Driving Massive Structural Data

Intel® Solid-State Drives and Intel® Xeon® processors help Taobao get the high power, capacity, and reliability and low total cost of ownership (TCO) it needs for its OceanBase* database

Challenges

- **New database system.** As Taobao's e-commerce business grew, the mass of data seriously challenged its traditional relational database. It needed a new database system.
- **New hardware platform.** To enable its New OceanBase high-performance distributed database system to work efficiently, Taobao urgently needed a new hardware platform that could meets its fast-growing needs.

Solutions

- **Powerful new hardware platform.** Taobao deployed a new hardware platform including Intel Solid-State Drives and Intel Xeon processors.
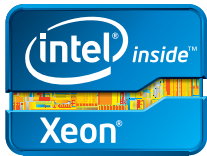
Impact

- **Power for the future.** The new Intel hardware platform delivers the high performance, capacity, reliability and low TCO Taobao needs for its OceanBase database system. It enables Taobao to process its massive data load more efficiently. The new platform is the foundation for providing quality e-commerce services for customers.

As China's most popular online shopping retail platform, Taobao (taobao.com) has nearly 500 million registered users, with more than 60 million visitors daily. It sells more than 800 million items daily, with 48,000 sold every minute. At the end of 2011, Taobao's single-day transaction volume peaked at 5.2 billion RMB, creating 2.7 million direct employees. With its expanding scale and the increasing number of users, Taobao has transformed from a single consumer-to-consumer (C2C) net mart into a comprehensive retail and business center that includes C2C, group purchase, distribution, auction, and other e-commerce models. Currently, it is one of the key e-commerce trading platforms worldwide.

With the rapid development of Internet technology, Taobao's e-commerce business activities are also booming, making shopping quick, convenient and low-cost. As a leader in this sector, Taobao provides quick and convenient Internet business services for hundreds of millions of users at a rate of tens of thousands of transactions per minute. With so many transactions, Taobao's business data has also been sharply expanding from gigabytes (GB) to terabytes (TB)—a figure that will likely increase to hundreds of terabytes. Since data is at the heart of e-commerce, and the company's traditional relational database has been unable to handle so much data, how to build a new database to deal with this massive data became a huge challenge for Taobao.

To meet the challenge, Taobao has been engaged in R&D for its OceanBase high-performance distributed database system since 2010. This system, which can handle hundreds of billions of records and several hundreds of TBs of strutural data, is the driving force for Taobao to handle future business development. However, with the gradual deployment of the new database system, Taobao's hardware platform also needed to be stronger. The existing Serial Attached SCSI* (SAS*)–based hard disk's hardware platform couldn't handle the massive data processing requirements of the new system while providing high performance, capacity, and reliability and low total cost of ownership (TCO). Taobao's challenge was how to build a new, stronger hardware platform for the OceanBase database system's high performance.

To address the challenge, Intel provided Taobao with a new hardware platform based on Intel Solid-State Drives and Intel Xeon processors.

> "Taobao's new-generation OceanBase high-performance distributed database system, based on Intel Xeon processors and Intel Solid-State Drive 320, uses random reading capability and a range of technologies such as 25nm MLC technology, power-off protection, and built-in redundancy, as well as the Intel Xeon processors' powerful processing capability. All these strengths have helped our OceanBase high-performance distributed database system deliver maximum performance and serve us with high performance, capacity, and reliability and low TCO."
>
> Dr. Yang Zhen-Kun
> Senior Researcher, Taobao

The OceanBase high-performance distributed database system hardware platform, based on Intel Solid-State Drives and Intel Xeon processors, helps Taobao provide better quality services for its users

"We need to provide high capacity, low cost and highly consistent and reliable structural data storage and access services in the context of Taobao's rapidly growing data," said Dr. Yang Zhen-Kun, Taobao's senior fellow researcher and the system architect of its OceanBase high-performance distributed database. "Taobao's previous database is a relational database. Since, by nature, it is a single-computer system, it is limited in size. To break this bottleneck, a distributed database is a good solution. Taobao is engaged in e-commerce, so there should be zero errors or omissions in the user's products and the transaction. The display of the webpage should also be as fast as possible. So we need a more consistent and faster database system. Moreover, large numbers of small transactions require minimal cost for each transaction. We developed the new OceanBase high-performance distributed database system on the hardware platform of Intel Solid-State Drive 320. This new database system has made performance breakthroughs and also handled tens or hundreds of times as much data as the relational database. Moreover, Intel Solid-State Drives have guaranteed its query performance and response time and have excellent performance in terms of cost-effectiveness, environmental protection, and energy conservation."

High performance

In the process of e-commerce, every transaction is completed by the high-speed data on the Internet, which needs a powerful background database system. Consequently, for the huge number of transactions Taobao handles, high-performance processing power is essential to success. Equipped with Intel Xeon processors and Intel Solid-State Drive 320, the new system delivers the superior performance Taobao needs. Under Baidu's own testing, the processing performance has improved by five times compared to traditional hard drives, decreasing the number of servers/hard drives needed by a quarter.

High capacity

With so much data to process, Taobao's OceanBase high-performance distributed database system demands high capacity from its hardware platform. Using industry-leading 25nm multi-level cell (MLC) technology, Intel Solid-State Drive 320, with its large capacity and low cost, builds the strong foundation for Taobao's new database system. Compared with previous single-level storage (SLC), Intel Solid-State Drive with MLC technology has doubled storage capacity for the new OceanBase high-performance distributed database system and reduced costs.

High reliability

Data is the heart of e-commerce businesses. And ensuring the safety and reliability of the database system is a key challenge for the hardware platform. Intel Solid-State Drive 320 meets the challenge. Intel has the industry's leading NAND/SSD production and quality control technology. It has also added two important technologies: power-off protection for enterprise-class application environments and built-in redundancy. The power-off protection technology can help the database system avoid data loss in the case of an unexpected electrical accident, which means data can still be written even when the power is off. The built-in redundancy technology provides a redundant NAND flash memory array in the specified storage capacity. When an accident occurs, it can automatically reconfigure the accident NAND flash memory array, reducing the possibility of data loss and further enhancing the reliability of the database system. Compared with the three-year guarantee period of other brands' products, Intel guarantees its solid state drives up to five years, giving stronger quality assurances to its users.

Low TCO

"The previous database system of Taobao's My Favorites* was 16 computers in each of two computer rooms equipped with Intel Xeon processor L5630 and 300G*10 SAS hard drives. Now we are changing to 12 computers in each of two computer rooms, equipped with Intel Xeon processor E5520 and Intel Solid-State Drive 320 300G 10, " said Li Zhen, the project manager of OceanBase. "The unique characteristics of Intel Solid-State Drives enable us to use fewer computers and enjoy guaranteed performance for the database system at the same time. This not only helps to reduce each server's power consumption, but also reduces our TCO. Meanwhile, with the rapid growth of Taobao's business, data and customer visits are also growing accordingly. However, our power and data center cannot grow at the same speed. Therefore, from now on, we will be able to reduce the growth rate in the number of servers while improving our service capacity and performance. Otherwise, we not only couldn't afford the cost, but also couldn't find computer room and racks. From the point of energy efficiency, we will have to work in a low-carbon and environmentally-friendly way."

Find the 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).**

Features of OceanBase High-Performance Distributed Database System

- **Functions**

  The OceanBase system can handle hundreds of billions of records and several hundreds of TBs of data with a relational database's main functions including cross-row and cross-sheet business, query within a scope, and join. It also supports automatic fault tolerance and online expansion in the event of a server failure and disregards the tables in database.

- **Characteristics**

  As a distributed system, OceanBase replicates data among several servers (typically three to four replicas) and thus achieves the high reliability of a high-end server by using commodity servers.

  Real-time or near-real-time synchronized slave clusters in a MAN (metropolitan area network) or WAN (wide-area network) help to counter against various kinds of human-led and/or natural disasters. Random write of disk is abandoned and thus makes solid state drives a perfect match to OceanBase. There is no sharding from users' view as it is made by OceanBase internally.