



Maximize Data Analytics Performance

Rethink the data warehousing equation and empower data-driven business decisions

Skechers USA Inc. is a global leader in lifestyle footwear with more than US\$2 billion in annual revenue. Retail daily sales and inventory data is loaded each night into the company's data warehouse system. About 15 business analysts use this data to understand retail sales and market trends, and to decide on product marketing and promotions.

With more than a billion rows in fact tables, analytical queries could take anywhere from a few minutes to 45 minutes to complete. Time lost waiting for answers is a productivity loss because it limits the number of business analytical queries being executed. This puts more pressure on DBAs to tune each query.

Skechers isn't the only company facing these hurdles. We live in a data-rich world, and the opportunities that come with unlimited data are epic. Retailers have access to fine-grain demographics, buying history, social media identities, online community involvement, information from public records, and more. Manufacturers record and track supplier metrics in exquisite detail, from performance against delivery deadlines to breakage and discount levels. Law enforcement agencies rely on enormous amounts of data and information, locally, nationally, and internationally. Criminal and scientific data that is fast and easy to access allows law enforcement to nail down airtight cases.

When it's used in a timely manner to drive critical business decisions, data analysis equals competitive advantage. The key phrase is "when it's used in a timely manner."

Companies are finding that as data volumes grow, they overwhelm storage systems and the analytics engines that turn data into game-changing insight. In this article, we'll look at how and why the tremendous potential of unlimited data can become a huge challenge, and how Intel and IBM are partnering to help organizations structure data delivery and data management processes that turn unlimited data into unlimited business value opportunity.

Companies Strive for Rapid Analysis

Trying to keep up with unlimited data puts a huge strain on an entire organization, starting with slow analytics response times. As organizations amass more and more data, even basic queries can take a substantial amount of time. More sophisticated queries—particularly those generated by business analytics applications—take even more time. In extreme cases, the query may return results already out of date.

Businesses and public agencies are devoting growing amounts of their IT budgets to handle these vast quantities of data. Data tools and infrastructures have become quite efficient, but even the most advanced technologies cannot keep pace with unlimited data. Creating schemas, indexing, building cubes and summary tables, and creating partitioning strategies all require extensive planning and human effort. Where a single database administrator (DBA) once managed a centralized data center for the entire company, there now may be many DBAs



As seen in IBM Data Management magazine

IBM
data
management
KNOWLEDGE. PERFORMANCE. RESULTS.

managing and tuning data warehouses for just one business unit of a large organization.

And then there is the pure capital cost of the hardware needed to manage growing data. Data warehouses are often large, centralized systems—replacements and upgrades can be expensive, risky, and labor intensive. But with critical decisions dependent on the collation and analysis of diverse data sources, no company can afford to fall behind the warehousing and analytics performance curve.

IBM and Intel In-memory Optimizations

As organizations struggle to maintain query performance in the face of unlimited data, a common bottleneck is the hard-disk I/O throughput. The tipping point has been reached: hard-disk I/O speeds cannot increase quickly enough to keep up with unlimited data, especially as the total amount of storage needed increases. In fact, most strategies intended to accelerate data warehouse queries—including compression, optimization, buffering, and index tuning—are designed to minimize disk I/O.

However, other areas of information technology are advancing rapidly. RAM and solid-state storage become less expensive almost by the day; processor technology continues to make great strides. Processor-to-RAM access speed is increasing at a much higher rate than processor-to-disk access speed. Modern processors are not only faster and less costly than their predecessors, but have been engineered to fully utilize large quantities of RAM and take efficient advantage of larger on-die caches.

That's why Intel and IBM have worked together to develop a cost-effective way for organizations to apply the advanced technologies of the Intel® Xeon® processor E7 family to meet the challenges around data delivery and data management for solution optimization.

Answers at the Pace of Your Digital Investments

IBM® Informix® Warehouse Accelerator helps increase data warehouse performance by exploiting features such as single instruction multiple data (SIMD) and a larger on-chip cache per Intel processor; implementing novel compression and query processing algorithms; and using larger low-cost memory and affordable systems with terabyte memory. The solution is tightly integrated with IBM Informix to accelerate queries transparently to the analytics applications. Data is loaded completely into system memory in a compressed form, using a special columnar scheme developed by IBM. As queries come in, the Informix database passes them to the Informix Warehouse Accelerator, which quickly performs the queries and returns the results.

By compressing and caching the data in-memory, and processing queries on compressed data, Informix Warehouse Accelerator can dramatically increase performance. As processors add more cores and bigger on-chip caches, Informix Warehouse Accelerator is able to capitalize on this trend by running parallel instructions, optimizing algorithms to exploit each on-chip cache. This enables massive scaling and significant performance gains.

With the Informix Warehouse Accelerator query processing engine, you can scan billions of rows of data in seconds. The deep columnar data representation, query processing on compressed data, and innovative algorithms eliminate the need for indexes. Once the basic memory and storage configuration is specified during installation, consistent plans, elimination of disk I/O, fast scans, and joins help eliminate run-time tuning. All data is stored in-memory with just a copy of the in-memory image on a disk. There is no need to plan and create storage spaces for tables and indexes.

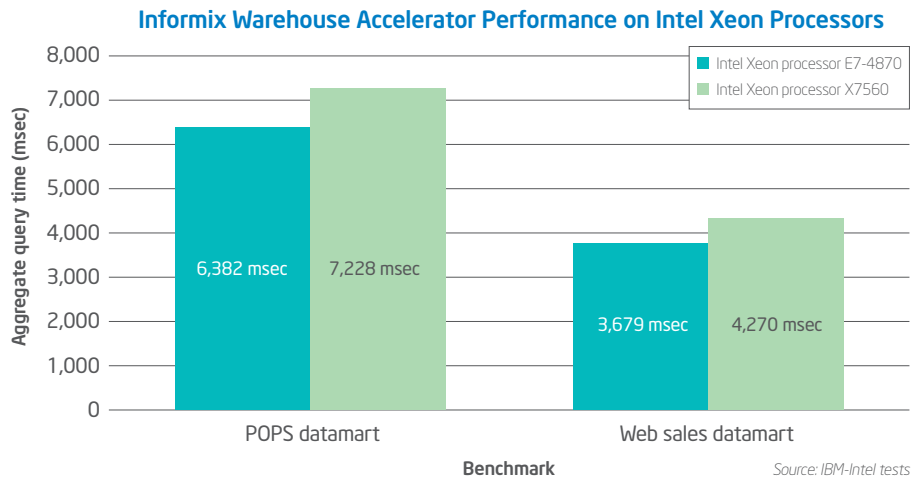
Finally, the accelerator is designed to run on commodity hardware and to easily scale with a growing infrastructure. The accelerator runs on all hardware platforms supported by Informix, and significant data warehouse acceleration can be achieved while minimizing hardware costs.

Informix Warehouse Accelerator is already delivering real-world results. “Complex inventory and sales analysis queries on the enterprise warehouse took anywhere from a few minutes to 45 minutes to run,” says Ashutosh Khunte, manager of data management services for Skechers USA. “When we ran those same queries using Informix Warehouse Accelerator, they finished in 2 to 4 seconds. That means they ran from 60 to 1,400 times as quickly, with an average acceleration factor of more than 450—all without any index or cube building, query tuning, or application changes!”

Informix Warehouse Accelerator Runs Best on the Intel Xeon Processor E7 Family

Transforming traditional disk I/O-based query execution to in-memory execution requires a new level of processor support. The Intel Xeon processor E7 family delivers the high processor performance needed to accelerate even the most demanding warehouse queries, and it supports large quantities of system RAM, enabling organizations to maximize the benefits of the Informix Warehouse Accelerator in-memory technology.

The Intel Xeon processor E7 family is engineered to meet demands for rapidly increasing compute capacity, faster performance, and high availability. It provides up to 10 cores, up to 20 hyper-threads, and 30 MB of last-level cache per processor, all in a four-socket server. The Intel Xeon processor E7 family supports 32 GB dual in-line memory modules (DIMMs), enabling application workloads that require larger memory capacities. Together with Informix Warehouse Accelerator, the Intel Xeon processor E7 family eliminates



the need for disk I/O and dramatically improves the speed of access to complex data sets.

Also, with the Intel Xeon processor E7 family, organizations do not need to introduce additional infrastructure security capabilities. Protection against hardware errors and malicious software attacks is automatically managed by advanced, built-in features. For example, Machine Check Architecture (MCA) enables the CPU to coordinate with the operating system to gracefully recover from memory errors.

Finally, advanced compression capabilities in Informix Warehouse Accelerator enable the Intel Xeon processor E7 family to handle more data in the same memory configurations without sacrificing performance. With Intel® Turbo Boost Technology, processor performance can be enhanced beyond normal operating constraints when thermal, power, and current limits have not been exceeded.

The Intel Xeon processor E7 family—together with in-memory software solutions—demonstrates innovation that can lead to a more efficient performance experience for your data warehousing implementations. Extensive testing by Intel and IBM shows that the Intel Xeon processor E7 family provides a significant performance boost (see Figure 1).

Figure 1: The Intel Xeon processor E7-4870 works with IBM Informix Warehouse Accelerator to deliver increased performance over previous generations of processors for datamarts of widely varying sizes. The proof of performance and scalability (POPS) datamart consists of two fact tables with five dimensions, one with 350 million rows and one with 1 billion rows. The web sales datamart is relatively small, consisting of a single fact table with 7 million rows and 12 dimensions.

Benefit from Better Insight, Exactly When It's Needed

Together, Informix Warehouse Accelerator and the Intel Xeon processor E7 family combine to give organizations the tools they need to analyze unlimited data. Reducing elapsed query times by up to 1,400 times—as Skechers did—means that answers to critical questions are available precisely when they're needed, making knowledge workers more effective and businesses more agile.

For example, marketing teams constantly look for ways to effectively deliver targeted messages to specific audiences. Internet-based marketing has created the expectation of an immediate response to inbound inquiries. As a result, experienced marketing personnel end up devoting excessive time to developing response programs that meet the needs of diverse demographic profiles—yet the resulting programs are often ineffective. System access to large information stores, including huge Internet log files, is crucial for automated, immediate, and personalized marketing communications. Organizations can more fully leverage their investment in analytics applications when the data queries they generate are run and completed at a faster rate.

Making best use of sophisticated analysis and decision-support tools gives organizations a competitive edge. For example, traffic records, criminal records, and other data collected by

federal, state, and local governments can help a law enforcement official identify a wanted criminal during a routine traffic stop. But to do so, the officer must be able to access and query vast stores of complex data in many different repositories, and get accurate, up-to-date results in seconds. Informix Warehouse Accelerator running on Intel Xeon processors can help deliver the appropriate answers in time and in context.

High Performance without the High Price Tag

Despite the need for fast, immediate data access, budget-conscious organizations must balance the benefits of high performance against the resources required to support it. Informix Warehouse Accelerator tips the scale in the right direction. It requires no workload tuning, determines join orders automatically, and requires no application changes—removing rote administration tasks and freeing DBAs to work on higher-value, strategic projects. Also, the in-memory data-handling capabilities of the Intel Xeon processor E7 family help data professionals save time when managing physical schema design.

With Informix Warehouse Accelerator running on the Intel Xeon processor E7 family, organizations can turbocharge their data warehouse performance with scalable, readily available, off-the-shelf hardware. Instead of constantly adding more processing power and more storage,

organizations can acquire commodity components designed to work together to overcome performance bottlenecks and make better use of existing infrastructures.

The Intel Xeon processor E7 family is engineered for reliability and security as well as power efficiency. Informix Warehouse Accelerator delivers significant data warehouse performance improvements by fully exploiting the features of these multicore, large-cache processors to eliminate disk I/O.

Organizations need better options to handle unlimited data. Informix Warehouse Accelerator optimized for the Intel Xeon processor E7 family enables knowledge workers to quickly extract business insight from more data using fewer resources. In a time of unlimited data, IBM and Intel created the innovative Informix Warehouse Accelerator to deliver a rich, competitive advantage in today's data-loaded, computing-intensive environment.

Learn More

More about IBM Informix Warehouse Accelerator: www.ibm.com/software/data/informix/warehouse

More about the Intel Xeon processor E7 platform: www.intel.com/itcenter/products/xeon

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, ibm.com, and Informix are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

* Other product, company, or service names may be trademarks or service marks of others.

Copyright © 2011 Intel Corporation. All rights reserved.

0711/KN/TDA/XX/PDF

325892-001 US