snowflake®

5 QUESTIONS TO ASK WHEN CONSIDERING A MIGRATION TO SNOWFLAKE

A Guide to Migration Evaluation and Readiness

CHAMPION GUIDES

EBOOK

TABLE OF CONTENTS

INTRODUCTION

There are many reasons why users of legacy databases might consider migrating from their current platform to a cloud-based alternative like Snowflake. As with any software, legacy databases have limitations. Many users cite pains such as slow, inefficient queries, scalability issues, and climbing licensing costs when operations grow. For users who have struggled with the complexity, maintenance, and collaboration difficulties on these systems, the Snowflake Data Cloud presents a compelling alternative.

Several key features set the Snowflake Data Cloud apart from other services:

- **A single platform:** Enable a full spectrum of use cases across organizations—all against the same copy of data and supporting a variety of data formats (structured, semi-structured, and unstructured), languages (SQL, Python, Java, and Scala), and architecture patterns—whether a data warehouse, data lake, a hybrid of the two, or data mesh.
- **It just works:** Snowflake's differentiated architecture and service delivery model ensures that the platform just works. Automations and optimizations keep customer data efficient, governed, and secure at scale. Snowflake also provides leading performance and concurrency out of the box;t users need not worry about resource contention or tuning. And with Snowflake's unique cross-cloud technology layer, Snowgrid, businesses' ecosystems are connected across regions and clouds for global scale, powering cross-cloud collaboration, data governance, and business continuity.
- **Price performant with right-sized resources:** Snowflake continuously improves price performance. Its proprietary, near-instant compute elasticity eliminates wasteful overprovisioning and costly delays in cluster spin up and resizing operations. Ongoing, transparent updates to its performance engine mean price for performance gets better over time. With Snowflake's real-time consumption-based pricing, users can ensure they're only charged for what they use and that any performance improvements immediately translate to lower costs.

Across thousands of clients, dozens of vendor partners, and more than a decade of experience, Snowflake has gained significant insight into the migration process—and what makes it most successful. We've come to understand why businesses seek a change and the most important questions to ask before they begin migrating from their current database to Snowflake. In this ebook, we outline five key questions enterprises should ask as they assess the need for migration and their readiness to make the move.

1. **Does my warehouse have fast analytical performance?**
2. **Am I experiencing problems with workload management?**
3. **Is my platform truly scalable?**
4. **Can I quickly enable data sharing?**
5. **Am I charged for more than I actually need?**

1: DOES MY WAREHOUSE HAVE FAST ANALYTICAL PERFORMANCE?

Fast analytical performance is crucial for effective data analysis, data-informed decision-making, and business competitiveness. It enables analysts to work efficiently, scale the volume of data they manage, and flag errors quickly.

Analytical technology has changed dramatically in the last decade, significantly accelerating time to business value. Cloud-based analytics have made it possible for organizations to acquire computing power for short periods, making projects easier to execute and allowing even small businesses to reap the benefits of a data warehouse. As options for databases proliferate, speed has become an important differentiator.

Legacy databases are often transactional systems by nature, and many engineers have found that they struggle to deliver analytical workloads at the required pace when working with large data volumes.

Online transaction protocol (OLTP) databases are often optimized to work with a specific set of rows. This makes it easy to write a query that retrieves the data needed, but querying across millions of rows is time-consuming since each row must be read from a disk and processed by the central processing unit (CPU) before the next row can be read. This inefficiency, coupled with developers who don't have enough experience with languages like SQL to write queries, can lead to major problems when trying to scale.

Alternatively, when data is uploaded with Snowflake, it is reorganized into an optimized, compressed columnar format. The Snowflake platform manages storage from organization and file size to compression and metadata. Since columnar databases use less memory to output data, more data can be stored, making queries faster. This structure is essential to process big data and is a requirement for accelerated data analytics.

2: AM I EXPERIENCING PROBLEMS WITH WORKLOAD MANAGEMENT?

When long-running and complex queries run on the same on-premises machine, compute and storage are tightly coupled without separation. In this structure, either compute suffers in favor of storage prioritization, or vice versa. This resource-sharing can create scalability and speed issues.

In practice, data engineers are locked into running workloads at specific times of day; they can't run them all at the same time due to limited compute resources. A simple scheduling mistake can damage—or even bring down—the server. That's a lot of pressure.

As data analytics have shifted from pure transaction processing in recent years, more platforms, including Snowflake, have separated compute and storage. The benefits are three-fold:

- **Scalability.** The legacy model of tightly-coupled compute and storage makes it difficult to scale the database. When database nodes are added, the data in the cluster must be "rebalanced," which is resource-intensive and often requires downtime. Conversely, when storage and compute are separated, they can be scaled independently for different use cases, usually in a matter of seconds.
- **Availability.** With all data stored remotely, there are no grave consequences if a compute node fails. Recovery is quick and painless since the remaining nodes in any given cluster can still access all data through the network.
- **Cost.** Separation of compute and storage means that businesses can pay for what they use as they use it. By some estimates, this could **save as much as 66%** in compute costs in industries like retail, which have peak demand periods such as Black Friday and that don't need to be provisioned year-round. This structure virtually eliminates the complexity and cost of data copies while providing a single source of truth.

This new paradigm in data platforms means you can trigger compute resources of any scale to query the shared storage since the queries will not affect other workloads that have their own dedicated compute resources. It's all about smart allocation of resources.

3: IS MY PLATFORM TRULY SCALABLE?

Every organization wants to scale with systems that can keep up with growth. While legacy databases are capable of scaling, doing so typically requires bringing down the server and spinning up a new one over the course of roughly an hour—which simultaneously puts business on hold. Disruptions equal timeouts on the client side, lost income, and frustrations.

Alternatively, scaling horizontally—for more concurrent queries—often requires costly hardware upgrades that can't be scaled down. Vertical scaling allows companies to increase the processing power of a single server or cluster, but they risk losing data in the event of a hardware or software error. Both methods carry risk and there are limitations to how much machines can be upgraded.

Snowflake's elastic performance engine was designed for seamless scaling. Typically accomplished in seconds, enterprises gain out-of-the-box concurrency without contention, tuning, or the need to manage the system. This way, data teams can do what they need to do without managing the systems themselves. Snowflake can support any number of users, jobs, or data with reliable multi-cluster resource isolation.

Resizing warehouses to improve query performance and reduce queuing can be done in Snowflake at nearly any time, even while running and processing statements. Warehouses can be started or stopped at any time, and new ones can be quickly created to support any activity. This flexibility maximizes economic value, allowing users to continually optimize price for performance.

Snowflake has key management, tuning, and vacuuming baked into the software, eliminating the stress of overseeing and maintaining the system and reducing the total cost of ownership. This near-zero maintenance is automated and is an online operation, so downtime rarely needs to be scheduled.

4: CAN I QUICKLY ENABLE DATA SHARING?

Data sharing ensures business teams have broader access to the most relevant and near-real time data necessary to accelerate business growth. In a legacy system, it is extremely common for databases to build additional logic into the data pipeline that allows users to copy data sets across environments.

Many legacy database users encounter friction when trying to share key data sets through traditional processes like file transfer protocol (FTP), the standard network protocol used for the transfer of computer files between a client and server on a computer network. It can be difficult to run FTP through any kind of firewall and many gateways struggle to support inbound connections, creating complications and delays.

Management of traditional FTP systems can be risky, as many cannot encrypt data. That lack of security means the sender loses some control of data sent via FTP, and has little say in who accesses the data and how long the files are usable. These systems often feature little by way of automation, costing teams many hours to physically oversee the movement. These complexities coupled with database inconsistencies, authenticity headaches, and the difficulty of sharing large volumes of data add up to frustrating, expensive, and time-consuming data exchange processes.

Snowflake puts data sharing on top of supported features so users can effortlessly pull in data sets from other Snowflake accounts and pick subsets of data to share with others, eliminating the need to build ETL jobs to accompany data sets across warehouses. Permissions are easily managed for each data set so they are accessible to the people who need them, when they need them.

Snowflake also boasts governed sharing of terabytes of data without the cost or time investment of physically moving it, adding to a suite of features built for functionality. With these tools at their disposal, data engineers don't have to build incoming and outgoing pipelines to share data.

Designed for modern, data-driven enterprises, Snowflake was built with the understanding that successful businesses must exchange data to innovate.

- Share data easily and securely across clouds, companies, teams, departments, functions, and business units. When both companies are in the same Snowflake Cloud Region, this data share occurs near instantly
- Easily set up security and governance, with permissions and roles built in
- Share data, views, and dashboards as needed for better collaboration and decision-making
- Improve planning and use insights to respond to customer needs, manage resources, and reduce time to market on new services
- Replicate data across different cloud vendors via Snowflake's cross-cloud capabilities, enabling seamless data-sharing across public clouds—no additional work required

5: AM I CHARGED FOR MORE THAN I ACTUALLY NEED?

In most legacy databases, users can stop or pause servers when not in use, but there is no sufficient solution to monitor and detect user demand for unique servers. As a result, many businesses run them around the clock and pay the associated price. If they pause a server, businesses are billed hourly even if the server only runs for a few minutes.

The technology stack needed to support a modern data warehouse is a major investment. Pricing for technologies varies greatly based on an organization's needs and goals, but cost tends to increase with time and growth. Paying for a set amount of storage, computing time, and computing power incurs significant yearly costs and requires a great deal of management that can consume an analytics team's time that would be better spent analyzing data and using it to drive competitive advantage.

Cost-effective and efficient resources allow enterprises to focus on their bottom lines without sacrificing the tools needed to innovate. That's why Snowflake's usage model is based on paying only for what businesses need, with a structure designed around consumption instead of peaks.

Snowflake Compute is billed per second, with a one-minute minimum, saving businesses significant cost. Snowflake's speed also supports savings, with workloads spun up instantaneously, making the per-second pricing extra meaningful. With this efficiency starting at an average $23 per **automatically compressed terabyte per month,** expensive premium storage is no longer required for effective data management.

TAKING THE NEXT STEP

As organizations create and consume data at unprecedented rates, effective data management isn't about adding complexity—it's about simplification. Snowflake was designed with that ethos at heart. The key value of simplicity extends to the migration process.

To streamline databases and make sharing, scaling, and analysis as seamless as possible, reach out to Snowflake Professional Services, who can equip you with tooling such as SnowConvert. The system was designed to quickly and easily migrate tables, views, procedures, macros, join indexes, and proprietary scripts from existing systems to Snowflake quickly and easily.

Ready for a change? Connect with Snowflake's Professional Services Team to learn more about how Snowflake can level up your operations.

Additional Resources

Migrate to the Data Cloud with ease. Learn more about Snowflake's Migration Accelerated initiative.

ABOUT SNOWFLAKE

Snowflake enables every organization to mobilize their data with Snowflake's Data Cloud. Customers use the Data Cloud to unite siloed data, discover and securely share data, power data applications, and execute diverse AI/ML and analytic workloads. Wherever data or users live, Snowflake delivers a single data experience that spans multiple clouds and geographies. Thousands of customers across many industries, including 590 of the 2022 Forbes Global 2000 (G2K) as of April 30, 2023, use Snowflake Data Cloud to power their businesses.

Learn more at **snowflake.com**