SNIA
Data Protection and Capacity Optimization Committee

CONTRIBUTED ARTICLES PROGRAM

Primary considerations on Data Protection and the Cloud

By David Hill, Mesabi Group

What is data protection in the cloud? That is not an easy question to answer since it comes in various forms, and the tools and technologies to provide data protection are extremely numerous and can be used in different combinations. From IT's perspective, a large number of choices can make cloud more difficult than traditional schemas. Still, we can cut the Gordian knot of cloud complexity with three steps that will help guide further exploration of data protection more easily.

First, we need to understand in general what data protection types the cloud might solve. Second, we need to understand that big decisions related to the cloud and data protection include what is managed internally and what is managed by a third party. Third, we need a rough basis for putting together an inventory of workloads that an organization can consider when moving to the cloud. We can then use this foundation as a basis for future discussion of data protection in the cloud.

Untangling Terminology Confusion Related to Data Protection

Backup, recovery, business continuity (BC), and disaster recovery (DR), are among the issues that are bandied around in discussions involving data protection (as are other terms, such as combined BC/DR and high availability (HA)). Keep in mind that definitions matter. If you and a vendor define issues and processes differently, what you get may not be what you want or need.

Backup is the easiest and the most familiar process for most situations. A backup is a data protection copy of data derived from the production copy (which is the official working copy of the data). A backup copy is the one that is used to recover data that is needed to restart an application correctly.

Disaster recovery (DR) is the recovery of the entire relevant IT infrastructure at a remote (i.e., secondary) site after a primary (i.e., production) site has become unavailable for an unacceptable period of time. Yes, the data is important, but so is the recovery of servers and their applications as well as any networking capabilities that are needed.

Business continuity (BC) is about both operational recovery (OR) and disaster recovery (DR). Operational recovery from a specific problem at a primary site — such as a server, application, or disk failure — may be absolutely critical, but it requires an emergency room, fire drill response as opposed to invoking (often with an official declaration of disaster) wide-spread disaster relief that affects all applications. Very few recovery events are caused specifically by a disaster (thank heavens!). Instead, most are operational recoveries. While some of the DR infrastructure may be helpful in some cases for providing operational recovery, simply being able to recover data from a backup in the cloud may be all that is necessary.

How Cloud Definitions Impact Data Protection

The National Institute of Standards and Technology (NIST) in Special Publication 800-145 define public cloud as "The cloud infrastructure is provisioned for open use by the general public. …It exists on the premises of the cloud provider." In contrast, private cloud is defined as "The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises."

We know this, but consider it in relation to data protection and the implication of a third party's key role in providing a service, such as backup-as-a-service (BaaS), recovery-as-a-service (RaaS), or disaster recovery-as-a-service (DRaaS), even if no such label is used. That implies a level of trust towards that service provider that must be in place from the very beginning of a cloud engagement.

If an organization builds and hosts its own private cloud without service provider help that is commendable, but is not very different from traditional implementations. Now, some vendors, such as backup/recovery vendors, have products that can work in traditional, private, and public infrastructures, but our focus of data protection in the cloud will focus on services-provided. While the definition of private cloud seems correct, the roles of the organization and a third party are not decoupled.

A term called "managed private cloud" solves this problem. A managed private cloud is where a service provider supports specific services in a cloud for each organization individually. This is in contrast to multi-tenancy where a public cloud provides isolated access to the same pool of infrastructure to multiple organizations. Thus, if a service provider delivers DRaaS and uses its own data center as a cloud, then that cloud can be referred to a managed private cloud (even though technically it could simply be called a private cloud).

What Requirements for Data Protection might be supported by a Cloud?

Let's say that an enterprise wants to consider what to do with cloud services. First, IT has to make an inventory of all workloads. What workloads are run in-house? Of those workloads are any candidates for traditional outsourcing or moving to a cloud? If they move, their BC/DR requirements would move with them. If they do not move, what functions for backup, DR, and BC need to be performed for each application?

Rudyard Kipling classically defined six serving men (what, why, when, how, where, and who) of journalism but we can use this model for this non-journalistic purpose. Take each workload and fill in the six related blanks (obviously this just a starter for thinking and not a full methodology).

Table 1: Apply the Six Serving Men to Each Application's Need for Data Protection

		Backup	Disaster Recovery	Business Continuity
What	What needs to be done?			
Why	What is the business case for doing data protection?			
When	When does the data protection need to be provided, such as continuously or at discrete intervals?			
How	Will it be a managed private or public cloud or some combination?			
Where	Will it be on-premises, off-premises, and what combination?			
Who	Who takes responsibility for the different functions?			

Source: Mesabi Group, October 2014

Filling out the table for each workload helps put you in the driver's seat when dealing with vendors (as they can have a good story, but unless you know what you want you may become mesmerized by what they say). Thus you can get a rough understanding of what you need and use this as a basis in contacting vendors.

Mesabi Musings

The cloud is, of course, a hot topic in the IT world. But how does data protection fit in the cloud? Although it is not the end all and be all of data protection, for purposes of the cloud, we have to distinguish among backup, disaster recovery, and business continuity and the parts or functions of the business they relate to. Otherwise, the terms can be bandied about to create confusion and unnecessary complexity.

Then we have to understand that data protection in the cloud is a service provided by a third party. That service can be provided through either a managed private cloud or a public cloud. Finally, we have to get a general understanding of whether the data protection requirements for each workload might fit in the cloud. Then each organization has enough to get started in evaluating alternatives for data protection in the cloud. Now, this is only a start towards understanding data protection in the cloud, but it should be enough to get you on your way.