

# Healthcare Information System Architecture Design Based on Big Data

Zifan Zhang<sup>1, a</sup>

<sup>1</sup>International School, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>a</sup>zhangzifan1612@163.com

**Keywords:** Big data, information system, healthcare.

**Abstract:** The rapid development of medical information leads to the exponential increase of medical data. Medical big data has brought huge pressure to the existing hospital information systems. With the emerging of various kinds of unstructured data, the existing medical information system fails to meet the requirements of large data in storage space, storage speed, and storage structure, so it has to partially discard data, resulting in the loss of valuable medical data. However, the data integrity of the existing system is not enough, and the data processing speed is slow, hence it is obviously unable to meet the needs of users on data processing efficiency. Based on this, this paper puts forward the structure of intelligent medical system for large data, and analyzes the infrastructure layer, the information processing layer and the service provider layer of the three aspects.

## 1. Introduction

With the development of science and technology, diversified medication indicates the frequent usage of devices with various sensors. X-ray and magnetic resonance imaging have caused exponential growth of medical data, which are abundant appearances of text, pictures, videos and other semi-structured and non-structured data. Big data and unstructured data present a huge challenge to the health system. On the one hand, with the continuing emergence of all kinds of unstructured data, the existing medical information system fails to meet the requirements of big data in storage space, storage speed, and storage structure, so a considerable amount of data has to be gave up, causing loss of a large number of valuable medical data. On the other hand, with the deep understanding of big data, hidden values in medical big data becomes a treasure [1]. However, the data integrity of the existing system is not enough, and the data processing speed is slow, it is obviously that it is unable to meet the users' needs of the data processing efficiency. These years, more and more countries play an active part in the reform of medical information, with the rapid development of technology and a large amount of research investment, big data analysis of massive data is no longer elusive. The world-renowned consulting firm McKinsey pointed out in its report that US medical services can obtain \$300 billion additional worth through big data analysis without considering the institutional obstacles. The analysis and application of big data can greatly improve the efficiency and effect of medication [2].

The big data analysis technology is applied in the field of health care. The medical data were analyzed by using data mining and analysis technologies and combined with the traditional medical data, achieving precise and personalized health care services. In clinical application, the use of big data technology analysis can achieve disease pattern analysis, clinical effect comparison, susceptible population (high risk) analysis, personalized treatment, clinical decision support, remote patient monitoring (such as the analysis of chronic disease effect, public health monitoring) and etc. For example, through the analysis of electronic medical records of patients in the country, medics can quickly determine whether a disease outbreak happens or whether a new infectious disease is discovered. If it is, the public health department can start epidemic surveillance and response procedures in a timely manner. Therefore, the use of medical big data can reduce the rate of infectious diseases and improve public health monitoring. In terms of management, medical big data can be used for disease classification, resource management, quality control and other operations [3]. Under this background, this thesis proposes the wisdom medical system structure of big data, and analyze

the three aspects from processing infrastructure layer, data layer and information service layer to provide a reference for upgrade of hospital information systems as well as the construction of regional medical society.

## 2. General Structure of Systems

Medical information is not simple computerization, but it is based on information sharing, and information sharing is in each medical institution and among medical institutions, community information systems, health administrative departments and medical insurance institutions. It is helpful to patients during the whole process of seeking medical advises, while improving the efficiency of medical work, and bring convenience for all kinds of decision analysis of management personnel, giving full play to the value of application of information technology in the medical industry, and helpful to the utilization of medical resources.

The intelligent medical information system, based on big data, is divided into three layers: infrastructure layer of information acquisition and storage, information integration layer of data processing and analysis, and application Service Provider that provides services for providers of medical institutions and medical software, the overall structure of the system is shown in Figure 1.

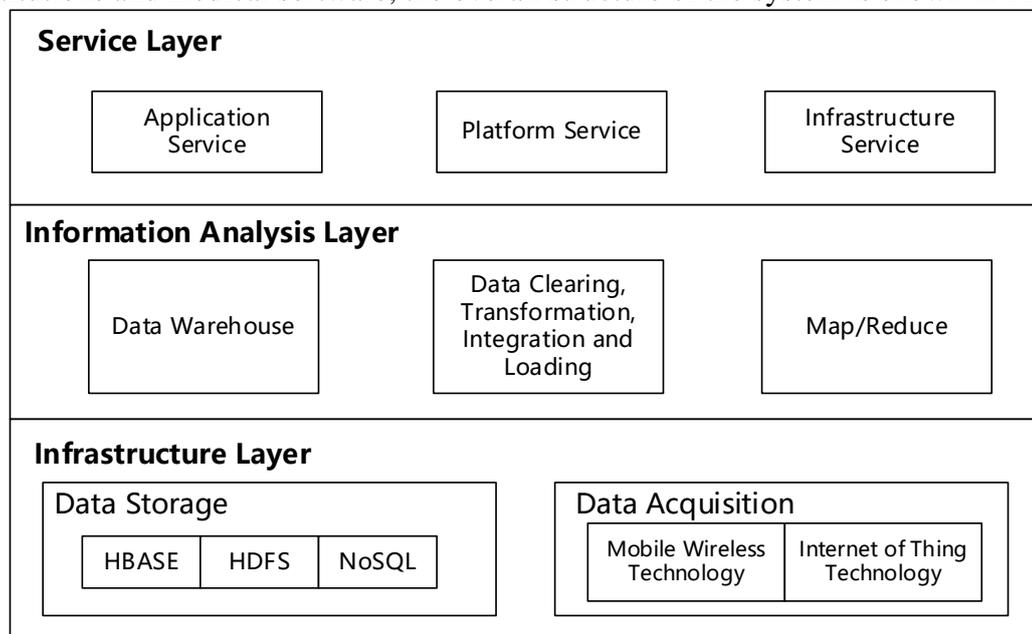


Figure 1 Architecture of the healthcare information system based on big data

## 3. Design on Each Layer

### 3.1 Infrastructure Layer of Medical Big Data

With the popularization of hospital information system, a lot of medical information is generated every day, including image information (as CT, MR), vital signs, clinical examination, diagnostic information and other information. These resources are precious properties of the hospital, and have important value in the patient's condition diagnosis, clinical comprehensive display and medical research and other aspects. The basis of the use of medical big data is effective access and storage of these information. The infrastructure layer contains the collection layer, storage layer and analysis layer.

As the existing hospital information system is a desktop information system with HIS serving as core, and cannot cover all aspects, the closed management has breakpoints, which cannot meet the requirements of "Data completeness", thus the information collection layer is established to obtain medical information completely. The acquisition layer mainly uses mobile wireless technology (smart PDA, Pad, Mobile trolley, wireless AP) and things networking technology (Zigbee wireless sensor network, RFID tags, various types of environmental sensors) to form a floor unified platform

for the Internet of Things to achieve the integrity, convenience and authenticity of information collection.

The huge amount of data in the medical system causes situations that one single machine cannot handle, and it is necessary to take the distributed storage system. In addition, the medical data contains a lot of unstructured and semi-structured data, which produces problems to traditional relational database, but Hadoop can well deal with this difficulty. However, Hadoop is not designed for replacing the existing relational database systems [4]. Hadoop and relational databases have both advantages and disadvantages in processing massive data, regarding to the situation and problems of its own. It is easier to research and analyze complex data using Hadoop, but relational database is good at dealing with intensive computing tasks. In practical applications, Hadoop and existing relational database systems are deployed together in a new way to form a powerful combination.

Therefore, in the medical industry, we will store highly consistent structured data in the traditional relational database. Unstructured data and low consistency data, which are hard for relational database to deal with, are stored in HDFS or distributed NoSQL database. Hadoop is designed to store huge amounts of data, and in accordance with any way to deal with large data and according to the needs of the data exchange system to arbitrary transfer. Therefore, the data from the traditional database system can be imported into Hadoop, hence the traditional relation database system can be designed to handle data intensive task, while Hadoop can complete complex and time-consuming big data analysis, so the combination of two storage methods has great advantage to improve the efficiency of the system [5].

### **3.2 Analysis Layer of Medical Big Data**

The main purpose of the big data analysis layer is to provide convenient and easily-used data analysis interface to the application layer, and help the application program to use the distributed big data computing platform, so as to simplify the difficulty of the developing application. Therefore, the big data analysis layer needs the function of reading data and data cleaning, operation and returning results as required. As medical data has features of large amount, complex structure, high processing speed, if the data was stored in the distributed system, the traditional big data analysis tools and algorithms cannot fully utilize the advantage of distributed system. In order to solve his problem, information analysis layer encapsulates some extensible classical algorithm, and according to the hospital's business needs, several other medical data analysis tools were implemented using basic algorithms, facilitating the development of intelligent applications.

The traditional data mining model is a single task serial execution structure, the size of data can set a bottleneck on data processing ability, so we need to use distributed computing to solve this problem. However, Hadoop distributed computing platform use concurrent processing mode, hence big data will be decomposed into small sets of data. These data sets are distributed to each node, and performed on each node by mapping and protocol function. Therefore, in order to deploy the data mining to distributed computing platform, we need to modify Hadoop data mining model, making it better adapt to the parallel computing and MapReduce programming model of distributed computing platform, and improve the security and stability of data mining architecture.

The data mining model is deployed to the Hadoop distributed computing platform, Mahout tool library package of Hadoop platform is wrapped in mining layer of improved model, Mahout can do parallel processing to the traditional data mining algorithm, and it is suitable for MapReduce operation of distributed computing platform, so we can realize the operation of request of the data mining algorithm by calling Mahout to provide the data mining algorithm.

### **3.3 Service Provider of Medical Big Data**

Service provider layer of medical big data adopts distributed computing model. Therefore, the intelligent medical system of big data can provide three kinds of services: IaaS (Infrastructure as a service), PaaS (Platform as a service) and SaaS (Software as a service). Considering the requirements of medical information of medical institutions and patients are diverse and flexible, IT services which is provided by distributed platform could meet the application requirements on different levels of application of medical institutions.

IaaS encapsulates hardware equipment and other basic resources as service for users to use. Through the building of Hadoop cluster environment of infrastructure, it can provide services for the medical software developers. This is the usage of all computing infrastructure, including the processing of CPU, memory, storage, network and other basic computing resources. The users do not manage or control any distributed computing infrastructure, but can control the storage space, the deployment of the application, they may also have limited access to network components (such as routers, firewalls, load balancers and so on).

PaaS provides working environment to help creating application service by providing integrated development tools and software for medical software vendors, medical institutions and medical management department and other users, and support medical institutions of the distributed computing platform for subsequent develop and use. By using a variety of interface calls provided by this platform, software development kit (SDK), data mining engine and other basic services, this service can rapidly develop various types of software to meet self-need and seamlessly integrated into the existing platform.

SaaS mainly provides basic functions of software which is used to support the operation of medical institutions (such as HIS, LIS, PACS, EMR, HER, medical management platform, vital signs monitoring system, personalized medical service, financial management system, material management system and so on), so the medical institutions can quickly obtain the basic business functions of software, and obtain more professional software service with less IT input.

#### 4. Conclusions

Through the study of medical information technology, big data technology and the application of big data in the medical industry, this paper makes a full exposition of the intelligence of the medical system structure based on big data, and the structure includes the overall architecture of large data intelligent medical system model, the acquisition and storage of medical data, the analysis of medical big data, the use of medical big data, and etc. This study provides a reference for the system upgrade of hospital information and the construction of social regional medical treatment.

#### 5. References

- [1] Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama*, 309(13), 1351-1352.
- [2] Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The big-data revolution in US health care: Accelerating value and innovation. *Mc Kinsey & Company*, 1-13.
- [3] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 1.
- [4] Nandimath, J., Banerjee, E., Patil, A., Kakade, P., Vaidya, S., & Chaturvedi, D. (2013, August). Big data analysis using Apache Hadoop. In *Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on* (pp. 700-703). IEEE.
- [5] Pavlo, A., Paulson, E., Rasin, A., Abadi, D. J., DeWitt, D. J., Madden, S., & Stonebraker, M. (2009, June). A comparison of approaches to large-scale data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data* (pp. 165-178). ACM.