Introduction to Cloud Computing (CS 524)

(Lab Assignment 4)

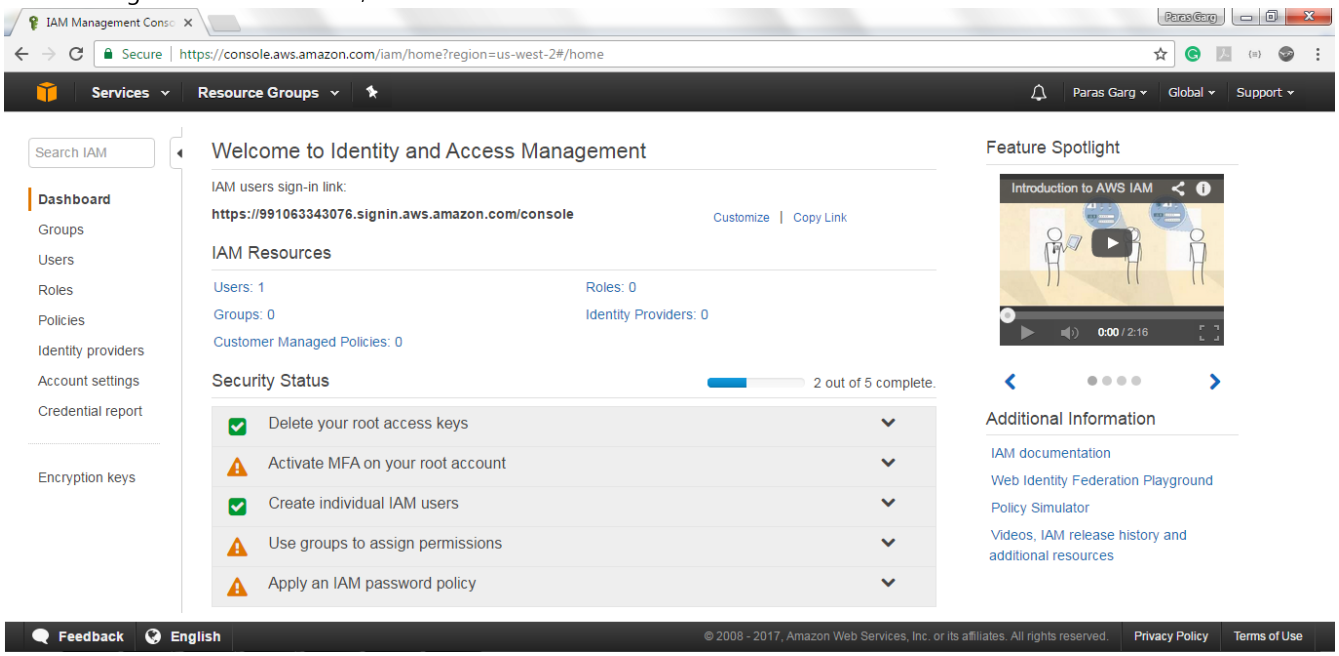
Prof. Igor Faynberg

Student Name: **Paras Garg**

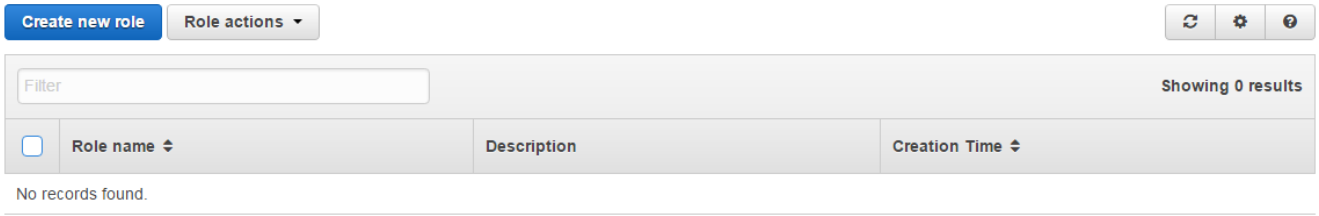
Course Section: **CS 524-A**

Step to create IAM role

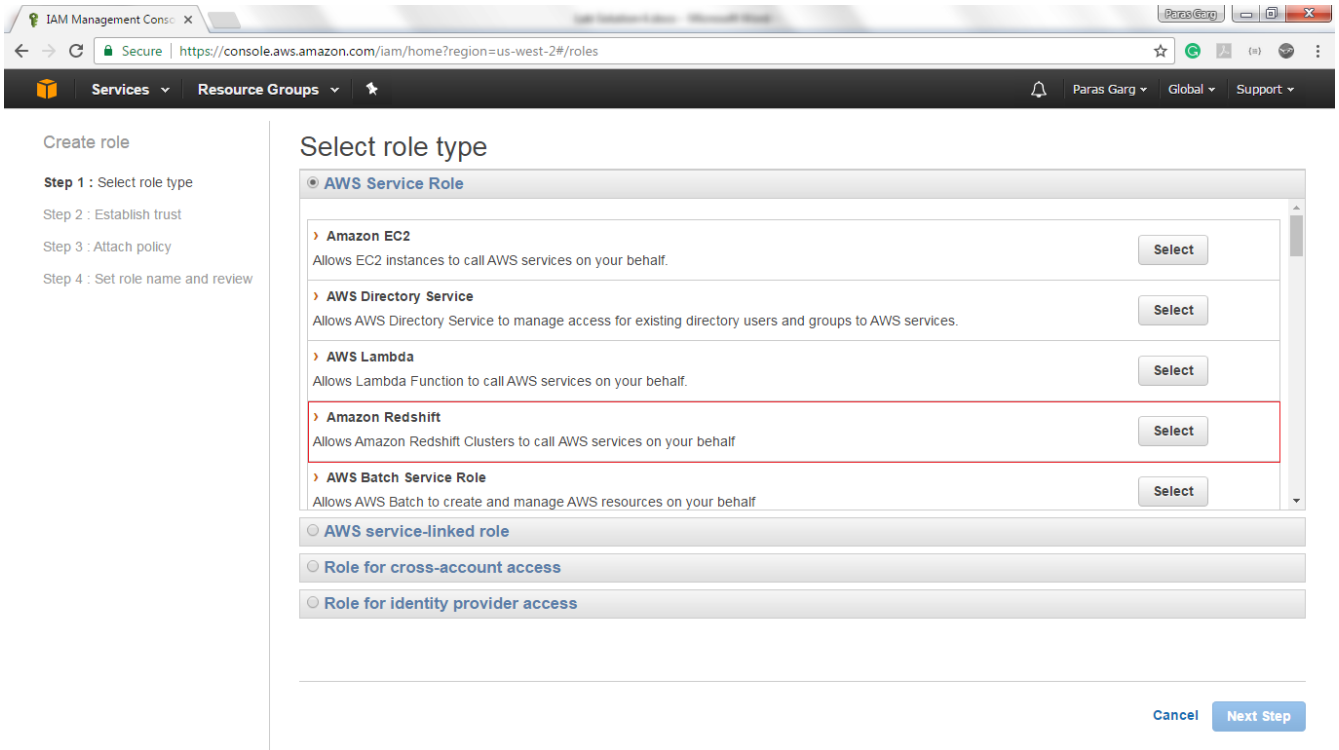
- After login into AWS RedShift, chose IAM service from the dashboard to land on IAM dashboard.



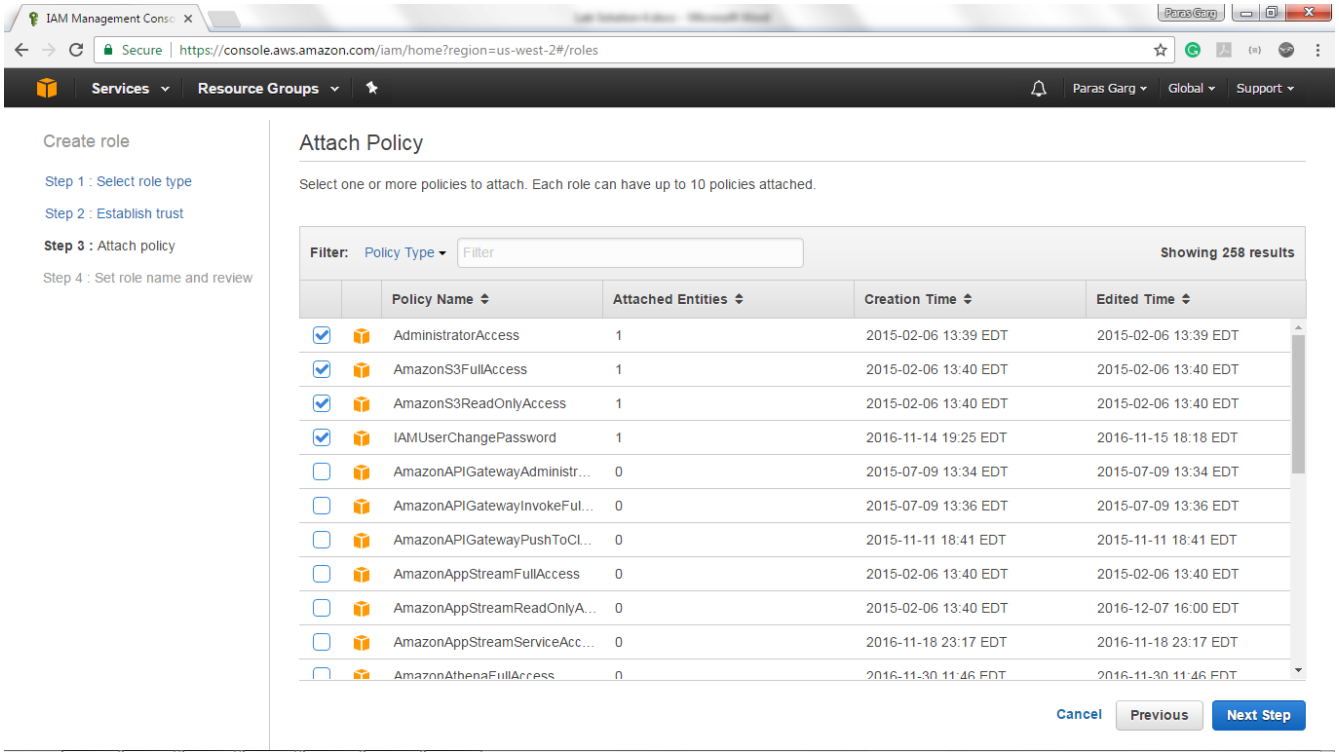
- Then clicked on Roles tab to create a new role for AWS RedShift, and clicked on **Create new role** button



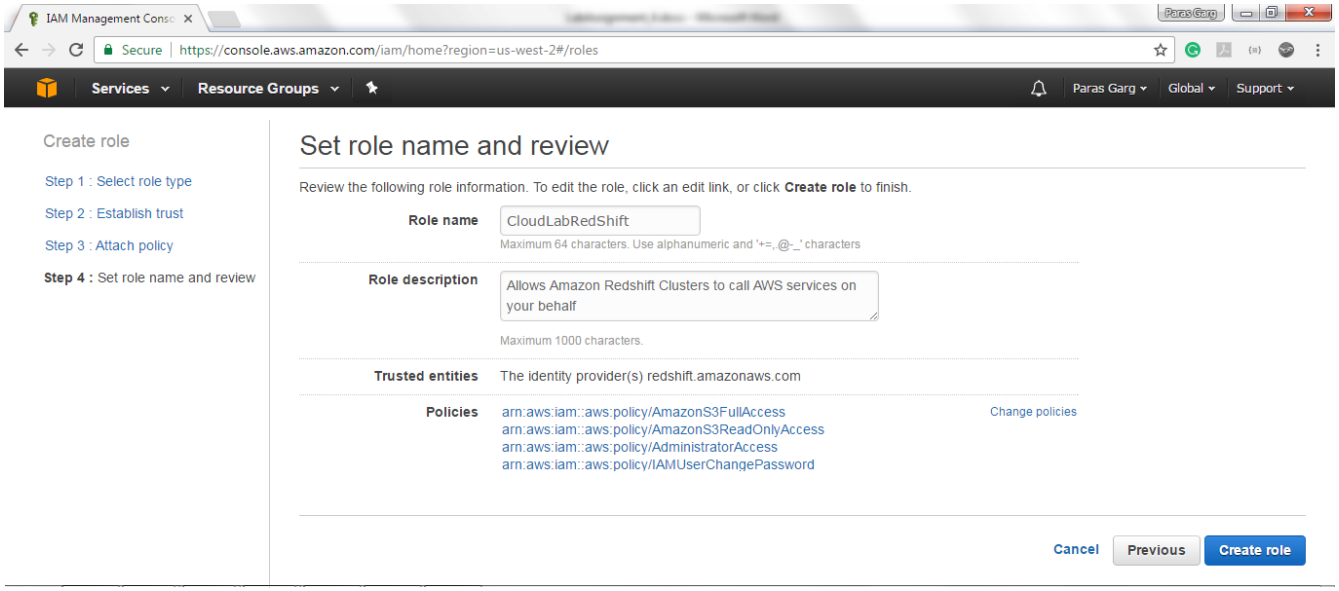
- Then under AWS Service Role, I have selected Amazon RedShift.



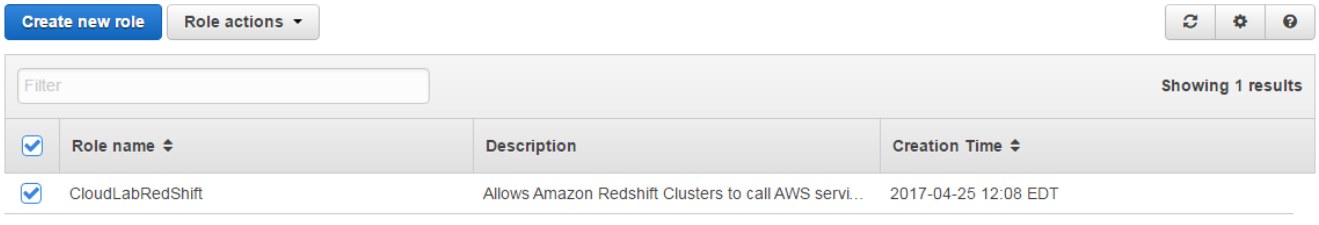
- Now attaching required policies to the selected role.



- In final step, prompting role name, reviewing other entities and then clicked on create role button.

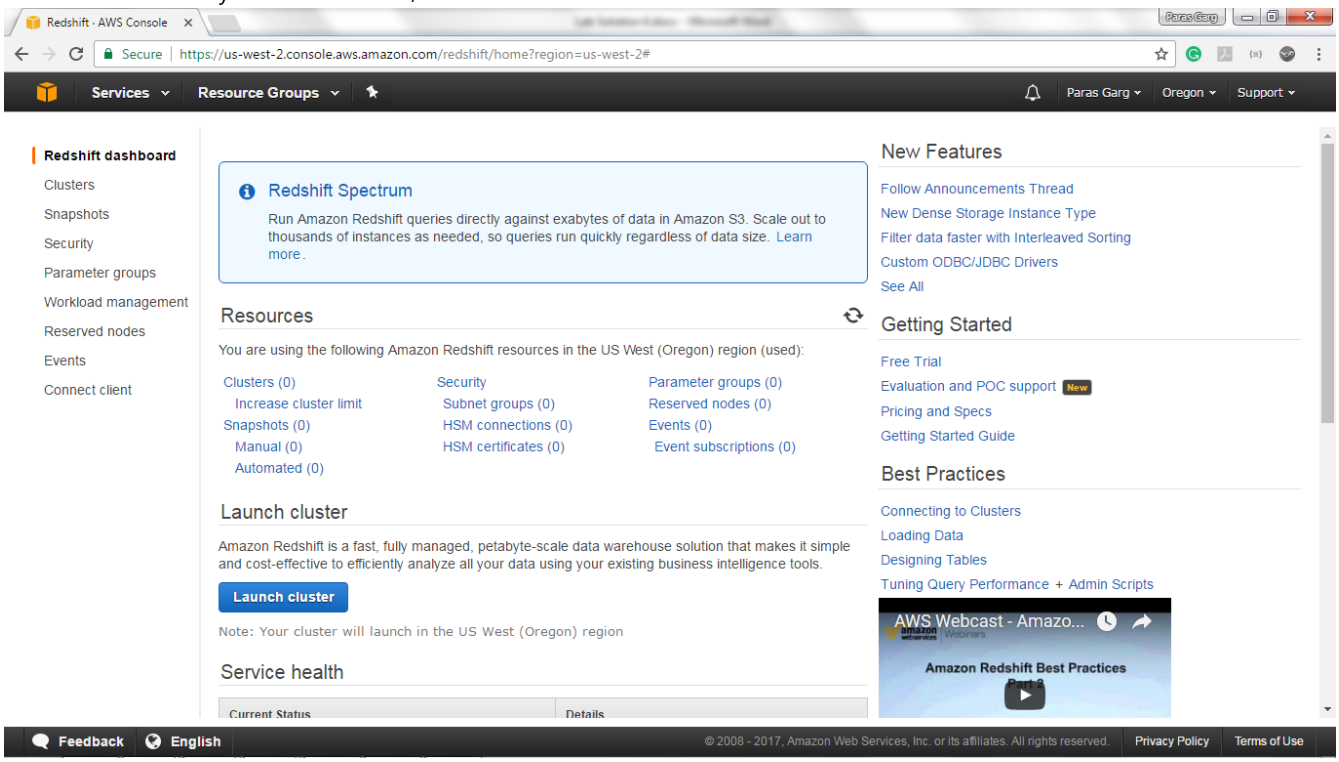


- Now, under the roles tab, we can check that the role has been created successfully

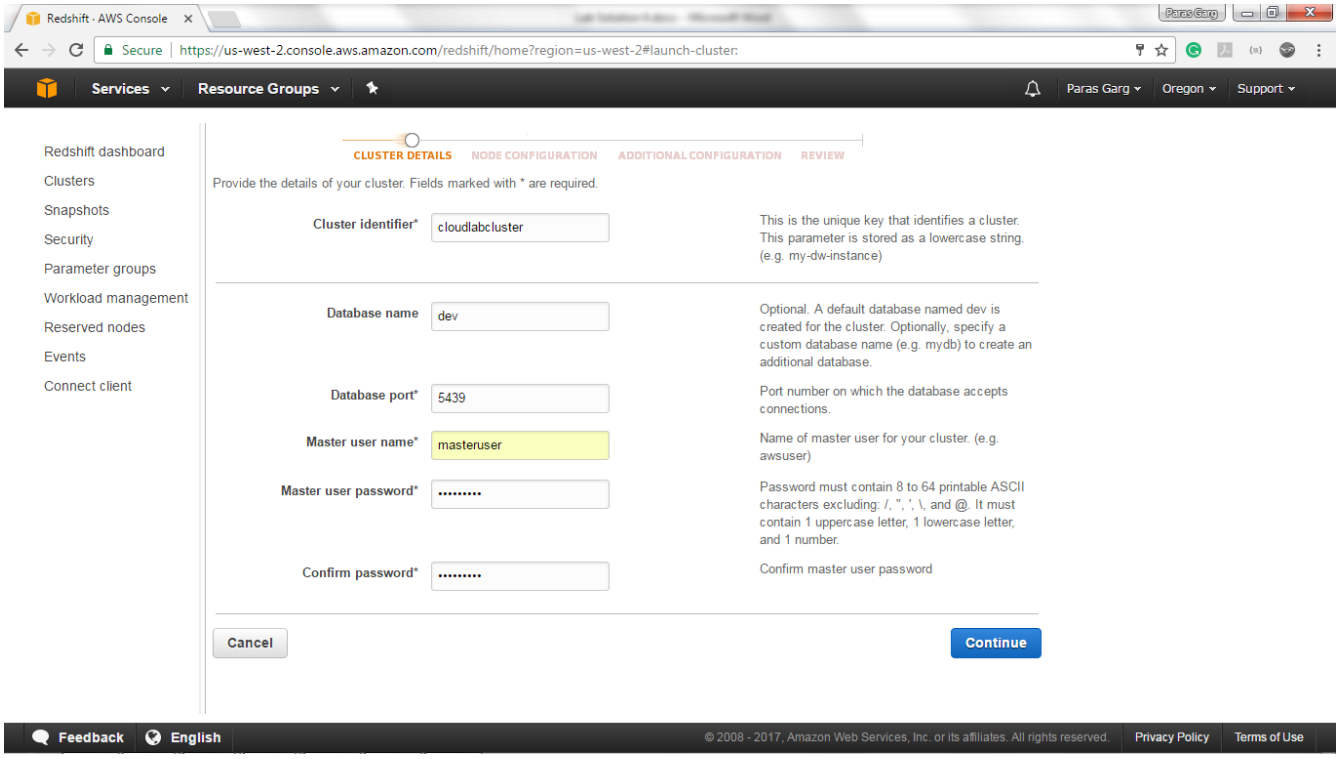


Step to create and launch RedShift cluster

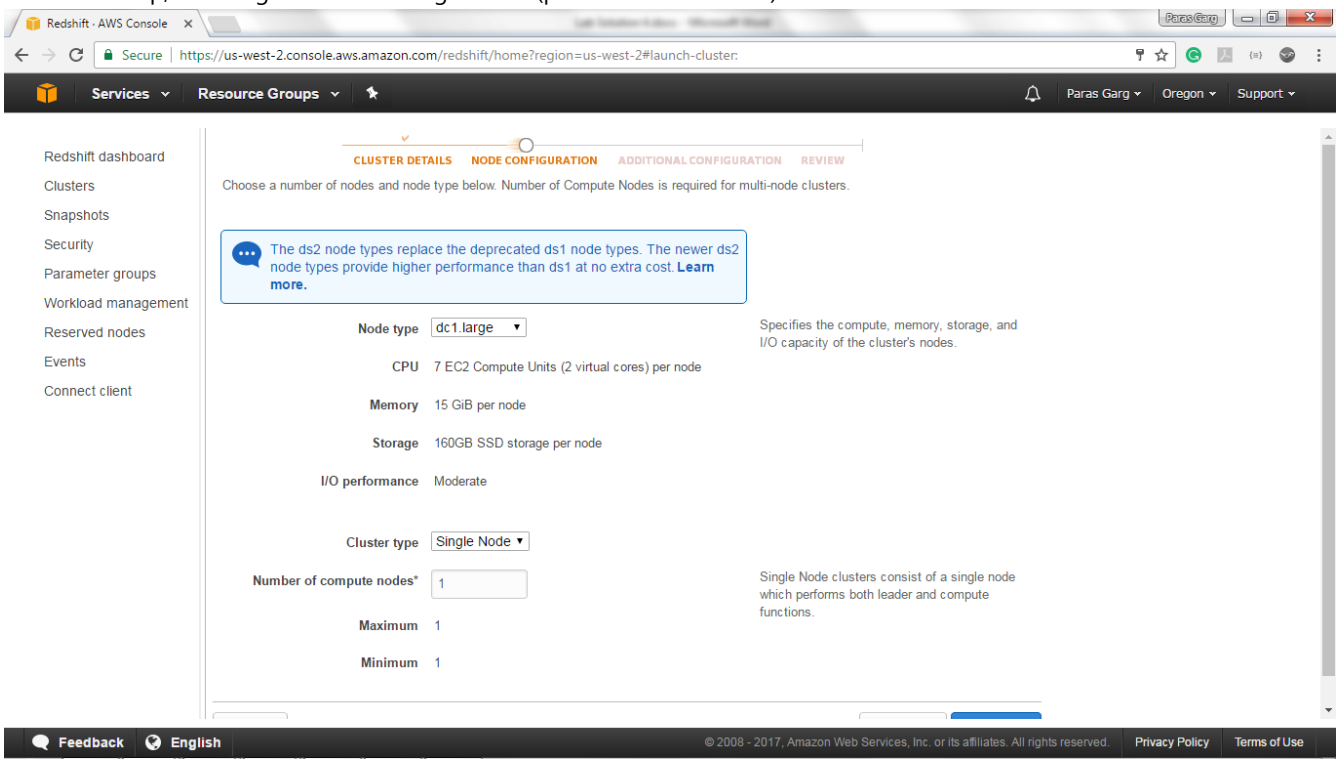
- After successfully created IAM role, I chose for RedShift service from the AWS dashboard to access RedShift dashboard.



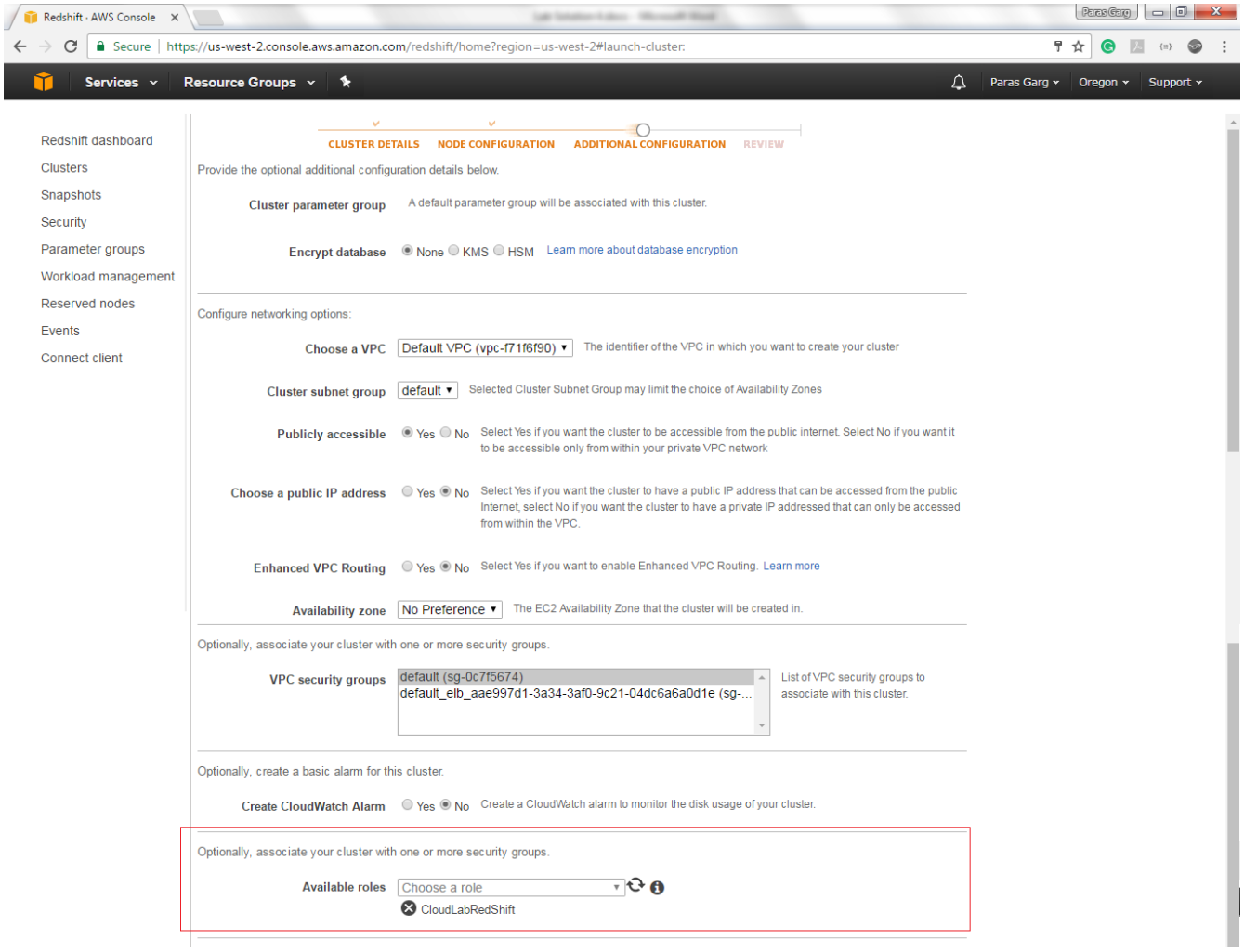
- After clicking to **Launch cluster** button in previous step, I have filled the cluster details where Database name and Master user name has been set to default value mentioned in AWS website. Then clicked on **Continue** button.



- In this step, checking for Node Configuration. (passed default values)



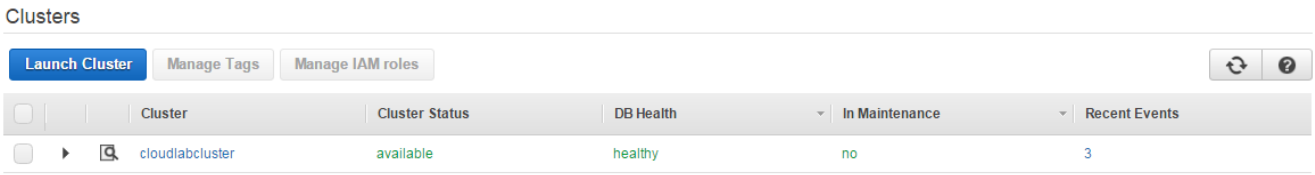
- In Additional Configuration Settings, I have set recently created IAM role.



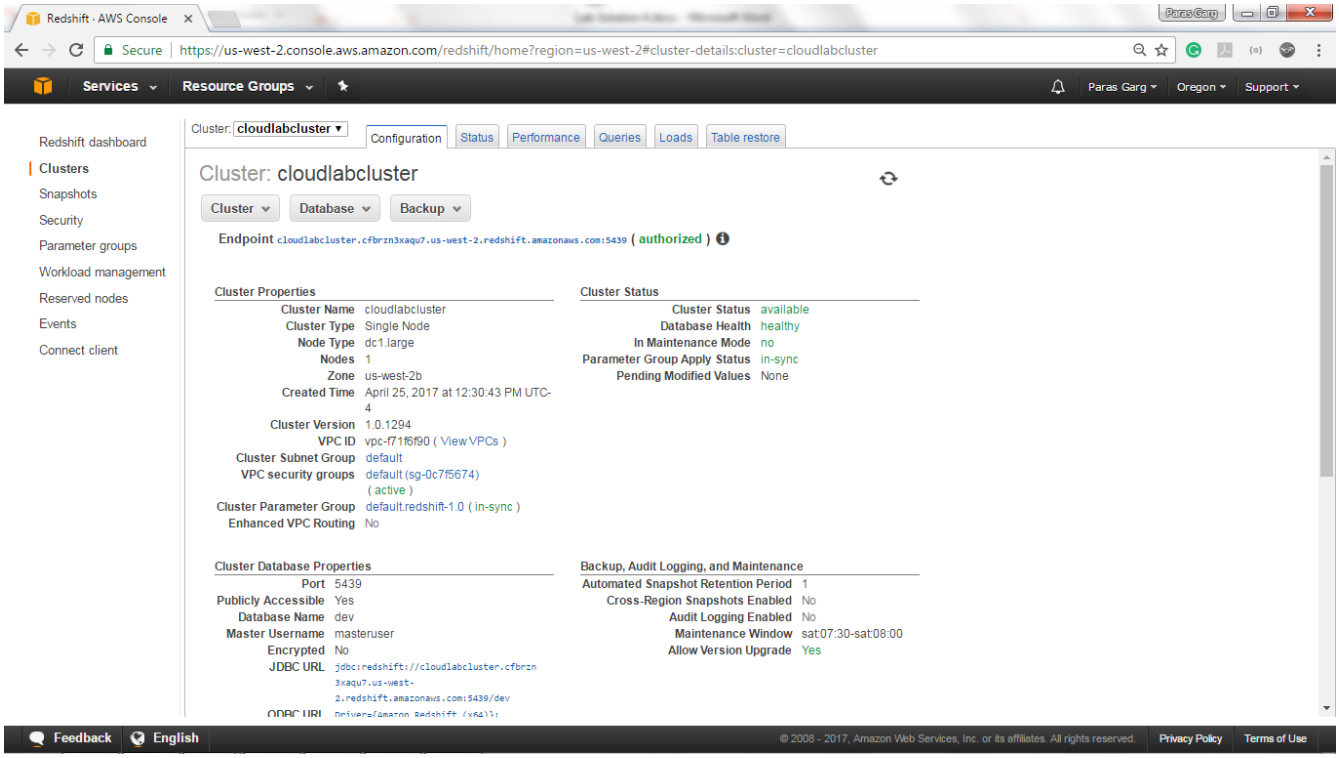
- In this step, we are allowed to review all the information and configuration that we passed in previous steps, and then finally clicked on Launch cluster to create one.

- After successfully launching a cluster, a notification popped-up

- We can also check and verify the created cluster under Clusters tab in RedShift dashboard.

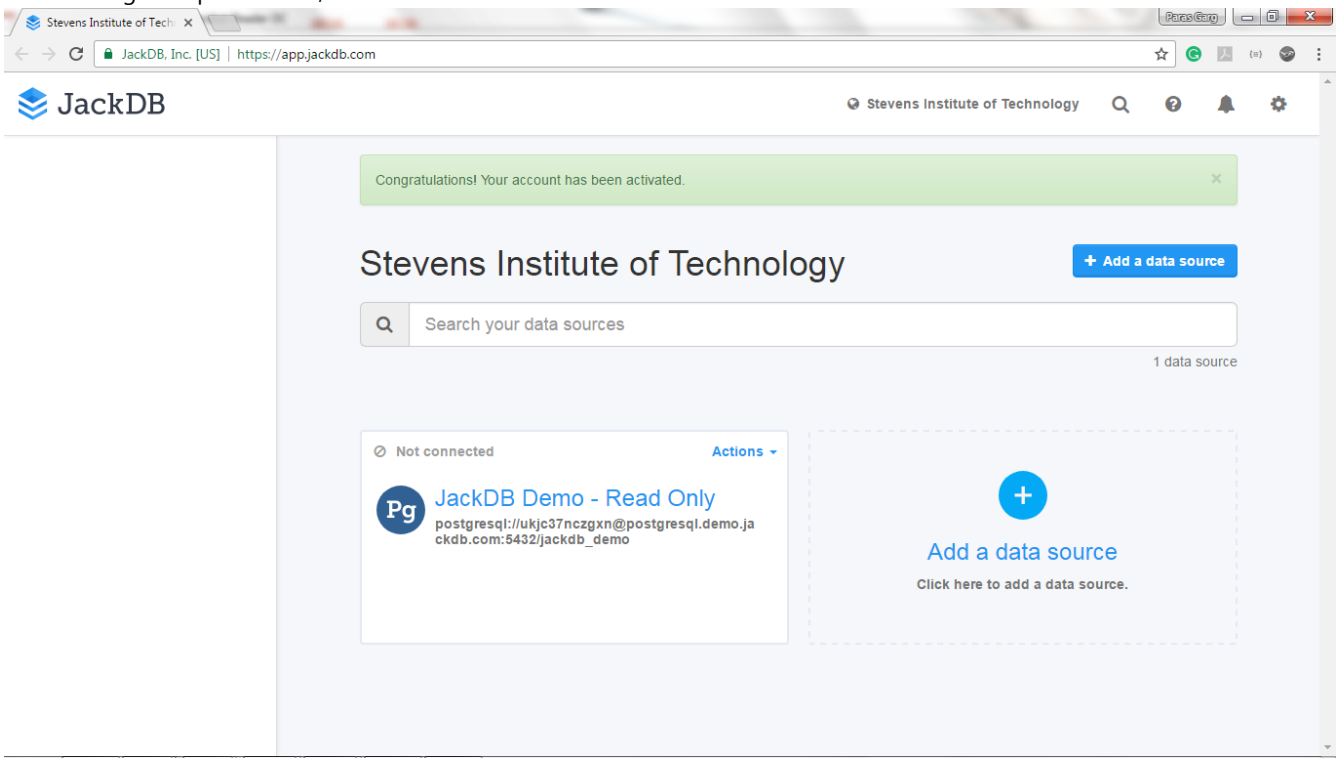


- We can also check all the configuration of the cluster we have create under Configuration tab.

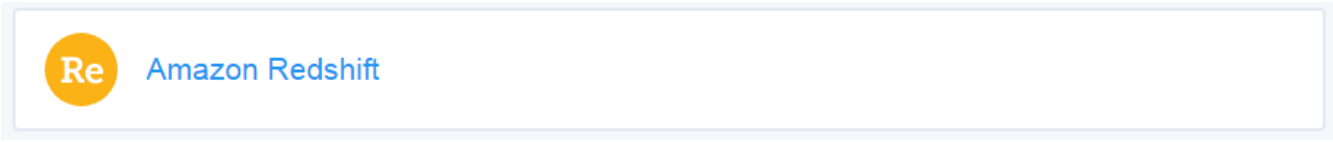


Step to enable secure connection to JackDB

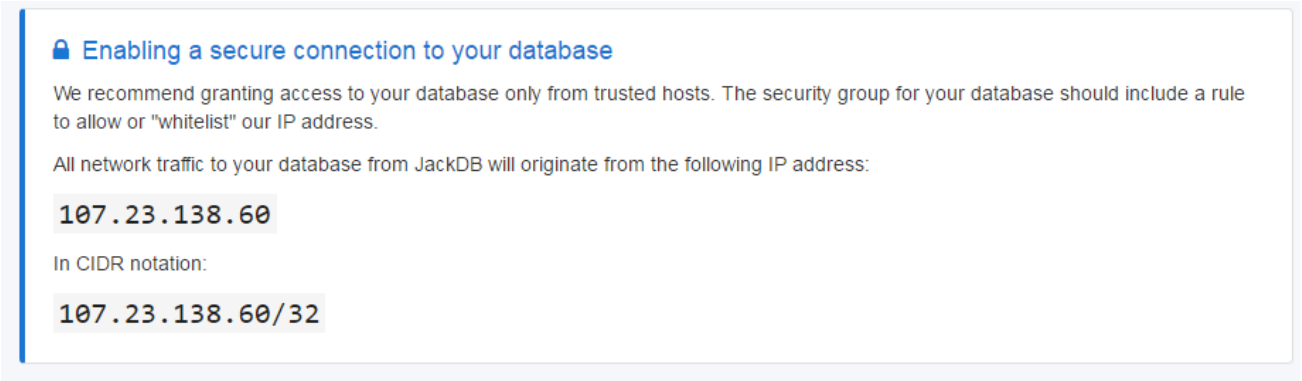
- After signed up in JackDB, I clicked on **Add a data source** button to add a data source.



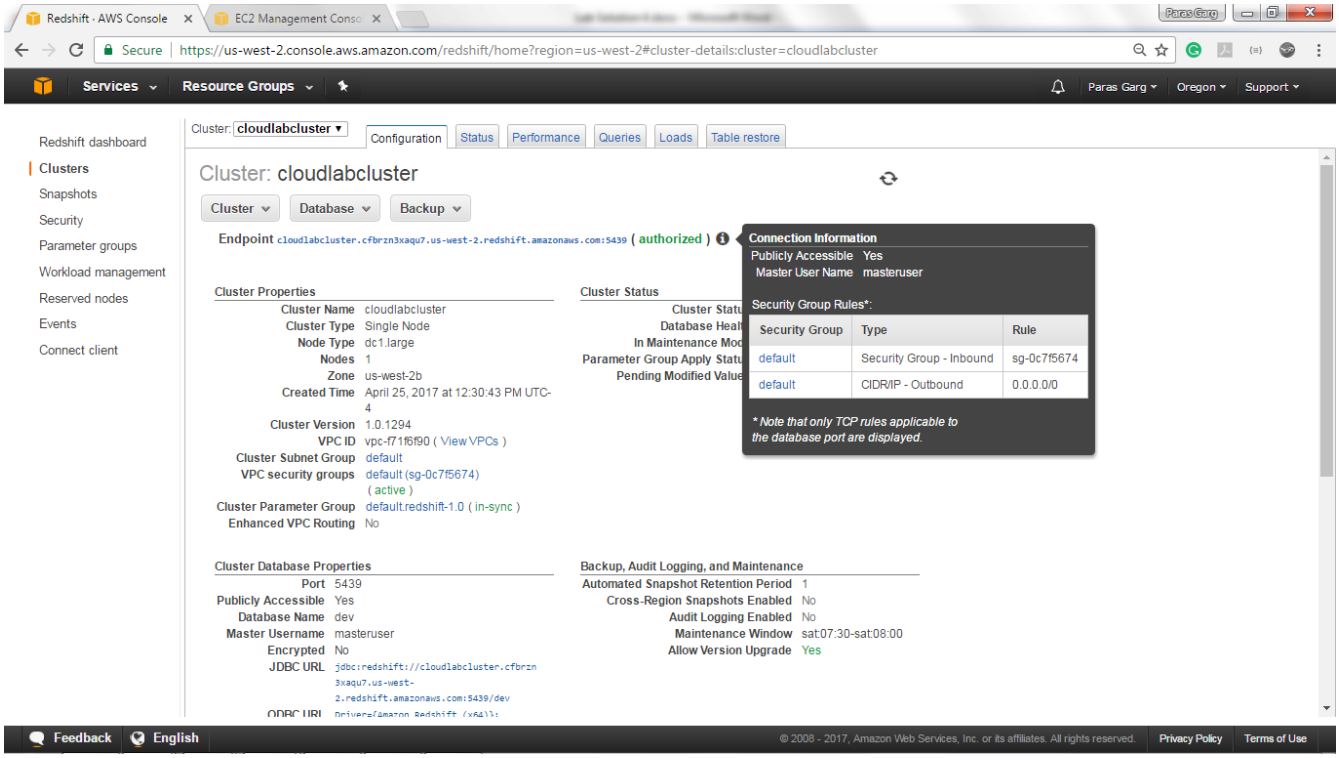
- Then selected Amazon RedShift data source



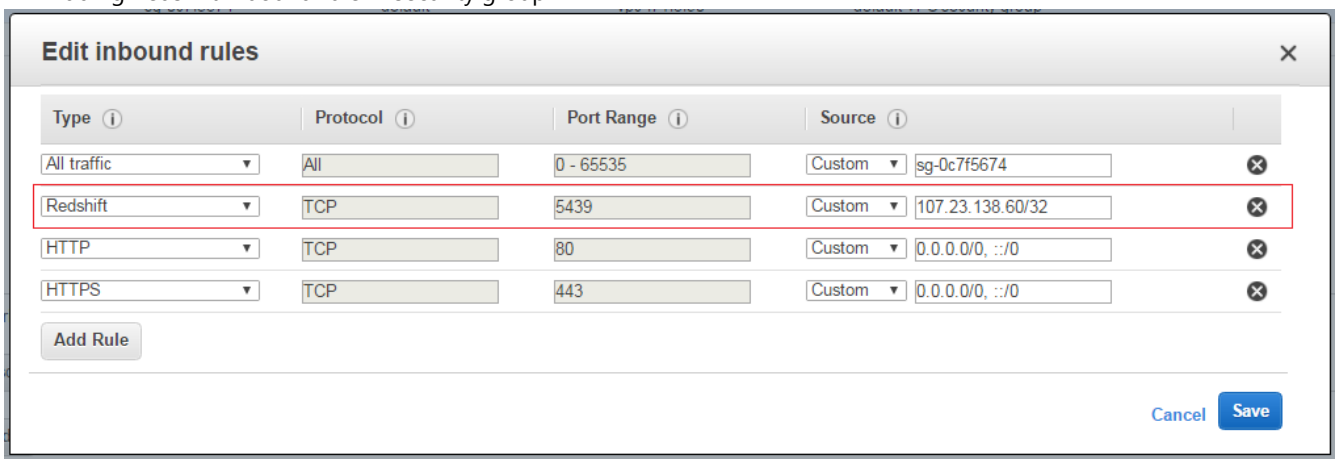
- Now, I have enabled a secure connection to my JackDB database. The IP address is mentioned below.



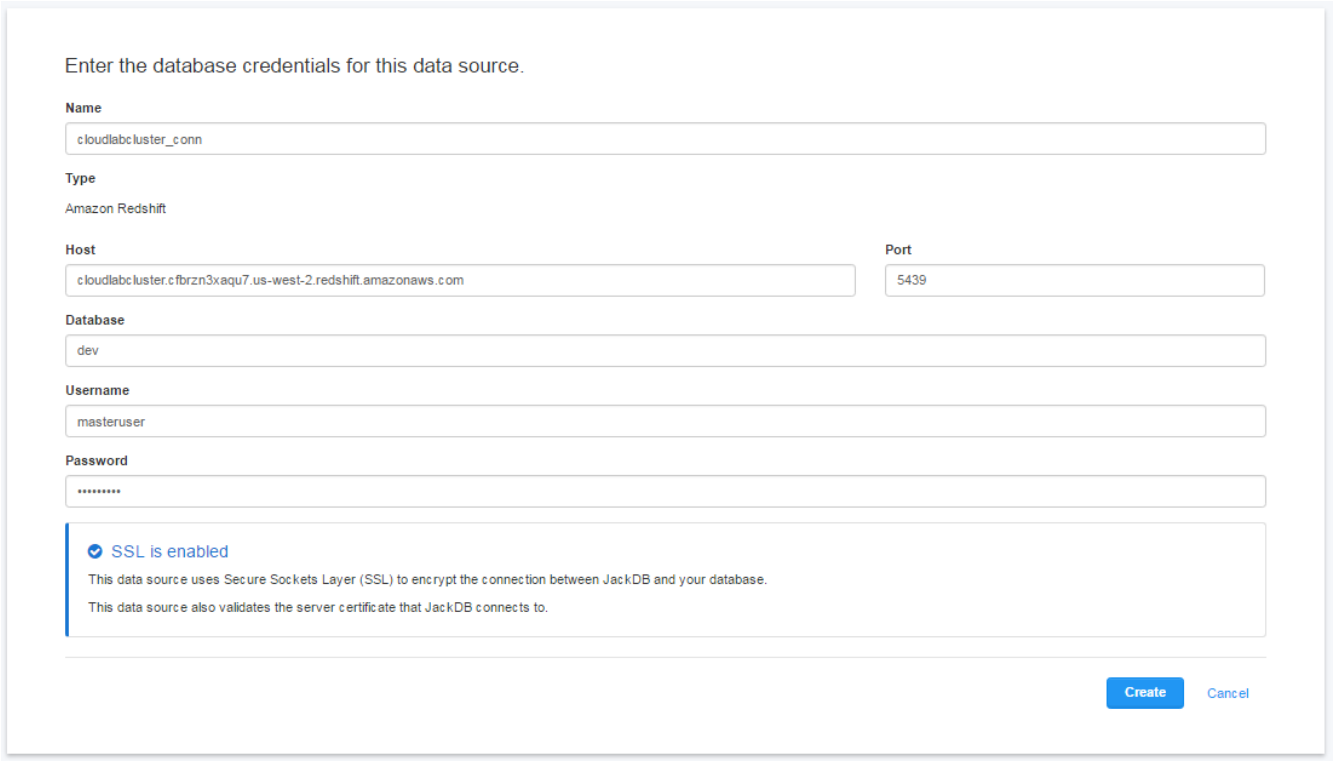
- As there is no inbound rule set to access my cluster. Adding, JackDB ip in the security group inbound rules.



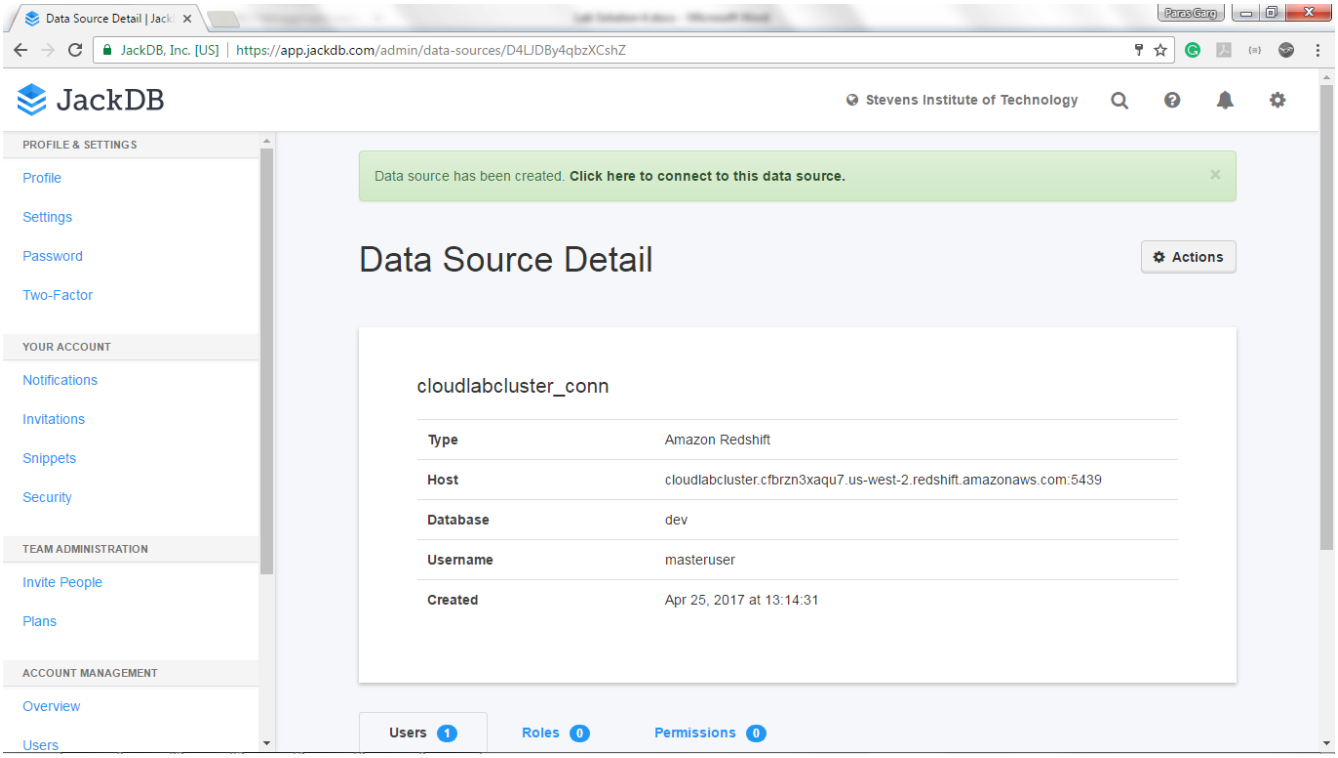
- Adding Redshift inbound rule in security group



- After adding security group, filling my cluster information to connect JackDB to RedShift cluster.



- Now, data source has been successfully created.



Step to load and check sample data

- Executing command for initial table creation in RedShift through JackDB

CloudLab-JackDB | Execute | Export | Explain

```
1  create table users( userid integer not null distkey sortkey,
2    username char(8), firstname varchar(30), lastname varchar(30), city varchar(30),
3    state char(2), email varchar(100), phone char(14),
4    likesports boolean, liketheatre boolean, likeconcerts boolean, likejazz boolean,
5    likeclassical boolean, likeopera boolean, likerock boolean, likevegas boolean,
6    likebroadway boolean, likemusicals boolean);
```

Executed successfully

Updated 0 rows in 0.836 seconds.

- Executing command to load sample data from S3 into my database table. Here we have to use our AWS credentials which we created in previous labs. We can also obtain this credentials from Your Security Credentials in AWS.

CloudLab-JackDB | Execute | Export | Explain

```
1  copy users from 's3://awssampledbuswest2/tickit/allusers_pipe.txt'
2
3  CREDENTIALS 'aws_access_key_id=AKIAJCYDAYVDUGT3KZ5A;aws_secret_access_key=OY0xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx1KZ'
4    delimiter '|';
```

Executed successfully

Updated 0 rows in 39.497 seconds.

- Executing commands 1 - to check whether the sample data has been loaded.

CloudLab-JackDB | Execute | Export | Explain

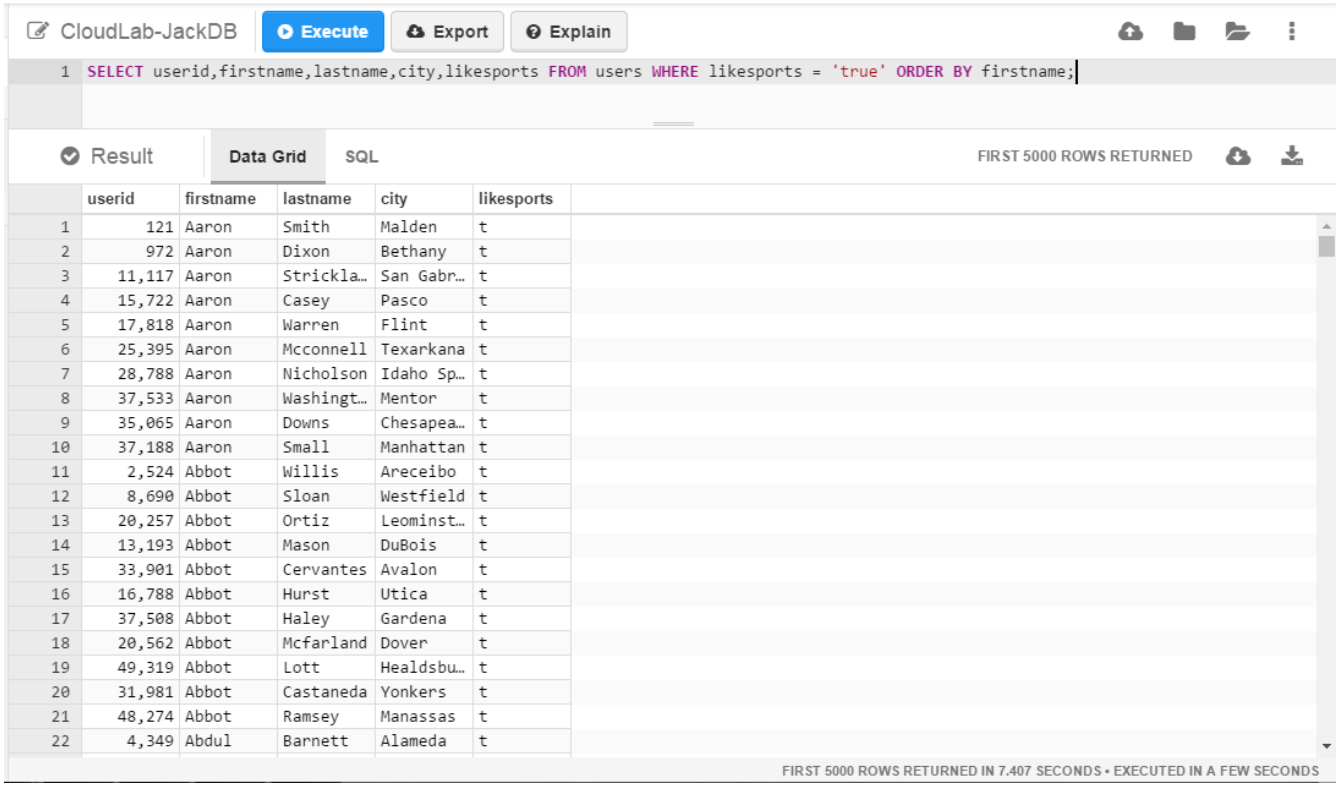
```
1  SELECT * from users;
```

Result | Data Grid | SQL — FIRST 5000 ROWS RETURNED

	userid	username	firstname	lastname	city	state	email	phone	likesports	liketheatre	likeconcerts	likejazz	likeclassi
1	2	PGL08LJI	Vladimir	Humphrey	Murfrees…	SK	Suspendi…	(783) 49…	null	null	null	t	t
2	4	XDZ38RDD	Barry	Roy	Omaha	AB	sed@lacu…	(355) 45…	f	t	null	f	nu
3	5	AEB55QTM	Reagan	Hodge	Forest L…	NS	Cum@accu…	(476) 51…	null	null	t	f	nu
4	7	OWY35QYB	Tamekah	Juarez	Moultrie	WV	elementu…	(297) 87…	null	null	null	t	t
5	9	MSD36KVR	Mufutau	Watkins	Port Orf…	MD	Integer.…	(725) 71…	t	f	null	f	t
6	10	WKW41AIW	Naida	Calderon	Waterbury	MB	Donec.fr…	(197) 72…	f	f	f	null	f
7	15	OWU78MTR	Scarlett	Mayer	Gadsden	GA	lorem.ip…	(189) 88…	t	f	t	null	nu
8	16	ZMG93CDD	Kieran	Drake	Hot Spri…	BC	molestie…	(192) 91…	null	t	t	null	f
9	18	VDP05MXU	Germaine	Valdez	Kokomo	WY	cursus.I…	(998) 87…	null	t	t	null	t
10	19	CXQ97IWP	Amal	Landry	Lomita	NT	euismod@…	(891) 52…	null	f	t	null	t
11	22	RHT62AGI	Hermione	Trevino	Walnut	WI	non.just…	(245) 11…	null	t	null	f	f
12	23	HOH86KXG	Vielka	Glass	Fort Wor…	MB	non.ante…	(170) 58…	null	null	null	null	nu
13	24	KJQ48LNY	Driscoll	Herring	Buffalo	CT	eros@bla…	(219) 70…	null	null	null	t	nu
14	25	IFJ28JVS	Ezra	Perry	Florence	NU	et@Nuncq…	(818) 37…	null	f	f	null	nu
15	27	KOY02CVE	Martina	Wiley	El Cerri…	NT	augue@Do…	(163) 56…	null	t	null	f	f
16	28	GOK50CHC	Kellie	Mcdowell	Henderson	NU	cursus@p…	(409) 92…	null	null	t	null	nu
17	29	HUH27PKK	Helen	Avery	Garland	PE	in.fauci…	(385) 92…	null	f	null	t	nu
18	33	PFD07GEF	Shafira	Glenn	Starkvil…	WA	at@Duis.…	(313) 67…	t	f	t	t	f
19	35	QRV60YSK	Kitra	Blanchard	Bellflow…	SK	quam@fac…	(319) 83…	f	null	null	null	nu
20	37	ETK12XWE	Renee	Hardin	Gardena	YT	risus.Do…	(337) 46…	f	null	null	f	nu
21	38	ZID76UPY	Lana	Ruiz	Bradbury	BC	molestie…	(479) 27…	null	t	f	null	f

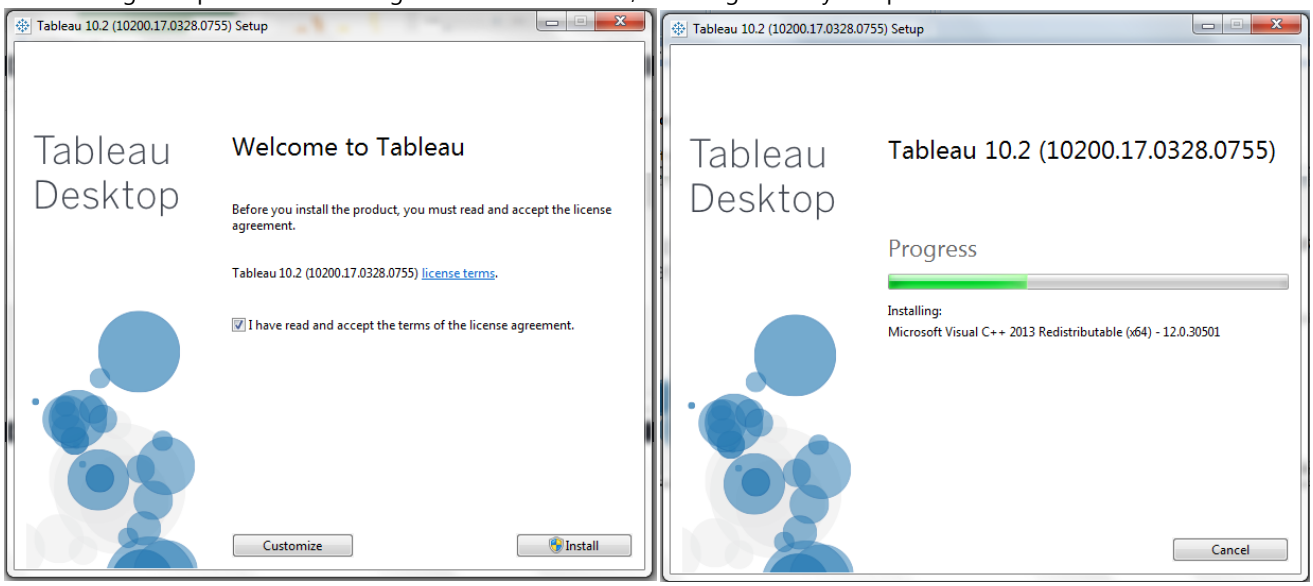
FIRST 5000 ROWS RETURNED IN 4.528 SECONDS • EXECUTED IN A FEW SECONDS

- Executing commands 2 - to check whether the sample data has been loaded.

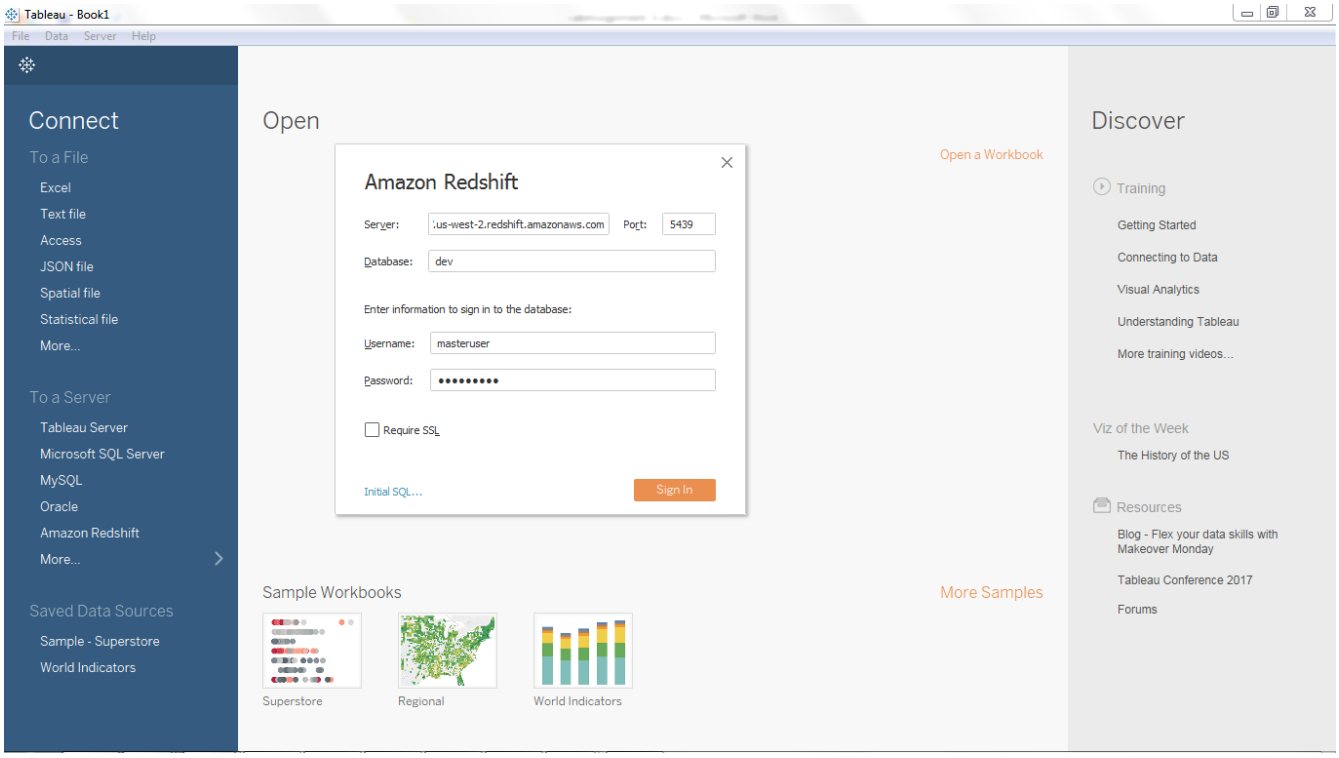


Step to connect and access from Tableau software

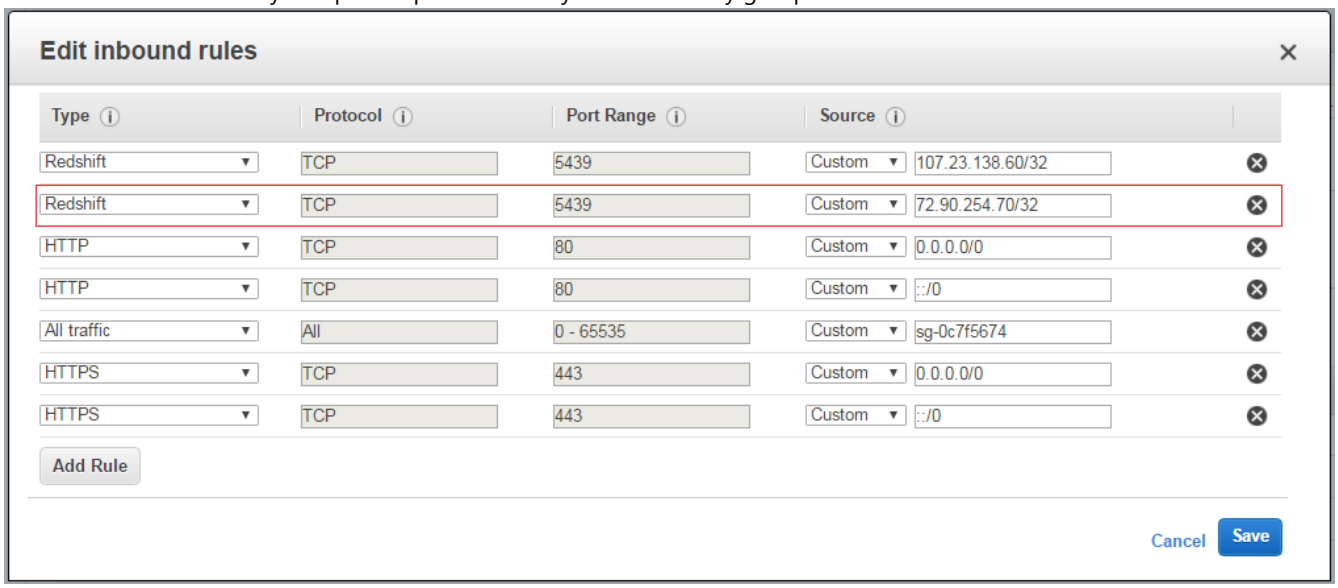
- After signed up and downloading the Tableau software, installing it on my computer



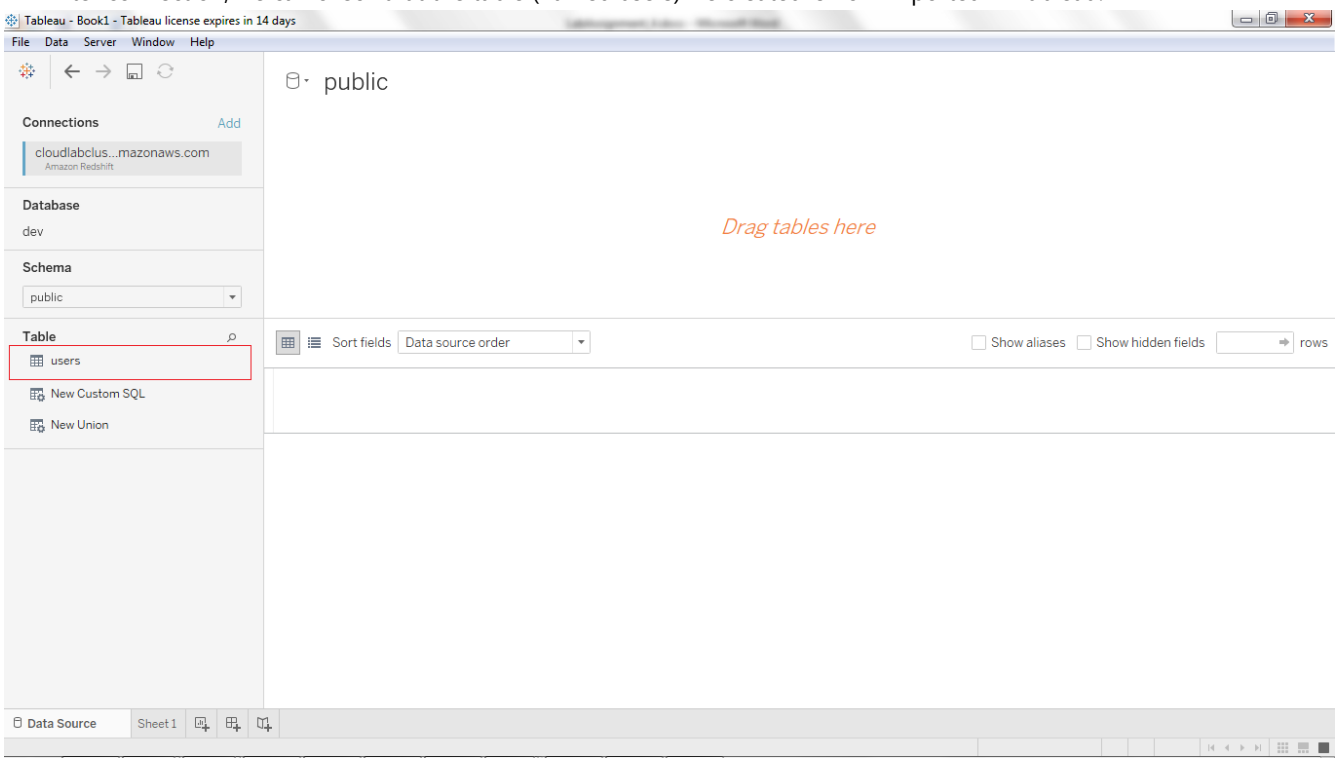
- After installing the software, connecting the software to Amazon Redshift by prompting AWS RedShift credentials. Then, clicked on sign in.



- While connecting Tableau to Amazon RedShift, I got an error saying unable to communicate with RedShift. To debud the error I have added my computer's public IP to my cluster security group. Then it connected.

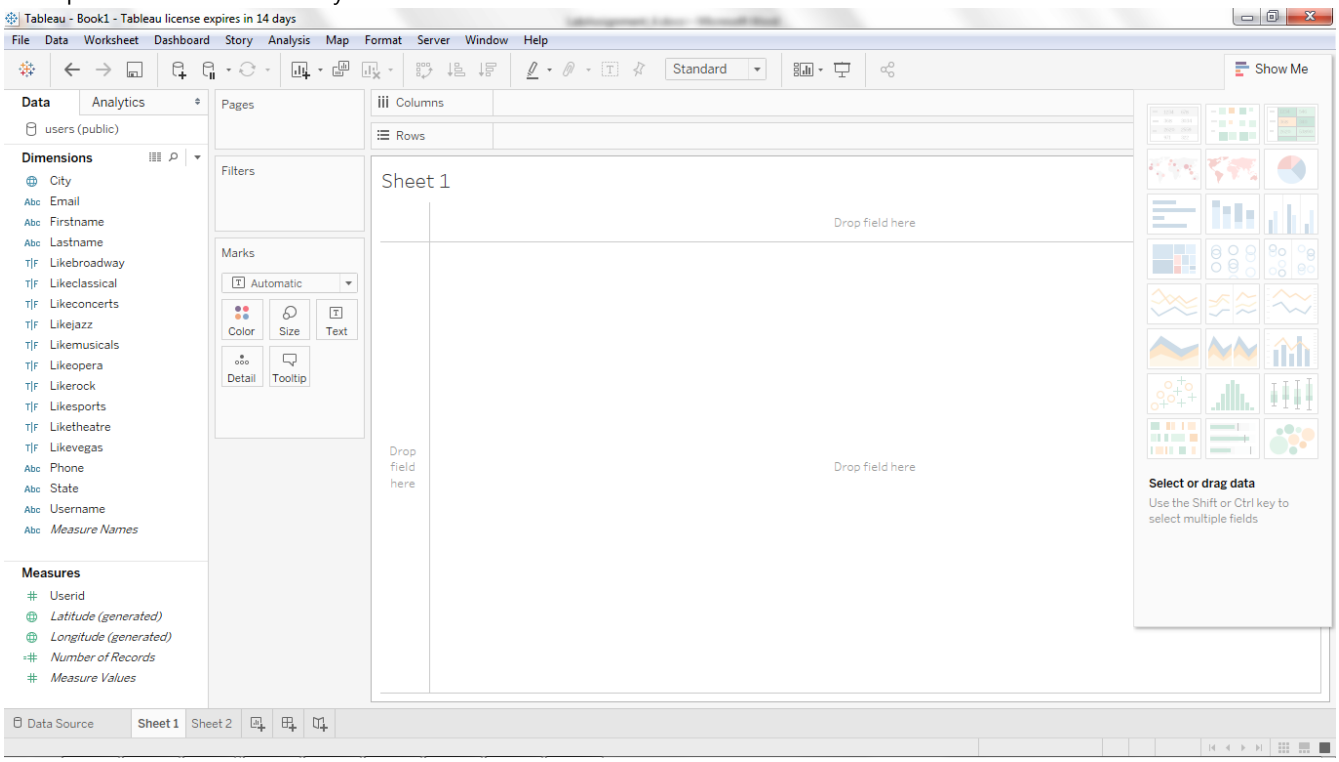


- After connection, we can check that the table (named users) we created is now imported in Tableau.

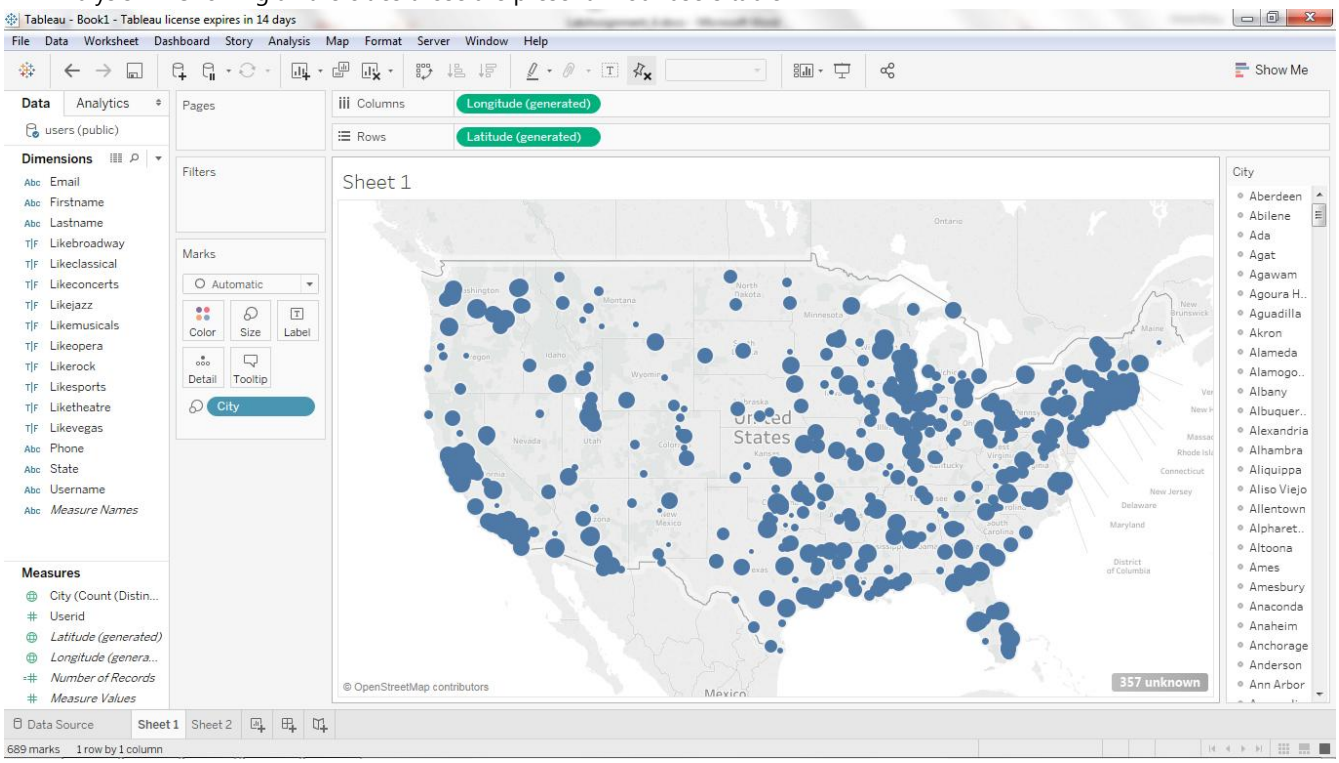


Step to analyze data and create charts using Tableau

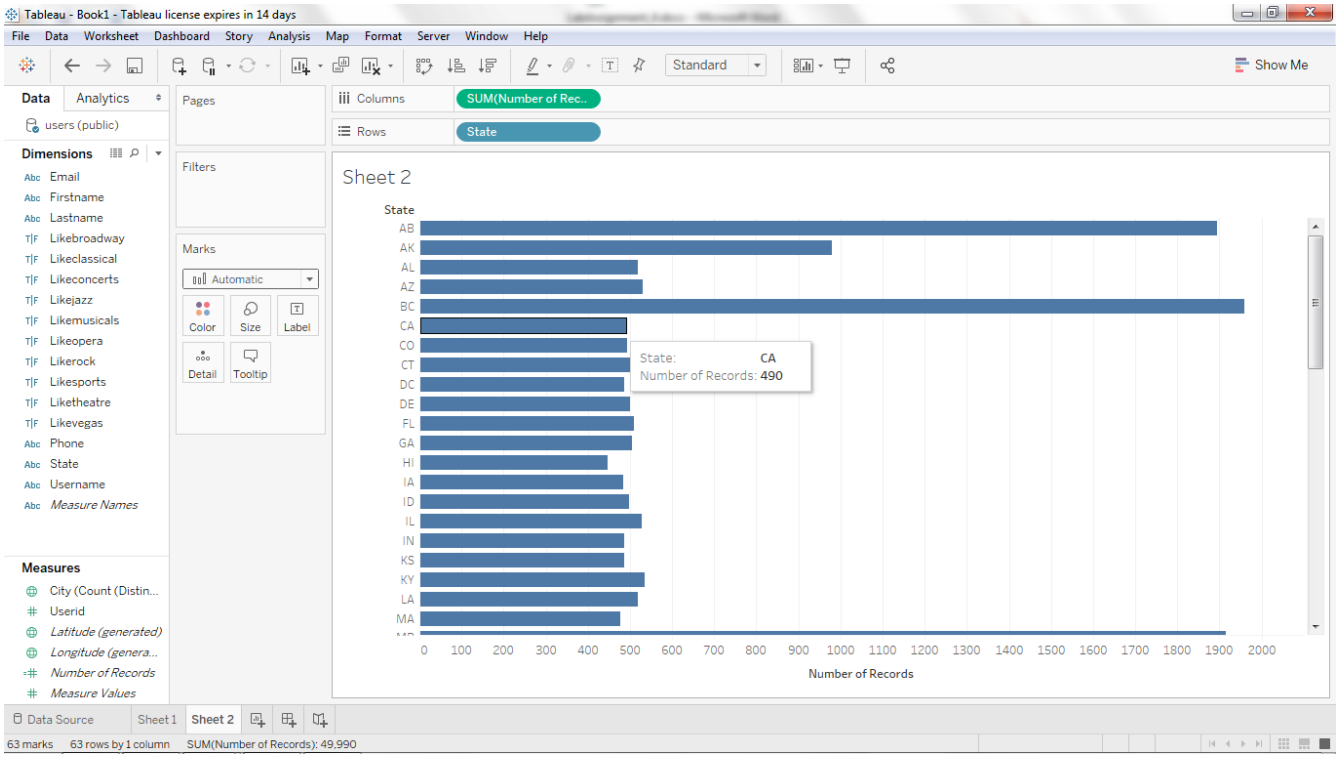
- Opened new sheet for analysis and chart creation



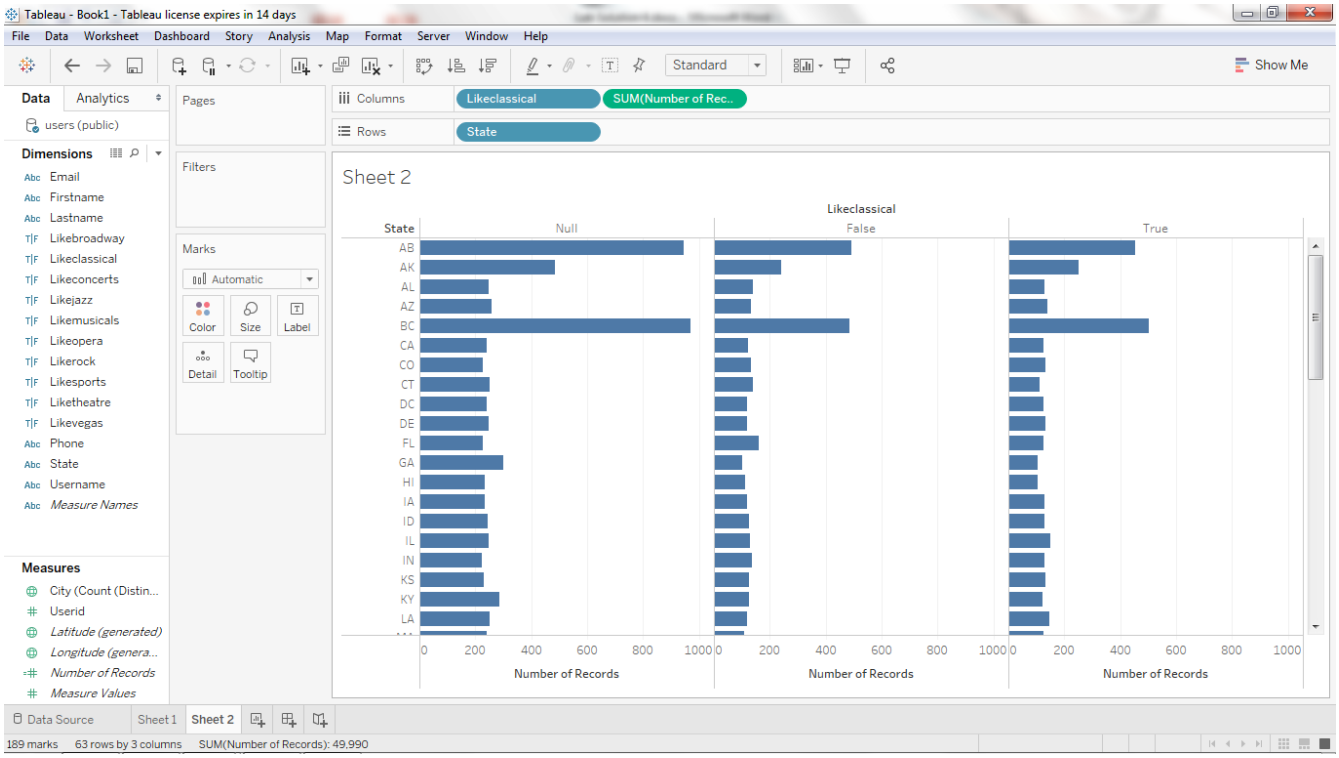
- Analysis 1 – Showing all the cities those are present in our users table



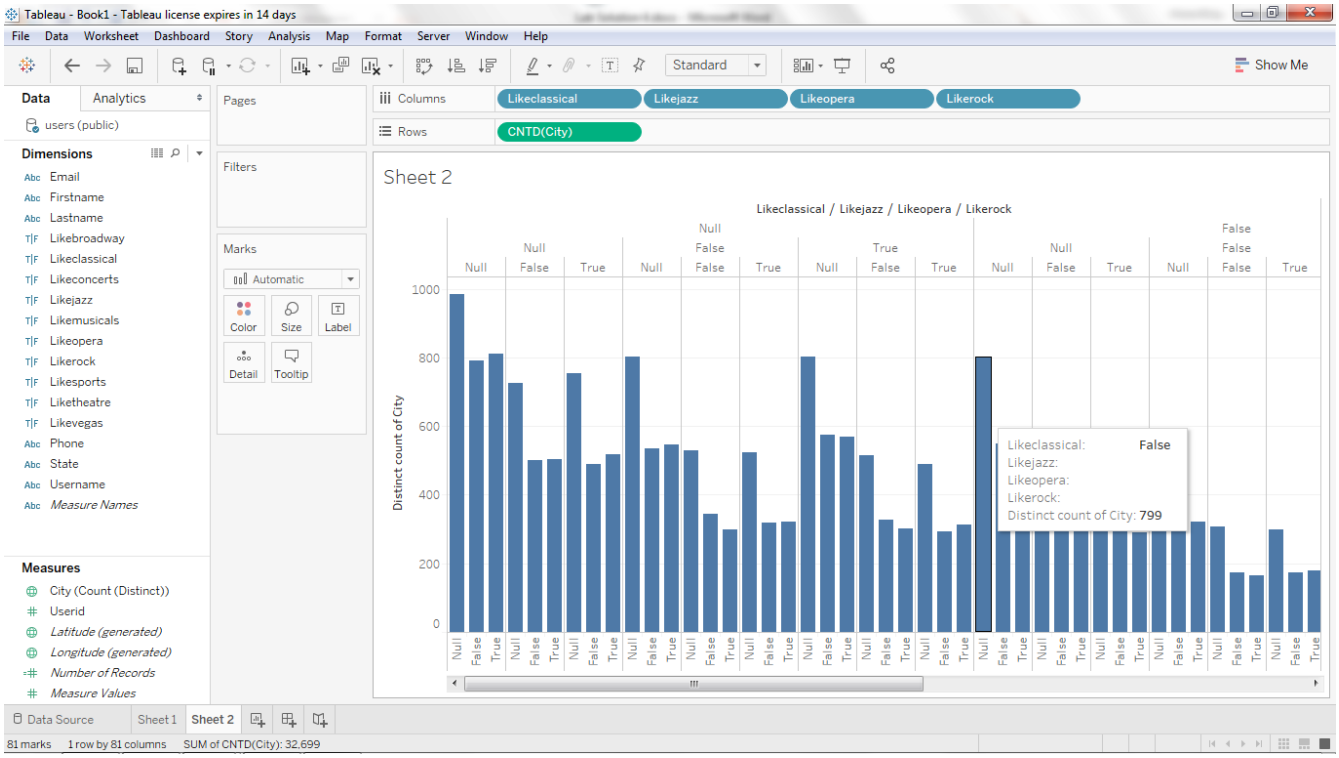
- Analysis 2 – It shows the list of state are present in number of records in users table



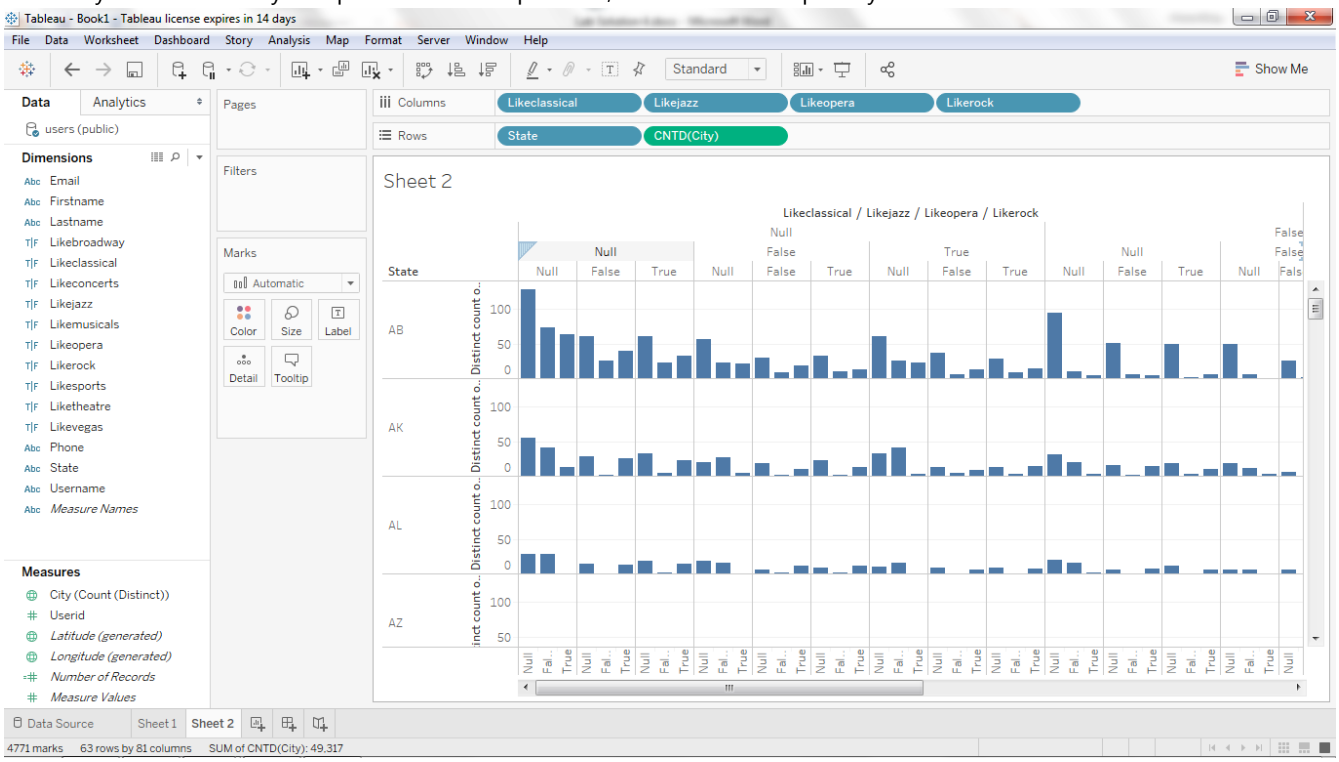
- Analysis 3 – In this report, I have combined more dimensions, that shows the total records have likeclassical as respect to each state.



- Analysis 4 – This report show the interest in different kind of music like in different cities.



- Analysis 5 – This analysis report is extend of previous, it bifurcate the report by states too.



- Now after performing the entire tasks. I have deleted the cluster successfully.