

NOSQL OPERATIONAL CHECKLIST

NEW APPLICATION REQUIREMENTS ARE DRIVING A DATABASE REVOLUTION

There is a new breed of high volume, highly distributed, and highly complex applications that have database requirements beyond what the relational database can deliver. Traditional relational databases were never designed to cope with modern application requirements — including massive amounts of unstructured data and global access by millions of users on mobile devices that require geographic distribution of data.

These new applications and operational requirements have forced Enterprises to look for and create solutions to solve their emerging needs. Because of this, Enterprises are more willing than ever to explore alternatives to the relational database model.

ENTERPRISES ARE RAPIDLY ADOPTING NOSQL DATABASES

Enterprises are increasingly choosing NoSQL databases to fulfill the requirements for their highly scalable applications. These requirements are driven by the need for storing massive amounts of unstructured data, high performance, fault tolerance, as well as the growing complexity of Enterprise applications.

According to 451 Research, the NoSQL database market is expected to grow from a little under \$200 million in 2012 to over a billion dollars by 2016. This five-fold growth in 5 years is being fueled by organizations adopting NoSQL databases for their Enterprise applications.

451 also says that a key-value data store is the only currently available database technology capable of providing the combination of scalability, performance and robustness for extremely large databases, those beyond the capability of RDBMS.

WHY RIAK KV FOR OPERATIONAL SIMPLICITY?

SCALABILITY — Easily scale up and down with no manual sharding

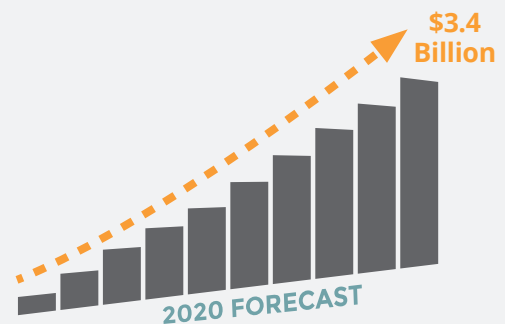
GLOBAL AVAILABILITY — Ensures consistently fast apps anywhere and on any device

FAULT TOLERANCE — Automated data replication and self-healing

PERFORMANCE — 5x faster response times than relational DB

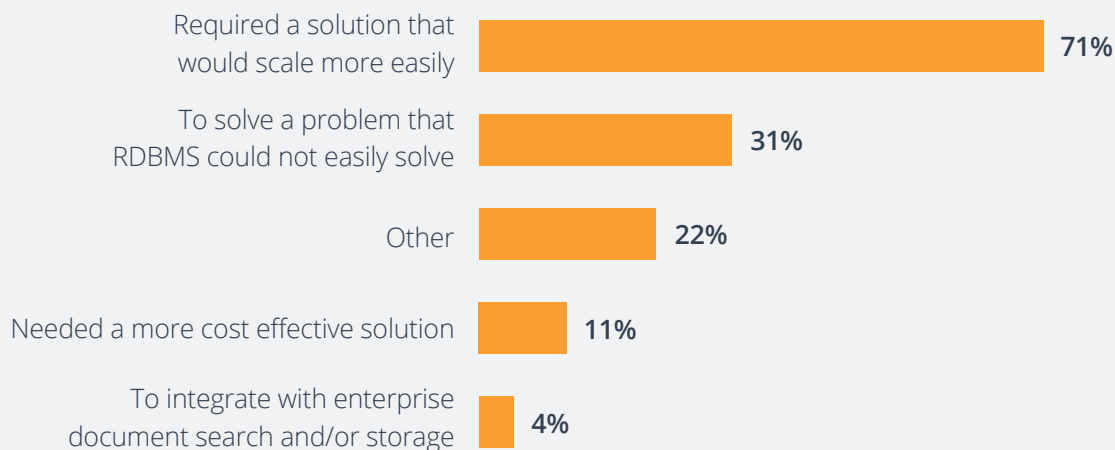
The gap between Enterprise requirements and traditional relational database capabilities has driven organizations to look for other database solutions.

GLOBAL NOSQL MARKET



In a recent TechValidate survey, 71% of respondents said that they chose NoSQL because they needed a solution that would scale more easily.

WHY DO ENTERPRISES CHOOSE NOSQL?

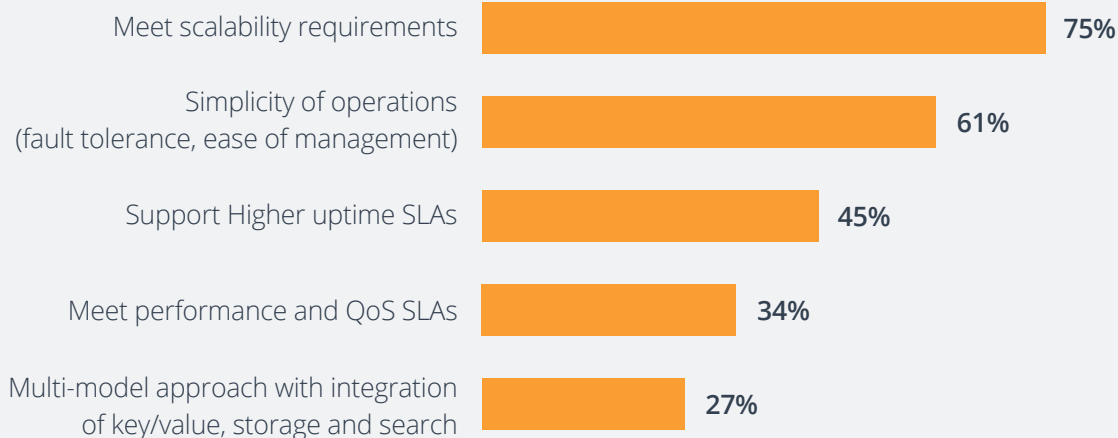


BEYOND NOSQL, CHOOSING THE RIGHT SYSTEM FOR OPERATIONAL SIMPLICITY AND SCALABILITY

The companies that pioneered the NoSQL movement did so in the context of strict requirements around availability, scalability, fault tolerance, and global footprints. These same requirements along with operational simplicity must be considered by the Enterprise when adopting and implementing NoSQL solutions.

In a recent TechValidate survey, respondents said that NoSQL priorities for them were meeting scalability requirements and simplicity of operations.

THE IMPORTANCE OF FAULT TOLERANCE



OPERATIONAL CHECKLIST

SCALABILITY AND PERFORMANCE

Typically, all applications and databases work at small scale. However, many modern applications need the ability to scale both up and down.



*To **scale vertically** (or **scale up**) means to add resources to a single node in a system, typically involving the addition of (bigger and faster machines) (bigger) CPUs or memory (more cores) to a single computer (machine).*

*To **scale horizontally** (or **scale out**) means to add more nodes to a system (or cluster), such as adding a new computer to a distributed software application. An example might involve scaling out from one Web server system to three.*

- Wikipedia: <http://en.wikipedia.org/wiki/Scalability>

Many Enterprises have found that traditional relational databases don't scale as needed and are not cost effective when scaling up. Often, Enterprises need to provision hardware for peak load leaving wasted resources when not running at peak.

Enterprise applications need to be able to scale out linearly to ensure a positive return on investment as additional commodity hardware is added. NoSQL databases enable scale out by adding more servers and resources, but manual sharding of the database may be required, which is a complex operational task. However, not all NoSQL databases are alike — massive scale distributed systems enable scale out, scale up, and scale down predictably and linearly. While traditional relational databases

Things to consider when choosing the right NoSQL solution for your company:

- Scalability
- Performance
- Global Availability
- Simplicity of Operations
- Fault Tolerance

scale by using larger and larger servers, Riak scales linearly simply by adding additional commodity servers without the need to manually shard.

It is also important for your NoSQL system to perform well with fast response times. In the case of a virus checking notification service that serves 50M connections per day, the traditional Oracle database generated response times that were 300 ms for a business critical operation. After switching to Riak, the service saw their response times reduce to 60 ms for that same operation.

The Weather Channel ingests more than 20TB of data a day. They capture 2.2 million current-weather-condition data points from around the globe four times per hour. The company's new consolidated platform, called SUN (Storage Utility Network), will capture 2.25 billion (with a "b") weather data points 15 times per hour. Riak met their critical requirements enabling them to scale both quickly and cost effectively.

“ *When you're globally distributing massive amounts of data across Amazon nodes or Google Compute nodes, you want something that's simple to use and configure. Cassandra, for example, is great at distributing data, but it's complicated and complex to run. Riak was built to handle massive data movement, replication, and data-synchronization on a cloud-based, globally distributed data platform.* ”

— Bryson Koehler, Executive VP and CIO
at the Weather Company

MANUAL SHARDING

With traditional relational databases, it is hard to grow when your dataset exceeds the capacity of a single server. While it is possible to manually shard your database, this can often be difficult and time consuming. Manual sharding puts an extra operational burden on your staff requiring them to separate and manually manage chunks of data onto different servers.

GLOBAL AVAILABILITY

Many Enterprises have additional complexity due to requirements for global availability. These Enterprises need to provide a globally consistent user experience. The implementation of global availability and global user experience must be easy to operationalize.

To achieve global availability, you need to deliver content or services out of multiple locations — often geographically disbursed. This is true whether you own and deploy the equipment in physical data centers, or implement with cloud services across different geographical locations.

The importance of global availability can be seen in Rovio's deployment of Angry Birds. In addition to Rovio having a global audience, it has functions wherein the latency needs to be as low as possible. Rovio also needs to ensure its high service levels can be achieved in a cost-effective fashion.

For example, in May of 2013, Angry Birds (Rovio) had over a billion downloads and 263 million monthly active users. By early 2014, they had grown to approximately two billion downloads.

Rovio is using Riak KV as the database supporting its mobile gaming platform, including features such as payments, game state storage, and push notifications. Riak KV enables a

function to maintain game state across devices so that users can continue their game when they switched to different devices. Additionally, Rovio used Riak KV to implement a “wallet” function to enable low-latency user purchases from any device, anywhere. Riak Enterprise allows Rovio to serve low-latency content at massive scale to concurrent users all over the world.

“ Providing the infrastructure for hundreds of millions of users is no small feat. The world is becoming much more connected, and people are using more devices than ever before. Keeping track of those data types and scaling to meet demand cost-effectively can be a huge challenge. With Riak KV, Basho has provided us with the fast, scalable and flexible foundation needed to address the challenges associated with cross-platform entertainment. This has been done while keeping operational costs affordable and while providing the best possible experience to our global fan base.

– Juhani Honkala, Vice President of Technology at Rovio

END CUSTOMER EXPERIENCE FOR GLOBAL USERS REQUIRES GEO-DATA LOCALITY

Many Enterprises' applications are not just being used globally, but are relied upon for fast access to rapidly changing and critical data. End customer experience matters. When access is slow, application usage drops along with productivity. When data operations are closer to end users, they receive better response times and a better end user experience. The data locality capabilities of distributed systems enables data operations close to end users.

Tapjoy's experience with the challenges of a global customer base where fast access is required is a great demonstration of the power of distributed systems. Tapjoy is a mobile advertising and monetization platform that allows end users to select personalized advertisements that they can engage with in exchange for rewards. At Tapjoy, over 460 million mobile users participate in advertisements in exchange for virtual currency.

Tapjoy's original infrastructure was built on SimpleDB; however, with billions of requests coming in on an average day, they

started to experience performance issues due to latency. The farther away their users were from the application and database infrastructure, the performance and end user experience got worse. Their SimpleDB infrastructure also imposed limits on the size and location of data being stored.

With their growth straining their data store, they wanted to find a new solution that would guarantee performance and uptime, even with peak traffic. To ensure deep user engagement and ultimately to optimize monetization, Tapjoy needs their data replicated across multiple datacenters so that any mobile device can access it.

When choosing a database to address their requirements, Tapjoy considered Riak KV, Cassandra and HBase. Tapjoy selected Riak for its features, performance, and operational ease. After migrating to Riak KV, Tapjoy witnessed a tenfold increase in performance. They now generate 2ms response times for 5 billion operations per day.

OPERATIONAL SIMPLICITY AND FAULT TOLERANCE

For applications using NoSQL databases, Operational Simplicity — especially involving fault tolerance — is as important a consideration as scalability. Fault tolerance is key for supporting uptime SLAs, performance SLAs, and Quality of Service SLAs.

Surviving failure conditions does not equate to fault tolerance, because a truly fault tolerant system will keep the application fully available — meaning both reads and writes will continue during a partition.

Bump Technologies' NoSQL deployment highlights how important fault tolerance is to operational simplicity. After deploying Riak KV, when a node has a problem, they no longer have an emergency on their hands. They simply reboot the machine and Riak will read-repair as necessary. Bump ultimately chose Riak KV for its operational ease-of-use, ability to scale writes, and availability under failure conditions.

While most NoSQL vendors talk about having no Single Point of Failure (SPOF), not all approaches to fault tolerance are equal. Distributed systems, like Riak KV, by their very nature have a masterless architecture that helps support fault tolerance. With well-designed distributed systems, your system, application, and database can survive server, network, or data center failures.

“ *As a start-up offering these monitoring services it is vitally important Dataloop.IO works with a platform able to cope with the huge demand being placed upon it. Any losses in data or disruptions to the service would be hugely damaging to our growing reputation. Riak KV has given us exactly what we need.* ”

– Colin Hemmings, Co-Founder and Chief Architect of Dataloop.IO

“ *Tapjoy can't have downtime ever, planned or unplanned... Riak is a critical piece of our business, and it's a huge relief that it just works.* ”

– Weston Jossey, Head of Operations at Tapjoy

Dataloop.IO, the fast-growing provider of infrastructure monitoring software for online services, chose Riak KV to underpin its metric-crunching cloud monitoring solution. With architecture required to scale to millions of metrics being generated every minute, Dataloop.IO selected Riak KV to support the analysis, visualization and interrogation of metrics as they arrive in real time.

Working with Dataloop.IO, Riak KV hosts the huge amount of time-series data required, in addition to providing memory backing to prevent data loss and enabling the company to manage its complex event processing system. Riak's ability to scale up or down depending on the number of customers and servers enables Dataloop.IO to grow its client base without impacting performance.

SELF-HEALING

For Operational Simplicity, it's important to select a NoSQL database that:

- Enables you to eliminate Single Points of Failure (SPOF)
- Has self-healing cluster technology
- Offers self-healing that works both “at rest” and “after a network partition”

SUMMARY

NoSQL databases are increasingly required for Enterprise mission critical and distributed applications. Operational considerations can make or break a NoSQL deployment, making ease of scalability, true fault tolerance, seamless global availability, and operational simplicity all key considerations when choosing a NoSQL system.

Remember these questions when choosing a NoSQL database:

- Can the database scale up, down, and out with linear performance?
- When adding capacity is manual sharding required?
- Does the database continue to do both reads and writes DURING a failure condition?
- After a failure, does the database recover automatically and is the system self-healing?
- Will my end-users have a consistent and fast experience no matter where they are or what device they are using?

Choosing a solution based on operational considerations makes great business sense. If the database is easy to scale, fault tolerant, and globally available, this will reduce the time and energy your team spends on database operations leaving more resources to focus on other aspects of growing your business. Riak KV Enterprise is especially suited to fulfill your Enterprise NoSQL requirements including operational simplicity.

NEXT STEPS

You can read more about [Riak KV and Riak KV Enterprise](#), including more in-depth technical details for developers and operators, at the documentation portal. Riak KV is available open source for download at <http://basho.com/resources/downloads/>.

If you are interested in Riak KV Enterprise, please contact us. We would love to talk to you about your possible use case and how Riak KV can work for you.

For more information visit www.basho.com or follow us on [Twitter](#) at www.twitter.com/basho.

