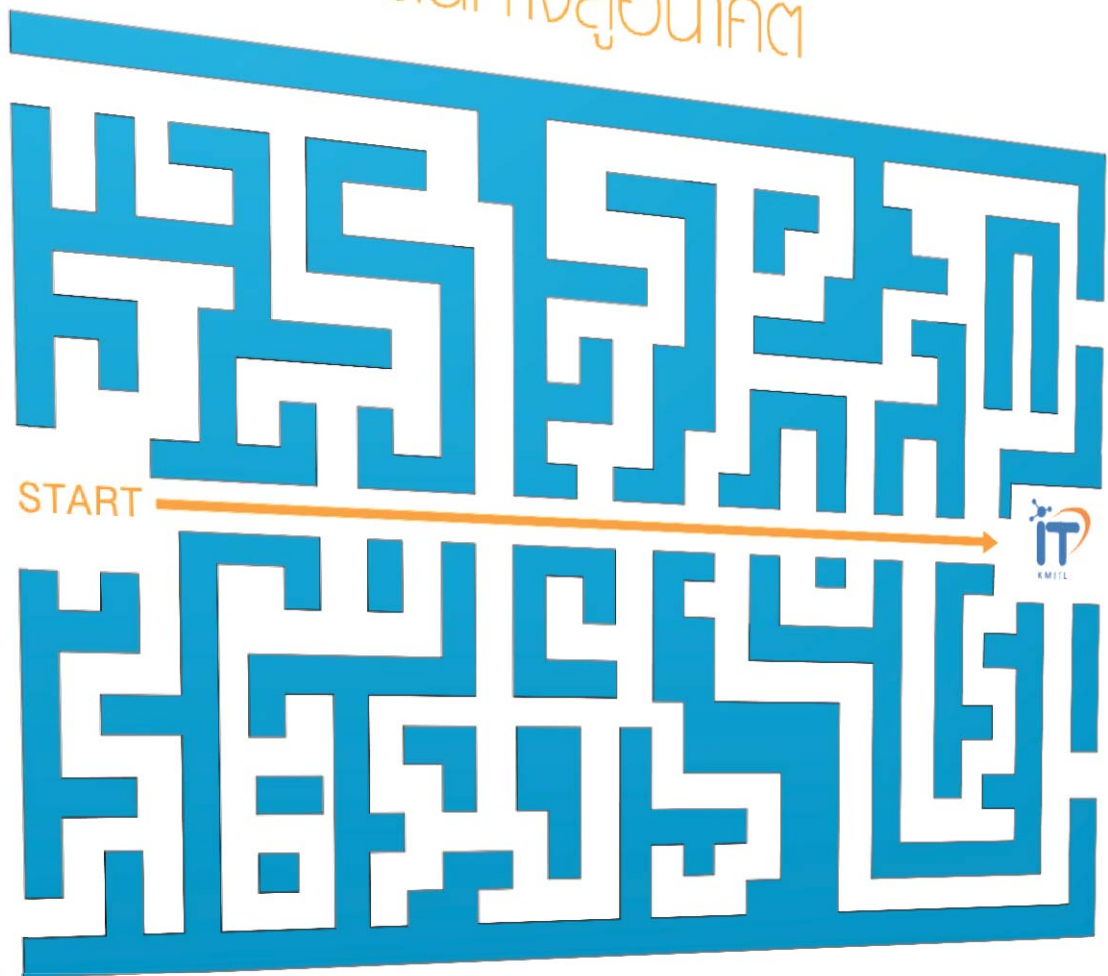


# IT ตีวน้อง '54 สอบตรง

เส้นทางสู่อนาคต



## DATABASE

by P'Golf, P'Bank

## DATABASE (ฐานข้อมูล)

### ความหมายของข้อมูล

ข้อมูลคือ ข้อเท็จจริงที่มีการรวบรวมไว้และมีความหมาย อาจเกี่ยวข้องกับคน สิ่งของ หรือเหตุการณ์อื่นๆ

### คุณสมบัติของข้อมูลที่ดี

#### ความถูกต้อง (Accuracy)

ข้อมูลที่ดีจะต้องมีความถูกต้อง เพื่อให้สามารถนำไปใช้ประโยชน์ได้ ข้อมูลที่ไม่เป็นจริง และมีความคลาดเคลื่อนอยู่มาก อาจก่อให้เกิดความเสียหายเมื่อนำข้อมูลนั้นมาใช้ ดังสำนวนที่ว่า

“ใส่ขยะเข้าไปก็จะได้ขยะกลับออกมา ( Garbage In Garbage Out : GIGO )”

#### มีความเป็นปัจจุบัน (Update)

เนื่องจากข้อมูลปกติเป็นข้อมูลที่คงที่ เว้นแต่จะมีการแก้ไขโดยผู้ใช้ อีกทั้งเหตุการณ์ต่างๆ มักเปลี่ยนแปลงหรือเกิดขึ้นใหม่อยู่ตลอดเวลาหากข้อมูลที่มีอยู่ล้าสมัย อาจจะทำให้ผลลัพธ์ผิดพลาดได้

#### ตรงตามความต้องการ (Relevance)

จะต้องมีการสำรวจขอบเขตของข้อมูลที่จะนำมาใช้ให้สอดคล้องตรงตามความต้องการให้มากที่สุด

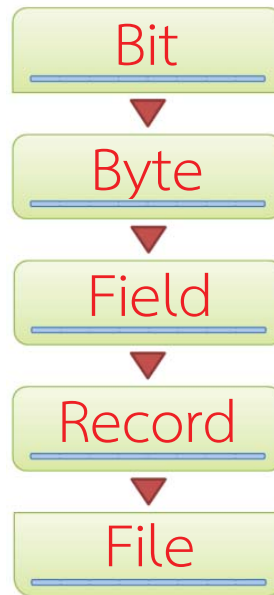
#### ความสมบูรณ์ (Complete)

การนำเอาข้อมูลมาใช้ จะต้องมีความสมบูรณ์ของข้อมูลมากพอ จึงจะทำให้เกิดประโยชน์อย่างเต็มที่

#### สามารถตรวจสอบได้ (Verifiable)

ข้อมูลที่ได้มา อาจมาจากหลายแหล่งข้อมูล ดังนั้นหากข้อมูลที่ได้มาสามารถตรวจสอบแหล่งที่มา หรือมีหลักฐานอ้างอิงแล้ว จะทำให้ป้องกันข้อมูลที่ไม่เกิดประโยชน์ และอาจนำผลเสียมาให้

## การแบ่งลำดับชั้นข้อมูล (Hierarchy)



\*\*\*เรียงจากหน่วยข้อมูลที่เล็กที่สุด ไปใหญ่ที่สุด\*\*\*

- บิต ( Bit : Binary Digit )

เป็นลำดับชั้นหน่วยข้อมูลที่เล็กที่สุด เป็นเลขฐานสองที่มีเพียงค่า 1 กับ 0 เท่านั้น

- ไบต์ ( Byte )

คือบิตที่มารวมกันหลายๆบิต

- ฟิลด์ หรือเขตของข้อมูล ( Field )

ประกอบด้วยข้อมูลไบต์ตั้งแต่ 1 ตัวขึ้นไป

- เรคอร์ด ( Record )

คือกลุ่มของฟิลด์ที่มีความสัมพันธ์กัน และนำมาจัดเก็บรวมกัน

- ไฟล์ หรือแฟ้มตารางข้อมูล ( File )

คือการนำเอาข้อมูลหลายเรคอร์ดมารวมกัน

id	Family - Code name	CPU Model	Native Frequency [GHz]	Core
0	i3 - Clarkdale	550	3.20	2
1	i5 - Clarkdale	680	3.60	2
2	i5 - Lynnfield	750	2.66	4
3	i7 - Lynnfield	880	3.06	4
4	i7 - Bloomfield	980x	3.33	6

Record

Field

จากตัวอย่างด้านบนแสดงข้อมูลซีพียู เห็นได้ว่าประกอบด้วยข้อมูลจำนวน 5 필ด์ ด้วยกัน ชื่อของ필ด์ต่างๆ ได้แก่

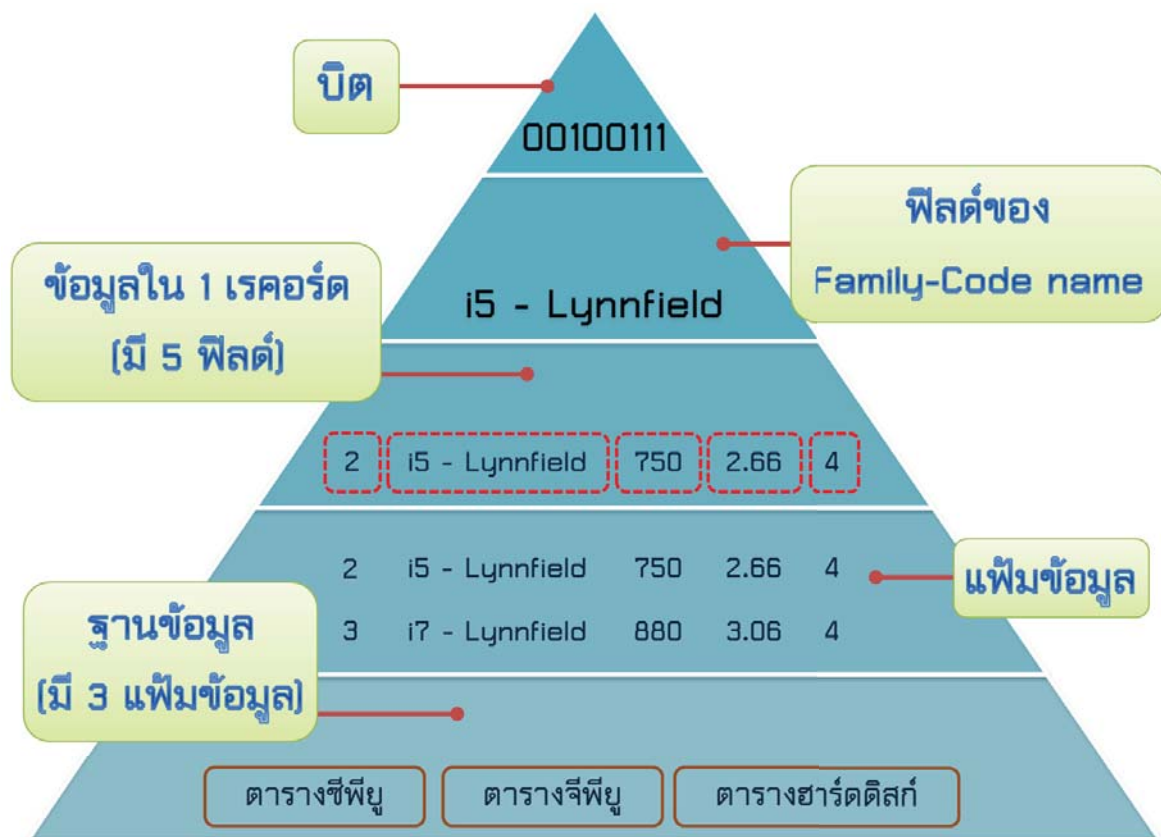
1. id (필ด์ที่ 1)
2. Family-CodeName (필ด์ที่ 2)
3. CPU Model (필ด์ที่ 3)
4. Native Frequency (필ด์ที่ 4)
5. Core (필ด์ที่ 5)

ซึ่งข้อมูลของซีพียูแต่ละตัวก็คือ เรคอร์ด นั่นเอง และในตารางนี้ก็มีทั้งหมด 5 เรคอร์ด

필ด์ที่ไม่มีข้อมูลซ้ำกันเลยจะเรียกว่า คีย์ฟิลด์ ( key field ) ซึ่งจะใช้เป็นตัวอ้างอิงแต่ละเรคอร์ด  
จากตัวอย่างดังกล่าว มีคีย์ฟิลด์คือ id ซึ่งไม่มีข้อมูลซ้ำกันเลย ตรงข้ามกับฟิลด์สามารถมีข้อมูลที่ซ้ำกันได้

## ฐานข้อมูล (Database)

ฐานข้อมูลเกิดจากการรวบรวมเอาแฟ้มตารางข้อมูลหลายๆแฟ้มที่มีความสัมพันธ์กันมาเก็บรวบรวมกันไว้ที่เดียว โดยจะมีการเก็บคำอธิบายเกี่ยวกับโครงสร้างฐานข้อมูลเรียกว่า พจนานุกรมข้อมูล ( data dictionary )



## การจัดโครงสร้างของแฟ้มข้อมูล (File Organization)

โครงสร้างแฟ้ม	ข้อดี	ข้อเสีย	สื่อที่ใช้เก็บ
แบบเรียงลำดับ (sequential file)	<ul style="list-style-type: none"> <li>- เสียค่าใช้จ่ายน้อย</li> <li>- ใช้งานได้ง่าย</li> <li>- เหมาะกับงานประมวลผลที่มี การอ่านข้อมูลแบบเรียงลำดับ และมีปริมาณมาก</li> <li>- สื่อที่ใช้เก็บมีราคาถูก</li> </ul>	<ul style="list-style-type: none"> <li>- เสียเวลาหาข้อมูล</li> <li>- ข้อมูลที่ใช้ต้องมีการจัดเรียง ก่อนเสมอ</li> <li>- ไม่เหมาะสมกับงานที่ต้องแก้ไข เพิ่ม ลบข้อมูลเป็นประจำ</li> </ul>	เทปแม่เหล็ก (magnetic tape)
แบบสุ่ม (random file)	<ul style="list-style-type: none"> <li>- ทำงานได้รวดเร็ว</li> <li>- เหมาะกับงานที่แก้ไข เพิ่ม ลบ เป็นประจำ</li> </ul>	<ul style="list-style-type: none"> <li>- ไม่เหมาะกับงานประมวลผลที่มี ปริมาณมาก</li> <li>- การเขียนโปรแกรมเพื่อค้นหา ข้อมูลจะซับซ้อน</li> <li>- ไม่สามารถเข้าถึงข้อมูลแบบ เรียงลำดับได้</li> </ul>	จานแม่เหล็ก (magnetic disk)
แบบลำดับดัชนี (indexed sequential file)	<ul style="list-style-type: none"> <li>- รองรับการทำงานประมวลผลสองแบบ คือแบบลำดับ และแบบสุ่ม</li> <li>- เหมาะกับงานที่แก้ไข เพิ่ม ลบ เป็นประจำ</li> </ul>	<ul style="list-style-type: none"> <li>- สิ้นเปลืองเนื้อที่ในการจัดเก็บ</li> <li>- การเขียนโปรแกรมเพื่อค้นหา ข้อมูลจะซับซ้อน</li> <li>- การทำงานช้ากว่าแบบสุ่ม</li> <li>- มีค่าใช้จ่ายสูง</li> </ul>	จานแม่เหล็ก (magnetic disk)

## แนวคิดของการใช้ฐานข้อมูล

- ลดความซ้ำซ้อนกันของข้อมูล (Reduced data redundancy)
- ลดความขัดแย้งของข้อมูล (Reduced data inconsistency)
- การรักษาความคงสภาพของข้อมูล (improved data integrity)
- ใช้ข้อมูลร่วมกันได้ (Shared data)
- ง่ายต่อการเข้าถึงข้อมูล (Easier access)
- ลดระยะเวลาการพัฒนาระบบงาน (Reduced development time)

นอกจากนี้ฐานข้อมูลยังช่วยในเรื่องการรักษาความปลอดภัยของข้อมูลได้ด้วย

## เครื่องมือสำหรับการจัดการฐานข้อมูล (DBMS)

การจัดการฐานข้อมูลด้วยคอมพิวเตอร์โดยปกตินี้จะโปรแกรมที่เรียกว่า ระบบการจัดการฐานข้อมูล (DBMS : Database Management System) ที่นิยมใช้งานกันในปัจจุบันได้แก่

- |                        |                 |
|------------------------|-----------------|
| ➤ Oracle               | ➤ Cache'        |
| ➤ Sybase               | ➤ PostgreSQL    |
| ➤ Microsoft SQL Server | ➤ Interbase     |
| ➤ Microsoft Access     | ➤ Firebird      |
| ➤ MySQL                | ➤ Pervasive SQL |
| ➤ DB2                  | ➤ SAP DB        |



## ลักษณะของ DBMS

