

Automatic Backup System for Virtualization Environment

Idris Winarno, Muzaki Nurus Sani

Politeknik Elektronika Negeri Surabaya
Kampus PENS, Jl. Raya ITS Sukolilo, Surabaya, 031-5947280
E-mail: idris@pens.ac.id, vnyx14@gmail.com

Abstract

Virtualization is a technology lately much discussed and considered as the proper way to cut costs in the construction of a data center. One example of the implementation of virtualization technologies is to using VMware. Another tools for virtualization are Xen and OpenVZ, but VMware is more flexible than Xen or OpenVZ because VMware can run a variety of operating systems. Although it has the advantage, virtualization technology also has a vital weakness, virtualization technologies could be analogous by putting all the eggs in a basket. This means that if the master server problem, all systems inside the virtual machine can not be used. However, it can be anticipated by provide backup facilities that run continually and automatically. VMware itself has had an application to backup/replicate virtual machines. However, that application is not free yet.

This research has been design and creates a web-based software for backing up virtual machines on VMware. So it made easier for users and admins to perform periodic backups of virtual machines.

From the test results has been done, it can be seen that used disk type thin or zeroed thick make process backup faster, system can't work well when virtual machine has snapshot, scheduling system and restoring system has worked well, physical ability data storage influence system.

Keywords: Virtual machine, virtualization, VMware, Backup, Data Center.

1. INTRODUCTION

The data center is a facility used for the placement of a set of servers or computer systems and storage which is conditioned by the power supply, air-conditioning, fire prevention and is usually equipped with physical security system [1]. At first, the data center is always built in a large room to assist the operation of a company. The super computers are still too complex for the maintenance and operation. Until in the 1960s IBM introduced a mainframe virtualization technology in its products, which can be built on a mainframe

virtual machines-virtual machine by utilizing the resources of a very large IBM mainframe. Virtualization itself can be interpreted as the creation of a virtual version of a form or something physical, such as operating systems, storage devices/data storage or network resources [2]. With virtualization technology makes it easier to implement data center on company facilities. Virtualization technology has many advantages such as [2] :

- **Efficiency**, by using the concept of virtualization, an agency can cut procurement costs for servers.
- **Security**, concept of virtualization has protection on various system resources, namely by eliminating the division of resources directly, so there is no problem of protection in virtualization technology. Virtualization technology is a perfect vehicle for operating systems research and development. With virtualization technology, if there is a change in one part of the machine, it is guaranteed not to change other components.
- **Recovery**, virtualization solutions can be used for system recovery plan that requires portability and flexibility across platforms.

In addition to the above advantages of virtualization technology also has drawbacks, virtualization technologies could be analogous to putting all the eggs in one basket. This means that if the master server problem, all systems inside the virtual machine can not be used. It can be anticipated by providing automatic backup facility and periodically or by applying the principle of fail over/clustering.

VMware itself has had an application to backup/replicate virtual machines that Veeam Backup & Replication. However, these applications are still paid. Through this research, is expected to be a tool to facilitate the users especially network admin perform backups automatically and periodically.

2. RELATED WORKS

Reviewing of several papers that there is:

Xiaoying Wang , Zhihui Dua, Yinong Chen, Sanli Li (2007) [1], in a paper entitled "Virtualization-based autonomic resource management for multi-tier Web applications in shared data center". The paper explains that they develop a new architecture of resource management framework for multi-tier applications based on virtualization mechanisms to isolate different application environments for security reasons and to provision shared resources effectively and efficiently to meet different service quality targets with The minimum operating costs.

Xiangzhen Kong, Chuang Lin, Yixin Jiang, Wei Yan, Xiaowen Chu (2010) [2], in a paper entitled "Efficient dynamic task scheduling in virtualized the data centers with fuzzy prediction". The paper explains that the virtualization also brings some challenges, particularly for resource management and task scheduling. This paper proposes an efficient dynamic task scheduling scheme for virtualized data centers. Considering the availability and responsiveness

performance, the general model of task scheduling for virtual data centers is built and formulated as a two-objective optimization. Procedure graceful fuzzy prediction method is given to model the uncertain workload and the vague availability of the virtual server nodes, by using type I and type-II fuzzy logic systems.

Hai Jin, Wei Gao, Song Wua, Xuanhua Shi, Xiaoxin Wub, Fan Zhou (2010) [3], in a paper entitled "Optimizing the live migration of virtual machine by CPU scheduling". This paper explains that they are addressing the problem of downtime during the migration of virtual machine by designing optimization scheme for live migration, which according to pre-copy speed, vCPU operating frequency can be reduced so that at certain phases of memory pre-copy the remaining dirty can achieve the desired small amount . VM downtime during the migration can be limited. This scheme works for the scenario where the migrated application has a high memory writing speed or the pre-copy speed is slow for example, due to low network bandwidth between the migration.

3. Method

GhettoVCB is backup agent used to backup the virtual machine server ESX (i). GhettoVCB created using the basic shell programming so it can work on UNIX OS. These scripts perform backups of virtual machines residing on ESX (i) server 3.5/4.x/5.x using methodology similar to VMware VCBtool. The script takes a snapshot of the virtual machine is running, backing up the master VMDK (s) and then when you're done, delete the snapshot until the next backup process. The only caveat is that it utilizes the resources available to the Service Console of ESX server or Busybox Console (Tech Support Mode) from the ESXi server running the backups as opposed to following the traditional method of offloading virtual machine backups through a VCB proxy [6].

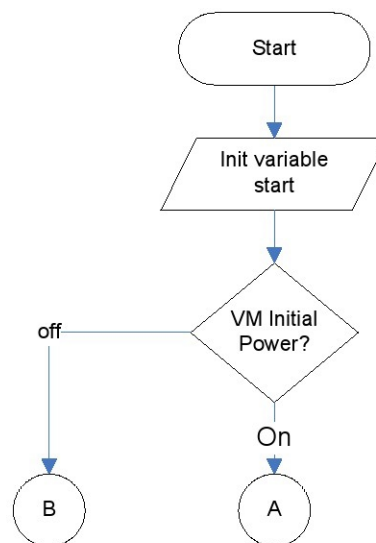


Figure 1. Flowchart GhettoVCB

This script has been tested on ESX and ESXi 3.5/4.x/5.x 3.5/4.x/5.x and supports the following backup media: LOCAL STORAGE, SAN and NFS[7]. Script is a non - interactive and can be setup to run via cron. Currently, this script accepts a text file that lists the names of the virtual machine display (s) to be supported. Moreover, one can specify the folder that contains the configuration data on a per VM basis for granular control over backup policies.

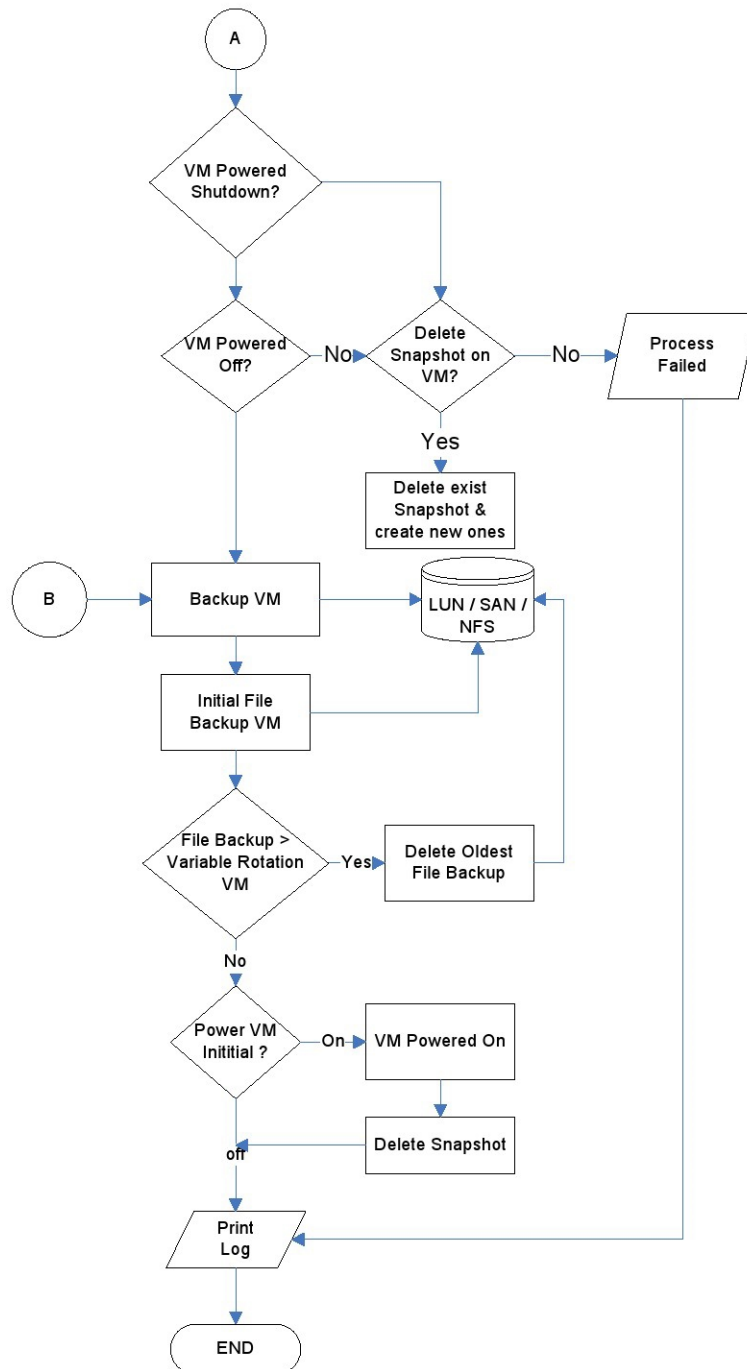


Figure 2. Flowchart GhettoVCB

Explanation of the flowchart (Figure 1 and Figure 2) such as:

- When running the agent, the script will initialize variables-variables that have been entered.
- After that, the agent will check whether the VMs to be backed up alive/dead. If you die, then the agent directly to backup VMs.
- If the VMs are alive then the agent will see the variable that has been entered to determine whether to turn off or not. If still on the agent will do the snapshotting process on the VMs.
- Then the agent will perform backups and store them into a predetermined storage.
- Agent will check the amount of rotation in the variable and the number of files stored on the storage. If more of the storage agent will remove the oldest data.

4. SYSTEM DESIGN

There is a design of the whole process in the manufacture of the system. As we know, before a system previously made sure no idea of the background to make such a system, it has been described by the authors in the introduction. Then to make a sketch of the system also needed a system that would be made, the authors have translate the planning system. Once the idea and design of existing.

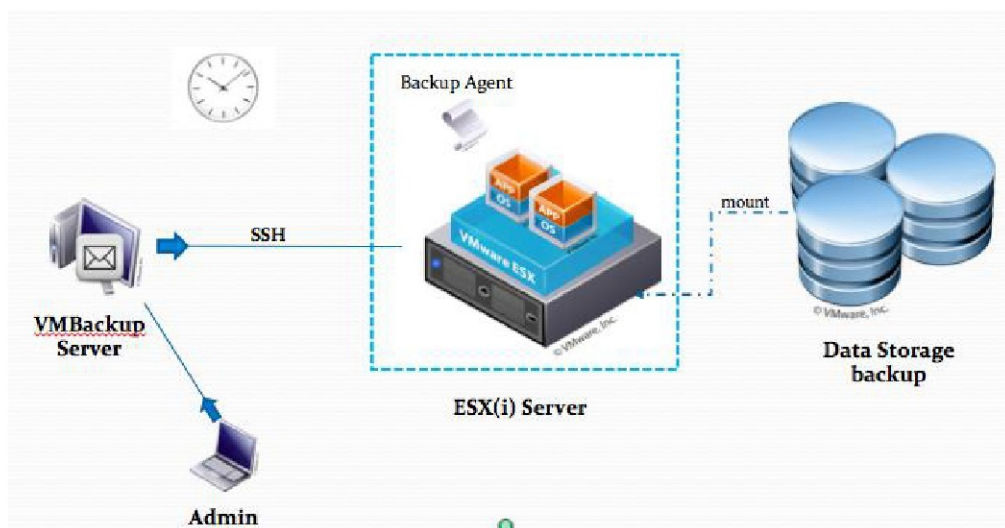


Figure 3. Work principle form backup system virtual machine

From Figure 3 can explain the working principle of the system in more detail as follows:

1. Users access the web server and provide the server command to perform a backup on a virtual machine.
2. Then, web server execute remote commands for VMware via SSH protocol and then run the backup agent.

3. Agent runs backup program to back up VMware host servers and put in backup storage that has been determined.
4. After processing user can get result in form either logs or emails.

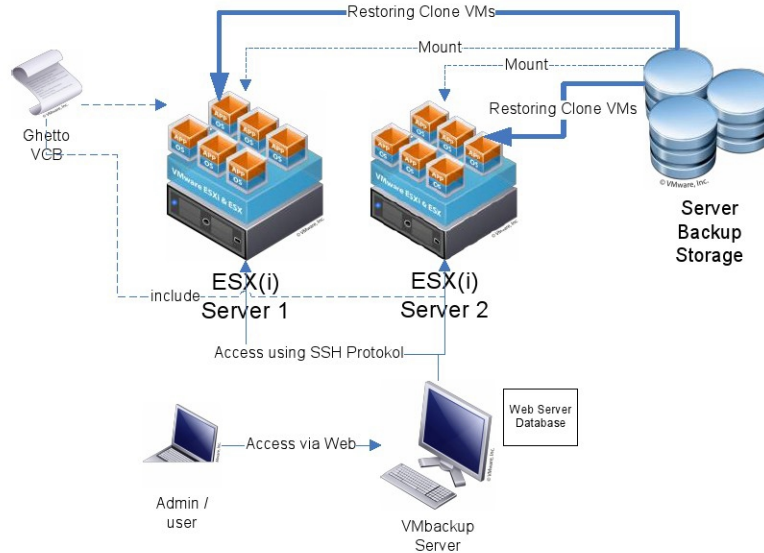


Figure 4. Work principle form restore system virtual machine

From Figure 4 can explain the working principle of the system in more detail as follows:

1. Users access the web server and provide the server command to restore broken virtual machine.
2. Then, web server execute remote commands for VMware via SSH protocol and then run a backup agent to restore virtual machine’s clone which contained in the backup storage server.
3. Agent runs restore program to restore Vmware virtual machine and place in ESXi Server that has been determined.

5. EXPERIMENT AND ANALYSIS

We build the system using PHP code[4] and here are the result of application virtual machine. Login page to access the configuration system shown on Figure 5.

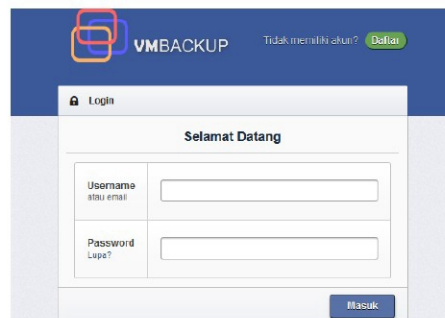


Figure 5. Login form

Configuration Page on Figure 6 would be shown to select to configure name of the virtual machine, backup format and rotation count.

VM Server	Select VM Server
Conf Name	Default <small>Note. Same name will update existing data.</small>
VM Backup Volume	/vmfs/volumes/datastore/BACKUPS
Disk Backup Format	Select an Option <small>Note. If you are using the 2gbparse on an ESXi 5.1 host, backups may fail.</small>
VM Backup Rotation Count	3

Figure 6. Configurationpage

One of the feature of the backup system is backup schedule that we can configure dynamicly shown on the Figure 7. And the history of the scheduling backup show on the Figure 8.

Scheduled

Hour: Date: Month: Day of Week:

VM Server: VMbackup Configuration:

VM Backup Volume : /vmfs/volumes/datastore/BACKUPS	Unmount NFS : -
Disk Backup Format : thin	NFS Server : -
VM Backup Rotation Count : 2	NFS Mount : -
Power VM Down Before Backup : No	NFS Local Name : -
Enable Hard Power Off : No	NFS VM Backup Dir : -
Iter To Wait Shutdown : 3	Email Log : 1
Power Down Timeout : 5	Email Server : -
Snapshot Timeout : 15	Email Server Port : -
VM Snapshot Memory : No	Email To : vnyx14@gmail.com
VM Snapshot Quiesce : No	Email From : -
Allow VM(s) With Snapshot to be Backedup : No	List Type : Whitelist
Enable Non Persistent NFS : No	Whitelist / Blacklist : vm_01

Figure 7.Scheduled backup page

VMs Backup History							
Search: <input type="text"/>			Show entries: 5				
No	IP Address	VMs	Request time	Start time	Finish Time	Status	Log
1	192.168.170.130	vm_02	2014-01-13 05:00:00	2014-01-15 06:11:01	2014-01-15 06:17:27	Success	
2	192.168.170.130	vm_01	2014-01-13 05:00:00	2014-01-15 06:00:10	2014-01-15 06:11:01	Success	
3	192.168.170.130	vm_01	2014-01-14 11:35:44	2014-01-14 11:35:44	2014-01-14 11:46:17	Success	
4	192.168.170.130	vm_01	2014-01-14 11:21:58	2014-01-14 11:21:59	2014-01-14 11:22:00	Failed	
5	192.168.170.130	vm_01	2014-01-14 11:17:21	2014-01-14 11:17:23	2014-01-14 11:17:24	Failed	

Showing 1 to 5 of 16 entries

First 1 2 3 4 Last

Figure 8. Backup’s result

Here are details of the testing that has been done, we try to compare the performance in different disk format of the virtualization (Table 1) and made some condition of the virtual machine to test the backup system (Table 2).

Table 1. Testing against different disk format of virtual machine

Format	Size VM	Backup Size	Provizioned Size	Duration Backup
Thin	4 GB	1.11 GB	4 GB	00:01:56
Zeroed Thick	4 GB	0.77 GB	4 GB	00:01:53
Eager Zeroed Thick	4 GB	4 GB	-	00:07:19

Table 2. Testing against condition of the virtual machine.

Test	Start Condition		Process Condition		Finish Condition		Backup Status
	Power	Snapshot	Power	Snapshot	Power	Snapshot	
1	On	No	On	Yes	On	No	Succeed
2	On	Yes	On	On	On	No	Succeed
3	On	Yes	On	Yes	On	Yes	Fail
4	On	No	Off	No	On	No	Succeed
5	Off	No	Off	No	Off	No	Fail

Note :

1. Test 1 used configuration “powering off No”, “hard powered off No”, “Allow VMs with snapshot backed up No”.
2. Test 2 used configuration powering off No”, “hard powered off No”, “Allow VMs with snapshot backed up Yes”.
3. Test 3 used configuration “powering off No”, “hard powered off No”, “Allow VMs with snapshot backed up No”.

4. Test 4 used configuration “powering off Yes”, “hard powered off Yes”, “Allow VMs with snapshot backed up No”.
5. Test 5 used configuration “powering off No”, “hard powered off No”, “Allow VMs with snapshot backed up No”.

Eager Zeroed Thick[5]: An eager zeroed thick disk has all space allocated and wiped clean of any previous contents on the physical media at creation time. Such disks may take longer time during creation compared to other disk formats. The entire disk space is reserved and unavailable for use by other virtual machines.

Zeroed Thick[5]: A thick disk has all space allocated at creation time. This space may contain stale data on the physical media. Before writing to a new block a zero has to be written, increasing the IOPS on new blocks compare to Eager disks. The entire disk space is reserved and unavailable for use by other virtual machines.

Thin[4]: Space required for thin-provisioned virtual disk is allocated and zeroed on demand as the space is used. Unused space is available for use by other virtual machines.

From test table in Table 1, Table 2, Table 3 and Table 4 can be analyzed that system was working as it should according to design system has been made . Both in testing the backup system or restore system .

Table 3. Testing against scheduled backup and restore.

Process	Backup status
Scheduled Backup	Succeed
Restore	Succeed

Table 4. Testing against different storage type.

Storage	Start Time	Finish Time	Duration
NFS	01:07:37	02:54:52	01:47:15
NFS (FreeNAS)	04:56:45	05:01:30	00:04:45
iSCSI (FreeNAS)	12:32:56	12:37:59	00:05:03

In Table 1 it can be seen that every disk format has a significant difference, for type zeroed thick and thin discs are not all sectors are in virtual machine backed up, but that only disk has data backed up so that it can speed up backup’s process. As for the type of eager zero thick all sectors are backed up so that backup’s capacity greater as original virtual machine. In Table 2 it can be seen that if the configuration enable parameter "Allow VMs with snapshots to be backups", snapshot on the virtual machine will be deleted. In Table 3 it can be seen that the scheduling system and restore virtual machine works well. In Table 4 it can be seen that significant differences occurred in the backup process using the 3 models of storage . On the NFS server backup process longer than NFS (FreeNAS)[8][10]/iSCSI (FreeNAS) because disk backup is shared. While on NFS (FreeNAS)/iSCSI(FreeNAS) is dedicated.

6. CONCLUSION

From the test results and analysis we conclude:

1. Backing up a virtual machine use thin disk or zeroed thick type faster than using the type of the eager zeroed thick.
2. This backup system will not work if virtual machine has snapshots. However, the system will create a snapshot that can be used when a failure occurs during backup process.
3. Scheduled system has worked well. Either restore system.
4. Physical ability of storage can affect performance backup process.

For the future works, this system would be implement for dynamic storage system and implement an algorithm to find out the best performance of the virtual machine backup environment.

REFERENCES

- [1] Xiaoying Wang , Zihui Dua, Yinong Chen, Sanli Li, "**Virtualization-based Autonomic Resource Management for Multi-tier Web Applications in Shared Data Center**", *Journal of Systems and Software*, Volume 81, Issue 9, Pages 1591–1608, September 2008.
- [2] Xiangzhen Kong, Chuang Lin, Yixin Jiang, Wei Yan, Xiaowen Chu, "**Efficient Dynamic Task Scheduling in Virtualized Data Centers with Fuzzy Prediction**", *Journal of Network and Computer Applications*, Volume 34, Issue 4, Pages 1068–1077, July 2011.
- [3] Hai Jin, Wei Gao, Song Wua, Xuanhua Shi, Xiaoxin Wub, Fan Zhou, "**Optimizing the Live Migration of Virtual Machine by CPU scheduling**", *Journal of Network and Computer Applications*, Volume 34, Issue 4, Pages 1088–1096, July 2011.
- [4] Lukmanul Hakim, **Membongkar Trik Rahasia Para Master Php**, Lokomedia, 2011.
- [5] **Thick vs Thin Disks and All Flash Arrays.**
<http://blogs.vmware.com/vsphere/2014/05/thick-vs-thin-disks-flash-arrays.html>. Accessed on 03-02-2014.
- [6] **Vmware Communities.**
<http://communities.vmware.com/docs/DOC-8760>. Accessed on 06-01-2014.
- [7] **Network File Sistem (NFS) + Samba.**
http://lecturer.eepis-its.edu/~idris/files/admin_jarkom/nfs_samba.pdf
Accessed on 27-01-2014.
- [8] **Belajar membuat server NAS dengan FREENAS.**
<http://iwing.wordpress.com/black-box/belajar-membuat-server-nas-dengan-freenas/>. Accessed on 06-01-2014.
- [9] **Configuring iSCSI for Vsphere esx(i).**
<http://www.everything-virtual.com/2013/05/configuring-freenas-8-3-iscsi-and-nfs-storage-for-vmware-vsphere/>. Accessed on 03-02-2014.

[10] **How to Setup Synologi NAS as VMware Server Datastore.**

<https://www.synology.com/en-uk/support/tutorials/508>. Accessed on 03-02-2014.