Management Information System

Lecture 01 : Databases & Information Management

File organization concepts

Computer system uses hierarchies

- ❑ **Bit :** Smallest unit of data a computer can handle
- ❑ **Byte:** Group of bits, represents a single character
- ❑ **Field:** Group of bytes put into a word
- ❑ **Record:** Group of related fields
- ❑ **File:** Group of records of same type
- ❑ **Database:** Group of related files

File organization concepts (contd..)

- **Record**: Describes an entity

- **Entity**: Person, place, thing about which we store information
  - **Attribute**: Each characteristic, or quality, describing entity
    - E.g. Attribute **Course and Grade** belong to entity **STUDENT**

The Data Hierarchy
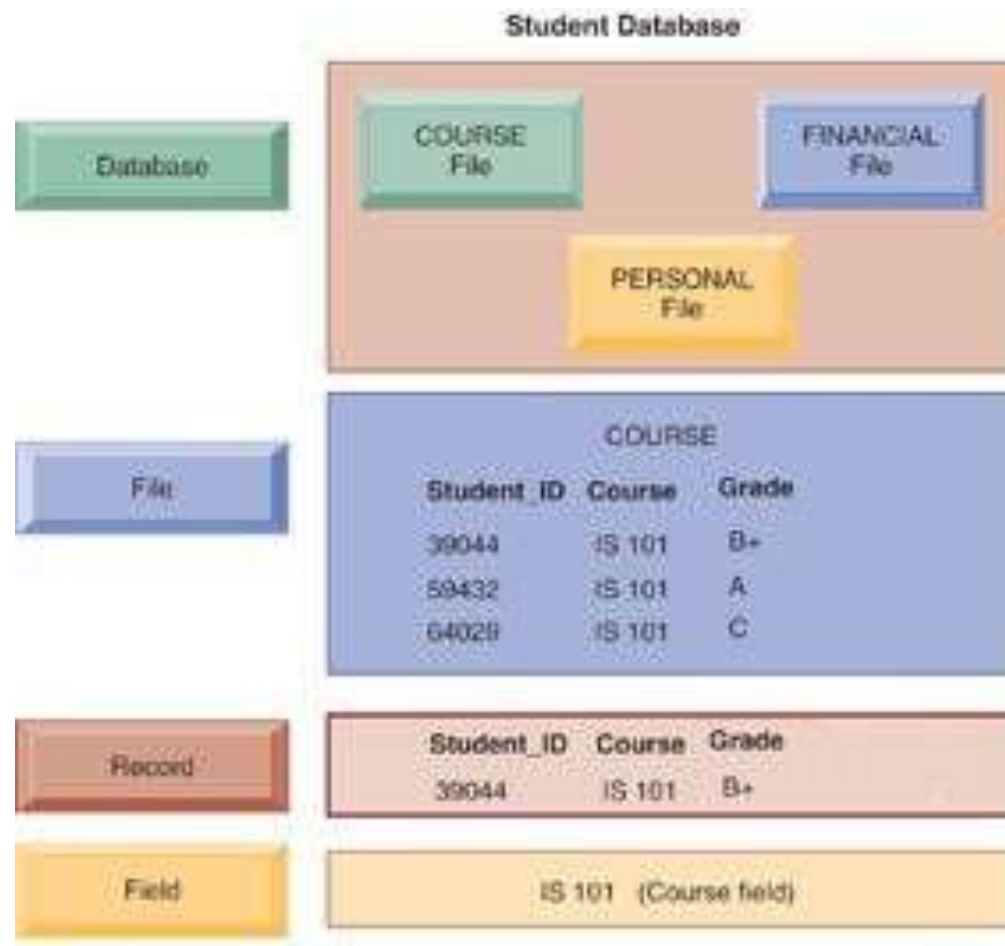
Problems with the traditional files

- **Data redundancy and inconsistency**
    - **Data redundancy:** Presence of duplicate data in multiple files
    - **Data inconsistency:** Same attribute has different values

- **Poor security**

- **Lack of data sharing and availability**
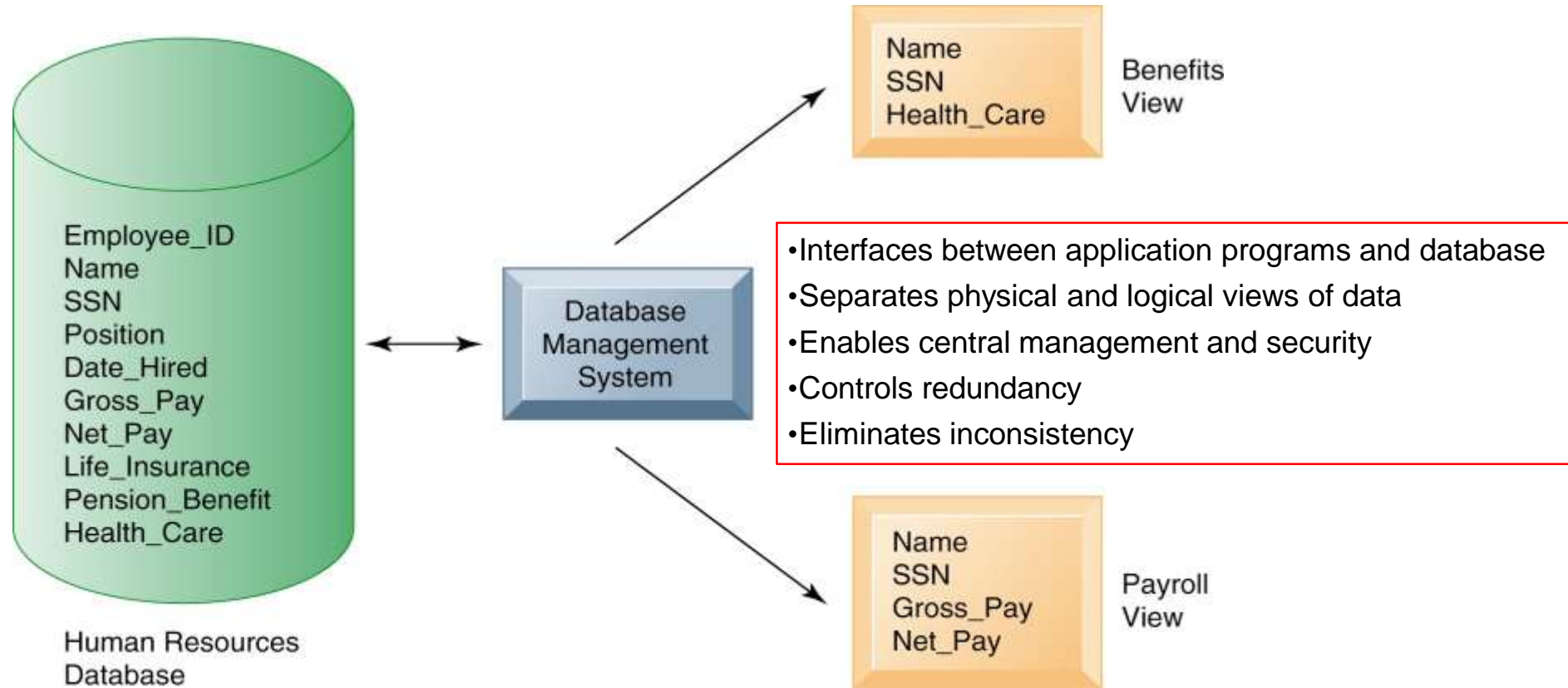
The Database Approach to Data Management

- **Database**
  - Collection of data organized to serve many applications by centralizing data and controlling redundant data
- **Database management system**
  - Interfaces between application programs and database
  - Separates logical and physical views of data
  - Solves problems of traditional file environment
    - Controls redundancy
    - Eliminates inconsistency
    - Enables central management and security

Human Resources Database with Multiple Views

Benefits View
- Name
- SSN
- Health_Care

Human Resources Database
- Employee_ID
- Name
- SSN
- Position
- Date_Hired
- Gross_Pay
- Net_Pay
- Life_Insurance
- Pension_Benefit
- Health_Care

Database Management System
- •Interfaces between application programs and database
- •Separates physical and logical views of data
- •Enables central management and security
- •Controls redundancy
- •Eliminates inconsistency

Payroll View
- Name
- SSN
- Gross_Pay
- Net_Pay

A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

Relational DBMS

- **Relational DBMS**
  - Represent data as two-dimensional tables called relations or files
  - Each table contains data on entity and attributes
  - Examples: Microsoft Access, Oracle Database, MySQL, Microsoft SQL Server.
- **Table**: Grid of columns and rows
  - **Rows (tuples):** Records for different entities
  - **Fields (columns):** Represents attribute for entity
  - **Primary key**: Field in table used for key fields or unique identification
  - **Foreign key**: Primary key used in second table as look-up field to
             identify records from original table

Relational Database Tables
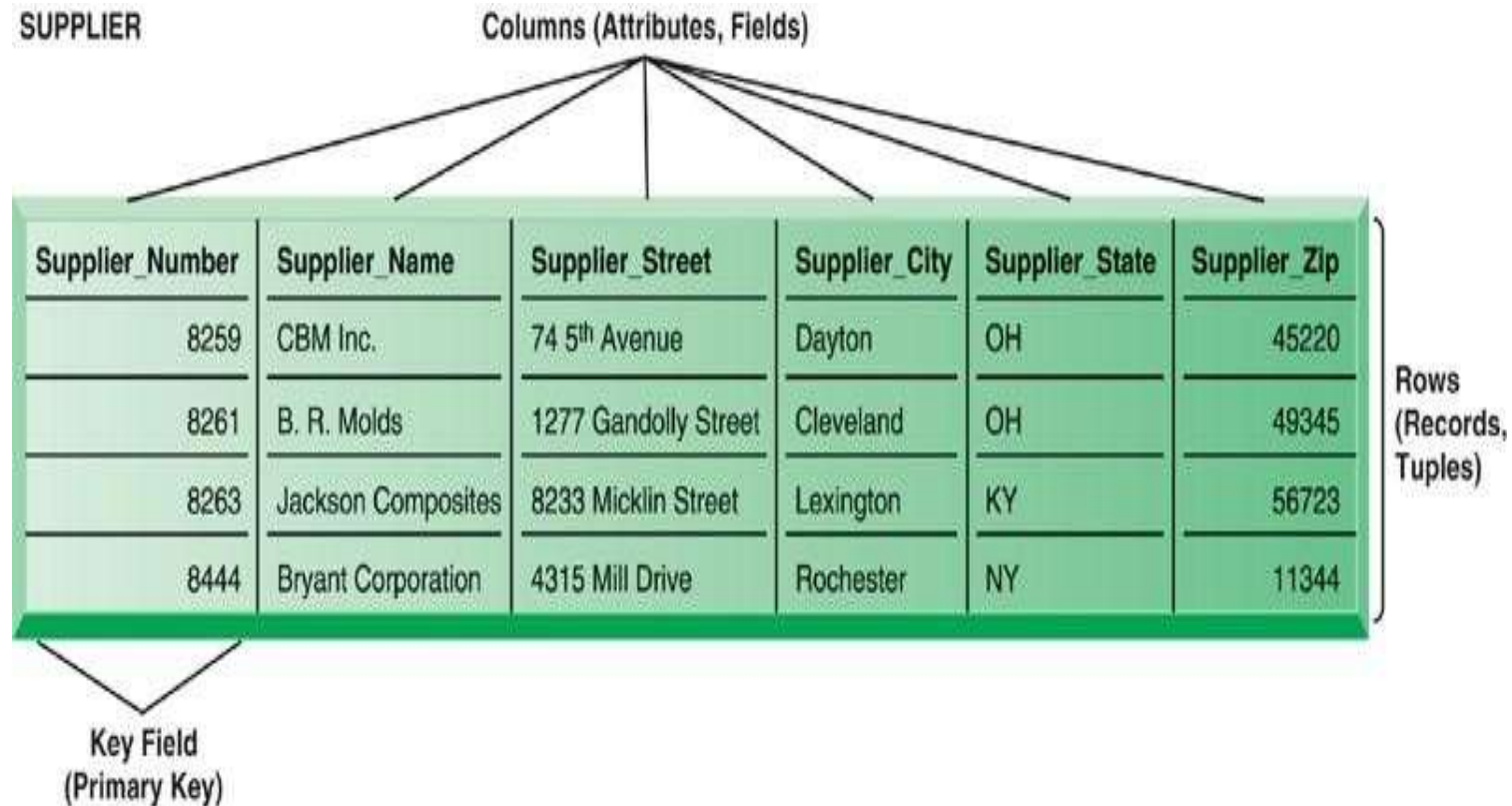
A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier_Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

Relational Database Tables ....

PART

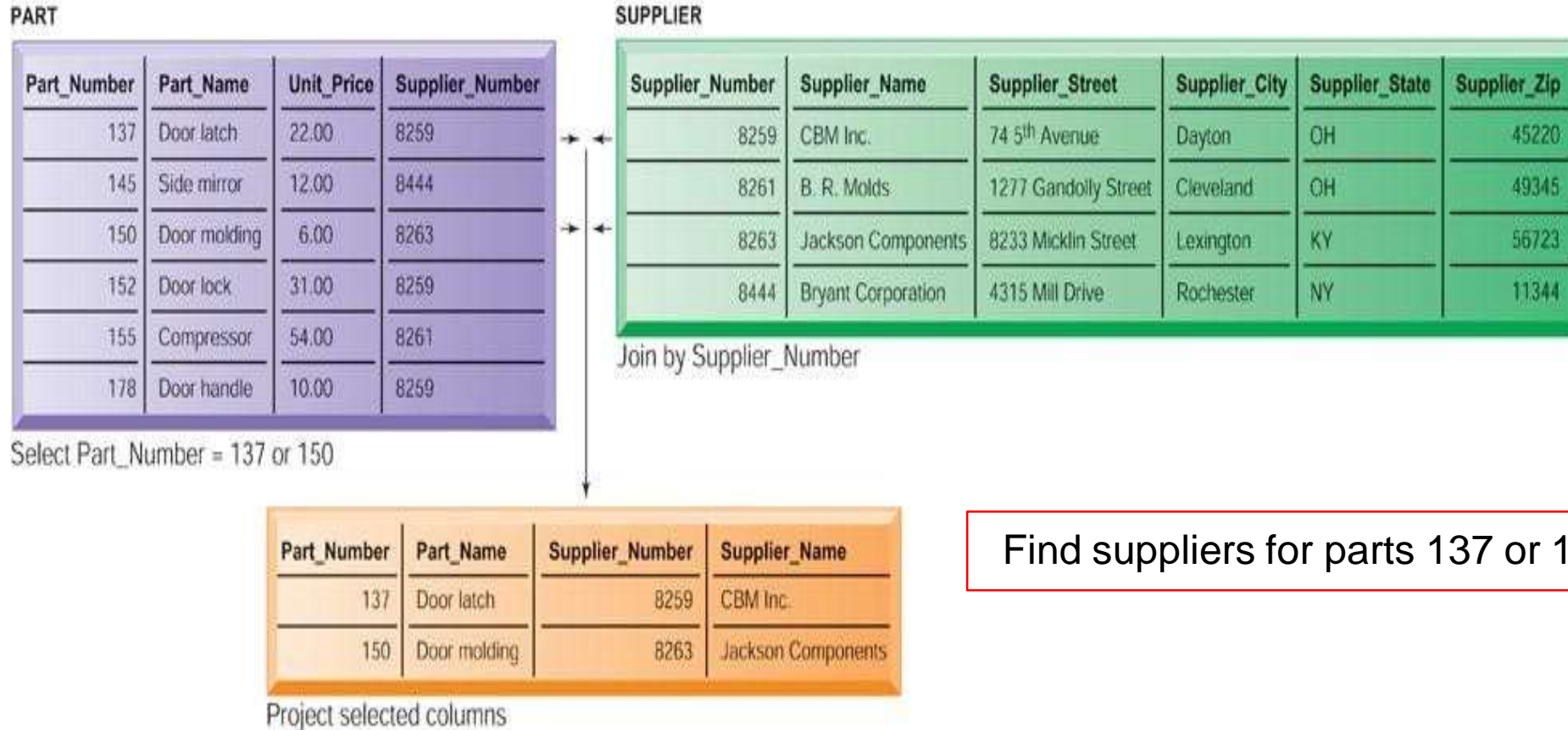| Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key

Foreign Key

Operations of a Relational DBMS

Three basic operations used to develop useful sets of data

- **SELECT**: Creates subset of data of all records that meet stated criteria

- **JOIN**: Combines relational tables to provide user with more information than available in individual tables

- **PROJECT**: Creates subset of columns in table, creating tables with only the information specified

The Three Basic Operations of a Relational DBMS

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Select Part_Number = 137 or 150

SUPPLIER

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Components	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Join by Supplier_Number

Part_Number	Part_Name	Supplier_Number	Supplier_Name
137	Door latch	8259	CBM Inc.
150	Door molding	8263	Jackson Components

Project selected columns

Find suppliers for parts 137 or 150

The select, project, and join operations enable data from two different tables to be combined and only selected attributes to be displayed.

Capabilities of DBMS

- **Data manipulation language:** Used to add, change, delete, retrieve data from database
  - Structured Query Language (SQL)
  - Microsoft Access user tools

- Many DBMS have **report generation capabilities** for creating polished reports

Example of an SQL Query

SELECT: Lists the desired fields that have to be included in the query
FROM: Lists the tables from where the data has to be drawn
WHERE: Specifies the values of the fields that have to be included or the
conditions that have to be met to include the field.

```
SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number,
SUPPLIER.Supplier_Name
FROM PART, SUPPLIER
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND
Part_Number = 137 OR Part_Number = 150;
```

Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150

Management Information System

Data Warehouses & Business Intelligence

E-R (Entity-Relationship Diagrams)

- E-R Models are descriptions of the business requirements of data from the user perspective.

- They are often the first step in database design.

- A set of diagrammatic tools used to create these models are called E-R diagrams.

- An E-R diagram consists of data *entities, relationships* between these entities and *attributes* that describe the entities and relationships.

- Relationships identify natural links or associations between entities.

E-R (Entity-Relationship Diagrams) contd...
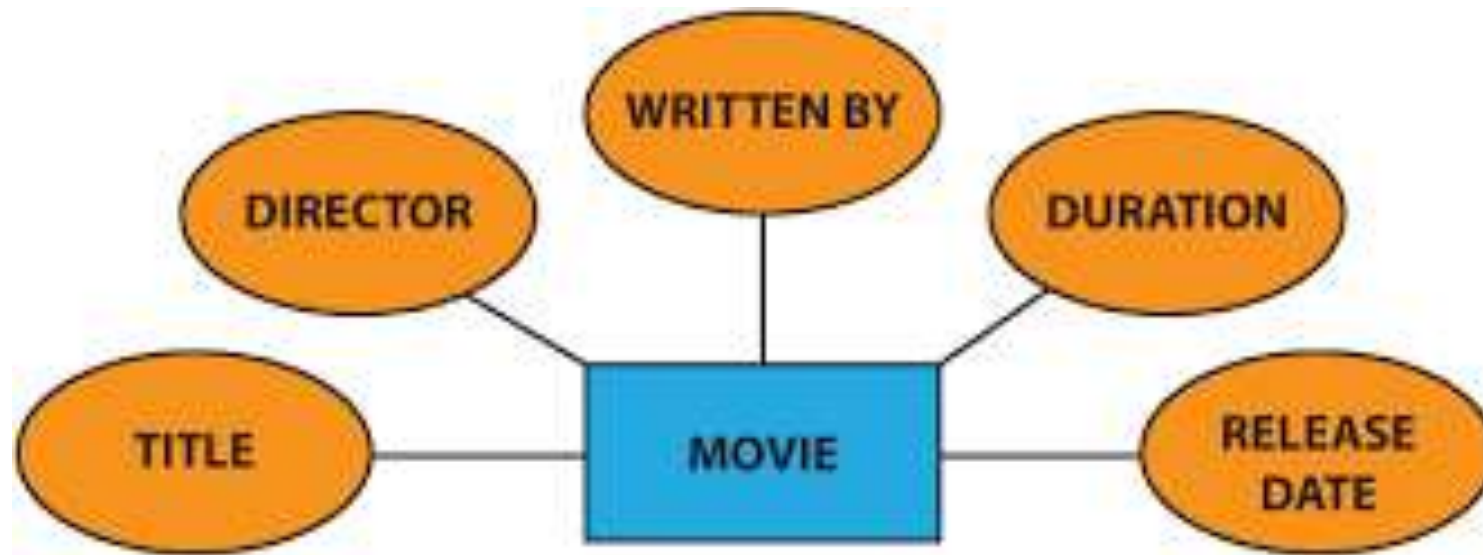
Movie entity with attributes

E-R (Entity-Relationship Diagrams) contd..

- The number of entities in a relationship represent the *cardinality* of a relationship.

Cardinality of relationship
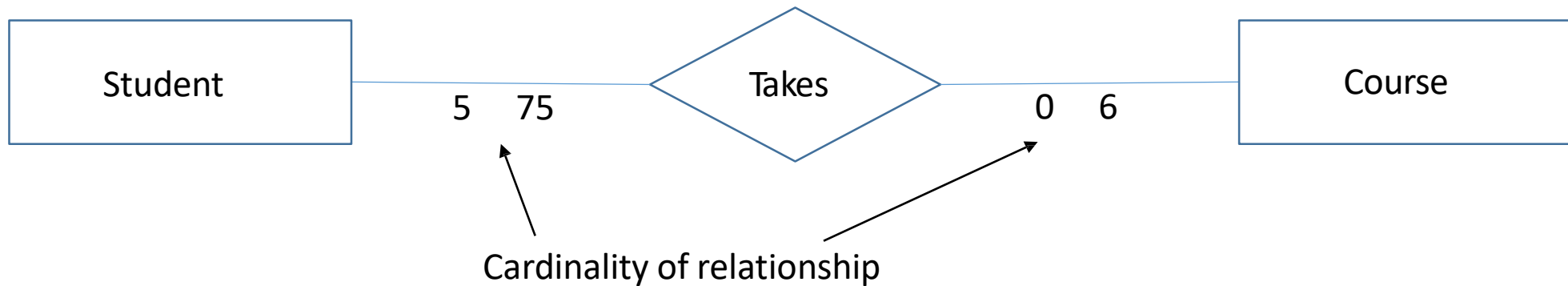
E-R (Entity-Relationship Diagrams) contd..

Various Types of Relationships for E-R Diagrams

| Employee | 1 N | has | 1 | Supervisor |

| Customer | 1 | places | 1 N | Order |

| Order | 1 | has | 1 | OrderDetails |
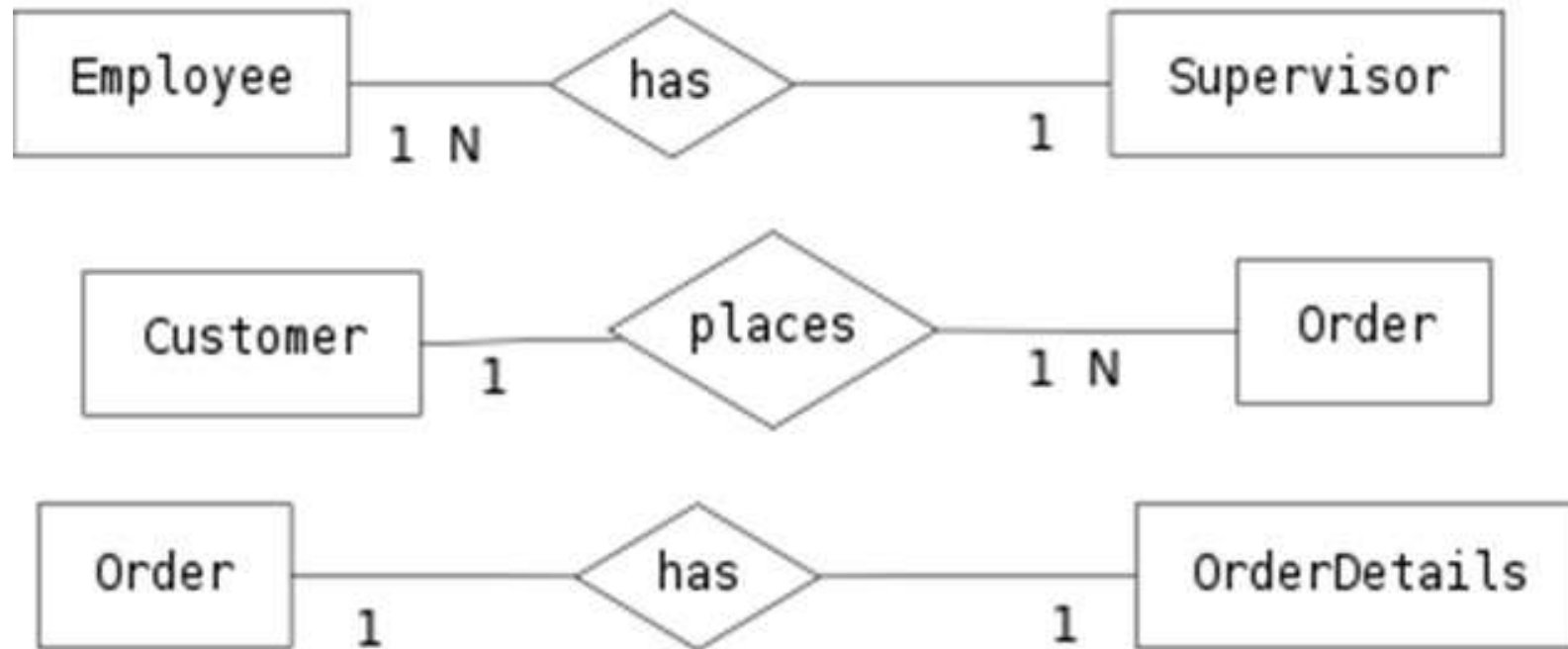
Normalization

Streamlining complex groupings of data to minimize redundant data elements.

Students Engg. Mechanics

Roll No.	Name	Dept.	HoD	Dept. Contact no.
101	Sachin	Electrical	Prof. X	1234567
102	Rahul	Mechanical	Prof. Y	4567899
103	Saurav	Electronics	Prof. Z	6789048
104	Virat	Mechanical	Prof. Y	4567899
105	Dhoni	Electrical	Prof. X	1234567
106	Anil	Mechanical	Prof. Y	4567899

Normalization (contd..)

Students Engg. Mechanics

Roll No.	Name	Dept. Id
101	Sachin	001
102	Rahul	002
103	Saurav	003
104	Virat	002
105	Dhoni	001
106	Anil	002

Department

Dept. Id	Dept. Id	HoD	Dept. Contact no.
001	Electrical	Prof. X	1234567
002	Mechanical	Prof. Y	4567899
003	Electronics	Prof. Z	6789048

Using Databases to Improve Business Performance & Decision Making

- Very large organizations with huge amounts of data felt the need for:
  - Consolidating much of the data from various databases into a whole that could be understood clearly.
  - Focusing on the use of data for decision making, as opposed to simply for running transactions.

Using Databases to Improve Business Performance & Decision Making (contd..)

- This gave rise to special capabilities and tools required for analyzing large quantities of data:
  - Data warehouses
  - Data marts
  - Data mining

Data Warehouses

- Stores current and historical data from many core operational transaction systems

- Consolidates and standardizes information for use across enterprise, but data cannot be altered.

- To create a data warehouse, data is extracted from transactional tables, pre-processed to remove unwanted data types and then loaded into tables in the data warehouse.

- Warehouses differ from transaction databases as users can run complex queries on them.

- Data warehouses provide querying, analysis, and reporting tools.

-

Data Mart

- Subset of data warehouse with summarized or highly focused portion of firm's data for use by specific population of users.

- Typically focuses on single subject or line of business.

- Used to identify problems and find solutions pertaining to a particular domain.

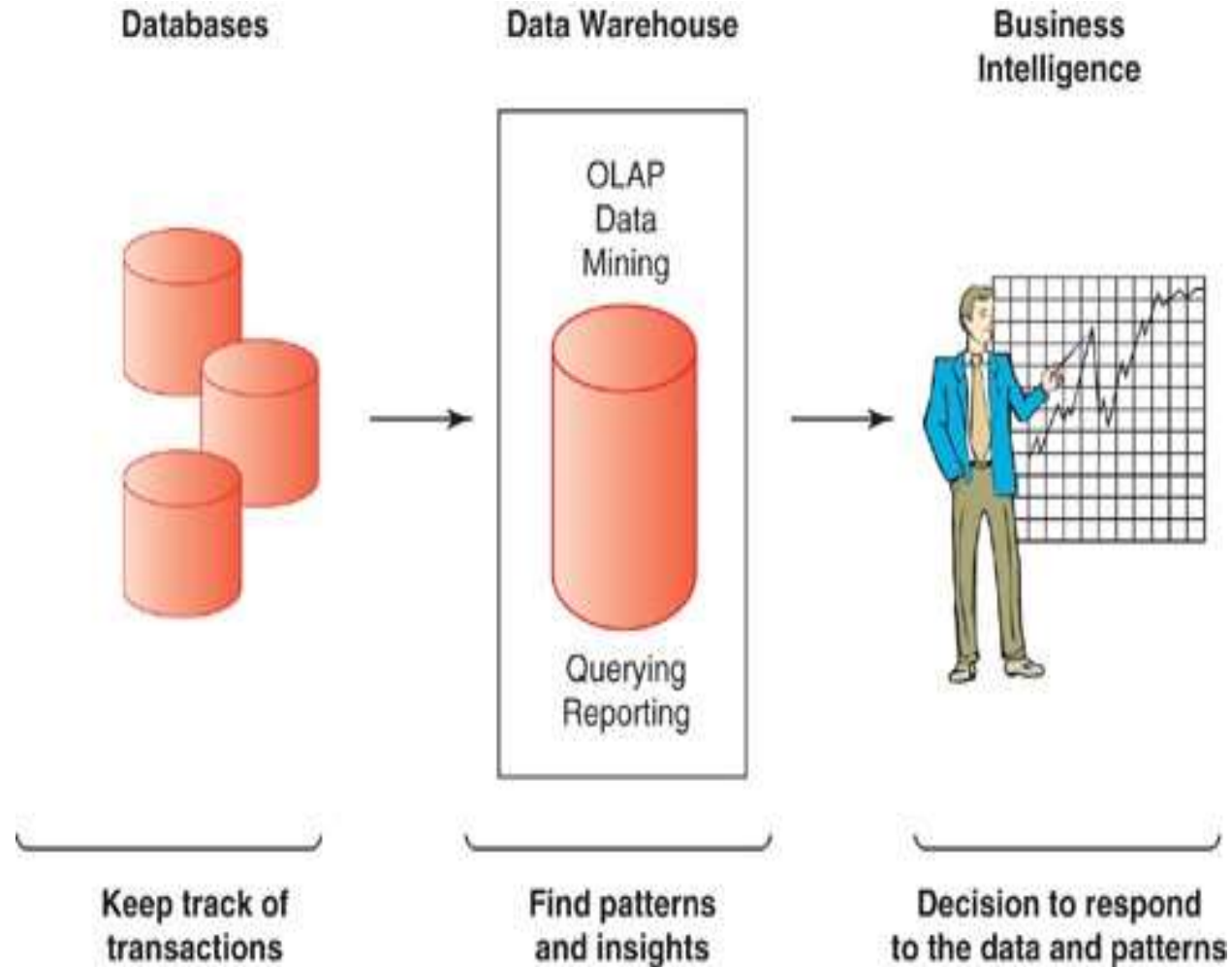- Example: Sales data mart.

Components of a Data Warehouse
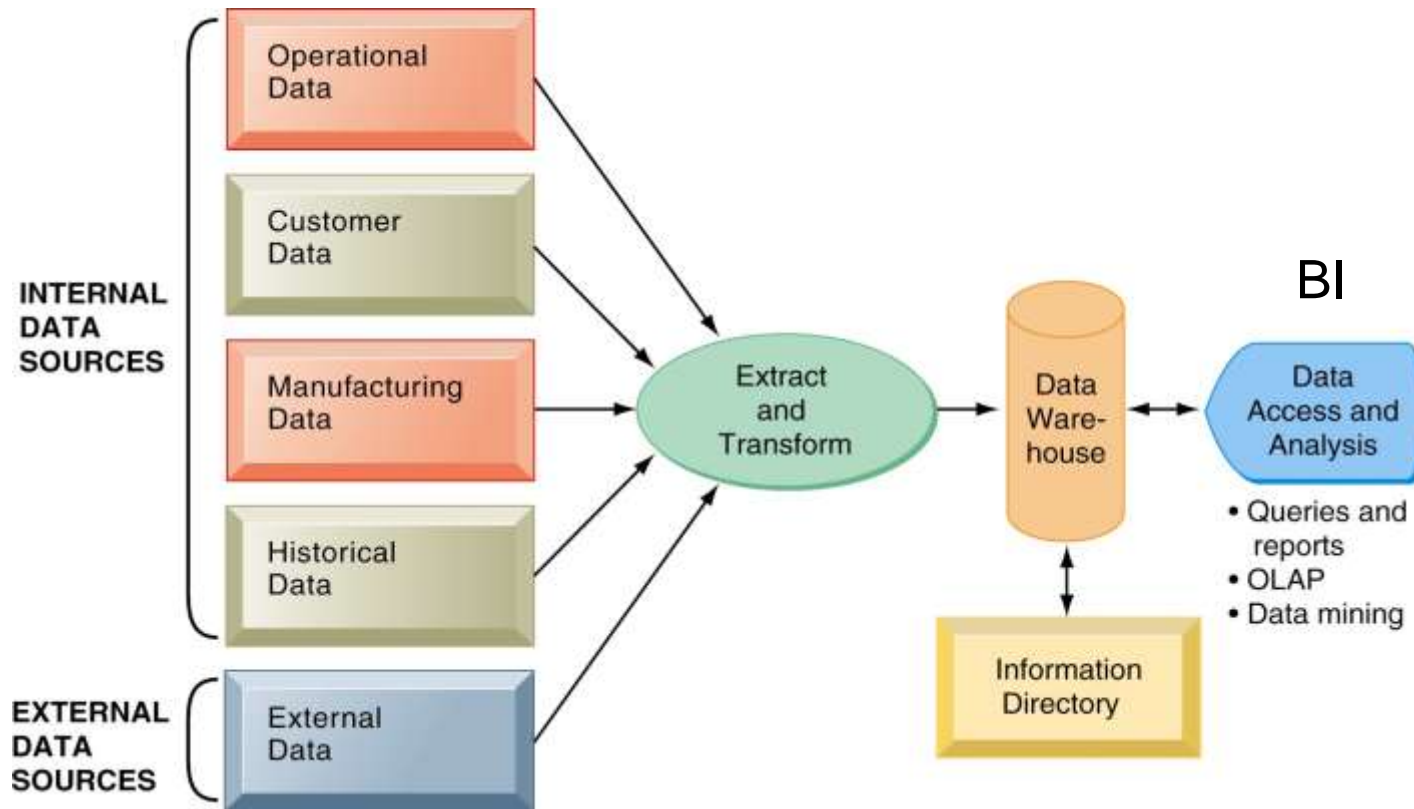
The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and reorganized into a central database designed for management reporting and analysis. The information directory provides users with information about the data available in the warehouse.

Business Intelligence

- Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions
  - E.g. Harrah's Entertainment analyzes customers data to develop gambling profiles and identify most profitable customers
- Principle tools to derive BI from data in a warehouse include:
  - Software for database querying and reporting
 - Online analytical processing (OLAP)
  - Data mining

What drives Business Intelligence?

Databases

Data Warehouse

Business Intelligence

OLAP
Data
Mining

Querying
Reporting

Keep track of
transactions

Find patterns
and insights

Decision to respond
to the data and patterns

Online analytical processing (OLAP)

- Supports multidimensional data analysis
- Enables viewing data using multiple dimensions
  - Each aspect of information (product, pricing, cost, region, time period) is different dimension
  - E.g. How many cycles were sold in Eastern India in June?
- OLAP enables rapid, online answers to ad hoc queries

Multidimensional Data Model
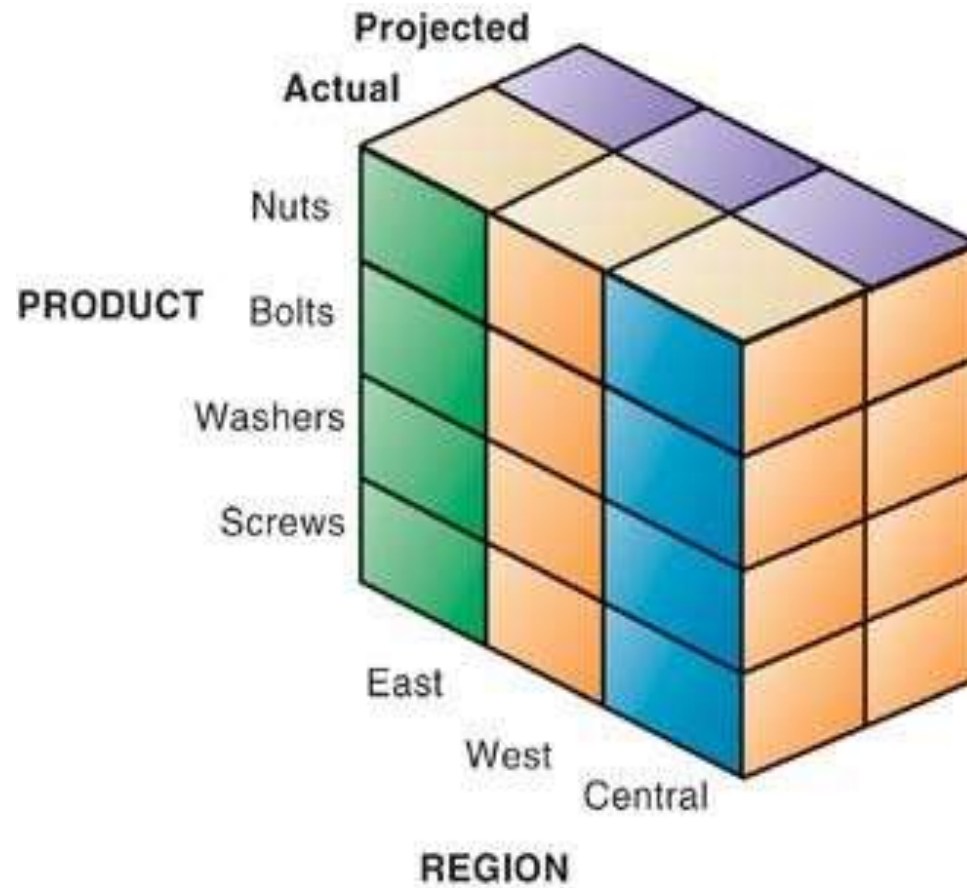
- What were the actual sales of Bolts in Central India?

Management Information System

Introduction to Data Mining

Why mine data?

- Huge amounts of data being collected and warehoused
  - Walmart records 20 million items in transactions per day
  - Health care transactions: multi-gigabyte databases
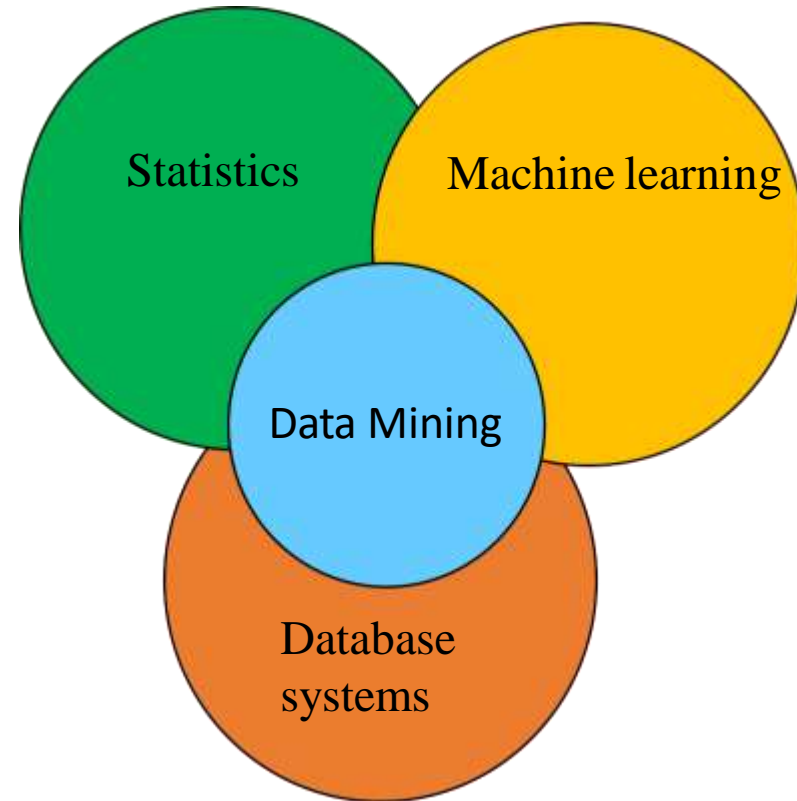
- Source of competitive advantage

What is data mining and KDD?

- Knowledge discovery in databases (KDD) is the non-trivial process of identifying **valid**, potentially **useful, understandable** and ultimately **actionable patterns** in data.

- Data mining is a step in the KDD process of applying data analytics and discovery algorithms

Knowledge Discovery Process

Integration   **Understanding**

Interpretation & Evaluation

Data Mining

Raw Data

Transformation

Selection & Cleaning

Knowledge

Patterns and Rules

Transformed Data

Target Data

DATA Ware house
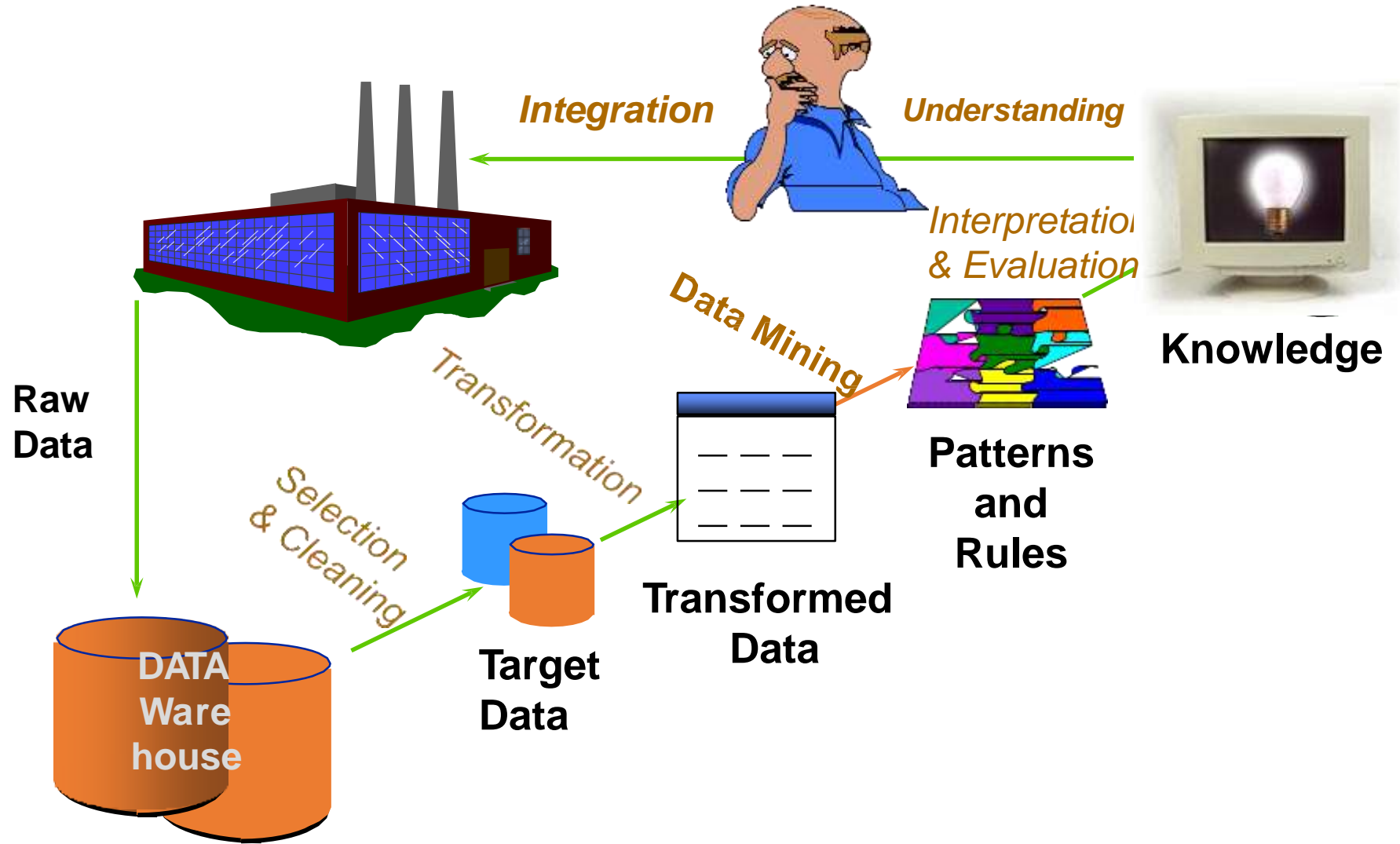
Data Mining Applications

Typical Applications

Customer Segmentation

Propensity to Buy

Profitability Modeling & Profiling
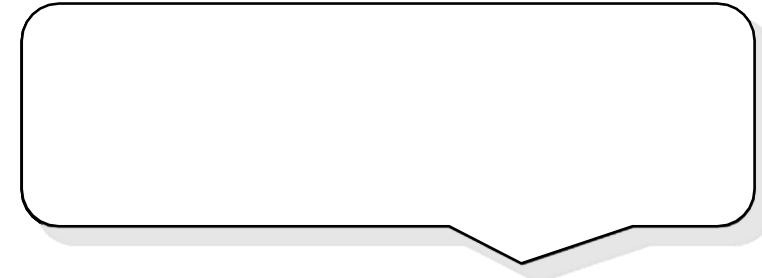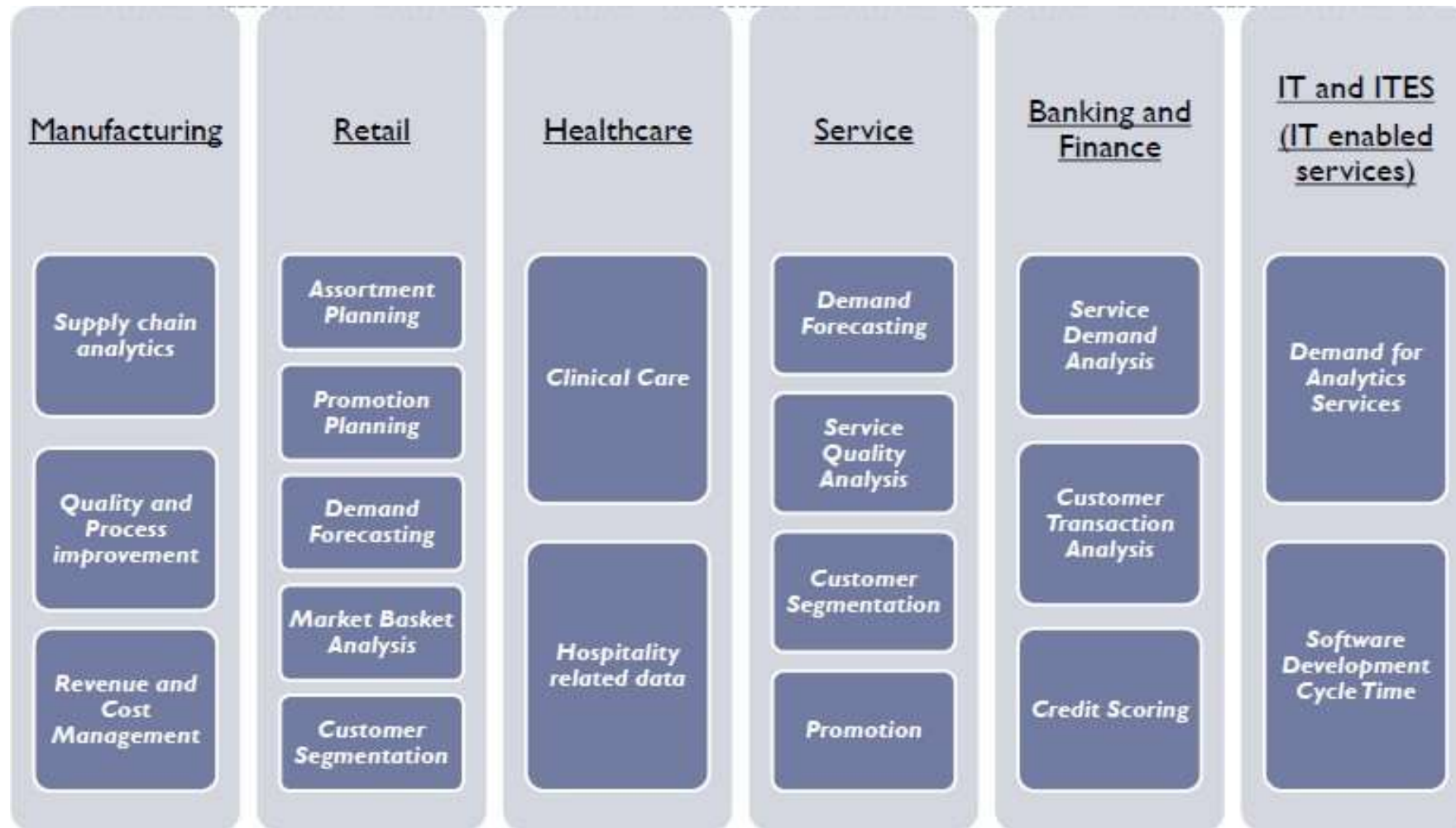
Customer Attrition

Channel Optimization

Fraud Detection

Detect and prevent fraud to minimize loss.

Industry wide applications of analytics

Manufacturing	Retail	Healthcare	Service	Banking and Finance	IT and ITES (IT enabled services)
Supply chain analytics	Assortment Planning	Clinical Care	Demand Forecasting	Service Demand Analysis	Demand for Analytics Services
Quality and Process improvement	Promotion Planning		Service Quality Analysis	Customer Transaction Analysis	
Revenue and Cost Management	Demand Forecasting	Hospitality related data	Customer Segmentation	Credit Scoring	Software Development Cycle Time
	Market Basket Analysis		Promotion		
	Customer Segmentation				

**Primary sources of data and secondary sources to be used in solving these analytical problems

Some other applications….

- **Marketing**
  - Which customers are likely to respond to this campaign?
  - Which customers are likely to be profitable ?
  - Who might want to buy this product ? ('Cross-selling')
  - Which web-pages are customers visiting before buying products or before leaving our site ? Which types of customers visit which pages ?
- **Telecommunications**
  - Which customers are vulnerable to attrition (at risk of churning) ?
  - Based on these symptoms, where are problems located in the network ?
- **Finance and Insurance**
  - Which customers are credit risks / insurance risks ?
  - Which claims or credit transactions are fraudulent ?
  - Which stocks are likely to perform well in the next 3 months, and why ? When should we buy and sell, given the likely performance and transaction costs ?

Some other applications (Contd..)

- **Healthcare**
  - Which patients may take longer to recover ?
  - What is the likely cause of this illness ?
  - Which patients are at risk of disease (and might benefit from medication)? Pfizer pharmaceuticals used data mining to construct a predictive model that was then embedded in their online cholesterol health risk assessment, which tells patients their cholesterol risk score. High risk patients can consult their doctors and request Lipitor, Pfizer's cholesterol medication.

- **Retail**
  - Which products do customers buy together (or in sequence) & which do they not buy together ? ('Category management'.)
  - What characterizes customers at various stores ?
  - What items are bought for cash, on credit, or by check ?
  - What type of customer buys this item, or this product type ?

Data Mining Applications (Contd..)

- **Quality Control**
  - Which shipments are high-risk and need to be inspected ?
- **Customer Support**
  - Which tasks schedule (ordering) is optimal (or good enough) ?
  - Which customer service representative should I assign to a task ?
  - What documents or people are likely to be helpful to the customer in solving their problem ?

Real-life Applications…

At Flipkart...

- Forecast demand for each SKU.

- Predict customer cancellations and returns.

- Predict what a customer is likely to purchase in the future?

- How to optimize the delivery system?

Management Information System

Data Analytics Tools and Techniques

Divorce360.com

Target predicts customer pregnancy

https://www.youtube.com/watch?v=XH1wQEgROg4

Categories of Data Analytics

Descriptive Analytics

Describe, summarize and analyze historical data (What happened?)

Diagnostic Analytics

Identify causes of trends and outcomes (Why it happened?)

Predictive Analytics

Predict future outcomes based on the past (What could happen?)

Prescriptive Analytics

Recommend 'right' or optimal actions or decisions (What should be done?)
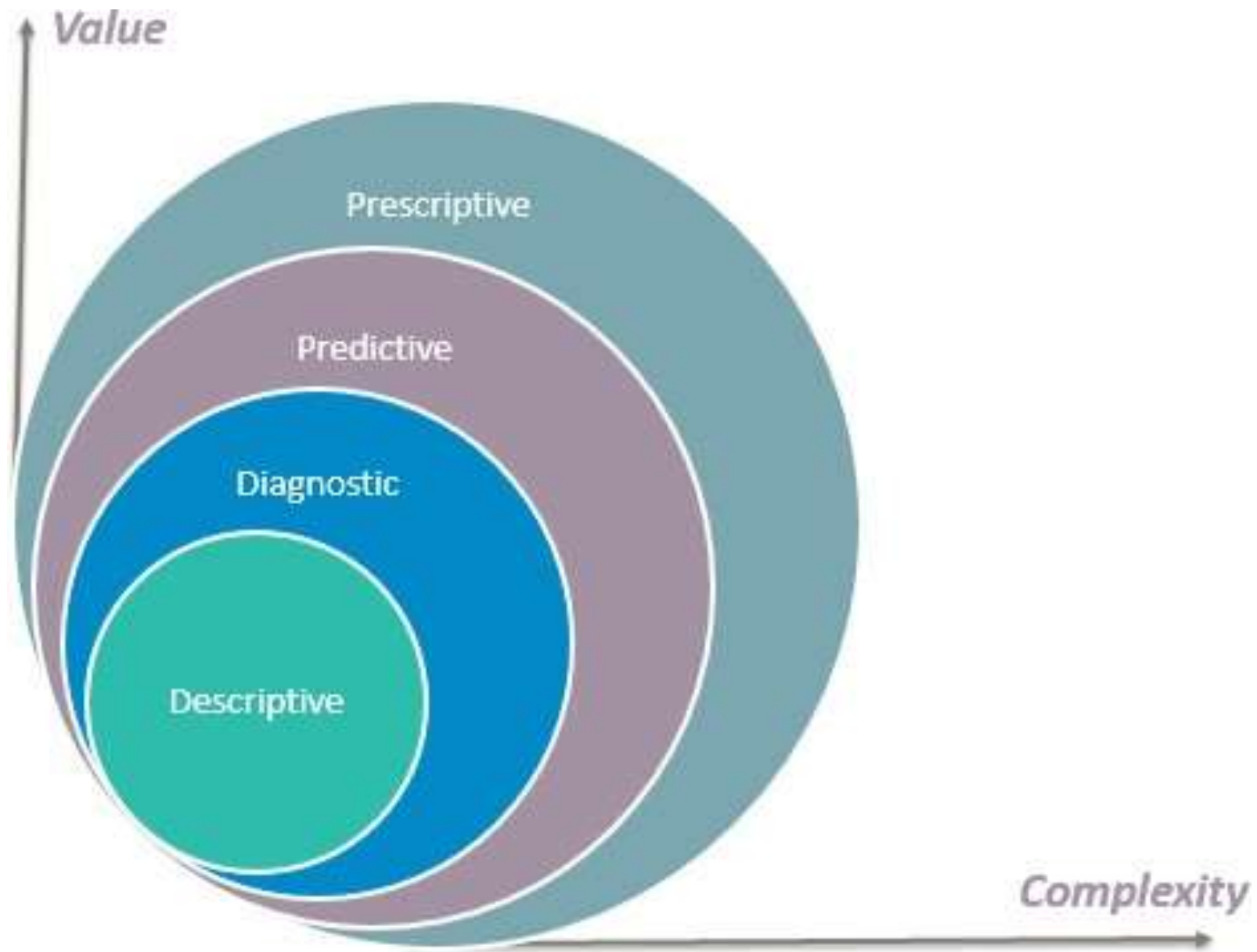
Descriptive Analytics Applications

▸ Most shoppers turn towards right when they enter the a retail store.

▸ Conversion rate of women shoppers is higher than male shoppers among electronic gadgets purchasers (Radio Shack).

▸ Strawberry pop-tarts sell 7 times more during hurricane compared to regular period (Wal Mart).

▸ Women car buyers prefer women sales person.

Diagnostic Analytics Applications

➤ Why customers liked your social media campaign or why they didn't?

➤ Why certain products were popular at a certain time, at a certain place?

Predictive Analytics Applications

‣ Which product the customer is likely to buy in his next purchase (recommender system).

‣ Which customer is likely to default in his/her loan payment.

‣ Who is likely to cancel the product that was ordered through e-commerce portal.

Prescriptive Analytics Applications

▸ What is the optimal product mix?

▸ What is the optimal route for a delivery truck.

▸ Best markdown pricing for fashion products.

▸ Optimal assignment of aircraft to flight.

▸ How to manage the fleet of vehicles owned by a company for employee drop and pick up?
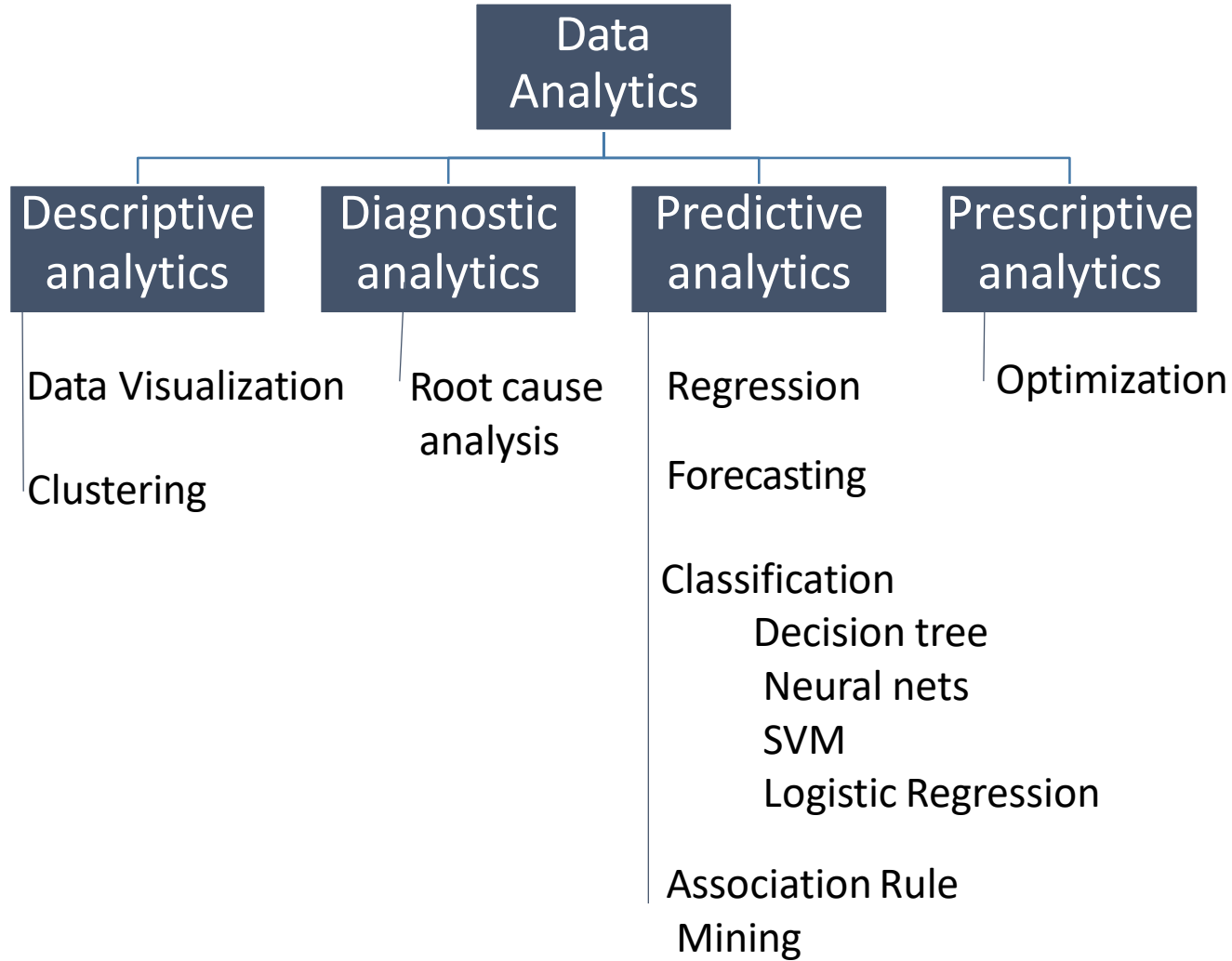
Descriptive Analytics Techniques

➢ **Data Visualization:** Graphical representation of **data**. By using visual elements like charts, graphs, and maps, **data visualization** tools provide an accessible way to see and understand trends, outliers, and patterns in **data**.

➢ Software for Data Visualization: MS Excel; Power BI, Tableau; QlikView

Descriptive Analytics Techniques (contd..)

- **Clustering:** Given a set of data points, each having a set of attributes, find clusters such that:
  - data points in one cluster are more similar to one another
  - data points in separate clusters are less similar to one another.

- Application in Market Segmentation: Subdivide a market into distinct subsets of customers based on their geographical and lifestyle related information where any subset may conceivably be selected as a market target to be reached with a distinct marketing mix.

Intercluster distances are maximized

Intracluster distances are minimized

12
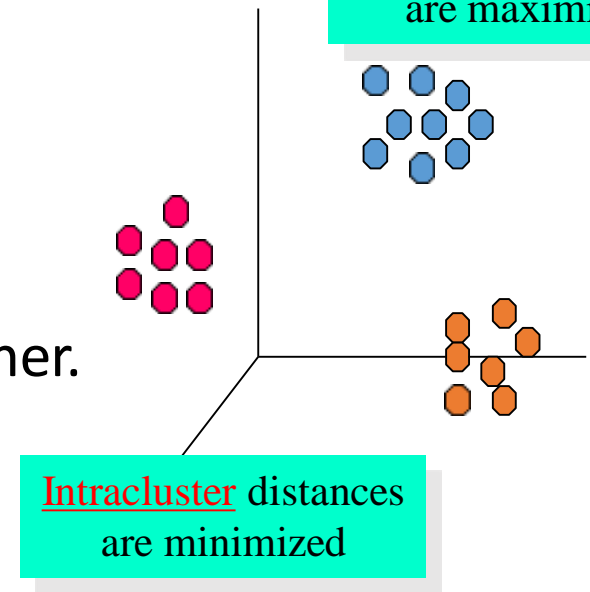
Predictive Analytics Techniques

- **Regression analysis:** Regression analysis is a statistical technique used to describe relationships among variables. The simplest case to examine is one in which a variable Y , referred to as the dependent or target variable, may be related to one variable X, called an independent or explanatory variable.
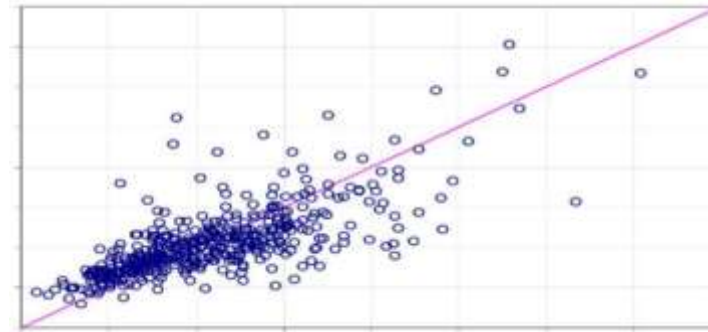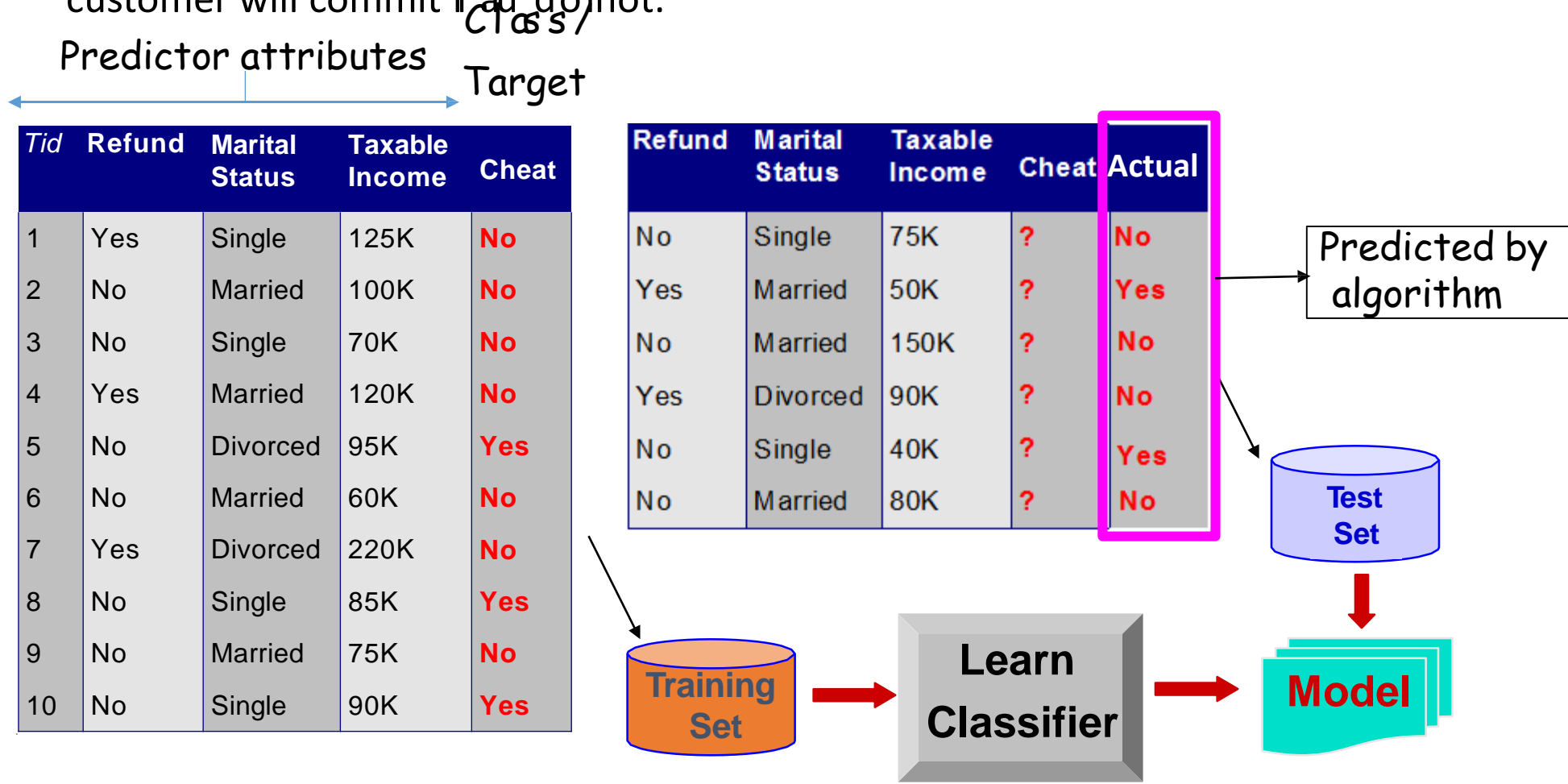

- **Applications:**
- Predicting events that are yet to occur:
    - Demand analysis or number of units consumers will purchase in the next quarter based on past trends
    - Number of shoppers who will watch a particular advertisement.
    - Number of policy holders who will be involved in accidents in the next year.

Predictive Analytics Techniques (contd..)

- **Classification:** Data defined in terms of <u>attributes</u>, one of which is the <u>class or target</u> and others are predictor variables.

- Find a model for class/target attribute as a function of the values of other(predictor) attributes, such that previously unseen records can be assigned a class as accurately as possible. Given data is usually divided into training and test sets.

- Training Data: used to build the model

- Test data: used to validate the model (determine accuracy of the model).

- Classification Techniques:
  - Decision tree
  - Neural Networks
  - Support Vector Machines
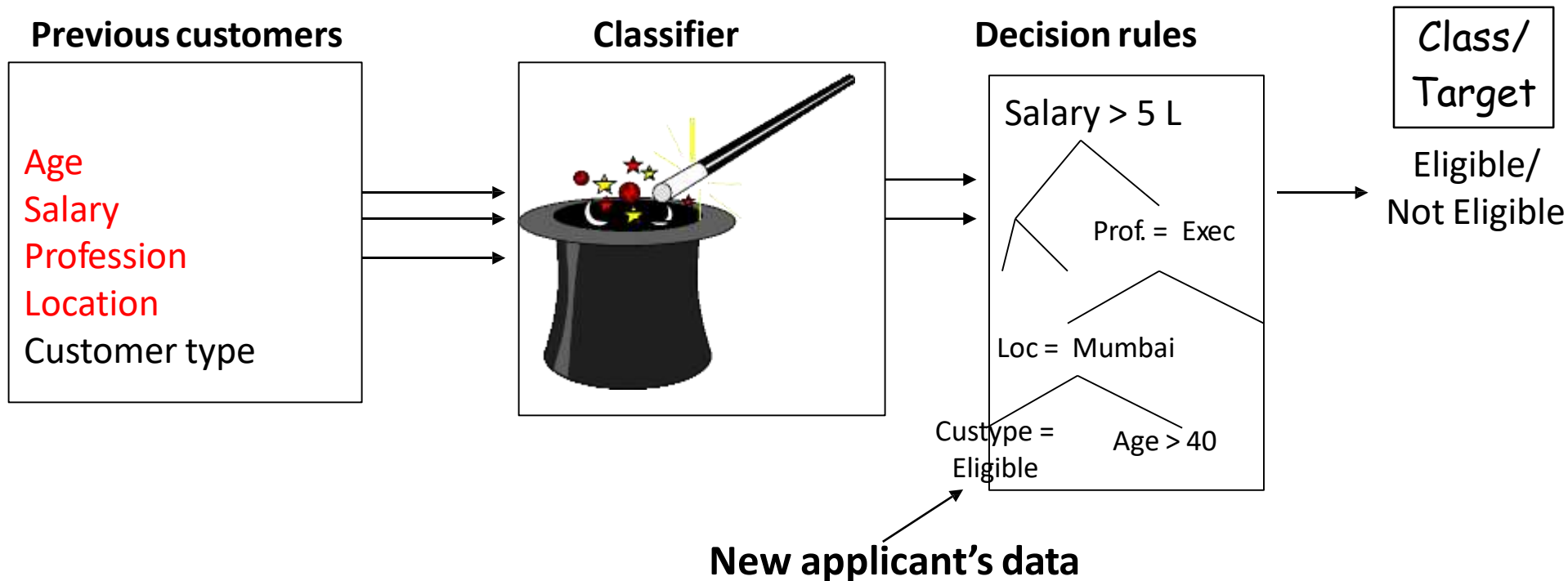 - Bayesian Classifier

Predictive Analytics Techniques (contd..)

- **Application 1:** Given old data about fraudulent customers and their background, predict whether new customer will commit fraud or not.

Class/
Target

Predictor attributes

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	**No**
2	No	Married	100K	**No**
3	No	Single	70K	**No**
4	Yes	Married	120K	**No**
5	No	Divorced	95K	**Yes**
6	No	Married	60K	**No**
7	Yes	Divorced	220K	**No**
8	No	Single	85K	**Yes**
9	No	Married	75K	**No**
10	No	Single	90K	**Yes**

Refund	Marital Status	Taxable Income	Cheat	Actual
No	Single	75K	?	No
Yes	Married	50K	?	Yes
No	Married	150K	?	No
Yes	Divorced	90K	?	No
No	Single	40K	?	Yes
No	Married	80K	?	No

Predicted by algorithm

Test Set

Training Set → Learn Classifier → Model

Predictive Analytics Techniques (contd..)

- **Application 2:** Given old data about customers and payments, predict new applicant's loan eligibility.

Previous customers

Age
Salary
Profession
Location
Customer type

Classifier

Decision rules

Salary > 5 L

Prof. = Exec

Loc = Mumbai

Custype = Eligible

Age > 40

New applicant's data

Class/ Target

Eligible/ Not Eligible

Predictive Analytics Techniques (contd..)

- **Association Rule Mining:** Finding frequent patterns:
  - Analysing item sets in customers basket or transactions and identifying frequently occurring item sets, which can be basis for recommendations.
  - Known as **Market Basket Analysis** in Retail.
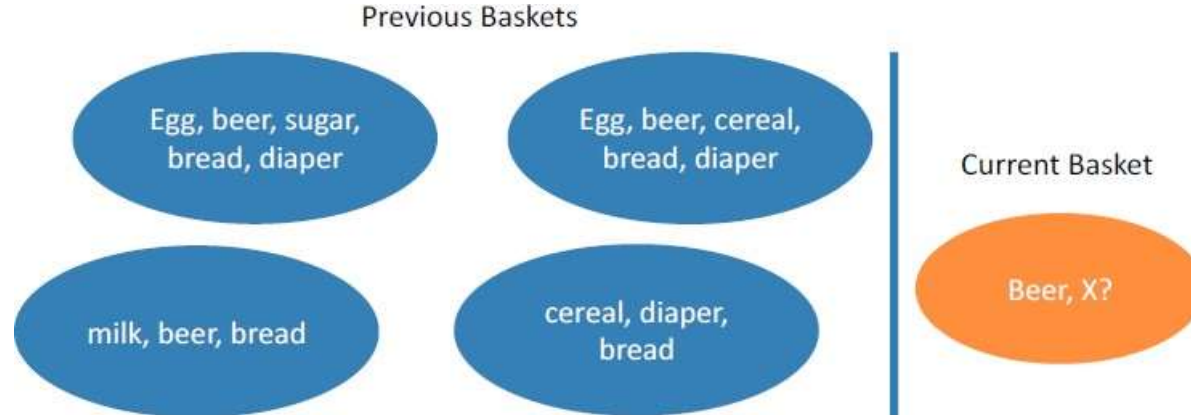
  - Applications:
    - Store layout: Place the frequently occurring items either in close proximity or as far apart as possible; give combo offers.
    - Warehousing and inventory management.

THANK YOU!