

AHEAD

Protecting & Preserving Data in the Cloud



Over the past few years, we've seen a huge portion of enterprise workloads begin shifting to the cloud. While this is certainly no secret, there have been some significant changes in what types of workloads organizations are choosing to host there. Historically, the public cloud has been almost exclusively reserved for non-persistent, non-enterprise-critical workloads. Today, however, we're seeing an increasing number of enterprises using the cloud to host some of their most business-critical workloads and applications. In fact, Gartner asserts that "by 2025, over 95% of new digital workloads will be deployed on cloud-native platforms – up from just 30% in 2021."

More workloads in the cloud means more **data** in the cloud. But cloud service providers (CSPs) are protecting that data, so out of sight, out of mind—right? Actually, the answer is no. What may come as a surprise to many organizations is that most major CSPs follow what is called a **shared responsibility model** when it comes to data protection in the cloud. And while each CSP has slightly different policies in this regard, the bottom line is that enterprises are typically on the hook for protecting and preserving the data they store in public clouds.

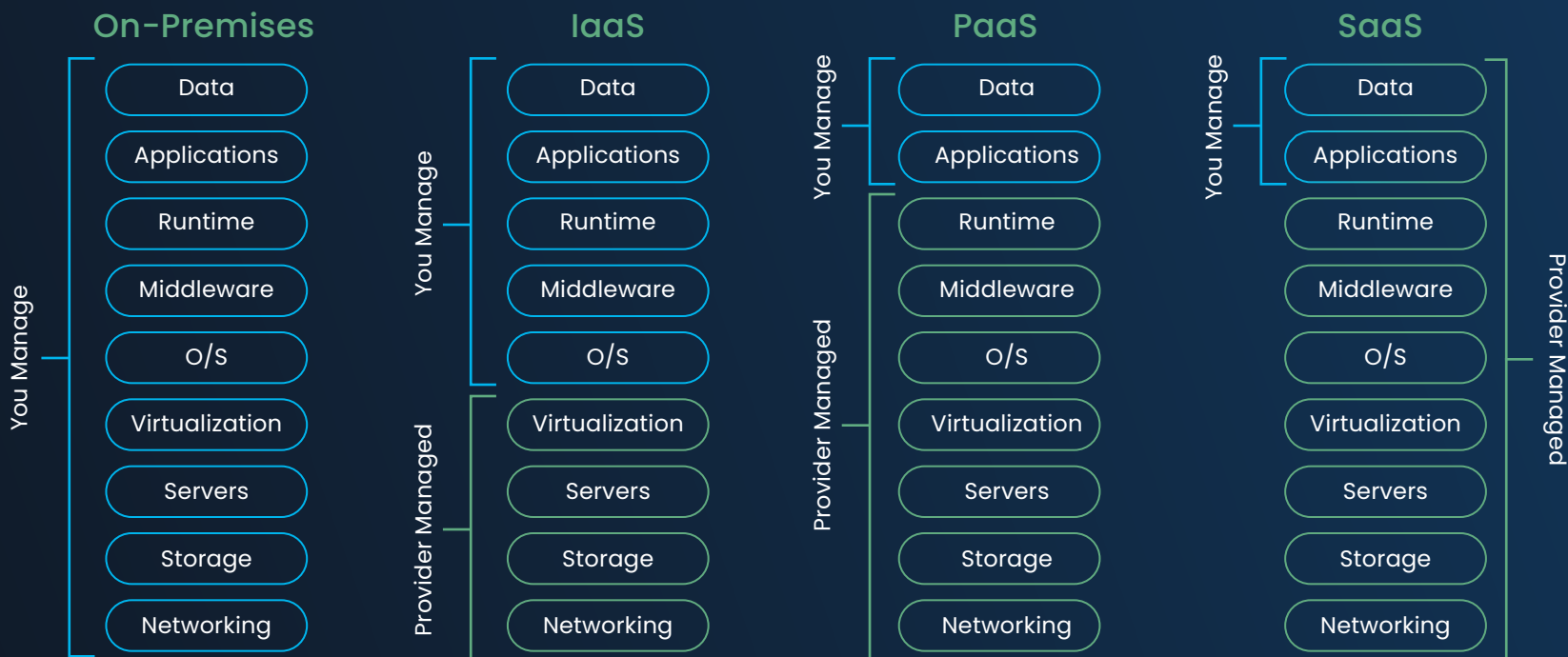
Below, we'll examine the implications of a shared responsibility model, the many challenges related to protecting cloud data, and the requirements for cloud data protection that cannot be overlooked.



Understanding the Shared Responsibility Model

Far too often, organizations put themselves at risk by making the assumption that data protection is included within the scope of service from their chosen CSP. While CSPs are invaluable to modern enterprises and provide a wide range of other services related to cloud infrastructure, data protection is commonly not within their purview. Thus, it is crucial for organizations to know exactly what is covered by their CSP and, perhaps more importantly, what is not.

Below, we've laid out a common breakdown of the shared responsibility model that demonstrates how even as organizations increase their CSP engagement from infrastructure to platforms to software as a service, data tends to remain the responsibility of the enterprise itself.



Data Protection Challenges in the Cloud

So, with the understanding that data protection and preservation in the cloud is the responsibility of the organization and not the CSP, let's examine some of the most prevalent challenges facing enterprises in this space.



Cloud Data Protection Requirements

When choosing a cloud data protection solution, there are a number of different requirements that organizations must take into consideration to ensure that their data is truly secure:

Data/Service Types

- Virtual Instances/OS
- Kubernetes/Containers
- Code Repositories
- Data Stores/Platforms
- Supporting Infrastructure

Discovery/Search

- Automated Detection of Rapidly Provisioned/Deprovisioned Cloud Resources
- Leverage Tagging for Automation
- Catalog/Search for Granular Restores

Disaster Recovery

- Accessible Data Copies in a Different Region or Provider
- Automated Restore Capabilities

Cyber Resilience

- Immutability
- Analytics
- Security Controls
- Segmented Copy

Governance

- GDPR Region Lock
- Legal Hold
- Industry Compliance Requirements
- SLA Enforcement
- Tagging/Auditing

Ephemeral

- Rapidly Scale Up/Down Resources
- Convert Formats & Quickly Spin Up Instances

Storage & Retention

- Use Multiple Classes of Storage in an Automated Fashion
- Long-term Retention using Cost Appropriate Storage

Management

- Enable Self-service via APIs
- Integration into ESM Platform
- Consistent Management & Operations across all Clouds (including Policy Creation/Enforcement)
- Consolidated Reporting

Cloud Focused

- Leverage Native Cloud APIs
- Little-to-no Backup Infrastructure to Manage

Areas of Cloud Data Protection

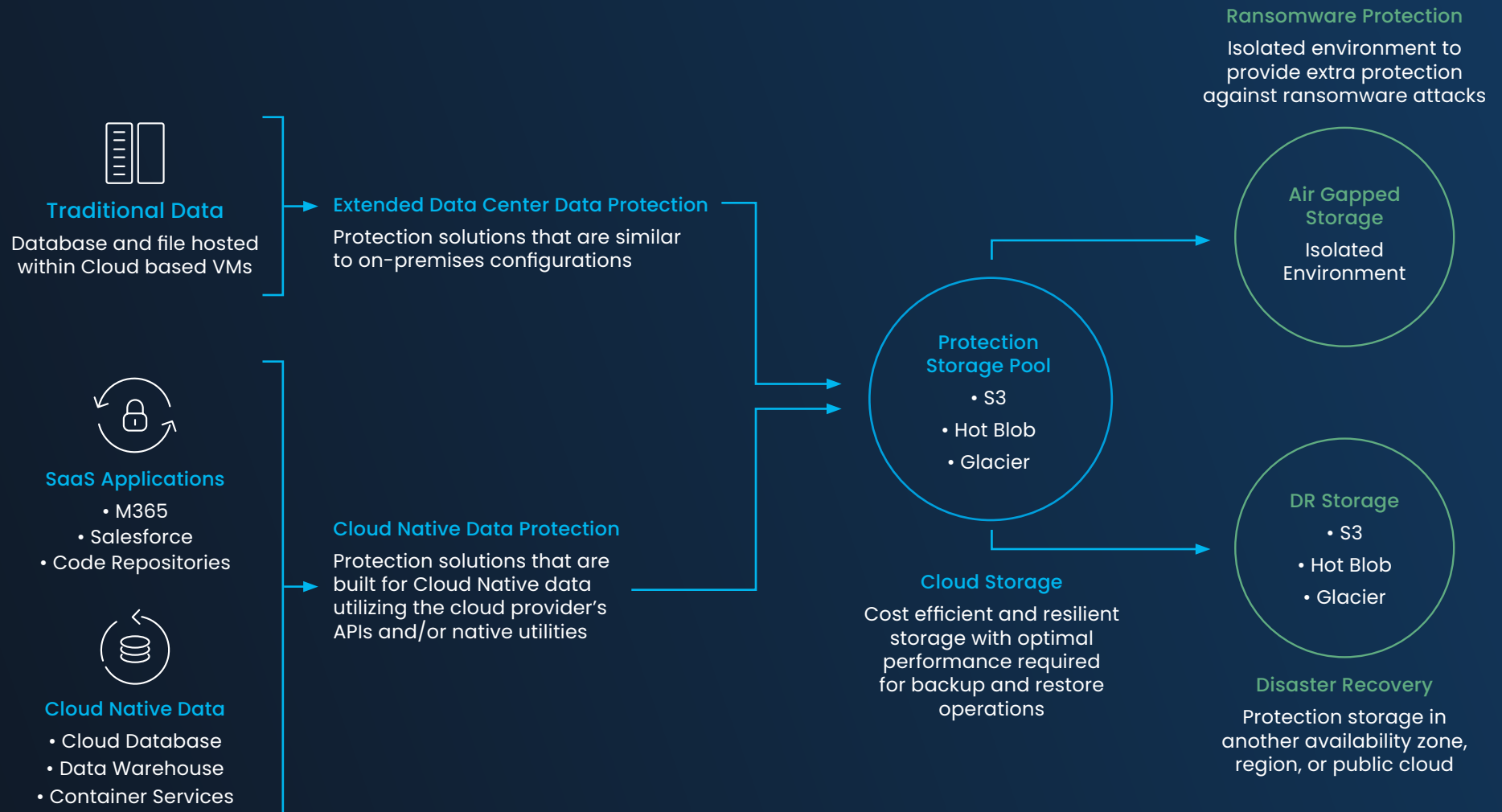
For most organizations, their cloud world looks a lot like their on-prem world, where they're using virtual machines or instances and running their applications or databases inside of these virtual instances. For others, they are leveraging SaaS platforms, which likely include things beyond what most think of with M365 or Salesforce. They may also have a SaaS offering for critical data like a code repository. And as organizations get more comfortable with the cloud, they're turning to more cloud-native services, such as databases, data warehouses, and containers. In many cases, enterprises are likely to end up using all three.

No matter what types of services they are leveraging, there are some commonalities that exist for the majority of organizations. Most will end up with a protection storage pool that banks backed-up data – frequently in the form of snapshots stored in an object storage platform. It is vital for organizations to have a DR copy of this data, which may—depending on your RTOs and RPOs—require a different solution than what is used for standard backup tooling.

And finally, for critical data, it is recommended that enterprises have an air-gapped storage environment that is isolated from production. Bad actors can compromise cloud credentials and potentially gain access to primary storage protection pools, so the best practice here is to leverage a storage environment that is less likely to be discovered and has very strict security controls.



Areas of Cloud Data Protection



As Code Data Protection

While it is entirely possible to provision resources manually in the cloud, most organizations will choose to do so by leveraging automation. In those cases, there are three main areas that are critical to protect:

1. Kubernetes Environments

In the past, this has been less of a concern, as organizations have the option to redeploy and may not have much to back up in the case of stateless applications. However, it's becoming more common for enterprises to run applications with stateful and persistent data, which must be protected. Whether an organization is protecting from a configuration error or a disaster situation, they'll want to ensure that critical Kubernetes objects—such as namespaces, pods, services, persistent volumes and persistent volume claims, secrets, and ConfigMaps—are protected. Backing up container images is also worth consideration.

2. Code Repositories

There is a common misconception that since everything is stored in a code repository—which can be rebuilt from scratch—there is less need for protection in this area. But what happens if the data in that code repository is lost? Malicious actors or something as simple as human error can lead to lost code repository data. Thus, it is vital to ensure that code repositories are backed up—especially if they are part of the organization's disaster recovery plan.

3. CI/CD Pipeline

Organizations spend a ton of time building and tuning their CI/CD pipeline so that they may deploy code at a rapid pace. Leaving the CI/CD pipeline unprotected can lead to significant downtime, which impacts developer productivity and could delay critical product releases. While organizations can certainly rebuild the pieces, much of the customization and tuning work can be lost if not protected. Backing up the CI/CD pipeline also gives organizations the option to restore rather than rebuild, which is a much quicker fix in a space where speed is king.



Platform Data Protection

Another area which requires attention is the data within cloud data platforms. People often say that “data is the new oil,” and they may be correct. Data has become perhaps the most valuable commodity in the modern business world—and its inherent value makes it tough to replicate. While it’s always an option to redeploy code or rebuild servers and containers, recreating data is extremely difficult, if not impossible altogether. The sheer number of data platforms available today is incredible, but often overwhelming. When it comes to protecting information within cloud data platforms, there are a few key things to keep in mind:

API Integration

Integration with native snapshots and protection APIs is a great way to simplify the backup and recovery process.

Rapid Scaling

Data protection software integration enables rapid scaling of protection policies as cloud-native data sources are created.

Replication Policies

Replication policies within the data protection software allow copies to be made for disaster recovery, ransomware protection, or migrating the databases for use on-premises or in other cloud environments.



Key Takeaways

- Cloud-native data protection presents a number of new and unique challenges to enterprises
- While some on-premises concepts and practices are still relevant, the cloud requires a different approach
- It is vital to understand exactly what is covered by CSPs – most are not responsible for protecting data within the cloud
- Partnering with a company like AHEAD can help ease the process of protecting cloud-native data



2022 AHEAD SPRING summit

To learn more, check out the [on-demand sessions](#) from the 2022 AHEAD Spring Summit or get in touch with our data protection experts today.

National Hubs

ATLANTA

1117 Perimeter Center
W406
Atlanta, GA 30338

CHICAGO

401 Michigan Ave.
#3400
Chicago, IL 60611

SAN FRANCISCO

2000 Crow Canyon Place
Suite 250
San Ramon, CA 94583