Case Study

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

HPE Vertica enables fast data queries and meets risk management requirements

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
• 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.
“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 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

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.