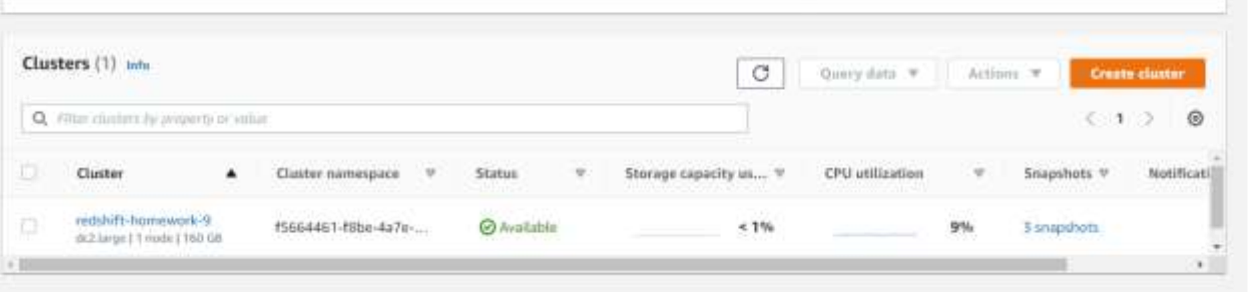
Data base systems homework 9

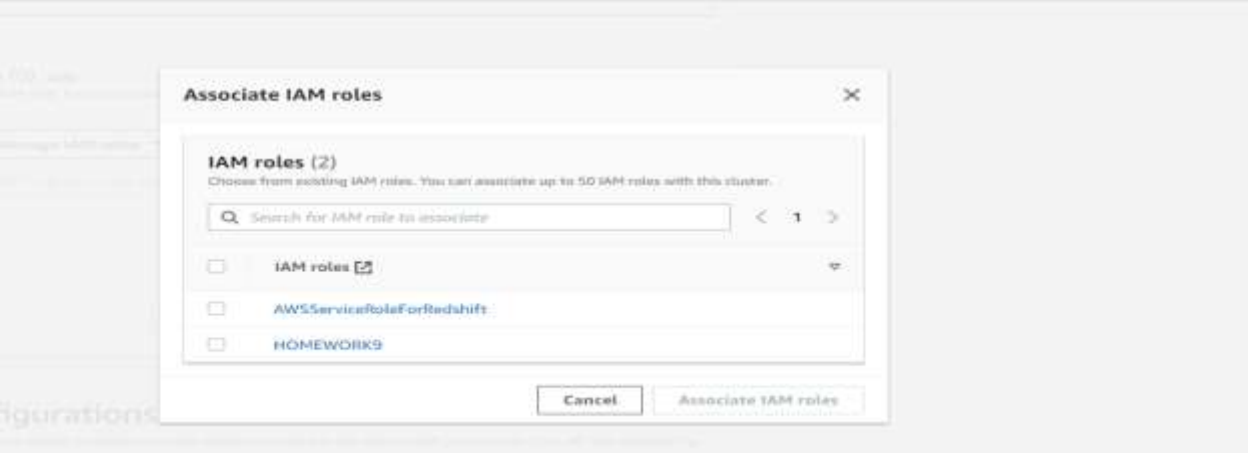
1. Create a cluster named 'redshift_hw9' (only use free tier)



Clusters (1) Info

Cluster	Cluster namespace	Status	Storage capacity us...	CPU utilization	Snapshots	Notificati
redshift-homework-9 dc2.large | 1 node | 160 GB	f5664461-f8be-4a7e-...	✓ Available	< 1%	9%	5 snapshots	

2. Create/Show the required IAM roles for the task.



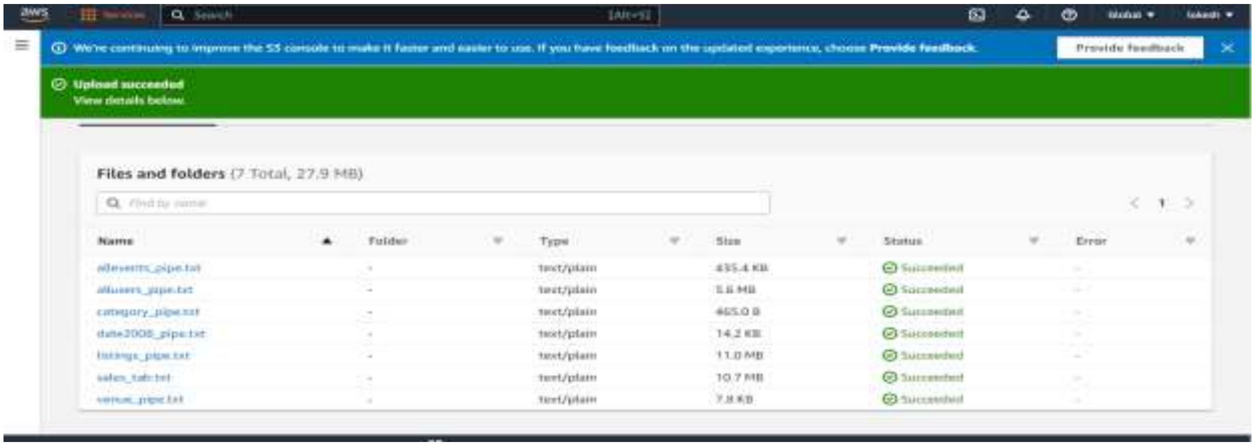
Associate IAM roles

IAM roles (2)

Choose from existing IAM roles. You can associate up to 50 IAM roles with this cluster.

	IAM roles
□	AWSServiceRoleForRedshift
□	HOMEWORK9

3Create an S3 bucket and load the tickitdb files.



We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose **Provide feedback**.

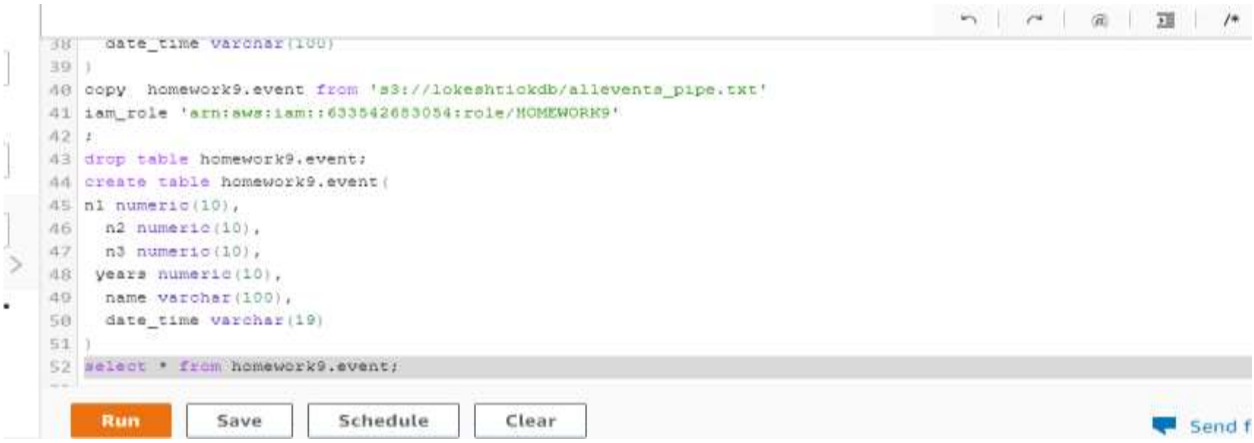
Upload succeeded
View details below.

Files and folders (7 Total, 27.9 MB)

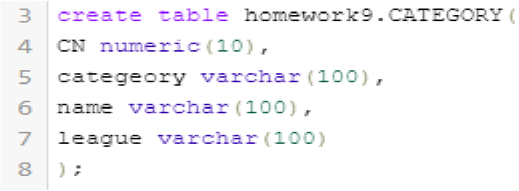
Name	Folder	Type	Size	Status	Error
allevents_pipe.txt	-	text/plain	455.4 KB	✓ Succeeded	-
allusers_pipe.txt	-	text/plain	5.6 MB	✓ Succeeded	-
category_pipe.txt	-	text/plain	465.0 B	✓ Succeeded	-
date2008_pipe.txt	-	text/plain	14.2 KB	✓ Succeeded	-
listings_pipe.txt	-	text/plain	11.0 MB	✓ Succeeded	-
sales_tab.txt	-	text/plain	10.7 MB	✓ Succeeded	-
venue_pipe.txt	-	text/plain	7.8 KB	✓ Succeeded	-

Data base systems homework 9

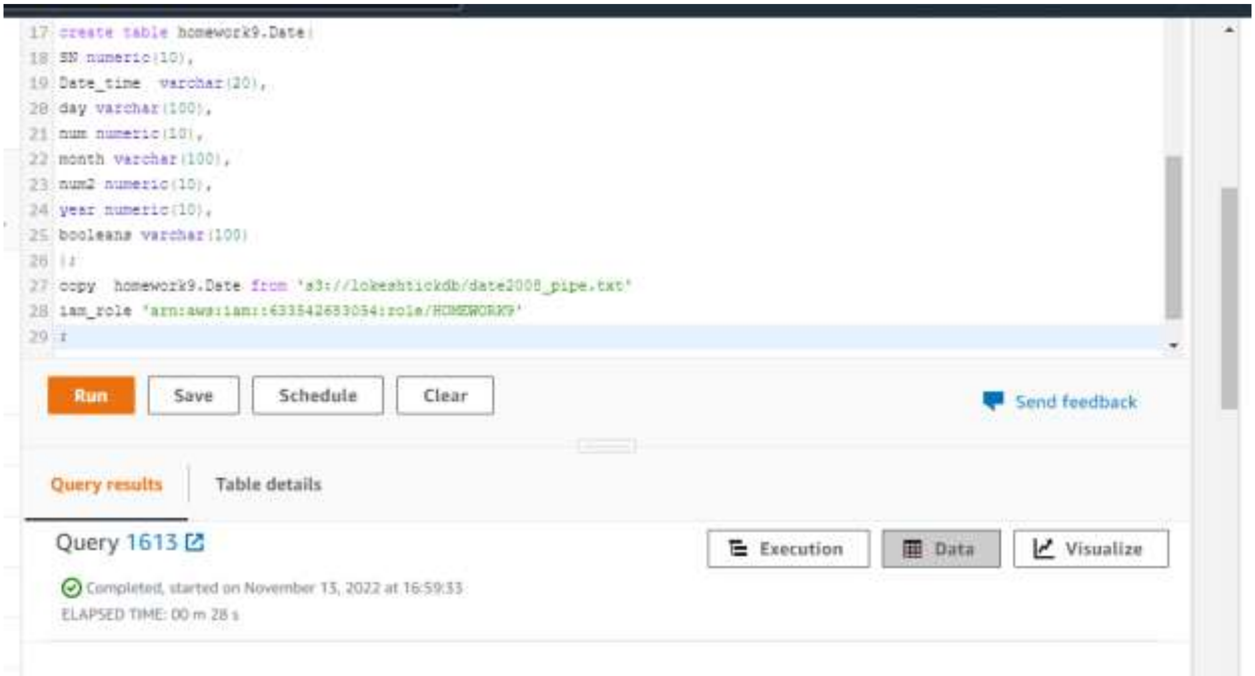
4.Create tables on your cluster. (CATEGORY table, DATE table, EVENT table, VENUE table, USERS table, LISTING table, SALES table)



```
  date_time varchar(100)
)
copy  homework9.event from 's3://lokeshtickdb/allevents_pipe.txt'
iam_role 'arn:aws:iam::633542683054:role/HOMEWORK9'
;
drop table homework9.event;
create table homework9.event(
n1 numeric(10),
  n2 numeric(10),
  n3 numeric(10),
 years numeric(10),
  name varchar(100),
  date_time varchar(19)
)
select * from homework9.event;
```



```
create table homework9.CATEGORY(
CN numeric(10),
categeory varchar(100),
name varchar(100),
league varchar(100)
);
```



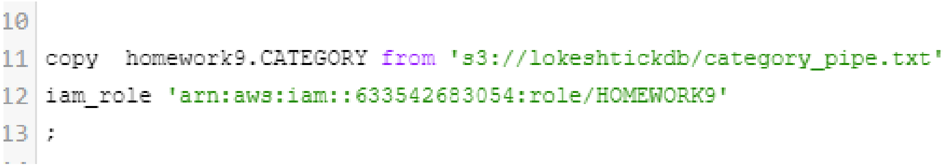
```
create table homework9.Date(
SN numeric(10),
Date_time  varchar(20),
day varchar(100),
num numeric(10),
month varchar(100),
num2 numeric(10),
year numeric(10),
booleans varchar(100)
);
copy  homework9.Date from 's3://lokeshtickdb/date2008_pipe.txt'
iam_role 'arn:aws:iam::633542683054:role/HOMEWORK9'
;
```

Query 1613

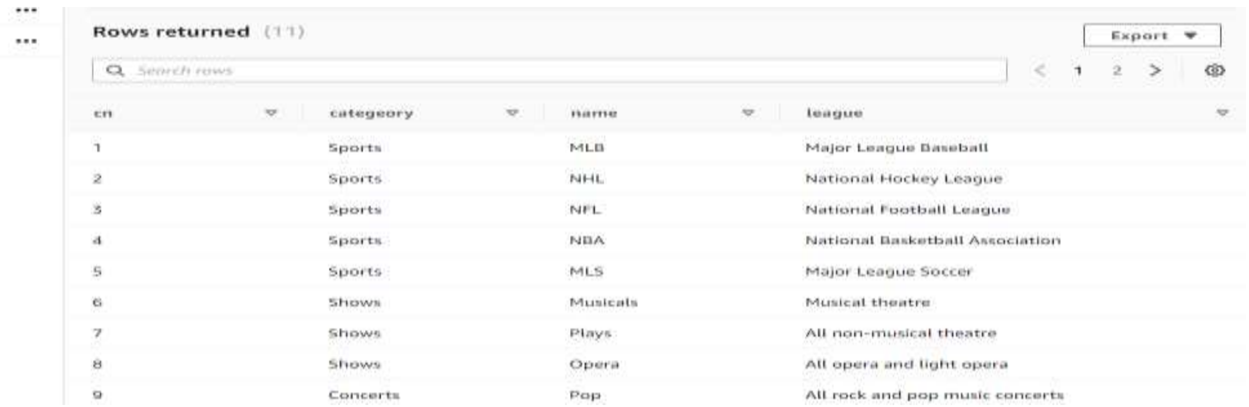
Completed, started on November 13, 2022 at 16:59:33

ELAPSED TIME: 00 m 28 s

5. Copy the data from S3 to Redshift cluster 'redshift_hw9'.



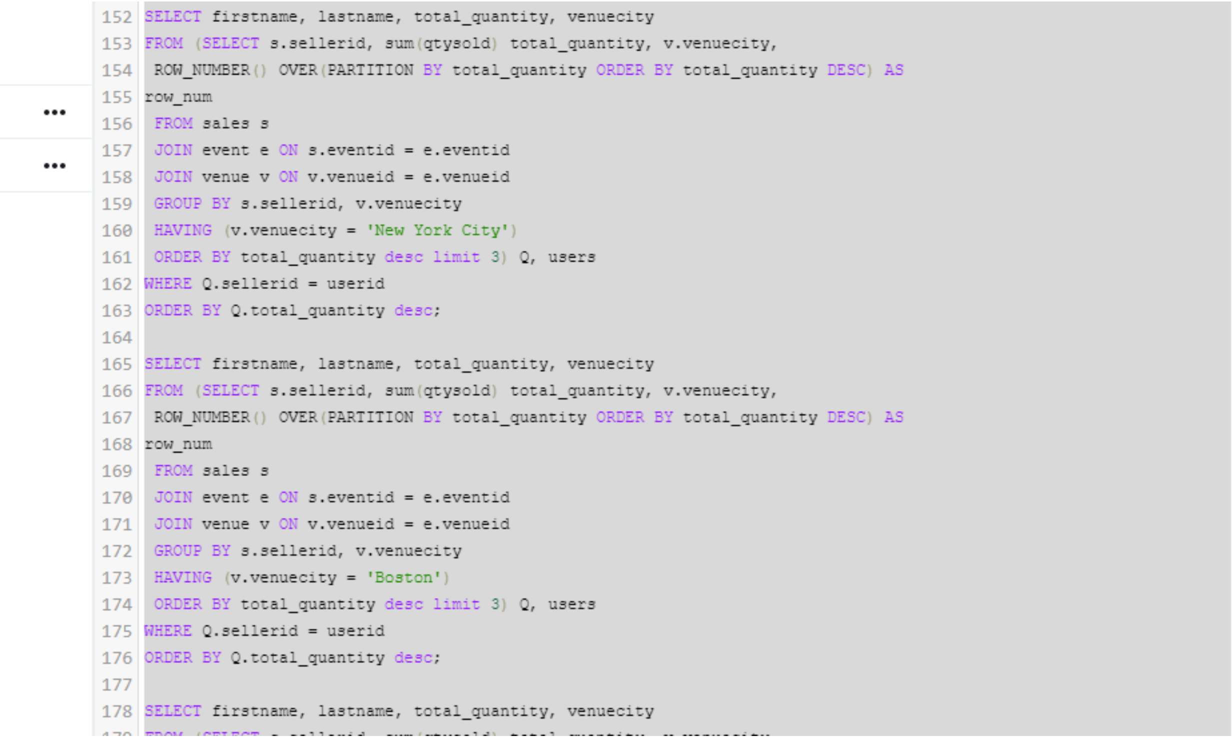
```

copy  homework9.CATEGORY from 's3://lokeshtickdb/category_pipe.txt'
iam_role 'arn:aws:iam::633542683054:role/HOMEWORK9'
;
```



Rows returned (11)

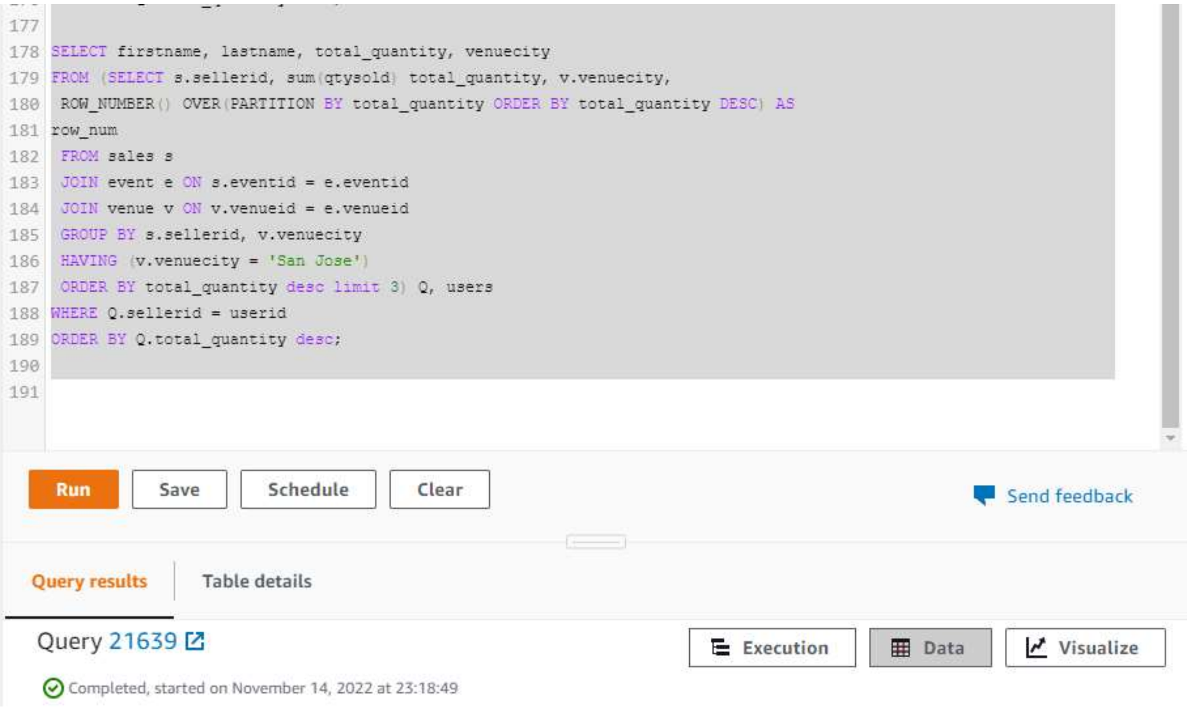
cn	category	name	league
1	Sports	MLB	Major League Baseball
2	Sports	NHL	National Hockey League
3	Sports	NFL	National Football League
4	Sports	NBA	National Basketball Association
5	Sports	MLS	Major League Soccer
6	Shows	Musicals	Musical theatre
7	Shows	Plays	All non-musical theatre
8	Shows	Opera	All opera and light opera
9	Concerts	Pop	All rock and pop music concerts

6.Write a query in the Redshift query editor to find the top 3 sellers in **San Diego, New York, Boston, and San Jose** based on the number of tickets sold in 2008. (Use Window functions)



```
SELECT firstname, lastname, total_quantity, venuecity
FROM (SELECT s.sellerid, sum(qtysold) total_quantity, v.venuecity,
 ROW_NUMBER() OVER(PARTITION BY total_quantity ORDER BY total_quantity DESC) AS
row_num
 FROM sales s
 JOIN event e ON s.eventid = e.eventid
 JOIN venue v ON v.venueid = e.venueid
 GROUP BY s.sellerid, v.venuecity
 HAVING (v.venuecity = 'New York City')
 ORDER BY total_quantity desc limit 3) Q, users
WHERE Q.sellerid = userid
ORDER BY Q.total_quantity desc;

SELECT firstname, lastname, total_quantity, venuecity
FROM (SELECT s.sellerid, sum(qtysold) total_quantity, v.venuecity,
 ROW_NUMBER() OVER(PARTITION BY total_quantity ORDER BY total_quantity DESC) AS
row_num
 FROM sales s
 JOIN event e ON s.eventid = e.eventid
 JOIN venue v ON v.venueid = e.venueid
 GROUP BY s.sellerid, v.venuecity
 HAVING (v.venuecity = 'Boston')
 ORDER BY total_quantity desc limit 3) Q, users
WHERE Q.sellerid = userid
ORDER BY Q.total_quantity desc;

SELECT firstname, lastname, total_quantity, venuecity
FROM (SELECT s.sellerid, sum(qtysold) total_quantity, v.venuecity,
```

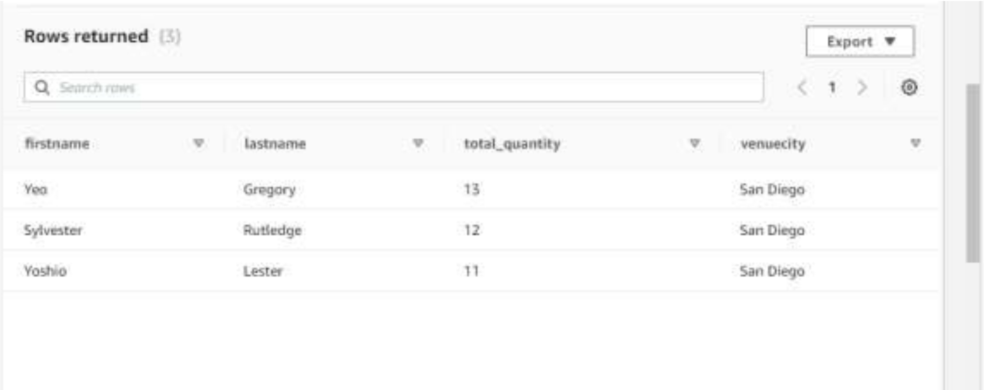
Data base systems homework 9



```
SELECT firstname, lastname, total_quantity, venuecity
FROM (SELECT s.sellerid, sum(qtysold) total_quantity, v.venuecity,
 ROW_NUMBER() OVER(PARTITION BY total_quantity ORDER BY total_quantity DESC) AS
row_num
 FROM sales s
 JOIN event e ON s.eventid = e.eventid
 JOIN venue v ON v.venueid = e.venueid
 GROUP BY s.sellerid, v.venuecity
 HAVING (v.venuecity = 'San Jose')
 ORDER BY total_quantity desc limit 3) Q, users
WHERE Q.sellerid = userid
ORDER BY Q.total_quantity desc;
```

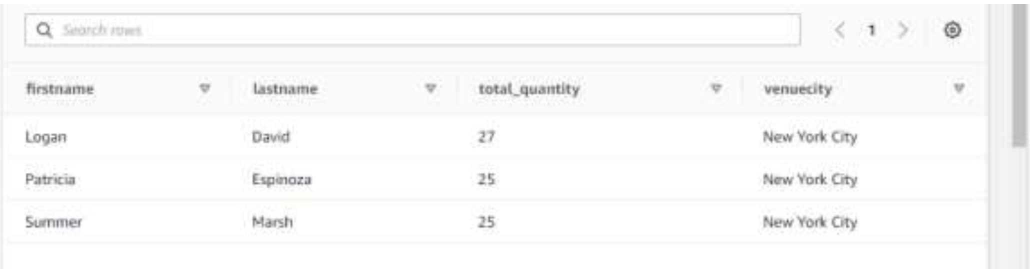
Query 21639

Completed, started on November 14, 2022 at 23:18:49



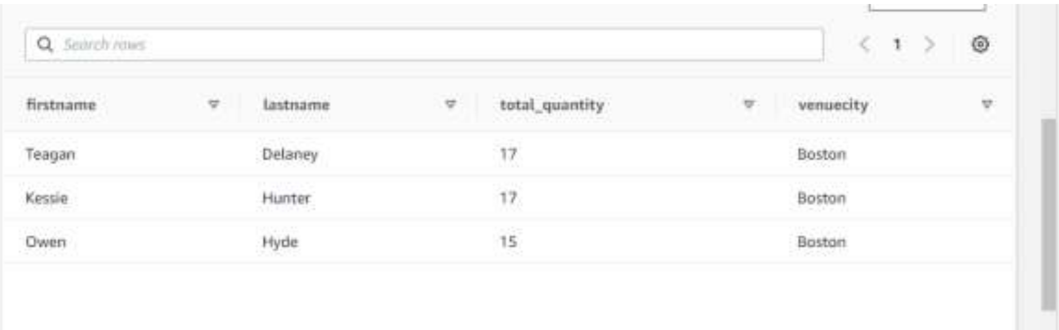
Rows returned (3)

firstname	lastname	total_quantity	venuecity
Yeo	Gregory	13	San Diego
Sylvester	Rutledge	12	San Diego
Yoshio	Lester	11	San Diego

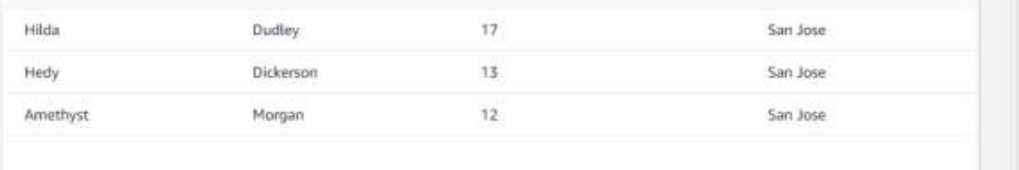


firstname	lastname	total_quantity	venuecity
Logan	David	27	New York City
Patricia	Espinoza	25	New York City
Summer	Marsh	25	New York City

Data base systems homework 9

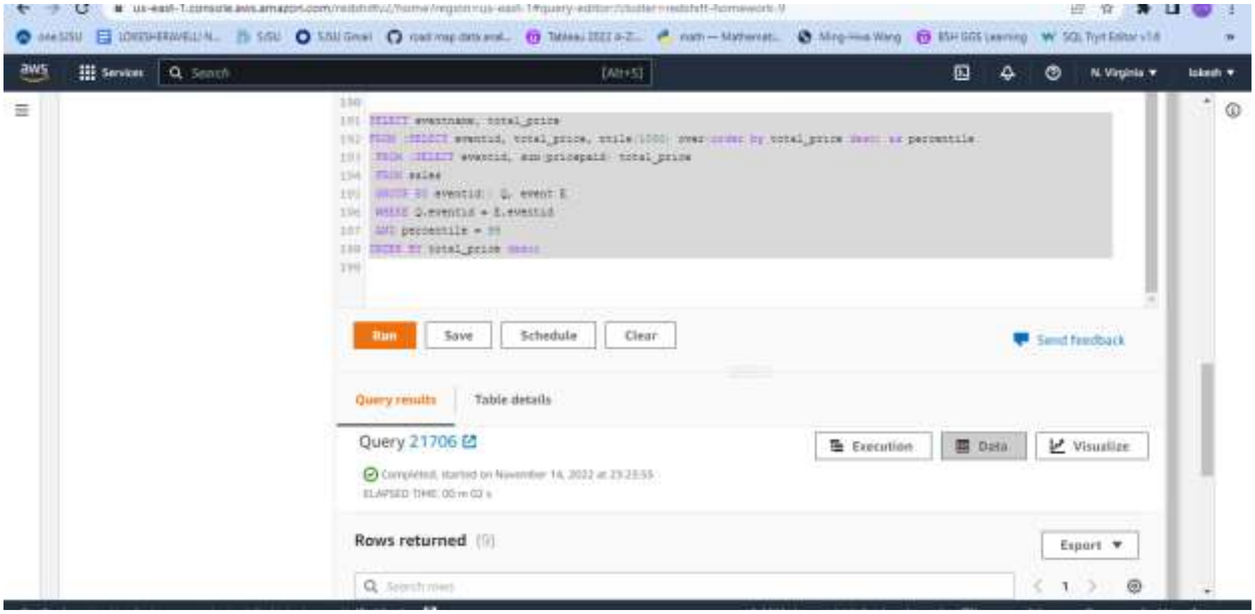


firstname	lastname	total_quantity	venuecity
Teagan	Delaney	17	Boston
Kessie	Hunter	17	Boston
Owen	Hyde	15	Boston



Hilda	Dudley	17	San Jose
Hedy	Dickerson	13	San Jose
Amethyst	Morgan	12	San Jose

7) Write a query in the Redshift query editor to see events with the lowest sales.

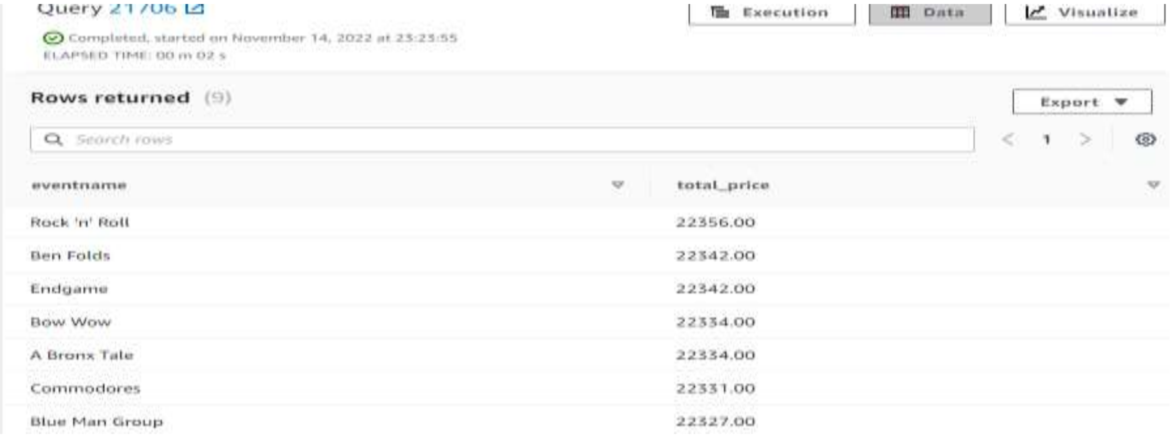


```
SELECT eventname, total_price
FROM (SELECT eventid, total_price, ntile(1000) over(order by total_price desc) as percentile
 FROM (SELECT eventid, sum(pricepaid) total_price
 FROM sales
 GROUP BY eventid)) Q, event E
 WHERE Q.eventid = E.eventid
 AND percentile = 99
ORDER BY total_price desc;
```

Query 21706

Completed, started on November 14, 2022 at 23:23:55
ELAPSED TIME: 00 m 02 s

Rows returned (9)



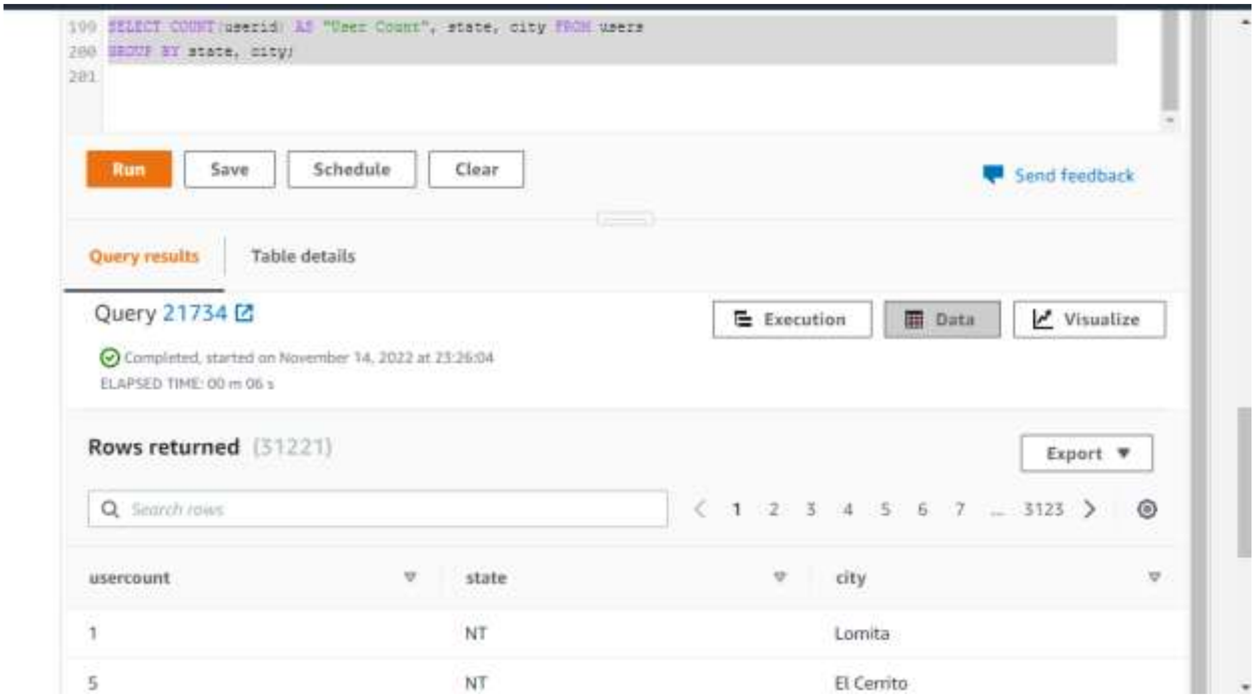
Query 21706

Completed, started on November 14, 2022 at 23:23:55
ELAPSED TIME: 00 m 02 s

Rows returned (9)

eventname	total_price
Rock 'n' Roll	22356.00
Ben Folds	22342.00
Endgame	22342.00
Bow Wow	22334.00
A Bronx Tale	22334.00
Commodores	22331.00
Blue Man Group	22327.00

8) Write a query in the Redshift query editor to count the number of users in each state and city combination.



```
SELECT COUNT(userid) AS "User Count", state, city FROM users
GROUP BY state, city;
```

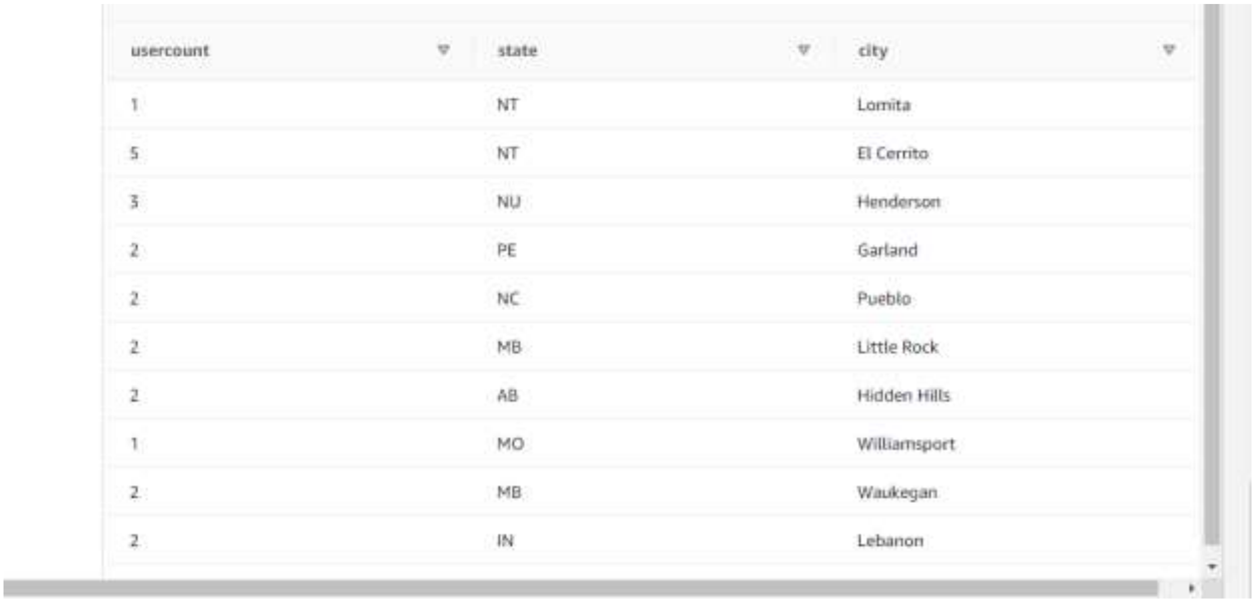
Query 21734

Completed, started on November 14, 2022 at 23:26:04

ELAPSED TIME: 00 m 06 s

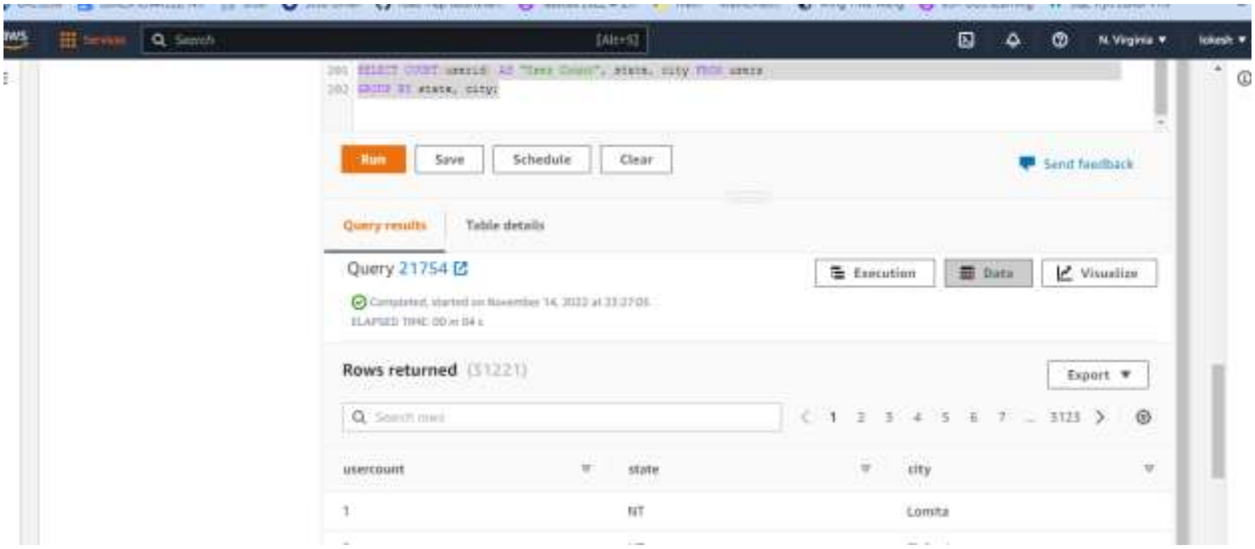
Rows returned (31221)

usercount	state	city
1	NT	Lomita
5	NT	El Cerrito



usercount	state	city
1	NT	Lomita
5	NT	El Cerrito
3	NU	Henderson
2	PE	Garland
2	NC	Pueblo
2	MB	Little Rock
2	AB	Hidden Hills
1	MO	Williamsport
2	MB	Waukegan
2	IN	Lebanon

Data base systems homework 9



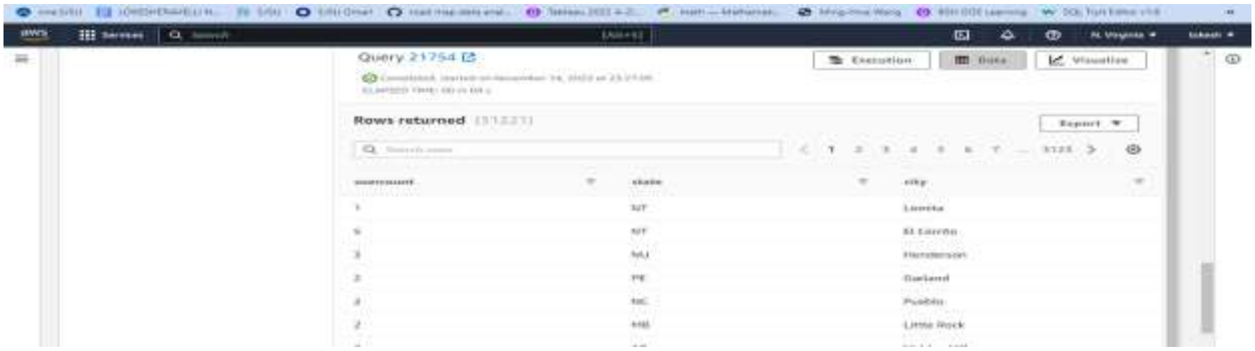
```
SELECT COUNT userid AS "User Count", state, city FROM users
GROUP BY state, city;
```

Query 21754

Completed, started on November 14, 2022 at 23:27:05
ELAPSED TIME: 00 m 04 s

Rows returned (31221)

usercount	state	city
1	NT	Lomita



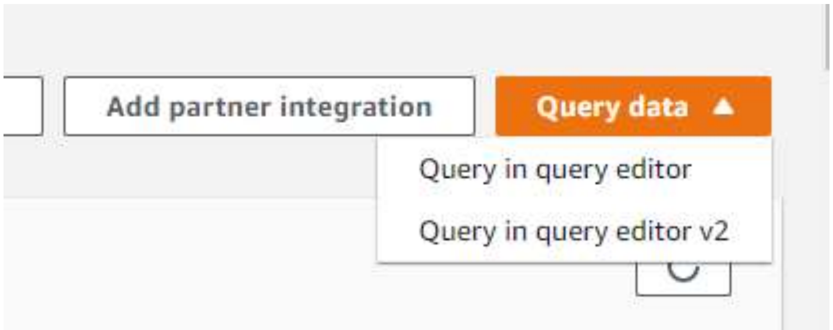
Query 21754

Completed, started on November 14, 2022 at 23:27:05
ELAPSED TIME: 00 m 04 s

Rows returned (31221)

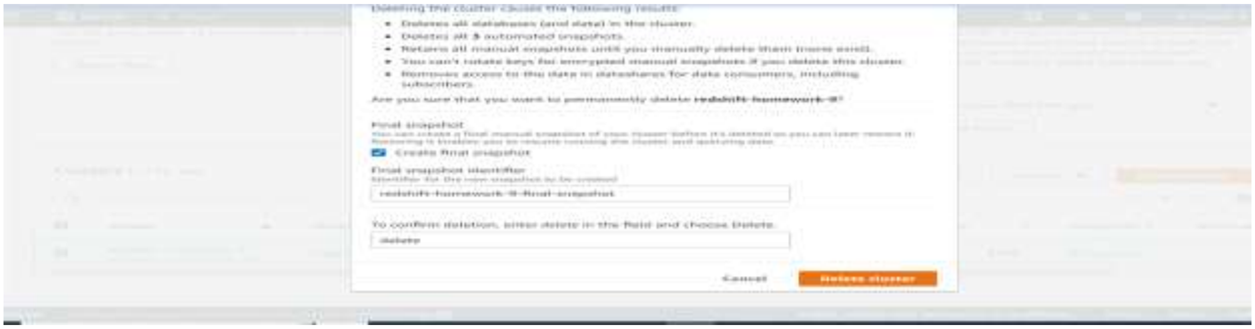
usercount	state	city
1	NT	Lomita
5	NT	El Cerrito
3	NU	Henderson
2	PE	Garland
3	NC	Pueblo
2	MB	Little Rock

9) What are Database Querying Options available in Amazon Redshift?



10) Delete the cluster 'redshift_hw9', IAM role and S3 bucket

Data base systems homework 9



Deleting the cluster causes the following results:

- Deletes all databases (and data) in the cluster.
- Deletes all 3 automated snapshots.
- Retains all manual snapshots until you manually delete them [illegible].
- You can't rotate keys for encrypted manual snapshots if you delete this cluster.
- Removes access to the data in datashares for data consumers, including subscribers.

Are you sure that you want to permanently delete **redshift-homework-9**?

Final snapshot

You can create a final manual snapshot of your cluster before it's deleted so you can later restore it. Restoring it enables you to resume running the cluster and querying data.

☑ Create final snapshot

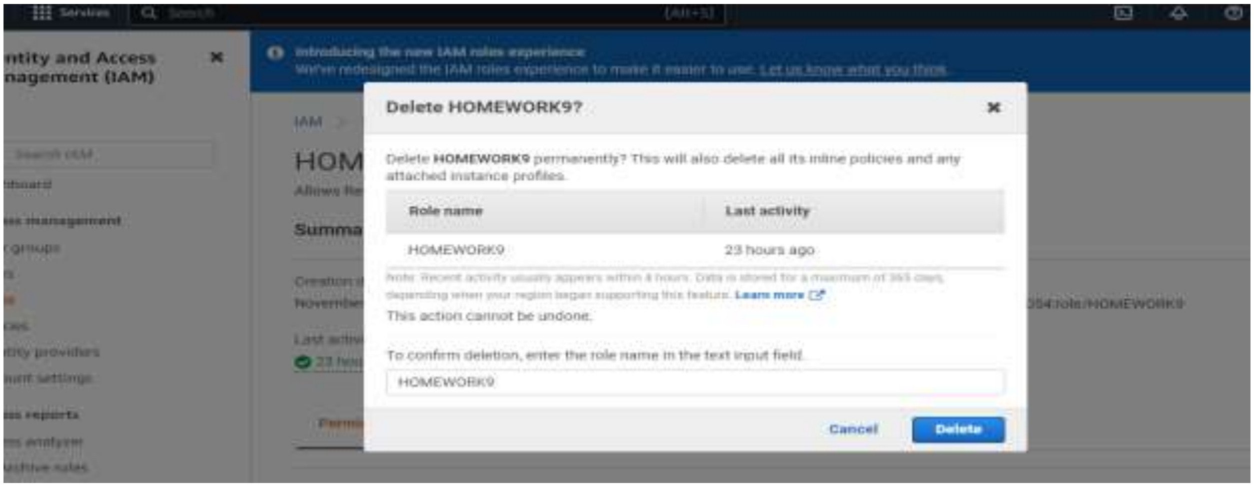
Final snapshot identifier

Identifier for the new snapshot to be created

redshift-homework-9-final-snapshot

To confirm deletion, enter delete in the field and choose Delete.

delete



Introducing the new IAM roles experience

We've redesigned the IAM roles experience to make it easier to use. Let us know what you think.

Delete HOMEWORK9?

Delete **HOMEWORK9** permanently? This will also delete all its inline policies and any attached instance profiles.

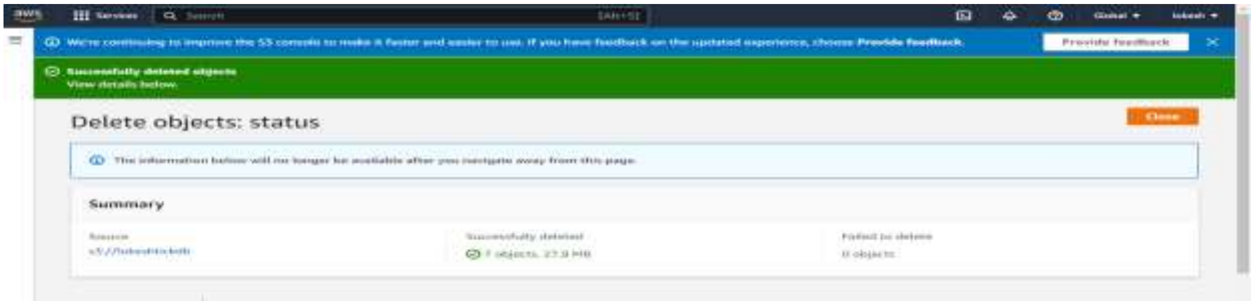
Role name	Last activity
HOMEWORK9	23 hours ago

Note: Recent activity usually appears within 4 hours. Data is stored for a maximum of 365 days, depending when your region began supporting this feature. Learn more

This action cannot be undone.

To confirm deletion, enter the role name in the text input field.

HOMEWORK9



We're continuing to improve the S3 console to make it faster and easier to use. If you have feedback on the updated experience, choose **Provide feedback**.

Successfully deleted objects

View details below.

Delete objects: status

The information below will no longer be available after you navigate away from this page.

Summary

Source	Successfully deleted	Failed to delete
s3://[illegible]	7 objects, 27.9 MB	0 objects