CLOUDERA

WHITE PAPER

HIPAA COMPLIANCE FOR YOUR CLOUDERA DATA PLATFORM

How Cloudera Ensures Personal Health Information Data Security and Governance Standards

CDP

Table of Contents

Overview

As part of the Health Insurance Portability and Accountability Act (HIPAA), the HIPAA Privacy Rule addresses the use and disclosure of individuals' health information called protected health information (PHI). The Privacy Rule protects all "individually identifiable health information" stored or transmitted by a covered entity or its business associates, in any form or media.

PHI is a core component of HIPAA and includes demographic data such as name, address, birthdate, Social Security number or other information that is useful in identifying an individual and can be used to relate his or her:

- Past, present or future physical and mental health or condition
- Provision of health care
- Payment for the provision of health care

According to the American Health Information Management Association (AHIMA), during a hospitalization, an average of 150 people have access to a patient's medical records, including x-ray technicians, nursing staff and billing personnel, among others. While many of these individuals have a legitimate need to see all or part of a patient's records, prior to HIPAA and the Privacy Rule, there wasn't any regulation that specified:

- Which personnel could access patient records
- What type of information approved personnel could see
- Which actions approved personnel were permitted to perform with the information once they had access to it

The objective of the Privacy Rule is to strike a balance between protecting individuals' health information and facilitating the flow of health information that is required to provide and promote high quality health care.

HIPAA and the Privacy Rule have important data security and governance implications that have led to the application of technologies to control access to personal health information. These applications include the authentication and authorization of individuals who have access to patient information and the establishment of audit trails of those accessing or modifying information at different levels.

Covered Entities

The organizations that are subject to the Privacy Rule are known as covered entities. These include:

- Health plan organizations that provide or pay for the cost of medical care
- Health care providers including hospitals, physicians, dentists and any person or organization that furnishes, bills, or is paid for health care services
- Health care clearinghouses including billing services, repricing companies, community health management information systems and value-added networks and switches

Penalties for Non-Compliance

The Office for Civil Rights (OCR) in Health and Human Services (HHS) is responsible for implementing and enforcing the Privacy Rule. OCR may impose a penalty on a covered entity for failure to comply with the Privacy Rule. Penalties take into consideration factors such as the date of the violation, whether the covered entity knew or should have known that it was in non-compliance, or whether the covered entity's failure to comply was due to willful neglect.

In 2009 Congress passed the Health Information Technology for Economic and Clinical Health Act (HITECH). HITECH strengthens HIPAA enforcement by greatly increasing the penalties for violations and authorizing states' attorneys general for its enforcement. HITECH also outlines the requirements for first federal data security breach notification, and mandated HHS to conduct privacy and security audits.

	FOR VIOLATIONS OCCURRING PRIOR TO 2/18/2009	FOR VIOLATIONS OCCURRING ON OR AFTER 2/18/2009
Penalty Amount	Up to $100 per violation	$100 to $50,000 or more per violation
Calendar Year Cap	$25,000	$1,500,000

A person who knowingly obtains or discloses individually identifiable health information may face a criminal penalty of up to $50,000 and up to one-year imprisonment. The criminal penalties increase to $100,000 and up to five years imprisonment if the wrongful conduct involves false pretenses, and to $250,000 and up to 10 years imprisonment if the wrongful conduct involves the intent to sell, transfer, or use identifiable health information for commercial advantage, personal gain or malicious harm.

Breach Notification

The HIPAA Breach Notification Rule requires covered entities to notify the affected individuals, the Secretary of HHS, and in certain circumstances the media of a breach of protected health information. A breach is defined as an impermissible use or disclosure that compromises the security or privacy of the protected health information.

Also the business associate is responsible for notifying the covered entities, in case the breach occurs at or by the business associate.

Building Support for HIPAA Compliance with Your Cloudera Data Platform

The Cloudera Data Platform (CDP) forms the core of the modern data architecture that allows enterprises to store, process and gain insight from massive amounts of structured, semi-structured, and unstructured data across infrastructures ranging from on-premises to hybrid and multi-cloud. Trusted by hundreds of healthcare customers and proven in highly security-conscious environments, CDP provides robust security capabilities to help businesses meet HIPAA requirements.

CLOUDERA DATA PLATFORM (CDP)

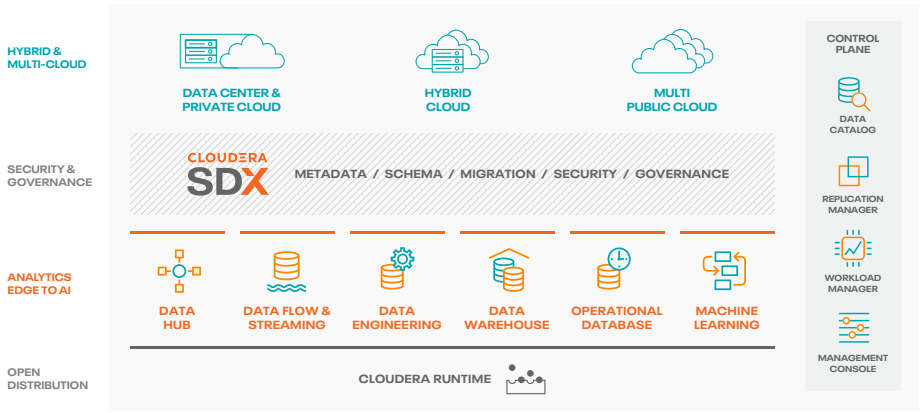

Figure 1. Cloudera Data Platform

Based on a holistic approach to protection, the Cloudera security framework revolves around five pillars: administration, authentication/perimeter security, authorization, audit and data protection. Rather than applying protection as an afterthought, Cloudera prevents gaps and inconsistencies through a bottom-up platform approach that makes it possible to enforce and manage security across the stack through a central point of administration and management built into the DNA of CDP.

The Cloudera SDX (Shared Data Experience) is a differentiating platform capability that provides centrally managed and consistently applied data security, governance and control across all CDP deployments. As such, it is a critical capability that embodies the Cloudera security framework and helps healthcare organizations efficiently develop and manage multi-function analytics across their complete IT landscape.

The Cloudera security architecture implemented in SDX begins with Ranger, Knox and Kerberos. Ranger serves as the central interface for security administration. Users can create and update policies, which are then stored in a policy database. Ranger plugins consisting of lightweight Java programs are embedded within processes of each cluster component. These plugins pull in policies from a central server and store them locally in a file. When a user request comes through the component, these plugins intercept the request and evaluate it against the security policy. Plugins also collect data from the user request and follow a separate thread to send this data back to the audit server.

This platform approach ensures that each of the five security pillars complement each other effectively to enable comprehensive protection.

	HIPAA TECHNICAL SAFEGUARDS—REQUIREMENT SUMMARY			
	Authentication w/ Perimeter Security	Authorization	Audit	Data Protection
CDP	• Kerberos • Perimeter security with Knox	• Fine grained access control with Ranger	• Centralized audit reporting with Ranger and Atlas	• Wire encryption • HDFS encryption with Ranger KMS

Figure 2. Comprehensive security in CDP

Administration

Ranger as part of SDX provides a single pane of glass to define, administer and manage security policies consistently across all the components of the platform. Ranger enhances the productivity of security administrators and reduces potential errors by empowering them to define security policies once and apply them consistently and without further configuration or effort to all the applicable components across the platform from a central location.

Authentication And Perimeter Security

The Knox Gateway ensures perimeter security for Cloudera customers, providing a way for users to reliably identify themselves and then have that identity propagated throughout the cluster to access resources such as files and directories, and to perform tasks such as querying Hive table data from their SQL editor or BI tool, all facilitated through the single sign-on (SSO) capability Knox, as part of the platform, provides. Cloudera uses Kerberos, an industry standard, to authenticate users and resources within the platform. Cloudera has also completely hidden Kerberos setup, configuration and maintenance. With Knox, enterprises can confidently extend the various REST APIs to new users without Kerberos complexities, while also maintaining compliance with enterprise security policies. Knox provides a central gateway for REST APIs that have varying degrees of authorization, authentication, SSL and SSO capabilities to enable a single access point for the complete platform.

Authorization

Ranger manages fine-grained access control through a rich user interface that ensures consistent policy administration across data access components. Security administrators have the flexibility to define security policies for a database, table and column or a file, and administer permissions for specific LDAP based groups or individual users. Rules based on dynamic conditions such as data classification (through Atlas), time or geography can also be added to an existing policy rule. The Ranger authorization model is highly pluggable and can be easily extended to any data source using a service-based definition. Ranger works with standard authorization APIs in each platform component and is able to enforce centrally administered policies for any method of accessing the data lake. The combination of Ranger's rich user interface with deep audit visibility makes it highly intuitive to use, enhancing productivity for security administrators.

Audit

Atlas is a set of core foundational governance services that enables enterprises to effectively and efficiently meet their compliance requirements within the complete platform and allows integration with the complete enterprise data ecosystem.

Ranger also provides a centralized framework for collecting access audit history and easily reporting on this data, including the ability to filter data based on various parameters. Together with Atlas, this makes it possible for users to gain a comprehensive view of data lineage and audit data access, with an ability to query and filter audit information based on data classification, users or groups, as well as other filters.

Data Protection

CDP adds a robust layer of security by making data unreadable in transit over the network or at rest on a disk. Cloudera has long since introduced HDFS encryption, complemented with a Ranger-embedded open source key management store (KMS). Ranger provides security administrators with the ability to manage keys and authorization policies for KMS. Cloudera is also working extensively with its encryption partners to integrate HDFS encryption with enterprise-grade key management frameworks. With Cloudera, customers have the flexibility to leverage either an open source key management system (KMS), or use enterprise wide KMS solutions provided by the partner ecosystem.

Encryption in HDFS, combined with KMS access policies maintained by Ranger, prevents rogue Linux or platform administrators from accessing data and supports segregation of duties for both data access and encryption.

How Cloudera Helps You Meet Key HIPAA Requirements

A covered entity must maintain reasonable and appropriate technical and physical safeguards to prevent intentional or unintentional disclosure of protected health information in violation of the Privacy Rule.

HIPAA TECHNICAL SAFEGUARDS—REQUIREMENT SUMMARY

Standards	HIPAA Section	Implementation Specifications	Required (R) or Addressable (A)[1]	Cloudera Competency
Access Control	§ 164.312(a)(1)	1. Unique User Identification	R	Kerberos, Apache Knox & Apache Ranger with LDAP/AD integration
		2. Automatic Logoff	A	Ranger
		3. Encryption and Decryption	A	Wire Encryption HDFS Encryption with Ranger
Audit Control	§ 164.312(b)	Hardware & Software Procedural Controls	R	Cloudera Replication Manager, Ranger, Apache Atlas
Integrity	§ 164.312 (c) (1)	Mechanism to Authenticate Electronic PHI	A	Ranger, Kerberos & Knox
Person or Entity Authentication	§ 164.312 (d)	Procedures to Verify That a Person or Entity Seeking Access to Electronic PHI is the One Claimed	R	Ranger & Knox
Transmission Security	§ 164.312 (e)(1)	1. Integrity Controls	A	Replication Manager
		2. Encryption	A	Wire Encryption HDFS Encryption with Ranger

[1] Covered entities are required to comply with every Security Rule classified as "Standard." Certain implementation specifications within those standards are classified as "addressable," while others as "required." The required implementation specifications must be implemented. The addressable designation permits covered entities to determine whether the implementation specification is reasonable and appropriate for that covered entity.

HIPAA REQUIREMENT	HOW CLOUDERA CAN HELP
Access Controls: A covered entity must implement technical policies and procedures that allow only authorized persons to access electronic protected health information (e-PHI)	CDP simplifies implementation of Kerberos for authentication within a cluster. Knox provides authentication for REST-based services and can be integrated with AD/LDAP or other authentication mechanisms. Integrating Kerberos and Knox with LDAP/AD also enables organizations to identify unique users and provide SSO across the platform.
Audit Controls: A covered entity must implement hardware, software, and/or procedural mechanisms to record and examine access and other activity in information systems that contain or use e-PHI	Ranger provides a centralized framework to collect access audit history and easily report on this data, including the ability to filter information based on various parameters. The combination of Ranger and Atlas enables users to gain a comprehensive view of data lineage and access audit, with an ability to query and filter audit information based on data classification, users or groups, and other filters. CDP supports wire encryption for all access protocols as well as for Solr, Kafka and YARN, with dynamic attributes such as geolocation, time and data to drive security policy decisions.
Integrity Controls: A covered entity must implement policies and procedures to ensure that e-PHI is not improperly altered or destroyed. Electronic measures must be put in place to confirm that e-PHI has not been improperly altered or destroyed	Ranger provides fine-grained access control across the complete platform, including elements like HDFS, Hive, HBase, Storm, Knox, Solr, Kafka and Yarn. Security administrators have the flexibility to define security policies for a database, table and column or a file, and administer permissions for specific LDAP based groups or individual users. With the introduction of dynamic tag-based policies enabled through the integration of Ranger with Atlas (data governance solution), users now have the flexibility to enforce both role-based (RBAC) and attribute-based (ABAC) access control. For further details, please refer to the section titled Dynamic Metadata-based Access Policies for Real-Time Access Control .
Person or Entity Authentication: Implement procedures to verify that a person or entity seeking access to e-PHI is the one claimed	CDP supports implementation of Kerberos for authentication within a cluster. Knox provides authentication for REST-based services and can be integrated with AD/LDAP or other authentication mechanisms to provide platform wide SSO.
Transmission Security: A covered entity must implement technical security measures that guard against unauthorized access to e-PHI that is being transmitted over an electronic network	CDP supports encrypting network traffic to provide privacy for data movement, as well as for data at rest. CDP supports wire encryption for all access protocols as well as for Solr, Kafka and YARN, with dynamic attributes such as geolocation, time and date to drive security policy decisions. CDP further supports the ability to encrypt files stored in the platform, including a Ranger-embedded open source key management store (KMS). Ranger provides security administrators with the ability to manage keys and authorization policies for KMS. With CDP, customers have the flexibility to leverage open source KMS or use enterprise wide KMS solutions provided by the partner ecosystem.

Dynamic Metadata-based Access Policies for Real-Time Access Control

Atlas provides data governance capabilities in the platform and is designed to exchange metadata both within and outside of the stack. By reconciling both logical data models and forensic events, enriched by business taxonomy metadata, Atlas enables a scalable set of core governance services. These services enable enterprises to effectively and efficiently address their compliance requirements by providing:

- Dataset exploration, curation and provenance through lineage
- Metadata-driven data access control
- Indexed and searchable centralized audit for operational events
- Comprehensive data lifecycle management from ingestion to disposition
- Metadata interchange with other metadata tools

By integrating Ranger with Atlas, Cloudera empowers enterprises to rationalize compliance policies at runtime. Now Atlas' data classification schemes can be leveraged by Ranger to enforce flexible attribute-based policies that prevent violations from occurring as mandated by HIPAA.

Ranger's centralized platform empowers data administrators to define security policies once based on Atlas metadata tags or attributes defined by a data steward or administrators, and let the platform (through SDX) apply the policies in real-time to the entire hierarchy of assets. Some data stewards can focus on data discovery and tagging while another group can manage compliance policy. This decoupling of explicit policies offers two important benefits:

- **Dynamic policy enforcement:** data analysis-driven tags can be enforced immediately
- **Reusability:** one policy can be applied to many assets, simplifying management

Ranger enforces both RBAC and ABAC to create a flexible security profile that meets the needs of data-driven enterprises. The initial set of policies being constructed within the community are defined as:

- **Attribute-based access controls:** For example, a column in a particular Hive table is marked with the metadata tag "PII" or "HIPAA-Sensitive." The same tag may be used in various other data sources, and the tag itself is used to assign permissions around data access (denied, permitted in original form, permitted in anonymized form, etc.) to different groups. This is an evolution from role-based entitlements, which require discrete and static one-to-one mappings.
- **Prevention of certain dataset combinations:** It's possible for two data sets—for example, one consisting of account numbers and the other of customer names—to be in compliance individually, but pose a violation if combined. Administrators can apply a metadata tag to both sets to prevent them from being combined, helping avoid such violations.
- **Time-based access policies:** Administrators can use metadata to define access according to time windows in order to enforce compliance with regulations such as SOX 90-day reporting rules.
- **Location-specific access policies:** Similar to time-based access policies, administrators can define entitlements differently by geography. For example, a U.S.-based user might be granted access to data while still in a domestic office. If this user travels to Switzerland and tries to access the same data, different geographical context would apply to trigger a different set of privacy rules to be evaluated and enforced.

These policies can be used in combination to create very sophisticated and dynamic security access policies for each user at any point in time and location, and are highly relevant for HIPAA compliance. Of course, the reach that Ranger provides in terms of authorization for the full spectrum of platform components allows organizations to consistently define and apply data access policies based on metadata regardless of the route by which the user or an application attempts to access the data itself.

About Cloudera
At Cloudera, we believe that data can make what is impossible today, possible tomorrow. We empower people to transform complex data into clear and actionable insights. Cloudera delivers an enterprise data cloud for any data, anywhere, from the Edge to AI. Powered by the relentless innovation of the open source community, Cloudera advances digital transformation for the world's largest enterprises.

Learn more at cloudera.com

Connect with Cloudera
About Cloudera:
cloudera.com/more/about.html

Read our VISION blog:
vision.cloudera.com
and Engineering blog:
blog.cloudera.com

Follow us on Twitter:
twitter.com/cloudera

Visit us on Facebook:
facebook.com/cloudera

See us on YouTube:
youtube.com/user/clouderahadoop

Join the Cloudera Community:
community.cloudera.com

Read about our customers' successes:
cloudera.com/more/customers.html

Physical Safeguards

These HIPAA safeguards are related to establishing policies and procedures that enable organizations to respond to an emergency that might damage electronic systems containing PHI such as fire, vandalism, system failure or natural disaster.

Cloudera's Replication Manager is part of CDP and plays a major role in addressing HIPAA physical safeguards as they relate to data replication, business continuity, and lineage tracing by deploying a framework for data management and processing. Both migration of analytical workloads, as well as planning for disaster recovery, means replicating data. Replication Manager moves data together with its security and governance policies, delivering:

- Complete backup and disaster recovery
- Migration from legacy Cloudera clusters or third party applications to CDP
- Migration of existing on-premises CDP workloads to cloud
- Flexibility to dynamically deliver data and workloads to hybrid cloud infrastructure
- Convenient creation of development and test systems

HIPAA REQUIREMENT	HOW CLOUDERA CAN HELP
Data Backup Plan: Establish and implement procedures to create and maintain retrievable exact copies of electronic protected health information	Replication Manager provides organizations with the capability to mirror data sources such as HDFS directories or Hive tables, and produce an exact copy of the data as well as keep both copies synchronized. Companies can also mirror between on-premises storage systems and public cloud infrastructures like Amazon S3 or Microsoft Azure.
Disaster Recovery Plan: Establish procedures to restore any loss of data	In addition to data replication, Replication Manager provides functionality to trigger processes for retry, and handle late data arrival logic. In addition, Replication Manager can mirror data sources like file systems or Hive and HCatalog on clusters using recipes that enable users to re-use complex workflows
Emergency Mode Operation Plan: Establish procedures to enable continuation of critical business processes related to protecting the security of electronic protected health information while operating in emergency mode	Replication Manager provides an alerting mechanism for a variety of backup and disaster recovery events to let administrators monitor the health of their data pipelines. All events are logged and administrators can view the information from the log or capture them using a custom interface. Each event logged provides the following information: • Date: Date of action • Action: Event name • Dimensions: List of name/value pairs of various attributes for a given action • Status: Result of the action • Time-taken: Time in nanoseconds for a given action to complete

Conclusion

To realize the full strategic benefits of Big Data without increasing the risk of compliance violations or data breaches, companies need to ensure that their analytic and data environments meet the strict requirements laid out in the HIPAA standard. The best way to accomplish this is through a holistic approach based on a platform that provides for all five essential pillars of security—administration, authentication/perimeter security, authorization, audit and data protection. With SDX and CDP, Cloudera provides the robust capabilities needed to address key HIPAA requirements and facilitates the creation and maintenance of a fully compliant modern data architecture. Hundreds of healthcare customers trust Cloudera to power their Big Data strategy. To learn more about how Cloudera can help you address key security and data protection challenges in your organization, please visit our website.

CLOUDERA

Cloudera, Inc. 395 Page Mill Road Palo Alto, CA 94306 USA cloudera.com