

Case Study

iArchives

Intel® IOP333 I/O
Processor
Intel® Core™2 Duo
Processors

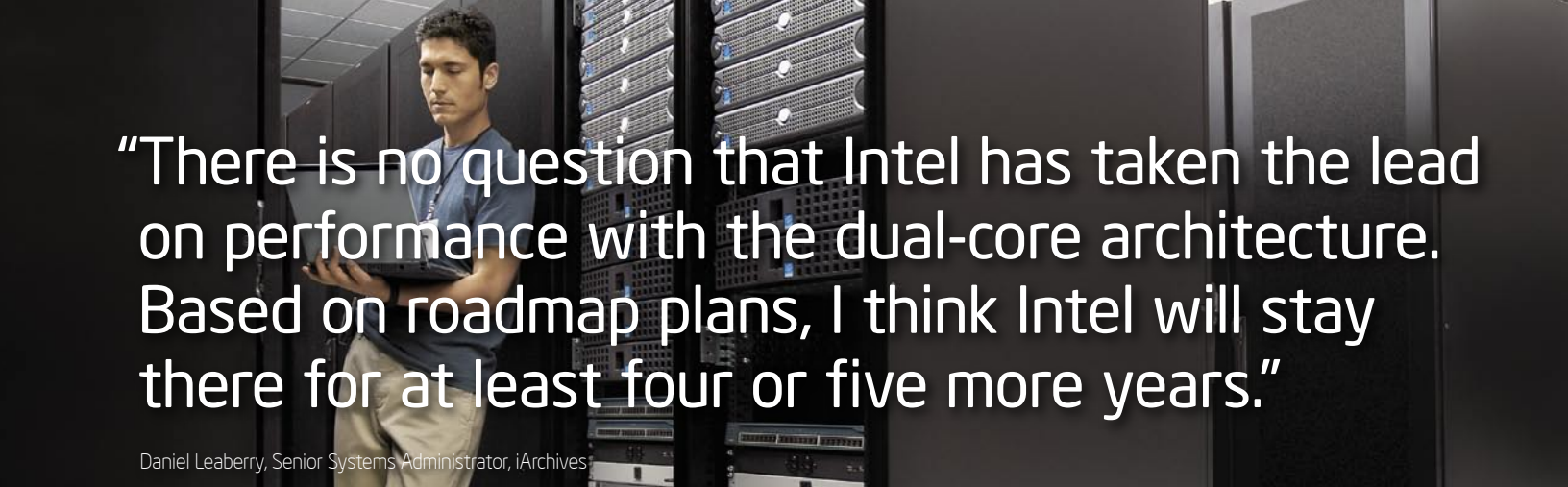
I/O and CPU Systems Get a Workout Converting Paper into Searchable Digital Archives

RAID 6 plays a critical role in data protection

Challenge ▪ Scanning and processing over 3 million 90 MB images each month, iArchives must store and protect as much as 110 TB of data at any given time. With a system of this size, sustaining throughput, disbursing thousands of jobs at a time, and minimizing power consumption and thermal output can be challenging. After suffering a drive loss during a RAID 5 rebuild, iArchives wanted a more reliable data protection scheme.

Solution ▪ iArchives uses 250 CPUs to process and store customer images. Storage is disbursed into 250 GB SATA drives that are configured in arrays no larger than 5 TB each. iArchives recently began using hardware-accelerated RAID 6 to protect data even during drive rebuilds. They also began using dual-core processors to maximize performance while lowering power consumption, running 80 processor cores on a single 30 amp circuit.

Products ▪ Each Areca ARC-1260 RAID 6 host adapter with the Intel® IOP333 I/O processor provides maximum sustained I/O throughput and supports up to 65 CPUs. iArchives also recently started using Dell PowerEdge*1955 server blades, each outfitted with two Intel® Core™2 Duo processors to minimize power consumption and thermal output.



“There is no question that Intel has taken the lead on performance with the dual-core architecture. Based on roadmap plans, I think Intel will stay there for at least four or five more years.”

Daniel Leaberry, Senior Systems Administrator, iArchives

“20 of those blades, with 80 processor cores, can be on one 30 amp circuit. That’s four times more than could be run on the old servers with single-core chips on a 30 amp circuit.”

Daniel Leaberry
Senior Systems Administrator
iArchives

Background: Turning Paper Assets into Digital Databases

Now that digital publishing has become commonplace, it can be difficult to appreciate the fact that many valuable records still exist only in paper form. Most legal documents are still recorded on paper. Newspapers, magazines, and other mass publications are still printed. And everything that was produced prior to the early 1980s exists on paper because there was no means for digital publishing.

But compared to digital databases, paper archives can be rather unwieldy. Beyond a large physical footprint, these “libraries” can be cumbersome to search, are subject to deterioration over time, and are difficult to share over a wide geographic area. That is why many organizations are looking for ways to convert their most important paper assets into searchable digital archives.

Using state-of-the-art scanning equipment, iArchives creates a high-resolution image of a paper document and then saves it as a TIFF, PDF, or other customer-specified file type. The patented iArchives OCR software and feature-rich solution enriches the image through enhancement, text indexing, and metadata file tags that make digital documents highly searchable.

Because iArchives displays the digital version of a document in its original format (such as a newspaper page), users can view and gather information in the original context. Users can also access surrounding information like photos and incidental data, plus see a true representation of signatures, typography, and other unique characteristics. Unlike microfilm, digital documents can be stored in almost any database selected by the customer and can be searched using almost any preferred search engine.

iArchives performs the scanning, indexing, tagging, and sorting service for its customers, storing the data only until the project is complete. After that, the completed digital archive is turned over to the customer for management and maintenance.

System Requirements for the iArchives Solution

Let’s consider for a moment the scope of the iArchives solution and the networking and storage system the company developed to support its business. iArchives currently processes over three million 90 MB images each month. At any given time, iArchives must store—and protect—nearly 110 TB of newly minted digital data.

“Scanning such large images takes a lot of straight sequential I/O,” explains Daniel Leaberry, senior systems administrator at iArchives. “We also need a lot of processing power to handle the data—both in I/O and the CPU. We’re currently running 250 CPUs in our network.”

With so many processors and so much demand for I/O, iArchives has to worry about power consumption and thermal output. That means the company needs high-performance servers that optimize the power ratio, making sure that sustained throughput can be maximized with a minimum number of controllers.

Besides performance, Leaberry also worries about protecting data. “With so many images, so much data, and the fact that our business involves processing massive amounts of unique data, iArchives has to do everything necessary to protect against data loss in the event of a drive failure,” Leaberry says.

"iArchives used to rely on RAID 5 with hot-swap and drive-rebuild capabilities," Leaberry recalls. "But we lost a drive during a RAID 5 rebuild, which is the one time when a system is vulnerable. iArchives was lucky in that most of the data was already scanned and stored elsewhere, so not much was lost. RAID 5 didn't offer enough protection for the iArchives solution, so we switched to the dual-parity drive scheme offered with RAID 6. That way customer data is protected even if we lose a drive during a rebuild."

Elements of the iArchives Solution

Switching to RAID 6 was one of the more recent changes made by Leaberry to improve the iArchives storage and data protection solution. In making the move, he evaluated several RAID options and eventually chose the Areca ARC-1260 RAID 6 host adapter powered by the Intel® IOP333 I/O processor. "The ARC-1260 RAID 6 host adapter with the Intel processor—running hardware-accelerated RAID 6—achieved 20 percent greater sustained throughput than our existing RAID 5 controller," says Leaberry. "Our infrastructure can now support up to 65 simultaneous processor threads with the Areca controller running the dual-parity scheme in RAID 6. We're very happy with that."

The Areca RAID host adapters with Intel® I/O processors and integrated RAID 6 have delivered everything Leaberry expected, and then some. "We can saturate a 3 GB link with this system, which means a tremendous amount of data can be put through at one time. There's even headroom for more."

"Areca's driver support is really good, especially for open source drivers, which really helps because iArchives relies heavily on open source products," adds Leaberry. In fact, iArchives is planning to cluster its servers using open source Luster clustering software, which will allow for better resource allocation among the system processors.

Another core element of the iArchives solution are the server blades that support the image scans. iArchives recently upgraded about half of its network to new Dell server blades, each running two Intel® Core™2 Duo

About the Areca ARC-1260 SATA II RAID 6 Host Adapter

The ARC-1260 SATA II RAID 6 host adapter is a high-performance, 16-port PCI-Express bus to SATA II disk array host adapter. When properly configured, the SATA host adapter can provide non-stop service with a high degree of fault tolerance through the use of RAID technology and advanced array management features.

The onboard Intel® IOP333 I/O processor integrates the RAID 6 engine, offering hardware acceleration of dual-parity RAID 6 data protection for better performance. The integrated RAID 6 feature cuts Areca's engineering and development effort and saves on component costs and internal PCI bandwidth. Areca and its customers also benefit from an emphasis on architectural commonality in Intel I/O processors, enabling Areca developers to capitalize on existing code base when developing a new adapter.

processors. "20 of those blades, with 80 processor cores, can be run on one 30 amp circuit," Leaberry notes. "That's four times more than could be run on the old servers with single-core chips on a 30 amp circuit." Because iArchives runs so much data through its system, increased power savings is a huge benefit that makes the company's solution much more robust.

Leaberry also likes the multiple cores in the Intel processors because the iArchives application is engineered to take advantage of them. "We're a multi-threaded business, and the multi-core architecture of the new Intel Core 2 Duo processors means we can hand out more jobs without straining the network. And we're very excited about Intel's forthcoming quad-core processors."

"We have experience with Intel® CPUs, and we know Intel builds a good processor."

Daniel Leaberry
Senior Systems Administrator
iArchives

“Intel Has Taken the Lead on Performance”

Leaberry says that the presence of the Intel I/O processor on the Areca card was one of the reasons that iArchives chose the ARC-1 260 when moving to RAID 6. “We have experience with Intel CPUs, and we know Intel builds a good processor. We figured the hardware-based RAID would be good, too. Plus, the fact that Areca is using Intel I/O processors assures us that the company isn’t spending all of its time designing a custom ASIC. That means Areca has the resources to dedicate to making a better card, and that’s a win for all of us.”

Although Leaberry says he had a preference for Intel I/O processors, it was the performance test between the RAID 5 card and the Areca ARC-1 260 with hardware-accelerated RAID 6 that made the decision. “20 percent better performance is amazing, especially since the dual-parity drive scheme in RAID 6 usually slows things down a bit.”

Apparently the recent iArchives decision to buy Dell PowerEdge* 1955 blade servers was also driven by the performance of the Intel Core 2 Duo processors. “There is no question that Intel has taken the lead on performance with the dual-core architecture,” observes Leaberry. “Based on roadmap plans, I think Intel will stay there for at least four or five more years.”

He notes that the Intel processor-based Dell PowerEdge blade servers are the only kind that iArchives is buying now. “We are going to stick with PowerEdge for a few years—we are pretty sure that Intel has a good lead on performance for the next five years or so.”

Leaberry adds that iArchives makes decisions based on the best product available, not necessarily out of brand loyalty. “We did not blindly choose Intel. We just want the best-performing product, and Intel has it right now—and will for awhile.”

Growing the Network

When talking about the growth of iArchives and the processing and storage network that serves the company’s expanding customer base, Leaberry notes that more Dell servers and more Areca controllers will be added in the near future. About half of the current network is supported by the Areca ARC-1 260 controllers, and eventually the whole storage system will be moved over to the RAID 6 solution. More multi-core Intel processor-based blade servers are also planned.

“We are really excited about Intel’s forthcoming quad-core processors,” says Leaberry, hinting perhaps that servers based on them are also in their future plans.

For more information, visit: intel.com/go/storage, iarchives.com, and areca.com.tw.

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 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. Intel products are not intended for use in medical, life saving, or life-sustaining applications. Intel may make changes to specifications and product descriptions at any time, without notice.

Information regarding third party products is provided solely for educational purposes. Intel is not responsible for the performance or support of third party products and does not make any representations or warranties whatsoever regarding quality, reliability, functionality, or compatibility of these devices or products.

Copyright © 2007 Intel Corporation. All rights reserved.

Intel, Intel. Leap Ahead., the Intel. Leap Ahead logo, and Intel Core Duo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

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

