บทที่ 3

3. ขั้นตอนการดำเนินงานของโครงงาน

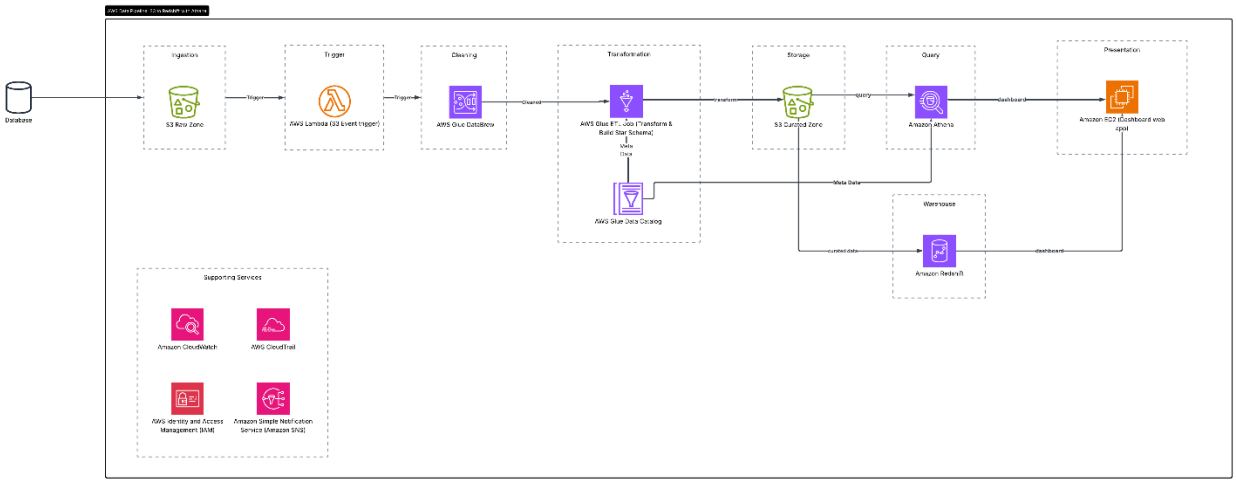
3.1 ความต้องการของระบบ

จากการศึกษากระบวนการจัดการข้อมูลของธุรกิจอีคอมเมิร์ซ (ชุดข้อมูล Olist) ระบบคลังข้อมูลบนระบบคลาวด์ที่พัฒนาขึ้นมีความต้องการดังต่อไปนี้

1. ด้านการนำเข้าและทำงานอัตโนมัติ: ระบบต้องสามารถนำเข้าข้อมูลที่ทยอยเข้ามาใหม่สู่พื้นที่ S3 Raw Zone ได้ และต้องสามารถตรวจจับไฟล์ใหม่เพื่อสั่งรันกระบวนการ ETL ได้โดยอัตโนมัติ (Automated Pipeline) โดยไม่ต้องใช้คนกดสั่งงาน
2. ด้านการทำความสะอาดและแปลงข้อมูล: ระบบต้องสามารถจัดการข้อมูลที่สูญหาย (Missing Value) หรือผิดรูปแบบ และแปลงโครงสร้างข้อมูลให้อยู่ในรูปแบบ Star Schema ที่เหมาะสมกับการวิเคราะห์เชิงธุรกิจ
3. ด้านการจัดเก็บข้อมูล: ระบบต้องจัดเก็บข้อมูลที่พร้อมใช้งานไว้ใน Amazon S3 (Curated Zone) และ Amazon Redshift
4. ด้านการแสดงผลและสืบค้น: ระบบต้องรองรับการสืบค้นข้อมูลผ่าน Amazon Athena และ Amazon Redshift เพื่อเปรียบเทียบประสิทธิภาพ และต้องมีเว็บแอปพลิเคชัน (Dashboard) เพื่อแสดงผลข้อมูลเชิงลึกให้เข้าใจง่าย
5. ด้านข้อจำกัด: ระบบทั้งหมดต้องสามารถทำงานได้ภายใต้ข้อจำกัดของ AWS Academy Learner Lab (เช่น งบประมาณ, สิทธิ์ LabRole, จำนวน Concurrency ของ Service)

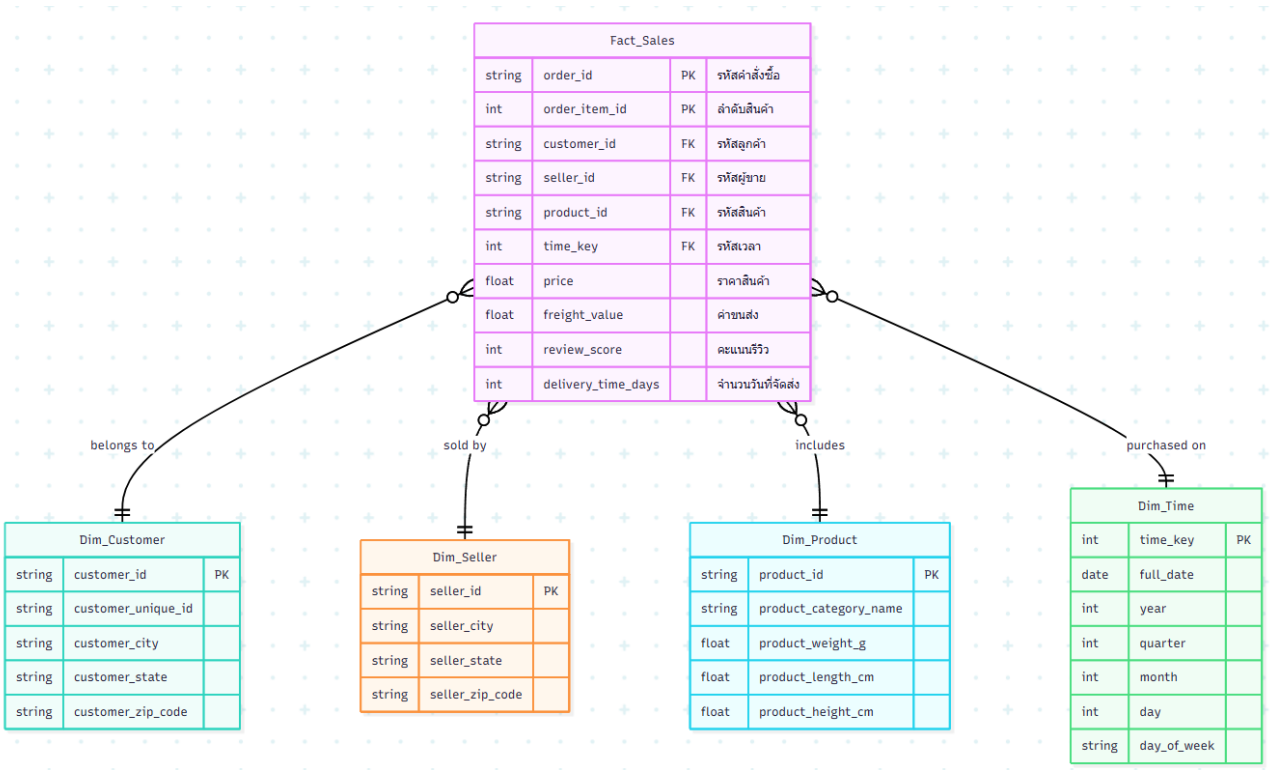
3.2 การวิเคราะห์และออกแบบระบบ

1. การออกแบบโครงสร้างการใช้งานของ AWS (Cloud Architecture Design)



รูปที่ 3.1 โครงสร้างการใช้งาน AWS

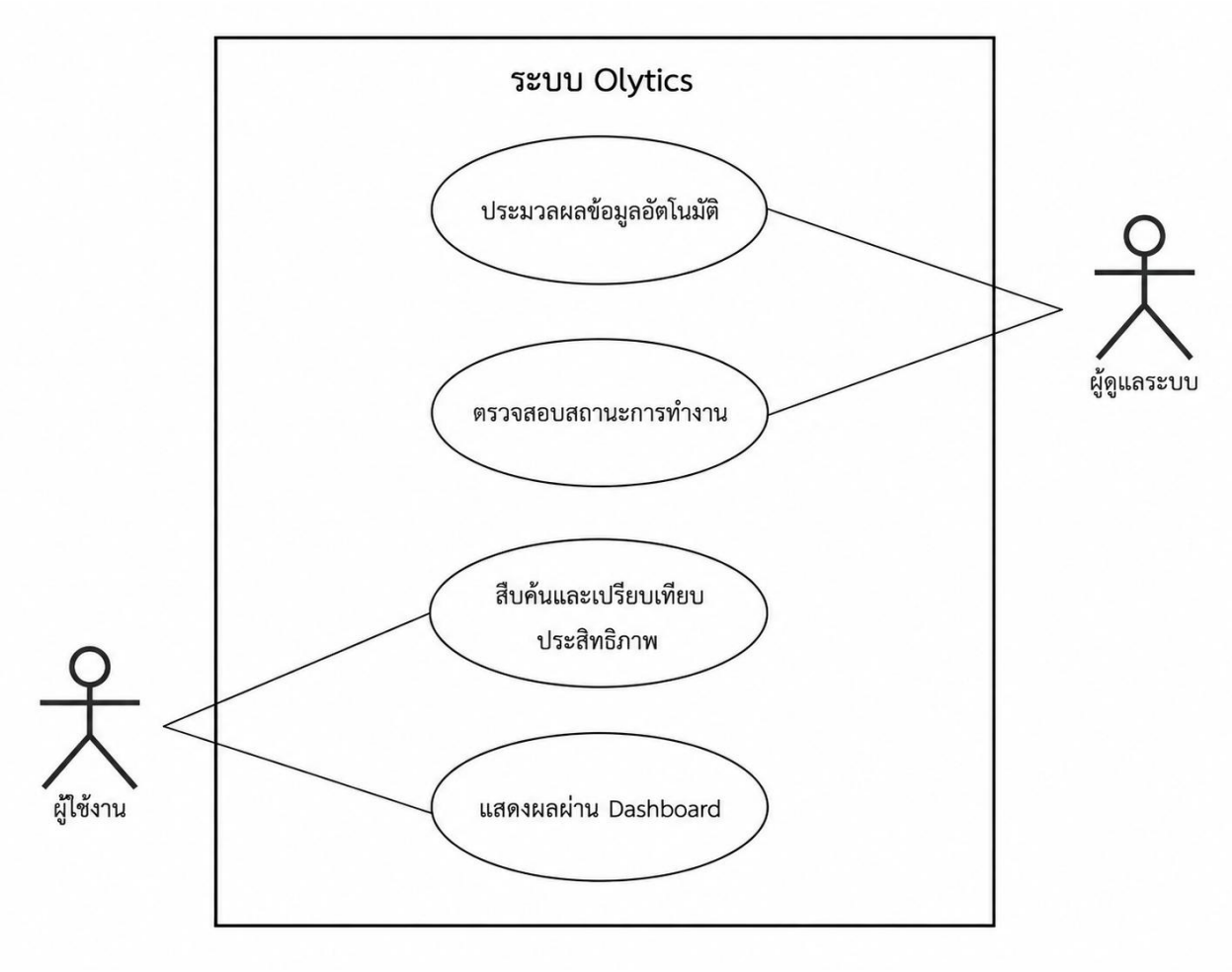
2. การออกแบบโครงสร้างข้อมูล (Data Modeling / Star Schema)



รูปที่ 3.2 ตาราง Star Schema

3.3 หลักการทำงานของระบบ

1. Use Case Diagram



รูปที่ 3.3 Use Case Diagram

2. Use Case Description

Use Case Description : ประมวลผลข้อมูลอัตโนมัติ

Use Case Name : ประมวลผลข้อมูลอัตโนมัติ	ID : UC-01
Actor : ผู้ดูแลระบบ	
Stakeholders and Interests : ผู้ดูแลระบบ, ผู้ใช้งาน	
Brief Description : Use case นี้ครอบคลุมกระบวนการนำเข้าข้อมูลดิบ ตรวจจับไฟล์ใหม่ ทำความสะอาดข้อมูล แปลงข้อมูลให้อยู่ในรูปแบบ Star Schema และจัดเก็บลงคลังข้อมูล โดยกระบวนการทั้งหมดทำงานต่อเนื่องกันโดยอัตโนมัติเมื่อมีข้อมูลชุดใหม่เข้าสู่ระบบ	
Normal Flow of ประมวลผลข้อมูลอัตโนมัติ	
ผู้ดูแลระบบ	ระบบ
1.ผู้ดูแลระบบดึงข้อมูลจากฐานข้อมูลต้นทางแบ่งออกเป็นชุดแล้วอัปโหลดเข้าสู่ S3 Raw Zone	
	2. ระบบตรวจพบไฟล์ใหม่ผ่าน S3 Event Notification และส่ง Event ไปยัง Lambda
	3. Lambda เรียก Glue Job ให้เริ่มทำงาน
	4. Glue DataBrew ทำความสะอาดข้อมูล ลบข้อมูลซ้ำ และจัดการค่าว่าง

	5. Glue ETL Job แปลงข้อมูลให้อยู่ในรูปแบบ Star Schema พร้อมคำนวณค่าที่จำเป็น เช่น ระยะเวลาจัดส่ง
	6. ระบบจัดเก็บผลลัพธ์ไว้ใน S3 Curated Zone และโหลดเข้า Amazon Redshift
	7. Glue Crawler อัปเดต Glue Data Catalog
	8. ระบบบันทึกสถานะการทำงานว่าเสร็จสิ้น
Alternate Condition : 3-a: Glue Job มีงานอื่นกำลังทำงานอยู่ (Concurrency เต็ม) 3-a-1: ระบบบันทึก Log ข้อผิดพลาดผ่าน Amazon CloudWatch และส่งการแจ้งเตือนไปยังผู้ดูแลระบบผ่าน Amazon SNS 5-a: ข้อมูลมีรูปแบบผิดพลาดจนไม่สามารถแปลงได้ 5-a-1: ระบบบันทึก Error Log ผ่าน Amazon CloudWatch และแจ้งเตือนผู้ดูแลระบบผ่าน Amazon SNS แล้วหยุดกระบวนการเฉพาะไฟล์นั้น 5-a-2: ผู้ดูแลระบบตรวจสอบ Log และแก้ไขข้อมูลต้นทางก่อนนำเข้าใหม่	

ตารางที่ 3.1 Use Case Description ประมวลผลข้อมูลอัตโนมัติ

Use Case Description : ตรวจสอบสถานะการทำงาน

Use Case Name : ตรวจสอบสถานะการทำงาน	ID : UC-02
Actor : ผู้ดูแลระบบ	
Stakeholders and Interests : ผู้ดูแลระบบ, ผู้ใช้งาน	
Brief Description : Use case นี้เป็นการตรวจสอบสถานะและ Log การทำงานของ ETL Pipeline ผ่าน Amazon CloudWatch เพื่อยืนยันความถูกต้องของการทำงาน หรือค้นหาสาเหตุเมื่อเกิดข้อผิดพลาด นอกจากนี้ยังสามารถตรวจสอบกิจกรรมการเรียกใช้ AWS API ย้อนหลังผ่าน AWS CloudTrail เพื่อวัตถุประสงค์ด้านความปลอดภัยเพิ่มเติม	
Normal Flow of ตรวจสอบสถานะการทำงาน	
ผู้ดูแลระบบ	ระบบ
1. ผู้ดูแลระบบเข้าสู่ Amazon CloudWatch Console	
	2. ระบบแสดงรายการ Log ของ Glue Job และ Lambda ที่ทำงานผ่านมา
3. ผู้ดูแลระบบเลือกดูรายละเอียด Log ของงานที่ต้องการตรวจสอบ	
	4. ระบบแสดงสถานะการทำงาน (สำเร็จ / ล้มเหลว) และรายละเอียดข้อผิดพลาด (ถ้ามี)

Alternate/Exceptional Flows 4-a: พบข้อผิดพลาดใน Log 4-a-1: ผู้ดูแลระบบตรวจสอบสาเหตุจากรายละเอียด Error 4-a-2: ผู้ดูแลระบบแก้ไขต้นเหตุ แล้วสั่งรัน Glue Job นั้นใหม่ด้วยตนเองผ่าน AWS Console

ตารางที่ 3.2 Use Case Description ตรวจสอบสถานะการทำงาน

Use Case Description : สืบค้นและเปรียบเทียบประสิทธิภาพ

Use Case Name : สืบค้นและเปรียบเทียบประสิทธิภาพ	ID : UC-03
Actor : ผู้ใช้งาน	
Stakeholders and Interests : ผู้ใช้งาน	
Brief Description : Use case นี้เป็นการสืบค้นข้อมูลชุดเดียวกันผ่าน Amazon Athena และ Amazon Redshift เพื่อเปรียบเทียบผลลัพธ์ทั้งด้านเวลาในการประมวลผลและรูปแบบค่าใช้จ่าย	
Normal Flow of สืบค้นและเปรียบเทียบประสิทธิภาพ	
ผู้ดูแลระบบ	ระบบ
1. ผู้ใช้งานเขียนคำสั่ง SQL สำหรับสืบค้นข้อมูลที่ต้องการวิเคราะห์	
2. ผู้ใช้งานรันคำสั่งผ่าน Amazon Athena	
	3. ระบบสืบค้นข้อมูลจาก S3 Curated Zone โดยตรงและแสดงเวลาที่ใช้พร้อมปริมาณข้อมูลที่สแกน
4. ผู้ใช้งานรันคำสั่ง SQL เดียวกันผ่าน Amazon Redshift	
	5. ระบบสืบค้นข้อมูลจากคลังข้อมูล และแสดงเวลาที่ใช้ในการประมวลผล
6.ผู้ใช้งานเปรียบเทียบผลลัพธ์ทั้งสองด้านเวลาและค่าใช้จ่าย	

Alternate Condition : 3-a: Athena แจ้งเตือนปริมาณข้อมูลที่ต้องสแกนเกินกว่าที่เหมาะสม 3-a-1: ผู้ใช้งานปรับคำสั่ง SQL ให้จำกัดขอบเขตข้อมูลแล้วสืบค้นใหม่

ตารางที่ 3.3 Use Case Description สืบค้นและเปรียบเทียบประสิทธิภาพ

Use Case Description : แสดงผลการวิเคราะห์ผ่าน Dashboard

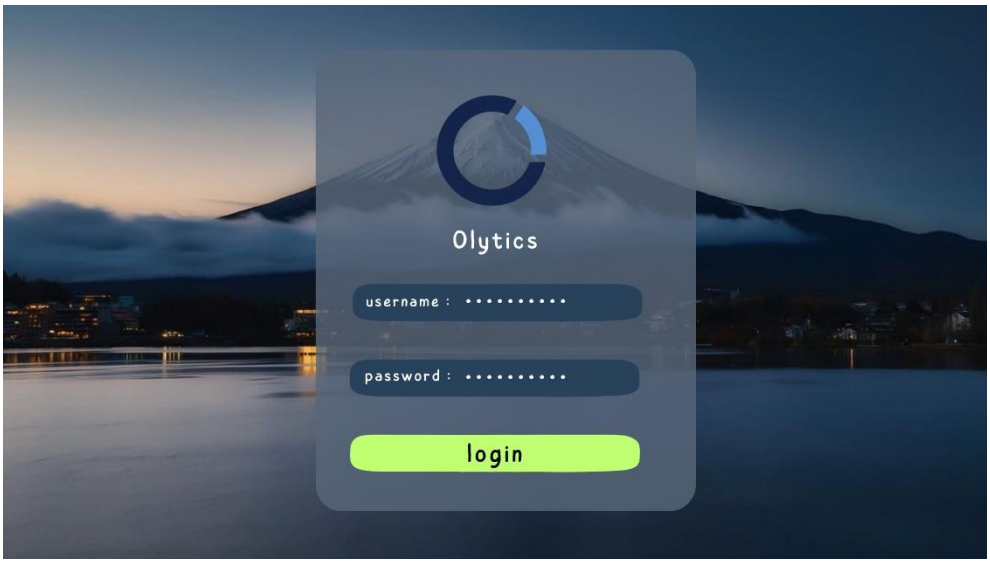
Use Case Name : แสดงผลการวิเคราะห์ผ่าน Dashboard	ID : UC-04
Actor : ผู้ใช้งาน	
Stakeholders and Interests : ผู้ใช้งาน	
Brief Description : Use case นี้เป็นการเข้าดูผลการวิเคราะห์ข้อมูลเชิงธุรกิจผ่านเว็บแอปพลิเคชัน Dashboard ที่พัฒนาขึ้นด้วย Next.js โดยดึงข้อมูลจาก Amazon Redshift	
Normal Flow of แสดงผลการวิเคราะห์ผ่าน Dashboard	
ผู้ดูแลระบบ	ระบบ
1. ผู้ใช้งานเข้าสู่เว็บแอปพลิเคชัน Dashboard ผ่านเบราว์เซอร์	
	2. เว็บแอปพลิเคชันร้องขอข้อมูลจาก Amazon Redshift ผ่าน API ที่พัฒนาขึ้น
	3. ระบบแสดงผลข้อมูลในรูปแบบกราฟและตาราง เช่น แนวโน้มยอดขายรายเดือน หมวดสินค้ายอดนิยม และความสัมพันธ์ระหว่างระยะเวลาจัดส่งกับคะแนนรีวิว
4. ผู้ใช้งานเลือกกรองข้อมูลตามช่วงเวลาหรือหมวดหมู่ที่ต้องการ	
	5. ระบบแสดงผลข้อมูลที่กรองแล้ว

Alternate Condition : 2-a: ไม่สามารถเชื่อมต่อกับ Amazon Redshift ได้ 2-a-1: ระบบแสดงข้อความแจ้งเตือนข้อผิดพลาดให้ผู้ใช้งานทราบ

ตารางที่ 3.4 Use Case Description แสดงผลการวิเคราะห์ผ่าน Dashboard

3. User Interface (Prototype)

หน้าเข้าสู่ระบบ



รูปที่ 3.4 หน้าเข้าสู่ระบบ

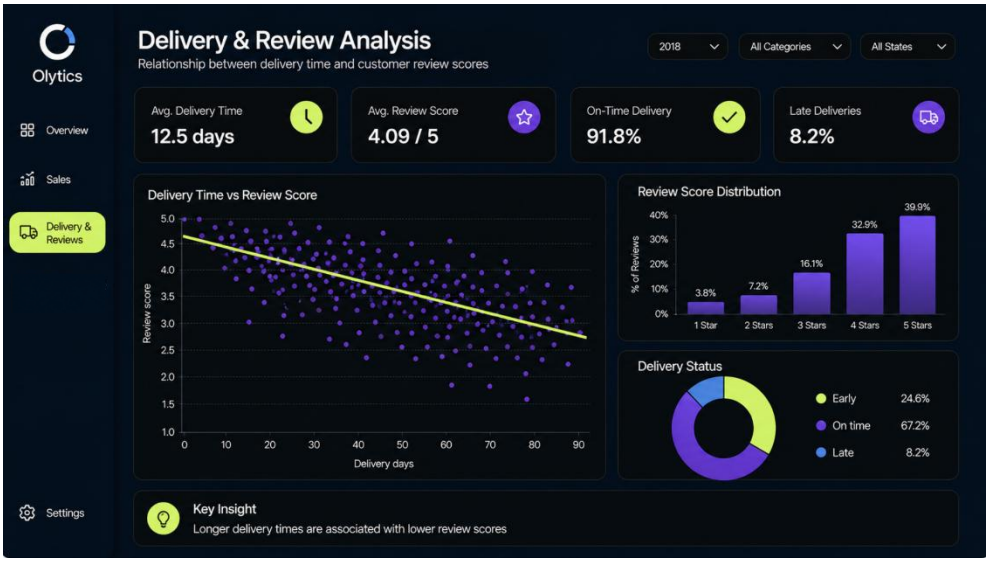
หน้าภาพรวมระบบวิเคราะห์ข้อมูล

รูปที่ 3.5 หน้าภาพรวมระบบวิเคราะห์ข้อมูล

หน้าวิเคราะห์ยอดขาย

รูปที่ 3.6 หน้าวิเคราะห์ยอดขาย

หน้าวิเคราะห์การจัดส่งและรีวิวลูกค้า



รูปที่ 3.7 หน้าวิเคราะห์การจัดส่งและรีวิวลูกค้า