# IBM, Intel Accelerate Terabyte-Class XML Database Processing

Enterprises with large XML processing needs don't have to sacrifice performance or scalability to take advantage of XML-based technology, according to recent Intel tests on servers based on the new Intel® Xeon® processor E7 family. Combined with IBM® DB2® technology, the new Intel Xeon processor E7-4870—which achieved more than 17,750 transactions per second on the Transaction Processing over XML (TPoX) 2.0 benchmark—creates a platform for fast, efficient processing of the type of XML database transactions that are increasingly common in enterprise applications.

## Strong Database Performance for New XML Processing Demands

These findings are good news for the continued success of web applications, service-oriented architectures (SOAs), and electronic data exchange between organizations. In particular, XML has emerged as a de facto standard format for electronic business records and messages, because this data format is self-describing and platform-independent, and therefore makes an easy platform for document exchange.

In addition, XML is an extensible and flexible data format, which makes it adaptable for evolving business needs and changing data processing requirements. The flexibility XML offers makes rapid, systematic content repurposing and reuse relatively easy.

## DB2 pureXML: New, Faster XML Database Storage

Moreover, the trend toward storing XML permanently in databases is rapidly accelerating, driven by new requirements for auditing and compliance; the need for a more flexible and suitable data format than a rigid relational database schema can provide; and the desire to simplify applications and boost their efficiency with XML storage. Companies in virtually every industry—insurance, finance, publishing, engineering, and healthcare, to mention a few—are now using IBM DB2 pureXML® to align their database strategies with their web applications and SOAs, thereby achieving greater agility and interoperability.

DB2 pureXML also manages complex vehicle information in the automobile sector, as well as the integration of sales figures in retail companies and order management in the telecommunications industry. And government agencies employ DB2 pureXML to manage electronic forms where the variability is so great that only the XML data format can reasonably handle them.

## Speed and Agility with Intel Processors, IBM Technology

Even extremely performance-sensitive businesses such as those that use stock-trading programs—where a few microseconds could mean the loss or gain of millions in revenue—are implementing XML-based systems. The directors of these businesses know they can now have flexible systems that can be changed quickly and easily as business demands change. The speed and agility they need to meet shifting business demands is available thanks to continued breakthrough performance by Intel microprocessors and IBM DB2 pureXML technology.

IBM data management
KNOWLEDGE. PERFORMANCE. RESULTS.

| | Intel Xeon X7460 | Intel Xeon X7560 | Intel Xeon X7560/SSDs | Intel Xeon E7-4870/SSDs |
|---|---|---|---|---|
| TPoX transactions per second (TTPS) | 6,654 | 13,743 | 14,271 | 17,757 |
| Users | 220 | 420 | 420 | 440 |
| CPU % | 94 | 96 | 98 | 94 |
| Average I/O latency[1] (ms) | 6.17 | 7.15 | 1.57 | 3.72 |
| Scalability | N/A | 2.07 | 1.04 | 1.24 |
| Processors | 4 | 4 | 4 | 4 |
| L3 cache per processor (MB) | 16 | 24 | 24 | 30 |
| Cores | 24 | 32 | 32 | 40 |
| Threads | 24 | 64 | 64 | 80 |
| Frequency (GHz) | 2.67 | 2.27 | 2.27 | 2.4 |

*Table 1: TPoX performance statistics for the Intel Xeon family of processors.*

For example, some of the world's leading investment companies are storing and querying XML messages with financial trading information in DB2 pureXML—often using financial XML standards such as FpML or FIXML—which is exactly what the TPoX 2.0 benchmark simulates.

And when companies keep XML in persistent storage, they typically need to insert, index, query, and update XML, all with the same performance, scalability, and ACID properties (atomicity, consistency, isolation, and durability) that relational databases have long offered for traditional relational data.

In response to these needs, DB2 9.1 and later versions offer sophisticated XML capabilities with support for SQL/XML, XQuery, XML Schemas, XSLT, and other XML-related standards. And as the latest tests show, DB2 pureXML—running on servers built on the latest Intel processors—can handle the extra demands of these new standards with speed and near-perfect scalability.

## The TPoX 2.0 Benchmark: Simulating a Real-World Trading Application

The TPoX 2.0 benchmark mimics a financial trading application, with traders placing buy and sell orders and checking the quickly changing prices and availability of securities. These TPoX tests operate on a terabyte of XML data—a midrange scale factor setting for XML transactional databases—and provide a formidable exercise of server and database technology. For more information on TPoX,

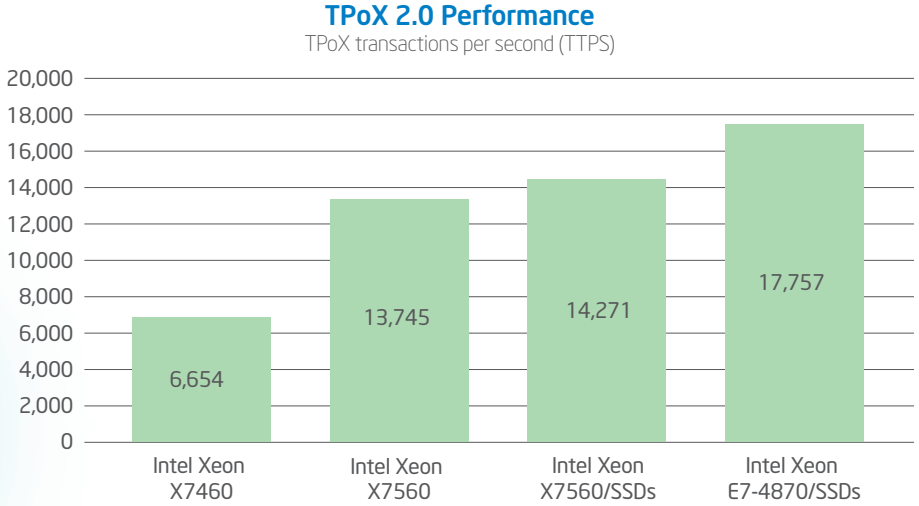


*Figure 1: Performance of different processors in the Intel Xeon processor family. The new Intel Xeon E7-4870, tested with solid-state drives, achieved more than 17,750 transactions per second on the Transaction Processing over XML (TPoX) 2.0 benchmark.*

see the sidebar, "The TPoX 2.0 Benchmark: Simulating XML-Based Database Applications," and check out the benchmark and associated documentation at http://tpox.sourceforge.net.

IBM and Intel conducted these tests for the Intel Xeon processor X7460, Intel Xeon processor X7560, and the latest addition to the Intel Xeon processor family, the Intel Xeon processor E7-4870. Because the throughput of the benchmark is pushed higher with the increasing performance of new generations of Intel processors, the performance demand on the storage system increases accordingly. To better manage this phenomenon, the Intel Xeon processor E7-4870 in this test uses Intel® Solid-State Drives (Intel® SSDs). To provide a fair comparison, the Intel Xeon processor X7560 was also tested with the Intel® X25-E SSD.

## Scalable, High Performance for XML-Based Transaction Processing

The new Intel Xeon processors faced stiff competition from their predecessors, which were already extremely fast. But the Intel Xeon processor family X7560—sporting a radically new architecture and simultaneous multi-threading for quicker processing—has only 25 percent more cores than its predecessor, the Intel Xeon processor family X7460, yet demonstrated twice the performance of the previous generation on the terabyte benchmark (see Table 1).

The Intel Xeon processor E7-4870 follows Intel's tradition of continuous improvement by increasing performance an extra 24 percent, largely due to 25 percent more cores than the X7560, making it ideal for high-speed XML database access and database server consolidation. Figure 1 depicts these results.

## The TPoX 2.0 Benchmark: Simulating XML-Based Database Applications

Like many enterprise applications, the TPoX 2.0 benchmark requires processing large amounts of complex information extremely quickly. Here are the benchmark's most important characteristics:

- TPoX is an application-level XML database benchmark that executes a concurrent multiuser workload to simulate financial transaction processing. In TPoX, financial orders are represented as FIXML messages, a real-world XML format used in financial applications.
- TPoX is Java*-based, database agnostic, and open source, with an extensible workload driver and an XML-based configuration to express the transaction mix.
- The database schema includes three tables, all with a single XML column: **security**, **custacc** (customer account), and **orders**.
- The workload mix comprises four types of transactions: queries, inserts, updates, and deletions. Seventy percent of the workload is queries; thirty percent is inserts, updates, and deletes.
- The listed score is taken in a steady state and reported as "TPoX transactions per second" (TTPS).
- TPoX includes a workload driver that spawns parallel threads that simulate concurrent database users. Each "user" connects to the database and submits a mix of transactions. The transactions are picked randomly from a set of transaction templates. At runtime, parameter markers in the templates are replaced by actual values drawn from configurable random value distributions. The workload driver collects and reports performance metrics, such as min/max/avg response time and overall throughput.

The following rules are used in defining the **update**, **delete**, and **insert** transactions:

- Customer accounts are updated to reflect trades (execution of orders).
- New orders arrive continuously, and old orders are pruned from the system eventually, at the same rate.
- Security prices are updated regularly during a business day.
- The turnover of customers is low, with few insertions and deletions of accounts.
- The number of securities remains fixed, with no deletions or insertions.

Visit http://tpox.sourceforge.net for more information.

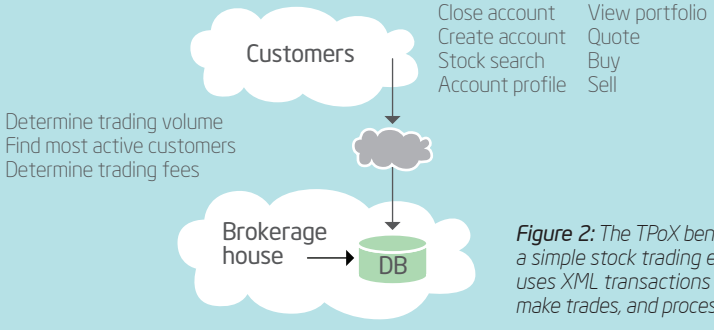


*Figure 2: The TPoX benchmark simulates a simple stock trading example, which uses XML transactions to check prices, make trades, and process orders.*

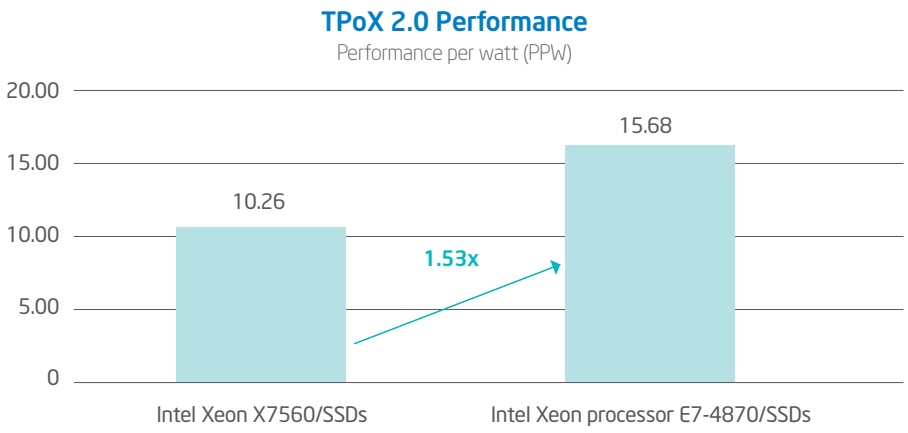


*Figure 3: Performance per watt for the Intel Xeon processor E7-4870 and its predecessor in the Intel Xeon processor family.*

| | Intel Xeon processor X7560/SSDs | Intel Xeon processor E7-4870/SSDs |
|---|---|---|
| TPoX transactions per second (TTPS) | 14,271 | 17,757 |
| Steady state power (watts) | 1,391 | 1,133 |
| Active idle power (watts) | 910 | 716 |
| Power per watt (PPW) (TTPS/steady state power) | 10.26 | 15.68 |
| PPW improvement | N/A | 1.53 |

*Table 2: Power statistics for the Intel Xeon processor family.*

The advantages of the Intel Xeon processor E7 family are not limited to continuing the excellent performance and scalability demonstrated by its predecessors on XML workloads. They also include the power efficiency advantages of the latest Intel platforms.

As Figure 3 shows, on TPoX 2.0, the Intel Xeon processor E7-4870 demonstrates a 53 percent improvement on performance per watt (PPW)—also known as "performance per power"—over the previous generation of Intel Xeon processors. This is because the Intel Xeon processor E7-4870 maintains the same thermal design point (TDP) as the Intel Xeon processor X7560, but also uses low-voltage memory due to improvements in the Intel® 7512 Scalable Memory Buffer, and so puts out less waste heat than previous generations of microprocessors. Other power statistics for the benchmark tests are also shown in Table 2.

## More Power for Industrial-Strength Database Workloads

The TPoX 2.0 benchmark results show that transactional XML workloads can benefit up to 24 percent in performance and up to 53 percent in performance per watt when moved from Intel Xeon processor X7560 to Intel Xeon processor E7-4870 platforms.

Even more important, these results show that IBM DB2 pureXML, running on servers based on the Intel Xeon processor family, readily exhibits the high performance levels needed to handle XML-heavy transaction processing on large databases. The results also confirm what many savvy enterprise IT managers already know from experience: each generation of Intel processors leaps to ever-higher processing speeds to handle the flexible, information-driven applications that run today's businesses, providing the power they have come to expect from IBM and Intel.

## Learn More

More about IBM DB2 pureXML: www.ibm.com/software/data/db2/xml

To learn more about the Intel Xeon processor E7 platform, visit www.intel.com/itcenter/products/xeon

[1] The I/O latency is the average time that a disk I/O request that DB2 makes to the operating system stays in the OS queue, plus the time for servicing by the disk device. It is obtained by using the await statistic from the iostat Linux* tool, measured for all disks in the one-hour steady state interval. What is reported here for each run is the average await statistic across all logical drives in the system.

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.



325444-001US