VMware® Fault Tolerance Recommendations and Considerations on VMware vSphere™ 4

WHITE PAPER

vmware®

Table of Contents

Introduction 3
The VMware solution 3
Overview of VMware® Fault Tolerance 4
VMware® vLockstep technology 4
Transparent failover 4
Lifecycle of a fault-tolerant virtual machine 5
General FT Requirements and Recommendations 6
Cluster and host requirements 6
Storage requirements 6
Networking recommendations 7
VMware FT Usage Scenarios 8
Different methods of protecting virtual machines 8
VMware FT on-demand 8
Patching hosts running VMware FT virtual machines 8
Recommendations for Reliability 9
Uniformity of Hosts 10
Placement of Fault Tolerant Virtual Machines 11
Timekeeping Recommendations 12
Windows guest operating system time synch 12
Linux guest operating system time synchronization 12
ESX server time synchronization 13
VMware FT Application Recommendations 13
Example 1: High availability for a multi-tiered SAP application 13
Example 2: High availability for the Blackberry Enterprise Server 14
Summary of Performance Recommendations 15

Introduction

As dependencies on computing increase, ensuring that applications are highly available becomes more critical. Many enterprise hardware components have built-in redundancies, such as RAID storage, ECC memory, and multiple power supplies in servers. However, customers are increasingly interested in protecting the entire operating environment from complete hardware failures, and not just the failure of individual components.

Some existing fault-tolerant systems address this requirement, but these systems typically require either proprietary hardware or modified software and usually support only a handful of operating systems. More often, administrators use software-based solutions such as clustering to provide high availability. However, such clustering solutions are difficult to set up and are often tied to specific applications and operating systems.

The VMware solution

A key philosophy within VMware has been to add a range of high-availability features to its virtualization platform. All of these features are simple to set up and are agnostic to the operating system and applications running within the virtual machine.

Table 1 lists the high availability features introduced by VMware and the protection they provide.

AVAILABILITY FEATURE	PROTECTION	DOWNTIME	INTRODUCED IN
NIC Teaming	Network link/switch failures	No	ESX 2.0 (2003)
Storage Multi-pathing	Storage link failures	No	ESX 2.0 (2003)
VMotion™	Server maintenance	No	ESX 2.0 (2003)
VMware® High Availability (HA)	Host and guest failures	Yes	ESX 3.0 (2006)
VMware® Storage VMotion™	Storage maintenance	No	ESX 3.5 (2007)
VMware vCenter™ Site Recovery Manager	Site failure	Yes	ESX 3.5 (2008)
VMware® Fault Tolerance	Host failures	No	ESX 4 (2009)

Table 1: Timeline of Availability Features

These features provide additional levels of protection that are complimentary to the built-in protection provided by the hardware. As an example, RAID protects disk failures, and storage multi-pathing in VMware ESX™ 4 provides additional protection from storage link failures. Likewise, NIC teaming uses a group of NICs to provide greater throughput in some cases during normal operations and redundancy for failover during failure events. In fact, VMware ESX 4 has become the de facto platform foundation for the virtual datacenter operating system in part because of its trusted ability to provide very high availability.

The newest addition to this list is the ground breaking feature - VMware Fault Tolerance (FT). VMware FT is a feature available with VMware vSphere™ 4 (i.e., ESX 4 and vCenter™ Server 4) that allows a virtual machine to continue running even when the underlying physical server fails. It is a software solution that runs on commodity hardware and does not require any modifications to the guest operating system or applications running inside the virtual machine. This paper gives a brief overview of how VMware FT works and provides a list of requirements. The paper also describes several usage scenarios along with application and reliability recommendations. Finally, a summary of performance recommendations is included.

Note: All references to ESX in this document also apply to VMware® ESXi™.

Overview of VMware® Fault Tolerance

VMware FT is a leading-edge technology that allows virtual machines to continue running even when server failures occur. When VMware FT is enabled on a virtual machine (called the Primary VM), a copy of the Primary VM (called the Secondary VM) is automatically created on another host, chosen by VMware Distributed Resource Scheduler (DRS). If VMware DRS is not enabled, the target host is chosen from the list of available hosts. VMware FT then runs the Primary and Secondary VMs in lockstep with each other – essentially mirroring the execution state of the Primary VM to the Secondary VM. In the event of a hardware failure that causes the Primary VM to fail, the Secondary VM immediately picks up where the Primary VM left off, and continues to run without any loss of network connections, transactions, or data. This section provides the highlights of VMware FT. Full details can be found in the following white paper
http://www.vmware.com/resources/techresources/1094.

VMware® vLockstep technology

VMware FT keeps the Primary and Secondary VMs in lockstep using VMware vLockstep technology. vLockstep technology ensures that the Primary and Secondary VMs execute the same x86 instructions in an identical sequence. Here, the Primary VM captures all nondeterministic events and sends them across a VMware FT logging network to the Secondary VM. The Secondary VM receives and then replays those nondeterministic events in the same sequence as the Primary VM, typically with a very small lag time. (The vLockstep architecture is shown in Figure 1.) As both the Primary and Secondary VMs execute the same instruction sequence, both initiate I/O operations. However, the outputs of the Primary VM are the only ones that take effect: disk writes are committed, network packets are transmitted, and so on. All outputs of the Secondary VM are suppressed by ESX. Thus, only a single virtual machine instance appears to the outside world.

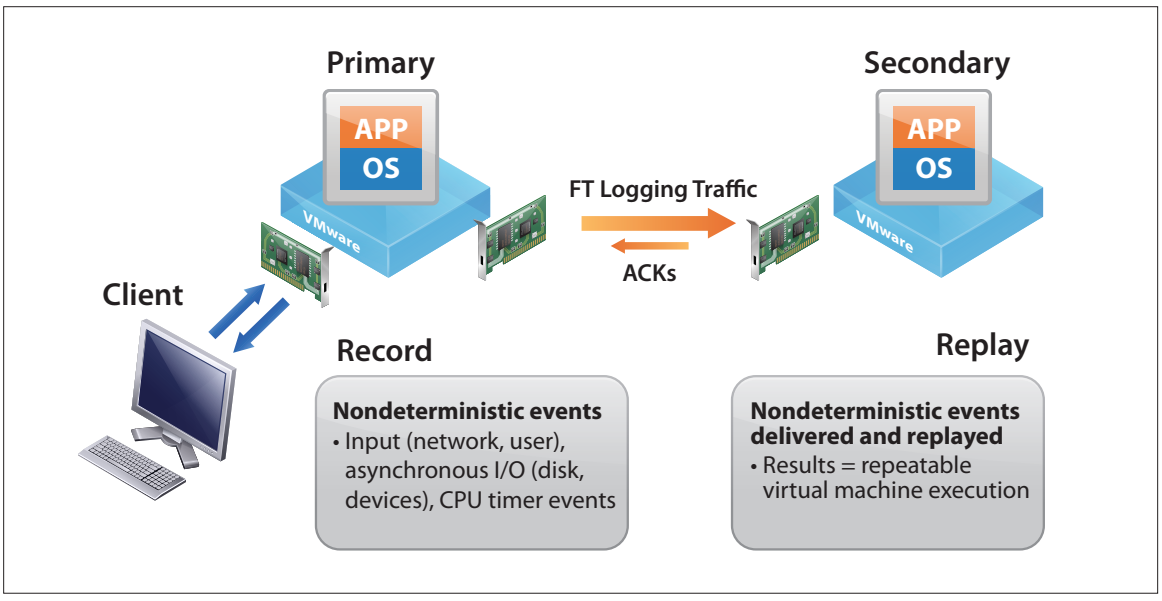

Figure 1: vLockstep architecture

Transparent failover

Along with keeping the Primary and Secondary VMs in sync, VMware Fault Tolerance must rapidly detect and respond to hardware failures of the physical machines running the Primary or the Secondary VM. When vLockstep technology is initiated, the ESX hypervisor starts sending heartbeats over the FT logging network between the ESX hosts where the Primary and Secondary VMs reside. This allows VMware FT to detect immediately if a host fails and execute a transparent failover where the remaining VMware FT virtual machine continues running the protected workload without interruption.

To illustrate the transparent failover functionality, consider a VMware HA cluster of three ESX hosts, two of which are running a Primary and Secondary VM. If the host running the Primary VM fails as shown in step 1 of Figure 2, the Secondary VM is immediately activated to replace the Primary VM, shown in step 2 of Figure 2. A new Secondary VM is created and fault tolerance is re-established in a short period of time as shown in step 3 of Figure 2. Unlike the initial creation of the Secondary VM where DRS chooses the target ESX host, for failovers VMware HA chooses the target ESX host for the new Secondary VM. Users experience no interruption in service and no loss of data during the transparent failover.

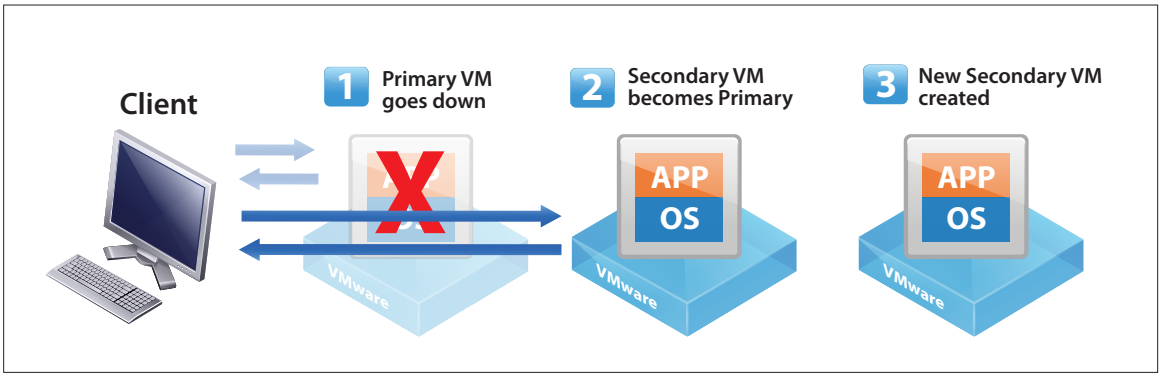

Figure 2: Transparent failover

Lifecycle of a fault-tolerant virtual machine

Turning on and enabling VMware FT for a virtual machine affects the virtual machine's lifecycle, but it is entirely transparent to the end-user client and does not disrupt client connections or the client's workload. The following steps outline the lifecycle of a VMware FT virtual machine:

1. Administrator selects a virtual machine in either the powered-on or off state and turns on VMware FT.
2. The virtual machine becomes the Primary VM and a Secondary VM is automatically created and assigned to an ESX host, sharing the same disk as the ESX host running the Primary VM.
3. If the Primary VM is already powered-on when VMware FT is turned on, its active state is immediately migrated using a special form of VMotion to the Secondary VM on an automatically chosen ESX host. If the Primary VM is powered-off then the migration of its active state to the Secondary VM occurs right after the Primary VM is powered on.
4. The Secondary VM stays synchronized with the Primary VM through VMware vLockstep technology.
5. If the ESX host running the Primary VM goes down, the Secondary VM will immediately "go live" and become the Primary VM.
6. VMware HA automatically starts a new Secondary VM on another available host to restore protection.
7. The Secondary VM is powered off when the Primary VM powers off or when VMware FT is disabled. The Secondary VM is removed altogether when VMware FT is turned off.

General FT Requirements and Recommendations

VMware FT has a list of requirements in order to perform well. Please refer to the Fault Tolerance Configuration Requirements section in the vSphere Availability Guide http://www.vmware.com/pdf/vsphere4/r40/vsp_40_availability.pdf.

The following list summarizes the requirements as well as some configuration and deployment recommendations.

Cluster and host requirements

- VMware FT can only be used in a VMware HA cluster.
- Ensure that all ESX hosts in the VMware HA cluster have identical ESX versions and patch levels. vLockstep technology only works between Primary and Secondary VMs on hosts running identical versions of ESX. Please see the section on Patching hosts running VMware FT virtual machines for recommendations on how to upgrade hosts that are running FT virtual machines.
- Please refer to http://kb.vmware.com/kb/1013637 when mixing ESX and ESXi hosts in the same VMware HA cluster.
- ESX host processors must be VMware FT capable and belong to the same processor model family. VMware FT capable processors required changes in both the performance counter architecture and virtualization hardware assists of both AMD and Intel. These changes could only be included in recent processors from both vendors: third-generation AMD Opteron™ based on the AMD Barcelona, Budapest and Shanghai processor families; and Intel® Xeon® processors based on the Penryn and Nehalem micro-architectures and their successors. For details please refer to http://kb.vmware.com/kb/1008027.
- VMware FT does not disable AMD's Rapid Virtualization Indexing (i.e., nested page tables) or Intel's Extended Page Tables for the ESX host, but it is automatically disabled for the virtual machine when turning on VMware FT. However, virtual machines without FT enabled can still take advantage of these hardware-assisted virtualization features.
- VMware FT is supported on ESX hosts which have hyper-threading enabled or disabled. Hyper-threading does not have to be disabled on these systems for VMware FT to work.

Storage requirements

- Shared storage required – Fibre channel, iSCSI, or NAS.
- Turning on VMware FT for a virtual machine first requires the virtual machines' virtual disk (VMDK) files to be eager zeroed and thick-provisioned. During the process of turning on VMware FT, a message will state this requirement. The message asks whether or not it should convert the virtual disk to the supported format of eager-zeroed and thick-provisioned. The user must convert the virtual disk at this time in order to proceed with turning on VMware FT. Alternatively, the user may wish to convert the virtual disks before they turn on VMware FT to allow for a quicker VMware FT turn-on process at a later time. So, thin-provisioned or lazy-zeroed disks could be converted during off-peak times through two methods:
- Use the `vmkfstools --diskformat eagerzeroedthick` option in the vSphere CLI when the virtual machine is powered off. Please see the vSphere Command-Line Interface Installation and Reference Guide for details: http://www.vmware.com/pdf/vsphere4/r40/vsp_40_vcli.pdf
- Set `cbtmotion.forceEagerZeroedThick = "true"` flag in the .vmx file before powering on the virtual machine. Then use VMware Storage VMotion to do the conversion:

1. In the left-hand pane of vSphere Client, right-click the virtual machine who's disk you wish to convert and select **Migrate**.
2. In the Migrate Virtual Machine dialog box select **Change datastore**, which is the option that initiates Storage VMotion. Click **Next**.

3. Choose a datastore listed that has enough free space to hold the virtual disk. Click **Next**.
4. Select **Thick format**. **Click Next**.
5. Click **Finish**.

- Backup solutions within the guest operating system for file or disk-level backups are supported. However, these applications may lead to the saturation of the VMware FT logging network if heavy read access is performed. In fact, saturation of the FT logging network could occur for any disk-intensive workload. The resulting network saturation may affect and lower the performance of the VMware FT-enabled virtual machine. Do not run a lot of VMware FT virtual machines with high disk reads and high network inputs on the same ESX host. See the section on Placement of fault tolerant virtual machines for a calculation on the amount of network bandwidth required for VMware FT logging.
- VMware Consolidated Backup will be supported in a future release.

Networking recommendations

- At a minimum, use 1 GbE NICs for VMware FT logging network. Use 10 GbE NICs for increased bandwidth of FT logging traffic.
- Ensure that the networking latency between ESX hosts is low Sub-millisecond latency is recommended for the FT logging network. Use vmkping to measure the latency.
- VMware vSwitch settings on the hosts should also be uniform, such as using the same VLAN for VMware FT logging, to make these hosts available for placement of Secondary VMs Consider using a VMware® vNetwork Distributed Switch to avoid inconsistencies in the vSwitch settings.

Baseline recommendation:

Preferably, each host has separate 1 GbE NICs for FT logging traffic and VMotion. The reason for recommending separate NICs is that the creation of the Secondary VM is done by migrating the Primary VM with VMotion. This can produce significant traffic on the VMotion NIC and could affect VMware FT logging traffic if the NICs are shared. In addition, it is preferable that the VMware FT logging NIC has redundancy, so that no unnecessary failovers occur if a single NIC is lost. As described in the steps below, the VMware FT logging NIC and VMotion NIC can be configured so that they will automatically share the remaining NIC if one or the other NIC fails.

1. Create a vSwitch that is connected to at least two physical NICs.
2. Create a VMware VMkernel connection (displayed as VM kernel Port in vSphere Client) for VMotion and another one for FT traffic.
3. Make sure that different IP addresses are set for the two VMkernel connections.
4. Assign the NIC teaming properties to ensure that VMotioand FT use different NICs as the active NIC:
   - For VMotion: Set NIC A as active and NIC B as passive.
   - For FT: Set NIC B as active and NIC A as passive.

Note that it is possible to run VMware FT with just a single NIC. The vSwitch stack is flexible enough to route all the traffic (e.g., console, virtual machine, VMware FT, VMotion) through one NIC. However, this configuration is strongly discouraged, since VMware FT will perform better and more reliably with redundancy at all levels of the system.

For details on what VMware FT logging traffic consists of please refer to the section on Placement of fault tolerant virtual machines.

Not supported:
Source port ID or source MAC address based load balancing policies do not distribute FT logging traffic. However, if there are multiple VMware FT host pairs, some load balancing is possible with an IP-hash load balancing scheme, though IP-hash may require physical switch changes such as ether-channel setup. VMware FT will not automatically change any vSwitch settings.

VMware FT Usage Scenarios

VMware FT can be used in conjunction with other technologies to provide a range of continuous and high availability options. The following list provides recommendations for using VMware FT in different scenarios.

This section describes the ability to run VMware FT and VMware HA virtual machines together, a process called VMware FT On-Demand, and finally the best approach for patching hosts that are running VMware FT virtual machines.

Different methods of protecting virtual machines
Different virtual machines require different levels of protection. Since VMware FT can only be turned on for virtual machines in a VMware HA cluster, by default all virtual machines in a VMware HA cluster are protected by either VMware HA or VMware FT. VMware FT can be used to protect mission-critical workloads, while VMware HA protects the other workloads by restarting the virtual machine in the event of a virtual machine or ESX host failure.

Running VMware FT and VMware HA virtual machines on the same ESX host is fully supported. VMware HA also helps protect VMware FT virtual machines in the unlikely case where the ESX hosts running the Primary and Secondary VMs both fail. In that case, VMware HA will trigger the restart of the Primary VM as well as re-spawn a new Secondary VM onto another host. Note that if the guest operating system in the Primary VM fails, such as resulting from a blue screen in Windows, the Secondary VM will experience the same failure. The VMware HA feature called VM Monitoring will detect this Primary VM failure through VMware Tools heartbeats and VMware HA will automatically restart the failed Primary VM and re-spawn a new Secondary VM.

Together, VMware HA and VMware FT deliver the highest possible availability to virtualized environments.

VMware FT on-demand
The process of turning on VMware FT for a virtual machine takes on the order of minutes. Turning off VMware FT occurs in seconds. This allows virtual machines to be turned on and off on-demand when needed. Turning on and off VMware FT can also be automated by scheduling the task for certain times using the vSphere CLI. During critical times in your datacenter, such as the last three days of the quarter when any outage can be disastrous, VMware FT on-demand can be scheduled to protect virtual machines for the critical 72 or 96 hours when protection is vital. When the critical period ends VMware FT is turned off again, and the resources used for the Secondary VM are no longer allocated.

Patching hosts running VMware FT virtual machines
VMware recommends applying the latest patch and update bundles to ESX 4 hosts using VMware Update Manager or the esxupdate utility. Please see Best Practices for Patching ESX http://www.vmware.com/resources/techresources/1075 and the ESX 4 Patch Management Guide http://www.vmware.com/pdf/vsphere4/r40/vsp_40_esxupdate.pdf for further details on ESX patching. When ESX hosts are running VMware FT virtual machines, the ESX hosts running the Primary and Secondary VMs must be running the same ESX version and patch level. This requirement must be carefully considered when updating the ESX hosts. The following two approaches are recommended for patching ESX hosts with FT virtual machines.

The first approach is suggested for environments where disabling VMware FT for virtual machines can be tolerated for the amount of time required to update all ESX hosts in the cluster. Perform this approach as follows:

1. For each virtual machine protected by VMware FT in the cluster, right-click the virtual machine, highlight **Fault Tolerance** and select **Disable Fault Tolerance** (**note:** turning off VMware FT would work but turning it back on later would take longer).
2. Proceed to update the hosts in the cluster using Update Manager or esxupdate as normal. These hosts should be patched or upgraded to the same set of baselines. After patching, a visual inspection of the system and patch levels using Update Manager or esxupdate is recommended.
3. After updating all hosts in the cluster to the same version and patch level right-click each virtual machine you wish to protect with VMware FT, highlight **Fault Tolerance**, and select **Enable Fault Tolerance**.

Please note that the performance data of the Secondary VM will be lost when you turn off VMware FT for the virtual machine. This data is not lost when you disable VMware FT.

The second approach can be used to ensure that VMware FT is disabled for virtual machines for a minimal amount of time, but it can only be applied to clusters with four or more hosts. In this approach, virtual machines are unprotected only for the time it takes to disable FT, VMotion the virtual machine to another host, and then re-enable VMware FT. Using three hosts is possible but FT would be disabled for a longer period of time, as described in the first approach above. The following steps outline this approach:

1. Use VMotion to manually migrate all VMware FT virtual machines onto at most half the ESX hosts in the cluster. Assuming you follow the four-host minimum requirement, then all VMware FT virtual machines would be migrated to at least two ESX hosts. If there are not enough resources on half the ESX hosts in the cluster to provide the required service level agreements for those virtual machines' workloads, then use the first approach mentioned above.
2. Run VMware Update Manager or esxupdate on the hosts that are not running the FT virtual machines. These hosts should be patched or upgraded to the same set of baselines. After patching, a visual inspection of the system and patch levels using Update Manager or esxupdate is recommended.
3. For each FT virtual machine, disable VMware FT (**note:** turning off VMware FT would work but turning it back on later would take longer) and migrate the Primary VM to the newly updated hosts. Enable VMware FT immediately after the virtual machine is migrated to an updated host. The Secondary VM will be automatically placed on a host with the same system and patch level.
4. Run Update Manager or esxupdate on the remaining un-updated hosts. These hosts should not be running any VMware FT virtual machines. Again, these hosts should be patched or upgraded to the same set of baselines.
5. Use VMotion to manually migrate FT virtual machines to the newly updated hosts.

Thus, the first approach is simpler and is targeted at environments where clusters have three ESX hosts or fewer and can tolerate VMware FT being disabled for a short duration. The second environment is targeted towards larger environments where VMware FT can be disabled for a minimal amount of time.

Recommendations for Reliability

Removing single points of failure from your environment is the most important practice in increasing reliability. Reduce single points of failure by implementing multiple NICs, multiple HBAs, multiple power supplies, storage RAID, etc. Fully-redundant NIC teaming and storage multi-pathing are recommended to improve reliability as shown in Figure 3 below. VMware FT does attempt a failover if the Primary VM loses all paths to fibre channel storage and the Secondary VM still has connection to fibre channel storage, but customers should not rely on this. Instead they should implement fully-redundant NIC teaming and storage multi-pathing. For NIC teaming, please see the section on Networking Recommendations. For details on storage multi-pathing please refer to the Fibre Channel SAN Configuration Guide http://www.vmware.com/pdf/vsphere4/r40/vsp_40_san_cfg.pdf.

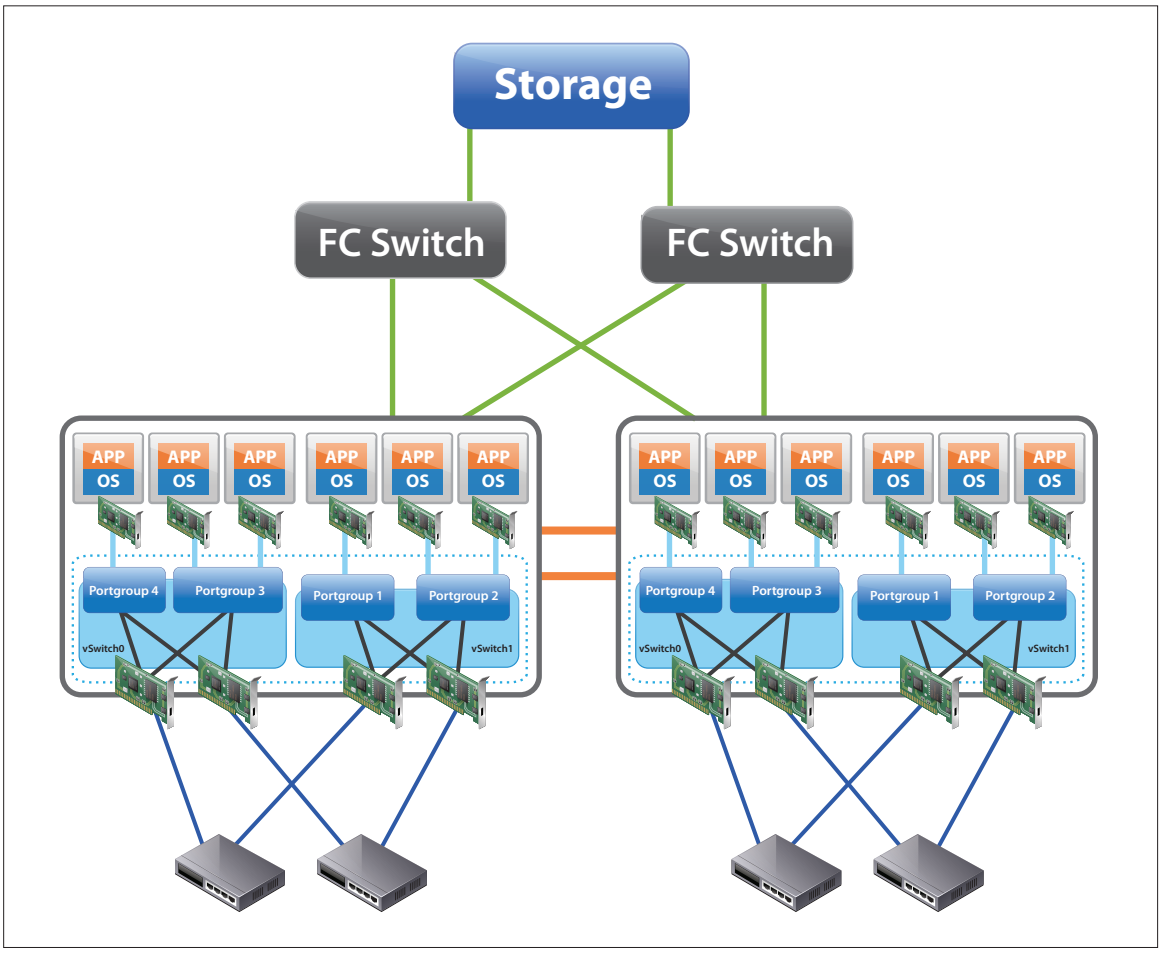

Figure 3: Diagram of networking and storage redundancy

Other recommendations to improve reliability include:

1. Ensuring VMotion and VMware FT logging NICs use a private network.
2. Using vNetwork Distributed Switches for all networks and hosts.
3. Minimizing VMotion migrations of the Primary or Secondary VMs to reduce network and compute resources required by VMware FT. The administrator may also prefer to keep the Primary and Secondary VMs on specific hosts.
4. Ensuring that ESX hosts deliver consistent CPU cycles by making the power management usage consistent among hosts.
5. When using network-attached storage (NAS), ensure that the NAS device itself has sufficient resources

Uniformity of Hosts

The ESX hosts in your cluster should be as uniform to each other as possible – as described in the Cluster and host requirements section. For better performance, the hosts running the Primary and Secondary VMs should operate at roughly the same processor frequencies in order to ensure the highest level of fault tolerance. Processor speed differences greater than 400 MHz in frequency may become problematic for CPU-bound workloads.

In VMware HA environments, it is difficult to predict where the Primary and Secondary VMs will be powered on. Therefore, it is best if all the hosts in the cluster have approximately uniform processors. Even in cases where the processor frequency is roughly the same between two hosts, platform power management settings such as power capping and enforced low frequency modes to save maximum power – which do not adjust based on workload – can cause the frequencies to vary greatly. CPU frequency scaling may cause the Secondary VM to run slower than the Primary VM and will cause the Primary VM to slow down. It is therefore recommended that BIOS-based power management features be used consistently across hosts and that certain settings should be avoided on hosts with VMware FT virtual machines.

VMware® Distributed Power Management (DPM) will not recommend a host for power off unless it can successfully recommend VMotion migrations of all virtual machines off that host. Since VMware FT virtual machines are VMware DRS disabled and cannot be migrated by VMotion recommendations, VMware DPM will not recommend powering off any host with running VMware FT virtual machines. However, VMware DPM can still be enabled on a VMware HA cluster running VMware FT virtual machines and will simply provide power on or off recommendations for hosts not running VMware FT virtual machines.

Placement of Fault Tolerant Virtual Machines

VMware FT creates Secondary VMs and places them onto another ESX host. If VMware DRS is enabled, DRS decides the target host for the Secondary VM when VMware FT is turned on. If DRS is not enabled, the target host is chosen from the list of available hosts. After a failover, VMware HA decides the target host for the new Secondary VM. When enabling VMware FT for many virtual machines, you may want to avoid the situation where many Primary and Secondary VMs are placed on the same host. The number of fault tolerant virtual machines that you can safely run on each host cannot be stated precisely because the number is based on the ESX host size, the virtual machine size, and workload factors, all of which can vary widely. VMware does expect the number of supportable VMware FT VMs running on a host to be bound by the saturation of the VMware FT logging network. Given this, it is recommended that no more than four Primary and Secondary VMs be placed onto the same ESX host. For running more than four VMware FT virtual machines on a host, refer to the following:

As described in the section on VMware vLockstep technology, the VMware FT logging network traffic depends on the amount of nondeterministic events and external inputs that are recorded at the Primary VM. Since the bulk of this traffic usually consists of incoming network packets and disk reads one could calculate the amount of networking bandwidth required for VMware FT logging using the following:

VMware FT logging bandwidth ~= (Avg disk reads (MB/s) x 8 + Avg network input (Mbps)) x 1.2 [20% headroom]

The above calculation reserves an additional 20 percent of networking bandwidth on top of the disk and network inputs to the virtual machine. This 20 percent headroom is recommended for transmitting nondeterministic CPU events and for TCP/IP overhead. You can measure the characteristics of your workload through the vSphere Client. Click the **Performance** tab of the virtual machine to see disk and network I/O. Figure 4 below illustrates the traffic that makes up the VMware FT logging bandwidth.

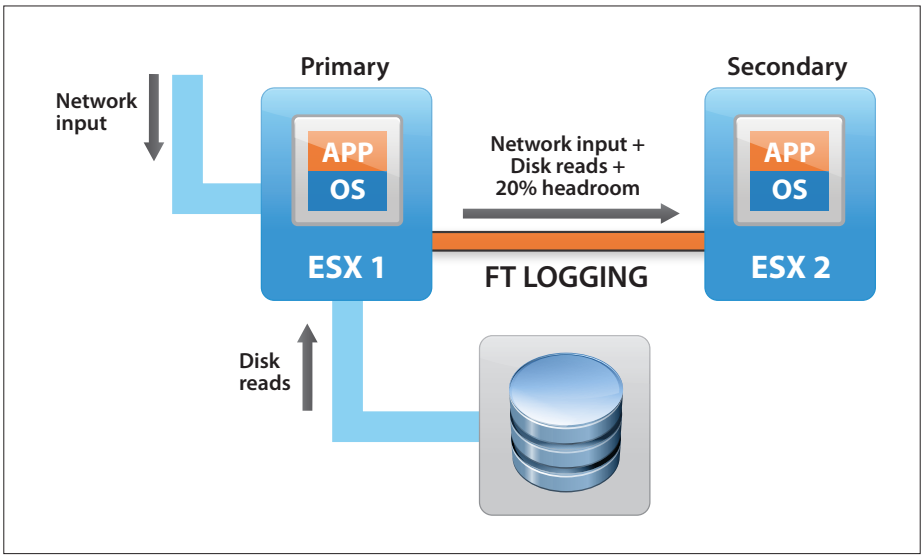

Figure 4: Components of FT logging traffic

When running multiple VMware FT virtual machines on the same ESX host, mix Primary and Secondary VMs together. The bulk of the VMware FT logging traffic flows from the Primary VM to the Secondary VM. Much less traffic flows from the Secondary VM to the Primary VM. Therefore, the bandwidth of the VMware FT logging NICs will be better utilized if each host has a mix of Primary and Secondary VMs, rather than all Primary VMs or all Secondary VMs. Also, the Secondary VM does not perform any I/O to the virtual machine network and disk. So, the utilization of the virtual machine network and disk will also be more balanced if a host has a mix of Primary and Secondary VMs.

Timekeeping Recommendations

In order to avoid time mis-match issues of a virtual machine after an VMware FT failover, perform the following steps:

1. Synchronize the guest operating system time with a time source, which will depend whether the guest is Windows or Linux.
2. Synchronize the time of each ESX server host with a network time protocol (NTP) server.

Windows guest operating system time synch

For Windows Server 2003 guest operating systems, synchronize time with the appropriate domain controllers within their Microsoft Active Directory (AD) domain. In turn, each domain controller should sync their clock with the primary domain controller emulator (PDC Emulator) of the domain. All PDC Emulators should be time synchronized with the PDC Emulator of the root forest domain. Finally the PDC Emulator of the root forest domain should be time synchronized with a stratum 1 time source such as an NTP time server or a hardware atomic clock. Please refer to your Active Directory (AD) documentation for configuration details and for time synchronization of other versions of Windows. If AD is not being used in your environment, synchronize time directly with the NTP time server or another reliable external time source. Please refer to your Windows documentation for details.

Linux guest operating system time synchronization

For Linux guest operating systems, synchronize time with an NTP server by performing the following steps:

1. Open the VMware Tools Properties dialog box from within the guest.

2. Under Miscellaneous Options, make sure "Time synchronization between the virtual machine and the ESX Server" option is not checked.
3. Synchronize time with an NTP time server. Please refer to Installing and Configuring Linux Guest Operating Systems for configuration details. http://www.vmware.com/resources/techresources/1076

ESX server time synchronization
Set the ESX host time to synchronize with an NTP server to ensure the ESX host time is accurate. Perform the following steps to synchronize an ESX host to an NTP server:

1. In vSphere Client, click the **Configuration** tab for the ESX host you want to synch.
2. Click **Time Configuration** under Software.
3. Click **Properties** on the upper, right-hand corner.
4. In the **Time Configuration** dialog box as shown in Figure 5, click **Options.**
5. In the NTP Daemon (ntpd) Options dialog box as shown in Figure 5, click **NTP Settings** on the left pane.
6. Click **Add...** to add the address of the NTP server. Click **OK**.
7. In the NTP Daemon (ntpd) Options dialog box, click **General** on the left pane and select a Startup Policy. Click **OK**.
8. In the Time Configuration dialog box, make sure the NTP Client is enabled. Click **OK**.

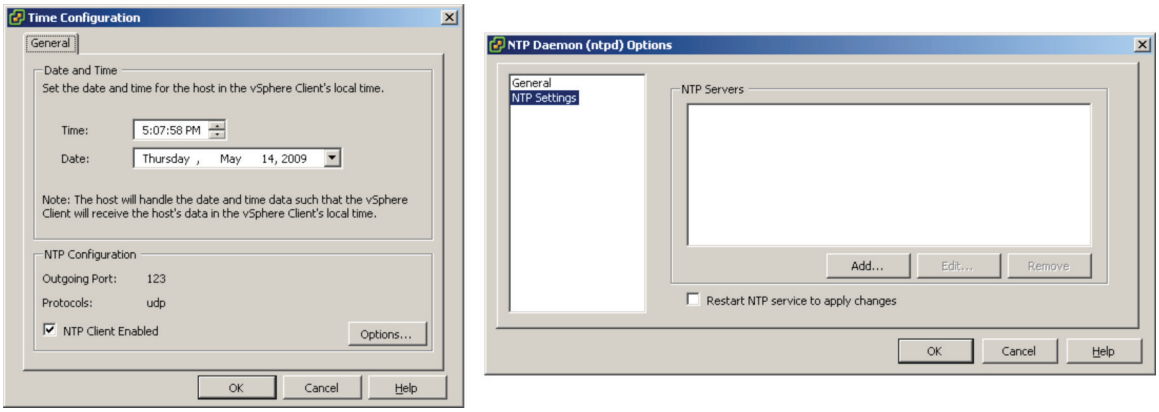

Figure 5: Screenshots for synchronizing ESX with an NTP server

If your guest operating system is very time-sensitive, then synchronize the guest operating system directly with the NTP server. The method to do this varies depending on the guest operating system. Please consult your guest operating system documentation for details.

VMware FT Application Recommendations

VMware FT can be used to protect a wide range of mission-critical, tier-1 applications against server failures. Here are a few example recommendations for protecting applications with FT.

Example 1: High availability for a multi-tiered SAP application
SAP NetWeaver 7.0 is a service-oriented application and integration platform that serves as the foundation for all other SAP applications. Within this multi-tiered SAP NetWeaver 7.0 application, the ABAP SAP Central Services (ASCS) instance is a single point of failure. (ABAP stands for Advanced Business Application Programming.) ASCS is a group of two servers: the Message Server and the Enqueue Server.

1. The Message Server handles all communications in the SAP system. Messaging Server failures cause internal communications between SAP dispatchers to fail. Other problems include failures in user logon and in batch job scheduling.
2. The Enqueue Server manages the logical locks for SAP documents and objects during transactions. Enqueue Server failures result in automatic roll backs of all transactions holding locks and SAP updates that are requesting locks will be aborted.

Since the ASCS is a single point of failure, it requires a high availability solution. For moderate use cases of client connections, a single vCPU virtual machine running ASCS will suffice. Running these services on a single vCPU virtual machine on another host will allow it to be protected with VMware FT. Figure 6 shows the layout of the three-host cluster.

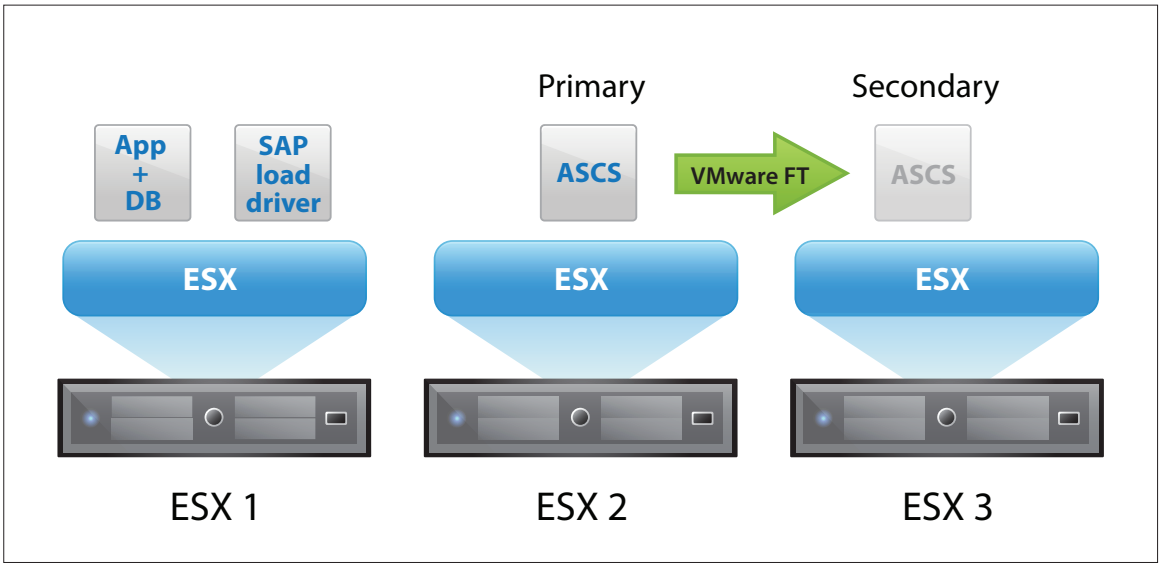

Figure 6: VMware Fault Tolerance on SAP NetWeaver 7.0

ESX #1: Virtual machine with two vCPUs running the database and SAP Central Instance (minus the Message and Enqueue Servers). **Note:** This host is also running an SAP-specific load driver benchmark called the Sales and Distribution (SD) Benchmark. This benchmark was used to validate continuous transaction execution with VMware FT during host failover.

ESX #2: Virtual machine with one vCPU running ASCS (i.e., the Message and Enqueue Servers). This virtual machine has VMware FT turned on and acts as the Primary VM.

ESX #3: Virtual machine with one vCPU acting as the Secondary VM for the ASCS.

Upon failure of either ESX #2 or #3, VMware FT allows the virtual machine on the other host to immediately takeover execution. Thus, the ASCS services will not lose any data and will not experience any interruption in service. This can be tested by manually checking lock integrity via SAP transaction SM12, the SAP lock management transaction. If ESX #1 fails the database (protected via VMware HA) will temporarily go down but will not force a client disconnection for users logged onto separate dialog instance virtual machines (not shown above). The client will only experience a pause until the database comes back online either when the host is rebooted or when the database virtual machine is rebooted on another host through VMware HA.

Example 2: High availability for the Blackberry Enterprise Server

The Blackberry Enterprise Server (BES) 4.1.6 for Microsoft Exchange enables push-based access in delivering Exchange email, calendar, contacts, scheduling, instant messaging, and other Web services to Blackberry devices. Running BES in a single vCPU virtual machine can support up to 200 users that receive an average

of 100-200 email messages per day. Unless there is a failover mechanism in place, the loss of BES due to hardware failure will result in the disruption of Blackberry users' ability to synch with Exchange. VMware FT can be turned on for the BES virtual machine as shown in Figure 7 to provide continuous availability that can survive ESX host failures.

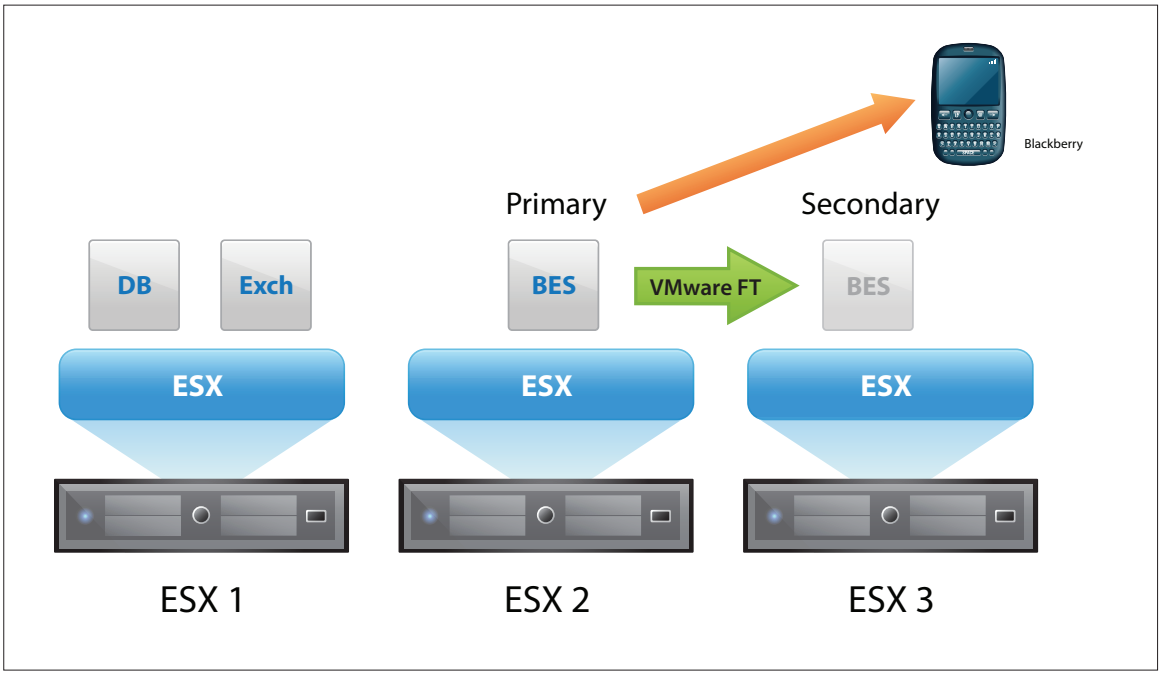

Figure 7: VMware Fault Tolerance on Blackberry Enterprise Server 4.1.6

ESX #1: Virtual machine with two vCPUs running the database and Microsoft Exchange server.

ESX #2: Virtual machine with one vCPU running BES 4.1.6. This virtual machine has VMware FT turned on and acts as the Primary VM.

ESX #3: Virtual machine with one vCPU acting as the Secondary VM for BES 4.1.6.

A failure of either ESX #2 or #3 results in no loss of email delivery to the Blackberry device. VMware FT ensures that the BES workload is uninterrupted. Currently there are a number of different methods to protect BES from failure, ranging from simple backup plans to having offline stand-by servers prepared. However, VMware FT is the only software solution to offer uninterrupted protection for BES service while remaining cost-effective and user-friendly.

Summary of Performance Recommendations

The following points summarize the recommendations provided in "VMware Fault Tolerance Architecture and Performance" and "Performance Best Practices for VMware vSphere™ 4." Please refer to those documents for full details.

- For each virtual machine there are two VMware FT-related actions that can be taken: turning FT on/off and enabling/disabling FT.

  "Turning on FT" prepares the virtual machine for VMware FT by prompting for the removal of unsupported devices, disabling unsupported features, and setting the virtual machine's memory reservation to be equal to its memory size (thus avoiding ballooning or swapping).

"Enabling FT" performs the actual creation of the Secondary VM by live-migrating the Primary VM.

Note: Turning on VMware FT for a powered-on virtual machine will also automatically "Enable FT" for that virtual machine.

Each of these operations has performance implications.

- Do not turn on VMware FT for a virtual machine unless you will be using (i.e., Enabling) VMware FT for that machine. Turning on VMware FT automatically disables some features for the specific virtual machine that can help performance, such as hardware virtual MMU (if the processor supports it).
- Enabling VMware FT for a virtual machine uses additional resources (for example, the Secondary VM uses as much CPU and memory as the Primary VM). Therefore make sure you are prepared to devote the resources required before enabling VMware FT.
- The live migration that takes place when VMware FT is enabled can briefly saturate the VMotion network link and can also cause spikes in CPU utilization.
- If the VMotion network link is also being used for other operations, such as VMware FT logging, the performance of those other operations can be impacted. For this reason, it is best to have separate and dedicated NICs for FT logging traffic and also for VMotion, especially when multiple VMware FT virtual machines reside on the same host.
- Because this potentially resource-intensive live migration takes place each time FT is enabled, it is recommended that VMware FT not be frequently enabled and disabled.
- Because VMware FT logging traffic is asymmetric (the majority of the traffic flows from Primary to Secondary VM), congestion on the logging NIC can be avoided by distributing primaries onto multiple hosts. For example, on a cluster with two ESX hosts and two virtual machines with VMware FT enabled, placing one of the Primary VMs on each of the hosts allows the network bandwidth to be utilized bi-directionally.
- VMware FT virtual machines that receive large amounts of network traffic or perform lots of disk reads can create significant bandwidth on the VMware FT logging NIC. This is true of machines that routinely do these things as well as machines doing them only intermittently, such as during a backup operation. To avoid saturating the network link used for logging traffic, limit the number of VMware FT virtual machines on each host or limit disk read bandwidth and network receive bandwidth of those virtual machines.
- Make sure the VMware FT logging traffic is carried by at least a 1 GbE-rated NIC (which should in turn be connected to at least 1 GbE-rated infrastructure).
- Avoid placing more than four VMware FT-enabled virtual machines on a single host. In addition to reducing the possibility of saturating the network link used for logging traffic, this also limits the number of live-migrations needed to create new Secondary VMs in the event of a host failure.
- If the Secondary VM lags too far behind the Primary VM (which can happen when the Primary VM is CPU bound and the Secondary VM is not getting enough CPU cycles), the hypervisor may slow down execution on the Primary VM to allow the Secondary VM to catch up. This can be avoided by making sure the hosts on which the Primary and Secondary VMs run are relatively closely matched with similar CPU make, model, and frequency. It is recommended to disable certain power management settings that do not allow for adjustments based on workload. As another alternative, enabling CPU reservations for the Primary VM (which will be duplicated for the Secondary VM) will help ensure that the Secondary VM gets CPU cycles when it requires them.
- Though timer interrupt rates do not significantly affect VMware FT performance, high timer interrupt rates create additional network traffic on the FT logging NIC. Therefore, if possible, reduce timer interrupt rates as described in the "Guest Operating System CPU Considerations" section of "Performance Best Practices for VMware vSphere™ 4."

vmware®

VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com