The Register®

How backup fits into your data-protection strategy

The over-arching rule remains 3-2-1

Joseph Martins Tue 3 Nov 2020 // 07:00 UTC

Data is like water. It has to circulate around your organisation, otherwise your survival is in peril. And while you need it in the short term, day to day, you also have to plan ahead, for the times when that flow might be disrupted. But there's a point at which the analogy breaks down – data tend to begets more data. And this also has to be managed and stored.

So where do backup and archive fit in these days? To understand this, let's recap what each is. Although the two terms often appear next to each other, they are not interchangeable. In fact, it's better to think of the two as separate but equal parts of an overall data protection system.

The crucial point to grasp is that backups are – today – generally focused on keeping track of constantly changing business information and may well be short term and quickly overwritten. An archive, by contrast, is all about long term retention of the data. Clearly, a bank that has no access to its customer or financial data will struggle to do banking. But for a broader range of organisations, those real time analytics and machine learning inference jobs aren't going to happen if the right data is not available at the right time.

Data resiliency

Let's examine what you need to take into account when putting together the backup element of a data protection system for your storage infrastructure.

Production enterprise storage systems may include some data protection mechanisms. For example: snapshotting, which enables rolling back to an earlier version in the case of a problem, and data replication, which ensures that if a component goes down, there is still a copy of the data available on the system.

But these alone aren't going to give you the sort of robust data protection a truly data-centric organisation needs. For example: in the event of data corruption, or even worse having your system hijacked by ransomware, you may be left with two equally useless datasets. And just how many snapshots – and replicated copies – of your data do you want to keep on your very fast, and very expensive, production storage? It's not just a question of disk space, but the networking, compute and management overhead it inevitably sucks up too.

This raises the issue of recovery point objective (RPO) and recovery time objective (RTO). The former term defines how much data you're prepared to lose in the event of an outage, while the latter sets out how quickly you want to be back up and running. Data replication or snapshotting might help you work around a small disruption, depending on your replication or snapshotting schedule. A bigger threat, like data corruption or ransomware, might require reverting to a full backup. If you have one, of course.

The issue of ransomware also illustrates why airgaps in your storage infrastructure are crucial.

Air-gapping should be a key component of any comprehensive data protection infrastructure. Anything connected stands a chance – however remote – of being accessed by bad actors. In such incidents, companies face the prospect of not just losing business, but also of losing trust amongst their customers. However, when you physically separate live data and backup data – creating an air gap – your backup data is offline and therefore unreachable for data thieves and hackers. Storing the data on removable media – most likely tape – and moving the data to a secure location off-site hardens defences further. In the event of a cyber attack, organisations can, with the appropriate backup software, quickly recover clean data from the air gap, to restore operations.

Cloud backup

It makes sense to demand that your chosen backup hardware and software is able to take advantage of the cloud This potentially offers a limitless amount of storage for backup , snapshots and other copies of data. But depending on your broader infrastructure strategy, you also need to consider whether the cloud is the best medium to use for rapid data recovery. With the best will and the widest pipe in the world, restoring your on-prem production data from the cloud could be a tortuously slow process.

If you're looking for a framework to help formalise your data protection strategy, the over-arching rule remains 3-2-1. That is, you have three copies of the data, two of which are on a different storage media – for example, one on an appliance, and one in the cloud. One should be in a remote location, and this could be offline data, such as tape, or it could be in the cloud.

Another formula is bronze, silver and gold, where bronze assumes one copy in one location, and a retention period measured in days. The focus here is on fast backup. Silver assumes up to two locations, with one or two copies on disk, and at least one copy on tape. Gold level assumes two locations, two or more copies on disk, and at least two copies on tape, with one of these in a safe. The retention period here would be months and represents the maximum focus on data availability and fast recovery.

All this will help you begin formulating a data protection policy. But what do you need to put it into practice?

Appliances

Unless your data needs are very modest, you'll likely want to consider dedicated backup appliances, which work in conjunction with backup software suites to harvest data. Those appliances might be considered integrated or target appliances. The former offer a tightly integrated combination of pre-configured hardware with backup software suite such as Commvault/Veritas. The latter are not tied to a specific backup application and can be used with any backup software suite.

If virtual machines, as well as the cloud, are a significant part of your workloads, you might also need to consider specialist services such as Veeam, which are designed to back up virtual environments – whether within the cloud or on-prem – and enable fast recovery.

Large organisations, with multiple locations, will likely want to take a consolidated backup approach, with a large central target appliance, supplemented with cloud backup or dedicated local appliances for satellite locations.

As for what's inside your appliances, while tape remains the cheapest option when it comes to storing data, in reality, disk will likely be your favoured backup option, particularly when it comes to restoring data quickly. Production storage has increasingly moved to flash, which as well as being faster and more reliable than traditional spinning media, also offers benefits in terms of power consumption and space. While flash drives remain more expensive than traditional drives, they are also increasingly finding their way into backup hardware, particularly where the ability to recover quickly to get the business back up and running is at a premium.

And keeping price performance in mind, it's worth remembering that cloud storage, as well as offering challenges when it comes to quickly recovering data back to on-prem systems, can also quietly spiral in cost, particularly if you're not being selective about what you're backing up there.

Your data storage infrastructure may, and probably should, encompass a range of formats and platforms, both on-prem, remotely, and in the cloud, all geared towards whatever your organisation's mission happens to be. Your data protection system will probably be as diverse. While the 3-2-1 rule is useful, your own organisation's characteristics mean you'll want to tailor it to ensure you have the right balance of data protection, data availability, and price performance.

So putting a comprehensive backup strategy in place requires careful planning above and beyond the raw investment in appliances and software. But once you've nailed your SLAs, the high degree of automation state of the art software and hardware offer will ultimately pay dividends, beyond the reassurance that data is backed up and available quickly, because that high level of automation massively reduces the possibility of human error.

You should be wary of anyone that promises you a one-size fits all solution. One way to work out if it's worth continuing a conversation with a potential data protection partner? Ask them about the difference between archive and backup.

Sponsored by Fujitsu

Published on "The Register", November 03,2020

https://www.theregister.com/2020/11/03/your_data_protection_strategy/

If you want to know more about different storage solutions, check out the other articles published on "The Register" or explore the storage data jungle: www.fujitsu.com/data-jungle/storage/

The key role of storage in building a datacentric organisation
https://docs.ts.fujitsu.com/dl.aspx?id=57cc1295-ba0b-46dc-8882-40db4fe8d188

The persistent importance of traditional storage
https://docs.ts.fujitsu.com/dl.aspx?id=d82ce000-1188-4d69-985a-d9b102cc14c3

SDS or HCI? How to figure the right fit for the right workloads
https://docs.ts.fujitsu.com/dl.aspx?id=981e167e-2f7e-4cd9-baa6-2f73a47ec2f9

Your archive contains a treasure trove of information. Use it
https://docs.ts.fujitsu.com/dl.aspx?id=6ed0abd2-3d22-43d6-b073-3386dd2a0077