Deep Science Publishing
https://doi.org/10.70593/978-81-988918-5-3_4

DeepScience
Open Access Books

Chapter 4: Architecting cloud-based infrastructure for scalable and secure health data storage

4.1. Introduction

The protection of electronic health (eHealth) data is of utmost importance for individual patients and society as a whole. eHealth data systems must be designed in a way that data privacy, confidentiality, and integrity are preserved. By considering patient data as sensitive data and applying the data protection principles at a very early stage of the technology life cycle, privacy and confidentiality will be ensured throughout the data storage and processing. The most sustainable way of achieving these ends is by means of safe and sustainable digital infrastructures that allow individual patients to control the lifecycle of their own patient data, and that is particularly important for sensitive or classified data. Current eHealth applications are rarely equipped with effective data protection mechanisms. The desired result is that individuals control their own sensitive data through trust-building relationships. Achieving this sustainable form of safe and secure health data storage cannot be done by mere statements about data protection. Instead, we need to build secure and sustainable cloud-based digital infrastructures complying with data protection and cyber security requirements (Murdoch & Detsky, 2013; Kankanhalli et al., 2016; Kruse et al., 2016).

We present secure and sustainable cloud-based digital infrastructures by introducing a novel architecture for eHealth data storage: ArchiDelta. It relies on a micro service architecture using storage micro services for the decentralized storage and the encrypted versioning of health data under the complete control of the individual patient's process in the desired manner, but without any direct involvement in the implementation of the respective data security measures, enabling the patients to define and implement processing protocols controlling the complete lifecycle of the related patient data. The combination of DeltaLake, IPFS, and Docker provides the technological basis on which new eHealth solutions can emerge, using the collaboration and the feedback of application developers and individual patients as well as other investors. In this way, ArchiDelta supports the co-creation of data protection-enhanced eHealth services.The responsibility of security depends on the service model chosen for deploying cloud resources and the data type being processed. Business associate agreements with cloud vendors define, comprehend, and clarify responsibilities regarding security considerations. It is possible to have security features built on top of cloud storage services; however, customers are responsible for configuring and managing those security features. Thus, security information should also be considered regarding data integrity and availability, as data loss is imminent at any level. Service-level agreements with vendors provide you with the guarantee of uptime, but not data backups needed for sensitivity (Wang, Alexander, & Lee, 2016; Wang, Kung, & Byrd, 2018).

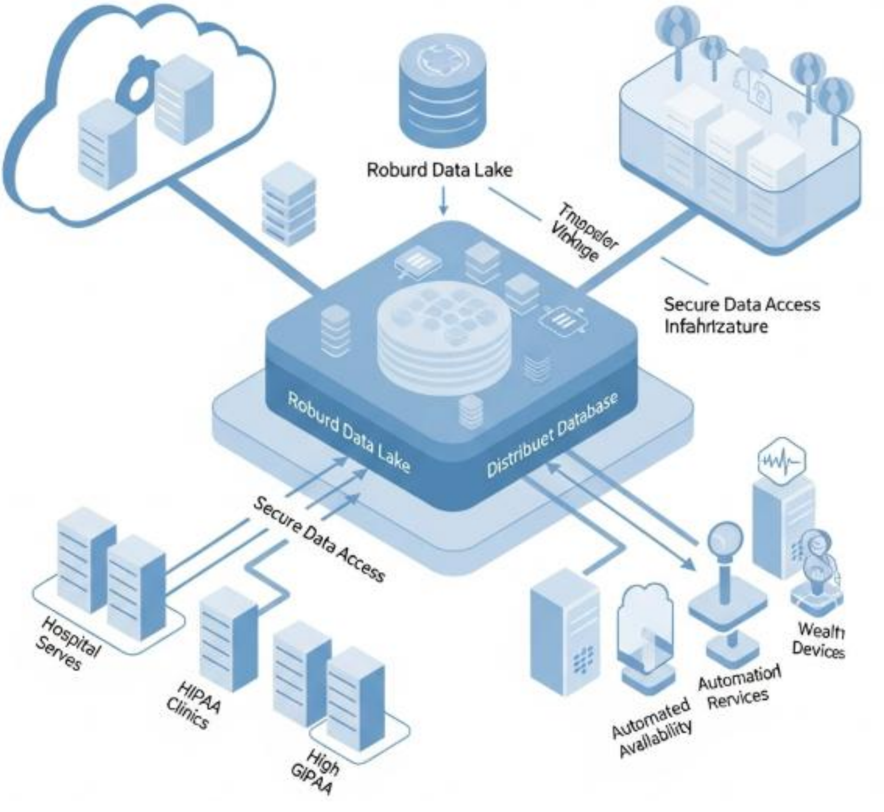

Fig 4.1: Architecting Cloud-Based Infrastructure for Scalable and Secure Health Data Storage

4.1.1. Background and Significance

Health data are sensitive due to protected health information (PHI) and personally identifiable information (PII). Storage and transmission of PHI should follow security and privacy policies and procedures. However, many healthcare organizations still face a challenge in achieving compliance. Ensuring compliance with internal policies and procedures, related state laws, and Privacy and Security Rules protects health data from unauthorized access and misuse. There are additional requirements to consider with compliance during business associate agreement negotiations for cloud storage.

While data transmissions between on-premises and cloud environments can be secured and storage access with secure passwords and permissions can enable the right role-based access, encryption at the application and storage levels can address confidentiality at the source and storage, respectively. Cloud storage vulnerability still exists with the application APIs if data is unencrypted and it relies on security by obscurity.

4.2. Cloud Computing Fundamentals

Cloud computing appears to be the next generation of computing, presenting new procedures, models, architectures, and technologies dealing with managed and relatively cheap on-demand access to the resource, such as applications, storage, servers, networks, platforms, and processes, over the Internet. It aims to facilitate users to store, share, and manipulate their data on the cloud rather than their local machines. Since then, cloud computing has garnered impacts for a wide range of consumers and enterprises alike, leading to the rapid growth of innovative services and business models. With its quick scalability and flexibility, and pay-as-you-go pricing, cloud technology has been widely recognized by both the information technology (IT) industry and enterprises.

Cloud computing technology indeed provides a fast, effective, and economical way for enterprises to accomplish their IT goals. With cloud services readily available, enterprises are free from the burdens of developing, implementing, and managing their own expensive IT system infrastructure, optimization of application performance, and predictable enterprise budget expenditure. Cloud computing is an abstracted, virtualized resource pool residing on top of a farm of physically distributed data centers. By means of virtualization, management overhead can be simplified and resource utilization improved. Cloud offers numerous advantages, the principal one being economic. In cloud services, clients pay for resources on-demand and only for what they use, thus reducing CAPEX on hardware and OPEX on the running of the hardware.

4.2.1. Types of Cloud Services

The most common cloud services are software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS). Software as a service is a model where users access the software on demand from the vendor. The vendor is responsible for the management and security of the software. The service is usually charged based on a subscription model, with the SaaS vendor handling all these tasks and providing just the service the user needs. Platform as a service provides a cloud-based environment for users to develop, deploy, and manage applications. PaaS allows users to create applications without worrying about having the necessary infrastructure for development locations or application management. PaaS provides services for the entire ASP lifecycle: applications design, development, testing, deployment, and hosting. Infrastructure as a service is a model where users outsource the entire management, maintenance, and ownership of IT infrastructure to a vendor that provides virtualized resources.

4.2.2. Definition and Key Concepts

Cloud computing plays a crucial role in providing Infrastructure as a Service (IaaS); that is, resources offered as services over the Internet. Such resources may include servers, storage spaces, and computing power. These resources are critical assets for many organizations creating or using Digital Health platforms, given the considerable amount of Digital Health data involved. Consequently, organizations depend on third-party IaaS providers to support their infrastructures. The attraction of IaaS lies in its potentially considerable reduction in cost, compared to owning and operating a large physical infrastructure. The prices come down with an IaaS provider's scale-effect and ability to negotiate low unit-power costs with energy providers. Moreover, the conversion of Capex into Opex means improved cash flow and accounting. IaaS is appealing also because it involves lower operational complexity and risk. With IaaS, the burden of maintaining and securing the computing equipment and physical facilities moves from the resource user to the IaaS provider. Because of the global availability of Infrastructures, advanced technical capabilities, and experience in resource provisioning among IaaS providers, Digital Health infrastructure becomes more available than what a single organization could offer.

Cloud computing involves the use of networking to enable programs and services to be accessed via the Internet. In a cloud-based model, an organization's computing infrastructure becomes a resource that can be easily resized or reconfigured according to its specific needs.

4.2.3. Types of Cloud Services

A main consideration when architecting cloud-based infrastructure for secure and scalable health data storage is the specific services offered by cloud providers and how these fit in our overall architecture. Customers are presented with many options from cloud providers, and these services can be classified broadly into three categories — Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS). SaaS is the most restrictive — the cloud provider manages everything for you, and the customer has little control over architectural design decisions — but it is also the most convenient and least expensive. PaaS services offer a more fine-grained level of control over design decisions but also require the customer to do more work, managing application dependencies and possibly even the design and structure of the underlying database. IaaS is the most flexible, allowing customers to provision virtually any type of virtualized resource, but it also requires the customer to do the most work including managing all scaling and redundancy issues. The more control the customer has, the more configuration, management, maintenance, and development work becomes the responsibility of that customer.

Choosing the correct type of service for components of the application being developed can have a significant impact on the speed of deployment, resources required, and reliability of systems utilizing cloud resources. In the case of the cloud-based health data infrastructure example given, we have chosen to primarily utilize IaaS resources with the aim of having maximum control during the development and launch phases of the product, but we also heavily utilize multiple SaaS services, such as authentication and user management, to minimize the amount of time dedicated to building small systems that are not core functionalities of our product. The web application framework we have chosen to utilize also offers a PaaS option for deployment; utilizing these PaaS resources would allow for a very quick launch timeline and minimal management while still allowing us to focus on the specific features of the product that we are building.

4.3. Health Data Storage Requirements

Health data storage needs to fulfill several requirements that arise from the nature of health data. In particular, these data are often sensitive and confidential in nature. Storage design also needs to comply with several statutory requirements, which to a great extent focus on guaranteeing the safety, integrity, and confidentiality of health data.

Regulatory Compliance

The growing adoption of cloud computing in national health systems, and the continuous growth in the volume of data collected and stored, raise several concerns regarding the compliance of cloud services with privacy regulations and their technical provisions. Several national health authorities and third parties have developed directives concerning the purchase, delivery, and use of cloud services by patient data processors and controllers.

Among privacy regulations, the General Data Protection Regulation applies to all EU member states. Introduced in 2018, it adopts technological neutrality as a guiding principle, meaning cloud computing specifically is not mentioned, but big data processing is. This way the processing of large-scale personal data is strictly regulated.

Compliance with Code of Conducts on GDPR Internal Onboarding or with technical norm implementations defined in Article 25 may also play a pivotal role. Such documents revolve around security measures, breaches, anonymization techniques, and other recommendations.

Health data storage from cloud services are regulated and ensured in cases where either the health data owner, the research center or both use cloud services that are compliant with the required compliance statement. Data processing of the principal investigator, his team, or the activities of the utilizing cloud service provider while storing and hosting health data are subjected to compliance with the regulatory rules. Nevertheless, the consideration of regulatory compliance is usually focused on the system utilized to store health data without paying attention to all other persons and institutions involved in the project and having access to the cloud service.

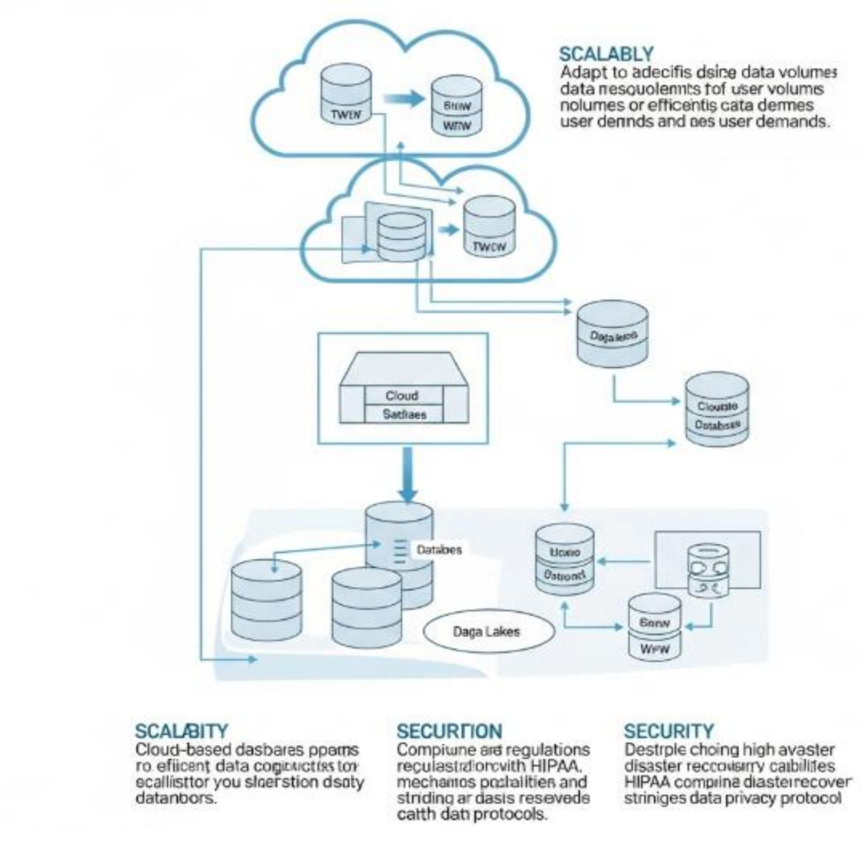
TWEN
Bmw
WrIW
SCALABLY
Adapt to adecifis dsire data volumes data nesquolennts fof user volums nolumes or efficentis cata demes user dennds and oes user demands.
TWCW
Dataleso
Cloud
Satliaes
Cloutde
Databses
Databes
Uoxo
Ostrond
Daga Lakes
Bnnw
WPW
SCALABITY
Cloud-based dasbares ppems ro efiicent data cogiructies to ecallisttor you slaerstion dsaty datanbors.
SECURITION
Compiune and regulations reculastrdforcvith HIPAA. mechanos podalities and striding ar dasis resevede catth dath protocols.
SECURITY
Destrple choing high avaster disaster reccoviry caibilites HIPAA compina diasteirecover striniges data privacy protocol

Fig 4.2: Health Data Storage Requirements

4.3.1. Regulatory Compliance

Storing and managing health data on cloud-computing-based health systems is a huge responsibility of owners of such systems. Such data should be secured and maintained in accordance with both regulatory and industry standards. Public cloud service providers do publicly advertise their compliance with certain top regulatory standards but storage and management of health data on such services do not make the whole service automatically compliant with such regulations. Hence, utilizing a public cloud-based solution to store or share health data for hybrid or multi-institutional research does not automatically provide regulatory compliance of such a research project. The cloud-based service utilized for data storage evolution for a certain research project should be supported with both adequate technical implementations and documented data handling procedures and practices. Utilization of cloud-based services, even from the same public service provider that offers compliance with certain regulations, is not a trivial task for health data owners and is facing data sharing issues, especially in multi-organizational research projects. Each institution and their reply on the data owner, usually the principal investigator, should ensure regulatory compliance while performing joint research that involves not only health data but also cloud computing services used for data storage and processing. These considerations act as a major roadblock for the evolution of inter-institutional research, and the only possible solution to resolve such problems is to carefully document the research plan, address how all regulatory requirements are managed, and put this described data handling plan in an agreement that has to be signed by all parties involved in the research project.

4.3.2. Data Privacy and Security

In the past decade, data security has found its way to the forefront of the internal and external mindscape of organizations within, and outside the healthcare charter. As the omnipresence of data has manifested in organizations harvesting any and all data streams with an outward-facing funnel, the sheer value of these data streams have led to an impetus for tighter controls from a legal and regulatory side. Collection of user data on the Internet, in the form of advertising cookies, has long since been a point of contention. Recently however, the flow has been reversed, and regulators have taken a stance of preventively activating control measures that automatically limit the amount of data that can be collected, due to privacy concerns. The introduction of regulations in Europe made companies actively check for the legitimacy of data collection if their user base touched European shores. This legal stance against the often-abused free-for-all on the Internet has now found its way into other regions. The California Consumer Privacy Act has become the de facto gold standard as its key components are now taken by copycat states, and introduced the concept of a digital rights toolkit for users. Recently, we have also seen the introduction of big-tech monopoly deactivation attempts on both sides of the big pond in the form of regulations.

As the recent breach of security around the use of health data through the pandemic has demonstrated, thievery and hijacking of sensitive health data has also found a very prominent and rich vein for cybercriminals through the use of illegal hacks or exploits. Organizations within the healthcare sector that store and analyze this data have had to deal with far-reaching consequences as many of the current security encryptions were inadequate or outdated.

4.4. Designing Scalable Cloud Architecture

Scalability is one of the main attributes for any cloud service provided by a cloud provider and a responsible solution architect must design and implement a highly scalable and reliable cloud based solution. Cloud provides a large amount of resources; however, traditional monolithic solutions may not be able to take advantage from the cloud scalability. Moreover, realization of a planned scalability may not be straightforward at all times. Thus, scalable cloud resources must be strategically designed into solution architecture. Solution design must be made so that components exposed publicly may be scaled easily and have availability triggers to provision large instances behind them and additional nodes of backend infrastructures with a trigger based on defined thresholds. Normally, the backend components are the ones employed for processing information workloads. Data persistence backends such as databases and storage components may become bottlenecks if not designed correctly from the onset. For databases, in particular, certain design aspects must be taken into account such as partitioning strategy, size and layout of the tables, and frequency of backups during design phase. The choice of caching technology is also important. Caching removed the bottleneck problem at the backend, but the wrong choice may create additional scaling problems at the caching layer. Decision is to either prefer no caching and move the load at the data persistence layer with automatic triggers to scale up the resources or investing in additional time for defining caching strategy upfront.

Microservices Architecture

Traditional monolithic applications must be re-architectured into cloud native microservices that provide a horizontal scaling capability. For services that are exposed via public networks, scaling is mostly automatically triggered based on thresholds defined. For scaling backend services, the extraction of process workloads must trigger additional instance creations. In addition, careful design must minimize backend communication. Microservices should more or less share the same data persistence models both for efficiency and fast throughput.

4.4.1. Microservices Architecture

Cloud computing enables explosive growth in the volume of data generated by health care. Maintaining the system infrastructure that ingests, processes, stores, and makes the health data available for facility and patient-centric operations usually requires the data center owners to maintain a huge Virtual Machines (VMs) based Infrastructure. Microservices architecture is an effective alternative topology of the design that allows cloud data and service providers to effectively build, secure, scale, and manage the data services required during facility-based operations and real-time patient healthcare. The microservices architecture provides a way of developing and running enterprise-class applications that are made up of heterogeneous applications, communicating through modern protocols and able to perform and scale the services they support independently of one another. This allows business and data service providers to reduce the costs of the service infrastructure while increasing the operational control and enable patient-centric services by providing the Health Data APIs and Micro DaaS functionalities for real-time operations.

The Microservices architecture is based on the idea of decomposing large monolithic applications into independent services usually focused on a specific business capability and interacting through an event-based mechanism or API mechanisms. The microservices architecture supports the agile and container deployment model, API configuration, Dev Ops, and performance-oriented microservice development methods performed by a small team of developers. The massive reduction in operational load for intra business IT infrastructure makes the architecture popular with the enterprises. The architecture supports the agile development process but requires a service-oriented communications mechanism using lightweight protocols over HTTP for communication through traversing the OSI-1 and OSI-2 stacks.

4.4.2. Load Balancing Techniques

Load balancing is a very important consideration when designing cloud architecture hosting scannable enterprise applications. As with most infrastructural overhead, this should be more of an abstract concern at higher levels of the hierarchy. However, with the rapidly changing technology landscape, there are becoming higher levels of concerns.

Static Load Balancing

Static load balancing tries to balance loads at assignment time based on knowledge of future requests. Static load distribution techniques usually rely on statistical information relating to the expected requests. The distributions may apply to simple parameters such as total request volume or to more complex parameters, such as relative proportion of request volumes for different service types, predictive of execution times. This information may be very complicated and difficult to collect.

Dynamic Load Balancing

Dynamic load balancing considers all requests at the time of incomings of each request and hence allows all requests to be assigned to an appropriate processor such as load balance at the time of assignment. Cost for load assignment by predicting execution times in relative load, pertaining not directly to load computation, but indirectly by doing so on response times, may be considerable.

Load balancing options within a working cluster or cloud affect performance and power, trailing implications with redundant servers. Costs of solutions vary greatly depending on volume, peak loads, and service. Simple static configuration is always available. Completely automatic, dynamic solutions offer great simplicity and economy when volumes do not justify a more robust tooling requirement. Manual adjustment to simple services is normally sufficient to handle load, scaling up and down by hand, tools to provide more queues or smaller target wait, or refresh rates of the complexity.

4.5. Security Considerations in Cloud Storage

The cloud paradigm introduces a number of new security risks that are rarely present in traditional IT environments. When choosing an infrastructure or platform vendor, some level of trust must be placed in the vendor to handle potential insider threats and protect

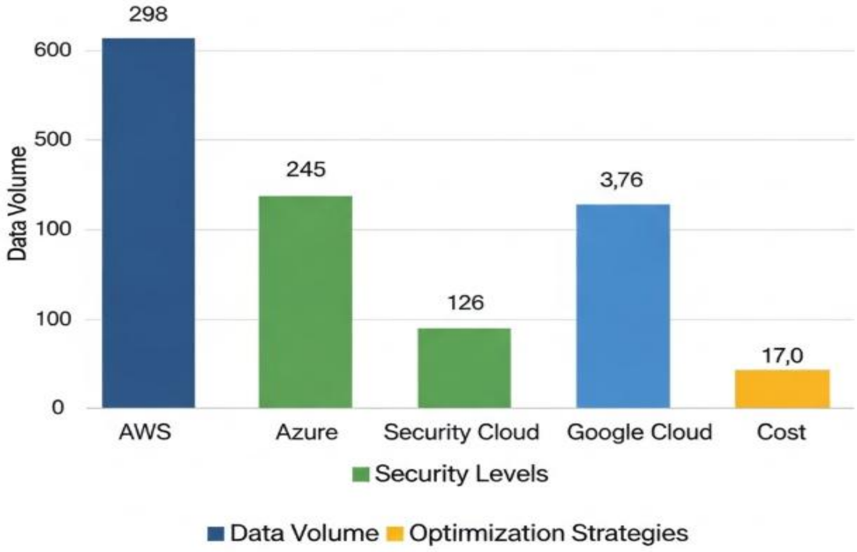

Fig : Cloud-Based Infrastructure for Scalable and Secure Health Data Storage both the physical and virtual assets that comprise the cloud service. Trust is further extended when the cloud service deals with sensitive data such as health records. For added protection when dealing with insecure environments, security best practices must be followed, such as encryption and controlling access to sensitive data, which is usually accomplished through user authentication and access controls. Additional measures should be used if they are supported by the vendor.

Security of a cloud storage system is often reliant on the mechanisms used to protect the data. Traditional data protection methods that are often ignored in the cloud paradigm are data encryption and control of data access. Most cloud storage solutions are designed to make it easy for a large number of users to share and view data.

4.5.1. Encryption Practices

The security problem is manifested through the majority of attacks on sensitive health data. Security is normally achieved through a combination of technical and procedural approaches. The most common technical approach for health data is encryption. It provides protection to sensitive health data over the cloud storage and also during the transmission phase. Moreover, encryption provides security to sensitive health data, not only from eavesdroppers but also from any unauthorized access. Usually symmetric encryption is required for real time sensitive applications however with the advent of large storage overhead and computation capabilities asymmetric encryption is becoming more and more useful. There are several encryption models present for health data.

An elliptic curve based encryption system for women's health records combines the advantages of elliptic curve encryption with the advantages of facial recognition for user authentication. After carefully analyzing the work we can see that this encryption system is not sufficient for cloud health data storage. Although this encryption system is providing confidentiality and security it is not solving the problem of efficient user access when dealing with a large number of users. In addition to this only patient data is secured, not cloud service data. The main design strength is evident when we consider the work from a GUI-based approach in the context of wearable health data from patients. A novel key-managed encryption model to share cloud data securely is also presented.

4.5.2. Access Control Mechanisms

A second major class of cloud storage-oriented security concerns relates to the governance of access to sensitive health information and the trustworthiness of those who are granted access. In particular, the cloud storage service must allow for the creation of fine-grained access control lists that specify allowed operations for each entry in the data store. Use of versatility in this aspect is key because, for example, while it is generally appropriate to restrict reading access to all patient data to the patient in question as well as a few other clinical staff members, the security considerations for which specific clinical staff members are permitted to enter new information for which patients is very likely to differ between public community health data and patient data coming from an individual's engagement with a health system. The range of potential differences in use cases argued by such clinical collaborators and researchers is fairly diverse and applies both to their collective interests in looking at patient data as well as their interests in who else at those organizations could be entrusted with such governance.

The cloud storage service's access control implementation should also impose constraints on reads of summarized or de-identified information. At a minimum, enforced de-identification should vary appropriately based on the size of the summary dataset, which could either be defined as the entire patient cohort for a clinical study or similar to summary tasks that are central to the building block described.

4.6. Data Backup and Disaster Recovery

Organizations must have an ongoing plan for data backup and disaster recovery to ensure quick restoration of health data in the event of unintentional loss or damage. Data loss can occur for many different reasons including equipment failure, environmental hazards, software failure, cyber attacks, or human error; and with so much important health information stored digitally today, loss of this data can come at great expense. Furthermore, the associated costs with not being able to access health information can adversely affect overall patient care and safety. A secure work offsite and a clear, accurate process of restoring IT operations can assist organizations with reducing this data loss risk. Cloud computing offers a viable option for health data backup. Backup and disaster recovery services are generally offered by cloud service providers and are often less expensive when compared to traditional on-premises solutions.

Data backup services place copies of health data in a cloud repository while disaster recovery services utilize cloud resources to back up an organization's IT infrastructure. While similar, backup and disaster recovery services differ in that backup services are typically used for routine systems where the recovery objective is within hours, whereas disaster recovery services are for critical systems that need to be restored within minutes to avoid business interruption. For organizations that have critical systems, they should consider cloud disaster recovery services to adequately prepare for a disaster. Choosing a service with a Recovery Time Objective of minutes or seconds, versus a backup type service that recovers files between hours to days, should be considered during the selection process.

Cloud Backup and Disaster Recovery as a Service providers offer solutions for both security and reliability. A secure cloud backup and DRaaS solution encrypts data in transit and at rest with encryption key management in compliance with the organization’s security policies, while also offering a configuration that ensures a solid level of security and compliance. A reliable cloud backup and DRaaS solution has a fault tolerant design with data durability audited on a regular schedule.

4.6.1. Backup Strategies

Data security puts trust in our systems creating complexity from a relationship requiring assurance and accountability. The most useful way to build confidence that data has not been corrupted requires the use of multiple strategies across several different systems. When other means of verifying the data's integrity have failed, we want to know with confidence that we can meet our obligations and restore valid services. This requires an interdisciplinary approach among different storage technologies and disciplines.

Bespoke solutions are sometimes required. For example, standard storage solutions often pack data onto physical devices where the striping across many devices creates look-up tables where information may be lost. Metadata loss can falsify the stored data's validity. A healthy backup needs to be totally separate from a primary system and possibly even into a different technology. For example, cloud or tape backup should be held to validate data on disk systems. This might also mitigate failure from hardware, malware, or terrorist activities. Enterprise solutions sometimes have proprietary requirements or designs hiding or removing integrity verification processes buried in the technology. Using validation checks on the systems or devices of the logical roads to data can help provide confidence corporations are not locked into bad vendor relationships. Sanity checks can also be run using external non-standard tools to help verify integrity.

Procedures need to be built to respond when potential problems are alerted so they can be evaluated and repaired quickly. Addressing potential issues also helps mitigate distrust in the operations and procedures validating the systems. As the complexity of our systems increases, confidence demands redundancy from regular integrity checks on organizations' data.

4.6.2. Disaster Recovery Planning

Disaster recovery (DR) is an essential aspect of IT infrastructure planning and implementation, and to a larger degree, how that infrastructure is monitored once it is operating. DR includes complete, real-time infrastructure capture, documentation, and regular declaration of the point of contact for reinstallation, monitoring, and/or completion of the DR process. Capturing the precise configuration/architecture of a deployed infrastructure, at a point in time navigation into the options, installable images, scripts, settings, schedules, and budgets of all components is no small amount of work—even in a small environment. In today's world of user- and customer-driven eServices expectations, outages are intolerable harms and infra operations suffer financial impacts aptly termed "User Frustration Costs" (UFCs) millions $/hr.

Using only operating systems or software-based virtualization, storage, user authentication, and security services is in many cases an integrated implementation just waiting for an environmental snafu to provide the expectations and UFCs of a transient Public Cloud Networking infrastructure. However, the DR burden is alleviated somewhat by cloud infrastructure operations' explicit deployment of entirely functional DR-equipped/dedicated Virtual Private Clouds and other more-or-less under the user's explicit control systems and services such as Landing Zones and Secure Network Constructs. Depending on the degree of assumed risk and allocated budget, a developed and deployed DR will use any or all of the DR options ranging from having no DR or very limited budget DR systems, all of which provide diminished reserves of stored and operating functionality.

4.7. Conclusion

In this chapter, we defined the requirements for health data storage architectures and discussed design considerations for cloud-enabled, health-data storage infrastructure. The analysis of existing industry offerings revealed two confident trends. First, there is a major shift from localized IT infrastructures embedded into end-user facilities to cloud trust anchors. Public cloud infrastructures are ceasing to be trust anchors however. Cloud infrastructures run by centralized public cloud providers are ceasing to be trust anchors due to multiple compliance, security and breakage risks. Healthcare organizations have started to deploy high availability, private cloud infrastructures by combining enterprise solutions from cloud software providers with enterprise solutions from cloud hardware providers. Health data storage is the primary driver of this refresh. Health collection and storage is being singled out for policy consideration. This is prompting federal authorities to devote financial stimulus to specially-designed cloud infrastructures that could become federally-secured utility storage services.

Second, ransomware has emerged as a credible risk for health data storage. Attacks aimed at available data assets have started to proliferate and become industrialized. The potential for ransoming of data assets, and the costs of such attacks, have intensified the investment and innovation ambitions of cloud infrastructure providers who offer data-specific cloud solutions. Emerging investment and innovation momentum focus on four areas: capabilities for mandatory data encryption, management and control; recovery point and RTO objectives for disaster recovery – ambitious objectives and service level guarantees; multi-cloud investments for data resiliency management; and immutable data storage. These trends are shaping an ecosystem of federally-rooted infrastructures that offer high-availability data storage and cyber immune data storage and access capabilities to healthcare organizations. The proposed health data storage as a utility service is a step toward a practical realization of this vision.

4.7.1. Emerging Trends

During the last couple of decades, the adoption of Cloud Storage has been rapidly increasing, with many new providers entering the market, improving both the offers and the service quality. New technologies such as event driven design with microservices and function-as-a-Service enabled new architectures that have a reduced complexity and scale down Business applications focused on document and simple file management. On the other hand, Health Data has become more centralized across specializations, and higher volumes have increased the need for higher performance services. The design study focuses on the synergy between the two trends, by designing a cloud –native novel architecture that enables the interoperation of the growing number of Health data storage and management cloud solutions, starting with the File Store Service, but also the Database Services as the Document DB, the Relational DB, the NoSQL Store, and the specialized services as the DICOM Store, the Data Lake Service, the Blockchain Service.

The services, as primitives provided to developers and organizations seeking to build Healthcare Software Applications Hosted in the Cloud, would be based on the more recent developments of the Distributed and Federated Cloud architecture solutions, and of the Serverless Computing Technology. The basic primitives behind the proposed Federated Multi-Cloud Healthcare Storage Solution Architecture are the newly formed Cloud File Systems that are built above and that allow the secure remote access to all other Storage Services. Whereas highly secure Data can be shared, with lower complexity and high performance between the remote BoTs, other Data, less sensitive or that use flawed security mechanisms can be shared by lower priority channels or with a corrected priority if they need to be exchanged continuously with immutable access control and auditing mechanisms. Finally, other Data are only exchanged using a federated Cloud strategy that relies on a Workflow engine that has been properly integrated with the Cloud Storage and that can trigger actions in 3rd Party Applications that can reuse the Cloud Storage Services APIs.

References

Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. JAMA, 309(13), 1351–1352. https://doi.org/10.1001/jama.2013.393

Kankanhalli, A., Hahn, J., Tan, S. B., Gao, G., & Tan, C. S. (2016). Big data and analytics in healthcare: Introduction to the special section. Information Systems Frontiers, 18(2), 233–235. https://doi.org/10.1007/s10796-016-9633-1

Wang, L., Alexander, C. A., & Lee, S. (2016). Big data analytics and health care: A systematic review. Journal of Biomedical Informatics, 61, 303–313. https://doi.org/10.1016/j.jbi.2016.04.004

Kruse, C. S., Goswamy, R., Raval, Y., & Marawi, S. (2016). Challenges and opportunities of big data in health care: A systematic review. JMIR Medical Informatics, 4(4), e38. https://doi.org/10.2196/medinform.5359