Golden Copy Storage

What, why, when and how?

in association with

Western Digital.

Introduction

Every organization believes that the data it holds is important, and most will also recognize the need for comprehensive data visibility. Users might need up-to-date information to make business decisions or real-time operational adjustments, say, or to meet legal requirements by preparing compliance reports for auditors and regulators.

And yet most organizations also recognize that users often experience problems and delays when looking for the information they need (Figure 1). Worse still, the cost can be high if they fail to find it in time. As well as direct business losses and of course the frustration it causes to users, failure or delay could damage customer and investor confidence, bring punitive fines for regulatory non-compliance, and much more.
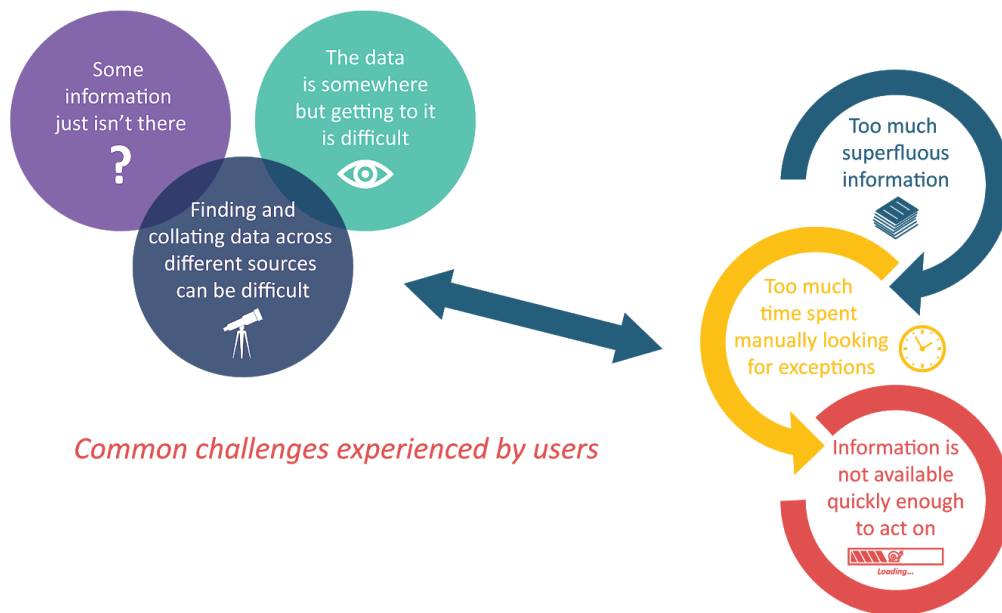

Figure 1 Finding data in a timely manner can be difficult

The challenge, therefore, is to enable authorized users to find accurate data in a timely fashion, while also ensuring that their access to it is effectively governed. One potential solution is to build a "Golden Copy" data store - a storage infrastructure solution that addresses these challenges by acting as a central system of record for the organization.

This paper will discuss why such a system is needed and what to bear in mind when implementing one. More importantly, it will look at how business pressures and technological developments have made the need for such systems more urgent than ever before.

What is a Golden Copy?

The concept is older than IT, and it refers to those individual data records which have been identified as the definitive version of that data. Golden Copy can, therefore, be described as "the single version of the truth" for any specific record.

Identifying such golden records is often a complex task that involves reconciling various versions of a data record held in different systems. This can be a tricky process just in IT terms, but when you add internal politics into the mix, agreeing on a consistent, repeatable process to identify and store the golden data can be both challenging and time-consuming.

Golden Copy storage

Extending the Golden Copy concept to encompass as many enterprise data sets as possible is the idea behind Golden Copy storage. It enables the organization to build a single storage set that any authorized user can use to quickly locate information and make operational decisions as part of everyday business. Some organizations have already started to build such repositories, but may have limited them to certain specific "systems of record". The Golden Copy storage platform has the potential to be far broader than that, perhaps moving to become the System of Ultimate Record.

Golden Copy data pools should not be confused with Big Data projects or Data Warehouses which are usually targeted at pooling together information in order to be able to run specific analytics. Indeed, one of the biggest challenges that Big Data projects often face is simply finding and cleansing data sources to create a version of the data suitable for use in the project. Data reconciliation is still a major challenge, and will likely always be so.

Why does it matter now?

Given that none of the issues mentioned above are new, why is the Golden Copy idea something to look at now? There are several reasons, some of which have hugely grown in importance recently.

Data growth: More and more information is being created daily, and much of it must then be stored safely. Until recently most of the new data came from users undertaking their everyday business tasks, or computer systems interacting with customers, suppliers or channel partners.

But now there is a new source of data creation: IoT. This is the data generated by the operational technologies used in business operations, whether that's recording data from customers' electricity supply systems, the industrial plants that create the products the company sells, or tracking the movement of stock through the supply and distribution chains.

These systems can generate huge volumes of data every month, some may even do so every minute and hour of the day. And there is value in these new data sets, both on their own and in combination with other enterprise information. The goal, naturally enough, is to generate

business value from this new mountain of data. Add in the potential impact of AI and Machine learning, and it is clear that new platforms for data storage are required.

The need for speed: Today's businesses must be ready to react quickly to developing challenges and bring new services to market fast. Doing this with mountains of data means paying attention to the challenges of data collection, cleansing and storage. The pressure users face to deliver results quickly equates to a need for storage systems to be equally responsive.

Data quality: For users to be confident in their data analysis, they have to be certain that the information and systems they work with are accurate and up to date. Having the trust of users is essential.

Finding data quickly: Just as important as the accuracy of the data is the ability to locate the information needed quickly and be able to access it without laborious request processes.

Real-time Business Intelligence for operations

Putting all of these elements together allows an organization to undertake business analytics in near real-time. And in today's markets, the ability to make fast and effective use of accurate information means you can make forward-looking operational business decisions, rather than simply producing backward-looking quarterly or annual reports.

"Just in time" business decision-making, based on a broad picture of what is really going on, has long been regarded almost as nirvana. The time has come to put in platforms ready to support this, and enterprise-grade Golden Copy storage could provide organizations with the foundation on which to build such systems.

The characteristics required

Given the business potential for Golden Copy storage, what would such a platform look like? Many of the desired characteristics follow logically from the demands previously highlighted:

- Performance - The platform needs to deliver a fast response to data requests despite holding a huge volume of data, potentially of many different types.

- Service quality flexibility - A Golden Copy storage platform needs to be able to store very large volumes of data and deliver it according to need at access time. This potentially requires the platform to be able to use many types of storage mediums but move data between them automatically based on policy.

- Access - Data held must be easily accessed by other systems as well as users without requiring IT professional assistance.

- Data integrity - The platform must be capable of storing data across its service tiers without risking any form of data corruption or loss, potentially over many years.

- Search and discovery - Given such a repository could be terabytes or petabytes in size, data must be easy to find by a variety of means.

- Management - Any data repository of scale demands excellent management tools, and the ability to automate processes. In particular, the platform should be able to automate all data movement between its tiers as and when required. It may even be possible to include historical data that may currently be held in cold storage in a Cloud or on Tape.

- Security - Different data elements will need different levels of protection, and the storage platform must be able to deliver these without impacting performance.

- Availability - A Golden Copy data store is likely to be required to operate 24x7, without service or maintenance interruptions.

- Trustworthy - If users do not trust the data held in the store, they will not use it. The same applies if the data cannot be found and accessed easily.

How these characteristics are delivered may also be impacted by the state of the existing IT infrastructure, namely how prepared is it to either form part of the Golden Copy platform or to act as feeds into it (Figure 2).
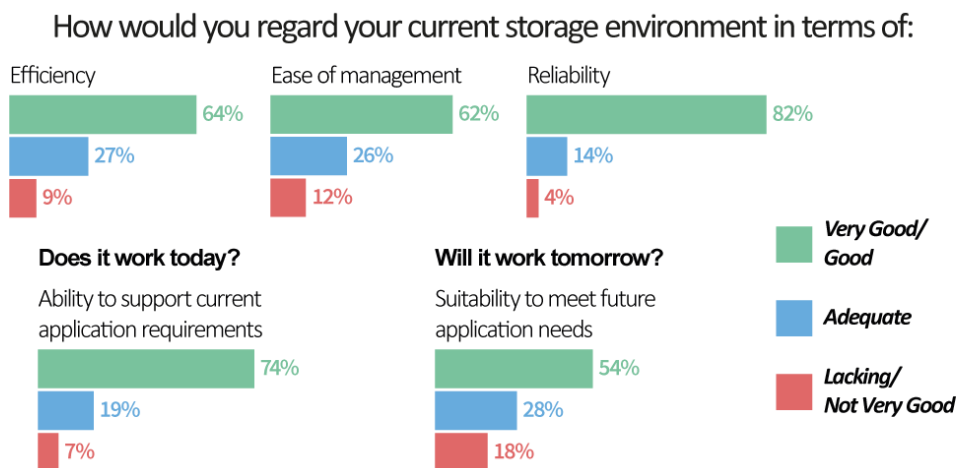

How would you regard your current storage environment in terms of:

Efficiency	Ease of management	Reliability
64%	62%	82%
27%	26%	14%
9%	12%	4%

Does it work today?

Ability to support current application requirements
74%
19%
7%

Will it work tomorrow?

Suitability to meet future application needs
54%
28%
18%

Very Good/Good

Adequate

Lacking/Not Very Good

Figure 2 Are storage platforms ready for tomorrow?

What does Golden Copy storage not deliver?

Clearly no single storage platform will have this vast range of functionality built-in. Multiple technologies must come together into a coherent whole. Your requirements will vary, but these could include data discovery and classification technologies, process automation software, data access, security and protection tools, plus of course fast search and access solutions.

People and process considerations

A complete package of storage platform plus software tools is still not enough: there are also important people and process issues to consider. For example, simply setting metrics to monitor the quality of the data held and the service quality experienced by users could well bring some internal politics into play.

Other key issues relate to data quality, usage and governance, all of which are central to providing users with confidence in the platform. These all come down to data stewardship - notably who does it and what processes need to be put in place.

In essence, if you are going to eliminate, or at least reduce, data silos, you have to have a repeatable, well understood and accepted process, that needs to be automated lest it consumes too much time and effort. This process gets data from where it is created, reconciles it with other data sources that may hold the same core information, and then places it in the Golden Copy store in a way that permits it to be easily found.

The stewardship element also extends to controlling data usage. This can be a very complex process in today's business world, as while it may be straightforward to set access controls for each individual data record, significant issues can arise under modern regulations when multiple data records are combined.

Addressing these matters requires people from every part of the business (including IT, of course) to work together to establish good processes and a method to modify things as times change. It is likely that both end users and IT Professionals will require ongoing education.

How to start

If you plan to build a Golden Copy platform, a major question is how should you go about things? Do you go 'big bang' and create a single repository and move every record in the organization or do you take a more measured approach (Figure 3)?

Big bang! *Single initiative to modernize across the board*

Draw a line *Modern platform for new apps, don't touch legacy*

Steady shift *Modern platform for new apps, move legacy over time*

Opportunistic *Just let new technology creep along as required*

Avoid/ignore *Leave things as they are for the foreseeable future*
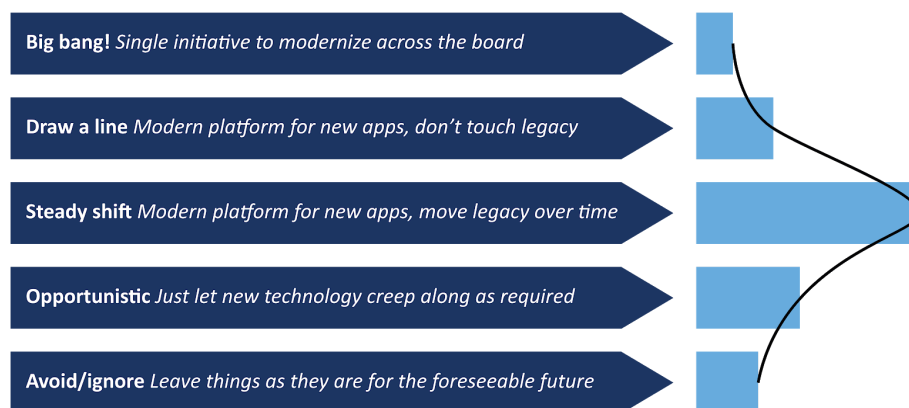
Figure 3   Typical major infrastructure change options

In practice, there are many ways to start a major IT infrastructure project, but most boil down to variants of those shown above. Depending on your situation avoiding or ignoring the problem may not be an option, but equally, it is very rare for an enterprise to attempt such a significant project as a single big transition. The most likely approach is building a new platform, and populating it at a steady pace over time.

For the new project to be effective, it must not only work, but it must be seen to be successful. Indeed, a key element of rolling out beyond the first project or two will be showing other business groups and users the benefits that using the Golden Copy store brings. In this way it can act as a magnet, encouraging users to become part of it.

This makes picking the initial use case even more important than ever. For example, some early Golden Copy successes have targeted business processes where faster access to data previously held in a variety of sources has obvious business benefits.

Clearly selecting a storage infrastructure with all of the required capabilities is step one, but to this must be added the software to allow automated data collection, reconciliation and placement in the store. It is also important that tools be used to monitor the system, and not only its physical well being.

The next step is to ensure that governance and data usage policies are put in place and, even more important, kept up to date as data sources are added or regulatory conditions change.

And lastly, don't forget that, however easy the Golden Copy system you implement is, users will need training. Not only must they learn how to use it effectively, but they also need to know not to misuse - whether accidentally or deliberately - the wide range of information it holds.

In summary

A Golden Copy data store is not a new idea. The potential benefits of such a platform are easy to understand, but until now many of the components required to build such systems have been too complex, technically immature and expensive to consider, except for very limited subsets of data. Today, the technology has matured greatly, making these capabilities a realistic proposition for a growing range of organizations.

But it can be difficult to make the business case for investment unless real benefits can be identified and delivered quickly. Even more important will be to ensure that everyone involved understands that Golden Copy storage is not a data lake. Instead, it is a platform where users can easily find the data they need, and then use it effectively without having to learn complex data handling skills.

Business needs something like Golden Copy storage because change happens so fast now. In addition, the IT ecosystem is sufficiently mature to make building an enterprise system of ultimate record a realistic proposition.

About Freeform Dynamics

Freeform Dynamics is an IT industry analyst firm. Through our research and insights, we help busy IT and business professionals get up to speed on the latest technology developments and make better-informed investment decisions.

For more information, and access to our library of free research, please visit www.freeformdynamics.com or follow us on Twitter @FreeformCentral.

About Western Digital

Western Digital creates environments for data to thrive. As a leader in data infrastructure, the company is driving the innovation needed to help customers capture, preserve, access and transform an ever-increasing diversity of data. Everywhere data lives, from advanced data centers to mobile sensors to personal devices, our industry-leading solutions deliver the possibilities of data. Western Digital data-centric solutions are comprised of the Western Digital®, G-Technology™, SanDisk® and WD® brands.

The Data Center Systems business unit is a leading provider of data center infrastructure solutions for enterprise, private-cloud, and edge data center environments. We have primary operations in 8 countries, offices throughout the Americas, Asia Pacific, Europe, and the Middle East, and power the data center storage infrastructure of many Global 1000 companies.

Our solutions are used by thousands of customers in the capture, preservation, mobilization, and transformation of business-critical Big Data and Fast Data sets generated by a wide range of enterprise, business, industrial, and cloud applications.

Learn more at www.westerndigital.com or follow us on Twitter @westerndigidc