



NoSQL Innovators – Part 2

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.

This research note addresses 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 document-oriented and graph NoSQL 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.



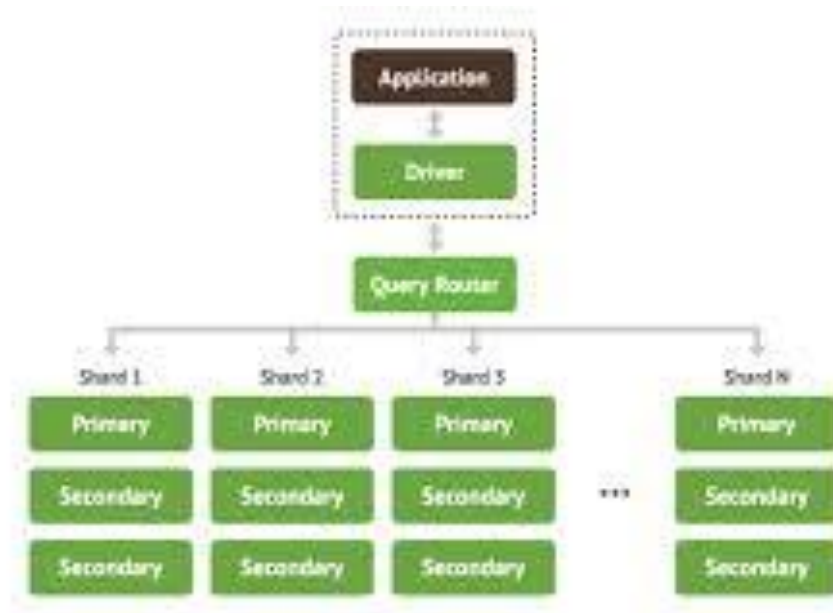
<i>Disruptive NoSQL DB Solution Providers</i>	
<i>Distributed Databases</i>	
<i>Document-Oriented NoSQL DB</i>	
<i>Graph NoSQL DB</i>	
<i>In-Memory NoSQL DB</i>	
<i>NoSQL DB Solutions, & Services</i>	

Document-Oriented Databases

As the category name implies, DBs in this group – also referred to as document stores – are optimized to handle documents and other forms of unstructured or semi-structured data such as emails, instant messages and the like. As with non-relational DBs in other categories, document-oriented DBs can also possess attributes associated with other DB categories, including distributed nodes, graph capabilities and near real-time analytics.

MongoDB

MongoDB (formerly 10gen) is the developer of MongoDB, which has the largest community of any open-source database distribution. MongoDB executives attribute the large following to a few key attributes: ease of solving both easy and hard problems, agile development, sufficient for most workloads and a transparent business model that makes it easy to do business. MongoDB's enterprise version includes Kerberos authentication, SNMP support and user training. Both community and enterprise versions feature a JSON data model with dynamic schemas for improved document handling, auto-sharding of objects to enable horizontal scaling, replication and availability, and rich document querying and search capabilities. MongoDB is deployable on-premise, in the cloud or as a hybrid solution, is supported by IBM services and is the most popular NoSQL DB on Amazon AWS. MongoDB announced October 4, 2013 that it secured \$150 million, the largest funding round ever for any Database vendor – NoSQL or otherwise.

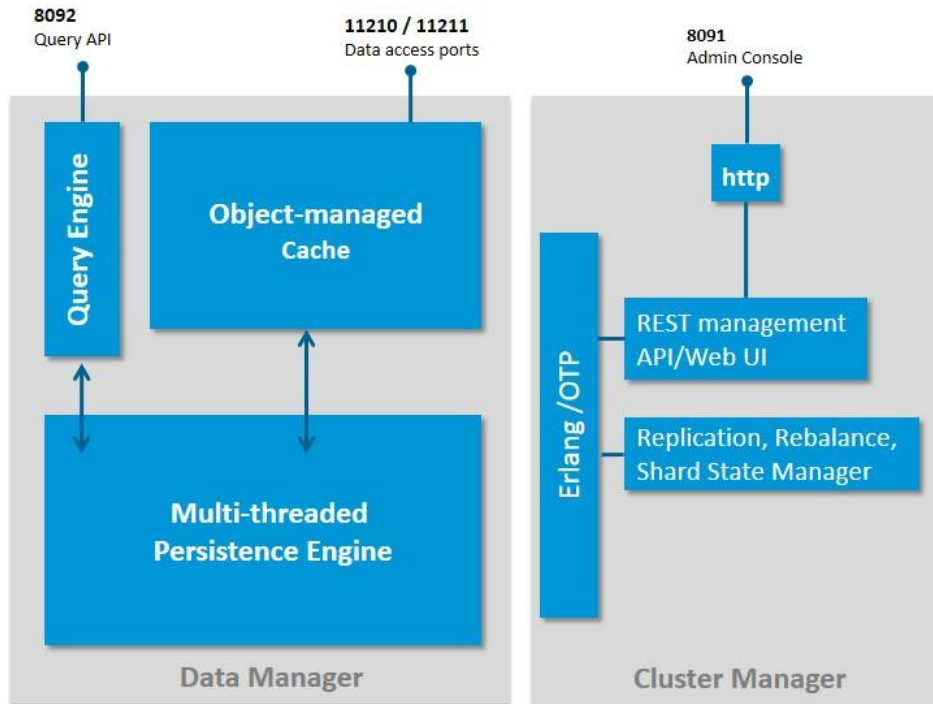


Couchbase

Couchbase is laser focused on being "the platform of choice for the most demanding web and mobile applications at the world's largest enterprises." Couchbase Server is a NoSQL document database optimized for interactive web and mobile applications. A flexible JSON data model makes it easy to modify applications without being constrained by a fixed database schema. "With sub-millisecond, high-throughput reads and writes, Couchbase Server delivers consistent high performance for web and mobile apps. It is easy to scale out, and supports live cluster topology changes with zero downtime." Couchbase recently announced its JSON anywhere mobile strategy with the first NoSQL database for mobile. A recent funding round brings total money raised to \$55 million. Funds will be used to further expand international sales and marketing operations and support key strategic product initiatives.

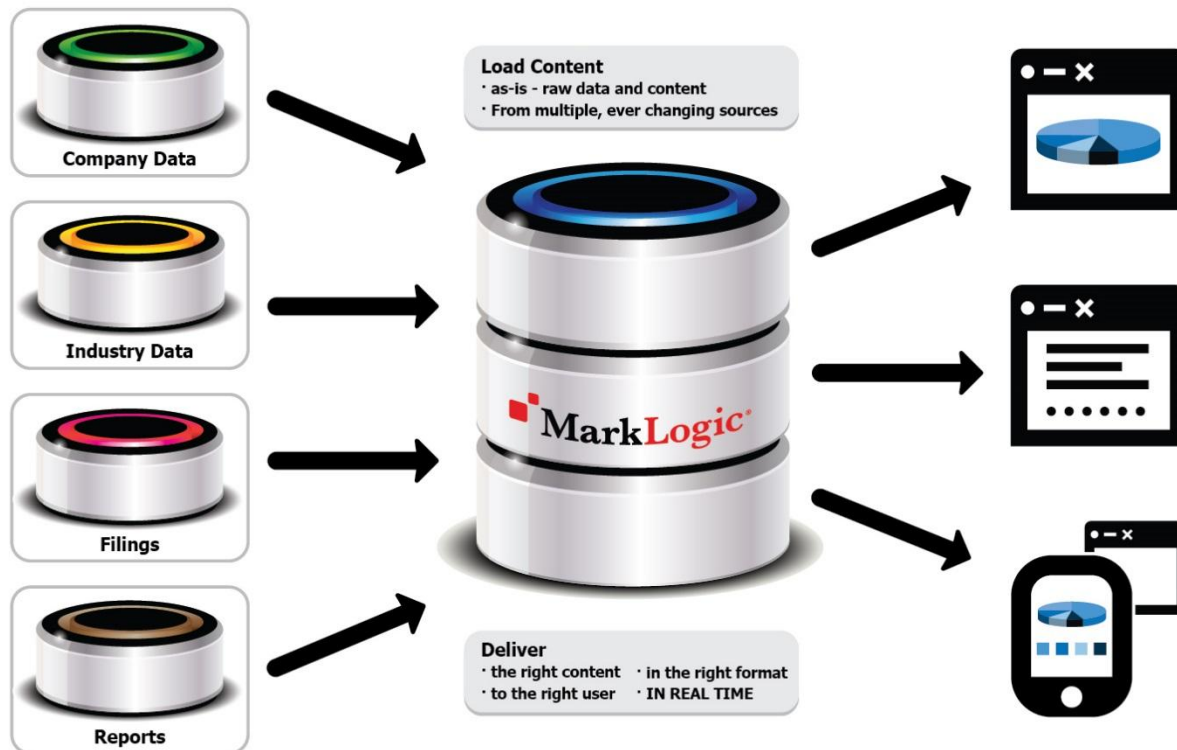


Couchbase Server Architecture



MarkLogic

MarkLogic is the market share leader in the Hadoop/NoSQL market segment as measured by Wikibon. A dozen years ago, MarkLogic embraced XML and XQuery as document markup and access standards for multi-terabyte scale collections. Today, MarkLogic Server ingests a variety of other document formats, including PDF and JSON, is "schema-agnostic" and has developed or partners with a variety of query, search and analytics programs to find information within enterprises' document stores. Some clients report that they have replaced SQL with XQuery although MarkLogic supports both, as well as keyword and faceted searches, enabling non-technical users to more easily find information within documents or search meta-data associated with images, sound and multimedia files. MarkLogic supports ACID transactions and has developed a REST API, a native Hadoop bi-directional connector and semantic indexing and search capabilities along with other enhancements to support its latest 7.0 release.



Graph NoSQL Databases

While databases outside this category, such as Sqrrl, support graph capabilities, these four providers specialize in this segment. Some graph DB vendors are also appropriate for use cases beyond purely graph database apps, including YarcData, which has an in-memory discovery analytics capability. Graph DBs are often paired with other types of databases to dramatically improve performance and relevance for e-commerce, fraud detection or knowledge-based applications.

Franz Inc.

Franz has "expert knowledge in developing and deploying Semantic Web technologies (i.e., Web 3.0) and providing Common Lisp (programming language)-based tools that offer an ideal environment to create complex, mission-critical applications." AllegroGraph and Allegro CL, Franz's object-oriented development system, with AllegroCache are distinct scalable platforms used by startups and Fortune 100 companies for knowledge-based applications or for social media analysis. "AllegroGraph is a modern, enterprise, high-performance, persistent graph database. It uses efficient memory utilization in combination with disk-based storage, enabling it to scale to billions of quads while maintaining superior performance." AllegroGraph supports SPARQL, Jena, Sesame, ACID Compliant, RDFS++, and Prolog reasoning from numerous client



applications. Franz also provides training, services and support for Lisp-based programming environments and has built connectors to MongoDB and other popular databases as well as to search and BI (business intelligence) tools.

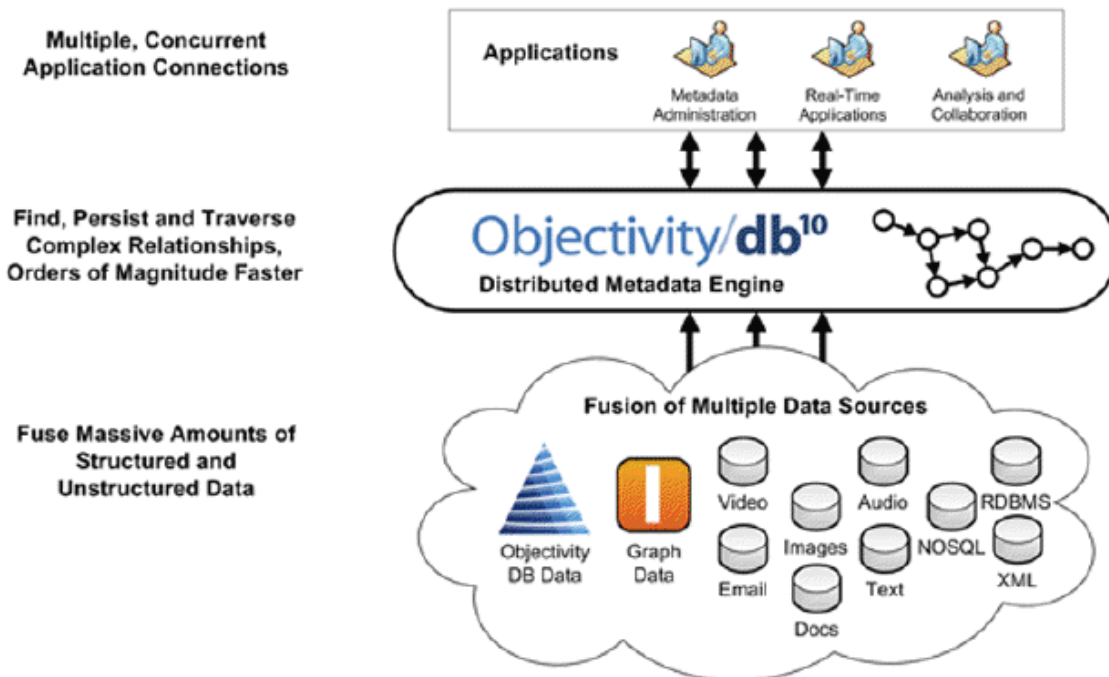


Neo Technology

Neo Technology developed, open-sourced and now supports Neo4j, which has the largest ecosystem of any graph database, with over 500,000 downloads. Its enterprise version supports high-availability clustering, ACID requirements and delivers what Neo4j's CEO Emil Eifrem refers to as a "run-time, real-time transaction environment" for OLTP and other mission-critical use cases. Eifrem believes the most powerful cognitive model for developing relationships between seemingly disparate data types is the whiteboard, and the Neo4j graph model mimics that whiteboard friendliness. "Query performance with connected data sets can literally be 1,000 times faster than traditional DBs because it's a native graph database." Social networking, identity & access management, geo routing, dependency analysis and fraud detection apps have all adopted graph DBs due to their speed and ease of use. In Eifrem's view, the need for fast, intuitive, visually compelling applications is driving their growth. Neo4j also works well with several NoSQL DBs.



The Distributed Metadata Engine

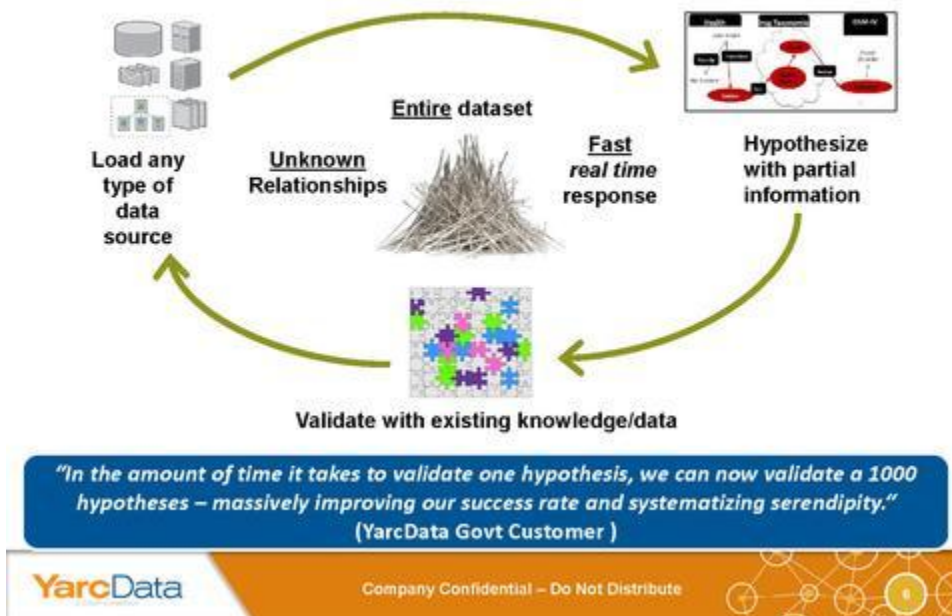


YarcData

YarcData focuses on in-memory discovery analytics as opposed to just search. A wholly owned subsidiary of supercomputer manufacturer Cray Inc., YarcData turnkey appliances help solve complex Big Data problems suitable for graph DBs. Its purpose-built Urika appliance has 512 TB of shared memory along with 8,000 processors that offer a performance boost of two to four orders of magnitude over traditional RDBMSs. Urika is particularly well suited for sifting through massive amounts of unstructured or rich text data sets as its triple store database architecture – similar to the Semantic Web – is ideal for uncovering hidden relationships within constantly changing and varied data sources. Use cases include personalized and evidence-based medicine, fraud detection, cyber security, financial risk management, and baseball analytics.



DISCOVERY through FAST hypothesis validation



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.