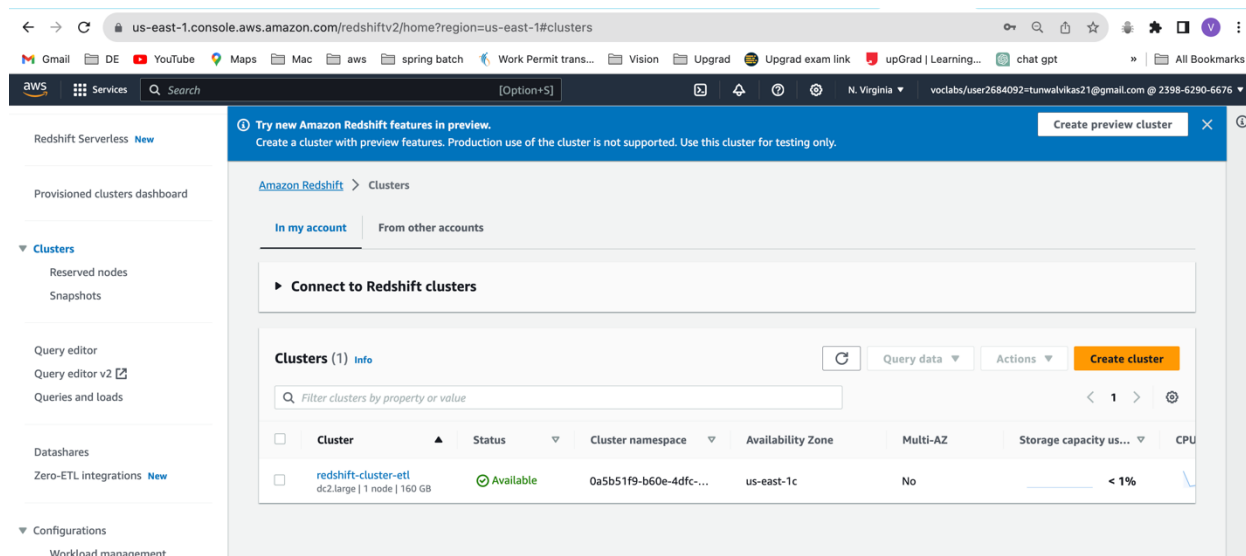


Redshift Cluster setup

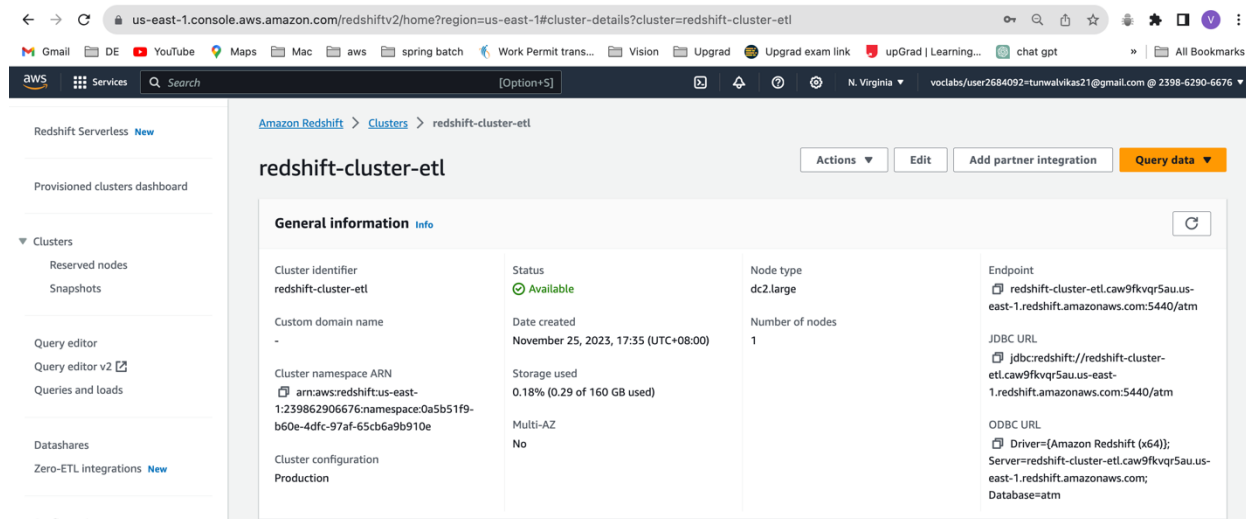
Redshift Cluster created - redshift-cluster-etl



The screenshot shows the AWS Redshift console interface. On the left, there's a navigation menu with options like 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Databases', and 'Configurations'. The main content area displays a list of clusters. A table shows the details of the cluster 'redshift-cluster-etl'.

Cluster	Status	Cluster namespace	Availability Zone	Multi-AZ	Storage capacity us...	CPU
redshift-cluster-etl dc2.large 1 node 160 GB	Available	0a5b51f9-b60e-4dfc-...	us-east-1c	No	< 1%	

Cluster General Information about Endpoint, ARN, JDBC and ODBC URL etc.



The screenshot shows the 'General information' page for the cluster 'redshift-cluster-etl'. It provides detailed information about the cluster's configuration and status.

Cluster identifier	Status	Node type	Endpoint
redshift-cluster-etl	Available	dc2.large	redshift-cluster-etl.caw9fkvqr5au.us-east-1.redshift.amazonaws.com:5440/atm
Custom domain name	Date created	Number of nodes	JDBC URL
-	November 25, 2023, 17:35 (UTC+08:00)	1	jdbc:redshift://redshift-cluster-etl.caw9fkvqr5au.us-east-1.redshift.amazonaws.com:5440/atm
Cluster namespace ARN	Storage used	Multi-AZ	ODBC URL
arn:aws:redshift:us-east-1:239862906676:namespace:0a5b51f9-b60e-4dfc-97af-65cb6a9b910e	0.18% (0.29 of 160 GB used)	No	Driver=(Amazon Redshift (x64)); Server=redshift-cluster-etl.caw9fkvqr5au.us-east-1.redshift.amazonaws.com; Database=atm
Cluster configuration			
Production			

Cluster performance Metrics

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#cluster-details?cluster=redshift-cluster-etl

Services Search [Option+S]

Redshift Serverless **New**

Provisioned clusters dashboard

▼ Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Zero-ETL integrations **New**

▼ Configurations

- Workload management
- Subnet groups
- HSM
- Manage Tags

▼ AWS Partner Integration

Informatica Data Loader

Cluster performance Query monitoring Zero-ETL integrations Resource Policy Schedules Maintenance Properties

► **Recommendations (0)**

To improve performance and decrease operating costs, the Amazon Redshift Advisor provides recommendations.

► **Alarms (0)**

CloudWatch alarms are triggered when a metric threshold is met.

► **Events (1)**

Amazon Redshift tracks events that occur on your cluster.

Cluster metrics (11/11)


Search Cluster metrics

Last hour Data for every 5 minutes Average

< 1 2 >

CPU utilization


The percentage of CPU utilization.



■ Shared

Percentage disk space used

The percent of disk space used.



■ Shared

us-east-1.console.aws.amazon.com/redshiftv2/home?region=us-east-1#cluster-details?cluster=redshift-cluster-etl

Services Search [Option+S]

Redshift Serverless **New**

Provisioned clusters dashboard

▼ Clusters

- Reserved nodes
- Snapshots

Query editor

Query editor v2

Queries and loads

Datashares

Zero-ETL integrations **New**

▼ Configurations

- Workload management
- Subnet groups
- HSM
- Manage Tags

▼ AWS Partner Integration

Informatica Data Loader

Cluster metrics (11/11)

Search Cluster metrics

Last hour Data for every 5 minutes Average

< 1 2 >

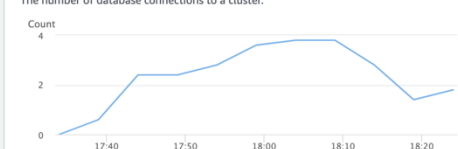
Auto vacuum space freed

Space reclaimed by auto vacuum in all tables.

No data available.

Database connections

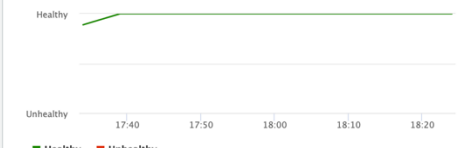
The number of database connections to a cluster.



■ Cluster

Health status

Indicates the health of the cluster.



■ Healthy ■ Unhealthy

Query duration

The average amount of time to complete a query.

No data available.

Query History and listings

Query runtime

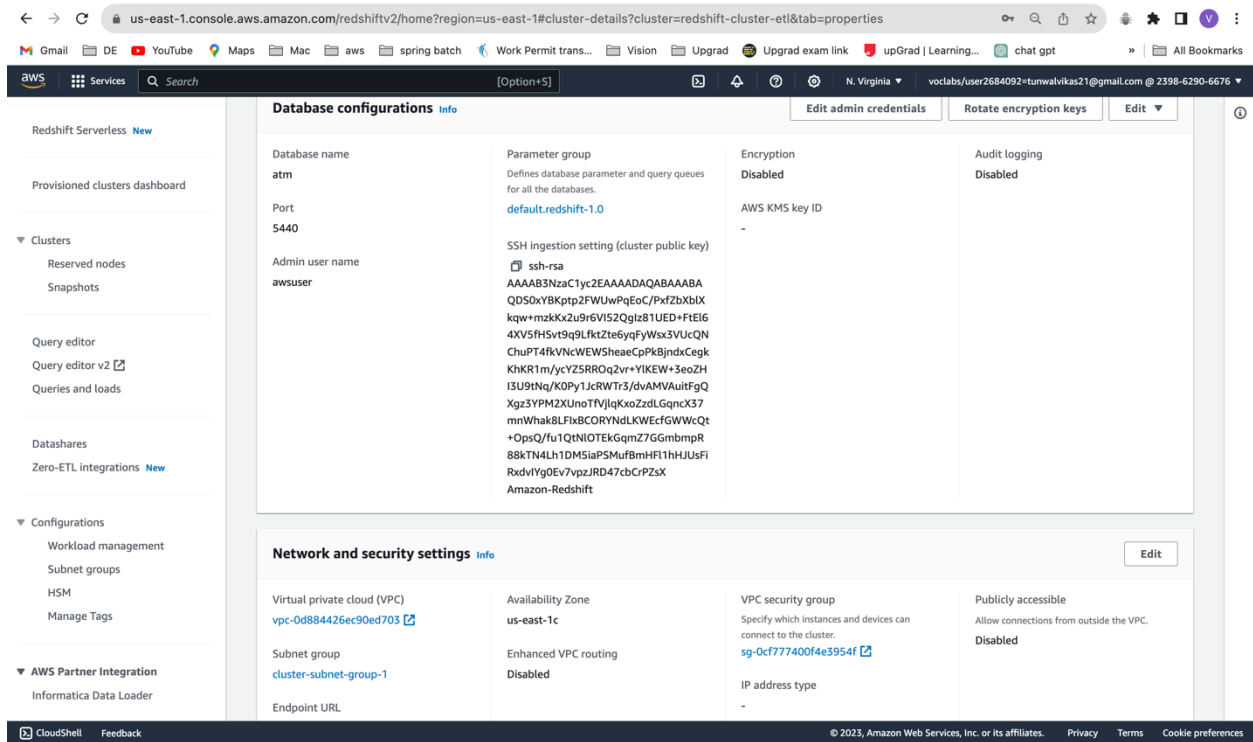
The query activity on a timeline. Use this graph to see which queries are running in the same timeframe. Choose a query to view more query execution details.

Queries and loads (85)

Filter queries

<input type="checkbox"/>	Start time	Query	Status	Duration	SQL
<input type="checkbox"/>	Nov 25th, 2023 05:57:59 PM 33 minutes ago	1973	Completed	19 sec	-- Copying the data to fact_atm_trans table copy atm_data.fact_atm_trans from 's3://atm-bucket-s3/fact_atm_trans/part-00000-02e7af69-eaa2-429d-9faf-eb4 ...
<input type="checkbox"/>	Nov 25th, 2023 06:26:16 PM 5 minutes ago	2482	Completed	19 sec	SELECT a.atm_number, a.atm_manufacturer, l.location, d.weekday, COUNT(trans_id) AS total_transaction_count FROM atm_data.fact_atm_trans f inner JOIN at ...
<input type="checkbox"/>	Nov 25th, 2023 06:21:34 PM 10 minutes ago	2405	Completed	13 sec	SELECT d.year, d.month, COUNT(trans_id) AS total_transaction_count, SUM(case when atm_status = 'inactive' then 1 else 0 end) as inactive_count, CASE when co ...
<input type="checkbox"/>	Nov 25th, 2023 05:57:29 PM 34 minutes ago	1947	Completed	13 sec	-- Copying the data to dim_date table copy atm_data.dim_date from 's3://atm-bucket-s3/dim_date/part-00000-48a9e4b1-639e-4555-87b0-42a1d9ae5800-c ...
<input type="checkbox"/>	Nov 25th, 2023 05:59:55 PM 31 minutes ago	2002	Completed	12 sec	-- Copying the data to dim_location table copy atm_data.dim_location from 's3://atm-bucket-s3/dim_location/part-00000-775974fa-a051-424f-b2a8-2d244 ...

Database configurations:- Database name 'atm', user awsuser and password 'AWSUser1'.



The screenshot shows the AWS Redshift console for a cluster named 'atm'. The left sidebar contains navigation options like 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Datashares', 'Configurations', and 'AWS Partner integration'. The main content area is divided into two sections: 'Database configurations' and 'Network and security settings'.

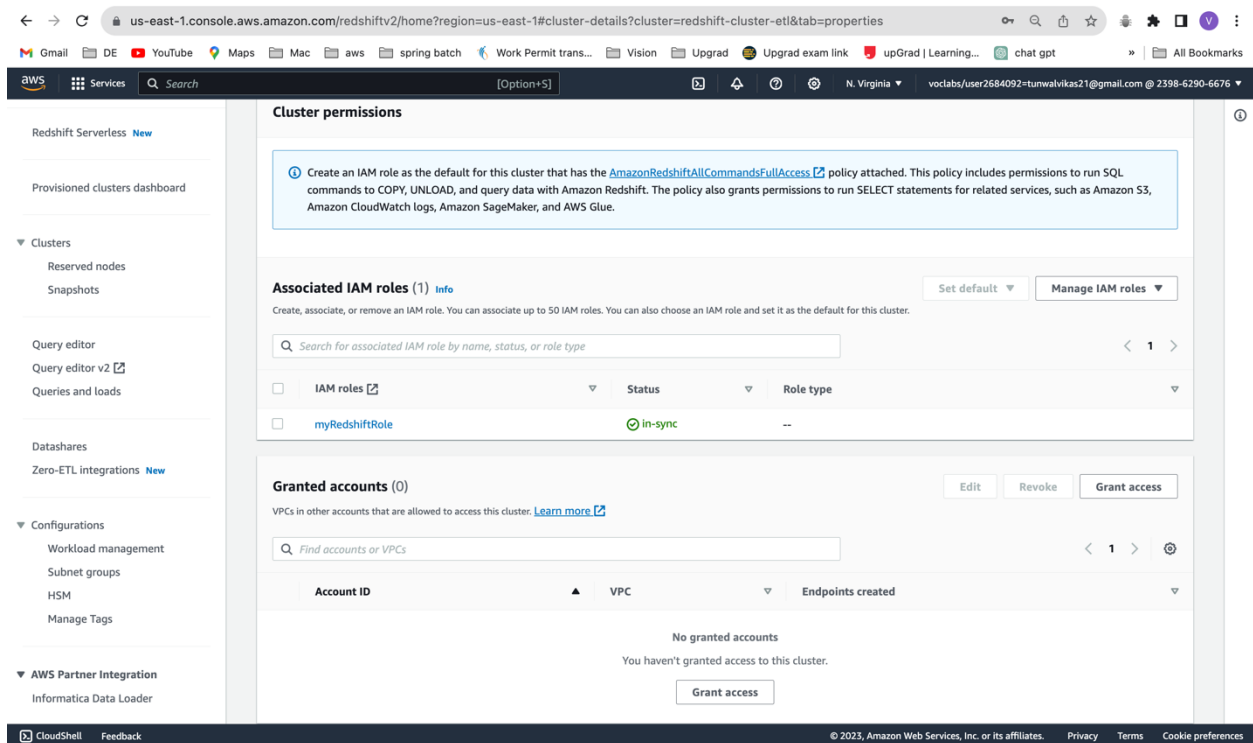
Database configurations:

- Database name: atm
- Parameter group: default.redshift-1.0
- Port: 5440
- Admin user name: awsuser
- Encryption: Disabled
- Audit logging: Disabled

Network and security settings:

- Virtual private cloud (VPC): vpc-0d884426ec90ed703
- Availability Zone: us-east-1c
- Subnet group: cluster-subnet-group-1
- Endpoint URL: -
- VPC security group: sg-0cf777400f4e3954f
- Publicly accessible: Disabled

Custom IAM role to allow Redshift the S3 access



The screenshot shows the AWS Redshift console for a cluster named 'atm'. The left sidebar contains navigation options like 'Redshift Serverless', 'Provisioned clusters dashboard', 'Clusters', 'Query editor', 'Datashares', 'Configurations', and 'AWS Partner integration'. The main content area is divided into two sections: 'Cluster permissions' and 'Granted accounts'.

Cluster permissions:

- Create an IAM role as the default for this cluster that has the [AmazonRedshiftAllCommandsFullAccess](#) policy attached. This policy includes permissions to run SQL commands to COPY, UNLOAD, and query data with Amazon Redshift. The policy also grants permissions to run SELECT statements for related services, such as Amazon S3, Amazon CloudWatch logs, Amazon SageMaker, and AWS Glue.

Associated IAM roles (1):

IAM roles	Status	Role type
myRedshiftRole	In-sync	--

Granted accounts (0):

VPCs in other accounts that are allowed to access this cluster. [Learn more](#)

Find accounts or VPCs

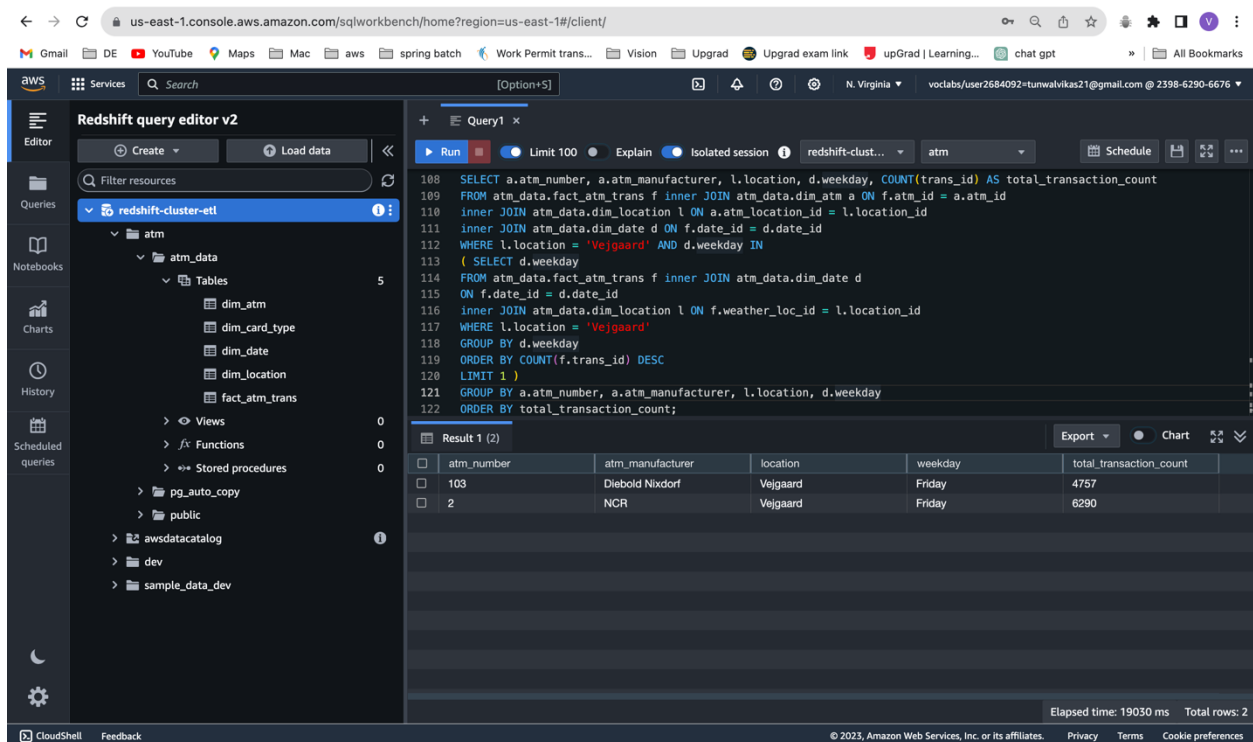
Account ID | VPC | Endpoints created

No granted accounts

You haven't granted access to this cluster.

[Grant access](#)

Setting up a database in the Redshift cluster and running queries to create the dimension and fact tables



The screenshot shows the AWS Redshift query editor v2 interface. The left sidebar displays the database schema, including a tree view of the 'atm' database and its 'atm_data' tables. The main editor shows a SQL query for 'Query1' that joins fact and dimension tables. The bottom right of the editor displays the query results for 'Result 1 (2)', showing two rows of data with columns: atm_number, atm_manufacturer, location, weekday, and total_transaction_count.

atm_number	atm_manufacturer	location	weekday	total_transaction_count
103	Diebold Nixdorf	Vejgaard	Friday	4757
2	NCR	Vejgaard	Friday	6290

Queries to create the various dimension and fact tables with appropriate primary and foreign keys:

Creating schema

```
create schema atm_data;
```

Creating location dimension table

```
--  
create table atm_data.DIM_LOCATION  
(  
location_id int not null DISTKEY SORTKEY, location varchar(50),  
streetname varchar(255), street_number int, zipcode int,  
lat decimal(10,3),  
lon decimal(10,3), PRIMARY KEY(location_id) );
```

Creating atm dimension table

```
--  
create table atm_data.DIM_ATM  
(  
atm_id int not null DISTKEY SORTKEY, atm_number varchar(20), atm_manufacturer  
varchar(50), atm_location_id int,  
PRIMARY KEY(atm_id),  
FOREIGN KEY(atm_location_id) references atm_data.DIM_LOCATION(location_id) );
```

Creating date dimension table

```
--  
create table atm_data.DIM_DATE  
(  
date_id int not null DISTKEY SORTKEY, full_date_time timestamp,  
year int,  
month varchar(20),  
day int,  
hour int,  
weekday varchar(20), PRIMARY KEY(date_id) );
```

Creating card type dimension table

```
--  
create table atm_data.DIM_CARD_TYPE  
(  
card_type_id int not null DISTKEY SORTKEY, card_type varchar(30),  
PRIMARY KEY(card_type_id) );
```

Creating atm transactions fact table

```
-- Creating atm transactions fact table
create table atm_data.FACT_ATM_TRANS (
  trans_id bigint not null DISTKEY SORTKEY, atm_id int,
  weather_loc_id int,
  date_id int,
  card_type_id int,
  atm_status varchar(20),
  currency varchar(10),
  service varchar(20),
  transaction_amount int,
  message_code varchar(225),
  message_text varchar(225),
  rain_3h decimal(10,3),
  clouds_all int,
  weather_id int,
  weather_main varchar(50),
  weather_description varchar(255),
  PRIMARY KEY(trans_id),
  FOREIGN KEY(weather_loc_id) references atm_data.DIM_LOCATION(location_id),
  FOREIGN KEY(atm_id) references atm_data.DIM_ATM(atm_id),
  FOREIGN KEY(date_id) references atm_data.DIM_DATE(date_id),
  FOREIGN KEY(card_type_id) references atm_data.DIM_CARD_TYPE(card_type_id) );
```