What is Amazon Redshift?

02:07 PM

Amazon Redshift is a fully managed, petabyte-scale data warehouse service in the cloud. You can start with just a few hundred gigabytes of data and scale to a petabyte or more. This enables you to use your data to acquire new insights for your business and customers.

The first step to create a data warehouse is to launch a set of nodes, called an Amazon Redshift cluster. After you provision your cluster, you can upload your data set and then perform data analysis queries. Regardless of the size of the data set, Amazon Redshift offers fast query performance using the same SQL-based tools and business intelligence applications that you use today.

Amazon Redshift concepts
Following are some key Amazon Redshift concepts:

Cluster – The core infrastructure component of an Amazon Redshift data warehouse is a cluster.

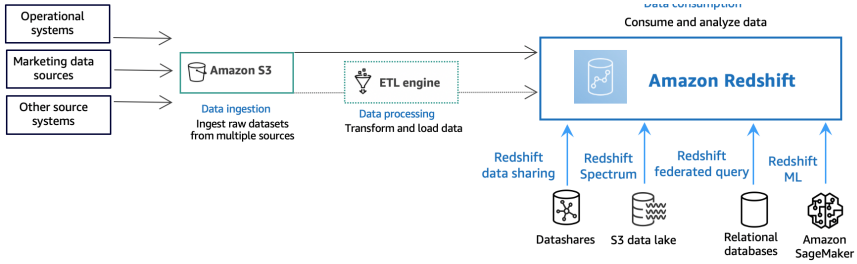
A cluster is composed of one or more compute nodes. The compute nodes run the compiled code.

If a cluster is provisioned with two or more compute nodes, an additional leader node coordinates the compute nodes. The leader node handles external communication with applications, such as business intelligence tools and query editors. Your client application interacts directly only with the leader node. The compute nodes are transparent to external applications.

Database – A cluster contains one or more databases.

User data is stored in one or more databases on the compute nodes. Your SQL client communicates with the leader node, which in turn coordinates running queries with the compute nodes. For details about compute nodes and leader nodes, see Data warehouse system architecture. Within a database, user data is organized into one or more schemas.

Amazon Redshift is a relational database management system (RDBMS) and is compatible with other RDBMS applications. It provides the same functionality as a typical RDBMS, including online transaction processing (OLTP) functions such as inserting and deleting data. Amazon Redshift also is optimized for high-performance batch analysis and reporting of datasets.



Lab: Create a sample Amazon Redshift cluster

https://docs.aws.amazon.com/redshift/latest/gsg/rs-gsg-launch-sample-cluster.html

https://docs.aws.amazon.com/redshift/latest/gsg/data-loading.html

1. Sample Clustor for learning
2. Bring your own data set for prod ready