



backupify™ WHAT IS BACKUPIFY HOW IT WORKS

HOW VALUABLE IS YOUR DATA?

What would happen if you lost critical Google Apps data due to user error or hacked accounts? Use Backupify and get the peace of mind from knowing that your data is always protected.

- Over 100 Terabytes of Backups
- Over 250 Million Archived Documents
- Over 1 Billion Archived Email Messages

Get independent backups of all your Google Apps domain data – where it can't be stolen, corrupted or deleted.

Features

- ✓ 25 GB Per-User Storage for \$3 / Month
- ✓ Restore Any Lost Content w/ Single Click
- ✓ Setup Takes Minutes
- ✓ Fully Searchable Archives
- ✓ Notifications of Backup Status
- ✓ Dashboard of Your Backups
- ✓ Nightly Backups using Amazon S3

Customer Testimonials

"Every once in a while you read a horror story of someone getting locked out of some online service, seemingly randomly, and causing a huge disaster for themselves. Paying Backupify a few bucks to back up all that data automatically minimizes that risk."

Chris Coyler, Wufoo, and egg-sticks.com
JUNE 28, 2011

"There is really no other solution on the market that provides the functionality of Backupify to help take the stress out of using social media for government purposes."

Keoni Hayes, City of Rockford, IL

In the News

Backupify Re...
Apps Backu...
Sto...
Cloud Comp...
SM
Backupify Ba...
Million Email...
TechCrunch

Cassandra's elasticity has been essential in supporting Backupify's ever-changing big data needs.

Company backupify

Data size
21+ clusters

Challenge

The need for a database that can scale horizontally according to rapidly changing data storage needs, handle extremely high write-loads, reliably back up data on a set schedule, and manage the data sharding process

Solution

The elastically scalable Apache Cassandra platform, which allows Backupify to store, in the cloud, massive amounts of active data coming from the cloud, and provides no single point of failure

Backupify is a cloud-based utility that enables businesses and consumers to backup, search and restore the content of popular online applications such as Google Apps, Gmail, Facebook, Twitter, Blogger and others – essentially backing up the cloud into another cloud. The company's data protection service provides customers with an extra layer of insurance against unintentional or malicious data deletion, and makes regulatory compliance easier for many.

As of September 2011, the Cambridge, Mass.-based company was storing more than 200 terabytes of data for more than 175,000 users – primarily backing up data in Google Apps accounts such as Gmail and Google Docs.¹ "Google Apps is where most business data resides that users want to protect, so that is a primary focus for us," says Matt Conway, Backupify's vice president of engineering. "But keeping up with it all takes work. Gmail is especially difficult because there's so much data to protect."

In 2010, following a round of angel funding based on a prototype, Conway and his team began building a new platform to handle Backupify's projected growth. "I did some architectural design and realized we would need a database that could scale horizontally because the amount of data we'd be pulling down and writing would be really large," Conway explains. "It also became apparent early on in our search process that a relational database wouldn't be an option for us – particularly because of sharding (horizontal partitioning in a database) and the management involved with that process."

After evaluating other solutions,

including the Riak distributed database and the MongoDB open source document-oriented database, Conway decided to implement Cassandra because he was impressed with its underlying architecture. "Apache Cassandra was actually one of the first solutions I investigated because I knew it was a very good fit for high write-loads, and it essentially does the sharding for you," he says. "It was just a better design all around – more truly horizontally scalable and with less management overhead – and there's no single point of failure. I looked at Cassandra's architecture and thought, 'Yeah, that's how you do it.'"

Freeing IT to focus on product instead of infrastructure

Cassandra's elasticity has been essential in supporting Backupify's ever-changing big data needs. The company started out with seven Cassandra clusters and then expanded to 21 then 42 and back to 21 again in a matter of days. "We've been able to scale up and down really quickly according to our needs," Conway says. "I feel confident that as Backupify continues to grow and serve more users, we'll be able to add more Cassandra nodes without significantly changing our management overhead."

While Conway says that Backupify experienced a few growing pains when moving to Cassandra, a NoSQL database, from a MySQL open source database and an Amazon SimpleDB relational database, he says working out issues during the implementation process was far more efficient for the company and its IT resources than building actual infrastructure. "Without Cassandra, our



¹"Backupify Closes \$5 Million in Round Led by Avalon Ventures," by Arik Hesseldahl, All Things Digital, Sept. 8, 2011: <http://allthingsd.com/20110908/exclusive-backupify-closes-5-million-in-round-led-by-avalon-ventures/>

“Eliminating downtime ensures we can backup customer data around the clock. And since we have reliable, redundant and scalable low-balance data storage, if a node goes down for any reason, Cassandra just retries another one.”

—MATT CONWAY, BACKUPIFY,
V. P. OF ENGINEERING

engineers would’ve had to create something that could scale to our needs,” says Conway. “That would’ve prevented us from focusing on building product and solving problems for Backupify’s users, which are far more important tasks.”

The appeal of open source

The fact that Apache Cassandra is a truly open source project also appealed greatly to an experienced architect like Conway, who pays close attention to what his industry peers are saying about the merits of different database solutions. “Cassandra is actually backed by many respected people in the community,” he says. “We’re not talking about some dude in a basement coding up product and making it available. Apache Cassandra is a truly open source project, and that carries a lot of weight. I can tell you it heavily influenced my decision – as did the cost.”

Since implementing Cassandra, Conway and his team have used the DataStax’s OpsCenter as needed to enhance their management and monitoring of Backupify’s cluster ring. The platform provides sophisticated visualizations of Backupify’s Cassandra clusters, allowing the IT team to view and interact with the cluster ring and see overlaid health and status information. “The DataStax OpsCenter

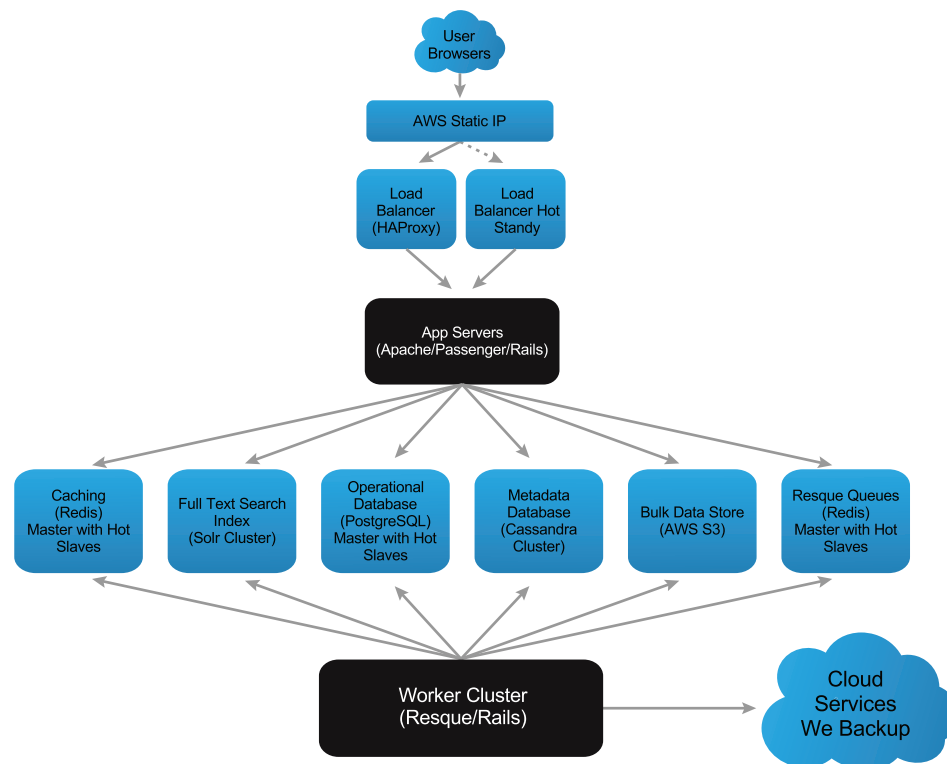
dashboard provides a window into the performance of our cluster,” says Conway. He adds that the platform likely will be even more useful when Backupify eventually upgrades to Cassandra 0.8.

Another important benefit of using Cassandra, according to Conway, is eliminating downtime. “Cassandra provides us with reliable, redundant and scalable low-balance data storage. If a node goes down for any reason, Cassandra just retries another one,” he says. “This ensures customer backups continue as scheduled, without data loss, even in the event of node failures. It’s a very big deal for us that our backups happen reliably – and this is due in large part to Cassandra.” Additionally, IT managers who buy backup for their on-premise apps want it for their cloud apps too, which is why our growth has been faster than anticipated, and Cassandra is

part of the reason we could accommodate that growth easily.”

Looking beyond backup

Backupify is a searchable service, which means there is potential for the company to provide business intelligence to its users someday. “At some point we’d like to be able to say to our users, ‘We’re collecting your data and backing it up for you. Now, is there some kind of service we can run for you against that data?’” says Conway. “That’s what we’re thinking about: Giving people more insight into their data, if they want it.” And because Cassandra is proving to be so reliable and flexible – and not labor-intensive – Backupify’s IT team can focus their attention on development projects just like this that could make all the difference to the company’s future growth and success.



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