

StorHouse®



Intelligent Storage Virtualization Software

**Ensuring data availability, accessibility,
integrity, and cost-efficiency throughout
the information lifecycle**



FileTek, Inc.

9400 Key West Avenue
Rockville, MD 20850
Phone: 301.251.0600

International Headquarters:

FileTek Ltd

1 Northumberland Avenue
London WC2N 5BW
Phone: +44 (0) 207.872.5583
intsales@filetek.com

For more information: info@filetek.com
www.filetek.com

©2009 FileTek, Inc. All rights reserved. FileTek and StorHouse are U.S. registered trademarks of FileTek, Inc. Other trademarks included herein are the property of their respective owners.

Table of Contents

About Server and Storage Virtualization	1
About StorHouse.....	2
Business Drivers for Storage Virtualization	3
Computing the Total Cost of Ownership	4
Managing the Shift from Gigabytes to Petabytes	5
Intelligent Virtual Storage Environments	6
Integrating New Storage Technologies.....	6
Providing Universal Access	7
Providing Affordable Scalability	8
Managing Storage Efficiency and De-Duplication.....	9
Maintaining Data Integrity	10
Simplified and Efficient Backup and Recovery.....	10
Lowering System Administration Overhead	11
Summary	13
Footnotes.....	14



StorHouse: Intelligent Storage Virtualization Software

Ensuring data availability, accessibility, integrity, and cost-efficiency throughout the information lifecycle

Virtualization is an outstanding way to achieve exceptional business value from IT investments. By introducing a software abstraction layer between the physical components of a computing infrastructure and the applications that use them, virtualization platforms transparently pool hardware resources and properly assign them to individual applications.

This whitepaper discusses business drivers for storage virtualization and describes the characteristics of an intelligent virtual storage implementation. It also introduces the FileTek StorHouse® virtualized storage platform and describes the intelligence features it provides.

About Server and Storage Virtualization

Server and storage virtualization are critical components of a well-engineered enterprise computing environment. Server virtualization is an established protocol that delivers clear cost savings. For example, many organizations use server virtualization to make a single, powerful enterprise-class server appear on the network as a set of individual servers that can host multiple applications. In this context, virtualization software enables companies to deploy numerous right-sized servers and only incur the capital and management costs of a single physical machine.

Storage virtualization is the abstraction of physical storage locations. In other words, instead of interacting directly with a storage device, applications interact with virtualization software that tracks the physical location of data and makes it available regardless of the logical location.

Storage virtualization is similar in concept to server virtualization because it implements an intelligent software abstraction layer over a variety of heterogeneous storage technologies to create a sharable pool of storage resources. These resources can include SAN, RAID, SATA disk, and high-speed tape, among others. Furthermore, these storage elements can be comprised of existing storage assets and new storage devices that were purchased as part of the move to a virtualized environment. Once these diverse storage resources are pooled and under management, IT organizations can create and assign virtual storage environments from within the pool for use by specific applications and users. By intelligently blending multiple storage technologies, administrators can assign each application the appropriate storage quantity and characteristics that deliver the best performance at the lowest cost-per-terabyte.

About StorHouse

StorHouse from FileTek is a unique storage virtualization platform that can archive, retrieve, and back up massive amounts of relational and file-based information using an automatically managed pool of traditional and alternative storage devices. Organizations deploy StorHouse for digital preservation initiatives, active archive applications, database extension systems, and native file format backups of terabytes to petabytes of data residing on operational systems.

Figure 1 illustrates a sample StorHouse architecture.

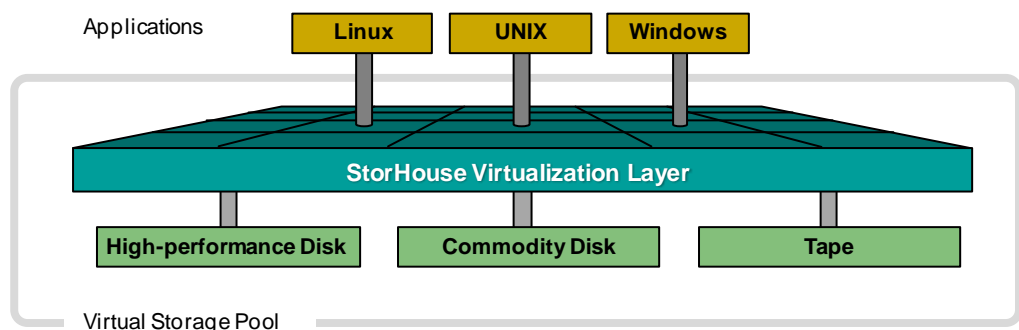


Figure 1: Sample StorHouse architecture

StorHouse provides an intelligent storage virtualization and data management layer that complements existing IT and storage infrastructures. It enables organizations to leverage existing investments in storage technology by matching storage resources to storage requirements. With StorHouse, administrators can transparently introduce new storage, retire old storage, and migrate data between diverse storage devices – all while maintaining 100% uptime and accessibility.

In cloud computing terminology, StorHouse is an “internal cloud” that abstracts storage away from diverse users and applications throughout the enterprise. In fact, neither users nor applications require information about the resources in the virtualized storage pool or expertise about how to manage them. The controlling technology consists of storage virtualization and data management layers that reside in the StorHouse cloud infrastructure. These software components dynamically scale storage capacity and the virtual file system that StorHouse presents to client applications.

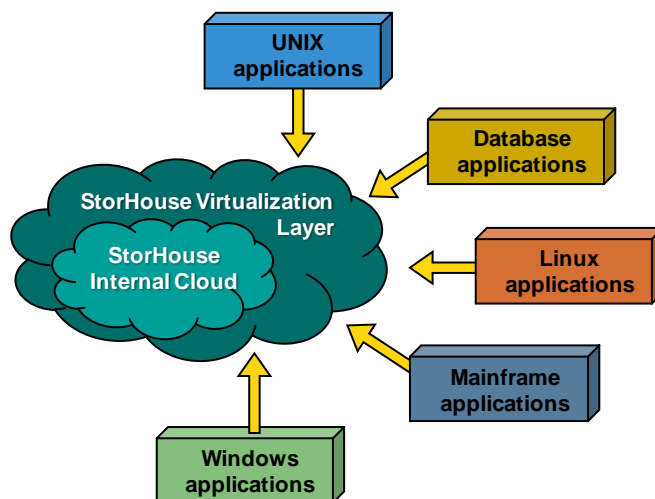


Figure 2: StorHouse as an Internal Cloud

Business Drivers for Storage Virtualization

The business drivers for storage virtualization are similar to those for server virtualization. CIOs and IT managers must cope with shrinking IT budgets and growing client demands, improve asset utilization, use IT resources more efficiently, and ensure business continuity. In addition, they are faced with ever-growing volumes of data and mounting constraints on power, cooling, and floor space utilization.

Computing the Total Cost of Ownership

With the amount of electronically generated information continuing to expand at an estimated compound annual growth rate of almost 60%¹, organizations are looking for ways to benefit from the cost-effectiveness and efficiencies of different media types. As data progresses through the information life cycle, its business value and access requirements change over time. Therefore, to achieve an appropriate cost/performance model for enterprise storage, information should ideally be migrated across various storage technologies to consistently and properly align the cost of storage with the varying value of data.

Many industry-recognized consultant groups have researched the total cost of ownership (TCO) of tape and disk as components of a storage infrastructure. While disk-based storage wins in terms of reduced access times for data, tape-based storage provides many advantages in terms of cost-per-terabyte of storage and ongoing power and cooling expenses.

Figure 3 shows the results of a TCO cost-comparison study performed by the Data Mobility Group.² The study examined the seven-year cost projections for three storage models: disk only, tape only, and an intelligently managed disk and tape blend. The initial storage environment for all configurations was 125 terabytes. Each year, the environment was estimated to add 25 TB of new data and grow at a rate of 20%.

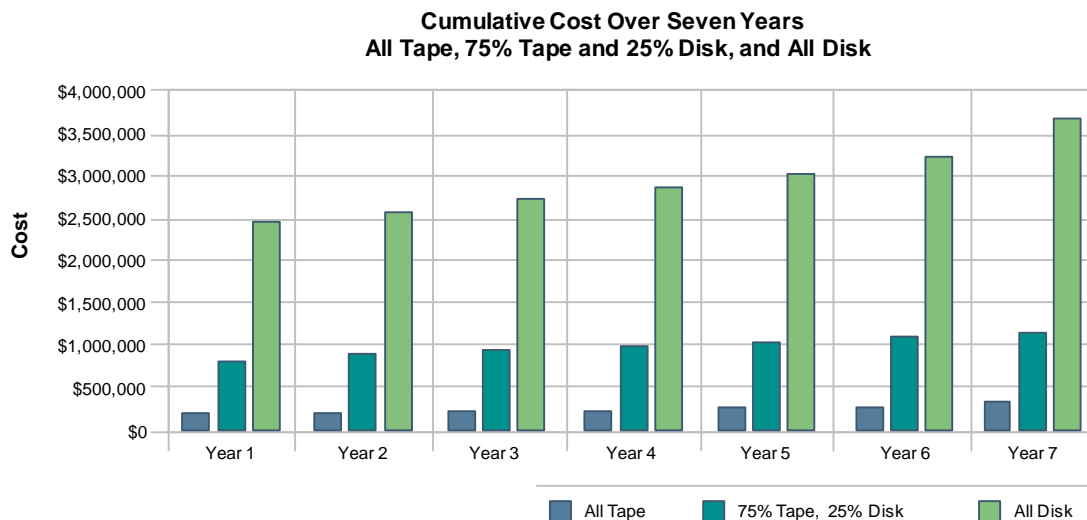


Figure 3: TCO Cost Study

As Figure 3 indicates, a storage environment that combines disk and tape can provide outstanding storage performance characteristics at a small fraction of the overall cost of a pure disk-only approach. Using storage virtualization, organizations can dramatically reduce power and cooling costs by blending different media types (high-performance tape to high-speed disk) with drastically differing power requirements and heat dissipation profiles. StorHouse can easily provide this technology mix in its virtual storage pool.

Managing the Shift from Gigabytes to Petabytes

Today, the world measures data volumes in terabytes and petabytes rather than megabytes and gigabytes. It's sometimes difficult to comprehend what this order-of-magnitude shift represents. Just think about it. If an organization were to write data at a modest rate of 50 megabytes per second, it would take just under 21 seconds to write one gigabyte. At the same speed, it would take almost 250 days to write one petabyte. Without the watchful eye of intelligent storage management software layers, very large data volumes could easily consume too many resources and cause extended outages during routine data management activities.

Because of the ever-growing volume of data, computing infrastructures require more intelligence to handle even routine data, system, and storage management tasks. They must be able to benefit from new technology as it becomes available, support innovative digital content formats, respond to the requirements of heterogeneous user communities, and affordably scale from terabytes to petabytes without performance degradation.

StorHouse clearly provides these and other features with its storage virtualization layer. It is a hardware agnostic, all software solution that can be deployed as a component of an overall storage infrastructure. The system provides a comprehensive, enterprise-class, storage virtualization layer that supports one or more storage locations. These locations can scale two-dimensionally to support increased storage capacity and number of stored objects. In addition, the StorHouse virtualized storage environment can span a wide variety of media types, including tape, disk, solid state devices, and most other generally available storage technologies. Because of this virtualization capability, StorHouse enables organizations to choose the best technology mix, thereby satisfying cost-versus-performance requirements while maintaining complete hardware vendor independence.

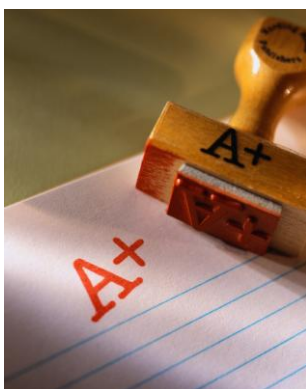


Intelligent Virtual Storage Environments

An intelligent virtual storage environment is about more than just storing data. It's about protecting information amidst technological advancements, efficiently managing escalating data volumes created in a greater variety of formats, and ensuring data integrity and accessibility over time.

Key intelligence features include:

- Automatically integrating new storage devices and technologies into an existing storage infrastructure and phasing out old ones as required
- Providing universal access to storage using standard interfaces and protocols



- Scaling the storage environment from terabytes to petabytes and beyond in an effective and affordable manner
- Managing storage cost-efficiency and data de-duplication
- Ensuring data integrity and uninterrupted data availability
- Supporting simplified and efficient backup and recovery
- Reducing administration overhead with easy-to-use, graphical-based system management and monitoring tools

The remainder of this paper describes these features and explains how StorHouse provides them all.

Integrating New Storage Technologies

Intelligent virtual storage environments integrate new storage devices with ease and ensure data persistence. To sustain persistence, they must protect current data from impending loss, threat, media degradation, and technology obsolescence.

Additionally, they support dependable data retrieval and maintenance mechanisms that remain compatible with future technologies.

When adopting new storage technologies, a major obstacle is the service interruption that occurs when moving data from old to new media. Without virtualization, there are two choices: Write new data to new media and leave old data in place, or migrate data from old media to new media in its entirety or in manageable chunks.

Both choices are fraught with problems. With choice one, how will applications know that data is stored in more than one location (which, in some cases, is not even an option)? In choice two, which data should organizations take offline, migrate, and then bring back online? What size maintenance window (weeks, months, or even years) should be scheduled to move a very large archive?

The StorHouse virtualized storage platform provides a better, more manageable method for adding new technology. With StorHouse, organizations can insert new storage technologies and phase out old ones transparently without service disruption, even during the migration process. Using a graphical management and monitoring tool, administrators can tell StorHouse to move one class of data to a new storage component. Throughout the migration process, StorHouse ensures continuous data availability and verifies the integrity of the newly written information.

StorHouse enables organizations to choose the best technology mix, thereby satisfying cost-versus-performance requirements while maintaining complete hardware vendor independence.

Providing Universal Access

An intelligent virtual storage system provides universal access to sustain uninterrupted data accessibility and availability. StorHouse provides this capability through its standard connectivity protocol interface layers, which include NFS, NTFS and CIFS. Using these standard interface protocols, even the largest StorHouse virtual storage environment can appear on a network as a single, unified file share. Users and applications can access the file share through traditional drive letter mapping (V: \) or a server-oriented file path (\\<SERVERNAME>\<SHARENAME>).

Because of these open systems protocols and established IT connectivity conventions, organizations can easily add StorHouse to heterogeneous computing environments that include UNIX, Linux and/or Windows-based platforms without modifying existing applications or re-training users. Furthermore, these diverse networked platforms can concurrently and securely access information on the same StorHouse system, thus enabling data sharing across the enterprise.

Providing Affordable Scalability

Intelligent virtual storage systems must also provide affordable scalability to accommodate potentially terabytes to petabytes of data. Furthermore, to keep storage costs down, they must also successfully align the cost of storage and individual media performance characteristics with the changing value of data. Most non-virtualized systems typically ignore these characteristics and support only random access media such hard disks, or to a lesser extent, certain optical technologies to store growing data volumes.



In contrast, StorHouse not only offers affordable scalability but also helps reduce overall storage and storage administration costs. The system provides a strategically assigned technology blend that satisfies the various performance requirements of different data sets. A simple example is a mix of disk and tape. Data that requires sub-second access time can be stored on disk while older, less frequently accessed data can be stored on tape. In addition, StorHouse can seamlessly migrate data between different media types should access requirements change.

Typically, storage vendors present either a physical or logical block device to a host operating system. Then, it's up to the operating system to create a file system on the block device. In effect, storage vendors are making the operating system responsible for managing the state and data integrity of the storage system.

When an operating system creates a file system, it also creates data structures called inodes that store metadata (for example, permissions, ownership, etc.) for each file in the file system. On many file systems, the number of available inodes is fixed at file system creation, thus restricting the maximum number of files that can be supported. As file systems grow in number of objects, the operating system uses more and more resources to ensure object integrity and maintain the inode structure.

Storage vendors publish the absolute maximum number of files a file system can store as well as a recommended maximum. While a file system may be able to theoretically contain 2^{64} files, if that file system becomes corrupt, the recovery time for such a large number of files would be intolerable in terms of system downtime and data inaccessibility. Even file systems that support journaling derive the maximum number of supported files based on how much downtime an organization can tolerate rather than on the calculated maximum number of files.

StorHouse takes a fresh, innovative approach to managing large numbers of files. This approach not only virtualizes physical devices but also implements a virtual file

system with no inode constraints, pointers, or operating system-imposed limitations on the volume of information or the number of stored objects. To provide this unprecedented level of scalability, the StorHouse virtual file system uses a replicated relational model that enables the storage of trillions of records without impacting the integrity or stability of the repository. StorHouse remains a cost-effective, accessible, performance-oriented solution well beyond the benchmark that defines when most file systems and relational databases begin to fail.

Managing Storage Efficiency and De-Duplication

Storage efficiency is critical when considering the total cost of ownership of a storage infrastructure. For simplicity, consider a legacy backup strategy that performs a weekly full backup and daily incremental backups of server-based business information. In this very simple case, when a single file changes on a daily basis, over the course of a one week, the backups will generate seven copies of a single file under management. Extrapolate this example over months or years, and with millions of files, it's easy to see that most backup data under management is redundant.



Intelligent virtual storage systems ensure that critical data is always available and easy to locate. Therefore, it follows that retaining irrelevant data is a waste of money, time, and storage resources.

StorHouse is an intelligent virtual storage platform that understands the importance of data availability and de-cluttered storage. To support business-centric, efficient storage, StorHouse provides data de-duplication tools to manage, identify, and remove duplicate copies of data. For example, rather than store every incremental backup copy of a file, administrators can define an enterprise virtual storage policy that retains only the most recent three (or five or 10) versions on StorHouse. The number of versions under management is a policy decision that can be configured according to long-term business objectives. When the number of retained versions is limited, StorHouse automatically deletes older versions as new ones are added to the file system, recovers the deleted space, and makes it available for reuse. This high level of StorHouse automation and intelligence makes traditional backup applications obsolete and eliminates cumbersome restores.

If the primary copy on the operational system becomes corrupted or is accidentally deleted, applications and users can access the backup copy directly from the StorHouse virtual file system. No restores are necessary because every StorHouse

file is available and accessible in native file format. Business continuity is never compromised because there is no service interruption.

StorHouse provides additional benefits. It continuously monitors media utilization to ensure efficient operation. And, to help minimize storage costs, the system automatically migrates data to the most appropriate storage location according to business retention/performance parameters. When using serial tape, StorHouse stores data in a compact format to approach near 100% media efficiency.

Maintaining Data Integrity

The potential for data corruption increases proportionally with the size of the storage environment. Even though contemporary storage solutions are incredibly reliable, intelligent storage management platforms go a step further. They support built-in preventive maintenance features that continually monitor, detect, and correct data integrity problems such as silent corruption and bit rot before they occur.

With traditional systems, administrators typically do not know if a file is corrupted until a user or application tries to access and use it. Although there is an immediate need for the file, it remains inaccessible until it can be restored and recovered, which could take days to weeks. And worst case, what happens next if the backup copy is also corrupted? If this happens, the file may no longer exist in a usable format.



StorHouse prevents the preceding scenario from ever occurring because data integrity is an integral function of the storage virtualization layer. The system physically stores all checksums required to identify a corrupt file as secure metadata with the file. Using this mechanism, StorHouse can constantly self-validate the health of the data it manages. Using a set of comprehensive background processes, StorHouse can proactively identify corrupt content and perform auto-repair functions long before the content is ever needed or accessed. Furthermore, if desired, StorHouse can maintain multiple file copies in one or more geographically dispersed locations for even greater levels of content assurance and disaster recoverability.

Simplified and Efficient Backup and Recovery

Intelligent storage virtualization systems simplify backup and recovery operations to utilize storage resources more efficiently and reduce the time-to-data. Providing these objectives is a major StorHouse strongpoint.

StorHouse uses a native file format backup method – an approach that saves content in the same format used by the originating application. Unlike backup applications

that store data in a vendor-specific, proprietary format, StorHouse provides direct access to all data at any time. Native file format backups eliminate the need to restore data to disk prior to access, reduce recovery times, sustain uninterrupted data availability, and provide a more reliable, secure, and cost-effective way to safeguard and protect critical enterprise information.

In addition, StorHouse merges archive and backup operations into a single system that provides comprehensive digital preservation attributes, uninterrupted archive accessibility, and the security and assurances of reliable backups. Applications and users have routine access to all StorHouse data at any time. Once a file is backed up to StorHouse, its status varies with its access requirements. For example, based on changing user needs, a backup file can transfer transparently and automatically to an archive status without any physical movement or additional media management. When access requirements diminish to “I will never need to access the file again but must keep it for security, compliance, and recovery purposes,” the file regains its backup status. The significant business value of such high-level system automation, transparency, data availability, and access flexibility distinguishes StorHouse from other backup and archive applications on the market today.

StorHouse remains a cost-effective, accessible, performance-oriented solution well beyond the benchmark that defines when most file systems and relational databases begin to fail.

Lowering System Administration Overhead

Intelligent storage virtualization systems also reduce system administration overhead because they are efficient data and storage managers. The StorHouse virtual storage system is no exception. StorHouse promotes efficient management and reduces overall operating expenses. A part-time administrator can manage a large StorHouse environment because on-going tasks such as data replication and migration can be completely automated. Once the administrator configures control policies, the StorHouse rules-based policy engine takes over and manages all data movement automatically.

StorHouse provides an easy-to-use, comprehensive system administration and monitoring tool called StorHouse/CCi. This software supports command-line and browser-based interfaces that use the same underlying command structure. The command-line interface is helpful for automating tasks through scripting and for integrating StorHouse with third-party tools and solutions. The graphical user interface is ideal for managing one-off events and other exceptions. Typical one-off

operations are disk evacuations that support disk upgrades, server upgrades, or operating system maintenance activities and disk, tape, or server expansions that provide more capacity. Some one-off events (for example, an overall change to the storage characteristics for a particular class of information) are best handled by modifying the rules that direct the policy engine.

Figure 4 illustrates a sample StorHouse/CCi system management screen.

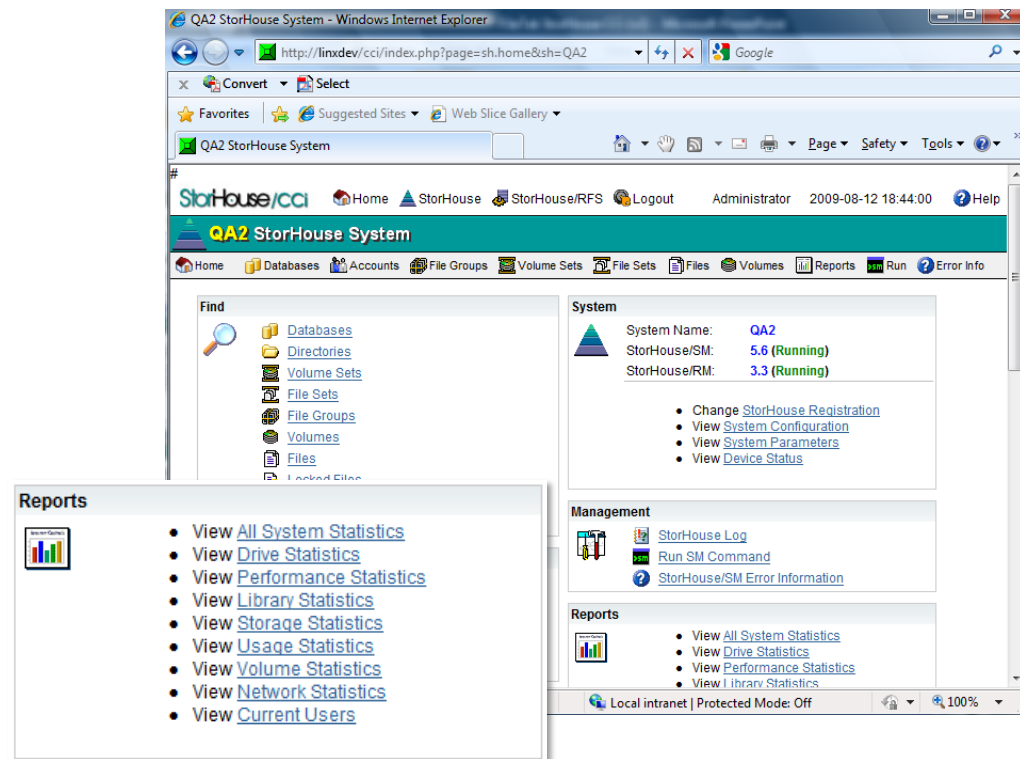


Figure 4: Sample StorHouse/CCi Administration Screen

Administrators can use StorHouse/CCi to create policies that enable automated system monitoring and alerting on a 24/7 basis. Once these policies are defined, StorHouse will send email alerts when thresholds are met and begin fixing the problem condition automatically.

Summary

Today, IT organizations are faced with many challenges, including how to manage increasing data volumes, seamlessly integrate emerging technologies, plan for long-term access requirements, and affordably scale from terabytes to petabytes and beyond. Storage virtualization is a technique that helps organizations overcome these challenges because it greatly simplifies managing heterogeneous storage environments.

An intelligent storage virtualization system supports the following features:

- Automatically integrating new storage devices and technologies into an existing storage infrastructure and phasing out old ones as required
- Providing universal access to storage using standard interfaces and protocols
- Scaling the storage environment from terabytes to petabytes and beyond in an effective and affordable manner
- Managing storage cost-efficiency and data de-duplication
- Ensuring data integrity and uninterrupted data availability
- Providing simplified and efficient backup and recovery protocols
- Reducing administration overhead with easy-to-use, graphical-based system management and monitoring tools

StorHouse from FileTek is an intelligent virtual storage platform that satisfies all of these requirements. It provides the most comprehensive and advanced storage virtualization capabilities on the market today. In fact, StorHouse takes virtualization to the next level by integrating data integrity management as a standard operation and then virtualizing the entire storage infrastructure, including the file system.

StorHouse benefits include:

- Automatic individual user file restores without administrator or IT involvement – no need to restore from tape to disk
- High availability to all data at all times

- Automated data access during recovery
- Continuous data protection against silent corruption and bit rot through monitoring, failure analysis, and self-healing
- Decreased need for costly D2D backup solutions
- Automated data and storage management, including data migration, replication, recovery, backup, and retention
- Shareable archive data between systems and applications
- Smaller more manageable file systems
- Capability to perform compliance deletes
- File versioning over time
- Built-in technology growth path that preserves investments in existing hardware and supports adding new storage options, including disk and tape-based hardware, at any time
- Ability to replicate data automatically to other StorHouse systems
- Lowest total cost of ownership and cost per terabyte or petabyte of data

For more information about how a StorHouse virtual storage environment can reduce total cost of ownership, ensure efficient operation, automate storage management, provide uninterrupted data availability, promote business continuity, and support affordable scalability to petabytes and beyond, contact a FileTek sales representative at info@filetek.com or call 301 251-0600.

Footnotes

¹EMC Press Release, “New Study Forecasts Explosive Growth of the Digital Universe – Spotlights Worldwide Phenomenon of ‘Digital Shadow,’” March 11, 2008, http://www.marketwire.com/mw/rel_us_print.jsp?id=831713.

² Data Mobility Group, White Paper, “Is Tape Cheaper Than Disk?” Diane McAdam, October 20, 2005.