



The toxic terabyte

How data-dumping threatens business efficiency

Contents

- 2 Too much of a good thing**
- 5 When terabytes turn toxic**
- 7 Taming the data beast**
- 9 IBM Global Technology Services,
ILM and the end-to-end solution**
- 10 Time for a data detox**

As companies, government departments and other organisations accumulate information at an accelerating rate, they face growing costs and inefficiencies that threaten their ability to function. The answer lies not just in new ways of applying IT technology and services, but also in changes in individual and organisational behaviour.

Too much of a good thing

This year, electronics manufacturers will produce more transistors – at least 26,000 million million of them – than the world’s farmers grow grains of rice. Packed on to the chips that power personal computers (PCs), mobile phones and a host of other devices, the fundamental building blocks of information technology (IT) will each cost about the same as one printed newspaper character.

These figures are a reflection of Moore’s Law, first advanced by Intel** founder Gordon Moore in 1965. He said that each new design of chip could be expected to do twice as much as its predecessor, leading to an exponential rise in performance matched by a corresponding fall in the cost of computing power. Moore turned out to be right, and his prediction of a doubling in performance every 18-24 months holds good to this day.

The rise in chip performance is generating a vast and expanding store of data.

At first sight, the proposition of ever more power for less and less cost looks like a good thing, helping to support the advance of pervasive computing for all kinds of desirable purposes. But there’s a downside, and one that is becoming very hard to ignore. For computer systems not only process data, they also store it, and in increasingly vast quantities.

It is projected that just four years from now, the world’s information base will be doubling in size every 11 hours. So rapid is the growth in the global stock of digital data that the very vocabulary used to indicate quantities has had to expand to keep pace. A decade or two ago, professional computer users and managers worked in kilobytes and megabytes. Now schoolchildren have access to laptops with tens of gigabytes of storage, and network managers have to think in terms of the terabyte (1,000 gigabytes) and the petabyte (1,000 terabytes). Beyond those lie the exabyte, zettabyte and yottabyte, each a thousand times bigger than the last.

Highlights

Some observers have likened what is happening to the Industrial Revolution, when economies made the first move away from individual craftsmanship and towards the production line, with its potential for quantum increases in output. Except now it is not pots and pans or cars that are being produced in their thousands, but data bits in their millions, billions and trillions.

The trend is ever upwards. Processor-based devices sell to a rapidly expanding mass market. People are becoming more and more used to capturing and storing still and moving images, music and other entertainment content. Businesses, governments and other organisations now depend for their very existence on networks and databases, to the point where survival of the information stock can matter more than the temporary loss of an office or factory.

At the simplest level, company e-mail systems spawn large amounts of data. Business e-mail – some of it important to the enterprise, some much less so – is estimated to be growing at a rate of 25-30% annually. And whether it's relevant or not, the load on the system is being magnified by practices such as multiple addressing and the attaching of large text, audio and even video files.

The data explosion affects all areas of commerce.

No organisation is immune. The data explosion affects the whole of commerce, from manufacturing to financial services.

One industry, aerospace, is undergoing a transformation in the way it designs and builds aircraft and their major subsystems. Where once a big aircraft project depended on thousands of pen-and-paper drawings created by skilled technical draughtsmen, now the information underlying a new design is all-electronic, created by powerful computer-aided design and manufacturing (CAD/CAM) software. It's a great advance on the older methods, allowing much faster prototyping and helping to eliminate errors before they become expensive at the final-assembly stage. But it also throws up mountains of data.

Another technique now part of the aerospace mainstream is computational fluid dynamics (CFD) – the use of very powerful computers to simulate things like the airflow around aircraft and into engines at hypersonic speeds. While CFD helps to eliminate a lot of costly and risky testing, it also gives rise to huge data sets running to many terabytes.

Highlights

Typical of the data challenge facing the financial services industry is the practice of quantitative analysis – mathematical modelling of how a particular security, a complex trade, or an entire market will behave in the future. A key input is the historic price of an asset, and it is not uncommon to use 20 years' worth of such information. Originally the analysts looked at daily data sets – opening and closing prices plus daily volumes – running to several gigabytes in size. Now they need to work with the price and volume for each and every trade of a particular stock over a number of years, and the data sets have reached the terabytes.

Even bigger masses of information will result from the efforts of some nations to digitise whole populations. National health services are moving to digitise patient records, including the results of diagnostic procedures such as X-rays and MRI scans, while the British Government aspires to a national identity database covering more than 60 million people.

The creation of data is growing at an exponential rate.

From the schoolgirl with her MP3 player and picture-phone to the middle manager broadcasting memos by e-mail, from the doctor calling up comparative images of a tumour to the policeman checking a licence plate, the citizens of the industrialised societies are using and creating data at a rate that's growing exponentially.

Most of this rising tide of information is being stored – on the laptops and smartphones of individuals, on company servers, in offsite archives and data warehouses. Finding a physical home for it all is still straightforward – the technology of storage is advancing as fast as that of processing, and prices are falling at least as quickly.

But the ready availability of somewhere to dump data after its immediate window of usability is masking a problem that could over the next decade gravely affect the profitability of businesses and the efficient functioning of health services, police and security forces, local and national governments, and many other types of organisation.

Highlights

When terabytes turn toxic

Knowledge is power – but only if it can be extracted quickly and efficiently from an ever-growing mass of data. Businesses and other organisations now see their information stocks snowballing beyond their ability to manage them and beginning to work against the health of the enterprise by damaging efficiency and bottom lines.

The stock answer to the data pile-up is more cheap storage and lots of it. But reflexively pumping everything and anything into an apparently limitless reservoir hurts the organisation in three ways:

- 1. It becomes harder and harder to retrieve information promptly*
- 2. More people are needed to manage increasingly chaotic data dumps*
- 3. Networks and application performance are slowed by excess traffic as users search and search again for the material they need.*

The traditional data storage solution is not as cheap as it seems.

As these penalties of the keep-everything culture make themselves felt, organisations are beginning to look at the true cost of throwing hard disks at the problem and finding that the solution is not as cheap as they once thought. The power bills are no longer negligible, and the likelihood of mandatory controls on CO₂ emissions could create a whole new source of cost in the future.

Finally, there are those who believe that data is accumulating at such an accelerating rate that the time will come when it will outstrip storage technology as it is now understood – that no amount of disks will be enough to soak up the deluge.

That doomsday is probably some time off. But companies are already starting to find that problems with information retrieval aren't just a nuisance – they cost real money.

Highlights

E-mail has proved to be one of the first sources of this corporate pain. Once seen as nothing more than a quick and flexible communications tool, e-mail is now estimated to be the platform for as much as 75% of company intellectual property. E-mail documents figure in some 75% of all cases of corporate litigation. Sheer weight of usage means that the medium has in many organisations become the primary record repository, a fact recognised by legislation requiring the long-term retention of messages.

E-mail storage and retrieval is just one problem facing organisations today.

Companies are now learning the hard way about the need to take e-mail storage seriously. Five US banks were recently fined US\$1.25 million each when they failed to retrieve e-mails that were demanded of them. One Fortune 500 company had to spend US\$750,000 to dig e-mails out of an archive in response to a legal subpoena. A pharmaceuticals company was forced to devote time and people to searching through 30 million messages for a court case.

Regulatory insistence on data retention looks set to continue unabated in the future. Along with factors like the introduction of megabit-rated mobile communications services for consumers, citywide wireless Internet access and ultra-broadband wireless networking inside homes and offices, this regulatory insistence will add still more momentum to today's roaring inflation in the demand for data.

Companies and other organisations face an increasingly urgent choice about how to respond to this enterprise-threatening challenge. They can carry on dumping, creating ever bigger and more incoherent 'data pits' and paying a soaring price when they need to retrieve items of value. Or they can face up to the problem and find out what it takes to actively manage information from cradle to grave, weeding out the mass of ephemera early on and keeping only what is likely to be of long-term value.

Highlights

Information lifecycle management (ILM) presents a solution to the problem of data-dumping.

Taming the data beast

As the world at large has woken up to the need for wiser stewardship of the planet and its resources, so the IT industry has understood that the present approach to data creation and storage is simply unsustainable. Its response, which it regards as part – though not all – of the solution, is information lifecycle management (ILM).

The principles of ILM were defined by the Storage Networking Industry Association (SNIA), which includes in its membership IBM and other world-leading IT vendors. It is a process for managing information all the way from conception to disposal, based on its intrinsic value to the company and in a way that makes the most efficient use of storage while minimising the cost of retrieval.

In other words, ILM is a declaration of war on data-dumping. It's designed to eliminate low-value information as early as possible before putting the rest into actively managed long-term storage in which it can be quickly and cheaply accessed. An ILM solution is ultimately executed by hardware and software, but the optimum start, although there are others, is with development of the first filters, the working practices and the policies that determine the business value, origin and fate of the various types of data circulating on the company network.

While some vendors are willing and able to support policy development, the best judge when it comes to deciding which data is to live, and for how long, must be the company itself. At first sight, the obvious candidate for the job is the chief information officer (CIO), who after all has been wrestling with the underlying problem for years.

Highlights

Most CIOs have long since shaken off the ‘technician’ tag once associated with the role and have broadened their view of the business. But the person calling the shots on data retention policies must know the business from top to bottom and have a good understanding of future business direction. Setting data policy is not so much a technology issue as one bound up with the very nature of the company, and what it does now and what it may want to do in the future. Ambitious CIOs who want to reposition their current role as a business role, and who have sufficient board sponsorship will relish the challenge.

Another solution could be the appointment of a board-level ‘information czar,’ possibly an ambitious CIO looking for a wider role. The czar’s task would be to bridge the gulf between the IT team and top management, and to take on responsibility not only for the storage infrastructure but also for the information in it.

With the support of the board, CIOs can play a vital role in putting policies in place to address the problem.

Reinvention of the CIO is not the only change in company behaviour that’s needed. It is clear, for instance, that the threat posed by data accretion simply has not registered with most senior managers. And if it has, it is still regarded as an IT problem, to be solved by the CIO instead of demanding strategic decisions at the highest level. Individual employees tend to have a similar view. What happens on the network is IT’s business, not theirs. And in companies where everyone has a laptop, people are far too busy doing their jobs to attend to the minutiae of responsible data management.

In ILM, the IT industry has a basic set of tools to address the problem. Depending on the supplier, the package will include not just hardware and software but also help with the development of new working practices and policies, examples of best practice, and decision support for managers who have to decide which data to keep and which to discard. Above all, it will make it so easy for employees and company to do the right thing that they won’t think twice.

Highlights

A change in behaviour is required.

In the end, individuals and organisations change their behaviour only when it becomes obvious that they need to for their own good. In recent years, ecological campaigners have been very successful in waking up the population at large to the many dangers of breakneck consumption of natural resources. Though they may have a way still to go with some national governments, they have sounded the alarm on climate change, made waste recycling commonplace in many societies and given fresh impetus to the biofuels industry. Who is – or should be – beating the drum for a ‘greener’ approach to data creation and storage?

SNIA and its members have taken the lead with ILM, and the products and services based on it. Certain industry analysts are widely respected and undoubtedly have a role to play in winning hearts and minds. As the champion of national commerce, a body like the Confederation of British Industry (CBI) might be expected to have something to say. National governments likewise, though the indications are that they have not yet woken up to this threat to their economies.

IBM Global Technology Services, ILM and the end-to-end solution

IBM’s range of hardware, software, services and consultancy makes it uniquely capable of helping its customers to achieve end-to-end, top-to-bottom ILM. This contrasts with other approaches to ILM that continue to emphasise data storage over the working practices, policies, architectures and long-term service support that provide the glue for a complete solution.

IBM Global Technology Services is helping companies implement effective ILM strategies.

As a supplier of outsourced IT operations, IBM Global Technology Services is now developing complete ILM strategies for several leading corporations around the world. It is also walking the talk, having begun its own journey towards implementation of full, integrated ILM. IBM Global Technology Services is extending the knowledge gained to help large and medium-sized organisations define and implement their strategies – often beginning with smaller projects that maximise returns.

Highlights

Recent IBM successes.

- A US hospital found that its storage costs fell by half and storage capacity grew 500% after the installation of an IBM disk system
- Cost of ownership halved as a result of increased utilisation when a Fortune 500 power company implemented an IBM tiered-storage solution
- A US law enforcement agency deployed an IBM-based nationwide identification application and cut the cost per suspect identified by more than 80%
- A US regional bank was running into trouble with data storage – its stock was doubling every 12-18 months and disk utilisation was a seriously uneconomic 28%. The addition of storage visualisation software (SVC) from IBM boosted utilisation to 80% and slashed costs
- A leading state healthcare insurer used IBM technology to consolidate its storage, improving performance threefold and cutting backup times by 50%
- A global food company put 15 months of data, equal to three million documents, into an IBM-based archive. Back-office processes ran 20-25% faster and there was an initial cost saving of US\$70,000, with more in prospect.

Time for a data detox

It's tempting to think of data accumulation as a disaster in the making, a sudden and violent step change leading to system collapse and business failure. That fate could indeed befall the particularly unprepared. But the more likely outcome for most is a steady but remorseless loss of momentum, like an old-fashioned sailing ship trailing a growth of weed from its hull.

The lengthening data trails now being dragged along by the world's businesses and other organisations will soon come to hurt profitability or delivery of the mission to the point where the problem just cannot be ignored any longer.

Highlights

The IT industry believes it has part of the solution in the form of ILM – a set of principles designed to actively manage down the amount of data entering storage, and to ensure that what is kept can be retrieved quickly and economically, in the immediate future and in the long term. ILM makes evident good sense, but there is much to be done before it can take effect.

True ILM solutions address the problem at both human and business levels.

First, the industry must set its own house in order. Too many vendors pay lip service to the cradle-to-grave concept behind ILM, while continuing to offer nothing but storage and yet more storage. True ILM solutions address the problem at the human and business levels, as well as providing software and hardware tools.

Second, the effort to win business hearts and minds must be redoubled. The problem has not yet registered with senior managements, who remain ignorant of the looming threat to their profitability. CIOs, the senior professionals who should be sounding the warning, either do not have the ear of the board or are so busy slapping on sticking plasters that they can't give storage strategy the attention it demands. On the shop floor, the individual employee regards storage as an infinite resource and access to it a basic human right.

Third, and most difficult, individual businesses and economies at large need to think hard about the data overhead created by internal control processes and external regulation. Good regulation produces benefits outweighing its cost of implementation. Could it be that in some cases rule-setting and law-making have reached the point where they are doing more economic harm than good?

Sixty years into the information age, and just a quarter of a century after computers began to enter the mainstream of business and domestic life, a dismaying fact is becoming evident. We're used to thinking of information as power, as a prime source of business advantage. Now it may be about to turn into a weakness, accumulating at such rate that it could clog the arteries of commerce.

When it comes to storing information, "better safe than sorry" is no longer good enough – it's time now for the IT industry and business to begin implementing ILM in its fullest form. It may not yet be five minutes to midnight, but the shadows are lengthening.



More information

For more information on infrastructure solutions from IBM Global Technology Services, please visit:

ibm.com/solutions/itsolutions

or e-mail:

tonyr_cox@uk.ibm.com

IBM United Kingdom Limited

76 - 78 Upper Ground
South Bank
London
SE1 9PZ

The IBM home page can be found on the Internet at **ibm.com**

IBM and the IBM logo are trademarks of International Business Machines in the United States, other countries, or both.

** Intel is a trademark of Intel Corporation or its subsidiaries in the United States and other countries.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. Copying or downloading the images contained in this document is expressly prohibited without the written consent of IBM.

This publication is for general guidance only.

© Copyright IBM Corporation 2006
All Rights Reserved.

Contributors

Paul Coles – Paul is a Storage Solutions Architect with extensive experience. He has architected, designed and implemented solutions for many customers across Europe.

Tony Cox – Tony has extensive infrastructure services experience across both systems and storage management in both Specialist Sales and Consulting roles. He helps clients address their information and storage management challenges.

Chris Mackey – Chris is an Infrastructure Consultant and has worked across a broad range of industries in service delivery, infrastructure and processes.

Simon Richardson – Simon is an IBM Service Delivery Manager and has extensive client-facing experience, including work as an Infrastructure Storage Consultant.