Data protection in an era of massive data breaches

Data protection of digital data is a fundamental and mandatory responsibility for all organizations. Therefore, organizations need to understand the basic principles and concepts of data protection, especially in our current era of massive data breaches. To satisfy that need, the Storage Networking Industry Association (SNIA) has developed a technical whitepaper to provide the industry with a vendor-neutral overview of the relevant best current practices for data protection at the storage level.

By Thomas Rivera, CISSP, Chair of SNIA Data Protection and Capacity Optimization (DPCO)

Data protection is traditionally viewed as the execution of backup operations that are assured of providing data recovery if a loss of the original data (production data) occurs. In fact, data protection encompasses much more than backups and recovery techniques, such as dealing with issues related to data corruption and data loss, data accessibility and availability, as well as compliance with retention and privacy rules and regulations.

There are many factors to consider when it comes to data protection at the storage level. The main areas fall into three data protection “drivers”. These are data corruption and data loss, accessibility and availability, and compliance. Protected data must meet intended uses for all three drivers. Preventing data corruption and data loss ensures that the data is what the organization expects it to be when the data needs to be used. Accessibility and availability relate to the data being made available in a timely manner for intended uses. Compliance ensures that the data usage meets all associated legal and regulatory requirements.

Data corruption and data loss

Data must be protected both logically (to prevent data corruption from hacking or other external threats) and physically (in the case of data loss or the irreversible failure of a storage device). Physical prevention of data loss from hardware failure on a random-access storage system can use techniques such as RAID or erasure coding.

Backup and recovery are two of the traditional cornerstones to data protection for both physical and logical reasons. Backup relates to the processes of providing a copy of the data at a point in time and recovery refers to the ability to restore data for intended application use according to the organizational Service Level Agreements (SLA). One approach on a storage system itself is through the use of snapshots. These snapshots may serve as the basis for the data that is copied to a backup target storage system. Other approaches include the use of continuous data protection (CDP) or to use a public or private cloud as a backup service.

Cloud backup refers to backing up data to a remote, storage-as-a-service (public, private or hybrid). A cloud backup service is not a pre-defined, fixed solution and must be considered in the overall context of a business data protection or disaster recovery strategy. Cloud-based

backup appeals to many businesses because it offers a low-cost way to protect business data off-site but there are many different considerations to be aware of when planning such an implementation.

Moving the business data into the cloud is the easy part. Getting it back when you really need it is when things can get challenging. For this reason, it is important to understand all the imperatives before embracing public cloud-based data backup as part of the data protection and disaster recovery strategy.

When considering backup to a cloud provider, it is imperative to define the business requirements. These requirements may include business demands, SLAs and Quality of Service (QoS) levels for the backup data, along with the required skills for deployment of the cloud-based backup technology. It is also important to pay careful attention to the network design that will connect the cloud service provider to your data center. What network currently exists, what security does it offer, is there enough bandwidth redundancy, latency, etc.

Replication and mirroring are also used to make copies of data. Replication refers to point in time copies whereas mirroring provides for continuous writing of data to two or more targets. Replication may be used for both physical and logical data protection while mirroring is a physical data protection approach.

An archive is an official set of more or less fixed data that is managed separately from more active production data. As such, copies have to be made for data protection purposes, but more active measures, such as standard backup or mirroring are not necessary.

Accessibility and Availability

For accessibility and availability, Business Continuity Management (BCM) includes the processes and procedures for ensuring ongoing business operations. One key aspect of BCM is Disaster Recovery (DR), which involves the coordinated process of restoring systems, data, and the infrastructure required to support ongoing business operations after a disaster occurs. But a BCM plan also includes technology, people, and business processes for recovery.As part of accessibility and availability, basic infrastructure redundancies need to be provided, including UPS systems to provide redundancy for power in case of a power outage and extra network and power connections.

Compliance

Compliance includes the application of specific technologies that allow for the ability to secure data for meeting the appropriate rules and regulations typically related to data retention, authenticity, immutability, confidentiality, accountability, and traceability, as well as the more general problem of data breaches. There are a number of technologies that relate to compliance including:

- Long term retention of archival information is useful for integrity, immutability, authenticity, confidentiality, and provenance purposes
- Encryption provides support for confidentiality and integrity reasons.
- Data sanitization, i.e., electronic data shredding, provides for the proper deletion of data at the end of its life-cycle
- Monitoring and reporting gathers access information to determine if data tampering or data breaches have been attempted or have taken place

The two sides of data protection

Data protection is an important component of any Information Technology (IT) system, and the methods used for data protection and how they are configured have important inter-relationships with other aspects of the data center. By its nature, data protection has two sides: the backup or replication side and the restore or recovery side.

The backup side of data protection is the process or processes performed on a regular, or even a continuing basis to create one or more copies of an organization's primary data at a particular point in time. Backup processes may well differ from one type or subset of data to another, and they must be chosen with care to minimize the impact on the availability of primary data to all applications and users that need it. The backup must also provide for recovery of data in the way prescribed by the organization's Service Level Agreements (SLAs) with regard to each set of data. Thus traditional daily backups (copies of data to a different media) may be used for some subsets of data, while a real-time mirroring process may need to be used for other, highly critical data sets, in order to facilitate faster restores of the data.

Successful recovery operations are the result of having put appropriate backup processes in place, and recovery of lost or corrupted data is vital to an organization's health. A recovery operation may be required just to replace a file that a user accidentally deleted or a corrupted set of data (operational recovery), or to replace a major portion of a data center or an entire data center, in case of a disaster such as a multiple device failure, a virus, denial of service attack, or the destruction of a data center by a fire or flood (disaster recovery).

There are two important considerations or objectives for data recovery that in turn determine how it needs to be backed up; they are the Recovery Point Objective (RPO) and the Recovery Time Objective (RTO). RTO and RPO are important factors in deciding what backup or replication strategy the business needs to use, and they need to be a part of any organization's SLAs with regard to data protection.

Data protection and digital archives

Although a digital archive represents another set of copies of primary data like those intended for backup or disaster recovery (DR), an archive is more immutable in nature, with changes either not allowed, or strictly controlled by a journaling process. Also, archives themselves require data protection; they are not intended to be used for data protection.

Archives may be divided into two types based on their intended longevity, with those intended to last more than ten years being considered a long-term archive's. Long-term archives typically require different methods for storage, security and management.