FROM THE SOLARWINDS BLOG

Archiving vs. Backup—What's the Difference?

By Brien Posey

Brien Posey is a third-party content contributor and not a SolarWinds employee.

Although the terms "backup" and "archiving" are sometimes used interchangeably, they refer to two completely different processes. Both deal with long-term data storage, yes, but that is where the similarities end.

Backups are designed to provide an organization with a way of recovering its data following some sort of catastrophe. As such, a backup can be thought of as a restorable, point-in-time copy of an organization's data or of entire systems.

Conversely, archives are not designed to act as a rescue mechanism following a data loss event. Instead, archives are typically meant to be used as a tool for reducing storage costs.

THE CASE FOR ARCHIVING

It's no secret that organizations accumulate data over time. The problem with data accumulation are the costs associated with storage. If an organization were to allow data to accumulate indefinitely, then that organization would have to periodically add primary storage to accommodate all the new data being created.

Data archiving is based on the idea that data can become stale over time. This is especially true for unstructured data. A five-year-old PowerPoint® presentation, for example, probably doesn't get used every day. Data archiving systems are designed to identify data that has not been accessed recently, and then move that data off the organization's primary storage and onto a less expensive medium.

Data archiving solutions vary considerably, but can generally be grouped into two different categories:

1. Online archiving
   These solutions move data onto typically less expensive storage, but keep the data online and accessible to its owner. These types of solutions might, for example, be based around cloud storage or an on-premises storage array that uses high-capacity, but low-performance, hard drives.

2. Offline archiving
   Offline archiving can often be more cost-effective than online archiving, but it usually sacrifices easy accessibility. Offline archives are commonly written to tape.
   In the case of an online archiving solution, a user might not even realize their data has been archived. Opening an archived document might take a little bit longer than the user would expect, but the document will remain available for use.
   Offline data archiving solutions, however, sometimes give the illusion that files have been deleted. In some cases, archived files are still displayed within the file system, but an icon or other indicator shows that the file has been archived. In either case, the user is usually unable to access the archived file without getting assistance from the IT department.

WHY ARCHIVING AND BACKUP ARE VERY DIFFERENT

You might ask yourself, "Why would an organization even bother with offline archiving? Why not just delete data that is unlikely to be needed again?"

The reasons for retaining aging data vary from one organization to the next. In some cases, an organization may be required by law to retain data for a specific length of time. In other cases, there may be a corporate policy stipulating that aging data must be retained in case it is ever needed in the future.

So why can't you use backup software to accomplish this? Why not just back up the aging data and then remove it from primary storage?

WHY TIME MATTERS

Backup solutions tend to overwrite—or retire—backups after a specific length of time. Depending on how an organization's backups are set up, the system may only retain backups for a month or two. Archives, on the other hand, can often store data for decades.

Furthermore, an archive system can sometimes automate the process of identifying aging data and moving that data to the archive. If you were to use backup software as a solution for archiving data, this would typically largely require a manual process.

In other words, backup and archiving serve two completely different purposes. Backups are designed for recovering data following a catastrophic event, while archives are built to move aging data off primary storage to help reduce storage costs.

Brien Posey is a 13-time Microsoft® MVP with over two decades of IT experience. Prior to going freelance, Posey was a CIO for a national chain of hospitals and healthcare facilities and served as a network engineer for the United States Department of Defense at Fort Knox. Posey has also worked as a network administrator for some of the largest insurance companies in America.

You can follow Brien on Twitter® at @BrienPosey