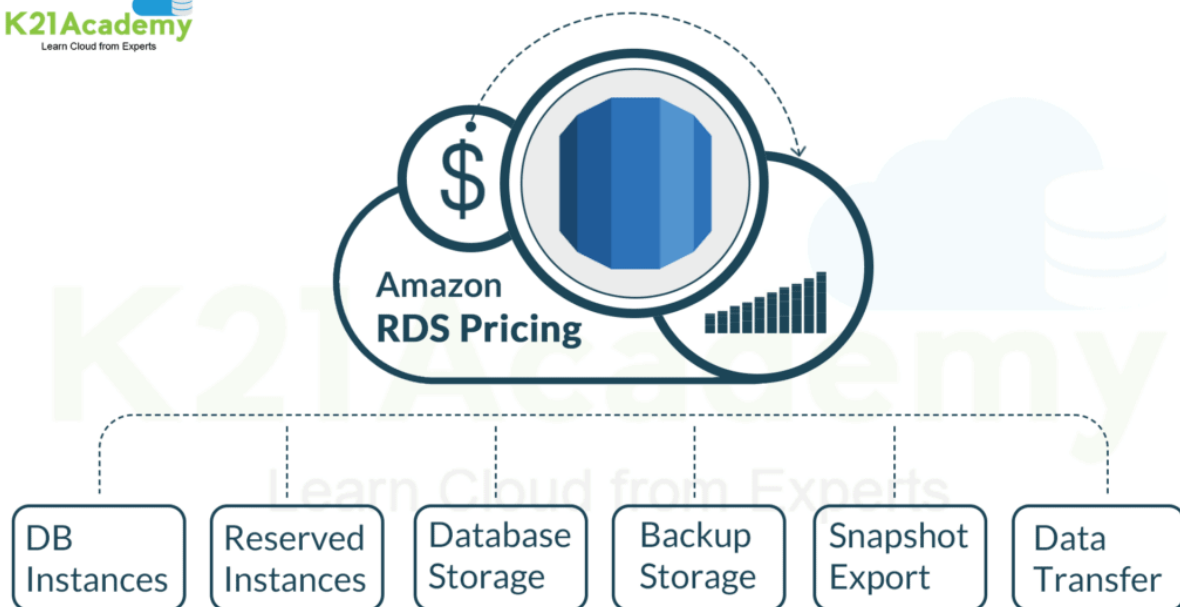# Amazon RDS

What is Amazon RDS?

Amazon RDS is a **Relational Database Service** by Amazon Web Services. A relational database is a type of database that stores data in tables with rows and columns. Amazon RDS provides an industry-standard relational database that is cost-efficient with resizable capacity. It gives users solutions for easier setup, operation, and scale of a relational database in the cloud.

Features of Amazon RDS

- It manages software patching, automatic failure detection, backups, and recovery.
- Offers flexibility to opt for automated backups or manual backup snapshots. In case of data loss, you can restore the backups at any time.
- Flexibility to pick your database engine among **MySQL**, **MariaDB**, **PostgreSQL**, **Oracle**, and **Microsoft SQL Server**.
- Enable more security with **AWS IAM** (Identity and Access Management) to ensure that access remains only for the selected users with limited permissions.

Amazon RDS Pricing

Amazon RDS is free to try, and you can test your environment with basic settings. If you are setting up Amazon RDS for your organization with more frequent usage, then the charges depend on the usage of the resources. There are various database engines on which the Amazon RDS pricing depends like **Amazon Aurora**, **My SQL**, **PostgreSQL**, **Maria DB**, **Oracle**, and **Microsoft SQL Server**. On running Amazon RDS, the charges are made on the following parameters.

- On-Demand DB Instances
- Reserved Instances
- Database Storage

- Backup Storage
- Snapshot Export
- Data Transfer

Configuration

Engine type **Info**

- ○ Amazon Aurora
- ● MySQL
- ○ MariaDB
- ○ PostgreSQL
- ○ Oracle
- ○ Microsoft SQL Server

DB instance size

○ Production	○ Dev/Test	● Free tier
db.r6g.xlarge	db.r6g.large	db.t2.micro
4 vCPUs	2 vCPUs	1 vCPUs
32 GiB RAM	16 GiB RAM	1 GiB RAM
500 GiB	100 GiB	20 GiB
1.017 USD/hour	0.231 USD/hour	0.020 USD/hour

Snowflake vs Redshift : Modern-Day Data Warehouse Solutions

Both Snowflake and Redshift are two big names in this field, and they both offer similar services. They're extensive data analytics databases capable of reading and analyzing massive amounts of data. In this post, I will share everything you need to know about **Snowflake vs Redshift.**
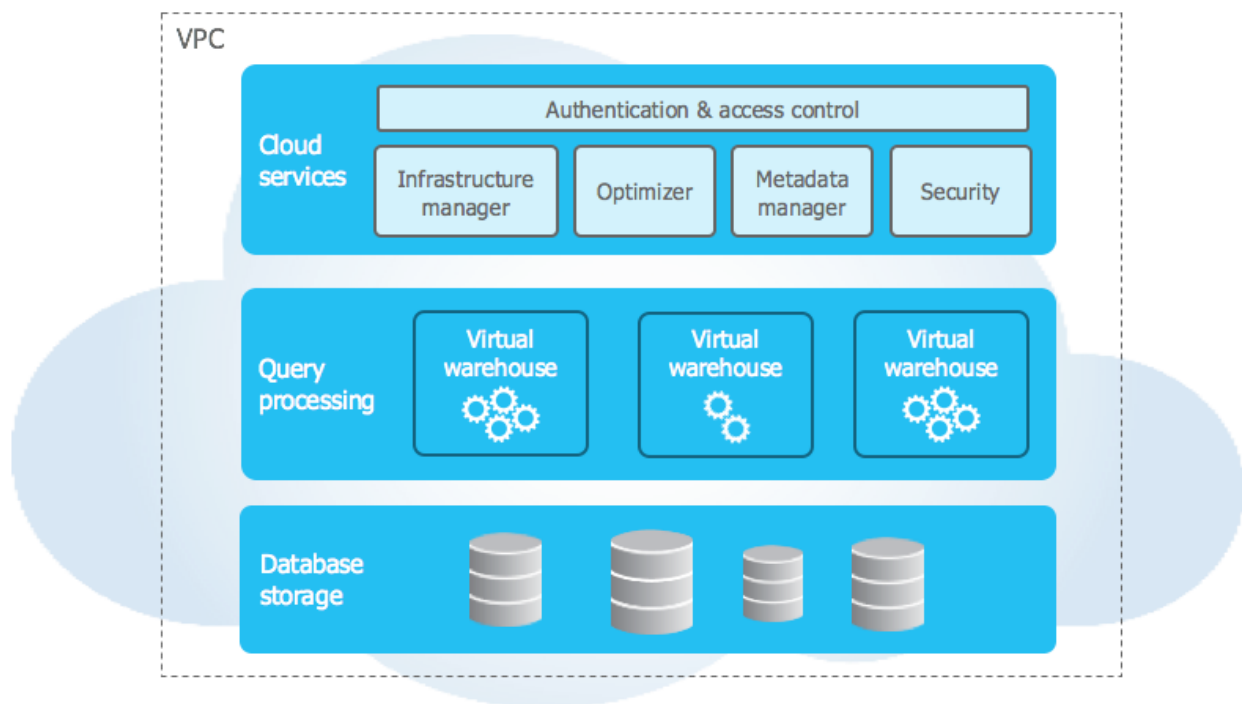
What is Snowflake?

Snowflake is a **cloud-based data warehouse** that runs on **Amazon Web Services** or **Microsoft Azure**. It's great for enterprises that don't want to devote resources to the setup, maintenance, and support of in-house servers because there's no hardware or software to choose, install, configure, or manage. And utilizing any ETL tool, data can be migrated into Snowflake. Snowflake's design and data exchange capabilities, however, set it distinct. Customers can utilize and pay for storage and computing separately thanks to the Snowflake architecture, allowing storage and computing to scale independently.

How does Snowflake work? – Architecture

Snowflake follows a straightforward, **flexible architecture**. It separates the storage and computes tasks, so businesses that need a lot of storage but don't need a lot of CPU cycles, or vice versa, don't have to pay for an integrated package that includes both. Users can **scale up or down** according to their needs and only pay for the resources they utilize. Storage is charged in terabytes per month, whereas computation is charged per second.

Source: Snowflake Docs

Snowflake Architecture covers the following three layers.

Database Storage
When data is imported into Snowflake, it is reorganized into a columnar format that is internally optimized and compressed. Snowflake saves the optimized data to the cloud.

Snowflake is in charge of all aspects of data storage, including organization, file size, structure, compression, metadata, statistics, and other aspects of data storage. Customers cannot see or access the data objects stored by Snowflake; they can only access them through SQL query operations executed through Snowflake.

Query Processing
The processing layer is where queries are executed. Snowflake employs "virtual warehouses" to process queries. Each virtual warehouse is an MPP compute cluster made up of numerous compute nodes procured from a cloud provider by Snowflake.

Each virtual warehouse is its own compute cluster, with no other virtual warehouses sharing compute resources. As a result, the performance of one virtual warehouse has no bearing on the performance of others.

Cloud Services
The cloud services layer organizes the entire system and uses ANSI SQL. It reduces the requirement for data warehouse management and tuning to be done manually. This layer contains the following services:

- Authentication
- Infrastructure management
- Metadata management
- Query parsing and optimization
- Access control

Features of Snowflake

Let's quickly go through the features of Snowflake.
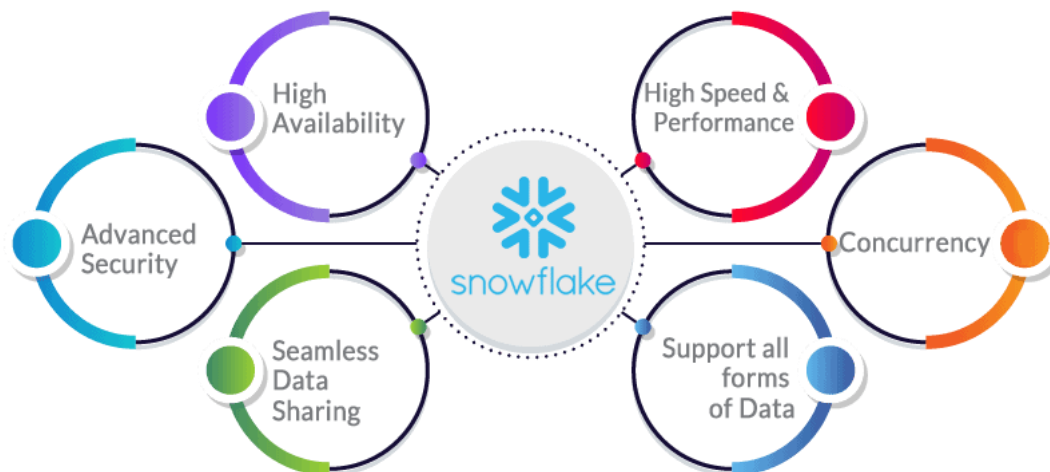
- **High Availability and Advanced Security:** Snowflake is designed to run constantly and suffer component and **network failures with minimal impact on clients.**
- **Free Data Sharing:** Snowflake's architecture allows users to **share data with one another**. It also enables businesses to share data with anyone, whether or not they are a Snowflake customer.
- **High Performance:** Because of the cloud's elasticity, you can scale up your virtual warehouse to take advantage of more compute resources if you need to **load data faster** or perform a large number of queries.

- **Concurrency:** Snowflake's innovative **multicluster architecture** addresses concurrency issues: queries from one virtual warehouse never influence queries from another, and each virtual warehouse may scale up or down as needed.
- **Support Structured and Semi-structured Data**: For analysis, you can aggregate structured and semi-structured data and load it into a cloud database without first converting or transforming it into a set relational schema.

Disadvantages of Snowflake

As we move deeper to know more about Snowflake vs Redshift, let's have a quick look at the cons of Snowflake.

- Snowflake is the wrong choice if you're running a firm that relies on on-premise technology that is difficult to combine with cloud-based services.
- When you establish a virtual warehouse, a minute's worth of Snowflake credits will be spent, but you will be charged by the second after that.
- The SQL editor in Snowflake should also be upgraded to handle autocomplete functions far better than it does today.

What is Redshift?

Amazon Redshift is a **cloud-based, fully managed petabyte-scale data warehousing service.** Starting with a few hundred gigabytes of data, you may scale up to a petabyte or more. This allows you to gain fresh insights for your company and customers by analyzing your data. The first step in creating a data warehouse is to set up an Amazon Redshift cluster, which is a collection of machines. You can upload your data set and then run data analysis queries after you've provisioned your cluster. Regardless of the size of the data set, Amazon Redshift provides rapid query performance using the same SQL-based tools and business intelligence applications you're already familiar with.

How does Redshift work? – Architecture

Let's try to understand the working of Redshift through its components.

Redshift Cluster: The basic infrastructure component of Redshift is a cluster of nodes. A cluster typically consists of one leader node and several compute nodes. There is no additional leader node when there is only one compute node. Each **compute node** has its own processor, memory, and storage disc. Client applications are unaware of the existence of compute nodes and never interact with them directly.
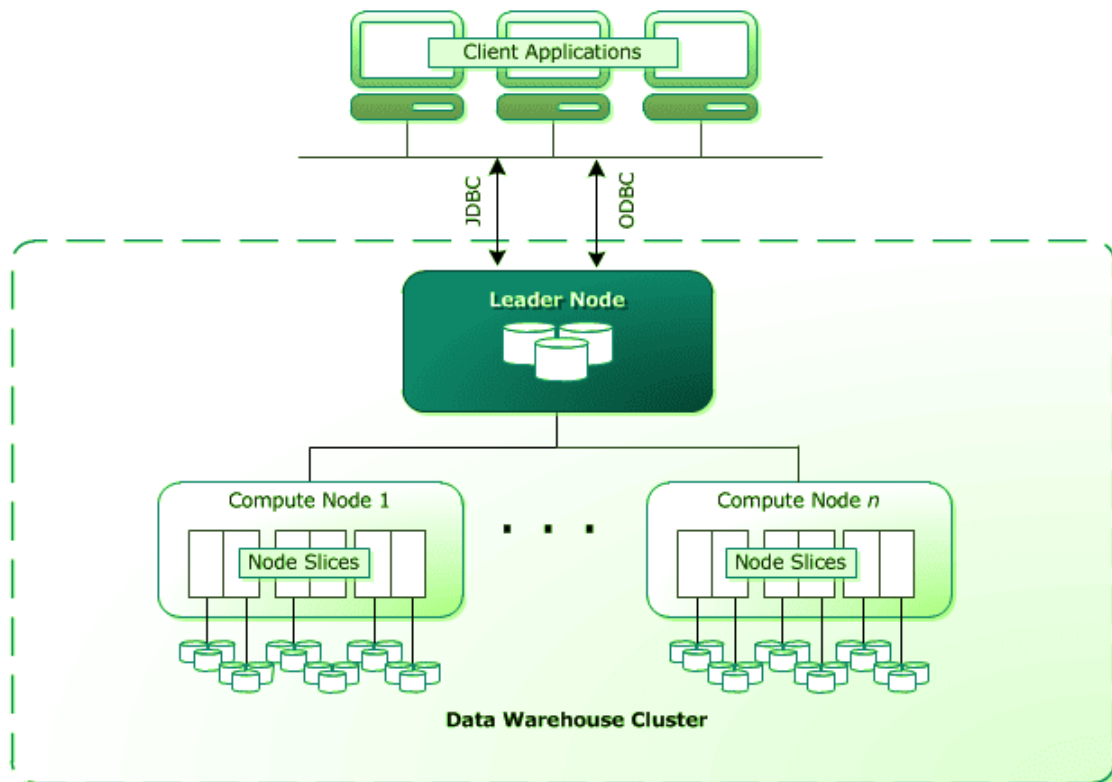
The **leader node** is in charge of all communications with the client apps. The leader node is also in charge of compute node coordination. The leader node is also in charge of query parsing and execution plan development. When the leader node receives a query, it constructs an execution plan and distributes the produced code to the compute nodes. Each compute node is given a piece of the data. The leader node is in charge of aggregating the results in the end.

Users can choose between two types of nodes in Redshift: **Dense Storage nodes and Dense Compute nodes.** Customers can choose them based on the nature of their needs, whether they are stored or compute-intensive. The number of nodes in Redshift's cluster can be increased, individual node capacity can be increased, or both.

Features of Redshift

Let's have a quick look at the features of Redshift.

- Amazon Redshift is a **fully managed service** that requires very little user participation.
- It has the capacity to **store a Petabyte of data.**
- Scaling requires very little effort and is only limited by the customer's financial ability to pay.
- Amazon **Redshift is faster** than most data warehouse systems, and it has a distinct advantage when it comes to running sophisticated queries repeatedly.
- **On-demand pricing** is available through Amazon Redshift. This, together with the ability to create clusters from snapshots, can help customers better manage their budgets. It also offers a lot of flexibility when it comes to selecting node types for various workloads.
- It comes with a **full set of security features**, including lighting.

Difference between Snowflake vs Redshift

Snowflake	Redshift
Snowflake Elastic **Data Warehouse** is a cloud-based **data storage and analytics** service offered by Snowflake Computing.	Amazon Redshift is a cloud-based, fully managed **petabyte-scale** data warehousing service.
Snowflake is available on the **AWS Marketplace and Microsoft Azure** and has some extremely amazing on-demand features.	Redshift should be your first pick if you're working with the Amazon ecosystem.
Security and compliance **options vary by tier.**	Security and compliance are enforced in a **comprehensive fashion** for all users.

More **automated database maintenance** features.	**Manual maintenance.**
Snowflake offers **instant scaling.**	Redshift **takes minutes to add more nodes.**
Snowflake is a bit more expensive as it charges separately for warehousing and computation.	Amazon's **Redshift is less expensive** than Snowflake.

Disadvantages of Redshift

Now we will have a look at the cons of Redshift.

- This isn't a good fit for transactional systems.
- While you wait for AWS to release a new patch, you may need to revert to an older version of Redshift.
- The cost of Amazon Redshift Spectrum is determined by the number of bytes scanned.
- Redshift is out of date in terms of features and data types, and the dialect is very similar to PostgreSQL 8.
- Hanging queries in external tables can cause complications.