IDC
Analyze the Future

I D C I V I E W

Extracting Value from Chaos

June 2011

By John Gantz and David Reinsel

Sponsored by EMC Corporation

Content for this paper is excerpted directly from the IDC iView "Extracting Value from Chaos," June 2011, sponsored by EMC. The multimedia content can be viewed at http://www.emc.com/digital_universe.

State of the Universe: An Executive Summary

As we mark the fifth anniversary of our annual study of the digital universe, it behooves us to take stock of what we have learned about it over the years. We always knew it was big – in 2010 cracking the zettabyte barrier. In 2011, the amount of information created and replicated will surpass 1.8 zettabytes (1.8 trillion gigabytes) - growing by a factor of 9 in just five years.

But, as digital universe cosmologists, we have also uncovered a number of other things — some predictable, some astounding, and some just plain disturbing.

While 75% of the information in the digital universe is generated by individuals, enterprises have some liability for 80% of information in the digital universe at some point in its digital life.

The number of "files," or containers that encapsulate the information in the digital universe, is growing even faster than the information itself as more and more embedded systems pump their bits into the digital cosmos. In the next five years, these files will grow by a factor of 8, while the pool of IT staff available to manage them will grow only slightly.

Less than a third of the information in the digital universe can be said to have at least minimal security or protection; only about half the information that should be protected is protected.

The amount of information individuals create themselves — writing documents, taking pictures, downloading music, etc. — is far less than the amount of information being created about them in the digital universe.

The growth of the digital universe continues to outpace the growth of storage capacity. But keep in mind that a gigabyte of stored content can generate a petabyte or more of transient data that we typically don't store (e.g., digital TV signals we watch but don't record, voice calls that are made digital in the network backbone for the duration of a call).

So, like our physical universe, the digital universe is something to behold — 1.8 trillion gigabytes in 500 quadrillion "files" — and more than doubling every two years. That's nearly as many bits of information in the digital universe as stars in our physical universe.

However, unlike our physical universe where matter is neither created nor destroyed, our digital universe is replete with bits of data that exist but for a moment — enough time for our eyes or ears to ingest the information before the bits evaporate into a nonexistent digital dump.

This is not to diminish the value of the temporary existence of these bits that can serve a variety of purposes during their short lives, such as driving consumption (to increase ad revenue from Web site traffic) or real-time data analytics (to optimize existing operations and create entirely new markets).

What are the forces behind the explosive growth of the digital universe? Certainly technology has helped by driving the cost of creating, capturing, managing, and storing information down to one-sixth of what it was in 2005. But the prime mover is financial. Since 2005, the investment by enterprises in the digital universe has increased 50% — to $4 trillion. That's money spent on hardware, software, services, and staff to create, manage, and store — and derive revenues from — the digital universe.

In an information society, information is money. The trick is to generate value by extracting the right information from the digital universe — which, at the microcosmic level familiar to the average CIO, can seem as turbulent and unpredictable as the physical universe.

In fact, thanks to new tools and technologies, and new IT and organizational practices, we may be on the threshold of a major period of exploration of the digital universe. The convergence of technologies now makes it possible not only to transform the way business is conducted and managed but also to alter the way we work and live.

Considerations

New capture, search, discovery, and analysis tools can help organizations gain insights from their unstructured data, which accounts for more than 90% of the digital universe. These tools can create data about data automatically, much like facial recognition routines that help tag Facebook photos. Data about data, or metadata, is growing twice as fast as the digital universe as a whole.

Business intelligence tools increasingly are dealing with real-time data, whether it's charging auto insurance premiums based on where people drive, routing power through the intelligent grid, or changing marketing messages on the fly based on social networking responses.

New storage management tools are available to cut the costs of the part of the digital universe we store, such as deduplication, auto-tiering, and virtualization, as well as to help us decide what exactly to store, as in content management solutions.

An entire industry has grown up to help us follow the rules (laws, regulations, and customs) pertaining to information in the enterprise. It is now possible to get regulatory compliance systems built into storage management systems.

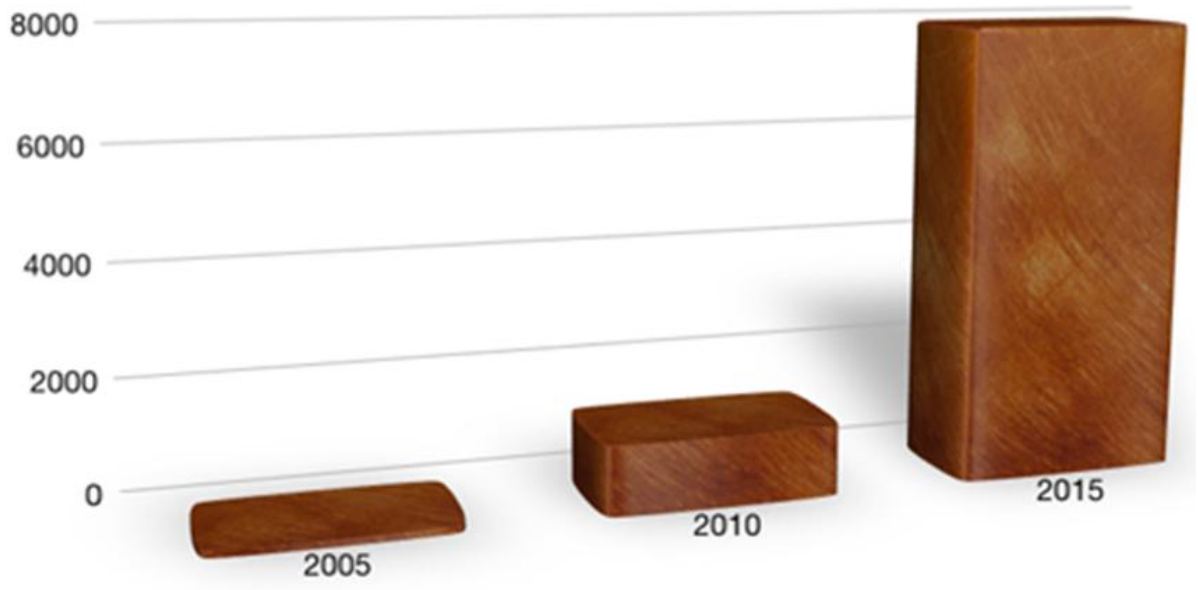
New security practices and tools can help enterprises identify the information that needs to be secured and at what level of security and then secure the information using specific threat protection devices and software, fraud management systems, or reputation protection services.

Cloud computing solutions — both public and private and a combination of the two known as hybrid — provide enterprises with new levels of economies of scale, agility, and flexibility compared with traditional IT environments. In the long term, this will be a key tool for dealing with the complexity of the digital universe (see Figure 1).

Cloud computing is enabling the consumption of IT as a service. Couple that with the "big data" phenomenon, and organizations increasingly will be motivated to consume IT as an external service versus internal infrastructure investments.

Figure 1

A Decade of Digital Universe Growth: Storage in Exabytes

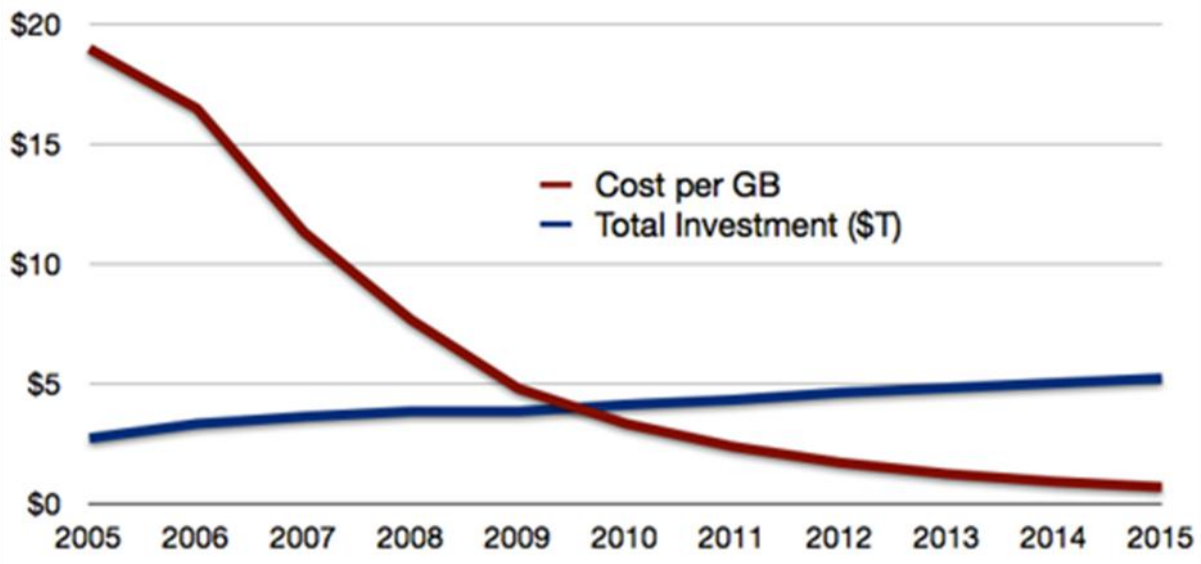


Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

This period of "space exploration" of the digital universe will not be without its challenges. But for the "astronauts" involved — CIOs and their staff—– it represents a unique, perhaps once-in-a-career opportunity to drive growth for their enterprises. They will need to lead the enterprise in the adoption of new information-taming technologies, best practices for leveraging and extracting value from data, and the creation of new roles and organizational design. Each step will require organizational change, not just a few new computers or more software. The success of many enterprises in the coming years will be determined by how successful CIOs are in driving the required enterprisewide adjustment to the new realities of the digital universe (see Figure 2).

Figure 2

The Digital Universe Growth Paradox: Falling Cost and Rising Investment



Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

Journey to the Cloud

As the digital universe expands and gets more complex, processing, storing, managing, securing, and disposing of the information in it become more complex as well.

Consider this: Over the next decade, the number of servers (virtual and physical) worldwide will grow by a factor of 10, the amount of information managed by enterprise datacenters will grow by a factor of 50, and the number of files the datacenter will have to deal with will grow by a factor of 75, at least. Meanwhile, the number of IT professionals in the world will grow by less than a factor of 1.5.

As a result, the skills, experience, and resources to manage all these bits of data will become scarcer and more specialized, requiring a new, flexible, and scalable IT infrastructure, extending beyond the enterprise. Today we call it cloud computing.

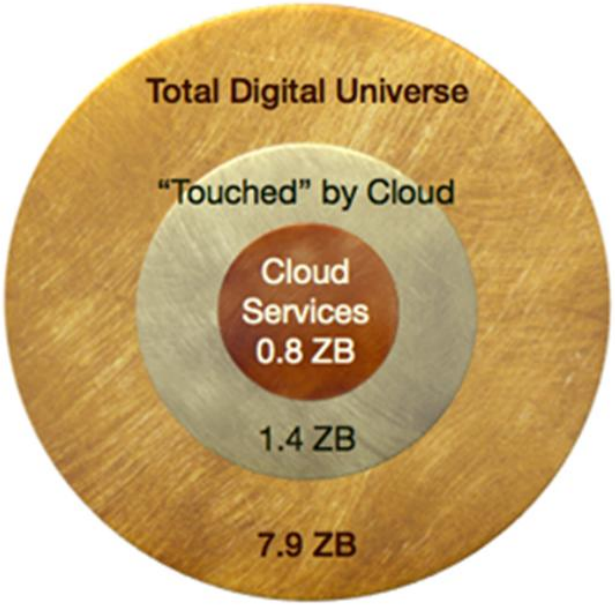
And while cloud computing accounts for less than 2% of IT spending today, IDC estimates that by 2015 nearly 20% of the information will be "touched" by cloud computing service providers — meaning that somewhere in a byte's journey from originator to disposal it will be stored or processed in a cloud. Perhaps as much as 10% will be maintained in a cloud.

Much of the current movement to cloud architectures is being enabled by pervasive adoption of virtualization. Last year was the first year in which more virtual servers were shipped than physical servers. IDC estimates that today nearly 10% of the information running through servers is doing so on virtualized systems and expects that number to grow to more than 20% in 2015. This percentage increases along with the size of the organization. Some larger environments today operate with 100% virtualized systems.

Of course, cloud services come in various flavors — public, private, and hybrid. For organizations to offer their own cloud services, they have to do more than just run virtual servers. They must also allow for virtualized storage and networking, self-provisioning, and self-service — and provide information security and billing. Few enterprises are here yet, so the impact of private clouds on the digital universe today is small (see Figure 3). But by 2015, when the virtualized infrastructure is more common, the rate of growth will accelerate.

Figure 3

The Digital Universe and the Public Cloud, 2015



Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

The challenges for cloud adoption include:

- Data preparation for conversion to cloud
- Integrated cloud/noncloud management
- Service-level agreements and termination strategies
- Security, backup, archiving, and disaster control strategies
- Intercountry data transfer and compliance
- Organizational politics

The latter is nontrivial. Many of the most successful virtualization projects succeed, in part, because CIOs have developed opt-in and opt-out strategies for internal departments that may be reluctant to share responsibility for what was once considered "their" information or data. Converting to cloud computing means changing the status quo — always a difficult task, even if there are good reasons to do so.

Big Value from Big Data

Big data is a big dynamic that seemed to appear from nowhere. But in reality, big data isn't new. Instead, it is something that is moving into the mainstream and getting big attention, and for good reason. Big data is being enabled by inexpensive storage, a proliferation of sensor and data capture technology, increasing connections to information via the cloud and virtualized storage infrastructures, and innovative software and analysis tools. Big data is not a "thing" but instead a dynamic/activity that crosses many IT borders. IDC defines it this way:

> *Big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.*

Big data is a horizontal cross-section of the digital universe and can include transactional data, warehoused data, metadata, and other data residing in ridiculously large files. Media/entertainment, healthcare, and video surveillance are obvious examples of new segments of big data growth. Social media solutions such as Facebook, Foursquare, and Twitter are the newest new data sources. Essentially, they have built systems where consumers (consciously or unconsciously) are providing near continuous streams of data about themselves, and thanks to the "network effect" of successful sites, the total data generated can expand at rapid logarithmic rates.

It is important to understand that big data is not only about the original content stored or being consumed but also about the information around its consumption. Smartphones are a great illustration of how our mobile devices produce additional data sources that are being captured and that include geographic location, text messages, browsing history, and (thanks to the addition of accelerometers and GPS) even motion or direction (see Figure 4).

Figure 4

Big Data Is Not the Created Content, nor Is It Even Its Consumption — It Is the Analysis of All the Data Surrounding or Swirling Around It

Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

Capture and analysis of this "swirling vortex" of data is a definite big data opportunity but also a source of consternation for datacenter managers. Datacenter architectures and organizational models will need to evolve as big data applications pervade a company's infrastructure. The IT architectural and organizational approach used in clustered environments like a large Hadoop grid is radically different from the converged and virtualized IT environments driving most organizations' datacenter transformation strategies.

Big data will inject high-velocity requirements associated with capture and analysis, as well as results/predictive reporting. With big data, IT is best organized around the specific opportunity and/or capability rather than merely a set of shared services that serve both traditional and newer uses. Most IT disciplines — from infrastructure to applications to governance — are ideally part of a single integrated team and work closely with users of big data in ways that are very distinct from traditional enterprise IT approaches.

The cloud providers will play a key enabling role in nearly every facet of the big data space. First, they will be among the most important collectors of data streams and content. Second, they will be among the most aggressive users of big data systems to run their own businesses. Third, they will also be in a position to enable big data use by technically savvy, but resource constrained, organizations (through simple, temporary provisioning of large compute and data pools). For example, cloud-based big data platforms will make it practical for smaller engineering and architectural firms to access massive compute resources for short, semipredictable time periods without having to build their own big data farms.

The ultimate value of a big data implementation will be judged based on one or more of three criteria:

- Does it provide more useful information?
- Does it improve the fidelity of the information?
- Does it improve the timeliness of the response?

Netflix suggesting your next movie rental, dynamic monitoring of embedded sensors in bridges to detect real-time stresses and longer-term erosion, and retailers analyzing digital video streams to optimize product and display layouts and promotional spaces on a store-by-store basis are a few real examples of how big data is involved in our lives today.

Big data represents both big opportunities and big challenges for CIOs. Almost every CIO aspires to make IT a more valued asset to the organization. And IT is front and center in big data projects, which are typically at the boundaries of the business where many of the most significant business expansion or cost reduction opportunities lie.

Big data also poses a number of challenges. As noted earlier, big data buildouts can be disruptive to current datacenter transformation plans. In addition, big data deployments require new IT administration and application developer skill sets. People with these skills are likely to be in short supply for quite a while. The biggest challenge, however, is the cultural challenge. Today, many of these big data projects are best described as "junior science projects" with a small core of servers and storage assets. However, unless managed closely, these small projects can quickly turn into the next "Manhattan project" with companywide and industrywide business, organizational, and legal consequences.

Challenges aside, the opportunities abound. Taking a lead in big data efforts provides the CIO with an opportunity to be the most significant strategic partner for a business unit or even drive a transformation of the entire enterprise.

Trust in the Digital Universe

Last year, for the first time, we sized the amount of information in the digital universe that requires some level of security. The frightening realization is that the amount of information that needs to be secured is growing faster than our ability to secure it as employees leverage more mobile devices, consumers knowingly (and unknowingly) share more personal data, and companies find new ways to mine this data.

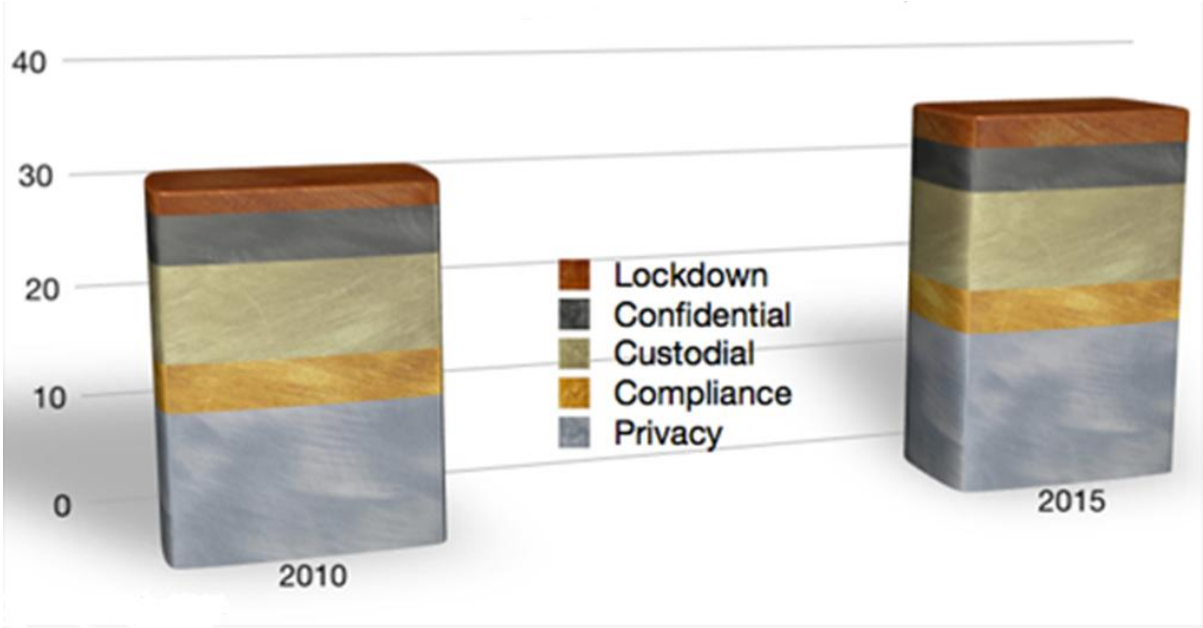
For the sake of understanding the degree of security in the digital universe, we have classified information that requires security into five categories, each requiring successively higher levels:

- **Privacy** only — such as an email address on a YouTube upload
- **Compliance** driven — such as emails that might be discoverable in litigation or subject to retention rules
- **Custodial** — account information, a breach of which could lead to or aid in identity theft
- **Confidential** — information the originator wants to protect, such as trade secrets, customer lists, confidential memos, etc.
- **Lockdown** — information requiring the highest security, such as financial transactions, personnel files, medical records, military intelligence, etc.

In 2010, 28% of the digital universe required some level of security (see Figure 5). Note that this is information that *needs* security. It may not have it.

Figure 5

The Need for Information Security: Percentage of the Digital Universe



Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

The growing amount of information that requires security is driven largely from two main sources: corporations (including employees) and consumers. Compounding this are government regulations and company policies and processes that mandate the capture and preservation of various data.

More organizations are adopting policies where users have control over the laptops, tablets, and smartphones that they could use to conduct their business (or personal) affairs. Given that IT resources increasingly are drawn from a combination of private and public infrastructures, the pressure is on for these organizations to figure out a way to manage security and compliance obligations across multiple many-to-many relationships.

The notion of trust in how information is used, shared, archived, and managed is critical in this complex and highly fluid environment. Trust relates to the origin of the information; the integrity of the processes and the computing systems that generate, capture, and manage the information; as well as the credentials and identities of the individuals and business entities that touch or have access to the information.

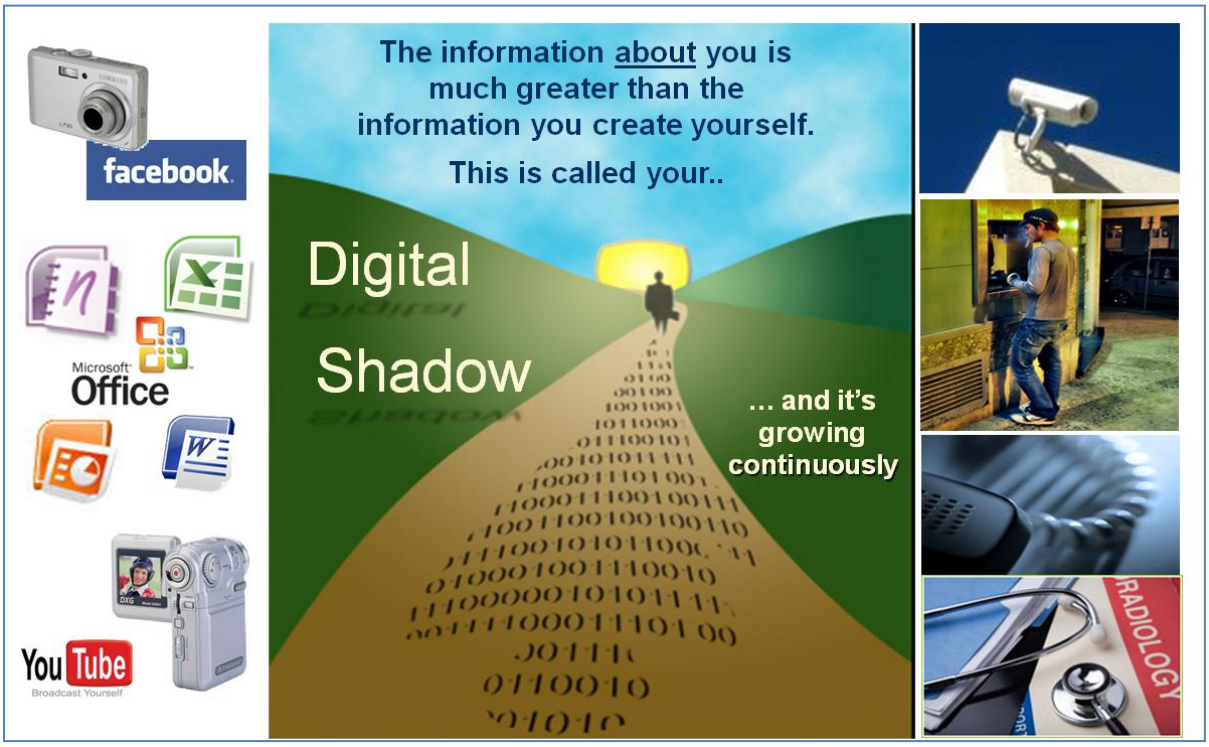
We are seeing this discussion around trust unfolding before us today. Online data collection is becoming more invasive, data mining analytics and big data make it possible for businesses to profile individual consumers, and individuals are expanding their digital shadow through their use of mobile device applications and their participation in social networking sites. As a result, there are increasing calls from advocates, academics, and regulators to amend the current privacy and data protection regimes.

A few years ago, we introduced the concept of the digital shadow (see Figure 6). This shadow is growing faster every year, and most of the time without our knowing it. Our digital shadow is made up of information we may deem public but also data that we would prefer to remain private. Yet, it is within this growing mist of data where big data opportunities lie — to help drive more personalized

services, manage connectivity more efficiently, or create new businesses based on valuable, yet-to-be-discovered intersections of data among groups or masses of people.

Figure 6

Fueling the Digital Universe: The Digital Shadow



Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

Organizations that manage their own private cloud, or those that provide public cloud services, must find ways to mitigate unnecessary or accidental exposure of secure information. Employees and consumers (individuals) cannot be trusted to do the right thing all the time. Instead, organizations must integrate policies and processes to create an environment of trust that is automated and thorough and not sustained by siloed and manual operations. The following specific steps are needed:

- Enterprises must understand the dependencies across processes and compute resources.
- Enterprises need to be able to identify potential policy conflicts, as well as analyze the impact of new technologies and processes on their overall risk and compliance posture.
- With IT organizations managing very complex network topologies and vendor relationships, they need to be able to correlate dependencies across events and changes to configuration standards across functional systems. Further, they need to be able to track dependencies across people, processes, information, and the underlying compute resources.
- Enterprises have to be able to demonstrate that they are consistently meeting their regulatory and legal obligations.

The convergence of these developments — mobile computing, consumerization of IT, cloud computing, big data, and advanced data mining technologies — is compelling many organizations to transition from a "chasing compliance" mindset to a risk management mindset. A risk management framework would allow an organization to tailor its responses to the criticality of the information, business process, and business relationships. Because resources are limited, a programmatic risk management framework also allows an organization to prioritize its resources so that it is able to identify and respond more quickly to the most critical threats to its security and compliance postures.

Call to Action

Since 2007, IDC's Digital Universe Study has highlighted the mismatch between the rapid growth of the digital universe and the very slow growth of staff and investment to manage it. This year, the study highlights an *additional* issue that promises to define for CIOs and business executives much of the next 10 years of their careers: the mismatch between the value of *some* data and the value of *other* data. In the digital universe, it is now possible to run factories from afar, tap vast stores of social networking traffic for meaning, analyze customers in efficient ways that were impossible even a few years ago, and create smart cities, buildings, and homes. But this requires sifting through the data molecules (while they exist) in the digital universe to identify the ones that matter and creating organizational value out of them.

And there's the rub. CIOs and their staff may be able to manage the new tools of search and discovery, information classification and management, information security, and even information disposal, but organizations have to be prepped and ready to deal with the nuggets that can now be pulled out of the digital universe.

So the call to action from this year's study has two components — one technical, one organizational.

The technical calls to action are:

- Investigate the new tools for creating metadata — the information you will need to understand which data is needed when and for what. Big data will be a fountain of big value only if it can speak to you through metadata.
- Master virtualization, not just server and storage virtualization but also application virtualization. Start working on self-provisioning and self-service, including metering and billing.
- Move what you can to the cloud — it's inevitable. But it will require a new level of commitment and rigor toward managing the process. It's more than simply an outsourcing contract.
- Determine which big data projects will have the most "bang," along with the requisite data sets and analytical tools. From there, formulate an enterprise data strategy that overcomes legacy data integration limitations and layers in the new required tools and techniques.
- Stay very close to the latest information security strategies and practices.
- Be aggressive in developing and managing advanced storage management tools.

The organizational calls to action are:

- Set the strategy and build, with other C-level executives, a process for migrating to shared resources — virtualization today, public and private clouds tomorrow. This will take leadership, coercion, cajoling, and politicking.
- Begin laying the organizational groundwork today with the specific analytical and managerial skill sets, mindsets, and processes necessary to extract the most value possible from big data.

- Work to make your company a data-driven organization. New tools — from smartphones and iPads to executive dashboards and real-time business intelligence — can make this easier, but the blessings of the digital universe will be lost on the organization that won't take action based on what the data tells it. If you don't, you can be sure that the competition will.
- Press your suppliers and business partners to help. We are still coming out of a global recession, but we are 10 years into a technology renaissance. Companies that aren't at least fast followers of those pushing the envelope of data and information utilization and management will simply fall behind.

The combination of post-recession business growth, a technology renaissance, and the growth of the digital universe this next decade creates a once-in-a-career opportunity for CIOs and their staff to drive change and growth for their organizations. The growth of the digital universe may be a challenge, but it is also a propellant for new and exciting uses of data.

A B O U T T H I S P U B L I C A T I O N

This publication was produced by IDC Go-to-Market Services. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Go-to-Market Services makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com