Your Role in GDPR and Data Protection

NetApp Solution Overview

An Architecting IT White Paper
October 2017
Chris M Evans

architectingIT

GDPR AND YOU

CONTENTS

Executive Summary

The adoption of GDPR (General Data Protection Regulation) by the EU in 2018 will represent a huge challenge to any business that currently stores and uses personal information. GDPR introduces much stricter rules and penalties on how information used to identify an individual can be stored and processed. The EU has moved to introduce new rules that put the individual firmly in control of their own personal information and as a result introduce significant challenges for CIOs and their IT teams in reviewing the way in which data is managed by their organisation.

As businesses look at how they can build out a GDPR strategy, storage and data management vendors need to step up and assist with hardware products and solutions that are more aligned to the needs of business in meeting their GDPR commitments. As an existing solutions partner to thousands of existing IT teams, NetApp already offers a broad range of storage products and solutions. Increasingly, this portfolio is being re-engineered for a data management rather than data storage perspective.

At the heart of NetApp's evolution is a concept called the Data Fabric, a strategy to bring all the existing solutions together in a data-centric ecosystem. When the Data Fabric concept is fully realised, information will be fully portable across multiple hardware platforms and the public cloud, while meeting the needs of customers in terms of cost, security and operational efficiency. Some of this functionality will be delivered by NetApp directly and some by partners.

In terms of developing a strategy for GDPR, NetApp is evolving to put data, rather than storage at the heart of their hardware and software offerings. Data is the core asset that needs to be managed with GDPR oversight and NetApp is well positioned to meet the needs of their customers as the legislation comes into force and as the IT industry moves ever forward to a hybrid world.

Background

The modern world is increasingly conducted online. Businesses collect and store information on their customers; we use email and social media platforms daily and many people choose to share all their personal moments with potential strangers. We have come to rely on the Internet as our source of all knowledge and as a place to conduct transactions from banking to ordering food, clothing and anything else Amazon and other online retailers could possibly consider selling to us.

The core of this world is based on data. Personal data about you and me that identifies what our habits are, what TV and music we like and where we choose to holiday. Businesses track every engagement with customers, from keeping credit card details on file to make that next purchase easier, to looking at how we all interact with websites and online shops.

Look back 20+ years to October 1995 when the EU introduced the Data Protection Directive and we see a very different world. There was no Google (the company wasn't founded until 1998). There were no smartphones. Facebook, Twitter, Instagram and Snapchat weren't even a dream in someone's imagination. At the time, the Data Protection Act in the UK sought to ensure that any personal data was collected and managed in accordance with strict principles. However, there was no provision to cater for ensuring organisations disposed of data once they were finished with it, or for the owner of that data to determine how it should be used.

Across the world none of the data protection regulators could have imagined that in only two decades we would be collecting data on people's movements through GPS and using that information to schedule taxis, provide driving directions or even to work out how busy shops are. Even if that data didn't specifically identify an individual, modern AI or machine learning can work out a person's identity, given enough information.

Data or information has become the core asset for the majority of modern businesses. Individuals rely on that information being accurate and securely held, at arms' reach from hackers or others who would seek financial gain from exploiting data breaches. As a result, data protection legislation across the EU needed to be harmonised and brought up to date, with additional provisions that protect the end user, putting them in control of their own data assets.

As we move into this new world, the systems used to store, track and protect data become ever more important. Businesses need the ability to identify personal data, store it securely and ensure that any breaches of access can be contained and measured. Data management vendors, including NetApp must move forward and focus on data rather than storage management, offering tools and solutions to deliver to personal data management for the 21st Century.

GDPR What It Means to You

General Data Protection Regulation or GDPR is currently being introduced across the European Union and enforcement applies from 25th May 2018. The aim of the new regulation is to standardize, improve and strengthen data protection laws and bring them into a modern context that more accurately reflects the way in which data is used by businesses and other organisations. GDPR replaces the previous Data Protection Directive (DPD), which used seven principles to determine how personal data should be managed. For individual member EU countries, the previous directive needed to be incorporated into local law. However, the regulations don't require local legislation and so will automatically come into force across the whole of the EU from May 2018. EU Member States are still able to amend regulations through the use of "opening clauses" – more on this later.

GDPR Evolution

GDPR updates and improves on the data protection directive in several ways that reflect the change in the use of technology since DPD was first implemented in the mid-1990s. The first is in the definition of personal data. Previously the definition was anything that could directly identify a "natural person" and so comprised common concepts like name, images and photographs, email addresses, phone numbers and so on. Today, individuals can be identified in many more ways, such as via IP addresses, mobile IMEI and SIM card IDs, through website cookies and a range of biometric data.

Collection of this range of information allows businesses to build up profiles on individuals, which is a step further from the idea of simply being able to identify an individual person. This data can be used to predict how individuals might act, what their buying preferences are and if used maliciously to exploit their financial standing. GDPR article 22 specifically allows profiles and the data processes (algorithms) used to make decisions based on this data to be contestable by the individual. This means that businesses need to be concerned not just with getting consent from individuals to collect their data, but agreement that any data processed can be used in automatic processing – not just what but also "how".

Further strengthening on the position of the end user or consumer are being introduced. Individuals will have the "right to be forgotten" under Article 17, a provision that was previously implemented by some Member States but now is consistent across the EU. This means that unless a business can provide a legitimate reason to retain an individual's data, the individual can request that their data is erased by the business, without "undue delay". We'll discuss this point in more detail later, however the implications from a technical perspective alone will be very challenging for most IT organisations.

Introducing the Data Processor

Under DPD, personal data was under the management of a Data Controller, typically a business collecting and using data. GDPR extends the definition with the introduction of a Data Processor (Article 28), which can be another organisation or individual that processes data on behalf of the Data Controller. The interesting extension of the definition here is in how the changes will apply to the use of public cloud. In effect, public cloud providers could be classed as Data Processors and therefore be required to meet the same strict requirements in managing data applied to Data Controllers.

Change of Scope

GDPR now applies to any organisation within or outside of the EU that is processing the data of any EU citizen. This greatly expands the scope of the regulations compared to DPD, effectively making any organisation that interacts with the EU liable to follow the terms of GDPR. This means almost any business needs to have a plan in place to meet GDPR, whether they are directly trading in the EU or not.

Penalties

Under DPD, individual member states could adopt their own rules regarding breach notification and penalties. With the adoption of GDPR, the rules for reporting a breach have been standardised across all EU countries. Any breach must be reported to the local supervisory authority within 72 hours of the organisation being aware of it. Failure to notify can result in a fine of up to 10 million Euros or 2% of global turnover from the previous years' business. Penalties increase for negligent or intentional violation of GDPR of up to 20 million Euros or 4% of turnover.

It's clear to see that the introduction of GDPR places a significant additional burden on businesses to understand the data they collect and manage, but also to adequately safeguard that data and see it as a resource that they have a privilege to use, rather than a right to own. The individual gains significant rights, which businesses, through their IT organisations, must honour.

GDPR as a Baseline

We mustn't forget that in lots of ways GDPR is merely a baseline of data protection standards that provide harmonisation across the whole of the EU. However, individual member states can apply additional stringent rules to data protection and many do. Typically, we see the use of this data adding extra protection for the customer/consumer, while creating headaches for the business. For example, several countries don't permit any personal data to be taken out of country. In Germany, regulations on data protection can also be implemented by individual

regions, making the task of compliance extremely complex. New GDPR rules will allow the use of "opening clauses", where individual Member States have the option to modify the provisions of certain GDPR articles. This allows more restrictive rules to be introduced and could be used, for example to bring in tighter rules on the management of employee data. What this means is that although the overall standards of data protection will increase, businesses will still have to honour individual Member States rules.

Implementing GDPR

So how does any business go about building a strategy for implementing GDPR and other data protection policies? There are a few key attributes that need to be followed.

Know Your Data

Do you know what and importantly where all your data is? Look back 20 to 25 years and almost all IT systems were installed in corporate data centres or perhaps bureaux or co-location services. Today data can be spread across multiple data centres, multiple geographies and the public cloud. We collect and access data across many different devices; including mobile phones and tablets.

Remember that data doesn't just refer to information stored in primary systems. Personal data is also kept in backups, replicated copies for BC/DR, snapshots and any additional copies taken to seed test systems. While we would expect these copies to be cleansed to obfuscate obvious data like names and addresses, the process may not extend to other content and so still make an individual identifiable.

This fluidity in having data in many places provides significant business advantage to reducing costs and delivering efficient services, however it also introduces risk because data assets are so dispersed.

Identification of data doesn't just mean knowing what storage platform it exists on. Data in systems needs to be mapped to usage, based on the application and information provided by the business. In some instances, this process can be automated but in most cases, will require the business to provide information on what level of personal data is being stored.

Many public-facing websites will collect data on global customers. Across the EU Member States, this means keeping data on individuals' locations geographically, as rules may be implemented more strictly in some countries than in others. The result is a requirement to apply multiple rules at varying levels, depending on the

individual. Efficient storage management will therefore form a basis for efficient data management processes.

Identifying the Individual

As we've already discussed, the range of data that can be used to identify a single individual is now wide-ranging and complex. However, there are two main ways to classify personal information. There is data generated by the user themselves. This can include information created on mobile devices, but also data collected as part of interactions on websites, as part of telephone calls or other IT systems. There is also data generated on behalf of a user. This type of information covers scenarios where the data on an individual is collected by a third party and could be for data entry of information on forms, logging service and helpdesk calls.

We can see that the two examples could comprise endless scenarios, making much of the information generated by a business potentially classified as including personal data. This means businesses need to consider applying best practice management to all their data, including encryption (either at rest or in flight), implementing strict access policies and auditing access to content as much as possible.

Data Management from NetApp

NetApp has been providing storage and data management solutions for over quarter of a century. Initially the first platform sold by the company was a hardware appliance running a custom storage operating system called Data ONTAP (now branded as ONTAP). The ONTAP platform pioneered the concept of a filer or dedicated file server for storing unstructured file content.

Today NetApp offers a wide portfolio of hardware solutions, software defined storage platforms and a suite of software tools that together deliver a comprehensive data management strategy. With a view to GDPR, how does this portfolio of products help customers ensure they can meet the upcoming compliance regulations?

A Platform for All Requirements

Modern IT applications use many different data platforms across multiple storage protocols, including traditional databases on block or file systems and unstructured data in object stores. NetApp has technology solutions that can be deployed as dedicated appliances, as software-defined or in public cloud.

Table 1 - NetApp Platform Offerings

Platform	Block	File	Object	Appliance	SDS	Cloud
ONTAP	✓	✓		AFF/FAS	ONTAP Select	ONTAP Cloud for AWS, Azure & NPS
SolidFire	✓			SF Series	Element X	
E-Series	✓			E2000/E5000		
StorageGRID			✓	SG5600 Series	VSA & Containers	
AltaVault		✓		AVA400/800	VSA	Cloud Appliance

With a common storage operating system base, platforms such as ONTAP enable data to be moved between physical locations using existing well-tried and efficient tools such as SnapMirror (more on this later). Running a virtual storage appliance in the public cloud may seem like an expensive choice simply to gain the benefit of efficient replication however the benefits are much more than that.

The ability to manage data is based on having good metadata – information describing the content. Having a common underlying architecture means that the metadata managing the content is consistent and shareable across all deployments. This enables both efficiency (for example, only moving changed data) and consistency – keeping track with data updates across multiple locations.

IT teams can also benefit from consistent use of platform features, such as those used for data protection, providing a consistent implementation of business continuity/disaster recovery (BC/DR) policy, regardless of where the data is located.

Looking at the hardware portfolio from a GDPR perspective, there's no intrinsic benefit in the way one system works compared to another, however as NetApp introduces interoperability between systems, then metadata can be retained with data as it moves between platforms, ensuring GDPR attributes are retained against data copies, snapshots or information that is otherwise moved around the infrastructure.

NetApp Private Storage

One particularly interesting solution for the public cloud is NetApp Private Storage (NPS). NPS deploys traditional ONTAP hardware in co-location sites close to public cloud storage providers and connects directly into the network. This provides low latency private storage that can be connected to cloud compute instances. NPS provides a number of benefits for the customer:

- Data is managed outside of the scope of the cloud provider on dedicated hardware, enabling the customer to better meet their data compliance rules. This includes the ability to encrypt data at a disk and volume/LUN level.
- NPS offers higher performance capabilities than shared storage from the cloud service provider, with the ability to both meet the requirements of demanding workloads and implement more granular levels of quality of service.
- Enable the Implementation of higher levels of resiliency. ONTAP already provides high levels of resiliency, however NPS can be integrated with other services such as SnapMirror to deliver unique cloud-to-cloud availability.

At first glance it may seem excessive to deploy a hardware solution to complement public cloud. After all, public cloud is about buying IT services, not infrastructure. However, the very definition of cloud implies that the cloud provider exposes no access to the infrastructure used to deliver the solution and as a result the customer can only use existing interfaces to store, retrieve and search data. This also applies to the metadata associated with volumes, file shares and objects.

The use of NPS provides customers more control over their data assets and it's not surprising that the use of private storage connected to public cloud is now offered from other vendors. Today NPS is available on Amazon AWS, Microsoft Azure and IBM SoftLayer/Bluemix clouds and delivered in conjunction with close to 300 global partners.

For NetApp customers concerned about data sovereignty (both geographically and in multi-tenant deployments), NPS offers a compelling strategy that leverages existing skills and the ecosystem of processes built up by the customer, while retaining the capability to migrate to and from public cloud with ease. As previously discussed, data stored on NetApp systems can be assigned metadata tracking usage and compliance rules, which is potentially made harder with data in the public cloud.

StorageGRID Webscale

The majority of data growth continues to be in unstructured content and is expected to continue to increase in the share of data in the enterprise as we move into a world where more content is machine generated. StorageGRID Webscale is a software platform that allows the geographically dispersed storage of objects (such as files and media) across multiple solutions including virtual appliances (VMware vSphere, OpenStack), containers (Docker) and on NetApp appliances.

Key features of the platform include:

- **Global Name Space** – the ability to see a single view (and single security model) of object storage across all public and private data endpoints.
- **Policy-driven data placement** – the ability to move data between physical tiers of storage and locations, based on business-focused policies.

- **Support for S3 and SWIFT protocols** – S3 has become the de-facto standard for object storage and is rapidly becoming a standard pathway for data exchange.
- **NAS Protocol Bridge** – provide the capability to store and retrieve the same data across both file and object protocols.

From a GDPR perspective, StorageGRID provides for the implementation of a single view of object data across all data centres, with a single security model applied to the data. The implementation of multiple protocols to access the same content reduces the need to replicate data between platforms, reducing the risk of having to maintain multiple copies of data.

Using policy-driven data placement, data can be moved between physical geographies (as well as the obvious range of storage tiers) in order to comply with GDPR rules. This could be implemented by storing data in separate pools or buckets that match geographic boundaries.

Storage Management Tools

Data management requires robust tools for deploying, monitoring and managing both storage infrastructure and the data being stored. NetApp has brought together a suite of tools under the OnCommand umbrella that provide the capabilities of managing both on and off-premises resources.

OnCommand Insight (OCI) has evolved from an infrastructure management tool to a full analytics-based solution for optimising hardware resource utilisation. This enables customers to better manage resource consumption and make more informed purchasing decisions about the technology they need in the data centre.

As a support tool to GDPR, OnCommand Insight helps NetApp customers ensure that data is highly available and all storage resources identified and reported on. Insight also extends to the public cloud, offering hybrid resource management for Amazon and Azure storage products.

There is still work to do in this area, as OCI is still very much infrastructure-based. The scope of data collected needs to be extended to cover data within applications – looking at database platforms and how this translates to physical storage.

The Data Fabric – Bringing it All Together

In 2014, NetApp introduced the concept of the Data Fabric, an architecture and strategy for managing an organisation's data rather than physical storage on which the data sits. The basis of the Data Fabric addresses the idea that the core asset of a business survives and outlives the hardware on which it is stored. Technology

solutions therefore need to work together to manage data as a single entity, rather than dealing with data in a fragmented and disjointed way.

At the outset, the data fabric from NetApp was a concept and a goal for bringing together data assets in a way that allows them to be managed more consistently. There are already product features that enable this. SnapMirror for ONTAP as one example is a data replication tool that allows volumes and file shares to be replicated between physical and logical infrastructure. As a protocol, SnapMirror is highly efficient as it tracks data updates through metadata, enabling data consistency to be maintained by only shipping changed blocks between systems. As a single feature, SnapMirror in itself doesn't implement any GDPR compliance, however what it does do is provide a method for ensuring any replicated or cloned data is still associated with the original application and any GDPR compliant assessment that has already been done.

As part of the data fabric initiative, SnapMirror is being extended to allow data from SolidFire appliances to be replicated to ONTAP systems, including hardware FAS appliances, ONTAP Select and ONTAP Cloud. This is one step in implementing data mobility and reducing the boundaries between NetApp and 3rd party hardware platforms.

At this stage is it probably fair to say that the NetApp Data Fabric is in a rapid state of development and currently implemented by a set of discrete tools and solutions. The platform has a number of specific goals, that meet the upcoming needs of GDPR. These include data security, the ability to deploy on multiple physical and cloud storage platforms and support for a range of application platforms, including server virtualisation and containers. Data Fabric technologies and solutions that exist today include;

- **ONTAP Cloud** – an instance of ONTAP running in public cloud, with the same features and functionality as the commercial appliance version.
- **NetApp Cloud Sync Service** – a tool that enables unstructured data on-premises to be kept synchronised with a copy in Amazon S3 (Simple Storage Service)
- **AltaVault Cloud Integration** – the ability to take data from a range of existing primary storage solutions and store in both public and private cloud using NetApp AltaVault.
- **Cloud Control** – the ability to backup and restore data held in Storage as a Service cloud offerings, such as Microsoft Office 365.

With such a huge existing install base, it would be unreasonable to expect NetApp to transform overnight to a Data Fabric solution that fully integrates with every existing hardware platform and software solution. However, great progress is being made, with the promise of much more to come.

Summary

Data management as a discipline is clearly different from storage management, which is focused more on physical assets. NetApp has a strong background in storage management, with (as discussed) an increasing portfolio of products and solutions.

As the IT industry embraces cloud, physical resource management becomes less important and the focus changes to managing data because the (public) cloud provider takes responsibility for the hardware. Even in private cloud environments, the hardware is being sufficiently abstracted from the data to allow the two to be treated as separate functions.

The move towards data management is a gradual one, that for NetApp represents a developing transformation of the business. Existing customers can't be left behind, however the needs of hybrid infrastructure needs to be addressed.

The work to date on projects such as the Data Fabric represent a transition to data management. NetApp provides the framework onto which certain features can be integrated (for example OCI), however many parts of the data management story will be delivered in conjunction with partners and that means exposing capabilities of the hardware platforms through which partners can deliver added value.

No other vendor has refocused their business on addressing the data needs of the customer and as such, NetApp stands out once again as a leader in transforming the storage industry.

More Information

For additional technical background or other advice on replication technologies, contact enquiries@brookend.com for more information.

Architecting IT is a brand name of Brookend Ltd and independent consultancy, working for the business value to the end customer.

Email: architectingit@brookend.com
Twitter: @architectingit

The Author

Chris M Evans has worked in the technology industry since 1987, starting as a systems programmer on the IBM mainframe platform. After working abroad, he co-founded an Internet-based music distribution company during the .com era, returning to consultancy in the new millennium. Chris writes a popular blog at http://blog.architecting.it, attends many conferences and invitation-only events and can be found providing regular industry contributions through Twitter (@chrismevans) and other social media outlets.