Dealing with data storage

Richard Bray

DATA STORAGE USED to be easy. When paper overflowed the filing cabinets, the oldest folders went into boxes and someone from the central records department took them away. The equivalent today is usually every bit as simple – disk drives are so inexpensive, it's tempting just to keep chaining them together forever.

Unfortunately, it is not that simple. Ever-growing RAIDs (redundant array of inexpensive disks) are only part of the answer and hardware is only one cost component of data storage. In fact, the hardware may soon be much less expensive than the associated software and services.

It's almost a cliché to note that 'not all information is created equal,' but the idea does lead to information lifecycle management (ILM). Consider its exact opposite, the search engine, Google. While the search engine gathers every stray thought on the Internet and throws it in a big, searchable bag, ILM works from a set of policies and procedures to ensure that every bit and byte is in the right place; while Google is open 24 hours a day to anyone with a computer, ILM decides who gets to see what; and while Google faces the daunting task of archiving the entire Internet forever, ILM sets out to keep only important information only as long as necessary, and no longer.

Governments are all about information. They collect, produce, modify, distribute and, above all, they store it. The kind of information they collect will increasingly dictate how they store it. Take a police investigation, for example. No matter where it is on the network, or what kind of data it is, the storage system must deliver information quickly to investigators' desktop computers.

Police databases might store lists of witnesses and descriptions of physical evidence, with predetermined slots for each piece of information, and different ways to search through it. Individual officers would write and circulate reports and statements, keeping copies on their desktop hard drive; this kind of information would be sorted into folders, and retrieved by file name. Emails related to the case might remain on the departmental mail server. Photographs, X-rays, audio tapes and video recordings that exist in the real world would all have their digitized equivalent somewhere on the network, available for viewing but not for modification. Unlike a database or a word-processed file, this kind of 'fixed content' does not describe itself with the text it contains, so to make it searchable, someone must enter a detailed description.

Fixed content is a looming nightmare for many storage managers, because the files are not only big, they tend to multiply. People would much rather send the same photographs or videos to a dozen other people by email than put them in a common folder or website and email just the address. Well-designed storage management systems can work quietly behind the scenes on network, finding and deleting all the copies of a file while redirecting all requests to the remaining one. That might slow the growth for a while but digital cameras, websites and video-conferencing are all gathering and creating content faster.

Data storage is an area where the purchasing decision has to mesh very closely with organizational goals. Is there legislation or policy covering the retention of records? Is it likely to change? Can the existing network support new storage processes? Is security both adequate and user-friendly?

If you are thinking about a data storage decision, you might want to wait until companies in the United States figure out Section 404 of the *Sarbanes-Oxley Act*. After a wave of corporate criminality badly shook consumer confidence, the US Congress legislated strict controls on financial records management. The deadlines for large and small companies have changed, and may change again, but at press time companies with over $75 million in sales had until November 15, 2004 to comply, while smaller companies had until June 15, 2005.

There are plenty of ways to fail the compliance test. At one recent conference, a senior auditor listed some: processes and their documentation don't match; undocumented processes, or documented processes are not followed; customized and 'home-grown' applications are insecure; former employees and consultants retain access; too many people with too much access, and no way to record who did what and when; developers test new processes on 'live' systems; users can bypass controls to reach critical data; and, duties are not segregated, so some employees can improperly grant themselves access.

All this means that right now, the data storage industry is very busy looking after some highly motivated buyers. Once the dust dies down, a good selection of data storage solutions should be available.

Richard Bray is an Ottawa-based freelance writer specializing in the IT sector. He has been published in magazines and newspapers in Australia, the US and Canada. Before freelancing, he worked as a producer, reporter and senior writer for CBC in Toronto.