NoSQL Innovators – Part 3

RFG POV: NoSQL providers can be divided into five categories: distributed databases, document-oriented NoSQL databases, graph NoSQL databases, in-memory databases, and NoSQL database solutions and services. Across these dimensions there are now more than 50 vendors that have entered the NoSQL DB software and services space. As is the case with most nascent technology markets, more companies will emerge and others will buy their way into the market, fueling the inevitable surge of consolidation. This three-part research note series will address 21 NoSQL innovators that are providing leading edge solutions in the above categories. IT executives will need to understand the NoSQL categories, definitions, alternatives and select a minimum set that best meets corporate needs.

NoSQL innovators – part 3 contains a short list of companies anticipated to disrupt the database space over the next five to seven years arranged in somewhat different categories from the previously defined NoSQL taxonomies and based more on use case within the enterprise than on data model.

This group is also distinguished by added capabilities or functionality beyond just providing a simple data store with the inclusion of analytics, connectors (interoperability with other DBs and applications), data replication and scaling across commodity servers or cloud instances.

This research note discusses in-memory NoSQL databases and NoSQL database solutions and services. Not all of the covered solutions are strictly NoSQL-based, including NuoDB and Starcounter, two providers that refer to their databases as "NewSQL"; and Virtue-Desk, which refers to its DB as "Associative." All three get lumped into the NoSQL category because they offer alternatives to traditional RDBMS solutions.

In-Memory NoSQL Databases

In-memory databases are typically faster than disk storage-optimized databases because they rely primarily on processing data within a computer's main memory (DRAM). This obviates the need to swap data in and out of memory from hard disk storage devices (HDD) and eliminates seek time when querying data. A major game-changer for database architectures has been the introduction of non-volatile storage devices (no moving parts) such as Flash or SSD drives. In-memory DBs can also have attributes associated with other NoSQL categories – e.g., Aerospike is also a distributed, highly available NoSQL DB.

Flash drives can be 100x faster than traditional HDD spinning disk drives and 10x (or more) smaller with the same capacity, allowing for increasingly larger data sets to be stored and managed in-memory – or in very close proximity to it. Flash storage is also much more affordable than DRAM. Due to their speed and compact form factor, multi-terabyte Flash drives are now being installed on computer PCIe (peripheral component interconnect express) and DIMM (dual in-line memory module) cards as well. Real-time, or near real-time updates are possible with in-memory solutions that are able to handle a mixture of live and archived data based on use cases such as online transactions.

Aerospike
Aerospike is an in-memory and a distributed NoSQL DB used for hyper-scale low latency applications that need 100 percent availability. Aerospike takes a hybrid approach

by blending computer memory (DRAM) and Flash – indexes are stored in DRAM and data can be stored in DRAM or Flash memory accessed via a native file system. "Aerospike 3 with queries, user defined functions and aggregations greatly simplifies deployment by compressing the database stack, processing all data within one database layer and eliminating the need for caching and queuing technologies." Aerospike believes simplicity is the way to scale: clusters can be smaller and fewer components need to be kept in sync, therefore, applications are easier to manage while maintaining availability as new nodes are added or servers fail. Aerospike is deployed as a user profile store by real-time bidding firms and other platforms in advertising, ecommerce, mobile and gaming. Aerospike also follows the ACID-compliant transaction standard.
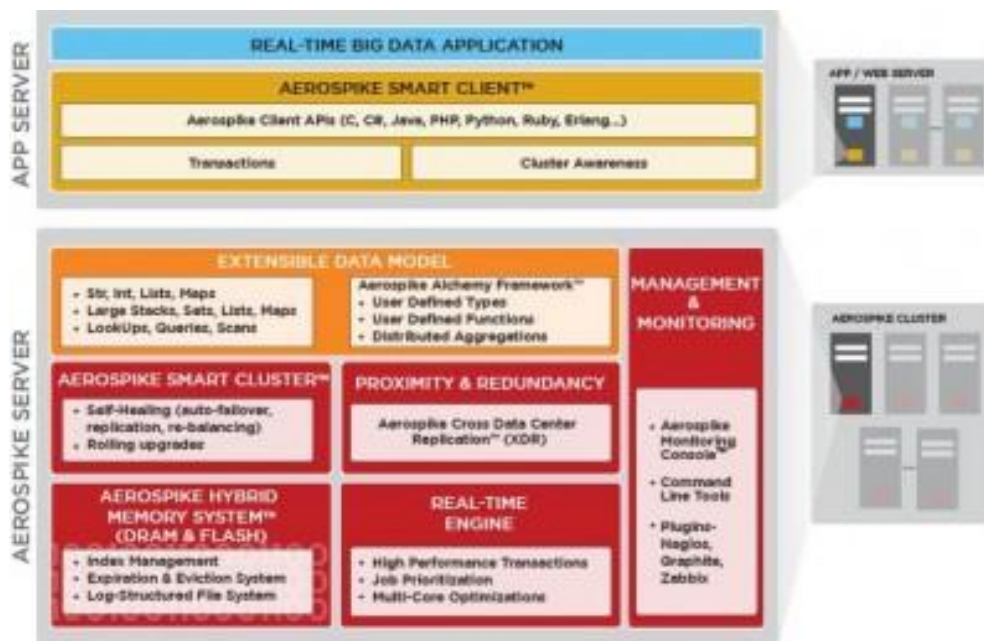
GridGain

GridGain is determined to change the way companies compute. "Just putting data online in-memory is not transformational enough" says GridGain VP of Product Management Jon Webster. "Making Flash look like memory is key. Send the compute to the data. Then memory and data are local to the processing. With GridGain, data movement is minimized." In business for just two years, GridGain's in-memory DB platform has been in development since 2005, attracting a number of household names as clients along with $10 million in a recent funding round. GridGain offers two classes of products: a high performance (HPC) in-memory computational model for risk analysis utilizing historical and streaming data, and accelerators for enabling a new class of in-memory products to enhance the performance of popular open-source solutions, including MongoDB and HDFS.
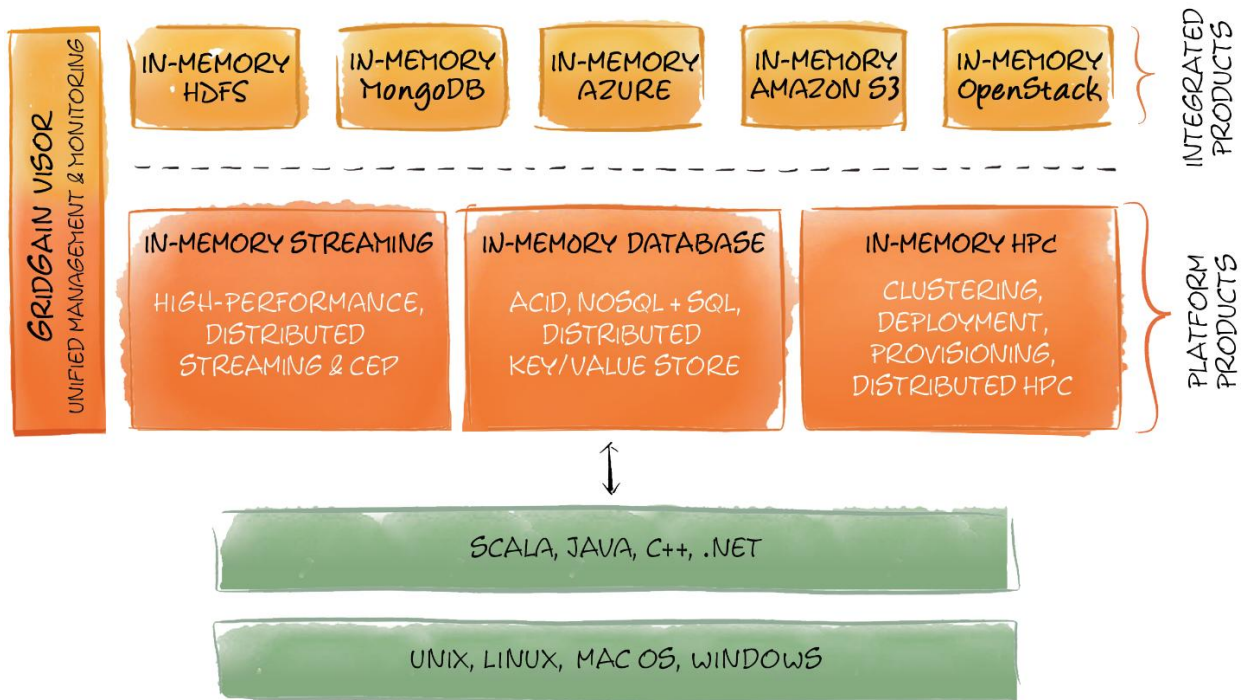
Starcounter

Starcounter is a "scale-in," NewSQL, in-memory DB capable of 3 million web requests on a single server. With ACID compliance, Starcounter 2.0 performs up to 300k writes per second utilizing a multicore server and 100k ACID write TPS on one core and scales reading transactions linearly with 500k TPS per core – making it the "world's fastest consistent" DB. Starcounter has also delivered a new solution dubbed VMDBMS, which is an integration between the application run-time virtual machine and the DBMS. "This makes our solution substantially faster than other in-memory, high-performance databases because data resides all the time in RAM and is not copied back and forth between the database and the application. A native object .NET API completely removes the object-relational mapping (ORM) and reduces the lines of code needed to implement an application." Integrated into Starcounter is a web service supporting JSON/Rest, "enabling performance all the way from the core database out to the end user clients."

Your alternatives

OldSQL

- The market dominating old relational databases (OldSQL) can only process tens of thousands database transactions per second (TPS) on a single machine.

+ They support ACID transactions which guarantee consistency.

NoSQL

+ NoSQL databases distribute the data over several machines and in this way can process many more database transactions per second.

- They cannot provide data consistency and the simplicity that it brings.

NewSQL

+ Starcounter (NewSQL) can process millions of database transactions per second on a single machine.

+ These database transactions are ACID and thus consistency is guaranteed.
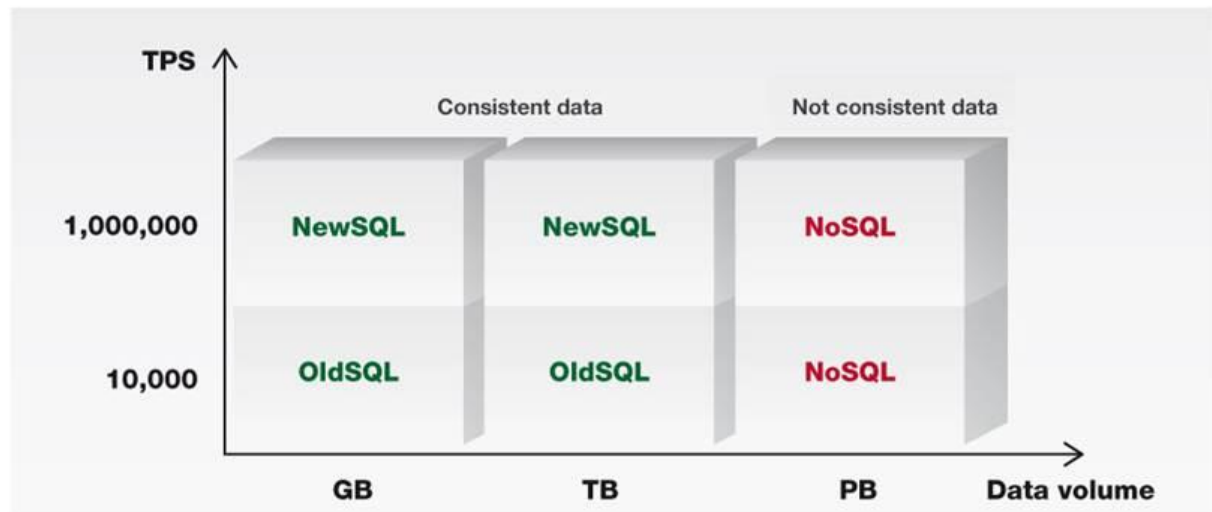
NoSQL Database Solutions and Services

While all of the 21 companies profiled in this report provide database related solutions and some level of enterprise services, this group is characterized by its diversity of services provided from startups who have developed NoSQL integration tools enabling existing DB solutions, to established services firms specializing in DB implementations and consulting services, to Oracle which has dominated the RDBMS landscape for two decades. Oracle could make a good argument for being included in at least four categories as they offer a variety of database options including relational, open-source MySQL (acquired in the Sun deal), a version of open-source Berkeley DB and a variety of business intelligence, DB query tools and high performance storage solutions. Since Oracle defies classification, by default they land here.

28msec

28msec delivers "information agility via an Information Processing Platform that quickly extracts data out of *any* source and transforms that data into a valuable commodity – actionable information." Sold as a service or software, 28.io is a query "layer" designed to consolidate data silos leveraging JSON, XML, relational, object or flat-file protocols and data formats. CTO Matthias Brantner calls the platform "an accelerant for NoSQL. 28.io doesn't store data; it connects databases. Customers get a 360° view of their data

when they write one query that goes to all databases – and Facebook and Twitter if required." When dealing with large datasets, single queries are parallelized and automatically balanced across multiple servers. Founders include senior Oracle architects and the author of XQuery and JSONiq. A key partnership is with the creator of XBRL financial reporting.

Altoros

Altoros is a vendor-independent services organization focused on Hadoop/NoSQL/Cloud Foundry PaaS enablement. Altoros configures open-source cloud platforms such as Cloud Foundry, query optimization tools, OS/hypervisors, NoSQL/Hadoop clusters on AWS, OpenStack and vShere clouds. Altoros has completed more than 25 performance benchmarks of various Hadoop, NoSQL and NewSQL solutions to support organizations of any size seeking intelligence and deployment advice. CEO Renat Khasanshyn believes that platform as a service (PaaS) solutions, such as Cloud Foundry, make infrastructure as a service (IaaS) a commodity: "While IaaS brings value, PaaS could bring 2x to 3x as much, while removing lock-in into IaaS platforms." The Altoros team of 300 engineers and consultants is split between the U.S. and Europe and Latin America. Partners include Pivotal, Cloudera and Hortonworks.

Cloud Foundry Consulting

NoSQL & Hadoop Integration

Cloud Automation

Multi-cloud Cloud Foundry PaaS

NoSQL/Hadoop as a Service

Multi-cloud, Highly Available templates

FACTS ON ALTOROS
Our mission is to transform companies and ourselves

70% Male

30% Female

+5 Join us on an average month

100% Services Revenue

0% License or Referral Revenue

Oslo · Boston · London · Minsk · Sunnyvale, CA · Copenhagen · Zurich · Santa Fe · Buenos Aires

ALTOROS

GigaSpaces

GigaSpaces is "the pioneer of a new generation of application virtualization platforms and a leading provider of end-to-end scaling solutions for distributed, mission-critical application environments and cloud enabling technologies." GigaSpaces' complementary solutions include XAP Elastic Application Platform, an in-memory data grid, and Cloudify, an open-source PaaS solution developed by GigaSpaces that quickly moves organizations to the cloud with little or no application code changes necessary. According to CEO Nati Shalom, "Cloudify provides the equivalent of Amazon OpenWorks on OpenStack without the vendor lock-in." OpenStack is the most widely used open-source software for building private and public clouds. GigaSpaces offers a silo-free architecture along with deployment services to support rapid adoption of cloud-based, high-performance applications.
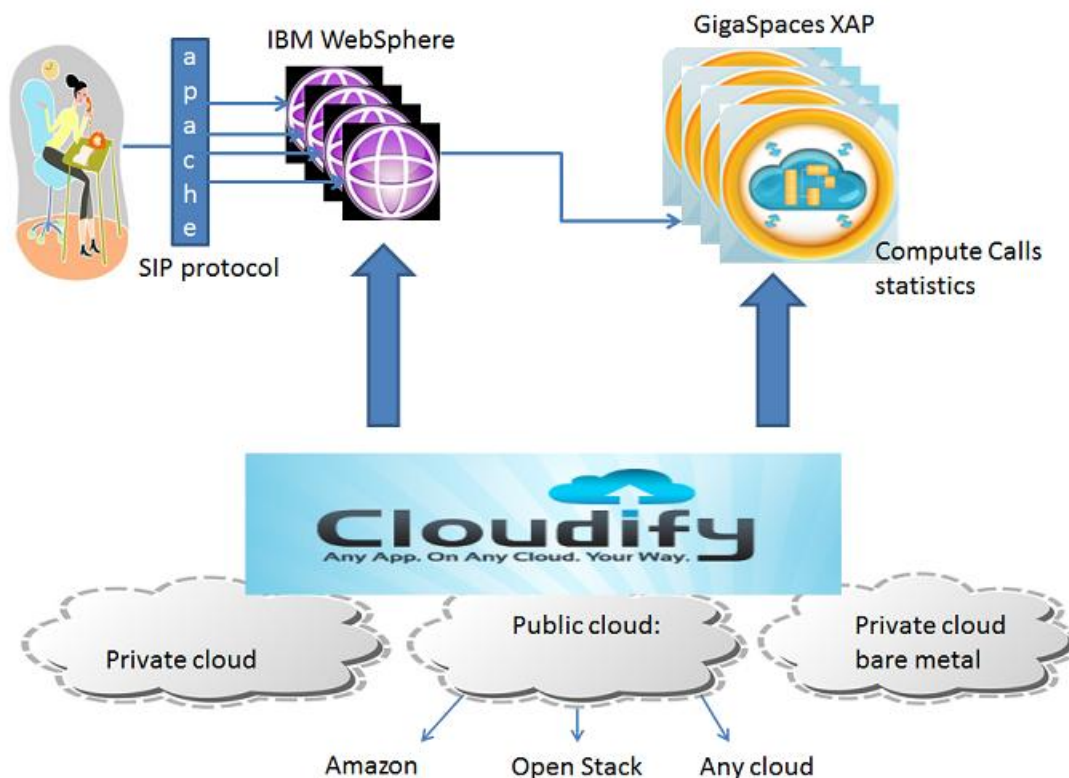
Oracle

Oracle is the market share leader for RDBMS and open-source SQL-centric databases. Oracle's foray into the NoSQL space is based on the Berkeley DB open-source distribution. "Berkeley DB provides a collection of well-proven, building-block technologies that can be configured to address any application need, from the hand-held device to the datacenter, from local solutions or worldwide distributions, from KBs to PBs." Director of Product Management David Segleau described "the first NoSQL appliance," which includes up to 300 TB of disk and a starter rack with 6 dual-core servers with redundant InfiniBand switches, and which offers the Cloudera distribution, including Apache Hadoop to acquire and organize data. Segleau states, "We continue to enhance our NoSQL key-value store (JSON or Graphic) with enterprise-class features, including auto-failover, sharding, query load balancing, smart topology and data distribution."
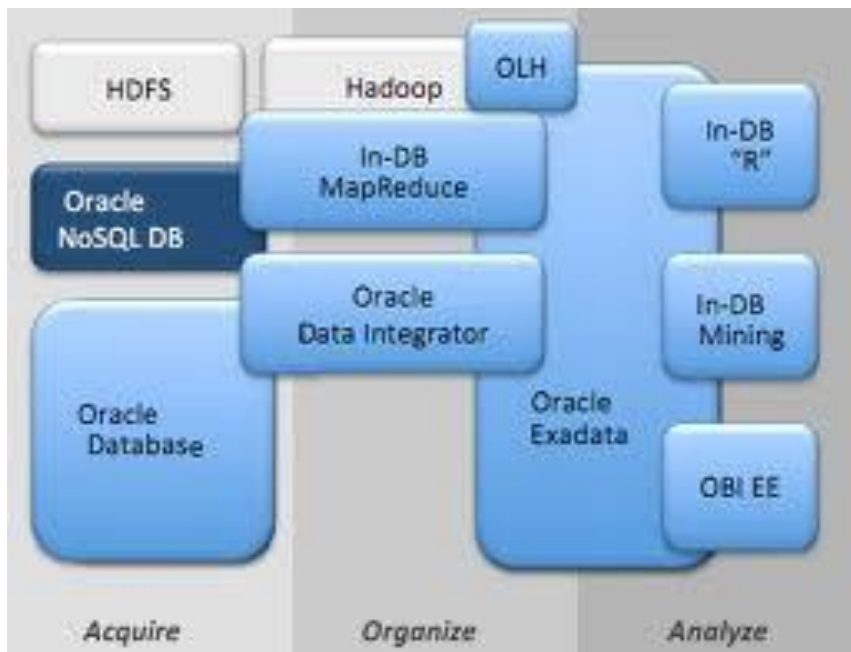
Qubole

Qubole is a "Big Data as a Service" solution. Founded by the creators of Apache Hive and former managers of Facebook's data infrastructure team, Qubole is used by some of the largest brands in social media, online advertising, gaming and other data-intensive enterprise organizations and also completed an initial funding round of $7 million earlier this year. Its flagship 100 percent cloud-based solution, Qubole Data Service (QDS), "provides a fast, auto-scaling Hadoop service built for the cloud, with built-in data connectors and a graphical user-interface for Hive, Pig, Oozie and Sqoop – all integrated in an easy-to-use and easy-to-operate web service." Qubole's turnkey solution has a flexible, pay-as-you-go model that provides end-users with the ability to scale up and down as needed without the need of a technical team or a large capital expenditure.

Integration with Other Data Platforms
Connectors and Data Movement Services
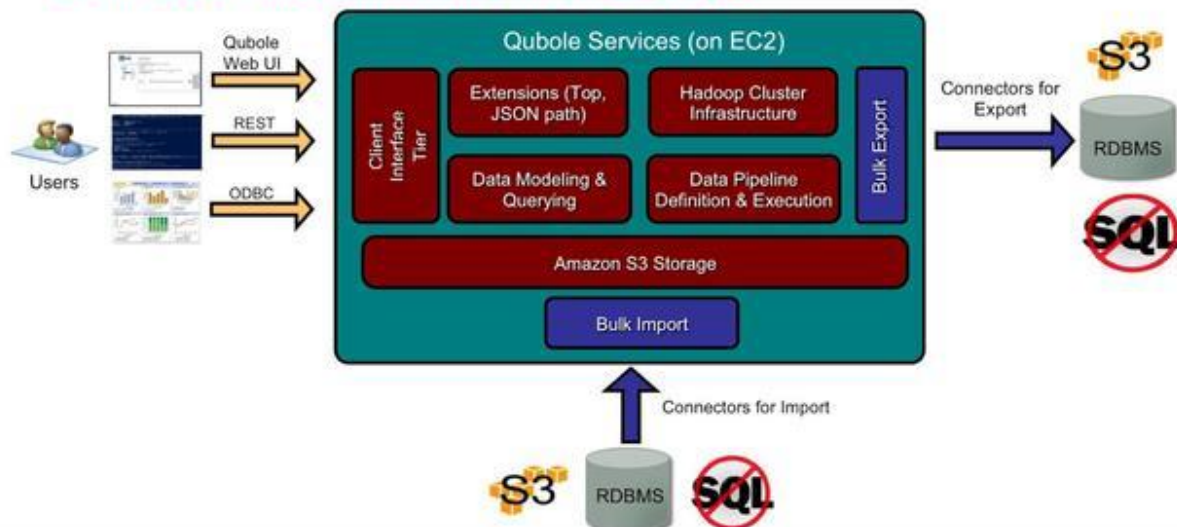
Conclusion

Since no one type of NoSQL database neither satisfies all business requirements, innovators and venture capitalists will continue to invest in newer NoSQL iterations and variations. This will just add to the confusion over the next four or five years while all this slowly sorts out. Thus, while the market remains immature and the options are myriad, IT executives cannot wait before selecting the right NoSQL platforms.

> **RFG POV:** The NoSQL wave of database technology is immature and expanding and a myriad of options exist to confound IT executives and slow down decision-making. The clear trend for non-relational database deployment is for enterprises to acquire multiple DBs based on application-specific needs – what could be referred to as software-defined database adoption. IT executives and data architects should understand the variety of options and then map them to current and future business and technical requirements for each application type where a NoSQL database might apply.

Additional relevant research is available. Interested readers should contact Client Services to arrange further discussion or interview with Mr. Gary MacFadden, Principal Research Analyst.