CIO Perspectives

Safely Lifting and Shifting Enterprise IT to the Public Cloud
CIOs that we talk to say organizations need to develop an enterprise cloud strategy and that both the business side and IT side need to work together in order to achieve success. Interestingly, these CIOs see a division of labor taking place with their business counterparts. They see the business side leading demand for Software-as-a-Service (SaaS) while IT leading demand for Platform-as-a-Service (PaaS) or Infrastructure-as-a-Service (IaaS).

So far, CIOs say their adoption of PaaS and IaaS has been muted. They have gotten past the stage of saying no one is doing it, but wide scale public cloud adoption is held back today by real and perceived capabilities and compliance issues. To respond to these risks, many CIOs have chosen to discriminate between what goes to the public cloud and what goes to the private cloud. Typically, this is done based upon the perceived sensitivity of data and because – fairly or not – on premise, private cloud is judged to be safer than the public cloud.

With this said, CIOs believe most organizations will predominantly be using public cloud within the next 5-10 years. Summarizing this point of view, one CIO said the following:

“The end game for most with cloud is ultimately public cloud. Hybrid is just an interim step for most organizations.”

So, what are the tangible reasons for holding back the wide-scale use of public cloud today? Clearly, the inertia of prior IT projects and investments continue to hold back some. One CIO said, “You don’t spend years building
something and making it safe for the business to overnight dump it all into production on public cloud, ready or not.” But most CIO believe that, longer term, public cloud remains compelling as more investment is needed to respond to digital disruptors in nearly every industry.

As the financial pressures to shift capacity to the public cloud intensifies, CIOs insist that they are personally accountable for securing their enterprises. CIOs say for this reason security should not be baked into the cloud after the fact. They feel that storage and applications moved to the cloud need to be hardened prior to getting there. But a wholesale move to public cloud requires something different. One CIO suggests that cloud data governance is needed. Clearly, there is enhanced awareness of the fact that information security is needed in the public cloud model. However, there remains a general perception that public cloud is not safe unless it is under the originating businesses’ lock and key. CIOs insist that there must be due diligence on security and compliance.
Legitimate Concerns for Cloud

There are many questions to be asked of cloud providers separate and distinct from what we will call later a “cloud protection architecture.” These are just a few of the most obvious ones:

- How does the cloud provider handle encryption and encrypted data?
- Do our users have exclusive access to their data?
- Does our data get commingled with data from the cloud vendor’s other clients?
- Does the cloud provider satisfy all compliance requirements including specific statutory regulations for all jurisdictions or all enterprise policies?
- Are data stored so that it is physically protected as well?
- Does the cloud provider mine the data that it stores for its own purposes?
- Is the cloud provider fully auditable?
- Does the cloud provider provide breach notifications according to our company’s privacy policies and statutory requirements?
- Is the cloud provider’s overall security capabilities sufficient?
- Does the cloud provider have data transfer capabilities and sufficient security for the data transfer?

These are a lot of questions but they are here so that you go into a cloud relationship with your eyes wide open.
Most CIOs admit that they have been throttling their company’s move to public cloud. They say that it is important that they think differently about on premise or private cloud than public cloud. When pushed on this issue, CIOs say they need to think differently for two reasons: service delivery and cybersecurity.

With the former, the use of virtualization approaches likely yield indistinguishable results. However, if you are lifting and shifting existing capacity, many want to know that their cloud vendor can provide like or better capabilities to an internal datacenter. How do you do this without testing end-to-end system performance with real data? How do I do this in a way that does not expose sensitive data to a partner or my DBAs and testing resources? Another major difference can occur when there is a service delivery event or incident. For some organizations, there is a nuclear submarine approach with “all hands-on deck.” But can this quality of service be bought from cloud providers?

So this brings us to the last remaining issue: cybersecurity. If you could completely control access to data in the public cloud would you still need to have data centers? Or more importantly, would this capability accelerate your move to public cloud if this were the case? Could a cloud secured by yourself allow you to start recouping the sunk costs in datacenters and IT operations? CIOs we have talked to say the answer is yes. So, in this brief, it is our goal to show you how your architecture can move from hybrid cloud with more private cloud and less public cloud to almost all public cloud; and how this can be done with better data security than you or your cloud vendor provides natively.
Developing a Cloud Protection Architecture

Security concerns have clearly prevented many of those entrusted with sensitive data from taking advantage of all that the public cloud offers. For them, the ability to directly control the protection of data as it flows in and out of the cloud is vital. “What has been missing,” one CIO said, “is an architecture for data security within the cloud.”

With this capability, CIOs say that they would move faster in making use of public cloud. CISOs extend the CIOs thought here by saying what is needed is “secure transport” for data, identity, and access management within the cloud. Clearly, with the capacity to control and protect data in the cloud, organizations can start to reduce datacenter investment. These cost reductions will come from eliminating people and infrastructure.

But how do we get there? What is needed is an architecture that can protect data at all points. This includes the movement of data from legacy datacenters and then up into the cloud. This approach needs to allow data security to be centrally managed not by the public cloud vendor but by the enterprise and their application/data owners. Without this, the company is left at the risk of the security policies and procedures of the public cloud owner. A Gartner public cloud analyst said recently that, “Public cloud users today must accept that if their public cloud vendor is hacked then their keys have likely been exposed.”

Effectively, what is needed is the ability for the cloud provider to be hacked with no sensitive data loss to the enterprise customer. This means that organizations need to think holistically about data and how it flows, and then architect a solution that works with these data assets, at rest and as they move from one platform to another. This needs to occur when data moves between on-premises and cloud, as well as when data moves from one platform to another in the cloud.
For the most part, cloud is just another place to store and process information. However, there are some compute patterns that are specific to cloud computing and managed services like decoupling storage from compute. For example, Amazon S3 is emerging as a central cloud storage service that provides data to multitudes of computing platforms including Amazon EMR. As such, we can think of a cloud data security strategy just like a typical enterprise data security strategy, with additional provisions for the cloud computing patterns that do not exist on-premises.

So, all things considered, what are the key components of a cloud security strategy? An effective strategy will make use of the following concepts: Separation of duty, fine-grained data protection, centralized policy management, pervasive enforcement points, and transparency and automation. Let’s review each of these.

**SEPARATION OF DUTIES AND EXTENSIVE AUDIT:**

Best practices dictate that administrators do not have access to data. This prevents the problem that many have with access control systems where those with privileged access become targets for hackers trying to break in. Access to secure assets is logged with the user, place, time, and action. This way there is auditability for organizations that need to show compliance to auditing organizations.

**FINE-GRAINED PROTECTION:**

Here protection is ascribed to certain pieces of data and/or the person trying to access it. This solves the problems with typical coarse-grained protection. Coarse-grained protected data cannot flow from one platform to another without losing its protection. This is a problem when data moves between clouds or from on-premises to a public cloud. In the cloud, we also want to decouple storage from compute. In this approach, data
storage and consumption take place on separate platforms. Centrally managed fine-grained protection is the only viable solution other than permanent de-identification. Permanent de-identification of course is not satisfactory for almost all uses cases.

**CENTRALLY MANAGED POLICIES:**

This is the best way to manage sensitive data elements, protection, unprotection, and access privileges. This means that data policies are consistently applied as they transverse on premises and cloud systems. If we want to protect data in S3 and unprotect in EMR, then centrally managed policies are the only viable solution.

**PERVASIVE ENFORCEMENT POINTS:**

You want consistent enforcement across all systems. One of the strengths of cloud computing today is that many on-premises platforms are replicated in the cloud, allowing re-hosting and lift-and-shift. Support on a wide variety of platforms, traditional DBMS, big data, EDWs, and NoSQL DBs, is essential.

**TRANSPARENCY AND AUTOMATION:**

This is the ability to inject data protection and unprotection without having to modify application code or adding specific polling. You want the data security related operations to be declarative, transparent to applications and users, and automatic, e.g. poll for changed data and protect them, or as data is on the move.
Build Solutions Based on Data Flows in the Cloud

If we had a solution with all the above key components, what would that solution look like? Where should it protect data, where would it unprotect the data, and how should it manage all the rules and policies around protection, unprotection, and access?

Here are the guiding principles we recommend:

1. The protection of data should occur before landing or immediately after landing in the cloud.

2. Protected data should stay protected wherever it flows. Protection should follow data as it moves across increasing complex data ecosystems.

3. The unprotection of data should occur only as the data is needed by authorized users. When using Privacy by Design principles, this is called “Privacy as the Default Setting.”

4. Data policy enforcement should be consistent. Enterprise level policies should be centrally managed by data or role. These rules and policies should apply on-premises as well as in the cloud. Thereby, coherent security control can be delivered across a hybrid ecosystem.
Given the above strategy, some key questions to ask include: In the cloud, where does data land and how? and Where do you protect data? One distinctive computing pattern in the cloud is the idea of de-coupling storage from compute. In many use cases, there is value to allowing the compute loads to go up and down, while data are persisted in cloud storage and are provided to computing platforms as needed. With this said, how do you get your data into cloud storage? These are the places where we want to apply protection:

**DIRECT DEPOSIT OF DATA IN CLOUD STORAGE:**

For data architects, they have a choice here. They can encrypt data in transit or instead data can be protected upon landing. The safest option protects data in transit and use. It is better to have two layers of security in protecting your cloud data where you are always in control. In Figure 1, we show two examples of how data protection can be implemented for Amazon S3.

**INGESTION, AS IN THE CASE OF DATA REFINERIES:**

Amazon EMR, for example, sometimes is used as an ingestion engine. In this case, data can be protected by EMR and then deposited in cloud storage, Amazon S3. In Figure 2, we show an example of using EMR to land and refine data, including applying protection, and then deposit the data onto S3 protected and ready for consumption.
Figure 1. Protecting data in Cloud Storage
In both scenarios, data is protected and de-identified via fine-grained protection. We show that the protected data can then flow in two ways:

1. Fine-grained protected data can be used for analytics without access control constraints, it can flow freely across an extended corporate ecosystem, as whoever accessing the data will not see the sensitive data values but can see information that is sufficient for analytics, especially aggregated trending analysis.

2. Fine-grained protected data can be unprotected upon consumption, as shown in Figure 2, under the control of centrally managed policies, with only authorized users able to see unprotected data.

Figure 2. Protecting data on Amazon S3 during ingestion using EMR
As shown in the above figure, the centralized policy manager enables role-based protection/unprotection and Separation of Duties. Separation of Duties holds that the data policy makers do not have access to sensitive data. If Amazon S3 is a cornerstone of your cloud deployment, then you must have a strategy to ensure that data sitting in S3 is protected and ready to be dispensed.

SOLUTION EXAMPLE: BIG DATA IN THE CLOUD, SECURELY

In the last few years, there has been a dramatic increase in the number of organizations running big data workloads in public cloud, mainly on AWS. This movement of big data workload from on-premises to the cloud is perhaps due to the high total cost of ownership of on-premises Hadoop clusters. Today, AWS EMR (Amazon Web Services Elastic MapReduce) in fact runs more workloads than either Cloudera and Hortonworks. By leveraging S3 as central data hub for all AWS EMR workloads, users can process big data sets at scale. The problem is that often these big data systems contain significant amounts of sensitive personally identifiable information (PII) or protected health information (PHI) data. To operate these in the public cloud or even on premises it is essential that this data be protected.

In Figure 2, we showed how EMR protects data as a data refinery. In the same picture, we also show that EMR can consume protected data from S3, either as de-identified data, or unprotected to access original values as needed. Remember, EMRs as a computing platform may be transient, spinning up and down. The advantage of the cloud is that one can easily create these massively powerful computing platforms. To keep up with that speed of innovation, it is important that data can flow at low friction and yet securely, as shown in Figure 2.
The examples shown above can be easily extended as more workloads move to the cloud. In the diagram below, data flows to a new platform. All we need to do is to deploy a protector so that the new platform operates under the central policy management and enforcement, as shown in the diagram below.

Figure 3. Protegrity support for cloud computing journey
Overview of Protegrity

Protegrity provides a Data Security Platform that systematically overcomes the people, process, and technology challenges faced by IT leaders in an increasingly complex cloud data ecosystem. This matters because IT leaders tell us that real and perceived public cloud data security issues are what hold them back. They need to control sensitive data as it enters and is used in the public cloud. This means if your cloud vendor is hacked your sensitive data is still secured. It also means because you can protect your data from internal and external actors with ill intent, you can use your real data in the migration process to public cloud storage and applications.

By doing this, Protegrity allows customers to go beyond compliance to ensure data security and privacy without compromising business processes. Forward thinking CIOs recognize that in contrast with the “big iron days” when all data was in one place, an ecosystem approach to data is now essential to business success. In this cloud-first environment, data protection needs to move from the perimeter to the point of use to be effective. In a world with only virtual boundaries, security focus needs to shift from the systems to the data itself. Protegrity allows CIOs and other IT leaders to actively manage data and control access to it as it moves to the cloud, enabling them to finally take advantage of all efficiencies and cost savings that the cloud has to offer.
CIOs need the ability to maintain processes, compliance, security, and privacy in the public cloud. Protegrity empowers them to do this by enabling everyone involved – from the data stewards creating policies for access and use of data, to the CISO securing and protecting data assets and the analysts seeking data insights – to securely perform their duties. Protegrity enforces security policies to protect sensitive data and control access to it, regardless of where it flows, is used, or rests. One CIO illustrated the importance of this by saying, “You know those flight maps in the airline magazines? Those are our data movement maps. We have in our environment, data flying all over the place.”

Protegrity solves data compliance problems by allowing protection to be by design, through centralized management of security and privacy policies that control access and protect data itself using anonymization and pseudonymization technologies. With these in
place, data protection is automated enterprise-wide. Today protecting
data needs to be a 360-degree conversation and attention “needs to be
given not to the pieces but to the whole enchilada,” according to one
CIO. This allows organizations for the first time to consistently protect
sensitive information throughout its lifecycle in complex IT ecosystems,
maximizing data security and regulatory compliance, on premise or in
the cloud, without compromising their day-to-day operations.

Protegrity’s Data Security Platform seamlessly integrates protection
into all existing business applications and business processes, allowing
organizations and their third-party associates to ensure appropriate,
quick, and easy access to critical data while providing defense against
insider threats or external attacks. Protegrity simplifies the process by
which organizations move beyond reactive compliance with a mandated
need for privacy and data security to proactive, risk-adjusted protection.
This goes a long way toward inspiring client trust that their private and
sensitive information is protected against misuse and disclosure.

Protegrity’s solutions allows specific sensitive data to be fully visible,
partially visible, or completely protected depending on each user’s role
in the organization. This enables data analytics while preserving privacy
and security for different applications and parties. The foundation for
protecting sensitive data in the enterprise is the data protection policies
each organization creates, based upon their needs and circumstances
and the following business questions:

1. What data shall be protected?
2. Who shall have access to it?
3. When (days/times) shall it be accessible?
4. Where in the enterprise shall the policy be enforced?
5. Audit of access and process attempts by whom, to what data,
   where and when.
Protegrity has been helping companies for more than 15 years to protect their data as a core business competency. We are used for public cloud environments by leading enterprises in the financial services, healthcare, entertainment, and gaming industries. Each has expansive uses cases for their data that go beyond simply cutting costs and at the same time, involves ensuring that their sensitive customer data is not compromised.

www.protegrity.com