



UNDERSTANDING VMWARE INFRASTRUCTURE NEEDS: SAVE MONEY



Kobi Shamama
EMC Corporation

Table of Contents

Understand the Problem 3
Understand organization requirements, needs, and regulations..... 5
Understand NAS Performance vs. SAN Performance10
Virtual Machine Performance analysis14
Build your Environment and Calculate ROI18
Author Biography21

Disclaimer: The views, processes, or methodologies published in this article are those of the author. They do not necessarily reflect EMC Corporation’s views, processes, or methodologies.

Understand the Problem

Benefits and advantages of virtualization technology are well known and widely discussed. Today, virtualization technology can be found in nearly every data center, especially in new data centers where we found that cloud technology and “IT as a service” is well implemented.

So why is it important to keep your hand on the pulse when the organization is already saving money by using virtualization technology?

The ease with which we can build new virtual machines and expand the environment with just a few clicks creates tremendous growth for relevant infrastructure in multiple aspects; whether in connectivity, cables, physical servers, or new employees. However, although organizations are using virtualization technology, it doesn't mean that the money spent is invested wisely.

So how can the organization be more efficient?

If virtual machines in the data center could be characterized in SLA or performance needs, managers—together with the IT team—could make wiser decisions on how to expand or build their virtual environment. Characterizing virtual machines with simple decisions and definition, the organization can define if virtual machines will work over expensive and complex infrastructure or more cheap and simple infrastructure.

The SLA can be found anywhere, not just in the technology world. Take for example, a car factory; not all the cars are the same model or made from the same materials. The car model is predefined by category and the parts, engine, and investment in the car differs from model to model.

In the same way of thinking, why not characterize virtual machines? Should the testing application server be placed on the same infrastructure as the organization production SQL? Ethernet cables, cards, and switches are far cheaper than fibers. Why not mix those technologies wisely?

While placing a cluster of ESX servers on Ethernet infrastructure can surely reduce performance compared to the fiber infrastructure, do we really need all the virtual machines on the fiber infrastructure?

I say no!

By investigating performance needs and defining the SLA, we can separate our infrastructure and define if each virtual machine will work on expensive powerful resources or more inexpensive and simple resources.

To be more efficient, organizations need to understand the differences between the technologies and how they work in order to create a new virtual infrastructure.

After understanding the issues, the obvious questions are:

- How much money can be saved?
- How can I combine those two infrastructure technologies?
- Is it relevant for me?

This article will guide IT technical personnel as well as Managers and CTOs on how to answer those questions and wisely decide when the organization will need to expand or build new infrastructure for connectivity between the physical servers and the storage.

While I chose VMware as the virtual infrastructure, this article is relevant for all virtual infrastructure technologies an organization may choose to use, such as VMware, Hyper-V, XEN servers, or any other.

Understand organization requirements, needs, and regulations

A key measurement for smart choices is to understand the requirements and needs of the organization. While determining the ROI of using cheaper infrastructure, the organization SLA and regulation needs to be investigated and make sure changes are possible. Remember; business and regulation determine the technology which will be used, not the other way around.

Assume the following scenario:

- The organization has reports system with several servers that generate financial reports at the end of the every quarter.
- While measuring the workload of the entire system build from several servers it's shown that, 90% of the time, the system is idle except at the end of the quarter. But, for the CTO and the organization, this system is highly critical and should get the lowest response time whenever it is up and running.

So, where should these servers be placed?

If we go strictly by the technical numbers, they should be placed on Ethernet infrastructure. However, on the other hand, the organization requires that they be placed on fiber infrastructure.

Assume another scenario where just a few virtual machine candidates are found. Should they be separated to a different ESX cluster with only one ESX server?

To choose an efficient solution, follow the flow chart in Figure 1:

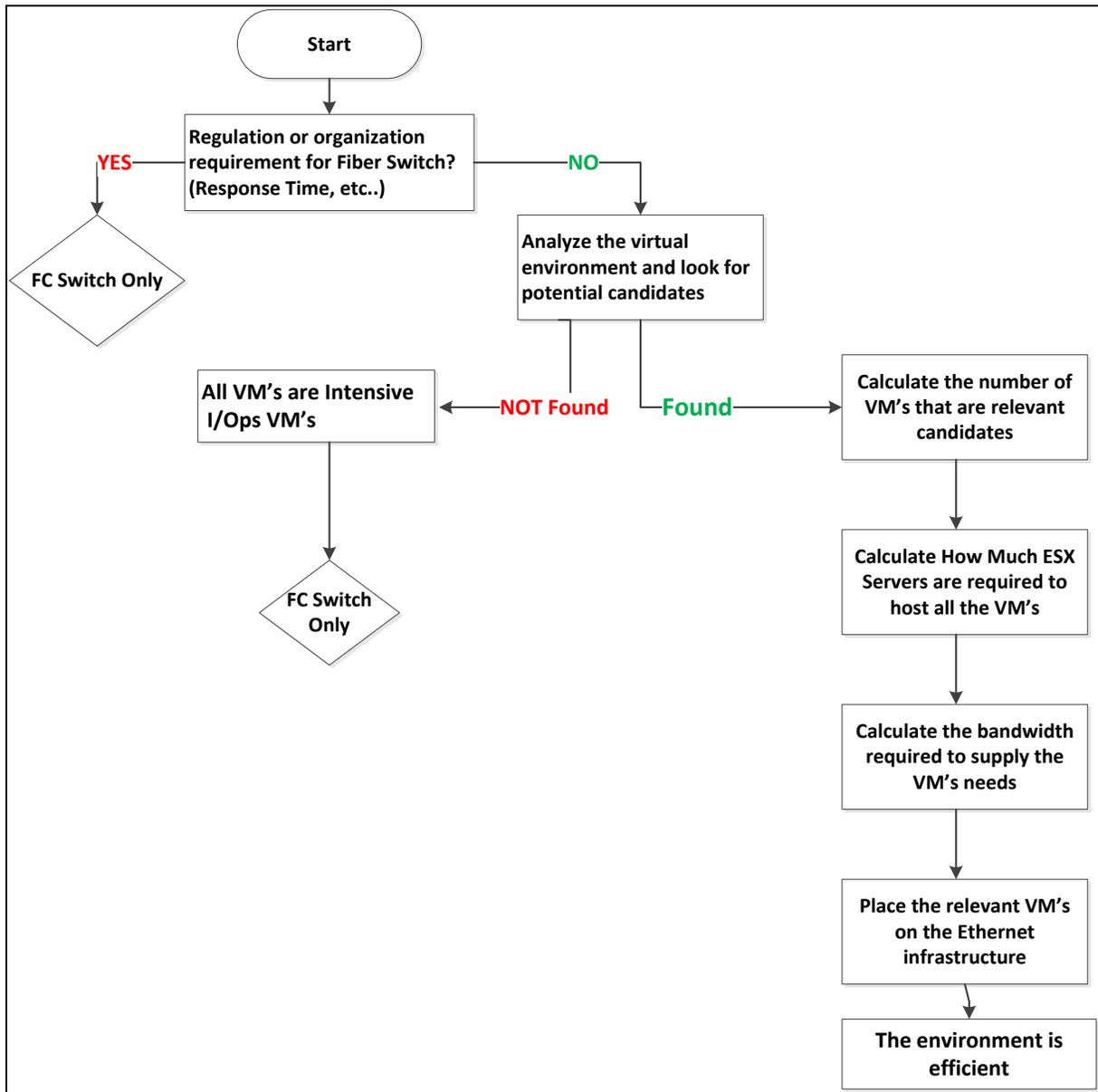


Figure 1

Step 1: Regulation and Organization Requirements

As described in the beginning of this section, if organization requirements and definitions are suitable only for fiber switch, migrating to Ethernet infrastructure isn't suitable for your environment, as shown in Figure 1. If no such organizational restriction applies, continue with the green arrow.

Step 2: Virtual Machines performance analysis

To migrate to an Ethernet environment, relevant candidates need to be found. The virtual machines need to be analyzed and performance calculated so that the organization can wisely decide where every virtual machine should be placed based on performance and SLA.

For further information on how to calculate performance in VMware environments, see the section, Build your Environment and Calculate ROI.

Step 3: Calculate how many virtual machines will be migrated

After calculating and measuring performance for each virtual machine, summarize the entire performance requirements. Those numbers will help you sufficiently align with the bandwidth and IOPS required.

Step 4: Calculate how many physical servers will be needed to host the virtual machines

Calculate the number of physical servers needed in order to host all the virtual machines. The calculation of the physical servers needed **shouldn't** be based on bandwidth required but on the virtual RAM and CPU that is required.

Consult with the VMware team on the amount of memory, CPU, and servers required to host all the virtual machines in the new environment.

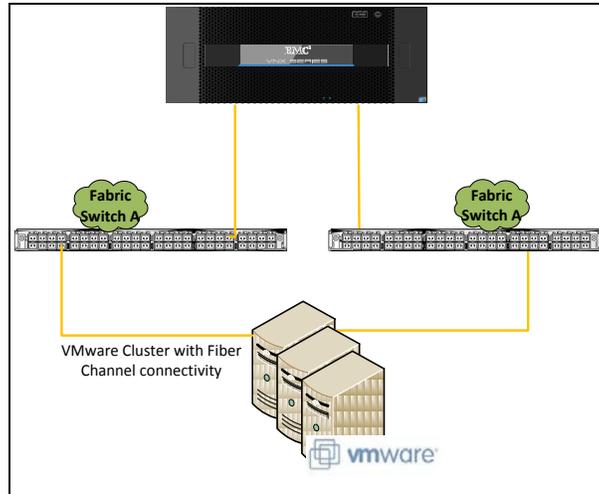
Step 5: Build the new Infrastructure

After calculating the number of physical servers needed, it's time to plan and build the new environment and get to best ROI for your money.

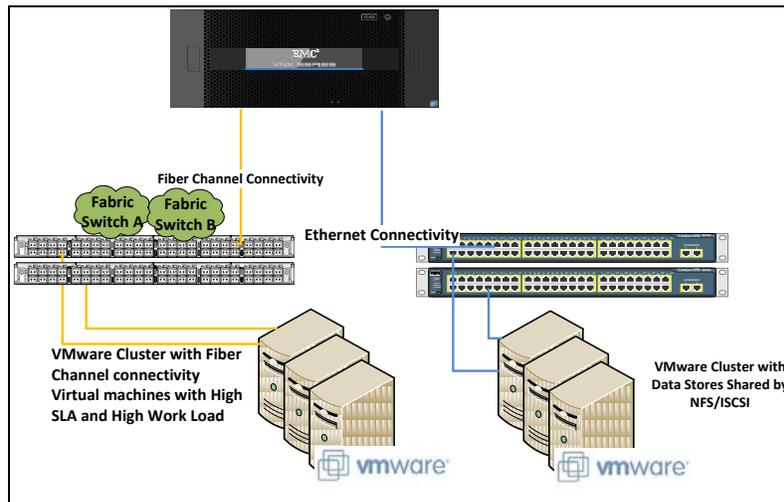
- Build the environment and test it for benchmarking.
- Migrate the virtual machines to the new environment and begin calculating the ROI!

The followings sections will help you divide your environment into two different infrastructures separated by SLA and understand every step on the journey to an efficient virtual infrastructure:

Before the change, all servers in the cluster are connected to the storage via Fiber Channel:



After the changes, the environment separated into two different clusters defined by SLA.



The following flow chart will help you navigate and understand the full process to build a more efficient environment for your organization:

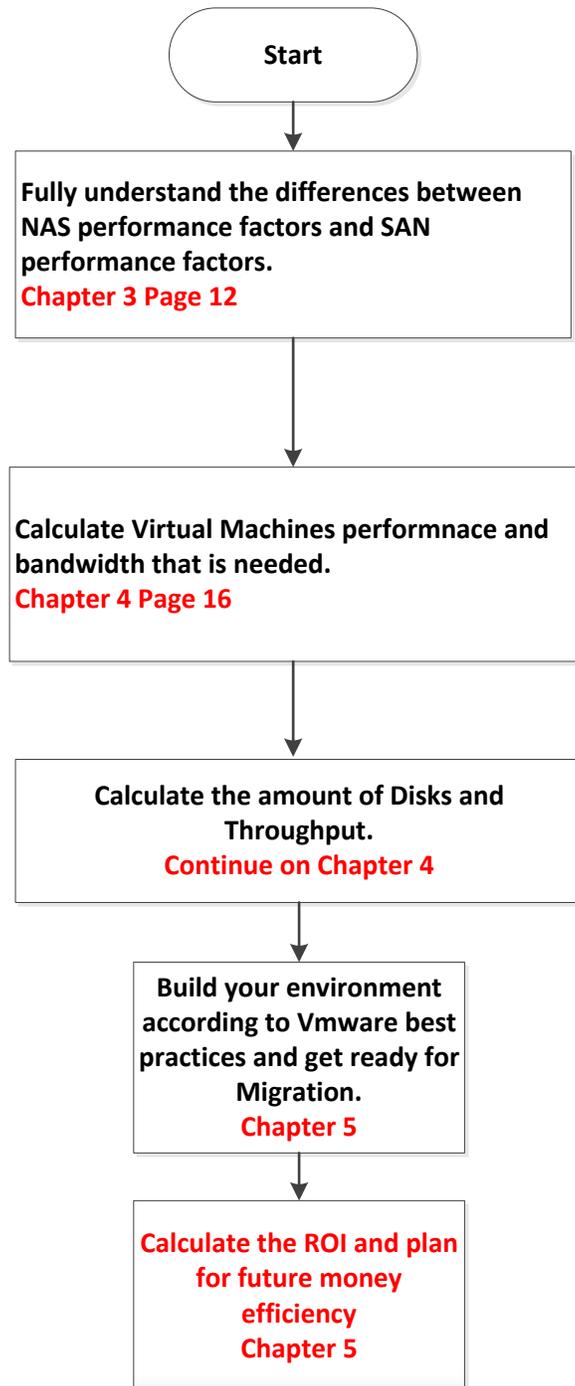


Figure 2

Understand NAS Performance vs. SAN Performance

The common opinion about NAS is that it's much slower than SAN environments. However, we need to be careful about what we mean when talking about performance. It's important to understand where the NAS is slower than SAN.

If talking about I/O, remember that the storage array—and the disks on the storage array—can produce the same amount of I/O whether using SAN or NAS. For example, RAID group from SATA disks (6+2) can deliver the same I/O in NAS or SAN.

Most believe that the main differences between the two are the cables and the Ethernet switch that can supply 1Gb bandwidth instead of 4-8Gb bandwidth for a fiber switch. However, with the increasing wave of FCoE that provides 10Gb bandwidth, this assumption is no longer relevant.

The real answer to the performance differences in is the protocol used in each.

The NAS environment enables connectivity to the storage array and sending data over Ethernet cables. Data transferred over Ethernet cables are delivered over TCP/IP, until they reach the storage array; it doesn't matter if NFS or iSCSI is being used.

The SAN environment is connected with special cards, called HBA, via Fiber Channel cables. Fiber Channel cables usually connect to a fiber switch and then to the storage array. FCP is the protocol used to transfer the data.

To fully understand the differences between the networks, it's important to understand why TCP/IP has degraded performance compared to Fiber Channel.

TCP/IP retransmission

TCP enable the communication between applications and servers over IP. TCP is responsible for separating the data into IP packets that are then delivered over the network from one server to another.

When a session between two sides begins, a “three-way handshake” initializes and synchronizes the connection. For every packet sent using TCP/IP, acknowledgment is sent back to the sender. If the acknowledgment is not received, the sender will retransmit the data. Data retransmission makes TCP/IP much slower than Fiber Channel. As well, calculating bandwidth is done differently than for Fiber Channel.

When do retransmissions occur?

- For every packet sent, a counter is started. If the counter reaches zero before the acknowledgment is received, the sender will retransmit the packet.
- If packets are sent out of order, the packets retransmit.
- Loss of data or connection during transmission will cause a retransmit.

Another major player that impacts TCP/IP performance is the round trip timer (RTT), which is the time it takes the packet to reach the destination and for the acknowledgment to arrive at the source side.

If this loop takes a long time, maximum bandwidth will degrade. If the loop takes only a short time, bandwidth will get better. This is why it's important to have a strong, fast, and reliable network infrastructure.

After the retransmission issue in TCP/IP is understood, it's time to calculate maximum bandwidth available.

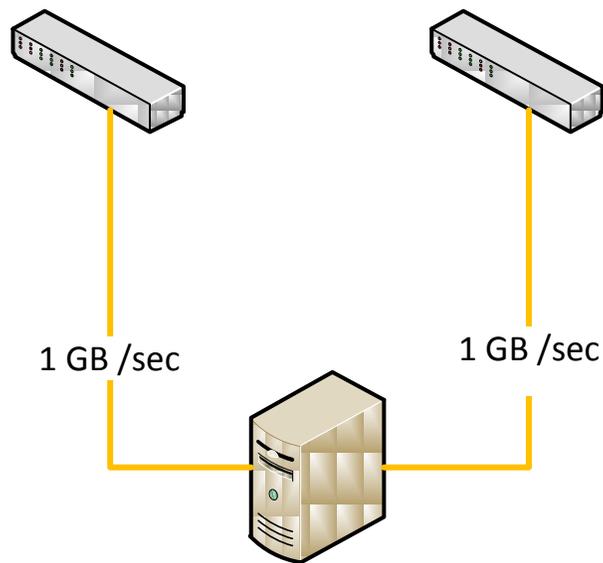
How to calculate maximum bandwidth in a Fiber Channel environment

Network bandwidth is always calculated in **bits**.

File and data calculated in **bytes**.

Assume that a server is connected via 2 Fiber Channel cables with two HBA cards.

1. Translate the bandwidth into bytes: $8\text{Gb}/8\text{bits} = 1\text{GB}$; every fiber channel cable can deliver 8Gb
2. Total bandwidth available for the server is: $2\text{ HBA cards} \times 1\text{GB} = 2\text{GB/sec}$



Virtual Machine Performance analysis

Important to know before calculating performance:

The virtual machine performance gathering process should be predefined and fully understood. It's important to understand what to measure and what to collect when talking about virtual machine performance, as it could become very confusing for the storage administrator and the VMware administrator.

Few virtual machines are hosted on one datastore (LUN), and the I/O generated on that LUN is made by the entire virtual machine workload hosted on that LUN. Still, while analyzing all LUN's does provide an overall view of the entire virtual environment, it's not what is needed.

Rather, the performance that must be collected is how much I/O each virtual machine is creating. By collecting this data, it will be possible to get a good, reliable view of the environment and workload of each virtual machine.

There are a lot of tools and methods to collect performance about virtual machine I/O such as VCops, shell scripts of esxtop utility that keeps historical data (esxtop is online tool). VMware also allow you to enable specific counters for Windows virtual machines to collect counters about the disk I/O each virtual machine is creating from the performance monitor utility.

A common method used to measure I/O is to calculate the average I/O that a virtual machine is producing. This method can be very confusing because when calculating average workload, peaks and bursts I/O not made by the daily workload of the organization is also calculated and can present a misleading overview of the environment.

My recommendation is to collect data statistics from the virtual machine itself and calculate the workload on 95% of the time the virtual machine is working; this provides a more reliable measurement.

You can use any tool that monitors virtual machine performance. There are also a number of companies that offer specific product for virtual machine performance analysis.

Calculate the bandwidth and I/O of the candidates

Once all of the data has been collected and the candidates are found, the number of disks and the bandwidth needed must be calculated.

When calculating performance, separate between Front-end I/O and Back-end I/O.

Front-end I/O is the amount of I/O made between the servers and the storage array.

Back-end I/O is the amount of I/O made between the storage controller and the disks connected to it.

The number calculated by the virtual machines is the number of I/O made by the virtual machine to the storage array. The virtual machine and the application residing on it are completely unaware of the RAID level of the disks the LUN is residing on and we cannot take only those numbers into consideration because of the RAID penalty effect.

RAID penalty – XOR calculation is made for every write to the RAID group that the host is made to the LUN. The write cannot be complete until the write to the parity completes; this is the RAID penalty. Of course, the penalty only applies to the write operation and not the read operation.

The penalty of the different RAID type is:

RAID 0 – 0

RAID 10 – 2

RAID 5 – 4

RAID 6 – 6

Assume that the virtual machines that are planned to migrate into the environment produce 5000 I/O; 1000 I/O of write and 4000 I/O of read. The write ratio is 20% (1000/5000) and the read ratio is 80% (4000/5000).

Now let's assume that we are using RAID 5; the number of I/O that we will need to answer will be calculated as $(1000 * 4) + 4000 = 8000$. The number of I/O that the spindles will need to answer is 8000 and not 5000.

After calculating the number of I/O that need to be answered, it's time to plan the number of disks in the array. There are several disk types and every disk can produce a different amount of I/O.

The following table shows the number of I/O that every disk can produce (as presented in EMC documentation.) The numbers shown are only a rule of thumb and guidelines that help to calculate the number of disks required.

Disk Type	I/O
NL-SAS	90
SAS 10K	140
SAS 15K	180
Flash	3500

Number of disks = Back-end IO/IO for disk

Take, for example, if using with SAS disks 15K, each can produce 180 I/O. Hence, the number of the disks needed will be $8000 / 180 = 44.44$, which results in 45 disks (always round up the number) in RAID 5 being required to produce the I/O amount requested.

Now let's calculate how much bandwidth is needed from the servers to the storage. Whether migrating from a Fiber Channel environment or building a new environment from scratch, the bandwidth available from the servers to the storage can be a real bottleneck unless it's calculated properly.

As mentioned earlier, bandwidth on Fiber Channel and bandwidth on Ethernet are calculated differently. In addition to the I/O produced by the virtual machines, it must be verified that enough bandwidth can be supplied for all throughput made by those virtual machines. The throughput is the amount of MB that can be transferred to the storage array in a given amount of time.

Every application performs I/O with known block size. The more block size that is small, the more I/O the application will perform. Small block size carry small amount of data so the application will need to send a lot of I/O to send all the data.

So before calculating the throughput required, the entire throughput made by those virtual machines and the average block size must be summarized. These data can be collected in the same way used to collect the I/O.

Recalling the example of the virtual machine candidates that produce 5000 front-end I/O, we know that the average block size is 8 KB and the number of disks needed for I/O is 45 (as calculated earlier).

The throughput that can be produced by those disks will be calculated as:

$$\text{MB/s} = (\text{Raw I/O of one disk} * \text{Number of disks}) * \text{Block size} / 1024$$

$$\text{MB/s} = (45 * 180) * 8 / 1024 = 62.28 \text{ MB/s}$$

That means that the amount of data that can be produced by those disks is 62.28 MB/s.

So where is the bottleneck? Well, now that we are building the new environment, we need to ensure that the Ethernet cable with the TCP/IP retransmit degradation (as mentioned earlier), can handle the amount of throughput.

What do I mean by that?

Take for example a big parking area capable of hosting 2000 cars. The parking area has only one lane for entering and one lane for exiting. Consequently, the number of cars that can enter the parking area in a given amount of time isn't dependent on the parking area; it depends on the lane and how many cars can enter in a day.

It's the same for throughput. The classic Ethernet cable connected to the switch can provide 1Gb bandwidth and we need to calculate how many cables will be needed to transfer this

amount of data. To calculate the number of cables needed to answer the performance and bandwidth needed, do the following:

$(\text{Application Block Size} * 8 \text{ Bits}) / \text{RTT} = \text{Maximum throughput that can be produced}$. For example, $\text{RTT} = 150$.

One cable of Gigabit can produce the following: $1000/8 * 0.15 = 18.75 \text{ MB}$.

Hence, at least four Ethernet cables are needed for carrying all the data from the server to the storage (~ 62; calculated before / ~18 MB).

Build your Environment and Calculate ROI

After all the numbers are calculated and the number of servers/cables/disks needed have been determined for your environment, build it!

Prepare the environment according to the best practices document and perform ATP test to make sure the environment is reliable.

The reason for taking all these steps is to reach the final step where you can calculate the ROI.

The investment made by the organization needs to separate into two parts:

Education – IT teams need to learn how to prepare and design this kind of solution and project.

Infrastructure – New cables and switches, storage, and sometimes servers need to be taken into considered.

So when can ROI be measured? Almost immediately. Upon the very first day that you transfer and separate the environment, you could see the money saved by the organization:

	Fiber Switch	Ethernet Switch	Comments
Connectivity Card	HBA	Ethernet/ISCSI HBA	HBA card much more expansive than Ethernet card
Cables	Fiber	Ethernet	Fiber cables are much more expansive than Ethernet cables
Gibics	Needed for every port	Not needed	Every port that is used in the FC switch need FSP , unlike Ethernet switch
Maintenance	Unique role and knowledge	Special role not needed	Zoning and SAN maintenance requires special knowledge in the organization where in every site the network team can just configure special VLAN or jumbo frames in the switch - the knowledge and personnel are already in the organization

Table 1

Over time, the organization will see the real money saved by IT. Every organization will be able to more easily plan for growth in the future—and plan well for it.

The money saved per virtual machine after separating the fiber channel environment and Ethernet environment will be completely different. Organizations will find it much cheaper to store the virtual machines on cheaper infrastructure in those cases where the virtual machine doesn't require special SLA or high performance.

Figure 3 shows how the ROI to the organization increases over time.

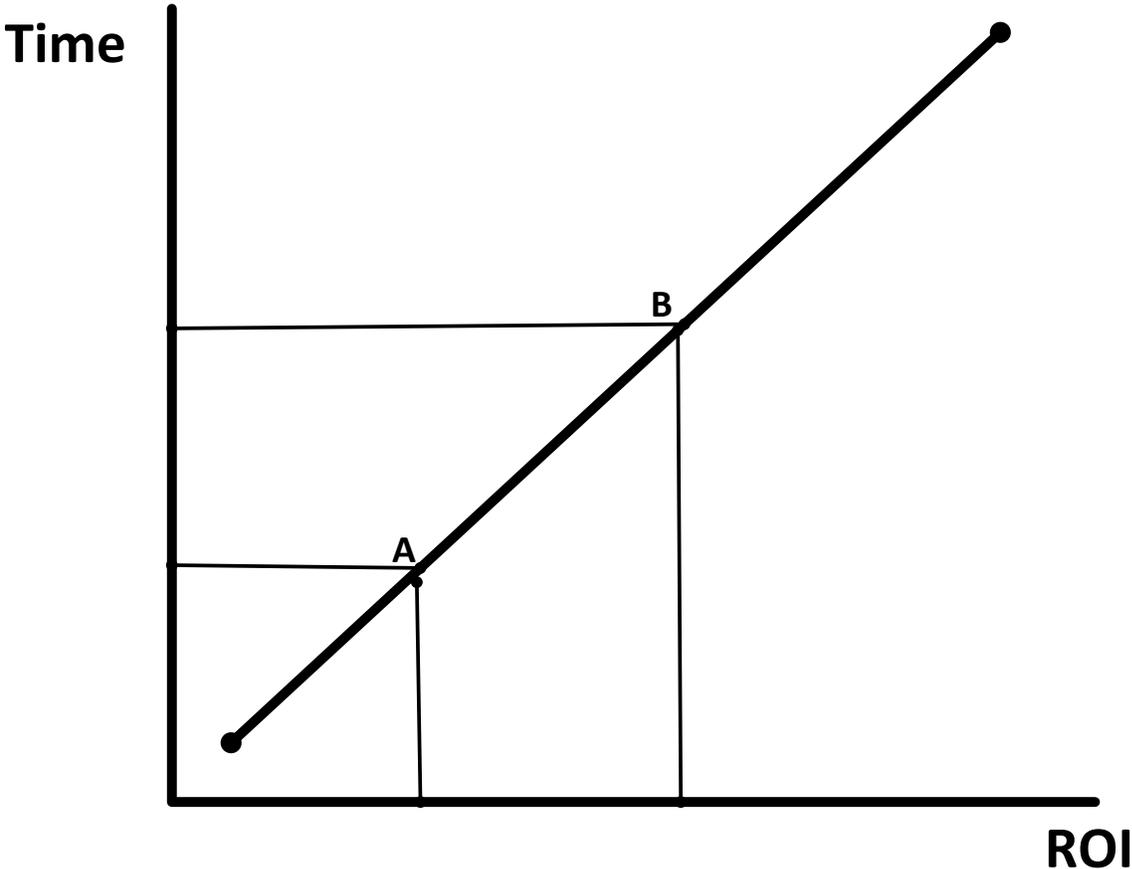


Figure 3

In summary, I would like to share a sentence that I believe in and compelled me to write this article:

“Business choices create technology needs; not the other way around”

Author Biography

As an implementation specialist with EMC for VNX system since December 2010, I have seen many customer sites and environments and performed design, architecture, and implementations.

Prior to EMC, I joined the Israel Defense force and spent 4.5 years, building my knowledge of virtual solutions and infrastructure.

My long-term vision is to combine financial efficiency in the technology world and help organizations make smart and economical decisions aligned with technical achievements.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.