

How Pink OTC Markets Cuts Reporting Time and Operating Costs with a Market Data Warehouse Built on Vertica and Syncsort

Overview

Pink OTC Markets Inc. operates both the OTCQX and Pink Sheets Over-the-Counter equity marketplaces. The third largest equity trading marketplace in the U.S., Pink OTC Markets needed to improve its reporting capability to stay ahead of the millions of records its trading system amasses daily. By supplementing its traditional relational database management system (RDBMS) with a market data warehouse built on technology from Vertica and Syncsort, Pink OTC Markets dramatically cut reporting time while saving on hardware and programming costs. The Vertica and Syncsort solution allows Pink OTC Markets to publish new applications and reports much faster, operate more efficiently, and compete more effectively against larger, deep-pocketed competitors.

The Application

Pink OTC Markets connects OTC (Over-the-Counter) marketplace participants, providing a quotation and trading platform for FINRA member broker-dealers, market data for OTC participants, and disclosure services for issuers with securities traded on the OTC market. Using Web-based dashboards, customers can analyze, get quotes and trade more than 8,000 U.S. equities and a growing number of international equities available as American Depository Receipts (ADRs).

To keep this diverse set of clients happy, Pink OTC Markets needs to quickly generate reports for up-to-the-minute market analysis. Customers expect “web-time” response and 24x7 uptime.

Delivering fast, reliable analytics is directly tied to Pink OTC Markets’ profitability.

The Challenge

Staying ahead of demand was becoming a challenge for Pink OTC Markets, given the explosive growth of data. As of early 2008, its database had 600 million rows and was growing by two to three million records a day. To enable real-time queries, the company stored data for popular queries in memory and used a complex combination of data instructions and Java applications to move the data around and query it.

The Vertica Solution at a Glance

The Company



<http://www.pinkotc.com>

The Industry

Financial Services

The Application

Real-time market analytics

Benefits

- *Ad hoc* querying of detail data in billions of records – no cubes or summary tables required
- 100X -100,000X faster than traditional RDBMS
- Programmers can create new reports in minutes not days, using SQL skills – has been able to retire much of its Java code
- Can cost-effectively store billions of records and scale easily by simply adding CPUs – no additional licensing fees
- Highly available and highly reliable – no downtime in a year of production use
- Reduced administrative costs and competitive licensing terms provided superior economics over open-source solution

Data also had become heavily silo-ed. Internal users were forced to work with multiple databases when they did their analyses, as were programmers when they improved reports or created new reporting products.

Pink OTC Markets wanted a comprehensive repository for all its data – one that provided a holistic view across all data sources and enabled faster time-to-market for new applications and reports. The repository had to accommodate fast, ad-hoc queries *and* real-time data loading. If the repository were built correctly, it would reduce the load on the transactional applications by offloading ad-hoc queries and reporting. It would also enable new insights into data and allow Pink OTC Markets to create data marts for different reporting functions.

Pink OTC Markets leverages best-of-breed technologies to create an efficient marketplace. The company knew that choosing the right data-warehousing technologies – and the right vendors – was critical. The company expected a solution that was faster, better, less expensive and quick to deploy.

The Solution

Pink OTC Markets first considered a data warehouse based on its traditional RDBMS platform, but rejected it as too costly and complex. The company then zeroed in on the Vertica Analytic Database and Infobright, an open-source MySQL database engine.

The Vertica Database is a high-speed relational SQL database management system purpose-built for analytics and business intelligence. It uses a column-oriented, MPP architecture, aggressive compression and other modern technologies to achieve dramatically faster performance than traditional databases while running on industry-standard hardware. Vertica handles queries very fast (usually 50x-200x faster than other databases), inserts data at high speeds (hundreds of thousands of rows per second and terabytes per hour) and reduces hardware costs by 50%-90%.

Pink OTC Markets conducted a head-to-head comparison of the Infobright and Vertica databases. Although the databases performed comparably out of the box, Vertica performed dramatically faster once Pink OTC Markets' database team optimized Vertica for its workload (a simple automated process). When querying large tables, Vertica was 5X to 1000X faster than Infobright. (It was 100X to 100,000X faster than the incumbent RDBMS.)

**“Our first proof-of-concept with Vertica was very encouraging. When we loaded in some of our tables and ran our initial queries, we were blown away by how fast the Vertica Database was.” –
Rahul Bose, Director of Applications Development, Pink OTC Markets**

“Our first proof-of-concept with Vertica was very encouraging,” recalls Rahul Bose, Pink OTC Markets' Director of Applications Development. “When we loaded in some of our tables and ran our initial queries, we were blown away by how fast the Vertica Database was.”

Other advantages of the Vertica Database were ease of administration; scalability; and high availability. “We were able to install Vertica right out of the box on a three-node cluster and it worked seamlessly,” says Bose. “It’s been in production now for a little less than a year, and we haven’t had any outages. We can take a node off to do maintenance on the box and the cluster still stays up, so we haven’t had any downtime.”

Another significant advantage: Vertica offered Pink OTC Markets a complete solution, through its partnerships with Equant Solutions and Syncsort. Systems integrator Equant designs high-performance business-intelligence and data warehousing solutions for demanding applications in the financial-services industry and other industries. Syncsort offers DMExpress, a high-speed extract-transform-load (ETL) software package.

For Pink OTC Markets, Equant designed a data model that was based on a combination of a star schema and a snowflake schema. Both schemas are supported by Vertica.

“Vertica has this great shared-nothing architecture,” says Ramesh Allu, Equant Solutions project lead on the Pink OTC Markets deployment. “We were able to create a three-node production cluster that has built-in redundancy and will save hundreds of thousands of dollars a year in storage-area network costs. Vertica also scales linearly, which means that if our three-node cluster can’t handle demand we can just add one more node to the cluster and we are good to go.”

“Vertica’s compression also saves a lot on the cost of disk space,” says Allu. “We are achieving a four-to-one compression ratio when moving data from our OLTP platform to Vertica. Finally, since customers pay by the terabyte with Vertica, there’s no extra cost for adding more CPUs, or for making hardware changes, or for building-in replication, high availability or disaster recovery. We pay per terabyte of production data – period.”

When the design was ready to be implemented, Equant selected an ETL tool, evaluating both open-source and commercial alternatives. Syncsort’s DMExpress was dramatically faster than the other tools at providing real-time market data integration and loading – 10x to 1000x faster at bulk-loading large tables.

Together, these technologies create an easy-to-use data warehouse that enables Pink OTC Markets’ analysts and developers to load, transform and generate complex reports on large volumes of data with minimal effort. “Using Syncsort, we were able to load 800 million records into Vertica in less than half an hour,” says Equant’s Allu. “This would have taken any other database or ETL hours.”

The speed and compression of the Vertica Database enables ad hoc querying of detail market data, eliminating the need to build cubes or summary tables. Programmers can build new reports in minutes instead of days, using SQL skills instead of Java programming.

The Final Word

Bose concludes: "Pink OTC Markets leverages best-of-breed technologies to drive a highly efficient OTC market at lower cost than traditional securities exchanges. Our business is connecting traders, improving price discovery and providing better intelligence to our customers. The Vertica-Syncsort solution solved a big data analytics challenge at a fraction of the cost and with much better performance than traditional database systems."

Try the Vertica Analytic Database Yourself

If you would like to learn more about how the Vertica Analytic Database can help your company transform business analytics, or to request an evaluation copy, please visit www.vertica.com.