

DATA-DRIVEN HEALTHCARE DONE RIGHT

How to reinvent your analytics with data warehousing built for the cloud





TABLE OF CONTENTS

- 2 Introduction
- **3** Move your data into a modern data warehouse built for the cloud
- 4 Best of all worlds: Unlimited, cost-effective scalability and performance
- **5** Up, down and out: How a cloud data warehouse should scale
- 6 User profiles: Solutions for common demands
- 7 The big question: What about security?
- 8 But wait, there's more: Additional security factors
- Security and compliance: Meeting stringent industry requirements
- **10** Tools to get your data into the cloud...
- **11** Large-scale physical data transfers
- **12** TCO: Budgeting and cost management
- **13** ROI: It's TCO and much, much more
- 14 Your 5-step journey to data-driven champion
- **15** Find out more

INTRODUCTION

Being a data management professional in the healthcare industry is never easy. Business and clinical users from all levels and departments want useful insights from your organization's data. If that's not enough, they want that same level of insight from data that has yet to arrive, or just won't fit, into your traditional healthcare data warehouse. And they expect immediate response times to both simple and complex queries.

The healthcare sector has unique data challenges. Efficient healthcare delivery requires secure and controlled data sharing among hospitals, payers, service providers, health insurance entities, and the government, especially in the move toward value-based services.

In addition, healthcare organizations must minimize costs associated with storing, managing, and analyzing data while having access to comprehensive analytics.

Bombarded with demands, you must juggle legacy technology located on premises, in the cloud, or both. Too much of your time is spent fixing operational snafus that shouldn't happen. You need the ability to link databases between clients and their partners through secure connections that allow query into both datasets in real time, with no delay for data to move from one repository to another. You also need the potential to store data in a manner that takes full advantage of core architecture to be able to query across all genomic, claims, and EHR datasets. This data sharing capability can provide valuable insights.

What if your organization could easily and affordably give your business intelligence staff, doctors, administrators, executives, partners, and patients the data they want, when they want it? And what if you could shift from managing day-to-day issues to championing strategic technology and data initiatives?

MOVE YOUR DATA INTO A MODERN DATA WAREHOUSE BUILT FOR THE CLOUD

Free your data from legacy constraints with a modern cloud data architecture that delivers fundamental improvements in performance, concurrency, and simplicity, while maintaining compliance and security. Deliver the insight from all your data with query times in seconds or minutes, not hours or days. What's the result? You'll have the time you need to work on strategic initiatives. In addition to providing many organizational and IT advantages, a cloud data warehouse that takes advantage of modern cloud architecture can increase performance by 200x for a tenth of the cost of legacy data warehouse systems located on premises or migrated to the cloud. Here's what you need to know to move to a modern cloud data warehouse.



THE FAST PATH TO CHANGE

This guide contains the information you'll need to understand and succeed with modern cloud data warehousing. It will start you on a path to transform your company's data analytics with a cheat sheet on five key topics crucial to getting started with cloud data warehousing:



BEST OF ALL WORLDS: UNLIMITED, COST-EFFECTIVE SCALABILITY AND PERFORMANCE

Analyze vast amounts of varying data with speed

Today, your business users want it all.

- Easy access to very large and disparate data sets
- Integration of flexible data types
- The ability to access all levels of data to get answers fast, no matter how complex the query
- The ability to share data across departments and with partners

User demands push legacy data warehouses beyond their limits. This means you're constantly juggling processes so the system doesn't crash. But too often, contention for limited resources forces you to reschedule jobs or kill a user's query entirely. When that happens, you're no one's champion.

WELCOME TO LIMITLESS RESOURCE FLEXIBILITY

With the arrival of cloud architecture, you can scale storage and compute resources powerfully and cost-effectively. Cloud data warehousing can give you unlimited resources and the elasticity to access any scale of compute horsepower, paying only for what you need by the month, week, day, or hour. You can avoid the legacy problem of overprovisioning for peak demand and getting stuck with an underutilized system the rest of the time. In addition, cloud storage can cost a fraction of what you paid for storage devices that currently live in your data center. But note that only a modern data warehouse built for the cloud, one that truly separates compute from storage, can effectively capitalize on everything cloud architecture offers.

Here are some of the many issues a modern cloud data warehouse can alleviate:

- Competition between users' queries and data integration activities that degrade performance.
- Users forced to only analyze data subsets or aggregates to avoid further strain on the system.
- The complexity of data movement, refinement and transformation in legacy environments.

TRADITIONAL VS. CLOUD-BASED ARCHITECTURE

TRADITIONAL ARCHITECTURES



SHARED DISK SHARED STORAGE SINGLE CLUSTER



SHARED NOTHING DECENTRALIZED LOCAL STORAGE SINGLE CLUSTER

BUILT-FOR-THE-CLOUD ARCHITECTURE



MULTI-CLUSTER, SHARED DATA CENTRALIZED SCALE-OUT STORAGE MULTIPLE, INDEPENDENT COMPUTE CLUSTER

UP, DOWN AND OUT: HOW A CLOUD DATA WAREHOUSE SHOULD SCALE

Multiple ways to achieve peak performance

Cloud data warehousing can give you unlimited compute resources dynamically and without lag time. There are multiple ways to easily scale up, down and out (concurrency) to meet demand and only pay only for what you use:

• START A NEW COMPUTE CLUSTER:

Let's say you have a workload that normally accesses the same amount of data all the time. You've chosen the right size cluster and it runs those queries in a timely manner. But what happens when users expand their queries from last month's data to include five years of data? You'd be better served with a new and bigger cluster.

However, adding more compute nodes to the existing cluster may not be the best solution. The good news is, with the right kind of cloud data warehouse, you can quickly define and spin up a new cluster in just a few minutes, flat.

• KEEP A PRE-DEFINED COMPUTE CLUSTER IN SUSPENDED MODE, READY TO GO:

If you have a regular event that requires a burst of compute resources, legacy architectures pressure you to build or add a new cluster from scratch every time you need it. This consumes time and money. But the right cloud data warehouse can provide a predefined compute resource in suspended mode to switch on whenever you want. When users are done with the resource, you can put it in "sleep" mode until it's needed again. Better still, you can configure the resource to go to sleep mode automatically after a predetermined time of no activity. After the cluster turns off, you don't pay for an unused resource.

DYNAMICALLY START AN ADDITIONAL CLUSTER TO HANDLE CONCURRENT USERS:

The number of concurrent users and queries on a cluster can surge, creating a queue. A true cloud data warehouse can automatically scale concurrency by transparently creating a new cluster that load balances with the first. The mirrored cluster has access to the same data as the original. When the load subsides and the queries catch up, the second cluster will automatically spin down. Ideally, you can determine this concurrencybuilding capability and cost by specifying the maximum number of clusters that can be provisioned automatically.

MANUALLY RESIZE AN EXISTING COMPUTE CLUSTER:

If you want to keep tight control over compute resources and costs, manual resizing provides an alternative to auto-scaling. Let's say you have an existing cluster spun up with four nodes. But you know there's a surge in data coming and you want to give that cluster more compute horsepower, but for a specific period of time. You can resize that existing cluster, specifying eight, 16 or 32 nodes, while it's still running. When the surge subsides, scale the cluster back to its original four-node configuration.

These are just some of the possibilities available with a data warehouse that truly takes advantage of cloud architecture. Its exponentially better performance, compared to on-premises and "cloudified" data warehouses, is ready to meet insatiable user demands for data volumes, velocity and variety unimagined just 10 years ago.

USER PROFILES: SOLUTIONS FOR COMMON DEMANDS

Warehousing strategies to get you out of reactive mode



The CFO who needs an answer right now.

Start a new, ad hoc compute cluster fast, so a finance analyst can run queries within minutes without impacting other users. Your results come from the company's single source of data, so you're confident it's correct.



The compliance auditing and reporting team that makes a big push at the end of every quarter.

Keep a predefined cluster for user, patient, and compliance data in suspended mode, ready to go. Activate this predefined resource at a moment's notice. When the end-ofquarter rush is over, easily put it back into suspended mode.



The curious valuebased care executive looking for the most efficient and cohesive patient experience.

When users start adding multiple dimensions to their queries and analysis, complexity escalates quickly. Having a separate cluster for the value-based care (VBC) executive ensures that a complex ad hoc query doesn't impact other users or workloads.



The data scientist who wants to stress test a theory.

For example, if the data scientist expands the analysis to cover three years instead of one, you can manually resize an existing cluster to deliver the necessary compute power. When the data scientist is satisfied with the analysis, you can scale the cluster back to its original size to keep costs low until the next big test.



The patient services team that wants to find out who has the best patient response scores.

If a manager typically looks at a particular data set, but a large group of team members suddenly hits the same data set with additional queries, the system needs to respond quickly. Automatic concurrency scaling can start a second cluster, accessing the same data, and then load-balance the queries with the first cluster. When the flurry of queries subsides, the second cluster automatically shuts down.



The purchasing and supplies analyst scrutinizing inventory turnover trends.

This analyst often works with a rolling six months of data but now is running queries on data going back two years. When you get complaints that queries are too slow, you can start a new properly sized cluster in just a few minutes, or you can simply resize the existing cluster to meet the increased demand.

THE BIG QUESTION: WHAT ABOUT SECURITY?

The cloud can be more secure than on-premises solutions

For years, organizations have considered the cloud to be more vulnerable than on-premises solutions, even though hackers continue to breach major corporate data centers. Certainly, security remains the top concern for organizations migrating sensitive data to the cloud. In addition, healthcare organizations must follow and report on strict compliance guidelines such as the Health Insurance Protection and Accountability Act (HIPAA).

Getting cloud security right is critical. If you carefully evaluate and choose cloud offerings that prioritize security and compliance, you can benefit from some of the best security defenses and compliance tools in the industry—better than what most organizations have in place to protect their legacy systems.

Moving to a secure cloud data warehouse means you don't have to devote resources to security. Instead you can redirect those resources to focus on other strategic efforts.

Here are a few of the top measures a cloud data warehouse should offer to provide state-of-the-art security and compliance:

ENCRYPTING DATA IN TRANSIT AND AT REST

If an unauthorized user gains access to your data, they must not be able to read it. The modern cloud data warehouse should protect data in transit and at rest, whenever it is sent over a network or stored on disk. This includes data files persistently stored, query results, and the content of a local disk cache.

In addition, an advanced cloud data warehouse solution should use the latest industry-standard encryption algorithms. The Advanced Encryption Standard, AES, with 128-bit keys, is the minimum best practice for symmetric encryption. For enhanced security, the most robust cloud data warehouses use AES-256.

Unlike legacy security architectures, encryption shouldn't impact query or load performance of a cloud data warehouse. However, it's rare for enterprise data centers to be encrypted to this high degree because of the cost, time and scarcity of expertise.

SECURITY FOR COMPLIANCE REQUIREMENTS

Organizations with compliance needs must choose a multi-tenant service that implements isolation at multiple levels. Ideally they should run their data warehouses inside a virtual private cloud (VPC), a logically isolated network section within a cloud service such as the Amazon Web Services (AWS) cloud. The VPC enables isolation and limits access to its internal components. For customers who have HIPAA, Payment Card Industry (PCI), or other compliance requirements, customers should implement additional security features to meet specific compliance requirements. Such a solution should include HIPAA support, be PCI compliant, and feature an enhanced security policy.

SECURITY BY DESIGN





AUTHENTICATION

ACCESS CONTROL



DATA ENCRYPTION



EXTERNAL VALIDATION

BUT WAIT, THERE'S MORE: ADDITIONAL SECURITY FACTORS

Exploring encryption key management and multi-factor authentication

The encryption cipher is only one crucial security factor. Also consider key management and multifactor authentication.

KEY MANAGEMENT

Key management governs the lifecycle of the encryption keys, which includes the generation, storage, distribution, use, and disposal of the keys. Ideally, you would use a key hierarchy so that the root keys encrypt secondary keys—about one per data partition. These, in turn, encrypt even more granular keys such as one per table.

For any data warehouse, you should limit the amount of data covered by an individual encryption key and limit the time the key is used. This is an industry best practice delivered through key rotation and data rekeying.

- Key rotation is a method to periodically generate a new encryption key to protect newly inserted data.
- Rekeying is the ability to go back to previously stored data, re-encrypt it with freshly generated encryption keys, and dispose the old encryption keys.

Both mechanisms are necessary to manage the complete lifecycle of encryption keys to meet the highest industry standards. A cloud data warehouse that provides key management functionality nearly eliminates the responsibility of customers who are currently implementing encryption configuration and management with a legacy system. Even if key management wasn't available with your legacy data warehouse, consider it mandatory to operating securely in the cloud.

MULTIFACTOR AUTHENTICATION (MFA)

As a best practice, anyone accessing data in a cloud data warehouse should do so using multi-factor authentication (MFA). After logging in with a username and password, the user will need a second authentication mechanism. This can be a random code generated by an app on a user's smartphone. Together, these factors prove that users are who they say they are – a strong measure to ensure only authorized parties gain access to data in the cloud.

SECURITY AND COMPLIANCE: MEETING STRINGENT INDUSTRY REQUIREMENTS

The importance of third-party verification

Industry-specific standards provide an additional layer of assurance for data security concerns. A cloud data warehouse provider should comply with the following standards:

- 1. System and Organization Controls 2 (SOC 2): The American Institute of CPAs has developed the SOC 2 report, which evaluates an organization's information systems for security, availability, processing integrity, confidentiality, and privacy.
- 2. HIPAA: Protected health information (PHI) is subject to the privacy and security rules under HIPAA. Cloud service providers storing PHI must adhere to HIPAA regulations for:
 - Security & privacy
 - User access tools
 - Encryption
 - Data location
 - Return of data
 - Contingency planning & disaster recovery
 - Service Level Agreements (SLAs)

For most cloud infrastructure providers, such as Amazon Web Services (AWS), properly managing PHI data requires additional security controls detailed in AWS's business associate agreement (BAA). These controls often extend critical protections, such as encrypting data at rest, to exceed HIPAA specifications because cloud service providers want to reduce their risk. Healthcare organizations that store their data in modern warehouses built for the cloud reap the benefits of these additional protections.

- 3. PCI: PCI compliance requires adherence to a set of specific security standards developed to protect credit card information during and after a financial transaction. All card brands require PCI compliance. For card information stored in a cloud warehouse, the vendor must:
 - Build and maintain a secure network
 - Protect cardholder data
 - Maintain a vulnerability management program
 - Implement strong access control measures
 - Regularly monitor & test networks
 - Maintain an information security policy

THE IMPORTANCE OF THIRD-PARTY VERIFICATION

Some regulations allow providers to "self-attest" as proof of compliance. To ensure your data has the highest security, expect a cloud data warehouse vendor to use an independent provider to conduct penetration tests. A "pen test" is an attempt to evaluate IT infrastructure security by safely trying to exploit vulnerabilities that can exist in operating systems, service and application flaws, improper configurations, or risky end-user behavior. This type of security validation provides further assurance the security of a cloud data warehouse vendor meets industry standards and your organization's expectations

TOOLS TO GET YOUR DATA INTO THE CLOUD...

... from wherever it resides now

Getting data into any data warehouse has historically been a slow and tedious process. That means business analysts may not always have the most current data. In addition, your organization, like many others, may be moving major applications to the cloud. Now you have to contend with accessing that data in the cloud, too. Fortunately, many tools are available to help you migrate existing, on-premises application data into a cloud data warehouse. Popular choices include graphical pipeline tools that allow you to extract, transform, and load (ETL) data from both cloud services and traditional on-premises database sources into your modern cloud warehouse.

THE BIG MOVE: BULK-LOADING FILES

A bulk-load approach can work best for your initial transfer if you have many terabytes of data that live on storage devices within your organization. The data is loaded into thousands of manageable files and bulk loaded in parallel into the cloud data warehouse. Modern cloud data warehouses should provide easily-deployed, supplemental resources to speed up loading.

INCREMENTAL UPDATES

After you've transferred your data, you should capture incremental changes. Expect to use one of the many modern ETL or ELT (extract, load, transform) tools for data transfer, available in commercial software and open source options. Additional options include the newer change data capture (CDC) and advanced data replication tools. Depending on the nature and source of the data, you might also want to use newer cloud-based data streaming tools and platforms. These tools are often well suited for incremental migrations rather than bulk loads.

LARGE-SCALE PHYSICAL DATA TRANSFERS

When you have a vast amount of data

If you have hundreds of terabytes or even petabytes of data, it's sensible to use physical appliances from cloud providers to migrate your data. You can load large amounts of encrypted data from your on-premises data center onto the appliance and ship it to the cloud services provider, which then uploads the encrypted data to a staging area in the cloud.

This method allows the transfer of about a petabyte of data per week to the cloud. You can then easily move the data over to your cloud data warehouse provider.

Data transfer appliances are often faster and more cost effective than trying to push all that data to the cloud via the Internet. They're a logical choice if all of your data would take longer than a week to upload to a cloud data warehouse.

THE EXABYTE OPTION

Companies today generate so much data it's often measured in exabytes. (A single exabyte is equal to one billion gigabytes.) Some cloud service providers now offer a "drive-up" option for this level of data transfer. Literally speaking, it's a device housed in a large shipping-like container attached to a semitruck. With ten of these trucks, you can have an exabyte of data moved to the cloud in six months.



TCO: BUDGETING AND COST MANAGEMENT

Planning a move to the cloud business model

Moving your data warehouse to the cloud will save you the significant expense of buying, maintaining and securing an on-premises system. But how do you budget for an entirely different cost model?

SAAS IS NOT JUST A TECHNOLOGY SHIFT

Switching from on-premises technology to a cloud solution also changes how an organization budgets and records costs. An on-premises solution is often a large up-front capital expense that's treated as an asset and depreciated over time. A cloud solution is a much smaller up-front transaction that's treated as an operating expense deducted every month from an organization's corresponding revenues. Each solution has different expenses for ongoing management and maintenance.

When moving to a new technology solution, IT managers must show that the new technology meets an organization's needs and budget. At the heart of any solution purchase is the total cost of ownership (TCO)—a financial estimate that details the costs of a system over its expected lifetime.

TCO: CLOUD VS. ON-PREMISES

With a TCO estimate, the first-year expense for a cloud-based solution should be a fraction of the cost of buying for on premises. But when estimating for five, 10, or more years of ownership, organizations often debate whether the cost of the cloud solution will eventually exceed an on-premises alternative.

With larger or organization-sized solutions, the initial and ongoing expenses of an on-premises solution are likely to cost more than the aggregate usage charges of a cloud-based solution.

After incurring the huge upfront costs of an onpremises purchase, an organization must factor in the ongoing maintenance, repair, and eventual replacement of its IT infrastructure. This includes building and maintaining one or more data centers, which can cost many millions of dollars plus any financing costs. In addition, annual maintenance and support fees for on-premises systems can exceed 20 percent of the initial purchase price of the appliance and data warehouse software. Finally, the salaries of the highly specialized people who support a data center and all related functions can easily exceed 50 percent of the TCO.

In short, a simple but exhaustive TCO analysis for a cloud-based data warehouse is likely to reveal a substantial savings.



ROI: IT'S TCO AND MUCH, MUCH MORE

Evaluating the ROI for your cloud data warehouse

TCO and ROI are inextricably linked. A TCO estimate reveals how much a cloud data warehouse will cost, and how much it will save an organization compared to an on-premises alternative. But those savings represent only one aspect of ROI. Consider many other factors when thinking about the ROI differences between on premises and cloud.

Your ROI estimate should additionally include:

- Deployment times: A cloud-based data warehouse can go live in weeks or just a few months, depending on the size of the project and the migration strategy to the cloud. With a true SaaS data warehouse, an organization can see benefits much sooner with a cloud solution. Note that solutions based on Infrastructure-as-a-Service (IaaS) or Platform-as-a-Service (PaaS) approaches can take much longer. Time saved directly translates into money saved.
- Software upgrades: On-premises solutions and solutions that create on-premises clouds take a standard "waterfall" development approach to functionality updates. To enable periodic updates,

IT usually must take the system down or place it in maintenance mode, which results in lost time and money. To avoid downtime, IT may keep using a specific version of the software, which prevents the organization from using the latest technologies and security features.

With a modern cloud data warehouse, monthly updates prevent any disruptions.

- The people factor: As previously described, the number of people required to maintain an on-premises data warehouse and supporting infrastructure can be an enormous expense. Cloud alternatives enable an organization to significantly reduce or nearly eliminate this expense depending on the level of functionality, automation, and vendor management of the solution.
- Pay only for what you use: An on- premises data warehouse forces you to buy enough storage space and compute horsepower to handle demand on your busiest day of the year. What about the other 364 days? With the right cloud data warehouse, there's tremendous opportunity to pay only for

what you use, when you use it. Additionally, the cost of the actual storage and compute resources should be significantly lower with a cloud solution thanks to cloud economies of scale.

These factors and more should be part of your ROI calculations when comparing your on-premises option to a cloud data warehouse. If you consider more than one cloud data warehouse vendor for your next purchase, don't forget to evaluate the differences among those alternatives, since no two data warehouses are the same, whether on-premises or in the cloud.

YOUR 5-STEP JOURNEY TO DATA-DRIVEN CHAMPION

A cloud data warehouse cheat sheet



CHAMPION GUIDES

FIND OUT MORE

Become a champion today with modern cloud data warehousing

Data warehousing has been re-thought and reborn in the cloud for the modern, data-driven organization. Find out how you challenge the status quo and become an IT champion, giving users the benefits they dream of with data warehousing built for the cloud.

Stop juggling problems. Take the first step of taking control of your data analytics operation.

Visit snowflake.com

Want more informative content to help transform your organization into a data driven enterprise?

Visit Snowflake's Resource Library www.snowflake.comt/resources

20



ABOUT SNOWFLAKE

The Snowflake Cloud Data Platform shatters the barriers that prevent organizations from unleashing the true value from their data. Thousands of customers deploy Snowflake to advance their businesses beyond what was once possible by deriving all the insights from all their data by all their business users. Snowflake equips organizations with a single, integrated platform that offers the only data warehouse built for any cloud; instant, secure, and governed access to their entire network of data; and a core architecture to enable many other types of data workloads, including a single platform for developing modern data applications. Snowflake: Data without limits. Find out more at **snowflake.com**.



© 2019 Snowflake. All rights reserved.