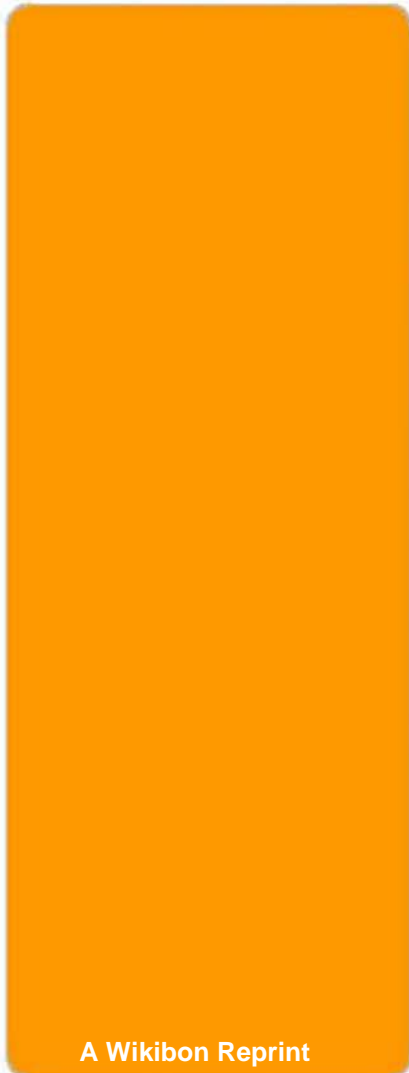




Wikibon.org

In-Memory Databases Put the Action in Actionable Insights

Jeff Kelly



View the [live research note on Wikibon](#).

The “Big” in Big Data tends to get a lot of attention, and for good reason. Data volumes are increasing exponentially year-over-year, with Google Chairman Eric Schmidt famously declaring that humanity creates in two days the same amount of information created from the dawn of civilization up through 2003. Another prediction is that total data created per year will hit 44 zettabytes in 2020.

Whether these estimates are precise or not, it’s clear that data volumes are exploding and traditional databases and data management tools are not up to the job of storing and processing it. Enter Hadoop, the open source software framework for storing and processing huge volumes of data across distributed commodity machines. In just a few short years, Hadoop has evolved from a niche tool for Web 2.0 companies to one of the foundational technologies in the emerging modern data architecture.

But exploding data volumes and new methods for storing and processing Big Data are just one part of a larger story focused on business value. There is no value in storing and processing large volumes of data for its own sake. In order to derive business value from Big Data, practitioners must also have means to quickly (as in sub-milliseconds) analyze data, derive actionable insights from that analysis, and execute the recommended actions.

The last part of this equation, action, is critical to generating business value from Big Data. It doesn’t do a retailer much good to determine through Big Data analysis that it can upsell Customer X to more profitable products and services, for example, if the retailer can’t then operationalize that insight and actually take the actions necessary (targeted advertisements, for example) to make it happen.

While Hadoop is ideal for storing and processing large volumes of data affordably, it is less suited to this type of real-time operational analytics, or Inline Analytics, workload. For these types of workloads, a different style of computing is required. The answer is in-memory databases.

In-Memory Database	Vendor	Type
Aerospike	Aerospike	Hybrid DRAM/SSD/disk
DB2 with BLU Acceleration	IBM	Hybrid DRAM/SSD/disk
GemFire	Pivotal	Fully in-memory
Oracle Database In-Memory	Oracle	Fully in-memory
SAP HANA	SAP	Fully in-memory

Figure 1: Partial list of In-memory databases on the market as of December 2014 (Source: Wikibon)

In-memory databases differ from traditional database systems (and Hadoop for that matter) in that they store data in dynamic random-access memory (DRAM) rather than on rotating disks. Reading and writing data to DRAM is dramatically faster than reading and writing data to disk, meaning in-memory databases are well positioned to support data-centric applications with high performance requirements.

In-memory databases aren't necessarily new, but they are increasingly accessible to mainstream enterprises. That's because the price of memory has gone down dramatically over the last decade-plus to the point where the price/performance ratio of in-memory databases makes them economically feasible for use-cases when the benefit of extremely fast performance outweighs the costs of storing data completely in DRAM. New database design approaches - including columnar architecture, data skipping capabilities and high-rates of data compression – add to the performance capabilities of many in-memory databases.

Enterprises and start-ups alike are starting to take notice. One of the first industries to deploy in-memory database technology at scale is the ad-tech business. Anyone who's been online has probably interacted with ad-tech. When a user logs onto a commercial website, that user is nearly-instantaneously presented with any number of advertisements, each highly-targeted to the user based on his or her browsing behavior, geo-location and other relevant data sources. Under the covers, an auction has just taken place among advertisers, who bid to place an ad in front of that user based on his or her perceived value. All this takes place thanks to lightning-fast analytics made possible, in part, by in-memory databases that support Inline Analytics.

Ad-tech is just one of many use cases that in-memory databases support. Any operational business process that requires sub-millisecond analytic performance is a candidate. These include Industrial Internet use cases such as automating the optimization of industrial equipment performance based on real-time environmental, mechanical and market conditions. On the commercial side, any consumer mobile app that is monetized through targeted advertisements and

personalized, time-sensitive marketing offers are possible in-memory database candidates.

Beyond automating intelligence actions, in-memory databases also support analytic applications that serve human end-users, providing near-instantaneous query responses. Business analysts, for example, may use analytic applications supported by in-memory databases to crunch large volumes of sales, accounting or other financial data. Marketing analysts similarly can benefit from near-real-time query responses when analyzing marketing campaign data. In these cases, the fast query response times enabled by in-memory databases allow end-users to maintain their line of thought, getting answers to queries and then iterate again and again.

Database vendors have taken two approaches to in-memory database design. Some, such as SAP, have developed in-memory databases that exclusively store data in DRAM providing significant performance advantages versus disk-based databases. An example is HANA, SAP's all-in-memory database the company developed from scratch to support both its traditional ERP and new analytic applications.

Others, such as IBM, take a hybrid approach to in-memory databases, storing the most frequently accessed data in DRAM and other data on SSD or disk. This hybrid approach, when architected correctly, allows practitioners the best of both worlds: the performance of a fully in-memory database at a price point closer to that of disk-based databases. Practitioners can make use of data in DRAM for the most performance-intensive applications but leverage less expensive disk for other applications. IBM's BLU Acceleration for DB2 is an example of an in-memory database that takes this approach.

In the real world, in-memory databases and Big Data platforms such as Hadoop play complementary roles. Data Scientists at enterprises across verticals are increasingly turning to the latter to perform Deep Data Analytics, scouring huge volumes of historical data from multiple sources to discover hidden patterns and insights that can impact the business and building predictive models based on those insights. Those models are then deployed and operationalized, enabling real-time Inline Analytics, using in-memory database which support intelligent automation as well as end-user analytic applications.

Action Item: Obviously, Deep Data Analytics is an important element of what is collectively called Big Data. But Inline Analytics, often supported by in-memory databases, is often where the "action" occurs, where transactions take place and where the real money is made. Enterprise Big Data practitioners must bring these two sides of the equation together to get full value from all that Big Data they are collecting and storing.