Transforming your Information Infrastructure with IBM’s Storage Cloud Solution
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Executive summary</td>
<td>3</td>
</tr>
<tr>
<td>Introduction</td>
<td>4</td>
</tr>
<tr>
<td>A Story or three for inspiration</td>
<td>6</td>
</tr>
<tr>
<td>Oops, Out of Disk Space</td>
<td>6</td>
</tr>
<tr>
<td>Goodbye to Old Storage</td>
<td>6</td>
</tr>
<tr>
<td>Disaster Recovery Cloud Style</td>
<td>7</td>
</tr>
<tr>
<td>Is Cloud Storage Right for You?</td>
<td>7</td>
</tr>
<tr>
<td>Storage Cloud Service</td>
<td>8</td>
</tr>
<tr>
<td>Implementing a Cloud Storage Solution</td>
<td>9</td>
</tr>
<tr>
<td>Dynamic Storage Management</td>
<td>9</td>
</tr>
<tr>
<td>Concurrent, Multiprotocol Data Access</td>
<td>11</td>
</tr>
<tr>
<td>New Levels of Manageability</td>
<td>12</td>
</tr>
<tr>
<td>Conclusions</td>
<td>14</td>
</tr>
</tbody>
</table>
Executive Summary

Reducing the total cost of ownership (TCO) for data storage is on the top of the list for IT managers today. Purchasing storage hardware is only the beginning and often not the most expensive part of the solution. It is not likely that an organization will be able to reduce the amount of data stored so to reduce the TCO for your storage infrastructure requires improving overall efficiently by reducing the daily maintenance costs, streamlining upgrades, and the ability to dynamically introduce new technologies to your existing infrastructure to continue to improve efficiency. Achieving storage operational efficiencies requires that you have the ability to mix server and storage technologies. To non-disruptively add and remove storage and move data as requirements fluctuate over time without downtime or weekend work for your IT staff. In addition to a mixed hardware and storage environment you need to provide multiple levels of service to meet various business needs for availability, efficiency or regulatory compliance requirements across your organization. To achieve these goals requires a storage technology that is flexible, scalable and easy to manage. This is where Cloud Storage comes into the picture. Cloud storage provides these features for the storage environment by presenting storage as a service. Storage cloud service means decoupling the node running the application from the application data residing on the storage.

Cloud storage today is provided by an evolving set of technologies and has many definitions across the industry. For example, a public implementation of cloud storage is a service that allows you to access the storage over the public internet. It can provide services like home PC backup-up over the internet. When you “sign up” for this type of service you agree to a service level that defines that the file will always be available and protected against disasters. In addition when signing up for such a service you don’t ask questions like “Is my data stored on SATA drives?” The Cloud storage approach allows the customer to focus on what is really important which is “will my data be there when I need it?” This is storage as a service. Today’s cloud storage solutions supply this type of service to more applications than just backing up a home PC. Cloud storage can provide data management services for mission critical applications including relational databases, web applications and scalable analytics.
This paper defines what is meant by the term “storage cloud service”, it provides insight as to what tools are available today to implement a storage solution and a glimpse of what you can look forward to in the not so distant future.

Introduction

Cloud storage is an emerging technology and an integral part of the future of data storage. IBM has introduced a cloud storage technology built on proven infrastructure components and offered as a solution. This cloud storage solution is called IBM Scale out File Services. IBM’s Cloud Storage Solution is a second generation cloud storage technology. The first generation was developed for internal use at IBM and is currently in use by thousands of IBM employees around the world storing millions of files. Lessons learned from this initial internal implementation were applied to IBM’s Cloud Storage Solution resulting in a solution that is quickly helping customers realize the benefits of cloud storage.

There are many vendors promising cloud storage solutions in the market that are built on brand new unproven software and hardware combinations. Many of these solutions are promising levels of scalability and performance that has never been tested. On the other hand, the IBM cloud storage solution is built on proven IBM technologies. This includes the proven reliability of IBM System x® Servers and TotalStorage® disk technologies paired with industry leading software including the IBM General Parallel File System (GPFS™) and Tivoli® Storage Manager. These technologies have been integrated into a solution developed based on years of internal use providing file services to IBM’s more than 300,000 employees provides an enterprise ready cloud storage solution.
Public Cloud vs. Private Cloud

There are two main scenarios for storage clouds that IBM customers can choose to pursue based on their business drivers and technical strategy. The two scenarios are defined as Public Storage Cloud and Private Storage Cloud, the key differentiators of these scenarios is that the public storage cloud is designed for customers who do not want to own, manage, or maintain the storage environment thus reducing their capital and operational expenditures cost around storage. The public storage cloud provides for variable billing options and shared tenancy of the storage cloud, giving customers the flexibility to manage the usage and growth of their storage needs. This is the industry standard view of a storage cloud offering and comparable to storage cloud offering by other vendors. The private cloud has fixed charges and dedicated tenancy, so it is designed for enterprise customers who want flexibility around ownership, management, and maintenance of the storage cloud.

Public Cloud:
Similar to a rent model, IBM dictates the choice of technology and cloud location, shared infrastructure with variable monthly charges, dynamic physical capacity at customer level and security measures to isolate customer data. In a public cloud IBM owns the physical assets, facilities, and standard contracts with multiple service level agreements (SLA) to meet specific needs. Public cloud solutions work well for cross industry solutions that storing from tens of Terabytes (TB) to multiple Petabytes PB of data.

Private Cloud:
Similar to a purchase or lease model, with a private cloud customers have the choice of technology and location on dedicated infrastructure with fixed monthly charges and physical capacity at the customer level. Each application can utilize dynamic capacity by sharing the cloud storage among multiple applications. Further the Private Cloud provides built-in security thru platform dedication, choice of asset ownership, and custom service level agreements (SLA) to meet specific needs. Private clouds work well for cross industry solutions storing tens of Terabytes (TB) to multiple Petabytes PB of data.
A Story or Three for Inspiration
To introduce the idea of cloud storage let’s take a look at a few situations organizations face each day and how a cloud storage environment can help.

Oops, Out of Disk Space
It’s Friday evening at 4:59 PM and the VP of Marketing is informed 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. The Marketing VP calls the IT support department to report the incident and inform them of the urgency of the issue. He is informed by the IT support department that the Marketing department has fully utilized their pool of storage. The IT representative quickly finds additional space on an existing storage pool that was just added the other day. The IT representative uses the cloud storage management interface to dynamically expand the storage allocation for the marketing department to include space in this new storage pool. The marketing VP is informed that the additional space has been added and they will have enough storage capacity to finish their work. At 5:02PM the situation is resolved.

Goodbye to Old Storage
It is Tuesday afternoon and your IT staff has just finished the plugging in a brand new storage server and it has been added to your storage cloud. With the addition of the new higher density storage it is now time to retire the older less efficient storage. With a few clicks existing data is being redistributed to the new storage and the old storage disks are being emptied. On Thursday afternoon, after running the data migration process “behind the scenes” for a couple days, your existing data and new data (added during the migration process) is now evenly distributed over the newly added storage and the old storage is empty and ready for removal. This was all done while the application data remained online.
Disaster Recovery Cloud Style
You have told your storage cloud that you need to maintain three copies of each file across three separate sites at all times. Good thing you did. One afternoon there is a flood at one of your sites destroying all of the IT equipment and quickly taking the entire site offline. This is when your cloud polices take effect. When the site goes down client requests for file data 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 goes to work populating that site with the data from the newly selected primary site automatically restoring the original configuration. Now that the applications are satisfied and running you can focus your energy on drying out the flooded data center.

Is Cloud Storage Right for You?
Imagine that your data is always available, always accessible at the required performance and you never run out of space. If this sounds good than a cloud storage solution may be a good fit in your organization. Although not all of these technologies exist today for the system administrator there are solutions that are well along the path to making them a reality. When reviewing possible cloud storage solutions you need to determine whether it provides:

- Dynamic Storage Management
- Scalable Capacity and Performance
- Concurrent, Multi-protocol Data Access
- Centralized Management
Storage Cloud Service

Storage Area Network (SAN) technology introduced storage connected using a network topology which was a great improvement to disks attached using 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 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 OS supports dynamically growing a file system. A cloud storage architecture takes this idea further by decoupling the server and the storage. In a cloud storage infrastructure when a host needs more space, 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 just whatever you had available in the local storage server attached to the host. It is the proper storage type for the application based on performance and availability with quality of service guarantees. When more of a particular type of storage is required it is added to the “cloud”, assigned characteristics like level of performance and reliability and made available for use.

Tiered storage is a great idea that many organizations have not implemented because they lack a tightly integrated infrastructure to manage data movement between the tiers and the high performance tools to effectively manage large amounts of data. It was not always practical, due to SAN complexity and tool limitations, to provide three different classes of storage to each host or application. An effective cloud storage solution makes managing tiers of storage a practical reality. 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.

To the host, the storage cloud is accessed in a common manner regardless of the use of the storage. Access is through standard network file access protocols and the data is available concurrently through multiple protocols as required.
Implementing a Cloud Storage Solution

Storage cloud service describes the end user or host view into the storage cloud. Storage administrators that are tasked with implementing a Cloud Storage solution require a set of tools that allow them to provide this end user experience. For the storage administrator storage clouds contain a mix of servers and storage types grouped together into pools. Each pool can contain disk, tape or other data storage technology. Each storage pool can be designed to provide predefined levels of performance and availability.

There are multiple areas of functionality a solution must provide to successfully implement a cloud storage infrastructure. IBM Scale out File Services (IBM’s Cloud Storage Solution) provides multiple key features that enable the storage administrator to effectively manage a cloud storage infrastructure.

Dynamic Storage Management

IBM’s Cloud Storage Solution provides storage pooling with a tightly integrated policy based storage management system. Cloud storage needs to support multiple types of storage and storage connection mechanisms. With IBM’s Cloud Storage Solution you have the ability to mix storage technologies like Fibre Channel disks, SAS and SATA disk technologies.

The ability to contain multiple pools of storage in a single namespace is just the first step to implementing a storage pooling strategy. To make the use of storage pools effective requires a mechanism to efficiently and dynamically migrate data from pool to pool with minimal overhead or impact to data access.

In IBM’s Cloud Storage Solution file data can be migrated from pool to pool automatically based on policies. There are two main features of this solution that makes IBM’s Cloud 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 IBM’s Cloud Storage Solution are direct disk to disk data copies with 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 providing high performance with great flexibility. These migrations can be done online without interruption in data access. The data movement can be done slowly by a single node or very quickly by running migration operations in parallel by all nodes in the IBM’s Cloud Storage Solution.

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 IBM’s Cloud Storage Solution processes rules and starts moving data very quickly. This allows you to apply a set of policies on a set of 1 billion files in about 15 minutes.

To be flexible a Cloud storage solution needs to scale beyond the SAN. There are limits with a SAN infrastructure as to how many hosts can share a single disk array and 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. The IBM Cloud Storage Solution allows you to grow a single namespace beyond the SAN. To achieve this it utilizes a block level network based protocol. This block level network protocol (similar in concept to iSCSI) 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 allows for great flexibility by enabling growth beyond a SAN and the ability to easily adopt new storage technologies as they become available and integrate them directly into your existing Cloud Storage Solution.
Extremely scalable capacity and performance keeps you from getting stuck. Successfully implementing a cloud storage 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. There are some cloud storage offerings that were introduced with some interesting technologies but had limited applicability because of their inability to scale at a single “entry point” into the cloud. Effective scalability of a cloud storage solution requires that you can efficiently match front end processing power with the data storage to supply the right balance for the application. Cloud storage scalability does not mean simply having the ability to place 1 PB 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 cloud storage solution must be able to support billions of files, multiple petabytes of data while dynamically providing room to grow and support for future technologies.

IBM’s Cloud Storage Solution provides the capacity and performance scalability and the flexibility required to effectively support a very large namespace and a high performance cloud “entry point”. The IBM Cloud Storage Solution 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 IBM Cloud Storage Solution allows the infrastructure to grow as needed. This growth can be within a single data center or across geographically distributed processing centers.

Being able to create large global namespace is one issue, managing it is another. IBM’s Cloud Storage Solution provides many tools that make managing a large file system practical.

**Concurrent, Multiprotocol Data Access**

So how is storage provided as a service today? A cloud storage solution provides storage services through multi-protocol access to a common set of data using multiple standard interfaces. This includes a mix of hosts accessing a set of data over network protocols including NFS and CIFS while concurrently providing data access to SAN attached hosts.
IBM’s Cloud Storage Solution provides standards based network access to data using multiple protocols including CIFS and NFS 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 Cloud storage solution.

In addition to the standard network protocols IBM’s Cloud Storage Solution provides the ability to add Linux®, AIX® or Windows® nodes running your own applications. These nodes can access a common set of data directly over the SAN, if required along with the standard network file protocols. This allows high performance access to the data concurrently with client network access to the same data. This allows you to have high performance access to the data 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**

Successfully implementing a cloud storage solution requires new levels of manageability. A large cloud storage solution needs to be manageable by a small team of system administrators.

To fulfill the administration requirement IBM’s Cloud Storage Solution provides a single point of administration through a web based tool. All of your IBM’s Cloud Storage Solution clusters can be administered from a single interface.

Using this management interface you can monitor events across the entire solution. All of the events are collected in a central event log. The administrator can define what event types are emailed, for example, to a set of administrators. The web based management tool can be used to collect and view information on multiple aspects of the solution and graph the results over time. For example you can collect CPU
utilization or file system utilization and look at it over the last month to determine whether you need to add more nodes or storage. For integration with other monitoring solutions IBM's Cloud Storage Solution provides an SNMP interface that allows for browsing of cluster information and a set of traps that can be monitored from many standard tools.

To enable a large global namespace, IBM's Cloud Storage Solution allows you to dynamically grow the environment 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.

A new paradigm of storage management is possible with IBM's Cloud Storage Solution provides. Today, administrators typically create dozens of small file systems and assign these spaces to the application and end users as space is requested. With other solutions this is typically done because the file system size is limited or there are insufficient tools to manage a large namespace for backup and policy based tiered storage operations. With IBM's Cloud Storage Solution 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 utilize available storage using features including over-provisioning, quota management and high performance reporting tools. In addition you can free up system administrators time by delegating some space management tasks to other people for project creation and share definition.

Security is essential in a Cloud storage environment. IBM's Cloud Storage Solution provides multiple methods of securing the data and integrating with existing environments. For example IBM's Cloud Storage Solution can participate in a Microsoft® Active Directory domain.
Conclusions
Implementing a cloud storage solution can optimize your storage infrastructure providing improved availability, flexibility and increased efficiency which can greatly reduce your storage total cost of ownership (TCO). To optimize your investment when planning your cloud storage implementation you should start with a thorough assessment of your environment. A recent Gartner publication titled Invest in Storage Professional Services, Not More Hardware concluded that “By improving storage utilization, IT departments can delay incremental storage acquisitions.” IBM is a leading provider of storage related services that can help customers streamline their storage environment and reduce unnecessary overhead.

The IBM Cloud Storage Solution along with the experience of IBM IT services makes an unbeatable combination and helps to ensure a successful implementation of your next generation cloud storage infrastructure. With IBM IT services and the IBM Scale out File Services Cloud Storage Solution you can start implementing the future of storage today.
For More Information
To learn more about the IBM Storage Systems and Cloud Computing, please contact your IBM marketing representative or IBM Business Partner, or visit the following Web site: http://www.ibm.com/ibm/cloud/

About the Author
Scott Fadden
GPFS/SoFS Technical marketing

Scott has more than 15 years experience in developing, designing and implementing complex IT solutions.