White Paper

NetApp Data Fabric and the Essential Data Security Controls for Hybrid Clouds

By Doug Cahill, ESG Senior Analyst, Cybersecurity

March 2016

This ESG White Paper was commissioned by NetApp and is distributed under license from ESG.

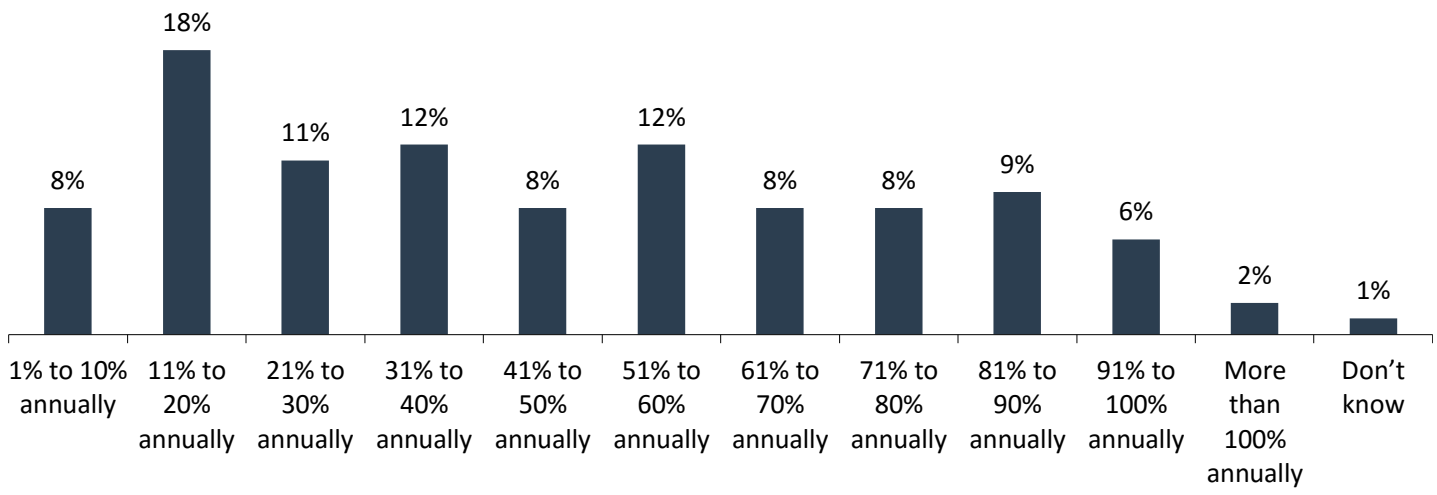
Contents

The State of Data in a High-risk World

There is no question that the digital data explosion has created a wealth of benefits for businesses, but the corollary to this rapid data growth is the ever-increasing challenge to secure and manage all that data. New data sets are being merged with preexisting ones for big data analytics, providing business intelligence leveraging new forms of organizational intellectual property to improve short-term decision making, while guiding longer-term strategic direction. With these trends in mind, ESG recently conducted a survey of 200 North American-based IT professionals with day-to-day data storage responsibilities at midmarket and enterprise organizations on the topic of securing the data volumes that they are tasked with securing, regardless of location.

Software-defined data centers allow for a level of agility that clearly provides competitive advantage, and emerging trends, such as the Internet of Things (IoT), create additional data sets to secure and manage. Indeed, ESG's *2016 IT Spending Intentions Survey* reveals that managing data growth is—once again—one of the five most commonly cited overall IT priorities for 2016.[1] So at what rate is storage capacity increasing annually in an attempt to keep pace with these continually growing volumes of data? As seen in Figure 1, the plurality of respondents (18%) cited overall annual storage capacity growth rates in the 11% to 20% range, though nearly half (45%) reported storage capacity growing in excess of 50% per year.[2]

FIGURE 1. Annual Data Storage Capacity Growth Rate



Source: Enterprise Strategy Group, 2016

Not only are diverse types of unstructured data growing at a phenomenal rate, but data is being created and stored in an increasingly distributed fashion—across on-premises infrastructures, colocation centers, and public clouds—which further complicates the security challenge. The totality of these distributed environments represents today's hybrid cloud, the new normal of the modern data center. To manage these disparate infrastructures, IT organizations require the ability to arbitrate data location across these storage locations while ensuring the security of the data both at rest and in flight.

[1] Source: ESG Research Report, *2016 IT Spending Intentions Survey*, February 2016.

[2] Source: ESG Research Survey, *Security Considerations for Storing and Managing Data*, January 2016. All ESG research references and charts in this white paper have been taken from this research survey unless otherwise noted.

Additionally, organizations have been implementing storage tiering for a long time, adding more complexity to data management as well as applying security policies. And the use of online, nearline, and offline storage tiers as architected to prioritize the placement of increasingly diverse data sets now includes cloud-resident storage, adding another dimension to the security imperative.

It is important to recognize that all these types of data are assets, often representing substantial intrinsic value to their organizations, and they are at risk of loss by virtue of being exposed to insidious cybersecurity threats. Acting upon the need to secure data assets from compromise in today's hybrid cloud norm is further complicated by the use of traditional data security controls that are not purpose-built to keep pace with the exponential growth rate of distributed data sets characterized by the multiplicity of unstructured data flooding corporate data stores.

Strategically, all organizations must be positioned to embrace modern technologies and methodologies from big data analytics to software-defined infrastructure and DevOps while also securing their most sensitive data assets in an operationally efficient manner. The interconnected nature of today's infrastructures creates additional dimensions of complexity and security challenges, which require a new approach—a data fabric that protects data assets independent of location, allows data encryption, and facilitates agile data movement.

The Data Security Landscape—Problems and Challenges

Data at Risk: External and Internal Threats

While destructive cyber-attacks, such as distributed denial of service (DDoS), that can take down a website have a high profile in the media, and are still important to CIOs and other top IT leaders, the vast majority of today's security breaches are those that target and steal corporate data assets.

When it comes to the theft of this valuable data, there are fundamentally two types of attackers with malicious intent: external and internal bad actors. Such adversaries employ a variety of methods to penetrate organizational infrastructures, with the most common being socially engineered spear-phishing attacks that prey on human gullibility as an exploitable vulnerability. Spear-phishing entails bogus emails that appear to be from an individual or business that the user knows and trusts; they look legitimate, but are not. The external criminal thrives on familiarity. They know users' names, email addresses, and, after researching social media and other sources, have some personal knowledge about the user, leading to a well-engineered and targeted attack campaign.

These attacks are successful when a gullible recipient clicks on a link or opens an attachment that results in the introduction of malicious software (malware) to the user's system. Compromised users thus become the proxy, and their systems the entry point, for the external attacker whose actual target is sensitive corporate data. To get to this ultimate destination, attackers will move laterally across an organization's infrastructure to the data center systems storing the data assets of interest to the bad actor. What constitutes sensitive data that is at risk of loss as a result of such attacks varies by industry and includes, for example, personal health information (PHI) at health care organizations, personnel information, credit card holder data, and various types of intellectual property (IP). Intellectual property includes software source code, engineering designs, new product specifications, scientific research findings, patent research, and more. In all cases, IP represents that data which is strategically fundamental to that company's business—quite literally, the crown jewels. It is important to understand how cyber attacks operate to effectively protect an organization's own IP against loss.

Such attacks typically follow the cybersecurity kill chain, which is exemplified by a military concept in which the military would find, track, engage, and attack the target. In today's security parlance, the term serves as a model for the seven steps that mark the typical process of a cyber-threat: reconnaissance, weaponization, delivery, exploitation, installation, command and control, and exfiltration. Attackers can infiltrate IT systems and networks at any point in the kill chain, but most commonly they will start with the first phase—reconnaissance. In the spear-phishing example, this phase is about

getting to know the target to identify soft spots as the attack vector. For example, if the reconnaissance mission learned that a Director of Marketing at a Boston company was a baseball fan, she could then be sent an alluring phishing email message with a link to a Red Sox website. Once this fraudulent website is opened, malicious code downloads, providing the next step in access to valuable, business-critical data. Since the Director of Marketing is not actually the target, but the entry point, the attacker moves laterally to ultimately meet his prime objective—exfiltration of the sensitive data that's typically stored on the network.

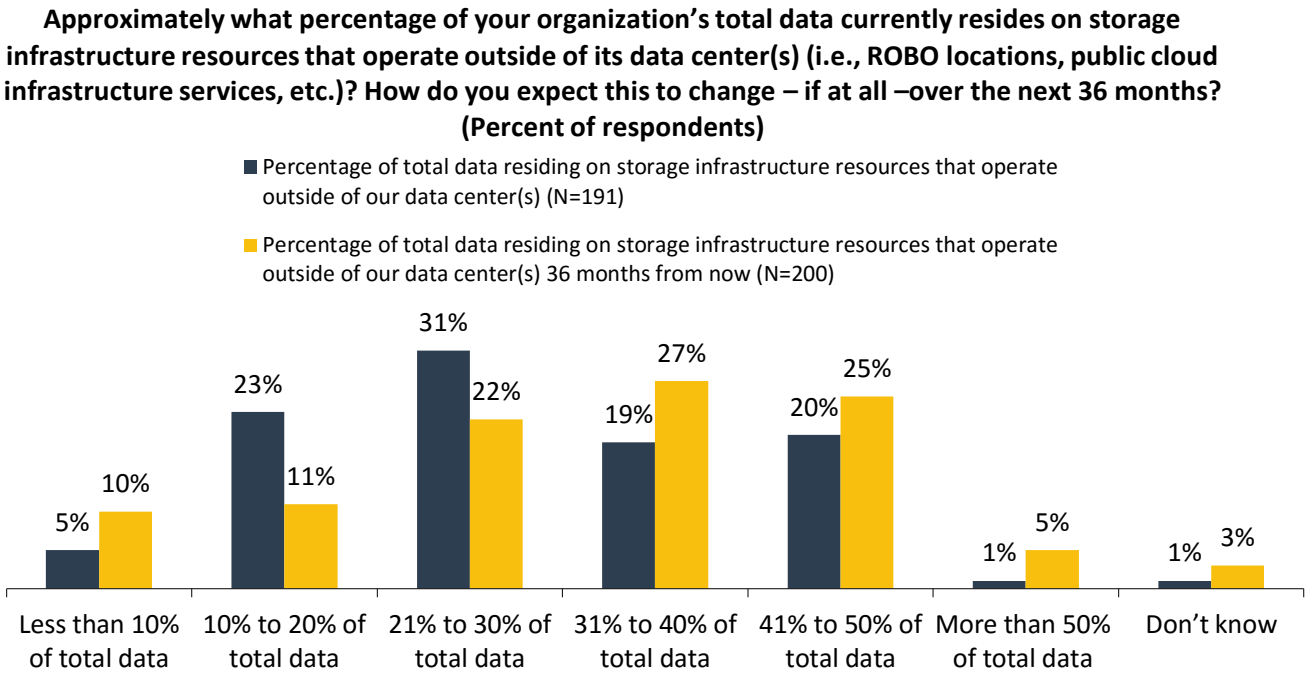
Internal individuals, such as disgruntled employees or employees who are leaving an organization, are another type of nefarious attacker, but in this case are employed by the target organization, and thus are intent on stealing their own company's data. Insider threats have a more intimate knowledge of their organization's applications and infrastructure, and are therefore especially difficult to protect against as their approach of "low and slow" often does not trip breach detection alarms.

The key to protection in today's hostile threat landscape is to employ a defense in depth strategy that employs multiple security controls and countermeasures to protect data assets from compromise. Since not all data is of the same value to an organization, this approach should include a risk prioritization strategy of identifying those data sets that require the strongest level of security. In the event of an attack, a well-designed strategy of this kind can also help minimize the adverse impact of the attack, and give security practitioners time to deploy new or updated countermeasures to mitigate against a recurrence.

Data Security Challenge: Distributed and Tiered Storage

It is a commonly accepted tenet of security that you cannot secure what you cannot see. With that in mind, many companies strive to keep their security controls colocated as close as possible to the data for maximum visibility and control. The use of tiered storage, especially cloud storage services, makes visibility challenging and thus applying controls difficult. ESG's research indicates that the amount of data currently stored in locations such as public cloud IaaS and remote/branch offices is significant today and will only increase over the next 36 months. Specifically, 40% of organizations currently store more than 30% of their total data outside of data center-resident resources, a figure which is expected to increase to 57% within 36 months (see Figure 2).

FIGURE 2. Current and Expected Percentage of Data Residing on Storage Infrastructure Resources Outside of Organization's Data Center



Source: Enterprise Strategy Group, 2016

King County in Washington State is an illustrative example of an organization using the cloud as a back-end target for nearline and offline data storage. As a governmental entity, King County must maintain compliance with a number of laws and security policies, some of which require not only encrypting data before it is moved offsite to the cloud, but also that King County, and not the cloud service provider (CSP), be custodian of the encryption keys. The county is meeting this requirement by leveraging NetApp's Data Fabric to encrypt its data before it is transmitted to the cloud provider while still maintaining control of its associated encryption keys.

Data Location Challenge: 3D View of Data Movement

Organizations need to easily move their data to the right location at the right time based on automated policies, not only for data management and protection purposes, but also to comply with regional data sovereignty regulations. The fluid nature of data movement in a hybrid cloud context makes applying consistent security policies challenging, but three factors exemplify why organizations will want to arbitrate the location of their data:

- Cost: Data should be stored on the system best aligned with the respective accessibility and recoverability service level agreements (SLA) required.
- Compliance: Regional laws and regulations that govern data sovereignty and privacy are such that customers in countries with such laws are required to store data in the same country as the company that created and owns the data.
- Incident Response: In the context of the cybersecurity kill chain where an organization's most important data assets are at risk, organizations that have detected an incident, and are in a response posture, may want to move certain data sets to an uncompromised environment.

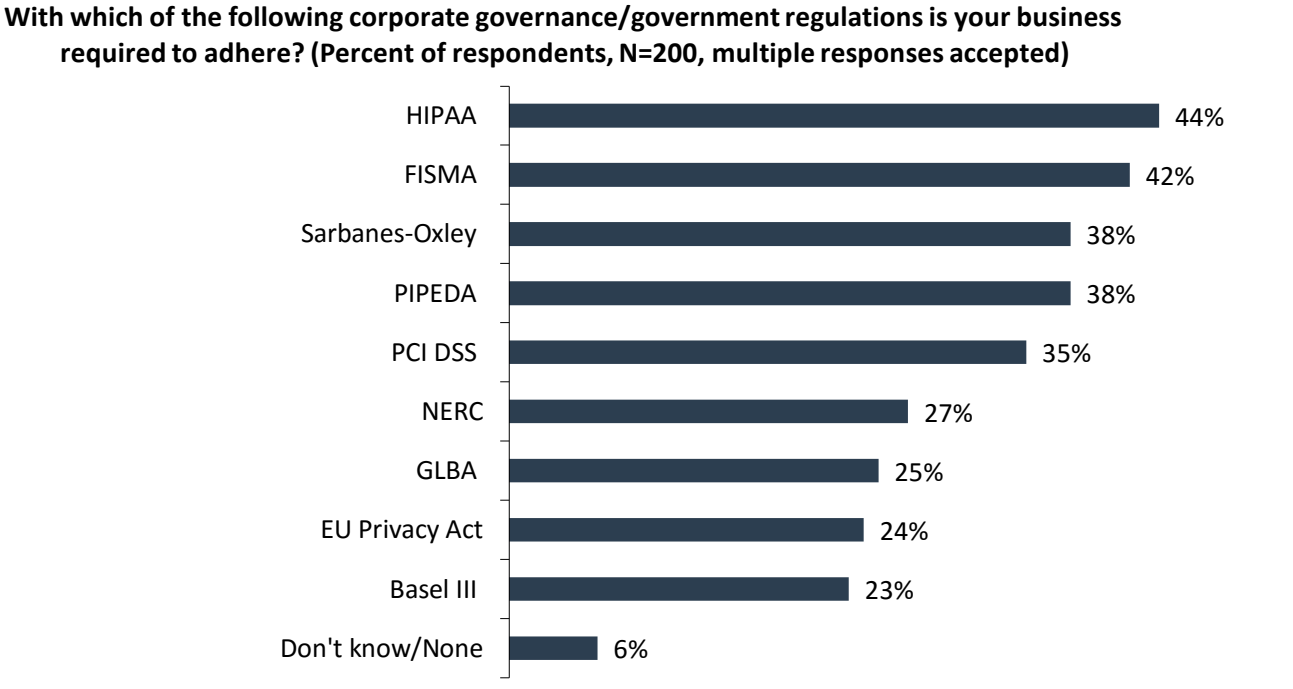
Regulatory Challenge: The Compliance Tax

Public and private sector regulations have put companies on notice that they have to stay within the boundaries of the requirements and law when it comes to how they secure their infrastructure and data. After achieving compliance with a regulation, organizations are required to demonstrate ongoing compliance, typically, at a minimum through an audit trail, and, depending on the regulation, regular audits conducted by a third party. Such is the case with PCI DSS and the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the latter of which provides data privacy and security provisions for safeguarding medical information.

There are overhead costs incurred by meeting regulatory requirements in the form of both CapEx and OpEx expenditures because new products may need to be purchased and new processes implemented. In some cases, organizations view compliance simply as a cost of doing business. For a retailer processing credit cards, for example, these financial outlays come with the territory. In other cases, however, depending on the perspective of the organization and the state of the business, companies may view such compliance requirements as business enablement—an opportunity to grow by complying with certain regulations. For example, HIPAA's business associate agreement (BAA) enables different organizations within a healthcare ecosystem or supply chain to do business together, provided they both adhere to HIPAA guidelines.

According to ESG's research, the five most common corporate governance or government regulations to which respondents' organizations are required to adhere include: *HIPAA* (Health Insurance Portability Accountability Act) (44%), Information Protection *FISMA* (Federal Information Security Management Act) (42%); *Sarbanes-Oxley* (Corporate and Auditing Accountability and Responsibility Act) (38%); *PIPEDA* (Personal Information Detection and Electronic Documents Act) (38%); and *PCI DSS* (Payment Card Industry Data Security Standard) (35%). The applicability of these regulations spans the public sector as well as multiple private sector industries, making securing data in a hybrid cloud for compliance a broad-based requirement.

Figure 3. Regulatory Compliance Requirements



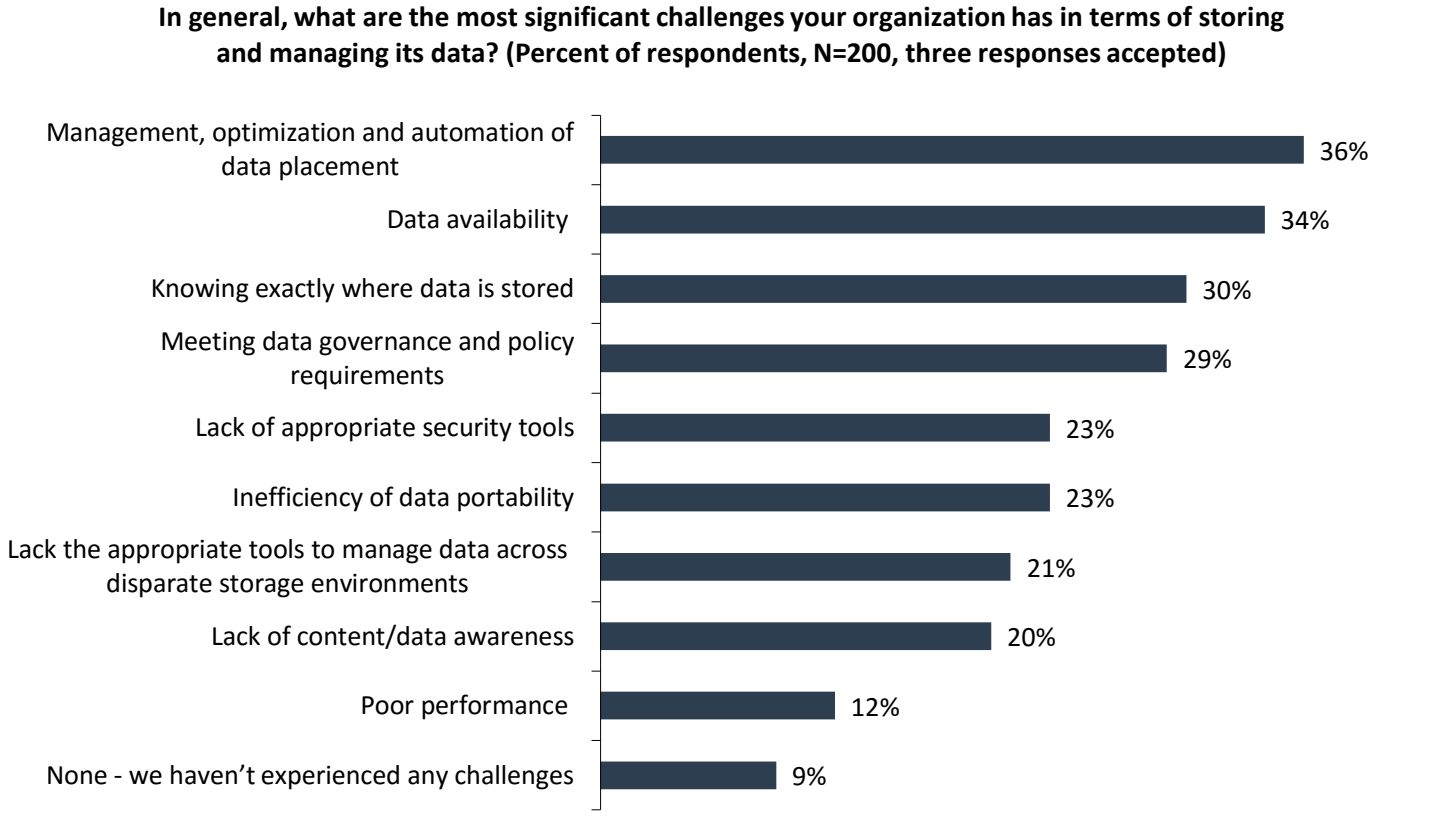
Source: Enterprise Strategy Group, 2016

Location-agnostic Data Security Provides the Solution

Visibility via a Platform

Based on ESG's *Security Considerations for Storing and Managing Data* research results, it comes as no surprise that companies are concerned about managing and securing their data given the variety of data location choices and the requirement to seamlessly and automatically move data sets. When asked about their most significant challenges regarding data storage and management, the most commonly cited response was management, optimization, and automation of data placement (see Figure 4). Other top responses included data availability (34%), knowing exactly where data is stored (30%), and meeting data governance and policy requirements (29%). These responses indicate that while many organizations operate in a hybrid cloud, automated data placement and security are essential from the perspective of many IT pros.

Figure 4. Regulatory Compliance Requirements



Source: Enterprise Strategy Group, 2016

Given the complex challenges associated with burgeoning data creation and movement of the digital era, there is a need to provide visibility via location-agnostic, federated platforms that are deployed and managed to work in concert. Companies require these unified platforms to be based on a fabric that enables IT environments to operate without constraints—but with security controls—across multiple, best-of-breed, hybrid cloud environments.

Examples of such highly distributed and multi-tiered storage use cases with strong security requirements are emerging Internet of Things (IoT) management platforms. IoT data is increasingly strategic to many businesses because it creates new insights, with examples including instrumentation readings from sensors on factory floors and energy farms, customer loyalty data from retail store systems, the status of health care systems, and a multitude of other use cases, all of which

represent data assets that must be securely moved. Based on a hybrid cloud architecture, IoT platforms capture streams of such data from IoT systems that are staged at the edge for secure movement to the cloud, requiring a data management fabric to streamline both the transfer and data encryption.

Border-agnostic Security and the Data Encryption Imperative

With the advent of cloud computing and end-user mobility, the traditional corporate network perimeter has become less structured and thus increasingly amorphous. Organizations, therefore, need to put a perimeter around their data in the form of encryption to obfuscate it from unintended viewers. ESG research highlights the importance of security, with 96 percent of respondents to the *Security Considerations for Storing and Managing Data* survey stating that they either strongly agree or agree that data encryption is a critical security control for protecting data across all their organizations' storage resources.

Certain regulations, including PCI DSS and HIPAA, require the use of cryptographic encryption as part of their respective data security and privacy requirements. For example, merchants that process credit card transactions are subject to the 12 requirements of PCI DSS, number 3 of which states that encryption must be applied to credit card holder data and the management of encryption keys must be documented. From an international perspective, organizations in certain countries will need to be mindful of their local data privacy regulations, and employ encryption solutions that allow them to create business opportunities while also complying with those regulations.

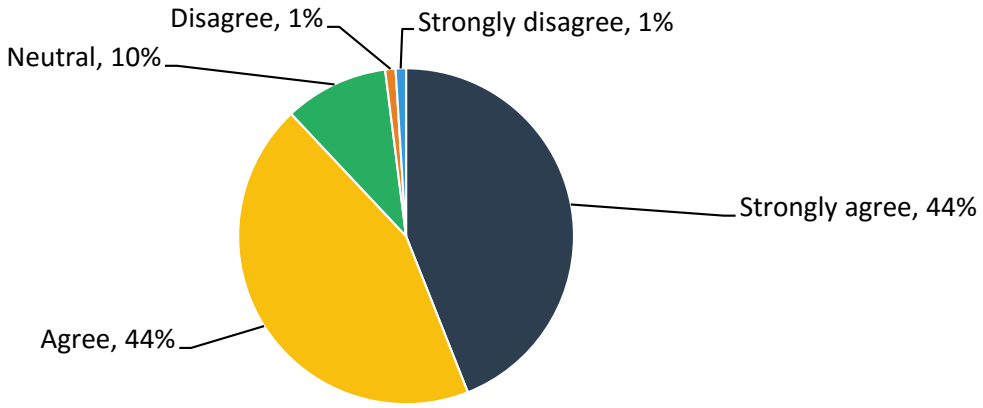
Long understood as a foundational cybersecurity best practice, the use of data encryption only increases in relevance for the fluidity of data movement in a hybrid cloud.

Key Management Is Key

Encryption and key management are top-of-mind issues that are inextricably linked. The encryption key is the passcode or secret that unlocks encrypted data because it makes the data readable, making the management of such keys as important a consideration as encrypting the data. For certain types of data assets and for certain industry regulations, it is strongly desired, if not required, that organizations are the custodians of encryption keys. And some organizations may prefer a separation of duties where already-encrypted data is moved to and stored in the cloud, or they may even rely on their cloud service provider to encrypt the data, while they "own" and manage the keys themselves. And since keys could be colocated with cloud storage but managed by the customers, some organizations may also want to consider separation of location between the data and the actual keys. Highlighting the importance of key management as a central data security concern, 88% of ESG survey respondents said they either strongly agree or agree that having custody of and managing data encryption keys for both on-premises and cloud-resident storage was a gating factor in their storage purchasing decisions (see Figure 5).

FIGURE 5. Sentiment toward Ability to Have Custody of Data Encryption Keys for On-premises and Cloud-resident Storage in Terms of Storage Purchasing Decision

Please indicate your organization's level of agreement with the following statement: when it comes to data security for our distributed storage environments, the ability to have custody of and manage the data encryption keys for both our on-premises and cloud-resident storage is–or would be–a gating factor in our storage purchasing decisions. (Percent of respondents, N=200)



Source: Enterprise Strategy Group, 2016

The bottom line here is that organizations want and sometimes require governance over encryption keys. The sensitive nature of certain data sets, as well as regulations with which organizations must be compliant, require a demonstrated level of governance over encryption keys. According to Bob Micelli, IT Enterprise Manager at King County, "We manually manage one set of keys, and we copy them to a secure location so they're under our governance. This is an absolute requirement so we not only meet our own security policies, but also prescriptive industry regulations such as those from the Criminal Justice Information System database (CJIS)."

Introducing NetApp Data Fabric

The NetApp Data Fabric is a federated platform that is purpose-built to satisfy the requirements for securing data assets in highly distributed, hybrid cloud environments. By separating logical and physical control planes, Data Fabric is location-agnostic, providing a wide breadth of coverage whereby the solution's automated, policy-driven data movement capabilities are integrated with leading cloud services providers (CPSs) via native connectors. Breadth of coverage and data location awareness enhances the security applicability of Data Fabric's data movement functionality by providing both the visibility and control to expedite moving data in the midst of a cybersecurity incident.

In addition to its data movement capabilities, NetApp's Data Fabric provides enterprise-grade encryption capabilities. The solution can not only automate the encryption of data, but also provides key management to allow user custodianship of the keys, a security best practice required by some industry regulations. Data Fabric's applicability for compliance use cases is further bolstered with a robust audit trail feature.

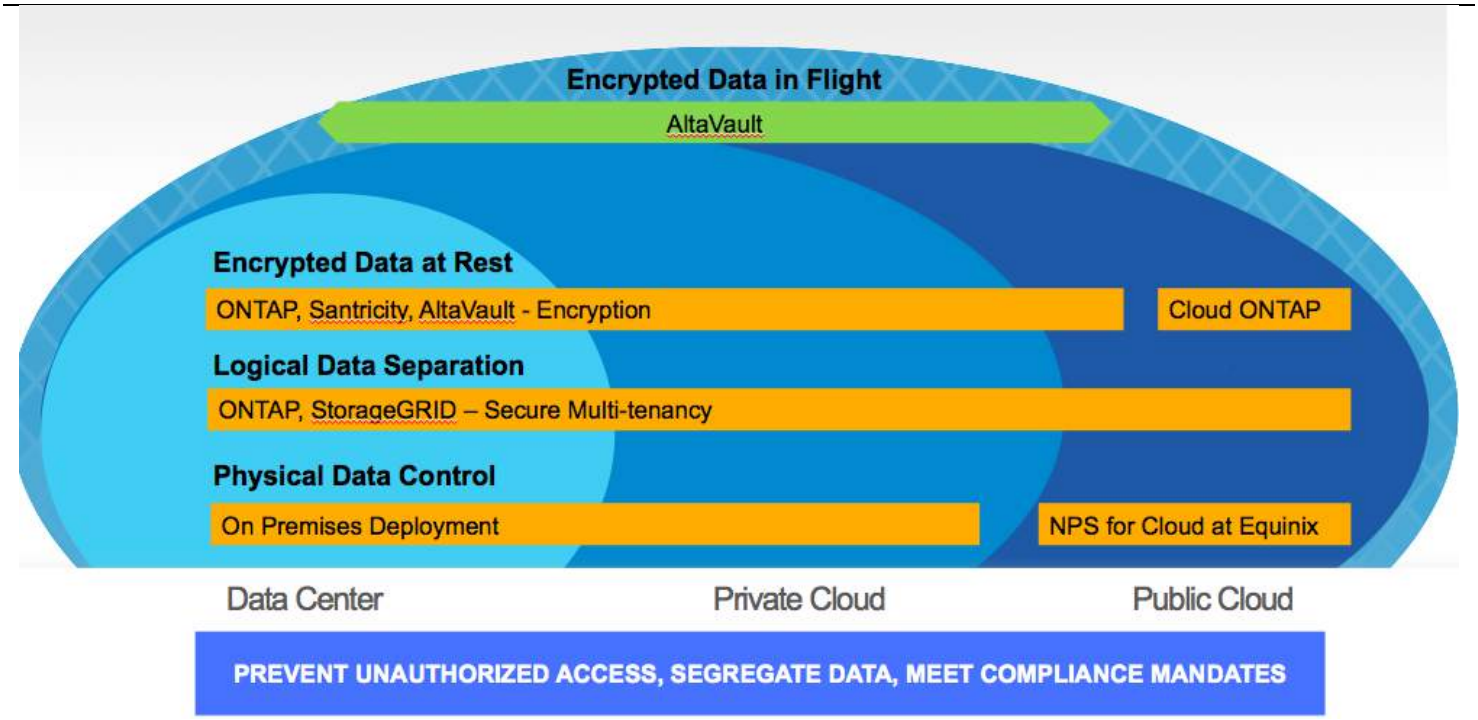
NetApp Clustered Data ONTAP, Cloud ONTAP, NetApp AltaVault, and NetApp Private Storage (NPS) provide the foundation of the NetApp Data Fabric. Other enabling technologies are built on Clustered Data ONTAP solutions and combined with integrated data transport from NetApp SnapMirror and NetApp SnapVault, as well as data management from NetApp OnCommand data management software. By implementing these products, users can begin to create a common set of data services across flash, disk, and multifarious clouds. NetApp AltaVault provides cloud-integrated, enterprise-class

protection by encrypting data at rest and in flight to prevent against network-based compromises, while NPS enables customers to employ cloud services while maintaining ownership of their data on their own systems in close proximity to clouds.

NetApp Data Fabric addresses today's notable data management and security challenges by enabling IT organizations to securely span data centers and clouds while assuring control of their data assets. It accommodates multiple different architectures, including pools of virtualization, bare metal servers, and private and public clouds, such as the King County on-premises environment that utilized cloud as a target for its nearline and offline data storage.

Figure 6 shows how NetApp Data Fabric advances security in today's hybrid cloud data center.[3]

FIGURE 6. NetApp Data Fabric Security for Hybrid Cloud Data Centers



Source: NetApp, 2016

The Bigger Truth

Agility and security need not be mutually exclusive. Agility is at the core of both cloud computing and the data explosion that is spawning innovative new business models. Realizing the benefits of software-defined infrastructures and data-driven business models are essential elements of success for all organizations, but it must not come at the expense of security. Nor can it ignore two immutable truths: In the modern hybrid cloud data center, data will continue to be created at an exponential rate, and an organization's most sensitive data will be vulnerable to cybersecurity threats from both external and internal adversaries. To go fast securely, organizations require automated data movement and comprehensive data security controls that address these realities. Proven in hybrid cloud data center environments, the NetApp Data Fabric solution meets these data movement and data security requirements, allowing customers to realize agility without compromising security.

[3] Source: *NetApp Security Solutions Portfolio*, November 2015.

Enterprise Strategy Group is an IT analyst, research, validation, and strategy firm that provides actionable insight and intelligence to the global IT community.

 www.esg-global.com

 contact@esg-global.com

 P. 508.482.0188