Amazon Redshift

Cloud Data Warehousing Service

Case Study



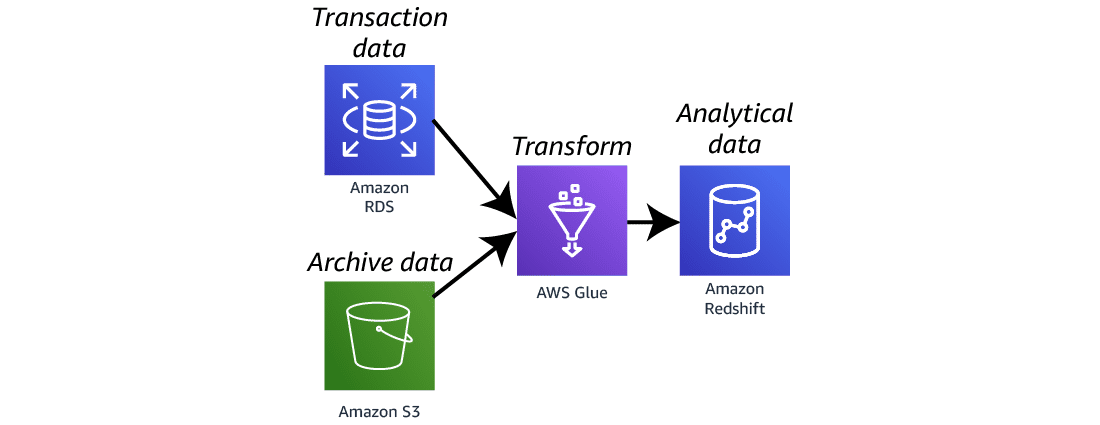
Introduction

Amazon Redshift is an enterprise-level, petabyte scale, fully managed data warehousing service. Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse service that makes it simple and cost-effective to efficiently analyse all your data using your existing business intelligence tools. It is optimized for datasets ranging from a few hundred gigabytes to a petabyte or more and costs less than $1,000 per terabyte per year, a tenth the cost of most traditional data warehousing solutions.

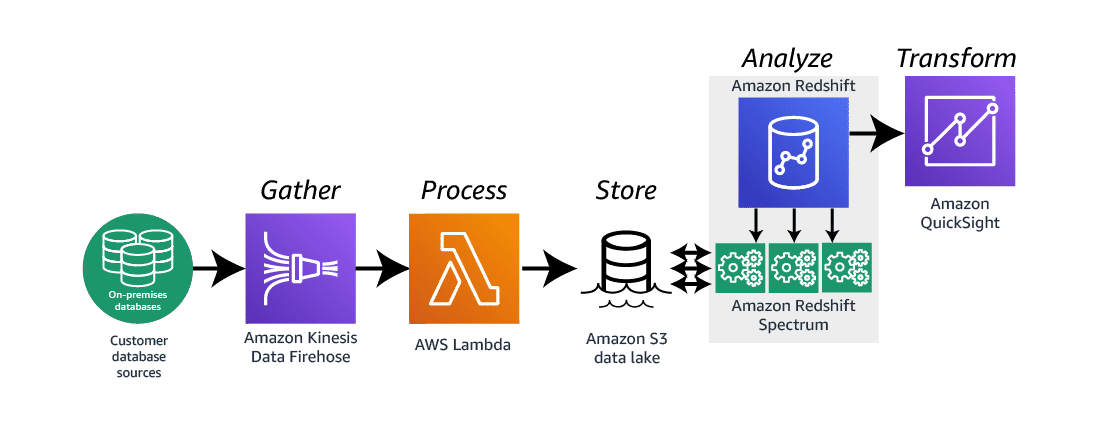
Function of Redshift

* It supports Online Analytical Processing (OLAP) type of DB workloads and analyzes them using standard SQL and existing Business Intelligence (BI) tools (AWS QuickSight or Tableau).
* It is used for executing complex analytic queries on semi-structured and structured data using query optimization, columnar-based storage, and Massively Parallel Query Execution (MPP).
* Redshift Spectrum helps to directly query from the objects (files) on S3 without actually loading them.
* It has the capability to automatically copy snapshots (automated or manual) of a cluster to another AWS Region

**Rich data platform architecture**



**Event-driven data analysis architecture**



Step 1

Getting started with Amazon Redshift

Amazon Redshift cluster and shows you how to both load and query your data.

**Preparing the environment**

* **Create a VPC**
* **Create the subnets**
* **Create an Amazon S3 bucket**
* **Create an Amazon S3 endpoint**
* **Create an IAM role**
* **Create the security groups**
* **Create the subnet group**

1. **Creating Cluster and testing the database**

**Sign in to the AWS console**

Graphical user interface, website

Description automatically generated

**Search Redshift in search bar and select it**

A screenshot of a computer

Description automatically generated

**On left hand side you can see the Cluster select that and then click on Create Cluster Button**

Graphical user interface, text, application, email

Description automatically generated

**Cluster Configuration**

* **Name or cluster identifier**
* **Select free trial**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, Word

Description automatically generated

**Cluster is created**

**Now you can select the cluster by clicking on its name**

Graphical user interface, text, application

Description automatically generated

**Step 2**

**After cleating cluster now it’s time to query run**

**Select Query data/query in query editor**

Graphical user interface, application

Description automatically generated

**We need to connect to database to run queries**

**Using Query Editor 2 you can see the different interface as same as Postgress SQL in that we have databases, schemas, tables, etc.**

* **Create database**
* **Create tables, schemas**
* **Load data**
* **Select database**
* **Run Query**
* **See the result in downside**

A screenshot of a computer

Description automatically generated

**Create**

Graphical user interface, text, application, chat or text message

Description automatically generated

**Load data;**

* **We can load data from S3 Bucket**
* **Click on Browse S3**

Graphical user interface, application

Description automatically generated

**Choose S3 bucket**

Table

Description automatically generated

**Check on required data file / object which is shown in list**

**Click on choose in down right-side corner.**

Graphical user interface, text

Description automatically generated

Graphical user interface, application

Description automatically generated

**Save Query**

**You can also save query**

Graphical user interface

Description automatically generated

**You can find those saved queries in Queries**

Graphical user interface, application

Description automatically generated

**Chart**

**You can also plot different charts for data and queries visualization**

**You can also save chart and visualize it later.**

Chart

Description automatically generated

**Different chats you can prepare**

* **Bar**
* **Scatter**
* **Histogram**
* **Area**
* **Etc….**

Graphical user interface, application, Word

Description automatically generated

**You can also export the chart and share with respected one.**

Graphical user interface, text, application, email

Description automatically generated

**Delete Cluster**

**By going to cluster dashboard, you can delete cluster of Amazon Redshift in Action button.**

Graphical user interface, text, application, email

Description automatically generated

**Write Delete word in it and click on delete cluster**

**Your cluster is deleted**

1. **Connecting your cluster by using SQL Workbench/J**

**Amazon Redshift doesn't provide or install any SQL client tools or libraries, so you must install any that you want to use with your clusters. If you already have a business intelligence application or any other application that can connect to your clusters using a standard PostgreSQL JDBC or ODBC driver, then you can skip this section. If you don't already have an application that can connect to your cluster, this section presents one option for doing so using SQL Workbench/J, a free, DBMS-independent, cross-platform SQL query tool.**

Install SQL Workbench/J

* Follow the link download the SQL Workbench/J

[**https://www.sql-workbench.eu/downloads.html**](https://www.sql-workbench.eu/downloads.html)

* **download this link Zip file and extract it in one folder**

[**https://www.sql-workbench.eu/Workbench-Build128-with-optional-libs.zip**](https://www.sql-workbench.eu/Workbench-Build128-with-optional-libs.zip)

Graphical user interface, text, application, email

Description automatically generated

**Extracted folder look like this**

Graphical user interface, application, table

Description automatically generated

**Click** **SQL Workbench Application**

**Note**

SQL Workbench/J requires the Java Runtime Environment (JRE) be installed on your system.

If not please install it otherwise it will give you an option to download it and install it click on that button “Download JRE”.

The SQL Workbench connection dashboard look like this.

You must provide following details of your Redshift cluster

* Set Profile name
* Driver – which can be downloaded from AWS management console after cluster is created.
* URL – Endpoint to connect
* Username and Password.
* Click on connet

Graphical user interface, text, application, email

Description automatically generated

Download JDBC driver shown in Redshift dashboard.

Graphical user interface, text, application

Description automatically generated

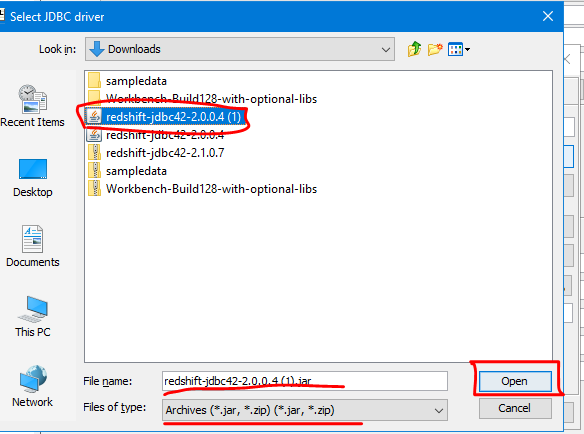
**Select that .jar file that we have downloaded just above, browser in Manager driver**

* **Click on file button that I have red marked to browsed .jar file.**

Graphical user interface, text, application

Description automatically generated

**Select redshift JDBC .jar file**



Graphical user interface, application, Word

Description automatically generated

**Provide all details**

* + **Username**
  + **Password**
  + **URL – check url from cluster details**

Graphical user interface, text, application, email

Description automatically generated

**Open redshift cluster by clicking on its name**

Graphical user interface, text, application, email

Description automatically generated

**Go to general information of cluster**

**Copy the endpoint of cluster**

Graphical user interface, text, application

Description automatically generated

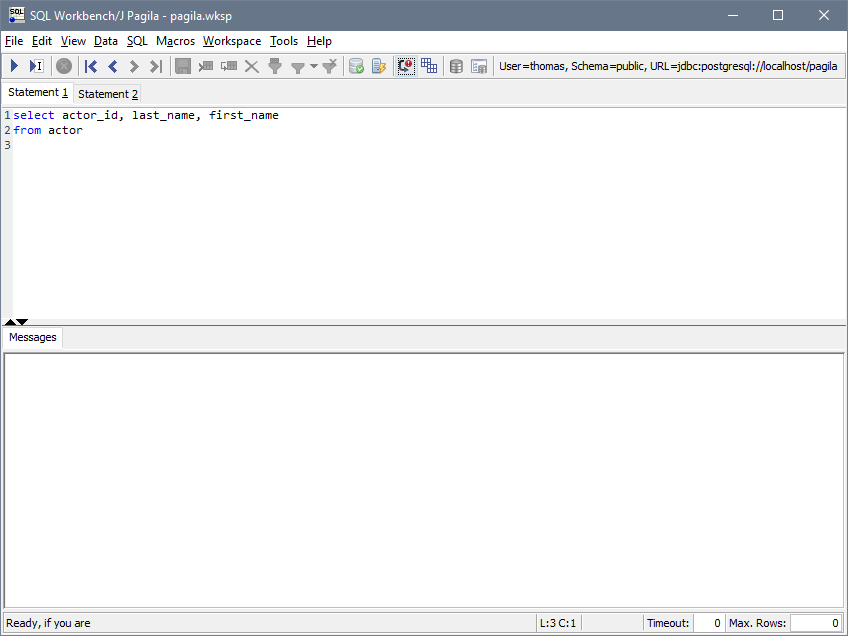
**Past it into the URL field on SQL Workbench /J**

Graphical user interface, text, application, email

Description automatically generated

**Click on okay is**

* **Now you can run query on it**
* **You can create table and run query and also do other**
* **Data analysis**



**This is how we can connect to SQL Workbench / J to our cluster and run Queries.**

**If you want don’t want to use this created cluster, we can Delete it**

* **Go to redshift cluster**
* **Click on Action**
* **Select Delete**
* **You can also Resize, Reboot , Pause.**

Graphical user interface, application

Description automatically generated

**To confirmation delete enter delete word in field and click on delete cluster.**

Graphical user interface, text, application, email

Description automatically generated

**Your cluster is deleted**