

OVUM OPINION

Oracle mainstreams its Hadoop platform with Cloudera OEM deal

Reference Code: OI00201-002

Publication Date: January 2012

Author: Tony Baer

OVUM VIEW

Summary

Oracle's announcement of the availability of its new Big Data Appliance came with one surprise – the Hadoop distribution would in fact be an OEM of Cloudera's offering. The deal fleshes out Oracle's Hadoop distribution plan, which until now was a blank slate. Instead of a staging ground for Oracle Exadata, Oracle Big Data Appliance will be a full-fledged Hadoop platform. The Oracle deal is clearly a big win for Cloudera, coming on the heels of another OEM arrangement with Dell that bundles a reference Cloudera Hadoop distribution (but not the support subscription) and Cloudera Enterprise on specially equipped PowerEdge servers. Besides expanding Cloudera's market penetration, it reinforces Cloudera's strategy to differentiate the Hadoop stack from competing alternatives from IBM and EMC. Having a definitive stack will be critical to give third parties a consistent target, which will be one of the prerequisites for Hadoop to evolve to a mainstream enterprise analytic data platform.

Cloudera fills Oracle Hadoop's missing links

As Ovum stated when Oracle announced the Big Data Appliance in October 2011, Oracle's strategy was not necessarily to counter Cloudera, but to provide its own alternative to offerings from IBM and EMC. Oracle's initial announcement was vague as it had not yet lined up its Hadoop distribution. Based on an offering that was unclear about its support for different components of the



Apache Hadoop stack and heavy on utilities for extracting, loading, and transforming data from Hadoop to Oracle Exadata, we originally believed Oracle was positioning Hadoop primarily as an exploratory analytics and staging platform. Along with Hadoop, Oracle was planning to offer its own NoSQL database, a distributed key-value store based on an enhancement of BerkeleyDB, as an alternative to Hadoop HBase and other NoSQL solutions.

With Oracle's announcement of general availability of the Big Data Appliance came a surprise: that the Hadoop engine would be Cloudera's distribution and that the bundling would also include Cloudera Enterprise (the part that Cloudera charges for), with the exception that Oracle (not Cloudera) would be the primary point of support. Cloudera is not simply giving away a free distribution for market share; this is a business agreement with a revenue stream. Oracle in turn gets a full-fledged Hadoop platform that can run with any third-party tools designed for the Apache stack, and is not simply a staging area for Exadata.

In addition to the Big Data Appliance, Oracle is releasing Oracle Big Data Connectors, which is licensed separately at \$2,000 per processor for Oracle targets. Although Oracle had not disclosed pricing previously, its October announcement described the offering as part of a product stack that would bind its Hadoop implementation more closely to Oracle Exadata. Oracle's release groups these connectors as a separate bundle that is not Exadata specific. Instead, the tooling bundle supports extraction, loading, and transformation from Hadoop (from the Big Data appliance or any other Hadoop cluster) into Oracle Database (or other databases with separate licensing of Oracle Data Integrator) or other SQL targets, as well as dynamic SQL access from Oracle Database to data stored in HDFS.

Clearly, Oracle's release of Oracle Big Data Appliance signifies a full commitment to Hadoop as a first-class citizen of the Oracle data platform. Its price, \$450,000 for 216 CPU cores backed by 648TB of storage and the same Infiniband backplane used by Oracle Exadata and Oracle's other engineered systems, is definitely competitive.

Oracle strengthens Cloudera's play to define Hadoop

With a three-year head start in the market to package and deliver commercial support for Hadoop, Cloudera has played its cards well. Having drawn \$75m in four rounds of venture funding to date (including \$40m announced as recently as November 2011), Cloudera enters 2012 from a position of strength in its goal to assume the same role for Hadoop that Red Hat plays for Linux: to become the de facto standard commercial provider of the stack. Cloudera also has used its first-mover advantage in the field effectively. It currently has several dozen alliances, mostly with BI, data warehousing, data integration, and specialty analytic providers. Now it can add Oracle to the list.



Cloudera had the Hadoop management and support market to itself until 2011, when IBM, EMC, and Hortonworks (the rest of the Yahoo Hadoop team that originally did not join Cloudera) joined the fray. With new players came a splintering of the Hadoop technology platform. While Cloudera and Hortonworks are focusing on supporting a pure Apache stack, IBM and EMC have proprietary alternatives for specific Hadoop components. More to the point, it is a debate that is more than academic: the outcome will determine whether third-party tool providers will have to write to one or multiple Hadoop or Hadoop-like targets. Ultimately, third-party support will be one of the key prerequisites for Hadoop to become established as a mainstream enterprise analytics platform. Ovum expects that by early 2013, the market will have rendered its verdict regarding its expectations for a core Hadoop platform, and in turn for a definable target for third parties.

Oracle's entry prompts the question: scale up or scale out for Hadoop

Hadoop emerged as open source technology because it was highly scalable across clusters of commodity Intel x86 servers. A key ingredient of Yahoo's and Facebook's early successes with Hadoop was that they could simply keep scaling out their commodity clusters as Hadoop's scalability was extremely linear. Their Hadoop farms have grown to tens of thousands of servers.

From an infrastructure cost standpoint, expanding Hadoop was fairly economical. However, the price was the need for skills and tooling to configure and manage huge cluster farms. Although Internet firms tend to have deep engineering staffs that can design their way around clusters, that is not necessarily the case for enterprise IT. One solution is to rely on a cloud service provider to host, manage, and expand (or contract) the infrastructure – as long as the data in question is allowed to run off premise. The alternative solution is what Oracle, EMC, and others are proposing: run Hadoop in large appliances, which Oracle calls “engineered systems.” Although Oracle's engineered systems can be scaled in increments or fractions of a rack, adding capacity involves more hardware than just a few commodity servers here or there. The counter argument is that appliances reduce the cost of management.

For enterprise IT organizations, the question of appliance scale-up versus commodity scale-out boils down to a pay-now-or-pay-later strategy. We do not believe most enterprises have the expertise to run their own clusters.

APPENDIX

Disclaimer

All Rights Reserved.



No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, Ovum (an Informa business).

The facts of this report are believed to be correct at the time of publication but cannot be guaranteed. Please note that the findings, conclusions and recommendations that Ovum delivers will be based on information gathered in good faith from both primary and secondary sources, whose accuracy we are not always in a position to guarantee. As such Ovum can accept no liability whatever for actions taken based on any information that may subsequently prove to be incorrect.