Top 10 Archiving Qualification Questions

1. How much data is being stored in your network and what percentage of the total is archive data?

Quantifying the volume of data in a network is an important place to begin. It seems a simple question, but many system administrators do not have a good understanding of exactly how much data they are trying to manage. Capturing the total data volume requires determining the actual disk utilization minus duplication, orphaned data and system objects.

The next step is to identify what percentage of the total data volume is active data and what percentage is inactive or archive data. Active data is defined as data currently being created, modified or frequently accessed. Inactive data can be defined as data that is no longer being modified and is less frequently accessed. Some inactive data may have very little value to an organization, but other inactive data is valuable and must be retained for long-term availability. Valuable inactive data can be classified as archive data and requires an archival storage strategy if it is to be reliably stored for long periods of time.

Archive data typically has very different storage priorities and requirements and by identifying this data and separating it from active data, a unique strategy can be defined to ensure that the archive data is retained for as long as required and its authenticity is protected.

Administrators are often surprised at how much of their online storage is consumed by archive data. Separating archive data from active data can have major benefits in administration and cost reduction. For example, since archive data is no longer being modified, it is not necessary for it to be part of a normal backup cycle. Archive data can be placed outside the standard backup schedule so that backup operations can focus on protecting the active data that is changing on a frequent basis. Moving archive data to non-volatile storage media dramatically reduces the average backup window, administrative overhead and the total cost of the backup process.

2. What is the growth of your archive data on a weekly or monthly basis?

The long-term nature of an archive means that you have to design for current requirements while planning for future growth. Having said this, archive growth calculations can be difficult. It may not be adequate to base your growth estimate on current data types since history has demonstrated that the archive net is being cast ever wider to include new data types not previously considered. Estimates need to include the growth rate of currently identified archive data, as well as, the potential addition of new records that may need to be managed in the future.

Constantly growing volumes of archive records demand solutions with high initial capacity and the flexibility to scale over time. The QStar storage management software has been designed specifically to accommodate archive growth with an open ended file system. This strategy provides incremental archive capacity while protecting your initial hardware investment. Using removable media is another important advantage to growing archives. Rather than purchasing new or expanding an existing RAID system, organizations have the option of taking older data sets off-line to make room for newer data. This is simply not an option with magnetic disk based archives without purchasing additional hardware to transfer the data. Whether expanding an existing library or taking data off-line, both techniques are non-disruptive and cost-effective strategies to incrementally expand archive capacity.

3. Is your organization subject to external industry regulations or internal corporate policies on record retention and authenticity?

Another way of clearly identifying archive data is by regulations and policies that define the processes and retention periods for specific data types. More and more organizations are required to adhere to government and industry regulations on data storage for legal or public safety purposes. In addition to external regulations, organizations often have their own data retention policies to ensure availability of corporate knowledge and to protect against potentially damaging litigation. Understanding the role that regulations and policies play within your environment is essential to establishing a defendable data archive.
4. What is the profile of your archive data (i.e. its format, size, and retention period and access requirements)?

Once the volume and scope of your archive data has been determined, the next step is to sub-divide the archive by establishing profiles for each primary data type. A profile is a set of attributes that characterize a document format, size, retention period, and access requirements.

Archives being used for a single task may only require two or three profiles since the document types may be very limited. By contrast, archives being used by many applications will require additional profiles to reflect the wider range of document attributes. The objective of profiling is to add structure to previously unstructured data. Well-defined profiles allow the administrator to properly manage both the physical and logical storage of the archive data.

For example, healthcare institutions handle both patient records, as well as, high volume digital images generated by modern medical equipment such as x-rays, MR or CT scans. Most of this data must be retained for the lifetime of the patient. Fast data retrieval is a high priority in order to ensure that medical staff has ready access to patient information. Data integrity is of foremost importance as it is an essential criterion in maintaining a secure and unadulterated Healthcare Enterprise Archive.

Engineering and manufacturing companies are also dealing with large and complex data, such as 3-D models, simulations, mock-ups and other technical documents, but the retention period and access requirements attributes will be different from those of medical records.

One advantage of data classification is the ability to physically group similar profiles on the same piece of archive media. By doing this, older or less frequently accessed records could be removed from a library making room for newer data while still retaining the media for potential access. For example, an engineering company can choose to archive all design, technical and administration records for one or more related projects on a single piece of media, allowing projects to be taken off-line for future reference or distribution to other site locations.

QStar archive solutions are designed to track and manage off-line as well as on-line storage, providing flexible options for companies to meet demanding data profile requirements.

5. How quickly do archive records need to be accessed to meet your business requirements?

Since data written to most archives is part of a background process, the ingestion rate of records into an archive is often not a critical path. By contrast, timely retrieval of archived records is business critical. While it is true that most archive data does not require the same access performance as active data, just how quickly do your archive records need to be available? Is a 5-minute retrieval time adequate or do you need to have access in 10 seconds? The answer to this question will be closely tied to your business requirements and the discovery period mandated by regulations. It is also a critical factor in choosing the proper hardware. Failure to meet access time service level agreements can have a significant impact on the overall success of products or services that are dependent on your archive.
QStar archive solutions provide random access performance even to magnetic tape archives, with our sophisticated built in cache management. All directory information is stored in the cache as well as all recently touched files, returning them at hard disk speed even if they have been written to the archive media, saving organizations time. Once the archive threshold has been hit for a file in cache, it is deleted from the cache automatically, but can be brought back from the archive media if needed.

6. What is the maximum number of concurrent users likely to be accessing your data archive at any given point in time?

Timely retrieval of archived records is business critical and an important factor that will impact on data access time is the number of concurrent users accessing the archive. The ability to predict access patterns and user levels is important to properly configure the archive hardware. An archive with insufficient “horse power” may cause request queuing, impacting the overall performance of the archive and consequently the business operations.

To accommodate a range of concurrent access requirements, rotational media libraries such as optical will offer quick exchange times and fast seek operations for faster retrieval of single files. By contrast, tape drives have higher exchange, seek and rewind times. Understanding the access demands on the archive allows the library to be properly configured to meet demanding business requirements from simultaneous users without expensive hardware overkill.

7. What is the disaster recovery strategy for your archive?

By definition, archive data has intrinsic value to a company. Lost data could cost an organization a huge amount of time and money to recreate from scratch and some records, such as legal or historic documents, can simply not be replaced. Even the most secure storage media cannot protect your data from fire, flood or sabotage. If your archive records are valuable to your business you cannot afford to have a single copy. For this reason, it is important to implement a disaster prevention strategy that provides for a secondary copy of your archive in another geographic location. A disaster prevention strategy adds cost to the overall solution; however, what is the cost of losing your archive?

No media can protect data from site disasters like flood and fire entirely, so like all other media there should be at least one secure secondary copy of your valuable archive data. There are two typical disaster recovery strategies: mirrored systems and offline vaulted media and QStar can support both. QStar provides a disk/tape copy utility so that disks and tapes may be copied and moved outside the library for offline vaulted security. This is a cost effective option that magnetic disk archives cannot support without additional hardware and administrative overhead. Where fast access to disaster prevention is required, QStar provides a mirror strategy with QStar Data Director. Data Director can manage archive libraries in two separate independent locations to provide virtually uninterrupted failover. In addition the data is protected within the network with all network security intact.

8. What is your plan for maintaining older media and migrating data to newer storage technologies over the life of the information?

Applying short-term IT solutions to a long-term archive can prove to be inefficient, costly and dangerous to the longevity of the data. The time periods involved in an archive pose unique considerations that are not always part of the typical IT infrastructure. Since archive data may need to be available for many decades, you need to plan for the maintenance of your archive environment and for the potential migration of data from one storage architecture to another. The choice of storage technology, both software and hardware, will have a major impact on maintenance overhead and migration frequency. Deploying products that have not been designed with a long-term support strategy will result in higher costs, increased maintenance and they may compromise the authenticity of the archive data. Choosing products that conform with industry-standards is an important consideration, since international standards provide an assurance for quality and interoperability.
9. Who within your organization is responsible for ensuring that archival storage compliance is being properly addressed?

For organizations that place a high priority on data compliance, there is a growing trend to define a role that is responsible for the compliance process. Often referred to as the “Compliance Officer”, this person takes responsibility for researching compliance requirements and for the design and implementation of both procedures and technology to meet compliance obligations.

For smaller organizations this role may be a part time responsibility, but it still makes good sense to establish the position since it can serve more than one purpose. First and foremost, it will help to ensure that the organization is meeting all necessary compliance regulations. Secondarily, it demonstrates corporate commitment to legal and regulatory bodies. Thirdly, it better prepares an organization to respond professionally in the event of litigation. Those organizations with a Compliance Officer are in a stronger position to defend the integrity of their audit trails and data, avoiding the potential of costly penalties and jail sentences.

The more tools a Compliance Officer has at his disposal, the better able an organization is to demonstrate data integrity. Clearly documented audit trails are fundamental and while this is largely a question of procedure, Write-once media uses an irreversible recording technique so that data can never be changed or altered, which facilitates audit trail management. In addition, QStar provides a unique, unalterable, serial number on each media formatted. Application can include this serial number as part of the audit trail data, recording when and on which piece of media records have been written. This extra level of authentication prevents someone from writing modified records to a new piece of media in an attempt to tamper with archived data.

10. How do you calculate the Total Cost of Ownership (TCO) for creating and maintaining a long-term data archive?

Cost is always a key factor in choosing any IT solution, but here again the criteria by which you judge the cost of a data archive is different than that for active data storage. Since an archive needs to operate for many years, you cannot simply look at the initial acquisition cost and because an archive consists of many components you cannot compare it to the raw storage cost of magnetic disk, tape and optical. To fairly compare different technology options, you must look at all the pieces necessary to meet the longevity and authenticity requirements demanded by an archive environment.

When estimating the TCO of a particular configuration or when using TCO to compare different technology options it is important to factor in as many of these costs as possible and to do so in an even handed way.

QStar has been designed specifically to manage archival storage. Scalability, longevity and authenticity are inherent traits of the QStar solution and thus QStar supports a wide range of archive hardware and front end applications. In addition, QStar's long-term support commitment to the technology is in keeping with the needs of organizations that must maintain their archive for many years in the future.

These key design benefits have a direct impact on the TCO of archive management. When factoring in all the appropriate costs over the years of operation, QStar mitigating many of the costs over the long term operation of a digital archive.