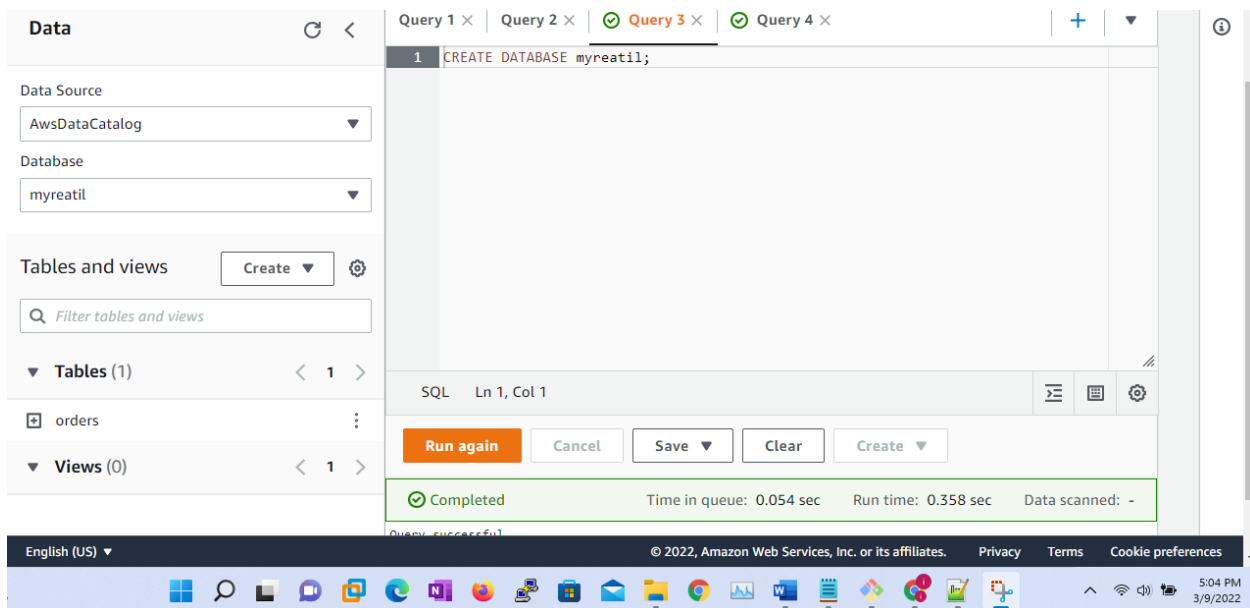


# Create Database and Tables using Athena

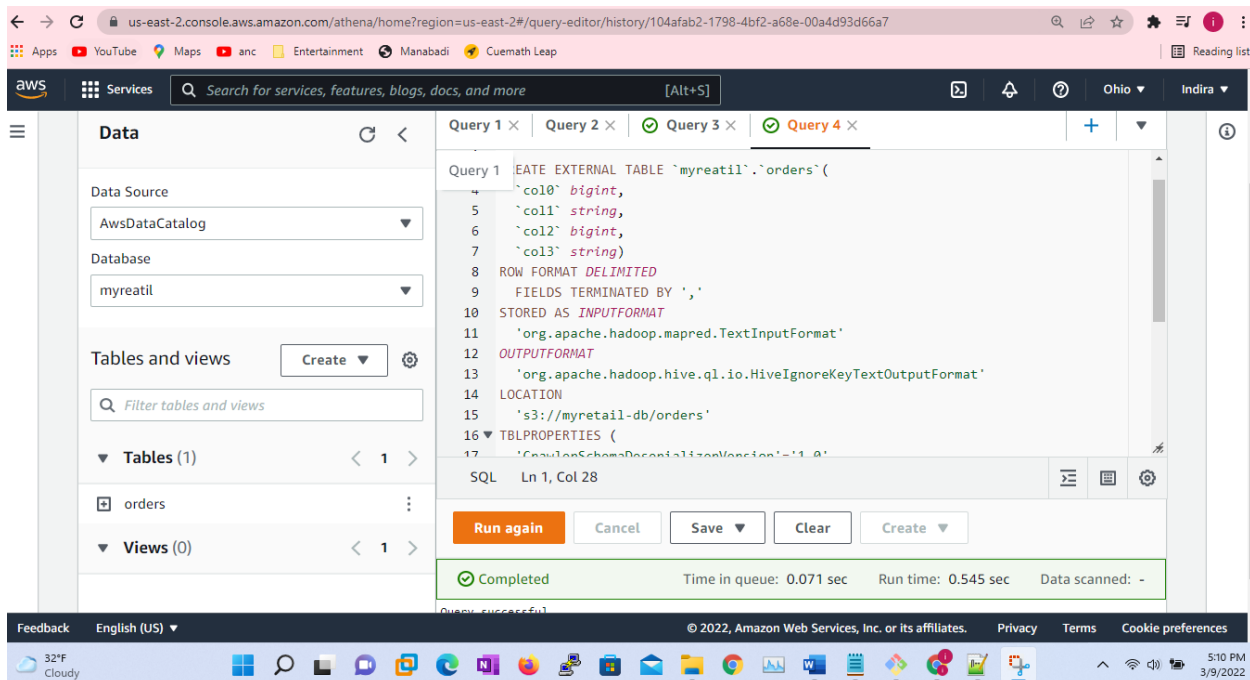
Using the Athena created the Database name 'myretail'



Database is successfully created. Created the tables under database "myretail"

Select Data Source: AWSDataCatalog

DataBase: myretail

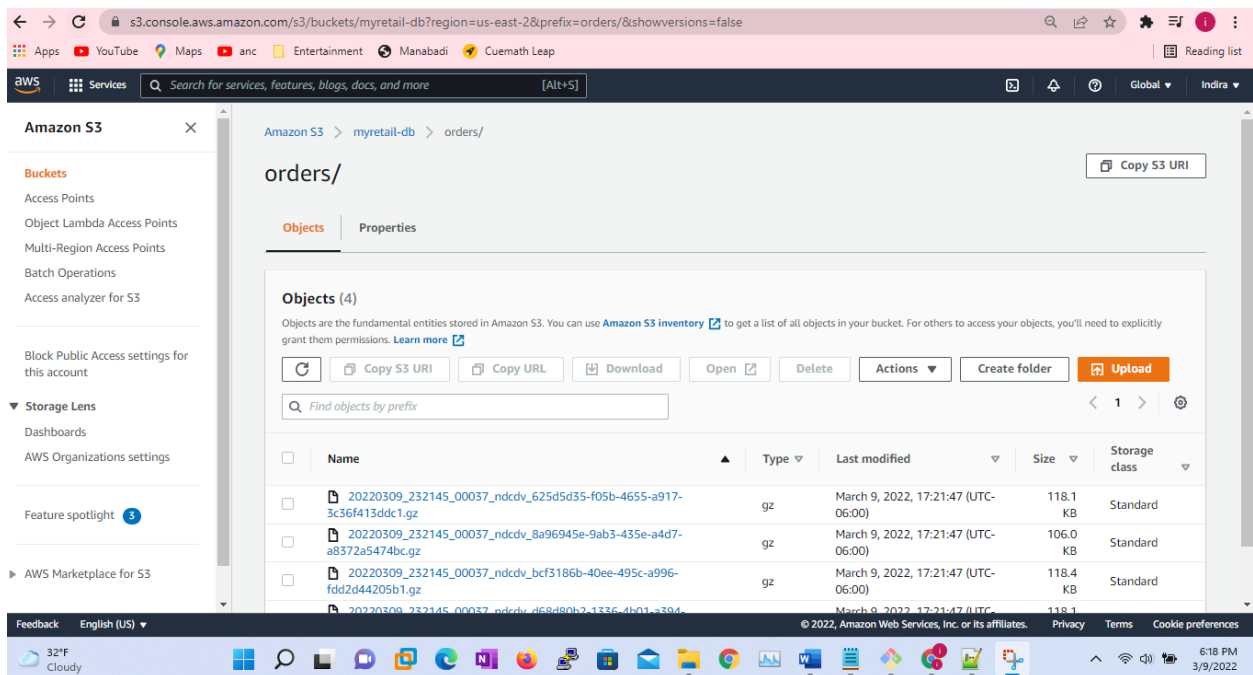


Insert data into the tables by using the commands

```
Query 1 × | Query 2 × | ✓ Query 3 × | ✗ Query 4 ×

1 CREATE DATABASE myreatil;
2
3
4 INSERT INTO myreatil.orders
5 SELECT * FROM retail_db.orders
6
7 SELECT * FROM myreatil.orders LIMIT 10
8
9 SELECT COUNT(*) FROM myreatil.orders
```

Verified that data is inserted into the s3 bucket



## Using CTAS (CREATE TABLE AS SELECTED) to create table using Athena

Syntax: CREATE TABLE myreatil.order\_items

AS

SELECT \* FROM retail\_db.order\_items

SELECT count(\*) FROM myreatil.order\_items

SELECT count(\*) FROM myreatil.order\_items LIMIT 10

When it comes to CTAS to create table to stage the data into some location and then download it and take it further. By default it uses parquet file format to different file format.

```
CREATE TABLE myreatil.order_items
WITH(
  format = 'TEXTFILE',
  external_location = 's3://myretail-db/orders_items/',
  field_delimiter = ','
)
AS
SELECT * FROM retail_db.order_items
```

The screenshot shows the AWS S3 console interface. The main content area displays the 'myretail-db' bucket. Under the 'Objects' tab, there are two objects listed:

	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	orders_items/	Folder	-	-	-
<input type="checkbox"/>	orders/	Folder	-	-	-

The console also shows a sidebar with navigation options and a top bar with the AWS logo and search bar. The bottom status bar indicates the date and time as 11:13 PM on 3/9/2022.

## Created partitioned table using Athena

```
1 CREATE TABLE myreatil.orders_part(  
2     order_id INT,  
3     order_customer_id STRING,  
4     order_date STRING,  
5     order_status STRING  
6 ) PARTITIONED BY ( order_month INT )  
7     STORED AS parquet  
8     LOCATION 's3://myretail-db/orders_items/'  
9  
10 SELECT count(*) FROM myreatil_orders
```

us-east-2.console.aws.amazon.com/glue/home?region=us-east-2#table:catalog=083798296286:name=orders\_part:namespace=myreatil

Services Search for services, features, blogs, docs, and more [Alt+S]

**AWS Glue**

**Data catalog**

- Databases
- Tables
- Connections
- Crawlers
- Classifiers
- Schema registries
- Schemas
- Settings

**ETL**

- AWS Glue Studio
- Jobs - New
- Jobs (legacy)
- ML Transforms

**Input format** org.apache.hadoop.hive ql.io.parquet.MapredParquetInputFormat  
**Output format** org.apache.hadoop.hive ql.io.parquet.MapredParquetOutputFormat  
**Serde serialization lib** org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe

**Serde parameters** serialization format 1

**Table properties** EXTERNAL TRUE transient\_lastDdlTime 1646893049

**Schema** Showing: 1 - 5 of 5

	Column name	Data type	Partition key	Comment
1	order_id	int		
2	order_custom...	string		
3	order_date	string		
4	order_status	string		
5	order_month	int	Partition (0)	

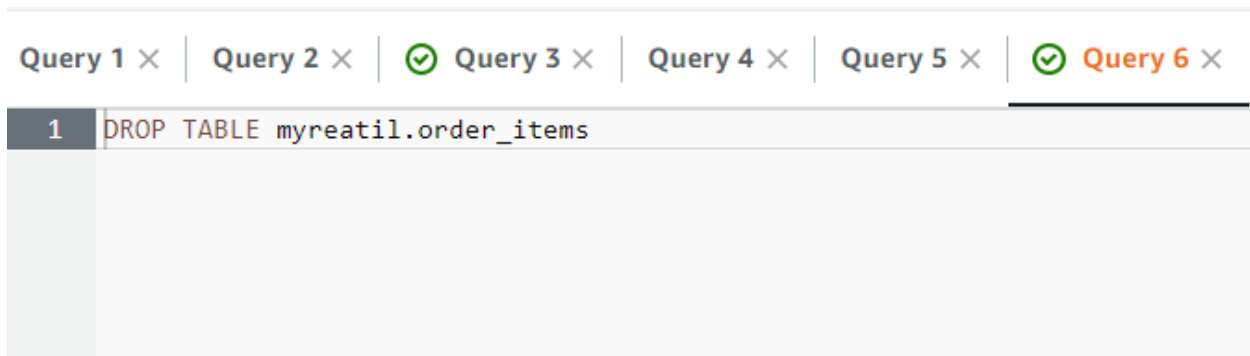
Feedback English (US) © 2022, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

23°F Snow 11:33 AM 3/10/2022

Figure 1Created the table Order\_part with Order\_month as partition

## Drop the Table in Athena

To drop the table, we use DROP TABLE command:



To clean up the data use below command through AWS CLI

```
aws s3 rm s3://myretail-db/order_items --recursive
```

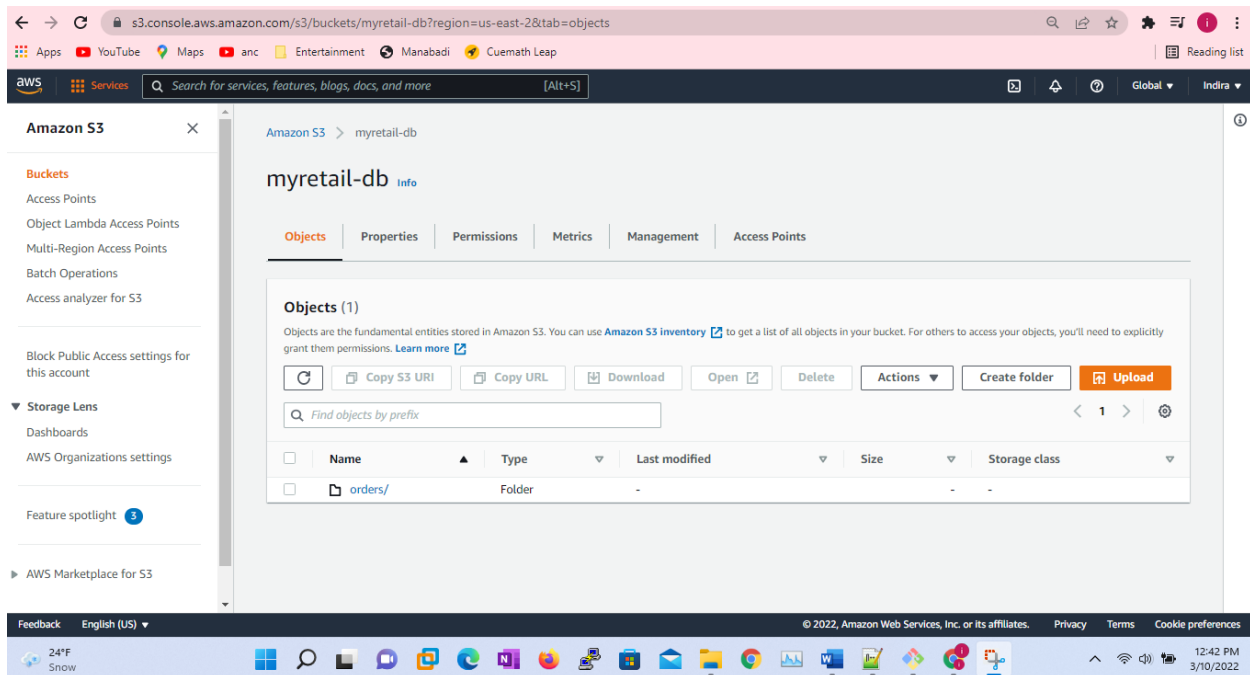


Figure 2 Deleted the order\_item in s3 bucket

## Drop Partition table using Athena

Command: DROP TABLE myreatil.orders\_part

Cleaning up the data in s3: `s3://myretail-db/order_part --recursive`