Big Data and Cloud based Healthcare System

Kiran Tekwade[1], Shweta Bapure[2], Pratiksha Ganage[3], Nikita Parakh[4]

[1] *Student, Computer Engineering, Sinhgad Academy Of Engineering, Maharashtra, India*
[2] *Student, Computer Engineering, Sinhgad Academy Of Engineering, Maharashtra, India*
[3] *Student, Computer Engineering, Sinhgad Academy Of Engineering, Maharashtra, India*
[4] *Student, Computer Engineering, Sinhgad Academy Of Engineering, Maharashtra, India*

ABSTRACT

The aim of this paper is to provide information about big data and cloud used in health care system. This paper includes guidance as well as strategies for design to help evaluation and comparison between cloud computing from different cloud providers. Big data refers to datasets that are not only big but also high and very large in variety and velocity, which makes them difficult to handle using tools and techniques. Amount of data produced within health informatics has grown to be quite vast and analysis of these data information can improve quality of healthcare offered to patient. However there are number of issues that arises when dealing with cloud computing ,heath care consumers must have clear understanding of unique benefits and risks associated with cloud computing. Including Big data analytics in health care system, it provides users which have potentials to advance personalized care for improving patient details and avoid unnecessary costs.

Keyword: - *Big Data, Cloud, Healthcare System, MongoDB, desktop based medical management, patient database on cloud.*

1. Introduction

Data is the building block upon which any organization or system depends. Imagine a world without any kind of data storage; a place where every detail about a person or organization, every transaction which have performed, or every aspect which can be documented is lost directly after use. Organizations would not be able to extract valuable information and knowledge, achieve detailed analyses, as well as provide new opportunities and advantages. With the increase in storage capabilities and methods of data collection, large amounts of data have become easily available. Every minute, more data is being created and it needed to be stored and analyzed in order to extract value. Furthermore, data has become cheaper to store, so organizations they necessity to get as much value as possible from the large amounts of stored data. The size, variety, and rapid change of such data require a new category of big data analytics, as well as different storage and analysis methods. The contribution of this paper is to provide an analysis of the available literature on big data analytics as well as cloud computing.

The information is in the form of big data, so- called, not only for the sheer volume but for its complexity, diversity. For instance researchers can mine the data to see what treatments are most operative for particular conditions, identify patterns related to drug side effects or hospital readmission and gain other important information that can help patient and reduce cost. Fortunately, recent technologies advances in the industry have enhanced their ability to work with such data even though the files are enormous and often have different data base structures and technical characteristics. Today it is a necessity to used cloud in health care. Different cloud providers, taking into account different requirements from various sectors including medical, hospitals, research, insurance companies and governments. To use cloud computing, health care consumers must have a clear understanding of different benefits and risks associated with cloud computing. They have to also set the realistic expectations with their cloud provider.

There are different models of service delivery such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) and cloud deployment models such as private, public, and hybrid.

2. Big Data and its evolution [2, 3]

The amount of data produced is increasing at a very rapid rate day by day. At the start of the computer era the size of the data was measured in KB (Kilo Byte) to MB (Mega Byte), GB (Giga Byte), TB (Tera Byte), PB (Peta Byte), EB(Exa Byte) an now recently the digital world is dealing data of sizes in ZB(Zeta Byte).

So far there is not a single standard definition defined for the term big data, in big data [3] is defined as it is a collection of datasets that are so large as well as complex that it becomes extremely difficult to process using on-hand management tools or the customary data processing applications , Big data is defined as large and difficult or complex datasets made up of variety of structured and also unstructured data which are too big, too fast or too hard to be managed by customary techniques[2]. Therefore, big data analytics is where advanced analytic techniques are applied on big data. Analytics based on large data samples reveals and leverages business transformation. However, the bigger the set of data, the more tough it becomes to manage [3].

2.1 Characteristics of Big Data [3]

Big data is data whose scale, distribution, diversity, and/or timeliness require the use of new technical architectures, analytics, and tools in order to permit insights that unlock new sources of business value. There are three main features characterize big data and those are : volume, variety, and velocity, or the three V's. The volume of the data is its size, and how enormous it is. Velocity states the rate with which data is varying, or how often it is created. Finally, variety consists of different formats and categories of data, as well as the different kinds of uses and ways of analyzing the data [3].

1. Volume
2. Velocity
3. Variety
4. Veracity
5. Value

There has been growth in the rate of advent of data which is because of the following main reasons-

a. Increasing automation process

b. Increasing connectivity between multiple systems

c. Increase in the social interaction between the people.

3. MONGODB[6]

MongoDB is a powerful, flexible, and also scalable general-purpose database .It combines the capacity to out with features such as secondary indexes, range queries, sorting, aggregations, and geospatial indexes. MongoDB is a document-oriented database, MongoDB isn't a relational one. The primary reason for affecting away from the relational model is to make scaling out easier, but there are some other advantages as well.

3.1 Characteristics of MongoDB [7, 8]

MongoDB is intended to be a general-purpose database, so besides from creating, reading, updating, and deleting data, it provides an ever-growing list of unique features

3.1.1. Indexing

MongoDB supports simple secondary indexes, a change of fast queries, and provides unique, compound, geospatial and full-text indexing capabilities as well.

3.1.2. Aggregation

MongoDB supports an "aggregation pipeline" that permits you to build complex aggregations from modest pieces and allow the database to enhance it.

3.1.3. Special collection types

MongoDB supports time-to-live collections for data that should finish at a certain time, such as sessions. It supports fixed-size collections also, which are helpful for holding recent data, such as logs.

3.1 4. File storage

MongoDB supports an easy-to-use protocol for storing large files and file metadata. In addition to combine documents by collection, MongoDB groups gatherings into databases.

4. JSP [9]

Java Server Pages (JSP) is a technology that benefits software developers create enthusiastically generated web pages based on HTML, XML, or other document kinds. Released in 1999 by Sun Microsystems, JSP is alike to PHP and ASP, but it uses the Java programming language. To arrange and run Java Server Pages, a well-matched web server with a servlet container, such as Apache Tomcat or Jetty, is required. JSPs are generally used to deliver HTML and XML documents, but through the use of Output Stream, they can distribute other types of data as well. JSP pages use numerous delimiters for scripting functions. A scriptlet is a fragment of Java code that is run when the user requests for the page. Other delimiters contain <%= ... %> for expressions, where the scriptlet and delimiters are exchanged with the result of evaluating the expression, and directives, denoted with <%@ ... %>.

Java code is not compulsory to be complete or self-contained within a single scriptlet block. It can straddle markup content, which provides that the page as a whole is syntactically correct. For example, any of Java if/for/while those blocks is opened in one scriptlet they must be correctly closed in later scriptlet for the page for successfully compilation.

The compiled pages, as well as any of dependent Java libraries, in which contains Java byte code rather than any machine code. Like any other Java program, they must be executed within a Java virtual machine (JVM) that relates with the server's host operating system to provide abstract, platform-neutral environment.

5. Cloud [1, 4, 5]

The basic concept of the cloud , based on the services they propose, from application service provisioning, grid and service computing, to Software as a Service [4, 5]. Even though the specific architecture, the dominant idea of this computing model is that customers' data, which can be of individuals, organizations or enterprises, is processed remotely in unidentified machines about which the user not responsive. The ease and productivity of this approach, however, comes with privacy and security risks. Confidentiality of data is the main obstacle in implementation of cloud services. A huge data centers are recognized in cloud computing, but the deployment of data and services are not trustworthy. These create innumerable new security challenges and these challenges are vulnerabilities in accessibility, virtualization and web, such as SQL injection, cross site scripting, physical access issues, privacy and control issues trendy from third parties having physical control of data, issues related to identity and credential , issues related to data confirmation, changing and privacy, data loss and theft, issues related to integrity and also IP spoofing.

5.1 Characteristics of Cloud Computing [5]

Cloud computing provides five necessary characteristics as defined by NIST (National Institute of Standards and Technology)[8].

5.1.1. On-demand self-service. A consumer can unilaterally offer computing capabilities.

5.1.2. Broad network access. Capabilities are accessible over the network and accessed through a standard mechanism that promotes use by mixed thin or thick client platforms.
5.1.3. Resource pooling. The provider's computing resources are pooled to assist multiple consumers, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

5.2 Cloud Deployment Models [5]

Finally, NIST identifies four different deployment models for the foregoing service models. These deployment models are as follows:
5.2.1. Private cloud. The cloud infrastructure is operated exclusively for one organization. It may be achieved by the organization or a third party. Possibly this may be the most secure type of infrastructure, depending on the nature of the controls deployed.
5.2.2. Community cloud. In this model, the cloud infrastructure could be shared by several organizations and supports a specific community or interest group that has shared apprehensions (e.g., mission, security requirements, policy, and compliance considerations). It may be achieved by the organizations or a third party and may happen on premise or off premise.

5.2.3. Public cloud. The cloud infrastructure is made accessible to the general public or a huge industry group and is owned by an association selling cloud services.

5.2.4. Hybrid cloud. The cloud infrastructure is a configuration of two or more clouds (private, community, or public) that remain exclusive entities but are bound together by standardized or proprietary technology that allows data and application portability (e.g., cloud bursting for load-balancing between clouds).

5.4 Benefits:
Cloud computing proposals significant profits to the healthcare sector: doctor's clinics, hospitals, and health clinics require speedy access to computing and large storage amenities which are not provided in the traditional settings. Cloud provides to all the necessities thus providing the healthcare organizations an incredible chance to improve services to their customers, the patients, to share information more effortlessly than ever before, and develop operational efficiency at the equivalent time.

5.5. The five phases of managing health data in the Cloud
5.5.1) Acquire – The cloud services provider must have the ability to obtain data from various sources in their local format and encrypt data as it travels to the cloud provider's data center.

5.5.2) Transform – In the increasingly two-way environment of healthcare, data from diverse sources they must be normalizing and Synthesized into any format that permits all of the data to be used in any analytical applications.

5.5.3) Manage & Securely Store – On a related note, the cloud services provider should have the prospective to support numerous databases and Big Data suite options. It is also critical to work with a cloud services provider that recognizes the explicit wishes adjacent protected health information.

5.5.4) Analyze – While many will present some or all of the first above three phases, top-tier cloud services suppliers will have the ability to support healthcare organizations with incorporating Big Data analytics tools that allows subject matter experts to investigate into the data and use it to answer the organization's most critical questions.

5.5.5) Present – While the skill to slice and dice data is supreme, so too is presenting the information in user-friendly manner across altered systems and devices, including PCs, tablets, smart phones and the web.

This paper will enlighten new ability for healthcare analytics, the key engine for the "Big Healthcare Data" projects, interchange their data to cloud services provider with an exclusive Focus on health data management.

6. Deployment Model

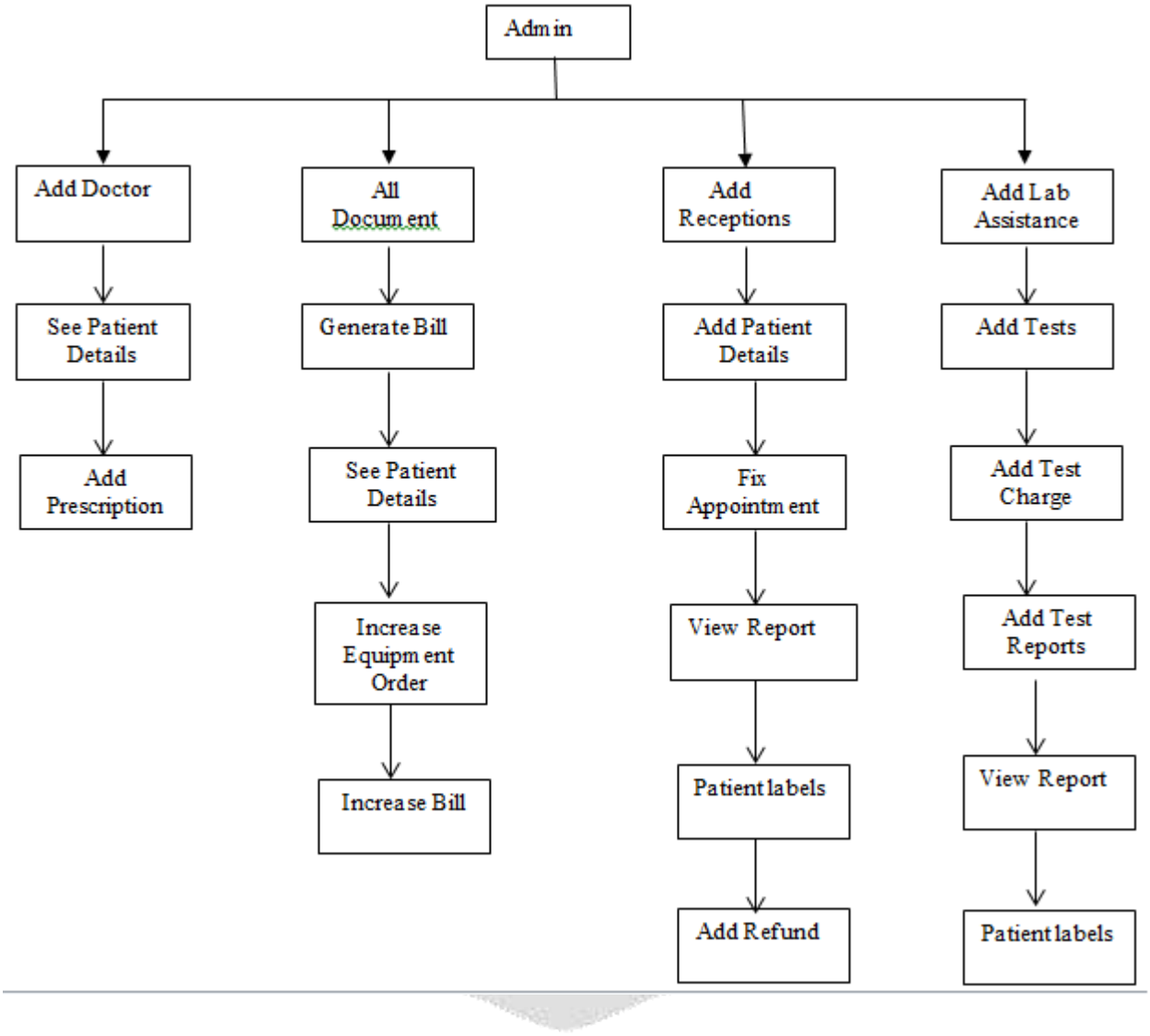

Fig 6.1. Deployment Diagram.

7. CONCLUSIONS

The system which brings about General Hospital has approved that the manual method of caring patient records should be change to computerized patient medical records which will help them to reduce inefficiency and unreliability associated with the manual method. In this system the all information are store like Appointment, Selection of surgery, medicine, patient report, billing etc. System will be on cloud so users can see data from anywhere in the world. This system's aim is reducing the documentation type work.

Healthcare is moving to mutual and shared model focused on patient outcomes. As dominant analytics are at the center of these changes, now it's a time to arrange these new advances in our system.

8. REFERENCES

[1] T. Mather, S. Kumaraswamy, and S. Litif, Cloud Security and Privacy: An enterprise perspectives on Risks and Compliance (Theory in Practice). O' Reilly, 2009.
[2]. Ms.Manisha saini, Ms.Pooja Taneja, Ms.Pinki Sethi "Big Data Analytics: Insights and Innovations" International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X
[3]. Alejandro Zarate Santovena "Big Data: Evolution, components, challenges and opportunities" pp 27-33
[4] Cloud Computing: Clash of the clouds. the economist., 2009.
[5] Flavio.L and R. D.P. Transparent Security for Cloud. in SAC '10: Proceedings of the 2010 ACM Symposium on Applied Computing. 2010.

[6]MongoDB.Available. http://www.mongodb.org

[7] K. Chodorow and M. Dirolf, MongoDB: The Definitive Guide. O'Reilly Media, 2010

[8] P. Eelco, M. Peter, and T. Hawkins, The Definitive Guide to MongoDB: The NoSQL Database for Cloud and Desktop Computing, vol. 44, no. 5. Apress, 2010,p.328.

[9] JSP 1.1 Syntax Reference