SECURING ELECTRONIC HEALTH RECORDS ON CLOUD

Khilari Shital[1], Jadhav Kiran[2], Kalokhe Hemlata[3], Langote Shital[4]

[1] *Khilari Shital, Information technology, SRES COE kopargaon, Maharashtra,India*
[2] *Jadhav Kiran, Information technology, SRES COE kopargaon, Maharashtra,India*
[3] *Kalokhe Hemlata, Information technology, SRES COE kopargaon, Maharashtra,India*
[4] *Langote Shital, Name Information technology, SRES COE kopargaon, Maharashtra,India*

ABSTRACT

We present a cloud-based approach for the design of interoperable electronic health record (EHR) systems. Cloud computing environments provide several benefits to all the stakeholders in the healthcare ecosystem (patients, providers, payers, etc.). Lack of data interoperability standards and solutions has been a major obstacle in the exchange of healthcare data between different stakeholders. We propose an EHR system—cloud health information systems technology architecture (CHISTAR) that achieves semantic interoperability through the use of a generic design methodology which uses a reference model that defines a general purpose set of data structures and an archetype model that defines the clinical data attributes. CHISTAR application components are designed using the cloud component model approach that comprises of loosely coupled components that communicate asynchronously. In this paper, we describe the high-level design of CHISTAR and the approaches for semantic interoperability, data integration, and security.

Keyword: - *Cloud EHR, data integration, electronic health records, healthcare.*

---

1. INTRODUCTION

Healthcare ecosystem consists of the healthcare providers (doctors, physicians, specialists, etc.), payers (health insurance companies), pharmaceutical companies, IT solutions and services firms, and the patients. The process of provisioning healthcare involves massive healthcare data which exists in different forms (structured or unstructured) on disparate data sources (such as relational databases, file servers, etc.) and in different formats. When a patient is admitted to a hospital, his/her information is entered into electronic health record (EHR) systems. Physicians diagnose the patient and the diagnostic information (from medical devices such as CT scanners, MRI scanners, etc.) is stored in EHR systems. In the diagnosis process, the doctors retrieve the health information of patients and analyze it to diagnose the illness. Doctors can take expert advice by sharing the information with consulting specialists. The cloud can provide several benefits to all the stakeholders in the healthcare ecosystem through systems such as health information management system, laboratory information system, radiology information system, pharmacy information system, etc. With public cloud based EHR systems hospitals do not need to spend a significant portion of their budgets on IT infrastructure. Public cloud service providers provide on-demand provisioning of hardware resources with pay-per-use pricing models. Thus, hospitals using public cloud-based EHR systems can save on upfront capital investments in hardware and data center infrastructure and pay only for the operational expenses of the cloud resources used. Hospitals can access patient data stored in the cloud and share the data with other hospitals. Patients can provide access to their health history and information stored in the cloud (using SaaS applications) to hospitals so that the admissions, care, and discharge processes can be streamlined. Physicians can upload diagnosis reports (such as pathology reports) to the cloud so that they can be accessed by doctors remotely for diagnosing the illness. Patients can manage their prescriptions and associated information such as dosage, amount, and frequency, and provide this information to their healthcare provider. Health payers can increase the effectiveness of their care management programs by providing value added services and giving access to health information to members.

2. LITERATURE SURVEY

VistA [2] is the most widely used EHR system in the United States. There are three main distributions of VistA outside VA—WorldVista [6] (GPL licensed), OpenVistA [3] (AGPL licensed), and vxVista [7] (EPL licensed). OpenEHR [8] is an EHR system that is designed for achieving semantic interoperability. OpenEHR puts special emphasis on semantic interoperability to improve the quality of data exchanged between different stakeholders in the healthcare ecosystem. OpenEHR is based on a two-level modeling approach in which a reference model constitutes the first level of modeling, while the formal definitions of clinical content in the form of archetypes and templates constitute the second level. To enable interoperability of healthcare data, various solutions have been developed that allow integrating data from different sources. Mirth Connect [9] is an open source integration engine that supports a variety of messaging standards and protocols for connecting to external systems and databases. FM projection [10] is a set of tools that allows inspecting VistA File Manager data and structures using SQL like representations. In our previous work [11], we proposed a data collection framework for collecting big sensor data in a cloud. For CHISTAR, we propose a similar approach for data collection that is based on a cloud-based distributed batch processing infrastructure. Since EHR systems handle massive healthcare data, benchmarking the performance of such systems is important to ensure the effectiveness of such systems in provisioning healthcare. For testing cloud-based systems such as EHRs we proposed an approach for prototyping and benchmarking cloud-based systems in our previous work. A similar approach will be used for evaluating the performance of CHISTAR.

CHISTAR uses the cloud component model approach for application design described in our previous work [14]. For the design of mobile applications that can utilize the capabilities of the next generation of cellular networks, CHISTAR adopts the guidelines described by Radio *et al*. [15].

3. PROPOSED CLOUD-BASED EHR SYSTEM

Fig. -1 shows the layered architecture of the proposed CHISTAR system. The infrastructure services layer consists of the cloud instances (for load balancers, application servers, Hadoop master, and slave nodes, etc.) on which CHISTAR is deployed. The information services layer consists of a data integration engine (that allows integrating data from multiple disparate data sources into the cloud), models for data storage and clinical concepts, and the data governance module. The application services layer provides various services such as EHR service, demographic service, archetype service, and terminology service. The presentation services layer consists of smart and connected healthcare applications (web and mobile based).

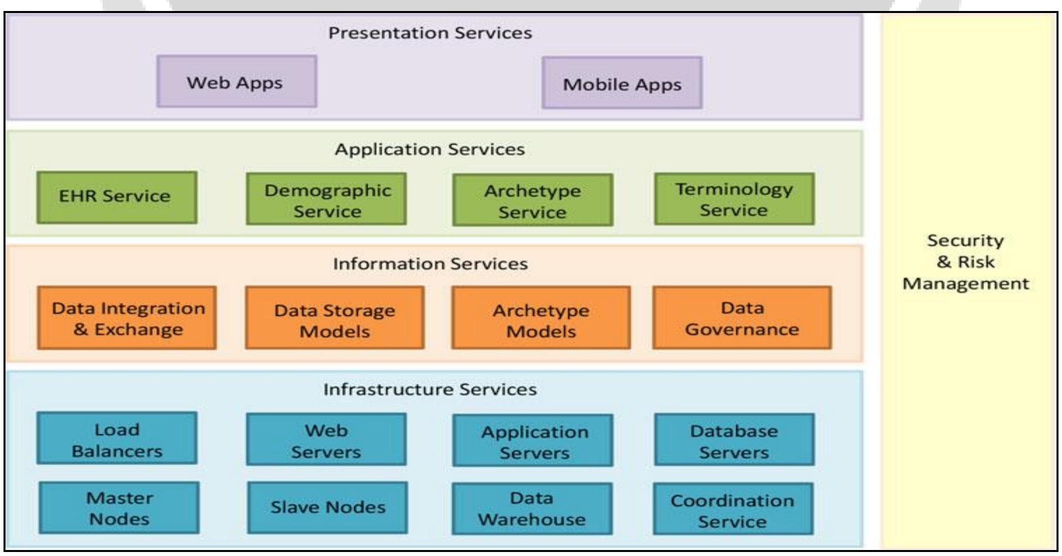

Fig. -1 Architecture of proposed CHISTAR system.

3.1 The key design principles of CHISTAR are described as follows.

- ***Semantic Interoperability***

Semantic interoperability is defined as the ability to share, interpret, and make effective use of information exchanged. CHISTAR achieves semantic interoperability by using a generic design approach. CHISTAR uses a two-level modeling approach which separates data from clinical knowledge. A two-level modeling approach for an EHR system consists of a data storage model and an archetype model [8]. Data storage model defines entities for data storage and represents the semantics of storing data. Archetype model defines the clinical concepts. Archetype model represents the domain-level structures and constraints on the generic data structures defined by the data storage model. Two-level modeling approach makes the system more robust as the software need not be changed whenever there is a change in the clinical knowledge.

- ***Data Integration***

Healthcare data exist in various forms (structured or unstructured) on different data storage systems such as relational databases (RDBMS such as MySQL, Oracle, etc.), file servers (as text, image, video files, etc.) and EHR standards (such as HL7 messages). The proposed data integration engine is based on Hadoop Map Reduce framework

- ***Security***

The biggest obstacle in the widespread adoption of cloud computing technology for EHR systems is security and privacy issues of healthcare data stored in the cloud, due to its outsourced nature. In U.S., organizations called covered entities (CE), that create, maintain, transmit, use, and disclose an individual's protected health information (PHI) are required to meet health insurance portability and accountability act (HIPAA) requirements. HIPAA requires CE to assure their customers that the integrity, confidentiality, and availability of PHI information they collect, maintain, use, or transmit is protected. HIPAA was expanded by the health information technology for economic and clinical health act (HITECH), which addresses the privacy and security concerns associated with the electronic transmission of health information.

Following are the key security features.

1) ***Authentication:*** CHISTAR adopts single sign on (SSO) for authentication. SSO enables users to access multiple applications after signing in only once, for the first time. When a user signs in, the user identity is recognized and there is no need to sign in again and again to access related applications.
2) ***Authorization:*** Authorization services include policy management, role management, and role-based access control.
3) ***Identity management:*** Identity management services provide consistent methods for identifying persons and maintaining associated identity attributes for the users across multiple organizations.
4) ***Securing data at rest:*** CHISTAR adopts 256-bit Advanced Encryption Standard (AES-256) which is a data encryption standard. All CHISTAR data that are stored in HBase are first encrypted with AES-256 encryption and then inserted into HBase.
5) ***Securing data in transit:*** All transmission of data is protected with HTTP over secure socket layer (SSL) encryption technology.
6) ***Key management:*** All keys for encryption are stored in a data store in the cloud which is separate and distinct from the actual data store. Additional security features such as key rotation and key encrypting keys are also used. Keys can be automatically or manually rotated. The key change frequency can be configured. In the automated key change approach, the key is changed after a certain number of transactions (i.e., accesses to a patient's records). All keys are themselves encrypted using a master key.
7) ***Data integrity:*** Data integrity ensures that the data are not altered in an unauthorized manner after it is created, transmitted, or stored. CHISTAR uses message authentication codes (MAC) to detect both accidental or deliberate modifications in the data. MAC is a cryptographic checksum on the data that is used to provide an assurance that the data have not changed. Computation of MAC involves the use of (1) a secret key that is known only to the party that generates the MAC and the intended recipient, and (2) the data on which the MAC is computed.
8) ***Auditing:*** Regulations such as HIPAA/HITECH require that log data on the accesses to PHI be maintained for accountability purposes. CHISTAR logs all read and write accesses to patient health records. Logs include the user involved, type of access, timestamp, actions performed, and records accessed.

- **Component-Based Architecture**

  CHISTAR adopts the Cloud component model approach for application design described in our previous work [14]. Cloud component model allows identifying the building blocks of a cloud application which are classified based on the functions performed and type of cloud resources required. Each building block performs a set of actions to produce the desired outputs for other components. The model is represented as a component map in which the columns represent various functions of the application, and rows represent cloud resources.

- **Evaluation**
  We deployed CHISTAR on the amazon elastic compute cloud (EC2) infrastructure. In this deployment, tier-1 consists of web servers and load balancers, tier-2 consists of application servers, and tier-3 consists of a cloud-based distributed batch processing infrastructure such as Hadoop [18]. HBase is used for the database layer. HBase [19] is a distributed nonrelational column oriented database that runs on top of HDFS. HBase provides a fault-tolerant way of storing large quantities of sparse data. HDFS is used for the storage layer for storing healthcare data in the form of flat files, images, etc. Hive [20] is used to provide a data warehousing infrastructure on top of Hadoop. Hive allows querying and analyzing data in DFS/HBase using the SQL-like Hive query language (HQL). Zookeeper [29] is used to provide a distributed coordination service for maintaining configuration information, naming, providing distributed synchronization, and providing group services.

4. RESULTS

We used the amazon simple queuing service (SQS) [28] for message queues between various components of CHISTAR. Amazon SQS offers a reliable, highly scalable, hosted queue service for storing messages. To store the intermediate status,

we used Amazon SimpleDB [28] as the status database. Amazon SimpleDB is a highly available and flexible nonrelational data store. CHISTAR components communicate asynchronously using the SQS messaging queues and store the state externally in a SimpleDB database.

In order to evaluate the scalability of CHISTAR, we performed a series of experiments with very large data sets (upto 1 000 000 patient health records). The data sets for the experiments were generated synthetically. The patient record data used for experiments consisted of diagnosed problems, medications, vital signs, etc., for patients. Fig. -2 shows the average response time for the CHISTAR application for four different deployment configurations and varying number of patient health records. The results shown in Fig. -2 were obtained with 100 users accessing the CHISTAR application simultaneously. We observe that response time increases as the number of records increase.

Fig. -3 shows the average response time for the CHISTAR application for four different configurations and varying number of simultaneous users. The results in Fig. -3 were obtained with 10 000 patient health records in the CHISTAR application. With increase in number of users the mean request arrival rate increases since CHISTAR services higher number of requests per second, therefore an increase in response time is observed.

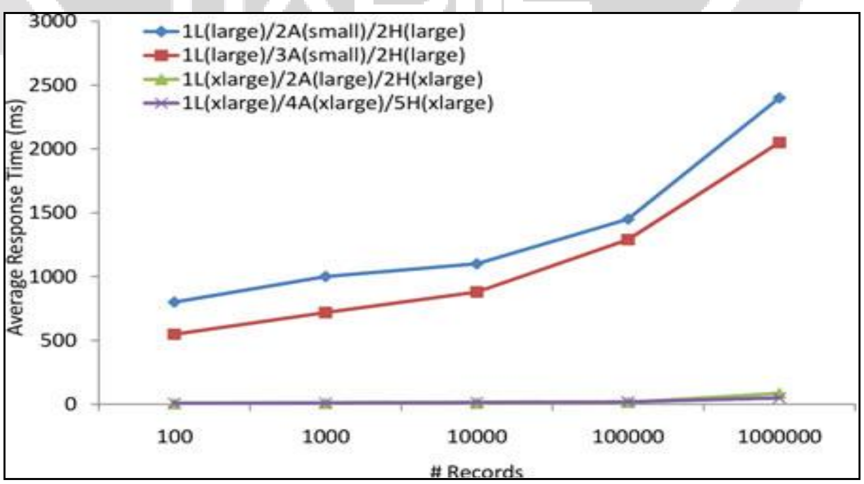

Fig. -2. Average response time for CHISTAR for varying number of patient records with 100 simultaneous users.

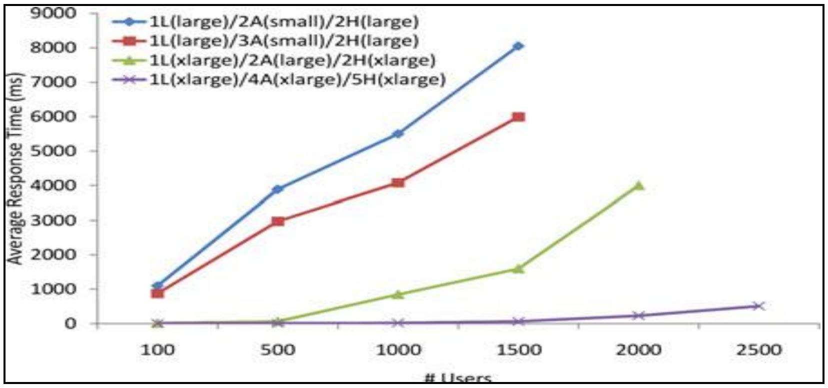

Fig. -3. Average response time for CHISTAR for varying number of simultaneous users with 10 000 patient records.

5. ADVANTAGES OVER EXISTING SYSTEM

In this section, we describe the advantages of cloud-based EHR systems over client–server EHRs that are based on a dedicated hosting model.

1) Interoperability: CHISTAR has better interoperability as compared to client–server-based EHR systems such as VistA. To achieve interoperability, CHISTAR adopts a two-level modeling approach for separation of information from the clinical knowledge. Furthermore, the data integration engine of CHISTAR allows integrating data from disparate data sources such as MySQL servers, JDBC servers, Oracle, file servers, and different EHR standards (HL7 messages, HL7 CDA documents, etc.) into a cloud-based storage.

2) Scalability: Cloud-based EHRs such as CHISTAR have better scalability as compared to client–server EHRs. CHISTAR adopts the cloud component model approach for application design which provides better scalability by decoupling application components and providing asynchronous communication mechanisms. Since components are designed to process requests asynchronously, it is possible to parallelize the processing of requests. Using cloud component model, CHISTAR can leverage both horizontal and vertical scaling options.

3) Maintainability: CHISTAR has better maintainability as compared to client–server-based EHR systems. The functionality of individual components of CHISTAR can be improved or upgraded independent of other components. Loose coupling allows replacing or upgrading components, without changing other components. Since CHISTAR has loosely coupled components, it is more resilient to component failures. In case of client–server-based EHR systems with tightly coupled components, failure of a single component can bring down the entire application.

4) Portability: Cloud-based EHR systems such as CHISTAR have better portability. By designing loosely coupled components that communicate asynchronously, it is possible to have innovative hybrid deployments in which different components of an application can be deployed on cloud infrastructure and platforms of d ifferent cloud vendors.

5) Reduced Costs: Client–server EHR systems with dedicated hosting require a team of IT experts to install, configure, test, run, secure, and update hardware and software. With cloud based EHR systems, organizations can save on the upfront capital investments for setting up the computing infrastructure as well as the costs of managing the infrastructure as all of that is done by the cloud provider. Though hardware maintenance overhead is reduced, organizations still need to pay for the software maintenance and support costs. Additional cost benefits come by scaling cloud resources up (or scaling out) only for those components which require additional computing capacity.

6. CONCLUSIONS

In this paper, we described the design of a cloud-based HER system—CHISTAR which addresses the problems faced by traditional client–server EHR systems. CHISTAR adopts a two level modeling approach for achieving semantic interoperability. The data integration engine of CHISTAR allows aggregating healthc are data from disparate data sources. CHISTAR supports advanced security features and addresses the key requirements of HIPAA and HITECH. CHISTAR has better interoperability, scalability, maintainability, portability, accessibility, and reduced costs as compared to traditional client–server EHR systems. Future work will focus on the development of a cloud based information integration and informatics (III) framework for healthcare applications. III framework will allow development of smart and connected healthcare applications backed by massive scale healthcare data integrated from heterogeneous and distributed healthcare systems within a scalable cloud infrastructure.

7. REFERENCES

[1]. KPMG, The Cloud Changing the Business Ecosystem [Online]. Available: www.kpmg.in
[2]. (2012). VistA Monograph [Online]. Available: www.va.gov/vistamonograph
[3]. Medsphere OpenVistA. (2012). [Online].Available: http://medsphere.org/community/project/openvista-server
[4]. Medsphere Systems Corporation. (2010). From VistA to OpenVista—Enhancing an Already Valuable Tool [Online]. Available: www.medsphere.com/vista-to-openvista, 2010
[5]. Medsphere Systems Corporation. (2010). From MUMPS to Java: OVID Unleashes Power of Open-source Health IT [Online]. Available: http://www.medsphere.com/ovid-white-paper
[6]. WordVistA. (2012). [Online]. Available: http://worldvista.org
[7]. vxVistA. (2012). [Online]. Available: https://www.vxvista.org
[8]. OpenEHR. (2012). [Online]. Available: http://www.openehr.org
[9]. Mirth Connect. (2012). [Online]. Available: http://www.mirthcorp.com/products/mirth-connect
[10]. Medsphere FM Projection. (2012). [Online]. Available: http://medsphere.org/community/project/fm-projection
[11]. A. Bahga and V. K. Madisetti, "Analyzing massive machine maintenance data in a computing cloud," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 10, pp. 1831–1843, Oct. 2012.
[12]. A. Bahga and V. K. Madisetti, "Synthetic workload generation for cloud computing applications," J. Softw. Eng. Appl., vol. 4, no. 7, pp. 396–410, 2011.
[13]. A. Bahga and V. K. Madisetti, "Performance evaluation approach for multi-tier cloud applications," J. Softw. Eng. Appl., vol. 6, no. 2, pp. 74–83, 2013.
[14]. A. Bahga and V. K. Madisetti, Rapid Prototyping of Advanced Cloud- Based Services and Systems, IEEE Comput., under review, Aug. 2012.
[15]. N. Radio,Y. Zhang, M. Tatipamula, and V. K.Madisetti, "Next-generation applications on cellular networks: Trends, challenges, and solutions," Proc. IEEE, vol. 100, no. 4, pp. 841–854, Apr. 2012.
[16]. J. Patrick, R. Ly, and D. Truran, "Evaluation of a persistent store for openEHR," in Proc. HIC HINZ, 2006, pp. 83–89.
[17]. HL7. (2012). [Online]. Available: http://www.hl7.org
[18]. Apache Hadoop. (2012). [Online]. Available: http://hadoop.apache.org/mapreduce
[19]. Apache HBase. (2012). [Online]. Available: http://hbase.apache.org
[20]. Apache Hive. (2012). [Online]. Available: http://hive.apache.org
[21]. hQuery. (2012). [Online]. Available: http://projecthquery.org