



# Knowledge Discovery and Data Mining

Data Mining vs. OLAP

# Acknowledgement

- All the material in this presentation is taken from the Internet.
- A simple search of “Data Mining vs. OLAP” on Google would provide link to all the original sources.



# DM vs. OLAP

- OLAP and data mining are used to solve different kinds of analytic problems:
  - OLAP provides summary data and generates rich calculations. For example, OLAP answers questions like "How do sales of mutual funds in North America for this quarter compare with sales a year ago? What can we predict for sales next quarter? What is the trend as measured by percent change?"
  - Data mining discovers hidden patterns in data. Data mining operates at a detail level instead of a summary level. Data mining answers questions like "Who is likely to buy a mutual fund in the next six months, and what are the characteristics of these likely buyers?"

## DM vs. OLAP (Cont'd)

- OLAP and data mining can complement each other. For example, OLAP might pinpoint problems with sales of mutual funds in a certain region.
- Data mining could then be used to gain insight about the behavior of individual customers in the region. Finally, after data mining predicts something like a 5% increase in sales, OLAP can be used to track the net income.
- Or, Data Mining might be used to identify the most important attributes concerning sales of mutual funds, and those attributes could be used to design the data model in OLAP.

## DM vs. OLAP (Cont'd)

- Both data mining and OLAP are two of the common Business Intelligence(BI) technologies.
- Business intelligence refers to computer-based methods for identifying and extracting useful information from business data.
- Data mining is the field of computer science which, deals with extracting interesting patterns from large sets of data. It combines many methods from artificial intelligence, statistics and database management.
- OLAP (online analytical processing) as the name suggest is a compilation of ways to query multi-dimensional databases.

## DM vs. OLAP (Cont'd)

- OLAP is a class of systems, which provide answers to multi-dimensional queries.
- Typically OLAP is used for marketing, budgeting, forecasting and similar applications. It goes without saying that the databases used for OLAP are configured for complex and ad-hoc queries with a quick performance in mind.
- Typically a matrix is used to display the output of an OLAP. They often use methods of aggregation on multiple tables to obtain summaries.
- For example, it can be used to find out about the sales of this year in Wal-Mart compared to last year? What is the prediction on the sales in the next quarter? What can be said about the trend by looking at the percentage change?



## DM vs. OLAP (Cont'd)

- Although it is obvious that Data mining and OLAP are similar because they operate on data to gain intelligence, the main difference comes from how they operate on data.
- OLAP tools provides multidimensional data analysis and they provide summaries of the data but contrastingly, data mining focuses on ratios, patterns and influences in the set of data.

## DM vs. OLAP (Cont'd)

- The main difference between OLAP and data mining is how they operate on the data.
- OLAP tools provide multidimensional data analysis-- that is, they allow data to be broken down and summarized (such as by regional sales).
- For example, OLAP typically involves the summation of multiple databases into highly complex tables. OLAP tools deal with aggregates--OLAP technology basically comes down to the operation of data via addition. For example, OLAP can tell you about the total number of widgets sold in all the ZIP codes in the country.



## DM vs. OLAP (Cont'd)

- Data mining, on the other hand, is about ratios, patterns and influences in a data set. As such, data mining is division.
- Data mining can tell you about the factors influencing the sales of the widgets in those ZIP codes.
- This is not to say that both OLAP and data mining should not be used in conjunction to gain a powerful insight into your company databases, customer information file, data marts and data warehouse.