The how and why of protecting today's big data and open source workloads

By Mark Coletta

13 May 2019

As organisations grow in size, complexity and the services they offer so too does the diversity of the workloads they must run. Modern companies operate huge applications, some of which must process billions of requests each minute from all around the world. Alongside this data flood, organisations must also contend with new kinds of data - from images to videos to social media posts, much of this is unstructured and difficult to tabulate in traditional databases.

Mark Coletta is senior product manager at Veritas Technologies

In response to this, organisations are increasingly embracing big data and open source databases which are advanced and flexible enough to scale with their needs. However, outdated attitudes to data security present a worrying chink in the armour.

The data protection measures of the past are no longer fit to maintain the workloads of today. You shouldn't assume your big data or open source databases are secure because you periodically replicate them to another server.

While many big data and open source databases offer some form of protection – such as replication, snapshots, or even built-in recovery tools – they lack the point-in-time backup and recovery capabilities needed to achieve enterprise-grade data protection. The stakes are too high to let a workload go down, so organisations must cover all the bases from backup and recovery to analysis and data management.

Causes for concern

Time is of the essence. Big data and open source solutions have already become crucial to businesses. According to IDC, big data will soon become a $260bn market and more than a third of organisations work with scalable big data solutions. At the same time, open source ranks among the most popular databases companies use for mission-critical applications.

5 SME big data myths demystified

15 Apr 2019

Big data and open source environments demand a range of security requirements, including the point-in-time recovery of

historical analysis and fast data recovery. Back up is critical – these workloads carry critical data and services no organisation can afford to lose.

The primary reason to create backups is to protect against accidental or malicious data loss due to logical or human error. In today's digital economy data is precious. Losing it can lead to gaps in valuable insights or missing out on crucial opportunities. With GDPR now in force, it also poses the risk of severe regulatory penalties and reputational damage. Backups with off-site copies protect against site outages and complete disaster situations that result in the destruction of equipment and data environments.

Yet there are also benefits from a resilience point of view. According to a Ponemon Institute study, the mean cost of an unplanned outage per minute is $8,851, or approximately $530,000 per hour. As many big data or open source database environments are utilised for mission-critical applications, this could have a significant compliance or financial impact your organisation.

Bringing AI to the data centre

Jim Holland 17 Apr 2019

Many companies may see replication solutions as sufficient to the task, but they don't go quite far enough. Replication does provide real-time to near-real-time protection, but it doesn't protect from logical or human error – whether accidental or malicious – that can result in data loss. Replication can also lead to expensive and resource-hungry clustering, using up unnecessary storage space when it may already be in short supply.

Manual recovery similarly has its shortcomings as a protection measure. While it's sometimes possible to reconstruct data from the original data sources through manual recovery, in most situations the data will either be lost or unavailable from the source. That or the reconstruction process is time-prohibitive.

The point of protection

As enterprises increasingly rely on non-traditional workloads for their mission-critical applications, it's more important than ever to understand the need for reliable backup and recovery. To reduce the complexity of these environments, a single unified strategy is crucial.

It's not unusual for some organisations to see data growth of between 40% and 60% a year. To keep up, you need backups that run as fast as possible without disrupting production activity. Businesses should look to modern, parallel streaming architectures to eliminate bottlenecks and optimise storage for these demanding scale-out, multi-node workloads. Big data workloads can grow drastically in a short space of time, so it's also important that solutions can scale automatically and be responsive as the needs of these workloads evolve.

Securing your data across hybrid and multi-cloud environments

Modeen Malick 12 Apr 2019

Proper protection doesn't have to be just another cost of business. By connecting all environments under one system, it becomes easier to source and utilise previously siloed data. According to IDC, organisations that can analyse all relevant data and deliver actionable insight will generate $430bn more in productivity benefits by 2020 than their less analytically advanced peers. Backup, once an afterthought for big data volumes, is now invaluable.

Next generation big data and open source workloads enable digital transformation in enterprises. Without them, companies would not be able to generate the insight or develop the innovative features and applications they depend on for relevancy. To stay competitive and relevant to customers, it's vital that organisations make proper protection their top priority.

ABOUT THE AUTHOR

Mark Coletta is senior product manager at Veritas Technologies