



Everything You Need to Know About Migrating Your Data Warehouse to Amazon RedShift

A Whitepaper



Data is the new oil. Businesses are leveraging it to devise strategies, build new business models, focus on product and service innovation, improve operational efficiency, and increase customer satisfaction and competitive advantage.

Especially in the post-pandemic world, the need to innovate, build business resilience, and anticipate trends has spurred the growth of data warehouse solutions.

The Need for Modern Data Warehouse

One [market research](#) report estimates the global data warehousing market to grow from \$21.18 billion in 2019 at a CAGR of 10.7% to touch \$51.18 billion by 2028.

Data warehousing refers to building a data warehouse to store data from heterogeneous sources. The report further identifies the factors spurring this growth as:

- The need to improve data quality and the decision making capability
- Deliver improved business intelligence solutions to the end-users
- Gain competitive advantage

However, for many, data is also a challenge. It is available across the organization but distributed in silos. It is available in many formats and can be a combination of structured and unstructured. This poses a limitation in being able to gain a holistic view.

Scalability is a limitation posed by the traditional approach to data management. It involves high costs and additional servers for computing and storage. Load balancing can also be difficult during peak periods.



Businesses need a scalable, cost-effective, secure, and compliant solution to break the silos and work with a single source of truth. This is offered by modern, cloud-based data warehouses such as Amazon Redshift.

Amazon Redshift, a fully managed data warehouse, is becoming a popular choice for businesses wanting to migrate to cloud-based solutions that promise speed and scalability. Redshift facilitates data analytics at scale with good performance and low costs.

Redshift—A Modern Data Architecture

One of the key advantages of Redshift is the access to unlimited storage and the ability to integrate data from multiple sources. It facilitates secure access to data with data governance controls for running analytics or machine learning.

Some of the key benefits that make Amazon Redshift a popular choice include:

Data Analytics

Amazon Redshift allows access to data from across organizational data warehouses and data lakes for analytics without compromising security and governance policies. Amazon Redshift Spectrum can be used to query data without prior loading or preparation.

The results of the query can be saved in the data lake using data lake export.

As Amazon Redshift is integrated with the data lake, it makes analytics in real-time and use cases of ML/AI possible without the need for re-architecture.

Data sharing capabilities enable collaboration internally and externally across Amazon Redshift clusters, providing access to a single source of truth. ML models can be created, trained, and deployed using Amazon Redshift ML.



Performance at Scale

Amazon Redshift can scale up quickly without the need for defining sort keys and distribution keys or tuning the data warehouse. It can self-tune, self-learn, and is capable of auto-refresh, materialized views, and auto-query rewrite to match scaling needs for data or the number of users.

Its concurrency scaling feature helps to scale up compute resources in case of additional concurrent users to manage the extra load. Since storage and compute are decoupled, each can be scaled independently, based on need.

Self-Service

Being a fully managed data warehouse, Redshift does not require much performance optimization or infrastructure management. Users can run commands using common SQL syntax customize responses using different visualizations.

Data can also be accessed securely from any application without the need to configure software drivers or manage database connections. Users can also access data from the BI tool leveraging Amazon Redshift's integration capabilities.

Factors to Consider when Migrating to Amazon Redshift

Data warehouse migration is a complex process with many associated risks related to time, cost, and resources. But with the right data warehouse migration strategy, this challenge can be overcome.

The factors to keep in mind when devising the strategy are the current data warehouse landscape and desired level of transformation required when migrating to Amazon Redshift.



When creating your migration strategy, consider the following:

Size

It is important to have a good understanding of the data domains and sources that need to be migrated to Amazon Redshift. This understanding is essential to size the migration project optimally.

The objects, tables, and databases will determine the size of the source data warehouse that will be migrated.

Data Transfer

During the migration process, data needs to be transferred from the source data warehouse servers to AWS. This can be done over a network interconnection such as

AWS Direct Connect linking the source location with AWS or do an offline transfer using tools or services such as AWS Snow Family.

Rate of Data Change

The frequency with which your data is updated or changed will influence the intervals at which the updation must happen. This will ensure synchronization between the source and the target.

In cases where the rate of change is high, the service switching between the source and Amazon Redshift must happen within a shorter time window to ensure it is updated in that interval.





Data Transformation

Data mapping and schema change will help transform heterogeneous data during the migration.

How complex the process of data transformation is will influence the time needed for an iteration of migration.

Migration and ETL Tools

It is important to select the right tool for migration and extract, transform, and load (ETL) since they also have an impact on the migration process.

AWS provides a variety of tools and services that can help make the migration smooth and seamless.

Migration Strategies

The migration strategy also should be considered before embarking on the migration journey. Choose between one-step, two-step migration, or wave-based migration.

AWS Snowball belonging to the AWS Snow Family services can be used for delivering datasets to Amazon Simple Storage Service (Amazon S3) to be loaded into Amazon Redshift.

When continuous replication is not needed, one-step migration is a strategy you can adopt. This is ideal for databases that have comma separated value (CSV) files or columnar format such as Parquet.

To ensure data is consistent with the source, conduct a test in the destination Amazon Redshift database. Once all data has been validated, the database switch over to AWS can happen.



Where continuous operation is required, choose the two-step migration strategy. Since the source database is continuously undergoing changes, the target is synchronized using continuous replication.

The two steps in this migration strategy include:

Initial Data Migration

When the usage is low, extract data from the source database. Use the one-step migration strategy then to transfer it to Amazon Redshift.

Changed Data Migration

Before changing over, propagate the changed data after the initial migration to the target and synchronize the two databases.

Perform the necessary tests to validate the target database, and if it passes the tests, switch over.

For large-scale data warehouse migration, use wave-based migration. In this case, complex migration projects are broken down into manageable smaller logical and systematic waves to mitigate risks and reduce the complexity.

Begin with a workload containing a certain number of data sources and subject areas of medium complexity. During each subsequent wave, add more data sources and subject areas.

For a period of time, source and target warehouse will have to be run in parallel before retiring the source data warehouse.

The data warehouse migration to AWS had been a manual process that added to the complexity. Now, AWS Schema Conversion Tool (AWS SCT) is available for automating the migrations.



Some of the features of AWS SCT include:

- Macro conversion automation
- Case-sensitive database identifiers supported
- Gain competitive advantage
- Recursive common table expressions (with clauses)
- Automatic table optimization

Indium—An Enabling Migration Partner

Indium Software is an AWS partner with long experience in data migration and app modernization.

Our cross-functional team of domain and technology experts studies our client's existing data warehouse and devises a strategy to ensure minimum disruption and maximum performance during the migration to Redshift.

We use proven processes and technology accelerators to speed up acceleration and modernize legacy systems.

Leveraging existing tools and devising optimal strategies, we provide bespoke solutions best suited to your business needs, breaking barriers to innovation and ensuring agility.

To know more about how Indium can help you with migration to Amazon Redshift, contact us now.



USA

Cupertino | Princeton
Toll-free: +1-888-207-5969

INDIA

Chennai | Bengaluru | Mumbai | Hyderabad
Toll-free: 1800-123-1191

UK

London
Ph: +44 1420 300014

SINGAPORE

Singapore
Ph: +65 6812 7888

www.indiumsoftware.com



For Sales Inquiries
sales@indiumsoftware.com



For General Inquiries
info@indiumsoftware.com

