China PnR builds scalable, real-time data warehouse analytics platform

HPE Vertica enables fast data queries and meets risk management requirements

Objective
Improve data processing performance and meet regulatory, risk management and real-time query requirements

Approach
Carried out detailed market research and screened two completely different solutions

IT Matters
- Overall data processing performance has improved by a factor of ten, data extraction speed has improved by 300 percent and data queries are 100 times faster

- Volume of data used for analysis has expanded from a few months' worth to 3-5 years' worth and further expansion can be made without a time limit

- HPE Vertica has high scalability and is easy to manage so total cost of ownership is reduced

Business Matters
- Financial risks are mitigated with support from core financial applications for risk control, money laundering and fraud prevention. Business and regulatory requirements are met and real-time data queries for various complex requests are provided

- Analytics provide more precise opportunities to understand clients and users for better service and focused marketing

Financial services organization, China PnR, suffered a data explosion. Its existing tool could no longer cope with loading and aggregating data and performing real-time queries, so China PnR began to use HPE Vertica to build a data warehousing analytics platform. This would enable transactions and user data to be loaded, accessed and analyzed continuously in real-time and improve performance by a factor of ten, to meet business development requirements and prevent financial risks.

Challenge

Increased data volumes
China PnR Co. was established in 2006 with an investment of nearly one billion yuan. The company provides integrated financial services such as financial payment and investment banking for small and micro enterprises, financial institutions industry clients and investors. China PnR is headquartered in Shanghai with branch offices in more than 30 cities. It has three main subsidiaries: China PnR Data, China PnR Finance and China PnR Technology.

"HPE Vertica has a much higher data processing capability than previous generation systems. Its query speed is hundreds of times faster than that of a traditional database, and its R support allows us to take a solid step forward in terms of data applications."

— Yi Juan, general manager, China PnR Data Second Research and Development Department

The company has moved into new financial sectors for almost two years, providing services for more than 800 P2P (Peer to Peer) platforms through its payment, settlement and account systems, operational risk control and data services. In 2014, China PnR's trading volume broke through the 1.5 trillion yuan mark, giving it an annual compound growth rate of over 300 percent.

Rapid development of the company's services has led to an explosive growth in data volumes. It needs to complete the collection, warehousing, consolidation, aggregation and reporting of data. This requirement is, in addition to performing risk management on this data analysis and meeting the requirement of real-time queries. However, the analysis processing performance of China PnR's original solution, based on traditional data warehousing, was low. It was slow and unable to meet the requirements of real-time queries or large-scale compression. The surge in data volumes has had a major impact on back-end systems. In addition, the original database system used up valuable server and storage resources.

The risk management of financial services organizations is a very important issue for both corporate managers and high-level regulatory authorities. The focus of the People's Bank of China's regulatory authorities are focused on preventing and resolving systemic financial risks in order to maintain financial stability in China. In the face of an endless stream of innovations in financial services, the regulatory instruments used by the People's Bank have also changed from the traditional, on-site methods of checking reports to real-time requests querying all types of data.

China PnR urgently needed to consolidate its business transaction and user information data. It needed to re-establish a user-centric data warehousing analytics system capable of continuous real-time data access and analysis. The company developed strict data warehousing standards to meet the following conditions:

• Fast data transfer: Processing speed requirements will be increasingly higher due to constantly growing data volumes.

• High scalability: The new data warehouse must be able to economically support geometric increases in data volume, without the purchase requirements of an expensive new storage system.

- Data analysis capabilities: The new data warehouse must store data, and also support a variety of data analysis requirements, ranging from simple to complex, from SQL to statistical models.

- Stability and reliability: The new data warehouse needs to be reliable, fault-free and easy to maintain.

- Simple management: The new data warehouse must have automated basic processing functions and allow employees to perform maintenance, hardening and task optimization without the need for external resources.

Based on these requirements, China PnR carried out detailed market research and screened two completely different solutions: HPE Vertica, a business application, and Hadoop, an open-source product.

Solution

Improved scalability
HPE Vertica supports the open source R programming language that is used for data modelling and data analysis.

These applications can be integrated seamlessly to enable China PnR to combine a Vertica analytics platform with various R data models for high-performance analytics performed in a simple, scalable way. Following POC testing and evaluation, China PnR ultimately chose HPE Vertica as the data warehousing analytics platform that would achieve its project objectives.

Yi Juan, the general manager of China PnR Data's Second Research and Development Department, says: "Performance is our most important consideration. China PnR's IT and corporate business development need to be synchronized and become more forward-looking. We must always be at the forefront of technology. This is because real-time queries and other features will only be implemented when the basic data warehousing system is faster, more secure and more efficient. This will enable the company to gain insight quickly and to respond to every kind of change with agility and speed."

HPE Vertica default settings process large numbers of requests quickly, and its 10+ thread parallel loading is able to achieve near real-time loading. It does not have any disk I/O bottlenecks and is able to perform uploads and queries at the same time enabling more people to access information quickly and easily. HPE Vertica is able to perform data compression processing efficiently, which significantly reduces database storage space.

Customer at a glance

Software
• HPE Vertica

"The establishment of the HPE Vertica data warehousing analytics system consolidates our business transaction and user information data. It enables operators to perform queries and analysis on this data more quickly and easily than before. The product's stable operating efficiency prevents data loss during high-speed processing, and the system is able to expand at any time."

— Yi Juan, general manager, China PnR Data Second Research and Development Department

HPE Vertica also provides scalability, says Yi Juan. He believes this is one of the product's main highlights, and fits with the needs of his company: "We started by running HPE Vertica on three rack servers, and the system has not lost any data during two years' of operation."

By using additional cluster nodes, HPE Vertica has successfully expanded performance and data volumes, and is processing more data. China PnR now plans to increase the three-node foundation to six nodes by the end of 2015. This process will be fully automated and without any disruption to service processes.

Benefit

Quicker data extraction

HPE Vertica data warehousing analytics system has consolidated all of China PnR's business transaction and user information data. And, it has enabled operators to perform queries and analysis on data more quickly and easily than before. The product's stable operating efficiency prevents data loss during high-speed processing and the system can expand to meet increases in business or data volumes.

HPE Vertica has now become China PnR's main data repository, storing approximately 300 million pieces of data. It delivers a much higher data processing capability than previous generation systems and an overall improvement in data processing performance by a factor of ten.

Data volume for analysis has expanded from a few months' worth to 3-5 years' worth. Batch processing has gone from taking 10 hours down to taking 40 minutes, while incremental data can be used for analysis within 1-5 minutes.

The speed of data extraction has gone from 4-5 hours down to 1-2 hours, an improvement of 300 percent.

Data queries have become 100 times faster. Every complex query can be completed in 2-3 seconds; this even includes the time taken to operate the mouse. Responses are achieved in seconds, and several operators are allowed to submit queries at the same time.

Learn more at
hpe.com/software

Sign up for updates

★ Rate this document

Hewlett Packard Enterprise