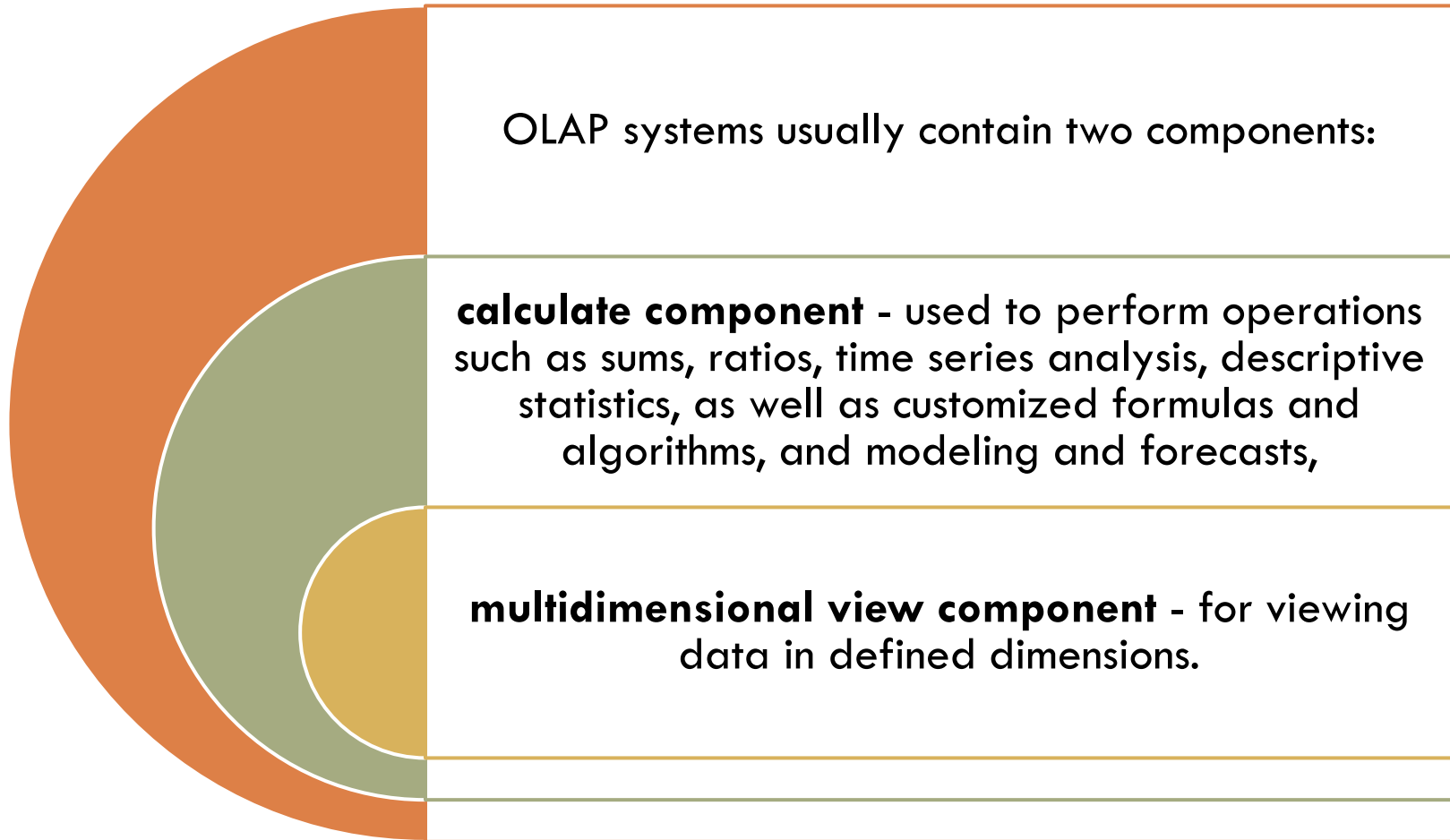
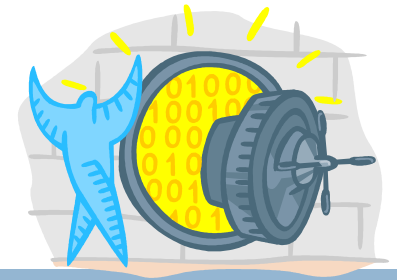# Types of OLAP systems

Types of OLAP:

MOLAP (multidimensional),

ROLAP (relational),

HOLAP (hybrid).

Others are **LOLAP** or **DOLAP** and **WOLAP**.

MOLAP

MOLAP - multidimensional OLAP, using multidimensional database management systems for analytical processing.

They work like a spreadsheet.

Disadvantages:

☐ no scalability of dimensions, which makes it necessary to provide dimensions when designing a warehouse,

☐ no query language standards for these systems.

Advantages:

☐ very high work efficiency.

University of Gdańsk

ROLAP

ROLAP - relational OLAP, supported by a relational database management system.

Disadvantages:

☐ Low efficiency.

☐ There is a metadata server between the database and the analytical system.

Advantages:

☐ Since they are based on relational systems, the scalability problem does not arise here.

☐ The query language has been standardized and is commonly known as SQL.

University of Gdańsk

HOLAP

HOLAP - hybrid OLAP, uses the MOLAP architecture to place and view data in aggregates.
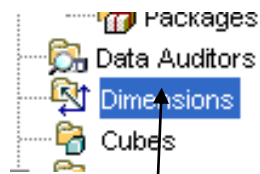
More detailed data is stored in a relational database.

Characteristics:

The system performance is higher than that of ROLAP systems, but still less than that of MOLAP systems.

OLAP – Oracle database

Packages
Data Auditors
Dimensions
Cubes

Create Dimension - Step 4 of 8: Levels

Specify the levels in the default hierarchy:

	Name	Description
1	POZIOM_1	Domyslny
2	POZIOM_2	Dla szczegolowej analizy

Creating dimension

Hierarchy

Create Dimension - Step 2 of 8: Storage Type

The storage type for a dimension is driven by the warehouse implementation. Choose ROLAP if you have high volumes of data combined with high refresh rates or if you have detailed high volume data. Choose MOLAP to store aggregated data that is used for analysis.

Storage type:
- ROLAP: Relational storage
- MOLAP: Multidimensional storage

Storage technology

University of Gdańsk

OLAP – Microsoft database

University of Gdańsk

Operational versus analytical activities
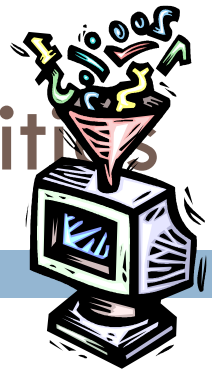
| |
|---|
| Operational activities vs. **Decision-making and analytical activities** |
| Often performed vs. **Less frequent** |
| More predictable vs. **Less predictable** |
| Smaller amounts of data in the query vs. **Larger amounts of data in the query** |
| Query on raw data vs. **Query on transformed data** |
| Requires real-time data vs. **Requires current data and historical** |
| Low complexity of data sources vs. **High complexity of data sources** |

University of Gdańsk

Multidimensional and relational data store
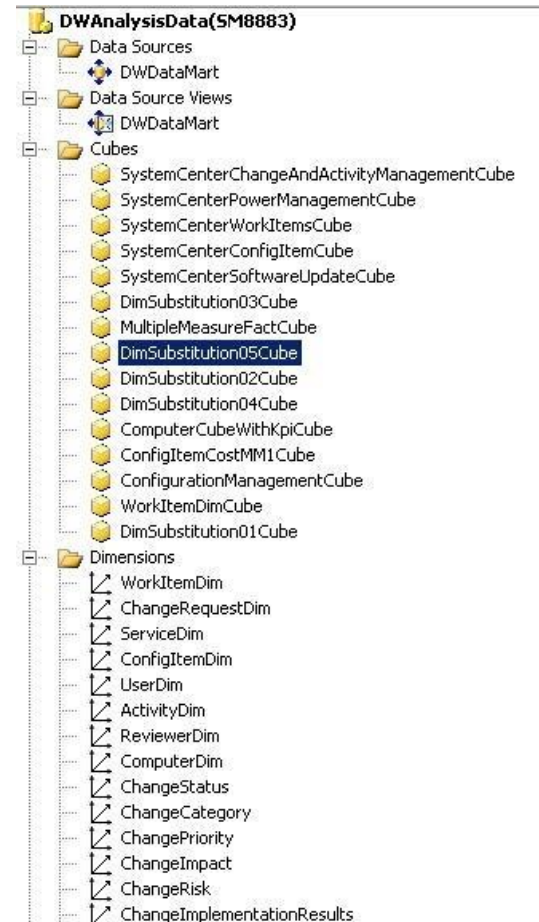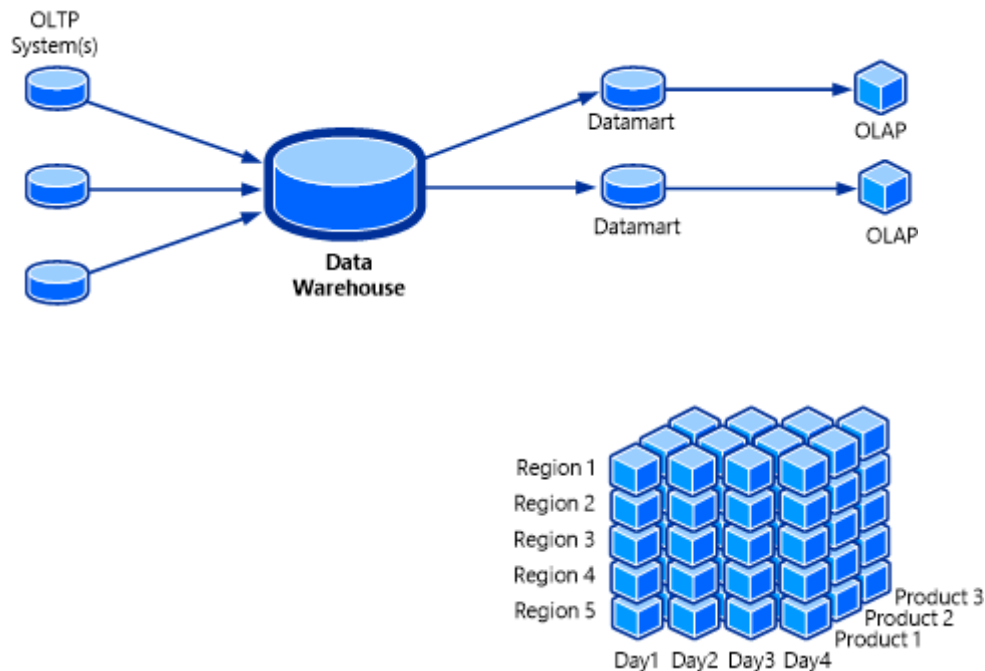
Source: docs.microsoft.com

University of Gdańsk

19 Data exploration
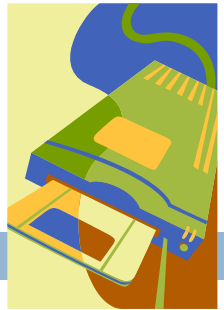
Data Mining

Text Mining

Duo Mining

Data Mining

Data Mining allows for automatic data analysis.

Data mining is about efficiently finding previously unknown dependencies and relationships between data.

University of Gdańsk

Comparison: reporting, OLAP, Data Mining

OLTP	OLAP	Data Mining
Extraction of detailed and summary data	Summaries, trends and forecasts	Discovering knowledge of hidden patterns and insight
„information"	„analysis"	„ insight and prediction"
Who has been buying from the fund in the past three years?	What is the fund's average revenue from buyers in a given region and year?	Who will buy from the fund in the next six months and why?

University of Gdańsk

Data Mining - functions
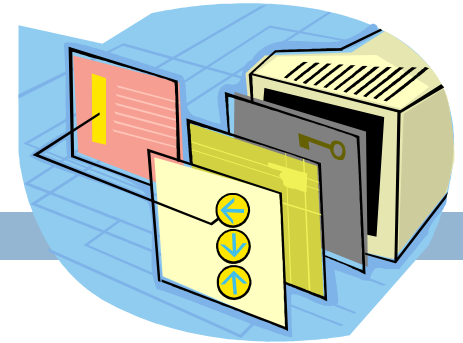
Two basic functions:

☐ **verification of hypotheses** when there is some idea about the significant relationship between data elements,
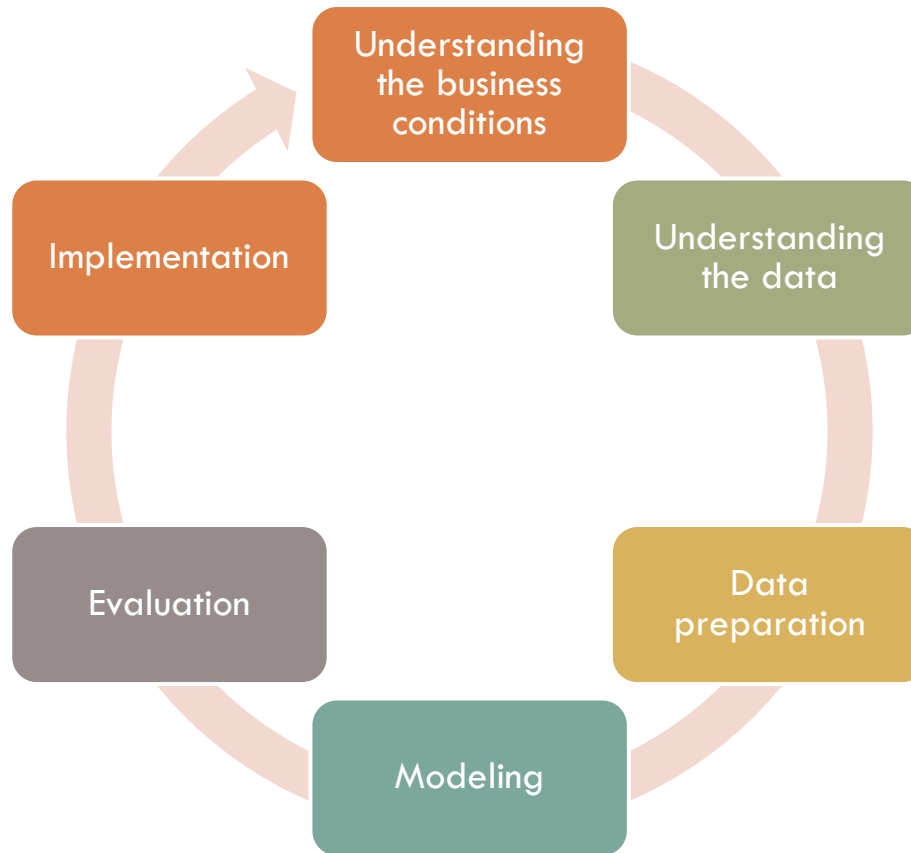
☐ **knowledge discovery** where there may be hitherto unknown but meaningful relationships between data items that are difficult to deduce.

Data Mining process phases

```
                    Understanding
                     the business
                      conditions

   Implementation                        Understanding
                                           the data

      Evaluation                          Data
                                        preparation

                        Modeling
```

University of Gdańsk

Data mining method classes
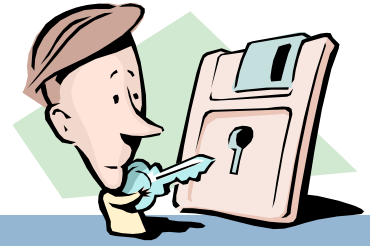
association discovery – a class of methods that allows to discover unknown dependencies and associations between methods

clustering – these methods are designed to find a finite set of object classes in the database with similar characteristics

discovering sequence patterns – discovering certain patterns of behavior

discovering classification – finding the relationship between the classification of objects and their characteristics

discovering similarities in time courses – finding similarities in time series describing specific processes

detection of changes and anomalies – finding differences between current and the expected data values

University of Gdańsk

Text Mining

Text Mining is a method that allows to perform automatic analysis of the content of text documents

It usually works by summarizing information, keywords, and analyzing the content of the document

Web Mining is a subclass of Text Mining, which focuses on analyzing the content of web pages
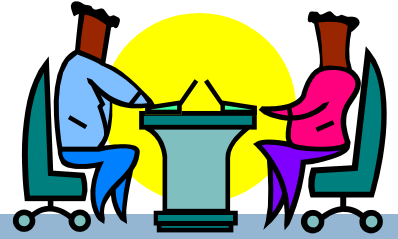
University of Gdańsk

Text Mining

Text Mining searches for patterns in unstructured data

It consistently uses techniques such as semantic analysis and artificial intelligence

The essence of Text Mining is automatic pattern discovery and thus learning and automatic creation of a document description

University of Gdańsk

Duo Mining

Goal

- Duo Mining was proposed as an integration of data and text mining tools

Example

- When analyzing the creditworthiness of a given customer, both operations can be analyzed within the customer's bank account (structured data) as well as any documents provided by the customer (unstructured data)

University of Gdańsk

28 Emerging technologies of data storage

University of Gdańsk

NoSQL

- ☐ Variety of types (column data storage, key-value data storage, graph storage, document storage)
- ☐ Unstructured data storage
- ☐ Are different types, the most common is JSON-like documents
- ☐ Fields instead of columns
- ☐ Collections instead of tables

- ☐ Unstructured data storage was also proposed by W. Inmon in Data Warehouse 2.0

University of Gdańsk

Question

Which of the following are types of OLAP systems:

- ☐ ROLAP
- ☐ XOLAP
- ☐ HOLAP
- ☐ BOLAP

University of Gdańsk