CHAPTER 1

Why Itanium Processors? Benefits of the New Processor Family

"The best, simplest way to describe the 'Mission Statement' for the Itanium family of microprocessors is: Creation of an architecture that can address future business needs with the best price-performance and flexibility."

—HP/Intel design team

In This Chapter:

- The Itanium architecture 'Mission Statement'
- The Itanium Processor Family
- Benefits of Itanium processors for different platforms
- Handling Legacy Software, Hardware, and Data
- Reasons for Itanium Processor Performance Value
- Itanium Processor Family in the Enterprise Environment

In this day and age, most of us are familiar with the Silicon-valley buzzword, the 'killer app'—the creation of a software application that can do so much more than its predecessors that it's bound to create tremendous demand. However, it's less often that you hear about a 'killer chip' that will enhance information technology infrastructure to the degree that it is nothing short of a revolution. Yet that is exactly what the Intel® Itanium® processor family[1] architecture is capable of—by building

1. Intel and Itanium are registered trademarks of Intel Corporations or its subsidiaries in the United States and other countries.

on, and then breaking away from established computing principles so that it becomes the industry's new application performance leader.

What you're seeing with Itanium development is the result of heavy investments by both Intel and Hewlett-Packard in a brand new microprocessor family. Yet this is not a matter of breaking with their established system and environments. It's an evolution, not a revolution—one that provides for maximum transition and flexibility—adding value with minimal risk to customers.

Itanium Mission Statement

The Itanium processor family came about for several reasons, but the primary one was that the processor architecture advances of RISC were no longer growing at the rate seen in the 1980's or the 1990's. Yet, customers continued to demand greater application performance, due to the following developments:

- Increased users and demand (internet)
- Higher bandwidth tasks (streaming)
- Requirements for secure processing (SSL)
- Larger hardware requirements (Very Large Data Bases)
- Support for multi-OS environments (virtual data center, computing as a utility)

The Itanium processor family was developed as a response to address the future performance and growth needs of business, technical, and scientific users with greater flexibility, better performance, and a much greater 'bang for the buck' in the price-performance arena.

The Itanium Processor Family

The Itanium architecture was designed to be the new industry standard in high performance processor architecture for the next twenty years. The Itanium processor captures the best-in-class application performance for technical and enterprise computing out of the box today. However, given the continued development of chips and compilers for this unique architecture, its long-term future is also bright.

The Itanium processor starts out with a strong price/performance ratio and is designed to take advantage of scalability gains. This is assisted by broad industry support led by both HP and Intel, who created the initial microprocessor architecture from an alliance of their best R&D resources. This new instruction set architecture, based on HP/Intel co-developer knowledge, has enabled the creation of a strong microprocessor 'roadmap' as indicated in Figure 1-1.

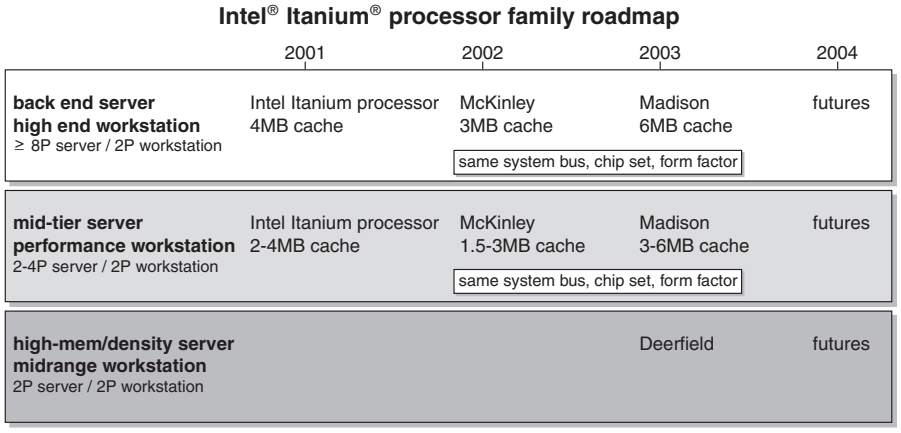

Figure 1-1 The Itanium Processor Family Roadmap

How Will I Benefit if I Run a RISC-Based System?

Enterprises that run RISC-based systems gain an immediate benefit from the switch to an Itanium®-based platform. The lower hardware costs and multi-vendor OS strengths of the new industry-standard architecture, Explicitly Parallel Instruction Computing (EPIC), provide for the following benefits:

- Multi-OS support
- Lower overall Cost of Ownership for enterprise IT
- Assurance of leading performance and scalability over the long-term

How Will I Benefit if I Run an IA-32 Based System?

IA-32 based system users will see immediate performance gains when taking on more complex workloads and processing large amounts of data. Areas that are less sensitive to performance can transition to the new architecture on an as-needed basis. Organizations that move to an Itanium-based platform can address current performance issues points today and gain familiarity with architecture that will be able to keep up with scalable demands in the near and intermediate future. These demands include:

- Greater memory addressability
- More complex applications and computing environments
- Secure web server transactions
- Computer aided design such as Mechanical Analysis
- Very large memory databases
- Industry-standard servers in every tier of the data center

Itanium RAS Features

The information technology industry's term RAS is one that applies directly to Itanium. 'RAS,' or 'Reliability—Availability—Serviceability' provides an excellent example of the benefits that Itanium brings to clients. In each case, Itanium can provide a benefit that is either unique or best in its class compared to less advanced processors.

Reliability

'Reliability' refers to the ability of the hardware to avoid failing. With the Itanium processor family, this ability is built directly into the processor. The prime example of the improved reliability of the processor is the built-in 'error correcting memory.' In a large system with megabytes of cache memory, a little thing like an alpha particle coming through the atmosphere can hit one of the memory circuits and change one of the bits. A change in the bit means that the value in the memory has also been altered.

An alpha particle strike is an example of a purely random error that occurs from environmental causes even though no "hard" failure has happened. It's not the result of a bad design or a component failure. However, the error correcting memory will fix the problem on the fly guarding against this effect.

This is because error correction introduces of 'parity check' that can tell whether a given bit should be 'on' or 'off', and even fix the piece of data.

Availability

A system with 100% 'Availability' is a system that is always up and never down. At present, the 100% threshold is only theoretically attainable, in part due to the unpredictability of random errors, natural disasters, required maintenance, upgrades and of course 'hard' and 'soft' errors. Typically, the reason systems tend to crash and go down is due to these last two kinds of errors.

Hard Errors and Soft Errors

A 'hard' error, as the name implies, is where something is physically changed on the system, or a piece of hardware crashes, freezes, or burns up. This can happen to a power supply if it shorts out, a coolant fan that seizes up, or a component that short-circuits. Hard errors also include outside interference that results in downtime—such as a lightning strike that causes a power surge, resulting in knocked out equipment.

A 'soft error' is of the type described in the previous section on reliability. Here a cosmic ray or electrical noise on a bus will unintentionally and randomly reset a data bit. But the parity and error correcting code circuits (ECC) of the Ita-

nium processor will actually fix these errors as they are detected and keep the user's computing environment safe in spite of these unexpected events. The on-chip parity checks and ecc will detect and correct both hard and soft errors.

Software Errors

Software Errors, as the name implies, are errors that occur in the software. They typically occur when software has encountered something in the programming or data that confuses the program or is unanticipated by the programmer, so the program doesn't know where it is going, or what it should do next.

Sometimes these errors could be initially caused by faulty hardware (a hard error leading to a software error) or by programming mistakes. For example, this could be caused by a situation where a program is trying to branch to an address in memory and that address happens to be in the middle of a block of data or in open memory.

Ideally, the system should be able to catch these software errors without having to stop and wait for some kind of operator or other intervention. The best result is that the system has a place to go and 'recover' from the error and not crash. An example would be allowing the program to jump to a new address. This allows it to back up and take a look at what happened to keep executing at a slower speed, or even go into the execution process to generate a recovery in a graceful way. It may even go into another routine that 'branches' around the area with the problem.

Serviceability

Let's say that as a network administrator you find out that nothing is being received over a specific network node. Investigating, you find an I/O card has gone out on the system. On a typical Itanium-based system, you can pull out the card while the system is still running and plug another one in—it's 'instantly serviceable.' You don't have to shut the machine off.

Hewlett-Packard computers typically have system component redundancies for subsystems like fans, and power supplies. We tend to build the redundancy into the areas that statistically tend to have the most failures. Thus if the fan fails, the entire unit doesn't burn up. Similarly, built into the architecture is the enabling capability to allow for OLR (Online Replacement). This is another way of saying that you can pull out a component (e.g., I/O card) while a system is running, and keep on going. Instead of engaging in costly downtime, replacement can be done on the fly.

This is a capability that has been built into the architecture. In a multi-processor situation, you can turn off one processor and substitute in a new one. The ability to service the machine while it is still running reduces your down time even further. It's rather like being able to replace a flat on your car while in the middle of your freeway commute.

Reasons for Itanium®-based Platform Value

The Itanium architecture achieves a more difficult goal than a processor that could have been designed with 'price as no object'. Rather, it delivers near-peerless speed at a price that is sustainable by the mainstream corporate market. Some of the features that this processor brings to the table follow below. Other reasons, such as the parallel architecture, are important enough to discuss in their own separate sections in this chapter, which follow.

- Floating-point performance for compute intensive applications
- EPIC technology for maximum parallelism & HW/SW synergy
- Scalability from 1-way to 128-way+
- 64-bit addressing and high bandwidth

Highly Parallel Architecture

The highly parallel architecture of the Itanium architecture allows the processor to get more things done during each 'clock cycle'. There are typically two ways to speed up the performance of a processor. The one that the general public understands and follows is to increase the 'clock speed' of a processor. For years, this has been the initial benchmark as to how to measure the speed of a chip—a 500Mhz Pentium versus a 200Mhz Cyrix chip, for example.

Members of the Itanium processor family will have a clock speed of one gigahertz and greater. Of course, we're all excited about being able to hasten the clock speed to such a degree. Increasing the clock rate will improve performance, but the processor is still bound by the architecture that defines its operation. For example, the processor may be an early implementation of RISC, which will attempt to get only a single instruction completed per cycle.

Imagine the performance gain from completing two instructions per clock cycle instead of one. By doing two things in parallel, it's the equivalent of doubling the clock speed on the microprocessor. Consider that if you had a 500Mhz chip that could process two instructions per clock cycle, you would end up with a one-gigahertz microprocessor without the costs associated with the higher clock rate.

This ability to make these leaps of performance was the reason we made the step in RISC to move to what's called 'superscalar out-of-order execution RISC'. The end goal was specifically to be able to do more things in parallel, and create a speed gain where previously, the main method was to keep increasing clock speed and changing the manufacturing process to "shrink" the chip

HP and Intel created EPIC—an explicitly parallel instruction computing architecture, which is the heart of why the Itanium processor family can create its unique price-performance signature. This allows us to go from a slightly more than one instruction per cycle to up to 6 instructions per cycle with the current generation

of processors. This level of parallelism is probably not likely over an entire application, but by getting a boost in even a couple of critical loops, the overall effect on the speed of processing can be significant.

In performance tests in the lab, we've seen 1 gigahertz Itanium processors running alongside 2.6-gigahertz chips, and the Itanium processors are getting more usable application work done—because of the unique architecture.

The magic of EPIC is in its parallelism. Of course, it does take some work in the compiler, and it does take some sorting out to queue up the information to the microprocessor; you have to know that you can do certain things in parallel. And the future of this technology is just beginning. As compilers get better, and we continue our research into the methods of properly utilizing parallel architecture, we'll be able to get even more speed and performance gains out of this processor family.

Investment Protection

The triple protection that Itanium architecture offers is for legacy software, legacy hardware, and legacy data. This was the second major goal we had when we began the project. We wanted to protect the customer's investment in IT infrastructure as much as possible.

Legacy Software

It turns out that when you analyze an organization's IT investment, the bulk of it is not in the hardware. Most of the investment beyond the people is in the software and the legacy data that is stored. So one of the goals when developing Itanium was being able to run—even though it would be at a lower speed than a native binary—the existing applications coming from both the IA-32 world as well as the PA-RISC world.

HP built this investment protection for PA-RISC into HP-UX, our version of the UNIX®[2] operating system. For the IA-32 world, the backward compatibility is also implemented, allowing even more flexibility. This allows Itanium-based systems to run Linux applications and Microsoft Windows applications without change.

The idea is that the user won't have to immediately go out and update new software applications. Most customers will want to update some applications in order to take full advantage of the phenomenal speed gains the Itanium processor family makes available, but in some cases, it's just not a critical issue. For example, the application in question could be a utility routine that runs in the background, say to monitor events and generate reports. An example would be HP's Openview suite of system management tools. In other words, it would be a program that would stand

2. UNIX is a registered trademark of The Open Group.

to benefit extremely little from an performance upgrade, and yet would be annoying or laborious to re-compile or re-write.

A second problem that IT organizations often run into is in their own proprietary software. This is where the source code has been lost and the person who wrote it is no longer available to consult. The cause of this loss is irrelevant. What matters is that there is no one around to redesign the program on short notice. With the investment protection, built into Itanium-based systems so that a redesign isn't necessary, this issue is neatly avoided.

Legacy Hardware

Hewlett-Packard's long-standing policy on legacy hardware was extended to the Itanium architecture before the project even began in earnest. The goal here was to build systems that we could upgrade from PA-RISC microprocessors by simply replacing the systems board with future Itanium processors installed. So most of the Hewlett Packard PA-RISC systems that we've built and launched in the past year have the ability to be upgraded to versions of the Itanium processor family.

Legacy Data

Enterprises have huge databases built over time on their IT systems. This data, if built under a PA-RISC and HP-UX environment, is stored in a "big endian" format. If you move to a system, which is "little endian", the user can't simply plug the older disk with its data into the new system. (See Appendix A for a definition of "endian.") Instead, the data will need to be converted before being used. In order to protect the data of our enterprise customers on HP-UX systems, as well as data that is stored in "little endian" on Windows or Linux based systems, we made sure that Itanium was endian neutral and could handle data in either format.

Choice and Breadth of Operating Systems and Applications

A real frustration for customers is when the application needed is not available under the operating system that is currently installed and running on a system. For example, you may be running Microsoft Windows, but the application you want runs only under Unix. Often, you're left with only one option—going back to the application vendor and asking them if they can develop a version of the application that will work under your installed operating system.

Many IT organizations today are concerned with the mix of computers using different operating systems on a single network. This has been caused in part by the gradual dominance of the front-end applications by Windows, while the back end continues to use UNIX or other non-Microsoft systems. There's a legitimate concern that a major investment in one type of operating system will result in a severe disadvantage should the investment prove to be in the 'wrong' one.

Under the Itanium processor family, a maximum amount of choice is preserved. You can continue running UNIX or OpenVMS on the back end, for instance, but if the world switched over to Windows or Linux the following day, the impact is minimized. Of course, you'd have to get new applications, but instead of having to replace your hardware you could phase in a transition without the major capital expense and hassle involved in complete replacement.

Enterprise Technology

The basic definition of 'enterprise computing' is the type of computing that is done in large companies on high-bandwidth networks, as opposed to 'personal' computing, which tends to take place on stand-alone machines. Most enterprise computing is done in large corporations, so the emphasis is on which systems work best with business-to-business (B2B) applications. These B2B applications include programs written for supply chain management, customer relationship management, communications/sales tracking, and project/milestone management.

This form of computing can also be determined by the following characteristics:

- Large databases with the need to update, analyze, and mine the data constantly.
- The need for reliability/available/scalability 24 by 7.
- Large applications.
- Hundreds or thousands of users.

These requirements are also expressed graphically, as in Figure 1-2.

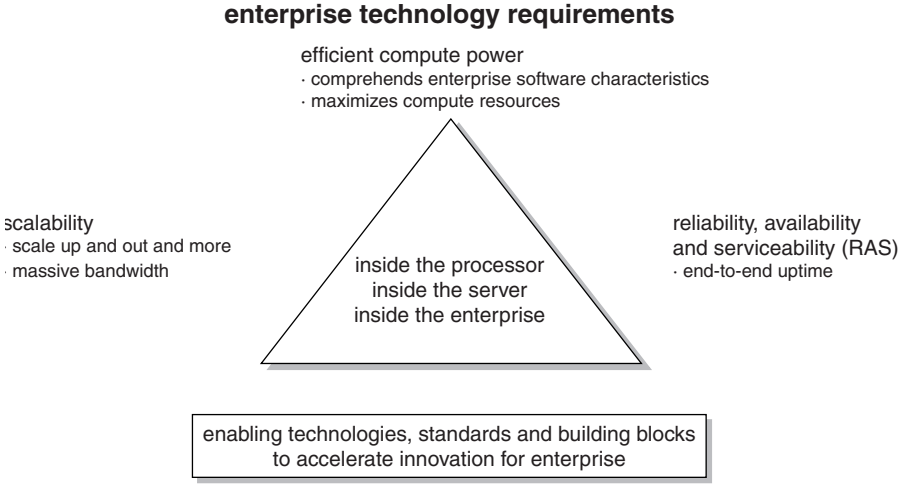

Figure 1-2 Diagram of Enterprise Technology Requirements

How Itanium Architecture Affects Enterprise Computing

The Intel Itanium architecture was intended to extend further into the enterprise by improving upon capabilities of today's architectures. Enterprise IT environments are extremely heterogeneous, with multiple system, applications, and operating systems. An example of this can be found at Hewlett-Packard itself. It's not uncommon to find UNIX servers, which need to talk to a Windows or Linux servers, which in turn service PCs that are running the latest version of Microsoft Windows.

A multiple, heterogeneous environment is much more the norm today than one which is dedicated exclusively to a single operating system or set of enterprise applications. Itanium's ability to handle just about any operating system that is run on it makes it a natural fit for today's mixed network environments

In Summary

- The Itanium architecture was developed to meet and exceed future performance and growth needs of enterprise business, technical, and scientific users with greater flexibility and future headroom. The idea was to create an architecture capable of greater speed and also unbeatable price-performance.
- The Intel Itanium processor is faster than most of the present day competition. And, with future compiler improvements and newer chip designs, it will outpace all other existing architectures.
- No matter if you run RISC or IA-32 based systems, you'll still be able to reap performance gains from Itanium-based systems.
- An Itanium-based platform also provides superior performance in all three areas of 'RAS': 'Reliability—Availability—Serviceability'.
- The Itanium processor uses EPIC, or explicitly parallels instruction computing architecture. This allows a system to go from one instruction per cycle to 6 or even more instructions per cycle.
- Investment protection was another goal of Itanium-based solutions. The triple protection that Itanium-based systems offer is for legacy software, legacy hardware, and legacy data. This is due in part to Itanium's ability to work with almost any operating system and application.
- Enterprise computing is an ideal place for Itanium power. By improving upon capabilities of today's architectures, Itanium can ensure flexibility and choice in the enterprise environment.