How will the database incumbents respond to NoSQL and NewSQL?

Analyst: Matt Aslett

The acquisition of **MySQL AB** by **Sun Microsystems** in January 2008 appeared to signal that open source databases were on the brink of opening up a new battleground against the proprietary database giants. In announcing the deal, Sun signaled its intention to provide the support and development resources required for MySQL to challenge the established vendors in supporting mission-critical, high-performance applications on Web-based architectures. Needless to say, reality was somewhat different as Sun faced wider problems of its own and eventually succumbed to takeout by **Oracle** (Nasdaq: ORCL) in April 2009, in doing so handing ownership of the leading commercial open source database to the database heavyweight.

We had previously argued that MySQL was very much the crown jewel of the open source database world thanks to its focus on Web applications, its lightweight architecture and its fast read capabilities, which made it potentially complementary technology for all of the established database players. Additionally, if Oracle's major rivals were seeking an obvious alternative to MySQL in 2009, they were out of luck.

Just two years later, however, the database market is awash with open source databases with lightweight architectures targeted at Web applications. Not only have the likes of **Monty Program** and **SkySQL** emerged to provide alternative support for MySQL and its forks, but there are also a large number of products available under the banner of NoSQL, which emerged in mid-2009 as an umbrella term for a loosely affiliated collection of non-relational database projects. We have also seen the emergence of what we have termed 'NewSQL' database offerings, with companies promising to deliver the scalability and flexibility promised by NoSQL while retaining the support for SQL queries and/or ACID (atomicity, consistency, isolation and durability), or to improve performance for appropriate workloads to the extent that the advanced scalability promised by some NoSQL databases becomes irrelevant.

From MySQL to NoSQL

Despite being a good match for many read-intensive applications, MySQL does not provide predictable performance at scale, particularly with a few writes thrown into the mix. The memcached distributed memory object-caching system can be used – and has been widely adopted – to improve performance but does not provide any persistence and lacks consistency. To some extent, the rise of NoSQL has been driven by the inadequacies of

MySQL to perform at scale without manual intervention, even with the performance boost added by memcached. This is most obviously true in the case of **Schooner Information Technology's** Membrain and **Couchbase's** Membase projects, both of which add persistence capabilities to turn memcached into a key value store. **RethinkDB** is working on similar capabilities.

However, we also see other NoSQL databases such as **Basho Technologies'** Riak, **10gen's** MongoDB and **DataStax**-supported Apache Cassandra being deployed for applications that previously ran on MySQL, or for which MySQL would previously have been the natural choice. For example, while **Facebook** continues to run its core applications on MySQL using the InnoDB storage engine and memcached, it found that MySQL was unable to deliver the performance required for large data sets, and created what became Apache Cassandra to power its inbox search and selected Apache HBase for its Messages application, which was updated in late 2010 to combine chat, email and SMS. Similarly, content discovery service **StumbleUpon** adopted HBase following problems with MySQL failover, **Digg** replaced its MySQL cluster with Apache Cassandra, and **Wordnik** replaced MySQL with MongoDB.

Of course, not every MySQL application is suitable for a NoSQL database, and it is worth noting that like Facebook, many major NoSQL users also continue to use MySQL, including **Twitter**, which backtracked on a planned migration of its core status table to Apache Cassandra in 2010. Twitter continues to use MySQL, but is adopting Apache Cassandra for newer projects.

The emergence of NewSQL

MySQL deployments are also a primary target for the new breed of relational database vendors promising to improve the performance and scalability of MySQL. Examples include **Tokutek** and **ScaleDB**, which act as storage engines for MySQL to improve its performance and scalability, respectively; and **CodeFutures** and **ScaleBase**, which both offer products to manage the sharding of data across multiple databases while maintaining compatibility with existing applications and tools. We must also mention **Continuent's** database-clustering and replication technology.

Additionally, Schooner Information Technology offers Schooner MySQL Enterprise with InnoDB and Active Cluster, which features an optimized MySQL distribution for read-write-intensive online transaction-processing workloads as well as read-intensive online analytical-processing applications, while **ScalArc's** iDB appliance is designed to improve the performance of MySQL with a transparent database query-caching layer. **GenieDB** offers a pluggable storage engine for MySQL to improve the scalability of MySQL to meet the demands of hosted and cloud-based Web applications, while **Xeround** provides its Xeround MySQL Cloud DB SQL database as a service for cloud-computing platforms.

Other NewSQL players are eyeing MySQL deployments as a replacement opportunity. **Clustrix's** Sierra database technology supports the MySQL protocol, which it uses to ensure compatibility with existing MySQL applications, but it is primarily designed to replace MySQL rather than complement it. **VoltDB** is another firm offering a scalable database product and although it is not directly targeting MySQL applications, the biggest interest in

VoltDB is coming from companies that are reaching the scalability and performance limits of MySQL.

Oracle holds the cards

Many of the NewSQL players offering complementary capabilities to MySQL – Tokutek, GenieDB and Xeround, for example – could be seen more as potential acquisition targets for Oracle, rather than competitive threats. And it seems likely that Oracle could buy additional capabilities to protect and extend the MySQL installed base.

Despite fears from some open source supporters that Oracle would use its ownership of MySQL to squash what could be considered a potential rival, we have argued that the company was more likely to see the benefits of positioning MySQL as a complementary offering to Oracle Database and use it as a competitive weapon against **Microsoft's** (Nasdaq: MSFT) SQL Server for departmental and Web applications. And that is precisely what Oracle has done: combining the MySQL and InnoDB development teams and releasing MySQL 5.5 with significant performance improvements, especially on Windows. Oracle has recently been touting its performance gains and favorable cost comparisons with Microsoft's SQL Server.

As for NoSQL, we do not believe that Oracle sees a significant threat from the various NoSQL players at this stage, but if it does, the company already has assets at its disposal to respond thanks to its ownership of Berkeley DB, which it obtained when it bought **Sleepycat Software** in February 2006. While Berkeley DB is best known as an open source embedded database, there has been a notable increase in the number of references being made to it by Oracle employees as a key value store, which is what the technology is at heart.

Oracle could also make use of HandlerSocket, which is an open source project designed to enable MySQL to act like a NoSQL database. HandlerSocket is a MySQL plug-in that interfaces directly with the InnoDB storage engine to boost performance for high-load queries without the need for memcached and was recently added to MySQL consulting and services provider **Percona's** Percona Server distribution.

Oracle is also making increased use of MySQL Cluster, which uses the NDB shared-nothing storage engine to provide high-performance scalability. MySQL Cluster is not new – it has been widely used in the telecom industry and was originally developed by **Alzato**, a spin-off of **Ericsson** (Nasdaq: ERIC) that was acquired by MySQL in October 2003. However, Oracle recently made MySQL Cluster available as a subscription offering for the first time and is also positioning it for use with Web and e-commerce applications.

MySQL, YourSQL and OurSQL

Mention of Percona throws up yet another competitive threat for Oracle's MySQL. The company's investment in MySQL has not stopped the launch of new MySQL support providers and offerings, as other players have seen an opportunity to exploit concern about Oracle's plans. MySQL creator Monty Widenius formed Monty Program around his MariaDB fork, while a group of former MySQL employees established SkySQL to offer support and services for MySQL. Additionally, Percona recently raised the level of its support offering. Meanwhile, the Drizzle fork of the MySQL code base was launched in

2008 to strip out some of the advanced enterprise functionality in an effort to create a more back-to-basics version of the code for Web and cloud applications. Drizzle just reached general availability.

M&A opportunities

If Oracle's major rivals were out of luck in seeking an obvious alternative to MySQL in 2009, today they are spoiled for choice. Would any of them make a play for one of the NoSQL or NewSQL vendors, or for one of the MySQL support providers?

The fact that the vast majority of the NoSQL databases are open source and few are designed to operate in Windows/.NET environments means that Microsoft is an unlikely candidate to go shopping, and in any case we believe that Microsoft's primary offering for lightweight Web applications is increasingly going to be its SQL Azure cloud database. Some of the NewSQL providers that complement existing database deployments while providing additional scalability would be a more natural choice, and it is interesting to note that Microsoft is already working with ScalArc on a version of the latter's iDB technology for SQL Server.

IBM (NYSE: IBM) is the other main relational database provider, and we can see the company adding both NoSQL and NewSQL offerings to its database portfolio – particularly those that provide an alternative path for existing MySQL users. Similarly, **Red Hat** (NYSE: RHT) would be an obvious contender to buy an open source NoSQL database or one of the alternative MySQL support providers, although the company has consistently maintained that it is not interested in entering the database market, and is also developing the Infinispan data grid project, which is likely to be positioned as the distributed data grid layer for cloud-computing environments alongside Red Hat's open source operating system, virtualization and Java development technologies.

Another potential acquirer that could be said to have already made its move is **VMware** (NYSE: VMW), which bought **GemStone Systems** and the GemFire data fabric layer through its **SpringSource** division in May 2010. We believe that VMware will increasingly position vFabric GemFire as a strategic alternative to existing databases as the primary data management layer in the vFabric Cloud Application Platform. Additionally, the company hired Redis creator Salvatore Sanfilippo and contributor Pieter Noordhuis in 2010 and became a sponsor of the key value database. It remains to be seen how VMware might make commercial use of Redis, although its SpringSource business is developing integration between Spring-powered applications and Redis.

Hewlett-Packard (NYSE: HPQ) also has to be considered a potential suitor. The company is in the process of acquiring analytic database provider **Vertica Systems** but lacks an operational database, and new CEO Leo Apotheker recently commented that traditional relational database products are decreasingly relevant for the sort of PaaS offering the company intends to build. While that would appear to put a NoSQL database in the firing line, we think that HP would be more likely to move for a NewSQL or data grid provider that complements its existing database products, making migration to any new PaaS platform easier and protecting the company's ongoing relationship with Microsoft. Apotheker also maintained that HP is not interested in owning its own database but if it were, a vendor like

ScaleBase would be a good fit, as would **EnterpriseDB** given its Oracle compatibility and designs on the cloud relational database-as-a-service opportunities.

Note: This report is excerpted from a forthcoming 451 long-form report that contains analysis of the drivers behind the development and adoption of NoSQL and NewSQL databases, the evolving role of data grid technologies, and associated use cases. It will be available soon from the Information Management and CAOS practices.