

REPORT REPRINT

Reining in data copies is an ongoing struggle

JULY 28 2020

By Henry Baltazar

In our Voice of the Enterprise: Storage, Data Management and Disaster Recovery survey, respondents highlighted multiple challenges related to having too many copies of data, such as increased administrative costs and the risk of a data breach.

THIS REPORT, LICENSED TO STORAGE GUARDIAN, DEVELOPED AND AS PROVIDED BY 451 RESEARCH, LLC, WAS PUBLISHED AS PART OF OUR SYNDICATED MARKET INSIGHT SUBSCRIPTION SERVICE. IT SHALL BE OWNED IN ITS ENTIRETY BY 451 RESEARCH, LLC. THIS REPORT IS SOLELY INTENDED FOR USE BY THE RECIPIENT AND MAY NOT BE REPRODUCED OR RE-POSTED, IN WHOLE OR IN PART, BY THE RECIPIENT WITHOUT EXPRESS PERMISSION FROM 451 RESEARCH.



S&P Global Market Intelligence

Introduction

It is a key requirement for organizations to have multiple copies of a production data set in their environments because these data copies add to the resilience and scalability of an organization's workload processing capabilities. Despite this market reality, it is important to note that having too many copies of a data set can lead to a string of unfortunate consequences, based on the findings of our recent Voice of the Enterprise: Storage, Data Management and Disaster Recovery 2020 survey, in which respondents tell us that, on average, they maintain five copies of their data sets at their organizations. With companies struggling to conserve resources during the COVID-19 outbreak, the need to rein in data sprawl has increased in importance.

451 TAKE

Data copies are necessary not only for backup and disaster-recovery purposes, but also to fuel a number of key business processes, including test/dev and business intelligence. With the rising importance of digital transformation and the drive to accelerate software development to create business value, IT infrastructure teams are being asked to make it easier for developers and business stakeholders to get copies of production data to meet their requirements. Initiatives for self-service data consumption models are creating difficult challenges since staffers must find a way to balance the demanding provisioning requirements of business stakeholders and developers with the need to maintain compliance guidelines and the possibility of uncontrolled data growth. As organizations seek to automate more of their infrastructure and data provisioning, the need for intelligent copy data management will likely rise since out-of-control copy creation could lead to infrastructure resource shortages and skyrocketing costs.

How many copies are too many?

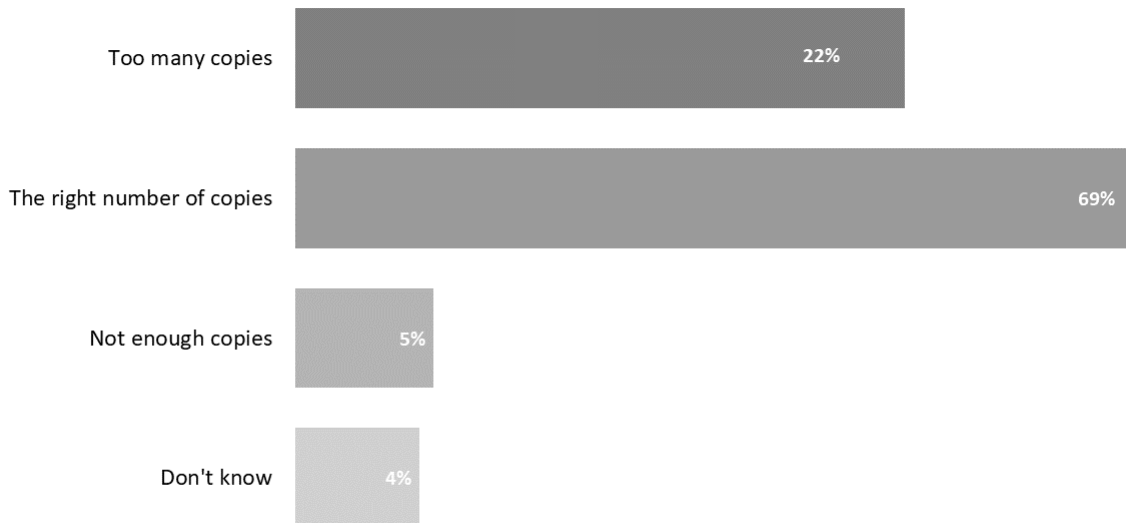
In the survey, the mean number of copies at organizations was 5.22, with a median value of four. The most popular selection for respondents was three copies, which was selected by 24.6% of respondents. Looking at the most extreme cases, 11.8% of respondents claimed to have 10 or more copies of data within their organizations.

While excessive data copies are a potential problem that organizations should be on the lookout for, only 21.7% of respondents thought they were storing too many copies of data, while another 4.9% thought they actually did not have enough copies of data (see Figure 1).

REPORT REPRINT

Figure 1: How would you describe the number of copies of business data your organization has across all locations?

Source: 451 Research's Voice of the Enterprise: Storage, Data Management & Disaster Recovery 2020



Base: All Respondents
Sample Size: n=203
Source: 451 Research's Voice of the Enterprise: Storage, Data Management & Disaster Recovery, 2020
Q. In your opinion, how would you describe the number of copies of business data your organization has across all locations (e.g., in primary storage, backups, archives)?

'Definitely two [copies], in some cases three [of our business data across the organization].. That's fine. It used to be more than that. It used to be four and five, but not anymore.'

– Midlevel manager, 5,000-9,999 employees, \$5bn-9.99bn revenue, Financial Services

In a discussion with a manager at a financial services organization, their team was able to cut down the number of data copies from four or five copies to two or three copies, to rein in their storage consumption. On the other side of the spectrum, we spoke with IT managers that had issues eliminating excessive data copies.

'... We store working data, working tables and then the final tables multiple times. There's backups. There's your active copies, the DR. So just in that, there's many copies. And then a given transaction may be stored by multiple systems for multiple reasons. ...It's too many [copies]. Getting to eliminate a lot of that gets to be very difficult.'

– Food, Beverage & Agriculture, 10,000-49,999 employees, \$2.50bn-4.99bn, IT Engineering

There are a number of factors that could be leading to the storage of an excessive number of copies, and that point to gaps in an organization's data management policies and enforcement. Unenforced or missing data deletion and retention policies are big contributors to the copy data issue. In this study, 26% of respondents claim that even though their organization has data deletion policies, these policies are not followed consistently, with another 21% saying they have no deletion policies whatsoever.

Looking at the issue from the data retention side, 29% of respondents claimed that despite having data retention policies, their organization was not consistently following them, with an additional 14% saying they have no data retention policies. Vendors and service providers should provide tools and services to give customers insight to help identify and prioritize content.

REPORT REPRINT

Developers are looking to create applications as quickly as possible to generate business value, and often look to create and test applications while using production data, which inevitably leads to the spawning of additional copies of data. With the rise of self-service data consumption models, this threat of out-of-control copy data management could become far more significant in future years.

An interview with an IT engineer at a large consumer retail vendor revealed the extremes to which excessive data copies could spiral out of control.

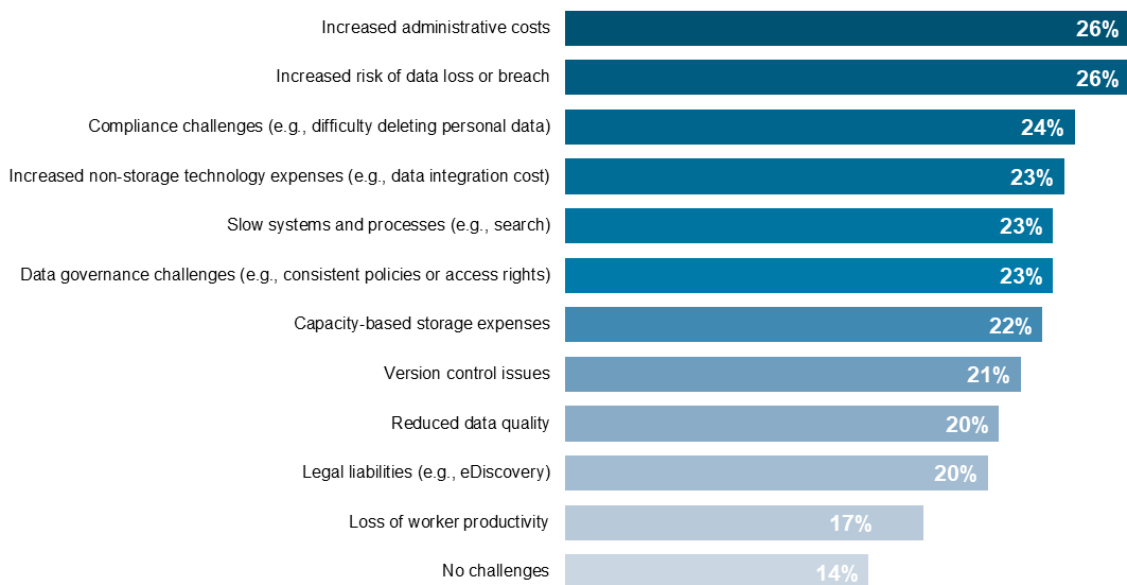
'[There are] probably 40-50 copies [of business data]. Way too much data. Now, there's obviously a little bit of iterations between some of it. But they have data duplicated so many places, and they back that up, and then they're replicating it, then they're backing up the replica.... Too much of the same copy.... Every time they open something up and look for something, they're finding multiple copies. It is a mess.'

– Consumer Retail, 10,000-49,999 employees, \$2.50bn-4.99bn, IT Engineering

In this scenario, the creation of data copies combined with the use of traditional backup and replication data protection capabilities led to the creation of dozens of data copies. While backups and replication are necessary to protect production data, applying these capabilities to all data sets, including noncritical ones, is wasteful, especially if the most relevant production data set is already adequately protected.

Figure 2: Which of the following data-related challenges is your organization experiencing by having multiple copies of its business data?

Source: 451 Research's Voice of the Enterprise: Storage, Data Management & Disaster Recovery 2020



Base: All Respondents

Sample Size: n=197

Source: 451 Research's Voice of the Enterprise: Storage, Data Management & Disaster Recovery, 2020

Q. Which of the following data-related challenges, if any, is your organization experiencing by having multiple copies of its business data? Please select all that apply.

Having too many copies of data in your environment can lead to a number of different negative outcomes (Figure 2). In this study, increased administrative costs was tied with increased risk of data loss or breach, with 26% of respondents selecting each of those as top challenges. Compliance challenges were another major negative for having too many copies of data.

REPORT REPRINT

Inefficiently hoarding data often produces negative consequences, as well, especially if an organization does not do a good job of assessing the value of a file. Twenty-one percent (21%) identified version control issues and 20% named data quality as negative consequences of having too many copies of data.

Recommendations

The data copy control challenge is not a problem that can be solved by haphazardly purging copies of data from an environment. Storage and infrastructure professionals need insight from their business stakeholders and data management tools to determine which copies of data are relevant and which can be eliminated without impacting other operations. As we see in this customer example, organizations keep copies of data for the sake of convenience and a feeling of safety. What they should be doing instead is investing effort into identifying and maintaining the right copy of data to eliminate many of the issues we mentioned.

‘We have these multiple copies. And even the retention for that, those copies are not properly understood.... In any organization, there is a huge overemphasis to be on the safe side, to keep copies rather than just having one right copy.’

– Financial Services, 100,000+ employees, \$10bn+, Senior Manager

Many of the key storage vendors currently have data management and discovery tools to help customers locate and analyze data within their environments, and this will continue to be an area where we expect to see innovations. That being said, purchasing a tool in vacuum without the support of business stakeholders or the resources to properly deploy them could lead to unsatisfying outcomes. In this particular interview with a IT engineer from a large communications, media and publishing organization, even though they felt that a data management software tool would be able to reap benefits, the lingering fear was that the tool would quickly become ‘shelfware’ if it was not properly implemented and integrated into the business.

‘[For data management] we might have to buy a product, but the product would probably pay for itself in a year, but it’s hard to get the buy-in on buying the product.... There seems to be a little bit of a reluctance to buy tools because there’s a feeling that we’re going to buy this tool and it’s just going to sit there and collect dust, and nobody’s going to use it.’

– Communications, Media & Publishing, 10,000-49,999 employees, \$5bn-9.99bn, IT Engineering

This example shows the importance of getting executive buy-in when implementing this sort of solution. On the vendor side, just selling the tool alone is not enough to ensure customer success. Vendors need to be more proactive with their onboarding and training programs, to ensure that their tools are put in a position where they can deliver value immediately, and where they can become integrated within a customer’s infrastructure and processes.

As organizations become more automated in the future, policies for provisioning data copies and the proper elimination of excessive ones will need to be integrated with data management tools and processes to ensure that the convenience and speed of automation do not lead to wasted infrastructure resources.