

Introduction to Cloud Computing (CS 524)

(Lab Assignment 4)

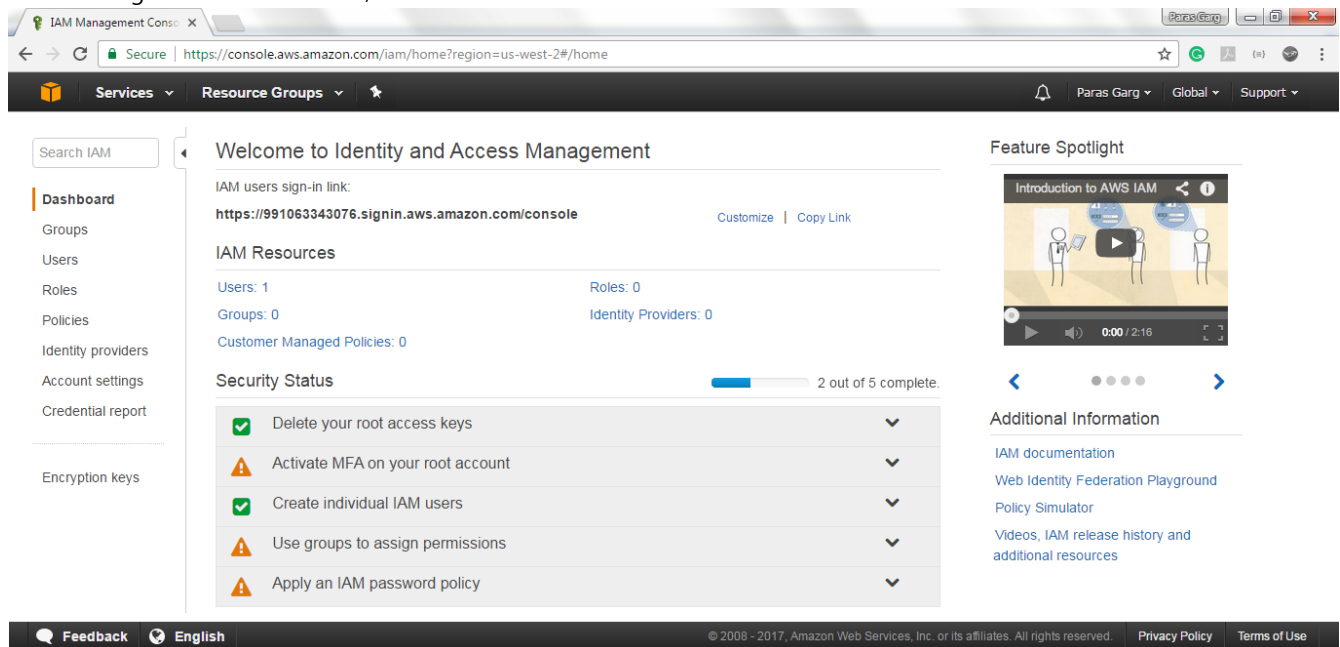
Prof. Igor Faynberg

Student Name: **Paras Garg**

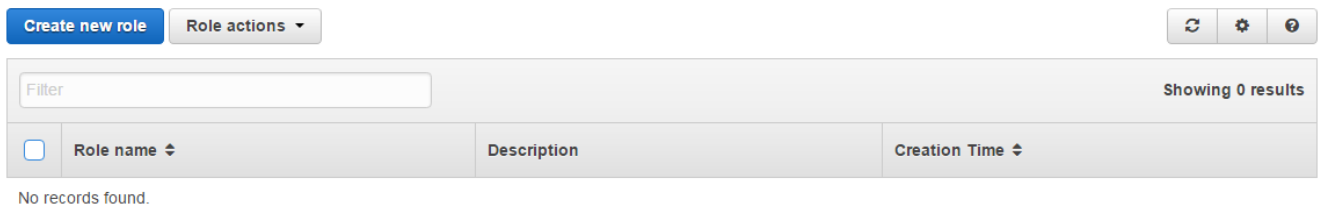
Course Section: **CS 524-A**

Step to create IAM role

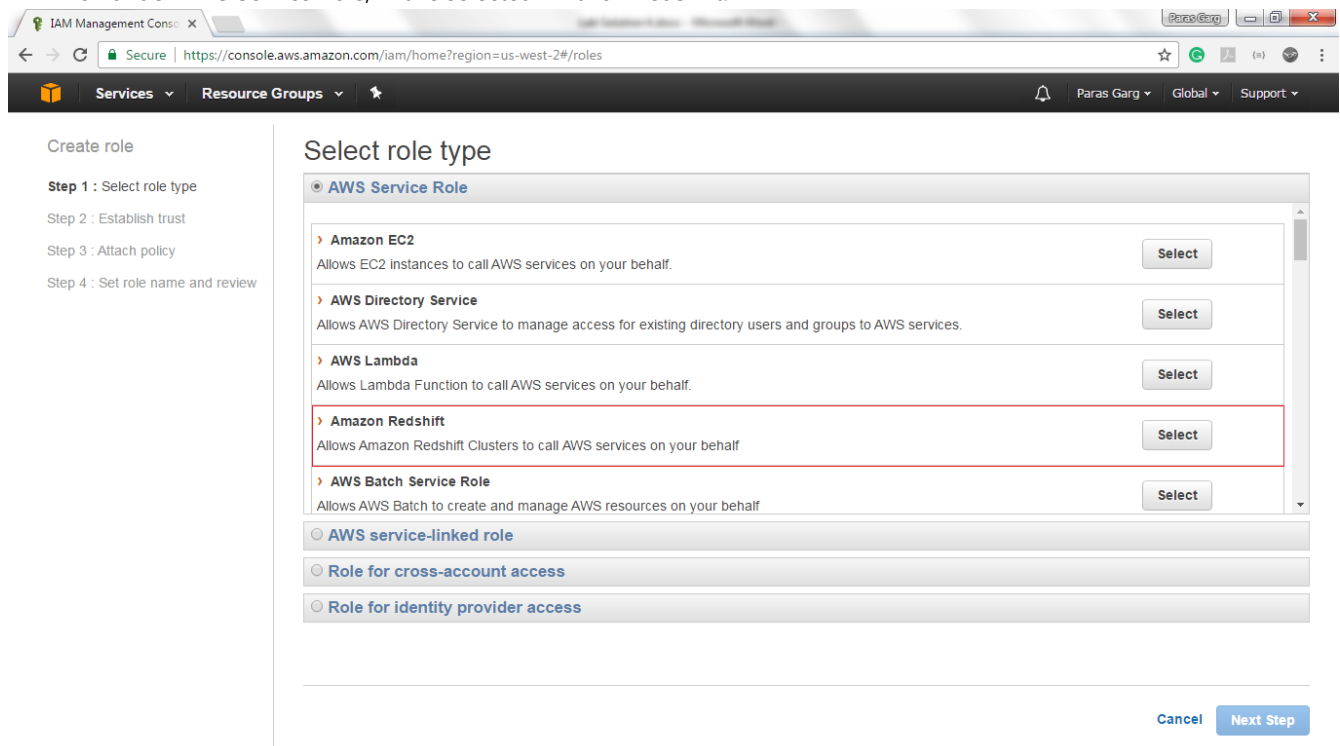
- After login into AWS RedShift, chose IAM service from the dashboard to land on IAM dashboard.



- 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.

The screenshot shows the 'Attach Policy' step in the AWS IAM console. The left sidebar indicates the progress: Step 1: Select role type, Step 2: Establish trust, Step 3: Attach policy (current), and Step 4: Set role name and review. The main content area is titled 'Attach Policy' and includes the instruction: 'Select one or more policies to attach. Each role can have up to 10 policies attached.' Below this is a table of available policies. The table has columns for 'Policy Name', 'Attached Entities', 'Creation Time', and 'Edited Time'. The first four policies are selected with checkboxes: AdministratorAccess, AmazonS3FullAccess, AmazonS3ReadOnlyAccess, and IAMUserChangePassword. The bottom right of the table has 'Cancel', 'Previous', and 'Next Step' buttons.

Policy Name	Attached Entities	Creation Time	Edited Time
<input checked="" type="checkbox"/> AdministratorAccess	1	2015-02-06 13:39 EDT	2015-02-06 13:39 EDT
<input checked="" type="checkbox"/> AmazonS3FullAccess	1	2015-02-06 13:40 EDT	2015-02-06 13:40 EDT
<input checked="" type="checkbox"/> AmazonS3ReadOnlyAccess	1	2015-02-06 13:40 EDT	2015-02-06 13:40 EDT
<input checked="" type="checkbox"/> IAMUserChangePassword	1	2016-11-14 19:25 EDT	2016-11-15 18:18 EDT
<input type="checkbox"/> AmazonAPIGatewayAdminstr...	0	2015-07-09 13:34 EDT	2015-07-09 13:34 EDT
<input type="checkbox"/> AmazonAPIGatewayInvokeFul...	0	2015-07-09 13:36 EDT	2015-07-09 13:36 EDT
<input type="checkbox"/> AmazonAPIGatewayPushToCL...	0	2015-11-11 18:41 EDT	2015-11-11 18:41 EDT
<input type="checkbox"/> AmazonAppStreamFullAccess	0	2015-02-06 13:40 EDT	2015-02-06 13:40 EDT
<input type="checkbox"/> AmazonAppStreamReadOnlyA...	0	2015-02-06 13:40 EDT	2016-12-07 16:00 EDT
<input type="checkbox"/> AmazonAppStreamServiceAcc...	0	2016-11-18 23:17 EDT	2016-11-18 23:17 EDT
<input type="checkbox"/> AmazonAthenaFullAccess	0	2016-11-30 11:46 EDT	2016-11-30 11:46 EDT

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

The screenshot shows the 'Set role name and review' step in the AWS IAM console. The left sidebar indicates the progress: Step 1: Select role type, Step 2: Establish trust, Step 3: Attach policy, and Step 4: Set role name and review (current). The main content area is titled 'Set role name and review' and includes the instruction: 'Review the following role information. To edit the role, click an edit link, or click **Create role** to finish.' Below this are four sections: 'Role name' (CloudLabRedShift), 'Role description' (Allows Amazon Redshift Clusters to call AWS services on your behalf), 'Trusted entities' (The identity provider(s) redshift.amazonaws.com), and 'Policies' (arn:aws:iam::aws:policy/AmazonS3FullAccess, arn:aws:iam::aws:policy/AmazonS3ReadOnlyAccess, arn:aws:iam::aws:policy/AdministratorAccess, arn:aws:iam::aws:policy/IAMUserChangePassword). The bottom right has 'Cancel', 'Previous', and 'Create role' buttons.

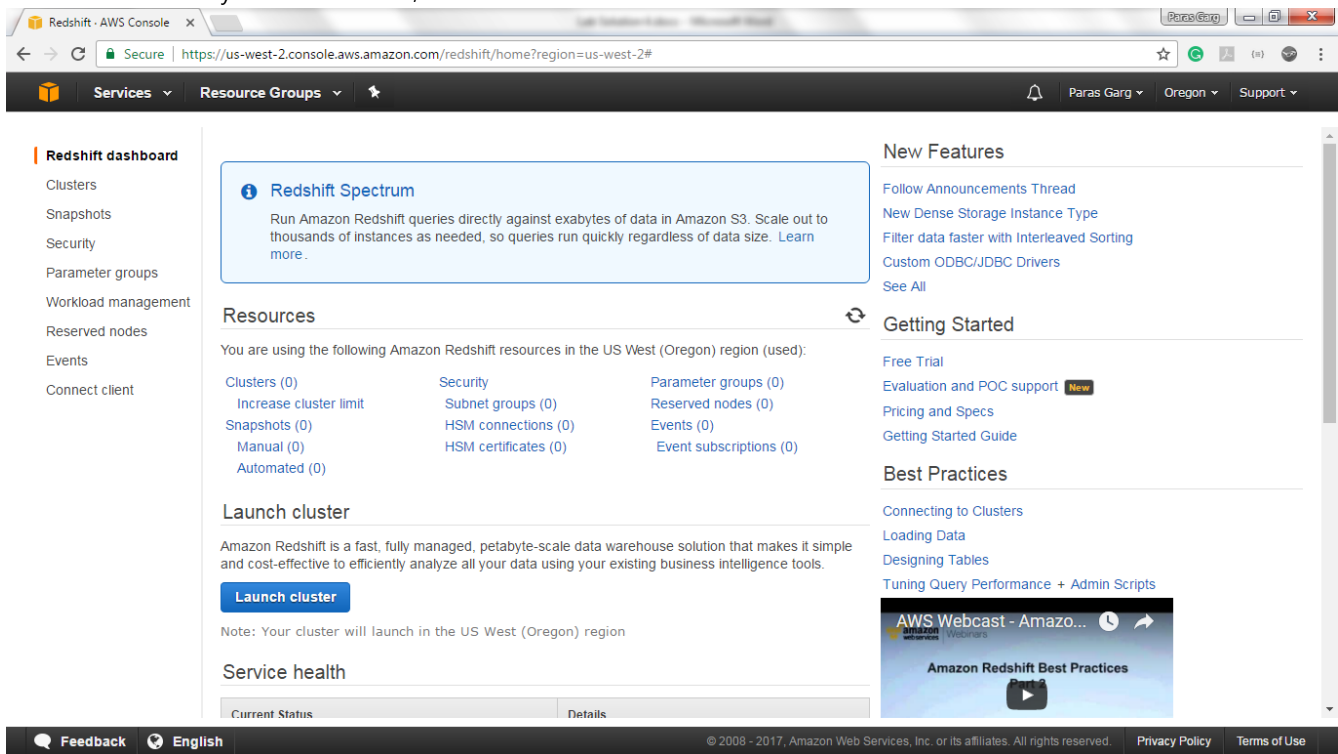
- Now, under the roles tab, we can check that the role has been created successfully

The screenshot shows the 'Roles' tab in the AWS IAM console. At the top, there are buttons for 'Create new role' and 'Role actions'. Below is a table of roles. The table has columns for 'Role name', 'Description', and 'Creation Time'. The first role listed is 'CloudLabRedShift' with the description 'Allows Amazon Redshift Clusters to call AWS servi...' and creation time '2017-04-25 12:08 EDT'. The bottom right of the table has 'Showing 1 results'.

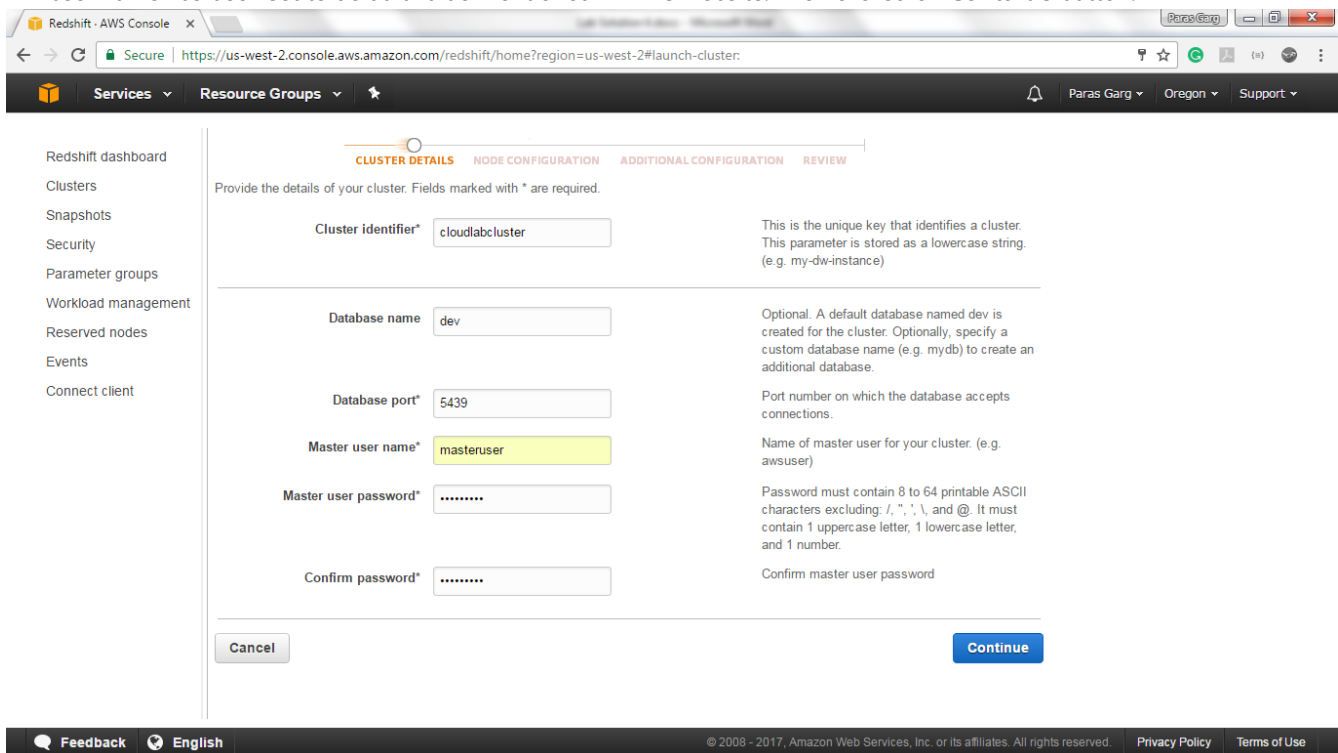
Role name	Description	Creation Time
<input checked="" type="checkbox"/> CloudLabRedShift	Allows Amazon Redshift Clusters to call AWS servi...	2017-04-25 12:08 EDT

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 on **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.



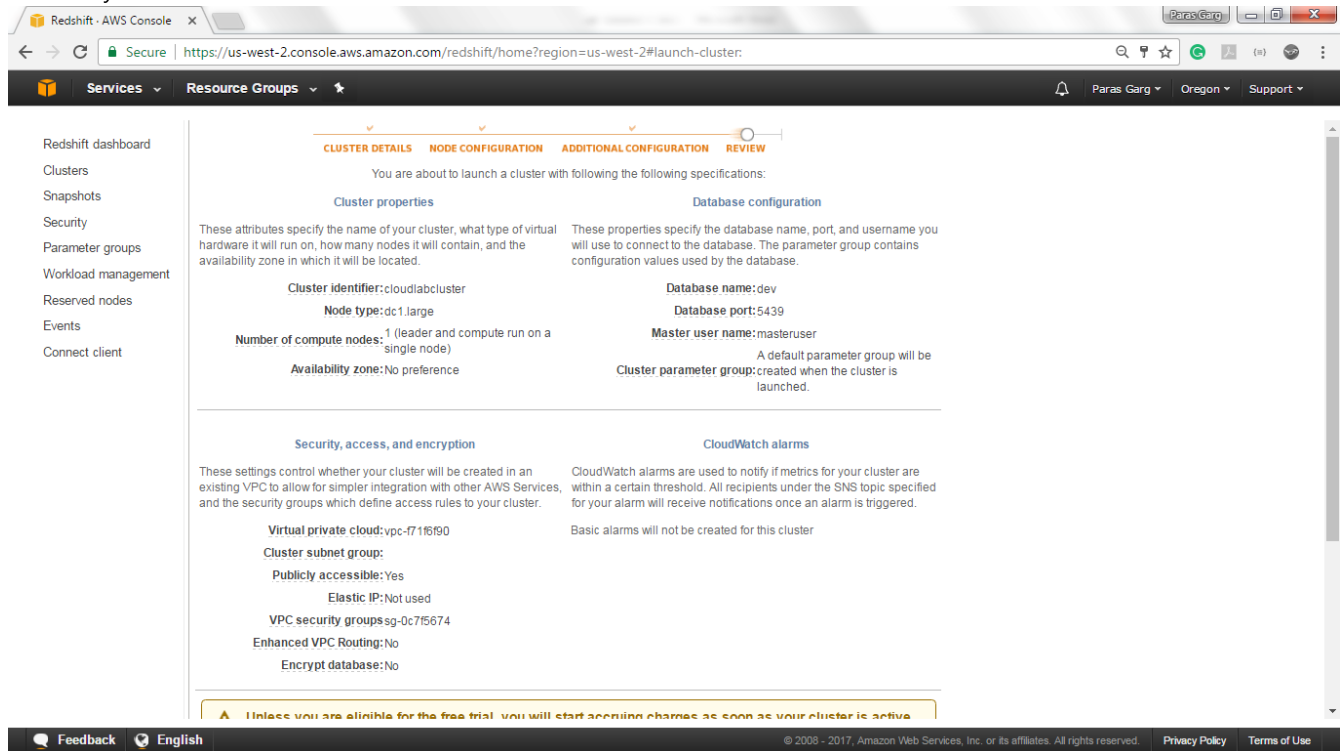
- In this step, checking for Node Configuration. (passed default values)

The screenshot shows the AWS Redshift console with the 'NODE CONFIGURATION' step selected. The interface includes a sidebar with navigation links like 'Redshift dashboard', 'Clusters', 'Snapshots', 'Security', 'Parameter groups', 'Workload management', 'Reserved nodes', 'Events', and 'Connect client'. The main content area has a progress bar with four steps: 'CLUSTER DETAILS', 'NODE CONFIGURATION' (active), 'ADDITIONAL CONFIGURATION', and 'REVIEW'. Below the progress bar, a message states: 'Choose a number of nodes and node type below. Number of Compute Nodes is required for multi-node clusters.' A blue callout box contains the text: 'The ds2 node types replace the deprecated ds1 node types. The newer ds2 node types provide higher performance than ds1 at no extra cost. [Learn more.](#)' The configuration fields are: 'Node type' set to 'dc1.large' (with a description: 'Specifies the compute, memory, storage, and I/O capacity of the cluster's nodes.'), 'CPU' set to '7 EC2 Compute Units (2 virtual cores) per node', 'Memory' set to '15 GiB per node', 'Storage' set to '160GB SSD storage per node', 'I/O performance' set to 'Moderate', 'Cluster type' set to 'Single Node', 'Number of compute nodes*' set to '1' (with a description: 'Single Node clusters consist of a single node which performs both leader and compute functions.'), 'Maximum' set to '1', and 'Minimum' set to '1'. The footer includes 'Feedback', 'English', and copyright information for Amazon Web Services, Inc. or its affiliates.

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

The screenshot shows the AWS Redshift console with the 'ADDITIONAL CONFIGURATION' step selected. The interface is similar to the previous screenshot, with the 'ADDITIONAL CONFIGURATION' step now active in the progress bar. The main content area has a message: 'Provide the optional additional configuration details below.' The configuration fields are: 'Cluster parameter group' with a note 'A default parameter group will be associated with this cluster.', 'Encrypt database' with radio buttons for 'None' (selected), 'KMS', and 'HSM', and a link 'Learn more about database encryption', 'Configure networking options:' section with 'Choose a VPC' set to 'Default VPC (vpc-f71f6f90)' (with a note: 'The identifier of the VPC in which you want to create your cluster'), 'Cluster subnet group' set to 'default' (with a note: 'Selected Cluster Subnet Group may limit the choice of Availability Zones'), 'Publicly accessible' with radio buttons for 'Yes' and 'No' (selected 'No', with a note: 'Select Yes if you want the cluster to be accessible from the public internet. Select No if you want it to be accessible only from within your private VPC network.'), 'Choose a public IP address' with radio buttons for 'Yes' and 'No' (selected 'No', with a note: 'Select Yes if you want the cluster to have a public IP address that can be accessed from the public Internet, select No if you want the cluster to have a private IP address that can only be accessed from within the VPC.'), 'Enhanced VPC Routing' with radio buttons for 'Yes' and 'No' (selected 'No', with a link 'Learn more'), 'Availability zone' set to 'No Preference' (with a note: 'The EC2 Availability Zone that the cluster will be created in.'), 'Optionally, associate your cluster with one or more security groups.' section with 'VPC security groups' set to a list containing 'default (sg-0c7f5674)' and 'default_elb_aae997d1-3a34-3af0-9c21-04dc6a6a0d1e (sg-...)' (with a note: 'List of VPC security groups to associate with this cluster.'), 'Optionally, create a basic alarm for this cluster.' section with 'Create CloudWatch Alarm' radio buttons for 'Yes' and 'No' (selected 'No', with a note: 'Create a CloudWatch alarm to monitor the disk usage of your cluster.'), and a red-bordered box at the bottom with the message 'Optionally, associate your cluster with one or more security groups.' and 'Available roles' set to 'Choose a role' (with a note: 'List of IAM roles to associate with this cluster.') and a list containing 'CloudLabRedShift'.

- 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



Cluster **cloudlabcluster** is being created

Note: Your cluster may take a few minutes to launch.

[View your cluster on the Clusters dashboard.](#)

Close

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

Clusters

Launch Cluster Manage Tags Manage IAM roles					
<div> <div></div> <div></div> <div></div> </div>					
	Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
<input type="checkbox"/>	cloudlabcluster	available	healthy	no	3

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

The screenshot displays the AWS Redshift console interface. The top navigation bar includes the AWS logo, 'Services', 'Resource Groups', and a user profile 'Paras Garg' in the 'Oregon' region. The main content area is titled 'Cluster: cloudlabcluster' and features tabs for 'Configuration', 'Status', 'Performance', 'Queries', 'Loads', and 'Table restore'. The 'Configuration' tab is active, showing various cluster details.

Cluster: cloudlabcluster

Endpoint: cloudlabcluster.cfbrzn3xaqu7.us-west-2.redshift.amazonaws.com:5439 (authorized)

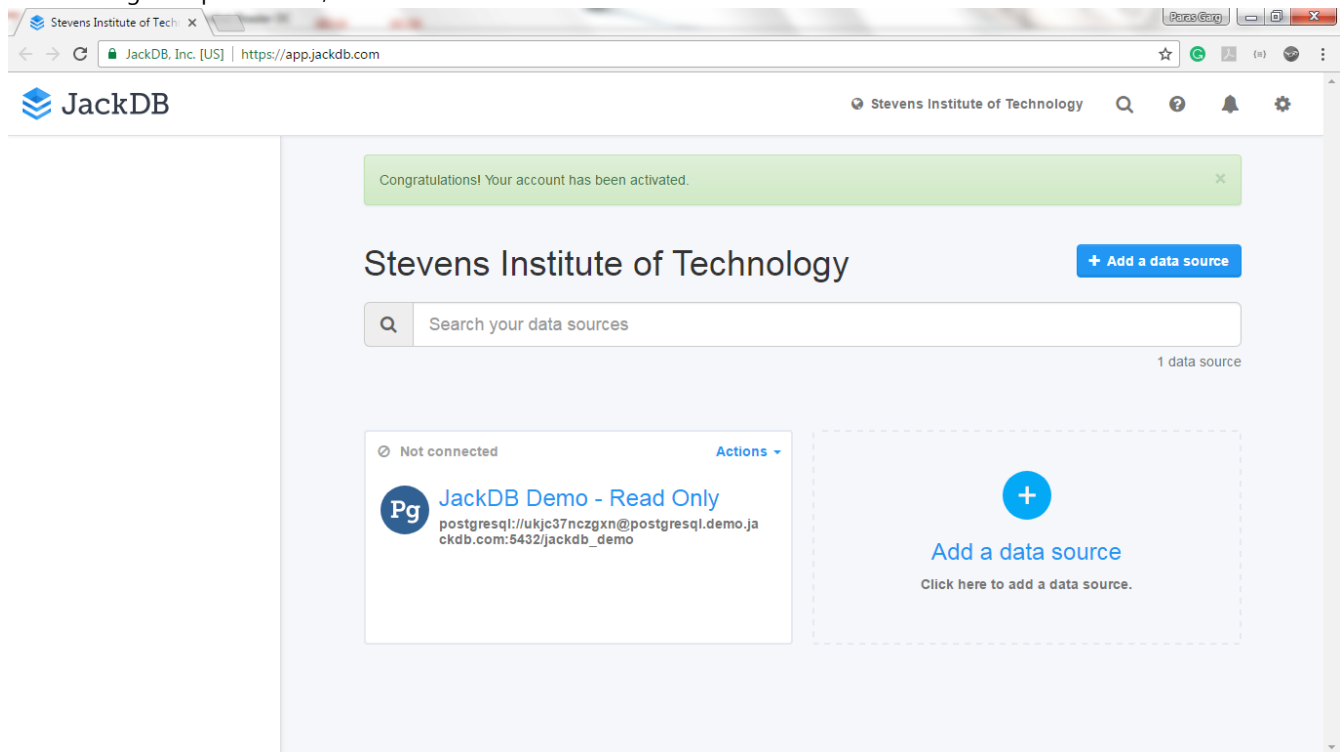
Cluster Properties		Cluster Status	
Cluster Name	cloudlabcluster	Cluster Status	available
Cluster Type	Single Node	Database Health	healthy
Node Type	dc1.large	In Maintenance Mode	no
Nodes	1	Parameter Group Apply Status	in-sync
Zone	us-west-2b	Pending Modified Values	None
Created Time	April 25, 2017 at 12:30:43 PM UTC-4		
Cluster Version	1.0.1294		
VPC ID	vpc-f71f6f90 (View VPCs)		
Cluster Subnet Group	default		
VPC security groups	default (sg-0c7f5674) (active)		
Cluster Parameter Group	default:redshift-1.0 (in-sync)		
Enhanced VPC Routing	No		

Cluster Database Properties		Backup, Audit Logging, and Maintenance	
Port	5439	Automated Snapshot Retention Period	1
Publicly Accessible	Yes	Cross-Region Snapshots Enabled	No
Database Name	dev	Audit Logging Enabled	No
Master Username	masteruser	Maintenance Window	sat:07:30-sat:08:00
Encrypted	No	Allow Version Upgrade	Yes
JDBC URL	jdbc:redshift://cloudlabcluster.cfbrzn3xaqu7.us-west-2.redshift.amazonaws.com:5439/dev		
ODBC DRI	drivers/amazon-redshift (x64)		

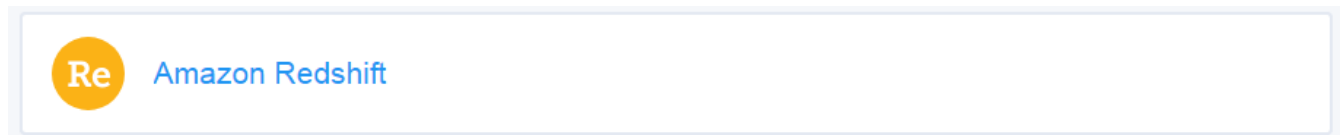
The footer of the console includes a 'Feedback' button, the language 'English', and copyright information: '© 2008 - 2017, Amazon Web Services, Inc. or its affiliates. All rights reserved.' along with links to 'Privacy Policy' and 'Terms of Use'.

Step to enable secure connection to JackDB

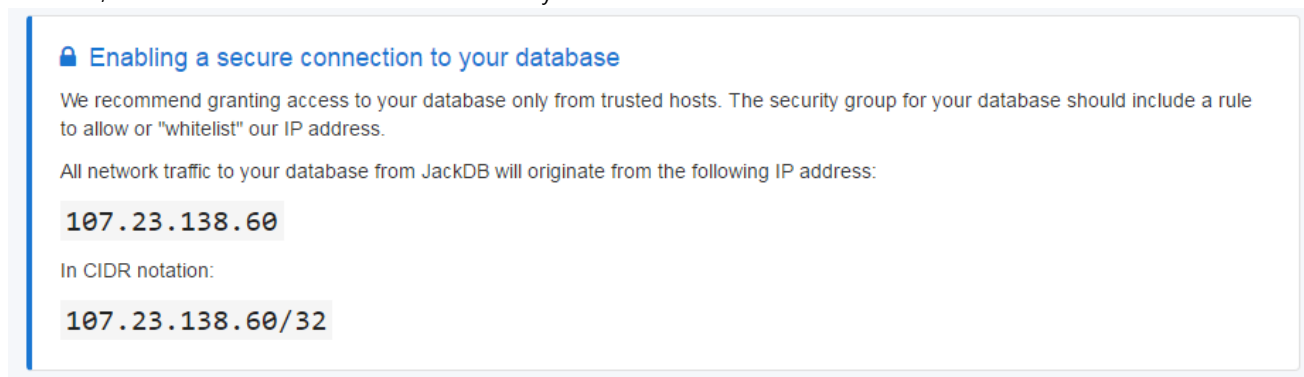
- After signed up in JackDB, I clicked on **Add a data source** button to add a data source.



- 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.

The screenshot shows the AWS Redshift console interface. The main content area displays the configuration for a Redshift cluster named 'cloudlabcluster'. The cluster is in the 'Available' state. The 'Connection Information' modal window is open, showing the following details:

- Endpoint:** cloudlabcluster.cfbrzn3xaqu7.us-west-2.redshift.amazonaws.com:5439 (authorized)
- Publicly Accessible:** Yes
- Master User Name:** masteruser
- Security Group Rules:**

Security Group	Type	Rule
default	Security Group - Inbound	sg-0c7f5674
default	CIDR/IP - Outbound	0.0.0.0/0

A note at the bottom of the modal states: '* Note that only TCP rules applicable to the database port are displayed.'

- Adding Redshift inbound rule in security group

The screenshot shows the 'Edit inbound rules' dialog box. The dialog contains a table with the following rules:

Type	Protocol	Port Range	Source
All traffic	All	0 - 65535	Custom sg-0c7f5674
Redshift	TCP	5439	Custom 107.23.138.60/32
HTTP	TCP	80	Custom 0.0.0.0/0, ::/0
HTTPS	TCP	443	Custom 0.0.0.0/0, ::/0

The 'Redshift' rule is highlighted with a red border. At the bottom of the dialog, there is an 'Add Rule' button and 'Cancel' and 'Save' buttons.

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

Enter the database credentials for this data source.

Name
cloudlabcluster_conn

Type
Amazon Redshift

Host
cloudlabcluster.cfbrzn3xaqu7.us-west-2.redshift.amazonaws.com

Port
5439

Database
dev

Username
masteruser

Password

☒ **SSL is enabled**
This data source uses Secure Sockets Layer (SSL) to encrypt the connection between JackDB and your database.
This data source also validates the server certificate that JackDB connects to.

[Create](#) [Cancel](#)

- Now, data source has been successfully created.

JackDB, Inc. [US] | https://app.jackdb.com/admin/data-sources/D4LDBy4qbzXCshZ

Stevens Institute of Technology

JackDB

PROFILE & SETTINGS

- Profile
- Settings
- Password
- Two-Factor

YOUR ACCOUNT

- Notifications
- Invitations
- Snippets
- Security

TEAM ADMINISTRATION

- Invite People
- Plans

ACCOUNT MANAGEMENT

- Overview
- Users

Data source has been created. [Click here to connect to this data source.](#)

Data Source Detail

[Actions](#)

cloudlabcluster_conn

Type	Amazon Redshift
Host	cloudlabcluster.cfbrzn3xaqu7.us-west-2.redshift.amazonaws.com:5439
Database	dev
Username	masteruser
Created	Apr 25, 2017 at 13:14:31

Users **1** Roles **0** Permissions **0**

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://awssampleduswest2/ticket/allusers_pipe.txt'
2
3 CREDENTIALS 'aws_access_key_id=AKIAJCYDAYVDUGT3KZ5A;aws_secret_access_key=0Y0xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx1KZ'
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	likeclass
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 Sprin...	BC	molestie...	(192) 91...	null	t	t	null	f
9	18	VDP05MXU	Germaine	Valdez	Kokomo	WY	cursor.s...	(998) 87...	null	t	t	null	t
10	19	CKQ97IWP	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	cursor.s@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.

CloudLab-JackDB
Execute
Export
Explain

1 SELECT userid,firstname,lastname,city,likesports FROM users WHERE likesports = 'true' ORDER BY firstname;

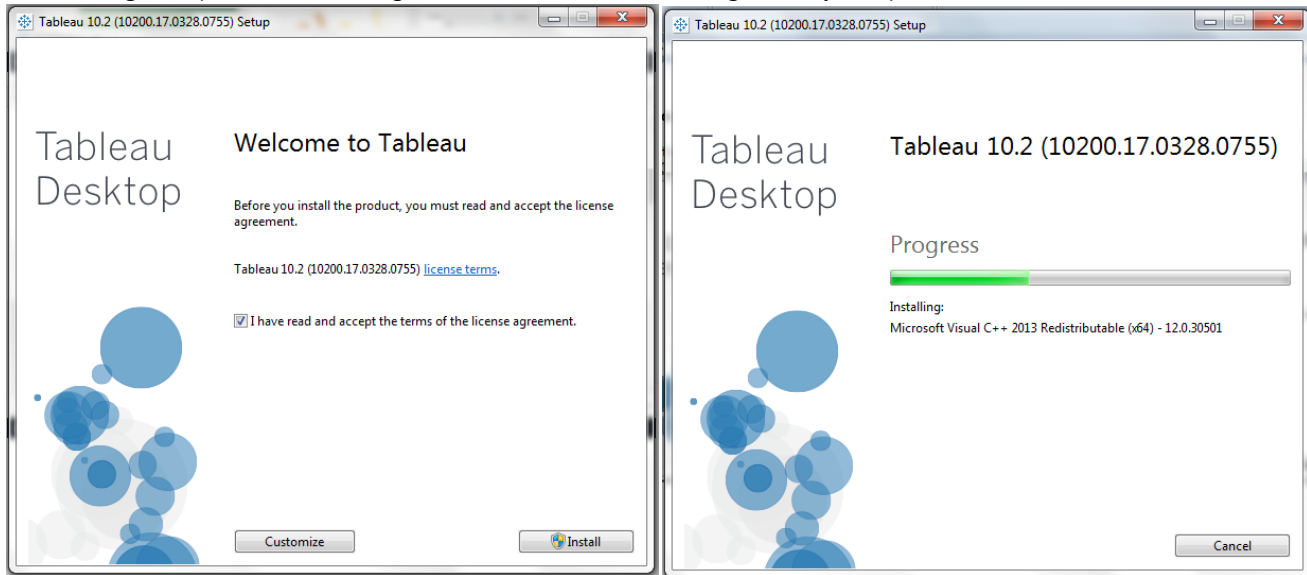
Result
Data Grid
SQL
FIRST 5000 ROWS RETURNED

	userid	firstname	lastname	city	likesports
1	121	Aaron	Smith	Malden	t
2	972	Aaron	Dixon	Bethany	t
3	11,117	Aaron	Strickla...	San Gabr...	t
4	15,722	Aaron	Casey	Pasco	t
5	17,818	Aaron	Warren	Flint	t
6	25,395	Aaron	Mcconnell	Texarkana	t
7	28,788	Aaron	Nicholson	Idaho Sp...	t
8	37,533	Aaron	Washingt...	Mentor	t
9	35,065	Aaron	Downs	Chesapea...	t
10	37,188	Aaron	Small	Manhattan	t
11	2,524	Abbot	Willis	Areceibo	t
12	8,690	Abbot	Sloan	Westfield	t
13	20,257	Abbot	Ortiz	Leominst...	t
14	13,193	Abbot	Mason	DuBois	t
15	33,901	Abbot	Cervantes	Avalon	t
16	16,788	Abbot	Hurst	Utica	t
17	37,508	Abbot	Haley	Gardena	t
18	20,562	Abbot	Mcfarland	Dover	t
19	49,319	Abbot	Lott	Healdsbu...	t
20	31,981	Abbot	Castaneda	Yonkers	t
21	48,274	Abbot	Ramsey	Manassas	t
22	4,349	Abdul	Barnett	Alameda	t

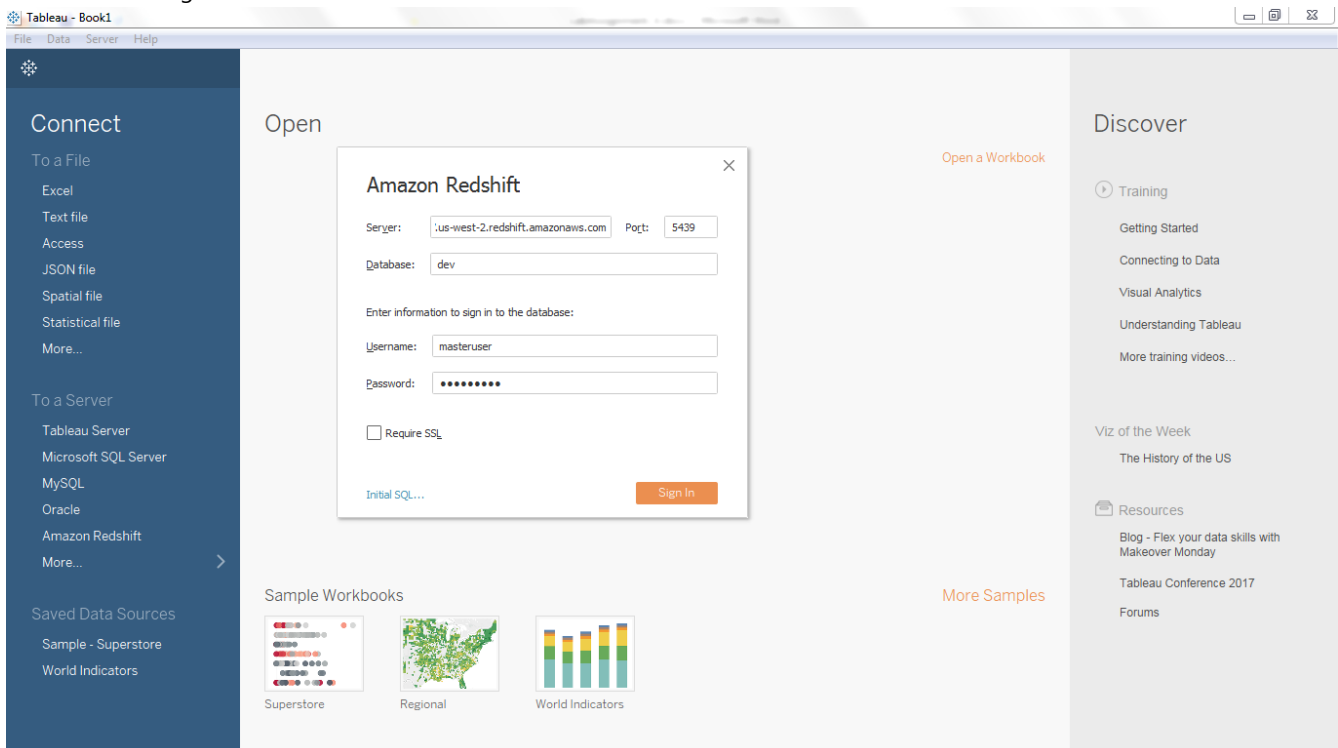
FIRST 5000 ROWS RETURNED IN 7.407 SECONDS • EXECUTED IN A FEW SECONDS

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 debug the error I have added my computer's public IP to my cluster security group. Then it connected.

Type	Protocol	Port Range	Source
Redshift	TCP	5439	Custom 107.23.138.60/32
Redshift	TCP	5439	Custom 72.90.254.70/32
HTTP	TCP	80	Custom 0.0.0.0/0
HTTP	TCP	80	Custom :::/0
All traffic	All	0 - 65535	Custom sg-0c7f5674
HTTPS	TCP	443	Custom 0.0.0.0/0
HTTPS	TCP	443	Custom :::/0

Buttons: Add Rule, Cancel, Save

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

Tableau - Book1 - Tableau license expires in 14 days

File Data Server Window Help

public

Connections: cloudlabclus...amazonaws.com Amazon Redshift

Database: dev

Schema: public

Table: users (highlighted)

New Custom SQL, New Union

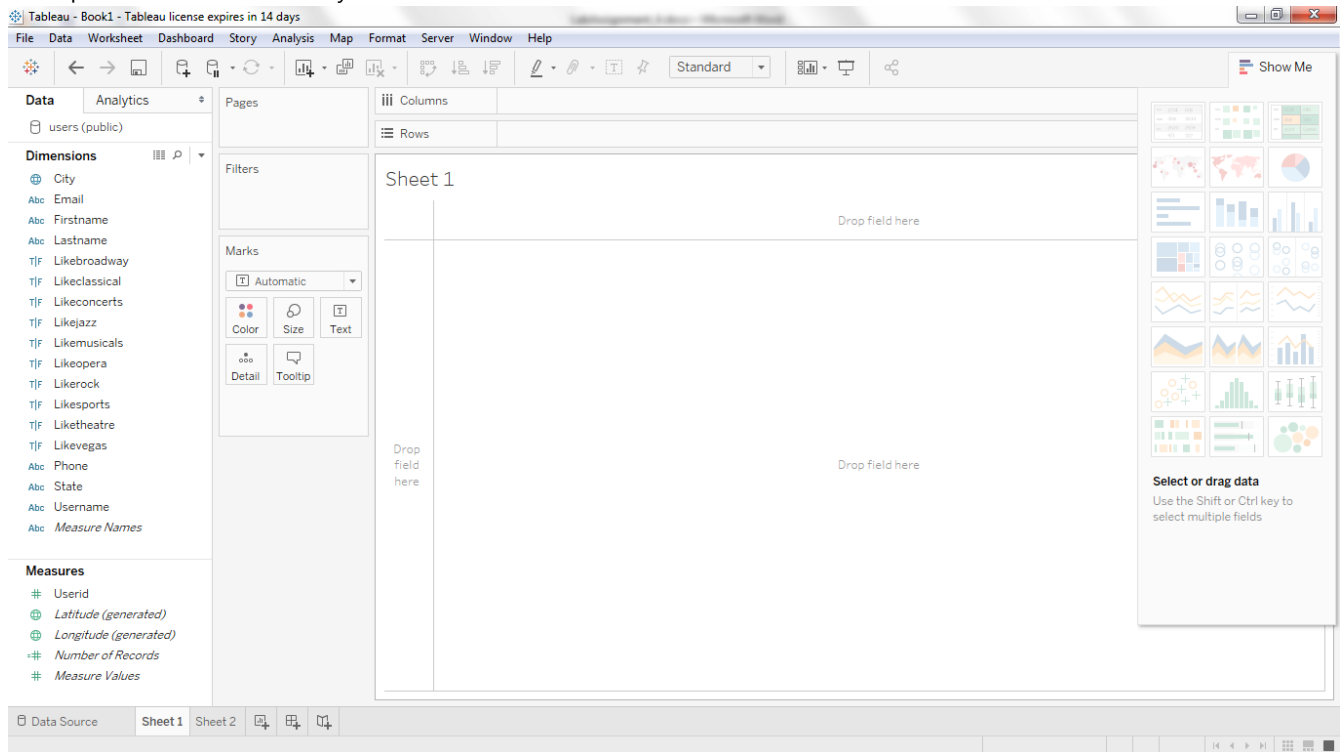
Sort fields: Data source order

Show aliases, Show hidden fields, rows

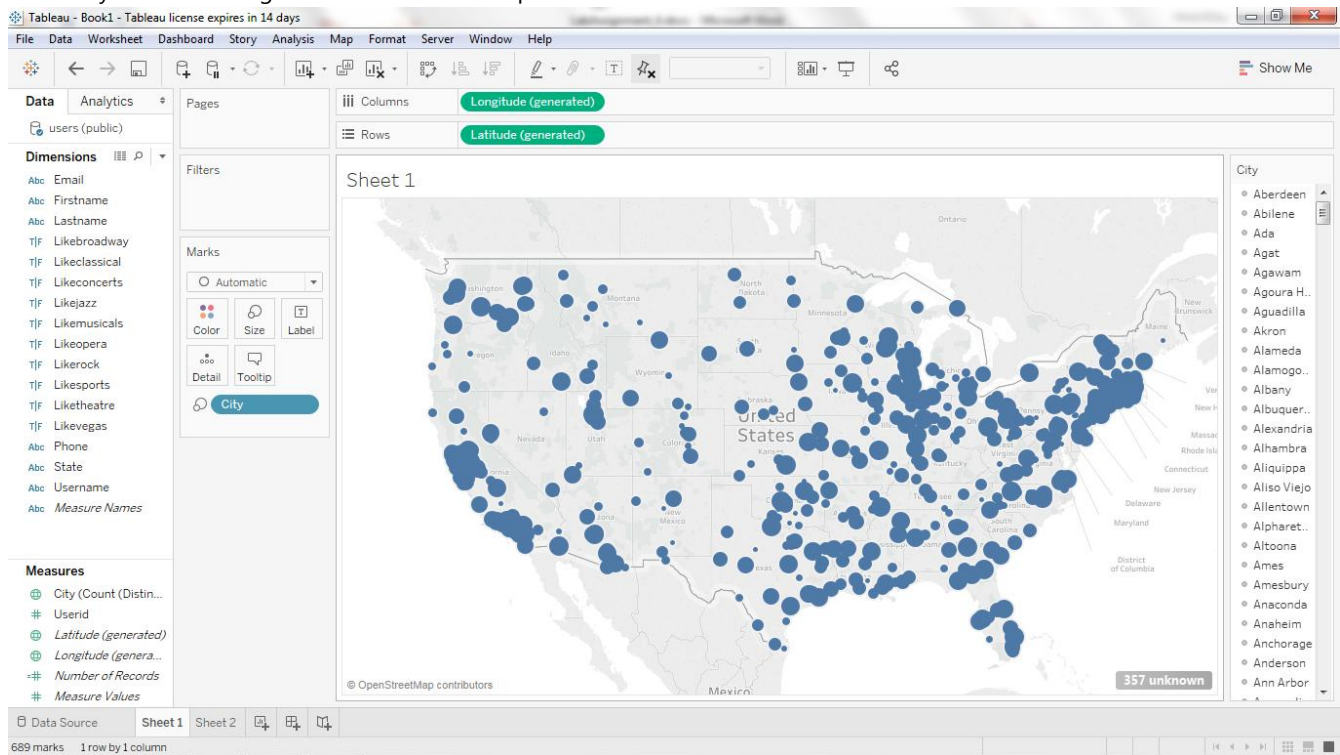
Data Source, Sheet 1

Step to analyze data and create charts using Tableau

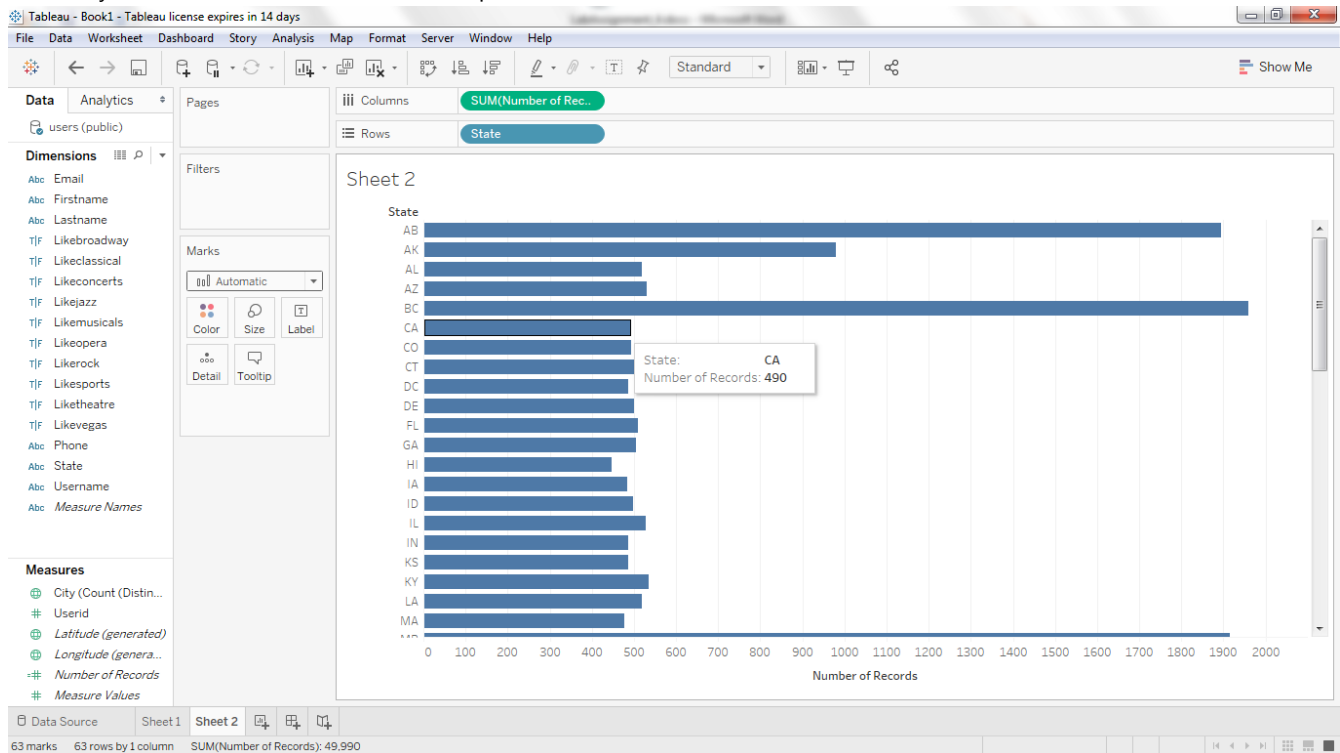
- Opened new sheet for analysis and chart creation



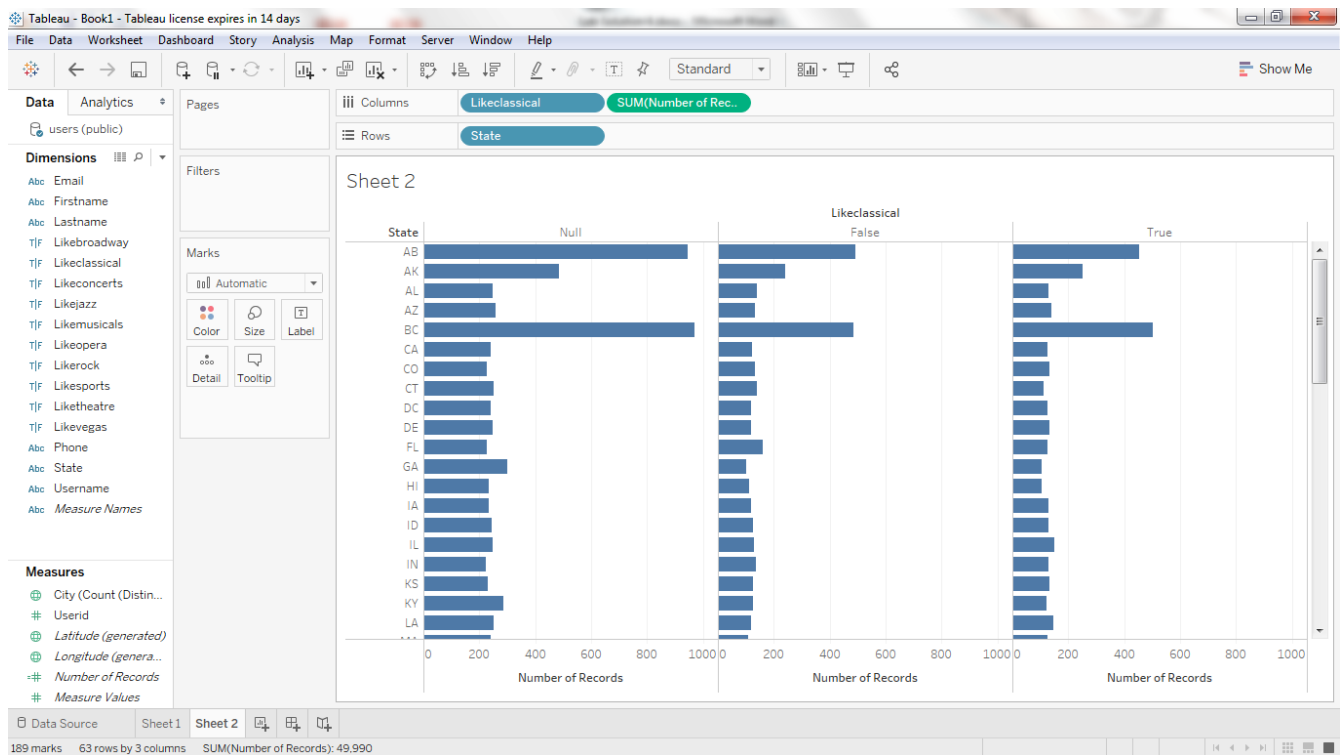
- Analysis 1 – Showing all the cities those are present in our users table



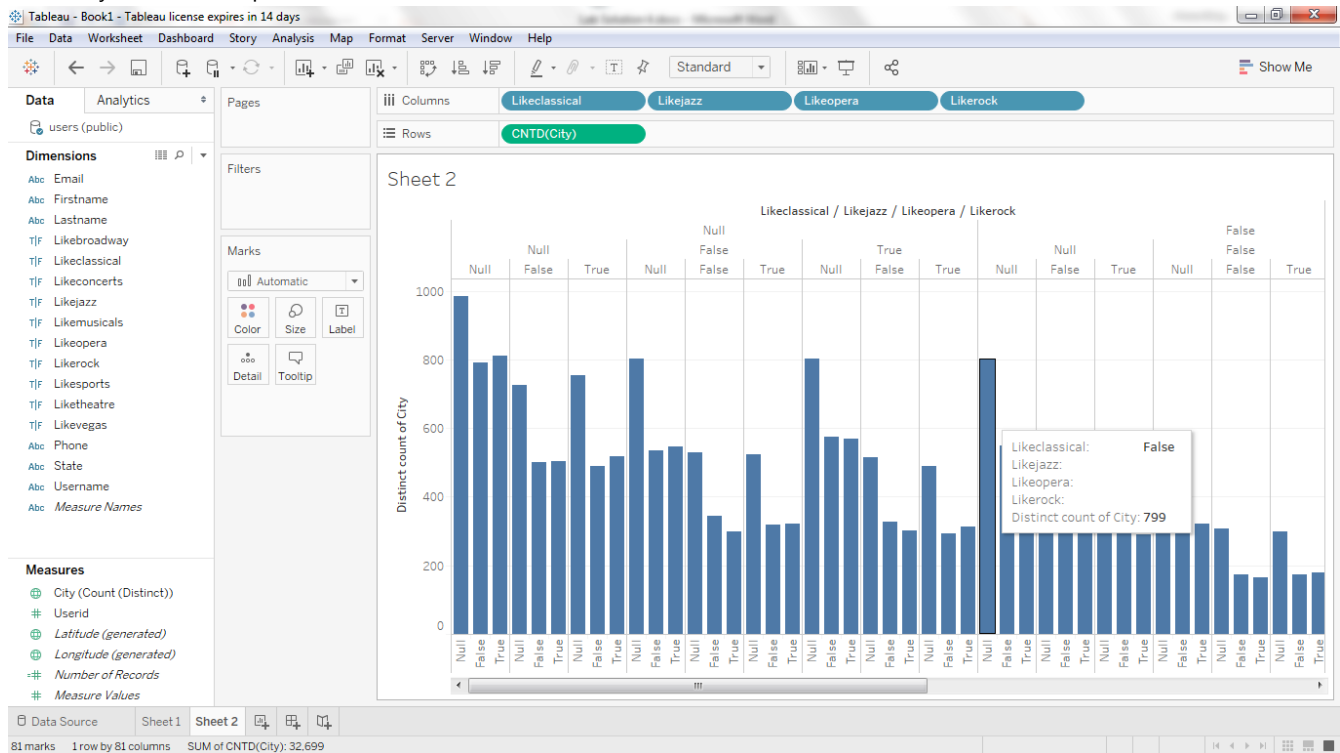
- Analysis 2 – It shows the list of state are present in number of records in users table



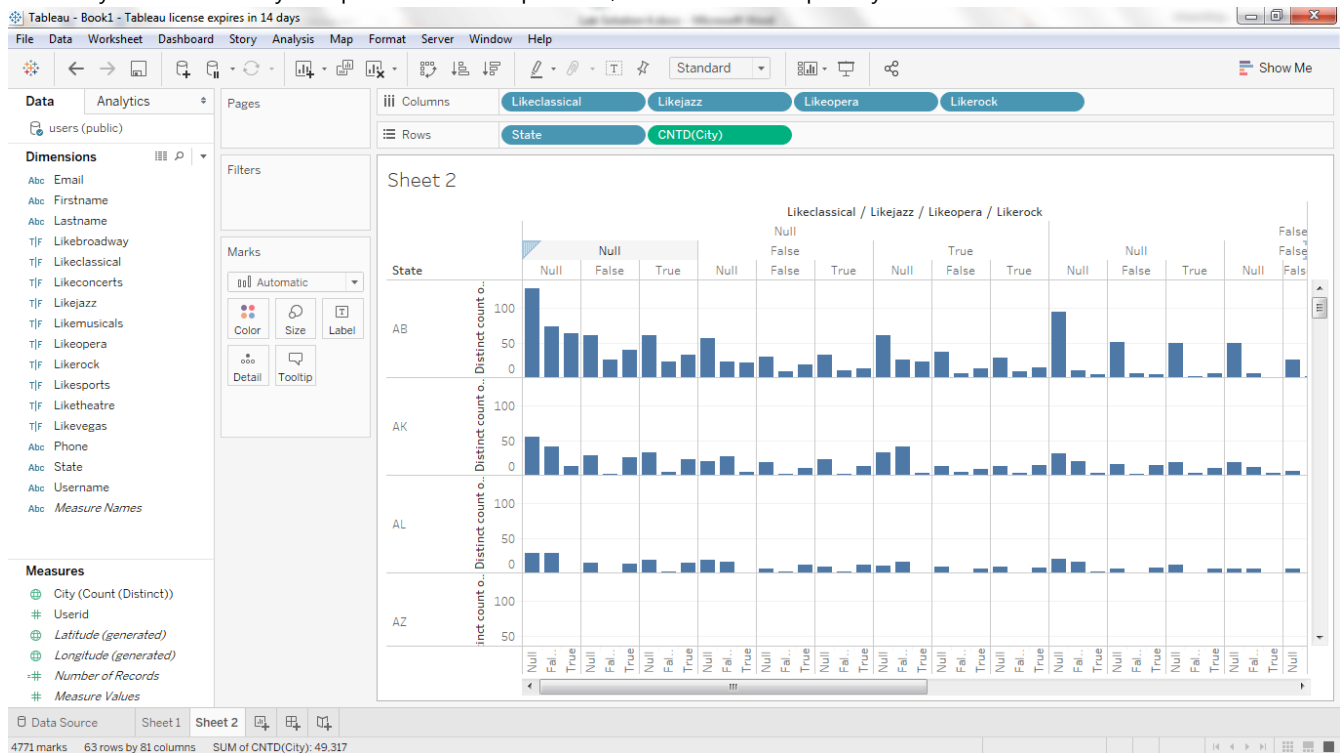
- Analysis 3 – In this report, I have combined more dimensions, that shows the total records have likeclassical as respect to each state.



- 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.