

เอกสารประกอบ

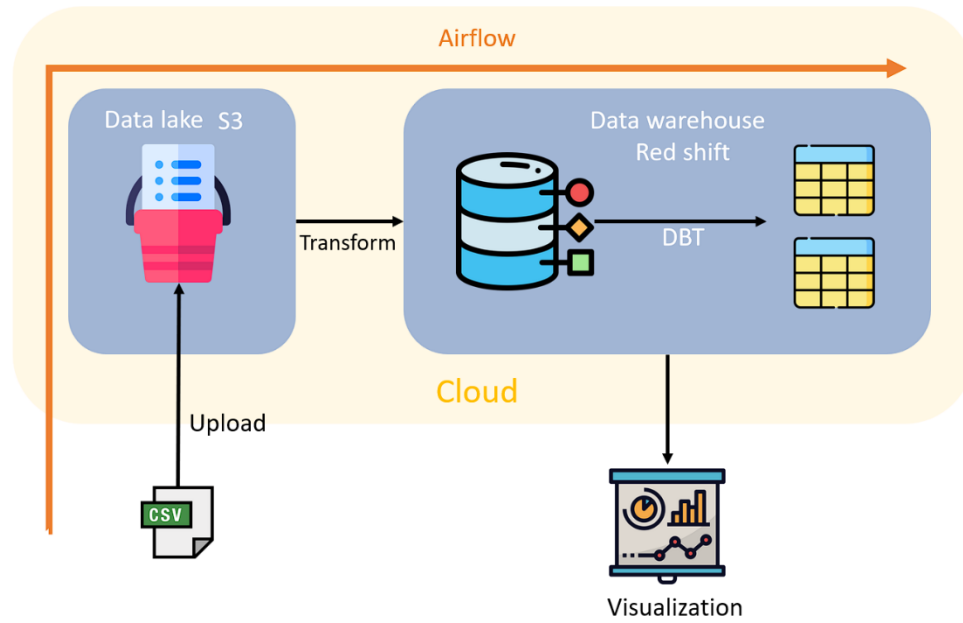
ปัญหาและที่มา

เนื่องจากในแต่ละวันทางร้าน Pizza จะต้องมีการสต็อกวัตถุดิบในการทำพิซซ่าแบบวันต่อวัน โดยจะต้องกำหนดปริมาณให้พอดีกับช่วงเวลาที่เปิดทำการในแต่ละวัน ซึ่งเมนูพิซซ่าต่างๆของทางร้านนั้นโดยปกติแล้วจะมีเมนูที่ขายดีและเมนูที่ขายไม่ดีทำให้การวัตถุดิบที่สต็อกไว้อาจจะไม่พอทำเมนูพิซซ่าที่ขายดี และอาจจะทำให้วัตถุดิบที่ใช้ทำเมนูพิซซ่าที่ขายไม่ดีนั้นเหลือทิ้ง

ดังนั้นจึงได้จัดทำแดต้าแพลตฟอร์มสำหรับร้านพิซซ่าเพื่อใช้ในการเก็บข้อมูลการขายพิซซ่าในแต่ละวันและนำไปวิเคราะห์เพื่อแก้ไขปัญหาที่กล่าวมาในข้างต้น และยังสามารถนำไปวิเคราะห์พฤติกรรมของลูกค้าได้ด้วยว่าชอบเมนูพิซซ่าชนิดใดไม่ชอบเมนูพิซซ่าชนิดใดซึ่งจะทำให้สามารถสร้างโปรโมชั่นที่สามารถดึงดูดลูกค้าได้สามารถนำไปวิเคราะห์ได้อีกว่าจะทำอย่างไรให้ลูกค้ามีความสนใจกับเมนูที่ไม่ได้รับความนิยม และยังสามารถนำไปวิเคราะห์เพื่อปรับสูตรหรือคิดค้นเมนูพิซซ่าใหม่ๆที่สามารถดึงดูดลูกค้าได้อีกด้วย

Data Pipeline

- ใช้ AWS สำหรับเก็บข้อมูลบน Cloud
 - ใช้ S3 สำหรับ Data lake เพื่อเก็บ RAW Data
 - ใช้ Redshift สำหรับ Data warehouse เพื่อเก็บข้อมูลที่มาจาก S3 โดยจะมีการทำความสะอาดข้อมูลมาแล้ว
- ใช้ DBT สำหรับสร้างตารางออกมาเพื่อวิเคราะห์เฉพาะบางส่วนที่สนใจ
- ใช้ Airflow สำหรับสร้าง Schedule เพื่อใช้ในการดึงข้อมูลไปที่ S3, สร้าง Tables ที่ Redshift และ Transform ข้อมูลลงเก็บที่ Redshift อัตโนมัติ



Data modelling (Data warehouse)

ตาราง pizzasale	
order_details_id	int
order_id	int
pizza_id	int
quantity	int
order_date	DATE
order_time	TIME
unit_price	float
total_price	float
pizza_size	text
pizza_category	text
pizza_ingredients	text
pizza_name	text

ตาราง select_pizza_num_of_pizzaname	
Pizzaname	text
Num_of_pizza	int

ตาราง select_total_num_by_hour	
hour	int
Num_of_pizza	int

ตาราง select_total_num_by_month	
month	int
Num_of_pizza	int

ตาราง select_num_of_pizzaname_by_month	
month	int
Pizzaname	text
Num_of_pizza	int

ตาราง select_num_of_pizzaname_by_hour	
hour	int
Pizzaname	text
Num_of_pizza	int

ขั้นตอนการดำเนินการ

1. สร้าง S3 Bucket ใน AWS โดยกำหนดชื่อ และทำการ uncheck ที่ Block all public access เพื่อให้สามารถอัปโหลดไฟล์จากที่อื่นเข้ามาได้

Amazon S3 > Buckets > Create bucket

Create bucket [Info](#)

Buckets are containers for data stored in S3. [Learn more](#)

General configuration

Bucket name

Bucket name must be globally unique and must not contain spaces or uppercase letters. [See rules for bucket naming](#)

AWS Region

US East (N. Virginia) us-east-1 ▼

Copy settings from existing bucket - *optional*
Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Block Public Access settings for this bucket

Public access is granted to buckets and objects through access control lists (ACLs), bucket policies, access point policies, or all. In order to ensure that public access to this bucket and its objects is blocked, turn on Block all public access. These settings apply only to this bucket and its access points. AWS recommends that you turn on Block all public access, but before applying any of these settings, ensure that your applications will work correctly without public access. If you require some level of public access to this bucket or objects within, you can customize the individual settings below to suit your specific storage use cases. [Learn more](#)

☐ **Block all public access**
Turning this setting on is the same as turning on all four settings below. Each of the following settings are independent of one another.

- ☒ **Block public access to buckets and objects granted through *new* access control lists (ACLs)**
S3 will block public access permissions applied to newly added buckets or objects, and prevent the creation of new public access ACLs for existing buckets and objects. This setting doesn't change any existing permissions that allow public access to S3 resources using ACLs.
- ☒ **Block public access to buckets and objects granted through *any* access control lists (ACLs)**
S3 will ignore all ACLs that grant public access to buckets and objects.
- ☒ **Block public access to buckets and objects granted through *new* public bucket or access point policies**
S3 will block new bucket and access point policies that grant public access to buckets and objects. This setting doesn't change any existing policies that allow public access to S3 resources.
- ☒ **Block public and cross-account access to buckets and objects through *any* public bucket or access point policies**
S3 will ignore public and cross-account access for buckets or access points with policies that grant public access to buckets and objects.

- สร้าง Cluster Redshift ใน AWS โดยกำหนดชื่อ Cluster และต้อง check Edit publicly accessible เพื่อให้สามารถเข้าถึงจากภายนอกได้

Cluster configuration

Cluster identifier
This is the unique key that identifies a cluster.

The identifier must be from 1-63 characters. Valid characters are a-z (lowercase only) and - (hyphen).

What are you planning to use this cluster for?

☒ **Production**
Configure for fast and consistent performance at the best price.

☐ **Free trial**
Configure for learning about Amazon Redshift. This configuration is free for a limited time if your organization has never created an Amazon Redshift cluster.

Choose the size of the cluster

☒ I'll choose

☐ Help me choose

Node type [Info](#)
Choose a node type that meets your CPU, RAM, storage capacity, and drive type requirements.

Number of nodes
Enter the number of nodes that you need.

Edit publicly accessible

Publicly accessible

☒ **Turn on Publicly accessible**
Allow public connections to Amazon Redshift.

⚠ When you turn on this feature, clients can connect to the database from outside the VPC.

Elastic IP address
Select the Elastic IP address for connecting to the cluster.

Cancel **Save changes**

- นำ Credential AWS ไปใช้ในการเขียนโปรแกรมเพื่อเข้าถึง AWS โดยสามารถได้รับตามภาพด้านล่างประกอบไปด้วย

- 1) Aws_access_key_id
- 2) Aws_secret_access_key
- 3) Aws_session_token

ALLv1-24728 > Modules > Learner Lab > Learner Lab

```

ddd_v1_w_odR_1383608@runweb69854:~$ cat ~/.aws/credentials
[default]
aws_access_key_id = 
aws_secret_access_key = FR
aws_session_token = FwoGZXIvYXZzEEQaDLBty3MVKWj70A1WMSLKAdy2p1K6TGuCvLh+K/qv1LpnAr9uKV1W5I+sBd2FjXTkvbTD+Lmsr/3uMn/Dn8v+
AhycNc+DF
SqIPpv92g
D5ZV3uJ0bRokU60j3ZzPKtKU/1w3K7ssh1jx2Q==
ddd_v1_w_odR_1383608@runweb69854:~$

```

- หลังจากรันโปรแกรมจะต้อง Set connection ให้กับ Airflow เพื่อให้สามารถเข้าใช้งาน Redshift ได้ โดย Set ค่าดังภาพด้านล่าง

Connection Id * redshift

Connection Type * Postgres
Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.

Description

Host pizzasalecius.orgjtkftimj.us-east-1.redshift.amazonaws.com

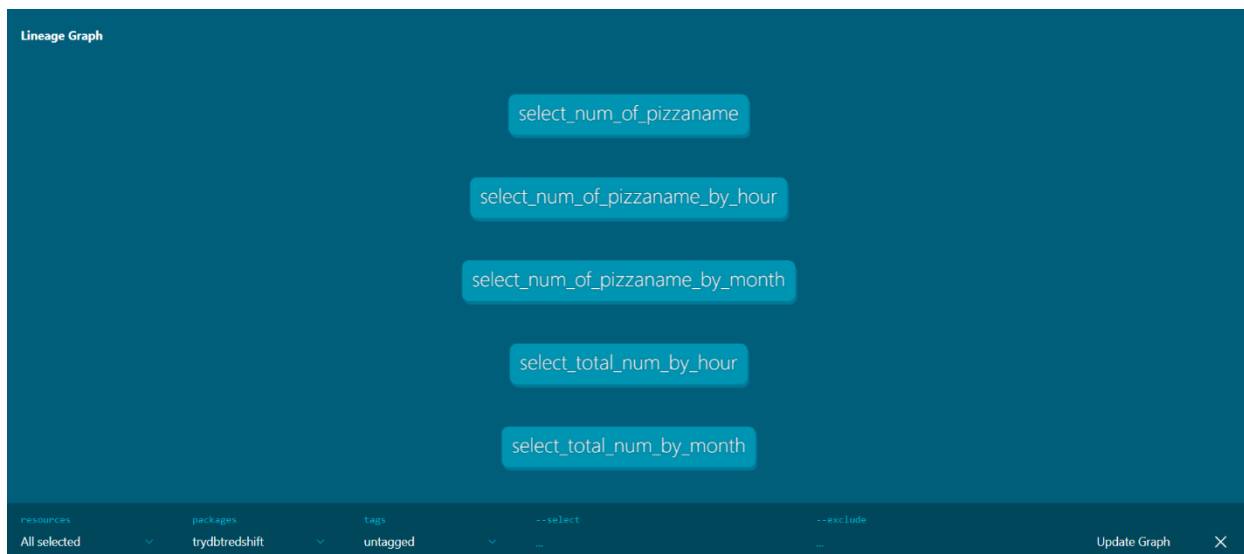
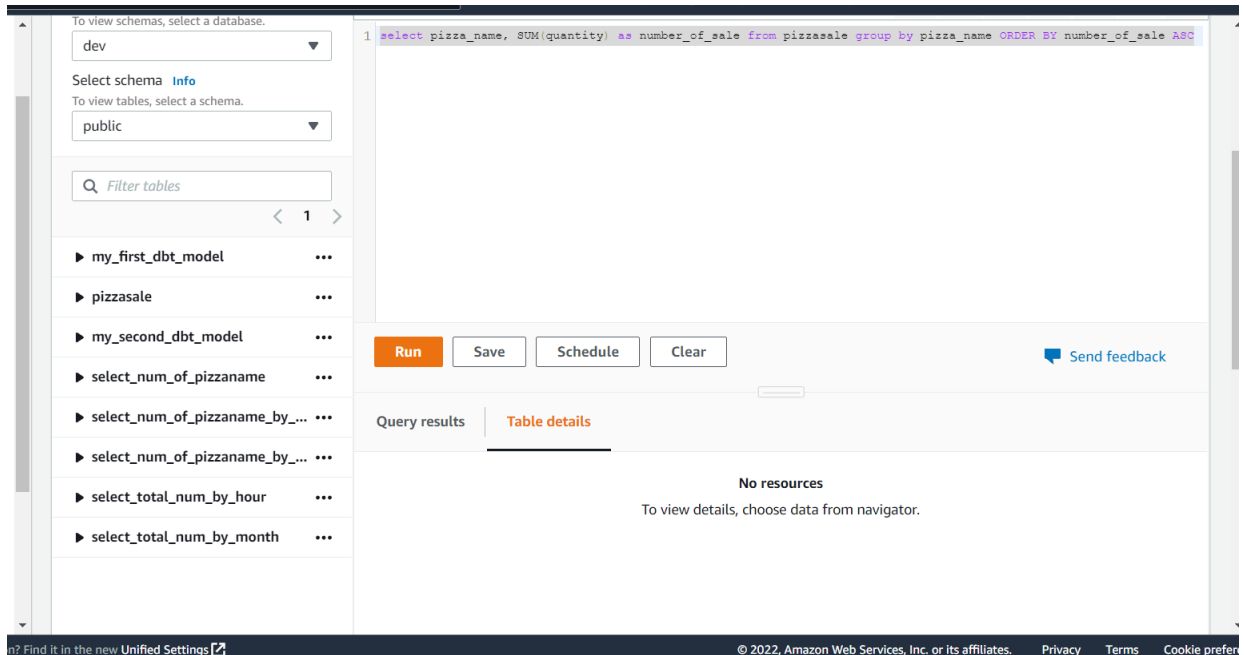
Schema dev

Login awsuser

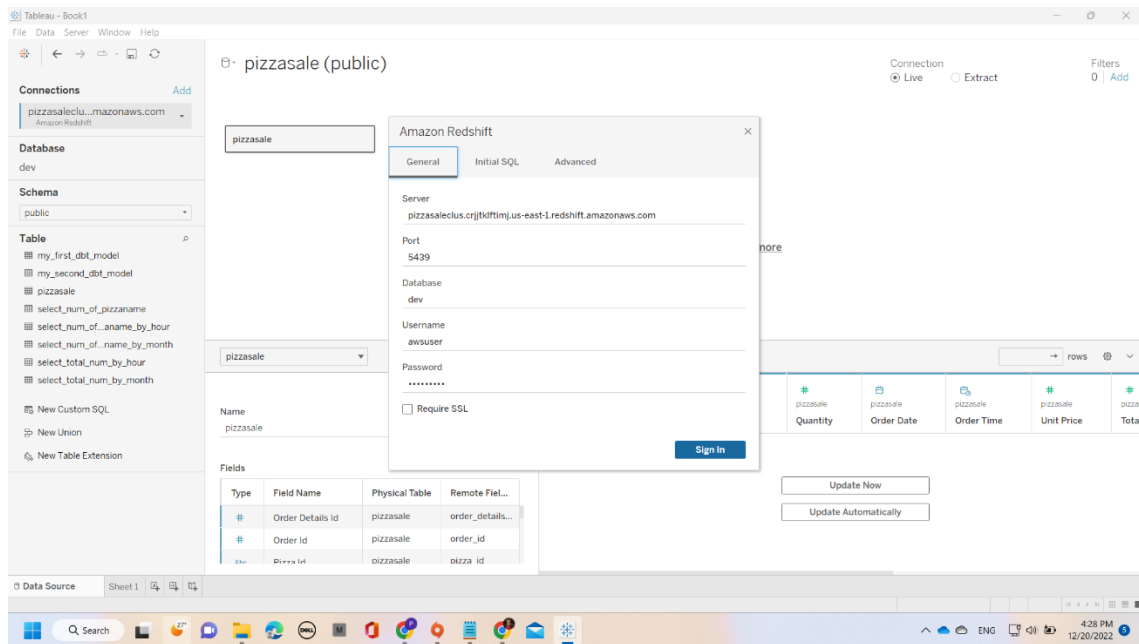
Password

Port 5439

5. จากนั้นจะได้ไฟล์ RAW Data อยู่ใน S3 และได้ Table พร้อมข้อมูลทำความสะอาดเรียบร้อยแล้วอยู่ใน Redshift
6. จากนั้นใช้ DBT ในการสร้างตารางสำหรับนำไปวิเคราะห์เพิ่มเติม



7. นำข้อมูลใน Redshift ไปสร้างเป็น Dashboard ด้วย Tableau โดยมีการกำหนดค่าเพื่อดึงข้อมูลดังภาพด้านล่าง



Data visualization

