

# บทที่ 2

## การประมวลผลข้อมูล (Data Processing)

# Outline

- ความหมาย
- ประเภทของการประมวลผล
- การประมวลผลข้อมูลด้วยคอมพิวเตอร์
- องค์ประกอบของข้อมูล

# การประมวลผล (Processing)

- การดำเนินการกับข้อมูลดิบ (Raw Data) เพื่อให้ได้สารสนเทศ (Information)



# ข้อมูลดิบ

- ข้อมูลที่รวบรวมมาจากแหล่งข้อมูลต่าง ๆ
- ตัวเลข ตัวอักษร รูปภาพ เสียง อุณหภูมิ ความเร็ว
- คะแนนสอบของนักศึกษา 1 กลุ่ม

# ข้อมูลสารสนเทศ

- ข้อมูลที่ผ่านการประมวลผลแล้ว สามารถนำไปใช้ประโยชน์ได้
- ค่าเฉลี่ยคะแนนสอบของนักศึกษา 1 กลุ่ม

# ประเภทของการประมวลผล

- การประมวลผลด้วยมือ (Manual Data Processing)
- การประมวลผลอัตโนมัติ (Automatic Data Processing)
  - Online Processing
  - Batch Processing

# Online Processing

- คือการประมวลผลที่ได้ผลลัพธ์ทันที (Real-time) เช่น ระบบสำรองที่นั่งในเครื่องบิน ฝาก/ถอน ATM
- ไม่สามารถคาดการณ์ได้ว่าจะต้องประมวลผลเมื่อใด เช่น เครื่อง ATM ไม่รู้ว่าจะมีคนมากดเงินเวลาไหนบ้าง
- มักเป็นการประมวลผลแบบ Time Sharing คือมีผู้ใช้เข้าสู่ระบบประมวลผลมากกว่า 1 คนในขณะเดียวกัน
- ใช้สื่อที่มีการเข้าถึงแบบโดยตรง เช่น ฮาร์ดดิสก์

# Online Processing (ต่อ)

- ข้อมูลที่ใช้ประมวลผลต้องเป็นปัจจุบัน (Up to date)
- สามารถทำรายงานได้ทันที ได้ผลลัพธ์ที่เป็นปัจจุบัน
- จากการทำงานที่กล่าวมา จึงต้องมี I/O Device ที่ต้องเชื่อมต่อ กับ CPU ตลอดเวลา (Online)



# Batch Processing

## การประมวลผลแบบ Batch Processing

ตัวอย่าง: การ Add-Drop รายวิชา

- ไฟล์หลัก (Master Files): มีการเปลี่ยนแปลงน้อย
  - ไฟล์รายการ ข้อมูลรายวิชา
- ไฟล์รายการเปลี่ยนแปลง (Transaction Files): มีการเปลี่ยนแปลงบ่อย
  - รายการ เพิ่มรายวิชา (Add)
  - รายการ ถอนรายวิชา (Drop)

ซึ่งสำนักทะเบียนจะมีการประมวลวันละ 2 รอบ

# Batch Processing (ต่อ)

- ประมวลผลตามรายการเปลี่ยนแปลงที่รวบรวมเป็นชุด (Batch) แล้วประมวลผลตามคาบเวลาที่ละรายการตั้งแต่ระเบียบแรกจนหมด Batch นั้นๆ
- ประมวลผลตามคาบเวลา (Period of time) ทุกวัน/ทุกเดือน
- เป็นการประมวลผลแบบลำดับ (Sequential)
- สามารถใช้สื่อที่มีการเข้าถึงข้อมูลเป็นแบบลำดับได้ เช่น เทปแม่เหล็ก จึงทำให้มีค่าใช้จ่ายต่ำ

# Batch Processing (ต่อ)

- มีการจัดเก็บระเบียบข้อมูลเพื่อรอการประมวลผลตามคาบ เวลา จึงเกิดภาวะข้อมูลไม่ทันต่อเหตุการณ์ (ไม่ Update) เช่น เก็บข้อมูลขอ drop-add แล้วประมวลผลพร้อมกันภายหลัง
- รายงานหรือผลลัพธ์ที่ได้ล่าช้าไม่ทันต่อเหตุการณ์ (Past Information) เช่น รายงานยอดขายสินค้าเมื่อสิ้นเดือนที่แล้ว
- จากการประมวลผลที่กล่าวมา I/O Device จึงไม่จำเป็นต้องเชื่อมต่อกับ CPU ตลอดเวลา

# การประมวลผลด้วยคอมพิวเตอร์

- การเก็บข้อมูล (Data Collection)
- การประมวลผลข้อมูล (Data Processing)
- ผลลัพธ์ (Output / Information)



# การเก็บข้อมูล

- วิธีการเก็บข้อมูล
- ลักษณะของข้อมูลที่ดี
- การเข้ารหัสข้อมูล
- การตรวจสอบความถูกต้องของข้อมูล
- การเก็บข้อมูลในสื่อบันทึกข้อมูล

# วิธีการเก็บข้อมูล

- สังเกต (Observation)
- วัดค่า (Measurement)
- เอกสาร (Document)
- สัมภาษณ์ (Interview)
- แบบสอบถาม (Questionnaire)
  - Manual
  - Internet

# ลักษณะของข้อมูลที่ดี

- ถูกต้อง (Accuracy)
- ปัจจุบัน (Updated)
- สมบูรณ์ (Completeness)

# การแทนข้อมูลด้วยรหัส

- ประหยัดเนื้อที่ในการเก็บข้อมูล เพราะเก็บรหัสแทนข้อมูล
- มีความเร็วในการค้นหาข้อมูลชนิดรหัสสูงกว่าข้อมูลเดิม
- ใช้ตัวเลข หรืออักขระ เป็นรหัสแทนข้อมูล
- ข้อมูลที่จะเข้ารหัสได้ต้องถูกจัดในกลุ่มที่รู้จำนวน เช่น ข้อมูลเพศมี 2 กลุ่ม คือ ญ/ช จะใช้ตัวอักขระ 1 ตัวแทนข้อมูล หรือข้อมูลคณะของ ม.ช.ที่มีมากกว่า 10 คณะ จะใช้อักขระตัวเลข 2 หลักแทน (รหัส 00-99 มีถึง 100 รหัส)



# การแทนข้อมูลด้วยรหัส

นาย	สมชาย	คนไทย	เชียงใหม่	12,000 บาท
-----	-------	-------	-----------	------------



M	สมชาย	คนไทย	053	12
---	-------	-------	-----	----

การเข้ารหัสจะทำกับข้อมูลที่มีซ้ำกันมากๆ จะได้ประโยชน์

# การตรวจสอบความถูกต้องของข้อมูล

- Edit/ Update การแก้ไข/ปรับปรุง:เพิ่ม ลบ เปลี่ยนแปลง
- Verify การทวนสอบขั้นตอนการนำเข้าข้อมูล
- Validation ความถูกต้องอย่างสมเหตุสมผลของข้อมูล
  - Range Check เป็นการเช็คที่เกี่ยวข้งกับข้อมูล 1 field เช่น field คำตอบ มีข้อมูลคำตอบแบบสอบถาม 4 ตัวเลือก ข้อมูล ที่เป็น คำตอบ ต้องเป็น ก-ง เท่านั้น
  - Relation Check เป็นการเช็คข้อมูลที่มีสัมพันธ์กัน 2 field เช่น ข้อมูล เงินเดือนตำแหน่ง programmer ของบริษัท ก ไม่ต่ำกว่า 10,000.- บาท หรือข้อมูล คำนำหน้าชื่อ นาย ต้องมีข้อมูล เพศเป็น ชาย

# การเก็บข้อมูลในสื่อบันทึกข้อมูล

- ข้อมูลจริง หมายถึงข้อมูลที่เกิดขึ้นจริง จากการทำงาน
- การเปลี่ยนสื่อบันทึกข้อมูล หมายถึงการนำข้อมูลจริงบันทึกลงสื่อคอมพิวเตอร์ที่เหมาะสมเพื่อการประมวลผลข้อมูลต่อไป หรือการนำข้อมูลไปบันทึกยังสื่ออื่น เช่น การสำรองข้อมูล (Back up)
- ข้อมูลในสื่อบันทึกข้อมูล หมายถึงการนำข้อมูลจริงบันทึกลงในสื่อคอมพิวเตอร์ตามหลักการหรือวิธีการของ file หรือฐานข้อมูล (Database)

# ข้อมูลจริง

ชื่อ	แผนก	ชั่วโมง	อัตรา
นายสมชาย คนไทย	คอมพิวเตอร์	12	60
นางสาวแพร มีบุญ	บัญชี	20	50
นางศศิวิมล คนดี	ประชาสัมพันธ์ 5		80
...	...	...	...

# ข้อมูลในสื่อบันทึกข้อมูล

M สมชาย คนไทย	1	12	60
N แพร มีบุญ	2	20	50
O ศศิวิมล คนดี	3	5	80
...	...	...	...

↑  
ระเบียน (Record)

↑  
ค่าแอททริบิวต์  
(Attribute)

# การประมวลผลข้อมูล

- คำนวณ (Calculating)
- ปรับปรุง (Updating)
- รายงาน (Reporting) / เอกสาร (Document)
- สรุป (Summarizing)
- เรียงลำดับ (Sorting)
- ค้นหา (Searching)
- แบ่งกลุ่ม (Grouping)
- แบ่งประเภท (Classification)

- นำกลุ่มของค่าแอททริบิวต์ที่มีอยู่/จัดเก็บ มาคำนวณหาค่าต่างๆ ตามที่ต้องการ
  - หาค่าเฉลี่ย (Mean)
  - หาค่ามัธยฐาน (Median)
  - หาค่าฐานนิยม (Mode)

# ปรับปรุง(Update)

- นำค่าแอททริบิวต์ที่มีอยู่มาคำนวณ แล้วแทนที่
- การปรับค่าอัตราเงินเดือนของพนักงานขึ้นจากเดิมอีก 20 %
- การปรับค่าเงินสุทธิของบัญชีธนาคารของลูกค้าแต่ละบัญชี ด้วยการนำค่าสุทธิเดิม – จำนวนเงินที่ถอนวันนี้ หรือ +จำนวนเงินที่ฝากวันนี้



# รายงาน

ชื่อ	แผนก	อัตรา	ชั่วโมง	ค่าจ้าง
นายสมชาย คนไทย	คอมพิวเตอร์	60	12	720
นางสาวแพร มีบุญ	บัญชี	50	20	1,000
นางศศิวิมล คนดี	ประชาสัมพันธ์	80	5	400
...	...	...	...	...

# สรุปผล

ชื่อ	แผนก	อัตรา	ชั่วโมง	ค่าจ้าง
นายสมชาย คนไทย	คอมพิวเตอร์	60	12	720
นางสาวแพร มีบุญ	บัญชี	50	20	1,000
นางศศิวิมล คนดี	ประชาสัมพันธ์	80	5	400
...	...	...	...	...
			72	5,140

# สรุปผล (ต่อ)

แผนก	อัตรา	ชั่วโมง	ค่าจ้าง
คอมพิวเตอร์	60	22	1,320
บัญชี	50	33	1,650
ประชาสัมพันธ์	80	15	1,200
		70	4,170

# เรียงลำดับ

- ตัวเลข      เช่น คะแนนสอบ
- ข้อความ    เช่น ชื่อตามอักขระ ก ข ค
- รหัส        เช่น รหัสสินค้า

# ค้นหา

- ค้นหากระเป๋านที่ต้องการจากเงื่อนไขที่กำหนด
- ค้นหานักเรียนที่  $GPA < 1.75$
- ค้นหาพนักงานที่ เงินเดือน  $\leq 5,000$  หรือ อยู่ฝ่ายบัญชี
- ค้นหาสินค้าที่ ราคา = 100 และ เป็นอาหาร

# แบ่งกลุ่ม

- แบ่งระเบียบยันท้งหมดออกเป็นกลุ่มจากเงื่อนไขที่กำหนด
- แบ่งแบบสุ่ม เช่น จับฉลากแบ่งกลุ่มนักเรียน
- แบ่งแบบน้ำหนักของแต่ละกลุ่มมีค่าใกล้เคียงกัน เช่น แบ่งกลุ่มนักเรียนตาม GPA โดยที่แต่ละกลุ่มมีคนเรียนเก่งและเรียนอ่อนพอ ๆ กัน

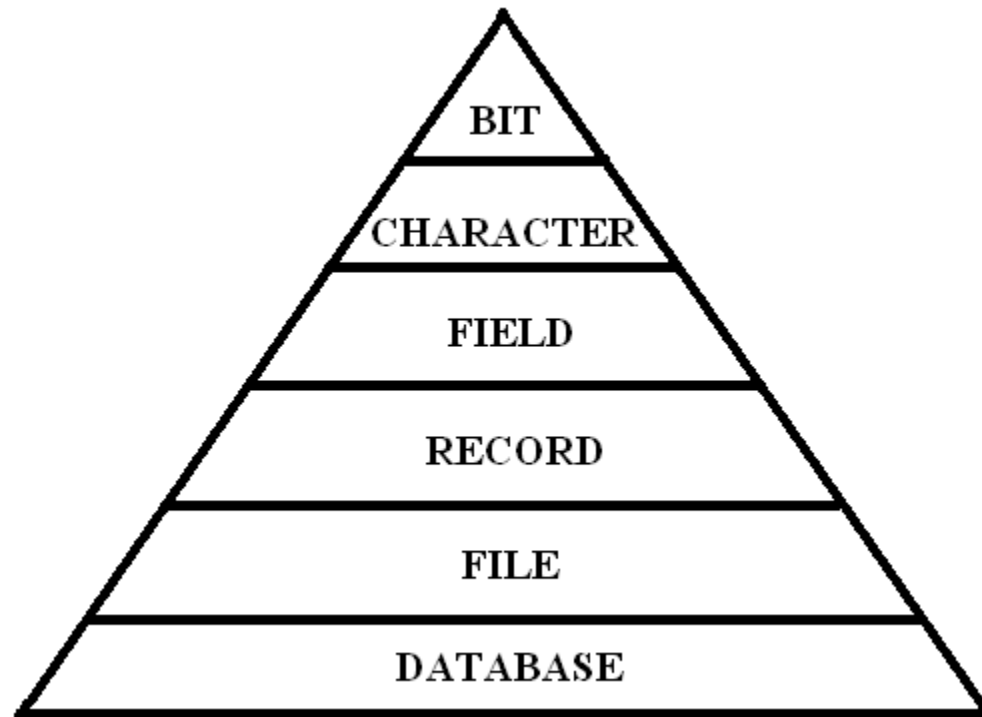
- แสดงผล แล้วนำผลลัพธ์ไปใช้เลย
- Feedback กลับไปแก้ไขได้ในกรณีที่มีข้อผิดพลาด คือมีการพิจารณา Output เพื่อการประมวลผลอีก
  - กลับไปเก็บข้อมูลใหม่
  - กลับไปการประมวลผลข้อมูลใหม่

# องค์ประกอบข้อมูล (Data Organization)

- บิต (Bit)
- อักขระ (Character)
- ขอบเขตข้อมูล (Field)
- ระเบียน (Record)
- แฟ้มข้อมูล (File)
- ฐานข้อมูล (Database)



# องค์ประกอบข้อมูล (ต่อ)



# บิต (Bit)

- ตัวเลข 0 และ 1
- หน่วยพื้นฐานที่จัดเก็บในคอมพิวเตอร์

# อักขระ (Character)

- กลุ่มของบิต เช่น 011101011
- 1 อักขระ นับเป็น 1 ไบต์
- ในปัจจุบัน โดยทั่วไป 1 ไบต์ (อักขระ) มี 8 บิต
- เช่น 01101011 แทน F เป็นต้น

# ขอบเขตข้อมูล (Field)

- กลุ่มของอักขระหรือข้อความ (String) เช่น Cat
- เลขจำนวน (Number) คือข้อมูลที่นำมาประมวลผลทางคณิตศาสตร์ได้ เช่น เลขจำนวนเต็ม (Integer) เช่น 1,004 หรือ เลขจำนวนจริง (Real) เช่น 3.25
- อักขระเดี่ยว (Character) หมายถึง อักขระ A-Z, 0-9, สัญลักษณ์พิเศษ เช่น \*

# ทะเบียน (Record)

- กลุ่มของขอบเขตข้อมูลที่เกี่ยวข้องกัน

นายพร มีบุญ	แผนกบัญชี	20 ชั่วโมง	50 บาท
-------------	-----------	------------	--------

# แฟ้มข้อมูล (File)

- กลุ่มของระเบียบที่มีรูปแบบเดียวกัน สีฟ้า คือ field name/  
attribute name สีดำคือ ข้อมูล

ชื่อ	แผนก	ชั่วโมงทำงาน	วันลา	เวลาเริ่มงาน
นายพร มีบุญ	คอมพิวเตอร์	35	10	10.00
นางสาวศศิวิมล คนดี	คอมพิวเตอร์	35	10	10.00
นาง ปณิตดา ดวงคำ	ประชาสัมพันธ์	40	15	8.00

# ฐานข้อมูล (ต่อ)

- กลุ่มของแฟ้มข้อมูล/ตารางข้อมูลที่สัมพันธ์กัน เช่นฐานข้อมูลพนักงาน มี 2 ตารางข้อมูล ดังนี้

ชื่อ	แผนก
นายพร มีบุญ	คอมพิวเตอร์
นางสาวศศิวิมล คนดี	คอมพิวเตอร์
นาง ปณิตดา ดวงคำ	ประชาสัมพันธ์

แผนก	ชั่วโมงทำงาน	วันลา	เวลาเริ่มงาน
คอมพิวเตอร์	35	10	10.00
ประชาสัมพันธ์	40	15	8.00

# แฟ้มข้อมูล VS ฐานข้อมูล

- ลักษณะของแฟ้มข้อมูล
- ลักษณะของฐานข้อมูล



# ลักษณะของแฟ้มข้อมูล

- โปรแกรมต้องรู้โครงสร้างของแฟ้มข้อมูล (Data dependence)
- ค่าใช้จ่ายน้อยเหมาะสำหรับการประมวลผลงานขนาดเล็ก
- ใช้บันทึกข้อมูลเรื่องใดเรื่องหนึ่งเท่านั้น
- โดยปกติใช้กับงานเพียงด้านเดียวเท่านั้น
- ถ้าใช้หลายที่ในหน่วยงานข้อมูลจะมีความซ้ำซ้อน

# ลักษณะฐานข้อมูล

- ข้อมูลเป็นอิสระจากโปรแกรม(Data Independence) ใช้โปรแกรมภาษา 4 GL เช่น Access, MySQL ซึ่งเรียกว่าเป็นโปรแกรมจัดการฐานข้อมูล
- ข้อมูลมีโครงสร้างที่แน่นอน เหมาะกับองค์กรที่มีข้อมูลมาก จึงทำให้เกิดค่าใช้จ่ายสูง
- มีชุดเดียวใช้ได้ทั้งหน่วยงาน (Shared Data)
- มีภาษาสอบถาม (Query Language) ใช้ง่าย สะดวก