# Cloud-based Healthcare data management Framework

**Mohemmed Sha M[1]\***, **Mohamudha Parveen Rahamathulla[2]**

[1]Department of Computer Science, College of Arts and Sciences, Prince Sattambin Abdulaziz University,
Wadi Al Dawaser–Saudi Arabia
[e-mail: sahalshas@gmail.com]

[2]Department of Medical Lab Sciences, College of Applied Medical Sciences, Prince Sattambin Abdulaziz
University
Wadi Al Dawaser –Saudi Arabia
[e-mail: mahamudhaparveen@gmail.com]
*Corresponding author:Mohemmed Sha M

## *Abstract*

Cloud computing services changed the way the data are managed across the healthcare system that can improve patient care. Currently, most healthcare organizations are using cloud-based applications and related services to deliver better healthcare facilities. But architecting a cloud-based healthcare system needs deep knowledge about the working nature of these services and the requirements of the healthcare environment. The success is based on the usage of appropriate cloud services in the architecture to manage the data flow across the healthcare system.Cloud service providers offer a wide variety of services to ingest, store and process healthcare data securely. The top three public cloud providers- Amazon, Google, and Microsoft offers advanced cloud services for the solution that the healthcare industry is looking for. This article proposes a framework that can effectively utilize cloud services to handle the data flow among the various stages of the healthcare infrastructure. The useful cloud services for ingesting, storing and analyzing the healthcare data for the proposed framework, from the top three cloud providers are listed in this work. Finally, a cloud-based healthcare architecture using Amazon Cloud Services is constructed for reference.

# 1. Introduction

Every nation has the vision to improve the health and quality of life of its citizens by providing transparent and easy to access healthcare facilities. The advancements in the field of Information and Communication Technology (ICT) encourage almost all the fields including the healthcare industry for its expansion. eHealth is the practice of ICT in the health care system. This motivates the use of emerging web technologies and enables the improvement of health care broadly. Implementing eHealth through web applications and its portals, open the door for a patient-centric healthcare system throughout the world. But most of the healthcare systems follow standalone Hospital Information Systems (HIS) and uses different standards, which leads to high maintenance cost. Interoperability among these systems is difficult and it is considered as a big challenge to provide highly usable services of healthcare at a low cost.

Currently, the rate of data generation in the healthcare industry is at an ever-increasing speed. The various sources of healthcare data are from hospitals, insurance companies, and public datasets. Healthcare organizations try to use the power of the cloud to manage healthcare data wisely to improve the quality of healthcare. Developing healthcare systems using cloud computing resources from third-party providers will ease the situation [1].

Healthcare organizations have different types of data, from structured Electronic Health Record (EHR) data to unstructured patient data. Some of these data need real-time access and others need to be stored for a long time. Depends on the requirement, appropriate web services are used to manage the data at various level. The services needed for cloud application development is provided by different service providers like Amazon, Microsoft, Google, VM Ware, etc... These services are mainly grouped under certain domains such as compute, storage, database, security, management tools, and migration. These web services changed the way the data ingested, stored, processed in the cloud environment.

Designing a suitable architecture and usage of appropriate web service to manage the data is very important for the success of any healthcare system. This article proposes a framework that supports the healthcare system which uses several cloud-based services for its functionality. The main stages of the proposed architecture are data ingestion, storage, and processing. In each stage, the appropriate web services are used to complete certain functionality. Depend on the data need to be processed the appropriate cloud service from top cloud provides can be used to ingest, store and analyse the data. Security and Identity management is the main concern in all the stages of this healthcare framework.

This work is organized as follows. Section 2 reviews the existing cloud computing healthcare systems. In section 3 the proposed healthcare architecture and the related cloud services are listed. In the next part, an architecture that uses Amazon Cloud services is designed and its functions were discussed. Finally, this article is concluded with the importance of cloud computing resources in developing healthcare systems.

# 2. Literature Review

Currently, almost all the countries are facing challenges to provide high quality and easy to access healthcare facilities at an affordable cost. The developments in the field of Information and Communication Technology (ICT) have raised the expectation in the healthcare sector

that leads to a secure and cost-effective eHealth system. eHealth is the use of ICT to support the delivery of electronic services in the healthcare industry. It motivates the health managers to design and implements advanced healthcare services in the health sector to monitor the status ofpublic health and also to keep track of the diseases.

A number of surveys on the cloud-based healthcare applications specify that cloud is the perfect solution to satisfy the healthcare needs. Also, cloud provides powerful computation, efficient storage and advanced analytics on the healthcare data [2]-[4]. There are several researches that focus on the usage of cloud computing resources in the field of healthcare. They project the possibilities of building an advanced healthcare system using the latest cloud computing service offered by different cloud service providers [5]-[8].Alex in his research discussed the chances and challenges in implementing cloud computing in the healthcare sector [9]. Lidong and Cheryl present the service models of cloud computing that can support the different medical services in the healthcare industry [10].Sungyoung et al. proposed an architecture for a cloud-basedsystem that supports the mid-sized hospital to manage their service at low cost [11]. Changming et al. list the challenges in implementing the cloud computing service in the medical field [12].

Top cloud providers are interested in offering advance cloud services to handle the data flow in various stages of the healthcare system. They invent new services to match the advancementsin the medical field. Google cloud expertise in offering healthcare Application Programming Interfaces (APIs) to manage bigdata, machine learning and related analytics for better insights into the healthcare data [13]. Amazon Simple Storage Service (S3) for storing and Amazon Elastic Compute Cloud (EC2) for computing data offer a great opportunity for healthcare organizations to manage their data securely [14]. Microsoft Azure provides a wide range of service to deal with medical data management. The latest services from Azure supports Fast Healthcare Interoperability Resource (FHIR) to manage healthcare data in the cloud [15]. Constructing a suitable architecture is very important for any healthcare organization to use these services. The healthcare data stored in the cloud storage in the desired format should be processed and analyzed using these services to get some business values and insights.

## 3. Framework to manage healthcare data

The proposed cloud based Healthcare framework is shown in **Fig. 1** In the healthcare environment, the data are coming from both clinical and non-clinical setup. The clinical data include genome sequencing data, images from devices such as x-ray, ultrasound and Magnetic Resonance Imaging (MRI), digital data from medical equipments and data from sensors and wearable devices. The non-clinical data may include data from surveys, social media, and insurance claim [16]. The proposed architecture mainly concentrates on the clinical data from healthcare organizations.

The main stages of the proposed cloud-based healthcare architecture are data ingestion, data storage, and data processing. Throughout all the stages a set of cloud services can manage the data to activate the workflow. The initial step of the healthcare lifecycle is called data ingestion. In this stage, the raw data is pulled from various data sources of the healthcare setup. Depending on the use case, varieties of cloud services are used to support data ingestion. After cleansing /filtering the received data, the data is stored in the cloud. Depend on the analytics or

processing needed on the data for the future, the data stored with appropriate cloud storage. The last stage is processing the data to get some insights. The stakeholders of the healthcare sector are the end-users of this cloud-based system. In each stage, the appropriate cloud services from a cloud provider can be used to manage the data to satisfy the task. These stages are discussed in detail in the following section.
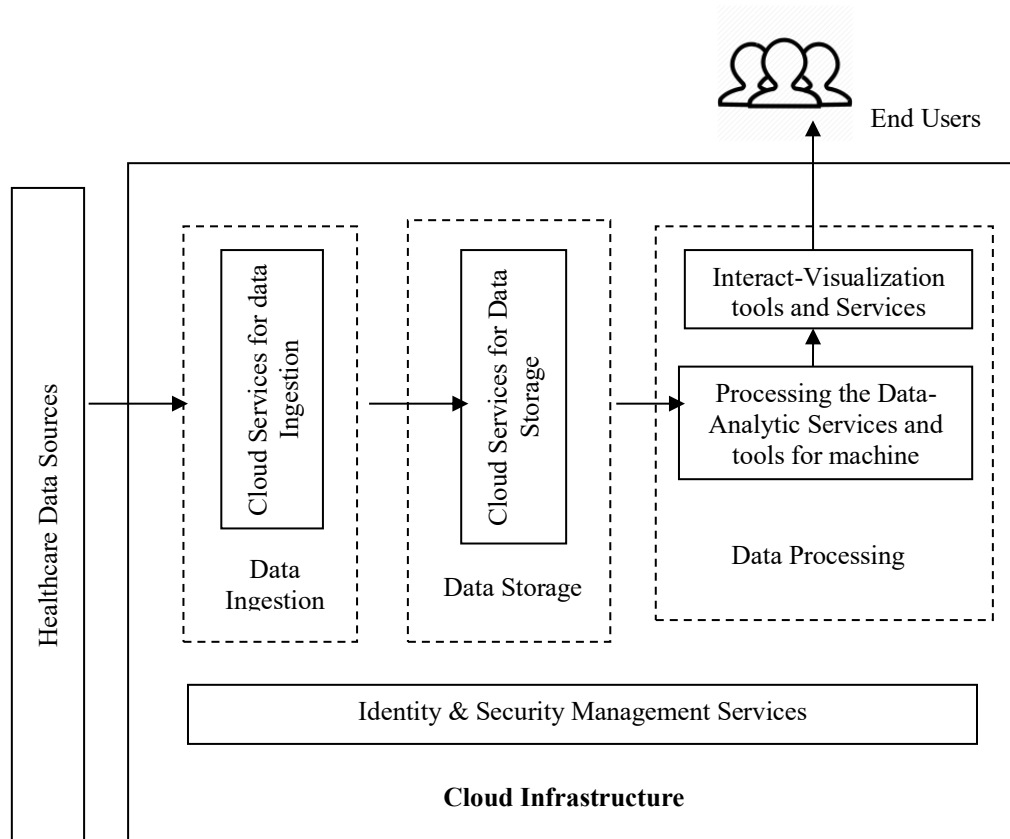


**Fig. 1.** Healthcare data management Framework

## 3.1 Data Ingestion Layer

The data ingestion layer is the initial and important layer in the cloud-based healthcare architecture because the computing services in the processing layer completely depend on the consistency and accessibility of the data stored in the cloud storage. In the data ingestion stage of the cloud healthcare system, the raw data from various sources are retrieved and passes to the cloud data storage. This retrieval mechanism may depend on the data source, its size, and latency. Cloud service provides offers different tools and services to ingest the data from various sources securely. The main sources are streaming data from the hospital system, databases from on-premises database servers, and files such as scanned images and x-rays stored in the server [17].  Based on the format of the data from the sources, the data flow is classified into categories such as streaming data, Batch data and application data.**Fig. 2** shows the data ingestion from various sources of the healthcare environment.
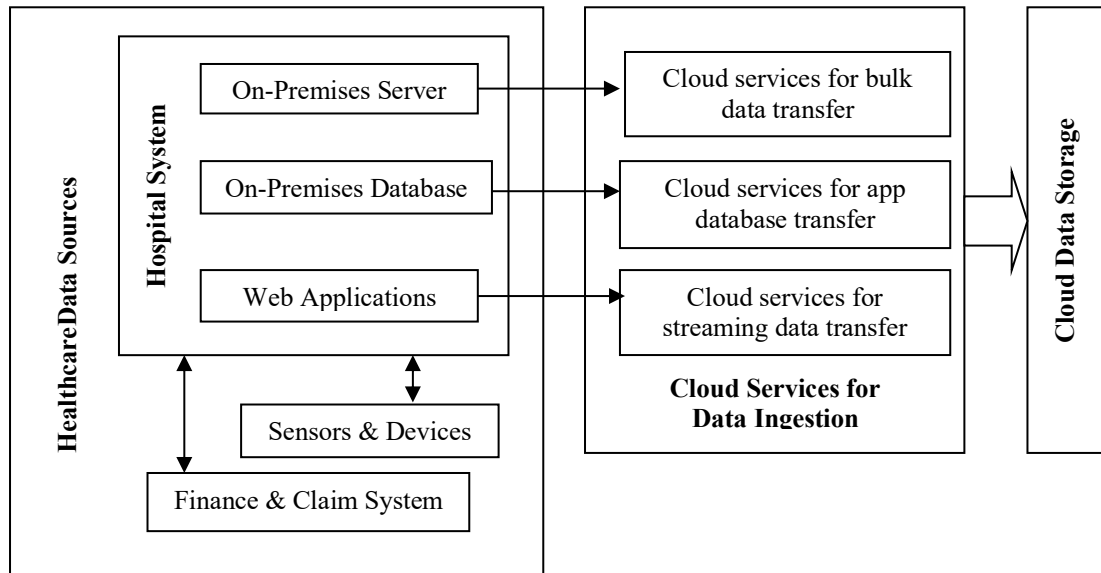
**Fig. 2.** Data Ingestion in Healthcare Environment

### 3.1.1 Streaming Data

Streaming data is the short message that passed to the data storage immediately when it is generated. Data generated from devices, data collected from mobile apps are passed asynchronously to the data storage for further processing. Providers have different types of web services for collecting streaming data from various sources. In the healthcare environment, the streaming data may include data from mobile applications and data from the medical Internet of Things (IoT) devices. Depend on the use case, the customer can select the appropriate service to ingest the data in the desired format. Amazon uses the cloud service called Kinesis for ingesting the streaming data from various healthcare environments such as hospital system that has Electronic Medical Records (EMR), Electronic Health Records (HER), Lab results and financial details. Cloud Pub/Sub is the service from Google to ingest streaming data in the healthcare environment. These services move the data messages to the storage as well as to the cloud data flow for real-time processing.

### 3.1.2 Bulk Data Ingestion

In some situation, a huge volume of data needs to be transferred to the cloud setup. This data may be located as a file or database format in the on-premises server. In the healthcare system, files such as diagnostics scanned images and X-rays are digitally stored in huge volume in the on-premises server for further analytics. Top cloud service providers offer services that can connect the on-premises server to the cloud storage for transferring bulk data. Amazon provides the device called snowball that can be used to ship petabytes of data securely to the cloud system. Google uses Cloud Transfer Service and Cloud Appliances for transferring on-premises data to the cloud storage.

### 3.1.3 Application/Hybrid data Ingestion

Data generated from application events are collected and sent by the web services to the cloud storage. Processing and analyzing the event-driven data will provide deep insights,used for healthcare decision making. An app data can be ingested to the object data storage as well as to a database in the cloud. Normally files are stored in the object data storage and structured data are stored in the cloud databases from applications. Google offers Cloud Pub/Sub to ingest real-time messages from healthcare application to cloud storage. Microsoft clouds provide the service called Event Grids to publish the data to Azure function. The Azure function will import bulk data into the cloud storage. Amazon Glue is the service which can transfer data from an on-premise database to Amazon S3 or other cloud database storage services. **Table 1** list the important cloud services used to ingest healthcare data into the cloud storage in the desired format.

**Table 1.** List of cloud Services for Healthcare data ingestion

| Data Types | Amazon | Google | Microsoft |
|---|---|---|---|
| Streaming Data | Amazon Kinesis | Cloud Pub/Sub | Data Factory |
| Batch Data | AWS Snowball devices, AWS Glue | Transfer Appliance, Cloud Transfer Service, DICOMweb | Event Grids, Azure Functions |
| Application/ Hybrid Data | AWS Storage Gateway, AWS Direct Connect. | Cloud Pub/Sub, App Engine, Compute Engine, Kubernetes Engine, Stackdriver Logging | Event Grids, Azure Functions |

### 3.2 Data Storage Layer

The data retrieved from various sources are stored in the cloud storage for further processing. Structured, semi-structured and unstructured are stored in the cloud in the cloud in a suitable format for analytics. Depend on type and nature of format appropriate cloud service can be used for storage. The ingested Electronic Health Record (HER) data may be in HL7 (Health Level Seven), FHIR (Fast Healthcare Interoperability Resources) or DISCOM (Digital Imaging and Communications in Medicine) formats [18].
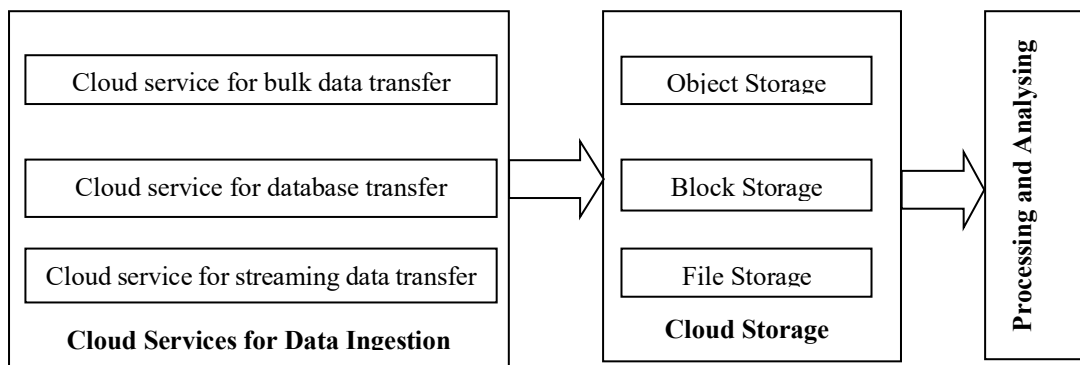


**Fig. 3.** Cloud Storage layer for Healthcare Data

In the healthcare industry files are commonly used to store diagnostic scanned images. Such bulky unstructured data can be stored as managed object storage. The function of cloud storage layer is shown in **Fig. 3**.

The top cloud provider offers HIPPA (Health Insurance Portability and Accountability Act of 1996) eligible storage services for securely storing healthcare data. Google cloud storage, Amazon S3 bucket, and Microsoft Data Lake Store are some of the bulk storage options for unstructured healthcare data.

**Table 2.** List of cloud Services for Healthcare data Storage

| Type of Storage | Amazon | Google | Microsoft |
|---|---|---|---|
| Object Storage | Amazon Simple Storage Service(S3), Amazon Glacier | Cloud Storage, Cloud SQL | Azure storage account, Blob Storage,Data Lake Store |
| Block Storage | Amazon Elastic Block Store, Amazon RDS,MariaDB, PostgreSQL, Oracle, and SQL Server, Dynamo DB | Cloud Data store,Cloud Bigtable,Bigquery, Cloud Firestore,Cloud Spammer | Persistent Disk, SQL Database, Table Storage, Azure Cosmos DB |
| File Storage | Amazon Elastic File System | Cloud File Store | Azure Files |

Cloud providers offer a variety of database storage services for SQL and NoSQL data from the healthcare setup [19]. In the proposed framework Amazon S3 act as the central storage and other storage classes will interact with Amazon S3 to maintain operational services with processing and analytics services. Azure Files support fully managed secure and shared access to the files in the cloud environment [20]. For example, archival data can be stored in Amazon Glacier from Amazon S3 for future usage. **Table 2** list the main cloud services used for data storage.

## 3.3 Data Processing and Analyzing

Once data is uploaded to the cloud storage, powerful analytic and visualization tools can be used to process the data into desirable insights. Cloud services connect the healthcare systems and the applications built on the cloud. These advanced cloud services support the processing of existing data formats in the healthcare system to have advanced analytics. Google Cloud Services supports streaming data processing with Cloud Data Flow, advanced scalable analytics using BigQuery and machine learning with Cloud Machine Learning Engine [21]. The healthcare industry must have to analyze the data from all the sources to have some business intelligence as well as management insights. **Fig. 4** shows the processing and analysing layer of the proposed framework.
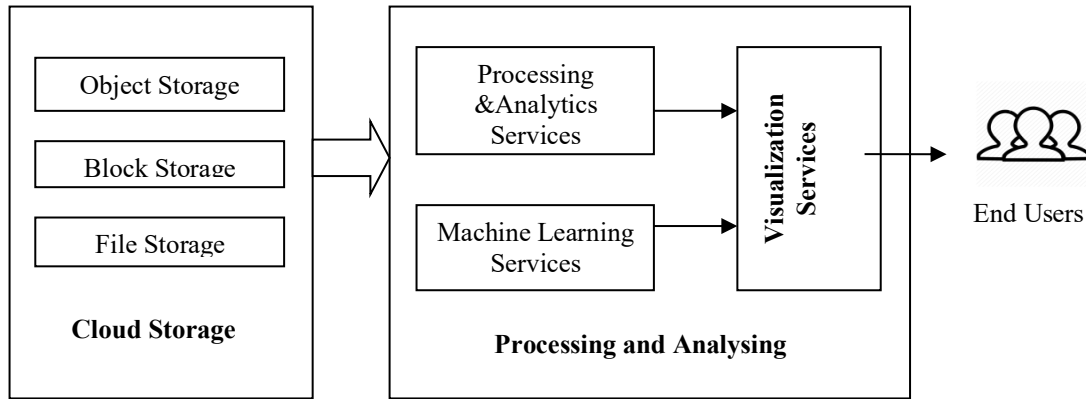
**Fig. 4.** Cloud data processing and analysing services

AWS provides over 100 machine learning and analytics services recognized by HIPPA to improve decision making and healthcare outcomes.Amazon Kinesis is the cloud service to analyze streaming data at real-time. Google provides a scalable data management and visualization tools that can the analyzed data to build machine learning models.Microsft Azure released a blueprint and provide related services for healthcare organizations to ingest, store and analyze sensitive healthcare data. Google Cloud provides services and machine learning tools to centralize and analyze the healthcare data to unlock the key data insights.There are several cloud services supported by top providers to analyze and process the healthcare data stored in the cloud. **Table 3** lists some of these services suitable for this purpose.

**Table 3.** Cloud Services for Healthcare data processing and Visualization

| Types of Processing | Amazon | Google | Microsoft |
|---|---|---|---|
| Streaming Data | Amazon Redshift | Cloud Dataflow,Cloud Dataproc | Cognitive Services,Azure Data Explorer |
| Scalable Analytics | Amazon EMR, Amazon Athena, EC2, Lambda, Amazon ElasticSearch | Bigquery, Cloud ML, Cloud Dataprep | Azure Data Explorer,Azure Databricks,Analysis Services |
| Machine Learning | Amazon Sage Maker | Cloud Machine Learning Engine,AutoML, CloudML | Machine Learning Studio, Application Insights |
| Data Visualisation | Amazon QuickSight | Cloud DataLab, Google Data Studio | Power BI |

## 4. Reference architecture with Amazon services

A reference architecture based on Amazon cloud services is constructed and shown in **Fig. 5**. Here Amazon S3 is acting as the Data Lake which can be the central storage for the healthcare data. Data from different data sources are ingested with services like Amazon Kinesis, Amazon direct connect into the data cloud storage. The huge volume of healthcare data from the on-premise server can be directly transferred to the S3 bucket using the device Snowball. Variety of data from the clinical setup can be sent to Amazon S3 through AWS storage gateway using Direct Connect [22]. Depend on the analytics needed the data moved different cloud storage options in the desired format from the data lake Amazon S3. Finally, the analytics and visualization services process the data to have healthcare insights.

In this architecture, the Amazon cloud data ingestion service Kinesis is used to retrieve the streaming data from the hospital setup. For loading huge volume of data from the on-premise server to the data lake in the cloud the Amazon Snow ball device is used. Amazon Direct Connect service load variety of data to Amazon S3 through the storage gateway. Non-clinical data such as data from scocial media and IoT devices are ingested using the services Kinesis Firehorse and Amazon IoT.
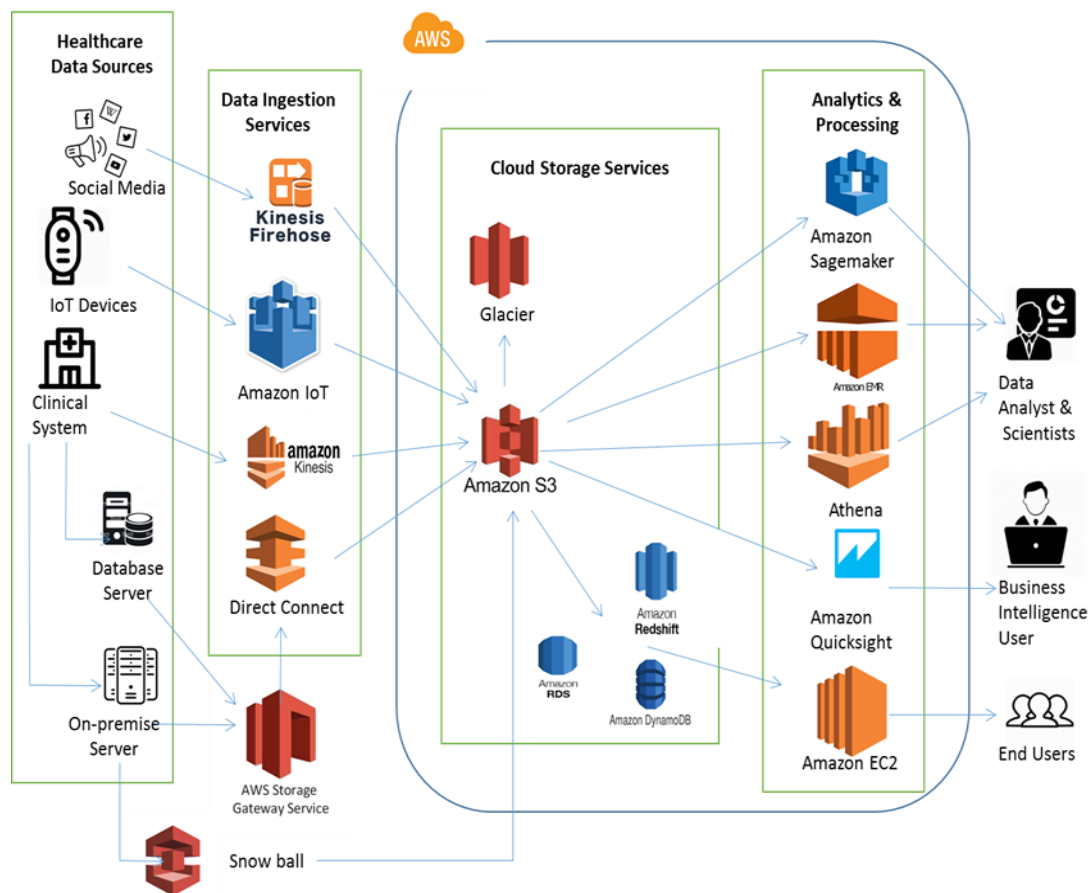


**Fig. 5.** Cloud-based healthcare architecture using Amazon services

Here, Amazon S3 acts as the primary storage because of its unlimited scalability. This storage is highly durable and centralized in nature. Unlike traditional data storage options, S3 supports the decoupling of data storage and data processing. Moreover, S3 provides a cost-effective solution to store all types of data formats from the healthcare industry. Depend on the analytics needed, the data stored in Amazon S3 is further forwarded to the cloud storage options such as Amazon RDS, Amazon DynamoDB and Amazon RedShift. The data need to be stored for long time is stored with the cloud archival service Amazon Glacier.

Amazon EC2 is the main compute service for processing the healthcare data stored in the various storage options. Based on the requirements of the stakeholders, the data is processed by Amazon EC2 and accessed through appropriate portals.

Analytics will be done using the services Amazon Athena and Amazon EMR for the deep insights. Amazon Quicksight is the advanced visualization tool which will display analytic results to the users in the desired format. AWS offers advanced machine learning and analytics tools to improve patient care which will lead to next-generation healthcare systems.

## 5. Challenges and Future Research

Hospitals and other healthcare organizations are using different software standards to manage their data. Before implementing the cloud services, the compatibility with the existing IT infrastructure must be analyzed. Ingesting the data in different formats from various data sources is a big challenge for any cloud-based application. Moreover, the data ownership of sensitive healthcare data is the main concern of the implementation of cloud technology in the healthcare sector. The selection of appropriate services suitable for data ingestion and storage from cloud providers is not an easy task. Healthcare data contain more sensitive information and hence appropriate security measures should be taken to protect data at each stage of data management. Cloud providers offer several security services that can secure the data from unauthorized access. Identifying those services at each layer of this architecture will be the future expansion of this research.

## 6. Conclusion

Cloud computing is the main solution for most of the problems the current healthcare industry is facing. These services can tackle the data stored across the healthcare system to meet the interoperability challenges. Data uploaded from multiple sources of the healthcare sector into cloud storage can be analyzed and visualized using advanced cloud offering services. These services can acquire the full potential of the huge data stored in the cloud to have great insights for management decisions. This research highlights the data handling in each layer of the proposed healthcare framework with appropriate cloud services. The latest cloud offerings support healthcare data in FHIR format to improve the interoperability among organizations. Also, the big data analytics services on huge healthcare data will motivate medical research. So the cloud-based healthcare system enables organizations to streamline its functions for faster and improved medical services.

# References

[1] Aziz HA and Guled A, "Cloud Computing and Healthcare Services," *Journal of Biosensors & Bioelectronics*, vol.7, no.3, pp.1-4, 2016. Article (CrossRef Link)

[2] C.Pino and R. Di Salvo, "A Survey of Cloud Computing Architecture and Applications in Health," in *Proc.of 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013)*, pp. 1649-1653, March, 2013. Article (CrossRef Link)

[3] Fernaz Narin Nur and Nazmun Nessa Moon, "Health care system based on Cloud Computing," *Asian Transactions on Computers*, vol. 2, no. 5, pp. 9-11, November, 2012. Article (CrossRef Link)

[4] Lena Griebel, Hans-Ulrich Prokosch, Felix Kopcke, Dennis Toddenroth, Jan Christoph, Ines Leb, Igor Engel and Martin Sedlmayr, "A scoping review of cloud computing in healthcare," *BMC Medical Informatics and Decision Making*, vol. 15, no. 17, March, 2015. Article (CrossRef Link)

[5] Fangjian Gao, Scott Thiebes, and Ali Sunyaev, "Rethinking the Meaning of Cloud Computing for Health Care: A Taxonomic Perspective and Future Research Directions," *Journal of Medical Internet Research*, vol. 20, no. 7, e10041, July, 2018. Article (CrossRef Link)

[6] Lingkiswaran Devadass, Sugalia Santhira Sekaran and Rajermani Thinakaran, "Cloud Computing in Healthcare," *International Journal of Students' Research In Technology & Management*, vol.5, no.1, pp.25-31, May, 2017. Article (CrossRef Link)

[7] Vaibhav Kamal Nigam, Shubham Bhatia, "Impact of Cloud Computing on Health Care," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 5, pp. 2804-2810, May, 2016.

[8] Maulik Parekh, Saleena B, "Designing a Cloud-based Framework for HealthCare System and applying Clustering techniques for Region Wise Diagnosis," *Procedia Computer Science*, vol. 50, pp. 537-542, 2015. Article (CrossRef Link)

[9] Alex Mu-Hsing Kuo, "Opportunities and Challenges of Cloud Computing to Improve Health Care Services," *Journal of Medical Internet Research*, vol. 13, no. 3, e67, July-September, 2011. Article (CrossRef Link)

[10] Lidong Wang, Cheryl Ann Alexander, "Medical Applications and Healthcare Based on Cloud Computing," *International Journal of Cloud Computing and Services Science*, vol. 2, no. 4, pp. 217-225, August, 2014.

[11] Sungyoung Oh, Jieun Cha, Myungkyu Ji, Hyekyung Kang, Seok Kim, Eunyoung Heo, Jong Soo Han, Hyunggoo Kang, Hoseok Chae, Hee Hwang, and Sooyoung Yoo, "Architecture Design of Healthcare Software-as-a-Service Platform for Cloud-Based Clinical Decision Support Service," *Health Informatics Research*, vol. 21, no. 2, pp. 102-110, April, 2015. Article (CrossRef Link)

[12] Changming Chen,Yaping Guo, Changbo Zhu, Xia Zhang, "Research on the Application of Cloud Computing in Medical Field," *Big Data and Cloud Innovation*, vol. 1, no. 1, pp. 1-7, 2017. Article (CrossRef Link)

[13] Google cloud, "Data lifecycle," Retrieved June 11, 2019. Article (CrossRef Link)

[14]  Amazon Web Services, "Store, Protect, Optimize Your Healthcare D ata in AWS," Retrieved June 19, 2019. Article (CrossRef Link)

[15] Microsoft Azure, "Azure Security and Compliance Blueprint - HIPAA/HITRUST Health Data and AI," Retrieved July 3, 2019. Article (CrossRef Link)

[16] Amazon Web Services, "Architecting for HIPAA Security and Compliance on Amazon Web Services," pp. 1-44, July, 2019. Article (CrossRef Link)

[17] Clear Data, "The 5 Phases of Health Data Management in the Cloud," Retrieved July 5, 2019. Article (CrossRef Link)

[18] Nathan Regola, Nitesh V Chawla, "Storing and Using Health Data in a Virtual Private Cloud," *Journal of Medical Internet Research*, vol.15, no.3, e63, March, 2013. Article (CrossRef Link)

[19] HIT Infrastructure, "Healthcare Data Storage Options: On-Premise, Cloud and Hybrid Data Storage," Retrieved July 7, 2019. Article (CrossRef Link)

[20] Microsoft Azure, "Azure Security Documentation," Retrieved June 19, 2019.

[21] Google Cloud, "Getting to know the Google Cloud Healthcare API: Part 1," Retrieved June 20, 2019. Article (CrossRef Link)

[22]  Amazon Web Services Architecture, "Store, Protect, Optimize Your Healthcare Data with AWS: Part 2," Retrieved June 19, 2019. Article (CrossRef Link)

**Dr. Mohemmed Sha M** received M.C.A in Computer Applications in the year 2000 from Manonmaniam Sundaranar University, India. He completed his M.Phil in Computer Science in 2005 from Bharathidasan University, India and Ph.D in Computer Applications from Manonmaniam Sundaranar University, Tirunelveli, India. His research focuses on Evaluating the Quality of Web services and Big Data Analytics. His reputed publications deals with selection of quality oriented web services based on its functionality and Big Data analytics in health care systems.He served as a principal of Achariya School of Business and Technology, Pondicherry for 6 years. Now he is working as an assistant Professor in Prince Sattam bin Abdulaziz University, Saudi Arabia.

**Dr. Mohamudha Parveen Rahamathulla** has completed her Masters in Medical Microbiology from University of Madras and PhD in Medical Microbiology from the Institute of National Importance - JIPMER, India. She has published many research papers in reputed journals. Her expertise and research interest includes Clinical Bacteriology and eHealth. She is serving as Peer reviewer in more than dozen reputed, peer-reviewed Medical and Science Journals. She is also an Assistant and Associate Editor in a couple of Journals. She is an active academician and researcher, presently working as Assistant Professor and Head of the Department in Department of Medical Lab Sciences, College of Applied Medical Sciences in Wadi Al Dawaser, Prince Sattam bin Abdulaziz University, Saudi Arabia.