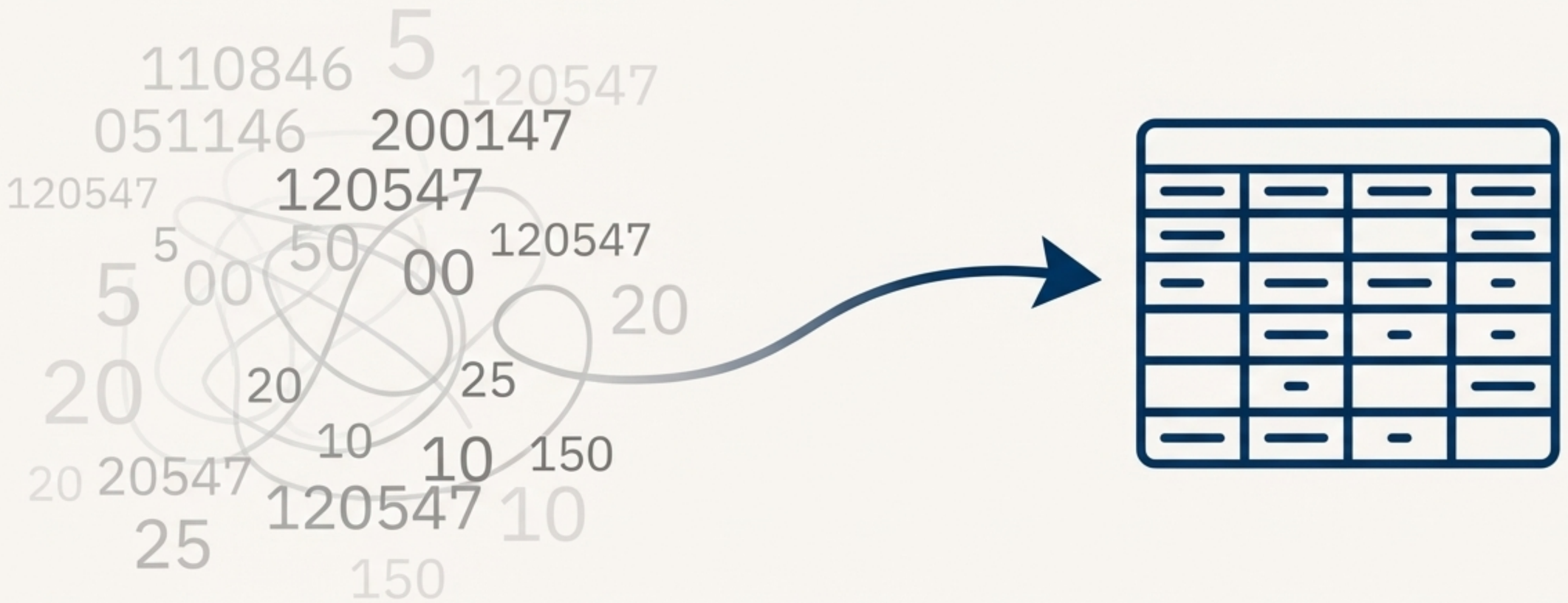
จากความยุ่งเหยิงสู่ความชัดเจน: การเดินทางของข้อมูล

ทำความเข้าใจพื้นฐานของการจัดการข้อมูล จากข้อเท็จจริงดิบสู่สินทรัพย์ที่มีค่า



NotebookLM

ข้อมูล (Data) คืออะไร?

- **นิยาม:** ข้อมูล คือ ข้อเท็จจริงที่มีการรวบรวมไว้และมีความหมาย อาจเกี่ยวข้องกับคน สิ่งของ หรือเหตุการณ์อื่นๆ
- **เป้าหมาย:** เป็นส่วนพื้นฐานสำหรับการประมวลผลเพื่อให้ได้สารสนเทศ (Information) สำหรับช่วยตัดสินใจและนำไปใช้ประโยชน์

```
110846   5    50   00
051146  20    25   00
200147  10     5   00
120547   3   150   00
290647   8    75   00
300947  13    15   00
051047  22   250   00
```

ข้อมูล

วันที่สั่ง	จำนวนที่สั่ง	ราคาต่อหน่วย
11/08/46	5	50.00
05/11/46	20	25.00
20/01/47	10	5.00
12/05/47	3	150.00
29/06/47	8	75.00
30/09/47	13	15.00
05/10/47	22	250.00

รายการสินค้าที่สั่ง

NotebookLM

วัตถุดิบของเรามาจากไหน: แหล่งข้อมูล

แหล่งข้อมูลภายใน
(Internal Data Sources)

แหล่งกำเนิดข้อมูลอยู่ภายในองค์กร

- ยอดขายประจำปี
- ข้อมูลผู้ถือหุ้น
- รายงานกำไรขาดทุน
- รายชื่อพนักงาน
- ข้อมูลผลิตภัณฑ์ใหม่

แหล่งข้อมูลภายนอก
(External Data Sources)

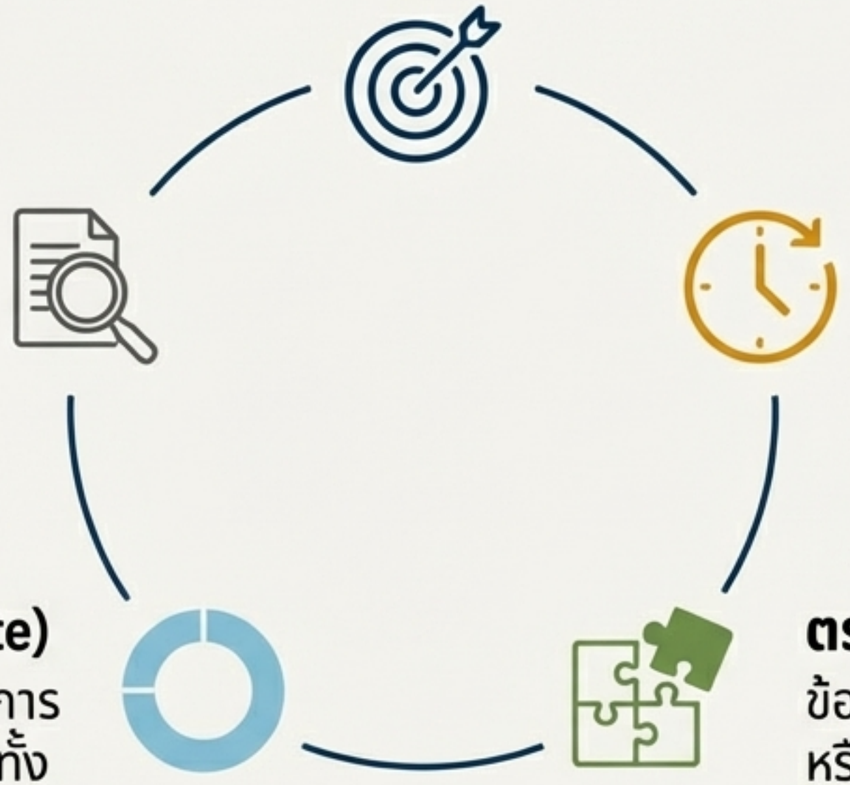
แหล่งกำเนิดอยู่นอกองค์กร

- ข้อมูลลูกค้า
- เจ้าหนี้
- อัตราดอกเบี้ย
- กฎหมายและอัตราภาษีของรัฐบาล
- ข้อมูลบริษัทคู่แข่ง

NotebookLM

การควบคุมคุณภาพ: คุณสมบัติของข้อมูลที่ดี

ความถูกต้อง (Accuracy)
ข้อมูลต้องเป็นจริงและเชื่อถือได้ ป้องกัน
ผลลัพธ์ที่ผิดพลาด (Garbage In, Garbage Out)



มีความเป็นปัจจุบัน (Update)
ข้อมูลต้องมีการปรับปรุงแก้ไขให้
ทันสมัยอยู่เสมอเพื่อไม่ให้คลาดเคลื่อน

ตรงตามความต้องการ (Relevance)
ข้อมูลที่นำมาใช้ต้องสอดคล้องกับปัญหา
หรือสิ่งที่ต้องการวิเคราะห์มากที่สุด

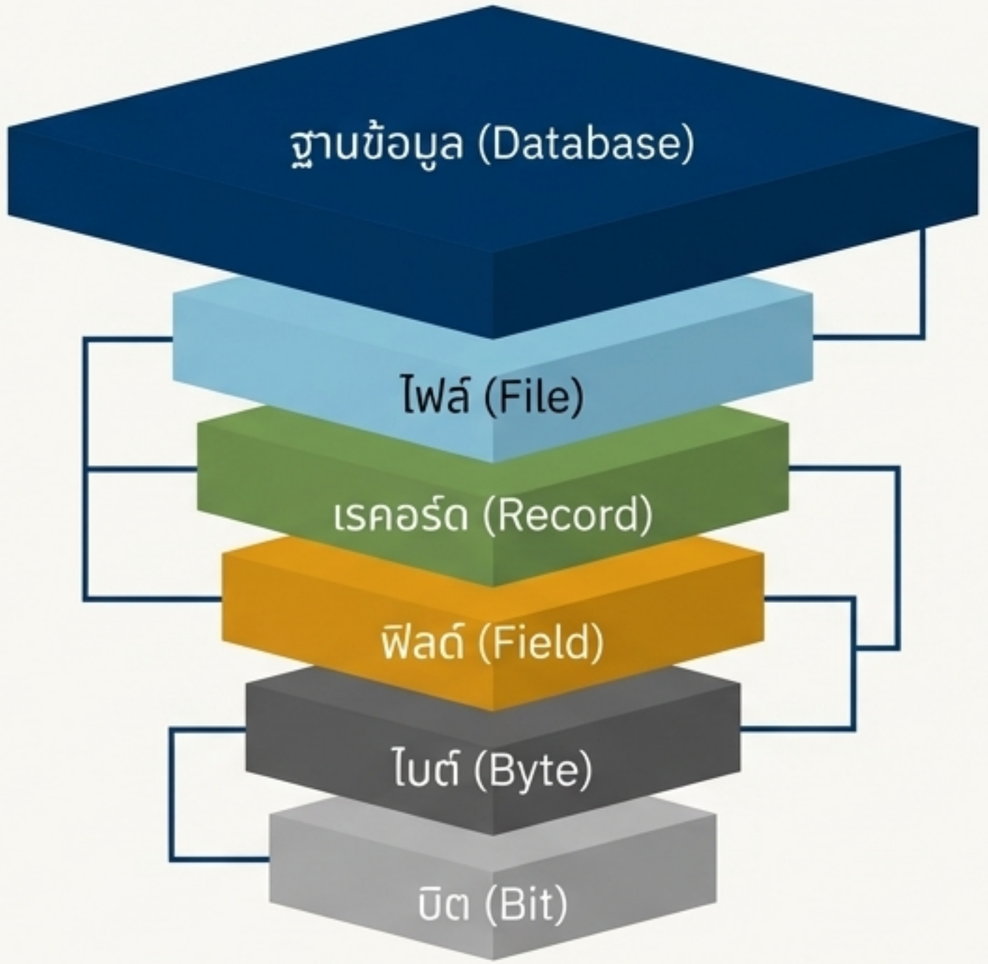
ความสมบูรณ์ (Complete)
มีความครบถ้วนในทุกมิติที่จำเป็นต่อการ
ใช้งาน เช่น ข้อมูลประชากรควรมีครบทั้ง
อายุ เพศ การศึกษา

สามารถตรวจสอบได้ (Verifiable)
สามารถตรวจสอบแหล่งที่มาหรือ
หลักฐานอ้างอิงได้ เพื่อป้องกัน
ข้อมูลที่ไม่น่าเชื่อถือ

NotebookLM

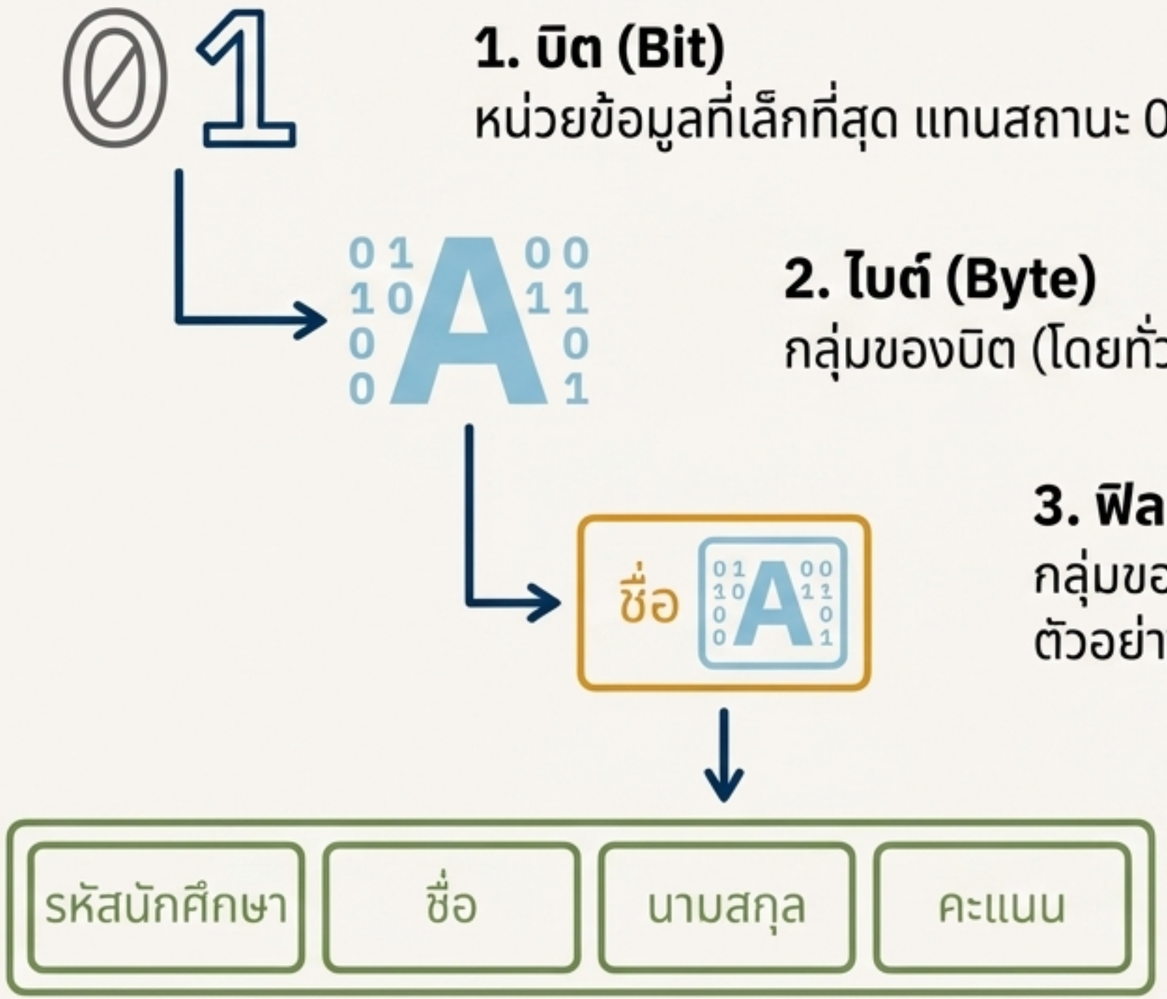
พิมพ์เขียวแห่งข้อมูล: ลำดับชั้นของการจัดการข้อมูล (Hierarchy of Data)

ในคอมพิวเตอร์ ข้อมูลไม่ได้ถูกเก็บอย่างกระจัดกระจาย แต่มีการจัดเรียงอย่างเป็นระบบตามลำดับชั้นที่ชัดเจน
จากหน่วยที่เล็กที่สุดไปจนถึงโครงสร้างที่ซับซ้อนที่สุด



NotebookLM

จากเล็กที่สุดสู่โครงสร้างที่ใหญ่ขึ้น: ส่วนประกอบของข้อมูล



1. บิต (Bit)
หน่วยข้อมูลที่เล็กที่สุด แทนสถานะ 0 หรือ 1 ของสัญญาณไฟฟ้า

2. ไบต์ (Byte)
กลุ่มของบิต (โดยทั่วไป 8 บิต) ที่รวมกันแทน 1 ตัวอักษรหรือสัญลักษณ์

3. ฟิลด์ (Field)
กลุ่มของไบต์ที่รวมกันเพื่อแสดงความหมายอย่างใดอย่างหนึ่ง
ตัวอย่าง: 'รหัสนักศึกษา', 'ชื่อ', 'คะแนน'

4. เรคอร์ด (Record)
กลุ่มของฟิลด์ที่เกี่ยวข้องกัน เพื่ออธิบายข้อมูลของสิ่งใดสิ่งหนึ่ง
ตัวอย่าง: ข้อมูลทั้งหมดของนักศึกษาหนึ่งคน (รหัส, ชื่อ, นามสกุล, คะแนน)

NotebookLM

การรวบรวมพิมพ์เขียว: จากไฟล์สู่ฐานข้อมูล

5. ไฟล์ (File)

การรวมกลุ่มของเรคอร์ดประเภทเดียวกันไว้ด้วยกัน

- ตัวอย่าง: ‘แฟ้มตารางข้อมูลคะแนนนักศึกษาวิชาคอมพิวเตอร์’ ที่ประกอบด้วยเรคอร์ดของนักศึกษาหลายๆ คน

รหัสนักศึกษา	ชื่อ-นามสกุล	คะแนน
B6001	สมชาย ใจดี	85
B6002	วิไลลักษณ์ สุขเกษม	92
B6003	มานะ อดทน	80
...	...	...
...	...	...

6. ฐานข้อมูล (Database)

การรวมเอาแฟ้มข้อมูลหลายๆ แฟ้มที่มีความสัมพันธ์กันมาเก็บไว้ที่เดียวกันอย่างเป็นระบบ

- ตัวอย่าง: ฐานข้อมูลนักศึกษา อาจประกอบด้วย ‘ไฟล์ประวัติส่วนตัว’, ‘ไฟล์ผลการเรียน’, ‘ไฟล์ข้อมูลทางการเงิน’

NotebookLM

จัดระเบียบห้องสมุด: 3 วิธีหลักในการจัดโครงสร้างแฟ้มข้อมูล (File Organization)

เมื่อเรามีข้อมูลที่มีโครงสร้างแล้ว คำถามต่อไปคือ เราจะจัดเก็บและเรียกใช้ข้อมูลเหล่านั้นอย่างไรให้มีประสิทธิภาพ? โครงสร้างแฟ้มข้อมูลคือวิธีการกำหนดรูปแบบการจัดเก็บบนสื่อบันทึกข้อมูล เพื่อให้เข้าถึงได้อย่างรวดเร็วและถูกต้อง

แบบเรียงลำดับ (Sequential)

เหมือนเทปคาสเซ็ท
(ต้องเล่นไปตามลำดับ)

แบบสุ่ม (Direct/Random)

เหมือนเครื่องเล่น MP3
(เลือกเพลงไหนก็ได้ทันที)

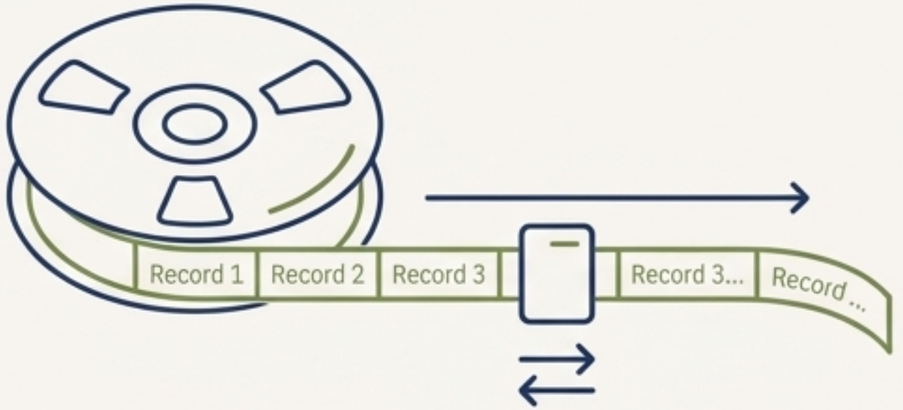
แบบลำดับเชิงดรรชนี (Indexed Sequential)

เหมือนหนังสือที่มีดรรชนี
(อ่านเรียงหน้าหรือเปิดหาจากดรรชนีก็ได้)

NotebookLM

วิธีการแบบดั้งเดิม: การเข้าถึงข้อมูลแบบเรียงลำดับและแบบสุ่ม

โครงสร้างแบบเรียงลำดับ (Sequential File)



- **การทำงาน:** จัดเก็บข้อมูลเรียงต่อกันไป การค้นหาต้องเริ่มจากเรคอร์ดแรกเสมอ ข้ามลำดับไม่ได้
- **เหมาะสำหรับ:** งานที่ต้องการอ่านข้อมูลต่อเนื่องกันในปริมาณมาก เช่น การประมวลผลเงินเดือน
- **อุปกรณ์จัดเก็บ:** เทปแม่เหล็ก (Magnetic Tape)

โครงสร้างแบบสุ่ม (Direct/Random File)

- **การทำงาน:** สามารถเลือกอ่านข้อมูลเรคอร์ดใดก็ได้โดยตรง ไม่ต้องผ่านเรคอร์ดแรกๆ
- **เหมาะสำหรับ:** งานที่ต้องการเข้าถึงข้อมูลอย่างรวดเร็วและมีการแก้ไขบ่อย
- **อุปกรณ์จัดเก็บ:** จานแม่เหล็ก (Magnetic Disk) เช่น ฮาร์ดดิสก์, CD-ROM

NotebookLM

ทางออกที่ยืดหยุ่นกว่า: แบบลำดับเชิงดรรชนี (Indexed Sequential)

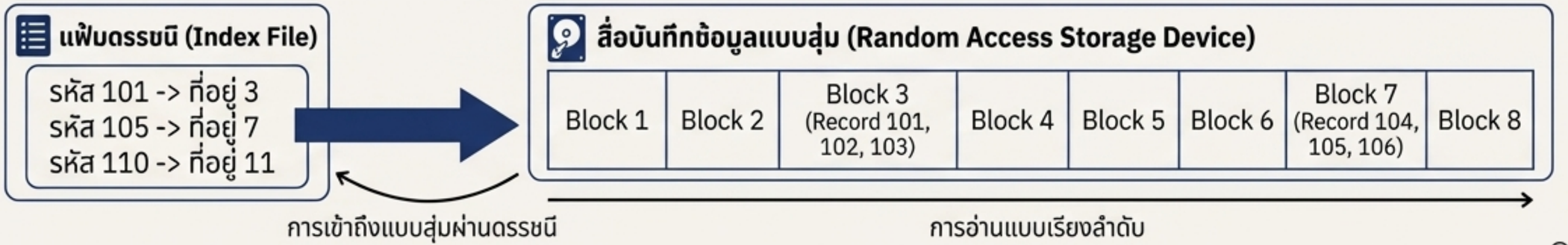
เป็นการรวมความสามารถของทั้งสองแบบเข้าด้วยกัน โดยใช้กระบวนการที่เรียกว่า ISAM (Index Sequential Access Method) ทำให้สามารถเข้าถึงข้อมูลได้ทั้งแบบเรียงลำดับและแบบสุ่ม

การทำงาน

1. ข้อมูลถูกจัดเก็บเรียงลำดับกันบนสื่อบันทึกแบบสุ่ม (เช่น จานแม่เหล็ก)
2. มีการสร้าง ‘แฟ้มดรรชนี’ (Index File) แยกต่างหาก ซึ่งทำหน้าที่เหมือนสารบัญของหนังสือ
3. แฟ้มดรรชนีจะชี้ไปยังตำแหน่งของข้อมูล ทำให้สามารถกระโดดไปหาข้อมูลที่ต้องการได้โดยตรง (แบบสุ่ม) หรือจะอ่านไปทีละเรคอร์ดก็ได้ (แบบเรียงลำดับ)

เหมาะสำหรับ

ระบบที่ต้องการความยืดหยุ่นสูง เช่น ธุรกรรมออนไลน์ ที่ต้องทั้งค้นหาข้อมูลลูกค้าเฉพาะรายและออกรายงานสรุปตามลำดับ



NotebookLM

เปรียบเทียบโครงสร้างแฟ้มข้อมูล: เลือกเครื่องมือให้เหมาะกับงาน

ลักษณะ (Characteristic)	แบบเรียงลำดับ (Sequential)	แบบสุ่ม (Random)	แบบลำดับเชิงดรรชนี (Indexed Sequential)
ลักษณะ	ค้นหาตามลำดับตั้แต่ต้น	ค้นหาได้ทันที	รองรับทั้งแบบลำดับและสุ่ม
ข้อดี	ค่าใช้จ่ายน้อย, เหมาะกับงานประมวลผลชุดใหญ่	ค้นหาเร็ว, เหมาะกับงานที่แก้ไขบ่อย	ยืดหยุ่นสูง
ข้อเสีย	ค้นหาช้า, ไม่เหมาะกับงานแก้ไข	ไม่เหมาะกับการอ่านข้อมูลมากๆ, โปรแกรมซับซ้อน	เปลืองเนื้อที่เก็บดรรชนี, ทำงานช้ากว่าแบบสุ่ม
สื่อที่ใช้จัดเก็บ	เทปแม่เหล็ก (Magnetic Tape)	จานแม่เหล็ก (Magnetic Disk)	จานแม่เหล็ก (Magnetic Disk)

NotebookLM

ปัญหาของโลกเก่า: เมื่อข้อมูลถูกเก็บแยกจากกัน (File Processing)

ในระบบดั้งเดิม แต่ละแผนกมักจะสร้างและจัดเก็บแฟ้มข้อมูลของตนเองแยกจากกันโดยสิ้นเชิง เช่น ฝ่ายบุคคลมีข้อมูลพนักงาน ฝ่ายขายมีข้อมูลการขาย

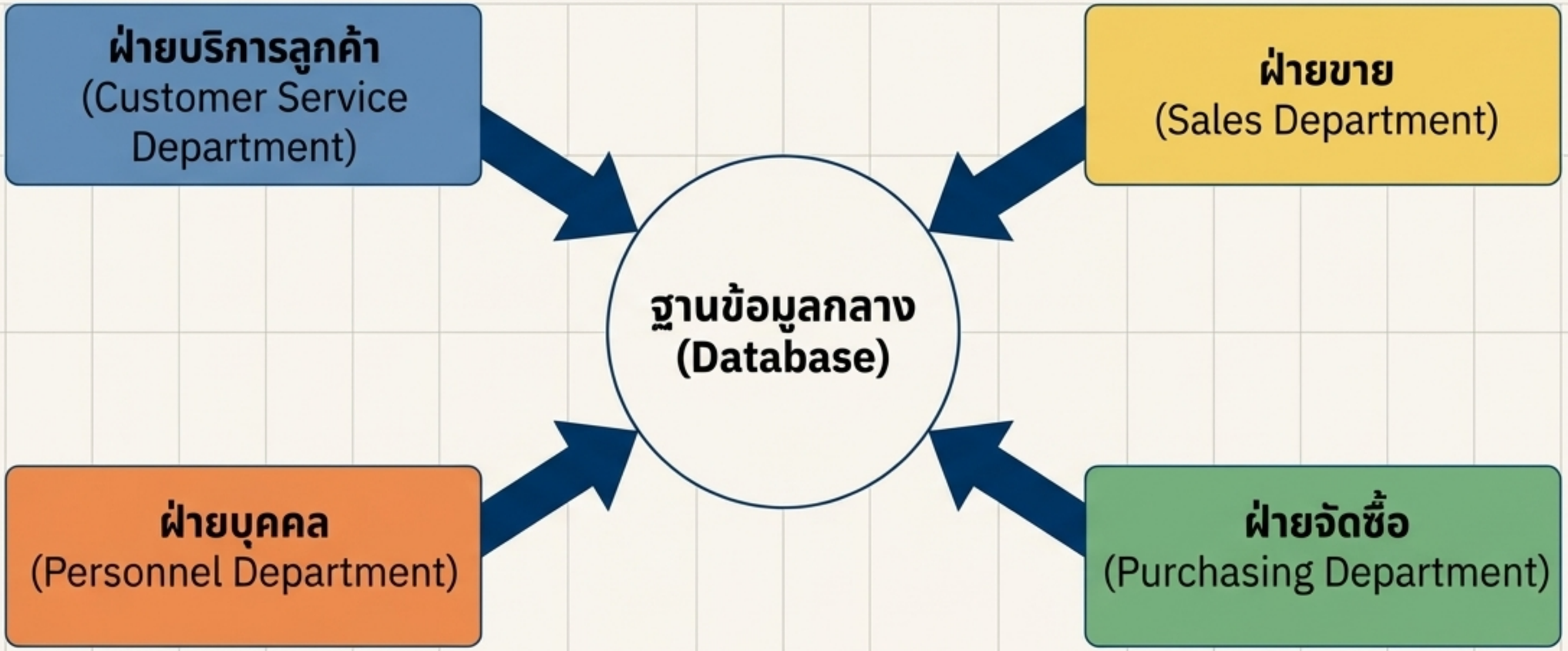
Key Problem*: สิ่งนี้ทำให้เกิดปัญหาใหญ่คือ **ความซ้ำซ้อนของข้อมูล (Data Redundancy) ข้อมูลชุดเดียวกันถูกจัดเก็บไว้หลายที่ ทำให้สิ้นเปลืองพื้นที่ และที่สำคัญกว่านั้นคือ...

NotebookLM

การปฏิวัติสู่ความชัดเจน: พลังของระบบฐานข้อมูล (Database Systems)

ระบบฐานข้อมูลถูกสร้างขึ้นเพื่อแก้ปัญหาความซ้ำซ้อนและความขัดแย้ง โดยการรวบรวมแฟ้มข้อมูลที่สัมพันธ์กันมาจัดเก็บไว้ที่ "ฐานข้อมูลกลาง" เพียงแห่งเดียว

The Solution: ทุกแผนกที่ได้รับอนุญาตจะสามารถเข้าถึงและใช้ข้อมูลจากแหล่งเดียวกันได้ (a single source of truth) ทำให้การค้นหา เรียกใช้ และปรับปรุงข้อมูลทำได้อย่างสะดวกและเป็นระเบียบ



NotebookLM

6 ประโยชน์หลักของการใช้ฐานข้อมูล

แนวคิดของการใช้ฐานข้อมูล

ลดความซ้ำซ้อนของข้อมูล (Reduced Data Redundancy)

จัดเก็บข้อมูลไว้ในที่เดียว ไม่ต้องเก็บซ้ำซ้อนในแต่ละแผนก

รักษาความคงสภาพของข้อมูล (Improved Data Integrity)

สามารถกำหนดกฎเพื่อควบคุมความถูกต้องของข้อมูลได้ (เช่น ชนิด, ความยาว)

ใช้ข้อมูลร่วมกันได้ (Shared Data)

ทุกฝ่ายสามารถเรียกใช้ข้อมูลชุดเดียวกันได้ตามสิทธิ์ที่กำหนด

ลดความขัดแย้งของข้อมูล (Reduced Data Inconsistency)

เมื่อแก้ไขข้อมูลที่ส่วนกลาง ทุกฝ่ายจะเห็นข้อมูลที่ถูกต้องตรงกันเสมอ

ง่ายต่อการเข้าถึงข้อมูล (Easier Access)

มีกลไกและภาษามาตรฐาน (เช่น SQL) ในการเข้าถึงข้อมูล

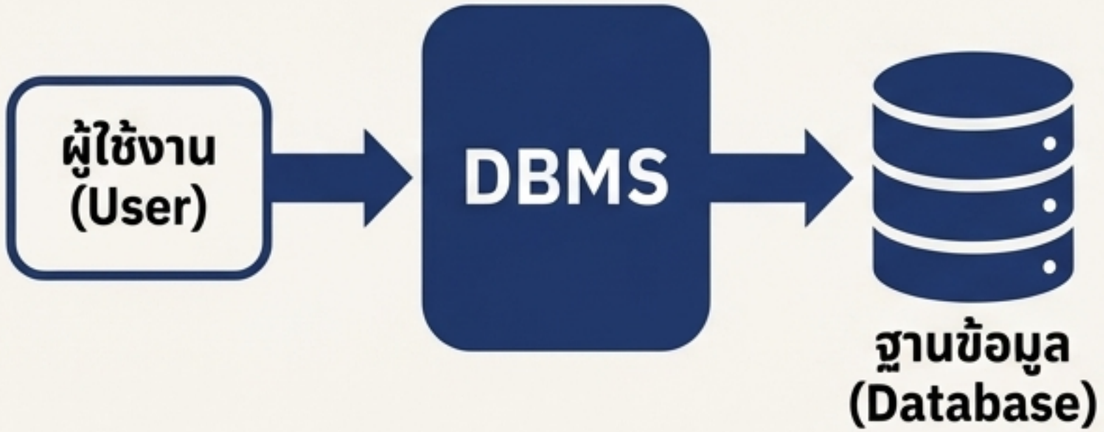
ลดเวลาพัฒนาระบบ (Reduced Development Time)

นักพัฒนาระบบทำงานได้เร็วขึ้น เพราะปัญหาข้อมูลลดลง

NotebookLM

เครื่องมือสู่ความชัดเจน: ระบบจัดการฐานข้อมูล (DBMS)

DBMS (Database Management System) คือซอฟต์แวร์ที่ทำหน้าที่เป็นตัวกลางระหว่างผู้ใช้งานและฐานข้อมูล ช่วยให้เราสามารถจัดการและติดต่อกับข้อมูลได้อย่างมีประสิทธิภาพ



ความสามารถหลักของ DBMS

สร้างฐานข้อมูล (Create Database): ออกแบบและสร้างโครงสร้างของฐานข้อมูล

เพิ่ม/เปลี่ยน/ลบข้อมูล (Add, Change, Delete Data): จัดการข้อมูลในฐานข้อมูลให้เป็นปัจจุบัน

จัดเรียงและค้นหา (Sort and Retrieve Data): ค้นหาและดึงข้อมูลที่ต้องการได้อย่างรวดเร็วผ่านภาษาคิวรี (Query Language) เช่น SQL

สร้างฟอร์มและรายงาน (Create Forms and Reports): นำเสนอข้อมูลในรูปแบบที่เข้าใจง่ายเพื่อการวิเคราะห์และตัดสินใจ

NotebookLM