IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Volume-4, Issue-7 E-ISSN: 2347-2693

A Survey on Storage Virtualization and its Levels along with the Benefits and Limitations

Pratik Rajan Bhore

Department of Computer Engineering, Bharati Vidyapeeth University, Pune, India

Available online at: www.ijcseonline.org

Received:18/Jun/2016 Revised: 26/Jun/2016 Accepted: 16/Jul/2016 Published: 31/Jul/2016

***Abstract*—** This study paper revolves around the impact of virtualization at the various layers of storage stack. There is a rapid growth in the storage capacity, and hence the processing power in the respective enterprises storage appliances coupled with the requirements for high availability and it needs a Storage Area Network (SAN) architecture for providing the storage and performance elements here. The Storage Virtualization provides us with a combination and management of storage resources for Storage Area Network with multiple servers as well as the storage devices. The main aim for storage virtualization is its necessity to be inexpensive and not affect the performance. Currently, the storage virtualization is displayed at three different architectural levels: (1) the storage device, (2) the host and (3) the SAN fabric hardware as a central management unit. This paper further provides us with more information on the storage virtualization levels that are part of its architecture. Each of these levels that we will see grants us with particular advantages and benefits but is also limited in their capabilities gets considered as their drawbacks or limitations.

***Keywords*—** Storage, Performance, Virtualization, Network, Storage Virtualization, Storage Area Network (SAN), Network Attached Storage (NAS), Server, Storage Device (Sub-System), Host, SAN Fabric, Virtual Machine Monitor/ Hypervisor

I. INTRODUCTION

A. *Virtualization*

Virtualization is known as the creation of a virtual version of a resource (device) like a server, a storage device, a network or an operating system where the framework is dividing the device into single or more executing environments.

For example, even partitioning of a hard disk drive is considered as virtualization. The reason behind this is, we have one single hard disk drive, and we divide it into two partitions which give us two different hard drives. In the end explaining to us what the term virtualization is all about [1][2][3][4].

Virtualization gets divided into types such as:

1) Server Virtualization (Full and Para-Virtualization) (consists of Virtual Servers)
2) Client Virtualization

Moreover, also

3) Storage Virtualization (consists of Virtual Storage)

The main area of focus here is the most important type of virtualization i.e. the Storage Virtualization. Hence we will study it in detail further. We analyze its levels which are an integral part of its architecture. Moreover, also we will explain its benefits as well as the limitations or drawbacks it faces in the various levels of storage virtualization.

B. *Server Virtualization*

In Server virtualization, we see that one individual server performs the work of many servers combined, as it partitions out the resources or devices of a respective individual server across the multiple environments. The hypervisor/ virtual machine monitor (VMM) allows it to host many apps and operating systems locally.

C. *Client Virtualization*

In Client virtualization, we see that the system administrator carries out the monitoring in a practical manner and also updates the customer's devices such as desktops, laptops, and mobile devices. It gives the improvement of the client device management and helps the client device to defend itself from the cyber criminals and remain secure.

D. *Para-Virtualization*

In Para-Virtualization, we see that there are virtual calls made, which carry out the substitution between ISA (instruction set architecture) of the host. It shows that there

is communication between the virtual machine monitor and operating system (guest). It must be carried for improvement in efficiency as well performance. Accessing devices are better in Para-Virtualization compared to full virtualization since all devices matched in full virtualization.

E. Full Virtualization

In Full Virtualization, we see that the Virtual machine monitor creates a remote environment between the virtual server (guest) and server hardware (host). Operating systems can directly access the hardware controllers as well as the peripheral devices without awareness of the virtual environment or its modified requirements.

F. The Need for Storage Virtualization

Storage Virtualization is essential in carrying out the necessary operations of dividing the available storage space into Virtual Volumes irrespective to the physical layout of the actual available storage components. (e.g. Disk Drives, RAID subsystems)

In Storage Area Network (SAN) there is an essential need for management of large chunks of storage data in a uniform way and centrally located. There is a rapid growth in storage capacity and processing power in most of the enterprise installation activities. Its coupling gets done with the need for high availability and the everyday operations, which requires the Storage architecture to allow them the seamless addition of storage without any downtime. These strategic goals are achieved at best because of virtualization of the storage. The primary need for Storage Virtualization is because of its capability to expand. The importance of this ability can be justified while handling the necessary activities of e-commerce websites. It enables the property of data migration and mobility. Also, Storage Virtualization allows the usage of software apps instead of human admins to perform the various activities and allocations. Data management gets significantly carried through Storage Virtualization applications such as Snapshot Remote Mirroring, Virtual Tapes. It requires creation and expansion of its Virtual Volumes for the use. In summary, a SAN cannot be useful without Storage Virtualization. Hence it is vital and needed the most in Storage Area Network.

II. LITERATURE REVIEW

A. Storage Virtualization

In Storage Virtualization, we observe the creation of abstraction of storage from physical to logical. There are three types of data storage seen in virtualization they are namely; Direct Attached Storage (DAS), Network Attached Storage (NAS) and Storage Area Network (SAN). In DAS we see data storage where the storage devices are directly attached to the server. In NAS we see a connection formed through the network. The NAS hence can be used for operations such as file or device sharing and backup for storage among devices [5][6][7][8]. Moreover, lastly in SAN, we see that there are storage devices that get shared with different servers though a highly accelerated network.

B. Storage Area Network

The Storage Area Network gets defined as an individual network which grants us access to the shared data storage. SANs have been initially used to improve the storage resource devices, such as disks, tapes, and optical disks that can be accessed by the servers so that these support devices can get locally connected to the Operating System. SAN has its individual network of various storage resource devices that cannot access from the LAN by other support equipment.

C. Network Attached Storage

Network-attached storage (NAS) has a data storage server which gets connected to a network that provides data access to a similar bunch of clients. NAS is known for giving access to various files. It is indeed manufactured many times as a computer device which is a unique purpose-built network. The NAS systems contain multiple storage devices attached to the network, which gets arranged as logical, redundant storage containers such as Redundant Array of Independent Disks also known as RAID.

III. LEVELS OF STORAGE VIRTUALIZATION

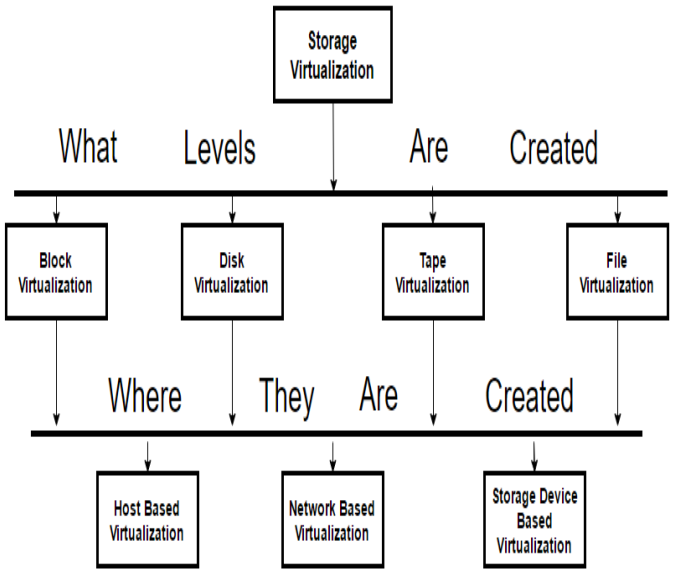

Fig.1 Storage Virtualization Taxonomy

A. *Block Virtualization*

Block Virtualization is a storage virtualization technique that ties the many of the free storage devices together which also known as pooling. Its presentation gets done as one single individual storage device to the host. It carries out address re-mapping from physical to logical address as redirecting takes place here of the IO devices. Its deployment gets implemented in storage area network. It provides non - disruptive data mobility as well as data migration where there is no interference in accessing data or storage device from any mobile location. It gives improved resource devices as well as it enables to reduce cost.

B. *Disk Virtualization*

Disk Virtualization is a storage virtualization technique where the device gets to use a shared disk manager in the virtual world. Logical disks also known as Virtual disks are created by the virtualization device and are made visible to the necessary host or server, thus providing a common area for the management of all the shared volumes or spaces in the virtual world. Here thin provisioning is carried out i.e. the allocation is done accordingly to the demands. It gets transported to maximize storage utilization [9]. The physical storage gets safely stored into the mapping table. Moreover, it can translate the storage address during mapping from physical to logical storage address in the virtual environment. Here Disk expansion in size can also be carried by adding more physical storage to the mapping table. Similarly, disks can also be shrunk in size here by eliminating few of the physical storage from the mapping table [10][11].

C. *Tape Virtualization*

Tape Virtualization is a storage virtualization technique which gets used for backup and recovery reasons. A Virtual Tape / Virtual Tape Library (VTL) present a storage component (generally a hard disk drive) which serves the purpose of a TL or TD for the providing us with existing backup component. The advantages of tape virtualization include that storage is shared and faster in data restoration process. Most of the VTLs in today's world use SAS or SATA disk technologies as their main storage component as they have low cost. The use of these technology enclosures increases the scalability of the component by permitting it the addition of more drives and also increases its storage capacity. The use of VTL also reduces streaming issues that affect the efficiency in tape drives.

D. *File Virtualization*

File Virtualization is a storage virtualization technique which operates on file level in a computer device. It also carries pooling of records where it unites many of the storage devices all into one single logical pool. It is an important part of the file area network (FAN) as well as the network file management (NFM) fields [11][12][13]. Its primary goal is its wish to protect the users as well as the system administrators from the compound storage areas or environments. Other similar goals that add up are the simple managing that is done here, also the efficiency that it provides in capacity usage and allocation as well as the reduction in costs of management.

E. *Host Based Virtualization*

It requires additional components which are running on the host, as an individual task or process. In some of the cases, volume management is built into the OS, and in others, it is offered as a separate product. The traditional physical device driver handles the Volumes presented to the host system. However, a software layer such as the volume manager is just above the instrument driver, and it intercepts the IO operation requests and provides it with the metadata lookups and IO operational mapping [14].

F. *Network Based Virtualization*

It is a storage virtualization that operates on network-based device components and uses Internet Small Computer

System Interface (iSCSI) or Fiber Channel (FC) networks to form a connection as Storage Area Network. These types of device components are available and implemented in the shape of virtualization. The virtualization device components work in the SAN and provide us with the layer of abstraction between the host that is performing the IO operation and the storage controller which provides the storage capacity [15].

G. Storage Device (Sub-System) Based Virtualization

Similar to the host-based virtualization, several operations have existed since many years and have recently gets named as virtualization. Simple data storage components, like single hard drives, do not provide us with virtualization. However, even the simplest disk technology provides a logical to physical translation, as they use Redundant Array of Independent Disk solutions to join more than one drive in a single array. Advanced disk controllers consist of cloning, snapshots, and remote replications. Usually, these components do not provide us with the advantages of data migration or replication across different storage, as each user wants to use their necessary protocols.

IV. STUDY FINDINGS

Storage Virtualization is a great innovation in the world of technology, and it has plenty of benefits no doubt, but meanwhile, there are a few limitations as well which also exist and get considered.

A. Benefits

Storage Virtualization carries out enhancement of the Network Attached Storage (NAS) where multiple storage devices get connected via a large accelerated network. Hence it gives a significant capacity. This capability gets expanded with the increase in demand for storage. There is no disruption in the data migration. The same benefits get seen during achievement of data mobility [16]. The utilization of storage is optimized and improved. It needs to get improved for handling the big data that gets stored. Load balancing is carried here in Storage Virtualization while integrating large chunks of data in datacenters. The replication of data takes place which is handy during disaster recovery. It creates various data replicas that are useful during such phenomenon. Data is stored and properly managed. It is a necessity to keep management and organization in this big amount of data. Pooling of various storage devices together is done. It ties all these storage devices together via a network. The most significant benefit is that there has been a reduction in the management costs. It also reduces the Storage infrastructure problems where everything in connected virtually. Moreover, hence no physical space is needed in large amounts. It increases the efficiency in the storage of data. It undergoes address space remapping where the physical storage gets abstracted to logical storage during the IO redirection requests. It carries the metadata which stores all physical storage locations that are mapped. These places all get stored in a mapping table. It is real management and organization at its best. It carries out disk management here. All the drives connected are managed over here over a single network. It has different storage devices connected to a network i.e. virtually or in a practical manner hence it has small capital expenses. It also has a higher availability rate and is available when we need resources. That is giving service to the user 24 x 7. It provides efficiency in resource utilization and consumes less energy. It is beneficial for energy saving purposes. Also, all the energy is utilized properly for processing purposes. It consists of huge disk memory space for storage. It gets seen in data centers for example. It has a fast loading and backup speed. The same amount of work can be completed with fewer servers since they are collectively working well together. Hence proving that it is work efficient and time-saving in storing and processing of a large amount of data.

These are the benefits of Storage Virtualization derived from surveying the existing architecture.

B. Limitations

Data leakage can occur here due to the insecure environment or handle done by some random party. Data leakage is becoming very common these days where while transferring data for storage or processing purposes has become unsafe. The Data Integrity and Confidentiality get preserved for stopping the data leakage [17]. It consists of various security issues such as intrusion or hacking which need to be detected and prevented. Many hackers want to violate the privacy of others and want to access their data without authorization. It causes into intrusion and hence affecting the security of stored data. It has data remanence issues where the deleted data gets used by the user itself. Some delicate information is removed by the user after it uses. This data can be retrieved from the stored spaces and used in blackmailing the user. Privacy threats get handled here. Here the private life of stored data should be preserved. The resources get managed appropriately and according to their demand. It gets based on a need basis. It shows backing out is carried after a failed attempt of implementation. It gives the limitation of internally moving the operations with the user support. It consists of various complexities that need attention. The metadata needs to be managed to keep track of all the mapped storages. All this

data needs to be properly handled and organized. Performance and scalability are the most important of these issues, where they need to be optimized otherwise the work done is low. Performance gets based on various parameters such as latency, throughput, bandwidth and memory. These parameters get considered for achieving high performance in virtualization. If not, then the performance is low. Here violation of licensing agreements occurs sometimes. The benefit of reduction in management costs and an increase in some disks can boost people to add some servers, which will create a server sprawl, in which there will be too many servers to be managed. The network system is much more complex here in storage virtualization. If one storage system fails while getting connected during processing, then they all fail. If one of the servers gets infected or corrupted, then the whole network gets affected and compromised.

These are the limitations of Storage Virtualization derived from surveying the existing architecture.

V. CONCLUSION AND FUTURE SCOPE

This paper discussed the Storage Virtualization concept in brief and its relation as well as use with the storage topologies such as Storage Area Network and Network Attached Storage. We also discussed the concept of Levels in Storage Virtualization along with their benefits and limitations. Finally, considering its high performance as well as storage, high availability, efficiency, its capability to expand and scalability of utilizing the data proves that it has humongous future scope in the new world of technology. Moreover, hence Storage Virtualization has been proposed for future consideration and will grow bigger in the coming days.

ACKNOWLEDGMENT

I would like to thank Dr. Naveen Kumar for his complete guidance and the Computer Engineering department of BVUCOE, Pune as well. I would also like to thank my family and friends for their constant support and confidence in me.

REFERENCES

[1] Jun-wei Ge, Yong-long Deng, Yi-qiu Fang, "Research on Storage Virtualization Structure in Cloud Storage Environment" , Multimedia Technology (ICMT), International Conference, IEEE, **2010**.

[2] Aameek Singh, Madhukar Korupolu and Dushmanta Mohapatra, "Server-Storage Virtualization: Integration and Load Balancing in Data Centers", IEEE, **2008.**

[3] Sandeep Kumar and Syam Kumar P, "Secure and efficient design and implementation of out-of-band storage virtualization", IEEE, **2015**.

[4] S. D. Joshi, Naveenkumar Jayakumar, Farid Zeimfar, "Workload Characteristics Impacts on File System Benchmarking", Int. J. Adv. Res. Computer Sci. Software Eng., vol. 4, no. 2, pp. 39–44, **2014**.

[5] N. Jayakumar, S. Singh, S. H. Patil, and S. D. Joshi, "Evaluation Parameters of Infrastructure Resources Required for Integrating Parallel Computing Algorithm and Distributed File System," IJSTE - Int. J. Sci. Technol. Eng., vol. 1, no. 12, pp. 251–254, **2015**.

[6] N. Jayakumar, T. Bhardwaj, K. Pant, S. D. Joshi, and S. H. Patil, "A Holistic Approach for Performance Analysis of Embedded Storage Array."

[7] D. T. Jayakumar and R. Naveenkumar, "Active Storage," Int. J. Adv. Res. Computer Sci. Software Eng. Int. J, vol. 2, no. 9, pp. 62–70, **2012**.

[8] M. G. Kakamanshadi, M. J. Naveenkumar and S. H. Patil, "A Method to Find Shortest Reliable Path by Hardware Testing and Software Implementation," Int. J. Eng. Sci., **2011**.

[9] N. Kumar, S. Angral and R. Sharma, "Integrating Intrusion Detection System with Network monitoring," Int. J. Sci. Res. Publ., vol. 4, no. 5, pp. 1–4, **2014**.

[10] P. D. S. D. J. Naveenkumar J, "Evaluation of Active Storage System Realized through Mobility RPC," Int. J. Innov. Res. Computer Commun. Eng., vol. 3, no. 11, pp. 11329–11335, **2015**.

[11] S. D. Joshi, Naveenkumar J, "Evaluation of Active Storage System Realized Through Hadoop," Int. J. Computer Sci. Mob. Computer, vol. 4, no. 12, pp. 67–73, **2015**.

[12] J. Naveenkumar, R. Makwana, S. D. Joshi, and D. M. Thakore, "Offloading Compression and Decompression Logic Closer to Video Files Using Remote Procedure Call," J. Impact Factor, vol. 6, no. 3, pp. 37–45, **2015**.

[13] N. J. Rishikesh Salunkhe, "Query Bound Application Offloading: Approach Towards Increase Performance of Big Data Computing," J. Emerg. Technol. Innov. Res., vol. 3, no. 6, pp. 188–191, **2016**.

[14] R. Salunkhe, A. D. Kadam, N. Jayakumar, and S. Joshi, "Luster A Scalable Architecture File System: A Research Implementation on Active Storage Array Framework with Luster File System." in ICEEOT, **2015**.

[15] R. Salunkhe, A. D. Kadam, N. Jayakumar, and D. Thakore, "In Search of a Scalable File System State-of-the-art File Systems Review and Map view of new Scalable File system.," International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - **2016**, pp. 1–8.

[16] Naveenkumar J, Sagar S Lad, S. D. Joshi, "Comparison study on Hadoop's HDFS with Lustre File System,"

Int. J. Sci. Eng. Appl. Sci., vol. 1, no. 8, pp. 491–494, **2015**.

[17] G. Sullivan, "Virtualization for Storage Efficiency and Centralized Management."

[18] T. A. Faculty, S. Sivathanu, and I. P. Fulfillment, "Enhancing Storage Performance in Virtualized Environments Enhancing Storage Performance in Virtualized" no. August, **2011**.

[19] G. Praveen, "Analysis of Performance in the Virtual Machines Environment," International J. Adv. Sci. Technol., vol. 32, no. pp. 53–64, **2011**.

[20] Press, "Virtualization." Summary, "How Processor Core Count Impacts Virtualization Performance and Scalability.

[21] Salunkhe R, Kadam A.D, Jayakumar N and Thakore D, In Search of a Scalable File System State-of-the-art File Systems Review and Map view of new Scalable File system.

[22] Raval K.S, Suryawanshi R.S, Naveenkumar J and Thakore D.M, The Anatomy of a Small-Scale Document Search Engine Tool: Incorporating a new Ranking Algorithm. International Journal of Engineering Science and Technology, 1(3), pp.5802-5808, **2011.**

[23] Archana R.C, Naveenkumar J and Patil, S.H, Iris Image Pre-Processing and Minutiae Points Extraction. International Journal of Computer Science and Information Security, 9(6), p.171, **2011**.

[24] BVDUCOE, B.B, Iris Image Pre-Processing and Minutiae Points Extraction. International Journal of Computer Science & Information Security, **2011**.

[25] Jayakumar M.N, Zaeimfar M.F, Joshi, M.M and Joshi S.D, 2014. INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY (IJCET). Journal Impact Factor, 5(1), pp.46-51.

[26] Naveenkumar J and Joshi S.D, 2015. Evaluation of Active Storage System Realized through Mobility RPC.

[27] Jayakumar D.T and Naveen Kumar, S.D Joshi, "International Journal of Advanced Research in Computer Science and Software Engineering", Int. J, 2(9), pp.62-70, **2012**.

[28] Jayakumar N, Singh S, Patil S.H and Joshi S.D, Evaluation Parameters of Infrastructure Resources Required for Integrating Parallel Computing Algorithm and Distributed File System.

[29] Jayakumar N, Bhardwaj T, Pant K., Joshi S.D and Patil S.H, A Holistic Approach for Performance Analysis of Embedded Storage Array.

[30] Naveenkumar J, Makwana R, Joshi S.D and Thakore D.M, Offloading Compression and Decompression Logic Closer to Video Files Using Remote Procedure Call, Journal Impact Factor, 6(3), pp.37-45, **2015**.

[31] Naveenkumar J, Makwana R, Joshi S.D and Thakore D.M, Performance Impact Analysis of Application Implemented on Active Storage Framework. International Journal, 5(2), **2015**.

[32] Salunkhe R, Kadam A.D, Jayakumar N and Thakore D, In Search of a Scalable File System State-of-the-art File Systems Review and Map view of new Scalable File system.

[33] Salunkhe R, Kadam A.D, Jayakumar N and Joshi S.D, Luster A Scalable Architecture File System: A Research Implementation on Active Storage Array Framework with Luster File System.

[34] Jayakumar N, Reducts and Discretization Concepts, tools for Predicting Student's Performance.

[35] Jayakumar M.N, Zaeimfar M.F, Joshi M.M and Joshi S.D, International Journal of Computer Engineering & Technology (IJCET), Journal Impact Factor, 5(1), pp.46-51, **2014.**

[36] Kumar N, Angral S and Sharma R, Integrating Intrusion Detection System with Network Monitoring. International Journal of Scientific and Research Publications, 4, pp.1-4, **2014**.

[37] Namdeo J and Jayakumar N, Predicting Students Performance Using Data Mining Technique with Rough Set Theory Concepts. International Journal, 2(2), **2014**.

[38] Naveenkumar J, Keyword Extraction through Applying Rules of Association and Threshold Values. International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), ISSN, pp.2278-1021.

[39] Kakamanshadi G, Naveenkumar J and Patil S.H, 2011. A Method to Find Shortest Reliable Path by Hardware Testing and Software Implementation. International Journal of Engineering Science and Technology (IJEST), ISSN, pp.0975-5462.

[40] Naveenkumar J and Raval K.S, Clouds Explained Using Use-Case Scenarios.

[41] Naveenkumar J, S.D Joshi, Evaluation of Active Storage System Realized Through Hadoop. International Journal of Computer Science and Mobile 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**.

[43] Sagar S Lad, S.D Joshi, N.J, Comparison study on Hadoop's HDFS with Lustre File System. International Journal of Scientific Engineering and Applied Science, 1(8), pp.491–494, **2015**.

[44] Angel, S. et al., End-to-end performance isolation through virtual datacenters. 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14), (Vdc), pp.233–248. Available at: http://dl.acm.org/citation.cfm?id=2685048.2685067, **2014**.

[45] Cherkasova L & Gardner R, Measuring CPU overhead for I/O processing in the Xen virtual machine monitor. Proceedings of the USENIX annual …, pp.387–390. Available at: http://static.usenix.org/events/usenix05/tech/general/full_papers/short_papers/cherkasova/cherkasova.pdf. **2005**.
[46] Paper W, Storage Intelligence in the Network : EMC's Perspective, **2003**.
[47] VMware, Virtualizing Performance-Critical Database Applications in VMware vSphere. VMware, p.8, **2009**.
[48] https://www.cs.uic.edu/~jbell/CourseNotes/OperatingSystems/13_IOSystems.html
[49] Durairaj M and Kannan P, A Study On Virtualization Techniques and Challenges in Cloud Computing, International Journal Of Scientific & Technology Research Volume 3, Issue 11, November **2014**.
[50] 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**.
[51] 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.**

AUTHORS PROFILE

Pratik Rajan Bhore received his B. Tech in Computer Engineering (2015) and Diploma in Network Security (2015) from Bharati Vidyapeeth University's College of Engineering, Pune, India where he is pursuing his last year of M. Tech in Computer Engineering. (2016-17). His research interests include Virtualization in Storage, Storage Area Network, Nanorobotics, Big Data, Software Engineering and Software Defined Storage.

