



## **Best Practice Paper**

### **An Installation Guide:**

Fujitsu Switches & Servers  
Myricom Cards and Drivers  
Microsoft Compute Cluster Solution (CCS)

*A High Performance Computing (HPC) Solution*

**Fujitsu would like to thank the following contributors:**

**Markus Fischer from Myricom**  
**Greg King from King Charter, Inc.**

**Version: 2a April 2007**

## TABLE OF CONTENTS

Introduction .....	1
Step-by-Step in Brief.....	2
Step-by-Step in Greater Detail.....	4
Step 1: Hardware & Software .....	4
Step 2: Install Fujitsu switches and Fujitsu servers in the rack.....	5
Step 3: Configuring the Fujitsu switch .....	6
Step 4: Setup of Fujitsu servers & install Windows Server 2003 x64 .....	7
Step 5: Install Active Directory on the head node.....	11
Step 6: Install Microsoft Computer Cluster Solution (CCS) .....	12
Step 7: Validate the cluster.....	15
Step 8: Install Myricom drivers.....	16
Step 9: Finishing the setup .....	18
Appendix A: Power and cooling .....	20
Appendix B: Greater detail on Fujitsu Switches .....	21
Appendix C: Fujitsu Server Setup in Greater Detail.....	30
Appendix D: Myricom Driver Performance.....	43
Appendix E: Glossary of Acronyms & Terms .....	48
Appendix F: Other Supportive info & Links .....	51

## **Introduction**

Fujitsu, Myricom and Microsoft have come together to deliver a world class High Performance Computing (HPC) solution. This paper details the path to installing a Fujitsu HPC solution<sup>1</sup> for a maximum of 254 nodes<sup>2</sup>. The solution includes Fujitsu servers<sup>3</sup> and 10 Gigabit Ethernet switches, Myricom cards and drivers, Microsoft Windows Server 2003 x64, Active Directory and Microsoft's Compute Cluster Solution (CCS). Once completed, the cluster will be ready to deploy applications.

### ***First Cluster***

It is intended that the individual using this document understands the fundamentals of setting up servers and switches as well as installing software; However, since this may be the first time setting up a cluster, this document provides greater detail on certain topics so as to ensure the installer's success.

### ***In the Details***

Some steps simply say to install an item while others lead to more detailed instructions. The more detailed instructions and attachments have been provided as an extra guide in the event that detailed help beyond an unassisted installation is required.

### ***Other Papers***

There are certain automation options that are not covered in this document. For example, it will be necessary to manually install software on every server, rather than running a PXE<sup>4</sup>. A future paper will address PXE as well as many other solutions.

### ***Assistance***

Additional papers on a range of HPC topics are forthcoming. If additional information is required regarding the installation of this solution, contact Fujitsu.

### ***Feedback***

Please send Fujitsu any and all feedback.

---

<sup>1</sup> This paper is the first in a series of best practice papers by Fujitsu on HPC. Check back for future papers on an array of HPC topics including parallel file systems, PXE "pixie booting", ADS, SQL, among others.

<sup>2</sup> This is the current practical upper threshold for Microsoft's CCS. For more than 254 nodes, one possible option would be to install SQL on the head node. For more information, visit:  
<http://support.microsoft.com/kb/930057/en-us>

<sup>3</sup> This paper addresses servers not blades. A future paper will address the setup of blades.

<sup>4</sup> Pre-boot eXecution Environment (PXE) Pronounced "pixie" and sometimes referred to as "pixie boot"  
[http://en.wikipedia.org/wiki/Preboot\\_Execution\\_Environment](http://en.wikipedia.org/wiki/Preboot_Execution_Environment)

## **Step-by-Step in Brief**

The following is an abbreviated version of the step-by-step instructions for building a compute cluster. The same steps are explored in greater detail starting on page 4.

### **Step 1: Hardware & Software**

A minimum of one Fujitsu 10 Gb-E switch and two Fujitsu servers are required. Microsoft Compute Cluster Solution (CCS), Primergy ServerView Suite ServerStart, Myricom drivers on a CD, and a few other tools are also required. See page 4.

Ensure that a suitable rack, power and environment for the high performance compute cluster are available and ready for the setup. See Appendix A: Power and cooling on page 20.

### **Step 2: Install Fujitsu switches and Fujitsu servers in the rack**

Install the switch and servers into the rack. As to the number of 10 Gb-E switches that will be needed, one port will be dedicated to each server. See page 5.

### **Step 3: Configuring the Fujitsu switch**

What matters most is that the switch is configured for “jumbo frames.” Please see the instructions on page 6 as well as Appendix B: Greater detail on Fujitsu Switches on page 21.

### **Step 4: Setup of Fujitsu servers & install Windows Server 2003 x64**

It is essential that the 64 bit version of Windows Server 2003 is installed on every server in the cluster. See page 7.

There are a number of important considerations that must be factored during the setup, so be certain to review the details of the entire section beginning on page 7 prior to installing the operating system.

(e.g., each server must have a unique name. Name each server according to established protocols. If there is no naming convention, establish and document a straightforward protocol now. Then nominate one of the servers to be the “head node,” all of the other servers will be “compute nodes.”)

### **Step 5: Install Active Directory on the head node**

Once Windows Server 2003 x64 has been successfully installed on each node, install Active Directory onto the head node ONLY. See page 11.

### **Step 6: Install Microsoft Computer Cluster Solution (CCS)**

Now that Windows Server 2003 x64 is installed on all nodes, and Active Directory has been installed on the head node, it is time to install Microsoft Compute Cluster Solution (CCS). Since Windows Server 2003 x64 has been installed already, the version of CCS to be installed is the Compute Cluster Pack (CCP) and not the Compute Cluster Edition (CCE). Please review the step-by-step instructions on page 12.

### **Step 7: Validate the cluster**

Now that CCP has been installed, it is time to validate the cluster. A set of simple instructions can be found on page 15.

### **Step 8: Install Myricom drivers**

Now that the cluster is running, it is time to install the Myricom drivers. It will be important to review the detail on page 16 as well as Appendix D: Myricom Driver Performance.

### **Step 9: Finishing the setup**

Now that the Myricom drivers are installed, it is time to complete the CCS installation. See page 18.

**Congratulations! The compute cluster is complete.**

For greater detail on the above steps, proceed to the next page.

## Step-by-Step in Greater Detail

### Step 1: Hardware & Software

Although it is possible to setup and validate the servers and switch without first installing the equipment into a rack, it is strongly recommended that switches and servers be correctly installed into quality server racks.

The following hardware is required:

Quantity	Description
1 or more	Quality Server rack with power
1	Fujitsu 10 gigabit Ethernet switch (e.g., XG700, XG2000, XG2000C)
2 or more	Fujitsu servers (e.g., RX220)
1 per server	Myricom Network Interface Cards (NICs)
2	1 Gb-E switches (for setup as well as access to legacy networks)
1	Serial cable (for use during setup)
1	Ethernet cable (for use during setup)
1	Notebook computer (for setup) –or- KVM w/ serial and Ethernet
	CX4 cables to connect the servers to the switches
1	USB Memory Stick (aka, a “Pen drive” or “Thumb drive”)
	Network access
	Internet access

The following software is required:

Quantity	Description
License	Microsoft Windows 2003 Server x64 Standard Edition or higher
License	Microsoft Computer Cluster Solution (CCS): Compute Cluster Pack (CCP) <sup>5</sup>
CDs 1 & 2	Primergy ServerView Suite ServerStart
All	Myricom Drivers (on a CD is preferable, but access online is functional)

Ensure that there is a suitable rack, power and environment for the high performance compute cluster that is being setup. For a table that details the ideal conditions, go to: **Appendix A: Power and cooling.**

<sup>5</sup> There are two versions of CCS. The version covered in this paper is the Compute Cluster Pack (CCP). CCP assumes that a license of Windows Server 2003 x64 is already installed on the server. The second version of CCS is the Compute Cluster Edition (CCE) which includes a restricted version of Windows Server 2003.

## **Step 2: Install Fujitsu switches and Fujitsu servers in the rack**

As to the number of 10 Gb-E switches that are needed, there must be one port dedicated to each server. Count the number of servers that are being setting up, and that dictate the number of switches that are needed. For example: The XG700 has 12 ports. If there are 62 servers, then a minimum of six XG700 switches will be needed; which will leave 10 available ports for expansion on the sixth switch.

As to the servers, it will be necessary to nominate one of the servers to be the “head node”; nominating the server in the first “U” of the rack is the simplest solution. All other servers will become “compute nodes.”

Install the two 1 Gb-E switches for use during setup and for access to legacy networks. One switch will be a private network and the other will be public. Cable the servers to these two switches.

### Step 3: Configuring the Fujitsu switch

Do the following:

1. Power up the switch. Wait for status LED to stop blinking. Confirm that Alarm LED does not light.
2. Connect serial cable to management serial port. For more information on accessing the switch (and other info on the switch) see **Appendix B: Greater detail on Fujitsu Switches**.
3. Start a terminal or telnet session and login to switch.

The default login and password is as follows:

```
xg login: admin  
Password: password
```

Enter the following text when prompted:

```
xg> enable  
xg# configure terminal
```

Then enter the configuration commands, one per line. End with CNTL/Z.

```
xg(config)# bridge jumbo-frame 9216  
xg(config)# exit  
xg# copy running-config startup-config
```

4. Execute the "show bridge" command. Inspect the output to assure that the cut-through switching is enabled AND be certain that jumbo-frames are enabled and frame size is set to 9216 bytes.
5. Execute the "show interface" command. Inspect the output to assure that all ports report flow control is RX and TX.
6. Execute the "show log error" command and verify no errors have occurred in the switch.
7. Execute the "show maintenance" command and verify that no system dumps have occurred (this command normally does not generate any output).
8. The switch is ready for connection to the cluster servers.
9. If a problem occurs, refer to Chapter 6 and 7 of the Fujitsu Switch User Guide for troubleshooting assistance.

## **Step 4: Setup of Fujitsu servers & install Windows Server 2003 x64**

Once the Fujitsu hardware has been assembled as described in Step 3, it is time to install Microsoft Windows onto each server.

### **Important Notes:**

- #1 Microsoft Compute Cluster Solution requires that each server has a unique name. Give each server a logical and unique name. For example: name the server in the top “U” slot in the first rack “Fujitsu001” and the second “U” “Fujitsu002” and so on through to “Fujitsu254.”
- #2 A head node (verses a compute node) must be nominated. When naming the head node, add an identifier to the name of the head node. For example: “Fujitsu001head” and then the others would be “Fujitsu002” through “Fujitsu254.” To simplify things, it is recommended that the server in the top slot of the rack, or the first “U,” becomes the head node; if for any reason a server other than the first “U” is nominated to be the head, please write the name and number of the server down and keep this information in or near the rack (a physical note might be place on or near the server itself).
- #3 Once Windows Server 2003 x64 is successfully installed onto the head, Active Directory will also be installed on the head. This will be addressed in the step-by-step instructions below.<sup>6</sup>
- #4 Windows Server 2003 x64 will need to be installed onto each server separately. A future paper will explain how to use PXE and ADS to dynamically install Windows Server 2003 x64 across many nodes.<sup>7</sup>
- #5 Once the installation of Windows Server 2003 is completed and each server restarts, updates to Windows will need to be run on each server. For limited bandwidth situations, it is strongly recommended that the download of updates on each system is timed in such a way that downloads are consecutive and not simultaneous (i.e., one after another and not all at the same time).

---

<sup>6</sup> This document assumes that the cluster is being built in a self-contained environment where Active Directory does not currently exist. It is possible to integrate Microsoft Compute Cluster Server 2003 within an existing Active Directory environment. That process is out of scope for this document.

<sup>7</sup> Windows Compute Cluster Server 2003 natively supports automated deployment of compute nodes via Remote Installation Service (RIS). When deploying via RIS, it is necessary to integrate the proper storage drivers for Fujitsu server into the compute node system images. Those integration processes are out of scope for this document. For further information, consult Microsoft Compute Cluster Server 2003 documentation and <http://technet2.microsoft.com/WindowsServer/en/Library/96969653-bd0f-44d7-af4f-f95c3016d2be1033.mspx> for details.

### Required Software

✓ Primergy ServerView Suite ServerStart CDs 1 & 2
✓ Microsoft Server 2003 x64
✓ Fujitsu, Myricom, and Microsoft drivers

### Required Hardware

✓ Terminal access to the Fujitsu server (e.g., via a keyboard video mouse (KVM) clamshell terminal in an assembled rack)
✓ At least two Fujitsu servers (e.g., Fujitsu RX220) Note: Each server will need to be installed separately.
✓ At least one Fujitsu 10 Gb-E switch with at least one available port for each server (e.g., XG series such as the XG700)
✓ A removable USB flash drive (a.k.a., a “thumb driver” or “pen drive”) for use during the installation.

### Setup the BIOS

1. When the server starts up, hold down F2 to enter the Bios setup.
2. Under “Main,” Go to “Boot Option,” to “Boot device priority.” Set 1st Boot Device to CD/DVD, and 2nd Boot Device to RAID.
3. Under “Advanced,” make certain that option is either “Default” or “Enabled.”
4. Exit out and save all changes.

### Setup the Server & Install Windows

5. Put a removable USB flash drive in one of the USB ports on the Fujitsu server that is being installed.
6. Open the CD/DVD drive on the Fujitsu server and insert the CD:
7. Primergy ServerView Suite ServerStart CD 1
8. Power on the server.
9. A prompt to set up the array is next, choose: Auto setup (There will be an option to customize setup. This paper addresses an auto setup.)
10. Next will be a prompt to save a configuration file. Select the radio button for “Status Backup Removable Media.” Click “Create.” Select the

removable USB flash drive and click “Ok.” (Special Note: Make certain that the USB removable drive is in the USB port of the server on which the installation is being performed.)

11. Leave all options as default and click “Ok.”
12. Click on “Initialization of ServerStart core running.” (There may be a runtime error, if so simply click “Ok.”)
13. Select “Prepare and/or initiate an operating system installation”.
14. Select “Microsoft Windows.”
15. Select “Microsoft Windows Server 2003.”
16. Select “Prepare and initiate an unattended installation of Microsoft Server 2003 x64.” (Note: Again, 64 bit servers and software are required.) Click “Ok” to start the installation.
17. Click to “Open.” Open changes to continue.
18. Click to “Start Wizard.”
19. In the wizard, it will say: “System successfully detected.” Check “Rack Model” if the servers are in a rack.
20. Click “Next” through all screens (e.g., SKU Wizard, Boot Dog, BMC IP Configuration, etc.) without making any edits, until presented with the screen: “Configuration for Disks & RAID controller”.
21. On the screen for “Configuration for Disks & Raid controller,” select the Radio Button for “Drive View,” click “Add Partition” and click “Next.”
22. At the “Windows Installation” screen, enter a unique password for the administration account. Document this password in a secure location other than this server. Click “Next.”
23. On the “Computer Identification” screen, select the radio button for “Operating System source CD ROM.”
24. Optionally check the box for an R2 installation. (Special Note: The R2 installation CD is required to install R2. Access to the network to perform a network installation of R2 is not available at this stage.)
25. Click “Next.”
26. Set the time zone on the “Time Zone” screen. Click “Next” until presented with the “User Name” screen.

27. At the “User Name” screen, assign a computer name and enter a valid Windows Server 2003 product key. (Note: This number is located with the Microsoft Windows documentation and may be located on the sleeve or jewel case for the installation CD.)
28. Click “Next” until presented with the screen for “Save ServerStart Configuration File.” Save the configuration file on the USB removable drive. (Again, make certain that the USB removable drive is in the USB port of the server on which the installation is being performed.)
29. Click to “Start the Installation of Windows Server 2003 x 64.”
30. When prompted, remove disk 1 and insert disk 2 and click “Ok.”
31. The server will restart.
32. Once restarted, a prompt will be presented to select “ServerStart status backup media”. Select the radio button for “Removable Media” and click “Ok.”
33. After the unattended installation runs, there will be a prompt to insert the Windows CD. Click “I agree” to install Windows.
34. Remove all media and click “Ok.”
35. **Congratulations:** Once the installer finishes, Windows Server 2003 is completely installed.
36. On the first restart of Windows Server 2003 make certain that each server is connected to the Internet so that Windows updates and QFE’s (an RIS hotfix) will run. Also install .NET 2.0 and MMC 3.0.
37. **Hint regarding the next step:** At the beginning of the installation instructions for Active Directory it will be necessary to create two logical partitions on the head node. This step is essential, so be sure to review and perform that step.

For more granular detail on the setting up of Fujitsu servers, please see **Appendix C: Fujitsu Server Setup in Greater Detail.**

## Step 5: Install Active Directory on the head node

The cluster must have at least two servers. One server will be the head node, the second and any additional servers will be the compute node(s).

Once Windows Server 2003 x64 is successfully installed on each node, Active Directory will need to be installed onto the head node ONLY.

1. There MUST be two logical partitions on the head node. If there are not two logical partitions on the head node at this point, create the partitions now. It is recommended that the partitions be allocated 50/50 (meaning each partition should be half of the available drive space). One partition is for Windows 2003 Server, Active Directory (AD) and CCS, the other partition is for RIS and image storage. *These instructions assume that there is a minimum of 250GB drives in each server.*
2. Insert the Windows Server 2003 installation CD 1 into the head node CD drive.
3. Go to the “Start” menu, and select “Configure your Server Wizard.”
4. Select the radio button for “Typical configuration for the first server.” Click “Next.”
5. Do nothing to the NetBios. Click “Next.”
6. Do nothing to the “Summary.” Click “Next.”
7. There will be a notification that the server will “Restart” when the installation is complete. Click “Ok.”
8. A RAS (Remote Access Server) dialog will be presented. Click “Ok.” (Special Note: As it is not compatible with other components in the CCS solution, RAS will be removed later.)
9. Click “Next,” then click “Finish.”
10. **Simple test:** Log out and back in to make certain that AD works.

When logging out & logging back in, make certain that this is performed under a domain account so that domain authentication can be verified.

In the logon dialogue window, type in the user account and password, and in the “domain” field select the new domain.

## **Step 6: Install Microsoft Computer Cluster Solution (CCS)**

With Windows Server 2003 x64 installed on all nodes, and Active Directory installed on the head node, it is now time to install Microsoft Compute Cluster Solution.

As a reminder, since Windows Server 2003 x64 has been installed, it is essential to install the Compute Cluster Pack (CCP) and not the Compute Cluster Edition (CCE).

As a reminder, a basic cluster requires a minimum of two servers. One server will become the head node, while the second (and any additional) server will become the compute node(s).

**Special Note:** CCS runs on top of Windows Server 2003 Standard x64 Edition or higher. If Windows Server 2003 x64 is not yet installed and running on all servers, please do so now. At this point, Windows 2003 Server x64 should already be installed on all servers. Additionally, Active Directory should be installed onto the head node.

**Background:** CCP provides the components that turn a server into a cluster node that can then serve as either a head node or a compute node. CCP is a combination of interfaces, utilities, and management infrastructure that enables high performance computing (HPC) on Windows servers.

1. Make certain that all servers are on the same physical network so that they can see and “ping” each other.
2. It is now time to install the Compute Cluster Pack (CCP) on all servers: both head and compute nodes.
3. To get started, insert the CCP installation CD into the head node’s CD drive.
4. Make certain that the head node server is selected.

**Simple test:** Remove all CDs or DVDs from all of the servers/nodes except for the head node. Then, from the KVM, if a CD is present, the head node is in fact selected.

### **Create a New User**

5. Create a new user by going to the “Start” menu, then to “All programs,” then to “Administrator Tools,” then to “Active Directory Users and Computers.”
6. In the Active Directory Users and Computers window, go to the “Users Folder” and right click using the KVM’s mouse.

7. Click on “User” to create a new user.

**Important Note:** Make sure that all boxes are unchecked before clicking “Next.”

8. Name the new user. For example: Cluster Administrator
9. Click “Next” and then click “Finish.”
10. Now that the cluster administrator has been created, it is time to run the setup. Go to the “Start” menu, to “Run,” to “Browse CD,” to “CCP,” then double click on the “Setup.exe.”
11. Read and accept the Software License Agreement by selecting the radio button and click “Next.”

**Special Note:** If the company does not accept the terms of the software license for any reason, do not accept the terms and the installation will end here.

12. Select the radio button to “Create a new compute cluster with this server as a head node” and then click “Next.”
13. Then install components via the Microsoft Compute Cluster Pack Installation Wizard.
14. Once the above has been completed, a list of flags and errors will be presented in the Cluster Deployment Task To Do List.
15. To address the flags and errors, go to the “Networking” tab, to “Configure,” to the “Cluster Network Topology Wizard” and click “Next.”
16. Select a “Public Network Adapter” that is connected to the Gb-E network switch. Click “Next,” then click “Finish,” then click “Close.”
17. Ignore the “RIS” section of the ToDo List, go to “User Management,” to “Manage cluster users and Administrators” and click “Next.”
18. Add users or click “Next” to bypass the creation of new users.
19. Click “Add an Administrator” and type the path in this format “domain/user” for example: fujitsu\clusteradministrator (The domain in this example is named “fujitsu” and the user created above is named “clusteradministrator”).
20. Click “Next,” click “Finish,” click to close.
21. **Congratulations:** The head node is now completed.

**It is now time to join the domains on all compute nodes.**

**Repeat the following on each node:**

22. Join each compute node to the domain that was created previously. This operation will need to be manually accomplished on each node.<sup>8</sup>

Go to the “Start” menu, right click on “My Computer,” select “Properties,” go to the “Computer Name” tab, click “Change,” select the radio button “Domains,” type a domain name (e.g., fujitsu), click “Ok.”

**Important Note:** Follow the prompts carefully. The prompts require that the credentials be entered for an existing account that has permissions to join machines to the domain.

23. Confirmed that each node in the intended cluster has been properly joined to the domain (e.g., fujitsu).

Insert the CCP installation CD into the CD drive of the first compute node (e.g., fujitsu002).

24. Go to the “Start” menu, to “Run,” to “Browse CD,” to “CCP,” to “Setup.exe.” (Start > Run > Browse CD > CCP > Setup.exe)
25. Select the radio button option to “Accept” the license agreement, and click “Next.”
26. Select the radio button option to “Join this server to an existing compute cluster as a compute node.” Type the name of the head node (e.g., fujitsu001 or fujitsu001head). Click “Next.”
27. Install the .NET framework, click “Next.”
28. Install Microsoft Compute Cluster Pack.
29. Repeat this installation on all compute nodes. (See footnote #8.)

---

<sup>8</sup> A future paper will describe how to “push images” for unattended installations across many nodes.

## **Step 7: Validate the cluster**

Now that CCP has been installed, it is now time to validate the cluster by doing the following:

1. From the head node, go to the “Start” menu, to “Microsoft CCP,” to “Compute Cluster Administrator.”
2. In the “Compute Cluster Administrator” window, a list of the nodes that are “Pending Approval” will be presented. Select all. Right click and select “Approve.”
3. All nodes will now show as “Paused.” Select all again. Right click and select “Resume.”
4. All nodes will now show as “Ready.”
5. Go to the “Start” menu and select “Run Command.” Select all, right click, select “Run a command,” type: ipconfig/all then hit “Start.” A prompt to authenticate credentials will be presented. Enter a valid ID and password, and click “Ok.”
6. The cluster should return a command line output; showing the results of “ipconfig /all” as seen from each of the nodes in the cluster.
7. It is now possible to control all nodes in the cluster from the head node.

## Step 8: Install Myricom drivers

Now that the cluster is running, it is time to install the Myricom drivers.

The drivers for Myri-10G NICs are available from Myricom's [www.myri.com](http://www.myri.com) web site. The MXoE drivers for Windows operating systems are linked from the [www.myri.com/scs](http://www.myri.com/scs) page. Please contact [help@myri.com](mailto:help@myri.com) for a login/password combination.

**Important Note:** Build a CD with all current drivers. Then manually install all drivers on each node.

For MXoE for Windows, the 10G driver installation consists of 3 parts:

1. Install the NDIS driver (mx\_setup.exe -install <full path to mx.inf>)
2. Install the MXoED (MX over Ethernet Daemon) (mxoed\_setup.exe -install <full path to mxoed.inf>)
3. Run Tunesettings

The WSD-MX proxy is also available from [www.myri.com/scs](http://www.myri.com/scs). The installation is performed by:

Install WSD-MX (inst\_wsd -install)

By default, Myricom's network enables jumbo packets with a 9-kiloByte MTU. Please make sure that the switch is configured the same way. Alternatively the configuration panel allows the customization of the packet size to an MTU of 1500 Bytes.

The README file coming with the software package has further details on the configurable options.

Install Windows CCS and its services first. In this case the clusrun command can be effectively used to deploy MX software on all nodes in one step.

After the installation is complete, it is recommended running mx\_info, which provides information about the available hosts for MXoE.

### Performance Results:

The following results were obtained on a cluster of 4 nodes running MXoE 1.2.1 drivers with Myri-10G 10GBase-CX4 Network Interface Cards (NICs) and a Fujitsu XG700 10-Gigabit Ethernet switch, which has 10GBase-CX4 ports. All

nodes were running Windows CCS, Service Pack 1, and the performance results were always conducted including the switch.

The hardware information for the test configuration is:

CPU: AMD Opteron 252, 2.6 GHz

Memory: 1GB RAM

NIC: Myri-10G with MXoE 1.21 in x8 PCI Express (PCIe) slot

Software Stack: NDIS 5.1, Winsock Direct WSD-MX 1.0.2

Switch: Fujitsu XG 700, 9000 MTU, flow control enabled

More information on Myricom drivers, including performance indicators and benchmarks, can be found in **Appendix D: Myricom Driver Performance**.

## **Step 9: Finishing the setup**

Once the Myricom driver installation is complete, the following will be true:

1. The head node and all compute nodes are able to communicate with each other over Gb-E.
2. Myricom 10Gb network adapters are installed, but no IP addresses have been assigned to them.
3. CCS is aware of the Public network only.

Ensure that:

1. The head node and all compute nodes are communicating with each other over Gb-E and 10Gb-E.
2. Myricom 10Gb network adapters are installed, and have valid IP addresses.

It is possible to visit each node and manually enter a valid IP address, subnet information, and DNS name resolution information –however– the preferred method is to take advantage of the CCS Topology Wizard to configure the ICS (Internet Connection Sharing). ICS will manage the Myricom 10Gb adapter IP addresses.

To use the CCS Topology Wizard, do the following:

1. Logon to the head node.
2. Start the Compute Cluster Administrator MMC.
3. Go to the ToDo List.
4. Click on Configure Cluster Network Topology (wizard) under the Networking section.
5. Select "All nodes on public and private networks" in the pull down window.
6. The wizard will ask which adapter should be assigned to the public network. For the public network, choose the Gb-E adapter. For the private network, choose the Myricom 10Gb-E adapter.
7. Enable ICS. There may be an error regarding RRAS being installed. If this happens, simply go into the Services control panel (Start-->Right Click "My Computer"--> Manage --> Services and Applications --> Services, Set "Routing and Remote Access" to "Disabled").
8. Click Next through the wizard to finish up.

It is recommended to wait for approximately 10 minutes, and then repeat the cluster validation test. When ipconfig/all is run against the whole cluster, each node should show at least two valid IP addresses – one from each adapter.

**The compute cluster is now complete.**

## Appendix A: Power and cooling

Category	Caution
Danger of Device Damage	Do not place an XG switch on its side or stack up XG switches (on a table or floor). Always correctly install an XG switch in a quality server rack.
	Do not install an XG switch in an unstable place (such as on a slanted surface or a place that is subjected to vibrations).
	Do not place any objects on top of an XG switch.
	Do not use an XG switch as a working surface.
	Install the XG switch inside a rack inside of a building. Using an XG outside may damage it.
	Do not use an XG in areas of extremely high temperature, low temperature, or an area where the temperature goes up and down suddenly.
	Do not expose the XG to seawater.
	Do not use the XG in a place where chemicals are being sprayed or may otherwise come in contact with it.
	Do not use the XG near objects which generate strong magnetic fields, such as microwave ovens.
	Do not use the XG with foreign objects (liquids and/or pieces of metal) inside it.
	When moving the XG, be sure to remove the power plug from the outlet first.
Danger of electromagnetic interference	Do not use the XG near a radio or a TV. (Doing so will interfere with the radio and TV reception.)
Danger of electric shock	Do not open the cover unless performed by a maintenance engineer. When performing maintenance on the XG, be sure to remove the power plug from the outlet first.
Danger when rack-mounting	Only use the XG if the temperature inside the rack is 40°C or less. Ignoring this may damage the XG.
	Ensure that the rack is sufficiently ventilated and that excess heat is properly exhausted.
	Check that the configuration of devices in the rack does not overload the power supply.
	To ensure the stability of the rack, fix it to the wall or floor as appropriate.
	Do not install the XG in a rack if it would make the rack unstable.
	Check that all units installed in the rack are correctly connected to a grounded power source.
	When removing the XG from a rack, be sure to hold it by both sides. At least two people should work together.
Danger when cleaning	When cleaning the XG, only use a soft cloth and wipe it gently.

## Appendix B: Greater detail on Fujitsu Switches

### Accessing the switch

There are two ways to access the XG series switch to configure it for operation, run commands, and collect statistical data: a serial connection via the front panel RS-232 DB-9 connector; or a 10/100 Ethernet connection via the front panel RJ-45 connector.

The following functions are exclusive to the LAN Management port:

- Transfer firmware image files to the switch
- Transfer startup configuration files to and from the switch
- Transfer diagnostic dumps or error logs from the switch

All these functions are performed with a TFTP client/server connection, a network based application.

### *Regarding the Initial Installation and Configuration*

When an XG switch is initially installed, the default settings of the LAN management port are not set to any particular IP address. The switch must be initially configured with a serial connection before the LAN management port can be used.

An IP address compatible with the client (or management network) must be defined before the LAN management port can be used.

The client side of the serial port connection settings:

Item Setting	Value
Baud rate	9600 bps
Character size	8 bit
Parity	None
Stop bits	1 bit
Flow control	None
Emulation	VT100
Character set	ASCII
Transmission:	CR (carriage return) only
Reception:	LF is added

After communications has been established, the baud rate can be changed to either 9600, 19200, 38400 and 57600 (bps) using the "baud-rate" command.

Invoke and configure the client system terminal emulation program (Hyperterminal) for the serial port settings shown in the table.

Connect the client to the XG switch using a crossover cable (which was supplied with switch). Power on the switch

Hit the enter key until the xg login prompt is displayed. As a default out of the box, the username is “admin” and the password is “password.”

If more information is needed than is supplied in this paper, please refer to the User Guide Chapter 3, Section 3.1 for LAN Management configuration instructions. Also refer to Chapter 4 for Feature and Function configuration commands. More details on the commands can be found in Chapter 5.

**Special Note:** If the switch being installed has been used previously, then it is advised that the switch be restored to the factory default settings before proceeding. This procedure is described in 1.3 to reset the switch the factory default settings.

If the switch being installed is new out of the box, then a factory reset is not needed. In this case, proceed directly to 1.4 to configure the Simple Network Management Protocol (SNMP).

### **Restore factory defaults on the switch**

**Special note:** This step is only necessary if the switch has been used previously and is not new out of the box. If the switch is new out of the box, then proceed to the instructions for setting up the Management LAN / Ethernet port and switch.

1. If not already done, install the switch into a rack. Connect the notebook or KVM to the switch via a serial cable.
2. Open a terminal window or create a telnet session and login to the switch console port and execute the following commands:

```
xg> enable
```

```
xg# reset factory-default
```

The following prompt will be displayed:

```
“Do you want to restart system to factory-default? (y/n) ”
```

If "y" or "Y" is entered, the contents of the startup-configuration are reset to the factory defaults and the system will restart.

To cancel the process, respond to this message with any keys other than "y "or "Y."

3. The switch boot process will be displayed. When completed, the login prompt will appear.
4. Login to the Console port and proceed with the port and switch configuration as detailed below.

### **Configure SNMP**

1. Open a terminal window or create a telnet session and login to the Console or Management LAN port.
2. Enter the configuration mode by executing the "enable" and "configure terminal" commands:

```
xg> enable  
xg# configure terminal
```

3. Configure for SNMP server access using the following command.

```
xg(config)# snmp-server access host 192.168.1.10 community xgpublic
```

**Important Note:** The network's actual SNMP server IP address and community name must be provided. In the example given "192.168.1.10" is the SNMP server IP address and "xgpublic" the community name.

4. The switch location and administrator contact information can be inserted in the SNMP messaging by invoking the following two commands. The location and contact information can be any text string, including white/blank spaces, up to 255 characters in length. In the example given "3F West" and "administrator tel:012-3456-7890" are the respective location and contact text strings.

```
xg(config)# snmp-server location 3F West
```

```
xg(config)# snmp-server contact administrator tel:012-3456-7890
```

5. Execute an "exit" command to leave the configuration mode. The prompt will change to xg#.
6. Check the configuration by executing the "show snmp-server" server command.

```
xg# show snmp-server
```

7. Verify that the host IP address, community name, and optional location, and contact are correct.
8. Save the configuration with the "copy" command.

```
xg# copy running-config startup-config
```

9. If any additional SNMP functions (SNMP traps) or RMON extensions are desired, refer to Section 5.16 and 5.17 of the XG series User Guide.

### Configuring the port and switch simultaneously

It is now time to setup Gb-E network management using SNMP. If the intention is to move forward with the option to configure both the Ethernet port and switch at the same time via the serial port, do the following:

1. Connect a serial cable from the notebook computer or KVM to the switch that is to be configured.
2. Create a telnet or terminal session and paste in the following:

Begin copying below:

```
# =====
# Switch Configuration for Microsoft CCS
# This script must be edited before use.
# Insert:
# 1. NameOfSwitch
# 2. IP address and subnet mask for management port
# 3. IP address of gateway (Optional, but, if SNMP
#    server on other side of router, use router IP address)
# 4. Change port range according to switch used
#
# note: command sequence after login
# login: admin
# password: password
#
enable
configure terminal

# insert name of switch in next two command lines
hostname NameOfSwitch
```

```
banner login Welcome to NameOfSwitch

management-lan ip 172.25.100.11/24 default-gw 172.25.100.1

telnet-server

# configure telnet session timeout

terminal timeout vty 25

# configure serial console timeout

terminal timeout console 25

# size console window if telnet session host fails to set values

line console

terminal window 80 24

# this completes management port configuration
# switch configuration follows

bridge forward-mode cut-through

bridge jumbo-frame 9216

# change port range 1 12

interface port range 1 12

flowcontrol send-receive

# exit to config mode ("xg(config)>" prompt)

exit

# this completes the switch configuration

# save settings for next startup

copy running-config startup-config

exit
```

```
# exiting to operator EXEC mode ("xg>" prompt)
# use the reset command or power-cycle the switch
# follow instructions to check configuration
```

*End copying text above.*

3. **Before hitting RETURN!** Modify the text. Locate the 24th line of the text, identify the "management-lan ip" line where there is a sample IP address and correlating sample gateway IP address.
4. Replace those sample addresses with an actual static IP address for the switch along with the correct corollary IP address for the gateway.
5. **Protocol advisory:** Write down the new static IP and gateway addresses. Keep the information along with other critical administrator information (such as serial numbers, logins and passwords) in the network administrator's handbook. If a dedicated handbook does not yet exist in a secure location (such as the same place where the routine backups are stored off site in a fire proof safe) then it is strongly recommended that one be created immediately along with a best practices protocol and routine.
6. **Special note:** If at any point the switch is locked or the IP address is forgotten, use the serial port to gain access.
7. **Congratulations:** The Ethernet / Management LAN port and the switch have been successfully configured.

### **Configuring the Ethernet / Management LAN port ONLY**

It is now time to setup Gb-E network management using SNMP. If the intention is to move forward with the option to configure the LAN Management (aka Ethernet) port first AND THEN configure the switch via the newly configured LAN Management port, then do the following:

1. Using a notebook or existing server, connect a serial cable from the notebook or server to the switch to be configured.
2. Create a terminal session and paste in the following text:

Begin copying below:

```
# =====
# Switch Configuration for Microsoft CCS
# This script must be edited before use.
# Insert:
```

```
# 1. NameOfSwitch
# 2. IP address and subnet mask for management port
# 3. IP address of gateway (Optional, but, if SNMP
# server on other side of router, use router IP address)
# 4. Change port range according to switch used
#
# note: command sequence after login
# login: admin
# password: password
#

enable
configure terminal

# insert name of switch in next two command lines
hostname NameOfSwitch

banner login Welcome to NameOfSwitch

management-lan ip 172.25.100.11/24 default-gw 172.25.100.1

telnet-server

# configure telnet session timeout

terminal timeout vty 25

# configure serial console timeout

terminal timeout console 25

# size console window if telnet session host fails to set values

line console

terminal window 80 24

End copying above.
```

3. **Before hitting RETURN!** Modify the text. Locate the 24th line of the text, identify the “management-lan ip” line where there is a sample IP address and correlating sample gateway IP address.
4. Replace those sample addresses with an actual static IP address for the switch along with the correct corollary IP address for the gateway.

5. **Protocol advisory:** Write down the new static IP and gateway addresses. Keep the information along with other critical administrator information (such as serial numbers, logins and passwords) in the network administrator's handbook. If a dedicated handbook does not yet exist in a secure location (such as the same place where the routine backups are stored off site in a fire proof safe) then it is strongly recommended that one be created immediately along with a best practices protocol and routine.
6. **Special note:** If at any point the switch is locked or the IP address is forgotten, use the serial port to gain access.
7. **Congratulations:** The Ethernet / Management LAN port has now been configured.

### **Configuring the switch using the Ethernet / Management LAN port**

Once the Ethernet / Management LAN port has been configured, the switch can be configured via the Ethernet / Management LAN port.

1. Disconnect the serial cable from the switch and notebook or KVM. Connect an Ethernet cable to the notebook or KVM and the Ethernet port switch.
2. Open a telnet session or terminal window and copy and paste the following text:

Begin copying below:

```
# this completes management port configuration  
# switch configuration follows
```

```
bridge forward-mode cut-through
```

```
bridge jumbo-frame 9216
```

```
# change port range 1 12
```

```
interface port range 1 12
```

```
flowcontrol send-receive
```

```
# exit to config mode ("xg(config)>" prompt)
```

```
exit
```

```
# this completes the switch configuration  
  
# save settings for next startup  
  
copy running-config startup-config  
  
exit  
  
# exiting to operator EXEC mode ("xg>" prompt)  
# use the reset command or power-cycle the switch  
# follow instructions to check configuration  
  
End copying above.
```

3. **Congratulations:** The switch has now been configured!

## Appendix C: Fujitsu Server Setup in Greater Detail

### ***Initiate the boot sequence***

In order to initiate the configuration phase and OS-installation, the server system has to be booted from the ServerStart CD. In some circumstances it will be necessary to adjust several settings to do this:

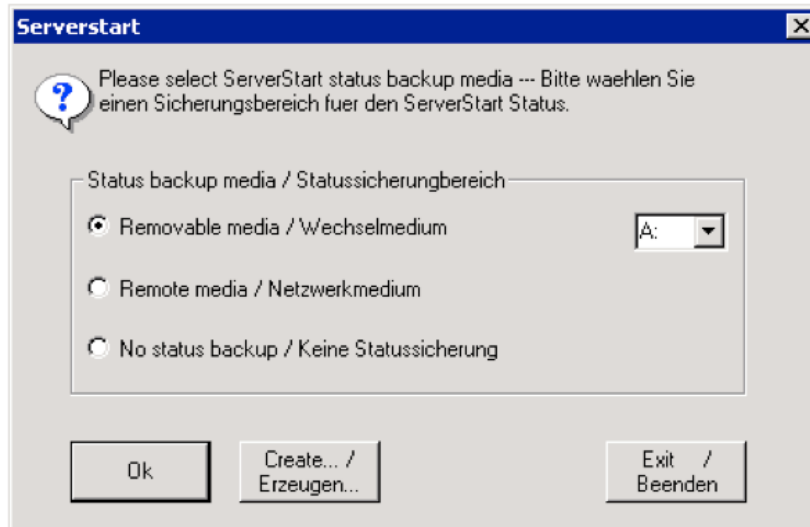
1. Insert the ServerStart CD into the CD-ROM drive and press the power button on the front panel of the server blade. Wait a few seconds until the monitor is activated.
2. When prompted, press the [F2] key to enter the server BIOS Setup.
3. In the Phoenix BIOS Setup Utility, select the Boot submenu and activate the Boot from CD-ROM Drive option.
4. Press the [ESC] key and choose Exit Saving Changes from the Exit submenu.

### **Booting the target system from the ServerStart CD System startup**

5. Ensure that the CD-ROM drive is the first boot device to be accessed when the system is started. Further information on setting up the boot properties can be found in section .Setting up the CD-ROM drive as boot device (Windows only).
6. Power up the server and insert the ServerStart CD in the drive (for BX server systems, see section .Initiate boot sequence.)
7. Press the Reset switch on the front of the device.

**Special Note:** If the PRIMERGY server does not have a Reset switch, power the server down and then up again. ServerStart is now started on the target system from the CD.

8. Choose whether to save the configuration data on a local removable medium or on a remote medium on the network.



9. Choose a status backup medium

**Please note:** If the *No status backup* option is chosen, all configuration data will be lost after a reboot.

10. Click the “Ok button.

#### ***Local removable medium (floppy disk, USB pen) as backup medium***

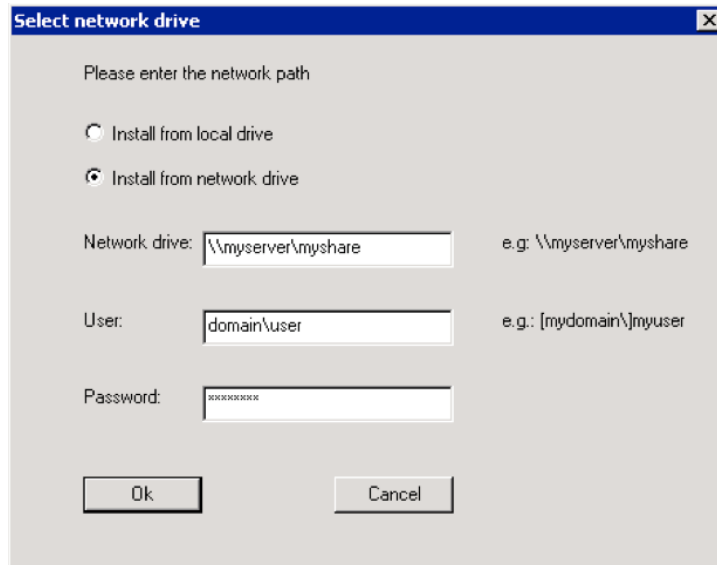
11. Insert a floppy disk or connect a USB pen.

**Special note:** The backup medium must not be write-protected.

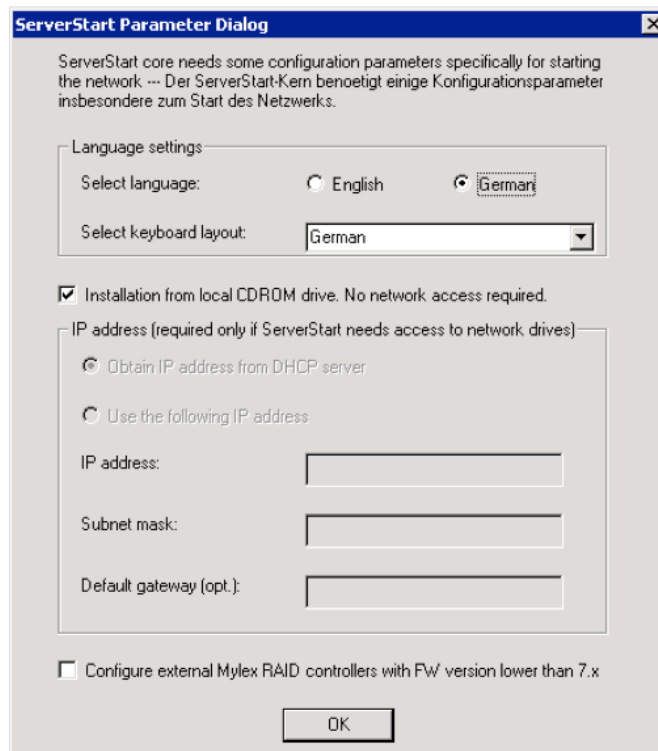
#### ***Remote medium as backup medium***

**Special note:** Instead of a local removable medium (floppy disk or USB pen) it is possible to specify a remote medium.

- To do this, define the required share settings and select the Remote Media option. When the “Ok” button is clicked, an additional window is opened in which a network drive can be selected:



- In the subsequent dialog box, enter the preferred language for both the user interface and keyboard:



14. Enter the location of the source data for the operating system, service packs and additional applications. If the installation is to be performed from a remote medium, enter the IP address and subnet mask. Alternatively, have the IP address assigned via DHCP.
15. It is possible to configure an external LSI/Mylex controller with a firmware version lower than 7.x, to do so select the check box Configure external Mylex RAID Controllers.

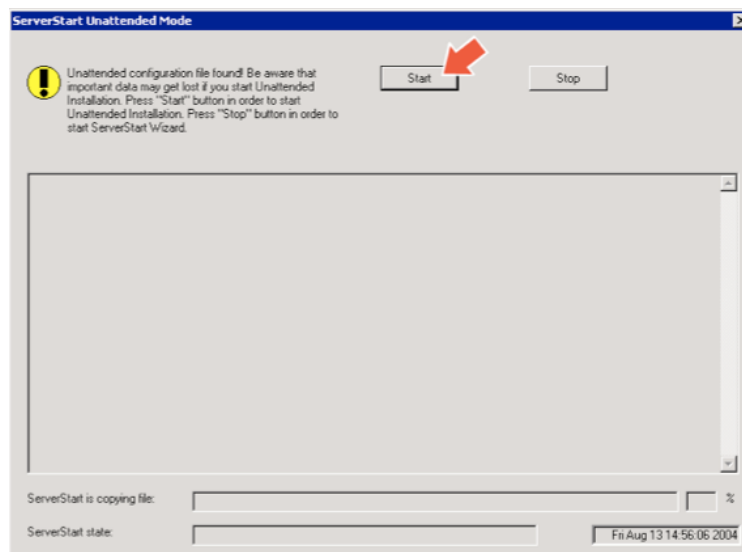
**Special Note:** *Explanation:*

Choose between LSI/Mylex controllers FW6 and FW7. The two variants require different drivers:

In the case of internal LSI/Mylex controllers, ServerStart automatically determines the drivers that have to be installed for RAID configuration during the ServerStart session. In the case of the external Mylex controllers with firmware versions 6 and 7, ServerStart is not able to automatically determine the drivers that have to be loaded for RAID configuration. By default, version 7.x drivers are loaded. However, to load the FW6 drivers it will be necessary to check the corresponding box.

16. Click OK.

**Special Note:** If the inserted removable medium already contains a valid *SerStart-Batch.ini* configuration file then immediately initiate an unattended installation of the operating system if desired:



17. Click “Stop” to continue configuration and replace the existing configuration file.

18. Click “Start” to initiate the installation process.

The target system hardware is now analyzed. The detected configuration data is then used as the basis for the configuration. All the steps detailed below are carried out automatically.

**Special Note:** In the following cases, configuration mode is started instead of unattended installation:

19. The server floppy disk does not contain a valid configuration file.

20. The user halts unattended installation.

In these cases, perform the installation in the configuration mode as described below.



Click [here](#) to prepare an **operating system installation** for a PRIMERGY Server

Further Functions:



Tools



Installation of Drivers



Information



Applications



ServerManagement Software



FloppyBuilder



Firmware



Exit ServerStart

## Initiating the configuration phase

21. Choose the desired menu item:

- For configuring in Guided Mode:

*Click here to prepare an operating system installation for a PRIMERGY Server - MS Windows operating systems - <preferred Windows version> - Prepare and initiate an unattended installation.*

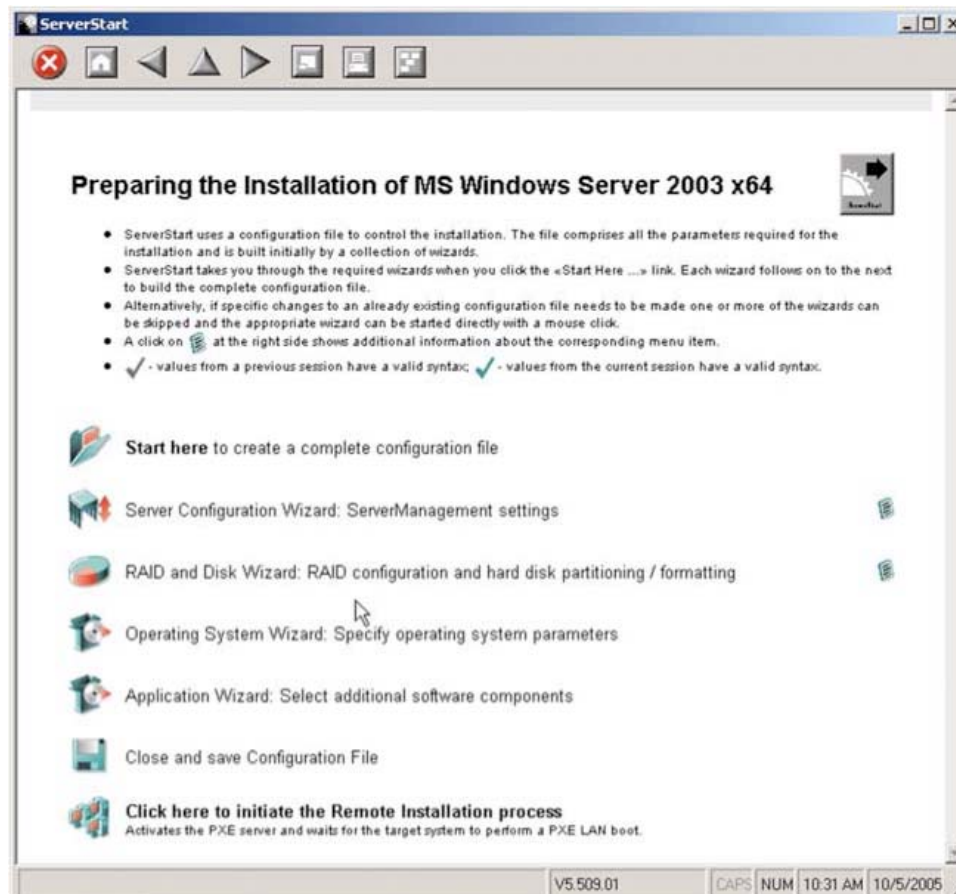
- For configuring in *Expert Mode*:

**Special Note:** This configuration mode requires expert knowledge.

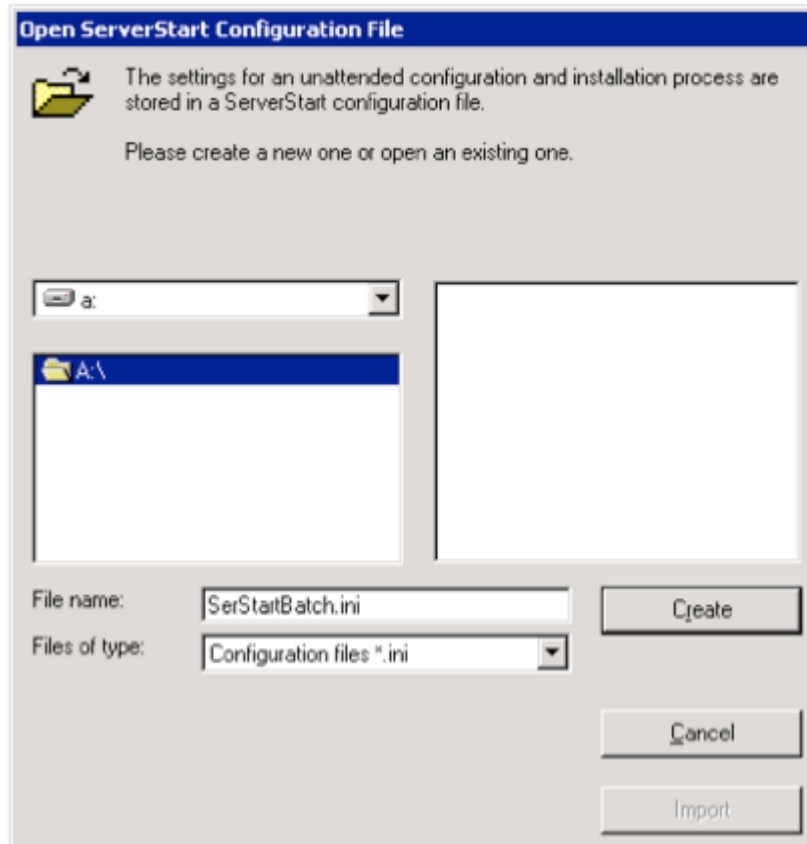
*Prepare and/or initiate operating system installation - MS Windows operating systems - <preferred Windows version> - Install MS Windows interactively*

## Opening/creating a configuration file

22. Click “Start here to create a complete configuration file.”



## Opening/creating a configuration file



23. Confirm the preset file name SerStartBatch.ini and click Create followed by Continue or select and edit an existing configuration file.

**Special Note:** The removable medium must not be write-protected.

### Local configuration of a single server in Guided Mode (Windows) on the target system

**Special Note:** To achieve the greatest possible level of security in the detection of the existing hardware components, use ServerStart in Guided Mode to perform configuration on the target system for installation.

In *Guided Mode*, wizards perform the configuration of the server hardware and hard disk arrays. The installation phase is initiated without any need for a restart once the configuration has been defined. To this end, it is necessary to have the media containing the operating system, service packs and additional applications.

**Special Note:** Any missing specifications are loaded as default settings from the Default Configuration File. This configuration file is located on the ServerStart CD and can be processed by the administrator.

## Selecting a server system

### ***Server detection successful***

The PRIMERGY server system is automatically detected at the start of the configuration operation.

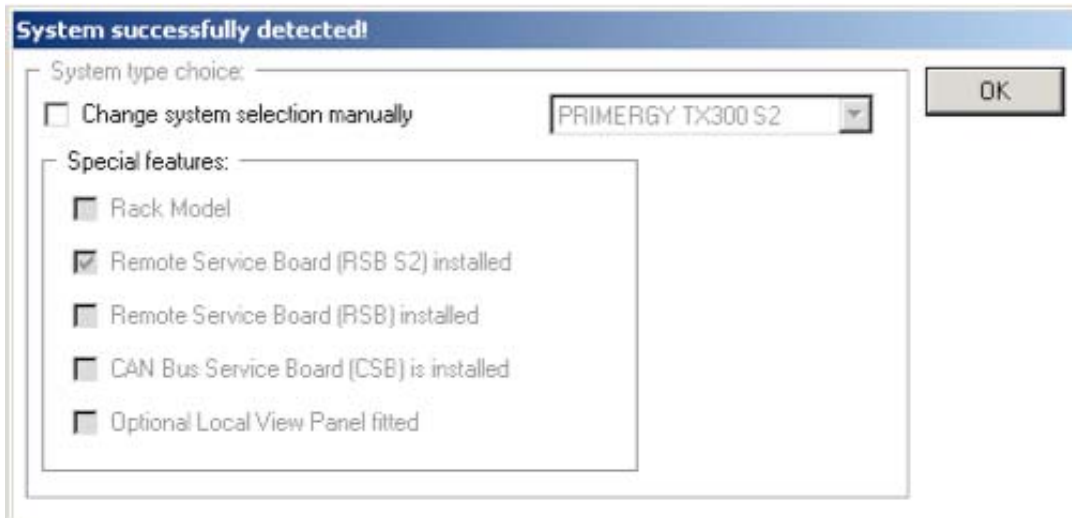


Figure: Server detection successful (Guided Mode)

24. Click the OK button to confirm the detected server model.

Proceed as follows to include a not yet installed service board or LocalView panel in the configuration:

25. Select the Change system selection manually check box to correct the preselection.

26. Add the desired components by selecting the corresponding check boxes.

### Server detection not successful

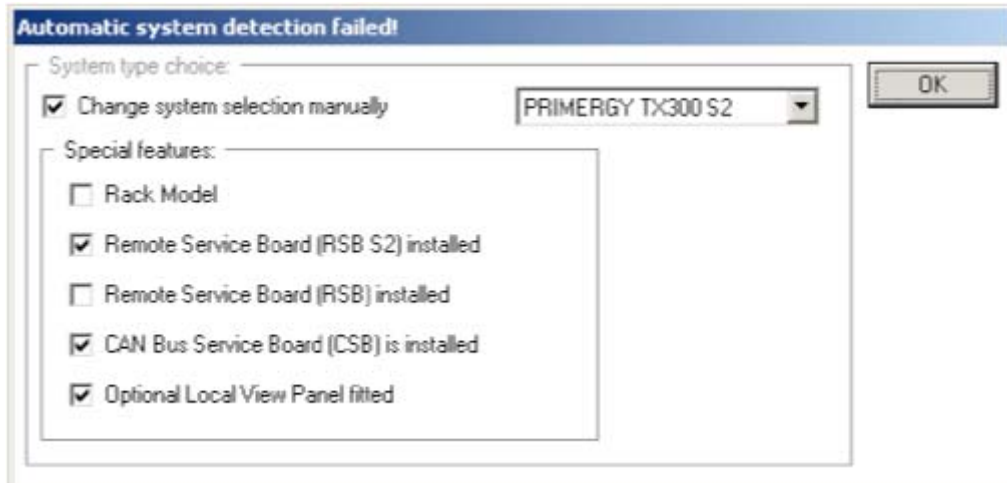


Figure: Server detection unsuccessful (Guided Mode)

Proceed as follows if automatic server detection does not return any result:

27. Select the PRIMERGY server model from the drop-down list.
28. Check the corresponding box if the inclusion of an installed service module or a LocalView panel in the configuration is desired.
29. Click “Ok.”

**Special Note:** If the server model is not present in the drop-down list then use a more recent ServerStart version for the installation. The table below indicates the required version: Supported PRIMERGY models.

<b>ServerStart</b>	<b>Supported PRIMERGY models</b>
Version 6.606	PRIMERGY Econel 30, Econel 40, Econel 50, Econel 50 Edition, Econel 100, Econel 200
	PRIMERGY C150
	PRIMERGY H250, H450
	PRIMERGY L250
	PRIMERGY P250
	PRIMERGY R450
	PRIMERGY RX100, RX100 S2, RX100 S3
	PRIMERGY RX200, RX200 S2, RX200 S3
	PRIMERGY RX220
	PRIMERGY RX300, RX300S2, RX300 S3
	PRIMERGY RX600, RX600 S2
	PRIMERGY RX800
	PRIMERGY T850
	PRIMERGY TX150 (incl. SATA), TX150 S2, TX150 S3, TX150 S4
	PRIMERGY TX200, TX200 S2, TX200 S3
	PRIMERGY TX300, TX300 S2
	PRIMERGY TX600, TX600 S2
	PRIMERGY S60, S80
	PRIMERGY BX300 Pentium III, BX300 Pentium M
	PRIMERGY BX620, BX660
PRIMERGY BX620 S2, BX620 S3	
PRIMERGY BX630	

**Configuration wizards**

The configuration wizards will now support the specification of installation parameters for server management, RAID environment, hard disks, operating system and additional applications.

30. Enter the required specifications in each of the wizards. The Next button will be the next input mask. Once one configuration section is complete, the next wizard is started.

**Special Note:** For detailed descriptions of the input masks see the ServerStart online help.

## **RAID and Disk wizard: Defining the RAID parameters, Partitioning and formatting the hard disks**

This wizard allows the specification of the manufacturers and types of RAID controllers, the desired RAID level and the number of hard disks available. In this menu, it is also possible to define partitions for the hard disk drives of the PRIMERGY server system plus the associated file systems. An overview of the hard disks installed in the system and their partitions is provided.

## **Installation wizard for MS Windows: Defining the system parameters**

In this menu, define regional settings for the operating system to be installed.

### *Windows Installation*

In this menu, specify country settings.

### *Computer Identification*

In this menu, identify the server to be installed in the network.

### *Installation Directory and Time Zone*

In this menu, define the desired time zone and the directory in which the operating system is to be installed.

### *User Name*

In this menu, specify the user who will normally work on the server system to be installed.

### *Display Settings*

In this menu, define the resolution, refresh rate and color depth in which the Windows user interface is to be displayed the first time it is started.

### *Network Protocol*

In this menu, specify the network protocols to be used.

### *Software Components*

In this menu, select the Windows components to be preinstalled by ServerStart in addition to the operating system itself.

### *Services*

In this menu, select the services to be preinstalled by ServerStart in addition to the operating system itself.

### *Additional Properties*

In this menu, make settings for error logs, remote maintenance and system recovery.

**Special Note:** These input masks are only available for configuring Windows 2003.

### Application Wizard

In this menu, select additional software components to be preinstalled by ServerStart in addition to the operating system. Also define customer-specific scripts that will be executed after the operating system installation is complete. For more information see section “Customer-specific scripts following a Windows installation.”

**Special Note:** If the intention is to manage the server system or the server blade via RemoteView, then Fujitsu Siemens ServerView SNMP agents must be installed. These and the ServerView and GlobalFlash packages are contained on the PRIMERGY ServerView Suite Software CD. For a description of how to install the server management software, see the ServerView installation guide on the ServerBooks CD.

### Completing the configuration

When all the ServerStart wizards are finished, then the configuration phase is complete.

### Saving the configuration file

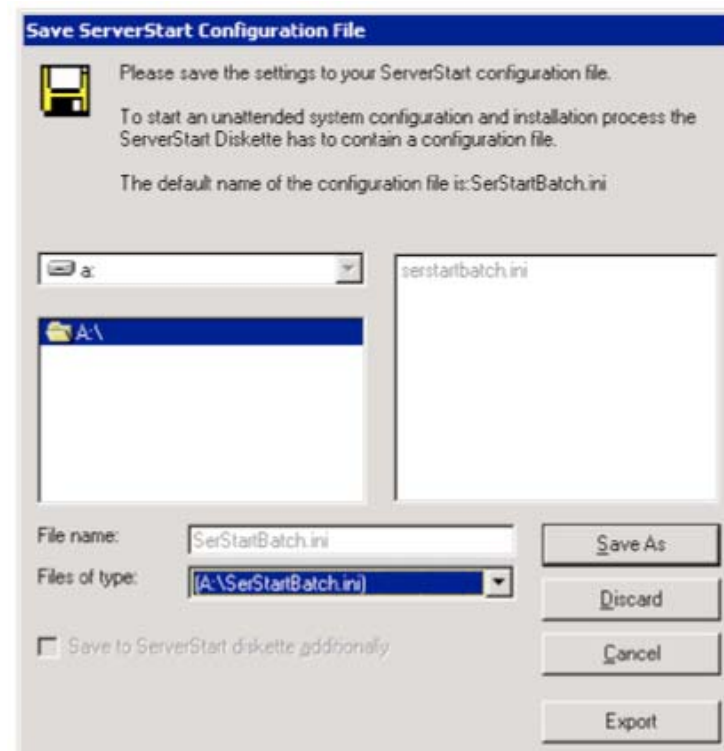


Figure: Saving the configuration file

31. Once all of the required steps are completed, save the configuration file under the suggested name `SerStartBatch.ini` on the inserted removable medium or on a network drive.

32. Continue according to one of the following sections:

*For immediate local server installation*

see section “Local installation following configuration on the target system”

*For local installation of one or more server systems with an identical hardware configuration (replication)*

see section “Local installation following configuration on a Windows system (Replication)”

## **Remote OS installation using ServerStart**

### **Local installation following configuration on the target system**

**Special Note:** The preceding configuration phase is described in section “Local configuration of a single server in Guided Mode (Windows) on the target system” (Guided Mode) and section “Local installation of a single server in Expert Mode (Windows)” (Expert Mode).

33. Leave all inserted media unchanged and, in the ServerStart user interface, choose the menu item Start Installation:

The installation process is started according to the parameters in the configuration file. It runs unattended, including all necessary restarts. User interference is only necessary to change data carriers (operating system, service packs, applications) or in case of incorrect or incomplete hardware recognition.

**Special Note:** Exception: Necessary information that was not specified in the configuration phase, will be requested during the installation.

During the installation of the operating system, ServerStart automatically integrates drivers for system components that are not contained in the operating system. During the installation of other operating systems, ServerStart reports missing drivers.

Make sure that all installation steps called up run correctly and are closed without errors. To repeat a faulty operating system installation, reset the *ServerStart* status configuration using the taskbar.

## Appendix D: Myricom Driver Performance

### NTTTC

The tcp benchmark is a well-known benchmark for Sockets. Originally developed for different versions of UNIX, Microsoft provided an improved version that reflects the capabilities of the Winsock2 Architecture. For this purpose, overlapping send and recv allows for a more efficient way of pipelining messages. The NTTTC sender and receiver are included in the Myri-10G driver package and are good indicators whether the infrastructure has been set up correctly.

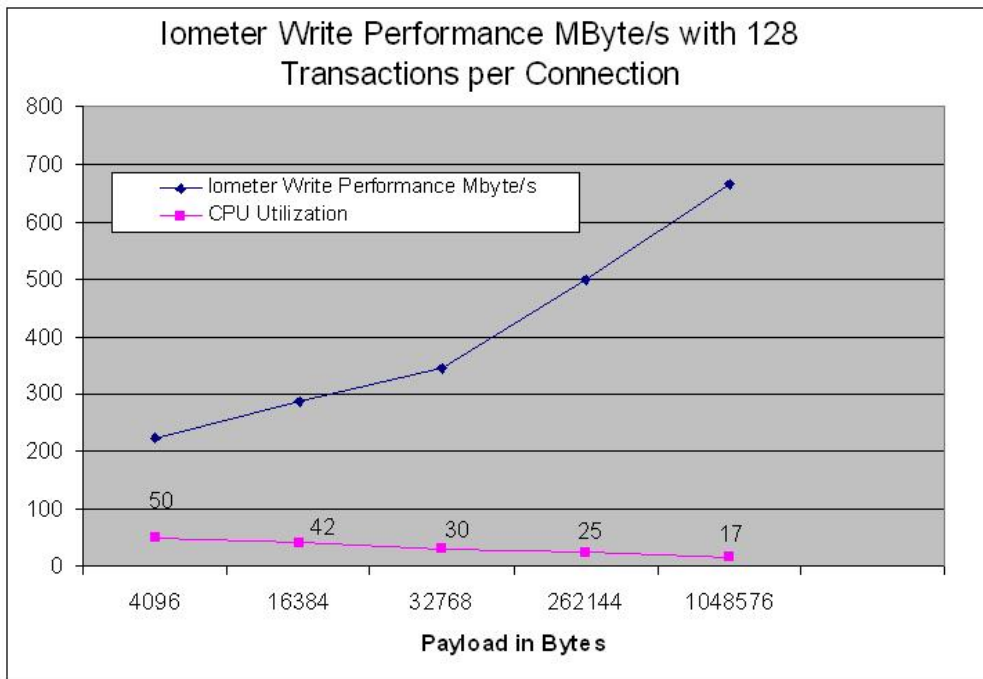
When running NTTTC the performance should be close to the following:

Windows 2003

9000B MTU, TSO:	9.6Gb/s, 11% CPU utilization
1500B MTU, TSO:	8.7Gb/s, 10% CPU utilization

### lometer

lometer is a well-known tool measuring disk and network I/O. The following graph depicts the results of running a CIFS test between two nodes. The test specified 128 transactions for the payloads ranging from 4096 Bytes to 1048576 Bytes.



## mx\_pingpong

mx\_pingpong is a low level benchmark that reports the latency for a given message size as the round-trip time divided by 2 (RT/2). Its performance metrics using MX are the base for other middleware such as MPICH-MX.

When running mx\_pingpong between two nodes connected through the switch, the latency for a 0-Byte message is reported as 3.983µs. A direct back-to-back setup reports 3.51µs.

```
mx_pingpong -S 0 -E 4194305 -M 2 -d SNOW1:0
```

```
Starting pingpong send to host SNOW1:0
```

```
Running 1000 iterations.
```

Length	Latency(us)	Bandwidth(MB/s)
0	3.983	0.000
1	4.096	0.244
2	4.077	0.491
4	4.111	0.973
8	4.123	1.940
16	4.130	3.874
32	4.133	7.743
64	4.800	13.335
128	5.079	25.204
256	6.686	38.292
512	7.452	68.706
1024	9.245	110.757
2048	10.602	193.171
4096	14.917	274.586
8192	19.177	427.178
16384	27.098	604.620
32768	41.163	796.064
65536	96.778	677.175
131072	150.814	869.097
262144	256.161	1023.358
524288	468.010	1120.248
1048576	894.658	1172.041
2097152	1745.842	1201.227
4194304	3669.813	1142.920

It is impressive that the ~1200 MByte/s (9.6 Gigabit/s) asymptotic data rate for large user-level messages closely approaches the 10 Gigabit/s channel rate.

## Intel MPI Benchmark

The Intel MPI Benchmarks (IMB) are a set of typical MPI operations, and measure PingPong, PingPing, and SendRecv latency and data rate along with

those of Exchange and other collective operations. In the following report are the PingPong performance results. While different MPI versions are available for Windows, Windows CCS comes with MSMPI, which uses the Sockets interface for communication.

Starting with Windows 2000 Advanced Server SP2 and Windows Server 2003 a Winsock Direct Switch is available to transparently map Socket function calls to networks with system-area-network properties. Myricom provides a proxy to the Winsock Direct Switch, and the following results are reported for WSD-MX:

```
#-----
# Benchmarking PingPong
# ( #processes = 2 )
#-----
```

#bytes	#repetitions	t[usec]	Mbytes/sec
0	1000	13.82	0.00
1	1000	14.09	0.06
2	1000	14.78	0.13
4	1000	14.84	0.26
8	1000	14.72	0.52
16	1000	14.79	1.03
32	1000	14.87	2.05
64	1000	16.42	3.72
128	1000	16.72	7.30
256	1000	17.00	14.36
512	1000	17.85	27.35
1024	1000	19.82	49.28
2048	1000	22.03	88.66
4096	1000	28.42	137.44
8192	1000	34.98	223.33
16384	1000	54.19	288.35
32768	1000	100.16	312.00
65536	640	161.82	386.22
131072	320	237.87	525.49
262144	160	403.00	620.34
524288	80	624.68	800.41
1048576	40	1057.07	946.01
2097152	20	1954.48	1023.29
4194304	10	3974.00	1006.54

```
#=====
```

## Low-Latency 10-Gigabit Ethernet with Myri-10G:

Myricom supplies optional message-passing software for Myri-10G network-interface cards (NICs) to deliver low latency and low host-CPU utilization over 10-Gigabit Ethernet.

Myricom extended its Myrinet Express (MX) software, already widely used in High-Performance Computing (HPC) clusters interconnected with Myrinet, to work also over 10-Gigabit Ethernet. “MX over Ethernet” operates by kernel bypass with Myricom’s dual-protocol Myri-10G network-interface cards and standard 10-Gigabit Ethernet switches to achieve latencies 5 to 10 times lower than with TCP/IP over Ethernet. The MX over Ethernet (MXoE) protocols are open.

**How it Works.** Myricom’s Myri-10G solutions introduced a convergence at 10-Gigabit/s data rates of Myrinet, the most successful specialty network for HPC applications, and mainstream Ethernet. Dual-protocol Myri-10G NICs initially achieved optimal performance running MX software with Myrinet network protocols through Myri-10G switches. MX’s kernel-bypass techniques achieve low latency and low host-CPU utilization by allowing application programs to communicate directly with firmware in the programmable Myri-10G NICs. Now, the availability of MXoE extends MX’s advantages to standard 10-Gigabit Ethernet switching. OEMs and cluster integrators can achieve HPC performance with mainstream Ethernet technology.

MXoE uses 10-Gigabit Ethernet as a layer-2 network with an MX EtherType to identify MX frames (packets). The EtherType identifies the protocol of an Ethernet frame and is defined for the Internet Protocol (IP), Address Resolution Protocol (ARP), AppleTalk, and many other protocols. All of these protocols can be carried concurrently on the same Ethernet network. Myri-10G NICs carry TCP/IP and other traffic along with MX traffic, but achieve the best performance by circumventing TCP/IP. MX provides its own, highly efficient, reliability layer.

MXoE is ‘plug-and-play’ with any 10-Gigabit Ethernet switch, although the best performance is with low-latency switches such as the Fujitsu XG700 or XG2000. In the Performance section of this article are results of several benchmarks conducted by measuring TCP/IP together with MX performance carried over Ethernet.

In addition to low latency, MX exhibits host-CPU utilization that is dramatically lower than the typical TCP/IP utilization and service demand reported in standard benchmarks such as NTTTCP. For example, the host-CPU utilization for MPI communication for MXoE can be as low as 1 $\mu$ s of host-CPU time at the sender or receiver to transfer messages up to 2 Kbytes. Even applications that are not sensitive to latency can benefit from MXoE due to the savings in host-CPU load.

These MX/Ethernet results show that for small clusters 10-Gigabit Ethernet is capable of performance formerly associated only with specialty cluster interconnects. Due to performance losses in building larger networks by connecting multiple Ethernet switches (a limitation of the Ethernet spanning tree computed by Ethernet switches), these solutions will be limited to smaller clusters that can be served with a single 10-Gigabit Ethernet switch. However, 10-Gigabit Ethernet will be used for larger clusters as 10-Gigabit Ethernet switch technology advances.

This MXoE innovation provides strong new evidence that 10-Gigabit Ethernet is a good choice to interconnect the small HPC clusters.

## Appendix E: Glossary of Acronyms & Terms

<b>x64</b>	
<b>64 bit</b>	Refers to the microprocessor's architecture. It is essential that the hardware and software for this solution be 64 bit or x64 and not the older 32 bit versions of the hardware and software.
<b>.exe</b>	The file extension for an "executable" application as opposed to a file that is used by an application (such as a .doc for a Word document, .xls for an Excel file, .dll for a system resource, etc.).
<b>AD</b>	
<b>ADS</b>	"Active Directory Service," commonly called "Active Directory." AD or ADS allows an administrator to assign policies and deploy programs.
<b>BIOS</b>	The "Basic Input Output System" controls the behavior of the data that comes in and out of the computer. The BIOS is the first software to run when the computer is powered-on.
<b>Cluster</b>	A group of computers and/or servers that work together.
<b>CCS</b>	"Compute Cluster Solution" is software by Microsoft that joins many servers together to work as one.
<b>CCP</b>	"Compute Cluster Pack" is a version of CCS that is designed to be installed on top of an existing license of Windows Server 2003 x64.
<b>CCE</b>	"Compute Cluster Edition" is a version of CCS that includes a restricted version of Windows Server 2003.
<b>CNTL/Z</b>	When the "Control" key is held down and then the letter "Z" key is pressed.
<b>CPU</b>	The "Central Processing Unit" is the neural core of a computer.
<b>CX4</b>	Cabling that is copper-based and designed to work with 10 Gb-E solutions.
<b>Driver</b>	A software component that is used to interact with a hardware device.
<b>Gb-E</b>	"Gigabit Ethernet." This paper addresses both 1 Gb-E and 10 Gb-E switches. The 1 Gb-E switch is used to connect the cluster to the internet and network, while the 10 Gb-E switch is essential for the cluster itself to work.

<b>HPC</b>	“High Performance Computing.”
<b>ICS</b>	“Internet Connection Sharing” means that many computers can share a single connection.
<b>ID</b>	“Identification.”
<b>KVM</b>	“Keyboard Video Mouse” refers to the notebook looking device that is part of the rack where the servers and switch are installed.
<b>LAN</b>	“Local Area Network.”
<b>LED</b>	“Light Emitting Diode.”
<b>MPI</b>	
<b>Benchmark</b>	“Message Passing Interface” is an independent standard for message passing on parallel machines.
<b>MTU</b>	“Maximum Transmission Unit.”
<b>MX</b>	A message passing system.
<b>MXoE</b>	Identifies MX packets (frames) using Ethernet as a layer-2 network with an MX EtherType.
<b>NIC</b>	“Network Interface Card.”
<b>NTTTC</b>	Sometimes written as NTttcp, NTTCP is a LAN testing utility.
<b>Node</b>	A computer or server that has been specially configured to perform functions as well as receive and/or relay data.
<b>Head node</b>	A server that controls a cluster of servers.
<b>Compute node</b>	A server that is part of a cluster.
<b>Partition</b>	A division on a hard drive. A partition on a hard drive can be likened to the separation of one room into two. The size of each room can be specified.
<b>Ping</b>	A query that is sent as a test that sends data in the form of a packet.
<b>PXE</b>	“Pre-boot eXecution Environment.” Pronounced “pixie” and sometimes referred to as “pixie boot.”

<b>Protocol</b>	In the case of hardware and software, protocols are the rules that govern communication between devices.
<b>Rack</b>	A large metal frame with slots for servers.
<b>RAID</b>	“Redundant Array of Independent Devices” is a data storage and management scheme for multiple drives.
<b>RRAS</b>	“Routing and Remote Access.”
<b>RX and TX</b>	Receive and Transmit.
<b>SNMP</b>	“Simple Network Management Protocol.”
<b>SQL</b>	“Structured Query Language.” Pronounced “sequel” or S-Q-L, it is a computer language that is used to create, update, send and receive data in a database.
<b>Server</b>	A computer that is designated to receive and deliver data.
<b>Switch</b>	A hardware device that controls communications between devices.
<b>Terminal</b>	A console, that is a text entry and display device.
<b>Telnet</b>	Originally the “teletype network” telnet simply refers to a communications protocol that is used on the Internet or on a local area network.
<b>TFTP</b>	“Trivial File Transfer Protocol” uses a very small amount of memory to transfer very small amounts of data.
<b>TSO</b>	“Time Sharing Option” is an option of the MVS operating system that provides interactive time sharing from remote terminals.
<b>“U”</b>	A “U” is a unit. In a server rack, each slot where a server can be installed is called a “U.”
<b>USB flash drive</b>	A small removable data storage device that is sometimes called a “thumb drive” or “pen drive”. A USB flash drive is a memory stick that plugs into a “Universal Serial Bus” port on a computer or server
<b>WSD</b>	“Winsock Direct” operates over MX; thus WSD-MX.
<b>XG</b>	XG is the name of the switch referred to in this paper.

## **Appendix F: Other Supportive info & Links**

A definition of High Performance Computing

[http://en.wikipedia.org/wiki/High\\_performance\\_computing](http://en.wikipedia.org/wiki/High_performance_computing)

An overview of High Performance Computing (HPC) from Microsoft

[http://download.microsoft.com/download/9/e/d/9edcdeab-f1fb-4670-8914-c08c5c6f22a5/HPC\\_Overview.doc](http://download.microsoft.com/download/9/e/d/9edcdeab-f1fb-4670-8914-c08c5c6f22a5/HPC_Overview.doc)

A definition of computational clusters

[http://en.wikipedia.org/wiki/Computer\\_cluster](http://en.wikipedia.org/wiki/Computer_cluster)

Information on Active Directory

<http://www.microsoft.com/ad>

Introduction to Microsoft Compute Cluster Solution

<http://www.microsoft.com/hpc/>

Product Overview of Windows Compute Cluster Server 2003

<http://www.microsoft.com/windowsserver2003/ccs/overview.mspx>

Pre-boot eXecution Environment (PXE)

(Pronounced "pixie" and sometimes written referred to as "pixie boot")

[http://en.wikipedia.org/wiki/Preboot\\_Execution\\_Environment](http://en.wikipedia.org/wiki/Preboot_Execution_Environment)

Information on Remote Installation Service (RIS)

<http://support.microsoft.com/kb/325862>

Information on Installing Compute Cluster Pack

<http://technet2.microsoft.com/WindowsServer/en/library/2711035c-7452-4831-81a0-1038608a3e581033.mspx>

Reviewer's Guide for Windows Server 2003 Compute Cluster

<http://www.microsoft.com/windowsserver2003/ccs/reviewersguide.mspx>

For information on pricing and licensing, please visit:

<http://www.microsoft.com/windowsserver2003/ccs/howtobuy/pricing/default.mspx>

For CCS product requirements, please visit:

<http://www.microsoft.com/windowsserver2003/ccs/sysreqs.mspx>

For frequently asked questions (FAQ) about CCS, please visit:

<http://www.microsoft.com/windowsserver2003/ccs/faq.mspx>

For additional information or help, please visit:

<http://windowshpc.net/>

Information on Migrating Unix applications

<http://www.microsoft.com/technet/solutionaccelerators/cits/interopmigration/unix/hpcunxwn/ch11hpc.mspx>