

## สารบัญ

	หน้า
สารบัญ.....	(6)
รายการตาราง.....	(9)
รายการภาพประกอบ.....	(10)
บทที่	
1. บทนำ.....	1
1.1 ความเป็นมา.....	1
1.2 การตรวจเอกสารและงานวิจัยที่เกี่ยวข้อง.....	2
1.2.1 การตรวจเอกสาร.....	2
1.2.2 งานวิจัยที่เกี่ยวข้อง.....	2
1.3 วัตถุประสงค์.....	3
1.4 วิธีการดำเนินการวิจัย.....	3
1.5 ขอบเขตงานวิจัย.....	5
1.6 ขั้นตอนการดำเนินงาน.....	5
1.7 ระยะเวลาดำเนินงานและแผนการดำเนินงาน.....	5
1.8 เครื่องมือและอุปกรณ์.....	6
1.9 ประโยชน์ที่คาดว่าจะได้รับ.....	6
2. คลังข้อมูลและการค้นหาข้อมูล.....	7
2.1 ฐานข้อมูลดำเนินการและคลังข้อมูล.....	9
2.1.1 ฐานข้อมูลดำเนินการ (Operational Database).....	9
2.1.2 คลังข้อมูล (Data Warehouse).....	9
2.2 สถาปัตยกรรมของคลังข้อมูล.....	12
2.3 แบบจำลองเชิงแนวคิดของคลังข้อมูล.....	14
2.3.1 โครงสร้างแบบดาว (Star Schema).....	15
2.3.2 โครงสร้างฟลิคหิมะ (Snowflake Schema).....	16
2.4 การเข้าถึงข้อมูล.....	17
2.4.1 แฮชชิ่ง (Hashing).....	17
2.4.1.1 วิธีแก้ปัญหาคollision (Collision Resolution).....	19
2.4.1.2 ข้อดีของแฮชชิ่ง.....	21
2.4.1.3 ข้อจำกัดของแฮชชิ่ง.....	21

## สารบัญ (ต่อ)

	หน้า
2.4.2 ดัชนี (Indexing).....	21
2.4.2.1 ดัชนีแบบ B-Tree.....	21
2.4.2.1.1 ข้อดีของดัชนีแบบ B-Tree.....	22
2.4.2.1.2 ข้อจำกัดของดัชนีแบบ B-Tree.....	22
3. ดัชนีแบบบิตแมป (Bitmap Index).....	23
3.1 ดัชนีบิตแมปแบบพื้นฐาน (Simple Bitmap Index).....	24
3.1.1 การสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปแบบพื้นฐาน.....	25
3.1.2 ข้อดีของดัชนีบิตแมปแบบพื้นฐาน.....	26
3.1.3 ข้อจำกัดของดัชนีบิตแมปแบบพื้นฐาน.....	26
3.2 ดัชนีบิตแมปแบบช่วง (Interval Bitmap Index).....	26
3.2.1 การสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปแบบช่วง.....	28
3.2.2 ข้อดีของดัชนีบิตแมปแบบช่วง.....	29
3.2.3 ข้อจำกัดของดัชนีบิตแมปแบบช่วง.....	29
3.3 ดัชนีบิตแมปแบบกระจาย (Scatter Bitmap Index).....	29
3.3.1 การสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปแบบกระจาย.....	32
3.3.2 ข้อดีของดัชนีบิตแมปแบบกระจาย.....	32
3.3.3 ข้อจำกัดของดัชนีบิตแมปแบบกระจาย.....	33
3.4 ดัชนีบิตแมปแบบเข้ารหัส (Encoded Bitmap Index).....	33
3.4.1 การสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปแบบเข้ารหัส.....	35
3.4.2 ข้อดีของดัชนีบิตแมปแบบเข้ารหัส.....	35
3.4.3 ข้อจำกัดของดัชนีบิตแมปแบบเข้ารหัส.....	35
3.5 เปรียบเทียบดัชนีบิตแมปแบบทั้ง 4 ชนิด.....	35
4. ดัชนีบิตแมปแบบคู่กัน.....	39
4.1 นิยามคำศัพท์และสัญลักษณ์ที่เกี่ยวข้อง.....	39
4.2 การคิดค้นดัชนีบิตแมปแบบคู่กัน.....	40
4.3 ขั้นตอนวิธีการสร้างดัชนีบิตแมปแบบคู่กัน.....	53
4.4 ขั้นตอนวิธีการสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปแบบคู่กัน.....	55
4.5 ข้อดีของดัชนีบิตแมปแบบคู่กัน.....	56
4.6 ข้อจำกัดของดัชนีบิตแมปแบบคู่กัน.....	56

## สารบัญ (ต่อ)

	หน้า
5. การวิเคราะห์และผลการทดลอง.....	57
5.1 ค่าใช้จ่ายเกี่ยวกับพื้นที่ที่ใช้ในการจัดเก็บดัชนี.....	57
5.2 ค่าใช้จ่ายเกี่ยวกับเวลาที่ใช้ในการสอบถามแบบค่าเท่ากัน.....	60
5.3 ประสิทธิภาพในแง่ Space-Time Trade-off.....	64
6. บทสรุปและข้อเสนอแนะ.....	68
6.1 บทสรุป.....	68
6.2 ข้อเสนอแนะและงานในอนาคต.....	71
บรรณานุกรม.....	72
ภาคผนวก.....	75
ก. การเตรียมข้อมูลเพื่อใช้ในการทดลอง.....	75
ข. การเขียนโปรแกรมเพื่อทดลองสอบถามแบบค่าเท่ากันบนดัชนีบิตแมป.....	77
ข.1 ขั้นตอนวิธีทำงานหลักของการสอบถามบนดัชนีบิตแมปทั้ง 5 แบบ.....	78
ข.2 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบพื้นฐาน.....	79
ข.3 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบช่วง.....	80
ข.4 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบกระจาย.....	83
ข.5 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบเข้ารหัส.....	84
ข.6 ขั้นตอนวิธีการสอบถามบนดัชนีบิตแมปแบบคู่กัน.....	86
ประวัติผู้เขียน.....	88

## รายการตาราง

ตาราง	หน้า
1-1 ตารางแสดงระยะเวลาดำเนินงาน	6
2-1 ตารางแสดงการเปรียบเทียบฐานข้อมูลดำเนินการและคลังข้อมูล	10
3-1 ตารางแสดงลักษณะที่สำคัญของดัชนี-bitแมปแบบพื้นฐาน แบบช่วง แบบกระจาย และแบบเข้ารหัส	37
5-1 ตารางแสดงพื้นที่ที่ใช้ในการจัดเก็บดัชนี-bitแมปทั้ง 5 ชนิด	58
5-2 ตารางแสดงจำนวน-bitแมปเวกเตอร์ที่อ่านและจำนวนครั้งในการดำเนินการ ตรรกะระหว่าง-bitแมปเวกเตอร์เมื่อมีการสอบถามแบบค่าเท่ากันของดัชนี -bitแมปทั้ง 5 ชนิด	61
5-3 ตารางแสดงเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันบนดัชนี-bitแมปทั้ง 5 ชนิด	62
6-1 ตารางแสดงลักษณะที่สำคัญของดัชนี-bitแมปทั้ง 5 ชนิด	69
6-2 ตารางสรุปประสิทธิภาพที่ต้องการสำหรับการสอบถามแบบค่าเท่ากัน และดัชนี-bitแมปที่ควรเลือกใช้	70

## รายการภาพประกอบ

ภาพประกอบ	หน้า
2-1 แสดงสถาปัตยกรรมของคลังข้อมูล (Architecture of Data Warehouse )	12
2-2 แสดงข้อมูลหลายมิติ (Multidimensional Data)	15
2-3 แสดงโครงสร้างแบบดาว (Star Schema)	15
2-4 แสดงโครงสร้างแบบผลึกหิมะ (Snowflake Schema)	16
2-5 แสดงการเข้าถึงข้อมูลโดยตรง	17
2-6 แสดงการใช้ฟังก์ชันแฮชในการเทียบค่าไปยังสล็อตบนตารางแฮช	18
2-7 แสดงวิธีการแก้ปัญหาคารชนกันด้วยเซต	19
2-8 แสดงวิธีการแก้ปัญหาคารชนกันด้วยลิเนียร์โพรบิง	20
2-9 แสดงตัวอย่างดัชนีแบบ B-Tree	22
3-1 แสดงแผนภาพการลงรหัสดัชนีบิตแมปแบบพื้นฐาน เมื่อ $C = 15$ ใช้ 15 บิตแมปเวกเตอร์	24
3-2 แสดงดัชนีบิตแมปแบบพื้นฐานบนแตริบิวต์ A เมื่อ $C = 15$ ใช้ 15 บิตแมปเวกเตอร์	25
3-3 แสดงแผนภาพการลงรหัสดัชนีบิตแมปแบบช่วง เมื่อ $C = 15$ ใช้ 8 บิตแมปเวกเตอร์	27
3-4 แสดงดัชนีบิตแมปแบบช่วงบนแตริบิวต์ A เมื่อ $C = 15$ ใช้ 8 บิตแมปเวกเตอร์	28
3-5 แสดงแผนภาพการลงรหัสดัชนีบิตแมปแบบกระจาย เมื่อ $C = 15$ ใช้ 8 บิตแมปเวกเตอร์	30
3-6 แสดงดัชนีบิตแมปแบบกระจายบนแตริบิวต์ A เมื่อ $C = 15$ ใช้ 8 บิตแมปเวกเตอร์	31
3-7 แสดงแผนภาพการลงรหัสดัชนีบิตแมปแบบเข้ารหัส เมื่อ $C = 15$ ใช้ 4 บิตแมปเวกเตอร์	33
3-8 แสดงดัชนีบิตแมปแบบเข้ารหัสบนแตริบิวต์ A เมื่อ $C = 15$ ใช้ 4 บิตแมปเวกเตอร์	34
3-9 แสดงแผนภาพการลงรหัสดัชนีบิตแมปแบบทั้ง 4 ชนิด เมื่อ $C = 15$	36
4-1 แผนภาพแสดงความสัมพันธ์ระหว่างบิตแมปเวกเตอร์กับค่าข้อมูลของ แตริบิวต์ที่มีคาร์ดินอลิตี้ 15 ใช้ 6 บิตแมปเวกเตอร์	42

## รายการภาพประกอบ(ต่อ)

ภาพประกอบ	หน้า
4-2 แผนภาพแสดงการลงรหัสดัชนีบิตแมปแบบคู่กัน เมื่อ $C = 15$ ใช้ 6 บิตแมปเวกเตอร์	43
4-3 แสดงการลงรหัสข้อมูลของแอมป์บิตในการสร้างดัชนีบิตแมปแบบคู่กัน เมื่อ $C = 15$ ใช้ 6 บิตแมปเวกเตอร์	44
4-4 แสดงบิตแมปเวกเตอร์แรกและบิตแมปเวกเตอร์ที่สองที่ถูกลงรหัส '1' บนดัชนีบิตแมปแบบคู่กัน เมื่อ $C = 15$	46
4-5 แสดงค่าของ $r$ ซึ่งเป็นคำตอบของสมการ $\sum_{i=0}^{r-1} i \leq v \leq \left(\sum_{i=0}^r i\right) - 1$	51
4-6 แสดงดัชนีบิตแมปแบบคู่กันบนแอมป์บิต A เมื่อ $C = 15$ ใช้ 6 บิตแมปเวกเตอร์	54
4-7 แสดงการสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปแบบคู่กัน กรณีการสอบถาม " $A = 3$ "	56
5-1 แสดงแผนภาพการลงรหัสดัชนีบิตแมปแบบทั้ง 5 ชนิด เมื่อ คาร์ดินอลิตี้ ( $C$ ) เท่ากับ 15	57
5-2 กราฟแสดงการเปรียบเทียบพื้นที่ที่ใช้ในการสร้างดัชนีบิตแมปทั้ง 5 ชนิด เมื่อแอมป์บิตที่นำมาทำดัชนีมี 1,000,000 เรคอร์ด ( $N = 1,000,000$ )	59
5-3 กราฟแสดงการเปรียบเทียบพื้นที่ที่ใช้ในการสร้างดัชนีบิตแมปทั้ง 3 ชนิด เมื่อแอมป์บิตที่นำมาทำดัชนีมี 1,000,000 เรคอร์ด ( $N = 1,000,000$ )	59
5-4 กราฟแสดงเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันบนดัชนีบิตแมปทั้ง 5 ชนิด เมื่อคาร์ดินอลิตี้มีค่าเท่ากับ 25, 50 และ 150	63
5-5 กราฟแสดงความสัมพันธ์ระหว่างพื้นที่ที่ใช้ในการสร้างดัชนีบิตแมปและเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันของดัชนีบิตแมปทั้ง 5 ชนิด เมื่อคาร์ดินอลิตี้มีค่าเท่ากับ 25	65
5-6 กราฟแสดงความสัมพันธ์ระหว่างพื้นที่ที่ใช้ในการสร้างดัชนีบิตแมปและเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันของดัชนีบิตแมปทั้ง 5 ชนิด เมื่อคาร์ดินอลิตี้มีค่าเท่ากับ 50	65
5-7 กราฟแสดงความสัมพันธ์ระหว่างพื้นที่ที่ใช้ในการสร้างดัชนีบิตแมปและเวลาที่ใช้ในการสอบถามแบบค่าเท่ากันของดัชนีบิตแมปทั้ง 5 ชนิด เมื่อคาร์ดินอลิตี้มีค่าเท่ากับ 150	66