



การพัฒนารายงานเพื่อเพิ่มศักยภาพทางด้านระบบฐานข้อมูล

Report Development for Business Intelligence

นายศตวรรษ เหล่าอากาศ

โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีไทย – ญี่ปุ่น

พ.ศ. 2554

การพัฒนารายงานเพื่อเพิ่มศักยภาพทางด้านระบบฐานข้อมูล
Report Development for Business Intelligence

นายศตวรรษ เหล่าอากาศ

โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีไทย – ญี่ปุ่น
พ.ศ. 2554

คณะกรรมการสอบ

.....ประธานกรรมการสอบ
(อาจารย์ ปรีวัตร คงกำเนิด)

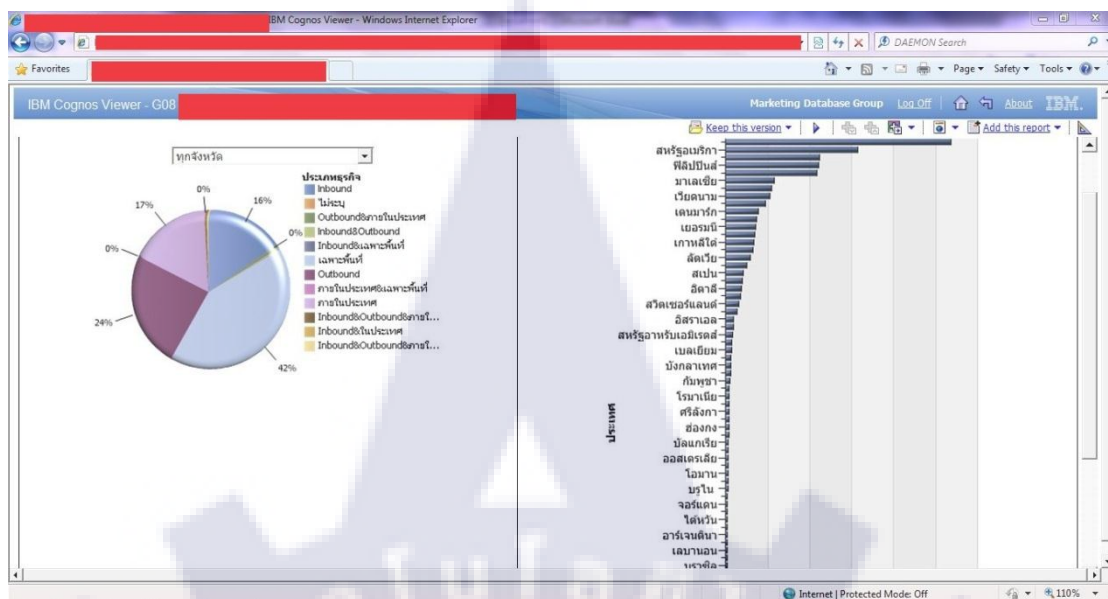
.....กรรมการสอบและอาจารย์ที่ปรึกษา
(อาจารย์ วรวิทย์ จิตขจรวานิช)

.....กรรมการ
(อาจารย์ ดร. วิมล แสนอุ้ม)

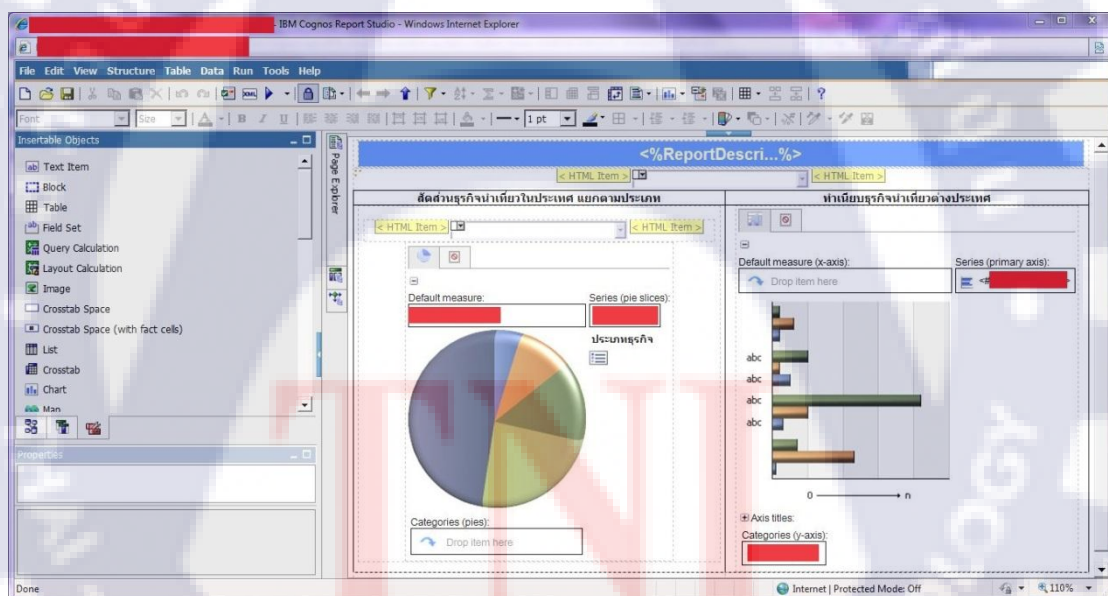
ลิขสิทธิ์ของสถาบันเทคโนโลยีไทย – ญี่ปุ่น

บทสรุป

ชื่อโครงการ	การพัฒนารายงานเพื่อเพิ่มศักยภาพทางด้านระบบฐานข้อมูล Report Development for Business Intelligence	
ผู้เขียน	นาย ศตวรรษ เหล่าอากาศกุล	
คณะวิชา	วิศวกรรมศาสตร์ สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อ.วรวิทย์ จิตขจรวานิช	
พนักงานที่ปรึกษา	นาย ดิณณภพ ภูพิจิตร	
ชื่อบริษัท	บริษัท เอไอเอส จำกัด	
ประเภทธุรกิจ/สินค้า	การให้บริการโฮสต์ติ้ง และ บริการระบบไอทีด้วยผลิตภัณฑ์ของออราเคิล (Oracle)	
งานที่ปฏิบัติ	1.ทำรายงานเพื่อเพิ่มศักยภาพให้ระบบฐานข้อมูลขององค์กรรัฐแห่งหนึ่ง โดยใช้ IBM Cognos Report Studio 2.จัดทำ User Manual ให้แก่องค์กรเพื่อประกอบการใช้งานระบบ	
ผลที่ได้รับจากการดำเนินงานและประโยชน์ที่ได้รับ	ได้รายงานในลักษณะต่างๆตามที่ลูกค้าต้องการ ทำให้ลูกค้าสามารถนำรายงานไปประกอบในการตัดสินใจในด้านต่างๆได้ และทำให้ความเป็นระบบระเบียบของฐานข้อมูลขององค์กรเป็นไปในทิศทางที่เข้าใจง่ายขึ้น ลดความซับซ้อน ยุ่งยาก ให้แก่ระบบ	



รูปตัวอย่างรายงาน



รูปตัวอย่างการสร้างรายงาน

สารบัญ

หน้า

บทคัดย่อ

ข

สารบัญ

ง

รายการตาราง

ช

รายการรูปประกอบ

ซ

บทที่

1. บทนำ

1

1.1 ชื่อและที่ตั้งของสถานประกอบการ

1

1.2 ลักษณะธุรกิจของสถานประกอบการหรือการให้บริการหลักขององค์กร

2

1.3 รูปแบบการจัดองค์กรและการบริหารองค์กร

3

1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย

3

1.5 พนักงานที่ปรึกษา และ ตำแหน่งของพนักงานที่ปรึกษา

3

1.6 ระยะเวลาที่ปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย

3

1.7 วัตถุประสงค์หรือจุดมุ่งหมายของการปฏิบัติงานหรือ

โครงการที่ได้รับมอบหมายให้ปฏิบัติงานสหกิจศึกษา

4

1.8 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย

4

1.8.1 Technical Skill

4

1.8.2 Soft skill

4

2. ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

5

2.1 ทฤษฎีSoftware

5

สารบัญ(ต่อ)

หน้า

บทที่

2. ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน (ต่อ)

2.2 Business Intelligence (BI) 6

2.3 IBM Cognos 10

3. แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน

17

3.1 แผนงานปฏิบัติงาน 17

3.2 รายละเอียดงานที่ได้รับมอบหมายให้ปฏิบัติ 18

3.3 ขั้นตอนการดำเนินงาน 18

3.3.1 ศึกษา Tools IBM Cognos 18

3.3.2 สร้าง/ปรับปรุง รายงานให้ลูกค้า 18

3.3.3 จัดทำ User Manual 18

3.3.4 ทดสอบและแก้ไขรายงาน 19

4. ผลการดำเนินงาน การวิเคราะห์และสรุปผลต่างๆ

20

4.1 สรุปผลการดำเนินงาน 20

4.2 ผลการวิเคราะห์ข้อมูล 20

4.2.1 ผลจากการสร้างรายงาน 20

4.2.2 ผลจากการทำ User Manual 21

5. บทสรุปและข้อเสนอแนะ

22

5.1 บทสรุป 22

5.2 สรุปผลการปฏิบัติงาน 22

5.2.1 สรุปผลการศึกษา BI และ IBM Cognos Report Studio 22

สารบัญ(ต่อ)

หน้า

บทที่

5. บทสรุปและข้อเสนอแนะ (ต่อ)

5.2.2 สรุปผลการสร้างรายงานจาก IBM Cognos Report Studio 22

5.3 ปัญหาและอุปสรรค 23

5.4 ข้อเสนอแนะ 23

เอกสารอ้างอิง

24

ภาคผนวก

25

ก. ความรู้เพิ่มเติม

25

ประวัติผู้วิจัย

40

TNI

NICHI INSTITUTE OF TECHNOLOGY

รายการตาราง

ตาราง

หน้า

3.1 ตารางปฏิบัติงานที่ได้รับมอบหมาย

17



รายการรูป

รูป	หน้า
1.1 แผนที่ตั้งบริษัท เอไอเอส จำกัด	1
1.2 แผนผังองค์กรบริษัทเอไอเอส	2
2.1 ภาพแสดงกระบวนการของ OLTP	5
2.2 กระบวนการทำงานของระบบ BI	8
2.3 ส่วนประกอบของ Report studio	11
2.4 รูปแบบรายงาน	11
2.5 หน้าต่างเข้า Log in ทาง IE	12
2.6 หน้าแรกของ Cognos	12
2.7 ส่วนของ Data Item	13
2.8 ส่วนของ Toolbox	13
2.9 รูปแบบประเภทต่างของ Chart	14
2.10 ลักษณะ Chart ต่างๆ	14
2.11 ส่วน Properties ของ List	15
2.12 ส่วน Properties ของ Crosstab	15
2.13 ส่วน Properties ของ Map	16
2.14 ส่วน Properties ของ Text , Date และ Time	16
4.1 หน้าแรกของ Report Studio	20
4.2 ตัวอย่างรายงานที่ได้ทำการสร้าง	21
ก.1 แสดงสัดส่วนการจัดลำดับความ สำคัญภายในองค์กรพบว่า Data Integration เป็นสิ่งที่องค์กรให้ความสำคัญเป็นหลัก จากการสำรวจ 672 องค์กร	26
ก.2 แสดงอุปสรรคที่มีต่อการทำ Data Integration จากผลสำรวจของ 672 องค์กรทั่วโลก	27

รายการรูป

รูป	หน้า
ก.3 แสดงสัดส่วนของงบประมาณสำหรับ โครงการ Data Integration	27
ก.4 แสดงการจัดระดับ 4 ระดับของ Enterprise Business Integration	29
ก.5 แสดงส่วนประกอบและแนวทางในการ ทำ Data Integration	30
ก.6 แสดงเทคนิคการรวบรวมข้อมูล (Data Integration Techniques: consolidation, federation and propagation)	32
ก.7 แสดงการทำงานในลักษณะ Push และ Pull ของ Data Consolidation	32
ก.8 แสดงกระบวนการทำงานของระบบ ETL	34

TNI

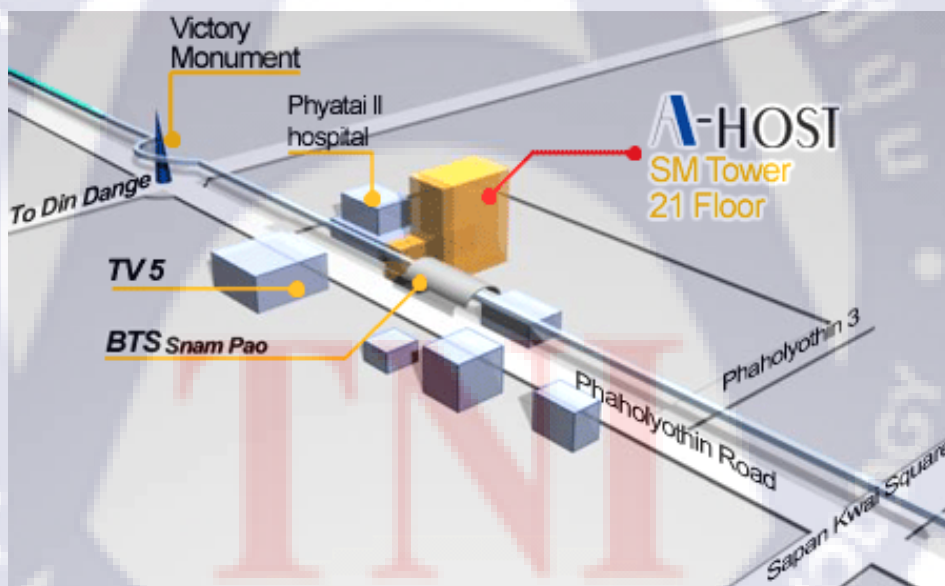
NICHI INSTITUTE OF TECHNOLOGY

บทที่ 1

บทนำ

1.1 ชื่อและที่ตั้งของสถานประกอบการ

ชื่อสถานประกอบการ	:	บริษัท เอโฮสต์ จำกัด
ที่ตั้งสถานประกอบการ	:	975/53-55 ชั้น 21 ตึก SM Tower ถนน พหลโยธิน แขวง สามเสน ใน เขตพญาไทย กรุงเทพฯ 10400
โทรศัพท์	:	(66)2298-0625-32
โทรสาร	:	(66)2298-0053



รูปที่ 1.1 แผนที่ตั้งบริษัท เอโฮสต์ จำกัด

1.2 ลักษณะธุรกิจของสถานประกอบการหรือการให้บริการหลักขององค์กร

เอไอเอสก่อตั้งบริษัทเมื่อปี 2542 ในฐานะหนึ่งบริษัทในเครือของบริษัท เมโทร ซิสเต็มส์ คอร์ปอเรชั่น (มหาชน) จำกัด และเป็นผู้เชี่ยวชาญด้านบริการจัดวางระบบไอที และบริการเสริมต่างๆ สำหรับลูกค้าตั้งแต่ธุรกิจขนาดย่อมไปจนถึงขนาดกลาง

ธุรกิจหลักของบริษัท เอ-ไอเอส คือ การให้บริการโฮสติ้ง และบริการระบบไอทีด้วยผลิตภัณฑ์ของออราเคิล (Oracle) ซึ่งเป็นซอฟต์แวร์สำหรับการวางแผนบริหารทรัพยากรขององค์กร (ERP) ระดับแนวหน้าของโลก

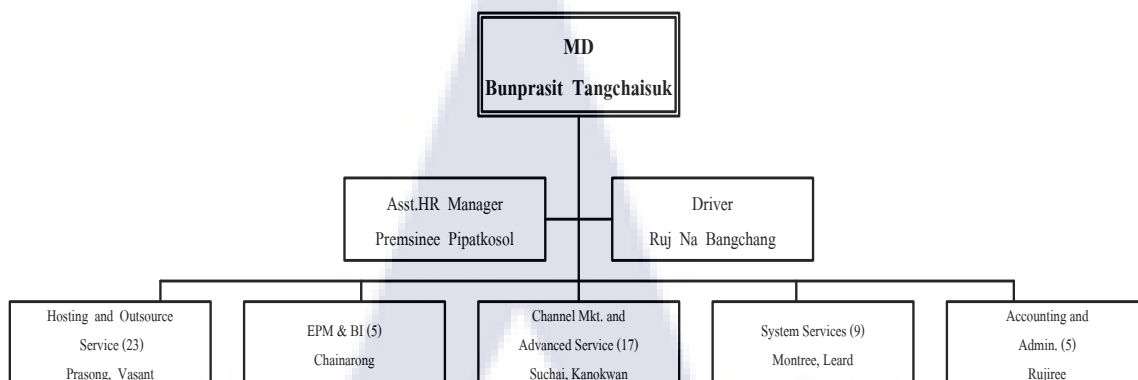
เอไอเอสถือกำเนิดขึ้นจากกลุ่มผู้เชี่ยวชาญด้านไอทีท่ามกลางภาวะเศรษฐกิจตกต่ำทั่วภูมิภาคเอเชียแต่เอ-ไอเอสก็สามารถเติบโตได้อย่างรวดเร็ว และมั่นคงตั้งแต่แรกก่อตั้ง ด้วยจุดแข็งในฐานะผู้บุกเบิกธุรกิจโฮสติ้งเซอร์วิส พร้อมทั้งนำธุรกิจแนวใหม่อย่างการให้บริการแอปพลิเคชัน หรือ ASP (Application Services Providing) เข้ามาให้บริการแก่องค์กรธุรกิจเป็นรายแรกในเมืองไทย

ธุรกิจการให้บริการแอปพลิเคชัน ในรูปแบบ ASP ของเอ-ไอเอสไม่เพียงแต่ให้บริการด้านแอปพลิเคชันด้านการดำเนินธุรกิจทางอิเล็กทรอนิกส์ระดับโลกของออราเคิลพร้อมโครงสร้างพื้นฐานทางเทคโนโลยีสารสนเทศเท่านั้น แต่ยังมีบริการที่ครอบคลุมตั้งแต่การให้คำปรึกษา การสนับสนุนและการให้บริการทั่วไปอย่างพร้อมครบบครัน รวมทั้งยังมีความยืดหยุ่นสูงมีการปรับเปลี่ยนบริการและทรัพยากรให้เหมาะสมกับความต้องการ และสภาพงานที่แตกต่างกันของลูกค้าแต่ละราย

ในการดำเนินธุรกิจของเอ-ไอเอสตลอดระยะเวลา 10 ปี ไม่เพียงแต่ในฐานะผู้บุกเบิกธุรกิจโฮสติ้งและธุรกิจการให้บริการแอปพลิเคชันในรูปแบบ ASP เท่านั้น แต่เอ-ไอเอสยังได้ทำการติดตั้งระบบไอที รวมทั้งผลิตภัณฑ์ของออราเคิลให้กับลูกค้าจนประสบความสำเร็จมาแล้วเป็นจำนวนมาก ซึ่งหลายรายเป็นหนึ่งในร้อยบริษัทชั้นนำของประเทศไทย แต่ที่สำคัญกว่านั้นก็คือการที่ เอ-ไอเอสได้สานสัมพันธ์กับลูกค้า และพันธมิตรทางธุรกิจอย่างแนบแน่นจนกลายเป็นหุ้นส่วนทางกลยุทธ์ และเป็นผู้สนับสนุนสำคัญที่มีส่วนช่วยผลักดันให้ธุรกิจของลูกค้าเติบโตสู่ความสำเร็จ

ปัจจุบันเอ-ไอเอสเป็นหนึ่งในบริษัทลูกของบริษัท ดราโก้วัน จำกัด ซึ่งเป็นบริษัทมหาชนที่อยู่ในตลาดหลักทรัพย์แห่งประเทศไทย

1.3 รูปแบบการจัดองค์กรและการบริหารองค์กร



รูปที่ 1.2 แผนผังองค์กรบริษัทเอไอเอส

1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย

ตำแหน่ง : Application Support / Programmer
 หน้าที่งานที่ได้รับมอบหมาย : ช่วยเหลือในการทำ Report Cognos, ETL ข้อมูล, Mapping ข้อมูล

1.5 พนักงานที่ปรึกษา และ ตำแหน่งของพนักงานที่ปรึกษา

พนักงานที่ปรึกษา : นาย ดิณณภพ ภูพิจิตร
 ตำแหน่ง : Assistant Consultant

1.6 ระยะเวลาที่ปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย

เริ่มต้นปฏิบัติงานสหกิจศึกษา วันที่ 1 มิถุนายน พ.ศ. 2554
 สิ้นสุดงานสหกิจศึกษา วันที่ 30 กันยายน พ.ศ. 2554

1.7 วัตถุประสงค์หรือจุดมุ่งหมายของการปฏิบัติงานหรือโครงการที่ได้รับมอบหมายให้ปฏิบัติงานสหกิจศึกษา

- เพื่อให้ลูกค้าสามารถนำรายงานไปใช้ในการตัดสินใจได้
- เพื่อให้ระบบฐานข้อมูลขององค์กรลูกค้า มีความเป็นระบบระเบียบ

1.8 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย

แบ่งเป็น 2 กลุ่มดังนี้

1.8.1 Technical Skill

- สามารถรู้ถึงโครงสร้าง การใช้งานของโปรแกรม IBM Cognos Reports Builder, IBM Cognos Framework Manager, Oracle Warehouse Builder, SQLdeveloper และองค์ประกอบต่างๆที่ จะต้องใช้
- สามารถรู้ถึงความสามารถของโปรแกรมว่าทำอะไรได้บ้าง รวมถึงการที่จะนำไปประยุกต์ใช้จริงๆ

1.8.2 Soft skill

- ได้ประสบการณ์จริงในการทำงานร่วมกับคนอื่นๆ
- การทำงานเป็นทีมซึ่งแบ่งแต่ละหน้าที่ของแต่ละคนว่าทำส่วนไหนกันบ้าง
- การจัดการเรื่องตารางเวลาของตัวเอง

บทที่ 2

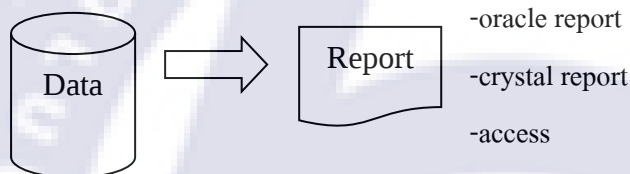
ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

2.1 ทฤษฎีSoftware

Software ทั้งโลกแบ่งออกเป็น 2 ประเภท คือ

1. **OLTP** (online Transaction Processing) คือ เป็นระบบที่สร้างขึ้นมาเพื่อรองรับงาน Operation ในแต่ละวัน เช่น บัญชีแยกประเภท สั่งซื้อส่งขาย นั่นก็อาจจะหมายความว่า OLTP คือ ข้อมูลที่มีการ Insert / Update / Delete อยู่ตลอดเวลา และอาจจะเป็นจำนวนมากด้วยใน 1 วัน

- ERP (Enterprise resource planning)
- POS (point of sale)



รูปที่ 2.1 ภาพแสดงกระบวนการของ OLTP

2. **OLAP** (online analysis processing) เป็น software ในการวิเคราะห์

- **BI** (Business Intelligence) ซอฟต์แวร์ที่นำข้อมูลที่มีอยู่เพื่อจัดทำรายงานในรูปแบบต่างๆ ที่เหมาะสมกับมุมมองในการวิเคราะห์ และตรงตามความต้องการของผู้ใช้งาน และใช้สำหรับวิเคราะห์ข้อมูล ของงานในมุมมองต่างๆ ตามแต่ละแผนก

- **Data mining** จะช่วยค้นหาแนวโน้มในอนาคต ความสัมพันธ์ของข้อมูล
- **DSS** (Decision Support System) การคำนวณโดยใช้ค่าความสำคัญของข้อมูล

*OLAP จะเกิดขึ้นได้ก็ต่อเมื่อ มี Database ดังนั้นจึงจำเป็นต้องมี OLTP ก่อน

2.2 Business Intelligence (BI)

ในยุคปัจจุบันที่เทคโนโลยีมีการเปลี่ยนแปลงอย่างรวดเร็ว และตลอดเวลา เช่นเดียวกัน ระบบธุรกิจก็มีการแข่งขันกันค่อนข้างรุนแรง และมากขึ้นด้วย จึงเป็นสิ่งที่เราหลีกเลี่ยงไม่ได้เลยว่า การที่องค์กรจะอยู่รอดได้นั้นจะต้องมีการใช้ข้อมูลสารสนเทศที่ทันสมัยและทันทั่วถึง เพื่อสนับสนุนการตัดสินใจอย่างรวดเร็วและสามารถนำไปวางแผน หรือ ได้ตอบปัญหา เชิงธุรกิจได้ทันต่อเหตุการณ์ ให้กับผู้บริหารระดับสูงขององค์กร

การที่จะได้มาซึ่งข้อมูล สารสนเทศเหล่านั้น หนึ่งจำเป็นต้องมีการแสวงหาหนทาง ในการเก็บรวบรวมข้อมูลให้ได้มาก เพราะว่าข้อมูลเหล่านั้นมิใช่ข้อมูล ภายในองค์กรเท่านั้น ซึ่งอาจจะเป็นข้อมูลขององค์กร ที่เป็นคู่แข่งหรือเป็นข้อมูลของ องค์กรอื่นๆ ที่อยู่ในธุรกิจเดียวกันกับเราก็เป็นไปได้ สองการเลือกสรรข้อมูลสารสนเทศที่มีคุณค่าจากกองข้อมูลที่มีขนาดมหึมา เพื่อให้แน่ใจว่าระบบข้อมูลสารสนเทศที่พัฒนาขึ้นมานั้นเป็นข้อมูลสารสนเทศที่ สามารถตอบสนองต่อความต้องการของผู้บริหารระดับสูงขององค์กรได้ เพื่อเอาชนะอุปสรรคเหล่านี้จึงจำเป็นต้องมีระบบที่สามารถช่วยเตรียมข้อมูลที่ลึกซึ้ง และมีคุณค่าทางกิจกรรมทางธุรกิจให้แก่องค์กรได้

ปัจจุบันการวางแผนทางกลยุทธ์ของบริษัทนั้นจำเป็นต้องใช้ข้อมูลมากมาย ซึ่งการวิเคราะห์ข้อมูล ทางด้านการตลาด การขาย การเงิน การผลิตนั้นจะต้องทันกับเหตุการณ์ซึ่งมีข้อมูลเกิดขึ้นเป็นประจำทุกวัน ดังนั้นการจัดทำรายงาน จะต้องมีการแก้ไขบ่อย และมีความยุ่งยาก

Business Intelligence คือ ซอฟต์แวร์ (Software) ที่นำข้อมูลที่มีอยู่เพื่อจัดทำรายงานในรูปแบบต่างๆ ที่เหมาะสมกับมุมมองในการวิเคราะห์ และตรงตามความต้องการของผู้ใช้งาน และใช้สำหรับวิเคราะห์ข้อมูล ของงานในมุมมองต่างๆ ตามแต่ละแผนก เช่น

- วิเคราะห์การดำเนินงานของบริษัทฯ เพื่อการตัดสินใจด้านการลงทุนสำหรับผู้บริหาร
- วิเคราะห์และวางแผนการขาย / การตลาด เพื่อประเมินช่องทางการจำหน่าย ฯลฯ
- วิเคราะห์สินค้าที่ทำกำไร สูงสุด / ขาดทุนต่ำสุด เพื่อการวางแผนงานด้านการตลาด และการผลิต
- วิเคราะห์ปัจจัยที่มีผลกระทบต่อยอดขายของสินค้า ฯลฯ

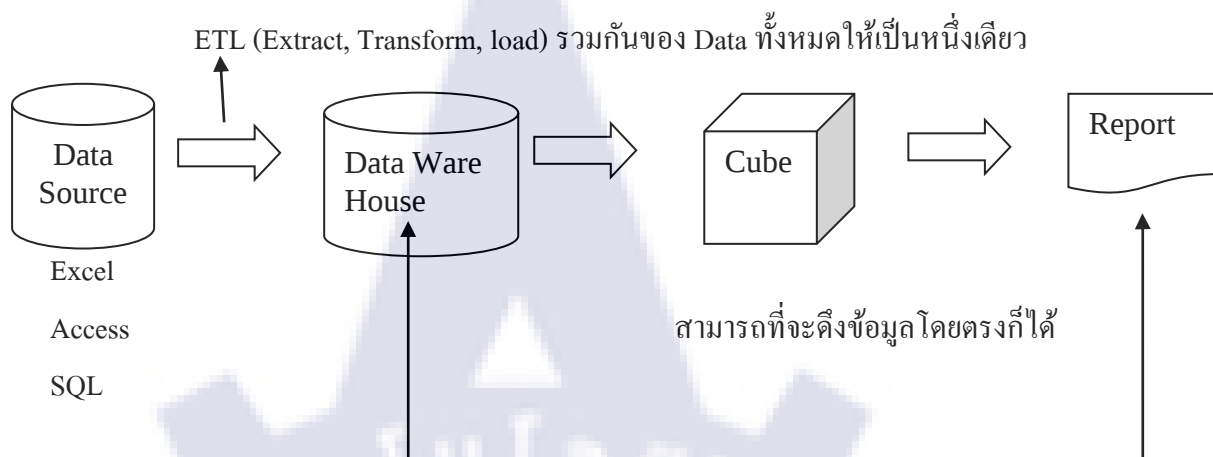
Business Intelligence จะประกอบไปด้วยระบบข้อมูล และ โปรแกรมแอปพลิเคชัน ด้านการวิเคราะห์ มากมายหลายระบบ เช่น

- **ดาต้าแวร์เฮ้าส์ (Data Warehouse)** คือฐานข้อมูลขนาดใหญ่ที่รวบรวมข้อมูลทั้งจากแหล่งข้อมูลภายในและ ภายนอกองค์กร โดยมีรูปแบบและวัตถุประสงค์ในการจัดเก็บข้อมูลซึ่งจำเป็นต้องมีการออกแบบฐาน ข้อมูลให้สอดคล้องกับการนำข้อมูลที่ต้องการนำมาใช้งาน
- **ดาต้ามาร์ท (Data Mart)** คือ คลังข้อมูลขนาดเล็กมีการเก็บข้อมูลที่มีลักษณะเฉพาะเจาะจง เช่น เก็บข้อมูลส่วนของการเงิน ส่วนของสินค้าคงคลัง ส่วนของการขาย เป็นต้น ซึ่งทำให้การจัดการข้อมูลการนำเอาข้อมูลไปสร้างความสัมพันธ์และวิเคราะห์ต่อ ก็ง่ายขึ้น
- **การทำเหมืองข้อมูล (Data Mining)**คือการนำคลังข้อมูลหลักมาประมวลผลใหม่ มาแสดงผลเฉพาะสิ่งที่สนใจโดยกระบวนการในการดึงข้อมูลออกจากฐานข้อมูลจะมี สูตรทางธุรกิจ (Business Formula)และเงื่อนไขต่างๆเข้ามาเกี่ยวข้องและผลลัพธ์ในรูปแบบที่แตกต่างกัน เช่นเป็นแผนภูมิในการตัดสินใจ (Decision Trees) เป็นต้น
- **เครื่องมือที่ใช้ในการวิเคราะห์ข้อมูลในหลายมิติ (OLAP)** คือการสืบค้นข้อมูลที่ใช้สามารถเลือกผลลัพธ์ออกมาในรูปแบบของตารางหรือ กราฟ โดยสามารถวิเคราะห์ข้อมูลใน มุมมองหลากหลายมิติ (Multi-Dimensional) โดยที่ผู้ใช้สามารถที่จะดูข้อมูลแบบเจาะลึก(Drill Down) ได้ตามต้องการ
- ระบบสืบค้นและออกรายงานต่างๆ (Search, Report)

Business Intelligence ยังมีจุดเด่นเพิ่มขึ้นอีกในด้าน

- ใช้งานง่ายเพียงแค่คลิกเมาส์ก็สามารถเปลี่ยนแปลงรายงานได้โดยไม่ต้องมีการคีย์ข้อมูลใหม่ ซึ่งผู้ใช้สามารถถาม ตอบคำถามทางธุรกิจได้หลายมุมมองเพียงในเวลาไม่กี่นาที ซึ่งช่วยการตัดสินใจแม่นยำ และรวดเร็วกว่าคู่แข่ง ทั้งในเชิงกว้าง และเชิงลึก
- สามารถดึงข้อมูลจากฐานข้อมูลที่หลากหลายภายในองค์กรมาทำการวิเคราะห์ เช่น Excel, FoxPro, Dbase, Access, ORACLE, SQL Server, Informix, Progress, DB2 เป็นต้น โดยไม่มีการเขียนโปรแกรมเพิ่มเติมใดๆ

Business Intelligence (BI)



รูปที่ 2.2 กระบวนการทำงานของระบบ BI

ETL ย่อมาจาก Extract transform and load คือการดึงข้อมูลจาก data source ต่างๆเข้าสู่ Data Warehouse แบ่งเป็น 3 ขั้นตอนหลักดังนี้

- **Extract** คือการดึงข้อมูลจากแหล่งข้อมูลที่แตกต่างกัน
- **Transform** การนำข้อมูลที่ได้อาจจากการ Extract มาจัดรูปแบบให้ถูกต้องสอดคล้องกัน เช่น Data Mapping การทำให้ข้อมูลที่มีความหมายเดียวกันแต่อยู่ในรูปแบบที่แตกต่างกันให้อยู่ในรูปแบบเดียวกัน Data cleansing การตรวจสอบและแก้ไขข้อมูลให้ถูกต้อง
- **Load** การนำข้อมูลที่ผ่านการ Transform แล้ว เข้าสู่ Data Warehouse

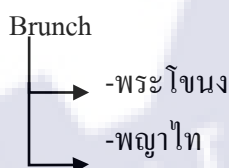
Tool ที่ใช้ในการ ETL คือ Pentaho Data Integration

DataWareHouse คือการนำ data หรือ data หลายๆที่เข้าด้วยกัน เพื่อที่จะนำมาทำระบบสนับสนุนการตัดสินใจ

Cube คือโครงสร้างข้อมูล (Data structure) รูปแบบหนึ่งเพื่อให้วิเคราะห์ข้อมูลได้เร็วขึ้น ซึ่งใช้กับ OLAP โดยส่วนใหญ่ ในการทำ OLAP นั้น ได้ใช้ Cube เป็นพื้นฐานในการวิเคราะห์ข้อมูลเป็นหลายมิติ หรือ dimensions นั้นเอง โดยส่วนใหญ่ Cube จะถูกสร้างในส่วนของ Data warehouse นั้นเอง เช่น SSAS ของ SQL Server หรือใน Business Information Warehouse ของ SAP นั้นเอง และถ้าระบบ Business intelligence นั้น จะนำ Cube ที่สร้างไปเป็นข้อมูลในการสร้างรายงานต่างๆ โดยสามารถนำ Cube มาทำเป็น Array 2 มิติ เพื่อนำเสนอข้อมูลในมิติต่างๆได้ ในคำนิยามของ ที่ได้มีคนให้ไว้กับ Cube เช่น

Cube นั้นใช้อธิบายความสัมพันธ์ต่างๆ ผ่าน Fact table กับ Dimension Table และ Cube เป็น กลุ่มของ กลุ่มตัววัดที่ตั้งแต่หนึ่งขึ้นไปโดยมีความสัมพันธ์กับ มิติ ซึ่งเราสามารถลองเล่น Cube ได้จาก Tool ต่างๆ นะครับ เช่น BW, SSAS, Essbase เป็นต้น

- Dimension มุมมอง (มีได้มากกว่า 1)
- Measure หน่วยวัด ต้องเป็นตัวเลขเท่านั้น
- Hierarchy จัดระบบตามลำดับชั้น, ลำดับชั้น เช่น



ข้อดี Cube

- เพื่อเพิ่มประสิทธิภาพให้ดีขึ้น เร็วขึ้น
- สะดวกใช้ง่าย

ข้อเสีย Cube

- ต้นทุนสูงในเรื่อง Hardware และ Software
- เปลืองเนื้อที่ในการเก็บข้อมูล

Tool ของ **Business Intelligence (BI)**

Business Intelligence Tools คือ ประเภทของโปรแกรมซึ่งใช้ในการทำออกแบบรายงาน, วิเคราะห์และนำเสนอข้อมูล โดยการนำข้อมูลมาจาก Datawarehouse หรือ Datamart ซึ่งเป็น โปรแกรมที่เริ่มมีความต้องการในองค์กรมากขึ้น เช่น

IBM	=> Cognos
Oracle	=>Hyperion, Oracle Business Intelligence Enterprise Edition(OBIEE)
Microsoft	=>SQL server OLAP
SAP	=>BOC (Business Object + Crystal report + Excelious)
SAS	=>Data mining

2.3 IBM Cognos

Cognos ปัจจุบันเป็นโปรแกรมของบริษัท IBM ที่ไปซื้อบริษัท Cognos (อยู่ในประเทศแคนาดา) มาตั้งแต่ปี 2008 ซึ่งโด่งดังในเรื่องของรายงานธุรกิจอัจฉริยะ (Business Intelligence) และระบบคลังข้อมูล (Data Warehouse) ก็เป็นโปรแกรมระดับแรกๆ ที่นิยมกันใช้ทั่วโลก

Cognos BI เป็นชุดโปรแกรมซึ่งประกอบด้วย หลายๆ โปรแกรมรวมกัน และใช้งานต่างกัน เพื่อให้ครอบคลุมประกอบของ BI ซึ่งประกอบด้วย

- **Report Studio** สำหรับใช้สร้างรายงานระดับ Professional ตัวนี้สำคัญมากและใช้มากที่สุด เพื่อสร้างรายงานให้ผู้บริหารทั้งหลายดู

- **Query Studio** สำหรับใช้ออกรายงานประเภทต้องการทันที ซึ่งจะเห็นข้อมูลบนหน้าจอเลย เหมาะสำหรับดึงข้อมูลเพื่อตอบโจทย์ผู้บริหาร ในกรณีที่ไม่ได้ทำรายงานรองรับไว้ ข้อมูลเป็นลักษณะ Listing ก็มีข้อจำกัดในการใช้งานเยอะเหมือนกัน

- **Analysis Studio** ชื่อก็บอกแล้วว่าใช้สำหรับการวิเคราะห์ ก็คล้ายๆ กับ Query Studio นั้นแหละ แต่จะปรับมุมมองได้ สลับแกนได้ จะเห็นข้อมูลทันทีเหมือน Query Studio แต่สามารถดูข้อมูลในรูปแบบ Crosstab คือเป็นตารางในแนวดิ่ง และแนวนอน

- **Metric Studio** สำหรับการทำ Balance Scorecard และ KPI (Key Performance Indicator) ซึ่งเป็นส่วนสำคัญของการใช้ประเมินผลตามแต่ผู้บริหารจะกำหนด

- **Metric Designer** ใช้สำหรับสร้างโครงสร้างของข้อมูลที่จะนำเข้า Metric Studio

- **Event Studio** ใช้สำหรับให้ระบบทำการแจ้งผู้เกี่ยวข้อง เมื่อระบบตรวจสอบและเข้าเงื่อนไขที่กำหนดไว้

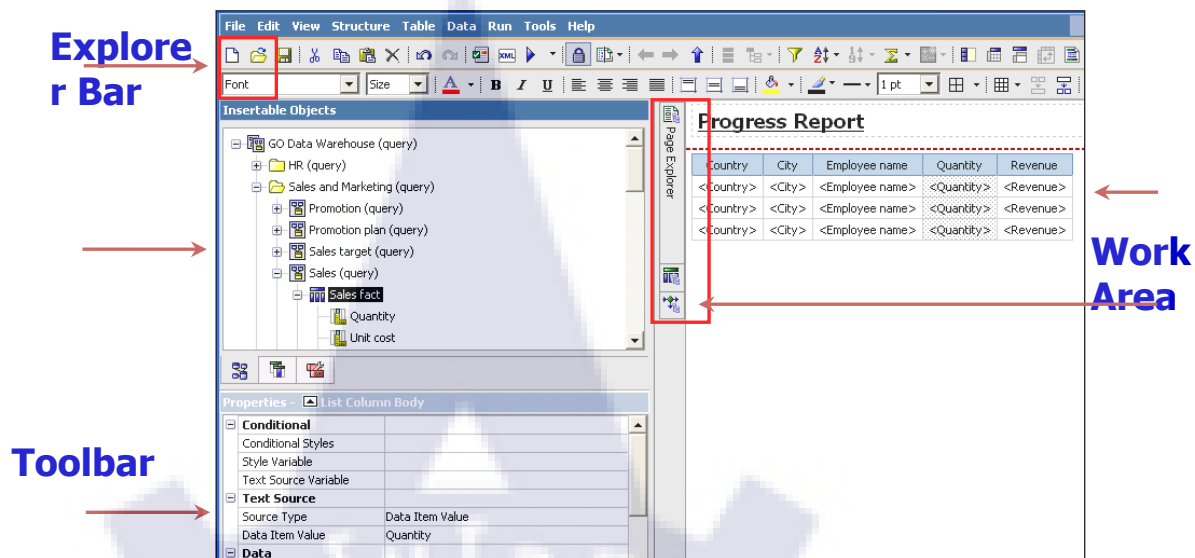
- **Framework Manager** ใช้สำหรับผูกโครงสร้างข้อมูลจากระบบต่างๆ เพื่อนำไปใช้ในการออกรายงาน ไม่ว่าจะเป็น Report/Query/Analysis Studio ถ้าผูกผิดข้อมูลที่ใช้ก็เพี้ยนไปหมด

- **PowerPlay** เป็นโปรแกรมเรียกดูข้อมูลที่เป็น PowerCube เท่านั้น ซึ่งก็เหมือนๆ กับ Analysis Studio แต่เป็น Windows Application

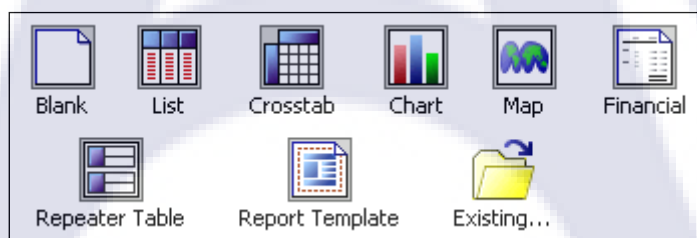
- **Transformer** เป็นโปรแกรมสร้าง PowerCube ซึ่งเมื่อนำไปออกรายงานก็จะรวดเร็ว กว่าข้อมูลรูปแบบอื่น เพราะจะถูกสร้าง Index ในมุมมองต่างๆ ไว้หมดแล้ว

- **Data Manager** เป็นโปรแกรมทางด้าน ETL (Extract Transform and Load) ใช้ดึงข้อมูลจากระบบต่างๆ เข้า Data Warehouse

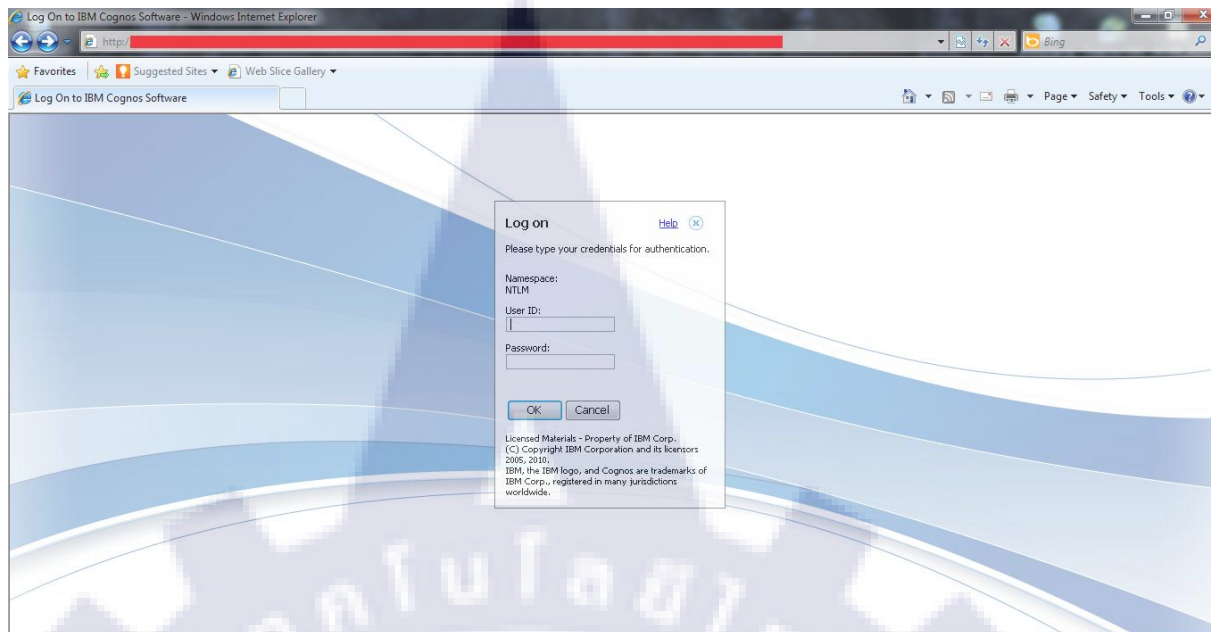
ส่วนต่างๆใน Cognos Report Studio



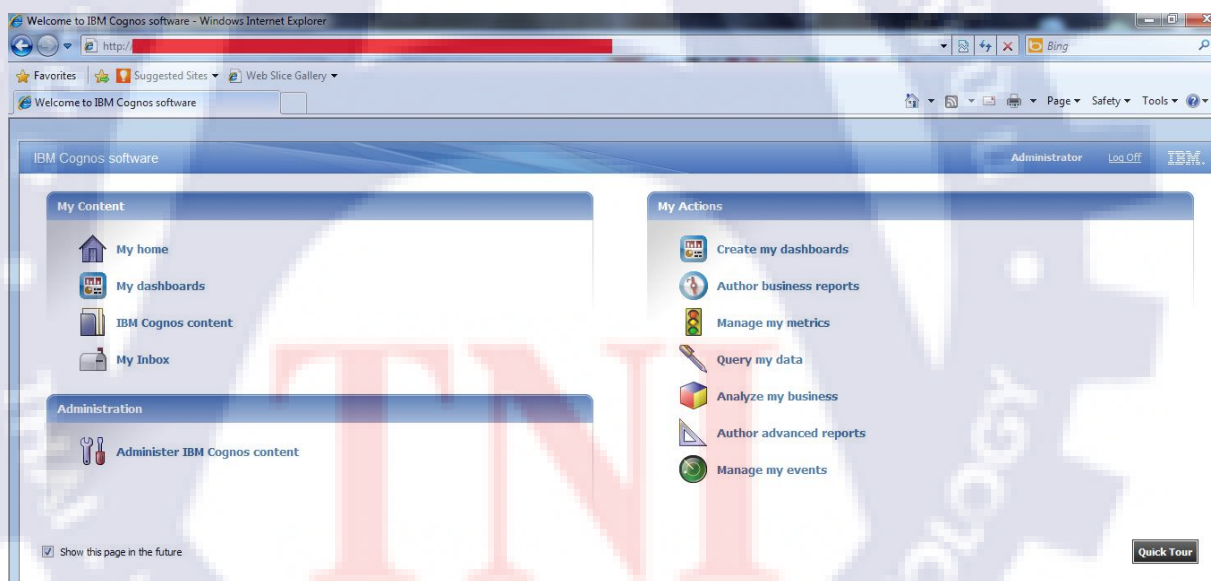
รูปที่ 2.3 ส่วนประกอบของ Report studio



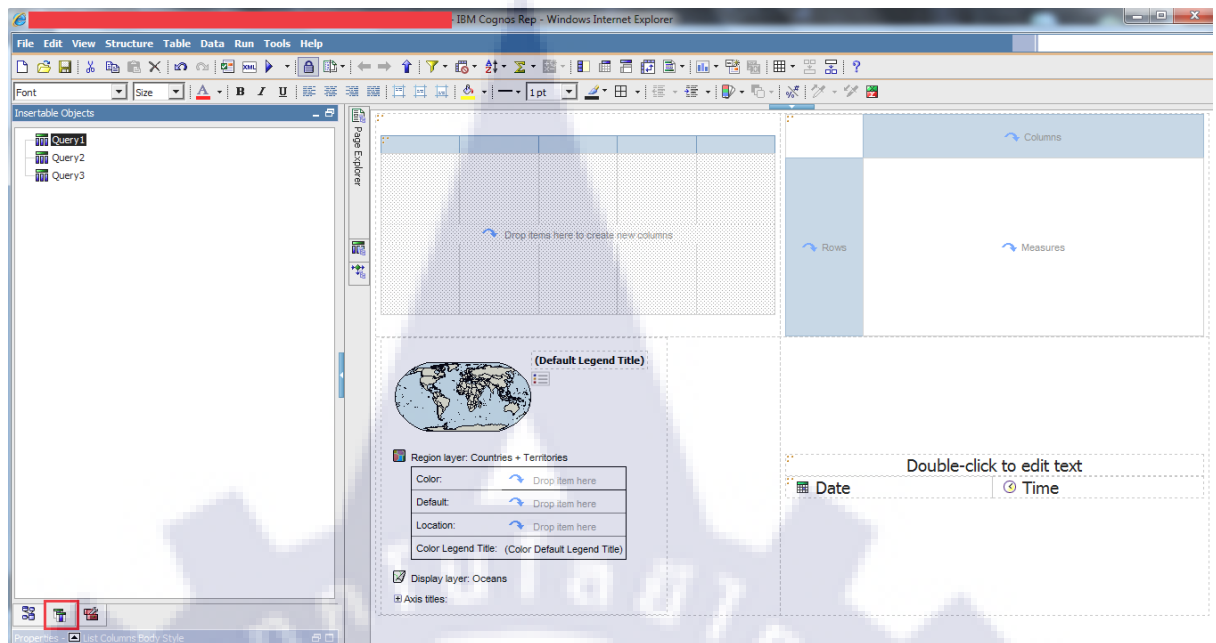
รูปที่ 2.4 รูปแบบรายงาน



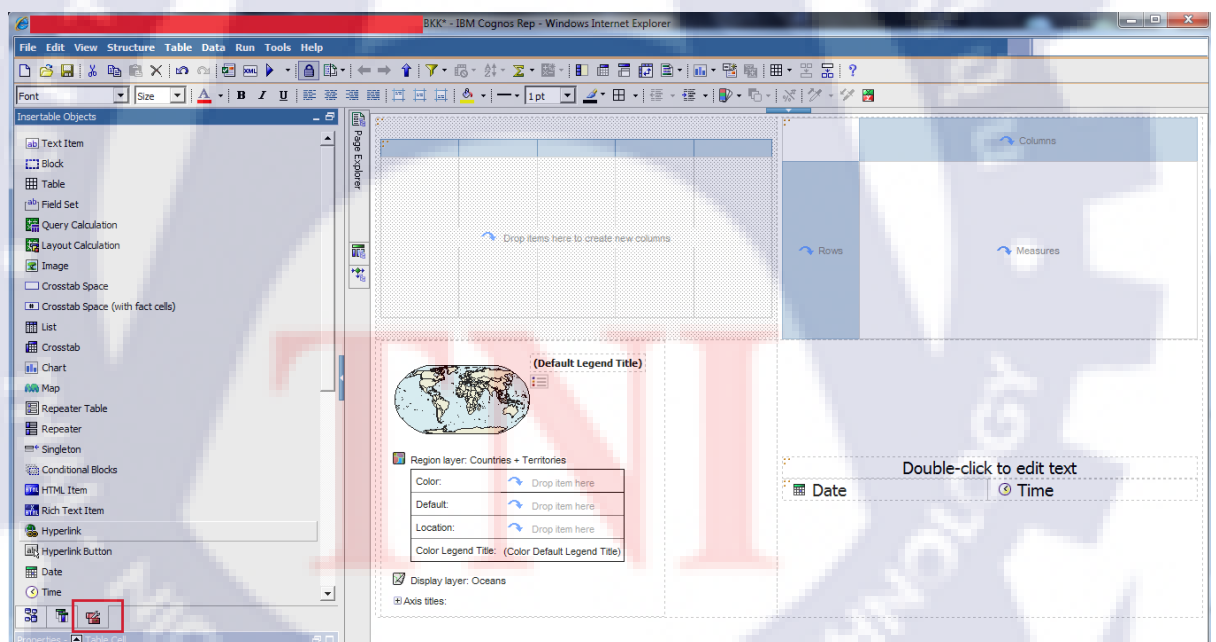
รูปที่ 2.5 หน้าต่างเข้า Log in ทาง IE



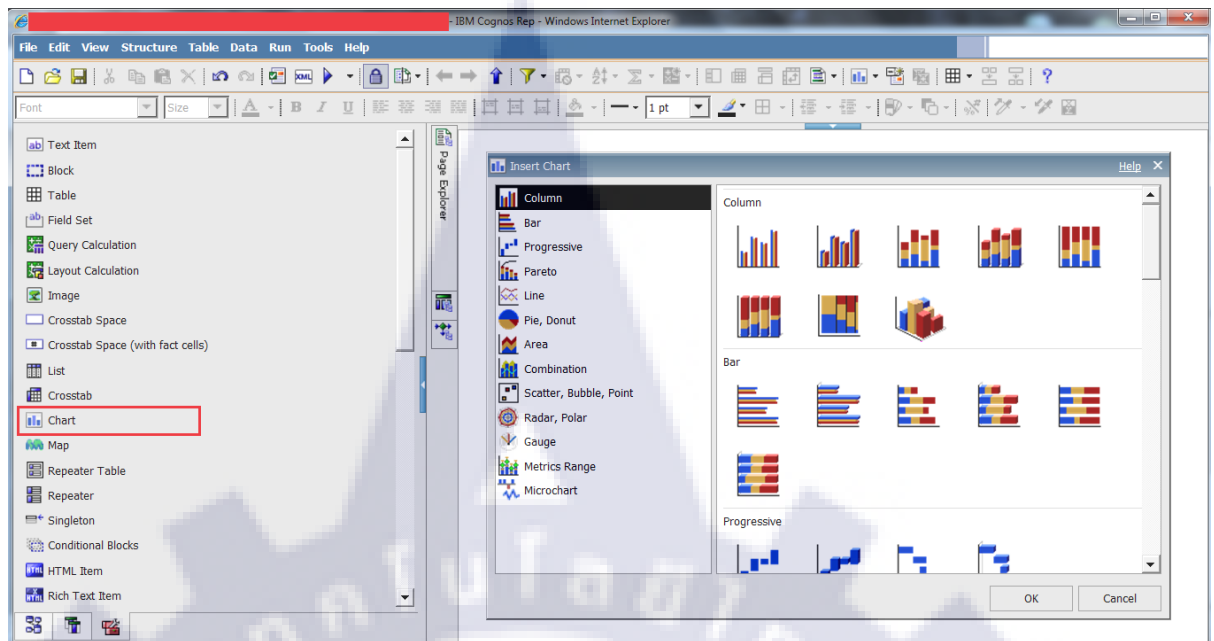
รูปที่ 2.6 หน้าแรกของ Cognos



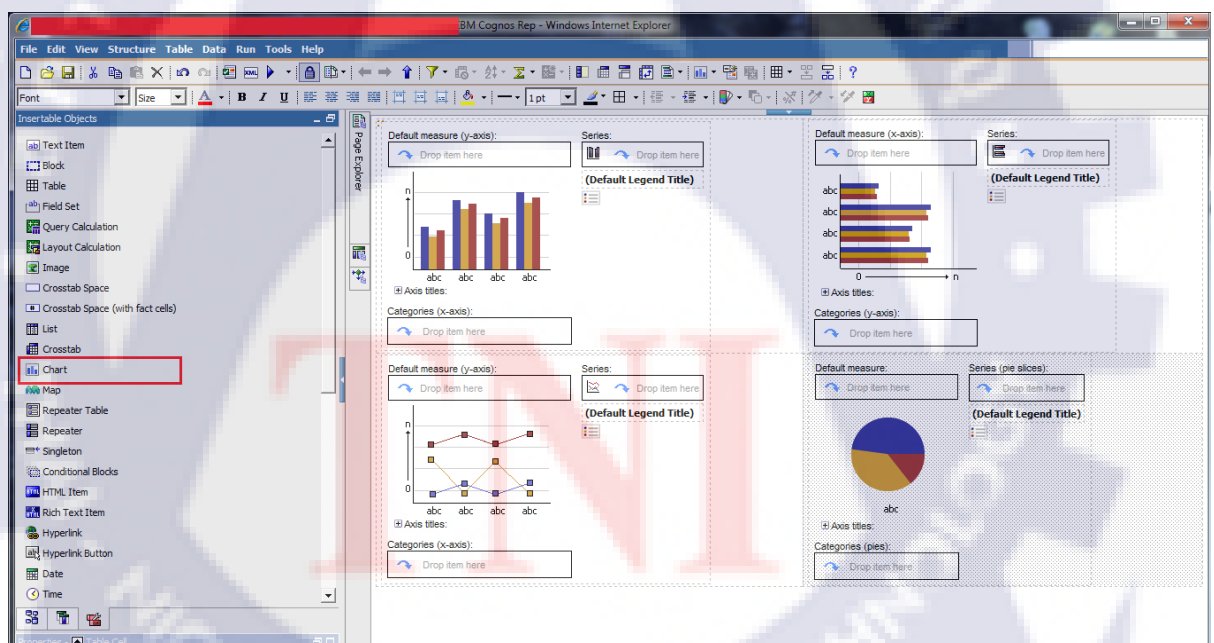
รูปที่ 2.7 ส่วนของ Data Item



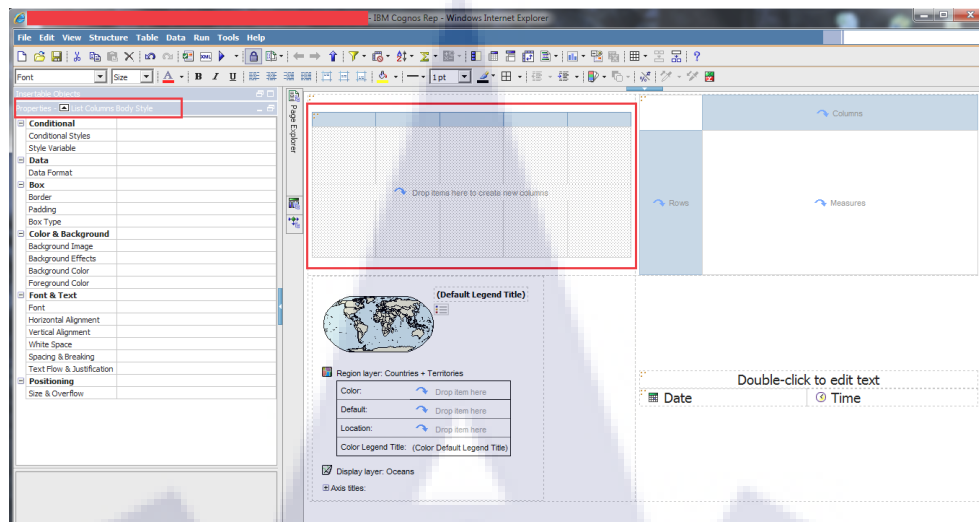
รูปที่ 2.8 ส่วนของ Toolbox



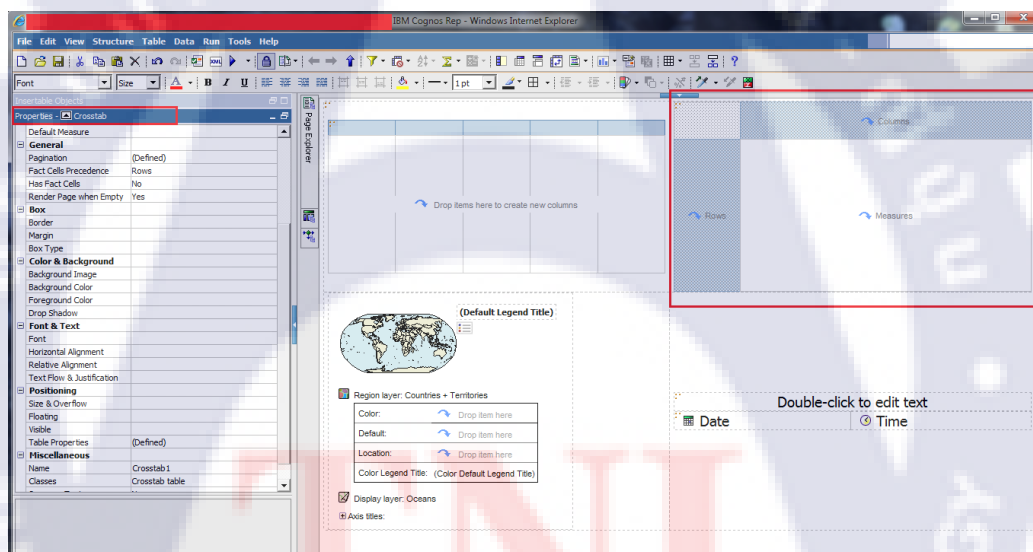
รูปที่ 2.9 รูปแบบประเภทต่างของ Chart



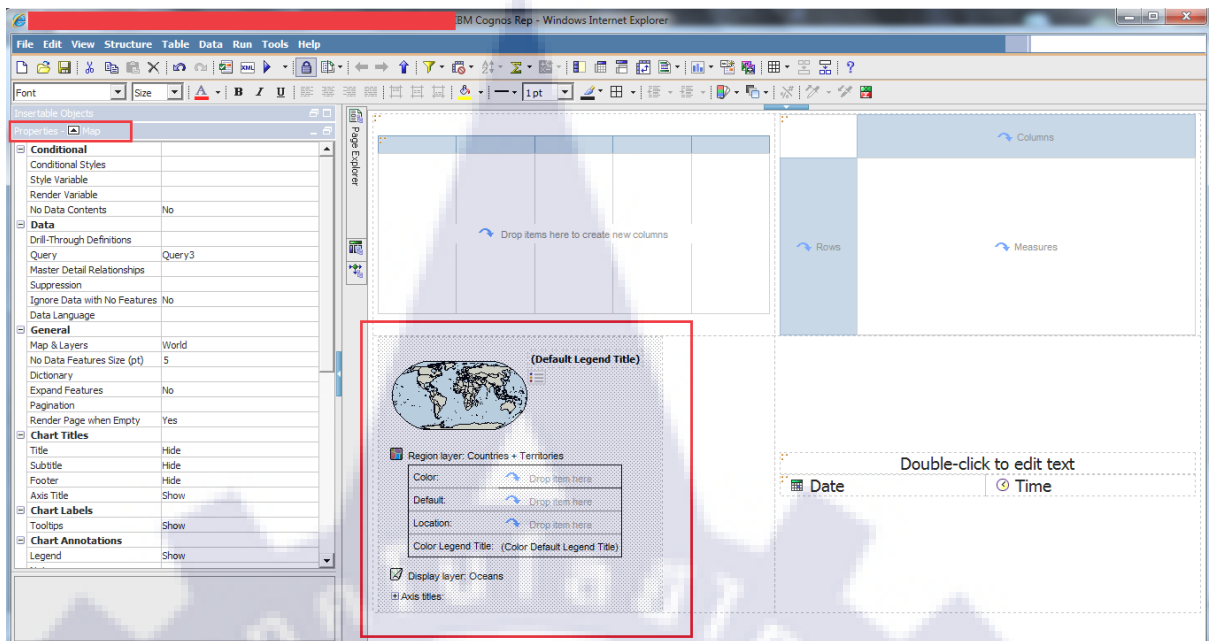
รูปที่ 2.10 ลักษณะ Chart ต่างๆ



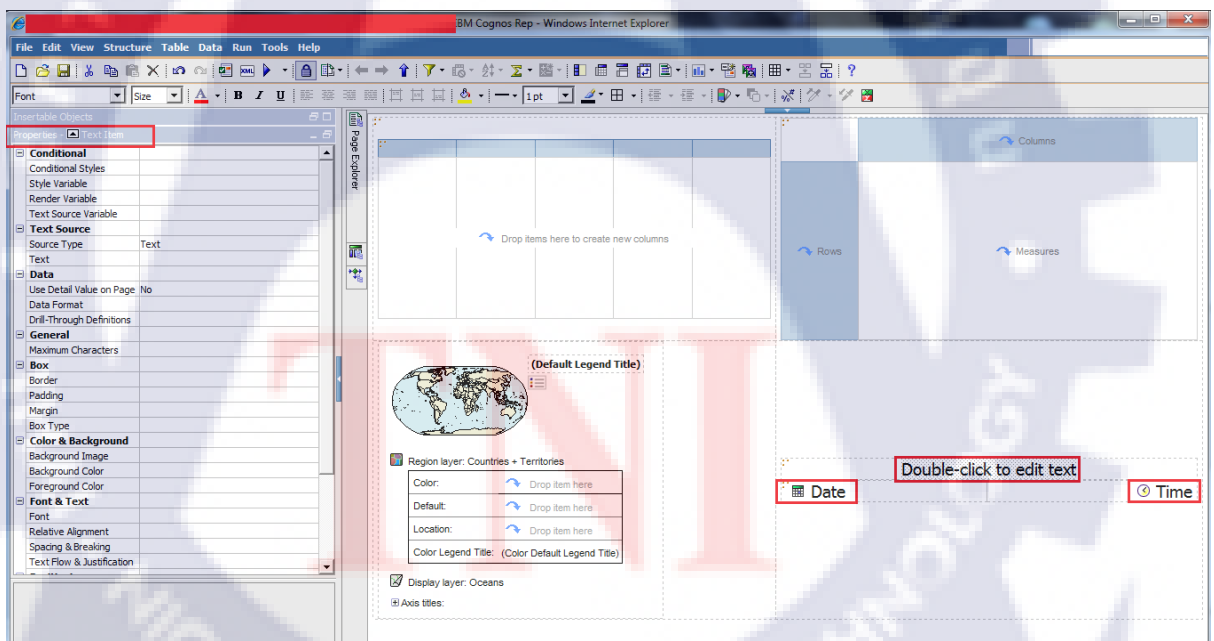
รูปที่ 2.11 ส่วน Properties ของ List



รูปที่ 2.12 ส่วน Properties ของ Crosstab



รูปที่ 2.13 ส่วน Properties ของ Map



รูปที่ 2.14 ส่วน Properties ของ Text , Date และ Time

บทที่ 3

แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน

3.1 แผนงานปฏิบัติงาน

ตารางที่ 3.1 ตารางปฏิบัติงานที่ได้รับมอบหมาย

รายละเอียด	เดือนที่ 1	เดือนที่ 2	เดือนที่ 3	เดือนที่ 4
ศึกษา Tools IBM Cognos	■			
	■			
สร้าง/ปรับปรุงรายงาน ให้ลูกค้า		■	■	
		■	■	
จัดทำ User Manual				■
				■
ทดสอบและแก้ไข รายงาน				■
				■

สีแดง หมายถึง แผนระยะเวลาปฏิบัติงานที่คาดการณ์ไว้

สีฟ้า หมายถึง ระยะเวลาปฏิบัติงานจริง

3.2 รายละเอียดงานที่ได้รับมอบหมายให้ปฏิบัติ

- ศึกษา tools IBM Cognos และทำการทดลองสร้างรายงาน
- ทำการสร้างรายงานตามความต้องการของลูกค้า มีการใช้ ภาษา SQL ควบคุมในการสร้างรายงานไปด้วย
- จัดทำ User Manual วิธีใช้งานระบบให้แก่ลูกค้า

3.3 ขั้นตอนการดำเนินงาน

3.3.1 ศึกษา Tools IBM Cognos

ทำการศึกษา IBM Cognos ในส่วนของ Report Studio โดยนำแบบฝึกหัดของทางบริษัทมาทำ โดยเริ่มจากการทำความเข้าใจ ทฤษฎีของ BI เบื้องต้น และได้ทำการใช้ Data ที่ทำการ Mock up ไว้ มาทดลองสร้างรายงาน

3.3.2 สร้าง/ปรับปรุง รายงานให้ลูกค้า

รับ ความต้องการ มาจากลูกค้า และทำการนำข้อมูลขององค์กรลูกค้ามาสร้างรายงาน ซึ่งมีพนักงานในบริษัทได้ทำการ ETL ไว้เรียบร้อยแล้ว ทำการสร้างรายงานและนำไปเสนอต่อลูกค้าเพื่อรับ Comment และนำมาปรับปรุงรายงาน

3.3.3 จัดทำ User Manual

เมื่อรายงานในระบบคงที่แล้ว ก็ทำการจัดทำเอกสาร User Manual ซึ่งเป็นคู่มือการใช้งานระบบไว้ให้ลูกค้า

3.3.4 ทดสอบและแก้ไขรายงาน

เชื่อว่าลูกค้ามีความต้องการ ใหม่ๆหรือเปล่า และทำการทดสอบรายงานในระบบ จากนั้นทำการปรับปรุงแก้ไข



บทที่ 4

ผลการดำเนินงาน การวิเคราะห์และสรุปผลต่าง ๆ

4.1 สรุปผลการดำเนินงาน

จากการปฏิบัติงานที่ได้รับมอบหมายเป็นระยะเวลา 4 เดือน ได้ข้อสรุปตามขั้นตอนดังนี้

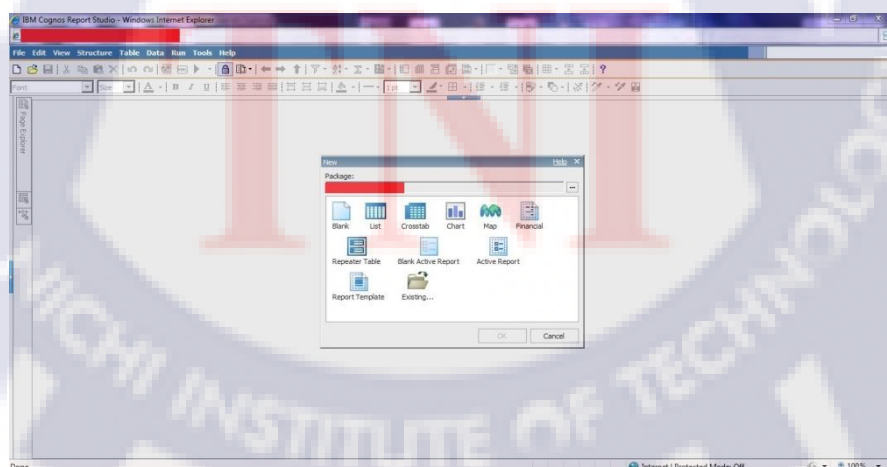
- ศึกษา Tools IBM Cognos และทำการทดลองสร้างรายงาน
- ศึกษา ความต้องการของลูกค้า และ ระบบข้อมูลของลูกค้า
- นำ ความต้องการของลูกค้า มาทำรายงาน พร้อมทั้งปรับปรุงแก้ไขรายงานตลอดเวลา
- จัดทำ User Manual ให้แก่ลูกค้า

4.2 ผลการวิเคราะห์ข้อมูล

จากการดำเนินงานได้วิเคราะห์และสรุปดังนี้

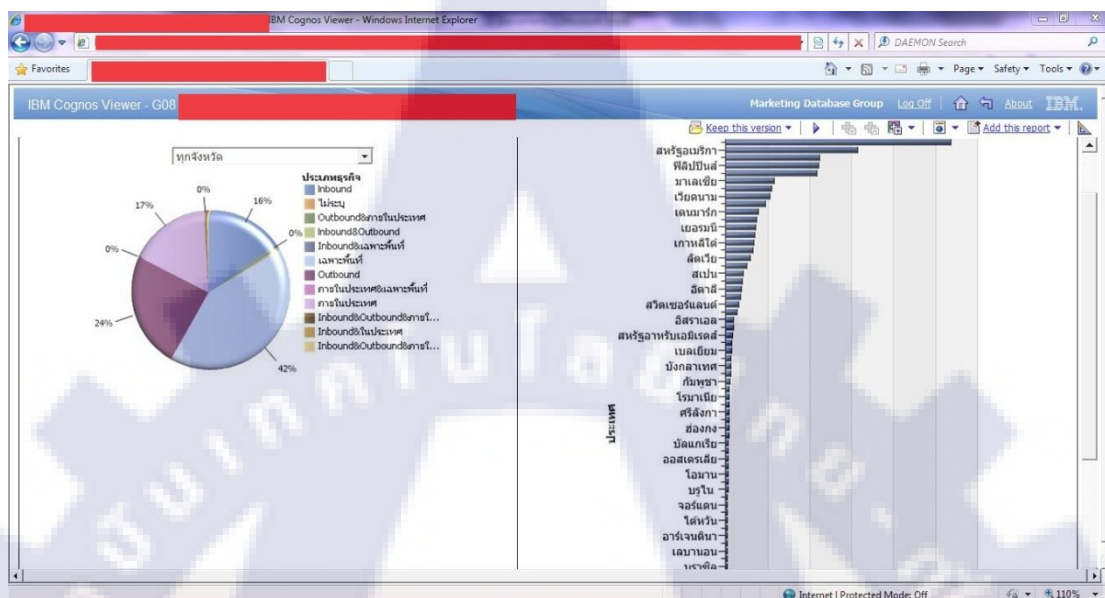
4.2.1 ผลจากการสร้างรายงาน

ในส่วนของ Tools IBM Cognos นั้น สามารถใช้สร้างรายงานให้มีลักษณะได้หลายแบบ โดยใช้ในส่วนของ Report Studio ซึ่งจะมีลักษณะหน้าต่างดังนี้



รูปที่ 4.1 หน้าแรกของ Report Studio

เมื่อทำการรับ ความต้องการ มาจากลูกค้าก็ทำการสร้างรายงาน Prototype เพื่อนำเสนอให้ลูกค้าดูก่อน แล้วรับ ความต้องการ ใหม่มาปรับปรุงตลอดเวลา รวมไปถึงการสร้างรายงานใหม่ๆ ตามความต้องการของลูกค้า



รูปที่ 4.2 ตัวอย่างรายงานที่ได้ทำการสร้าง

4.2.1 ผลจากการทำ User Manual

หลังจากที่รายงานในระบบมีความคงที่แล้ว ได้ทำการจัดทำเอกสาร User Manual หรือคู่มือประกอบการใช้งานแก่ผู้ใช้งาน โดยหลังจากที่ทำเสร็จแล้วได้มีการนำไปใช้จริง ซึ่งพบว่าลูกค้าสามารถนำไปประกอบการใช้งานได้ตรงตามเงื่อนไข

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

จากการปฏิบัติงานในการสร้างรายงานเป็นไปตามที่คาดหวังเอาไว้ ลูกค้าสามารถนำไปใช้และตรงตามความต้องการของลูกค้า รวมไปถึง User Manual ที่สามารถนำไปประกอบการใช้หรือการสอนได้อย่างมีประสิทธิภาพ ซึ่งอาจจะมีความซับซ้อนในรายงานบางตัว แต่ User Manual สามารถตอบข้อสงสัยได้อย่างครบถ้วน

5.2 สรุปผลการปฏิบัติงาน

5.2.1 สรุปผลการศึกษา BI และ IBM Cognos Report Studio

จากการที่ได้ศึกษาและทดลองทำแบบฝึกหัดแล้ว พบว่า การที่มีระบบ Business Intelligence หรือ BI จะช่วยทำให้องค์กรต่างๆ มีประสิทธิภาพในการคาดการณ์และช่วยตัดสินใจได้เป็นอย่างดี เพราะ การทำระบบ BI จะทำให้สามารถมองเห็นข้อมูลที่ทางองค์กรนั้นๆ มีอยู่เป็นรูปร่างที่สามารถจับต้องได้ ทำให้มองเห็นแนวโน้มที่คาดว่าจะเกิดในอนาคต และทำให้สามารถนำข้อมูลเหล่านี้ไปช่วยในการตัดสินใจวางแผนการตลาดได้อีกด้วย ในส่วนของ Report Studio นั้น จากการทดลองศึกษา พบว่า ผู้ใช้จำเป็นจะต้องมีความรู้เรื่องมุมมองในฐานข้อมูล หรือเรียกว่า CUBE ก่อน จึงจะทำการใช้ Tools ได้อย่างเข้าใจมากยิ่งขึ้น

5.2.1 สรุปผลการสร้างรายงานจาก IBM Cognos Report Studio

จากการสร้างรายงานโดยใช้ IBM Cognos Report Studio พบว่าสามารถทำรายงานออกมาได้หลายรูปแบบไม่ว่าจะเป็น Crosstab Table, Chart ต่างๆ ซึ่งแสดงผลออกมาได้ดีเยี่ยม และมีความสวยงาม และสามารถสร้างรายงานได้ตรงตามความต้องการของลูกค้า

5.3 ปัญหาและอุปสรรค

- จากการปฏิบัติงานที่ได้รับมอบหมาย ทำให้ต้องไปทำงานที่องค์กรของลูกค้า ทำให้มีความยุ่งยากด้านการนำอุปกรณ์เข้า หรือ ออก จากสถานที่ จำต้องขอใบอนุญาตต่างๆ รวมไปถึงการเข้าวง LAN เพื่อเข้าถึงข้อมูลอีกด้วย
- ปัญหาทางด้าน Hardware ก็เป็นอีกหนึ่งปัญหาที่ทำให้การปฏิบัติงานชะงัก เช่น เครื่องเซิร์ฟเวอร์ของทางลูกค้าเสีย, ทางองค์กรมีการทำระบบ LAN ใหม่ ทำให้ไม่สามารถใช้ระบบ LAN ได้ส่งผลให้ไม่สามารถปฏิบัติงานได้
- ทางลูกค้าขาดความเข้าใจในเรื่องระบบ BI จึงทำให้การอธิบายเป็นไปได้ยากพอสมควร ทำให้การปฏิบัติงานมีความล่าช้าในบางครั้ง

5.4 ข้อเสนอแนะ

ในการสร้างรายงานโดยใช้ IBM Cognos Report Studio นั้น ผู้ที่ทำการสร้างจำเป็นต้องมีความรู้ทางด้าน BI มาก จึงทำให้การใช้ Tools เป็นไปได้ยากสำหรับผู้ที่เป็น Admin ของทางลูกค้า จึงจำเป็นต้องมีการอบรมต่อเนื่อง รวมไปถึงการศึกษาจาก User Manual ควบคู่ไปด้วย

เอกสารอ้างอิง

ความสำคัญของ Data Integration (2553).[ออนไลน์].

เข้าถึงได้จาก : <http://kanompia.blogspot.com/2010/08/data-integration-with-etl.html>.

(วันที่สืบค้นข้อมูล : 13 มิถุนายน 2554).

IBM Cognos Data (2550).[ออนไลน์].

เข้าถึงได้จาก : <http://www-01.ibm.com/software/data/cognos/>.

(วันที่สืบค้นข้อมูล : 16 มิถุนายน 2554).

คู่มือการใช้งาน Cognos.(2553).[ออนไลน์].

เข้าถึงได้จาก : www.md.go.th/it/Manual/Manual_Cognos.pdf.

(วันที่สืบค้นข้อมูล : 20 มิถุนายน 2554).

ภาคผนวก ก.
ความรู้เพิ่มเติม

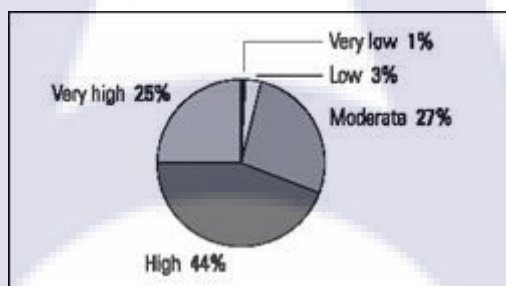
TNI

ความสำคัญของ Data Integration

จากข้อมูลของ TDWI (The Data Warehouse Institute) ในปี 2003 ซึ่งมีการสำรวจข้อมูลจากองค์กรใหญ่ในโลก 672 องค์กร เกี่ยวกับเรื่อง *Evaluating ETL and Data Integration Platforms* ซึ่งมีเป้าหมายเพื่อหาข้อมูลเกี่ยวกับ เทคนิค, เทคโนโลยี, การประยุกต์ใช้ และ เครื่องมือที่นิยมในปี 2003 ได้ผลออกมาดังต่อไปนี้

ความท้าทายของ Data Integration

การรวบรวมและประสานข้อมูลจากหลายๆ แหล่งเข้าด้วยกัน ถือการใช้งานเป็นงานที่ยากและก่อให้เกิดการทะลักหรือระเบิดของข้อมูลภายใน องค์กรหลายๆ ที่ ซึ่งงานต่างๆ เหล่านี้ถือว่าไม่ง่ายเลย จากข้อมูลการสำรวจพบว่า 69% ขององค์กร มีความเห็นว่า Data Integration จัดความสำคัญในระดับ very high หรือ high ดัง แสดงในภาพ



รูปที่ ก.1 แสดงสัดส่วนการจัดลำดับความ สำคัญภายในองค์กรพบว่า Data Integration เป็นสิ่งที่องค์กรให้ความสำคัญ หลัก จากการสำรวจ 672 องค์กร

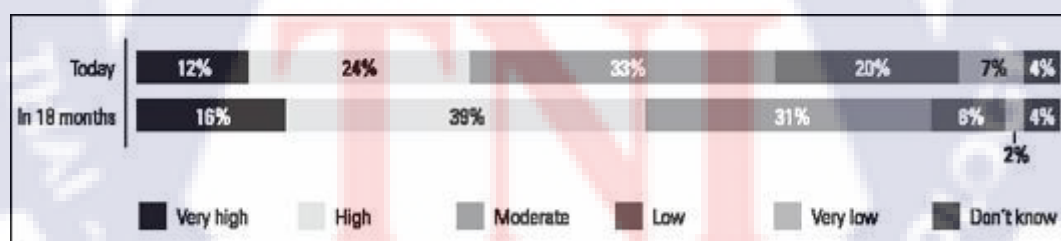
ประเด็นสำคัญที่มีผลต่อความสำเร็จ ของ Data Integration



รูปที่ ก.2 แสดงอุปสรรคที่มีต่อการทำ Data Integration จากผลสำรวจของ 672 องค์กรทั่วโลก

งบประมาณสำหรับโครงการ Data Integration

จากข้อมูลข้างต้นจะพบว่า แม้ว่าการขาดแคลนงบประมาณสนับสนุนจะมีผลอย่างมากเป็นอันดับ 2 ต่อความสำเร็จของโครงการ ผลการสำรวจแสดงให้เห็นว่าบริษัทหรือองค์กรต่างๆ กำลังเพิ่มงบประมาณอย่างต่อเนื่องเพื่อโครงการ Data Integration นี้ และ ยังพบอีกว่า 36% ของบริษัทต่างๆ จัดลำดับความสำคัญในงบประมาณของ Data Integration ให้เป็น Very high หรือ high ผลสำรวจพบว่ามีปริมาณการเพิ่มงบประมาณถึง 55% ของเรื่องนี้ภายใน 18 เดือน



รูปที่ ก.3 แสดงสัดส่วนของงบประมาณสำหรับ โครงการ Data Integration พบว่ามีการเพิ่มของบุคลากรและงบประมาณในระดับ very high และ high ถึง 55% ภายใน 18 เดือน จากการสำรวจ 672 องค์กร

จากข้อมูลเบื้องต้นทำให้เราสรุปใน ประเด็นของการจัดสรรงบประมาณได้ว่า:

องค์กรใหญ่ มีแนวโน้มที่จะมุ่งเข้าหากระบวนการในการทำ Data Integration Architecture ซึ่งบริษัทเหล่านี้มีแหล่งข้อมูลมากมายและ อยู่กระจ่ายกันหลายที่ รวมทั้งบางแหล่งข้อมูลเป็นรูปแบบเก่าแก่ บริษัทเหล่านี้มุ่งความสนใจที่จะซื้อหรือทำให้ได้มาซึ่งเครื่องมือในการรวบรวมข้อมูล (Integration) และ บริษัทเหล่านี้จ่ายเงินเพื่อซื้อเครื่องมือที่ทันสมัย ทำงานรวดเร็วและดีที่สุด ดูเหมือนว่าบริษัทเหล่านี้จะมีศูนย์กลางของการรวบรวมข้อมูลด้วย (Data Integration Competency Center)

องค์กรขนาดกลาง จะมุ่งความสนใจไปที่การรวบรวมข้อมูลจากมุมมองของ Business Intelligence มากกว่า และ มุ่งสนใจว่าองค์กรจะสามารถนำเอาข้อมูลมาใช้ได้อย่างไร เนื่องจากว่าองค์กรขนาดกลางจะมีข้อมูลเก่า (Legacy) น้อยกว่าองค์กรใหญ่ ซึ่งทำให้้องค์กรขนาดกลางมีความสนใจในเครื่องมือที่ดีและทันสมัยที่สุดน้อย กว่าองค์กรใหญ่

บทบาท และหน้าที่ของ Data Integration ในองค์กร

กระบวนการรวบรวมข้อมูลขององค์กร สามารถเกิดขึ้นได้ภายใต้ 4 ขั้นตอนในระบบไอที (Information Technology) คือ data, application, business process และ user interaction เทคโนโลยีและเครื่องมือ หลายชนิดที่มีความเหมาะสมกับงานนี้แต่เพื่อให้การจัดหมวดหมู่และจัดการง่าย ขึ้น เราจะจำแนกเครื่องมือต่างๆ ออกได้เป็น 4 ระดับดัง ต่อไปนี้

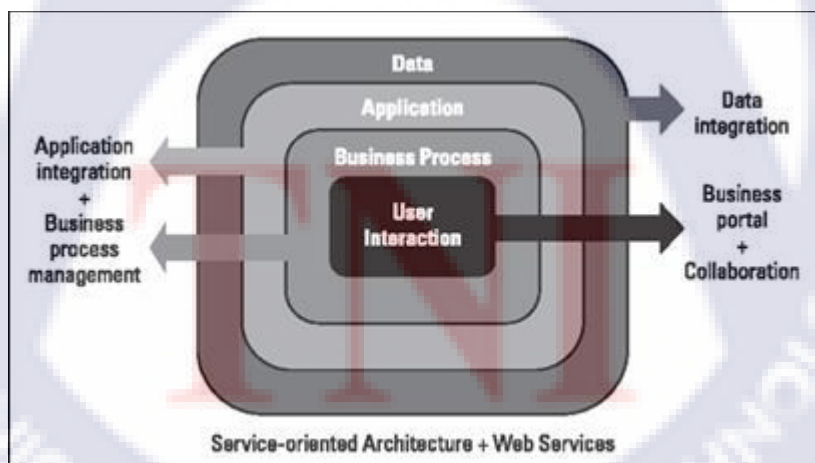
Data Integration: ให้ ประโยชน์ในการแสดงภาพรวมข้อมูลของธุรกิจ ที่กระจายอยู่ในส่วนต่างๆ ในองค์กร ภาพรวมวงกว้าง ณ จุดเดียวนี้สามารถสร้างได้โดยใช้เครื่องมือเทคนิค และ เทคโนโลยี แตกต่างชนิดกัน มันอาจจะเป็นภาพรวมทางกายภาพของข้อมูลที่ได้จากแหล่งข้อมูลหลายๆที่รวมกัน หรือ มันอาจจะเป็นข้อมูลที่มีการรวบรวมเสมือน (Virtual federate) ให้เสมือนว่าข้อมูลมีการอยู่ร่วมกัน แต่ว่าแท้จริงแล้วทางกายภาพอยู่กันคนละที่ก็ทำได้

Application Integration: เปิด ช่องทางให้เกิดภาพรวม ณ จุดเดียวของมุมมองธุรกิจที่มีต่อโปรแกรมประยุกต์ของธุรกิจ ที่ข้อมูลนั้นอาจจะอยู่ในหรืออยู่นอกองค์กรก็ได้ วิธีการทำ application integration นั้นทำได้หลากหลายวิธี หลายเทคนิค และ เครื่องมือ ขึ้นอยู่กับความต้องการ ความซับซ้อน ของโปรเจกนั้นๆ

Business Process Integration: เปิดโอกาสให้เกิดภาพรวม ณ จุดเดียวขององค์กรที่มีต่อกระบวนการทางธุรกิจ(Business Process) เครื่องมือที่ช่วยในการออกแบบกระบวนการทางธุรกิจจะช่วยให้ นักพัฒนาสามารถ วิเคราะห์, สร้างต้นแบบ, และ แบบจำลองเสมือนจริงของกระบวนการทางธุรกิจ และ กิจกรรมที่เกี่ยวข้องได้ ประโยชน์อีกประการสำคัญของกระบวนการรวบรวมงานทางธุรกิจคือ กระบวนการต่างๆที่สร้างขึ้นมามีเครื่องมือเหล่านี้อิงกับกระบวนการทาง ภาพภาพเดิมที่มีอยู่ในองค์กรอยู่แล้ว

User Interactive Integration: เปิดโอกาสให้ผู้ใช้งานขั้นสุดท้าย (End User) เข้าถึงข้อมูลทางธุรกิจที่ผ่านกระบวนการ สังเคราะห์แล้ว (business process, applications และ data) ส่วนที่ติดต่อกับผู้ใช้งานนี้จะใช้งานได้ง่าย และไม่เหมือนเครื่องมือที่ใช้ติดต่อกับระบบที่นักพัฒนาเข้าถึง อันจะช่วยให้ผู้ใช้งานเกิดความคล่องตัวในการใช้งาน และ ธุรกิจดำเนินไปได้อย่างสะดวก รวดเร็ว

กระบวนการ 4 ชั้น ของการรวบรวมธุรกิจนี้จะไม่สามารถกระทำอย่างโดดๆได้ แต่จะต้องกระทำการรวบรวมไปพร้อมๆกัน ซึ่งสิ่งที่กล่าวมาทั้งหมดนี้จะกระทำภายในสถานที่เรียกว่า Operational Data Store (ODS) หรือ Data Warehouse ผล ผลิตปลายทางของข้อมูลเหล่านี้มักจะส่งตรงถึงผู้ใช้ขั้นสุดท้ายในรูปแบบที่ เรียกว่า Dashboard ซึ่งมีกระบวนการทำงานภายใต้ระบบที่เรียก ว่า Enterprise Portal

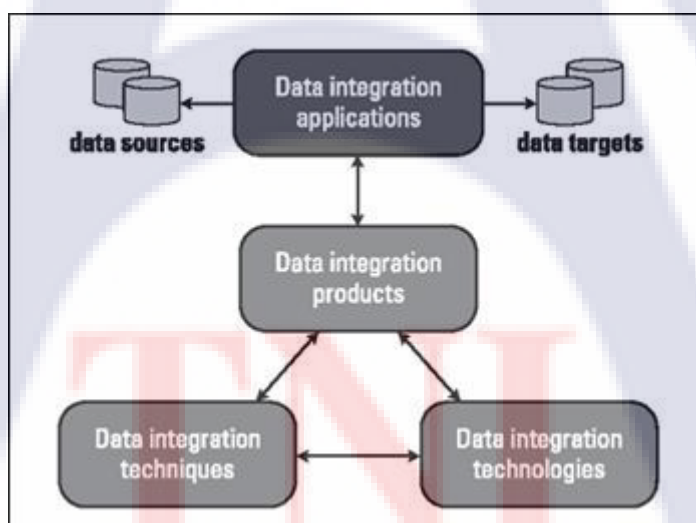


รูปที่ ก.4 แสดงการจัดระดับ 4 ระดับของ Enterprise Business Integration ภาพจาก BI Research

คุณลักษณะ เฉพาะของการควมรวมข้อมูล(Characteristics Data Integration)

Data Integration จำเป็นต้องใช้ Framework ของ application, เทคนิค ขั้นสูง, เทคโนโลยี และ ผลิตภัณฑ์ต่างๆ เพื่อสร้างระบบต่างๆ นี้

- **Applications** คือ โปรแกรมเฉพาะกิจหรือ โปรแกรมที่พัฒนาโดยผู้ขาย (vendor) และ นำมาใช้ให้เป็นประโยชน์กับการทำงาน
- **Products** หมายถึงโปรแกรมสำเร็จรูปที่พัฒนาขึ้นมาและนำมาเพื่อการขาย คือ นำมาพร้อมใช้นั่นเอง
- **Technologies** คือ เทคโนโลยีที่ใช้เพื่อการสร้าง integration technique
- **Techniques** คือ แนวทางปฏิบัติที่อิงตามแต่ละเทคโนโลยีเพื่อกระทำงานในด้าน Data Integration ให้ลุล่วง



ภาพที่ ก.5 แสดงส่วนประกอบและแนวทางในการ ทำ Data Integration

เทคนิคการควมรวมข้อมูล (Data Integration Techniques)

มีแนวทางของเทคนิคหลักๆอยู่ด้วยกัน 3 แนวทางคือ consolidation, federation และ propagation

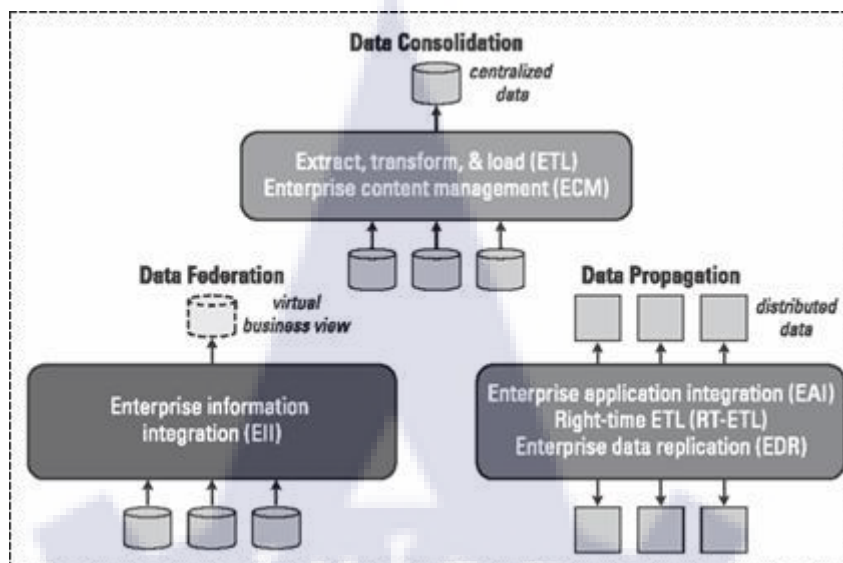
Data Consolidation คือการรวบรวมข้อมูลจากหลายๆ แหล่งข้อมูล หลายระบบและรวบรวมให้เป็นข้อมูลหนึ่งเดียว(single persistent data store) ข้อมูลนี้จะนำไปใช้ในกระบวนการอื่นๆ ต่อไปในอนาคตในงาน Data Warehouse หรือ อาจจะใช้ในงานรูปแบบอื่นใดนอกเหนือจาก Data Warehouse ก็เป็นไปได้

ด้วย วิธีการ Data Consolidation จะมีปัญหาเรื่องความช้าของ ข้อมูล (Delay or Latency) ระหว่างเวลาที่เกิดขึ้นกับข้อมูลจริงในระบบ เนื่องจากว่า ข้อมูลที่เกิดจากการใช้งานจริง เกิดอยู่ตลอดเวลาและตัวข้อมูลนั้นก็ยังไม่ได้ นำเข้าสู่ระบบ ซึ่งหากนำข้อมูลเข้าสู่ระบบก็จะเกิดปัญหาในการประมวลผลที่ค่อนข้างต้องใช้ เวลาตามมามาก ความช้าของข้อมูลนี้มีโอกาสเกิดในระดับ วินาที นาที ชั่วโมง หรือ หลายวัน ขึ้นอยู่กับธรรมชาติของงาน หรือ ความใหญ่โตของข้อมูล นั้น ดังนั้นจึงเกิดคำว่า “Near real time” คือ ความพยายามทำให้ข้อมูลใน Data Warehouse มีความใกล้เคียงกับสิ่งที่กำลังเกิดขึ้นจริงๆ ให้ ได้มากที่สุดเท่าที่จะทำได้ ข้อมูลที่เป็นแบบ “Real time” นั้น มีความยากที่จะทำ ได้ด้วยวิธีการแบบ “Data Consolidation” เนื่องจากการกระจายของ ข้อมูลของข้อมูลอยู่ในหลายๆที่นั่นเอง

พื้นที่ ใช้เก็บข้อมูลปลายทาง (ส่วนมากข้อมูลจะช้ากว่าปกติหนึ่งวัน เป็นต้น) จะสร้างโดยการใช้ batch เพื่อดึงข้อมูลbatch นี้จะดึง(pull) ข้อมูลจากต้นทางภายในระยะเวลาที่กำหนด

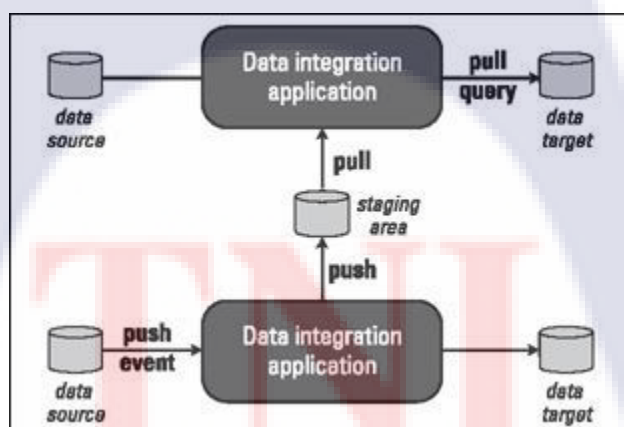
ด้วยวิธีการดึงข้อมูลแบบ Snapshot นี้เรา จะได้ข้อมูลที่เป็นปัจจุบัน (ณ เวลานั้น) ของข้อมูล ซึ่งการดึงนี้จะไม่ทำให้เกิดการเปลี่ยนแปลงข้อมูลในระบบ และ อีกประการคือ ณ ขณะที่ดึงข้อมูล นั้นข้อมูลที่กำลังถูกดึงออกมาสามารถที่จะเปลี่ยนแปลง ณ ขณะนั้นได้ตลอดเวลา แต่วาก็ถือว่า การดึงในจุดนี้จะไม่มีข้อมูลที่เรากำหนดให้มีการดึงติดมาด้วย

ข้อมูล ที่มีความแตกต่างเนื่องจากเวลาต่ำคือข้อมูลที่มีการแก้ไขได้สดผ่านทางระบบ Online หรือ ระบบที่มีการปรับปรุงข้อมูลเพื่อให้ทันสมัยจากต้นทางไปยังปลายทางอยู่เสมอๆ เรียกว่า ระบบผลัก (Push) ซึ่งเราสามารถ应用技术 ที่เรียกว่าChange Data Capture(CDC) คือ จะมีการตรวจสอบว่าข้อมูลชุดไหนบ้างที่มีการเปลี่ยนแปลง และการผลักข้อมูลมาก็จะผลักเฉพาะข้อมูลที่มีการเปลี่ยนแปลง ณ เวลาสุดท้ายที่เราทำการบันทึกไว้เท่านั้น จะไม่มีการผลักมาทั้งหมดทั้งก่อน ด้วยเทคนิคแบบนี้จะทำให้เราได้ข้อมูลมาในแบบ Incremental มีขนาดเล็กถึงลง รวดเร็วกว่าเดิมมาก



รูปที่ ก.6 แสดงเทคนิคการรวบรวมข้อมูล (Data Integration Techniques: consolidation, federation and propagation) จาก BI Research

ระบบ การ Consolidate ข้อมูลในรูปแบบ Push และ Pull สามารถใช้งานควบคู่กันได้โดยอิงตาม ขนาด และ ความเหมาะสมของสภาพแวดล้อม และ ขนาดของข้อมูล



รูปที่ ก.7 แสดงการทำงานในลักษณะ Push และ Pull ของ Data Consolidation

Data Federation จัดหาให้ได้ มาซึ่งภาพรวม ณ จุดเดียวของข้อมูลจากต้นทางหลายๆชุด จำเป็นต้องมีเครื่องมือเพื่อช่วยในการดึงข้อมูลจากหลายๆ ชุดที่มีอยู่และมาจัดรูปแบบให้อยู่ในรูปแบบเดียวกันเพื่อให้่ายต่อการนำไป ใช้ต่อ

Data Propagation คือเครื่อง มือที่ใช้ในการคัดลอกข้อมูลจากแหล่งหนึ่งไปสู่อีกแหล่งหนึ่ง โดยต้นทางและปลายทางไม่จำเป็นจะต้องเป็นชนิดของข้อมูลชนิดเดียวกัน เครื่องมือชนิดนี้ ยกตัวอย่างเช่นเครื่องมือประเภท Enterprise Application Integration (EAI) และ Enterprise Data Replication (EDR)

เทคโนโลยี เพื่อการรวบรวมข้อมูล (Data Integration Technologies)

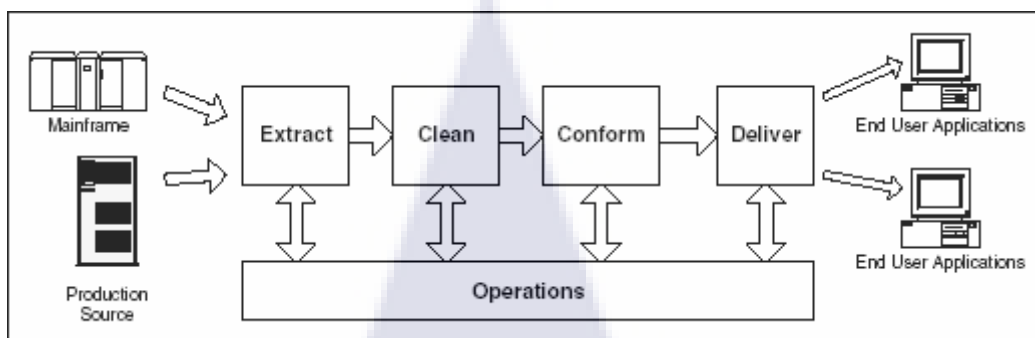
จากที่ได้กล่าวมาแล้วว่า เป้าหมายของการรวบรวมข้อมูลเข้าด้วยกันคือเพื่อนำข้อมูลจากหลายๆ แหล่ง เข้ามาจัดรูปแบบให้คล้ายกันเพื่อนำข้อมูลเหล่านั้น ไปใช้ประโยชน์ มีเครื่องมือต่างๆมากมาย ที่ใช้เพื่อช่วยในการรวบรวมข้อมูลเข้าด้วยกัน แต่เนื้อหาส่วนนี้จะมุ่งเน้นไปที่เครื่องมือที่ เรียกว่า ETL (Extract, Transform and Load)

ETL (Extract, Transform and Load)

ETL (Extract-Transform-Load) คือกระบวนการหนึ่งในระบบ Data Warehouse โดยระบบที่ ออกแบบเอาไว้จะดึงข้อมูลออกมาจากหลายๆ ที่, นำกระบวนการตรวจสอบคุณภาพของข้อมูลมา ประยุกต์ใช้, มีการเชื่อมโยงและปรับข้อมูลให้เป็นไปในรูปแบบเดียวกันเพื่อให้ ข้อมูลจากหลายๆ แหล่งสามารถใช้งานร่วมกันได้ และท้ายที่สุดทำการส่งมอบ (Delivery) ข้อมูลเหล่านั้นในรูปแบบที่ ง่ายต่อการใช้งาน เพื่อใช้ในการตัดสินใจขององค์กร

โดยมีกระบวนการหลักๆ ที่เกี่ยวข้องกับกระบวนการต่างๆ ดังต่อไปนี้

- **Extract** - กระบวนการดึงข้อมูลจากแหล่งของข้อมูลภายนอก
- **Transforming** – แปลง ข้อมูลเพื่อให้ได้ตรงตามกับความต้องการ ซึ่งเป็นกระบวนการที่ต้อง ใช้วิธีการเชิงคุณภาพ
- **Loading** – นำ ข้อมูลเข้าสู่ระบบปลายทางที่ต้องการ ซึ่งโดยทั่วไปจะหมายถึงระบบ Data Warehouse หรือ ฐานข้อมูลอื่นใด



รูปที่ ก.8 แสดงกระบวนการทำงานของระบบ ETL

โดยจะแยกอธิบายเป็นหัวข้อ ๆ ได้ดังต่อไปนี้

Extract (กระบวนการดึงข้อมูล)

เป็นกระบวนการเริ่มต้นของระบบ ETL จาก แหล่งของข้อมูล โดยทั่วไปแล้วระบบ Data warehouse จะประกอบด้วย ข้อมูลจากหลายๆ ที่ ข้อมูลที่อยู่ต่างที่กันนั้น อาจอยู่ในรูปแบบที่แตกต่างกันด้วยยกตัวอย่างเช่น อาจอยู่ในรูปแบบของฐานข้อมูลคนละชนิด หรือ ไม่ใช่ฐานข้อมูลแท้จริงซึ่งอาจจะเป็นระบบไฟล์ข้อมูลธรรมดา (Flat Files) หรือ ในอีกกรณีคือเป็นข้อมูลในฐานข้อมูลที่ไม่ใช่ระบบ RDBMS (Relational Database System) เช่นข้อมูลในรูปแบบของ Information Management System (IMS) หรือ ข้อมูลอื่นใดนอกระบบเช่น Virtual Storage Access Method (VSAM) หรือ Indexed Sequential Access Method (ISAM) หรือแม้กระทั่งการดึงจากข้อมูลรูปแบบ อื่นๆ ที่ดึงกันสด ๆจากระบบอื่นเลยเช่น Web Spider หรือ Screen scraping ข้อมูลจะถูกดึงเข้าสู่ระบบแบบ on-the-fly เป้าหมายของกระบวนการ Extract นี้คือดึงข้อมูลเข้ามาสู่รูปแบบมาตรฐานเดียวกัน เพื่อให้เหมาะสมต่อการแปลงรูปร่างของข้อมูลในขั้นตอนถัดไป เนื้อหาของกระบวนการนี้จะยุ่งเกี่ยวกับกระบวนการอ่านข้อมูลที่ถูกดึงเข้าสู่ ระบบ ส่งผลให้เกิดกระบวนการตรวจสอบข้อมูลว่า ตรงตามรูปแบบที่กำหนดไว้เบื้องต้นหรือไม่ ทั้งรูปแบบของข้อมูล และ โครงสร้างของข้อมูล ซึ่งถ้าหากเกิดข้อมูลนั้นไม่ได้มาตรฐานหรือรูปแบบที่กำหนดเอาไว้เบื้องต้น ก็เกิดกระบวนการปฏิเสธ (Reject) ข้อมูลนั้นๆ

Transform (กระบวนการแปลงข้อมูล)

ขั้นตอนการแปลงข้อมูลนี้จะมีการใช้กฎหรือฟังก์ชัน(Function) มากมายเพื่อที่จะแปลงข้อมูลให้ได้ตามที่ต้องการก่อนที่จะ นำข้อมูลเข้าไปยังปลายทาง ข้อมูลจากต้นทางบางแหล่งข้อมูลมีความ

จำเป็นน้อยมากหรือแทบจะไม่ต้องการ การแปลงข้อมูลเลย แต่ในบางแหล่งอาจต้องการ กระบวนการที่ซับซ้อน ซึ่งจะกินทรัพยากรระบบและเวลาในการประมวลผล ทั้งนี้ ความซับซ้อนของข้อมูลนั้นขึ้นอยู่กับความต้องการของธุรกิจ หรือ เป้าหมายของการนำข้อมูลไปใช้งาน โดยมี กระบวนการตัวอย่างต่อไปนี้

- **Selection** คือ เลือกเฉพาะสคัมภ์ (Column) ที่ต้องการที่จะ Load ยกตัวอย่างเช่น ถ้าต้นทางข้อมูลมี 3 สคัมภ์(Column) หรือจะเรียกว่ามี 3 attributes เช่น roll_no, age และ salary จะมีการแปลงข้อมูล และ เลือกที่จะไม่มีการแปลงข้อมูลหากว่าสคัมภ์ (column) salary มีค่าเป็น Null
- **Translation** คือ การแปลงข้อมูล ยกตัวอย่างเช่น หากข้อมูลต้นทางมีการเก็บข้อมูลเพศว่า 1 เป็นเพศชาย และ 2เป็นเพศหญิง จะต้องมีการแปลจากโค้ดที่กำหนดก่อนหน้านี้ให้ 1 = Male และ 2 = Female กระบวนการนี้เรียกว่า data cleaning หรือ กระบวนการทำความสะอาดข้อมูล
- **Encoding free form** ยกตัวอย่างเช่นการ mapping จาก “Male” ไปเป็น “1” และ “Mr” ไปเป็น “M”
- **Deriving a new calculated value** คือกระบวนการคำนวณเพื่อให้ได้ข้อมูลที่ต้องการยกตัวอย่างเช่น $\text{sale_amount} = \text{qty} * \text{unit_price}$
- **Filtering** คือ กระบวนการกรองเฉพาะข้อมูลที่กำหนด
- **Sorting** คือ กระบวนการเรียงข้อมูลที่ต้องการ
- **Joining** คือ กระบวนการเชื่อมโยงข้อมูลระหว่างตารางข้อมูล
- **Aggregation** คือ กระบวนการรวบรวม และ สรุปชุดข้อมูล ยกตัวอย่างเช่น การรวมยอด (summarize) ข้อมูลจากหลาย ๆ ระเบียบจนได้มาเป็น ยอดขายรวม เป็นต้น
- **Transposing or pivoting** คือการสลับทิศทางตำแหน่งของการแสดงข้อมูล เช่นการย้ายระเบียบไปเป็นสคัมภ์ หรือ ย้ายสคัมภ์มาเป็นระเบียบ เพื่อให้ง่ายต่อการนำข้อมูลไปใช้

- **Splitting column into multiple columns** คือ กระบวนการหั่นข้อมูลออกเป็นสคัมภ์ย่อย ยกตัวอย่างเช่นถ้าข้อมูลเป็นข้อมูลวันที่เช่น 2010/08/13 10:58:02 ก็จะมีการหั่นเป็นสคัมภ์ย่อย ๆ ได้ เป็น 2010,08,13,10,58,02
- **Disaggregation** คือ กระบวนการยกเลิกการรวมข้อมูล ยกตัวอย่างเช่นแทนที่จะเก็บข้อมูลเป็นระเบียบ แต่นำข้อมูลนั้นมาเชื่อมต่อกันเป็นระเบียบเดียว
- **Lookup and validate data** กระบวนการ Lookup ข้ามตารางเพื่อตรวจสอบความถูกต้องของข้อมูล
- **Applying any form of simple or complex data validation** กระบวนการนี้หากพบว่าการตรวจสอบข้อมูลผิด พลาดจะมีการปฏิเสธข้อมูลบางส่วนหรือทั้งหมดของข้อมูล แล้วแต่การกำหนดของผู้ที่ออกแบบการประมวลผลนี้

Load (ขั้นตอนการนำข้อมูลเข้าสู่ ระบบ)

กระบวนการ โหลดข้อมูลเข้า โดยทั่วไปจะนำข้อมูลเข้าไปในระบบ Data Warehouse ทั้ง นี้ขึ้นอยู่กับความต้องการขององค์กร หรือ ธุรกิจว่าจะให้ข้อมูลไหลไปในทิศทางใด บางองค์กร หรือ บางงานจะมีการสะสมของข้อมูล ความถี่ของการนำข้อมูลเข้าสู่ระบบ อาจจะมีการล้างข้อมูลแล้ว ทับข้อมูลใหม่ โดยทั่วไปแล้วข้อมูลของ Data Warehouse จะมีการใช้กันปีต่อปี เมื่อขึ้นปีใหม่แล้ว จะมีการล้างข้อมูลของปีเก่า และ เก็บไว้ในระบบข้อมูลสำรอง เนื่องจากว่ากระบวนการนำข้อมูลเข้า จะต้องปฏิสัมพันธ์กับฐานข้อมูล (Database) ดังนั้นจะต้องมีประเด็นเรื่องของ Database Constraints, Referential Integrity, Database Trigger เข้ามาเกี่ยวข้อง ด้วยในกระบวนการนำข้อมูลเข้า ซึ่งสิ่ง เหล่านี้รวม ๆ แล้วเรียกว่า กระบวนการควบคุมคุณภาพของข้อมูล (Data Quality performance of ETL process)

ตัวอย่างการใช้ ETL

- สถาบันทางการเงินต้อง การทางการเงินของลูกค้าจากหลายๆแผนกและในแต่ละแผนกอาจจะ ข้อมูลที่เก็บเอาไว้ ในรูปแบบที่แตกต่างกันออกไป แผนกที่ดูแลลูกค้าอาจจะลูกค้าที่เรียงตามรายชื่อ ซึ่งแผนกบัญชีอาจใช้วิธีค้นหาลูกค้าด้วยหมายเลข ID ระบบ ETL สามารถที่จะรวม ข้อมูลที่แตกต่างกัน

กันนี้พร้อมกับเชื่อมโยง ตรวจสอบคุณภาพของข้อมูลและนำเสนอในรูปแบบที่แตกต่างจากที่สอง แผนกใช้กันอยู่

หรืออีกทางคือบริษัท หรือองค์กรใช้ ETL เพื่อย้ายข้อมูลจากระบบหนึ่งไปอีกระบบหนึ่ง อย่างถาวร ยกตัวอย่างเช่น ข้อมูลจากระบบประมวลผลคำ(Word Processing) ซึ่งต้อง การจะแปลง ให้เป็นรูปแบบของตัวเลขและตัวอักษรเพื่อเก็บในฐานข้อมูล ด้วยวิธีการเช่นนี้ก็สามารถทำได้ ผ่านทาง ETL

วัฏจักรของ ETL ในชีวิตจริง

วงจรการใช้งาน ETL ในชีวิตจริงประกอบด้วยกระบวนการดังต่อไปนี้

- Cycle initiation – เริ่ม ตั้งต้นวงจร
- Build reference data – สร้างข้อมูลอ้างอิง
- Validation – ตรวจสอบ ความถูกต้อง
- Transform – แปลงรูป ข้อมูล กระบวนการนี้รวมทั้ง การ ทำความสะอาดข้อมูล นำกฎและความ ต้องการของธุรกิจมาประยุกต์ใช้ ตรวจสอบความสมบูรณ์ของข้อมูล รวมผลข้อมูลหรือยกเลิกการ รวมผลข้อมูล
- Stage – คือกระบวนการ โหลดตารางข้อมูลตามลำดับขั้น
- Audit report ตรวจสอบ ความถูกต้องของข้อมูลยกตัวอย่างเช่น ความถูกต้องของข้อมูลที่สามารถ เข้ากันได้กับกฎเกณฑ์และความต้องการเบื้องต้น ของธุรกิจ
- Publish – เผยแพร่ ข้อมูลไปยังปลายทางเช่นตารางข้อมูล
- Archive – กระบวนการ เก็บข้อมูลให้เป็นเอกสารสำคัญ
- Clean up – ทำความ สะอาดร่องรอยที่หลงเหลือจากการทำงาน

สิ่งที่ท้าทายของระบบ ETL

กระบวนการทำงานกับ ETL มีความสลับซับซ้อนและละเอียดอ่อนต่อข้อมูลมาก การทำงานอาจจะประสบปัญหาหากนักหาคใช้ระบบ ETL ที่ออกแบบมาไม่ดี รวมทั้งปริมาณข้อมูลอาจจะเกินกว่าที่ระบบจะรองรับได้ ซึ่งต้องใช้ผู้ออกแบบ(designer) มานั่งเสียเวลาตรวจสอบระบบ Data Warehouse เป็นระบบที่มีข้อมูลจากหลายๆแหล่งต้องปฏิสัมพันธ์ด้วย ระบบ ETL เป็นระบบเดียวที่เป็นตัวกลางในการประสานงานระบบต่างๆ เข้าด้วยกัน การออกแบบระบบ ETL เพื่อการใช้งานจะต้องคำนึงถึงความสามารถในการขยายขนาด (Scalability) ของ ETL ด้วย เพราะเมื่อปริมาณข้อมูลมีขนาดใหญ่แล้วจะมีปัญหาหรือไม่ ระบบ ETL บางระบบมีความสามารถในการรองรับข้อมูลกระจายไปอยู่ในหลาย ๆ Hard disk ได้และสามารถกระจายการทำงานของ batch ให้ไปทำงานแยกตามเครื่องอื่น ๆ เพื่อช่วยประมวลผลได้ด้วย

ประสิทธิภาพ

ผู้ขาย ETL มีการ benchmark จำนวนระเบียบสูงสุดที่ระบบตนสามารถทำได้ อยู่ที่ระดับ TB(terabyte) ต่อ ชั่วโมง หรือที่ระดับ ~1 GB ต่อ วินาที โดยการ Server ที่มีพลังสูง มี CPUs หลายตัว และมี Hard Drive หลายตัว และมี Gigabit Network Interface รวมทั้งมีหน่วยความจำความเร็วสูงในปริมาณมากด้วย

ในชีวิตจริงนั้น กระบวนการทำงานที่ช้าที่สุดของระบบ ETL นั้นเกิดจากขั้นตอนการโหลดข้อมูลเข้าและออกจากฐานข้อมูล ระบบฐานข้อมูลอาจจะทำงานช้าเพราะระบบจะต้องคำนึงถึงการเข้าถึงพร้อมๆกันจากหลายๆแหล่งข้อมูล(Concurrency) ซึ่งจะก่อให้เกิดความไม่ถูกต้องของข้อมูลตาม มาหากว่าไม่มีตัวนี้ อีกประเด็นสำคัญคือการจัดการกับระบบ Index แนวทางปฏิบัติในการทำงานเพื่อให้ได้ประสิทธิภาพที่สูงที่สุดมีแนวหลักคร่าวๆดังนี้

- ดึงข้อมูลออกหรือนำ ข้อมูลเข้าตรงๆให้มากที่สุดแทนที่จะใช้การ Query ข้อมูลเพื่อลดการใช้ทรัพยากรขณะทำการ Query

ขั้นตอนการแปลงข้อมูลให้กระภายนอกฐานข้อมูล

- ใช้ bulk load operation ให้มากที่สุดเท่าที่จะทำได้

ถึงแม้ว่าจะใช้กระบวนการ bulk load ก็ตาม ฐานข้อมูลก็ยังเป็นคอขวดประเด็นของปัญหาใน ETL อยู่เช่นเดิมวิธีการง่าย ๆ ในการเพิ่มประสิทธิภาพของระบบคือ

- Partition tables – พยายาม สร้างหรือทำ partition ให้เป็นขนาดเดียวกันเสมอ
- กระทำการตรวจสอบข้อมูล ในระดับของ ETL ก่อนที่จะโหลดข้อมูลเข้าไปฐานข้อมูล และ พยายามยกเลิก integrity checking (disable constraint) ในตารางข้อมูลที่จะโหลดเข้า
- ยกเลิกการใช้ Trigger เพื่อลดการทำงานมากเกินไปของ Database
- Generate ID ใน ETL แทนที่จะให้ฐานข้อมูลทำ
- หากตารางนั้นมี Indexes อยู่ ให้ทำการ drop indexes ก่อน แล้วโหลดข้อมูลเข้า หลังจากทีโหลด แล้วจึงมาสร้าง Indexes เหล่านั้นใน ภายหลัง

ข้อผิดพลาดของการทำงานใน ETL คือ การมี dependencies ระหว่างงานใน ETL มากเกินไป ยกตัวอย่างเช่น งาน A ผูกกับงาน B และ งาน C ผูก กับ B ฯลฯ ดังนี้

$A \Rightarrow B \Rightarrow C \Rightarrow D \Rightarrow E \Rightarrow F$

หากเป็นเช่นนี้จะเกิดลูกโซ่ที่จะต้อง รอกงานให้แล้วเสร็จจาก A ถึง F เป็นระยะเวลานาน ซึ่งหากเลื่องได้ก็ไมควรจะผูกงานให้ยาวมากเกินไป ควรจะกระจายงานออกเป็นกลุ่มย่อย ๆ แล้วแยกกันทำงาน จะทำได้เร็วและจัดการได้ดีกว่า

ประวัติผู้วิจัย

ชื่อ-สกุล	นายศตวรรษ เหล่าอากาศ
วัน เดือน ปีเกิด	24 กรกฎาคม 2533
ประวัติการศึกษา	
ระดับประถมศึกษา	ชั้นประถมศึกษาตอนต้นและตอนปลาย พ.ศ. 2539 โรงเรียนบ้านดอนศรีเสริมกสิกร
ระดับมัธยมศึกษา	ชั้นมัธยมศึกษาตอนต้นและตอนปลาย พ.ศ. 2545 โรงเรียนสตรีศรีน่าน
ระดับอุดมศึกษา	คณะวิศวกรรมศาสตร์ สาขาวิชา วิศวกรรมคอมพิวเตอร์ พ.ศ. 2550 สถาบันเทคโนโลยีไทย – ญี่ปุ่น
ทุนการศึกษา	กองทุนเงินกู้ยืมเพื่อการศึกษาที่ผูกกับรายได้ในอนาคต(กรอ.)
ประวัติการฝึกอบรม	- ไม่มี -
ผลงานที่ได้รับการตีพิมพ์	- ไม่มี -