

Enhanced Data Warehouse Solutions Offer Greater Performance and Scalability at Lower Costs



TABLE OF CONTENTS

Introduction1

The Need for Better Data Warehouse Solutions2

Perfecting the Appliance Model for Data Warehouses2

Microsoft SQL Server 2008 R2 Parallel Data Warehouse..... 2

The Basis for a Better Hub-and-Spoke Architecture.....4

SQL Server Fast Track Data Warehouse..... 4

Enhanced Performance on Intel® Xeon® Processor 5600 Series5

Power When You Need It, Energy Savings When You Don't..... 5

The Memory and Data Bandwidth to Maximize Performance 5

Security with Less Performance Cost 5

Summary.....6

Introduction

Data warehouses are essential to providing employees fast access to the complete and up-to-date information they need to make better decisions, improve operational efficiency, build profitable customer relationships, develop differentiated product offerings, maximize corporate performance, and increase competitiveness. The more effectively and widely a data warehouse’s business intelligence (BI) solutions can be leveraged, the more benefit these solutions can provide.

Many organizations find the cost and complexity of data warehouses challenging. Consequently, they set up and rely instead on data marts – smaller analytical data stores that focus on specific business functions for a specific part of an organization. This can be an effective solution for a while, but a number of shortcomings crop up over time:

- No definitive source of corporate data – just a group of marts
- Each data mart must be built from scratch when a new need for data appears
- Data marts tend to be brittle and need to be destroyed and rewritten when basic business requirements change
- Once organizations go down the path of multiple data marts, there’s generally no easy or graceful migration path to move to a data-warehouse-centric environment

Microsoft and Intel are now ready to help organizations turn a new page in data warehousing. Using Microsoft® SQL Server® 2008 R2 Enterprise, you can now quickly deploy small data warehouses or data marts with the Microsoft Fast Track Data Warehouse option – a set of new scalable reference architectures for data marts and warehouses that provide an appliance-like experience. Then, as data grows to hundreds of terabytes and the need for a full-scale data warehouse arises, these data marts or warehouses can become part of a scale-out solution using Microsoft SQL Server 2008 R2 Parallel Data Warehouse– a highly scalable appliance offered by leading technology manufacturers for enterprise data warehousing. With these solutions, Microsoft has solved some of the hardest aspects of such a conversion, including the automatic synchronization of data managed centrally in the data warehouse with data in the marts.



Intel plays a key role in these solutions by enabling great performance at a low price per terabyte on platforms based on the Intel® Xeon® processor 5600 series. Working together, Microsoft, Intel, and a variety of technology manufacturer partners have solved many of the data bottlenecks that have prevented organizations from getting the full benefit of the performance potential of their BI solutions.

The Need for Better Data Warehouse Solutions

In the past, organizations have taken either a centralized or a decentralized approach to data warehouse needs. Both these approaches as traditionally practiced have had serious shortcomings. Centralized enterprise data warehouses (EDWs) tend to be extremely expensive and inflexible. As a result, business units become frustrated at the inability of the EDW to meet their needs at a sensible cost and within a reasonable time frame. On the other hand, decentralized data marts, while more responsive to business unit needs, often result in many versions of the same data. This makes it difficult to keep data consistent across the enterprise.

Whether monolithic or chaotically distributed, these traditional BI solutions tend to lock an organization into a major investment in proprietary architecture and result in spiraling infrastructure costs. At some point, a forklift upgrade is often required, triggered by factors such as performance degradation, lack of storage space, changes in business practices, a desire to reduce IT's carbon footprint, increases in users, or any combination of these issues. Such experiences have led many organizations to seek a more scalable, agile EDW environment.

For some, hub-and-spoke architectures, sometimes referred to as "federated" EDWs, have appeared to be the perfect compromise. They match the business structure of most large enterprises by offering a centralized EDW and a set of dependent data marts. The hub allows the enterprise as a whole to set and enforce common standards while enabling analysis and reports that cut across business units. The data mart spokes allow business units to meet their own needs quickly at relatively low cost and also conform to the IT requirements of the overall enterprise. Unfortunately, hub-and-spoke systems have been notoriously difficult to implement, until now.

Perfecting the Appliance Model for Data Warehouses

An easier way to implement hub-and-spoke architecture is the use of data warehouse appliances. A data warehouse appliance is a package that includes all of the necessary hardware and software in a preconfigured solution that dramatically simplifies running and managing a data warehouse. An even better idea is to build this appliance on an industry-standard hardware platform.

Microsoft SQL Server 2008 R2 Parallel Data Warehouse

SQL Server 2008 R2 Parallel Data Warehouse is a state-of-the-art example of an appliance based on an industry-standard hardware platform. This highly scalable appliance for EDWs is the next step in the evolution of the data warehouse appliance originally created by DATAlegro. It uses massively parallel processing (MPP) to deliver high performance and scalability on SQL Server 2008, Windows Server® 2008 R2, and industry-standard hardware. The MPP architecture enables better scalability (scale out rather than scale up), better and more predictable performance, and a lower cost per terabyte than other EDW solutions.

Using SQL Server 2008 R2 Parallel Data Warehouse, you can scale a data warehouse from tens to hundreds of terabytes. What's more, the use of MPP eliminates the physical limits imposed by CPU, memory and storage upon speed and scale that many EDW solutions using symmetric multi-processing (SMP) architecture face. Query processing occurs entirely within one physical instance of a database. There's no competition for resources.

The Microsoft implementation of the appliance model in partnership with various hardware vendors uses Intel Xeon processor-based platforms that have been pre-tuned and optimized for sequential I/O. All components are balanced against each other to reduce any potential performance bottlenecks (from the storage area network, or SAN, to each processor core). Equally important, all server and storage components are mirrored for enterprise-class redundancy. To maximize performance over the MPP architecture, a control node routes queries from applications to all compute nodes, then collects and returns the result.

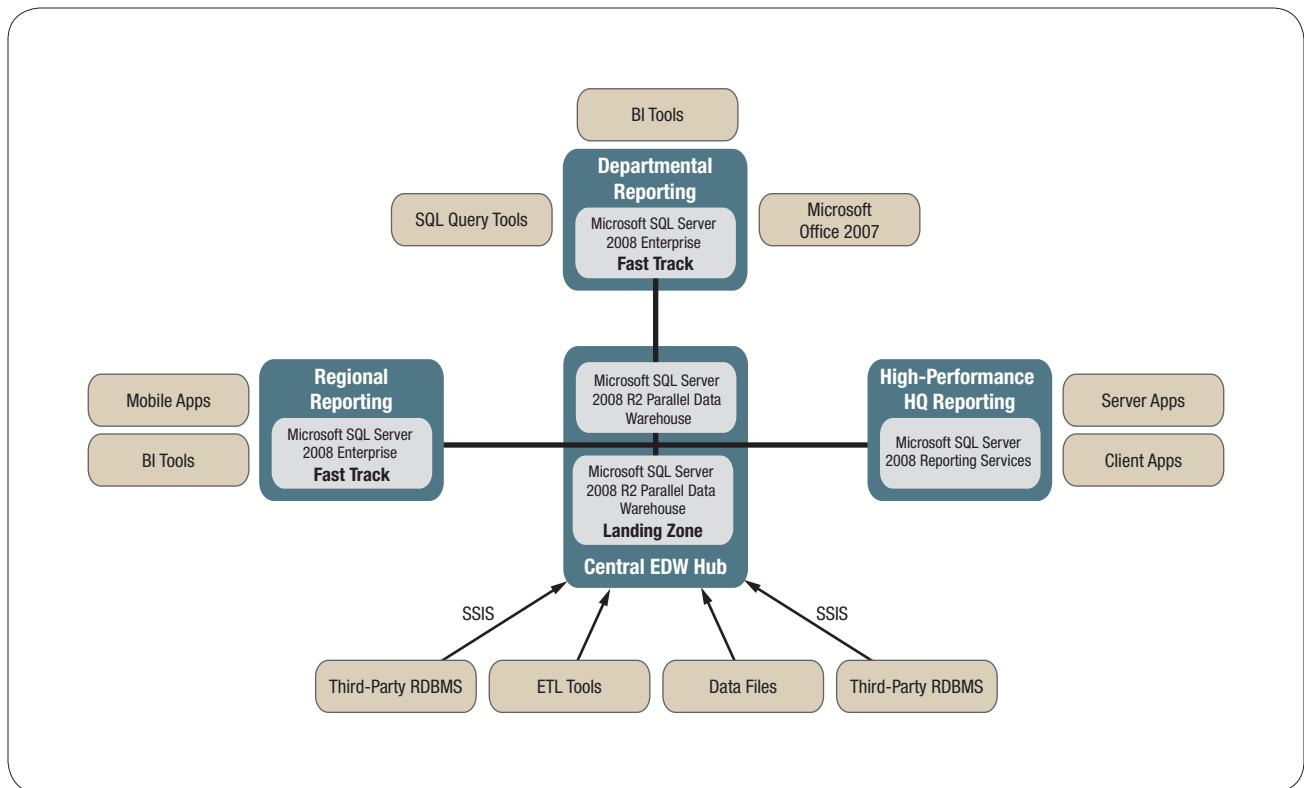


Figure 1. As a massively parallel processing appliance (MPP), Microsoft Parallel Data Warehouse acts as an enterprise hub that publishes data as needed to various business units or departments (spokes). For dedicated high-performance requirements, individual business units can redeploy SQL Server® 2008 data marts as spokes.

A key advantage of MPP architecture is the ability to leverage multiple concurrent resources to resolve SQL Server set operations against distributed data. Each instance works in parallel on its own distribution of a single user query. SQL Server 2008 R2 Parallel Data Warehouse supports up to 10 parallel instances of SQL Server DBMS per data rack. A four-rack system would support up to 40 instances. SQL Server 2008 R2 Parallel Data Warehouse's use of MPP eliminates contention per user query and dedicates a full set of resources for each query in an "ultra shared nothing" design to deliver predictably fast results. Resource and data independence are maintained in each database management system (DBMS) instance. Because data is evenly distributed across multiple nodes and processing occurs in parallel, queries can be handled many times faster than on single SMP database servers. What's more, as data volumes grow, scaling up requires just adding additional capacity to the

appliance. You can scale out continually to add new resources with another box/instance. As you add more nodes, system performance scales linearly.

SQL Server 2008 R2 Parallel Data Warehouse supports third-party BI solutions and provides seamless integration with Microsoft BI products, including:

- Microsoft Office and Microsoft PowerPivot for Excel®
- Microsoft SQL Server 2008 Analysis Services
- Microsoft SQL Server 2008 Reporting Services
- Microsoft SQL Server 2008 Integration Services
- Microsoft® SharePoint® Server 2010

The Basis for a Better Hub-and-Spoke Architecture

SQL Server 2008 R2 Parallel Data Warehouse revolutionizes the hub-and-spoke solution by enabling it to include not only MPP appliances, but SMP instances of SQL Server 2008 and SQL Server Analysis Services. This makes it possible to redeploy existing SMP data marts and plug them into the MPP cluster as spokes, receiving data from the MPP hub. The SMP data marts are viewed as nodes within a grid. The result is a highly flexible, affordable, and scalable platform that, for the first time, makes large-scale hub-and-spoke EDW architectures a practical reality. Organizations get the benefits of central control and governance with the agility of decentralized data marts.

SQL Server Fast Track Data Warehouse

To help organizations accelerate their data warehouse roadmap, Microsoft offers its SQL Server Fast Track Data Warehouse. This set of scalable reference architectures and best practices for data warehousing takes the guesswork out of hardware and architectural decisions for the database layer of SQL Server-based BI solutions and speeds deployment or migration of data warehouses. Organizations can choose the right performance, scalability, and pricing to suit their business needs. Individual marts can scale from four to 48 terabytes using compression

capabilities in SQL Server 2008 R2 Enterprise.

Like SQL Server 2008 R2 Parallel Data Warehouse, SQL Server Fast Track Data Warehouse is designed to provide resource and data independence for each DBMS instance. Because data is evenly distributed across multiple nodes, queries can be many times faster than on single SMP database servers.

SQL Server Fast Track Data Warehouse enables organizations to deploy a high-performance, enterprise-class warehouse without the high cost of traditional systems or all the time-consuming hardware testing and tuning. Pre-configured servers, storage and networking using industry-standard hardware come specifically balanced and optimized for warehousing and better “out of the box” performance. Organizations that build data warehouses using Microsoft Fast Track Data Warehouse reference architectures can easily incorporate them as spokes in a parallel data warehouse hub and spoke system. This offers organizations greater flexibility to implement an architecture that meets their unique data demands. Organizations that are looking for a cost-effective configuration can implement a single hub with Fast Track Data Warehouse spokes powered by the Intel processor 5600 series. For massive data workloads, multiple hub-and-spoke architectures can be implemented and supported with the Intel® Xeon® processor E7-8800/4800/2800 family for greater performance. Either solution enables you to maintain enterprise data in a parallel data warehouse hub and publish the data as

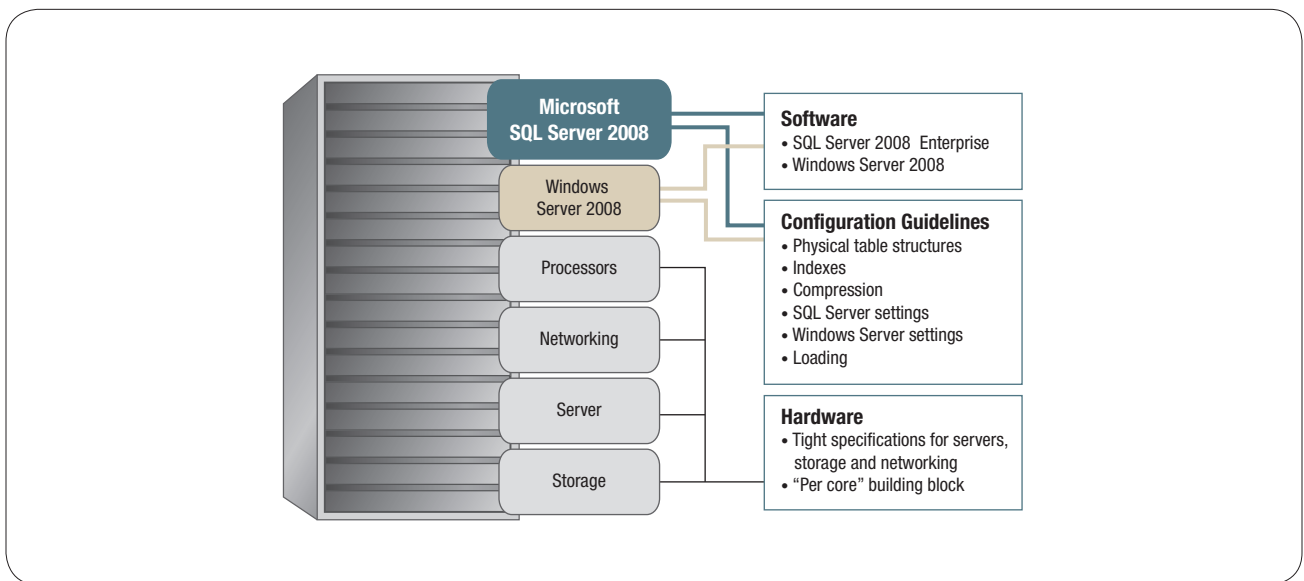


Figure 2. Fast Track Data Warehouse Components

needed to data marts supporting departments and business units.

Enhanced Performance on Intel® Xeon® Processor 5600 Series

For SQL Server 2008 R2 Parallel Data Warehouse appliances and SQL Server Fast Track Data Warehouse reference architecture, Microsoft specifies platforms using the Intel Xeon processor 5600 series or the Intel Xeon processor E7-8800/4800/2800 family for more demanding Fast Track Data Warehouse infrastructures. The Intel Xeon processor 5600 series provides great performance at a low price per terabyte, as well as the advanced system-level features and functionality – such as independent CPU, memory, I/O scaling, and hard partitioning – necessary for balanced platforms designed to process queries within one physical instance of a database and partition large tables across multiple physical nodes.

The Intel Xeon processor 5600 series supports server designs with two, four, or six sockets without the need for customized OEM chipsets and delivers scalable performance across all these configurations. The Intel Xeon processor 5600 series per-socket computing resources and total system scalability make it perfect for SQL Server 2008 R2 Parallel Data Warehouse.

With the Intel Xeon processor 5600 series, organizations can get greater performance and value from SQL Server Fast Track Data Warehouse as well. What distinguishes a Fast Track Data Warehouse from a typical online transaction processing (OLTP) database is that all components – from CPU to disk – are balanced for online analytical processing (OLAP), with a primary emphasis on eliminating potential performance bottlenecks. Optimized for sequential I/O rather than random I/O, each Fast Track Data Warehouse spoke is designed to provide up to 200 MB/s per CPU core. With the right processors, you can make fast work of each query, plus maximize energy cost savings. Data marts built on Fast Track configurations using servers based on the Intel Xeon processor 5600 series can target everything from a smaller departmental footprint of query activity and data volume to departments requiring the handling of higher concurrency and query complexity and optimized at more than 10 TB of compressed storage.

Intel Xeon processors rewrite the book on how processors work

and interact with system memory and other server components. Consequently, these processors are ideal for a parallel, ultra shared nothing architecture where each node has dedicated CPU, memory and storage, and runs its own instance of SQL Server.

Power When You Need It, Energy Savings When You Don't

The Intel Xeon processor 5600 series automatically regulates power consumption and intelligently adjusts server performance according to application demand, maximizing both energy cost savings and performance. The secret to this surprising combination is Intel's 32nm microarchitecture. The Intel Xeon processor 5600 series uses Intel Turbo Boost Technology[®] to intelligently adjust performance to application needs (automatically increasing core frequencies above rated values for heavy workloads when thermals allow) and Intel Intelligent Power Technology[®] to reduce energy costs by automatically shifting the CPU and memory into the lowest available power state. In fact, the Intel Xeon processor 5600 series delivers the same performance as previous generation servers but uses up to 30 percent less power.¹ Available with up to six cores, the Intel Xeon processor 5600 series provides up to 12 processing threads through its Intel Hyper-Threading Technology[†] (Intel[®] HT Technology) to further maximize performance and efficiency.

The Memory and Data Bandwidth to Maximize Performance

To deliver top performance for bandwidth-intensive applications like DBMS, the Intel Xeon processor 5600 series features Intel QuickPath Technology. Each processor has its own dedicated memory that it accesses directly through an integrated memory controller. In the Intel Xeon processor 5600 series, this scalable, shared memory architecture delivers memory bandwidth leadership with up to 4.4x the bandwidth of previous-generation processors,² as well as high-speed (up to 25.6 GB/s) point-to-point connections between processors, and between processors and the I/O hub. This ensures the data bandwidth necessary to keep each core running at capacity.

Security with Less Performance Cost

For faster handling of encrypted data, the Intel Xeon processor

5600 series includes seven instructions (Advanced Encryption Standard-New Instructions or AES-NI) that will be used by the AES algorithm. This set of instructions delivers over 3x the encryption and decryption rate compared to prior AES processes. An instruction called PCLMULQDQ will perform carry-less multiplication for use in cryptography. These instructions will allow the processor to perform hardware-accelerated encryption, not only resulting in faster execution but also protecting against software targeted attacks.

Summary

SQL Server 2008 R2 Enterprise enables organizations to quickly deploy small data warehouses or data marts on industry-standard hardware with the Microsoft Fast Track Data Warehouse option. As an organization's data grows to hundreds of terabytes and the need for a full-scale data warehouse arises, these data marts (and existing SMP data marts) can become part of a scale-out solution using SQL Server 2008 R2 Parallel Data Warehouse powered by the high performing Intel Xeon processor 5600 series. The unique synergy between these processors and Microsoft data warehouse solutions deliver Intel and Microsoft platform leadership, with a higher performance platform that allows IT to support more applications and users per server, so a business can continue to grow without outgrowing its data warehouse capacity. As data volumes grow, scalability requires simply the addition of capacity to the appliance. There is no need for a "forklift upgrade" where the entire appliance must be upgraded. Deploying these platforms, organizations can significantly reduce the effort required to deploy

your data warehouse solution, realize better price/performance, boost business agility, and take advantage of the business insights you can gain from better and faster access to your data.

Parallel Data Warehouse appliances include carefully balanced, pre-assembled hardware from leading hardware vendors and software. This appliance model significantly accelerates your time to value and reduces deployment costs. Through carefully balanced configuration, tight specifications, and testing, and the use of industry-leading Intel Xeon processors, the Parallel Data Warehouse MPP appliance is designed to deliver predictable performance. Microsoft reference configurations suit a number of different usage scenarios, including reporting and ad hoc queries. As data grows and business needs change, Microsoft solutions and Intel-based servers offer excellent options to support more users and evolving business needs.

Learn More

For more information on Microsoft SQL Server 2008 R2 Parallel Data Warehouse, visit www.microsoft.com/sqlserver/2008/en/us/parallel-data-warehouse.aspx

For more information on Microsoft SQL Server 2008 Fast Track Data Warehouse, visit www.microsoft.com/sqlserver/2008/en/us/fasttrack.aspx

For more information on the Intel Xeon processor 5600 series, visit www.intel.com/itcenter/products/xeon/5600

¹ Hyper-Threading Technology requires a computer system with an Intel processor supporting Hyper-Threading Technology and an HT Technology enabled chipset, BIOS and operating system. Performance will vary depending on the specific hardware and software you use. See <http://www.intel.com/info/hyperthreading/> for more information including details on which processors support HT Technology.

² Intel® Turbo Boost Technology requires a platform with a processor with Intel Turbo Boost Technology capability. Intel Turbo Boost Technology performance varies depending on hardware, software and overall system configuration. Check with your platform manufacturer on whether your system delivers Intel Turbo Boost Technology. For more information, see <http://www.intel.com/technology/turboboost>.

³ Intel® Intelligent Power Technology requires a computer system with an enabled Intel® processor, chipset, BIOS and for some features, an operating system enabled for it. Functionality or other benefits may vary depending on hardware implementation and may require a BIOS and/or operating system update. Please check with your system vendor for details.

⁴ Source: Intel internal measurements based on SPECint_rate_base2006 (Linux). Baseline Configuration and Score on Benchmark Supermicro* pre-production system with two Intel® Xeon® processor X5570 (2.93 GHz, 8MB L3, 6.4 GT/s, Quad-core 95W TDP) BIOS rev 01/06/2010, C3 Disabled, C6 Enabled, Turbo Enabled, HT enabled, NUMA Enabled, Automatic Fan, 24 GB (6x4 GB DDR3-1333 DR registered ECC), 1x150 GB 10K RPM SATA HDD, 1x800w PSU, SLES 11 kernel: 2.6.27.19-5-default x86_64. Source: Intel internal testing as of Feb 2010. SPECint_rate_base2006 score: 249. New Configuration and Score on Benchmark Supermicro* pre-production system with two Intel® Xeon® processor L5640 (2.26 GHz, 8MB L3, 5.86 GT/s, Hex-core 60W TDP) BIOS rev 01/06/2010, C3 Disabled, C6 Enabled, Turbo Enabled, HT enabled, NUMA Enabled, Automatic Fan, 24 GB (6x4 GB DDR3-1333 DR registered ECC), 1x150 GB 10K RPM SATA HDD, 1x800w PSU, SLES 11 kernel: 2.6.27.19-5-default x86_64. Source: Intel internal testing as of Feb 2010. SPECint_rate_base2006 score: 257.

⁵ STREAM_MP-Triad benchmark results comparing 2S Intel® Xeon® processor E5472 (12M Cache, 3.00 GHz, 1600 MHz FSB) 16 GB memory (8x2 GB FB DDR2-800) based server platform to 2S Intel® Xeon® processor X5680 (12M Cache, 3.33 GHz, 6.40 GT/s Intel® QPI) based server platform, 24 GB memory (12x4 GB DDR3-1333). Source: Intel internal measurements SSG TR#942 and TR#1071 as of February 12, 2010.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order. Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web Site www.intel.com.

Copyright © 2011 Intel Corporation. All rights reserved. Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries.

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

