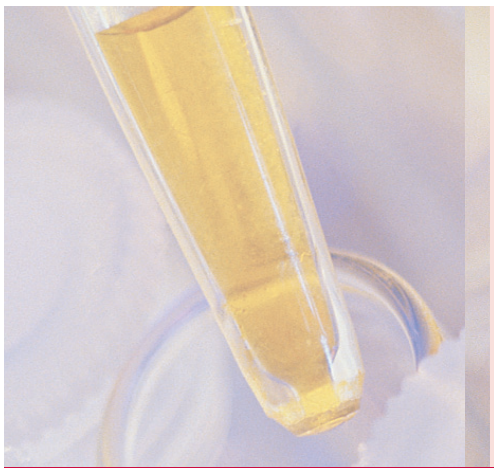


Laboratory Product Focus

AN INSIDERS GUIDE TO DATA STORAGE

Keith Vickers, Mark Hiley, Mark Brierley.

Data is one of the highest forms of currency for companies today with many going out of business purely due to a failure to manage and protect information correctly.

The same applies to the scientific research community where faster processing, larger bandwidth and more efficient ways of accessing data are improving the speed of new discoveries all at increasingly affordable prices.

STORAGE FOR SCIENCE

Storage technology has evolved and matured dramatically in the last 10 to 15 years as industries have increased the amount of data they produce and process. As the technology has developed to meet the need to store more data, speed and available bandwidth have increased while connection costs between remote sites continue to fall. Also, software now exists for data to be moved easily and efficiently, even between systems remotely.

As the cost of storage continues to come down, putting more data on disk, rather than the original tape-based storage systems, is both practical and economical. Having significant amounts of data stored on tape in large computer rooms makes accessing the right data particularly time consuming. Advances in disk technology now means the difference between access times of 10 milliseconds, rather than 10 minutes, or even 10 hours, which was the case only a few years ago. The impact these timescales have on the development and execution of a research project is considerable.

Scientific research and wider projects have now moved from being largely theoretical to increasingly more practical as technology enables experiments to be carried out in virtual environments with fast access to data. Researchers now have rapid insight into problems, can tackle larger issues and complete vital research projects more quickly. Crucially, sophisticated storage making data readily available means that researchers and scientists can review data as many times as needed to check its output quality.

SGI® InfiniteStorage TP9700 Storage System

Keith Vickers, SGI Pre-Sales Manager, SGI

Another important development for the scientific community is the ability for data sharing between researchers at different institutions. Government initiatives are funding the creation of central storage systems that enable researchers to feed in data and accesses information from various establishments across the country. When looking to buy a storage system, departments now need to consider compatibility with these larger systems. With the appropriate infrastructure, right data formats and processing power, storage can make an enormous difference to the amount of information available to other institutions, in addition to creating a more attractive work environment for employees.

> ***"BUT, WITH A MULTITUDE OF OPTIONS AVAILABLE, WHERE DO SCIENCE ORGANISATIONS START? HOW DO YOU CHOOSE THE RIGHT STORAGE SOLUTION TO MATCH YOUR INSTITUTION'S INDIVIDUAL NEEDS?"***

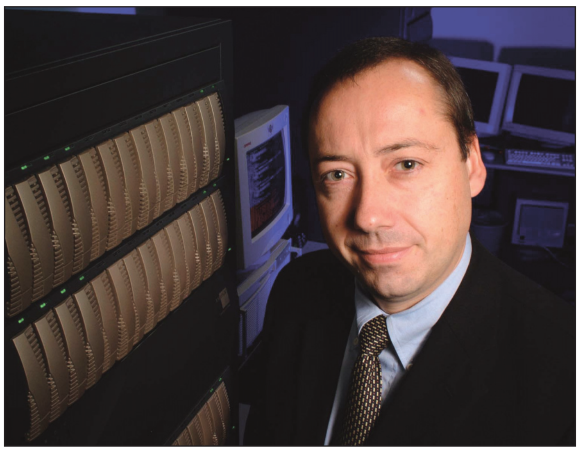
PRICE VERSUS PERFORMANCE

With all these considerations, where does a modestly sized research department start when purchasing a storage system? Speed of access will undoubtedly be a priority. With increasing amounts of data to process, a system that can automate functions is needed, allowing scientists time to concentrate on the research itself. These solutions also have to be simple to use and implement. With limited IT resources, most departments will want a 'plug and play' system which needs to work from day one, as well as being quick to set up, operate and configure. The department team are researchers, not IT experts, after all.

Understandably, with limited funds available for many research and science departments, the first consideration when purchasing storage is price. However, one of the most common mistakes comes from the false belief that similar systems will yield the same performance. Often departments opt for the cheapest when in fact performance differences between systems can vary by up to 50%. In essence, the differentiator between systems is the storage controller, which manages the disk drives and delivers the physical storage space to the servers or applications that need it. The storage controller effectively controls the bottleneck to performance and its design can determines the overall performance of the system. This should be the key consideration when purchasing storage. There are vendors in the market selling storage systems with very low cost controller technologies, but it often worth paying a small premium to avoid problems further down the line.

Author Details:

Keith Vickers, SGI Pre-Sales Manager, SGI
Mark Hiley, Director Northern EMEA, LSI Logic
Mark Brierley, Technical Consultant, LSI Logic



Mark Hiley, Director Northern EMEA, LSI Logic

Another important consideration is the efficiency of the system. As mentioned earlier, ease of use and timesavings are the crucial factors for the smooth and efficient running of a project and department overall.

A system that has zero downtime, or in other words is running 24 hours a day, has untold benefits. Any reconfiguration or change that needs to be made can be performed without having to shut the system down, meaning no extra time is taken or data unavailable for lengthy periods. Downtime implications are often not taken into account when buying a system before its too late. With data becoming ever more important for an organisation to thrive, the levels of storage required to protect that data should be given due consideration and the best possible back-up functionality. It is these crucial, but often overlooked features and capabilities that may attract a premium in the short-term, but protect an organisation, as well as improving work life quality, for the long-term.

Many research and science departments have adopted cluster-based server technology, which has, traditionally, been a pinch point when attaching storage. The first clusters on the market had storage per node and data had to copied across all the nodes. As clusters matured, storage became more sophisticated and network file systems, from the nodes to a head node, could be attached to the storage at the back-end. Today, more changes are being made to technology. For example, Infiniband, where the interconnect of the nodes and the data can be centralised and away from the cluster itself. At present, Infiniband beats almost all other connections for current and planned transmission rates, achieving high speeds based on the strength of its architecture. Using Infiniband, a processor can be coupled with a central memory and rudimentary extensions together with other identical elements to build a functioning server system. Such an end node can communicate with every other node in the same 'subnet'. Infiniband enables the sharing of this interconnect to deliver blistering data simultaneously between processor nodes and storage.

BUYING THE RIGHT DISK STORAGE

Disk is now the dominant storage technology, but mistakes can still be made when buying a storage system with the wrong kind of disk. Currently, the cheapest is the SATA (Serial ATA) disk which is only suited to certain types of applications and thus not always the most appropriate, despite claims from its vendors. Unfortunately, many organisations discover this too late and try to shoehorn the SATA storage into the wrong type of environment. In a worst case scenario, this causes a loss of data and severely disrupts uptime.

The alternative is a system based on fibre channel disks. These yield better performance and more availability, while providing the basis for networked storage – something that is now becoming a reality for smaller enterprises due to its increasing affordability. However, the perception remains that these disks are expensive, and so smaller organisations are reluctant to invest. The good news is that costs are being driven down with the introduction of SAS (Serial Attached SCSI), which is the replacement technology for SCSI, the protocol for connecting physical disk drives to servers and storage controllers. What makes SAS different from SCSI is that it is networkable and more robust, whilst giving the ability to network storage cheaply. Also, as the market for SAS technology matures and completely replaces SCSI, the volumes will increase, the cost per drive will decrease and it will ultimately be easier to deploy networked storage systems. So, SAS can provide a relatively easy means of starting to deliver a high-speed low-cost fibre channel SCSI with improved connectivity.

Mark Brierley, Technical Consultant, LSI Logic

UTILITY STORAGE MODEL

Looking ahead, there are new storage models emerging with the flexibility to control storage costs further. 'Capacity on Demand' is a storage services model, which operates in much the same way as a utility provider, offering storage capability as and when needed. These 'Capacity on Demand' storage providers need to start understanding the end-user's business, what type of research they are doing, and the demands on the infrastructure to develop a tailored service with the flexibility for growth. This model also allows for the development of a tiered approach specific to individual storage requirements. In other words, there may be the need for fast available storage space for key data, but a second storage requirement for archive data, which will require limited access. As the pricing structure is such that you pay for what you use, both come at different price points. The third layer could be a disaster recovery requirement, which could be taken offsite or copied. Because of the economies of scale it provides and greater capability, along with advent of SAS, this service model has significant benefits for smaller organisations.

WHAT THE FUTURE MAY HOLD

This article has discussed clear advances in storage performance increasingly at a lower cost point particularly in the areas of networked storage. What this means for research departments is that it will be cost effective to deploy storage area networks instead of directly attaching storage to specific servers and even outsourcing storage as a 'Capacity on Demand' service becomes increasingly more affordable.

Key technologies such as SAS and newer and faster controllers coming onto the market are driving change. The latest generation of storage users are starting to use four gigabit and, in just two to three years, eight gigabit will not be unrealistic.

The UK government's Energy Review and upcoming legislation is also forcing the industry to change. It is now the responsibility of business and industry to reduce carbon emissions and storage manufacturers will have to adapt regulations. With these governmental pressures, the pace of adoption and new proven cost-effective technologies, the science community is looking forward to driving vital research using even more sophisticated storage systems.