SOLUTION BRIEF
Intel® Xeon® Processor E5 Family
Intel® Distribution for Apache Hadoop* Software
Healthcare
Big Data and Cloud Computing

Big data builds a healthy city

Jinzhou City boosts healthcare initiatives with an optimized regional data center using UFIDA Medical healthcare solutions based on Intel® Xeon® processor E5-2600 and big data solutions

UFIDA 用友

With a population of about 3.1 million, Jinzhou is a major port city and regional center in Liaoning, China. In a bid to become a healthy city, the city government established a municipal healthcare data center to house the digital versions of the residents' health cards. This will be used to access various healthcare services when they become available online. The data center will also hold electronic medical records, public health information and other related data that would be integrated and centrally managed. Twenty years worth of data are expected to be accumulated, and the volume of digital content to be created could hit petabyte levels. As Jinzhou moves toward becoming a healthier city, personal health data is expected to greatly increase as residents become more health conscious while more healthcare services are rolled out.

CHALLENGES

- **Meet high-volume data needs.** Increase data center storage to accommodate growing health records, electronic medical files, and public health management data.
- **Accommodate new types of data.** Address the need to store complex and volatile data formats and other types of data that evolve continually, without having problems in data exchange and interoperability.
- **Address future data challenges.** Improve efficiency in data management to meet future challenges such as mass storage, backup, scalability, speed in data access and data exchange.

SOLUTIONS

- **Build a data center using UFIDA Medical healthcare solutions running on Intel® Xeon® processor E5 family.** Take advantage of UFIDA Medical's expertise in building healthcare solutions at a provincial and municipal level, using servers based on Intel Xeon processor E5-2600 to allow an increase in capacity and performance to accommodate high-volume data.
- **Utilize Intel® Distribution for Apache Hadoop* software.** Employ Intel big data solutions for storing new types of data and new methods of storing such data while ensuring that the data center is able to perform quick searches on all documents.
- **Employ horizontal scaling.** Using big data architecture, horizontal scaling can be easily implemented to meet future needs and keep up with the growing amount of data.

BUSINESS VALUE

- **Enabled rapid data search.** Allowed doctors to quickly search for patient background, medical history, past treatment plans, and prescription information for better diagnosis to avoid medication duplication and adverse drug reactions.
- **Supported high levels of concurrency while retrieving large amounts of data.** Enhanced individual access to health records (either as residents or as patients) and helped establish effective methods for health management and enabled a shift in philosophy towards prevention.
- **Improved flexible data analysis, modeling and data mining.** Data can be widely used in various areas such as disease control, regulation of clinical practices, medical quality control, and so on, as the data center supports flexibility in data analysis, modeling and data mining, as well as providing near-real-time analytics for regional healthcare management.

TECHNOLOGY RESULTS

- **Enhanced storage for massive data.** Implemented a storage solution to handle massive amounts of data, targeted at managing and storing more than 100 million documents.
- **Achieved dynamic expansion of data format.** Allowed easy adjustment and expansion of data formats using Hbase* column-based storage.
- **Realized rapid search and retrieval from large databases.** The data center supports a high number of concurrent data retrieval requests and real-time retrieval of health records by both doctors and patients.
- **Ensured centralized analysis of data.** Built an open framework that improved near-real-time statistical analysis of data.

Jinzhou regional healthcare data center, based on the Intel® Xeon® processor E5 family running on Intel® Distribution for Apache Hadoop software, is an optimized platform for big data analytics and intelligence for healthcare cloud computing systems

"Intel has made advances in big data solutions, and their experience has been of great help to us. Through this project, we have been able to benefit from the use of big data in ways we have never dreamed of, in terms of value for money and timely response to queries."

Xu Chunhua
Chief Scientist,
Ufida Smart Healthcare Institute
Vice President,
Ufida Health IT System Co. Ltd.

Big data to boost healthcare

As part of the nationwide move towards the use of information technology in its twelfth Five-Year Plan, the Ministry of Health in 2010 devised a strategy codenamed 3521-2. The 3 refers to the three levels of information on public health: at a national level, provincial level, and city level. The 5 refers to the five areas that information technology would support: medical care, public health, availability of treatment and access to medicines, and the integrated management of all of these areas. The 2 stands for the two basic repositories of data: one for health records and the other for medical files. The 1 is the single specialized healthcare information network to be established. The last 2 represents the two areas of information security and standard specification.
As part of the larger initiative to use information technology at a regional level, one of the core programs of this strategy was to establish a municipal-level information system.

To establish its own regional healthcare information system, the Jinzhou City Government knew the importance of employing big data technologies to meet this goal. Big data technologies are integral in the management of patients' healthcare records and related services. For a medical practitioner or public health administrator, access to the full medical history of a patient is a must, especially in the treatment of chronic diseases. Having information on the disease's progression over time and past treatment plans help in the diagnosis and creation of current treatment plans. Data on past allergies and adverse drug reactions also help to eliminate diagnostic mistakes and accidents.

The ability to amass vast amounts of medical and health-related data that can be mined and analyzed provides a more scientific basis for management decisions and regulatory implementations.

Traditionally, clinical studies would only be based on limited sample size. Having large numbers of detailed health records can reduce the workload in clinical studies, improve the quality and quantity of clinical studies, and make processing and managing data more efficient.

Apart from easy access to detailed health records, big data technologies can also help with clinical diagnostics and research. Access to a large number of health records is also essential in public health decision-making as the government works to improve public health management.

The demands of a regional healthcare data center

To support its goal of becoming a healthy city in line with the nationwide healthcare strategy, Jinzhou City Government sought the expertise of Yonyou Medical and Healthcare Information Systems Limited (UFIDA Medical) to be the main solutions architect for its regional healthcare data center. UFIDA Medical has extensive experience in providing healthcare solutions at a provincial and municipal level.

Initially, UFIDA Medical tried to use traditional relational databases to set up the data center, but it soon encountered problems inherent to the model:

- **High-volume data.** With a population of over three million, Jinzhou is considered a mid-sized city. The regional center houses three municipal-level databases for health records, electronic medical files, and public health management. As city population continues to rise, the data center is expected to reach petabyte class within 20 years. Using traditional relational databases to house such massive data is not practical, since there are limitations when they are deployed for big data storage. Performance issues can also be expected when single table sizes reach 500GB.
- **Complex and volatile data formats.** The regional healthcare database houses many types of complex data that need to be stored, both structured and unstructured. Using a traditional

relational database poses problems such as accessing or storing PACS imaging, B ultrasound, pathological analysis and other services that produce unstructured data of varying sizes, and evolving formats such as HL7 Clinical Document Architecture* (CDA*) compliant Extensible Markup Language* (XML*) format, which stores electronic clinical records.

- **Data exchange and interoperability.** New challenges in these areas may appear due to complexities inherent to the healthcare industry, so it would be hard to establish a data standard.
- **Future data challenges.** The data center should be able to face inevitable challenges in the future, which include mass storage, backup, scalability, the speed at which a specific data can be accessed, and efficiency in data exchange, among others.

Building a data center on Apache Hadoop

To overcome these challenges, UFIDA worked with Intel to devise a logical framework that would help analyze and explore the requirements in-depth and provide clear direction for constructing a big data analysis solution. In building the regional healthcare data center using big data solutions from Intel, UFIDA set the following targets:

- Based on the projected volume of data, the data center needs to be able to perform quick searches on all documents.
- As medical systems continue to be upgraded and healthcare standards improve, the data center needs to accommodate solutions for storing new types of data and new methods of storing such data.
- The data center should be able to scale horizontally without affecting the application layers. The basic operational layers need to be segregated from the higher layers. Adding or upgrading hardware should be transparent to the software layers to allow for increase in capacity and performance capabilities.

The region's healthcare providers and agencies access services on the information platform and supply data using the HL7 CDA* format. This document format is based on XML and designed specifically for the healthcare industry. The collected CDA documents include personal health records, summary of outpatient visits, prescription records, diabetes case studies and diabetes follow-ups. The uploaded documents are parsed using XML, and metadata is extracted from the records. The extracted metadata and parsed XML documents are stored in

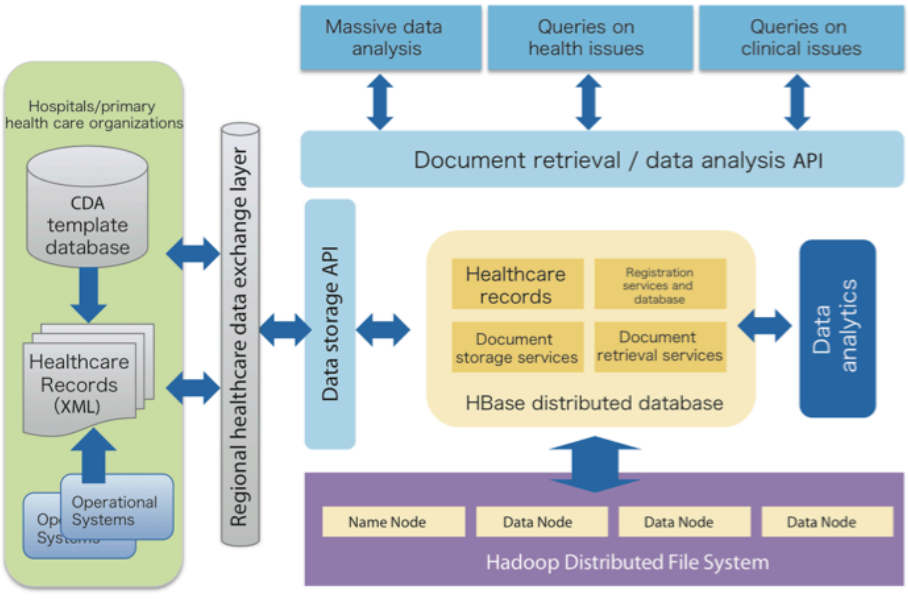

Figure 1. Data processing flow in the Intel Distribution for Apache Hadoop software-based regional health information platform

- Health records are fundamental to each person and should cover the entire lifecycle of a city's residents. Therefore, the planning stages of building a database for these records are very important.
- Utilizing big data solutions such as Intel Distribution for Apache Hadoop software ensures that solutions for big data management, search and retrieval, and analysis can be built successfully.
- Compared to traditional databases and solutions based on smaller host computers, big data solutions from Intel provide a more flexible environment at better value.
- A data center built on big data infrastructure creates a highly scalable, integrated graphical interface that provides a convenient method for node management.

the healthcare repository, which is in the form of an Hbase* database. Designed using big data architecture, the regional healthcare information platform processes data and provides document retrieval and analysis services. Through this document access and data analytics interface, individual analysis is provided to residents or patients, while other types of analyses are provided for doctors and healthcare administrators. The system is able to manage and analyze the massive amount of data it holds in different ways to meet needs of different users.

"We have built the regional healthcare project in Jinzhou based on a one card, one file concept, using the residents' health cards and health records. Health records are fundamental to each person, so it needs to be built up over the long term, and should cover the entire lifecycle of the city's residents. Therefore, the initial stages of planning of the data center were important," explains Xu Chunhua, chief scientist of Ufida Smart Healthcare Institute and vice president of Ufida Health IT System Co. Ltd.

After repeated load testing for each service, high-volume data tests, and other performance tests, Intel and UFIDA optimized the systems using a variety of techniques. The Jinzhou regional healthcare data center operates servers based on the Intel Xeon processor E5-2600, using Intel Distribution for Apache Hadoop software. This infrastructure supports a comprehensive suite of big data solutions for regional healthcare and is designed to handle a large number of records (in excess of 100 million) and perform real-time analyses. The tests confirmed that the servers based on the Intel Xeon processor E5 family running on Hadoop delivered a platform for big data analytics and intelligence for healthcare cloud computing systems. The infrastructure ensured that solutions for big data management, search and retrieval, and analysis can be built successfully.

Big-data optimized regional healthcare data center

Intel and UFIDA Medical were able to build software solutions for big data that allowed Jinzhou's regional healthcare data center to create storage solutions for massive amounts of data. It manages and stores more than 100 million documents. With Hbase, which uses column-based storage for easy adjustment and expansion of data formats, the data format was dynamically expanded. Since big data supports a high number of concurrent data retrieval requests, rapid search and retrieval from large databases was enhanced, particularly supporting real-time retrieval of health records by both doctors and patients. The big-data-optimized data center also promoted centralized analysis of data, thanks to the open framework for statistical analysis and near-real-time statistical analysis. More importantly, the big data, architecture-based data center allowed easy expansion with the horizontal scaling that could be easily implemented to meet future needs and keep up with the growing amount of health data.

Adopting big data solutions also brought Jinzhou economic benefits. Compared to traditional databases and solutions based on smaller host computers, the big data solution from Intel provided a more flexible environment at better value, reducing the cost of operating host computers. The ability to scale horizontally also significantly reduced management costs. Moreover, the highly scalable, integrated graphical interface provided a convenient method for node management. Since it is optimized, the data center can take advantage of available hardware solutions, thereby reducing upgrading costs.

But more than the technological and economic advantages, Jinzhou doctors and patients benefit immensely from the big data solution optimized for processing large amounts of data. The data center enables doctors to rapidly search patient background, medical history, past treatment plans, and prescription information, allowing doctors to better diagnose their patients, thereby avoiding medication duplication and adverse drug reactions. The data center also supports high levels of concurrency while retrieving large amounts of data, which makes individual access to public records easier for residents or patients. Combined with better health education for residents, the data center helps establish effective health management, enabling a shift in philosophy towards prevention. It also supports flexible data analysis, modeling and data mining. Furthermore, it provides near-real-time analytics for regional healthcare management, allowing the data to be widely used in areas such as disease control, regulation of clinical practices, and medical quality control, among others.

"We have been able to benefit from the usage of big data in ways we have never dreamed of, particularly in terms of value for money and timeliness in the response to queries. The experience and expertise of UFIDA Medical and Intel, which has made advances in big data solutions, were of great help to us," says Xu.

Find a solution that's right for your organization. Contact your Intel representative, visit Intel's Business Success Stories for IT Managers **(www.intel.com/itcasestudies)** or explore the Intel.com IT Center **(www.intel.com/itcenter)**.

SOLUTION PROVIDED BY:

This document and the information given are for the convenience of Intel's customer base and are provided "AS IS" WITH NO WARRANTIES WHATSOEVER, EXPRESS OR IMPLIED, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, AND NONINFRINGEMENT OF INTELLECTUAL PROPERTY RIGHTS. Receipt or possession of this document does not grant any license to any of the intellectual property described, displayed, or contained herein. Intel® products are not intended for use in medical, lifesaving, life-sustaining, critical control, or safety systems, or in nuclear facility applications.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

© 2013, Intel Corporation. All rights reserved. Intel, the Intel logo, Intel Xeon, and Intel Xeon Inside are trademarks of Intel Corporation in the U.S. and other countries.

*Other names and brands may be claimed as the property of others.

1212/JAY/PMG/XX/PDF

328895-001US