SearchBusinessAnalytics.com

Experts sound off on 'big data' technology, Hadoop and data warehouses

By Nicole Laskowski

While Yahoo jumped into the "big data" spotlight yesterday with [its decision to spin off Hortonworks](#), a Hadoop support company, just last week a panel of vendors and consultants was speaking about the importance -- and pitfalls -- of [big data analysis and the tools](#) that go with it.

A panel of technology vendors, analysts and consultants speaking at [Enzee Universe 2011](#) held in Boston, suggested that [big data](#) technology now has a place in the [enterprise software toolbox](#). Harnessing data that's quickly multiplying in depth and breadth from unstructured and semi-structured sources, in addition to structured sources they're already are tapping into, really *can* make a difference.

Hadoop, MapReduce and the enterprise data warehouse alternatives

There's no shortage of technology vendors joining Yahoo in its efforts to [target big data](#). Apache Hadoop, an open source project in which Yahoo is the largest contributor, and Google MapReduce, software developed to analyze big data sets and a major component of the Hadoop framework, are becoming increasingly popular. Cloudera, a startup that offers Hadoop support, services and software will compete with Yahoo's Hortonworks. Additionally, EMC is creating a for-pay Hadoop offering based on technology from the startup MapR Technologies.

> When bright, shiny toys emerge on the horizon, we buy them and bring them back, and now we're starting to see the backlash around Hadoop.
>
> Shawn Rogers, vice president, Enterprise Management Associates

In some cases, [MapReduce and Hadoop](#) work well, providing a grid for large computational work, according to Usama Fayyad, CEO of the data strategy, technology and consulting firm Open Insights and former chief data officer and executive vice president of Yahoo's research and data strategy division. In other cases, they aren't needed, but businesses "overly enamored with the technology" are using them anyway. Fellow panelist Shawn Rogers added that the repercussions for broadly deploying Hadoop are beginning to surface.

"When bright, shiny toys emerge on the horizon, we buy them and bring them back, and now we're starting to see the backlash around Hadoop," said Rogers, a vice president at Enterprise Management Associates.

[Hadoop deployments](#) may not be entirely necessary. Enterprise data warehousing platforms with shared-nothing massively parallel processing architecture enable in-database analytics performance and high-performance data management, according to James Kobielus, a senior analyst from Forrester Research Inc. and an audience member during the panel discussion.

Kobielus, in his work on an upcoming Forrester report, asked early Hadoop adopters if they first "considered using the tried-and-true approach of a petabyte-scale EDW."

"Many of the case studies did, in fact, consider an EDW such as those from Teradata and Oracle," Kobielus wrote after the conference. "But they chose to build out their 'Big Data' initiatives on Hadoop for many good reasons. … By using Apache Hadoop, they could avoid paying expensive

software licenses, could give themselves flexibility to modify source code to meet their evolving needs, and could avail themselves of leading-edge innovations coming from the worldwide Hadoop community."

Kobielus agreed with Fayyad and Rogers' assertion that not all tasks require Hadoop and MapReduce, adding that although he believes Hadoop will eventually become the "pre-eminent scientific analytic platform," it currently lacks real-time integration and robust high availability traditionally found in an EDW.

Early Hadoop users identify specific big data tasks

While that may be the case, some businesses have etched out specific tasks for Hadoop software within their data management system toolbox.

Arup Ray, the director of data warehousing and business intelligence development and architecture for Intuit, said his company uses Hadoop as an extract, transfer and load (ETL) engine and for ad hoc analysis. Intuit is also leveraging analytics using Netezza technology.

The growing interest in big data technology

Read the Whatis.com definition of big data

Still confused? You'll want to read 'big data demystified'

Read more about how big data is a bigger deal than previously thought

'Big data' tops most annoying industry buzzwords list

Still other companies, such as T-Mobile have rejected going the route of Hadoop altogether.

"We talked about playing around with it, but we're really just happy to stay with Netezza," said Christine Twiford, the manager of network technology systems at T-Mobile. The telecommunications operator traded in its Oracle appliance about five years ago for Netezza, enabling data to be loaded 50% faster. Plus, data can now be explored without having to first know the question and queried in new ways, according to Twiford.

And, although the buzz around Hadoop and MapReduce continues to grow, T-Mobile isn't alone. In a recent SearchBusinessAnalytics.com survey of business and IT professionals, end users and consultants, only 1% said their current data warehousing architecture included Hadoop. Thirteen percent of respondents indicated they had plans to add Hadoop within the next year, which corresponds to the least anticipated growth when compared with other emerging technologies related to the data warehouse.

Regardless of how the tools are being marketed these days, the race to become the big data analytics wonder drug continues. The new IBM Netezza High Capacity Appliance was rolled out at last week's conference. With its ability to analyze 10 petabytes of data in minutes, there's no question that the new appliance, Netezza's first release since it was acquired by IBM last fall, is aimed at big data.

Is open source the wave of the future?

Even so, as Forrester's Kobielus pointed out during the question-and-answer portion of the panel discussion, open source tools, such as Hadoop and the programming language R, have successfully pushed at the boundaries of big data analytics.

"Does the future belong to open source for big data?" he asked panelists.

"Open source has done a good job of kicking the door open," Rogers of Enterprise Management Associates said. "But they come to market in a less mature way."

Rogers mentioned Hadoop, Pentaho and Jaspersoft as examples, adding that open source, which he proclaimed to be a fan of, moved at a different speed than proprietary options.

"It's a great way to explore new frontiers," Open Insights' Fayyad said. "But the ethos of open source cannot keep up with the demands of the mainstream."

29 Jun 2011