



การดึงข้อมูลจากแหล่งข้อมูลต้นทาง สู่คลังข้อมูลปลายทางด้วยเครื่องมือ

IBM INFOSPHERE DATASTAGE

ETL FROM DATA SOURCE TO DATA WAREHOUSE USING

IBM INFOSPHERE DATASTAGE

นางสาวอินทร์ภากร กุลเลิศประเสริฐ

โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยี ไทย-ญี่ปุ่น

พ.ศ. 2561

การดึงข้อมูลจากแหล่งข้อมูลต้นทาง สู่คลังข้อมูลปลายทางด้วยเครื่องมือ

IBM INFOSPHERE DATASTAGE

ETL FROM DATA SOURCE TO DATA WAREHOUSE USING

IBM INFOSPHERE DATASTAGE

นางสาวอินทร์ภากร กุลเลิศประเสริฐ

โครงการสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีไทย - ญี่ปุ่น

ปีการศึกษา 2561

คณะกรรมการสอบ

..... ประธานกรรมการสอบ

(ผศ. ดร. ประจักษ์ เน็ดโหม)

..... กรรมการสอบ

(ผศ. ดร. นรินทร์ วีไลสกุลยง)

..... อาจารย์ที่ปรึกษา

(อาจารย์อดิศักดิ์ เสือสมิง)

..... ประธานสหกิจศึกษาสาขาวิชา

(อาจารย์สลิล ชีวกิจการ)

ลิขสิทธิ์ของสถาบันเทคโนโลยีไทย - ญี่ปุ่น

ชื่อโครงการ	การดึงข้อมูลจากแหล่งข้อมูลต้นทาง สู่คลังข้อมูลปลายทางด้วยเครื่องมือ IBM InfoSphere DataStage ETL FROM DATA SOURCE TO DATA WAREHOUSE USING IBM INFOSPHERE DATASTAGE
ผู้เขียน	นางสาวอินทร์ภากร กุลเลิศประเสริฐ
คณะวิชา	เทคโนโลยีสารสนเทศ สาขาวิชา เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา	อาจารย์อดิศักดิ์ เสือสมิง
พนักงานที่ปรึกษา	นายอภิสิทธิ์ แซ่ตั้ง
ชื่อบริษัท	บริษัท เอ-โฮสต์ จำกัด
ประเภทธุรกิจ/สินค้า	ให้บริการทางด้าน Oracle และ IBM Product และ Hosting Service

บทสรุป

ในยุคที่เทคโนโลยีมีการแข่งขันสูง องค์กรธุรกิจส่วนใหญ่จึงต้องมีการวางแผนการจัดเก็บข้อมูลและการนำข้อมูลไปใช้อย่างมีประสิทธิภาพ ระบบธุรกิจอัจฉริยะ (Business Intelligence) เป็นเทคโนโลยีในการนำเสนอข้อมูล ซึ่งช่วยในการตัดสินใจเพื่อสร้างกลยุทธ์ทางธุรกิจ จึงเป็นทางเลือกหนึ่งในการวางแผนกลยุทธ์ทางธุรกิจ ผู้จัดทำโครงการจึงได้รับมอบหมายให้ศึกษา IBM Infosphere DataStage ที่เป็นเครื่องมือในการทำ ETL (Extract Transform Load) ซึ่งเป็นกระบวนการหนึ่งในธุรกิจอัจฉริยะ

จากการดำเนินการทำ ETL นั้นเป็นการนำข้อมูลจากหลายๆแหล่งมาปรับเปลี่ยนให้อยู่ในรูปแบบที่ต้องการและตรวจสอบคุณภาพของข้อมูลก่อนนำข้อมูลทั้งหมดไปเก็บในฐานข้อมูลขนาดใหญ่

กิตติกรรมประกาศ

ในการที่ข้าพเจ้าได้มาศึกษา ณ บริษัท เอ-โฮสต์ จำกัด ตั้งแต่วันที่ 4 มิถุนายน 2561 ถึง วันที่ 28 กันยายน 2561 ได้ทำให้ข้าพเจ้าได้เรียนรู้ประสบการณ์และได้รับความรู้จากการทำงานจริง และส่งผลให้ข้าพเจ้าสามารถนำสิ่งต่างๆที่ได้จากการมาศึกษาเหล่านั้นมาใช้พัฒนาทักษะของตนเอง สำหรับรายงานการปฏิบัติงานสหกิจศึกษาในครั้งนี้สำเร็จลุล่วงได้ด้วยดีจากความร่วมมือและการสนับสนุนจากหลายฝ่ายดังนี้

1. คุณพิชานน จะรัมย์พันธ์
2. คุณระพีพัฒน์ ธนะพัฒน์
3. คุณนฤตยา โสมศิริกุล
4. คุณปาริชาติ ทองคอนน้อย
5. คุณอภิสิทธิ์ แซ่ตั้ง

และบุคลากรท่านอื่น ๆ ที่ไม่ได้กล่าวนามทุกท่านที่ได้ให้คำแนะนำช่วยเหลือในการจัดทำรายงานข้าพเจ้าใคร่ขอขอบพระคุณ ผู้มีส่วนเกี่ยวข้องทุกท่านที่มีส่วนร่วมในการให้ข้อมูลเป็นที่ปรึกษาในการทำรายงานฉบับนี้จนเสร็จสมบูรณ์ ตลอดจนให้การดูแลและให้ความเข้าใจเกี่ยวกับชีวิตของการทำงานจริงข้าพเจ้าขอขอบพระคุณไว้ ณ ที่นี้

TNI

นางสาวอินทร์ภากร กุลเลิศประเสริฐ

ผู้จัดทำ

สารบัญ

บทสรุป.....	ก
กิตติกรรมประกาศ.....	ข
สารบัญ.....	ค
สารบัญรูป	จ
สารบัญตาราง	ซ
บทที่ 1 บทนำ.....	1
1.1 ชื่อและที่ตั้งของสถานประกอบการ	1
1.2 ลักษณะธุรกิจของสถานประกอบการ หรือการให้บริการหลักขององค์กร	2
1.3 รูปแบบการจัดองค์กรและการบริหารองค์กร	4
1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย	5
1.5 พนักงานที่ปรึกษา และ ตำแหน่งของพนักงานที่ปรึกษา.....	5
1.6 ระยะเวลาที่ปฏิบัติงาน.....	5
1.7 ที่มาและความสำคัญของปัญหา	5
1.8 วัตถุประสงค์หรือจุดมุ่งหมายของโครงการ	6
1.9 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย	6
1.10 นิยามศัพท์เฉพาะ	6
บทที่ 2 ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน.....	7
2.1 ระบบปัญญาธุรกิจ (Business Intelligence: BI)	7
2.1.1 จุดเด่นของ BI คือ	7
2.1.2 กระบวนการทำงานของ BI	8
2.1.3 ประโยชน์ของ BI	10
2.2 เทคโนโลยีที่ใช้ในการปฏิบัติงาน	10
2.2.1 Microsoft SQL Server.....	10
2.2.2 WinSCP	12
2.2.3 IBM Infosphere Information Server	13
2.2.4 IBM Infosphere Information DataStage	13
2.2.5 IBM Infosphere Information Analyzer	16
บทที่ 3 แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน	17

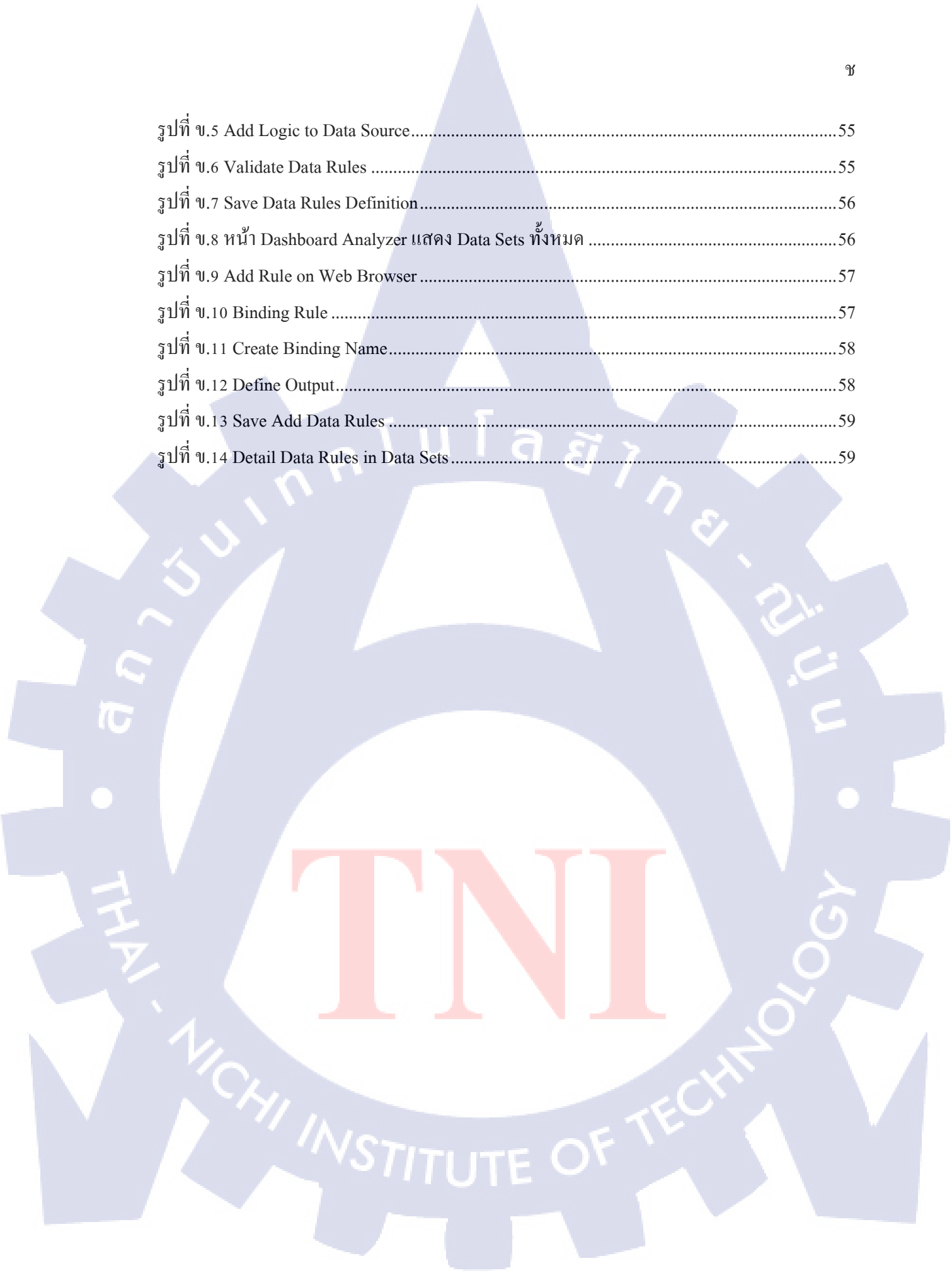
3.1 แผนงานการฝึกงาน	17
3.2 รายละเอียดที่นักศึกษาปฏิบัติในการฝึกงาน.....	17
3.3 ขั้นตอนการดำเนินงานที่นักศึกษาปฏิบัติงาน	18
3.3.1 ศึกษา Concept Business Intelligence	18
3.3.2 ศึกษาและค้นคว้า IBM InfoSphere Information DataStage	18
3.3.3 ศึกษาและค้นคว้า IBM InfoSphere Information Analyzer	18
3.3.4 ติดตั้งโปรแกรมเพื่อศึกษาและลองใช้งาน	18
3.3.5 จัดทำเอกสารคู่มือในการใช้งานระบบ.....	18
บทที่ 4 สรุปผลการดำเนินงาน การวิเคราะห์และสรุปผลต่างๆ.....	19
4.1 ขั้นตอนและผลการดำเนินงาน.....	19
4.1.1 ภาพรวมของระบบ	19
4.1.2 ETL Process : กระบวนการการนำเข้าข้อมูล	20
4.1.3 ตรวจสอบความถูกต้องของข้อมูล	31
4.2 ผลการวิเคราะห์ข้อมูล	31
4.3 วิเคราะห์และวิจารณ์ข้อมูลโดยเปรียบเทียบผลที่ได้รับกับวัตถุประสงค์และจุดมุ่งหมายใน การปฏิบัติงานหรือการจัดทำโครงการ	32
บทที่ 5 บทสรุปและข้อเสนอแนะ.....	33
5.1 สรุปผลการดำเนินงาน	33
5.2 แนวทางการแก้ไขปัญหา	33
5.3 ข้อเสนอแนะจากการดำเนินงาน	34
เอกสารอ้างอิง	35
ภาคผนวก.....	37
ภาคผนวก ก.....	38
ภาคผนวก ข.....	52
ประวัติผู้จัดทำโครงงาน.....	60

สารบัญรูป

รูป	หน้า
รูปที่ 1.1 แผนที่ตั้ง บริษัท เอ-โฮสต์ จำกัด	1
รูปที่ 1.2 A-HOST Proud Awards	3
รูปที่ 1.3 คณะผู้บริหารบริษัท เอ-โฮสต์ จำกัด แต่ละแผนก	4
รูปที่ 2.1 บทบาทของ BI.....	7
รูปที่ 2.2 ระบบงานของ BI	8
รูปที่ 2.3 ETL Extract-Transform-Load	9
รูปที่ 2.4 กระบวนการทำงานของ BI.....	10
รูปที่ 2.5 Microsoft SQL Server	11
รูปที่ 2.6 WinSCP	12
รูปที่ 2.7 การทำงานของ WinSCP	12
รูปที่ 2.8 IBM InfoSphere Information Server	13
รูปที่ 2.9 การทำงานของ DataStage.....	15
รูปที่ 2.10 IBM InfoSphere Information Analyzer	16
รูปที่ 4.1 โครงสร้างการนำเข้าและจัดเก็บข้อมูล	19
รูปที่ 4.2 Create Project	20
รูปที่ 4.3 การนำเข้าข้อมูลจากต้นทางเข้าสู่ Data Gateway	21
รูปที่ 4.4 การนำเข้าข้อมูลจาก Data Gateway สู่ Data Warehouse	21
รูปที่ 4.5 การนำเข้าข้อมูลจาก Data Gateway ออกรายงาน	22
รูปที่ 4.6 การนำเข้าสถานะการนำเข้าของข้อมูล	22
รูปที่ 4.7 Sequence File Stage.....	23
รูปที่ 4.8 Properties ของ Sequence File Stage.....	23
รูปที่ 4.9 Format ของ Sequence File Stage	24
รูปที่ 4.10 Column ของ Sequence File Stage	24
รูปที่ 4.11 ตัวอย่างข้อมูลใน Sequential File Stage	25
รูปที่ 4.12 Data Rules Stage.....	25
รูปที่ 4.13 การกำหนดขอบเขตของข้อมูล	26
รูปที่ 4.14 Transformer Stage	26

รูปที่ 4.15 Transformer Properties	27
รูปที่ 4.16 Column ใน Transformer Stage.....	28
รูปที่ 4.17 Parameter ที่เกี่ยวข้องกับ ETL Process	29
รูปที่ 4.18 การจัดลำดับการนำเข้าข้อมูล	29
รูปที่ 4.19 ตรวจสอบความถูกต้องของข้อมูล.....	31
รูปที่ ก.1 Setting to Run Program.....	39
รูปที่ ก.2 แสดงหน้า Command เพื่อบอกสถานะการติดตั้ง.....	39
รูปที่ ก.3 หน้า Browser เข้าสู่หน้าติดตั้งโปรแกรม	40
รูปที่ ก.4 เลือกภาษาที่ต้องการติดตั้ง	40
รูปที่ ก.5 หน้า Getting Start รายละเอียดของการติดตั้ง	41
รูปที่ ก.6 Firewall Requirement.....	42
รูปที่ ก.7 Check requirements.....	43
รูปที่ ก.8 Setting Regidit.....	43
รูปที่ ก.9 Regidit Editor	44
รูปที่ ก.10 Set Value Ntfs	44
รูปที่ ก.11 Early Requirements Check.....	45
รูปที่ ก.12 Installation Directory	45
รูปที่ ก.13 Installation Type Selection.....	46
รูปที่ ก.14 Product Selection	47
รูปที่ ก.15 Software License Agreement	47
รูปที่ ก.16 Desktop Shortcut Creation Option.....	48
รูปที่ ก.17 System Requirements Check	49
รูปที่ ก.18 Response File Review	49
รูปที่ ก.19 Pre installation Summary for the Current Computer	50
รูปที่ ก.20 Post Installation Summary for the Current Computer	51
รูปที่ ก.21 Software about ETL	51
รูปที่ ข.1 Data Quality	53
รูปที่ ข.2 Data Rule Definition	53
รูปที่ ข.3 Data Rule Definition Description	54
รูปที่ ข.4 Rule Logic.....	54

รูปที่ ข.5 Add Logic to Data Source.....	55
รูปที่ ข.6 Validate Data Rules	55
รูปที่ ข.7 Save Data Rules Definition.....	56
รูปที่ ข.8 หน้า Dashboard Analyzer แสดง Data Sets ทั้งหมด	56
รูปที่ ข.9 Add Rule on Web Browser	57
รูปที่ ข.10 Binding Rule	57
รูปที่ ข.11 Create Binding Name.....	58
รูปที่ ข.12 Define Output.....	58
รูปที่ ข.13 Save Add Data Rules	59
รูปที่ ข.14 Detail Data Rules in Data Sets	59



TNI

สารบัญตาราง

ตาราง

หน้า

ตารางที่ 3.1 ตารางปฏิบัติงานในโครงการ.....17

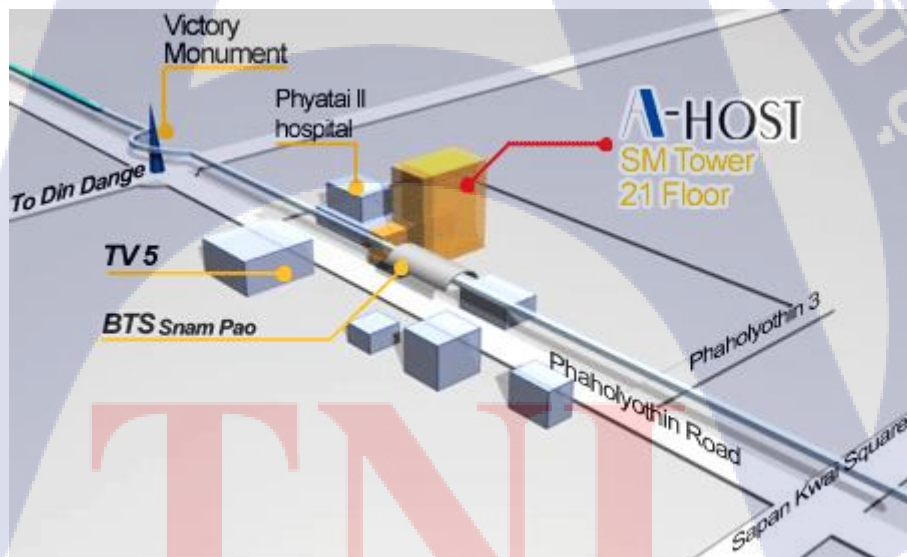


บทที่ 1

บทนำ

1.1 ชื่อและที่ตั้งของสถานประกอบการ

ชื่อสถานประกอบการ	:	บริษัท เอ-โฮสต์ จำกัด A-Host Company Limited
ที่ตั้งสถานประกอบการ	:	979/53-55 ชั้น 21 อาคาร SM Tower ถนนพหลโยธิน แขวงสามเสนใน เขตพญาไท กรุงเทพฯ 10400
โทรศัพท์	:	(66) 2298-0625-32
โทรสาร	:	(66) 2298-0053



รูปที่ 1.1 แผนที่ตั้ง บริษัท เอ-โฮสต์ จำกัด

1.2 ลักษณะธุรกิจของสถานประกอบการ หรือการให้บริการหลักขององค์กร

เอ-โฮสต์ก่อตั้งบริษัทเมื่อปี 2542 ซึ่งเป็นบริษัทหนึ่งในเครือของบริษัท เมโทร ซิสเต็มส์ คอร์ปอเรชั่น (มหาชน) จำกัด และเป็นผู้เชี่ยวชาญด้านบริการจัดวางระบบไอที และบริการเสริมต่างๆ สำหรับลูกค้าตั้งแต่ธุรกิจขนาดย่อมไปจนถึงขนาดกลาง

ธุรกิจหลักของบริษัท เอ-โฮสต์ คือ การให้บริการโฮสติ้ง และบริการระบบไอทีด้วยผลิตภัณฑ์ของออราเคิล (Oracle) และไอบีเอ็ม (IBM) ซึ่งเป็นซอฟต์แวร์สำหรับการวางแผนบริหารทรัพยากรขององค์กร (ERP) ระดับแนวหน้าของโลก

เอ-โฮสต์ถือกำเนิดขึ้นจากกลุ่มผู้เชี่ยวชาญด้านไอทีท่ามกลางภาวะเศรษฐกิจตกต่ำทั่วภูมิภาคเอเชีย แต่ เอ-โฮสต์ ก็สามารถเติบโตได้อย่างรวดเร็ว และมั่นคงตั้งแต่แรกก่อตั้ง ด้วยจุดแข็งในฐานะผู้บุกเบิกธุรกิจโฮสติ้งเซิร์ฟเวอร์ พร้อมทั้งนำธุรกิจแนวใหม่อย่างการให้บริการแอปพลิเคชัน หรือ ASP (Application Services Providing) เข้ามาให้บริการแก่องค์กรธุรกิจเป็นรายแรกในเมืองไทย

ธุรกิจการให้บริการแอปพลิเคชัน ในรูปแบบ ASP ของเอ-โฮสต์ไม่เพียงแต่ให้บริการด้านแอปพลิเคชันด้านการดำเนินธุรกิจทางอิเล็กทรอนิกส์ระดับโลกของออราเคิลพร้อมโครงสร้างพื้นฐานทางเทคโนโลยีสารสนเทศเท่านั้น แต่ยังมีบริการที่ครอบคลุมตั้งแต่การให้คำปรึกษา การสนับสนุนและการให้บริการทั่วไปอย่างพร้อมพร้อมครบครัน รวมทั้งยังมีความยืดหยุ่นสูงมีการปรับเปลี่ยนบริการและทรัพยากรให้เหมาะสมกับความต้องการ และสภาพงานที่แตกต่างกันของลูกค้าในรายได้ระดับต่างๆ

ในการดำเนินธุรกิจของเอ-โฮสต์ตลอดระยะเวลา 10 ปี ไม่เพียงแต่ในฐานะผู้บุกเบิกธุรกิจโฮสติ้ง และธุรกิจการให้บริการแอปพลิเคชันในรูปแบบ ASP เท่านั้น แต่เอ-โฮสต์ยังได้ทำการติดตั้งระบบไอที รวมทั้งผลิตภัณฑ์ของออราเคิลให้กับลูกค้าจนประสบความสำเร็จมาแล้วเป็นจำนวนมาก ซึ่งหลายรายเป็นหนึ่งในร้อยบริษัทชั้นนำของประเทศไทย แต่ที่สำคัญกว่านั้นก็คือการที่ เอ-โฮสต์ ได้สานสัมพันธ์กับลูกค้า และพันธมิตรทางธุรกิจอย่างแนบแน่นจนกลายเป็นหุ้นส่วนทางกลยุทธ์ และเป็นผู้สนับสนุนสำคัญที่มีส่วนช่วยผลักดันให้ธุรกิจของลูกค้าเติบโตสู่ความสำเร็จ

ปัจจุบันเอ-โฮสต์เป็นหนึ่งในบริษัทลูกของบริษัท ดราก้อนวัน จำกัด ซึ่งเป็นบริษัทมหาชนที่อยู่ในตลาดหลักทรัพย์แห่งประเทศไทย

เป็นเวลากว่า 10 ปีที่เอ-โฮสต์ และออราเคิลได้ดำเนินธุรกิจร่วมกันอย่างใกล้ชิด และถือเป็นพันธมิตรทางธุรกิจที่มีความแนบแน่นกันมานับตั้งแต่ก่อตั้งบริษัทปี 2542 จวบจนกระทั่งในปัจจุบัน

ในปี 2011-HOST ได้ก้าวไปข้างหน้าเพื่อความท้าทายทางธุรกิจใหม่ที่จะเป็น "พันธมิตรทางธุรกิจของไอบีเอ็มพรีเมียร์ (IBM Premier Business Partner)" เป็นตัวแทนจำหน่ายฮาร์ดแวร์ของไอบีเอ็มและผลิตภัณฑ์ซอฟต์แวร์ ที่จะสามารถเสริมสร้างผลิตภัณฑ์และผลงานบริการของเราเพื่อให้ลูกค้าได้รับเทคโนโลยีที่ดีที่สุดอยู่ตลอดเวลาและช่วยส่งเสริมการเจริญเติบโตของลูกค้าอย่างรวดเร็วและมั่นคง

ด้วยความมุ่งมั่นในการนำเสนอผลิตภัณฑ์ของออรากิล และ ไอบีเอ็ม ผ่านการให้บริการแอปพลิเคชันในรูปแบบของ ASP ในฐานะที่เอ-โฮสต์เป็นผู้บุกเบิกธุรกิจดังกล่าว และเพิ่มศักยภาพในการดำเนินธุรกิจของลูกค้าได้อย่างเต็มประสิทธิภาพ และเหมาะสม ทำให้ได้รับรางวัลแห่งความสำเร็จและได้รับการยกย่องมาอย่างต่อเนื่อง



รูปที่ 1.2 A-HOST Proud Awards

1.3 รูปแบบการจัดองค์กรและการบริหารองค์กร



รูปที่ 1.3 คณะผู้บริหารบริษัท เอ-โฮสต์ จำกัด แต่ละแผนก

1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย

ตำแหน่งงานที่ได้รับมอบหมายในการปฏิบัติงานคือ Assistant Technical Consultant หน้าที่งานที่ได้รับมอบหมายคือ ศึกษาค้นคว้า IBM Infosphere DataStage , IBM Infosphere Information Analyzer , เป็นผู้ช่วย Develop ใน Project Master Data Management และจัดทำเอกสาร

1.5 พนักงานที่ปรึกษา และ ตำแหน่งของพนักงานที่ปรึกษา

พนักงานที่ปรึกษา : นายอภิสิทธิ์ แซ่ตั้ง
ตำแหน่ง : Assistant Technical Consultant
โทรศัพท์ : 064-9349269
E-mail : apisit@a-host.co.th

1.6 ระยะเวลาที่ปฏิบัติงาน

ปฏิบัติงานสหกิจเป็นเวลา 4 เดือน ตั้งแต่วันที่ 4 มิถุนายน 2561 ถึง วันที่ 28 กันยายน 2561 ก่อนเริ่มปฏิบัติงานสหกิจมีการอบรมเป็นเวลา 1 เดือน 1 สัปดาห์ ตั้งแต่วันที่ 17 เมษายน 2561 ถึง วันที่ 25 พฤษภาคม 2561

1.7 ที่มาและความสำคัญของปัญหา

ในปัจจุบันมีข้อมูลหลากหลายฐานข้อมูลให้เลือกใช้และแต่ละองค์กรก็จะมี Server เป็นของตัวเองในแต่ละที่ที่จะมีข้อมูลอยู่มากมายจึงเกิดปัญหาในการจะดูข้อมูลของแต่ละที่ได้พร้อมๆกันทั้งหมด เพราะจะเกิดปัญหาหลายอย่าง เช่น สิทธิในการเข้าถึงข้อมูลการเชื่อมต่อของระบบ Network จึงมีกระบวนการในการจัดการปัญหาเหล่านี้คือ การทำ ETL เป็นกระบวนการในการดึงข้อมูลที่ต่าง Database ต่าง Server มาเก็บรวบรวมอยู่ใน Data Warehouse เพื่อที่จะเอาข้อมูลเหล่านี้ ออกรายงาน ได้อย่างรวดเร็ว เช่น การสรุปยอดขายของแต่ละสาขา หรือข้อมูลต่างๆ ในเชิงสถิติ เป็นต้น เพราะฉะนั้นการนำ ETL Database จึงมีความจำเป็นอย่างยิ่งในการจัดการข้อมูล

1.8 วัตถุประสงค์หรือจุดมุ่งหมายของโครงการ

การศึกษาในหัวข้อ การจัดการข้อมูลหลักด้วยเครื่องมือ IBM Infosphere DataStage มีวัตถุประสงค์เพื่อการจัดการข้อมูลจากหลายๆแหล่งข้อมูลมาปรับเปลี่ยนให้อยู่ในรูปแบบเดียวกันเพื่อให้ใช้ข้อมูลร่วมกันได้ และนำข้อมูลเข้าสู่คลังข้อมูล เพื่อนำไปใช้วิเคราะห์หรือพัฒนาระบบต่อไป

1.9 ผลที่คาดว่าจะได้รับการปฏิบัติงานหรือโครงการที่ได้รับมอบหมาย

1. ได้นำความรู้ที่ได้จากการปฏิบัติงานสหกิจไปใช้ต่อยอดในอนาคต
2. ได้รับความรู้และทักษะเฉพาะทางในสายงานนี้มากยิ่งขึ้น
3. มีความรับผิดชอบและสามารถทำงานภายใต้แรงกดดันของการทำงานจริง
4. ได้รับแนวทางการใช้ชีวิตจากการทำงานจริงเพื่อนำไปปรับปรุงให้ดียิ่งขึ้นในอนาคต
5. ระบบสามารถใช้งานได้จริง
6. ระบบที่ทำการนำเข้าสู่คลังข้อมูลสามารถนำไปต่อยอดได้
7. ระบบสามารถโอนย้ายข้อมูลตามความต้องการของลูกค้าได้

1.10 นิยามศัพท์เฉพาะ

1. BI (Business Intelligence) หมายถึง ชุดของแนวคิดและกระบวนการทัศน์ที่จะพัฒนากระบวนการตัดสินใจของธุรกิจโดยอาศัยข้อมูลที่เป็นข้อเท็จจริงจากฐานข้อมูล
2. Data Warehouse หมายถึง การจัดเก็บข้อมูลที่นำมาจากแหล่งข้อมูลภายในองค์กรจากฐานข้อมูลการใช้งานประจำวันหรือฐานข้อมูลปฏิบัติการ (Operational Database) หรือมาจากฐานข้อมูลภายนอกองค์กร (External Database) ข้อมูลในคลังข้อมูลจะถูกนำมาใช้เพื่อสนับสนุนการตัดสินใจบริหารงานของผู้บริหาร เช่น ระบบสนับสนุนการตัดสินใจ (Decision Support System) หรือ ระบบการจัดการสารสนเทศ (Management Information System)
3. ETL (Extract Transform Load) เป็นกระบวนการดึงข้อมูลจากข้อมูลจากหลายๆแหล่งข้อมูลมาปรับเปลี่ยนให้อยู่ในรูปแบบเดียวกันและนำข้อมูลเข้าสู่คลังข้อมูล

บทที่ 2

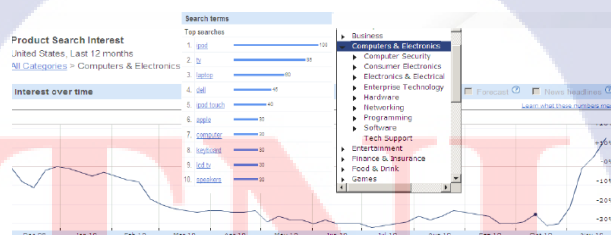
ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

ในการปฏิบัติงานสหกิจศึกษาครั้งนี้ เป็นการนำความรู้ทางด้านทฤษฎีและเทคโนโลยีมาใช้ในการปฏิบัติงานทุกส่วนตลอดการปฏิบัติงานสหกิจศึกษา ซึ่งเป็นการนำความรู้ทั้งที่เคยเรียนมาประยุกต์ใช้และเป็นการศึกษาเรียนรู้สิ่งใหม่ ๆ ที่ได้จากการปฏิบัติงาน

2.1 ระบบปัญญาธุรกิจ (Business Intelligence: BI)

Business Intelligence คือ การนำเสนอข้อมูล เพื่อช่วยในการวางแผน การตัดสินใจหรือตอบคำถามเชิงธุรกิจให้กับผู้บริหาร ทำให้องค์กรสามารถคาดการณ์หรือพยากรณ์ความต้องการของผู้บริโภคได้อย่างถูกต้องแม่นยำ ส่งผลให้ประสิทธิภาพการทำงานขององค์กรสูงขึ้น

Business Intelligence หรือที่เรียกว่า “ BI ” นั้นได้ถูกคิดค้นโดย Howard Dresner ในช่วงปี 1990 Howard ได้ให้ความหมาย คำว่า “ ระบบปัญญาธุรกิจ ” (Business Intelligence) หรือ BI หมายถึง “ชุดของแนวคิดและกระบวนการที่ซึ่งพัฒนากระบวนการตัดสินใจของธุรกิจโดยอาศัยข้อมูลที่เป็นข้อเท็จจริงจากฐานข้อมูล”



รูปที่ 2.1 บทบาทของ BI

2.1.1 จุดเด่นของ BI คือ

- 1) ใช้งานง่าย ผู้ใช้ไม่จำเป็นต้องมีความรู้ด้านฐานข้อมูล ก็สามารถใช้งานได้ เพียงเลือกข้อมูลที่ต้องการก็สามารถได้ผลลัพธ์ตามที่ต้องการ
- 2) ข้อมูลมีความถูกต้องแม่นยำ ทำให้สามารถใช้ข้อมูลเพื่อช่วยในการตัดสินใจได้รวดเร็วกว่าคู่แข่ง ทั้งในเชิงกว้าง และเชิงลึก

3) สามารถดึงข้อมูลจากฐานข้อมูลที่หลากหลายมาทำการวิเคราะห์ โดยไม่มีการเขียนโปรแกรม



รูปที่ 2.2 ระบบงานของ BI

2.1.2 กระบวนการทำงานของ BI

1) การกำหนดแหล่งข้อมูล (Data Sources) โดยแหล่งข้อมูลสามารถแบ่งออกเป็น 2 ประเภท คือ

- แหล่งข้อมูลภายใน (Internal Data Sources)
- แหล่งข้อมูลภายนอก (External Data Sources)

2) การออกแบบคลังข้อมูล (Data Warehouse Design) มีอยู่ 2 แบบ คือ

- คลังข้อมูลแบบ Star Schema หรือ Multidimensional Schema
- คลังข้อมูลแบบ Snowflake Schema

3) การคัดเลือก

ปรับเปลี่ยนข้อมูลให้เหมาะสม เพื่อนำข้อมูลเข้าสู่คลังข้อมูลโดยกระบวนการ ETL ETL ย่อมาจาก Extract, Transform, and Load คือ

การดึงข้อมูลจาก Data Source ต่าง ๆ เข้าสู่ Data Warehouse แบ่งเป็น 3 ขั้นตอนหลัก ดังนี้

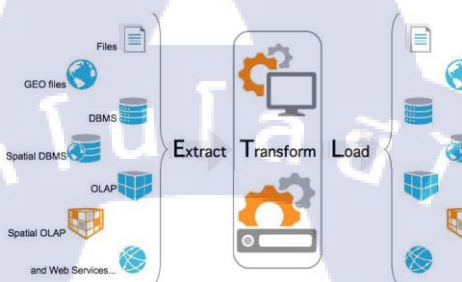
3.1) Extract คือ การดึงข้อมูลจากแหล่งข้อมูลที่แตกต่างกัน

3.2) Transform คือ การนำข้อมูลมาจัดรูปแบบให้ถูกต้องสอดคล้องกัน เช่น

- Data Mapping การทำให้ข้อมูลให้อยู่ในรูปแบบเดียวกัน
- Data Cleansing การตรวจสอบและแก้ไขข้อมูลให้ถูกต้อง

3.3) Load คือ การนำข้อมูลที่ผ่านการ Transform แล้ว เข้าสู่ Data Warehouse โดยทั่วไปแล้ว ETL Tools สามารถทำ

- Data Cleansing เป็นการตรวจสอบความถูกต้องของข้อมูล รวมทั้งกำจัดข้อมูลที่ผิดพลาด
- Data Transformation การแปลงข้อมูล หรือจัดรูปแบบข้อมูลเพื่อให้สามารถนำไปวิเคราะห์ได้ง่ายขึ้น
- Data Loading and Refreshing เป็นการกำหนด schedule ได้ว่าจะให้โหลดมาทุก ๆ กี่วัน หรือทุก ๆ เท่าไหร่ รวมทั้งยังสามารถกำหนด Storage ปลายทางได้อีกด้วย



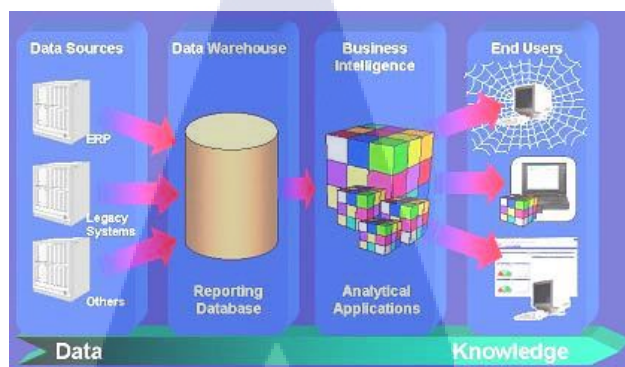
รูปที่ 2.3 ETL Extract-Transform-Load

4) การทำข้อมูลให้อยู่ในรูปแบบ Multidimensional Model หรือ Cube

รูปแบบการทำให้ข้อมูลเกิดมิติขึ้นในหลาย ๆ ด้านก่อนจะนำไปสร้างเป็นรายงานในรูปแบบต่าง ๆ โดยอาศัยเครื่องมือที่ช่วยในการ Query ข้อมูลที่ต้องการ เช่น Oracle Essbase, Cognos Transformer

5) การออกแบบและสร้างรายงาน (Interactive Report)

รายงานที่น่าสนใจมักจะเป็นผลการดำเนินงานตามตัวบ่งชี้การดำเนินงานต่าง ๆ ของหน่วยงานหรือการติดตามค่าเป้าหมายของการดำเนินงาน ที่สำคัญการนำเสนอรายงานมักจะอยู่ในรูปแบบของกราฟเพื่อทำให้เกิดความเข้าใจได้ง่ายผ่าน Dashboard ที่ผู้ใช้สามารถเข้าถึงผ่านหน้าเว็บไซต์ที่จัดทำไว้ เช่น OBI Report, Cognos Report เป็นต้น



รูปที่ 2.4 กระบวนการทำงานของ BI

2.1.3 ประโยชน์ของ BI

- ช่วยเพิ่มศักยภาพในการตัดสินใจให้ถูกต้องและรวดเร็วจากข้อมูลที่มีอยู่ โดยเห็นภาพพจน์ของข้อมูลที่มีก่อนการตัดสินใจ
- เพิ่มประสิทธิภาพในการแลกเปลี่ยนข้อมูลภายในองค์กร โดยสามารถแลกเปลี่ยนข้อมูลภายในผ่านเครือข่ายได้ในแบบอัตโนมัติ
- ลดต้นทุนทั้งด้านเงินและเวลาในการเข้าถึงข้อมูลองค์กร
- ช่วยให้ผู้ใช้สามารถตอบคำถามทางธุรกิจที่สำคัญ และช่วยให้สามารถ รวบรวมและปรับข้อมูลตามต้องการเพื่อสร้างแนวคิดที่แตกต่าง
- ดำรวจข้อมูลทุกชนิดจากทุกแง่มุม
- วิเคราะห์ข้อเท็จจริงและคาดการณ์แนวโน้มเชิงยุทธวิธีและเชิงกลยุทธ์

2.2 เทคโนโลยีที่ใช้ในการปฏิบัติงาน

2.2.1 Microsoft SQL Server

เป็นโปรแกรมในการบริหารจัดการฐานข้อมูลที่ดีที่สุดของ Microsoft โดยเป็นในรูปแบบของ Relational Database Management System หรือที่เรียกว่า RDBMS ซึ่งจะเป็นบริหารข้อมูลให้กับผู้ใช้บริการต่าง ๆ รองรับการทำงานได้จำนวนมาก และมีความสามารถมากมาย เทียบเท่ากับระบบฐานข้อมูลอื่น ๆ เช่น Oracle, DB2, Informix เป็นต้น มีคุณสมบัติเด่นเรื่องของ User Interface ที่ใช้งานได้ง่าย ภายใต้การจัดการของ Microsoft SQL Server นั้น จะมีฐานข้อมูลที่ Microsoft SQL

Server จะต้องใช้ และมีมาตั้งแต่ต้น นับตั้งแต่ Install กันเลย โดยฐานข้อมูลนี้จะอยู่ในกลุ่มของ System Database ซึ่งแต่ละฐานข้อมูลจะมีหน้าที่ในการทำงานต่าง ๆ กันดังนี้

2.2.1.1 System Database ของ Microsoft SQL Server

Master Database

มีความสำคัญมากที่สุด ใช้สำหรับเก็บข้อมูลที่สำคัญของระบบ เช่น Meta Data พวกร User, Login Information, Error Message, Linkedserver รวมถึงบอกตำแหน่งใน Primary File ในแต่ละข้อมูลอีกด้วย โดยหากฐานข้อมูลนี้มีปัญหา ก็จะส่งผลกระทบต่อการทำงานของ Microsoft SQL Server เลยทีเดียว

MSDB Database

สำคัญ รองจาก Master เก็บข้อมูลเกี่ยวกับ Alert, Job, Schedule ซึ่งถูกใช้โดยบริการของ SQL Server Agent ซึ่งเป็นบริการที่ทำงานอัตโนมัติ เช่น Backup ข้อมูลอัตโนมัติแต่ละวัน การแจ้งเตือนเมื่อระบบมีปัญหา โดยจะทำการส่ง Email ให้กับผู้ดูแลระบบ เป็นต้น

Distribution Database

ใช้ในการ Replication ของฐานข้อมูล เพื่อให้ข้อมูลของ Microsoft SQL Server แต่ละที่มีความตรงกัน

Model Database

เป็นฐานข้อมูลที่เป็นต้นแบบ (Database Template) กรณีที่เราสร้าง Database ใหม่ Microsoft SQL Server จะเอา Model Database นี้มาเป็นตัวตั้งต้น

TempDB Database

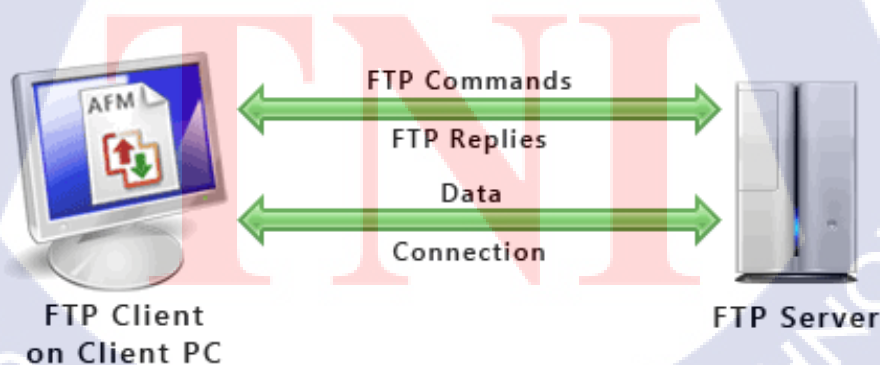
ไว้ในการเก็บข้อมูลที่เป็นชั่วคราว Temporary สำหรับ process ที่จำเป็นต้องการนำ data มาพักไว้ก่อนแล้วค่อยนำไปทำอย่างอื่นต่อ ลบทุกครั้งที่มีการ Restart Service หรือ Shutdown ระบบ

2.2.2 WinSCP



รูปที่ 2.6 WinSCP

WinSCP (วินเอสซีพี) เป็นโปรแกรมสำเร็จรูป FTP Client (เอฟทีพี ไครแอนท์) ที่เอาไว้ใช้ถ่ายโอนข้อมูล คัดลอกไฟล์ข้อมูล Upload (อัปโหลด) Download (ดาวน์โหลด) จากเครื่องลูกข่าย หรือ Client (ไครแอนท์) หรือคอมพิวเตอร์ของผู้ใช้ไปยังเครื่องแม่ข่ายหรือ Server สำหรับผู้ดูแลระบบหรือดูแลเว็บไซต์ จัดเป็นโปรแกรม FTP (เอฟทีพี) SFTP (เอสเอฟทีพี) โปรแกรม WinSCP (วินเอสซีพี) สามารถดาวน์โหลดไปใช้ได้ฟรีเพราะเป็น Open Source (โอเพ่นซอร์ส) บนระบบปฏิบัติการ Windows (วินโดว) โดยใช้มาตรฐาน FTP: File Transfer Protocol (เอฟทีพี: โพรโทคอล) ในการถ่ายโอนข้อมูล โปรแกรม FTP (เอฟทีพี) จะแบ่งเป็น 2 ส่วน FTP Server ติดตั้งไว้บน Server (เซิร์ฟเวอร์) เพื่อให้บริการ Client (ไครแอนท์) ที่ติดต่อมาเข้ามา FTP Client จะติดตั้งไว้บนเครื่อง Client (ไครแอนท์) หรือเครื่องของผู้ใช้ทั่วไป เพื่อติดต่อไปยัง Server (เซิร์ฟเวอร์) เพื่อถ่ายโอนข้อมูลจาก Client (ไครแอนท์) ไปยัง Server (เซิร์ฟเวอร์)



รูปที่ 2.7 การทำงานของ WinSCP

2.2.3 IBM InfoSphere Information Server

InfoSphere Information Server ให้บริการด้านการประมวลผลข้อมูลขนาดมหาศาลได้ มีความสามารถในการส่งแพลตฟอร์ม ที่มีความยืดหยุ่นสูงและมีสเกลการทำงานขนาดสูงซึ่งมีความแตกต่างในปริมาณของข้อมูลอยู่ในนั้น(มีทั้งขนาดใหญ่และเล็กรวมอยู่ในแพลตฟอร์มนั้น)

InfoSphere Information Server นั้นให้ความสามารถในการยืดหยุ่นข้อมูลต่างๆ ให้เข้ากับความต้องการในการรวบรวมข้อมูลของคุณได้ จากการรวบรวมข้อมูลเบื้องต้นไปสู่การประเมินคุณภาพข้อมูลและการจัดการข้อมูลเพื่อส่งข้อมูลที่เชื่อถือได้ในการนำไปประกอบการตัดสินใจในการทำธุรกิจของคุณ เช่น Big Data , Analytics , Data Warehouse Modernization และ Master Data Management



รูปที่ 2.8 IBM InfoSphere Information Server

2.2.4 IBM InfoSphere Information DataStage

IBM InfoSphere DataStage เป็นเครื่องมือที่ใช้รวบรวมข้อมูลของ IBM InfoSphere Information Server โดยการออกแบบ พัฒนา และใช้งาน job ต่าง ๆ ที่ทำหน้าที่โอนถ่ายและแปรรูปข้อมูล ซึ่ง data flow และ framework ของ job ต่าง ๆ จะมีการแสดงผลในรูปแบบของรูปภาพ graphic ทำให้เข้าใจได้ง่ายและเป็นมิตรต่อผู้ใช้งาน นอกจากนี้ DataStage ยังรองรับการดึงข้อมูล

จากแหล่งต่าง ๆ ที่มีระบบต่างกัน และสามารถส่งต่อข้อมูลที่ปรับแต่งอย่างเหมาะสมแล้วไปยัง data warehouse, data mart, operational data store, real-time web service และ messaging system หรือแอปพลิเคชันอื่น ๆ ขององค์กรได้โดยตรงเพื่อความสมบูรณ์และความถูกต้องของข้อมูล นอกจากนี้ DataStage ยังมี prebuilt functions ต่าง ๆ เพื่อช่วยลดระยะเวลาพัฒนาระบบและช่วยตรวจสอบความสอดคล้องของข้อมูล พร้อมทั้งรองรับกระบวนการทำงานทั้งแบบ ETL และ ELT นอกจากนี้ DataStage ยังถูกพัฒนาขึ้นเพื่อรองรับการขยายตัวขององค์กรโดยใช้เทคนิคการประมวลผลแบบขนาน

DataStage Features

- การเรียกใช้ข้อมูลโดยใช้ข้อมูลจากหลายๆ Project ร่วมกัน ด้วย Metadata Repository
- สามารถใช้ร่วมกับเครื่องมือ IBM InfoSphere FastTrack เพื่อลดระยะเวลาในการพัฒนา Project ได้
- สามารถใช้ร่วมกับเครื่องมือ IBM InfoSphere Information Analyzer เพื่อกำหนดกฎเกณฑ์ของข้อมูลต่าง ๆ ได้
- สามารถใช้ร่วมกับเครื่องมือ IBM InfoSphere Information Services Director เพื่อติดตั้งและใช้งาน Job ในรูปแบบของ Service ได้

DataStage's Roles

ใน DataStage จะมีการกำหนดบทบาท (Role) ต่าง ๆ ให้กับผู้ใช้เพื่อเป็นการกำหนด ควบคุม และจำกัดสิทธิต่าง ๆ ของผู้ใช้ บทบาทของผู้ใช้สามารถจำแนกออกเป็น 4 บทบาทดังนี้

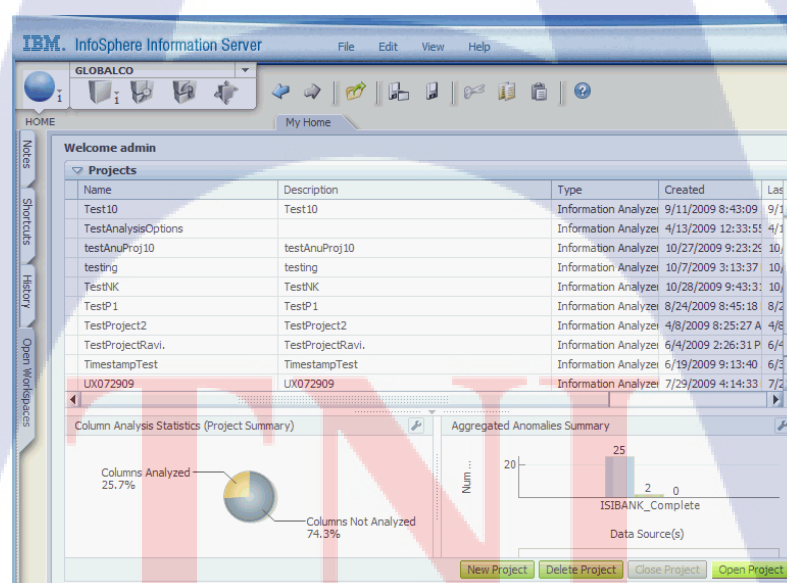
1. Administrator หรือ ผู้ดูแลระบบ เป็นบทบาทสูงสุด มีสิทธิในการกำหนดการตั้งค่าต่าง ๆ ได้ทั้งหมด สามารถเพิ่ม ลบ และประกาศใช้งาน Project ได้ สามารถบริหารจัดการสิทธิต่าง ๆ ของผู้ใช้ได้ โดยผู้ดูแลระบบจะทำงานผ่านทาง Command Interface ของ DataStage Repository
2. Manager หรือ ผู้จัดการระบบ มีสิทธิในการเข้าชมและแก้ไข Project ต่าง ๆ รวมถึงการจัดเก็บและบริหารจัดการ Metadata ที่สามารถนำมาใช้ใหม่ได้ โดยที่ผู้จัดการระบบจะทำงานผ่านทาง Main Interface ของ DataStage Repository

2.2.5 IBM InfoSphere Information Analyzer

IBM Analyzer เป็นโปรแกรมที่ใช้สำหรับการวิเคราะห์ข้อมูลให้ความสำคัญของคุณภาพ และความมั่นคงของข้อมูลให้ข้อมูลมีความแม่นยำมากขึ้นโดยการวิเคราะห์ Inconsistencies , Redundancies และ Anomalies ข้อมูลในแต่ละคอลัมน์ ตาราง Analyzer ยังช่วยสร้างทางเลือกในการตัดสินใจในการทำธุรกิจโดยการวิเคราะห์โครงสร้างข้อมูลเพื่อนำข้อมูลไปใช้อย่างมีประสิทธิภาพ

IBM Analyzer สามารถวิเคราะห์ข้อมูลได้หลายรูปแบบ ไม่ว่าจะเป็น Column Analysis , Key-Cross Domain Analysis , Baselines Analysis , Data Quality เป็นต้น และ Analyzer ยังสามารถกำหนดขอบเขตของคุณภาพข้อมูลได้ด้วย Data Quality ซึ่งใช้ Data Rules ในการจำกัดขอบเขตของข้อมูลชุดนั้นๆ

ในเครื่องมือ IBM Analyzer สามารถใช้งานร่วมกับ IBM DataStage ได้ โดยใช้ตรวจสอบคุณภาพของข้อมูลว่าเป็นไปตามที่ต้องการหรือไม่ โดยใช้ Data Rules Stage ซึ่งสามารถกำหนด Data Rules ได้ผ่านทางหน้า Console ของ IBM Analyzer



รูปที่ 2.10 IBM InfoSphere Information Analyzer

บทที่ 3

แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน

3.1 แผนงานการฝึกงาน

หัวข้องาน	เดือนที่ 1	เดือนที่ 2	เดือนที่ 3	เดือนที่ 4
สร้างชุดข้อมูลจำลองสำหรับทำ Demo BI Report				
ศึกษาและค้นคว้า Component ใน IBM InfoSphere Information DataStage				
ศึกษาและค้นคว้า IBM InfoSphere Information Analyzer				
เป็นผู้ช่วยในการ Develop ใน Project Master Data Management				
จัดทำเอกสาร Data Gateway				
จัดทำเอกสาร Master Data Management				
จัดทำเอกสารเสนอโครงการ (Project)				

ตารางที่ 3.1 ตารางปฏิบัติงานในโครงการ

3.2 รายละเอียดที่นักศึกษาปฏิบัติในการฝึกงาน

ในการปฏิบัติงานในโครงการสหกิจศึกษาที่บริษัท เอ-โฮสต์ จำกัด ได้รับหน้าที่ในตำแหน่ง Programmer ของแผนก Channel Service BI ลักษณะงานที่ได้รับมอบหมายคือ ศึกษาและค้นคว้าเครื่องมือ IBM InfoSphere Information DataStage และ IBM InfoSphere Information Analyzer เพื่อเตรียมความพร้อมสำหรับเป็นผู้ช่วยในการ Develop ใน Project Master Data Management และจัดทำในส่วนของเอกสาร Master Data Management และเอกสาร Data Gateway นอกจากนี้ยังได้รับมอบหมายให้ปฏิบัติงานจริงในโครงการของลูกค้าตามที่ได้รับ Requirement เพิ่มเติมในโครงการ Data Gateway

3.3 ขั้นตอนการดำเนินงานที่นักศึกษาปฏิบัติงาน

3.3.1 ศึกษา Concept Business Intelligence

ศึกษาภาพรวมของระบบปัญญาธุรกิจ เช่น Data Source , Data Warehouse , ETL , ELT , Cube , และ Report ทำความเข้าใจในกระบวนการทำงานและเครื่องมือต่าง ๆ ที่จำเป็นต่อการใช้งาน เพื่อเตรียมพร้อมในการปฏิบัติงานจริง

3.3.2 ศึกษาและค้นคว้า IBM InfoSphere Information DataStage

ศึกษาการใช้งานและ Component ต่าง ๆ ของเครื่องมือ IBM InfoSphere Information DataStage จาก Workshop ในเอกสารที่ได้รับมาและค้นคว้าเพิ่มเติมจากอินเทอร์เน็ต

3.3.3 ศึกษาและค้นคว้า IBM InfoSphere Information Analyzer

ศึกษาการใช้งานและ Component ต่าง ๆ ว่าตัวโปรแกรมสามารถใช้งานส่วนใดได้บ้างและนำมาใช้อย่างไรให้เกิดผลเพื่อศึกษาผลลัพธ์ที่เกิดขึ้นจากการวิเคราะห์ข้อมูลด้วยเครื่องมือ IBM Analyzer ส่วนส่วนประกอบดังนี้ Data Analysis , Data Quality และ Report เป็นต้น

3.3.4 ติดตั้งโปรแกรมเพื่อศึกษาและลองใช้งาน

เป็นการติดตั้งตัวโปรแกรมเพื่อสร้างโปรเจกในการทดสอบใช้งานในส่วนของเครื่องมือต่างๆในโปรแกรม โดยจะมีกระบวนการการติดตั้งเครื่องมือ ติดตั้ง Software อื่นๆที่จำเป็นสำหรับการใช้งานร่วมกันของโปรแกรม

3.3.5 จัดทำเอกสารคู่มือในการใช้งานระบบ

จัดทำเอกสารคู่มือการใช้งานระบบเพื่อเป็นกรณีศึกษางานตัวโปรแกรม

บทที่ 4

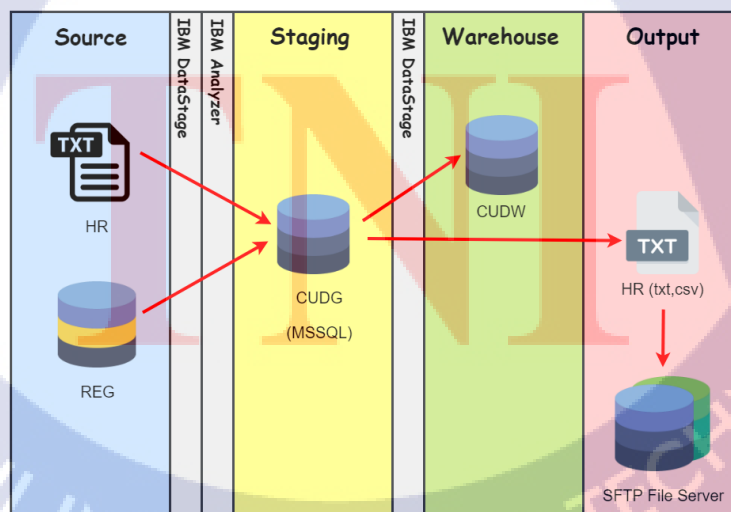
สรุปผลการดำเนินงาน การวิเคราะห์และสรุปผลต่าง ๆ

4.1 ขั้นตอนและผลการดำเนินงาน

ในช่วงระยะเวลาหนึ่งเดือนแรกของการปฏิบัติสหกิจศึกษา จะเป็นช่วงในการศึกษาตัวเครื่องมือ และข้อมูลที่จำเป็นต่อการใช้งานเครื่องมือ IBM DataStage และ IBM Analyzer รวมไปถึง Software ที่เกี่ยวข้องก่อนที่จะทำการจัดการข้อมูล เช่น Microsoft SQL Server , WinSCP เป็นต้น และในช่วงเข้าเดือนที่สองเป็นต้นไปได้เริ่มทำการสร้าง Jobs เพื่อจัดการและวิเคราะห์ข้อมูลตามแผนงานที่ได้วางไว้ มีขั้นตอน ดังนี้

4.1.1 ภาพรวมของระบบ

ได้ออกแบบระบบเพื่อทำการจัดลำดับการนำเข้าข้อมูล เป็นขั้นตอนแรกของการทำงาน โดยจะเริ่มจากการวางแผนว่าต้องการนำเข้าข้อมูลอย่างไร ใช้เครื่องมืออะไรในการทำงาน และออกรายงานข้อมูลออกมาได้อย่างถูกต้องมีผลลัพธ์ออกมาตรงตามความต้องการ ข้อมูลปลายทางต้องมีความสม่ำเสมอเหมือนข้อมูลต้นทาง โดยประกอบไปด้วย 4 ส่วน ดังนี้



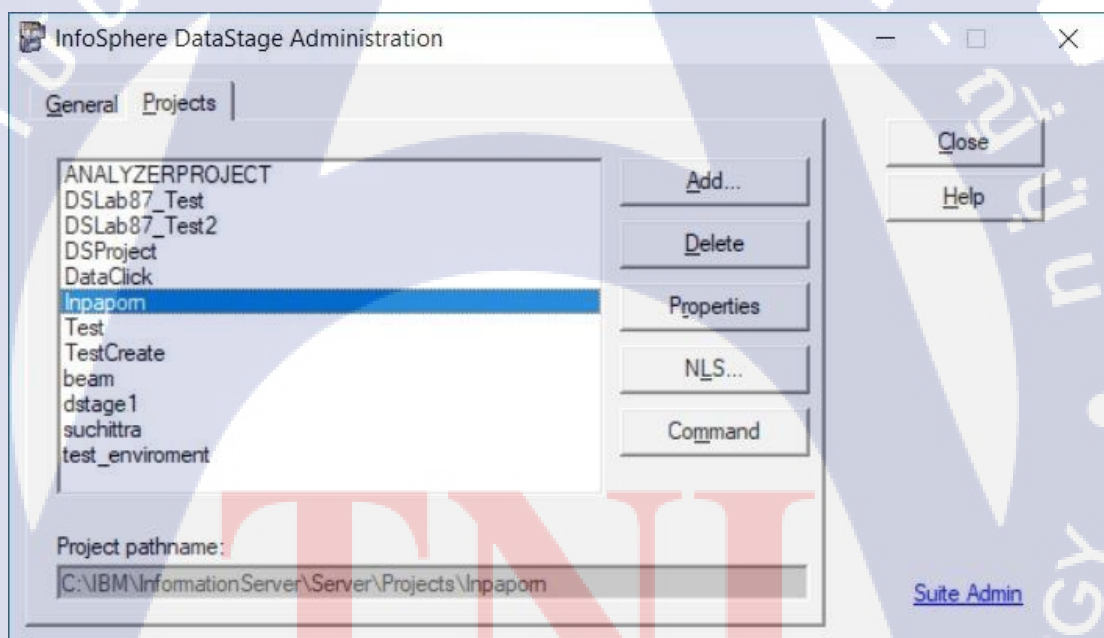
รูปที่ 4.1 โครงสร้างการนำเข้าและจัดเก็บข้อมูล

4.1.2 ETL Process : กระบวนการการนำเข้าข้อมูล

เป็นขั้นตอนที่เริ่มกระบวนการทำ ETL โดยเริ่มจากการเชื่อมต่อกับฐานข้อมูล สร้างพื้นที่สำหรับทำ Project จากนั้น จะเป็นการนำข้อมูลมาปรับให้เข้ากับคลังข้อมูล โดยปรับเปลี่ยนข้อมูลให้อยู่ในรูปแบบที่ถูกต้อง แล้วสร้าง Sequence Jobs เพื่อนำไปจัดลำดับการนำเข้าข้อมูล ซึ่งเป็นขั้นตอนที่นำ Object ต่างๆ มาผูกรวมกัน และรันเป็นลำดับ การสร้าง Sequence Jobs นั้นจะอยู่ในขั้นตอนสุดท้ายในการทำงานของ ETL Process ทั้งหมด โดยจะมีขั้นตอนการทำงานดังนี้

4.1.2.1 Create Project

เริ่มต้นด้วยขั้นตอนการสร้าง Project เพื่อสร้างพื้นที่ในการทำงาน

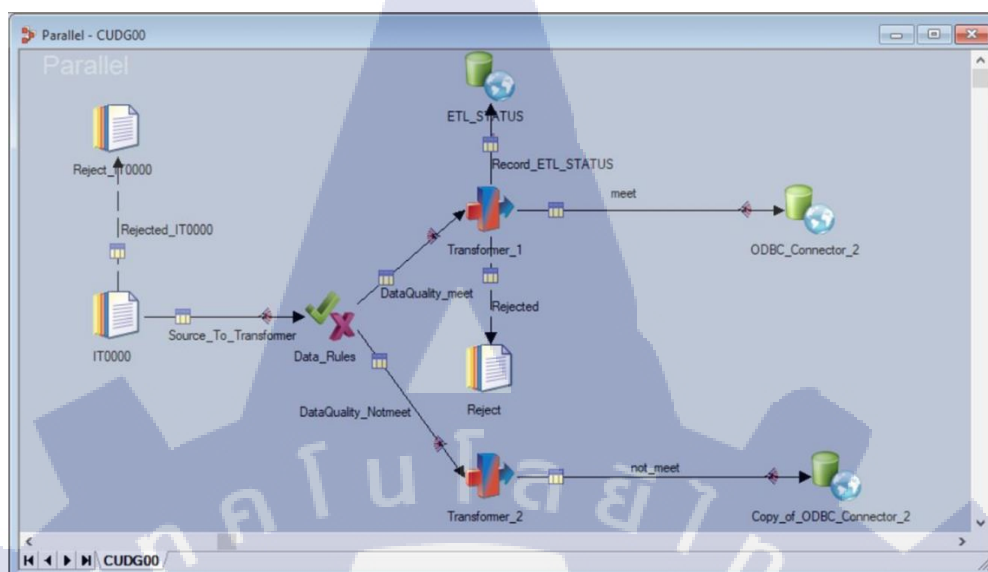


รูปที่ 4.2 Create Project

4.1.2.2 Create Parallel Jobs

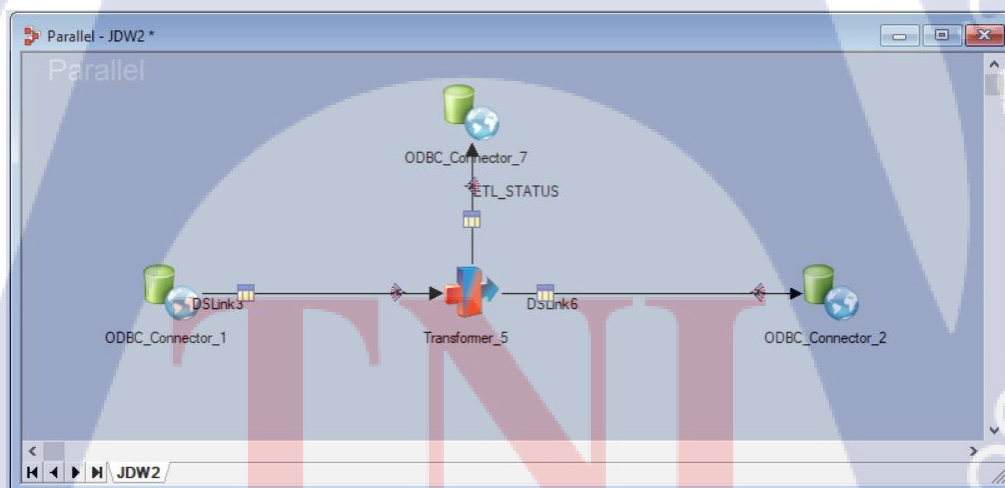
เป็นขั้นตอนในการสร้าง Parallel Job เพื่อทำการนำเข้าข้อมูล ปรับปรุงข้อมูล และนำข้อมูลออกไปใช้งาน โดยจะประกอบด้วย 3 Jobs หลักๆดังนี้

1. Job Flow สำหรับการนำเข้าข้อมูลเข้าสู่ Staging Database



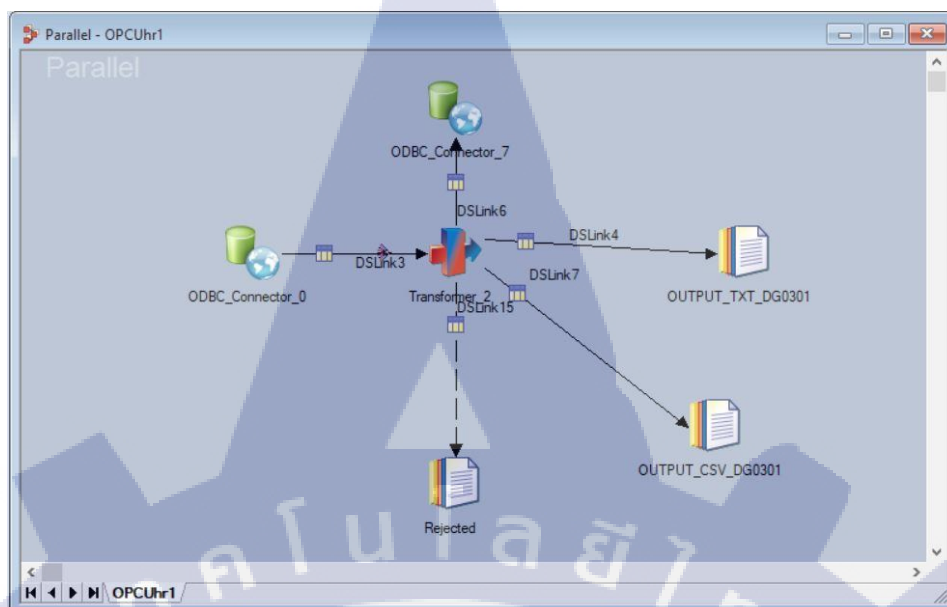
รูปที่ 4.3 การนำเข้าข้อมูลจากต้นทางเข้าสู่ Data Gateway

2. Job Flow สำหรับการนำเข้าข้อมูลเข้าสู่ Data Warehouse



รูปที่ 4.4 การนำเข้าข้อมูลจาก Data Gateway สู่ Data Warehouse

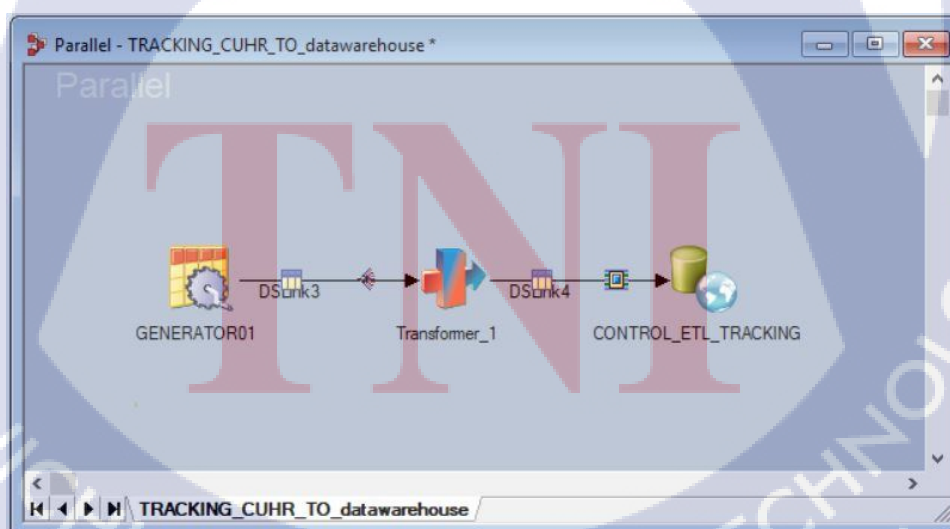
3. Job Flow สำหรับการนำข้อมูลส่งออกเป็นไฟล์เพื่อนำข้อมูลไปวิเคราะห์



รูปที่ 4.5 การนำเข้าข้อมูลจาก Data Gateway ออกรายงาน

4.1.2.3 Create Other Jobs

เป็นขั้นตอนสร้าง Job Flow ที่มีการนำเข้าสถานะการ Run jobs ของข้อมูล การปรับปรุงข้อมูล และการนำข้อมูลส่งออก ประกอบด้วยกระบวนการนำเข้าข้อมูล ดังนี้



รูปที่ 4.6 การนำเข้าสถานะการนำเข้าของข้อมูล

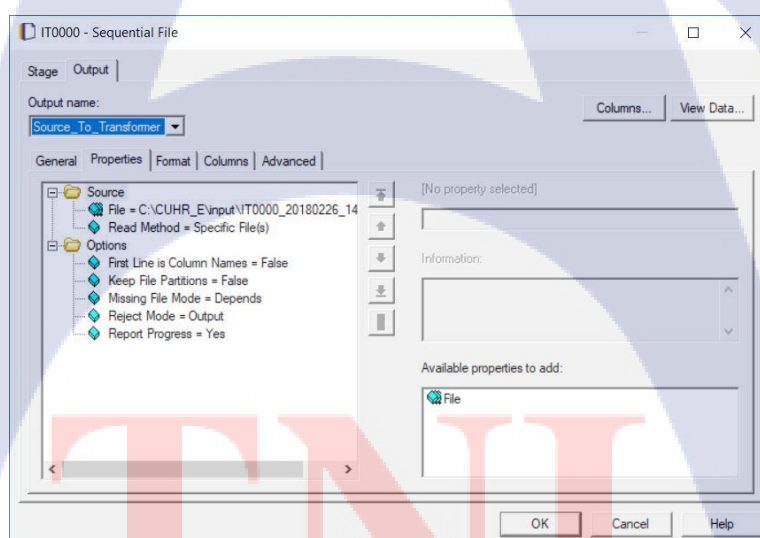
4.1.2.4 Palette ที่ใช้ในการทำงาน

1. Sequence File Stage

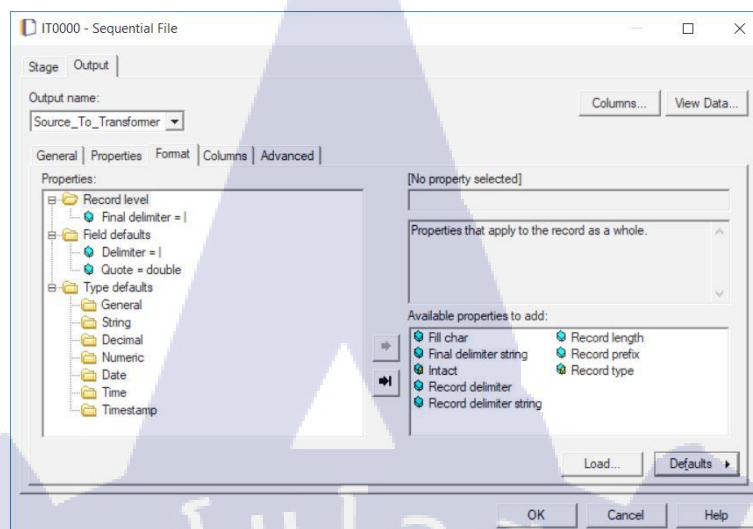


รูปที่ 4.7 Sequence File Stage

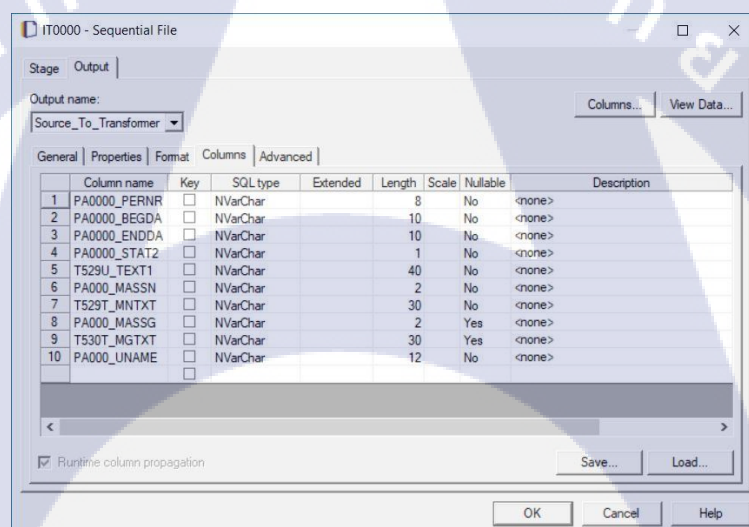
เป็น Stage สำหรับการอ่านไฟล์ที่เป็นไฟล์ txt ซึ่งการใช้งานนั้น หลัก ๆ แล้วจะต้องกำหนดตำแหน่งของไฟล์ที่ต้องการอ่านข้อมูลในไฟล์, Format ของข้อมูลในไฟล์กำหนดในส่วนของ Delimiter ข้อมูล และ Column ที่ต้องการนำข้อมูลเข้าและนำข้อมูลออก



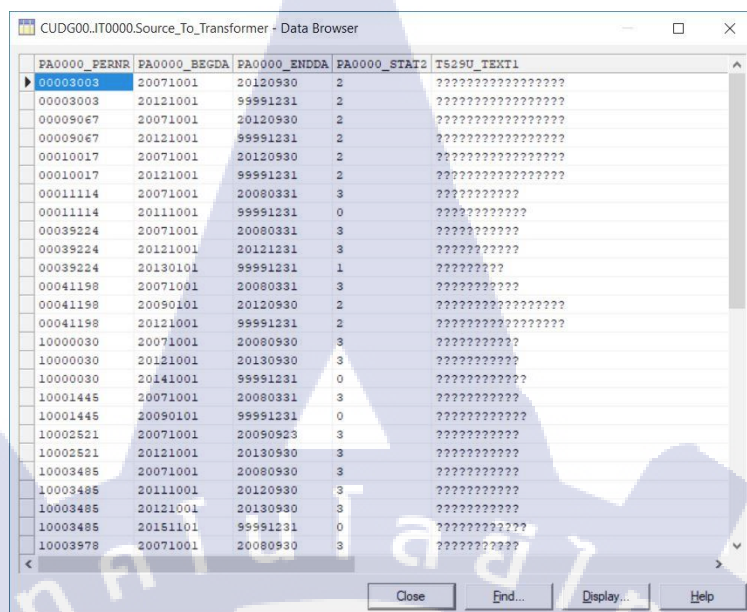
รูปที่ 4.8 Properties ของ Sequence File Stage



รูปที่ 4.9 Format ของ Sequence File Stage



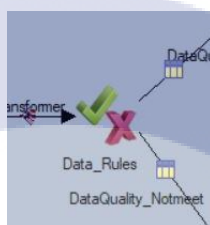
รูปที่ 4.10 Column ของ Sequence File Stage



PA0000_PERNR	PA0000_BEGDA	PA0000_ENDDA	PA0000_STAT2	TS29U_TEXT1
00003003	20071001	20120930	2	??????????????
00003003	20121001	99991231	2	??????????????
00009067	20071001	20120930	2	??????????????
00009067	20121001	99991231	2	??????????????
00010017	20071001	20120930	2	??????????????
00010017	20121001	99991231	2	??????????????
00011114	20071001	20080331	3	???????????
00011114	20111001	99991231	0	???????????
00039224	20071001	20080331	3	???????????
00039224	20121001	20121231	3	???????????
00039224	20130101	99991231	1	???????????
00041198	20071001	20080331	3	???????????
00041198	20090101	20120930	2	??????????????
00041198	20121001	99991231	2	??????????????
10000030	20071001	20080930	3	???????????
10000030	20121001	20130930	3	???????????
10000030	20141001	99991231	0	???????????
10001445	20071001	20080331	3	???????????
10001445	20090101	99991231	0	???????????
10002521	20071001	20090923	3	???????????
10002521	20121001	20130930	3	???????????
10003485	20071001	20080930	3	???????????
10003485	20111001	20120930	3	???????????
10003485	20121001	20130930	3	???????????
10003485	20151101	99991231	0	???????????
10003978	20071001	20080930	3	???????????

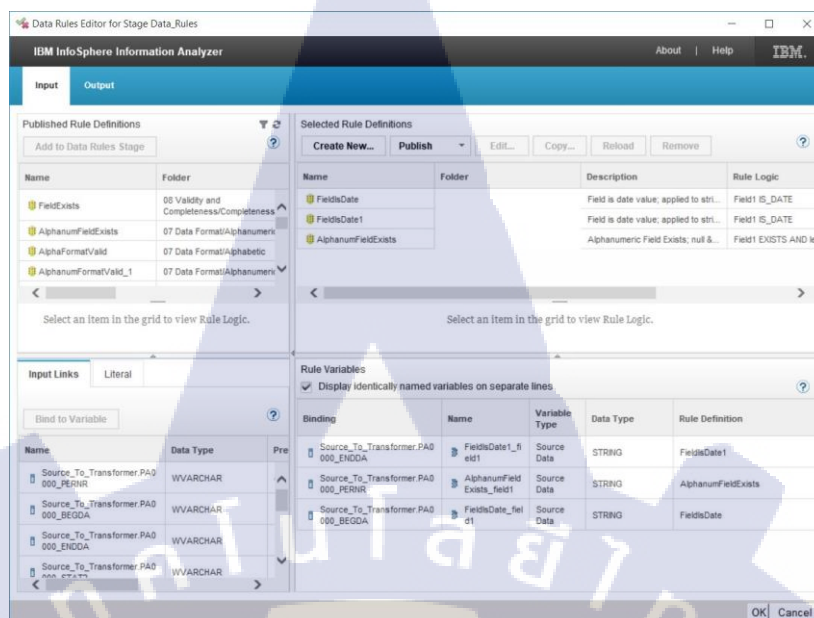
รูปที่ 4.11 ตัวอย่างข้อมูลใน Sequential File Stage

4.1.2.1 Data Rules Stage



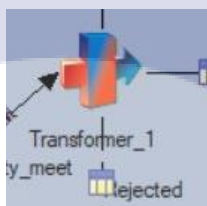
รูปที่ 4.12 Data Rules Stage

เป็น Stage สำหรับการกำหนดขอบเขตของข้อมูลเพื่อวัดคุณภาพของข้อมูลหากข้อมูลไม่ตรงตามขอบเขตนั้นก็จะมีไม่เข้าสู่ฐานข้อมูลปลายทาง



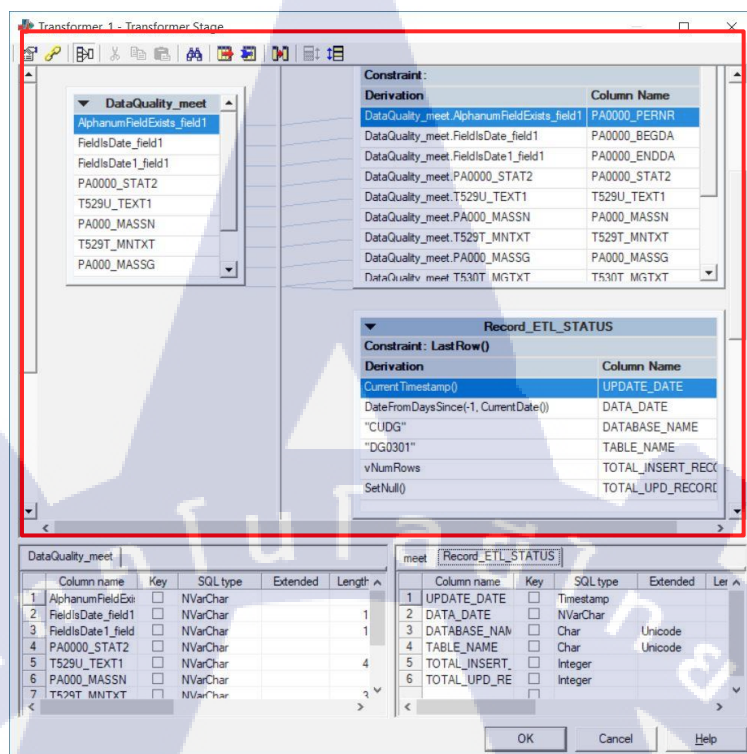
รูปที่ 4.13 การกำหนดขอบเขตของข้อมูล

4.1.2.2 Transform Stage



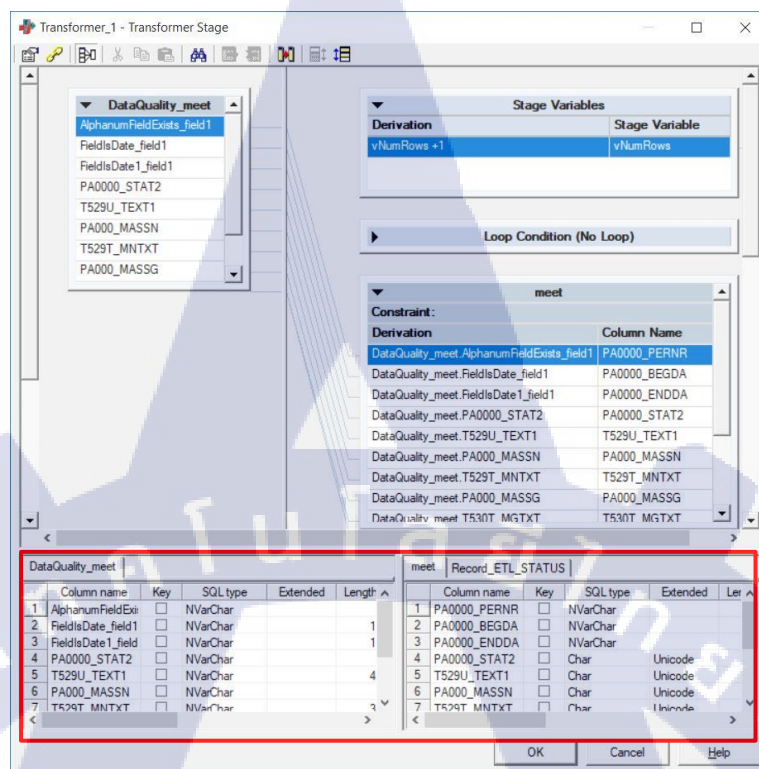
รูปที่ 4.14 Transformer Stage

เป็น Stage ที่ใช้สำหรับการปรับเปลี่ยนรูปแบบข้อมูลให้ออกมาในรูปแบบที่ถูกต้อง สามารถ Mapping Column จากต้นทางสู่ปลายทาง การใช้ Expression รวมไปถึงถ้าใน Column ไม่มีการกำหนดค่าในแต่ละคอลัมน์ สามารถกำหนดเป็นค่า Function ได้



รูปที่ 4.15 Transformer Properties

รูปแบบและขนาดของข้อมูลจากฝั่งต้นทางที่อยู่ทางฝั่งซ้าย และข้อมูลที่ส่งออกซึ่งอยู่ทางฝั่งขวา การตั้งค่าสองฝั่งนั้นไม่จำเป็นต้องตรงกันหรือมีจำนวนเท่ากัน



รูปที่ 4.16 Column ใน Transformer Stage

4.1.2.3 Setting Parameter

Parameter ที่เกี่ยวข้องกับกระบวนการนำเข้าและส่งออกข้อมูลเพื่อเก็บสถานะการนำเข้าข้อมูล มีดังนี้

Parameter Set : ENV ประกอบด้วย

ENV = Environment : DEV = CUDG_DEV

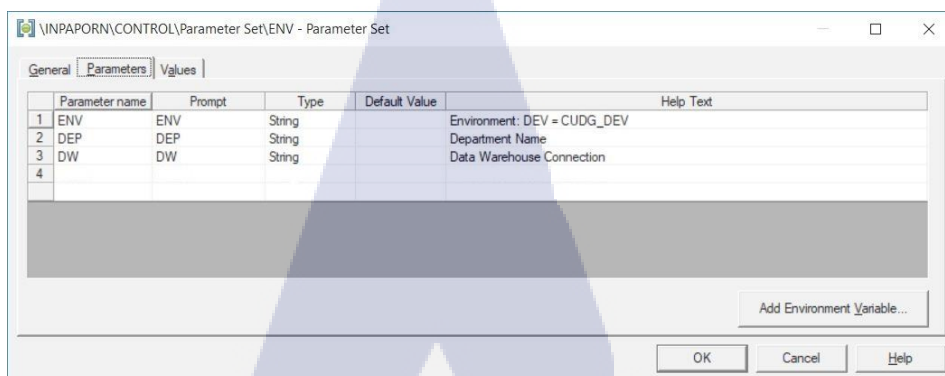
DEP = Department Name

DW = Data Warehouse Connection

และ Parameter อื่นๆ ดังนี้

CURR_DATE = เก็บค่าวันที่ปัจจุบัน

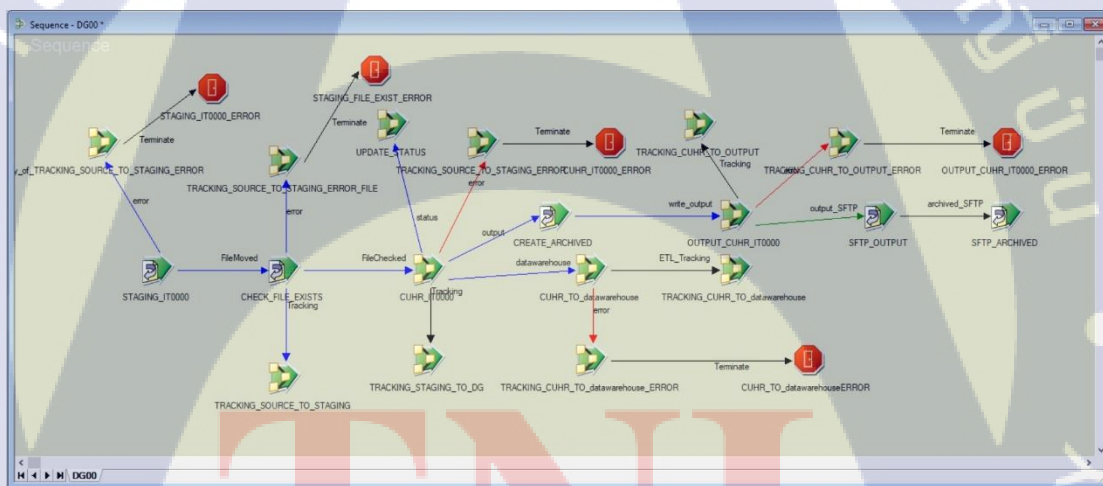
ETL_DATE = เก็บค่า Timestamp ของการทำงาน



รูปที่ 4.17 Parameter ที่เกี่ยวข้องกับ ETL Process

4.1.2.4 Create Sequence Jobs

เป็นขั้นตอนการจัดลำดับการนำเข้าข้อมูล เพื่อเช็คว่าการดำเนินการที่เกิดขึ้นเมื่อ Sequence Jobs รัน จะต้องมีการทำงานที่สอดคล้องกับ Parallel Jobs



รูปที่ 4.18 การจัดลำดับการนำเข้าข้อมูล

Job Name	Type	Description
STAGING_IT0000	Execute Command	เป็นการย้ายไฟล์ข้อมูลจากต้นทางมายัง Staging
CHECK_FILE_EXISTS	Execute Command	เป็นการตรวจสอบไฟล์ต้นทางที่ทำการย้ายมายัง Staging
CUDG_DG0301	Sequence Job	เป็น Sequence Job สำหรับนำข้อมูลจากไฟล์ IT0000.txt เข้าสู่ตาราง DG00 ในฐานข้อมูล CUDG
CREATE_ARCHIVED	Execute Command	เป็นการเรียกใช้ Script ทำการสร้างโฟลเดอร์วันที่ปัจจุบันลงใน Archived
OUTPUT_DG0301	Sequence Job	เป็น Sequence Job สำหรับนำข้อมูลจากตาราง DG00 ในฐานข้อมูล CUDG ส่งออกเป็นไฟล์ DG00.txt และ DG00.csv
SFTP_OUTPUT	Execute Command	เป็นการ run script เรียกดูข้อมูลใน SFTP
SFTP_ARCHIVED	Execute Command	เป็นการ run script อัปโหลดข้อมูลลง SFTP Server และ Delete ข้อมูลในเครื่อง SFTP File Server
CUDW_DG0101	Sequence Job	เป็น Sequence Job สำหรับนำข้อมูลจากตาราง DG00 ในฐานข้อมูล CUDG เข้าสู่ตาราง DG00 ในฐานข้อมูล datawarehouse
ETL_TRACKING_DG_TO_DW	Sequence Job	เป็นการบันทึกสถานะการ Run Job ของส่วนนั้นๆลงใน ETL_TRACKING
UPDATE_STATUS	Sequence Job	เป็นการอัปเดตสถานะของข้อมูลเมื่อมีการแก้ไขงาน
STAGING_IT0000_ERR OR	Terminator Activity	เป็นการส่งคำขอหยุดไปยังงานทั้งหมดตามลำดับหากมีการผิดพลาดในการทำงาน

4.1.3 ตรวจสอบความถูกต้องของข้อมูล

ในส่วนของการตรวจสอบนี้หลังจากที่ทำ ETL Process เสร็จเรียบร้อยแล้ว จะต้องทำการเช็คข้อมูลที่อยู่ในคลังข้อมูลว่าตรงตามความต้องการหรือไม่ ถ้าหากไม่ตรงต้องทำการแก้ไขเพื่อให้ตรงตามความต้องการ

	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE	JOB_ID	SALARY	COMMISSION_PCT	MANAGER_ID	DEPARTMENT_ID
1	100	Steven	King	SKING	515.123.4567	17-JUN-03	AD_PRES	24000			90
2	101	Neena	Kochhar	NKOCHHAR	515.123.4568	21-SEP-05	AD_VP	17000		100	90
3	102	Lex	De Haan	LDEHAAN	515.123.4569	13-JAN-01	AD_VP	17000		100	90
4	103	Alexander	Hunold	AHUNOLD	590.423.4567	03-JAN-06	IT_PROG	9000		102	60
5	104	Bruce	Ernst	BERNST	590.423.4568	21-MAY-07	IT_PROG	6000		103	60
6	105	David	Austin	DAUSTIN	590.423.4569	25-JUN-05	IT_PROG	4800		103	60
7	106	Valli	Pataballa	VPATABAL	590.423.4560	05-FEB-06	IT_PROG	4800		103	60
8	107	Diana	Lorentz	DLORENTZ	590.423.5567	07-FEB-07	IT_PROG	4200		103	60
9	108	Nancy	Greenberg	NGREENBE	515.124.4569	17-AUG-02	FI_MGR	12008		101	100
10	109	Daniel	Faviet	DFAVIET	515.124.4169	16-AUG-02	FI_ACCOUNT	9000		108	100
11	110	Chen	Jchen	JCHEN	515.124.4269	28-SEP-05	FI_ACCOUNT	8200		108	100
12	111	Ismael	Sciarra	ISCIARRA	515.124.4369	30-SEP-05	FI_ACCOUNT	7700		108	100
13	112	Jose Manuel	Uman	JMURMAN	515.124.4469	07-MAR-06	FI_ACCOUNT	7800		108	100
14	113	Luis	Popp	LPOPP	515.124.4567	07-DEC-07	FI_ACCOUNT	6900		108	100
15	114	Den	Raphaely	DRAPHEAL	515.127.4561	07-DEC-02	PU_MAN	11000		100	30
16	115	Alexander	Kvoo	AKHOO	515.127.4562	18-MAY-03	PU_CLERK	3100		114	30
17	116	Shelli	Baida	SBAIDA	515.127.4563	24-DEC-05	PU_CLERK	2900		114	30
18	117	Sigal	Tobias	STOBIAS	515.127.4564	24-JUL-05	PU_CLERK	2800		114	30
19	118	Guy	Himuro	GHIMURO	515.127.4565	15-NOV-06	PU_CLERK	2600		114	30
20	119	Karen	Colmenares	KCOLMENA	515.127.4566	10-AUG-07	PU_CLERK	2500		114	30
21	120	Matthew	Weiss	MWEISS	650.123.1234	18-JUL-04	ST_MAN	8000		100	50
22	121	Adam	Frip	AFRIPP	650.123.2234	10-APR-05	ST_MAN	8200		100	50
23	122	Payam	Kaufling	PKAUFLIN	650.123.3234	01-MAY-03	ST_MAN	7900		100	50
24	123	Shanta	Vollman	SVOLLMAN	650.123.4234	10-OCT-05	ST_MAN	6500		100	50
25	124	Kevin	Mourgos	KMOURGO	650.123.5234	16-NOV-07	ST_MAN	5800		100	50
26	125	Julia	Nayer	JNAYER	650.124.1214	16-JUL-05	ST_CLERK	3200		120	50
27	126	Irene	Mikkilineni	MIKKILIN	650.124.1224	28-SEP-06	ST_CLERK	2700		120	50
28	127	James	Landy	JLANDRY	650.124.1334	14-JAN-07	ST_CLERK	2400		120	50
29	128	Steven	Markle	SMARKLE	650.124.1434	08-MAR-08	ST_CLERK	2200		120	50

รูปที่ 4.19 ตรวจสอบความถูกต้องของข้อมูล

4.2 ผลการวิเคราะห์ข้อมูล

จากการที่ทำการศึกษา ติดตั้งตัวโปรแกรม และทำการนำเข้าข้อมูล โดยมีวัตถุประสงค์ให้นำเข้าข้อมูลอย่างมีประสิทธิภาพ ข้อมูลมีความคงเดิมครบถ้วน ระบบงานนี้สามารถนำไปต่อยอดได้ ซึ่งได้ดำเนินการจนเสร็จบรรลุตามวัตถุประสงค์เป็นที่เรียบร้อยแล้ว โดยสามารถนำเข้าข้อมูลและวิเคราะห์ขอบเขตคุณภาพของข้อมูล ซึ่งได้ผลลัพธ์ ดังนี้

4.2.1 สามารถทำการย้ายข้อมูลได้อย่างครบถ้วน

4.2.2 สามารถคงคุณภาพของข้อมูลได้เหมือนกับข้อมูลจากต้นทาง

4.2.3 สามารถนำระบบงานไปพัฒนาต่อได้

4.3 วิเคราะห์และวิจารณ์ข้อมูลโดยเปรียบเทียบผลที่ได้รับกับวัตถุประสงค์และจุดมุ่งหมายในการปฏิบัติงานหรือการจัดทำโครงการ

จากวัตถุประสงค์ การศึกษาในหัวข้อ การจัดการข้อมูลหลักด้วยเครื่องมือ IBM Infosphere DataStage เพื่อการจัดการข้อมูลจากหลายๆแหล่งข้อมูลมาปรับเปลี่ยนให้อยู่ในรูปแบบเดียวกัน เพื่อให้ใช้ข้อมูลร่วมกันได้ และนำข้อมูลเข้าสู่คลังข้อมูล เพื่อนำไปใช้วิเคราะห์และพัฒนาระบบต่อไป หลังจากการทำโครงการสามารถเปรียบเทียบผลที่ได้รับวัตถุประสงค์ในการจัดทำโครงการระบบเป็นไปตามวัตถุประสงค์ที่กำหนดไว้ แต่บางขั้นตอนมักจะมีปัญหาอยู่บ่อยครั้งแต่ก็สามารถแก้ไขปัญหาได้ในส่วนของ

TNI

บทที่ 5

บทสรุปและข้อเสนอแนะ

5.1 สรุปผลการดำเนินงาน

จากการที่นักศึกษาได้เข้าปฏิบัติงานสหกิจศึกษา ณ บริษัท A-HOST จำกัด ระหว่างวันที่ 4 มิถุนายน 2561 ถึง วันที่ 28 กันยายน 2561 ในตำแหน่ง Programmer ที่แผนก Channel Service BI ได้มาทำงานในส่วนของ ETL Architecture โดยใช้เครื่องมือ IBM InfoSphere Information DataStage ในการนำเข้าข้อมูล และ เครื่องมือ IBM InfoSphere Information Analyzer ให้ข้อมูลออกมามีคุณภาพและมีประสิทธิภาพตรงตามความต้องการของผู้ใช้งาน รวมถึงการเป็นผู้ช่วยในการ Implement ระบบให้แก่ลูกค้า การจัดทำเอกสารคู่มือ การจัดบันทึกการประชุม ซึ่งงานที่ได้รับมอบหมายนี้ ได้รับคำแนะนำ และเทคนิคในการทำงานจากพนักงานที่ปรึกษา

5.2 แนวทางการแก้ไขปัญห

ปัญหาที่พบระหว่างการศึกษและการจัดการชุดข้อมูลหลักด้วยเครื่องมือ IBM InfoSphere Information DataStage และ IBM InfoSphere Information Analyzer ในส่วนแรกจะเป็นในส่วนของการทำความเข้าใจในการทำงานของเครื่องมือต่างๆ เนื่องจากในส่วนของเอกสารที่ได้รับจากพนักงานที่ปรึกษาที่นำมาใช้ในการศึกษาเกี่ยวกับเครื่องมือมีรายละเอียดเกี่ยวกับ Palette ที่ใช้ในการจัดการข้อมูลมีไม่เพียงพอต่อการทำความเข้าใจ และด้วยเครื่องมือ IBM InfoSphere Information Analyzer เป็นเครื่องมือใหม่และไม่มีเอกสารเกี่ยวกับการใช้งานหรือ Workshop ทำให้การทำงานหรือการ Testing เป็นไปได้ยากและล่าช้า จึงต้องค้นคว้าข้อมูล ทำคู่มือการใช้งาน และลองทดสอบอยู่สม่ำเสมอ และส่วนที่สองจะเป็นในส่วนของการแก้ไขข้อผิดพลาดที่เกิดจากการ Run Jobs ของ IBM InfoSphere Information DataStage และ Test Quality ของ IBM Information Analyzer เนื่องจากตัวเครื่องมือทั้งสองนั้นเป็นเครื่องมือที่ใช้จัดการตัวข้อมูลดังนั้นปัญหาที่เกิดขึ้นส่วนใหญ่จะเป็นข้อผิดพลาดเกี่ยวกับ Case Sensitive และการ Connect ตัว ODBC แก้ไขปัญหาโดยการที่ตรวจสอบอย่างถี่ถ้วนและรอบคอบเกี่ยวกับ Datatype หรือการ Setting ต่างๆให้เรียบร้อย และส่วนสุดท้ายจะเกี่ยวกับเครื่องมืออื่นๆที่มีการใช้งานร่วมกันในการทำงาน เช่น WinSCP, FreeFTPd เป็นต้น เนื่องจากไม่เคยใช้งานมาก่อนทำให้เสียเวลาในการศึกษาเพิ่มเติม

5.3 ข้อเสนอแนะจากการดำเนินงาน

1. ควรมีการศึกษาเครื่องมือสำหรับการทำ ETL และวางแผนเพื่อเตรียมความพร้อมในการทำงานต่างๆ
2. ควรมีความละเอียดรอบคอบในการทำงาน
3. ควรศึกษาเกี่ยวกับเครื่องมือพื้นฐานที่จำเป็นในการใช้งานในเรื่องเกี่ยวกับ File Transfer และการ Coding



เอกสารอ้างอิง

- [1] IMD Company Limited , [2010] , Business Intelligence [Online] , Available : <http://www.imd.co.th/knownledges.php?id=1> [2018, June 04].
- [2] IBM Corporation , [2008] , DataStage [Online] , Available : <file:///C:/Users/Cherth/Desktop/red%20book%20dstage.pdf> [2018 , June 12].
- [3] IBM Corporation , [2014], Sequence Activity [Online] , Available : https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ds.design.doc/topics/c_ddesref_Activity_Stages.html [2018 , July 06].
- [4] IBM Corporation , [2011] , Analyzer [Online] , Available : https://www.ibm.com/support/knowledgecenter/SSZJPZ_8.7.0/com.ibm.swg.im.iis.productization.iisinfsv.overview.doc/topics/cisoiaoverview.html [2018 , July 10].
- [5] IBM Corporation , [2014] , Analysis Column [Online] , Available : https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ia.report.doc/topics/c_column_analysis_statistics_reports.html [2018 , July 18].
- [6] IBM Corporation , [2014] , Data Rules [Online] , Available : https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.7.0/com.ibm.swg.im.iis.ia.quality.doc/topics/dq_executables_overview.html [2018 , July 23].
- [7] IBM Corporation , [2011] , Type of check (Data Rules) [Online] , Available : https://www.ibm.com/support/knowledgecenter/SSZJPZ_8.7.0/com.ibm.swg.im.iis.ia.quality.doc/topics/dq_types_rule_checks.html [2018 , July 25].
- [8] IBM Corporation , [2015] , Mail Server [Online] , Available : https://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/com.ibm.swg.im.iis.event.doc/topics/t_configuring_email_provider.html [2018 , July 26].
- [9] WinSCP.net , [2000-2018] , WinSCP [Online] , Available : <https://winscp.net/eng/index.php> [2018 , August 03].
- [10] IBM Corporation , [2015] , Data Rule Stage [Online] , Available : https://www.ibm.com/support/knowledgecenter/SSZJPZ_8.7.0/com.ibm.swg.im.iis.ia.dsrules.doc/topics/dr_rules_and_rulesets.html [2018, August 10].

[11] IBM Corporation , [2014] , Quality DataStage [Online] , Available :

https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.3.0/com.ibm.swg.im.iis.ds.nav.doc/containers/cont_iisinfo_ds_and_qs.html [2018 , September 09].

[12] IBM Corporation , [2014] , Methodology [Online] , Available :

<http://publibfp.boulder.ibm.com/epubs/pdf/c1942930.pdf> [2018 , September 19].



สถาบันเทคโนโลยีไทย-ญี่ปุ่น

ภาคผนวก

TNI

THAI - JAPANESE INSTITUTE OF TECHNOLOGY

ภาคผนวก ก

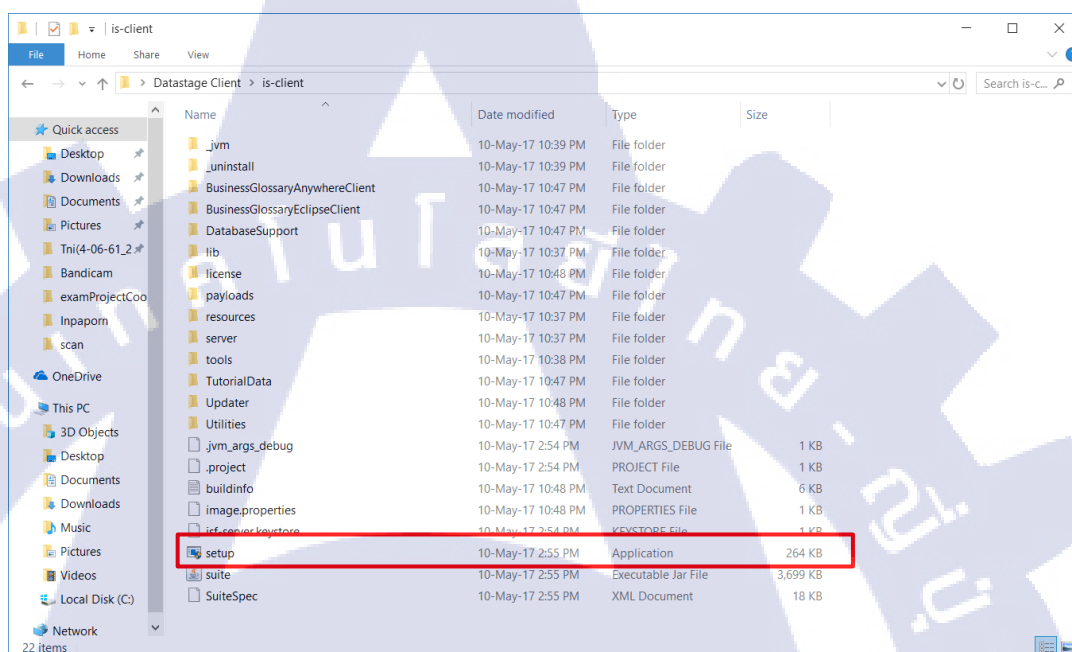
การติดตั้งโปรแกรม IBM InfoSphere Information Server

TNI

IBM InfoSphere Server Installation Guide

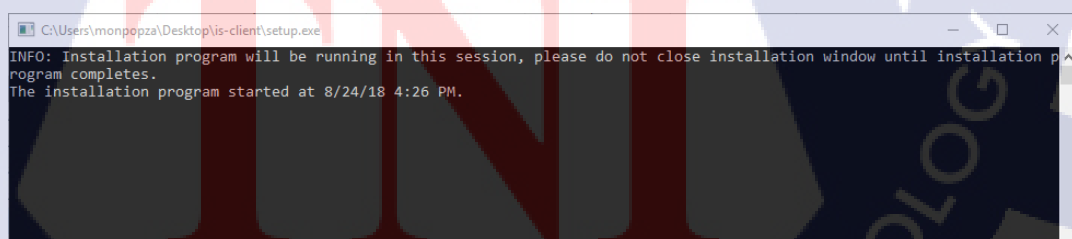
กระบวนการติดตั้ง มีดังนี้

1. ทำการ Double Click ที่ไฟล์ Set Up เพื่อทำการรันโปรแกรม



รูปที่ ก.1 Setting to Run Program

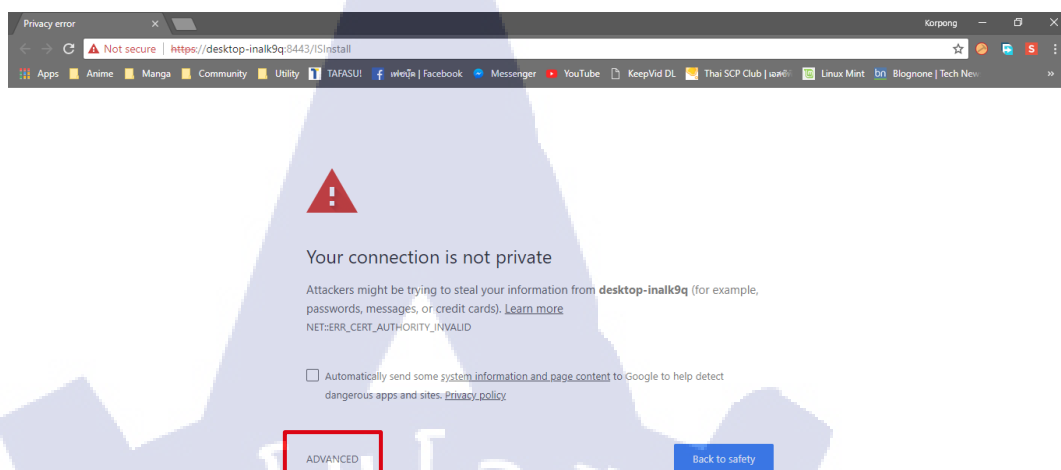
2. เมื่อ Double Click ที่ไฟล์ Set Up แล้ว จะปรากฏหน้าต่าง Command เพื่อแสดงสถานะว่าตัวติดตั้งโปรแกรมนี้กำลังอยู่ในขั้นตอนการติดตั้ง



*** หมายเหตุ อย่าปิดหน้าต่างนี้จนกว่าจะทำการติดตั้งเสร็จเรียบร้อยแล้ว ***

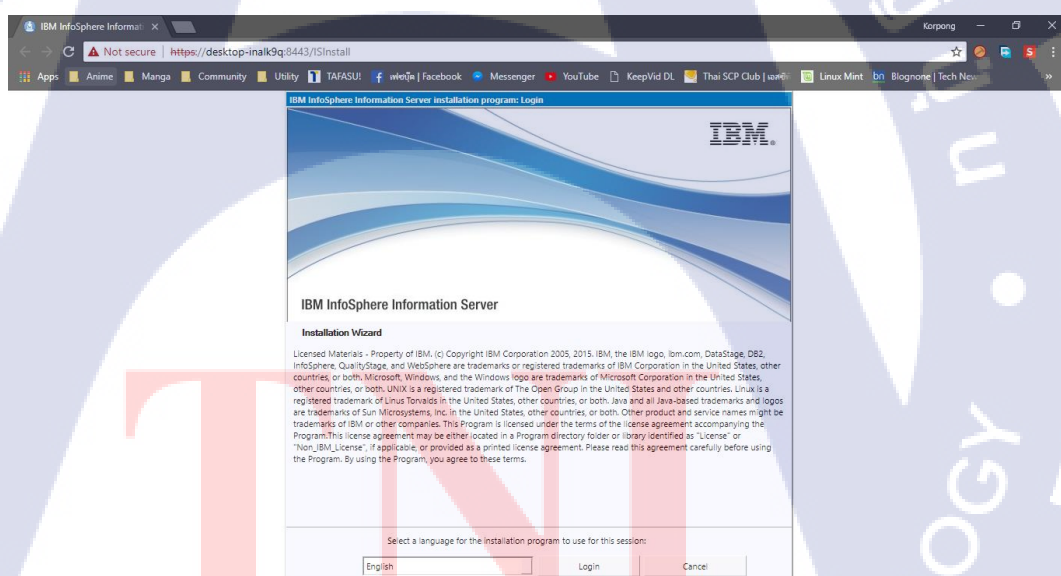
รูปที่ ก.2 แสดงหน้าต่าง Command เพื่อบอกสถานะการติดตั้ง

3. จากนั้นจะปรากฏหน้าต่าง Browser ให้อัตโนมัติดังภาพนี้ โดยให้คลิก Advanced จะปรากฏ link เพื่อเข้าสู่หน้าติดตั้งโปรแกรม



รูปที่ ก.3 หน้า Browser เข้าสู่หน้าติดตั้งโปรแกรม

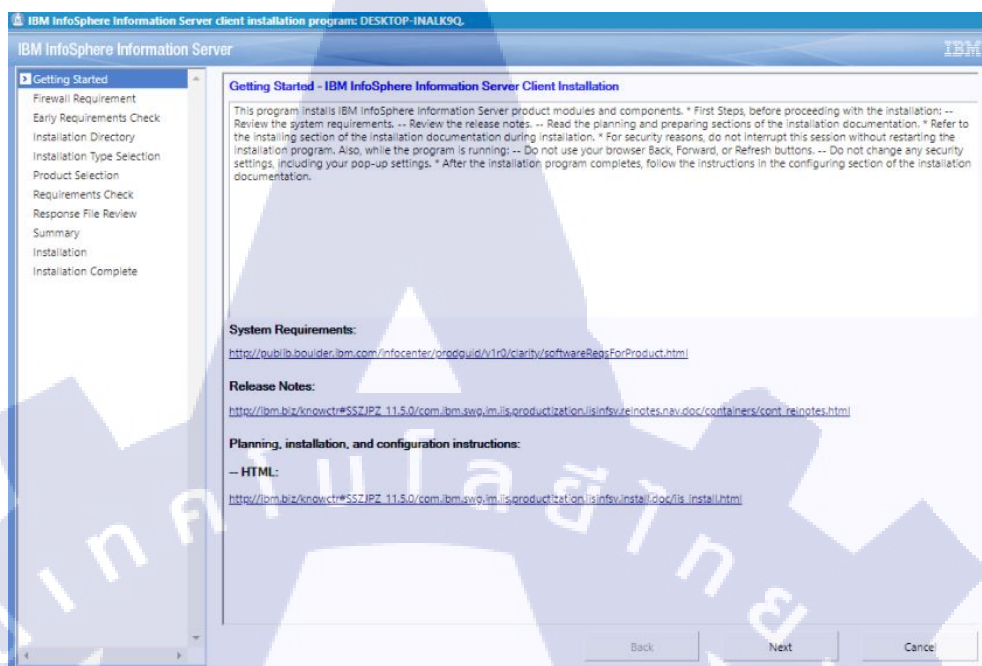
4. จากหน้านี้ให้เลือกภาษาที่ต้องการสำหรับเพื่อทำการติดตั้ง และ Click ที่ Login เพื่อเข้าสู่หน้าถัดไป



รูปที่ ก.4 เลือกภาษาที่ต้องการติดตั้ง

5. Getting Started

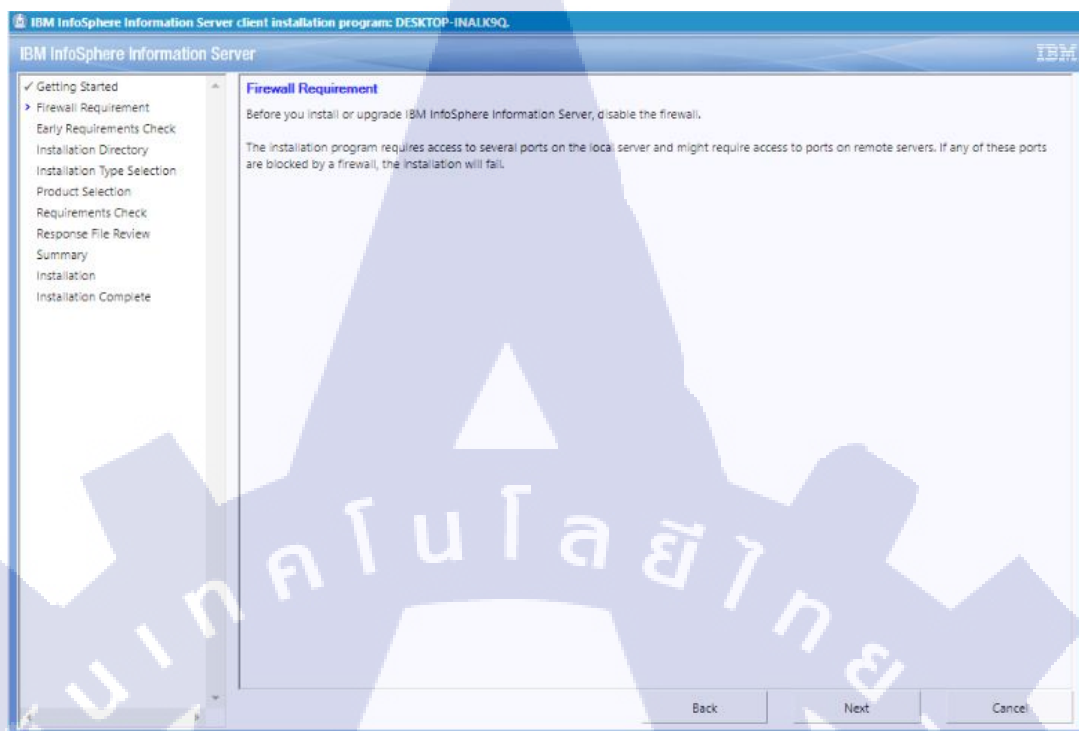
แสดงรายละเอียด ข้อกำหนด และคำแนะนำต่างๆของการติดตั้ง เมื่ออ่านข้อกำหนดแล้วให้กด Next เพื่อดำเนินการต่อ



รูปที่ ก.5 หน้า Getting Start รายละเอียดของการติดตั้ง

6. Firewall Requirement

หากจะทำการติดตั้งหรืออัปเดต IBM InfoSphere Information Server ให้ทำการปิดการใช้งาน Firewall จากนั้นกด Next เพื่อดำเนินการต่อ

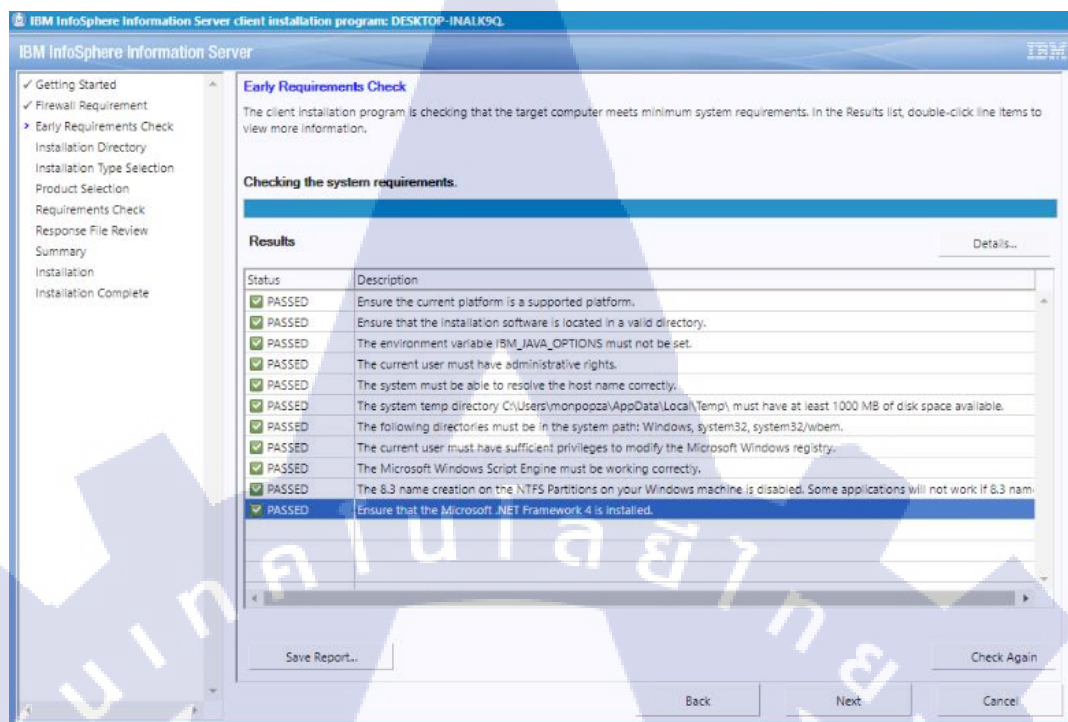


รูปที่ ก.6 Firewall Requirement

7. Early Requirements Check

Check requirements ก่อนการติดตั้งโปรแกรม โดยตรวจสอบว่าเครื่อง Client ตรงตามความต้องการขั้นต่ำ และสามารถผลลัพธ์ได้โดยการ double-click line items เพื่อดูรายละเอียดทั้งหมด จากนั้น กด Next เพื่อดำเนินการต่อ

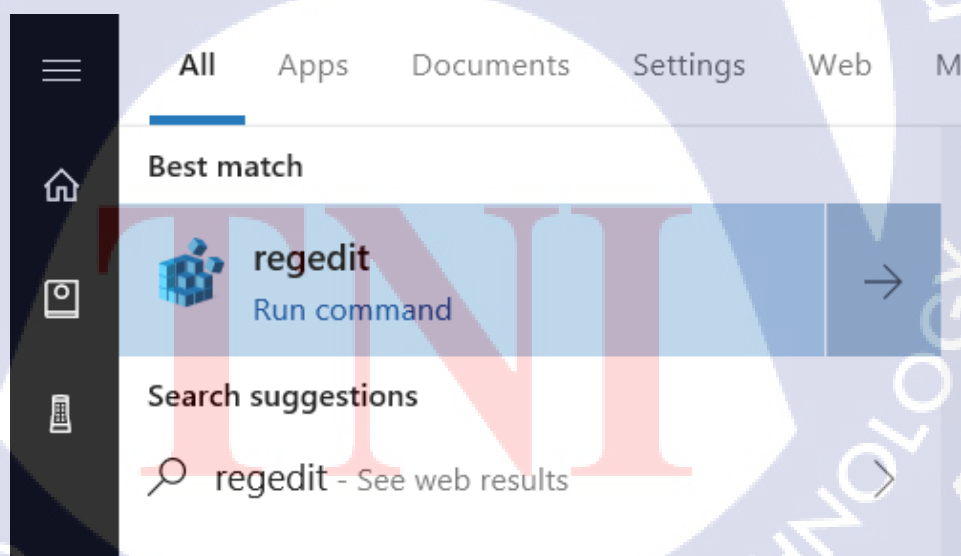
TNI



รูปที่ ก.7 Check requirements

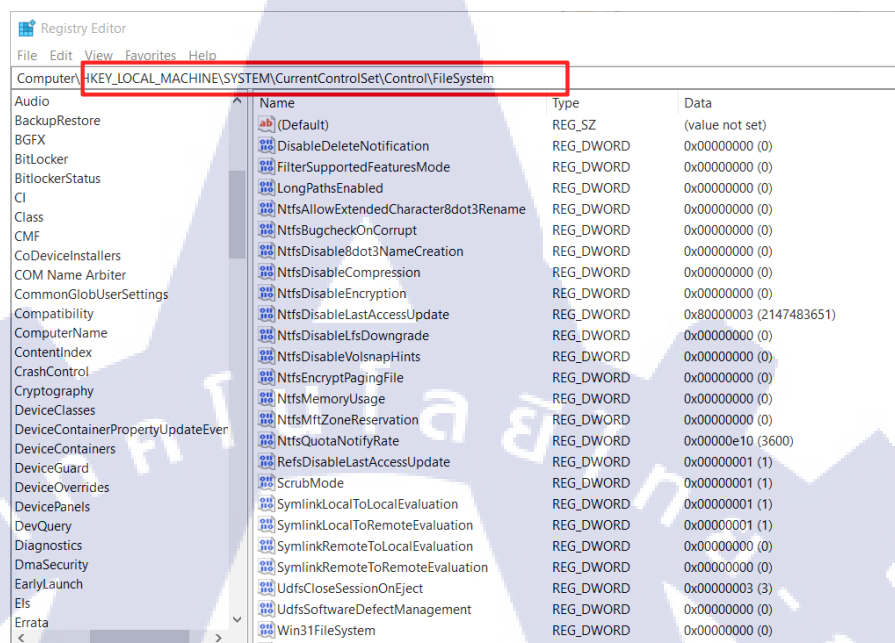
* หมายเหตุ หาก Status ในหน้านี้ในส่วนที่วงไว้เกิดการ Warning ให้ทำการ ดังต่อไปนี้ *

7.1 ทำการเข้าไปยังหน้า Start ค้นหาคำว่า regedit



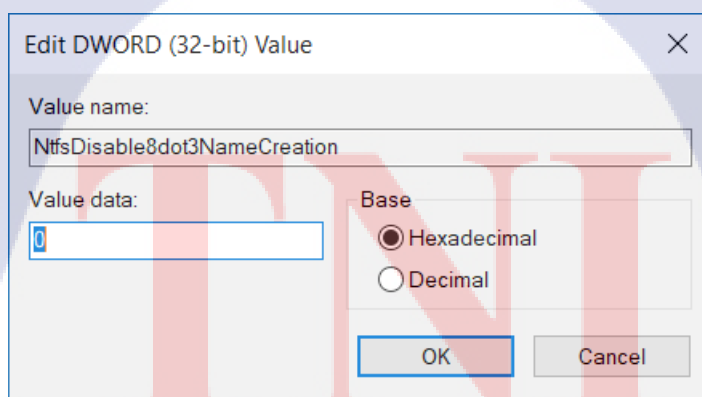
รูปที่ ก.8 Setting Regidit

7.2 จากนั้นเข้าไปยัง Path ตามที่วงไว้ดังภาพ



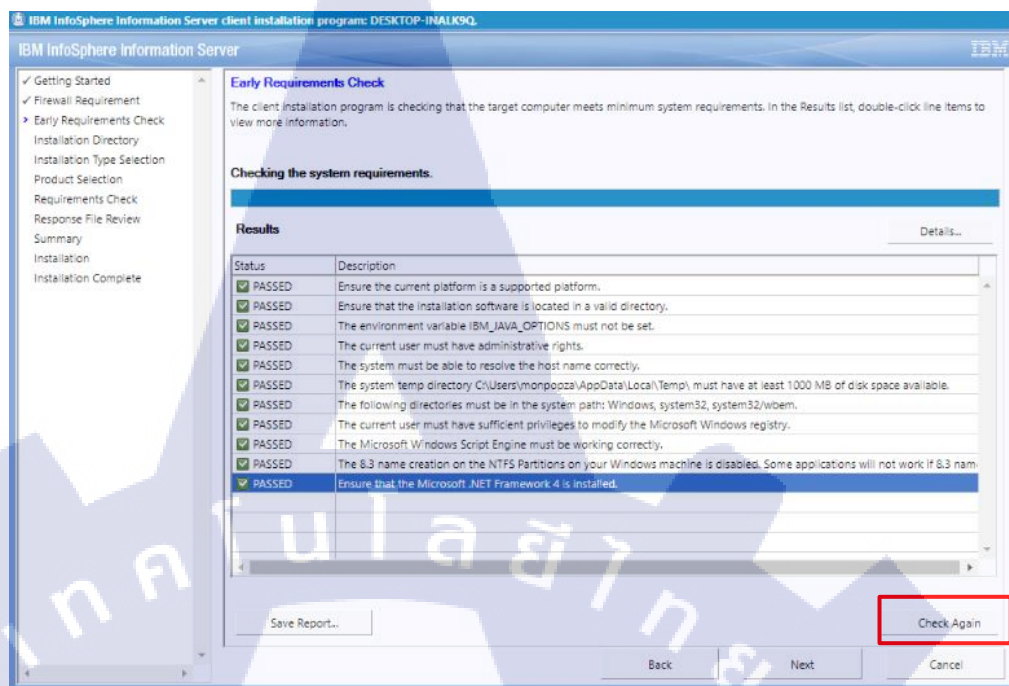
รูปที่ ก.9 Regidit Editor

7.3 ทำการคลิก REG_DWORD ที่ชื่อว่า NtfsDisable8dot3NameCreation และเปลี่ยนค่าให้เป็น 0 จากนั้นกด OK



รูปที่ ก.10 Set Value Ntfs

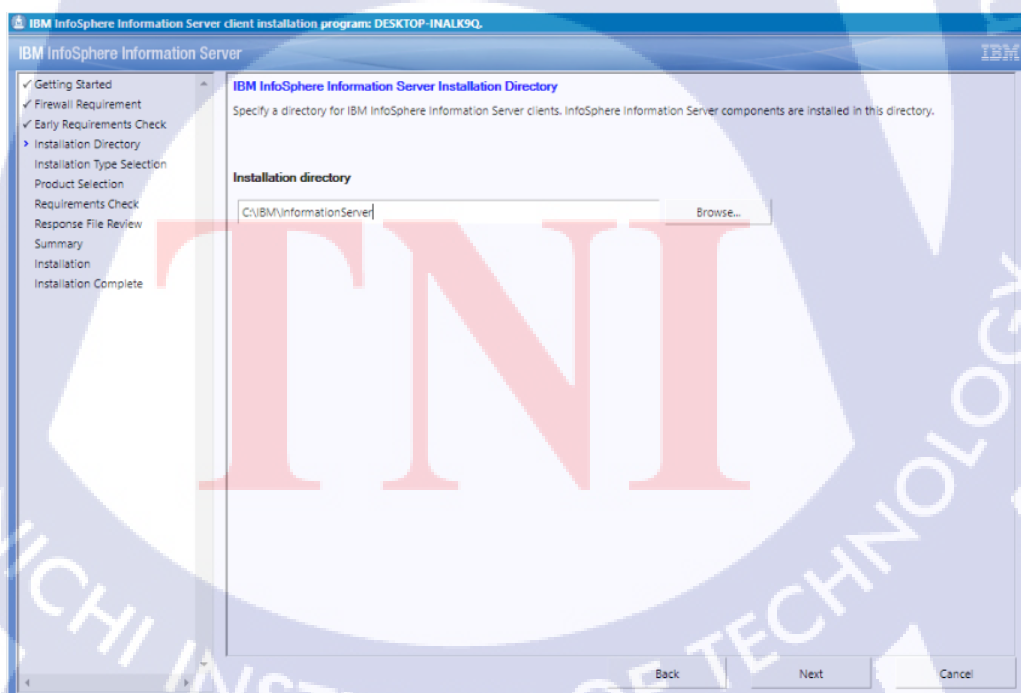
7.4 กลับมายังหน้า Early Requirements Check เพื่อทำการ Check อีกครั้งเป็นอันเสร็จขั้นตอน



รูปที่ ก.11 Early Requirements Check

8. Installation Directory

เลือก Directory ในการจัดเก็บโปรแกรมที่ติดตั้ง จากนั้นกด Next เพื่อดำเนินการต่อ

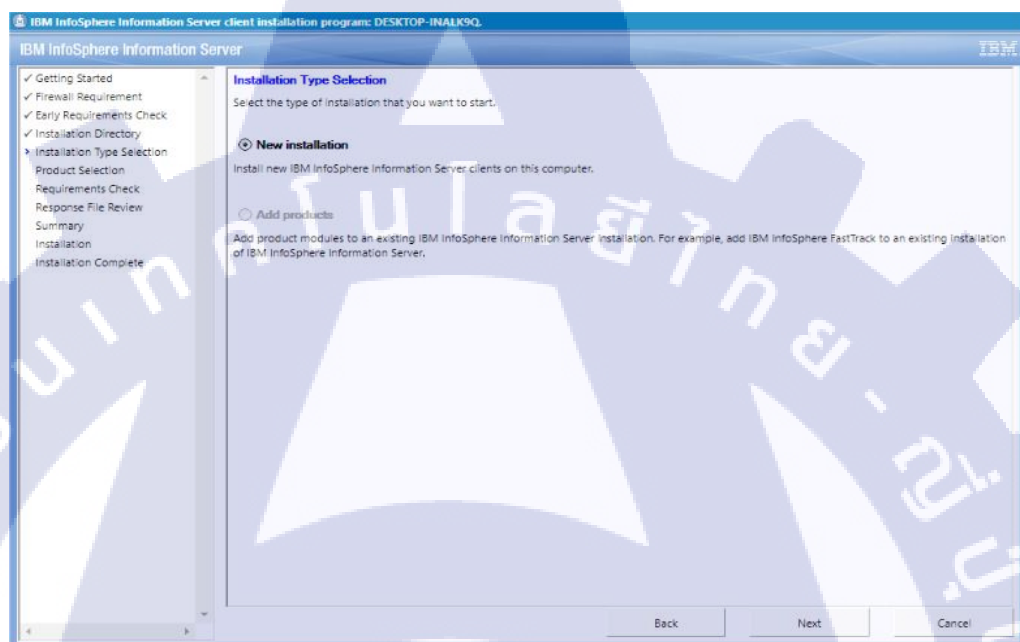


รูปที่ ก.12 Installation Directory

9. Installation Type Selection

เป็นการเลือกประเภทการติดตั้ง

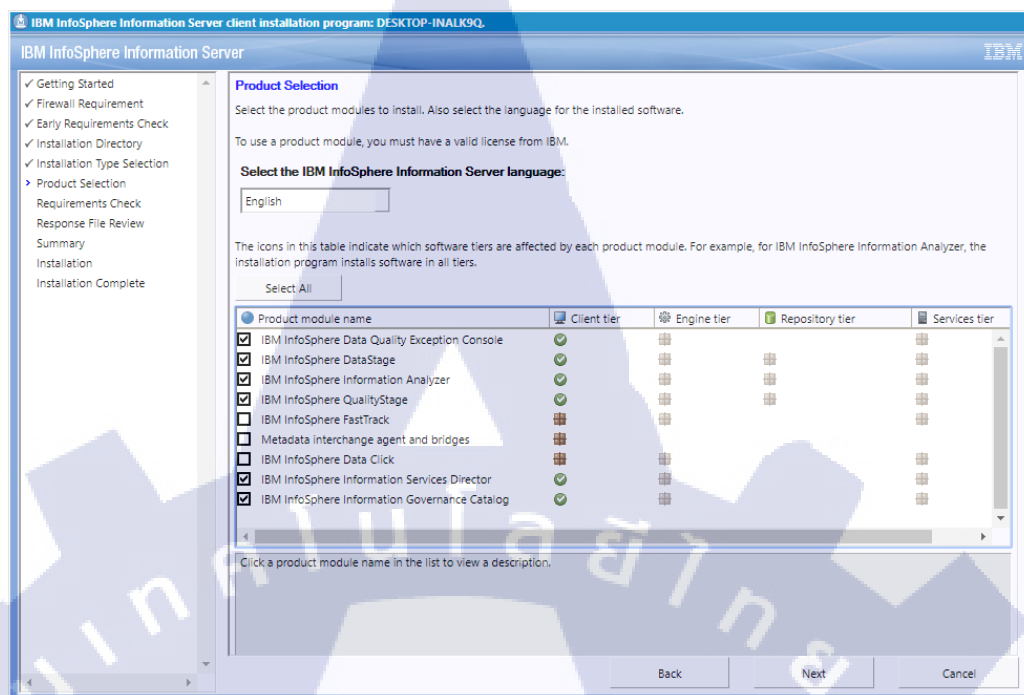
- กรณีที่เป็นการติดตั้งครั้งแรกให้เลือก New Installation
- กรณีเคยติดตั้งแล้วต้องการเพิ่ม product ให้เลือก Add products จากนั้นกด Next เพื่อดำเนินการต่อ



รูปที่ ก.13 Installation Type Selection

10. Product Selection

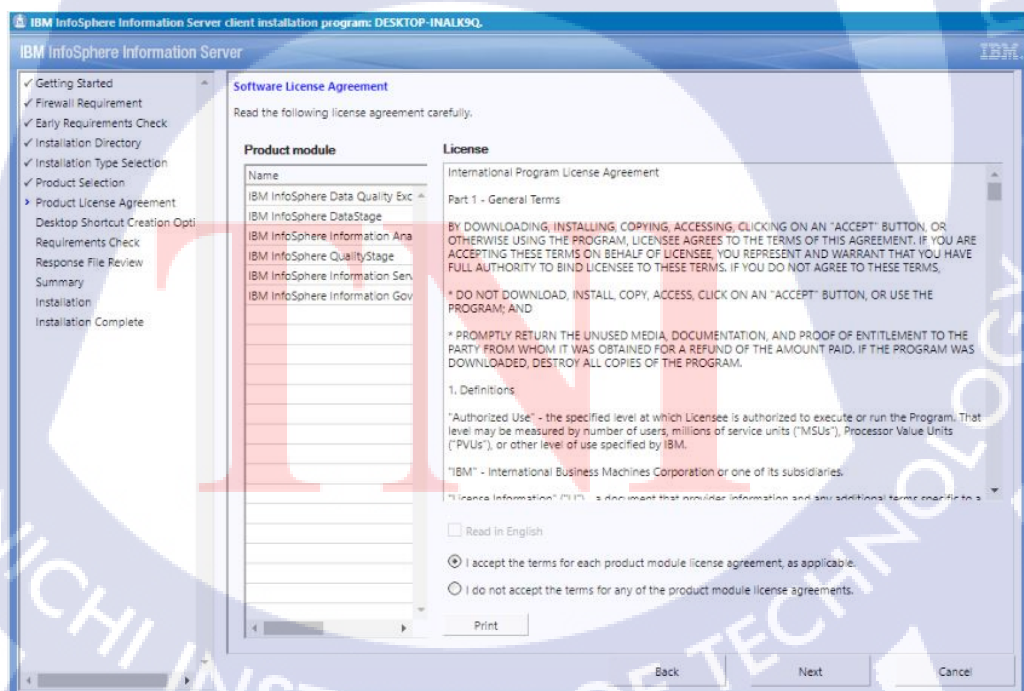
เลือกภาษาในการติดตั้ง และเลือกโปรแกรมที่ต้องการใช้งาน ในที่นี้ให้เลือกแค่ IBM InfoSphere DataStage จากนั้นกด Next เพื่อดำเนินการต่อ



รูปที่ ก.14 Product Selection

11. Software License Agreement

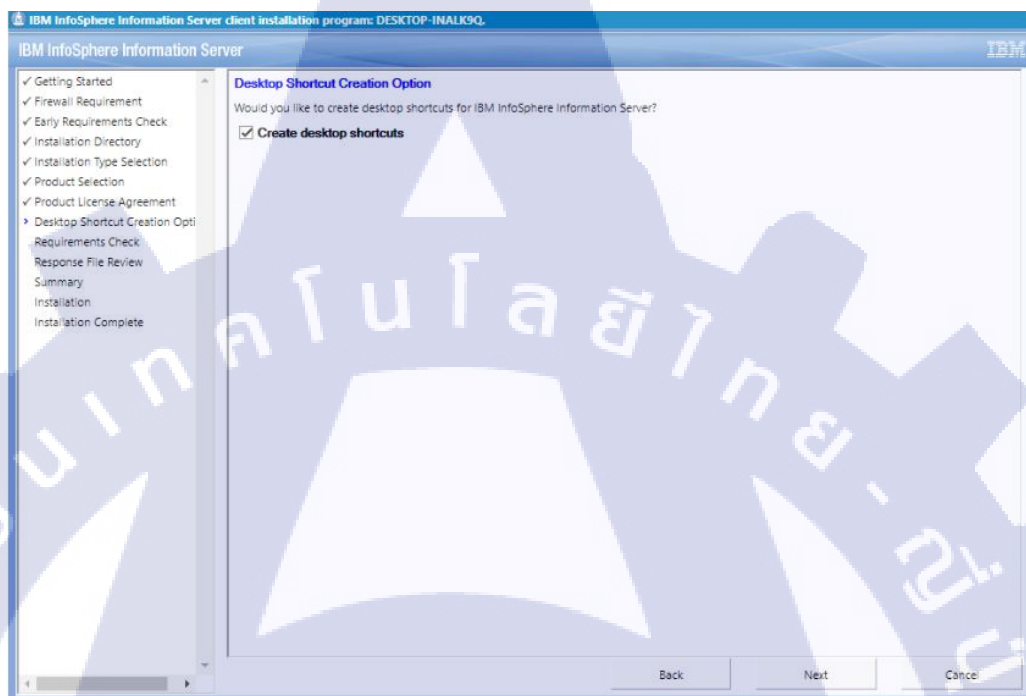
สำหรับ Software License Agreement เมื่ออ่านข้อตกลงแล้ว ให้เลือก accept เพื่อยอมรับข้อกำหนด จากนั้นกด Next เพื่อดำเนินการต่อ



รูปที่ ก.15 Software License Agreement

12. Desktop Shortcut Creation Option

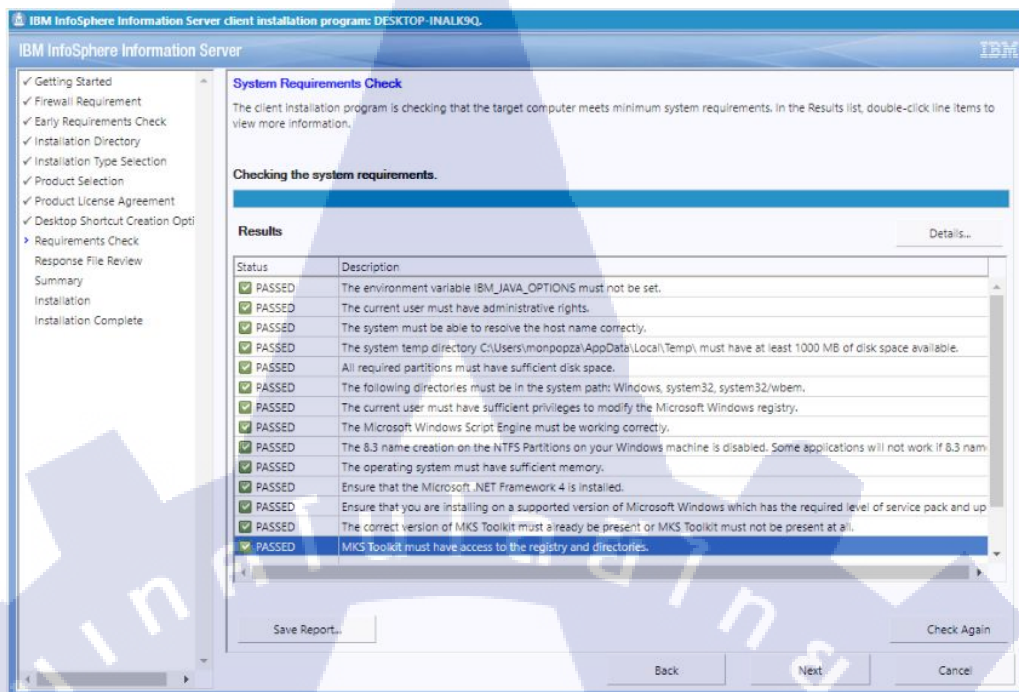
หากต้องการให้ icon ของเครื่องมือทั้งหมดใน IBM InfoSphere Information Server ปรากฏอยู่บน desktop ให้คลิกเลือก Create desktop shortcuts จากนั้นกด Next เพื่อดำเนินการต่อ



รูปที่ ก.16 Desktop Shortcut Creation Option

13. System Requirements Check

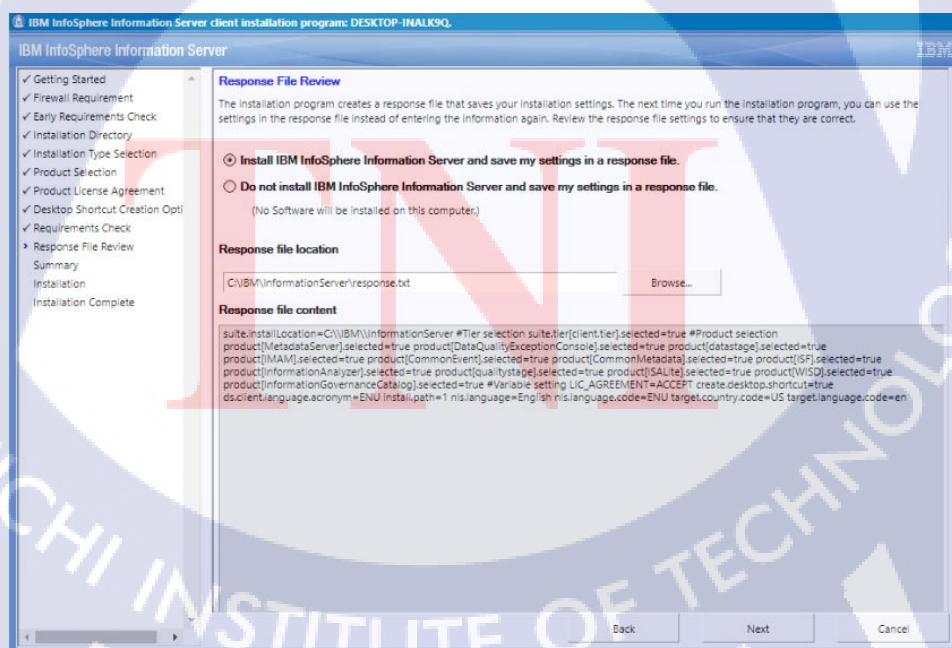
Check Requirements ของระบบอีกครั้งก่อนเริ่มทำการติดตั้ง และสามารถดูรายละเอียดเพิ่มเติมได้โดย double-click line items จากนั้นกด Next เพื่อดำเนินการต่อ



รูปที่ ก.17 System Requirements Check

14. Response File Review

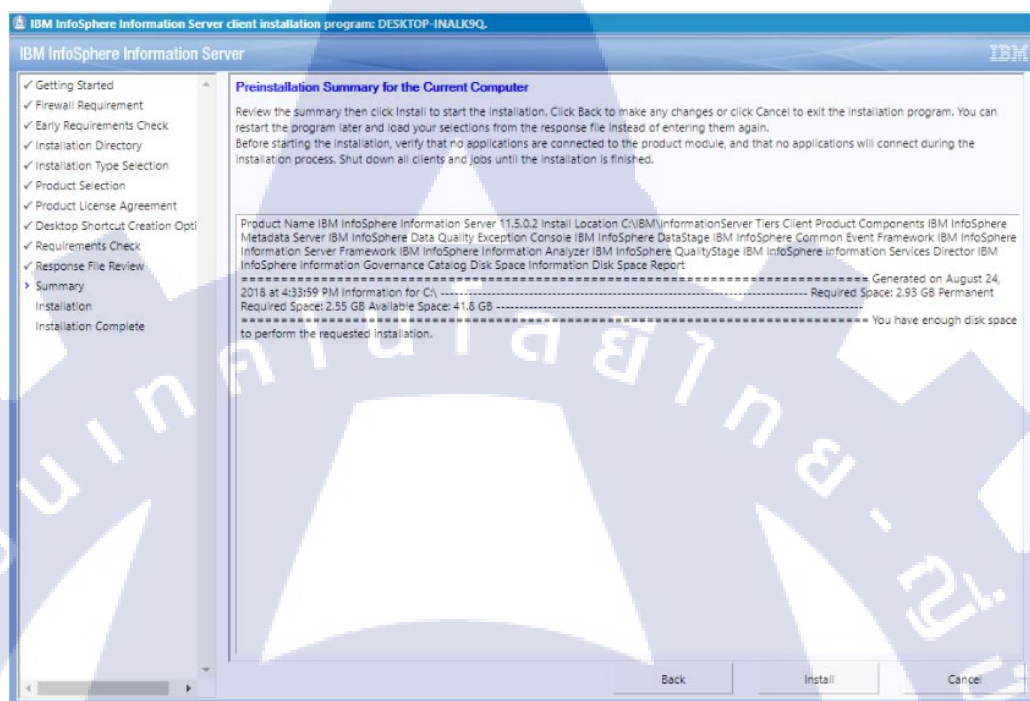
ในขั้นตอนนี้ให้เลือก Install IBM InfoSphere Information Server and save my settings in a response file เพื่อทำการติดตั้งและบันทึกการตั้งค่าที่เลือกไว้ใน response file และเลือก Response file location เพื่อทำการจัดเก็บ response file



รูปที่ ก.18 Response File Review

15. Pre installation Summary for the Current Computer

แสดงผลสรุปก่อนการติดตั้งโปรแกรมของเครื่องคอมพิวเตอร์ปัจจุบัน เมื่อตรวจสอบข้อมูลเรียบร้อยแล้ว ให้เลือก Install เพื่อเริ่มการติดตั้งโปรแกรม

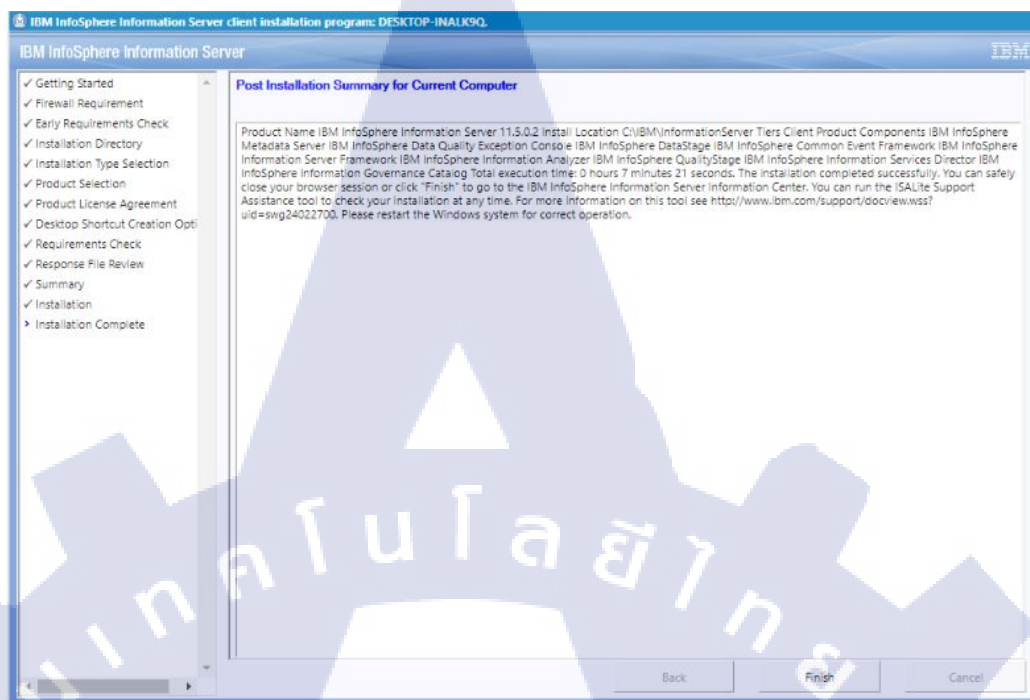


รูปที่ ก.19 Pre installation Summary for the Current Computer

16. Post Installation Summary for the Current Computer

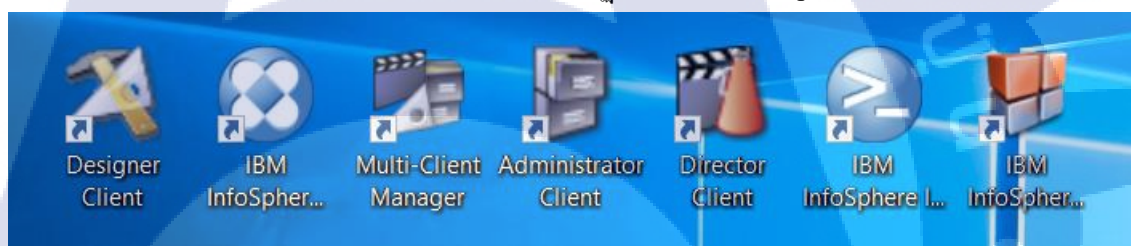
แสดงผลสรุปหลังการติดตั้งโปรแกรม เมื่อตรวจสอบเรียบร้อยแล้ว ให้เลือก Finish เพื่อสิ้นสุดการติดตั้ง

TNI



รูปที่ ก.20 Post Installation Summary for the Current Computer

17. เมื่อดำเนินการติดตั้งโปรแกรมเรียบร้อยแล้ว จะปรากฏ icon บน desktop ดังภาพ



รูปที่ ก.21 Software about ETL



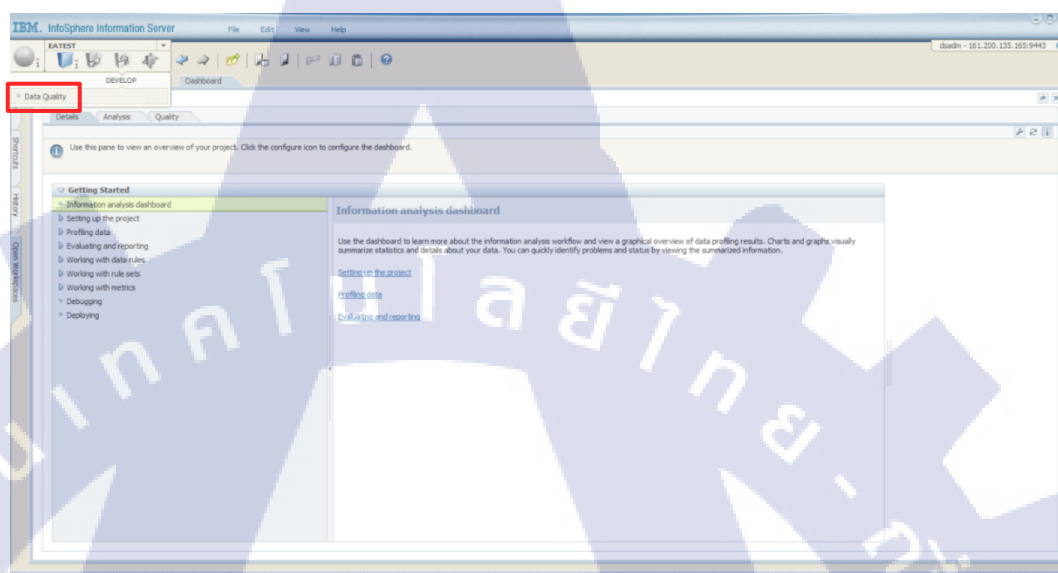
ภาคผนวก ข

การสร้าง Data Rule เพื่อจำกัดขอบเขตของข้อมูลด้วย IBM Analyzer

TNI

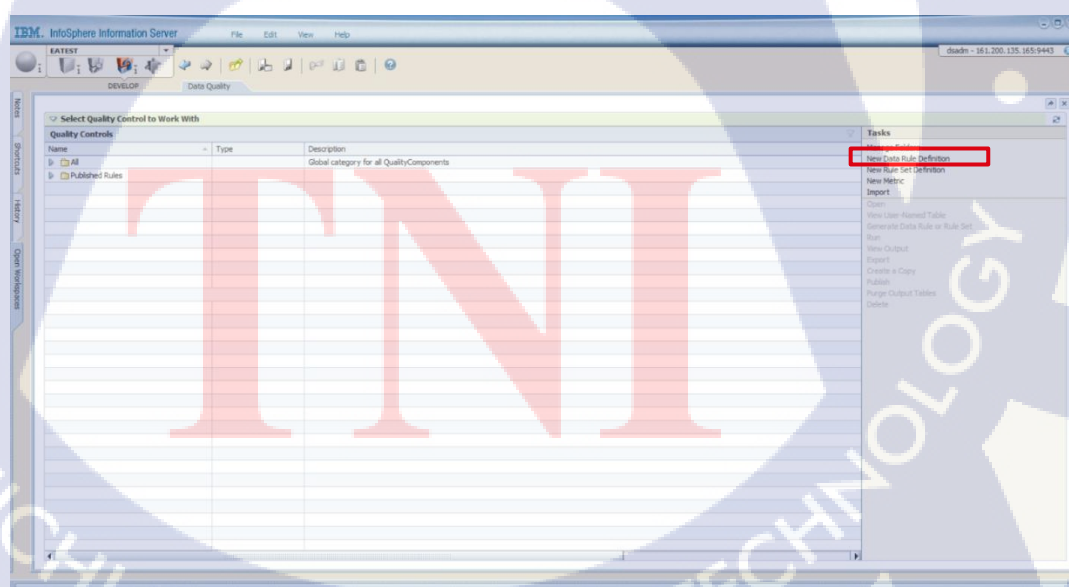
Create Data Rule and Add Rule on Web IBM Analyzer

1. ไปที่แถบ Navigator Menu คลิกไอคอน Develop และคลิกเลือก Data Quality จากแถบ Menu



รูปที่ ข.1 Data Quality

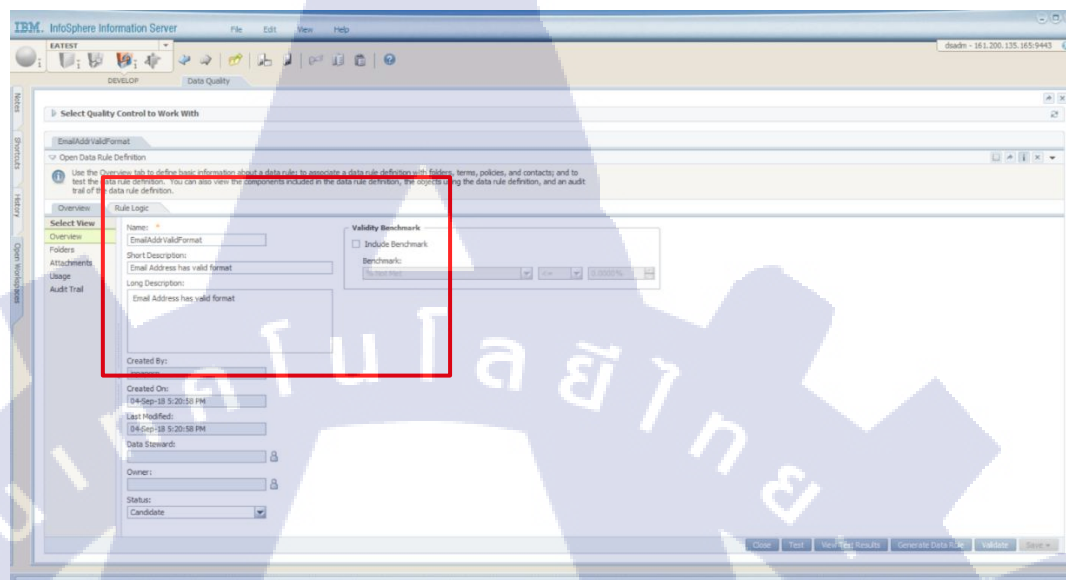
2. ในหน้า Workspace ของ Data Quality ให้คลิกที่ New Data Rule Definition ที่แถบ Tasks ทางด้านขวาของ Workspace เพื่อทำการสร้าง Data Rule Definition



รูปที่ ข.2 Data Rule Definition

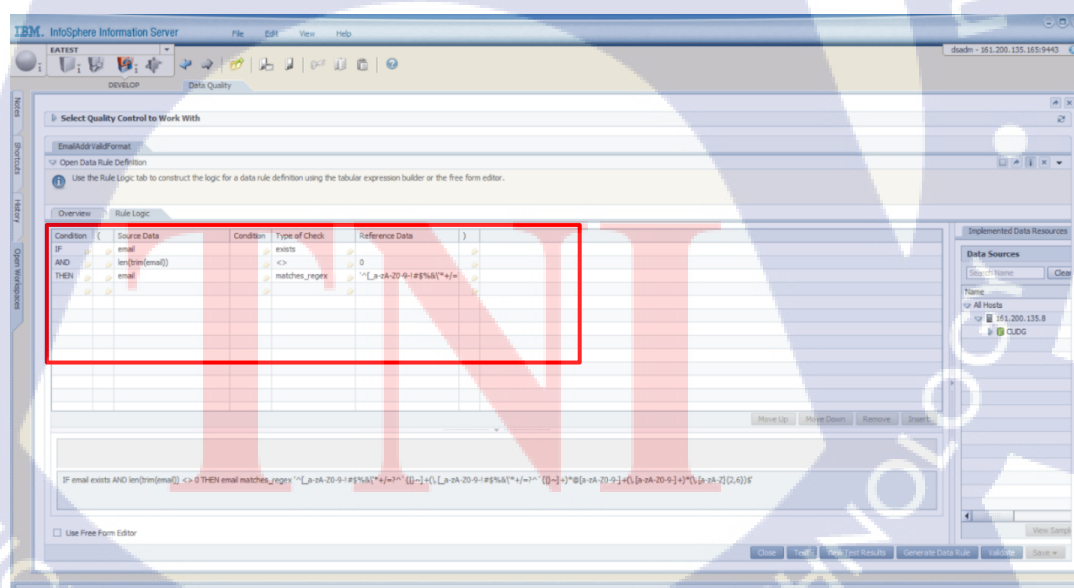
- สร้าง Rule Definition โดยให้กำหนดชื่อของ Rule ใน Overview Tab

*ควรใส่ Description สำหรับ rule เพื่อบรรยายเกี่ยวกับ rule ที่สร้างขึ้นว่าใช้เพื่อกำหนดขอบเขตอะไร



รูปที่ ข.3 Data Rule Definition Description

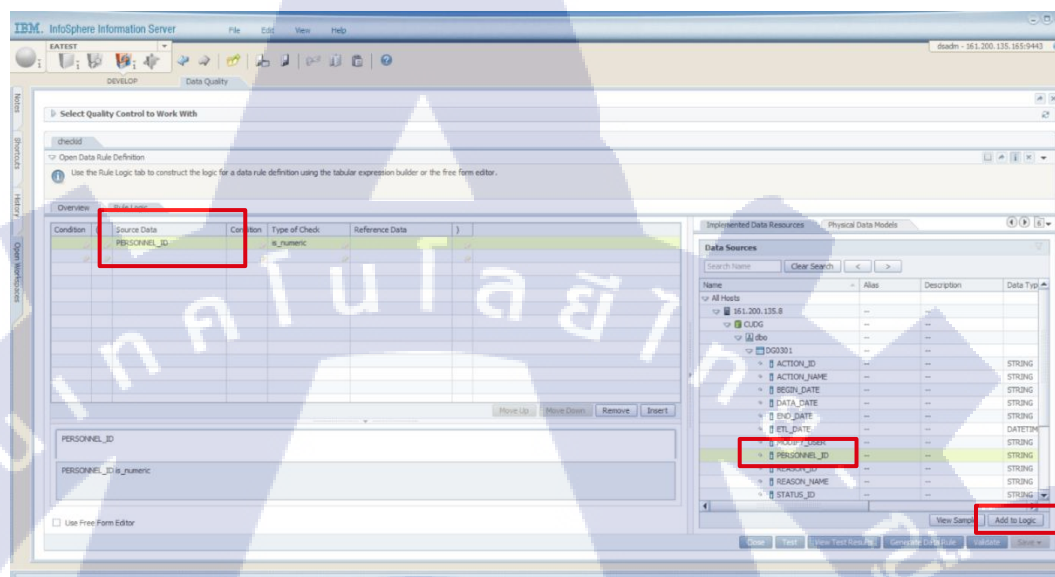
- จากนั้นเลือกเปลี่ยนไปที่แถบ Rule Logic เพื่อกำหนด Condition ที่ต้องการ วิเคราะห์ใน Rule



รูปที่ ข.4 Rule Logic

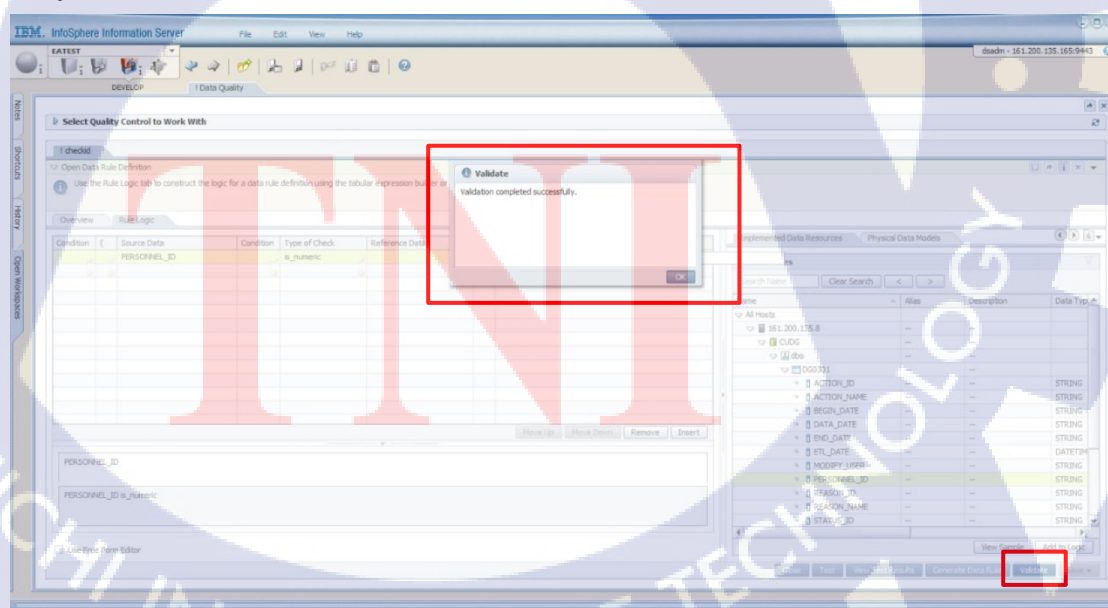
- ในการกำหนด Rule Logic ต้องประกอบไปด้วย Data Source และ Type of check อย่างน้อย 1 รายการ

6. Source Data ใน Rule Logic สามารถ Add column จาก Data Set ที่มีอยู่ หรือ สามารถสร้างตัวแปรขึ้นมาได้
7. โดยคลิกที่ Source Data และเลือก Column จาก Data Source ใน Tasks ทางด้านขวา จากนั้นเลือกคลิกที่ Add Logic



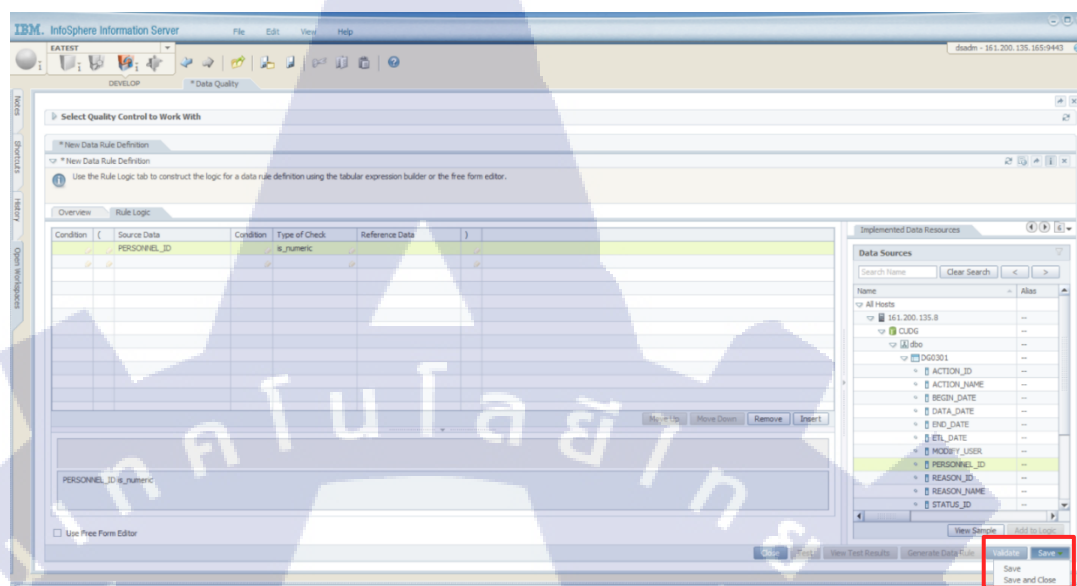
รูปที่ ข.5 Add Logic to Data Source

8. ถัดมาเมื่อกำหนด Rule Logic เสร็จแล้ว ให้คลิกที่ Validate เพื่อเช็คว่ Rule ที่ทำการสร้างนั้นถูกต้องหรือไม่



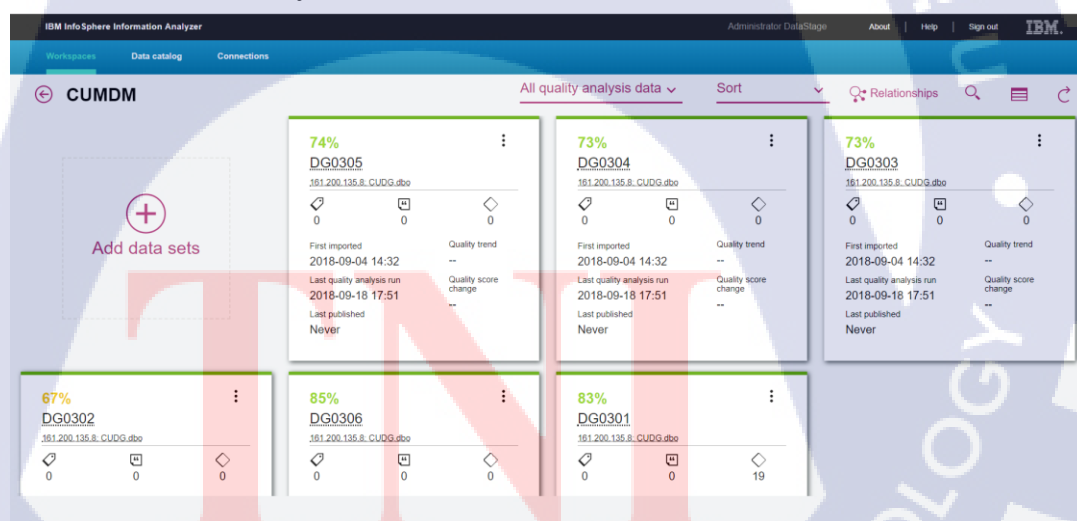
รูปที่ ข.6 Validate Data Rules

9. เมื่อ Validate เสร็จเรียบร้อยแล้วให้คลิก Save ฟังก์ชันด้านล่างสุด เพื่บันทึก Rule Definition ให้เรียบร้อยเป็นอันเสร็จขั้นตอนการสร้าง Data Rule Definition



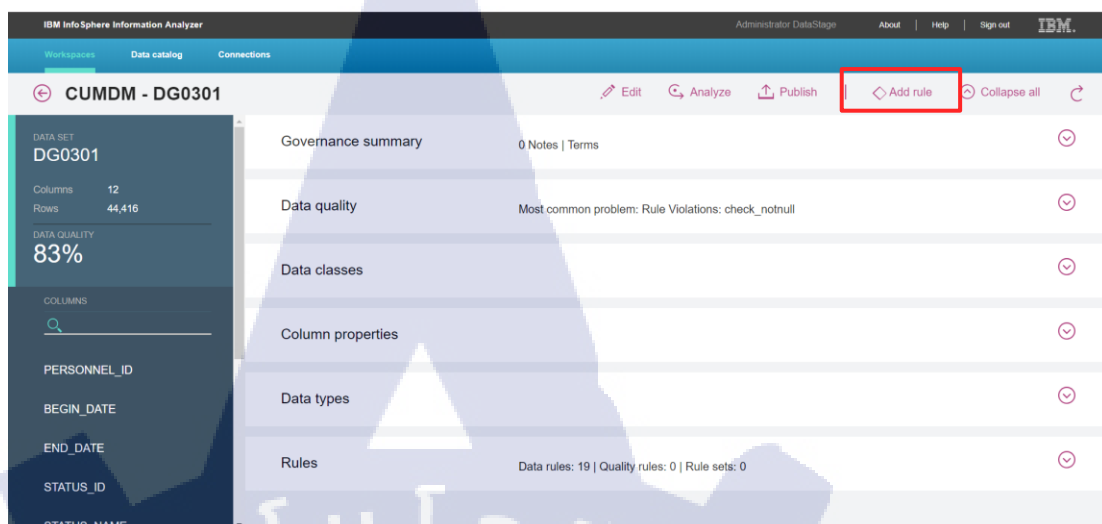
รูปที่ ข.7 Save Data Rules Definition

10. หลังจากการสร้าง Data Rule Definition แล้วให้เข้าสู่ IBM InfoSphere Analyzer บนหน้าเว็บ เพื่อทำการ Add Rule เข้าสู่คอลัมน์หรือตารางที่ต้องการทำ Data Quality



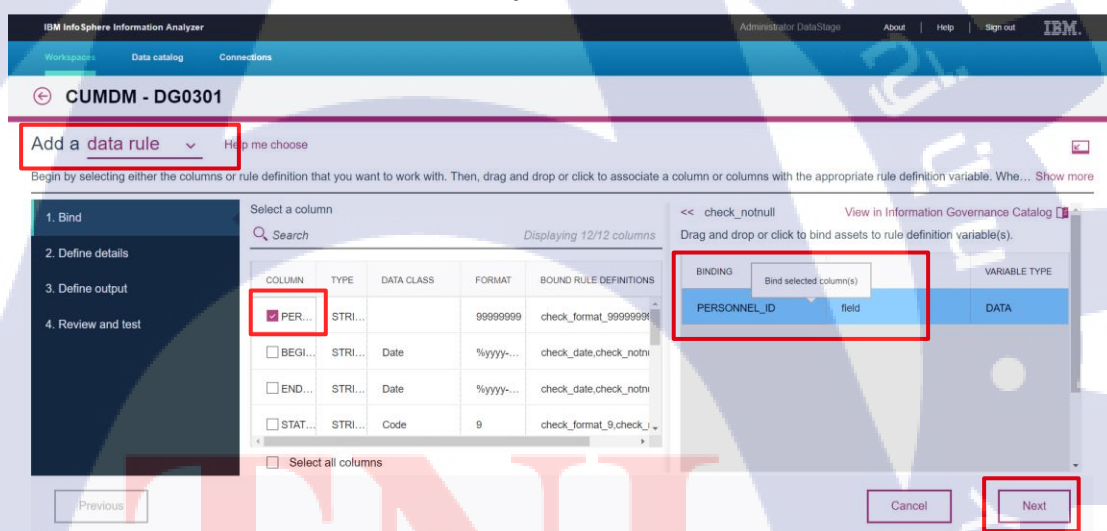
รูปที่ ข.8 หน้า Dashboard Analyzer แสดง Data Sets ทั้งหมด

11. คลิกเลือกตารางที่ต้องการ Add Rule จากนั้นให้เลือก Add Rule ทางฟังก์ชันขวา



รูปที่ ข.9 Add Rule on Web Browser

12. จากภาพให้เปลี่ยนประเภทของ Rule เป็น Data Rules จากนั้นทำการเลือก Rule Definition และ Column เพื่อทำการ Binding Rule ให้จับคู่เข้ากับ Column ที่ต้องการ จากนั้นกด Next



รูปที่ ข.10 Binding Rule

13. ให้ทำการกำหนดชื่อของ Data Rule ที่ทำการ Binding แล้ว

IBM InfoSphere Information Analyzer

Administrator DataStage About Help Sign out IBM.

Workspaces Data catalog Connections

CUMDM - DG0301

Add a **data rule** Help me choose

Begin by selecting either the columns or rule definition that you want to work with. Then, drag and drop or click to associate a column or columns with the appropriate rule definition variable. Who... Show more

1. Bind 2. Define details 3. Define output 4. Review and test

Details for data rule 1 of 1 Define details for next data rule >>

Name * NotNull

Description check personal id not null Bind selected column(s)

Created by Administrator

Steward dsadm Owner dsadm

Previous Cancel Next

รูปที่ ข.11 Create Binding Name

14. เป็นหน้ากำหนด output ซึ่งแสดง output ที่ทำการ binding ไว้ก่อนหน้านี้แล้ว จากนั้นให้ทำการกด Next

IBM InfoSphere Information Analyzer

Administrator DataStage About Help Sign out IBM.

Workspaces Data catalog Connections

CUMDM - DG0301

Add a **data rule** Help me choose

Begin by selecting either the columns or rule definition that you want to work with. Then, drag and drop or click to associate a column or columns with the appropriate rule definition variable. Who... Show more

1. Bind 2. Define details 3. Define output 4. Review and test

Output Content for 1 of 1 Define output for next Data Rule >>

Columns

COLUMN

☐ PERSONNEL_ID

☐ BEGIN_DATE

☐ END_DATE

Add

☐ Select all columns

Selected output

OUTPUT NAME	BINDING	TYPE
PERSONNEL_ID	PERSONNEL...	field column

Advanced options >>

Previous Cancel Next

รูปที่ ข.12 Define Output

15. ในส่วนของหน้านี้จะเป็นการทวนขั้นตอนการ Set Rule ทั้งหมดว่า Add rule ใดเข้ากับ Column ใด เมื่อตรวจสอบแล้วให้ทำการกด Done เป็นอันเสร็จขั้นตอนการ Add Rules

รูปที่ ข.13 Save Add Data Rules

16. ในหน้า Data Set ชุดข้อมูลนี้ ส่วนของ Data Rules จะแสดงรายละเอียดต่างๆ ดังนี้

RULE DEFINITION	RULE DETAILS	RULE STATUS	EXECUTION STATUS	BINDINGS	LAST RUN TIME	NUMBER MET
check_notnull EXPRESSION: field exists and field is not null and field is not null	NotNull_...	Candidate	Not Started	PERSONNEL_ID	--	--
	NotNull_...	Candidate	Not Started	BEGIN_DATE	--	--
	NotNull_...	Candidate	Not Started	END_DATE	--	--
	NotNull_...	Candidate	Not Started	STATUS_ID	--	--

รูปที่ ข.14 Detail Data Rules in Data Sets

ประวัติผู้จัดทำโครงการงาน

ชื่อ – สกุล

อินทร์ภากร กุลเลิศประเสริฐ



วัน เดือน ปีเกิด

16 มีนาคม 2540

ประวัติการศึกษา

ระดับประถมศึกษา

โรงเรียนวิชัยวิทยา ปี 2546-2551

ระดับมัธยมศึกษา

โรงเรียนสตรีวัดอัมพรสวรรค์ ปี 2552-2557

ระดับอุดมศึกษา

คณะเทคโนโลยีสารสนเทศ

สาขาเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีไทย – ญี่ปุ่น ปี 2558-2561

ทุนการศึกษา

- ไม่มี -

ประวัติการฝึกอบรม

Co.,Ltd.

Training Pre-Cooperative Education โครงการสหกิจฯ บริษัท A-HOST

ผลงานที่ได้รับการตีพิมพ์ - ไม่มี -

TNI