Search Innovation Insights | **Search**

- Sign In

WIRED
Innovation Insights
New Thinking for a New Era

New Thinking for a New Era

- Home
- About
- Blogs
- Forum
- Media
- Members

- All Blog Posts
- My Blog
- Add

Big Data Does Not Equal Big Memory

- Posted by
- View Blog

It's one of the big ironies of big data: For an industry defined by the ability to process diverse and multi-structured forms of data in varied ways, we seem remarkably myopic when it comes to swallowing hype. I'm talking about the kind you get around individual analytic components as "the next big thing" to singlehandedly manage big data. Like physicists pining for a "theory of everything" to explain the whole universe, it seems like we're fascinated by the prospect of a single solution or silver bullet to handle it all. Unfortunately, great innovations can easily be oversold and over applied amidst a super heated big data world. I have written other pieces on the danger of overselling and under delivering.

The Overselling of In-Memory

I'm starting to see overselling happening with in-memory, or big memory. To be sure, high density and lower cost memory options are a game-changing innovation for the data processing and integration industry.

Leveraging these newer technologies, software can sidestep relatively slow disk I/O and leverage lightning-fast performance of memory based storage devices. But even the best in memory systems can't be all things to all data. To start, while it's getting cheaper, in-memory is still two orders of magnitude more expensive than other storage media. There are also persistence issues, meaning if there's a power surge, you lose all your data unless you have another copy somewhere.

As with all good innovations, I think in-memory spurs the kind of excitement that can lead us to overestimate what the technology can or should be doing. InformationWeek's big data guru Doug Henschen suggested as much when -- in his recent comprehensive and clear-eyed survey of the 16 Top Big Data Analytics Platforms -- he called out one in-memory provider as "tone deaf, incessantly promoting 'real-time' whether there's a clear need for that level of performance or not."

Big memory is really good for running certain kinds of analytic regressions involving some pieces of big data, but these specific applications don't automatically mean in-memory is your all-in-one big data solution. That's because all data is not created equal. Some data you need to store, but not process a lot. Some data you need to store, but not process in real-time. Some data is so hot you need to process it in real-time streams. So there are different categories of use cases, and in-memory is one of the tools you should have in your toolbox, but not the only tool.

And, that's the point. Big data doesn't equal just big memory; it doesn't equal just Hadoop; and it doesn't equal just relational databases. This is not a one size fits all kind of environment!

The Era of Liquid Analytics

Those who embrace the silver bullet trap are confusing the architecture with the desired experience. Think of an automobile: the engine, drive train, suspension and many other components operate seamlessly out of sight in order to provide you with an efficient, hassle free experience getting from one place to another. You shouldn't have to think about everything that's going on under the hood, but that complexity is essential to keep you moving. We ought to consider analytics the same way, but we usually don't. So sticking with our auto analogy: how do we square the seamless experience with the needed complexity under the hood? Try a "liquid" approach to data.

Liquid analytics involves algorithms and systems that automatically channel the appropriate data to the right part of your analytic architecture. Think of the many ways you can slice data along multiple axes: hot or cold multi-temperature data; real time vs. historical data; high value vs. low value; decision-oriented data for business users vs. discovery-oriented information for data artists. Good liquid analytics "just works" to make the appropriate data available to the questioner, no matter where it sits.

For industry players, a certain leap of faith is required to adopt the "best of breed", open source approach that's required for seamless operations across multiple analytic platforms. As Jeff Kelly says on Wikibon , no single analytic tool can handle everything. One of his hypotheticals, for instance, positions SAP, Teradata and other vendors in a larger context. "SAP HANA is not a platform for loading, processing, and analyzing huge volumes -- petabytes or more -- of unstructured data, commonly referred to as big data. Therefore, HANA is not suited for social networking and social media data analytics," Kelly writes. "For such uses cases, enterprises are better off looking to open-source big-data approaches such as Apache Hadoop or LexisNexis HPCC Systems, or even MPP-based next generation data warehousing appliances like EMC Greenplum or Teradata AsterData."

I think the message is clear: Liquid analytics and a "best of breed" mindset are needed for big data to thrive and start delivering real ROI. Every company should be working to build the cooperative spirit and the intelligent and integrated systems needed for this approach to work.

████████████████████████

Views: 660

Tags: [Analytics](), [Big](), [Data](), [Gnau](), [In-Memory](), [Intelligent](), [Liquid](), [Memory](), [Scott](), [Teradata]()

[0 members favorited this]()

Share   Tweet   G+

Like 0

- [< Previous Post]()

Comment

You need to be a member of Innovation Insights to add comments!

Welcome to
Innovation Insights

Sign In

Members