5 Ways Big Data Disrupts Your Existing Data Warehouse

by Avi Kalderon

JULY 23, 2015

While the enterprise data warehouse has become a staple for stability and accuracy in delivering corporations mission-critical data, big data has disrupted the status quo with data lakes and other approaches that increase the agility and fidelity of the data an enterprise consumes. We call this approach "Data-First" and it's changing the way businesses think about their existing data warehouses.

1. Constraints of Today's Data Warehouses

Traditional data warehouses are typically rooted in relational database technology (a.k.a. RDBMS). Warehouses were created in order to optimize complex workloads churning large quantities of data by machines that are expensive to procure and operate. The relational model was originally created in order to optimize performance and storage costs. The techniques applied during the modeling of the warehouse were geared towards creating structures that stage data for specific, well-known reports and dashboards. The data needs to be organized for quick retrieval, so relational schemas were, and still usually are, a great design. But the increasing demand for data cut in many different ways has revealed limits to the warehouse's ability to respond quickly as these would present continuous pervasive changes to its design.

2. Brute Force Number-Crunching

Big data technology (a.k.a. NoSQL) tends to circumvent these limitations with brute force, employing a large number of machines that each crunch numbers independently of the others. Using commodity hardware and largely open-source software, enterprises achieve a higher scale of compute and storage capability at a lower price point. Cost can drop further as some payloads absolve data modelers and engineers from optimizing the data structures upfront.

Instead, the scale and better price point open the door to deliver results from queries not known ahead of time, allowing analysts to retrieve answers without suffering through the lag of rigorous requirements that feed the engineering process to change the design of the warehouse. Much like a search engine, big data platforms create virtual structures on top of your data at the time of inquiry, permitting richer, more agile and nimbler analysis. We call this the "Data-First" approach. Keep your data in its original form and apply structures when needed instead of sorting and storing it physically in specific models and relationships ahead of time. In other words – worry more about the data and less about how it is stored.

3. Data Management is Key

Neither of the two approaches (RDBMS and NoSQL) relieves an organization from meticulously governing their data by classifying, defining, profiling and understanding the semantics that surround their specific content. *Data-first does not mean no data management*. It is just that the layout of the data and the payloads possible through the big data paradigm allow users to explore and analyze faster than before. Data Management is still a necessity.

4. Unleashing the Power of the Machine to Learn Your Data

With the advent of machine learning and pattern recognition algorithms, you can use your big data platform compute resources to learn your content and propose virtual structures for different domains of data. For example: finding similarities in patterns of data between data sources would allow your database to propose that two columns from two differently-named systems do contain the same type of information. SSN = SOC_SEC = PERSON_ID in three different sources if the data pattern = 999-99-9999.

By harnessing the power of the machines correlations can be made much faster accelerating the task of data modeling and facilitating data governance activities. That's just one example of how some of the new paradigms enable faster time to deliverable. We saw a similar evolution when Oracle PL_SQL was used by smart programmers to generate dynamic data-driven PL_SQL programs on the fly. This was the genesis of Metadata, and with the various algorithms and machine learning capabilities available today, we can see big data technology taking this model a huge step forward.

5. Data as the Source of Design

It is reasonable to assert that these capabilities can be achieved rather easily in todays' RDBMS- based databases and that's true. The relational warehouse still dominates the enterprise and is quite capable. But the power of big data technology is how it treats the data as the source of the design rather constraining it to rigid structures.

That is how the 'Data-First' approach truly shortens the journey to valuable business insights, which is already proving to accelerate enterprises' time-to-decision.

Avi Kalderon is a partner and practice leader for Big Data and Analytics with NewVantage Partners where he heads up NewVantage Partners' [Big Data Fast Track](#) program, the result of three years of Big Data strategy and execution with more than a dozen Fortune 1000 and industry leading clients. He has extensive experience as a business leader and technology executive, most recently as senior VP of the architecture and advanced technology group for FINRA (Financial Services Regulatory Authority), the largest independent securities regulator in the US.

For further information, contact Randy Bean, Managing Partner, NewVantage Partners at rbean@newvantage.com.