EXCLUSIVE INTERVIEW BARBARA ADEY, VP CONVERGED INFRASTRUCTURE, HP

HPC REVIEW

The reference media in high-performance IT solutions

www.hpcreview.com

HPC | BIG DATA | CLOUD | STORAGE | VISUALIZATION | VIRTUALIZATION | NETWORKS | WORKSTATIONS

#5 GLOBAL EDITION

QUANTUM COMPUTING

READY FOR PRIME-TIME?

LAB REVIEW

FUJITSU ETERNUS CD 10 000

HP 3PAR STORESERV 20800

STORAGECRAFT SHADOW PROTECT 5

VIEWPOINT

BRINGING LUSTRE RELEVANCE TO THE ENTERPRISE

HOW TO

OpenMP Device Constructs

TECH ZONE

BEYOND DEEP LEARNING AND NEURAL NETWORKS

4 Blades

4 GPGPU

HighServer XLR4 Blade

Virtualization Server with GPUs / Co-Processors

- High density
- Low power consumption
- Easy management

Supports the

Intel® Xeon® Processor E3-1200 V4

Intel Inside®. Powerful Solutions Outside.

GIGABYTE™

Available at:

CARRI Systems

intel inside™ XEON®

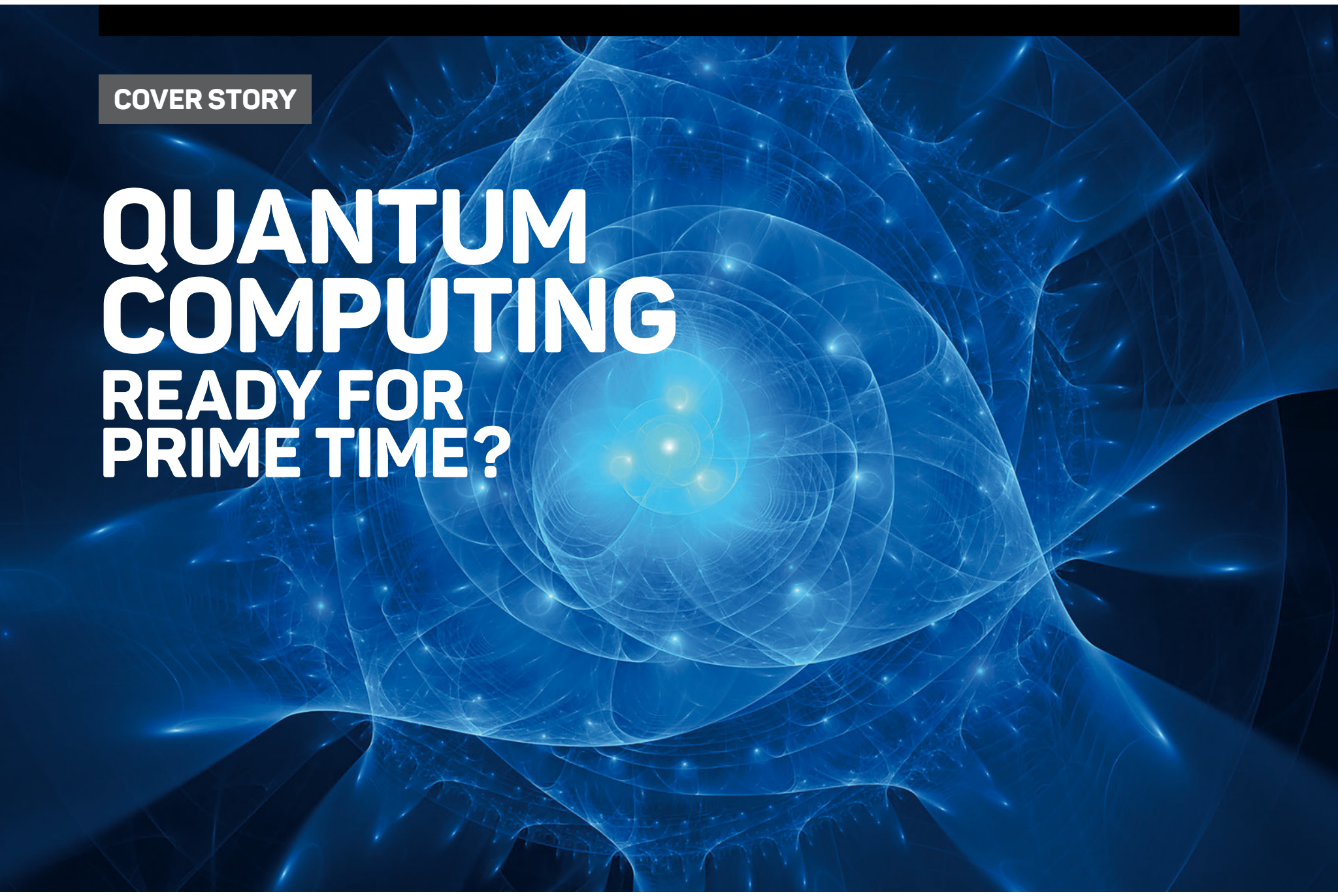
Citius, Altius, Fortius !

Faster, higher, stronger.... that could be the motto of the IT industry as a whole, which has since its inception in the last century, made human activities progress dramatically. This issue's cover story is a perfect example of how the research in the academic and industrial world unite to create a new architecture based on quantum principles. Given the progress made in recent months, a tangible success seems at hand. Because the transition from a binary computer towards a quantum computer is a race against time!

The end of Moore's Law ? According to numerous sources, Moore's law could cease to work by 2020, the size of transistors approaching that of the atoms that compose them, rendering current photolithography techniques impossible to use onwards. Not only that, even if current technologies would allow venturing into the infinitely small, at such diminutive scale quantum effects disrupt traditional transistors that make up the current processors. Therefore, the race to master this new state of matter and lead to the next generation of computers runs full swing.

This is the story of a journey full of pitfalls, but also recent achievements that sheds a light of hope on the dawn of a new era in computing.

Happy reading!



COVER STORY

QUANTUM COMPUTING
READY FOR PRIME TIME?

ASUS®
IN SEARCH OF INCREDIBLE

intel

ESC4000 G3

Green500 Champion Successor

http://www.asus.com/Server-Workstation/

Thoughts on the Exascale Race: HPC has become a mature market

As the HPC community hurtles toward the exascale era, it's good to pause and reflect. Here are a few thoughts...

The DOE CORAL procurement signaled that extreme-performance supercomputers from the U.S., Japan, China and Europe should reach the 100-300PF range in 2017-2018. That's well short of DOE's erstwhile stretch goal of deploying a trim, energy-efficient peak exaflop system in 2018 or so, but still impressive. It would appear to leave room for one more pre-exascale generation before full-exascale machines begin dotting the global landscape in the 2020-2024 era.

An exaflop is an arbitrary milestone, a nice round figure with the kind of symbolic lure the four-minute mile once held. And as NERSC Director Horst Simon pointed out many moons ago, there are three temporal stages to these computing milestones that have occurred about once a decade. First will come peak exaflop performance, then a Linpack/TOP500 exaflop, and finally the one that counts most but will likely be celebrated least: sustained exaflop performance on a full, challenging 64-bit user application.

STEVE CONWAY
is IDC Research Vice President, HPC

A peak exascale system is merely an "exa-size" computer, to cite the term Chinese experts used in an SC13 conference talk. It's a show dog without a repertoire of tricks. A system that completes a Linpack run at exascale shows at least that a major fraction of the system can be engaged to tackle a dense system of linear equations. The path to the third stage — sustained exaflop performance on challenging user applications — is where many of the biggest hurdles lie. Prominent among these, as is well-known, are scaling the software ecosystem, providing enough reliability and resiliency to finish exa-jobs, and supplying enough IO to keep the heterogeneous processing elements busy. These are the same challenges advanced users face today, only more so.

The fact that governments have met HPC market forces partway is ultimately a good thing for all parties. It means that many of the government-supported advances for exascale computing will sooner or later benefit the mainstream HPC market, including SMEs that buy only a rack or two of technical servers.

The IO challenge is particularly nasty. In recent decades, HPC systems have become extremely compute-centric ("f/lopsided"). This increasing imbalance has aggravated the memory wall and narrowed the breadth-of-applicability for each succeeding generation of high-end supercomputers, especially for data-intensive simulation and the growing importance of advanced analytics. Fortunately, strategies are under way to alleviate (but not fix) this issue, including more capable interconnect fabrics, burst buffers and NVRAM, tighter linkages between CPUs and accelerators, clever data reduction methods, and more besides. But no one should expect supercomputers to return to the more balanced status of yesteryear. IDC vendor studies show that the basic architecture of HPC systems is unlikely to change in the next five to seven years, although configurations and some components will shift.

Not long ago, a fundamental premise underlying advanced supercomputer development was that evolutionary market forces were too slow and governments needed to stimulate revolutionary progress. The idea was that the government would do the heavy lifting to pave the way, and the mainstream HPC market would follow to take advantage of the revolutionary advances. In our annual HPC pre- dictions, IDC back then pointed out the risk that the government-supported high-end HPC market might split off as its own ecological niche, while the mainstream market continued to evolve on its own inertial path.

That split, though still possible, has not happened. Instead, government officials, for the most part, have realized that they are no longer the primary drivers of HPC. Market forces have usurped that role. The worldwide HPC mar- ket's diversification and ten-fold expansion in the past three decades, from $2 billion to more than $20 billion, has removed the government from the kingpin position it once held. Government officials in most HPC-exploiting countries have inflected their strategies to take better advantage of market forces, especially technology commoditization and open standards.

The fact that governments have met HPC market forces partway is ultimately a good thing for all parties. It means that many of the government-supported advances for exascale computing will sooner or later benefit the mainstream HPC market, including SMEs that buy only a rack or two of technical servers. That, in turn, means that savvy government officials can help justify the skyrocketing investments needed for extreme-scale supercomputers by pointing to ROI that benefits the large mainstream market, including industry and commerce. Government-driven advances can be used both to out-compute and to outcompete.

So, it appears that at least through the early exascale era, vendors will continue to build Linpack machines, because most government buyers will continue to see superior Linpack performance as a mark of leadership. Things might develop differently if more leading sites followed the example of the NCSA "Blue Waters" procurement, where the overwhelming stress was on the assessed needs of user appli-

The long-term good news is that HPC has become a mature market, one driven by market forces. That gives strong assurance that the market will behave rationally over time. Demand, in the form of buyer and user requirements, will increasingly win out.

cations and Lin- pack performance was not even reported. That was a deliberate decision, because "Blue Waters" is also a competent Linpack machine at heart and could have recorded impressive Linpack results. The point here is that, if lots of buyers gave primary consideration to user requirements in the procurements, this should lead to better system balance and wider applicability over time.

At the high end of the supercomputer market, money talks, too. Government funding appetites will play a major role in determining the sequence in which the entrants cross the exascale finish line. In earlier times, the global supercomputer race pitted "muscle cars" from the U.S. and Japan against each other, and these monoliths featured lots of custom technology. But today, as a successful Arnold Schwarzenegger once advised a neophyte bodybuilder, "it's not the size of your muscles that counts; it's the size of your wallet." Among governments, the U.S. is still the largest funder and the Obama Administration's budget request puts a high priority on exascale funding — although Congress has not approved this yet. The EU has been ramping up exascale funding, although not as fast as China, and Japan is likely to give everyone a run for their money.

World-leading supercomputers have not exactly morphed from muscle cars to family sedans yet, but they've been on that path — and it's generally a healthy one. The adoption of industry standards has been necessary for the expansion and democratization of the HPC industry, for broader collaboration, for better reliability, and for preserving and leveraging investments in software and hardware development. It's hard to imagine how vendors could make exascale muscle cars affordable, even for government buyers with the deepest pockets. The "Blue Waters" and CORAL procurements, among others, prove that, in the era of evolutionary HPC systems, important innovations can be pursued on behalf of users.

Governments around the world have increasingly recognized that HPC is a trans- formational technology that can boost not only scientific leadership, but also industrial and economic competitiveness. Accompanying this recognition is the notion that HPC is too strategic to outsource to another country, meaning to the U.S. in most cases. Exascale initiatives in Asia and Europe are promoting the development of indigenous technologies, often in conjunction with non-native components.

I've been talking so far about hardware, but we've said for some years at IDC that software advances will be more important than hardware progress in determining future HPC leadership. It's gratifying to see national and regional exascale initiatives increase funding for exascale software development, although the amounts still seem unequal to the task.

The long-term good news is that HPC has become a mature market, one driven by market forces. That gives strong assurance that the market will behave rationally over time. Demand, in the form of buyer and user requirements, will increasingly win out.

When data grows & demand increases, but budgets don't.

Meet the all-new **Saber 1000**

SATA 3.0 Enterprise SSD series.

Features

- OCZ's Barefoot 3 controller, in-house firmware, and next generation A19nm NAND from Toshiba
- Delivers sustained performance and consistent I/O responses for high-volume deployment enterprise SATA SSDs
- Supports Power Failure Management Plus (PFM+) to protect against unexpected power loss events
- Central management capability via OCZ StoragePeak 1000 software suite
- Available up to 960GB

Delivers great performance for today's typical datacenter applications

Read Cache & Indexing

Virtual Desktop Infrastructure (VDI)

Front-End Web Server

Media Streaming

Video on Demand (VoD)

Decision Support System (DSS)

Cloud Infrastucture

Customer Relationship Management (CRM)

Online Archiving

Video Editing/ Photo Sharing

Enterprise Content Management (ECM)

Virtual Tape Library (VTL)

OCZ STORAGE SOLUTIONS | A Toshiba Group Company

BECHTLE bechtle.co.uk | CCL cclonline.com | SPAN.COM span.com

ocz.com

Hp storage strategy: full flash ahead!

Following the announcement of its new 3PAR StoreServ 20800 full Flash storage arrays at the HP Discover conference, HP recently detailed its strategy for storage. The goal is to help companies accelerate their digital transformation.

The new 3PAR StoreServ 20800 product range is available in two versions, hybrid (3PAR StoreServ 20800) and full flash (3PAR StoreServ 20850). It is obviously the latter is emphasized, with an extremely low announced cost: 1.5 dollar per gigabyte. The advantages put forward on the hp StoreServ are its affordability, hyper scalability, high performance and high resistance.

According to hp, a successful digital transformation for an enterprise depends on making the right choices for storage. In order to deploy an IT infrastructure on demand, to se-

cure the digital assets of the company and as a way to improve productivity and helping the company to turn data into value-added information.

THE OBSTACLES THAT SLOW ADOPTION OF MODERN STORAGE SYSTEMS

Among the obstacles that have slowed the adoption of full flash storage arrays, are questions of a technical nature. Should it be a block-, file- or object-oriented file management system? Should there be converged or hyperconverged infrastructure elements? Should storage resources need to remain on premise within the company? Or be put in the cloud, or adopt a hybrid approach? Should storage remain traditional, hybrid or full flash? With what benefit and for which application context? Finally, should the new storage infrastructure be software defined, or Open Source? As we see, this cascade of choices is what determines the adoption of a strategic storage system for any company wishing to make its storage infrastructure evolve.

A SCHEME TO GUIDE THE COMPANY'S STORAGE STRATEGY

To guide companies to opt for the strategy that best suits its business and thereby its needs, HP recommends to choose the infrastructure according to four criteria arranged in two axes. On the X axis is on one end the SLA criterium for highly critical missions, and on the other the cost if economic considerations are the strongest. To complement this decision tree, the Y axis has on each end the two market approaches: complete solutions or hardware components.

A GUIDED COMPREHENSIVE SOLUTION BY SUCCESSIVE CHOICES

HP offers for its part two broad guidelines, the first for a software defined approach with StoreVirtual and the second for a system defined approach with its 3PAR offer. The second step is to complete this first brick with a converged (CS700 range) or a hyperconverged server (CS HC-200) to build up a solution for an optimal deployment. All are bound by the StoreOnce Data Services. Finally, HP recommends the deployment of common and standardized management interfaces like OneView integrating OpenStack. The alignment of all of these elements, according to Thierry Auzelle, hp France's storage is to allow «a real-world TCO calculation», lack of which explains why the investment decision has been delayed for so long.

AN AGGRESSIVE OFFER

Last year, the previous generation of 3PAR arrays reached more than 900,000 IOPS with a latency of 0.3 msec, for a cost of 2 dollars per Giga-byte using 1.92 TB SSDs. HP emphasizes its leading market position in the field of express indexing, with a capacity of 1.4 Peta-bytes of usable capacity. New this year is the upcoming availability of SSD 3.84 TB boasting a capacity boosted by 20% using Adaptive Sparing, a patented algorithm with a guaranteed thin provisioning mechanism and a 50% inline deduplication algorithm (done by a proprietary ASIC). The result is a multiplication of the density by 8, a 73% reduction of the price and an IOPS performance multiplied by 4. 3PAR Storeserv 20850 reaches 3.2 million IOPS with a latency of 0.2 to 0,8ms. Capacity on a rack is 5.5 Peta-bytes (280TB by 2U chassis) and can reach 12 Peta-bytes. Streamed asynchronous replication capabilities for long distances with up to one second allowed between the target and origin. A data control scheme named persist checksum ensures the integrity of transactions from one end to another. **JOSCELYN FLORES**

An IDC insight into flash technology adoption trends

During this presentation, Sebastien Lamour, an IDC consultant presented some figures on the challenges and perspectives of IT managers as part of the results of the 2015 Storage Observatory, a recurring survey of 18 companies with over 500 employees.

The finding about flash technology is that it introduces a disruption in storage technologies, doubled the following conclusions:

- The virtual infrastructure is at the heart of the future of the data center
- Growth in the volume of data continues as energy and floor space is limited
- The storage needs are massive and will remain so
- New approaches to significant efficiency gains appear
- New applications (VDI, I / O, transactional) present very varied needs

Sébastion Lamour concludes that new storage architectures have become necessary. In a context where the data center has become the first point of contact for a company and its storage architecture, these data centers operate 2-3 on average.

A TYPOLOGY OF EVOLVING DATA CENTERS

The types of data centers has evolved from the rest, with a concentration around large data centers, whose number increased by 15% between 2011 and 2014, a trend that is expected to accelerate and be around +31% between 2014 and 2017. Meanwhile, the smaller datacenter reduce in number: -18% between 2011 and 2014 and -6% expected between 2014 and 2017. The medium-sized data center for their part remain fairly stable, with a reduction of order -5% between 2011 and 2014, and projected to be -1% between 2014 and 2017.

The average physical servers running in the French data centers is 82 servers per datacenter. The installed base of servers deployed in French companies is at 1.08 million at present, against 1.11 million in 2014 and 1.21 million in 2011. A logical consequence of virtualization, which has evolved to reach 214 virtual servers per company, with an estimated growth of 14% in 2017. And a virtualization rate increasing accordingly: from 70% in 2015 it should increase to 78% in 2017.

Another interesting point is the evolution of the primary storage in data centers in France. Between 2015 and 2017, it increased from 105 to 117 TB per datacenter (+ 14% growth). To face this inexorable increase in data, new deduplication technologies are becoming essential but are not yet sufficiently deployed.

HIERARCHY OF PRIORITIES: FRENCH COMPANIES CAN DO BETTER

Another finding of the study: the priorities of CIOs do not place data storage in very good position. Currently, they are as follows (in order of priority 1, 2 and 3):

- Computer security: 30% / 30% / 19%
- Reducing costs: 22% / 18% / 14%
- Improved system performance for business: 14% / 11% / 21%
- Consolidation and virtualization systems: 14% / 13% / 16%
- The optimization and exploitation of data: 5% / 11% / 13%

The second observation is that the level of maturity

of French companies is very different. This maturity level is the adoption filter for innovative and disruptive offerings like full flash storage.

Thus, only 6% of companies have adopted full flash storage as of now. 30% of them have mixed flash-based storage and traditional disks. The storage in public cloud for its part is used by only 5.8% of companies. Conversely, converged storage offerings are a bit less established (up to 11%) than the software defined storage architectures (18%).

Flash storage technologies benefit of substantial investment intentions, since 36% of companies plan to spend between 10 and 50% of their storage needs on the flash. 26% expect less than 10%, and 36% of companies have no opinion on the subject.

DIFFICULTIES AND PRIORITIES IN THE NEW STORAGE LANDSCAPE

The difficulties identified by businesses on their existing storage infrastructure in this study are:

- Lack of performance for IO stream for 21% of businesses
- Lack of storage capacity to 27% of businesses
- insufficient time to conduct comprehensive backups for 21% of businesses

The priorities identified for the companies:

- Enhance or establish a Disaster Recovery Plan for 51% of businesses
- Ensure compliance and data retention for 58% of businesses
- Improving storage performance = 48%

The favorable challenges to adoption of Flash technology

- Speed and speed of access to data for 91% of businesses
- Low latency and playback time to 63% of businesses
- Energy consumption gains in 11% of them

The hindrances identified in flash storage adoption

- per-Gigabyte price for 58% of any company
- Lack of potential gains identified for 26% of businesses
- Lack of internal skills: 7%

SUMMARY AND RECOMMENDATIONS

The dependence of IT in business is becoming stronger. The use of historical approaches no longer suffices to achieve business goals, so it becomes imperative to work in order to align IT objectives with trade issues: SLAs, efficiency and costs.

Approaches to be considered to meet these business requirements:

- integrated and converged systems
- Find a superior integration with the next wave of hypervisors
- benefit from flash technology on the main I/O layers
- Adopt the quality of service and support as a differentiator
- Energy consumption has become a criterion increasingly important to take into account
- Emblematic workloads that can benefit from Flash are databases, virtualized environments, virtual desktops (VDI), electronic messaging.

Finally, Sebastian Lamour believes that the price and lack of enterprise-class robust software capable of performing the following operations: snapshots, replication, and provisioning of QoS are the major obstacles to the adoption of flash. However, Sebastian Lamour recommends not to stop at the only speed and cost criteria. High availability is crucial in the context of critical applications. For widespread adoption of flash infrastructure three elements must be reconciled : cost, performance and criticality.

Barbara Adey,
VP Business Development, Converged Data Center infrastructure, HP

HP designs the future of infrastructure

Through the API Composable Infrastructure and Redfish, hp offers the industry to become a partner to reinvent the business infrastructure of tomorrow.

In an exclusive interview during her visit to Paris, Barbara Adey. HP vice president in charge of the Converged Infrastructure business unit detailed the fundamentals of hp's composable Infrastructure initiative which she unveiled at the HP Discover conference in Las Vegas last June.

Over the past five years, it became clear that applications are the engine of our digital economy. The pressing question remains, how can companies migrate applications that sustain the businesses to the new generation of applications that determine the future of these companies? Hence, the need for a new class of infrastructure built to support future business needs quickly became a reality.

HP believes that this new infrastructure design will be «Composable», i.e. built on flexble and powerful pools of compute, storage and network, broken down so they can be quickly integrated, decomposed and recomposed in an on-demand software model to meet the specific needs of a given application or workload. More importantly, HP is developing this new infrastructure design class from existing solutions without requiring a new architecture.

Barbara Adey is Vice President Business Develoment Converged Datacenter Infrastructure (CDI), HP. Her role is to generate revenue, market share and margins. Before joining HP in January 2014, Barbara Adey was senior director of product management in the Security Technology Group at Cisco Systems. She was head of operations for the wireless routing group, security and Cisco, and previously worked in business and sales strategy. Barbara started her career in software development at Nortel Networks and has held marketing and consulting positions.

HPC Review: You have unveiled the composable infrastructure initiative at the recent Discover hp in the month of June. What is it about?

Barbara Adey: this is a new approach to anticipate and develop tomorrow's infrastructure. Our focus is on business CIOs and administrators, but also the datacenters that have everything to gain in improving their infrastructure responsiveness. Our operational focus is twofold. The first concerns the company's applications, which currently suffer from a lack of flexibility, even so when designed for the cloud. The second aspect that has focused all our attention was to 'thin' all the elements that make up the infrastructure of the company as of datacenter. That is, compute resources, disk and network.

The continuous supply of applications and services required by the new business models requires a rule-based automation for both applications and infrastructure through development, testing and production activities.

HPC: What is the purpose of your initiative Composable Infrastructure?

BA: The objective is to provide a form of business agility to companies that need to transform their business to more flexibility and responsiveness, by providing infrastructure to a form of intelligence that is currently lacking. To speed the path to this new class of infrastructure, HP has developed a Composable Infrastructure API that manufacturers make available to its partners through the HP Composable Infrastructure Partner dedicated program. The combination of both will allow the initial automation and orchestration tools to be integrated with current infrastructure. Among the partners that have already confirmed their engagement are Chief Software, Docker, Puppet Labs, Ansible, and VMware. In dealing with the physical infrastructure in the form of API commands, software vendors and developers can control their infrastructure to build new workflows with the HP Composable infrastructure. This single open API is integrated natively in HP OneView, which automates provisioning, configuration and monitoring of the HP infrastructure. By integrating HP Composable API, software vendors can provide solutions that enable customers to reduce the time spent managing their environments. By providing interoperability with this API, software vendors can support customer requirements for traditional IT environments and the digital economy growing rapidly. It should be noted that beyond the HP equipment, the initiative aims to eventually cover third party infrastructure elements.

HPC: Can you detail us the expanse and functional benefits of the Composable Infrastructure Partner Program?

BA: The continuous supply of applications and services required by the new business models requires a rule-based automation for both applications and infrastructure through development, testing and production activities. HP's Composable API enables integration into automation tools for Test / Dev and Production to drive a more aligned and responsive delivery of IT services to meet the needs of businesses. The HP Infrastructure Composable partnership program which is part of HP Alliance One, provides a set of tools and resources that enable software vendors and developers to create interoperability between HP OneView and other software for access programmatic infrastructure. Access to the program is available to all members of the Alliance One. As the initiative continues to unfold in the coming quarters, we look forward to bringing new partners on board and support our customers as they embrace a new style of infrastructure to meet their changing business needs.

HPC: How is this Composable Infrastructure initiative compatible with developments in SDS and SDN from hp?

BA: Our SDS and SDN developments rely on hardware resources which constitute the company's infrastructure. They are fully complementary, since the management layer and the intelligence-aware related Composable Infrastructure API determine the allocation of resources and bandwidth required by applica-

HP Dell, Emerson and Intel have come together with the intention of establishing Redfish as industry standard to facilitate the transition of clients to servers defined by software and thus make it compatible with the new requirements of Conduct activity based on an agile and responsive IT.

tions and virtualized business environments. The set provides administrators the required tools to simplify and automate the allocation but also the release of resources based on application requirements. Let us not forget that one of the growing criticism against the current infrastructure is the waste of resources induced by allocating resources that are no longer in use, but continue to function (called overprovisioning). A simple six month timeout will allow administrators to keep an eye on their resource consumption, and thereby reduce investment by adjusting the resources in line with the actual needs of the business with a high level of granularity. From this point of view the operation as a unified pool has its meaning.

HPC: Can you tell us more about Redfish, which HP helped to make it an industry standard?

BA: Redfish is an emerging industry specification for the management of the datacenter for customers who require independent interfaces of suppliers, based on light and modern paradigms llike REST. The objective is to realize the benefits such as improved scalability and security. Redfish is for companies considering modern approaches of data center management software.

HP Dell, Emerson and Intel have come together with the intention of establishing Redfish as industry standard to facilitate the transition of clients to servers defined by software and thus make it compatible with the new requirements of Conduct activity based on an agile and responsive IT. The design of Redfish around the REST API is in line with HP's strategic plans for the management of the data center.

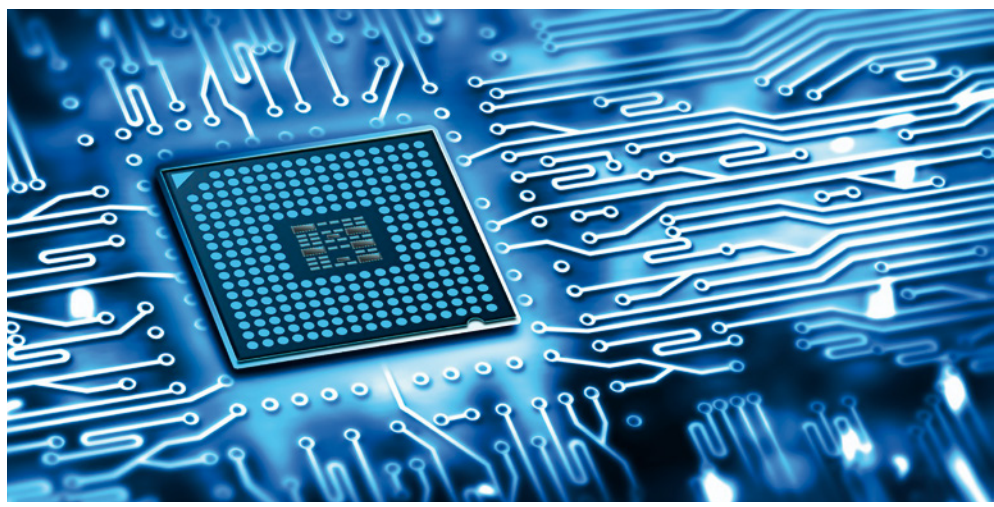
Standardization of data center management can reduce costs, improve security and productivity. Redfish uses modern REST based interface to improve and expand access to data and analysis, system-to-system communication, remote management, security features and scalability. This specification eliminates the need to know protocols and use tools specific software environments of data centers and server management.

HPC: Finally, can you detail the upcoming strategic developments aimed at enterprise infrastructure and data centers?

BA: Certainly. We have already defined the four steps ahead. The first is devoting a unified API, and is at the heart of our Composable Infrastructure initiative. The second initiative is to make available to our customers new infrastructure natively compatible with this API. Our third objective is to help to ensure continued provision of services. Finally, we also intend to build on the progress made on The Machine to integrate our infrastructure development strategy in the coming years. One could imagine the elements of a whole new genre to accelerate infrastructure performance, like datacenter-ready ultrafast memristor memory filled integrated enclosures... but we're not there yet. INTERVIEW BY **JOSCELYN FLORES**

IBM, NVIDIA and Mellanox launch a design center for Big Data and HPC

IBM, Nvidia and Mellanox have announced plans to launch a new European OpenPower design centre with the purpose of developing high performance computing (HPC) apps. The centre, which will be situated in Montpellier, France, will work alongside the Jülich Supercomputing Center in Germany, which was launched at the end of last year. Unlike the Jülich centre, however, this new facility will focus on HPC apps built using the open source Power architecture, as part of the OpenPOWER Foundation.



FOCUS ON THE DEVELOPMENT OF HIGH PERFORMANCE COMPUTING APPS

Developers from all three companies will work together to create the applications using IBM Power CPUs, Nvidia's Tesla Accelerated Computing Platform and Mellanox's InfiniBand networking solution. The collaboration will also help software developers learn advanced HPC skills, boosting the development of new technologies in big data to serve customers around the world.

«Our launch of this new centre reinforces IBM's commitment to open-source collaboration and is a next step in expanding the software and solution ecosystem around OpenPower,» said Dave Turek, IBM's vice president of HPC Market Engagement. «Teaming with Nvidia and Mellanox, the centre will allow us to leverage the strengths of each of our companies to extend innovation and bring higher value to our customers around the world,» he added.

AN ALLIANCE AIMED TOWARDS EXASCALE COMPUTING

Stefan Kraemer, director of HPC Business Development, EMEA, at NVIDIA added: «Increasing computational performance while minimizing energy consumption is a challenge the industry must overcome in the race to exascale computing. «By providing systems combining IBM Power CPUs with GPU accelerators and the NVIDIA NVLink high-speed GPU interconnect technology, we can help the new Center address both objectives, enabling scientists to achieve new breakthroughs in their research,» Kraemer concluded.

«The new POWER Acceleration and Design Center will help scientists and engineers address the grand challenges facing society in the fields of energy and environment, information and health care using the most advanced HPC architectures and technologies,» said Gilad Shainer, VP marketing, Mellanox. «Mellanox IB networking solutions offer more than a decade of experience building the world's highest performing networks, and are uniquely based on an offload-architecture. Only Mellanox offloads data movement, management and even data manipulations (for example Message Passing - MPI collective communications) which are performed at the network level, enabling more valuable CPU cycles to be dedicated to the research applications.»

As founding members of the OpenPOWER Foundation, IBM, NVIDIA, and Mellanox share a common vision to bring a class of systems to market faster to tackle today's big data challenges.

AMD FirePro™ S9150 Server GPU

The GPU of choice for high-performance computing

The most compute-intensive workloads in data analytics or scientific computing are no challenge for the AMD FirePro™ S9150 server GPU. With support for OpenCL™ 2.0, 16 GB GDDR5 memory, and up to 2.53 TFLOPS of peak double-precision and up to 10.8 GFLOPS-per-watt peak double-precision performance, the choice is clear.

The AMD FirePro™ S9150 features:

- Up to 5.07 TFLOPS peak single-precision floating point performance
- Up to 2.53 TFLOPS peak double-precision floating point performance
- 16 GB of GDDR5 memory
- Up to 320 GB/s memory bandwidth
- AMD Graphics Core Next (GCN) Architecture
- Full Rate Double Precision
- AMD STREAM Technology
- AMD PowerTune Technology
- OpenCL™ 2.0 support

THE GREEN 500

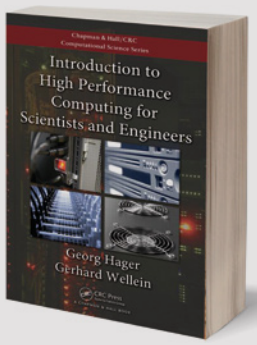
#1 on the November, 2014 Green500 list [1]

Please visit ***www.fireprographics.com/s-series*** *to find out where you can get the AMD FirePro S9150.*

1. AMD FirePro™ S9150 server GPU powers the #1 supercomputer on the November, 2014 Green500 list.
For more details, please visit http://www.green500.org/news/green500-list-november-2014

2. AMD FirePro™ S9150 max power is 235W and delivers up to 2.53 TFLOPS peak double and up to 5.07 peak single precision floating point performance. Nvidia's highest performing single-GPU server card in the market as of March 2015 is the Tesla K40, max power of 235W, with up to 1.43 TFLOPS peak double and up to 4.29 peak single-precision compute performance. FP-97

BOOKS



Introduction to High Performance Computing for Scientists and Engineers

Written by high performance computing (HPC) experts, Introduction to High Performance Computing for Scientists and Engineers provides a solid introduction to current mainstream computer architecture, dominant parallel programming models, and useful optimization strategies for scientific HPC. From working in a scientific computing center, the authors gained a unique perspective on the requirements and attitudes of users as well as manufacturers of parallel computers. After discussing parallel computing on a theoretical level, the authors show how to avoid or ameliorate typical performance problems connected with OpenMP. They then present cache-coherent nonuniform memory access (ccNUMA) optimization techniques, examine distributed-memory parallel programming with message passing interface (MPI), and explain how to write efficient MPI code. The final chapter focuses on hybrid programming with MPI and OpenMP. **Georg Hager, Gerhard Wellein, Chapman-Hall / CRC Press (360 pages, £43.34)**

Designing Scientific Applications on GPUs

Many of today's complex scientific applications now require a vast amount of computational power. General purpose graphics processing units (GPGPUs) enable researchers in a variety of fields to benefit from the computational power of all the cores available inside graphics cards. From physics and mathematics to computer science, this book explains the methods necessary for designing or porting your scientific application on GPUs. It will improve your knowledge about image processing, numerical applications, methodology to design efficient applications, optimization methods, and much more. This book covers many numerical applications, including obstacle problems, fluid simulation, and atomic physics models. The last part illustrates agent-based simulations, pseudorandom number generation, and the solution of large sparse linear systems for integer factorization. Some of the codes presented in the book are available online. Raphael Couturier, Chapman-Hall / CRC Press (498 pages, £50.99)

The Green Computing Book: Tackling Energy Efficiency at Large Scale

Edited by one of the founders and lead investigator of the Green500 list, this book explores seminal research in large-scale green computing. It begins with low-level, hardware-based approaches and then traverses up the software stack with increasingly higher-level, software-based approaches. In the first chapter, the IBM Blue Gene team illustrates how to improve the energy efficiency of a supercomputer by an order of magnitude without any system performance loss in parallelizable applications. The next few chapters explain how to enhance the energy efficiency of a large-scale computing system via compiler-directed energy optimizations, an adaptive run-time system, and a general prediction performance framework. The book then explores the interactions between energy management and reliability and describes storage system organization that maximizes energy efficiency and reliability. It also addresses the need for coordinated power control across different layers and covers demand response policies in computing centers. The final chapter assesses the impact of servers on data center costs. Wu-chun Feng, Chapman-Hall / CRC Press (353 pages, £46.74)

MOOCS

Data Analyst Nanodegree

This Nanodegree program is the most efficient curriculum to prepare you for a job as a Data Analyst. You will learn to:
· Wrangle, extract, transform, and load data from various databases, formats, and data sources
· Use exploratory data analysis techniques to identify meaningful relationships, patterns, or trends from complex data sets
· Classify unlabeled data or predict into the future with applied statistics and machine learning algorithms
· Communicate data analysis and findings well through effective data visualizations

You will work with your peers and advisors on projects approved by leading employers as the critical indicators of job-readiness. We designed these projects with expert Data Analysts, Data Scientists, and hiring managers. You'll also have the opportunity to prepare for your new career with reviews of your online presence (resume, LinkedIn, portfolio), prepare for interviews, take part in workshops covering topics like networking and salary negotiation as well as take part in a new program facilitating job placement.

Next enrollment opens August 11 at 12:00 pm Eastern Time. You can start the program immediately after you enroll.

$200/month After 1 week free trial
9–12 months
Minimum 10hrs/week.
Work on your own schedule.

https://www.udacity.com/course/data-analyst-nanodegree--nd002

Cryptography I

Cryptography is an indispensable tool for protecting information in computer systems. This course explains the inner workings of cryptographic primitives and how to correctly use them. Students will learn how to reason about the security of cryptographic constructions and how to apply this knowledge to real-world applications. The course begins with a detailed discussion of how two parties who have a shared secret key can communicate securely when a powerful adversary eavesdrops and tampers with traffic. We will examine many deployed protocols and analyze mistakes in existing systems. The second half of the course discusses public-key techniques that let two or more parties generate a shared secret key. We will cover the relevant number theory and discuss public-key encryption and basic key-exchange. Throughout the course students will be exposed to many exciting open problems in the field. The course will include written homeworks and programming labs. The course is self-contained, however it will be helpful to have a basic understanding of discrete probability theory.

Length: 6 weeks
Effort: 5 - 7 hours per week
Institution: University of Stanford
Languages: English
Subtitles : Portuguese and english
Link : https://fr.coursera.org/course/crypto

THE HPC OBSERVATORY

$44 Billion

This is the projection of the worldwide turnover of HPC in 2020. Market Research Media research firm expects the area of supercomputing will grow an average of 8.3% annually to reach $ 44 billion in 2020 . This sector will also generate 220 billion dollars in cumulative sales over the period 2015-2020. *Source : http://www.marketresearchmedia.com/*

THE TOP 3 OF THE TOP 500

1 Tianhe-2
National Supercomputing Center in Canton: 33863/54902 TFlops Manufacturer: NUDT**, architecture** Xeon E5-2692 Xeon Phi 31S1P +, TH Express-2

2 Titan
OOak Ridge National Laboratory, USA: 17590/27113 TFlops Manufacturer: Cray XK7**, architecture** Opteron 6274 + Nvidia Tesla K20X Cray Gemini Interconnect

3 Sequoia
Lawrence Livermore National Laboratory, USA: 17173/20133 TFlops Manufacturer: IBM Blue Gene / Q **architecture** PowerPC A2

The TOP500 ranks every six months the 500 most powerful supercomputers in the world. The two selected values, RMAX and RPEAK represent the computing power and maximum theoretical Linpack.

THE TOP 3 GREEN 500

1 5271,81 MFlops/W
GSI Heimholtz Center (Germany)
57.15 kilowatts consumption

2 4945,63 MFlops/W
High Energy Accelerator KEK (Japan)
37.83 kilowatts consumption

3 447,58 MFlops/W
GSIC Center, Tokyo Institute of Technology (Japan)
35.39 kilowatts consumption

Green 500 rankis the most energy efficient supercomputers in the world. Energy efficiency is assessed by measuring performance per watt. The unit is here MFLOP / Watt.

HPC Today

The global media in high-performance IT

Subscribe now > for FREE

Subscribe now and receive, every month,
an expert, actionable coverage of HPC and
Big Data news, technologies, uses and research...

+ get yourself instant access to exclusive contents and services

www.hpctoday.com

IS QUANTUM COMPUTING READY FOR PRIME-TIME?

As Moore's Law demonstrates, the IT sector is undoubtedly the most dynamic since its inception 70 years ago. If computing architectures have followed from the start the Von Neumann model, researchers and R & D departments have always pursued the dream of next generation computers. The quantum computer is such a dream, slowly becoming reality as the technical barriers are slowly but enduringly overcome.

If we could build a quantum computer with only 50 quantum bits (qubits) instead of four, none of the TOP500 supercomputers today could outstrip its performance – IBM

The search for a successor to current architectures has a technical origin. According to the Moore's law, the size of transistors will approach that of the atom as soon as 2020. At this scale, quantum effects disrupt the operation of the components. The current photolithographic engraving technology may well prevent any architectural evolution by conventional means. Just as the multiplication of execution cores within the processors succeeded the race for clockrate during the past decade, architectural developments are at the heart of the quantum approach and aims to multiply the capacity of the future computers.

HOW A QUANTUM COMPUTER WORKS

Quantum computing is based on quantum bits or qubits. Unlike traditional computers, wherein these bits can only have a value of zero or one, a qubit can be a zero, one, or both values simultaneously. This representation of information allows qubits to process information in a manner that no classical computer does, enjoying phenomena such as tunneling and quantum entanglement. As such, quantum computers can theoretically be able to solve some problems in days where traditional architectures could take up to millions of years to solve the same problem. The promise is phenomenal! But the pitfalls many.

WHAT IS A QUANTUM COMPUTER USEFUL FOR?

Unlike current computers based on the Von Neumann model and which are based on the exchange of the four elements that distinguish it, namely the arithmetic unit processing, the control unit, the memory and the input output devices, a quantum computer requires few inputs and low outputs. It therefore lends itself for calculations whose complexity lies in the combinatorial algorithms. We find these problems in scheduling, operational search calculations, bioinformatics, and cryptography. This low volume of input-outputs to process and solve problems further predisposes them to remote use through the Internet.

A DIZZYING POTENTIAL

If large quantum computers (more than 300 qubits) could be built – which is unlikely at this moment - they would be able, according to David Deutsch, a British physicist and professor of physics at Oxford University, to simulate the behavior of the universe itself. They could also solve cryptanalysis problems in a much shorter time than a conventional computer, because increasing linearly (N) with the size N of the key, and not exponentially (To 64n, for example) as with sequential brute force methods. Quantum computers require different calculation techniques to be programmed, while still using classical linear algebra methods. But this technology suffers of reliability issues.

The reason being that the qubit, the quantum information storage unit, is fragile and sensitive to changes of the temperature and the magnetic field, thereby generating errors. And this occurs even if the work is carried out at a temperature close to absolute zero.

D-WAVE: THE HAMPERED PIONEER

D-Wave announced officially on February 13, 2007 to have created the world's first 16-qubit

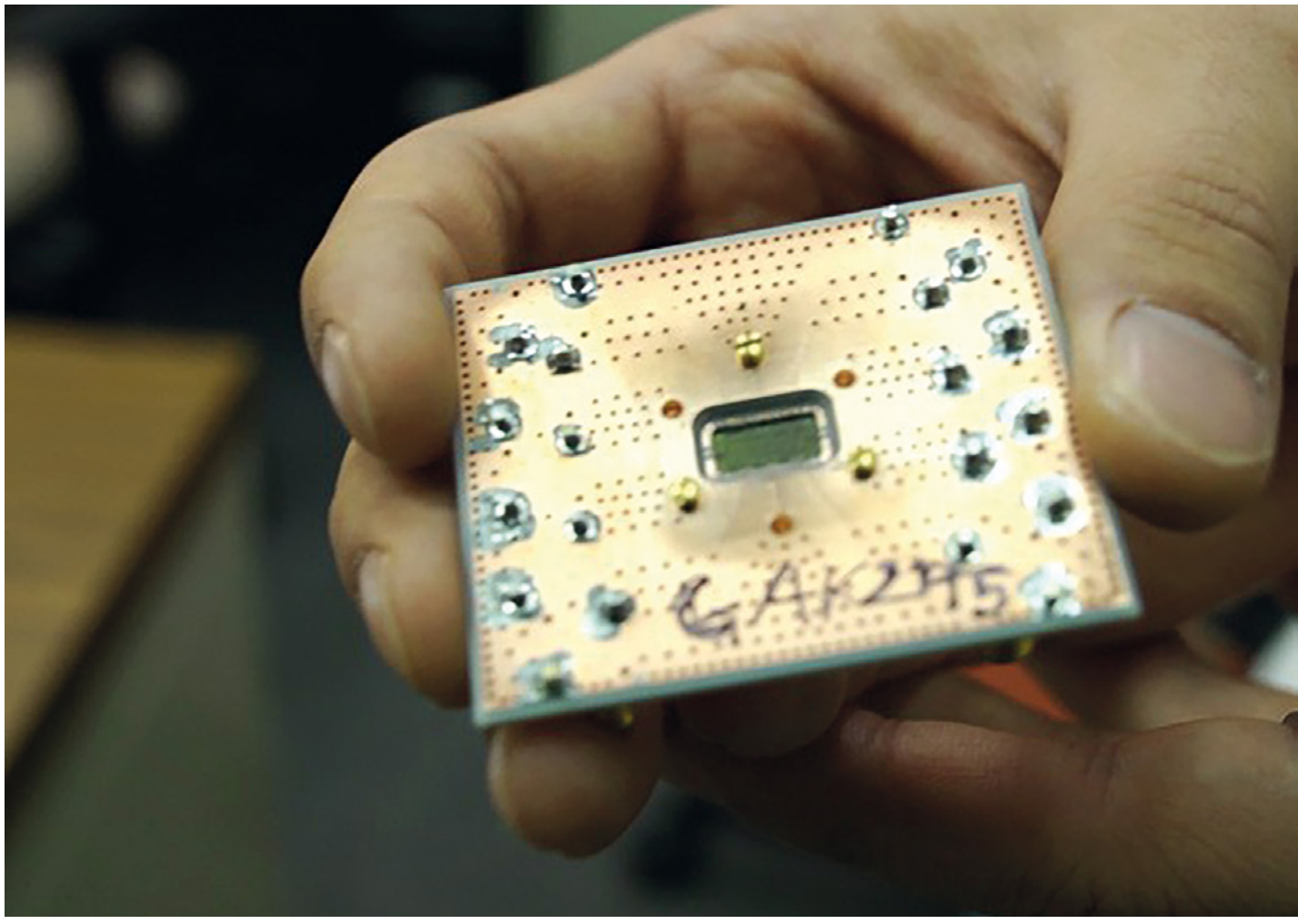
quantum computer. This computer would however be limited to certain quantum optimized calculations. The combinatorial problems (like sudoku) are however solved slower than with a regular computer. At the heart of the D-Wave system 2 (adopted by NASA and google) lies the SQUID, a proprietary quantum transistor that can manipulate superconductive qubits, which are the basic bricks of a quantum computer. SQUID is an acronym meaning Superconducting Quantum Interference Device. The term «interference» refers to electrons, acting as waves within quantum waves, which in turn create interference patterns which give rise to quantum effects. The reason why quantum effects such as electron waves are held in place in the structure – allowing them to act as a qubit - Is due to the properties of the material of which they are made. Squid consists of metallic niobium (As opposed to conventional transistors made from silicon). When the metal is cooled, it becomes superconductive, and begins to exhibit quantum effects. The superconducting structure encodes their states like tiny magnetic fields, pointing up or down. We call these states +1 and -1, and correspond to the two states that the qubit can adopt. The use of quantum mechanics can control one, several or all of the qubits to adopt the superposition of these two states. Although D Wave 2 contains 512 qubits this quantum computer suffers from two flaws: their lifespan is limited and the computational scope is limited to specialized tasks like Machine Learning, pattern recognition and detection of abnormalities, Financial Analysis and verification and validation process. Furthermore it is not independent and must operate besides a traditional computer, which makes D-Wave2 a quantum coprocessor still far from universal quantum computer pursued by Google, IBM and Microsoft.

GOOGLE: SOLVING THE CORRECTION ERROR AND CORRUPTION PHASE

The University of California, combined with Google, has developed an error correction mechanism in qubits, the quantum equivalent bits. But these errors are today one of the main obstacles to the design of quantum computers. The teams of Physics professor John Martinis of the University of California, associated with Google since last September, just crossed a significant step: the reliability of quantum computers.

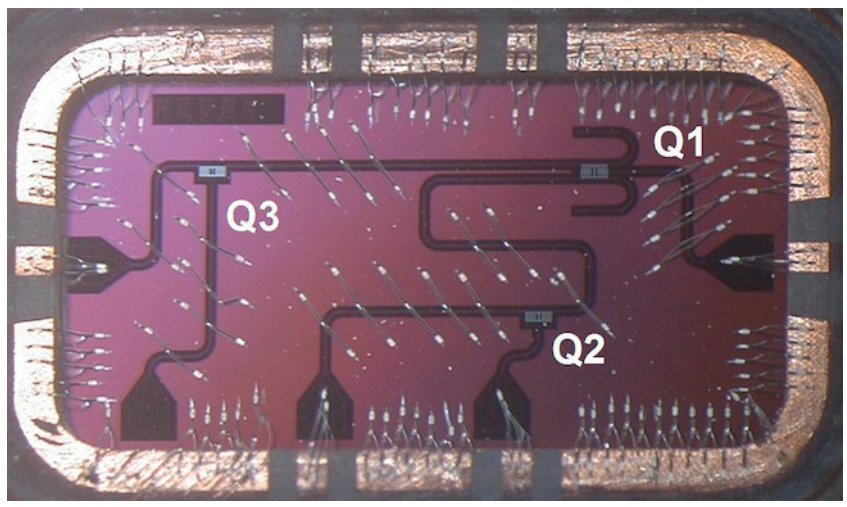
The teams of the University of California and Google have managed to program a chip containing 9 qubits able to monitor each other to detect bit inversion errors. The method imagined does not correct these errors, but prevents them contaminating the steps of a calculation. The team of John Martinis explains that the mechanism they imagined reduced the error rate by a factor of 2.7 when 5 qubits are used, and by a factor of 8.5 when the

9 elements are active. «Other studies have yet be conducted before we can say error-free quantum calculations are possible, «says Daniel Gottesman, who works on error corrections at the Perimeter Institute in Canada. If the inversion of bits that treats John Martinis, can be supported by classic algorithms, another type of error, the alteration of a qubit property called the qubit phase, requires much more complex calculations. In the MIT Technology Review, Austin Fowler, an engineer at Google, ensures that the teams of Mountain View and the University of California are working specifically on the subject of phase alteration as well as an error detection mechanism over 9 qubits.



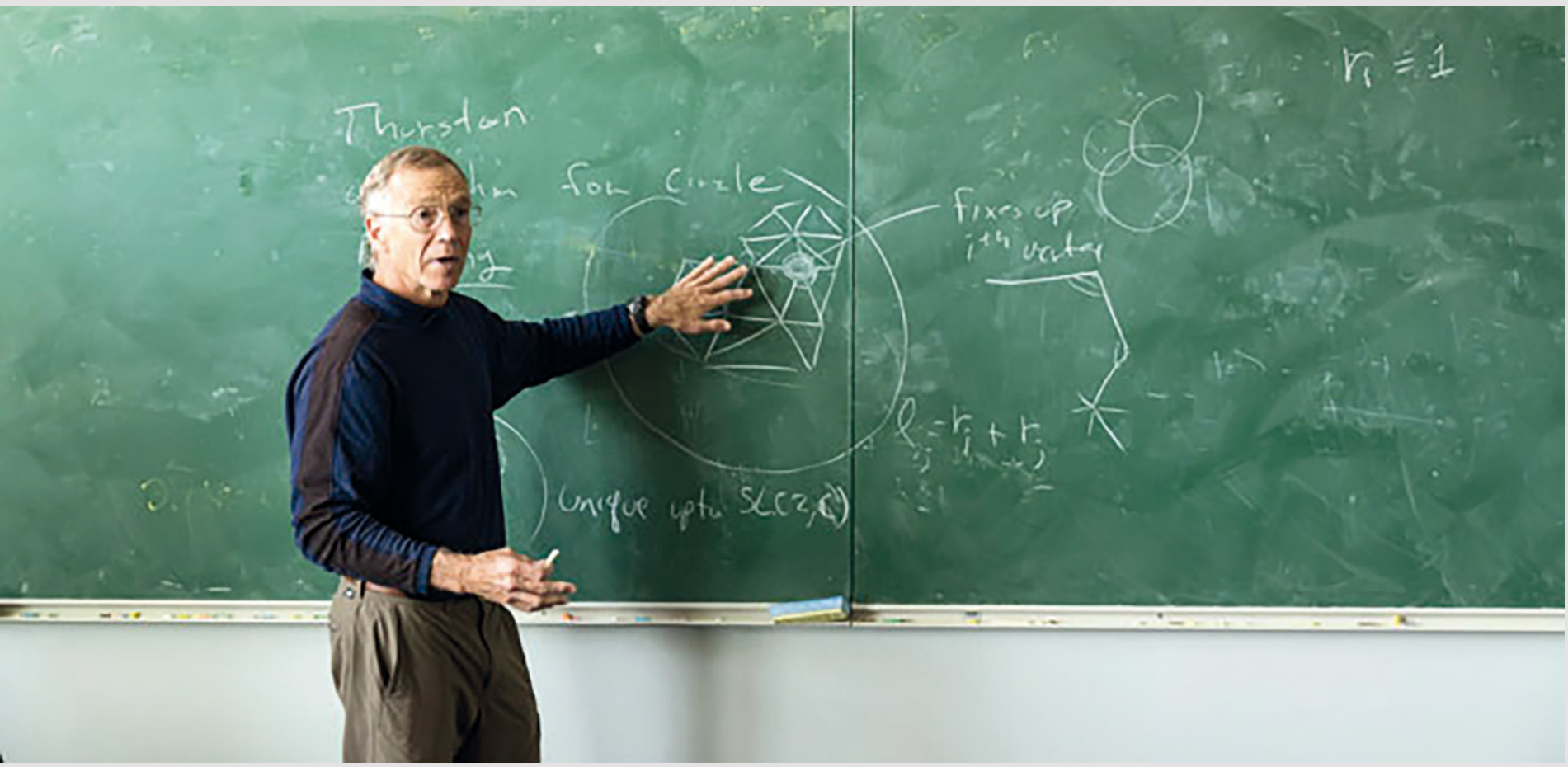
IBM: SUPERCONDUCTIVITY AND SQUARE DESIGN

As we see, a quantum computer will only work when quantum decoherence is eliminated, namely the emerging errors in the calculations because of the ambient temperature, or even electromagnetic radiation. The qubits are extremely sensitive, their simple measurement can change their condition. It is very possible to have flip type bit errors, which equates to obtain the opposite state (1 instead of 0, for example). Or still fall on a turnaround phase error, which can flip the state sign (+ instead of -). Quantum error correction is necessary in any large-scale reliable quantum computer design. So far, there was only possible to detect either of these two phenomenons simultaneously. The IBM solution is a quantum bit circuit which is based on a square lattice of four supercooled superconductive qubits placed on a chip with an area of about 60mm². The square shape allows the circuit to work for the resolution of the quantum error correction. This form also enables scaling by adding more qubits.



Previous work in this area, using linear arrangements, “only allowed observation of the bit inversion errors hence providing incomplete information on the quantum state of a system» says Jay Gambetta from IBM’s Quantum Computing Group. «Our work allows us to solve this obstacle by detecting the two types of quantum errors. Even better, they are transferable to larger systems as the qubits are arranged in a square configuration,

Station Q, Microsoft Research's quantum research hub

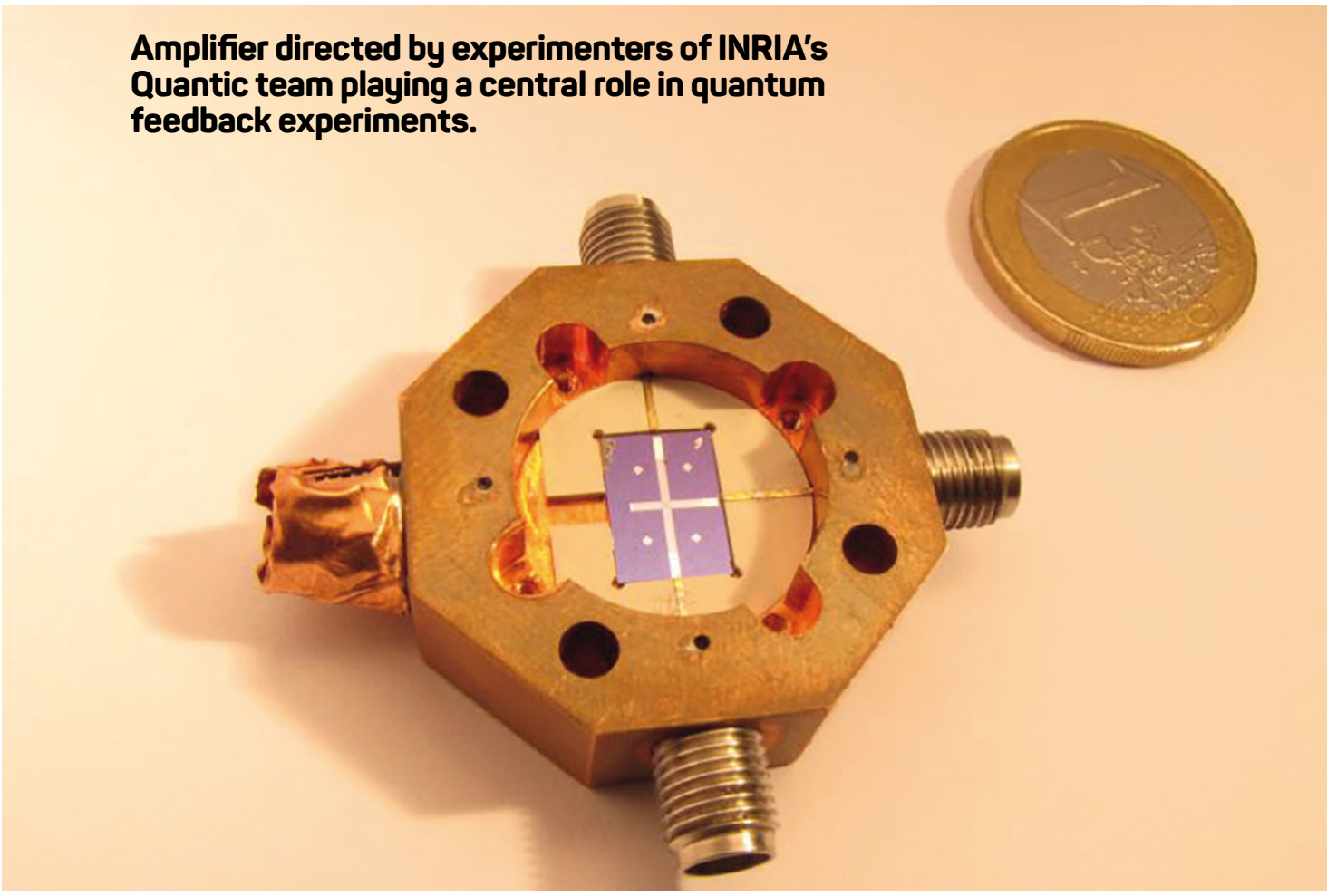


Station Q is located on the campus the University of California, Santa Barbara. This quantum research facility is headed by Michael Freedman, whose main interest since ten years is the quantum computer. According to Freedman such a computer should be able to solve problems that would take thousands of years – even more - in less time than it takes to make a mug of coffee. With unheard of uses in Machne Learning, medicine, chemistry, Cryptography, science of materials and engineering. A universal quantum computer could enable mankind to understand and control the elements that constitute our universe. «Quantum computing «says Freedman «Could have enormous consequences on calculation. The truth is that we don't know it yet. «The main research axis at Station Q differs from those of D-Wave Systems, Google and IBM and covers topologic qubits derived from a particle called Majorana fermion, believed to be more reliable and better suited to mass production than those designed with other techniques.

as opposed in a linear array. The next step is to design and manufacture a handful of reliable superconducting qubits with low error rates. Once done, we could very well be on the way to a complete quantum computer. If we could build a quantum computer with only 50 bits quantum bits (qubits) instead of four no TOP500 supercomputers today could not succeed to surpass its performance - Which would be absolutely amazing. "

INRIA: MISSION QUANTIC

Mazyar Mirrahimi, research director at the French research institute INRIA and head of the Quantic Mission estimates it is necessary to obtain a full-blown quantum computer, to



Amplifier directed by experimenters of INRIA's Quantic team playing a central role in quantum feedback experiments.

couple several quantum systems, which seems very difficult in quantum optics. This problem led researchers in the field of mesoscopic physics to prefer superconductive circuits at low temperatures in order to create the logic gates and memories needed by a quantum computer. We feel the need of a systems engineering that obeys quantum rules, that we could call «quantum engineering laws". In terms of impact, Mazyar Mirrahimi considers that there is a whole field of possible applications including metrology measurement accuracy improvement, as we did it for the atomic clock. We could consider, for example, improving and stabilizing the amplitude measurements of a magnetic field. Some applications like quantum cryptography and quantum communications are based on these laws and are rather easier to achieve. There are already industrial prototypes that can communicate information by optical fiber in a quantum encrypted way by polarizing the light photons. Solving problems not accessible to current computers involves to be able to manipulate thousands of qubits. However, quantum superposition (The possibility for a qubit to be simultaneously 0 and 1) is very fragile. To go beyond 9 qubits implies an increase in the risk of developing new noise sources, in particular, correlated noise (which can affect many qubits at a time). We speak here of the problem of «scalability»: how to significantly increase the size of the quantum system without deteriorating the properties of its subsystems? This is the heart of the problem in developing a universal quantum computer. As we see, the potential of the quantum computer is such that it alone justifies the investment made by the world's biggest players and research institutes. If it has not the ability to revolutionize day to day computing, the acceleration potential in research and specific treatments is mind-blowing. The state of research is nevertheless encouraging and allows the academic and industrial sectors to work together to solve problems on the way. The results may still be years away, but for such a young sector the progress in only a decade are very encouraging ! **JOSCELYN FLORES**

How to significantly increase the size of the system without damaging the properties of its subsystems?

Lab Review

How do we test ?

HPC Labs

HPC Labs is the technical unit of the HPC Media group and totally independent of the manufacturers. HPC Labs' mission is to develop methodologies and materials testing and software metrics in the high performance IT world. Capitalizing on best practices in the field, these tools are based on several decades of joint experience of the laboratorys' management.

HPCBench Solutions

Specifically designed for HPC Review, the HPCBench Solutions assess not only performance but also other equally important aspects in use, such as energy efficiency, sound volume, etc. To differentiate synthetic protocols like Linpack, these protocols allow direct comparison of solutions pertaining to the same segment, resulting in a single index taking into account the specific hardware or software tested. For example, an SSD will be tested with the HPCBench Solutions> Storage, while a GPU accelerator will be tested with the HPCBench Solutions> accels. Rigorous and exhaustive, these protocols allow you to choose what will be for you, objectively, the best solution.

HPC Bench Global index 9 108

A single synthetic index to help you compare our test results

HPC REVIEW AWARD

A technical recognition Award

Fujitsu Eternus CD10000

When HyperScale reaches towards Exascale

With the Eternus CD10000, Fujitsu launches the world's first storage platform designed to grow as long as online data will be needed. An offer adapted to organizations in need of petabyte capabilities for Big Data processing.

The ETERNUS CD10000 is a new storage offering that, unlike conventional disk arrays, uses conventional server nodes and open source software (RedHat Ceph on CentOS). This scale-out solution can store a total capacity of 56 Peta-bytes. It is particularly destined to companies and operators who assemble today an OpenStack private cloud infrastructure.

One response for three needs

As a platform with virtually unlimited storage capacity, CD10000 is designed to meet the exponential growth of data. With a total volume of data generated and stored online that continues to multiply, organizations face three major problems: increased demand on scala-

The Eternus CD10000 comes in the form of a fully autonomous 42U rack containing all the servers, storage arrays, disks and connectivity.

bility, greater complexity and associated costs, and to circumvent the physical limitations to be able to move data between storage systems without disruption. Together, these three factors advocate in favor of a new approach to enterprise storage with volumes of data that can be counted in tens of Peta-bytes. To give some perspective, a Peta-byte equals about 100,000 hours of full HD 1080p video.

A CONCENTRATED DISTRIBUTED ARCHITECTURE IN A RACK

The Eternus CD10000 comes in the form of a fully autonomous 42U rack containing all the servers, storage arrays, disks and connectivity. The operating system used is Ceph, an open source software that can expose the contents as a file, block and object from a cluster of storage nodes distributed objects on which data is distributed. In theory, Ceph is capable of handling up to 1 exabyte (1000 Peta-bytes, or 1 million Terabytes), leaving a substantial margin of evolution.

THREE CHOICES OF NODES TO ADDRESS THREE TYPES OF MISSIONS

In hardware terms, the Eternus CD10000 can amount up to 224 nodes which communicate via InfiniBand dual channel connections of 40 Gbits / s. The CD10000 may have three types of nodes:

• Base node - each containing two Xeon processors and 12.6To raw capacity with 16 900 GB 10000 rpm SAS drives and 2.5-inch SSD PCI Express for caching, logging and metadata.

• Capacitive node - 252To each containing 60 3.5-inch SATA drives to 14 900GB 7,200 rpm SAS drives.

• Performance oriented node - each containing 34.2To raw capacity using SAS 2.5-inch 10000 rpm and PCIe SSD.

The network interfaces through a 10Gbit Ethernet link and its resources are exploitable via KVM, Swift and S3.

ADJUSTABLE CAPACITY AND RESILIENCE

Usable capacity of the system depends of course on the number of data replicas (eg two or three) in place to protect against data loss. In a system of 224 nodes, the total number of discs amounts to nearly 13,400 discs. To improve redundancy and data resilience, Fujitsu decided to bypass RAID and uses a replication mechanism and self-healing with an almost zero downtime according to Fujitsu. The level of protection may be further improved by increasing the number of copies, one to two, three or four depending on their criticality.

A scalable platform designed to last

The Eternus CD10000 is designed to evolve with the emergence of newer technologies. The CD1000 HyperScale architecture allows to add, change and update the storage nodes without service interruption. The racks can notably be upgraded without stopping the machine, the software undertaking to move data to use the new nodes, operation after which the old racks can be décommissionned. Fujitsu is actively working with partners and customers to add applications CD10000 system to equip it to cloud services, file synchronization, and data archiving.

RAMON LAFLEUR

StorageCraft ShadowProtect Desktop 5.2.3

HPC REVIEW AWARD

Developed by the former teams of the legendary partitioning tool Partition Magic, ShadowProtect Desktop is a software entirely dedicated to the backup and restoration of disks and partitions. With flawless reliability and efficiency.

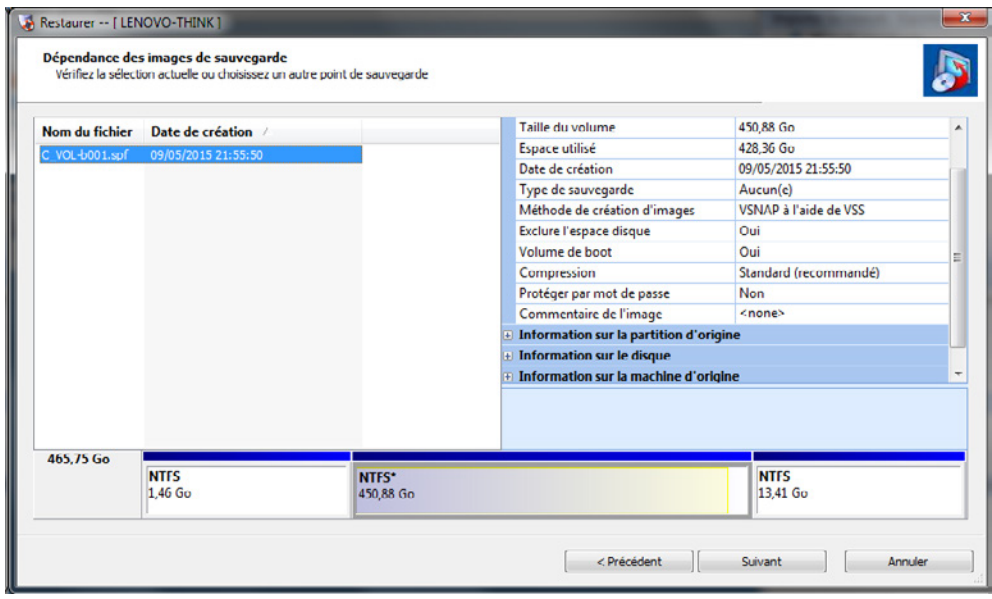
ShadowProtect Desktop is dedicated to the creation and operation of disk images that contain a complete snapshot of a partition on your disk, which you can use to restore your entire system in a single operation. It is also possible to explore the contents of a disk image to copy individual files and folders. The Desktop application is designed for efficiency for usability, but even moderately experienced user can use it. All one need to know is to navigate the Windows file system.

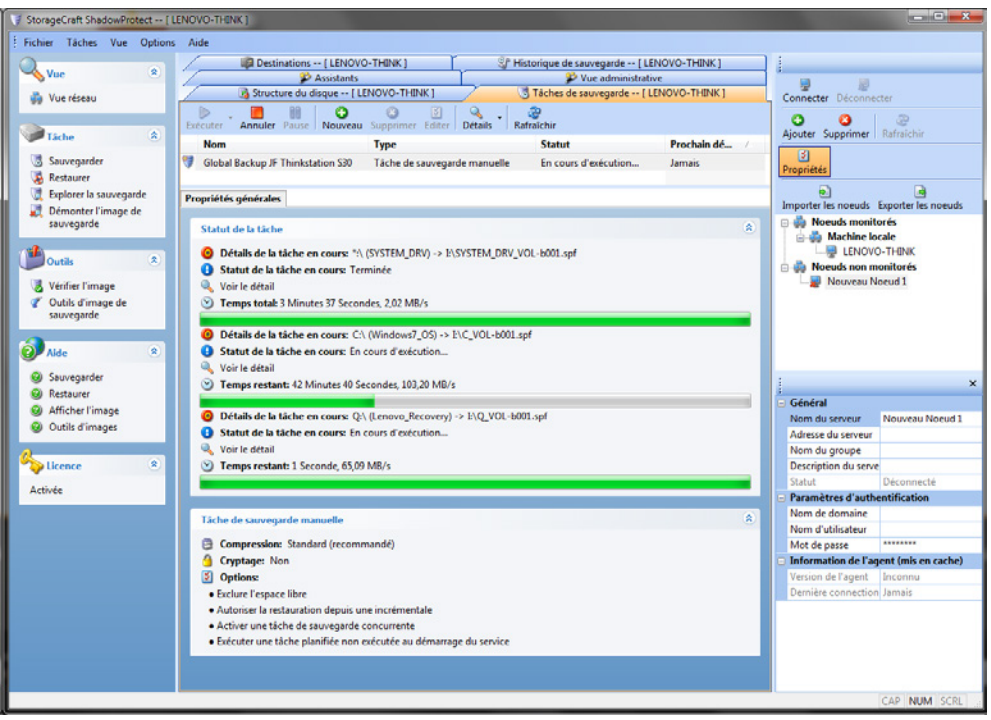
A WIZARD TO GET STARTED

At startup, the software offers a window with three wizards that will first guide you to your first backup. Just follow the instructions to select a backup partition and the location where you want to save the backup. You can opt for an external USB drive, but the destination selection dialog allows you to navigate to another system or network storage device. You also have the leisure, if it is essential for you to keep fast access to old versions of your files, to save the backup on another partition on your primary hard drive.

EXTENDED CONFIGURATION SETTINGS

The next dialog box allows you to schedule your backup. You can change the default setting and choose the day of the week for a full backup. Note that you also here the choice of the type of backup. The software simultaneously allows two types of backups, full and incremental. The incremental backup method uses continuous snapshots (default is every hour on weekdays) and boasts abundant options to adjust the desired backup time ranges. Additionally, you can also opt for weekly or monthly backups. The next option allows you to encrypt your backups and to choose the compression rate (zero, standard, high). An "advanced" button gives access to several settings as a slider to adjust the performance of the software and the use of resources (the default being maximum). This is also where you can enable simultaneous backups, useful when they involve multiple partitions or multiple disks.





AUTOMATIC PURGE FUNCTIONS

To prevent all backups to eventually occupy too much disk space, the advanced settings allow the application to retain a specific number of backups, the former being deleted on a FIFO (first in, first out) basis. This allows to keep only the last two full weekly backups (and all associated incremental backup sets), for example. The wizard summarizes all the selected settings and options prior to launch the selected tasks in the background.

WHAT TO DO WITH BACKUPS

You can do two things with your saved backup images: restore them as usual backup sets, but also explore the data inside them. To explore a saved disk image, just run the ad hoc assistant from the home window and select the image you want to explore. If there are incremental backups associated with a full backup, they will also appear. Which provides access to a file version to a specific date and time. ShadowProtect distinguishes itself by the clarity of the operation by displaying at this stage and graphically this partition – which is essential in the presence of backup disks with multiple partitions.

EXPLORE TO RUN

Once the desired partition chosen, you can explore the content as if it were a real hard drive. A "write" option allows programs to run from the image, but it also modifies the files contained in the image. However, the actual contents do not actually change, this setting actually creates a delta file that records changes in the files and applies the delta file to the image when you open it again, so that you have access to the changed version of the files including the changes made to them. When you explore the image again, an option allows you to ignore the delta file, so your original backup remains unchanged.

EXPLORE TO RESTORE

If your drive has multiple partitions, you can restore an image to another partition. You can also restore it to your existing Windows partition should any problem arise (faulty start, corrupt system, virus ...). To do this, you will need a recovery CD or USB key, but you will need to download the necessary files from the StorageCraft website. Two versions exist, the first being Windows PE (which requires the download of the large Microsoft Windows administration and deployment kit and, around 2 GB). The simplest option is to download the "cross-platform" version of the Linux recovery utility and burn it to a CD using the instructions provided. This tool allows you to start over and back up, restore, or explore backup images. Fortunately the documentation, abundant, is entirely available online, as well as access to a knowledge base and a user forum. Note the recent availability of a version for Linux certified on Ubuntu 12.04, Red Hat and CentOS 6 6.

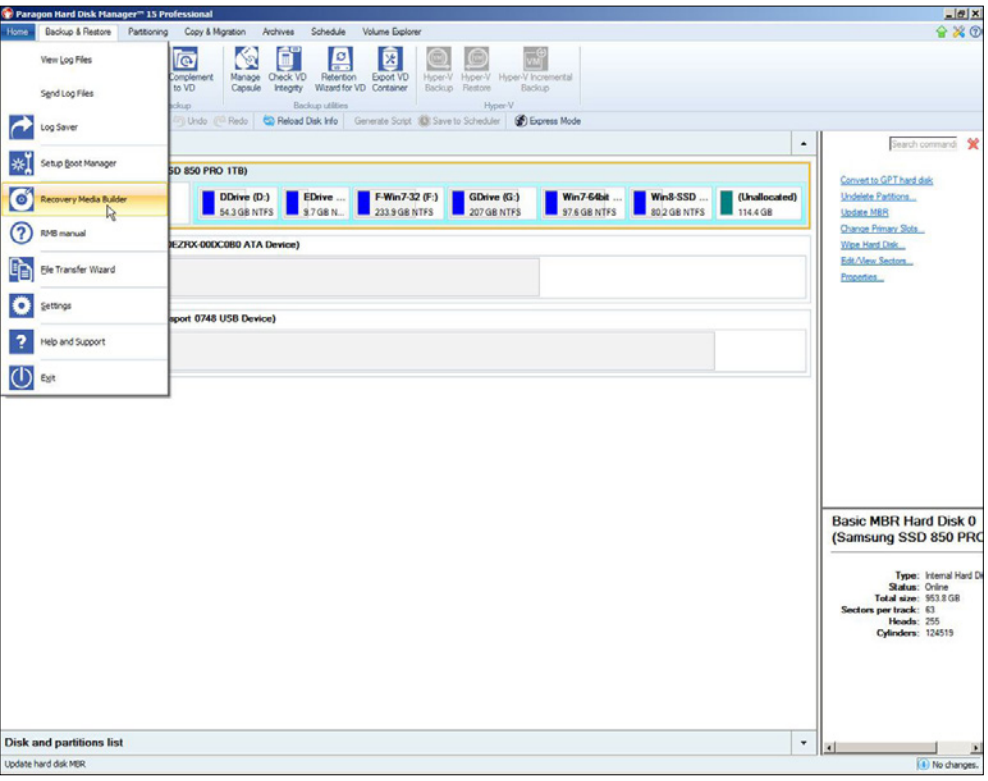
CONCLUSION

ShadowProtect does not surprise initially by its look or its list of features, but what it does it does very well. Rich options for backup and providing multiple modes of restoration, ShadowProtect Desktop 5 proved efficient, reliable and flexible. Making it probably the best in its sector. **JOSCELYN FLORES**

Paragon Hard Disk Manager 15 Business

HPC REVIEW AWARD

Year after year, Paragon Hard Disk Manager obtains distinctions in the disk management category, extending its functional scope over the years.



Paragon is a historical actor of disk management software, with this fifteenth version. In this regard, and in light of its field of intervention consisting of “live” date, the main criterion to consider is reliability. Suffice to say that we are not attracted to exclusive features that can leave you with a unusable disk. Instead, we reckon it is better to turn to software that supports standard industry characteristics in a fast and convenient way. Paragon Hard Disk Manager 15 Business turns out to be the most complete application to manage hard disks, and, apart from one or two minor drawbacks, the software does the work with a minimum of hassle.

PHILOSOPHY AND CHARACTERISTICS

The key function of the application is disk partitioning. Create, resize, delete and copy are supported. Ahead of its competitors, it integrates advanced backup and storage functions such as virtual disks supported by Windows Server and Hyper-V virtual machines. The complex and potentially destructive operations like copying a drive to another are made through well designed wizards to anticipate your likely choice, but leave you the flexibility to specify the options. The bootable Recovery Media Builder media creation feature lets you create a USB drive or a bootable CD based on Linux or Windows PE to start over and begin a restore in case your hard disk is unbootable. From the start, the software directs you to achieve your first backup, and it is prudent to follow the advice.

EXPLORE PHYSICAL AND VIRTUAL VOLUMES? EASY!

The Volume Browser function allows you to extract files from physical disks or virtual disk images. A uniquely designed wizard copies an existing physical system as a virtual “guest” system for use with either Hyper-V (2008 or 2012), Microsoft Virtual PC, or even Oracle VirtualBox, VMware Workstation or Fusion. Similar functions are available via Microsoft Hyper-V, but the Paragon wizard is easier to

use and has a wider range of options. By default, the application saves Paragon backup images to the native native Paragon virtual disk format, but you have the ability to record to Hyper-V, VirtualBox or VMware formats. One option in this version lets you back up only the media files, emails, or other documents to the virtual disk format to facilitate read-only access to the files when you need them. The software also has unique features such as an optional boot manager that supports up to sixteen different operating systems, reserved for special situations involving non-standard operating systems. Windows users will be content with the Microsoft launcher and Linux users will prefer GRUB 2 (which can also manage mixed Windows-Linux systems), but the alternative proposed by Paragon is just as reliable and efficient, and open to alternative operating systems.

A SOLID CONTENDER

For our tests, we have reproduced a typical use case and used a workstation containing two SSDs. The objective is to copy the contents of the first over the second largest capacity SSD (480 against 256 GB). To perform this operation we started on the bootable USB key generated by the software after clicking a few buttons, and about 45 minutes later the operation was completed. We then exchanged the SSDs and could start normally. For our second test, we used a notebook PC and made a copy of the internal drive on a connected USB SSD. Again the operation went without a hitch. Note that on both copy operations, the only notable problem was a «Estimated time» message that kept changing wildly, first predicting that the operation would take five minutes, then gradually increase the time to 50 minutes, then suddenly dropping to 25 minutes, and finally ending in twenty minutes. Unfortunately, this is quite usual, and no competitor does better in this regard. Regarding the cloning of boot disks, better let the software do its job with enough time ahead - and not do it in the morning before going to work!

The third test was to use the application to resize partitions on a workstation. As with other advanced partitioning software, the application is able to do most partition creation and modification operations without rebooting. However more complex partition mergers, demergers or moves only become effective after a restart, the operation being accomplished before starting Windows. With the same inaccuracy in the time required to carry these operations out, but the key is that these operations are conducted smoothly. Unusual opportunity, a copy of a hard drive or partition can be planned in advance.

NEW FEATURES

This new version adopts a very «Windows 8»-like interface with toolbars and icons «flat design». Convenient if you ever wish to use the software in touch mode, but the mouse is equally suitable. Among the new features a virtual disk Export Wizard that converts virtual disks from one format to another, for example, from the Microsoft Virtual PC format (now obsolete) to VMware or Hyper-V, or anywhere from any combination of source and target. The full or incremental backup functions can be a great time saver for administrators in charge of a fleet of machines. Another innovation incorporates an algorithm designed specifically for SSD. Another welcome note is the existence of an operating system installation wizard, which allows an administrator to prepare a bootable disk while remaining on the original PC.

In conclusion, Hard Disk Manager 15 Business is a top choice that goes well beyond simple partitioning features it already offers, and it gains interest with its P2V and V2P conversion functions and backup physical volume virtual. Paragon has done an excellent job. **JOSCELYN FLORES**

HP 3PAR StoreServ 20800

At the HP Discover event in Las Vegas, the company announced its 3PAR StoreServ 20000 Storage series offering 25 percent lower cost-per-GB, along with a new class of massively scalable flash arrays and flash optimized data services. Enough to lower HP's $18 cost-per-GB from two and a half years ago to $1.50 per useable GB today.

Although the 3PAR StoreServ 20800 is a «flash-first design», hp also offers to its customers an hybrid alternative with the 20850 model which uses an adjustable mix of flash storage and conventional 15 krpm spinning disks.

The 20850 model can be deployed with as few as two controllers, but it easily scales out to eight controllers with up to 3.6 TB of cache and 1,024 SSDs. This provides between 1.92 TB and 3,932 TB of raw capacity (12 PB total useable) and a rack density of 5.5 PB. HP offers a range of individual SSDs from 480 GB up to 3.84 TB to propel the top speed up to 3.2 million IOPS and 75 GBps of throughput. The addition of deduplication services expands the useable capacity of the storage pool, and thus lowers the cost per addressable GB. The 20800 can be deployed with as few as two controllers, and can scale up to eight controllers. The architecture offers from 1.8 TB of on-node cache up to 33.8 TB. The system supports up to 1,920 total drives with any mixture of SSDs and HDDs.

MIXED SYSTEM FOR TIERING APPLICATION

This mixed SSD and HDD system is useful for tiering applications, such as storing snapshots and bulk data on the HDD tier and keeping high-performance workloads on the flash tier. A small 20800 configuration begins at 1.92 TB of raw capacity and scales up to an impressive 6 petabytes of raw storage. The system supports the same range of SSDs as the 20850, but it also adds a slew of HDD options that include 600 GB 15K, 1.2 TB 10K and the capacity-heavy 6 TB 7,200 RPM HDDs.

HP also added several new features to address flash-specific challenges. Asynchronous replication supports the demanding RPO (Recovery Point Objectives) brought on by the speed of flash systems by lowering them to one-second intervals.

HIGH-END FEATURES AND LOWER COSTS TO GRAB MARKETSHARE

The StoreServ 20800 features new end-to-end high-availability features, including the Persistent Checksum product for HP ProLiant Servers. This system leverages the new T10 PI (Protection Information) standard to run checksums on all data from the host down to the end storage.

IDC forecasted a 46 percent compound annual growth rate for all-flash arrays over the next five years, and HP is looking to grab a portion of the lucrative market with pricing under $100,000 for the 20800 and 20850 platforms. 15K HDDs have already taken a bludgeoning at the hands of flash, and with flash-based systems coming in at $1.50 per GB, it appears that 10K HDDs are now under duress as well. With new all-flash arrays that provide lower overall TCO and competitive CAPEX with disk-based systems, the trend is sure to accelerate. **PAUL ALCORN**

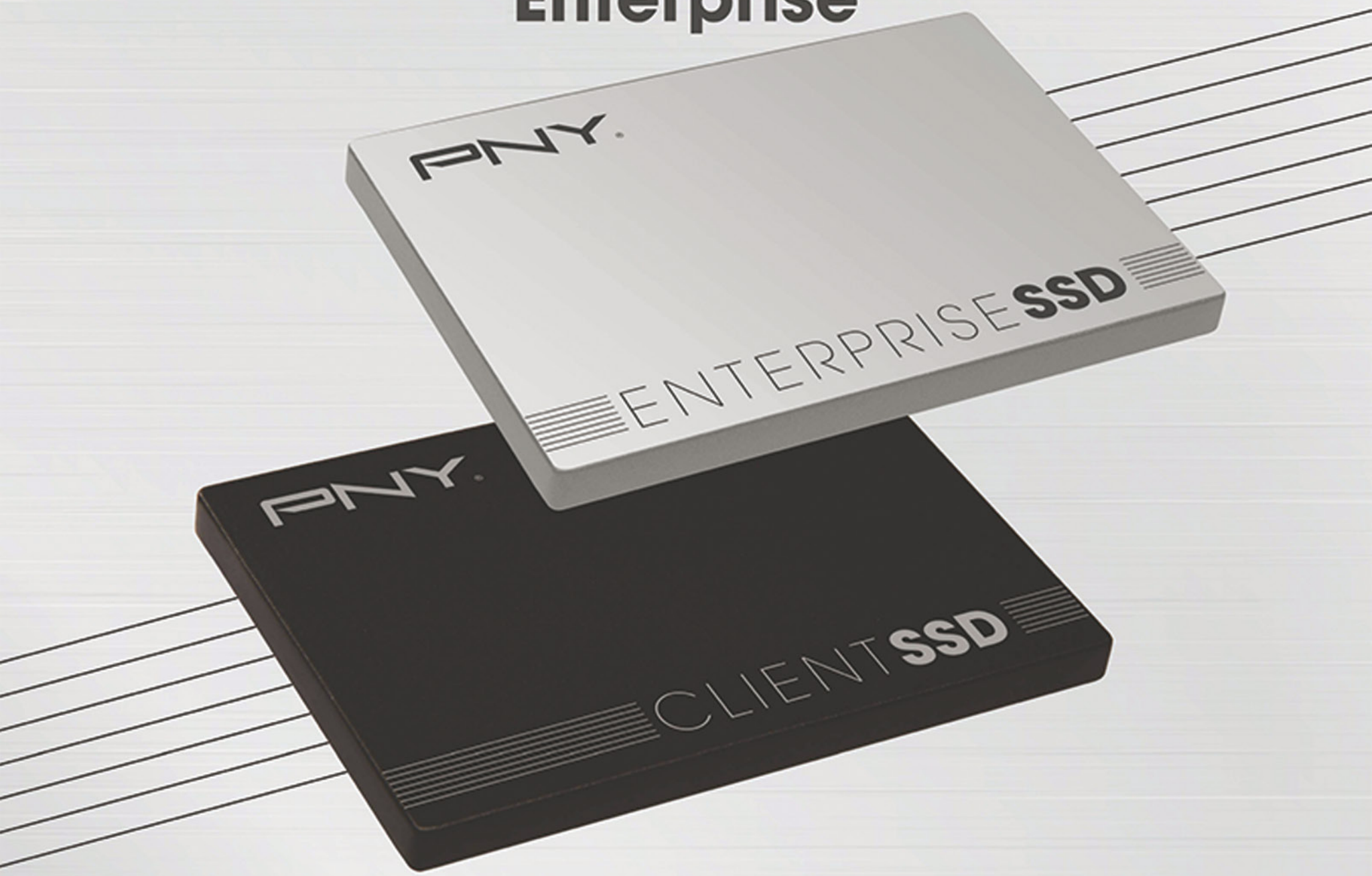
PNY

Professional Solutions

NVIDIA Quadro® / NVIDIA Tesla® / NVIDIA Grid®/ SSDs

Solid State Drives

EP7000 Series
Enterprise



CL4000 Series
Client

WWW.PNY.EU

PNYPROFR@PNY.EU

OpenMP Device Constructs

Many-core devices, with their many small, power-efficient processing units, provide massive threading and SIMD processing. The value of this unprecedented hardware parallelism is widely acknowledged by industry, but has been slowly adopted, partly due to the lack of an open standard. Nevertheless, some device-specific and general APIs, such as CUDA, OpenACC, and OpenCL provide solutions to users for porting their codes to devices.

OpenMP is a well-known open standard for shared memory multiprocessing. Recently, the OpenMP language committee has extended the standard to include support for heterogeneous, non-shared-memory computing. OpenMP extensions now provide the ability to run code on both the host and a device in a «work sharing» manner within a single program. The execution model starts on a host processor. Sections of code encapsulated by OpenMP target directives are launched for execution on a device, while optionally allowing the host to execute in parallel with the device. The host controls all the allocation of device memory, transfer of data, queuing target executions on a queue, and managing their completion.

Significantly, OpenMP now provides a single, parallel model for threading, worksharing, device targeting, teams, and SIMD execution. A single paradigm provides a portable platform for development and a highly composable platform for integrating heterogeneous executions within a single program.

PROGRAMMING DEVICES WITH OPENMP

1 CODE EXECUTION ON A TARGET DEVICE

In OpenMP there is a host device and a set of target devices. Program execution begins on the host device. A thread encountering a target directive on the host device will execute subsequent code statements in the target region on a target device. Variables accessed in the target region are mapped to the device and the target region is executed by the target device. By default, the host thread that encountered the target region waits for the target device to complete the execution of the target region.

```
void vec_mult(int N)
{
  int i;
  float p[N], v1[N], v2[N];

  init(v1, v2, N);
  #pragma omp target
  for (i=0; i<N; i++)
    p[i] = v1[i] * v2[i];
  output(p, N);
}
```

Figure 1: *Execute code on a target device*

In Figure 1, the for-loop following the target directive is executed on a target device. The host thread waits for the completion of the target region and then continues with the execution of the function output. The variables p, v1,

v2, I, and N are mapped to the target device at the beginning of the region and then mapped from the device at the end of the region.

If a function is called from a target region, a declaration of that function must appear between a declare target / end declare target directive pair. This tells the compiler to generate a device version of the function along with the host version.

```
#pragma omp declare target
extern void fib(int N);
#pragma omp end declare target

#define THRESHOLD 1000000
void fib_wrapper(int n)
{
#pragma omp target if(n > THRESHOLD)
{
fib(n);
}
```

Figure 2: *Call a function from a target region*

In Figure 2, the function fib() is called from a target region and the declaration of fib() appears between a declare target / end declare target directive pair.

2 THE DEVICE DATA ENVIRONMENT

Each device has a device data environment containing the variables currently mapped to the device. In the mapping process, each variable referenced in a target construct is allocated a corresponding variable in the device data environment. By default, the corresponding variable is initialized with the value of the original variable on entry to the target region, and the original variable is assigned the value of the corresponding variable on exit from the region. (The original variable is the variable that the host sees, outside of any target device data environment.)

The mapped variable model supports both shared and distributed memory systems between host and target devices. Depending on the underlying hardware memory system, a mapped variable might require copies between host and target device memories, or no copies if the host and target device share memory. Even if memory is shared, a pointer translation of a memory coherence operation might still be required when mapping a variable. When an original variable in a host data environment is mapped to a corresponding variable in a device data environment, the mapped variable model asserts that the original and corresponding variables may share storage. Writes to the corresponding variable may alter the value of the original variable. Therefore, a program cannot assume that mapping a variable results in a copy of that variable.

MAPPING VARIABLES TO A DEVICE

The map clause specifies how original variables are mapped to their corresponding variables in a device data environment. The map clause has a map-type that can be used to optimize the mapping of variables. The to map-type indicates that on entry to the region the corresponding variable is initialized with the value of the original variable. The from map-type indicates that on exit from the region the original variable is assigned the value of the corresponding variable. The tofrom map-type is the default and combines the behaviors of the to and from map-types. The alloc map-type neither initializes the corresponding variable on entry to the region nor assigns the original variable on exit from the region.

```
extern void init(float*, float*, int);
extern void output(float*, int);
void vec_mult(int N)
{
  int i;
  float p[N], v1[N], v2[N];

  init(v1, v2, N);
  #pragma omp target map(to: v1, v2)
map(from: p)
  #pragma omp parallel for
```

```
  for (i=0; i<N; i++)
    p[i] = v1[i] * v2[i];
  output(p, N);
}
```

Figure 3: *Using map-types and map clauses*

In the map clause of Figure 3, the to map-type indicates that on entry to the target region the corresponding variables in the device data environment v1 and v2 are initialized with the values of the original variables in the host's data environment. On exit from the target region, the values are not copied and the device storage is removed. The from map-type for the variable p specifies that storage is created on the device without the values of the original variable, and on exit the corresponding (device) values are copied to the original variable storage before freeing the storage.

ARRAY SECTIONS

An array section designates a subset of the elements of an array. Fortran has built-in support for array sections, but C and C++ do not. Some OpenMP clauses such as the map clause accept array section syntax that can occur in place of an array subscript. The syntax in C and C++ is variable[<lower-bound> : <length>], where variable has a type of array or pointer, and lower-bound and length are integral expressions that specify a contiguous set of elements. Array sections are useful for describing the memory behind a pointer in C/C++ or for mapping a slice or sub-section of a large array. Note that in Fortran the lower-bound and upper-bounds are specified as the upper and lower extent of the array section. For C/C++, the lower-bound is the start of the subsection and the length is the size of the array section. This difference from Fortran was chosen because C/C++ deals with lengths of arrays and not upper bounds.

The notation [:] is a shorthand for a whole array dimension if the size of the dimension is known from the array declaration. Both <lower bound> and <length> must be specified if either is needed.

The compiler must be able to determine the shape and size of an array object. If the base of an array section has incompletely specified dimensions (such as a pointer variable), the length of the array section must be specified explicitly.

PERSISTENT DATA REGION

As with any storage, performance may be optimized by minimizing data transfers. Also, since creating space on a remote device can be expensive, reusing the same allocated space across target invocations (making the storage persistent) can enhance performance. The structured block of a target data directive in C/C++ and the code block terminated by the end target data directive in Fortran create a region in which the storage and the data within the storage persist across enclosed target directive executions. The map clause is used to designate data transfer to the device (map-type to) at the beginning of the region and transfer data back from the device (map-type from) at the end of the region. The tofrom map-type is a combination of to and from. The alloc map-type designates only storage creation without any transfer.

The program below illustrates persistence and data transfers at the boundaries of the target data region.

```
#define N 1000
int main(){
{
 int i, a[N], b[N], C[N], D[N];

 for(i=0; i<N; i++) { a[i]=i; b[N-i]=i; }

 #pragma omp target data map(to:    a) \
                         map(tofrom:b) \
                         map(from:  c) \
                         map(alloc: d) \
 {
   #pragma omp target
   for (i=0; i<N; i++) { d[i] = a[i] + b[i]; }
```

```
  #pragma omp target
  for (i=0; i<N; i++) { b[i]=d[i]*i; c[i]=d[i]/i; }
 }

 printf(«b[0] b[N-1] c[0] c[N-1] %d %d %d %d
\n», \
        b[0],b[N-1],c[0],c[N-1]);
}
```

***Figure 4:** Persistent data with the Target Data directive*

Note, there is no implicit mapping of the arrays on the target execution directives. The variable i is implicitly mapped, though. Target update directives are used within the target data region to transfer (assign) data in the original list items (host values) to the corresponding device storage with the to clause. Transfers in the reverse direction are specified with the from clause.

The next program illustrates "update" transfers, and uses array sections of dynamic data (arrays created by malloc) on the host. Here, v1, v2, and p array sections are created on the device, and v1 and v2 values on the host are copied to the device at the beginning of the target data region. After the first target execution, the host modifies its v1 and v2 arrays, and then copies the values to the device with the target update directive. The new values are used in the second target execution directive. At the end of the target data region, the p array section is copied from the device to the host.

```
extern void init(float *, float *, int);
extern void init_again(float *, float *, int);
extern void output(float *, int);
void vec_mult(float *p, float *v1, float *v2, int N)
{
  int i;
  init(v1, v2, N);

#pragma omp target data map(to: v1[:N], v2[:N])
map(from: p[0:N])
{
#pragma omp target
#pragma omp parallel for
for (i=0; i<N; i++)
    p[i] = v1[i] * v2[i];

init_again(v1, v2, N);
#pragma omp target update to(v1[:N], v2[:N])

#pragma omp target
#pragma omp parallel for
for (i=0; i<N; i++)
   p[i] = p[i] + (v1[i] * v2[i]);
}
  output(p, N);
}
```

***Figure 5:** Updating persistent storage with the Target Data directive.*

3 ACCELERATED WORKSHARING

The teams construct creates a league of thread teams in which the master thread of each team begins execution of the region. Each master thread is an initial thread, and executes sequentially, as if enclosed in an implicit task region defined by an implicit parallel region surrounding the entire teams region.

```
int main (int argc, char ** argv)
{
  int nteams = 4;
#pragma omp target
  {
#pragma omp teams num_teams(nteams)
    {
      int teamSize = omp_get_num_threads();
#pragma omp distribute
      for (int t=0; t<nteams; t++) {
#pragma omp parallel for
        for (int i=0; i<teamSize; i++) {
          int nTeams = omp_get_num_
teams();
          int myTeam = omp_get_team_
num();
          int me     = omp_get_thread_num();
          int teamSize = omp_get_num_
```

```
threads();
printf («Iteration: %d -- thread %d of %d in
team %d of %d\n», i, me, teamSize, myTeam,
nTeams);
   // system with 240 threads would print the
following for iteration 0
  // Iteration: 0 -- thread 0 of 60 in team 0 of
4
  // Iteration: 0 -- thread 0 of 60 in team 1 of 4
  // Iteration: 0 -- thread 0 of 60 in team 2 of
4
  // Iteration: 0 -- thread 0 of 60 in team 3 of
4
}}}}}
```

Figure 6: *Teams and Distribute constructs*

4 ASYNCHRONOUS EXECUTION

As usual, the nowait clause on a target directive means that the encountering thread does not wait at some form of an implicit barrier or wait. A thread that encounters a target region without a nowait clause will launch a target execution and wait for the target execution (and the data transfers) to complete before continuing execution beyond the target construct. That is, the encountering thread blocks until the target execution is done. This guarantees that the data mapped from the device will be in place for execution after the target construct.

A target nowait clause enables a target region to be launched in the background and allows a thread to immediately continue execution after the target region. The thread is free to execute serially, or to create or simultaneously participate in parallel regions on the host. Essentially, the target execution can proceed asynchronously as a host thread executes code after the target construct. One might think of the asynchronous behavior as similar to that of asynchronous I/O.

```
#pragma omp target                    #pragma
omp target  nowait
device_work();                  device work();
// host thread blocks here          // host
thread continues execution here
// until device_work completes   // while
device_work executes
```

Figure 7: *Blocking (left) and Asynchronous(right) Target Execution*

Technically, the details of how a nowait clause accomplishes asynchronous execution of a host thread and a target execution is based on the execution model for tasks. The target construct executes as though it is enclosed by a task; this generated task is a target task. Without a target nowait clause, a target task is executed immediately by the encountering thread (i.e., the task in undeferred) and waits at a scheduling point for device execution and data transfers to complete.

When a nowait clause is present, the target task is queued for execution and the task of the encountering thread may resume execution after the target construct before the target task completes execution. The underlying mechanisms for data transfers and launch of the device executable are implementation defined. That is, OS threads may do this work. Hence, in a serial region where only a single OpenMP thread is available, the mechanics of running, monitoring and terminating a target execution on the device (launch and cleanup), and transferring data may proceed without interrupting the single OpenMP host thread until it is time to resume the target task waiting at the scheduling point.

Completion of the target task, and hence the target execution, is guaranteed by a taskwait directive. In the code snippet below, the master (serial) thread waits for completion of its child task (the target task), guaranteeing completion of the target execution and data transfers.

```
int main(){
  #pragma omp declare void device_work(int
ia[])
  int ia[2];
```

```
  ia[0]=1; ia[1];

  #pragma omp target nowait
  device_work(ia);              // ia changed on
device

  printf(«ia[0]=%d\n», ia[0]); // race condition

  #pragma omp taskwait

  printf(«ia[0]=%d\n», ia[0]); // device values
}
```

Figure 8: *Asynchronous Target Execution*

ADVANCED FEATURES

1 DEPEND OFFLOADING

The depend offloading functionality provides the ability to offload a task onto the device and makes the thread immediately available to participate in worksharing. Asynchronous offloading is integrated with the OpenMP tasking model, providing the ability to order tasks on the host and offload tasks.

```
#pragma omp target depend(out: b) map(a)
 { task1(a); }
#pragma omp target update depend(out: a)
map(a)
 { task2(a); }
#pragma omp task depend(in: b)
 task3_on_host();
#pragma omp target depend(inout: a, b)
map(a,b)
{
 #pragma omp task
    task4(a);
 #pragma omp task depend(out: b)
    task5();
 #pragma omp task depend(in: b)
    task6();
}
```

Figure 9: *Target Depend Clauses*

2 UNSTRUCTURED DATA DIRECTIVES (TRANSFERS AND PERSISTENT DATA)

The unstructured data directives provide a rich set of data allocation and data movement. These directives enable allocation of device memory in one routine, deallocation in a different routine, and conditional data motions.

```
class myClass {
 myClass(){
    commonData = malloc()
   #pragma omp target enter data
map(alloc(commonData))
 }
 ~myClass(){
  #pragma omp target exit data
map(release(instanceData))
 }
transfer_to() {
   #pragma omp target exit data
map(to(instanceData))
 }
 transfer_from() {
   #pragma omp target exit data
map(from(instanceData))
 }

 private:
 float *commonData;
 int length;
}
```

Figure 10: *Unstructured Data Directives*

PROGRAMMING GUIDELINES

Programs that are good candidates for offloading have these characteristics:
1. There is a high level of parallelism either in threads or vectors
2. The data transfer should be minimal.
3. The code should not execute significant amounts of I/O

In a heterogeneous computing environment, an application can be tuned to execute particular components of work on specific devices. For example, serial components might exe-

To extend the lifetime of the variable beyond this lexical scope so variables can be allocated in one routine and freed in a different routine, two new standalone target constructs have been introduced: “target enter data” and “target exit data”.

cute best on a general purpose, high frequency CPU whose architecture is suited for logic and branched code. Code with highly parallel and vectorizable components might share the work across the host and the device(s). Targeted execution on a device may incur significant overhead for data initialization and transfer. In this case it is beneficial to overlap data communication with computation or reuse data storage across many target executions. OpenMP 4.0 provides the ability to reuse data across multiple offloads and a future extension will enable asynchronous (overlapping) data transfers.

FUTURE OF DEVICE CONSTRUCTS

The OpenMP committee continues to enrich the language with new features that enable efficient use of the target devices. The committee has released Technical Report 3 (TR3) in anticipation of the OpenMP 4.1 release. The key new features in the TR3 are:

NON-STRUCTURED DATA ALLOCATION

In OpenMP 4.0, variables are mapped to a target device for the duration of the lexical scope where the construct is used. To extend the lifetime of the variable beyond this lexical scope so variables can be allocated in one routine and freed in a different routine, two new standalone target constructs have been introduced: “target enter data” and “target exit data”. “target enter” begins the lifetime of the variable in the target device and “target exit data” ends the lifetime of the variable.

ASYNCHRONOUS OFFLOAD

In OpenMP 4.0, all target tasks are synchronous. The thread which encounters the target task starts the execution of the target task on the target device and waits for the target task to complete. A new clause “wait” has been added to the target construct to enable the encountering thread to resume execution before the target task completes its execution.

DEPEND CLAUSE ADDITION

A new “depend” clause has been added to the device constructs that enables synchronization between device constructs, and also between device constructs and task constructs. The behavior of this “depend” clause is the same as the “depend” clause in the task construct and will be described in the article on tasking.

MAP CLAUSE EXTENSIONS

A new map type “delete” has been added to the existing map-type to enable the user to free the specified object from the device environment unconditionally. This can be used to release the object in an inner nesting of a target region.Map type modifier “always” has been added to the map-type to modify the default behavior of the map-type. The “always” modifier will enable the user to force a transfer where the transfer may not have occurred due to the presence rule. In OpenMP 4.0, this would have required the user to use an “update” construct to transfer the data.

APPENDIX ON OPENMP TARGET SYNTAX

C/C++	Fortran
#pragma omp target [clause[[,] clause],...] new-line structured-block *where clause is one of the following:* **device(integer-expression) map([map-type :] list) if(scalar-expression)**	**!$omp target [clause[[,] clause],...] structured-block !$omp end target**

C/C++	Fortran
#pragma omp target data [clause[[,] clause],...] new-line structured-block *where clause is one of the following:* **device(integer-expression) map([map-type :] list) if(scalar-expression)**	**!$omp target data [clause[[,] clause],...] structured-block !$omp end target data**

C/C++	Fortran
#pragma omp target update [clause[[,] clause],...] new-line *where clause is one of the following:* **to(list) from([list) device(integer-expression) if(scalar-expression)**	**!$omp target update [clause[[,] clause],...]**

C/C++	Fortran
#pragma omp declare target new-line Declarations-definition-seq #pragma omp end declare target new-line	*For variables, functions and subroutines* **!$omp declare target data([list)** *For functions and subroutines* **!$omp declare target**

C/C++	Fortran
#pragma omp teams [clause[[,] clause],...] new-line structured-block *where clause is one of the following:* **num_teams(integer-expression) thread_limit(integer-expression) default(shared | none) private(list) firstprivate(list) shared(list)**	**!$omp teams [clause[[,] clause],...] structured-block !$omp end teams**

C/C++	Fortran
#pragma omp distribute [clause[[,] clause],...] new-line for-loops *where clause is one of the following:* **private(list) firstprivate(list) collapse(n) dist_schedule(kind[, chunk_size])**	**!$omp distribute [clause[[,] clause],...] do-loops [!$omp end distribute]**

C/C++	Fortran
#pragma omp distribute simd [clause[[,] clause],...] new-line for-loops	**!$omp distribute simd [clause[[,] clause],...] do-loops [!$omp end distribute]**

C/C++	Fortran
#pragma omp distribute parallel for [clause[[,] clause],...] new-line for-loops	**!$omp distribute parallel do [clause[[,] clause],...] do-loops [!$omp end distribute parallel do]**

C/C++	Fortran
#pragma omp distribute parallel for simd [clause[[,] clause],...] new-line for-loops	**!$omp distribute parallel do simd [clause[[,] clause],...] do-loops [!$omp end distribute parallel do simd]**

Bringing Lustre Relevance to the Enterprise

High Performance Computing (HPC) technologies are coming to the enterprise to help with Big Data as well as the emerging HPC in the enterprise workloads. Common to these workloads are exponentially expanding data requirements. So, enterprises are facing one of their biggest challenges—efficiently serving up those massive amounts of data to the compute complex. Company IT departments are experiencing bottlenecks in their storage I/O, network I/O, or both with their existing NFS and local file systems. The file systems need to be more performant to keep up with the workload demands, but scaling NFS systems is difficult.

There is a lot of complexity around managing a bunch of NFS servers, breaking up the namespace across servers, which potentially results in isolated storage domains. "One company in the Financial Service Industry," said Brent Gorda , General Manager of Intel's High Performance Data Division, the group that develops Intel® Lustre* software, "has a hundred NFS servers, and users have to first remember on which server their data resides before they can even start processing it." Some organizations may attempt to mask this complexity with a complicated tangle of automount maps; this normalizes the namespace to a certain degree, but does not solve the problem of unequal distribution of data and IO workloads across the storage domain. The logical solution is a clustered approach with a parallel file system, and Lustre is the leading parallel file system in performance and scalability. So enterprises are looking at Lustre. But it has had a reputation of not being very enterprise-friendly.

ABOUT THE AUTHOR

Ken Strandberg *is a technical story teller. He writes articles, white papers, seminars, web-based training, video and animation scripts, and technical marketing and interactive collateral for emerging technology companies, Fortune 100 enterprises, and multi-national corporations. Mr. Strandberg's technology areas include Software, HPC, Industrial Technologies, Design Automation, Networking, Medical Technologies, Semiconductor, and Telecom.*

EQUIPPING LUSTRE FOR THE ENTERPRISE

"Lustre is an open source project, but Intel is a driving force behind it, having led every community release since 2010, starting with Whamcloud," stated Bret Costelow, Director of Global Sales for Lustre Solutions at Intel. "Each year, Intel developers continue to spend significant work investing in specific features that make Lustre more relevant and appealing in enterprise workloads."

"Big Data and HPC storage in the enterprise is not the same as in academia and the big labs"

"Big Data and HPC storage in the enterprise is not the same as in academia and the big labs," according to Malcolm Cowe, product manager for Intel Enterprise Edition for Lustre software. "In the labs, there might be six or more full-time engineers managing their Lustre file system. Enterprises can't afford that level of dedicated resources." In the enterprise, file systems have to be easily managed, reliable, and available with minimal attention. And, enterprise IT leans heavily on automation across their infrastructure. "Storage systems need to run without interruption and intervention as much as possible." To address enterprise requirements, Intel has done some interesting things to and with Lustre.

There is more to standing up an enterprise-class 50 petabyte, 2 terabyte/second Lustre file system with 50,000 drives than just acquiring the servers and installing the open source software. For mission critical workloads, reliability, availability, and even disaster recovery need to be built into the solution, so there is no downtime. For example, Australia's Bureau of Meteorology relies on data being always accessible for numerical weather prediction to assist air traffic, fire suppression planning, and weather emergencies, where lives can depend on its availability 24/7.

While Intel and the Lustre community have created highly stable software, enterprise-class reliability is largely based on investment in system hardware. Intel Lustre solution architects work with a large partner ecosystem to develop high availability building block modules based on established design patterns, so there's no single point of failure in the server infrastructure. The partners then deliver the solution to their customers. "Investing in and working with a partner ecosystem and providing a credible support infrastructure through that partnership is a critical part of our reliability and serviceability model," remarked Costelow.

POLISHING LUSTRE'S UNDESERVED REPUTATION

"Lustre has an undeserved reputation for being difficult," stated Cowe. "It comes from an open source history, where the focus is on those writing the code rather than those consuming it. So, we have made and continue to make Lustre as accessible as possible with features like Intel Manager for Lustre (IML)." IML is a web-based interface that makes it straightforward to install and easy to manage the system. The software takes IT industry best practices based on established hardware system design patterns and automatically deploys the file system for highly reliable and available operation.

"Lustre was originally architected to quickly serve massively large data sets—in the gigabyte to terabyte range—in a single file," explained Gorda. "By working in parallel, the object storage servers feed up that data incredibly fast and efficiently. With those types of workloads, Lustre did not need high performance meta data servers." In Life Sciences and the Financial Services Industry (FSI), it is a different situation, one for which Lustre was not originally designed.

In Life Sciences and FSI, the large data sets can be made up of a massive number of smaller files. For example, in genomics, sequencers can create millions of files of a few hundred megabytes that are joined together into a 6 TB data set describing a specie's genome. The file system must find these millions of files through meta data accesses in order to create the entire data set. Until recently, Lustre's upper limit of about 60k file creates per second made it less efficient for these types of workloads. With the latest release of Lustre

with the added feature of parallel metadata servers, the file system has scaled out to over 1 million file creates per second, giving Life Sciences and FSI a significant performance benefit. According to Malcolm Cowe, the Intel team and the open source Lustre community continue to innovate methods for even greater scalability for these types of workloads. "We're working on different striping methods, distributed transactions, and a project called Data on Meta data to enhance even further small file performance on Lustre," remarked Cowe.

ENABLING LUSTRE FOR BIG DATA AND THE CLOUD

The performance Lustre delivers is ideal for Big Data workloads and the Cloud.

"Everybody associates Big Data with Hadoop," claimed Gorda. "But there's more to Big Data. We support Hadoop, but we go well beyond it." This is good news for HPC users with a Lustre file system who want to try Hadoop without scaling out an entire Hadoop platform with replicated local storage. Intel created an Hadoop interface to HPC job schedulers, like Slurm, so the Hadoop job looks like an HPC job. And they have written a file system interface to Hadoop that takes out the Hadoop Distributed File System (HDFS) and puts Lustre in, effectively removing the need for local storage and opening the door to running Map/Reduce with Lustre. PayPal uses Lustre and Big Data for real-time fraud detection. "Intel's work with Lustre on Big Data is a huge enabling for our HPC customers who want to run Hadoop workloads on their data," said Gorda.

HPC is emerging in the Cloud, so that companies who temporarily need massively parallel computing capacity can take advantage of it. Intel has enabled these companies to leverage Lustre in the cloud with their release of Intel Cloud Edition for Lustre software. Amazon Web Services (AWS) uses the Intel cloud version to offer high-performance, scalable storage using Lustre in their Elastic Compute Cloud (EC2). AWS is able to deploy a production Lustre file system in ten to 15 minutes, according to their web site. SAS, the Business Analytics software company, delivers clustered analytics services through the Amazon Web Services Marketplace and recommends using Lustre on AWS for their analytics services.

TEAMING WITH THE COMMUNITY TO GO BEYOND TRADITIONAL LUSTRE USAGES

With Intel's and the community's work, Lustre now also supports Hierarchical Storage Management for customers who need to balance their requirements for performance, scalability, and capacity. And they do a lot of work to integrate other protocols, such as SMB using Samba, and NFS, in order to mix Lustre with other networks.

There's a lot of interest by enterprises to integrate more security functionality into Lustre. Intel and the Lustre community are working on developing access controls with SELinux to provide fine-grained secure access to data by applications. "And we're also looking at Kerberos to do authentication and authorization of nodes, plus over the wire network encryption," indicated Cowe.

"The fact that the majority of sites we work with, and the majority of the community, have moved forward to a 2.5.x Lustre code base is a strong endorsement of the advances we've made over the past several years to add enterprise-class features and stability to the code," said Gorda. "And there's more to come." Intel is working on the next generation of storage technology to support Exascale computing. "With upcoming Intel technologies and the software we're working on, we'll be able to support not terabytes, but petabytes per second," stated Gorda.

"Today, we are seeing signs of encouragement in Lustre's ability to provide value for enterprise environments," said Gorda. "We believe this is the tip of the iceberg and definitely a sign of great things to come for the Lustre community."

Micheal Feindt, Blue Yonder's founder and Chief Scientific Advisor

Beyond deep learning and neural networks

Blue Yonder combines world class data science with professional business software, deep business knowledge in many verticals and a cloud service offering in order to serve companies in predicting the near future (including uncertainty and risk calculation) and optimize and automatize decisions.

In this exclusive interview, Prof. Dr Micheal Feindt, Blue Yonder's founder and Chief Scientific Advisor, tells us more about the company's predictive analysis technologies and how a new kind of decision-making automation could soon revolutionize the way science – as well as business – are made.

HPC Review: Tell us a little bit about how Blue Yonder was founded and what it provides.

Michael Feindt: Blue Yonder was founded in 2008 from my earlier company Phi-T and the OTTO group, world's second largest distance selling company after Amazon. In late 2014, the international growth private equity firm Warburg Pincus announced an investment of $ 75 mio to Blue Yonder, the largest real tech investment in Europe in 2014.

Blue Yonder combines world class data science with professional business software, deep business knowledge in many verticals and a cloud service offering in order to serve companies in predicting the near future (including uncertainty and risk calculation) and optimize and automatize decisions.

Usually, this has deep consequences and is really disruptive.

HPC Review: What was the rationale for offering Blue Yonder's predictive analytics as a cloud service rather than as an application to be run on in-house resources?

Feindt: Our experience shows that for many companies it is very difficult to attack large data science projects because they need expertise in many fields — their business experience, but also mathematics, statistics, machine learning, software engineering, data handling, as well as hardware and operations. We feel that it is way easier, faster and efficient if we cluster all this additional expertise in Blue Yonder so that our customers can concentrate on their core business and outsource all mathematical and technical complexity. Also, exceptional data scientists like to work in a larger group of excellent data scientists.

In contrast to the usual large on-premise-software projects the initial investment is much lower, the time to market is reduced drastically, and the failure rate is zero.

HPC Review: The service the company offers is based on the NeuroBayes algorithm, which you developed. Can you describe in layman's terms what the algorithm does?

Feindt: The core of NeuroBayes is a neural network of the second generation, with Bayesian regularization, that next to simple classification is able to predict individualized complete probability distributions for real valued quantities, which are the basis for optimal decisions. However, over time more and more robust preprocessing steps were introduced, and many more other efficient algorithms were developed at Blue Yonder. Generalizability, robustness, learning and prediction speed, and scalability are important design criteria for our algorithms. Recently, we also focused on interpretability (understanding)

The development is going into the opposite direction: we do not try to mimic the human brain, but to build efficient, robust and fast models. Bayesian statistics is a key ingredient. The «neural» character is not important in the recent development, but the possibility to predict conditional densities is.

and the reconstruction of causal effects from historic data. So, the Blue Yonder algorithm library is much more than the original NeuroBayes algorithm.

In all cases, we analyze large complex systems and automatically learn from past examples what observable quantities (of any kind) Review can say about another quantity in the near future, e.g. the number of Granny Smith apples in the XYZ Supermarket in Baker Street tomorrow. The prediction is given in form of a probability density, i.e. each possible future (number of apples sold tomorrow in that shop) is assigned a probability. Of course, this distribution should be as narrow as possible, but not narrower. The real future must be described correctly by this. Thus, statistical statements are individualized. On this basis mathematically optimal decisions can be taken, given we know the cost function of deviations of the future realization from our decision.

HPC Review: How is the technology different from other neural network schemes we've heard about – like the ones being used by search engine companies to classify images?
As already stated, the development is going into the opposite direction: we do not try to mimic the human brain, but to build efficient, robust and fast models. Bayesian statistics is a key ingredient. The "neural" character is not important in the recent development, but the possibility to predict conditional densities is.

Hierarchical models play a role, but we find it often more convenient (and way faster) to use our own neural networks for defining the hierarchy instead to let a deep network learn it.

HPC Review: What kinds of organizations are using the Blue Yonder offering? Can you talk about some specific problems that are being addressed?
NeuroBayes was originally developed for experimental elementary particle physics, and still is used very successfully at experiments at CERN (Geneva), Fermilab (USA) and KEK (Japan).

One of the latest developments was the completely automatic reconstruction chain for B factories - here automation was more than 2 times as efficient in reconstructing B mesons than 400 physicists in 10 years together. The other is the implementation of the NeuroBayes expert in hardware for the next generation B-factory experiment Belle II: Here, more than 10 billion decisions will be taken directly at the sensor array in order to find out which parts of the detector should be read out to the computers at all.

Large businesses in retail, travel and transport, and industry are the most important customers of Blue Yonder offerings. Blue Yonder performs demand predictions and even complete decision automation for the complete supply chain for many retailers and CPGs. Dynamic pricing in Internet and also brick and mortar shops is another hot topic with very large potential. In marketing optimization we invented a new algorithm in order to predict whether a customer will change his behavior by getting sent a catalogue.

The term «prescriptive analytics» stands for not only delivering the prediction, but also to optimizing the decision to be taken on this basis and give that as a recipe to the human expert. The fun stuff is the observation that cognitive biases also let the human decide wrongly in this case. Often gut feeling overrules the machine decision.

HPC Review: For businesses in general, how can the technology help improve their operation? In particular, how much can predictive analytics automate business processes?
One of the killer applications is the prediction of demand of each single article in each single store each day and the computation of optimal orders. Especially for perishable food this is of great value -

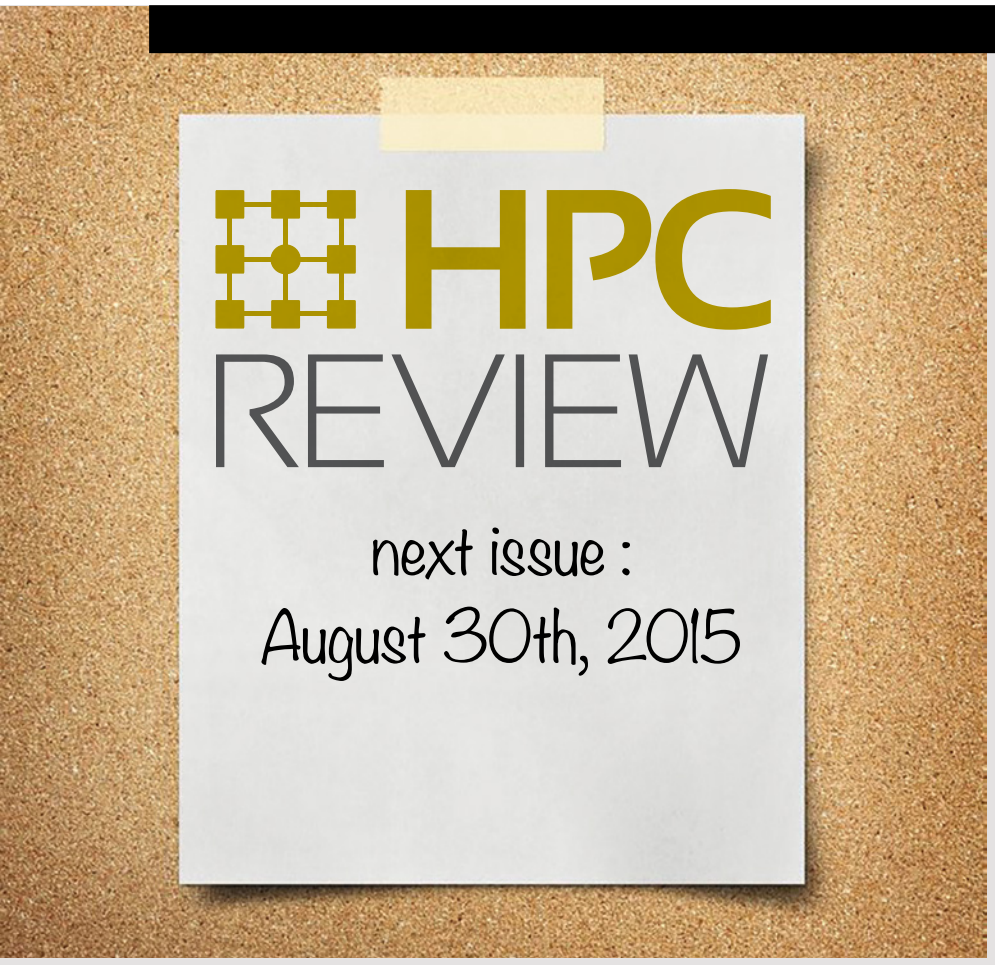
for one customer we could avoid food waste in the order of 25 mio € in one year, but simultaneously have less out-of-stock situations. The secret is individualization - to optimize not only one strategic goal, but to break it down to the thousands or millions of operational decisions each day. No human can take into account so many factors on so many articles each day in order to get it right.

The term "prescriptive analytics" stands for not only delivering the prediction, but also to optimizing the decision to be taken on this basis and give that as a recipe to the human expert. The fun stuff is the observation that cognitive biases also let the human decide wrongly in this case. Often gut feeling overrules the machine decision. The full potential, up to four times the effect of prescription, is only gathered if a full automation (99.9%) is realized, and only exception handling is left to the human expert.

There are many other examples like this, they all go into the same direction. I am very confident that more and more operational decisions, and thus also white collar work, will be automated.

HPC Review: Some business people will be apprehensive about relying on this level automation for decisions that were traditionally under the control of humans? What would you say to placate those fears?
That's right. Trust has to be built, and references obviously help. But already Lenin knew: To accept anything on trust, to preclude critical application and development, is a grievous sin. Thank god you can measure the improvement. In A/B tests you can prove that modern algorithms perform better than human decisions on many, even classically contradictory KPIs. The usual thing is to start the test with small groups (shops, articles), prove that it really works there, and then to roll it out fully in a few steps, with control at each stage. This way risk is minimized and trust is built.

But independent of all rational arguments often there is quite some resistance against any innovation and change — especially in hierarchical systems. It needs C-level sponsors and good concepts on how to communicate and organize the change. The danger in not going for automation or delay it too long is the sharp competition: the advantage is so large that it might be killing complete companies if only the competitors get more efficient.



LAB REVIEW

Nutanix hyperconverged appliance

Lenovo Thinkpad W550s mobile workstation

HOW TO

Securing the enterprise digital assets

EN COUVERTURE

Flash Storage

Breaking the ultimate barriers

VIEWPOINT

Journey to CERN The ultimate infrastructure

HPC Media
The HPC and Big Data Global Media

HPC Media
11, rue du Mont Valérien
F-92210
St-Cloud, France

CONTACTS
editorial@hpcreview.com
subscriptions@hpcreview.com
sales@hpcreview.com

PUBLISHER
Frédéric Milliot

EDITOR IN CHIEF
Joscelyn Flores

EXECUTIVE EDITOR
Ramon Lafleur

CONTRIBUTORS TO THIS ISSUE
Amaury de Cizancourt
Steve Conway
Roberto Donigni
Ramon Lafleur
Stacie Loving
Renuke Mendis
Marc Paris

ART DIRECTOR
Bertrand Grousset

SUBMISSIONS
We welcome submissions. Articles must be original and are subject to editing for style and clarity