The HIPAA Security Technical Safeguard Rule and the Snowflake Elastic Data Warehouse

The Snowflake Elastic Data Warehouse has been designed from the start with a comprehensive, integrated approach to data security. It provides industrial-strength security capabilities complemented by processes and safeguards to provide data security that can meet important compliance and regulatory requirements, all without requiring an army of security experts or interfering with user access to data.

In this paper we'll examine in technical detail how Snowflake's data warehouse enables customers to demonstrably meet HIPAA data security requirements. Snowflake's Enterprise for Sensitive Data edition, required for data subject to HIPAA requirements, has been developed to enable customers to meet these requirements while opening the door to the advantages of a data warehouse in the cloud

THE HIPAA SECURITY RULE FOR TECHNICAL SAFEGUARDS

Organizations that use electronic protected health information (ePHI) must comply with HIPAA Security Rules, which include administrative, physical, and technical safeguards. Snowflake's Elastic Data Warehouse uses technology that helps customers meet the requirements of the Technical Safeguards,[1] and to evidence compliance to risk assessors, auditors, and reviewers. This section recaps the Technical Safeguards rule and poses some questions that assessors might ask or the risk manager may be concerned about.

Access Control Standard

The Access Control Standard requires that access to ePHI be restricted "to allow access only to those persons or software programs that have been granted access rights…."[2] This standard has one Required implementation specification, *unique user identification,*[3] and two Addressable implementation specifications, *automatic logoff*[4] and *encryption / decryption*.[5] Here are some of the questions that assessors might ask where Snowflake technology can help.

Unique user identification

* Does the technology support unique user identification?
* Does this include software processes that access ePHI? How is this done?
* Are administrative users authenticated with multiple factors? Non-administrative users? What factors?

Automatic logoff

* Does the technology support automatic logoff of users or processes after a period of inactivity?

Encryption and decryption

* Does the technology support encryption of the data?
* Does an administrator have to decide what to encrypt and what not to encrypt?
* Is data at rest always encrypted, including backups?
* Is data in flight encrypted as soon as it is received, if not already encrypted? In flight between any two devices once received? In flight between virtual machines within the same physical machine? Before transmission to external networks?
* Are the encryption method and key lengths sufficiently strong to satisfy NIST recommendations?
* Are the keys strongly protected?
* Can the keys be changed easily and frequently?

Audit Control Standard

The Audit Control Standard requires that organizations "record and examine activity in information systems that contain or use electronic protected health information."[6] This standard does not have any implementation specifications. Here are some questions that assessors might ask.

* Does the technology support recording of all access to ePHI, by persons and processes?
* Does the recording include who the actor was, what the information accessed was, what privileged actions were taken, and when it happened?
* Does recording audit data require administrator action? Could an administrator defeat logging, or forget to turn it on?
* Does the system automatically supply access and

[1] https://www.gpo.gov/fdsys/pkg/CFR-2015-title45-vol1/pdf/CFR-2015-title45-vol1-sec164-312.pdf
[2] 164.312(a)(1)
[3] 164.312(a)(2)(i)
[4] 164.312(a)(2)(iii)
[5] 164.312(a)(2)(iv)
[6] 164.312(b)

activity reports?

* Can the system automatically detect and report unauthorized access attempts?
* Can the system provide sufficient data to identify suspicious or out-of-pattern access attempts?
* How long is the audit data retained? Can it be altered by anyone?

Integrity Standard

The Integrity Standard requires that organizations "protect electronic protected health information from improper alteration or destruction".[7] Some questions could be:

* Does the technology provide a way to limit information access privileges (read, write, modify, delete) based on the role of the person or entity?
* Are changes to the access privileges of a user, entity or role logged?
* Does the technology provide a way to detect unauthorized alteration of data?
* Are messages received by processes individually authenticated?

Person or Entity Authentication Standard

The Person or Entity Authentication Standard requires organizations to have "procedures to verify that a person or entity seeking access [to ePHI] is the one claimed."[8] This standard has no implementation specifications. Potential questions from assessors include:

* Does the technology support authentication using multiple authentication factors?
* Are administrative users required to authenticate using multiple factors?
* Is multi-factor authentication available to non-administrative users?
* Are processes required to authenticate?

Transmission Security Standard

Finally, the Transmission Security Standard requires that organizations "guard against unauthorized access [to ePHI] that is being transmitted over an electronic communication network."[9] This standard has two implementation specifications, both Addressable: *integrity controls* ("electronically transmitted [ePHI] is not improperly modified without detection");[10] and to *"encrypt [ePHI] whenever deemed appropriate."*[11] The last clause can be problematic because it can put the manager at risk of disagreeing with an auditor as to what she "deems" is a cost-effective security method. Some questions are:

Integrity controls

* Does the technology provide a way to protect the integrity of data in transit?
* Are messages received by processes individually authenticated?

Encryption

* What data is encrypted and what is not?
* Is data in flight between any two devices in the data center encrypted, even behind the firewall, and even between virtual machines within the same physical machine?
* Other encryption questions as listed under the Access Control standard.

ENCRYPTION HAS ITS CHALLENGES

Data encryption is a critical part of protecting data. However, encryption is by no means universal due to a number of challenges. In order to understand the advantages of Snowflake's encryption technology and practices, let's first consider some of these challenges. Then we'll review how Snowflake addresses them.

Weak and obsolete algorithms

Cryptographic algorithms that were once considered strong have later been shown to be vulnerable.[12] The only safe choice is an algorithm that is currently considered best in class. As NIST states: "Whenever possible, AES [Advanced Encryption Standard] should be used for the encryption algorithm because of its strength and speed."[13]

Increasing attacker compute power

7 164.312(c)(1)
8 164.312(d)
9 164.312(e)(1)
10 164.312(e)(2)(i)
11 164.312(e)(2)(ii)
12 For example, the RC4 stream cipher was believed for many years to be strong. https://en.wikipedia.org/wiki/RC4
13 NIST SP800-111, p. 4-4

Attackers have been aggressive in deploying rapidly increasing computer power for cryptographic attacks. Algorithms and key lengths may become susceptible in the future to new attacks based on advances in mathematics and computer power. NIST advises, "Organizations should consider how easily the solution can be updated when stronger algorithms and key sizes become available in the future."[14] It should be not only possible but easy to re-key an encrypted data store with a new key.

Selective encryption

Because encryption has been computationally expensive, and because encryption "bolted on" may require expensive modification to applications, some organizations have opted to encrypt only the subset of data elements (like particular columns in a table) that are considered sensitive. However, what is considered sensitive can change. Customers, market sectors and regulators differ. HIPAA requirements for de-identification of patient data are much more restrictive than the common understanding of the Gramm-Leach-Bliley Act. The straightforward solution would be to encrypt everything, even images, but historically that's been too costly.

Aging keys

The longer a key is used, the more data is accessible if the key is compromised, and the longer an attacker can apply compute power to crack while it's still in use. Organizations need to be able to change the key to keep the data safe.[15]

Difficulty of rekeying

Given that we should change keys when necessary, how is this to be done for data at rest? If the data has a short shelf life, the organization could use a new key going forward, and destroy the data encrypted with the old key after its age-out period. This approach probably requires the application to be able to recognize and handle multiple keys, which may not be practical. But this does not work for data that must be retained for long periods of time—one HIPAA rule requires medical records to be retained for six years.[16] Another approach is to decrypt everything with the old key and re-encrypt with the new key, taking care to include all copies and backups. With terabytes of data, this re-encryption in software can take days or weeks.

Locking users out

Data may not be available to users or applications during the rekeying. If the rekeying can be done overnight or over a long weekend, denying access to users might be acceptable. But with terabytes of data, a long lockout time is not acceptable.

Customer-specific keys

When the data store includes information from multiple organizations, as in a multi-tenant cloud environment, one or more unique keys for each organization are needed. Customers of a multi-tenant service need assurance that no other customer has the same keys.

Key protection

Even the strongest cryptographic algorithms can be weakened if the keys are poorly protected. The keys themselves should be encrypted and not accessible by unauthorized persons or processes, and resistant to destructive attacks or forces.

Proper management of cryptographic keys can be a significant burden on an organization.

SNOWFLAKE ENCRYPTION TECHNOLOGY MEETS THE CHALLENGES

To ensure that data stored in Snowflake is always encrypted, Snowflake has integrated data encryption into its product. Snowflake's hierarchical key management technology[17] and integrated key management (more on that in the next section) solve many of the challenges customers have with encryption.

Snowflake currently uses 256-bit Advanced Encryption System (AES) encryption. There is currently no known and practical way to crack AES-256 without some weakness in the way keys are protected.[18]

The figure nearby shows how hierarchical key

14 NIST SP800-111, p. 4-4

15 NIST has recommendations for key changes in Special Publication 800-57. http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r4.pdf

16 https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNMattersArticles/downloads/SE1022.pdf

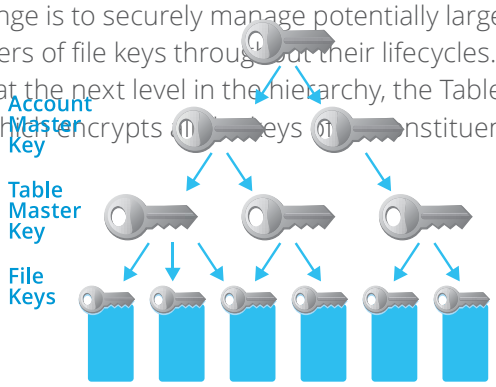
17 Hierarchical key management is described in http://www.snowflake.net/resource/industrial-strength-security-by-default/.

18 Bruce Schneier, a leading expert and author on encryption, believes that not even the NSA can crack AES. https://www.schneier.com/blog/archives/2012/03/can_the_nsa_bre.html

management works. Let's start at the bottom. Each file has its own key, and if files are small enough, their keys can be changed at any time with only minimal disruption to any user or application. This meets several challenges of rekeying—provided there is a good way to manage many keys:

* Keys can be replaced with new keys, keeping ahead of potential attackers.
* Keys can be replaced well before the end of their crypto-periods—or almost any time, just to be safe.
* Users are not locked out during rekeying, and applications remain performant.
* Every customer, data source, application, data table and file has its own key.

Providing these four benefits is why Snowflake has implemented hierarchical key management. The next challenge is to securely manage potentially large numbers of file keys throughout their lifecycles. This is done at the next level in the hierarchy, the Table Master Key, which encrypts all keys constituent files.



Snowflake employs a hierarchical key model to securely encrypt data

Here is how a file key is changed: use the Table Master Key to decrypt the file key, generate a new file key, lock the file, decrypt the file, re-encrypt the file with the new key, unlock the file, and encrypt the new file key with the Table Master Key. In this process, access to the other files in the table is uninterrupted, and access to the one table is interrupted only briefly.

In a similar way, Table Master Keys can be changed. At the next level of the hierarchy, all Table Master Keys for a particular customer are encrypted with an Account Master Key that is unique to the customer. The Table Master Key is like the key to a key safe that contains other keys. Changing the Table Master Key is like changing the key to the key safe. Since file keys are not changed when the Table Master Key is, user access to the data is not interrupted.

Snowflake manages the keys used to encrypt data, including changing keys. There are two types of key changes: key rotation, in which "new data gets fresh keys," and re-keying, in which "old data gets fresh keys." Snowflake rotates and re-keys on prescribed intervals. Snowflake's management of keys relieves the customer of the cost and responsibility for protecting and changing keys.

At the top of the hierarchy, all Account Master Keys are encrypted and stored in a hardware security module (HSM). This is a device that is dedicated to Snowflake (not accessible to any non-Snowflake personnel), and meets or exceeds the highest level of requirements for HSMs stated by NIST.[19] The device used by Snowflake has been tested and validated to meet requirements by an independent testing laboratory accredited by NIST.

Snowflake provides an in-depth description of hierarchical key management in another paper.[20]

So far, we have discussed how Snowflake's hierarchical key management technology solves these problems:

* Weak and obsolete algorithms
* Increasing attacker compute power
* Aging keys
* Difficulty of rekeying
* Locking users out
* Customer-specific keys
* Key protection

These benefits accrue to the extent that data is encrypted, but as we've said before, historically it has been too expensive—computationally and financially—to do the straightforward thing and encrypt everything.

ENCRYPTION ALWAYS AND EVERYWHERE

Snowflake from day one has taken the design approach of encrypting everything and everywhere. Highly scalable and affordable cloud compute power makes this possible. Snowflake's divide-and-encrypt strategy, enabled by hierarchical key management, allows

[19] Specifically, evaluation assurance level 4 plus, EAL4+. See FIPS PUB 140-2, http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf
[20] http://www.snowflake.net/blog/encryption-key-management-in-snowflake/

Snowflake to assure that

all of your data at rest will be strongly encrypted all of the time.

That goes for backups too, and even for the cached results of any queries run on the data. The customer does not need to decide what to encrypt, or take the risk of not encrypting something that is later found to need it. The performance and availability impact on users for encrypting all data at rest is minimal.

Snowflake also encrypts data of HIPAA-covered customers in transit at all times within its control. All external connections between the Snowflake Elastic Data Warehouse and customers are encrypted with HTTPS using TLS 1.1 or higher.[21]

All transmissions between components in the Snowflake environment, both physical devices and virtual machines, are encrypted. Some organizations have taken the position that the network "behind the firewall" is trusted, so data in transit there does not need to be encrypted. However this position is falling out of favor, and so in order to avoid all doubt Snowflake encrypts all data in flight, either with HTTPS or AES. Even data in flight between virtual machines sharing the same physical hardware is encrypted using an algorithm based on AES-256, and even though HIPAA customers run on physical machines dedicated to Snowflake.[22] Snowflake assures that

all of your data will be strongly encrypted all of the time in transit.

To summarize the two key attributes of Snowflake encryption:

In a Snowflake Elastic Data Warehouse, all customer data is encrypted at all times at rest and in transit.

SNOWFLAKE IDENTITY AND ACCESS MANAGEMENT

Organizations responsible for protecting ePHI under HIPAA rules are required to implement administrative and physical safeguards, as well as technical safeguards. Organizations depend to some degree on their technology providers to be able to implement safeguards. Compliance sometimes requires a combination of customer procedure and enabling technology. This section describes how Snowflake identity and access management technology enables the customer to comply with the Access Control, Audit Control, Integrity, and Person or Entity Authentication Standards.

To recap, the **Access Control Standard** requires organizations "to implement policies and procedures … to allow access only to those persons or software programs that have been granted access rights …."

The Elastic Data Warehouse implements role-based access control (RBAC). In RBAC, the organization assigns persons (for instance according to job title) to functional roles the person must perform, and then defines the access to information, systems, and privileges that should be permitted to each role. Although most commonly one person may have only one role, RBAC allows for multiple roles per person, as well as multiple persons per role. For example, it may be desirable for a system administrator to have one role to administer production systems, another for test systems, and a third for general business use. The customer can use the RBAC in the Elastic Data Warehouse to flexibly assign and efficiently administer access privileges according to its policies and procedures. All users—not just those with administrative privileges—subject to the HIPAA security rule must use[23] multifactor authentication to access the Elastic Data Warehouse in order to enforce strong authentication. Authentication of users requesting access from trusted smart mobile devices includes a random one-time password. Processes requesting access to data or other processes in the Elastic Data Warehouse are also authenticated.

To further assure that access to databases and data warehouses is restricted to authorized users, customers have the option to white-list authorized source IP addresses. Snowflake also supports SAML 2.0 for federated single sign-on (SSO) services. SAML (Security Assertion Markup Language) is a standard by which two organizations can exchange authorization and authentication data. With SAML, authorized users do not have to log in separately to the Snowflake Elastic Data Warehouse once they have authenticated in their corporate network.

21 Enforcement of higher versions of TLS may depend on the end user's browser.
22 Snowflake requires Enterprise Edition for Sensitive Data to customers covered by HIPAA rules.
23 Assuming Enterprise Editon for Sensitive Data

The Access Control Standard also has two implementation specifications that Snowflake technology supports, *unique user identification* (Required) and *automatic logoff* (Addressable).

The unique user identification implementation specification under the Access Control Standard requires the organization to "assign a unique name and/or number for identifying and tracking user identity." This means that an action taken in a system, like one shown in a log entry, must be traceable to one and only one person. Snowflake's Elastic Data Warehouse supports unique user identification.

The automatic logoff implementation specification requires "procedures that terminate an electronic session after a predetermined time of inactivity." Snowflake addresses this by automatically terminating inactive user sessions.

Snowflake's identity and access management features provide:

* Role-based access control
* Strong two-factor authentication for all users, not just administrators
* One-time random passwords for mobile users
* Optional white-listing for users' source IP addresses
* Support for federated single sign-on with SAML 2.0
* Authentication of processes as well as users
* Automatic logoff

The **Audit Control Standard** requires that organizations "record and examine activity in information systems that contain or use electronic protected health information." Logging processes in the Snowflake Elastic Data Warehouse automatically log all accesses to data, what database or data warehouse was accessed, who accessed it, the location (IP address) and user id of the requestor, and the date and time of access. All administrative actions taken through the web portal are logged. No action by the customer is needed to set up and start logging. Logs are automatically backed up and are tamper-proof, so even administrators cannot defeat logging. Logs are retained for six months. Read access to the last 100 log entries is available to authorized customer administrators through the web portal. (Snowflake will fulfill customer requests for older log data on an individual-case basis.) Upon request, Snowflake can also provide customer administrators with automated reports showing daily accesses and other activity.

In support of the **Integrity Standard** to "protect [ePHI] from improper alteration or destruction," Snowflake uses encryption based on TLS 1.1 or higher on all transmissions, internal and external. The TLS protocol includes message authentication, which prevents undetected loss or alteration of the message in transit. If, in spite of these technical safeguards, data has been corrupted by accidental or other errors, the customer can instantly restore or query any previous version of a table or database as of an arbitrary point in time within a standard retention period by using Snowflake's Time Travel feature. The customer can specify an extended retention period at the time the table or schema is created.[24] Snowflake employs file integrity monitoring (FIM) and host-based intrusion detection (HIDS) to protect software processes and thereby the data they access from corruption. These protections are monitored and managed internally by Snowflake.

The customer can use RBAC to limit access privileges according to its need-to-know and least-privilege policies and procedures. Changes to access privileges of users, roles, and processes are logged. Reports of all access attempts are automatically sent to customer administrators who have requested such reports. Using these reports, suspicious and out-of-pattern activity may be detected.

In sum, Snowflake technology supports the Integrity Standard by:

* Authenticating all messages using Transport Layer Security
* Logging all access attempts
* Logging all privileged actions
* Automatically reporting log entries daily

THE SNOWFLAKE ADVANTAGE

Recent advances in technology offer healthcare payers and providers with tantalizing and unparalleled

[24] For more on Time Travel please see http://info.snowflake.net/ContinuousDataProtection_WP.html. Time Travel is optional.

opportunities to leverage data from many sources to reduce cost, improve the quality of care, and offer new services. An ever-increasing amount of data is available, not only from electronic health records, patient monitors, and social media but also from new sources including genomics, proteomics, and patient wearables. Cloud computing and advanced analytics platforms provide affordable and scalable ways to mine this data.

At the same time, cyber criminals are finding more ways and having more success in stealing and monetizing patient data. Intensifying regulatory scrutiny reflects the public's concern for the safety and privacy of its healthcare data.

The challenge for leadership is:

How can we take advantage of our data for analytics, and position ourselves for the future, while safeguarding patient data and leveraging the economy and scalability of cloud computing? Can we do all this without exposing ourselves to too much risk, and without having to hire more security staff?

The organization cannot outsource its responsibility to implement the security controls mandated by the HIPAA rules. But what it can do is select the right business partner, one that has implemented a comprehensive set of safeguards and capabilities, and that has made it easy to leverage those protections.

The technology in Snowflake's Elastic Data Warehouse enables customers to meet the requirements of the HIPAA Security Rule in many ways.[25] In fact, Snowflake compliance to relevant HIPAA requirements has been attested by an independent third party auditor according to AICPA standards. Key features include:

* Industrial-strength encryption, always on, for all data at rest and in transit
* Cryptographic key management
* Strong, two-factor authentication for all users
* Authentication of all messages and processes
* Logging of all access attempts and all privileged actions
* Daily activity reports

At Snowflake, security is architected into the technology from the beginning. This is how we make it cost-effective, seamless, and non-disruptive for our customers.

25 Snowflake has implemented technology in certain products and services to meet HIPAA requirements. Snowflake requires its Enterprise for Sensitive Data edition for customer environments covered by HIPAA rules.

About Snowflake

Snowflake Computing, the cloud data warehousing company, has reinvented the data warehouse for the cloud and today's data. The Snowflake Elastic Data Warehouse is built from the cloud up with a patent-pending new architecture that delivers 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 can be found online at snowflake.net.

Copyright © 2016 Snowflake Computing, Inc. All rights reserved.
SNOWFLAKE COMPUTING, the Snowflake Logo, and SNOWFLAKE ELASTIC DATA WAREHOUSE are trademarks of Snowflake Computing, Inc.

WP-HSTSRSEDW-1_1_062016

snowflake

www.snowflake.net | @SnowflakeDB