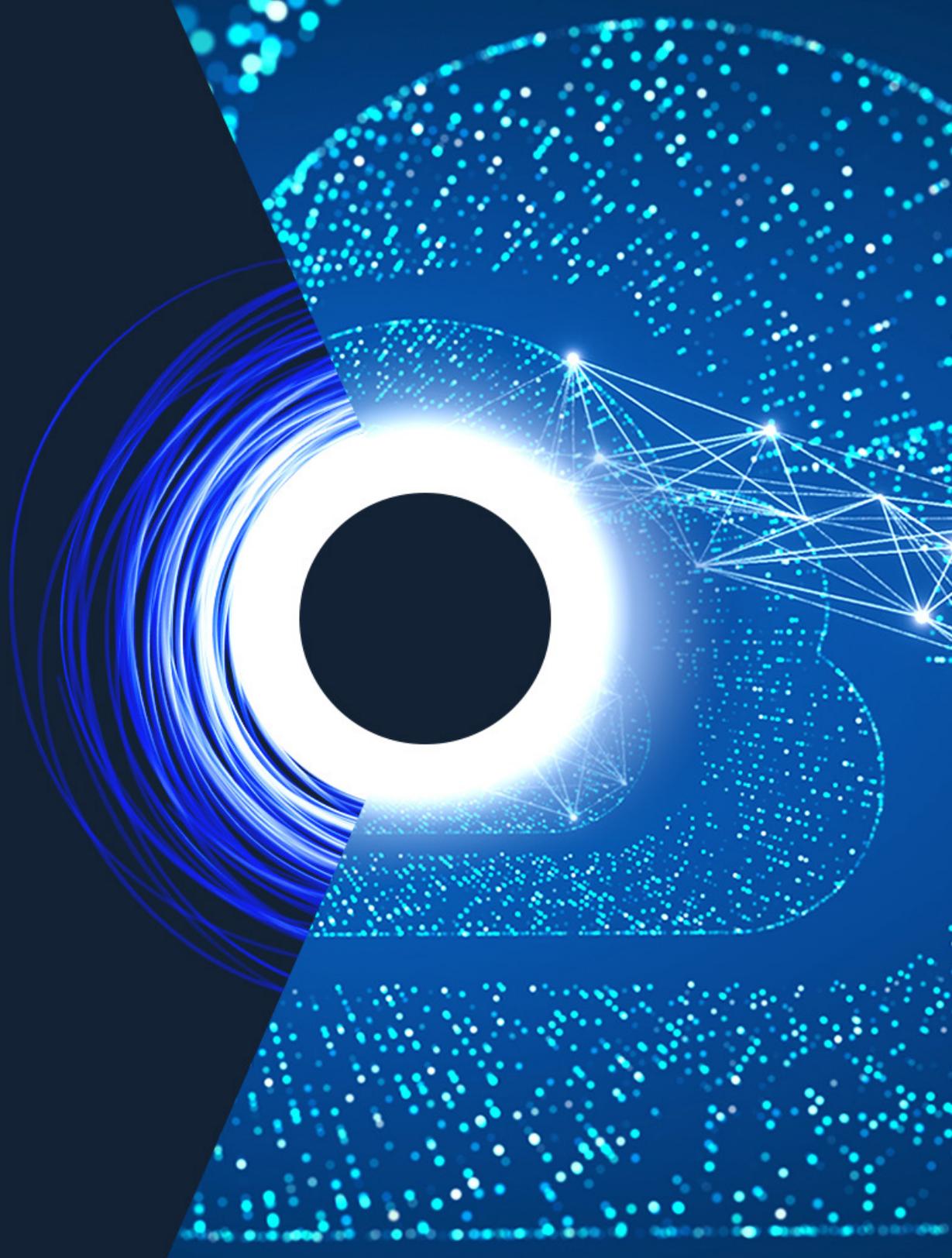


# 9 Reasons to Jump-Start Your Cloud Data Warehouse on AWS



---

## Contents

<b>What is Cloud Data Warehousing?</b>	<b>3</b>
<b>Preparing for the Data Warehouse Migration Process</b>	<b>7</b>
<b>Using an iPaaS to Migrate</b>	<b>9</b>

## What is Cloud Data Warehousing?

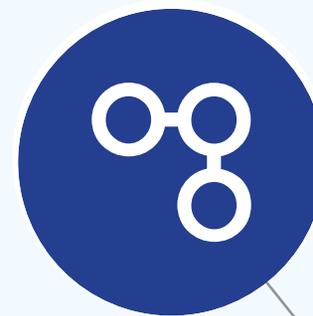
Cloud data warehousing is a Data Warehouse as a Service (DWaaS) approach that simplifies the time-consuming and costly activities that are typical of managing, administering, and tuning on-premises data warehouses. Cloud data warehousing does more than simply host warehouse databases in the cloud. Processing is also cloud-based, making data scalability and process scalability notable DWaaS benefits.

SnapLogic's Intelligent Integration Platform as a Service (IIP) empowers business and IT teams to quickly and easily move high volumes of data into and out of Amazon Web Services hosted cloud data warehouses including Amazon Redshift, Snowflake and Databricks Delta Lake, at any scale. With SnapLogic and AWS, data flows securely, without friction or impediment, across an entire organization, regardless of the source or application, bringing the best of the cloud to AWS customers.

In fact, SnapLogic has a strong foundation of moving data into Amazon Redshift. To date, SnapLogic has delivered over:

- 215B transactions using the Redshift Snap
- 135K distinct pipelines using Redshift Snap Packs
- 31TB of data in and out of Redshift per month
- 200 billion records a month to-and-from Redshift

But cloud data warehousing is only one piece of the puzzle. The rate of cloud adoption continues to accelerate, with more than 90% of the year-over-year revenue growth in the market in 2020 attributable to the cloud. At the same time, the core technologies that comprise data management solutions continue to proliferate and advance. That's why today's modern organizations require modern integration tools that align with the strengths of the cloud data warehouse. Together, the integration tool and cloud data warehouse offer a multitude of advantages. For example, [Kaplan Test Prep](#) was able to ingest data from 50+ applications into their data warehouse and reduced time to create data reports from 4 weeks to 3 days. The data team was able to produce an enterprise-standard technology connecting systems across the business while enriching data reporting to fuel innovation in products and services.



## If you're still wondering whether you should move to cloud data warehousing, here are 9 reasons to do so:



### Scalability

The volume of data in a warehouse typically grows at a steady pace as time passes and history is collected. Sudden upticks in data volume occur with events such as mergers and acquisitions, and when new subjects are added. The workload grows as more data is processed and as the number of concurrent users expands. Scalability adapts to growth, adding resources incrementally as data and workload increase.



### Elasticity

Processing workloads in a data warehouse fluctuates, sometimes with remarkable differences between peaks and valleys. Cloud elasticity enables the data warehouse to quickly expand and contract data and processing capacity as needed with no impact to infrastructure availability, stability, performance, and security.



### Managed Infrastructure

Eliminating the overhead of data center management and operations for the data warehouse frees up resources to focus where value is produced – using the data warehouse to deliver information and insight.



### Cost Savings

Operating an on-premises data center is expensive and includes the costs of staffing, servers and hardware, networking, floor space, power, and cooling. Data warehouse operating expense in each of these areas is eliminated or substantially reduced with cloud data warehousing.



### Speed of Processing

When timeliness of data matters, cloud elasticity meets the demand for fast processing. Data warehousing is process intensive: data acquisition, changed data detection, data transformation, data cleansing, data integration, data aggregation. The many processes involved in moving data from original sources to a data warehouse are complex and interdependent. A single bottleneck slows the entire pipeline and an unexpected spike in data volume amplifies the need for speed.



### Speed of Deployment

Data warehouse enhancements and modifications are seemingly endless. A continuous flow of projects adds new subject areas, builds new data marts, and adjusts to changing business needs and data sources. Projects frequently have to wait for infrastructure upgrades to expand data capacity, increase processing capacity, or support additional development and test environments. Organizations can eliminate project delays and accelerate deployment by leveraging the elasticity of the cloud.

**Magellan**  
HEALTH



With SnapLogic in place, we've automated and put into production more than 100 pipelines in the last few years. We also have about 22 SaaS solutions in our environment. Our job is to make sure the integration happens very seamlessly and SnapLogic allows us to do that better and faster. You can't ask for anything more."

— CIO at Magellan Health



## Fault Tolerance and Disaster Recovery

Today's data warehouses are business critical and many are mission critical. But many disaster recovery and business continuity plans address transactional systems and overlook the data warehouse – in part because the complexities of data warehousing make disaster recovery planning particularly difficult. Cloud data warehousing consolidates the entire data warehouse – data, processing, and state – as a single virtual entity. The entire virtual entity can be copied, stored offsite, and made live on another host in a matter of minutes.



## Security and Governance

Deploying the data warehouse in a virtual private cloud (VPC) environment inherently adds a level of security. VPC is an on-demand pool of shared computing resources within a public cloud environment, providing isolation between the different tenants sharing those resources. Some cloud data warehouse providers build security and governance capabilities into their architecture with features such as role-based access to data and built in data encryption.



## Relational Database Management Systems in the Cloud

Hadoop and NoSQL databases provide the means to deploy semi-structured, multi-structured, and unstructured data in the cloud. But what about relational data? Today, mainstream cloud data warehousing implements relational database management systems (RDBMS) in the cloud, putting relational data close to big data and close to cloud analytics. The benefits of data and process proximity now extend to the relational data warehouse. In the future it is likely that we'll see data warehouses built with non-relational databases. Migrating your existing data warehouse to the cloud is a step toward that future.

**Now that you've seen the benefits, how do you prepare to migrate a data warehouse to the cloud?**

# Preparing for the Data Warehouse Migration Process

Moving an existing data warehouse to the cloud is not quick, and it isn't easy. It is certainly not as simple as exporting data from one platform and loading to another. Tactically and technically, data warehouse migration is a process of many steps to migrate all of the components. When planning for cloud migration here are some key items to consider:

- **Migrating Schema** – Before moving warehouse data, table structures and specifications need to be migrated. Will these structural changes be made as part of the migration? Do indexing or partitioning need to be considered?
- **Migrating Data** – Moving very large volumes of data is process intensive, network intensive, and time consuming. How long will it take to migrate and what can be done to accelerate the process? Should restructuring be done as part of schema migration? Does data need to be transformed as part of the data migration? Can this transformation be done in stream or should it be pre-processed and then migrated?
- **Migrating ETL** – Moving data may be the easy part when compared to migrating ETL processes. Will changes to the code base need to be done to optimize platform performance? Are changes to data transformations required to sync with data restructuring? Should data flows remain intact or should they be reorganized? Will reduced data latency and delivery of near real-time data be part of migration? Would it make sense to migrate ETL processing to the cloud as well? If so, is there a utility to convert ETL code?

- **Rebuilding Data Pipelines** – With any substantive change to data flow or data transformation, rebuilding data pipelines may be a better choice than migrating existing ETL. Can individual data transformations be isolated and packaged as executable modules? Are the dependencies among data transformations to construct optimum workflows fully understood? What advantages can be gained – performance, agility, reusability, and maintainability – by rebuilding ETL as modular data pipelines using modern, cloud-friendly technology?
- **Migrating Metadata** – Source to target metadata is a crucial part of managing a data warehouse, knowing data lineage, and tracing and troubleshooting when problems occur. How readily will this metadata transfer to a new cloud platform? Are all of the mappings, transform logic, dataflow, and workflow locked in proprietary tools or buried in SQL code? Is export and import functionality required? Can metadata be reverse engineered? Or must it be rebuilt from scratch?
- **Migrating Users and Applications** – The final step in the process is migrating users and applications to the new cloud data warehouse with no interruption of business operations. What security and access authorizations need to be created or changed? Which BI and analytics tools should be connected? What communication is needed and with whom?

A typical enterprise data warehouse contains a large amount of data describing many business subject areas. Migrating an entire data warehouse in a single pass is usually not realistic. Incremental migration is the smart approach when “big bang” migration isn’t practical. Migrating incrementally becomes imperative when undertaking significant design changes as part of the effort.

**Once you’ve prepared for the migration, the next step is to select a migration tool.**

## Using an iPaaS to Migrate

There are many data integration tools available, however, an integration platform as a service (iPaaS) is one of the most efficient and scalable. SnapLogic is an example of data integration technology built for the cloud. The SnapLogic Intelligent Integration Platform provides a fast, multi-point, and modern approach to on-ramping your data to all major cloud data warehouses running on AWS such as Amazon Redshift, Snowflake, and Databricks Delta Lake. SnapLogic's intelligent platform as a service (iPaaS) can rapidly transform and load data into a broad range of cloud data warehouses keeping the data synchronized with source systems on a real-time and batch basis. SnapLogic Intelligent Integration Platform delivers a flexible connectivity layer that integrates any combination of data sources using pre-built, intelligent connectors called Snaps. These Snaps make migrating data to AWS fast and efficient. With support for multiple data warehouse destinations, SnapLogic delivers a faster time to value.

SnapLogic helps automate the movement of data into AWS, reducing time to value. Our no-code/low-code UI maximizes productivity and makes it easier for business and IT users alike, to move data into cloud data warehouses such as Amazon Redshift, and leverage that data across the organization for analytics, Machine Learning, and AI. SnapLogic moves data at enterprise scale.

Want to learn more about the SnapLogic iPaaS and how to get started with AWS cloud data warehousing? Visit [snaplogic.com/partners/amazon-web-services](https://snaplogic.com/partners/amazon-web-services). Want to get started with SnapLogic on AWS? Visit [SnapLogic in Marketplace](#).

*Some of the content in this ebook was originally published in our white paper, [Jump-Start Your Cloud Data Warehouse](#).*





## Get Your Demo Now

SnapLogic powers the automated enterprise. The company's self-service, AI-powered integration platform helps organizations connect applications and data sources, automate common workflows and business processes, and deliver exceptional experiences for customers, partners, and employees. Thousands of enterprises around the world rely on the SnapLogic platform to integrate, automate, and transform their business. Learn more at [snaplogic.com](https://snaplogic.com).