NOSQL DATABASES CONTINUE TO EXPAND AND CO-EXIST WITH SQL

TechTarget® | SearchData Management

F

OR COMPANIES WITH growing amounts of unstructured data, NoSQL databases are becoming integral parts of day-to-day business. However, SQL still maintains an important presence. In this expert e-guide, explore the challenges that NoSQL databases are rising to meet, and how the latest developments in NoSQL database technology are interweaving with traditional relational databases.

SPONSORED BY *rackspace.*
the #1 managed cloud company

NOSQL DATABASES DENT RELATIONAL SOFTWARE'S DATA PROCESSING DOMINANCE

Jack Vaughn

In 1922, automaker Henry Ford famously wrote that his customers could have a car painted any color they wanted -- as long as it was black. Until recently, IT managers, application developers and business executives faced similarly limited choices in selecting database technologies. Relational databases built on top of the SQL programming language were the dominant engines powering corporate IT and business systems, with no real challengers in sight.

But things have changed. Starting in the mid-2000s, SQL's absolute supremacy was undone by the likes of Yahoo, Google, Facebook, Amazon.com and eBay. At those Internet giants and other companies, the need to run colossally scalable Web applications with varied and fast-changing data requirements prompted efforts to find alternatives to mainstream relational databases. That ushered in first a stream, and over the past few years a torrent, of new technologies that eschewed rigid SQL development principles in favor of more flexible and scalable data designs. Those databases are spread across several distinct

SPONSORED BY

product categories based on different data models. But they share a pithy um-brella term with a stake-in-the-ground sound: NoSQL.

All in the NoSQL Family

NoSQL databases are geared toward managing large sets of varied and frequently updated data, often in distrib-uted systems or the cloud. They avoid the rigid schemas associated with relational databases. But the architec-tures themselves vary and are separated into four primary classifications.

Document databases
Store data elements in document-like structures that encode information in formats such as JSON. Common uses include content management and monitoring Web and mobile applications.
EXAMPLES: Couchbase Server, CouchDB, MarkLogic, MongoDB

Graph databases
Emphasize connections between data elements, storing related "nodes" in graphs to accelerate querying. Common uses include recommendation engines and geospatial applications.
EXAMPLES: InfiniteGraph, Neo4j

Key-value databases
Use a simple data model that pairs a unique key and its associated value in storing data elements. Common uses include storing clickstream data and application logs.
EXAMPLES: Aerospike, DynamoDB, Redis, Riak

Wide column stores
Also called table-style databases—store data across tables that can have very large numbers of columns. Common uses include Internet search and other large-scale Web applications.
EXAMPLES: Accumulo, Cassandra, HBase, Hypertable, SimpleDB

The truth is, though, that the NoSQL movement isn't really an up-against-the-wall revolution seeking to eradicate relational databases. Yes, some NoSQL vendors do talk like that's their ultimate goal. But the term NoSQLhas been softened to also mean "not only SQL," in recognition of the fact that many of the databases do incorporate some elements of SQL. More substantively, NoSQL technologies aren't positioned as wholesale replacements for relational soft-ware -- they tend to be built for specific uses, usually involving large data sets that need to be accessed and updated frequently. And that's how things are play-ing out on the ground thus far: NoSQL databases have become must-have items for companies with fast-growing vaults of Web, social media, demographic and machine data, but often they're sharing data processing and analysis workloads with SQL-based software.

For example, Crittercism Inc. is a startup that helps organizations monitor the performance of their mobile applications, based on real-time data collected from more than 800 million mobile devices. In application performance man-agement parlance, a user interaction with an app is called a request; Crittercism pulls in information about more than 30,000 requests per second, a rate that adds up to nearly 3 billion a day. That has created a pool of more than 20 tera-bytes of data -- and the total only keeps growing, said Lars Kamp, vice president

SPONSORED BY

of business development at the San Francisco company.

Included in the mix is data on application errors, crash diagnostics and what Crittercism calls "network breadcrumbs" documenting the trail of network calls and other processing events leading up to app problems. That data "is very unstructured and non-uniform, and varies widely from customer to customer and application to application," said Mike Chesnut, the company's director of operations engineering.

MEETING THE OLD WAY HALFWAY

The sheer amount of information involved, and its variable nature, mandated a fresh approach to formatting the data. Using relational software would have required substantial processing overhead to maintain a database schema that could accommodate all of the information, plus frequent downtime for making changes to the schema, Chesnut said; he added that the company had to be able to modify how it collects and stores data "on the fly, often several times a day." Kamp was even blunter: "Crittercism as a company would not have been possible 10 years ago," when SQL was the only choice, he said.

Enter MongoDB, a NoSQL database running on the Amazon Web Services cloud. Like other NoSQL technologies, it offered schema design flexibility. That

made it possible for Crittercism to store the error and crash data in a single "collection" -- the MongoDB equivalent of a relational table -- without imposing a strict schema on the information. In turn, the lack of a fixed data structure with uniform fields has enabled the company's performance management service to "evolve organically" to meet the needs of different customers, Chesnut said.

Crittercism also uses Amazon.com's DynamoDB NoSQL database to store data on a specific request path that requires particularly fast performance, according to Chesnut. But there's SQL in the company's database architecture, too. A PostgreSQL open source database holds highly relational operations data, and all of the information is summarized in a SQL-based Amazon Redshift data warehouse for analysis and reporting. Chesnut and his colleagues aren't NoSQL purists: "We're very engaged with exploring any and all technology offerings that can help us solve our problems and better serve our customers," he said.

Recent surveys show that NoSQL databases are making inroads with big data users -- but overall, adoption is still relatively low. For example, TechTarget's 2013 Analytics & Data Warehousing Reader Survey found that 21% of 222 respondents with active or in-the-works big data programs were using or planning to deploy NoSQL systems as part of the efforts. Another survey conducted
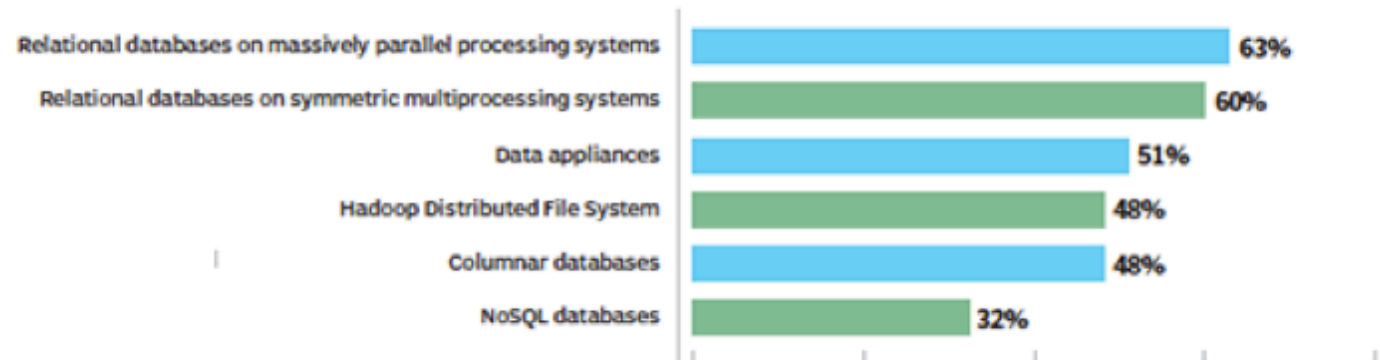
SPONSORED BY **rackspace.**
the #1 managed cloud company

last year by Enterprise Management Associates Inc. and 9sight Consulting produced an almost identical result: In that case, 22% of the 259 respondents said they had NoSQL platforms in place. In a third survey, done by The Data Warehousing Institute, 32% of 189 respondents said their organizations were using NoSQL software. Even there, though, NoSQL technology was last on the adoption list, trailing behind relational databases, data appliances, columnar software and big-data fellow traveler Hadoop (see Figure 1).

What's in Your Big Data Environment?
Percentages of organizations using various data platforms for big data management, in order of prevalence:

Platform	Percentage
Relational databases on massively parallel processing systems	63%
Relational databases on symmetric multiprocessing systems	60%
Data appliances	51%
Hadoop Distributed File System	48%
Columnar databases	48%
NoSQL databases	32%

SOURCE: THE DATA WAREHOUSING INSTITUTE'S MANAGING BIG DATA; BASED ON RESPONSES FROM 189 IT PROFESSIONALS, CONSULTANTS AND BUSINESS USERS WITH BIG DATA MANAGEMENT EXPERIENCE, FROM AN ONLINE SURVEY CONDUCTED IN 2013

Greater penetration of data centers is expected going forward: Analyst

SPONSORED BY *rackspace.* the #1 managed cloud company

group Wikibon forecast last year that worldwide revenue for NoSQL software and services would grow from $286 million in 2012 to $1.825 billion in 2017. And venture capitalists are betting big on that kind of growth.MongoDB Inc., which leads the development of its namesake database, raised $150 million in new funding last fall. That came shortly after $45 million and $25 million funding rounds by DataStax Inc. and Couchbase Inc., two other NoSQL vendors.

RELATIONAL PLAYERS HIT FROM BOTH SIDES

Even the big relational database vendors have gotten into the NoSQL game. Oracle introduced a NoSQL database in late 2011 and was one of the lead sponsors of last year's NoSQL Now! conference; Oracle representatives gave a keynote speech and led two technical sessions at the event. Last June, IBM added support for MongoDB's application programming interface to its DB2 relational database, enabling users to store data there in the JavaScript Object Notation (JSON) format. DB2 can also handle graph and XML data, and IBM in March acquired Cloudant Inc., a NoSQL vendor that runs a hosted version of the JSON-based CouchDB database. Microsoft offers a NoSQL data store as part of its Windows Azure cloud platform.

Application-driven data needs and the growing move toward cloud

computing are creating a wider opening for NoSQL methods, said Carl Olof-son, a database analyst at market research company IDC. For IT managers and business executives, though, he compared buying into NoSQL with investing in a new stock that doesn't have a lot of market history.

"Most of the NoSQL databases are new. They still need to be battle tested," Olofson said. "If you're constantly changing data definitions and you can't change your relational database fast enough, you might look at NoSQL. But there is risk."

For one thing, NoSQL technologies typically don't provide full ACID capa-bilities -- atomicity, consistency, isolation and durability -- for guaranteeing transaction integrity, as relational databases do. In addition, they often lack enterprise-class services in areas such as disaster recovery, security and data quality, according to Olofson. Like other analysts, he also expects a whittling of the well-populated ranks of NoSQL vendors as the market matures.

"NoSQL databases are really good for handling XML and JSON data, which includes a lot of things Java developers are working on these days," said Wayne Eckerson, a TechTarget industry analyst and president of consultancy Eck-erson Group Inc. In particular, they're well suited to high-performance Web applications "with a high volume of reads and writes," Eckerson said. But, he

added, they aren't such a good fit for "long-running queries" and other complex analytics jobs.

NOSQL SOFTWARE PROVIDES SPEED BOOST

That maps to the database architecture at Exelate, a marketing data services and technology provider that uses a diverse range of tools to supply information on household demographics and purchases to online advertisers and publishers. "Data is what we do," said Elad Efraim, co-founder and chief technology officer at the New York company. That makes performance paramount, he added. And while Exelate didn't start out with NoSQL technology when it was founded seven years ago, the need for speed eventually led Efraim and his team to deploy Aerospike, an in-memory NoSQL database that has helped scale the company's infrastructure to rapidly handle as many as one trillion real-time data transactions a month.

Aerospike provides a high-performance repository for data on the user session activity of website visitors that is constantly being updated, Efraim said. "We're talking about a large-scale system with a very high capacity of reads and writes that have to complete in some milliseconds. It's very important for us to make sure we can access the data in a way so that it can be made available [to

our customers] for decision making."

The database runs on servers at four fully replicated data centers world-wide, indexing everything to memory and holding it in the server cluster for further processing. From there, the data can be mined and correlated to other information in analytics and back-office systems. To make that happen, though, Exelate's applications don't solely use NoSQL software. One layer above the Aerospike repository is a "pretty standard" MySQL relational database that lets customers aggregate data, Efraim said. The company also uses an IBM Netezza appliance and relational database as a data warehouse for analytics uses.

To put things in Henry Ford's terms, users like Exelate and Crittercism no longer have to limit themselves to basic-black relational databases -- and they're taking advantage of NoSQL's new color choices to drive applications that mainstream relational software isn't suited for. But SQL black isn't going completely out of style with IT shoppers. For now, the two technologies are likely to share space in database garages.

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.