

สถาปัตยกรรม OLAP

(On-line Analytical Processing)

สมชาย เล็กเจริญ*

ภาพ ปัจจุบันสถาปัตยกรรม OLAP ที่มีอยู่ในตลาดมีอยู่หลายชนิด อาทิ ROLAP HOLAP MOLAP และ DOLAP จึงทำให้เกิดความสับสนว่าแต่ละชนิดมีการจัดเก็บข้อมูลที่เกิดการประมวลผลของ OLAP และมีหลักเกณฑ์อย่างไรในการเลือกใช้ OLAP แต่ขั้นตอนแรกต้องพิจารณาลักษณะข้อมูลที่ OLAP ใช้ในการประมวลผล

ขั้นตอนพิจารณาข้อมูล

ข้อมูลที่ใช้ใน OLAP ได้มาจากแหล่งอื่นๆ ฉะนั้นจำเป็นต้องเก็บข้อมูลเหล่านั้นแยกไว้ต่างหากที่รู้จักดีคือ Data Warehouse และ Data Mart การที่ต้องมีการเก็บข้อมูลเข้าซ้อนด้วยเหตุผลดังนี้

1. ความรวดเร็วในการประมวลผล

ข้อมูลจากหลายๆ แหล่งมาจัดเก็บรวมกันทำให้ปริมาณข้อมูลมากยิ่งขึ้นไปเรื่อยๆ จึงจำเป็นต้องมีแหล่งจัดเก็บแยกไว้ต่างหาก เมื่อ OLAP ต้องการข้อมูลก็จะทำการดึงข้อมูลมาใช้ ทำให้รวดเร็วในการประมวลผล

2. ข้อมูลจากหลายๆ แหล่ง

OLAP ต้องการข้อมูลจากแหล่งต่างๆ รวมทั้งข้อมูลจากภายนอกและจากเครื่องพีซี ซึ่งข้อมูลเหล่านั้น มีระบบการจัดเก็บที่แตกต่างกัน เช่น ซอฟต์แวร์ที่จัดเก็บข้อมูลที่แตกต่างกัน ฉะนั้นจำเป็นต้องเปลี่ยนให้มาอยู่ในระบบเดียวกัน ก่อนเพื่อลดความซับซ้อนของข้อมูลลง

3. จัดข้อผิดพลาดของข้อมูล (clean-sing data)

ปกติการประมวลผลของระบบงานมักจะมีข้อผิดพลาดของข้อมูลรวมอยู่ด้วย ฉะนั้นจึงจำเป็นต้องจัดข้อผิดพลาดของข้อมูลนั้นเสียก่อน เช่น ฟิวด์ที่เป็นทางเลือก (optional field) มักจะเป็นสาเหตุเล็กๆ น้อยๆ ทำให้เขียนโปรแกรมผิดพลาดได้

4. การปรับเปลี่ยนข้อมูลให้เหมาะสม

มีเหตุผลหลายประการที่จำเป็นต้องมีการปรับเปลี่ยนข้อมูลให้เหมาะสมก่อนที่จะนำมาวิเคราะห์แต่ต้องไม่มีผลกระทบกับการประมวลผลระบบ transaction ตัวอย่างสำหรับเหตุผลที่ต้องปรับเปลี่ยนข้อมูลให้เหมาะสม เช่น การบริหารและการปฏิบัติการของบริษัทเปลี่ยนแปลงหรือต้องการข้อมูลที่ไม่มีในข้อมูลการทำงานประจำวัน (operational data) เช่น การโฆษณาทางทีวี หรือลักษณะทางประชากร เป็นต้น

5. ความทันสมัยของข้อมูล

ข้อมูลมาจากหลายแหล่ง ฉะนั้นแต่ละแหล่งของข้อมูล ก็จะมีการปรับปรุงข้อมูลให้ทันสมัยในช่วงเวลาที่ต่างกัน เช่น ข้อมูลจากแหล่งที่หนึ่งมีการปรับปรุงข้อมูลทุกสิ้นเดือน ข้อมูลแหล่งที่สองมีการปรับปรุงข้อมูลทุกสิ้นสัปดาห์ ฉะนั้นเพื่อให้ OLAP สามารถวิเคราะห์ข้อมูลได้อย่างถูกต้องและเที่ยงตรง จึงจำเป็นต้องจัดเก็บข้อมูลเหล่านั้นไว้ใน data warehouse

* อาจารย์ประจำภาควิชาการจัดการสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยรังสิต

6. ข้อมูลในอดีต

ส่วนใหญ่แล้วการประมวลผลของ OLAP จะมีมิติของเวลาเข้ามาเกี่ยวข้องด้วยโดยใช้สถิติ time series analysis มาวิเคราะห์ ฉะนั้นจำเป็นต้องมีการเก็บข้อมูลที่ผ่านมาในอดีตไว้ใน data warehouse

7. สรุปข้อมูล

โดยปกติข้อมูลการทำงานประจำวันต้องการรายละเอียดมาก แต่ข้อมูลสำหรับการตัดสินใจต้องการเพียงการสรุปผลเท่านั้น

8. การปรับปรุงข้อมูล

ข้อมูลทำงานประจำวัน (operational data) ต้องมีการปรับปรุงให้ทันสมัยตลอดเวลา แต่ OLAP ต้องการข้อมูลที่ไม่เปลี่ยนแปลงตลอดเวลาเพราะจะทำให้ผลการวิเคราะห์ข้อมูลผิดพลาด จึงจำเป็นต้องมีที่จัดเก็บข้อมูลจากงานประจำวัน

ลักษณะการจัดเก็บข้อมูลที่ OLAP สามารถเรียกใช้

รูปแบบการจัดเก็บข้อมูลที่ทรงประสิทธิภาพมีอยู่ 3 ชนิดที่ OLAP สามารถเข้าถึงข้อมูลได้

1. **ฐานข้อมูลเชิงสัมพันธ์ (Relational Database)** โดยทั่วไปแล้วระบบงานประจำมีการจัดเก็บข้อมูลแบบ RDBMS และ data warehouse ก็มีการจัดเก็บข้อมูลแบบ RDBMS ซึ่งอาจจะมีโครงสร้างข้อมูลแบบ Star schema และอาจจะเป็นได้ทั้ง normalized และ denormalized ก็ได้

2. **ฐานข้อมูลหลายมิติ (Multidimensional Database)**

ข้อมูลที่จัดเก็บในฐานข้อมูลหลายมิติอาจมาจากฐานข้อมูลเชิงสัมพันธ์ หรือระบบงานปัจจุบันโดยจะแปลงการจัดเก็บข้อมูลเสียใหม่โดยมีโครงสร้างการจัดเก็บแบบ array โดยทั่วไปแล้วฐานข้อมูลหลายมิติ จะยอมให้สิทธิการเขียน

ข้อมูลลงในฐานข้อมูล ในช่วงเวลาหนึ่งเพียงคนเดียว แต่อนุญาตให้หลายคน เข้าค้นหาข้อมูลในเวลาเดียวกัน หรือมิฉะนั้นก็อนุญาตให้ค้นหาข้อมูลเพียงอย่างเดียว

3. เก็บข้อมูลไว้ที่ client ในลักษณะของ file (Client-base files)

ในการยินยอมให้ client ดึงข้อมูลจำนวนไม่มากนักมาเก็บไว้ซึ่งเหมาะกับการประมวลผลแบบกระจายหรือการสร้างคำสั่งให้ข้อมูลปรากฏบน Web

สรุปได้ว่า ความสามารถในการจัดเก็บข้อมูลที่มีขนาดใหญ่ คือ ฐานข้อมูลเชิงสัมพันธ์ รองลงมา คือฐานข้อมูลหลายมิติ ส่วนการเก็บข้อมูลไว้ที่ client นั้นบรรจุข้อมูลได้น้อยที่สุด

ชนิดของ OLAP

1. **MOLAP (Multidimensional OLAP)** เป็นซอฟต์แวร์ที่ใช้ดึงข้อมูลจากฐานข้อมูลหลายมิติ (Multidimensional Database) ซึ่งมีประสิทธิภาพการทำงานได้รวดเร็วกว่า ROLAP แต่เหมาะกับขนาดข้อมูลจำนวนไม่มากนัก เช่น Data Mart

2. **ROLAP (Relational OLAP)** เป็นซอฟต์แวร์ที่สามารถดึงข้อมูลจาก data warehouse ซึ่งใช้แนวคิดการจัดเก็บข้อมูลแบบฐานข้อมูลเชิงสัมพันธ์ ฉะนั้น ROLAP จึงเหมาะกับข้อมูลขนาดใหญ่ และมีกลไกในการค้นข้อมูลที่ มีประสิทธิภาพโดยสามารถใช้คำสั่ง SQL หลายคำสั่ง (multiple SQL Statement) มาสร้างผลลัพธ์ออกมาในลักษณะของมิติข้อมูลได้ ทั้งทฤษฎีและปฏิบัติจะเห็น ROLAP มีความสามารถในการวิเคราะห์ข้อมูลทางธุรกิจได้ดีกว่า MOLAP โดยเฉพาะอย่างยิ่งการเรียกดูข้อมูลเพียงบางส่วน (Slicing and dicing) จะทำได้ดีกว่า MOLAP

3. **DOLAP (Desktop OLAP)** จากประสิทธิภาพของ PC (Personal Computer)

ผู้ผลิตภัณฑ์ประเภท OLAP พบว่าเครื่องคอมพิวเตอร์ประเภท Client จะเข้ามาเกี่ยวข้องมากขึ้น ดังนั้นจึงผลิตซอฟต์แวร์ DOLAP ทำงานบนเครื่อง client สามารถคำนวณข้อมูลทางธุรกิจและนำเสนอออกมาหลายมิติ และสามารถนำเสนอผล แสดงบน Web ได้

4. **HOLAP (Hybrid OLAP)** ผู้ผลิต OLAP พยายามให้ซอฟต์แวร์ของตนมีประสิทธิภาพสามารถเข้าถึงข้อมูลที่จัดเก็บไว้ในฐานข้อมูลเชิงสัมพันธ์ และฐานข้อมูลหลายมิติ ซึ่งเรียก OLAP ประเภทนี้ว่า Hybrid OLAP เช่น Oracle Express

ตารางแสดงสถาปัตยกรรม OLAP

เป็นตารางแสดงความสัมพันธ์ระหว่าง ลักษณะการจัดเก็บข้อมูล เช่น RDBMS, Multidimensional database server และ client file กับกลไกในการค้นคืนข้อมูล เช่น SQL, multidimensional server engine และ client engine เมื่อรวมกันแล้วจะได้ 9 ช่อง ซึ่งบางช่องแสดงถึงตัวอย่างชื่อผลิตภัณฑ์ และชนิดของ OLAP

Multidimensional processing options	Multidimensional data storage options		
	RDBMS	Multidimensional database server	client files
SQL	1 - Micro Strategy DSS Agent		
Multidimensional server engine	2 - IBM DB ₂ OLAP Server - IA DecisionSuite - Oracle Express (ROLAP model) - Seagate Holos (ROLAP mode)	4 - SAS CFO Vision - Commander Decision - Arbor Essbase - Oracle Express - Gentia - Seagate Holos - Pilot Analysis Server - Applix TM1	
Client engine	3 - Informix MetaCube - PowerPlay (Server option)	5 - Commander FDC - Dimensional Insight - Hyperion Enterprise - Hyperion Pillar	6 - BrioQuery - Cognos PowerPlay - Business Objects - Personal Express - TM1 Perspectives

● ผลิตภัณฑ์ที่ใช้ ROLAP หมายเลข 1, 2, 3

● ผลิตภัณฑ์ที่ใช้ DOLAP หมายเลข 6

● ผลิตภัณฑ์ที่ใช้ MOLAP หมายเลข 4, 5

● ผลิตภัณฑ์ที่ใช้ HOLAP หมายเลข 2, 4

หลักพิจารณาเลือก OLAP

จากที่ได้กล่าวมาแล้วจะเห็นว่า OLAP มีสถาปัตยกรรมแตกต่างกันไปซึ่งแต่ละชนิดของ OLAP มีทั้งจุดอ่อนและจุดแข็ง ฉะนั้นจึงมีหลักเกณฑ์ในการเลือกสถาปัตยกรรม ดังนี้

1. เปรียบเทียบประสิทธิภาพของสถาปัตยกรรม OLAP

OLAP ต่างชนิดกันจะมีประสิทธิภาพการทำงานแตกต่างกัน ฉะนั้นจำเป็นต้องเลือกให้เหมาะสมกับงาน ซึ่งจะยกข้อเด่นของฐานข้อมูลหลายมิติซึ่งใช้กับ MOLAP และ ROLAP มาให้พิจารณา

1.1 ลักษณะเด่นของฐานข้อมูลหลายมิติ (Mullidimension Database -- MDD)

- สามารถปรับปรุงเปลี่ยนแปลงข้อมูลในฐานข้อมูลได้ในส่วนของ ROLAP นั้นใช้โครงสร้างแบบ star จึงเหมาะสมกับการนำข้อมูลมาประมวลผลเพียงอย่างเดียว และไม่เหมาะกับการปรับปรุงข้อมูล

- MOLAP จะมีฟังก์ชันเกี่ยวกับงานธุรกิจ เช่น ค่าเสื่อมราคา อัตราการแลกเปลี่ยนเงิน อัตราดอกเบี้ย ส่วน ROLAP ไม่มีฟังก์ชันเหล่านี้

- MOLAP อนุญาตให้คำนวณข้ามมิติของข้อมูลหรือเซลล์ได้ (cross-dimensional calculations) แต่ ROLAP ไม่สามารถทำได้

- MOLAP อนุญาตให้ผู้ใช้สามารถเขียนฟังก์ชันเพิ่มเติมได้ (user-defined functions) ส่วน ROLAP ไม่มีในส่วนนี้

1.2 ลักษณะเด่นของ ROLAP

- ROLAP สามารถดึงข้อมูลจาก data warehouse มาวิเคราะห์ได้ ซึ่ง data warehouse สามารถจัดเก็บข้อมูลขนาดใหญ่ได้ ส่วน MOLAP ใช้กับฐานข้อมูลขนาดเล็ก เช่น Data Mart

- ROLAP สามารถเปลี่ยนแปลงในการสร้างมิติได้อย่างรวดเร็ว เช่น มีการเปลี่ยนแปลงรหัสสินค้าลูกค้า หรือเจ้าของในส่วนของ MDD ไม่สามารถทำได้

- ROLAP สามารถแสดงข้อมูลในลักษณะมิติได้ไม่จำกัด (Fluid Dimensionality) แต่ MOLAP มีข้อจำกัดด้านนี้

- ปัจจุบันนี้ส่วนมากข้อมูลจะถูกจัดเก็บแบบ RDBMS ฉะนั้นจึงเหมาะสมกับการใช้ ROLAP

2. พิจารณาถึงความเหมาะสมกับความต้องการ

ปัจจุบันมีหลายบริษัทที่พยายามแข่งขันกันเพื่อพัฒนา OLAP ให้มีความสามารถแตกต่างกันเพื่อเป็นกลยุทธ์ในการขายอย่างหนึ่ง ฉะนั้นผู้ใช้งานจะต้องพิจารณาความเหมาะสมและความต้องการในระบบงานของตน ดังนั้นจึงขอยกตัวอย่างบางปัจจัยที่ต้องพิจารณา ดังนี้

2.1 การนำ OLAP มาพัฒนาใช้กับงานนั้น อาจต้องพิจารณาความเหมาะสมที่ว่าถ้าต้องการพัฒนาระบบงานที่ไม่ซับซ้อนมาก ไม่ควรเลือก OLAP ประเภทที่มีฟังก์ชันการใช้งานมาก ๆ เพราะต้องเสียเวลามากในการศึกษาซอฟต์แวร์

2.2 ด้านอินเตอร์เน็ต OLAP บางตัวยอมให้เข้าถึงข้อมูลได้โดยผ่าน Web Browser ฉะนั้นต้องพิจารณาว่าท่านต้องการ OLAP ที่มีความสามารถดังกล่าวหรือไม่

2.3 การติดต่อกับซอฟต์แวร์อื่น Meta Cube ของ Informix มี OLE2 เพื่อสำหรับติดต่อกับซอฟต์แวร์อื่น

2.4 สามารถใช้งานได้นบนเครือข่าย (Network) ต้องพิจารณาว่า OLAP ที่ซื้อมันสามารถทำงานในระบบ client server ได้หรือไม่ และต้องพิจารณาความจำเป็นทั้งปัจจุบันและอนาคตด้วย

3. ความสามารถติดต่อกับฐานข้อมูลขนาดเล็กและใหญ่ (Scalability)

ROLAP มีขีดความสามารถที่เกี่ยวข้องกับฐานข้อมูลขนาดใหญ่ได้ อย่างเช่น data warehouse

4. พิจารณาในเรื่องอนาคต

ต้องนึกอยู่ในใจเสมอว่าเมื่อท่านเลือกซื้อ OLAP ตัวใดตัวหนึ่ง มันจะอยู่กับท่านไปอีกนาน ฉะนั้นต้องพิจารณาให้รอบคอบก่อนว่า

- ผลิตภัณฑ์นั้นมีคุณภาพอะไรบ้าง
- ผลิตภัณฑ์นั้นมีคุณภาพใกล้เคียงกับมาตรฐานหรือไม่

- ผู้ผลิตซอฟต์แวร์ตัวนั้นยังคงพัฒนาหรือปรับปรุงคุณภาพต่อไปหรือไม่

สถาปัตยกรรมของ OLAP ที่มีจำหน่ายปัจจุบันนั้นมีหลายชนิด เช่น ROLAP MOLAP HOLAP และ DOLAP ซึ่งผู้ใช้จะต้องทราบว่าหน่วยงานของตนเก็บข้อมูลในลักษณะใดจึงจะสามารถเลือกซื้อ OLAP ได้อย่างถูกต้องและต้องพิจารณาถึงความสามารถของ OLAP แต่ละชนิดว่าเหมาะสมกับงานธุรกิจของตนเองหรือไม่ตลอดจนต้องติดตามความก้าวหน้าของ OLAP เพื่อจะได้มาปรับปรุงคุณภาพของ OLAP ที่มีอยู่ในปัจจุบันให้มีประสิทธิภาพมากยิ่งขึ้นไปเรื่อยๆ

บรรณานุกรม

- Barquin, Ramon C. **Planning and Designing the data warehouse.**
New Jersey : Prentice Hall PTR, 1997
- Business Intelligence Ltd. **OLAP architectures**
- Microstrategy, Inc. **Two architecture for on-line analytical Processing**
(<http://www.strategy.com/>)