Virtual Infrastructure (Journey from Physical Servers to the Cloud)

Subash Thota
Data Architect

Abstract
Today's IT world has a widespread infrastructure hosting various applications for internal requirements and customer requirements. Management of such distributed architecture becomes a pain with growing needs of a customer. Hence there needs to be a one-stop solution to consolidate all infrastructures at one centralized location. The planning consideration and the analysis that need to be followed by converting such distributed architecture to a centralized, one-stop solution is being presented in this white paper. This white paper provides the recommendations for planning the migration upon existing physical servers to a Cloud Infrastructure. That includes the recommendations followed for Network and Compute capacity on the virtual environment.

Keywords: *Cloud Computing, On-Premise, Business Intelligence, Cloud Strategy*

Virtualization can be referred to as the conversion of a physical landscape of server environment into a virtual environment. The virtual version of the server will be designed to have the same capacity regarding computing hardware to meet its functional requirements. Servers implemented in a virtual environment will be described as VM – Virtual Machine. These VMs will be built using an underlying hardware which is referred to as Host. Virtualization uses software-based technology to make use of the underlying equipment on the host. This software is described as a hypervisor. The hypervisor will interact with the hardware resources on the host and provide the computing requirements dynamically for the virtual machines. The architecture will have many such virtual machines upon one single host (hardware) thereby reducing the physical footprint. These virtual machines will run as multiple logical resources. Hence the goal of virtualization is to manage to host workloads in a more scalable manner.

Deciding to virtualize
The benefits of virtualization are significant, but these benefits will depend on the understanding of business requirements and the expectations. Any organization deciding to virtualize intend to be aligned to a goal or a set of goals that will be measured. It is a wise practice first to know what exactly virtualization will offer as results for your requirement. A virtualized application will consume more resources than before and will perform well only when hosted on hardware which is faster than its original device. This additional resources consumed will depend on the hardware platform available, how the configuration has been done and the type of virtualization engine that is used. Hence it is necessary to ensure every computer resource like storage, memory; CPU, network bandwidth, etc. are in place. For critical applications, it is recommended to plan for worst case scenario. On the other side, we should avoid dedicating more resources than required since this will have an opposite impact on other hosted virtual machines.

Scenarios where to avoid virtualization
Regardless of the feasibility to virtualize servers and applications, there are few instances where virtualization does not add any gain. Below are few examples.

1. Applications that make use of USB keys for licensing
2. Applications that use high-end hardware such as video handling equipment
3. Application that consumes high compute resources such as CPU, disk, memory, and network
4. Large database servers

Implementation Plans
As like any other project, it is essential to analyze and understand the outcome and the goals of virtualization. The below need to be identified before commencing a virtualization project 1. Why is this virtualization being done? 2. What is getting virtualized? (entire platform or only applications) 3. What will be the estimated budget and how much will be my cost saving? 4. What are the risks involved? 5. What is the scope of virtualization? (will this be a one-shot activity or will there be multiple phases?) 6. Dependent changes – what will be the relevant changes that will need to be done about this virtualization? The list can get comprehensive depending upon the level of virtualization being done. However, a strategic analysis as like above can be termed as a best practice.

Deciding the starting point

There is nothing as such called a starting point instead, the best kick off for a virtualization can be to identify the areas where maximum value can be achieved. One of the best practices to start is to first virtualize development environment, quality assurance, testing environment, support and demo centers. Next best step is to decide then which business applications would make the best fit for consolidating to a virtual platform. It's good to start first with virtualizing resources that have low severity and less resource utilization levels Ensure that the targeted application supports virtualization. It's critical that you get the necessary support from the executive management and relevant stakeholders who host and manage the application. The required hardware like memory, CPU resources to be made available for virtualization. It is recommended to have a centralized storage for files during virtualization.

The preceding step is to analyze who are the market players providing virtualization platform. Few of them given below

1. IBM
2. Citrix
3. Microsoft
4. VMWare
5. Amazon

Industry standards and trends to be analyzed and then the appropriate virtualization platform to be decided. Specific features of the virtualization platform may or may not add value to requirement. Hence it is recommended to compare the features that each of these platforms offers.

Tools to manage the infrastructure

Each vendor provides tools that facilitate to manage their systems efficiently. But it's essential to ensure that the machine provides a broader level of support that enables to complete the business process. For this we need to identify the interrelationship between the components that may affect the services. Example a problem on one of the services on virtualization platform may modify one or more functions of the application. Few other challenges that relate to management of the environment are provided below.

1. Easy creation of VMs and migration of existing VMs
2. Managing the entire lifecycle of the VMs that include patching, OS upgrade, and provisioning of new VMs
3. Monitor and track performance utilization in real time and through historical data
4. Platform that provides tools for back up and disaster recovery of the virtual hosting environment
5. Platform that provides tools to manage security
6. Smart Management

- Dynamic allocation or change in resource allocated to virtual machines
- Moving virtual machines between hosts when required
- Dynamic Provision of virtual machines when triggered by policies

All above primary capabilities would need to be managed in the right manner. Else this will result in potential risks.

Managing VM Image deployment phases

Deployment of virtual machines happens in a virtual environment through virtual machines image files which in other words are called as VM templates. There can be phases to manage the different versions of these images. However, it is essential that appropriate change controls are in place to manage those images. Below are some phases to achieve the images.

Development

The phase to have the images while in the stage of development. This includes fresh models as well as updates to existing models.

Test / QA This phase can be used to test and verify the fresh images that are developed. If further changes or corrections required, they can be transferred back to the development phase. Approved images from this can then be moved into "Available images."

Available images This is the library where the master images are stored for approved usage. The master images will be implemented in the form of VM templates. When the OS image is to be deployed, the image based on this template to be transferred to configuration phase. If there are any modifications to be made they need to be transferred to the development phase. If a version is not to be used or to be retired, it needs to be moved to Archive.

Archive

To have the images that have reached end of life

Configuration The phase where images are configured before deployed for production. Configuration steps may include the renaming of the hostname or adding a step to include additional software to be implemented. Once the server is configured, it is moved to production.

Production

Production is the phase where the images are deployed. If the requirement is to upgrade an image this need to be moved back to available images phase.

Planning Storage Resources

Storage is one of the most important areas to analyze during a virtualization requirement. The estimated storage should not only contribute to the success of current requirement. However, must turn out to get scalable in future. Below are few constraints which turn out to create performance bottlenecks.

Multiple I/O requests accessing the same disk

While there are multiple I/O streams trying to communicate with the same disk, there will be delays in processing such requests. The best example for this is in case of a database. Also, the same kind of issue may come if there are multiple virtual machines using the same physical disk. This risk can be mitigated by ensuring each VM has its physical disk. There need to be a SAN storage based on fiber channel where the disks can be allocated or distributed among virtual machines. Storage solution with multiple disks preferably disks of smaller size is recommended. This is because multiple disks will provide optimized performance due to a large number of heads and do concurrent read / write operations.

Storage for individual VMs

On traditional physical hosting, the entire disk storage attached to the server was allocated. This means the server will significantly have more space than it requires thereby having lots of unused storage space. By using an external storage system, it becomes easier to manage by allocating the disk storage on a need basis. Allocating storage space on External SAN storage on a need basis can save more storage space.

Estimating compute resource requirement

Once the virtualization approach is finalized, the next step will be to estimate the hardware required to support the workload. The recommended method is to estimate the hardware as equivalent to a current resource available in physical hosting. Well planned virtual environments turn out to be highly flexible where there is a scenario of resource shortage it is easy to plug in a host server and transfer the excess load.

The first step on sizing the compute resources is to gather the list of applications targeted to be virtualized. Then the footprint of existing physical hardware to be noted. This includes memory, CPU, internal hard disk capacity, storage size attached, the network for each software that is hosted. The minimum processing resources may be lesser than the actual resource present on the hardware. Still, the actual full capacity of hardware in the current state must be accounted for computing calculations.

Estimating peak loads

For effective virtualization of the existing application, it is highly recommended to analyze the existing workloads and performance. This will help to understand how and when the application uses the hardware resources. CPU, RAM, Network, and storage are the key resources to be looked upon. It is important to consider the number of transactions (inbound and outbound) that are happening, the storage utilization, Memory and the network utilization of the application. All these criteria can be monitored over a period of one month to observe the peak utilization. This will help to establish the extent of computing resource the application will need at different times. This will eventually simplify the resource allocation or scheduling of resources for the application during its peak utilization. Below are few examples that may cause the resources to shoot up.

Restarting of the application platform

It is likely to have a significant impact when there is one server getting restarted. The load may fall on another server when there is a load balancing configured. This will cause a big impact if multiple servers are restarted without proper planning.

Reoccurring events

While there are multiple processes or functions triggered at the same time, this may result in peak utilization. Example multiple users logging in at the same time into an application. Multiple users were running reports at the same time on the application or multiple users logging out of the application during the close of business.

A virtual hosting infrastructure will have its base hardware under single physical hardware platform. During unusual scenarios, it is likely that other components will also get affected due to performance issues. Hence the hardware capacity planning to be done in such a manner to accommodate such peak loads. Hypervisors can share the overallocated resources when they are available. However, while there are expectations of guaranteed response times, it is recommended to dedicate resources appropriately.

Planning considerations on CPU Speeds

There are various architectures upon which physical CPU is built. Quad, Hex-core, Octa-core, hyperthreading, etc., There will be a difference in performance when compared to a single dual-core CPU and two separate dual-core CPU. Hence it is important to have CPUs built on similar architecture within a virtual infrastructure. Because CPUs with varying design may not be able to take the loads while the processing requirements are being transferred between each other. It is important to have fast CPUs however on the other

side it is recommended that the CPU has large L2 /L3 cache. The next factor is the motherboard which must be of a high-class standard. It is critical that the bus that has an interface between the CPU, memory and the other components on the same board is with high performance. In a virtual infrastructure, the CPU load is the most prominent resource that will be shared between VMs. The VM kernel is responsible for scheduling the available physical CPU resources so that the virtual machine will have access to the CPU.

While there is a CPU request from one virtual machine and if the CPU is busy serving other virtual machines, then the virtual machine must wait until the CPU reverts after serving other virtual machines. This will result in slowness on the virtual machine while it waits for CPU resources. Hence it is essential to ensure adequate CPU resources are available on the host system (underlying physical hardware). While estimating the number of CPU cores for a host server below are few things to be considered.

1. The hardware on the host server always adds overhead to the CPU usage. This overhead is usually at an average of 10%. Hence it is recommended to have minimum 10% buffer on the CPU cores while capacity planning is done for CPU

2. While there are cases where all virtual machines hosted on the same host will hit the peak at the same time, there needs to be a sufficient number of physical CPUs. It is recommended to have virtual CPU allocated to the host equal to number of physical CPUs

3. CPUs with large amount of L2 / L3 cache can be efficiently used for minimizing the need of virtual CPUs

There are many virtualizations supported functions available in the BIOS. It is recommended to enable all such options on the BIOS of the hardware being used. Also, there are few virtualization functions supported by the hypervisor. It is recommended to enable these options. (Example – AMD /V, Intel VT or hyperthreading)

Physical Memory – RAM

Virtualization always adds the overload for both underlying hosts and the virtual machines on top of it. This overhead depended upon factors such as OS, number of VCPU and allocated RAM. A Guest operating system as a VM consumes approximately 200 MB as overhead for VM Ware based platform. (Source – vSphere Resource management guide)

In the case of Microsoft Hyper V – should assign 32 MB memory as overhead for the first GB of virtual RAM and another 8 MB for an addition of each additional GB. Hence misconfigurations may lead to oversubscription of memory. But this depends on the virtualization engine that is being used. Hence memory allocation must be handled with precaution.

For additional details, please refer URLs under Memory Resource Management in VMware ESX Server and v Sphere Resource Management Guide in the reference section

NICs and Bandwidth

Network Interfaces are shared between the VMs. The virtual machines have access to a pool of teamed NICs, and this can be of advantage when it comes to performance and fault tolerance. Since two physical network interfaces are teamed, an outage on one of the physical NIC does not cause any impact on the availability and performance.

Below are some recommendations

1. For best utilization of shared resources, it is recommended to use multiple server grade NICs
2. 2 Physical separate NICs to be teamed and deployed for high availability of the port. E.g., two dual-port NIC cards are best preferable when compared to one quad port NIC card
3. Having multiple NIC interfaces allows the physical host to schedule and distribute the load between them
4. It is recommended to have dedicated NICs for service console and vMotion or live migration tools
5. To minimize the CPU overhead, it is recommended to analyze what are the offline mechanisms supported by the virtual engine and the NIC cards deployed

There are stateless features like Large Send Offload, Checksum offload and VLAN tagging that are supported by virtual engines.

For more details, please refer the link under - Performance Tuning Guidelines for Windows Server 2008 R2 mentioned at reference section

Storage

I/O speeds become a constraint resulting in poor disk performance. Hence it is recommended to use high-performance physical SAN based disks in a virtual environment. It is preferable to use SAN disk based on HBA fiber channel based. Below are the benefits of using a SAN-based storage system

1. Flexibility and ease of management for natural allocation of storage
2. The physical disk environment will be separated from the Virtual Machines. So, the machines can be moved between physical disks. This helps in high availability solutions like vMotion
3. Currently used SAN storages have large cache that will increase the I/O performance

Disk performance becomes a hit when there are multiple virtual machines trying to access one single physical disk or fewer disks at the same time. It is preferable that the VM demanding for disk performance has a dedicated access to one single LUN (Logical Unit Number). At the same time while high-performance disks are used each LUN can be shared between few virtual machines. While doing this, it is essential to continuously monitor the disk performance Page files are used as additional boosters on top of the physical RAM. Placing this page file on a poor performing disk is not going to add any benefit. Below measures can be followed by the placement of the page file

1. Page file must not be placed in conjunction with memory intensive application data
2. Page file not to be placed on the drive allocated for fault tolerance because this will slow down the write access 3. It is best to use physical array off disks for paging

Recommendation on multiple vCPUs

1. While required it is recommended to use smaller VMs with lesser configurations when compared to one single large configuration VMs
2. Multiple vCPUs to be allocated only when the application gets benefited due to such settings. It is recommended to test the application with one vCPU and add as when required
3. While the schedule is continuously being enhanced in every version of ESX, it is better to use the latest version of ESX server
4. It is recommended to have necessary to have physical CPUs available whenever there is a requirement for reliable performance

For benefits on performance, it is recommended to use the 64-bit operating system. Ensure application compatibility before choosing a 64-bit operating system

Conclusion

A simple explanation on virtualization can be termed as cutting up one single piece of hardware into multiple logical pieces. Virtualization resources are limited only to one single virtual machine. Hence while a hacker compromises a program, on one virtual machine, the impact will be restricted to only that virtual machine. And the hacker will not have access to applications running on other virtual machines.

There are numerous advantages of virtualization regarding flexibility, scalability, and management. However, without proper planning and execution on all the above, it will turn out to become a negative impact. Virtualization can be expected to grow to the next generation which will stand more promising than the current scenarios. Virtualization has also stepped into mobile computing where a single mobile can be used with two profiles.

This will ensure data security between the two profiles. With the growing demands in the market, virtualization has turned out to be the one-stop solution for any customer. With these practices, skills, and recommendation adopted in the market, virtualization will be the right choice for customers expecting exponential growth.

References

[1] Memory Resource Management in VMware ESX Server http://www.vmware.com/pdf/usenix_resource_mgmt.pdf
[2] Li, H., Spence, C., Armstrong, R., Godfrey, R., Schneider, R., Smith, J., White, J. (2010), Intel Cloud Computing Taxonomy and Ecosystem Analysis, Printed in USA 0210/KC/KC/PDF, 1-4
[3] Linthicum, D. S. (2009), Moving to Cloud Computing Step-by-Step, The Linthicum Group
[4] Malcolm, D. (2009), The five defining characteristics of cloud computing, ZDNet
[5] McKendrick, J. (2007), BI, Delivered from the Cloud, Ebizq Net, The Insider's Guide to Business and IT Agility
[6] Menken, I. (2009), Cloud Computing - The Complete Cornerstone Guide to Cloud Computing Best Practices:
[7] Concepts, Terms, and Techniques for Successfully Planning, Implementing and Managing Enterprise IT Cloud
[8] Computing Technology, Publisher Emereo Pty Ltd, 172 pages, ISBN-10: 1921573007, ISBN-13: 978-1921573002 Avanade (2009), 2009 Global Survey of Cloud Computing
[9] Thota, S., 2017. Big Data Quality. Encyclopedia of Big Data, pp.1-5. ttps://link.springer.com/referenceworkentry/10.1007/978-3-319-32001-4_240-1
[10] Birst (2010), Why Cloud BI? The 9 Substantial Benefits of Software-as-a service Business Intelligence, Birst, Inc.
[11] Ghilic-Micu, B., Stoica, M., Mircea, M. (2008), A framework for measuring the impact of BI solution, The 9th
[12] Hinchcliffe, D. (2009), Eight ways that cloud computing will change business, ZDNet
[13] Sheelvant, R. (2009), ten things to know about cloud computing strategy
[14] The Cloud Promise, Scientific and Academic Publishing http://article.sapub.org/10.5923.j.ac.20170703.02.html
[15] Verizon (2009), Start Packing. You're Moving to the Cloud – and We Can Help, Verizon Business Wayne, W. E. (2009), Implementing BI in the Cloud, The Data Warehousing Institute
[16] Marinela Mircea, Bogdan Ghilic-Micu, Marian Stoica, Combining Business Intelligence wit Cloud Computing to Delivery Agility in Actual Economy.
[17] Winans, T., Brown, J.S (2009), Cloud Computing, 1-39
[18] Berkowitz, J. (2009), Cloud Computing (Part 1): Advantages, Types and Challenges, CRM Mastery Weblog
[19] vSphere Resource Management Guide http://www.vmware.com/pdf/vsphere4/r40/vsp_40_resource_mgmt.pdf
[20] Performance Tuning Guidelines for Windows Server 2008 R2 http://www.microsoft.com/whdc/system/sysperf/Perf_tun_srv.mspx

About the author

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