AWS REDSHIFT VS SNOWFLAKE: WHAT'S THE DIFFERENCE & HOW TO CHOOSE?

Snowflake and Amazon Redshift are two popular cloud-based data warehousing platforms that offer outstanding performance, scale, and business intelligence capabilities. Certainly, both platforms offer similar core functionalities, such as:

- Relational management
- Security
- Scalability
- Cost efficiency

The key differences, however, are their pricing models, deployment options, and user experience.

In this article, we'll help you decide whether AWS Redshift or Snowflake is right for you. Let's compare these two solutions based on their similarities, differences, and use cases. We also highlight how each platform addresses common challenges faced by businesses looking to implement a data warehouse.

AWS Redshift and Snowflake are two popular data warehouses. How do you know which one's right for you? Let's take a look.

Choosing the right data warehouse

First, let's briefly look at data warehouses in general.

Data warehouses (DWH) are large repositories of data, collected from different data sources, which

organizational typically use for analytical insights and business intelligence. An efficient data warehouse relies on an architecture that offers consistency by collecting data from different operational databases, then applying a uniform format for easier analysis and quicker insights.

One of the fundamental purposes of data warehouses is to enable quick access to historical data and context, thereby helping decision-makers to optimize strategies and improve bottom lines.

Implementing the right data warehousing solution is key to gaining a competitive advantage in today's data-centric business world. Leveraging an efficiently provisioned business intelligence framework, a data warehouse supports business outcomes such as:

- Increased bottom line
- Efficient decision making
- Enhanced customer service
- Improved analytics

The most important characteristics of any efficiently designed data warehouse is to ensure that it:

- Has consistent schemas across different tables that returns expected results against a query.
- Supports multi-table querying out of the box, which allows users to generate ad hoc reports without writing custom code or creating custom table views.

Some key factors to consider when selecting a warehousing platform include:

- Business goals
- Cost models
- Simplicity of integration
- Cloud-readiness
- Adherence to security and compliance standards

What is Snowflake?

Snowflake is a cloud-based, software as a service (SaaS) data platform that allows

- Secure data sharing
- Unlimited scaling
- A seamless multi-cloud experience

The platform relies on a virtual warehouse framework that leverages third party cloud-compute resources such as AWS, Azure, or GCP. The option to choose high-performance cloud platforms allows real-time auto-scaling to organizations who are looking to:

- Run faster workloads
- Process large query volumes on the elastic cloud

As compared to legacy DWH solutions, Snowflake offers a non-traditional approach to data warehousing by abstracting compute from storage. That means data can reside in a central repository while compute instances are sized, scaled, and managed independently.

Snowflake manages all aspects of data administration for a simpler, more flexible warehousing solution that provides various capabilities of enterprise offerings.

The Snowflake analytics platform leverages a custom [SQL query engine](#) and three-layer architecture to support [real time analytics](#) of [streaming big data](#). Its flexible architecture allows users to build their own analytical applications without having to learn new [programming languages](#).

(Check out our [Snowflake Guide](#).)

Benefits of Snowflake

- Organizations don't need to install, configure, or manage the underlying warehouse platform, including hardware or software
- Integrates with most components of the data ecosystem
- Separates configuration, management and charges for storage and compute instances
- Offers an intuitive, powerful SQL interface
- Enables account-to-account data sharing
- Simple to set up and use

When to use Snowflake

Snowflake is considered the perfect data warehouse solution for situations when…

- The query load is expected to be lighter.
- Workload requires frequent scaling.
- Your organization requires an automated, [managed solution](#) with zero operational overhead to manage the underlying platform.

Now let's turn to Redshift.

What is Amazon Redshift?

AWS Redshift is a data warehousing platform that uses cloud-based compute nodes to enable large scale data analysis and storage. The platform employs column-oriented databases to connect business intelligence solutions with SQL-based query engines. By leveraging [PostgreSQL](#) and Massively Parallel Processing (MPP) on dense storage nodes, the platform delivers quick query outputs on large data sets.

While offering faster query processing, Redshift also offers multiple options for efficient management of its clusters. These include:

- Interactively using the AWS CLI or Amazon Redshift Console
- Amazon Redshift Query API
- AWS Software Development Kit

Amazon Redshift is a fully managed warehousing platform that allows organizations to query and combine petabytes of data with optimized price performance. The Advanced Query Accelerator (AQUA) offers a cache that boosts query operations performance by up to 10x, allowing businesses to gain new insights from every data point in the application/system.

(Explore our hands-on [AWS Redshift Guide](#).)

Benefits of AWS Redshift

- Offers a user-friendly console for easier analytics and query
- A fully managed platform that requires little effort towards maintenance, upgrading and administration
- Integrates seamlessly with the AWS services ecosystem
- Supports multiple data output formats
- Works seamlessly with SQL data using PostgreSQL syntax

When to use Redshift

AWS Redshift is considered the perfect data warehouse solution for situations when…

- Your organization is already using AWS services.
- Workloads run structured data.
- The application has a high query load.

Structured data
- Difficult to collect
- Affordable to collect, process
- Limited insights
- Purpose-driven
- Requires active participation
- Transparency promotes privacy

Unstructured data
- Easy to collect
- Pricier to collect, process
- Nearly infinite insights
- Reusable
- Requires presence only
- Lack of transparency, privacy

AWS Redshift vs Snowflake: A quick comparison

Let's look at the clear differences between the two.

- Snowflake is a complete SaaS offering that requires no maintenance. AWS Redshift clusters require some manual maintenance
- Snowflake separates compute from storage, allowing for flexible pricing and configuration. Redshift allows for cost optimization through Reserved/Spot instance pricing.
- Snowflake implements instantaneous auto-scaling while Redshift requires addition/removal of

nodes for scaling.

- Snowflake supports fewer data customization choices, where Redshift supports data flexibility through features like partitioning and distribution.
- Snowflake supports *always-on* encryption that enforces strict security checks while Redshift provides a flexible, customizable security model.

Similarities between Snowflake & Redshift

- Both support Massive Parallel Processing (MPP) for faster performance.
- Both the platforms connect BI solutions to databases using column-oriented databases.
- Data in both warehouses is accessed using SQL based query engines.
- Both Snowflake and Redshift are designed to abstract data management tasks so users can easily gain insights and improve system performance using data-driven decisions.

Choosing Snowflake or Redshift

In the modern data-driven world, data warehousing solutions allow organizations to store large sets of operational data and make holistic analytical decisions to improve system performance.

DWHs are designed to store vast amounts of structured or semi-structured data to provide fast retrieval times and easy analytics.

Redshift and Snowflake are two top cloud-based data warehouses that offer powerful data management and analysis options. Both the platforms also offer:

- High availability with minimal downtime
- Scalability through replication across multiple servers

While both the platforms are highly popular and each outclasses the other meagerly in offering benefits, the choice between the two platforms depends on business demands, resources, bundled services, and specific use cases.

Related reading

- BMC Machine Learning & Big Data Blog
- MongoDB vs DynamoDB: Comparing NoSQL Databases
- How To Load Data to Amazon Redshift from S3
- Big Data vs Data Analytics vs Data Science: What's The Difference?
- The State of SaaS Today: Growth Trends & Statistics
- Data Ethics for Companies