



Successfully implementing a private storage cloud to help reduce total cost of ownership

Contents

- 2 Executive summary**
- 3 What is a storage cloud?**
- 3 A story—or three—for inspiration**
- 3 *Out of disk space*
- 4 *Goodbye to old storage*
- 4 *Disaster recovery, cloud-style*
- 4 Private storage cloud architecture:**
- An overview**
- 6 IBM and storage cloud technology**
- 6 *Dynamic storage management*
- 8 *Scalable capacity and performance*
- 8 *Concurrent, multiprotocol data access*
- 9 *New levels of manageability*
- 10 Getting up and running**
- 11 Conclusion**

Executive summary

Reducing the total cost of ownership (TCO) for data storage is a top priority for most IT managers today. The initial storage hardware purchase is only the beginning, and often not the most expensive part of the solution. With data continuing to increase exponentially, it is not likely that you will be able to reduce the amount of data stored. Therefore, reducing your TCO for your storage infrastructure requires you to improve your overall efficiency—including daily maintenance, upgrades and the dynamic introduction of new technologies to an existing infrastructure to improve scalability. This includes the ability to mix server and storage technologies, nondisruptively add and remove storage, and move data as requirements fluctuate over time without incurring downtime. You also need to be able to provide multiple levels of business service to meet availability or regulatory compliance requirements across your organization.

Achieving these goals requires a storage technology that is flexible, scalable and easy to manage. For many organizations, private storage cloud is the answer. By decoupling the node running the application and the storage of data, a private storage cloud solution can provide the flexibility, scalability and manageability needed to support explosive data growth while reining in costs.

This paper defines what is meant today by the term private storage cloud and provides insight as to the types of technologies you can look forward to in the not-so-distant future. We will examine how IBM Smart Business Storage Cloud can facilitate a successful implementation of your next-generation storage cloud infrastructure—one that can offer improved availability, flexibility and efficiency to help reduce your overall TCO.

What is a storage cloud?

A storage cloud is an emerging technology and an integral part of the future of data storage. Designed for affordability and efficiency, a flexible storage cloud solution can optimize your available storage capacity, enabling fast, easy access to storage where and when you need it.

There are two types of scenarios available for a storage cloud: public or private. An example of a public implementation of a storage cloud would be a service that allows you to back up your video files to the Internet so that end users will be able to access them on demand. An example of a private implementation of a storage cloud would be enabling hospitals and facilities within a network to access stored images from a medical imaging provider.

While the public storage cloud provides for variable billing options and shared tenancy, the private cloud offers fixed charges and dedicated tenancy for enterprises looking for flexibility around ownership, management and maintenance of the storage cloud.

A story—or three—for inspiration

To better illustrate the concept of storage cloud technology, let's take a look at a few situations storage administrators face each day and how a storage cloud environment can help.

Out of disk space

It's Friday evening at 4:59 p.m. and the phone rings. It is the vice president of marketing, and he sounds concerned. He explains that the marketing campaign that is supposed to release Monday morning is on hold because the designers ran out of storage space for the artwork. You look and see that the marketing department has fully utilized its pool of storage. You quickly identify that there is still plenty of space on a storage pool you added the other day. Using the management interface, you can dynamically expand the storage allocation for the marketing department to include space in this new storage pool. You tell the marketing vice president that the additional space has been added, and they will have enough to finish their work. At 5:02 p.m. you head home.

Goodbye to old storage

It is Tuesday afternoon and you have just finished plugging in a new storage server. You add it to your storage cloud. With the addition of the new higher density storage, it is now time to retire the older, less dense storage. With a few clicks, your existing data is redistributed to the new storage and the old storage disks are emptied. On Thursday afternoon, after running the data migration process behind the scenes for a couple of days, your existing data and new data are now evenly distributed over the newly added storage, and the old storage is empty and ready for removal. Throughout the entire redistribution and removal, the application data remained online and available—keeping your employees productive and your customers happy.

Disaster recovery, cloud-style

You set up your storage cloud to maintain three copies of each file across three separate sites at all times. Good thing you did, because one afternoon a flood at the site that runs all of the IT equipment takes the entire site offline. This is when your cloud policies take effect. When the site goes down, client requests are automatically redirected to a secondary site. This secondary site now becomes the primary site. With a few clicks you tell the grid that the original primary site is permanently offline. At this point, the cloud quickly identifies space for the new third copy of the data and starts to populate that site with the data from the newly selected primary site, automatically restoring the original configuration. With your applications taken care of, you can focus your energy on drying out the flooded data center.

Private storage cloud architecture: An overview

Storage area network (SAN) technology introduced the ability to connect storage through a network, which was a great improvement to disks attached using small computer system interface (SCSI) or another type of direct connection. For the first time, the SAN allowed storage administrators to more easily, and with far less wiring, attach multiple hosts to shared storage or multiple storage servers. Although SAN was a great improvement over locally attached

devices within the SAN, there is still a direct relationship between a host and the associated storage. For example, if you need more space, the administrator could add a logical unit number (LUN), zone it to the host and expand the file system. This works well in cases where there is extra capacity on the SAN and the host operating system supports dynamically growing a file system.

A private storage cloud architecture takes this idea further by decoupling the server and the storage. When a host needs more space in a storage cloud infrastructure, the administrator clicks to assign more space to that host from a pool of available storage, and the application continues. The additional space is not simply whatever is available in the storage server attached to the host. Instead, it is the proper storage type for the application based on performance, availability and quality of service levels. When more of a particular type of storage is required, it is added to the cloud, assigned characteristics such as level of performance and reliability, and made available for use.

Many organizations have not implemented tiered storage yet because they lack a tightly integrated infrastructure between the tiers and high-performance tools to effectively manage the data. It was not always practical to provide three different classes of storage to each host or application due to SAN complexity and tool limitations. An effective private storage cloud solution makes managing tiers of storage a practical reality. To the host, the private storage cloud is accessed in a common manner through standard network file access protocols, regardless of the use of the storage. The data is available concurrently through multiple protocols. By providing access to pools of storage to a variety of applications, you can more efficiently utilize the storage available instead of fragmenting across hosts.

IBM and storage cloud technology

While many vendors offer storage cloud solutions in the market with promising levels of scalability and performance, often these solutions can be built on relatively new software and hardware combinations. In 2007, IBM introduced a second-generation storage cloud technology called IBM Storage Optimization and Integration Services – scale out file services (SOFS). Like SOFS, IBM Smart Business Storage Cloud, the next generation of storage cloud solutions from IBM are built on the proven reliability of IBM System x® servers and IBM TotalStorage® disk technologies paired with industry-leading software, such as the IBM General Parallel File System (GPFS) and IBM Tivoli® Storage Manager. These technologies have been integrated into a storage cloud solution based on years of internal use providing file services to IBM's more than 300,000 employees. The result is a flexible, standards-based and enterprise-ready storage cloud solution that can support advanced virtualization. Private storage cloud solution technology and services from IBM address multiple areas of functionality, including:

- Dynamic storage management
- Scalable capacity and performance
- Concurrent, multiprotocol data access
- New levels of manageability

The next few sections discuss these capabilities in greater detail.

Dynamic storage management

With support for multiple types of storage and storage connection mechanisms, the storage cloud solution from IBM enables you to mix storage technologies like Fibre Channel disks, serial attached SCSI expanders (SAS) and serial ATA (SATA) disk technologies. The ability to contain multiple pools of storage in a single namespace, however, is just the first step to implementing a storage pooling strategy. Making storage pools effective requires a mechanism to efficiently and dynamically migrate data from pool to pool with no additional overhead or impact to data access. File data in the storage cloud solution from IBM can be migrated from pool to pool automatically based on predefined policies.

Two main features make this storage solution uniquely capable of high-performance storage management: disk-to-disk migrations over the SAN, and extremely fast scanning of file metadata. Disk-to-disk migrations in the storage cloud solution from IBM are direct disk-to-disk data copies with virtually no change to the file in the user namespace. These disk-to-disk copies can take place over a SAN, InfiniBand or a TCP/IP connection to provide high performance with great flexibility. Migrations can be done online without interrupting data access. And data movement can be done slowly by a single node or very quickly by running migration operations in parallel by all nodes in the solution system.

Before any data is moved, the policies must be evaluated and candidate files identified for migration, deletion or change in replication status. With the ability to process more than 1 million files per second, the storage cloud solution from IBM processes rules and starts moving data very quickly, allowing you to apply a set of policies on a set of 1 billion files in about 15 minutes.

In addition, the private storage cloud solution from IBM allows you to grow a single namespace beyond the SAN to facilitate optimal flexibility. Typically there are limits with a SAN infrastructure to how many hosts can share a single disk array, and how many hosts can concurrently access a single LUN, for example. Growing beyond a SAN requires the ability to tie the namespace together using standard networking technologies, including TCP/IP and InfiniBand technologies. The private storage cloud solution from IBM utilizes a block-level network-based protocol (similar in concept to iSCSI), which allows you to tie multiple building blocks of nodes and storage together in a single file system over a TCP/IP or InfiniBand connection. This approach allows for great flexibility by enabling growth beyond a SAN. It also enables you to easily adopt new storage technologies as they become available and integrate them directly into your existing storage cloud solution.

Scalable capacity and performance

Successfully implementing a storage cloud architecture requires the ability to provide the required level of performance for the most demanding applications and scalability so that you are not forced into managing islands of data. Effective scalability of a storage cloud solution requires that you can efficiently mix front-end processing power with the data storage to supply the right balance for the application. Storage cloud scalability does not mean simply having the ability to place 1 petabyte of storage behind a pair of processing nodes or to spread data across separate appliances, creating islands of data, in an attempt to achieve the required level of performance. To be effective, a storage cloud solution must be able to support billions of files and multiple petabytes of data while dynamically providing room to grow and support future technologies.

The private storage cloud solution from IBM provides the capacity, performance scalability and flexibility required to effectively support a very large namespace and a high performance cloud entry point. A storage cloud solution from IBM today can support up to 512 billion files and hundreds of petabytes of data. With the ability to store billions files in a single managed namespace, the storage cloud solution from IBM allows the infrastructure to grow as needed. This growth can be in a single data center or across geographically distributed processing centers.

Concurrent, multiprotocol data access

So how is storage as a service actually provided today? A storage cloud solution provides multiprotocol access to a common set of data using multiple standard interfaces. This includes access from a mix of hosts to a set of data over network protocols including network file system (NFS) and common Internet file system (CIFS) while concurrently providing data access to SAN-attached hosts.

The private storage cloud solution from IBM provides standards-based network access to data by protocols, including CIFS and NFS access to a common set of data. A single set of data can be shared concurrently using multiple network protocols over 3 to 25 nodes or more. This provides the extreme scalability and reliability required to support a storage cloud solution.

In addition to the standard network protocols, the private storage cloud solution from IBM provides the ability to add Linux® nodes that are running your own applications. These nodes can access a common set of data directly over the SAN, if required. This allows high performance access to the data concurrently with client network access to the same data and for file management, backup or application integration. This level of integration provides a key component to implementing a successful cloud infrastructure.

New levels of manageability

The private storage cloud solution from IBM provides a single point of management through a Web-based administration tool. Through this management interface, you can administer clusters and monitor events across the entire solution, which are collected in a central event log. The administrator can define what event types are e-mailed, for example, to a set of administrators. You can also use the management tool to collect and view information on many solution dimensions and graph the results over time. For example, you can collect CPU or file system utilization and review it over the last month to determine whether to add more nodes or storage. The storage cloud solution from IBM provides an SNMP interface that integrates with other monitoring solutions to enable you to browse cluster information that can be monitored from many standard tools.

The private storage cloud solution from IBM provides for growing environments by allowing you to add and remove nodes and storage while the data remains available to the end users. Nodes can be added that are the same as current nodes or upgraded nodes. The solution supports a mix of node and connection types for maximum flexibility and future protection.

The private storage cloud solution from IBM also provides a new mode of space management. Typically administrators create dozens of small files system and apply these spaces to the users as space is requested—either because file system size is limited or there are insufficient tools to manage a large namespace for backup and hardware security module (HSM) operations. With the storage cloud solution from IBM, each storage pool can be multiple petabytes in size, and these pools can contain billions of files. This allows the system administrator to more

effectively use available storage through features such as over-provisioning, quota management and high performance reporting tools. In addition, you can free system administrators' time by delegating some space management operations to other people for project creation and share definition.

Getting up and running

Deciding to implement a private storage cloud solution is one thing. The actual implementation is another, particularly if you'd prefer not to divert IT staff from other activities. If you'd like help getting started—or just need help maintaining your solution—IBM offers a number of storage services designed to help you quickly and reliably implement, manage and maintain a scalable storage cloud system. Delivered by highly experienced professionals located at your site or through a network of global and regional delivery centers, IBM provides best practices and a single point of contact for asset, configuration, performance and problem call management.

The services support model from IBM uses a layered approach that can include:

- **Consulting**— evaluating criteria, information lifecycle management and archiving
- **Performance tuning** – troubleshooting and fixing performance issues related to existing environment or new application or user group deployments
- **Design and implementation** – solution design and sizing, deployment and provisioning
- **Ongoing cloud services** – a single point of contact and assistance during the operational phase as well as 24x7 monitoring and full maintenance, including daily housekeeping, upgrades and policy management
- **Optional cloud management services** – hosting, real-time hardware and software monitoring and management, Web portal interface for ticketing and cloud report presentment, and management of other hosted components such as tape libraries, servers and applications
- **Expert technical support** – access to extensive technology expertise and consultation for advanced support assistance on demand

Conclusion

Private storage cloud technology is rapidly advancing to the point where data is continually accessible at the required performance, with never-ending space to store it. As we edge closer to that reality, increasing numbers of organizations have recognized that a private storage cloud is the right fit for their organization.

Implementing a storage cloud solution can greatly optimize your storage infrastructure and increase efficiency to help reduce your storage TCO. As you plan your storage cloud implementation, you will want to start with a thorough assessment of your environment to make sure you fully optimize your investment. IBM is a leading provider of storage-related services that can help customers streamline their storage environments and reduce unnecessary overhead. IBM Smart Business Storage Cloud, along with IBM Storage Optimization and Integration Services – scale out file services, make an unbeatable combination to facilitate a successful implementation of your next-generation storage cloud infrastructure.



© Copyright IBM Corporation 2009

IBM Global Services
Route 100
Somers, NY 10589 U.S.A.

Produced in the United States of America
November 2009
All Rights Reserved

IBM, the IBM logo and ibm.com, System x, Tivoli and TotalStorage are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.