## NoSQL Innovators – Part 1

**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.

Oracle has publicly committed to its Berkeley DB open-source version of NoSQL, while IBM offers support for Hadoop and MongoDB solutions as part of its InfoSphere information management platform as well as Hadoop enhancements for its PureData System, and Microsoft supports a variety of NoSQL solutions on its Windows Azure cloud-based storage solution. Suffice to say, the big three RDBMS vendors are pragmatic about the future of databases. Sooner or later, expect them all to make NoSQL acquisitions.

Meanwhile, these research notes address 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 grouping 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 the distributed databases. 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.

**Note:** One could argue that other categories such as [http://en.wikipedia.org/wiki/Embedded_database Embedded Databases] could also be included. In over 20 hours of interviews, only two NoSQL solution providers – Oracle Berkeley DB and Virtue-Desk – mention embedding their databases within applications. In the case of Virtue-Desk, its solution is written entirely in Assembler and can be embedded in "any" device that has more the 1MB of memory – the DB is only 600k installed.

## Distributed Databases

This group is made up of four venture-backed, pure-play NoSQL vendors and two NoSQL alternative solutions in NuoDB and Virtue-Desk. Distributed databases are critical for web-scale or cloud-based applications where multiple nodes are needed to ensure availability and partition tolerance – an AP solution in CAP Theorem jargon. (See RFG research on Evolution of NoSQL.) These databases also have capabilities consistent with other categories of DBs, including relational and graph. Use case examples include ecommerce applications for Amazon and other web stores, gaming programs such as Angry Birds, Netflix, and applications for government agencies and scientific research organizations.

**Basho**

Basho is the progenitor and primary curator of Riak, a non-relational open-source DB (under the Apache license since 2011) based on Amazon Dynamo. With Riak CS (cloud storage) release 1.4 in March 2013, Basho boosts its distributed cloud enablement

capabilities to allow customers to support well over 100 nodes, making it easier to distribute data. Riak CS VS (virtual store) also supports multi-tenant object stores for CSPs looking to extend their portfolios. Basho boasts a 6x performance advantage over the most popular NoSQL DBs, superior availability and data replication characteristics compared with any kind of database, and dramatic TCO (total cost of ownership) over traditional RDBMSs – as much as 80 percent savings. According to CTO Justin Sheehy, "One of Riak CS's key draws is that customers can use it to build private or public cloud storage that is API-compatible with Amazon S3."
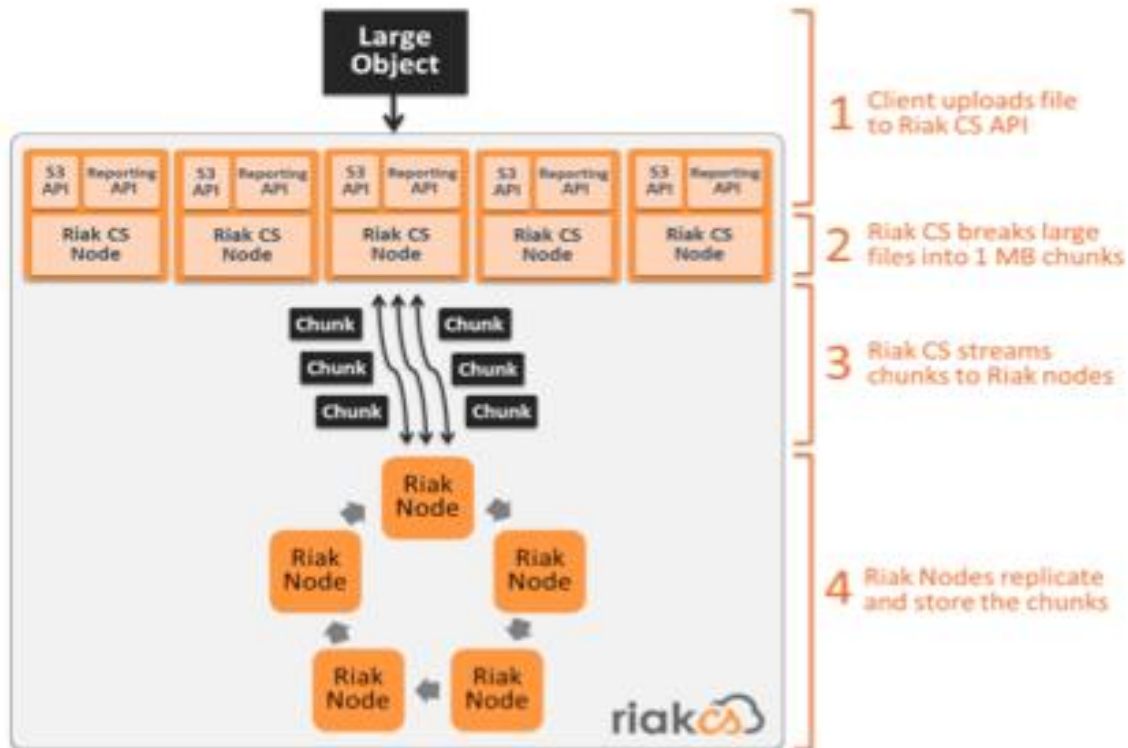


### DataStax

DataStax is laser focused on providing its customers with availability and performance when powering their online applications. The release of its Enterprise (DSE) 3.1 Cassandra-based platform offers more powerful tools and easier deployment capabilities that allow up to 10x more Cassandra data per node. According to CEO Billy Bosworth, DSE 3.1 improves "enterprise confidence" by providing analytics and search capabilities for online applications that rely on real-time data, such as tracking and displaying shopper search trends, with the ability to personalize data within the same cluster. DataStax is aggressively addressing security for the bulk of its customer use cases, including integration with Kerberos, and embracing industry standards such as Hive and ODBC for directing data between systems. A recent large investment round will help accelerate company growth worldwide.

**DataStax Enterprise** – Hot Data in Context
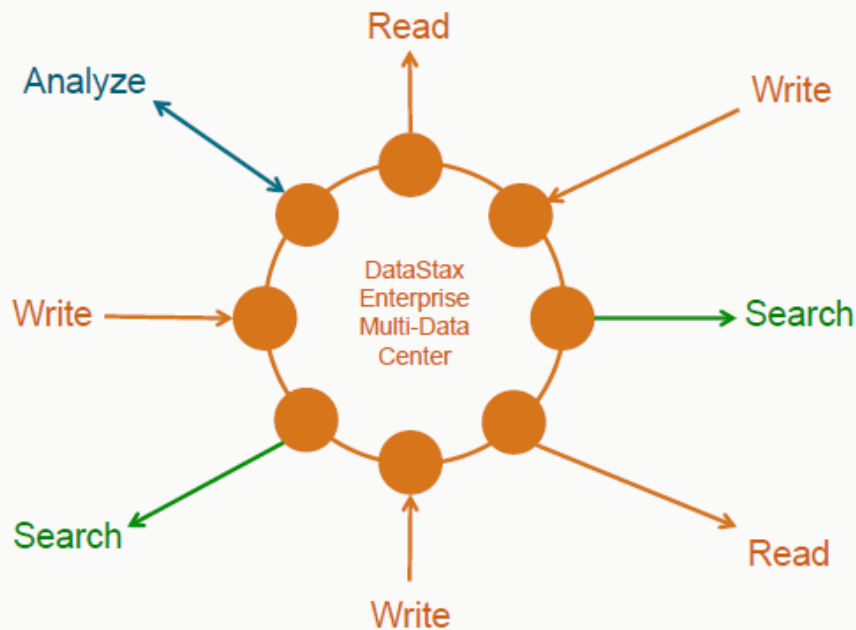
**FoundationDB**

FoundationDB is a new entrant into the NoSQL space that emphasizes scalability and consistency while supporting ACID (Atomicity, Consistency, Isolation, Durability) transactions and giving up "perfect" availability. Community and enterprise versions became generally available in August 2013. Co-founder David Rosenthal describes the process as "turning many computers into a single powerful database. By using a shared-nothing distributed architecture, FoundationDB scales out by adding more machines to a cluster rather than just scaling up by increasing capacity of a single machine." Utilizing a key-value store data model, FoundationDB has been providing its more than 2,000 beta users with multiple "real transactions" without sacrificing speed. Service contracts are priced starting at $99 per month per server process including support for document, graph and SQL and can be effectively used for both operational (OLTP) and analytic (OLAP) workloads.
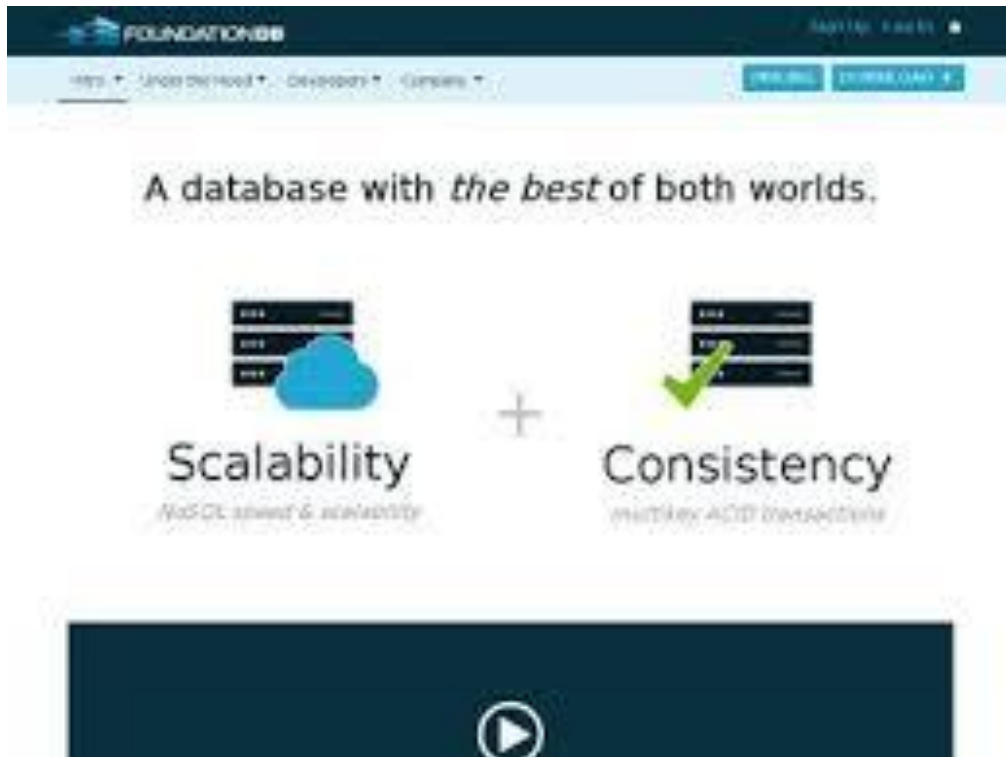
### NuoDB

NuoDB describes itself as "the only distributed cloud database that is 100% SQL, 100% ACID and 100% elastically scalable." Launched in February 2013, NuoDB targets web-scale distributed processing analytics and OLTP use cases offering a single "logical" database across several virtual machines or geographies while maintaining the high performance expected from scale-up DB solutions. NuoDB touts its easy administration that allows users to "seamlessly" switch between on-premise and cloud deployments running on Azure or AWS and in Windows or Linux configurations. While not an in-memory DB, NuoDB's architecture allows for up to 100 transaction engines, which NuoDB states doubles the throughput with every added engine. The DB synchronizes storage nodes across all storage engines so each node has a full copy of the data to provide built-in high availability and replication capabilities.
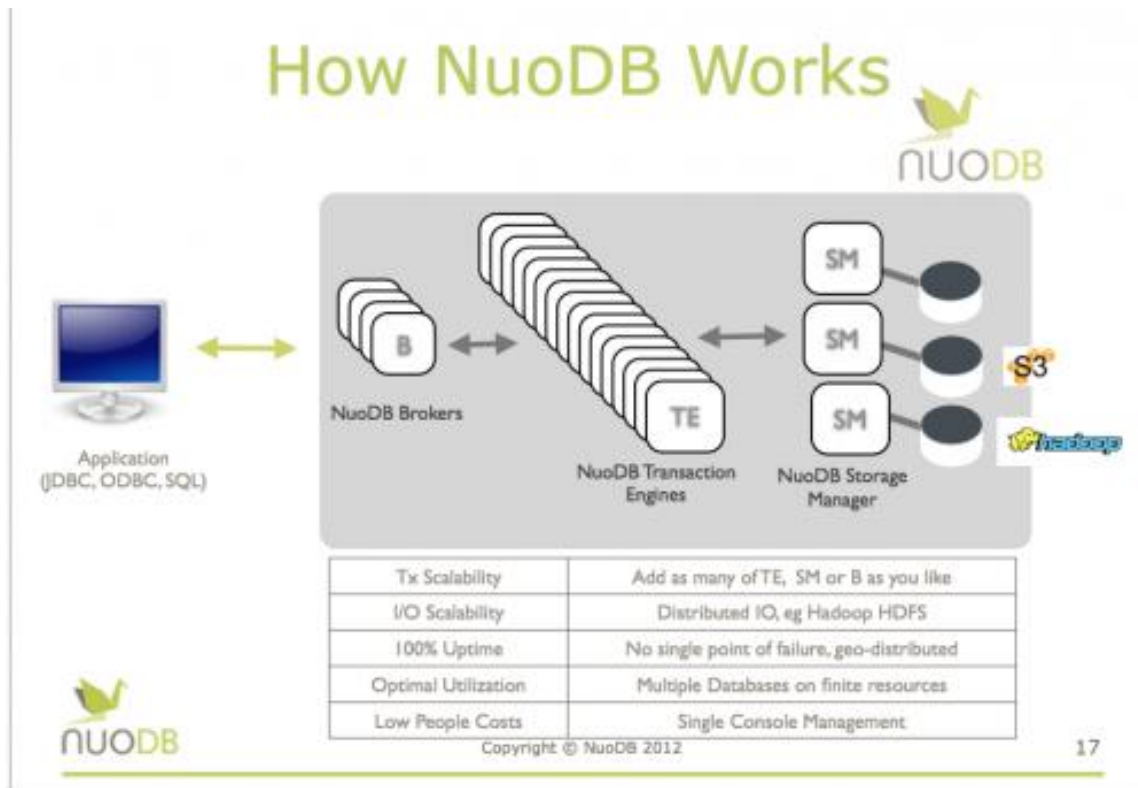
## Sqrrl

Sqrrl is the creator of Sqrrl Enterprise, "a highly secure, massively scalable Big Data platform powered by Apache Accumulo and Hadoop." The genesis of Accumulo started in 2008 with the National Security Agency's (NSA) search for a petabyte-scale, BigTable-like (Google) data store with flexible schemas and fine-grained security features (aka "cell-level security") that could very quickly handle both structured and unstructured data running on commodity hardware. Sqrrl extends Accumulo "with additional data ingest, security, and real-time analytical features." Sqrrl Enterprise also cuts across multiple NoSQL categories by adding document store and graph store capabilities to Accumulo. Use cases include fraud and risk applications in government, financial services, healthcare, and cyber security. Accumulo is widely deployed and used throughout the U.S. Department of Defense and Intelligence agencies, and now has users across the Federal Government and in various commercial sectors.

# Sqrrl Enterprise Data Flow



Sqrrl Data, Inc. All Rights Reserved
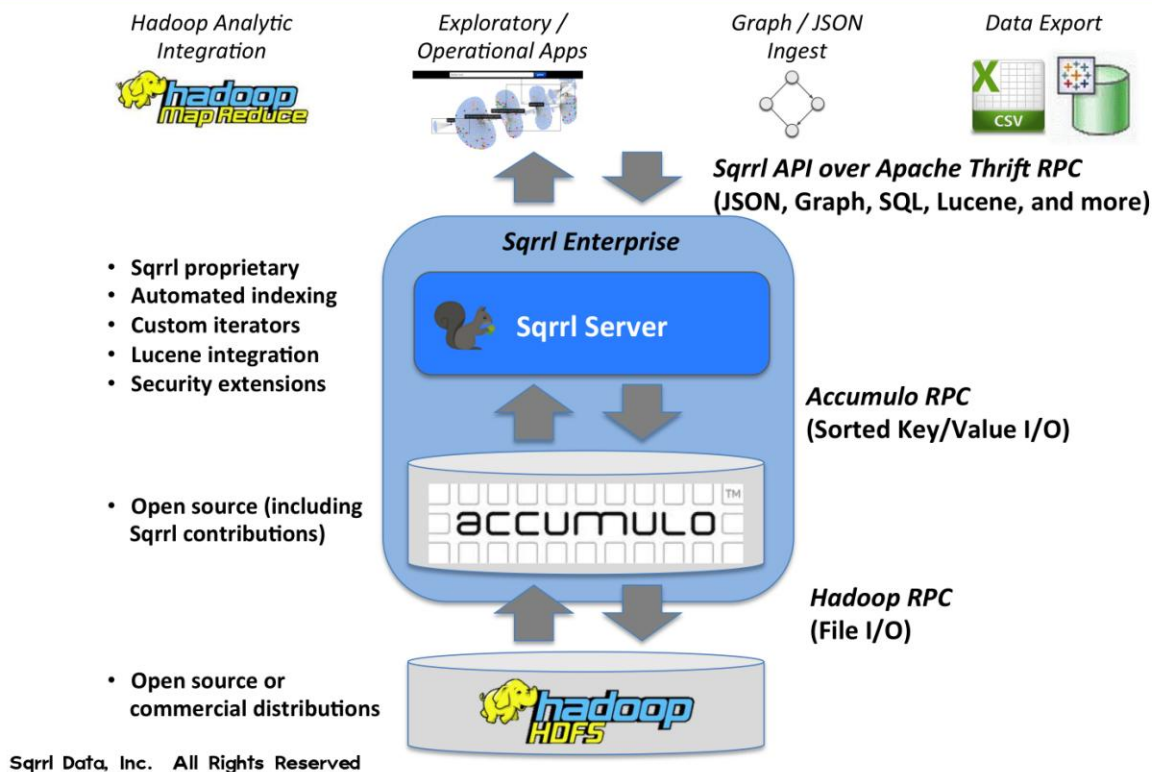
**Virtue-Desk**

Virtue-Desk provides what it refers to as an "associative" database, which it claims is 100x faster than SQL on reads. Dubbed AtomicDB, the solution has no views, no indexes, no tables and no whitespace, and there are no queries to write. Used for more than a decade by the U.S. Navy to keep track of 70 million parts and patented in 2011, AtomicDB extends its practical use to healthcare and financial services applications, including multi-petabyte-scale document stores, and is intended to be used by business analysts and other non-technical users. The system does not "hold" data but "associations" with the data, and "the data is the structure" left resident in its existing location with application-class libraries acting as APIs. CTO Jean Michel LeTennier states, "AtomicDB is much simpler and more cost effective to implement than traditional RDBMSs or even NoSQL solutions."
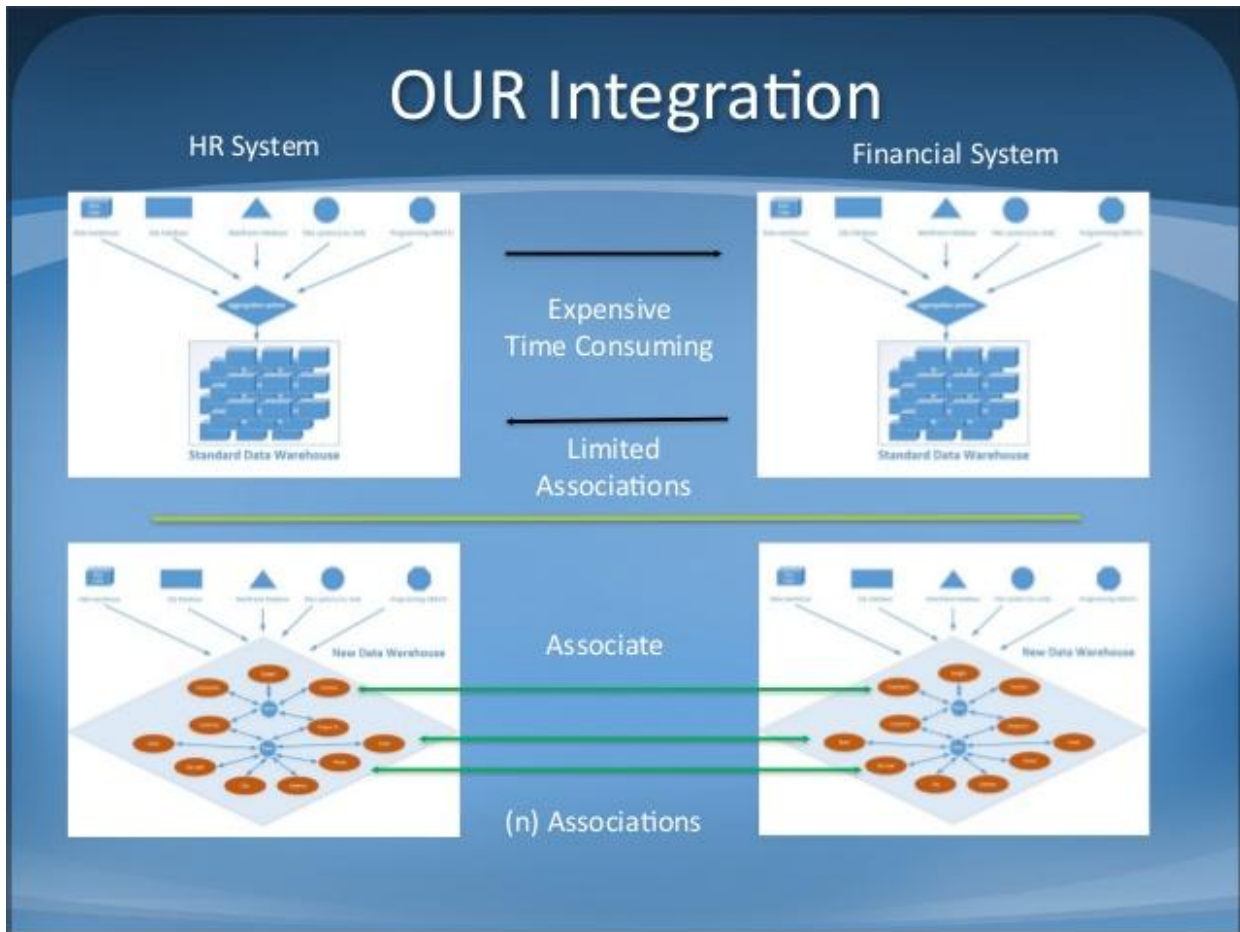
Wednesday, August 28, 13

## 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 rapidly changing 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.*