DATA GENOMICS INDEX

A report on the true makeup of storage environments from the experts in unstructured data.

2017

NOV

VERITAS™

An organization's ability to develop and execute against its digital strategy is key to succeeding in the modern era. Companies leading the way into digital transformation have profit margins up to 57% higher than those who choose to remain in the past.[1] Digital transformation requires organizations to harness the power of their information in a number of new ways. To do so, they need to have a clear understanding of the data they have under management. At this time, organizations are still wrestling with an ever-growing and ever-evolving data estate.

As organizations grow and redefine their business roadmaps, some data loses relevance—it goes dark. Dark data directly and indirectly impacts performance, revenues and business visibility. This underscores the importance of addressing behavioral trends within the corporate data universe.

The Data Genomics Project is an initiative designed to question the status quo in data management. Veritas founded this initiative to bring a community of like-minded data scientists, industry experts and thought leaders together with the purpose of uncovering the true nature of the unstructured data that we create, store and manage.

The primary output of the Data Genomics Project is the annual Data Genomics Index (DGI)—a benchmark report on the composition of real storage environments. In working with 86% of the Fortune 500 and backing up, archiving or analyzing exabytes of data for customers, Veritas is in a unique position to understand the defining characteristics of a typical enterprise data estate.

Our analysis focuses on anonymized metadata aggregated from Veritas Information Map™. Through this analysis, we can surface accurate details about the real composition of data environments.

Compared to the 2016 DGI, this year's analysis draws from a substantially larger data pool. The 2017 DGI analyzed more than 31 billion files to generate insights into the composition of enterprise data. Data was sourced from across the globe and from a wide variety of industries.

2017 DGI Data Population

billion

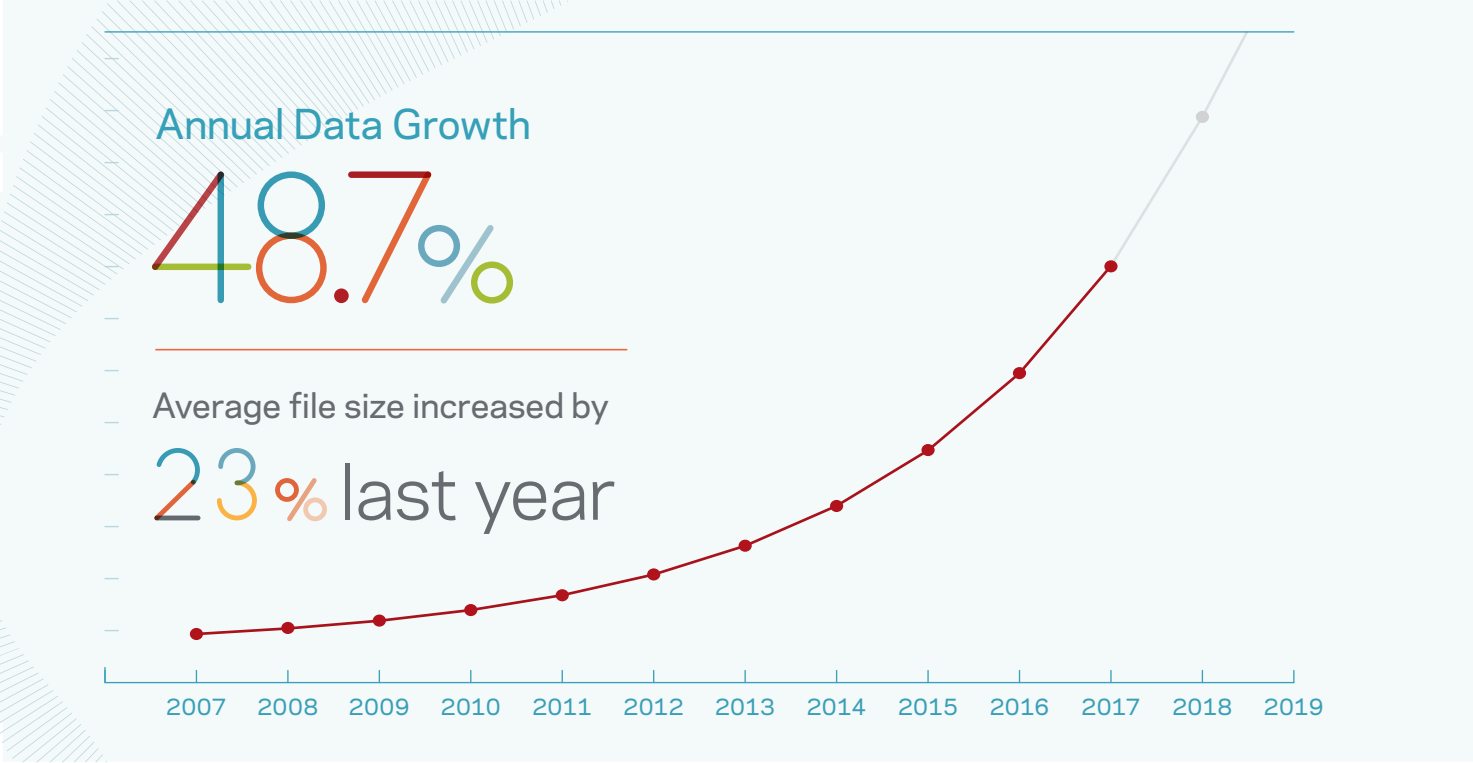
31 billion files

20 million GBs of data

8,000+ file type extensions

14 countries[2]

10 industries[3]



Today's Data Snapshot: Data Density on the Rise

The proliferation of cloud-based apps and emerging technologies such as AI, IoT and blockchain[4] is disrupting the way data is created, managed, and utilized. Organizations are seeking to use these new techniques to generate more powerful insights that allow them to redefine the customer experience and deliver more valuable solutions to their customers.

More than 50% of large organizations will compete using advanced analytics and proprietary algorithms by 2018.[5] This will give way to new opportunities to collect, analyze, and potentially monetize data.

These trends translate into higher file density, a composite measure of average file size and file variety. We expect these trends will continue to evolve throughout 2017 and beyond.

Tectonic Shifts in Virtual Real Estate

You can't manage what you don't measure. As new technologies gain momentum and speed up data creation, a digital transformation roadmap has become critical to align corporate strategy, people and processes. Companies now find themselves in a race to fully understand the real value of their business data.

New rules of engagement among corporations, customers and employees have given way to new technologies. Shorter forms of communication have drastically diversified the data universe composition. As a result, the distribution of stored file types has shifted.

Unfortunately, the road to transformation remains imperfect. Leaders should formulate their data management strategies based not only on recently created data, but also on the redundant, obsolete and trivial (ROT) data they are already storing.

The Main Characters

Y-o-Y Changes

Storage Capacity Increase[8]

Backup GBs	+98%
Compressed GBs	+89%
Unknown GBs	+28%

Storage Capacity Decrease[9]

Image GBs	-61%
Video GBs	-50%

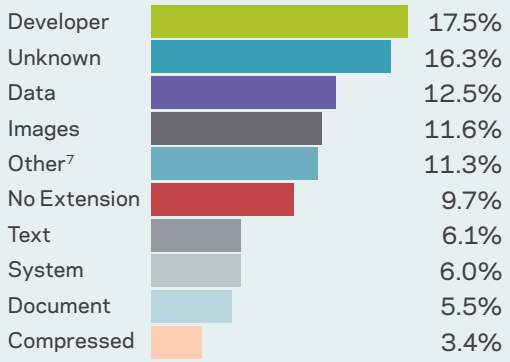
Almost

50%

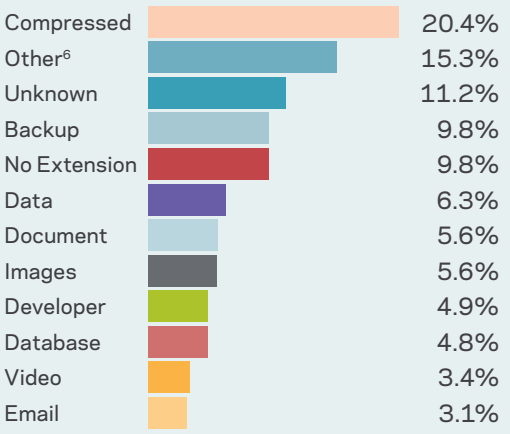
of Storage Capacity has been captured by 4 file types

- Compressed
- Unknown
- Backup
- No Extension

File count

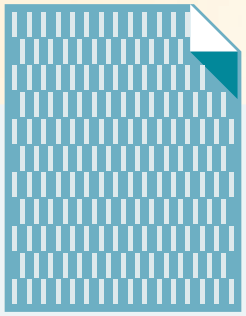


Storage capacity



Over the last 10 years, newly-created **backup** files have increased in size by

2,586%



36.2 MB
average file size in 2017

Old Habits Die Hard: Is Backup a "Save It All" Strategy?

Both backup and compressed files are expanding their presence within the global storage landscape. While the rate of data growth explains this acceleration, it also begs the question whether organizations really need to be storing copies of all of their data.

World-class data management is not just about optimizing storage space, it requires organizations to distinguish mission critical information from ROT data—as well as visualizing and acting upon it at the right time. Additionally, corporations face increasingly strict regulatory responsibilities—like FOIA and GDPR—that demand promptness and accuracy when gathering supporting documents for litigation or information request processes. With this in mind, compressing old files and maintaining a "save it all" strategy no longer seems like a solution that organizations can afford. Companies should let go of old habits and embrace innovations found within new data management tools to classify data, assign consistent retention policies and automate these workflows to achieve an elegant and efficient solution to this conundrum.

Files catalogued as "Unknown" have soared during the last year, both in **storage capacity** and in **file count.**

Unknown file count

+51%

since 2016

The Mystery of the Unknown

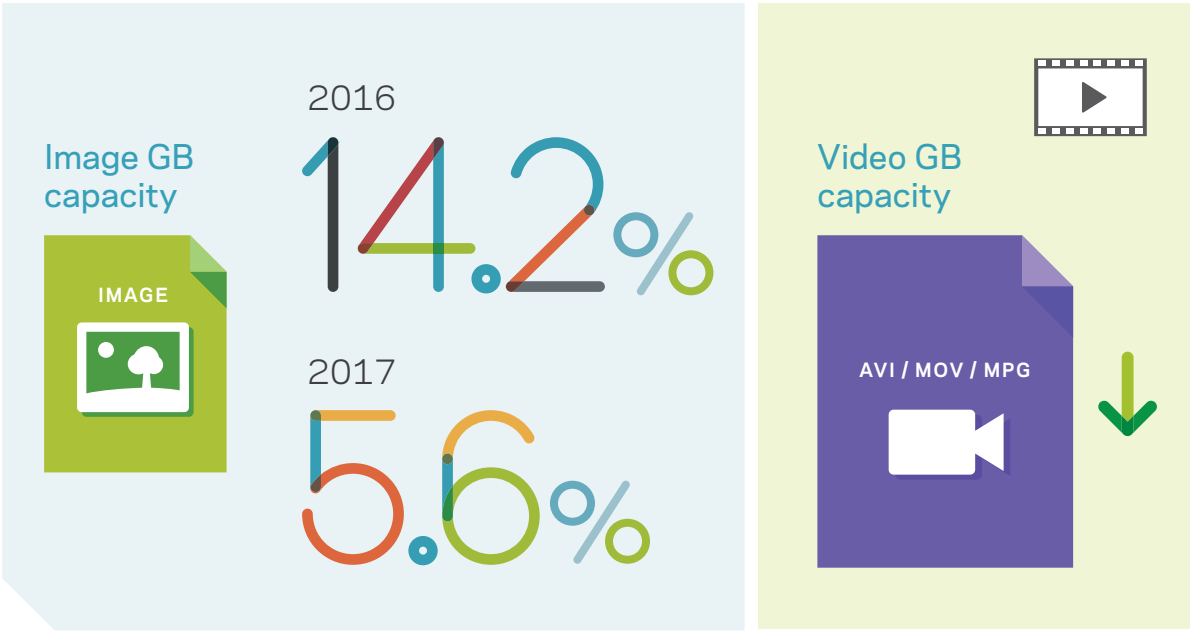
New product innovation has fueled the growth of Unknown files, and digital transformation lies at the core of this trend. Corporations seeking to enhance the customer experience are using emerging technologies to build new products and services. Custom apps are being developed to extract value from increasing volumes of customer data and deliver those products and services to consumers. This trend manifests itself through a surge in new and diverse types of files, catalogued as "Unknown" in our analysis. The rising challenge relies now on implementing the right tools to manage these soaring volumes of known and unknown unstructured data, while supporting innovation and mitigating the risks associated with it.

Media Management Seems to be on Track

With average image file size remaining virtually unchanged throughout the last 10 years, a reduction in overall Image GB capacity from 14.2% of the data environment in 2016 to 5.6% in 2017 suggests that companies are being more conscious about what they are storing—in terms of images, at least.

Video File GB capacity has seemingly dropped as a result of new information governance policies aiming to prevent employees from taking up corporate storage space with personal files.

Organizations are not standing still, yet the road to digital transformation still presents significant challenges, and the implementation of new technologies might drastically disrupt these trends.

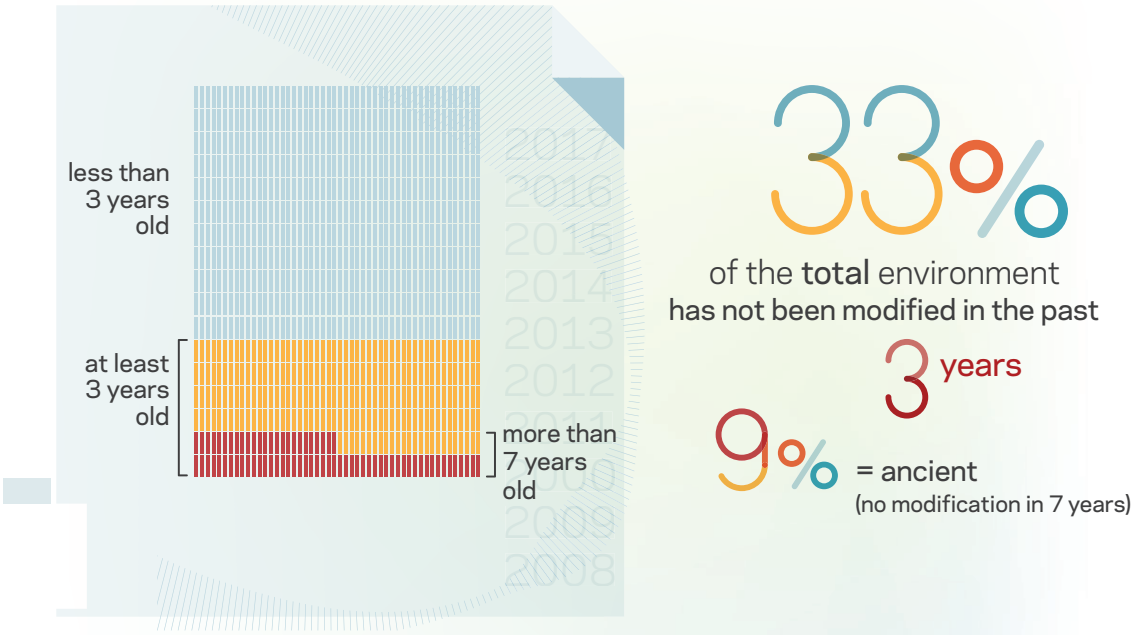


redundant obsolete trivial data

ROT vs. ROI: Shedding Light on Dark Data

Smart data managers maximize the utility of their data by always tracking when data was created, where it resides, when and how it was modified, why it's (still) there and how to leverage its value. Here's where the real advantage of having the right data management tools arises. In order to transform data into information, it must be given relevance and purpose. And that's easier said than done.

Achieving this goal will empower organizations to drive action and make the right decisions. Harnessing the potential of data utility is more than understanding the past. It's about creating a benchmark for the future. Data is no longer a resource. It's a foundation.



Data's Most Wanted List

In light of some advancements that organizations have made regarding better data governance, there continue to be certain file types that pose challenges for individual industries. Here's a glimpse into the worst offenders.

Image files have long prevailed in corporate data ecosystems across industries including Consumer Products, Education, Government and Healthcare. This comes as no surprise, since each of these industries, in its own way, bases business value on this type of file. Image file size has not varied significantly over time however. Corporations fitting this profile should watch out for image file count.

Organizations that rely almost exclusively on images will likely continue to do so. However, the introduction of new technologies will load storage repositories with more than images, and it is possible that the surge of Unknown files will eventually fuel the image file count. In Healthcare, applications for AI focusing on imaging and diagnostics are on the rise. Healthcare-focused AI is proliferating among organizations of all sizes, leading to the prediction that this tendency won't disappear any time soon.

Government organizations should also pay attention. With AI gaining greater momentum in the public sector, government agencies should arm themselves with the necessary tools to face the potential data storage repercussions of implementing this technology along with other emerging projects, like Smart Cities, that rely almost exclusively on data. The motivation to do so is definitely there. Deloitte estimates that $41.1 billion of US public spending dollars could be saved yearly by having an AI-augmented government .[13]

Industries with highest Image Storage Capacity

Images also prevail in Education. Despite AI facing initial challenges to launch within this segment, adoption is growing. Enhanced applications like Deep Learning require an increasingly robust data management infrastructure, adding large volumes of data to the storage ecosystem.

Within the financial services industry, document files have historically taken a leading role, since most operations are paper-based with a digital wrapper. There is a certain degree of skepticism as to how long this trend will prevail, especially given the surge of Smart Contracts. Moreover, this industry is facing a number of transformations, particularly related to blockchain transactions. Given that 80% of banks worldwide will have invested in this technology by the end of 2017, and considering the elevated risks of poor data governance within such a highly regulated industry, we should expect CIOs to address this topic sooner than later.

Document files make up

of stale data in the Financial Services industry

The Time is *Now*

Data growth continues to accelerate, but in new ways. A "save-it-all" strategy is short-sighted and has already proven ineffective: compressed files will never compensate for the soaring volume (and density) of data that's being created. Backup and Compressed files will never be able to easily grant Risk, Legal, and Information Managers the visibility they need to better manage their data and their business.

Digital transformation is defined by the value of information—how to maximize it, leverage it or even identify it. To stay ahead of the curve, companies must now incorporate data management tools into their action plans and acknowledge how much a solid (or mediocre) digital strategy can impact their business goals.

Many companies still see digital transformation as a tool to primarily increase revenues or decrease costs. These causes are certainly worthy but somewhat myopic, since they fail to recognize the real goal behind digital transformation: boost innovation, embed the latest technologies into products and redefine the customer experience. This in turn impacts customers and employees, since it has the possibility of dramatically accelerating growth, performance, and business operations. A profound understanding of the human experience that surrounds digital strategies—and the right tools to achieve it—form what's now become the key to turning technology into transformation.

VERITAS

About Veritas Technologies

Veritas Technologies empowers businesses of all sizes to discover the truth in information—their most important digital asset. Customers utilize the Veritas platform to fast track digital transformation, accelerate multi-cloud data management and solve pressing IT and business challenges, from data protection and workload portability—with no cloud vendor lock-in—to storage optimization and compliance readiness. Eighty-six percent of Fortune 500 companies rely on Veritas today to reveal data insights that drive competitive advantage. Learn more at www.veritas.com or follow us on Twitter at @veritastechllc.

Legal:

[1] https://hbr.org/2017/01/what-the-companies-on-the-right-side-of-the-digital-business-divide-have-in-common
[2] Represented Countries: USA, UK, Brazil, South Africa, Sweden, Canada, France, Chile, The Netherlands, Malaysia, Panama, Ireland, Australia, Italy.
[3] Represented Industries: Consumer Discretionary, Consumer Staples, Education, Financials, Government, Healthcare, Industrials, Information Technology, Telecomms, Utilities.
[4] Blockchain is a new technology, first developed as the infrastructure to Bitcoin, that allows transactions to be conducted and recorded on shared, immutable and globally distributed digital ledgers managed by peer-to-peer networks of connected computers.
[5] http://www.gartner.com/newsroom/id/3546517
[6] Corresponds to file types representing less than 3% of total storage space.
[7] Corresponds to file types representing less than 3% of total number of files
[8] Shows file types with greatest increase in storage capacity over a 1-year period (2016-2017)
[9] Shows file types with greatest decrease in storage capacity over a 1-year period (2016-2017)
[10] Veritas Data Genomics Index Report 2017
[11] Return On Information
[12] Veritas Data Genomics Index Report, 2017
[13] Veritas Data Genomics Index Report, 2017
[14] https://dupress.deloitte.com/dup-us-en/focus/cognitive-technologies/artificial-intelligence-government-summary.html
[15] https://www.weforum.org/reports/the-future-of-financial-infrastructure-an-ambitious-look-at-how-blockchain-can-reshape-financial-services