

# ENTERPRISE SOARING SUCCESS

Mayank Shama meets Robin Schumacher of DataStax to talk about building success with open source while nurturing the community.

**B**ack in the days when virtually all open source software was licensed under the GPL, building a business around it was quite a challenge. MySQL was one of the first projects that attempted to create a true-blue 'enterprise open source' product by creating new licensing terms and adding exceptions to work around restrictive licensing. The proliferation of permissive licenses opened the flood gates, both for the use of open source software in the enterprise and for setting up professional support business around these products.

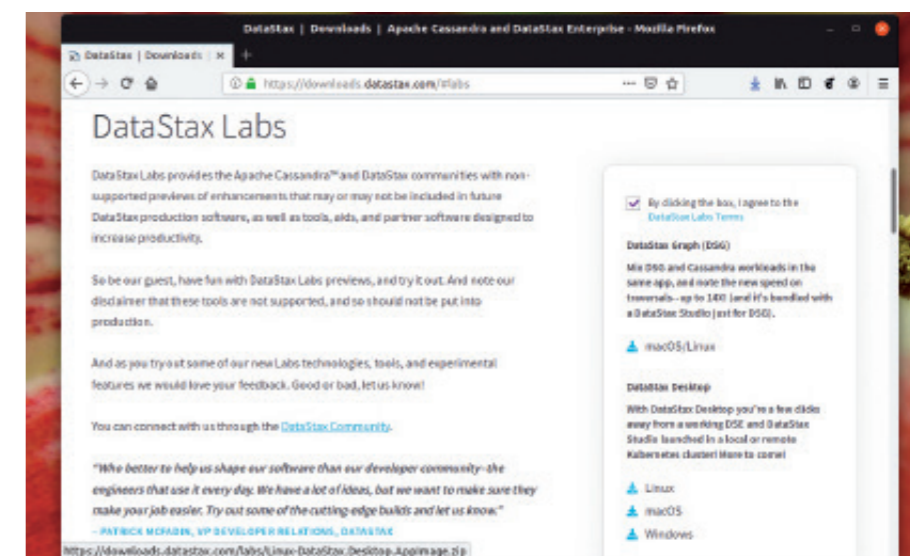
Despite the fact that these days virtually all major open source software has a corporate backer that works with the community, mixing the two without burning your fingers is quite a balancing act. We caught up with Robin Schumacher, VP of Products at DataStax, at the company's Accelerate conference in Washington, to better understand its relationship with the open source Apache Cassandra community. We wanted to see how it contributes to the open source project, and what it takes to offer commercial solutions based on a freely available piece of software.

Robin is a been-there, done-that kind of guy and has worked up and down the corporate food chain. He has been tinkering with all kinds of databases in one form or another for about three decades, first as a database administrator and then making software for DBAs, before heading enterprise product teams at companies such as MySQL AB and EnterpriseDB.

## Paid or not paid?

Robin came to DataStax as one of its earliest employees in 2011, when his good friend Billy Bosworth was tapped to be CEO. He has overseen the company's enterprise products almost from the start, and has been responsible for adding the features you'd expect from an enterprise grade database. "When I got here, we had no security," he says. "There was none. So if you go to a government agency or a financial institution, they're not going to let you in the front door. And so you need to be laying these big pieces of the puzzle that's going to make up your data platform for the enterprises."

Robin cut his teeth on rolling open-source software at MySQL AB – a relationship, he fondly tells us, which



Robin pointed out that, for the first time, DataStax has put up some of its new under-development apps on its website, for anyone to try without signing up for the Early Access Program.

began with him having reindeer meat with the company's co-founder David Axmark in Sweden. Just as Oracle was closing on MySQL AB, he moved into a similar role at a PostgreSQL company called EnterpriseDB. He tells us that his process of turning open source software into enterprise products has evolved over time from MySQL AB to EnterpriseDB to DataStax now. It's critical for a responsible

the ability to have IDs and passwords, and manage the permissions of the various objects. Once they have this list of necessary features that's customary for any database, those were the things they gave back to the community.

"But then we asked, is it normal for a database to have low-level access control? No. That's typically something government and financial institutions

## HOW DATASTAX BUILT UP ITS TRUST

"If you go to a government agency or a financial institution, they're not going to let you in the front door."

company to first reinforce the open source software by contributing the necessary features upstream.

At DataStax he's helped to develop a set of guidelines that help the company determine what it gives back to the community, and what features it retains for paying customers. For him, the segregation process begins with the classification of features: "I like to start with the foundation, where you begin by asking the question in terms of the features you are starting to look at: is the functionality usual and customary for the software's original purpose in the marketplace?"

He illustrates this point by saying that it's imperative for any database to have

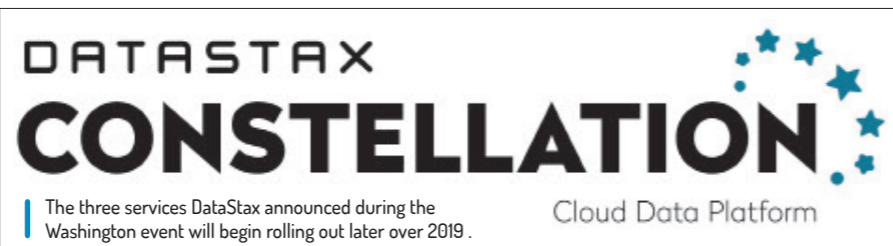
need, and so we kept that feature for our paying customers.

"So we start with that rule of thumb and begin to work out from there," Robin continues. "Then you begin asking some of the big overarching questions. We added analytics as you always need to do real-time analytics on your data. You need to look up data, search of course, secure your data; you bet, be able to manage your data via some visual tools and things, make it easy to handle complex rollouts; absolutely. And then you just keep moving down the line."

Connecting the dots all the way to the latest cloud-centric releases at the conference, Robin says that now that the company has put in the big pieces,

it's time to ease the experience: "Our current roadmap makes it even cleaner in terms of bringing some of these very powerful functions to bear inside of the platform itself. A couple of the first steps is our advanced performance feature set, where it's all transparent to the users. You don't have to worry about doing anything special in your app to make it run faster. You don't have to do anything to your cluster to make sure all the nodes are synchronised with the proper data. It's all handled for you in an automatic fashion."

DataStax unveiled its cloud platform strategy (see box below) during the conference, led by the DataStax Constellation platform, which Robin likens to Microsoft Azure: "Constellation is the overall platform, and within Constellation you are going to have a variety of different services that support cloud applications. So it's kind of like Azure; as Azure is to Microsoft, Constellation is to us. Under



Azure you have a variety of different services that developers and database pros take advantage of. It's the same thing with us."

### No free lunches

During our discussions, Robin shared that DataStax has a very active Open Source committee that's made up of product representatives, engineering and the developer relations and community team that represents the open source interest: "We have very healthy and strong debate about what we are going to be doing

about open source, what are we going to give back to open source. And it not only concerns the Cassandra database, but also things like our drivers. We currently maintain all the open source drivers, but we also have a set of Enterprise drivers; what the differentiation between those looks like, and why." He adds that having a strong committee has helped DataStax not only walk the fine line between being very good open source stewards, but at the same time has helped differentiate its enterprise products to help keep its accounts in the green.

One of the decisions DataStax had to take to ensure its viability came about with the release of DataStax Enterprise (DSE) 6.0. With that release the company prioritised its development efforts behind DSE, and relaxed a number of guidelines that Robin helped develop, in order to retain functionality for its enterprise offering. For instance, one of the guidelines said that open source Cassandra would never run slower, or scale less, than DSE. That changed with the DSE 6.0 release, which included a set of optimisations and utilities that enabled it to not only run faster than the open-source Cassandra, but also helped it to scale much better.

When asked how the larger Cassandra community reacted to the news, Robin said the reaction was actually quite muted. "I was pretty happy to see that we had, I think, only one or two people out on the Cassandra mailing list that complained about that." Robin concedes that the reaction would have been quite different back in his MySQL days, and that's because the open source community these days has much better appreciation for what a vendor like DataStax does: "They understand that you can't get something for nothing all the time."

### Mature community

He recalls an incident from back in the MySQL days when he had someone at an open source event walk up to him and say that they should just put out everything for free. "I just answered him: when you stop taking a pay cheque, so will I." Robin continues: "It doesn't work that way. You need a strong vendor that's incentivised

to be able to buttress these open source projects. And every open source project out there has one or two that do this type of thing."

He's glad that the Cassandra community has been very understanding of the differentiation DataStax has had to build between the products: "They understand that it's a business that has to be run, and they also understand that when that business is successful we are continuing to give back to open source. There are features now in DSE that likely in three to five years we will transfer to open source as we continue to innovate."

One of the developments that DataStax hasn't yet contributed back to the

points out that all graph databases – be it JanusGraph, DSE Graph, Amazon Neptune, Cosmos DB or IBM Graph – are all built on top of Apache TinkerPop, which is the open source graph framework – and DataStax is the largest code contributor to that project.

### Engaging the community

Another bit of the company's engagement with open source that intrigued us was the DataStax Distribution of Apache Cassandra (DDAC). It was announced towards the end of 2018 as essentially the company's release of the open source version, and sticks out among DataStax's proprietary DSE offerings.

to use open source software is to avoid vendor lock-in. "And then thirdly, while some customers understand the need for DataStax Enterprise for large-scale production systems, they might have smaller departmental systems or large development environments where they don't necessarily need the extra firepower that DSE provides."

It's a combination of these three reasons that led DataStax to offer DDAC: "What we have done is taken Cassandra and certified it just as we do for DSE. We supplied the bulk loader and we give them that hand to hold, the support, all at an extremely attractive cost. Believe me, they are spending pennies to save tons of money."

Talking to Robin and the other head honchos at the company, especially its co-founder and CTO Jonathan Ellis, you get the feeling that the company has its heart in the right place. Ellis was the project chair of Apache Cassandra from its beginning, before stepping down in 2016. Besides contributing code, bug-fixes and documentation, the company does its bit through various initiatives such as certifications, meet-ups and more, to nurture the open source community around the Cassandra open source project.

There were also a large number of individual Cassandra users and developers at the conference, and talking to them it was clear that they all saw DataStax as a staunch ally and a committed contributor to the open source project. **LXF**

## » ROLLING THE DATABASE AS A SERVICE

A couple of times during our conversation, Robin alluded to the fact that DataStax has now set its eyes on helping users eradicate some of the day-to-day operational complexity from their workflow. The DataStax Apache Cassandra as a Service is one of the initiatives announced during the conference that'll help it meet this lofty objective.

The database-as-a-service (DBaaS) product aims to simplify app deployment and development provisioning by reducing the process to just a few clicks. Robin tells us that to make the cloud offering more enticing, DataStax has rolled it along with enterprise grade security and offers it on usage-based pricing terms. He adds that it'll also get the self-healing features that ensure users don't have to deal with certain administrative functions that are a headache to deal with in the open-source Cassandra.

But it isn't just a simple port of DataStax Enterprise (DSE) to the cloud: "You're right, there are some things we've ported over from DSE to the new service, but there are things that are just simply not required in the cloud that you may need on-premise. As an example, we have something called DSE multi-instance, and this enables you to take a particular large piece of hardware and run

multiple copies of our software at the same time. But you don't need that in the cloud, because it's already containerised and already contains the virtual machines and everything else. So there's no reason to take all that and put it over to our service. The same is true of some of the other features that we have there."

Some of the DBAs we spoke to wondered how Constellation compared to InstaClustr, which offers a similar platform for managing Apache Cassandra deployments. "They kind of do the same thing, in a sense, yes," agrees Robin. But he adds that InstaClustr takes the open source Cassandra and provides a web console for managing it: "When you look at what they do verses what our distribution offers, we have quite a bit more enterprise capability."

One set of features he points out that is missing in InstaClustr are the advanced security capabilities, because it's based off open source Cassandra's minimalist feature set: "So for people who have critical needs to protect sensitive data, they want to make sure they are able to track who changed what when, who looked at what when; when they need to interface with all the security packages that are out there that their app might need to adhere to... they are going to find that in our service."

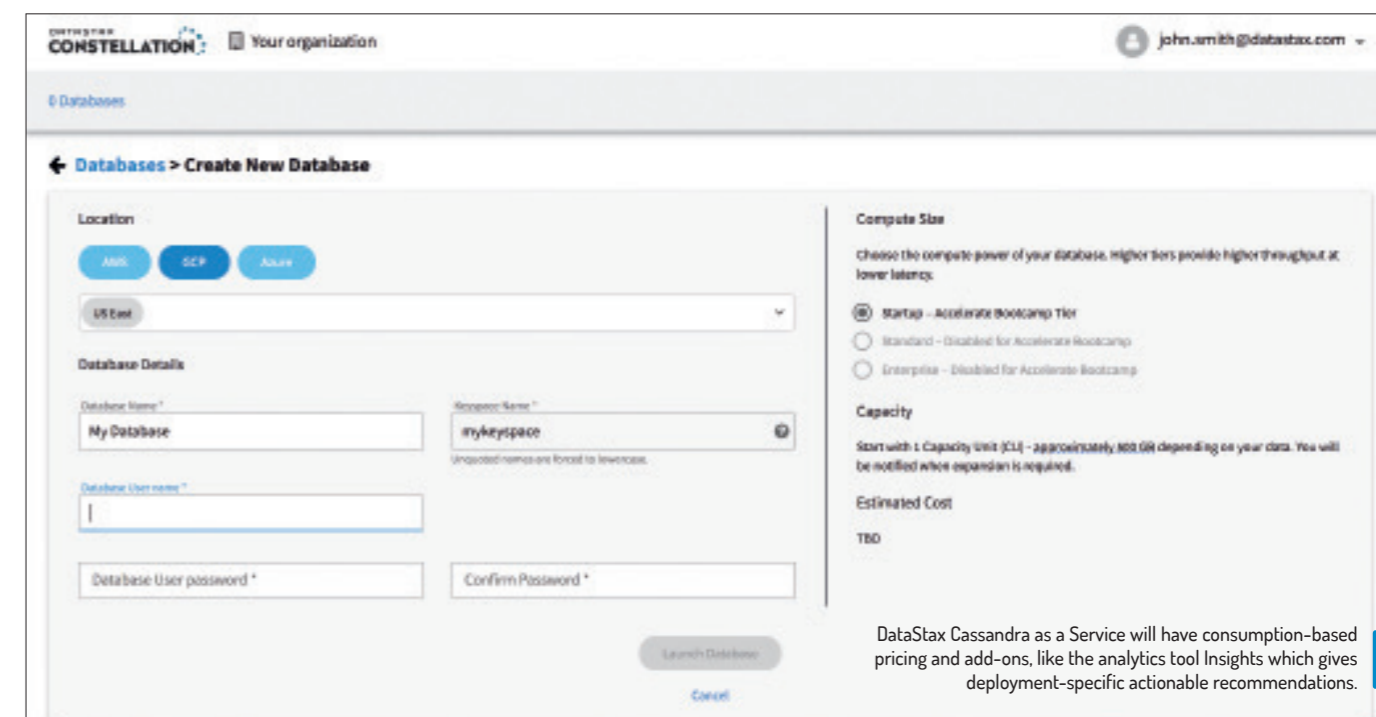
## TALK IS CHEAP, BEER LESS SO

"Back in the MySQL days with large companies like Facebook, Yahoo! and Google – they find out open source is free, like a free dog."

Cassandra project is the Graph database technology it got through the acquisition of Aurelius LLC in 2015. Robin points out that at the time of acquisition, DataStax followed through the commitment it made to the open source community that it would see through the 1.0 release of the TitanDB database.

The project was later forked and rechristened JanusGraph and moved to the Linux Foundation. However, Robin

"A lot of folks, when they say they believe in open source, what they really mean is that they are looking for a zero-cost or a low-cost product. And I get that," explains Robin. "But they quickly find out – and I'll tell you that I saw this back in the MySQL days with large companies like Facebook, Yahoo! and Google – they find out open source is free, like a free dog." Besides cost, another popular reason that companies want



DataStax Cassandra as a Service will have consumption-based pricing and add-ons, like the analytics tool Insights which gives deployment-specific actionable recommendations.