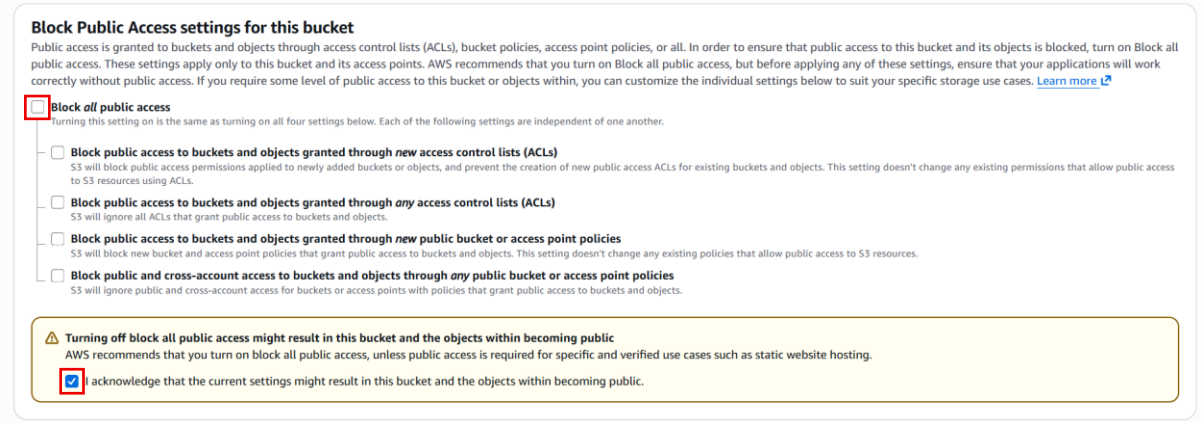
Load Data from S3

To begin with the Lab

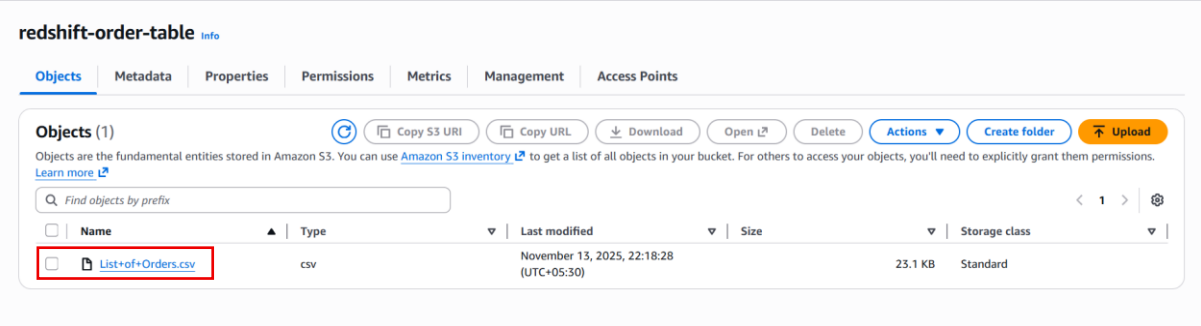
Summary of the Lab

In this lab, you learn how to **load data from Amazon S3 into Amazon Redshift**. First, an S3 bucket is created, and a CSV file (e.g., orders.csv) is uploaded. Then, a target table is created in Redshift using the **CREATE TABLE** command, and data is imported using the **COPY** command, specifying the S3 path, IAM role, region, and delimiter options. Alternatively, Redshift's graphical **Load Data wizard** can be used to import data without SQL by selecting the S3 bucket, configuring file settings, and assigning roles. The data successfully loads into the Redshift table.

- Setting Up an S3 Bucket
- A new S3 bucket is created to store the data that will be loaded into Redshift.
- Default settings are used for simplicity.
- Public access restrictions are disabled for this demonstration.



- After creation, the bucket appears under the specified region (e.g., US East 1).
- A CSV file (such as orders.csv) is uploaded to the bucket.



- Loading Data into Redshift
- Data loading is performed using the COPY command, a standard SQL command in data warehouses.
- The COPY command copies data from S3 into a Redshift table.

- Before executing the COPY command, a target table must exist in the database.
- The command structure includes the table name, S3 file path, IAM role, region, delimiter, and header options.

```
COPY public.orders
FROM 's3://redshift-order-table/List+of+Orders.csv'
IAM_ROLE 'arn:aws:iam::463646775279:role/service-role/AmazonRedshift-CommandsAccessRole-20251113T225531'
DELIMITER ','
IGNOREHEADER 1
REGION 'us-east-1';
```


COPY public.orders

FROM 's3://redshift-order-table/List+of+Orders.csv'

IAM_ROLE 'arn:aws:iam::463646775279:role/service-role/AmazonRedshift-CommandsAccessRole-20251113T225531'

DELIMITER ','

IGNOREHEADER 1

REGION 'us-east-1';

- Creating the Target Table
- A table must be created before using the COPY command.
- The CREATE TABLE command is used to define the table structure, specifying columns and data types.
- The table is created under the default “public” schema if no other schema is specified.
- The new table can be verified in the Redshift Query Editor’s schema view.

```
CREATE TABLE orders (
    order_id VARCHAR(20),
    order_date VARCHAR(20),
    customer_name VARCHAR(100),
    state VARCHAR(50),
    city VARCHAR(50)
);
```


CREATE TABLE orders (

order_id VARCHAR(20),

order_date VARCHAR(20),

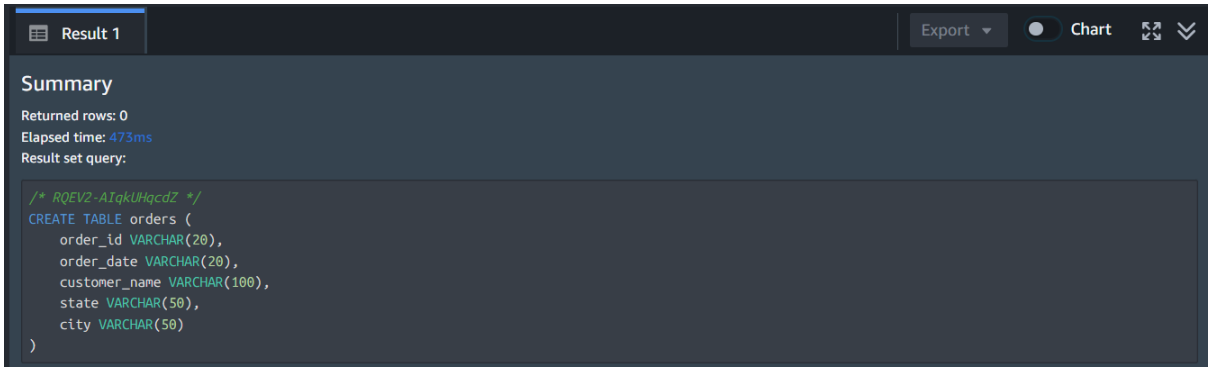
customer_name VARCHAR(100),

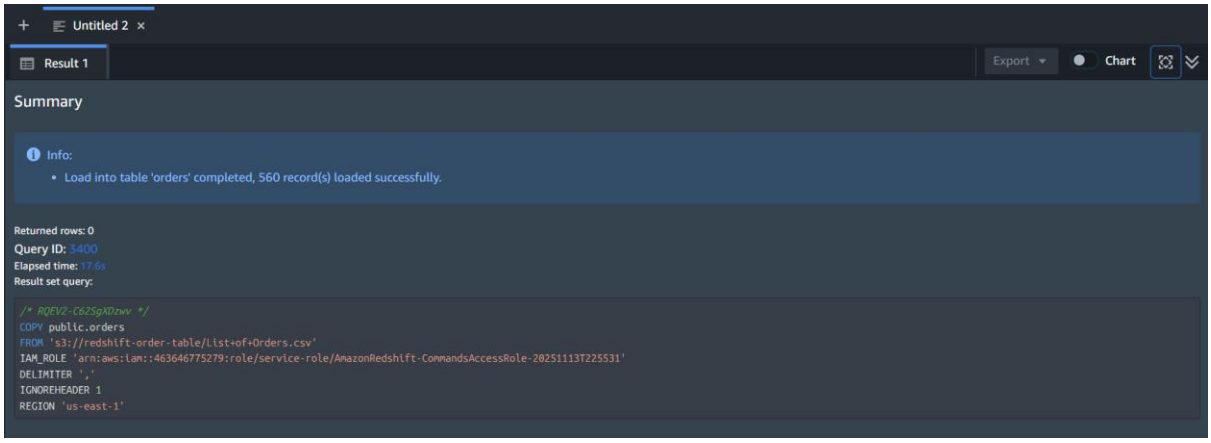
state VARCHAR(50),

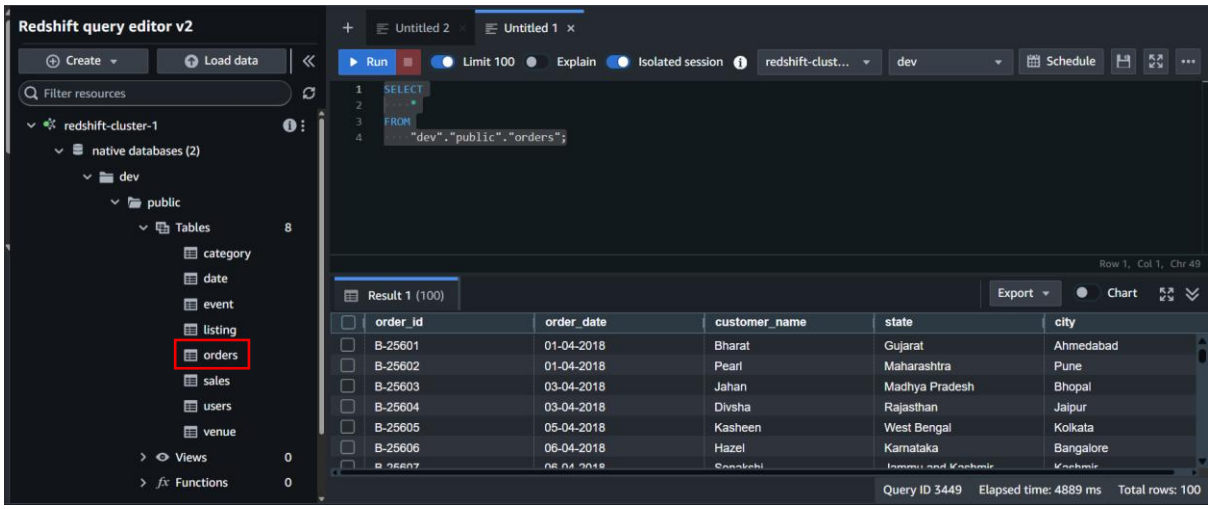
city VARCHAR(50)

);

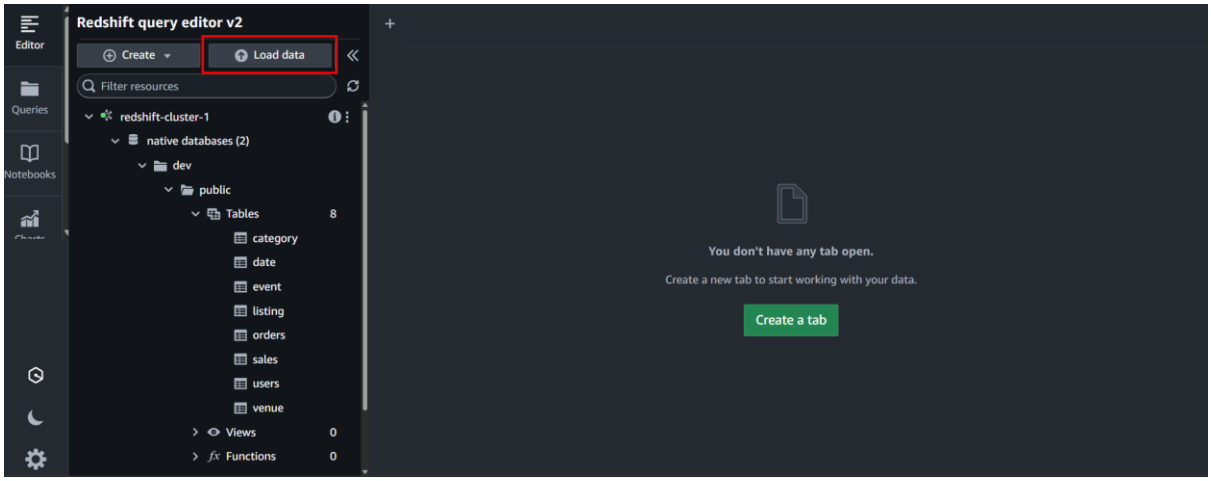
- Here we created the table first and then we have updated the data from s3 to table.
- We used the above commands to create table and load the data from the s3 bucket



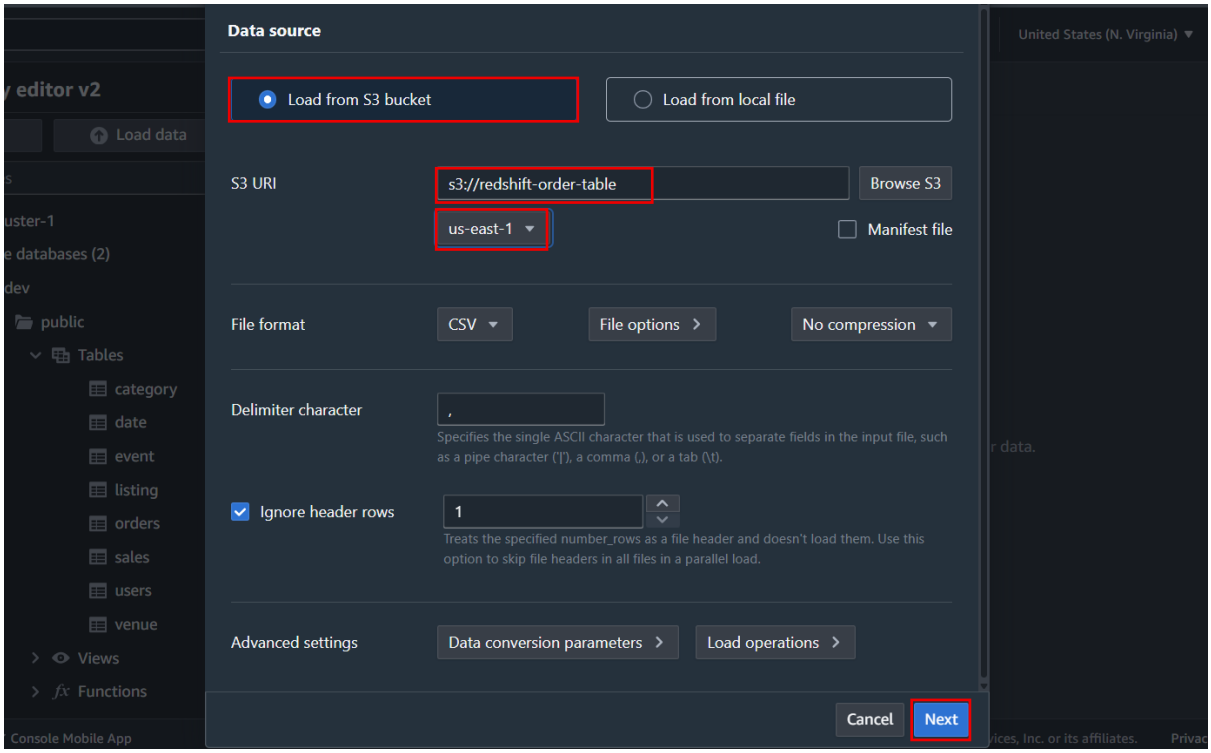




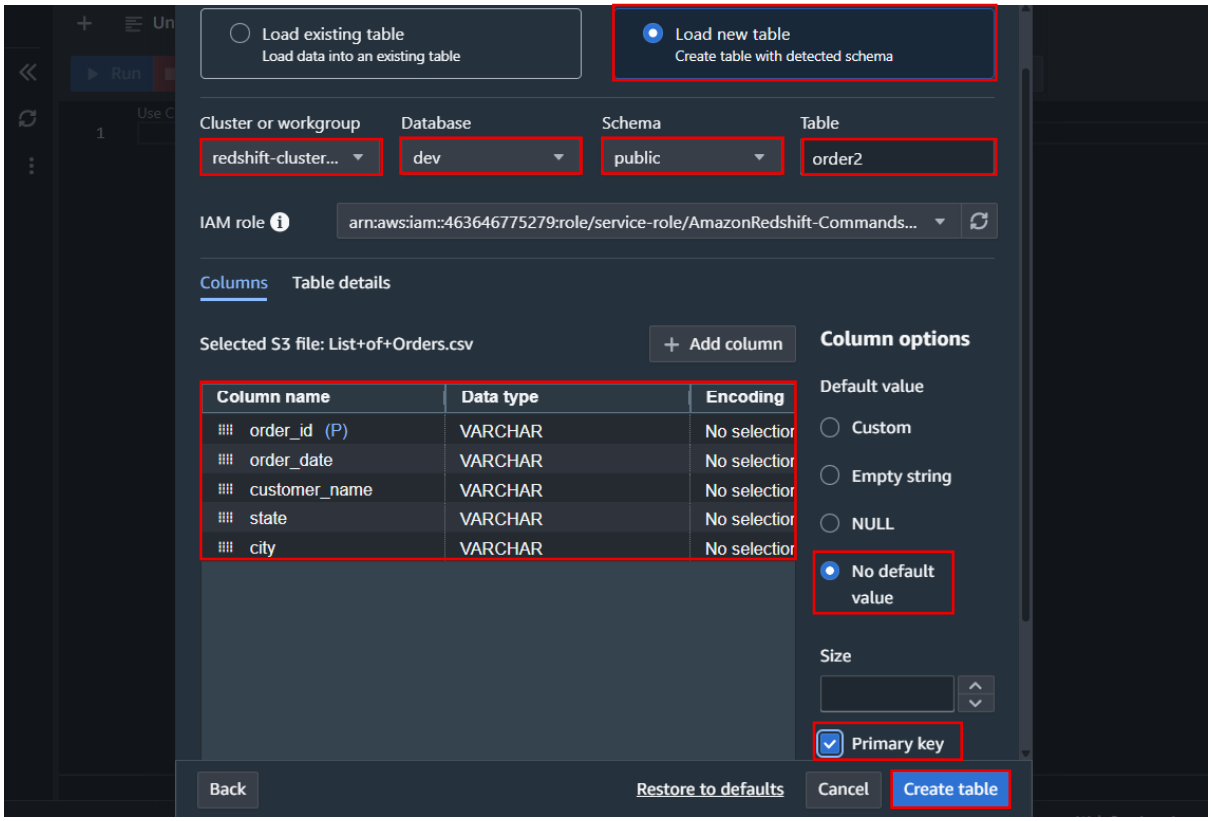
- Alternative Method – Using the Redshift Interface
- Redshift provides a graphical method to load data using a built-in wizard.
- The “Load Data” option in Query Editor allows users to import data without writing SQL.

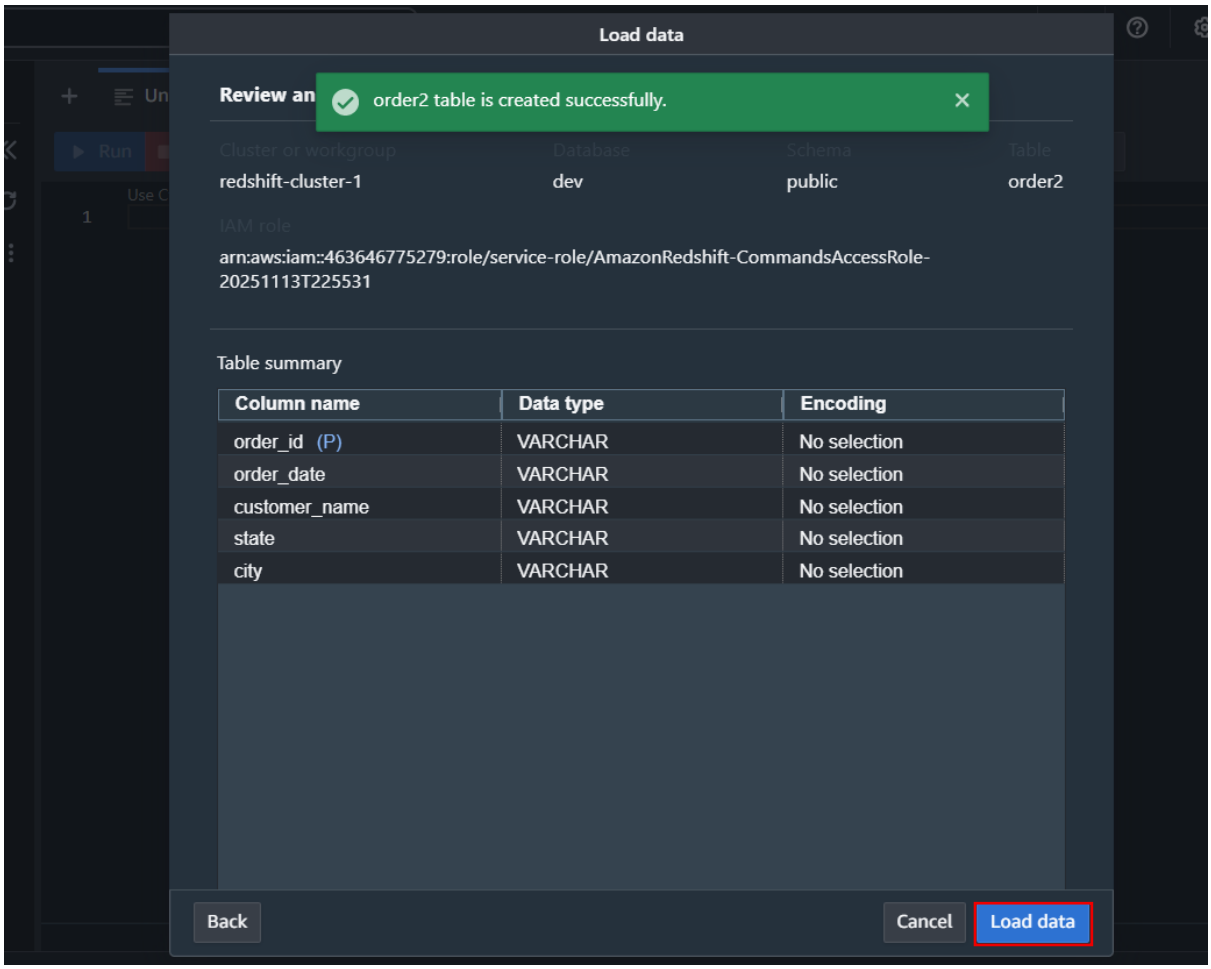


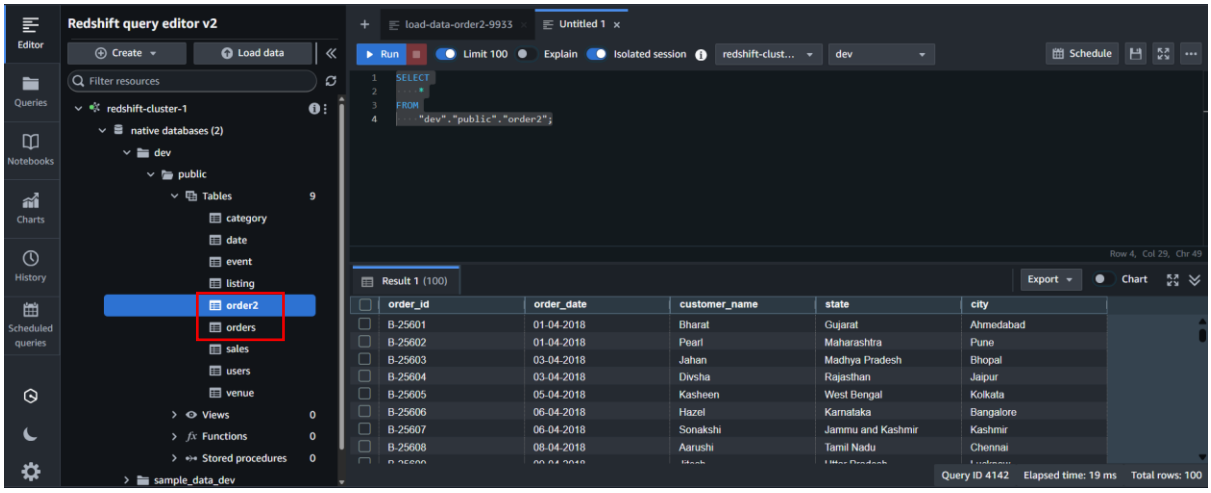
- Selecting “Load data from S3.”
- Choosing the S3 bucket and region.
- Configuring file options like delimiter and headers.



- Selecting whether to load into an existing table or create a new one.
- Defining column names, data types, and constraints (e.g., primary key or not null).
- Selecting the IAM role for access permissions.
- After setup, Redshift automatically generates and executes the COPY command.


Load existing table
Load data into an existing table
Load new table
Create table with detected schema
Cluster or workgroup
redshift-cluster...
Database
dev
Schema
public
Table
order2
IAM role
arn:aws:iam::463646775279:role/service-role/AmazonRedshift-Commands...
Columns
Table details
Selected S3 file: List+of+Orders.csv
Add column
Column name
Data type
Encoding
order_id (P)
VARCHAR
No selection
order_date
VARCHAR
No selection
customer_name
VARCHAR
No selection
state
VARCHAR
No selection
city
VARCHAR
No selection
Column options
Default value
Custom
Empty string
NULL
No default value
Size
Primary key
Back
Restore to defaults
Cancel
Create table


Load data
Review an
order2 table is created successfully.
Cluster or workgroup
redshift-cluster-1
Database
dev
Schema
public
Table
order2
IAM role
arn:aws:iam::463646775279:role/service-role/AmazonRedshift-CommandsAccessRole-20251113T225531
Table summary
Column name
Data type
Encoding
order_id (P)
VARCHAR
No selection
order_date
VARCHAR
No selection
customer_name
VARCHAR
No selection
state
VARCHAR
No selection
city
VARCHAR
No selection
Back
Cancel
Load data



- You can see the table order2 with the data uploaded from the s3 bucket.