IBM’s dashDB forges data warehouse in the cloud

Data lake meets warehouse in hybrid data architectures

As the Hadoop data lake gains more definition and deployments, it’s beginning to look like something that will coexist with existing data warehouse technology. This new view of hybrid data architectures has implications for data design, skills and planning. In addition, although Amazon’s Redshift led the way in cloud data warehouses, now IBM hopes to catch up on the wings of dashDB.
IBM'S DASHDB FORGES DATA WAREHOUSE IN THE CLOUD

Jack Vaughan, Senior News Writer

Amazon’s Redshift led the way in cloud data warehouses. Now IBM hopes to catch up on the wings of dashDB.
Each quarter, the editors at SearchDataManagement recognize a data management technology for innovation and market impact. The product selected this quarter is IBM’s dashDB.

PRODUCT: IBM DASHDB
Release date: IBM dashDB is a software service that provides frequent, often monthly, releases. The service does not adhere to any numerically designated release schedule.

WHAT IT DOES
IBM’s cloud-based data warehouse service, dashDB, integrates an in-memory version of DB2 with BLU Acceleration, row-based Netezza analytics technology, and connectors to IBM’s Cloudant NoSQL database and other analytics
WHY IT MATTERS

Elastically scalable cloud data services have risen in importance in recent years in response to applications generating large amounts of data, sometimes in hard-to-anticipate spurts. The cloud has become a natural home to such data, because so much of that data is created via Web and mobile applications hosted on the cloud. The cloud database also can offer economic benefits, because it reduces an organization’s need to build and maintain on-premises data centers. IBM’s entry acts as a cloud hub for several of its key analytics offerings. Last fall, IBM added massively parallel processing and R programming language support to a still growing list of dashDB enhancements. In March, dashDB added support for mail-in drive data migrations and node-level high availability.

WHAT USERS SAY

IBM dashDB’s ability to run data models in-memory in the database provides valuable performance dividends over other cloud approaches, according to Shiv Sehgal, a solutions architect at RSG Media in New York City. He said he and his colleagues have begun to implement dashDB together with IBM's...
Cloudant NoSQL database to analyze unstructured data related to Web users’ behavior.

RSG offers software and services that help cable, entertainment and other companies efficiently deliver advertising and content, and a variety of log and social data are now part of that mix. Sehgal said dashDB’s connectors to NoSQL stores and diverse analytical tools simplify integration development efforts. He indicated that support for R and REST interfaces would enable RSG Media to create customer dashboards that allowed users to do “what-if” analysis from their own desktops with developer intercession.

**DRILLDOWN**

- Predictive modeling algorithms built into the database include linear regression, decision tree clustering, k-means clustering and Esri-compatible geospatial extensions.

- Supports encryption of data at rest, as well as monitoring dashboards that display database performance.
IBM has been enhancing dashDB via a series of guides that show ways to use this cloud data warehouse along with a variety of other systems it supports. This includes Watson Analytics, Spark, SPSS and Cognos Business Intelligence.

**PRICING**

Pricing for dashDB, as hosted on IBM’s Bluemix platform as a service, ranges from an entry level plan with no charge for up to 1 GB of data storage or $50 per month for up to 20 GB of data storage to an enterprise plan priced at $7,370 per instance for dedicated instances with 256 GB for “storage-dense” applications.
DATA LAKE MEETS WAREHOUSE IN HYBRID DATA ARCHITECTURES

Jack Vaughan, Senior News Writer

A new view on hybrid data architectures, in which data lakes and warehouses coexist, emerged at EDW 2016. The hybrid approach has implications for data design, skills and planning.

As the Hadoop data lake gains more definition and deployments, it’s beginning to look like something that will coexist with existing data warehouse technology. Such a view of hybrid data architectures emerged in sessions at the Enterprise Data World 2016 conference in San Diego, Calif.

“It’s not an ‘all or nothing’ thing. It’s a ‘both’ thing,” consultant Joe Caserta told EDW 2016 attendees. “The enterprise data warehouse will not go away. Even when we are doing Hadoop and Spark and all the other shiny new things, it is still there.”

But data lakes are finding a place in big data science and analytics applications. Caserta, president and CEO of Caserta Concepts in New York, said Hadoop-based data lakes are typically built first of all to handle large and quickly
arriving volumes of unstructured data. The data lake is a key part of big data trends that will bring change to data professionals’ familiar practices, according to Caserta and others.

“What we used to do with data warehouses was first to create data models, but that has changed,” Caserta said. With data lakes, the models come after the fact. “We don’t do it right away anymore,” he said.

**ANALYTICS AND APPLICATIONS**

One reason for that is the data lake’s association with real-time data streaming. As analytics become more closely tied to operational applications, and part of real-time decision making, data is required to be accessible as soon as it’s created, Caserta said. That, too, makes it very different from data warehouse work, which continues to be the foundation for necessary business reports.

This view was shared by Tom Place, director of data management at payment processing, retail data security and e-commerce services provider First Data Corp. He sees a distinction between uses for the data lake and data warehouse, as well as a need for both in data architectures.

“The data warehouse is really designed for slowly changing data -- daily summaries, weekly summaries and monthly summaries of known, structured...
data,” Place said. “On the other hand, the data lake is being designed for quickly changing data -- data that tells you what happened one minute ago or five minutes ago.”

Like Caserta, Place is seeing selective rollups of unstructured data from the lake going into the warehouse.

**DATA RESERVOIR DAYS**

As data lakes evolve, their days as a simple, undifferentiated refuge for data may be nearing an end. Caserta and Place both see different degrees of data governance being applied to different levels of data in the data lake.

The divisions are based on the purposes -- and skills -- of advanced analytics users. For Place, data consumers at Atlanta-based First Data comprise business analysts and data scientists, but also specialists in product innovation and product refinement. Example applications range from business reporting to fraud prevention.

Place said he actually prefers the term data reservoir to data lake. In his view, a reservoir conveys the idea that ingested data will be worked on.

“A data lake itself is just a collection of raw data that you don’t understand. It can be something you can’t manage and you can’t validate for your users,” he
said. “With a reservoir, that data becomes well governed, well understood and well managed. And, you can actually do more valuable things with the data.”

UP FROM THE SANDBOX

As a term, data lake is far from universally welcome. It’s not a favorite of Luminita Vollmer, senior IT architect for data and business intelligence delivery at Thrivent Financial, an insurance and investment management company in Minneapolis. She told an EDW 2016 crowd she preferred the common development term sandbox, because much of the data lake’s use is experimental. Still, in a session on the prospects of data warehousing, she told participants to look at their present data warehouse with a view toward how their organizations will use tools of the future, including NoSQL databases and predictive analytics software. Hadoop, she said, has already found a place in the data architectures of many organizations.

Like others, Vollmer said that a new spectrum of data analytics users is emerging. Things are different than they were when the enterprise data warehouse was the only game, she said, and that will affect the way data management teams are organized going forward.

“You have to have some people that support present systems and some
people doing some research,” Vollmer said. “That is a change in the way we do things.”
IBM’s dashDB forges data warehouse in the cloud

Data lake meets warehouse in hybrid data architectures

FREE RESOURCES FOR TECHNOLOGY PROFESSIONALS
TechTarget publishes targeted technology media that address your need for information and resources for researching products, developing strategy and making cost-effective purchase decisions. Our network of technology-specific Web sites gives you access to industry experts, independent content and analysis and the Web’s largest library of vendor-provided white papers, webcasts, podcasts, videos, virtual trade shows, research reports and more—drawing on the rich R&D resources of technology providers to address market trends, challenges and solutions. Our live events and virtual seminars give you access to vendor neutral, expert commentary and advice on the issues and challenges you face daily. Our social community IT Knowledge Exchange allows you to share real world information in real time with peers and experts.

WHAT MAKES TECHTARGET UNIQUE?
TechTarget is squarely focused on the enterprise IT space. Our team of editors and network of industry experts provide the richest, most relevant content to IT professionals and management. We leverage the immediacy of the Web, the networking and face-to-face opportunities of events and virtual events, and the ability to interact with peers—all to create compelling and actionable information for enterprise IT professionals across all industries and markets.