A HEALTH RESEARCH COLLABORATION CLOUD ARCHITECTURE

Mugonza Robert[1] and Basaza-Ejiri. H. Annabella[2]

[1]Department of Computer Science, Mbarara University of Science and Technology
[2]College of Computing and Engineering, St. Augustine International University

ABSTRACT

Cloud computing platforms used for research collaborations within public health domains are limited in terms of how service components are organized and provisioned. This challenge is intensified by platform level challenges of transparency, confidentiality, privacy and trust. Addressing these collaboration issues will necessitate that components are reorganized. There is a need for secure and efficient approaches of reorganizing the service components, with trust to support collaboration related requirements. Through iterative design, a reliable and trust-aware re-organization of cloud components – the Collaboration cloud architecture is achieved. We utilize SOA, Privacy-By-Design principles and insights from blockchain to enforce trust. We illustrate its potential with multi-layer security process flow based on server-to-client tokens, Role-based Access Control and the traditional authentication – username and password with assurance for privacy and trust. The architecture shows promise towards data governance and the overall management of internal or external data flows.

KEYWORDS

Cloud, Computing, the blockchain, health, research, collaboration, architecture

1. INTRODUCTION

Collaboration over the cloud is an opportunity of co-authoring and making available computer files for sharing through cloud computing, with users uploading files that they wish to share amongst themselves[1]. Best collaboration tools and applications have been identified as those that possess these usage characteristics. These tools and applications need to:- provide for real-time commenting and messaging, enhancing the speed of object delivery; provide for resource monitoring enabling identification of when others are active on documents owned by other persons; be secure and address trust relationships during data sharing for example, enabling permissions to be set on resources and managing of activity profiles, activity feeds and email alert profiles to keeping up-to-date with the latest activities per resource usage. Above all, they need to enable collaboration, file sharing internally and externally through a firewall of some sort with compliance to underlying security and existing Information Technology (IT) compliance frameworks; and more so, ensure full audit-ability of shared resources while at the same time reducing workarounds for sharing and collaborations on much larger files. When observed from health and biomedical areas, the developers of applications, providers and Health IT will often face highly demanding requirements for confidentiality, application and system-wide transparency, and ultimately the resulting system's reliability as detailed in our literature survey and empirical review studies [2, 3].More so, while computing resources have advanced and grown exponentially[4], multidisciplinary collaborations are growing too dramatically and the

DOI: 10.5121/ijccsa.2018.8503

resulting data as well with enormous capability for creating new insights for increased innovativeness, discovery and quality of life.

Guided by the National Institute of Sciences and Technology (NIST) [5] and the European Union Agency for Network and Information Security (ENISA), Cloud computing enables on-demand and convenient broad network access to shared pool of configurable computing resources like servers, storage, networks, applications, and services, rapidly provisioning them with minimal service provider interaction. And besides its benefits, the computing approach's challenges: security, privacy, and trust, continue to impede its wide acceptance among collaborations. In this research, our interests revolve around an efficient way of re-organizing the cloud service components to support collaboration related needs.

2. RELATED WORKS

2.1. SERVICE ORIENTED PROVISIONING

In the perspective of a service-oriented computing, a number of service architectures exist. This support for service orientation presented in architectures including Service-oriented Architectures (SOA)[6-8], shared architecture and the Open Systems Interconnection (OSI) architecture[3], and now, the Cloud Computing Open Architecture (CCOA)[9]. We utilize the architectural notions to put to life to a new body of knowledge of Collaboration Cloud Architecture (CCA), well suited for Health and biomedical research collaboration.

Service-oriented Architectures advance computing in a sense that services are provided to the other components by application components, through a communication protocol over a network. The term is often applied to software development to allow application developers to combine software functionality to form applications, built purely from existing. The concept of service orientation has led to computing developments including Seeded Cloud and the "Three tire web business" architecting. Its usage in computing has afforded a number of service characteristics: componentization, reusability, extensibility, and flexibility. Service orientation is guided under a number of predefined standards as reviewed with [4, 8-12].

2.2. SOA DESIGN PRINCIPLES

Service-oriented provisioning will adhere to SOA design principles. The principles give a base for any architectural development though these combined reflect upon service orientation for the case for a cloud computing architecture (CCA). For a cloud architecture, Zhang [9] presents a number of design principles upon which a CCA can build. The design of the Collaboration Cloud is based on these. These briefly include cloud ecosystem enablement; infrastructure management via virtualization, service orientation, provisioning and subscription, quality analytics, information architecting and management among the many others. These design principles are of key affordance, as they help isolate details of one module/ functionality, supporting reusability and ensure the ultimate service orientation benefits.

With resources (services) offered over the internet broadly ranging from data, hardware, database, software and platform among others, the public health and biomedical research domains introduce specific resource demands: the need to provide continuous and longterm infrastructural support for data and its related functions including extending simple data functions like sort to include data-intensive algorithms, mathematical and statistics; and more so, the need to leverage the existing tools' support for innovative and advanced techniques for data mining – Extraction, Transform and Loading, and massive storage.

More closely, Demirkan[13] discusses that there are often data challenges for organizations in managing large amounts of data. It is envisaged and yet still observed that these data continue to grow exponentially due to a decrease in costs for storage, digitization of data and information, and collection devices including mobile phones, sensors networks, laptops, data systems, sequencing machines. For example public health collaborations including the cancer studies, Remote monitoring studies on Human Immune Virus (HIV) and Tuberculosis (TB) that participated in the empirical study in [2], present ever increasing datasets of patients adherence behavior and we believe that this data will surpass the available capabilities of the resource storing the datasets.

Apart from the challenges identified, digitizing data has got numerous opportunities across domains/sectors allowing for context-specific aggregation and analysis of the data. For instance device information accumulated of TB patients can ease the works of the doctors to accurately managing the disease, identifying dozes missed, increasing adherence, and through this, the spread will be curbed down as well as the health cost for the community where the patient comes from, thus ensuring and assuring the overall health care quality and efficiency.

2.3. SUPPORTING CYBER-INFRASTRUCTURE

More recent funded developments in biomedical and public health have attempted to address domain researcher concerns including – Cyber Infrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA)[17]; CyberHealth for Aggregation, Research and Evaluation(CARE)[18] – that Utilizes the seeded approach to cloud computing to provide for discovery, access and visualization of data; the Integrating Data for Analytics and Anonymization and Sharing (iDASH) - focuses on design of algorithms and tools for data sharing in a privacy-preserving mechanism [19]; the Cyber Infrastructure for Comparative and Effectiveness research (CyCORE) project - focuses on improvements in Comparative Effectiveness Research data through a proposal to build a platform for improved data collection, storage and sharing in cancer clinical trials [14]; CitiSense – a geospatial improvement in air quality using wireless personal exposure monitoring systems [20]; among others. These initiatives are still at their infancy stage and not suitable for situations that require appropriate access to datasets under the collaboration. The existing cloud environments are further limited with the inability to openness, lack trust with regards to sharing of sensitive data, complicated collaboration between scientists from across the world, and re-use of micro data for secondary research as well their ability to provide for consultations and actual usability with and for all stakeholders, limiting support for mutual respect of values, privacy and trust plus actions for reliable partnerships.

2.4. SERVICE LEVEL PROVISIONING

Cloud provisioning is further challenged with often inabilities to interoperate among providers [14], and due to the over-dependence onto one provider, their design intensifies data lock-in

issues[11], the inflexible nature of cloud application interfaces caused by limited User Interface (UI) frameworks [6, 15]. Often implementation and deployment lack standard SLAs required for business service deployment, making it difficult to extend multi-tenancy support across multiple collaboration clients. The reservoir architecture is a development that presents support for federation of multiple cloud providers and guarantees the appropriate SLAs[16].
SLA based provisioning suggested in [7, 17] and SOA based approaches [9, 11, 18] can provide base support for leveraging clouds for reliability, trust, integrity and appropriate access to resources in collaboration based clouds – next to the use of service brokering. The resulting architecture will ensure that supply and demand of cloud resources are regulated to achieve market equilibrium and promote Quality of Service (QoS) based resource allocation mechanisms. Key components of this architecture include the SLA based Resource Allocator consisting of a service request examiner, service request dispatchers, service access controller and virtual machines (VMs) monitors. It is based on the feedback from the monitoring performed by VM and service request monitors, the request dispatchers route requests from consumers to resource providers, directly meeting the desired QoS.

2.5. PRIVACY BY DESIGN (PBD)

The other challenge of the collaboration based on cloud computing is Privacy. Adopting Privacy by Design foundational principles in[19] will ensure that collaboration data is protected while in transit and within the cloud, maintains appropriate access to it, while at the same time ensuring its integrity. Addressing the identified collaboration environment challenges will necessitate that service components are reorganized and devise better approaches that observe these principles. Through this, we seek an efficient way of reorganizing the service components to support collaboration based cloud requirements.

The rest of this paper is organized as follows. Section-2 describes the required collaboration cloud requirements. Section-3 describes the proposed architecture. Section-4 discusses the architectural designs approaches the resulting views, the application use cases, and implication of the design to data governance. The last section describes our next steps and concludes as well.

3. REQUIREMENTS ELICITATION

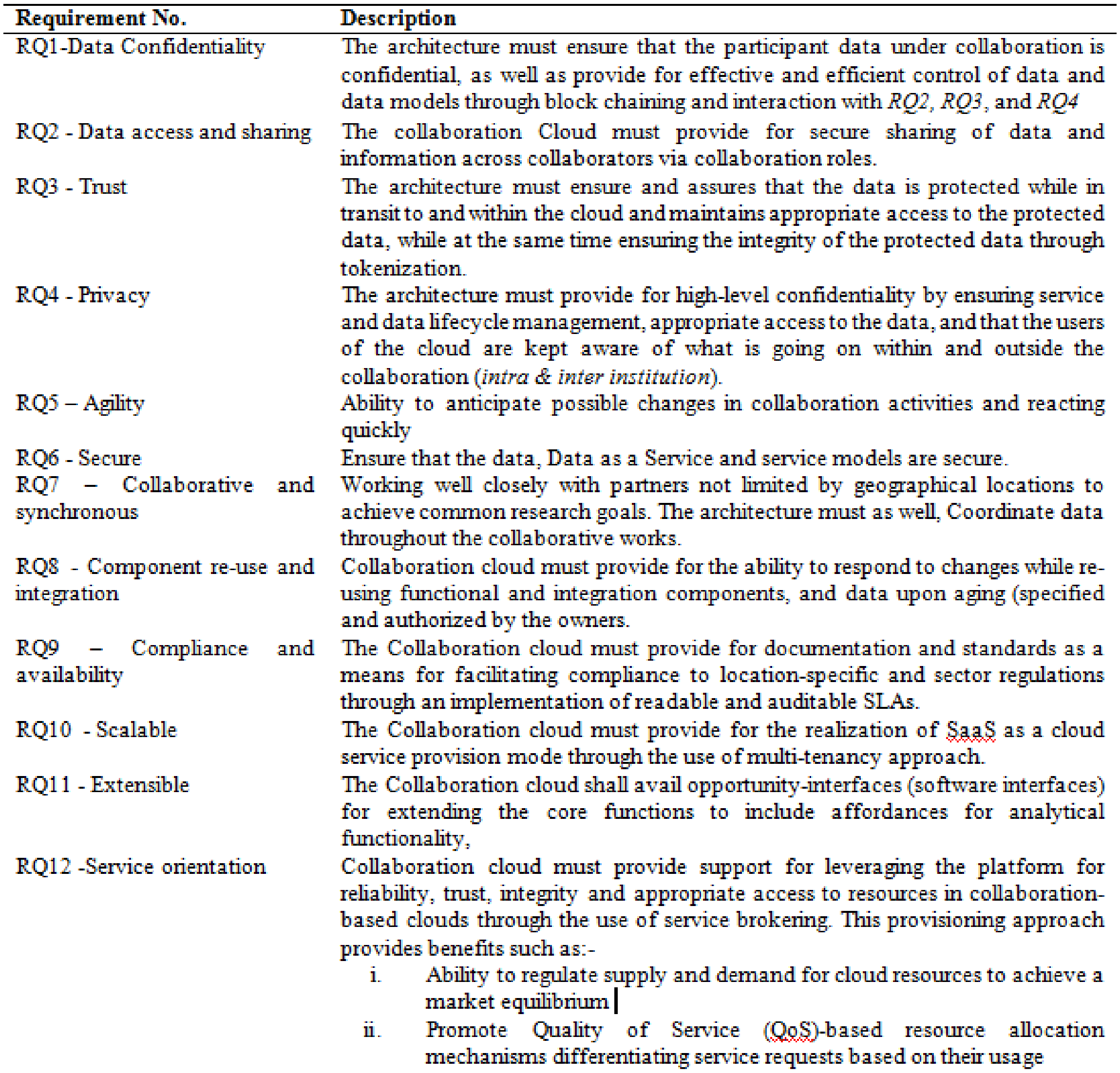
Input for the Collaboration Cloud Architecture originates from the complexity of public health and biomedical health research, the computing ecosystem: researchers themselves, the providers of the ‘tools and applications' referred as services; as well a strong body of knowledge and information from manuscripts and research articles, and proceedings and reports from conference, workshop, and symposia. More so, the empirical reviews conducted on issues affecting health research collaborations dependent on the cloud in[3] are pivotal here. Challenges identified in the surveys and reviews especially concerns relating losing control over sensitive data outside the trust boundary[20, 21]. The proposed architecture aims to address many of the related security, trust and privacy requirements collected from our recent studies.

Key requirements for the architecture included:- the need for trust, adequate security, increased data control, ownership, confidentiality, integrity, access, increased SLA integration, availability, ensuring flexible data collection facilities, storage and accessibility to a broader collaboration datasets, appropriate sharing, and increased privacy among others and summarized in Table-1.

These requirements are furthered with usability concerns arising from a provisional analysis, a case for mixed collaborations, with researchers:-Principal investigators and research assistants differ from the northern partners along with a number of perceived characteristics. These include - type, format, and form of data involved; the type of storage; the level of research skills and knowledge; accessibility to data and level of involvement.

Articulating such characterization is vital and critical in establishing successful health research collaborations, and more so, enables translation of the needs into complete and traceable requirements for collaboration systems.

Table-1: A summary of the requirements for the Collaboration Cloud.



Requirement No.	**Description**
RQ1-Data Confidentiality	The architecture must ensure that the participant data under collaboration is confidential, as well as provide for effective and efficient control of data and data models through block chaining and interaction with *RQ2, RQ3*, and *RQ4*
RQ2 - Data access and sharing	The collaboration Cloud must provide for secure sharing of data and information across collaborators via collaboration roles.
RQ3 - Trust	The architecture must ensure and assures that the data is protected while in transit to and within the cloud and maintains appropriate access to the protected data, while at the same time ensuring the integrity of the protected data through tokenization.
RQ4 - Privacy	The architecture must provide for high-level confidentiality by ensuring service and data lifecycle management, appropriate access to the data, and that the users of the cloud are kept aware of what is going on within and outside the collaboration (*intra & inter institution*).
RQ5 – Agility	Ability to anticipate possible changes in collaboration activities and reacting quickly
RQ6 - Secure	Ensure that the data, Data as a Service and service models are secure.
RQ7 – Collaborative and synchronous	Working well closely with partners not limited by geographical locations to achieve common research goals. The architecture must as well, Coordinate data throughout the collaborative works.
RQ8 - Component re-use and integration	Collaboration cloud must provide for the ability to respond to changes while re-using functional and integration components, and data upon aging (specified and authorized by the owners.
RQ9 – Compliance and availability	The Collaboration cloud must provide for documentation and standards as a means for facilitating compliance to location-specific and sector regulations through an implementation of readable and auditable SLAs.
RQ10 - Scalable	The Collaboration cloud must provide for the realization of SaaS as a cloud service provision mode through the use of multi-tenancy approach.
RQ11 - Extensible	The Collaboration cloud shall avail opportunity-interfaces (software interfaces) for extending the core functions to include affordances for analytical functionality,
RQ12 -Service orientation	Collaboration cloud must provide support for leveraging the platform for reliability, trust, integrity and appropriate access to resources in collaboration-based clouds through the use of service brokering. This provisioning approach provides benefits such as:- i. Ability to regulate supply and demand for cloud resources to achieve a market equilibrium ii. Promote Quality of Service (QoS)-based resource allocation mechanisms differentiating service requests based on their usage

4. THE PROPOSED COLLABORATION CLOUD ARCHITECTURE

The design of the Collaboration Cloud Architecture has benefited from research meetings[2, 5, 22, 23], the researchers' recommendations – including the provision of a flexible platform that takes into consideration for security in electronic data capture, the increasing need for secure and unified storage, and reliability in data sharing in collaborations can be achieved more reliably and equitably using cloud computing though this time, innovatively.

The design of the Collaboration Cloud given the service requirements specified in section II is achieved through an iterative design methodology. The design approach though similar to incremental approaches, follows conventional prototyping, through testing, analysis and then refinement of the product or process. We utilize this approach under consideration of design requirements, and integration of a number of architectural design principles, including Service Orientated Architectures (SOA), privacy-by-design and blockchain design principles. The integration of approaches contributes a characteristic cloud environment suitable for research collaboration.

Initial attempts to sort and reorganize the cloud provisioning components presented a vague prototype. This underwent the iterative design process through the refinement phase of the 1st cycle. Subsequent iterations were conducted with improvements, producing various versions at least trying to meet the collaboration cloud requirements.

The resulting architecture was tested on application use cases to satisfy the elicited design requirements including the reliability of the cloud environment. At this point, we didn't test for usability since the focus was directed to key functional components.

5. ARCHITECTURAL DESIGN – VIEWS

Architecting the Collaboration Cloud for Public Health and Biomedical research requires meeting the following objectives as proposed by Zhang[9]. Ensuring a consistent and effective cloud architecting will require a) devising innovative and scalable approaches and enabling provisioning of configurable resources of the cloud; b) propose common and shareable services for the construction of the cloud, provide business services and other cloud offerings to consumers in a unified approach; c) maximize the potential business value of the cloud via an extensible IT infrastructure and management system.

The architectural design observes seeding as a multi-tenancy approach for scalability and seeding to ensure the overall platform reliability while providing the expected services to researchers. We argue that the possibility of unifying storage (access, sharing, and reference) through establishing a singleton cloud storage for every instance of the collaboration where every participating institution in the collaboration is offered logic units for collecting, storing and making accessible the collected data to all parties under the collaboration, - as seeding. Seeding as an approach considers that the institutional logic container instances of the common storage assigned to the collaboration as *'seeds'* as opposed to the approach proposed by attempting to architect a health CI while observing the principles and characteristics of cloud computing. Their approach describes seeds as datasets and technical tools to be used for interacting with the data and associated knowledge communities. These efforts are vital, as they attempt a converging heterogeneous data. The approach does not provide for assuring privacy, integrity, and the overall platform reliability.

5.1. THE TIRED VIEW OF THE COLLABORATION CLOUD ARCHITECTURE

Based on the SOA, Block-chain and the Seeding approaches, the Storage cloud architecture for research collaborations self-organizes its components and services into service tires as shown in fig-2. L-R, Tire-1 is the client, accessible via a web browser, is the point of interaction of the researcher with the cloud system providing affordance for data uploads, downloads, and other data functionalities.

Tier-2, the service broker charged with interfacing services from the server for the clients, is enabled by NoSQL. This generic code enables service level brokering through the implementation of web service calls, prescribed well in Web service definition language (WSDL). Unlike the standard three tier web architecture; the open service broker replicates the service level contract/agreement with sync function at both the collaboration instance and at the server.

The 3rd Tier, the server is an application developed with functionality for service provisioning including facilities for shared storage, access, sharing among others; implementing closely with the service broker, SLA based authentication similar to the blockchain ledgers.

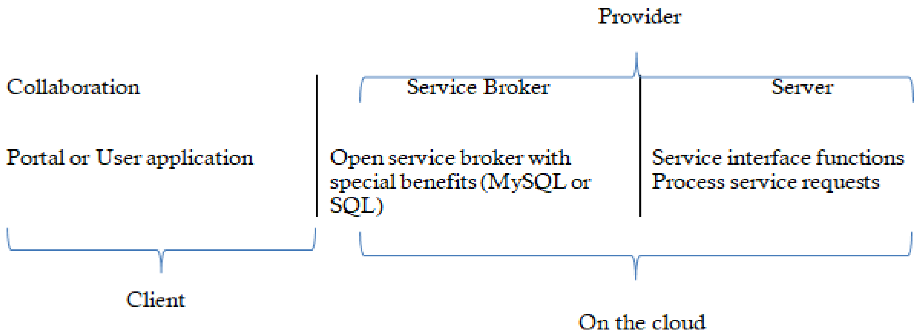

Provider
Collaboration
Service Broker
Server
Portal or User application
Open service broker with special benefits (MySQL or SQL)
Service interface functions Process service requests
Client
On the cloud

Fig-1: The 3-Tire view of the Health Cloud Collaboration architecture

Tier-2 & 3 if combined, form a single comprehensive tire - the provider Tier, ensuring that service provisioning and a service request examiner, service request dispatchers, service access control and virtual machines (VMs) monitors/ virtualized cloud server instances.

5.2. The logical view of the Collaboration Cloud Architecture with seeding.

The goals of this architecture is a) to provide for trust and security in terms of integrity, confidentiality, access control, auditability, and the application's overall reliability; b) provide for flexible electronic data capture, the increasing need for secure and unified storage, and reliability in data sharing in collaborations; while providing for the core functionalities more reliably and equitably. Multi-tenancy coupled with measures for privacy and trust is one cloud computing characteristic that attempts to address these goals.

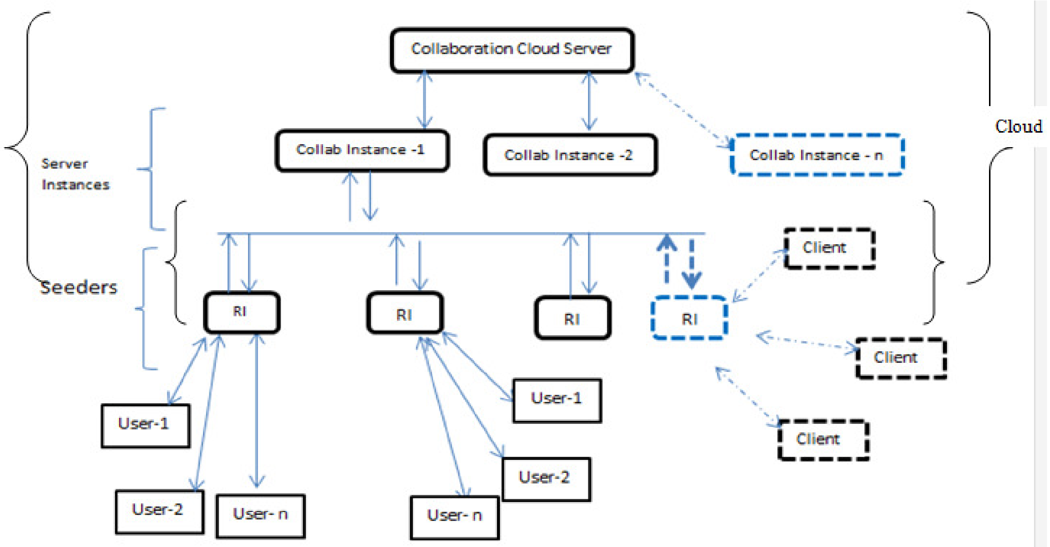


Fig-2: The seeded architecture for research collaboration with multi-tenancy

The Collaboration cloud is organized as follows:

a) The Server Application through use of virtualization and instantiation technologies like Docker and VMware is provisioned to various collaboration projects upon request for a collaboration environment by teams of researchers represented by their research institutions, the *collaboration instance*.
b) The collaboration cloud server instance is a logical view of the server application almost as a local application providing an environment for data storage, access, sharing and transfer; houses the open services broker –*services management* Including exposing the services and providing a standardized means for consuming the cloud services. It serves as well as a registry-like facility required for implementing a ledger, as a means for restricting and when drawing a new collaborating institution. The instance provides affordance for common storage and collaboration core functions.
c) The seeders also called clients – minute software code downloaded and installed on the application user's computer that is maintained like blocks connected to a collaboration server instance. The clients connect to the collaboration instances through the use of

tokens and identifiers (IDs), accessible to all researchers being served by a common storage and collaboration cloud functions – sharing, download, among others.

d) The application users – are simple devices at the bottom of our SaaS model that interact with the cloud for storage and application logic - basic data handling tools and cloud functionality (provisioned as services). These interact with the client software authenticated using a combined access control measures: "*Traditional Authentication*" technology and Role-based *Access Control (RBAC).* With traditional authentication assigns privileges directly to users using – *username and password* to access the institutional server while RBAC ensures that privileges are granted directly to roles. For instance, abstract positions in research collaboration may include different assignments and responsibilities, and access different items of the collaboration. RBAC implementation is advantageous in a number of ways. It reduces management costs of establishing and managing access control policies as compared to the previous one – traditional authentication list. This as well, will enable scalability of similar policies for a wider user base and thus suitable for inter-organization collaboration settings.

The collaboration Cloud Architecture for biomedical and public health collaboration utilizes a multi-layer security framework that is based on server-to-client tokens with component IDs and the traditional authentication for identification. The instantiation enabled either through dockering or Virtual machines, provides for a multi-tenant based approach for accessing similar resources housed in the Collaboration Cloud server without jeopardizing the privacy and security of other tenants.

More so, similar to blockchain illustrated by Guy and Oz[24], the Collaboration Cloud functions as a series of connected nodes like in Fig-4. We have organized it in a sense that each node "RI" serving as a client with a task of validating and relaying transactions (synchronizing) with the Collaboration instance resources. Each node RI is an administrator with similar roles as every other member under the collaboration and joins based on the SLA agreed with during the creation of the collaboration and membership requirements, also as specified in the collaboration goals and objectives. An automated access code, a hash code is provided by the collaboration instance through a tokenization process. And while serving as an admin, each member is assigned a role in the collaboration. All the transactions in the collaboration including the addition of a new member are recorded and shared, visible open to all collaborators.

Similar to the blockchain DB, a distributed database is maintained, built on NoSQL, often reconciled synchronously with seeds. Each member contributes to the distributed resources of the collaboration including the distributed Database, hosted in the Collaboration Instance through a synchronizing a copy of the data that is at the RI's server. Thus, every member maintains their own database whilst the core/main database is as well populated through replication.

More so, the Collaboration SaaS is multifaceted with mixed database communication strategies, implementing both a centralized and decentralized data flows, just in case, with requirements for guaranteed data availability and zero data alteration effect from external threats including the hackers and failures at the RIs. This is theoretically and practically feasible. Reconciliation based computing with the sync function, will enable the Central DB to update any RI DB in case of malicious damage and assuming the central DB has been attacked, self-healing has been implemented to enable clean wipe and data reconciled at the with the various RIs for uniformity, integrity, and data quality assurance.

Independent data collection service access - The collaboration cloud exposes and avails data collection facilities, integration towards unified storage and sharing as services and this can be harnessed appropriately well by the use cases in regards to a) providing flexibility during data collection and integration from heterogeneous data sources (enabling for eliciting from broader data sources like remote sensors, wearable devices, patient mobile devices including also phones); b) equitable and dependable sharing of resources among collaborators and the external world with assurance for increased integrity and reliability. For example, applying the blockchain approach (of the open ledger) on the research collaboration within the cloud architecture, collaborators are assured of integrity while data is hosted and on the go (being shared) through implementing of an independent registry at the research institution (RI) server that syncs to the open collaboration registry provided by the collaboration server instance (also called the service broker). Using the approach, each collaborator accesses their own data-store (minute copies of collaboration specific data) and interfaces for the collaboration cloud services, implementing provider level privacy.

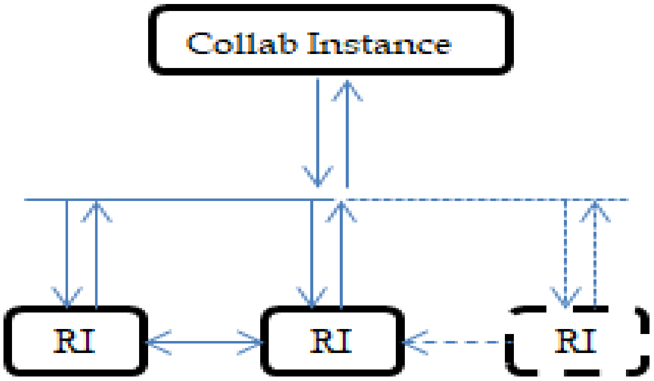


Fig-3: A network of computing nodes under the collaboration

Access to the collaboration data is however made possible through an encryption mode specified by the service broker registry, through which collaborators can as well monitor and audit actions on the overall Collaboration data. All institutional data collected is maintained majorly at the RI server and synced at the Collaboration instance storage allocated to the collaboration, specified by the service broker.

Adding an extra researcher to the collaboration implies adding a new member to the blockchain as illustrated in fig-3. This will require account creation for the requester. This proceeds with double authorization (system based and lead collaborator based authorization). And based on the authorization, extra configurations for a co-researcher within the collaboration implemented as role-based service access distribution, updating the open registry as well. The Collaboration SaaS introduces new and various benefits and implements researcher recommendation for assured availability and transparency due to zero alteration with data flow and its models among others.

5.3. APPLICATION USE CASES

Based on the designed architecture, two examples need to show feasibility, the relevance of the architecture: a) cancer genomics - esophageal squamous cell carcinoma; b) Real-time TB monitoring. These are public health and biomedical collaboration studies involving researchers from diverse continents collaborating to achieve set study objectives and arrangements. We illustrate the potential for the privacy and trust-aware cloud – the Collaboration cloud with case studies of a) and b) mentioned herein and to follow.

a) Cancer genomics - Esophageal Squamous Cell Carcinoma (ESCC) study

The cancer genomics study is an NIH funded study[25] aimed at investigating how biomass fuels like polycyclic aromatic hydrocarbons; and diet affect the ESCC risk with a special focus on the how DNA methylation can act as a mediator. The study contributes extensively to their recent scientific study on how being a male, taking alcohol with exposure to smoking can increase the risks of acquiring ESCC. A quick exploration of the study informs vast contributions to cancer genomics, gastroenterology, and nutrition expertise. The reporter presents that achieving the objective will require the researchers to explore genetic, molecular signatures of ESCC. More still, the study involves measurement and analysis of nutritional data, cancer, and genetic epidemiology, with outcomes enriching research, academia and practice. In this study, data collection made via tools designed in CSPRO using tablets that are also further utilized on a third party – proprietary tools. It is further reported that storage is also conducted on separate offline tools every after two weeks to update the data that is hoped to be shared upon completion of usage policies and agreements with the institutions under the collaboration. Findings from the study are vital to the public health of sub-Saharan Africa and other settings where ESCC is believably rampant. Upon completion, in other words, the study will innovatively contribute to alternative approaches to targeting multifactorial risk modification necessary currently to prevent and improve quality of life by putting an end to disease progression.

Cloud computing can transform the face of health and biomedical research collaboration, enriching both research and practice; though focusing on availing flexibility during and in data collection, in the integration of data storages and during sharing while at the same time providing for trust, integrity, confidentiality, access control, auditability and the collaboration environment’s overall reliability; and the core functionalities more reliably and equitably.

b) Real-time TB monitoring

Poor TB medication adherence has greatly retarded the heavily invested-in efforts of donors and enervated the initiatives to alleviating communities off this disease burden. With the multi-country efforts observed at work in ensuring TB is successfully treated and transmission is further prevented, while non-adherence to medication continues to become a grounded TB treatment burden due to its developed resistant strand. The study[26] presents how the non-resistant TB is quite expensive to treat and difficult for patients to comply with due to the longer duration and related side effects. Real-time TB adherence monitoring also a National Institute of Health (NIH) funded study, is an innovative approach that has worked elsewhere but not explored resource-limited settings. The study is grounded in the use of adherence monitoring using a device, SMS reminders, and social support to improve TB medication adherence. The project utilizes a Wise pill device with sensors to monitor patients' adherence to TB medication, with SMSs of reminders forwarded to patients through the CommCare platform and ‘Yo' SMS service. Device information is collected and stored on CommCare provided the environment. The stored data is then made accessible to the team for analysis when due.

The collaboration cloud presents a flexible model that observes trust while providing data collection, storage and facilitates sharing more reliably, of course with assurance for privacy and integrity. The two use cases: Real-time TB monitoring and the Cancer genomics project studies can take advantage of the platform and utilize platform core functionalities. Also part of the platform is that now, collaborators can openly share data as a resource and choose when to avail the data for public use, that is to say, secondary research thus contributing to research, enabling innovation and advancement of public health to treat some of the world's disease burdens through improvements in algorithms for analysis, treatment, drugs development, and knowledge bases.

5.4. IMPLICATIONS FOR DATA GOVERNANCE

Informed by requirements elicitation, data governance is amongst the key concerns of the Collaborations. The proposed Cloud Architecture will contribute with affordances for control of ownership; enforce privacy, security, compliance, usability, availability and the overall management of internal or external data flows.

a) Control of ownership – the architecture ensures that ownership is preserved ownership and increased control through right based access control. At the inception of collaboration, configurations are set for the collaboration coordinator, allowing the rest to actively serve as members of the collaboration. With the retention of a collaboration dataset access, support for researcher recommendations for an environment that enables for consultations with and for all stakeholders, supporting mutual respect of values, strategies, and actions for the reliable partnership of people affiliated by a geographic location, shared interest, or similar circumstances.

b) Privacy, security, and compliance – the blockchain format integration within the architecture enables for increased privacy, confidentiality, integrity through the use of a common registry and provisioning for what happens when a new collaborator is added, thus increasing trust among collaborators. The architecture as well ensures privacy and security throughout the entire collaboration.

c) Availability and usability – the seeding concept in fig-3 enable collaborators to retain a copy of their dataset when they contribute to the collaboration dataset on their server. The data flow indicates that the sync function will run until the collaboration has aged. The data is transformed for external sharing using a hashing algorithm and archived to be availed for public access. Just like any endeavor, the collaboration established is limited by time. The SaaS enables collaborators to estimate the collaboration period through the use of SLAs, seeking to articulate when data could be availed for public use; enabling for and re-use of samples and data for secondary research. Prior to availing access to data for re-use, the data is hashed to observe confidentiality, privacy and compliance agreements like HIPPA and country-specific data protection policies and laws.

And in case of malicious loss and damage, the collaboration is still assured access to their data due to the sync function on multiple storages: Collaboration storage and that of the seeding participants.

d) Management of internal and external data flows –At the ‘seekers’ level, an independent registry at the research institution (RI) server is implemented supported by the application's based role based service access. The members at the research institution headed by a principal research/investigator (PR/I) have specific roles to the collaboration cloud data. For example, access to the overall data is allocated to the principal research while the rest including the data manager and research assistants are availed opportunities for managing their local data copies that sync to the open collaboration data stores provided by the collaboration server. Multi-level security internally is ensured through observing that each collaborator accesses their own data-store (minute copies of collaboration data). Access to the collaboration data is however made possible through an encryption mode specified by the service broker registry, through which collaborators can as well monitor and audit actions on the overall Collaboration data. Upon agreed policy and well standardized SLAs, data age is determined, and data is prepared with hashing for confidentiality, hiding some sensitive details of the data sets prior to having them shared.

6. CONCLUSION

In this study, we have attempted inquiry on how we derive a cloud computing architecture that supports scalability, flexibility, and reliability with data sharing in collaborations. We have observed numerous data sharing concerns including security, integrity, transparency and the overall data quality, and capabilities for internal and external data sharing among researchers; ensuring privacy and enforcing trust relationships for data stationed or in transit. The conventional tools used by collaborators in biomedical and public health are limited and continue to face numerous challenges like the view and nature of health-related data is characterized with demanding requirements for confidentiality, application and system-wide transparency, availability, reliability and especially the increasing concerns of data lock-in. The absence of innovative approaches to cloud-based research collaborations in these domains will accelerate the collaboration concerns and impede the computing paradigm’s wide adoption.

We also observed that cloud computing can help leverage the provision for electronic data capture, the increasing need for secure and unified storage, and reliability in data sharing in collaborations can be achieved more reliably and equitably. Innovative implementation of the computing paradigm will eliminate provisioning challenges of unavailability and lack of Service Level Agreement (SLA) support, data lock-in, and inability to interoperate amongst provider clouds prevent deployment, plus the inability to extend multi-tenancy support across multiple collaboration clients, resulting into unreliable, unreliable and inefficient.

In this paper, we propose an efficient collaboration cloud that utilizes the PbD foundational principles to assure optimum privacy; the multi-level tenancy coupled with measures for privacy and trust to enable secure and scalable service provisioning; Service orientation to expose and consume collaboration based cloud resources with multi-level and role-based access control; and seeding to ensure the overall platform reliability while providing the expected services to researchers. Our approach utilizes some of the blockchain conceptual approaches to leverage trust and privacy including confidentiality of Collaboration data. We argue that the possibility of unifying storage (access, sharing, and reference) through establishing a singleton cloud storage for every instance of the collaboration where every participating institution in the collaboration is offered logic units for collecting, storing and making accessible the collected data available to all parties under the collaboration.

The collaboration cloud is able to expose and avail data collection facilities, integration towards unified storage and sharing as services and this can be harnessed appropriately well by the use cases in regards to a) providing flexibility during data collection and integration from heterogeneous data sources (enabling for eliciting from broader data sources like remote sensors, wearable devices, patient mobile devices including also phones); b) equitable and dependable sharing of resources among collaborators and the external world with assurance for increased integrity and reliability. For instance applying the blockchain approach (of the open ledger) on the research collaboration within the cloud architecture, collaborators are assured of integrity while data is hosted and on the go (being shared) through implementing of an independent registry at the research institution (RI) server that syncs to the open collaboration registry provided by the collaboration server instance (also called the service broker). Using our approach, each collaborator accesses their own data-store (minute copies of collaboration specific data) and interfaces for the collaboration cloud services, implementing provider level privacy. Access to the collaboration data is however made possible through an encryption mode specified by the service broker registry, through which collaborators can as well monitor and audit actions on the overall Collaboration data. All institutional data collected is maintained majorly at the RI server and synced at the Collaboration instance storage allocated to the collaboration. The proposed architecture features SLA and service-oriented provisioning of collaboration clouds, role-based encryption and ensures enhanced privacy and trust through a service registry with seeding approach to synchronize collaborator data with the Collaboration instance data. Future work under this study shall include the actual design and development of the SaaS prototype to implement the designed architecture. Evaluation of the SaaS prototype will undergo a number of experiments, testing for whether the actual architectural design meets the collaboration requests, usability and several enhancements for the collaboration cloud functions so as to address collaboration scalability, flexibility and reliability with data sharing and provides further support for data analysis functions as required for Health Research Collaboration based on Cloud Computing. The proposed architecture control of ownership; enforce privacy, security, compliance, usability, availability and the overall management of internal or external data flow.

ACKNOWLEDGMENTS

We acknowledge the African Development Bank – Higher Education Science and Technology (ADB-HEST) programme for the financial support rendered towards this study. We are grateful to reviewers for the constructive feedback, it helped transform our first write-up of the article. You shall always be of value to us.

REFERENCES

[1] T. Bradley. (2011, 2018-3-21). The Cloud, Day 10: Storing Data in the Cloud. PCWorld.

[2] R. Mugonza and A. H. Basaza-Ejiri, "Survey: Cloud Computing architectures for Health and Biomedical Research Collaboration," In press|.

[3] R. Mugonza and A. H. Basaza-Ejiri, "Issues Affecting Health Research Collaborations based on Cloud Computing," International Journal of New Technology and Research, vol. 4.

[4] C. Baru, N. Botts, T. Horan, K. Patrick, and S. S. Feldman, "A seeded cloud approach to health cyberinfrastructure: Preliminary architecture design and case applications," in System Science (HICSS), 2012 45th Hawaii International Conference on, 2012, pp. 2727-2734.

[5] P. Mell and T. Grance, "The NIST definition of cloud computing," 2011.

[6] A. Arsanjani, "Service-oriented modeling and architecture," IBM developer Works, vol. 1, p. 15, 2004.

[7] R. Buyya, S. K. Garg, and R. N. Calheiros, "SLA-oriented resource provisioning for cloud computing: Challenges, architecture, and solutions," in Cloud and Service Computing (CSC), 2011 International Conference on, 2011, pp. 1-10.

[8] T. Erl, SOA: principles of service design vol. 1: Prentice Hall Upper Saddle River, 2008.

[9] L.-J. Zhang and Q. Zhou, "CCOA: Cloud computing open architecture," in Web Services, 2009. ICWS 2009. IEEE International Conference on, 2009, pp. 607-616.

[10] M. H. Valipour, B. AmirZafari, K. N. Maleki, and N. Daneshpour, "A brief survey of software architecture concepts and service-oriented architecture," in Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on, 2009, pp. 34-38.

[11] W.-T. Tsai, X. Sun, and J. Balasooriya, "Service-oriented cloud computing architecture," in Information Technology: New Generations (ITNG), 2010 Seventh International Conference on, 2010, pp. 684-689.

[12] E. Newcomer and G. Lomow, Understanding SOA with Web services: Addison-Wesley, 2005.

[13] H. Demirkan and D. Delen, "Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud," Decision Support Systems, vol. 55, pp. 412-421, 2013.

[14] N. Nikzad, N. Verma, C. Ziftci, E. Bales, N. Quick, P. Zappi, K. Patrick, S. Dasgupta, I. Krueger, and T. Š. Rosing, "Citisense: Improving geospatial environmental assessment of air quality using a wireless personal exposure monitoring system," in Proceedings of the conference on Wireless Health, 2012, p. 11.

[15] W.-T. Tsai, Q. Huang, J. Elston, and Y. Chen, "Service-oriented user interface modeling and composition," in e-Business Engineering, 2008. ICEBE'08. IEEE International Conference on, 2008, pp. 21-28.

[16] B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth, and J. Caceres, "The reservoir model and architecture for open federated cloud computing," IBM Journal of Research and Development, vol. 53, pp. 4: 1-4: 11, 2009.

[17] E. Badidi, "A broker-based framework for integrated SLA-aware saas provisioning," arXiv preprint arXiv:1605.02432, 2016.

[18] N. Sonwalkar, "The first adaptive MOOC: A case study on pedagogy framework and scalable cloud Architecture—Part I," in MOOCs Forum, 2013, pp. 22-29.

[19] A. Cavoukian, "Privacy by design," Take the challenge. Information and privacy commissioner of Ontario, Canada, 2009.

[20] L. Chen and D. B. Hoang, "Novel data protection model in healthcare cloud," in High-Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on, 2011, pp. 550-555.

[21] M. Li, S. Yu, K. Ren, and W. Lou, "Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings," in International conference on security and privacy in communication systems, 2010, pp. 89-106.

[22] A. H. B.-E. Robert Mugonza, "Title," unpublished‖.

[23] H. O. D. Kasule, "The 9th Annual National Research Ethics Conference," Uganda National Council for Science and Technology2017.

[24] G. Zyskind and O. Nathan, "Decentralizing privacy: Using blockchain to protect personal data," in Security and Privacy Workshops (SPW), 2015 IEEE, 2015, pp. 180-184.

[25] N. T. D. I. Health®. (2018, 23 June 2018). Polycyclic Aromatic Hydrocarbons Exposure And Dietary Risk Of Esophageal Squamous Cell Carcinoma In Uganda. Available: https://projectreporter.nih.gov/project_info_description.cfm?aid=9392596

[26] N. T. D. I. Health®. (2018, 23rd June 2018). Real-Time Tuberculosis Medication Adherence Intervention In Rural Southwestern Uganda. Available: https://projectreporter.nih.gov/project_info_description.cfm?aid=9321387&icde=40018563&ddparam=&ddvalue=&ddsub=&cr=1&csb=default&cs=ASC&pball=

AUTHORS

Robert Mugonza is a lecturer in the Department of Computer Science, Faculty of Computing & Informatics, Mbarara University of Science and Technology. He holds an MSc in Advanced Computer Science from the University of Leeds and a Bachelor of Computer Science from Mbarara University of Science and Technology. He is a member of the IEEE Computer Society and IEEE Cloud Computing. His research areas of interest are around the Internet of Things, Cloud Computing, Computational Modelling, Design and Analysis of Systems, and Software Engineering.

Dr. Annabella Habinka Basaza-Ejiri is an Associate Professor in the College of Computing and Engineering, St. Augustine International University. She holds a Ph.D. in Information Systems from Groningen University in the Netherlands, MPhil. S& T from Stellenbosch University in South Africa, a Master's of Science degree in Information System from Makerere University and a Bachelor of Computer Science from Mbarara University of Science and Technology. She is a member of the ACM, IEEE, and Association for Information Systems computer Society with areas in Software Engineering; Cloud computing, Information Systems, and Decision Support.