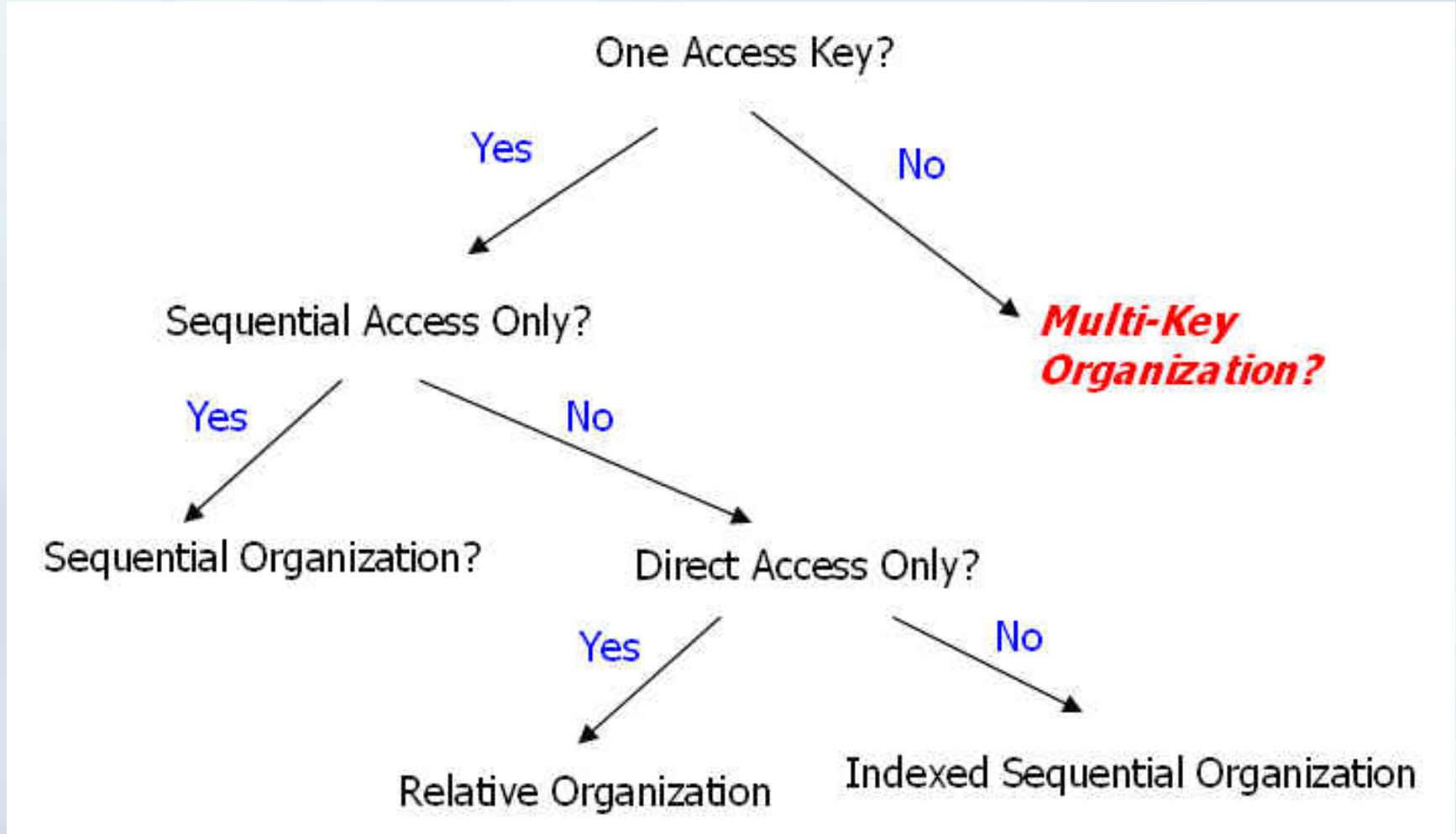


# บทที่ 8

## การจัดแฟ้มข้อมูลแบบคีย์หลายตัว

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์

# ประเภทของการจัดโครงสร้างไฟล์



# ลักษณะโครงสร้างของ Multiple-Key

- จากหัวข้อที่ผ่านมา จะมีการพูดถึงเฉพาะไฟล์ประเภทต่างๆ ที่ใช้ค่าคีย์เดี่ยวๆ (Single Key File) ได้แก่ ไฟล์แบบ Sequential ซึ่งจัดเรียงลำดับเรคอร์ดตามค่าของคีย์หนึ่งฟิลด์ ไฟล์แบบ Relative ซึ่งเราสามารถเข้าถึงตัวเรคอร์ดได้โดยตรง โดยระบุค่าคีย์เฉพาะค่าใด ค่าหนึ่ง และไฟล์แบบ Index Sequential ซึ่งเป็นไฟล์ที่สามารถเข้าถึงได้โดยตรง หรือเข้าถึงแบบจัดเรียงไปตามลำดับก็ได้ โดยการใส่ค่าคีย์ที่สามารถจัดเรียงลำดับเรคอร์ดและเป็นทั้งคีย์ที่ใช้ระบุถึงเรคอร์ดใดก็ได้

# องค์ประกอบของ multiple key file

- Primary key
- Secondary key
- Non indexed

# ตัวอย่าง

- ระบบบุคลากร เข้าถึงแบบสุมโดยใช้ รหัสบุคลากร เข้าถึงแบบสุมโดยใช้รหัสแผนก เข้าถึงแบบสุมโดยใช้เงินเดือน
- ระบบคลังสินค้าของบริษัทให้เช่ารถ ซึ่งมีรายละเอียดเป็น รหัสรถ ยี่ห้อ แบบ รุ่น ระยะทางวิ่งมาแล้ว และสี โดยจัดให้สามารถเข้าถึงข้อมูลโดย ใช้ รหัสรถ ยี่ห้อ แบบ ค้นหาข้อมูล

# เรคอร์ดของลูกค้าธนาคาร

Acct. no.	Name		Group-Code		SOCND	Balance	Overdraw limit
	Last	First	Branch	Type			

# กลุ่มผู้ใช้ระบบธนาคารที่ต้องการเข้าถึงข้อมูลในลักษณะแตกต่างกันคือ

- พนักงานเก็บเงิน ต้องการตรวจสอบเรคอร์ดบัญชีโดยผ่าน เลขบัญชี (Account No.)
- พนักงานฝ่ายสินเชื่อต้องการดูข้อมูลทั้งหมดตาม Overdraw Limit ที่กำหนด หรือผ่านทาง SOCNO
- ผู้จัดการสาขา ต้องการดูข้อมูลผ่าน Branch และ Type ของ Group-Code
- พนักงานธนาคาร ต้องการรายงานประจำวัน สัปดาห์สำหรับลูกค้าทั้งหมดที่เรียงตาม เลขบัญชี
- ลูกค้าธนาคาร (ผู้ถือบัญชี) จะต้องสามารถเข้าถึงเรคอร์ดของตัวเองโดยผ่าน เลขบัญชีหรือคีย์รวมของ NAME SOCNO และ Type Code

# ต้องมีแฟ้มบัญชีหลาย ๆ ชุดดังนี้

- Indexed Sequential account file เรียงตามคีย์เลขบัญชี (บริการ พนักงานธนาคาร และลูกค้ำธนาคาร)
- Sequential account file เรียงตาม OVERDEAW-LIMIT (บริการพนักงานฝ่ายสินเชื่อ)
- Direct file ใช้ SOCNO เป็น คีย์ (บริการพนักงานฝ่ายเงินกู้)
- Sequential file อีกชุดหนึ่งเรียงตาม GROUP-CODE (บริการผู้จัดการสาขา)
- Direct file อีกชุดหนึ่งใช้ NAME SOCNO และ TYPE code เป็น คีย์รวม (บริการผู้ถือบัญชี หรือลูกค้ำธนาคาร)



- ในการสร้างความผูกพันจะมีการผูกพันโดย
  - ให้คีย์รองเป็นตัวแปรต้น และ
  - ไพรมารีคีย์ หรือ Relative Address เป็นตัวแปรตาม เพื่อช่วยขยายความสามารถในการเรียกใช้ข้อมูลของระบบ

# ตัวอย่างสินค้าคงคลังของบริษัทให้เช่ารถยนต์

Primary Key	Secondary Key		Non Indexed		
Car ID	Brand	Style	Model	Mile	Color
C1	Dodge	2 DR	Omni	25,000	Green
H1	Honda	4 DR	Accord	32,000	Yellow
F1	Ford	2 DR HB	Escort	54,000	White
GM1	Cadillac	4 DR	Fleetwood	9,000	Red
GM2	Oldsmobile	2 DR	Delta	28,000	Blue
H2	Honda	2 DR HB	Accord	11,250	Brown
F2	Lincoln	4 DR	Continental	38,000	Black
C2	Dodge	2 DR	Aspen	7,000	Silver
F3	Ford	2 DR	Thunderbird	33,000	Green
GM3	Cheroleet	2 DR	Camaro	63,000	Blue
GM4	Cadillac	2 DR	Cimaron	11,000	White
GM5	Oldsmobile	4 DR	98	3,400	Red
T1	Toyota	2 DR HB	Celica	57,000	Blue

# ความหมาย Multiple-Key

- Multiple-Key วิธีการที่สร้างคีย์หลายๆ แบบ เพื่อให้สามารถเข้าถึงเรคอร์ดในไฟล์ ทำให้การสอบถาม หรือเข้าถึงข้อมูลแบบสุ่ม สามารถเข้าถึงได้หลายฟิลด์
- *ลักษณะการจัดไฟล์แบบนี้มี 2 แบบ*
  - **Inverted**
  - **Multilist**

# ความจำเป็นของ Multiple Access

- ในระบบสารสนเทศที่เน้นการโต้ตอบ (Interactive Information System) ส่วนใหญ่จะต้องการไฟล์ที่ใช้ค่าคีย์หลายคีย์ในการเข้าถึงและเรียกใช้เรคอร์ดในไฟล์
- วิธีการ อาจสร้างไฟล์หลายๆไฟล์ที่มีข้อมูลซ้ำๆ กัน เพื่อเข้าถึงข้อมูลแบบต่างๆกัน
- *ปัญหา คือ เปลืองเนื้อที่ เกิดความซ้ำซ้อนของข้อมูล และ ในกรณีแก้ไขข้อมูล อาจแก้ไขไม่ครบถ้วน และเกิดความขัดแย้งได้*

# ลักษณะทั่วไปของ Multiple Key

- วิธีการแก้ไขปัญหาการเข้าถึงข้อมูลโดยผู้ใช้อย่างต่างกัน จัดให้ไฟล์มีลักษณะการเข้าถึงข้อมูลได้หลายทิศทาง เป็น Multiple Access Path ใช้วิธีการเข้าถึงได้หลายรูปแบบตามความต้องการของผู้ใช้
- แนวคิดพื้นฐานของการจัดไฟล์แบบนี้ คือสร้างหลายดัชนี (Multiple Indexes) ที่จะสามารถชี้ เรคอร์ดที่ต้องการได้จากหลายทิศทาง
- ค่า Index ที่ใช้ในการจัดการไฟล์ Multiple Key จะมีมากกว่า 1 คีย์ มี Primary Key และ Secondary Key ต่างๆ ที่ต้องการค้นหาด้วย

# การจัดโครงสร้างแบบ Inverted File

- โครงสร้างแบบ Multiple-Key เป็นการจัดไฟล์ที่อาศัยความสัมพันธ์ผกผัน (Inverse Relationship) ระหว่างข้อมูลในเรคอร์ดกับค่า Primary Key ของเรคอร์ดนั้น ไฟล์ของข้อมูลจะมีค่าดัชนีผกผันกับค่าคีย์ของเรคอร์ดในไฟล์นั้น เหมือนกับดัชนีของหนังสือ ถือได้ว่ามีค่าผกผันกับหน้าของหนังสือ

# การจัดโครงสร้างแบบ Inverted File (ต่อ)

## ไฟล์แบบ Inverted ประกอบด้วย 2 ส่วน

- Directory or inversion indexes ประกอบด้วยรายการผกผัน (Inverse List) ของแต่ละ Secondary Key ซึ่งนำมาใช้เรียกข้อมูลในส่วนของ Data Record Area โดยอาศัย Primary Key เข้าไปดึงข้อมูลออกมา
- Data Record Area : อาจจัดเก็บแบบ Relative, sequential, or indexed sequential ที่มีค่า Primary Key เป็นการเข้าถึงข้อมูลแบบสุ่ม

# การจัดโครงสร้างแบบ Inverted File (ต่อ)

- รายการชื่อฟิลด์ใน Inverted List อาจจัดเก็บเป็นไฟล์ย่อยๆ ไฟล์ละ Secondary Key ก็ได้ หรืออาจจัดเก็บรายการฟิลด์ที่เป็น Secondary Key ทั้งหมดไว้ใน 1 ไฟล์ ก็ได้ แต่ละหน่วย (Entry) ในรายการ Inverted List ของแต่ละ Secondary Key จะประกอบด้วย 1 ฟิลด์ ที่เป็น Secondary Key ในไฟล์



# Creating Inverted list

Brand	Primary key
Dodge	C1

Brand	Primary key
Dodge	C1
Honda	H1

Brand	Primary key
Dodge	C1
Honda	H1
Ford	F1
Cadillac	GM1
.	.
.	.

Style	Primary key
2 DR	C1

Style	Primary key
2 DR	C1
4 DR	H1


Style	Primary key
2 DR	C1
4 DR	H1
2 DR	F1

Style	Primary key
2 DR	C1
4 DR	H1, GM1,
2 DR	GM2

# Directory of Brand and Style

Brand	Primary Keys
Dodge	C1, C2
Honda	H1,H2
Ford	F1,F3
Cadillac	GM1,GM4
Oldsmobile	GM2,GM5
Lincoln	F2
Chevrolet	GM3
Toyota	T1

## key Data file contents

key	Model	Mile	Color 
C1	Omni	25,000	Green
H1	Accord	32,000	Yellow
F1	Escort	54,000	White
GM1	Fleetwood	9,000	Red
GM2	Delta	28,000	Blue
H2	Accord	11,250	Brown
F2	Continental	38,000	Black
C2	Aspen	7,000	Siler
F3	Thunderbird	33,000	Green
GM3	Camaro	63,000	Blue
Gm4	Cimaron	11,000	White
Gm5	98	3,400	Red
T1	Celica	57,000	Blue

Style	Primary keys
2 DR	C1,C2,F3,GM3,GM4
4 DR	H1,GM1,GM2,F2,GM5
2 DR HB	F1,H2,T1

# ตัวอย่างข้อมูล ข้อมูลระบบนักศึกษา

รหัส	ชื่อ-สกุล	เพศ	อายุ	รหัสภาควิชา	รหัสวิชาเรียน	คะแนน
103	AAA	F	18	Com	111	75
115	BBB	F	18	Phy	111	83
120	CCC	M	19	Com	111	90
135	DDD	M	18	Com	150	77
137	EEE	M	19	Bio	111	65
142	FFF	F	18	Phy	150	75
150	HHH	M	18	Chm	109	70

# สร้าง Inverting File และ Data File

- ต้องการค้นข้อมูลแบบสุ่มด้วย *รหัสนักศึกษา*
- ต้องการค้นข้อมูลแบบสุ่ม ว่านักศึกษาที่สังกัดภาคนั้นๆ มีใครบ้าง สามารถเลือกค้นได้จาก *รหัสภาควิชา*
- ต้องการค้นข้อมูลแบบสุ่ม มีใครบ้างที่เรียนวิชารหัสต่างๆ สามารถเลือกค้นได้จาก *รหัสวิชาเรียน*

# สร้าง Inverting File และ Data File (ต่อ)

รหัสวิชา **รหัสตน.**

111 103, 115, 120, 137

150 135, 142

109 150

รหัสภาค **รหัสตน.**

Com 103, 120, 135

Phy 115, 142

Bio 137

Chm 150

รหัส	ชื่อ-สกุล	เพศ	อายุ	รหัสภาควิชา	รหัสวิชาเรียน	คะแนน
103	AAA	F	18	Com	111	75
115	BBB	F	18	Phy	111	83
120	CCC	M	19	Com	111	90
135	DDD	M	18	Com	150	77
137	EEE	M	19	Bio	111	65
142	FFF	F	18	Phy	150	75
150	HHH	M	18	Chm	109	70

# สอบถามข้อมูล แล้วจะค้นอย่างไร

- ถามข้อมูลเกี่ยวกับนักศึกษา ถ้าทราบรหัส ค้นได้จาก Data File
- ถามว่ามีใครบ้างอยู่ภาควิชา Com
- ถามว่ามีใครบ้างอยู่ภาควิชา Bio
- ถามว่า มีใครบ้างเรียนวิชา 111
- ถามว่า มีใครบ้างเรียนวิชา 111 และอยู่ภาค Com
- *ถ้าบอกชื่อ จะค้นได้ อย่างไร ควรทำอะไร เพิ่มอย่างไร เพื่อให้การทำงานมีประสิทธิภาพมากยิ่งขึ้น*

## **ประโยชน์ของ Inverted file**

- สามารถค้นหาข้อมูลได้ตามเงื่อนไขได้ง่าย เช่น ต้องการหารถ ยี่ห้อ Dodge style 2 ประตุน
- ประหยัดเนื้อที่ ในส่วนของ ข้อมูล ไม่ต้องเก็บส่วนของ Secondary key

## **ข้อด้อยของ Inverted file**

เสียเวลา ในการค้นหา หากมีค้นหาในหลายเงื่อนไข

# การออกแบบ Inverted File

## ควรพิจารณา

- ส่วนของ Directory or inversion indexes ว่าต้องการค้นหาข้อมูลผ่านเขตข้อมูลใดบ้าง ควรสร้าง ส่วน inverted เท่าที่จำเป็น
- ส่วนของ Data Record Area ซึ่งใช้ได้ทั้งแบบ Relative, sequential, or indexed sequential



## การสร้างองค์กรแฟ้มหลายคีย์แบบ Multi-list

- Multi-List File สามารถทำงานได้เช่นเดียวกับ Inverted File แต่สร้างได้ง่ายกว่า ถึงแม้ว่าจะมีประสิทธิภาพด้อยกว่า และเพื่อลดปัญหาของ Inverted List ในเรื่อง ความยาวของรายการที่อยู่ของ เรคอร์ดที่มีขนาดแตกต่างกัน ทำให้เรามีการจัด Inverted List ออกเป็นหลาย Inverted list ด้วยการหาทางเชื่อม Secondary key เดียว ลักษณะการเชื่อมโยงกันของ Secondary key นี้ เราเรียกว่า Multi-List

Brand	Primary key	length
BENZ	A45	4
HONDA	A23	5
MASDA	Y56	2
NISSAN	G45	4
TOYOTA	B34	5

**Multi-list ของ Brand**

Style	Primary key	length
2 DR	A23	8
2 DR HB	B34	4
4 DR	A45	5
4 DR HB	G45	3

**Multi-list ของ Style**

Record address	Primary key	Secondary keys				NON-Indexed data		
	Car ID	brand	Next	Style	Next	Model	Mileage	Color
1	A23	HONDA	G44	2 DR	Y56	CIVIC	25,000	WHITE
2	A45	BENZ	G32	4 DR	Y44	E500	34,000	GREEN
3	B34	TOYOTA	Y44	2 DR HB	T65	COLOLA	13,500	BLUE
4	G45	NISSAN	W43	4 DR HB	A58	SUNNY SUPER	12,340	GREEN
5	Y56	MASDA	K25	2 DR	G44	SETIA	45,250	GRAY
6	G44	HONDA	T65	2 DR	T55	CIVIC	5,600	BROWN
7	Y44	TOYOTA	W12	4 DR	W43	CAMRY	45,000	SIVER
8	T65	HONDA	A58	2 DR HB	G32	ACCORD	56,000	SIVER
9	W43	NISSAN	T55	4 DR	D36	TIANA	25,000	WHITE
10	T55	NISSAN	L54	2 DR	W12	SUNNY NEO	15,000	RED
11	W12	TOYOTA	D36	2 DR	K25	YARIS	25,000	YELLOW
12	A58	HONDA	Q88	4 DR HB	F45	CIVIC	38,000	GREEN
13	K25	MASDA	-	2 DR	L54	FAMILIA	13,500	BLUE
14	G32	BENZ	K39	2 DR HB	Q88	E180	18,340	GREEN
15	D36	TOYOTA	F54	4 DR	L68	CAMRY	44,250	GRAY
16	L54	NISSAN	-	2 DR	K39	MURAN	5,600	BROWN
17	F45	TOYOTA	-	4 DR HB	-	KLUGER	45,000	YELLOW
18	K39	BENZ	L68	2 DR	-	300SE	56,000	SIVER
19	L68	BENZ	-	4 DR	-	220SE	35,000	WHITE
20	Q88	HONDA	-	2 DR HB	-	ELEMENT	25,000	RED

## Data ของ Multi-key File

## เปรียบเทียบแฟ้มข้อมูล inverted file และ multi - list ส่วนที่เหมือนกัน

1. ทั้งคู่ต่างก็มี Secondary Key Index และมี Index Entry สำหรับทุกๆ ค่าของ Secondary Key
2. โครงสร้างของดรรชนีของทั้งคู่ อาจเป็นตารางหรือ Tree
3. Entry ของส่วนคีย์และส่วน Pointer อาจเรียงหรือไม่เรียงตามลำดับก็ได้
4. Pointer อาจเป็น Direct หรือ Indirect Address ก็ได้

## ความแตกต่างของดรรชนีของ 2 แบบนี้ คือ

1. Inversion Index มี Pointer สำหรับทุกๆ เรคอร์ดของคีย์ที่กำหนด จึงมีขนาดไม่คงที่
2. Multi-List Index มี Pointer เฉพาะเรคอร์ดแรกของคีย์ที่กำหนด เท่านั้น ดังนั้น Multi – List Index จึงมีขนาดคงที่
3. ส่วนของแฟ้มข้อมูลจะไม่ถูกระทบถ้าใช้ Inversion Index แต่ หากใช้ Multi – List Index โครงสร้างของแฟ้มจะกลายเป็น Linked List ของเรคอร์ดสำหรับคีย์ที่กำหนด
4. Multi-List จะเหมาะกับการถามคำถามที่ไม่สลับซับซ้อนมาก ซึ่ง หากคำถามซ้อนกันหลายชั้นแบบ Inversion Index จะง่ายกว่า