

Capstone-4

Bibliotheca E-Book Subscription Recommendation

Batch A | Group 4

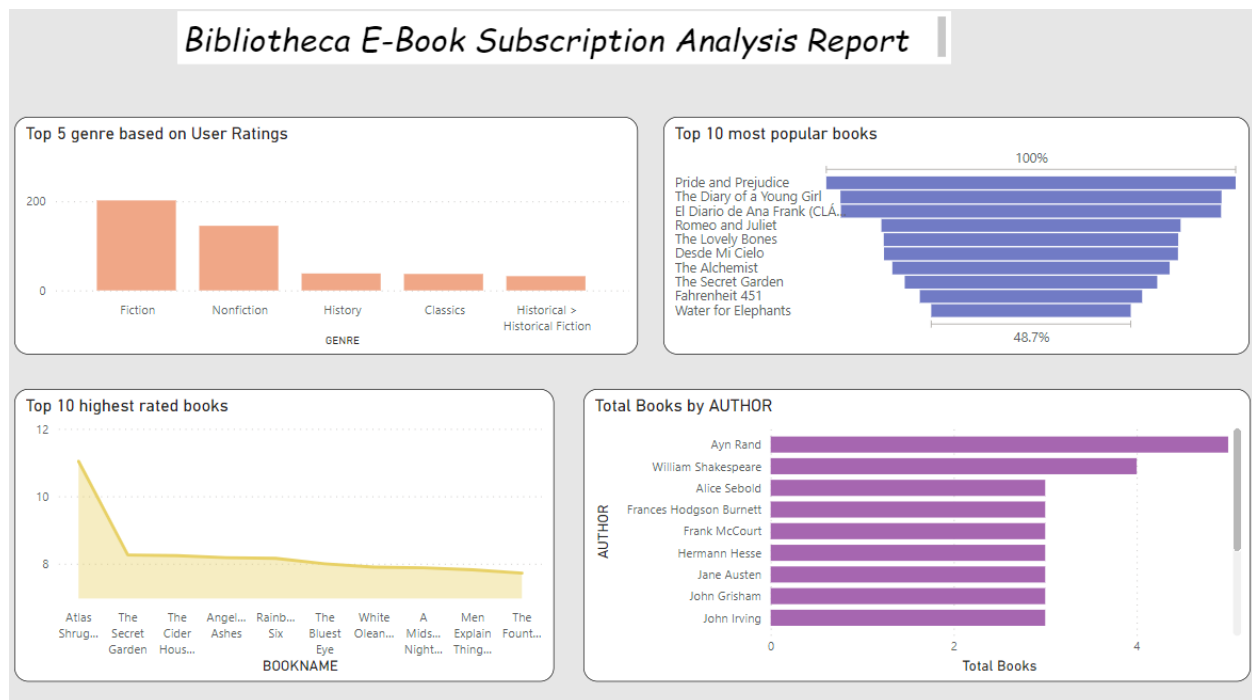
Group Members

- Gokul Prakash Babu
- Gomathi Sorna Ganesh
- Gourav Kumar
- Grace Kollabathula
- Gudla Sabarish

CHECKPOINT 3 | TASK 3.2 | DATA ANALYSIS ON CLOUD

Snapshot 1: Analysis of Data using Power BI through AWS

Based on the data provided the following report was prepared.



The following steps are followed to connect Amazon RedShift to Power BI

Step 1: Move the Datasets to AWS s3

The screenshot shows the Amazon Redshift console interface. The breadcrumb navigation is "Amazon Redshift > Clusters > grp4clusters". The cluster name "grp4clusters" is displayed at the top. Below it, there are buttons for "Actions", "Edit", "Add partner integration", and "Query data". The "General information" tab is selected, showing a table with the following data:

Cluster identifier	Status	Node type
grp4clusters	Available	dc2.large

Additional details shown include:

- Cluster namespace: c2cd10ba-548d-4b18-803b-a9a6ef160505
- Date created: March 22, 2022, 18:58 (UTC+05:30)
- Storage used: -
- Number of nodes: 1
- AQUA: Not available

On the right side, there is a "Endpoint copied" notification and a list of URLs:

- Endpoint: grp4clusters.creiu9scb0qp.us-east-1.redshift.a...
- JDBC URL: jdbc:redshift://grp4clusters.creiu9scb0qp.us-e...
- ODBC URL: Driver=(Amazon Redshift (x64)); Server=grp4c...

Below the general information, there are tabs for "Cluster performance", "Query monitoring", "Schedules", "Maintenance", and "Properties". The "Database configurations" section is visible at the bottom, with buttons for "Change admin user password", "Rotate encryption keys", and "Edit".

Step 2 : Create Redshift Instance

This screenshot is identical to the one above, showing the Amazon Redshift console interface for the cluster "grp4clusters". It displays the "General information" tab with details such as cluster identifier, status (Available), node type (dc2.large), cluster namespace, date created, storage used, number of nodes, and AQUA status. It also shows the endpoint, JDBC URL, and ODBC URL on the right side.

Step 3: Creating tables required using RedShift query editor and using copy command to move the data from storage to data warehouse(Redshift).

The screenshot shows the AWS Redshift Query Editor v2 interface. The left sidebar displays the database structure with a tree view showing 'grp4clusters' and 'dev' database. The main editor area shows a SQL query with the following code:

```
1 create table BookCatalogue(  
2 BookID nvarchar(10),  
3 State nvarchar(50),  
4 SubState nvarchar(50)  
5 );  
6  
7 copy BookCatalogue from 's3://grp4bucket/BOOKSCATALOGUE.csv'  
8 credentials 'aws_iam_role=arn:aws:iam::937145466380:role/RedshiftS3Redadwrite'  
9 delimiter ',' region 'us-east-1';  
10  
11 select * from BookCatalogue;  
12
```

The query is executed, and the results are displayed in a table with 100 rows. The table has three columns: bookid, state, and substate. The data is as follows:

bookid	state	substate
BookID	STATE	SUBSTATE
174285	New Mexico	New Mexico
176660	Oregon	Oregon
140353	Arizona	Arizona
178378	New York	New York
128711	California	Orange
125646	Oregon	Oregon

The status bar at the bottom indicates 'Elapsed time: 101 ms' and 'Total rows: 100'.

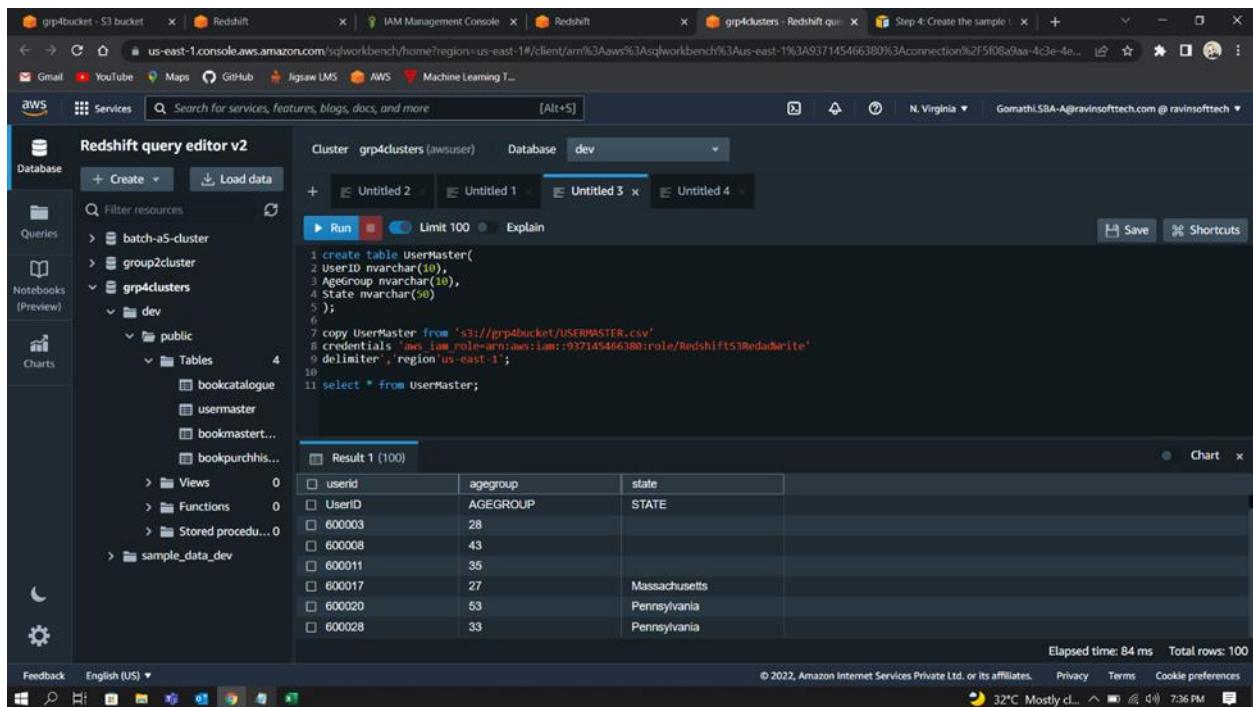
The screenshot shows the AWS Redshift Query Editor v2 interface. The left sidebar displays the database structure with a tree view showing 'grp4clusters' and 'dev' database. The main editor area shows a SQL query with the following code:

```
1 create table BookPurchaseHistory(  
2 BookID nvarchar(10),  
3 UserID nvarchar(10),  
4 SubState nvarchar(50)  
5 );  
6  
7 copy BookPurchaseHistory from 's3://grp4bucket/BOOKSPURCHASEHISTORY.csv'  
8 credentials 'aws_iam_role=arn:aws:iam::937145466380:role/RedshiftS3Redadwrite'  
9 delimiter ',' region 'us-east-1';  
10  
11 select * from BookPurchaseHistory;
```

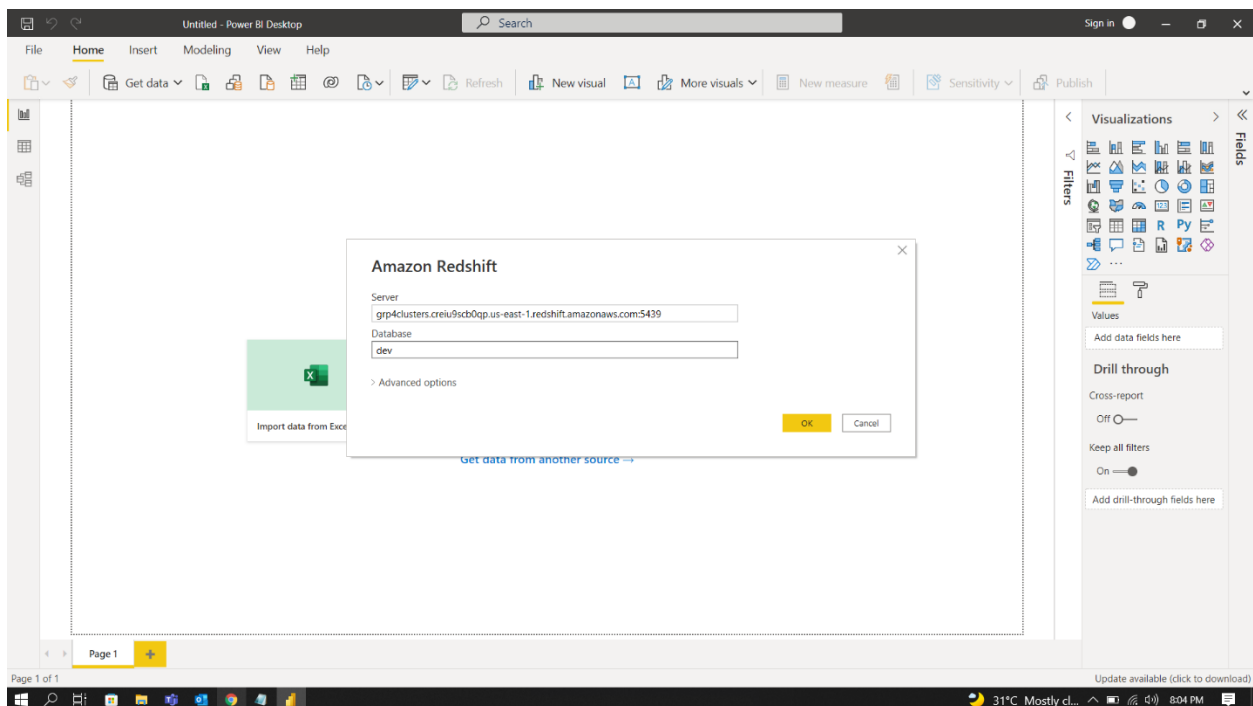
The query is executed, and the results are displayed in a table with 100 rows. The table has three columns: bookid, userid, and substate. The data is as follows:

bookid	userid	substate
BookID	UserID	SUBSTATE
140361	608502	Santa Clara
118603	673204	San Diego
170523	641650	Oklahoma
111924	619531	Orlando
178056	628661	Orange
112488	622287	Hawaii

The status bar at the bottom indicates 'Elapsed time: 80 ms' and 'Total rows: 100'.



Step 4 : Connecting the Redshift data to Power BI using the End point ID of Redshift Cluster , database name and user credentials and performing the tasks mentioned.



FileHomeInsertModelingViewHelp

Get data

Search

Sign in

Visualizations

Fields

Filters

Values

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Navigator

Display Options

Amazon Redshift [2]

catalog_history

public [5]

bookcatalogue

bookmastertest

bookpurchhistory

bookvisithistory

usermaster

Select Related Tables

LoadTransform DataCancel

bookid	userid	substate
BookID	UserID	SUBSTATE
140361	608502	Santa Clara
118603	673204	San Diego
170523	641650	Oklahoma
111924	619531	Orlando
178056	628661	Orange
112488	622287	Hawaii
102767	616482	Houston
106620	653094	Georgia
129631	646225	Los Angeles
131877	620864	Orlando
122740	608550	Mississippi
178261	630815	San Diego
118564	609462	Riverside
169501	623120	Georgia
147734	672182	Houston
146503	630676	Connecticut
120402	644478	Houston
162671	652052	Houston
135629	605753	San Diego
139213	616228	Orange
143688	645021	San Diego
158772	606663	Santa Clara
117918	646377	San Antonio

Page 1 of 1

Update available (click to download)

31°C Mostly cl... 8:05 PM