



Company

Netflix

Data Size

Seven clusters in production by end of 2011

Challenge

To maintain high availability of member information and quality of streaming video data for a rapidly growing company, while creating flexibility and agility for Netflix's technology teams

Solution

The open-source and scalable Apache Cassandra data platform, which lets Netflix quickly create and manage data clusters and minimize the chance of failures and outages

With its streaming service rapidly gaining popularity with its members, Netflix's choice of Cassandra is helping the company meet heavy demand for highly available data.

Netflix, the world's leading subscription service for movies and TV, launched in 1997 as a provider of DVDs by mail. However, since it launched its "Watch Instantly" streaming delivery service in 2007, the Los Gatos, Calif., company has steadily shifted resources toward the streaming model, since it offers benefits to both the business and to customers (no postage costs or delivery wait times).

At the same time, the company considered the data and storage demands that streaming media would require. (DVDs require warehouse storage – streaming movies require storage of information.)

"We had a single data center, which meant we had a single point of failure," explains Adrian Cockcroft, cloud architect at Netflix. "We were approaching limits on traffic and capacity. Now that people can watch Netflix streaming programming from their phones, from Wii devices, Roku boxes and many others, the demand for availability increases all the time. We have more customers every quarter, more customers are using streaming, and they're using streaming at a greater rate."

Data was growing as fast as the customer base, Cockcroft says: The number of API requests in January 2011 was 37 times higher than requests in January 2010. Netflix stores the entire viewing history of all 25 million members, including the position at which they stopped viewing a streaming program, so that this position can be retrieved when they return to watch the program. As of July 2011, Netflix offered 12,000 streaming video titles.

The fear was that outages or poor-quality streaming experiences might occur, which would damage the company's

reputation. "We knew we had to get out of the data center so we could keep running and keep growing," Cockcroft says.

Taking Advantage of Cloud Flexibility

In 2010, Netflix began moving its data to the cloud – specifically, Amazon Web Services (AWS) – and the company is now nearly 100 percent cloud-based. The next step in Netflix's plan to build in greater agility was to replace its Oracle SQL database, which managed these massive data stores.

"For us, the problem with a central SQL database was that everything was in one place – which is only convenient until it fails," Cockcroft explains. "And because these databases are expensive, you tend to put everything in there. Then everything fails at once."

Central SQL databases also inhibited the exchange of data around the world – a capability that Netflix needs, since it plans to expand its streaming service to Latin America and the Caribbean by the end of 2011. "SQL databases work best when everything is inside one machine," says Cockcroft.

Another problem with the SQL database, Cockcroft says, is that schema changes required system downtime. "Every two weeks, we'd have at least 10 minutes of downtime to put in the new schema," he explains. "The limitations of a SQL database impacted our availability and scalability."

It became clear to the Netflix team that the company needed a database that could match the flexibility of the cloud. A NoSQL solution – specifically Apache

“With Cassandra, we get better business agility, and we don’t have to plan capacity in advance, we don’t need to ask permission of other people to build things for us, and we don’t worry about running out of space or power.”

—ADRIAN COCKCROFT, Netflix

Cassandra – offered the globally distributed data model Netflix needed, along with the flexibility to create and manage data clusters quickly. To begin the process, Netflix scraped data from Oracle and placed it into Amazon’s SimpleDB distributed database, eventually transitioning that data into Cassandra. (Some data remains in SimpleDB, such as configuration services.)

As of mid-2011, Netflix is using six major applications with Cassandra. Three are currently in production, and three others are in beta and expected to be placed into production soon. The applications include Netflix’s subscriber system, AB testing, and viewing history service (including positions at which members stopped watching a streaming program).

Each cluster has a multiple of 12 nodes. In addition to the six clusters for each application in production, Netflix has a shared Cassandra cluster with 12 nodes, used for smaller applications that don’t need their own cluster.

Cassandra Creates Business Agility

“With Cassandra, we get better business agility,” Cockcroft says. “We don’t have to plan capacity in advance, we don’t need to

ask permission of other people to build things for us, and we don’t worry about running out of space or power.”

This agility is a tremendous boon to a company that plans to expand its worldwide reach. “I can create a Cassandra cluster in any region of the world in 10 minutes,” says Cockcroft. “When the marketing guys decide we want to move into a certain part of the world, we’re ready.”

The regular downtime that needed to occur with SQL is no longer necessary. “There are no schema to change in Cassandra, therefore, there’s no downtime,” Cockcroft says. In addition, latency is roughly the same for Cassandra as it was for Oracle – a few milliseconds for simple queries.

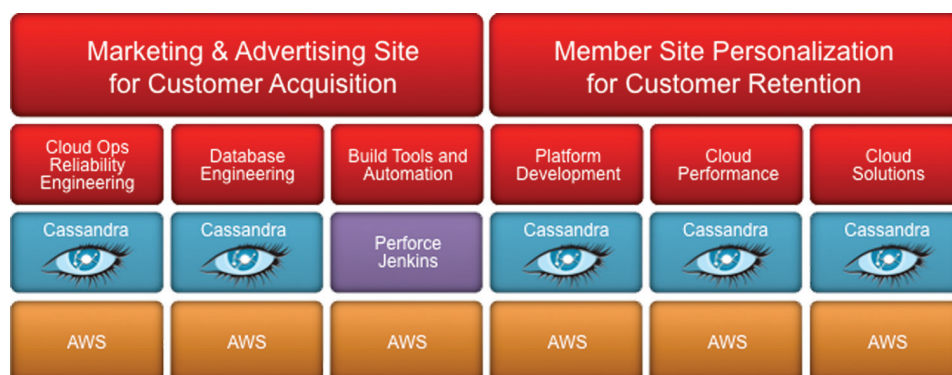
Cassandra does require a bit more management, Cockcroft says, but the benefits outweigh the negatives. “You have many different systems instead of one

system,” Cockcroft explains. “However, they tend to break at different times, so you’re losing small pieces of the system, rather than the whole system at once. Also, each store is simpler to administer.”

Cassandra’s open-source model adds to its flexibility for Netflix. “We can implement our own backup and recovery systems and our own replication strategy,” Cockcroft explains. In fact, Netflix has contributed to making Cassandra even better: “We treat it as a mutable toolkit. If we think Cassandra isn’t doing something well, our engineers will simply change it.”

With its streaming service rapidly gaining popularity with its members, Netflix’s choice of Cassandra is helping the company meet heavy demand for highly available data. “Cassandra is a key infrastructure component of our globally distributed streaming product,” Cockcroft says. “It makes it possible for us to deliver the service our members rely on.”

Netflix “NoOps” Organization



For more information please go to www.datastax.com



DATASTAX

270 East Lane #1
Burlingame, CA 94010

About DataStax DataStax is the developer of DataStax Enterprise, a distributed, scalable, and highly available database platform that delivers optimal performance either on premise or in the cloud for modern enterprise applications that manage both real-time and analytic workloads. The company has over 100 customers, including leaders such as Netflix, Cisco, Rackspace and Constant Contact, and spanning verticals including web, financial services, telecommunications, logistics and government. DataStax is backed by industry leading investors, including Lightspeed Venture Partners and Crosslink Capital and is based in Burlingame, CA with offices in Austin, TX and Stamford, CT. For more information, visit www.datastax.com