# Rocket

# 5 Questions to Answer Before Investing in Archiving

It's no news flash that data storage will continue to grow at a faster rate than IT budgets. A study published December 2012 by IDC projects installed raw storage to exceed 7 million petabytes by 2017. And it appears they were pessimists. To compound this data growth challenge, regulatory requirements keep extending the length of time data (that is material to the business or potentially subject to any legal or regulatory discovery) be kept available. Currently, it must be kept for at least ten years and maybe much longer in some industries. The bottom line is that IT organizations need to strategically plan and implement solutions to properly manage the ever increasing mountain of data that needs to be stored and retained.
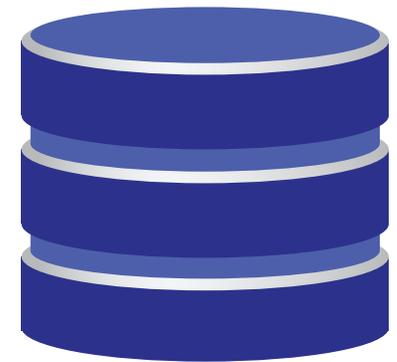
The most powerful strategy to cost effectively manage the retention and disposition of data is to implement a data archiving system. Archiving keeps primary storage clear of "cold" data (not accessed for 6 months or more), improves backup performance, makes restores quicker when disaster strikes, and properly retains data for future reuse or compliance and regulatory needs. When the time comes to access the data, a good archive solution can provide immediate, granular access to data for easy retrieval.

On the operational side, it's fiscally irresponsible not to have a proper archiving implementation in every IT department because not only does it provide these infrastructure benefits, but it lowers the overall IT cost of storing and protecting data. Archiving data also provides the future flexibility to expand storage at the time and place of the company's choosing rather than on past purchase patterns.

## Planning Your Storage Archive

An archival storage system is straightforward to implement and there are many ways to begin that don't require years of committee and policy setting meetings to establish what to do with each data type and how long it is to be retained. It is easy to get bogged down in discussions on data type and how important each is relative to the other and how it varies by department.  Grab some data and move it. No end user will know and you'll start getting benefits immediately. If the task seems too daunting to do manually or with scripts, there is software available that can identify and monitor data activity and characteristics for the purposes of selecting what to archive.

As you prepare to archive data, there are some important questions to ask yourself to make sure that the implementation goes smoothly and stands up to the growth needs of the IT infrastructure and the business users.

## Archiving Preparation Questions

1.  **What type of data are you storing - cold, compliance, or both?** Most of the data kept in an archive is cold data that is rarely accessed by users. These implementations are designed to keep a balance between

preserving data at the lowest cost and having the most actively used and important data on the best performing (and likely the highest cost) storage. The choice of storage is primarily based on cost and performance.

Archiving data for some organizations is a business requirement for regulatory compliance or potential legal discovery. Still others are seeing data as a potential gold mine of information that can be researched for profit as analytical tools are developed in the future. These uses require an object storage repository that can ensure data integrity and provide information on legal custody and who has access to the data.

Understanding the possible future use and need for different data types will also be useful for the creation of the policies for the removal of data from primary storage, where it is to be retained, and how long it is to be retained.

2.  **What type of storage technology?** Storage is a combination of hardware and medium (tape, hard disk (HDD), solid state (SSD), or combination). Most people are familiar with the failure rates of these devices, the need to update firmware and software, and the need to change interfaces as they are upgraded over time.

    The medium also has a useful life. The longest useful life is likely to remain tape for a long time. LTO tape that is used monthly will last up to 17 years. With infrequent use, then removed and properly stored, it has the potential to last 100 years.

    HDD, separate from manufacturing defects and random errors, has shown it can run flawlessly for three years. After that point the various mechanical components that spin and actuate begin to break down and by 6 years 50% of the drives have failed. Powered down drives don't necessarily last longer. The disk head will land on the platter when spun down and stick to the disk surface if it rests too long. When the drive spins up again and the head has stuck to the surface, the media will be damaged and data lost.

SSD, while they have no mechanical parts, still store data electronically and those cells "burn out" and degrade in capacity. All SSD vendors have a strategy of over provisioning and relocating data to good cells so there is no effect on performance. Longevity, however, is controversial and not well understood yet in terms of long term retention. Vendors currently warrant SSDs for three years with some studies proclaiming they will last longer than HDDs.

The bottom line is that ALL these mediums wear out with use. There must be a plan to periodically test the data on the chosen medium to ensure its continued readability and to migrate the data to newer media as the hardware, software, and interfaces change over time or the media itself degrades.

3.  **What is the data format?** Long term usability of data not only requires that all the "bits" are accurately preserved and that there is hardware that can read the data from the storage medium but that there is also an application that can interpret the data in its file, database, or object format. Many technologists are surprised by how quickly these formats change and are unaware of the vendors' responsibility for backward compatibility. Typically vendors will maintain at least two major revisions of backward compatibility, but that may only last two to five years. It is mandatory to have a process that not only checks the readability of the data, but also the usability by the application responsible for displaying the data.

4.  **Is the archive in the Cloud or Data Center?** Data Center storage and systems cost should be well understood by the IT department. Technology is now to the point where it is feasible to consider cloud as a long term archive. When considering cloud, there are some compelling positives and some challenging negatives to consider.

On the positive side, it is pay as you grow convenience turns capital expense into operational expense which may be an important dynamic for some companies. Another advantage is that the cloud service provider will

have a strategy to constantly balance the type, age, and interface of the storage device so you can be assured that the data is always available and readable. You are still responsible to ensure the data is application readable, but everything else should be properly handled by the Cloud storage provider and that assurance should be part of your selection due diligence.

The biggest downside is that as data grows and time passes, the cost of your cloud-based archive will likely grow.

When considering cloud storage, it is also important to understand the time and cost of moving data BACK to the data center. Trickling data to the cloud is simple and cost effective. When data is needed quickly for a disaster recovery operation, system restore, or legal compliance, the bandwidth of the link between the data center and cloud can be too slow for the amount of data that has to be moved and the timeframe it is needed. Considering a typical WAN speed of 100Mbit/sec, theoretically 1TB per day can be transferred between sites. Multiple links, higher speed connections, and dedicated lines can increase this number – at a cost. Also, remember mileage may vary due to overhead that can slow down throughput (other processes/users sharing the connection, hops, latency, line quality, sun spots…). Likely in an emergency, the data center will need more than 1TB quickly. Even with no interference, 100TB of data can take over 3 months to move back to the data center without a dedicated, high speed link or physically taking the data back to the data center.

5.  **What do you know about your data?** Most IT departments can tell you how much storage capacity they have, but few can tell you how much data they have by application, owner, and age. Tools are available to scan the environment with no additional overhead or hardware purchases. This scanning software can identify key metadata parameters and produce reports to help plan the archive system. These are the critical metrics for recalling archived data and therefore, also make the best guide for what data to send to the archive.

Understanding what is happening in the IT department at the data level is one of the greatest benefits for IT management and administrators of developing an archiving system. They can then discuss with business managers the departments, people, and applications that are consuming the most data, what the growth rate is, and when they are likely to run out of capacity. They can show them the cost of their current data retention policies and work with them to reduce cost or expand services.

Managers can have weekly, monthly, and quarterly reports and charts show up in their email or made immediately available to them by logging into a reporting system. These reports can provide the information necessary to create an ROI analysis for new system purchases and the detail required to smoothly implement a chargeback system. Most importantly, managers have a constant health check on how their data environment is growing and being managed and the confidence that they can respond quickly and appropriately to challenges as they inevitably arise.

## Archiving is Essential

Archiving is the essential data management service that protects the high performance, vitality, and health of the IT environment and the longevity of the data. The truth is no IT department is going to fall apart tomorrow if they don't archive today or this week. It's the sustained non-performance that leads to extensive data atrophy, sluggish systems, and extraordinary costs to upgrade and migrate to new systems or respond to legal and regulatory requests. Archiving protects against these challenges and is straightforward to begin implementing by answering the questions above.

## Archive and Backup Management

Rocket Software has extensive software capabilities and talented engineers with hundreds of years of collective experience in Archive and Backup Management. Software, data assessment services, and consulting are available to help support implementing or augmenting your active archiving system. We can work with you to understand your environment, what's working, where the holes may be, and build fact based proposals that save money and have relevant ROI based on your operation – not generic numbers. You can find out more by calling us at +1-650-237-6246 or visiting the website http://www.rocketsoftware.com/solutions/archive-and-backup-management.

# Rocket

www.rocketsoftware.com

info@rocketsoftware.com