บทที่ 3

การพิจารณาการจัดแฟ้มข้อมูลหรือวิธีปฏิบัติข้อมูล

วิธีการจัดแฟ้มข้อมูลหรือการปฏิบัติข้อมูลมีหลายแบบด้วยกัน แบบที่แพร่หลายและรู้จักกันดีที่สุดคือ แฟ้มข้อมูลแบบต่อเนื่อง (SEQUENTIAL FILE) เป็นแฟ้มข้อมูลที่บันทึกข้อมูลแต่ละระเบียน (RECORD) เรียงไปตามลำดับของส่วนซึ่งใช้เป็นรหัส (KEY FIELD) ลักษณะการจัดแฟ้มข้อมูลแบบนี้ทำให้ต้องปฏิบัติข้อมูลแต่ละระเบียนไปตามลำดับ ตั้งแต่ระเบียนแรกไปอย่างต่อเนื่องจนถึงระเบียนสุดท้าย ซึ่งอาจทำเป็นงวด (BATCH) แล้วแต่จะกำหนดขึ้น

นอกจากนี้ยังมีแฟ้มข้อมูลแบบอื่น ซึ่งเราสามารถจะปฏิบัติกับข้อมูลไปที่ระเบียนต่าง ๆ ในแฟ้มข้อมูลโดยตรง ไม่ต้องปฏิบัติต่อเนื่องอย่างแบบแรกคือ แฟ้มข้อมูลแบบสุ่ม (RANDOM FILE หรือ DIRECT ACCESS FILE) และแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี (INDEXED SEQUENTIAL FILE) ซึ่งจะกล่าวถึงเป็นลำดับต่อไป

การที่จะตัดสินว่าจะเลือกวิธีปฏิบัติข้อมูลในงานต่าง ๆ อย่างไรจึงจะทำให้เกิดประสิทธิภาพสูงสุดนั้น ย่อมขึ้นกับสภาวะหรือลักษณะของงานนั้น ๆ หลักการที่จะใช้ประเมินว่าสภาวะงานแบบใด ควรเลือกวิธีจัดแฟ้มข้อมูลในการปฏิบัติงานอย่างไร มีหลักที่จะพิจารณา 6 ประการด้วยกันคือ

1) ปริมาณรายการที่จะเปลี่ยนแปลงในแต่ละงวดของการปฏิบัติข้อมูล
2) อัตราความเร็วในการตอบสนองต่อการเปลี่ยนแปลง (RESPONSE TIME TO INQUIRIES)
3) ความคงสภาพของแฟ้มข้อมูล
4) ขนาดของแฟ้มข้อมูล (FILE SIZE)

5) ความเกี่ยวพันหรือสัมพันธ์ระหว่างแฟ้มข้อมูลต่าง ๆ

6) ความปลอดภัยและระบบสำรองข้อมูล (FILE SECURITY)

3.1 การจัดแฟ้มข้อมูลแบบต่อเนื่อง (SEQUENTIAL FILE)

แฟ้มข้อมูลแบบต่อเนื่อง มีลักษณะการบันทึกข้อมูลโดยเรียงข้อมูลแต่ละระเบียนตามลำดับของส่วนที่ใช้เป็นรหัส ระเบียนทั้งหมดจะถูกบันทึกในตัวกลางเก็บข้อมูลแบบใดแบบหนึ่งก็ได้ เช่นเทปแม่เหล็ก บัตรเจาะรู หรือเทปกระดาษ การปฏิบัติหรือเปลี่ยนแปลงข้อมูลแต่ละระเบียนในแฟ้มข้อมูลแบบนี้ ต้องทำอย่างต่อเนื่องตั้งแต่ระเบียนแรกเป็นต้นไปจนถึงระเบียนสุดท้าย แม้ระเบียนใดที่ไม่ต้องการแก้ไขก็จะต้องถูกอ่านตรวจสอบด้วยแล้วจึงจะละเว้นไม่แก้ไข ข้ามไปยังระเบียนต่อไป วิธีปฏิบัติข้อมูลแบบต่อเนื่องนี้จึงเท่ากับเป็นการปฏิบัติกับข้อมูลในแฟ้มข้อมูลทุกรายการนั่นเอง จากการประเมินโดยใช้หลักทั้ง 6 ประการที่กล่าวข้างต้น สามารถประเมินวิธีจัดแฟ้มข้อมูลแบบต่อเนื่องออกมาเป็นข้อ ๆ ดังนี้

1) ปริมาณรายการเปลี่ยนแปลง หมายถึงปริมาณของรายการต่าง ๆ ในแฟ้มข้อมูลที่จะถูกแก้ไขหรือเปลี่ยนแปลงโดยระเบียนที่ส่งมาเป็นงวด ๆ

จากการใช้วิธีปฏิบัติข้อมูลแบบต่อเนื่อง รายการในแฟ้มข้อมูลหลักเดิม (OLD MASTER FILE) จะถูกอ่านขึ้นมาทุกระเบียนไม่ว่าจะมีการแก้ไขหรือไม่ แล้วจึงถ่ายทอดไปยังแฟ้มข้อมูลหลักใหม่ (NEW MASTER FILE) โดยวิธีนี้ไม่ว่าจะเปลี่ยนแปลงรายการเพียงสอง สามระเบียนก็จะต้องใช้เวลาในการปฏิบัติงานเกือบจะเท่าเดิมเพราะเวลาที่คอมพิวเตอร์ใช้ปฏิบัติต่อรายการที่เปลี่ยนแปลง เมื่อเทียบกับเวลาในการอ่านและบันทึกข้อมูลแล้ว ความเร็วของตัวกลางประมวลผลมีความเร็วกว่าความเร็วของอุปกรณ์อ่านและบันทึกมาก

ในการปฏิบัติข้อมูลแบบนี้ค่าใช้จ่ายในการปฏิบัติข้อมูลต่อรายการเปลี่ยนแปลงหนึ่ง ๆ จะยิ่งลดลงถ้ามีรายการเปลี่ยนมากขึ้น ดังนั้นอาจไม่เป็นการประหยัดที่จะใช้วิธีปฏิบัติข้อมูลแบบต่อเนื่องกับแฟ้มข้อมูลที่มีรายการเปลี่ยนแปลงน้อย และต้องการแก้ไขรายการบ่อยครั้งในช่วงนั้น ๆ

2) อัตราความรวดเร็วในการตอบสนองต่อการเปลี่ยนแปลงหมายถึงช่วงเวลานับแต่รายการเปลี่ยนแปลงเกิดขึ้นจนถึงเวลาที่ได้แก้ไขรายการนั้นกับรายการในแฟ้มข้อมูลหลักเรียบร้อยแล้ว ถ้าต้องการให้ช่วงเวลาตอบสนองสั้นเท่าใดก็จะต้องทำการปฏิบัติข้อมูลถี่มากขึ้นเท่านั้น ถ้าทำการปฏิบัติข้อมูลในช่วงสั้นขึ้นจะทำให้ปริมาณของรายการเปลี่ยนแปลงในแต่ละงวดลดน้อยลง ทั้งนี้ค่าใช้จ่ายต่อรายการหนึ่ง ๆ จะสูง ดังนั้นถ้าแฟ้มข้อมูลในงานใดต้องการให้ช่วงเวลาในการตอบสนองสั้นมาก เช่น แฟ้มข้อมูลที่ต้องการแก้ไขรายการตลอดเวลาย่อมไม่เหมาะสมที่จะปฏิบัติข้อมูลแบบต่อเนื่อง

3) ความคงสภาพของแฟ้มข้อมูล หมายถึงปริมาณการเพิ่มหรือลดของจำนวนระเบียนในแฟ้มข้อมูล ระหว่างช่วงการปฏิบัติข้อมูลงวดหนึ่ง ๆ ยกตัวอย่างเช่น แฟ้มข้อมูลหลักของบัญชีเงินเดือนความคงสภาพของแฟ้มข้อมูลจะขึ้นกับอัตราหมุนเวียนของพนักงานว่ามีการรับเข้าใหม่หรือลาออกมากน้อยเพียงใด ในการปฏิบัติข้อมูลแบบต่อเนื่อง การเพิ่มหรือลดของจำนวนระเบียนในแฟ้มข้อมูลหลักทำได้ง่าย เพราะใช้วิธีถ่ายทอดข้อมูลจากแฟ้มข้อมูลหลักเดิมไปยังแฟ้มข้อมูลหลักตัวใหม่อยู่แล้ว การเพิ่มหรือลดจำนวนระเบียนจึงเป็นเพียงผลพลอยได้จากการปฏิบัติข้อมูลปกติ

4) ขนาดของแฟ้มข้อมูล หมายถึงปริมาณของข้อมูลในแฟ้ม เนื่องจากการบันทึกข้อมูลลงในเทปแม่เหล็กจะเสียค่าใช้จ่ายต่อไบท์ (BYTE) น้อยกว่าการใช้สื่อกลางชนิดอื่น เช่น บัตรเจาะรู หรือแผ่นจานแม่เหล็ก ดังนั้นถ้าแฟ้มข้อมูลใดมีปริมาณข้อมูลมาก และเป็นข้อมูลที่ไม่สู้จะมีการเคลื่อนไหวบ่อยนัก จึงนิยมบันทึกในเทปแม่เหล็ก เพราะจะประหยัดกว่า

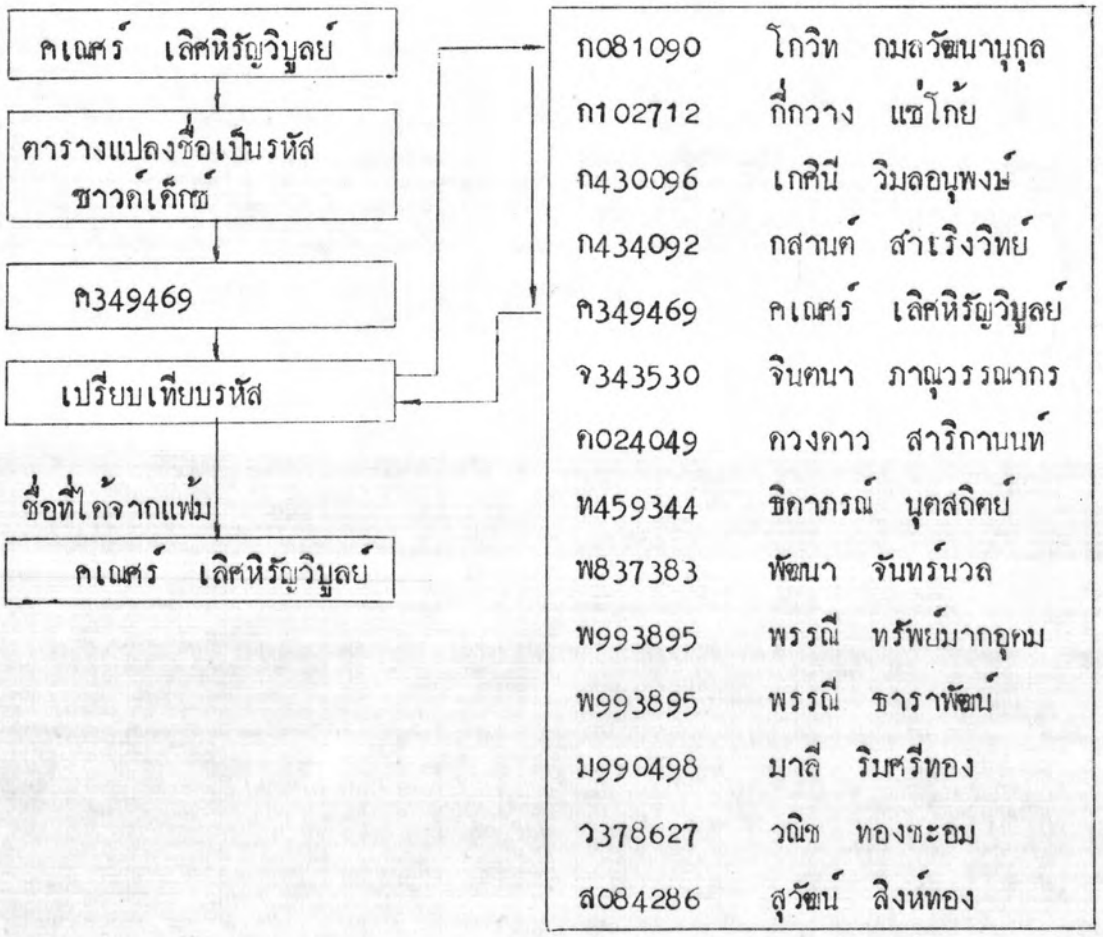
5) ความเกี่ยวพันของข้อมูลในระหว่างแฟ้มข้อมูลต่าง ๆ ถ้าแฟ้มข้อมูลต่าง ๆ มีความสัมพันธ์หรือเกี่ยวพันกันอย่างใกล้ชิด และในการปฏิบัติข้อมูลต้องใช้แฟ้มข้อมูลที่เกี่ยวข้องพร้อมกันแล้ว การใช้วิธีปฏิบัติแบบต่อเนื่องอาจไม่เหมาะสม เพราะระบบการเรียงรหัสในแต่ละแฟ้มข้อมูลอาจต่างกัน และถ้าจะทำได้ก็อาจจะต้องออกแบบโปรแกรมอย่างซับซ้อนซึ่งจะทำให้การปฏิบัติงานขาดประสิทธิภาพ

6) ความปลอดภัยและระบบสำรองข้อมูล ถ้าใช้การปฏิบัติข้อมูลแบบต่อเนื่องโดยใช้ตัวกลางบันทึกข้อมูลที่เป็นเทป ย่อมเป็นการสร้างระบบสำรองข้อมูลในตัวอยู่แล้ว แฟ้มข้อมูลหลักเดิมหลังการแก้ไขข้อมูลในงวดนั้นแล้ว ไม่ได้ถูกลบทิ้งจึงอาจเก็บไว้เป็นแฟ้มข้อมูลสำรองได้โดยอัตโนมัติ แต่ถ้าเป็นแฟ้มข้อมูลแบบต่อเนื่องซึ่งบันทึกในแผ่นจานแม่เหล็ก ระบบสำรองข้อมูลและรักษาความปลอดภัยของข้อมูลอาจทำได้ลำบากหลังการแก้ไขข้อมูลในแฟ้มข้อมูลแล้ว มักนิยมถ่ายทอดข้อมูลจากจานแม่เหล็กขึ้นไปบันทึกไว้บนเทปแม่เหล็กเป็นครั้งคราว เพื่อให้มีข้อมูลสำรอง ข้อเสียที่สำคัญที่สุดของการจัดแฟ้มข้อมูลแบบต่อเนื่องบนแผ่นจานแม่เหล็ก ก็คือละเลยความสามารถพิเศษของจานแม่เหล็กที่อาจบรรจุข้อมูลแบบสุ่มได้

เนื่องจากการปฏิบัติข้อมูลแบบต่อเนื่องมีทั้งข้อดี ข้อเสียแฝงอยู่ในตัว ดังนั้นจึงมีผู้คิดหาวิธีปฏิบัติข้อมูลใหม่ เพื่อหลีกเลี่ยงการปฏิบัติข้อมูลแบบต่อเนื่องกับบางงานซึ่งมีข้อเสีย วิธีดังกล่าวคือ วิธีปฏิบัติข้อมูลแบบสุ่ม และวิธีปฏิบัติข้อมูลแบบต่อเนื่องมีดัชนี อย่างไรก็ตามการที่จะปฏิบัติข้อมูลด้วยสองวิธีนี้ต้องอาศัยอุปกรณ์บรรจุข้อมูลที่ปฏิบัติถึงข้อมูลทุกรายการได้โดยตรง ซึ่งได้แก่แผ่นจานแม่เหล็ก เป็นต้น

ตัวอย่างแฟ้มข้อมูลแบบต่อเนื่องและวิธีค้นหา แสดงไว้ในรูปที่ 3.1 แสดงการเก็บระเบียนของชื่อและรหัสแบบซาวด์เด็กซ์ โดยในการค้นหาจะใช้ชื่อเข้าไปค้นหาซึ่งจะต้องมีการแปลงให้เป็นรหัสแบบซาวด์เด็กซ์ก่อน

ในรูปที่ 3.1 ใช้ชื่อ คเณศร์ เลิศหิรัญวิบูลย์ เป็นชื่อใช้ค้นหา โปรแกรมจะแปลงชื่อเป็นรหัสซาวด์เด็กซ์แล้วใช้เป็นรหัสค้นหาข้อมูลในแฟ้มข้อมูลหลัก การเปรียบเทียบจะต้องอ่านข้อมูลตั้งแต่ระเบียนแรกเปรียบเทียบเรียงตามลำดับมาจนถึงระเบียนที่มีรหัสเป็น ค349469 ซึ่งตรงกับรหัสของชื่อที่ใช้ค้นหาได้ระเบียนที่เป็น คเณศร์ เลิศหิรัญวิบูลย์ ออกมาจากแฟ้มข้อมูลหลัก



รูปที่ 3.1 แสดงแฟ้มข้อมูลแบบต่อเนื่องและวิธีค้นหา

3.2 การจัดแฟ้มข้อมูลแบบสุ่ม (RANDOM หรือ DIRECT ACCESS)

ถ้าแฟ้มข้อมูลที่ใช้จัดอยู่ในประเภทงานที่มีปริมาณของรายการเปลี่ยนแปลงน้อย หรือต้องการระบบที่จะเข้าปฏิบัติกับรายการใด ๆ ในแฟ้มข้อมูลได้เร็ว การจัดระบบแฟ้มข้อมูลแบบสุ่มจะเป็นทางเลือกที่เหมาะสม ซึ่งการจัดแฟ้มข้อมูลแบบสุ่มจะทำให้ความยากง่ายในการเข้าถึงรายการต่าง ๆ ในแฟ้มข้อมูลใกล้เคียงกัน และมีความสามารถที่จะเข้าถึงรายการใด ๆ ในแฟ้มข้อมูลได้โดยตรง โดยไม่ต้องอ่านรายการอื่นที่ไม่เกี่ยวข้องก่อน ดังเช่นการเข้าถึงรายการที่ต้องการในแฟ้มข้อมูลแบบต่อเนื่อง (ในทางปฏิบัติจริง ๆ ของแฟ้มข้อมูลแบบสุ่มอาจต้องอ่านรายการอื่นบ้างก่อนเล็กน้อย แต่อนุโลมให้ถือว่าเป็นการปฏิบัติเข้าถึงรายการใด ๆ โดยตรงได้)

อุปกรณ์บันทึกข้อมูลที่จัดอยู่ในประเภทปฏิบัติข้อมูลแบบสุ่มหรือโดยตรงได้ คือแผ่นจานแม่เหล็ก และกระบอกแม่เหล็ก อย่างไรก็ตามช่วงเวลาที่ใช้ในการอ่านข้อมูลหนึ่ง ๆ ยังขึ้นอยู่กับตำแหน่งที่ข้อมูลนั้นบรรจุอยู่ ตัวอย่างเช่น ถ้าอ่านข้อมูลที่บันทึกอยู่ในไซลินเดอร์ (CYLINDER) เดียวกัน การอ่านรายการหนึ่ง ๆ อาจใช้ช่วงเวลาเพียง 1,000 ไมโครวินาที แต่ถ้ารายการที่ต้องการอ่านบันทึกในไซลินเดอร์ถัดไป ทำให้ต้องขยับหัวอ่านบันทึกด้วย ช่วงเวลาที่อ่านอาจกลายเป็น 100,000 ไมโครวินาที

สรุปแล้วเวลาที่ต้องเสียไปในการอ่านรายการแต่ละรายการจะมีดังนี้

1) ขั้นแรกต้องเสียเวลารอให้หัวอ่านบันทึกขยับไปอยู่ในตำแหน่งของไซลินเดอร์ที่ต้องการก่อน เวลาที่เสียไปในช่วงนี้เรียกว่า เวลาในการค้นหา (SEEK TIME)

2) เนื่องจากมีแผ่นจานแม่เหล็กหลายแผ่นในไซลินเดอร์หนึ่ง ๆ ดังนั้นจึงต้องเลือกหัวอ่านบันทึกที่เหมาะสม ช่วงเวลานี้เป็นช่วงของเวลาในการเลือกหัวอ่านบันทึก (HEAD SELECTED TIME)

3) เวลาที่เสียไปในการคอยให้รายการที่ต้องการเวียนมาอยู่ในตำแหน่งของหัวอ่านบันทึก (LOTATIONAL DELAY หรือ LATENCY)

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

4) เมื่ออ่านข้อมูลที่ต้องการแล้วย่อมจะต้องมีการถ่ายทอดไปยังจุดหมายที่ต้องการ เวลาที่ใช้ไปเรียกว่า เวลาในการถ่ายทอด (TRANSFER TIME)

ประเมินผลการปฏิบัติข้อมูล สำหรับแฟ้มข้อมูลแบบสุ่ม มีดังนี้

1) ด้านปริมาณของรายการเปลี่ยนแปลง การจัดแฟ้มข้อมูลแบบสุ่มโดยการใช้รหัสค้นหา แม้จะสามารถค้นหารายการในแฟ้มข้อมูลได้รวดเร็ว แต่เวลาโดยเฉลี่ยในการค้นหาแต่ละรายการจะใกล้เคียงกันเพราะรายการกระจัดกระจายอยู่ทั่วไป ดังนั้นค่าใช้จ่ายในการค้นหาเพื่อแก้ไขข้อมูลต่อหนึ่งรายการจะคงที่ ไม่ว่าจะแก้ไขจำนวนมากเท่าใดก็ตาม ทั้งนี้ต่างกับการปฏิบัติข้อมูลแบบต่อเนื่อง ซึ่งถ้าในงวดหนึ่ง ๆ มีจำนวนรายการที่ต้องการค้นหาเพื่อแก้ไขมาก ค่าใช้จ่ายในการปฏิบัติข้อมูลต่อรายการหนึ่ง ๆ จะลดลงเป็นลำดับ

2) ด้านอัตราความเร็วในการตอบสนองต่อการเปลี่ยนแปลง เนื่องจากค่าใช้จ่ายในการค้นหาต่อรายการคงที่เสมอ ดังนั้นไม่ว่าจะกำหนดให้เวลาตอบสนอง (RESPONSE TIME) เป็นทุกวินาทีหรือทุกวัน ค่าใช้จ่ายต่อรายการแก้ไขจะคงที่ ทั้งนี้ผิดกับการใช้วิธีแบบต่อเนื่อง ซึ่งยิ่งต้องการให้อัตราการตอบสนองรวดเร็วเพียงใด ก็ยิ่งต้องเสียค่าใช้จ่ายสูงขึ้นเท่านั้น การใช้เทปแม่เหล็กอัตราความเร็วของการตอบสนองยิ่งมีข้อจำกัด เพราะในการทำงานอาจต้องหมุนเทปกลับเพื่อค้นหาข้อมูลซึ่งย่อมทำได้ช้า ดังนั้นจึงไม่อาจกำหนดช่วงเวลาของการตอบสนองให้ถี่เกินข้อจำกัดนี้

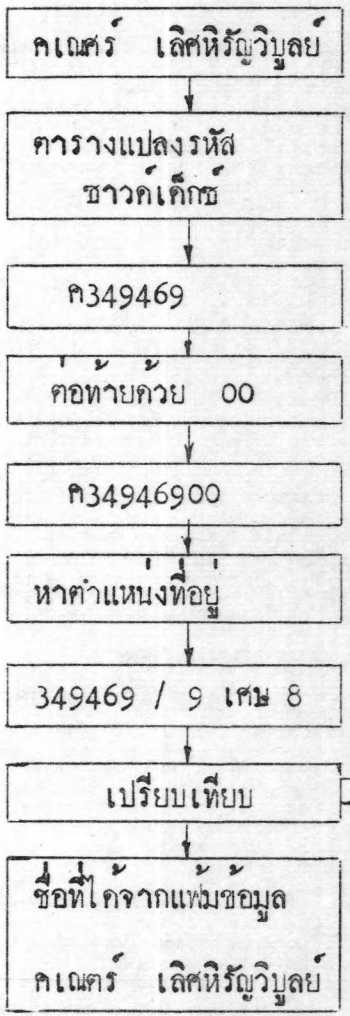
3) ด้านความคงสภาพของแฟ้มข้อมูลในการบรรจุข้อมูลแบบสุ่ม การเพิ่มหรือลดจำนวนรายการในแฟ้มข้อมูลทำได้ง่ายมาก อย่างไรก็ตามต้องมีระบบควบคุมมิให้เมื่อลบรายการใดแล้วเผลอลบที่อยู่ (ADDRESS) ซึ่งเชื่อมโยงไปยังพื้นที่รับข้อมูลส่วนเกิน (OVERFLOW AREA) เพราะจะทำให้รายการส่วนเกินนั้นสูญหาย เมื่อเก็บรักษาแฟ้มข้อมูลนานไปจนพบว่ามีรายการส่วนเกินมากหรือได้ลบรายการจนมีที่ว่างทิ้งเปล่ากระจัดกระจายภายในช่วงเวลาที่เหมาะสม อาจมีการลบล้างที่ว่างเปล่าเหล่านี้ โดยถ่ายทอดข้อมูลในแฟ้มข้อมูลไปยังแผ่นจานแม่เหล็กที่ใหม่ ซึ่งมีขนาดเล็กหรือใหญ่กว่าเดิม เพื่อจัดระบบใหม่ตามความเหมาะสม

4) ขนาดของแฟ้มข้อมูล ถ้าข้อมูลที่ไม่สู้จะมีการเคลื่อนไหวหรือไม่ต้องการระบบตอบสนองต่อการเปลี่ยนแปลงที่รวดเร็วและแฟ้มข้อมูลมีขนาดใหญ่ การใช้ระบบข้อมูลแบบสุ่มก็ไม่สู้จะเหมาะสมนักควรจะใช้บันทึกข้อมูลลงเทปเป็นแบบต่อเนื่องจะเป็นการเหมาะสมกว่าและประหยัดกว่า เพราะเทปถูกกว่าแผ่นจานแม่เหล็กราว 10 เท่า

5) ความเกี่ยวพันกันของข้อมูลในระหว่างแฟ้มข้อมูลต่าง ๆ การบรรจุข้อมูลแบบสุ่ม ทำให้การปฏิบัติข้อมูลที่มีการเกี่ยวพันกันเป็นไปอย่างสะดวกเพราะอาจเข้าถึงข้อมูลใด ๆ ก็ได้ในเวลาอันรวดเร็ว นอกจากนั้นยังอาจบรรจุแฟ้มข้อมูลหลายแฟ้มข้อมูลไว้ในแผ่นจานแม่เหล็กเดียวกัน ซึ่งเปิดโอกาสให้สามารถรวมการปฏิบัติข้อมูลแบบต่าง ๆ ไว้ในงวดการปฏิบัติข้อมูลครั้งเดียวกันได้สะดวก

6) ความปลอดภัยและระบบสำรองข้อมูล การบันทึกข้อมูลและการปฏิบัติข้อมูลที่บรรจุในแผ่นจานแม่เหล็กย่อมขาดระบบสำรองข้อมูลโดยอัตโนมัติเช่นการใช้เทปแม่เหล็ก

ตัวอย่างแฟ้มข้อมูลแบบสุ่มและวิธีค้นหา แสดงไว้ในรูปที่ 3.2 ซึ่งแสดงถึงการเก็บระเบียนของชื่อและรหัสแบบซาวค์เด็กซ์ โดยใช้รหัสแบบซาวค์เด็กซ์เป็นรหัสสำหรับค้นหา แฟ้มข้อมูลแบบสุ่มซึ่งแสดงในรูปที่ 3.2 นี้ใช้เก็บระเบียนของชื่อซึ่งมีรหัสค้นหาที่แปลงจากชื่อโดยใช้การจัดกลุ่มที่ 2 เช่นเดียวกับรหัสในแฟ้มข้อมูลแบบต่อเนื่อง แต่รหัสที่ใช้ในแฟ้มข้อมูลแบบสุ่มจะเพิ่มความยาวของรหัสเป็นตัวเลขอีก 2 ตัวต่อท้าย คือถ้าชื่อใดที่แปลงแล้วได้รหัสซ้ำกัน เช่นซ้ำกัน 3 ชื่อรหัสของระเบียนแรกตัวท้ายจะใส่เป็น 00 ระเบียนที่ 2 ตัวท้ายจะใส่เป็น 01 ระเบียนที่ 3 ตัวท้ายจะใส่เป็น 02 ถ้ารหัสที่แปลงได้ไม่ซ้ำกันเลยมีตัวเดียวจะใส่ตัวเลขท้ายรหัสเป็น 00 ในรูปตัวอย่างให้แต่ละที่เก็บ (HOME BUCKET) เก็บระเบียนได้ 2 ระเบียนข้อมูล (BUCKET CAPACITY = 2) และไม่ได้แสดงที่เก็บข้อมูลส่วนเกินไว้ (OVERFLOW BUCKET) ตำแหน่งที่อยู่ของระเบียนจะอยู่ที่เก็บใดใช้เศษจากการหารตัวเลขในรหัสตำแหน่ง 2 ถึง 7 ด้วย 9 (ซึ่งรหัสนี้แปลงได้จากชื่อ) เป็นตัวชี้ ในที่นี้ต้องการค้นหาระเบียนชื่อ คเณศร์ เลิศหิรัญวิบูลย์ แปลงเป็นรหัสแล้วได้ ค34946900 หาตำแหน่งที่เก็บโดยหาค่าตัวเลขรหัสตำแหน่งที่ 2 ถึง 7



0	ก08109000	โกวิท กมลวัฒนานุกูล	จ34353000	จินตนา ภาณุวรรณากร
1	ค02404900	ควงดาว สาริกานนท์	ส08428600	สุวัฒน์ สิงห์ทอง
2	ท45934400	ธิดาภรณ์ นุตสถิตย์		
3	ม99049800	มาลี ริมศรีทอง		
4	ก10271200	กี่กวาง แซ่โก้ย	ก43409200	กสานต์ สำเริงวิทย์
5				
6				
7	พ99389500	พรรณี ทรัพย์มากอุดม	พ99389501	พรรณี ธาราพัฒน์
8	ค34946900	คเณศร์ เลิศหิรัญวิบูลย์		
9				

รูปที่ 3.2 แสดงแฟ้มข้อมูลแบบสุ่มและวิธีค้นหา

ด้วย 9 (349469/9) ได้เศษ 8 จะไปค้นหาชื่อในที่เก็บหมายเลข 8 ได้ชื่อ คเณศ เลิศหิรัญวิบูลย์ ต่อไปจะบวกค่า 1 เข้าที่ 2 ตัวท้ายของรหัสได้ค่า ค34946901 ใช้เป็นรหัสค้นหาไปยังที่เก็บที่ 8 อีกครั้ง แต่ปรากฏว่าในแฟ้มไม่มีรหัสนี้อยู่ ถือว่าไม่พบข้อมูลหรือไม่มีข้อมูลที่มีรหัสซ้ำจะสิ้นสุดการค้นหา

3.3 การจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี (INDEXED SEQUENTIAL FILE)

เนื่องจากการจัดแฟ้มข้อมูลแบบต่อเนื่องและแบบสุ่ม มีทั้งข้อดีและข้อเสียไปคนละด้านซึ่งมักจะขัดกัน ในที่สุดจึงมีผู้พยายามจะประนีประนอมโดยเอาลักษณะการจัดระบบของทั้งสองแบบมาผสมกันโดยเรียกว่า วิธีปฏิบัติข้อมูลแบบต่อเนื่องมีดัชนี (INDEXED SEQUENTIAL ACCESS METHOD) หรือเรียกย่อ ๆ ว่า 'ISAM'

ระบบการจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี ช่วยให้สามารถทำการปฏิบัติข้อมูลแบบต่อเนื่องเป็นงวดได้ ซึ่งทำให้ประหยัดค่าใช้จ่าย และในขณะเดียวกันก็สามารถเข้าถึงรายการใด ๆ ในแฟ้มข้อมูลได้คล้าย ๆ กับแบบสุ่มซึ่งทำได้รวดเร็ว แนวคิดในการจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี มีหลักการเช่นเดียวกับการค้นหาคำในดิกชันนารี ถ้าทำแบบต่อเนื่องเราจะต้องใช้เวลาในการอ่านตั้งแต่เริ่มต้นจนกว่าจะถึงคำที่ต้องการทุกครั้ง แต่ถ้าดิกชันนารีเล่มนั้นมีสารบัญ เราอาจเริ่มต้นค้นจากการดูสารบัญว่าศัพท์ที่ขึ้นต้นด้วยตัวอักษรที่เราต้องการค้นหานั้นเริ่มจากหน้าที่เท่าใด แล้วพลิกไปที่หน้านั้นโดยตรง จากนั้นจึงค้นหาจนพบคำที่ต้องการ จะเห็นได้ว่าวิธีนี้ช่วยตัดทอนเวลาที่ใช้ในการค้นหาลงได้มากมายหลายเท่า เมื่อนำแนวคิดนี้มาใช้ในการจัดระบบแฟ้มข้อมูล ในขั้นแรกจะบันทึกข้อมูลแบบต่อเนื่องให้รายการเรียงตามส่วนที่ใช้เป็นรหัส จากนั้นจะสร้างดัชนีไว้ที่จุดเริ่มต้นของการค้นหาให้ทำหน้าที่คล้ายสารบัญ การจัดระบบดัชนี มักจะจัดในลักษณะ ดัชนีย่อยตามลำดับขั้น (INDEX HIRACHY)

ประเมินผลการปฏิบัติข้อมูลสำหรับแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี มีดังนี้

1) ด้านปริมาณของรายการเปลี่ยนแปลง จะสังเกตได้ว่าวิธีจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี จะไม่ประหยัดเมื่อเทียบกับวิธีการจัดแบบต่อเนื่อง ถ้าปริมาณการเปลี่ยนแปลงในงวดการปฏิบัติข้อมูลหนึ่ง ๆ มีมากเมื่อมีรายการเปลี่ยนแปลงมากเกือบทุกรายการในแฟ้มข้อมูลแล้ว ดัชนีจะเป็นสิ่งไม่จำเป็น ทำให้เสียค่าใช้จ่ายในการสร้างโดยใช่เหตุ เพราะต้องเข้าถึงทุกรายการในลักษณะเรียงกันอยู่แล้ว ในทำนองเดียวกันถ้ารายการเปลี่ยนแปลง

มีน้อย การใช้วิธีจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนีย่อมต้องเสียค่าใช้จ่ายแพงกว่าการใช้วิธีการจัดข้อมูลแบบสุ่ม เพราะต้องเสียค่าใช้จ่ายในการสร้างดัชนี และการจะเข้าถึงรายการใด ๆ จะต้องเสียเวลาอ่านดัชนีก่อน

สรุปแล้วการจัดระบบแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี เป็นวิธีอยู่ระหว่างกลาง ถ้าปริมาณรายการเปลี่ยนแปลงมีน้อยแล้ว จะสู้วิธีการจัดแบบสุ่มไม่ได้ และถ้าปริมาณรายการเปลี่ยนแปลงมีมากก็จะสู้วิธีการจัดแบบต่อเนื่องไม่ได้

2) อัตราความเร็วในการตอบสนองต่อการเปลี่ยนแปลง วิธีจัดระบบแบบต่อเนื่องมีดัชนี เป็นวิธีที่อยู่ระหว่างกลาง กล่าวคือถ้าต้องการให้อัตราการตอบสนองต่อการเปลี่ยนแปลงถี่ การจัดระบบแบบต่อเนื่องมีดัชนีจะต้องเสียค่าใช้จ่ายแพงกว่าการจัดแบบสุ่ม แต่ถ้าอัตราการตอบสนองต่ำ การจัดระบบแบบต่อเนื่อง จะประหยัดกว่าเพราะไม่ต้องสร้างดัชนี

3) ด้านความคงสภาพของแฟ้มข้อมูล ถ้ามีการเพิ่มหรือลดจำนวนรายการในแฟ้มข้อมูลอยู่เสมอ ส่วนที่กันไว้สำหรับเป็นพื้นที่รับข้อมูลส่วนเกินอาจเต็ม ทำให้ต้องล้นไปอยู่ในไซลินเดอร์อื่น ทำให้หัวอ่านบันทึกต้องเคลื่อนไหว ดังนั้นต้องมีช่วงเวลาที่สมควรทำการจัดแฟ้มข้อมูลใหม่ (REORGANIZE) เพื่อจัดระบบดัชนีขึ้นใหม่อยู่เสมอเหตุนี้จึงพอจะสรุปได้ว่า ถ้ามีการเพิ่มหรือลดจำนวนรายการในแฟ้มข้อมูลอยู่เสมอครั้งละจำนวนมาก การใช้วิธีจัดแฟ้มข้อมูลแบบต่อเนื่องจะมีประสิทธิภาพ

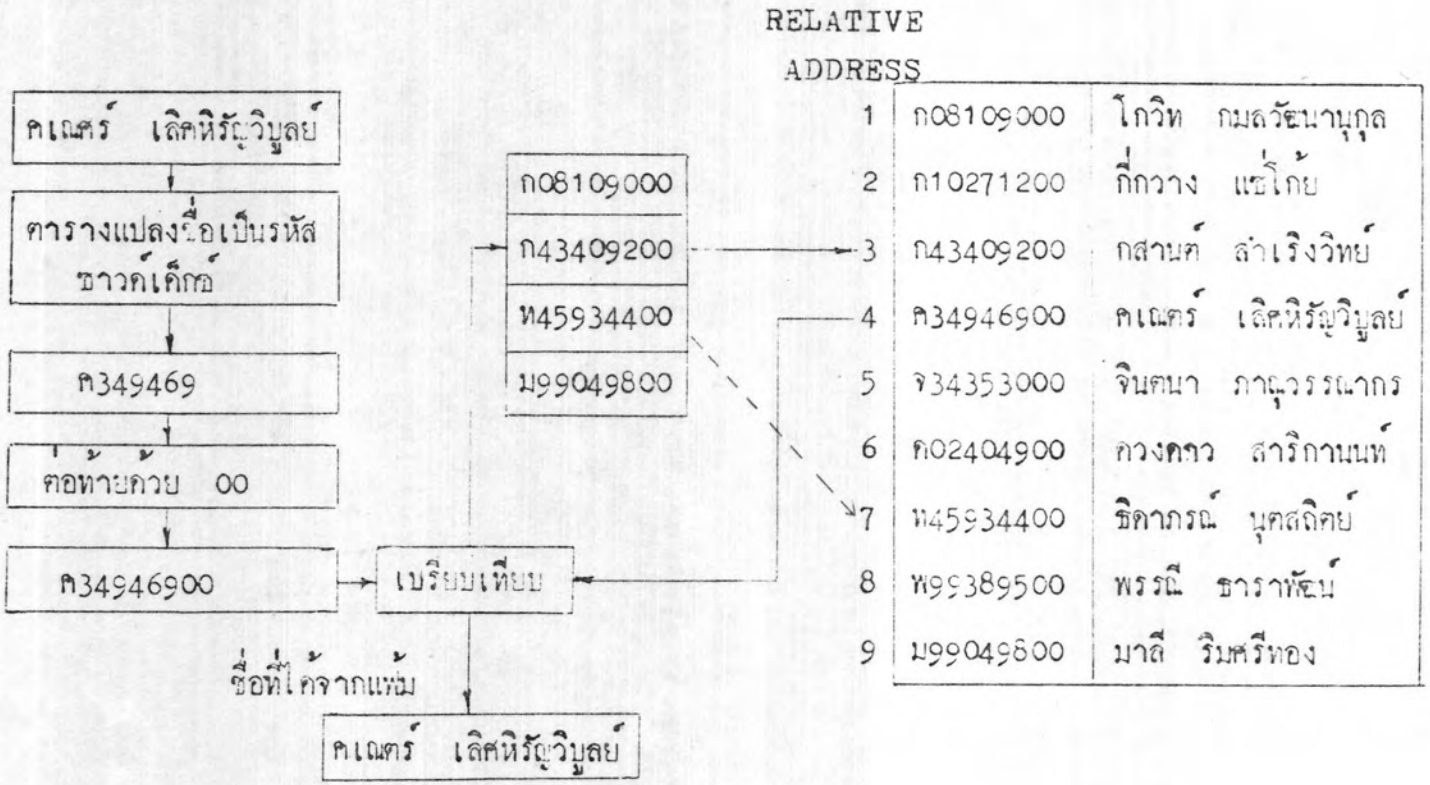
โดยสรุปวิธีจัดระบบแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี เปิดโอกาสให้ได้ผลประโยชน์จากการปฏิบัติข้อมูลเป็นงวดแบบต่อเนื่อง แต่ในขณะเดียวกันก็สามารถเข้าถึงรายการใด ๆ ในแบบกึ่งการปฏิบัติโดยตรง วิธีการจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี เป็นวิธีจัดระบบข้อมูลแบบที่อยู่ระหว่างกลางของแบบต่อเนื่องและแบบสุ่ม

ระบบการจัดแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี เป็นที่นิยมแพร่หลายในวงการธุรกิจ ซึ่งต้องการทราบข้อมูลบางข้อมูลที่บรรจุอยู่ในอุปกรณ์จัดเก็บข้อมูลในทันทีทันใด เมื่อมีความต้องการซึ่งโดยมากมักจะแสดงข้อมูลที่ส่งออกมาทางจอภาพ การจัดระบบแบบนี้เรียกว่า

ออนไลน์ (ONLINE)ซึ่งหมายถึงการจัดระบบการทำงานของเครื่องคอมพิวเตอร์ในลักษณะที่เครื่องสามารถเข้าถึงข้อมูลใด ๆ ซึ่งบรรจุอยู่ในอุปกรณ์จัดเก็บได้โดยตรงด้วยตัวของมันเองตลอดเวลา

ตัวอย่างแฟ้มข้อมูลแบบต่อเนื่องมีดัชนีและวิธีค้นหา แสดงไว้ในรูปที่ 3.3 ซึ่งแสดงการเก็บระเบียนของชื่อและรหัสแบบชาวด์เด็กซ์ โดยใช้รหัสแบบชาวด์เด็กซ์เป็นรหัสสำหรับค้นหา การเก็บข้อมูลจะเรียบตามลำดับจากน้อยไปหามาก มีการแบ่งช่วงข้อมูลออกเป็นส่วน ๆ ซึ่งอาจแบ่งโดยส่วนของไซลินเดอร์ หรือ แทรคของแผ่นจานแม่เหล็ก ในที่นี้ไม่ได้แสดงที่เก็บข้อมูลส่วนเกิน

การค้นหาจะใช้รหัสที่แปลงได้จากชื่อ คเณศร์ เลิศหิรัญวิบูลย์ เข้าไปเปรียบเทียบกับรหัสในส่วนที่เก็บดัชนี ซึ่งที่เก็บดัชนีแรกจะมีค่าต่ำสุดของรหัสที่เก็บไว้ในที่เก็บข้อมูลแรกของแฟ้มข้อมูล ส่วนที่เก็บดัชนีอื่นจะเก็บค่าสูงสุดของรหัสในแต่ละที่เก็บหรือแต่ละแทรคไว้ เพื่อเปรียบเทียบจะได้รู้ว่าอยู่ในส่วนที่เก็บใด ก็จะไปทำการค้นหาในส่วนที่เก็บนั้น เพื่อนำข้อมูลตามรหัสออกมา ในรูปรหัส ค34946900 เปรียบเทียบกับส่วนที่เก็บดัชนีจะเห็นว่ามากกว่ารหัสของส่วนที่เก็บข้อมูลที่ 1 แต่น้อยกว่ารหัสของส่วนที่เก็บข้อมูลที่ 2 ดังนั้นข้อมูลที่ต้องการจะอยู่ในส่วนที่เก็บข้อมูลที่ 2 ก็ไปทำการค้นหาตั้งแต่ต้นส่วนที่เก็บข้อมูลที่ 2 จนพบข้อมูลที่ต้องการ คือ ค34946900 ได้ คเณศร์ เลิศหิรัญวิบูลย์ ออกมาจากนั้นจะค้นหาต่อไปแบบต่อเนื่องจนกระทั่งรหัสทั้ง 7 ตัวแรกนั้นเปลี่ยนก็จะได้ชื่อรหัสซ้ำจนครบนั่นคือจะได้ชื่อที่มีรหัสซ้ำออกมาจนครบ แต่ในที่นี้พออ่านต่อก็ได้รหัสเปลี่ยนเลย ดังนั้นจะได้ชื่อ คเณศร์ เลิศหิรัญวิบูลย์ ออกมาชื่อเดียว สิ้นสุดการค้นหา



รูปที่ 3.3 แสดงแฟ้มข้อมูลแบบต่อเนื่องมีดัชนีและวิธีค้นหา

3.4 พิจารณาเปรียบเทียบการจัดแฟ้มข้อมูล

จากการประเมินการจัดแฟ้มข้อมูลทั้ง 3 แบบที่กล่าวไว้ข้างต้น การจัดแฟ้มข้อมูลแต่ละแบบต่างมีข้อดีข้อเสียทั้งนั้น การเลือกใช้แฟ้มข้อมูลกับงานต่าง ๆ ควรพิจารณาให้เหมาะสม เพื่อให้สะดวกกับการใช้งานและประหยัด

สำหรับรหัสฮาวด์เด็กซ์ซึ่งเป็นรหัสที่สร้างขึ้นจากการแปลงชื่อเป็นรหัสประกอบด้วยตัวอักษรนำหน้าหนึ่งตัวและตามด้วยตัวเลขอีกแล้วแต่จะกำหนด ตัวอักษรและตัวเลขทุกตัวมีหลักการแปลงโดยการพิจารณาจากอักษรแต่ละตัวนั้นอยู่ในกลุ่มเสียงใด อักษรที่อยู่กลุ่มเสียงเดียวกันก็จะแทนด้วยตัวอักษรหรือตัวเลขเดียวกัน จากวิธีการสร้างรหัสดังกล่าว ถ้ามีจำนวนชื่อมาก ๆ โอกาสที่รหัสซ้ำกันจะมีมากด้วย ฮาวด์เด็กซ์เหมาะกับการประมวลผลจำพวกต้องการค้นหารายการต่าง ๆ จากชื่อ หรือต้องการค้นหารายการแต่จำเลขประจำรายการนั้นไม่ได้ก็อาจจะใช้ชื่อเป็นตัวในการค้นหาโดยตรงเลย ซึ่งการค้นหาแต่ละครั้งมีรายการที่ต้องการค้นหาไม่มากนัก

จากข้อดี ข้อเสีย ของการจัดแฟ้มข้อมูลแบบต่อเนื่อง ที่กล่าวมาแล้ว จะเห็นได้ว่าไม่สะดวกที่จะนำมาใช้กับการจัดแฟ้มข้อมูลเพื่อใช้กับฮาวด์เด็กซ์ เนื่องจากว่า ฮาวด์เด็กซ์นั้นเหมาะกับงานที่ต้องการการตอบสนองเร็ว จำนวนรายการแก้ไขแต่ละครั้งอาจจะมีไม่มากนักและไม่ได้เข้ามาในเวลาเดียวกันหมด

ดังนั้น ถ้าจัดแฟ้มข้อมูลแบบต่อเนื่องแล้วทุกครั้งที่มีรายการแก้ไขเข้ามา จะต้องทำการปฏิบัติตั้งแต่ต้นแฟ้มข้อมูลทุกครั้ง ทำให้เสียเวลาในการปฏิบัติข้อมูลมาก ยิ่งถ้าเป็นแฟ้มข้อมูลที่มีขนาดใหญ่มาก เช่นงานทะเบียนที่ดิน หรืองานธนาคาร เป็นต้น แฟ้มข้อมูลที่มีขนาดใหญ่มากนี้ ถ้าหากจัดแฟ้มข้อมูลแบบต่อเนื่องแล้วการปฏิบัติข้อมูลแต่ละครั้ง จะใช้เวลาในกลรปฏิบัติมาก นอกจากว่าจะเก็บรายการแก้ไขนั้น ๆ รวมไว้มาก ๆ แล้วทำการปฏิบัติข้อมูลทีเดียวซึ่งไม่สะดวก เพราะงานบางอย่างลูกค้าไม่ต้องการจะรอ เช่นงานธนาคาร เป็นต้น

i 17211943

ลองมาพิจารณาแฟ้มข้อมูลแบบสุ่มดู จะเห็นว่าแฟ้มข้อมูลแบบสุ่มนี้เหมาะกับงานประมวลผลผ่านสายที่ต้องการการตอบสนองข้อมูลแต่ละรายการคงที่เสมอ และเข้าถึงรายการนั้นได้โดยตรง น่าที่จะใช้กับซาวด์เด็กซ์ได้ แต่เมื่อพิจารณาในข้อที่ว่า ผู้ใช้จะต้องรับผิดชอบในการจัดการข้อมูลในพื้นที่รับข้อมูลส่วนเกินเอง ก็ออกจะเป็นการไม่สะดวกที่จะใช้กับซาวด์เด็กซ์ เนื่องจากว่าซาวด์เด็กซ์อาจจะต้องมีรหัสที่ซ้ำกันเกิดขึ้น ซึ่งข้อมูลส่วนที่มีรหัสซ้ำกันนี้ถ้าหากว่าไม่อาจบรรจุในพื้นที่ปกติแล้วก็จะต้องตกไปอยู่ในพื้นที่รับข้อมูลส่วนเกิน ซึ่งข้อมูลส่วนนี้ผู้ใช้จะต้องรับผิดชอบในการจัดการข้อมูลเอง

อีกวิธีหนึ่งที่จะพิจารณาคือแฟ้มข้อมูลแบบต่อเนื่องมีดัชนี แฟ้มข้อมูลแบบนี้มีลักษณะรวมของแฟ้มข้อมูลทั้ง 2 แบบที่กล่าวมา แฟ้มข้อมูลแบบนี้สามารถที่จะปฏิบัติเข้าถึงรายการข้อมูลต่าง ๆ ได้โดยตรง เช่นเดียวกับแฟ้มข้อมูลแบบสุ่ม สามารถใช้กับงานประมวลผลผ่านสายได้ ถึงแม้ว่าการให้ผลตอบสนองข้อมูลอาจไม่ดีเท่าแบบสุ่ม เพราะต้องเสียเวลาอ่านรายการอื่น ๆ ในกลุ่มซึ่งอยู่ใกล้กันบ้าง แฟ้มข้อมูลแบบนี้รายการต่าง ๆ จะเรียงลำดับกันอยู่ เมื่อเข้าถึงรายการหนึ่งโดยใช้ดัชนีเป็นตัวนำเข้าแล้ว สามารถที่จะประมวลผลต่อกับแฟ้มข้อมูลนี้ในลักษณะต่อเนื่องได้เลย

ดังนั้น การใช้แฟ้มข้อมูลแบบนี้จะสามารถแก้ข้อเสียของแฟ้มข้อมูลแบบสุ่มได้ โดยที่ผู้ใช้ไม่ต้องจัดการกับข้อมูลที่อยู่ในพื้นที่รับข้อมูลส่วนเกินเอง และในการเปลี่ยนแปลงแก้ไขข้อมูลก็ทำได้ง่ายกว่าการจัดแฟ้มข้อมูลแบบสุ่ม และในแฟ้มข้อมูลแบบนี้ถึงแม้ว่าจะมีรหัสซ้ำกันมากมายดังเช่นในซาวด์เด็กซ์ ก็สามารถจะถูกค้นหาออกมาได้หมดได้ โดยการเข้าถึงรายการแรกโดยดัชนี และเข้าถึงรายการต่อ ๆ ไปแบบต่อเนื่อง ซึ่งในการสร้างแฟ้มข้อมูล เราสามารถสร้างเลขเรียงลำดับ (SERIAL NO.) ต่อกันไปได้

สรุป จากการพิจารณาแฟ้มข้อมูลทั้ง 3 แบบจะเห็นได้ว่าแฟ้มข้อมูลแบบต่อเนื่องมีดัชนีเป็นแฟ้มข้อมูลที่ได้เปรียบที่สุดเหมาะที่จะใช้กับการประมวลผลข้อมูลซึ่งใช้ซาวด์เด็กซ์เป็นรหัสค้นหา เพราะสามารถให้ผลตอบสนองที่รวดเร็ว และข้อมูลทุกรายการที่เกี่ยวข้องหรือมีโอกาสที่จะเป็นรายการของรหัสที่แท้จริง จะถูกนำออกมาจากแฟ้มข้อมูลเพื่อแสดงได้หมดและสามารถที่จะเข้าถึงรหัสขนาดเท่าไรไปก็ได้ เช่นเริ่มจากขนาดรหัส 3 จนถึงขนาดรหัส 8 ก็จะทำได้โดยเริ่มจากขนาดรหัส 3 ตัวเช่นต้องการรหัสที่ขึ้นต้นด้วย ก01 ก็สามารถเข้าถึงรายการเริ่มแรกได้โดยเอา ก01000000 เป็นตัวเริ่มต้นค้นหา แล้วก็จะอ่านรายการอื่นต่อไปเรื่อย ๆ จนกระทั่ง 3 ตัวแรกเปลี่ยนดังนี้เป็นต้น ซึ่งจะเห็นได้ว่าการประมวลผลแบบนี้จะเป็นการประมวลผลที่ถูกต้องและสมบูรณ์ที่สุดเหมาะที่จะนำมาใช้กับซาวด์เด็กซ์มากที่สุด