

# Data Mining & Data Warehouse

Associate Professor Dr. Raed Ibraheem Hamed

University of Human Development,  
College of Science and Technology  
Department of Computer Science

2016 – 2017

Department of CS- DM - UHD



# Introduction

---

- Motivation: Why data mining?
- What is data mining?
- Data Mining: On what kind of data?
- Data Mining and Business Intelligence
- Data, Information, Knowledge and Decision Making  
Action Cycle
- The Primary Tasks of Data Mining
- Basic Data Mining Tasks

# Why Data Mining?

---

- 1) The Explosive Growth of Data: from terabytes to petabytes.
- 2) We are drowning in data, but the strongest need for knowledge!
- 3) “Necessity is the mother of invention”—Data mining—Automated analysis of massive data sets

# What Is Data Mining?

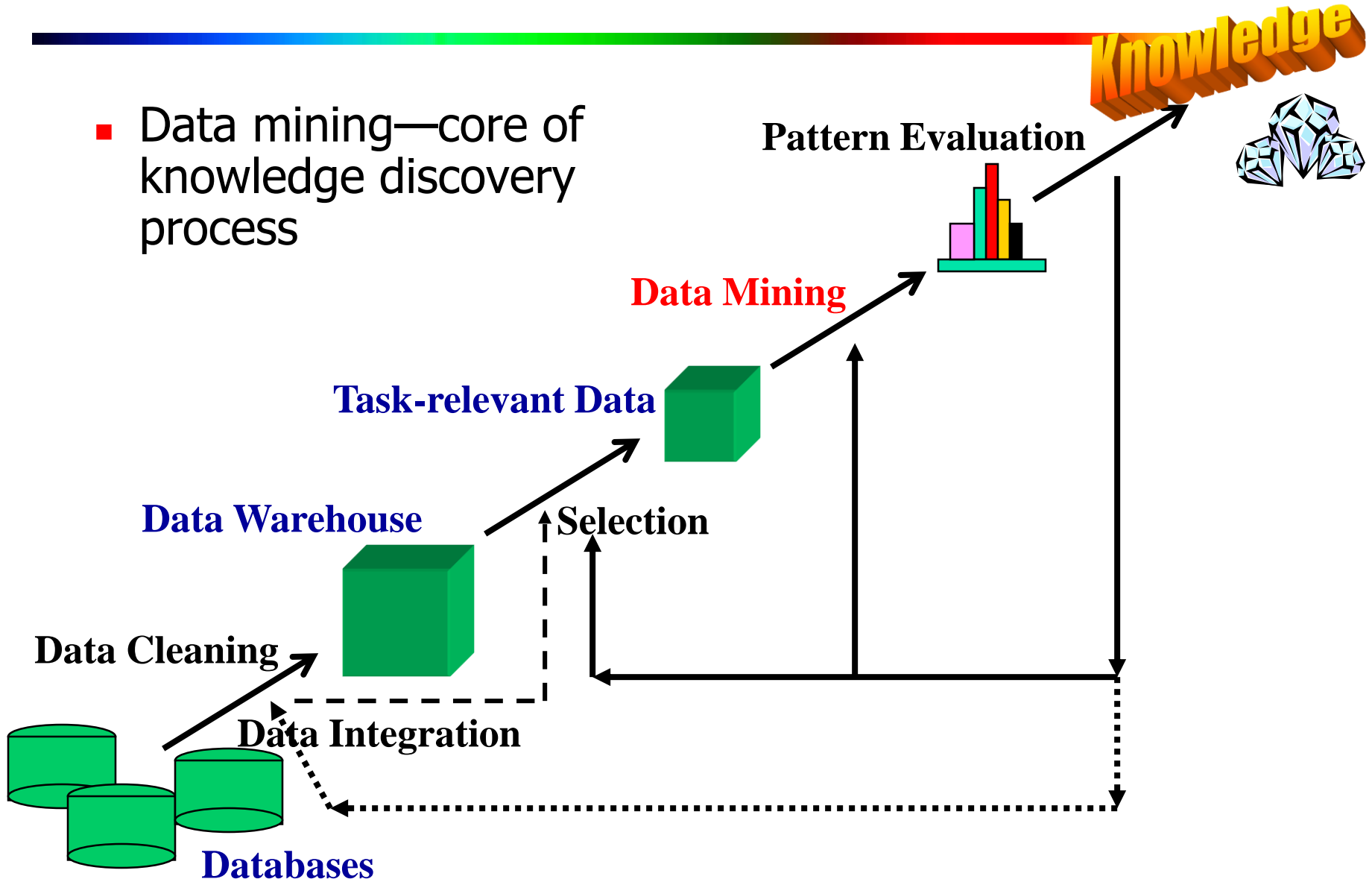


- Data mining (knowledge discovery from data)
  - Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data.
- **Alternative names**
  - Knowledge discovery in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, information gathering, business intelligence, etc.

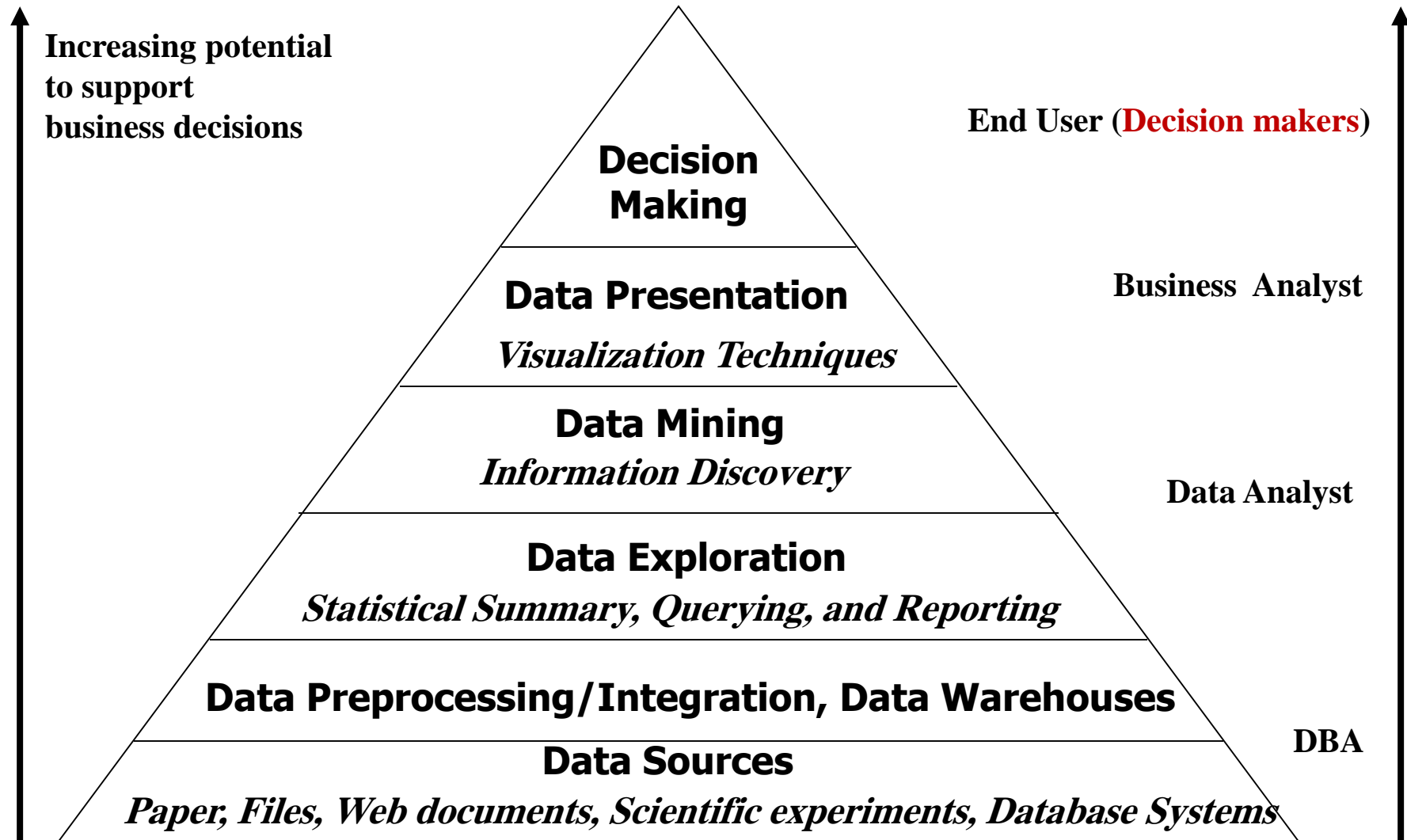


# Knowledge Discovery (KDD) Process

- Data mining—core of knowledge discovery process



# Data Mining and Business Intelligence

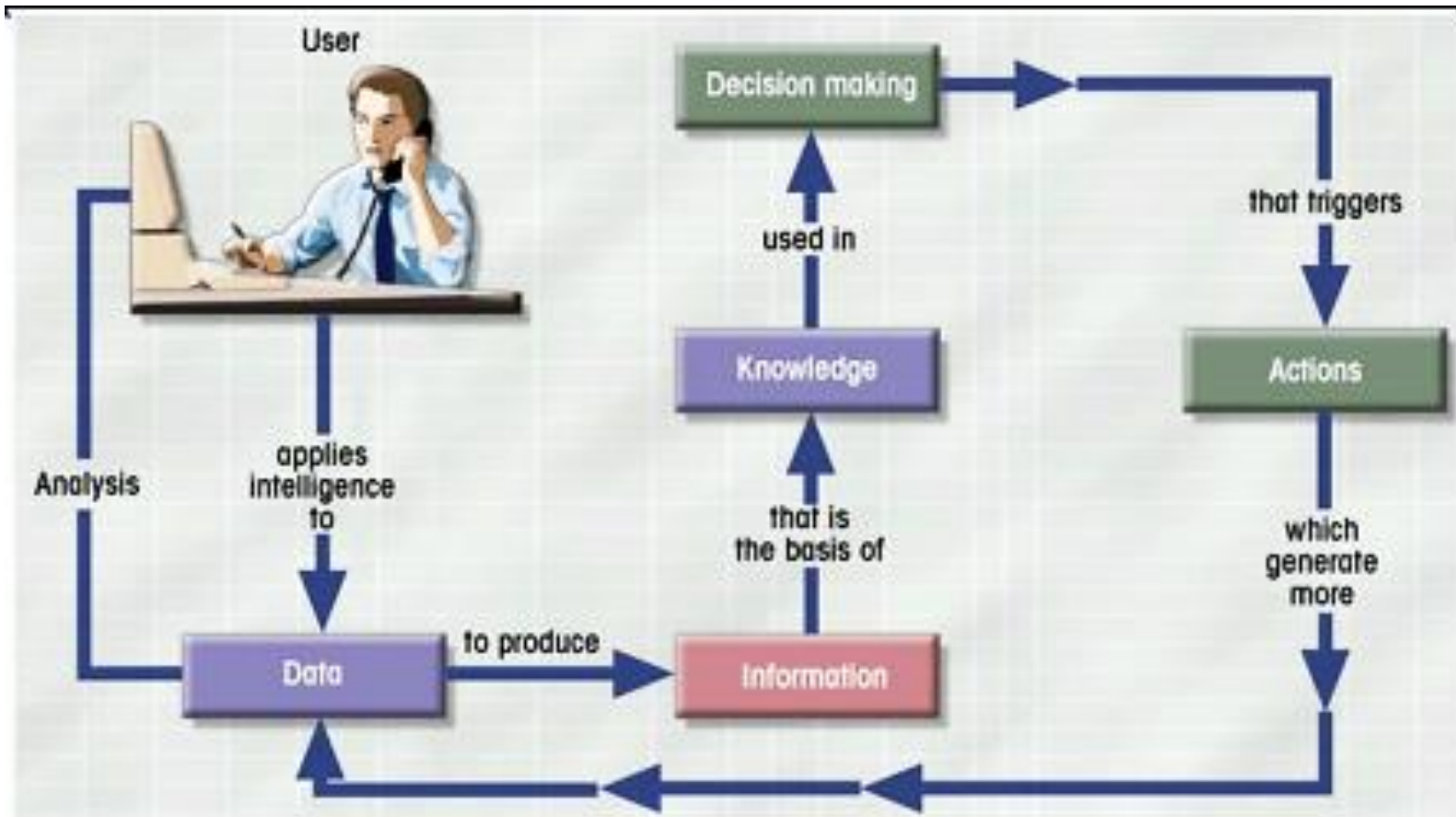


# Data Mining and Business Intelligence

---

- ✓ **DBA**: Gather the data from different sources like paper based data, transactional data, other form of data. DBA put different form of data in to same format and build a **Data warehouse**.
- ✓ **Data Analyst**: Once Data is ready, job role for data analyst is to understand the requirement from business users ( decision makers)
- ✓ **Business Analyst**: work on different tools to create a sophisticated reports and graphs which gives the end user flexibility to view the key processes within the organization.
- ✓ **End user**: Are also known as a **Decision makers** who view these ports and take the decisions about the business.

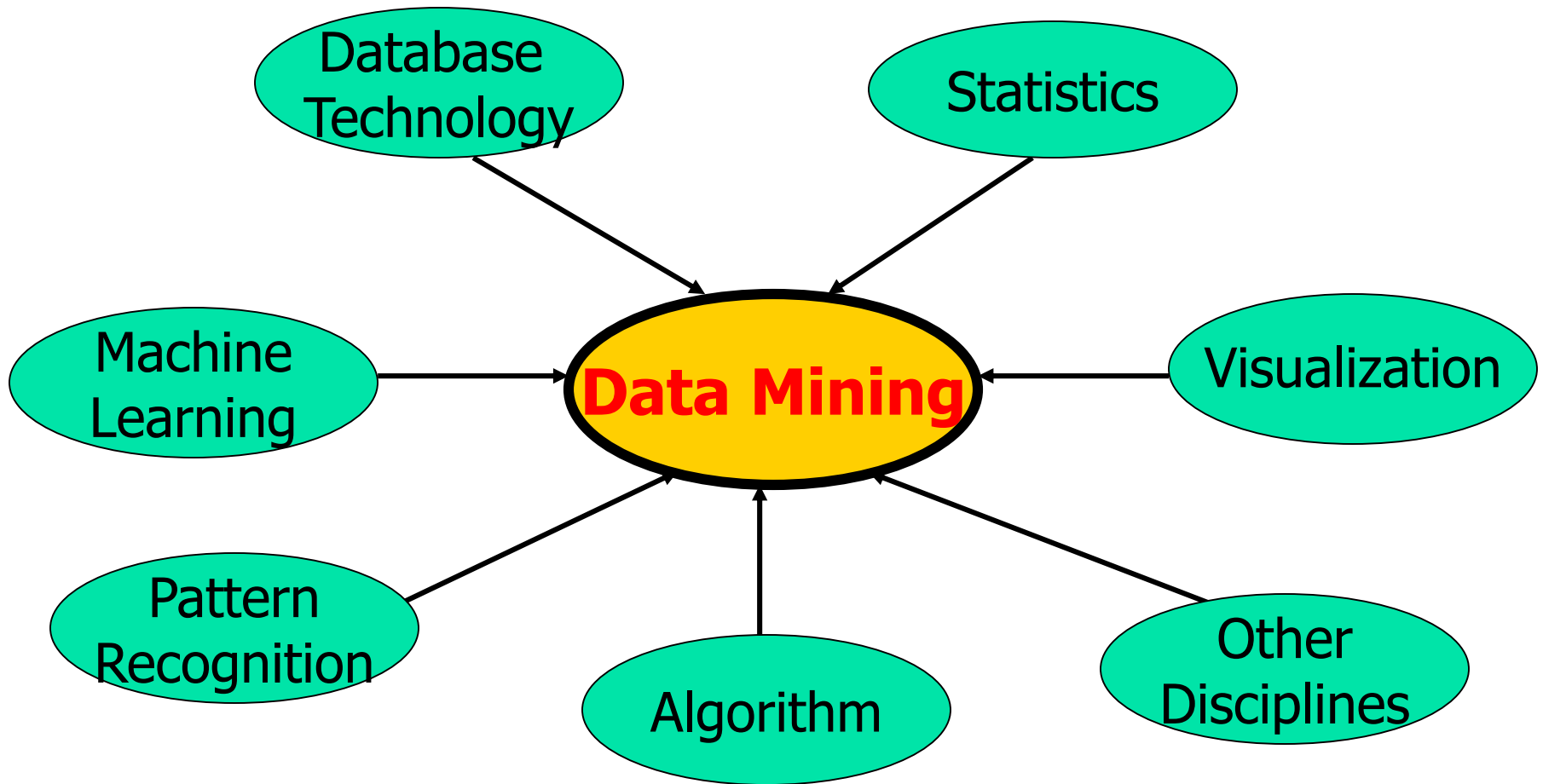
# Data, Information, Knowledge and Decision Making Action Cycle





# Data Mining: Confluence of Multiple Disciplines

---



# Why Not Traditional Data Analysis?

---

- Huge amount of data
  - Algorithms must be highly scalable to handle such as tera-bytes of data
- High-dimensionality of data
  - Micro-array may have tens of thousands of dimensions
- High complexity of data
  - Data streams and sensor data
  - Time-series data, temporal data, sequence data
  - Structure data, graphs, social networks and multi-linked data
  - Heterogeneous databases and inheritance databases
  - Multimedia, text and Web data
  - Software programs, scientific simulations
- New and sophisticated applications

*Thank  
you*

