

คลังข้อมูล เพื่อการจัดการข้อมูลสมัยใหม่

DATA WAREHOUSING
FOR MODERN DATA
MANAGEMENT

รัฐสิทธิ์ สุขะหุต

คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

ประวัติและผลงาน



รองศาสตราจารย์ ดร.รัฐสิทธิ์ สุขะหุต

ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

การศึกษา

- Ph.D. in Information Systems, University of East Anglia, UK, พ.ศ. 2544
- Master Degree in Information Systems, Hawaii Pacific University, USA, พ.ศ. 2539
- Bachelor Degree in Computer Science, University of Hawaii at Hilo, USA, พ.ศ. 2538

ประสบการณ์ทำงาน/งานบริหาร

- ผู้อำนวยการสำนักบริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยเชียงใหม่
- อาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่
- กรรมการอำนวยการประจำสำนักบริการเทคโนโลยีสารสนเทศ มหาวิทยาลัยเชียงใหม่
- กรรมการอำนวยการประจำสำนักหอสมุด มหาวิทยาลัยเชียงใหม่
- กรรมการบริหารสถานีวิทยุเสียงสื่อสารมวลชน คณะการสื่อสารมวลชน มหาวิทยาลัยเชียงใหม่
- ที่ปรึกษาด้านด้านเทคโนโลยีสารสนเทศ มูลนิธิโครงการหลวง
- อนุกรรมการเทคโนโลยีสารสนเทศ สถาบันวิจัยพัฒนาพื้นที่สูง องค์การมหาชน
- กรรมการประจำสำนักจัดกิจเพื่อการศึกษา มหาวิทยาลัยราชภัฏเชียงใหม่
- กรรมการประจำสำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏลำปาง
- กรรมการวิพากษ์หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏลำปาง
- กรรมการวิพากษ์หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยแม่โจ้

ความเชี่ยวชาญพิเศษ

- การประมวลผลภาษาธรรมชาติ การสืบค้นข้อมูล
- การออกแบบและการพัฒนาฐานข้อมูล
- คลังข้อมูลและธุรกิจขนาดเล็ก
- การบริหารโครงการพัฒนาซอฟต์แวร์

คลังข้อมูล เพื่อการจัดการข้อมูลสมัยใหม่

DATA WAREHOUSING
FOR MODERN DATA
MANAGEMENT

รัฐสิทธิ์ สุขะหุต
ภาควิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

คลังข้อมูลเพื่อการจัดการข้อมูลสมัยใหม่

บรรณาธิการ: จีรายุทธ ไชยจรรวณิช
ISBN: 978-616-398-310-7
ผู้แต่ง: รัฐสิทธิ์ สุขะหุต
เจ้าของและผู้จัดพิมพ์: ศูนย์บริหารงานวิจัย สำนักงานมหาวิทยาลัยเชียงใหม่
โทรศัพท์: 0 5394 3603-4
โทรสาร: 0 5394 3600
<http://cmupress.cmu.ac.th>
E-mail: cmupress.th@gmail.com

พิมพ์ครั้งที่ 1: กุมภาพันธ์ 2560
พิมพ์ครั้งที่ 2: มิถุนายน 2561
ราคา: 450 บาท

ข้อมูลทางบรรณานุกรม

รัฐสิทธิ์ สุขะหุต.

คลังข้อมูลเพื่อการจัดการข้อมูลสมัยใหม่ -- พิมพ์ครั้งที่ 2 -- เชียงใหม่
: ศูนย์บริหารงานวิจัย มหาวิทยาลัยเชียงใหม่, 2561.
278 หน้า.

1. การจัดการฐานข้อมูล. I. ชื่อเรื่อง.

005.74

ISBN 978-616-398-310-7

ออกแบบและพิมพ์: ดองสาม ดีไซน์

©สงวนลิขสิทธิ์ พ.ศ. 2537 ห้ามลอกเลียนแบบ ทำซ้ำ ตัดแปลง ส่วนหนึ่งส่วนใดของหนังสือเล่มนี้ รวมทั้งการจัดเก็บถ่ายทอโดยไม่ว่ารูปแบบหรือวิธีการใดๆ ด้วยกระบวนการทางอิเล็กทรอนิกส์ การถ่ายภาพ การบันทึกหรือวิธีการอื่นใดโดยไม่ได้รับอนุญาต

กรณีต้องการสั่งซื้อเป็นจำนวนมาก กรุณาติดต่อ สำนักงานมหาวิทยาลัยเชียงใหม่
โทรศัพท์: 0 5394 3605 โทรสาร: 0 5394 3600
<http://cmupress.cmu.ac.th>, E-mail: cmupress.th@gmail.com

พัฒนาการของเทคโนโลยีสารสนเทศตั้งแต่อดีตจนถึงปัจจุบันได้ส่งผลเอื้ออำนวยให้การขับเคลื่อนทางเศรษฐกิจและรูปแบบสังคมโลกเปลี่ยนแปลงไปอย่างรวดเร็ว ข้อมูลธุรกรรมและกิจกรรมต่างๆ ของสังคมโลกได้รับการบันทึก จัดเก็บ เชื่อมโยง แลกเปลี่ยนและใช้ประโยชน์ร่วมกันในวงกว้าง ปริมาณข้อมูลดังกล่าวนี้วันวันจะมีขนาดเพิ่มมากขึ้นเรื่อยๆ ความต้องการทางเศรษฐกิจและสังคมที่จะใช้ประโยชน์จากข้อมูลดังกล่าวก็มีมากขึ้น และซับซ้อนขึ้นตามลำดับ ประเด็นความท้าทายทางเทคโนโลยีนี้ทำให้เกิดศาสตร์ใหม่ในวงการวิชาการที่เป็นที่รับทราบโดยทั่วไปในนาม Data Science

หนังสือเรื่อง คลังข้อมูลเพื่อการจัดการข้อมูลสมัยใหม่ หรือ Data Warehouse for Modern Data Management ซึ่งเรียบเรียงโดย ผศ.ดร.รัฐสิทธิ์ สุชะหุด อาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ถือเป็นส่วนหนึ่งในพัฒนาการของเทคโนโลยีสารสนเทศของประเทศไทย ผู้เขียนได้แสดงให้เห็นเทคโนโลยีการจัดการข้อมูลสมัยใหม่ในภาพกว้าง ได้ให้ความรู้ความเข้าใจพื้นฐานและอธิบายรายละเอียดเทคนิคเชิงลึกด้านการจัดการคลังข้อมูลได้อย่างถี่ถ้วน หนังสือเล่มนี้จะเป็นประโยชน์และเหมาะแก่การเรียนรู้ของนักศึกษา คณาจารย์ และบุคคลทั่วไปที่อยู่ในวงการวิชาชีพเทคโนโลยีสารสนเทศเป็นอย่างยิ่ง

รองศาสตราจารย์ ดร.จิรยุทธ ไชยจารุณิช

คำนำ (พิมพ์ครั้งที่ 1)

ปัจจุบันในยุคสมัยของโลกสังคมดิจิทัลและสื่อออนไลน์ที่มีการพึ่งพาเทคโนโลยีสารสนเทศมากยิ่งขึ้น วิทยาการข้อมูลได้เข้ามามีบทบาทสำคัญองค์กรต่างๆ ทั้งหน่วยงานในภาครัฐและเอกชนที่มีการจัดเก็บข้อมูลในรูปแบบของสื่ออิเล็กทรอนิกส์ต่างให้ความสำคัญต่อการจัดเก็บข้อมูลสำหรับนำข้อมูลมาเพื่อประกอบการวางแผน วิเคราะห์ และเป็นข้อมูลสนับสนุนการตัดสินใจของผู้บริหาร แนวคิดของการจัดการข้อมูลสมัยใหม่เป็นศาสตร์ที่ว่าด้วยการบริหารจัดการข้อมูลอย่างเป็นระบบ เป็นการผสมผสานระหว่างวิทยาการคอมพิวเตอร์ หลักสถิติ คณิตศาสตร์ และซอฟต์แวร์ และต้องอาศัยความรู้ทางด้านการจัดการแบบบูรณาการ ซึ่งจำเป็นต้องให้ความสำคัญต่อการออกแบบโครงสร้างและสถาปัตยกรรมข้อมูลที่ดี เพื่อให้การใช้ข้อมูลสารสนเทศเกิดประโยชน์สูงสุดต่อองค์กร

หนังสือคลังข้อมูลเพื่อการจัดการข้อมูลสมัยใหม่เล่มนี้ได้เรียบเรียงจากเอกสาร ตำรา และการค้นคว้าความรู้เพิ่มเติมจากแหล่งข้อมูลต่างๆ รวมถึงการสอดแทรกความรู้และประสบการณ์ของผู้เขียน เพื่อให้เนื้อหามีความทันสมัยโดยถ่ายทอดเป็นหนังสือเล่มนี้ ซึ่งมุ่งเน้นการนำเสนอแนวคิดของการพัฒนาคลังข้อมูล เหตุผลและความจำเป็นของการพัฒนาคลังข้อมูลเพื่อมุ่งเน้นการพัฒนาข้อมูลไปสู่ข้อมูลสารสนเทศสำหรับผู้บริหาร และได้อธิบายถึงขั้นตอนการออกแบบคลังข้อมูลโดยอาศัยไดเมนชันโมเดล ตลอดจนแนวทางการประยุกต์ใช้ข้อมูลที่จัดเก็บภายในคลังข้อมูล การประมวลผลข้อมูลเชิงมิติโดยอาศัยเครื่องมือโอแลป นอกจากนี้ยังได้อธิบายถึงภาษาเอ็มดีเอ็็กซ์ที่เป็นโปรแกรมภาษาที่ใช้ในการค้นคืนข้อมูลเชิงมิติเพื่อนำไปสู่การใช้ประโยชน์จากข้อมูลเชิงสารสนเทศโดยใช้เครื่องมือธุรกิจอัจฉริยะ และท้ายสุดได้อธิบายถึงการบำรุงดูแลรักษาระบบคลังข้อมูล การบริหารความเสี่ยงและความต่อเนื่องทางธุรกิจ ซึ่งต้องคำนึงถึงความมั่นคงปลอดภัยของข้อมูล เนื้อหาของหนังสือเล่มนี้ได้มีการเรียบเรียงและนำเสนออย่างละเอียดทุกขั้นตอนพร้อมภาพประกอบเนื้อหาแบ่งเป็นบทต่างๆ สุดท้ายนี้ผู้เขียนหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์สำหรับอาจารย์ นักศึกษาสาขาวิชาวิทยาการคอมพิวเตอร์ วิศวกรรมคอมพิวเตอร์ และสาขาเทคโนโลยีสารสนเทศ รวมถึงผู้ที่สนใจในเนื้อหาที่เกี่ยวกับการคลังข้อมูลและเทคโนโลยีฐานข้อมูลต่อไป

ผู้ช่วยศาสตราจารย์ ดร.รัฐสิทธิ์ สุชะหุด

คำนำ (พิมพ์ครั้งที่ 2)

ในโลกของเทคโนโลยีข่าวสารที่ข้อมูลมีบทบาทสำคัญต่อการดำเนินธุรกิจ ทำให้หลายองค์กรปัจจุบันเริ่มหันมาให้ความสนใจต่อการจัดเก็บข้อมูลของตนเองมากขึ้น การให้ความสำคัญต่อการบริหารจัดการข้อมูลเหล่านี้จึงเป็นสิ่งจำเป็น ข้อมูลที่เก็บสะสมมาตั้งแต่อดีตกลายมาเป็นข้อมูลขนาดใหญ่ที่มีปริมาณมหาศาล ข้อมูลเหล่านี้กลายมาเป็นสินทรัพย์ที่สำคัญขององค์กร หากนำมาใช้ให้เกิดประโยชน์ ข้อมูลสารสนเทศที่ผ่านกระบวนการสังเคราะห์ทำให้ผู้บริหารขององค์กรเข้าใจสถานการณ์และผลประกอบการขององค์กรตนเอง สามารถนำเอาข้อมูลมาใช้ในการวางแผน การวิเคราะห์ และตัดสินใจ รวมถึงการใช้ประโยชน์ข้อมูลเพื่อสร้างความได้เปรียบต่อคู่แข่ง

คลังข้อมูลเป็นแนวคิดของการบริหารจัดการข้อมูลขนาดใหญ่เพื่อการนำข้อมูลไปใช้ประโยชน์เพื่อการวางแผนและสนับสนุนการตัดสินใจสร้างความได้เปรียบเชิงกลยุทธ์ในการแข่งขันในโลกธุรกิจ คลังข้อมูลเป็นการออกแบบโครงสร้างข้อมูลที่รองรับข้อมูลที่มีปริมาณมาก เป็นข้อมูลที่มีการเชื่อมโยงจากข้อมูลหลายแหล่ง สามารถเพิ่มประสิทธิภาพในการวิเคราะห์ การวิเคราะห์ คลังข้อมูลเป็นการจัดเก็บข้อมูลย้อนหลังในอดีต ต่างจากฐานข้อมูลระดับปฏิบัติการที่ข้อมูลมีความเป็นปัจจุบัน เนื้อหาของหนังสือเล่มนี้ได้อธิบายภาพรวมของความสำคัญของการบริหารจัดการข้อมูลขนาดใหญ่ ประกอบด้วยขั้นตอนการวิเคราะห์ ออกแบบ และพัฒนาไปสู่โมเดลการสร้างมุมมองข้อมูลเชิงมิติ ตลอดจนการนำเอาข้อมูลที่จัดเก็บภายในคลังข้อมูลมาใช้ประโยชน์ ทั้งการคิวรีข้อมูลขนาดใหญ่โดยใช้ภาษาเอ็มดีเอ็กซ์ และการพัฒนาไปสู่ข้อมูลสารสนเทศสำหรับผู้บริหารโดยใช้เครื่องมือธุรกิจจากฐานข้อมูล รวมถึงการบำรุงดูแลรักษาคลังข้อมูลและการบริหารความเสี่ยง

หนังสือคลังข้อมูลเพื่อการจัดการข้อมูลสมัยใหม่เล่มนี้ได้เรียบเรียงจากเอกสาร ตำรา และการศึกษาค้นคว้าเนื้อหาเพิ่มเติมจากแหล่งข้อมูลต่างๆ รวมถึงการสอดแทรกความรู้จากประสบการณ์จากผู้เขียนเอง เพื่อให้เนื้อหาที่ถ่ายทอดมีความทันสมัยและสอดคล้องกับสถานการณ์โลกปัจจุบัน หนังสือเล่มนี้เหมาะสำหรับอาจารย์ นักวิจัย และนักศึกษาที่เกี่ยวข้องกับสาขาเทคโนโลยีสารสนเทศ วิทยาการคอมพิวเตอร์ วิศวกรรมคอมพิวเตอร์ และคอมพิวเตอร์ธุรกิจ ซึ่งผู้เขียนหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์แก่ผู้ที่สนใจในการบริหารจัดการข้อมูลขนาดใหญ่เพื่อพัฒนาเป็นคลังข้อมูลต่อไป

รองศาสตราจารย์ ดร.รัฐสิทธิ์ สุขะหุด

สารบัญ

คำนิยม	I
คำนำ (พิมพ์ครั้งที่ 1)	II
คำนำ (พิมพ์ครั้งที่ 2)	III
สารบัญ	IV
สารบัญภาพ	XIV
สารบัญตาราง	XVIII
บทที่ 1 แนวคิดของการจัดการข้อมูลสมัยใหม่	1
1.1 วัตถุประสงค์	2
1.2 บทนำ	2
1.3 พัฒนาการข้อมูลในปัจจุบัน	4
1.4 แนวคิดของการจัดการข้อมูลสมัยใหม่	6
1.5 ข้อมูลและข้อมูลสารสนเทศ	9
ประเภทข้อมูล	10
สถาปัตยกรรมข้อมูลสารสนเทศ	11
มุมมองผู้บริหารต่อข้อมูลสารสนเทศ	13
1.6 คลังข้อมูลและข้อมูลสารสนเทศ	14
ฐานข้อมูลปฏิบัติการ	15
การประมวลผลข้อมูลระดับทรานแซคชันแบบออนไลน์	16
ระบบฐานข้อมูลเชิงสัมพันธ์	18
การประมวลผลออนไลน์เชิงวิเคราะห์	20
1.7 การเตรียมความพร้อมสำหรับโครงการคลังข้อมูล	23
แนวคิดการเชื่อมโยงข้อมูลธุรกิจ	23
การวางแผนการดำเนินงาน	24
การสื่อสารกับคนในองค์กร	24
ภาวะข้อมูลท่วมท้น	25
ข้อมูลสกปรก (dirty data)	25

1.8	สรุป	27
1.9	คำถามท้ายบท	28
บทที่ 2	คุณลักษณะสำคัญของคลังข้อมูล	29
2.1	วัตถุประสงค์	30
2.2	บทนำ	30
2.3	คลังข้อมูล (data warehouse)	31
	ประวัติคลังข้อมูล	32
	แนวคิดของการพัฒนาคลังข้อมูล	33
2.4	โครงสร้างพื้นฐานของระบบเครือข่ายและเครื่องแม่ข่าย	34
2.5	คุณสมบัติสำคัญของคลังข้อมูล	35
	Subject-oriented data	35
	Integrated data	36
	Time-variant data	37
	Non-volatile data	37
2.6	สถาปัตยกรรมคลังข้อมูล	38
2.7	ความละเอียดของข้อมูล (data granularity)	39
2.8	ความแตกต่างระหว่างคลังข้อมูลและดาต้ามาร์ท	42
2.9	แนวทางการพัฒนาคลังข้อมูลสำหรับองค์กร	44
	แบบบนลงล่าง Top-down approach	44
	แบบล่างขึ้นบน Bottom-up approach	45
	องค์ประกอบแหล่งข้อมูล	45
	องค์ประกอบพื้นที่พักข้อมูล	47
	องค์ประกอบพื้นที่จัดเก็บข้อมูล	49
2.10	การใช้ประโยชน์จากคลังข้อมูลในเชิงสารสนเทศ	50

สารบัญ

2.11 เมตาตาต้า	52
เมตาตาต้าในระดับปฏิบัติการ	52
เมตาตาต้าสำหรับการสกัดและการแปลงข้อมูล	52
เมตาตาต้าในระดับผู้ใช้งานข้อมูล	52
2.12 สรุป	53
2.13 คำถามท้ายบท	53
บทที่ 3 การวิเคราะห์ความต้องการทางธุรกิจ	55
3.1 วัตถุประสงค์	56
3.2 บทนำ	56
3.3 มุมมองข้อมูลเชิงมิติ	57
ธรรมชาติมิติข้อมูลทางธุรกิจ	58
การวิเคราะห์มิติข้อมูลในเชิงธุรกิจ	61
มิติข้อมูลแบบลำดับขั้น	64
3.4 การกำหนดเป้าหมายของการพัฒนาลังข้อมูล	65
3.5 ความน่าเชื่อถือของข้อมูล	67
3.6 การจัดเก็บความต้องการของผู้ใช้งาน	68
3.7 การประมวลผลความต้องการเพื่อนำไปสู่การออกแบบคลังข้อมูล	71
3.8 สรุป	73
3.9 คำถามท้ายบท	74
บทที่ 4 แบบจำลองข้อมูลหลายมิติ	75
4.1 วัตถุประสงค์	76
4.2 บทนำ	76
4.3 แบบจำลองข้อมูลเชิงมิติ	76

4.4 การออกแบบแบบจำลองข้อมูลเชิงมิติ	78
การวิเคราะห์แบบจำลองเชิงมิติ	78
การสอบถามข้อมูลเชิงมิติ	80
4.5 โครงสร้างสกีมาของโมเดลเชิงมิติ	82
โครงสร้างแบบดาว	82
โครงสร้างแบบเกล็ดหิมะ	85
โครงสร้างแบบกาแล็กซี	86
4.6 แนวคิดการออกแบบเชิงการวิเคราะห์เชิงมิติ	88
4.7 คุณสมบัติของไทม์ไลน์	90
4.8 คุณสมบัติของแฟก	91
ตารางแฟกที่ไม่มีค่าคำนวณ	92
ความละเอียดของข้อมูล	92
การกำหนดคีย์ตัวแทน	93
4.9 มิติเวลา	94
4.10 เมตาตาต้า	97
4.11 มุมมองข้อมูลเชิงสรุป	98
4.12 การออกแบบคลังข้อมูลตามแนวคิดของ Ralph Kimball	98
4.13 ประสิทธิภาพการทำงานของคลังข้อมูล	103
การสร้างดัชนีข้อมูล	104
การแบ่งพาดิชนของฐานข้อมูล	104
การสร้างข้อมูลแบบจัดกลุ่ม	105
การสร้างวิวแบบสรุป	105
4.14 สรุป	106
4.15 คำถามท้ายบท	106

สารบัญ

บทที่ 5	กระบวนการสกัด-แปลงและการโหลดข้อมูล	109
5.1	วัตถุประสงค์	110
5.2	บทนำ	110
5.3	กระบวนการสกัด แปลง และโหลดข้อมูล	111
5.4	ความท้าทายของกระบวนการสกัดและการแปลงข้อมูล	112
5.5	การสกัดข้อมูล	113
	ขั้นตอนการสกัดข้อมูล	113
	ขั้นตอนสำคัญในการสกัดข้อมูล	114
	เทคนิควิธีการการสกัดข้อมูล	115
	ความถี่ในการสกัดข้อมูล	116
5.6	การแปลงข้อมูล	117
	คุณภาพของข้อมูล	117
	ความถูกต้องในค่าต่างๆ ที่เป็นไปได้ของข้อมูล	118
	ความถูกต้องจากความสัมพันธ์ของข้อมูล	120
	การทำความสะอาดข้อมูล	121
	ข้อมูลที่มีความผิดพลาด	123
5.7	การโหลดข้อมูล	124
	การปรับปรุงคลังข้อมูลให้เป็นปัจจุบัน	125
	การโหลดข้อมูลตามช่วงเวลา	125
5.8	วงจรการทำงานของ ETL	127
5.9	สรุป	129
5.10	คำถามท้ายบท	129
บทที่ 6	พื้นที่พักข้อมูล	131
6.1	วัตถุประสงค์	132
6.2	บทนำ	132

6.3	สถาปัตยกรรมของพื้นที่พักข้อมูล	132
6.4	การทำงานของพื้นที่พักข้อมูล	136
6.5	การทำหน้าที่เป็นที่สำรองข้อมูลชั่วคราว	136
6.6	กิจกรรมที่เกิดขึ้นภายในพื้นที่พักข้อมูล	137
	การเชื่อมโยงข้อมูล	138
	การปรับข้อมูล	138
	การลดความขัดแย้งข้อมูล	138
	การตรวจสอบช่วงเวลาของการประมวลผลข้อมูล	139
	การตรวจสอบการเปลี่ยนแปลงข้อมูล	139
	การทำความสะอาดข้อมูล	139
	การประมวลผลข้อมูลสรุปเบื้องต้น	140
	การจัดเก็บข้อมูลที่เก่าในอดีต	140
6.7	สรุป	140
6.8	คำถามท้ายบท	141
บทที่ 7	การประมวลผลออนไลน์เชิงวิเคราะห์	143
7.1	วัตถุประสงค์	144
7.2	บทนำ	144
7.3	แนวคิดการประมวลผลออนไลน์เชิงวิเคราะห์	144
7.4	การประมวลผลโอแลป	146
	การ Roll-up	148
	การ Drill-down	149
	การ Slice and dice	150
	การ Pivot	151
7.5	สรุป	152
7.6	คำถามท้ายบท	153

สารบัญ

บทที่ 8	การออกแบบทางกายภาพ	155
8.1	วัตถุประสงค์	156
8.2	บทนำ	156
8.3	การออกแบบทางกายภาพ	156
	มาตรฐานการพัฒนา	157
	จัดทำแผนโครงการ	157
	การกำหนดพาร์ติชันโครงสร้างข้อมูล	158
	การกำหนดตัวเลือกการจัดกลุ่ม	158
	การจัดทำดัชนี	159
8.4	องค์ประกอบที่ต้องพิจารณาในการออกแบบ	159
	ส่วนประกอบของแบบจำลองทางกายภาพ	161
	การจัดเก็บข้อมูลทางกายภาพ	164
	พื้นที่จัดเก็บโครงสร้างข้อมูล	165
	การจัดทำดัชนีคลังข้อมูล	165
8.5	สรุป	170
8.6	คำถามท้ายบท	170
 บทที่ 9	 ภาษาเอ็มดีเอ็กซ์	 171
9.1	วัตถุประสงค์	172
9.2	บทนำ	172
9.3	ภาษาเอ็มดีเอ็กซ์	172
	ฟังก์ชัน MEMBERS	175
	ฟังก์ชัน CHILDRENS	176
	ฟังก์ชัน DESCENDANTS	177
	ฟังก์ชัน ADDCALCULATEDMEMBERS	177
9.4	การสร้างคำสั่งแบบ Complex query	178

9.5	การกำหนดเงื่อนไขสำหรับการคิวรี	178
	การสร้างค่าคำนวณใหม่	179
	ฟังก์ชัน CURRENTMEMBER และ PARENT	181
	ฟังก์ชัน FIRSTCHILD, LASTCHILD และ CURRENTMEMBER	181
9.6	การท่องตามโครงสร้างแบบลำดับชั้น	182
	ฟังก์ชัน ANCESTER	182
	ฟังก์ชัน EXCEPT	182
	ฟังก์ชัน GENERATE	183
	ฟังก์ชัน LEAD	184
9.7	ฟังก์ชันที่ทำงานกับช่วงเวลา	184
9.8	การกรองข้อมูล (filter) และการเรียงลำดับ (sort)	186
	ฟังก์ชัน FILTER	187
	ฟังก์ชัน ORDER	187
9.9	การวิเคราะห์ผลจากการจัดอันดับสูงสุดและต่ำสุด	188
9.10	ฟังก์ชันสำหรับการคำนวณและเงื่อนไข	189
9.11	สรุป	191
9.12	คำถามท้ายบท	192
บทที่ 10	เหมืองข้อมูล	193
10.1	วัตถุประสงค์	194
10.2	บทนำ	194
10.3	แนวคิดของการทำเหมืองข้อมูล	194
10.4	เหมืองข้อมูลคืออะไร	195
10.5	หลักการทำเหมืองข้อมูล	197
10.6	การเตรียมความพร้อมก่อนการทำเหมืองข้อมูล	198
10.7	การประมวลผลข้อมูลเชิงมิติ OLAP กับเหมืองข้อมูล	201

สารบัญ

10.8	เหมืองข้อมูลและคลังข้อมูล	204
10.9	เทคนิคการทำเหมืองข้อมูล	205
	เทคนิคการจัดกลุ่ม	206
	เทคนิคต้นไม้ตัดสินใจ	209
	การหาเหตุผลจากหน่วยความจำ	213
	การวิเคราะห์การเชื่อมโยง	216
	กฎความสัมพันธ์	217
10.10	ประโยชน์ที่ได้รับจากการทำเหมืองข้อมูล	222
10.11	สรุป	225
10.12	คำถามท้ายบท	225
บทที่ 11	ธุรกิจอัจฉริยะ	227
11.1	วัตถุประสงค์	228
11.2	บทนำ	228
11.3	แนวคิดธุรกิจอัจฉริยะ	228
11.4	องค์ประกอบที่สำคัญของเครื่องมือธุรกิจอัจฉริยะ	229
11.5	การวิเคราะห์ข้อมูลแบบมุมมองหลายมิติ	230
11.6	รายงานในเครื่องมือธุรกิจอัจฉริยะ	233
11.7	สรุป	238
11.8	คำถามท้ายบท	238
บทที่ 12	การบำรุงดูแลรักษาค้างข้อมูล	239
12.1	วัตถุประสงค์	240
12.2	บทนำ	240
12.3	แนวคิดของการบำรุงรักษาระบบคลังข้อมูล	240
	บทบาทและความรับผิดชอบต่อการพัฒนาค้างข้อมูล	241
	การรักษาความปลอดภัยข้อมูล	243

สารบัญ

12.4	ทรัพยากรฮาร์ดแวร์ที่เกี่ยวข้องกับคลังข้อมูล	244
12.5	เทคโนโลยีเวอร์ชวลไลเซชัน	244
12.6	การสำรองและการกู้คืนข้อมูล	247
12.7	การล้างข้อมูล	248
12.8	การบริหารจัดการความเสี่ยง	249
12.9	สรุป	251
12.10	คำถามท้ายบท	251
บรรณานุกรม		253
ดัชนี		254

ตัวอย่าง

สารบัญภาพ

รูปที่ 1.1	พัฒนาการของเทคโนโลยีเว็บ	3
รูปที่ 1.2	ตัวอย่างเอเจนท์ที่ใช้ในแต่ละอุตสาหกรรม	3
รูปที่ 1.3	เครือข่ายการสื่อสารที่เชื่อมหากันบนโลกออนไลน์	4
รูปที่ 1.4	สถาปัตยกรรมข้อมูลสารสนเทศ	12
รูปที่ 1.5	โครงสร้างระบบ SAP อีอาร์พี	18
รูปที่ 1.6	โครงสร้างฐานข้อมูลร้านเช่าวิดีโอ ในระดับ conceptual design	20
รูปที่ 2.1	ความแตกต่างระหว่างฐานข้อมูลปฏิบัติการคลังข้อมูล	33
รูปที่ 2.2	การรวบรวมข้อมูลจากแหล่งข้อมูลต่างๆ เป็นคลังข้อมูล	36
รูปที่ 2.3	สถาปัตยกรรมโดยรวมของระบบคลังข้อมูล	38
รูปที่ 2.4	ระดับความละเอียดข้อมูลที่จัดเก็บภายในคลังข้อมูล และทิศทางการไหลของข้อมูล	40
รูปที่ 2.5	โครงสร้างสถาปัตยกรรมของมิติเอเตอร์	48
รูปที่ 2.6	การไหลของข้อมูลสู่คลังข้อมูลตามช่วงเวลา	50
รูปที่ 3.1	แสดงภาพโครงสร้างองค์กรในฝ่ายการตลาด โดยแบ่งออกเป็นลำดับชั้น	59
รูปที่ 3.2	ไฮเปอร์คิวบ์แสดงมิติของข้อมูลทางธุรกิจ	62
รูปที่ 3.3	มิติของข้อมูลในแต่ละประเภทของธุรกิจ	63
รูปที่ 3.4	ตัวอย่าง Information Packages สำหรับข้อมูลการขาย	72
รูปที่ 4.1	ตัวอย่างแบบจำลองข้อมูลเชิงมิติ (dimension model)	77
รูปที่ 4.2	แสดงขั้นตอนจากความต้องการสู่การออกแบบโมเดล ข้อมูลเชิงมิติ	78
รูปที่ 4.3	แบบจำลองข้อมูลสำหรับ Automaker sales	79
รูปที่ 4.4	โมเดลข้อมูลที่มีโครงสร้างแบบดาว (star schema)	83
รูปที่ 4.5	โมเดลข้อมูลที่มีโครงสร้างแบบเกล็ดหิมะ (snowflake schema)	85
รูปที่ 4.6	โมเดลข้อมูลที่มีโครงสร้างแบบกาแล็กซี (galaxy schema)	87

สารบัญภาพ

รูปที่ 4.7	การเชื่อมโยงข้อมูลระหว่างแฟก Enrollment และ Time table ในโครงสร้างกาเล็กซี่	87
รูปที่ 4.8	โครงสร้างข้อมูลเชิงมิติสำหรับห้างสรรพสินค้า	89
รูปที่ 4.9	ตัวอย่างตารางแฟกที่ไม่มีค่าคำนวณ	92
รูปที่ 4.10	การสร้างคีย์ตัวแทน CustomerKey สำหรับข้อมูลลูกค้า	94
รูปที่ 4.11	ตัวอย่างตารางไทม์เม้นชันสำหรับมิติเวลา (time dimension)	95
รูปที่ 4.12	แสดงความสัมพันธ์ระหว่างตารางไทม์เม้นชันและตารางแฟก	100
รูปที่ 5.1	การออกแบบขั้นตอนการทำ ETL	111
รูปที่ 5.2	ความสัมพันธ์ข้อมูลสายการบินที่สัมพันธ์กับเวลา	116
รูปที่ 5.3	การออกแบบขั้นตอนการทำ ETL	125
รูปที่ 5.4	แสดงวงจรการทำงานของ ETL	128
รูปที่ 6.1	สถาปัตยกรรมของพื้นที่พักข้อมูล	133
รูปที่ 6.2	การเตรียมการก่อนการนำเข้าข้อมูล	137
รูปที่ 7.1	คิวบ์แสดงมุมมองการวิเคราะห์ข้อมูลแบบมิติไทม์เม้นชัน	145
รูปที่ 7.2	คิวบ์แสดงมุมมองการวิเคราะห์ข้อมูลแบบมิติไทม์เม้นชัน	146
รูปที่ 7.3	การสร้างมุมมองข้อมูลในลักษณะของการ Roll-up	148
รูปที่ 7.4	การสร้างมุมมองข้อมูลในลักษณะของการ Drill-down	149
รูปที่ 7.5	การสร้างมุมมองข้อมูลในลักษณะของการ Slice	150
รูปที่ 7.6	การสร้างมุมมองข้อมูลในลักษณะของการ Dice	151
รูปที่ 7.7	การสร้างมุมมองข้อมูลในลักษณะของการ Pivot	152
รูปที่ 8.1	ขั้นตอนการออกแบบทางกายภาพ	157
รูปที่ 8.2	กระบวนการแปลงแบบจำลองทางตรรกะเป็นแบบจำลองทางกายภาพ	161
รูปที่ 8.3	องค์ประกอบทางกายภาพของคลังข้อมูล	161
รูปที่ 8.4	ตัวอย่างนิยามโครงสร้างใน SQL	162
รูปที่ 8.5	แบบจำลองทางตรรกะและแบบจำลองทางกายภาพ	163

สารบัญภาพ

รูปที่ 8.6	โครงสร้างข้อมูลในคลังข้อมูล	165
รูปที่ 8.7	ตัวอย่างดัชนี B-Tree	166
รูปที่ 8.8	ตัวอย่างดัชนีบีตแมป	168
รูปที่ 8.9	การสืบค้นข้อมูลจากดัชนีบีตแมป	168
รูปที่ 9.1	ไคเมนชันโมเดลสำหรับ Sales	174
รูปที่ 10.1	สนับสนุนการตัดสินใจดำเนินการทำเหมืองข้อมูล	196
รูปที่ 10.2	การค้นหาข้อมูลที่มีคุณค่า	198
รูปที่ 10.3	ขั้นตอนการค้นหาความรู้	199
รูปที่ 10.4	กระบวนการค้นหาความรู้	201
รูปที่ 10.5	ความแตกต่างระหว่าง OLAP และ Data mining	202
รูปที่ 10.6	การทำเหมืองข้อมูลในสภาพแวดล้อมของคลังข้อมูล	205
รูปที่ 10.7	การกระจายตัวของข้อมูลลูกค้าตามอายุและรายได้เฉลี่ยของลูกค้าทั้งหมด 100 คน	208
รูปที่ 10.8	ผลลัพธ์จากการแบ่งกลุ่มลูกค้าทั้งหมด 100 คนออกเป็น 3 กลุ่ม โดยพิจารณาจากข้อมูลอายุและรายได้เฉลี่ย	208
รูปที่ 10.9	ต้นไม้การตัดสินใจสำหรับผู้ซื้อเครื่องคอมพิวเตอร์โน้ตบุ๊ก	209
รูปที่ 10.10	ต้นไม้การตัดสินใจสำหรับการออกไปเล่นกีฬากลางแจ้งโดยพิจารณาจากสภาพอากาศ	213
รูปที่ 10.11	การกระจายของข้อมูลดอกไอริส 3 ชนิด จำนวน 15 ดอก	214
รูปที่ 10.12	การกระจายของข้อมูลดอกไอริส 3 ชนิด จำนวน 15 ดอก และข้อมูลดอกไอริสที่ต้องการทำนายชนิดพร้อมเพื่อนบ้านที่ใกล้ที่สุด 3 ตัว	216
รูปที่ 11.1	โครงสร้างตารางข้อมูลแบบ OLTP	231
รูปที่ 11.2	แบบจำลองโครงสร้างข้อมูลเชิงมิติสำหรับข้อมูลการขายสินค้า	232
รูปที่ 11.3	ตัวอย่างเครื่องมือ BI ที่แสดงผลรายงานแบบตาราง	234
รูปที่ 11.4	ตัวอย่างเครื่องมือ BI ที่แสดงผลรายงานแบบแผนภูมิ	234

สารบัญภาพ

รูปที่ 11.5	ตัวอย่างเครื่องมือ BI ที่แสดงผลรายงานแบบตัวชี้วัด	235
รูปที่ 11.6	การแสดงผล Dashboard ภาพรวมสารสนเทศ	236
รูปที่ 11.7	ตัวอย่างรายงานของ Prelytis ที่สนับสนุนการทำงานแบบ Multi-platform	237
รูปที่ 12.1	กิจกรรมที่เกี่ยวข้องกับการดูแลรักษาคลังข้อมูล	243
รูปที่ 12.2	สถาปัตยกรรมของเครื่องแม่ข่ายเสมือนที่ใช้อุปกรณ์การจัดเก็บร่วมกัน	245
รูปที่ 12.3	การสำรองข้อมูลแบบ Incremental backup	248

ตัวอย่าง

สารบัญตาราง

ตารางที่ 1.1	ความแตกต่างระหว่างข้อมูลปฏิบัติและคลังข้อมูล	22
ตารางที่ 9.1	รายละเอียดของตารางไคเมนชันสำหรับคิวบ์ Sales	173
ตารางที่ 9.2	รายละเอียดของตารางแฟก Sales	174
ตารางที่ 10.1	OLAP จะใช้ในการวิเคราะห์ข้อมูลที่ผ่านมา และเหมืองข้อมูลสำหรับการทำนายอนาคต	203
ตารางที่ 10.2	ความแตกต่างของ OLAP และ Data mining	203
ตารางที่ 10.3	การประยุกต์ใช้เหมืองข้อมูลเพื่อการวิเคราะห์ ในแต่ละลักษณะงาน	206
ตารางที่ 10.4	ข้อมูลสภาพอากาศ	211
ตารางที่ 10.5	แสดงระยะห่างระหว่างข้อมูลดอกไอริส	215
ตารางที่ 10.6	แสดงสินค้าที่ลูกค้าจำนวน 10 คน ซื้อจากร้านค้าแห่งหนึ่ง	218
ตารางที่ 10.7	แสดงเซตรายการที่ประกอบด้วยแอทริบิวต์ 2 แอทริบิวต์และค่าสนับสนุนของแต่ละเซตรายการ	219
ตารางที่ 10.8	กฎความสัมพันธ์ที่สร้างจากเซตรายการที่ประกอบด้วย 2 แอทริบิวต์และค่าความเชื่อมั่น	222
ตารางที่ 10.9	แนวทางการประยุกต์เทคนิคเหมืองข้อมูล	223

01

บทที่ 1

ตัวอย่าง

แนวคิดของการจัดการข้อมูล
สมัยใหม่

1.1 วัตถุประสงค์

- เพื่ออธิบายถึงแนวคิดของการบริหารจัดการข้อมูลสมัยใหม่
- เพื่ออธิบายโครงสร้างของข้อมูลเพื่อนำไปสู่ข้อมูลสารสนเทศ
- เพื่อให้เห็นความแตกต่างระหว่างฐานข้อมูลปฏิบัติการและคลังข้อมูล
- เพื่ออธิบายถึงแนวทางการใช้ประโยชน์ข้อมูลในเชิงการวิเคราะห์และการตัดสินใจสำหรับผู้บริหาร

1.2 บทนำ

ปัจจุบันเทคโนโลยีสารสนเทศมีบทบาทสำคัญเป็นอย่างมากในองค์กรธุรกิจ ทั้งภาครัฐและเอกชน เทคโนโลยีถูกนำมาใช้เป็นเครื่องมือในการเพิ่มประสิทธิภาพการทำงาน การบริหาร เพื่อสนับสนุนให้องค์กรบรรลุเป้าหมายและพันธกิจขององค์กร อีกทั้งยังเป็นการเพิ่มขีดความสามารถในการพัฒนาและเป็นการสร้างความได้เปรียบเหนือคู่แข่ง จากการเติบโตของเทคโนโลยีสารสนเทศและการสื่อสารที่มีพัฒนาการเพิ่มขึ้นอย่างรวดเร็ว ทำให้เกิดการสื่อสารแลกเปลี่ยนข้อมูล โดยเฉพาะเทคโนโลยีสมาร์ทโฟน (smart phone) เกิดเป็นแอปพลิเคชันบนระบบเครือข่ายอย่างมากมาย การเกิดชุมชนแบบออนไลน์และเครือข่ายสังคมออนไลน์ (social media) ทำให้มีปริมาณข้อมูลที่เพิ่มมากขึ้น ประกอบกับความหลากหลายของข้อมูลที่ต้องดำเนินการมีปริมาณที่เพิ่มขึ้นอย่างรวดเร็ว ก่อให้เกิดข้อมูลขนาดใหญ่ที่อยู่บนโลกออนไลน์ที่ไร้ข้อจำกัดในเรื่องของเวลาและสถานที่ ระบบการสืบค้นข้อมูลออนไลน์หรือเสิร์ชเอ็นจิน (search engine) เป็นเครื่องมือที่ได้รับความนิยม ทำให้การเข้าถึงข้อมูลที่อยู่บนโลกอินเทอร์เน็ตทำได้โดยง่าย โดยจากรูปที่ 1.1 แสดงถึงพัฒนาการเริ่มต้นจากการจัดเก็บและการเผยแพร่ข้อมูลที่มีการจัดเก็บข้อมูลในรูปแบบของไฟล์หรือฐานข้อมูลอยู่ภายในคอมพิวเตอร์แบบตั้งโต๊ะ (PC era) จนถึงการเผยแพร่ข้อมูลข่าวสารผ่านทางเว็บไซต์ (world wide web - web 1.0) การสื่อสารในโลกสังคมออนไลน์ (the social web - web 2.0) การค้นคืนข้อมูลเชิงความหมาย (the semantic web - web 3.0) จนถึงปัจจุบันอยู่ในยุคสมัยของเว็บชาญฉลาด (the intelligence web - web 4.0) ซึ่งมีทิศทางการพัฒนาเว็บในยุคถัดไปที่ข้อมูลบนโลกออนไลน์มีความเชื่อมโยงกันมากขึ้นในลักษณะของเครือข่ายเชิงความหมาย นำไปสู่การพัฒนาซอฟต์แวร์แอปพลิเคชันที่มีความชาญฉลาด โดยอาศัยเทคโนโลยีการประมวลผลทางภาษาธรรมชาติ (Natural Language Processing: NLP) ที่คอมพิวเตอร์จะมีความสามารถในการเข้าใจภาษามนุษย์ในเชิงความหมายมากขึ้น การแบ่งปันข้อมูลข่าวสารหรือโซเชี่ยลบุ๊กมาร์ก (human social tagging – “folksonomies”) เพื่อการจัดเก็บ แบ่งหมวดหมู่ และการค้นหาข้อมูลบนโลกออนไลน์มีความสะดวกมากยิ่งขึ้น และรวมถึงแนวคิดการจัดเก็บข้อมูลโดยอาศัยแนวคิดออนโทโลยี (ontology) เพื่ออธิบายความสัมพันธ์โครงสร้างความรู้แบบลำดับชั้นเชิงวัตถุ (object hierarchical data structure) และการนำเสนอความรู้ (knowledge representation)