

Informatica® PowerCenter® on Itanium® 2-based Platforms:

Meeting performance and scalability imperatives for next-generation enterprise data integration.

WHO

Informatica Corporation is a leading provider of data integration software used to manage enterprise information assets.

WHAT

Performance measurements demonstrate how effectively Informatica® PowerCenter® takes advantage of the power of Itanium® 2-based servers, providing near-linear scalability with widely varying sizes of data sets.

WHY

Running PowerCenter on an Itanium 2-based platform when large data volumes are present or highly complex transformations are required can result in a substantial performance improvement.

WHY INTEL

Itanium 2-based servers provide essential enterprise computing capabilities, including exceptional performance, reliability, and scalability along with a choice of operating systems, software, and hardware.

MEETING ENTERPRISE DATA CHALLENGES WITH FLEXIBILITY AND POWER

To meet the challenge of managing ever-larger stores of data while satisfying real-time demand for information, today's companies must implement large-scale data warehouse and business analysis implementations. Those implementations must integrate seamlessly with disparate enterprise systems, and they must scale effectively to meet the needs of tomorrow. Choosing both hardware and software that is up to the task of meeting these criteria is vital for success.

The Intel® Itanium® 2 processor, Intel's highest-performing and most reliable server platform, brings 64-bit computing resources to bear on the largest enterprise computing tasks. This architecture enables businesses to support enterprise data-integration products and databases with tremendous volumes of data in highly competitive environments while moving beyond the limitations of high-cost, proprietary RISC platforms. The Itanium 2-based platform provides massive compute resources, mainframe-class reliability, and headroom for growth, while supplying substantially better return on investment and lower cost of ownership than competing proprietary platforms.

Informatica® PowerCenter®, an advanced enterprise data integration platform, takes excellent advantage of the features and capabilities of Itanium 2-based platforms to help organizations access, transform, and integrate data from a large variety of systems and deliver that information to other transactional systems, real-time business processes, and people. Enterprise data can be accessed by systems throughout the enterprise, with an agile, trustworthy, and timely data solution. PowerCenter does not need any intermediate files or staging to process the data thus scaling to true-64 bit performance. Scalability is a key PowerCenter strength, helping businesses to meet the needs of increasing data volumes and user demand and decreasing batch windows.

Taken together, Itanium 2-based systems and Informatica PowerCenter provide the basis for excellent

data-warehousing and enterprise data integration solutions. Enterprise architects who need to deliver very large data sets throughout the enterprise and respond rapidly to an atmosphere of change will benefit from considering this combination. The flexibility, performance, and scalability described in this solution brief make a compelling case for any modern enterprise.

Ongoing Imperatives in Performance and Scalability

Data integration platform performance has become a key technical requirement for success in many organizations. Companies faced with ever-increasing data volumes and transformation complexity must consider solutions that not only meet their present needs, but that also promise to scale and adapt for emerging needs that often cannot be predicted.

IT organizations are accustomed to having the rest of the company look to them to support open-ended requirements. Those requirements can evolve rapidly, due to a host of factors including normal business expansion, mergers and acquisitions, compliance requirements, and strategic initiatives. Too often, the ability to respond effectively to these efforts is hampered by a complex environment of systems that were not designed to share data with each other. In such cases, the ability to make systems work together at all may become the goal, whereas ensuring maximum performance and profitability would be the preferable focus.

A robust enterprise data integration environment provides the means to dynamically support the need for various parts of the business infrastructure to seamlessly share data, anywhere, anytime. Rather than having the topologies of existing systems act as key design considerations, emerging business needs are able to take their rightful place as the driving force behind building out new solutions.

In day-to-day operations, organizations must integrate data and manage their metadata using a robust enterprise data integration platform. Informatica PowerCenter provides the means to provide consumers

“The Itanium processor-based HP server has been a critical part of Informatica’s product development efforts, and that shows in the performance of our systems working together. There are a number of things that are complementary about both of our technologies to support high-volume processing of data.”

*Ivan Chong, Vice President
Product Marketing at
Informatica Corporation*

with accurate, on-time data by improving their ability to dynamically adapt to changing demands. Because of the requirements of very large data volumes and complex processing, one of the primary focuses of PowerCenter is on performance. Because those demands constantly grow, PowerCenter also strives to meet the needs of the enterprise in terms of scalability.

Informatica PowerCenter Helps Businesses Take Charge of Their Information

Informatica PowerCenter is a single, unified enterprise data integration platform that enables companies and government organizations of all sizes to access, discover, and integrate data from virtually any business system, in any format, and deliver that data throughout the enterprise at high speed.

Organizations rely on PowerCenter to serve as the foundation for mission-critical enterprise data integration initiatives including data warehousing, data migration/consolidation, data synchronization, master data management, and cross-enterprise data integration. PowerCenter helps organizations derive business value from all their data so that they can reduce IT costs and complexity, streamline operations, and drive revenue growth.

The metadata-based approach embodied in Informatica PowerCenter helps enterprise architects to unlock the value of information that resides in disparate applications and databases that have grown over time and that do not natively interact with one another. The single data environment enabled by the platform allows for universal transforming, profiling, integrating, cleansing, and reconciling data. Unique metadata management capabilities help boost efficiency to meet changing market demands. PowerCenter also provides robust functionality for real-time, on-demand, or periodic data updates, helping to ensure that the right data reaches the right destination, at the right time.

PowerCenter helps organizations keep costs in check by dramatically simplifying systems design and re-use of data. Developer productivity is greatly aided by the presence of open APIs and SDKs that simplify the addition of new data sources or transformations to the platform. PowerCenter further simplifies design processes by making it easy to sample and profile data, reuse objects across teams and projects, and leverage metadata. It facilitates collaboration across global development teams, sites, and projects by providing granular version control and automated configuration.

PowerCenter supports mission-critical IT resources securely by means of complete user authentication, privacy management, secure data transport, high availability and automated recovery. As discussed in the performance figures later in this document, scalability as processor resources increase achieves near-linear levels, providing excellent return on hardware invest-

ment. Various processing options, such as data-smart parallelism, partitioning, and the ability to scale out to computing grids, deliver flexibility in meeting increased demands.

Itanium 2 Microarchitecture Combines Massive Resources and Innovative Design

The largest mission-critical, data-intensive enterprise environments benefit from the headroom and performance provided by Itanium 2-based systems. In addition to breaking through the 4GB barrier of supported memory imposed by 32-bit architectures, the Itanium 2 microarchitecture has a number of features that support the needs of enterprise-scale systems. A few of the most noteworthy include predication, speculation, and the processor’s explicitly parallel design.

Predication allows the Itanium 2 processor to eliminate the performance deficits associated with branch mis-prediction. In most hardware architectures, processors increase performance of branched sections of code by making informed guesses about which branch will be executed. The processor then executes that branch of code in advance, so that the output of the appropriate operations is available when it is needed. The success of this model demands that the prediction algorithms are accurate enough that the deficits associated with guessing incorrectly are outweighed by the benefits of guessing correctly.

Regardless of how accurate the branch-prediction mechanisms are, it is inevitable that some degree of performance deficit is always associated with branch prediction. The Itanium 2 microarchitecture’s predication capability does away with this deficit by executing both branches in parallel and then discarding the result that is not needed. The Itanium 2 processor does also employ the traditional branch prediction approach. It does both. The compiler determines in each case which approach is most efficient.”, we use the text “The Itanium 2 processor does also employ the traditional branch prediction approach, in addition to predication. The compiler determines in each case which approach is most efficient.

Speculation, another innovative characteristic of the Itanium 2 microarchitecture, further improves computation performance. Because of the fact that processor speed is dramatically faster than memory speed and the speed of the bus that connects the processor to system memory, there is a marked latency associated with the processor having to access information from memory. Thus, it is highly desirable to access that information well before it is needed.

Hardware architectures typically provide mechanisms to ‘pre-fetch’ data from system memory and store it in the processor cache, so that data is available when it is needed to support computations. Most architectures are able to pre-fetch data only after the code branch

that will require it has entered the execution pipeline. In contrast, the Itanium 2 microarchitecture's speculation capability, in addition to extremely large processor caches, allows the processor to load data much further in advance of when it is needed, reducing the incidence of the processor having to wait for data to arrive from memory, and further improving performance.

Explicitly Parallel Instruction Computing (EPIC) architecture allows applications to present instructions to the Itanium 2 processor in bundles of up to three instructions each, and the explicit parallelism of the processor design allows parallel execution of those instructions. That parallelism is being dramatically added to by the emerging generation of multi-core platforms. By providing multiple execution cores in a single processor package, multi-core chips add new levels of parallelism to hardware architectures, which is expected to compound the already dramatic performance and scalability of enterprise solutions such as those based on the Informatica PowerCenter platform.

A Winning Combination of Hardware and Software Architectures

The combination of Informatica PowerCenter and Itanium 2-based servers supports very high throughput and scalability for large data volumes. Unlike many competing solutions, Power Center requires no intermediate staging or hashing files for processing. This characteristic removes a key dependency on disk input/output (I/O), which is typically a gating factor in achieving performance and scalability in enterprise data integration platforms.

PowerCenter makes very good use of the architectural characteristics of the Itanium 2 microarchitecture to provide increased efficiency in memory-intensive transformations such as aggregator, joiner, and lookup.

Particularly valuable to this increased efficiency is the fact that applications running on Itanium 2-based systems are not subject to the 2GB memory limitations imposed on those applications by 32-bit architectures. Since the cache does not need to spill over to disk at the 2GB level, PowerCenter running on Itanium 2-based systems avoids slowdown associated with disk I/O, even with very large lookups.

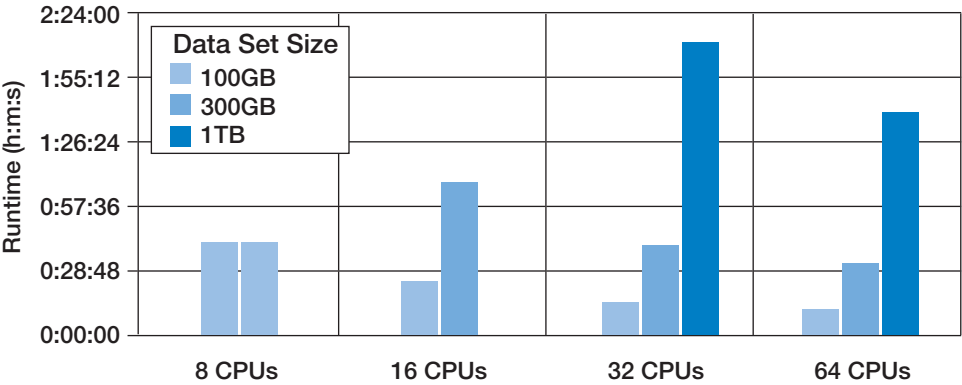
Predication, speculation, and the explicitly parallel design of the Itanium 2 processor architecture further enhance the capability of the system to support the largest data loads with high performance. PowerCenter reader and writer threads also improve disk I/O by using larger data blocks when passing data to and from the database. Equally important to performance itself, the platform scales gracefully under increasingly large data sets and numbers of processors.

PowerCenter's multithreaded partitioning allows the environment to distribute work efficiently across all available processors, increasing CPU utilization and throughput. PowerCenter also provides very robust native support for grid computing environments. The server grid feature allows for multiple heterogeneous servers to be managed through one console via a single metadata repository while providing robust, automated load-balancing. These characteristics enable PowerCenter to deliver high return on hardware investment within the data-centric enterprise.

Performance Tests Show Near-Linear Scalability

In order to demonstrate the scalability of PowerCenter solutions on Itanium 2-based systems, Informatica conducted throughput tests in permutations of increasing numbers of processors and data sets varying in size between 100GB and 1TB. The results of this testing are shown in Figure 1, below.

Figure 1. Scalability of PowerCenter transform operations on workloads 100GB to 1TB in size, on a system with 8 to 64 Itanium 2 processors¹.



Intel Technology

Intel® Centrino™ Mobile Technology

Intel® Mobilized Software Initiative



The data sets for this testing were generated using the Dbgen utility developed by the Transaction Processing Performance Council (TPC). Mappings were built to represent the typical complexity of the average business user, including data type transformations, expressions, lookups, and aggregates. This result demonstrates near-linear performance with all three data-set sizes as the number of processors increases.

This dramatic performance increase with each increase in the number of processors being brought to bear on the problem suggests excellent return on hardware investment. It also exemplifies that headroom that is achievable simply by adding processors to the environment. Thus, this study suggests that customers can readily scale out their environments by adding new hardware, without the need to redesign their infrastructures.

CONCLUSION

By providing a standards-based environment that aggregates enterprise data from disparate sources and enables it to be accessed from in a timely and accurate manner from anywhere in the enterprise, Informatica PowerCenter helps to solve key business challenges. The Itanium 2 microarchitecture is well-suited to hosting this environment, due to its massive compute power and headroom, including large caches and features such as predication, speculation, and an explicitly parallel design.

PowerCenter makes excellent use of the features of the Itanium 2 microarchitecture, and its advanced design removes key bottlenecks that challenge most enterprise data integration platforms. Chief among these are the removal of dependency on disk input/output and using large-scale memory resources efficiently to support large data lookups without cache spillover.

The combination of Itanium 2-based systems and PowerCenter allows excellent scalability, including native support for grid computing, that provides a winning combination for businesses that must provide agile support for data-intensive emerging business needs, meeting open-ended requirements among disparate systems.

Informatica Corporation

100 Cardinal Way

Redwood City, CA 94063

Phone 650.385.5000

www.informatica.com

TAKING ACTION

To find out more about Informatica technology, visit www.informatica.com/products/powercenter.

To learn more about the Intel Itanium 2 processor, visit www.intel.com/products/server/processors/server/itanium2/index.htm.

For more information about Informatica Corporation, visit www.informatica.com.

THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY EXPRESS OR IMPLIED WARRANTY OF ANY KIND INCLUDING WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT OF INTELLECTUAL PROPERTY OR FITNESS FOR ANY PARTICULAR PURPOSE.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, reference www.intel.com/performance/resources/limits.htm or call (U.S.) 1-800-628-8686 or 1-916-356-3104.

1 HP Integrity Superdome* server with 64 Itanium 2 processors at 1.5GHz, 256GB RAM, HP-UX* 11i 11.23 Ver. 2.0, Informatica PowerCenter 7.1, Oracle 9i R2, HP StorageWorks* XP128 Disk Array.

Intel, the Intel logo, and Itanium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Informatica and PowerCenter are registered trademarks of Informatica Corporation.

*Other names and brands may be claimed as the property of others.

Portions Copyright © 2005, Intel Corporation. All rights reserved.

Portions Copyright © 2005, Informatica Corporation. All rights reserved.

Production code: 1005/DPS/ITF/HOP/1K Part No. 310168-001US