

## Assignment6: E-Commerce Data Pipeline

- 1)ran python codes from local in terminal for mock data generator
- 2)uploaded to s3 the data using uploadtos3 function using terminal
- 3)created redshift tables from the below commands

```
CREATE TABLE ecommerce_data.dim_customers (  
    customer_id VARCHAR(20) ENCODE lzo,  
    first_name VARCHAR(255) ENCODE lzo,  
    last_name VARCHAR(255) ENCODE lzo,  
    email VARCHAR(255) ENCODE lzo,  
    membership_level VARCHAR(255) ENCODE lzo  
)  
DISTSTYLE KEY  
DISTKEY (membership_level)  
SORTKEY (customer_id , first_name, last_name);  
  
INSERT INTO ecommerce_data.dim_customers (customer_id, first_name, last_name,  
email, membership_level)  
VALUES  
( 'CUST00001', 'John', 'Doe', 'john.doe@example.com', 'BR'),  
( 'CUST00002', 'Jane', 'Smith', 'jane.smith@example.com', 'SL'),  
( 'CUST00003', 'Michael', 'Lee', 'michael.lee@example.com', 'GL'),  
( 'CUST00004', 'Alice', 'Broun', 'alice.brown@example.com', 'BR'),  
( 'CUST00005', 'David', 'Miller', 'david.miller@example.com', 'SL');  
  
CREATE TABLE ecommerce_data.dim_products (  
    product_id VARCHAR(20) ENCODE lzo,  
    product_name VARCHAR(255) ENCODE lzo,  
    category VARCHAR(100) ENCODE lzo,  
    price DECIMAL(10,2) ENCODE delta,  
    supplier_id VARCHAR(20) ENCODE lzo  
)  
DISTSTYLE KEY  
DISTKEY (category)  
SORTKEY (product_id , product_name);  
  
INSERT INTO ecommerce_data.dim_products (product_id, product_name, category,  
price, supplier_id)  
VALUES  
( 'PROD00001', 'Laptop', 'Electronics', 799.99, 'SUP00001'),  
( 'PROD00002', 'Smartphone', 'Electronics', 499.99, 'SUP00002'),  
( 'PROD00003', 'Headphones', 'Electronics', 79.99, 'SUP00001'),  
( 'PROD00004', 'T-Shirt', 'Clothing', 19.99, 'SUP00003'),  
( 'PROD00005', 'Jeans', 'Clothing', 49.99, 'SUP00003'),  
( 'PROD00006', 'Coffee Maker', 'Home Goods', 99.99, 'SUP00002'),
```

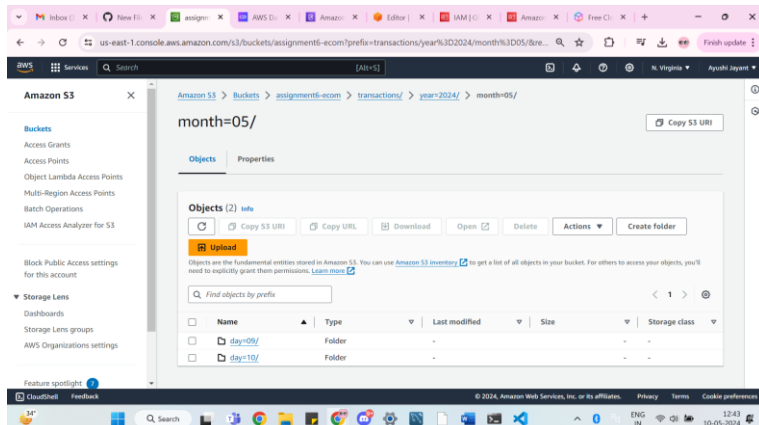
```
( 'PROD0007', 'Microwave', 'Home Goods', 149.99, 'SUP00001'),
( 'PROD0008', 'Vacuum Cleaner', 'Home Goods', 199.99, 'SUP00002'),
( 'PROD0009', 'Book', 'Other', 14.99, 'SUP00003'),
( 'PROD00010', 'Movie', 'Other', 19.99, 'SUP00002');
```

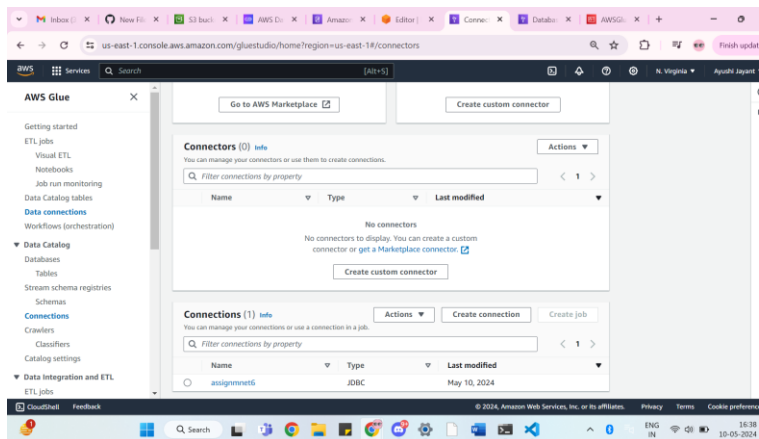
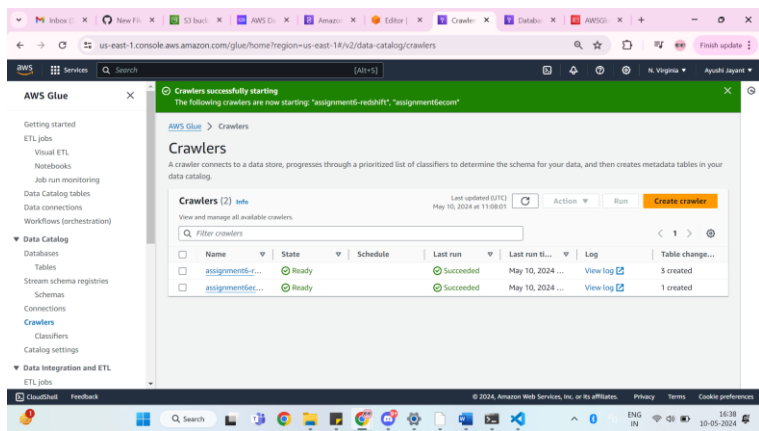
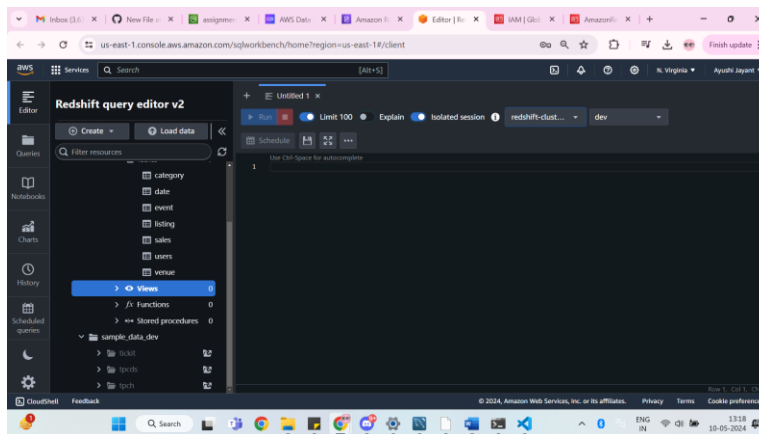
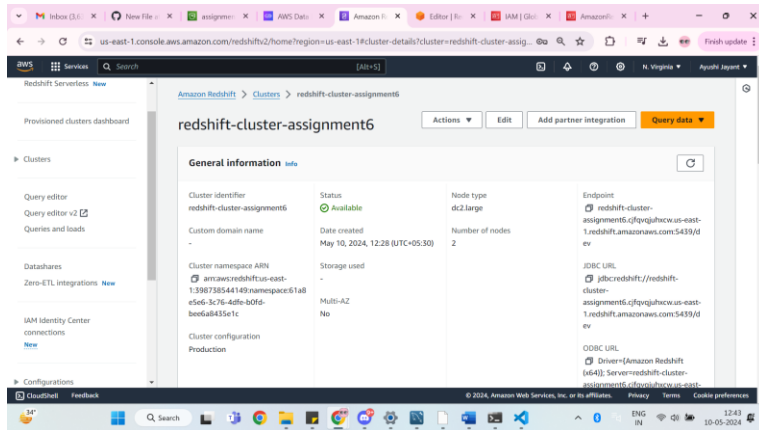
```
CREATE TABLE ecommerce_data.fact_transactions (
  transaction_id VARCHAR(20) ENCODE lzo,
  customer_id VARCHAR(20) ENCODE lzo,
  customer_email VARCHAR(100) ENCODE lzo,
  product_id VARCHAR(20) ENCODE lzo,
  product_name VARCHAR(255) ENCODE lzo,
  quantity INT ENCODE LZ0,
  price DECIMAL(10,2) ENCODE DELTA,
  supplier_id VARCHAR(20) ENCODE lzo,
  transaction_date DATE ENCODE BYTEDICT,
  transaction_type VARCHAR(20) ENCODE LZ0,
  payment_type VARCHAR(20) ENCODE LZ0,
  status VARCHAR(20) ENCODE lzo
)
```

DISTSTYLE KEY

DISTKEY(transaction\_type)

SORTKEY(transaction\_id);





us-east-1 console.aws.amazon.com/glue/home?region=us-east-14/v2/data-catalog/tables

### AWS Glue

You can now enable auto-compaction for managed tables in Glue Data Catalog. To learn more, visit the documentation.

#### Tables

A table is the metadata definition that represents your data, including its schema. A table can be used as a source or target in a job definition.

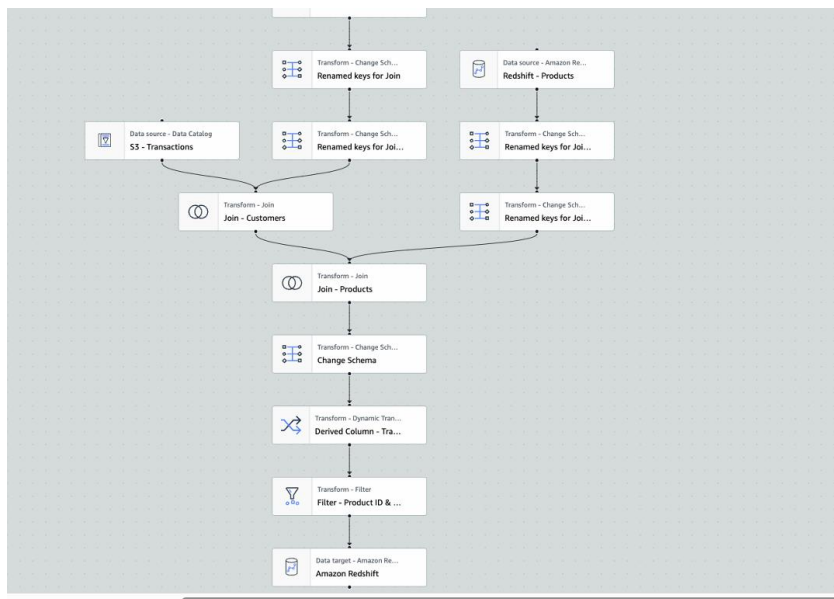
Tables (4)

Last updated ETLCT May 10, 2024 at 11:08:38

View and manage all available tables.

Filter tables

<input type="checkbox"/>	Name	Database	Location	Classification	Deprecated	View data	Data quality
<input type="checkbox"/>	assignment6_eom	assignment6	s3://assignment6-e...	CSV	-	Table data	View data quality
<input type="checkbox"/>	dev_commerce_da	asos-redshift	devcommerce_dat	redshift	-	-	View data quality
<input type="checkbox"/>	dev_commerce_da	asos-redshift	devcommerce_dat	redshift	-	-	View data quality
<input type="checkbox"/>	dev_commerce_da	asos-redshift	devcommerce_dat	redshift	-	-	View data quality



Run Limit 100

```
1 select * from ecommerce_data.fact_transactions;
```

Row 1, Col 48, Chr 47

Result 1 (50)

	transaction_id	customer_id	customer_email	product_id	product_name	price
<input type="checkbox"/>	TXN45630831	CUST00004	alice.brown@example.com	PROD00001	Laptop	1599.98
<input type="checkbox"/>	TXN45431435	CUST00004	alice.brown@example.com	PROD00004	T-Shirt	79.96
<input type="checkbox"/>	TXN51236145	CUST00001	john.doe@example.com	PROD00002	Smartphone	1999.96
<input type="checkbox"/>	TXN28945659	CUST00002	jane.smith@example.com	PROD00001	Laptop	799.99
<input type="checkbox"/>	TXN92488593	CUST00003	michael.lee@example.com	PROD00004	T-Shirt	79.96
<input type="checkbox"/>	TXN81697662	CUST00002	jane.smith@example.com	PROD00003	Headphones	239.97
<input type="checkbox"/>	TXN28809082	CUST00001	john.doe@example.com	PROD00001	Laptop	1599.98
<input type="checkbox"/>	TXN87866731	CUST00002	jane.smith@example.com	PROD00004	T-Shirt	59.97
<input type="checkbox"/>	TXN1728098	CUST00002	jane.smith@example.com	PROD00005	Lease	99.98

Elapsed time: 34 ms Total rows: 50