eGuide

FROM IDG

CIO COMPUTERWORLD CSO InfoWorld NETWORKWORLD

VERITAS™

presents

Data Management

Data: It's the lifeblood of today's enterprise. This eGuide is aimed at helping you consider the role of data in your organization, and how to best manage, secure, and protect this most precious asset.

Why data governance?

Data governance used to be a nice-to-have, but due to the increasing focus and importance of data and analytics, it's becoming a necessity that helps to drive data management across the enterprise.

BY NANCY COUTURE | Enterprise Data Management (EDM) seems to be at the top of many organizations' strategies for 2018, as the importance of data to organizations continues to grow exponentially. EDM plans may include modernizing an existing data warehouse to enable near-real-time data, building a big data environment to support deeper analytics, focusing on and increasing digital capabilities and associated analytics, moving existing data and analytics to the cloud, increasing analytics capabilities in the organization, or most likely a combination of these.

Data governance is a key component of EDM, and is also taking on a higher level of importance. Some of the key trends that are causing a greater need for data governance include:

- Increasing data volumes from more and more sources, causing data inconsistencies that need to be identified and addressed before decisions are made using incorrect information
- More self-service reporting and analytics (data democratization), creating the need for a common understanding of data across the organization
- The continuing impact of regulatory requirements such as GDPR, making it even more important to have a strong handle on what data is where, and how it's being used
- An increasing need for a common business language to enable cross-departmental analysis and decisions

Regardless of the type of data an organization is managing—data warehouse, data lakes, big data—a strong data governance capability is important. It will enable proactive management of data.

For example, a financial institution I worked with had very poor, inconsistent customer data. All of the customers with first, middle and last names had multiple differences, and addresses were inconsistent. This type of situation makes it very difficult to do any type of customer analytics, from identifying cross-sell opportunities to tracking and understanding customer experience. Data governance can be a first step in identifying the issues, defining standards, and implementing changes in the business to align with these standards.

My prior article provided ideas on the initial (discover) phase of a data governance initiative. With the right assistance, an organization can begin data governance activities during the discover phase (typically a handful of weeks), then move right into implementation. The discover phase should have resulted in a set of recommendations and a roadmap to follow, both for immediate next steps as well as longer-term considerations. The

With the right assistance, an organization can begin data governance activities during the discover phase (typically a handful of weeks), then move right into implementation.

discover phase should have also provided some clarity on the initial areas of focus, such as:

- Identifying key sources of truth and ensuring they are consistently used
- Improving data quality across these key data sources
- Developing standard business language and a business glossary for key data sources
- Enhancing organizational metadata
- Developing standards and policies for data access
- Identifying business priorities for EDM initiatives
- Developing and providing data and report certification
- Developing and sharing an enterprise data framework to provide data visibility and alignment

The actual implementation of data governance can take weeks or even months, depending upon the level of attention and engagement there is across the organization. One important thing to recognize is that most successful data governance programs become part of the day-to-day processes over time. So, as you embark on implementation, ensure that what you put in place is scalable and sustainable.

- Start with a mission and objectives that include the initial focus area.
- Get the structure and the right people in place, from executives to stewards.
- Establish the data governance charter and policies from the start.
- Ensure there's visibility of decisions and progress across the organization.
- Implement with the goal of developing processes that can be embedded in the day-to-day business operations.
- Keep it as simple as possible, with the goal of less bureaucracy and more progress.

Data governance used to be a nice-to-have, but due to the increasing focus and importance of data and analytics, it's becoming a necessity that helps to drive data management across the enterprise.

Data management: beyond just storing bits

As unstructured data grows, opportunity lies in the growing field of data management.

BY STEVE PAO | Data is the new oil. As enterprises increasingly recognize the value of digital assets, infrastructure considerations have evolved well beyond simply storing data into the growing field of data management.

Traditional views on managing data typically involve block storage for database and application servers. But videos, images, sensor data, and other unstructured file data that can't be easily stored in traditional relational databases are bigger and much harder to manage.

Real opportunity lies in data management, and unstructured data requires a different approach. With unstructured data predicted to grow at a compound annual growth rate of 29.8 percent through 2021, according to IDC Research, data backup and archive must evolve.

Here are four higher-level functions of data management.

1. Data protection

Anyone who has lost their smartphone without backing up their data has felt the pain of data loss. For consumers, data protection is an issue of securing their privacy and digital assets of personal value, such as photos with family and friends.

For enterprises, data loss is not only inconvenient and painful, but can be devastating for business. Understandably, enterprises want to protect their valuable intellectual property and other critical data. After all, it's what has helped them achieve their success.

With data protection in mind, organizations seek secondary storage that backups and archives quickly—meaning no risk of missing backup windows or any need for backup windows at all—and restores data on-demand.

2. Data movement

Growing data sizes and more sophisticated computing paradigms have created a need for data movement.

For example, my iPhone integrates with a variety of cloud services, including iCloud and Dropbox. Data movement policy can determine what photos get replicated to iCloud and what is retained locally on the phone. Policy conditions, such as available capacity on the phone itself, can determine whether high-resolution or lower-resolution versions of photos are stored locally and what is stored in the cloud. Performing these functions depends on a central catalog of all the data and its awareness across tiers of storage.

Similarly, enterprises in every industry—from media and entertainment to bio-IT—have large amounts of data, performance issues, and a desire to constrain local capacity growth, but at a scale much larger than for individual consumers.

3. Discovery

Data has become hard to find. That's because it's difficult to know what's scattered across hundreds of file systems and

With unstructured data predicted to grow at a compound annual growth rate of 29.8 percent through 2021, data backup and archive must evolve.

hybrid cloud architectures.

As the vast majority of data is machine-generated, manual cataloging simply isn't possible. Today, some applications to manage digital assets generate their own catalogs, but federating those silos requires an appropriate ecosystem.

In the consumer world, separate catalogs exist on the iPhone for Photos, Music, App Store, and other applications. However, searching through those catalogs is possible using Spotlight Search for iOS's federated search, because the iPhone provides a search indexing ecosystem for all applications to utilize.

As in the consumer world, separate applications for unstructured file data may exist in the enterprise in the form of document management systems, digital asset managers, laboratory information management systems, and other applications. Other times, applications may not exist at all to manage file data for workflows. No easy system exists to let users know what data is available or where it lives.

Opportunities exist to provide a common layer in which to store, catalog, and manage metadata for search and discovery of digital assets in the enterprise.

4. Learning

Every data management application needs a set of analytics to help users know what's there, and application scenarios often involve analyzing data to create more useful metadata for classification.

Back in the consumer world, Photos for iOS catalogs photos based on EXIF data generated by the camera app, such as the time and GPS coordinates where the photos were taken. Beyond simple cataloging, Photos for iOS can also classify the photos to identify faces and places through machine learning. When you search your photo files for "London," Photos parses GPS coordinates from the image files using this classification and pulls up all photos taken in London.

In the enterprise world, these same opportunities exist to "decorate" metadata in a common way across applications. Opportunity exists to move from a closed world, where decorated data often lives in proprietary applications or SQL databases off to the side, to an API-driven world, where metadata is stored in common object formats and index stores are accessible via common APIs. Making data more easily accessible to software paves the way for valuable applications, such as using machine learning for auto-classification of data.

While traditional views on enterprise data focus on storage and data protection, the real opportunity in today's digital economy lies in the higher-level functions of data management.

In a previous article, "3 reasons to embrace horizontal scaling for secondary storage," I referenced a comment made by Taneja Group Founder Arun Taneja. "Data protection as a discipline has been sleeping for decades," Arun wrote, "but data management may be the wakeup call it needs."

When you have a common data management layer, you gain so much power. From an enterprise perspective, wouldn't it be beneficial for different applications to rely on the same underlying data management functions?

New innovations in data management make it possible. Tying all these data management functions together and eliminating silos benefits your enterprise, your IT team, and your business users.

Data preparation is the key to big data success

Common barriers to big data adoption, and how to adapt and overcome those challenges.

BY MICHAEL JUDE | Big data it is often hyped, but I encourage taking a more realistic stance. I've seen many organizations attempt to adopt big data solutions and ultimately fail. I fear these missteps may eventually sour the market on adopting big data solutions. This would be unfortunate, because I view big data as a transformational capability, an essential part of a new IT infrastructure. To ensure greater success, this post presents common barriers to adoption that I've observed and provides insight into how to adapt and overcome these challenges.

One of the primary barriers to big data success is the lack of a data preparation strategy. Data preparation includes all the steps necessary to acquire, prepare, curate, and manage the data assets of the organization. Sound data is the foundation for actionable insights delivered by advanced analytic applications. If the data is tainted, then conclusions based on it become questionable—moreover, debatable—and big data, if not backed by accurate intelligence, can add to confusion and organizational turmoil.

It's not unlike the Hippocratic oath: "First, do no harm." The worst outcome of a big data undertaking would be to make poor decisions because of bad information, but be really confident!

Interestingly, most companies contemplating big data and, unfortunately, vendors selling such solutions rarely consider the implications of data preparation. Building the hardware infrastructure and software to support a big data lake can be complex and expensive, leading adopters to conclude that this is the most challenging element of the big data equation. However, once the infrastructure is in place, they are often dismayed to discover that the big data infrastructure is simply the tip of the iceberg. Collecting and managing trusted data can be much more expensive, especially if the big data project begins with a poorly understood idea of what data will ultimately be required.

So, what constitutes a good data preparation strategy? The following six-step process will help:

1. Identify your decision set.

Knowing the context of a company's decision-making is an essential first step. It defines the data sets you will use to support a decision, how the data will be manipulated, and ultimately the analytical process that will define insight generation. Many assume that if data is simply cleansed and curated effectively, any analytic process can be supported. This is not true. Organizational leaders need to define the end game first, and data preparation will be much simpler.

2. Select the data sources to support the desired decisions.

Granted, you cannot know in advance every possible data source that might be needed, but it is possible to identify the primary data sources that will need to be used. These will not only define

Once infrastructure is in place, adopters are often dismayed to discover that the big data infrastructure is simply the tip of the iceberg.

the types of data available but will largely define the kinds of data cleansing that will need to be done.

3. Choose the right vendor of data cleansing technology.

You will want technology that not only offers solutions that accommodate your initially identified data types but also offers a platform that feeds your existing cross-organizational analytic tools. In an analytics-driven company, many different levels in an organization will be using tools to inform decisions. It is essential that the data-preparation tool provides a platform, accessible by all, to enable access to curated and trusted data. Only by starting from the same basic data set will decisions be consistent across the organization.

4. Assess and ingest additional data sets.

As noted in Step Two, it is not possible to predefine every data set that might be necessary to support a decision; additional data sets are constantly being discovered that might be useful. Assessing new data sources is ongoing and an important part of decision-making.

5. Identify any new analytic tools that will produce the desired insights.

There are many excellent analytic packages on the market, from simple statistical tools all the way to very advanced machine-learning-based applications. They each provide different insights and require different degrees of data cleansing. For example, machine learning may be able to handle data in an essentially native mode, while a statistical tool might need very clean data, with every field reconciled.

6. Extend data preparation to incorporate new data into existing data sets.

Many data sets are dynamic and evolving. As new data is discovered or becomes available, data preparation must be conducted to ensure its ready availability.

By recognizing that data preparation is a necessary first step to a big data value, an enterprise can significantly reduce the costs of big data while accelerating the delivery of actionable insights that drive good decision-making. Fortunately, the data preparation market is growing rapidly, with Frost & Sullivan projections of over $9 billion globally by 2025; finding solutions and expertise to address the need for a data preparation strategy are readily available.

The case for data-centric audit and protection in the Information Age

Data-centric Audit and Protection (DCAP) applies security to only specific pieces of data—making it scalable while having minimal impact on business processes.

BY TRAVIS WRIGHT | Chief information officers and chief information security officers, as well as anyone interested in protecting their enterprise data, find themselves at a turning point in history. As the rise of the Information Age has made data more valuable than ever before, hackers have never posed a more severe societal threat than they do today. The examples of major security breaches are so numerous even from only the past year that there's no need to list them specifically. The havoc they've wreaked is unimaginable.

Identity and access management provides a certain level of protection, but its walls are crumbling and access control and encryption aren't enough. In addition, new regulations such as the General Data Regulation Protection (GDPR) and other international standards are starting to place the responsibility for security breaches squarely on the shoulders of the enterprises that are breached. Finally, hackers aren't giving in, growing up, or going away anytime soon, especially considering the rise of the Internet of Things (IoT) and the security vulnerabilities it represents.

One of the problems CIOs and CISOs face is that enterprise IT security has traditionally been thought of as the IT department's domain and was thus walled off from other areas of the company. However, it's now necessary for security to be integrated throughout an organization's business processes, with input and involvement from all its key decision-makers. Segregated and silo-based security models are failing because they don't scale and are impossible to implement across an entire enterprise.

Organizations need to take a data-centric approach to remain secure and compliant in the modern Information Age.

Enter DCAP and why it matters

Data-centric Audit and Protection, or DCAP, is a means of protecting your organization's data privacy. It emphasizes the security of data itself rather than the security of networks, hardware, or software. One of DCAP's main benefits is that data security can be applied to just the specific pieces of data that need to be protected, making it scalable across the enterprise while having little to no impact on a company's business processes. Thus, it aligns data security with business strategy.

It's now necessary for security to be integrated through-out an organization's business processes, with input and involvement from all its key decision-makers.

DCAP generally involves several processes. These include the ability to know where sensitive data is stored; defining policies for how data is managed within a business context; defending data against unauthorized access or usage; and data monitoring and auditing to ensure that there are no deviations from normal behaviors that would indicate malicious intent.

Gartner identifies five areas or silos where data can be protected through DCAP. They are:

- **Infrastructure as a Service (IaaS)**—Cloud computing to provide virtual computing resources through the internet
- **Software as a Service (Saas)**—Cloud computing to distribute software and host applications through the internet
- **Database Management Systems (DBMS) and Database as a Service (DBaas)**—On-premise or cloud-based systems for creating, retrieving, updating and managing data
- **Big data**—Data sets that are large and complex and are thus incompatible with traditional data processing applications
- **File storage**—Data storage in which large pools of data are stored on the cloud or across multiple physical and virtual locations

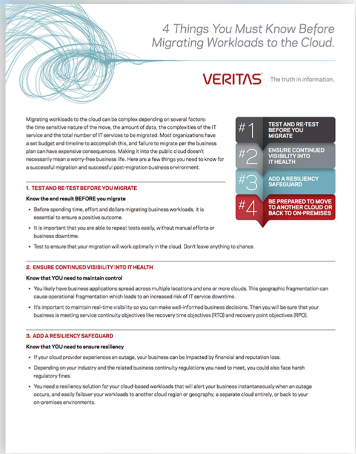
Who are the top players in DCAP, and why?

Gartner recently released its Market Guide for Data-Centric Audit and Protection for 2017. This report is intended for those who are concerned about the privacy of their data. Its purpose is to help them evaluate technology to meet their security needs. It's not a ranking, but is instead a means to identify the capabilities of vendors in the marketplace and indicate the coverage options that each one offers.

According to Gartner, "The exponential growth in data generation and usage across multiple data silos is rendering current data security methods obsolete, requiring significant changes in both architecture and product selection approaches."

To this end, Gartner identifies several security features that should be offered by DCAP vendors for one or more data silos. These features include:

- **Data classification and discovery**—Classify and discover sensitive data in both on-premise systems and cloud-based storage in IaaS, SaaS and DBaaS
- **Data security policy management**—Set, monitor and control privileges of unique user identities (including highly privileged users such as administrators and developers) with access to the data
- **Monitoring user privileges and data access activity**—Use behavior analytics techniques to monitor users when accessing data in real time, generate customizable security alerts, and block unacceptable user behavior, access patterns or geographic access, etc.
- **Auditing and reporting**—Create auditable reports of user access to data and security events with customizable details that can address defined regulations or standard audit process requirements
- **Behavior analysis, alerting and blocking**—Prevent specific data access by individual users and administrators. This may also be achieved through encryption, tokenization, masking, redaction or blocking
- **Data protection**—Provide a single management console that enables the application and orchestration of data security policies consistently across multiple data repository formats



4 Things You Must Know Before Migrating Workloads to the Cloud.

VERITAS The truth in information.

Migrating workloads to the cloud can be complex depending on several factors: the time sensitive nature of the move, the amount of data, the complexities of the IT service and the total number of IT services to be migrated. Most organizations have a set budget and timeline to accomplish this, and failure to migrate per the business plan can have expensive consequences. Making it into the public cloud doesn't necessarily mean a worry-free business life. Here are a few things you need to know for a successful migration and successful post-migration business environment.

#1 TEST AND RE-TEST BEFORE YOU MIGRATE
#2 ENSURE CONTINUED VISIBILITY INTO IT HEALTH
#3 ADD A RESILIENCY SAFEGUARD
#4 BE PREPARED TO MOVE TO ANOTHER CLOUD OR BACK TO ON-PREMISES

1. TEST AND RE-TEST BEFORE YOU MIGRATE

Know the end result BEFORE you migrate

- Before spending time, effort and dollars migrating business workloads, it is essential to ensure a positive outcome.
- It is important that you are able to repeat tests easily, without manual efforts or business downtime.
- Test to ensure that your migration will work optimally in the cloud. Don't leave anything to chance.

2. ENSURE CONTINUED VISIBILITY INTO IT HEALTH

Know that YOU need to maintain control

- You likely have business applications spread across multiple locations and one or more clouds. This geographic fragmentation can cause operational fragmentation which leads to an increased risk of IT service downtime.
- It's important to maintain real-time visibility so you can make well-informed business decisions. Then you will be sure that your business is meeting service continuity objectives like recovery time objectives (RTO) and recovery point objectives (RPO).

3. ADD A RESILIENCY SAFEGUARD

Know that YOU need to ensure resiliency

- If your cloud provider experiences an outage, your business can be impacted by financial and reputation loss.
- Depending on your industry and the related business continuity regulations you need to meet, you could also face harsh regulatory fines.
- You need a resiliency solution for your cloud-based workloads that will alert your business instantaneously when an outage occurs, and easily failover your workloads to another cloud region or geography, a separate cloud entirely, or back to your on-premises environments.

There are many vendors in the Gartner report that offer partial coverage across single or multiple data silos, but organizations may want to select a vendor that offers protection capabilities for all silos through DCAP. There are only two vendors in the Gartner report that meet this criteria: Protegrity and Informatica.

Increasingly, DCAP is becoming a critical component of an organization's ability to protect its most sensitive data. DCAP vendors are quickly adding capabilities, but few vendors have achieved the broad coverage for both on-premises and cloud-based data protection. Regardless, the purpose of DCAP is to integrate data security with business processes for the benefit and protection of the organization as a whole.

By integrating a comprehensive and holistic DCAP solution into your enterprise, you'll be laying the foundation for the most impenetrable walls of the Information Age.

Don’t fall for the “pluggable cloud” siren call

As enterprises build more and more multicloud environments, they want the ability to plug and unplug clouds as needed. But it doesn’t work that way.

BY DAVID LINTHICUM | People once made requests for hybrid cloud because of the perception of flexibility. Now they make multicloud requests, for the same reasons. Multicloud is just part of a cloud architecture that uses more than two clouds, private and/or public. However, most multicloud deployments involve more than two public clouds, typically AWS, Microsoft, and sometimes one other, such as Google.

Although the concept of having “pluggable clouds” is not at all new, I get more and more inquiries about multicloud patterns that promote the notion of pluggable clouds.

A pluggable cloud is a multicloud setup where you can swap out the public or private clouds without having to change much of the underlying application dependencies. The term is often used to describe any multicloud architecture where changing out clouds is something that enterprises do to deal with price and functionality changes.

Is this type of architecture even feasible? Consider the facts.

First, this can only work if you use a cloud service broker (CSB), a cloud management platform (CMP), or other tools that provide abstraction away from the native cloud services. Otherwise it becomes too complex to manage the native cloud services of each public cloud provider, because you have to deal with each native cloud service on its own terms.

Second, you need to understand the “pluggable” requirement. If the expectations are that you can unplug AWS and plug in Alibaba, for example, without significant alterations to how the applications and data storage systems use those services, you’re smoking something. In reality, there are vast differences between how AWS does storage, and how Microsoft or Google or Alibaba does storage. Even if you do a great job creating abstraction and orchestration layers, there is a great deal of work needed to make it actually work. I’m not sure “pluggable” would be the word I’d use.

Third, while it might be possible to make your multicloud setup pluggable, you would do so at the expense of services. You would be forced into a least-common-denominator approach, which means that you’ll only use the basic functions to make your workloads work across cloud providers. In other words, you’ll only use a fraction of what the clouds can offer in terms of services such as storage and compute.

Keep in mind that both CSBs and CMPs are proven tools that can manage multicloud complexity. Just be aware that the use of these tools does not mean you can add and remove public clouds without significant remapping of cloud services to your workloads.

> ***If the expectations are that you can unplug AWS and plug in Alibaba, for example, without significant alterations to how the applications and data storage systems use those services, you’re smoking something.***

IT resiliency and the problem with SaaS: What is your risk profile?

Many organizations are recognizing that cloud providers are more equipped to handle security, maintenance and testing of a cloud environment, so they're offloading more responsibilities to these providers.

BY JEFFREY TON | It's hard to find a company today not using some sort of Software-as-a-Service (SaaS) application for business efficiency. Whether it's for marketing, accounting, sales or IT, all departments find value in the accessibility and ease of these solutions to streamline otherwise complex processes.

However, the ease of use SaaS applications provide the organization may overshadow some of the risks, such as downtime and security. As is the case for any third-party solution, it's key to understand the risk profile both on your side and the provider's end. What does daily use look like? What do you examine?

Knowing the creators and owners

Sometimes the third-party provider who is selling you the SaaS application is not the creator of the technology or platform. If this is the case, what do you know of the creator and how the technology has been created? Consider the security aspects of the design and whether this technology has secured itself, by design, against common threats and ramifications.

What is the IT stance of the service provider? Unfortunately, it's all too common for a company to become overly dependent upon a SaaS application. Then, when the application goes down on the provider's end, the company is left helpless. Your company should have an IT resiliency plan, and the provider should also have one. It's key to check that they do.

Knowing who has access

Consider who can share information and how it is shared. Especially if your organization is under a regulatory framework, be careful that such actions don't risk a breach of compliance. For example, if your organization uses Google Docs or Dropbox to share and store information cross-functionally, what sort of information is shared and what could happen if that information is exposed? If there is a breach or disruption to your SaaS application, what happens to your data? Does the provider keep any of your data transferred through their technology? If so, for what reason?

There may not be any right answers here—it's just important

The first step to closing security gaps is understanding that using cloud—any cloud—means you are using a third-party provider in some way.

to know exactly what level of risk your organization is putting itself under. If you identify any unnecessary avenues for sharing information, close them off. Best to be careful, rather than risk data loss and exposure.

What contractual aspects exist around who retains what data from your organization? Can the SaaS provider limit your access to your data due to a dispute? What if they go out of business? It's paramount to put everything in writing, even service level agreements (SLAs).

Knowing who owns what responsibilities

Make sure your IT Disaster Recovery (DR) plan delineates between you and the SaaS provider the order of operations for execution and what to do if personnel are missing. Note all responsibilities in your DR playbook and share it with the provider. The best way to understand these intricacies is to test for a disruption scenario, executing your DR plan to be sure you can actually recover in a real event. Keep your SaaS provider abreast of your testing and ask to see their testing results.

Does your SaaS provider have a DR plan? Where is your data stored, exactly? How many of your SaaS applications share a common backend datacenter (AWS Zone 1, for example)? It is important to understand the architecture, so you have a complete picture of your risk profile.

Welcome to the age of the cloud

The widespread use and accessibility of the cloud has transformed the expectations of day-to-day business. Customers demand an always-on marketplace, which means the cloud isn't something that's going away. Now downtime is no longer an option, as it impacts the bottom line and reputation. For this reason, it's best to embrace the cloud, but to do so cautiously to ensure everything remains secure against any and all threats. Never make a leap to the cloud without first laying out a strategy to get there (here's a Forbes article I wrote on this subject).

Oftentimes, organizations that own the deployment and management of their cloud prefer to do so in their own datacenters, but this could present its own set of risks. The first step to closing security gaps is understanding that using cloud—any cloud—means you are using a third-party provider in some way. For example, if you have an AWS cloud, you're in business with Amazon.

Cloud providers usually have the resources and bandwidth to properly manage the essential maintenance aspects and testing, given that this is their core business model and, thus, they perform work in the cloud all day every day.

A small-to-medium-sized business, on the other hand, may have only a limited number of IT team members and resources. And these personnel may have other, more pressing responsibilities.

For this reason, many organizations are recognizing that cloud providers are more equipped to handle security, maintenance and testing of a cloud environment, so they are offloading more responsibilities to these providers with the goal of getting out of the datacenter business altogether.

2017
Truth in Cloud
Report

The rise of multi-cloud:
Combatting misconceptions
and realigning data
management responsibilities

VERITAS
The truth in information.