International Journal of Advanced Trends in Computer Science and Engineering

Expand the Capabilities for Backups a Paas-Type Virtualization System

Rosen Hristev[1], Magdalena Veselinova[2]

[1] University of Plovdiv PaisiiHilendarski, Plovdiv 4002, Bulgaria blv, 236, Bulgaria, hristev@uni-plovdiv.bg
[2] University of Plovdiv PaisiiHilendarski, Plovdiv 4002, Bulgaria blv, 236, Bulgaria, veselinova@uni-plovdiv.bg

ABSTRACT

Currently, resources in data centers are used extremely inefficiently. Storage systems are loaded on average about 25%, and servers and network resources - up to 30%. After implementing virtualization, the resource load level in a well-managed server environment increases by 30% to 90%. Virtualization undoubtedly provides many advantages in an infrastructure. One of the most important is the ability to easily create and manage backups of virtual machines, as well as quick recovery if necessary after disasters or accidents. Recovery time is many times faster than when applications and the operating system are hosted on a real server, while the loss of information with proper management is from zero to minimal. The available weekly and daily backups in Proxmox VE are not always flexible enough to properly organize backups in an IT infrastructure. In most companies and organizations there are virtual and real servers that play a significant role, but the data in them, as well as operating systems change very rarely. With existing methods, weekly backups need to be set up to ensure the reliability of the data and to recover quickly in the event of a disaster or accident. The paper aims to research and propose approaches which can extend the bult-in backup process by adding monthly backups for Proxmox VE. The research discusses the optimization of the process of creating backups to reduce network traffic between nodes and storage, as well as optimizing stored storage data.

Key words : Virtualization, backup, monthly backups, recover from backup.

1. INTRODUCTION

The process of business development leads immensely to the development of its IT infrastructure. Currently, resources in data centers are used extremely inefficiently. Storage systems are loaded on average about 25%, and servers and network resources - up to 30%. After implementing virtualization, the resource load level in a well-managed server environment increases by 30% to 90%.
Virtualization as technology allows reducing the cost of electricity, hardware, reducing the physical space required for storage of computer equipment, allows easy scalability and most importantly, in times when the financial value of the hardware is many times less than the value of the data are stored on it, with a properly built infrastructure virtualization allows quick recovery to normal operation. Research about optimizing server resources is made in [1]. Good performance evaluation is made in [2] for three selected hypervisors one of which is the subject of our research. In [3] it is provided and appropriate assessment to five well-known hypervisors built in virtualized cloud computing servers.
In order to achieve the goal of rapid disaster recovery, it is necessary to have properly structured backups so as to ensure minimal loss of information.

2. BACKUP VIRTUAL MACHINES

The implementation of virtualization in IT infrastructures is becoming common practice for companies of different sizes. The first environments for backup and recovery were not designed for virtual environments, and the recovery process for physical and virtual servers is different.
Virtualization undoubtedly provides many advantages in an infrastructure. One of the most important is the ability to easily create and manage backups of virtual machines, as well as quick recovery if necessary after disasters or accidents. Recovery time is many times faster than when applications and the operating system are hosted on a real server, while the loss of information with proper management is from zero to minimal.
According to VMware, the main problems of backing up virtual environments are the large volumes of backups (because it is more convenient to save all virtual machines at once) and their verification, ensuring the integrity of data from a variety of applications and managing processes in a complex virtual environment, especially when the backup has to cover a growing number of virtual machines. With an existing backup mechanism, issues related to the integration of different solutions inevitably arise. Companies face the dilemma: whether to back up virtual infrastructure will be performed by the same system that is already used for the physical environment, or should a separate one be created. Finally, areas of responsibility often need to be shared between infrastructure owners and applications.

Regardless of the used hypervisor, for example, Hyper-V, VMware or Proxmox, the approaches to virtual machine backup (VM) are the same. Backup can be done at the host system level (or as it is also called at the hypervisor level) or at

the virtual machine level. Some products even provide the ability to run a virtual machine directly from the backup, which reduces disaster recovery time. However, in order not to lose all data, if a VM recovery fails, this option should provide for a backup replica.

Virtual machine-level backups are usually made in the same way as physical server backups. If this method of backup is used, then the infrastructure largely loses the advantage of fast recovery after a crash, and the process of returning to normal work is the same as with a real server.

In this research, we will consider the built-in methods for creating backups in Proxmox VE, as well as describe an approach to extend the built-in features.

3.BUILT-IN BACKUP OPTION FOR PROXMOX VE

Proxmox Virtual Environment (Proxmox VE) is an open-source virtualization environment that is Debian Linux based with a modified RHEL kernel. Proxmox VE was developed by Dietmar and Martin Maurer for the reason that OpenVZ does not offer a backup tool and no graphical user interface management. The first public version of Proxmox was released in April 2008, at which time it was one of the few platforms to offer full virtualization, managed by a web user interface, similar to paid virtualization environments.A detailed presentation of the capabilities, setup and configuration of Proxmox VE is made in [4] and [5].

Proxmox VE comes with a built-in virtual machine backup implementation, which is available on local directories, GlusterFS, NFS, CIFS, Proxmox Backup Server and CephFS. In order to use the host-level backup functions of any of the above storage types, it must be specified that its contents may also contain a "VZDump backup file". The "Max Backups" variable determines how many backups will be saved for each machine. If there is "0" then there is no limit to the number of backups. By default, the backup settings are stored in "Datacenter" -> "Backup", and there are the settings for all physical servers that are in the cluster.

Figure1: Create backup

The "Add" button opens the backup manager, as it is shown in Figure 1. By default, the list shows all virtual machines that are in the cluster. From the "Node" drop-down menu, you can choose to limit to virtual servers that are located on only one physical server. "Storage" controls which of the available disk spaces that are allowed to contain the VZ Dump backupfile will be backed up. The next drop-down menu selects which day of the week the backup will be made, and more than one day can be selected, ie. Daily and weekly backups are allowed by default.

"Start time" determines the time at which the backup procedure will begin. There is no universal formula, but in the general case, backups are made at the least busy hours.

The "Selection Mode" selects whether the virtual machines marked in the list will be backed up or not. The third option here allows all machines to start backing up at the same time, which is not recommended due to the high load that follows. How many backups to start making at the same time is determined by the infrastructure and organization.

If an SMTP server is set up, Proxmox VE can notify you via email when backups are made or when for some reason they cannot be made.

Proxmox VE supports 3 options for compressing the backed up copies, as we are not obliged to use compression and we can keep the copies unfolded, LZO is also supported, which is weaker compression, its advantages are that it will not load the server as much and will backup virtual machine faster. With GZIP, the encryption algorithm will load the CPU while making a copy, so it will be slower but take up less disk space. There are three types of backup modes. With "Snapshot", the virtual machine will continue to run and its applications can be accessed, an invisible backup will be made without affecting the functioning virtual server, this is the slowest method to back up. In the "Suspend" method, the virtual machine will be paused, and the machine's RAM will be saved after the backup is created, and the virtual server will continue to run from where it left off. While the backup is being created, the machine and its applications cannot be accessed. With the "Stop" method, the virtual machine will be turned off and after a backup is created, it will be started automatically. This is the fastest backup method.

The "Enable" checkbox determines whether these backups will be active, ie. whether to perform.

In the management of Proxmox VE, when visiting the virtual server in the "Backup" menu, the existing backups for the virtual machine can be found, and from there a backup can be made at the moment. In order to make a backup, the number of existing ones must not exceed the number allowed by the storage, if the number is exceeded, then one of the existing ones must be deleted. For backups configured by "Datacenter", the oldest backup copy is automatically deleted if the maximum number of allowed stored copies of the storage is reached.
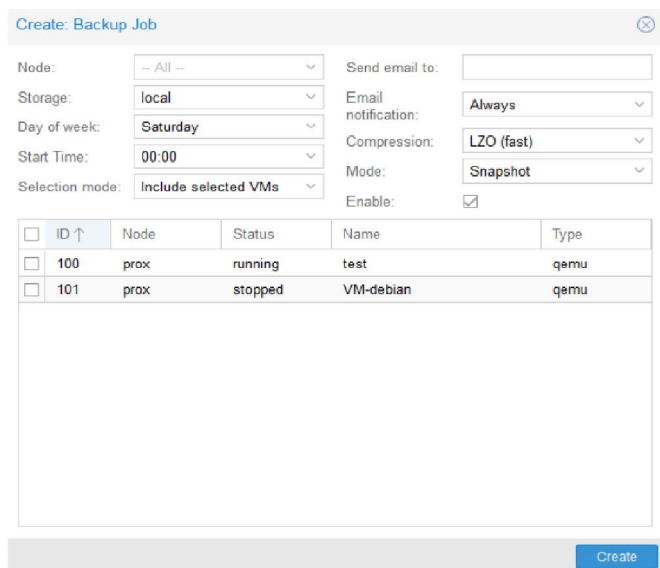
4. EXPANDING THE POSSIBILITIES FOR CREATING BACKUPS IN PROXMOX VE

The available weekly and daily backups in Proxmox VE are not always flexible enough to properly organize backups in an IT infrastructure. In most companies and organizations there are virtual and real servers that play a significant role, but the data in them, as well as operating systems change very rarely. With existing methods, weekly backups need to be set up to ensure the reliability of the data and to recover quickly in the event of a disaster or accident. Examples of such systems are external DNS - used to service Internet requests - when a domain/subdomain is searched, they return an IP address where the corresponding resource can be found. Proxy servers - "hide" where the server is actually located, which should be accessible, as requests go through them. If either of the two machines is unavailable in the general case, the user will not be able to reach the required resource. Virtual servers hosting corporate websites used as business cards. Other machines that are important but with rarely changing data are some of the servers that serve industrial machines.

Assuming that there is an external DNS server and a proxy server in infrastructure, both of which use one hard drive with a compression size of 30 GB, then with weekly backups for each of the virtual machines will be written to the storage 120 GB or for a year 1.44 TB or for both servers a total of 2.88 TB, which is unnecessary network load and hard drive waste. The goal is to create scripts for automatic monthly backup of virtual machines of this type, thus for the same infrastructure for one year for both virtual machines the amount of information that will be written to the disks will be 660 GB.

To organize a flexible and efficient method for backing up virtual servers in an IT infrastructure, a method for creating monthly backups in Proxmox VE virtualization environment has been developed. The developed method includes seven configuration files, one for each day of the week, two bash scripts and one cronjob.

/etc/cron.d/ contains a file called monthlybackup. Figure 2 below shows its contents.

```
00 03 1-7 * * root /opt/custombackup/checkday.sh >>
/var/log/montlybackup.log 2>&1
```
Figure 2: Cronjob to start monthly backup

The script will run every day from the first to the seventh date of each month at 3:00, and this will run with the rights of the superuser root script located in /opt/custombackup/ named checkday.sh, with the answers that "returns" the script will be saved in /var/log/monthlybackup.log.

Bash scripts and configuration files are stored in the /opt/custombackup/ directory.

The contentsof the first bash script - chekday.sh is shown in Figure 3.

```
#!/bin/bash
currentdir=$(dirname $0)
echo "Montly backup script for Proxmox VE"
day=''
if [ `date +%e` -le 1 ] && [ `date +%u` == 1 ]
then
```

```
   day='monday'
elif [ `date +%e` -le 2 ] && [ `date +%u` == 2 ]
then
   day='tuesday'
elif [ `date +%e` -le 3 ] && [ `date +%u` == 3 ]
then
   day='wednesday'
elif [ `date +%e` -le 4 ] && [ `date +%u` == 4 ]
then
   day='thursday'
elif [ `date +%e` -le 5 ] && [ `date +%u` == 5 ]
then
   day='friday'
elif [ `date +%e` -le 6 ] && [ `date +%u` == 6 ]
then
   day='saturday'
else
   day='sunday'
fi
$currentdir/backup.sh $day
```
Figure 3: Contents of checkday.sh

The script checks for which day of the week is `date +% u` returns the day as 1 is Monday and 7 is Sunday. Once the day is discovered, the variable day accepts the day as value, and then a script called backup.sh is called, which is in the same directory with the argument to the variable day.

In order for checkday.sh and backup.sh to be executed, they must have execution rights.

```
#!/bin/bash
currentdir=$(dirname $0)
day=''
if [[ $# -eq 0 ]]
then
  day='error'
  msg="You must specify day of week"
else
if [[ $1 == 'monday' || $1 == 'tuesday' || $1 == 'wednesday' ||
$1 == 'thursday' || $1 == 'friday' || $1 == 'saturday' || $1 ==
'sunday' ]]
  then
    day=$1
  else
    day='error'
    msg="Incorrect argument: $1"
  fi
fi
# Error check
if [[ $day == 'error' ]]
then
  echo ""
  echo "Error: $msg"
  exit 1
fi

echo `date +"%Y-%m-%d %T"`": running backup script
for the first $day of month"

VMids=`head -n1 $currentdir/$day.conf 2> /dev/null`
```

```
storage=`sed -n '2p' < $currentdir/$day.conf`
compression=`sed -n '3p' < $currentdir/$day.conf`
mode=`sed -n '4p' < $currentdir/$day.conf`

#check variables
if [[  $VMids -eq 0  ]] || [[  $storage -eq 0  ]] || [[
$compression -eq 0  ]] || [[  $mod -eq 0  ]]
then
    echo "You don't have configuration set for $day or your
configuration is not correct."
    echo "Please check $currentdir/$day.conf."
    echo "On the first row of the file you need to write down
the IDs of the VMs who you want to backup. On the second
row, you need to have the exact place to store the backup.
On the third row you need to have the compression
algorithm. On the fourth row you need to have the
archiving mode."
    exit 1
else
    echo "Script is running for $VMids Virtual Machine/s."
    echo "Backup is storing in the $storage storage, the
compression is $compression and the mode is $mode ."
fi

#do backup
ionice -c3 /usr/bin/vzdump $VMids --storage $storage
--compress $compression --mode $mode --quiet 1
```

Figure 4: Contents of backup.sh

When starting, backup.sh first checks if it is not started without an argument and whether the argument with which it was started is part of its list, then in /opt/custombackup or the directory in which it is placed it looks for the configuration file for the current day to be [day-of-the-week].conf. The contents of backup.sh is shown in Figure 4.

Then the data from the configuration of variables are assigned, the correctness of the data in the variables is checked and the vzump as an idle process is started.

VZDump is a software solution for archiving virtual machines, it is the application built into the installation of Proxmox VE through which the system organizes its backups. VZDump is fully compatible with all types of storage and guest operating systems.

The [day-of-the-week].conf file contains information about the configured virtual machines on which the archive will be made, on which of the available storages in the Datacenter the archive will be stored, it also contains a compression algorithm and what method will be used for the backup.

In the first line of the configuration are placed the IDs of the virtual machines on which the archive will be made. They should be separated by a space between them. The first line can also be "--all" which will back up all virtual machines.

The second line of the configuration contains the disk space on which the archive of the virtual machines will be stored. The storage must be added in advance in "Datacenter" -> "Storage", and its content must be allowed to contain "VZDump backup file."

The third line in the configuration defines the compression algorithm by which the archive will be created and stored, and

the possible options are:
- 0 - the virtual hard disk of the machine will be stored on the specified storage without compression, takes up too much space;
- LZO - fast compression;
- GZIP - slow compression, the file takes up less physical storage space, but more compute resources are used to compress it.

Table 1: Comparison of compression algorithms in archiving.

Algorithm	Archive size	Time to create the archive
0 (without compression)	4.98 GB	00:11:28
LZO	2.84 GB	00:03:03
GZIP	2.34 GB	00:07:00

Table 1 describes the size of the archives for the different types of compression of a virtual machine with a 10 GB hard disk, of which 3.3 GB is actually used.

The fourth line of the configuration file describes the method by which the configured backup is made. The line may contain:
- Stop - in this way the virtual machine or container is turned off while backing up. This leads to a long downtime.
- Suspend - This method uses rsync to copy the hard disk of the virtual server to a temporary location. After the backup is created, a second rsync is executed to copy the changed files, after which the virtual machine is started (returned) again. This method has minimal downtime, but requires more free disk space to store a copy of the disk.
- Snapshot - a way in which there is no downtime.

```
100 101
local
lzo
snapshot
```

Figure 5: Example for configuration file of Proxmox VE monthly archive system

The configuration file shown in Figure 5 will be backed up to virtual machines with IDs 100 and 101, which will be stored on local storage, the compression algorithm to be used is LZO, and the selected backup method is snapshot.

5. RECOVER A VIRTUAL SERVER FROM A BACKUP USING PROXMOX VE

Recovery of a virtual machine from an archive that is done using the extended monthly backup method or the built-in method is done in the same way, as in both methods the backups are made using VZDump. Recovery is done through the management web interface.

In the virtual machine itself, there is a menu for managing archives - "Backup", like it is shown below in Figure 6.
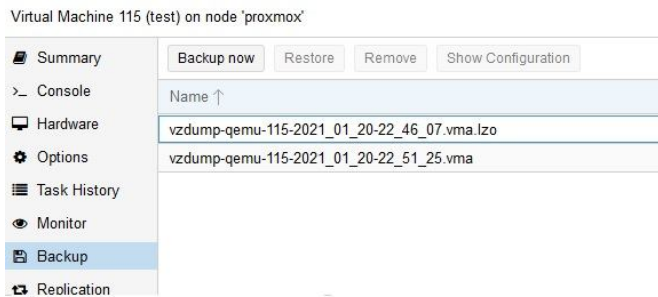
Figure6: Example for configuration file of Proxmox VE monthly archive system

An instant archive of the virtual machine can be made from the menu, and in order to be made, the storage limit for the number of saved archives must not be reached. If it is reached, one of the existing backups for the respective virtual machine must be manually deleted.

The backups are called vzdump-qemu-<id-virtual-server>-<year>_<month>_<date>-<hour>.

From the "Show configuration" button you can see the hardware configuration for the virtual server, which will also be restored.

To restore a virtual machine, it must first be turned off. Once the process is complete and the machine is started, it is loaded with the recovered files and settings.

6. IMPLEMENTATION AND TESTING OF SCRIPTS IN INFRASCTUCTURE

The scripts for expanding the system for automatic monthly archives have been successfully implemented and tested in 6 infrastructures. The implementation was achieved by the following algorithm:

1. A custombackup directory has been created in /opt/;
2. The scripts checkday.sh and backup.sh from the paper can be created in the newly directory custombackup, with executable rights;
3. .conf files are created in the directory as described in the previous point;
4. A cronjob has been created in /etc/cron.d/ which runs every month from the first to the seventh date, which starts checkday.sh, saving the execution logs in /var/log/ montlybackup.log.

The developed system has been tested on Proxmox VE 6.2 Proxmox VE 5.4 and Proxmox VE 3.1 on all three different versions, the implementation and subsequent testing passed successfully without problems.

7. CONCLUSION

Virtualization as technology plays a major role in IT infrastructure optimization, can maximize the use of resources and reduce the cost of electricity, hardware, physical space required for storage of computer equipment. Due to the high cost of a license, we consider open source-based virtualization like Proxmox VE. We consider the built-in methods for creating backups in Proxmox VE. Three options for compressing the reserved copies are considered. Three types of backup modes are described. A method for creating monthly archives in the Proxmox VE virtualization environment has been developed for organizing a flexible and efficient method for archiving virtual servers in a given IT infrastructure. A comparison of compression algorithms in archiving has been made. An approach for implementation of the considered method for archiving of virtual servers is described. Recovering a virtual server from a backup is discussed.

ACKNOWLEDGEMENT

REFERENCES

1. Edward Alia, Susandrib, Rahmaddenia, **Optimizing Server Resource by Using Virtualization Technology**, *Procedia Computer Science*, Volume 59, 2015, Pages 320-325.
2. Sultan Abdullah Algarni, Mohammad Rafi Ikbal, RoobaeaAlroobaea, Ahmed S. Ghiduk and Farrukh Nadeem, **Performance Evaluation of Xen, KVM, and Proxmox Hypervisors**, *International Journal of Open Source Software and Processes (IJOSSP)* 9(2), 2018.
3. Bao Rong Chang, Hsiu-Fen Tsai, Chi-Ming Chen, **Evaluation of Virtual Machine Performance and Virtualized Consolidation Ratio in Cloud Computing System**, *Journal of Information Hiding and Multimedia Signal Processing*, Volume 4, Number 3, July 2013.
4. Goldman, R., *Learning Proxmox VE*, Packt Publishing Ltd, 2016.
5. Simon M. C. Cheng, *Proxmox High Availability*, Packt Publishing Ltd, 2014.