



WHITE PAPER

The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things

Sponsored by: EMC Corporation

Vernon Turner
David Reinsel
April 2014

John F. Gantz
Stephen Minton

Content for this paper is excerpted directly from the IDC iView " The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things, " April 2014, sponsored by EMC. The multimedia content can be viewed at <http://www.emc.com/leadership/digital-universe/index.htm>.

EXECUTIVE SUMMARY: DATA GROWTH, BUSINESS OPPORTUNITIES, AND THE IT IMPERATIVES

This is the digital universe. It is growing 40% a year into the next decade, expanding to include not only the increasing number of people and enterprises doing everything online, but also all the "things" – smart devices – connected to the Internet, unleashing a new wave of opportunities for businesses and people around the world.

Like the physical universe, the digital universe is large - by 2020 containing nearly as many digital bits as there are stars in the universe. It is doubling in size every two years, and by 2020 the digital universe – the data we create and copy annually – will reach 44 zettabytes, or 44 trillion gigabytes.

Like the physical universe, it is diverse – created by everyone using a digital camera, by the more than 2 billion people and millions of enterprises living their lives and doing their work online, and by the millions of sensors and communicating devices sending and receiving data over the Internet. It includes Oscars-host Ellen DeGeneres' "celeb selfie" tweet that was viewed 26 million times across the Web during a 12-hour period; the more than one billion hours of TV shows and movies streamed from Netflix per month; the data collected by sensors connected to a giant gas turbine and its analysis, making electricity cheaper and cleaner; and the data streaming at 2.8 Gigabytes per second from the Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope.

But unlike the physical universe, the digital universe is created and defined by software, a man-made construct. It is defined by software that analyzes this ever-expanding universe of digital data, finding the hidden value and new opportunities to transform and enhance the physical world - keeping the Mars Rover roving, shipping money, or storing the pictures of our loved ones. And it is software that will both create new opportunities and new challenges for us as we try to extract value from the digital universe that we have created.

Hence the theme for this year's digital universe study: The expanding universe of business opportunities. As more of the world goes online, including the physical world, the more opportunity there is for enterprises (and consumers) to use data in new ways - to learn about customers, speed business cycles, flatten organizational structures, and transform themselves into companies designed for what we call the "Third Platform" of the digital age, a platform built upon a foundation of cloud computing, mobility, social networking, and Big Data.

To maximize that opportunity, though, requires following certain imperatives for IT organizations. Information security, for example, virtualized datacenters, seamless public and private cloud computing, next- generation analytics, new storage management technologies, new data access tools and processes, automatic tagging, and the ability to deal with real-time data. And IT organizations will have to help their enterprises become more data- and software-driven.asdfasdf

Today, the digital universe has reached a number of new thresholds: The data coming from embedded systems (e.g., MP3 players, traffic lights, MRI scanners) has grown to a level where it's starting to challenge established practices in datacenters; the migration to digital entertainment – movies and TV – is almost complete; and metadata (e.g., the data added to your email message describing when it was created), once tightly coupled with the data it describes, has grown into a category in and of itself, the fastest-growing subcategory of the digital universe.

This year's iteration of the study is based on IDC's estimates of the data generated by more than 40 types of devices, from RFID tags and sensors to supercomputers and supercolliders, from PCs and servers to cars and planes. Here is a summary of the new findings:

- From 2013 to 2020, the digital universe will grow by a factor of 10 - from 4.4 trillion gigabytes to 44 trillion. It more than doubles every two years.
- Between 2013 and 2020 the division of the digital universe between mature and emerging markets (e.g., China) will switch - from 60% accounted for by mature markets to 60% of the data in the digital universe coming from emerging markets.
- In 2013, two-thirds of the digital universe bits were created or captured by consumers and workers, yet enterprises had liability or responsibility for 85% of the digital universe.
- In 2013, only 22% of the information in the digital universe would be a candidate for analysis, i.e., useful if it were tagged (more often than not, we know little about the data, unless it is somehow characterized or tagged – a practice that results in metadata); less than 5% of that was actually analyzed. By 2020, the useful percentage could grow to more than 35%, mostly because of the growth of data from embedded systems.
- Of the useful data, IDC estimates that in 2013 perhaps 5% was especially valuable, or "target rich." That percentage should more than double by 2020 as enterprises take advantage of new Big Data and analytics technologies and new data sources, and apply them to new parts of the organization.
- In 2013, while about 40% of the information in the digital universe required some type of data protection, less than 20% of the digital universe actually had these protections.
- Data from embedded systems, the signals from which are a major component of the Internet of Things, will grow from 2% of the digital universe in 2013 to 10% in 2020.

- In 2013, less than 20% of the data in the digital universe is "touched" by the cloud, either stored, perhaps temporarily, or processed in some way. By 2020, that percentage will double to 40%.
- Most of the digital universe is transient - unsaved Netflix or Hulu movie streams, or Xbox One gamer interactions, temporary routing information in networks, sensor signals discarded when no alarms go off, etc. - and it is getting more so. This is a good thing, because the world's amount of available storage capacity (i.e., unused bytes) across all media types is growing slower than the digital universe. In 2013, the available storage capacity could hold just 33% of the digital universe. By 2020, it will be able to store less than 15%.
- In 2014, the digital universe will equal 1.7 megabytes a minute for every person on Earth.

These estimates encapsulate the opportunities, challenges, and paradoxes of the digital universe. There is a lot of valuable data in the digital universe, but it will take determination and skilled workforce to find and put to use. It will need to be protected, analyzed, and acted upon.

There is an abundance of technical solutions, and successful early adopters. But organizations must adapt - and adapt fast, given that the digital universe more than doubles every two years. The foundation of that adaptation will be in the datacenter, but the rest of the organization, the one filled with people, tradition, culture, and habits, must also adapt.

THE INTERNET OF THINGS: DATA FROM EMBEDDED SYSTEMS WILL ACCOUNT FOR TEN PERCENT OF THE DIGITAL UNIVERSE BY 2020

There have been three major growth spurts for the digital universe in modern memory. The first was when digital camera technology replaced film; the second, when analog telephony went digital; and the third, when TV went digital.

Now comes a fourth growth spurt -- the migration of analog functions monitoring and managing the physical world to digital functions involving communications and software telemetry.

Call it the advent of the Internet of Things (IoT). Fed by sensors soon to number in the trillions, working with intelligent systems in the billions, and involving millions of applications, the Internet of Things will drive new consumer and business behavior that will demand increasingly intelligent industry solutions, which, in turn, will drive trillions of dollars in opportunity for IT vendors and even more for the companies that take advantage of the IoT.

Why the IoT heralds a new era of computing is a matter of math. All earlier eras involved the computerization of enterprises or people, of which there are a finite number on the planet. This era involves the computerization, adding software and intelligence, to things - things as varied as cars and toys, airplanes and dishwashers, turbines and dog collars.

Yes, there is a finite number of things - at least big things - that might be computerized. But, by IDC's count, that number is already approaching 200 billion. And the number of sensors (e.g., the accelerometer in your smart phone) that track, monitor, or feed data to those things is already more than 50 billion, with scientists talking about trillion-sensor networks within 10 years.

Of course, not all of those 200 billion things are actually wired and communicating on the Internet, but some 14 billion are. And, by 2020, this number will grow to 32 billion connected devices.

IDC describes the IoT as a network connecting - either wired or wireless - devices (things) that is characterized by automatic provisioning, management, and monitoring. It is innately analytical and integrated, and includes not just intelligent systems and devices, but connectivity enablement, platforms for device, network and application enablement, analytics and social business, and applications and vertical industry solutions. It is more than traditional machine-to-machine communication. Indeed, it is more than the traditional Information and Communications Technology (ICT) industry itself.

The IoT will, in fact, subsume the ICT industry over time - and to good effect. The compound annual growth rate for spending on traditional ICT from 2013 to 2020 is just under 4%. Vendor revenues tied to the part of IoT that are not already in traditional ICT spending will grow at three times that rate. And that's just the revenue to the supply side. To the buyers and users of IoT technology and services, the payoff should be at least twice that - perhaps many more times.

The IoT, however, comes with its own challenges, including a lack of standards, the ability to scale globally, security concerns, and an immature ecosystem. For vendors, there is no homogeneous IoT market - each industry and application is different. For users, especially IT organizations, there can be issues of managing operational systems in an organization that might be culturally designed as a support organization, as well as dealing with the real-time demands of many IoT applications.

EMC and IDC see the IoT creating new opportunities for business in five main ways by enabling:

- **New business models:** The IoT will help companies create new value streams for customers, institute processes that speed time to market, measure market performance, and respond rapidly to customer needs, to name a few.
- **Real-time information on mission-critical systems:** With IoT, enterprises can capture more data about their processes and products in a timely fashion and radically improve their market agility to become software-defined enterprises, creating new revenue streams, improving operational efficiency, and increasing customer loyalty.
- **Diversification of revenue streams:** The IoT can help companies monetize additional services on top of traditional products, e.g., vending machine vendors offering inventory management to those who supply the goods in the machine.
- **Global visibility:** The IoT will make it easier for enterprises to see inside the business regardless of location, including tracking it from one end of the supply chain to the other. This will lower the cost of doing business in far-flung locales, and the assertion that "software is eating the world" will ring true everywhere.
- **Efficient intelligent operations:** Access to information from autonomous end points, as today's smart grid already supplies to utility companies, will allow all kinds of organizations to make on-the-fly decisions on pricing, logistics, sales, and support deployment, etc.

The impact of the IoT is already visible in the digital universe. Data just from embedded systems - the sensors and systems that monitor the physical universe - already accounts for 2% of the digital universe. By 2020 that will rise to 10%.

There is one final way to look at the importance of embedded systems - the load they will put on the IoT in terms of management. Computers don't just have to manage megabytes, they also have to manage the containers, or software-based digital "files" that the megabytes come in. Some containers are big, like a digital camera image or a 30-minute loop on a surveillance camera. But some are small. RFID tags and sensor "containers" may contain as few as 32 bytes. Because of this small signal size, the number of containers that must be managed from these embedded systems will dominate the digital universe in 2020. We're talking 99% of all "files" in the digital universe.

Because of the growth of embedded systems data in the digital universe, the number of "containers" is growing faster than the number of petabytes, from 28 quadrillion in 2010 to 4,200 in 2020.

Finally, a good portion of the digital universe will be generated by mobile devices and people - from 17% in 2013 to 27% in 2020 - but the percentage of mobile "things" in the IoT will be more than 75% by 2020.

Every few years in this industry we witness the emergence of a new "next big thing." The IoT is surely the next big thing in 2014.

HIGH VALUE DATA: FINDING THE PRIZE FISH IN THE DATA LAKE IS KEY TO SUCCESS IN THE ERA OF THE THIRD PLATFORM

The phenomenon is this: enterprises are finding new sources of data, new ways to analyze data, new ways to apply the analysis to the business, and new revenues for themselves as a result. They are using new approaches, moving from descriptive to predictive and prescriptive analytics and doing data analysis in real-time. They are also increasingly adopting self-service business intelligence and analytics, giving executives and frontline workers easy-to-use software tools for data discovery and timely decision-making.

Think of Under Armour embedding sensors into the shorts worn by NFL hopefuls at this year's tryouts. Think of John Deere, which changed its IT organization in 2010 and is now able to deploy new software to tractors in the field nearly every month, software that can help tractors drive straight across plowed fields (fields which may already have sensors in them tracking environmental conditions reporting back to smart phones or PCs). Think of EMI Music, which drives marketing of artists based on data from a database of a million music-listener surveys and information from Spotify music streams.

Call these intrepid enterprises "data-driven companies." Call them "software-defined companies." Call them what you will, they are on track to be successful in what IDC calls the era of the Third Platform.

It's early yet for definitive studies on the opportunities for this new breed of organization, but as reported in the Harvard Business Review, companies in the top third of their industry in the use of data-driven decision making are, on average, 5% more productive and 6% more profitable than their competitors [<http://hbr.org/2012/10/big-data-the-management-revolution/ar/2>]. A study by IDC found that users of Big Data and analytics that use diverse data sources, diverse analytical tools (such as

predictive analytics), and diverse metrics were five times more likely to exceed expectations for their projects than those who don't. The more data and analytics diversity, the better.

However, the digital universe may actually be an obstacle for companies trying to become data-driven. There is too much information, it is too diverse, and it is too effervescent.

The key is to find the part of the digital universe that is richer than others. To size the portion of the digital universe that one could call "target rich" data, IDC analyzed the 60 or so data streams rolling up into the digital universe and gave them subjective measures on 5 criteria:

- Ease of access - Can you obtain the data, or are they hopelessly locked away on end user PCs, shuttling about on closed-end data processing systems, or trapped in proprietary embedded systems?
- Real-time - Is the data available in real-time, or does much of it come too late to drive real-time decisions and actions?
- Footprint - Could top notch analysis of this data affect a lot of people, major parts of the organization, or lots of customers?
- Transformative - Could this kind of data, properly analyzed and acted upon, actually change a company or society in a meaningful way?
- Intersection synergy - Could this kind of data have more than one of the above attributes?

These are subjective measures, sure, but by assigning them value we can come up with a first-order approximation of how big this high-value data pool is, and how fast it is growing. For 2014, we peg this type of data at a little more than 6% of data that's useful if tagged, rising to 11% in 2020. That means we are talking about little more than 1% of the digital universe as a whole, an entirely more manageable area of discovery.

In 2014, the majority of this target rich data is general IT data, which includes all metadata. And that will continue to grow as Big Data projects expand and as metadata builds up, including metadata on metadata. But the embedded portion of the IoT will grow from less than 10% of this target rich data to more than 20% in 2020. The biggest decline in target rich data will come from surveillance data, as the analog-to-digital transition winds down, as compression algorithms improve, and as the installed base growth slows.

For enterprises, the news is this: While the size, diversity, and rapid expansion of the digital universe are never-ending challenges to deal with, they are also a source of opportunity for those with the tools and corporate will to take advantage of them.

THE IT IMPERATIVES

Before organizations can begin to turn the promise of the digital universe into reality, the following fundamental conditions in the management and deployment of IT will need to be met:

- Information security will need to be tightened.
- Information privacy will need to be ensured, particularly as more third-party consumer data is incorporated into the operational and analytical frameworks.
- Data warehouses will need to be upgraded or swapped out for more flexible data repositories that can handle various data types, automatic tagging, autonomous data "check-in," and many terabytes. These warehouses must be able to store the vast amount of data on the most efficient infrastructure, bowing to the reality that only a fraction of stored data is actually engaged at any given moment.
- Data analytic output will need to be driven to more parts of the organization, including real-time input to operational decision making.
- The datacenter will have to be virtualized and cloud computing incorporated into enterprise IT architectures.
- IT managers will have to enable and manage the explosive growth in mobile devices that increasingly request and send data from all over the world.
- IT managers will need to provide the appropriate security permissions to allow for data to be queried regardless of where it is stored, giving rise to virtual "data lakes."
- While the digital universe is doubling every two years in size, and the number of information "containers" every 20 months, the number of IT professionals on the planet may never double again - or at least not for 20 years. In the context of the digital universe, the number of gigabytes per IT professional will grow by a factor of eight between now and 2020, while the number of devices on the IoT grows by a factor of 2 (not counting virtual devices).

None of these challenges are insurmountable. Most companies are somewhere along on each of the vectors. There are more virtual servers installed now than physical servers. Data analytics software is already a \$40 billion market and growing 10% a year. Security products and services are a \$50 billion market. Cloud computing now accounts for 5% of total IT spending, growing to 10% by 2020. But real transformation to a data-driven or software-defined enterprise is an all-hands-on-deck imperative. IT alone can't make the transition.

Perhaps the first order of business is information security and privacy. The irony of the digital universe is that, while most of it (almost 70%) is created by the actions of individuals - taking pictures, watching digital TV, being captured on surveillance cameras in airports - enterprises have contact with, and therefore liability and responsibility for more of it (85%), such as account information, email addresses, location stamps, and so on.

Not all the digital universe needs tight security, of course -- old photos still on camera phones, digital TV shows watched and consumed, etc. But IDC estimates that at least 40% of it requires some level of security, from privacy protection to full-encryption "lockdown." Unfortunately, the amount of the digital

universe that needs protection, but which actually has protection, is less than half. Also unfortunately, the amount needing protection will grow as more of the IoT comes online.

On an enterprise level, the need for information security will reach outside the boundaries of the organization as third-party and public sector data become folded into ongoing analytics. Is the outside data clean? Have users opted in? Is it accurate? What are the supplier's privacy policies and assurances to those from whence the data came?

The next order of business is probably organization around data. Traditionally, data stored for analysis is kept in data warehouses and must be cleaned and prepped carefully for later retrieval. Data is cataloged in a way that makes it easy to analyze predefined variables according to generally known data types.

But the data needed to support business transformation in the era of the Third Platform is much messier. It tends to be unstructured, diversely formatted, of uncertain accuracy (and sometimes uncertain origin), of unpredictable value, and often flowing into repositories and demanding attention in real time.

The role and importance of the datacenter (be it private/public cloud-based or otherwise) is being catapulted in the era of the Third Platform. The exponential growth in mobile (and wearable) devices, with little to no embedded storage and yet integrated with incredible processing power, will require data anytime, anywhere, on any device, and of all types. Datacenters must be prepared to feed this inevitable voracious compute appetite with data.

IDC tracks a blizzard of technologies needed to deal with tomorrow's data, from relational databases to key value stores, text analytics, ontologies, categorizers, schema extractors, search indexes, parallel file systems, complex event processing engines, graph databases, and other technologies that organize and manage data at rest and in flight.

Together, these technologies and processes can slowly change the hidebound structure of today's data stores to more egalitarian and flexible stores for tomorrow, which some vendors call "data lakes." Data can flow in without prior classification and tools like Hadoop, parallel processing search programs, or complex event processing engines can process and classify the data later. Software is augmenting our limited capacity and bandwidth to make sense of the data deluge.

But these new software tools are not useful unless they are put to use in the service of specific business objectives. In a 2013 survey of 700 large North American enterprises, IDC found that fewer than 1% had achieved the highest level of Big Data usage - where the use of big data is operationalized and continuously providing process improvement realization. And these were the most likely candidates to be further along.

So that journey has just begun.

THE BUSINESS IMPERATIVES: A CALL TO ACTION

Becoming a data-driven or software-defined enterprise, if you aren't already one, may be no easy task. Think of how long it took brick-and-mortar companies to adapt to the online world, and how many are still struggling. Now, imagine the online world on steroids. That's the era of the Third Platform.

The challenges and opportunities of the digital universe are technical, for sure, but they are first and foremost organizational. Here are three steps all enterprises must take without delay if they want to thrive in the new era:

- Create a C-level position in charge of developing new digital business opportunities: Whether it is a new position (e.g., Chief Digital Officer), or an enhanced existing one (e.g., upgrading the CIO's current responsibilities), this new executive role is in charge of identifying and pursuing new revenue streams based on internal and/or external data and its analysis.
- Develop and continuously revise an executive-team understanding of the new digital landscape for your enterprise: Who are the new (and potential) digital competitors? How are you going to cooperate with others in your industry to anticipate and thwart digital disruption? What are the short- and long-term steps you must take to ensure a smooth and timely digital transformation?
- Design and execute a plan for accelerated investments in digital enterprise technologies and skills: Re-allocate resources across the business based on digital transformation priorities, invest in promising data collection and analysis areas, and identify the gaps in talent and skills required for success in the era of the Third Platform.

Increased investment in human capital and the new skills required today is the first order of business for all organizations. Another crucial action to take is launching an enterprise-wide initiative to determine where to focus the response to the opportunities presented by the Third Platform. What kind of new products and services based on data and analytics can the business develop and sell? What are the areas of the business where the Internet of Things will have the biggest impact? These and similar questions must be answered and acted upon, then reviewed and the answers revisited every six months as new applications and new sources of data become available. The Third Platform requires not a "strategic plan," but a continuously revised action plan.

The right skills, a laser-sharp focus, and a passion for continuous learning and adaptation are a must for enterprises to thrive in the digital universe of tomorrow. The digital universe will continue to expand at a rapid clip and flood already saturated infrastructures and the people that manage them. The Internet of Things will add new levels of complexity – and opportunity – on top of what Big Data has offered in just the last three years. The bright stars of the digital universe of tomorrow will be the enterprises that will ride the data flood to new heights of productivity and prosperity.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com

Copyright Notice

This IDC research document was published as part of an IDC continuous intelligence service, providing written research, analyst interactions, telebriefings, and conferences. Visit www.idc.com to learn more about IDC subscription and consulting services. To view a list of IDC offices worldwide, visit www.idc.com/offices. Please contact the IDC Hotline at 800.343.4952, ext. 7988 (or +1.508.988.7988) or sales@idc.com for information on applying the price of this document toward the purchase of an IDC service or for information on additional copies or Web rights.

Copyright 2014 IDC. Reproduction is forbidden unless authorized. All rights reserved.

