The Art of Eliminating Downtime to Achieve High Availability

by George Crump, Lead Analyst

Downtime, no matter what the cause, is becoming less acceptable to organizations of all sizes. Clearly, technology exists that can eliminate it. But with the reduction or elimination of downtime, the associated costs – when compared to once-per-night backups – increase. That means the trick in eliminating downtime is identifying which applications or data sets in your environment cannot experience downtime and then selecting a cost-effective solution that meets the goal of eliminating downtime. In turn, IT creates an enterprise where critical applications and services are always available no matter what threats impact the data center.

Understanding Your Four Zero-Downtime Options

Generally there are four types of protection options available to organizations. Each offers differing degrees of availability.

- **The Backup Process:** The first is the traditional, and well known, backup process. It typically runs once-per-night and attempts to capture all new data or changes in existing data since the last backup from the previous night.
- **Snapshot-based replication:** These are replication processes, typically built into a storage system, that leverage a snapshot process to identify data to copy to a remote site. A snapshot is a virtual view of a volume or file system from a particular point in time. After taking a snapshot, users can update the original volume or file system but the snapshot software preserves the view of that volume from when it took the snapshot. It does this by preserving original copies of any blocks that change after that point in time. Most organizations take snapshots at a range of 30 minutes to four hours.
- **Active replication:** These are typically stand-alone software solutions that replicate data to a remote site as users change or add data at the primary site. Since this solution is software-based, it can, in near real-time, replicate from any storage system to another storage system – including the cloud. The only risk of data loss is the data that is "on the wire," when a failure occurs.
- **Synchronous mirroring:** Makes sure writes are written to two different locations at the exact same time. It is less a common method but worth mentioning. With this option, changed or new data is written simultaneously to two storage arrays and is not considered complete until an acknowledgment is received from both arrays. While providing the ultimate in downtime reduction, it is also the most expensive and the most vulnerable. Synchronous mirroring requires a very high speed (and expensive) connection between locations and an almost identical storage system. Because of the requirement of an acknowledgement from the second site, the distance between the primary and secondary site must be relatively short. That means the likelihood of both the primary and secondary site getting caught in the same disaster is high. Finally, synchronous mirroring almost certainly eliminates the use of the cloud due to latency concerns.

The type of protection process impacts how close the organization can get to high availability. For example, the typical backup job will potentially lose 24 hours of data,. A snapshot-based replication process will typically lose an hour or more of data. But an active data replication process may only lose a few seconds to a few minutes of data, which for most organizations meets the high availability standard. Synchronous mirroring also meets the high availability requirement but again cost issues tend to severely limit its use.

How Recovery Time Impacts High Availability

The second component of a high availability strategy is time to recovery: How long it will take the application or data set to be moved or transformed from its protected format to a usable format. This is often called the Recovery Time Objective (RTO). A backup job may take many hours to restore data from backup storage to production storage. Snapshot and replication jobs typically store data in its native format, assuming these jobs are occurring to production suitable storage, then restart of the application should take only a few minutes.

Once the application is in place and running, the organization also needs to consider its data exposure window, which is essentially how much data needs to be re-entered. In the database use case, the exposure is the size of the transaction log that needs to be replayed (assuming they were protected separately). Recovery Point Objective (RPO) is the amount of data loss during the recovery. For applications or data sets that change rapidly or where data loss can't be recreated, organizations need to take steps to shorten the time between protection events.

All applications should get protection from an enterprise backup solution that can protect all of the organization's applications and data, and it's often true that most data only needs protection with that process. The organization should then identify which applications need an extra layer of protection. Organizations needing near-zero downtime should select something like snapshot or active replication. Even synchronous mirroring may have a role for some organizations that need true-zero downtime, despite its expense, for the most mission critical of applications.

Developing a Replication Target Strategy

Replication software, be it snapshot-based or active, needs to provide flexibility in the destination. For most near-zero downtime situations, it is advisable to have a secondary target on-premises and one in an alternative location either owned by the organization or in the cloud. The reason for the second on-premises storage system is most "disasters" are not data center loss events. Instead they are the failure of a server, application, network or storage system.

Architecturally, the replication software should replicate data as it changes from the first storage system to the second on-premises storage system and then a third time – to a storage system in a remote site or cloud. One of the values of active replication is its flexibility. The secondary storage system and DR site storage does not have to be the same as the primary storage system. It could be a lower cost, lower performance system since its primary role is to be a standby. Ideally the secondary system will snapshot data periodically because with active replication there is a chance that replication job will replicate corrupted data. The snapshot on the second system allows the organization to roll-back from that scenario.

The flexibility of active replication software is also critical for organizations wanting DR storage to be cloud-based since the cloud provider will almost certainly not run the same storage as the organization does on-premises. The enabling of the cloud as a disaster recovery site is an important justification for using active replication. Because the chance of an organization having a total data center failure is rare, the cost of a second site, where it pays for standby compute and storage, let alone the cost of the facility itself, is too burdensome. Most organizations that don't already have a second data center class facility will find having a DR site on "retainer" an ideal solution.

Deciding Between Snapshot Based Replication and Active Replication

Organization will need something more than just backup to meet the organization's expectations for disaster recovery. And most organizations will rule out synchronous replication due to costs. That means they will have to decide between snapshot-based replication and active replication. Snapshot-based replication is often factored into the price of the storage system, so it appears "free" to the customer. But there are hidden costs and areas of exposure to snapshot-based replication.

First, in most cases, the snapshot-based replication is included with the storage system hardware and requires a secondary storage system from the same vendor, often the exact same model and configuration. The problem is that in the three system configuration we describe, three of the exact same systems will get very expensive very quickly.

The second challenge with snapshot-based replication is the time lag in between protection events. While compared to a backup, a snapshot executed every four hours is an improvement, but it still may not meet the requirements of more critical applications. Active replication can deliver a much tighter RPO/RTO at an often significantly less expense in hardware and facilities costs.

Backup software vendors are starting to offer replication as a part of their backup processes. These still suffer the same time-lag concerns, and they essentially force you to commit to their backup solution. Look for a replication vendor that can complement the existing backup solution instead of replacing it, at least initially, with the option to add a backup solution from that vendor in the future.

Conclusion

The downtime expectations of organizations, their users and customers will increase but IT budgets will not. IT professionals need to look for new strategies to meet these new demands. Active replication combined with enterprise backup provides a very thorough data protection strategy and rapid recovery with minimal data loss of mission critical applications.

Sponsored by Carbonite