THE VERTICA® DATABASE
SIMPLY FAST.™

Powered by Vertica Solution Series

Vertica-Powered Glassbeam Delivers On-Demand Product Analytics

DBMS Improves Profitability of Innovative Software-as-a-Service

Based in Sunnyvale, Calif., Glassbeam offers the first software-as-a-service solution for analyzing product operational data. Using advanced visual-data analysis and data warehousing, Glassbeam turns any semi-structured data, such as device log files, into actionable business intelligence. By replacing its PostgreSQL database with the Vertica Analytic Database, Glassbeam achieved 50X - 1000X improvement in query speed with one-tenth the storage space. The company is now better positioned for growth because its service can handle more data, more types of queries and more concurrent users, and can scale inexpensively when needed.

The Application

Manufacturers of computers, network devices, medical instruments and other technology products have a big business opportunity.

Once installed in customer sites, these products continually collect operations data – log files, configuration changes, problem conditions and other event data – and send it back to the manufacturers either proactively via "call-home" applications or reactively by sending in "support bundles" on demand for support escalations analysis. Regardless, this data is a potential treasure-trove of information – for product marketers (who are wondering "how can I up-sell or cross-sell this customer?"), for engineers ("is the product working as planned?"), and for service and support departments ("how can I stay ahead of what's happening and sell the customer the best support plans?").

Getting fast answers to questions like this can mean millions of dollars of revenue for many companies. Industry experts estimate that product feedback from customer logs can be worth more than 4% of revenue in services for some enterprise vendors.

However, *capitalizing* on product intelligence has been challenging, for two reasons. First, analyzing the data is difficult: it's a mish-mash of semi-structured data – data that's meaningful to the human eye (with some work) but not easily readable by automated methods. Second, the vast amount of event data puts the capability out of the reach of most vendors, even those with sophisticated IT departments. It's just too costly and complex to collect, store and optimize the data.

The Vertica Solution At a Glance

The Customer

glassbeam™
product intelligence on demand

www.glassbeam.com

The Industry

Manufacturing

The Application

- First software-as-a-service (SaaS) application for analyzing product operational data
- Achieved up to 1000X improvement in query speed with one-tenth of the storage space
- Transitioned its proprietary analysis software from PostgreSQL to Vertica in only two man-months
- Can deliver higher service levels for ad hoc queries of 1TB+ data sets (95% of Vertica queries run in under 5 seconds); enables real-time analysis and inexpensive storage of historical data
- Lower TCO enables higher profitability: database runs on inexpensive hardware and Linux; can support more concurrent users with less hardware, scales easily and inexpensively

Enter Glassbeam. Founded in 2006, Glassbeam has created a commercial solution that meets both challenges. The solution combines patent-pending technology for parsing (recognizing, combining and normalizing) semi-structured data; a multi-terabyte data warehouse; and powerful visual analysis tools. Glassbeam delivers the solution as software-as-a-service (SaaS), eliminating the need for customers to invest in their own infrastructures and database experts. Gartner named Glassbeam one of its "cool vendors in business intelligence and performance management for 2009."

> *"Vertica gives orders of magnitude faster performance while consuming one-tenth the storage."*
>
> *– Larry Lancaster, CTO Glassbeam*

Says Glassbeam CTO Larry Lancaster: "We figured out how to make rigorous use of un-rigorously structured data. Our Semiotic Parsing Language – or SPL – is an interpreter that works on a declarative language that we've written. It unifies the parsing, ETL, database modeling, and OLAP tooling of semi-structured data into a single development step that is intuitive. Previously, getting to this point would have required multiple man-years of heavy data warehouse development, coordination among multiple teams, and a lot of resources."

The Problem

In spite of its innovative technology, however, Glassbeam had a problem. Customers' datasets were so large and complex that they were overwhelming the company's database technology (PostgreSQL), resulting in slow query times and high storage costs. Optimizing PostgreSQL was costing too much, making it impossible for Glassbeam to offer a competitively priced service and still be profitable.

A typical Glassbeam customer processes a billion events per year, or about a terabyte of data. The database uses 30 database objects, and queries typically involve an average of 14 foreign key references or 15 tables in a join (up to 38). Once Glassbeam runs the data through SPL, it becomes about 500 tables in the data warehouse. The three largest of the 500 tables add up to 2 billion rows, and the 500 tables have about 2,200 unique columns in total.

Getting fast query performance on such complex data was proving difficult. Glassbeam tried optimizing queries to run on PostgreSQL. The team partitioned all the tables, denormalized some of them, and created 700 indexes. However, as a result, the 1TB of source data became 9TB in the data warehouse. Meanwhile, query times were still intractably slow, even with only 10% of the production data loaded.

The Solution

The company's search for a more robust database led it to the Vertica Analytic Database. The database is a high-speed, relational SQL database management system (RDBMS) purpose-built for high-speed analytics and business intelligence. It uses a column-oriented, MPP architecture; aggressive compression; and other modern technologies to handle queries very fast and to operate inexpensively.

Optimizing Vertica for queries proved relatively simple. As a column-oriented database, the Vertica database is inherently optimized for queries, storing data the way it is queried (in columns). To further speed query performance, Glassbeam took advantage of a Vertica technique called projections, which matched well with the input for Glassbeam's software, an arbitrary acyclic schema. (Projections are collections of columns stored in various sort orders on multiple nodes.) Optimizing Vertica involved about 50 pre-joined projections, averaging 50 fields per join, and up to 750 join fields in some cases. "We were able to understand projections well enough to make them work for a given schema," says Lancaster. "Projections are not voodoo, and they work great. It's phenomenal to be able to do pre-join projections across a normalized data model without having to worry that it will sully the underlying databases. I'd strongly recommend projections to anyone trying to scale a database. It's been amazing."

With 10% of production data loaded – where Glassbeam stopped evaluating PostgresSQL – Vertica was delivering up to 1000X faster query performance: 30 *milliseconds* vs. 30 *seconds*. In addition, once fully optimized, Vertica used only 1TB of space – same as the source data – or about one-tenth the space used by PostgreSQL. "It's possible to get even greater savings, because Vertica only takes up a fraction of the source-data space, with the rest being used by our optimizations," says Lancaster. "Our design goal for page load time was less than two seconds, and we're beating that considerably. And it's all because of Vertica under the hood."

Glassbeam tested both databases on a single 32GB node with a 2TB SATA RAID disk.

Moving to the Vertica Database was relatively straightforward, with conversion of the Glassbeam interpreter taking only two man-months. Glassbeam is currently running Vertica in a three-node cluster in a SAN environment.

The Final Word

Lancaster concludes: "The products our customers sell generate a lot of data and they need to analyze this data quickly and accurately to make critical business decisions. For our offering to be successful, we needed a database backend with scalable performance and low operational costs. We evaluated PostgreSQL, Oracle and Vertica and found Vertica provided orders of magnitude faster performance while consuming one-tenth the storage."

The Vertica Database Advantage for SaaS and ISVs

More-differentiated solutions
Store more data, answer questions faster, provide real-time views and new insights

Faster time to market
Works out of the box, no tuning, standard integration with BI tools

Lower TCO
Support more concurrent users on less hardware and with less overhead; requires little or no administration, cuts storage costs by 70%, automatic fault tolerance and recovery, runs on Linux

Built for growth
Scales on inexpensive hardware, special flexible licensing and responsive support

Try the Vertica Analytic Database Yourself

If you would like to learn more about how the Vertica Analytic Database can help your company increase its profitability and competitive edge through better analytic data management, or to request an evaluation copy, please visit www.vertica.com.