TECHNOLOGY
PRIMER

Storage Technology Evolution and Alignment

A primer on implementing efficient multi-tiered architectures

Jon Collins and Tony Lock

August 2010

Executive summary

Few organisations today would deny their absolute dependency on both their computer systems and on the information they store. Rapid data growth continues for many businesses, and despite the current economic climate storage demands do not seem to be letting up. Trends such as increased user-created content coupled with mounting pressures from regulatory compliance are causing us to store more information over longer periods of time.

Against this backdrop, storage technologies continue to evolve and diversify to keep pace with data volumes. When it comes to managing live data, for example, developments in areas such as denser, faster and more power-efficient disks, along with the emergence of solid-state disks, are providing more options from a hardware perspective. Added to this we have more flexible management and software capability to allow storage virtualisation and the minimisation of data quantities through deduplication. Meanwhile, much is being made of disk-based backup, but tape is still recognised to hold a valid place in today's IT environments.

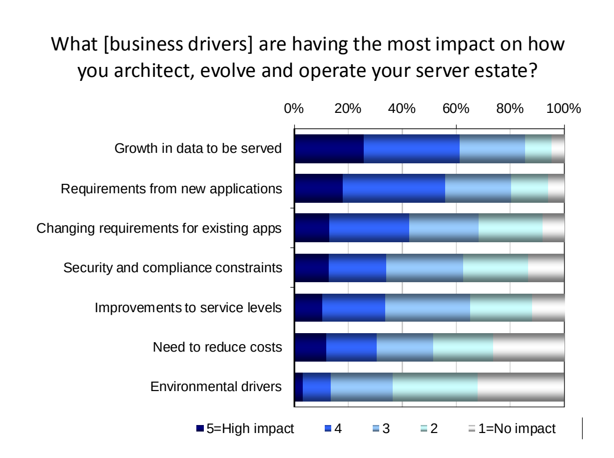
But this is not just about technology of course, and it is particularly not about believing the hype around any one particular solution. There are no magic bullets in storage and the reality is that a blend of technologies will be required in most business environments based on the so-called 'tiered storage' approach. With this in mind, it is important to ensure that a storage strategy is in place that reflects the requirements of the business. From there, balanced decisions can be made considering the value, cost and risk of any particular solution in a given context.

In order to help with this, the remainder of this paper considers the merits of the different technologies available on the market today, and how they might fit into a tiered storage architecture.

This paper was compiled and written on an independent basis by Freeform Dynamics, based on existing research gathered during 2008-2010. The paper was sponsored by Tandberg Data.

1. The unstoppable growth of data

To say 'data quantities are growing' would be an understatement. Even in the current financial climate the increase would appear to be unstoppable – in recent Freeform Dynamics research for example, we found that data growth was seen as the biggest factor when it came to the architecture, evolution and operation of today's server estates (Figure 1).



Data growth is the number one factor in data centre / computer room evolution[1]

Figure 1

Organisations are now grappling with the need to keep larger and larger volumes of not just business-critical information, but other less critical data as well, for longer and longer periods of time. Why is this? The reasons are many and varied, though the growth is largely centred around:

1. Sound business requirements to hold more detailed information on products, services, customers, suppliers and associated transactions that will inform business strategy, drive growth and increase service levels.
2. User-driven growth in unstructured content, through the increasing use of email, electronic documents, images, video and audio alongside more traditional structured content.
3. The need to mitigate risk influenced by both internal policies and externally imposed regulations, which necessitates the storage of various types of data for differing, though almost universally increasing periods of time.

The upshot of this is typically an ongoing need to purchase more storage devices and media to keep up with growing data volumes. As storage devices and technologies have then proliferated, IT departments have been faced with the associated cost and overhead of managing storage infrastructures that are ever increasing in complexity, with frequent disjoints and data that is often dispersed and even duplicated across the organisation. Added to this there is the challenge of providing protection against the very real threats of loss, theft or sabotage, as well as the need to deal with growing issues around power consumption, cooling and physical space availability.

Of course some IT departments have been able to manage things better than others, but if you are in the position of looking around you at a fragmented storage landscape across which you know certain types of data are probably sitting on the wrong type of storage for the wrong length of time, then rest assured you are not alone. As individual data requirements have been dealt with over the years, it is not uncommon for such a situation to arise, and if you recognise the issues and challenges we are talking about, then read on.

To begin with, let's look at a principle that is fundamental to bringing more order to the situation.

2. Not all data needs to fly first class

The value of data stored by an organisation varies considerably and often changes over time. One of the commonly heard rules of thumb, for example, is that up to 90% of data that is more than 90 days old is rarely or never accessed. And in some situations we find even more extreme cases, such as a recent study by the University of California at Santa Cruz [2] which showed that 90% of data stored to NAS was never accessed again, and another 6.5% was only accessed one more time.

With this in mind, while keeping all data at hand on high speed disks might seem ideal, in reality to do so represents unnecessary cost in terms of hardware, power, cooling and physical space. The key question here is why fly all data 'first class' when for quite a bit of it 'coach' will do? But in all too many cases, organisations are using premium storage to hold low value or infrequently accessed data. And, of course, we also see the converse – high performance systems dependent on mission critical data that resides on low spec commodity storage because that was what was available when the application was first provisioned.

The first step in reviewing or building a storage strategy is to consider the value of the various types of data that exist. It is beyond the scope of this paper to go into detail here, but the kind of techniques often used to determine storage requirements for specific data or artefacts include time based retention models, value based classification and information lifecycle management (ILM). The latter, for example, is based on the notion that a document or record may be very active at the beginning of its lifetime, e.g. when a purchase order is still open, but often reaches a point where it becomes inactive, e.g. when a purchase order is fulfilled, matched and payment has been made.

In most situations a mix of such approaches will be required, but whichever methods are used, this is secondary to the basic requirement of being able to distinguish between, for example, critical data associated with high performance applications that needs to be on fast resilient storage, versus static data that is rarely or never accessed which can be held on cheaper less resilient commodity disk, or even offline on tape. In reality, of course, we are dealing with a spectrum of requirements, and depending on your environment, it will probably be relevant to think in terms of different degrees of performance and resilience in between the two extremes.

3. Introducing the concept of tiered storage

The above way of thinking about your requirements paves the way to the implementation of a tiered storage strategy or architecture. Also known as hierarchical storage management (HSM), this kind of solution helps to balance the value delivered to the organisation with the cost of service delivery. At its most basic, the approach is based on defining different classes of storage based on the characteristics of the devices being used and then allocating data to those devices in a way that fits in with business needs and data retention policies.

Two ways of classifying storage have already been mentioned – performance, which defines throughput and access times, and resilience or fault tolerance, which clearly has a bearing on systems and information availability. With this in mind, it is common to define so-called 'primary' storage, for example, which is fast and resilient, but is also obviously going to be very costly. The question is then which data needs to be held on such primary storage, bearing in mind that you are simply throwing money away if it's used for static or low value data.

Beyond this, it is then common to define 'secondary' and 'tertiary' storage, referring to low cost commodity disk and offline storage respectively. Having said this, there is no reason why more tiers shouldn't be incorporated into the approach it if makes sense to define requirements in a more granular way, e.g. with the advent of solid state disks and semi-permanent caches on devices, storage experts now often refer to Tier Zero, representing the highest possible performance.

Beyond this type of initial mapping, an important part of the tiered storage idea is the migration of data between tiers as it ages or moves from one state to another. More of this later.

4. Online versus offline in more detail

Picking up on the question of online versus offline, such a decision is applicable in a number of scenarios, but before we look at this we need to introduce two concepts:

- Recovery Time Objective (RTO) – how quickly do you need to get this type of data back?
- Recovery Point Objective (RPO) – how recent must the data be in order to minimise impact to your business – minutes, hours or a few days?

As well as RPO and RTO, the third criterion is of course budget. This needs to include both initial purchase cost and other factors that add to the total cost of ownership of both the equipment, and the information it stores. With these factors in mind, the most commonly occurring scenarios to consider are:

Archiving of under-utilised content – Email and older versions of files can clog storage devices; removing them and saving them to offline storage can both free up valuable "productive" storage space and reduce the costs of the overall storage environment. This includes files that haven't been accessed for a period of time, are not required for compliance purposes, but may be required for future reference. Archiving this type of data will become more and more important as the general growth in user-generated content continues.

Data retention for compliance and e-discovery - separate from archival is the need to retain information for compliance and business governance reasons. New legislation requires mobile phone records of all customers to be held for months at a time in the telecom industry, for example, and existing legislation such as HIPPA in financial services or Sarbanes Oxley across the corporate sector in general requires data to be stored over many years. In the construction industry there is then the extreme requirement to hold information such as building plans for decades. Storing such information can be a costly burden on IT departments, particularly for information with no immediate operational relevance.

Data backup and restore - ensuring the timely restoration of data following a user error, system failure or other occurrence. The RTO and RPO are critical decisions to determining which storage technology to choose, with additional considerations including exactly what to backup and with what level of security.

Business continuity and disaster recovery – the key question here is, in the event of a significant business failure (whether caused by accident, natural events such as fire and flood, or even sabotage), what provision needs to be in place to get the business back up and running, and how quickly does it need to be deployed? In addition to the RPO and RTO, decisions on off-site storage locations need to be made to remove data from the same threats posed to the data centre.

Understanding the value of data in scenarios such as the above, underpins the whole concept of cost-effective, tiered storage architectures. We look at the options below.

5. Storage evolution

The evolution of storage in terms of capacity, access speed, and cost has been evident since early computers employed disk drives the size of washing machines (which initially promised to hold a whole 5 megabytes of capacity), and tape reels were as large as bicycle wheels. The past decade in particular, however, has seen a number of specific developments that are worth calling out:

- **Disk technology developments** have resulted in substantial reductions in the unit price of disks. As a result, it has become possible to consider disk as a backup and archiving target, for example in the form of virtual tape libraries (VTL).
- **Tape evolution** has also continued apace, for example with the latest generations of LTO Ultrium tape able to store up to 3 TB of data on a single cartridge.
- **Solid State Disks** are still currently too expensive for anything but top-end storage requirements, but they are expected to drop in price substantially over the next few years, which will also make them a potential alternative in many disk usage scenarios.
- **Storage virtualisation**, by inserting a layer of abstraction between physical storage and the applications that make use of it, enables more efficient allocation and use of storage. In principle this flexibility also extends to high-availability environments.
- **Data Deduplication** is where repeated files, or indeed patterns within files, are indexed and removed to enable faster data transfers and reduce backup storage capacity requirements. Today, deduplication is a hot topic among IT vendors and there are signs that the end user community is now beginning to look seriously at tools delivering such functionality, particularly as the costs of the technology reduce and it becomes better integrated with other storage management products.

This last item highlights the fact that dealing with the data accumulation problem is not just about storage devices and the tools we use to configure and manage them. Indeed data deduplication has an important role to play in the whole storage optimisation game, so it is worth exploring it a little further.

6. Data deduplication in more detail

This kind of solution has an impact across all data protection options. How does it work? In a nutshell, the idea behind deduplication systems is to find patterns at either a file or block level within data sets. A single master copy of the data is kept and all users are linked to the master copy. There are two modes of operation; 'source deduplication' is carried out on the platform where the data resides, and 'target deduplication' takes place on back-up and recovery servers. The former is appropriate for branch office / remote locations as it reduces the volume of data required to be transmitted over the network.

Users of data which has been 'treated' by a deduplication solution should detect no visible change to the data they access. Any 'repeated' data appears to be available exactly as always but to the system only a single copy is held in storage with very small 'tags' linking all other copies. This may save anything from 10 percent to, in extreme cases, over 50 percent of disk storage with the benefit of reducing the need to procure more storage capacity. Furthermore, as there is physically less data to hold, the associated protection and recovery processes can work much faster.

By reducing the amount of data that needs to be stored and potentially moved onto disk or tape, data deduplication has a potentially very important role to play in data protection where, as we know, backup windows have been under considerable pressure as data volumes grow and as user's acceptance of downtime shrinks. Reducing the amount of data to be copied can reduce backup times considerably. It has the additional benefit of allowing much larger scales of primary data to be held on a single tape. Given tape capacities today however, this benefit is usually perceived to be secondary to speeding up the backup process.

Mentioning tape brings us to the interesting question of where this storage medium fits both today and moving into the future.

7. Why tape still earns a place in the data centre

As disk-based backup solutions have gained momentum, attempts have been made to promote the idea that tape is "dead", or at best a legacy solution. However, research shows a very different picture, with only a minority of organisations adopting a disk-only approach to the storage of data.

The truth is that tape technology has continued to evolve. The first LTO tape drive was launched in 2000 with a storage capacity of 200GB, the latest LTO-5 Ultrium tape drives are capable of storing up to 3.0 TB of data on a single data cartridge. In summary LTO tape storage capacities have increased by 15 times in less than 10 years. For long-term data protection goals, modern tape offers a range of capabilities including:

- **Comparable write speeds** – modern LTO drives can reach continuous data transfer rates of up to 280 MB/s, which is as fast as most 10,000 rpm hard disks. With indexing it can also be quick and easy to retrieve data from archive.

- **Management and security** – LTO-4 and LTO-5 drives provide additional security features including hardware based data encryption plus sophisticated tape management and handling capabilities for tape automation environments.

- **Very low long-term costs** – Whilst the cost metrics associated with all storage platforms are falling it is clear that tape still holds many advantages over disk-based systems. Data held on tape does not need constant or intermittent use of electricity to hold the data nor is there any need for associated cooling systems that are usually needed alongside active disk systems. Thus for the vast majority of static, long term data storage the comparative costs, factoring in acquisition of equipment and media, along with electricity and data centre floor space, show tape based solutions to be less expensive than disk based systems.

- **Offline storage** – whilst access time may be slower, you don't need any ongoing investment in power to keep data stored on tape. Secondly, tape is rugged and easily portable for off-site remote storage, providing an air-gap and taking data well away from any threat to the data centre. Finally, tape is a convenient medium to retain a direct, ‘point in time’ copy of online information, protecting against data corruption and the threat of abuse.

- **Proven shelf life** – a factor that often falls under the radar of those charged with the long term storage of data, concerns the need to periodically refresh the storage medium as it nears its “natural end of life”. It is notable that data stored on modern tape systems can last for considerably longer than disk-based alternatives – up to 30 years in the case of LTO Ultrium tape media.

All in all, we have seen no evidence from our research or any other that significant numbers of organisations are looking to dispense with tape. Perhaps the most fundamental point however, is that as with many other areas of IT it’s not really a question of ‘either-or’. The real discussion is around where and how to best use all available storage technologies as part of a comprehensive, tiered storage architecture meeting the needs of RTO, RPO, cost and other factors.

Easy to say, but how does this pan out in practice? Let’s start with the basic question of disk versus tape.

8. Disk and tape – where do they fit?

The challenge for most organisations is to establish the characteristics of the wide range of backup, recovery and archiving options that are available today and highlight where each offers benefits.

At an architectural level these can be generalised under the headings of backing up to disk (Disk to Disk or D2D), backing up directly to Tape or an automated Tape Library, then there are Virtual Tape Libraries, employing disk and virtualisation technology for backup, and finally disk-to-disk-to-tape (D2D2T), which starts out by backing up to disk, then archiving to tape. The table below gives some idea of how the offerings can be distinguished.

Platform	Characteristics	Potential use scenarios
Tape drives & Tape Libraries	Low TCO Optimal power efficiency Long life Data Portability High capacity Fast streaming for batch data Automated storage (Libraries)	• Cost conscious environments • Long-term archiving • Storing infrequently accessed data • Off-site data storage • Automated storage • Disaster Recovery
Disk to Disk (D2D)	Fast backup and file restore Ease of use Data de-duplication More frequent RPO	• Short backup windows, frequent RPO • Fast recovery of frequently accessed data • Fast data recovery times for short RTO
Virtual Libraries	Faster data availability and accessibility Automation and consolidation of data Data de-duplication	• Rapid restore times • Shorter backup windows • Automated storage • Flexibility of configuration
Disk to Disk to Tape (D2D2T)	Combines best characteristics of both tape and disk solutions	• Covers all data protection and archiving scenarios

In order to decide which will be the best tool for the job, you first need to consider your own needs – we look at this in the next section.

9. Achieving a balance

So far in this paper we have set out available storage options, and given an indication of where they might be most suitable. But how should an organisation go about deciding the most appropriate mix of storage technologies for their own needs?

A first step concerns how to map data protection requirements onto appropriate storage capabilities. Each element of data protection has constraints which can be used to determine the types of capability that may be most appropriate. We consider the constraints in the following table – please note that it is not an exhaustive list or a technical discussion, however it should give an indication of what choices exist.

Scenario	Constraints	Considerations	Use tape when:	Use disk when:
Data backup and restore	• Quantity of data to be backed up • RPO and RTO • Length of time to store backups	The higher the backup capacity, the more the cost of media becomes a factor. How quickly do you need to get your data back? If backups are to be kept for very long periods, the solution may need to utilise low cost media with a long shelf-life	Budgets are constrained. Speed of access is less of an issue. Longer term storage is required.	Access times are paramount and cost is less of an issue. Disk-based backups are only retained for a short period (potentially before archiving to tape).
Business continuity and disaster recovery	• The allowable down-time for the business – RTO and RPO • Cost involved	If business down-time is critical to success, then systems and applications will need to be replicated on a secondary site, potentially in real-time.	Budgets are constrained. A recent point-in-time copy of information would be sufficient for disaster recovery.	Down-time needs to be kept to an absolute minimum and real-time replication is required.
Archiving of under-utilised or static content	• The allowable time for restore from archive • Cost of archival storage	Certain information may be required on a 'near-line' basis, that is, near online. For other information, offline access may suffice.	Offline access is sufficient, that is, a small amount of delay is seen as acceptable at restore time. Archival costs are to be minimised.	Near-line access is required, such that information must be restored from archive with minimal disruption, and cost is less of an issue.
Data retention for compliance reasons	• E-discovery time • Specific media requirements • Security	Discovery can be very expensive in terms of legal fees if information is not readily accessible. If data is to be retained for very long periods, the cost of re-copying from older media (disk or tape) becomes a factor. Some data retention laws stipulate the type of media to be used, for example concerning non-immutable media.	Lowest cost is required. A reliably long-term shelf life is needed to hold data over extended periods. WORM or Tape-based encryption is desirable (available with LTO-4 onward)	Fast time to e-discovery is an overriding factor.

With this level of understanding, you can consider what information is most important to you as a business, what constraints exist on that information, and therefore what are the most appropriate options for its storage within a hierarchical architecture. A good first step is to derive an 'ideal world' view of what your storage architecture needs to look like, before undertaking an assessment of your existing storage architecture. The resulting gap analysis can enable you to establish:

- The risks to the business of not having an appropriate mechanism in place
- The priorities in terms of what needs to be done first
- An action plan to set out shorter and longer term tasks to resolve any issues

With this in place, you will have a firm starting point to move your storage architecture forward, though of course we concur that storage does not exist in a vacuum. In most cases storage strategies must also take account of the storage management, content management and archiving software tools to be employed. Such considerations are outside the scope of this paper.

10. Conclusion

Despite advances in disk technology and 'hype' to the contrary, the demand for tape continues. Ongoing advances in tape performance, capacity, security (with encryption) and cost plus its long lifetime ensure that tape technology offers a compelling proposition in tiered storage architectures and looks set to do so for the foreseeable future.

As we have seen in this paper, the question is not whether to use one technology or another, but where and how to best use both tape and disk as part of a comprehensive data protection strategy especially when tiered storage architectures are considered. Each technology should be considered on its merits against the value of the data to the business over its life-cycle and the TCO involved. A hybrid storage approach that combines disk with tape allows customers to leverage the speed and high-availability features of disk along with the long-term retention, low-cost and scalability features of tape.

In conclusion, storage platforms will continue to evolve – we are already seeing the solid state and flash-based storage technologies reaching mainstream use where access times are paramount, and storage densities on both disk and tape continue to evolve too. Such things will undoubtedly lead to new storage options, each with their own capabilities, costs, benefits and constraints, with over-arching benefits from advancements in areas such as data deduplication. Success in data protection lies in considering each new technology from a storage architectural perspective, choosing specific technologies according to a range of data protection needs over time.

References

[1] Evolution of x86 Server Estates (November 2009)
Modernisation drivers and practicalities
http://www.freeformdynamics.com/fullarticle.asp?aid=902

[2] 2008 study by University of California looking into disk access patterns for the network of a large local business over 22TB of disk-based data

Please see www.freeformdynamics.com for a wide range of other papers, reports and research notes on the effective use of IT in a business context.

About Freeform Dynamics

Freeform Dynamics is a research and analysis firm. We track and report on the business impact of developments in the IT and communications sectors.

As part of this, we use an innovative research methodology to gather feedback directly from those involved in ITC strategy, planning, procurement and implementation. Our output is therefore grounded in real-world practicality for use by mainstream IT professionals.

For further information or to subscribe to the Freeform Dynamics free research service, please visit www.freeformdynamics.com or contact us via info@freeformdynamics.com.

About Tandberg Data

Tandberg Data is a leading global supplier of end-to-end data protection solutions for small and medium-sized businesses. The company's wide selection of business-grade storage solutions include disk-based solutions, tape drives, autoloaders, libraries, removable disk and tape media and software for mainstream configurations.

For further information please visit www.tandbergdata.com.