

พิมพ์ครั้งที่

2

# คลังข้อมูล เพื่อการจัดการข้อมูลสมัยใหม่

DATA WAREHOUSING  
FOR MODERN DATA  
MANAGEMENT

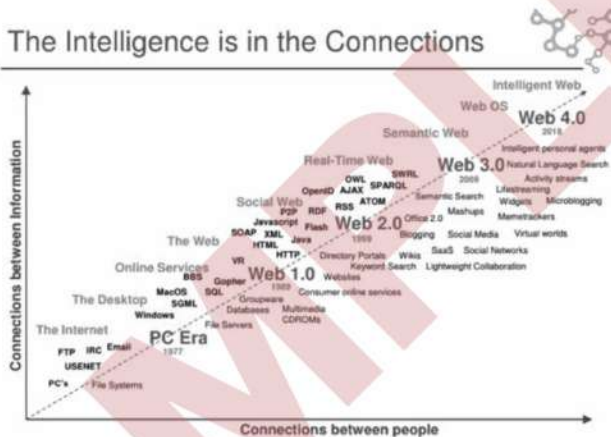
รัฐสิทธิ์ สุขะหุต

คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

## สารบัญ

|                                                 |          |
|-------------------------------------------------|----------|
| คำนิยาม                                         | I        |
| คำนำ (พิมพ์ครั้งที่ 1)                          | II       |
| คำนำ (พิมพ์ครั้งที่ 2)                          | III      |
| สารบัญ                                          | IV       |
| สารบัญภาพ                                       | XIV      |
| สารบัญตาราง                                     | XVIII    |
| <b>บทที่ 1 แนวคิดของการจัดการข้อมูลสมัยใหม่</b> | <b>1</b> |
| 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       |

language) นอกจากนี้ในยุคของ web 4.0 แอปพลิเคชันที่มีความฉลาดบนโลกออนไลน์ประเภทเอเจนต์ที่มีความฉลาด (Intelligent Agent: IA) จะทำให้เกิดการสื่อสารระหว่าง application-to-application ในลักษณะของการสื่อสารระหว่างเอเจนต์ด้วยกันเอง เพื่อแลกเปลี่ยนข้อมูลข่าวสารโดยไม่มีมนุษย์มาเกี่ยวข้อง ตัวอย่างเอเจนต์ในปัจจุบันได้แก่ การแลกเปลี่ยนข้อมูลบนเครือข่ายเว็บ การตรวจสอบราคาหุ้น การพยากรณ์สภาพอากาศ การเปรียบเทียบราคาโรงแรมที่พักในประเทศต่างๆ ดังแสดงในรูปที่ 1.2



รูปที่ 1.1 พัฒนาการของเทคโนโลยีเว็บ

ที่มา: <http://www.slideshare.net/novaspivack/web-evolution-nova-spivack-twine> (2009)

| Agent                | Percepts                  | Actions                    | Goals                    | Environment                   |
|----------------------|---------------------------|----------------------------|--------------------------|-------------------------------|
| ALVINN               | Images, Position          | Steering, Control speed    | Drive from A to B        | Roads, Vehicles, Hazards      |
| Google engine        | Web pages, Search queries | Navigate web, Gather info. | Find best search matches | Internet                      |
| Deep Blue            | Current board state       | Choose next move           | Win the game!            | Opponent, Game board          |
| Financial Forecaster | Stock market data         | Pick stocks to buy & sell  | Maximize investments     | Stock market, company reports |
| Medical Diagnostic   | Symptoms, Test results    | Tests, Treatments          | Healthy patient          | Patient, Hospital             |

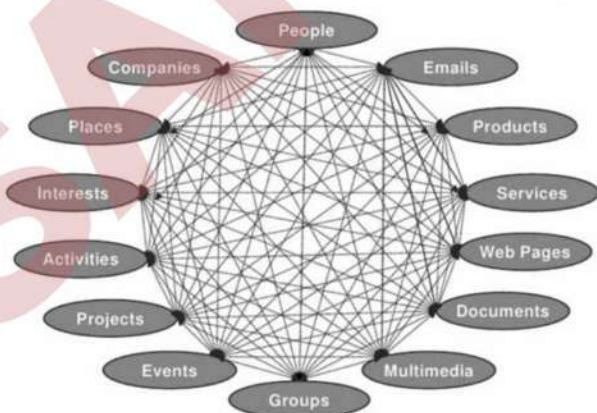
รูปที่ 1.2 ตัวอย่างเอเจนต์ที่ใช้ในแต่ละอุตสาหกรรม

ที่มา: [http://pages.cs.wisc.edu/~bsettles/cs540/lectures/02\\_agents.pdf](http://pages.cs.wisc.edu/~bsettles/cs540/lectures/02_agents.pdf) (2009)

การบริหารจัดการข้อมูลแบบบูรณาการภายใต้สภาพแวดล้อมและการเปลี่ยนแปลงในยุคโลกาภิวัตน์ ที่เทคโนโลยีอินเทอร์เน็ตมีบทบาทสำคัญในการสนับสนุนการทำงาน การทำพาณิชย์อิเล็กทรอนิกส์ การทำธุรกรรมอิเล็กทรอนิกส์ผ่านทางระบบเครือข่ายอินเทอร์เน็ต อาทิ การซื้อขายสินค้าแบบออนไลน์ ทำให้เกิดการสื่อสารแบบไร้พรมแดน ปัจจุบันจะพบว่าข้อมูลที่จัดเก็บภายในองค์กรมีปริมาณเพิ่มมากขึ้น รูปแบบการใช้ประโยชน์จากข้อมูลที่จัดเก็บไม่เพียงแต่มีการพัฒนาเป็นฐานข้อมูลเพื่อรองรับการทำงาน แต่จะต้องมองถึงการใช้ประโยชน์เชิงสารสนเทศ การวางแผนและประกอบกรตัดสินใจ ที่ต้องออกแบบโครงสร้างของข้อมูลที่จัดเก็บ เรียกว่าเป็นสถาปัตยกรรมโครงสร้างข้อมูลที่ต้องมีการวางแผน ซึ่งการวางแผนเป็นภาพรวมของระบบการบริหารจัดการข้อมูลตั้งแต่เริ่มต้น

### 1.3 พัฒนาการข้อมูลในปัจจุบัน

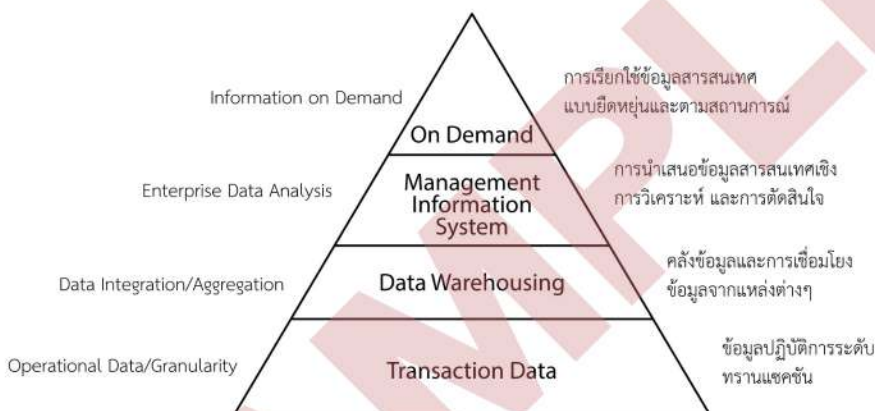
ปัจจุบันพัฒนาการเทคโนโลยีฮาร์ดแวร์มีขนาดที่เล็กลงเป็นอย่างมากเมื่อเทียบกับเทคโนโลยีในช่วง 10-20 ปีที่ผ่านมา แต่กลับมีประสิทธิภาพเพิ่มขึ้นแบบก้าวกระโดด อุปกรณ์ฮาร์ดแวร์มีความสามารถในการประมวลผล สามารถรองรับการทำงานแบบหลายๆ งานพร้อมกัน (multi-tasking) ระบบปฏิบัติการและซอฟต์แวร์ใหม่ๆ และพื้นที่จัดเก็บที่รองรับข้อมูลขนาดใหญ่ นอกจากนี้ความก้าวหน้าของเทคโนโลยีเครือข่ายและการสื่อสารที่ทำให้ข้อมูลเกิดการเชื่อมหากันแบบออนไลน์ ดังแสดงในรูปที่ 1.3 ปรากฏจากข้อจำกัดในเรื่องของสถานที่และเวลา ข้อมูลมีการสร้างขึ้นใหม่อยู่ตลอดเวลาทั้งในธุรกิจและในโลกของสังคมออนไลน์ การเปลี่ยนแปลงสำคัญคือ ผู้ใช้งานทุกคนกลายมาเป็นผู้สร้างข้อมูล (content creator) ซึ่งแต่เดิมจะถูกจำกัดอยู่เพียงแค่ผู้ดูแลระบบข้อมูล หรือผู้ดูแลเว็บไซต์ เท่านั้น



รูปที่ 1.3 เครือข่ายการสื่อสารที่เชื่อมหากันบนโลกออนไลน์

ที่มา: <http://www.slideshare.net/novaspivack/web-evolution-nova-spivack-twine> (2009)

**ระดับที่ 1 ข้อมูลระดับปฏิบัติการ (Transaction data)** เป็นฐานที่อยู่ในระดับล่างสุดของภาพปิรามิด เป็นข้อมูลที่เกิดจากข้อมูลปฏิบัติการที่มีพื้นฐานมาจากระบบงานพื้นฐาน ตัวอย่างเช่น ระบบบัญชีเงินเดือน หรือทะเบียนประวัติบุคลากร ข้อมูลเหล่านี้เป็นข้อมูลที่เกิดขึ้นจากระบบซอฟต์แวร์สำเร็จรูป ออกแบบมาเพื่อทำงานในงานนั้นๆ โดยข้อมูลนี้อาจจะถูกจัดเก็บอยู่ในรูปของฐานข้อมูลเชิงสัมพันธ์ หรือเป็นแฟ้มข้อมูลในรูปของไฟล์

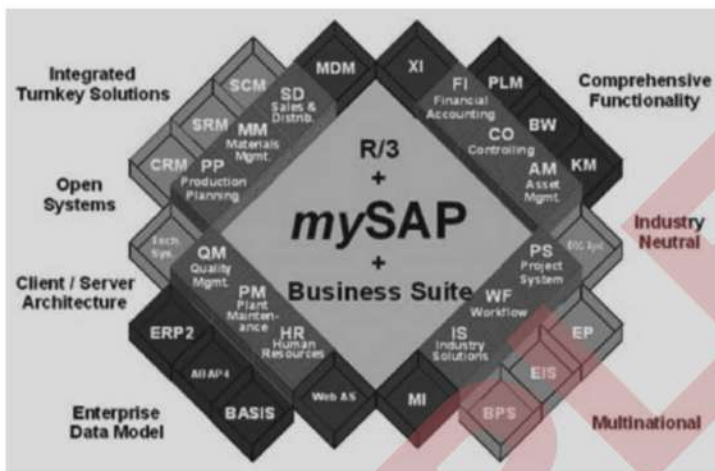


รูปที่ 1.4 สถาปัตยกรรมข้อมูลสารสนเทศ

**ระดับที่ 2 ระดับคลังข้อมูล (Data warehousing)** ที่มีการเชื่อมโยงข้อมูลจากข้อมูลระดับปฏิบัติการ ซึ่งหมายถึงแฟ้มข้อมูล และข้อมูลในระดับทรานแซกชัน โดยข้อมูลที่น่ามาจัดเก็บในคลังนั้นจะเป็นข้อมูลในเชิงสรุป (data aggregation) และการเชื่อมโยง (data integration) จากแหล่งข้อมูลต่างๆ

**ระดับที่ 3 เป็นการวิเคราะห์และการนำเสนอข้อมูลเพื่อการตัดสินใจ (Management information system)** โดยการนำเสนอข้อมูลในรูปของสารสนเทศและข้อมูลเพื่อการบริหาร เช่น ข้อมูลแผนและยุทธศาสตร์ขององค์กร เชื่อมโยงไปยังข้อมูลตัวชี้วัด เพื่อแสดงถึงประสิทธิภาพการทำงานขององค์กร

**ระดับที่ 4 เป็นการเรียกใช้ข้อมูลสารสนเทศแบบยืดหยุ่นหรือตามสถานการณ์ที่เกิดขึ้น (On demand)** โดยผู้บริหารสามารถเรียกดูรายงานได้ตามความต้องการ และสามารถปรับเปลี่ยนเงื่อนไขของการแสดงผลรายงานได้



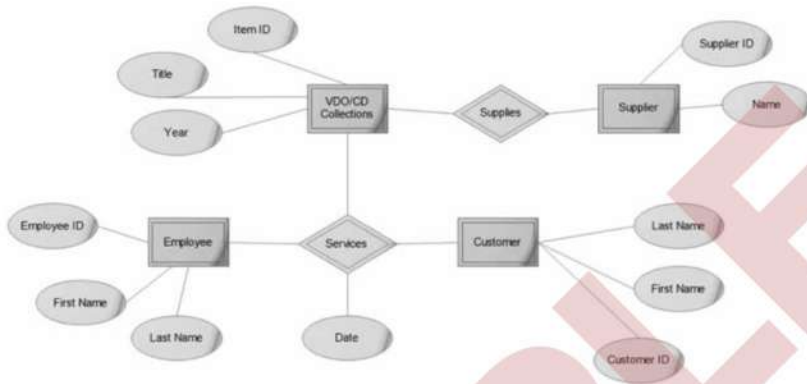
รูปที่ 1.5 โครงสร้างระบบ SAP อีอาร์ที

ที่มา: <http://sapbotrainingonline.com> (2014)

## ระบบฐานข้อมูลเชิงสัมพันธ์

ระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database Management System: RDBMS) เป็นระบบฐานข้อมูลที่มีโครงสร้างข้อมูลในรูปของตาราง (relation) โดยข้อมูลจะถูกจัดเก็บในรูปของแถว (row) และสตัมภ์ (column) และมาตรฐานการจัดเก็บข้อมูลที่มีกฎ เงื่อนไข ชนิด และรูปแบบการจัดเก็บ และรวมถึงการกำหนดความสัมพันธ์ข้อมูล อีกทั้งมีภาษาเอสคิวแอลที่ใช้สำหรับการควิรี่ข้อมูล ฐานข้อมูลเชิงสัมพันธ์เป็นฐานข้อมูลรองรับข้อมูลที่มีการเคลื่อนไหวอยู่ตลอดเวลา เช่น การเพิ่ม การลบ และการปรับปรุงข้อมูลในคลังสินค้า ซึ่งฐานข้อมูลเชิงสัมพันธ์เป็นฐานข้อมูลที่สนับสนุนการทำงานประจำของเจ้าหน้าที่ในระดับปฏิบัติการ โดยการปรับปรุงข้อมูลในฐานข้อมูลจะทำผ่านแอปพลิเคชันหรือโปรแกรมที่ออกแบบมาเฉพาะงานนั้นๆ เช่น โปรแกรมบัญชี สินค้าคงขาย หรือโปรแกรมขายหน้าร้าน

การออกแบบฐานข้อมูลเชิงสัมพันธ์จะอาศัยไดอะแกรมความสัมพันธ์เอนทิตี (entity-relationship diagram หรือ ER-diagram) การออกแบบจะมีการวิเคราะห์ฟังก์ชันการทำงาน การจัดเก็บข้อมูลในระดับปฏิบัติการซึ่งมีรายละเอียดข้อมูลที่มาก และเพื่อให้รองรับการจัดการข้อมูล ในลักษณะการเพิ่ม ลบ แก้ไข และการควิรี่ข้อมูล (create/read/update/delete) เรียกโดยย่อว่า CRUD การออกแบบฐานข้อมูลเชิงสัมพันธ์มุ่งเน้นการลดความซ้ำซ้อนของข้อมูลผ่านกระบวนการนอร์มัลไลซ์ (normalization) การควิรี่ข้อมูลในฐานข้อมูลเชิงสัมพันธ์ทำผ่านคำสั่งเอสคิวแอล การเชื่อมตารางข้อมูลต่างๆ โดยวิธีการเชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน เพื่อการประมวลผลข้อมูลและรายงานผลได้



รูปที่ 1.6 โครงสร้างฐานข้อมูลร้านเช่าวิดีโอในระดับ conceptual design

สำหรับการจัดการข้อมูลในมุมมองของคลังข้อมูลนั้น พบว่าโครงสร้างของระบบฐานข้อมูลเชิงสัมพันธ์มีข้อจำกัดในการค้นหาหรือการเรียกดูข้อมูลสำหรับข้อมูลขนาดใหญ่ ที่เน้นการนำข้อมูลไปใช้ในการวิเคราะห์และตัดสินใจทางธุรกิจ เนื่องจากการเชื่อมโยงข้อมูลจากหลายตารางทำให้ประสิทธิภาพในการประมวลผลของคิวรีลดลง แนวคิดของคลังข้อมูลมุ่งเน้นการคิวรีข้อมูลที่มีขนาดใหญ่มากกว่าการเพิ่ม ลบ แก้ไขข้อมูล ดังนั้นการออกแบบคลังข้อมูลจึงมีกระบวนการที่เน้นการดีนอร์มัลไลซ์ (denormalization) ข้อมูล ซึ่งเป็นกระบวนการเชื่อมตารางข้อมูลที่จัดเก็บแยกกันคนละตารางให้กลายเป็นตารางที่เพียงตารางเดียว โดยยอมให้ข้อมูลเกิดความซ้ำซ้อนของข้อมูล เพื่อประโยชน์ในการเรียกดูข้อมูล นอกจากนี้การออกแบบคลังข้อมูลจะอาศัยแบบจำลองข้อมูลเชิงมิติ (Dimension Model: DM) ซึ่งต่างจากแนวคิดของการออกแบบฐานข้อมูลเชิงสัมพันธ์ที่อาศัยการออกแบบความสัมพันธ์แบบเอนทิตี ซึ่งจะอธิบายรายละเอียดในบทต่อไป

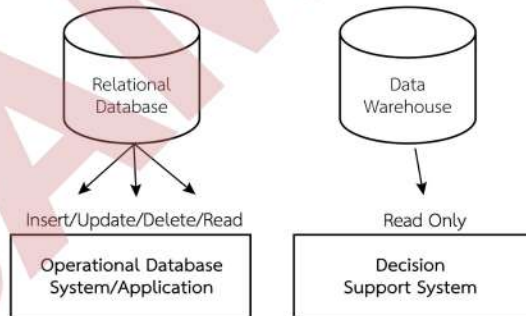
### การประมวลผลออนไลน์เชิงวิเคราะห์

การประมวลผลออนไลน์เชิงวิเคราะห์หรือโอแลป (OLAP) เป็นซอฟต์แวร์เครื่องมือที่ใช้สำหรับประมวลผลข้อมูลเชิงมิติ และการวิเคราะห์ข้อมูลแบบหลายมิติ เหมาะสำหรับข้อมูลที่มีปริมาณมาก โครงสร้างข้อมูลทีโอแลปนำมาประมวลผลจะอยู่ในรูปของมิติใดเมนชันโมเดล โดยมุมมองเชิงมิติสำหรับผู้บริหารเพื่อใช้ประโยชน์จากข้อมูลที่จัดเก็บเพื่อประกอบการตัดสินใจอย่างมีประสิทธิภาพ

## แนวคิดของการพัฒนาคลังข้อมูล

การพัฒนาคลังข้อมูลเป็นการออกแบบเพื่อรวบรวมข้อมูลที่มีคุณลักษณะเหมือนกันจัดอยู่ในชุดข้อมูลเดียวกัน ข้อมูลที่จัดเก็บภายในคลังข้อมูลเป็นข้อมูลทุติยภูมิ เนื่องจากเป็นข้อมูลที่จำเป็นต้องมีการรวบรวมและการเชื่อมโยงจากแหล่งต่างๆ ข้อมูลระดับปฏิบัติการ โดยข้อมูลที่นำมาเก็บภายในคลังข้อมูลนั้นจะต้องผ่านกระบวนการสกัด การแปลง และการโหลดข้อมูล อีกทั้งข้อมูลที่เก็บภายในคลังข้อมูลนั้นจะเป็นข้อมูลในเชิงสรุปเพื่อให้การประมวลผลข้อมูลมีความรวดเร็ว เป็นฐานข้อมูลที่ถูกออกแบบเพื่อรองรับการควิรีข้อมูลเชิงมิติ ข้อมูลที่จัดเก็บจะไม่มีมีการปรับปรุงแก้ไข ต่างจากฐานข้อมูลปฏิบัติการ การออกแบบคลังข้อมูลจึงเป็นการเตรียมความพร้อมของข้อมูลเพื่อการควิรีข้อมูลขนาดใหญ่ เพื่อใช้ประโยชน์ในเชิงการวิเคราะห์และการตัดสินใจ ตอบสนองต่อความต้องการผู้ใช้งานที่ต้องการระบบข้อมูลที่มีความยืดหยุ่นสูง

การพัฒนาคลังข้อมูลมีกระบวนการที่ซับซ้อน ซึ่งจำเป็นต้องมีการวิเคราะห์และออกแบบโครงสร้างสถาปัตยกรรมของข้อมูลเพื่อให้เหมาะสมกับรูปแบบและวัตถุประสงค์ของการนำข้อมูลไปใช้ประโยชน์ และให้สอดคล้องต่อการทำงานของเจ้าหน้าที่และบุคลากรอื่นๆ ที่เกี่ยวข้อง การพัฒนาจะต้องมองภาพโครงสร้างของข้อมูลโดยรวมซึ่งเป็นส่วนที่สำคัญ เนื่องจากการออกแบบโครงสร้างข้อมูลจะมีผลต่อการรวบรวมและจัดเก็บ การไหลของข้อมูลจากแหล่งข้อมูลต้นทาง รวมถึงการควิรีข้อมูลเพื่อนำไปใช้ประโยชน์ นอกจากนี้ยังมีผลต่อกิจกรรมต่างๆ ที่เกิดขึ้น



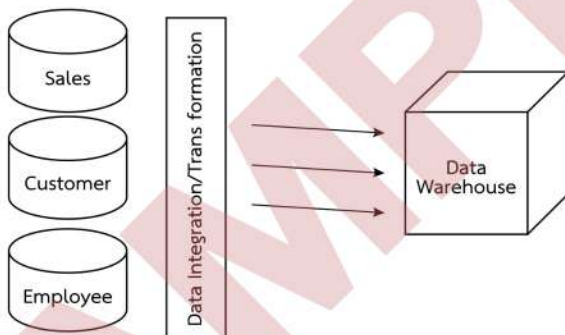
รูปที่ 2.1 ความแตกต่างระหว่างฐานข้อมูลปฏิบัติการคลังข้อมูล

รูปที่ 2.1 แสดงถึงความแตกต่างในการใช้ประโยชน์จากข้อมูลที่จัดเก็บอยู่ในฐานข้อมูลเชิงสัมพันธ์ และคลังข้อมูล ซึ่งจะเห็นว่าฐานข้อมูลเชิงสัมพันธ์นั้นสนับสนุนการทำงานเชิงปฏิบัติการ ซึ่งเป็นการเพิ่ม ลบ และแก้ไขข้อมูล ในขณะที่คลังข้อมูลจะใช้ข้อมูลเพื่อการอ่านเพียงอย่างเดียวและใช้เพื่อสนับสนุนการตัดสินใจ



## Integrated data

ข้อมูลที่น่ามาจัดเก็บในคลังข้อมูลจะมีการรวบรวมมาจากแหล่งต่างๆ แหล่งข้อมูลประกอบด้วยข้อมูลจากระบบการประมวลผลสายธุรกรรมข้อมูล ที่จัดเก็บโดยแอปพลิเคชัน ซึ่งอาจอยู่ในรูปของฐานข้อมูลเชิงสัมพันธ์ ไฟล์เอกสาร หรือในรูปแบบอื่นๆ เช่นไฟล์ที่มีรูปแบบเฉพาะ (proprietary data) แต่แหล่งที่มาของข้อมูลจะมีความแตกต่างกันทั้งในเรื่องของรูปแบบการจัดเก็บข้อมูล ชนิดข้อมูล หรือการกำหนดตารางและชื่อฟิลด์ การเชื่อมโยงข้อมูลเป็นคุณลักษณะที่สำคัญของคลังข้อมูล ที่จะรวบรวมข้อมูลโดยวิธีการสกัด (extract) ดักจับ (capture) ข้อมูลจากแหล่งข้อมูลตามช่วงเวลาเพื่อรวบรวมเป็นคลังข้อมูล โดยหนึ่งในกระบวนการที่สำคัญคือ การแปลงข้อมูล (transform) จากแหล่งต่างๆ ให้อยู่ในมาตรฐานเดียวกัน ก่อนที่จะมีการโหลดข้อมูลเข้าสู่คลังข้อมูล



รูปที่ 2.2 การรวบรวมข้อมูลจากแหล่งข้อมูลต่างๆ เป็นคลังข้อมูล

รูปที่ 2.2 แสดงถึงขั้นตอนการเชื่อมโยงข้อมูลจากแหล่งข้อมูลต่างๆ เพื่อให้เป็นคลังข้อมูล โดยมีขั้นตอนดังต่อไปนี้

1. การกำหนดแหล่งข้อมูล (Data source identification) เป็นการกำหนดแหล่งที่มาของคลังข้อมูลที่จะนำมาสู่คลังข้อมูล แหล่งข้อมูลแบ่งเป็น 2 ประเภท คือ 1) แหล่งข้อมูลภายในเป็นข้อมูลที่มีการพัฒนาและจัดเก็บภายในองค์กรเอง และ 2) แหล่งข้อมูลภายนอก ซึ่งเป็นข้อมูลที่มาจากภายนอกองค์กรที่นำมาประกอบการพัฒนาเป็นข้อมูลเพื่อการตัดสินใจ ได้แก่ ข้อมูลเชิงสถิติจากสถาบันต่างๆ บทวิเคราะห์ และบทความทางวิชาการต่างๆ
2. การศึกษารูปแบบข้อมูลในแต่ละแหล่งข้อมูล (Data source analysis) เป็นการวิเคราะห์โครงสร้าง รูปแบบการจัดเก็บข้อมูลในแต่ละแหล่ง

ปัจจุบันในยุคสมัยของโลกสังคมดิจิทัลและสื่อออนไลน์ที่มีการพึ่งพาเทคโนโลยีสารสนเทศมากยิ่งขึ้น วิทยาการข้อมูลได้เข้ามามีบทบาทสำคัญ องค์กรต่างๆ ทั้งหน่วยงานในภาครัฐและเอกชนที่มีการจัดเก็บข้อมูลในรูปของสื่ออิเล็กทรอนิกส์ต่างให้ความสำคัญต่อการจัดเก็บข้อมูลเพื่อนำข้อมูลมาใช้เพื่อประกอบการวางแผน วิเคราะห์ และเป็นข้อมูลสนับสนุนการตัดสินใจของผู้บริหาร แนวคิดของการจัดการข้อมูลสมัยใหม่เป็นศาสตร์ที่ว่าด้วยการบริหารจัดการข้อมูลอย่างเป็นระบบ เป็นการผสมผสานระหว่างวิทยาการคอมพิวเตอร์ หลักสถิติ คณิตศาสตร์ และซอฟต์แวร์ และต้องอาศัยความรู้ทางด้านการจัดการแบบบูรณาการ ซึ่งจำเป็นต้องให้ความสำคัญต่อการออกแบบโครงสร้างและสถาปัตยกรรมข้อมูลที่ดี เพื่อให้การใช้ข้อมูลสารสนเทศเกิดประโยชน์สูงสุดต่อองค์กร

DATA WAREHOUSING  
FOR MODERN DATA  
MANAGEMENT



CHIANG MAI  
UNIVERSITY PRESS

ISBN 978-616-398-310-7



9 786163 983107

ราคา 450 บาท