ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

Implementing Storage as a Service

[I]Varun Singh, [II]Ishan Srivastava, [III]Vishal Johri
[I,II,III]Dept. of CSE, Manipal Institute of Technology, Manipal, India

Abstract
In today's fast-paced and competitive environment, organizations must be agile and flexible to meet changing market requirements. This leads to rapid expansion and upgrade of resources while meeting stagnant IT budgets. Data storage is one such resource. Storage of data in an efficient and manageable way is of prime importance for majority of IT organizations. IT organizations have adopted storage technologies which have developed organically over time. This has resulted in a heterogeneous environment (both in terms of storage and backup technology) with isolated silos of data. These diverse tiers of storage and multiple backup technologies are either out of vendor support or nearing their end of service life. Current storage infrastructure deployment does not provide the ability to move data between tiers and across various data centres.There are similar issues present in the other technology domain as well. In order to improve the operations and reduce the cost, IT organizations have to address each of the technology areas. For the purpose of this paper, we have chosen storage as the first technology domain to transform into the new operating model. Our analysis is done for a mid-sized IT organization. We have reviewed the current state of the organization's storage by looking at the industry's best practises, technologies etc. and doing the gap analysis between the current state and future state. Also we have provided the implementation and experimental details in this paper.

Keywords
Storage as a Service, Cloud Computing, Software defined Storage, Data centre operations

I. Introduction

The importance, dependency, and volume of information for the business world continue to grow at astounding rates. Businesses depend on fast and reliable access to information critical to their success. Examples of business processes or systems that rely on digital information include airline reservations, telecommunications billing, internet commerce, electronic banking, credit card transaction processing, capital/stock trading, health care claims processing, life science research and so on. The increasing dependence of businesses on information has amplified the challenges in storing, protecting, and managing data. Legal, regulatory, and contractual obligations regarding the availability and protection of data further add to these challenges.

Data created by individuals or businesses must be stored so that it is easily accessible for further processing. In a computing environment, devices designed for storing data are termed storage devices or simply storage. The type of storage used varies based on the type of data and the rate at which it is created and used. Devices, such as a media card in a cell phone or digital camera, DVDs, CD-ROMs, and disk drives in personal computers are examples of storage devices. Businesses have several options available for storing data, including internal hard disks, external disk arrays, and tapes.

Historically, organizations had centralized computers (mainframes) and information storage devices (tape reels and disk packs) in their data centre. The evolution of open systems, their affordability, and ease of deployment made it possible for business units/departments to have their own servers and storage. In earlier implementations of open systems, the storage was typically internal to the server. These storage devices could not be shared with any other servers. This approach is referred to server-centric storage architecture. In this architecture, each server has a limited number of storage devices, and any administrative tasks, such as maintenance of the server or increasing storage capacity, might result in unavailability of information. The proliferation of departmental servers in an enterprise resulted in unprotected, unmanaged, fragmented islands of information and increased capital and operating expenses.

To overcome these challenges, storage evolved from server-centric to information-centric architecture. In this architecture, storage devices are managed centrally and independent of servers. These centrally-managed storage devices are shared with multiple servers. When a new server is deployed in the environment, storage is assigned from the same shared storage devices to thatserver. The capacity of theshared storage can be increased dynamically by adding more storage devices without impacting information availability. In this architecture, information management is easier and cost-effective. Storage technology and architecture continue to evolve, which enables organizations to consolidate, protect, optimize, and leverage their data to achieve the highest return on information assets.

In this paper, we have analysed the storage details of a mid-sized IT organization. Then based on company's need, we have designed and implemented the storage as a service for the organization.

II. Background and Overview

Our approach to the problem was to assess the current state, create a vision by looking at the industry's best practises; technologies etc. and doing the gap analysis between the current state and future state. We have also concentrated on allocating project to bridge the gap and transform the storage landscape into the future vision state.

There were three key initiatives under the storage transformation umbrella, and they were:

- Current state assessment
- Develop storage strategy and road map
- Implementation

A. Current state assessment

The current state assessment includes gathering details of current storage landscape and reviewing the same. The key challenge was due to the operating model, as the infrastructure services were outsourced to a third party and the asset database was not really up to date. In order to get the asset database up to date, a third party tool set was introduced. "HiSAT" tool from Hitachi Data System (HDS) was chosen as the tool to gather the storage details.

HiSAT

The tool was selected due to the ease of integration, as this was an out of the band tool, without having a requirement for an agent on each of theservers attached to storage. Hitachi engineers will install the application on to a laptop/desktop and plug them on to the data centre network. The HiSAT tool will scan the network and discover hosts and get the host, storage details. This provides not only the storage details but also the server inventory as well. HiSat stores all the data in SQL database, since the tool is a licensed product, the license was valid for only the duration of data gathering. HDS has provided a locked down version (the data gathering is disabled) of this tool to access the data and generate various reports. Some of these reports includedetails of all hosts discovers along with H/W, OS, Application, storage (both internal, external) etc.

HiSAT report provided a great level of information on the current storage landscape. The next step was to analyse the data,and outline the various technologies along with the products used in the storage landscape such as primary storage, data replication, backup, tape devices etc. This HiSAT report has confirmed that there were SAN islands created and there were no interconnectivity between these islands, this lead into sub-optimal utilisation of storage space. The key issues highlighted in the reports were

- Sub Optimal storage utilisation
- No connectivity between the SAN islands
- Lack of management tools
- 30% storage landscape has reached end of life and another 40 will reach end of life in another 12-18 months

Overall the report revealed that the storage landscape was under major threat and could impose a major risk in the overall application availability, thus having an impact on the business.

B. Develop storage strategy and roadmap

The current state assessment revealed the storage landscape was under major threat, imposing operational risks, and increased TCO. The first of storage strategy was to create a future state vision, which enables the storage tower to overcome the limitation of the current storage landscape.

Considering the organization's complex operating model including various lines of business (both regulated and non-regulated business), the vision statement was to "provide storage as a service" to the other IS systems and business. The main requirements of the strategy to implement storage as a service are:

- Global - Employ a global storage strategy. Develop a global storage Centre of Excellence with consistent methods, tools, policies, processes, investment plan and technical standards that support the company's "growth through acquisition" model.
- Aligned - Alignment of service levels with data value. A storage architecture that provides tiered classes of service based on business driven requirements for reliability, availability, scalability and performance
- Low TCO - Enduring support model with minimum TCO. Centralized monitoring and management of storage assets to make efficient use of labour and technical resources as well as vendor relationships.
- Predictable Costs - To be able to accurately estimate storage costs based on business requirements. To enable IT projects to accurately predict their storage costs by providing a catalogue of tiered storage options.
- Service Oriented - Storage as a service, deliver storage in an easy-to-request, packaged methodology to speed provisioning and meet service levels.
- Adaptive - A highly adaptive storage environment maximizes vendor independence, insulates applications from storage asset changes and increases responsiveness to changing business needs.
- Evergreen - An investment plan that includes technology upgrade and asset refresh.

The next step was to outline a strategy considering the above vision. This strategy should outline a series of projects necessary to bridge the gap between the storage services vision and where we are now. It looks at people, process and technology to describe the changes that are needed.

- People – The Storage Management Services team, its roles and responsibilities
- Process – The services and levels of service provided by the Storage Management Services Team and the processes used within the organization to provide those services at an acceptable cost.
- Technology – The software, hardware and applicable standards that are used by the Storage Services Management Team to execute these processes.

The following steps were needed to achieve the vision described above:

- Centralized Storage Management Responsibility - Centralize storage management to ensure the most cost effective and consistent approach across the enterprise. For efficiently manage storage infrastructure assets, a Storage Centre of Excellence need to be developed.
- Common Storage Infrastructure – Standardize the base technology infrastructure to enable delivery of adaptive storage services at a low predictable total cost. Define a financial plan for the storage environment that enables asset refresh to control costs.
- Virtualisation - Implement these storage services in a way that decouples applications from storage assets creating an adaptive storage environment.
- Tools & Processes - Equip the global storage management function with the tools necessary to efficiently monitor and control storage assets in a way that continuously lowers TCO.
- Storage Services – Define a standard portfolio of services at various cost points to be used by the business based on their requirements. Establish storage policies that reflect storage principles that are consistent with the identified storage service classes. Design the infrastructure and operational processes necessary to deliver these services. Minimize wasted resources and complexity by managing data over its entire lifecycle in a way that recognizes its changing business value. Ensure that placement of data is commensurate with the business value of the data at all points in its life cycle.

III. Implementation

A storage strategy has been defined, and it was time for action, the technology framework needs to be put in place and the service needs to be defined. The first step was to prepare the technology framework.

The new solution needs to be deployed and the existing data/application needs to be migrated. The success factor of this stage

Vol. 2, Issue 3 (July - Sept. 2014)
ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

depended on the overall storage system architecture. Some of the key challenges during the design were,

- Appropriate technology selection
- Data migration with very minimal or no impact on the overlaying applications

Solution design

The earlier stages captured/set the requirements, and the solution design stage analysed the requirements and initially created a conceptual design. A conceptual design was created based on these requirements and it was used to do the analysis to choose the right technology/vendor.

A logical design was created based on the conceptual design and the technologies/products identified.

Fig. 1 depicts the logical view of proposed solution. The details of each module as follows.

Storage Array Layer

Storage Array layerconsists of disk devices, where the data resides. Different tiers of storage together form the storage array layer. The disks (storage arrays) are classified as Tier1, Tier2, Tier3 based on cost, performance redundancy etc. An example of Tier1 storage is an enterprise class storage array, which offers the best performance, and redundancy, obviously the cost will also be in the higher range. The tier 2 storage is better performance and redundancy and moderate cost. Tier three storage is made of cheaper disks, and it's used where performance is not a criteria. Products used in various tiers are shown in Table 1.

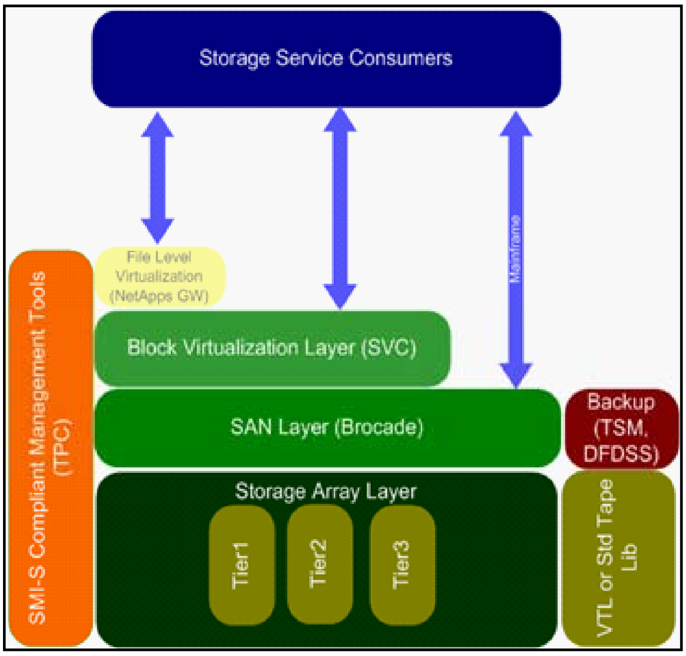

Fig. 1: Logical view of proposed solution

Table 1: Tier three storage

Tier	Product
Tier 1	IBM DS 8000, EMC Symmetrix
Tier 2	IBM DS4800, EMC Clarion
Tier 3	IBM DS4000, EMC Clarion/Celerra (SATA storage)

Table 1 shows some of the storage arrays used in various tiers, however, the usage of virtualisation made the disks as a commodity, which means there is no more locking on a particular vendor/ product. More details on this in the virtualisation section.

SAN Layer

It is the storage network layer, which interconnects the hosts and the storage. SAN layer has fibre channel predominantly, however, there is a small footprint of IP based storage using 1000BT Ethernet.

The fibre channel SAN layer is designed based on Core- Edge design, where edge switches will be connected to director acts as the core of the network. This design provides an optimal performance/scalability. Brocade switches are used for SAN layer, details are as follows:

Table 2: San Switches

Functionality	Product
Core	Brocade 48000 Director
Edge	Brocade 4100, 200E

The core switch has router blades, which allows interconnecting different SAN either using Fibre channel or using IP. This allowskeeping SAN within the data centre, but provides the data transfer between two SANs.

Virtualisation layer:

The virtualisationlayerseparates the application from the physical storage. Virtualisation is implemented between the storage and the host. Virtualisation of storage helps achieve location independence by abstracting the physical location of the data. The Virtualisation system presents to the user a logical space for data storage and itself handles the process of mapping it to the actual physical location. The advantages of storage virtualisation are,

- Improved storage utilisation
- Non-disruptive data migration between storage tiers or storage arrays
- Reduced management overheads, i.e. reducing the number of technologies
- Disks (Storage arrays) are commodity, no locking on to a particular vendor

IBM SAN Volume Controller (SVC) was chosen as the virtualisation appliance, as the underlying technology is mature and proven in enterprise. SVC is an in-band virtualisation appliance, which provides block level virtualisation.

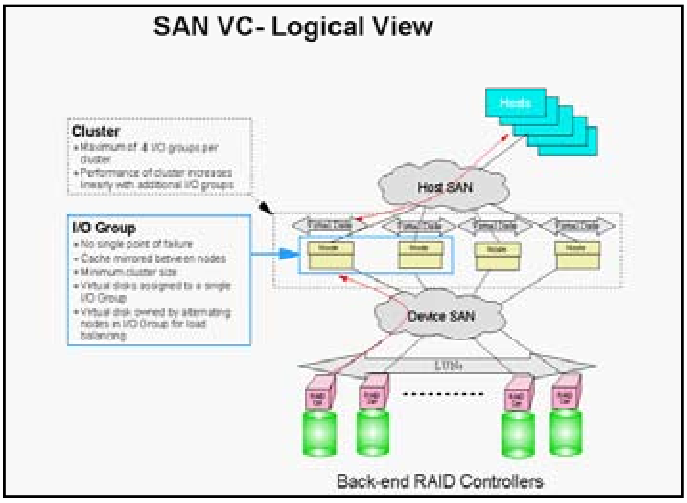

Fig. 2 : Logical view of SVC

Fig. 2 depicts the local view of SVC, disks (LUNs) from the storage array are presented to the SVC, and these disks are labelled as managed disk or MDisk. Local volumes are created on these managed disk and are presented to hosts are VDisk (Virtual disk). Fig. 3 shows the relation between MDisk and VDisk.
The SVC nodes are configured in a cluster to provide availability and load balancing; maximum of eight nodes (4 IO groups) can be configured in a cluster.

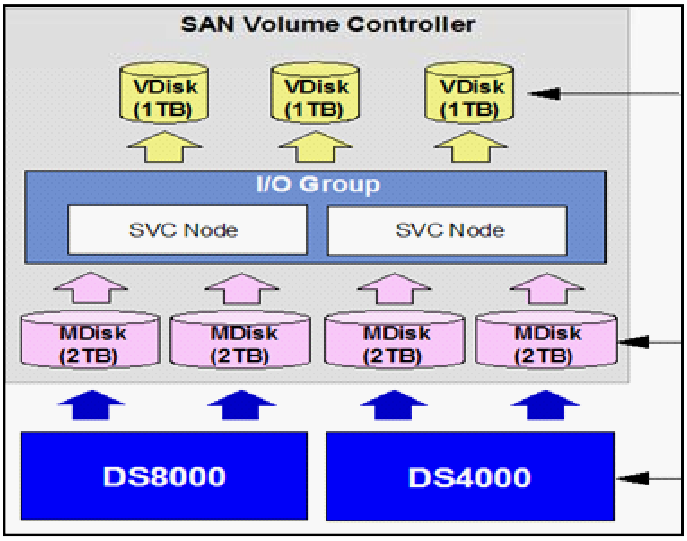

Fig. 3: the relation between MDisk and VDisk

The other advantage of SVC is the data replication used for HA can be brought in to the SVC layer; hence the replication is no longer locked down to a vendor specific solution, and avoids the symmetric configuration of storage arrays at both ends. IBM Flash copy provides a local replication and Metro mirror and Global mirror provides inter site data replication.

File level virtualisation

NetApp gateway is used for file level virtualisation; this allows the storage access via SVC using the standard building block. This design allows bringing in all the benefits of virtualised storage on top of file virtualisation. The primary usage of NAS gateway is for file serving purpose for shared areas or user's home drives.

Backup & Recovery

Tivoli Storage manager is identified as the backup tool due to the advance features such as "progressive incremental" backup, hierarchical storage management, wider support of agents including database, share point exchange etc. Also it provides an easy migration path from the older version of TSM, which was originally used. Virtual tape libraries (VTL) are used as the physical media. The VTL uses cheaper SATA disks and configure virtual tape drives and media. The backup software views the virtual drives hence there is no change in the way the backup software operates. The solution also lookat the data replication between the VTLs using TSM, which provides an offsite copy of data that can be used during a DR scenario.

Management layer

The end to end storage device management requires a tool/ tool set. IBM Totalstorage Productivity Centre (TPC) is used for management. TPC has different modules; tied together they provide the end to end storage management capabilities. The TPC models uses are shown in Table 3.

Table 3: TPC Model

Module	Area
TPC for data	Data usage from host
TPC for fabric	SAN fabric
TPC for disk	Storage Array

The technology framework is in place, and the next challenge is how the services are defined. Any service will have certain criteria, which it should meet. The usageof each of these services are different. The services are defined based on the three major criteria, performance, availability, and cost. Further analysis revealed that each service should have the following categories,

- Response time
- Availability
- Recovery time
- Data retention

Based on the categories five services were defined, each having a specific use case scenario as shown in Fig. 4 and they are:

- Mission Critical
- Business Critical
- Standard service
- Reference service
- Archive service

IV. Conclusion

Providing Storage as a service is extremely scalable and easy to manage. Based on the implementation provided in the paper,a service catalogue was created with the service details and an indicative cost. Any new project/application requiring storage will use this catalogue for their storage requirement. Storage as a service will help the organizations in reducing the RTB, greener data centres, cost transparency and storage as a commodity.

References

[1] *Jiyi WU1,2, Lingdi PING1, Xiaoping GE3,Ya Wang4, Jianqing FU1, 2010 International Conference on Intelligent Computing and Cognitive Informatics, "Cloud Storage as the Infrastructure of Cloud Computing"*
[2] *Storage Networking Industry Association.Cloud Storage Reference Model, Jun.2009*
[3] *Storage Networking Industry Association.Cloud Storage for Cloud Computing, Jun.2009*
[4] *C. Cachin, R. Haas and M. Vukolic, "Dependable storage in the Intercloud", Research Report RZ, 3783, 2010.*
[5] *H. Abu-Libdeh, L. Princehouse and H. Weatherspoon, "RACS: a case for cloud storage diversity", SoCC'10:Proc. 1st ACM symposium on Cloud computing, 2010, pp. 229-240*
[6] *E. Grosse, J. Howie, J. Ransome, J. Reavis and S. Schmidt, "Cloud computing roundtable", IEEE Security & Privacy, 8(6), 2010, pp. 17-23.*

Vol. 2, Issue 3 (July - Sept. 2014)
ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

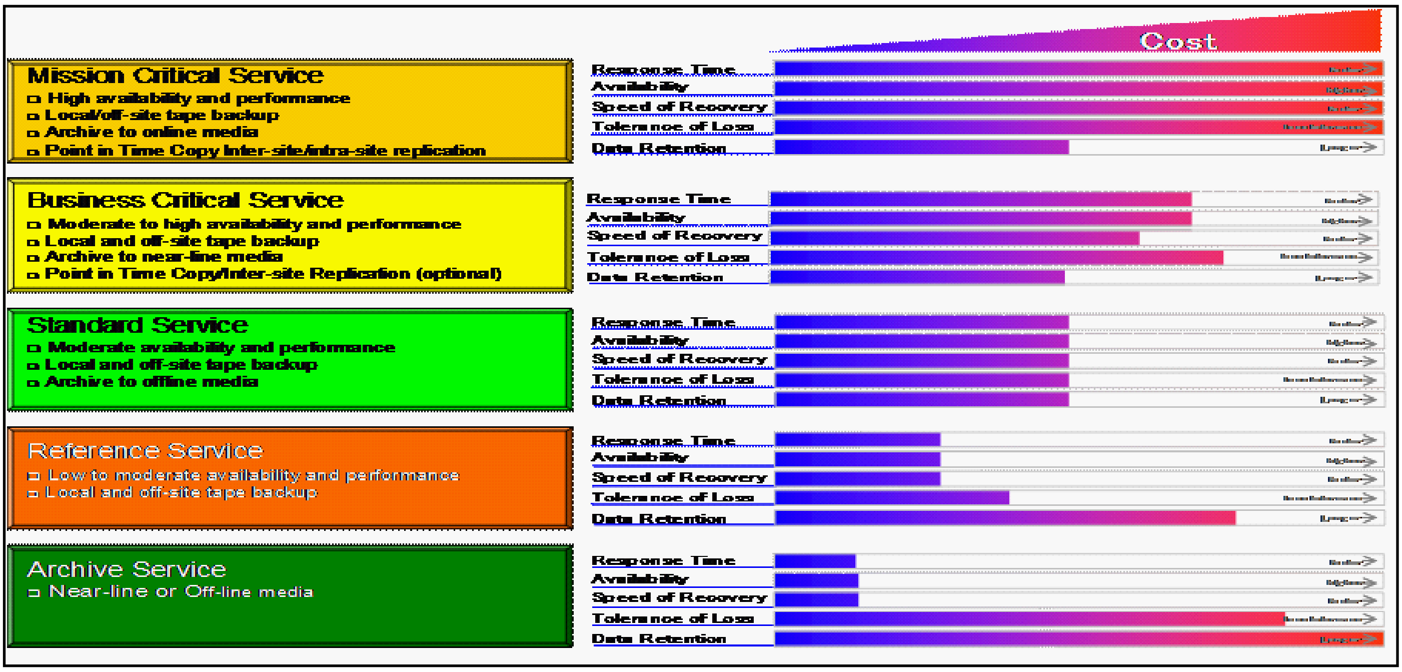

Fig. 4: Services and corresponding cost.