

Pivotal goes all-out open source for 'big data,' launches Hadoop alliance with Hortonworks

Analyst: Matt Aslett

20 Feb, 2015

Rumors of the demise of Pivotal's Hadoop distribution have been greatly exaggerated. However, the company is refining its Hadoop strategy via a strategic relationship with Hortonworks amid a shift to fully embrace an open source strategy for its Big Data Suite. Pivotal's Greenplum, GemFire and HAWQ products will all become open source this year as the company continues to transition customers to a single subscription for the various Big Data Suite components.

The 451 Take

Releasing Greenplum, GemFire and HAWQ as open source is a bold move for Pivotal, particularly since we believe they account for a significant proportion of the company's revenue. However, it has been clear since the launch of Big Data Suite that Pivotal's strategy was to deliver a single subscription-based portfolio of data-processing technologies, and as such it makes sense to make the underlying software as easily consumable as possible to accelerate adoption. It also means that the company now has the same licensing strategy for two of its three core businesses (Cloud Foundry and Big Data Suite, with the other being the Pivotal Labs agile development shop). There is clearly a lot of work to be done to actually deliver open-sourcing Greenplum, GemFire and HAWQ and we also think there is scope for customer confusion in relation to the company's Hadoop roadmap given the ongoing commitment to Pivotal HD and the strategic relationship with Hortonworks. But we would expect any wrinkles to be ironed out over time, driven by customer demand.

Context

Pivotal was founded by EMC and VMware in 2013 to bring together the various cloud- and big data-related offerings previously owned by its parent companies. While Cloud Foundry served as the common umbrella for the cloud-related initiatives, Pivotal inherited a more disparate set of data-related products and services and has taken longer to bring those under its strategic direction. The launch of the Big Data Suite in April 2014 was a big step in that direction – providing a single SKU for Pivotal HD, the Greenplum Database, the HAWQ query engine and the GemFire in-memory database.

The company generated approximately \$40m in sales from the Big Data Suite in 2014, and more than \$100m from its data-related software (Pivotal as a whole posted revenue of \$227m in 2014, up from \$179m in 2013). While the bulk of its data-related revenue comes from stand-alone deployments of its closed source software, the company is buoyed by its ability to generate sales of roughly \$40m from the open source Cloud Foundry, to the extent that it has decided to open-source the closed source components of the Big Data Suite – Greenplum, HAWQ and GemFire – this year. In addition, as it refines its Hadoop strategy, Pivotal has entered into a strategic relationship with Hadoop distributor Hortonworks and has become a founding member of the Open Data Platform Hadoop industry association.

Hadoop strategy

Several news reports recently suggested that Pivotal was set to retire its Pivotal HD distribution of Apache Hadoop, or even get out of the Hadoop business altogether, while laying off its Hadoop development team. None of these reports were accurate. It is true to say that the number of employees working on Pivotal's data products has been reduced – the company won't say by how much or to how many – but maintains that this was a process of consolidating development locations. It is also true to say that Hortonworks will now be providing escalation-level support for Pivotal HD 3, the latest version of Pivotal's Hadoop distribution, which suggests that at least some of the departing employees were involved in Hadoop support.

The support agreement with Hortonworks is part of a larger strategic relationship between the two companies that will see the Big Data Suite become available on the Hortonworks' Data Platform (as well as Pivotal HD), and the two companies collaborating with the other members of the Open Data Platform (ODP) on testing and certifying a common set of the various projects in the Hadoop ecosystem. Pivotal HD will be based on ODP but will continue to exist as a separate Hadoop distribution with additional capabilities (such as Spring XD and integration with Greenplum, HAWQ

and GemFire). Pivotal HD 3 also continues to include the Command Center management console, although that will be replaced over time by Apache Ambari. Pivotal reports that Hortonworks will recommend the deployment of Pivotal HD for use cases that require MPP analytics and in-memory data processing, while Pivotal will recommend Hortonworks Data Platform for all general-purpose Hadoop scenarios.

Open source strategy

Pivotal has also announced the launch of Big Data Suite 2.0, adding support for the Big Data Suite on Cloud Foundry, as well as the Spring XD framework, the Apache Spark in-memory analytics engine, the Redis key value store and the RabbitMQ message queue. However, the biggest news is that the major closed source components of the Big Data Suite – the Greenplum Database, the HAWQ query engine and the GemFire in-memory database – will all be made open source in 2015. Specifically, Pivotal will adopt the Apache License for HAWQ and GemFire and propose them as Apache Software Foundation projects, while the Greenplum Database will adopt the PostgreSQL License and be 'contributed' to the PostgreSQL community.

For the Greenplum Database, this will very much be a case of going back to the (open) source: Greenplum is based on PostgreSQL and was available via the Bizgres open source project between 2005 and 2008. Pivotal notes that it plans to work with members of the PostgreSQL community and is hopeful that it will embrace Greenplum's massively parallel architecture as a next major milestone in the development of PostgreSQL itself. Meanwhile, the company also plans to continue offering value-added functionality for Greenplum, HAWQ and GemFire as part of the 'enterprise' versions that will be available with the Big Data Suite subscription offering. Specifically, it will provide management tools and a new cost-based query optimizer for HAWQ and Greenplum, as well as continuous query and WAN support for GemFire.

Competition

From a data products perspective, Pivotal's core competition comes from established analytic database giants Oracle, IBM, Teradata and Microsoft, as well as SAP, all of which have also added Hadoop to their portfolios in recent years. In terms of matching the functionality of the Big Data Suite, IBM is probably the closest rival, although it does not offer a comparable single SKU. Pivotal's open source strategy will also be a differentiator in the long term and as such we would expect the closest competition for new customers to come from Hadoop distributors looking to go up the stack with advanced functionality and broader data platforms. Given the strategic alliance struck with Hortonworks, it is clear that Pivotal considers Cloudera its closest competitor, while we would also

expect it encounter MapR. In the wider cloud market, Pivotal vies with Amazon Web Services, as well as IBM, Microsoft, Oracle, SAP and Red Hat, among others.

SWOT Analysis

Strengths

Pivotal has already established itself as a key player in the cloud and data sectors thanks to mature technology and being given the backing of its parent companies to focus on innovation.

Opportunities

We believe that as enterprises begin to define their next-generation data management platforms, they will look to combine a variety of data-processing and storage methods and technologies, and Pivotal will be an attractive option – even more so now that the whole stack is open source.

Weaknesses

Any move to open-source a proprietary product has the potential to undermine existing revenue streams. The Pivotal team has done this before, but not at this scale.

Threats

The company faces competition from some of the biggest players in the industry, which have larger installed bases to leverage, as well as more nimble startups.

Reproduced by permission of The 451 Group; © 2015. This report was originally published within 451 Research's Market Insight Service. For additional information on 451 Research or to apply for trial access, go to: www.451research.com