

# FROM MAINFRAME TO NOSQL

MARKLOGIC WHITE PAPER • JANUARY 2016

The shortest, safest migration path for government systems



## THE OPPORTUNITY FOR TRANSFORMATION

The mainframe architecture has served the archiving, transaction, and system of record needs of government agencies for an impressively long time.

However, as we look to the new horizons of data applications, utilizing the legacy mainframe is a costly and time-consuming hurdle. Although maintaining the current installed system may seem to be the most affordable decision for the immediate financial cycle, analysis of the insidious increases in maintenance costs along with lost opportunities builds a strong case for migrating sooner rather than later.

The legacy mainframe is costly and slows data innovation. Organizations are losing ground every day the mainframe remains at the center of the enterprise. The trend toward reduced complexity, along with platform consolidation and environment rationalization, makes a compelling case for considering Intel® Xeon® processor-based solutions.

## DATA IS THE NEXT COMPETITIVE EDGE

Big data is the topic of conversation across all industries, and for good reason. As far back as 2011, a McKinsey & Company report found that the leaders in big data usage consistently outperformed the rest of their industry in both revenue and earnings<sup>1</sup>. Government agencies have a legacy of managing large repositories of critical data, and are at or near the top consumers of data amongst all enterprise categories<sup>2</sup>.

This push for big data solutions has led to more and more requests to conceive of information in a new way. This new way of looking at the mission can stretch the resources of the mainframe. In many cases, the task is entirely beyond the capabilities of the existing system of record.

Modern expectations require additional transaction and data capabilities. The new standard expects flexible, inclusive, and distributed data

<sup>1</sup> "Big Data, The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute, 2011.

<sup>2</sup> "2013 Global Storage Market Research Report" Frost & Sullivan, 2013

## THE GROWING COST OF DOING NOTHING

Future opportunity can be hard to quantify, but current costs are a serious and measurable issue. While there is a short-term cost to adopting a new technology capable of flexible and creative data implementation, the status quo will be increasingly more expensive. The maintenance of a legacy mainframe is becoming costlier by the year.



As an organization's legacy system ages and widespread use of mainframe technology fades further into history, it will cost more and more to maintain the same level of service.

As the skillsets to maintain the equipment and codebase sunset, programmers and administrators are more difficult to find and increasingly expensive. Following a major data breach, the United States Office of Personnel Management released its 2014 Strategic Information Technology Plan. Along with concerns about security, the report prioritizes migrating away from the legacy mainframe systems because cost of support is expected to rise 10–15 percent annually<sup>3</sup>. And this doesn't even take into consideration opportunity costs of not being able to use mainframe data in new big data solutions.

## THE ALTERNATIVE: DISTRIBUTED DATA CENTERS

While the administration and operation costs of legacy mainframes, including maintenance and new applications, are getting higher, the capabilities and return on investment (ROI) of Intel-based alternatives increase.

<sup>3</sup> Strategic Information Technology Plan, U.S. Office of Personnel Management, 2014.

The Intel-based data center solution ecosystem is being revolutionized by innovation in management, energy efficiency, and performance. Modern Intel-based data centers are able to run more services per staff hour, megawatt, and clock cycle than ever before.

## INTEGRATED DATACENTER TECHNOLOGY

Intel storage, network, and compute technologies are increasing the productivity and performance of data centers.

- **Scalable Processors Designed for Data Centers**  
For example, the Xeon E5-2600 v3 product family, which is designed for the workloads of a data center, includes numerous enhancements that provide performance increases up to three times that of the previous generation<sup>4</sup>, world-class energy efficiency, and enhanced security.
- **Intelligent and Efficient Storage** – The Intel SSD DC P3608 Series solid state drives (SSD) deliver high-performance and low latency. Non-Volatile Memory Express (NVMe) and eight lanes of PCIe 3.0 eliminate bottlenecks in high performance computing workflows, accelerate databases, and provide business insights through real time analytics. The unique dual controller architecture scales efficiently with multiple core Intel Xeon processors.
- **Integrated Network Systems** – With 1 billion Intel Ethernet ports shipped and more than 30 years of innovation, Intel offers a comprehensive portfolio that includes controllers, adapters, IP, and switch silicon. Since the inception of Ethernet, Intel has continued to help drive Ethernet technology by providing technological advances and active leadership in standards committees and industry associations.

## NOSQL: SKIPPING ETL AND SAYING YES

For decades, the art of data has been designing database schemas and the craft of extract, transform, and load (ETL) processes. Standardizing addresses, financial currencies, and other fields is traditionally a key part of any new data integration or consolidation project on a legacy mainframe system of record. Mainframe data often needs to be copied into a

different application-centric database or data warehouse, requiring an ETL process.

ETL projects need to be vetted for accuracy. The content and process need to be kept up-to-date. ETL can be resource intensive, both for staff and compute. For the agency, ETL is cost, delay, and constraints.

The transform portion in particular is resource intensive and complex. In the transforming of data, some aspects of the information can be lost or compromised. For instance, if a building height is calculated using different parameters in different databases and those need to be stored together, the variances might be lost.

When an organization adopts a NoSQL approach to data management, the database imports content in its native state. The ETL process is replaced by the significantly less involved “extract and load.” Sometimes termed “schema-agnostic,” Enterprise NoSQL document databases do not require extensive design, planning, or architecture before introducing a new data source.

## INTEGRATING DATA ACROSS BORDERS

In September of 2015, the US Department of State released a memo calling out their plans for extensive international data collaboration to help achieve sustainable development goals<sup>5</sup>. They plan to integrate data from around the world on topics of health, gender, agriculture, climate, and geospatial.

These initiatives will challenge the abilities of data integrators to bring together a variety of dissimilar data sets and relate them together. As stakeholders become aware of the capabilities of integrated data, they are turning to their information technology staff to create solutions.

## ENABLING EMERGING BIG DATA APPLICATIONS

With modern processing and data tools, organizations are discovering new value in their existing data sets. They create dynamic correlations by combining dissimilar data sources. Enterprise NoSQL is a

<sup>4</sup> As of September 8, 2014. New configuration: Hewlett-Packard Company HP ProLiant ML350 Gen9 platform with two Intel Xeon Processor E5-2699 v3, Oracle Java Standard Edition 8 update 11, 190,674 SPECjbb2013-MultiJVM max-jOPS, 47,139 SPECjbb2013-MultiJVM critical-jOPS. Baseline: Cisco Systems Cisco UCS C240 M3 platform with two Intel Xeon Processor E5-2697 v2, Oracle Java Standard Edition 7 update 45, 63,079 SPECjbb2013-MultiJVM max-jOPS, 23,797 SPECjbb2013-MultiJVM critical-jOPS <http://www.spec.org/jbb2013/results/res2014q1/jbb2013-20140121-00050.html>.

<sup>5</sup> “Harnessing the Data Revolution for Sustainable Development: U.S. Government Commitments and Collaboration with the Global Partnership for Sustainable Development Data” US Department of State, September 28, 2015. Found at <http://go.usa.gov/3zj7ms>

# MarkLogic / Enterprise NoSQL Database Platform

POWERFUL	AGILE	TRUSTED
<ul style="list-style-type: none"> <li>Native JSON Store</li> <li>Native XML Store</li> <li>Native RDF Triple Store</li> <li>Geospatial Support</li> <li>Full-text Search</li> <li>Flexible Indexes</li> <li>Bitemporal</li> <li>Real-time Alerting</li> <li>Semantic Inference</li> <li>Tiered Storage</li> <li>Server-side JavaScript</li> <li>Fully Transactional</li> </ul>	<ul style="list-style-type: none"> <li>Scalable and Elastic</li> <li>Cloud Ready (AWS)</li> <li>Hadoop and HDFS</li> <li>REST API</li> <li>SQL Support</li> <li>Multi-OS Support</li> <li>Schema Agnostic</li> <li>Samplestack</li> <li>MarkLogic Content Pump</li> <li>XA Transactions</li> <li>Ad-hoc Queries</li> <li>Index Across Data Types</li> </ul>	<ul style="list-style-type: none"> <li>Performance at scale</li> <li>LDAP and Kerberos Security</li> <li>Security Certifications</li> <li>Monitoring and Management</li> <li>Configuration Management</li> <li>24/7 Engineering Support</li> <li>ACID Transactions</li> <li>Flexible Replication</li> <li>Customizable Backup</li> <li>Customizable Failover</li> <li>Point-in-time Recovery</li> <li>Atomic Forests</li> </ul>

Enterprise NoSQL is a trusted platform that cost-effectively responds to modern business expectations.

compelling enabler of future data applications. An opportunity that requires the consumption of JSON, XML, unstructured files, text, RDF triples, geospatial, or any binary can be implemented rapidly without the pre-work of ETL and schema design.

A million heterogeneous documents can be corralled to answer a complex question such as those government agencies face. For example, the World Bank recently needed to determine “Where should the World Bank approve loans that will enhance agricultural incomes of small farmers and ensure increased access to services related to health, nutrition, sanitation and social entitlements?”<sup>6</sup> No matter the format each system uses, data and documents from a variety of sources can be combined to get a good result.

## MISSION-CRITICAL APPS ON NoSQL

Designed for durability and scalability, Enterprise NoSQL technology from MarkLogic supports massively scaled applications around the world. MarkLogic® Enterprise NoSQL is the platform for organizations in

need of a system of record capable of high volume and high flexibility. Many companies are already realizing the benefits of MarkLogic Enterprise NoSQL.

### TOP 5 BANK

32 million live deals	1,600 requests per second
-----------------------	---------------------------

### BBC STREAMING 2012 OLYMPIC COVERAGE

55 million global browsers	2.8 petabytes
----------------------------	---------------

### HEALTHCARE.GOV

5,000 highly complex transactions a second	11+ million Americans with health coverage
--	--

### CONDÉ NAST

7+ million rich media assets	<1 second search results
------------------------------	--------------------------

<sup>6</sup> World Bank Press Release, Dec 22, 2014: World Bank Approves \$75 Million Project in Andhra Pradesh, India – Expected to Improve Incomes and Health of Farmers found at <http://www.worldbank.org/en/news/press-release/2014/12/19/world-bank-75-million-project-andhra-pradesh-india-incomes-health-farmers>

## DISTRIBUTED AND FLEXIBLE DATA CENTER ARCHITECTURE

When dealing with field offices, new service offerings, and changing business climates, organizations need to be able to affordably grow and transform IT infrastructure. With a mainframe, one can add processing and storage, but adding secondary locations is not so easy. Architecting for bursts and peaks means essentially purchasing compute for the maximum need.

Organizations adopting Intel® Xeon® processor-based data centers can manage growth and opportunity using modern flexible computing techniques. Virtualization of servers and infrastructure allows for an unprecedented ability to maximize resource use.

Bursts in demand and new initiatives mean that an IT organization needs to be able to cost-effectively increase compute capacity, perhaps for only a brief period of time. Applications developed on Intel-based solutions have options. Along with reallocating resources within the data center, hybrid cloud solutions can extend the application to a platform as a service (PaaS) or an infrastructure as a service (IaaS) vendor. MarkLogic is architected to handle elasticity and new distribution deployments without downtime.

## MIGRATING THE SYSTEM OF RECORD FROM A LEGACY MAINFRAME TO NOSQL

### ENTERPRISE NoSQL + INTEL-BASED SOLUTIONS

- **Are ready for new integrations**
- **Are designed for distributed workloads**
- **Scale up and down elastically**
- **Lower management costs**
- **Lower development costs**
- **Lower TCO costs**

Achieving the promise of a new technology usually means a migration process. Migrating a system of record, the heartbeat of an organization, is no small task and not one to be taken lightly. As the market moves away from mainframe solutions, actionable guidance and help is available to the organization that embarks on a migration path.

Intel offers best practices for the migration planning and execution in its paper “Migration from UNIX/RISC and Mainframe to Intel-Based Solutions.”

### HSBC RE-HOSTS MODERN MAINFRAME APPLICATIONS

HSBC encountered performance and reliability issues for a key mainframe after deploying a new suite of loan applications. Migrating the applications to Intel Xeon processor-based servers delivered:

- A better experience for customers and branch office personnel due to improved performance and higher uptime
- Lower costs and greater headroom, with 70 percent lower monthly service charges and a 2,000 MIPS reduction in mainframe workloads
- Improved business flexibility with an infrastructure that is easier to scale and adapt

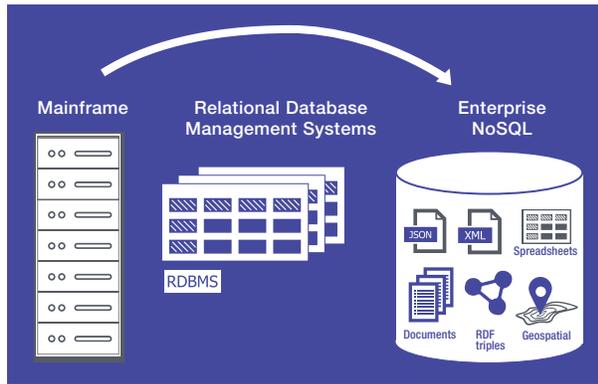
Read the full case study for more information about the HSBC methodology and migration experience at <http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/data-center-xeon-hsbc-whitepaper.pdf>.

For more guidance on migration to an Intel architecture data center, visit the Intel [IT Center for Server Migration](#)

“ MarkLogic has now serviced four of our county initiatives, and I know that there are several other applications we are considering using MarkLogic for as well.”

— Jim Callahan, Lead Application Developer, Fairfax County

## LEAPFROG RDBMS: ENTERPRISE NoSQL IS A MORE COMPELLING DESTINATION



Migrating to Enterprise NoSQL provides greater capabilities and a less disruptive migration path than RDBMS.

The legacy mainframe applications that are still running today have most likely survived despite previous challenges. Throughout the years, IT departments have evaluated migrations to relational database management systems (RDBMS). Many times, they chose to leave the mainframe in place.

Since the 1990s, relational databases have become the expected database solution for most data projects. Still, there is a huge base of applications still running on thousands of mainframe systems around the world. They were not migrated for very good reasons: migration is costly, disruptive, and risky.

Enterprise NoSQL, along with offering extended capabilities, has a safer, less intensive migration path and offers greater data agility to ensure the data is future-proofed for the next 40 or more years. With a more flexible information storage, less transformation means less time and less risk.

## CASE STUDY: IMPROVING PRODUCTIVITY AND TCO

One of the largest counties in the nation, Fairfax County, Virginia, sought to make it easier for county employees, land developers, and residents to access real-time information about zoning changes, county land ordinances, and property history. There were volumes of data in disparate databases and file systems and in different formats that couldn't be effectively searched.

After implementing MarkLogic, the county was able to launch an application that boosted the productivity of county employees and reduced TCO. In addition, it also gives residents an online source to access information on their computers or smart phones.



With Enterprise NoSQL, Fairfax County integrated data from multiple diverse sources to provide better service and reduce costs.

## MEASURING SUCCESS

For most government IT departments, a reliable system of record is the expected baseline. Surveyed IT departments have found that their internal customers can be the source of the most resistance to a migration. This is a reflection of the importance of reliability. Job one is to preserve or enhance the customer experience in line of business applications.

“MarkLogic not only provided us with a fast database engine to search content but allowed us to ingest ad hoc data, which during emergencies can't be planned for or schematized, that is then accessed by national headquarters for de-confliction and dissemination. Having a schema-less and flexible data model is a good foundation for our solution.”

—Will Lawrence, Chief Architect, FAA

For IT departments and executives, the financial bottom line is an obvious gauge of success. Reducing the cost of support and of developing new applications can be a path to realizing return on investment for the migration efforts.

Enterprise NoSQL on Intel-based solutions is an enabling technology for future applications: those currently planned and those not yet imagined. For the future, an organization can realize the value of an integrated data repository that provides access to all aspects of the organization's data profile. An Enterprise NoSQL solution is designed to quickly integrate new data sources as they appear, no matter the format or source.

### **IT CAN BE DONE; IT HAS BEEN DONE: COPYBOOKS TO XML**

While data is stored and managed differently in a modern NoSQL database, the XML-friendly structure allows for some logical migration paths. Ken Krupa, MarkLogic CTO, utilized the COBOL from his early career to [document a process](#) that will reliably ingest COBOL copybooks into the MarkLogic Enterprise NoSQL database.

applications. MarkLogic is headquartered in Silicon Valley and has offices throughout the U.S., Europe, Asia, and Australia. For more information, please visit [www.marklogic.com](http://www.marklogic.com).

## **ABOUT INTEL**

Intel (NASDAQ: INTC) is a world leader in computing innovation. The company designs and builds the essential technologies that serve as the foundation for the world's computing devices. As a leader in corporate responsibility and sustainability, Intel also manufactures the world's first commercially available “conflict-free” microprocessors. Additional information about Intel is available at [newsroom.intel.com](http://newsroom.intel.com) and [blogs.intel.com](http://blogs.intel.com), and information about Intel's conflict-free efforts is at [conflictfree.intel.com](http://conflictfree.intel.com).

## **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