eBook

How to Easily Migrate and Protect Sensitive Data in the Cloud

Introduction

Consolidating your business data in a cloud data warehouse or platform like Snowflake Data Cloud is a smart move that unlocks innovation and value. All your data in one place makes it easier to connect the dots in ways that were impossible or unimaginable before. For instance, a retail chain can optimize sales projections by analyzing weather patterns, or a logistics company can more accurately predict costs by accounting for the salaries of all the people involved in a shipment.

Getting those insights is a process that starts with moving the data and ensuring that it is an ideal state for analysis. An extract, transform, and load (ETL) technology partner simplifies moving or loading the data from each of your company's locations into a cloud data warehouse and transforming the data to make it analytics-ready in no time. Moving data is what these companies do best. Data governance and sensitive data security are not their priority, however, which is a tremendous concern when the most valuable data is often the most sensitive–both to the business and to bad actors. Confidential information like customer PII, which includes email, home addresses, or social security numbers, if breached, would create significant risk of legal exposure and to your reputation.

The need for high levels of data protection and secure access can cause significant tradeoffs in data usability and sharing, which adds risk and complicates matters for analytics teams. Even the built-in security and governance capabilities of data warehouses require a level of database coding expertise that is costly to implement and time-consuming to manage at scale. Distributed enterprises need a thoughtful yet simpler approach to protecting data in the cloud that keeps information airtight and doesn't slow down access and progress.

Use this guide to plan data migration from highly secure on-premises systems to the cloud that allows you to continually control and protect sensitive data.

Security Considerations:

- **Security considerations in planning your data move**
- **Selecting a data migration partner**
- **Selecting a data security partner**
- **Ready to migrate sensitive data to the cloud?**

Security Considerations in Planning Your Data Move

Before we migrate data to the cloud, let's understand why you need security and why some solutions fit better than others for your specific business needs. We all know that we must protect sensitive data in order to comply with appropriate regulations and maintain the trust of our customers. What is not always clear is that the same standards for storing and protecting data on-premises also apply to data in the cloud.

These requirements include using NIST-approved security or standards for at-rest data protection. At a minimum, we must ensure there's not a single door for hackers to get through, known as a single-party threat. If data can be de-obfuscated by just by one person, the protection method doesn't count. For example, simply reversing a medical record is not enough. Encryption meets this requirement because you need both the encrypted data and the key to unlock it in order to access the original data.

For data in the cloud, however, you need to rethink tooling and management decisions. Let's look at encryption again, but through a cloud lens. You'll quickly run into several issues:

- You can't expect to connect your on-prem key management solution to a cloud data warehouse like Snowflake and have it work at scale.
- Someone who gets the key can decrypt all the data stored in your centralized data warehouse.
- You also need fail-safes to prevent users with privileged access, like the Snowflake administrator, from being able to view the data if they're not supposed to.

To avoid these access and encryption issues, some security methods rely on transforming data through "one-way" techniques like hashing before storing the hash in the cloud. Hash codes ensure privacy and allow users to still know what the dataset comprises, for example, social security numbers. However, an authorized user who needs the real social security number won't be able to retrieve it, because once hashed, the data cannot be recovered in the cloud database.

Even anonymization techniques, such as storing the data as a range, limit the application of data. You might not need an individual anonymized data point today, but you may very well need it later. Your business may depend on allowing some authorized users to have access to the original data, while ensuring it is meaningless and opaque to everyone else.

If you must use sensitive data after you store it in the cloud–which is the endgame for analytics—then the preferred security solution is tokenization for its ability to balance data protection and sharing, which we explore in more detail in "Chapter 4."

Selecting a Data Migration Partner

WHEN CHOOSING AN ETL PROVIDER CONSIDER THE FOLLOWING PRIORITIES:

1. **Security-focused:** When daily data breaches make global news headlines, security should be everyone's responsibility. Look for pro-active, security-forward migration vendors. How can you tell? The ETL partner's security investments will reveal their focus on the NIST framework and certifications in key areas like SOC2 Type II, HIPAA, Consensus Assessments Initiative Questionnaire (CAIQ), and GDPR and CCPA privacy regulations, among others. Migration is the first step toward unlocking value from data in the cloud, and security must be airtight at the start of the journey.
2. **Universal:** Data today is everywhere, so your ETL solution must have universal connections and integrations. Look for cloud-native data migration platforms that offer multiple connectors across APIs, CRMs, databases, social media, and other services for your on-premises databases and cloud-based systems. To simplify data connectivity, they should offer a choice of pre-built connectors and custom connectors that you can create easily without coding required.
3. **Scalable and Flexible:** Your investment in data migration should allow you to be as elastic as the cloud warehouse itself. At the outset of your consolidation project, you may not know how much data you truly need to extract from all the locations in your distributed enterprise and move to the cloud. So make sure to choose a solution whose capabilities can scale up or down with your needs and vendor who won't lock you into commitments that are not commensurate with your needs. Solutions that leverage the power of the cloud to scale as your needs grow and consumption-based pricing that allows you to pay-as-you-go are ideal. Look vendors that will let you try it out, with technical support, and offer cloud-native solutions that can scale with your needs.

Pro Tip

If you're using an SI, GSI or consultant like Accenture or Slalom to help execute your data migration strategy...

Point number one applies there as well. Whenever data flows outside your company, it's critical to have visibility into who's accessing it and controls over who can access it. Look for a partner that has data governance and security controls in place before you start your project.

Selecting a Data Security Partner

You've decided on a cloud data platform and a migration partner. Now you need to make sure the data being loaded into the warehouse, including all the sensitive stuff, gets Fort Knox-level security that doesn't slow you down and allows you to start analysis and operations almost immediately. The strategy for choosing data security is three-fold:

1. **Low-latency:** Nobody enjoys watching paint dry or spinning hourglasses on their screen. To reduce the impact to your team, prioritize protection solutions that are low-latency and performance-aware. Some SaaS-based security solutions with an architecture based on AWS can be flexible and move resources to where your workloads are running to increase speed. For instance, if you're running out of the western U.S., a nearby point of presence (PoP) optimizes performance based on your location. While most of the workloads running in cloud data warehouses don't need the millisecond speeds required in financial services, it's always good to plan for potential future requirements.

2. **Functional:** The name of the game is analytics, so you'll want security solutions that make your data secure but still leave it available for use. As we mentioned in Chapter 2, tokenization is the best option for cloud data workloads. Tokenization has all the obfuscation benefits of encryption, hashing, and anonymization, while providing a much greater usability. Let's look at the advantages in more detail.

- Tokenization replaces plain-text data with a completely unrelated token that has no value if breached. There's no mathematical formula or key; the real data remains secure in a token vault.
- You can perform high-level analysis just like you could on real data, without having access to the real thing. In contrast, you have limited analytics capability on anonymized data and none on hashed and encrypted data. With the right tokenization solution, you can feed tokenized data directly from the warehouse into any application, without requiring data to be unencrypted and inadvertently exposing it to privileged users.
- Retaining the connection to the original data enables more granular analytics than anonymization. Anonymized data is hamstrung by the original parameters, such as age ranges, which might not provide enough granularity or flexibility for future applications. With tokenized data, analysts can create fresh slices of cloud data as needed, down to the original, individual PII.
- Tokenization combines the analytics opportunity of anonymization with the strong at-rest protection of encryption. Look for approaches that limit the amount of previously masked data that can revert to its original form (de-tokenization) and also issue notifications and alerts for de-tokenized data so you can ensure only approved users get the data.

3. **On-demand:** Governing and protecting data doesn't need to be complicated or costly. Apply controls to protect private information without writing code. Reduce expenses by working with vendors whose licensing doesn't charge you to use your data, only to protect it. Legacy platforms typically require enormous investments in time, resources, and money to even get started, which is not a fit today's usage-based, large-scale clouds. SaaS-based platforms that allow you to start with access controls at low or no cost for small organizations and grow to enterprise-class plans with advanced governance controls ultimately provide the most flexibility as you scale.

Move Data and Secure It Faster with ALTR, Matillion ETL and Snowflake

With ALTR and our ETL partners, you can safely and easily migrate your sensitive data to Snowflake and operate on that data at cloud scale. Move data faster by creating an approved sensitive data pipeline and sensitive data destination through an "Extract, Transform, Load, Tokenize" (ETLT) solution based on ALTR, ETL, and Snowflake technologies.

Matillion is a preferred ETL partner of ALTR. Matillion's ETL solution is optimized for Snowflake and takes advantage of the speed, scale, and performance of the cloud to make it easy to load all of an organization's data into Snowflake and transform it to become analytics-ready in no time.

1. ETL and Tokenization:

You can use an ETL solution, ALTR, and Snowflake together without changing your ETL workflows. The ETL solution loads sensitive data into a staging database in Snowflake and contacts ALTR for automatic tokenization of data in real-time. You can then copy loaded and tokenized data into Snowflake production databases for sharing and analysis. If any sensitive data within Snowflake needs to be transformed before analysis -- it can be detokenized, viewed and transformed in the ETL solution by an authorized user, and tokenized again with ALTR.

2. Operations and Analytics Post-Migration:

Using ALTR to protect your data, you can analyze and join tables based on tokenized columns without seeing the underlying data. ALTR's deterministic tokenization technology avoids detokenization to reduce risk overall. ALTR can also allow locked-down access, which includes auditing of every access and placing limits on read operations based on time or access (rate-limiting) to prevent data theft or accidental exposure.

The combination of ALTR, your ETL solution, and Snowflake allows you to get more value from your data in minutes, not months.

Ready to Migrate Sensitive Data to the Cloud?

A SaaS-based data control and security solution built for the cloud is tailor made for transferring and securing data seamlessly in the cloud. Once you unify your data in the cloud, you can immediately understand your data and how it is used. You've removed the complexity because there's no hardware between your users and the data, no risk of a platform change breaking connections and integrations, and no difficulty scaling with your cloud usage. You can add security measures like tokenization to protect private information and lower your data risk, while allowing collaboration and sharing. Simplifying control over your data is the fastest way to unleash value, to make it exponentially more functional for your business or mission.

If you're planning or kicking off your cloud data migration and want to learn how ALTR and our ETL partners can make it smooth and easy, let's schedule a demo.

Security Considerations:

- **Security considerations in planning your data move**
- **Selecting a data migration partner**
- **Selecting a data security partner**
- **Ready to migrate sensitive data to the cloud?**

Complete data control and protection

ALTR simplifies and unifies data governance and security, allowing anyone the ability to confidently store, share, analyze, and use their data. With ALTR, customers gain unparalleled visibility into how sensitive data is used across their organization. This intelligence can be used to create advanced policies to control data access. Through ALTR's cloud platform, customers can implement data governance and security in minutes instead of months.

Get started for free at get.altr.com/free