IDC
Analyze the Future

IDC - MULTIMEDIA WHITE PAPER

As the Economy Contracts, the Digital Universe Expands

May 2009

By John Gantz and David Reinsel

Sponsored by EMC Corporation

Content for this paper is excerpted directly from the IDC Multimedia White Paper, "As the Economy Contracts, the Digital Universe Expands," May 2009, sponsored by EMC. The multimedia content can be viewed at http://www.emc.com/digital_universe

Video transcription for John Gantz

For most of us sitting here in the middle of a global economic crisis there are only two numbers that matter, and both are mind-numbing. The first is the amount of money being spent to bail out the banks and get the economy going again, and the second is the amount of money we've lost in our 401(k)s, pension funds, or home equity. Ouch.

But we would like to tell you about some other mind-boggling numbers. How about this one: 3,892,179,868,480,350,000,000?

That number is the number of bits added to what we call the Digital Universe in 2008. 8 bits is a byte, a million bytes a megabyte, and million megabytes a terabyte, and so on. 487 exabytes, or 487 billion gigabytes, in 2008.

Here is one more mind boggling number. Five. In 2012, there will be five times as many bits created or captured and added to the Digital Universe as in 2008.

Like the physical universe, the Digital Universe is expanding. In fact, exploding.

Unfortunately, thanks to the economic crisis, the technology universe - the software tools, the new techniques and business practices, and the people - is not expanding. At least in comparison to the Digital Universe.

For three years now, IDC has been adding up the number of bits pumped into the Digital Universe each year.

We come up with the number by taking all the digital devices and applications IDC tracks - from digital cameras and supercomputers to emails and web searches, from

CAT scanners and RFID readers to the digital packets that contain the chopped up "ums" and "ahs" in your voice phone-calls and multiplying the number of those things times the amount of information they create or capture on an annual basis.

We have learned a few things over the years. Of course the size and growth of the Digital Universe remains its most enduring characteristic. But there are some other notable facts:

In 2007, the amount of digital information created in a year surpassed, for the first time, the amount of storage to deal with it. Of course we don't need to store all the bits created - like digital TV signals, phone-call routing information, or old email spam. But if we wanted to, we couldn't.

The Digital Universe is incredibly diverse. The objects in it range from two-hour digital movies to the tunes of "Happy Birthday" in a greeting card. In our physical universe, 98% of the known mass is invisible as interstellar dust or what scientists call "dark matter." In the Digital Universe we have our own dark matter - the tiny signals from sensors and RFID tags, and voice packets on the telephone lines, that make up less than 2% of the Digital Universe by gigabyte but account for more than 98% of the "units" or information "containers" in it.

Here's another gem. For all the information we create as individuals - taking pictures, updating our Facebook profile, sending emails, watching HDTV programs - an amount greater than that is created or captured about us in the Digital Universe. Health records and images, surveillance photos, transaction records, and credit reports. We call this our digital shadow.

We've also discovered that while more than 70% of the information in the Digital Universe is created by individuals - consumers and information workers toiling away far from the datacenter - at some point in time, enterprises have responsibility or liability for 85%. Think of Google getting sued by Viacom for a billion dollars for the copyright infringements of users posting content on YouTube, or NASA being forced to disclose personal emails as part of a criminal investigation of an astronaut love triangle.

This liability leads us to another conclusion. That the migration to software-as-a-service and eventually full cloud computing - where organizations take advantage of shared resources off premises - may put parts of an enterprise's Digital Universe in the hands of third parties who may be better equipped to manage, store, and secure that information - but it won't take the responsibility for that information away from the enterprise.

And this year we learned that an economic crisis doesn't stop the tsunami. And that's the main theme in this Multimedia White Paper. The impact of the economic crisis on the growth of the Digital Universe - which is no impact, unless to speed it up - and what it means for you.

The Economic Crisis and the Digital Universe

We are in a global economic crisis, and everything seems to have fallen – the stock markets, home prices, bank assets, and consumer confidence.

About the only growth rate that *hasn't* gone negative since the recession began is the creation of new digital information. People are still taking pictures, making phone calls, sending emails, blogging, and putting up videos on YouTube. Enterprises are still capturing daily transaction records, adding to their data warehouses. Governments are still requiring *more* information be kept and protected, forcing the migration to digital TV, and taking surveillance photos of their citizens.

In fact, since the economic crisis got hot and heavy in late 2008, we think both the amount and pace of information passed over the Internet, phone networks, and airwaves actually increased. The amount of information created in 2008 came in ahead of what we forecast last year by 16,000,000 gigabytes, or 3%.

This is a problem. While digital content is up, IT budgets are down. IDC forecasts that worldwide spending on servers and storage, where the heavy lifting of information management and storage takes place, will be down 6% in 2009.

In fact, over the next four years, while the Digital Universe grows by a factor of almost five, total IT budgets worldwide will only grow by a factor of 1.2, and IT staff by a factor of 1.1.

The economic crisis will create other challenges for those who must face the tsunami on behalf of their enterprises:

- The economic crisis will speed the migration to virtualization technologies. The number of servers installed over the next four years – if you count virtual servers – will double. That's not only more information-creation capacity, but redundant software (operating systems, applications software) replacing redundant hardware and creating more for IT organizations to manage.
- Meanwhile, other information-generation technologies and trends will continue, from the growth of mobile users by a factor of three, to the growth, by a factor of 3.6, of nontraditional IT devices, like wireless meters, automobile-navigation systems, industrial machines, RFID readers, and intelligent-sensor controllers.
- The economic crisis seems to be having no effect on the growth of Web 2.0 technologies in homes and businesses. IDC forecasts that the number of interactions between people via email, messaging, social networks, etc., will grow by a factor of 8 in the next 4 years.
- Most economic stimulus of most governments in the world actually call to *increase* the amount of information to be created as they are implemented,

from more access to broadband communications, to electronic patient records, smart electric grids, and smart buildings and autos. Machine-to-machine communications, like those on sensor networks, don't actually create much digital information (except video surveillance) because the signals are so small, but they *do* create information packages that must be managed, kept secure, and stored.

- The financial crisis will clearly lead to more regulation and government oversight, which, in a replay of Sarbanes-Oxley, will drive more mandated record-keeping compliance, and, hence, more digital information.

There are tools, services, and techniques that enterprises can use to deal with the challenges created by the growth of their own digital universes, but it will take organizational discipline to deal with them.

The Digital Universe and the Future of Storage

We used to think 100 megabytes of storage was more than we'd ever need. Then it was a gigabyte; now, it seems even a terabyte of storage (or 1,000 gigabytes) is not going to be enough to satiate our storage requirements at home, and certainly within our businesses, where thousands of terabytes of storage easily consume a datacenter's square footage.

There are basically two ways to increase datacenter capacity: install more drives or install larger capacity drives in place of existing smaller drives (a third, and supposedly common, alternative is to do both). Hard disk drive (HDD) manufacturers, and consequently storage system OEMs, continue to enable customers to do any of the options previously listed. However, while disk drive capacity continues to increase, it noticeably has slowed down over the past few years. This has created quite an issue for IT managers whose storage requirements continue, near unabated.

The natural conclusion is that customers are faced with increasing the number of storage arrays and the number of disk drives within their datacenters to keep up with capacity requirements.

To complicate things, not only are IT managers dealing with limited square footage, but also with limited power and cooling capacity. Every spinning disk consumes more energy and requires more cooling. IDC estimates that in 2008, an average datacenter disk drive cost US$36.29 to power and cool for a year.

One of the strategies to combat this increasing power consumption, and to manage costs in this economic downturn, is to move to more efficient platforms and to integrate more efficient storage technologies.

Server consolidation and virtualization has gone a long way in preserving precious datacenter floor space, as well as reducing significantly the costs to power and cool

servers. However, the consolidation of servers, as well as the virtualization of servers, has placed added stress on external storage, or storage area networks (SAN), to which these new aggregations of servers now attach. New storage typically is carved out of the SAN with every new virtualized server. This dynamic provides opportunities for bringing more efficiency to storage environments.

If servers can benefit from virtualization, can storage? The answer is a resounding "yes." Storage virtualization strategies allow the holistic use of storage, regardless of the type of data, type of attach (e.g., FC, iSCSI, etc.), or brand or performance of storage. The idea is to leave no byte stranded so that utilization rates can be increased from sub-40% in most of today's datacenters to over 60% without incurring any negative issues.

Once the data is housed in a common environment, the opportunities to remove duplicated data (which can consume a good deal of capacity) become quite compelling. Data deduplication can help conserve storage capacity in an IT or datacenter environment. This conservation helps to stave off the purchase of new storage capacity and also improves services like replication and backup by thinning the amount of data that needs to be managed.

However, the growth of data, and specifically unstructured data, has the real potential to introduce information management challenges. Some of the largest issues with this growing amount of unstructured data is not about how to store it, but instead, how to classify, search, and secure it… efficiently.

File system management software is one of the fastest-growing areas of interest as customers search for ways to manage their unstructured data. Managing unstructured data creates new challenges and opportunities. There are a number of technologies that can add value to a file-based repository. Areas where investments in technologies or partnerships will pay dividends include:

- Capacity optimization achieved through compression, deduplication, and tiering with file virtualization
- Security for access permissions delivered via strong authentication, identity management, or/and encryption
- Advanced information organization, management and security with data classification and indexing services and data loss prevention
- Datacenter and network infrastructure design and support to ensure the scalability and efficiency of the content repository
- Robust geographic dispersal through network acceleration, intelligent load balancing across geographies, and lower network latency

Challenges for the Enterprise

The Digital Universe will double every 18 months. It is probably safe to say that the challenges for enterprise IT organizations will more than double every year.

Consider just the matter of information security. As we pointed out earlier, *most* of the information IT organizations will need to keep secure is created outside the datacenter, and even outside the company. And more and more of that information originates from mobile users – workers, customers, suppliers, partners – adding another layer of complexity to the security equation.

To be specific, over the next four years, 600 million more people will become active on the Internet, and nearly two-thirds of all Internet users will utilize mobile devices at least some of the time. By 2012, there will be 850 million people buying and selling things on the Internet – all the more names, personal information, and transactions to keep secure – and twice as much Internet commerce taking place. By 2012, Internet commerce will reach $13 trillion, mostly in the form of sensitive business-to-business commerce.

Obviously, not all information in the Digital Universe needs to be made secure – like downloads of "American Idol" episodes to your personal digital video recorder– but much of it will be. And as increasingly-sophisticated cybercriminals find more ways to exploit information, some forms of content we don't consider sensitive today (such as personal blogs, Facebook data, and the like) will soon require higher levels of security or protection.

In fact, IDC has estimated that roughly a third of information created today is "security intense," meaning it requires high standards of protection, as, for instance, that "American Idol" file, which does not. What's more, we think that percentage will grow to closer to 45% by the end of 2012.

Security won't be the only challenge for IT organizations. The rules about what information to store and which to provide access to regulating authorities and auditors will grow more complex as a result of the economic crisis, making the total information that comes under the "compliance" umbrella grow to 35% of the Digital Universe.

Compliance and market pressures will also require some long-term strategies for storage and archiving – perhaps not to the 100-year standard of the movie industry, but certainly for 10 years or more. Here, too, the amount of preservation-intense information will grow faster than the Digital Universe as a whole.

If these are not pressing enough problems facing IT departments, there are some other trends that will add heat to the cauldron.

- The growth of the numbers and types of devices that aren't the traditional enterprise PCs, servers, storage systems, and network equipment, will drive changes in network and datacenter architecture and management. Where

today, most corporate computing traffic on networks is from the server to the client, more and more devices reporting in from the network edge will be reversing that trend. They will also be sending in much more diverse signals – voice packets, minutes of video surveillance, and sensor signals that need to be dealt with immediately. All, of course, need security, management, and storage, at least for a period of time.

- Along with these information types, IT departments will be inheriting the operational systems that generate them. Responsibility for physical security – guards, building access, perimeter monitoring, and video surveillance – has already started to merge with information security. Building automation systems, once managed on proprietary systems, will more and more run their control information over IT's IP networks. IDC expects that as more and more corporate systems – from factory floor automation and vehicle tracking to intelligent grid communications – roll onto IP networks, IT organizations will pick up responsibility for them.
- The migration toward software-as-a-service and cloud computing – where organizations take advantage of shared resources off premises – will also put information-related pressures on IT organizations. Although third parties may often be better equipped to manage, store, and secure information than their customers, the *responsibility* will still remain with the IT organization. Before they can take advantages of these new virtualization-base technologies, they will require IT practices, skills, and applications that are ready.
- The need to expand and accelerate training of existing IT staff as technology becomes more complex and skilled IT staff are lost to attrition and retirement.

The technology transformations under way before the economic crisis – the adoption of Web 2.0, migration to real-time business analytics, the move to virtualization – makes it clear that the number one critical success factor for IT organizations in the future would be speed: The ability to rapidly deploy new applications, timely integration of acquisitions, and flexibility and transparency in dealing with business units.

The economic crisis just turned up the voltage on the treadmill.

CIO Call to Action

The economic crisis, the growth of the Digital Universe, and the challenges already presented in this iView make it clear that CIOs and IT staff face some thorny issues. The laundry list of things to deal with is scary:

- Security
- Privacy
- Compliance
- eDiscovery
- Unstructured data

- Information classification
- Information management
- Information search, discovery
- Preservation
- Mobility
- Web 2.0
- New form factors, operational systems
- Budget constraints

And this is not to mention technical and departmental issues like changing architectures, performance worries, IT's own costs, and the lack of skilled IT personnel.

But in each item listed above, it soon becomes clear that IT is not alone. Dealing with each of these issues involves the business units, and deep commitment at that. Security means not just installing the latest security patches from Microsoft, but implementing data-loss prevention software and training employees not to leave sensitive data on their laptops. Information classification means agreement on which information is most important – and having enterprise-wide agreement, or at least top management agreement. Compliance means IT efforts integrated with the efforts of finance and legal departments, as well as outside auditors. Taking on new operational systems, like the phones or building automation, means making sure the business units understand the impact these thankless tasks will have on other areas of IT.

IDC believes that for IT organizations to face such overwhelming challenges, three actions are imperative:

- First. IT organizations will need to transform existing relationships with the business units. Simply having good service-level agreements won't be enough. These are the groups, when all is said and done, with the responsibility of dealing with customers if data is lost or breached, of setting information access policy, and of funding new initiatives. Leading-edge IT organizations are doing whatever they can to fuse their departmental DNA with that of the business units, from job rotation for employees in both camps and tying IT performance to business metrics, to sending IT staff out to meet customers and funding internal PR efforts.
- Second. Spearhead the development of organization-wide policies for information governance, including security, information retention, data access, and compliance. Extend these policies to business partners. Force the organization to mandate ongoing training. This includes educating business unit managers on areas IT might traditionally stay away from, such as the importance of reputation management software, competitive practices using business intelligence, and the brand risks from unmanaged employee

blogging or chaotic deployment of social networks department by department (as well as the brand risks for *not* using the new tools to reach customers).

- Third. Rush new tools and standards into the organization. Think of storage replication, deduplication, unstructured search, database analytics, ethical hacking, application virtualization, semantic web applications, and cloud computing.

The Digital Universe expands onward, economic crisis or not. Keeping up will be a challenge. But meeting the challenge will be an opportunity to differentiate your company and yourself from others.

The last time we saw a confluence of two such powerful trends – an explosion of new and potentially disruptive technologies and an economic crisis of this magnitude – was before the computer was invented. The challenge of a lifetime is also the chance of a lifetime.

The Digital Universe: A Snapshot

Like our own universe, the Digital Universe is both large beyond comprehension, expanding, and diverse.

At nearly 500 exabytes – or 500 billion gigabytes – the Digital Universe, if converted to pages of text and assembled into books, would stretch to Pluto and back 10 times. You could wrap the entire earth 8 times with all that paper.

At the current growth rate, that stack of books is growing 20 times faster than the fastest rocket ever made, the Atlas V that powered the NASA Pluto New Horizons spacecraft that left Patrick Air Force Base on January 19, 2006. It took that spacecraft 13 months to reach Pluto. It would take our growing stack of books 3 weeks.

But there is still really no way to explain how big the Digital Universe is, just as it's impossible to visualize Avogadro's number (which is bigger) or the number of stars in the universe (which is smaller).

It may be a little easier to explain the diversity of the Digital Universe, which contains everything from 30-minute digital TV shows coming in at 2 terabits a half hour to 128 bit RFID signals. In fact, the large files from digital and surveillance camera images account for the vast majority of the Digital Universe, in fact 88% of it. But by far, the number of "units" of information, sensor signals, RFID reads, and voice packets dominate, accounting for 98% of the total.

The Digital Universe is also messy. Because the image bits account for so much of the total, more than 95% of the data in the Digital Universe is unstructured, meaning its intrinsic meaning cannot be easily divined by simple computer programs. There are ways to imply meaning to unstructured data, and the Semantic Web project is

promising to develop the tools to help us do that in the future. But today, as anyone who has tried to manage his or her online photos knows, searching for the palm-tree-and-sunset picture in a digital stack of 2,000 digital vacation photos can be next to impossible.

Perhaps the final characteristic of the Digital Universe is its own economic footprint.

Add up how much money is spent on computers, storage, networks software, services and telecommunications, and IT and telecommunications staff and you get $6 trillion. Toss in medical equipment, entertainment and content creation, and electronic transaction service fees, and you can double that.

The Digital Universe is very large and continues to grow at a rapid pace.
