

AWS Certified Solutions Architect: Associate - 7.0 Databases and AWS

filename: amazon-acsaa-7-2-redshift

Title: Amazon Redshift

Subtitle: AWS Certified Solutions Architect: Associate

7.2 Amazon Redshift

- Basics
 - Petabyte-scale Data Warehouse Service
 - Designed for high OLAP performance
 - Based on PostgreSQL
 - Uses standard SQL for most operations
 - INSERT
 - UPDATE
 - SELECT
 - Some specialized commands
 - VACUUM (reorganizes data)
 - UNLOAD (exports data)
- Clusters
 - Components (Diagram)
 - Lead Node
 - Clients interact only with the lead node
 - 1 or more Compute Nodes
 - Transparent to the clients
 - Maximum varies based on node type
 - Six different types in two categories
 - Dense Compute
 - dc1.large
 - dc1.8xlarge
 - Dense Storage
 - ds1.large
 - ds1.8xlarge
 - ds2.large
 - ds2.8xlarge
 - Cluster can contain multiple databases
 - Client communications use JDBC or ODBC with the leader node
 - Query workload is distributed amongst the compute nodes by the leader
 - Uses a user defined *distribution strategy* to optimize data placement on the nodes
 - EVEN (Default): Data is uniformly spread amongst the compute nodes
 - KEY: Data is distributed based on the values of one column
 - ALL: A full copy of DB data is placed on each node
 - *Sort Keys* also help to improve query performance
 - Cluster can be resized at any time, but will be read-only during the resize
 - Redshift automatically applies *Compression Encoding* on each column in a table
 - Supports multiple methods of compression and will automatically choose the best one
- Other features
 - Snapshots
 - Automated and manual
 - Retention period is configurable
 - Security
 - IAM support at the service level
 - DB Users/groups at the DB level
 - SSL/VPC/ACLs at the network level
 - Encryption with AWS CloudHSM can provide data at rest security