A REVIEW PAPER ON STORAGE VIRTUALIZATION

Pradeep Nayak*[1], Kelvin Dmello*[2], Karthik Madakari T P*[3], Jahnavi*[4], Harshitha B*[5]

Alvas Institute of Engineering and Technology, Mijar, Karnataka, India- 574225.

Department of Information Science and Engineering.

ABSTRACT

Now days the trends is about the storage virtualization where storage virtualization helps in the data storage, memory management and security helps the companies to secure the information in the storage and you can backup it in the cloud it is best example for the storage virtualization and this review paper is all about how to make the storage virtualization more effective to the user or to the companies without the loosing the performance and to made inexpensive.

INTRODUCTION

WHAT IS STORAGE VIRTUALIZATION?

It involves giving hosts a logical representation of physical storage resources. Physical storage that is directly attached to the host appears and functions like logical storage. Storage virtualization advantages include a greater use of storage, Unaffected storage expansion or deletion.

HOW DOES STORAGE VIRTUALIZATION WORKS?

Physical storage hardware is replicated in a virtual volume during storage virtualization. A single server is used to join together several physical discs in order to provide a straightforward virtual storage system. The same server is assigned virtual storage or logical storage blocks, which assist in rerouting input/output (I/O) traffic.A virtualization layer that separates the physical discs from the virtual volume allows operating systems and programmes to access and utilise the storage. I/O requests are processed by virtual storage software, which then sends them to the relevant storage devices across the entire storage pool.Small data blocks or objects known as logical unit numbers (LUNs), logical volumes (LVs), or RAID groups are used to describe the physical discs itself. These chunks are displayed as a virtual disc to distant servers. Instead of the assortment of storage devices that make up the overall pool of storage in the virtualized environment, they appear to the server just like a physical disc.

VIRTUALIZATION

A clear and uniform Interface to complex functions is provided via virtualization. Understanding the complexity at the foundation is not necessary. For instance, a driver of an automobile doesn’t need to comprehend how the internal combustion engine or the accelerator pedal operate. Makes that process virtual. A method of abstracting physical resources into logical views is virtualization. It increases the capacity and utilisation of IT resources. Significantly reduces downtime and streamlines resource management by sharing and pooling resources. Better IT resource performance is the result.

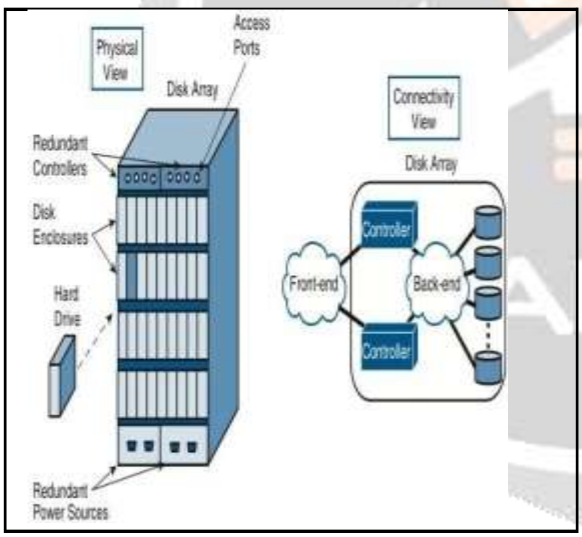
TYPES OF VIRTUALIZATION

Virtualization can take many different forms and is categorised as follows:

a. Virtual memory, which is separate from physical memory and is seen by each application,
b. Virtual Network - Each programme sees its own logical network, separate from the real network, in virtual networks.
c. Virtual servers - Virtual servers, which are separate from real servers, allow each programme to view its own logical server.
d. Virtual Storage – Independent of physical storage, each programme sees its own logical storage.

DISK ARRAYS

Computers in data centres lack internal disc storage. They're all located outside the server in a disc array. So that in the event of a server failure, Data are quickly accessible by other servers. Data centres do not offer JBODs (Just a bunch of discs), as they are challenging to operate. Disk arrays are a pool of redundant discs that are simple to manage. This is a representation of the disc array's components and structure.



COMPONENTS OF DISK ARRAY DATA ACCESS METHODS AND THEIR STRUCTURE

Applications can access data in three different ways:

1. Block Access: Block size is a predetermined number of bytes, such as 1, 4, or 16 sectors.

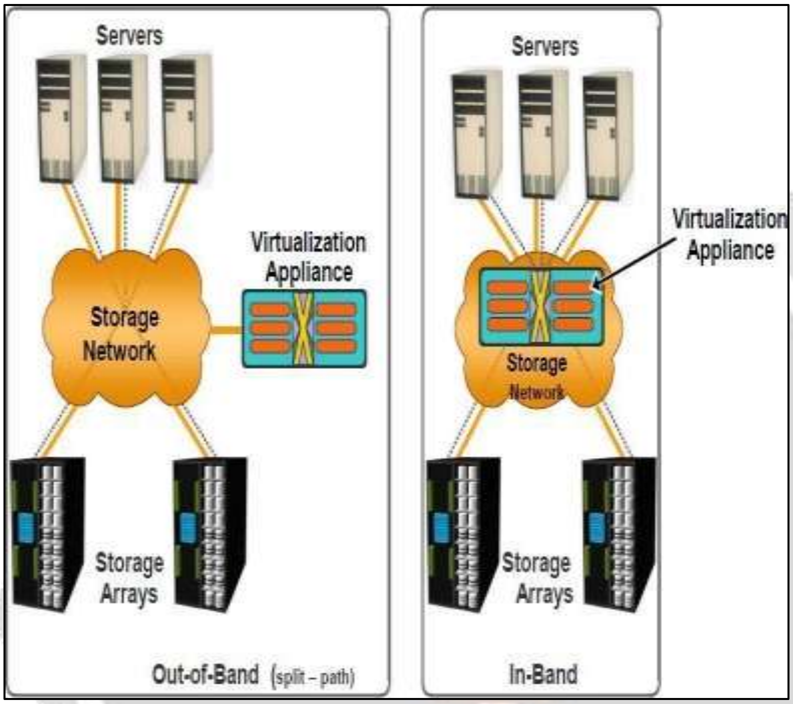
2. File Access: A collection of bytes containing the name, creation date, and other meta information. Could be contiguous or not. The way that meta-data is saved and files are organised is determined by a file system, such as FAT-32 (File Allocation Table) or NTFS (New Technology File System). Operating systems have different file systems.

3. Record Access: Used in databases for highly structured data. A certain format and collection of fields are used for each record. Accessed through Java Data Base Connection, Open Data Base Connectivity (ODBC), and Structured Query Language (SQL) (JDBC). Storage

CONFIGURATION OF STORAGE VIRTUALIZATION

Out-of-band deployment: The configuration of the virtualized environment is kept outside of the data flow.

-Virtualization appliances are hardware-based and fibre channel-optimized.



-Permits processing of data at network speed

-More scalability

Implementation in-band:- Virtualization is implemented in the data path

-Virtualization appliances run on general-purpose servers and are software-based.

-Data passing and storing through the appliance during processing adds to the latency.

-Lesser scalability; only appropriate in static environments with steady demands.

CHALLENGES FACED BY STORAGE VIRTUALIZATION

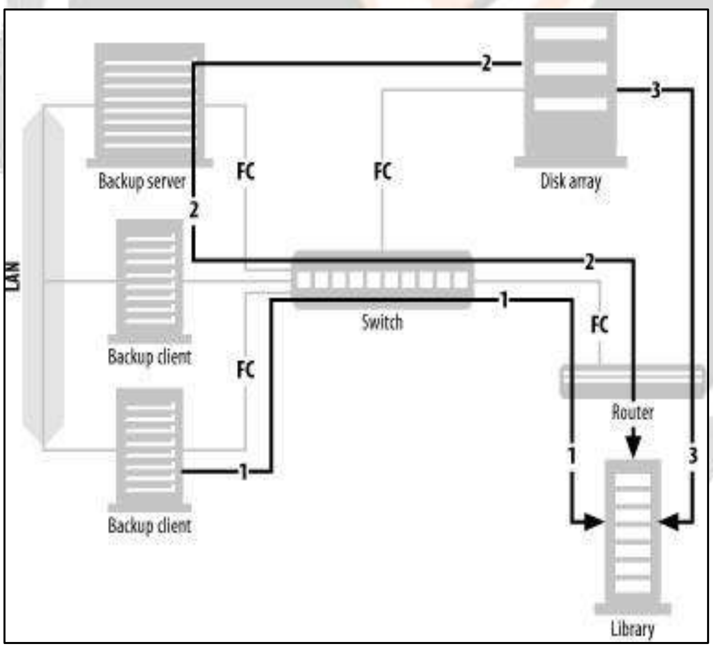
1. Scalability – In the absence of virtualization, each storage array is run separately to meet IOPS and capacity requirements for an application. Analysis of the environment as a whole is necessary for virtualization.
2. Functionality: A virtual environment must have the same or superior functionality. Must keep utilising arrays' current functionality
3. Manageability – A virtualization device disrupts a full view of the storage infrastructure. Integration with current management tools is necessary
4. Support – Multivendor Environment Interoperability

A REPRESENTATION OF ALL THE SAN ELEMENTS

Several servers, online storage (disc), and offline storage (tape or optical), all of which are coupled to a Fibre Channel switch or hub—typically a switch—make up a SAN. Any server in the SAN can be given complete read/write access to any disc or tape drive inside the SAN once the three servers in the preceding diagram are linked to it. This enables LAN-free, client-free, and server-free backups, each of which is denoted in the picture above by an arrow with a different number.

LAN-free backup: When numerous servers share a single tape library, LAN-free backups take place. Backing up to cello tape drives that a server connected to the SAN believes are locally attached is possible. SCSI-3 is used to transmit the data through the SAN

Client-free backups: Although a single computer is frequently referred to as a server, the backup system refers to it as a client. Data from a client can be backed up using SAN if the client's disc storage can create a mirror that can be split off and made visible to the backup server.



Server-free backup: The backup server; the backup client is never used to transfer data. This is known as client-free backup as a result. Client-free backups are depicted by arrow number 2, which depicts a data path from the disc array to the shared tape library after passing through the backup server, SAN switch, and router.

RECOVERY AND BACKUP FOR SAN

When it comes to backup and disaster recovery, SANs are especially useful. Data can be moved from one storage device to another inside a SAN without involving a server. This expedites backup and does away with the requirement to utilise server CPU cycles for backup. also numerous. SANs make use of Fibre Channel technology or other networking protocols to expand the geographic reach of their networks. This makes it more practical for businesses to store their backup data in far-off places.

WHY STORAGE VIRTUALIZATION REQUIRED

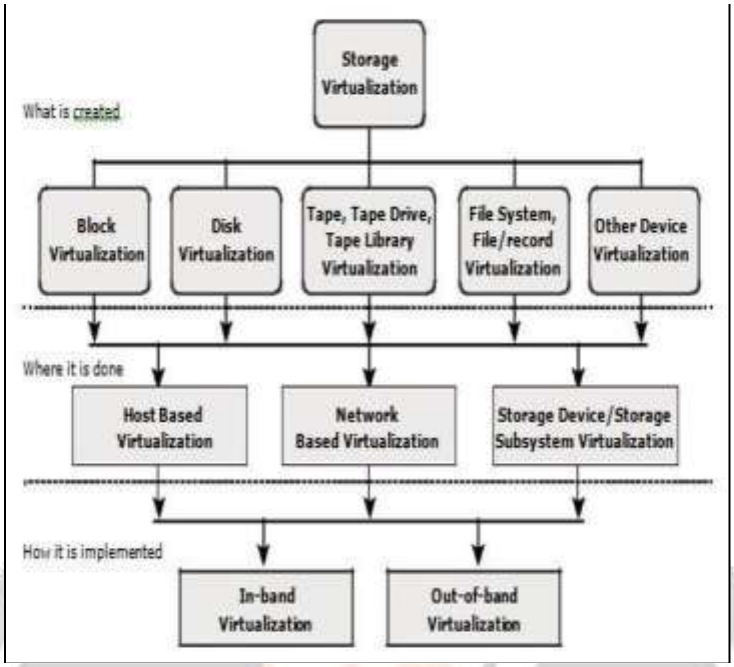
Regardless of the physical configuration of the storage space, storage virtualization is crucial for performing the necessary activities to divide the available storage space into virtual volumes. The real storage elements that are on hand. (For instance, disc drives and RAID tiers).Large amounts of storage data must be managed in Storage Area Networks (SAN) in a consistent manner and from a central location. With the majority of company installation activities, storage capacity and processing power are expanding quickly. Its combination with the need for high availability and routine activities necessitates that the storage design permit them to seamlessly expand storage without experiencing any interruptions. The virtualization of the storage allows for the greatest possible achievement of these strategic objectives. Because of its expansion potential, storage virtualization is mostly required. While managing the necessary tasks, the significance of this talent can be activities of online stores. It makes it possible for data to move around and be mobile. Moreover, Storage Virtualization enables the use of software applications rather than human administrators to carry out the various tasks and allocations. Storage virtualization tools such as Snapshot Remote Mirroring and Virtual Tapes play a key role in data management. For use, it needs to create and grow its virtual volumes. To sum up, Storage Virtualization is a must for a SAN. So, it is essential and most required in a storage area network.

DIFFERENT TYPES OF STORAGE VIRTUALIZATION

a. Storage Area Network - A network that gives us access to shared data storage is referred to as a storage area network (SAN). SANs were initially employed to enhance storage. The servers may access resource devices such drives, tapes, and optical discs so that these support devices can connect locally to the operating system. Other support equipment cannot access the SAN's separate network of diverse storage resource devices via the LAN.
b. Network Attached Storage - A data storage server for network-attached storage (NAS) is connected to a network that gives similar groups of clients access to data. NAS is renowned for providing access to a number of files. In actuality, it is frequently produced as a computer device and is a special, custom-built network. The NAS systems have a number of network-attached storage devices that are organised into logical, redundant storage containers like RAID, or Redundant Array of Independent Disks.

DIFFERENT LEVELS OF STORAGE VIRTUALIZATION

This type of storage where it provides a classification of storage virtualization where it defines what, where and how. The first level in storage virtualization “what” is created. It specifies the types of



and file type virtualization are the core area which are focus more by the developers. The second level is “where” virtualization can take place. This level requires a multilevel approach of all three levels of storage environment such as storage, server and storage network. An effective virtualization distributes the intelligence across three levels while centralising the management and control functions. Data storage functions such as caching and hardware scanning are the example for this type of storage. The third level of virtualization taxonomy specifies network level methodology, in band or out band.

SPACE MANAGEMENT FOR STORAGE DEVICE

Using a log-structured method, fragmentation is handled via garbage collection (GC) while space is allocated progressively in uDepot. We employ this strategy for three factors. First, it performs well on specialised storage, such as NAND Flash. Also, it is more effective than conventional allocation strategies. even for non-specialized storage such as DRAM. Third, co-designing GC and caches offers a variety of optimisation opportunities for caching, an essential use case for uDepot. The log-structured allocator of SALSA's user space port is used to implement allocation. Device space is divided into grains (usually sized equal to the blocks of the IO device), which are then divided into segments (default size: 1 GiB). There are two types of segments: index segments, which flush the index structure to expedite startup, and KV segments, which store KV records. To (sequentially) allocate and release grains, uDepot invokes SALSA. To move certain grains to free segments, SALSA runs GC and upcalls uDepot. A generalised version of the greedy and circular buffer (CB) algorithms, SALSA's GC enhances a greedy policy with the CB's ageing factor.

BENEFITS OF STORAGE VIRTUALIZATION

Network Attached Storage (NAS), in which several storage devices are connected over a sizable, faster network, is improved by storage virtualization. As a result, it offers a big capacity. As the rise in storage demand, this capacity expands. The data movement proceeds without interruption. When data mobility is achieved, the same advantages are seen. Storage use has been enhanced and optimised. To handle the big data that is saved, it needs to be enhanced. Large amounts of data are integrated in data centres while load balancing is carried out here in Storage Virtualization. Data replication occurs, which is useful for disaster recovery. It produces different data replicas that are helpful when such a thing occurs. Data is carefully managed and stored. Maintaining management and order in this massive amount of data is essential. It is done to pool several storage devices together. It connects every storage. Connecting devices through a network. Costs associated with management have decreased, which is the main advantage.

LIMITATIONS OF STORAGE VIRTUALIZATION

Because of the unsecure setting or handling by an uninvited party, data leakage may happen here. These days, data leakage is fairly widespread, and while. It is now risky to transport data for storing or processing. For the purpose of preventing data leakage, data confidentiality and integrity are maintained. It comprises of numerous security problems that must be identified and avoided, like incursion or hacking. Many cybercriminals want to invade the privacy of others and gain unauthorised access to their data. It results in intrusion, which affects the security of the data that is being stored. It has problems with data persistence, where the user can still use erased data.

The user can be blackmailed with this information by retrieving it from the saved areas. Threats to privacy are addressed here. Here, the saved data's private life should be preserved. The resources are properly handled in accordance with demand. It is given based on necessity. It demonstrates that following a failed effort at implementation, backing out is carried out. It imposes a restriction on moving activities internally while still providing user support. It has a number of intricate elements that demand attention. To keep track of all the mapped storages, the metadata needs to be controlled.

PROBLEM STATEMENT AND SOLUTION

The ability to control resource contention in the underlying storage infrastructure is one of the greatest problems with virtualized infrastructure that affects application performance. Over-provisioning storage platforms with IO and disc capacity and managing mission-critical and other IO-intensive applications in silos are two prominent strategies used to lessen the impact of this difficulty. This compromise has a negative influence on operational expenditures related to service assurance and results in a greater cost per gigabyte of storage given to each virtual machine.

SOL: These issues are addressed by Turbonomic's autonomic approach to virtualized storage management, which allows workloads to self-manage while exploring the whole stack of virtualized and cloud environments. Businesses and service providers can use Turbonomic to:20–30% lessening of recurring infrastructure costs for storage. Reduce operational expenses significantly by avoiding complex storage issues and their effects on workloads and, ultimately, end users. Application performance is ensured, and risk is decreased. Enables the reliable virtualization of IO-intensive applications, lowering the overall cost of delivering compute services to the business and customers. Determine which workloads require SSD access and when, then seamlessly integrate heterogeneous SSD/HDD storage configurations.

FUTURE IMPACTS ON STORAGE VIRTUALIZATION

By delivering highly scalable solutions, major cloud providers like AWS, Microsoft Azure, and Google Cloud have changed the storage business. They have altered how businesses handle and store information. Data storage is now necessary because of business applications like CRM, AI, IoT, collaboration tools, and cloud technology. 40% of businesses already utilise storage virtualization, and an additional 12% want to do so by 2022, according to a Spiceworks 2020 State of Virtualization Technologies report. According to the same report, 25% of firms intend to use storage virtualization technologies rather than purchase physical storage arrays in the future. This picture depicts how typical cloud virtualization typically looks. More and more businesses are using the cloud to store their data. after it Companies have some scepticism when it comes to cloud storage because of security issues. The most popular service is Azure, followed by Amazon AWS Storage, Google Cloud Storage, and then Microsoft Azure. According to a Spiceworks survey, 81% of firms employ a storage array (such as a SAN or NAS) nowadays. This includes hard disc only, hybrid, and all-flash storage systems. SSDs are becoming more and more common in businesses, and the SSD market is very competitive. The adoption of all-flash storage arrays is expanding quickly and will pick up steam during the coming several years. According to the Spiceworks analysis, the majority of firms may see a 57% or 60% growth in their storage needs for local shared storage and cloud storage, respectively, during the next two years.

CONCLUSION

This paper briefly covered the idea of storage virtualization and its relationship as well as application to storage topologies like Storage Area Network and Network Attached Storage. Integrated Storage. We also talked about the idea of levels in storage virtualization, as well as their advantages and disadvantages. Last but not least, taking into account its high performance and storage, high availability, efficiency, ability to expand, and scalability of using the data, it is clear that it has a huge future potential in the modern world of technology. Storage Virtualization has also been suggested for consideration in the future and will become more popular in the days to come. Here we also discuss about the benefits of the storage virtualization also the challenges or limitation it face in the future. The future aspects which may take huge impact on the people in the near future is also one of the topic we have covered in this review paper.

IMPLEMENTATION/METHODS USED IN STORAGE VIRTUALIZATION

HOST-BASED : Host-based virtualization is most frequently employed in HCI systems and cloud storage, where software is utilised to control traffic. With this approach, almost any device or array can be assigned physical storage. Virtual drives are presented to guest machines of any configuration by the host or a hyper-converged system made up of numerous hosts, such as virtual machines (VMs) in an enterprise, PCs accessing file shares, or servers accessing data via the cloud.

ARRAY BASED : In array-based storage virtualization, servers are physically placed, but neither the servers nor the users accessing the storage can see which array is being accessed. Here, a storage array performs the function of a primary storage controller by pooling storage resources from other arrays with the aid of virtualization software. The array can display several physical storage types as tiers rather than as a disjointed assortment of devices. The hard disc drives (HDDs) or solid-state drives (SSDs) in these tiers can be found on different arrays.

NETWORK BASED : Network-based storage virtualization is the most used type. Here, a network device connects every storage device to an FC or ISCSI SAN. Within their storage network, these connected devices appear themselves as a single virtual pool.

REFERENCE

1. Jiaxin Li, Yuxi Chen, Haopeng Liu, Shan Lu, Yiming Zhang, Haryadi S. Gunawi, Xiaohui Gu, Dongsheng Li, and Xicheng Lu. PCatch: Automatically Detecting Performance Cascading Bugs in Cloud Systems. In Proceedings of the 2018 EuroSys Conference (EuroSys), 2018.

2. X. Chen, W. Chen, P. Long, Z. Lu, and Z. Wang, ―Semma: Secure efficient memory management approach in virtual environment," in Advanced Cloud and Big Data (CBD), 2013 International Conference on, pp. 131-138, IEEE, 2013.

3. Computing, 4(12), pp.67–73, 2015. [42]Rishikesh Salunkhe, N.J, Query Bound Application Offloading: Approach Towards Increase Performance of Big Data Computing. Journal of Emerging Technologies and Innovative Research, 3(6), pp.188– 191, 2016.

4. Anjali Nigam and Vineet Singh, "Securing Data Transmission in Cloud using Encryption Algorithms", International Journal of Computer Sciences and Engineering Research Volume 4, Issue 6, June 2016.

5.Pritika Goel, "An Improved Load Balancing Technique in Weighted Clustering Algorithm", International Journal of Computer Sciences and Engineering Research Volume 4, Issue 6, June 2016.

6.Lakshmi N. Bairavasundaram, Garth R. Goodson, Bianca Schroeder, Andrea C. Arpaci-Dusseau, and Remzi H. Arpaci-Dusseau. An Analysis of Data Corruption in the Storage Stack. In Proceedings of the 6th USENIX Symposium on File and Storage Technologies (FAST), 2008.