บทที่ 9

พจนานุกรมข้อมูล (Data Dictionary)

วัตถุประสงค์ของบทนี้

1. ทราบถึงความหมายพจนานุกรมข้อมูล
2. การนำพจนานุกรมไปใช้ประโยชน์
3. การเขียนพจนานุกรมข้อมูล
4. แฟ้มข้อมูลประเภทต่างๆ
5. การประมาณการเพื่อสร้างพื้นที่บันทึกข้อมูล

พจนานุกรมข้อมูล (Data Dictionaries)

พจนานุกรมข้อมูล หมายถึงเอกสารที่ว่ารายละเอียดของโครงสร้างแฟ้มข้อมูล ตลอดจนบรรดารายการข้อมูลต่างๆ เช่นรายการชื่อ โครงสร้าง (Format) แหล่งที่เก็บ การนำไปใช้งาน ข้อจำกัดต่างๆของคุณลักษณะที่เป็นไปได้ของข้อมูลรายการต่างๆที่ปรากฏ ในกรณีที่ระบบเดิมที่เราศึกษาอยู่นั้นมีเอกสารชนิดนี้ปรากฏอยู่แล้ว ก็จะช่วยให้การวิเคราะห์ระบบเก่าเป็นไปได้อย่างอย่างดียิ่ง เพราะจะทำให้นักวิเคราะห์ระบบนำสิ่งนี้ไปสู่การติดต่อระหว่างระบบย่อยๆต่างๆ ในการส่งงต่อข้อมูลภายในองค์กร

พจนานุกรมข้อมูล เป็นระเบียบวิธีการในการเก็บชื่อข้อมูลหรือสารสนเทศ (Information) ต่างๆ พร้อมด้วยคำอธิบายสั้นๆ และรายละเอียดอื่นๆ ดังปรากฏรายการอธิบายต่อไปนี้

1. อธิบายคุณลักษณะของแฟ้มข้อมูล

1.1 ข้อมูลเบื้องต้นทั่วๆไปเกี่ยวกับแฟ้มข้อมูล

- ชื่อ (Name)
- ชื่ออื่นๆ (Alias or Synnonyms)
- สาระสังเขป (Description)

1.2 ประเภทของแฟ้มข้อมูล (File Type) แบ่งเป็น

1. แฟ้มแบบทีละลำดับ (Sequential File)
2. แฟ้มแบบสุ่ม (Random File) ซึ่งแบ่งเป็น
    - 2.1 แฟ้มโดยตรง (Direct File)
    - 2.2 แฟ้มเชิงดัชนี (Index Sequential File)
    - 2.3 แฟ้มแบบสัมพัทธ์ (Relative File)

ปัจจัยที่ใช้พิจารณาว่า จะสร้างแฟ้มข้อมูล รูปแบบใด คือ"การสืบค้น ข้อมูลมาใช้งาน" และความถี่ในการนำมาใช้งาน

1.3 ข้อมูลควบคุมแฟ้ม (Control Information) จะประกอบด้วย

- แหล่งที่เกิด (Source)
- วันที่ที่สร้างแฟ้ม (Date of Origin)

- ผู้ใช้ในระบบ (User)
- โปรแกรมใดบ้างที่ใช้แฟ้มข้อมูลนี้ (Program in which used)
- สิทธิการใช้แฟ้มข้อมูล (Acess Authorization)
- สิทธิในการปรับปรุงแฟ้มข้อมูล (Change Authorization)

1.4 สายความสัมพันธ์ของข้อมูล (Group Information)
- โครงสร้างของแหล่งที่เกิด (Parent Structure)
- โครงสร้างข้อมูลที่เกื้อกูล (Subsidiary Structure)
- โครงสร้างข้อมูลที่ซ้ำกัน (Repetitive Structure)
- ตำแหน่งที่เก็บบนสื่อ (Physical Location of Record File or Data Base)

2. รายละเอียดของข้อมูลแต่ละรายการ (A list of information typically recorded for each data element on a data dictionary)

2.1 ข้อมูลเบื้องต้นทั่วๆไปเกี่ยวกับข้อมูลแต่ละรายการ
- ชื่อ (Name)
- ชื่ออื่นๆหรือชื่อย่อ (Alias or Synnonyms)
- สาระสังเขป (Description)

2.2 รูปแบบของรายการข้อมูล (Format)
- ประเภทของรายการข้อมูล (Data Type)
- ความยาว (Length)
- รูปลักษณ์ (Picture)
- หน่วย (เช่นความยาว มีหน่วยเป็น นิ้ว หรือ ฟุต น้ำหนักมีหน่วนเป็น ปอนด์ กรัม หรือ หน่วยของการบรรจุผลิตภัณฑ์ เช่น 1 กล่อง หรือ 1 โหล เป็นต้น)

2.3 คุณลักษณะของการใช้รายการข้อมูลแต่ละรายการ (Usage Characteristics)
- ค่าที่เป็นไปได้ (Range of Value)
- ความถี่ของการใช้งาน (Frequencies of Use)
- เป็นรายการข้อมูลนำเข้า หรือ เป็นรายการข้อมูลนำออก (Input/Output /Local)
- ค่าของข้อมูลที่อยู่ในข่ายเงื่อนไข (Conditional Value) เช่นรายการ Reorder Level ในการควบคุมการสั่งซื้อสินค้ามีสิทธิที่จะน้อยกว่า 0 ได้ ถ้าเกิดมีการตกลงให้ใช้งานได้ตามข้อตกลง

ตัวอย่างของพจนานุกรมข้อมูลในระบบสินค้าคงคลัง (Inventory System) ทีแบ่งระบบงานออกเป็นส่วนๆ และงานส่วนหนึ่งที่เกี่ยวข้องก็คือระบบการสั่งซื้อ (Purchasing System) โดยที่ในส่วนของระบบนี้จะมีข้อมูลที่สามารถจะดำเนินงานได้มีรายละเอียดของแฟ้มที่เกี่ยวข้องดังที่เขียนในพจนานุกรมข้อมูลดังนี้

ภาพที่ 13.1 ตัวอย่างการเขียนพจนานุกรมข้อมูลของงานการจัดซื้อสินค้า รายละเอียดที่เกี่ยวข้องกับแฟ้มข้อมูล

Name	: **INVENTORY - EXCEPTION -** REPORT
Aluases	: **REORDER-REPORT,PURCHASING-REPORT**
Description	: **Group item representing daily list of parts to be reodered** .Sent **to purchasing**
Control Information :	
	Source : Order System
	Date of Origin **:** 30/12/1995
	Users : Purchasing Manager
	Programs in which used : REODER-REPORT PROGRAM
	Change Authorization **: Sale Manager**
Group Information **:**	
	Parent Structures : Stock File
	Physical **Location : File** (Magnetic **Tape**)
Location	**: Output to printer.**

ส่วนที่เกี่ยวข้องกับรายการข้อมูลแต่ละรายการในแฟ้มข้อมูลจะอธิบายดังรายการตัวอย่างนี้

Name	**PART - NUMBER**
Aliases	
Description	: **Key field that** uniquely identifies **a** specific **part** in **Inventory**
Format	**Alphanumeric** 8 **Characters** Pic x(8)
Location	**INVENTORY - EXCEPTION - REPORT**
	INVENTORY
	REORDER

Name	**REORDER - QUANTITY**
Ailases	: **PURCHASE QUANTITY**
Description	: **The number of** units **of a** given **part that are to be reordered**
	at a single time
Format	Numeric **5** digits
	Pic 9(5)
Location	: **INVENTORY-EXCEPTION-REPORT**
	INVENTORY
	REORDER

ประโยชน์ของการสร้างพจนานุกรมข้อมูล

1. ใช้เป็นเอกสารอ้างอิงการใช้งานเกี่ยวกับข้อมูลทั้งหลายที่ปรากฏในระบบ

2. เพื่อให้ข้อมูลอยู่ในรูปแบบที่เป็นมาตรฐาน สำหรับผู้ใช้ที่อยู่ในหน่วยงานเดียวกัน หรือต่างหน่วยงานก็นแต่จะต้องใช้ข้อมูลร่วมกัน

3. ในกรณีของทีมงานขนาดใหญ่ที่มีผู้เขียนโปรแกรมหลายๆคน จะได้ใช้ฐานพจนานุกรมข้อมูลเดียวกันได้อย่างถูกต้อง

4. ในมุมมองของการติดต่อระหว่างแต่ละระบบย่อยนั้น การมีพจนานุกรมข้อมูลจะนำพาไปสู่การเชื่อมโยงระหว่างระบบงานย่อยๆเหล่านั้นได้ ส่งผลให้ระบบงานรวมขององค์กรถูกต้อง

5. การมีพจนานุกรมข้อมูลจะช่วยในการลดความซ้ำซ้อนของข้อมูลในระบบ

6. ใช้เป็นสารสนเทศเพื่อช่วยในการบำรุงดูแลรักษาระบบข้อมูล

โปรแกรมสำเร็จรูปที่ใช้ในงานสร้างพจนานุกรมข้อมูล (Data Dictionary Software)

ด้วยเหตุผลและความจำเป็นดังที่กล่าวมาแล้ว ในการสร้างพจนานุกรมข้อมูล เราจึงจำเป็นจะต้องพึ่งพิงโปรแกรม ที่สร้างขึ้นมาเพื่อสร้างพจนานุกรมข้อมูล แทนที่จะดำเนินการสร้างโดยการเขียนออกมาในรูปแบบของเอกสารดังที่เคยปฏิบัติมาในอดีต ซึ่งสิ้นเปลืองเวลา แรงงานคน และปรับปรุงแก้ไขได้ยาก

ปัจจุบันนี้ เรามีโปรแกรมสำเร็จรูปเพื่อใช้งานในลักษณะดังกล่าวอยู่หลายโปรแกรมแล้วบางโปรแกรมก็สามารถทำงานได้อย่างค่อนข้างจะเป็นอิสระจากระบบเครื่องและจากระบบปฏิบัติการ แต่บางโปรแกรมก็ยังจะต้องพึ่งพิงอยู่กับระบบบริหารฐานข้อมูล และบางหน่วยงานก็ทำการพัฒนาโปรแกรมของตนเองขึ้นโดยเฉพาะ โดยยึดเอาระบบข้อมูลภายในองค์กรของตนเองเป็นหลัก

ในกรณีที่เราจะเลือกโปรแกรมสำเร็จรูปเพื่อทำงานดังกล่าวมาใช้ เราจะมีเกณฑ์ในการเลือกโปรแกรมสำเร็จรูปเหล่านี้มาใช้งานได้อย่างไร ข้อแนะนำในการเลือกโปรแกรมสำเร็จรูปเพื่อการสร้างพจนานุกรมข้อมูล มาใช้งานมีองค์ประกอบดังนี้คือ

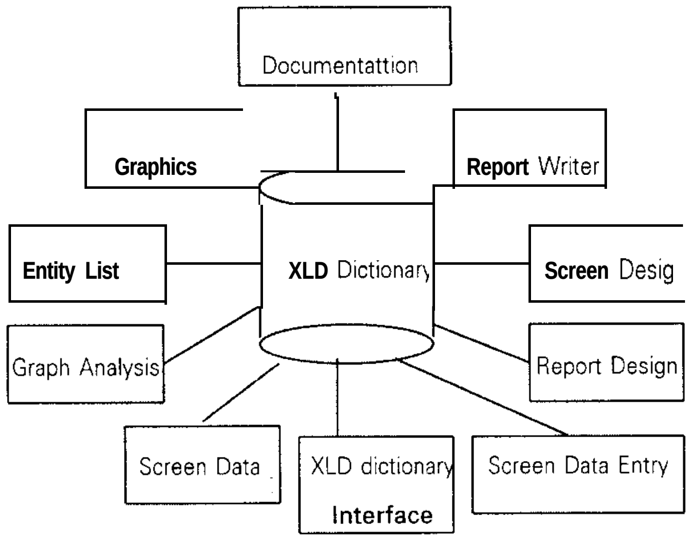
1. ปกติเราจะเห็นได้ว่าโปรแกรมที่เขียนกันขึ้นมาใช้ในระบบงานทั้งหลาย ล้วนแต่จะต้องมีการอ้างอิงถึงข้อมูลทั้งหมด หรืออ้างถึงข้อมูลเพียงส่วนจากแฟ้มข้อมูล หรือฐานข้อมูล ปรากฏอยู่ในส่วนหนึ่งของโปรแกรม นั่นก็คือถ้าเราสามารถเลือกโปรแกรมสำเร็จรูปที่สามารถทำการสร้างพจนานุกรมข้อมูล โดยอาศัยคำสั่งที่เกี่ยวข้องกับข้อมูลที่ปรากฏในโปรแกรมปฏิบัติงานอื่นที่เกี่ยวข้องได้ เราก็จะได้รับประโยชน์มาก ตัวอย่างเช่น การอ้างอิงข้อมูลในส่วนของ Data Division ในภาษา COBOL เป็นต้น

2 เลือกโปรแกรมสร้างพจนานุกรมข้อมูล ที่มีความสามารถในการสร้าง พจนานุกรมข้อมูลหุ่นจำลองให้ปรากฏบนจอภาพ โดยที่พจนานุกรมข้อมูลนั้นเราจะเรียกว่า แบบจำลองของพจนานุกรมข้อมูล (Simulated Data Dictionary) แล้วให้ผู้ใช้แก้ไขรายการต่าง ตามวัตถุประสงค์ที่ต้องการ

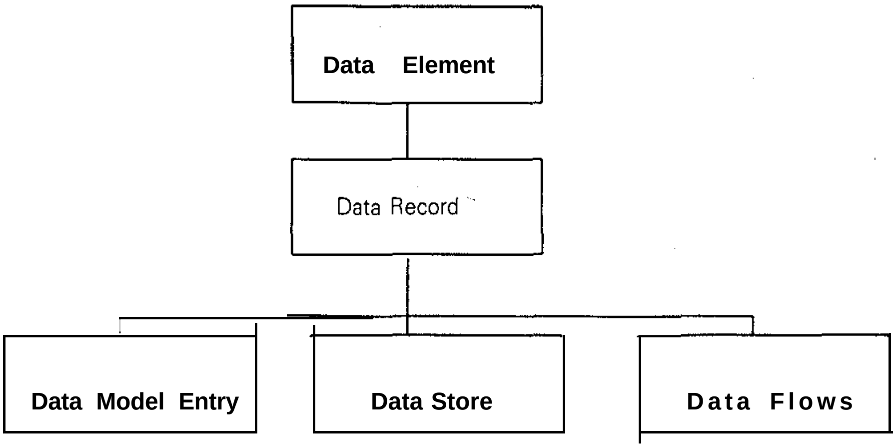
3. อีกแนวทางก็คือการเลือกโปรแกรมที่สามารถ โต้ตอบกับผู้ใช้ในการสร้างพจนานุกรมข้อมูลเฉพาะเรื่องได้อย่างชาญฉลาด

โปรแกรมสำเร็จรูปที่ใช้สร้างพจนานุกรมข้อมูล เท่าที่ปรากฏใช้งานทั่วไปนั้นมักจะอยู่ในรูปแบบของการสร้างพจนานุกรมข้อมูล ให้เป็น Source Code สำหรับการนำไปใช้งานนั้นๆ โดยพจนานุกรมข้อมูล นั้นมีความสามารถจะเพิ่มรายการข้อมูลใหม่ (add new data) หรือลบรายการข้อมูลเก่าที่ไม่ต้องการทิ้งได้ (delete old data)ทำให้ผู้ใช้ใช้งานได้อย่างสะดวกสบาย

ภาพที่ 13.2 แสดงถึงการใช้ประโยชน์จากพจนานุกรมข้อมูล



ภาพที่ 13.3 พจนานุกรมข้อมูลจะแสดงถึงโครงสร้างลำดับชั้นของข้อมูล ดังภาพ



การออกแบบแฟ้มข้อมูลและการคำนวณพื้นที่จัดเก็บข้อมูล (File Design and Space Estimate)

เป็นที่ทราบว่าภาระหนึ่งซึ่งถือว่าเป็นภาระที่หนักมากในงานการสร้างระบบใหม่ ก็คือการออกแบบแฟ้มข้อมูลเพื่อการใช้งานรวมถึงวิธีการจัดเก็บบนอุปกรณ์ที่จะต้องเลือกได้อย่างเหมาะสม เพราะการออกแบบแฟ้มที่เหมาะสมกับระบบงานนั้น นอกจากจะทำให้ระบบงานนั้นสามารถบรรลุเป้าหมายแล้วยังส่งผลให้ระบบงานนั้นมีประสิทธิภาพด้วย การออกแแบระบบโดยใช้เครื่องมือ คือ ผังระบบ หรือ ใช้ผังกระแสข้อมูลนั้น จะต้องทำการกำหนดเรื่องของแฟ้มข้อมูล หรือ ฐานข้อมูลพร้อมกันไปด้วย ส่วนรายละเอียดของแฟ้มข้อมูลหรือฐานข้อมูลเหล่านี้จะทำการออกแบบภายหลังในส่วนที่ว่า รายละเอียดของการออกแบบรายละเอียดของระบบ (Detail Design) ต่อไป

การออกแบบแฟ้มหรือฐานข้อมูลนั้นทำได้โดยการ นำเอาแหล่งที่เก็บข้อมูล (Data Store) เหล่านั้นมาออกแบบโครงสร้างแฟ้ม ข้อมูล รายการข้อมูล ประเภทของข้อมูล และขนาดของแต่ละรายการ ซึ่งสิ่งที่กล่าวถึงนี้ก็คือสิ่งที่จะปรากฏในพจนานุกรมข้อมูลนั่นเอง ตัวอย่างเช่นในระบบบัญชีเงินเดือนนั้น เราจะสร้างแฟ้มข้อมูลที่เป็นแฟ้มประเภทลำดับ (Sequential File) ดังรายการข้อมูลต่อไปนี้

Element name	Length	Format
Social Security Number	9	character
Name	20	character
Year-to-date gross pay	7	99999V99
Year-to-date federal tax	7	99999V99
Year-to-date state tax	7	99999V99
Year-to-date local tax	7	99999V99
Year-to-date FICA tax	7	99999V99
Year-to-dare net pay	7	99999V99

ในกรณีที่ต้องการนำแฟ้มข้อมูลไปใช้ในงานการสืบค้นหาสารสนเทศด้วย นอกเหนือไปจากการใช้แฟ้มข้อมูล ดังกล่าวเพื่อประมวลผลในเรื่องของเงินเดือนดังที่ปฏิบัติอยู่เป็นปกติ เราก็สามารถจะปรับเปลี่ยนรูปแบบของแฟ้มข้อมูลดังกล่าวให้เป็นแฟ้มเชิงดัชนี(Indexed Sequential File) ซึ่งมีโครงสร้างนี้จะประกอบด้วยแฟ้มที่เป็น แฟ้มดัชนี (Indexed File) และแฟ้มที่เก็บข้อมูล โดยที่ แฟ้มดัชนี (Indexed File) โดยจะทำหน้าที่ในการสืบค้น และนำสิ่งที่สืบค้นคือตำแหน่งที่ได้ไปหาจากตำแหน่งที่แท้จริงในตัวกลางที่บันทึก (Physical Location) ต่อไป

การออกแบบแฟ้มข้อมูล (File Design) ให้เหมาะสมกับระบบงาน จะต้องพิจารณาจากปัจจัยดังต่อไปนี้คือ

1. ลักษณะของการนำข้อมูลในแฟ้มไปใช้งาน โดยทั่วไปเราสามารถแบ่งระบบงานการประมวลผลข้อมูลออกมาได้ 2 แบบคือ การประมวลผลแบบกลุ่ม (Batch Processing) และการประมวลผลแบบทันทีทันใด (Online Processing) แต่ในบางครั้งก็จะต้องออกแบบแฟ้มข้อมูลนั้นให้มีความสามารถประมวลผลได้ทั้งสองแบบดังตัวอย่างที่ กล่าวมาแล้วข้างต้น ตัวอย่างของระบบงานที่มีการประมวลผลแบบกลุ่มก็จะใช้แฟ้มข้อมูลแบบทีละลำดับ เช่นงานประมวลผลข้อมูลระบบเงินเดือน การพิมพ์รายชื่อลูกค้าที่รับวารสารจากสำนักพิมพ์ การประกาศผลสอบนักศึกษาเป็นต้น โดยที่ข้อมูลแต่ละระเบียนในแฟ้มจะถูกนำมาปฏิบัติทุกระเบียนหรือเกือบทุกระเบียนข้อมูลในแฟ้ม

แฟ้มข้อมูลชนิดทีละลำดับนั้นจะมีประโยชน์อยู่หลายประการคือ

. ง่ายและสะดวกในการสร้างแฟ้มข้อมูล โดยที่ข้อมูลแต่ละระเบียนนั้นจะบันทึกเรียงไปทีละลำดับที่ส่งเข้าไป โดยที่ข้อมูลที่ส่งเข้าไปบันทึกนั้นอาจจะเรียงลำดับมาแล้วหรือยังไม่เรียงลำดับก็ได้

. ง่ายแก่การสืบค้นข้อมูล กรรมวิธีในการสืบค้นก็คือการค้นหาไปทีละระเบียนข้อมูล

. ใช้พื้นที่ได้อย่างมีประสิทธิภาพ (Space Efficiency) โครงสร้างข้อมูลที่เก็บในลักษณะนี้นั้น ไม่จำเป็นจะต้องกำหนดพื้นที่ในส่วนที่เป็นตารางสืบค้น (Dirctories) ไว้ เพราะตำแหน่งของข้อมูลไม่จำเป็นกับการใช้งาน ดังนั้นเราจึงสามารถประหยัดพื้นที่ส่วนนี้ไปได้

. การประมวลผลแบบกลุ่มนับเป็นการประมวลผลที่ประหยัดค่าใช้จ่ายกว่าการประมวลผลแบบทันทีทันใด

ข้อเสียของการเก็บแฟ้มข้อมูลชนิดทีละลำดับ

. เวลาที่ใช้ในการค้นหา (Searching Time) จะช้า เพราะเจตนาของการสร้างนั้นเพื่อที่จะทำการประมวลผลแฟ้มทั้งแฟ้ม ไม่ใช่เพื่อค้นหาระเบียนใดระเบียนหนึ่งตามลักษณะเฉพาะเรื่องที่ต้องการ เราสามารถเปรียบเทียบเวลาได้ว่าถ้าเราจะทำการประมวลผลระเบียนข้อมูลทั้งแฟ้มข้อมูลแล้ว การจัดเก็บข้อมูลแบบนี้นับว่าเป็นการเก็บที่เสียเวลาในการอ่านข้อมูลน้อยกว่าวิธีการที่เก็บข้อมูลแบบสุ่ม แต่ถ้าเราใช้เพื่อค้นหาระเบียนข้อมูลที่ต้องการแล้ว การจัดเก็บข้อมูลโดยวิธีนี้จะส่งผลให้เสียเวลามากในการสืบค้นข้อมูล โดยการสมมุติว่าภายในแฟ้มมีข้อมูลอยู่ n ระเบียนข้อมูลแล้วโอกาสที่เราจะค้นหาระเบียนข้อมูลใดๆที่ต้องการนั้นพบจะค้นหาโดยเฉลี่ยถึง $n(n+1)/2$ ครั้ง

. กรรมวิธีของการเพิ่มหรือลบระเบียนข้อมูลในแฟ้มข้อมูลแบบเรียงลำดับ โดยที่การลบหรือ การเพิ่มนั้น ไม่ได้กระทำในส่วนหรือหัวหรือท้ายแฟ้มข้อมูลก็จะกระทำได้ลำบากและเสียเวลามาก

2. เวลาที่ใช้ในการอ่าน/บันทึกข้อมูลในแฟ้มข้อมูล

3. พื้นที่ที่ใช้ในการจัดเก็บข้อมูล

4. สื่อบันทึกข้อมูลที่ใช้งาน ดังที่ทราบอยู่แล้วว่า ถ้าเราใช้เทปแม่เหล็กในการบันทึกข้อมูลแล้วการประมวลผลข้อมูลจะเป็นไปได้กรณีเดียวคือการดำเนินการทีละลำดับ ในขณะที่ถ้า

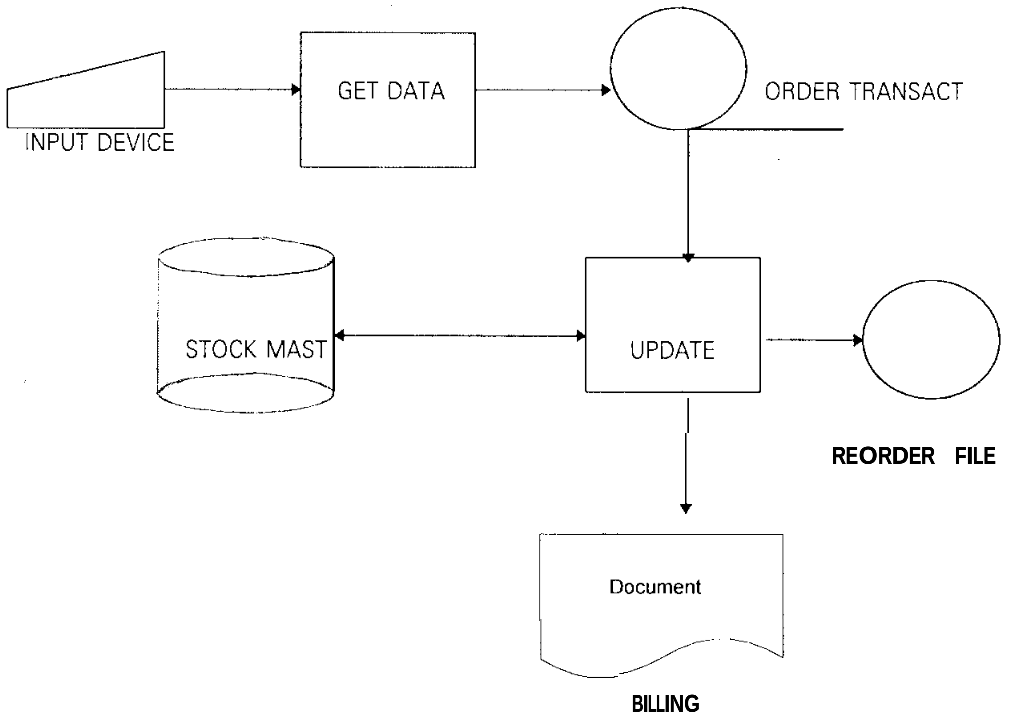
เป็น จานแม่เหล็ก (Disk) หรือจานแม่เหล็กชนิดอ่อน (Diskette) หรือแผ่น ซีดี (CD Rom) ก็จะสามารถจัดเก็บข้อมูลได้ทั้งแบบที่เป็นแฟ้มชนิดทีละลำดับหรือแบบสุ่มก็ได้

ดังที่ได้กล่าวมาแล้วว่าระบบการประมวลผลนั้นเรามี สองประเภทคือแบบกลุ่ม และ แบบทันที ทันใด ทั้งสองวิธีนี้ต่างก็เหมาะสมกับงานคนละประเภท เช่น การประมวลผลแบบทันทีทันใดที่เหมาะสม กับงานที่รอคอยไม่ได้ เช่นการ ฝาก - ถอนเงิน งานสำรองที่นั่งเครื่องบิน รถโดยสารเป็นต้น ในแง่ของความ ปลอดภัยที่เกิดจากการประมวลผลนั้น เราจะเห็นได้ว่าการประมวลผลแบบกลุ่มนั้นค่อนข้างที่จะปลอดภัย กว่าเพราะว่า ถ้าเกิดมีข้อผิดพลาดขึ้นมาในการประมวลผล เราจะสามารถสอบทานได้จากแฟ้มข้อมูลที่ ยังมีปรากฏ "ซาก" (Transaction) ปรากฏอยู่ ในขณะที่ถ้าเป็นการประมวลผลแบบทันทีทันใดแล้วการ จะตรวจค้นหรือแก้ไขจะทำได้ลำบากมาก เพราะไม่มีข้อมูลตรวจทานปรากฏอยู่แล้ว และข้อมูลได้ยังได้ กระจายเข้าไปอยู่ในแฟ้มข้อมูลแล้วนั่นเอง

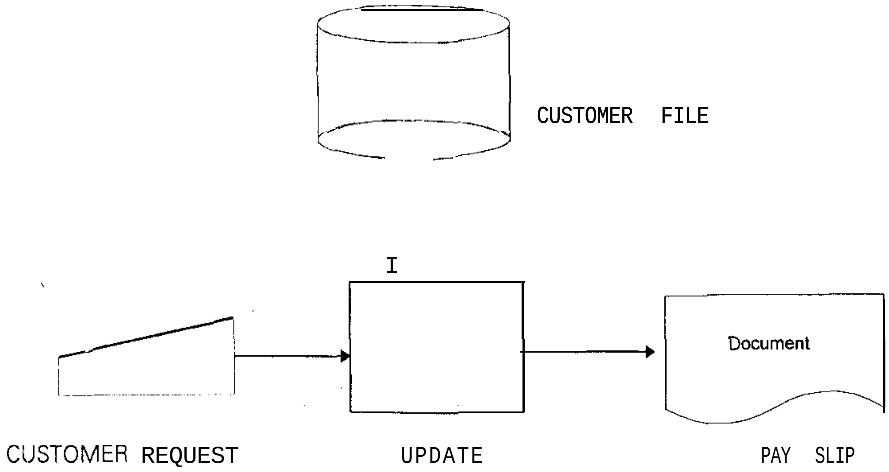
ถ้าเราเปรียบเทียบแฟ้มข้อมูลแบบสุ่มกับแฟ้มข้อมูลแบบทีละลำดับแล้วนั้นก็จะมีข้อดีข้อเสียแตก ต่างตรงข้ามกันดังรายละเอียดที่กล่าวมาแล้ว

ภาพที่ 13.4 ตัวอย่าง ผังระบบของการประมวลผลแบบกลุ่มในระบบงานการสั่งซื้อสินค้า

(Order System : Transaction Processing)

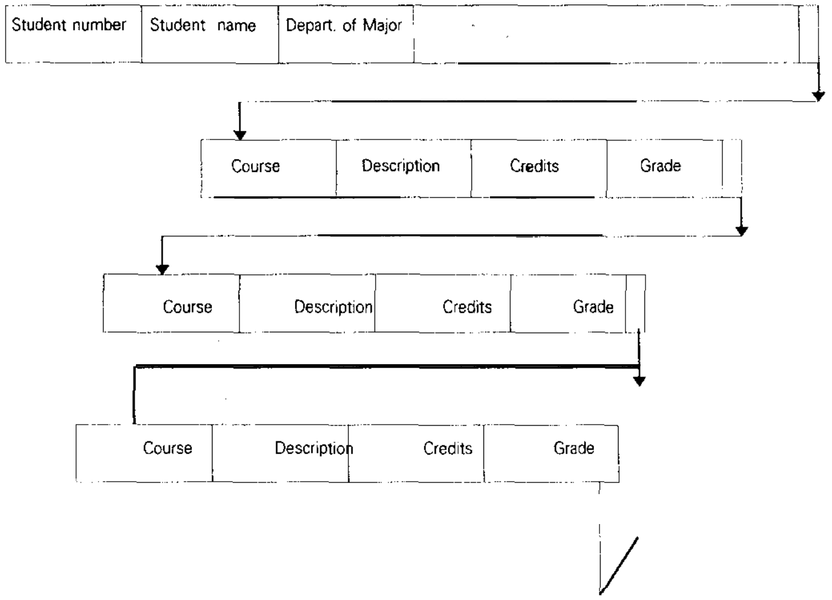


ภาพที่ 13.5 ตัวอย่าง ผังระบบของการประมวลผลแบบทันทีทันใดในระบบงานปรับยอด การฝาก - ถอนเงินของธนาคาร (Banking System)



แฟ้มข้อมูลที่เราออกแบบนั้นบางครั้งก็อาจจะต้องขยายไปเป็น "ระบบฐานข้อมูล " ถ้าหากว่าเราไม่สามารถออกแบบด้วยแฟ้มๆเดียวได้ให้ตรงกับวัตถุประสงค์ของการใช้งาน ความหมายของฐานข้อมูลก็คือแฟ้มข้อมูลที่มีความสัมพันธ์ต่อกัน ในปัจจุบันนี้มีโปรแกรมสำเร็จรูปที่บริหารฐานข้อมูลอยู่หลายโปรแกรมด้วย ตัวอย่างเช่น FOXBASE , PROGRESS และ ORACLE เป็นต้น

ภาพที่13.6 ตัวอย่างของระบบฐานข้อมูลของนักศึกษาที่ลงทะเบียนเรียนวิชาต่างๆ



รูปแบบของฐานข้อมูลแบบใดที่จะเหมาะสมนั้นขึ้นอยู่นักออกแบบฐานข้อมูล ซึ่งในหนังสือเล่มนี้จะไม่กล่าวเรื่องของฐานข้อมูลมากนักเพราะโดยปกติแล้วเรื่องบริหารฐานข้อมูลที่สลับซับซ้อนนั้นเรามักจะอาศัยโปรแกรมสำเร็จรูปที่ใช้บริหารฐานข้อมูล (Data Base Management System) อยู่แล้ว

นอกเหนือจากการออกแบบแฟ้มข้อมูลดังที่กล่าวมาแล้ว นักวิเคราะห์ระบบยังจำเป็นที่จะต้องดำเนินการประมาณการในเรื่องของพื้นที่ที่ใช้ในการจัดเก็บแฟ้มข้อมูลซึ่งจะต้องเผื่อไว้สำหรับการเพิ่มขนาดของข้อมูลต่อไปในอนาคตอีกด้วย ดังนั้นเราจึงควรจะคาดการว่าในอนาคตแฟ้มข้อมูลนั้นจะเติบโตขึ้นในอัตราร้อยละเท่าไรในแต่ละปีเพื่อรองรับการเก็บข้อมูลของระบบ การกำหนดพื้นที่ว่างที่เหมาะสมเพื่อรองรับการเติบโตของข้อมูลนั้น โดยเฉพาะกับการจัดการกับแฟ้มข้อมูลแบบสุ่ม เพราะถ้าหากแฟ้มดังกล่าวเกิดเพิ่มมากขึ้นเนื่องจากเราอาจมีลูกค้ามากขึ้น หรือขายสินค้ามากชนิดขึ้น ซึ่งจะทำให้เราค่อนข้างที่จะจัดการลำบากมากถ้าแฟ้มนั้นเป็นแบบสุ่มถึงแม้ว่าเราจะซื้ออุปกรณ์มาใหม่ เพราะโดยธรรมชาติของแฟ้มแบบสุ่มนั้นเรายังต้องอ้างอิงถึงตำแหน่งข้อมูลในอุปกรณ์ถึงแม้ว่าเราจะใช้ดัชนีชี้แบบตรรก (Logical Key) ก็ตาม แนวทางหนึ่งจะช่วยแก้ปัญหาในเรื่องนี้ก็คือจะต้องมีการคาดการณ์หรือพยากรณ์ปริมาณข้อมูลที่จะปรากฏในอนาคต

การพยากรณ์อัตราการเติบโตของข้อมูล โดยอาศัยแนวทางพยากรณ์ในทางสถิติเข้าช่วยแล้ว จึงทำการกำหนดว่าระบบที่สร้างขึ้นมานั้นควรจะอยู่ได้นานเท่าไร เช่น 3 ปี หรือ 5 ปี หรือมากกว่า แล้วนำจำนวนข้อมูลที่พยากรณ์ได้ไปประมาณการในเรื่องของการเลือกอุปกรณ์ต่อไป

การพยากรณ์ (Forecast) แบ่งออกเป็น 2 แนวทางคือ

1. ใช้เทคนิคทางสมการถดถอย (Regression Line) โดยการพิจารณาว่าตัวแปรมีขนาดของแฟ้มข้อมูลนั้นไปมีสัมพันธภาพกับตัวแปรอะไรบ้างในระบบ ตัวอย่างเช่น จำนวนลูกค้าของธนาคารจะสัมพันธ์กับอัตราการเจริญเติบโตทางเศรษฐกิจ และขนาดของประชากรที่เพิ่มขึ้น เป็นต้น เมื่อหาตัวแปรทั้งหลายที่คาดว่าจะเป็นมูลเหตุให้ จำนวนลูกค้าของธนาคารเพิ่มขึ้น จากนั้นนำข้อมูลของสิ่งที่คาดการณ์มาตรวจสอบโดยอาศัยวิธีการทางสถิติ เพื่อสร้างสมการในการพยากรณ์

$$Y_t = f(X_{1t}, X_{2t}, \ldots, X_{kt})$$

โดยที่ Y_t หมายถึง จำนวนของลูกค้าในปีที่ t

X_{1t} หมายถึงภาวะของเศรษฐกิจในปีที่ t

X_{2t} หมายถึงจำนวนประชากรที่ปรากฏในปีที่ t

X_{kt} หมายถึงดัชนีผู้บริโภคในปีที่ t

ภายหลังเมื่อใช้วิธีการทางสถิติคัดเลือกสมการถดถอยที่คิดว่าเหมาะสมที่สุดแล้ว เราก็นำไปใช้ในการพยากรณ์ปริมาณข้อมูลที่คาดว่าจะปรากฏในอนาคต เพื่อกำหนดขนาดของสื่อบันทึกและอุปกรณ์ที่ใช้ รวมทั้งโครงสร้างของแฟ้ม และการออกแบบรหัสข้อมูลที่เหมาะสม ตัวอย่างเช่นภายหลังการคัดเลือกแบบจำลองแล้วได้ผลออกมาเป็นรูปแบบพยากรณ์ดังนี้คือ

$$Y_t = 3 + 0.5 X_{1t} + 0.12 X_{2t}$$

โดยที่

X_{1t} : ภาวะเศรษฐกิจในปีใดๆที่ต้องการ

X_{2t} : จำนวนประชากรในปีใดๆที่ต้องการ

2. ใช้เทคนิคในเรื่องของเวลา (Time Series) ใช้ในกรณีที่วิธีที่ 1 ไม่เหมาะสม ด้วยเหตุที่ว่าข้อมูลนั้นขนาดเปลี่ยนไปตามเวลา หรืออีกนัยหนึ่งข้อมูลของปีที่ t จะขึ้นอยู่กับข้อมูลของ ปีที่ t - 1 ดังตัวแบบของแบบจำลองดังนี้คือ

$$Y_t = f(Y_{t-1})$$

ตัวอย่างเช่นเราสร้างสมการในการพยากรณ์โดยวิธีนี้แล้วได้แบบจำลองเป็นดังนี้คือ

$$Y_t = 3.1 + 0.15\,Y_{t-1}$$

โดยที่ Y_t = ปริมาณข้อมูลปี 2535

Y_{t-1} = ปริมาณข้อมูลปี 2534

ดังนั้น $Y_{2540} = 3.1 + 0.15\,Y_{2539}$

หมายเหตุ หลักการคำนวณพื้นเราใช้เก็บบันทึกข้อมูลนั้น ในสถานะการณ์ปกตินั้นเรามักจะคำนวณง่ายๆ กำหนดพื้นที่ส่วนที่เราจะต้องสำรองโดยการคำนวณจากสูตร loading factor ดังนี้คือ

$$\text{loading factor} = \frac{\text{จำนวนระเบียนข้อมูลที่ปรากฏ}}{\text{จำนวนพื้นที่ที่กำหนด}}$$

โดยที่ loading factor ที่เหมาะสมก็คืออยู่ในช่วง 70 % - 80%

การคำนวณขนาดของแฟ้มที่จะเติบโตในอนาคตนั้นจะมีผลต่อการกำหนดขนาดพื้นที่ที่เก็บ ซึ่งหมายถึงการเลือกขนาดของอุปกรณ์ที่เหมาะสม นอกจากนั้นยังทำให้เวลาที่เสียไปในการบำรุงดูแลรักษาแฟ้ม (Maintainance File) ลดลง ความหมายของการบำรุงดูแลรักษาแฟ้มข้อมูลนั้นมีความหมายในหัวข้อต่อไปนี้คือ

1. การปรับโครงสร้างของแฟ้มข้อมูล (Restructuring) ซึ่งหมายถึงการจัดการอย่างไรอย่างหนึ่งหรือหลายอย่างกับแฟ้มข้อมูลในลักษณะดังต่อไปนี้คือ

. การเปลี่ยนแปลงรูปแบบ (Format) ของรายการข้อมูลในแฟ้ม

. การเพิ่มรายการข้อมูลใหม่ๆที่ต้องการใช้งานเพิ่มเข้าไปในแฟ้ม เช่นเพิ่มรายละเอียดเกี่ยวกับครอบครัวเพื่อประกอบการคิดคำนวณหักเงินเข้ากองทุนสงเคราะห์

. การเปลี่ยนแปลงดัชนีสืบค้นจากแฟ้มแบบสุ่ม

. การปรับโครงสร้างแฟ้มจากแฟ้มเรียงลำดับแบบประเภท Single Record File มาเป็นแฟ้มประเภท Multiple Record File

2. การปรับปรุงเปลี่ยนแปลงข้อมูลแต่ละระเบียนในแฟ้ม ซึ่งการดำเนินการอาจจะประกอบด้วยกิจกรรมต่อไปนี้

. การเพิ่มระเบียนข้อมูล เช่นเพิ่มคนงาน ลูกค้ารายใหม่เข้ามาในระบบแฟ้มข้อมูล

. การลบระเบียนข้อมูลที่ไม่ใช้แล้วทิ้งไป เช่นหนังสือที่ไม่ปรากฏอยู่ในห้องสมุดอีกแล้ว

. การเปลี่ยนรายการข้อมูลบางรายการ เช่นลูกค้าที่รับวารสารจากสำนักพิมพ์ขอเปลี่ยนแปลงที่อยู่ คนงานมีการเปลี่ยนแปลงตำแหน่งหน้าที่ ฯลฯ

แบบฝึกหัดท้ายบท

1. การมีพจนานุกรมมีประโยชน์อย่างไรกับนักวิเคราะห์ระบบและผู้เขียนโปรแกรม
2. จงอธิบายถึงสารสนเทศที่เราจะต้องเขียนในพจนานุกรมข้อมูลว่าประกอบด้วยอะไรบ้าง
3. แฟ้มข้อมูลแบบเข้าถึงทีละลำดับกับแฟ้มข้อมูลแบบเข้าถึงแบบตรงแตกต่างกันอย่างไร และการนำไปใช้งานนั้นเหมาะสมกับระบบงานแต่ละประเภทอย่างไร
4. จงกล่าวถึงโปรแกรมสำเร็จรูปที่นำมาใช้ในการสร้างพจนานุกรมข้อมูล