

June 4, 2014, 11:58 AM ET

Hadoop Hits the Big Time

By **Thomas H. Davenport**

I'm speaking this week at the Hadoop Summit in San Jose, and it's quite a revelation for me. As a speaker on the business value of Big Data, I don't normally get invited to events like this, which tend to be highly technical. This one is technical too, but I guess they wanted to have a token business-oriented perspective.

I had always thought of the Hadoop-oriented conferences (Strata and HadoopWorld are the other key members of the genre) as being mostly for Silicon Valley tech companies. But Hadoop, the open-source tool for processing data across a cluster of connected servers, is clearly going mainstream. At this event I came across people from Verizon Wireless, American Express Co., Bank of America Corp., [Safeway Inc.](#), AT&T Inc., [Comcast Corp.](#), AIG Inc., [Corning Inc.](#)...you get the picture. These are big, established organizations that found the way to San Jose to learn more about Hadoop. Their presence here among the 3,200 attendees suggests that Hadoop is rapidly becoming an enterprise data processing and management tool.

I'll tell you a key reason why in a moment, but first I also want to point out that the vendors playing in the Hadoop world are also increasingly blue-chip. On the exhibit floor were such enterprise software titans as [Microsoft Corp.](#), [IBM Corp.](#), [SAP AG](#), [Teradata Corp.](#), [EMC Corp.](#), [Informatica Corp.](#), and [Cisco Systems Inc.](#) Also present were commercial analytics firms like [SAS](#) and [MicroStrategy Inc.](#), who have versions of their software that run on Hadoop now. HortonWorks, the Hadoop distribution vendor that hosts Hadoop Summit, has taken a strong partnering approach, and the exhibit floor suggests that the strategy is working. Cloudera Inc., the other large Hadoop vendor, is taking more of a go-it-alone approach, though they and other Hadoop providers like MAPR were also present at this gathering.

A single factoid will explain why Hadoop is growing rapidly in popularity among large corporations. I went to one presentation by [TrueCar](#), a website that tracks vehicle prices. They said that their previous cost for storing a gigabyte of data (including hardware, software, and support) for a month in a data warehouse was \$19. Using Hadoop, they pay 23 cents a month per gigabyte. That two orders-of-magnitude cost differential has got to be appealing to a lot of CIOs. There are performance improvements as well, of course, although they weren't quite as dramatic as the cost reductions in the examples I heard.

The downside of adopting Hadoop is that you have to put up with a good deal more technical complexity, and a ton of mumbo-jumbo. For example, the cutesy names for open-source languages and tools get old quickly (though I find the toy elephant that is the mascot of Hadoop oddly endearing). I missed both the "Pig on Tez" and "Pig on Storm" talks. I did catch a bit of

the “Spark on YARN” presentation, though I confess to have understood very little of it. If you enter the Big Data world come prepared to speak an entirely different language.

Despite the linguistic and technical complexity, a lot of people were talking about greater integration of Hadoop with enterprise software of various types, particularly SQL. That would make it much easier to query Hadoop datasets for non-technical analysts. I expect we will see many pairings of Hadoop at the back end for cheap storage and fast processing, and familiar tools like Excel, Tableau, and SAS at the front end for analysis and user interface.

And while I didn’t hear many of these companies say they were eliminating their data warehouses, much of the growth in data management terabytes and petabytes will obviously be on Hadoop. Database and data warehouse companies like IBM, Oracle Corp., and Teradata will have to scramble to figure out how to add value to Hadoop in an enterprise context.

Overall, I get a feeling of having glimpsed the future of enterprise IT by coming to this event. Like it or not, that future will be more open source, more technically edgy, more analytical, and—if I am typical—more difficult for businesspeople to understand.

Thomas H. Davenport is a Distinguished Professor at Babson College, a Research Fellow at the Center for Digital Business, Director of Research at the International Institute for Analytics, and a Senior Advisor to Deloitte Analytics.

CORRECTION: The number of attendees at the Hadoop Summit in San Jose is closer to 3,200 attendees, not 3,700 as previously reported.