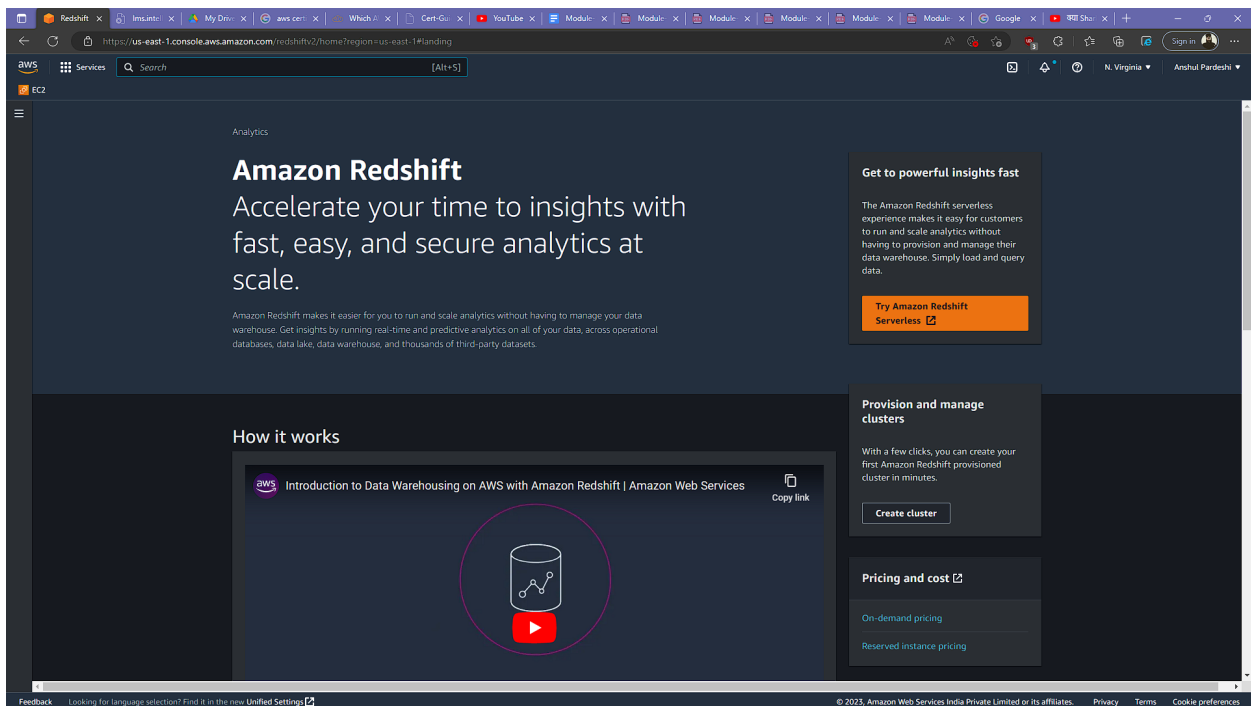


Module-6: Database Services Assignment - 4

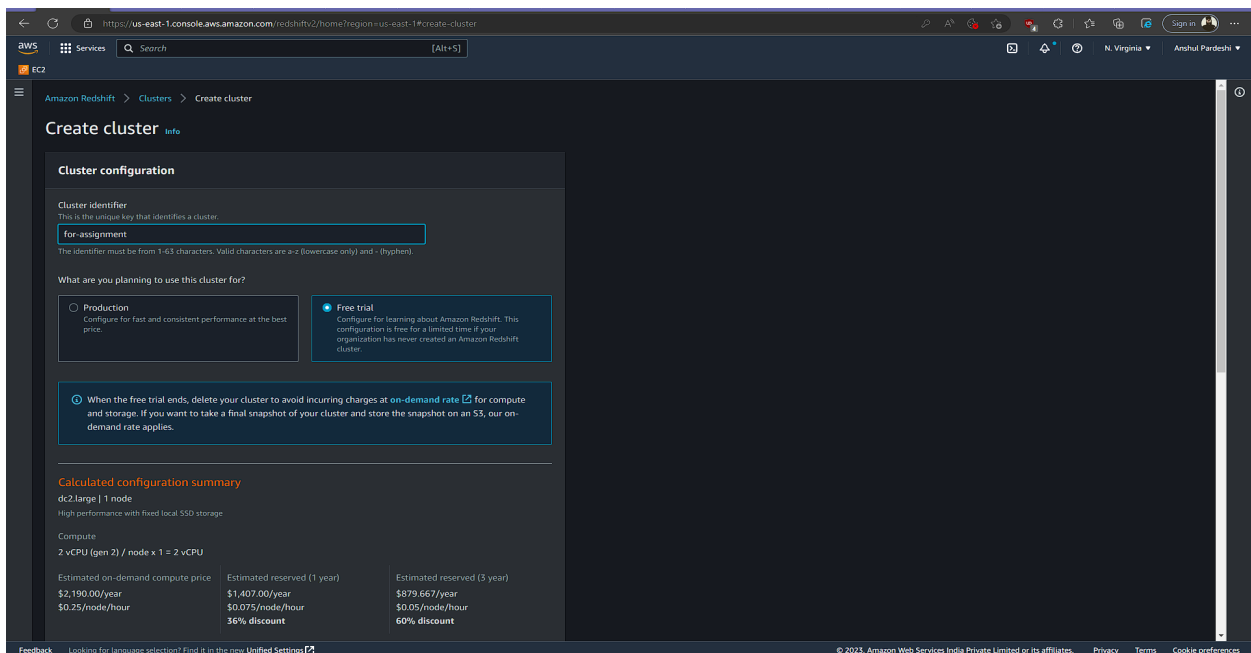
You have been asked to:

1. Create a Redshift Data Warehouse
2. Using the Query editor
 - a. Load some data
 - b. Query the data

Create a cluster by going to the Amazon Redshift dashboard.



Give it a name and let us choose a free tier for now.



Select create cluster.

The screenshot shows the 'Create cluster' page in the Amazon Redshift console. The page is divided into several sections:

- EC2**: A table showing pricing for different instance types. The selected instance type is `r5.xlarge` with a price of `$0.075/node/hour` and a `36% discount`. The storage capacity is `160 GB x 1 nodes = 160 GB`.
- Sample data**: A section with a toggle for 'Sample data is loaded with your Redshift cluster.' and a description of the 'Tickit' (28 MB) sample data set.
- Database configurations**: A section for configuring the database. It includes fields for 'Admin user name' (set to 'admin') and 'Admin user password' (set to 'drink7up'). There are checkboxes for 'Auto generate password' and 'Show password'.

At the bottom of the form, there are 'Cancel' and 'Create cluster' buttons.

Click on the cluster that you have created.

The screenshot shows the 'Provisioned clusters dashboard' in the Amazon Redshift console. The dashboard provides an overview of the clusters and their status.

- Resources overview**: A section showing the total number of nodes, on-demand nodes, reserved nodes, reserved nodes available, automated snapshots, and manual snapshots. The total number of nodes is 1, and the on-demand nodes are 1.
- Cluster overview (1)**: A table showing the status of the clusters. The cluster 'for-assignment' is shown with a status of 'Available'.
- Cluster metrics**: A section showing the number of queries, database connections, disk space used, and CPU utilization. The number of queries is 0.
- Datashares**: A section showing the number of datashares. There are 0 datashares.
- Alarms (0)**: A section showing the number of alarms. There are 0 alarms.
- Events (4)**: A section showing the number of events. There are 4 events.

At the bottom of the dashboard, there are buttons for 'Try Amazon Redshift Serverless', 'Purchase reserved nodes', and 'Create cluster'.

Click on Query in the query editor.

The screenshot shows the Amazon Redshift console interface. At the top, there's a navigation bar with 'AWS' and 'EC2' logos. Below it, a search bar and a 'Services' dropdown are visible. The main content area displays the details for a cluster named 'for-assignment'. The 'General information' tab is active, showing fields like Cluster identifier, Status (Available), Node type (dc2.large), and Endpoint. A 'Query data' button is highlighted in orange in the top right corner of the cluster details section. Below the general information, there are tabs for 'Cluster performance', 'Query monitoring', 'Schedules', 'Maintenance', and 'Properties'. The bottom of the console shows a footer with 'Feedback' and 'Looking for language selection? Find it in the new Unified Settings' link.

We can load data from a JSON format file or CSV. We created a JSONformate file named category and uploaded it in dev/public.

The screenshot shows the Amazon Redshift query editor v2 interface. The top bar includes a 'Database' dropdown and a 'Load data' button. The main area is a query editor with a 'Run' button and a 'Limit 100' toggle. The left sidebar contains a 'Queries' list with 'for-assignment' selected. The bottom of the editor shows a footer with 'Feedback' and 'Looking for language selection? Find it in the new Unified Settings' link.

so to query that data, we use command <select * from “dev”. “public”. “category”;>

The screenshot shows the AWS Redshift query editor v2 interface. The SQL query being executed is: `SELECT * FROM "dev"."public"."category";`. The interface includes a sidebar with navigation options like Database, Queries, Notebooks, Charts, and History. The main area displays the query and its execution status. The bottom status bar indicates the execution time and total rows.

Field	Type	NL
catid	smallint	NN
catgroup	character varying(10)	NULL
catname	character varying(10)	NULL
catdesc	character varying(50)	NULL

Data in tabular form has been displayed.

The screenshot shows the AWS Redshift query editor v2 interface with the query results displayed in a tabular format. The results are shown in a table with columns: catid, catgroup, catname, and catdesc. The table contains 11 rows of data. The bottom status bar indicates the execution time and total rows.

catid	catgroup	catname	catdesc
1	Sports	MLB	Major League Baseball
3	Sports	NFL	National Football League
6	Shows	Musicals	Musical theatre
8	Shows	Opera	All opera and light opera
11	Concerts	Classical	All symphony, concerto, a
2	Sports	NHL	National Hockey League
4	Sports	NBA	National Basketball Assoc...
5	Sports	MLS	Major League Soccer
7	Shows	Plays	All non-musical theatre
9	Concerts	Pop	All rock and pop music co...
10	Concerts	Jazz	All jazz singers and bands

