

Archiving Digital Content in the Cloud with Vertica

Sonian Inc. is one of a new breed of companies using the power of the Web to deliver critical business-automation functions to customers for less cost than traditional IT systems.

Sonian is a software-as-a-service (SaaS) provider of digital content archiving. Their product, Sonian Archive SA2, archives and indexes email, instant messages and other digital content from customers' communications servers and makes it easily searchable from a Web portal. Sonian's target customers include small-to-medium-sized businesses, large corporations and government agencies – all of whom need to securely store and quickly access large data sets for compliance and reporting purposes.

In the past, customers had to invest in expensive, proprietary IT systems or expensive hosted services to meet their archiving requirements. Today, they can outsource their archiving to Sonian, for a fraction of the cost.

To provide an enterprise-class solution at an affordable price, Sonian uses Amazon's Elastic Compute Cloud (Amazon EC2) and Amazon Web Services. Amazon EC2 gives Sonian access to a cluster of virtual servers that the application can harness in real time to process data quickly for Sonian customers. The Sonian product architecture was designed to scale inside the cloud, enabling Sonian to meet customers' storage and performance demands at low cost as its business grows.

In building out its product architecture, Sonian realized that it would eventually need a database that could scale to accommodate a large number of users doing lots of queries against large data sets – without compromising on performance. The database management system would need to cost-effectively store and analyze terabytes (and eventually petabytes) of customer data. The data includes both the content (for example, the complete content of an email message) as well as metadata, descriptive information that defines the content for indexing purposes. Sonian would be storing a large amount of data for each customer; for example, for a 6,000-employee health care organization, Sonian would be archiving and managing 100 terabytes of data.

"Our whole infrastructure has been designed from the ground up to scale inside of these cloud compute environments, so that puts unique requirements on the database that we use," explains Greg Arnette, Sonian's chief technology officer. "We saw that the MySQLs and the Postgresses just wouldn't work for what we needed – they just can't scale, and we knew that we would encounter capacity and performance problems as our data volumes grew. Although we could have built a home-grown system – and this was our original plan – maintaining and updating this kind of system would have been too expensive and time-consuming in the long run."

The Vertica Solution At a Glance

The Customer



www.soniannetworks.com

The Industry

Information Technology

The Application

Web 2.0 application for digital content archiving

The Benefits

- Fast, simple and cost-effective storage and *ad hoc* querying of terabytes (and eventually petabytes) of customer data
- Scales easily, by adding low-cost commodity servers
- Ease of integration with application's architectural platform (Amazon Web Services)

Vertica Analytic Database – Optimized for Cloud Computing

Sonian instead chose the Vertica Analytic Database 2.0, a commercial column-oriented database that was purpose-built for heavy querying of large data sets.

The Vertica Database organizes data on disk as columns of values from the same attribute, as opposed to storing it as rows of tabular records as in traditional relational databases. When a query needs to access only a few of those values – as in the Sonian application – it only needs to read those columns, making queries very fast. The Vertica Database also uses compression very aggressively, both of data on disk and of data “in motion” during queries; this further enhances query speed and saves on storage costs.

The Vertica Database had been built specifically for highly distributed computing environments (such as cloud computing), and it incorporates some of the same architectural philosophies that are behind Google, Amazon and Yahoo! So, Vertica was a perfect match – both technically and philosophically – for Sonian, says Arnette.

The Vertica Database runs on Linux and clusters of inexpensive off-the-shelf servers, enabling it to scale easily and inexpensively; it works with standard SQL and with popular reporting, ETL and analytic tools.

Arnette notes that Vertica is working with Sonian and other customers to fine-tune the database for the Amazon Web Services architecture. “That was important for us. We didn’t want to be blazing that trail by ourselves,” Arnette says.

The Vertica Database will be used to power the analytic engine behind Sonian Archive SA2. The combination of performance and cost-effectiveness will enable Sonian to meet its goal of providing enterprise-class archiving at a competitive price.

“We have introduced a disruptive pricing model into a market that previously only had premium offerings,” explains Arnette. “To be competitive, however, we need to keep our costs low without sacrificing performance. The Vertica Database and the Amazon cloud compute model are the right combination to give us the scalability we need while keeping costs in line.”

Arnette concludes: “Sonian and Vertica are both disruptive companies. Vertica is doing for analytic databases – really optimizing them for performance and scalability – what we’re doing for digital archiving. They really understand the customers’ pain point, as do we. Vertica is a great match for us.”

Try the Vertica Analytic Database Yourself

Getting started with the Vertica Analytic Database is easy. It supports SQL and integrates with ETL, and analytical and reporting tools as well as business intelligence applications via JDBC, ODBC and specific language bindings.

If you would like to learn more about how the Vertica Analytic Database can help your company more effectively perform data analysis please visit www.vertica.com or call +1-978-533-3500 to find out more.

7 Key Vertica Database Innovations

1. Column store architecture
2. Aggressive compression
3. Concurrent load and query
4. Automatic database design
5. High availability without hardware redundancy
6. Runs on commodity hardware
7. Scale by adding inexpensive servers