

## ความรู้เบื้องต้นเกี่ยวกับ คลังข้อมูล

## ข้อมูล (Data)

คือ ข้อเท็จจริงเกี่ยวกับสิ่งต่าง ๆ ซึ่งอาจเป็นคน สัตว์ สิ่งของ หรือเหตุการณ์ต่าง ๆ ที่ยังไม่ผ่านการประมวลผล โดยที่ข้อมูล อาจอยู่ในรูปของตัวเลข ตัวอักษร ข้อความ หรือ Multimedia ก็ได้

เนื่องจากข้อมูลเป็นส่วนประกอบที่สำคัญของระบบสารสนเทศรวมทั้งระบบสนับสนุนการตัดสินใจ จึงต้องเป็นข้อเท็จจริงที่ถูกต้อง ครบถ้วน สมบูรณ์ และน่าเชื่อถือ

## สารสนเทศ (Information)

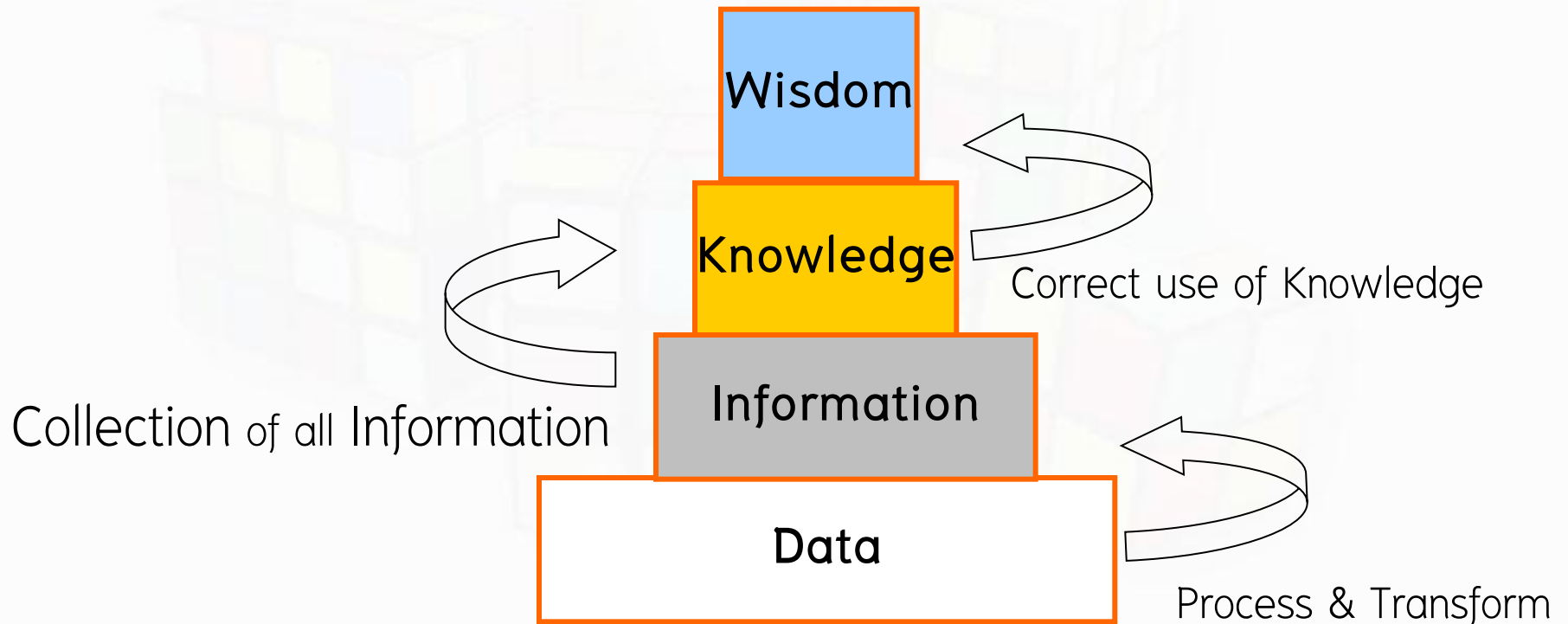
คือ ข้อมูลที่ผ่านการประมวลผล และจัดการให้มีความถูกต้อง  
ทันสมัย และสามารถนำไปใช้งานได้ตามที่ต้องการ



## องค์ความรู้ (Knowledge)

คือ สารสนเทศที่ถูกคัดเลือกเพื่อใช้ในการแก้ปัญหาต่าง ๆ ตามต้องการได้อย่างมีประสิทธิภาพ โดยอาศัยประสบการณ์ในการเลือกสารสนเทศมาใช้ในการแก้ไขปัญหาได้อย่างเหมาะสม

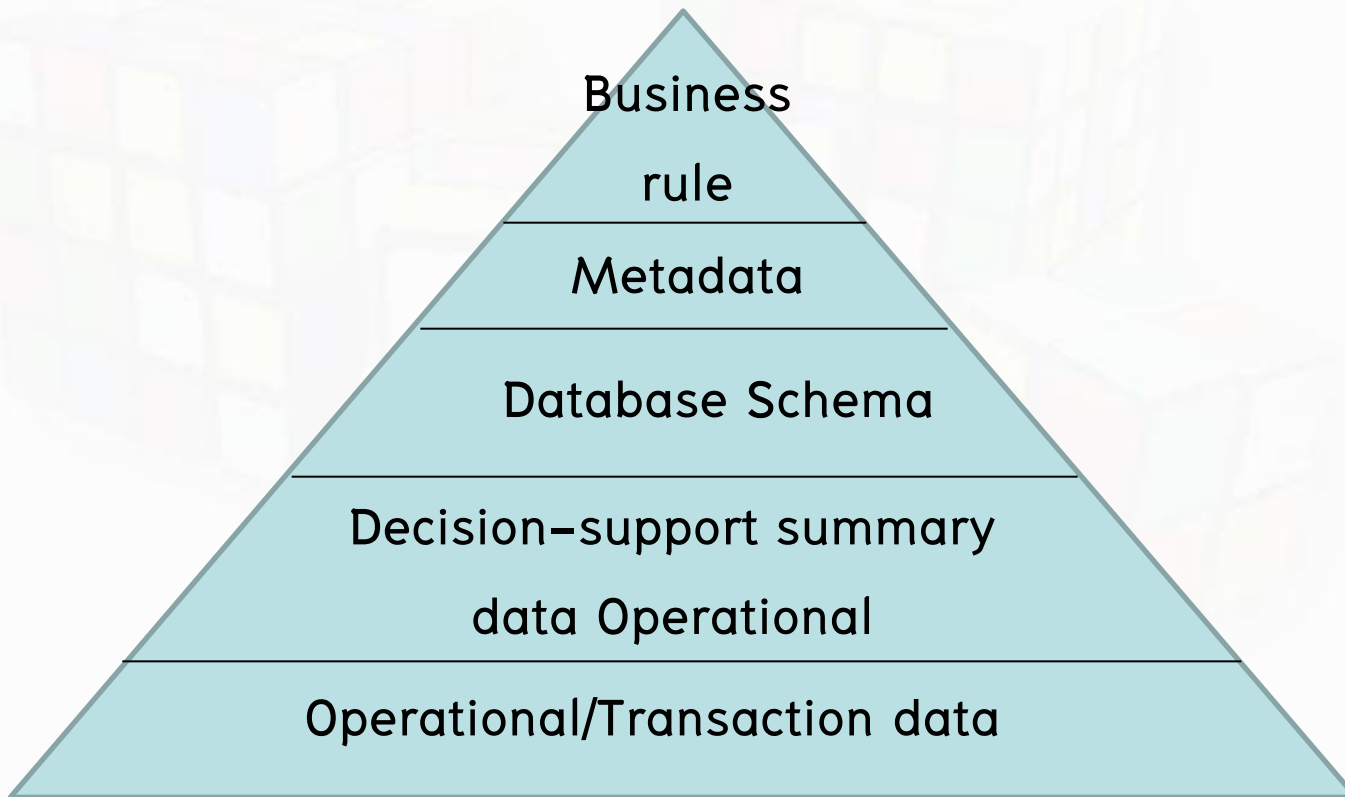
## องค์ความรู้ (Knowledge)



## สถาปัตยกรรมข้อมูล

- Operational/Transaction data
- Operational summary data
- Decision–support summary data
- Database Schema
- Metadata
- Business rule

## สถาปัตยกรรมข้อมูล (ต่อ)





## สถาปัตยกรรมข้อมูล (ต่อ)

### Transaction Data, The Base Level

ข้อมูลที่เป็นรายการเปลี่ยนแปลง หรือเป็นรายการแก้ไข (จากรายการใน แฟ้มข้อมูลหลัก) ตัวอย่างเช่นรายการซื้อ ขาย รายวัน รายการข้อมูลการโอนเงิน ของธนาคาร



## สถาปัตยกรรมข้อมูล (ต่อ)

### Operational Summary Data

ที่มีบทบาทเช่นเดียวกับ Transaction Data แต่แตกต่างกันคือ  
ได้ข้อมูลมาจาก Transaction มาทำการสรุปหรือนำข้อมูลของ  
คอลัมน์มาคำนวณหรือหาผลสรุป จะเก็บข้อมูลในรูปแบบ  
ข้อมูลสรุป เช่นระบบการชำระเงิน bill payments จะนำ  
คอลัมน์ยอดเงินแต่ละรายการมาคำนวณยอดรวมทั้งสิ้น หรือ  
ข้อมูลสรุปจำนวนสินค้าที่ขายทั้งหมด เป็นต้น

## สถาปัตยกรรมข้อมูล (ต่อ)

### Decision-Support summary Data

เป็นข้อมูลที่ใช้ในการตัดสินใจเกี่ยวกับธุรกิจ เช่น ข้อมูลที่เกี่ยวข้องกับทางการเงินจะใช้ดำเนินงานทางธุรกิจ ทำให้ต้องใช้ข้อมูลเชิงสรุปที่สนับสนุนประการการตัดสินใจ

## สถาปัตยกรรมข้อมูล (ต่อ)

Database Schema/Data model

โครงสร้างของข้อมูลหรือรูปแบบข้อมูล เพื่ออธิบาย  
ความสัมพันธ์ระหว่างข้อมูล มี 2 รูปแบบ

The Physical data Model

The Logical data Model

## สถาปัตยกรรมข้อมูล (ต่อ)

### Metadata

คือคำอธิบายรายละเอียดโครงสร้างของข้อมูล เช่น ประเภทของข้อมูลที่จัดเก็บในแต่ละคอลัมน์ ค่าและข้อกำหนดต่างๆในแต่ละคอลัมน์

## สถาปัตยกรรมข้อมูล (ต่อ)

### Business rules

เป็นกฎเกณฑ์หรือ นโยบายหรือกระบวนการพื้นฐานที่อธิบายถึงกิจการขององค์กรนั้นๆว่ามีกระบวนการขั้นตอนในการดำเนินงานอย่างไรบ้างอย่างถูกต้อง, ชัดเจนและเข้าใจได้ง่าย หรือเป็นข้อจำกัดและกฎเกณฑ์เงื่อนไขต่างๆทางธุรกิจที่ใช้ประกอบในการออกแบบการสร้าง Data Modeling

## คลังข้อมูลคืออะไร

คลังข้อมูลคือที่เก็บข้อมูลขององค์กรที่ได้รับการ  
ออกแบบเพื่อช่วยการตัดสินใจของฝ่ายบริหารในทางปฏิบัติ  
นั้น สิ่ง que เก็บอยู่ในคลังข้อมูลไม่ได้มีแต่เพียงข้อมูลเท่านั้น  
หากยังเก็บเครื่องมือสำหรับดำเนินการกับข้อมูล  
กระบวนการทำงานกับข้อมูล และทรัพยากรอื่นๆ



## คลังข้อมูลคืออะไร(ต่อ)

ระบบการจัดเก็บ และรวบรวมข้อมูลที่อยู่ในระบบปฏิบัติการต่าง ๆ ขององค์กร มักเป็นการนำเอาข้อมูลที่กระจัดกระจายเข้ามารวมไว้เป็นศูนย์กลางข้อมูลขององค์กร และสามารถเก็บข้อมูลย้อนหลังได้หลาย ๆ ปี ใช้เป็นข้อมูลการสนับสนุนการตัดสินใจ (Decision Support System)



## คลังข้อมูลคืออะไร(ต่อ)

คลังข้อมูล หมายถึง ฐานข้อมูลขนาดใหญ่ขององค์กรหรือหน่วยงานหนึ่งๆ ซึ่งเก็บรวบรวมข้อมูลจากฐานข้อมูลระบบงานประจำวัน หรือเรียกอีกอย่างว่า operational database และฐานข้อมูลอื่นภายนอกองค์กร หรือเรียกว่า external database โดยข้อมูลที่ถูกจัดเก็บในคลังข้อมูลนั้น มีวัตถุประสงค์ในการนำมาใช้งานและมีลักษณะของการจัดเก็บแตกต่างไปจากข้อมูลในฐานข้อมูลระบบงานอื่น โดยข้อมูลในคลังข้อมูลจะถูกนำมาใช้เพื่อสนับสนุนการตัดสินใจบริหารงานของผู้บริหาร โดยเฉพาะการเป็นข้อมูลพื้นฐานให้กับระบบงาน เพื่อการบริหารงานอื่น เช่น ระบบ DSS และระบบ CRM เป็นต้น

## คลังข้อมูลคืออะไร (ต่อ)

ระบบคลังข้อมูลเพื่อการบริหารได้ถูกออกแบบมาเพื่อใช้ในการเก็บข้อมูลขนาดใหญ่ โดยระบบข้อมูลเพื่อการบริหารนี้จะแยกข้อมูลออกจากฐานข้อมูลที่ใช้งานประจำวัน (Operational Database) ซึ่งข้อมูลสำหรับการบริหารโดยมากจะเป็นข้อมูลสรุป (Summary Data) ข้อมูลสรุปนี้อาจจะเป็นข้อมูลในอดีต ข้อมูลอ้างอิง หรือข้อมูล ณ ปัจจุบัน ซึ่งอาจได้มาจากข้อมูล Operational Database หรือมีการประมวลผลข้อมูลใน Operational Database ให้เป็นข้อมูลสรุป

## คลังข้อมูล

คลังข้อมูล ไม่ใช่ผลิตภัณฑ์ หรือระบบสำเร็จรูปที่สามารถหาซื้อมาใช้งานได้จากท้องตลาด แต่คลังข้อมูลเป็นหลักการ วิธีการ และแนวทางแก้ปัญหา เนื่องจากลักษณะงานขององค์กรแต่ละประเภทมีความแตกต่างกัน

## วัตถุประสงค์ของการสร้างคลังข้อมูล

วัตถุประสงค์ของการสร้างคลังข้อมูลคือ การแยกกลุ่มข้อมูลสารสนเทศที่ใช้ในการวิเคราะห์ทางธุรกิจออกจากฐานข้อมูลที่ใช้งานประจำวัน (Operational Database) มาเก็บอยู่ใน Relational Database Management Systems (RDBMS) ประสิทธิภาพสูง และทำให้การเรียกใช้ข้อมูลชุดนี้ทำได้อย่างยืดหยุ่น จากเครื่องมือที่อยู่บนเครื่องเดสก์ทอปทั่วไป โดยลด off-loading เพิ่มกลไกการช่วยตัดสินใจปรับปรุงเวลาที่ตอบสนอง (response time) รวดเร็วยิ่งขึ้นอย่างมากและผู้บริหารสามารถเรียกข้อมูลรายละเอียดที่จำเป็น ที่ถูกเก็บมาก่อนหน้านี้ (historical data) มาใช้ช่วยในการตัดสินใจทางธุรกิจแม่นยำขึ้น

## เป้าหมายในการสร้างคลังข้อมูล

1. คลังข้อมูลทำให้สามารถเข้าถึงข้อมูลขององค์กรได้ ผู้จัดการและนักวิเคราะห์ขององค์กรสามารถเชื่อมต่อเข้าไปยังคลังข้อมูลจากเครื่องคอมพิวเตอร์ของตนได้ ซึ่งการเชื่อมต่อสามารถทำได้ทันทีตามความต้องการและด้วยประสิทธิภาพสูง เครื่องมือที่มีให้กับผู้จัดการและนักวิเคราะห์ใช้งานง่าย สามารถออกรายงานได้ด้วยการคลิกปุ่มเดียว
2. ข้อมูลในคลังข้อมูลมีความถูกต้องตรงกันหมด คำถามเดียวกันต้องได้รับคำตอบที่เหมือนกันเสมอ ไม่ว่าผู้ถามจะเป็นใคร ถามเวลาใด



## เป้าหมายในการสร้างคลังข้อมูล

3. ข้อมูลในคลังข้อมูลสามารถถูกวิเคราะห์จากหัวข้อในธุรกิจประเภทนั้น โดยแบ่งข้อมูลหรือรวมข้อมูลมาวิเคราะห์ตามความต้องการ
4. คลังข้อมูลเป็นส่วนที่ผลิตข้อมูลจาก OLTP ข้อมูลไม่เพียงแต่ถูกรวบรวมมาไว้ที่ศูนย์กลางอย่างเดียว แต่จะถูกรวบรวมอย่างระมัดระวังจากแหล่ง ข้อมูลหลายๆแห่งนอกองค์กรด้วย แล้วมาปรับปรุงให้เหมาะสมกับการใช้งานเท่านั้น ถ้าข้อมูลเชื่อถือไม่ได้หรือไม่สมบูรณ์จะไม่ถูกอนุญาตให้นำ ไปใช้
5. คุณภาพของข้อมูลในคลังข้อมูลเป็นตัวผลักดันให้สามารถทำการ re-engineering ธุรกิจได้

## ประโยชน์ของระบบคลังข้อมูล

ในอดีตการนำข้อมูลที่มีอยู่ในลักษณะของ Operation Database มาใช้ในการวิเคราะห์นั้นมีปัญหาดังต่อไปนี้

1. บุคลากรทางด้าน Information Systems จำเป็นต้องเรียกข้อมูลจากฐานข้อมูลขนาดใหญ่ ซึ่งมีข้อมูลมากเกินไปเกินความต้องการ ส่งผลให้ประสิทธิภาพของ Transaction Operational Database ทำงานได้ช้าลง
2. ข้อมูลจะเป็นรูปแบบข้อมูลตารางเท่านั้น
3. ข้อมูลจะถูกนำเสนอในรูปแบบที่ตายตัว ไม่สามารถเปลี่ยนแปลงตามความต้องการของผู้ใช้



## ประโยชน์ของระบบคลังข้อมูล

4. ไม่ตอบสนองความต้องการของการตัดสินใจ เพราะข้อมูลสำหรับการตัดสินใจมีความสลับซับซ้อนสูง มีการรวมตัวกันของข้อมูลจากตารางต่างๆ หลายๆ ตาราง ข้อมูล ไม่ตอบสนองการสอบถามข้อมูล (Data Queries) สำหรับ ผู้ใช้
5. มีข้อมูลย้อนหลังน้อย (Historical Data)
6. ข้อมูลถูกจัดเก็บกระจายกระจายตามที่ตั้งต่างๆ ซึ่งยากต่อการเรียกใช้ หรือขาดความสัมพันธ์ทางธุรกิจอันอาจจะต้องเสียเวลาในการทำให้สอดคล้อง หรือเกิดความซ้ำซ้อนของข้อมูลได้

## ประโยชน์ของระบบคลังข้อมูล

จากปัญหาดังที่กล่าวมาข้างต้นคลังข้อมูลจึงได้ถูกออกแบบมาเพื่อตอบสนองงานในรูปแบบการตัดสินใจโดยการแยกฐานข้อมูลออกจาก Operational Database และเก็บข้อมูลในรูปแบบข้อมูลสรุป (Summary Data) ซึ่งข้อมูลสรุปนี้จะเลือกแต่เฉพาะข้อมูลที่มีความสำคัญต่อการตัดสินใจหรือเพื่อใช้ในการบริหารไปจนถึงการกำหนดแผนงานในอนาคต เช่น

## ประโยชน์ของระบบคลังข้อมูล

1. การเปรียบเทียบยอดขายระหว่างช่วงเวลาในอดีตกับปัจจุบันไปจนถึงการทำพยากรณ์ยอดขายในอดีต (Forecasting)
2. การหายอดขายสูงสุดหรือต่ำสุด
3. การเปรียบเทียบยอดขาย ต้นทุน กำไร ในรูปแบบตารางรายงานหรือรายงาน กราฟ

## ประโยชน์ของระบบคลังข้อมูล

ดังนั้นการจัดทำคลังข้อมูลจะมีความสำคัญมากขึ้นในอนาคต เพราะปัจจุบันนี้ผู้ใช้และผู้บริหารของหน่วยงานเริ่มมีเข้าใจความสำคัญของข้อมูลมากขึ้น และเริ่มตระหนักว่าหากนำข้อมูลมาวิเคราะห์ให้เข้าใจ สถานภาพหรือเหตุการณ์ที่เกิดขึ้นแล้วจะทำให้หน่วยงานหรือบริษัท สามารถตอบสนองต่อเหตุการณ์ที่เกิดขึ้นได้ดียิ่งขึ้น และจะทำให้หน่วยงานหรือบริษัททำงานบรรลุวัตถุประสงค์และเป้าหมายได้ดียิ่งขึ้น ตามไปด้วย

## คุณลักษณะของข้อมูลในคลังข้อมูล

**Subject Oriented** ข้อมูลจะต้องถูกสร้างขึ้นจากหัวข้อ (subject) ธุรกิจที่สนใจ เช่น ถ้าบริษัทประกันภัยต้องการใช้คลังข้อมูลฐานข้อมูลที่ได้จะต้องสร้างขึ้นจากประวัติลูกค้า, เบี้ยประกัน และการเรียกร้องแทนที่จะแยกตามชนิดของผลิตภัณฑ์ หรือบริการประกันภัย/ประกันชีวิต ข้อมูลที่สร้างขึ้นจะประกอบด้วยหัวข้อที่เก็บเฉพาะข่าวสารที่จำเป็น สำหรับกระบวนการตัดสินใจเท่านั้น

## คุณลักษณะของข้อมูลในคลังข้อมูล(ต่อ)

**Integrated** คือการรวบรวมข้อมูลจากหลายฐานข้อมูลปฏิบัติการเข้าด้วยกัน และทำให้ข้อมูลมีมาตรฐานเดียวกัน เช่น กำหนดให้มีค่าตัวแปรของข้อมูลในเนื้อหาเดียวกันให้เป็นแบบเดียวกันทั้งหมด

**Time-variant** ข้อมูลซึ่งใช้ตัดสินใจที่เก็บไว้จะต้องมีอายุประมาณ 5 ถึง 10 ปี เพื่อใช้เปรียบเทียบ หาแนวโน้ม และทำนายผลลัพธ์ในอนาคตได้

**Non-volatile** ข้อมูลจะไม่อัปเดตหรือถูกทำให้เปลี่ยนแปลงง่าย ๆ



## คลังข้อมูลแบ่งส่วนของข้อมูลออกเป็น 2 ส่วน คือ

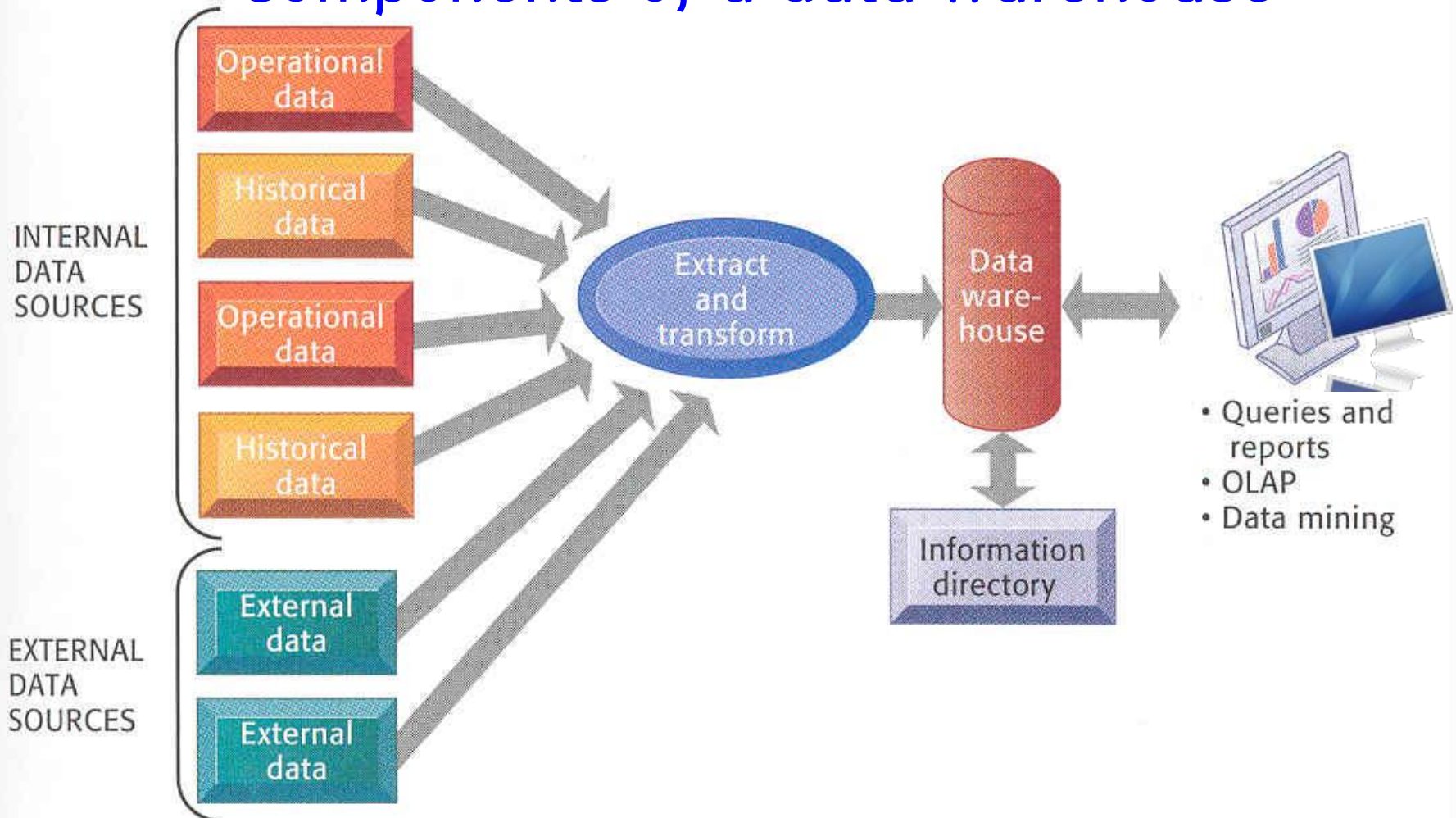
1. ข้อมูลเพื่อการปฏิบัติงาน เป็นข้อมูลที่เกิดจากการสั่งสมของกิจกรรมและผลการปฏิบัติงานขององค์กร การนำไปใช้งานต้องผ่านกระบวนการประมวลผลอย่างใดอย่างหนึ่ง
2. ข้อมูลเพื่อการวิเคราะห์ เป็นข้อมูลที่เกิดจากการใช้เครื่องมือที่มีอยู่ในการจัดการด้านคำนวณ และรวบรวมข้อมูลที่ประโยชน์ให้อยู่ในรูปแบบพร้อมใช้งาน



# Data Warehousing and Applications

Lect. Rattanawadee Panthong  
School of ICT

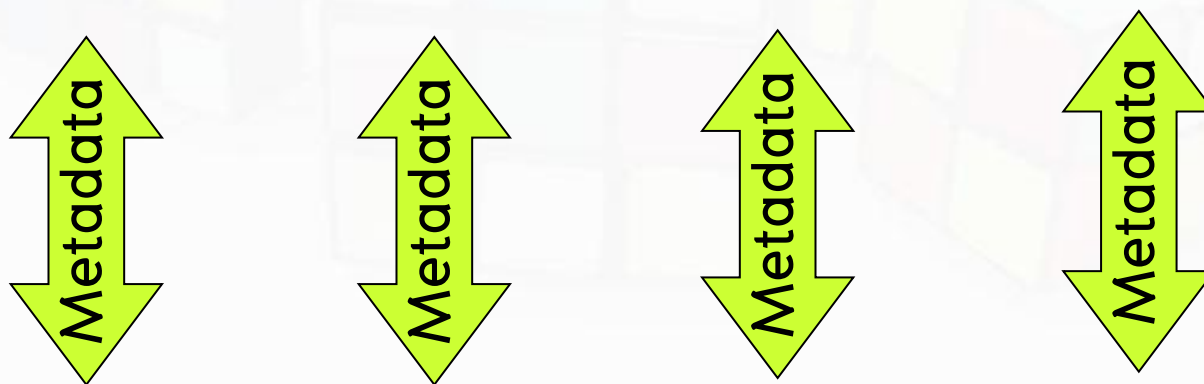
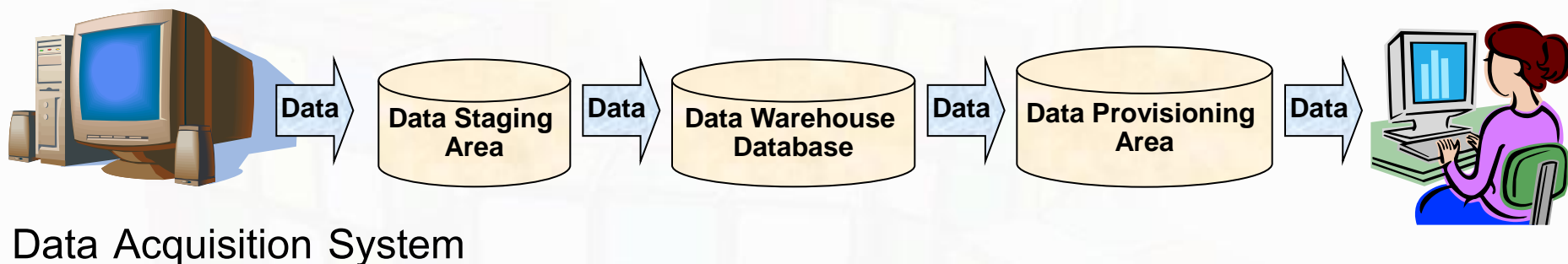
## Components of a data warehouse



## ส่วนประกอบของคลังข้อมูล

1. เครื่องมือสกัดแยกข้อมูล
2. ข้อมูลที่สกัดและแยกออกมาแล้ว
3. เมตาเดตาสำหรับบรรยายเนื้อหาข้อมูล
4. ฐานข้อมูลสำหรับคลังข้อมูล
5. เครื่องมือจัดการข้อมูลในคลังข้อมูล
6. โปรแกรมสำหรับจัดส่งข้อมูล
7. เครื่องมือวิเคราะห์สำหรับผู้ใช้
8. วัสดุและหลักสูตรการฝึกอบรม
9. ที่ปรึกษาด้านคลังข้อมูล

## สถาปัตยกรรมของคลังข้อมูล



Metadata Repository

## สถาปัตยกรรมของคลังข้อมูล

1. Data Acquisition System
2. Data Staging Area
3. Data Warehouse Database /Data Store
4. Data Provisioning Area /Data Mart
5. End User Terminal
6. Metadata Repository

## สถาปัตยกรรมของคลังข้อมูล

### 1. Data Acquisition System

จะทำหน้าที่รับข้อมูลจากภายนอก

### 2. Data Staging Area

Data Cleansing ลดความซ้ำซ้อนของข้อมูล

Filtering เลือกเฉพาะข้อมูลที่มีประโยชน์



## สถาปัตยกรรมของคลังข้อมูล

### 3. Data Warehouse Database /Data Store

Data Model การออกแบบจำลองข้อมูล การจัดเก็บข้อมูล

### 4. Data Provisioning Area / Data Mart

การรายงานผลลัพธ์ต่างๆที่จำเป็นสำหรับการวิเคราะห์ข้อมูล



## สถาปัตยกรรมของคลังข้อมูล

### 5. End User Terminal

เป็นส่วนที่ทำหน้าที่ดึงเอาข้อมูลที่ได้เตรียมไว้ใน Data Provisioning Area เพื่อนำเสนอผลลัพธ์ โดยจะมีเครื่องมือในการออกรายงานดังนี้

- Simple Report Tool
- Multi Dimensional Tools
- Data Mining Tools

## สถาปัตยกรรมของคลังข้อมูล

### 6. Metadata Repository

ใช้เก็บข้อมูลที่ใช้ควบคุมการทำงานในคลังข้อมูล โดย Metadata จะมีข้อมูลที่เกี่ยวข้องกับข้อมูลต่างๆใน Data Acquisition, System Data Staging, Area Data Warehouse และ Database Data Provisioning Area

## แหล่งข้อมูล (Data Source)

- แหล่งข้อมูลภายในองค์กร (Internal Data Source)
- แหล่งข้อมูลภายนอกองค์กร (External Data Source)

## แหล่งข้อมูล (Data Source)

- แหล่งข้อมูลภายในองค์กร (Internal Data Source)

คือแหล่งข้อมูลทั้งหมดภายในองค์กร เช่น แอปพลิเคชันต่างๆ ผู้บริหาร เจ้าหน้าที่ผู้ชำนาญการ หัวหน้างาน สินค้า พนักงาน หรือขั้นตอนการทำงาน

## แหล่งข้อมูล (Data Source)

- แหล่งข้อมูลภายนอกองค์กร (External Data Source)  
คือแหล่งข้อมูลทั้งหมดภายนอกองค์กร เช่นหน่วยงานรัฐบาล  
เอกชน สถาบัน เป็นต้น