



Improving the Office 365 experience for millions of customers with DataStax Enterprise (DSE) on Azure

"We're focused on what we can do for our customers, and we're finding the right implementation for the project, regardless of where it's built or what operating system it runs on. Using a combination of open source software and Microsoft Azure is delivering significant value to our customers."

Sean Usher, Senior Software Engineer, Customer Fabrics Platform, Microsoft



Millions of businesses worldwide use open source technology, and Microsoft is no exception. In fact, one in three Azure Virtual Machines runs Linux, and Microsoft is taking advantage of open source technologies to enhance its own cloud services. For example, the Customer Fabrics Platform Team at Microsoft recently chose DataStax Enterprise—the always-on data platform for cloud applications powered by Apache Cassandra—to improve service across multiple areas. Now, administrators work with DataStax Enterprise to gain insight into user behavior. Results include better support for Microsoft's fastest-selling product, and more satisfied, productive customers.

Gaining a better understanding of user behavior

With more than 85 million active users relying on its services, Office 365 wanted to ensure that its popular productivity platform provides the best possible user experience. To meet its goals, Office 365 turned to the Customer Fabric Platform Team at Microsoft to gain a better understanding of user behavior. “We wanted to build a system that would tell us how the customer was using our platform,” says Sean Usher, Senior Software Engineer, Customer Fabric Platform at Microsoft. “One of our main goals was to detect users having problems before they noticed it themselves.”

To gain insights from Office 365, the Customer Fabric Platform team had been running synthetic transactions that mimicked user behavior. However, testing every possible scenario became too costly. Then the team realized that it had more than 300 terabytes of anonymized data from client and server logs. To tap into that information, Usher and his team needed an always-on data platform that could scale quickly to handle an influx of information from every Office 365 server and client device.

Getting insight with DataStax Enterprise

After looking at a variety of options, the Customer Fabric Platform team chose DataStax Enterprise. DataStax Enterprise combines the power and resilience of a comprehensive real-time

data platform with the big data analytics capabilities of Apache Spark, an open source cluster computing framework. “DataStax Enterprise on Azure enables us to do everything we needed, including near-real-time data processing and longer-term batch processes,” explains Usher. “We’re also able to handle linear data growth without having to shard our clusters and manually manage data.”

Running on Azure Virtual Machines with the Linux Ubuntu operating system, the solution uses the distributed streaming platform Apache Kafka to ingest data from Office 365 servers and client devices. Then information flows into DSE’s Spark and Cassandra. The platform generates datasets that can be widely used for analytics, and it streams data for near-real-time consumption.

The pilot project, which included DataStax Enterprise in a 40-node cluster, focused on reducing support tickets. By targeting both server-side and client-side log data, the team could determine if a problem originated internally or from tenant administrators and solve problems before they escalated. Usher says, “Not only does that save us from getting support calls—which also saves us money—but it also protects the tenant administrator from being barraged by his or her users about broken services.”

The team improves the user experience in multiple other ways too. For example, administrators can see if users make multiple attempts to file a support ticket, indicating that either the page isn’t working or needs to be redesigned. Or, they can identify

“DataStax on Azure enables us to do everything we need, including near-real-time data processing and longer-term batch processes. We’re also able to handle linear data growth without having to shard our clusters and manually manage data.”

Sean Usher, Senior Software Engineer,
Customer Fabric Platform, Microsoft

Customer Name: Microsoft
Industry: Professional services—software engineering
Country or Region: United States
Customer Website: www.microsoft.com
Employee Size: 114,074
Partner Name: Datastax

Customer Profile:
Microsoft Corporation is a leading platform and productivity company for a mobile-first, cloud-first world. Reporting annual revenue of \$US85.32 billion in 2016, Microsoft offers an extensive portfolio of applications, devices, services, and cloud infrastructure, including Office 365.

which users are affected by a service incident and proactively contact them.

Simplifying implementation

The original 40-node cluster has expanded to more than 400 nodes, and the Customer Fabrics Platform team is confident that the highly scalable infrastructure can keep up with the rapid growth of Office 365. "In terms of load, we're handling about 400,000 write requests per second with DataStax Enterprise on Azure," says Anubhav Kale, Senior Software Engineer, Customer Fabrics Platform at Microsoft. "And the latency is very low at approximately 20 to 30 milliseconds per operation."

The team uses Azure Resource Manager templates to automatically deploy new clusters, and the open-source configuration management tool Chef to handle changes across multiple nodes. Usher says, "Nowadays, it's just a click of a button and everything deploys."

Enhancing customer experience satisfaction

The team expanded its DataStax infrastructure to include systems that monitor Microsoft Exchange activity for performance and security. Today, the Customer Fabrics Platform Team runs multiple DataStax Enterprise clusters on Azure.

In choosing DataStax Enterprise to enhance services, the Customer Fabrics Platform Team is putting customers before technology. "Our team is embracing open source to provide value to customers," says Usher. "We're not saying 'We can't do this yet because it's not supported on Windows.' We're focused on what we can do for our customers, and we're finding the right implementation for the project, regardless of where it's built or what operating system it runs on. Using a combination of open source software and Microsoft Azure is delivering significant value to our customers."

"In terms of load, we're handling about 400,000 write requests per second with DataStax Enterprise on Azure. And the latency is very low at approximately 20 to 30 milliseconds per operation."

Anubhav Kale, Senior Software Engineer,
Customer Fabrics Platform at Microsoft

Software

- Microsoft Products
 - Microsoft Office 365
 - Microsoft Azure Virtual Machines
- Open Source Technologies
 - Apache Cassandra
 - Apache Kafka
 - Apache Spark
 - Linux
 - Ubuntu

DataStax is a registered trademark of DataStax, Inc. and its subsidiaries in the United States, and/or other countries.

Apache Cassandra, Apache, Tomcat, Lucene, Solr, Hadoop, Spark, TinkerPop, and Cassandra are trademarks of the Apache Software Foundation or its subsidiaries in Canada, the United States, and/or other countries.

This case study is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY. Document published March 2017

