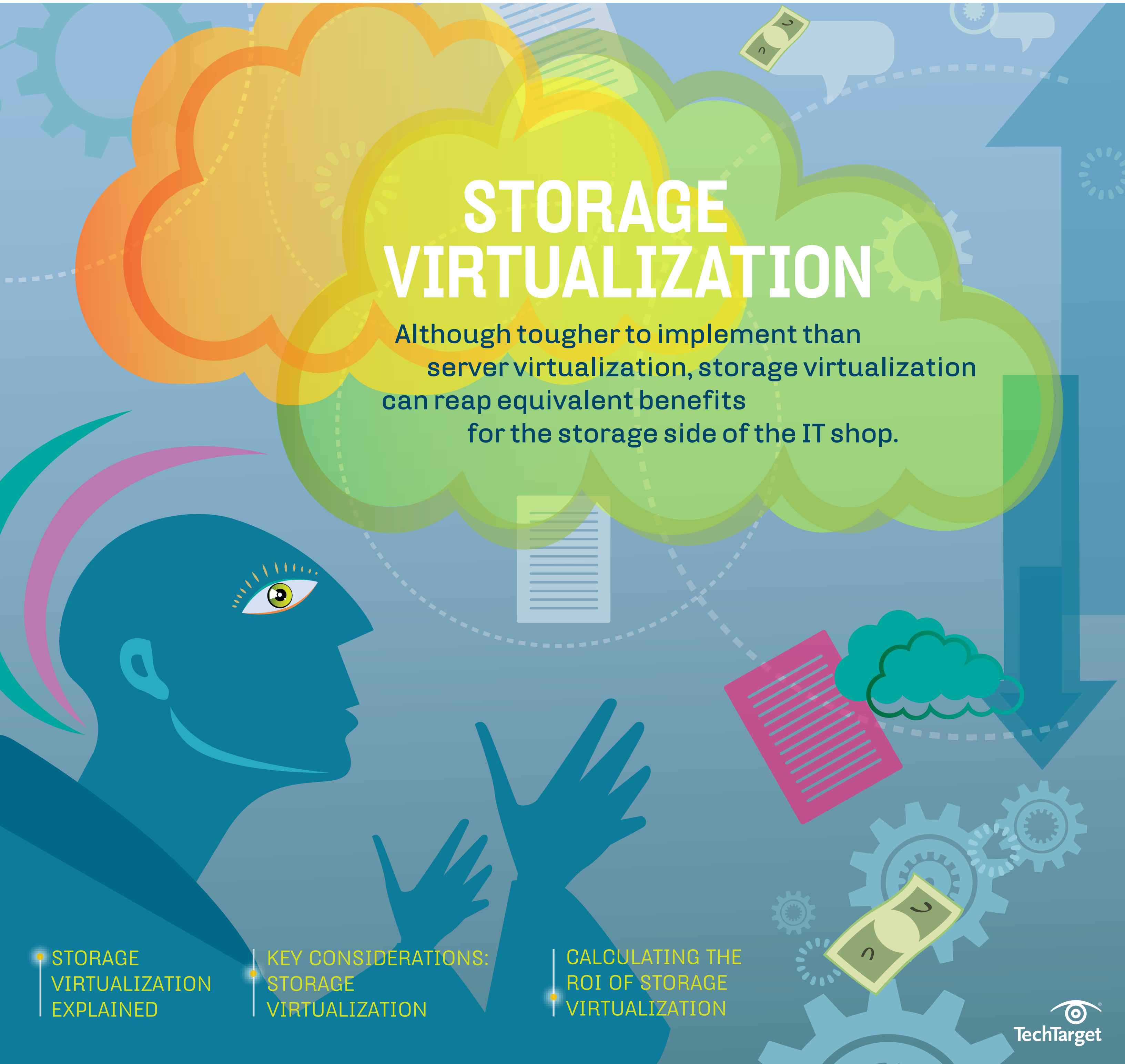
Managing the information that drives the enterprise

STORAGE

BUYER'S CHECKLIST



STORAGE VIRTUALIZATION

Although tougher to implement than server virtualization, storage virtualization can reap equivalent benefits for the storage side of the IT shop.

TechTarget

BUYER'S CHECKLIST:

Storage virtualization

Although overshadowed somewhat by server virtualization, storage virtualization technologies help companies use their existing storage assets more efficiently.

BY MANEK DUBASH

STORAGE VIRTUALIZATION WAS actually born a long time ago, when hard disks stopped exposing their physical layouts to operating systems and instead exposed a logical layout remapped internally to the physical structure. This technique has been used in different forms ever since; RAID, for example, is a form of block-level virtualization or abstraction of the underlying hardware.

More recently, storage virtualization has been developed and deployed on a much wider scale than just abstracting disk hardware from the operating system (OS). However, defining and implementing storage virtualization isn't as easy as with server virtualization. Still, the benefits of storage virtualization, and recent advances in its ease of use and deployment, are resulting in a growing uptake of the technology.

Modern storage systems are complex. They consist of a range of media types, typically from a variety of vendors, each with different characteristics and tradeoffs. For example, the fastest but most expensive component of an enterprise storage system is likely to be solid-state drives (SSDs), the next fastest would be 15,000 rpm SAS disks, then SATA disks, all the way down to tape or even non-rotating disks in a MAID configuration. These multiple systems, each consisting of maybe thousands of devices, might be spread over several data centers and geographies.

Such a huge number of assets and the data they contain would be very hard, if not impossible, to manage without some form of virtualization. The ultimate goal is to reduce costs by making the management of a heterogeneous data storage environment simpler and easier.

There are a number of technologies that fall under the storage virtualization umbrella but, at the highest level, storage virtualization is implemented in one or more of three main areas: host server, storage array, and network appliance or switch.

STORAGE VIRTUALIZATION IMPLEMENTATIONS

Host-level virtualization. The logical volume manager (LVM) as found in Linux is a good example of host-based storage virtualization. The LVM allows different elements of direct-attached storage (DAS) to be aggregated into a single pool, and managed and viewed by the host OS as a single entity. LVMs can also add such functions as software RAID, snapshotting and remote replication. Although the technology is in many cases free, host-level virtualization has limited application in an enterprise environment as each server's implementation must be managed separately, making it costly to maintain.

Storage-level virtualization. Virtualization of network-attached storage (NAS) is more common. Virtualization technology means the storage can be accessed using multiple file systems. For example, to Linux/Unix servers, NAS looks like an NFS share, while to Windows users it appears as a CIFS share. Where the data is (or what type or speed of disk it resides on) is invisible and, to most users, immaterial. It allows storage to be managed from a single location, and isn't host- or OS-specific.

Storage virtualization explained

Key considerations: Storage virtualization

Calculating the ROI of storage virtualization

Array-based virtualization. Array-based virtualization consists of appliances that can unify a disparate set of storage arrays, presenting them as a single system. This allows the lifecycle of legacy devices to be extended by adding features such as thin provisioning, snapshots and replication where such features may not exist or haven't been licensed on each external array. The unification of systems "behind" the appliance can also minimize the impact of operations that occur "behind" the appliance, such as data migration, on the rest of the network.

Network-level virtualization. This applies to storage-area networks (SANs) and consists of appliances on the network that abstract the storage, for example, to combine logical unit numbers (LUNs) from different SANs and present them as a single LUN. Alternatively, it can do the opposite and slice a LUN into smaller LUNs to enable smaller and thus faster restore operations, for example. Other functions include security and replication between SANs and across wide-area network (WAN) links.

Switch-level virtualization. More recently, storage virtualization technology has been implemented in network switches with the aim of reducing costs in large, heterogeneous environments. Since you need the switches anyway, implementing virtualization in the switch can unify storage management by reducing the number of boxes and vendor-specific interfaces, as well as the number of hardware purchases; it can also help to minimize any virtualization performance penalty.

Irrespective of which layer is involved, storage virtualization delivers a range of features and functions that give managers the flexibility to provision storage as required by the business, rather than having it imposed on them by the limitations of the technology.

STORAGE VIRTUALIZATION FEATURES

Flexibility, performance and management. Virtualization enables storage to be provisioned according to business requirements rather than technology limitations. Capacity, availability and performance are key storage attributes that business users care about. Using virtualization, managers can allocate storage to alter one or more of these attributes by, for example, adding blocks to arrays to make them bigger, striping across more blocks or disks to improve performance, or adding mirrored disks to enhance reliability.

A key driver for the growing interest in storage virtualization is the unchecked growth of data volumes. Estimates of the annual growth rate of unstructured data vary, but 50% is often quoted and a reasonable estimate. Even though storage technologies are getting cheaper and more capacious, organizations struggle to keep up with this level of growth.

Without virtualization, storage managers would have to spend now to accommodate data growth over the next three years, the typical length of an IT purchasing cycle. For most, this isn't a financially viable strategy. The alternative is to add disks as required, but this leads to high overheads, and a great deal of systems integration and data migration work.

Storage virtualization can help remove the management overhead by ensuring data isn't fragmented across different storage technology barriers but treated as a whole. For example, when buying more storage, matching new disks to existing RAID arrays or migrating data across arrays to make use of additional capacity are time-consuming, mundane tasks. Instead, virtualization enables organizations to integrate extra capacity into the storage pool by enlarging existing LUNs, rather than having to move and rearrange existing data.

Additionally, thin provisioning allows more storage to be allo-

(Continued on page 7)

Storage virtualization explained

Key considerations: Storage virtualization

Calculating the ROI of storage virtualization

Key considerations: Storage virtualization

ISSUE	WHY IT'S IMPORTANT	WHAT TO WATCH OUT FOR
Storage virtualization technology (in general)	Automates many management functions; helps simplify increasingly complex, heterogeneous multivendor systems	Another layer of software to manage; needs to scale without impairing performance; new practices to be learned
Works with existing SAN infrastructure	Retains investment in existing assets	Not replacing the SAN has a potential loss of benefits from newer techs
Scalability	Storage needs to be able to grow with the business	Potential for vendor lock-in; may lead to greater upfront investment than envisioned
Interoperability	Cross-vendor interoperation, ease of management	Potential for vendor lock-in
Ability to reverse installation	Flexibility to decommission the system if it doesn't deliver promised performance, interoperability or scalability	Potential for additional cost and complexity of product
ROI calculation	Enables better product/ technology choices	Many costs are hard to quantify
Host-level virtualization	Widespread; included in the OS; low cost	Must be managed individually for each server
Storage-level virtualization	Unites multiple NAS boxes; provides central management; relative simplicity	May need added features such as security; high level of abstraction may limit performance
Network-level virtualization	Combines advantages of host- and network-level virtualization; has widest reach across the data center; high performance	Potentially higher cost, additional level of management
Management of physical assets (adding and removing devices)	Reduces management load; greater storage deployment flexibility	Need to be aware of which physical asset is attached to which virtual storage pool

(Table continued on page 7)

Key considerations: Storage virtualization

(Table continued from page 6)

ISSUE	WHY IT'S IMPORTANT	WHAT TO WATCH OUT FOR
Thin provisioning	Allows overprovisioning to enable just-in-time storage purchases	Needs precision, planning and awareness of physical storage availability
Automated tiering	Lowers storage costs by reducing management overhead and optimizing storage utilization	Proprietary technologies can lead to vendor lock-in
Snapshotting	Automates backups	Can be costly feature on multiple file systems; can be CPU-intensive; may not work on distributed VMs
Replication	Data mirroring offers better protection than scheduled backups	Synchronous replication is resource-intensive and can be costly, while asynchronous replication is riskier
Virtual tape libraries	Better performance than tape; more easily managed; makes archiving more efficient	Additional hardware costs, extra layer of complexity

cated than is physically installed. This helps organizations to maximize the utilization of existing storage assets, rather than paying for energy to spin disks that will remain empty during much of their lifetime.

Automation, tiering and management. Virtualization also enables automation. IT managers and end users need no longer know or care where their data is stored. It might be local or remote, or spread across multiple partitions in a number of geographic locations but, to end users, it looks like a single storage system. This not only improves manageability and usability, but helps to eliminate single points of failure by distributing data across the storage environment.

Storage systems are frequently divided into tiers, with automated tiering a key element of storage virtualization. Automation is essential, as managing data manually to ensure that it resides on the most appropriate medium is an impossible task. The technology allows the most valuable data—usually a combination of the most recently created/used and/or mission-critical—to reside on the highest performing storage medium. As data ages, its value declines and the system cascades it down the chain from SSDs to high-rpm SAS disks, to SATA, and eventually to a near- or off-line archive. The process is invisible to end users and applications, and helps to deliver a consistent, cost-effective quality of service across the data storage infrastructure. The core advantage is that it matches the value of the data to the cost of the storage medium, so that only high-value data resides on high-performance storage, reducing overall costs.

Storage virtualization can help remove the management overhead by ensuring data isn't fragmented across different storage technology barriers.

A further benefit of storage virtualization in a multivendor environment is that storage managers no longer need detailed technical knowledge of each vendor's equipment because they're managing storage at a higher level and from a single console. In addition to ease of administration, consolidated management yields cost savings.

Data protection. Storage virtualization is a key enabling technology for options such as snapshotting and replication, which in turn enable the enterprise to implement affordable data protection strategies for backup and disaster recovery. The difference between virtualized storage replication and snapshotting is that with a physical operation all processes take place on the source array, which needs to have the capacity and performance to handle the snapshots and replicated data. With a virtualized array, data can be redirected to any other array or even a JBOD, as the destination is invisible to the source array.

Storage virtualization is now extending to tape libraries. A virtual tape library (VTL) is an array of disks presented as a tape library

Calculating the ROI of storage virtualization

ESTIMATING HOW MUCH a storage virtualization system is likely to save over a given time period is difficult. There are many variables involved, many of which boil down to the way existing storage systems are configured, who manages them and how.

Leaving aside issues specific to each organization, a key variable in this calculation is the number and volume of existing storage arrays. One of the main reasons for virtualizing storage is to automate processes and unite disparate systems (that were probably acquired at different times and for different reasons) under a single point of management with the aim of making them look like a single storage pool where appropriate.

Storage virtualization can save by reducing the time spent on mundane tasks such as adding and removing drives, and the resulting data migrations and RAID rebuilds. It's reasonable to assume that the bigger the storage environment, the larger those savings are likely to be. Other savings may result from reduced storage asset acquisition as a result of improved utilization of existing resources, leading to consequently lower energy and space footprints.

The alternative is to buy storage only from a single vendor whose technology is logically unified. However, a shop will need to be comfortable with its relationship with that vendor and its technology roadmap.

to backup applications. As well as being much faster than tape, a VTL offers other benefits. Often, tape libraries aren't capacity optimized as different backup applications use different formats, so a tape can be locked out for use by other applications even though it may not be full. With a VTL, a backup application can archive data from the disk arrays to a physical tape library in a single operation, which helps to maximize tape utilization. It can also optimize data streaming rates to eliminate shoe-shining—where the tape shuffles back and forth if data doesn't arrive in time to write the next block—which slows the process down and adds to tape wear. ■

Manek Dubash is a U.K.-based business and technology journalist with more than 25 years of experience.

TechTarget Storage Media Group

Storage magazine subscriptions:

www.SearchStorage.com

Storage magazine
275 Grove Street
Newton, MA 02466

editor@storagemagazine.com

TechTarget Inc.
275 Grove Street, Newton, MA 02466
www.techtarget.com

About TechTarget: TechTarget publishes media for information technology professionals. More than 100 focused websites enable quick access to a deep store of news, advice and analysis about the technologies, products and processes crucial to your job. Our live and virtual events give you direct access to independent expert commentary and advice. At IT Knowledge Exchange, our social community, you can get advice and share solutions with peers and experts.