VMWARE Virtualization - Physical to Virtual Migration

SUBASH THOTA
Data Architect& Division of Information Systems
Ashburn Virginia, USA

Abstract - *Resource optimization is one of the key drivers for deploying* **virtualization technology**. *Regardless of whethertheresourcesarecurrentlyunder-utilizedorover-utilized,itinfluencestheBusiness.Virtualization offers the potential for a reduction in total server hardware and better management of that hardware, but it is not without its own caveats. There are both advantages and disadvantages to using virtualization in any environment,anditiscriticalthatyouunderstandtherealityofw hatvirtualizationcanofferandreconcilethat with your expectations of how it can be used in yourenvironment.For example, you should not assume that applications are suddenly going to require fewer resources just because they are virtualized. On the contrary, the use of virtualization adds overhead, and virtualized applicationsoftenusemoreresourcesthanbefore.Theactuala mountofoverheaddependsonanumberof factors including the type of application, which virtualization engine is being used, what kind of hardware is available, and how it will be used. Before implementing virtualization you need to make sure that you have enough storage bandwidth/space, memory, CPU, network bandwidth and other resources to handle the applicationsANDthevirtualizationoverhead.Whenyouexhau stanysingleresourceonaserverthatisused for virtualization, all guest operating systems may be impacted and seen to wait for that single resource. Thus, low CPU or memory utilization alone is not an indication that additional work can be added - youmust verify that all required resources have availablecapacity.*

Keywords- **Big Data Analytics, Social Analytics, Storage Analytics, Data Management, Information Quality, Data Mitigation, Metadata, Data Profiling, VMWARE**

Introduction

Virtualization is a proven software technology that makes it possible to run multiple operating systems and applications on the same server at the same time. It's transforming the IT landscape and fundamentally changing the way that people utilize technology.Operating system virtualization is the use of software to allow a piece of hardware to run multiple operating system images at the same time.There are three areas of IT where virtualization is making headroads, network virtualization, storagevirtualization and server virtualization:

A virtual computer system is known as "virtual machine" (VM): a tightly isolated software container with an operating system and application inside. Each self-contained VM is completely independent. Putting multiple VMs on a single computer enables several operating systems and applications to run on just one physical server, or "host".A thin layer of software called a hypervisor decouples the virtual machines from the host and dynamically allocates computing resources to each virtual machine as needed.
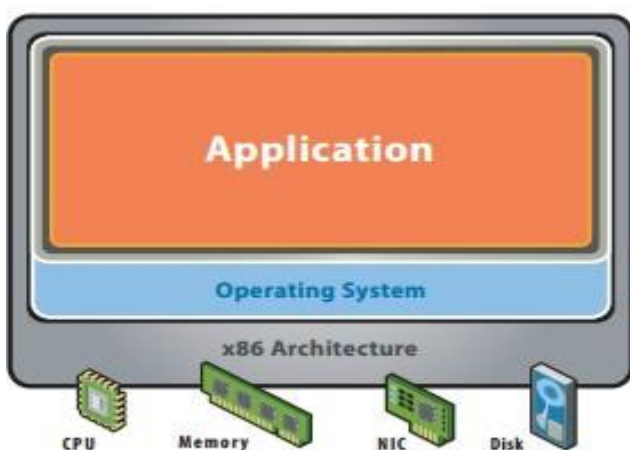
Before Virtualization:
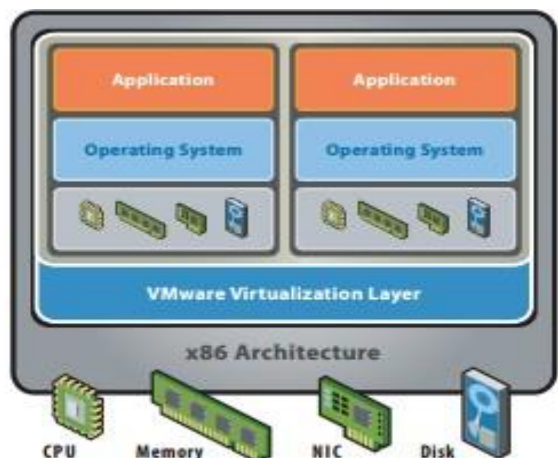
After Virtualization:

Advantages

Some common advantages offered by virtualization include:

- Improvement of overall resource utilization while providing

room for improved performance forpreferred applications (when larger/faster hardware is used forvirtualizationthanwas used for standaloneservers)

- More efficient use of existing servers - which may reduce

datacenter space requirements (for both the servers themselves and for their associated monitors and peripherals) and, by extension, the power and cooling costs for thosedatacenters.

- Ability to perform hardware and software maintenance on secondary systems, thereby minimizing (or completely avoiding) downtime on productionsystems
- Ability to easily create backups of existing systems can help support business continuity and aid in disasterrecovery
- Faster deployment of new logical servers and safer/faster migration of applications/infrastructure maximizes change control flexibility and minimizes disruption to productionsystems
- Ability to isolate applications in separate virtual environments can simplify support and reduce the likelihood of "application collisions" or incompatibilities created when several applications are hosted on onesystem

Disadvantages
Some disadvantages associated with virtualization include the following:

- Virtualization overhead can impact the performance of key workloads and virtualization generally makes workloads less predictable (especially when larger/faster hardware is not deployed for the virtualization deployment)
- New, relatively unproven abstraction layers may introduce potential security risks
- Consolidation of servers, as well as storage virtualization, may create a larger, single point of failure - making it easier to lose significant volumes of data
- Resources STILL need to be managed. If mismanaged the results are more far-reaching,and resource bottlenecks escalate at a more rapid pace
- Some vendors do not fully support virtualization,and you may need to reproduce a problem on a non-virtualized system to facilitate diagnosis (especially where specialized hardware drivers are involved).
- Deploying and managing virtualized environments requires skills that are less common but more in demand

ESXi Hardware Requirements
Refer   http://www.vmware.com/resources/compatibilityfor hardware compatibility.

ESXi Installation and Configuration Checklist
1. Change Boot order in BIOS as per requirement of booting device,e.g. SAN or Local Storage, iSCSI disk,etc.
2. Prepare Minimum 1 GB boot LUN from external storage if applicable. Otherwise, use localstorage.
3. Ready with VMWARE ESXi ISO (OR DVD) boot image or Vendor's specific customized ISO (OR DVD) bootimage.
4. It is considered that there is a valid FC/SATA connectivity between Physical Host andStorage.
5. Configure IP Address, Hostname, VLAN, DNS and Domain using DCUI interface ofESXi.
6. DisableIPv6.
7. ConfigureNTP
8. Disable SSH Service and change it to start and

stopmanually
9. Disable Shell Service and change it to start and stopmanually
10. Configure scratchpartition
11. Add license to ESXi server by login using vSphere client5.5.
12. Create DNS Host entry in exiting Domain for new ESXiServer
13. Add ESXi server in Domain so that Domain users / member of groups can login ESXi using Domain ID.
14. Add ESXi host in VCenter (if exist) outside the cluster withhostname.
15. Perform ESXi patching using VMware update manager. If no update manager available then perform manualpatching.
16. Reserve three IP Addresses for Management, Vmotion,andFT.
17. Add ESXi host in DV-switch on both Management and DataVLAN
18. Configure vmkernel ports for VMotion andFT
19. Get all Datastore of cluster mapped to new ESXihost
20. Add ESXi server inCluster
21. Test VMotion on new ESXihost

P2V Migration Checklist
Pre-Check List
1. Destination datacenter in DC is ready with appropriate VMware, Network, and storagesetup.
2. Use Local Administrator credential duringP2V
3. Verify and Get artifacts of Application and Database Status before initiating P2V/V2VActivity.
4. Source servers identified for migration will be cold booted first as per the current shutdown/reboot procedure to ensure the Server/Servicesconsistency.
5. The source machine should have 200 MB of free space on system disk (for VMware to capture the snapshots of the localmachine).
6. Make a record of the service status of source machine for post migrationcomparison.
7. Make a copy of Event logs for post migrationcomparison.
8. Ensure the TCP Ports 902,443, 445, 9089, 9090, 139 are opened between the source and destination machines and not being used by any otherapplication.
9. Ensure the UDP ports 137,138 are opened between the source and destination machines and not being used by any otherapplication.
10. Check that the data store on destination machine has sufficientspace.
11. Make sure that the local file sharing is turned off to ensure required ports are free for VMware migration.
12. Remove all the unwanted files from the source systemdisk.
13. Windows firewall should not block file and printersharing.
14. Stop all the antivirus services as it will cause the migration to be gotdelayed.
15. Stop all the application services communicating to the outer world (Isolation mode, the server shouldbe available for DC migration activity only over thenetwork)

16. Perform the graceful shut down of the database and Stop the Listenerprocess
17. Run VMware converter to start themigration
18. VM hostname provides without FQDN,E.g.<HOSTNAME01>. Remove <Domain>.com from hostname.
19. It is recommended that to select "Thick provision" during HDD layoutconfiguration.
20. Complete theMigration.

Post-Check List

1. Event log check for any new error other than existing onthe source.
2. Change the No of NIC, if needed as per new IP schema. Use VMXNET3 as a VMNIC.
3. Remove "USB device" from VM after P2Vcompleted.
4. Remove "Audio device" from VM after P2Vcompleted.
5. Remove "Floppy drive" from VM after P2Vcompleted.
6. Enable Memory and CPU hotplug on the VM. It is not recommended for Windows 2003 & older version.
7. Disable network card before powered onVM.
8. Reboot the migrated machine in safemode.
9. Reboot the migrated machine in normal mode and check whether any issues inlogin.
10. If there is anissue, then try to disjoin and join Domain. It is only recommended before VMgo live.
11. Remove all the unwanted hardware related software. (i.e., Hardware Vendorspecific)

12. Remove network cable from the Physical server.
13. Assign existing Physical server IP address to the VM and enable network card from VMware end.
14. Update the scripts or scheduled job, if applicable.
15. Check the status of license key if required, update as appropriate.
16. Install SSL certificates if applicable.
17. Install VMware tools.
18. Check that the machine is ok and there is no BSOD (Blue screen of Death) followed by areboot.
19. Check that the resource utilization for OS is normal.
20. Check the event log for any new error that might have arisen post-migration.
21. Check that network is properly accessible and all the required ports over the network are open.
22. Check that the application is accessible.
23. All the user accounts are migrated properly.
24. All the required services are up and running.
25. Finally, check network connectivity and RDP of the server.

Only for DB Servers

1. Update or Verify listener, tnsnames,andsqlnet configuration files and confirm the remote databaseconnectivity.
2. Perform database sanity to ensure databaseintegrity.
3. Check that the machine is ok and there is no BSOD (Blue screen of Death) followed by are boot.

Compare Virtualization

Overview	VMware vSphere Enterprise Plus	Microsoft Hyper-V Server 2008 R2 SP1 Enterprise	Citrix XenServer Enterprise Edition
Product Name	vSphere	Hyper-V	XenServer
Edition	Enterprise Plus	Enterprise	Enterprise Edition
Company	VMware	Microsoft	Citrix Systems, Inc.
Intended User	Enterprise	Small-Medium Business and Enterprise	Enterprise
Hypervisor Type	Bare Metal (Type 1)	Bare Metal (Type 1)	Bare Metal (Type 1)
Virtualization Type	Full Virtualization Hardware Assisted Virtualization Para virtualization	Full Virtualization Hardware Assisted Virtualization	Hardware Assisted Virtualization Para virtualization
Architecture	x86, x64	x86, x64	x86, x64, Power

Supported Storage	DAS FC FCoE iSCSI NAS NFS SSD for Swap	DAS FC Firewire iSCSI PATA SAS SATA SCSI USB	DAS eSATAiSCSI NAS NFS SAS SATA SCSI USB
Virtual Machine Limits	Virtual Disk Size - 2 TB	Virtual Disk Size - 2 TB	Virtual Disk Size - 2 TB
	RAM per VM - 1 TB	RAM per VM - 64 GB	RAM per VM - 128 GB
	Virtual CPUs per VM -32 vCPUs	Virtual CPUs per VM - 4 vCPUs	VirtualCPUsperVM- 16 vCPUs
	Virtual Disks per VM - 60	Virtual Disks per VM - 256	Virtual Disks per VM - 16
	Virtual NICs per VM - 10	Virtual NICs per VM - 12	Virtual NICs per VM - 7
Host Server Limits	VMs per Host - 512	VMs per Host - 384	VMs per Host - 130
	RAM per Host - 2 TB	RAM per Host - 1 TB	RAM per Host - 1 TB
	Virtual Disks per Host - 2048	NA	Virtual Disks per Host - 512
	Logical CPUs per Host - 160	Logical CPUs per Host - 64	Logical CPUs per Host - 160
	Virtual CPUs per Host - 2048	Virtual CPUs per Host - 512	Virtual CPUs per Host -4000
	Virtual CPUs per Core - 25	Virtual CPUs per Core - 12	NA

Compare Virtualization (Continued)

Overview	VMware vSphere Enterprise Plus	Microsoft Hyper-V Server 2008 R2 SP1 Enterprise	Citrix XenServer Enterprise Edition
Management Features	Capacity Planning/Management ChangeReports Dynamic Resource Allocation HighAvailability Live Migration P2VConversion Performance Metrics Performance Reports Power Management Real-Time Alerts SharedResourcePools Storage Migration ThinProvisioning VM Cloning VM Migration	Capacity Planning/Management Dynamic Resource Allocation Failover High Availability Live Migration Performance Reports Storage Migration	Asset Management Capacity Planning/Management Configuration Snapshots HighAvailability Live Migration Performance Reports PowerManagement Storage Migration Thin Provisioning VirtualFirewall

Supported Guest Operating System

Overview	VMware vSphere Enterprise Plus	Microsoft Hyper-V Server 2008 R2 SP1 Enterprise	Citrix XenServer Enterprise Edition
Supported Guest Operating Systems	Free BSD Mandrake Linux MS-DOS NovellLinuxDesktop Novell Netware Server Red Hat EnterpriseLinux AS Red Hat EnterpriseLinux ES Red Hat Enterprise Linux WS Red Hat Linux Small Business Server 2003 Solaris x86 Platform Edition Sun Java Desktop System SUSE Linux SUSE Linux Enterprise Server Turbo Linux Enterprise Server Turbo Linux Workstation Windows 2000 Professional Windows 2000 Server Windows 95 Windows 98 Windows for Workgroups Windows Me Windows NT Server WindowsNTTerminalServerEdition Windows NT Workstation Windows Server 2003 Enterprise Edition Windows Server 2003 Standard Edition Windows Server 2003 Web Edition WindowsXPHomeEdition Windows XPProfessional	Mandrake Linux RedHatEnterpriseLinuxAS RedHatEnterpriseLinuxES RedHatEnterpriseLinuxWS Red HatLinux SUSE Linux SUSELinuxEnterpriseServer Turbo Linux EnterpriseServer Turbo Linux Workstation Windows2000Professional Windows 2000 Server Windows 95 Windows 98 Windows for Workgroups Windows Me Windows NT Server WindowsNTTerminalServerEdition WindowsNTWorkstation Windows Server 2003 Enterprise Edition WindowsServer2003StandardEdition WindowsServer2003WebEdition WindowsXPHomeEdition Windows XP Professional	Mandrake Linux Novell Linux Desktop Red Hat Enterprise Linux AS Red Hat Enterprise Linux ES Red Hat Enterprise Linux WS Red Hat Linux SUSE Linux SUSE Linux Enterprise Server Turbo Linux Enterprise Server Turbo Linux Workstation Windows 2000 Professional Windows 2000 Server Windows 95 Windows 98 Windows for Workgroups Windows Me Windows NT Server WindowsNTTerminalServerEdition Windows NT Workstation Windows Server 2003 Enterprise Edition Windows Server 2003 Standard Edition WindowsServer2003Web Edition WindowsXPHomeEdition Windows XPProfessional

Issues and Mitigations (Servers)

S.No	Issues	Mitigation
1	P2VmigrationofADServerusingVMwareToolisnot recommended byVMware	New O.S. Win2K8 commissioned and AD / DNS Role migrated - (http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1006996)
2	P2V migration slowness observed due to VLAN issue. Migration was started with 5 Mbps speed.	TherewasVLANMappingissue.ESXiconnectedtoDMZ VLANsothatmigrationcompletedwith20Mbpsspeed.
3	Shares were inaccessible due to Hostname dependency formorethan1200sharedlinksafterP2VMigration	Hostname swapping activity performed with migrated VM server
4	NotabletologintorootcertificateServerusing DomainIDafterP2Vmigration.Unabletoinstall McAfeeAntivirus.Unabletoupdatediskdrivers.	Removedserverfromthe domain,removedoldroleCertificate Servicesfromthe server,rejoinedtheservertothe domain,loginwithdomainID,reinstalledroleCertificateServ iceson a serverwitha certificate,restoredtheCertificateAuthority configurationfromthebackup.
5	The serverresponsewasveryslowafterP2Vmigration, becauseofsomeDMZportblockissue	Investigatedtherequiredportandopeneditoverthe firewall.
6	Insufficient free space on Target LUN during P2V	Disksizereducedbycustomizingdiskconfigurationsetting duringP2V
7	The applicationwasnotworkingafterP2VusingsameIP configuration	IPaddresschangeafterP2VMigration,SinceApplication license,wasnothardcodedwithIPanditisbasedon FQDN.
8	DevicemanagerwasnotworkingonPhysicalserver henceonVMserverafterP2V.	DevicemanagerissueresolvedbyrepairingcorruptedDLL files,and then once again P2V activity performed successfully.
9	NetworkloopissueatswitchlevelafterP2Vmigratio n	SeparatedefaultgatewaysassignedoneachNICcard havingasperVLANandsubnetmask
10	UnabletocreateaVSSsnapshotofthesourcevolume duringP2V.Migrationfailedto1%after40Minutes	P2VConvertertoolreplacedwithv5.1andP2Vmigration completed

Issues and Mitigations (General)

S.No	Issues	Mitigation
1	Unable to query the live Linux source machine for SUSE 11	SUSE Linux 11 with UEFI is not supported by VMware converter, need to perform a backup restore
2	Unable to obtain hardware information for selected machine using Standalone VMware converter 5.5 and 6.0	This issue resolved by using VMware Standalone converter tool 5.1
3	IP address already in use error on a converted virtual machine (1012255)	Remove duplicate hardware from Registry as per below link (http://kb.vmware.com/selfservice/search.do?cmd=displayKC&docType=kc&docTypeID=DT_KB_1_1&externalId=1 01 2255)
4	MacAfee antivirus caused an issue during P2V migration	Antivirus is not getting updated post P2V requires reinstallation. We get faster speed during migration when antivirus is removed.
5	Virtual disk drivers were not updating after P2V due to McAfee Antivirus	Uninstalled the antivirus then restarted the server and updated virtual disk drivers successfully then and reinstalled the Antivirus
6	VM Converter created E1000 Network adapter after P2V by default which causes an issue for Network bandwidth and NIC working functionality.	Removed NIC card settings during P2V migration and installed separately VMXNET3 NIC card after P2V migration.
7	Windows 2003 Server was failed to boot after P2V migration	Disabled Memory\CPU hotplug option on VM as some Windows 2003 OS is not supported this feature
8	The server was unable to login using Domain ID but able to login using local Admin after P2V migration.	Most of the servers were disjointed and joined from the AD and issue resolved.
9	When tried to set the IP address on a network adapter, the following error message received --> The IP address XXX.XXX.XXX.XXX you have entered for the network adapter is already assigned to another adapter Name of the adapter. Name of adapter is hidden from the network and Dial-up Connections folder because it is not physically in the computer or is a legacy adapter that is not working. If the same address is assigned to both adapters and they become active, only one of them will use this address. This may result in incorrect system configuration. Do you want to enter a different IP address for this adapter in the list of IP addresses in the advanced dialog box?	There is a hotfix available to resolve the issue. Refer below link: https://support.microsoft.com/en-us/kb/269155

Issues and Mitigations (ESXI Host Wise)

S.No	Issues	Mitigation
1	UnabletoquerytheliveLinuxsourcemachinefor SUSE11	SUSELinux11withUEFIisnotsupportedbyVMwareconverter,need to perform a backuprestore
2	Unable to obtain hardware information for selected machine using Standalone VMware converter 5.5 and 6.0	This issue resolved by using VMware Standalone converter tool 5.1
3	IPaddressalreadyinuseerrororonaconvertedvirtual machine(1012255)	Remove duplicate hardware from Registry as per below link (http://kb.vmware.com/selfservice/search.do?cmd=displayKC&docType=kc&docTypeID=DT_KB_1_1&externalId=1012255)
4	MacAfee antivirus caused an issue during P2V migration	AntivirusisnotgettingupdatedpostP2Vrequiresreinstallation.We getfasterspeedduringmigrationwhenantivirusisremoved.
5	VirtualdiskdriverswerenotupdatingafterP2Vdue to McAfeeAntivirus	Uninstalledtheantivirusthenrestartedtheserverandupdated virtualdiskdriverssuccessfullythenandreinstalledtheAntivirus
6	VMConvertercreatedE1000Networkadapterafter P2Vbydefaultwhichcausesan issueforNetwork bandwidthandNICworkingfunctionality.	RemovedNICcardsettingsduringP2Vmigrationandinstalled separatelyVMXNET3NICcardafterP2Vmigration.
7	Windows2003ServerwasfailedtobootafterP2V migration	DisabledMemory\CPUhotplugoptiononVMassomeWindows 2003OSisnotsupportedthisfeature
8	The serverwasunabletologinusingDomainIDbutable tologinusinglocalAdminafterP2Vmigration.	MostoftheserversweredisjointedandjoinedfromtheADandissue resolved.
9	WhentriedtosettheIPaddressonanetwork adapter,thefollowingerrormessagereceived-->TheIPaddressXXX.XXX.XXX.XXXyouhaveentered forthenetworkadapterisalreadyassignedto anotheradapterNameofthe adapter.Nameofadapter ishiddenfromthenetworkandDial-upConnections folderbecauseitisnotphysicallyinthecomputer or isalegacyadapterthatisnotworking.Ifthesame addressisassignedtobothadaptersandthey becomeactive,onlyoneofthemwillusethis address.Thismayresultinincorrectsystem configuration.DoyouwanttoenteradifferentIP addressforthisadapterinthelistofIPaddressesin the advanced dialog boxes?	Thereisahotfixavailabletoresolvetheissue.Referbelowlink: https://support.microsoft.com/en-us/kb/269155

Benefits of Virtualization

- **Application availability** - Most server virtualization platforms offer a number of advanced features that are not found on physical servers, which helps with business continuity and increased uptime. Capabilities such as live migration, storage migration, fault tolerance, high availability, and distributed resource scheduling give virtual machines the ability to quickly recover from unplanned outages. The ability to quickly and easily move a virtual machine from one server to another is perhaps one of the greatest single benefits of virtualization with far-reaching uses.

- **Better disaster recovery**- Disaster recovery is quite a bit easier when the datacenter is virtualized. With up-to-date snapshots of virtual machines, one can quickly get back up and run. And should disaster strike the data center itself, virtual machines can be moved elsewhere. Having that level of flexibility means disaster recovery plan will be easier to enact and will have a much higher success rate.

- **Easier backups** – Not only full backups can be taken of the virtual server; one can take backups and snapshots of virtual machines. These virtual machines can be moved from one server to another and redeployed easier and faster. Snapshots can be taken throughout the day, ensuring much more up-to-date data. And because firing up a snapshot is even faster than booting a typical server, downtime is dramatically cut.

- **Faster server provisioning** - Server virtualization enables elastic capacity to provide system provisioning and deployment at a moment's notice. One can quickly clone a gold image, master template, or existing virtual machine to get a server up and running within minutes.

- **No vendor lock-in** - One of the nice things about virtualization is the abstraction between software and hardware. This means one does not have to be tied down to one particular vendor — the virtual machines don't really care what hardware they run on, so one is not tied down to a single vendor, type of, or even platform.

- **Save energy, go green**- Migrating physical servers over to virtual machines and consolidating them onto far fewer physical servers' means lowering monthly power and cooling costs in the datacenter.

- **Time spent on routine IT administrative tasks** – By virtualization there is are significant improvements in time spent on routine administrative tasks.

- **Isolate applications** - Server virtualization provides application isolation and removes application compatibility issues by consolidating many of these virtual machines across far fewer physical servers. This also cuts down on server waste by more fully utilizing the physical server resources and by provisioning virtual machines with the exact amount of CPU, memory, and storage resources that it needs.

- **Reduce the data center footprint -** Server consolidation with virtualization reduces the overall footprint of the entire data center. That means far fewer servers, less networking gear, a smaller number of racks needed -- all of which translates into less data center floor space required.

- **Reduced cost** - Hardware is most often the highest cost in the data center. By virtualization, amount of hardware used is reduced and hence cost reduction. Cost is also reduced due to lack of downtime, easier maintenance and less electricity used.

Conclusion

Virtualization is a powerful technique for extending the value of your hardware and software investments. Virtualization may be used to maximize utilization of expensive resources, enable the dynamic allocation of resources to high-priority applications, extend the effective lifespan of technology investments, improve the productivity of developers and quality assurance engineers, improve opportunities for collaboration, and simplify the management of diverse systems.

The adoption of virtualization is highly productive for many reasons. Collections of inefficient servers can be replaced with fewer machines; software can be tested while isolated in harmless virtual partitions; and data centers can gracefully and virtually conform to shifting work models, new technologies and changing corporate priorities.

The future of enterprise IT management will be based on virtual computing. Intel VT makes it possible to maximize computer utilization while minimizing all associated overheads of management, power consumption, maintenance and physical space.

References

[1] http://pubs.vmware.com/vsphere-55/topic/com.vmware.ICbase/PDF/vsphere-esxi-vcenter-server-551- installation-setup-guide.pdf http://www.vmware.com/in/virtualization/how-it-works

[2] Memory Resource Management in VMware ESX Server http://www.vmware.com/pdf/usenix_resource_mgmt.pdf

[3] vSphere Resource Management Guide http://www.vmware.com/pdf/vsphere4/r40/vsp_40_resource_mgmt.pdf

[4] Performance Tuning Guidelines for Windows Server 2008 R2 http://www.microsoft.com/whdc/system/sysperf/Perf_tun_srv.mspx

[5] Bi Data Analytics with R and Hadoop Vignesh Prajapati, Packt Publishing, 1st edition, 2013.

[6] Thota, S., 2017. Big Data Quality. Encyclopedia of Big Data, pp.1-5. https://link.springer.com/referenceworkentry/10.1007/978-3-319-32001-4_240-1

[7] Machine Learning with R Brett Lantz, Packt Publishing, 1st edition, October 2013.

[8] Hadoop For Dummies Dirk deRoos, Paul C. Zikopoulos, Bruce Brown, Rafael Coss, and Roman B. Melnyk, John Wiley & Sons, Inc., 1st edition 2014.

[9] Hadoop Beginner's Guide Garry Turkington, PacktPublishing, 2013.

[10] https://cwiki.apache.org/confluence/display/Hive/LanguageManual+ORC

[11] hortonworks.com/blog/announcing-apache-hive-0-12

[12] Subash Thota "Virtual Infrastructure (Journey from Physical Servers to the Cloud)". International Journal of Computer Trends and Technology (IJCTT) V52(1):4-8, October 2017. ISSN:2231-2803. www.ijcttjournal.org. Published by Seventh Sense Research Group.

About the Author

Subash Thota works as Data Architect and specializes in Big Data, Cloud, Data Integration and Data Analytics with significant experience in Project Management, Agile, and Data Governance. Subash has written several papers in the field of Big Data, the Cloud, and Analytics