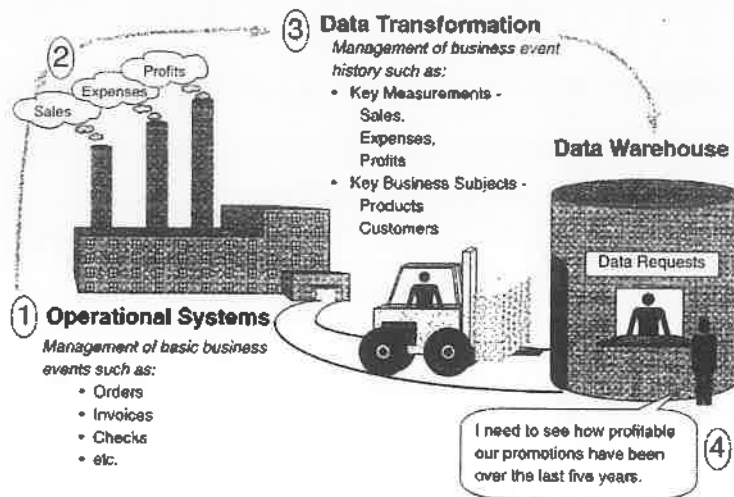


On-Line Analytical Processing

สมชาย เล็กเจริญ*

ซอฟต์แวร์เทคโนโลยีทางด้านอินเทอร์เน็ต และดาต้าแวร์เฮาส์กำลังเป็นที่นิยมทั้งปัจจุบันและอนาคต ซึ่งสามารถพิจารณาได้จากผู้ผลิตซอฟต์แวร์ จะต้องพัฒนาซอฟต์แวร์ของตนเองให้สามารถแสดงผลบน เวิลด์ไวด์เว็บ หรือสามารถเชื่อมต่อในลักษณะใดลักษณะหนึ่งกับอินเทอร์เน็ต แต่ในส่วนของระบบฐาน ข้อมูลนั้นผู้ผลิตซอฟต์แวร์ทางด้านฐานข้อมูลพยายามเพิ่มประสิทธิภาพให้มีประสิทธิภาพสามารถเปลี่ยนแปลง การจัดเก็บข้อมูลไปสู่ดาต้าแวร์เฮาส์ ดังรูปที่ 1 เช่น Oracle DB2 Informix Sybas เป็นต้น ถึงแม้ว่า จะมีดาต้าแวร์เฮาส์ที่ทรงประสิทธิภาพแล้วก็ตาม แต่ผู้ใช้ระดับผู้บริหารต้องการผลของการแสดงข้อมูลใน ลักษณะหลายมิติ หรือกราฟเพื่อใช้ประกอบการตัดสินใจ ฉะนั้นจึงต้องมีเครื่องมือประเภท On-Line Analytical Processing มาช่วยตอบสนองความต้องการ ดังกล่าว



รูปที่ 1 : แสดงการเปลี่ยนแปลงข้อมูลการทำงานประจำวัน (operational system)

เปลี่ยนแปลงการจัดเก็บแบบดาต้าแวร์เฮาส์ (Hammergren, Tom 1996 : 7)

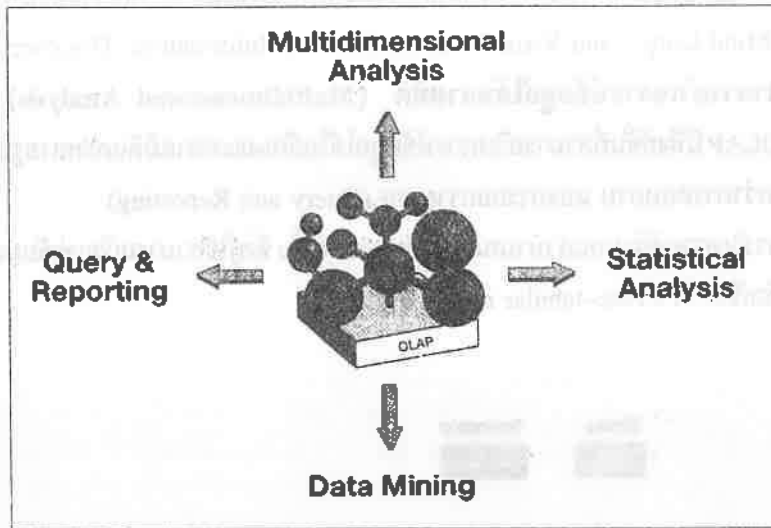
สามารถอ่านเพิ่มเติมได้จาก “รั้งลิตสารสนเทศปีที่ 3 ฉบับที่ 1 มกราคม-มิถุนายน 2540”

On-line Analytical Processing หรือเรียกย่อๆ ว่า OLAP เป็นซอฟต์แวร์ประเภทหนึ่งที่สามารถดึงข้อมูลที่จัดเก็บไว้ในดาต้าแวร์เฮาส์

และสามารถแสดงผลได้หลายรูปแบบ เช่น รายงานเชิงวิเคราะห์ รูปของกราฟ สเปรดชีท (Spreadsheet) เป็นต้น

* อาจารย์ประจำภาควิชา การจัดการสารสนเทศ คณะนิเทศศาสตร์ มหาวิทยาลัยรังสิต

โดยทั่วไปแล้ว OLAP มีความสามารถพื้นฐาน 4 ประการ ในการประมวลผลข้อมูล ดังรูปที่ 2



รูปที่ 2 : ความสามารถของ OLAP ในการวิเคราะห์ข้อมูล (Tanler,Richard 1997:94)

1. การสอบถามและการออกรายงาน (Query and Reporting)

OLAP ได้ถูกออกแบบให้มีความสามารถให้การสอบถามโดยมีคำสั่ง sq2 (Structured Query Language) ในการดึงข้อมูลจากฐานข้อมูลเชิงสัมพันธ์ (ปัจจุบันนี้โดยส่วนมากแล้ว OLAP ใช้กับฐานข้อมูลเชิงสัมพันธ์ (RDBMS) มากกว่าฐานข้อมูลเชิงมิติ (Multidimensional Database-MDD) และสามารถออกรายงานรูปแบบต่างๆ ตลอดจนสามารถแสดงกราฟประเภทต่างๆ ได้

2. สามารถใช้สถิติในการวิเคราะห์ข้อมูล (Statistical Analysis)

การใช้สถิติในการวิเคราะห์ข้อมูลจำนวนมากให้อยู่ในรูปที่ผู้ใช้สามารถเข้าใจได้ง่าย จึงทำให้ OLAP มีความซับซ้อนมากยิ่งขึ้น สถิติที่ใช้ใน OLAP เช่น หาค่าเฉลี่ย (mean or average) หาค่าความสัมพันธ์ของข้อมูล (correlation) การจัดกลุ่มของข้อมูล (factor analysis and clustering) การวิเคราะห์ความถดถอย (Regression) ตลอดจน

ใช้สถิติในการทำนายแนวโน้มของการขาย และส่วนแบ่งของตลาด โดยใช้สถิติ Time series analysis เป็นต้น

3. ดาต้าไมนิ่ง (Data Mining)

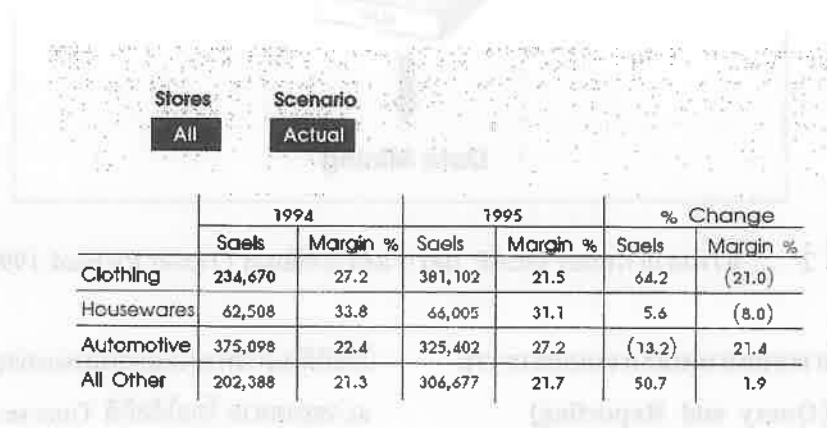
มีวิธีการเหมือนกับใช้สถิติในการวิเคราะห์ข้อมูล แต่เพิ่มฟังก์ชันที่ซับซ้อนมากยิ่งขึ้น เช่น Neural Network ที่จะกำหนดรูปแบบและความสัมพันธ์ภายในข้อมูลตลอดจนมีวิธีการจดจำและเรียนรู้รูปภาพของข้อมูลเพื่อสร้างแบบจำลองในการทำนายแนวโน้มอนาคตของธุรกิจ เช่น ต้องการทราบปัจจัยเกี่ยวกับฉ้อโกงและการกำหนดอัตราเสี่ยงของบัตรเครดิต (Credit Card) ดาต้าไมนิ่งทำการรวบรวมข้อมูลประวัติและพฤติกรรม การใช้บัตรเครดิตไปในทางที่ผิดของผู้ใช้ ถ้าพิจารณาให้ดีจะเห็นได้ว่ามีข้อมูล ข้อมูลที่เกี่ยวข้องที่เป็นไปได้จำนวนมากและมีตัวแปรจำนวนมาก ซึ่งถ้าใช้วิธีการวิเคราะห์ข้อมูลตามปกติจะทำได้ลำบาก ฉะนั้น ดาต้าไมนิ่งจะช่วยให้การวิเคราะห์ข้อมูลที่ซับซ้อนได้เป็นอย่างดี

ถึงแม้ว่าค่าตัวไมนิ่งจะเป็นสิ่งใหม่แต่ก็ได้ รับการพัฒนาอย่างรวดเร็วและเป็นปัจจัยหลักที่ผู้ผลิตซอฟต์แวร์ประเภทนี้ให้ความสนใจที่พัฒนาผลิตภัณฑ์ของตนเองให้มีค่าตัวไมนิ่ง เช่น IBM, HNC Software Inc., Data Mind Corp., Neo Vista Solutions Inc. และ Information Discovery Inc. เป็นต้น

4. สามารถวิเคราะห์ข้อมูลได้หลายมิติ (Multidimensional Analysis)

OLAP มีฟังก์ชันที่สามารถวิเคราะห์ข้อมูลได้ในลักษณะหลายมิติและหลายรูปแบบอย่างรวดเร็ว จึงทำให้เด่นกว่าการสอบถาม และการออกรายงาน (Query and Reporting)

การวิเคราะห์ข้อมูลแล้วนำเสนอในรูปแบบหลายมิตินั้น ซึ่งผู้ใช้สามารถเรียกดูข้อมูลในลักษณะตารางหลายมิติ หรือเรียกว่า Cross-tabular report ดังรูปที่ 3



	1994		1995		% Change	
	Saels	Margin %	Saels	Margin %	Saels	Margin %
Clothing	234,670	27.2	381,102	21.5	64.2	(21.0)
Housewares	62,508	33.8	66,005	31.1	5.6	(8.0)
Automotive	375,098	22.4	325,402	27.2	(13.2)	21.4
All Other	202,388	21.3	306,677	21.7	50.7	1.9

รูปที่ 3 OLAP แสดงผลหลายมิติ (Barquin, Ramon and Herb Edelstein 1997:206)

เมื่อพิจารณาแสดงรูปแบบออกมาถึง 5 มิติ คือ

มิติที่ 1 แสดงแผนกต่าง ๆ ในที่นี้มีแผนก Clothing, Housewares, Automotive และ All Other

มิติที่ 2 แสดงเรื่องของเวลา คือ ปี 1994 1995

มิติที่ 3 แสดงเรื่องการขาย คือ Sales และ Margin

มิติที่ 4 แสดงร้านขายปลีก คือ Store ในที่นี้แสดงผลรวม

มิติที่ 5 แสดงภาพรวม (Scenario ประกอบด้วย Actual, Plan, Update Forecasting เป็นต้น

หรืออาจต้องการข้อมูลที่มีรายละเอียดมากขึ้น เรียกว่า Drill-Down Report ซึ่งพิจารณาจากรูปที่ 2 ในส่วนของ Automotive ผู้ใช้อาจใช้เมาส์ (Mouse) ไปคลิกที่ Automotive เพื่อต้องการดูรายละเอียดให้มากยิ่งขึ้น ก็จะปรากฏรายละเอียดเกี่ยวกับ Repair, Accessories และ Audio ซึ่งอยู่ภายใต้หัวข้อของ Automotive จากรูปที่ 4

	Stores		Scenario			
	All		Actual			
	1994		1995		% Change	
	Saels	Margin %	Saels	Margin %	Saels	Margin %
Clothing	234,670	27.2	381,102	21.5	64.2	(21.0)
Housewares	62,508	33.8	66,005	31.1	5.6	(8.0)
Automotive	375,098	22.4	325,402	27.2	(13.2)	21.4
Repair	195,051	14.2	180,786	15.0	(7.3)	5.6
Accessories	116,280	43.9	122,545	47.5	5.3	8.2
Audio	63,767	8.2	22,071	14.2	(65.4)	7.3
All Other	202,388	21.3	306,677	21.7	50.7	1.9

รูปที่ 4 : OLAP แสดงผลแบบ Drill-Down Report (Barquin, Ramon and Herb Edelstein 1997:207)

หรือผู้ใช้ต้องการเรียกดูเพียงบางส่วน เรียกว่า Slice Report ซึ่งจะแสดงข้อมูลเพียงบางคอลัมน์และบางแถวของข้อมูลเท่านั้น ดังรูปที่ 5 แสดงเพียงคอลัมน์เดียวเกี่ยวกับการขายปี 1995 เท่านั้น

Stores	Scenario
	Actual
	1995
	Sales
Clothing	381,102
Housewares	66,005
Automotive	325,402
All Other	305,677

รูปที่ 5 : OLAP แสดงผลแบบ Slice Report (Barquin, Ramon and Herb Edelstein 1997:207)

และผู้ใช้สามารถเรียกดูจุดสำคัญตามความต้องการเรียกว่า Pivot Report ดังรูปที่ 6 ซึ่งต้องพิจารณาจากรูปที่ 5 ใช้เมาส์คลิกแล้วลาก (Dragged) จาก Scenario ไปถึง Sale OLAP จะรายงานรายละเอียดของ Senario ประกอบด้วย Actual, Plan และ Variance

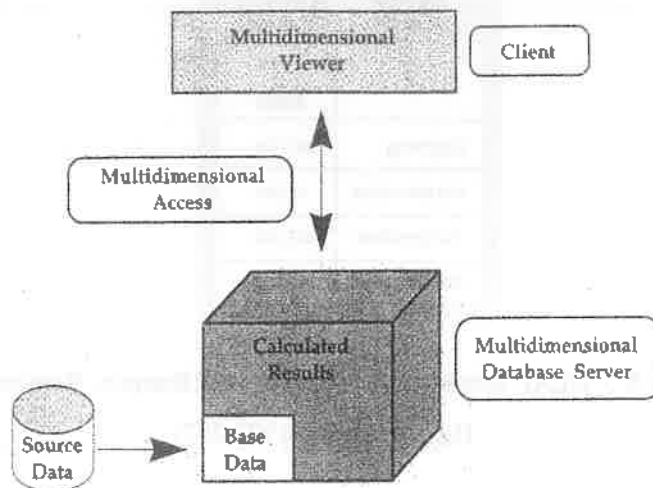
Stores
All

	1995			
	Sales			
	Actual	Plan	Variance	%
Clothing	381,102	350,000	31,102	8.9
Housewares	66,005	69,000	(2,995)	(4.3)
Automotive	325,492	300,000	25,492	8.5
All Other	305,677	350,000	(44,322)	(12.7)

รูปคค 6 OLAP แสดงผลแบบ Pivot Report (Barquin, Ramon and Herb Edelstein 1997:208)

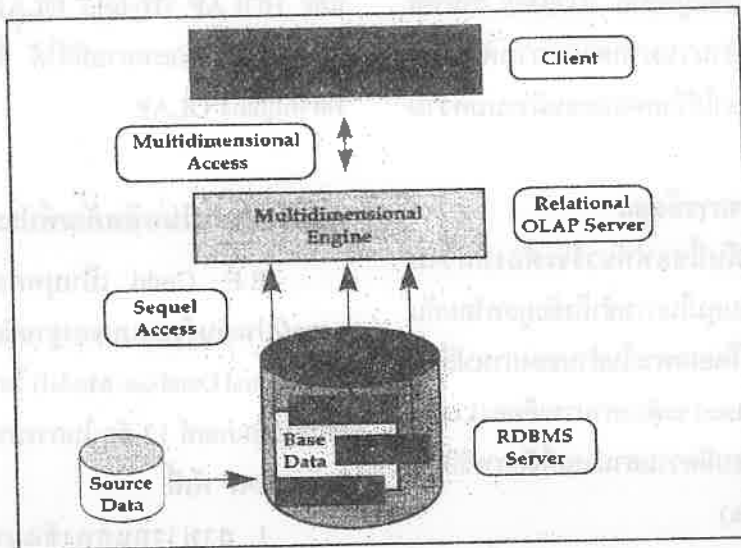
องคคคประกอบของ OLAP

OLAP ประกอบด้วยซอฟต์แวร์และข้อมูล นอกจากนั้น จะคคคมีความสามารถคคคข้อมูลจากฐานข้อมูลหลายมคคค (Multidimensional Database) คคครูปคค 7



รูปคค 7 องคคคประกอบของ OLAP (Singh, Harry 1998 : 173)

หรือฐานข้อมูลแบบสัมพันธ์ (Relational Database) โดยคคคใช้ ROLAP (relational OLAP) เป็นคคคคคคข้อมูล คคครูปคค 8



รูปที่ 8 รูปแบบของ Relational OLAP (Singh, Harry 1998:174)

ฉะนั้นจึงจำเป็นต้องเข้าใจถึงความแตกต่างของการทำงานทั้งระบบเพื่อที่สามารถเลือกใช้ที่เหมาะสมกับระบบงานธุรกิจของตนเอง ซึ่งมีหลักเกณฑ์พิจารณาดังต่อไปนี้

1. การจัดเก็บข้อมูลในระดับกายภาพ (Physical Data Storage)

ในสถาปัตยกรรมของฐานข้อมูลถือว่าเป็นระดับขั้นล่างสุด ซึ่งเกี่ยวข้องกับการจัดเก็บข้อมูลจริงในฮาร์ดดิสก์และผลลัพธ์ที่ถูกคำนวณไว้ล่วงหน้า ซึ่งการเข้าถึงข้อมูลดังกล่าวจะต้องผ่าน DBMS เพื่อควบคุมความปลอดภัยของข้อมูล เมื่อพิจารณารูปที่ 6 เป็นการจัดเก็บข้อมูลแบบฐานข้อมูลหลายมิติ (Multidimensional Database-MDD) ส่วนรูปที่ 7 เป็นการจัดเก็บข้อมูลแบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ซึ่งทั้งสองแบบการจัดเก็บข้อมูลประกอบด้วย ข้อมูล ผลลัพธ์จากการคำนวณล่วงหน้า ตรรกะที่ใช้ในการค้นคืนข้อมูลได้รวดเร็ว และเมตาดาตา (Metadata) ใช้อธิบายโครงสร้างทางกายภาพของข้อมูล (Physical Data Structure)

- การจัดเก็บข้อมูลในฐานข้อมูลหลายมิติ (Data Storage in an MDD)

ฐานข้อมูลหลายมิติ (Multidimensional Database) เป็นซอฟต์แวร์ที่ถูกออกแบบมาเพื่อประสิทธิภาพและความสะดวกในการจัดเก็บข้อมูลและการค้นคืนข้อมูลจำนวนมาก ตลอดจนแสดงและวิเคราะห์ข้อมูลได้หลายมิติ ฉะนั้นฐานข้อมูลหลายมิติจึงมีโครงสร้างการจัดเก็บข้อมูลในลักษณะของอะเรย์ (Array) และมีอินเด็กซ์เป็นตัวชี้ในแต่ละส่วนของอะเรย์

- การจัดเก็บข้อมูลใน ROLAP (Relational OLAP Data Storage)

ROLAP เป็นซอฟต์แวร์ใช้ในการดึงข้อมูลที่จัดเก็บไว้ในดาต้าแวร์เฮาส์เพื่อแสดงผลในรูปของหลายมิติ ฉะนั้นข้อมูลที่ป้อนเข้า ROLAP จะเป็นแบบคืนอร์มัลไลซ์ (Denormalized) ซึ่งมีโครงสร้างแบบสตาร์ (Star Schema)

2. ประสิทธิภาพของการจัดเก็บข้อมูล

ถ้าพิจารณาถึงประสิทธิภาพของการจัดเก็บและค้นคืนข้อมูล อาจตัดสินใจใช้ ROLAP

เพราะว่าการจัดเก็บข้อมูลแบบ RDBMS สามารถจัดเก็บข้อมูลได้จำนวนมากตามความต้องการสามารถบำรุงรักษาได้ง่ายตลอดจนมีระบบความปลอดภัยสูง

3. การจัดการข้อมูล

ในระดับนี้ซอฟต์แวร์จะต้องมีความสามารถในการควบคุมในการเข้าถึงข้อมูลพร้อมกันของผู้ใช้หลายคน โดยเฉพาะในส่วนของ การเปลี่ยนแปลงข้อมูล (Update) จะต้องสามารถล็อก (Lock) ข้อมูลไว้ได้ตลอดจนมีความสามารถที่จัดการข้อมูลภายในดิสก์ (Disk)

ในส่วนของ MDD จะต้องมีความรู้เกี่ยวกับโครงสร้างแบบหลายมิติ ดังนั้นการจัดการข้อมูลจะต้องขึ้นอยู่กับมิติของข้อมูล จึงทำให้การควบคุมความปลอดภัยของข้อมูลกระทำได้อย่าง ฉะนั้น MDD จะไม่สนับสนุนการเปลี่ยนแปลงของข้อมูลแบบทันทีทันใด

ในส่วนของ RDBMS สามารถควบคุมความปลอดภัยของข้อมูลโดยใช้หลักการลดคircular row และคอลัมน์ ดังนั้น ทำให้ ROLAP มีความปลอดภัยเกี่ยวกับข้อมูลไปด้วย เพราะ ROLAP ต้องไปดึงข้อมูลจาก RDBMS

4. การคำนวณและกระทำกับข้อมูลหลายมิติ

OLAP ทุกชนิดมีความสามารถกระทำกับข้อมูลหลายมิติและมีความสามารถคำนวณข้อมูลจำนวนมาก ๆ ตั้งแต่สิบถึงหลายร้อยรายการได้

แต่ MDD มีความสามารถเด่นกว่าในเรื่องการคำนวณข้อมูลหลายมิติและได้อย่างรวดเร็ว

5. สามารถแสดงผลมุมมองได้หลายมิติ

OLAP ทุกประเภท เช่น ROLAP (Relational OLAP), MOLAP (Multidimensional OLAP), DOLAP (Desktop OLAP)

และ HOLAP (Hybrid OLAP) จะต้องแสดงผลลัพธ์ในลักษณะหลายมิติได้ ซึ่งเป็นคุณสมบัติที่สำคัญของ OLAP

กฎการประเมินผลลักษณะที่ประเภท OLAP

E.F. Codd เป็นบุคคลที่มีชื่อเสียงในฐานะผู้ประดิษฐ์รูปแบบของฐานข้อมูลเชิงสัมพันธ์ (Relational Database Model) ในส่วนของ OLAP ได้ให้หลักเกณฑ์ 12 ข้อ ในการประเมินผลลักษณะที่ของ OLAP ดังนี้

1. สามารถแสดงข้อมูลได้หลายมิติ (Multi-Dimensional Conceptual View) OLAP จะต้องมีความสามารถในการแสดงผลที่ออกมาได้หลายมิติ และอนุญาตให้ผู้ใช้สามารถดูรายละเอียดของข้อมูลแบบเจาะลึก (Drill-Down) หรือดูข้อมูลเพียงบางส่วน (Slice and Dice)

2. สามารถเข้าถึงแหล่งข้อมูลภายนอก (Transparency) OLAP ต้องมีความสามารถเข้าถึงข้อมูลที่มาจากแหล่งที่มีโครงสร้างข้อมูลที่แตกต่างกัน และสามารถเพิ่มฟังก์ชันของสเปรตซิท

3. สามารถเข้าถึงแหล่งของการจัดเก็บข้อมูลทางกายภาพที่แตกต่างกันได้ (Accessibility) ปัจจุบันนี้มีการจัดเก็บข้อมูลในรูปแบบต่าง ๆ เช่น ฐานข้อมูลเชิงสัมพันธ์ (RDBMS) ฐานข้อมูลเชิงหลายมิติ (Multidimensional Database-MDD) ฐานข้อมูลเชิงระบบชั้น (Hierarchical Database) และฐานข้อมูลเชิงเน็ตเวิร์ค (Network Database) เป็นต้น ซึ่ง OLAP จะต้องมีความสามารถดึงข้อมูลที่มีการจัดเก็บลักษณะที่แตกต่างกันแล้วแปลงข้อมูลเหล่านั้นให้อยู่ในรูปแบบที่ OLAP ต้องการได้

4. ความรวดเร็วในการออกรายงานอย่างสม่ำเสมอ (Consistent Reporting Performance) การเพิ่มขนาดของดาต้าเบส หรือ

การออกรายงานที่แสดงมิติของข้อมูลมากขึ้นไม่มีผลต่อความเร็วของการออกรายงาน

5. สถาปัตยกรรมแบบไคลเอ็นท์-เซิร์ฟเวอร์ (Client-Server Achitecture) ปัจจุบันข้อมูลส่วนใหญ่จัดเก็บในระบบเมนเฟรม (Main-frame System) และมีการดึงข้อมูลผ่านเครื่องพีซี (Personal Computer) ฉะนั้น OLAP จะต้องสามารถทำงานภายใต้แวดล้อมของไคลเอ็นท์เซิร์ฟเวอร์

6. สามารถสร้างมิติ (Generic Dimensionality) OLAP สามารถเลือกจำนวนมิติที่ต้องการแสดงผลได้และต้องไม่เกี่ยวข้องกับโครงสร้างข้อมูลสูตรใช้คำนวณ หรือรูปแบบของการรายงาน

7. สามารถใช้บางมิติมาประมวลผล (Dynamic Sparse Matrix Handling) เดิมทีเครื่องมือในการวิเคราะห์ข้อมูลได้เพียงมิติของข้อมูล แต่ OLAP สามารถเลือกบางมิติของข้อมูลมาประมวลผลได้

8. สนับสนุนผู้ใช้หลายคน (Multi-User Support) OLAP สนับสนุนให้ผู้ใช้หลายคนสามารถค้นคืนข้อมูลหรือเปลี่ยนแปลงข้อมูล โดย OLAP จะมีระบบรักษาความปลอดภัยและรักษาความถูกต้องของข้อมูล

9. สามารถคำนวณข้ามมิติข้อมูล (Unrestricted Cross-Dimension Operations) ทุกรูปแบบของการคำนวณต้องสามารถคำนวณข้ามได้ทุกมิติของข้อมูล ทำให้ได้ผลลัพธ์ที่หลากหลายมากยิ่งขึ้น

10. การกระทำต่อข้อมูล (Intuitive Data Manipulation) การกระทำต่อข้อมูลจะกระทำโดยตรงกับเซลล์ควรจะใช้เมาส์เพื่อความสะดวก แต่ Dr. Codd ไม่ได้กล่าวโดยตรงว่าต้องใช้เมาส์

11. สามารถออกรายงานได้หลายรูปแบบ (Flexible Reporting) OLAP ต้องยืดหยุ่นในการออกรายงาน ซึ่งผู้ใช้สามารถดูการวิเคราะห์ข้อมูลในมิติใดก็ได้ตามความต้องการ

12. ต้องไม่จำกัดในเรื่องของมิติ (Unlimited Dimensions or Aggregation Levels) ยังไม่มีผลิตภัณฑ์ใดที่เข้ากฎเกณฑ์ข้อนี้ ซึ่ง Dr. Codd ได้แนะนำว่าอย่างน้อยต้องมี 15 มิติ และทางที่ดีควรจะเป็น 20 มิติ ของรูปแบบการวิเคราะห์

ทุกองค์กรธุรกิจต้องมีการวางแผนและการวิเคราะห์ข้อมูลในอดีตและปัจจุบันทั้งข้อมูลที่ใช้ในองค์กรและนอกองค์กร เพื่อช่วยในการตัดสินใจ การดำเนินงานด้านธุรกิจ ดังนั้นองค์กรธุรกิจจึงรู้จัก OLAP มีความสามารถช่วยในการวิเคราะห์ข้อมูลทั้งในอดีตและปัจจุบัน โดยนำเสนอที่ง่ายต่อผู้บริหารเข้าใจ เช่น ในรูปของตารางหลายมิติ กราฟ และสามารถเรียกดูเพียงบางส่วน (Slice and Dice) หรือต้องการดูรายละเอียดของข้อมูล (Drill-Down)

ผลิตภัณฑ์ทางด้าน OLAP ควรจะมีคุณสมบัติดังที่ E.F. Codd ได้ให้กฎเกณฑ์การประเมินไว้ทั้ง 12 ข้อ ซึ่งอาจจะมีหลายท่านได้วิจารณ์ว่าบางกฎเกณฑ์ของ E.F. Codd เป็นไปไดยากก็ตาม นอกจากนั้นแล้วจะต้องพิจารณาว่าจะใช้ OLAP ประเภทใดตั้งฐานข้อมูล เช่น ถ้าต้องการดึงข้อมูลจาก Data Warehoure ก็ต้องใช้ ROLAP แต่ต้องการดึงข้อมูลจากฐานข้อมูลหลายมิติต้องใช้ MOLAP หรือต้องการดึงข้อมูลทั้งสองประเภทต้องใช้ HOLAP ❖❖

บรรณานุกรม

Barguin, Ramon C. and Herbert A. Edelstein,ed. **Planning and designing the data warehouse.** New Jersey:Prentice Hall PTR, 1997.

Hammergren, Tom. **Data warehousing:building the corporate knowledge base.** London:International Thomson Computer Press, 1996.

Singh, Harry. **Data warehousing:concepts, technologies, implementations and management.** New Jersey:Prentice Hall PTR, 1998.

Tanler, Richard. **The intranet data warehouse:tools and techniques for building an intranet-enabled data warehouse.** New York : John Wiley & Sons, Inc. 1997.