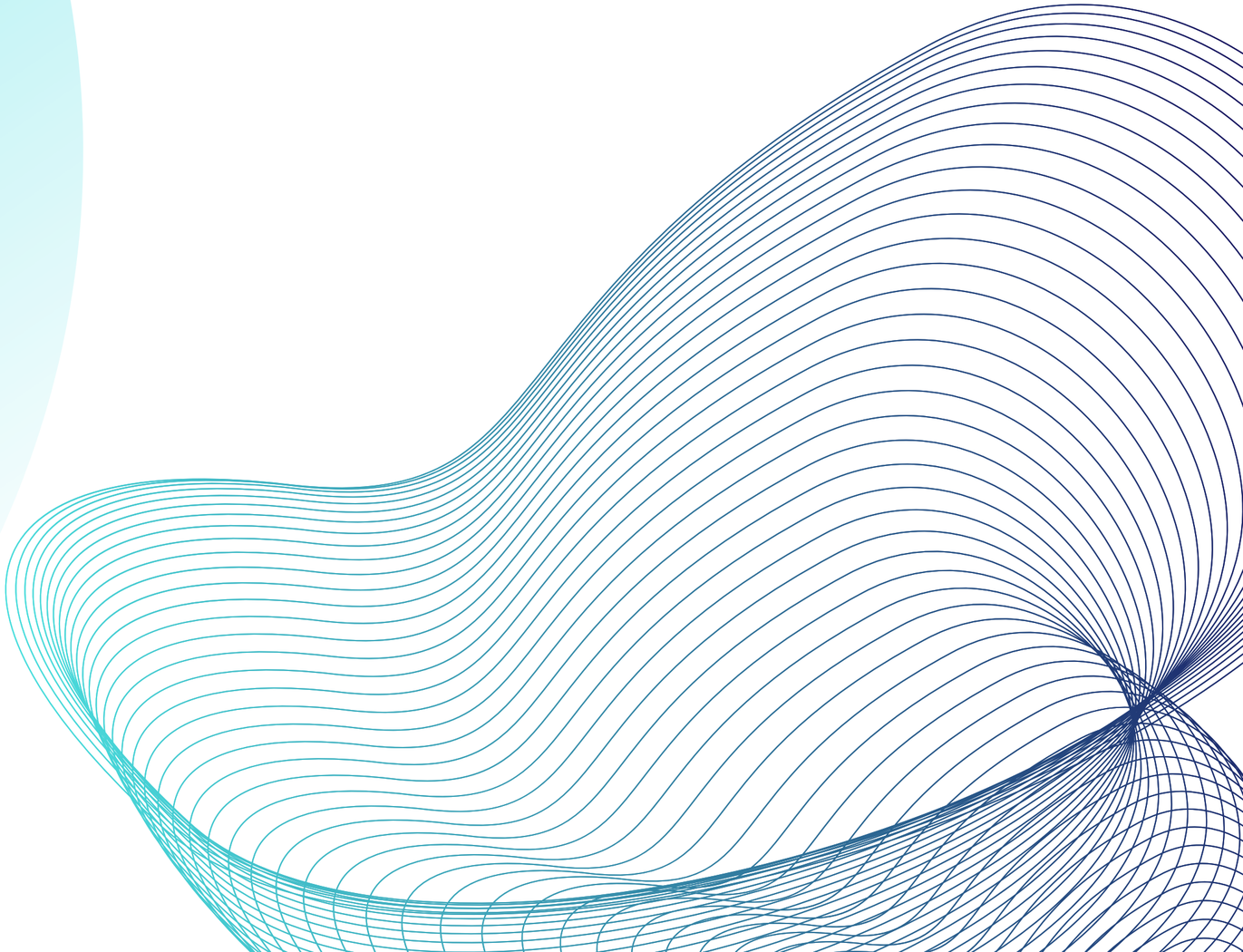
INTRODUCTION TO ARTIFICIAL INTELLIGENCE

PGS. TS. Nguyễn Thanh Bình

DATA-MINING PROCESS

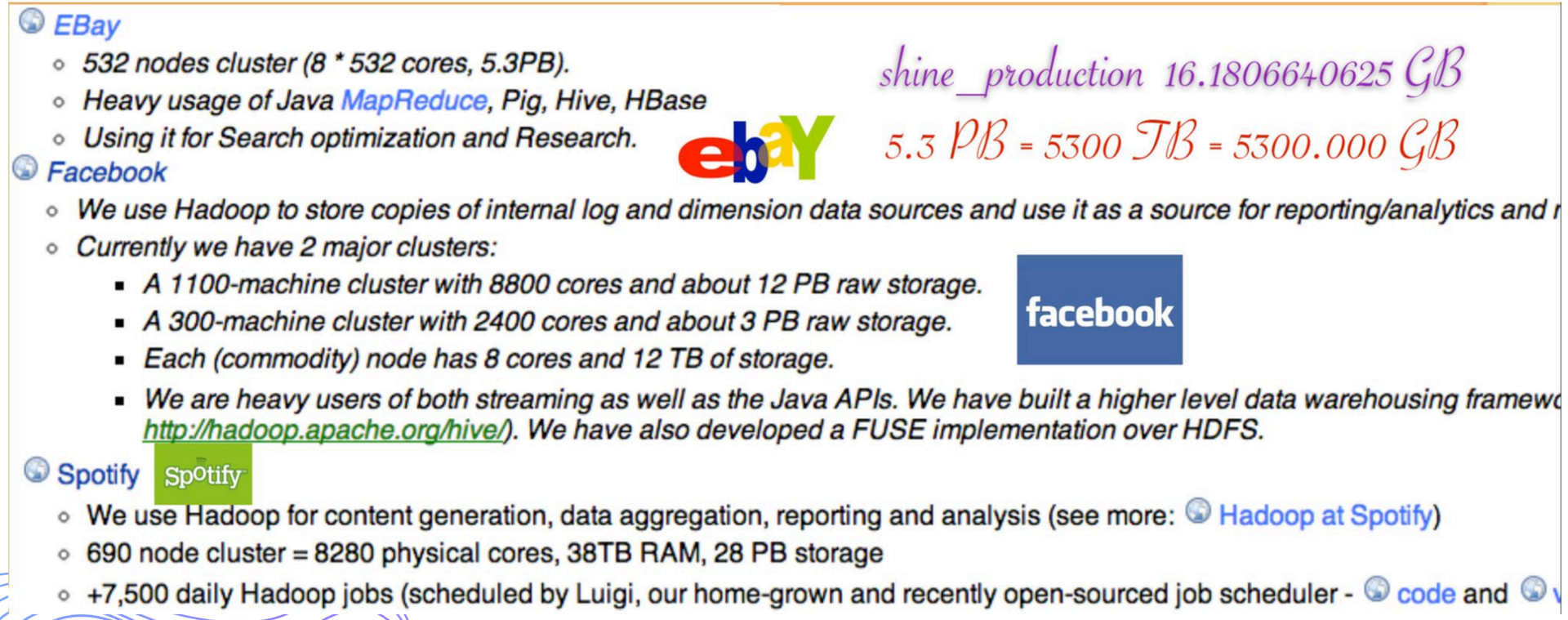


WHAT IS DATA-MINING?

- (Wikipedia) Extract information from a dataset and transform it into an understandable structure for further use.
- To discover the patterns and relationships in the data in order to help make better business decisions

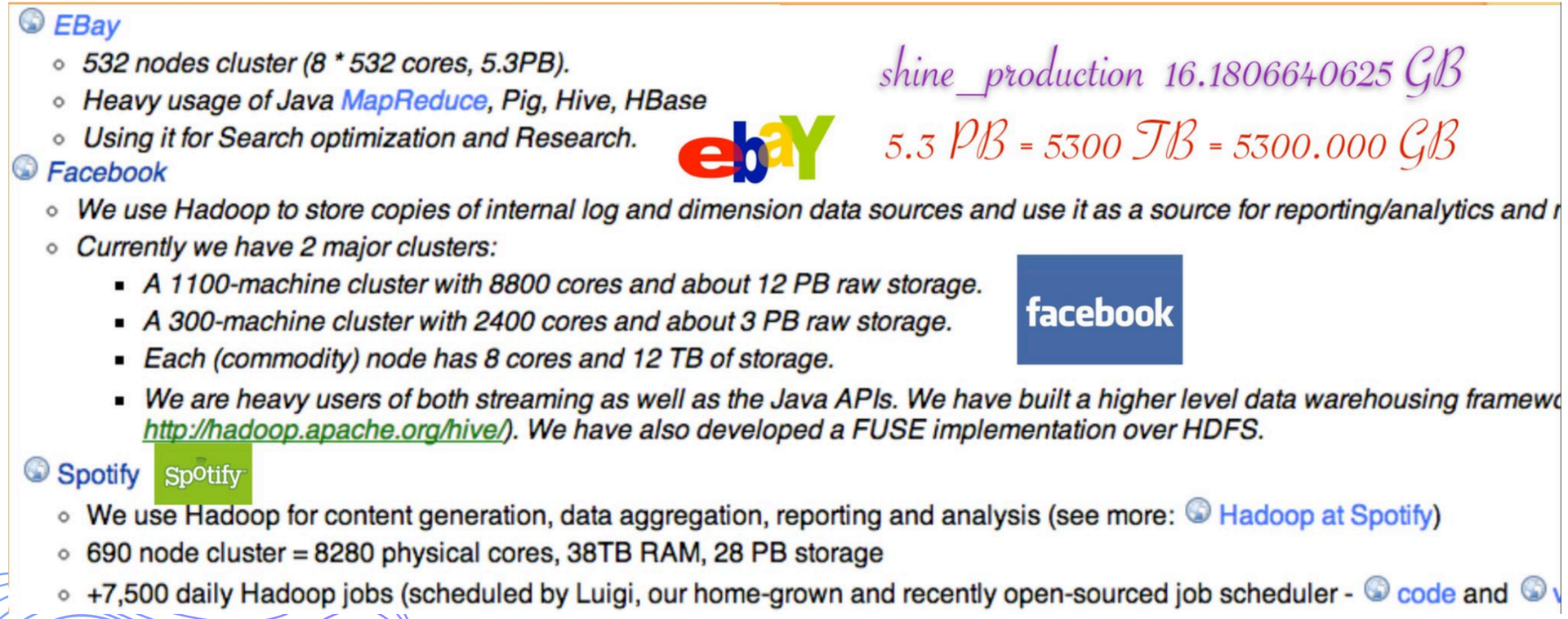
WHAT IS DATA-MINING?

- Databases today can range in size into the terabyte



WHAT IS DATA-MINING?

- Within these mass of data lies lots of crucial and hidden information.



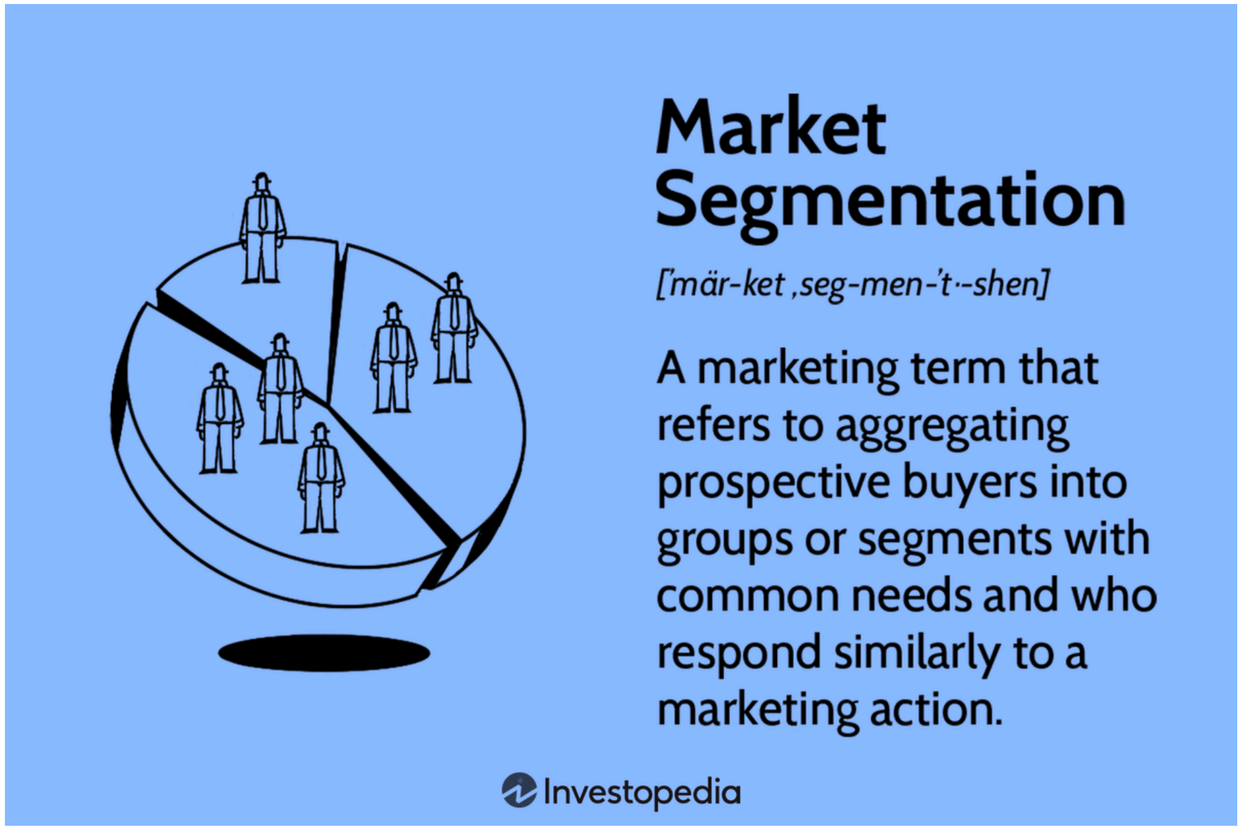
WHAT IS DATA-MINING?

WHAT IS DATA-MINING?

WHAT IS DATA-MINING?

WHAT CAN DATA MINING DO?

- **Market segmentation**: identify the common characteristics of customers who buy the same products from our company.

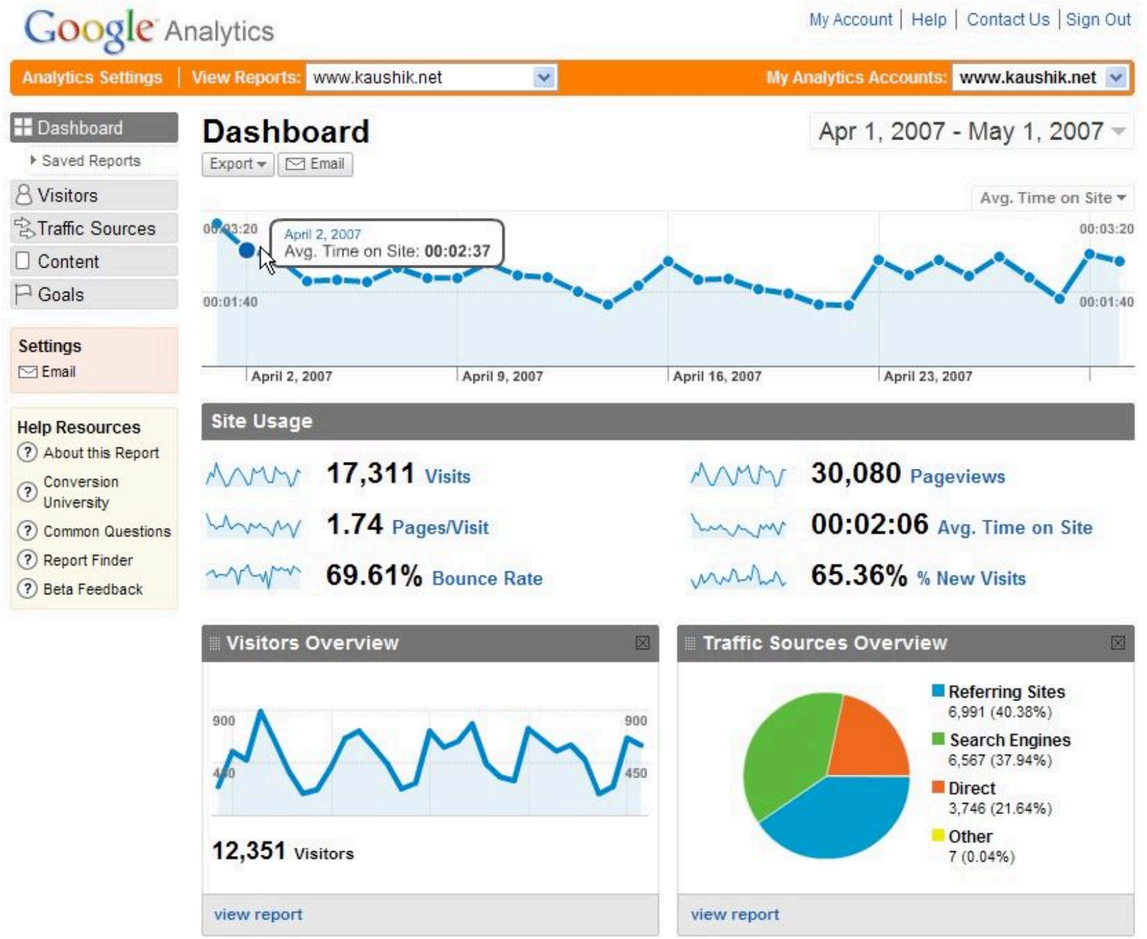


WHAT CAN DATA MINING DO?

- **Customer churn:** predict which customers are likely to leave our company and go to the competitor

WHAT CAN DATA MINING DO?

- **Interactive marketing**: predict what each individual accessing our Website is mostly likely interested in seeing

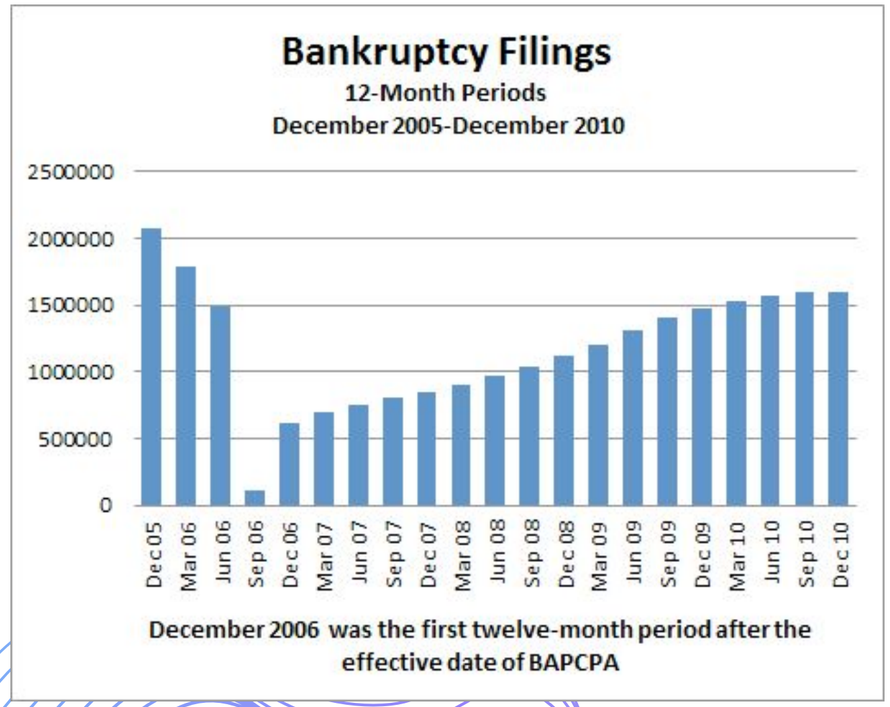


WHAT CAN DATA MINING DO?

- **Market basket analysis**: understand what products or services are commonly purchased together.

WHAT CAN DATA MINING DO?

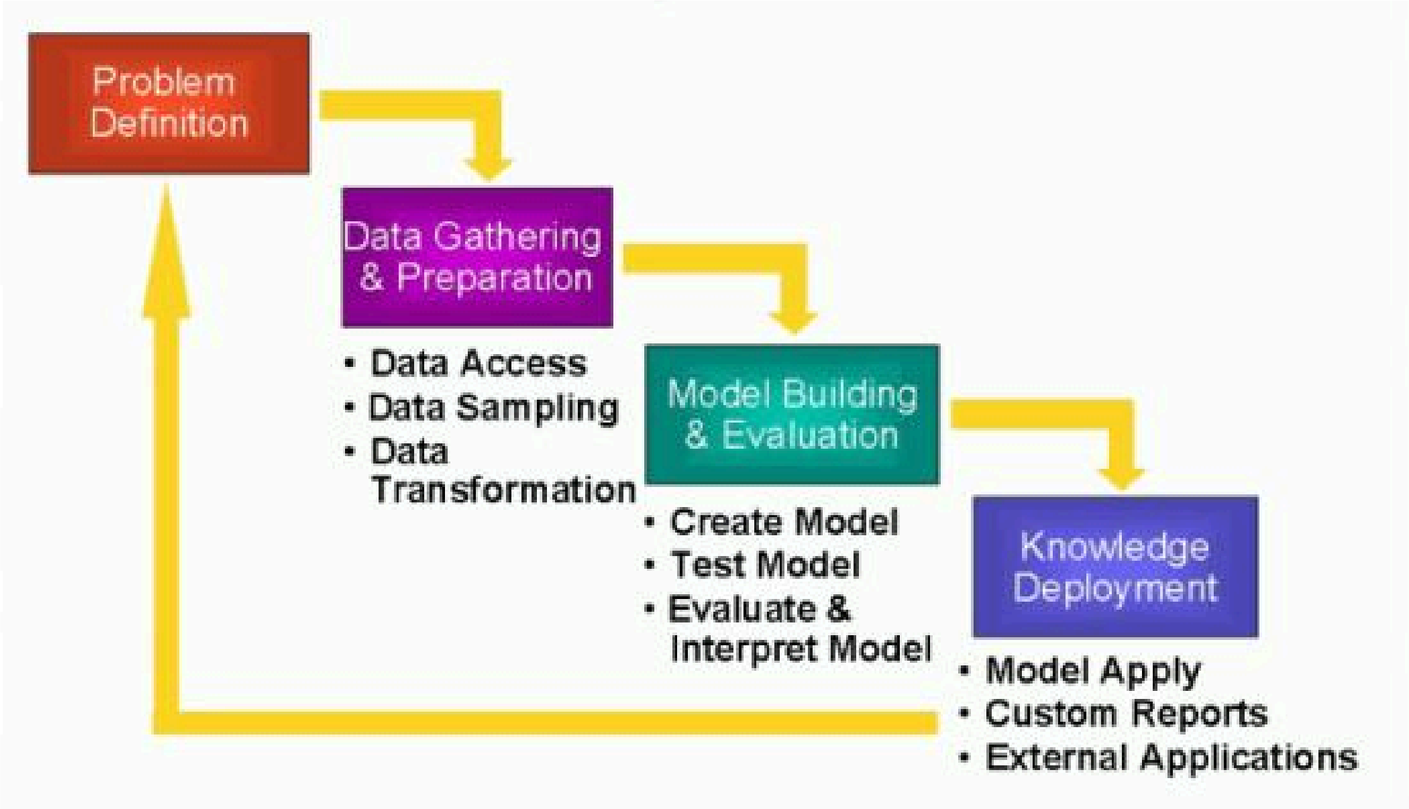
- **Automated prediction of trends and behaviors**: data mining automates the process of finding predictive information in a large database



WHAT CAN DATA MINING DO?

- **Automated discovery of previously unknown patterns**: data mining tools sweep through databases and identify previously hidden patterns.
    - *For example*: analyze the sales data in an company to identify seemingly unrelated products that are often purchased together.

DATA-MINING PROCESS

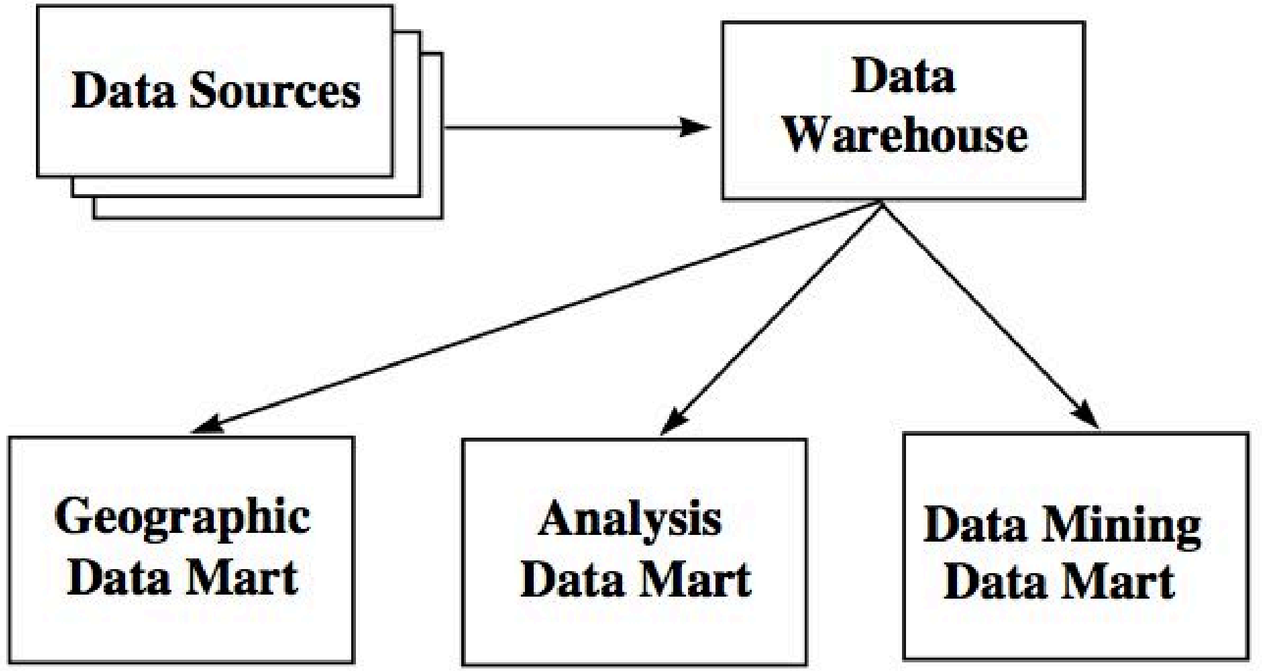


DEFINE BUSINESS PROBLEM

- Understand our data and business
- Clear statements of our objectives.
- *Examples*:
  - **Problem**: increase the response of a direct mail campaign.
    - Model 1: “Increasing the response rate”
    - Model 2: “Increase the value of a response”

BUILD DATA-MINING DATABASE

- Usually take anywhere from 50% to 90% time and efforts of the entire knowledge discovery process



EXPLORE THE DATA

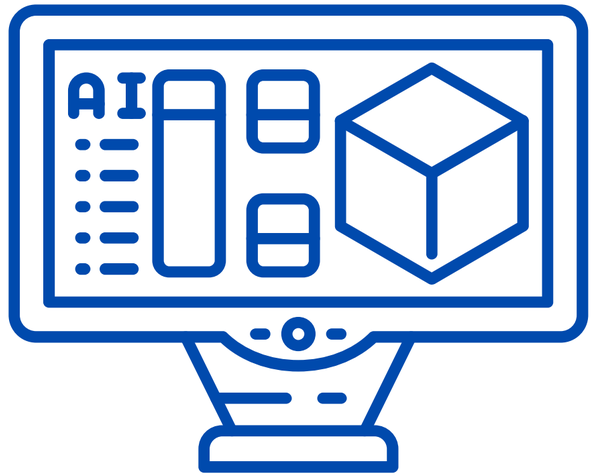
- Identify the most important fields in predicting an outcome and determine which derived values may be useful.

PREPARE DATA FOR MODELING

- This is the final data preparation step before building models.
- Have four main parts:
  - Select variables.
  - Select rows.
  - Construct new variables.
  - Transform variables.

DATA MINING MODEL BUILDING

- Explore alternative models to find the best one in solving our business problem.
- Choose a model type for making the prediction.
- Require a good training and validation protocol (supervised learning) for accurate and robust predictions



FINAL STEPS

- Evaluate the model and interpret the significance of its results.
- The accuracy rate applies only to which the model was built.
- When a data mining model was built and validated, it can be used to recommend actions or to apply the model to various data sets.

FINAL STEPS

- Evaluate the model and interpret the significance of its results.
- The accuracy rate applies only to which the model was built.
- When a data mining model was built and validated, it can be used to recommend actions or to apply the model to various data sets.