



Improving the ROI of Your Data Warehouse

Many organizations are struggling with a straightforward but challenging problem: their data warehouse can't affordably house all of their data and simultaneously support all of their analytics needs. Without an easy solution, some companies are storing significantly less data than they want to, and limiting the number of people or tools that can use their data. This vastly reduces the return on investment they see from their data warehouse, and sends a strong negative message to anyone with the curiosity and desire to ask a question or query the database.

In an effort to keep up with expanding datasets and users, companies with more budget flexibility attempt to keep up with demand by purchasing more and more database. This lowers the value organizations see from their data warehouse in a different way, by making each insight more expensive. What's more, the limitations of traditional database architectures ensure that each successive expansion in the data warehouse is progressively less effective.

To design a data warehouse that delivers on the promise of expanded data access and insight, traditional architectures have to be completely discarded, and a new, flexible, affordable and scalable architecture has to be implemented.

WHY IT'S HARD TO GET VALUE OUT OF MOST DATA WAREHOUSES

You have to buy enough compute capacity to match peak demand

Over time, most companies only use a tiny fraction of the compute capacity available within their data warehouse. Because they can't change compute resources on-demand, they have to ensure that they purchase more than enough capacity to match their usage peaks. But at every moment the warehouse isn't running at peak capacity, compute capacity is sitting idle and investment is being wasted.

To build a truly cost effective data warehouse, you need to be able to scale compute capacity to match demand, and effortlessly scale back down when usage peaks subside. The cloud can help solve this problem, but only if the underlying architecture of the warehouse supports it. For instance, most cloud data warehouses can only scale up by painstakingly stopping the data warehouse, adding compute and then restarting. This effectively negates the value of the ability to scale because it isn't practical to use in a time efficient way.

On the same token, "pricing by the query" data stores have also emerged attempting to resolve this issue. Although it might seem more attractive on the face of it, this addresses the capacity problem only by replacing it with completely unknown costs and unexpected service interruptions (when hard cost limits are hit).



Fig. 1: Fixed capacity data warehouses force organizations to buy more compute capacity than needed

Serving concurrent use cases is difficult and costly

Traditional data warehouses (including most built in the cloud) tightly couple storage and compute. Everyone that needs to use the data warehouse has to funnel through the same compute queue. Eventually, there are so many concurrent users and queries that the database becomes saturated, queries begin to fail, and people are unable to query the data they need when they need it.

Historically, there have been three methods for dealing with this problem: restricting access to the database, buying more compute capacity, or creating data marts with subsets of the data. All three options severely inhibit return on investment. Restricting access to the database makes it hard for people to get the data that they need when they need it. Buying more compute capacity can become cost prohibitive, and often marginally inefficient due to the **limitations of traditional Shared Disk and Shared Nothing architectures**. Creating data marts often requires more investment, and hampers the team with upkeep and management tasks.

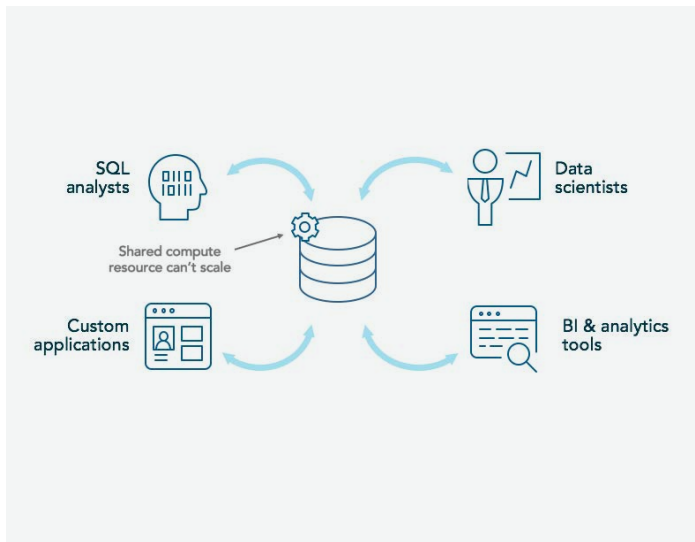


Fig. 2: Traditional data warehouses fail under concurrency, forcing expensive workarounds

Organizations need to store more data, and more diverse data, than ever before

The last problem is the simplest, and also one of the hardest to solve. More data is being stored than ever before, in traditional structured formats but also in increasingly popular semi-structured formats. Because traditional data warehouses don't have native support for semi-structured data, many companies are being forced to purchase noSQL tools like Hadoop to store and transform their JSON, Avro, Parquet and XML files. It goes without saying that this is an extremely expensive solution, since it requires additional fixed investment and massive amounts of time.



Fig. 3: Disparate data from a variety of sources necessitates expensive data silos

TAKING A DIFFERENT APPROACH

To empower the business with data -and do so at reasonable cost- the data warehouse must address each of these problems completely.

Matching compute capacity to demand

Instead of having to choose how much capacity they need to buy, organizations should be free to configure compute capacity as they need it, or have it automatically scale to their situation. If any individual use case requires more compute, it should be easy to provision that independently, without it affecting other activity on the database.

Regardless of the methods used to arrive at this state, every organization will need a platform that is easy to plan and budget for. If it's impossible to know what the costs from the database will be over a month, or even a single query, then it's impossible to understand the ROI of that tool.

Serving concurrent use cases affordably

To provide as much value as possible to the entire team, a cost effective data warehouse would be able to scale to as many concurrent use cases as needed, without additional fixed investment or time consuming overhead.

To do this, there needs to be a complete separation of storage and compute, with new and independent compute nodes able to spin up in moments to serve diverse needs. If there is any delay, or downtime, or lack of flexibility in delivering that, then the data warehouse will eventually fail to deliver the insight the organization needs at the time that it needs it.

Store as much data as you need to

To enable the highest return on investment, the data warehouse would enable anyone to store as much data as necessary, at a cost similar to or less than comparable raw storage solutions. Furthermore, it would natively support all data in any format to avoid the complexity and cost of disparate systems.

If there's any discussion about not storing data because of cost or type, then there's no way the data warehouse is going to provide long term value to the organization. Ideally, a modern data warehouse would take advantage of the cloud, and affordable cloud storage, to house massive amounts of diverse data without fear of the cost.

Complete separation of compute and storage is a key point here as well. Often, storage needs will expand or contract at a different rate than compute. If it's necessary to coordinate the two to address larger data volumes, pause the database, or shuffle the data in any way, then it will still be difficult to freely store large amounts of data.

BUILDING RETURN ON YOUR DATA WAREHOUSE INVESTMENT WITH SNOWFLAKE

Use virtual data warehouses to enable on-demand elasticity

With Snowflake, data is accessed through completely independent compute nodes that we call virtual warehouses. Each warehouse can scale up and down to match demand, even automatically. Crucially, any virtual warehouse can be automatically suspended or resized when peak usage has passed.

Instead of paying a fixed upfront cost, you simply pay for the amount of time that you use each warehouse. The smallest warehouse, an XS, uses 1 credit per hour (1 credit = \$1 or less), and each consecutive size larger doubles both the compute and cost per hour.

There's also support for multi-cluster virtual warehouses, which are unified compute resources comprised of multiple virtual warehouses. Multi-cluster warehouses can enable massive concurrent compute power for as long as you need it.

Virtual warehouses enable every organization to precisely match their computing power to their demand, and potentially save a significant amount over the fixed pricing of traditional data warehouses. Companies with brief but intense compute needs can see their spend lowered by 10 or even 100x compared to traditional solutions.

More importantly, analysts with larger queries and bigger questions have the freedom to query the database how they need to in order to find answers.

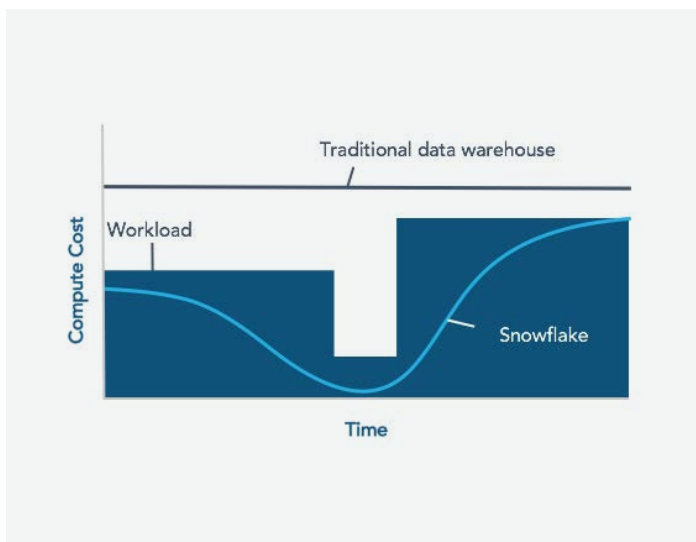


Fig. 4: Snowflake ensures that you only pay for the compute power you need

“The biggest change for us was that we didn’t take the additional \$5-10 million black eye required to double our existing on-premises all-in-one solution. When we did the proof of concept, Snowflake came in better on price compared to the industry leading solution, at comparable performance.”

– Michael Bigby, CTO, Research Now

Solve concurrency with as many virtual warehouses as you have use cases

With Snowflake, every virtual warehouse has access to the same shared database. This means that you can create as many virtual warehouses as you need to match the diverse requirements of any number of teams, and you’ll never need to create data marts or copy data. Since each virtual warehouse is priced by the hour, the same scalability that enables you to match compute capacity to supply can also enable you to serve concurrent use cases with their own independent peaks and troughs in demand.

The end result is significantly higher ROI. There’s no need to restrict data access, create data marts, or buy additional compute capacity. You can simply create the virtual warehouses you need to match demand, safe in the knowledge that you’ll only be charged for what you choose to use. Everyone with a reasonable need to access the database can be given the opportunity, unlocking the potential of your data and your people.

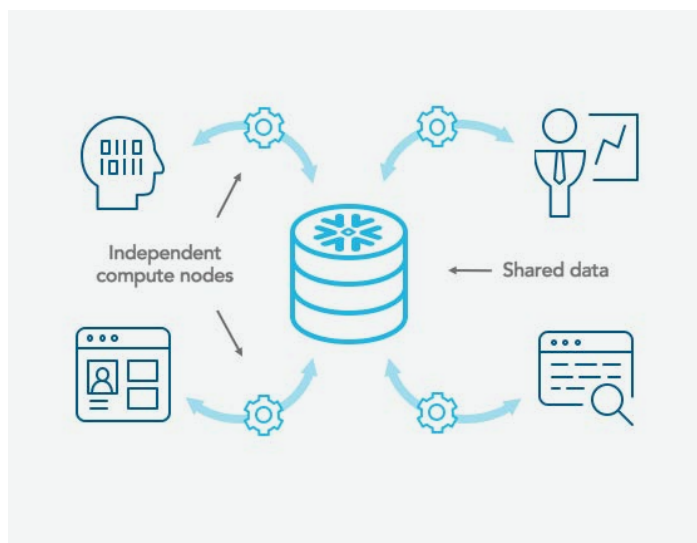


Fig. 5: Virtual warehouses help you to match the needs of many diverse people at low cost

“We had minimal configuration work to do with Snowflake; we did not have to worry about indexes or administration, because it’s a highly optimized SQL database already. Because the Snowflake data warehouse is truly elastic, we can increase and decrease compute power for different user needs that are temporary, with no changes to data or data locations”

– Rolfe Lindberg, Head of BI, DoubleDown Interactive

Store any kind of data at cloud storage prices

Snowflake was built to natively support both structured and semi-structured data, so there's no need to create an expensive, separate data store for your semi-structured data. Snowflake data storage is priced at a similar rate to Amazon S3, so you can affordably store massive amounts of data without needing to worry about cost.

Perhaps most importantly, because Snowflake seamlessly supports large and diverse data, you can focus on enabling analytics and insight, rather than figuring out where and how to store your data.

“We went from using 14 technologies to a stack of four powerhouses—AWS, Snowflake, Fivetran and MicroStrategy—that can be managed by a single person.”

— Joseph Bates, Head of Analytics, Sharethrough



Fig. 6: With Snowflake you can store structured and semi-structured data together, negating the need for expensive disparate silos.

CONCLUSION

Snowflake is the easiest solution for anyone looking to optimize the ROI of their data warehouse. By seamlessly matching your scale, concurrency and data storage needs at any moment, you can ensure that you only pay for the data warehouse you need, when you need it. By eliminating many of your cost constraints, you can encourage the types of questions and queries that will drive your team to bigger discoveries and accomplishments.

[Learn more about pricing.](#)

Snowflake is the only data warehouse built for the cloud. Snowflake delivers the performance, concurrency and simplicity needed to store and analyze all of an organization's data in one solution. Snowflake's technology combines the power of data warehousing, the flexibility of big data platforms and the elasticity of the cloud at a fraction of the cost of traditional solutions. Snowflake: Your data, no limits. Find out more at snowflake.net.

