# Data Storage is the Key to Business Success

*a report by*
**Patrick Khoo**

*Program Manager, Modular Connected Storage Architecture Group, Network Storage Technology Division, Data Storage Institute*

Patrick Khoo is Program Manager of the Modular Connected Storage Architecture Group in the Data Storage Institute's (DSI's) Network Storage Technology Division. He was the network administrator at the Office of Education Technology at Western Michigan University (WMU) before joining DSI in 1996, where he became a senior network specialist. Given the task of setting up a new research and development department to pioneer efforts to develop world-class competencies in networked storage technologies, which had not yet been explored in Singapore, he created and led a team of research engineers in the new Modular Connected Storage Architecture Group. His team went on to invent the innovative and new HyperSCSI Ethernet-based network storage protocol and demonstrated both high performance and diversity in application. A graduate of WMU, Mr Khoo has been active in developing, managing and finding networking solutions.

Today's world is an uncertain one. Between natural disasters, changing governmental policies and directions, regional conflicts, political scandals, terrorism, medical epidemics and even human negligence, it is difficult to see through this fog of unpredictability and chart a safe course for business. Contrary to popular belief, most business people can, in fact, deal with economic downturns, poor business sentiment, high running costs, aggressive competitors and the like. What they cannot handle is unexpected events, even good ones. Business people hate uncertainty and the inability to foresee or to know what is happening. Not knowing can mean the loss of a deal or even the loss of the entire business.

Knowledge and information are therefore the keys to survival and success in today's tough economic climate. Businesses have been jumping on the customer relationship management, knowledge management and enterprise resource planning (ERP) bandwagons in an effort to understand better the business processes, customers and forces at work. We now track everything, including (but not limited to) e-mails, documents, phone calls, stock inventory, customer orders, finances, logistics details, manufacturing quotas, meeting minutes, contact reports and Web clicks, etc. All this information is accumulated over days, months, weeks and even years in the hope of finding trends and patterns to help the business person cut a swathe through the fog of unpredictability and to prevent or reduce the uncertainty of doing business in an uncertain world.

Before businesses start looking into ways to extract trends and patterns from all that raw information, two major components must be addressed: the instruments of data collection and cabinets of data storage. Instruments of data collection relate to devices, terminals, computer systems or human beings from which information is generated. This includes point-of-sale terminals, clerks doing inventories, fax machines, temperature monitoring stations, quality assurance engineers and many more. These instruments of data collection generate lots of raw information that will be processed later in the hope of generating a trend of some kind. However, before and after being processed, this raw information needs to be kept safely and securely in some kind of device – a cabinet of data storage. These cabinets of data storage include filing cabinets, laptops, desktop PCs, ring-bound folders and mainframe computer systems.

While methods and tools for storing non-digital data have existed for centuries, digital data storage is almost half a century old, from the days of the first hard disk drive, the IBM RAMAC (random access method of accounting and control), in 1956. Not only is the advent of digital data relatively recent, but it is progressing faster than non-digital data. According to a Berkeley University study in 2001, it was estimated that humans have generated a total of 12 exabytes of data to date. What is truly shocking is that this data is projected to double within the next two and a half years alone. Having a good plan to store digital data is therefore imperative for the functioning of any business intending to cope with this deluge of information.

Technologies for storing and accessing digital data can be divided into two broad levels: an infrastructure base; and data and information management systems. An infrastructure base includes the fabrics, switches, interconnects, device management interfaces and, of course, hard disk drives and tape libraries. Information management systems cover systems and software such as tape management, back-up schedulers, storage virtualisers, file systems and metadata controllers, etc. What really makes things difficult is that so many complex systems and components are being used to save, store and retrieve digital data. The promise of a single unifying data storage and management solution is still a long way away.

In the meantime, we are forced to struggle with the battles of Fibre Channel versus Internet Small Computer System Interface (iSCSI), in-band versus out-band virtualisers, high costs, complex deployment and not enough trained and experienced administrators. It is important to look past the 'technobabble' and find the solution that best fits your needs. The only way to do this is to keep to the basics.

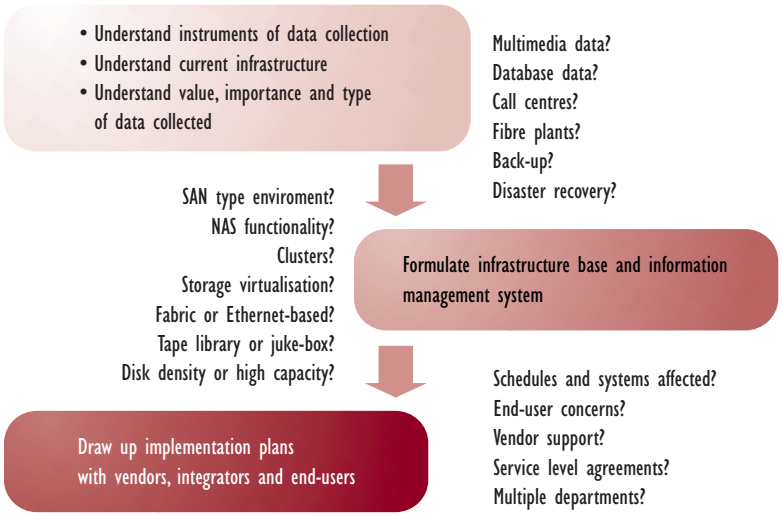
First, it is important to identify the instruments of data collection and the infrastructure that supports them. Is your primary data collector your e-mail system, the customer database or your call centre? This step is particularly important because different systems and types of data have different characteristics. For example, if you are supporting an e-mail system, then speed and performance are not as important as reliability and scalability. On the other hand, call centres demand information as fast and as accurately as possible to support their customers. Another factor to consider is whether there is a lot of physical paper to be scanned and entered into an ERP system. With regard to front-line customer-facing desktop systems, how many of them are there and are they distributed geographically with access to multiple database servers? How many servers are there that really need access to a large data plant? Are there a lot of existing cable plants that are near capacity?

The next step involves recognising the value and importance of the data generated from the instruments of data collection. This will determine the kind of cabinets of data storage that need to be deployed. Questions to ask concern: whether you need disaster recovery and no downtime; what about archiving for back-up and legal requirements; and are there more database systems or multimedia files? This will determine whether more physical disks are required at the expanse of larger capacity disks. Needing more physical drives for high transaction databases might mean larger high-density cabinets with 10,000 or 15,000 revolutions per minute SCSI drives, while media files might do just as well with large capacity 250GB Integrated Drive Electronics or Serial Advanced Technology Attachment-based disk arrays. If you have short back-up windows, perhaps you need to consider hard disk accelerated tape back-up solutions.

To identify the information management systems, the following questions should be asked. Do you have several Windows clusters to support? Perhaps you should consider using metadata controllers or file systems. Do you have a lot of desktop files or computer-aided designs to move around? Perhaps network attached storage (NAS) with storage area network (SAN) back-end storage is enough to manage the files in this case.

Once a suitable information management system, base infrastructure and data collection instruments have been identified, it is necessary to assemble all the components. Whether they are SAN, NAS, clusters, fabrics or metadata controllers, putting them all together will require getting buy-ins from management and users. After gaining support from them, a team of trained experienced professionals are needed who can deliver results. After all, you would not want to trust this job to just anyone.

*Figure 1: Making a Plan for Data Storage*




When putting the team together, both end-users and internal IT support staff must be included in the mix. Often, too much reliance is placed on the vendors and integrators to do all the work. The reason for this is to ensure that internal staff learn ownership of the new system. Do not underestimate the impact of the 'human factor' on successful day-to-day operations. Be wary also of solutions trying to support multiple departments or business units. Different data and different users may need different systems and solutions, as much as we would like to have a single unified solution.

To sum up, in order to help your business survive these uncertain times, lots of information must be obtained and then processed to help understand better the dynamics of the business, the customers and the processes within the entire value chain. To do this, data must first be collected and stored in a safe place.

In drawing up a plan for this to work, many technologies and solutions are available, but it is important to not lose sight of the final goal. Keep to the basics, understand the value of the data and its unique characteristics and then you can choose an architecture or solution that best fits your needs. The solution needs to be evaluated at two levels: the base infrastructure and the higher-level information management system. Finally, the human element of this grand scheme should not be underestimated or overlooked.

It is going to be a difficult progression, but one that will result in your business having a new level of competitive advantage over its competitors. This data storage solution will help supply you with the information, trends and patterns that can make it possible to chart a course into the future. ■