International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

Software Defined Storage Implementation Model to Data Centers

Andy Reyes Vargas, Alex Pacheco Pumaleque, Enrique Condor Tinoco, Marco de la Cruz Rocca

***Abstract:** This research focuses on deliver a SDS implementation model that facilitate the definition of requirements to be delivered in a timely manner to systems. Software Defined Storage (SDS) enables different hardware technologies to be homogenized and deliver independent storage. The technologies included on the implementation model are the open-source projects Red Hat Gluster Storage and Red Hat Ceph Storage. Both technologies are integrated into a virtualization platform, it is compared at a functional and economic level with respect to some enterprise solutions and it is shown that the same operations are performed as a viable technological alternative. The favorable results of the model are presented and the great benefit that open-source technologies have open scope that allows it to be integrated into other projects is presented.*

***Keywords:** Software Defined Storage, implementation model, virtualization, gluster, ceph, technological alternative, open-source.*

I. INTRODUCTION

Software Defined Storage (SDS) [1] is a technology where hardware and software are independent of a storage solution. Traditionally, storage providers such as HP, IBM, Dell [2] enhance the system functionalities at the hardware level, in this case the SDS applies these functionalities in the software level, allowing the hardware to only be a layer of block provisioning and the software to manage it. Therefore, the capacities can be expanded according to the needs of the client system and in the time required with the least amount of interruptions possible.

SDS can be implement a high-level storage[3] system with different server technologies, that is, a storage cluster can be implemented on physical or virtual servers[4] and can replicate the data to a disaster recovery data center waiting for the main data center be offline.

Manuscript received on April 02, 2020.
Revised Manuscript received on April 15, 2020.
Manuscript published on May 30, 2020.

* Correspondence Author

Andy Reyes Vargas*, IT Department, Open Nova IT Consulting, Lima, Peru. Email: andy.reyes@**opennova**.pe

Alex Pacheco Pumaleque, Professional School of Systems Engineering, National University of Cañete, Lima, Peru. Email: apacheco@undc.edu.pe

Enrique Condor Tinoco, Academic Department of Engineering and Computer Technology, Jose **Maria** Arguedas National University, Andahuaylas, Peru. Email: enricoti@gmail.com

Marco de la Cruz Rocca, Professional School of Systems and Computer Engineering, Daniel Alcides Carrion National University, Cerro de Pasco, Peru. Email: ocmar55@gmail.com

The proposed implementation model offers key functionalities for the organization such as:

- Automation of intra-datacenter data replication processes: Volumes created in the gluster and ceph groups are created with replication between the blocks of the servers in the same cluster
- Automation of inter-datacenter data replication processes: Local cluster volumes can replicate to another data center asynchronously, ensuring replication to a remote site at a controlled bandwith and preparing for a procedure disaster recovery.
- Scalability of systems from different providers: When implementing SDS, the nodes can be replaced by another available hardware, only clarifying that the HA function is limited, in case of having similar hardware, the HA[5] is fully functional.
- High availability of virtualization components: Option inherited from virtual systems hosted on physical systems that allow components to be deployed in active-backup or active-active mode to maintain component availability[6].
- Delivery of volumes in different access formats: The implementation of the gluster project for block storage for mixed platforms and specialization in virtualization and the ceph project that specializes in volumes for cloud platforms through instances, offer a very complete solution for the level of infrastructure as a service.

SDS implementation model offers some storage cases applied to the data center, such as:

- Replicated volume[7] for critical platform
- Volume distributed[8] for non-critical platform
- Replicated volume for cloud instances[9]
- Volume replicated for contingency data center asynchronously.

These cases are high-level requirements of a data center that are traditionally deploy with high-performance enterprise storage systems, however, now they are done with open-source projects.

II. SDS IMPLEMENTATION MODEL

The implementation model is divided into 2 parts, the first component virtualization and the second component implementation.

A. Component virtualization

Virtualization systems allow deployment and execution of guest systems with the least complexity by providing virtualization templates[10] with pre-installed primary applications, ready to customize only the unique parameters of the system, reducing administrator operations to deploy storage nodes.

Virtualization is implemented with redundancy of components and protocols capable of managing this redundancy in physical components, therefore, virtual systems inherit this redundancy from physical components and meet the high availability requirements of the organization.

Figure 1 shows a virtualization system for opennova.pe domain, which includes 3 hosts from the Red Hat Virtualization project[11].

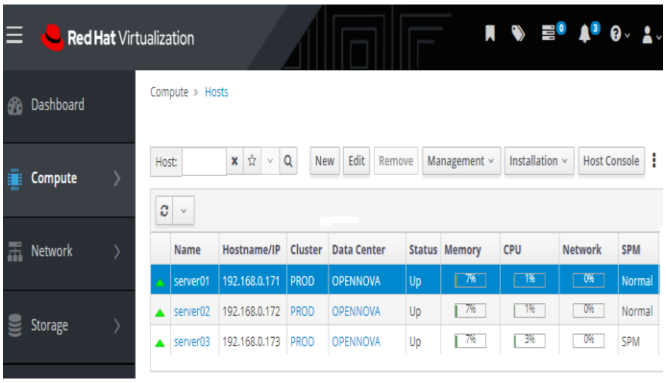

Figure 1. Virtual Data Center

In addition to hosts, it is important to implement redundant network connections configured in active-backup mode available as shown in Figure 2

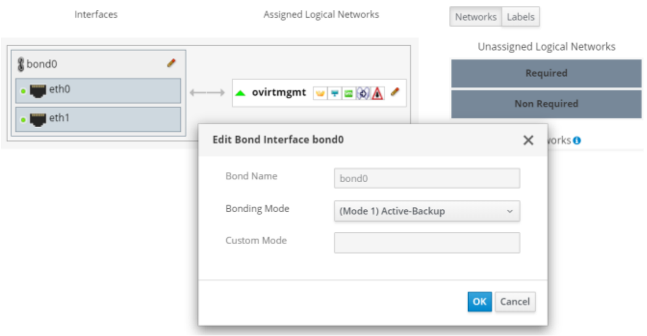

Figure 2. Setup host network

Finally, the virtualization templates of the gluster and ceph projects that simplify the SDS node implementation process as shown in Figure 3

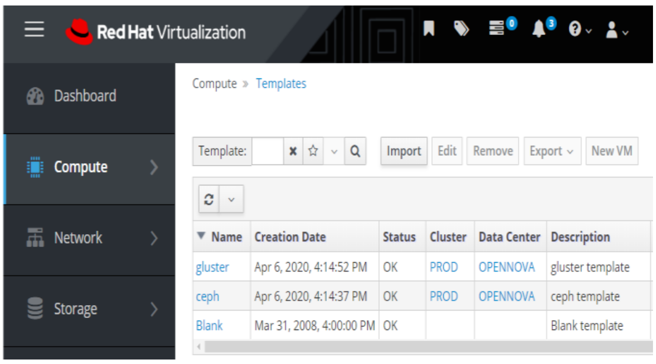

Figure 2. Gluster and Ceph templates

B. Component implementation

General-purpose storage is implemented with the Red Hat Gluster Storage[12] project, which has the necessary functionalities for the data center; these functionalities allow it to deliver volumes for multiple infrastructures such as virtualization technologies, Linux systems, Windows systems, containers, hyperconvergence and derivatives.

Cloud instance storage is implemented with the Red Hat Ceph Storage[13] project, which targets its volumes for cloud instances like Red Hat OpenStack project.

Creation of volumes begins with the design of those volumes, specifying their components and requirements, these details are shown in Table 1.

Table 1: Volume requirements

Volume	Type	Criticality	Client OS	Client platform
vol01	replicated	high	Linux	physical
vol02	replicated	medium	Windows	physical
vol03	distributed	low	Linux	virtual
vol04	replicated	high	Linux	container
vol05	replicated	high	Linux	instance

Considering the requirements of Table 1, the distribution of blocks is designed with the implementation model similar to Table 2 where X = used block and A = arbiter block[14].

Table 2: Volume design

	Node		
Volume	rhgs01	rhgs02	rhgs03
vol01	X	X	X
vol02	X	X	A
vol03	X	X	X
vol04	X	X	X
vol05	X	X	-

In Table 2 volumes 01, 02, 03 and 04 are created with the RedHat Gluster Storage project and volume 05 with RedHat Ceph Storage according to Figure 4 and Figure 5 respectively.

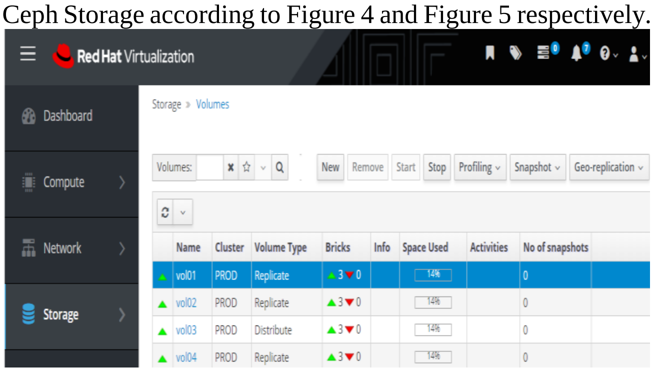

Figure 4. Gluster volumes

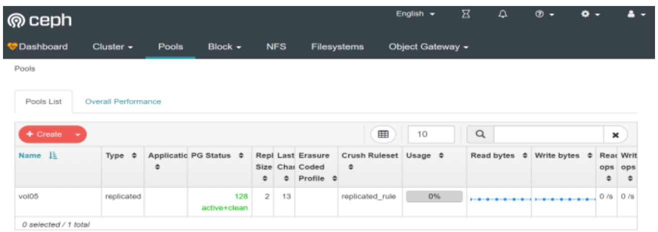

Figure 5. Ceph volumes

III. DISASTER RECOVERY

The SDS implementation model permits infrastructure to be prepared for recovery[15] in the event of a data center crash, protecting data from critical volumes such as Table 2 vol01.
Key functionality is geo-replication[16], which allows it replicates it through a link to another previously implemented data center, but with lower capacities.
The purpose of geo-replication is not saturate the link reserving a portion of the available bandwidth for remote replication. This guarantees the protection of data on the other site and its availability in the event of a disaster. In Figure 6 it is seen how the vol01 of the PROD cluster is replicated to the volume vol01_rep of the DISASTER cluster.

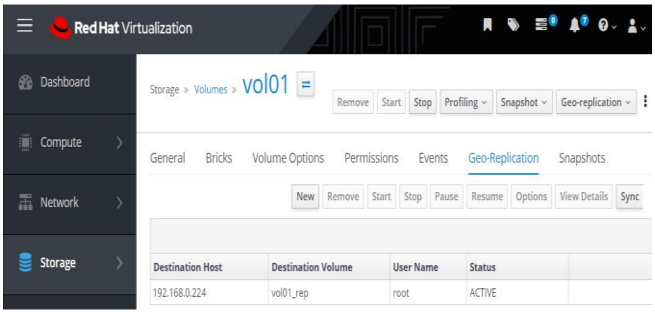

Figure 6. Geo-replication from vol01(PROD) to vol01_rep(DISASTER)

IV. BENEFITS AND RESULTS OF APPLYING THE IMPLEMENTATION MODEL

SDS control software allows managing access to hardware resources such as storage blocks and network connections. This intermediate layer adds the following benefits to the implementation model:

- ✓ Hardware independence: SDS allows implementing systems from different manufacturers, that is, a storage cluster can be built with pre-existing servers in the organization and if necessary, acquire new servers to increase the cluster's capabilities
- ✓ Profitability: reusing servers in the organization or hardening servers with new specific components generates considerable cost savings when compared to purchasing business storage systems.
- ✓ Adaptability: by separating the hardware from the storage systems, manager can expand the capacities of the volume over SDS online, migrate the volumes with change servers, generate backups or replicas in a transparent way, that is, the storage is completely dynamic.
- ✓ Integration: the projects chosen for the implementation model are characterized by integration with other projects through APIs[17], therefore, the possibilities of integration of the proposed model are very high.

In addition to the technical benefits, the economic benefit of the implementation model is important, since the open-source technologies of the RedHat company are not marketed as a license but as a subscription, it is a renewable online service if the organization requires it, generating considerable savings of costs, to estimate costs, the annual costs of the minimum subscriptions[18] for the implementation of the model are detailed in Table 3 if the storage is in the production phase and the organization requires it:

Table 3: Economic comparison between enterprise and open-source solutions in US$

Product	Quantity	Price	Total
VMware vSphere Enterprise Plus (includes vSan)	3	4500	18000
HPE Helion OpenStack	3	3385	10155
		Total	28155
Red Hat Enterprise Virtualization	3	1500	4500
Red Hat Gluster Storage (3 nodes)	1	10000	10000
Red Hat Ceph Storage (3 instances)	1	7500	7500
		Total	22000

V. CONCLUSIONS

- ✓ The implementation model allows integrating open-source projects using APIs, therefore it is recommended to continue investigating other independent SDS projects and integrate them into the proposed model.
- ✓ The application of open standards allows the implementation model to scale horizontally on platforms from different manufacturers, turning the implementation model into another open-source project.
- ✓ Although the model is open-ended for physical or virtual platforms, the researcher recommends using virtualization as a base platform because it generates a software layer that supports the high availability and scalability of components.

ACKNOWLEDGEMENTS

The author thanks Red Hat Certified Architect (RHCA) professionals for their work documenting and validating the integration of high-level open-source projects that are the main source of this research.

REFERENCES

1. NetApp. "What Is Software-Defined Storage?", (2020). [Online]. Available: https://www.netapp.com/us/info/what-is-software-defined-storage.aspx. [Accessed: 14-March-2020].
2. Gartner. "The Future of Software-Defined Storage in Data Center, Edge and Hybrid Cloud", (2019). [Online]. Available: https://www.gartner.com/en/documents/3913237/the-future-of-software-defined-storage-in-data-center-ed. [Accessed: 14-March-2020].
3. TrustRadius. "Software Defined Storage (SDS) Solutions", (2020). [Online]. Available: https://www.trustradius.com/software-defined-storage-sds. [Accessed: 14-March-2020].

4. VMware. "Server Virtualization", (2020). [Online]. Available: https://www.vmware.com/topics/glossary/content/server-virtualization. [Accessed: 14-March-2020].
5. VMware. "High Availability", (2020). [Online]. Available: https://www.vmware.com/products/vsphere/high-availability.html. [Accessed: 14-March-2020].
6. RedHat. "Bonding Modes", (2019). [Online]. https://access.redhat.com/documentation/en-us/red_hat_virtualization/4.3/html/technical_reference/bonding_modes. [Accessed: 14-March-2020].
7. RedHat. "Creating replicated volumes", (2019).
8. students enjoy learning and feel motivated to the
9. [Online]. https://access.redhat.com/documentation/en-us/red_hat_gluster_storage/3.5/html/administration_guide/sect-creating_replicated_volumes. [Accessed: 14-March-2020].
10. RedHat. "Creating distributed volumes", (2019). [Online]. https://access.redhat.com/documentation/en-us/red_hat_gluster_storage/3.5/html/administration_guide/creating_distributed_volumes. [Accessed: 14-March-2020].
11. AWS. "Amazon EC2 Instances", (2019). [Online]. https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Instances.html. [Accessed: 14-March-2020].
12. RedHat. "Virtualization templates", (2019). [Online]. https://access.redhat.com/documentation/en-us/red_hat_virtualization/4.3/html/virtual_machine_management_guide/chap-templates. [Accessed: 14-March-2020].
13. RedHat. "Red hat Virtualization", (2019). [Online]. https://access.redhat.com/documentation/en-us/red_hat_virtualization/4.3/html/product_guide/index. [Accessed: 14-March-2020].
14. RedHat. "Red hat Gluster Storage", (2019). [Online]. https://www.redhat.com/es/technologies/storage/gluster. [Accessed: 14-March-2020].
15. RedHat. "Red hat Ceph Storage", (2019). [Online]. https://www.redhat.com/es/technologies/storage/ceph. [Accessed: 14-March-2020].
16. Gluster Docs. "Arbiter volumes and quorum options in gluster", (2019). [Online]. https://docs.gluster.org/en/latest/Administrator%20Guide/arbiter-volumes-and-quorum/. [Accessed: 14-March-2020].
17. Gluster Blogs. "GlusterFS: Disaster Recovery", (2019). [Online]. https://www.gluster.org/glusterfs-disaster-recovery/. [Accessed: 14-March-2020].
18. RedHat. "Preparing to deploy geo-replication", (2019). [Online]. https://access.redhat.com/documentation/en-us/red_hat_gluster_storage/3.1/html/administration_guide/sect-Preparing_to_Deploy_Geo-replication#Exploring_Geo-replication_Deployment_Scenarios. [Accessed: 14-March-2020].
19. ResFullAPI. "What is Rest", (2019). [Online]. https://restfulapi.net/. [Accessed: 14-March-2020].
20. RedHat. "Red Hat Enterprise Linux subscription guide", (2019). [Online]. https://www.redhat.com/en/resources/Linux-rhel-subscription-guide. [Accessed: 14-March-2020].

AUTHOR PROFILE

Reyes Vargas, Andy, Computer and systems engineer and doctoral candidate, specialist in open source solutions by Redhat with the highest current RHCA degree, in addition to other certifications such as RHCE and RHCSA, and specialist in virtualized Data Centers with current certification of the vmWare company as VCP-DCV and the Redhat company as RHCVA.

Pacheco Pumaleque, Alex, Research professor focused on Smart Cities, Internet of Things, Cloud Computing and Educational Technology. Master in Strategic Management in ICTs Information and Communication Technologies, registered at Sunedu by the University of Piura UDEP. Miraflores headquarters. Past Director of the School of Systems Engineering of the National University of Cañete UNDC 2017-201

Condor Tinoco, Enrique Systems and Computer Engineer, soon to obtain the degree of Doctor in Systems Engineering; teacher at the José María Arguedas National University. Its objective is to make the lessons become practical exercises so that the students enjoy learning and feel motivated to the solutions of real problems.

De la Cruz Rocca, Marco Systems Engineer and PhD candidate, Master in Education at the Higher Level with ICT tools, Master in Finance, university professor, researcher and author of books in education and computer science.