

Construction and practice of Linfen regional healthcare big data intelligent platform

Xue-feng Su, Tie-zhong Wang, Bo Zhao, Shu-wei Zhang, Jun-feng Zheng, Hui Li, Wei-hong Guo, Lin-peng Wang, Peng Zhao, Ya-jie Zhou, Yue Liu
Linfen People's Hospital, Shanxi, China

Abstract. The construction and practice of Linfen Regional Health and Medical Big Data Intelligent Platform aims at creating a regional medical big data intelligent Platform that conforms to national standards and can better serve the life and health of local residents, and providing residents with health management services throughout the life cycle; better improving the operation efficiency of diagnosis and treatment, consultation and referral for medical institutions, and improving the service bearing capacity; providing data for health administrative departments and promoting the healthy development of Linfen regional economy and society.

1. Introduction

Health and medical big data refer to the relevant data generated by human health activities in social life, including the whole process of health activities and the health and medical big data throughout the whole life cycle. It has become an important basic strategic resource of China. The research shows that by 2020, the medical data rapidly increased to 35ZB, which was 44 times the data volume in 2009. At present, the phenomenon of "Data Island" in medical and health institutions is widespread, and the data format among institutions and Systems is ununiform. Establishing a regional Healthcare Big Data Platform and making scientific and rational use of Healthcare big data can improve the Healthcare service mode, better address the urgent needs of the people, and further promote social development.

2. Construction scheme of Linfen regional healthcare big data intelligent platform

The *Outline of the National Medical and Health Service System Planning (2015-2020)*, the *Key Tasks of Deepening the Reform of the Medical and Health System* in 2019, and the *Guidelines on Promoting and Standardizing the Application and Development of Health and Medical Big Data* issued by the State Council require to establish interconnected national, provincial, municipal and county population health information Platforms comprehensively, the application and development of Healthcare big data shall be integrated into the national big data strategic layout, and the construction of the national Platform for national health information and the provincial overall planning regional Platform shall be promoted continuously. A medical Big Data Platform shall be established urgently in Shanxi Province to solve the problems of data cleaning, storage, mining, analysis and utilization with the rapid development of mobile Internet, cloud computing and other information technologies.

2.1 Construction ideas of healthcare big data intelligent platform

Through the construction of Linfen Regional Big Data Platform, the medical data in the region will be cleaned, stored and mined, unified integration and management will be implemented, the problems of cleaning, collection, transmission, storage, processing, retrieval, calculation, analysis and utilization of medical data will be solved, the medical resources at the city, county, township and village levels in Linfen City will be gathered and integrated, and a medical and health Big Data Platform will be established, and the extraction and application of health data will be realized.

The top-level architecture design of the System is based on the national unified data exchange standards and business specifications. The System Platform is established through six layers according to the localization principle and the main index mechanism, and following the rules of platform level linkage, including data interface layer, platform data layer, service bus layer, business service layer, platform management and control layer and application display layer, to realize the interconnection of diagnosis and treatment information of medical and health institutions, and support cross-regional information query, medical service and comprehensive management.

The platform data deployment adopts the principle of localization of residents, and the data storage adopts hierarchical deployment. The information resource database of the platforms at all levels mainly includes medical records, electronic medical records, and examination and inspection results, etc.; the Platform saves the health records and other relevant information of the residents in the region. The database is built by MySQL database cluster, and the data and applications between regions are completely isolated.

The design of the System adopts the current advanced and stable technology, fully absorbs the advanced experience of manufacturers at all levels, and considers the localization situation, to meet the existing needs and grasp the future development direction; meanwhile, the design of the System also meets the needs of large concurrent transactions.

2.2 Architecture design of healthcare big data intelligent platform



Figure 1. System architecture of Linfen regional healthcare big data platform



Figure 2. Hardware architecture of Linfen regional healthcare big data platform

2.3 Healthcare big data intelligent platform sub-platform module

The System includes sub-platforms, such as platform management and control, two-way referral, family doctor and remote consultation, as shown in Fig. 3.



Figure 3. Interface of Linfen regional healthcare big data intelligent platform

2.3.1 Platform management and control

Platform management and control is the core of Linfen Regional Health and Medical Big Data Intelligent Platform, including organizational structure, personnel management, permission setting, etc., and it is a bridge to realize data interaction among sub-platforms in the Platform.

2.3.2 Two-way referral

The downward transfer channel is provided from the third-class hospital to the grass-roots hospital and the upward transfer channel is provided from the grass-roots hospital to the third-class hospital, the channel for patients is widened to seek medical treatment, unblock the upward transfer channel for severe patients, facilitate the lighter patients to seek medical treatment at home, and save medical costs.

2.3.3 Family doctor

The doctor end is provided to the community doctors to continuously record the vital signs of the contracted patients and form long-term and continuous health or treatment records; the patient end is provided to the patient to view his / her previous medical records and participate in health education.

2.3.4 Remote consultation

The third-class public hospitals in Linfen have successively entrusted several poorly managed first-class and second-class medical institutions. The hospital has sent several senior chief doctors to teach to improve the professional and technical level of each first-class and second-class medical institution. There are practical difficulties in on-site teaching with the increase in the number of entrusted medical institutions. The Remote Consultation System can be used for remote teaching, providing remote consultation for patients, saving time, cost and disease burden.

2.4. Construction characteristics of the healthcare big data intelligent platform

Linfen Regional Healthcare Big Data Intelligent Platform has the following characteristics in the design process:

The first is the distributed storage of tens of millions of data. The ten million level data distributed storage technology is a new database cluster based on MPP architecture. It focuses on industry big data and uses shared nothing architecture. It supports analytical applications through column storage, coarse-grained index and other big data processing technologies, combined with the MPP architecture's efficient distributed computing mode. The operating environment is mostly low-cost PC server, with the characteristics of high performance and high scalability.

Meanwhile, based on the technical expansion and encapsulation of Hadoop, relevant big data technologies are derived around Hadoop to deal with the data and scenarios that are difficult to handle in traditional relational databases, such as the storage and calculation of unstructured data, and make full use of the advantages of Hadoop open source. There will be more and more application scenarios with the continuous progress of related technologies. At present, the most common application scenario is to store and analyze Internet big data by expanding and encapsulating Hadoop. Hadoop Platform is better at unstructured and semi-structured data processing, complex ETL processes, complex data mining and calculation models.

At the level of the concurrent push system of millisecond data queue, it is divided into three parts: data layer, business layer, and application layer. The data layer is composed of multiple servers; the service layer sets one for each service (multiple load balancing or one server can be used, based on the traffic volume); each client of the application layer is linked to the business layer. The data layer holds global data; the business layer is responsible for processing business flow information and is the bridge between the application layer and the data layer; the application layer is directly connected to the business layer and does not directly communicate with the data layer. It mainly completes data presentation and friendly interface.

Intelligent data source sorting, medical database query, statistics and other functions are insufficient, and information cannot be used well. Data mining, also known as knowledge discovery in databases, mainly including three steps: data preparation, rule search and rule representation. Common functions include association analysis, clustering analysis, classification analysis, anomaly analysis, special group analysis and evolution analysis.

The response time of relational data is generally within 10ms or even shorter while the intelligent medical data cache is properly configured. When there are hundreds of millions of data and 1wtps, the response time should be within 10ms, which is almost impossible for relational data.

Local cache is the most used cache mode, with high speed and low cost. The use of the local cache is more limited for the cluster load balancing structure of the Web System, because there is no simple and effective way to update the local cache when the database data changes; However, if the local cache information is synchronized between different servers, the cost and performance may be difficult to accept due to the low timeliness and high access volume of the cache.

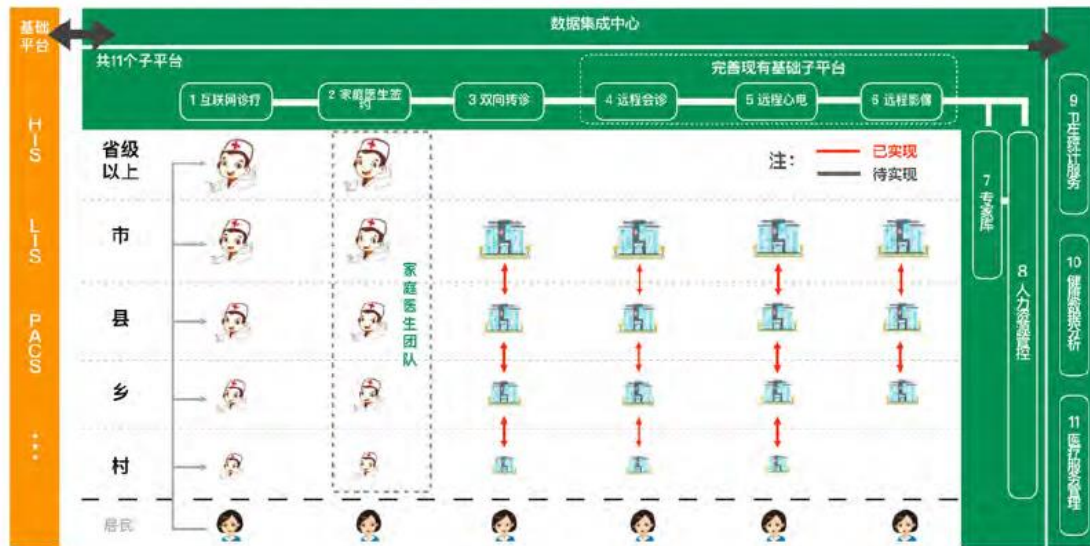


Figure 4. Regional healthcare big data intelligent platform (city, county, township and Village) four level linkage

The purpose of distributed cache is to provide higher TPS and scalability than RDB. Excellent distributed cache Systems include Memcached, Redis, and TAIR, etc. There is a big gap in the read and write performance compared with relational database and cache storage,. Therefore, try to transfer the read and write pressure from the database to the cache to protect the relational database when the technology and business are acceptable.

The database cache mainly refers to the query cache of the database, and the specific implementation details of each database are different. The basic principle is to use the hash value of the query statement as the key to cache the result set, which can improve the query efficiency of the database.

3. Risks and countermeasures of healthcare big data intelligent platform

(i) Data Risk and Countermeasures

There is a risk of massive medical data storage and disclosure. The risk of data loss shall be dealt with by establishing a remote loading mechanism, and the data to be encrypted shall be MD5 encrypted during the application process.

(ii) Information Security Risks and Countermeasures

There is a risk that the patients' medical records and other personal privacy data may be disclosed. The data between the software is transmitted through the encrypted channel to prevent the third party from reading. The possible network attacks by hackers are protected by professional equipment and realized by load balancing mode.

(iii) Structural Risks and Countermeasures

The original design architecture can not meet the process of massive data access. Meanwhile, the load balance is used to separate read and write. Meanwhile, a single host works in a multi-threaded mode.

4. Implementation and application effect of healthcare big data intelligent platform

As the leading hospital in Linfen, Linfen People's Hospital covers an area of 371 mu, with 1,800 beds and 2,432 employees. Adhering to the service tenet of " People's Health Oriented", and taking the provincial first square excellent hospital as the new goal, the hospital escorts the lives and health of 4.4 million people in Linfen. In 2018, the annual outpatient volume of the hospital reached more than 860,000, and the inpatient volume reached more than 70,000.

The hospital covers 17 counties and cities in the city, and the number of cooperative medical consortium units has reached 205, which is conducive to promoting the construction of four level regional Big Data Platformplatforms of city, county, township and village in the region. Meanwhile, it is exploring the construction of Shanxi Medical and Health Big Data Platform based on Linfen CityBig Data Platform.

With the hospital as the main body, we independently developed the Regional Healthcare Big Data Platform based on the combination of "Production" and "Research". Through the construction of Linfen Regional Big Data Platform, the medical data in the region will be cleaned, stored and mined, and unified integration and management will be implemented to realize the exchange of electronic medical records and inspection data among the four level medical institutions in the region.

For management organizations, the Big Data Platform can serve as the basis for strategic decision-making, provide services for the supervision and management of medical care, medicine and medical insurance, provide the basis for the rational flow and allocation of medical and health resources, provide support for hierarchical diagnosis and treatment, and assist decision makers to formulate targeted measures to solve problems.

For medical institutions, the Big Data Platform can assist clinical decision-making and scientific research, build channels for medical and health institutions at all levels to share diagnosis and treatment information, make diagnosis and treatment, consultation and referral more efficient, and realize "patient-centered" information interconnection.

For the public, the Big Data Platform can carry out health management for the whole life cycle of individuals and ensure the consistency, continuity and availability of health information.

5. Conclusion

By building the Big Data Platform to implement unified management of regional data, it is possible to integrate, optimize and utilize existing information system resources to the maximum extent, open up channels for information transmission of various medical and health institutions, realize collaborative linkage and data exchange of information Systems, improve resource utilization, and realize cross regional and cross institutional information sharing, business collaboration and overall linkage service capabilities, provide strong information support for the development and decision-making of medical and health services in the province, and further promote the informatization development of medical and health undertakings.