

Build a highly available application platform for J2EE, Part 2: Setting up the hardware platform

Skill Level: Introductory

[Eric Ye](#)

Senior Software Engineer
IBM

[Joseph Kim](#)

Software Engineer
IBM

[Nambi Yogalingam](#)

Software Engineer
IBM

[Lei Zhang](#)

Co-op
IBM

[Veronica Chiu](#)

Co-op
IBM

25 Mar 2005

In this series, you use existing hardware and software from across IBM divisions to produce a complete solution that offers high availability. Follow the steps to discover the baseline of what is possible using current technologies and existing hardware and software products. Then, you'll enhance the system design to take advantage of emerging technologies in the areas of automation, faster failure detection, and multisite failover. This tutorial shows how to set up the the hardware platform for this project with IBM BladeCenter, FAStT storage and IBM TotalStorage® SAN Volume Controller.

Section 1. Before you start

About this tutorial

This tutorial is the second installment in a [series](#) from the Continuous Computing team focusing on building a highly available solution platform for Java™ 2 Platform, Enterprise Edition (J2EE). Using existing hardware and software from across IBM divisions, you learn how to produce a complete solution that offers high availability throughout. In the first phase, we focus on establishing the baseline of what is possible using current technologies. Then, we enhance the design of the system to take advantage of emerging technologies in automation, faster failure detection, and multisite failover. In this tutorial, you'll set up the hardware platform for this project using IBM® BladeCenter™, FAStT storage, and IBM TotalStorage® SAN Volume Controller.

If you want to learn how to set up the hardware for a highly available platform, this tutorial is meant for you. You'll need basic knowledge of storage area networks (SANs), networking, and blade server and storage products to help you complete the tasks.

Prerequisites

To set up the environment described in this tutorial, you'll need:

- IBM TotalStorage SAN Volume Controller, 2 SAN Volume Controller nodes and software
- IBM eServer™ BladeCenter with 14 HS20 blade servers and fiber channel switch module
- IBM 2109 F-16 2109 SAN switch
- IBM FAStT900 storage system with 500GB of disk space
- Red Hat Enterprise Linux™ (RHEL) 3.0

Section 2. Introduction

Overview

The previous article in the series, "[Part 1: Delivering on the continuous computing promise](#)," describes the business need for building a highly available platform for J2EE. Enterprise-class organizations are increasingly dependent on their IT infrastructure to run their businesses. As a result, availability and robustness have become top concerns in managing IT infrastructures. IBM produces many products that can be used to craft these highly available solutions, but because the systems are not necessarily all going to be sold together, each product must be configured, out of the box, to run independently. The challenge is to take all assets and produce a ready-to-go solution to meet performance and scale requirements of customers such as AOL, Amazon, eBay, and Google and still meet the robustness needs of customers such as Schwab, Fidelity, American Airlines, or Daimler-Chrysler.

The charter of the Continuous Computing team is to build a highly available (HA) solution platform for J2EE applications that provides end-to-end availability. In the first phase, completed in January 2005, we integrated and configured existing IBM hardware and software products to showcase what is possible using technologies already available.

This tutorial shows you how to set up the hardware platforms for the project's first phase, which includes setting up and configuring IBM BladeCenter and blade servers, SAN Volume Controller, Fibre Array Storage Technology (FASSt) storage system, and SAN switches.

To enhance the availability level in the project environment, the hardware platform needs to meet several requirements:

- All components must be redundant with automatic failover to eliminate any single point of failure.
- The platform must be scalable so that as more capacity is needed, it can be expanded rather than replaced.
- The platform must be configurable so different types of storage options can be used.
- It needs to provide tools and interfaces that make it easy to automate management of the platform.

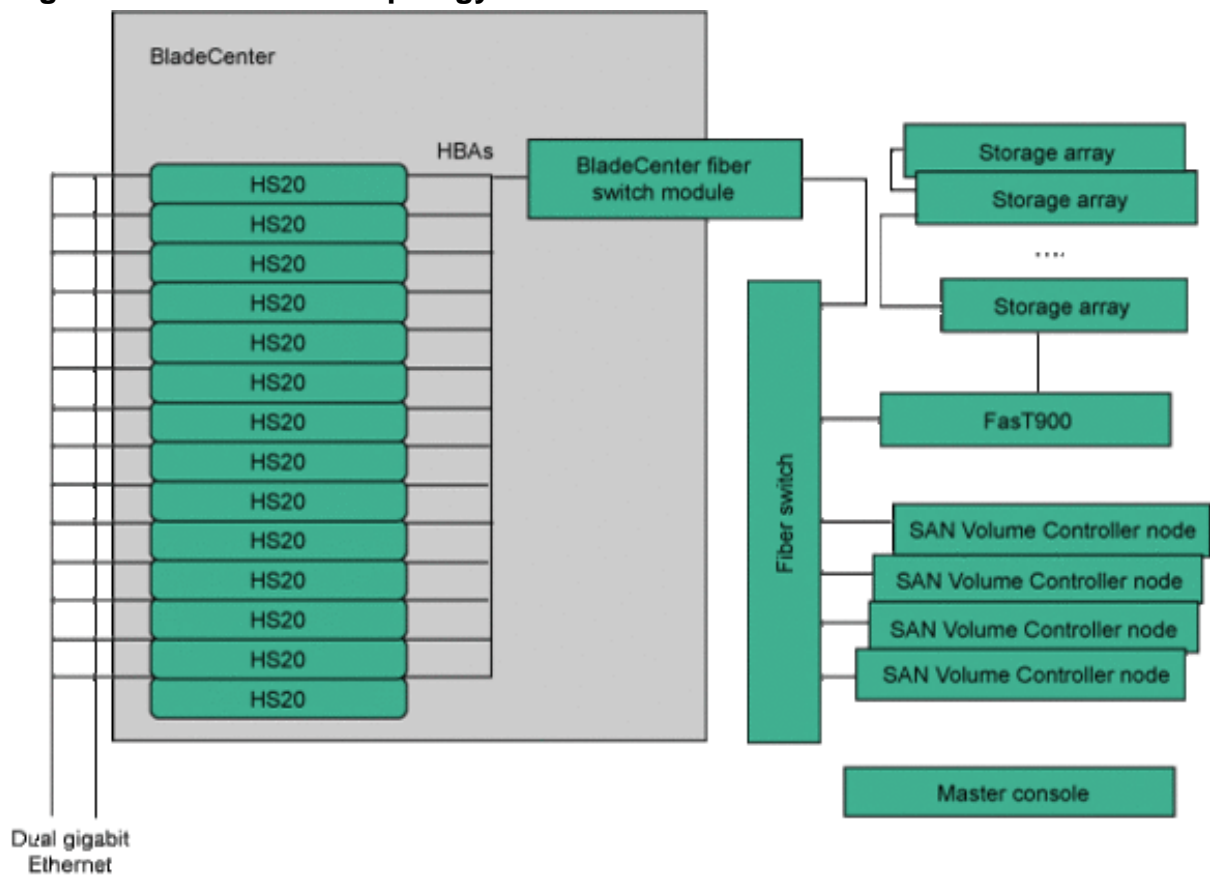
While it's possible to build the solution using standalone servers, it's more efficient to use a chassis and blade type of configuration to take advantage of common power supplies, backplanes, and connectivity. For our solution, it is recommended that all servers be diskless, using remote storage to maximize the safekeeping of data and to minimize the bring-up time required when a server needs to be replaced. For

hardware and network setup, we intend to accomplish:

- Virtualization of storage through SAN Volume Controller for better resiliency and flexibility
- "Diskless blade," so all blades will boot from and operate off virtual logical unit numbers (LUNS)
- Redundancy of hardware components leveraging IBM BladeCenter chassis technology

Figure 1 shows a sample topology chart for the Continuous Computing project environment.

Figure 1. Environment topology



Section 3. Set up FAStT, switches, and volume controller

Hardware models

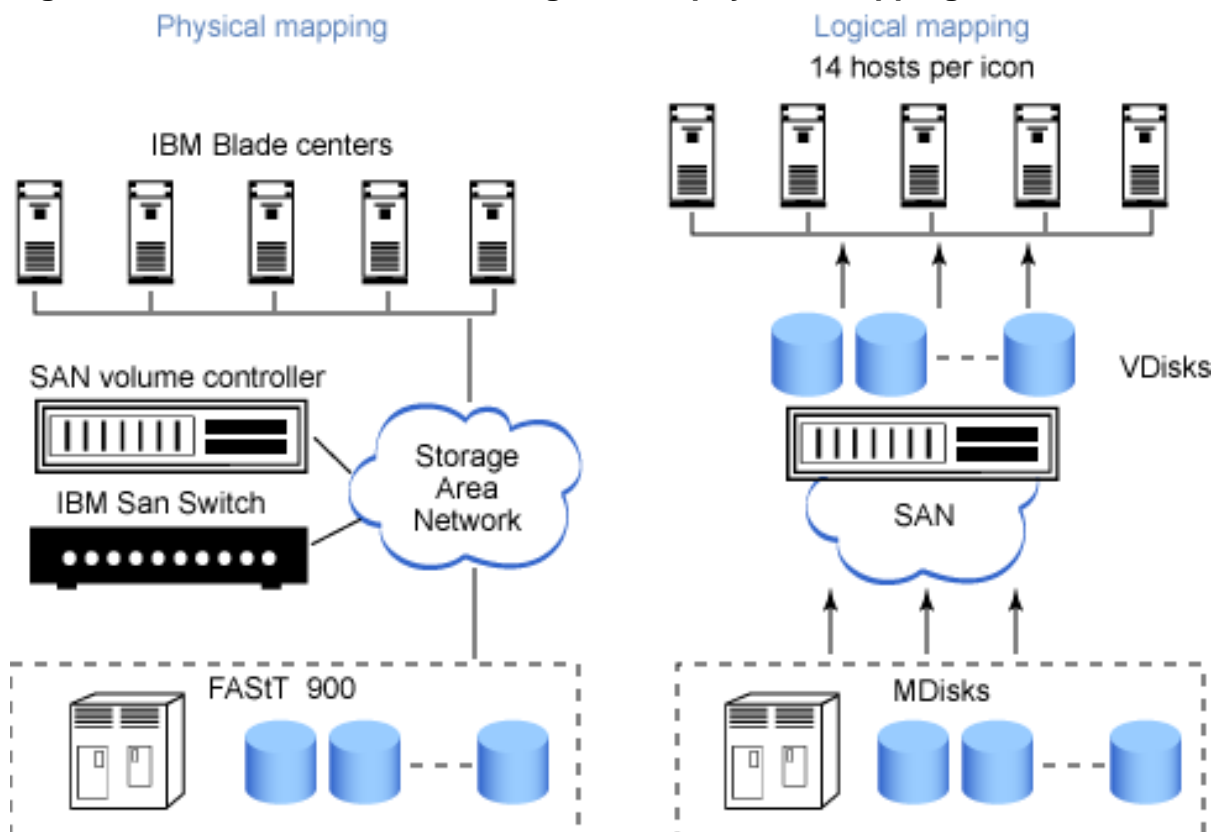
In the sample project environment, the following IBM hardware models are used:

- IBM TotalStorage SAN Volume Controller
- IBM eServer BladeCenter with fibre switch module
- IBM 2109 F-16 2109 SAN switch
- IBM FAStT 900 storage system

SAN topology and components

SAN Volume Controller, FAStT, SAN switches, and the blade center are connected in the topology shown in Figure 2.

Figure 2. SAN Volume Controller logical and physical mapping



The following list describes hardware components and concepts in Figure 2.

- A node is a single SAN Volume Controller. Two nodes are used for virtualization of storage. Each virtualization engine node is an independent Pentium-based server with four connections to the SAN fabric. SAN Volume Controller does not contain any internal battery backup units, so it's connected to an uninterruptible power supply to provide data integrity in case of power failures.
- Fourteen hosts running on BladeCenters with one for each blade servers. The BladeCenter system provides common resources shared by the blade servers. Each blade server's host bus adapter (HBA) is connected to SAN by a shared fibre channel (FC) switch module, Q-logic FC-SW-2 switch. The Continuous Computing project is using one BladeCenter.
- IBM SAN Switch 2109 F-16 is a 16-port SAN switch. The switch provides 2 gigabit per second port-to-port throughput with autosensing capability for connecting to existing 1 and 2 gigabit per second host servers, storage, and switches. The SAN consists of a single fabric through which hosts communicate with SAN Volume Controller. The seven shared BladeCenter FC switch modules and IBM SAN Switch 2109 are merged in the SAN fabric in interoperability mode.
- IBM FAStT solutions are designed to support the large and growing data storage requirements of business-critical applications. The FAStT Storage Server is a Redundant Array of Independent Disks (RAID) controller device that contains FC interfaces to connect the host systems and disk drive enclosures. The storage server provides high system availability by using hot-swappable and redundant components; it features two RAID controller units, redundant power supplies, and fans. All these components are hot-swappable, which assures excellent system availability. A fan or power supply failure will not cause downtime, and such faults can be fixed while the system remains operational. The same is true for a disk failure if fault-tolerant RAID levels are used. With two RAID controller units and proper cabling, a RAID controller or path failure will not cause loss of access to data. The disk enclosures can be connected in a fully redundant manner, which provides a very high level of availability.

For the host side FC connections, you can use up to four mini-hubs. The FAStT Storage Server can support high-end configurations with massive storage capacities (up to 33 Tb per FAStT controller) and a large number of heterogeneous host systems. It offers a high level of availability, performance, and expandability. IBM TotalStorage Server FAStT 900 coupled with EXP700 disk drive enclosures provide total storage capacity of 3.5 terabytes.

- Managed disks (MDisks), also called LUNs, are logical disks that a storage controller has offered on the SAN fabric. A managed disk may consist of multiple physical drives and provides usable blocks of physical storage to the SAN Volume Controller. Managed disk groups are the collection of managed disks.
- Virtual disks (VDisks) are logical disks presented to the SAN by SAN Volume Controller.

Set up the SAN environment

Now let's look at how to set up the SAN environment step by step.

1. Before you start, make sure the BladeCenter manage module is properly configured so the BladeCenter FC switch module can be configured. To access your switch module from an external environment, you may need to enable certain features, such as external ports and external management over all ports.
The BladeCenter SAN utility application is used to access and configure the BladeCenter Fibre Channel switch modules. The SAN utility can be installed on a BladeCenter HS20 blade server or an external network management workstation configured with a supported version of Linux.
2. Set up the management consoles for the IBM F-16 SAN switch and FAStT storage system. In our implementation, the FAStT Storage Manager client resides on a Linux management station that can communicate with the FAStT controllers using the network.
3. Connect both BladeCenter's IBM Qlogic SAN switch and FAStT storage system to IBM F-16 SAN switch through the FC interface. Make sure the SAN-attached hosts are connected to the SAN with two or more FC adapters for redundancy and fault tolerance; this ensures the hosts will have multiple paths to the storage.
4. Set the Qlogic SAN switch and F-16 SAN switch to interoperability mode.
5. On F-16 SAN switches, configure SAN zoning as:
 - Host zone: Visible and accessible to blade servers and SAN Volume Controller. The SAN File System client hosts will be in a zone that includes the SAN Volume Controller node FC ports. The client host's host bus adapter and the back-end storage must not be in the same

zone.

- Create a storage zone: The FC switch needs to be zoned to permit all SAN Volume Controller nodes to see all the storage controllers that contain Mdisks that it will manage. Any existing storage controllers that contain data to be migrated into the SAN Volume Controller will also need to be placed into this zone.
6. On the FASTT storage system, the storage arrays need to be configured. It is recommended that only one LUN be created on an array. The LUN should be sized to use all available storage space on the array. This single LUN will then be used to create smaller Vdisks, which will in turn be presented to the SAN File System as a system or user storage pool volume. The SAN Volume Controller nodes must be able to access all of the back-end LUNs, so the LUNs need to be mapped (LUN masking) to all the available SAN Volume Controller FC WWPNs. There are four ports on each SAN Volume Controller node, and at least two nodes in a cluster. To reduce the likelihood of an Mdisk or Mdisk group being unavailable, configure the FASTT arrays to provide redundancy. For example, only use RAID-1, RAID-3, or RAID-5.
 7. On the F-16 SAN switch, verify that the SAN switches have correctly detected the SAN Volume Controller node FC ports using the switch CLI or Web GUI displays.
 8. On SAN Volume Controller:
 1. Configure MDisks after checking if SAN Volume Controller can access the LUNs, which we will call MDisks. Configured in our storage subsystems, all the Mdisks should be unmanaged since they are not yet included in the SAN Volume Controller and will be assigned default names and ID numbers (for example, mdisk1). Use the `svcinfolsvdisk` command to list MDisks, and use `svctask mkmdisk` to include unmanaged Mdisks to SAN Volume Controller.
 2. Configure MDisk groups. SAN Volume Controller requires that Mdisks be put into managed disk groups before they can have Vdisks created on them. These Mdisk groups will form pools of storage, similar to an array, where you can create Vdisks. Use the `svctask mkmdiskgrp` command to create Mdisk groups, and use `svctask addmdisk` to add Mdisks configured in the previous step to the Mdisk group.
 3. Configure VDisks. Now that we have our pools of storage in the form of managed disk groups containing Mdisks, we can create the virtual disks from the Mdisk groups. We'll need to specify the Vdisk

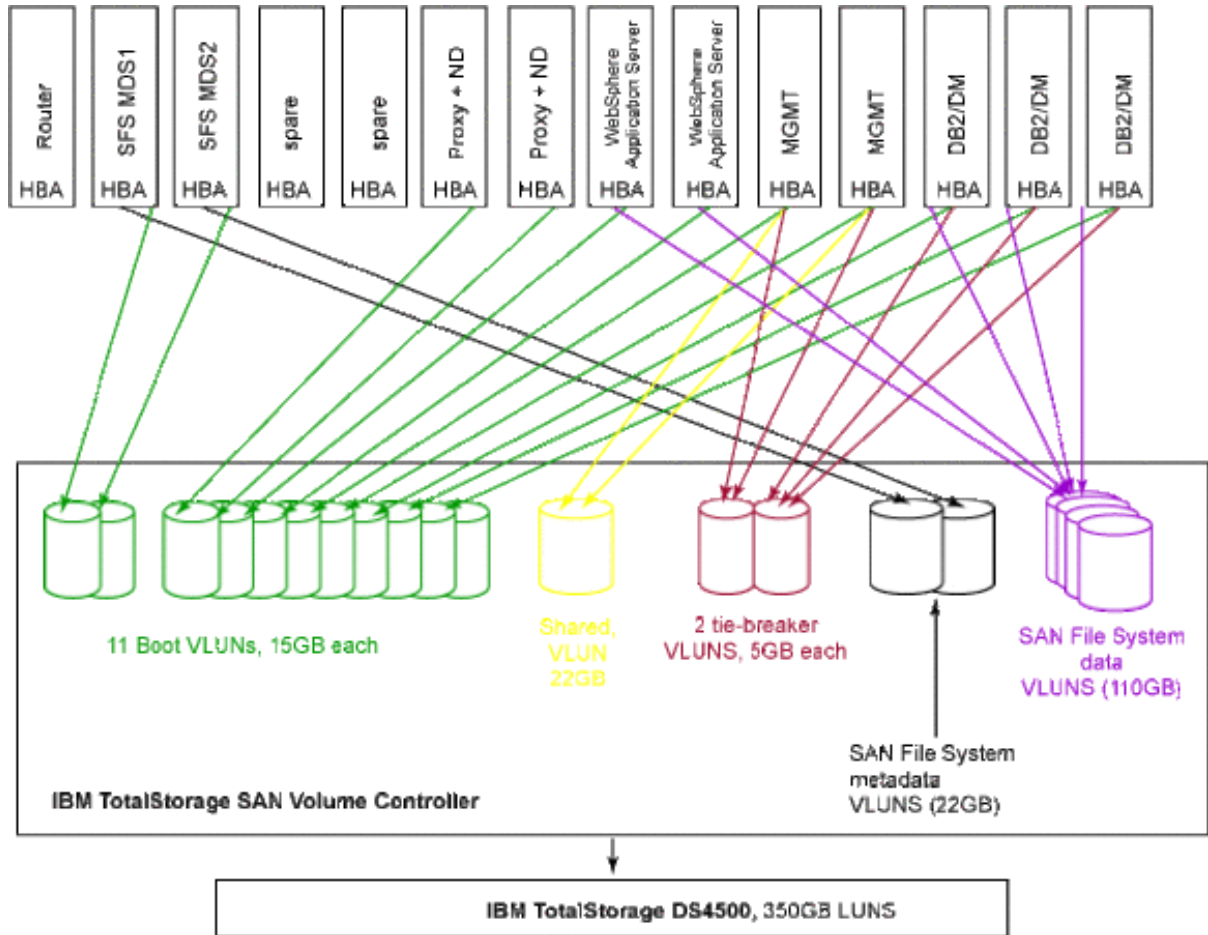
size and type, and can manually select the preferred node within the I/O group if needed. Use `svctask mkvdisk` to create the vdisk.

4. Create host objects. Now that the back-end storage has been defined in the form of Vdisks, it is time to allow the blade server client host servers access to the Vdisks. Do this by creating the SAN Volume Controller host objects to represent the hosts in the SAN that can access the storage.
 1. Use the `svcinfolshbaportcandidate` command to list available host bus adapter ports that can be accessed on the SAN.
 2. Use SAN switch displays or the command line to determine which WWPN belong to which host bus adapter port of the blade server.
 3. Use `svctask mkhost` to create host objects as required.
5. Assign VDisks to hosts corresponding to blade servers. Use `svctask mkvdiskhostmap` to map Vdisks to the host objects that represent the blade servers that need access to the corresponding storage.

Sample SAN storage allocation

Figure 3 shows a sample allocation of Vdisks (also called VLUNs) in our project environment.

Figure 3. VLUNs in the project environment



Section 4. Set up BladeCenter

Realize diskless blade

Our project environment consists of an IBM BladeCenter with 14 HS20 blades, and an Intel® 3.066 GHZ Xeon® processor. Two internal gigabit networks in the backplane provide intra-BladeCenter network connections among the 14 blades. The blades share a network module for connections to the outside world, and a fiber switch module to connect to the IBM 2109 F016 SAN switch, which is then connected to the SAN Volume Controller cluster of two nodes and FastT900 storage system with > 600 GB storage.

As mentioned in the introduction, a main design point in our hardware setup is to

realize diskless blade, which means all blades boot from VDisks instead of local disks. To set up Red Hat Enterprise Linux 3 on the blade to boot from the VDisk:

1. Power on the blade using the blade center management module. Connect to the blade console, and during the boot sequence, press **Ctrl+Q** to enter the QLogic bios configuration (FAST!UTIL) after the console displays that the QLogic driver is loading.
2. In the QLogic BIOS configuration utility, configure the host bus adapter card so that the blade can see the VDisks. To enable the assigned VDisk to be bootable:
 1. Select the first connection in the list and then choose **Scan Fibre Devices** to retrieve the port names. Write down the correct port name.
 2. Select **Configuration** and **Host Adaptor Settings**, to enable the Host Adaptor BIOS.
 3. Select **Selectable Boot Settings** to enable selectable boot. In the primary boot port field, enter the port names you wrote down in the preceding step.
3. Power off the blade using the management module Web admin console or command line to assign the media tray to the blade.
4. Power on the blade again to have the blade boot from the CD-ROM drive loaded with the RHEL 3.0 installation CD.
5. Install Linux RHEL 3 to the VDisk.
(In our environment, each VDisk is displayed as four disks. If this also happens in your environment, make sure the OS is installed to only one instance of the VDisk.)
6. Configure the boot sequence to ensure the blades boots from the VDisk chosen in the previous step. This means interrupting the normal boot sequence and entering the blade BIOS configuration utility. RHEL uses a label to identify which drive to boot off of. However, with four instances from the same VDisk, it will be confused during the boot process. So, after the install finishes, boot off the CD in rescue mode. Mount the drive that RHEL 3.0 is installed on. Change `/etc/grub.conf` and `/etc/fstab` such that it explicitly boots off a VLUN (for example `/dev/sda`) instead of the disk with "LABEL=/".

Finally, the FC card will take over the C: drive. Configure the blade's BIOS

to boot off hard disk 0.

7. If more than one VLUN is assigned to a blade server, you need a workaround. By default, the Linux host system expects a single LUN for each host bus adapter attached. You must modify the kernel so it can detect more than one LUN. To modify the kernel:
 1. Add the following to the `/etc/modules.conf` file:

```
options scsi_mod max_scsi_luns=255
options qla2300 ql2xmaxqdepth=4
```
 2. Issue the `mkinitrd` command to rebuild the RAM disk with the kernel being used.

Section 5. Summary and resources

Summary

In summary, to set up the hardware platform for the Continuous Computing project:

- Connect both the BladeCenter FC switch module and the FASTT storage server to IBM F-16 SAN switch through the FC interface.
- Set IBM Qlogic SAN Switch and IBM F-16 SAN switch to interoperability mode.
- On IBM SAN switches, configure SAN zoning as:
 - Host zone: visible and accessible to blade servers and SAN Volume Controller nodes
 - Storage zone: visible and accessible to SAN Volume Controller nodes and FASTT storage system
- On FASTT storage server, configure LUNs from disk arrays.
- On SAN Volume Controller:
 1. Configure MDisks from LUNs.
 2. Create Mdisk groups and add MDisks to the groups.

3. Configure VDIs on top of MDisk groups.
 4. Create hosts objects to represent blade servers that need access to VDIs.
 5. Assign VDIs to hosts corresponding to blade servers.
- On blade servers:
 1. Configure blade servers to boot from VDIs.
 2. Install RHEL 3.0 on VDIs.
 3. Configure blade servers to boot from the correct VDisk.

Resources

- Refer to these IBM Redbooks™ for general instructions to set up the SAN Volume Controller and FAStT storage, [Implementing Linux with IBM Disk Storage](#) and [IBM TotalStorage: Integration of the SAN Volume Controller, SAN Integration Server, and SAN File System](#).
- For instructions on how to set up the BladeCenter, view this [list of IBM Redbooks and Redpapers on BladeCenter](#).
- Innovate your business with the latest technology from IBM. [Get trial downloads of IBM products now](#).

About the authors

Eric Ye

Eric Ye is part of the IBM High Performance On Demand Solutions (HiPODS) team (formerly High Volume Web Sites). The HiPODS team continues to evolve its work from high-volume Web architectures to high-performance on demand operating environments, with an emphasis on optimizing IT resources and helping IBM customers move toward becoming on demand businesses. They learn customer pain points and requirements, and feed them back to IBM product groups and IBM Global Services. For technical questions, contact Eric at ye@us.ibm.com.

Joseph Kim

Joseph Kim is part of the IBM High Performance On Demand Solutions (HiPODS) team (formerly High Volume Web Sites). The HiPODS team continues to evolve its work from high-volume Web architectures to high-performance on demand operating environments, with an emphasis on optimizing IT resources and helping IBM customers move toward becoming on demand businesses. They learn customer pain points and requirements, and feed them back to IBM product groups and IBM Global Services.

Nambi Yogalingam

Nambi Yogalingam is part of the IBM High Performance On Demand Solutions (HiPODS) team (formerly High Volume Web Sites). The HiPODS team continues to evolve its work from high-volume Web architectures to high-performance on demand operating environments, with an emphasis on optimizing IT resources and helping IBM customers move toward becoming on demand businesses. They learn customer pain points and requirements, and feed them back to IBM product groups and IBM Global Services.

Lei Zhang

Lei Zhang is part of the IBM High Performance On Demand Solutions (HiPODS) team (formerly High Volume Web Sites). The HiPODS team continues to evolve its work from high-volume Web architectures to high-performance on demand operating environments, with an emphasis on optimizing IT resources and helping IBM customers move toward becoming on demand businesses. They learn customer pain points and requirements, and feed them back to IBM product groups and IBM Global Services.

Veronica Chiu

Veronica Chiu is part of the IBM High Performance On Demand Solutions (HiPODS) team (formerly High Volume Web Sites). The HiPODS team continues to evolve its work from high-volume Web architectures to high-performance on demand operating environments, with an emphasis on optimizing IT resources and helping IBM customers move toward becoming on demand businesses. They learn customer pain points and requirements, and feed them back to IBM product groups and IBM Global Services.