Intel and IBM Redefine the Limits of Data Warehousing

"The leadership performance of DB2 on new Intel Xeon processor 7500 series–based systems is the direct result of the deep collaboration between IBM and Intel. Together, we have achieved up to double the performance of previous-generation systems through full and efficient utilization of all available cores, threads, and memory to deliver even more value for our customers."

—Berni Schiefer
Distinguished Engineer,
Information Management Performance and Benchmarks, IBM

A performance-optimized data warehouse requires the right balance of database, server, and storage resources. The complexity involved in achieving this simple realization is the driving force behind a multi-year collaboration between IBM and Intel, who share the objective of delivering exceptional value—faster—to customers in need of business intelligence solutions. This is the underlying principle of the IBM® InfoSphere™ Balanced Warehouse™ portfolio, and a core requirement for next-generation analytics systems.

With more organizations choosing XML as their preferred transaction format, database administrators and warehouse architects are being challenged to integrate massive volumes of business-critical XML data into more sophisticated decision-support systems. Common design considerations for many of today's largest high-performance data warehouses already include the flexibility of the IBM DB2® pureXML® capability and the robustness of IBM System Storage™ solutions, but may not account for the latest innovations available from servers based on Intel architecture.

To this end, IBM and Intel have collaborated over multiple generations of products to optimize and measure the ever-increasing XML transaction processing capability of DB2 software on Intel® Xeon® processors, beginning nearly five years ago with single-core CPUs and humble 50 GB XML data sets. The latest performance tests reveal that multi-core Intel Xeon processors can indeed handle 1 TB XML repositories with ease.

But beyond transaction processing performance, measuring data warehouse performance is more about proving the solution's capacity to scale the performance of complex queries in tandem with the increasing data sizes those queries address. Can an Intel-based solution support a growing, multi-terabyte data warehouse without sacrificing query response times? To answer this question, IBM and Intel constructed a decision-support system based on high-performance components available in the market today, and extended the XML data size to 10 TB.

IBM DB2 9.7: The Heart of IBM InfoSphere Warehouse

IBM DB2 9.7 software ("DB2 9.7") represents the next generation of relational database servers, offering design innovations that make it well suited for demanding, mission-critical applications. For example, applications developed for service-oriented architectures should be able to scale with ease, be continuously available, and natively manage XML; DB2 software is engineered to handle all three requirements.

Sophisticated XML Data Management with IBM DB2 pureXML

Built on the high-performance XML data management capabilities of DB2 9.7, IBM InfoSphere Warehouse 9.7 software ("InfoSphere Warehouse 9.7") is optimized for data manipulation, query and retrieval, and data storage in a highly scalable, highly available, and secure architectural framework. Database design options for XML data—such

as hash partitioning, range partitioning, and multidimensional clustering—can help improve scalability and query performance, assist developers in exploiting parallel-processing environments, and simplify the addition and removal of time-sensitive data.

To support business collaboration and enhance application programmer and administrator productivity, DB2 9.7 offers IBM DB2 pureXML data schemas and scripts for major industry-specific XML messages, including Financial Information Exchange Markup Language (FIXML), Health Level Seven (HL7), Association for Cooperative Operations Research and Development (ACORD), News Markup Language (NewsML), and Human Resources XML (HR-XML).

IBM System Storage DS8700: Resiliency, Performance, Scalability, Value

The IBM System Storage DS8700 model ("DS8700") is the most advanced model in the IBM DS8000® lineup, offering a new level of performance for the company's flagship enterprise disk platform. Enabling up to 2.5 times the performance of the previous model, the DS8700 is designed to support demanding business applications with its unparalleled data throughput and resiliency features and five-nines availability (99.999 percent).[1] Moreover, with its tremendous scalability, new storage tier optimization, and broad server support, the DS8700 can help simplify the storage environment by consolidating disparate storage platforms onto a single system. At the same time, it provides the availability and performance levels required by your most important business applications.

The DS8700 model is a testament to IBM's reputation for outstanding quality and world-class engineering. And in today's challenging economic climate, having an enterprise disk platform that combines high levels of system and application availability with superior performance, flexibility, and total cost of ownership is essential.

Intel Xeon Processor 7500 Series: Powering Performance

Built to handle the most demanding applications, the Intel Xeon processor 7500 series delivers a quantum leap in enterprise computing performance. With up to 8 cores and 16 threads per processor, and a new high-bandwidth interconnect system that enables extreme scalability, this new platform is a top choice for supporting almost any high-performance data warehousing environment.

Green IT organizations can also take advantage of industry-leading virtualization performance and a server consolidation ratio of 20:1 to get more done with fewer servers, reduced power consumption, and less overhead.[2]

Mission-critical data warehouses will benefit from Intel Advanced Reliability Technology, a new feature in the Intel Xeon processor 7500 series that provides automatic detection and correction of errors, dynamic reassignment of workloads across CPUs, and even individual virtual machine recovery in consolidated environments—everything you need to help maintain data integrity, minimize downtime, and maximize productivity.

IBM DB2 and Intel Together: Doubling Performance

IBM and Intel have worked together for years to deliver multiple XML benchmarks featuring DB2 software on Intel Xeon processors, utilizing the industry-supported Transaction Processing over XML (TPoX) workload.

Figure 1 demonstrates that customers with DB2 9.7 deployed on previous-generation Intel Xeon processor–based servers can realize double the performance by upgrading to systems powered by the Intel Xeon processor 7500 series, such as the IBM System x® 3850 X5 server.[3]

Figure 1 also shows that the Intel Xeon processor 7500 series offers better virtualization efficiency. The amount of

"The new Intel Xeon processor 7500 series delivers to our mutual clients a new level of value for their business intelligence solutions. In collaboration with Intel, IBM has demonstrated InfoSphere Warehouse scaling linearly up to 10 TB of XML data on IBM System Storage DS8700—with similar query response times. This kind of efficient scalability improves the economics of solving complex business problems with IBM analytics solutions on Intel architecture."

—Salvatore Vella
Vice President of Development, Database Servers and Data Warehousing, IBM

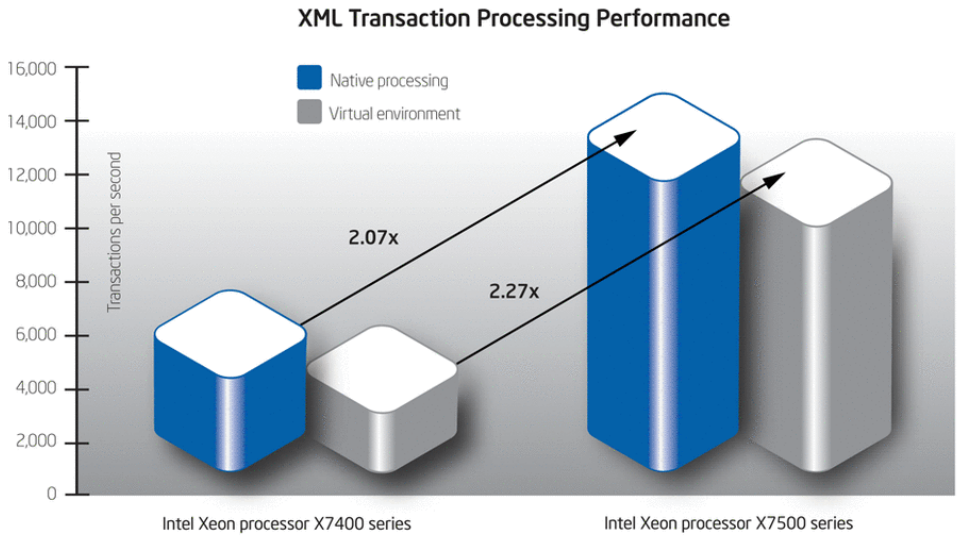

Figure 1: Compared to the previous-generation Intel Xeon processor 7400 series, XML transaction processing is doubled for both native and virtual environments on the Intel Xeon processor 7500 series.

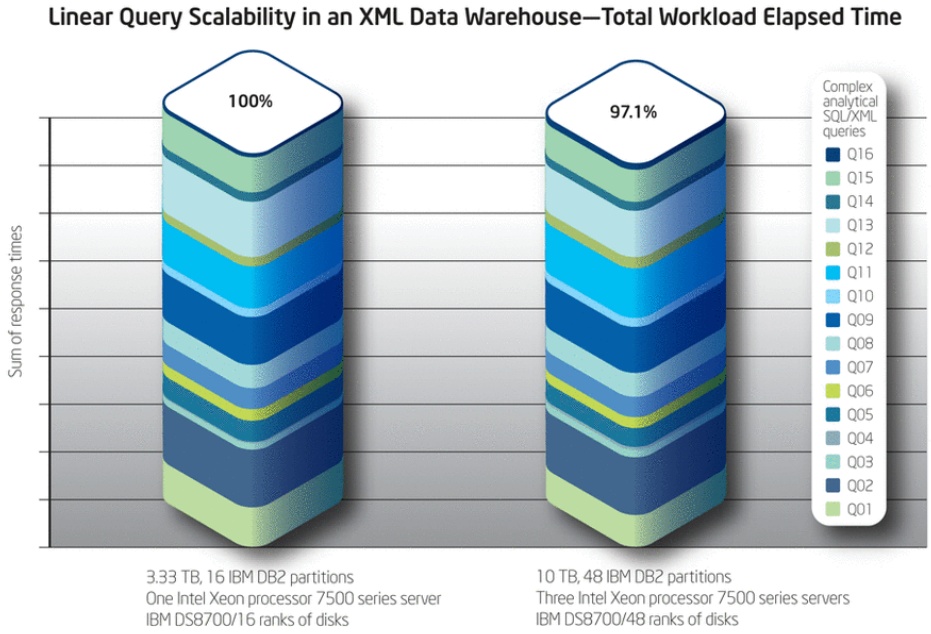

Figure 2: With the IBM-Intel solution, complex queries maintain consistent response times as both the size of the database and the number of servers triple.

processing overhead consumed by VMware on the Intel Xeon processor 7500 series is lower than the amount consumed on the previous Intel Xeon processor 7400 series–based hardware.[4]

These benchmarks are strong proof points of the value of the collaborative engineering efforts by IBM and Intel to maximize the innovations in DB2 9.7 and the Intel Xeon processor 7500 series.

Large XML Data Warehouses Become a Reality with Intel, IBM InfoSphere Warehouse, and IBM Storage

To determine if an Intel-based solution could sufficiently support a growing, multi-terabyte data warehouse, Intel and IBM paired InfoSphere Warehouse 9.7 with servers based on the Intel Xeon processor 7500 series and IBM System Storage DS8700. With this powerful reference configuration, the team studied response times for 16 complex analytical SQL/XML queries while growing the XML warehouse from 3.33 TB to 10 TB of raw data, representing up to 5.5 billion XML documents.

Figure 2 shows query response times at 10 TB that are nearly identical to those at 3.33 TB. In fact, IBM and Intel measured a difference in query performance of less than 3 percent between the two scale factors of the reference configuration.[5]

These initial tests demonstrate that IBM InfoSphere Warehouse 9.7 paired with the new Intel Xeon processor 7500 series and IBM System Storage DS8700 can provide linear query scalability as XML data warehouse size and complexity increase. Together, IBM and Intel have shown that as Intel compute resources are scaled in tandem with the data volume, the performance of the decision-support workload remains constant and predictable—a pivotal consideration for any high-performance data warehouse deployment.

Learn More

More about IBM DB2 9.7: www.ibm.com/db2

More on the Intel Xeon processor 7500 series: www.intel.com/itcenter/products/xeon/7500/index.htm

More about IBM System Storage DS8700: www.ibm.com/systems/storage/disk/ds8000/index.html

More about the TPoX 2.0 benchmark: http://tpox.sourceforge.net

---

[1] Source: IBM DS8000 performance team internal testing results. www-03.ibm.com/systems/storage/disk/ds8000/index.html

[2] Claim: "up to 20:1 server consolidation" based on comparison between 4S Intel® Xeon® MP CPU 3.3 GHz (single core with HT, 1 MB L2, 8 MB L3, Potomac) and 4S Intel® Xeon® X7560 2.26 GHz (8 core)–based servers. Calculation includes analysis based on performance, power, cooling, electricity rates, operating system annual license costs, and estimated server costs. This assumes 42U racks, US$0.10 per kWh, cooling costs are 2x the server power consumption costs, operating system license cost of US$900/year per server, per server cost of US$36,000 based on estimated list prices, and estimated server utilization rates. All US dollar figures are approximate. SPECint_rate_base2006* performance and power results are measured for X7560 and Xeon 3.3 GHz–based servers. Platform power was measured during the steady state window of the benchmark run and at idle. Performance gain compared to baseline was 20x.

–Baseline platform (measured score of 33.8): Intel server with four Intel Xeon MP CPU 3.3 GHz (single core with HT, 1 MB L2, 8 MB L3) processors, 16 GB memory (8x2 GB DDR2-400), two hard drives, one power supply, using Red Hat EL 5.3 x86_64 operating system.

–New platform (measured score of 705): Intel internal reference server with four Intel Xeon processor X7560 (24M cache, 2.26 GHz, 6.40 GT/s Intel QPI, Intel Hyper-Threading Technology, Intel Turbo Boost Technology), 256 GB memory (64x4 GB QR DDR3-1333), one hard drive, two power supplies, using SuSE* LINUX 11, cpu2006.1.1.ic11.1.linux64.binaries.nov242009.tar.bz2 binaries.

[3] Source: Intel internal measurements as of February 2010. All of the benchmark results presented are derived using the TPoX 2.0 mixed benchmark simulating an OLTP workload with users running a mixture of read-only queries, inserts, deletes, and updates over a two-hour period. The resulting database size was approximately 1 TB. Platform: server with Intel Xeon processor X7500, 2.27 GHz; 256 GB versus Intel Xeon processor X7400, 2.67 GHz; 64 GB. Both platforms were running IBM DB2 9.7 on Linux SLES10-SP2 64-bit with DB2 Compression and STMM.

[4] Same workload as footnote 3, except that IBM DB2 9.7 on Linux SLES10-SP2 64-bit was run with VMware* ESX* 4.0 Update 1. Platform: server with Intel Xeon processor X7500, 2.27 GHz; 256 GB versus Intel Xeon processor X7400, 2.67 GHz; 128 GB.

[5] Source: IBM internal measurements as of March 2010. All of the benchmark results presented are from a database with schema extended from the TPoX 2.0 schema. The benchmark was run with a series of 16 complex analytical SQL/XML queries. The resulting database size was built from approximately 10 TB data. Platform: three servers each with four Intel Xeon processor X7500, 2.27 GHz; 128 GB, running IBM DB2 9.7 on Linux RHEL 5.4 64-bit with DB2 Compression and STMM. DS8700 configured with 384 300 GB 15K RPM Fibre Channel drives.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments might vary significantly. Users of this document should verify the applicable data for their specific environment.

Intel, the Intel logo, and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. IBM, the IBM logo, Balanced Warehouse, DB2, DS8000, InfoSphere, pureXML, System Storage, and System x are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Copyright © 2010 Intel Corporation. All rights reserved.