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

Big data constructs intelligent healthcare

Intel works with Wonders Group to help Shanghai Health Bureau Information Center build a reliable and efficient health information search and intelligent alert system with Intel® technology

"Shanghai's health system has continued to lead the country in its use of information technology. We are dedicated to developing a point-line-plane integrated information platform for the community, region and city—namely, a three-level healthcare organization which will gradually support all public medical institutions over time. We will also use this system to cover all other medical institutions as part of this strategy. In achieving this goal, Intel and Wonders Group have given us a great deal of assistance. Not only have they introduced to us many advanced methods of big data analytics, they have also helped us deploy a reliable solution based on data center software products such as Intel® Distribution for Apache Hadoop* software."

Xie Wei
Deputy Director
Shanghai Health Bureau Information Center

The Shanghai Health Bureau Information Center has been engaged in the study and development of middle- and long-term plans for the Shanghai health information system since its founding in 1998. The Center mainly creates and regulates application management rules, technical specifications, and standards of information classification for constructing the health information system. The other basic functions include developing the health system's information network for effective networking for Medicare information, disease prevention, technologies, intelligence and administrative decisions; providing and maintaining public service information; organizing, coordinating, and guiding subordinate bodies to build up their information resources; carrying out training of modern management and computer management skills for the administrative officials and professionals in Shanghai's health sector; and providing information and services for the Health Department, the municipal government, and the bureaus under the government.

CHALLENGES

- **Limitation of traditional database system.** As the Shanghai health information system grows, the traditional database system cannot handle the massive amount of data, huge concurrent access, and great scalability. A new data processing model is needed.

SOLUTIONS

- **Hadoop-based system for data processing.** With the assistance of Wonders Group, the Shanghai Health Authority built a big data cluster with Intel® Distribution for Apache Hadoop* software and Intel® Xeon® processor E5-2640-based servers.

IMPACTS

- **Improved healthcare for the public.** The big data solution helps doctors to get quick access to health information of their patients, and drops the rate of medical re-exams from 80 to 30 percent.

- **Highly efficient health information system.** Shanghai Health Bureau System Information Center now runs a cloud-based, highly efficient and concurrent health information search and intelligent system, and delivers a retrieval service model that covers all the public medical institutions.

"For over 20 years, Shanghai has been developing a health information platform to streamline the access and delivery of healthcare information in the city," explained Xie Wei, deputy director of Shanghai Health Bureau Information Center. "Shanghai has a very clear roadmap for its health information system. We retrieve health information from three levels: from points such as the hospital information system (HIS) and the radiology information system (RIS), to the lines such as the tumor report system or the diabetic management information system that link all the institutions, and lastly, to a regional healthcare information system that covers all the regions of the city."

In such a developing process, managing and processing data, rather than collecting and obtaining data, has become a great challenge for Xie Wei and his team. As of today, 16 million data records swarm the backend of the database every day, equivalent to 4.3 billion data records in a year. The previous database architecture was obviously unable to meet the requirements of the information system with the high performance and efficiency needed to deal with such a

Developing a health information search and intelligent alert system that delivers highly efficient service to the public medical institutions in Shanghai

massive amount of data. Constructing a health information search and intelligent alert system demands new technology.

High scalability addresses ever-growing amounts of medical data

Doctors in Shanghai often have to serve several hundred patients every day. The time allocated for each patient may be only a few minutes. A health information search and intelligent alert system that can quickly retrieve the patient's medical history will help doctors work more efficiently and spend more time helping patients.

"Shanghai has nearly 120,000 doctor workstations sending data to the center's database every day, and this data volume is bound to increase due to urbanization," Xie said. "In fact, the pressure caused by the massive amount of data the Shanghai Health Bureau Information Center is facing is also a problem for every city's information center. It demands a scalable data platform that traditional databases cannot provide."

To cope with increasing data pressure, the Shanghai Health Bureau Information Center adopted a platform based on the Intel Xeon processor E5-2640 and Intel Distribution for Apache Hadoop software to build a health information search and intelligent alert system. The distributed system provides highly concurrent access to the data from thousands of doctor workstations.

Robust platform to address the challenges of big data

Intel Distribution for Apache Hadoop software includes Apache Hadoop and other software components with enhancements from Intel. The Intel Distribution for Apache Hadoop software is optimized for Intel Xeon processors, storage, and networking hardware components to ensure that the platform delivers balanced performance for the widest range of use cases. Intel Distribution for Apache Hadoop software uses a distributed database (HBase*) and distributed computing framework (MapReduce*) to provide high-performance and high-complexity data processing.

The Hadoop Distributed File System* (HDFS*) distributes data blocks across the cluster. Expanding the system's capacity only requires adding servers to the cluster. The distribution algorithm of the files can transfer the data blocks to the new servers without manual intervention. And the file system automatically expands to use the additional capacity. The distributed database (HBase) stores records in a flexible columnar database that also scales by simply increasing the number of the servers," explained Wang Xiaodong, Intel engineer. The MapReduce framework can conveniently process different kinds of data that enter the medical system, including structured, semi-structured, and unstructured data.

The Shanghai Health Bureau Information Center has a data center with more than 300 servers. Of these, 50 servers based on the Intel Xeon processor E5-2640 comprise the Hadoop cluster today, processing 5,000 concurrent access requests in one second. This data serves the entire health information search and intelligent alert system in Shanghai. This greatly helps the doctors in all the medical institutions in Shanghai to get quick access in the health information system, and make the rate of medical re-exams has also dropped from 80 to 30 percent. And the system has become an important component for Shanghai's smart city initiative.

Looking ahead, Shanghai Health Bureau Information Center plans to rely even more on big data technologies to seek breakthroughs in data analysis, data mining, and statistical analysis to provide better healthcare services for all its residents.

Intel Distribution for Apache Hadoop software offers key advantages:

- Intel provides a comprehensive set of software and hardware components for big data solutions.
- Intel Distribution for Apache Hadoop software is a stable platform that has been tested and verified in the production environment for this deployment.
- Intel Distribution for Apache Hadoop software offers solutions to existing problems during actual use as well as enhancements and optimizations to the open-source Apache Hadoop, delivering performance optimization, security, and manageability required for enterprise deployments.
- Intel cluster management and installation tools have simplified the installation and configuration of Hadoop to configure the cluster automatically for optimum performance, making full use of the cluster's computing ability.
- Based on previous experience with cloud computing deployments, professional services in all phases – from project planning to implementation – are crucial to help the client deploy complex systems with high scalability and high performance. In coordination with Intel's hardware department, it provides a comprehensive software and hardware solution.

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**).

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.

All performance tests were performed and are being reported by Shanghai Health Bureau Information Center. Please contact Shanghai Health Bureau Information Center for more information on any performance test reported here.

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.

© 2014, 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.

0414/SHA/PMG/XX/PDF 330014-001EN