

The Construction of City's Smart-healthcare Data System Based on Habase

*Zhen Niu**

Tianjin university of finance and economics, China

*Corresponding author: Zhen Niu, ziv.niu@hotmail.com

Abstract

To deal with information incompleteness or error and supervising difficulties caused by hardships in information synergy and sharing among current medical institutions, this article proposes a solution to construct a city's smart-healthcare data system. Through the design of synchronizing each medical institution's heterogeneous database to data center, the aforementioned problems are solved.

By analyzing the needs of city's medical data center, this article takes NoSQL database and hadoop as the basis of system construction. Meanwhile, through comparative analysis of several existing NoSQL databases, Habase is found suitable for storing and processing massive data. This article proposes a new database modeling thought, O-R (Object Relationship) modeling, by reference to E-R relational database modeling. This O-R modeling thought compensates for non-relational databases' deficiencies in front of relational databases. Utilizing this new modeling thought and the solution to synchronize heterogeneous databases identically to data center, the article completes the construction of the city's smart-healthcare data system.

Keywords: *habase, city's smart-healthcare data center, NoSQL, front-end web, O-R model*

1 Introduction

City's smart-healthcare means using medical treatment informationization and big data analysis to solve difficulties of citizens in seeking medical treatment and to avoid treatment delay or error caused by incomplete medical information. However, hard to timely update and exchange information, medical entities have turned into information islands, and thus the informationization of medical treatment fails to play its due role to realize the intelligence of city's smart healthcare. To solve this core issue, this article proposes a new smart-healthcare system construction based on Habase to exchange medical information and eliminate information islands.

Information islands refers to problems that local medical databases are hard to integrate and communicate. The most common causes of the problem are: 1. Different medical institutions or corporations conduct independent researches for their own needs, and their self-defined data formats lack identical standards; 2. Different medical institutions that define database standards produces many heterogeneous data and data models, resulting in the lack of data

exchange standards for the whole city's medical industry, and thus the incapability of information sharing among medical entities, which leads to the emergence of information islands. For the existence of this information islands, data are collected all for self-use. The absence of identical standards makes data repeatedly collected and massively redundant in database, causing the waste of resources.

Information islands must be smoothed out, in order to build a city's smart-healthcare data center that can be administered in a unified manner, convenient for information lookup and integration, as well as resource-conserving. To deal with this problem, this article proposes a solution to construct a city's smart-healthcare data center. This data center is constructed based on Habase through synchronizing heterogeneous databases of different medical institutions. Combining the O-R modeling thought, the design of this data center utilizes Habase's ability to autonomously define and modify data formats. After that, in the aspect of data format, heterogeneous data generated by heterogeneous databases is encapsulated as the standardized common data format of JSON to avoid the troubles caused by the difference of data format. Adopting the general transfer protocol of HTTP, under the thought of SOA architecture design, this design proposes to use Web Service to carry out synchronous data transmission; In obtaining incremental data for heterogeneous databases, this design takes an integrating method of time stamps, triggers and diary methods. Lastly, this article proposes the design of a general front-end web to read all the heterogeneous databases.

The structure chart is shown as *Fig. 1*.

2 Building Data System on the Basis of Habase

This section introduces related technologies to this article, provides basis for city's smart-healthcare data center construction by comparative analysis of these existing technologies, and creates foundation for data center's construction, data storage and quick index. Through the analysis in this part, this article concludes to base city's smart-healthcare data center construction on Habase.

2.1 NoSQL

A city's smart-healthcare data system not only has changeful data formats and massive amount of data, but also extends its scale at every moment. Under the condition of high flexibility, high performance, weak consistency and simple model, given the key, NoSQL database can perform better than other types of databases in looking up comparatively complex data value. Reviewing all the factors above, NoSQL is the most suitable database for the construction of city's smart-healthcare data system.

Synthesizing conclusions of section 2.1, the article concludes that city's smart-healthcare data system should be NoSQL database based on Hadoop, or in other words, that this database performs best (than other databases) when based on Habase. Therefore, after a series of consideration and comparison, this article selects Habase to construct medical data system.

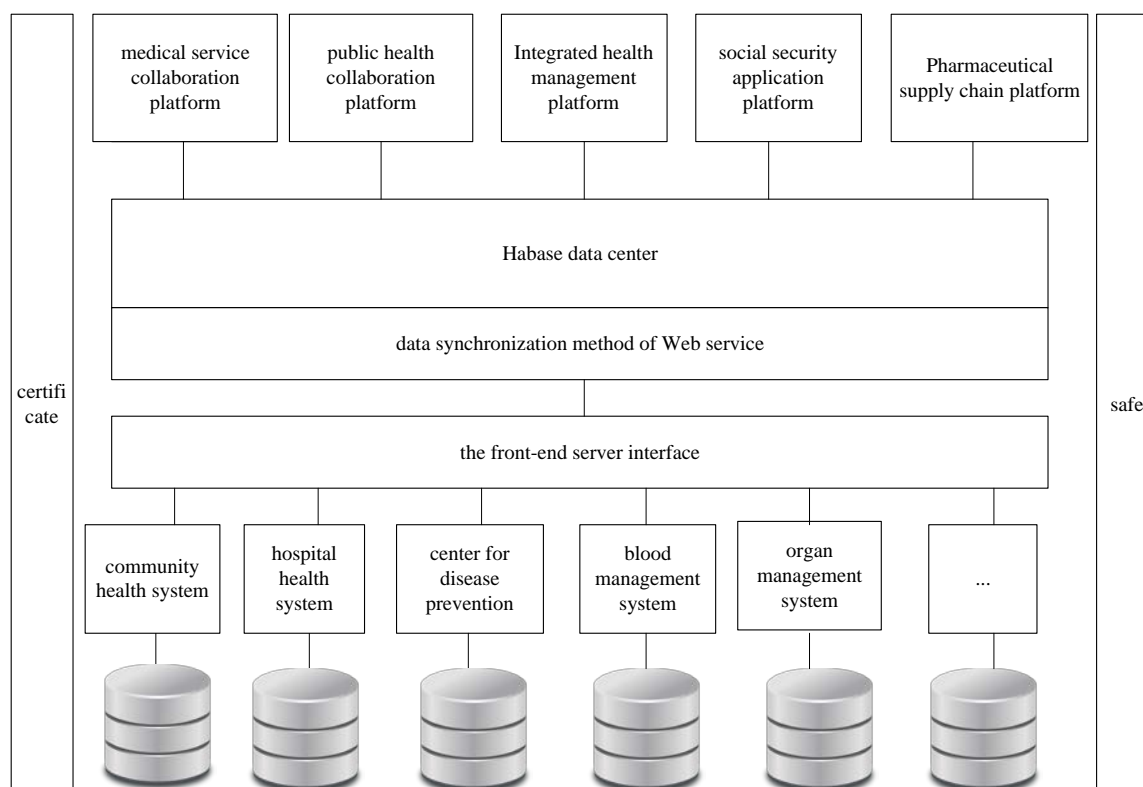


Fig. 1 – Software architecture diagram

2.2 Habase

Hbase is a hadoop-based new-type NoSQL database that supports distributed storage. It shows strong ability in massive data storage, performing efficiency and data analysis. This section will introduce the importance of hadoop and NoSQL to this database. Meanwhile, to meet the requirements of hadoop and NoSQL, this article suggests city's smart-healthcare center be based on Habase.²

3 The Construction of Medical Data System

As introduced above, this data system aims to eliminate information islands and provide a basis for cooperation and communication among medical institutions, or in another word, to put related data resource on somewhere every entity needing they can reach for sharing and analysis.

3.1 Constructing Database through O-R Model

Through analysis and compare in Section 2, this article demonstrates the rationality of the construction of a medical data system based on Habase. Despite the dynamic capability of Habase, considering the information sharing of a stable high-performance data center, a relatively stable table structure should be established for the system to facilitate information communication among medical institutions. This article uses health data elements published by the ministry of health to build this model. The logic of the model is displayed according to paradigm theory of relational database as the following chart:

Table 1 – Comparison of database types¹

Database type	Key-value	Document	Graphic	Column-based Storage
Advantage	1. Simple structure; 2.Easy for distributed deployment.	1.Not strict requirement on data structure. 2.Better lookup than that of Key-value's	Easy to extend to several servers	1.Able to cope with distributed file system storage of massive data. 2.Able to search the distributed database quickly which has the strong scalability.
Disadvantage	Data are hard to update when in unstructured storage.	Having no identical grammar for lookup.	Results can be obtained only after the analysis of the whole graph and computation, so this type is inapplicable to distributed cluster database.	
Example	Tokyo Cabinet	MangoDB/CouchDB	Neo4J/ Info Grid	Habase/ Riak
Application	Small scale database with quickly changing data, such as stock price storage and real-time communication messages.	Visual databases and databases capable of dynamic query.	Databases focusing on topology such as social network.	Databases with extremely massive data.

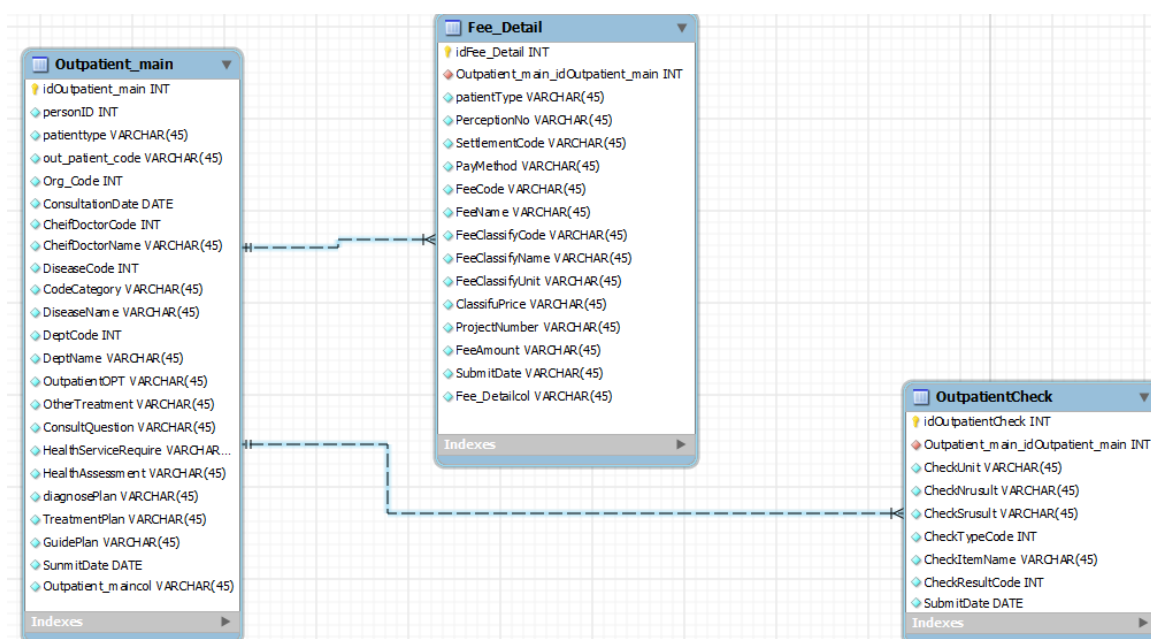


Fig. 2 – Medical data system based on E-R model

Unlike traditional databases, Habase database has no foreign keys between tables to establish external relations. However, it cannot be denied that the establishment of external relations makes E-R model more reliable and convenient than NoSQL database in data organization and sharing. Therefore, this article will introduce a new type of relationship model, O-R model(Object-relationship model) .

Referring to the results-oriented design thought of E-R model, this new type of relationship model makes problem domains transaction centers and proposes object-oriented method to build Habase data system model. It can be seen from the structure of Htable that Column Families of Htable are relatively stable. The number of elements in a Column Family can grow freely, but concrete Column Families cannot be freely modified. Thus, a Column Family can be seen as an object and each child column in the family can be taken as unfixed and scalable properties of this object. A Htable is generated by the combination of all the Column Families and their RowKeys. Then this Htable is a composite object named after Rowkeys. In some condition, there are behavioral relations between objects and these relations generates a series of data, namely relationship properties. Based on these relationship properties, multiple Htables can be created for the Habase system. Moreover, for the continuation of these relations, these Htables also have relationships among them. Therefore, referring to E-R modeling thought of relational databases, this article establishes relationships for these Htables. However, considering that Habase cannot establish relationships through foreign keys, despite the logic existence of these relationships, further analysis are needed for the system to guarantee that generated physical model can preserve these logical relationships.

Rowkey is the primary key of Htable. Data are stored and Habase is indexed according to this primary key. Rowkey is a byte array whose length can be extended to 64k. According to aforementioned, in returned relational data, physical storage relationships, no matter one-to

-one, one-to-many or many-to-many, are all additions of related fields to related tables. In Habase, this article changes this addition of related fields to the joins of Rowkeys so as to physically create association relationships. The consistency of these relationships needs to be controlled by programs, because Habase is unable to verify or constrain relationships.

This article call this model based on relationships among composite objects O-R model. Similar to E-R model, O-R model can be displayed in charts comprised of linking lines and block diagrams. Shown as dotted line, relationships between composite objects can be categorized into one-to-one, one-to-many and many-to-many as well.

To deal with the difficulty of building data dictionary for the lack of foreign keys, Habase can store this data directly, guaranteeing lookup efficiency and the storage stability at the price of data redundancy.

In the example of city's smart-healthcare data system, O-R model helps establish between individuals and institutions a behavioral relationship, which demonstrates the composition relations between objects.

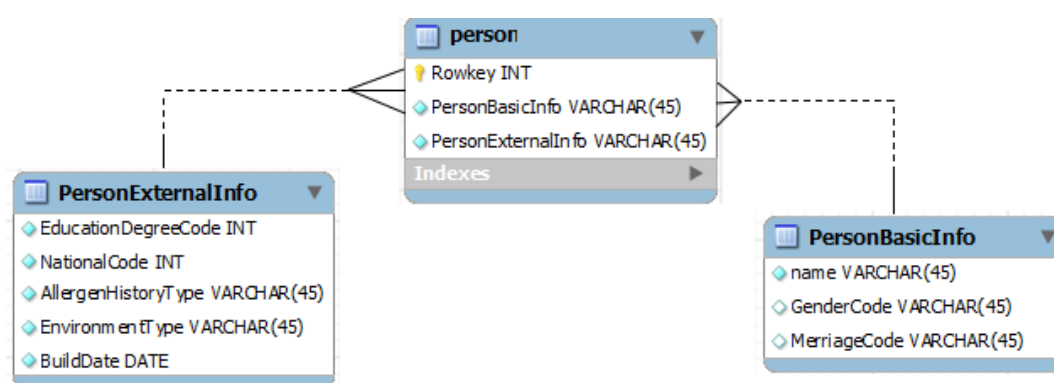


Fig. 3 – Individual information data model based on O-R model

3.2 Synchronizing Heterogeneous Databases to Smart-healthcare System

This article proposes adding front-end web into city's smart-healthcare data system, in order to connect VPN with subnet databases and tackle with two important problems in the integration of medical databases.

First problem is that networks used by those entities within the region are different since health services and relevant departments are distributed in different locations. The connection performance and speed of these networks cannot be in good condition, therefore VPN should be applied as the connection mode to enhance connection performance. Moreover, the application of VPN can realize network authentication and communication encryption mechanism and thus further guarantee the security of data synchronization. Secondly, without connections to internet, some medical institution's management system can be used in intranet only. Even when most medical institutions' management system can be connected to internet, they would not expose their databases to the public network for security concerns. Therefore, to connect these heterogeneous databases, just using VPN is not enough. There should be a

server that can both communicate with VPN and internal database, called the front-end web. Likewise, in the design of HBase data center, to guarantee its security it cannot be directly connected to the public network. The web server should serve as the middle tier along with other necessary safeguard measures such as the firewall. The network topology is displayed as below:

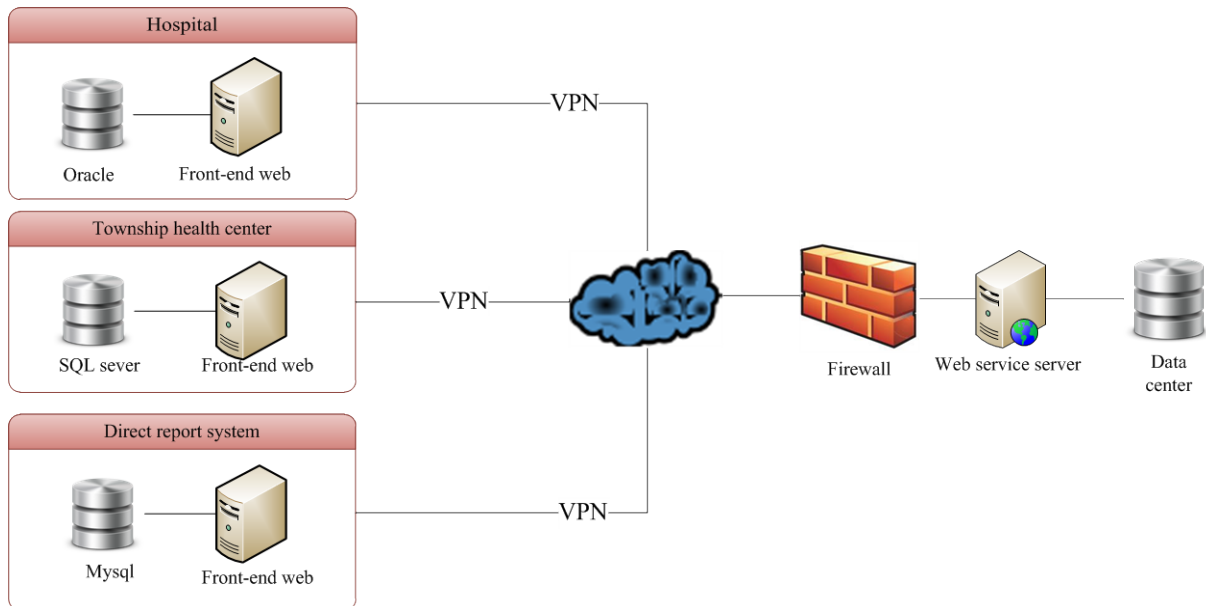


Fig. 4 – The network topology from heterogeneous databases to the data center

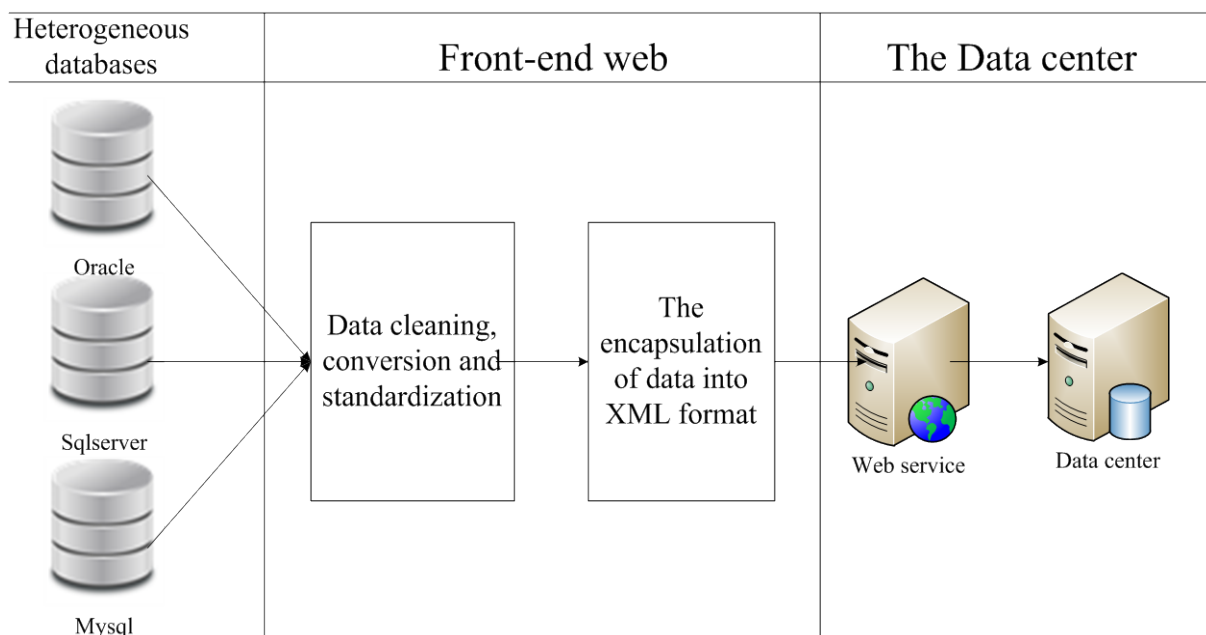


Fig. 5 – The procedure of data synchronization

As can be seen from Fig. 4, one function of the front-end web of this heterogeneous database is to connect heterogeneous data with HBase data center. So the front-end web should complete these tasks: Front-end web reads incremental data from heterogeneous database and transmits these data to the Web Service server after relevant processing; Web Service write these data into HBase data center when getting them and then front-end web does the logging.

Thus, to synchronize heterogeneous database to HBase mainly is to implement the accepting and writing operations on synchronized data of both front-end webs in heterogeneous databases and Web Service in the data center.³ The procedure of data synchronization is illustrated in *Fig. 5*.

Front-end web can tackle with above two problems smoothly and meanwhile protect each medical institution's database from attacks caused by exposure.

4 Conclusion

This article discloses the information island problem faced by current medical institutions and then proposes solutions to eliminate information islands and construct a city's smart-healthcare system. In the construction of this data system, by analyzing its features that it has a massive amount of data and a high requirement for lookup speed, this article proposes the design idea of basing its construction on Hbase. Meanwhile, this article makes two improvements on current Hbase database in light of its difficulties in satisfying the need of city's medical information system to connect databases of all medical institutions. On one hand, based on E-R model, this article proposes objects-oriented O-R modeling thought. On the other, this article proposes the application of the front-end web in the solution, not only to synchronize data in each medical institution's heterogeneous database transparently and identically to data system, but also to maintain the security of these databases.

Nevertheless, there are still certain deficiencies in the proposed solution. This deficiency, the lack of storage for special texts, graphs and videos, is an important direction for future improvement on this system. Only with efficient storage and communication of these special information can the completeness of city's smart-healthcare system be actually realized.

References

1. S. Tiwari. Professional NoSQL. Wrox, 2011
2. Z.Y. ZHANG, B. LIU. Research the Synchronization Mechanism for Distributed Heterogeneous Database[C]. IEEE International Conference on Information Management and Engineering, 2011
3. F. Curbera, F. Leymann, T. Storey, et al. Web services platform architecture: SOAP, WSDL, WS-policy, WS-addressing, WS-BPEL, WS-reliable messaging and more. Prentice Hall PTR, 2005