

Beyond Relational: Reimagine Your Data With Enterprise NoSQL

May 2014

Table of Contents

Executive Summary	3
Introduction	4
The NoSQL Paradigm Shift	6
Reimagining Big Data with MarkLogic	7
An Illustration: Relational Vs. Non-Relational	9
Getting Started with Enterprise NoSQL and MarkLogic	12
Resources	13

Modern enterprises face increasing pressure to deliver business value through technological innovation that leverages all available data. At the same time, those enterprises need to reduce expenses to stay competitive, deliver results faster to respond to market demands, use real-time analytics so users can make informed decisions, and develop new applications with enhanced developer productivity. All of these factors put big data at the top of the agenda.

Unfortunately, the promise of big data has often failed to deliver. With the growing volumes of unstructured and multi-structured data flooding into our data centers, the relational databases that enterprises have relied on for the last 40-years are now too limiting and inflexible. New-generation NoSQL ("Not Only SQL") databases have gained popularity because they are ideally suited to deal with the volume, velocity, and variety of data that businesses and governments handle today.

A paradigm shift in database technology is occurring in which organizations are rethinking what they can do with their most important asset: *data*. MarkLogic is at the forefront of this shift, having the only NoSQL database that is truly enterprise ready. With MarkLogic, organizations are making the world more secure, providing access to valuable information, creating new revenue streams, and reducing expenses.

Reimagining Your Data

Organizations adopting MarkLogic's Enterprise NoSQL database are achieving success:

- CMS provides access to health coverage for millions of Americans through healthcare.gov
- The BBC delivered 2.8 petabytes of content in one day via dynamic semantic publishing
- McGraw-Hill Financial analyzed 1,300,000 articles in one hour rather than five days
- CQ Roll Call reduced IT support time by 95%
- A top-5 investment bank replaced 20 systems with one for its derivatives trade store

Introduction

The Promise of Big Data

It was just a few years ago that the term big data began to show up in the media, and experts in the field rose from the database mines to become database pioneers exploiting the treasures of "big data." Ever since, the spotlight has been shining brighter as businesses seek to leverage their untapped information resources and entrepreneurs seek to win big on the promise of big data. With the world's data expected to

double every two years until 2020, it is not surprising that many businesses are anxious about whether they are truly getting the value that they hoped for out of their data resources.¹

The Problem with Big Data

Unfortunately, big data initiatives have often failed because organizations have tried to use existing information management practices and legacy technologies, which often collapse under the sheer weight of the data. All too often, enterprises try to use the same relational database they have been using but soon realize that the database is not only too slow, but in fact cannot handle the new requirements at all. Relational databases were great when information was highly structured and consistent, low in volume, entered through forms, and retrieved with a well-known set of queries. But, if you need to dig deep inside your data to understand context, analyze details, and assemble custom reports and views, these technologies quickly demonstrate their shortcomings.

"This notion of thinking about data in a structured, relational database is dead."

- Vivek Kundra, Federal CIO, July 21, 2009; Open Government and Innovations Conference

Businesses need a new approach to handling the volume, velocity, and variety of big data because the current relational model that has been the *status quo* for the past 40-years is not working:

- **Volume:** By 2020, the digital universe will grow to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child). The need already exists to process petabytes worth of data fast and with low overhead.²

¹ IDC, "Digital Universe Study". 2012.

² IDC, "Digital Universe Study". 2012.

- **Velocity:** 90% of the world's data was created in the past two years.³ Managing the stream of data in real-time and conducting real-time analytics is required because businesses and governments simply cannot afford to wait any longer.
- **Variety:** The world's data has too much variety, around 80% of which is unstructured. In one study, 64% of businesses said the primary reason for considering a new approach to big data was the diverse, new, and streaming data sources they now have to handle.⁴

Handling these three "V's" is something that relational databases have trouble with as relational data modeling is inherently different from the NoSQL approach. Unfortunately, it is not easy to get away from the relational approach. The leaders of the relational database market are Oracle, IBM, and Microsoft, and

together they make over \$24B per year.³ Oracle's relational database, including its multiple versions that have evolved over the past few decades, is arguably the single most successful information technology product of all time. That said, we believe the time is right for a change—there is a better way.

"Like a pair of spectacles... Occasionally a paradigm shift takes place when the spectacles get smashed, and you then put on a new pair that transforms everything into new shapes, sizes, and colors. Once this shift takes place a new generation is brought up wearing new glasses and accepting the new vision of truth. Through these new glasses, you see a whole new set of puzzles to be solved..."

- John L. Casti, **Paradigms Lost**⁶

³ SINTEF. "Big Data, for better or worse: 90% of world's data generated over last two years." ScienceDaily, 22 May 2013. <www.sciencedaily.com/releases/2013/05/130522085217.htm>.

⁴ New Vantage Partners, "Big Data Executive Survey: Themes & Trends". 2012 <<http://newvantage.com/wp-content/uploads/2012/12/NVP-Big-Data-Survey-Themes-Trends.pdf>>

⁵ Gartner, 2012.

⁶ Casti, John, *Paradigms Lost* (New York: William Morrow & Co., Inc., 1989), 39.

The NoSQL Paradigm Shift

A new generation needs a new generation database. NoSQL is gaining significant market traction⁷ because it solves the fundamental challenges with big data. It is not just an evolution of an existing technology, but a true paradigm shift in how businesses and governments think about and manage their data. With NoSQL, big data has moved from being a nebulous term to a current reality as a new generation database continues to out-perform its predecessors.

The primary reason is that NoSQL databases handle the variety of data very well—they are flexible and scalable to support the requirements today's data problems demand. NoSQL databases refer to the class of databases that are characterized by the following three features:

- **Scale easily across commodity servers, virtual machines or cloud instances:** NoSQL databases are built to scale horizontally, meaning it is very easy to distribute the data across a cluster of server instances. With many NoSQL databases, the gains in ease of scaling come at the cost of transactional consistency, but MarkLogic has a distinct advantage in that it can scale across large server clusters and maintain full read and write consistency.

- **Do not require rigid schemas be predefined before ingesting data:** There are countless examples of business value being locked up in data that does not fit neatly into rows and columns. Unstructured data, or any data that does not have a pre-defined data model, includes social media, health records, financial documents, journals, videos, web pages, and much more. NoSQL databases can manage this data and traditional structured data as well.
- **Use richer, more humanistic data models:** The problem with the typical approach to handling hierarchical information is that data is “shredded” into tables. A customer, derivative trade, or legal document is normalized into a model that satisfies referential integrity. But, just consider the typical first step when analyzing normalized data: *de-normalize* it! NoSQL document databases such as MarkLogic use hierarchical, tree-like data models based on XML or JSON to accommodate varying degrees of structure, organizing the data in a way that mirrors intuitive human perception.

These features help NoSQL databases achieve better performance, scalability, and flexibility, whether it is a document store (MarkLogic is a document store), column store, key-value store, or graph store. MarkLogic leverages the many advantages found with all NoSQL databases, but also solves key barriers to NoSQL adoption by providing enterprise features such as ACID-compliance and government-grade security.

⁷ Gartner, “Hype Cycle for Big Data, 2013”, July 31, 2013.

A New Data Model

Where Does Bob Live?

That question can be answered in two very different ways depending on how you store the data, illustrated in the simple example below. With a relational data model, answering the question “where does Bob live?” requires traversing across the customer, address, and state tables. The NoSQL document data model is human-oriented, but also very efficient for computers to analyze and process for big data applications. As data gets more complicated and less structured, the document model’s flexibility and human-oriented structure becomes even more important.

Relational Data Model

ID	ADDR_ST1	ADDR_ST2	ADDR_ZIP	ADDR_ST_CD
115	11 Main	Apt 7	11301	22
442	92 Locust		77112	33
934	33201 Alth	Unit 998	90271	9

ID	CST_NM	CST_AGE	ADDR_ID
112	Bob	25	642
133	Narraj	51	934
348	Jose	35	115

ID	ST_ABBR	ST_NAME
9	CA	California
22	NY	New York
33	TX	Texas

NoSQL Document Data Model

```

graph TD
    Customer[Customer] --> Name[Name: Bob]
    Customer --> Age[Age: 23]
    Customer --> Address[Address]
    Address --> Street[Street: 91 Locust St.]
    Address --> City[City: Arlington]
    Address --> State[State: TX]
    Address --> Zip[Zip: 77112]
  
```

Reimagine Big Data with MarkLogic

MarkLogic stands out among NoSQL databases because it has enterprise capabilities that organizations need to run mission-critical, operational applications: ACID transactions, government-grade security, policy management, scalability adapting to multiple cost-effective hardware choices, APIs and connectors for a wide variety of third-party services, and database management and monitoring tools. Today, MarkLogic's customers are reimagining their data to:

1. **Make The World More Secure**

A world-leading aerospace company built an entity management framework for the intelligence community powered by MarkLogic. The development team fully embraced MarkLogic as an application and analytics platform, and MarkLogic soon became the backbone of their architecture. They gained rich semantic search capability, real-time alerting to track incidents, and a simple

"The challenges around the velocity and volumes of data increase every day... there's new data sources coming online, from social media to transactions in our industry to our internal data. To be able to deal with that kind of volume of data, we need new paradigms."

- Adriaan Bouten, SVP, McGraw-Hill Financial

architecture that allowed for flexibility and scalability. This framework helps analysts manage, sort, and make sense out of huge volumes of fast-moving information.

2. **Provide Access to Valuable Information**

A top-5 investment bank developed a global, real-time, 24/7, single, unified, and accurate view of derivatives trades. MarkLogic replaced 20 disparate batch-processing servers with a single operational trade store that enables the bank to know its market and credit counterparty positions in real-time, providing the ability to act quickly to mitigate risk. The accurate and holistic view of the data allows the bank and its regulators to confidently rely on the metrics and results it reports.

3. **Create New Revenue Streams**

A 150-year old publisher used MarkLogic to consolidate 10 of their legacy systems into one customer-facing application. The new platform combines all of the existing systems on top of MarkLogic, and also implemented new features using an agile development approach. Just a little over a year later, the new platform was launched. The improved time-to-market enabled the company to increase acquisitions and partnerships with publishing partners.

4. **Gain Insights to Increase Market Share**

A leading provider of ratings, benchmarks, and analytics in the global capital and commodity

markets used MarkLogic in combination with Intel's distribution of Apache Hadoop and Tableau business intelligence software to conduct in-depth text analytics on 1,300,000 articles. Prior to using MarkLogic, the same analysis took five days and was done quarterly. With MarkLogic, the process only takes one hour, and the customer is now able to visualize data in new ways and quickly iterate on the approach.

5. **Reduce Bottom Line Expense**

A leading provider of congressional news, legislative tracking and advocacy services needed to accelerate updates of the comprehensive information database associated with all of their web properties. An independent researcher found that MarkLogic helped reduce IT support-time by 90 percent, improve line-of-business productivity, and increase customer revenues due to the improved user experience. They found a return on investment of 77%, an expected payback timeline of only 1.3-years, and a three-year cumulative net benefit of over \$1.3 million.

An Illustration: Relational Vs. Non-Relational

A Search and Discovery Application

To better illustrate some of the key differences between developing an application using a relational database versus MarkLogic, we will follow the example of building a functional web application for search and discovery.

Imagine an application that consolidates data and provides a rich search experience. It has a traditional search box, and will output usable results for further interaction. Your application also includes analytics features, like trending topics, or results based on demographic segments. You probably have a few data sources already available, like customer information, sales and contract documents, and external sources of industry information—all with facts and figures that sit in spreadsheets or tables.

Traditionally, you would bring all these data sources together by standing up a relational database to store the data, and bolting on a search engine such as Solr or Lucene to discover and understand the information.

The Relational Approach

In a relational database management system, you have to understand which rows and columns your data fits into before you start the build. This means you have to invest time and money to build out the perfect schema needed to support your application.

Once you have your data model, you normalize the data to fit into your application's database. At the same time, you pre-define the data to be indexed and pre-determine the ways in which the information can be searched.

But what happens when the users require another source of data or want another way to search it?

When you come upon the need for a new data source for your application, unless you included this information in your original schema, you now have to rebuild it from scratch and go through the whole ETL (extract, transform, and load) process all over again. This means going back and revisiting all those rows and columns to identify how your data model should incorporate this new information. You have to rebuild the data model, renormalize your data, reconfigure how the search engine indexes that new data, and reconnect it to your user's application. If any requirements change, you have to start all over again.



Figure 1: With relational architecture, time and effort is spent defining schema, and there is high probability of re-work

The traditional relational model has roadblocks that will stop you from achieving your goals and innovating. Whenever you need new application features or data sources, your data gets locked up while you rebuild your schemas. This model also impacts on business agility, as you cannot quickly respond to market changes or new business demands.

Relational databases and search engines also keep your ongoing costs high, as each change requires expensive investments in data modeling and internal build sessions while your competitors are out-maneuvering you. This applies across rebuilds resulting from new delivery methods such as mobile devices or special analytics layers. You either have to rebuild the schema, or build out an entire new platform.

With MarkLogic, there is a better way to organize the infrastructure stack for truly useful search and discovery applications that require less maintenance than traditional environments.

The MarkLogic Approach

To start, the database technology is completely different. MarkLogic takes all of the complexity out of your data by allowing you to load and store your data as-is—no upfront transformation needed. MarkLogic stores the data in the structure as-is, including JSON, XML or defined data models.

There are no schemas and no complex data models to build. Since you are not normalizing that data or extracting pieces to fit into a model, you index everything. That means you can search for it tomorrow, even

"It's fun to watch people use MarkLogic and have these 'Eureka' moments when they realize how easy it is to do things that used to be so complex."

- John O'Donovan, Director of Architecture & Development, Press Association

The other important part of the data system is MarkLogic's embedded search feature. Not only does it perform search and retrieval, but it also handles database-style queries and analytics. This engine was built especially to work with our indexes, and delivers sub-second search results across all data-sets.

In addition, MarkLogic includes application services, tools, and APIs out-of-the-box so that your team can develop and code in any language to deliver applications as quickly as possible, in any format needed. Whether you need web, mobile, or embedded applications, you can build them all using the single MarkLogic system by simply adding interfaces as you need them.

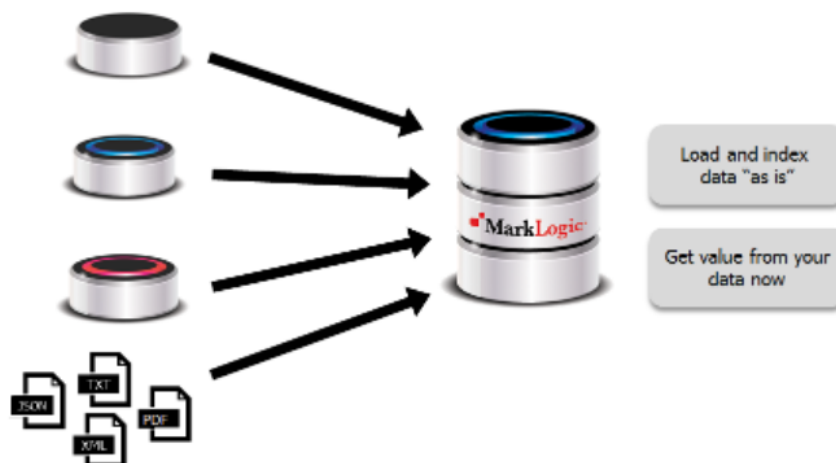


Figure 2: With MarkLogic Enterprise NoSQL Architecture, load data as is, whether it's JSON, XML or defined data models

Enterprise NoSQL

MarkLogic is the only Enterprise NoSQL database on the market. This means that in addition to providing the flexibility and agility needed to get the most out of your data, MarkLogic has the government-grade security, high availability, and back-up and recovery features that enterprises absolutely need for their mission-critical data systems.

MarkLogic is also the only NoSQL database with full ACID transaction support, which means it does not lose your data. And, it's the only NoSQL database with search included, which eliminates the development time and maintenance needed for a bolt-on search solution. MarkLogic is enterprise ready today, as opposed to just having enterprise features in a roadmap—something MarkLogic customers can testify to.

Meeting Our End Goal

With MarkLogic, innovation is never slowed down. Your team can respond to market demands, adding new features and functionality with speed. MarkLogic includes everything you need to build new applications, cutting the time needed to get your system up and running. This translates directly to reducing bottom line expenses. And, as data is migrated into MarkLogic, you are able to turn off expensive legacy systems—further reducing expenses.

Such completeness enhances the speed of development. Where some enterprises spend months or years trying to solve issues in a relational database, MarkLogic can help solve them quickly. And finally, because MarkLogic loads your data as-is and indexes your data in its entirety, your data is always available. You can create interfaces that allow your users to search and retrieve data as needed—future-proofing your business.

Getting Started with Enterprise NoSQL and MarkLogic

MarkLogic has worked with hundreds of organizations that have embraced change and have moved forward to create innovative, mission-critical applications, from leading publishing and media companies to world-leading trade banks to government agencies. For organizations beginning to develop big data strategies, we recommend some of the following steps to get started:

- **Establish a Sense of Urgency:** Seeing the need to change is critical to moving past any complacency in moving forward with the adoption of NoSQL as part of your vision for a big data strategy. With a sense of urgency established, it is then possible to establish your vision and gain buy-in to your development initiatives.
- **Build the Right Team:** It's not just about the technology, but about the people that are engaged with the initiative—including developers, administrators, and leadership. Everyone in the organization will benefit from MarkLogic, but it is important to identify who the real change agents are, making sure to include both decision-makers and implementers.
- **Begin with the Right Project:** First, it is important to weigh all of the costs of implementing a new technology to fully understand how MarkLogic can integrate with your existing infrastructure. Second, it may be best to begin with a modest yet still meaningful project that will help your team

gain familiarity with the technology if they are not familiar with NoSQL or MarkLogic.

- **Engage MarkLogic Early:** There are dedicated experts available who have experience deploying MarkLogic for all sorts of organizations. It is important to engage the MarkLogic team early as a technology partner, during the development phase, when support is most critical.

Resources

MarkLogic provides numerous resources to get you started quickly. Please visit marklogic.com or contact us at sales@marklogic.com.

Resources	
Free Developer License	developer.marklogic.com/products
Free Online Training	marklogic.com/services/training/
White Paper: Inside MarkLogic	marklogic.com/resources/inside-marklogic-server/
Presentation: Imagine What You Could Do With Enterprise NoSQL	marklogic.com/resources/imagine-what-you-could-do-with-enterprise-nosql/



About MarkLogic

For more than a decade, MarkLogic has delivered a powerful, agile, and trusted Enterprise NoSQL database platform that enables organizations to turn all data into valuable and actionable information. Organizations around the world rely on MarkLogic's enterprise-grade technology to power the new generation of information applications. MarkLogic is headquartered in Silicon Valley with offices in Washington D.C., New York, London, Frankfurt, Utrecht, and Tokyo. For more information, please visit www.marklogic.com.

© 2014 MarkLogic Corporation. All rights reserved. This technology is protected by U.S. Patent No. 7,127,469B2, U.S. Patent No. 7,171,404B2, U.S. Patent No. 7,756,858 B2, and U.S. Patent No 7,962,474 B2. MarkLogic is a trademark or registered trademark of MarkLogic Corporation in the United States and/or other countries. All other trademarks mentioned are the property of their respective owners. [SS-MLIH-13-06]

999 Skyway Road, Suite 200, San Carlos, CA 94070 ›US: +1 650 655 2300 ›INT'L: +1 877 992 8885
sales@marklogic.com › www.marklogic.com