



Pantheon Systems

With Apache Cassandra™, Pantheon Systems was able to develop a no-single-point-of-failure, distributable media storage system for all customers using its cloud-based web development platform for websites made with Drupal.

Pantheon Systems provides a cloud-based web development platform for websites made with Drupal – an open-source content management system. Companies can use Pantheon's web development tool and its web hosting and management service for developing, testing, hosting and managing their Drupal-based projects. The Pantheon platform provides a dashboard that enables all stakeholders to have visibility into their projects and to better manage quality control.

Pantheon's clients range from small consultancies that want to collaborate more effectively with their customers, to large organizations that seek a custom distribution of Pantheon's platform for every group involved with content management. One of Pantheon's largest customers is the University of California, Berkeley. Its IT department is using the Pantheon platform to manage a portfolio of Drupal-based sites, ensuring they adhere to a particular look and feel, meet all current security requirements, and that users can interact with the sites as intended.

San Francisco-based Pantheon relies on the Apache Cassandra™ database platform primarily for addressing two key big data challenges – scale and availability. The startup currently maintains two Cassandra-based production clusters; one is for its core application program interfaces (APIs). "All the actual platform data in Pantheon is persisted primarily to Cassandra," explains David Strauss, one of Pantheon's founders and its chief technology officer. "We could wipe out pretty much everything on Pantheon, but as long as the Cassandra store is there, we have our data."

By persisting into Cassandra, and having an interface on top of the platform that's written into a Twisted Python-based representational state transfer (RESTful) API and using a load balancer, Strauss says his team is able to "do any kind of upgrade we want while not taking anything down." He adds, "This is also good from an automated availability standpoint. If network links go down unexpectedly or one of the boxes dies, the API and all major functions of the platform stay online. There's really no other easy tool out there that lets things stay alive when a node goes down or up."

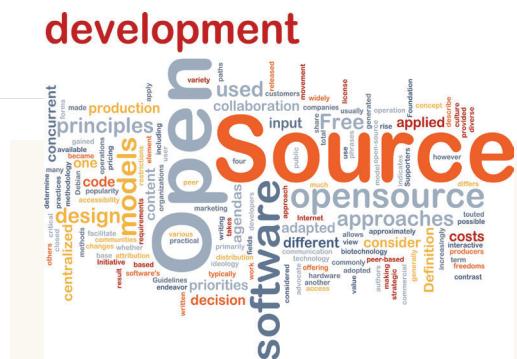
Strauss continues, "By doing a RESTful API where we put our data – which is basically how our whole core API works – we are able to send that data into Cassandra. If the data doesn't exist already, it gets created. If it does exist, it gets updated. That makes our API simpler. It makes our integration with the persistence layer simpler, too."

Chunking out' and scaling up

The other Cassandra cluster Pantheon maintains is for file storage, according to Strauss. He says, "We talked with several cloud platform providers, and they told us we'd never be able to get – without worlds and worlds of a pain – a multitenant file system backend that would be consistent for all of the things mounting it. Cassandra has allowed us to develop a no-single-point-of-failure, distributable media storage system for all our customers on the platform."

The Drupal open source content management system allows users to upload images, MP3 files and other assets that typically are less than 100MB. Strauss says Pantheon was able to build a fail system on top of Drupal that "chunks out" each file and then stores it into Cassandra.

"Cassandra's columns don't have to be tiny bits of data – they typically work fine for up to 10MB," he explains. "Chunking it in Cassandra allows us to have a file system that is highly available. We are able to add new nodes to the Cassandra cluster without disrupting the system – while providing higher capacity overall."



Company

Pantheon Systems

Data Size

2 production clusters

Business Challenge

Pantheon looked to provide a highly reliable, distributable media storage system for clients using Pantheon's cloud-based web development platform to manage their Drupal-based projects.

Technical Challenge

The need for a low-maintenance, NoSQL database solution to help address two of Pantheon's key big data challenges – scale and availability.

Solution

The Apache Cassandra™ platform, which has no single point of failure, scales in minutes, delivers on high availability requirements, and allows Pantheon to provide customers with a robust file storage solution and to develop new features for identifying and supporting emerging “power users” of the platform.

Peace of mind – and simplicity

Before deciding to implement a NoSQL solution, Strauss and his team considered using basic MySQL to address Pantheon's big data challenges – but their concerns about high availability led them to Cassandra. "I'm one of Drupal's two MySQL maintainers and I've built out some of the largest Drupal sites on the Internet," says Strauss. "However, I just don't have much faith in MySQL as a core system for maintaining something that's highly available. It requires too much babysitting. You need to constantly address glitches introduced into the system, rebuild replicas, and so on. I don't have to deal with that type of ongoing maintenance with Cassandra."

Cassandra also provides "unprecedented durability," according to Strauss. "We have a replication factor of three on the data, so anything being written to our media storage backend is also being written to three different boxes, which may or may not be on the same continent," he says. "Cassandra's durability is what gives us confidence about using it at the absolute core of the platform for our customers."

Cassandra's snapshot capability is another favorite feature for Strauss. "Being able to run the snapshot command with nodetool on the cluster, and knowing that every single node has produced a snapshot, allows me to proceed with data changes in the system without worrying about how I would restore copies," he says.

Strauss says Pantheon is making the most of other features in Cassandra to improve the service it provides to customers. As an example, the company has developed a user analytics feature that uses Cassandra counters to track when a user achieves "power user" status on the Pantheon platform.

"By using timestamps and activity markers, we'll know how many times a user has performed a certain action and when they last performed it," Strauss explains. "We can aggregate the data to find out where someone is in the conversion pipeline. This allows us to determine when and how to reach out with support that will help customers in their transition to becoming more active users of the platform."

About DataStax

DataStax powers the apps that transform businesses. DataStax powers over 250 Big Data apps for startups and 20 of the Fortune 100 with its flexible and massively scalable big data platform built on Cassandra, through multi-data centers.

DataStax Enterprise delivers enterprise-ready Cassandra, then goes one step further by integrating the best of breed Big Data technologies — Apache Hadoop for analytics, and Apache Solr for search across multiple datacenters and the cloud.

Top companies such as Adobe, HealthCare Anytime, eBay, and Netflix rely on DataStax to transform their businesses. Based in San Mateo, Calif., DataStax is backed by industry-leading investors: Lightspeed Venture Partners, Crosslink Capital and Meritech Capital Partners. For more information, visit <http://www.datastax.com/> and follow@DataStax.



"Cassandra's durability is what gives us confidence about using it at the absolute core of the platform for our customers."

*David Strauss
Co-Founder and
Chief Technology Officer
Pantheon Systems*