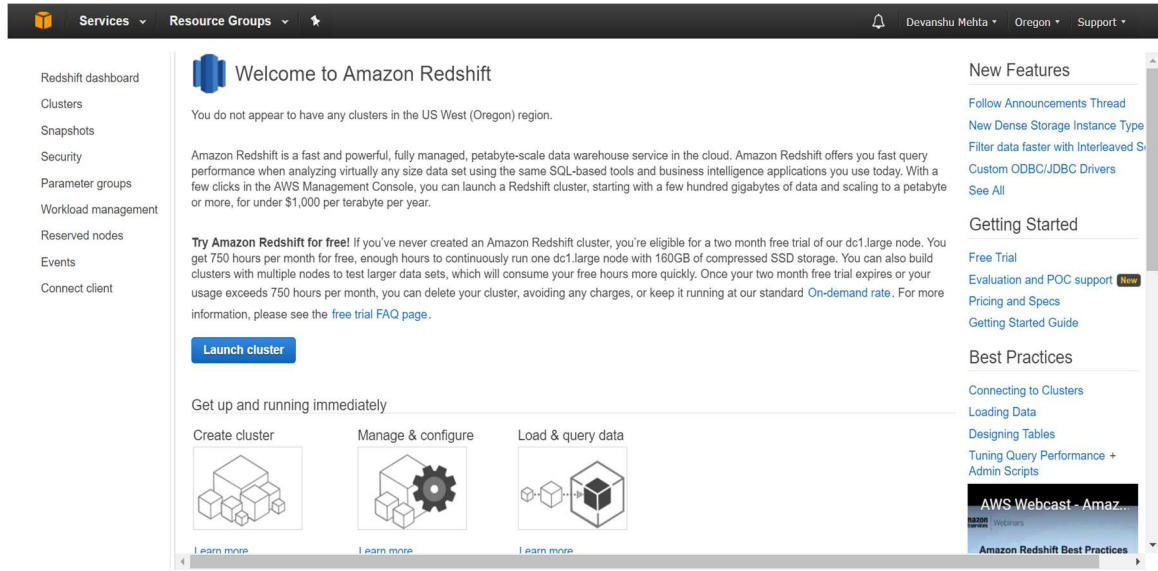


Lab Assignment 4

Solution

Steps to follow:

1. First, sign in to your console. And under services search for redshift and click on it.



2. After that click on launch a cluster. When you do that. You need to provide basic information about your cluster. This is for configuring your cluster. Make sure you remember this inputs because they will be used in the following steps.

The screenshot shows the 'Cluster Details' configuration page in the Amazon Redshift console. The page has a breadcrumb trail: 'CLUSTER DETAILS' > 'NODE CONFIGURATION' > 'ADDITIONAL CONFIGURATION' > 'REVIEW'. The 'CLUSTER DETAILS' tab is active. A message at the top says 'Provide the details of your cluster. Fields marked with * are required.' The form contains several input fields with corresponding descriptions: 'Cluster identifier*' (required, unique key for the cluster), 'Database name' (optional, default is 'dev'), 'Database port*' (required, default is 5439), 'Master user name*' (required, master user for the cluster), 'Master user password*' (required, must contain 8-64 printable ASCII characters), and 'Confirm password*' (required, must match the master user password). At the bottom, there are 'Cancel' and 'Continue' buttons.

Services Resource Groups

Devanshu Mehta Oregon Support

Redshift dashboard Clusters Snapshots Security Parameter groups Workload management Reserved nodes Events Connect client

CLUSTER DETAILS NODE CONFIGURATION ADDITIONAL CONFIGURATION REVIEW

Choose a number of nodes and node type below. Number of Compute Nodes is required for multi-node clusters.

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.](#)

Node type: **dc1.large** Specifies the compute, memory, storage, and I/O capacity of the cluster's nodes.

CPU: 7 EC2 Compute Units (2 virtual cores) per node

Memory: 15 GiB per node

Storage: 160GB SSD storage per node

I/O performance: Moderate

Cluster type: **Single Node** Single Node clusters consist of a single node which performs both leader and compute functions.

Number of compute nodes*: **1**

Maximum: 1 Minimum: 1

Cancel Previous Continue

- After configuring your cluster, it will take time to get created, so first the status of cluster will show “created” and after sometime it will get changed to “available”.

Services Resource Groups

Devanshu Mehta Oregon Support

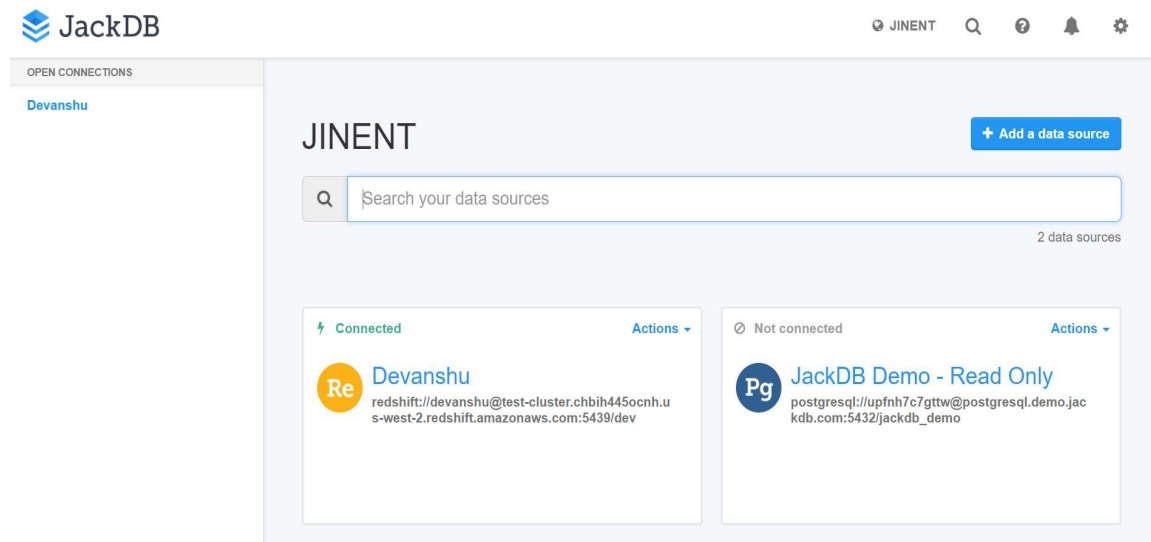
Redshift dashboard Clusters Snapshots Security Parameter groups Workload management Reserved nodes Events Connect client

Clusters

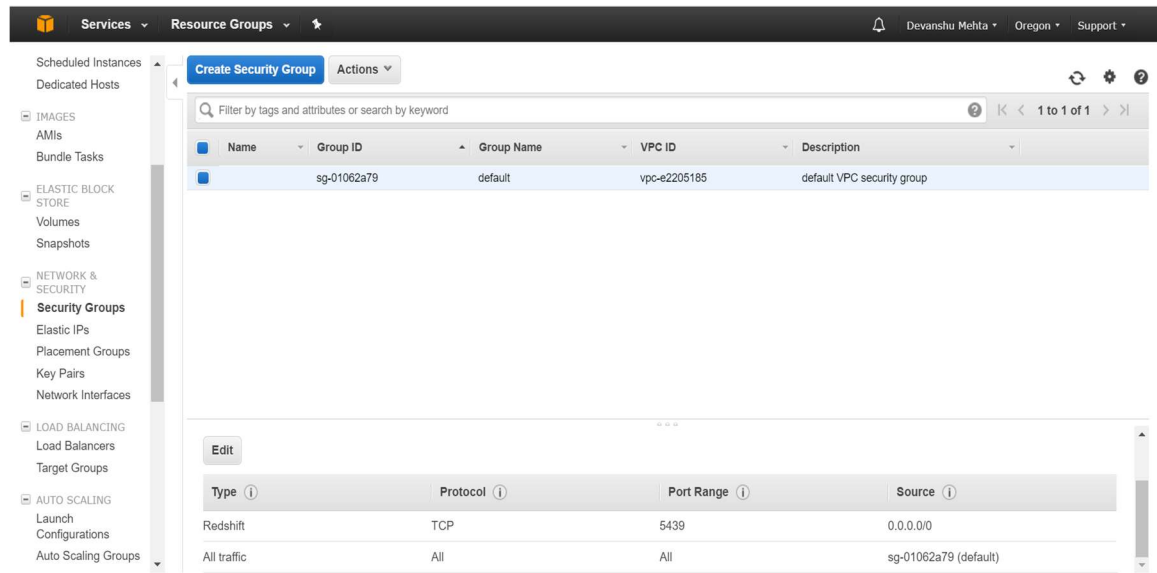
Launch Cluster Manage Tags Manage IAM roles

	Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
<input type="checkbox"/>	test-cluster	available	healthy	no	3

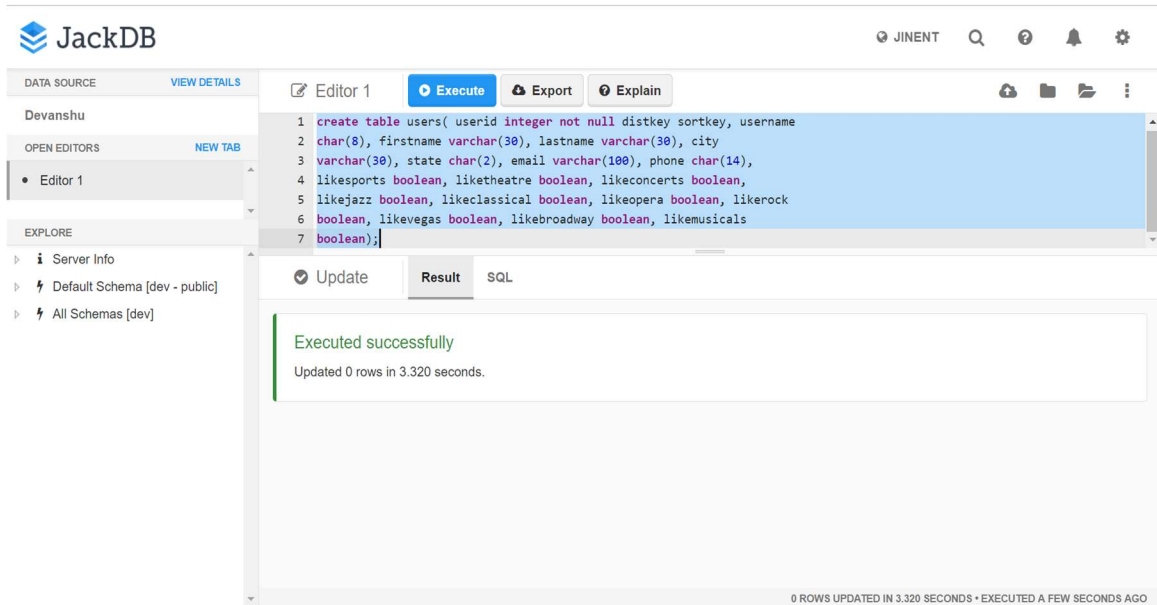
- Once it gets available click on that and you will get an ENDPOINT url which will be used as host name while configuring your JackDb database.
- Now go to www.jackdb.com and start your free trial. Now add Amazon Redshift as your data source. In that while configuring use your credentials you used while configuring your cluster, and provide the ENDPOINT url here as the host name. The below image shows the output once you do above steps.



- After that in your cluster dashboard go to security>security groups>click on amazon EC2 security groups>click on inbound>add rule>Amazon Redshift.



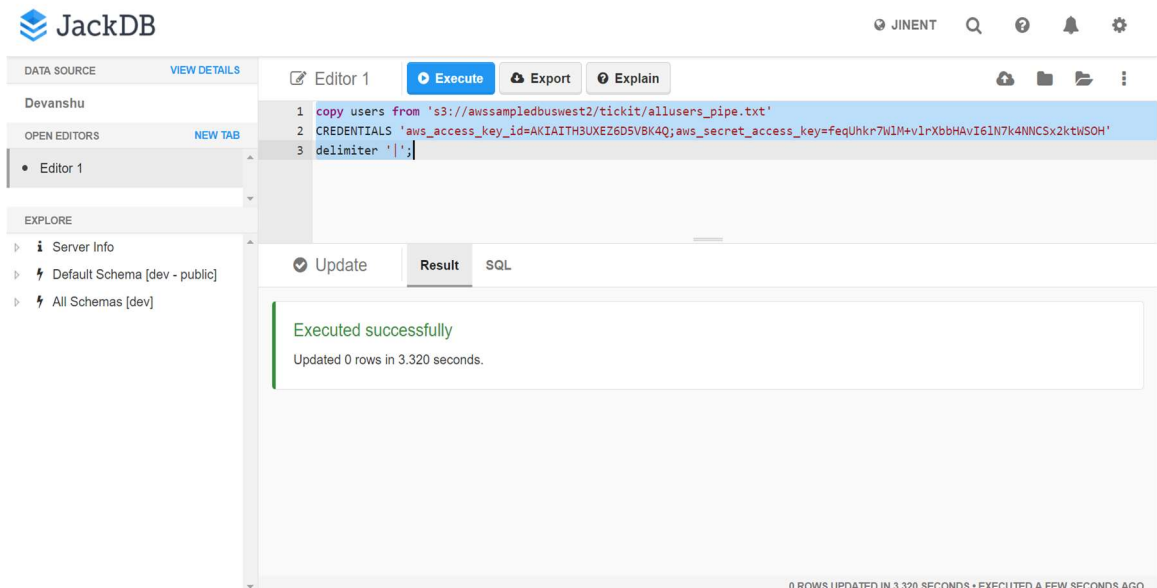
- Now in that provide source IP as "0.0.0.0/0", you can look it at the bottom right corner of above image. Now in JackDb click on create database.
- Now execute the commands given in assignment specifications one by one.
- The first command is for creating a table named "users".



The screenshot shows the JackDB web interface. On the left, the 'DATA SOURCE' section shows 'Devanshu' as the selected source. Below it, 'OPEN EDITORS' shows 'Editor 1'. The 'EXPLORE' section on the left lists 'Server Info', 'Default Schema [dev - public]', and 'All Schemas [dev]'. The main editor area, 'Editor 1', contains a SQL command to create a table named 'users' with various columns and data types. The command is highlighted in blue. Below the editor, the 'Update' tab is selected, showing a green message: 'Executed successfully' and 'Updated 0 rows in 3.320 seconds.' The 'Result' and 'SQL' tabs are also visible. At the bottom right, a status bar indicates '0 ROWS UPDATED IN 3.320 SECONDS • EXECUTED A FEW SECONDS AGO'.

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

10. The second command is to provide access ID key and secure access ID key to which will allow you to load data from S3.



The screenshot shows the JackDB web interface with the same layout as the previous one. The main editor area, 'Editor 1', contains a SQL command to copy data from an S3 bucket into the 'users' table. The command is highlighted in blue. Below the editor, the 'Update' tab is selected, showing a green message: 'Executed successfully' and 'Updated 0 rows in 3.320 seconds.' The 'Result' and 'SQL' tabs are also visible. At the bottom right, a status bar indicates '0 ROWS UPDATED IN 3.320 SECONDS • EXECUTED A FEW SECONDS AGO'.

```
1 copy users from 's3://awssampleduswest2/ticket/allusers_pipe.txt'
2 CREDENTIALS 'aws_access_key_id=AKIAITH3UXE26D5VBK4Q;aws_secret_access_key=fegUhr7WlM+v1rXbbHAvI61N7k4NNCSx2ktwSOH'
3 delimiter '|';
```

11. The third command is to display sample data inserted into the table from S3.

The screenshot shows the JackDB web interface. On the left, there's a sidebar with 'DATA SOURCE' (Devanshu), 'OPEN EDITORS' (Editor 1), and 'EXPLORE' (Server Info, Default Schema [dev - public], All Schemas [dev]). The main area has a 'Editor 1' tab with a SQL query:

```
1 SELECT * from users;
2 SELECT userid,firstname,lastname,city,likesports from users where
3 likesports = 'true' order by firstname;
4
```

Below the editor, there's a 'Result' tab showing a table with 14 rows. The columns are 'userid', 'firstname', 'lastname', 'city', and 'likesports'. The first 13 rows are visible, showing users with 'likesports' set to 't'.

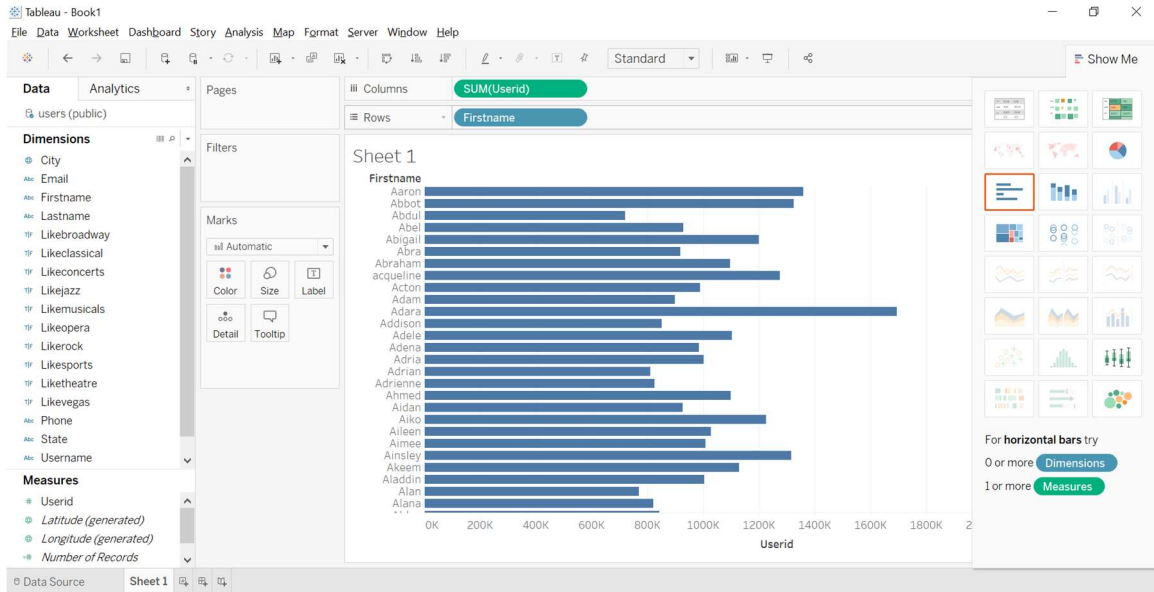
	userid	firstname	lastname	city	likesports
1	121	Aaron	Smith	Malden	t
2	972	Aaron	Dixon	Bethany	t
3	11,117	Aaron	Strickla...	San Gabr...	t
4	15,722	Aaron	Casey	Pasco	t
5	17,818	Aaron	Warren	Flint	t
6	25,395	Aaron	Mcconnell	Texarkana	t
7	28,788	Aaron	Nicholson	Idaho Sp...	t
8	37,533	Aaron	Washingt...	Mentor	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

At the bottom right, it says 'FIRST 5000 ROWS RETURNED IN 9.652 SECONDS • EXECUTED AN HOUR AGO'.

12. Once you get the data, install “Tableau” from www.tableau.com .
13. Now in tableau create your account to start a free trial. After creating the account click on Amazon Redshift to connect to your cluster. After clicking you will be asked for database name in your cluster, provide that and open it in the sheets of Tableau.

The screenshot shows the Tableau Desktop interface. On the left, there's a 'Connect' sidebar with options: 'To a File' (Excel, Text file, Access, JSON file, Spatial file, Statistical file, More...), 'To a Server' (Tableau Server, Microsoft SQL Server, MySQL, **Create**, Amazon Redshift, More...), and 'Saved Data Sources' (Sample - Superstore, World Indicators). The 'Create' option is circled in blue. The main area shows 'Open' (Book1), 'Discover' (Training, Viz of the Week, Resources), and 'Sample Workbooks'.

14. Now insert charts selecting certain dimensions and measures that are available from your table on the left side. The more fields you select more and more chart options will be available.



15. After creating this chart, I deleted my cluster and below is the image of delete event.

Services Resource Groups

Redshift dashboard

Clusters

Snapshots

Security

Parameter groups

Workload management

Reserved nodes

Events

Connect client

Events Subscriptions

Filter: All source types Search...

Viewing 5 of 5 events

Time	Event	Source ID	Source type	Event categories
Apr 25 12:39 PM	Amazon Redshift cluster 'test-cluster' has been deleted at 2017-04-25 16:39 UTC. A final snapshot was not saved.	test-cluster	cluster	management
Apr 25 10:32 AM	Amazon Redshift cluster 'test-cluster' has been created at 2017-04-25 14:32 UTC and is ready for use.	test-cluster	cluster	management
Apr 25 10:31 AM	Cluster restart is complete.	test-cluster	cluster	monitoring
Apr 25 10:31 AM	Cluster is being restarted.	test-cluster	cluster	monitoring
Apr 25 10:28 AM	Cluster parameter group 'default.redshift-1.0' was created at 2017-04-25 14:28 UTC.	default.redshift-1.0	cluster-parameter-group	configuration