Essentials about data archiving

White paper

Businesses are facing with explosive data grow, stringent compliance regulations and rising security risks. Intelligent data archiving as an integral part of a modern data-driven enterprise, can help solving these challenges.

Content

Executive summary

Digital transformation results in a huge amount of data, which becomes more and more valuable to businesses. This explosive data growth, along with stringent compliance requirements, rising security risks, and privacy and compliance concerns, is among the biggest challenges organizations face. In addition, there is an increasing need to protect, retain, access, and analyze this business-critical and valuable data. Therefore, businesses have to find the right balance and interaction between productive, backup and archive storage, in order to maximize their operational and business efficiency and agility in the hybrid IT infrastructure. Data archives are therefore becoming more and more an integral part of any modern, multitier data protection strategy, both to optimize cost and to mitigate the risk of data loss.

Trends and challenges for a data-driven engterprise

Data explosion

Internet of things, artificial intelligence, research and development, online sales, autonomous driving, healthcare data, video games, surveillance and streaming are only a short-cross section of the producers and evaluators of digital data. This digitalization, plus tight compliance requirements and service level agreements, results in an exponential data growth of 50 to 70 percent a year that data-driven enterprises have to deal with. Besides the storage of business-critical data, the tendency is to store more and more unstructured (sensor-) raw data from the edge for later data analysis. The data explosion requires organizations to manage their data across databases, virtual environments, and unstructured records from the edge to the core to the cloud. Analyst firms estimate that more than 60% of the stored data is inactive, meaning it has not been used for more than 30 days. Outsourcing this ‘cold’ data to an archive can help reduce the cost of data storage.

Compliance

Legal and security concerns have become more and more prominent. Businesses of all sizes are required to retain business data such as invoices, emails, customer and personnel data for a long period, due to regulatory compliance. The most important compliance regulations include the sarbanes-oxley act (SOX), health insurance portability and accountability act (HIPAA) and general data protection regulation (GDPR). Depending on the industry, there are different legal requirements for the retention of business data. Some companies keep data for decades, others for years. The dynamic regulatory environment requires constant attention from compliance, legal and IT departments to adapt effectively existing processes and technologies to evolving regulations, information security and privacy laws. Organizations must manage their responses to data privacy and protection legislation to avoid penalties for violating compliance, which include payments for damages, fines and voided contracts. Against this backdrop, the compliance topic is not only one for highly specialized, process-driven compliance managers, but is also a company-wide requirement to minimize the risk and cost. In addition, the use of intelligent data management

and analytic tools helps the responsible people to meet audit and compliance requirements. For each organization, a well-designed information and data protection policy is recommended for both the production data and the inactive data in the archive.

Cyberattacks

As data becomes more valuable to organizations, it also becomes an increasingly attractive target for theft or malware attack. The latest threat is ransomware, where critical data is encrypted until a ransom is paid in Bitcoin in exchange for the unlock key. Attacks of this nature are increasing, and infections can take hold in seconds. Organizations that are locked out of their data face a trail of damage that can take weeks or months to rectify. Some data protection specialists recommend extending the well-known data protection rule '3-2-1' with an additional offline copy. This new 'three-two-one-one' rule effectively mitigates the threat of data loss:

- Keep at least three copies of your data – the primary data and two copies – to avoid losing data to a faulty backup
- Store two copies on different types of storage media such as tape, disk, secondary storage, or cloud
- Keep one copy on a remote location in the event of system failure or local disasters
- Keep one copy offline on immutable storage spaces, in the event of infections within the network Especially for data critical to business operations, keeping an encrypted copy offline is a guaranteed way of keeping information safe from online attacks.

Small, highly portable tape cartridges are a convenient way to store data offline in a specific vaulting space within the tape library or physically in a safe or fire shelter. WORM (write-once-read-many) media or SoftWORM technology is also an alternative. Besides that, modern data protection software also provides immutable storage spaces on disk storage. These approaches provide a new layer of protection against ransomware and other types of online cyberattack. Tape storage helps to maintain the integrity and accessibility of business-critical data even in cases of cybercrime, such as ransomware attacks.

Therefore, more organizations include ransomware security technology in their modern hybrid infrastructure.

Defenition of data archiving

Retrospect

In the past, archives mainly existed to preserve paper documents for various purposes. Such paper archives required a lot of space, were difficult to manage, needs protection against fire and water damage, and were very costly. In addition, to find the required document or information could be a time-consuming task. Nowadays an increasing number of business processes are available in electronic form. This means you can use digital search functions, making it easier to find specific information. It also lowers costs throughout the information's archive lifecycle and permanently speeds up administration. Accordingly, electronic documents are increasingly replacing expensive-to-handle paper documents. But nevertheless, electronic documents must have the same permanent value as documentary evidence as would have been the case with a paper version. Digital archives must be trustworthy in order to legally safeguard business processes. When audits, assessments or legal issues arise, data must be provided quickly and completely. E-discovery and legal hold are of high importance in this context. What's more, it must be possible to prove the integrity and authenticity of the data at all times — in some cases for a period of more than 100 years.

General

Digital archives are not copies or backups of primary production data, but contain the original data from mostly non-frequently-used, inactive or cold primary information. This is older information that remains important for the business in the future or must be retained to fulfill regulatory compliance. Lower-cost secondary storage tiers are ideally suited for data archives. One important aspect of a modern archiving strategy is data insights, and the analysis of business data to inventory the data and identify which data should be archived or not. The archiving process moves the original data from costly online storage to low-cost secondary storage for long-term retention. Data archiving therefore reduces primary storage requirements, and allows businesses to maintain and monetize data that may be required for regulatory or later requirements and analysis. The role of the archived data is not to recover an application or business data, instead the main usage is for long-term retention and information retrieval – usually at the level of a file, e-mail, old database records or other individual piece of content for operational or regulatory requirements. In addition, archives are also suited as target for offloading inactive data that will otherwise participate in the daily backup stream and add an unnecessary burden to the overall backup process. Modern archiving solutions manage access to the archive. Depending on the critical information, some archiving systems enable writes and reads, others only reads. WORM (write once, read many) technology, realized in hardware media or as a software solution, protects the information from modification and thwart ransomware attacks. Intelligent archiving software automates the archiving process and moves "cold" data depending on the defined storage policies and retention period to low-cost tiers.

Difference between backup and archive

A backup is not an archive. Data archives are often confused with data backups, which are copies of data. Although both backup and archiving use higher-capacity storage media with lower performance and cost than primary storage, they serve different purposes. Archives act as a data repository for long-term data retention and contain data that is infrequently accessed, but still readily available. On the other hand, backups are secondary copies of active production data and apply to data protection and disaster recovery. Data backup helps businesses

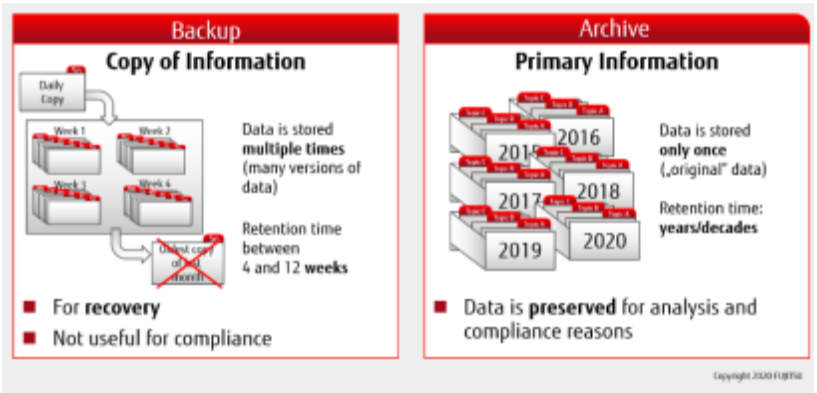
to reduce downtime in case of a disaster or system failure, and is metered as recovery time objective (RTO) and recovery point objective (RPO). Backups generally have a retention period of up to 30 or 60 days, depending on user data. A short recovery time, in case of data loss or corruption, is an important element of modern backup strategy.

More and more backup software suites also provide archiving capabilities, but be aware that they focus more on the copying of data (=backup) than the moving of data (=archive).

Backup versus Archive

Do you need both backup and archive? The answer is 'Yes, you need both'. Backup and archiving systems meet different requirements. Implementing both in accordance with a sophisticated information policy will provide governance and compliance benefits. In addition, for the highest SLA it is essential to have a backup of your archive in place as well!

- Backup is generally for restoring data after a data loss up to the level of a complete disaster recovery
- Archive is for long-term data retention, for compliance and security reasons or cost-effectively storing inactive data



Benefits of data archiving

Cost reduction

Data archives reduce the overall storage cost. Archiving moves data from primary storage to a lower storage tier, and frees up space on more expensive primary storage hardware. Primary storage arrays, no matter if hybrid or flash arrays, are typically more expensive than archiving hardware because user read/write activities and production applications require a high level of IOPS to meet operational demands. Data archiving solutions store the data on lower-cost media such as nearline SAS drives, tape or optical storage, or they move data into the cloud. The performance of such storage resources is generally slower, but they provide higher capacity and therefore the price per GB is much lower. In addition, offloading inactive data from more expensive primary and backup disk systems into an archive solution reduces the volume of active data. Data archiving therefore mitigates data growth and saves money for capacity upgrades of primary and backup storage.

Backup volume reduction

This offloading of old and inactive data also removes primary storage data that will otherwise participate in the daily backup stream and cause an unnecessary burden for the overall backup process. Archiving can remove tens of terabytes or more from the backup set. This reduces primary storage costs, as well as backup hardware and software costs. The amount of backup data tremendously shrinks, and results in faster backup and recovery processes. A reduced backup volume decreases the license cost of backup software, which is often metered in front end terabyte capacity. In addition, licensing cost of archiving capacity are usually lower than for backup.

Compliance

Governmental requirements and legal liability are key reasons to implement a data archiving strategy. Data archiving helps businesses meet compliance regulations by storing information long-term and immutable. Integrated WORM hardware and software capabilities guarantee a non-erasable, non-rewritable format of archived data to fulfill the requirements for compliance archiving. Besides data resiliency, archiving consolidates data for easy access in case of audit or inspections.

Risk mitigation

Today organizations have to be mindful of risk mitigation and avoid data loss or disruption. The most damaging risk for organizations are cyberattacks. The leading form is here ransomware and requires a strong disaster recovery strategy in place. Geographical separation and disconnection of networking (“air-gaps”) can mitigate the encryption or destruction of data by attacks like ransomware. Data archiving can provide both ways of separation. The use of WORM technology (write once, read many) for archiving data protects from ransomware attacks by making the data immutable. Administrators can set specific retention periods depending on business requirements. The read-only status thwart unauthorized access and data modification, and fulfills strengthened compliance regulations. Besides WORM, archives offer a so-called air-gap between live and archived data keeping the data securely offline until it is needed. Tape storage is optimally suited to keeping data offline and storing it in remote data centers or vaults. Remote locations increase data availability of business-critical information and mitigate the risk of data loss after fire, floods or other disasters. Archiving software features intelligent encryption routines and specific capabilities against ransomware to thwart criminal access to business data. Archive solutions create an additional level of security for data-driven businesses.

Use cases and best practices The data archive mainly contains unstructured data such as documents, videos, audio files, .pdfs, emails and

Use cases and best practice

more. Unstructured data is usually chaotic. Enterprises can have petabytes of data, which is not necessarily organized, easy accessible, retrievable and intelligently retained in a certain and simple way. In the past, an archive was created due to compliance regulations, and everybody hoped they would rarely if ever need to restore or search within the archived data. Unfortunately, this has dramatically changed.

Unstructured, archived data has now become the lifeblood of data-driven organizations. Nowadays an archive is not only used for compliance or company history, but more and more for data analysis, data mining and artificial intelligence. Read the analyst whitepaper from Freeform Dynamics about 'The importance of modern data archiving' to get more insights.

As digital transformation accelerates, the explosion in data has created many challenges for data-driven enterprises. Increasing

information governance a strategic imperative for business. Archiving of some data is no longer an obligation; the archiving of data is ready to

Conclusion

compliance regulations, ransomware threats, and data complexity have made data archiving and

become a fundamental enterprise resource and business enabler for data-driven enterprises.

Essentials about data archiving

White paper

For more information:

www-fujitsu.com/dataprotection

Contact
Fujitsu Technology Solutions GmbH
Mies-van-der-Rohe-Str. 8
80807 Munich
www-fujitsu.com/dataprotection