# HIPAA Compliant Cloud for Sensitive Health Data

Valentina Salapura
*IBM T. J. Watson Research Center, Yorktown Heights, NY, U.S.A.*

Keywords: Cloud Computing, Electronic Health Records (EHRs), PHI, HIPAA, Regulatory Compliance.

Abstract: Cloud environments offer flexibility, elasticity, and low cost compute infrastructure. Electronic health records (EHRs) require infrastructure which is regulated under several IT compliances with security and data persistence and restore. To enable customers to bring sensitive medical data in the cloud, we enabled the IBM Watson Health Cloud (WHC) for compliance with the U.S. federal electronic health record regulation. This paper briefly outlines how we create HIPAA- (Health Insurance Portability and Accountability Act) compliant cloud computing. We focus on the privacy and security rules for protecting Protected Health Information (PHI) and use data encryption for data-in-motion and data-at-rest. To meet HIPAA requirements for data persistence, we implement data back-ups, archiving service and disaster recovery plan. In this paper, we discuss various challenges and lessons learned for implementing the diverse set of compliance features required by HIPAA in the IBM WHC cloud.

## 1 INTRODUCTION

With cloud evolving from a compute-centric infrastructure into a data-and-compute infrastructure, more and more enterprises are moving into cloud platforms. These enterprises are bringing in terabytes of private data into the cloud platform in order to reduce the total cost of ownership of increasingly more demanding workloads. Such enterprise data include sensitive healthcare data and electronic health records (EHRs) that are governed under several IT regulatory compliance regimes such as HIPAA (U.S. Department of Health & Human Services, 1996), FERPA (US Department of Education, 1974), or GxP, to mention only a few.

Implementing regulatory compliance requirements in the cloud is not an easy task. It requires end-to-end compliance that puts additional requirements on the cloud provider. Compliance requirements range from defining processes for user onboarding (for a range of different users), user education and user access control, over data persistence and data restore measures to disaster recovery and to ensure business continuity, to implementing numerous IT security measures. With bringing highly sensitive data in the cloud, cloud computing security becomes more important than ever, as a single breach of such data can cause enormous damage.

HIPAA compliance regulation defines administrative controls, technical controls and physical controls, and policies and documentation. Administrative controls, for example, define how long and where HIPAA-related documents can be kept, how they can be accessed and modified, how information is shared, and how users of sensitive data are onboarded. Technical controls define how user activity is monitored and logged for auditing purposes, specify how user access is controlled, and how virus scans and password management are performed. It demands encryption of all sensitive data while these data are transferred via network (data-in-motion), or stored in computer memory and storage (data-at-rest). Physical control requires restrictions on physical access to servers and media where data is stored, as well as how these media are destroyed.

In this paper, we describe how we address regulatory compliance requirements for HIPAA, and implement them in IBM Watson Health Cloud (WHC). We are analysing security measures to further increase data security in the cloud to protect sensitive health data. We share our experience and give suggestions how to provide secure and regulated cloud infrastructure for hosting PHI data, and how to implement a HIPAA compliant cloud workload.



Salapura, V.
HIPAA Compliant Cloud for Sensitive Health Data.
DOI: 10.5220/0006356705960602
In *Proceedings of the 7th International Conference on Cloud Computing and Services Science (CLOSER 2017)*, pages 568-574
ISBN: 978-989-758-243-1

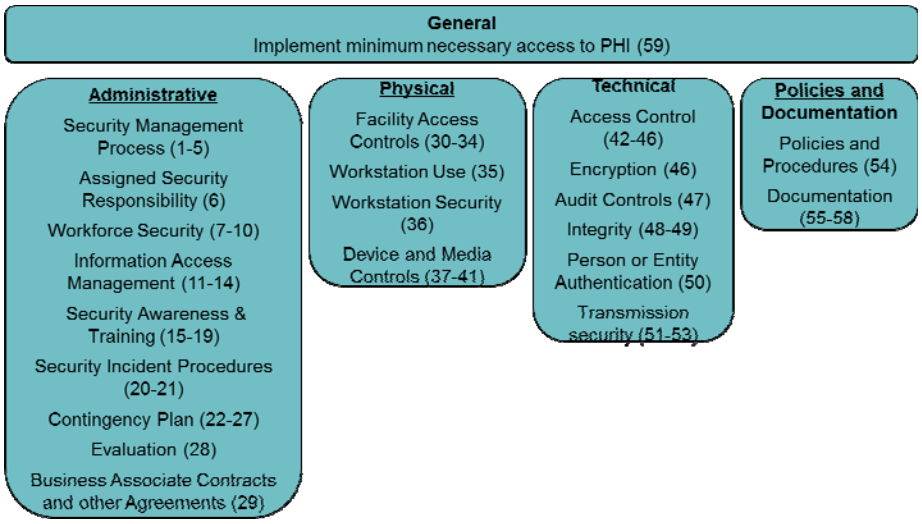


Figure 1: 59 HIPAA controls per category.

## 2 BACKGROUND

Cloud computing is being rapidly adopted across the IT industry to reduce the total cost of ownership of increasingly more demanding workloads. It is becoming the new de facto environment for many system deployments in a quest for more agile on-demand computing with lower total cost of ownership. Various enterprises, agencies and institutions are quickly adopting cloud computing which offers needed capacity and functionality without major investment in infrastructure, or operational expenses (Armbrust et al., 2010). To take advantage of cloud computing, companies handling electronic health records (EHR) demand to bring their business into the cloud.

However, applications handling EHR are constrained by federal regulatory legislation. Privacy and security of protected health information are regulated by the 1996 Health Insurance Portability and Accountability Act (HIPAA) to protect patient rights and allow them access to their records. The regulation requires administrative, physical, and technical protection measures to ensure the confidentiality, integrity, security, and access auditability of PHI data. Non-compliance with the regulation is sanctioned with penalties and fines. HIPAA requirements are defined in 59 controls, which are grouped into administrative controls, physical controls, technical controls, and policies and documentation, as presented in Fig. 1.

Handling sensitive medical data in the cloud requires adding IT management services such as user activity monitoring and auditing, data backup and restore methods, high availability and disaster recovery mechanisms, and methods for data protection, confidentiality, and integrity. Such features are typically offered by IT service providers in strategic outsourcing (SO) engagements, a business model in which the provider takes over several, or all aspects of management of a customer's data center resources, software assets, and processes.

This should be contrasted with unmanaged servers provisioned using basic Amazon Web Services (AWS) (Miller et al., 2010) and IBM's SoftLayer (SoftLayer) offerings, where the cloud provider offers automated server provisioning. To make servers managed, these cloud providers have partnered with other service providers that customers can engage to fill all the gaps up and down the stack. While this enables the user to add services to the provisioned server, the cloud provider assumes no responsibility for their upkeep or the additional services added. Therefore, it puts burden on the customer to obtain a fully managed solution for their sensitive workload rather than on the cloud service provider.

Several authors (Armbrust et al., 2010), (Schweitzer 2012), (Rodrigues et al., 2013) addressed the issue of bringing sensitive medical data into the cloud. These authors in general agree that the cloud computing offers to eHealth systems the opportunity to enhance their functionality and lower cost, but that it also implies risks in terms of the security and privacy of sensitive health records. The papers make suggestions for health care and cloud providers how to address security issues to

achieve acceptable privacy and security.

Another group of authors focus on implementing automated compliance management (Rodrigues et al., 2013), (Awad et al. 2015), (Khan and Bai, 2013). In (Awad et al. 2015), the architecture is proposed for business process compliance monitoring as a service during the run time. The authors of (Rodrigues et al., 2013) develop an approach to verify health regulatory compliance in the cloud to automatically check how the regulatory compliance is met. They represent health regulations as rules, and collect compliance related real-time data to verify the rules.

The IBM's Watson Health Cloud (WHC) (IBM Corporation) is a cloud offering that supports managed virtual and physical servers and provides a large number of managed services that are on par with the ones offered in high end SO contracts. It is built on IBM SoftLayer public cloud, and it implements managed services needed for regulated cloud environment to host PHI. Examples of such services are activity monitoring, security compliance, data encryption and key management, load-balancing, firewall services, resiliency, disaster recovery, and data backup and archiving. The installation, configuration, and run-time management of these services are fully automated and transparent to the user.

# 3 DATA LOCATION

Some of the biggest obstacles in ensuring regulatory compliance in a cloud computing environment are data location and user data isolation. While the use of private and dedicated data centers allows the detailed knowledge and control of computing environment, location of the data, and data protection method used, this control is not available for many cloud computing services.

By nature, cloud computing does not provide transparent information about the actual location of data, and where the computation is physically located. Thus, an application can be spread across multiple physical servers in a data center processing sensitive PHI data. This situation makes it difficult to make sure that sufficient data protection is in place to meet legal and regulatory compliance requirements.

To ensure isolation of user processing and sensitive data, dedicated physical servers can be used, and run as bare metal servers. As an alternative, to ensure isolation, private virtual instances can be used. These instances use virtual machines which are hosted on dedicated physical servers. The hosting servers are dedicated to a single account to ensure user and data isolation. IBM's SoftLayer cloud supports both these modes of operation.

Additionally, for some countries data cannot cross international borders. For the United States, regulated protected data has to physically stay within the legal jurisdiction (i.e., within U.S.). In these cases, a cloud provider has to ensure that data stays within the required geography, which limits how cross-data center data replication for disaster recovery or distributed computing is implemented.

# 4 DATA SECURITY RULES

The WHC is designed to host sensitive protected data and PHI as defined by HIPAA regulation in a secure and compliant manner. The methods for data protection system and processes include network security, data encryption for data-in-motion and data-at-rest, encryption key management, data confidentiality, data integrity, and user activity monitoring.

WHC data protection system uses state-of-the art technologies for protection from both external and internal threats. Moreover, a threat monitoring service monitors accesses to the data storage systems as well as to the external service interfaces.

## 4.1 Data Integrity

A compliant cloud solution has to implement data integrity measures to ensure that PHI data is not altered or destroyed in an unauthorized manner. Data integrity refers to the accuracy and consistency of data stored in a database, data warehouse, data mart or other data formats.

To assure data integrity of data-in-motion, WHC employs SSL/TLS with hashes and HMACs. The integrity of data-at-rest can be ensured by using digital signature schemes and message authentication codes HMACs. Each use of data leads to verification of integrity in cryptographic manner immediately after it is decrypted. Each new data object is processed to compute its digital signature/HMAC prior to its encryption and storage. The integrity codes are stored along with the data.

## 4.2 Data Confidentiality

To ensure data confidentiality, PHI data are

encrypted for both data-at-rest and during transmission over public networks.

Data-in-motion: When data is transmitted from an external client to an internal WHC service, two steps are carried out: the data is encrypted using a public key, and then the cipher-text of the data is transmitted over an encrypted channel using SSL/TLS. Such dual-encryption offers stronger protection especially if SSL/TLS has weaknesses, or when the key generated by the client device is not cryptographically strong.

Data-at-rest: Data-at-rest has to be protected on both externally accessible systems and on internally accessible systems such as local storage. Externally accessible systems, such as Object store and Cloudant, accept data in encrypted form over an encrypted channel. The storage service stores the data in encrypted form without decrypting them. Data-at-rest on internally accessible systems remains always encrypted. Examples of such systems is storage attached to VMs, and backup storage.

# 5 DATA PERSISTENCE AND DISASTER RECOVERY PLAN

To be able to restore data to a prior valid value, HIPAA demands implementation of data back-up. Compliant workloads must have a contingency plan to protect data in case of an emergency, and must create and maintain retrievable exact copies of electronic PHI.

SoftLayer offers several backup solutions. The cloud storage-based solutions offer storage that can persist independently from the workload instance that created it, the virtual machine where it ran, or even the whole data center. To align with HIPAA guidelines, WHC uses automated data snapshots to store data in cloud storage for long-term durability. Backed up data is replicated to a distinct location to provide isolation from failures of a single site. The backup can be accessed at any time.

To ensure compliance with controlled data location, data must not cross international borders. WHC uses SoftLayer data centers within U.S. as target location for cloud storage for data backup. At any point in time, multiple copies of the PHI data exist geographically distributed.

Some regulations demand that data is retained over the long term, even after the contract between a cloud provider and a user is terminated. Such a regulation is FDA Title 21 CFR Part 11: Electronic Records; Electronic Signatures. If PHI data are required to be kept for a large number of years, an archive solution for data should be considered. Archiving services are implemented by using lower tier storage, which is of lower cost than active storage or backup storage. Archived data has to be encrypted. In addition to data, the encryption keys have to be preserved, in order to be able to decrypt the data after a number of years. In case of auditing, archived data has to be retrieved as well as their logs documenting data modification over time.

A disaster recovery plan is required to protect PHI data and infrastructure in times of a disaster. A disaster is considered any unforeseen event, which directly or indirectly impacts systems availability beyond acceptable service levels that would result in the decision to recover the systems at an alternate site. Disaster includes situations in which the systems are unusable or inaccessible.

WHC implements a disaster recovery plan for workloads and PHI, as required by HIPAA. This involves maintaining highly available systems, and keeping both the data and system replicated off-site.

In the cloud environment, virtual server instances can be started quickly, and the whole system can be rebuilt quickly. A faster, more expensive recovery solution implements stand by instances similar to the primary site on a distinct site to create geographically diverse, fault tolerant systems that are highly resilient in the event of natural disasters. WHC performs periodic DR tests and includes various hosted workloads.

An alternative approach to DR is to implement a distributed computing environment, where applications are straddled across multiple data centers. This architecture requires continuous data replication between the data centers, and global network balancing which is session aware to ensure that all requests are distributed in accordance to session affinity. While this architecture is more complex to implement, it eliminates the need to implement a disaster recovery plan. Instead, disasters are avoided in a multi-active setup.

# 6 USER CONTROLS

To protect PHI as defined by HIPAA regulation in a secure and compliant manner, the cloud environment has to implement user access control, authorization and monitoring processes, and to ensure that users complete needed training.

### 6.1 User Access Control

WHC implements measures for fine-granular ID management, authorization and access control policies in the system to ensure that only authorized users can access systems.

Role-based access control: WHC defines multiple roles, and implements appropriate role and ID lifecycle management policies to create, activate, suspend, de-activate, and delete user IDs and roles.

Approvals: Obtaining access privileges requires approvals to confirm the business need for accessing the environment, as well as completing HIPAA training. The user access rights are renewed yearly.

Separation-of-duties: The access control policy ensures that systems hosting PHI data have separate admins than systems hosting keys and credentials. Additionally, hosted applications can implement additional application-level user roles, and manage their access at the application level.

Automatic log off: The access to the system is automatically disconnected after predetermined time of inactivity at all system levels - at the OS, hypervisor and at the application levels. The time of inactivity before disconnecting a user is typically set to 15 mins or 30 mins for all user roles.

### 6.2 User on Boarding and Training

The team members working in a regulated environment have to complete a HIPAA on-boarding process. This typically encompasses training to familiarize the users with PHI data and significance of protecting them. All users who might get in touch with PHI data need to complete HIPAA training which must be renewed annually. The proofs of training completion are to be kept for six years for auditing purposes. Similarly, the regulation defines off-boarding process for users leaving a HIPAA regulated project.

The proof collection and approval process is automated. Upon starting the on-boarding process, the information is automatically sent to the relevant parties, such as training links, or the approval links. The created approvals and training certificates are collected, and once all prerequisites are completed, the access to the system is granted. The system also implements yearly re-validation process.

### 6.3 User Activity Monitoring

When dealing with PHI, HIPAA requires that privileged user access and activity is monitored. Thus, we had to ensure that access are collected and retained for 270 days, as required by HIPAA.

All solutions running on WHC are included in the activity monitoring. Privileged user activities are monitored at the OS, hypervisor and network level. Each access to a system or query to retrieve PHI as well as non-PHI data is logged with details of the event – user ID, client IP address, time, type of query and so on. Such logs are forwarded to monitoring systems that analyse them in real time, and provide threat monitoring intelligence. The audit and security experts investigate anomalous and unauthorized events.

All activity logs are required to be retained for extended periods of time in case that an incident needs to be investigated. WHC provides user access monitoring logging and retention. Hosted workloads use this log retention system, but they must generate their own user activity logs for auditing purposes.

## 7 MORE RULES TO ENSURE

Additionally, physical control, policies, and more administrative controls have to be in place, which we briefly cover below.

### 7.1 Administrative and Legal Controls

Various regulations require respecting administrative and legal requirements and defined and described processes. For example, a HIPAA regulated project requires to have a person assigned with security responsibility who is responsible to ensure that all controls as specified by HIPAA regulation are covered. This person is approved by a project executive, and his/her name is communicated to the team.

The specified processes have to be documented. Documentation contains information about policies used, training proofs, and access approvals, and are retained for the required time, typically a number of years. The HIPAA regulation prescribes that this information is shared with the team, which can be done by using a wiki or a community web page.

From the legal point, all people who can access PHI data need to be covered by an employment or business associate contracts. Alternatively, other arrangements for end users and business associates need to be in place which will state that PHI information will be appropriately safeguarded, and that appropriate sanction policies are in place if he does not.

### 7.2 Physical Controls

HIPAA requires also several physical controls regulating facility access and management. These regulations define that only authorized users can access the physical servers in a cloud data center. Additionally, disposing of hard disks which were used for storing PHI data has to be controlled. Not only hard disks have to be cleared, but they also have to be physically destroyed. These physical aspects of ensuring compliance are carried out by the cloud provider. In our case, we rely on IBM SoftLayer HIPAA compliance (IBM Cloud Softlayer) to ensure implementing these controls when bringing sensitive patient records into the cloud.

## 8 LESSONS LEARNED

During the work on implementing HIPAA regulations for WHC, several points became apparent. One insight is that typical failures when hosting sensitive medical data from a non-cloud deployment are also relevant in the cloud. An example is a failure to implement timeout of a session after 30 min of inactivity, or incorrectly set permissions for a user that made data improperly visible. Fortunately, standard protection methods, such as data encryption or encrypted message transmission, are equally effective in the cloud. While the security threats are similar to those in a large data center, cloud environment imposes separation of responsibilities for security protection between a cloud provider and application developers. The cloud provider is responsible for the physical security and the cloud user for the application-level security.

Another observation was that bringing sensitive medical data into the cloud requires modification of how cloud resources are used. Some features which are guaranteed in a private data center where an administrator has full control of all resources, require special handling in the cloud. For example, WHC uses private virtual deployment where only WHC virtual machines are hosted on dedicated physical servers. Additionally, location of the PHI data in the cloud has to be controlled to ensure compliance with the regulation.

Implementing a HIPAA compliant cloud demands data persistence and business continuity techniques which increase cost. We implemented data backup, created a disaster recovery plan, and enabled data replication between multiple sites. While these measures ensure that data can be restored after an unforeseen event, or after accidental or malicious deletion or corruption of data, they increase cost of operation, and introduce complexity into the management of the cloud.

Auditability of a HIPAA regulated cloud includes a large amount of logging, tracking and persisting access logs. Providing the auditing service to track logs in centralized location is a more trusted solution than logging built within an application.

## 9 CONCLUSIONS

In this paper, we describe how we implemented a diverse set of compliance features required by HIPAA to enable bringing sensitive medical data in the IBM WHC cloud. The WHC cloud provides a dedicated, secure and regulated cloud infrastructure for hosting PHI data.

HIPAA demands conducting risk analysis, specifying security policies and incident procedures, limiting access to PHI data, servers and storage, encrypting data at rest and in motion, ensure data integrity, monitor system activity, and manage user identity. We implement security measures such as firewalls, intrusion prevention systems, anti-virus software, encryption, activity monitoring, identity and access management.

A cloud-specific multitenancy allows multiple customers to share compute and storage infrastructure. We isolate customers at the physical server level, control data location, and ensure data encryption and encryption key management. To meet HIPAA requirements for auditing and data persistence, we implement data back-up, disaster recovery plan, and an auditing system.

## REFERENCES

U.S. Department of Health & Human Services, 1996. “Health Insurance Portability and Accountability Act” HIPAA. [Online]. Available: https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996.

US Department of Education, 1974. “Family Educational Rights and Privacy Act (FERPA),” 1974. [Online]. Available: https://www2.ed.gov/policy/gen/guid/fpco./ferpa/index.html?src=rn.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M., 2010. “A View of Cloud Computing,” Communications of the ACM, vol. 53, no. 4, 2010.

Miller, F. P., Vandome, A. F., and McBrewster, J., 2010. Amazon Web Services. Alpha Press.
SoftLayer. [Online]. Available: http://www.softlayer.com/
Schweitzer, E. J., 2012. Reconciliation of the cloud computing model with US federal electronic health record regulations. Journal of the American Medical Informatics Association vol. 19 no. 2.
Rodrigues, J. J., de la Torre, I., Fernández, G., López-Coronado, M., 2013. "Analysis of the Security and Privacy Requirements of Cloud-Based Electronic Health Records Systems," Journal of Meical Internet Research, vol. 15, no. 8.
Awad, A., Sakr, S., Elgammal, A., 2015. "Compliance Monitoring as a Service: Requirements, Architecture and Implementation," International Conference on Cloud Computing (ICCC).
Khan, K. M., Bai, Y., 2013. "Automatic Verification of Health Regulatory Compliance in Cloud Computing," 15th International Conference on e-Health Networking, Applications & Services, IEEE.
IBM Corporation, IBM Watson Health Cloud for Life Sciences Compliance. [Online]. Available: https://www.ibm.com/us-en/marketplace/cloud-for-regulated-workloads.
IBM Cloud Softlayer, "Compliance." [Online]. Available: http://www.softlayer.com/compliance.