

Big Data as an Healthcare e-Health Service

Abinaya.K¹, PG Student¹, C. Chitra², M.E²

Department Of Computer Science and Engineering,

¹PG Student, Parisutham Institute of Technology and Science, Thanjavur, Tamilnadu, India.

²Asst.Professor, Parisutham Institute of Technology and Science, Thanjavur, Tamilnadu, India.

Abstract: Big Data is growing with large-volume, complex, growing data sets with several autonomous sources that transforming healthcare, business in all science and engineering, as healthcare becomes one of key driving factors during the innovation process. Here we introducing BDHeHS (Big Data Healthcare e-Health Service) to fulfill the Big Data applications in the e-Health service domain. BDHeHS will explain about why the existing Big Data technologies like Hadoop, MapReduce cannot be simply applied to e-Health services directly. Then we describe the additional capabilities as required in order to make Big Data services for e-Health become more practical healthcare service. Finally we report our design of the BDeHS architecture that supplies data flow management, exploratory management and resource management, and e-Health meaningful usages.

Keywords: Big Data Technologies, Big Data as a Service, Service Support Infrastructure, e-Health Data Operation Management, Exploratory and resource management..

I. INTRODUCTION

Bigdata and Digital healthcare solutions have promised to transform the whole healthcare process to become more efficient, less expensive and higher quality[1~8]. In the context of e-Health, numerous flows have generated slightly less than 1,000 peta bytes of data now (and may reach about 12 ZBs by 2020 in our own estimates) from various sources such as electronic medical records (EMR) systems, mobilized health records (MHR), personal health records (PHR), mobile health care monitors, genetic sequencing and predictive analytics as well as a large array of biomedical sensors and smart devices.

The electronic medical record (EMR) initiative has resulted data streams from all types of patients at the hospital, doctor's office, insurance office, institutions, government sectors, etc. A single patient stay generates thousands of data elements, including diagnoses, procedures, medical supplies, digital image, lab results billing, disease status, etc. All of these need to be verified, processed and combined into a large data sets to enable meaningful analysis. We are in situation to multiplying this by all the patient-stays across the health processing systems and combining it with the large number of points where data is retrieved and stored and the scope of the big data challenge begins to explore.

Other care provider facilities, such as external health data process including social media, smartphones, wearable sensor information on patient heart rate, brain activity, temperature, calories and lot of other clinical usage datata points, high-throughput and system-wide measurement of the systems of many biological body parts and their status, as well as other health-related information being carried among associates parties of insurance, government reporting, etc.

Additional data are in the e-Health communication and support infrastructure sources including the National Health Information Network (NHIN), Health Information Exchanges (HIE), Health Information Organizations (HIO) and Regional Health Information Organizations (RHIO).

As big data sources and volume of information and data sets increase, expectations also increase in utilizing those large volume of healthcare data to reduce costs, improve efficiency and boost outcomes and improve treatment. Finally we validate this BDHeHS to fulfill the Big Data applications in the healthcare service domain.

This paper is organized as follows. In section II, we describe the foundations and the additional BDHeHS functionality. To further extend the Big Data service approach into a national and global framework for e-Health, in Section III we describe our design of the BDHeHS architecture that supplies data flow management capabilities, exploratory management, and resource management. The final section concludes with a summary of our contributions.

II. BIG DATA FOUNDATIONS FOR E HEALTH

Figure 1 below depicts an environment that supports e-Health applications from individual outlets and test facilities as well as insurance providers and government agencies [9~12]. All generate tremendous data points interconnected with the national health information networks.

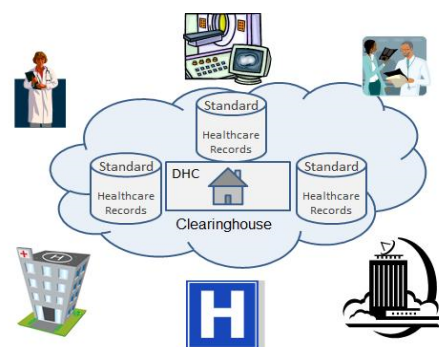


Figure 1. e-Health Big Data Service Environments

A. Big Data Solutions and Products

Big Data research requires knowledge about standards, filters, meta-data, techniques for storing, finding, analyzing, visualizing and securing data, and sector-specific editing of data. The predominant current technologies include MapReduce, Hadoop, STORM and alike with combinations or extensions.

1)STORM

Storm is a free and open source distributed realtime computation system. Storm makes it easy to reliably process unbounded streams of data. Storm can be used with any programming language (usually Java, but also Python and others as well). Storm integrates with the message queuing and database technologies. A Storm topology consumes streams of data and processes those streams in arbitrarily complex ways, repartitioning the streams between each stage of the computation however needed. Storm may achieve processing of over a million tuples processed per second per node as necessary for e-Health global flows.

To do realtime computation on Storm, a “topology” is created, which is a graph of computation. Each node in a topology contains processing logic, and links between nodes indicate how data should be passed around between nodes.

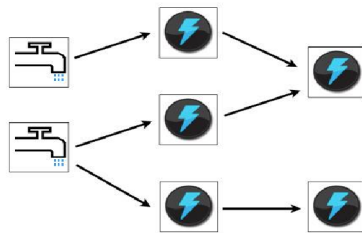


Figure 2. Illustration of STORM Architecture

Storm can integrate with any source of message queuing and database connections. It can also generate its own stream or read from somewhere like e-Health streaming data. The source of data stream is called a “spout” in the Storm architecture (see Figure 2 above).

C. Additional e-Health (Big Data) Capabilities

The BDeHS provides the services to access, organize, and glean discoveries from huge volumes of e-Health digital data. In order to augment the Big Data foundations to the ecosystems of e-Health, we identify additional key capabilities necessary for BDeHS services.

1) Data Federation and Aggregation

The various types have been described in the previous section, and those sources reflect the fragmentation of e-Health data among the various stakeholders, including payers, providers, labs, ancillary vendors, data vendors, standards organizations, insurance institutions and regulatory agencies. Solutions for big data will break the traditional model, in which all data is loaded into a warehouse. Data federation will emerge as a solution in which the big data architecture is based on a collection of nodes within and outside the enterprise and accessed through a layer that integrates the data and analytics.

2) Security and Regulatory Concerns

This is the most fundamental requirements that distinguish the Big Data services for e-Health. They deal with additional challenges, such as privacy, security and legal concerns, as well as questions about authenticity, accuracy and consistency.

The entire healthcare system can realize benefits from democratizing big data access and the cloud [2] makes exposing and sharing big data easy and relatively inexpensive. However, significant security and privacy concerns exist, including the Health Insurance Portability and Accountability Act (HIPAA). A credentialing process could

facilitate and automate this access, but there are complexities and challenges. Since providers, patients and other interested parties such as researchers need various secure accesses, data security policies have to control by group, role and function. Finally, the security of the data once it leaves the cloud also needs to be assured.

3) Data Operational Management

The operational management capabilities [8][9] include data interoperability management at a global scale, information timeliness to meet service level agreement, and operational support architecture to constantly automatic and improve the quality of services.

Big data solution architectures have to be flexible enough to cope with not only the additional sources but also the evolution of schemas and structures used for transporting and storing data. To ensure analytics are meaningful, accurate and suitable, metadata and semantic layers are needed that accurately define the data and provide business context and guidance, including appropriate and inappropriate uses of the data. This evolution of standards will eventually improve interoperability and data quality.

Data timeliness is a challenge in various healthcare settings, such as clinical decision support, whether for making decisions or providing information that guides decisions. Big data can make decision support simpler, faster and ultimately more accurate because decisions are based on higher volumes of data that are more current and relevant. As the data points and decision points are going beyond the humanly availability during a very limited window for clinical decision support, response time has to be capped to run a report or analytic query. Careful attention to data and query structure, scope and execution is needed to ensure that the constraints of the processing windows are observed while still obtaining the best possible answer.

In the Big Data ecosystems, streams of e-Health data containing complex and varied events without an overarching structure need to be addressed. In this case, those events have to be turned into meaningful measures in real time that are, in turn, suitable for rapid analysis. Security of data is inherently built-in when e-Health regulations were designed for promoting e-Health roll outs. As the operational infrastructure becomes extremely complex, operational management has become an integral part of the BDeHS when we design the solutions.

III. SOLUTIONS OF BIG DATA SERVICE FOR E-HEALTH

Our Big Data for e-Health Service solutions are evolved from our research of interoperable (data) flows as defined in [3,8,9] into streams as illustrated in the Figure 4 below.

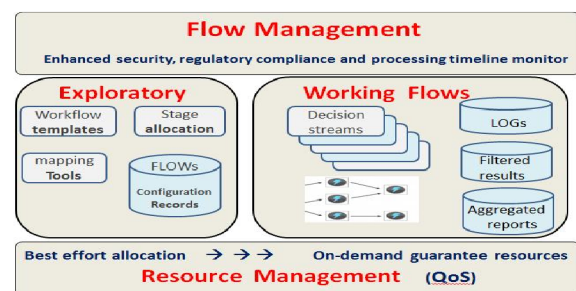


Figure 3. BDeHS Solution Diagram

A. Big Data Flows in e-Health Streams

Our new solution provides data streaming federation and decision points in supplement to the original flows of e-Health adaptation and message routing [3,8,9].

The center of the concept is around application flow set-up procedures that detail the processing stages including data fork points, stream joints, as well as event and message logging. All of which are specified in a policy format that controls the in-flight processing of data. A protocol command (e.g., transform/send/store data) may be triggered by a protocol type and the policy ID established during the flow setup processing.

A data flow in this e-Health environment is mapped into a stream with additional processing stages. First data formats are adapted via adaptation gateways into a common format as required by the e-Health metadata model, while security policies are tagged along with the flows.

Data Federation is extended into our own e-Health adaptation gateways [3][10] which admit data flows into the processing nodes with e-Health data processing logics (filtering, logging, aggregation, exploratory and iterative analysis, regulatory checkpoints, and so on). Inside the BDeHS infrastructure, processing nodes can further feed into another adaptation gateway for further flow analysis so that information may be further correlated with other flows.

Data sinking gateways provide the exit points of the e-Health data flows when anomalies (e.g., adverse treatment or drug effects) are detected on flight or when e-Health data security events have to be reported. Data logs from relevant flows can be directed into certain exit gateways so that aggregation can formulate dynamic solutions that require immediate decisions (e.g., responses to the spread patterns of an epidemic). In addition, aggregated data reports can be generated at any stages of the e-Health data streams.

In between the data entries and exits, a number of middle storage and processing stages supply data replications and parallel processing logics for additional data segmentation, summarization, (security and health regulation) policy enforcement, filtering, data transformation, header and trailer expansion, message split or union, state synchronization, and coordination with other distributed processing nodes.

One key benefits of our solution approach is in e-Health Data Consistency. Data adaptation and mappings promote consistency in self-reported data across the healthcare system to eliminate local discrepancies and increase the usefulness of data. As e-Health becomes global solutions, we plan to come up with new ways to further the adaptation layer to facilitate global scale usages. Aggregating data regionally and globally also provides healthcare researchers with larger populations for clinical studies, trending and disease monitoring for epidemics, as well as early detection and the potential for improved results.

Another benefit is in its flexibility to deal with Regionalization and/or Globalization of flows. External data will come from different medical systems in various regions and countries. Effectively working across these disparate data repositories can help identify local knowledge and best practices and leverage them regionally and globally.

The BDeHS e-Health-flows deal with increasing mobile Health application traffic as well. Demand for ubiquitous access to information mandates mobility and other

technologies that provide access on demand. As data becomes more current, it can be forked out of the flows into the mobility adaptation gateway into the hands of people with an immediate need for it, such as for clinical decision support. Quality of care and improved outcomes will be the ultimate benefits.

B. Security and Regulatory Compliance

Secure BDeHS service payloads and information flows have to be part of an e-Health solution. Possible messages with security concerns include summary of patient record exchange, terminology mediation, message handling (includes transformation, routing, logging and content based filtering), secure data delivery, and confirmation of meaningful usages.

Security framework in e-Health was already established in [8]. Once a secure e-Healthcare association is established, both end points may invite others to participate in a MPMD (Multiple Participations and Multiple Drop-offs) e-Healthcare flow. The e-Healthcare Service Associates Identifier(s) are correlated together so that each entity has the visibility to the relevant portion(s) of the communication payload, while linking the total messages intended for different recipients. When the data streams are routed into any Big Data analytic engines, patient-indefinable information are removed according to BDeHS security policies following the intent and security level of the stream sink points.

The major benefit of ID-removal ensures access to some de-identified data can simultaneously improve levels of self-reporting as well as data sharing. For example, Patient ID credentials (instead of the detail ID#, Name, Address information) are no longer presented from clinical offices to the lab facilities. Sending only the essential (minimum credential) data that are pre-agreed upon among the e-Healthcare processing parties also enhances performance of rapid forking of the streams into Big Data processing clusters. When we allow security associations with multiple entities, an issue has to be addressed as different parties may want to view a different subset of the records on-the-fly in order to collaborate in the care process. For example, some portions of the patient records may not be relevant to another e-Healthcare party (e.g., such as the lab with a processing entity that only needs a patient identity without the details of patient records).

When the BDeHS processing nodes recognized the communication message contains HIPPA conformance required fields, additional logging actions are invoked before a node can process or send the payload forward. As such, regulatory compliance specific actions are carried out during the flow. An example of e-Health network policy action is to do packet inspection inside a data stream to check the principal parties and the so called business associates (as defined in HIPPA and HIGHTEC Acts), and then to be conformant with HIPAA security rule of ensuring the appropriate log entries will have a whole message instead of a partial MDMP stream.

C. Data Operational Management and Data Timeliness

Because not all e-Health Big Data flows are batch process, we need to develop realtime flows that meets QoS (such as time constraints and throughput requirements) in order to provide on-flight e-Health event detection and aggregation

analysis.

The current BDeHS capabilities include policy actions further combined with QoS actions as future traffic patterns are identified with routing of e-Health contents being further improved. In other words, any routing policies may further reference application-oriented policies that have to be satisfied before the routing can be applied.

1) Centralized QoS Monitoring

Monitoring the BDeHS performance is the essential step in provisioning additional storage clusters and networking resources to guarantee QoS of e-Health Big Data applications. QoS Service Managers are devised to control and coordination of the end-to-end overall Big Data stream service views. The QoS manager handles end-point performance parameter requests. During a BDeHS flow, the QoS manager also monitors and assists in enforcement of end-to-end transmission performance, via accessing to logs and reports of functional performance feedback.

2) BDeHS Service Profile Management

Service profiles describe how to provision services for a specific domain or functionality such as e-Health participation levels, priorities, acceptable usage parameters, end-to-end test setup, and so on. The services describe the specific interfaces to be used among interconnection participants to locate and exchange health information. All are governed by QoS parameters which in turn support service level guarantees.

3) Data Interoperability Management

Our solution architecture is flexible enough to cope with not only the additional sources but also the evolution of schemas and structures used for transporting and storing data. To ensure analytics are meaningful, accurate and suitable, metadata and semantic layers are supported that accurately define the data and provide business context and guidance, including appropriate and inappropriate uses of the data. This evolution of standards will eventually improve data quality.

IV. PERFORMANCE EVALUTION

We Self-directed data sources with distributed and decentralized controls are main characteristic of Big Data applications. Each data source is able to generate and collect information without involving (or relying on) any centralized control. On this way the application of e-health care for heart disease data are uploaded to the cloud successfully and obtained the correct values of result for retrieval in milliseconds from various source of clusters in bigdata. While using mining algorithm it is tedious to achieve results in milliseconds.

Map reduce is a software framework that support the parallel processing, since it is a parallel process it splits the input data set into independent chunks which are processed by the map task in completely parallel manner.

This framework sorts the output of the maps, which are then input to the reduce task. Typically both input and output of job are stored in a file system.

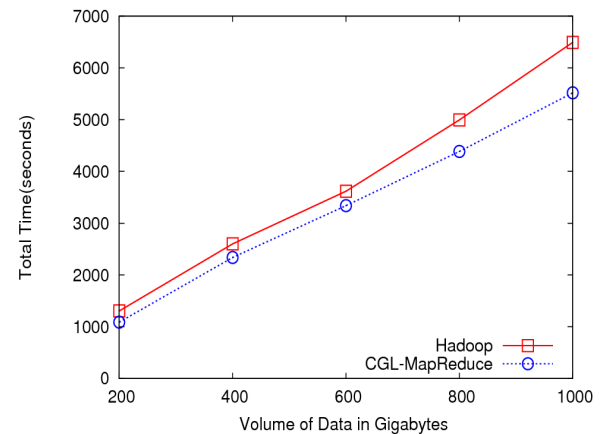


Fig. 4 Performance analysis of hadoop and map reduce

Figure 3 describes the performance analysis of hadoop and map reduce in cloud. The rectangle box describes the hadoop data where the circle belongs to map reduce.

V. CONCLUSION

We have presented a new BDeHS (Big Data for Health Service) approach to bring healthcare flows and data/message management intelligence into the ecosystems for better stream processing, operational management and regulatory compliance. By asserting security control into the service layer with native e-Health features our solutions ensure regulatory compliance.

Additional key benefits of our BDeHS operational features include: Big Data Stream setup and provision to support application flow associations, data federation and aggregation along the streaming paths, partitioning of e-Health messages along the processing. Security to meet stringent e-Health environments and regulations is an integral part of the infrastructure. Multiple flows can be managed by a common MPMD model. End-to-end regulatory oversights can be guaranteed.

Quality of Service guarantees includes realtime processing, reconfiguration and enhancement of processing cluster and network capacity, data interoperability management, as well as reporting capabilities.

The BDeHS service solution as reported in this paper has supplied the much needed detailed guidelines for overall Big Data services to achieve the meaningful usages of e-Health global solutions.

ACKNOWLEDGMENT

Sincere thanks to my guide Prof. C. Chitra Asst. Prof., Computer Science and Engineering Department, Parisutham Institute of Technology and Science, Thanjavur for her help and guidance enable me to propose this system.

REFERENCES

1. W. Liu and E.K. Park, "e-Health AON (Application Oriented Network)", IEEE International Conference on Computer Communication Networks, BMAN Workshop, Nausa, Bahamas, July 2013.
2. E.K. Park and W. Liu, "e-Healthcare Cloud Computing Application Solutions", IEEE ICNC2013, International Conference on Computing, Networking and Communications, San Diego, CA, January 2013.
3. W. Liu, E.K. Park and Udo R. Krieger, "e-Health Interconnection Infrastructure Challenges and Solutions Overview", IEEE HealthCom-2012, the 14th IEEE International Conference on e Health Networking, Application & Services, Beijing, China, October 2012.
4. C.G. Chute, "Obstacles and options for big-data applications in biomedicine: The role of standards and normalizations", 2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
5. D. Li, C. Tao, H. Liu and C. Chute, "Ontology-Based Temporal Relation Modeling with MapReduce Latent Dirichlet Allocations for Big EHR Data", Second International Conference on Cloud and Green Computing (CGC), 2012.
6. X. Lu, H. Tang, W. Cheng and T. Zhang, "Heterogeneous Data Source Middleware for Android E-Health Application", Eighth International Conference on Mobile Ad-hoc and Sensor Networks (MSN), 2012.
7. M. Diaz, G. Juan, O. Lucas and A. Ryuga, "Big Data on the Internet of Things: An Example for the E-health", "Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2012.
8. W. Liu and E.K. Park, "e-Healthcare Security Solution Framework", IEEE International Conference on Computer Communication Networks, MobiPST-2012, Munich, Germany, August 2012.
9. W. Liu and E.K. Park, "e-Health Service Characteristics and QoS Guarantee", IEEE International Conference on Computer Communication Networks, Workshop on Context-aware QoS Provisioning and Management for Emerging Networks, Applications and Services, Maui, HI, August 2011.
10. J. Yang, D. Tang and X. Zheng, "Research on the distributed electronic medical records storage model", International Symposium on IT in Medicine and Education (ITME), 2011.

AUTHOR DETAILS



K.Abinaya received B.E (CSE) from Periyar Maniammai University in 2013. She is currently pursuing M.E –Computer Science and Engineering in Parisutham Institute of Technology and Science.



C.Chitra received M.C.A.,M.E., working as an Asst.Professor in Dept. of Computer Science an Engineering, Parisutham Institute of Technology and Science, Thanjavur.