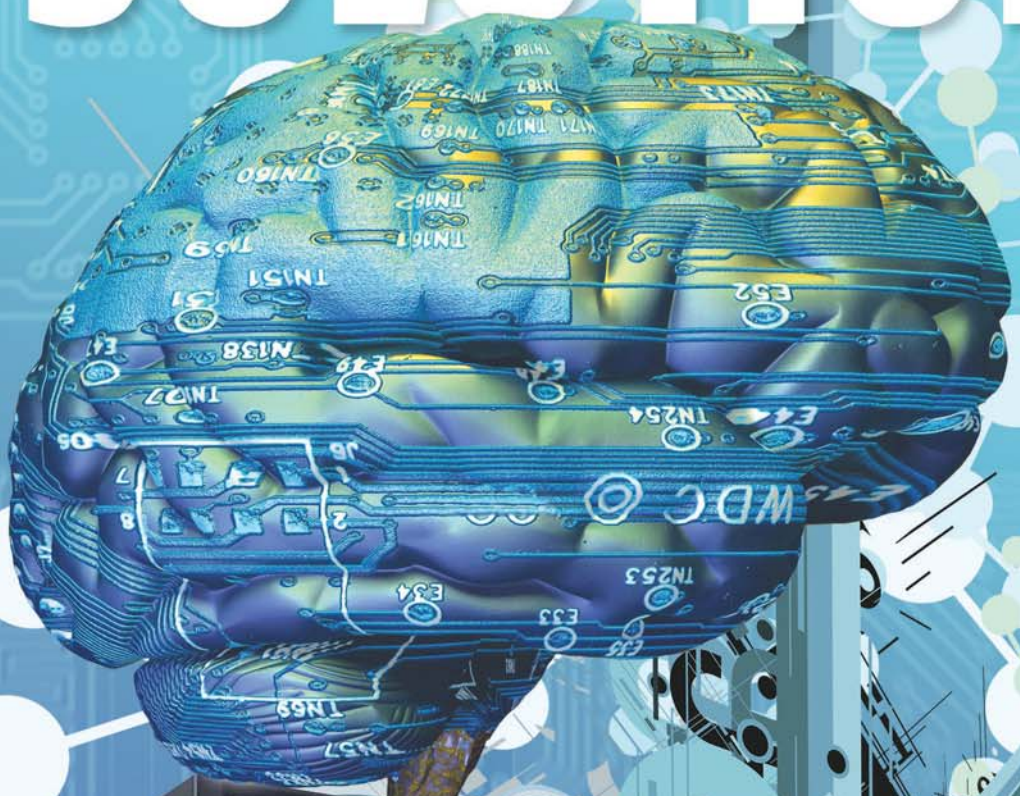
# STORAGE SOLUTIONS



French Institute for Brain and Spinal Cord Disorders gains petabyte storage

Promise Technology on choosing the right solution

PROMISE®
TECHNOLOGY, INC.

# Keeping Storage in Mind

When needing to manage petabytes of data, how do organisations decide which storage option is best? We find out

High-Performance Computing has increased the realms of scientific discovery, but as the volume of data being generated through research continues to grow at an exponential rate, organisations are faced with the question of storage. Moving forward, this is the biggest challenge faced by research centres that don't want to be limited by the speed and capacity of their solution. Founded in Paris by Professors Gérard Saillant, Yves Agid and Olivier Lyon-Caen, the French Institute for Brain and Spinal Cord Disorders (Institut du Cerveau et de la Moelle épinière, ICM) needed an advanced storage system that would combine high performance, data security and manageability.

## Common goals

The product of a new model in neuroscientific research, ICM is bringing together more than 800 researchers in the fields of molecular and cellular neurobiology, neurophysiology, cognitive sciences and related therapies. The main scientific objective of the institute is to understand the causes and mechanisms behind the major nervous system pathologies, and to offer new and specific treatments within a short-term perspective. Consolidating these activities in one place will allow researchers to take a multi-disciplinary approach to the discovery of new treatments for neurological disorders.

Thanks to progress in neuroscience research, the treatment of nervous system diseases, which are generally severe and often palliative, will become more effective. Potentially, such diseases could be prevented or cured. In order to achieve this objective, the information technology team under the direction of Dominique Bayle developed a vision of an architecture based on three pillars: High-Performance Computing, the network, and data storage. As the role of the ICM is to provide researchers with the technical infrastructure to perform their work in the best conditions possible, the primary guideline is that the infrastructure, and in particular the data storage, should not be a hindrance to research.

## The deciding factor

Following an invitation to tender, a high-performance computing solution proposed by NEC, Active Circle and Promise Technology was selected. Differentiating between vendors can often prove difficult, but Bruno Lecointe, sales manager Western Europe, from NEC Deutschland believes that while price will always be a necessary consideration, technical expertise should always be the most important factor. 'Within the storage market, companies are often focused on simply selling products, rather than being a true solutions provider,' he says.

'Because we don't operate through resellers, NEC is different and dealing directly with ICM meant that as the institute was doing its procurement we were able to point out the technical limitations and difficult requirements very quickly. Most importantly, we carefully examined ICM's individual needs, paying close attention to potential bottlenecks, and were clear on which of their requests were possible to fulfil, and which ones were not.'

> 'Within the storage market, companies are often focused on simply selling products, rather than being a true solutions provider'

NEC, Active Circle and Promise Technology provided an intelligent SAS/SATA based high-end storage solution which enables scientists to store their data with a high level of reliability and data protection, in a cost-effective solution. The enterprise-class, cable-less system design includes fully redundant, hot-swappable controllers, power and cooling units. All system components are constantly monitored using industry standard protocols to ensure continuous operation in case of failure or component malfunction. RAID storage by Promise Technology's Enterprise VTrak EClass systems equipped with Serial ATA (SATA) hard disk drives have been configured within RAID 6 disk arrays.

Petabytes of expected data meant that the capacity for handling an extremely high load of input/output operations (IOPs) was needed, and this was fulfilled by high-speed Serial Attached SCSI (SAS) host interfaces in a fully-redundant 'No Single Point of Failure' system architecture. The combination of enterprise-class RAID storage systems with cost effective,

but proven SATA hard disk drives provides the necessary performance and capacity ratio.

## Virtualisation

Installed on all storage nodes as a clustered file system with integrated security and tiered-storage management, the software provided by Active Circle essentially virtualises the storage. This meant that ICM's information technology team had an open choice when it came to selecting the hardware. The solution is compatible with standard hardware and is particularly well-adapted to the management of large volumes of data, natively protecting against errors, corruption and disasters.

One of the benefits of implementing a software solution such as this is that a limitless amount of disk arrays can be converted into one large file system. Philippe Boyon, marketing and business development at Active Circle, explains: 'With facilities like the Institut du Cerveau et de la Moelle épinière that have multiple servers and disk arrays, the ability to virtualise and consolidate all the data into one, or even a few file systems, can make a big difference to the pace of research.' He adds that providing researchers with quick and easy access to information is a primary concern, but beyond that is the issue of protecting the data. By managing the replication between file systems within the storage environment, the Active Circle software guarantees data protection without the user needing to implement any additional software. A further benefit of this software is that the file systems can be managed independently of the storage itself.

## Future growth

The storage solution delivered by NEC, Active Circle and Promise Technology has a capacity of almost three petabytes – necessary for storing the massive amounts of data generated from neuro and cellular imaging. While suitable now, that capacity will no longer fulfil requirements in the years to come. 'In this kind of research environment, each generation of experimental device increases the amount of data it produces – sometimes by a factor of 10. This means that IT departments are constantly trying to anticipate those exploding needs and requirements for capacity,' Boyon explains.

Currently under consideration at ICM is the introduction of tape storage in preparation for when a second tier of storage is required. The Active Circle software allows for the implementation of a multi-tiered storage architecture and so if a tape environment is required the institute will simply need to define a rule for migrating the data from disk to tape. The modular and scalable nature of the solution also means that the capacity can be increased on either disk or tape, while the hardware performance can be improved by adding servers.

NEC is currently performing some proactive maintenance on ICM's system in preparation for an increase in the number of petabytes. 'One of the key points is that we will be able to do this in a very short space of time, and without stopping the system,' says Bruno Lecointe. 'We delivered a solution that can easily be increased to six petabytes, and we will add in new disks and transparently grow the size of the file system in order to ensure that ICM has a storage system that meets its demands, now and in the future.'

**For more information, please visit:**

ICM | www.icm-institute.org

NEC High Performance Computing Group | www.nec.com
Bruno Lecointe | blecointe@hpce.nec.com

Active Circle | www.active-circle.com
Philippe Boyon | philippe.boyon@active-circle.com

Promise Technology | www.promise.com
Albrecht Hestermann | albrecht.hestermann@eu.promise.com

scientific computing world

Special supplement published on behalf of Promise Technology by *Scientific Computing World*
The Spectrum Building, The Michael Young Centre, Cambridge, CB2 8PD, UK
Tel : +44 (0) 1223 211 170 www.scientific-computing.com

# Talking Point

## Albrecht Hestermann,

Promise Technology's director of marketing EMEA, discusses the importance of choosing the right storage solution

At the heart of it, high-performance computing always comes down to a question of speed. Whether HPC systems are being deployed for engineering, research or development work, the request is always the same; make it faster. It's a wheel that all of us in the industry seem to be on. Computing strategies are changing and as the number of CPUs and blade servers increase, they create block-oriented data transfers and the real challenge is to have a storage system that is able to fulfil the huge number of input/output operations per second (IOPS) coming from these servers and CPUs. There are solutions on the market that provide an answer, such as our VTrak E-Class RAID storage that can manage up to 400,000 IOPS, but the demands placed on storage systems continue to grow.

One solution that has appeared in the past few years is the approach of having everything developed around parallel file systems so that metablock data and object block data are in different storage systems – the object data especially will be in several server blocks. At least one petabyte or more would be available, but in several blocks rather than one row, and supported by specialised storage servers in between. Having one exabyte block available is still a couple of decades away, however we will always have blocks that are combined in intelligent ways, such as Lustre solutions, parallel file solutions or virtualisation software, which are making these dedicated servers accessible to applications.

> 'Those scientists today who do work with supercomputers have a surprisingly large knowledge of CPUs, speed and data traffic, but what they lack is that consideration of how best to store the data'

One area that is receiving more attention is cloud computing. The issues differ from those associated with supercomputing, and the answers for cloud computing won't be found with large data streams. IOPS are important and if one storage system is not able to meet demands then of course we need to put the extra part in a parallel storage block. In some ways, cloud storage issues are easier to deal with because you can direct the storage to a dedicated server or application and when we are dealing with exabytes of data in the future, there will be a couple of blocks available instead of just one.

Entry-level computers that are coming onto the market from companies like Cray offer very high speeds and can be used for a range of applications, including private clouds. Looking at the industry from this point of view, it's possible that supercomputing, or rather the ideas that are coming out of supercomputing, will influence the cloud computing market. Of course, supercomputing strategy is firmly rooted in increasing the number of CPUs, servers and nodes. It's difficult to predict what will happen, but we are seeing a growing amount of companies within this market that are looking to offer entry-level enterprise applications around cloud computing.

Within this particular market we see hardware virtualisation platforms as being the most important. This is because there is a lot of technological insight at the heart of this storage in terms of balancing the capacity, the speed and the IOPs. There is also the matter of the deviation between the caches used by SSDs combined with primary and secondary storage. Here, intelligent hardware virtualisation platforms, such as Promise's VTrak S3000, are requesting more and the bottleneck is still on the application side going to storage or server farms. In this area, it seems that 10 Gb/s iSCSI are becoming more popular.

Ultimately, the supercomputing market is looking for the highest performance possible. Speed is not a luxury, it's a necessity and reducing the time that scientists take to do computational work makes a huge difference. Looking at the science market, we see that just 10 per cent have access to and use high-performance computers. The other 90 per cent still need to compute, of course, but they do so by using IT centres or platforms that have been developed by their own internal IT departments.

Those scientists today who do work with supercomputers have a surprisingly large knowledge of CPUs, speed and data traffic, but what they lack is that consideration of how best to store the data. Choosing a solution is not an easy decision and as each facility that uses high-performance computing will have its own set of requirements, what may suit one will not be the best option for another. Users will often view certain solutions as being either too expensive or too slow and as bottlenecks can occur as a result of the storage, it is important to seek expert advice. The success and speed of a system is often dependent on the storage, and no solution is equal to another.