

การสาธิตกระบวนการ ETL ของระบบ Business Intelligence ด้วยโปรแกรม IBM® InfoSphere® DataStage® Business Intelligence ETL Process Using IBM® InfoSphere® DataStage® Demonstration

นายศรัณยู ศรีวิจิตรานนท์

(

โครงงานสหกิ<mark>จศึกษ</mark>านี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร ปริญญาวิทย<mark>าศาส</mark>ตรบัณฑิต สาขาวิชาเท<mark>คโนโลยีสา</mark>รสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยี ไทย-ญี่ปุ่น

> พ.ศ. 2559 STITUTE O

การสาธิตกระบวนการ ETL ของระบบ Business Intelligence ด้วยโปรแกรม IBM® InfoSphere® DataStage® Business Intelligence ETL process using IBM® InfoSphere® DataStage® Demomonstration

นายศรัณยู ศรีวิจิตรานนท์

โครงงานสหกิจศึกษานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร วิทยาศาสตรบัณฑิต สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีไทย - ญี่ปุ่น ปีการศึกษา 2559

คณะกรรมการสอบ

.....ประธานกรรมการสอบ

(อาจารย์ สลิลา ชีวกิดาการ)

......กรรมการสอบ

(อาจารย์ ชาญ จารุวงศ์รังสี)

.....อาจารย์ที่ปรึกษา (อาจารย์ ดร. สะพ<mark>รั่งสิ</mark>ทธิ์ มฤทุสาธร)

.....ประช<mark>านสหกิจศึก</mark>ษาสาขาวิชา

(อาจารย์ อมรพัน<mark>ธ์ ชม</mark>กลิ่น)

ลิขสิท<mark>ธิ์ขอ</mark>งสถาบั<mark>นเทค</mark>โนโลยีไทย – ญี่ปุ่น

ชื่อโครงงาน	การสาธิตกระบวนการ ETL ของร	ระบบ Business Intelligence ด้วย
โปรแก	รม IBM® InfoSphere® DataStage	R
ผู้เขียน	นายศรัณยู ศรีวิจิตรานนท์	
คณะวิชา	เทคโนโลยีสารสนเทศ	สาขาวิชา เทคโนโลยีสารสนเทศ
อาจารย์ที่ปรึกษา	คร. สะพรั่งสิทธิ์ มฤทุสาทร	
พนักงานที่ปรึกษา	1. นางสาวปาริชาติ ทองคอนน้อย	
	2. นายระพีพัฒน์ ธนะพัฒน์	
ชื่อบริษัท	บริษัท เอ-โฮสต์ (ประเทศไทย) จำ	ากัด
ประเภทธุรกิจ/สินค้า	การให้บริการในด้าน Application	Hosting, การให้บริการในด้าน IT
Outsourcing, การติดตั้ง ERP. และ Business Intelligence Implementation		

#### บทสรุป

ปัจจุบัน ที่มีการนำระบบ Business Intelligence มาใช้มากขึ้น เพื่อช่วยเหลือในการตัดสินใจในการ บริหารต่าง ๆ ส่งผลให้มีการพัฒนาซอฟต์แวร์ขึ้นมามากมาย ดังนั้นความรู้เกี่ยวกับผลิตภัณฑ์ต่าง ๆ ที่ เกี่ยวข้อง จึงเป็นเรื่องที่จำเป็น ผู้จัดทำโครงงานจึงได้รับมอบหมายให้ศึกษา DataStage® ซึ่งเป็นหนึ่งใน เครื่องมือที่ใช้ในการทำ Business Intelligence ซึ่งใช้ในการแปรรูปข้อมูล นอกจากนี้ DataStage® ยังรองรับ การทำ ELT ซึ่งเป็นเทคนิคการแปรรูปข้อมูลที่พัฒนามาจากการทำ ETL เนื่องจากพัฒนาการของ Hardware ต่าง ๆ ซึ่งมีข้อดีในการย้อนหาด้นทางของข้อมูลได้ง่ายกว่า ETL จึงทำให้ DataStage® เป็นอีกหนึ่งโปรแกรม ที่น่าศึกษา เพื่อรองรับการให้บริการแก่ลูกก้าภายในอนากต

จากการคำเนินงานจัดทำร<mark>ะบบ</mark>จำลองนั้น ได้พบว่า Stage ต่าง ๆ ที่มีให้ใช้นั้น มีปริมาณที่มาก ทำให้ ไม่สามารถจะจัดทำคู่มือฉบับสมบูรณ์ได้ จึงเน้นไปที่ Stage ที่มีแนวโน้มจะได้ใช้งานบ่อย ๆ นอกจากนี้ใน การใช้งานบาง Stage นั้น จำเป็นต้องมี Library File แต่บางไฟล์ก็ไม่สามารถหามาติดตั้งได้ ทำให้ไม่สามารถ จำลองวิธีการใช้งานได้ ผู้จัดทำโครงงานจึงเปลี่ยนไปใช้ Stage อื่นที่มีลักษณะการใช้งานที่คล้ายกลึงกัน

Project's name	Business Intelligence ETL Process Using IBM® InfoSphere®
	DataStage® Demonstration
Writer	Mr.Sarunyoo Srivichitranond
Faculty	Faculty of Information Technology, Information Technology
Faculty Advisor	Dr. Saprangsit Mruetusatorn
Job Supervisor	1. Ms.Parichart Thongdonnoy
	2. Mr.Rapeepat Tanapat
Company's name	A-HOST Company Limited
Business Type / Product	Application Hosting Service, IT Outsourcing Service, ERP
Implementation, Business Intellig	gence Implementation

#### **Summary**

Nowadays, the Business Intelligence implementation is demanded by many businesses to support the managers' decision. There are many developers trying to meet with the demand of Business Intelligence software, thus there are many tools developed to gain the market share. As a company that provides the Business Intelligence implementation, the knowledge about Business Intelligence software is crucial for the company. DataStage® is one of the software applied to extract the information from data to use with the Business Intelligence later. The DataStage® also supports extraction methods such ETL and ELT.

While creating the demonstration, there are a lot of stages to work on, so this document covers only important stages and DataStage® signature stages. There are some stages that need library files in order to work, but the files are no longer published or distributed, therefore the substitute stage is selected and presented in this report instead.

### กิตติกรรมประกาศ

ในการที่ข้าพเจ้าได้มาสหกิจศึกษา ณ บริษัท เอ-โฮสต์ จำกัด ซึ่งเป็นความร่วมมือระหว่างทางบริษัท กับสถาบันเทคโนโลยีไทย-ญี่ปุ่น ตั้งแต่วันที่ 30 พฤษภาคม พ.ศ. 2559 ถึงวันที่ 30 กันยายน พ.ศ. 2559 ได้ทำ ให้ข้าพเจ้าได้เรียนรู้ประสบการณ์ต่าง ๆ ความรู้จากการทำงานจริง ซึ่งมีค่าอย่างมากมาย และส่งผลให้ข้าพเจ้า สามารถนำสิ่งต่าง ๆ เหล่านั้น มาใช้พัฒนาทักษะของตนเอง สำหรับรายงานการปฏิบัติงานสหกิจศึกษาใน ครั้งนี้ สามารถสำเร็จลุล่วงได้ด้วยดีจากความร่วมมือและการสนับสนุนจากหลายฝ่ายดังนี้

1.คุณบุญประสิทธิ์ ตั้งชัยสุข ที่เห็นความสำคัญของการสหกิจศึกษา และได้ให้โอกาสที่มีคุณค่ายิ่ง แก่ข้าพเจ้าในการเข้ามาสหกิจศึกษาที่บริษัทแห่งนี้

 2.คุณสุชัย เย็นฤดี ที่ได้จัดการฝึกอบรมให้ความรู้ก่อนปฏิบัติสหกิจ เพื่อฝึกความพร้อมก่อนการ ทำงานและให้คำแนะนำความรู้ต่างๆ ตลอดระยะเวลาที่ปฏิบัติงานสหกิจศึกษา

3.คุณระพีพัฒน์ ธนะพัฒน์, คุณปาริชาติ ทองดอนน้อย (พนักงานที่ปรึกษา) ที่ได้ให้ความกรุณารับ ข้าพเจ้าเข้ามาเป็นส่วนหนึ่งของแผนก Academy and BI ในการปฏิบัติงานสหกิจศึกษาในครั้งนี้ รวมทั้งให้ คำปรึกษาและคำแนะนำเกี่ยวกับความรู้หรือการทำงานต่าง ๆ

4.อาจารย์ คร. สะพรั่งสิทธิ์ มฤทุสาธร อาจารย์ที่ปรึกษา ที่ได้ให้ความช่วยเหลือในด้านต่าง ๆ ระหว่างทำสหกิจ ไม่ว่าจะเป็นการถ่ายทอดความรู้ และการให้คำแนะนำในด้านต่าง ๆ ตลอดระยะเวลาที่ได้ ปฏิบัติสหกิจศึกษา จวบจนกระทั่งเสร็จสิ้นกระบวนการสหกิจศึกษา

5.อาจารย์ อมรพันธ์ ชมกลิ่น ประธานสหกิจศึกษาประจำสาขาวิชาเทคโนโลยีสารสนเทศ ที่ได้ช่วย เตรียมความพร้อมในด้านต่าง ๆ ก่อนเริ่มการปฏิบัติสหกิจศึกษา และได้ช่วยให้กำปรึกษาในขั้นตอน กระบวนการต่าง ๆ ของสหกิจศึกษาเป็นอย่างดี

และทุกคนในแผนก Academy and BI รวมไปถึงบุคลากรของสถาบันเทคโนโลยีไทย-ญี่ปุ่น และ บุถคลท่านอื่น ๆ ที่มิได้กล่าวนามที่ได้ให้คำแนะนำช่วยเหลือในการปฏิบัติงานสหกิจและจัดทำรายงานฉบับ นี้ให้สำเร็จไปได้ด้วยดี ขอขอบคุณ <mark>ไว้ ณ</mark> ที่นี้

> นายศรัณยู ศรีวิจิตรานนท์ ผู้จัดทำ

## สารบัญ

จ

Ô

บทสรุป		ๆ
Summary		የ
กิตติกรรมประกาศ		3
สารบัญ		จ
สารบัญรูป		م م
สารบัญตาราง	น เ ล <i>ย</i> ั	ฑ

# บทที่

บทที่ 1 บทนำ
1.1 ชื่อและที่ตั้งของสถานประกอบการ [1]1
<ol> <li>1.2 ลักษณะธุรกิจของสถานประกอบการ หรือการให้บริการหลักขององค์กร</li></ol>
1.3 รูปแบบการจัดองค์กรและการบริหารองค์กรแบ่งตามประเภทการให้บริการที่รองรับ
1.4 ตำแหน่งและหน้าที่งานที่นักศึกษาได้รับมอบหมาย
1.5 พนักงานที่ปรึกษา และ ตำแหน่งของพนักงานที่ปรึกษา8
1.6 ระยะเวลาที่ปฏิบัติงาน
1.7 ที่มาและความสำคัญของปั <mark>ญหา</mark>
1.8 วัตถุประสงก์หรือจุดมุ่งหม <mark>ายขอ</mark> งโกรงงาน
<ol> <li>4.9 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงงานที่ได้รับมอบหมาย</li></ol>
1.10 นิยามศัพท์เฉพาะ

# สารบัญ (ต่อ)

บทที่ 2 ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน	10
ทฤษฎีที่เกี่ยวข้องกับการแปรรูปข้อมูล 2.1	10
2.2 เทคโนโลยีที่ใช้ในการปฏิบัติงาน	15
บทที่ 3 แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน	33
3.1 แผนงานการปฏิบัติงาน	33
3.2 รายละเอียคที่นักศึกษาปฏิบัติในการปฏิบัติงาน	
3.3 ขั้นตอนการดำเนินงานที่นักศึกษาปฏิบัติงาน	34
บทที่ 4 สรุปผลการดำเนินงาน การวิเคราะห์และสรุปผลต่าง ๆ	37
4.1 ส่วนประกอบต่าง ๆ ของ IBM® InfoSphere® DataStage®	37
4.2 วิธีการใช้งาน Stage ต่าง ๆ ของ DataStage®	50
บทที่ 5 บทสรุปและข้อเสนอแนะ	100
5.1 สรุปผลการดำเนินงาน	100
5.2 ปัญหาที่พบ	
5.3 แนวทางการแก้ไขปัญหา	101
5.4 ข้อเสนอแนะจากการคำเนินงาน	101
Ay sect	
เอกสารอ้างอิง	102

# สารบัญ (ต่อ)

ภาคผนวก กสไลด์ประกอบการนำเสนอ	8
ประวัติผู้จัดทำโครงงาน รับโลส	5
n n u la gin	
E TRIS	
10 AN INSTITUTE OF TECHNOL	

# สารบัญรูป

รูปที่ 1. 1 สถานที่ตั้ง บริษัท เอ-โฮสต์ จำกัด1
รูปที่ 1. 2 A-HOST Proud Awards
รูปที่ 1. 3 โครงสร้างขององค์กร แบ่งตามรูปแบบบริการที่รองรับ6
รูปที่ 2. 1 แผนภาพลักษณะการทำงานแบบ ETL10
รูปที่ 2. 2 แผนภาพลักษณะการทำงานแบบ ELT11
รูปที่ 2. 3 แผนภาพลักษณะการทำงานแบบ Data Pipelining12
รูปที่ 2. 4 แผนภาพลักษณะการทำงานแบบ Data Partitioning
รูปที่ 2. 5 แผนภาพลักษณะการทำงานแบบ Data Repartitioning13
รูปที่ 2. 6 แผนภาพลักษณะการทำงานแบบ Data Repartitioning13
รูปที่ 2. 7 แผนผังโครงสร้างของ InfoSphere®14
รูปที่ 2. 8 แผนผังขั้นตอนการพัฒนา Job17
รูปที่ 2. 9 ตัวอย่างการแสดง Stage ที่เกิดความผิดพลาดขึ้น
รูปที่ 2. 10 แผนผังขั้นตอนการทคสอบ Job19
รูปที่ 2. 11 ตัวอย่างการใช้งาน Peek Stage20
รูปที่ 2. 12 แผนผังขั้นตอนการติดตั้ง Job20
รูปที่ 2. 13 แผนผังวงจรชีวิตของ Job21
รูปที่ 2. 14 ตัวอย่างการใช้ Netezza Connector22
รูปที่ 2. 15 ตัวอย่างการนำข้ <mark>อ</mark> มูลจา <mark>ก W</mark> eb Service <mark>มาช่ว</mark> ยทำให้ข้ <mark>อม</mark> ูลมีคว <mark>ามส</mark> มบูรณ์มากขึ้น
รูปที่ 2. 16 ตัวอย่างการใช้งาน Join <mark>Stag</mark> e24
รูปที่ 2. 17 ตัวอย่างการใช้งาน Join <mark>Stag</mark> e ในการ Join ข้อม <mark>ูลจากฐา</mark> นข้อมู <mark>ล Or</mark> acle และ Teradata25
รูปที่ 2. 18 ตัวอย่างการใช้งาน Agg <mark>rega</mark> tor Stage25
รูปที่ 2. 19 ตัวอย่างการใช้งาน Pivot Enterprise Stage
รูปที่ 2. 20 ตัวอย่างการใช้งาน Change Capture Stage
รูปที่ 2. 21 ตัวอย่างการใช้งาน Lookup Stage27
รูปที่ 2. 22 ตัวอย่างการใช้งาน ILOG JRules Connector Stage27

รูปที่ 2. 23 ตัวอย่างการใช้งาน Standardize Stage
รูปที่ 2. 24 ตัวอย่างการใช้งาน Address Verification Stage
รูปที่ 2. 25 ตัวอย่างการใช้งาน One Source Match Stage
รูปที่ 2. 26 ตัวอย่างการใช้งาน Data Rules Stage29
รูปที่ 2. 27 ตัวอย่างการใช้งาน ISD Input Stage เพื่อดึงข้อมูลจาก Web Service
รูปที่ 2. 28 ตัวอย่างการใช้งาน WebSphere MQ Connector Stage
รูปที่ 2. 29 ตัวอย่างเชื่อมต่อระหว่าง Hive Stage กับ Netezza Connector Stage
รูปที่ 2. 30 ตัวอย่างการนำ Job ต่าง ๆ มาจัดเรียงเป็น Sequence Job
รูปที่ 4. 1 ตัวอย่างการใช้ Sort Stage ด้วยการประมวลผลแบบ Sequential (ซ้าย) และแบบ Parallel (ขวา)37
รูปที่ 4. 2 ตัวอย่างการตั้งค่า Preserve Partitioning เป็น Clear
รูปที่ 4. 3 สัญลักษณ์การเปลี่ยนรูปแบบการประมวลผลจาก Sequential เป็น Parallel (วงกลมทางซ้าย) และ
เปลี่ยนแปลงจาก Parallel เป็น Sequential (วงกลมทางขวา)
รูปที่ 4. 4 หน้าค่างเลือกประเภทงานที่จะสร้าง

รูปที่ 4. 16 กำหนดชื่อ Data Connection43
รูปที่ 4. 17 Data Connection Tab Parameter
รูปที่ 4. 18 เลือกประเภทของ Connector
รูปที่ 4. 19 ช่องกรอกรหัสระหว่างทำ Data Connection45
รูปที่ 4. 20 Path ที่ต้องการบันทึก Data Connection45
รูปที่ 4. 21 เมื่อเสร็จสิ้นการบันทึก Data Connection46
รูปที่ 4. 22 วิธีการสร้าง Parallel Job ใหม่
รูปที่ 4. 23 Canvas ของ Parallel Job47
รูปที่ 4. 24 การบันทึก Parallel Job47
รูปที่ 4. 25 กำหนด Path ที่ต้องการทำการบันทึก Parallel Job
รูปที่ 4. 26 เมื่อเสร็จสิ้นการบันทึก Parallel Job
รูปที่ 4. 27 ตัวอย่าง Data Definition
รูปที่ 4. 28 ตำแหน่งโฟลเดอร์ต่าง ๆ ที่บรรจุ Stage สำหรับ Parallel Job
รูปที่ 4. 29 การ Clear Preserve Partitioning เพื่อรองรับการประมวลผลแบบ Sequential
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)
รูปที่ 4. 30 รูป DB2 Connector Stage (Output)

รูปที่ 4. 40 หน้าต่าง Configure ของ Unstructured Data Stage (Input)	61
รูปที่ 4. 41 รูป Sequential File Stage (Output)	62
รูปที่ 4. 42 Output Properties ของ Sequential File Stage	62
รูปที่ 4. 43 รูป Sequential File Stage (Input)	63
รูปที่ 4. 44 Input Properties ของ Sequential File Stage	
รูปที่ 4. 45 ตัวอย่างข้อมูลที่บันทึกด้วย Sequential File Stage	65
รูปที่ 4. 46 รูป Transformer Stage	66
รูปที่ 4. 47 ส่วนประกอบต่าง ๆ ของ Transformer Stage	66
รูปที่ 4. 48 รูป Sort Stage	67
รูปที่ 4. 49 Stage Properties ของ Sort Stage	68
รูปที่ 4. 50 Stage Advanced ของ Sort Stage	68
รูปที่ 4. 51 รูป Remove Duplicates Stage	
รูปที่ 4. 52 Stage Properties ของ Remove Duplicates Stage	69
รูปที่ 4. 53 รูป Aggregator Stage	70
รูปที่ 4. 54 Stage Properties ของ Aggregator Stage	70
รูปที่ 4. 55 Output Mapping ของ Aggregator Stage	71
รูปที่ 4. 56 รูป Checksum Stage	72
รูปที่ 4. 57 Stage Properties ของ Checksum Stage	72
รูปที่ 4. 58 Output Mapping ของ Checksum Stage	73
รูปที่ 4. 59 ตัวอย่างผลลัพธ์ของ Ch <mark>ecks</mark> um Stage	73
รูปที่ 4. 60 รูป Row Generator Stage	74
รูปที่ 4. 61 Output Properties ของ Row Generator Stage	74
รูปที่ 4. 62 ตัวอย่างการใช้ Row Generator Stage ในการ Generate ข้อมูลในคอลัมน์ Name เ	ເດະ Salary75
รูปที่ 4. 63 รูป Peek Stage	75
INTUTE OF	

รูปที่ 4. 64 Stage Properties ของ Peek Stage76	
รูปที่ 4. 65 ตัวอย่าง output แบบ Job Log76	
รูปที่ 4. 66 หน้าต่างแสดงผลเมื่อเปิด Peek Stage Job Log76	
รูปที่ 4. 67 ตัวอย่างการวาง Peek Stage แบบ Output77	
รูปที่ 4. 68 ตัวอย่างข้อมูลจาก Database ที่บันทึกผลลัพธ์จาก Peek Stage แบบ Output77	
รูปที่ 4. 69 รูป Sample Stage	
รูปที่ 4. 70 ตัวอย่างการใช้งาน Sample Stage	
รูปที่ 4. 71 Stage Properties ของ Sample Stage79	
รูปที่ 4. 72 รูป Filter Stage	
รูปที่ 4. 73 Stage Properties ของ Filter Stage	
รูปที่ 4. 74 ตัวอย่างการใช้ Filter Stage แบบ Multiple Link	
รูปที่ 4. 75 รูป Funnel Stage	
รูปที่ 4. 76 ตัวอย่างการใช้ Funnel Stage81	
รูปที่ 4. 77 Stage Properties ของ Funnel Stage	
รูปที่ 4. 78 Stage Link Ordering ของ Funnel Stage	
รูปที่ 4. 79 ผลลัพธ์จากการใช้ Continuous Funnel83	
รูปที่ 4. 80 ผลลัพธ์จากการใ <mark>ช้ Sequence Funne</mark> l	
รูปที่ 4. 81 ผลลัพธ์จากการใช้ Sort <mark>Funn</mark> el	
รูปที่ 4. 82 รูป Lookup Stage85	
รูปที่ 4. 83 หน้าต่างการใช้งาน Loo <mark>kup</mark> Stage	
รูปที่ 4. 84 รูป Join Stage	
รูปที่ 4. 85 Stage Properties ของ Join Stage	
รูปที่ 4. 86 Stage Link Ordering ของ Join Stage87	
รูปที่ 4. 87 ตัวอย่างข้อมูลของ Left Table (DSLink23)	

รูปที่ 4. 88 ตัวอย่างข้อมูลของ Right Table (DSLink22)	
รูปที่ 4. 89 ผลลัพธ์จากการใช้ Full Outer Join	
รูปที่ 4. 90 ผลลัพธ์จากการใช้ Inner Join	88
รูปที่ 4. 91 ผลลัพธ์จากการใช้ Left Outer Join	
รูปที่ 4. 92 รูป Merge Stage	
รูปที่ 4. 93 Stage Properties ของ Merge Stage	90
รูปที่ 4. 94 ตัวอย่างข้อมูลจากปี 2014	90
รูปที่ 4. 95 ตัวอย่างข้อมูลจากปี 2015	91
รูปที่ 4. 96 Parallel Job ที่ใช้ในการทดสอบการทำงานของ Merge Stage	91
รูปที่ 4. 97 รูป Difference Stage	92
รูปที่ 4. 98 Stage Properties ของ Difference Stage	92
รูปที่ 4. 99 ตัวอย่างผลลัพธ์จากการใช้ Difference Stage	93
รูปที่ 4. 100 รูป Change Capture Stage	93
รูปที่ 4. 101 Stage Properties ของ Change Capture Stage	94
รูปที่ 4. 102 Stage Link Ordering ของ Change Capture Stage	94
รูปที่ 4. 103 ตัวอย่างการใช้งาน Change Capture Stage	95
รูปที่ 4. 104 รูป Change Ap <mark>ply Stage</mark>	95
รูปที่ 4. 105 Stage Properties (ซ้าย) และ Stage Link Ordering (บวา) บอง Change Apply Stage	96
รูปที่ 4. 106 ตัวอย่างผลลัพธ์ที่ได้จ <mark>ากกา</mark> รใช้ Chang <mark>e</mark> Apply Stage	96
รูปที่ 4. 107 ตัวอย่างการใช้งาน Ch <mark>ange</mark> Apply Sta <mark>ge</mark>	97
รูปที่ 4. 108 รูป Pivot Enterprise Stage	97
รูปที่ 4. 109 Stage Pivot Properties ของ Pivot Enterprise Stage	98
รูปที่ 4. 110 ข้อมูลก่อนแปรรูป	99
รูปที่ 4. 111 ผลลัพธ์หลังแปรรูป	99

## สารบัญตาราง

ตารางที่ 1.1 ตารางแสดงหน้าที่งานที่ได้รับมอบห	มาย	7
ตารางที่ 1. 2 ตารางอธิบายนิยามศัพท์เฉพาะ		9
ตารางที่ 3. 1 Gantt Chart ของแผนงานการปฏิบัติงาน	1	

# n í u í a æ n s væ e

# บทที่ 1

## ບກนຳ

ในบทนี้จะกล่าวถึงการแนะนำสถานประกอบการในด้านต่าง ๆ ตำแหน่งและหน้าที่ที่ผู้จัดทำ โกรงงานได้รับ รวมไปถึงวัตถุประสงก์ของโกรงงาน และกำศัพท์ต่าง ๆ ที่ได้ใช้ในโกรงงานฉบับนี้

## 1.1 ชื่อและที่ตั้งของสถานประกอบการ [1]

ชื่อหน่วยงาน บริษัท เอ-โฮสต์ จำกัด (A-HOST Company Limited)

- ที่ตั้ง เลขที่ 979/53-55 ชั้น 21 อาการSM Tower ถนนพหลโยธิน แขวงสามเสนใน เขตพญาไท กรุงเทพมหานกร 10400 (ดูแผนที่ประกอบได้ในรูปที่ 1.1)
- **โทรศัพท์** 02-298-0625 32
- แฟกซ์ 02-298-0053

เว็บไซต์ http://www.a-host.co.th

Marketing@a-host.co.th

อีเมล



รูปที่ 1. 1 สถานที่ตั้ง บริษัท เอ-โฮสต์ จำกัด \*1

#### 1.2 ลักษณะธุรกิจของสถานประกอบการ หรือการให้บริการหลักขององค์กร

บริษัท เอ-โฮสต์ จำกัด ได้ก่อตั้งขึ้นในปี พ.ศ. 2542 เป็นหนึ่งในผู้นำด้านการจดจำหน่ายผลิตภัณฑ์ ซอฟต์แวร์และบริการระบบเทคโนโลยีสารสนเทศในองค์กรแบบครบวงจร โดยเฉพาะ การให้บริการแอพ พลิเคชันแบบโฮสติง (Hosting) หรือ ASP (Application Service Provider) บริการดูแลระบบ ณ สถานที่ของ ลูกค้า (IT Outsourcing) รวมถึงบริการให้คำปรึกษาและจัดวางระบบด้วยแอพพลิเคชันสำหรับการวางแผน บริหารจัดการทรัพยากรองก์กร (Enterprise Resource Planning : ERP) ของ Oracle ซึ่งครอบคลุมทั้งองก์กร ขนาดใหญ่ องค์กรขนาดกลางและขนาดเล็ก ทั้งนี้ทางบริษัทยังได้รับการแต่งตั้งจาก Oracle ให้ดำเนินกิจการ โฮสติงในส่วนของระบบ Oracle Application เป็นรายแรกในประเทศไทยมาตั้งแต่ พ.ศ. 2542

ในการดำเนินธุรกิจของเอ-โฮสต์ตลอดระยะเวลา 10 ปี ไม่เพียงแต่ในฐานะผู้บุกเบิกธุรกิจโฮสติงและ ธุรกิจการให้ บริการแอพพลิเคชันในรูปแบบ ASP เท่านั้น แต่เอ-โฮสต์ยังได้ทำการติดตั้งระบบไอที รวมทั้ง ผลิตภัณฑ์ของ Oracle ให้กับลูกค้าจนประสบความสำเร็จมาแล้วเป็นจำนวนมาก ซึ่งหลายรายเป็นหนึ่งในร้อย บริษัทชั้นนำของประเทศไทย แต่ที่สำคัญกว่านั้น ก็คือการที่ เอ-โฮสต์ ได้กลายเป็นพันธมิตรทางธุรกิจอย่าง แนบแน่นจนกลายเป็นหุ้นส่วนทางกลยุทธ์และเป็นผู้สนับสนุนสำคัญที่มีส่วนช่วยผลักดันให้ธุรกิจของลูกค้า เติบโตสู่ความสำเร็จ

ปัจจุบันเอ-โฮสต์เป็นหนึ่งในบริษัทลูกของบริษัท คราก้อนวัน จำกัด ซึ่งเป็นบริษัทมหาชน ที่อยู่ในตลาด หลักทรัพย์แห่งประเทศไทย

เป็นเวลากว่า 10 ปีที่เอ-โฮสต์ และ Oracle ได้ดำเนินธุรกิจร่วมกันอย่างใกล้ชิด และ ถือเป็นพันธมิตรทาง ธุรกิจกันมานับตั้งแต่ก่อตั้งบริษัทปีพ.ศ.2542 จวบจนกระทั่งในปัจจุบันด้วยความมุ่งมั่นในการนำเสนอ ผลิตภัณฑ์ของ Oracle ผ่านการให้ บริการแอพพลิเคชันในรูปแบบของ ASP ในฐานะที่เอ-โฮสต์เป็นผู้บุกเบิก ธุรกิจดังกล่าว และเพิ่มศักยภาพในการดำเนินธุรกิจของลูกด้าได้อย่างเต็มประสิทธิภาพและเหมาะสม ทำให้ ได้รับรางวัลแห่งความสำเร็จและได้รับการยกย่องมาอย่างต่อเนื่อง

ทางเอ-โฮสต์ มีทักษะการทำงานและการบริการที่ครอบคลุมเทคโนโลยีชั้นสูงของ Oracle และได้รับ การแต่งตั้งให้เป็น OCAP (Oracle Certified Advantage Partner) รายแรกในไทย ซึ่งเป็นระดับสูงสุดของ Oracle ที่มีให้กับบริษัทคู่ก้าในปี พ.ศ. 2547 และต่อมาในปี พ.ศ. 2549 ได้รับรางวัล "ASEAN partner of the year" ดังที่ได้แสดงในรูปที่ 1.2



รูปที่ 1. 2 A-HOST Proud Awards \*1

ปัจจุบัน A-HOST มีประเภทของสินค้า และการบริการ ซึ่งสามารถแบ่งออกเป็น 3 กลุ่มใหญ่ ๆ ได้แก่

#### 1.2.1 Hosting & Outsource Services

เอ-โฮสต์ ได้ปรับปรุงและขยายการให้บริการ Hosting และ Outsource จนสามารถครอบคลุมความ ต้องการของลูกค้าได้หลากหลายโดยเรายึดหลักในการให้บริการที่เรียกว่า "Peace of Mind for the customer" ซึ่งหมายถึงการที่จะทำงานลูกค้าได้อย่างครบวงจร เพื่อที่ลูกค้าจะได้สามารถใช้งานระบบเทคโนโลยี สารสนเทศที่มีประสิทธิภาพได้อย่างสบายใจไร้ความกังวลต่อความเสี่ยงต่าง ๆ ไม่ว่าจะเป็นเรื่องของปัญหา ทางด้านเทคนิก การจัดการระบบและข้อมูลสำรองการปรับแต่งระบบให้ได้ประสิทธิภาพสูงสุด (Performance Tuning) และที่สำคัญที่สุดคือ การที่เข้ามารับภาระในด้านการบริหารจัดการบุคลากรทางด้าน ไอทีทั้งหมดแทนลูกค้า การใช้บริการ Hosting และ Outsource จะทำ ให้ลูกค้าสามารถทุ่มเทเวลา และทรัพยากรขององค์กร ให้กับธุรกิจที่เป็นแกนหลัก (Core Business) โดยทั่วไปแล้วบริการ Hosting และ Outsource ประกอบด้วย ส่วนประกอบและบริการย่อย ๆ ดังต่อไปนี้

- High Availability and High Performance IT Infrastructure

- Dedicated or Co-Location Service
- Disaster Site
- Oracle E-Business Applications (ERP, CRM, SCM)
- ERP Implementation Service
- System and Database Administration
- Help Desk

- On-Request Services i.e. On-site Support, Software Customization

ทั้งนี้การบริการ Hosting และ Outsource สามารถครอบคลุมได้ทั้งระบบที่ใช้เทคโนโลยีของ Oracle และ/หรือ ระบบที่ใช้เทคโนโลยีอื่น ๆ

#### 1.2.2 Oracle Core Technology Product and Advanced Services

เอ-โฮสต์เป็นผู้นำในการคำเนินธุรกิจในฐานะ Value-Added Distributor ให้กับ Oracle โดยไม่ เพียงแต่ทำหน้าที่ในการจัดจำหน่ายสินก้าในกลุ่ม Core Technology ของ Oracle ทุกประเภท แต่ยังมีทีม ผู้เชี่ยวชาญที่จะให้การสนับสนุน และการบริการเสริมอย่างกรบวงจรแก่บริษัท กู่ก้าและลูกก้า ไม่ว่าจะเป็น การร่วมจัดกิจกรรมทางการตลาด การฝึกอบรม การติดตั้งระบบและการให้กำปรึกษาเพื่อแก้ไขปัญหาต่าง ๆ

สินค้าบริการที่อยู่ในกลุ่มของ Oracle Core Technology Products and Advanced Services ได้แก่

- Oracle Database and database options

- Oracle Business Intelligence Suite
- Business Partner Development
- System Installation, Intelligence, and Optimization
- Oracle Fusion Middleware (รวมถึง BEA)
- สินค้าอื่น ๆ ทุกประเภทของ Oracle
- Marketing and Lead Generation Activities
- SOA-Based Development and Implementation

1.2.3 Oracle Enterprise Performance Management (EPM) Maz Hyperion Business Intelligence Products and Services

ความต้องการสูงสุดประการหนึ่งของผู้บริหารในการนำเอาระบบไอทีมาใช้ในองค์กร ไม่ว่าจะเป็น ภาครัฐหรือเอกชน คือ การทำให้ผู้บริหารสามารถได้ข้อมูลที่แสดงให้เห็นถึงสถานะในการดำเนินธุรกิจได้ อย่างแม่นยำ รวดเร็ว และนำเอาข้อมูลมาวิเคราะห์และวางแผนทั้งในระดับปฏิบัติการ และในระดับกล ยุทธ์ เพื่อให้ธุรกิจสามารถได้เปรียบ ปรับตัวตามสภาพเศรษฐกิจได้ในทุกสถานการณ์ Business Intelligence (BI) และเป็นระบบที่สนองตอบต่อความต้องการในลักษณะดังกล่าวได้เป็นอย่างดี เอ-โฮสต์ มีทีมงานที่ ปรึกษาที่มีประสบการณ์ทั้งทางด้านธุรกิจและทางด้านเทคนิก รวมถึงความเข้าใจในระบบอย่างลึกซึ้ง จึงทำ ให้สามารถให้บริการที่ปรึกษา เพื่อออกแบบติดตั้ง เชื่อมโยง และปรับใช้ระบบให้กับลูกค้าได้อย่างเป็น ประโยชน์ และประสิทธิภาพสูงสุดอีกทั้ง ยังมีความยึดหยุ่น และให้การตอบสนองที่รวดเร็วอีกด้วย

5

1.3 รูปแบบการจัดองค์กรและการบริหารองค์กรแบ่งตามประเภทการให้บริการที่รองรับ ทางสถานประกอบการณ์นั้นได้ให้บริการในด้านต่าง ๆ มากมาย [2] ผู้จัดทำโครงงานจึงได้ยกตัวอย่าง

บางส่วน ดังที่ได้แสดงในรูปที่ 1.3 ซึ่งผู้จัดทำโครงงานได้รับเลือกให้เข้าร่วมกับแผนก Academy and BI ซึ่ง สนับสนุนการให้บริการ BI Implementation Service



รูปที่ 1.3 โครงสร้างขององค์กร แบ่งตามรูปแบบบริการที่รองรับ

## 1.4 ตำแหน่งและหน้าที่งานที่<mark>นักสึ</mark>กษาได้รับ<mark>ม</mark>อบหมาย

ตำแหน่ง Assistance Software Professional

หน้าที่งานที่ได้รับมอบหมายทั้งหมดได้แสดงในตารางที่ 1.1

1	
a	ഴിച്ചില്
ตารางที่ 1. 1	ตารางแสดงหน้าทิ่งานที่ได้รับมอบหมาย

หน้าที่งานที่ได้รับมอบหมาย	ผถที่คาดว่าจะได้รับ
<ol> <li>สึกษากันกว้าคุณสมบัติของ</li> </ol>	- มีเอกสาร presentation
- BI Process	- ความรู้และเข้าใจ สามารถ present ได้
- ETL	
- Cube	
<ol> <li>สึกษาวิธีการติดตั้ง และการใช้งาน</li> </ol>	<ul> <li>มีเอกสารวิธีการติดตั้ง และการใช้งานโดยละเอียด</li> </ul>
- Oracle Database 11g	<ul> <li>สามารถสาธิตการติดตั้งและการใช้งานได้</li> </ul>
- Oracle Warehouse Builder 11g	- มีการทำ Demo ข้อมูลที่สามารถใช้งานได้
- IBM® Cognos Analytics	
- Oracle Business Intelligence Enterprise	
Edition 12c	
- IBM® InfoSphere® DataStage®	
3. ทำหน้าที่ผู้ทคสอบระบบ ( Tester ) ให้กับ	- สามารถช่วยกำหนด Test case ได้
ระบบต่าง ๆ ที่เกิดขึ้นของแผนก	<ul> <li>สามารถทดสอบระบบได้อย่างมีประสิทธิภาพ</li> </ul>
4. ช่วยSupport Project TAT BI 2015	<ul> <li>มีความรู้ความเข้าใจเกี่ยวกับข้อมูล การท่องเที่ยวเชิง</li> </ul>
	การตลาด
	- แก้ใขปรับปรุงรายงานและข้อมูล การท่องเที่ยวเชิง
	การตลาด ตามที่ได้รับมอบหมาย
	- ทำการเปิดเช็ครายงานของลูกค้า ทุกรายงานและETL
	log เพื่อเช็ <mark>กสถานะการใช้งาน</mark> เป็นประจำทุกเช้า
	<ul> <li>จัดทำเอกสาร Reconcile Report เพื่อเช็คข้อมูลของ</li> </ul>
T	ร <mark>ายงาน</mark> เป็น <mark>ป</mark> ระจำทุ <mark>กเคือ</mark> น
5. ช่วยพี่เลี้ยงใน class Training IB <mark>M®</mark> Cognos	- ช่วยเป็ <mark>น TA</mark> ในการ <mark>สอน I</mark> BM® Cognos Analytic ได้
Analytic	
6. มีการ Transfer งานทั้งหมด	- สามารถ Transfer งานทั้งหมดให้สหกิจรุ่นต่อไป
	- มีการทำ Archive เก็บเอกสารของงานไว้ที่ Server
7. จัดทำ Individual Work List	- นำเสนอ IWL เป็นประจำทุกเดือน

## 1.5 พนักงานที่ปรึกษา และ ตำแหน่งของพนักงานที่ปรึกษา

**ชื่อ** นางสาวปาริชาติ ทองคอนน้อย และ นายร<mark>ะพีพัฒน์ ธนะพั</mark>ฒน์

ตำแหน่ง Programmer

### 1.6 ระยะเวลาที่ปฏิบัติงาน

ปฏิบัติงานสหกิจศึกษาเป็นระยะเวลา 4 เดือนนับตั้งแต่วันที่ 30 พ.ค. 2559 – 30 ก.ย. 2559

# 1.7 ที่มาและความสำคัญของปัญหา

ในการทำรายงานต่าง ๆ ของทางแผนก Academy and BI นั้น จำเป็นต้องนำข้อมูลที่มีมาแปรรูปให้ เหมาะสมกับการนำเสนอเพื่อที่จะได้ตอบสนองต่อความต้องการของผู้ใช้งาน ซึ่งหนึ่งในเครื่องมือที่มีความ โคคเค่นในการแปรรูปข้อมูลนั้นก็คือโปรแกรม DataStage® ซึ่งเป็นหนึ่งในเครื่องมือของกลุ่ม InfoSphere® ที่สร้างขึ้นโคย IBM® ซึ่งโปรแกรมนี้มีจุดเค่นในเรื่องของการประมวลผลแบบขนาน ซึ่งส่งผลทำให้ กระบวนการแปรรูปข้อมูลเป็นไปได้อย่างรวดเร็วมากขึ้น ทางแผนก Academy and BI ได้เล็งเห็น ประสิทธิภาพของโปรแกรม จึงได้ทำการศึกษาการใช้งานและจัดทำกู่มือ เพื่อช่วยให้การทำความเข้าใจ โปรแกรม เป็นไปได้อย่างรวดเร็วมากยิ่งขึ้น

#### 1.8 วัตถุประสงค์หรือจุดมุ่งหมายของโครงงาน

- 1. เพื่อเปิดโอกาสให<mark>้ได้ลงมือปฏิบัติโก</mark>รง<mark>งานที่ม</mark>ือยู่จริงในบริษัท <mark>ตั้งแต่</mark>เริ่มต้นจนสิ้นสุด
- เพื่อจัดทำคู่มือการใช้งานโปรแกรม DataStage® ในระดับพื้นฐาน สำหรับลดระยะเวลาในการ เรียนรู้
- 3. เพื่อถ่ายทอดวิธีการใช้งา<mark>น St</mark>age ต่าง ๆ <mark>ข</mark>อง Data<mark>Stage®</mark> ให้แก่<mark>พนัก</mark>งานท่านอื่น ๆ
- ่ 4. เพื่อให้สามารถนำเสนอ<mark>วิธีกา</mark>รใช้งานแล<mark>ะ</mark>ผลลัพธ์ของ <mark>St</mark>age ต่<mark>าง ๆ ข</mark>อง DataStage® ได้

## 1.9 ผลที่คาดว่าจะได้รับจากการปฏิบัติงานหรือโครงงานที่ได้รับมอบหมาย

- 1. สามารถอธิบายความแตกต่างและการทำงานของกระบวนการ ETL และ ELT ได้
- สามารถเลือกวิธีการแปรรูปข้อมูลได้อย่างเหมาะสม
- 3. สามารถใช้งานและถ่ายทอดวิธีการใช้งานโปรแกรม IBM® InfoSphere® DataStage® ได้

## 1.10 นิยามศัพท์เฉพาะ

กำศัพท์ต่าง ๆ ที่ใช้ในรายงานฉบับนี้ล้วนแต่เป็นกำที่สามารถพบได้ทั่วไป แต่ในกรณีของโครงงานนี้ จะใช้ ในบริบทของการจัดการข้อมูลซึ่งแสดงในตารางที่ 1.2

C

## ตารางที่ 1. 2 ตารางอธิบายนิยามศัพท์เฉพาะ

คำศัพท์	คำอธิบาย		
bridge	ชิ้นส่วนที่ทำการแปลงนิยามข้อมูลจากรูปแบบหนึ่งไปเป็นอีกรูปแบบ โดยการจับคู่		
	องค์ประกอบต่าง ๆ ของนิยามข้อมูลเข้ากับค้นแบบมาตรฐาน		
connector	ชิ้นส่วนที่จัดเครียมการเชื่อมต่อข้อมูล และการรวมนิยามข้อมูล สำหรับฐานข้อมูล		
$\sim$	ภายนอก ซึ่ง โคยทั่วไปแล้ว ชิ้นส่วนแต่ละชิ้นจะถูกสร้างเพื่อรองรับการเชื่อมต่อกับ		
~	ฐานข้อมูลเฉพาะ		
data source	แหล่งของข้อมูล ตัวอย่างเช่นฐานข้อมูล หรือไฟล์ XML รวมไปถึงสารสนเทศต่าง ๆ ที่ใช้		
	เพื่อเข้าถึงข้อมูล		
job	วัตถุที่ออกแบบมาเพื่อประมวลผลองค์ประกอบต่าง ๆ ของโปรแกรม ไม่ว่าจะเป็น การ		
	เชื่อมต่อกับแหล่งข้อมูล การสกัดข้อมูล การแปรรูปข้อมูล และการบรรจุข้อมูลเข้าที่		
	แหล่งข้อมูลปลายทาง		
link	คือสัญ <mark>ลักษณ์</mark> ที่แทนทิศท <mark>างการไหลของข้อ</mark> มูล ระห <mark>ว่าง s</mark> tage ต่าง ๆ ภายใน job		
reference link	การเชื่อ <mark>มต่อ</mark> ที่เข้าสู่ Transformer หร <mark>ือ Look</mark> up stage ที่กำหนดตำแหน่งของตารางที่ใช้		
	สำหรับ <mark>การค้</mark> นหาข้อมูลเพื่ <mark>อ</mark> นำมาขยาย <mark>ความ</mark>		
stage	องค์ปร <mark>ะกอบ</mark> ของ job ที่ท <mark>ำห</mark> น้าที่อธิบายแ <mark>ห</mark> ล่งข้อม <mark>ูล ลำดั</mark> บการประมวลข้อมูล หรือระบบ		
	ปลายทางที่ได้นิยามตรรกะของการประมวลผล ที่ทำการเคลื่อนย้ายข้อมูลจาก link นำเข้า		
Če.	ไปยัง link ส่งออก		

## บทที่ 2 ทฤษฎีและเทคโนโลยีที่ใช้ในการปฏิบัติงาน

ในการปฏิบัติงานสหกิจศึกษาครั้งนี้ เป็นการนำความรู้ทางค้านทฤษฎีและเทคโนโลยีมาใช้ใน การปฏิบัติงานทุกส่วนตลอคการปฏิบัติงานสหกิจศึกษา ซึ่งเป็นการนำความรู้ทั้งที่เคยเรียนมาประยุกต์ใช้ และเป็นการศึกษาเรียนรู้สิ่งใหม่ ๆ ที่ได้จากการปฏิบัติงาน ซึ่งในหัวข้อนี้จะแบ่งออกเป็นสองส่วน ได้แก่ ส่วนที่เกี่ยวข้องกับแนวกิคในการจัคการข้อมูลในรูปแบบต่าง ๆ และแนวกิคของโปรแกรม และในส่วนที่ สองจะกล่าวถึงความสามารถต่าง ๆ ของโปรแกรม

#### 2.1 ทฤษฎีที่เกี่ยวข้องกับการแปรรูปข้อมูล 2.1.1 ETL และ ELT [3] Α В A B С Transformation C D Source Target D Engine EF E F ETL Pipeline

## **รูปที่ 2. 1** แผนภาพลักษณะการทำงานแบบ ETL \*3

รูปที่ 2.1 คือกระบวนการ ETL ซึ่ง<mark>ก็คือ</mark>การสกัด (Extract) แปรรูป (Transform) และจัดเก็บ (Load) ข้อมูล ซึ่ง ต่างจาก ELT ที่จะทำการสกัด (Extract) จัดเก็บ (Load) แล้วจึงแปรรูป (Transform) ข้อมูล ซึ่งแสดงในรูปที่ 2.2 ตรงที่ กระบวนการจัดเก็บแบบ ETL นั้นจะถูกแยกเครื่องออกเป็นสองหรือสามส่วน โดยแบ่งเป็นเครื่อง ต้นทาง (Source), เครื่องสำหรับแปรรูปข้อมูล (Transformation Engine), และเครื่องปลายทาง (Target) หรือ อาจจะรวมเครื่องต้นทางกับเครื่องที่ใช้แปรรูปข้อมูลเข้าด้วยกัน ซึ่งมีข้อดีคือใช้งบประมาณในการติดตั้งน้อย กว่าแบบ ELT แต่มีโอกาสเกิดปัญหาได้เนื่องจากการใช้วิธีการส่งข้อมูลแบบ Pipeline ซึ่งจะต้องนำข้อมูลมา เก็บไว้ก่อนที่ช่วง Transform ในกรณีที่ข้อมูลมีปริมาณมากเกินไปและส่งต่อได้ช้า ตัวอย่างเช่นการสั่ง จัดลำดับของข้อมูล ที่จะไม่สามารถส่งข้อมูลต่อได้ ถ้ายังอ่านข้อมูลไม่ครบ จึงทำให้เกิดการ Burst ได้



รูปที่ 2. 2 แผนภาพลักษณะการทำงานแบบ ELT \*3

เพื่อแก้ไขปัญหาการ Burst ของระบบ Pipeline จึงได้มีการพัฒนาวิธีการทำ ELT (Extract, Load, Transform) ขึ้นมา โดยวิธีนี้จะใช้งบประมาณในการติดตั้งระบบสูงกว่า และมักจะมีการวางโครงสร้างแบบแหล่งข้อมูล ต้นทางและปลายทางอยู่ในเครื่องเดียวกัน หรืออยู่คนละเครื่อง แต่เครื่องที่จัดเก็บข้อมูลปลายทางมีสมรรถนะ การประมวลผลที่สูง ทำให้สามารถแปรรูปข้อมูลได้อย่างสะควก ซึ่งวิธีนี้ยังมีข้อดีอีกอย่างคือความสะควกใน การสืบค้นหาแหล่งที่มาของข้อมูล

#### 2.1.2 Parallelism Basics

การประมวลผลของ IBM® InfoSphere® Information Server [4] นั้นใช้วิธีการ Pipeline Parallelism ร่วมกับ Partition Parallelism เพื่อทำให้สถาปัตยกรรมนั้นมีความยืดหยุ่น รองรับการขยายตัวของธุรกิจได้ โดยยัง รักษาประสิทธิภาพให้อยู่ในระดับสู<mark>งเส</mark>มอ ซึ่งเทคนิ<mark>ก</mark>ต่าง ๆ ที่ใช้ในการลำเลียงข้อมูลมีดังนี้

Data Pipelining คือ การเคลื่อนย้ายข้อมูลจากแหล่งข้อมูลไปประมวลผลตามลำคับขั้นที่ได้กำหนด
 ไว้ใน data-flow ซึ่งการทำแบบนี้ จะทำให้สามารถแปรรูปข้อมูลได้หลายครั้งโดยที่ไม่ต้องบันทึก
 ข้อมูลระหว่างขั้นตอน จุดเด่นของการใช้เทกนิกนี้ คือการใช้ความสามารถในการประมวลผลอย่าง

้เต็มประสิทธิภาพ เนื่องจากการถำเลียงที่ต่อเนื่องทำให้ใช้ส่วนประมวลผลต่าง ๆ ซึ่งแบ่งเป็น block ทำงานพร้อมกัน จึงทำให้ทำงานได้อย่างรวดเร็วดังรูปที่ 2.3



รูปที่ 2.3 แผนภาพลักษณะการทำงานแบบ Data Pipelining \*4

g

Data Partitioning คือ การนำข้อมูลที่จะทำการประมวลผล มาแบ่งเป็นชุดข้อมูลย่อย ๆ แล้วจึงแบ่งให้ หน่วยประมวลผลแต่ละหน่วย ทำงานแยกกัน โคยการแบ่งข้อมูลนั้นจะต้องไม่มีผลกระทบต่อความ ถูกต้องของข้อมูลที่ต้องการคำนวณ โดยความเร็วของการคำนวณจะเร็วขึ้นมากน้อยแก่ไหน ขึ้นกับ รูปแบบของข้อมูลที่ทำการแบ่งและวิธีการที่ใช้แบ่งก็ขึ้นอยู่กับรูปแบบข้อมูลที่เหมาะสมต่อ Stage ้นั้น ๆ ซึ่งโดยทั่วไปแล้ว ขั้นตอนนี้จะต้องสร้างด้วยการให้ผู้พัฒนา ออกแบบและสร้างวิธีการในการ แบ่งชุดข้อมูลเอง แต่ใน InfoSphere® นั้นได้สร้างมาให้แล้ว 10

ตัวอย่างเช่น รูปที่ 2.4 คือตัวอย่างการแบ่งงานโดยใช้ตัวอักษรเริ่มต้นของนามสกุล



รูปที่ 2. 4 แผนภาพลักษณะการทำงานแบบ Data Partitioning \*4

Dynamic Repartitioning คือ วิธีการในการจัดการข้อมูล ที่เกิดจากการผสมผสานระหว่าง Data
 Pipelining และ Data Partitioning ซึ่งจะได้รูปแบบตามรูปที่ 2.5



รูปที่ 2.5 แผนภาพลักษณะการทำงานแบบ Data Repartitioning \*4

ซึ่งวิธีการนี้ จะมีปัญหาในกรณีที่ใช้คอลัมน์ของข้อมูลในการจัดเก็บและการแบ่งชิ้นส่วนต่างกัน จึงได้พัฒนา ให้มีรูปแบบดังรูปที่ 2.6



รูปที่ 2. 6 แผนภาพลักษณะการทำงานแบบ Data Repartitioning \*4

ซึ่งจะมีการทำ Repartition หรือ<mark>ก็คือ</mark>การจับคู่ใหม่ระหว่างขั้นตอนทุกขั้นตอน โดยเทคนิคนี้ก็รวมมากับ InfoSphere® Information Server เ<mark>ช่นกั</mark>น ทำให้ผู้พั<mark>ฒ</mark>นาไม่จำ<mark>เป็นที่</mark>ต้องพั<mark>ฒนา</mark>ขั้นตอนนี้ด้วยตนเอง

(

#### 2.1.3 Information Server Backbone



รูปที่ 2.7 แสดงให้เห็นโครงสร้างของ Information Server ที่มี Application ต่าง ๆ [5] อยู่ในแถวบนสุด ซึ่ง Application ทั้งหมดนี้จะใช้ Service ในแถวกลางร่วมกัน โดยที่ทุก Application ใช้ Repository ที่แสดงอยู่ตรง มุมขวาล่างร่วมกัน โดยการจัดการ Application ต่าง ๆ สามารถทำได้ด้วยการใช้ Information Server Web Console หรือใช้ Client ของ Application นั้น ๆ ซึ่งแต่ละ Application จะมีลักษณะเด่นดังนี้

- Information Services Director ทำหน้าที่ในการรวบรวมข้อมูลและนำเสนอสารสนเทศในรูปแบบของ realtime service พร้อมทั้งสามารถนำหลักตรรกะที่เขียนไว้ กลับมาใช้ซ้ำได้

- Business Glossary ช่วยในการกำหนดกำศัพท์ทา<mark>งธุรกิจ</mark>ขององก์กร แล<mark>ะช่วย</mark>ในการสร้างและจัดการวิธีการ แบ่งหมวดหมู่ต่าง ๆ เพื่อช่วยในกา<mark>รจัดร</mark>ะเบียบให้ข้อมูล

- Information Analyzer ช่วยในการ<mark>ตรว</mark>จสอบและประเมินคุ<mark>ณภาพข</mark>องข้อ<mark>มูล อ</mark>อกแบบกฎต่าง ๆ ที่ใช้ควบคุม ข้อมูล และวิเคราะห์ข้อมูล เพื่อช่ว<mark>ยในก</mark>ารตีความข้<mark>อ</mark>มูล

- DataStage® ช่วยในการรวบรวมข้อมูลที่มาจากแหล่งข้อมูลหลายๆแหล่ง ซึ่งอาจจะติดตั้งอยู่บนระบบที่ แตกต่างกัน ซึ่งรวมไปถึง big data โดยใช้วิธีการทำงานแบบขนาน เพื่อประสิทธิภาพสูงสุด - QualityStage ช่วยในการสร้างและรักษาความสอดคล้องของข้อมูล เพื่อช่วยให้ข้อมูลอยู่ในรูปแบบที่ ถูกต้องเสมอ

 Metadata Workbench สามารถเข้าใช้ผ่านทาง browser โดยช่วยแสดงผลของ metadata ที่สร้างโดย Information Server ในรูปแบบของรูปภาพ เพื่อให้ผู้ใช้ สามารถใช้งานได้ง่ายขึ้น

\*ถึงแม้ว่า DataStage® และ QualityStage จะใช้ลิขสิทธิ์คนละชุดและนับเป็นสินค้าคนละชิ้น แต่ปกติแล้ว QualityStage จะถูกฝังอยู่ใน DataStage® ในรูปแบบของ Stage ต่าง ๆ

นโลยี

## 2.2 เทคโนโลยีที่ใช้ในการปฏิบัติงาน

#### 2.2.1 IBM® InfoSphere® DataStage®

#### 2.2.1.1 DataStage® Introduction

IBM® InfoSphere® DataStage® เป็นเครื่องมือที่ใช้รวบรวมข้อมูลของ IBM® InfoSphere® Information Server โดยการออกแบบ พัฒนา และ ใช้งาน job ต่าง ๆ ที่ทำหน้าที่โอนถ่ายและแปรรูปข้อมูล ซึ่ง data flow และ framework ของ job ต่าง ๆ จะมีการแสดงผลในรูปแบบของรูปภาพ graphic ทำให้เข้าใจได้ง่าย และเป็นมิตรต่อผู้ใช้งาน นอกจากนี้ DataStage® ยังรองรับการดึงข้อมูลจากแหล่งต่าง ๆ ที่มีระบบต่างกัน และสามารถส่งต่อข้อมูลที่ปรับแต่งอย่างเหมาะสมแล้วไปยัง data warehouse, data mart, operational data store, real-time web service และ messaging system หรือแอปพลิเกชันอื่น ๆ ขององค์กรได้โดยตรงเพื่อความ สมบูรณ์และความถูกต้องของข้อมูล นอกจากนี้ DataStage® ยังมี prebuilt functions ต่าง ๆ เพื่อช่วยลด ระยะเวลาพัฒนาระบบและช่วยตรวจสอบความสอดกล้องของข้อมูล พร้อมทั้งรองรับกระบวนการทำงานทั้ง แบบ ETL และ ELT นอกจากนี้ DataStage® ยังถูกพัฒนาขึ้นเพื่อรองรับการขยายตัวขององค์กรโดยใช้เทคนิค การประมวลผลแบบขนาน

#### 2.2.1.2 DataStage®'s Roles

ใน DataStage® จะมีการกำหนดบทบาท (Role) ต่าง ๆให้กับผู้ใช้ [6] เพื่อเป็นการกำหนด ควบคุม และจำกัด สิทธิต่าง ๆ ของผู้ใช้ บทบาทของผู้ใช้สามารถจำแนกออกเป็น 4 บทบาทดังนี้

- Administrator หรือ ผู้ดูแลระบบ เป็นบทบาทสูงสุด มีสิทธิในการกำหนดการตั้งค่าต่าง ๆ ได้ ทั้งหมด สามารถเพิ่ม ลบ และประกาศใช้งาน Project ได้ สามารถบริการจัดการสิทธิต่าง ๆ ของ ผู้ใช้ได้ โดยผู้ดูแลระบบจะทำงานผ่านทาง Command Interface ของ DataStage® Repository
- Manager หรือ ผู้จัดการระบบ มีสิทธิในการเข้าชมและแก้ไข Project ต่าง ๆ รวมถึงการจัดเก็บ และบริหารจัดการ Metadata ที่สามารถนำมาใช้ใหม่ได้ โดยที่ผู้จัดการระบบจะทำงานผ่านทาง Main Interface ของ DataStage® Repository
- Designer/Developer หรือ ผู้ออกแบบ/ผู้พัฒนาระบบ มีสิทธิในการสร้าง Job ต่าง ๆ โดย ผู้ออกแบบ/ผู้พัฒนาระบบจะมี User Interface แบบ Visual สำหรับการแสดงผล Data Flow
- 4. Director หรือ ผู้ดูแล มีสิทธิในการทดสอบ ติดตาม กำหนด และจัดการการทำงานของ Job ต่าง
  - ศ

#### 2.2.1.3 DataStage®'s Job

Job คือ การจัดกลุ่มของ Stage ซึ่งอาจจะมีหลาย Stage หรือมีเพียง Stage เดียวก็ได้ ในกรณีที่มีหลาย Stage ก็ จะมีการเชื่อมต่อด้วย Link โดยที่ Stage นั้นเป็นตัวแทนของแหล่งข้อมูล (Data Source) กระบวนการแปลง ข้อมูล (Transformation) และเป้าหมายที่จะทำการเก็บข้อมูล (Target) เมื่อนำ Job มาเรียงต่อกันเพื่อกำหนด ลำดับการทำงาน จะถูกเรียกว่า Job Flow

#### 2.2.1.4 Job Lifecycle

DataStage® ช่วยอำนวยความสะควกในการจัดการข้อมูล โดยใช้ Job ต่าง ๆ ซึ่งการพัฒนาระบบขึ้นมานั้นจะ มีทั้งหมด 4 ขั้นตอน ได้แก่ Develop, Test, Deploy, Operate

 Develop (พัฒนา) [7] คือขั้นตอนการจัดการ Job Flow ให้โอนถ่ายข้อมูลจากต้นทาง (Source) สู่ปลายทาง (Target) อย่างถูกต้องเหมาะสม โดยใช้ Designer client ในการจัดการ นอกจากนี้ ในกระบวนการนี้ สามารถตรวจสอบข้อผิดพลาด โดยการ compile Job Flow ที่เขียนขึ้นมา โดยในขั้นตอนการพัฒนานั้น สามารถแบ่งแยกย่อยได้อีก 3 ขั้นตอน ดังรูปที่ 2.8



- 1.1. Define the Job Flow หรือก็คือการออกแบบ Job Flow สามารถทำได้โดยจัดเรียง Stage ต่าง ๆ บน Designer Canvas และเชื่อมต่อ Stage ต่าง ๆ ด้วย Link ซึ่ง Link จะทำหน้าที่ส่งต่อ Metadata โดย stage เป็นตัวแทนของแหล่งข้อมูล (Data Source) กระบวนการแปลงข้อมูล (Transformation) และ เป้าหมายที่จะทำการเก็บข้อมูล (Target) ในขั้นตอนนี้ สามารถแบ่งแยกย่อยได้อีก 4 กระบวนการ ดังนี้
  - 1.1.1. Add source stages คือการกำหนดแหล่งข้อมูล (Data Source)
  - 1.1.2. Add transformation stages คือการกำหนดกระบวนการแปลงข้อมูล (Transformation)
  - 1.1.3. Connect the stages คือการกำหนด Link ซึ่ง Link จะทำหน้าที่ส่งต่อ Metadata
  - 1.1.4. Add target stages คือการกำหนดเป้าหมายที่จะทำการเก็บข้อมูล (Target)
- 1.2. Configure the Job หรือก็คือการตั้งค่า Job ซึ่งสามารถทำได้โดยการตั้งก่า Stage ต่าง ๆ เพื่อเจาะจง กระบวนการแปรรูปข้อมูล หรือเพิ่ม Job Parameter เพื่อให้รับค่าในเวลาที่เรียกใช้รายงานเพื่อเลี่ยง การกำหนดค่าล่วงหน้า หรือเพิ่ม Table Definition ที่เก็บข้อมูลเกี่ยวกับโครงสร้างข้อมูลซึ่งสามารถ Import จาก Database อื่น ๆ ได้ ไม่ว่าจะเป็นทาง Source หรือ Target
- 1.3. Review Compilation Errors เป็นการตรวจสอบ Stage ต่าง ๆ ว่ามีข้อผิดพลาดหรือไม่ ซึ่งสามารถตั้ง ค่าให้ Highlight Stage แรกที่มีปัญหาได้ นอกจากนี้ระบบยังช่วยระบุปัญหาเพื่อช่วยในการค้นหาจุด ที่ผิดพลาดและจุดที่ต้องทำการแก้ไข



รูปที่ 2. 9 <mark>ตัวอ</mark>ย่างการแส<mark>ค</mark>ง Stage ที่เกิด<mark>ก</mark>วามผิ<mark>ดพลา</mark>ดขึ้น \*7

ในรูปที่ 2.9 พบปัญหาคือ<mark>การตั้งก่</mark>าของ Tr<mark>an</mark>sformer Stage ชื่อ Trim\_Prep ไม่ครบถ้วน ดังนั้นระบบ จึงทำการ Highlight Stage ชื่อ Trim\_Prep 2. Test (ทดสอบ) [8] คือขั้นตอนทดสอบเพื่อ Debug Job ที่เขียนขึ้น โดยใช้เครื่องมือ Interactive Debugging ในการตรวจสอบ ขั้นตอนนี้ต่างจาก Review Compilation Errors ในขั้นตอนที่แล้วตรงที่ ขั้นตอนนี้จะ เน้นการตรวจสอบข้อมูลเป็นหลัก ซึ่งข้อมูลนั้นสามารถใช้ได้ทั้งแบบ Sample หรือจะ Generate ข้อมูล ขึ้นมาเพื่อทดสอบก็ได้ ในขณะที่ Review Compilation Errors ในขั้นตอนที่แล้วจะเป็นการค้นหาและ ตรวจสอบ Error ที่เกิดขึ้นระหว่าง Compile



<mark>รูปที่ 2. 10 แผนผังขั้นต</mark>อนการทดสอบ Job \*8

ขั้นตอนนี้สามารถแบ่งแยกย่อย<mark>ใด้ 2</mark> ส่วน ดังรูป<mark>ที่</mark> 2.10

- 2.1. Use debugging stages ก<mark>ือขั้น</mark>ตอนการกำหนดข้อมูลเพื่อที่จะทดสอบการทำงานของ Job ว่าทำงาน ถูกต้องตามที่ออกแบบเอาไว้หรือไม่ ซึ่งข้อมูลสำหรับทดสอบจะสามารถหามาได้ด้วย 2 วิธี
  - 2.1.1. Generate test data เป็นวิธีการ Generate ข้อมูลที่จะใช้ทดสอบขึ้นมา ซึ่งสามารถใช้ข้อมูลจาก ฐานข้อมูลอื่นมาประกอบ เพื่อช่วยเพิ่มรายละเอียดของข้อมูลให้มากขึ้น

2.1.2. Sample the data เป็นวิธีการทำสำเนาส่วนหนึ่งของข้อมูล เพื่อนำมาทคสอบ โคยใช้ Stage ที่มี ชื่อว่า Peek Stage ซึ่งมีการใช้งานเหมือนในรูปที่ 2.11



รูปที่ 2. 11 ตัวอย่างการใช้งาน Peek Stage \*8

- 2.2. Run in debug mode ในโหมดนี้จะสามารถสำรวจข้อมูลภายในคอลัมน์หรือค่าของตัวแปรต่าง ๆ ได้ โดยการใช้ Breakpoint ซึ่งก็คือการกำหนดตำแหน่งที่จะหยุดการทำงานชั่วคราว เพื่อตรวจสอบ ข้อมูลระหว่างการประมวลผลหรือแปรรูปข้อมูลได้ นอกจากนี้ Breakpoint ยังสามารถกำหนดแบบ เป็นเงื่อนไขได้ เช่นกำหนดให้หยุดการทำงานทุก ๆ 5 แถว
- Deploy (ติดตั้ง) [9] ขั้นตอนนี้ไม่ได้เกี่ยวข้องกับ DataStage® โดยตรง แต่เป็นการใช้เครื่องมือชื่อ Information Server Manager ในการค้นหาวัตถุต่าง ๆ ที่จำเป็นต่อการทำงานของ Job หรือกำหนดวัตถุอื่น ๆ ที่ต้องการที่จะติดตั้งลงบนระบบ โดยจะแบ่งแยกย่อยได้อีก 4 ขั้นตอนเหมือนรูปที่ 2.12 ดังนี้



- 3.1. Define the Package คือขั้นตอนนิยาม Package หรือก็คือการให้ความหมายหรือกำหนดลักษณะของ Package ว่าจะประกอบไปด้วยอะไรบ้าง ซึ่งการนิยามนี้จะทำเพียงหนึ่งครั้ง เมื่อมีความเปลี่ยนแปลง เกิดขึ้น Package สามารถที่จะ Rebuild, Recopy และ Redeploy ได้
- 3.2. Build the Package คือขั้นตอนสร้าง Package หลังจากที่ได้นิยามไว้แล้ว ก็จะทำการสำเนาข้อมูลมา จากต้นทาง โดยข้อมูลที่ทำสำเนามาเก็บจะเป็นไปตามข้อมูลที่เราได้กำหนดหรือได้นิยามเอาไว้ โดยขั้นตอนนี้สามารถที่จะเขียนทับข้อมูล Package เก่าทั้งหมด หรือจะอัพเดทข้อมูลใน Package โดยอ้างอิงจากการแก้ไขครั้งล่าสุด หลังจากที่ทำการสร้าง Package แล้ว Package จะมีสถานะที่ พร้อมสำหรับการนำไปติดตั้ง
- 3.3. Deploy the Package เป็นการนำ Package ที่ได้สร้างไว้ไปติดตั้งเข้ากับระบบ โดยจะแบ่งแยกย่อยได้ อีก 2 ขั้นตอน
  - 3.3.1.Deploy on the test system คือการนำ Packageไปติดตั้งบนระบบที่สร้างไว้เพื่อทดลองการ ทำงาน

3.3.2. Deploy on the production system คือการนำ Package ไปติดตั้งบนระบบจริง

4. Operate (คำเนินการ) [10] เป็นขั้นตอนที่ใช้เครื่องมือ DataStage® และ QualityStage ในการจัดการ ทรัพยากรของระบบและเรียกใช้ Job ที่ได้ทำการติดตั้งลงบนระบบแล้ว



รูปที่ 2. 13 แผนผังวงจรชีวิตของ Job \*10

รูปที่ 2.13 ได้แสดงแผนวงจรชีวิตของ Job ในขั้น Operate ซึ่งในขั้นตอนนี้จะมี Dashboard ช่วยแสดง workload ของ Job ที่ทำงานอยู่, Job ที่ทำงานเสร็จแล้ว, การใช้ทรัพยากรของระบบ นอกจากนี้ยังติดตาม จำนวน Job ที่ทำงานสำเร็จและ Job ที่ทำงานล้มเหลวรวมถึงสถานะของ Key Service ซึ่ง Job เหล่านี้ สามารถที่จะกำหนดลำดับความสำคัญและสามารถที่จะ Drill Down เพื่อดูรายละเอียดได้ และ มีระบบ แจ้งเตือนตามเกณฑ์ที่กำหนดผ่านทาง Dashboard
#### 2.2.1.6 Job Design

้ คือ ขั้นตอนการออกแบบ Job ให้สกัดข้อมูลมาจากหลายแหล่งข้อมูล แปรรูปให้เป็นข้อมูลที่ต้องการแล้วจึง จัดเก็บลงในระบบที่ต้องการ ซึ่ง Job นั้นประกอบด้วย Stage และ Link ดังนั้นการออกแบบ Job จึงจำเป็นที่ จะต้องรู้จัก Stage ต่าง ๆ ซึ่งมีตัวอย่างการใช้งานที่แบ่งหมวดหมู่ตามลำดับดังนี้

- 1. Extract and load data
- 2. Transform data
- 3. Enrich data
- 4. Cleanse data
- 5. Real-time processing
- 6. Big data processing
- 7. Combine jobs in a sequence job
- 1. Extract and Load Data [11]

ในหมวดนี้จะเกี่ยวข้องกับการย้ายข้อมูลต่าง ๆ ไม่ว่าจะเป็นการสกัดข้อมูลจากแหล่งข้อมูล (Source) หรือ การจัดเก็บข้อมูลเข้าคลังข้อมูล นอกจากนี้ยังรวมถึงการตรวจสอบความเปลี่ยนแปลงภายในฐานข้อมูล อีกด้วย

นโลยัไก

1.1. Extract Data from a Database to a File or Data Set

คือ การสกัดข้อมูลออกมาจากฐานข้อมูลแล้วนำไปเก็บในรูปแบบของ File หรือ Data Set

1.2. Load Data with High-speed Native Connectivity

คือ การนำข้อมูลที่อ่านจากแหล่งข้อมูลมาจัดเรียงและจัดเก็บลงฐานข้อมูลในรูปแบบ Native Format เพื่อเพิ่มความเร็วในการทำงาน เช่น การนำข้อมูลจากภายนอกบริษัทมาใส่ไว้ใน Netezza Table โดยใช้ Stage ที่มีชื่อว่า Netezza Connector โดยที่ในตัวอย่างได้ใช้ Native API เพื่อช่วยให้ ทำงานได้เต็มประสิทธิภา<mark>พดัง</mark>ที่ได้แสดงตัวอย่างในรูปที่ 2.14



รูปที่ 2. 14 ตัวอย่างการใช้ Netezza Connector \*11

1.3. Load Data in Real Time to a Data Warehouse

้ คือ การรับข้อมูลที่อยู่ระหว่างกระบวนการ Transaction มาแปรรูปแล้วจัดเก็บลงในคลังข้อมูล

1.4. Extract Data from SAP

คือ การสกัดข้อมูลออกมาจาก SAP ซึ่งสามารถใช้ Stage ที่มีชื่อว่า ABAP Extract ซึ่งข้อมูลที่ได้ สามารถนำไปแปรรูปด้วย Transformer Stage ก่อนที่จัดเก็บได้ ซึ่งรูปแบบที่จัดเก็บได้ก็มี หลากหลายรูปแบบ เช่น SAP Business Warehouse, ฐานข้อมูลขององค์กร, ไฟล์ ฯลฯ

1.5. Fetch Data from a Web Service to Enrich Data

คือ การนำข้อมูลจาก Web Service มาขยายความข้อมูลที่มีอยู่ ตัวอย่างเช่นในรูปที่ 2.15 ที่นำ Web Services Transformer Stage มาใช้ในการสกัดและแปรรูปข้อมูล Employee จาก Sequential\_File\_1 แล้วจึงส่งข้อมูลต่อให้ XML Stage ซึ่งจะทำการ Generate Row และ Column แล้วจึงนำไปเก็บใน Sequential\_File\_2



ร**ูปที่ 2. 15** ตัวอย่างการนำข้อมูลจาก Web Service มาช่วยทำให้ข้อมูลมีความสมบูรณ์มากขึ้น \*11

#### 1.6. Identify Database Changes

10

กือ การตรวจสอบความเปลี่ยนแปลงของฐานข้อมูล โดยใช้ InfoSphere® Change Data Capture (CDC) ซึ่งเป็นส่วนหนึ่งของ IBM® InfoSphere® Data Replication ซึ่งจะทำหน้าที่ในการจับความ เปลี่ยนแปลงที่เกิดขึ้นแล้วจึงส่งให้ DataStage®, ฐานข้อมูลปลายทาง หรือคิวข้อความ โดยหลักการ ทำงานคือ เก็บ Log ของการเปลี่ยนแปลงต่าง ๆ เอาไว้ ทำให้ลดภาระการทำงานของ ETL Server เนื่องจากไม่ต้องหาความเปลี่ยนแปลงที่เกิดขึ้น และแก้ไขเฉพาะข้อมูลที่มีการเปลี่ยนแปลงเท่านั้น 2. Transform Data [12]

ในหมวคนี้ จะเกี่ยวข้องกับตัวอย่างการแปรรูปข้อมูลในรูปแบบต่าง ๆ ดังนี้

2.1. Transform Data with Expressions

คือ การแปรรูปข้อมูลด้วย Expression โดยใช้ Transformer Stage เช่นการตัดช่องว่างหัวท้าย การนำ ข้อความมาต่อกัน การนำข้อมูลไปคำนวณ เป็นต้น

2.2. Join Data from Multiple Flat Files

คือ การอ่านข้อมูลจากหลายๆ ไฟล์แล้วนำมาเขียนเป็นไฟล์เดียว โดยการใช้ Join Stage ดังที่แสดงใน รูปที่ 2.16



2.3. Join Heterogeneous Sources

(0

กือ การอ่านข้อมูลจากหลายๆระบบ นำมารวมกัน และจัดให้มีรูปแบบเดียวกัน เพื่อสร้างฐานข้อมูล ที่ถูกต้อง แม่นยำ และเชื่อถือได้เพียงหนึ่งฐานข้อมูล ตัวอย่างเช่น ในตัวอย่างมีการนำข้อมูลมาจาก ฐานข้อมูล Oracle และ Teradata ซึ่งได้ทำการแปรรูปข้อมูลที่ได้จากฐานข้อมูล Oracle ก่อนหนึ่ง ครั้ง แล้วจึงนำมา Join กับข้อมูลจากฐานข้อมูล Teradata แล้วจึงนำข้อมูลทั้งหมดไปแปรรูปอีกครั้ง ก่อนจะทำการจัดเก็บโดยใช้ Netezza Connector Stage ดังรูปที่ 2.17



รูปที่ 2. 17 ตัวอย่างการใช้งาน Join Stage ในการ Join ข้อมูลจากฐานข้อมูล Oracle และ Teradata \*12

#### 2.4. Summarize Data by Common Characteristic

้คือ การสรุปข้อมูลโคยใช้ Aggregator Stage ซึ่งสามารถจัดกลุ่มข้อมูลที่ต้องการสรุปยอดได้ มากกว่า 1 กลุ่ม ตัวอย่างเช่น จัดกลุ่มด้วยวัน (วันจันทร์ ถึง วันอาทิตย์) และจัดกลุ่มด้วยเดือน เพื่อหา ผลสรุป ซึ่งอาจจะแสดงให้เห็นได้ว่า วันไหนของสัปดาห์เป็นวันที่ยุ่งที่สุด ซึ่งอาจจะแตกต่างกันไป ตามฤดูกาล ดังรูปที่ 2.18

C



รูปที่ 2. 18 ตัวอย่างการใช้งาน Aggregator Stage \*12

#### 2.5. Pivot Data in a Table

้ คือ การ Pivot Table โดยก<mark>ารใช้</mark> Pivot Enterprise Stage ซึ่งสามารถแบ่งออกมาเป็น Horizontal Pivot ้ซึ่งก็คือการนำข้อมูลประ<mark>เภทเ</mark>ดียวกันในห<mark>ล</mark>ายๆคอลัม<mark>น์ใน</mark>แถวเ<mark>ดียวกั</mark>น มากระจายออกเพื่อลด ้ปริมาณคอลัมน์ลง โค<mark>ยจำนวนแ</mark>ถวที่เพิ่มขึ้นจะสัมพันธ์กับจำนวนคอลัมน์ที่ลดลงไป กับ Vertical Pivot ซึ่งก็คือการนำข้อมูลที่เหมือนกันในหลายๆแถวมารวมกันเป็น 1 แถว และเพิ่มคอลัมน์ให้ สอคคล้องกับข้อมูลที่เปลี่ยนแปลง ดังรูปที่ 2.19



รูปที่ 2. 19 ตัวอย่างการใช้งาน Pivot Enterprise Stage \*12

2.6. Find Deltas from Yesterday's File

คือ การหาความเปลี่ยนแปลงที่เกิดขึ้นระหว่างไฟล์ เช่นไฟล์ของเมื่อวานกับไฟล์วันนี้ ซึ่งความ เปลี่ยนแปลงที่เกิดขึ้นนั้น จะสามารถเปรียบเทียบได้โดยการใช้ Change Capture Stage และจะได้ ผลลัพธ์เป็นข้อมูลของความเปลี่ยนแปลง ซึ่งในตัวอย่างคือการเปรียบเทียบไฟล์ของเมื่อวานและ วันนี้ โดยใช้ Change Capture Stage ในการหาความเปลี่ยนแปลง แล้วจึงส่งข้อมูลความเปลี่ยนแปลง ไปเก็บไว้ในฐานข้อมูล Teradata โดยใช้ Teradata Connector Stage ดังรูปที่ 2.20





<mark>รูปที่ 2. 20</mark> ด้วอย่างการใช้งาน Change Captur<mark>e Sta</mark>ge \*12

3. Enrich Data [13]

ในหมวดนี้จะเกี่ยวข้องกับการ<mark>เพิ่มร</mark>ายละเอียดข<mark>อ</mark>งข้อมูล เพื่<mark>อ</mark>ทำให้ข้อมูลมีความสมบูรณ์มากขึ้น โดย การนำข้อมูลจากแหล่งอื่นมาใช้ขยายความ ซึ่งขั้นตอนนี้ยังรวมถึงการกำหนดกฎทางธุรกิจต่าง ๆ เพื่อ ควบคุมการทำงานต่าง ๆ 3.1. Look Up Reference Data for Data Enrichment

คือ การเพิ่มรายละเอียดข้อมูลโดยนำมาจากแหล่งที่สามารถอ้างอิงถึงได้ โดยใช้ Lookup Stage ตัวอย่างเช่นในตัวอย่างที่นำข้อมูลลูกค้ามาจากฐานข้อมูล DB2 แล้ว Lookup เพื่อนำข้อมูลรหัส ประเทศจากฐานข้อมูล Oracle มาขยายความ แล้วจึงจัดเก็บในรูปแบบ Sequential File ดังรูปที่ 2.21



ร**ูปที่ 2. 21** ตัวอย่างการใช้งาน Lookup Stage \*13

3.2. Apply Business Rules

คือ การนำกฎที่เคยสร้างไว้มาใช้กับข้อมูลชุดอื่น ๆ เพื่อทำให้ข้อมูลเป็นไปในทางเดียวกัน โดยใช้ ILOG JRules Connector Stage ซึ่งสามารถเข้าถึงเซ็ตของกฎต่าง ๆ ที่อยู่ภายใน Operational Decision Manager ได้ ดังเช่นรูปที่ 2.22



ร**ูปที่ 2. 22 ตัวอ**ย่างการใช้ง<mark>า</mark>น ILOG JRules Connecto</mark>r Stage \*13

#### 4. Cleanse Data [14]

ในหมวดนี้จะเกี่ยวข้องกับการจัครูปแบบและตรวจสอบข้อมูล เพื่อให้ข้อมูลถูกเก็บในรูปแบบที่ต้องการ

4.1. Standardize Data

กือ การจัดรูปแบบของข้อมูล เพื่อทำให้ข้อมูลแต่ละ Record มีมาตรฐานเดียวกันเพื่อช่วยในการ ก้นหาข้อมูล โดยใช้ Standardize Stage ซึ่งมากับ IBM® InfoSphere® QualityStage ซึ่งสามารถ แก้ไขกฎต่าง ๆ ได้โดยการใช้ IBM® InfoSphere® QualityStage Standardization Rules Designer ซึ่งมีลักษณะเป็น web-base และเป็นเครื่องมือแบบ data-driven classification ดังรูปที่ 2.23



รูปที่ 2. 23 ตัวอย่างการใช้งาน Standardize Stage \*14

4.2. Verify International Addresses

คือ การตรวจสอบและจัดรูปแบบของข้อมูลเกี่ยวกับที่อยู่ ว่าเป็นสถานที่ ที่สามารถจัดส่งไปได้ หรือไม่ โดยใช้ Address Verification Stage ซึ่งมีมากับ IBM® InfoSphere® QualityStage Address Verification Interface ดังรูปที่ 2.24



<mark>รูปที่ 2. 24</mark> ตัวอย่างก<mark>าร</mark>ใช้งาน Addr<mark>e</mark>ss Ver<mark>ificat</mark>ion Stage \*14

4.3. Match Records to Identify Duplicates

คือ การหาข้อมูลที่ซ้ำกัน โดยใช้วิธีการจับคู่ข้อมูลที่มีการชี้ไปยังเป้าหมายเดียวกัน โดยในตัวอย่าง ได้มีการใช้ Match Frequency Stage ดังรูปที่ 2.25



รูปที่ 2. 25 ตัวอย่างการใช้งาน One Source Match Stage \*14

4.4. Validate Data with Data Rules

10

คือ การตรวจสอบคุณภาพของข้อมูลว่าเป็นไปตามที่ต้องการหรือไม่ โดยใช้ Data Rules Stage ซึ่ง สามารถกำหนดกฎเกณฑ์ได้ด้วย IBM® InfoSphere® Information Analyzer ดังรูปที่ 2.26



รูปที่ 2. 26 ตัวอย่างการใช้งาน Data Rules Stage \*14

5. Real-time Processing [15]

ในหมวดนี้จะเกี่ยวข้องกับการรวบรวมข้อมูลแบบ real-time ผ่านทาง web service

5.1. Respond to a Web Service Request

คือ การกำหนดการนำข้อมูลเข้า Job จาก web service ด้วย ISD Input Stage และส่งกลับไปยัง web service ด้วย ISD Output Stage โดยการใช้ InfoSphere® Information Services Director นอกจากนี้ ยังสามารถนำ Job นี้ไปใช้งานแบบ Service ได้ดังตัวอย่างในรูปที่ 2.27



รูปที่ 2. 27 ตัวอย่างการใช้งาน ISD Input Stage เพื่อดึงข้อมูลจาก Web Service \*15

#### 5.2. Verify Data between Message Queues

กือ การสร้าง Job ที่สามารถรับข้อมูลจาก Messaging Queue และจัดเก็บข้อมูลเข้า Messaging Queue ซึ่งตัวอย่างในรูปที่ 2.28 จะใช้ IBM® WebSphere MQ Connector Stage สำหรับการรับข้อมูล จากนั้นจึงนำมาจัดรูปแบบด้วย Standardize Stage, นำไปคัดแยกด้วย Data Rules Stage แล้วจึงนำไป เก็บข้อมูลเข้า Messaging Queue โดยใช้ IBM® WebSphere MQ Connector Stage ในกรณีที่ข้อมูล ได้รับการ Verified จาก Data Rules Stage และจัดเก็บในฐานข้อมูล DB2 ในกรณีที่โดน Reject



รูปที่ 2. 28 ตัวอย่างการใช้งาน WebSphere MQ Connector Stage \*15

6. Big Data Processing [16]

้ในหมวคนี้จะเกี่ยวข้องกับการแลกเปลี่ยนข้อมูล จากแหล่งข้อมูลขนาดใหญ่

- 6.1. Access Data on HDFS (Hadoop Distributed File System)
   คือ การใช้ Big Data File Stage ในการดึงและจัดเก็บข้อมูลจากไฟล์ HDFS
- 6.2. Augment Data with Hadoop-based Analytics

คือ การใช้ Hadoop-based analytic ในการประมวลผลก่อน จากนั้นจึงนำมาจัดเก็บในฐานข้อมูล ตัวอย่าง เช่นในในรูปที่ 2.29 ข้อมูลที่ผ่านการประมวลผลแล้ว จะสามารถเชื่อมต่อได้ โดยการใช้ Hive Stage แล้วจึงนำไปจัดเก็บลงฐานข้อมูล Netezza โดยใช้ Netezza Connector Stage



รูปที่ 2. 29 ตัวอย<mark>่างเชื่</mark>อมต่อระหว่<mark>า</mark>ง Hive Sta<mark>ge ก</mark>ับ Net<mark>ezza</mark> Connector Stage \*16

# VSTITUTE O

7. Combine Jobs in a Sequence Job [17]

การนำ Job มาต่อกันเพื่อกำหนดลำดับการทำงาน และวิธีการทำงาน เรียกว่า Sequence Job ซึ่งมี ความสามารถในการกำหนด Checkpoint Restart ทำให้สามารถ Restart เพื่อทำงานต่อจากขั้นที่ทำสำเร็จ ขั้นล่าสุดได้ทันที โดยในรูปที่ 2.30 Job ตัวอย่าง ได้มีการใช้ Job Activity Stage 3 Stage ในการ Extract, Transform, Load จากนั้นจึงใช้ Notification Activity Stage ในการส่งข้อความเตือนต่อไป ซึ่งในตัวอย่าง ยังมีการใช้ Exception Handler Stage ในการรองรับกรณีที่เกิด Error ต่าง ๆ โดยเมื่อมี Error เกิดขึ้น ก็จะ ทำงานตาม Routine Activity Stage



## บทที่ 3

## แผนงานการปฏิบัติงานและขั้นตอนการดำเนินงาน

ในบทนี้จะกล่าวถึงแผนการปฏิบัติงาน ซึ่งจะเจาะจงการทำโครงงานเป็นหลัก เนื่องจากงานที่ได้รับ ้มอบหมายต่าง ๆ นั้นเป็นการสนับสนุนการทำงานของพนักงานภายในแผนกเป็นหลัก และมีผลกับโครงงาน ที่ทำ ทั้งทางตรงและทางอ้อม ซึ่งผลลัพธ์ที่ได้ของโครงงานนี้ คือ User Manual ซึ่งมีความแตกต่างจากคู่มือ ของทาง IBM® เนื่องจาก โครงงานชิ้นนี้เน้นที่การตั้งก่าต่าง ๆ ของ Stage มากกว่ากู่มือของทาง IBM® ซึ่งมี นโลยี1ุก ลักษณะเป็นการทำตามขั้นตอนที่กำหนด

## 3.1 แผนงานการปฏิบัติงาน

## ตารางที่ 3. 1 Gantt Chart ของแผนงานการปฏิบัติงาน

งานที่ปฏิบัติ	ມີຖຸ	นา	ยน	กร	กฎา	คม	สิง	หา	คม		ก้	ันย	ายน	Î
1. ศึกษาและทำความเข้าใจ BI Concept										-		C	2	
2. ศึกษาโปรแกรมอื่นที่ใช้ในการทำ BI														
3. ศึกษา Concept ของ DataStage®														
4. ทำเอกสารแนะนำ DataStage®														
5. ศึกษาการใช้งาน Stage ต่าง ๆ ของ														
DataStage®														
6. นำเสนอวิธีการใช้งาน <mark>S</mark> tage <mark>ต่าง</mark> ๆ ของ													~	
DataStage®													(	)
7. จัดทำคู่มือการใช้งาน DataSt <mark>age®</mark>												C	)	
8. ตรวจสอบคู่มือ DataStage®											-			
9. แก้ไขปรับปรุงคู่มือ DataStage®								1		$\langle$	-			

## 3.2 รายละเอียดที่นักศึกษาปฏิบัติในการปฏิบัติงาน

เนื่องจากผู้จัดทำโครงงานได้เข้าร่วมสหกิจในช่วงที่แผนก Academy & BI ไม่ค่อยได้รับการแจ้ง ปัญหาจากลูกค้า และยังอยู่ในช่วงการทำ MA ให้กับลูกค้า ดังนั้นผู้จัดทำโครงงานจึงได้รับมอบหมายให้ทำ การตรวจสอบรายงานต่าง ๆ ของลูกค้า และทำการศึกษาเครื่องมือต่าง ๆ ที่ลูกค้าใช้งาน เพื่อเตรียมความ พร้อมในการรับมือกับการแก้ไขปัญหาที่ได้รับแจ้งจากลูกค้า นอกจากนี้ยังได้รับมอบหมายโครงงานในการ ทำการสาธิตกระบวนการ ETL ของระบบ Business Intelligence ด้วยโปรแกรม IBM® InfoSphere® DataStage® ซึ่งเป็นโปรแกรมที่ใช้ในการแปรรูปข้อมูล เนื่องด้วยศักยภาพในการประมวลผลที่สูงของ IBM® InfoSphere® DataStage® จึงทำให้ทางผู้ประกอบการต้องการทราบถึงความสามารถของโปรแกรม รวมถึงวิธีการใช้งาน เพื่อนำมาเปรียบเทียบกับโปรแกรมอื่น ๆ ที่ทางสถานประกอบการได้เป็นตัวแทน จำหน่ายต่อไป

## 3.3 ขั้นตอนการดำเนินงานที่นักศึกษาปฏิบัติงาน

#### 3.3.1 ศึกษาและทำความเข้าใจ BI Concept

ศึกษา Concept ต่าง ๆ ที่เกี่ยวข้องกับการทำ BI ได้แก่ Data Source, Data Warehouse, ETL, ELT, Cube, และ Report เพื่อเป็นการทำความเข้าใจกระบวนการทำงานขั้นพื้นฐาน

## 3.3.2 ศึกษาโปรแกรมอื่นที่ใ<mark>ช้ในการทำ BI</mark>

ศึกษาซอฟต์แวร์ต่าง ๆ เช่น Oracle Warehouse Builder 11g, IBM® Cognos Analytics, Oracle Business Intelligence Enterprise Edition 12c เพื่อสนับสนุนการทำงานของแผนก และทำความเข้าใจกับกระบวนการ ต่าง ๆ ซึ่งรวมทั้งการติดตั้ง

#### 3.3.3 ศึกษา Concept ของ DataStage®

ศึกษา Concept ต่าง ๆ ของ DataStage® ตัวอย่างเช่น Parallelism, Data Pipelining, Data Partitioning, Dynamic Repartitioning, Data Repartitioning ซึ่งช่วยในการทำความเข้าใจกระบวนการรูปแบบการทำงาน ต่าง ๆ ของ Stage ภายใน DataStage®

#### 3.3.4 จัดทำเอกสารแนะนำ DataStage®

จัดทำเอกสารเพื่อแนะนำความสามารถและอธิบาย Concept ต่าง ๆ ของ DataStage®

#### 3.3.5 ศึกษาการใช้งาน Stage ต่าง ๆ ของ DataStage®

ศึกษาการใช้งาน Stage ต่าง ๆ ของ DataStage® โดยประกอบไปด้วย Stage ต่าง ๆ ดังนี้ DB2 Connector Stage, Unstructured Data Stage, Sequential File Stage, Transformer Stage, Sort Stage, Remove Duplicates Stage, Aggregator Stage, Checksum Stage, Row Generator Stage, Peek Stage, Sample Stage, Filter Stage, Funnel Stage, Lookup Stage, Join Stage, Merge Stage, Difference Stage, Change Capture Stage, Change Apply Stage, และ Pivot Enterprise Stage ซึ่ง Stage ต่าง ๆ ที่เลือกมาล้วนแต่เป็น Stage พื้นฐาน ที่มีแนวโน้มจะใช้ งานสูง หรือเป็น Stage ที่เป็นเอกลักษณ์ของ DataStage®

#### 3.3.6 นำเสนอวิธีการใช้งาน <mark>Stage ต่าง ๆ</mark> ของ Data<mark>Stage</mark>®

นำเสนอวิธีการใช้งานต่อพนักงานที่ปรึกษา โดยเจา<mark>ะ</mark>จงไป<mark>ที่ความส</mark>ามารถ<mark>ของ</mark> Stage ต่าง ๆ รวมถึงข้อมูล ก่อนการแปรรูป และข้อมูลหลังกา<mark>รแป</mark>รรูป

## STITUTE O

### 3.3.7 จัดทำคู่มือการใช้งาน DataStage®

จัดทำกู่มืออธิบายการตั้งค่า Stage ต่าง ๆ ของ DataStage® พร้อมทั้งตัวอย่างการใช้งานและผลลัพธ์ ประกอบการอธิบาย

#### 3.3.8 ตรวจสอบคู่มือ DataStage®

นำคู่มือการใช้งานที่จัดทำมาใช้งาน เพื่อเป็นการตรวจสอบความถูกต้องของคู่มือที่ได้จัดทำขึ้น

## 3.3.9 แก้ใขปรับปรุงคู่มือ DataStage®

16

ปรับปรุงคู่มือที่จัดทำ ให้มีความถูกต้องแม่นยำมากขึ้น รวมทั้งปรับปรุงตามความเหมาะสมตามที่ได้ปรึกษา กับพนักงานที่ปรึกษา

นโล*ย*ั

## บทที่ 4

## สรุปผลการดำเนินงาน การวิเคราะห์และสรุปผลต่าง ๆ

### 4.1 ส่วนประกอบต่าง ๆ ของ IBM® InfoSphere® DataStage®

#### 4.1.1 Parallel and Sequential

ในการทำงานของ DataStage® นั้นจะมีวิธีประมวลผล 2 แบบ คือ Parallel และ Sequential ซึ่งการประมวลผล แบบ Parallel นั้นจะมีการประมวลผลที่รวดเร็วกว่า เนื่องจากมีการแบ่งข้อมูลออกเป็น Partition โดยไม่มี ผลกระทบต่อการคำนวณและการตีความข้อมูล ซึ่งทาง IBM® ได้สร้างตรรกะในการแบ่ง Partition เอาไว้ให้ แล้ว ดังนั้นผู้ใช้จึงไม่จำเป็นต้องจัดทำเอง แต่การประมวลผลแบบ Parallel นั้น บางครั้งจะได้ผลลัพธ์แตกต่าง จากที่ต้องการ ตัวอย่างเช่น Sort Stage เหมือนดั่งรูปที่ 4.1 ซึ่งทางซ้ายคือการใช้ Sort Stage ที่กำหนดให้ทำงาน แบบ Sequential ในขณะที่ทางขวาคือ Sort Stage ที่กำหนดให้ทำงานแบบ Parallel โดยสามารถกำหนด รูปแบบการทำงานได้โดยการ Double Click Stage ที่ต้องการกำหนด แล้วเลือก Tab Stage > Advance ซึ่งใน กรณีที่มี Preserve Partitioning มีค่าเป็น Default(Set) หรือ Set ก็จำเป็นจะต้องเปลี่ยนเป็น Clear เสียก่อน ดัง รูปที่ 4.2 และ 4.3



ร**ูปที่ 4. 1** ตัวอย่างการใช้ Sort S<mark>tage</mark> ด้วยการปร<mark>ะ</mark>มวลผล<u>แบบ S</u>equen<mark>tial (</mark>ซ้าย) และแบบ Parallel (ขวา)

ഫി			
Stage Input Output			
Stage name: Sort_7 General Properties Advanced NLS	Locale )		
Execution mode:	Combinability mode: (Auto)	Preserve partitioning: Clear	•
รปที่ 4. 2 ต <b>ัวอ</b> ย่	างการตั้งค่า Preserve Partiti	oning เป็น Clear	

38



ร**ูปที่ 4. 3** สัญลักษณ์การเปลี่ยนรูปแบบการประมวลผลจาก Sequential เป็น Parallel (วงกลม ทางซ้าย) และเปลี่ยนแปลงจาก Parallel เป็น Sequential (วงกลมทางขวา)

#### 4.2 Canvas

เมื่อทำการเปิดโปรแกรม InfoSphe<mark>re®</mark> DataStage® Designer และทำการ Log In แล้วจะเข้ามาที่หน้าต่างดัง รูปที่ 4.4 ให้เลือก Parallel Job ก็จะ<mark>ปราก</mark>ฏ Canvas สำหรับออกแบบ Parallel J</mark>ob ดังรูปที่ 4.5



ในส่วนของ menu bar ด้านบนจะมี icon 2 อันที่ใช้ในการกำหนดเกี่ยวกับ grid ดังรูปที่ 4.6 – 4.10

File Edit	View Repository Diag	Iram Debug	Import	Export	Tools	Window	Help														
1 - 💋	0 10 12 10 10	6   A U	<i>»</i> 🛛	🤹   👘	1	B 🕨 🚺	î [	田	<b>-</b>	₿ 6	1 🍛 (										
Repository			믜×																		
DSProject	1	Open quick	find	2							 	Para	illel -	Untit	iea. i	 	 	 	 	 	-
🖃 🔒 DS	Project		_1	I P																	<u>^</u>
Ē	U Test																				
🗎 🔅 🫅	_DSProject																				
🗎 🖻 🧰	Data Elements		LI																		
🕴 😐 🚞	IMS Databases (DBD)																				
📃 🔅 🚞	IMS Viewsets (PSB/PCB)																				
🗎 🖻 🚞	Jobs																				
🗎 🖻 🧰	Machine Profiles																				
📃 🕀 🧰	Match Specifications																				
📃 🕀 🧰	MNS Rules																				
🗎 🖻 🚞	Routines																				
🗎 🖻 🧰	Shared Containers																				× .
🗉 🕀 🚞	Stage Types			<																>	
😐 🗀	Standardization Rules			H H	• • U	Intitled.1															
🛱 🍋	Table Definitions						-				_										

รูปที่ 4. 6 ตำแหน่งของ Show grid lines และ Snap to grid

## Show grid lines

ร**ูปที่ 4. 8** เปิดใช้ Show grid lines เพื่อแสดงเส้นตาราง

## 🖪 Snap to grid

10

รูปที่ 4. 7 เปิดใช้ Snap to grid เพื่อถือกให้ stage ต่าง ๆอยู่ในกรอบ



## <mark>รูปที่ 4. 10</mark> เมื่อไม่ได้เปิดใช้ Snap to grid



รูปที่ 4. 9 เมื่อเปิดใช้ Snap to grid

#### 4.1.3 Repository

ในการสร้าง Folder ใน Repository สามารถทำได้โดย คลิกขวาที่ repository แล้วเลือก New > Folder ดังรูปที่ 4.11

Repository		<u> </u>	1
DSProject		Open quick find	1
Byrotext     Eind in     Export     Da     Da	this folder	Folder	

รูปที่ 4. 11 วิธีการสร้างโฟลเดอร์ใหม่

จากนั้นทำการตั้งชื่อ Folder ดังรูปที่ 4.12

T

Repository	
DSProject Open quick	<u>c find</u>
DSProject    U Test    DSProject    DSProject    Data Elements    IMS Databases (DBD)    IMS Viewsets (PSB/PCB)    Jobs	< III
Machine Profiles      Machine Profiles      Machine Profiles      Machine Profiles      Mover Folder      Mover Folder      Mover Folder      Mover Folder      Mover Folder      Stared Containers      Stage Types      Grading All      Grad	~

รูปที่ 4. 12 ตั้งชื่อโฟลเดอร์ใหม่

ในรูปที่ 4.13 จะสังเกตได้ว่า โฟลเดอร์ที่ถูกสร้างมา นั้น จะไม่ถูกเรียงตามตัวอักษร



รูปที่ 4. 13 เมื่อสร้าง โฟลเคอร์เสร็จและยังไม่ทำการรีเฟรช

เลือก repository ใน menu bar จากนั้นเลือก refresh เพื่อให้โฟลเดอร์ที่สร้างขึ้นใหม่ถูกจัดเรียงตามตัวอักษร ดังรูปที่ 4.14



ร**ูปที่ <mark>4. 14</mark> วิธีการ Ref<mark>r</mark>esh Repo<mark>sitory</mark> และผ<mark>ลลัพ</mark>ธ์** 

#### 4.1.4 Data Connection

ใน DataStage® นั้น สามารถที่จะ save Data Connection ได้ ซึ่ง Data Connection ก็คือข้อมูลที่เก็บรายละเอียด การเชื่อมต่อเกี่ยวกับฐานข้อมูล ซึ่งสามารถทำได้โดยการคลิกขวาในหน้าต่าง Repository จากนั้นจึงเลือก New > Other > Data Connection ดังรูปที่ 4.15



### รูปที่ 4. 15 สร้าง Data Connection ใหม่

กรอกชื่อที่ด้องการจะบันทึกในช่อง Data Connection Name แถ้วจึงเลือก Tab Parameters ดังรูปที่ 4.16

กดที่ปุ่ม ellipsis ดังรูปที่ 4.17

TC

□ \_U Project\Data Conne	ction\InsuranceDB	2 - Data Connecti	_ 🗆 X
<u>G</u> eneral <u>P</u> arameters			
Connect using Stage Type:			
Connection parameters			
Parameter name		Value	
Associated Parameter Set:			
	0	K Cancel	Help

รูปที่ 4. 17 Data Connection Tab Parameter

จากนั้นเลือกรูปแบบของฐานข้อมูลที่ต้องการเชื่อมต่อ (ในกรณีนี้เลือก DB2 Connector) แล้วกคปุ่ม Open

(โดย Connector ต่าง ๆจะอยู่ในโฟลเดอร์ Stage Types > Parallel > Database) ดังรูปที่ 4.18

2	Open	Contraction of the local division of the loc	x
DSProject	]	🗈 🗙 💣 🤉	Close quick find
Quick fin	d		×
Name to f	ind:		
			•
Types to	find:		
(All types	)		
🗖 Inclue	le descriptions		
			Find
	Project		_
÷		-	=
÷.	🛅 Mainframe		
	Parallel		
	Classic Federation		
	Cognos TM1 Connector		
	DB2 UDB API		
			~
Item name:			Open
DB2 Conne	ctor		opon
Item type:			Cancel
Parallel Sta	де Туре		Help

ร**ูปที่ 4. 18** เลือกประเภทของ Connector

กรอก Connection Name, Username ในช่องว่าง

เมื่อกลิกเพื่อจะกรอก Password จะมีหน้าต่างขึ้นมาดังรูปที่ 4.19 ก็กรอกรหัส 2 ครั้ง แล้วกด OK

Sequence Table Definition		🚇 \_U Project\Data Connection\InsuranceDB2 - Data Connecti 📒 🗖	x
Password - Encrypted		<u>G</u> eneral <u>P</u> arameters	
Encrypted String:	OK Cancel	Connect using Stage Type:	
	<u>H</u> elp	Parameter name Value AHSTEST	
outines		2 Usemame db2inst1 3 Password 4 Instance	
	-	Associated Parameter Set:	
tems from the Repository Tree or by using the	21	ОК Cancel H	elp

รูปที่ 4. 19 ช่องกรอกรหัสระหว่างทำ Data Connection

แล้วกด OK อีกครั้ง ก็จะมีหน้าต่างเปิดขึ้นมา เพื่อให้กำหนด Path ที่จะทำการ Save Data Connection ดังรูปที่

4.20 Save Data Connection A DSProject | 10 🖻 🗙 🖻 Quick fin × ame to fir Project\Data Connection\InsuranceDB2 - Data Connecti... 💻 🗖 • Types to fin (All types) Parameters • db2inst1 ОК Cancel Help Cancel path: roject\Data Conr รูปที่ 4. 20 Path ที่ต้องการบันทึก Data Connection

เมื่อกด Save ก็จะมี Data Connection ใน Path ที่ได้กำหนดไว้ ดังรูปที่ 4.21



ร**ูปที่ 4. 21** เมื่อเสร็จสิ้นการบันทึก Data Connection

#### 4.1.5 Create Parallel Job

(0

Parallel Job สามารถสร้างได้โดย คลิกขวาภายใน repository หรือกด File > New > Parallel Job ใน menu bar ดังรูปที่ 4.22



ร**ูปที่ 4. 22** วิธีการสร้าง Parallel Job ใหม่

## เมื่อเลือกแล้วก็จะปรากฏ canvas ของ parallel job ดังรูปที่ 4.23



รูปที่ 4. 23 Canvas ของ Parallel Job

## บันทึก Parallel Job ใด้โดยการเลือก File > Save As... ดังรูปที่ 4.24



STITUTE OF

🐔 Save Parallel Job As	x
DSProject 🗈 🗙 🖆 🤤	Close quick find
Quick find	×
Name to find:	
Types to find:	
Include descriptions	
	Find
⊡-/iि DSProject	
U Project	
Data Connection	
ie in table Definition	>
⊡	
Data Elements	
IMS Databases (DBD)     IMS Viewsets (PSB/PCB)	
E ⊡ Jobs	~
D Time xcel to DB2	Save
Folder path:	Cancel
\U Project\Parallel	Help

ตั้งชื่อ Parallel Job และกำหนด Path ที่ต้องการบันทึกแล้วกดปุ่ม Save ดังรูปที่ 4.25

รูปที่ 4. 25 กำหนด Path ที่ต้องการทำการบันทึก Parallel Job

Parallel Job ที่ทำการบันทึกจะปรากฏใน repository ดังรูปที่ 4.26



รูปที่ 4. 26 เมื่อเสร็จสิ้นการบันทึก Parallel Job

#### 4.1.6 Table Definition

ใน DataStage® Stage ต่าง ๆ จะมี Tab Output หรือ Input หรือทั้งสองอย่าง ขึ้นอยู่กับการออกแบบ โดยใน Tab เหล่านี้จะมี Tab Columns ที่จะกำหนดประเภทและขนาดของข้อมูลต่าง ๆ ที่ทำการนำเข้ามาภายในหรือ ส่งออกจาก Stage นั้น ๆ ซึ่งค่าที่ตั้งนั้น สามารถบันทึกเป็น Table Definition และสามารถนำมา Load เพื่อลด ระยะเวลาในการตั้งค่าได้ ดังรูปที่ 4.27 โดยการตั้งค่าต่าง ๆ นั้น จะถูกผูกกับ Link ที่ทำอยู่ ดังนั้น ถ้าลบ Link ก็จะต้องทำการตั้งค่าในส่วนนี้ใหม่



#### 4.2 วิธีการใช้งาน Stage ต่าง ๆ ของ DataStage®

Stage ต่าง ๆ สำหรับการใช้งานใน Parallel Job สามารถหาได้ใน DSProject > Stage Types > Parallel ดังรูป ที่ 4.28



ร**ูปที่ 4. 28** ตำแหน่งโฟลเดอร์ต่าง ๆ ที่บรรจุ Stage สำหรับ Parallel Job

## ซึ่งจะมีการกล่าวถึงการตั้งค่าและการใช้งาน Stage ต่าง ๆ เรียงตามลำคับคังต่อไปนี้

- 1. DB2 Connector Stage (Read)
- 2. DB2 Connector Stage (Write)
- 3. Unstructured Data Stage (Read)
- 4. Unstructured Data Stage (Write)
- 5. Sequential File Stage (Read)
- 6. Sequential File Stage (Write)
- 7. Transformer Stage
- 8. Sort Stage
- 9. Remove Duplicates Stage
- 10. Aggregator Stage

- 12. Row Generator Stage
- 13. Peek Stage
- 14. Sample Stage
- 15. Filter Stage
- 16. Funnel Stage
- 17. Lookup Stage
- 18. Join Stage
- 19. Merge Stage
- 20. Difference Stage
- 21. Change Capture Stage
- 22. Change Apply Stage
- 23. Pivot Enterprise Stage

การเพิ่ม Stage ลงไปใน Canvas ทำได้โดยลากจาก Repository ลงไปที่ Canvas โดยตรง เมื่อมี Stage ต่าง ๆ บน Canvas แล้ว ก็ต้องทำการสร้าง Link เพื่อเชื่อมต่อ ซึ่งสามารถทำได้โดยการคลิกที่ Stage ต้นทาง เพื่อให้ ขึ้นขอบรอบ ๆ Stage ที่เลือก แล้วนำ Cursor ไปวางไว้บนขอบ ซึ่ง Cursor จะเปลี่ยนเป็นรูป วงกลมสีดำสอง วงซ้อนกัน แล้วจึงลากไปยัง Stage ปลายทาง ซึ่งการตั้งค่า Stage ต่าง ๆ จะสัมพันธ์กับทิศทางของ Link

ุ น โ ล *ชี ไ ก* 

นอกจากนี้ Stage ต่าง ๆ ของ DataStage® ยังสามารถเจาะจงรูปแบบการประมวลผลว่าจะให้เป็นรูปแบบ Sequential หรือ Parallel ซึ่งสามารถเลือกใช้ได้ตามความเหมาะสม ซึ่งการตั้งค่านี้ สามารถพบได้ใน Tab Stage > Advance ซึ่งในบาง Stage ช่อง Preserve partitioning มีก่า Default เป็น Set ก็ต้องแก้ไขให้เป็น Clear ดังรูปที่ 4.29

# STITUTE O

all				
Stage Input Output				
Stage name: Sort_7				
General Properties Advanced NLS	6 Locale			
Execution mode: Sequential	Combinability mode: (Auto)	•	Preserve partitioning:	-

รูปที่ 4. 29 การ Clear Preserve Partitioning เพื่อรองรับการประมวลผลแบบ Sequential

#### 4.2.1 DB2 Connector Stage (Read)



รูปที่ 4. 30 รูป DB2 Connector Stage (Output)

## **IC**

เมื่อทำการลาก DB2 Connector Stage ลงมาใน Canvas ดังรูปที่ 4.30 และมีการกำหนด Link ให้นำข้อมูลออก จะมีการตั้งก่าต่าง ๆ ดังรูปที่ 4.31 ดังนี้

Connection ส่วนที่เจาะจงการเชื่อมต่อเข้ากับฐานข้อมูล เมื่อทำการกรอกแล้ว สามารถบันทึกไว้ใน รูปแบบ Data Connection ทำให้สามารถ Load การตั้งก่าต่าง ๆ มาใช้ได้ทันที เมื่อทำการกรอกข้อมูลกรบถ้วน แล้ว สามารถกดที่ปุ่ม Test เพื่อทำการทุดสอบการเชื่อมต่อได้

View Data สามารถใช้เพื่อดู<mark>ตัวอย่</mark>างข้อมูลที่จ<mark>ะ</mark>ทำการเรียก<mark>ออ</mark>กมาไ<mark>ด้ ดังรู</mark>ปที่ 4.32

Generate SQL ถ้าเลือกเป็น Yes จะทำการเปิดช่อง Table name และปิดช่อง Select statement ผลลัพธ์ที่ได้ คือ select \* from [Table name] Prefix for expression columns กำหนดชื่อให้คอลัมน์ที่เกิดจากการคำนวณด้วย Select statement Select statement ผู้ใช้สามารถกำหนดเงื่อนไขสำหรับการเรียกข้อมูลได้โดยตรง

Fail on size mismatch จะยกเลิกการทำงานถ้าขนาดของคอลัมน์ที่กำหนดใน DataStage® ไม่เหมือนกับที่ กำหนดไว้ในฐานข้อมูล

Fail on type mismatch จะยกเลิกการทำงานถ้าประเภทของคอลัมน์ที่กำหนดใน DataStage® ไม่เหมือนกับ ที่กำหนดไว้ในฐานข้อมูล

Duput   t name (downstream stage) USLink2 (Sequential_File_1) erai Properties Columns Advanced   Connection (Associated data conn Instance Detabase * User name	ction: /hsuranceDB2) AHSTEST	u la â	7.	Tet Loss Same
h nam (downsteam stage) SLink2 (Sequentia_Fic_1) eni Properties Columns Advanced ) Connection (Associated data conn Instance Database * User name	ction: InsuranceDB2) AHSTEST	u i a a	7	Tet Loss jäne
r name (cownamean stage) SLink2 (Sequential_File_1) ani Properties Columns Advanced Semection (Associated data comn Instance Database	ection: Insurance/DB2) AHSTEST		7	Tex Lost Save
SLink2 (Sequential_File_1) erail Properties [Columns   Advanced   Connection (Associated data conn Instance Database * User name	sclion: Insurance/DB2) AHSTEST			Test Load Save
eral Properties Columns Advanced Connection (Associated data conn Instance Database * User name	ction: (nsuranceDB2) AHSTEST		17	Test Load Save
eral Propeties Columns Advanced   Connection (Associated data conn Instance Database User name	AHSTEST		- 7	Test Load Save
eral Propeties [Columns   Advanced   Connection (Associated data conn Instance Database * User name	AHSTEST			Test Load Save
Connection (Associated data conn Instance Database • User name	AHSTEST			Test Load Save
Instance Database * User name	AHSTEST			
User name	AHSTEST			
User name				
	db/2inst 1			
Password				
UB2 client library file				
Atemate conductor settings	NO			
Use direct connections	NO			
Neep conductor connection alive	Tes			New Design
Usage				Mew Data
Generate SQL	No			
Table name				
Enable quoted identifiers	No			
Pretix for expression columns	EXPR			
V SUL				
Select statement	SELECT FROM D_TIME_SORT;			
Read select statement from file	No			
Transaction	NO			
Transdiction	2000			
	ZUUU			
Social	Note			
	C			
Isolation level	Cursor stability			
Auto commit mode	2000			
<ul> <li>Schema reconciliation</li> </ul>				
End on size mismatch	Yee			
Fail on time mismatch	Yee			
Enable LOB references	No			
Refore /After SQL	No			
Repatinization	None			
Lock wait mode	Use the lock timeout database configuration parameter			
Limit number of returned rows	No.			

รูปที่ 4. 31 Output Properties ของ DB2 Connector Stage

			View Data	
YEAR_CD	MONTH_KEY	MONTH_NM	MONTH_CD	
2011	1	JAN	201101	
2011	2	FEB	201102	
2011	3	MAR	201103	
2011	4	APR	201104	
2011	5	MAY	201105	
2011	6	JUN	201106	
2011	7	JUL	201107	
2011	8	AUG	201108	
2011	9	SEP	201109	
2011	10	OCT	201110	
2011	11	NOV	201111	
2011	12	DEC	201112	
2012	1	JAN	201201	
2012	2	FEB	201202	<u></u>
2012	3	MAR	201203	57 >
2012	4	APR	201204	
2012	5	MAY	201205	
2012	6	JUN	201206	

รูปที่ 4. 32 ตัวอย่างการ View Data ด้วย DB2 Connector Stage

#### 4.2.2 DB2 Connector Stage (Write)



รูปที่ 4. 33 รูป DB2 Connector Stage (Input)

เมื่อทำการลาก DB2 Connector St<mark>age ล</mark>งมาใน Ca<mark>n</mark>vas ดังรูปที่ <mark>4.3</mark>3 แล<mark>ะมีการ</mark>กำหนด Link ให้นำข้อมูลเข้า จะมีการตั้งค่าต่าง ๆ ดังรูปที่ 4.34 ดังนี้ Connection ส่วนที่เจาะจงการเชื่อมต่อเข้ากับฐานข้อมูล เมื่อทำการกรอกแล้ว สามารถบันทึกไว้ใน รูปแบบ Data Connection ทำให้สามารถ Load การตั้งค่าต่าง ๆ มาใช้ได้ทันที เมื่อทำการกรอกข้อมูลครบถ้วน แล้ว สามารถกดที่ปุ่ม Test เพื่อทำการทดสอบการเชื่อมต่อได้

View Data สามารถใช้เพื่อดูตัวอย่างข้อมูลที่อยู่ภายในฐานข้อมูลได้

Write mode กำหนดรูปแบบของการบันทึกข้อมูล เช่น Insert, Insert new rows only, Update, Delete, Insert then update, Update then insert, Delete then insert เป็นต้น

Generate SQL ถ้าเลือกเป็น Yes ปีดช่อง Select statement และจะสร้างโค้ดให้สัมพันธ์กับการตั้งค่าอื่น ๆ
 Prefix for expression columns กำหนดชื่อให้คอลัมน์ที่เกิดจากการคำนวณด้วย Select statement
 SQL statement ผู้ใช้สามารถกำหนดเงื่อนไขสำหรับการจัดการข้อมูลได้โดยตรง โดยจะเปิดให้กำหนดค่า
 ต่าง ๆ ที่สัมพันธ์กับ Write mode

 Table action กำหนดวิธีการจัดการข้อมูลใหม่กับ Table ซึ่งได้แก่ Append, Create, Replace, Truncate
 Fail on size mismatch จะยกเลิกการทำงานถ้าขนาดของคอลัมน์ที่กำหนดใน DataStage® ไม่เหมือนกับที่ กำหนดไว้ในฐานข้อมูล

Fail on type mismatch จะยกเลิกการทำงานถ้าประเภทของคอลัมน์ที่กำหนดใน DataStage® ไม่เหมือนกับ ที่กำหนดไว้ในฐานข้อมูล

Drop unmatched fields ถ้ากำหนดเป็น Yes จะไม่บันทึกข้อมูลที่มี Field ไม่ตรงกัน Fail on row error จะยกเลิกการทำงานถ้าพบแถวที่มีความผิดพลาคระหว่างการบันทึก

# STITUTE O

	DB2_Connector_0 - DB2 Connector
Innut	
in pos	
name (upstream stage)	
SLink8 (Sort_7)	
Properties   Column   Advanced   Destrict	
eral Properties Columns Advanced Partition	
onnection (Associated data connection	
Instance	AUCTECT
Jalabase	AN31531 do2art1
Password	
DB2 client library file	
Alternate conductor settings	No
Use direct connections	No
Keep conductor connection alive	Yes
Jsage	Mew Data
Write mode *	Inset
Generate SQL	No
Table name *	D_TIME_SORT
Enable quoted identifiers	No
Prefix for expression columns *	EXPR
V SQL	
Insert statement *	INSERT INTO D_TIME_SORT(YEAR_CD.MONTH_KEY.MONTH_NM.MONTH_CD) VALUES(ORCHESTRATE YEAR_CD.ORCHESTRATE.MONTH_KEY.ORCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHESTRATE.MONTH_KEY.ONCHEST
Update statement	
Delete statement	
Table satis	Reduct
Generate create statement at numme	Yee
Fail on error	Yes
Create statement *	
<ul> <li>Generate drop statement at runtime</li> </ul>	Yes
Fail on error	No
Drop statement *	
Generate truncate statement at runtime	Yes
<ul> <li>Transaction</li> </ul>	
Record count	2000
▼ Session	
Isolation level	Cursor stability
Auto commt mode	
Artay size	200
Fail on aire minmatch	Ym
Fail on type mismatch	109 Vas
Drop unmatched fields	Yes
Fail on row error	Yes
Insert buffering	Default
Atomic arrays	Auto
<ul> <li>Logging</li> </ul>	No
Log column values on first row error	
Log column values on first row error     Before/After SQL	No
Log column values on first row error     Before/After SQL Reoptimization	No None
Logging     Log column values on first row error     Before/After SQL Reoptimization     Lock wait mode	No None Use the lock timeout database configuration parameter
Logging     Log column values on first row error     Before/After SQL Reoptimization     Lock wait mode Pad character	No None Use the lock timeout database configuration parameter
Logolum values on first row error     Before/After SQL Reoptimization     Lock wait mode     Pad character     Unit parallelism	No None Use the lock timeout database configuration parameter No
Log column values on first row error     Befork/Atter SQL      Beoptimization     Lock wat mode     dard baracter     Linit: paralelem     Buk load to DB2 on 2/05	No None Use the lock timeout database configuration parameter No

รูปที่ 4. 34 Input Properties ของ DB2 Connector Stage

#### 56

#### 4.2.3 Unstructured Data Stage (Read)



รูปที่ 4. 35 รูป Unstructured Data Stage (Output)

Unstructured Data Stage เป็น stage ที่ใช้สำหรับการอ่านและเขียนไฟล์ Excel ดังรูปที่ 4.35 และรูปที่ 4.36 ซึ่ง สามารถตั้งค่าต่างๆได้ดังนี้

tage Output	Stage Output	
tage name	Output name (downstream stage)	
Instructured_Data_1	DSLink2 (BASIC_Transformer_3)	
General Properties Advanced	General Properties Link Ordering Adv	vanced   Columns
Document type Excel	Texcel	
Error action Fail	Skip null rows	Yes
Fail	<ul> <li>Runtime column propagation</li> </ul>	
Skin		E

รูปที่ 4. 36 Stage Properties (ซ้าย) และ Output Properties (ขวา) ของ Unstructured Data Stage

#### **Stage > Properties**

Document type [Excel] เ<mark>ลือกป</mark>ระเภทของ<mark>เ</mark>อก<mark>สาร</mark>

Error action [Fail, Skip] เลือกวิธีการรับมือเมื่อเกิดข้อผิดพลาด โดย skip จะไม่หยุดการทำงานเมื่อ เกิดข้อผิดพลาด แต่จะอ่านข้อมูลเท่าที่จะสามารถอ่านได้ ในขณะที่ Fail จะหยุดการทำงานของ Parallel Job เมื่อเกิดข้อผิดพลาดขึ้นระหว่างการอ่านข้อมูล
### **Output > Properties**

Skip null rows [Yes, No] การอ่านข้อมูล จะอ่านแถวที่มีค่าเป็น null ทั้งหมดด้วยหรือไม่ ถ้าต้องการ แถวที่มีค่าเป็น null ทั้งหมด ให้ทำการตั้งค่าเป็น No

Hidden columns [Extract, Skip] กำหนดว่าจะทำอย่างไรกับแถวที่ถูกซ่อนเอาไว้ ระหว่างอ่านแถวที่ ซ่อนเอาไว้ (Extract) หรือข้ามแถวที่ถูกซ่อน (Skip)

<b>a</b>	Unstructured_Data_1 - Unstructured Data	_ 0 ×
Stage Output		
Output name (downstream stage)		
DSLink2 (BASIC_Transformer_3)	▼ Confirme	
	Configuration Window for Unstructured Data stage	×
Data source		
File name: *	Link: DSLink2	
D:\InsuranceDemo\Data.xls	Read options 2 Import	3
Password:	Range option: Excel column Property Advanced	
	Specify the start row  Page expression: * Template data range: D_ACCOUNTING_DATE(A1:F1827	
Use template file for design time	D ACCOUNTING DATEIA1:F182	
Template file name:	Sheet names to skip:	JW
	Column #1 (b) MONTH CD	
Template password:	Column header:	
	First row of data ranges V	
	Load View	
		Import
	Map	
	map	4
	DataStage Column Excel Item Import Option	
	YEAR_CD Column #1 (YEAR Value	
	MONTH_KEY Column #4 (MONT Value	
		Move up
	M	ove down
		Insert
		Delete
		()
$\leq$		
<b>Y</b> .		
· · · ·		
รปที่ 4	. 37 หน้าต่าง Configure ของ Unstructured Data Stage (Output)	
Q.		

## จากรูปที่ 4.37

้ส่วนที่ 1 คือส่วนที่ระบุตำแหน่งของไฟล์ที่ต้องการจะอ่านและรหัสของไฟล์(ถ้ามี)

ส่วนที่ 2 คือส่วนที่กำหนดว่าจะทำการอ่านส่วนใหนของไฟล์ โดยจะประกอบไปด้วยชื่อของ Sheet หมายเลข แถวและตัวอักษรประจำคอลัมน์ พร้อมทั้งระบุว่าแถวแรกของไฟล์นั้นเป็นชื่อคอลัมน์หรือไม่ ซึ่งในส่วนนี้จะ สอดกล้องกับส่วนที่ 3 ตัวอย่างเช่น ถ้าไม่ได้ระบุ Range expression และ Sheet names to skip ในส่วนนี้แล้ว กด Load ส่วนที่ 3 ก็จะแสดงข้อมูลสำหรับนำเข้าทุก Sheet ทุกคอลัมน์ ซึ่งเมื่อสั่ง Import ในส่วนที่ 3 ช่อง Range expression ของส่วนนี้ก็จะทำการกำหนดให้โดยอัตโนมัติ

ส่วนที่ 3 ในส่วนนี้จะทำการเลือกคอลัมน์ที่ต้องการนำข้อมูลมาใช้ โดยทำการทำเครื่องหมายถูกที่คอลัมน์ที่ ต้องการแล้วจึงกด Import ก็จะทำการส่งชื่อคอลัมน์ต่อไปในส่วนที่ 4

ส่วนที่ 4 ส่วนนี้มีไว้เพื่อจับคู่ข้อมูลระหว่างคอลัมน์ของ Excel และคอลัมน์ที่ได้กำหนดไว้ใน Table Definition ของทาง DataStage® โดยค่าที่นำเข้าสามารถระบุได้ว่าให้เก็บเป็นรูปแบบใด ตัวอย่างเช่น expression ของ excel

## 4.2.4 Unstructured Data Stage (Write)

10



รูปที่ 4.38 รูป Unstructured Data Stage (Input)

Unstructured Data Stage ในฝั่งการ<mark>บันทึ</mark>ก จะมีการบันทึกค่า<mark>ตามรูป</mark>แบบที่ได้ตั้งไว้ใน Table Definition ดังรูป ที่ 4.38 ตัวอย่างเช่น ถ้าตั้งค่าคอลัม<mark>น์เป็น</mark> VarChar เมื่อทำการบันทึกลง Excel ก็จะถือว่าเป็นข้อความ ไม่ใช่ ตัวเลข ดังนั้นจึงไม่สามารถทำการคำนวนได้ โดยการบันทึกจะมีการตั้งค่าที่แตกต่างออกไปดังนี้

ji -	
Stage Input	
Stage name	
Unstructured_Data_1	
General Properties	Advanced
🛡 Usage	
Document type	Excel
Texcel	
Write mode	Create a file
1	Modify an existing file
1 <u>0</u> 1	Create a file

60

รูปที่ 4. 39 Stage Properties ของ Unstructured Data Stage (Input)

### Stage > Properties

จาก Error action จะเปลี่ยนเป็น Excel > Write mode ซึ่งสามารถเลือกได้สองแบบคือการแก้ไขไฟล์ที่มีอยู่แล้ว (Modify an existing file) และสร้างไฟล์ใหม่ (Create a file) ดังรูปที่ 4.39

## Configure

ในส่วนของหน้าต่าง configure นั้น แตกต่างจากขั้นตอนการอ่านโดยสิ้นเชิง โดยจะประกอบด้วยส่วนต่าง ๆ ดังนี้

Output file 1	Properties	2
Write method Specific file File name * D:\InsuranceDemo\test.xlsx	Set for all links	v
File update mode	Property	Value
Overwrite View	Column header Adjust column width	Column names V No V
U I N R	Sheet order Link name Sh DSLink4 She	a meet name meet 1 Move up Move down
		OK Cancel Help

รูปที่ 4. 40 หน้าต่าง Configure ของ Unstructured Data Stage (Input)

จากรูปที่ 4.40 ส่วนที่ 1 เป็นส่วนที่ระบุวิธีการบันทึกข้อมูลลงบนไฟล์ พร้อมทั้งตำแหน่งของไฟล์ และวิธีการ อัพเดท โดยวิธีการบันทึกจะมีแบบบันทึกลงไฟล์ที่เจาะจง (Specific file) และบันทึกแบบสร้างหลายไฟล์ (Generate multiple files) โดยการสร้างแบบหลายไฟล์ จะต้องกำหนดจำนวนแถวสูงสุดต่อ Excel Sheet ซึ่ง การตั้งค่าในส่วนนี้จะแสดงขึ้นมาอยู่ในส่วนที่ 2 ส่วนวิธีการอัพเดทจะประกอบด้วยการอัพเดท (Update) และ สร้างใหม่ (Create (Error if exists)) ซึ่งการสร้างใหม่จะไม่สามารถทำได้ในกรณีที่มีไฟล์ชื่อเดียวกัน

ส่วนที่ 2 ใช้ในการกำหนดการบันทึกหัวกอลัมน์และการปรับขนาดของกอลัมน์ตามขนาดของข้อมูล

ส่วนที่ 3 ใช้ในการจับกู่ Link ของ DataStage® กับ Sheet ของ Excel

#### 4.2.5 Sequential File Stage (Read)

		1
J	DS	Link2

รูปที่ 4. 41 รูป Sequential File Stage (Output)

Stage นี้ ใช้สำหรับการอ่านไฟล์ที่เป็นข้อความหรือไฟล์ txt ดังรูปที่ 4.1 ซึ่งการใช้งานนั้น หลัก ๆ แล้วจะต้อง กำหนดตำแหน่งของไฟล์, table definition ซึ่ง Stage นี้ยังรองรับการอ่านชื่อคอลัมน์ ดังรูปที่ 4.2 ในกรณีที่ ข้อมูลแถวแรกเป็นชื่อคอลัมน์ ซึ่งทำได้โดยแก้ First Line is Column Names = True

D	Sequential_File	e_1 - Sequential File	1- <b>-</b> ×	
Stage Ou	utput			
Output nam	ne:	C	olumns View Data	
DSLink2				
General	Properties Format   Columns   Advanced	- [No property selected]		
	Source File = D:\lnsuranceDemo\to_db2			
	Read Method = Specific File(s)     Options	+		
	<ul> <li>First Line is Column Names = False</li> <li>Keep File Partitions = False</li> </ul>	Information:		
	<ul> <li>Missing File Mode = Depends</li> <li>Reject Mode = Continue</li> </ul>			
	Report Progress = Yes		- \ \ \	
		Available properties to add:		
		Stile File		
		ОК	Cancel Help	
	รงไขี่ 4. 42 Output Propa	tios 201 Sequential File Sta		
	an 4. 42 Output Flope	thes our sequential file sta	1go	

#### 4.2.6 Sequential File Stage (Write)



63

รูปที่ 4. 43 รูป Sequential File Stage (Input)

รูปที่ 4.43 คือตัวอย่างของการจัดวาง Sequential File Stage เพื่อใช้ในรูปแบบการบันทึกไฟล์ ซึ่งมีหน้าต่าง การตั้งค่าดังรูปที่ 4.44 และผลลัพธ์ดังรูปที่ 4.45

File ระบุตำแหน่งของไฟล์ ชื่อไฟล์ และประเภทของไฟล์

File Update Mode [Append, Create (Error if exists), Overwrite] กำหนดวิธีการจัดการไฟล์ ได้แก่ เขียนต่อ จากของเดิม สร้างใหม่ซึ่งไม่สามารถสร้างได้หากมีไฟล์ชื่อซ้ำกัน และเขียนทับของเก่าตามลำคับ

Write Method [Specific File(s), Generate Multiple Files] กำหนดวิธีการบันทึกโดยจะแบ่งเป็น บันทึกข้อมูล บนไฟล์ที่เจาะจง และสร้างไฟล์ขึ้นมาหลายๆไฟล์ ในกรณีที่ต้องการแยกข้อมูลไปในแต่ละไฟล์

Cleanup On Failure [True, False] ในกรณีที่ตั้งค่าเป็น True ระบบจะทำการลบค่าที่บันทึกไปแล้วบางส่วน ในกรณีที่เกิดการทำงานผิดพลาด ส่งผลให้ไฟล์นั้นมีลักษณะเหมือนก่อนเริ่มทำการบันทึกข้อมูล

First Line is Column Names [True, False] ระบุว่าต้องการให้บันทึกชื่อคอลัมน์เป็นข้อมูลแถวแรกหรือไม่

Reject Mode [Continue, Fail, Output] ในกรณีที่เจอข้อมูลที่ไม่สามารถจัดเก็บได้ เช่น ข้อมูลไม่ตรงกับ table definition ที่ได้ระบุไว้ ก็สามารถเ<mark>ลือกวิ</mark>ธีรับมือได้สามวิธี ได้แก่ ไม่บันทึกข้อมูลส่วนที่มีปัญหา ยกเลิกการ ทำงาน และส่งต่อข้อมูลที่มีปัญหาไปให้ Stage ถัดไป

C	Sequential_File_9 - Sequential File	
1	Stage     Input       Input name:     Columns       DSLink8	View Data
	General       Properties       Partitioning       Format       Columns       Advanced         Image:       Target       Image:       Im	
	Available properties to add:	
•	OK Cancel รูปที่ 4. 44 Input Properties ของ Sequential File Stage	Help
THE		
	WSTITUTE OF TECH	

	to_db2 - Notepad	
File Edit Format View Help		
"2011","1","JAN","201101"		<u>^</u>
"2011","2","FEB","201102"		
"2011", "3", "MAR", "201103"		
"2011", "4", "APK", "201104"		
"2011", 5 , MAT , 201105		=
"2011", "7", "JUL", "201107"		
"2011","8","AUG","201108"		
"2011", "9", "SEP", "201109"		
"2011","10","OCT","201110	"	
"2011","11","NOV","201111		
"2011","12","DEC","201112	'_ \ II   a a .	
"2012", "1", "JAN", "201201"		
"2012", 2 , FED , 201202 "2012" "3" "MAR" "201202"		
"2012", "4", "APR", "201203		
"2012", "5", "MAY", "201205"		61
"2012","6","JUN","201206"		
"2012","7","JUL","201207"		
"2012","8","AUG","201208"		
"2012","9","SEP","201209"		
"2012", "10", "0(1", "201210 "2012", "11", "NOV", "201210		
2012, 11, NOV, 201211		<u> </u>
<b>`</b>		
		i i i i i i i i i i i i i i i i i i i
1/2		
	NSTITUTE OF	

#### 4.2.7 Transformer Stage



# **รูปที่ 4. 46** รูป Transformer Stage

Stage ดังรูปที่ 4.46 นี้ มีเพื่อการแปรรูปข้อมูลโดยการใช้ expression ต่าง ๆ รวมไปถึงการดัดแปลงแก้ไข กอลัมน์ ซึ่งการลบคอลัมน์ใน Stage อื่น อาจจะทำให้เกิดคอลัมน์ที่มีก่า null ทั้งหมด แต่ในการใช้ transformer stage จะไม่มีปัญหาในส่วนนั้น นอกจากนี้ ในการประมวลผลของบาง stage ตัวอย่างเช่น Pivot Enterprise Stage จะมีข้อมูลเป็น array เกินมาจากข้อมูลส่วนที่ต้องการ ก็ต้องใช้ transformer stage ในการตัดข้อมูลใน ส่วนนั้นออก มิฉะนั้นจะไม่สามารถบันทึกข้อมูลลงฐานข้อมูล DB2 ได้ ซึ่งมีหน้าต่างการใช้งานดังรูปที่ 4.47

	BASIC_Transformer_3 - Transformer Stage	e 🗕 🗖 🗙
🔗 📴 🕹 📾 🛤 🕞 🛲 🚺 📾 💷		
▼ DSLink2 ▲	Stage Variables	
CHANNEL_L4_CD Derivation	Stage Variable	
CHANNEL_L4_NM		
CHANNEL L3 CD		
CHANNEL 13 NM		
CHANNEL 12 CD		
	DSI ink4	
CHANNEL L1 CD		
	Columo Name	
DSURZ.CHANNEL_L4_NM	CHANNEL_L4_NM	
DSURKZ.CHANNEL_L3_CD	CHANNEL_L3_CD	
DSLink2.CHANNEL_L3_NM	CHANNEL_L3_NM	
DSLink2.CHANNEL_L2_CD	CHANNEL_L2_CD	
DSLink2.CHANNEL_L2_NM	CHANNEL_L2_NM	
DSLink2.CHANNEL_L1_CD	CHANNEL_L1_CD	
DSLink2.CHANNEL L1 NM	CHANNEL L1 NM	
<		
iLink2 DSLink4		
Column name Key SQL type Length Scale Nullable Column	name Key SQL type Length Scale Nullable Display	Data element Description 2
CHANNEL_L4_CI VarChar 10 Yes 1 CHANNE	EL_L4_CI VarChar 10 Yes	Value of CHANNEL_L4_CD
CHANNEL_L4_N VarChar 255 Yes 2 CHANNE	L_L4_N U VarChar 255 Yes	Value of CHANNEL_L4_NM
CHANNEL L3_CI L1 VarChar 10 Yes 3 CHANNEL	L_L3_CI VarChar 10 Yes	Value of CHANNEL_L3_CD
CHANNEL L2_N VarChar 200 Tes 4 CHANNEL CHANNEL L2_CL VarChar 10 Yes 5 CHANNE	L_L3_N VarChar 200 Tes	Value of CHANNEL L2_DM
CHANNEL L2 N VarChar 255 Yes 6 CHANNE	L L2 N VarChar 255 Yes	Value of CHANNEL L2 NM
CHANNEL L1 CI VarChar 10 No 7 CHANNE	L L1 CI 🗹 VarChar 10 No	Value of CHANNEL L1 CD
CHANNEL_L1_N VarChar 255 Yes 8 CHANNE	EL_L1_N VarChar 255 Yes	Value of CHANNEL_L1_NM
CH_WTD Decimal 5 4 Yes 9 CH_WTD	D Decimal 5 4 Yes	Value of CH_WTD
YEAR CD Integer 4 Yes 10 YEAR C	D Integer 4 Yes	Value of YEAR CD

รูปที่ 4. 47 ส่วนประกอบต่าง ๆ ของ Transformer Stage

ส่วนที่ 1 เป็นการแปรรูปข้อมูล โดยสามารถเลือกข้อมูลจากฝั่ง Input ทางซ้ายมือ มาใส่ในช่อง Derivation ทาง ขวามือเพื่อกำหนดข้อมูลที่จะถูกส่งออกจาก Transformer Stage ซึ่งในช่องนี้สามารถที่จะระบุ expression ต่าง ๆ เพื่อแปรรูปข้อมูลได้ นอกจากนี้ Transformer Stage ยังมี Stage Variable ที่สามารถกำหนดและนำมาใช้ได้ หลายครั้ง

ส่วนที่ 2 คือรูปแบบและขนาดของข้อมูลที่ทำการรับเข้ามาซึ่งอยู่ทางฝั่งซ้าย และข้อมูลที่ส่งออกซึ่งอยู่ทางฝั่ง ขวา การตั้งก่าสองฝั่งนั้นไม่จำเป็นต้องตรงกันหรือมีจำนวนเท่ากัน

4.2.8 Sort Stage



**รูปที่ 4. 48** รูป Sort Stage

Sort Stage ในรูปที่ 4.48 ใช้เพื่อจัดเรียงข้อมูล โดยที่ผู้ใช้สามารถกำหนด Key หรือก็คือคอลัมน์ที่ต้องการใช้ ในการเรียงได้ ซึ่งสามารถที่จะกำหนดหลาย ๆ Key เพื่อให้เรียงลำดับได้หลายชั้น ซึ่งการเรียงลำดับนั้น สามารถเลือกได้ทั้งแบบ จากน้อยไปมาก หรือจากมากไปน้อยได้ นอกจากนี้ ยังสามารถกำหนดได้ว่า Key ที่ ใช้ในการเรียงลำดับ สามารถมีค่าซ้ำกันได้หรือไม่ (Allow Duplicate) ดังรูปที่ 4.49 ซึ่งการตั้งค่าเป็น False จะ มีตัวเลือกเพิ่มเติมว่าจะทำการเก็บค่าแรกสุดหรือก่าล่าสุดเอาไว้ แต่อาจจะพบปัญหาคือค่าตั้งต้นของ Sort Stage มี Execution mode เป็น Parallel ซึ่งทำให้ผลลัพธ์ที่ได้อาจจะผิดเพี้ยนไปจากที่ต้องการ ซึ่งแก้ไขได้โดย การปรับเป็น Sequential และปรับการตั้งค่า Preserve partitioning ให้เป็น Clear ดังรูปที่ 4.50



#### 4.2.9 Remove Duplicates Stage



รูปที่ 4. 51 รูป Remove Duplicates Stage

Remove Duplicates Stage ดังรูปที่ 4.51 ใช้สำหรับลบ record ที่มีข้อมูลซ้ำกัน ซึ่งผู้ใช้เพียงแค่ระบุคอลัมน์ที่ ไม่ต้องการให้มีข้อมูลซ้ำกัน ดังรูปที่ 4.52 โดนการเลือกคอลัมน์ใน Keys that define Duplicates > Key และ กำหนดว่าต้องการเก็บ record อันแรกสุดหรืออันล่าสุด โดยการตั้งค่าที่ Option > Duplicate To Retain [First, Last]

	Remove_Duplicates_5 - Remove Duplicates
	Stage Input Output
	Stage name:
	Remove_Duplicates_5
	General Properties Advanced NLS Locale
17-1	- Keys that define Duplicates [No property selected]
	Key = MONTH_CD
	Continue - Continue - Cont
	Duplicate to Retain = nist     Information:
	Available properties to adu.
1	Si Key
	OK Cancel Help
	รูปที่ 4. 52 Stage Properties ของ Remove Duplicates Stage
	VSTITUTE OF

#### 4.2.10 Aggregator Stage



รูปที่ 4. 53 รูป Aggregator Stage

Aggregator Stage ดังรูปที่ 4.53 ใช้เพื่อหาผลรวมของค่าที่ต้องการของกลุ่มต่าง ๆ ซึ่งต้องกำหนดค่าต่าง ๆ ดัง รูปที่ 4.54 ดังนี้

Grouping Keys > Group กำหนดคอลัมน์ที่ระบุกลุ่ม ที่ต้องการหาผลรวม

Aggregation > Column for Calculation กำหนดคอลัมน์ที่ใช้ในการคำนวน

Aggregation > Column for Calculation > Sum Output Column กำหนดคอลัมน์ที่ต้องการใส่ผลลัพธ์ที่ได้

**Option > Allow Null Output** [True, False] กำหนดว่าผลลัพธ์สามารถเป็นค่า null ได้หรือไม่ ถ้าไม่ ก็จะทำ การแทนค่า null ด้วย 0

P	
Stage Input Output	
Stage name: Aggregator 9	
General Properties Advanced NLS Locale	
Grouping Keys Group = CHANNEL_L4_CD Group = CHANNEL_L4_NM Aggregations Aggregation Type = Calculation Column for Calculation = CH_WTD Sum Output Column = CH_WTD	
<ul> <li>☐ Options</li> <li>Allow Null Output = False</li> <li>Method = Hash</li> </ul>	

รูปที่ 4. 54 Stage Properties ของ Aggregator Stage

เมื่อกำหนดวิธีการหาผลรวมเสร็จแล้ว ใน Tab Output > Mapping จะมีการสร้าง Expression ให้โดยอัตโนมัติ ดังรูปที่ 4.55

Output name: DSLink24		
General Mapping Columns Advanced		
	Columns	
Expression		Column Name
DSLink21.CHANNEL_L4_CD		CHANNEL_L4_CD
DSLink21.CHANNEL_L4_NM		CHANNEL_L4_NM
Sum(DSLink21.CH_WTD)		CH_WTD

รูปที่ 4. 55 Output Mapping ของ Aggregator Stage

TC

#### 4.2.11 Checksum Stage



## รูปที่ 4. 56 รูป Checksum Stage

Checksum Stage ดังรูปที่ 4.56 จะทำการสร้างค่า Checksum โดยใช้ค่าภายในกอลัมน์ที่กำหนดมาคำนวณ ซึ่ง สามารถเลือกได้ว่าจะนำคอลัมน์ทั้งหมดมาคำนวณ (Use all columns) นำคอลัมน์ทุกคอลัมน์ยกเว้นคอลัมน์ที่ กำหนดมาคำนวณ (Use all columns except those specified) หรือนำมาคำนวณเฉพาะคอลัมน์ที่กำหนด (Use only columns specified) ดังรูปที่ 4.57

Stage Input Output Stage name: Checksum_12 General Properties Adv	Checksum	<ul> <li>Computation Mode</li> <li>Computation Mode</li> <li>Use all columns</li> <li>Use all columns exults all columns exults all columns exits</li> <li>Type: List</li> <li>Determines whether will be used to generate</li> <li>Available properties</li> </ul>	cept those specified specified er all input columns or a su erate the checksum.	□ × ubset ↓	
	รูปที่ <b>4. 57</b> Stage Prop	ок perties ปอง Checksum UTE O	Cancel Stage	Help	

เมื่อตั้งค่าเสร็จเรียบร้อยแล้ว ก็จะมีคอลัมน์ checksum เพิ่มขึ้นมาสำหรับ Mapping ดังรูปที่ 4.58

Stage   Ir	nput Output		
Output na	me:		
DSLink1	5 🔹		
General	Mapping Columns	Advanced	
	Colum	ins	
	Expression	Column Na	
	DSLink9.NAME	NAME	
	DSLink9.SALARY	SALARY	
G	Checksum()	checksum	
<b>•</b>			
	Output Monning	912 9 Chaolenne	. Star

# ซึ่งเมื่อคำนวณแล้วก็จะมีผลลัพธ์ดังรูปที่ 4.59

	NAME	SALARY	CHECKSUM
		SALART	
<u>.</u>		8	51f90eac0e9081dcce8837c982035540
2		20	15f8563068ec215f116a6c237bb20bcc
3		44	b6da2502797d77e885b97597a1be3d57
4	******	58	292dd2baf36fcff5546068ff7c0bfe0e
5	ΥΥΥΥΥΥΥΥ	60	4d5b14a598113b7c3ebf4a51bdb4f2e9
6	bbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbb	64	c38f9cdc27ed4ff46a534ac534982e54
7	33333333333333333333333333333333333	92	e15d23896f365514efc850d615b10f48
8	222222222222222222222222222222222222222	25	5a3808ad4b3ee4fd6a5de24fef149e61
9	DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD	39	e5a03a462d6b6b9b7e9da14cf8f2f024
10	LULL	47	7951ebbf6d654129b14f2d304ba626de
11	XXXXXXXXXXXXXXXXXXXXXX	59	ca5bca0b3aba74c7d9fa1b24e6085295
12	aaaaaaaaa	63	01f20795e9d9d4fb33c6e525cb6058a0
13	cccccccccccccccccccccccc	65	5e254007a8e07693bd38e7d9e1b97d3a
14	44444444444	93	fc829a538b8d53f7b39084a81ec2a6eb

รูปที่ 4. 59 ตัวอย่างผ<sub>ื</sub>ลลัพธ์ของ Checksum Stage

#### 4.2.12 Row Generator Stage



รูปที่ 4. 60 รูป Row Generator Stage

Row Generator Stage คังรูปที่ 4.60 เป็น Stage ที่มีไว้สร้าง Record ที่สุ่มข้อมูลขึ้นมา ซึ่ง Stage นี้มีการตั้งก่า หลัก ๆ คือ จำนวน Record ที่ต้องการ คังรูปที่ 4.61 – 4.62



	NAME	SALARY		CHECKSUM
1	111111111111	8		51f90eac0e9681dcce8837c982635546
2	սսսսսսսսսսսսսսսս	20		15f8563068ec215f116a6c237bb20bcc
3		44		b6da2502797d77e885b97597a1be3d57
4	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	58		292dd2baf36fcff5546068ff7c0bfe0e
5	ΥΥΥΥΥΥΥΥ	60		4d5b14a598113b7c3ebf4a51bdb4f2e9
6	bbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbb	64		c38f9cdc27ed4ff46a534ac534982e54
7	333333333333333333333333333333333333333	92		e15d23896f365514efc850d615b10f48
8	222222222222222222222222222222222222222	25		5a3808ad4b3ee4fd6a5de24fef149e61
9	DDDDDDDDDDDDDDDDDDDDDDDDDDDDD	39		e5a03a462d6b6b9b7e9da14cf8f2f024
10	IIII	47		7951ebbf6d654129b14f2d304ba626de
11	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	59		ca5bca0b3aba74c7d9fa1b24e6085295
12	аааааааааа	63		01f20795e9d9d4fb33c6e525cb6058a0
13		65		5e254007a8e07693bd38e7d9e1b97d3a
14	44444444444	93		fc829a538b8d53f7b39084a81ec2a6eb

รูปที่ 4. 62 ตัวอย่างการใช้ Row Generator Stage ในการ Generate ข้อมูลในคอลัมน์ Name และ Salary

### 4.2.13 Peek Stage



รูปที่ 4. 63 รูป Peek Stage

## TC

Peek Stage ดังรูปที่ 4.63 มีความสามารถในการดูข้อมูลต่าง ๆ ได้ ซึ่งสามารถวาง Peek Stage ไว้ระหว่าง Stage ที่ต้องการดูข้อมูล ซึ่งข้อดีของ Peek Stage ก็คือ สามารถดูข้อมูลได้ เมื่อมีการทำงานเสร็จแล้ว แม้ว่า Parallel Job จะล้มเหลวก็ตาม ซึ่ง Peek Stage จะมีวิธีแสดงผลลัพธ์สองวิธี ดังรูปที่ 4.64 คือ Job Log ดังรูปที่ 4.65 – 4.66 และ Output ดังรูปที่ 4.67 – 4.68 ในกรณีที่มีการทำงานแบบ Parallel Peek Stage จะมีการแบ่งปริมาณ ออกตามจำนวน Partition ที่มี

Rows > Number of Records (Per Partition) กำห<mark>น</mark>ดจำนวน reco</mark>rd ที่ให้แสดงผลต่อ 1 Partition

Columns > Peek All Input Columns [True, False] แสดงผลคอลัมน์ทั้งหมดหรือไม่

Partitions > All Partitions [True, False] แสดงผลทุก Partition หรือไม่

**Options > Peek Records Output Mode** [Job Log, Output] เลือกรูปแบบของ Output



ตัวอย่าง output แบบ Output

T



รูปที่ 4. 67 ตัวอย่างการวาง Peek Stage แบบ Output

	View Data	-
REC		_
NAME:bbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbb		Ì
NAME:dddddddddddddddddddddddddddddddd		
NAME:		
NAME:hhhhhhhhhhhhhhhhh SALARY:7		1
NAME:		i.
NAME:		
NAME:nnnn SALARY:13		
NAME:pppppppppppppppppppppppppppppppppppp		
NAME:mmmmr SALARY:17		
NAME: MINIMUM MINIMUM SALARY: 19		
NAME:зааааааааааааааааааааааааааааааааа		
NAME:cccccccccccccccccccccccccccccccccccc		
NAME:eeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeee		
NAME:gggggggggggggggggggggggggggggggggggg		
NAME:		
NAME:kkkkk SALARY:10		
NAME:mmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm		
NAME:000000000000000000000000000000000000		

ร**ูปที่ 4. 68** ตัวอย่างข้อมูลจาก Database ที่บันทึกผลลัพธ์จาก Peek Stage แบบ Output

4.2.14 Sample Stage



รูปที่ **4. 69** รูป Sample Stage

Stage ในรูปที่ 4.69 – 4.71 ใช้สำหรับการเลือกหรือสุ่มข้อมูลบางส่วน เพื่อนำไปทดสอบการทำงานของระบบ ต่อไป



รูปที่ 4. 70 ตัวอย่างการใช้งาน Sample Stage

## G

Options > Percent ใช้กำหนดปริมาณข้อมูลที่ต้องการสุ่มขึ้นมาเป็นตัวอย่าง (การสุ่มของ DataStage® มาจาก การคำนวณ ทำให้การสุ่มในแต่ละครั้ง ที่มีการตั้งค่าทุกอย่างเหมือนกัน จะได้ผลลัพธ์เหมือนกัน และการ กำหนดเป็น Percent ส่วนมาก มักจะไม่ได้จำนวน record ตรงกับการคำนวณตามปกติ เช่น ได้ 15 record จาก 25 Percent ของ 100 Record)

Options > Percent > Output Lin<mark>k Number</mark> ระบุ Link ที่จะถำเลียงข้อมูล ซึ่งหมายเลขของ Link สามารถดู ได้ที่ Tab Link Ordering Options > Sample Mode [Percent, Period] เป็นวิธีการเลือกสุ่มข้อมูล ซึ่งได้แก่ Percent ที่ได้อธิบายไว้ ข้างต้น และ Period ที่เป็นการกำหนดให้เลือกตัวอย่างมาทุก ๆ n แถว

Stage Input Output	
Stage name:	
Sample_3	
General Properties Advanced Link Ordering	
⊡- <sup>©</sup> Options	[No property selected]
Gutout Link Number = 0	
Sample Mode = Percent	Laforna Company
	Information:



4.2.15 Filter Stage



รูปที่ 4. 72 รูป Filter Stage

Stage ในรูปที่ 4.72 ใช้สำหรับการกรองข้อมูลตามเงื่อนไขที่ผู้ใช้กำหนด โดยมีการตั้งก่าต่าง ๆ ดังรูปที่ 4.73 – 4.74

Predicates > Where Clause กำห<mark>นด W</mark>here Clause Statement ที่ผู้ใช้ต้องการ<mark>ถ</mark>งไป

Predicates > Where Clause > Output Link กำหนด Link ในการส่งต่อข้อมูลที่ผ่านเงื่อนไข Where Clause ที่ Output Link เป็นสมาชิก Output Link สามารถดูได้ใน Link Ordering

Options > Output Rejects [True, False] สร้าง Link หรือไม่ สำหรับข้อมูลที่ขัดแย้งกับ table definition

Options > Output Row Only Once [True, False] กรณีที่ตั้งเป็น True record ที่ผ่านเงื่อนไขใดเงื่อนไขหนึ่ง แล้ว จะไม่ถูกนำไปประมวลผลกับ Where Clause อื่นซ้ำ ในขณะที่ ถ้าตั้งเป็น False จะนำ record ไป ประมวลผลกับทุก Where Clause ซึ่งถ้าถูกหลาย Where Clause ก็จะทำให้มี record ซ้ำกันได้



STITUTE O

4.2.16 Funnel Stage

10



รูปที่ 4. 75 รูป Funnel Stage

Stage ในรูปที่ 4.75 ใช้สำหรับการรวมแถวเข้าด้วยกัน โดยรูปที่ 4.76 จะเห็นได้ว่า มี Input Link เข้ามาใน Funnel Stage จำนวน 2 Link ซึ่งประกอบไปด้วยข้อมูลจำนวน 5 แถว และข้อมูลจำนวน 10 แถว และมี 1 Output Link ที่มีข้อมูลจำนวน 15 แถว



รูปที่ 4. 76 ตัวอย่างการใช้ Funnel Stage

โดยสามารถกำหนดวิธีการรวมข้อมูลได้ 3 วิธี ได้แก่ Continuous Funnel, Sequence, และ Sort Funnel ดังรูป ที่ 4.77

Stage name: Funnel_15	
General Properties Advanced Link Ordering NLS Locale	
<ul> <li>□· Coptions</li> <li>○ Funnel Type = Continuous Funnel</li> <li>○</li> <li>○<!--</td--><td>Funnel Type         Continuous Funnel         Sequence         Sort Funnel         Type: List         Continuous Funnel combines records as they arrive (i.e. no particular order); Sort Funnel combines the input records in the order defined</td></li></ul>	Funnel Type         Continuous Funnel         Sequence         Sort Funnel         Type: List         Continuous Funnel combines records as they arrive (i.e. no particular order); Sort Funnel combines the input records in the order defined

รูปที่ 4. 77 Stage Properties ของ Funnel Stage

ในส่วนของการกำหนด Link Ordering ดังรูปที่ 4.78 จะไม่มีผลกับการใช้ Sort Funnel เนื่องจาก Stage จะทำ การอ่านข้อมูลทั้งหมด เพื่อจัดเรียง ทำให้ลำดับของข้อมูลที่เข้ามานั้น ไม่ส่งผลต่อผลลัพธ์ แต่การใช้ Sort Funnel นั้น จะมีลักษณะเหมือนการใช้ Sort Stage ดังนั้น ผู้ใช้ควรปรับรูปแบบการประมวลผลให้เป็นแบบ Sequential เสียก่อน

Stage Input Output	
Stage name:	
Funnel_15	
General Properties Advanced Link Ordering	NLS Locale
Order the following input links:	0
Link label Link name	Ŧ
20 Cause4_Y2015	
1 Policy_N	-
	+

รูปที่ 4. 78 Stage Link Ordering ของ Funnel Stage

ต่อไปนี้จะเป็นตัวอย่างผลลัพธ์ที่เกิดจากการใช้ Funnel Stage รูปแบบต่าง ๆ ซึ่งให้สังเกตที่ลำดับการเรียง ของ กอลัมน์ CLIENT\_CD เป็นหลัก

รูปที่ 4.79 เป็นผลลัพธ์จากการใช้ Continuous Funnel ซึ่งมีการเรียงลำดับข้อมูลแบบ Record ใดที่ประมวลผล เสร็จก่อน ก็จะถูกบันทึกก่อน

		View Data	_ <b>□</b> X	
РК	CLIENT_CD	INSURE_CLIENT	POLICY_STS	CAUSE_TYPE_CO
201545RN	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	RN	4
2015411RN	11	Ratchaburi Electricity Generating Holding Public Co	. RN	4
2015413RN	13	Thai Agri Foods Public Company Limited	RN	4
2015415RN	15	Siam Steel International Plc.	RN	4
201111N	1	Sony Technology Thailand Co., Ltd.	N	1
201153N	3	Global Food Products Co., Ltd.	N	5
201145N	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	N	4
201157N	7	SIAM TAKASHIMAYA (THAILAND) CO., LTD.	Ν	5
201159N	9	Manulife Asset Management (Thailand)	Ν	5
2015418RN	18	TCC Land Co., Ltd.	RN	4
201152N	2	Jebsen & Jessen Group of Companies	Ν	5
201114N	4	Anek Ekburus & Alliance Co., Ltd.	N	1
201126N	6	Bosch Packaging Technology (Thailand) Co., Ltd.	N	2
201158N	8	Amata Corporation Public Co., Ltd.	N	5
2011110N	10	Honda Leasing (Thailand) Co., Ltd.	Ν	1
<		Ш		>
				Close

รูปที่ 4. 79 ผลลัพธ์จากการใช้ Continuous Funnel

รูปที่ 4.80 เป็นผลลัพธ์จากการใช้ Sequence ซึ่งมีการเรียงลำดับข้อมูล โดยเรียงตามลำดับของ Link Ordering ตัวอย่างเช่น จะทำการรับข้อมูลจาก Link Ordering ลำดับที่ 0 จนครบทั้งหมดก่อน แล้วบันทึก จากนั้นจึงอ่าน ข้อมูลจาก Link Ordering ล<mark>ำดับที่ 1 จน</mark>ครบ แล้วจึ<mark>งนำข้อมู</mark>ลไปต่<mark>อ</mark>ท้ายข้อ<mark>มูลเด</mark>ิม

10

		View Data		_ 🗆 X
PK	CLIENT_CD	INSURE_CLIENT	POLICY_STS	CAUSE_TYPE_CD
201545RN	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	RN	4
2015411RN	11	Ratchaburi Electricity Generating Holding Public Co.	RN	4
2015413RN	13	Thai Agri Foods Public Company Limited	RN	4
2015415RN	15	Siam Steel International PIc.	RN	4
2015418RN	18	TCC Land Co., Ltd.	RN	4
201111N	1	Sony Technology Thailand Co., Ltd.	N	1
201153N	3	Global Food Products Co., Ltd.	N	5
201145N	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	N	4
201157N	7	SIAM TAKASHIMAYA (THAILAND) CO., LTD.	N	5
201159N	9	Manulife Asset Management (Thailand)	N	5
201152N	2	Jebsen & Jessen Group of Companies	N	5
201114N	4	Anek Ekburus & Alliance Co., Ltd.	N	1
201126N	6	Bosch Packaging Technology (Thailand) Co., Ltd.	N	2
201158N	8	Amata Corporation Public Co., Ltd.	N	5
2011110N	10	Honda Leasing (Thailand) Co., Ltd.	N	1
<				>
	6	10.05	11	Close
				0.036

รูปที่ 4. 80 ผลลัพธ์จากการใช้ Sequence Funnel

รูปที่ 4.81 เป็นผลลัพธ์จากการใช้ Sort Funnel ซึ่งทำการอ่านข้อมูลทั้งหมดก่อน แล้วจึงทำการจัดเรียงตาม กอลัมน์ที่กำหนดให้เป็น Key แล้วจึงทำการบันทึกข้อมูล

TC

3	×			
PK	CLIENT_CD	INSURE_CLIENT	POLICY_STS	CAUSE_TYPE_C
201111N	1	Sony Technology Thailand Co., Ltd.	N	1
201152N	2	Jebsen & Jessen Group of Companies	Ν	5
201153N	3	Global Food Products Co., Ltd.	N	5
201114N	4	Anek Ekburus & Alliance Co., Ltd.	Ν	1
201545RN	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	RN	4
201145N	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	N	4
201126N	6	Bosch Packaging Technology (Thailand) Co., Ltd.	Ν	2
201157N	7	SIAM TAKASHIMAYA (THAILAND) CO., LTD.	N	5
201158N	8	Amata Corporation Public Co., Ltd.	N	5
201159N	9	Manulife Asset Management (Thailand)	N	5
2011110N	10	Honda Leasing (Thailand) Co., Ltd.	N	1
2015411RN	11	Ratchaburi Electricity Generating Holding Public Co	RN	4
2015413RN	13	Thai Agri Foods Public Company Limited	RN	4
2015415RN	15	Siam Steel International Plc.	RN	4
2015418RN	18	TCC Land Co., Ltd.	RN	4
<		and the second		>
				Close

รูปที่ 4. 81 ผลลัพธ์จากการใช้ Sort Funnel

#### 4.2.17 Lookup Stage



รูปที่ 4. 82 รูป Lookup Stage

รูปที่ 4.82 ใช้สำหรับการ Lookup ข้อมูลที่อยู่ต่าง Table เพื่อนำมาประกอบกับข้อมูลเก่า เพื่อช่วยให้ข้อมูลมี กวามสมบูรณ์มากขึ้น ตัวอย่างเช่นรูปที่ 4.83 ที่นำ CAUSE\_TYPE\_C จาก DSLink22 มาจับคู่กับ CAUSE\_TYPE\_C จาก DSLink23 แล้วจึงนำข้อมูลจากคอลัมน์ต่าง ๆ จาก DSLink22 มาแสดง และนำ กอลัมน์ CAUSE\_TYPE\_N ซึ่งเป็นชื่อของสาเหตุมาแสดงแทน CAUSE\_TYPE\_C ซึ่งเป็นรหัสตัวเลขของ สาเหตุ โดย Lookup Stage แตกต่างจากการใช้ Join Stage และ Merge Stage ตรงที่ Lookup Stage นั้นจะเป็น การนำ Reference Table เก็บไว้บน RAM และไม่มีการ Sort ข้อมูลก่อนเริ่มทำการจับคู่ทำให้มีความเร็วในการ ประมวลผลที่สูง แต่ในกรณีที่มีข้อมูลภายใน Reference Table ปริมาณมาก จะส่งผลให้ประสิทธิภาพด้อยกว่า Join Stage และ Merge Stage

							LOOKup_1	coonap bu	age				
7 🔗   BO   % 🖻		1   M   🖃 🕯						-					
·													
		-		DSLink22				-	DSLir	1k24			
		Key Expr	ession	Ban	ae Colum	n Nam		Derivatio	0	Column Name			
					РК			DSLink22	PK	РК			
					CLIENT	CD 1		DSLink 22	INSURE CLIENT				
					INCLID			DSLink22	POLICY STS				
					POLICY		_	DSLink22	VEAR CD	YEAR CD			
					FULIC	_313	-r /	DOLINK22.	CALLOS THE A	TEAR_CD			
					CAUSE	_TYPE	/	DSLink23.	CAUSE_TYPE_N	IM CAUSE_TYPE_			
					YEAR_	CD							
				_									
		•		DSLink23									
		Condition:										1 14	
		Condition: Key Express	sion	Кеу	Typ Colum	n Nam						()_	
		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum	n Nam						0	
		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum	n Nam E_DYF						$\mathcal{O}$	
		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum CAUSE	n Nam E_TYF _TYPE						S	
		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum CAUS CAUSE	n Nam E_TYF _TYPE						$\mathbf{O}$	
4		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum CAUSE	n Nam E_TYF _TYPE						ပ္ပ	
		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum CAUSE	n Nam E_TYPE _TYPE						$\hat{\mathcal{O}}$	
: <		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	C =	Typ Colum CAUSE	n Nam E_TYF _TYPE							
J CSLink22 DSLink23		Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum CAUS CAUSE	n Nam E_TYF _TYPE					à	0	
SLink22 DSLink23	Key SQLtype	Condition: Key Express DSLink22.CA	sion USE_TYPE_C	D =	Typ Colum CAUSE CAUSE	n Nam E_TYF _TYPE	pe Exter	ided Length	Scale Nullable		, Ō.		
SLink22 DSLink23 Column name	Key SQL type	Condition: Key Express DSLnk22.CA	sion USE_TYPE_C Length Sca 32	DSLink24	Typ Colum CAUSE CAUSE	n Nam E_TYFE TYPE	pe Exter	ided Length 32	Scale Nullable No		, O	0 0 ' (	
SLink22 DSLink23 Column name 1 PK 2 CLIENT_CD	Key SQLtype ☑ VarChar □ Integer	Condition: Key Express DSLink22.CA	sion USE_TYPE_C Length Sce 32 5	DSLink24 Column n: 1 PK 2 INSURE_C	Typ Colum CAUSE CAUSE ame Key LIENT	Nam ETYPE TYPE VarChar VarChar	pe Exter	nded Length 32 100	Scale Nullable No No	Value of INSURE_CLIE	NT		
SURK22 DSLInk23 Column name 1 PK 2 CLIENT_CD 3 INSURE_CLIENT	Key SQL type Ø VaChar Integer VaChar	Condition: Key Express DSLInk22 CA	sion USE_TYPE_C Length Sca 32 5 100	Key           D         =           DSLink24	Typ Colum CAUSE CAUSE ame Key LIENT TS	r Nam E_TYPE TYPE VarChar VarChar VarChar	pe Exter	ided Length 32 100 20	Scale Nullable No No No	Value of INSURE_CLIE Value of POLICY_STS	TO NT	о С	
SUINK22 DSLINK23 Column name I PK 2 CLIENT_CD 3 INSURE_CLENT 4 POLICY_STS	Key SQL type V VaChar Integer VaChar VaChar	Condition: Key Express DSLink22 CA	Length Sce 32 5 100 20	DSLink24 DSLink24 Column n 1 PK 2 INSURE_C 3 POLCY_S1 4 YEAR_CD	Typ Colum CAUSE CAUSE ame Key LIENT S	SQL ty     VarChar     VarChar     VarChar     VarChar     VarChar     VarChar	pe Exter	ided Length 32 100 20 5	Scale Nullable No No No No	Value of INSURE_CLIE Value of POLICY_STS Value of YEAR_CD	NT	0 ) (	
SLInk22 DSLInk23 Column name CLENT_CD CLENT_CD SURE_CLENT POLICYSTS CAUSE_TYPE_C	Key SQLtype M VarChar Integer VarChar Integer	Condition: Key Express DSLink22 CA	Length Sca 32 5 100 20 2	DSLink24 Column n 1 PK 2 INSURE_C 3 POLICY SD 5 CAUSE_TY	Typ Colum CAUSE CAUSE ame Key LIENT C TS PE_N C	SQL ty SQL ty VarChar VarChar Integer VarChar	pe Exter	nded Length : 32 100 20 5 100	Scale Nullable No No Yes	Value of INSURE_CLIE Value of POLICY_STS Value of YEAR_CD Value of CAUSE_TYPE	NT		

รูปที่ 4. 83 หน้าต่างการใช้งาน Lookup Stage

#### 4.2.18 Join Stage



Stage ในรูปที่ 4.84 ใช้ในการ Join table 2 table เข้าด้วยกันซึ่งสามารถมีได้หลาย Input Link แต่มีได้เพียงหนึ่ง Output Link ซึ่งการประมวลผลในแต่ละครั้ง Join Stage จะทำการ Sort ข้อมูลแบบ Ascending ก่อนที่จะทำ การ Join ซึ่งรูปแบบการ Join ที่มีให้เลือกนั้น ได้แก่ Full Outer, Inner, Left Outer, Right Outer ซึ่งสิ่งที่ต้อง กำหนดนอกเหนือจากวิธีการ Join คือ Key หรือกอลัมน์ที่ใช้ในการ Join ซึ่งการจะระบุว่า Input Link ใด เป็น Left หรือ Right สามารถดูได้ที่ Link Ordering ดังรูปที่ 4.85 – 4.91

Stage     Input     Output       Stage name:	
□··        Join Keys         □··	Join Type Inner Full Outer Inner Left Outer Right Outer Type of join operation to perform.
รูปที <mark>่ 4. 85</mark> Stage Propert	ies VOJ Join Stage

Stage	Input	Output			
tage Input Output tage name: Join_30 General Properties Advanced Link Ordering N Order the following input links: Link label Link name Link label DSLink23					
Join_30					
Genera	il   Pro	perties Ad	vanced	Link Ordering	NLS
Order	the foll	owing input	links:		
Link	label	Link name			Ŧ
10	eft	DSLink23			
- 🖊 R	ight	DSLink22			

Loc

#

รูปที่ 4. 86 Stage Link Ordering ของ Join Stage

1	А	В
1	CAUSE_TYPE_CD	CAUSE_TYPE_NM
2	1	Short Circuit
3	2	Accident
4	3	Lightning
5	4	Fire
6	5	Water Damage

14	<b>U</b>	ν		A
~ a lan / 0'	7 ຕາລຍາງຍ	າເລາເລຄເລາ	L aff Table	(DOT : -1-22)
111114 8		11 PLAT (1 - 11 PLA	Len ranie	$\mathbf{U}$ <b>J N</b> $\mathbf{U}$ <b>J N</b> $\mathbf{K}$ / <b>N</b>
	/ 1100014	00000000		(D)DDIIII(D)
0		a		· · · · · · · · · · · · · · · · · · ·

# T

PK	CLIENT_CD	INSURE_CLIENT	POLICY_STS	CAUSE_TYPE_CD	YEAR_CD	
201545RN	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	RN	4	2015	
2015411RN	11	Ratchaburi Electricity Generating Holding Public Co	. RN	4	2015	
2015413RN	13	Thai Agri Foods Public Company Limited	RN	4	2015	
2015415RN	15	Siam Steel International Plc.	RN	4	2015	
201111N	1	Sony Technology Thailand Co., Ltd.	N	<b>-</b> 1	2011	
201153N	3	Global Food Products Co., Ltd.	Ν	5	2011	
201145N	5	PROCUREMENT CENTRE (THAILAND) CO., LTD.	Ν	4	2011	
201157N	7	SIAM TAKASHIMAYA (THAILAND) CO., LTD.	Ν	5	2011	
201159N	9	Manulif <mark>e Asset</mark> Management (Thailand)	Ν	5	2011	
2015418RN	18	TCC Land Co., Ltd.	RN	4	2015	
201152N	2	Jebsen & Jessen Group of Companies	Ν	5	2011	
201114N	4	Anek E <mark>kburus</mark> & Alliance Co., Ltd.	Ν	1	2011	
201126N	6	Bosch Packaging Technology (Thailand) Co., Ltd.		2	2011	
201158N	8	Amata Corporation Public Co., Ltd.	Ν	5	2011	
2011110N	10	Honda Leasing (Thailand) Co., Ltd.	Ν	1	2011	

ร**ูปที่ 4. 88** ตัวอย่างข้อมูลของ Right Table (DSLink22)

PK	INSURE_CLIENT	CAUSE_TYPE_NM	POLICY_STS	YEAR_CD
201111N	Sony Technology Thailand Co., Ltd.	Short Circuit	N	2011
201114N	Anek Ekburus & Alliance Co., Ltd.	Short Circuit	N	2011
2011110N	Honda Leasing (Thailand) Co., Ltd.	Short Circuit	N	2011
201126N	Bosch Packaging Technology (Thailand) Co., Ltd.	Accident	N	2011
		Lightning		0
201545RN	PROCUREMENT CENTRE (THAILAND) CO., LTD.	Fire	RN	2015
2015411RN	Ratchaburi Electricity Generating Holding Public Co	. Fire	RN	2015
2015413RN	Thai Agri Foods Public Company Limited	Fire	RN	2015
2015415RN	Siam Steel International Plc.	Fire	RN	2015
201145N	PROCUREMENT CENTRE (THAILAND) CO., LTD.	Fire	N	2011
2015418RN	TCC Land Co., Ltd.	Fire	RN	2015
201153N	Global Food Products Co., Ltd.	Water Damage	N	2011
201157N	SIAM TAKASHIMAYA (THAILAND) CO., LTD.	Water Damage	N	2011
201159N	Manulife Asset Management (Thailand)	Water Damage	N	2011
201152N	Jebsen & Jessen Group of Companies	Water Damage	N	2011
201158N	Amata Corporation Public Co., Ltd.	Water Damage	N	2011

# ร**ูปที่ 4. 89** ผลลัพธ์จากการใช้ Full Outer Join

		View Data									
PK	_	INSURE_CLIENT	CAUSE_TYPE_NM	POLICY_STS	YEAR CD						
201111N		Sony Technology Thailand Co., Ltd.	Short Circuit	N	2011						
201114N		Anek Ekburus & Alliance Co., Ltd.	Short Circuit	N	2011						
2011110N		Honda Leasing (Thailand) Co., Ltd.	Short Circuit	N	2011						
201126N		Bosch Packaging Technology (Thailand) Co., Ltd.	Accident	Ν	2011						
201145N		PROCUREMENT CENTRE (THAILAND) CO., LTD.	Fire	N	2011						
201545RN	4	PROCUREMENT CENTRE (THAILAND) CO., LTD.	Fire	RN	2015						
2015411R	(N	Ratchaburi Electricity Generating Holding Public Co	Fire	RN	2015						
2015413R	(N	Thai Agri Foods Pu <mark>blic Co</mark> mpany Limited	Fire	RN	2015						
2015415R	(N	Siam Steel International Plc.	Fire	RN	2015						
2015418R	(N	TCC Land Co., Ltd.	Fire	RN	2015						
201152N		Jebsen & Jessen G <mark>roup o</mark> f Companies	Water Damage	N	2011						
201153N		Global Food Products Co., Ltd.	Water Damage	N	2011						
201157N		SIAM TAKASHIMAYA (THAILAND) CO., LTD.	Water Damage	N	2011						
201158N	2	Amata Corporation Public Co., Ltd.	Water Damage	N	2011						
201159N	1/2	Manulife Asset Management (Thailand)	Water Damage	Ν	2011						
	1										

ร**ูปที่ 4. 90** ผลลัพธ์จากการใช้ Inner Join

S.

PK	INSURE_CLIENT	CAUSE_TYPE_NM	POLICY_STS	YEAR_CD
201111N	Sony Technology Thailand Co., Ltd.	Short Circuit	N	2011
201114N	Anek Ekburus & Alliance Co., Ltd.	Short Circuit	N	2011
2011110N	Honda Leasing (Thailand) Co., Ltd.	Short Circuit	Ν	2011
201126N	Bosch Packaging Technology (Thailand) Co., Ltd.	Accident	N	2011
		Lightning		0
201545RN	PROCUREMENT CENTRE (THAILAND) CO., LTD.	Fire	RN	2015
2015411RN	Ratchaburi Electricity Generating Holding Public Co.	Fire	RN	2015
2015413RN	Thai Agri Foods Public Company Limited	Fire	RN	2015
2015415RN	Siam Steel International Plc.	Fire	RN	2015
201145N	PROCUREMENT CENTRE (THAILAND) CO., LTD.	Fire	Ν	2011
2015418RN	TCC Land Co., Ltd.	Fire	RN	2015
201153N	Global Food Products Co., Ltd.	Water Damage	N	2011
201157N	SIAM TAKASHIMAYA (THAILAND) CO., LTD.	Water Damage	Ν	2011
201159N	Manulife Asset Management (Thailand)	Water Damage	Ν	2011
201152N	Jebsen & Jessen Group of Companies	Water Damage	Ν	2011
201158N	Amata Corporation Public Co., Ltd.	Water Damage	N	2011

# รูปที่ 4. 91 ผลลัพธ์จากการใช้ Left Outer Join

4.2.19 Merge Stage

10



รูปที่ 4. 92 รูป Merge Stage

Stage ในรูปที่ 4.92 ใช้ในการรวมข้อมูลเข้าด้วยกัน ซึ่งประกอบด้วย 1 Master Table และหลาย Update Table โดยอาศัย Key หรือคอลัมน์ที่ต้องการในการ Merge ดังรูปที่ 4.93 ซึ่งก่อนจะทำการรวมข้อมูล จะต้องทำการ Sort แบบ Ascending เสียก่อน โดยข้อแตกต่างที่ชัดเจนกับ Join Stage คือ Merge Stage สามารถมีได้หลาย Output Link และยังรองรับการทำ Reject Link ซึ่งจะทำการเก็บ record ที่ไม่เข้าคู่กับ Master Table แต่จำนวน Link นั้น จะต้องมีจำนวนเท่ากับ Update Table นอกจากนี้ ข้อมูลใน Update Table จะต้องมี Key ที่ไม่ซ้ำกัน ซึ่งตัวอย่างของข้อมูลต้นทาง จะแสดงในรูปที่ 4.94 – 4.95

Ę	Merge_36 - Merge	
	Stage name: Merge_36 General Properties Advanced Link Ordering NLS Locale	
	Image Keys   Image Keys <td></td>	
2	Key	
	OK Cancel Help	

รูปที่ 4. 93 Stage Properties ของ Merge Stage

CHANNEL_L4_CD	CHANNEL_L4_NM	CHANNEL_L3_CD	CHANNEL_L3_NM	CHANNEL_L2_CD	CHANNEL_L2_NM	CHANNEL_L1_CD	CHANNEL_L1_NM	CH_WTD	YEAR_CD
10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10111	Asset Management	0.0410	2014
10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10112	Sales Department	0.0260	2014
10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10113	Consumer	0.0240	2014
10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10114	Corporate	0.0300	2014
10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10115	Other1	0.0290	2014
10	Bank	102	Sales Corporate	1021	Sales Corp. 1	10211	Comporate Business	0.0150	2014

ร**ูปที่ 4. 94** ตัวอย่างข้อมูลจากปี 2014

CHANNEL_L4_CD	CHANNEL_L4_NM	CHANNEL_L3_CD	CHANNEL_L3_NM	CHANNEL	_L2_CD	CHANNEL_L2_NM	CHANNEL_L1_CD	CHANNEL_L1_NM	CH_WTD	YEAR_CD
10	Bank	101	Sales Retail Office	1011		Sales Retail Office	10111	Asset Management	0.0400	2015
10	Bank	101	Sales Retail Office	1011		Sales Retail Office	10112	Sales Department	0.0300	2015
10	Bank	101	Sales Retail Office	1011		Sales Retail Office	10113	Consumer	0.0700	2015
10	Bank	101	Sales Retail Office	1011		Sales Retail Office	10114	Corporate	0.0500	2015
10	Bank	101	Sales Retail Office	1011		Sales Retail Office	10115	Other1	0.0100	2015
10	Bank	102	Sales Corporate	1021		Sales Corp. 1	10211	Corporate Business	0.0210	2015

รูปที่ **4. 95** ตัวอย่างข้อมูลจากปี 2015



รูปที่ 4. 96 Parallel Job ที่ใช้ในการทดสอบการทำงานของ Merge Stage

จากรูปที่ 4.96 มีการใช้ Transformer Stage เพื่อทำการกำจัดคอลัมน์ที่ไม่ด้องการออกไป และได้ทำการแก้ไข ชื่อคอลัมน์ เนื่องจาก Merge Stage จะถือว่าคอลัมน์ที่ชื่อเหมือนกันนั้น คือคอลัมน์เดียวกัน จึงทำให้เลือก ข้อมูลได้จาก Input Link ใด Link หนึ่งเท่านั้น ซึ่งในกรณีนี้ ได้ใช้คอลัมน์ CHANNEL\_L1\_CD เป็น Key ที่ใช้ ในการ Merge ได้ใช้ชื่อจาก Table ปี 2014 และนำ CH\_WTD14 มาจาก CH\_WTD ใน Table ปี 2014 และ CH\_WTD15 มาจาก CH\_WTD ใน Table ปี 2015

#### 4.2.20 Difference Stage



# รูปที่ 4. 97 รูป Difference Stage

Stage ในรูปที่ 4.97 ใช้เปรียบเทียบข้อมูลว่ามีความแตกต่างกันหรือไม่ โดยกำหนดคอลัมน์ที่จะใช้เป็น Key เพื่อใช้ในการจับคู่เปรียบเทียบ ดังรูปที่ 4.98 และกำหนดคอลัมน์ที่เป็น Value ที่จะใช้ในการเปรียบเทียบ นอกจากนี้ยังต้องกำหนดลำดับ Before และ After ใน Link Ordering เนื่องจาก Difference Stage จะแสดงผล โดยใช้ข้อมูลจาก Link ที่ตั้งเป็น Before เป็นพื้นฐาน



ซึ่งผลลัพธ์ที่ได้ ดังรูปที่ 4.99 จะออกมาในรูปแบบคอลัมน์ที่ชื่อ DIFF ซึ่งจะเก็บค่า 2 หรือ 3 โดย 2 มี กวามหมายคือ Value ที่นำมาเปรียบเทียบทั้ง 2 ชุดนั้นเหมือนกัน และ 3 มีความหมายคือ Value ของทั้ง 2 ชุด ที่ใช้ในการเปรียบเทียบนั้น มีค่าแตกต่างกัน ตัวอย่างเช่นรูปด้านล่างที่มี 1 Record ที่มีค่า CH\_WTD เหมือนกันทั้งในปี 2014 และปี 2015 จึงทำให้แสดงผลลัพธ์ออกมาเป็น 2

	select * from	U_DIFF; 2	= Same, 3 = D	lff							>	~
sal Ec	litor 🖹 Configurat	ion 🚺 Validation	Special Registers	🕼 Performance Me	trics							
Pro	perties 🔲 SQL R	esults 🖾						¢	• ⇔ ‡ [		₩ -	
	CHANNEL_L4_CD	CHANNEL_L4_NM	CHANNEL_L3_CD	CHANNEL_L3_NM	CHANNEL_L2_CD	CHANNEL_L2_NM	CHANNEL_L1_CD	CHANNEL_L1_NM	CH_WTD	YEAR_CD	DIFF	^
1	10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10111	Asset Managem	0.0410	2014	3	
2	10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10112	Sales Department	0.0260	2014	3	≡
3	10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10113	Consumer	0.0240	2014	3	
4	10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10114	Corporate	0.0300	2014	3	
5	10	Bank	101	Sales Retail Office	1011	Sales Retail Office	10115	Other1	0.0290	2014	3	
6	10	Bank	102	Sales Corporate	1021	Sales Corp. 1	10211	Corporate Busin	0.0150	2014	3	
7	10	Bank	102	Sales Corporate	1021	Sales Corp. 1	10212	SME-Medium	0.0120	2014	3	
8	10	Bank	102	Sales Corporate	1021	Sales Corp. 1	10213	SC11	0.0010	2014	3	
9	10	Bank	102	Sales Corporate	1021	Sales Corp. 1	10214	SC12	0.0020	2014	3	
10	10	Bank	102	Sales Corporate	1022	Sales Corp. 2	10221	SME-Small	0.0300	2014	2	
11	10	Bank	102	Sales Corporate	1022	Sales Corp. 2	10222	Recovery	0.0170	2014	3	
12	10	Bank	102	Sales Corporate	1022	Sales Corp. 2	10223	A Small	0.0130	2014	3	
13	10	Bank	102	Sales Corporate	1022	Sales Corp. 2	10224	Other2	0.0200	2014	3	
14	10	Bank	102	Sales Corporate	1023	Micro Branches	10231	SME-Micro	0.0500	2014	3	
15	10	Bank	102	Sales Corporate	1023	Micro Branches	10232	Branch	0.0200	2014	3	$\sim$
Fetch	ed 63 records, 63 rec	cords shown						ctivate Wind	ows			

รูปที่ 4. 99 ตัวอย่างผลลัพธ์จากการใช้ Difference Stage

### 4.2.21 Change Capture Stage



รูปที่ 4. 100 รูป Change Capture Stage

Stage ในรูปที่ 4.100 ใช้ในการหาค่าความเปลี่ยนแปลงคล้าย Difference Stage แต่ Change Capture Stage จะ เก็บค่าเฉพาะ Record ที่มีความเปลี่ยนแปลงโดยใช้ข้อมูลที่มาจาก Link After จาก Link Ordering เป็นพื้นฐาน และสามารถใช้ร่วมกับ Change Apply Stage ได้ โดยวิธีการใช้งานจะเหมือน Difference Stage โดยการ กำหนดคอลัมน์ Key ที่ใช้ในการจับคู่เพื่อเปรียบเทียบ และกำหนดคอลัมน์ Value เพื่อใช้สำหรับเปรียบเทียบ ดังรูปที่ 4.101 และ รูปที่ 4.102

**NSTITUTE** OF
Stars I have I O would	
Stage Input   Output	
Stage name: Change Capture 32	
General Properties Advanced Link Ordering NLS Locale	e
Change Keys	[No property selected]
Sort Order = Ascending	
Change Values     Value = CH WTD	Information:
Change Mode = Explicit Keys & Values     Drop Output For Copy = True	
Drop Output For Delete = False	
	Available properties to add:
Log Statistics = False	Key
รูปที่ 4. 101 Stage Properties ๆ	101 Change Capture Stage
Stage pame:	
[Change_Capture_32	
General Properties Advan	ced Link Ordering NLS Locale
Order the following input links	
	• 0 /
After DSLink23	
P AND DOLLARY	
	±

จากรูปที่ 4.103 จะเห็นได้ว่าข้อมูลที่นำมาเปรียบเทียบนั้นมีทั้งหมด 63 rows แต่เมื่อผ่าน Change Capture Stage แล้ว จะเหลือเพียง 62 rows ซึ่งทั้งหมดล้วนเป็น row ที่มีการเปลี่ยนแปลงค่า



รูปที่ 4. 103 ตัวอย่างการใช้งาน Change Capture Stage

4.2.22 Change Apply Stage



<mark>รูปที่ 4. 104</mark> รูป Change Apply Stage

Stage ในรูปที่ 4.104 ใช้สำหรับกา<mark>รนำค่</mark>าที่เปลี่ยนแปลงที่พบจาก Change Capture Stage มาอัพเดทข้อมูลเก่า โดย Link ที่มาจาก Change Capture Stage จะต้องอยู่ใน Link Change ใน Link Ordering ซึ่งการใช้งาน Change Apply Stage ทำได้โดยกำหนดคอลัมน์ Key เพื่อใช้ในการเปรียบเทียบ และคอลัมน์ที่มีการ เปลี่ยนแปลง ดังรูปที่ 4.105 และ 4.106

Stage   Input   Output		Stage Input Output
Stage name: Change_Apply_33		Stage name: Change_Apply_33
General Properties Advanced Link Ordering NL	S Locale	General Properties Advanced Link Ordering NLS Local
Change Keys Key = CHANNEL_L1_CD Soft Order = Ascending Change Values Value = CH_WTD Value = YEAR_CD Options Change Mode = Explicit Keys & Values Change Mode = Explicit Keys & Values Check Value Columns on Delete = True Log Statistics = False		Order the following input links:

รูปที่ 4. 105 Stage Properties (ซ้าย) และ Stage Link Ordering (บวา) ของ Change Apply Stage

CHANNEL_L2_CD	CHANNEL_L2_NM	CHANNEL_L1_CD	CHANNEL_L1_NM	CH_WTD	YEAR_CD
1011	Sales Retail Office	10111	Asset Manageme	0.0400	2015
1011	Sales Retail Office	10112	Sales Department	0.0300	2015
1011	Sales Retail Office	10113	Consumer	0.0700	2015
1011	Sales Retail Office	10114	Corporate	0.0500	2015
1011	Sales Retail Office	10115	Other1	0.0100	2015
1021	Sales Corp. 1	10211	Corporate Business	0.0210	2015
1021	Sales Corp. 1	10212	SME-Medium	0.0130	2015
1021	Sales Corp. 1	10213	SC11	0.0034	2015
1021	Sales Corp. 1	10214	SC12	0.0026	2015
1022	Sales Corp. 2	10221	SME-Small	0.0300	2014
1022	Sales Corp. 2	10222	Recovery	0.0150	2015
1022	Sales Corp. 2	10223	A Small	0.0120	2015
1022	Sales Corp. 2	10224	Other2	0.0030	2015
			a)		

รูปที่ 4. 106 ตัวอย่างผลลัพธ์ที่ได้จากการใช้ Change Apply Stage

รูปที่ 4.107 เป็นตัวอย่างของการใช้ Change Apply Stage โดยกำหนดให้อัพเดทคอลัมน์ CH\_WTD, YEAR\_CD เฉพาะคอลัมน์ที่มีความเปลี่ยนแปลงของ CH\_WTD โดยเปรียบเทียบระหว่างปี 2014 กับปี 2015 โดยในกรอบสีแดงคือค่าที่ไม่ได้มีความเปลี่ยนแปลงใด ๆ ทั้งสิ้น



รูปที่ 4. 107 ตัวอย่างการใช้งาน Change Apply Stage

4.2.23 Pivot Enterprise Stage



รูปที่ 4. 108 รูป Pivot Enterprise Stage

Stage ในรูปที่ 4.108 ใช้สำหรับการ Pivot Table ซึ่งสามารถทำได้ทั้งแบบ Vertical และ Horizontal ซึ่งการทำ แบบ Vertical คือการนำคอลัมน์มาใส่ในรูปแบบแถว ส่วน Vertical คือการนำแถวมาจัคใส่ในคอลัมน์ ซึ่งใน ตัวอย่างนี้จะเป็นการทำ Horizontal Pivot โดยการใช้งาน Pivot Enterprise Stage นั้น จะต้องกำหนดรูปแบบ ของการ Pivot ที่ Stage > Properties เสียก่อน แล้วจึงมากำหนด Pivot Properties ว่าต้องการให้จัดกลุ่มโดย คอลัมน์ใด และคอลัมน์ใดที่ต้องการทำ Pivot แล้วจึงระบุ Array Size ให้เหมาะสม ซึ่งในรูปที่ 4.109 ต้องการ จัดให้ข้อมูลรายเดือน มารวมกันเป็นรายปี ดังนั้นจึงตั้งก่าไว้ที่ 12 ซึ่งก็คือจำนวนเดือน

	]			Pivot_	Enterprise	e_19 - PxPivo	ot stage	_ <b>□</b> ×	
	Stage	linput Ou	utput						
	Stage	e name:							
	Pivot	t_Enterprise_1	9						
	Gen	neral Properti	es Pivot Pro	perties Adv	anced				
		Name	SQL Type	Length	Scale	GroupBy	Pivot	Aggregation functions required	
	1	YEAR_CD	Numeric	4		✓			
	2	MONTH_W1	Decimal	3	2				
						<u> </u>			
							81 7		
								$\boldsymbol{\rho}$	
								S	И
4	1								
t		Array Size	- IT	2					
		D Pivot In	dev.						
		Court	J Divert else els						
r.		Group by and	I FIVOI Check	boxes are de	signed for ve	ertical pivot		Load	
							ОК	Cancel Help	
									11

รูปที่ 4. 109 Stage Pivot Properties ของ Pivot Enterprise Stage

\*ข้อกวรระวังของการใช้ Pivot Enterprise Stage คือผลลัพธ์ที่ได้นั้น จะมีลักษณะเป็น Array ซึ่งเก็บลง ฐานข้อมูลไม่ได้ โดยมีวิธีการแก้ไขคือการวาง Transformer Stage ไว้หลัง Pivot Enterprise Stage ซึ่งภายใน Transformer Stage นั้น เพียงแค่ Mapping Output ให้เหมือนกับ Input (ไม่เกิดความเปลี่ยนแปลง)

1	YEAR_CD	MONTH_KEY	MONTH_NM	MO	NTH_CD	MONTH_WTD	
2	2015	1	January		201501	0.11	
3	2015	2	February		201502	0.06	
4	2015	3	March		201503	0.13	
5	2015	4	Arpril		201504	0.12	
6	2015	5	May		201505	0.08	
7	2015	6	June		201506	0.04	
8	2015	7	July		201507	0.08	
9	2015	8	August		201508	0.08	
10	2015	9	September		201509	0.04	
11	2015	10	October		201510	0.09	
12	2015	11	November		201511	0.07	
13	2015	12	December		201512	0.1	
14	2014	1	January		201401	0.06	
15	2014	2	February		201402	0.08	
16	2014	3	March		201403	0.07	
17	2014	4	Arpril		201404	0.1	
18	2014	5	May		201405	0.12	
19	2014	6	June		201406	0.04	
20	2014	7	July		201407	0.09	

# ตัวอย่างข้อมูล ประกอบด้วยข้อมูลก่อนการแปรรูป ดังรูปที่ 4.110 และข้อมูลหลังการแปรรูป ดังรูปที่ 4.111

**รูปที่ 4. 110** ข้อมูลก่อนแปรรูป

	YEAR_CD	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1	2011	0.04	0.06	0.04	0.09	0.08	0.07	0.12	0.13	0.08	0.08	0.10	0.11
2	2012	0.08	0.06	0.09	0.07	0.04	0.08	0.12	0.04	0.13	0.08	0.11	0.10
3	2013	0.04	0.04	0.10	0.11	0.07	0.06	0.13	0.09	0.08	0.12	0.08	0.0
4	2014	0.06	0.08	0.07	0.10	0.12	0.04	0.09	0.11	0.13	0.08	0.04	0.08
5	2015	0.11	0.06	0.13	0.12	0.08	0.04	0.08	0.08	0.04	0.09	0.07	0.10

รูปที่ 4. 111 ผลลัพธ์หลังแปรรูป

99

# บทที่ 5 บทสรุปและข้อเสนอแนะ

#### 5.1 สรุปผลการดำเนินงาน

การดำเนินงานในครั้งนี้ ผู้จัดทำโครงงานได้เริ่มจากการศึกษา concept ต่าง ๆ ของการทำ Business Intelligence ซึ่งมีการวัดผลโดยการตอบคำถามจากพนักงานที่ปรึกษา พร้อมกับได้รับมอบหมายให้ทำการ ตรวจสภาพรายงานของลูกค้า ซึ่งพนักงานที่ปรึกษาก็ได้ให้ทรัพยากรต่าง ๆ ที่จำเป็นต่อการทำความเข้าใจ ระบบของลูกค้า จากนั้นจึงได้รับมอบหมายโจทย์ที่ระบุรูปแบบและผลลัพธ์ข้อมูลต่าง ๆ ที่ต้องการ โดยให้ใช้ SQL ในการทำ เพื่อทดสอบความสามารถ และเสริมสร้างความเข้าใจขั้นตอนการทำ ETL

ในส่วนของการศึกษาซอฟต์แวร์ต่าง ๆ นั้น ถูกกำหนดโดยซอฟต์แวร์ที่ถูกค้าใช้งาน จึงทำให้ด้อง ศึกษาหลายซอฟต์แวร์ด้วยกัน ซึ่งจะทำตาม workshop ที่ได้รับ และมีการทบทวนกับพนักงานที่ปรึกษาเป็น ระยะ แล้วจึงทำการทดสอบโดยการตอบคำถามของพนักงานที่ปรึกษา ซึ่งช่วยส่งเสริมความเข้าใจใน ซอฟต์แวร์ที่เกี่ยวข้องกับ BI มากยิ่งขึ้น

สุดท้ายนี้กือการศึกษาและจัดทำกู่มือการใช้งาน IBM® InfoSphere® DataStage® ซึ่งเป็นซอฟต์แวร์ ที่เริ่มจะได้รับความสนใจ ซึ่งหน้าที่ที่ผู้จัดทำโครงงานได้รับมอบหมายกือ ศึกษาและอธิบายความสามารถ ต่าง ๆ ของ DataStage® ซึ่งจะมีการนำเสนอกับพนักงานที่ปรึกษาเป็นระยะ ๆ จนปรับปรุงให้เนื้อหาครบถ้วน เท่าที่เวลาจะเอื้ออำนวย

### 5.2 ปัญหาที่พบ

 ผู้จัดทำโครงงานนั้น ใช้ระยะเวลาในการระบุสาเหตุของปัญหาค่อนข้างนาน เนื่องจากระบบที่ศึกษา มีความซับซ้อน และระยะเวลาที่ค่อนข้างน้อย จึงทำให้ไม่เข้าใจระบบได้ดีพอที่จะระบุต้นเหตุของปัญหาได้ ทันที

 ใฟล์บางอย่างที่จำเป็นต่อการเชื่อมต่อฐานข้อมูลนั้น ไม่ได้มีมาพร้อมกับ IBM® InfoSphere® DataStage® และผู้ให้บริการที่เป็นเจ้าของไฟล์นั้น ก็ได้ยกเลิกการสนับสนุนไปแล้วเนื่องจากมี version ที่ ใหม่กว่า ซึ่งไม่สามารถใช้กับ DataStage® ได้

โน l ล *ยั* 

#### 5.3 แนวทางการแก้ไขปัญหา

เพื่อเป็นการลดระยะเวลาทำกวามเข้าใจระบบที่นั้น ผู้จัดทำโครงงานจึงได้เน้นไปที่การศึกษา BI concept และเครื่องมืออื่น ๆ ที่ใช้ในการทำระบบขึ้นมา เพื่อช่วยให้การทำความเข้าใจระบบ สามารถเป็นไป ได้อย่างรวดเร็วยิ่งขึ้น

 ในส่วนนี้ ทางผู้จัดทำโครงงานได้ปรึกษากับพนักงานที่ปรึกษาเพื่อหาแนวทางแก้ไข ซึ่งในกรณีนี้ กือการใช้ Stage อื่นทดแทน

#### 5.4 ข้อเสนอแนะจากการ<mark>ดำเนินงาน</mark>

ในกรณีที่เจอปัญหาที่มีความรู้สึกว่าแก้ไขไม่ได้ อาจจะเป็นเพราะขาดความรู้ที่จำเป็นต่อการทำงานชิ้นนั้น ดังนั้นนอกจากการหาความรู้และข้อมูลด้วยตนเองแล้ว ยังจำเป็นที่จะต้องมีความสามารถในการหาความรู้ พื้นฐานที่จำเป็นอีกด้วย ซึ่งการปฏิบัติสหกิจนั้น ย่อมได้เจอสิ่งใหม่ ๆ ที่ไม่เคยได้พบเจอเป็นเรื่องธรรมดา ดังนั้นการจะขอคำแนะนำจากพนักงานที่ปรึกษาจึงไม่ใช่เรื่องที่ผิด แต่ก่อนจะถามอะไร ก็ควรเรียบเรียงและ ตั้งเป้าหมายให้ชัดเจนเสียก่อน เพื่อความรวบรัดชัดเจน

#### เอกสารอ้างอิง

[1] A-HOST Company Limited, 2012, A-HOST Company Limited – Contact Us [Online], Available : http://www.a-host.co.th/index.php?option=com\_content&view=article&id=27&Itemid=128
[3 มิถุนายน 2559].

[2] A-HOST Company Limited, 2012, A-HOST Company Limited – A-HOST Services [Online],Available :

http://www.a-host.co.th/index.php?option=com\_content&view=article&id=44&Itemid=114

[1 ตุลาคม 2559]

[3] Ironside Group Inc., 2015, ETL vs. ELT – What's the Big Difference? [Online], Available : https://www.ironsidegroup.com/2015/03/01/etl-vs-elt-whats-the-big-difference/

[1 ตุลาคม 2559]

[4] IBM Corp., 2015, IBM Knowledge Center - Parallelism basics in IBM InfoSphere Information Server [Online], Available :

http://www.ibm.com/support/knowledgecenter/SSZJPZ\_11.5.0/com.ibm.swg.im.iis.productization.iisinfsv .overview.doc/topics/cisoarchparalbasics.html

[3 ตุลาคม 2559]

[5] IBM Corp. 2011, "Information Server backbone",

IBM InfoSphere DataStage Essential v8.7 Instructor Guide,

หน้า 1-7.

[6] IBM Corp., 2015, IBM Knowledge Center - IBM InfoSphere DataStage and IBM InfoSphere QualityStage roles [Online], Available :

http://www.ibm.com/support/knowledgecenter/SSZJPZ\_9.1.0/com.ibm.swg.im.iis.found.admin.common. doc/topics/c\_wdisad\_WebSphere\_DataStage\_project\_roles.html

[3 ตุลาคม 2559]

[7] IBM Corp., 2011, Develop [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_intro\_develop.html?view=embed

[6 ตุลาคม 2559]

[8] IBM Corp., 2011, Test [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_test.html?view=embed

[6 ตุลาคม 2559]

#### [9] IBM Corp., 2011, Deploy [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_intro\_deploy.html?view=embed

[6 ตุลาคม 2559]

[10] IBM Corp., 2011, Operate [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_intro\_operate.html?view=embed

[6 ตุลาคม 2559]

[11] IBM Corp., 2011, Extract and load data [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_samples\_extract.html?view=embed

[10 ตุลาคม 2559]

[12] IBM Corp., 2011, Transform data [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_samples\_transform.html?view=embed

[10 ตุลาคม 2559]

[13] IBM Corp., 2011, Enrich data [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_samples\_enrich.html?view=embed

[10 ตุลาคม 2559]

[14] IBM Corp., 2011, Cleanse data [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_samples\_cleanse.html?view=embed

[10 ตุลาคม 2559]

[15] IBM Corp., 2011, Real-time processing [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_samples\_realtime.html?view=embed

[10 ตุลาคม 2559]

[16] IBM Corp., 2011, Big data processing [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_samples\_bigdata.html?view=embed

[10 ตุลาคม 2559]

[17] IBM Corp., 2011, Combine jobs in a sequence job [Online], Available :

http://www.ibm.com/support/knowledgecenter/en/SSZJPZ\_11.3.0/com.ibm.swg.im.iis.ds.intro.doc/topics/ ds\_sequence\_job.html?view=embed

> กุกโนโลยั7 กุ

[10 ตุลาคม 2559]

10



รับโล ภาคผนวก ก. สไลด์ประกอบการนำเสนอ

CAN INSTITUTE OF TECH

 $\mathbf{S}$ 













# ประวัติผู้จัดทำโครงงาน

ชื่อ – สกุล

นายศรัณยู ศรีวิจิตรานนท์

วัน เดือน ปีเกิด 17 พฤศจิก

17 พฤศจิกายน พ.ศ. 2537

#### ประวัติการศึกษา

ระดับประถมศึกษา

ระดับมัธยมศึกษา

ประถมศึกษาตอนต้น – ประถมศึกษาตอนปลาย พ.ศ. 2544 – 2549

ิ ล ยั

โรงเรียนอัสสัมชัญแผนกประถม

มัธยมศึกษาตอนต้น – มัธยมศึกษาตอนปลาย พ.ศ. 2550 – 2553

โรงเรียนอัสสัมชัญ

มัธยมศึกษาตอนปลาย พ.ศ. 2554 – 2555

กศน. เขตยานนาวา

- ใม่มี -

ระดับอุดมศึกษา

คณะเทคโนโลยีสารสนเทศ สาขาเทคโนโลยีสารสนเทศ

ิส<mark>ถาบัน</mark>เทคโนโลยี<mark>ไทย – ญี่</mark>ปุ่น

ทุนการศึกษา

ประวัติการฝึกอบรม

1. Project-Based Learning by TNI & Osaka Institute of Technology

2. Network Cabling System by Interlink Communication PCL

3. Social Network Security by S-Generation

## ผลงานที่ได้รับการตีพิมพ์

(\*

- 1. S. Srivichitranond and Group, Design of Bus Tracking Framework Based on Android Mobile Application by using GPS, Proc. of the 2016 Int. Conf. on Business and Industrial Research, Bangkok, 12-13 May 2016.
- 2. T. Metkarunchit, K. Charoenpojvajana, S. Srivichitranond, Develop 3D Map Signal Strength Surveying System for Cellular Mobile Phone, Proc. of the 2016 Int. Conf. on Business and Industrial Research, Bangkok, 12-13 May 2016. Ĩ I N S

9