Redshift Standards and Support

Overview

Redshift is the standard database for data warehousing and data lake deployments at GP. Redshift is a highly available and scalable clustered environment based on the PostgreSQL database engine technology.

Redshift has many attractive features that make it a good choice for data warehousing deployments.

- Clusters can be scaled horizontally since compute nodes can all access the same underlying data.
- Data can be contained in the Redshift native tables, S3 buckets or be accessed from other database platforms directly allowing dynamic querying of heterogenous data sets.
- These datasets can be queried using well known PostgreSQL engine and SQL and procedure language.
- You can see a summary of more features here: AWS Redshift Features

You can review Redshift Application Development standards that were developed in collaboration by the cross business Data Architecture Group: GP Development Standards & Best Practices for Amazon RedShift

Support coverage, SLAs and service request models are the same for Redshift as for all databases that Database Services supports.

DBTS will consult with you on which instance type and instance size is best for your application implementation. This includes determining appropriate node sizing, performance considerations, appropriate security measures (setting of SSL) and recovery options. Engage a DBTS DBA early in your project to ensure the best deployment configuration and a smooth transition into implementation as we install and configure your database for use.

Implementation

Considerations

- Initial Size and Data Growth
- Processing Power and Memory Needs
- Instance type (RA3, DC)
- Standard Recovery Objectives
- Disaster Recovery Objectives and Tier Placement
- SSO Setup requirement
- Audit Setup for Redshift Instance

Standard Maintenance

- Backups (See Backup and Recovery Section)
- Adhoc Vacuum and Analyze
- Auto Vacuum and Analyze setup
- Standard Collections and Monitoring

Micro Account Deployment

DBTS prefers deployment of Redshift Instances on DBA owned micro accounts, this enables easy management, consolidation and security segmentation. For details of DBTS Micro Accounts see the **DBTS Micro-Account Details** section of the Standards, Roadmap & FAQ section.

DBTS acknowledges that there are situations where Redshift may need to be implemented in an application owned micro account due to throughput or AWS Service limitations. If you have reason to believe Redshift should be deployed in an application micro account, please consult DBTS early in your project as we will need to gain approvals and do security prework to enable us to deploy and support the implementation.

Production Specific settings

- Backups
- SSL setup
- Audit setup for Redshift instance to S3
- WLM and other policy settings

Backup and Recovery

Standard Backups: Redshift

- Snapshots are point-in-time backups of a cluster. There are two types of snapshots: automated and manual.
- Amazon Redshift stores these snapshots internally in Amazon S3 by using an encrypted Secure Sockets Layer (SSL) connection.
- When we restore from a snapshot, Amazon Redshift creates a new cluster and makes the new cluster available before all of the data is loaded, so you can begin querying the new cluster immediately. The cluster streams data on demand from the snapshot in response to active queries, then loads the remaining data in the background.

Automated snapshots

- By default, Amazon Redshift takes a snapshot about every eight hours or following every 5 GB per node of data changes, or whichever comes first.
- Alternatively, we can create a snapshot schedule to control when automated snapshots are taken. Automated snapshots are enabled by default when we create a cluster.
- Automated snapshots are deleted at the end of a retention period. The default retention period is one day, but we can modify it. (Maximum is 35 days)
- Amazon Redshift deletes automated snapshots at the end of a snapshot's retention period, when we disable automated snapshots for the cluster, or when we delete the cluster.
- We can't disable automated snapshots for RA3 node types. Only Amazon Redshift can delete an automated snapshot; you cannot delete them manually.

Manual snapshots

- We can take a manual snapshot any time. By default, manual snapshots are retained indefinitely, even after we delete your cluster. However, we can specify the retention period when we create a manual snapshot, or we can change the retention period by modifying the snapshot.
- AWS limits the number of manual snapshots you can take to 20 snapshots per account.
- Manual snapshots can be scheduled from Lambda.

Recovery

- In Redshift, customer must re-load source files that were loaded after their snapshot was taken.
- RPO is reliant on Redshift Snapshot schedules

- RTO is reliant upon the DB size / node-count throughput.
- Redshift Still does not support individual database backup and restore. (But this can be still done using DDL transfer)
- Table restores are possible from snapshot.

1 person liked this 85 Views Save for later