© 2015,
Scienceline Publication
www.science-line.com

ISSN: 2322-5114
Journal of World's Electrical Engineering and Technology
J. World. Elect. Eng. Tech. 4(2): 52-57, Dec 25, 2015

Study of Cloud Computing Techniques in Solving Healthcare's Big Data Problems

Mehdi Abedini and Ali Taeezadeh

Department of Computer Engineering, Qom Azad University, Qom, Iran

*Corresponding author's Email: mmehdiabedini@yahoo.com

ORIGINAL ARTICLE
PII: S232251141500009-4
Received 08 Jan. 2015
Revised 22 Jun. 2015
Accepted 20 Oct. 2015

Abstract *– Healthcare's data may improve the quality of care in medical area and the computerized software has as a great result reducing medication errors related to the patients treatment. Cloud Computing is a nascent technology where computing resources are made available in a pervasive and a convenient way as online services. Cloud computing provides potential opportunities for improving EHR adoption, healthcare services and research. In this paper, we describe a detailed overview of some of important healthcare cloud computing strategic planning model and discusses the limitations and strengths of each cloud discussed.*

Keywords*: Big Data , Cloud Computing , Healthcare, Framework*

INTRODUCTION

Big data shall mean the datasets that could not be perceived, acquired, managed, and processed by traditional IT and software/hardware tools within a tolerable time. Because of different concerns, scientific and technological enterprises, research scholars, data analysts, and technical practitioners have different definitions of big data. The following definitions may help us have a better understanding on the profound social, healthcare, economic, and technological connotations of big data [1].

Big Data is characterised by what is often referred to as a multi-V model, as depicted in Figure 1. Variety represents the data types, velocity refers to the rate at which the data is produced and processed, and volume denes the amount of data. Veracity refers to how much the data can be trusted given the reliability of its source, whereas value corresponds the monetary worth.

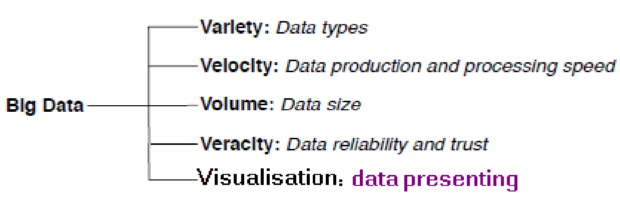

Fig 1. Some `Vs' of Big Data.

That a company can derive from employing Big Data computing. Although the choice of Vs used to explain Big Data is often arbitrary and varies across reports and articles on the Web|e.g. as of writing Viability is becoming a new V|variety, velocity, and volume are the items most commonly mentioned [2].

The rapid development of science and healthcare technology has yielded rapid and effective methods for verifying, detecting, preventing, and treating diseases. This phenomenon has generated big healthcare data, specifically:

1) A rapid accumulation in the number of medical records;

2) increased number of medical evaluation factors (e.g., items investigated during laboratory, biochemical, and genetic tests)

3) Diverse data types (e.g., texts and numerals, diagrams, tables, images, and handwritten documents);

4) Difficulties processing and managing data. Combined, these aspects of big data delay response times and increase costs [3].

Cloud computing technology perfectly matches such "big data" challenges by providing nearly unlimited storage resources on demand [4]. In healthcare, it is also gaining particular popularity by facilitating an inter organizational medical data sharing environment [5].

On the other hand, this paradigm also in volves many security and privacy risks that lead to concerns among patients and medical workers [6] who are being particularly afraid of losing the control over sensitive medical records while storing them on not fully trusted third party servers . Regulations such as HIPA Aalso call for a strong protection of medical records [7].

Recent advancements in cloud computing technology show promises of lower cost, higher scalability, accessibility, availability and disaster recoverability. Computing and storage in the cloud seem to be a natural solution to many problems we face today for long-term medical image archives. While researchers and policy makers are still actively studying the security, privacy, and liability issues involving sensitive medical information in the cloud, various technology vendors such as IBM and Amazon have started to provide solutions for early adopters.

In this paper, a study of cloud arciteture and comparison of Some `Vs' of Big Data solution in some important cloud in healthcare. Section 2 describe cloud computing, section 3 describe some cloud arcitecture in healthcare, section 4 comparison healthcare's clouds. Finally, Section 5 concluded this article.

Cloud computing

The cloud is a large group of interconnected computers extending beyond the enterprise. It is a paradigm shift from traditional desktop computing model that enables running of software programs from each computer. With the emergence of cloud computing platform the PC centric environment is transformed into document centric culture. The software's are not run from a personal computer rather stored on the server and is accessed through internet. The health care industry undergoes tremendous pressure to deliver quality service to patients and doctors across the globe [8].

According to the US National Institute of Standards and Technology (NIST) definition, ''cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction''. Its applications have been reported in business, industry, research, education, transportation and even national security. In healthcare, managers and experts also believe that it can increase Electronic Health Record (EHR) adoption, reduce in-house IT maintenance burdens and therefore improve healthcare services [9].

Cloud computing systems can be characterized as Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS). PaaS and IaaS are viable architectures that can be implemented and applied to CPS. IaaS applications provide system infrastructure as a service, which can allocate physical resources to serve their application's needs through programs. An example of IaaS applications can be found in Amazon's Elastic Compute Cloud (EC2) [10].

Cloud in healthcare

Cloudwave: Sahoo [11], describe a cloud computing platform called Cloud wave. Cloud wave is a big data platform for supporting real time access to multi-modal signal data used in multi-center collaborative epilepsy research. Cloud Wave will define and implement a new framework where specialized algorithms dynamically analyse cloud infrastructure and application behaviour, seamlessly integrate data pertaining to physical and virtual resources and IoT elements, and provide consolidated feedback to drive service evolution and adaptation.

The Cloud wave storage module extends HDFS and implements a Map Reduce-based data processing workflow to process EDF signal data generated from epilepsy centers and extracts data segments corresponding to specific signal channels. Cloud wave has developed a JavaScript Object Notation (JSON)-based expressive representation format called Cloud wave Signal Format (CSF) to store channel-specific data together with metadata information in a single file.

The Cloud wave query and visualization interface extends an open source visualization software called High charts. The visualization interface is integrated with the HDFS-based storage module to support efficient retrieval of data segments for multi-channel signal visualization. In Cloud wave, we have implemented two partitioning schemes that correspond to ontology classes modeled in the epilepsy domain ontology to support querying and retrieval of signal data segments for use in the visualization module.

Comprehensive cloud: Lin et al. [12] presented a comprehensive method for rapidly processing, storing, retrieving, and analyzing big healthcare data. The proposed methods and overall operational architecture are presented in Fig 2. In the present section, the MVC design pattern is used to explain the architecture in three steps.

In Model Layer They compiled all healthcare data of each patient into separate documents called patient-driven medical documents (PaMeDocs). Each patient has an independent PaMeDoc. PaMeDocs are used in document-oriented databases and their basic elements are key-value pairs such as (Birthday, "2013-12-31") where "Birthday" is the key and "2013-12-31" is the value. The key is used for identification and is unique within a PaMeDoc. One PaMeDoc can contain another PaMeDoc, forming a tree structure. The PaMeDoc storage platform can use an existing NoSQL database, preferably the document-oriented type.

In View Layer A Web-based user interface was adopted. The interface receives user requests and displays the results. In Controller Layer data reformulation, shading and using MapReduce, and conducting targeted queries and searches is explained. When restructuring these data into patient PaMeDocs, the basic information can be acquired from the ID table, after which the PK-FK relationship is determined to obtain the target data from the linked data tables. This recursive procedure continues until all patient PaMeDocs have been completed. When performing conditional searches, MapReduce uses the sharding-key to perform a targeted query, enhancing the search and computing performance.

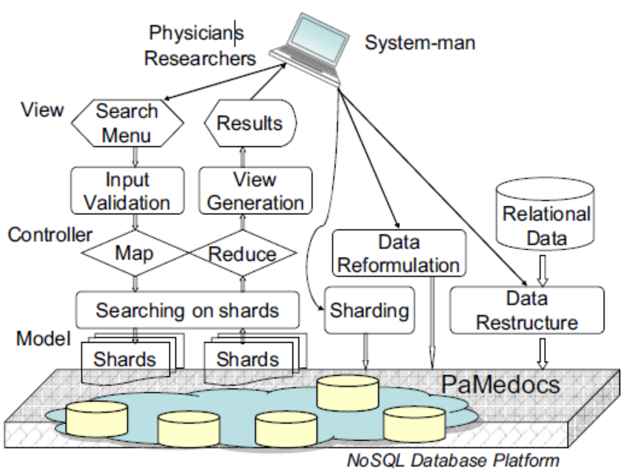

Fig 2. Overall operational architecture of the proposed methods

Virtual Cloud: Castiglione et al. [13] presented a Virtual infrastructure-less Cloud solution for secure management of 3D medical images, which operates in an almost completely transparent manner, regardless of computational and networking capabilities which users can avail in any given moment. This work proposes essentially two things. First, it introduces a novel engine for lossless dynamic and adaptive compression of 3D medical images. Second, define a ‘‘Virtual infrastructure-less Cloud’’ (VC) architecture for the management of these images, which is based on a peer-to-peer overlay network organization that dynamically provides the services offered by the aforementioned engine.

VC is logically composed by four main modules: Compression and Digital atermarking, Virtual Cloud, Storage and Front-end Interface. Compression and Digital watermarking concerns the part of data compression, such engine introduces two novel prediction models, namely, the intra-slice prediction model and inter-slice prediction model. The former exploits only the spatial redundancy, while the latter takes advantage of the whole three dimensional redundancy. It is important to point out that the neighboring pixels of the one which needs to be predicted, most likely have a similar intensity level, given that the human body part in exam is supposed to be the same with a good chance.

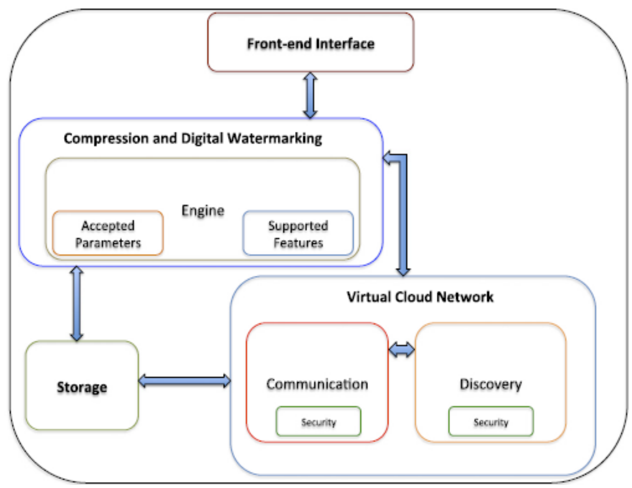

Fig2. The logical components of the whole solution

The Virtual cloud module provides a ‘‘global’’ mechanism for indexing and locating images within the VC, by allowing to know which node owns a given image. Moreover, such module is also able to detect nodes exposing certain features and consequently, which of them can perform certain tasks on behalf of the requesting user. These features may for example include the compression of images in particular formats or the embedding of a digital watermark with certain characteristics. Substantially, the component of discovery can be seen as the ‘‘glue’’ that holds together the various nodes belonging to the system. All nodes that want to become part of the VC, in order to share images or services, as well as for requiring them, have first to register themselves, together with the services or information they provide, within the cloud. After performing localization, the node which intends to access a specific image or a requested feature, can do that by means of the provided Communication Facility.

Storage module uses a storage system based on MySQL together with CryptDB. The architecture of

CryptDB can be subdivided into three main logical layers: Application Server (AS), CryptDB Proxy Server (CPS) and DBMS Server (DBMS).In particular, these internal components are able to intercept and manage SQL queries. Each layer can be hosted on a dedicated VM, in the case of sophisticated multi-node/multi-processor architectures, or all the three components can be executed as different processes (or set of processes) on a single VMlocated on the storage node. The main aim of the AS is to enable the interaction between the users and the other CryptDB layers. The fundamental assumption on which CryptDB relies is that all the queries pass through the CPS, connected to the AS. Therefore, the proxy server is responsible for encrypting and decrypting all the data traversing it, as well as for modifying some operators of the query which is currently under processing.

The Front-end Interface, running on end-user terminal nodes, acquires and displays patient records stored into the VC, while triggering the Discovery and data transfer operations. The main functionality of such module is providing medical experts and patients with an easy-to-use User Interface (UI) for managing healthcare information. Such interface enables to store, search and retrieve medical images, patient health records and patient-related medical data. In detail, this module allows the user to specify which images it intends to obtain, decodes and displays them along with other information regarding the patient status, in order to perform their relative accurate analysis and diagnostic assessment. As a consequence, the rendering as well as the visualization of 3-D medical images is in charge of the Front-end Interface, which is the ‘‘point of connection’’ between the end-user and the VC.

Task-level cloud: The major contribution of Zhang et al. [14] is a task-level adaptive MapReduce framework to process streaming data in healthcare. The framework is designed to scale in heterogeneous cloud platforms by applying four scaling theorems and scaling corollaries. Stream data scientific applications need to estimate the real time data arrival rate and plan for the computing resources accordingly. In this paper, propose two workload prediction methods and compared their benefits and performance in real life healthcare scientific applications. Real streaming data trace is used to justify the applicability of the framework. Paper reports the experimental results by showing the real time Map and Reduce tasks number variation, which matches perfectly with the variation of the streaming data arrival rate.

The Hadoop MapReduce is essentially a scheduling framework that processes data that can be sliced into different splits. Each Map task works on its own input data split without having to interact with other Map tasks at all. The Hadoop MapReduce framework can only be applied to process input data that have already existed. However, real-life scenarios of the state of-art big-data applications that typically require the input data be provisioned in streaming and be processed in real-time. Therefore, an enhanced MapReduce framework is required to cater for such a need. That is the motivation behind out design of the task-level adaptive MapReduce framework. An adaptive MapReduce framework is proposed to process the streaming data in real-time. A significant challenge here is how to address the incoming data streams with the varied arrival rate. There are four technical issues that we should consider when designing the adaptive framework. First, the framework should be horizontally and vertically scalable to process a mixture of varied workloads. Second, scaling the number of the Map and Reduce tasks should align with the scaling of the cluster size. Third, they also need to consider the heterogeneity of the processing capabilities of different Map tasks. Fourth, the optimal runtime Map and Reduce task count should be specified.

There are two ways to feed data split streams to the Map tasks. A proactive strategy caches streaming data locally first and pushes them every fixing period of time, for example every one minute. As an alternative option, data splits can also be pushed in a reactive way. In other words, a cache size is defined in HDFS before the input data starts to move in, whenever the cache usage hits a ratio, say 85%, the data splits begin to be pushed to the Map tasks.

The adaptive MapReduce framework starts from a novel runtime scheduler that feeds different Map tasks with different number of data splits. Suppose the first Map task is executed on a faster compute node and has processed two splits of the input data while the second Map task has processed only one. This leads to Map task one has four data splits while Map task two has two, and the total execution time of the Map stage is minimized.

Comparison healthcare’s clouds

Table 1 indicates that how each cloud solves healthcare's big data problems. There are 4 kinds of data in healthcare: EHR, 2D or 3D images, electrophysiological signal and streaming data. Each cloud works on only one kind of data and couldn’t store and process all of them. All clouds are not shown to be able to solve healthcare's big data problems. Most Clouds have used the map/reduce technique for processing data, so we can say that map/reduce is the only solution for processing Big Data at present. To have rapid respons, databases must be pre-processing, Usege of Nosql is also a way of increasing respons rate.

Cloud Wave will define and implement a novel architecture which incorporates the three main pillars. Implementation will be based on the Open Stack technology, while leveraging other open tools and standards. Focus on the electrophysiological signal to answer three questions: (1) What are the computational challenges that need to be addressed to effectively leverage healthcare big data; (2) Why should semantic computing approaches be used to address many of these challenges; and (3) How does a biomedical big data platform called Cloud wave use open source Hadoop technology stack with domain ontology to leverage biomedical big data for healthcare research. In comprehensive cloud Based on the PaMeDoc and shard design, Ajax and D3.js can be used to easily render various statistical distribution charts and data tables, specifically the timeline visual representation of individual patient records.

The information is dispersed in various shards, when the user searches the information, the query conditions are mapped to the corresponding sharding key and parallel searching is conducted on the corresponding shard. comprehensive cloud Although no fundamental theory or universal data modeling technique existed for use in NoSQL data bases, their strong expansibility and flexibility enabled abandoning relational data model for a novel cloud database.

IN Virtual Cloud the performance of such engine has been experimentally evaluated by demonstrating that it guarantees (on average) better performance in terms of Bits Per Pixel (BPP) than the most common 3D lossless compression algorithms available in literature. VC is responsible for transforming 3D images into an appropriate representation before being transferred to their requestors, in order to deal with heterogeneous devices and varying scenarios and network conditions, as well as providing the best information accessibility and perceived quality of data access.

In task-level cloud to process stream big-data in real-time, traditional parallelized processing frameworks, such as Hadoop MapReduce, Pregel, and Graphlab, are structurally constrained and functionally limited. The major difficulty lies in their designs are primarily contrived to access and process the static input data. No built-in iterative module can be used when the input data arrives in a stream flow. The existing frameworks are unable to handle the scenarios when the streaming input datasets are from various sources and have different arrival rates. Healthcare scientific applications vary the data acquisition frequency when the behavior of the person changes. For example, the data collected when a person is sleeping can be far less than the data collected when the person is running or swimming.

Table 1. Comparison of healthcare‘s cloud

Items	Cloudwave	Comprehensive cloud	Virtual cloud	Task-level cloud
Variety	electrophysiological signal data	EHR	3Dmedical images	streaming data
Velocity	MapReduce	MapReduce	Kademlia	MapReduce and Kalman
Volume	HDFS	NoSQL	Mysql	HDFS
Veracity			security	privacy
Visualisation	CSf files	PaMedocs files	Files in nodes	index files and Lucene documents
Emphasis	Manage signal data	rapidly searched and query process	secure and flexible access	Scalable cloud
Cloud	IaaS	IaaS	SaaS	PaaS

CONCLUSION

Recent advancements in cloud computing technology show promises of lower cost, higher scalability, accessibility, availability and disaster recoverability. In this paper, a study of cloud computing architecture and comparison of Some `Vs' of Big Data solution in some important cloud in healthcare. We presented the current research issues on Cloud Computing in healthcare and thus motivated the potential of cloud marketplaces for healthcare.

REFERENCE

[1] Chen M, Mao Sh, Liu Y, “Big Data: A Survey”, Springer, Mobile Netw Appl. 2014; 19:171–209.

[2] Marcos D. Assun, Rodrigo N. Calheirosb, Silvia Bianchic,Marco A. S. Nettoc, Rajkumar Buyyab, “Big Data Computing and Clouds:Trends and Future Directions”, Journal of Parallel and Distributed Computing, 2014; 75:140-184.

[3] Lin Ch-H, Huang L-Ch, Liu Ch-H, Temporal Event Tracing on Big Healthcare Data Analytics”, 2014; IEEE International Congress on Big Data (BigData Congress)

[4] R.F.Chong,Changingtheworld:big data and the cloud, (2012), ⟨http://www.theatlantic.com/sponsored/ibm-cloud rescue/archive/2012/09/changing-the-world-big-data-and-the-cloud/262065/⟩

[5] M.Poulymenopoulou,F.Malamateniou,G.Vassilacopoulos, E-EPR,”acloud-based architecture of an electronic emergency patient record”, 2011; in: Proceedings of Pervasive Technologies Related to Assistive Environments

[6] M. Li, S. Yu, Y. Zheng, K. Ren, W.Lou, ”Scalable and secure sharing of personal health records in cloud computing using attribute-basedencryption”, 2012; IEEE Trans. ParallelDistr. Syst. 24: 1131–143

[7] Benjamin Fabian , Tatiana Ermakova , Philip p Junghanns ,” Collaborative and secure sharing of healthcare data in multi-clouds”, 2015; Information Systems, 48: 132–150

[8] Chauhan R, Kumar A, The 4th IEEE International Conference on E-Health and Bioengineering, 2013.

[9] Mu-Hsing K, “A Healthcare Cloud Computing Strategic Planning Model ”,Computer Science and Convergence Lecture Notes in Electrical Engineering 2012; 114: 769-775.

[10] Chonka A, Abawajy J, “Detecting and Mitigating HX-DoS attacks against Cloud Web Services”,15th International Conference on Network Based Information Systems, 2012.

[11] Sahoo S.S., “Biomedical Big Data for Clinical Research and Patient Care: Role of Semantic Computing”, 2014; IEEE International Conference on Semantic Computing

[12] Lin Ch-H, Huang L-Ch, Chou S-Ch T. “Temporal Event Tracing on Big Healthcare Data Analytics”, 2014; IEEE International Congress on Big Data.

[13] Castiglione A, Pizzolante R, De Santis A, Carpentieri B, Castiglione A, Palmieri F, “Cloud-based adaptive compression and secure management services for 3D healthcare data”, Future Generation Computer Systems 2015; 43–44, 120–134.

[14] Zhang F, Cao J, Khan SU., Li K, Hwang K, “A task-level adaptive MapReduce framework for real-time streaming data in healthcare applications”, Future Generation Computer Systems 2015; 43–44, 149–160.