

**STORAGE**

## Data Streaming Company Tackles Big Data Storage Challenges

Data streaming company PubNub turned to NoSQL--and third-party help--to meet constantly increasing data storage and security needs.

[Karen D. Schwartz](#) | Apr 17, 2019



When Todd Greene first began thinking about starting the real-time data streaming network company that would become PubNub, it didn't immediately occur to him

probably don't know it. Its data streaming technology is part of more than 500 million [connected devices](#) and phones. It allows companies to send and manage real-time connected experiences with customers, between companies and apps, and between Internet-connected devices. For example, Peloton uses it to allow customers to share workouts in real time, while HubSpot uses it to help its sales associates keep tabs on prospects' activities in real time and chat with site visitors in real time. There is even a chain of laundromats using it to allow customers to reserve and pay for machines from their smartphones and monitor progress remotely.

When the data streaming company first started in 2012, the concept of a strong data storage layer wasn't even on the roadmap. Instead, the focus was on building a global messaging layer that would allow users to send messages from one device and have it be received by another device. Just as importantly, they wanted to be able to deliver messages in less than a quarter of a second. Today, they have shaved that time down to less than 40 milliseconds.

But as the company grew over the next few years, its customers wanted to be able to access messages that had been sent in the past.

“That was the first inkling that we needed a very strong security layer,” said Greene, CEO of the San Francisco company. “For example, we have hundreds of customers who build chat applications, and their customers wanted to be able to see the history of the chat they had yesterday, or from five days ago. Pretty quickly, we realized that our customers wanted us to provide that for them.”

When choosing the technology to power its storage layer back in 2013, Greene's team knew that it had to be massively scalable and able to handle global distribution of the database. And because PubNub wanted to be able to store data

ensure that PubNub wouldn't be caught in a vendor lock-in situation if its requirements changed over time, but Greene found [NoSQL](#) to be a good way to store time series data. Green considered Apache Cassandra the ideal form of NoSQL because he felt that it was emerging as the standard for storing the type of time series data PubNub needed to store at massive scale.

“We went through a pretty thorough reverse analysis of what the options were, how stable they were and what our cost would be to operate,” he explained. “We had a rule back in 2012, which we’ve long since broken, that said that whatever we design, we need to prove it out to 10 times our current traffic. By the end of 2013, we had a new rule: Anything we design must be linearly scalable in both cost and performance up to a few factors of magnitude. But we thought Apache Cassandra could handle it.”

Over time, however, data scale challenges emerged. As the company grew to manage more than twice the world’s global SMS traffic each year, PubNub continued to struggle to scale Cassandra. Latency and speed also were becoming bigger issues.

“The databases ended up getting so big that some types of repair and maintenance processes that had worked in the past were causing really high latencies or failed attempts to retrieve data,” Greene said. “But we’re a company that prides ourselves at being available five nines, and we needed help.”

The first step was getting back on track. PubNub contacted Instacluster, a company specializing in Cassandra migrations, scaling, management and optimization. Within a matter of days, working with PubNub’s team, Instacluster was able to get the system not only running, but also optimized. By changing some parameters, for example, it began to quickly start correcting itself.

## Instaclustr.

From its own offices, Instaclustr now operates PubNub's Cassandra clusters on its own servers. The PubNub staff also can consult with Instaclustr when they are thinking about changing anything in the environment.

“They are our first line of database operators,” he explained. “They are the people who, on our behalf, maintain and tune the database. They make sure our backups are operational. They make sure they can restore it quickly. So they work in tandem with our tech team to make sure that all of the database stuff is operating, and they do the operations of the database. They kind of maintain the servers on which the database is running.”

Since that switch last year, PubNub has met its 99.999% uptime guarantees to customers. That's important, since the company actually gives credits to its customers if it is down for more than 26 seconds per month.

TAGS: [DATA SECURITY AND ENCRYPTION](#)

**0 COMMENTS**

MENU

Q SEARCH

LOG IN

REGISTER



**Too Much Data? Copy--or Copy Data Management--That**  
FEB 20, 2019



**Tape Storage Is 'Still Here'**  
FEB 07, 2019

ITProToday™

an informa business

[About](#)

[Advertise](#)

[Contact Us](#)

[Sitemap](#)

[Ad Choices](#)

[Privacy Policy](#)

[Terms of Service](#)

[Cookie Policy](#)

Follow us:

