

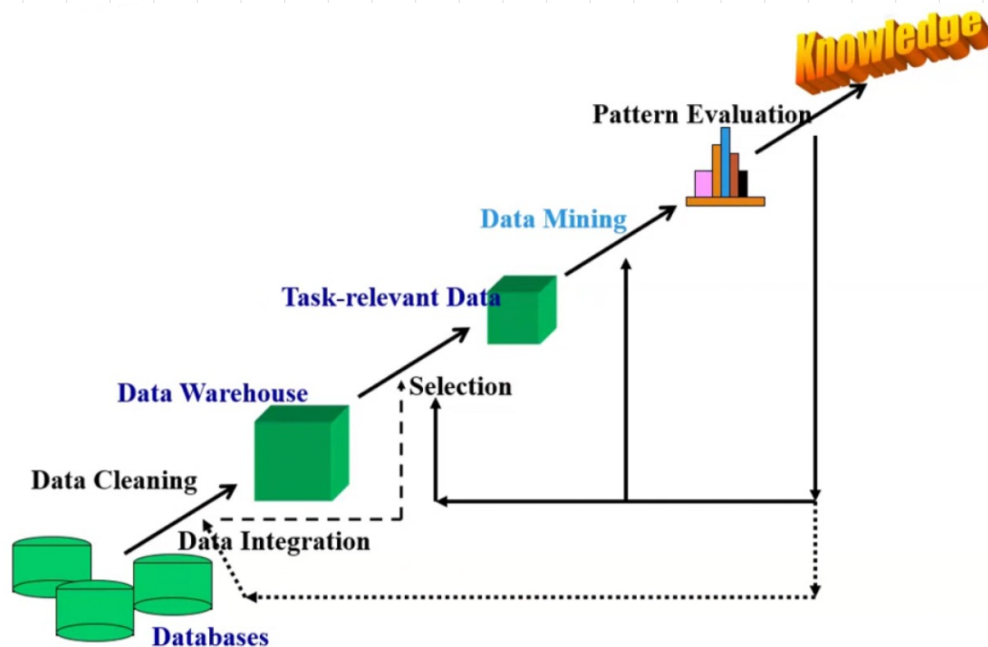
# Chapter 1. Introduction

**Data Mining** คือ การทำเหมืองข้อมูล ข้อมูลมีจำนวนเยอะมากในปัจจุบัน เนื่องจากทุกอย่าง  
สามารถเก็บข้อมูลได้ Ex. เห็นเซอร์ต่างๆ เว็บไซต์ ก็ดึงค่าข้อมูล ดังนั้น ถ้าคนที่มีความรู้เกี่ยวกับข้อมูลมากที่สุด  
คือ คนที่มี power มากที่สุด

ชื่อเต็มๆ "Knowledge discovery (mining) in databases (KDD)"

การสกัดองค์ความรู้จากฐานข้อมูล ที่เรามี นานาๆ นาน

## Knowledge Discovery (KDD) Process



1. Databases เราจะเก็บข้อมูลช่วงไหนให้มีประสิทธิภาพ
2. Data Cleaning มีการจัดการข้อมูลและทำความสะอาดข้อมูล
3. Data Integration เอาข้อมูลหลายๆแบบมารวมกัน เพื่อเก็บไว้ใน Data Warehouse
4. Data Warehouse คือข้อมูลที่จำเป็น ที่จะใช้ประโยชน์ต่างๆ จากฐานข้อมูลต่างๆมารวมกัน
5. Data Mining เลือกเอาข้อมูลที่จะสกัดเอาความรู้จริงๆ เอามาทำ Data Mining
6. **Pattern Evaluation** หลังจากทำ Data Mining จะได้รูปแบบที่ซ่อนอยู่ในข้อมูล
7. แล้วจากนั้นก็วัดผลว่าสิ่งที่สกัดออกมาคือ องค์ความรู้ที่แท้จริง แล้วจากนั้นจะได้เป็นองค์ความรู้

### Data Mining

- Pattern Discovery การสกัดหารูปแบบที่เกิดขึ้นในข้อมูล Association rule → Diaper → Beer  
สังเกตการซื้อของในใบเสร็จ คือ คนที่ซื้อผ้าอ้อม มักจะซื้อเปลี่ยนด้วย วิเคราะห์ว่าส่วนไหน จะเป็นจุดขาย ผู้ขายสามารถ  
นำ Pattern นี้ไปใช้ในการจัดการก็ได้
- Classification การจำแนกกลุ่ม เช่น กลุ่มคน ตามผลตอบรับนี้ หน้าที่ Classification คือหาข้อมูลไปทาบ  
ภาพรวมที่จะเป็นปกติไหม
- Cluster Analysis จัดกลุ่มข้อมูล โดยไม่สนใจว่าใครจะบอก ค้นหาข้อมูลที่คล้ายกัน มาอยู่ด้วยกัน