Simplifying the Complex World of Data

Whitepaper

We make IT work

Simplifying the Complex World of Data

Worldwide data will grow from 33 zettabytes this year to 175 zettabytes in 2025, a compounded annual growth rate of 61 percent, according to IDC.

New technologies will generate even more data; IoT devices are on track to create 90ZB of data over the next six years. This data is not just spread across departments but also multiple locations and with 81 percent of organisations using multiple cloud providers, data is becomingly increasingly complex to manage.

The challenge of data storage & backup

Data needs to be stored, protected and accessed on demand, allowing its transformation into information and ultimately intelligence. Our organisations deal with heavy workloads of business-critical information, so enterprise storage systems have to be scalable for workloads of hundreds of terabytes or even petabytes.

Code sprawl has contributed to the need for highly dense and scalable storage systems; including high-performance computing storage, converged infrastructure, hyper-converged storage infrastructure and scale-out and scale-up network-attached storage (NAS).

Underscoring the importance of enterprise data storage is the move towards big data, enabling organisations to make better decisions. As our data has grown so has the number of locations - we store it in-house as well as multiple cloud repositories such as Amazon Web Services (AWS), Google Cloud and Microsoft Azure.

More data brings the need for more backing up. Also the growing volume of data amplifies the age-old challenges, such as finding a suitable back up window. Increasingly large files also brings new challenges to restoration; for example, in 2008 an average mail box was 250Mb now Office

Data governance & compliance

Every year, businesses face new and complex compliance requirements. Passed down by government (or governments if they are operating in multiple regions) and regulatory agencies. These rules govern the conduct of businesses take to how data is stored, managed and protected.

From GDPR to Sarbanes-Oxley, there is now a regulatory environment that storage administrators need to be aware of and comply with. Data retention is one of the most crucial aspects of ensuring an organisation is compliant. This is complicated by different kinds of data being subject to different types of retention periods, causing many businesses to err on the side of caution by retaining large swathes of data “just in case” it is needed.

The increasing complexity around storage and retention, combined with ever growing volume, makes keeping on top of the data in a business an immense challenge. Without full insight into what is in it and where it is, a businesses data lake is at high risk of becoming a data swamp.

No company ever sets out to make a data swamp. Data lakes can deteriorate and become swamps unless an organisation makes and sticks to plans for regularly reviewing and cleaning their data. The start point of this is knowing what data there is within the organisation and where it resides; whether that is in-house or with third parties. Only then can the organisation decide what data it needs to keep, it can identify whether it needs to retain certain data for compliance reasons or whether it is currently holding data that is no longer needed.

Data governance & compliance

Every year, businesses face new and complex compliance requirements. Passed down by government (or governments if they are operating in multiple regions) and regulatory agencies. These rules govern the conduct of businesses take to how data is stored, managed and protected.

From GDPR to Sarbanes-Oxley, there is now a regulatory environment that storage administrators need to be aware of and comply with. Data retention is one of the most crucial aspects of ensuring an organisation is compliant. This is complicated by different kinds of data being subject to different types of retention periods, causing many businesses to err on the side of caution by retaining large swathes of data “just in case” it is needed.

The increasing complexity around storage and retention, combined with ever growing volume, makes keeping on top of the data in a business an immense challenge. Without full insight into what is in it and where it is, a businesses data lake is at high risk of becoming a data swamp.

No company ever sets out to make a data swamp. Data lakes can deteriorate and become swamps unless an organisation makes and sticks to plans for regularly reviewing and cleaning their data. The start point of this is knowing what data there is within the organisation and where it resides; whether that is in-house or with third parties. Only then can the organisation decide what data it needs to keep, it can identify whether it needs to retain certain data for compliance reasons or whether it is currently holding data that is no longer needed.

How Did We Lose Control of Data?

The level of sophistication around data and data storage has increased significantly over the last 20 years. There used to be little or no governance, minimal quality and what existed wasn’t well managed. Howeve every organisation has had to transform itself and now there is a new level of transformation required. It should be no surprise that if you do not have the basics like data quality under control, then whatever the expected outcome for the analytics is not going to be delivered.

It is not data volumes that pose the biggest challenge, but what is hidden in the data, both the good and the bad. Businesses cannot expect to adopt AI and machine learning and harness its benefits to gain a competitive advantage unless they have clean data to start with. In machine learning, data is the oil. Businesses face challenges in feeding the right data to machine learning algorithms or cleaning of irrelevant and error-prone data. But before they can set about creating a reliable data set, they need to locate the relevant data across the business.

Companies have three sets of data.

1. Business critical data; this is the data that the organisation needs to function.
2. Redundant data; this is the data which may be obsolete or trivial.
3. Dark data; this is the data that an organisation has no visibility of.

A company should only really be focusing on the first set - the critical business data - but policies and practices within their organisations mean that the value of the data is not considered when choosing to store it and back it up. The flexible, low cost nature of cloud storage has led to companies relaxing the policies on what that are keeping. However, this is leading to the retention of unstructured data, such as personal photos and videos or even personal IDS and legal documents, which will lead to compliance issues.

Data silos still exist in most organisations even after decades of trying to break them down they remain, impeding the free flow of data and restricting analytics. Silos rarely exist by design. They come about through legacy systems managed by people who believe they are the only ones that can keep them up and running. Many exist because of acquisitions and mergers never being fully integrated or they may exist on an office by office, country-by-country basis. But in the modern data storage environment they mean that it is hard to get a single view of what data a business has. This means that a company cannot make the decisions it should do about that data and it cannot make decisions based on the information held within that data when sometimes it does not even know that the data exists.

Lack of organisation wide oversight can lead to data, especially archives, to become orphaned and eventually forgotten altogether.

More data sources are being added almost daily so businesses have to deal with heterogeneous file data. With the coming explosion of IoT devices, which store very little data locally, this is going to only grow larger. Data captured by IoT devices is produced in a mix of data formats, including structured, semi-structured, and unstructured data. This data includes discrete sensor readings, device health metadata, or large files for images or video. There are over 200 image and video file formats alone. An organisation could deal with well over 1,000 file types which it will need to identify, manage and make decisions on and from.

The move towards increased use of the cloud for storage comes at a time of significant tightening of data protection regulations. Not only has the European Union's General Data Protection Regulation (GDPR) come into force, but other countries are introducing their own new regulations to follow suit which has prompted businesses to review how they gather and process information.

For compliance, the key issues facing Chief Information Officers are what types of data the organisation stores and where that data is. Once businesses ran their own in-house databases, archives and storage systems and were in a position to identify the location of their data. But the move to external locations for data storage and IT workloads and multi-vendor clouds has created a new challenge, businesses cannot outsource responsibility for compliance. Complying with laws such as GDPR and penalties for breaches falls squarely on the business, not the cloud vendor. ICO commissioner Elizabeth Denham has said "The law is clear – when you are entrusted with personal data, you must look after it. Those that don't will face scrutiny from my office to check they have taken appropriate steps to protect fundamental privacy rights," she cautioned.

Multi-vendor clouds also require a business to understand what data is with what cloud vendor if they are to avoid unnecessary replication of data and additional cost. There is also the issue of vendor lock in to contend with. Microsoft Azure Blob storage, for example, is different from Amazon AWS storage, so data is not transferable from one to the other. Running a multi-vendor cloud means addressing issue of not having a single view of data to enable optimisation of workloads, with each vendor having their own protocols, APIs and tools to manage data as well as the business onsite storage.

Storing data in the right locations can mean avoiding the cost associated downloading data from cloud storage. In most cases these fees are be nominal, however, if a business has multiple instances or needs to restore data after a disaster these fees can quickly become significant.

Cloud storage creates an issue with data management complexity relating to the number of storage devices or separate storage environments that a business must manage. It can also mean greater risk. A storage environment that is not effectively managed is more susceptible to data loss or data breach. This management is imperative because each account on a public cloud is a separate storage environment. A business may have hundreds of storage environments on a single cloud. It is possible that the IT department do not have visibility of some of these instances if they have been set up outside of their control, creating shadow IT, this introduces risk of a data breach if the in instance has been misconfigured.

Organisations of all sizes have an issue with duplicate data and data being saved by employees that has little or no value to the company. Many save more data than they will ever use, storing it in case it is needed one day. Their rational is "storage is cheap, so why not store as much data as possible?". However, when data is held without due care and attention it can become a security, privacy and compliance liability as well as opening up the business to uncontrolled costs.

An example of this is a company where a user had uploaded a ripped movie to a company-wide accessible folder on a tier-zero flash storage array. Two other colleagues copied it onto their home directories and a fourth copy was accessible to the internet. Each copy was backed up weekly and retained for seven years as per company data policy, resulting in that initial 5.5GB creating a terabyte of data over the seven years. The business estimated that this would have cost them over £8,000.

How to Get Visibility of Your Data

The IT team needs to protect data across an increasing complex environment. Yet, the lack of visibility across heterogeneous storage and back up platforms means that it is quite common for them to not know what they are responsible for. This does not need to be the case.

The IT team can take back control by getting a single pane of glass to view all the data stored and backed up wthin their organisation.

With data privacy and regulatory compliance becoming one of the most challenging aspects a company has to face, the ever-increasing amount of instructed data makes managing this a daunting and time-consuming undertaking, if it is even at all achievable. To tackle this challenge, it is essential to be able to first find and then filter specifics about what type of information is contained within a business's data infrastructure. Understanding the classification of data will help the administrators make informed decisions and take action to ensure that the right information is secure and available when needed.

It is possible to get a single view across the whole environment, in a dashboard that provides details about the data collected from both cloud and on-premise data sources and filters including location, content sources, owner, file type, sizes, extensions and classification tags.

This technology will interrogate the environment and uncover what data is stored in each location. If a business has a multi-cloud solution using AWS S3 and Microsoft Azure Blob storage and NetApp, also running Exchange then they can still get an aggregated view of their data all in one place.

SCC's ManageSMART Information Intelligence service allows a business to visualise their data, allowing authorised uses to dive into any location, file server, share, owner, classification or a combination of filters. It ensures data attributes such as type, ownership, access and age and unique classification tags are accessible and simple to understand.

The service addresses key metadata indexing, and its information classifier enables organisations to find the information held within their data infrastructure. The importance given to any piece of data will be dictated by the business and the regulations that it needs to comply with such as DPR and HIPPA. The solution gathers common file metadata information like file age, file type and file location and it can also provide much richer data by identifying information within your files that is traditionally more difficult to find out. By creating filters, a business can query data based on specific combinations, such as the file age and if it contains Personally Identifiable Information (PII).

When determining what information is important, administrators need to understand the details about regulatory requirements. Each of these has its own complex set of specific conditions that determine what information applies.
This process is simplified by using an information classifier which includes over 700 pre-configured data classification patterns and over 110 policies to help identify common data privacy and regulatory compliance principles. These include everything from U.K. Unique Tax Reference (UTR) to U.S. State Regulations—Criminal history, FCRA, FERPA, FFIECE, FISMA, IRS 1075 or SE.)

As one of the company's most powerful and riskiest assets, data cannot be utilised unless it can be organised. Once the data is located, a business can make evidence-drive decisions about it. For example, deleting stale data and data minimisation to comply with regulations.

All this can be achieved by deploying a virtual appliance on premises and it will enable clear visibility, targeted analysis and informed action on data to identify areas of data waste, risk and value. It gives the power to organise data and take informed action. Businesses can be confident that they are prepared to handle security concerns, new regulations and continuous data growth to regain control of their data.

To create a full data management strategy a business also needs a single pane of glass to give insights across any cloud, any storage, and any data protection solution enabling them to optimise storage cost, mitigate risk and streamline backup compliance across on-premises and hybrid clouds. Using the solution's management console allows businesses to view insights like cloud infrastructure utilisation, storage optimisation dashboard, chargeback, backup SLA status.

Take Back Control of Your Data

A lack of visibility in a storage environment, especially one that includes multi-cloud, is a business risk for security, compliance and governance, but it can also affect business costs.

SCC's ManageSMART Information Intelligence service gives a business full visibility of their data right across all their data stores. It will turn data into information and ultimately intelligence.

Working with SCC an organisation can identify if it currently has an issue with visibility of data, the SCC team will design a solution that resolves this issue, this may include rationalisation of cloud vendors or even introducing new storage and then SCC can deliver and manage this solution.

In a time where the volumes of data storage is about to grow expotentially lack of visibility will lead to increased uneccessary cost and potentially non-compliance. It is not an option.

Speak to one of SCC ManageSMART Specialists and start to take back control of your data.

SCC
James House
Warwick Road
Birmingham
B11 2LE

0203 451 8351
online@scc.com
scc.com

We make IT work