SOLUTION BRIEF

Big Data Analytics

Government, Financial Services, Healthcare, and more

Real-Time Performance For Enterprise Applications

"The Intel® Xeon® processor E5-2699 v4, combined with the Intel® Solid-State Drive Data Center P3700 Series, provides substantial gains in query performance for our enterprise NoSQL database and supports fast, scalable data ingestion. The benefits are significant for real-time searches, queries, and Internet-scale applications."

- Caio Milani, Director of Product Management, MarkLogic

The limitations of traditional databases have become a painful reality for businesses and government agencies working to integrate, analyze, and use their fast-growing data sets. NoSQL databases provide a path forward, but the great majority of these modern databases are not designed to address enterprise requirements.

With the MarkLogic* Enterprise NoSQL Database, the Intel® Xeon® processor E5-2699 v4 and the Intel® Solid-State Drive Data Center Family, you can:

- **Eliminate data silos** to provide a real-time, 360-degree view of your data.
- **Support real-time** applications at almost any scale.
- **Achieve up to 2.2x faster query performance**[3] to deliver better and more reliable user experiences with fewer servers.
- **Provide scalable, high-speed data ingestion** to support real-time analytics on massive amounts of fresh, fast-moving data.

The MarkLogic* Enterprise NoSQL Database running on Intel® architecture offers a compelling solution that can be implemented now. This blazingly fast data platform provides a flexible data model for storing, managing, and searching today's data. It also supports the levels of reliability, security, and data governance needed to replace enterprise relational databases and mainframes.[1]

For more than a decade, MarkLogic running on Intel architecture has been helping businesses build next-generation applications on a unified, 360-degree view of their data. Organizations as diverse as the BBC, a top 5 investment bank, and the U.S. Department of Health and Human Services have used it as a massively scalable data platform for real-time, public-facing applications.[2]

MarkLogic can be scaled out on Intel® Xeon® processor-based servers to provide a single, indexed data store for hundreds of terabytes of structured and unstructured data. As a schema-agnostic database with integrated support for semantics, MarkLogic simplifies application development and

allows all data types to be integrated as is, with no need for the cost and delays of traditional extract, transform, and load (ETL) procedures. Fresh data is instantly available for analysis and use in real-time environments.

The ability of MarkLogic to deliver extreme performance and scalability on standards-based Intel architecture offers compelling advantages. According to Jon Bakke, the executive vice president of Worldwide Field Operations for MarkLogic, "During the planning phases for Healthcare.gov, there was a moment when the government was ready to invest about $7.5 million in new hardware to run the MarkLogic instance, and I had the pleasure of recommending they not do that. We were confident that MarkLogic could work on the virtualized servers, the network-attached storage, and the infrastructure they had if it were configured properly. Once they were able to do that, we were proven right."

Dramatic Performance Gains for Users...and for IT

Up to 2.2x faster query performance[3] with the Intel® Xeon® Processor E5-2699 v4

The latest Intel® Xeon® processor E5-2600 v4 product family is ideal for the intense computing demands of MarkLogic. These processors provide up to 22 cores, 44 threads, and 55 MB of last-level cache. They also support new DDR4 memory options that, based on published specifications, are up to 12.5 percent faster than previous-generation DDR4 memory.[4] According to Caio Milani, director of product management for MarkLogic, "The faster memory and larger cache are a real advantage for query performance, especially when semantics are used to provide greater context within the stored data."

A recent benchmark study by Intel and MarkLogic quantified the performance benefits of this new processor family against the two previous Intel Xeon processor generations. A cluster of three servers powered by the Intel Xeon processor E5-2699 v4 provided up to 1.3x faster query performance versus a similarly configured cluster based on the Intel® Xeon® processor E5-2697 v3. The gains were much higher—up to 2.2x faster—compared to the Intel® Xeon® processor E5-2697 v2, which is still widely deployed in many data centers. (To provide additional context, Intel estimates that a comparable, new-generation Intel® Xeon® processor E5-2697 v4 would provide up to 1.1x faster performance than the Intel Xeon processor E5-2697 v3.)

Proven Performance Gains Across Multiple Processor Generations

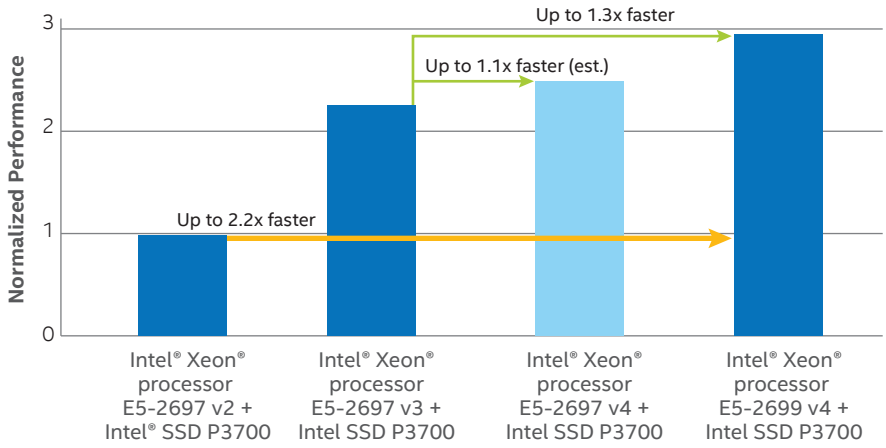

Figure 1. Through joint software optimization, MarkLogic and Intel continue to deliver substantial performance gains with every new Intel® Xeon® processor generation.[3]

These performance gains offer clear value in production environments. They can help businesses deliver better customer experiences, even as they contain costs and infrastructure complexity by using fewer servers to handle heavier workloads. Even more importantly, the additional headroom makes it easier to handle the kinds of unexpected spikes in demand that often come with big events, market shifts, and successful marketing campaigns. Enterprise applications are more likely to remain responsive during sudden surges that might otherwise overload or even crash the system.

Near-Instant Access to Massive, Fast-Moving Data Flows

High-speed data ingestion with the Intel® Solid-State Data Center P3700 Series

Fast ingestion of large data volumes is critical for supporting real-time, context-based applications, such as personalized content delivery and portfolio monitoring. MarkLogic is designed to load, tokenize, and index large volumes of data at high speed. The increased cache and core counts of the latest Intel Xeon processors can be valuable if complex data transformations are required. In most cases, however, storage I/O throughput and latency are most important for this process.

The Intel® Solid-State Drive (Intel® SSD) Data Center Family for PCIe* offers a valuable resource for accelerating data ingestion. Based on the NVMe* specification, these Intel SSDs support extreme levels of data throughput with low latency.[5] By using one or more Intel SSDs in combination with any number of commodity hard-disk drives (HDDs), organizations can meet demanding data ingestion requirements simply and cost-effectively. They can also scale performance and capacity independently as requirements grow, by adding additional Intel SSDs to scale performance and additional HDDs to scale capacity.

Ramping Up Value through Collaborative Optimization

MarkLogic and Intel engineers collaborate extensively to optimize MarkLogic software for the latest innovations in each new Intel Xeon processor generation. Together the team works to improve the efficiency of parallel execution across increasing numbers of cores, threads, and server nodes, and to make better use of the larger cache sizes. Intel engineers help to identify the performance-critical code segments that will benefit most from the new resources. These optimization efforts help to ensure that new innovations in each processor generation translate into real-world performance gains, so organizations get higher value from every new server.

Harness All Your Data

MarkLogic Enterprise NoSQL Database running on the Intel Xeon processor E5-2699 v4 and Intel SSDs provides a fast, flexible, and resilient foundation for meeting the data demands of enterprise applications. By supporting transactions, searches, and queries on a single, massively scalable data repository, this enterprise NoSQL data platform can help organizations achieve a 360-degree view of their data without sacrificing security or data governance.

Just as importantly, MarkLogic on Intel architecture can help organizations scale their data platform incrementally and almost without limit as demands grow, so they can integrate new data sources, handle increasing workloads, and take advantage of new opportunities.

Learn More

MarkLogic NoSQL Database

- Overview: www.marklogic.com/
- Customer success stories: www.marklogic.com/customer-type/testimonials/

Intel Xeon processor E5-2600 v4 product family

- Overview: www.intel.com/content/www/us/en/processors/xeon/xeon-processor-e5-family.html
- Software solutions: www.intel.com/XeonE5SoftwareSolutions
- Intel® Solid-State Drive Data Center Family: www.intel.com/content/www/us/en/solid-state-drives/data-center-family.html

[1] Read the MarkLogic white paper, "From Mainframe To NoSQL, Migrating the system of record to a hierarchical NoSQL solution," published January, 2016, www.marklogic.com/solutions/mainframe/

[2] For these and other MarkLogic customer success stories, visit the MarkLogic website at http://www.marklogic.com/customer-type/testimonials/

[3] Performance on the MarkLogic CPoX (Content Processing over XML) benchmark measured by Intel, February 8, 2016. Tests measured query performance for a large set of Wikipedia* documents in multiple languages using a cluster of three two-socket server nodes with three different server platforms: (1) a baseline server configured with 2 x Intel® Xeon® processor E5-2697 v2 (12 cores, 30 MB cache, 2.7 GHz), Result = 1860 queries per second; (2) a previous-generation server configured with 2 x Intel® Xeon® processor E5-2697 v3 (14 cores, 35 MB cache, 2.6 GHz), Result = 3109 queries per second; (3) a new server configured with 2 x Intel® Xeon® processor E5-2699 v4 (22 cores, 55 MB cache, 2.2 GHz), Result = 4113 queries per second (1.32x faster than the previous-generation server, 2.21x faster than the baseline server). All servers were configured with 128 GB of memory, 8 x hard-disk drives (1 TB, 7200 RPM), seven for data, one for the operating system, 1 x Intel® Solid-State Drive Data Center P3700 Series, 10 GbE Network, MarkLogic 8.0-3.2, CentOS* 6.6. Performance estimates of 1.13x for the Intel® Xeon® processor E5-2697 v4 versus the previous-generation server were based on the assumption that per-core performance would be the same on the Intel Xeon processor E5-2697 v4 as on the Intel Xeon processor E5-2699 v4.

[4] Source: Intel product specifications. The Intel® Xeon® processor E5-2600 v4 product family supports memory speeds up to 2400 MT/s versus maximum memory speeds of 2133 MT/s for the Intel® Xeon® processor E5-2600 v3 product family.

[5] For more information about the Intel SSD DC family, visit http://www.intel.com/content/www/us/en/solid-state-drives/data-center-family.html

Printed in USA 0315/KE/MESH/PDF ♻ Please Recycle 334092-001US