Community Cloud Architecture to Improve Use Accessibility with Security Compliance in Health Big Data Applications

Samaikya Valluripally, Murugesan Raju, Prasad Calyam,
Matthew Chisholm, Sai Swathi Sivarathri, Abu Mosa, Trupti Joshi
University of Missouri-Columbia, USA
svbqb@mail.missouri.edu,rajum@mail.missouri.edu,calyamp@missouri.edu,mrcd67@mail.missouri.edu,sss26x@mail.missouri.edu,mosaa@health.missouri.edu,joshitr@health.missouri.edu

ABSTRACT

The adoption of big data analytics in healthcare applications is overwhelming not only because of the huge volume of data being analyzed, but also because of the heterogeneity and sensitivity of the data. Effective and efficient analysis and visualization of secure patient health records are needed to e.g., find new trends in disease management, determining risk factors for diseases, and personalized medicine. In this paper, we propose a novel community cloud architecture to help clinicians and researchers to have easy/increased accessibility to data sets from multiple sources, while also ensuring security compliance of data providers is not compromised. Our cloud-based system design configuration with cloudlet principles ensures application performance has high-speed processing, and data analytics is sufficiently scalable while adhering to security standards (e.g., HIPAA, NIST). Through a case study, we show how our community cloud architecture can be implemented along with best practices in an ophthalmology case study which includes health big data (i.e., Health Facts database, I2B2, Millennium) hosted in a campus cloud infrastructure featuring virtual desktop thin-clients and relevant Data Classification Levels in storage.

KEYWORDS

Smart Healthcare, Cloud Architecture, Big Data Application, Security Standard Compliance, Electronic Health Records

ACM Reference format:
Samaikya Valluripally, Murugesan Raju, Prasad Calyam, Matthew Chisholm, Sai Swathi Sivarathri, Abu Mosa, Trupti Joshi. 2018. Community Cloud Architecture to Improve Use Accessibility with Security Compliance in Health Big Data Applications. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 6 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Health Care Systems collect big data using many entities such as instruments, billing records over several years. Increasingly, health care enterprises (both large organizations as well as individual providers) are taking advantage of state-of-the-art technologies in order to extract meaningful insights out of their big data. The goal in such data-driven health care eco-systems is to bridge the knowledge gaps in the information management between patients, health care providers and medical researchers as shown in Figure 1. The expectation is to leverage large health care related data sets to find new risk factors related to a disease, drive clinical research and ultimately inform clinical decisions for smart healthcare.

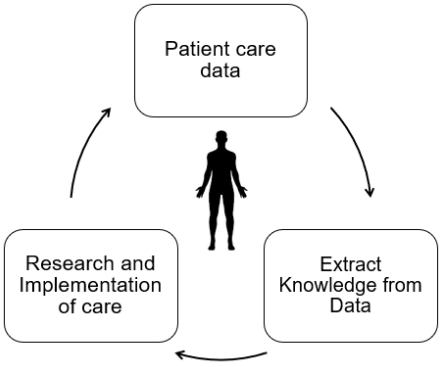

Figure 1: Big data analytics integration enables precision medicine within a Continuous Learning Healthcare System cycle.

With the current big data revolution especially in massive data collection and data availability for analysis/visualization, medical breakthroughs are possible in the areas of personalized medicine [1], knowledge discovery [2] by determining risk factors for diseases and testing hypotheses with relevant heuristics for clinical research [3]. In fact, biomedical big data has become one of the critical thrust areas for the US National Institutes of Health [4]. Moreover, electronic health records (EHRs) have transformed the patient data as well as diseases information availability to researchers and physicians. American Medical Informatics Association (AMIA) Genomics and Translational Bioinformatics Working Group [5] has identified that knowledge discovery and data mining as important components of clinical research informatics and next-generation Clinical Decision Support.

For successful adoption of big data analytics in healthcare applications as shown in Figure 1, several challenges around data sets need to be simultaneously addressed such as: compliance with stringent security standards, handling of the growing size/diversity of health care data, privileged access to clinical data or proprietary analytic tools, and high-speed processing of data sets with scalability considerations [6]. Particularly, health big data applications comprise of several lifecycle stages where data is transformed under different security requirements spanning multiple provider/consumer domains as illustrated in Figure 2. In addition, they not only use local private cloud resources, but distribute their data processing workflows across community and public cloud platforms.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

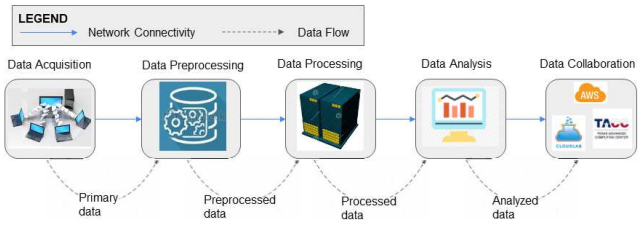

Figure 2: End-to-end lifecycle stages of a big data application with security requirements involving federated cloud resources.

In addition to the above broad challenges, specific issues that are related to data exchange and integration across the multiple domains include e.g., the lack of homogenization initiatives to handle: (a) multi-source data from clinicians, EHRs and Researchers, (b) diverse file formats with different data classification levels, and (c) security policy management for compliance with inter-organization security standards in the lifecycle stages i.e., billing (PCI DSS standard [7]), data access (HIPAA (Health Insurance Portability and Accountability Act) standard [8]), cyberinfrastructure configuration (NIST standard [9]). As a result, a health big data application owner today has the burden to evaluate computing, networking as well as storage options to implement existing security standards, identifying gaps in achieving scalable application performance, while having the flexibility to test bold hypotheses relevant to e.g., clinical care and drug discovery.

In this paper, we address the above challenges by proposing a community cloud architecture that helps end users (clinicians, researchers, data scientists) improve the use accessibility of health big data across the life cycle stages, while also supporting middleware frameworks that broker the multi-domain resources and align them with security standards compliance. Our main contributions in this work include:

- Threat modeling of a health big data application workflow
- Security Compliance with standards e.g., HIPAA, NIST
- Novel community cloud architecture supporting data transformation via a secure gateway and thin-client user access
- Best practices to implement our novel community cloud architecture in the context of an ophthalmology case study [11]

Our community cloud architecture is inspired by the work in [12] where, a cloudlet component is established between isolated/legacy data or instrument resources, and a private cloud to enhance user performance and also meet data/instrument provider's security requirements. Our architecture assumes that health big data applications will need to retrieve knowledge from billions of records while leveraging advanced technologies such as: relational database management system (RDBMS) [13], virtual desktop cloud [14] [15], and Hadoop file system [16]. The context for our big data problem assumption in smart healthcare stems from our case study that involves the Health Facts database at University of Missouri that allows data query via analytical software packages such as Statistical Analysis System (SAS) [17] to mine Electronic Health Records (EHR) information belonging to more than 50 million patients' clinical data. Researchers using Health Facts also use other sources of data e.g., I2B2, Millennium database within the University of Missour campus, as well as data from external archives.

The remainder of the paper is organized as follows: Section 2 presents the related work. Section 3 presents our problem formulation for the threat model and the health big data application performance/security requirements via a case study. Section 4 details our proposed community cloud architecture. Section 5 concludes the paper with a summary of best practices for our community cloud architecture implementation.

2 RELATED WORK

A data acquisition and analysis framework that captures, stores, correlates and coordinates real-time digital data in a trusted manner before the data is shared or widely accessed is proposed in [12]. In our work, we use a similar framework to store, process, analyze and distribute patient health data in a scalable manner or even in real-time within secure environment for data access. Authors in [12] propose a framework which is efficient in dealing with high-volume and fast-changing workload of heterogeneous types of data processing. Our work is alike in terms of dealing with diverse data sets at large scale. Our architecture addresses problems such as mismatched performance due to data access policies that improve security but introduce bottlenecks through stringent firewall rules and low responsive queries to protect against external threats.

Industry efforts such as in [18], [19], [20], [21] aim at proposing variants of community cloud architectures to deliver health care services in precision medicine with minimal cost, high clinical value, and high usability. In these architectures various computation capabilities along with HIPAA compliance are deployed to support the developing applications in health care. Our proposed solution also features high-speed computational capabilities in a campus computing environment, and outlines a novel HIPAA aligned compliance module within a multi-domain resource brokering framework to serve the needs of the end users in health care big data applications.

Existing works [22], [23] and [24] pertaining to security and dependability for community-cloud resources in data-intensive research communities mostly deal with security measures and point solutions to counter confidentiality, availability and integrity threats. They also do not consider end-to-end security design that helps in dynamic allocation and adaptation using such measures. Our method in this work for performance alignment with security standards is extended from our prior work in [27] that catered to bioinformatics application use cases. Health big data communities can benefit from our proposed solution approach of resource allocation based on multi-domain security requirements, and can augment their current approach of manual co-ordination of policies to achieve end-to-end security alignment.

3 HEALTH APPLICATION CASE STUDY

3.1 Health Facts based Study Motivation

In this section, we present a big data health application system that motivates our proposed architecture for an integrated high performance scalable solution. The Health Facts database was used in our prior study [11] to investigate the risk factors for age related ophthalmological illness such as cataracts. Amongst eye related illnesses, cataracts are the leading cause of blindness among people aged 40 and above worldwide. The economic burden of several age-related ophthalmological illness has been rising worldwide, however, the actual cause for increased prevalence of age-related ophthalmological illness is not yet known. An electronic medical record system is used in the Health Facts database used in our

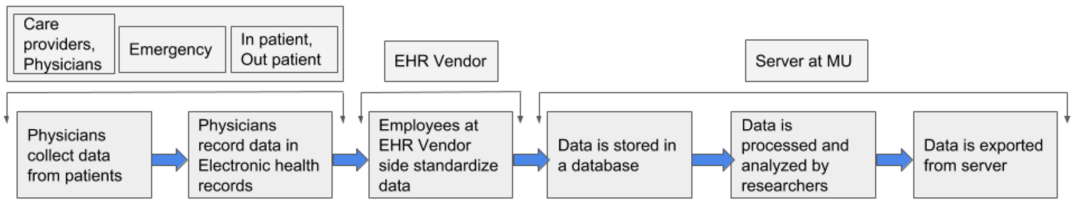

Figure 3: Pipeline diagram for a health big data application that illustrates the steps for collection of different source data and their analysis.

FACT Tables	Number of Records
HF_F_clinical_event	5,606,006,254
Hf_f_diagnosis	816,622,927
Hf_f_encounter	461,206,375
Hf_f_enc_history	309,634,504
Hf_f_implant_log	33,567
Hf_f_lab_procedure	4,293,382,357
Hf_f_medication	634,846,774
Hf_f_med_history	208,259,928
Hf_f_microbiology	156,748,165
Hf_f_micro_susceptibility	112,372,182
Hf_f_procedure	112,153,607
Hf_f_surgical_case	5,047,689
Hf_f_surgical_procedure	396,476

Table 1: High scale of records involving FACT tables accessed within a big data health application.

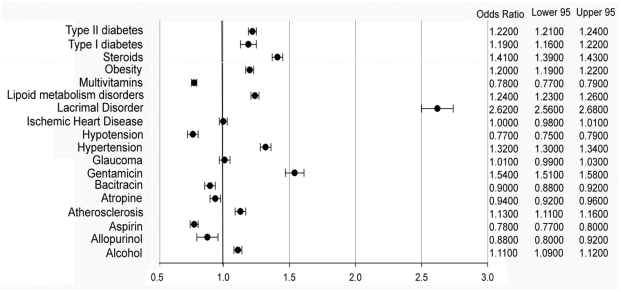

Figure 4: Risk analysis that highlights the nine risk factors pertaining to an ophthalmology study featuring a health big data application [11].

case study, which captures clinical events from hospital procedures, medication history, demographics, encounter, diagnosis, laboratory test, hospital information and billings. Over 600 individual sites from 90 health systems were participating in contributing data to the Health Facts database with billions of records as shown in Table 1.

Using the Health Facts database, the results obtained from statistical analysis [11] are shown in Figure 4. They pertain to nine risk factors with over a 20% increase in odds of developing the ophthalmological illness such as a cataract. With such an insight, researchers can have more informed methods in developing novel treatments and policies to restore sight to millions of people or prevent new ophthalmological illness cases.

The application workflow for Health Facts database use in our case study is shown in Figure 3. The lifecycle stages include acquisition, processing, analysis and collaboration. In order to have an effective big data analytics eco-system, the Health Facts database has to cater to a health big data application's performance and security requirements listed in the following sub-sections with an appropriate co-design of system considerations.

```
PROC SQL;
    CREATE TABLE PREJOINED AS
        SELECT IDSK.PATIENT_ID, IDSK.PATIENT_SK, IDSK.RACE, IDSK.GENDER,HF_F_ENCOUNTER.ADMITTED_DT_TM, HF_F_ENCOUNTER.ENCOUNT
        FROM IDSK AS L INNER JOIN HFFACTV3.HF_F_ENCOUNTER AS R
        ON L.PATIENT_ID=R.PATIENT_ID
        WHERE HF_F_ENCOUNTER.PATIENT_ID IN (SELECT PATIENT_ID FROM EYEVIS.EYE_ID);
QUIT;

PROC SORT DATA=PREJOINED NODUPKEY;
BY PATIENT_ID PATIENT_SK ADMITTED_DT_TM ENCOUNTER_ID WEIGHT HOSPITAL_ID;
RUN;

PROC SQL;
    CREATE TABLE JOINED AS
        SELECT PREJOINED.PATIENT_ID, PREJOINED.PATIENT_SK, PREJOINED.RACE, PREJOINED.GENDER, PREJOINED.ADMITTED_DT_TM, PREJOI
        FROM PREJOINED AS L INNER JOIN D_DIM_V3.HF_D_UNIT AS R
        ON L.WEIGHT_UNIT_ID=R.UNIT_ID;
QUIT;
```

Figure 5: Query execution of a big data health application case study: Health Facts using SAS (Statistical Analysis System) analytical tool

3.2 Performance Requirements

During the studies in [11] for determining the risk factors in health big data, significant performance hindrances were discovered that led to slow query responses. The usability was affected due to the inaccessibility of the data specifically to run large-scale queries for analytical purposes. The end users (clinicians, researchers) who are typically not experts in high-performance computing require automation for handling their queries that involve performing a data lookup at the high scale shown in Table 1. It took approximately 24 minutes to run a basic program using about 1% of the data available in the Health Facts database. For more complex programs using more of the data, it took 3 or 4 days to run as shown in a query captured in Figure 5. Such a performance output is prohibitively slow for researchers or clinicians who would need to wait excessively to test bold data-driven hypotheses and derive analytical insights for the future illness treatment or drug discovery. Such inefficient performance output in workflows can lead to wastage of valuable researcher time and cost, while also slowing the pace of innovation. In addition to such practical problems, additional bottlenecks in utilizing large health related databases can arise due to slow disk mechanisms used for data lookup, inadequate memory provisioning, and low-scale processing backends. Particularly when there are billions of records to be scanned within query look up in the tables, relevant user interfaces and appropriate system configurations need to be designed to obtain times for retrieval of relevant records in an efficient and effective manner for the users (researchers, clinicians, data scientists).

3.3 Security Requirements

The current security policies followed by major health data providers are a combination of policies given by EHR Vendors, IRB at local institutions [26] compliant with HIPAA, and server warehouse limitations within a campus environment. Each of the multi-source data (e.g., Health Facts, imaging data from scientific instruments, other healthcare vendors source data) are typically available in different formats and the researchers have the burden to ensure compliance in their studies across all these different source data with different

Type of Threat	Data Acquisition	Data Processing/Analysis	Data Collaboration	Exporting Data
Identity Spoofing	Illegitimate user can access data from database because of administrator error	Unauthorized users could be able to access data in computing sites due to malware	Access is required to store data in a repository	Clinicians/Re-searchers has to be authorized to access data from repository
Tampering of Data	Data is modified when in transit due to malware that might be present in the system	The data in transit from database to analytical tools could be modified because of human error	Data transit from the computing sites to storage in cloud	Data transferred from community cloud to a VDC should be encrypted
Elevation of Privilege	User is not permitted but able to access data from storage due to administrator error	Data analysis is performed by an underprivileged person	Person with low privileges may claim and access storage	Performance of resources could be hindered
Information Disclosure	Data flows in network get exposed if not encrypted	Disclose data using unauthorized access in computing sites	Not Applicable	Data could be exposed if not encrypted
Denial of Service	Compromise the database and illegally block access	Failure to access resources due to lack of permissions	Denial in data transfer to community cloud	Not Applicable

Table 2: Threat description for different lifecycle stages of a health big data application using a STRIDE model.

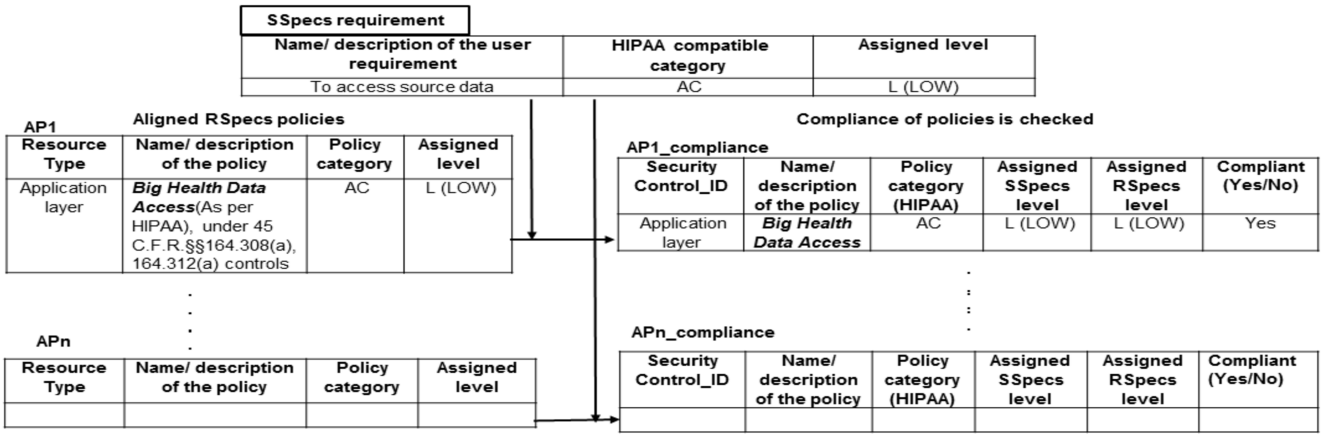

Figure 6: Checking for security compliance between HIPAA aligned user performance requirements (RSpecs) and security requirements (SSpecs).

standards. There is a need to homogenize these policies with a standard format that is compliant with NIST guidelines [9] for data classification levels as the data traverses across various domains with stringent policies [27]. Bearing in mind the challenges for compliance with diverse data sets with heterogeneous standards involving EHR Vendors, institutional policies and user requirements, it is essential to have end-to-end security alignment with a secure gateway for data access.

4 COMMUNITY CLOUD SOLUTION DESIGN

4.1 Threat model

The community cloud design has to consider different types of threats involved in the life cycle stages of a health big data application such as the one in our case study with the Health Facts database. For the threat modeling, we adapt the Microsoft's STRIDE model [28] as shown in Table 2 for an application using e.g., Health Facts database that is semi-automated currently at different lifecycle stages. In our investigation of the case study requirements, we have noted that a more common concern for users at the data-level relates to the data access levels (i.e., Loss of Confidentiality) for data-in-motion and data-at-rest and data-in-use within the lifecycle stages. There is concern of unintended users having access to data due to provisioning of excessive privileges amongst the various roles (e.g., clinician, researcher, data scientist, server admin). This is particularly important with healthcare data as it is typically confidential data that must be kept private. Loss of Integrity, where data may be corrupted due to administrator error in handling databases, or if a user action causes disk space to exceed for an analysis process. This is a major concern for those who are analyzing the data. Lastly, there is a threat of users not being able to access their data when needed (i.e., Loss of Availability) due to e.g., administrator error involving inadvertent system management actions which cause major changes to storage that may not get notified to users in a timely manner, which may then result in partial or full loss of their data. We also seek to study threat models for above issues in the context of a 'data safety' assessment that is personalized to the users depending on their security requirements. This results in loss of data and wasted labor hours, if a user is in the middle of an acquisition or analysis process.

4.2 Security Alignment and Compliance

One of the major hurdles for performance and security co-design driven resource management is that in most of the cases the domain security requirements are diverse. Thus, we need to homogenize the different data forms and align the security and performance

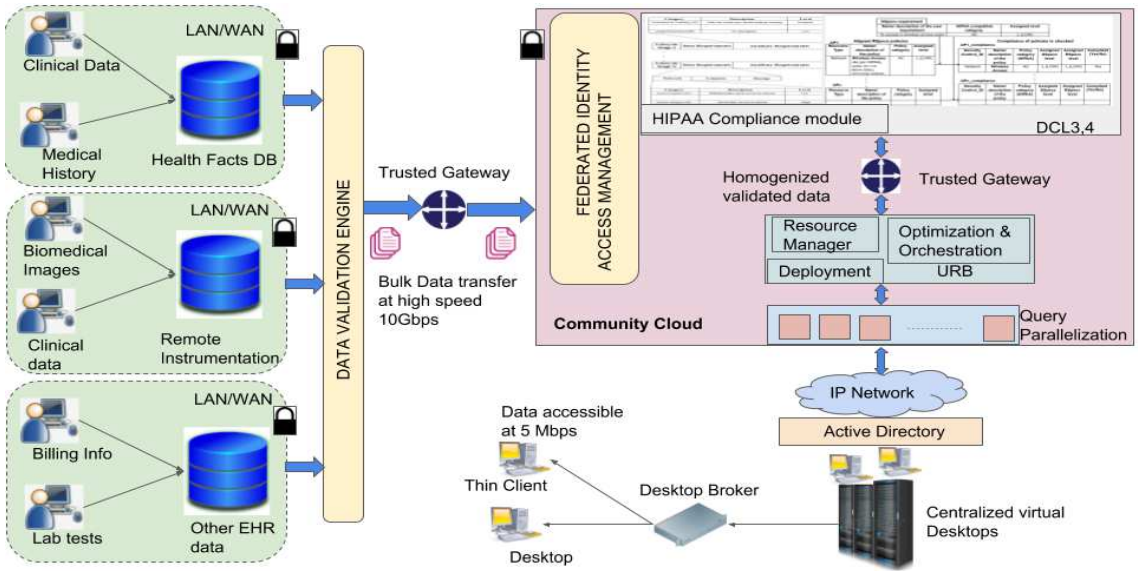

Figure 7: Integrated community cloud design with security compliance of heterogeneous data formats using a unified brokering structure to enhance the performance of querying data for analysis of a big data health application

requirements for compliance. However, the data classification levels for multi-domain resources are of different levels within enterprises i.e., they could be data classification level (DCL) 3,4. As a result, it becomes relatively difficult to align such disparity in classification levels. To address this issue, we use a novel alignment scheme detailed below to homogenize such diverse security requirements into homogeneous policy statements adapting the method discussed in [27]. As the Health Facts database comprises of sensitive healthcare data from day-to-day patient encounters, there is a need to utilize the dataset in a manner that ensures it is consistent with applicable regulations and has appropriate compliance levels when users are trying to access and analyze data. The policies and procedures for utilization of Health Facts database is discussed in Health Facts Access and Data governance [29].

SOM-R-007: *Users will access Health Facts through a secure server that is DCL3 compliant* is a Health Facts policy [29] which is about data governance that is compliant with HIPAA.

We term user requirements in a health big data application using various health data/resource providers as RSpecs and corresponding security requirements as SSpecs. The user requirements are to be aligned with providers' resource policies as shown in the Figure 6 in a manner that is HIPAA compliant. The alignment results in a homogenization of different data formats into a unified format that is HIPAA compliant and can be used with high-performance in a community cloud architecture within a Unified Resource Broker (URB) as shown in Figure 7. To check the compliance between the resource and security specifications, we consider the rules mentioned in NIST SP800-53 [9] where different security controls are classified as Low, Medium, and High. Now with the mapping document between NIST and HIPAA rules discussed in [10], we can apply similar categorical levels (Low, Medium, High) within our Health Facts based health big data application case study involving ophthalmological illness.

4.3 Architecture Components

Our proposed community cloud architecture shown in Figure 7 involves various components such as Source Data, Data Validation Engine, Secure and Trusted gateway, Unified Resource Broker (URB), Virtual Desktop Clients (VDC) with easy-to-use health big data application User Interface (UI). Data validation is performed for effective performance before uploading external data into the community cloud for a user study purpose. To ensure that existing security rules implemented in the Health Facts application are not violated, we make sure the Personal Identified information is not taken into the cloud platform as the researchers require mostly diagnosis information. Any sensitive data (personal information, Medical Record Number (MRN) that is tagged along with the extracted information by the researchers for the analysis, will be assigned a new temporary value for each new session the researcher tries to access the database. This ensures that none of the Personal Identified information is pulled into the cloud platform, and the data that is captured as part of the query run by the users is validated before use. In addition to this, as shown in Figure 7 a federated Identity Access Management (IAM) is used to authenticate a user who tries to access the analytical data with appropriate authorized privileges.

Considering the multi-format in this input data, our proposed community cloud architecture homogenizes the data classification levels and ensures all the alignment is of HIPAA compliance. After checking the compliance of the data with Health Facts policies (HIPAA compliant), to ensure secure transfer of data throughout the system, a secure gateway is used. This secure gateway has the implementation of appropriate security controls to enable the users to run the analysis on the queried data using relevant computation tools. To ensure high performance for the users at the VDC's as

shown in the Figure 7, our approach requires a query parallelization mechanism in the community cloud. The parallel querying module enables the users to run more than a query on the data pulled into the community cloud by the researchers. In the current Health Facts application case study, parallel queries are executed as different SAS sessions with equal share of resources to compute. However, this will lead to a bottleneck in performance as there might be a few queries which require less resources than the equal share currently being implemented in Health Facts application. Thus, our proposed architecture implements a prioritized queue for the queries to run using an URB as shown in Figure 7. The functionality of this URB is used to manage the resource needs of a user, by ensuring that the required amount of resources are available to service the users (researcher, clinicians, data scientists) in order to fulfill their analytics objectives. At the end-user side, data could be accessed by clinicians/researchers only through VDC which will be deleted for every new session to avoid any risk of unauthorized sensitive data downloads.

5 CONCLUSION AND FUTURE WORK

Smart healthcare with big data analytics has great potential to foster medical advancements in areas such as precision medicine, determination of risk factors, and biomedical research. In this paper, we proposed a novel community cloud architecture to automate any existing semi-automated big data health application that uses large healthcare related databases to test bold hypotheses. Our approach involves a co-design of high-scale performance and security compliance through alignment of user requirements and data provider policies via a Unified Resource Broker module. Using a case study involving a ophthalmological illness data analysis use case with multiple data sources (e.g., a Health Facts database, imaging data from scientific instruments, I2B2, Millennium), we describe how our community cloud architecture can mitigate the query response latency in running large-scale queries over billion transaction records, while also ensuring compliance with heterogeneity in the data classification levels in the various lifecycle stages of the health big data application.

Our future work is to extend and implement best practices related to several health bigdata systems such as “one health”. The “one health” efforts require analysis of the similarity in genomic information between veterinary and human health care data in order to develop new forms of treatments, and foster drug discovery. The best practices could also feature more intuitive user interfaces for non-experts in high-performance computation to easily query using parallel programming APIs (application programming interfaces).

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under award number OAC-1827177. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] K. Suh, S. Sarojini, M. Youssif, K. Nalley, N. Milinovikj, F. Elloumi, S. Russell, A. Pecora, E. Schecter, A. Goy, “Tissue Banking, Bioinformatics, and Electronic Medical Records: The Front-End Requirements for Personalized Medicine”, *Journal of Oncology*, 2016.

[2] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, “Knowledge Discovery and Data Mining: Towards a Unifying Framework”, *Association for the Advancement of Artificial Intelligence*, 1996.

[3] M. Bergner, “Quality of Life, Health Status, and Clinical Research”, *Advances in Health Status Assessment*, 1989.

[4] What is Big data?. [Online]. Available: https://datascience.nih.gov/bd2k/about/what [Accessed 10-2-2018].

[5] Genomics and translational bioinformatics trending advancements and their working groups.[online]. Available: https://www.amia.org/programs/working-groups/genomics-and-translational-bioninformations [Accessed 10-12-2018].

[6] J. Tenenbaum, P. Avillach, M. Benham-Hutchins, M. Breitenstein, E. Crowgey, M. Hoffman, X. Jiang, S. Madhavan, J. Mattison, R. Nagarajan, B. Ray, D. Shin, S. Visweswaran, Z. Zhao and R. Freimuth, “An informatics research agenda to support precision medicine: seven key areas”, *Journal of the American Medical Informatics Association*, 2016.

[7] PCI DSS standard. [Online]. Available at: https://www.pcisecuritystandards.org/pci_security [Accessed 10-15-2018].

[8] “An Introductory Resource Guide for Implementing the Health Insurance Portability and Accountability Act (HIPAA) Security Rule”, *NIST Special Publication 800-66 Revision 1*, 2008.

[9] “Security and Privacy Controls for Federal Information Systems and Organizations”, *NIST SP800-30 Technical Report*, 2013.

[10] “HIPAA Security Rule Crosswalk to NIST Cybersecurity Framework”, *https://www.hhs.gov/sites/default/files/nist-csf-to-hipaa-security-rule-crosswalk-02-22-2016-final.pdf*, 2016.

[11] M. Raju, M. Chisholm, A. Mosa, Greg Petroski, Chi-Ren Shyu, and Frederick W. Fraunfelder, “Investigating Risk Factors for Age-Related Cataract Using the Cerner Health Facts Database”,*Journal of Eye and cataract surgery*, 2017.

[12] P. Nguyen, S. Konstanty, T. Nicholson, T. Brien, A. Schwartz-Duval, T. Spila, K. Nahrstedt, R. Campbell, I. Gupta, M. Chan, K. McHenry, N. Paquin, “4CeeD: Real-Time Data Acquisition and Analysis Framework for Material-related Cyber-Physical Environments”, *17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, 2017.

[13] Introduction to relational database management systems in Oracle. [Online]. Available: https://docs.oracle.com/cd/E11882_01/server.112/e40540/intro.html [Accessed 10-14-2018].

[14] P. Calyam, S. Rajagopalan, S. Seetharam, A. Selvadhurai, K. Salah, R. Ramnath, “VDC-Analyst: Design and Verification of Virtual Desktop Cloud Resource Allocations”, *Elsevier Computer Networks Journal (COMNET)*, 2014.

[15] P. Calyam, S. Rajagopalan, A. Selvadhurai, S. Mohan, A. Venkataraman, A. Berryman, R. Ramnath, “Leveraging OpenFlow for Resource Placement of Virtual Desktop Cloud Applications”, *IFIP/IEEE Intl. Symposium on Integrated Network Management (IM)*, 2013.

[16] Apache hadoop. [Online]. Available: https://hadoop.apache.org/ [Accessed 10-14-2018].

[17] Statistical analysis and software tools (SAS) documentation. [Online]. Available: https://www.sas.com/en_us/home.html [Accessed 10-14-2018].

[18] S. Oh, J. Cha, M. Ji, H. Kang, S. Kim, E. Heo, J. Soo Han, H. Kang, H. Chae, H. Hwang, S. Yoo, “Architecture Design of Healthcare Software-as-a-Service Platform for Cloud-Based Clinical Decision Support Service”, *Healthcare Informatics Research*, 2018.

[19] Getting your data ready for precision medicine. [Online]. Available at: https://www.ibm.com/blogs/insights-on-business/healthcare/getting-data-ready-precision-medicine/ [Accessed 10-15-2018].

[20] Community cloud architecure for salesforce health care applications. [Online]. Available: https://www.salesforce.com/products/community-cloud/faq/ [Accessed 10-15-2018].

[21] Google cloud for Healthcare: new APIs, customers, partners and security updates. [Online]. Available at:https://www.blog.google/products/google-cloud/google-cloud-healthcare-new-apis-customers-partners-and-security-updates/ [Accessed 10-15-2018].

[22] C.L. Chen, C. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data”, *Information Sciences*, 2014.

[23] H. Takabi, J. Joshi, G. Ahn, “Security and Privacy challenges in cloud computing environments”, *IEEE Security and Privacy Journal, Vol. 8, No. 6, pp. 24-31*, 2018.

[24] L. Kaufman, “Data Security in the World of Cloud Computing”, *IEEE Security & Privacy, Vol. 7, No. 4)*, 2009.

[25] W. Pieters, T. Dimkov, D. Pavlovic, “Security Policy Alignment: A Formal Approach”, *IEEE Systems Journal, Vol. 7, No. 2, pp. 275-287*, 2013.

[26] Institutional Review Board at University of Missouri. [Online]. Available at: https://research.missouri.edu/irb/ [Accessed 10-11-2018].

[27] M. Dickinson, S. Debroy, P. Calyam, S. Valluripally, Y. Zhang, R.B. Antequara, T. Joshi, T. White, D. Xu, “Multi-cloud Performance and Security Driven Federated Workflow Management”, *IEEE Transactions in Cloud Computing*, 2018.

[28] Microsoft threat modeling tool preview and template documentation. [Online]. Available at:https://docs.microsoft.com/en-us/azure/security/azure-security-threat-modeling-tool-getting-started [Accessed 10-15-2018].

[29] “Health Facts access and data governance”, [online]. [Accessed 10-11-2018].