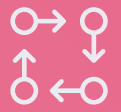
eBook

Level Up your Cloud Data Security with Tokenization

Advantages and implementation models



Introduction

A handful of “tokens” used to be as good as gold at the local arcade. It meant a chance to master skee-ball or prove yourself a pinball wizard. Today, “tokenization” is a data security technology for protecting sensitive data. By substituting a “token” with no intrinsic value in place of sensitive data, such as social security numbers or birth dates that do have value, usually quite a lot of it, companies can keep the original data safe in a secure vault, while moving tokenized data throughout the business.

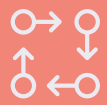
And today, one of the places just about every business is moving data is to the cloud. Companies may be using cloud storage to replace legacy onsite hardware or to consolidate data from across the business to enable BI and analysis without affecting performance of operational systems. To get the most of this analysis, companies often need to include sensitive data.

Three Risk-based Models for Tokenizing Sensitive Data in your Cloud Data Warehouse

But where do you start tokenizing data? We see three main models for tokenization on your cloud data journey. The best choice for your company will depend on the risks you’re facing:

Level 1: Tokenization just before moving to the cloud data warehouse

Level 2: Tokenization before moving through the ETL Process



Level 3: Full end-to-end tokenization

Why Tokenization is an Ideal Solution for the Cloud Data Warehouse

Companies that know they need to protect data in the cloud warehouse may be considering options like encryption or anonymization. Tokenization actually has some very clear, specific advantages over those two other options:

#1 Tokens have no mathematical relationship to the original data, which means unlike encrypted data, they can't be broken or returned to their original form. While many of us might think encryption is one of the strongest ways to protect stored data, it has a few weaknesses, including this big one: the encrypted information is simply a version of the original plain text data, scrambled by math. If a hacker gets their hands on a set of encrypted data and the key, they essentially have the source data. That means breaches of sensitive PII, even of encrypted data, require reporting under state data privacy laws. Tokenization on the other hand, replaces the plain text data with a completely unrelated "token" that has no value if breached. Unlike encryption, there is no mathematical formula or "key" to unlocking the data – the real data remains secure in a token vault.

2 Tokens can be made to match the relationships and distinctness of the original data so that meta-analysis can be performed on tokenized data. When one of the main goals of moving data to the cloud is to make it available for analytics, tokenizing the data delivers a distinct advantage: actions such as counts of new users, lookups of users in specific locations, and joins of data for the same user from multiple systems can be done on the secure, tokenized data. Analysts can gain insight and find high-level trends without requiring access to the plain text sensitive data. Standard encrypted data, on the other hand, must be decrypted to operate on, and once the data is decrypted there's no guarantee it will be deleted and not be forgotten, unsecured, in the user's download folder. As companies seek to comply with data privacy regulations, demonstrating to auditors that access to raw PII is as limited as possible is also a huge bonus. Tokenization allows you to feed tokenized data directly from Snowflake into whatever application needs it, without requiring data to be unencrypted and potentially inadvertently exposed to privileged users.

#3 Tokens maintain a connection to the original data, so analysis can be drilled down to the individual as needed. Anonymized data is a security alternative that removes the personally identifiable information by grouping data into ranges. It can keep sensitive data safe while still allowing for high-level analysis. For example, you may group customers by age range or general location, removing the specific birth date or address. Analysts can derive some insights from this, but if they wish to change the cut or focus in, for example looking at users aged 20 to 25 versus 20 to 30, there's no ability to do so. Anonymized data is limited by the original parameters which might not provide enough granularity or flexibility. And once the data has been analyzed, if a user wants to send a marketing offer to the group of customers, they can't, because there's no relationship to the original, individual PII.

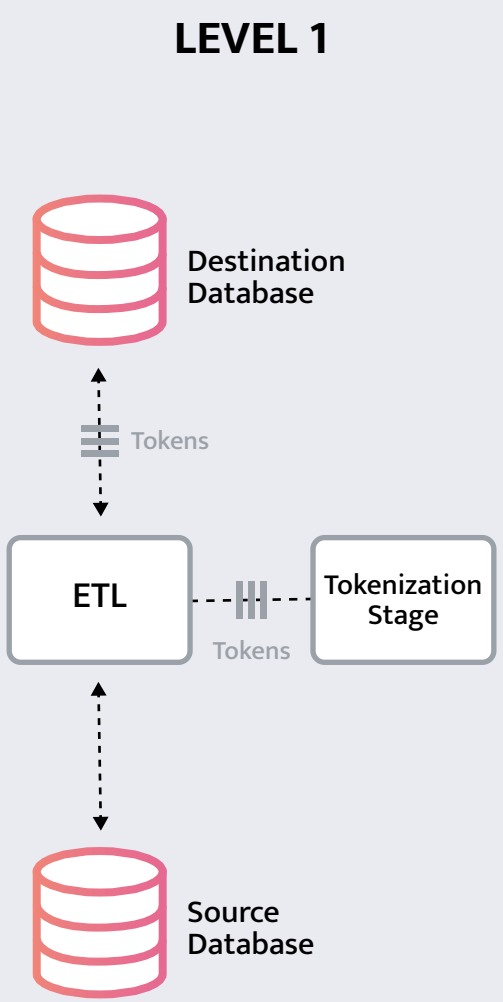
Tokenization essentially provides the best of both worlds: the strong at-rest protection of encryption and the analysis opportunity provided by anonymization. It delivers tough protection for sensitive data while allowing flexibility to utilize the data down to the individual. Tokenization allows companies to unlock the value of sensitive data in the cloud.

Level 1: Tokenization just before moving to the cloud data warehouse

If you're concerned about protecting sensitive data in your destination database, often a cloud data warehouse like Snowflake, you're not alone. While you may feel confident in the security measures in place in your own datacenter, storing sensitive data in the cloud is a different ballgame.

The first issue might be that you're consolidating sensitive data spread across multiple databases, each with its siloed and segmented security and log in requirements, into one central repository. This means a bad actor, dishonest employee or hacker just needs to sneak into one location to have access to all your sensitive data. It creates a much bigger risk. And this leads to the second issue: as more and more data is stored in prominent cloud data warehouses, they have become a top target for bad actors and nation states. Why should they target Salesforce or Workday separately when all the same data can be found in one place? The third concern might be about privileged access from Snowflake employees or your Snowflake admins who could, but really shouldn't, have access to the sensitive data in your offsite database.

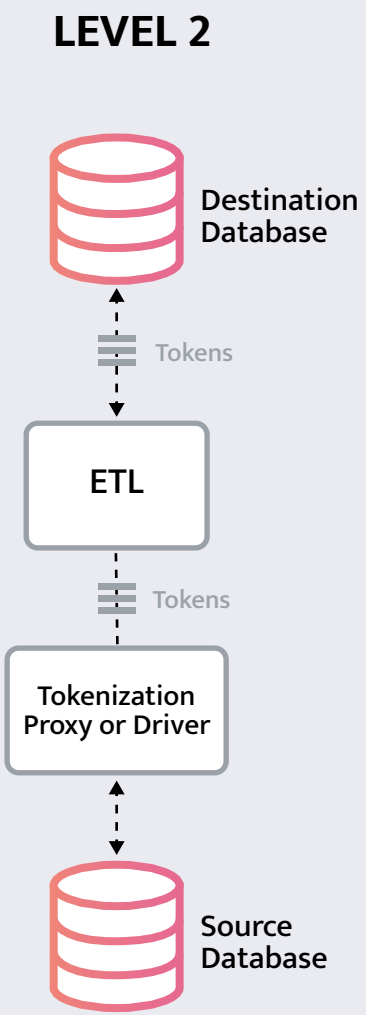
In these cases, it makes sense for you to choose "Level 1 Tokenization": tokenize data just before it goes into the cloud. By tokenizing data that is stored in the database, you ensure that only the people you authorize have access to the plain text data.



Level 2: Tokenization before moving through the ETL Process

As you're planning your path to the cloud, you may be concerned about data as soon as it leaves the secure walls of your datacenter. This is especially challenging for CISOs who've spent years hardening the security perimeter only to have control wrested away as sensitive data is moved to cloud data warehouses they don't own. If you're working with an outside ETL (extract, transform, load) provider to help you prepare, combine, and move your data, that will be the first step outside your perimeter causing concern. Even though you hired them, without years of built-up trust, you may not want them to have access to sensitive data. Or it may even be out of your hands—you may have agreements or contracts with your customers that specify you can't let any vendor or other entity have access without written consent.

In this case, "Level 2 Tokenization" is appropriate. This takes one step back in the data transfer path and tokenizes sensitive data before it even reaches the ETL. Instead of direct connection to the source database, the provider connects through the tokenization software which returns tokens. ALTR partners with SaaS-based ETL providers like Matillion to make this seamless for enterprises.

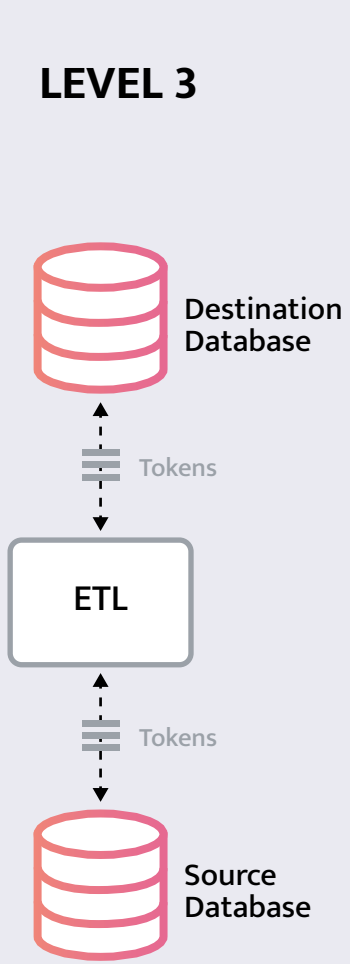


Level 3: Full end-to-end tokenization

If you're a very large financial institution classified as a "critical vendor" by the US government, you're familiar with the rigorous security required. This includes ensuring that ultra-sensitive financial data is highly secure – no unauthorized users, inside or outside the enterprise, can have access to that data, no matter where it is. You already have this nailed down on-prem, but we're living in the 21st century and everyone from marketing to operations is saying "you have to go to the cloud." In this case, you'll need "Level 3 Tokenization": full end-to-end tokenization of all your onsite databases all the way through to your cloud data warehouse.

As you can imagine, this can be a complex task. It requires tokenization across multiple on-premises systems before even starting the data transfer process. The upside is that it can also shine a light on who's accessing your data, wherever it is. You'll quickly hear from people throughout the company who relied on sensitive data to do their jobs when the next time they run a report all they get back is tokens. This turns into a benefit by stopping "dark access" to sensitive data.

Our customers found this so valuable, ALTR created a tool called "the observer" they can hook up to a source database to provide a report of every connection, every user, every IP address and every query run against the source database. The process forces companies to understand the full lifecycle of their data – where it's being used and by whom – and helps ensure sensitive data is truly secure.



Use the power of ALTR's tokenization platform

ALTR's tokenization platform provides unique data security benefits across your entire path to the cloud. Our SaaS-based approach means we can cover data wherever it's located: on-premises, in the cloud or even in other SaaS-based software like Salesforce. This also allows us to deliver innovations like new token formats or new security features more quickly, with no need to upgrade. Our tokenization solutions also range from the most fundamental level all the way up to PCI Level 1 compliant, allowing companies to choose the best balance of speed, security, and cost for their business. We've also invested heavily in IP that enables our database driver to connect transparently and keep data usable while tokenized. The drivers can, for example, perform the lookups and joins needed to keep applications that are unused to tokenization running.

With tokenization from ALTR, you can bring sensitive data safely into the cloud to get full analytic value from it, while helping meet contractual security requirements or the steepest regulatory challenges.

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