TruCluster Software Products

Software Installation

Part Number: AA-R88HA-TE

January 1998

Product Version:	TruCluster Production Server Software Version 1.5, TruCluster Available Server Software Version 1.5, and TruCluster MEMORY CHANNEL Software Version 1.5
Operating System and Version:	DIGITAL UNIX Version 4.0D

This manual describes how to install and configure TruCluster Production Server Software, TruCluster Available Server Software, and TruCluster MEMORY CHANNEL Software on the DIGITAL UNIX operating system.

Digital Equipment Corporation
Maynard, Massachusetts

Contents

Examples

Figures

Tables

About This Manual

This manual describes how to install the following TruCluster Software Products kits on the DIGITAL UNIX operating system:

- TruCluster Production Server Software (Production Server)
- TruCluster Available Server Software (Available Server)
- TruCluster MEMORY CHANNEL Software (MEMORY CHANNEL)

The manual also explains how to verify and troubleshoot the TruCluster software installation.

Audience

This manual is for system administrators who will install, configure, and administer the TruCluster Software Products kits. The instructions in this manual assume that you are experienced in maintaining your hardware, operating system, and network.

Organization

This manual is organized as follows:

Chapter 1	Describes the types of installations and the contents of the TruCluster Software Products kits.
Chapter 2	Provides step-by-step instructions to prepare for the installation.
Chapter 3	Provides step-by-step instructions for a full installation.
Chapter 4	Provides step-by-step instructions for a rolling or simultaneous upgrade installation.
Chapter 5	Describes how to verify the installation using the `clu_ivp` utility.
Chapter 6	Analyzes potential problems related to cluster software installation.

Appendix A	Contains examples of console boot messages following a successful cluster software installation, cluster formation, and cluster recovery.
Appendix B	Summarizes changes to system files related to TruCluster software installation.

Related Documents

Consult the following TruCluster manuals for assistance in configuration, installation, and administration:

- TruCluster Software Products *Release Notes*
- TruCluster Software Products *Hardware Configuration*
- TruCluster Software Products *Administration*
- TruCluster Production Server Software *Application Programming Interfaces*
- TruCluster Production Server Software *MEMORY CHANNEL Application Programming Interfaces*

In addition, you should have available the following manuals from the DIGITAL UNIX documentation set:

- *Installation Guide*
- *Release Notes*
- *System Administration*
- *Network Administration*

Reader's Comments

DIGITAL welcomes any comments and suggestions you have on this and other DIGITAL UNIX manuals.

You can send your comments in the following ways:

- Fax: 603-884-0120 Attn: UBPG Publications, ZKO3-3/Y32
- Internet electronic mail: `readers_comment@zk3.dec.com`

  A Reader's Comment form is located on your system in the following location:

  `/usr/doc/readers_comment.txt`

- Mail:

  Digital Equipment Corporation
  UBPG Publications Manager
  ZKO3-3/Y32
  110 Spit Brook Road
  Nashua, NH 03062-9987

  A Reader's Comment form is located in the back of each printed manual. The form is postage paid if you mail it in the United States.

Please include the following information along with your comments:

- The full title of the book and the order number. (The order number is printed on the title page of this book and on its back cover.)
- The section numbers and page numbers of the information on which you are commenting.
- The version of DIGITAL UNIX that you are using.
- If known, the type of processor that is running the DIGITAL UNIX software.

The DIGITAL UNIX Publications group cannot respond to system problems or technical support inquiries. Please address technical questions to your local system vendor or to the appropriate DIGITAL technical support office. Information provided with the software media explains how to send problem reports to DIGITAL.

Conventions

This manual uses the following typographical conventions:

`#`	A number sign represents the superuser prompt.
`% cat`	Boldface type in interactive examples indicates typed user input.
file	Italic (slanted) type indicates variable values, placeholders, and function argument names.
⋮	A vertical ellipsis indicates that a portion of an example that would normally be present is not shown.
`cat`(1)	A cross-reference to a reference page includes the appropriate section number in parentheses. For example, `cat`(1) indicates that you can find information on the `cat` command in Section 1 of the reference pages.
PS	Abbreviation for the TruCluster Production Server Software.
AS	Abbreviation for the TruCluster Available Server Software.
MC	Abbreviation for the TruCluster MEMORY CHANNEL Software.

1

Introduction

This manual describes how to install and configure the TruCluster Production Server, TruCluster Available Server, and TruCluster MEMORY CHANNEL products on the DIGITAL UNIX operating system.

Important Note

The procedures in this manual assume that each system's hardware and firmware are installed and configured as described in the TruCluster Software Products *Hardware Configuration* manual.

Do not begin the software installation until the hardware and firmware are installed and configured.

Clusters are complex entities. Before you begin the installation, please read this manual, paying close attention to the sections that deal directly with your type of installation. Understanding the various types of installations and becoming familiar with the general sequence of events can save time and prevent later problems.

This chapter provides the following information:

- Installation types
- License and subset information
- Base system mandatory and optional subsets
- General restrictions and considerations
- Location of online documentation

Throughout this manual, the following abbreviations are used to refer to TruCluster products:

PS	TruCluster Production Server Software (Production Server)
AS	TruCluster Available Server Software (Available Server)
MC	TruCluster MEMORY CHANNEL Software (MEMORY CHANNEL)

1.1 Installation Types

There are three basic installation types:

full installation

Perform a full installation the first time you install a TruCluster product, or when you cannot (or choose not to) perform an upgrade. A full installation does not preserve any existing databases or configuration parameters.

Note that all TruCluster MEMORY CHANNEL Software installations are full installations.

rolling upgrade

Perform a rolling upgrade for Production Server or Available Server when clients need uninterrupted access to services. Each system is removed in turn, upgraded, and returned to the Production Server or Available Server cluster. The existing available server environment (ASE) database is automatically copied to the upgraded system when a system is returned to an ASE, so you do not have to set up services after completing the upgrade. The cluster remains in operation during a rolling upgrade.

DIGITAL recommends that you complete a rolling upgrade procedure as quickly as possible, and minimize any hardware and software changes during the upgrade.

simultaneous upgrade

This procedure requires that you shut down the Production Server or Available Server cluster. You can save and restore an existing ASE database, which means that you do not have to manually set up services after the installation. The cluster is not in operation during a simultaneous upgrade; no services are available to clients.

Note

When installing a TruCluster product in a cluster environment, use the same installation type for each system. Do not attempt to mix and match installation types. For example, you cannot perform a full installation on one member and then perform a rolling upgrade on another member. Decide which type of installation best fits your site's requirements, and use that installation type for all systems.

When installed as a separate product, TruCluster MEMORY CHANNEL Software is not part of a cluster environment. Therefore, the concept of rolling or simultaneous upgrade does not apply. For this reason, all TruCluster MEMORY CHANNEL Software software installations are considered full installations. However, you must delete existing MEMORY CHANNEL software subsets before installing the latest version of the product.

Table 1–1 shows the installation types supported for each TruCluster product.

Table 1–1: Supported Installation Types

Installation Type	Production Server	Available Server	MEMORY CHANNEL Software
Full installation	Yes	Yes	Yes
Rolling or simultaneous upgrade from Version 1.4 of the current product	No	Yes (if on DIGITAL UNIX Version 4.0B)	not applicable
Rolling or simultaneous upgrade from Version 1.4A of the current product	Yes	Yes	not applicable

Table 1–2 shows the rolling upgrade paths for Production Server. Table 1–3 shows the rolling upgrade paths for Available Server.

Table 1–2: Production Server Rolling Upgrade Paths

Current Product	Intermediate Step	Intermediate Step	Intermediate Step	Final Step
Production Server Version 1.4A on DIGITAL UNIX Version 4.0B				Production Server Version 1.5 on DIGITAL UNIX Version 4.0D
Production Server Version 1.4 on DIGITAL UNIX Version 4.0A	Production Server Version 1.4 on DIGITAL UNIX Version 4.0B	Production Server Version 1.4A on DIGITAL UNIX Version 4.0B		Production Server Version 1.5 on DIGITAL UNIX Version 4.0D
TruCluster Software Version 1.0 on DIGITAL UNIX Version 3.2G	Production Server Version 1.4 on DIGITAL UNIX Version 4.0A	Production Server Version 1.4 on DIGITAL UNIX Version 4.0B	Production Server Version 1.4A on DIGITAL UNIX Version 4.0B	Production Server Version 1.5 on DIGITAL UNIX Version 4.0D

Table 1–3: Available Server Rolling Upgrade Paths

Current Product	Intermediate Step	Intermediate Step	Final Step
Available Server Version 1.4A on DIGITAL UNIX Version 4.0B			Available Server Version 1.5 on DIGITAL UNIX Version 4.0D
Available Server Version 1.4 on DIGITAL UNIX Version 4.0A	Available Server Version 1.4 on DIGITAL UNIX Version 4.0B		Available Server Version 1.5 on DIGITAL UNIX Version 4.0D
DECsafe Available Server Version 1.3 on DIGITAL UNIX Version 3.2G	Available Server Version 1.4 on DIGITAL UNIX Version 4.0A	Available Server Version 1.4 on DIGITAL UNIX Version 4.0B	Available Server Version 1.5 on DIGITAL UNIX Version 4.0D

This manual provides the information needed to upgrade from:

- TruCluster Production Server Software Version 1.4A (the entire cluster must be at Version 1.4A on DIGITAL UNIX Version 4.0B)
- TruCluster Available Server Software Version 1.4A (the entire ASE must be at Version 1.4A on DIGITAL UNIX Version 4.0B)
- TruCluster Available Server Software Version 1.4 (the entire ASE must be at Version 1.4 on DIGITAL UNIX Version 4.0B)

If your cluster is not running one of these versions, you must perform one or more intermediate upgrades before you can migrate to Version 1.5. The procedures for upgrading to Version 1.4 and Version 1.4A of the TruCluster Software Products are described in the documentation for those releases.

Note

You do not have to remove TruCluster subsets when updating the DIGITAL UNIX operating system from Version 4.0A to Version 4.0B. See Section 4.6 for information on updating the operating system without reinstalling a TruCluster product.

1.2 License and Subset Information

Each of the following TruCluster products has a separate license:

- TruCluster Production Server Software
- TruCluster Available Server Software
- TruCluster MEMORY CHANNEL Software

Note

You load and register *only one* TruCluster product license on each system; the licenses are mutually exclusive. For example, if you load an Available Server license, you cannot load a Production Server or MEMORY CHANNEL Software license on the same system. If you load a Production Server license, you cannot load an Available Server or MEMORY CHANNEL license on the same system.

The Production Server product includes both Available Server and MEMORY CHANNEL software; if you load a Production Server license, you can install and use the subsets for the Available Server and MEMORY CHANNEL products.

If you are setting up an available server environment (ASE) cluster, you will load an Available Server license on each member system. If you are setting up a Production Server cluster, you will load a Production Server license on each member system. The pre-installation procedures in Section 2.9 describe how to load a license product authorization key (PAK).

The set of subsets available for installation depends on which software license you register. Table 1–4 lists the subsets associated with each product license.

Table 1–4: Contents of TruCluster Subsets

Subset Name	Description	Contents	License		
TCRASE15*n*	TruCluster Available Server Software	`asemgr` utility and ASE daemons	PS	AS	
TCRCMS15*n*	TruCluster Cluster Monitor	Cluster Monitor and supporting daemons	PS	AS	
TCRCOMMON15*n*	TruCluster Common Components	Common installation checks and procedures for all TruCluster products	PS	AS	MC
TCRCONF15*n*	TruCluster Configuration Software	Configuration and verification software for all TruCluster products	PS	AS	MC
TCRDSVC15*n*	TruCluster Production Server Software	Connection manager and distributed raw disk (DRD)	PS		
TCRMAN15*n*	TruCluster Reference Pages	Reference pages for all TruCluster utilities and APIs, and MEMORY CHANNEL routines[a]	PS	AS	MC
TCRMCA15*n*	TruCluster MEMORY CHANNEL Software	MEMORY CHANNEL driver and distributed lock manager (DLM)	PS		MC

[a]Because all reference pages are distributed in a single subset, some may not apply to your product. Each reference page indicates which products support the described utility, command, or API.

Table 1–5 shows the approximate disk space requirements, in megabytes (MB), for each Version 1.5 TruCluster subset in the root (`/`), `/var`, and `/usr` file systems.

Table 1–5: TruCluster Subset Sizes

Subset	Root (/) File System (MB)	/usr File System (MB)	/var File System (MB)	Total (MB)
TCRASE150	0.0383	7.6754	0.8262	8.5399
TCRCMS150	0	4.9470	0.0575	5.0045
TCRCOMMON150	0.0221	0.7494	0.0078	0.7793
TCRCONF150	0	0.0080	0	0.0080
TCRDSVC150	0	4.7012	1.0376	5.7388
TCRMAN150	0	0.6338	0	0.6338
TCRMCA150	0	1.7072	0	1.7072
PS Totals	**0.0604**	**20.422**	**1.9291**	**22.4115**

Table 1–5: TruCluster Subset Sizes (cont.)

Subset	Root (/) File System (MB)	/usr File System (MB)	/var File System (MB)	Total (MB)
AS Totals	**0.0604**	**14.0136**	**0.8915**	**14.9655**
MC Totals	**0.0221**	**3.0984**	**0.0078**	**3.1283**

1.3 Base System Mandatory and Optional Subsets

Table 1–6 describes the TruCluster product dependencies on DIGITAL UNIX and layered product subsets. Subsets listed as *M* are mandatory; these subsets must be loaded on a system in order to install TruCluster products. Subsets listed as *O* are optional; the "Comment" column describes why you might choose to load one or more of these subsets. See the DIGITAL UNIX *Installation Guide* for more information on these subsets and their dependencies.

Table 1–6: Mandatory and Optional Operating System Subsets

Subset	Description	Mandatory / Optional	Comment
OSFBASE	Base System	M	
OSFBIN	Standard Kernel Objects	M	
OSFBINCOM	Kernel Header and Common Files	M	
OSFHWBIN	Hardware Kernel Modules	M	
OSFHWBINCOM	Hardware Kernel Header and Common Files	M	
OSFCLINET	Basic Networking Services	M	
OSFCMPLRS	Compiler Back End	M	
OSFNFS	NFS® Utilities	M	
OSFATMBASE	ATM Commands	O	Required if Asynchronous Transfer Mode (ATM) hardware is detected during the installation process.
OSFATMBIN	ATM Kernel Modules	O	Required if ATM hardware is detected during the installation process.

Table 1–6: Mandatory and Optional Operating System Subsets (cont.)

Subset	Description	Mandatory / Optional	Comment
OSFATMBINCOM	ATM Kernel Header and Common Files	O	Used for kernel software development.
OSFATMBINOBJECT	ATM Kernel Object	O	Used for kernel software development.
OSFDCMT	Documentation Preparation Tools	O	Required by TruCluster Reference Pages subset (TCRMAN).
OSFMAN*xxx*	Reference Pages subsets (base system)	O	Some TruCluster reference pages refer to the DIGITAL UNIX operating system reference pages.
OSFLSMBASE	Logical Storage Manager (LSM)	O	Required if setting up ASE services that use LSM.
OSFLSMBINCOM	Logical Storage Manager Kernel Header and Common Files	O	Required if setting up ASE services that use LSM.
OSFLSMBIN	LSM Kernel Modules	O	Required if setting up ASE services that use LSM.
OSFLSMX11	Logical Storage Manager GUI	O	Required to launch the LSM graphical user interface (GUI) from the Cluster Monitor.
OSFADVFS	POLYCTR AdvFS	O	Required for the POLYCENTER Advanced File System (AdvFS).
OSFADVFSBIN	POLYCTR AdvFS Kernel Objects	O	Required for AdvFS.
OSFRIS	Remote Installation Service	O	Required to use a cluster member as a RIS server.
OSFINET	Additional Networking Services	O	Required to use a cluster member as a RIS server.
OSFKTOOLS	Kernel Debugging Tools	O	Required to use the distributed raw disk (DRD) extensions to the kernel debugging utility, kdbx.

1.4 General Restrictions and Considerations

Note the following general installation restrictions and considerations:

- Do not install the product into a dataless environment.
- It is recommended that you have at least 64 MB of memory available on each member system.
- You must have superuser (root) privileges for the systems on which you will install the software.
- Back up all systems before beginning the installation process.
- If you generate a new system configuration file either by invoking the `doconfig` command without the `-c` option or by entering the `sizer -n` command, you must run the `ase_fix_config` utility after you enter the command. Section 3.7 describes the operation of the `ase_fix_config` utility.

  The `doconfig` program automatically runs the `ase_fix_config` script only when differences in the SCSI bus configuration are detected.
- After a cluster is created at, or upgraded to, Version 1.5, you cannot add a new system to the cluster unless that system is also at Version 1.5.

Section 4.1 lists additional restrictions for rolling upgrades.

1.5 Location of Online Documentation

Each book in the TruCluster Software documentation set is shipped as a set of Hypertext Markup Language (HTML) and graphics files in the `/TCR/doc/html` directory on the Associated Products Volume 2 CD-ROM. You can use the Netscape® Navigator™ World Wide Web browsing program to display these books.

If the DIGITAL UNIX installation program detects graphics capabilities on your system, it automatically installs Netscape Navigator. You can then invoke Netscape from an icon on the Common Desktop Environment (CDE) front panel or directly from the command line. Detailed help for Netscape is available through the help menus.

To access the TruCluster Software documentation from the viewer, click on the Open icon in the Netscape main window and enter the following file location in the Open Location: text entry field:

`file:/`*`mountpoint`*`/TCR/doc/html/BOOKSHELF.HTM`

2

Preinstallation Tasks

This chapter describes the tasks that you perform before loading the latest TruCluster subsets. No matter what type of installation you are performing, read this chapter. If you have not read Chapter 1, please do so before continuing.

To complete the preinstallation tasks, you need access to the following documents:

- TruCluster Software Products *Release Notes*
- TruCluster Software Products *Hardware Configuration*
- TruCluster Software Products *Administration*
- A TruCluster product authorization key (PAK) for each license
- DIGITAL UNIX *Installation Guide*
- DIGITAL UNIX *Release Notes*
- DIGITAL UNIX *System Administration*
- DIGITAL UNIX *Network Administration*
- Oracle7 Parallel Server *Concepts and Administration Guide* (if you are performing a rolling upgrade of a Production Server cluster and are running Oracle® Parallel Server™ (OPS) software)

Table 2–1 summarizes the preinstallation tasks. It lists the tasks in order, shows the TruCluster products to which each task applies, and provides pointers to necessary information.

Note

For Production Server, a cluster member does not have to be a member of an available server environment (ASE), although there must be at least one ASE in the cluster. If you are installing Production Server on a system that will not be part of an ASE, ignore the tasks related to ASE membership and services.

Table 2–1: Preinstallation Tasks

Task	Product			See:
Read the TruCluster Software Products *Release Notes*.	PS	AS	MC	TruCluster Software Products *Release Notes*
Check the hardware and firmware for installation readiness.	PS	AS	MC	TruCluster Software Products *Hardware Configuration*
Obtain any needed IP addresses (MC subnet, ASE services).	PS	AS	MC	Section 2.1
Back up the system.	PS	AS	MC	DIGITAL UNIX *System Administration*
If an upgrade, know the value of the system's `CLUSTER_NET` (cluster interconnect) variable in `/etc/rc.config`.	PS	AS		Use `rcmgr get CLUSTER_NET`
If an upgrade, know the value of the system's `ASE_ID` (ASE ID) variable in `/etc/rc.config`.	PS			Use `rcmgr get ASE_ID`
If an upgrade on a two-system, virtual-hub cluster, know the value of the system's `CNX_DISK` (tie-breaker disk) variable in `/etc/rc.config`.	PS			Use `rcmgr get CNX_DISK`
If a rolling upgrade and running Oracle® Parallel Server™ (OPS) software, stop all OPS database activity on the system.	PS			Oracle7 Parallel Server *Concepts and Administration Guide*
If a rolling upgrade, delete this member from the ASE. Run the `asemgr` utility on a system that is a member of the same ASE as the member you want to delete. You cannot run the `asemgr` utility on the member system you want to delete. If a simultaneous upgrade, put all services off line.	PS	AS		Section 2.2 and TruCluster *Administration*
Halt the system and set console variables.	PS	AS	MC	Section 2.3
If there are TruCluster subsets on the system, boot `/genvmunix` to single-user mode and deinstall these subsets. Decide whether or not to save an existing ASE database.	PS	AS	MC	Section 2.4

Table 2–1: Preinstallation Tasks (cont.)

Task	Product			See:
Install the DIGITAL UNIX operating system.	PS	AS	MC	Section 2.5 and the DIGITAL UNIX *Installation Guide*
Configure basic network services (optional for MC).	PS	AS	MC	Section 2.6 and the DIGITAL UNIX *Network Administration* manual
Configure a time service (optional for MC).	PS	AS	MC	Section 2.7 and the DIGITAL UNIX *Network Administration* manual
If performing a simultaneous upgrade and using a manually saved ASE database, restore the ASE database.	PS	AS		Section 2.8
Register *one* TruCluster product license per system.	PS	AS	MC	Section 2.9

2.1 Obtain IP Names and Addresses (PS, AS, MC)

Depending on which TruCluster product you install, you need to allocate some new Internet Protocol (IP) names and addresses and, for Production Server and MEMORY CHANNEL, a dedicated subnet. The following table shows the address requirements for the TruCluster products.

	Production Server	Available Server	MEMORY CHANNEL
Need dedicated MC subnet with a unique IP address for each system's interface to the subnet	Yes	No	Yes
Need IP addresses for ASE services that require IP addresses to be eligible for failover (these addresses are not on the dedicated MEMORY CHANNEL subnet)	Yes	Yes	Not applicable

The terminology used for network connections by the installation procedures depends on which TruCluster product you install, as follows:

- Production Server uses **cluster interconnect** when referring to a system's MEMORY CHANNEL interface. The IP name and address for this interface are used when establishing a system's membership in a cluster, and for establishing a system's membership in an ASE. Note that Production Server supports redundant MEMORY CHANNEL subnet interfaces (for failover). On any cluster member, these adapters share a single network address; only one `IFCONFIG` and `NETDEV` entry represent a system's cluster interconnect in its `/etc/rc.config` file, regardless of the number of physical MEMORY CHANNEL interfaces present.
- Available Server uses **member network interface** when referring to a system's primary network IP name and address. The IP name and address for this interface are used when establishing a system's membership in an ASE.
- MEMORY CHANNEL uses **MEMORY CHANNEL adapter** when referring to a system's MEMORY CHANNEL interface.

In addition to the terminology distinctions, there is a basic difference between Production Server and Available Server in the network connections between member systems:

- Production Server requires a dedicated MEMORY CHANNEL subnet, which is used only for intracluster communication. This MEMORY CHANNEL subnet is the primary network for the cluster; it is a separate and distinct physical network on which only cluster members reside.

  You need an IP address (subnet mask 255.255.255.0) and an IP name for the MEMORY CHANNEL cluster interconnect on each member system. See the DIGITAL UNIX *Network Administration* manual for guidelines on allocating IP addresses. Also see RFC 1918, in which the Internet Assigned Numbers Authority (IANA) reserves the following blocks of IP address space for use by private internets:

```
10.0.0.0      -   10.255.255.255   (10/8 prefix)
172.16.0.0    -   172.31.255.255   (172.16/12 prefix)
192.168.0.0   -   192.168.255.255  (192.168/16 prefix)
```

Note

On the MEMORY CHANNEL subnet, host number 42 is reserved by the Production Server installation procedure as an alias (`cluster_cnx`) for the connection manager service (`cnxmgrd`). The installation procedure automatically puts a host entry for `cluster_cnx` in the `/etc/hosts` file. You cannot assign host number 42 to another service or member system on the dedicated cluster subnet.

- For Available Server, standard network interfaces are used as primary and backup network interfaces. There is no dedicated MEMORY CHANNEL subnet.

For both Production Server and Available Server, the IP addresses used for highly available services must be accessible to clients requesting those services. Therefore, for a Production Server configuration, these addresses cannot be on the dedicated MEMORY CHANNEL subnet. (See the TruCluster *Administration* manual for information on setting up highly available services.)

Each product's installation procedure updates the `/etc/hosts` and `/etc/rc.config` file to reflect the IP names and addresses you supply during installation. For Production Server and Available Server, the installation procedure also assigns the IP name of the primary interface to the `CLUSTER_NET` variable in `/etc/rc.config`.

You must manually add entries for the remaining cluster or ASE members to the `/etc/hosts` file on each system. You can add these entries to `/etc/hosts` before installing the TruCluster software. You must add these entries before booting the kernel built during the TruCluster installation.

Because the ASE database uses unqualified host names, each entry must include an alias; for example:

```
123.123.123.123   clu14.abc.def.com  clu14
```

Note that the system's host name (the one displayed by the `hostname` program) does not change as a result of installing TruCluster software.

2.2 Prepare ASE Services and Members (PS, AS)

How you deal with available server environment (ASE) services and members depends on whether you are performing a rolling or a simultaneous upgrade. The following sections describe the differences between the two types of upgrade installations.

2.2.1 Rolling Upgrade

To prepare an ASE member for a rolling upgrade, run the `asemgr` utility on another system that is part of the same ASE and delete the member from the ASE.

Note

The `asemgr` utility inspects all services in the ASE database that have an automatic service placement (ASP) policy of

Favored Member or Restricted to Favored Member, and removes the deleted member from the list of favored members. If this makes any list empty, the following rules apply:

- A service with the favored policy effectively becomes one with the balanced policy.
- A service with the restricted policy cannot be started until its ASP policy is modified.

2.2.2 Simultaneous Upgrade

For a simultaneous upgrade, run the `asemgr` utility (on one member system only in each ASE) and perform one of the following tasks:

- If you will reuse the existing ASE database, put all services off line.
- If you will not reuse the existing ASE database, delete all services.

2.3 Halt System and Set Console Variables (PS, AS, MC)

To halt the system and set the console variables, follow these steps:

1. Halt the system. For example, to halt the system from multiuser mode with no other users on the system, enter the following command:

   ```
   # shutdown -h now
   ```

2. If your system supports the `bus_probe_algorithm` console variable, set its value to `new`. This ensures that peripheral component interconnect (PCI) devices are consistently probed on all member systems. To check the setting, enter the following command at the console prompt:

   ```
   >>> show bus_probe_algorithm
   bus_probe_algorithm  new
   ```

   If necessary, enter the following command to set the `bus_probe_algorithm` variable to `new`:

   ```
   >>> set bus_probe_algorithm new
   ```

3. In order to bring the system to a known state at each reboot, set the `boot_reset` console variable as follows:

   ```
   >>> set boot_reset on
   ```

2.4 Boot to Single-User Mode and Deinstall TruCluster Software (PS, AS, MC)

To deinstall existing TruCluster software, follow these steps :

1. From the console prompt, boot `/genvmunix` to single-user mode; for example:

   ```
   >>> boot -fl s -fi /genvmunix
   ```

2. Enter the `bcheckrc` command, which makes the `root` file system writable and mounts local file systems:

   ```
   # /sbin/bcheckrc
   ```

3. To make sure that the system's licenses are loaded and active, run the following LMF commands:

   ```
   # lmf reset
   # lmf list
   ```

4. Use the `setld -i` command to determine which TruCluster software subsets are installed. The following example shows a Version 1.4A installation, where *nnn* represents the TruCluster product kit number, for example `TCRCONF141`:

   ```
   #  setld -i | grep '^TCR'

   TCRASEnnn    installed  TruCluster Available Server Software(TruCluster(TM) \
                           Software)
   TCRCMSnnn    installed  TruCluster Cluster Monitor (TruCluster(TM) Software)
   TCRCOMMONnnn installed  TruCluster Common Components (TruCluster(TM) Software)
   TCRCONFnnn   installed  TruCluster Configuration Software (TruCluster(TM) \
                           Software)
   TCRDSVCnnn   installed  TruCluster Production Server Software(TruCluster(TM) \
                           Software)
   TCRMANnnn    installed  TruCluster Reference Pages (TruCluster(TM) Software)
   TCRMCAnnn    installed  TruCluster MEMORY CHANNEL(TM) Software(TruCluster(TM) \
                           Software)
   ```

5. Use the `setld -d` command to deinstall the subsets.

   To ensure that the subsets are deleted in an order that resolves any dependencies between subsets, delete all installed TruCluster subsets with one `setld -d` command. The following example shows how to delete existing subsets:

   ```
   # setld -d TCRCONFnnn TCRMANnnn TCRCMSnnn TCRDSVCnnn \
   TCRMCAnnn TCRASEnnn TCRCOMMONnnn
   ```

   If the deinstallation procedure detects an existing ASE database, `/var/ase/config/asecdb`, the procedure asks whether you want to save the existing database:

   ```
   Do you want to save the ASE database? [y]
   ```

For a rolling upgrade, do not save the ASE database. The remaining members of the ASE will repropagate the database when the member is returned to its ASE.

For a simultaneous upgrade, you usually want to save the existing ASE database so you can restart existing services after completing the upgrade.

Caution

If you answer **n**, the database is deleted. No copy is saved. If you are performing a simultaneous upgrade and delete the database on all systems, you will have to recreate all services after completing the upgrade.

If you save the existing ASE database and are performing a full installation of the DIGITAL UNIX operating system, either back up the database or copy it to another system so that the DIGITAL UNIX Version 4.0D installation does not overwrite the file.

2.5 Install the DIGITAL UNIX Operating System (PS, AS, MC)

Before starting the installation procedures described in the DIGITAL UNIX *Installation Guide*, read the following list and incorporate these tasks into the installation:

- Before installing the operating system, turn on the power to all member systems, the MEMORY CHANNEL hub (if one is included in your configuration), and external disks.
- Load all mandatory and any required optional subsets from Table 1–6.
- If you are installing the DIGITAL UNIX operating system on a system in an ASE, a disk that is connected to the shared SCSI bus might be listed in the root disk installation menu. Do not install the operating system on a disk that is connected to the shared bus because the data on the disk could be overwritten. DIGITAL recommends that the disks containing the standard system-level file systems (such as root, swap, `/usr`, and `/var`) reside on private, nonshared buses.
- If you install new hardware (for example, new MEMORY CHANNEL adapters) after you install or update the DIGITAL UNIX operating system, remember to boot `/genvmunix` and build a customized kernel. Otherwise, the system's kernel configuration file will not contain these hardware options, and the kernel you build during TruCluster installation will not recognize the new hardware. See the DIGITAL

UNIX *System Administration* manual for more information on configuring kernels.

- This step applies to systems connected to Asynchronous Transport Mode (ATM) networks. To configure support for ATM LAN Emulation (LANE), select the necessary options from the list displayed by `doconfig`. In the following partial list of `doconfig` options, those options required for LANE support are marked with an asterisk (*):

```
      IP Switching over ATM (ATMIFMP)
*     LAN Emulation over ATM (LANE)
      Classical IP over ATM (ATMIP)
*     ATM UNI 3.0/3.1 Signalling for SVCs
```

If the following statements apply to your upgrade path, boot `/genvmunix` after installing the DIGITAL UNIX operating system and before loading the TruCluster Version 1.5 software:

- You are performing an update installation of the DIGITAL UNIX operating system.
- You are performing a rolling or simultaneous upgrade of Production Server or Available Server.

Booting `/genvmunix` before upgrading the TruCluster product removes the possibility of a TruCluster kernel build failure caused by unresolved symbols.

2.6 Configure Basic Network Services (PS, AS, MC)

Using the information in the DIGITAL UNIX *Network Administration* manual, set up your network and configure basic network services such as the Berkeley Internet Name Domain (BIND) service and the Network Information Service (NIS).

If you configure BIND or NIS:

- To configure the members of the cluster as BIND servers, set up one member as a primary server and the rest as secondary servers. To configure the members of the cluster as NIS servers, set up one member as a master server and the rest as slaves. See the information on setting up BIND and NIS servers in the DIGITAL UNIX *Network Administration* manual.
- Because the Cluster Monitor's high-availability feature will not work if cluster members cannot resolve IP addresses, make sure that the `hosts` entry in `/etc/svc.conf` has the `local` service listed before the `bind` or `yp` services. For example:

```
hosts=local,bind,yp
```

- Make sure that user IDs (UIDs) and group IDs (GIDs) are consistent across all members of the cluster.

2.7 Configure a Time Service (PS, AS, MC)

Set up a distributed time service such as the Network Time Protocol (NTP) daemon (`xntpd`) on each member system. NTP provides highly accurate synchronization and tracks reliability of time sources. For information on the NTP daemon, see the DIGITAL UNIX *Network Administration* manual.

Note

If system times are not synchronized, checks that rely on accurate timestamps (for example, in determining which ASE database is the most recent) will fail.

Because the system times of cluster members should not vary by more than a few seconds, DIGITAL recommends that you not use the `timed` daemon to set the time.

Generally, there is some system in your environment that is considered most informed as far as time is concerned. This system may be getting time from some other system that is considered a reliable time source. If you want the cluster to act as a reliable time source, set up the time service as an ASE service.

2.8 Restore the ASE Database (PS, AS)

If you are performing a simultaneous upgrade, and if you manually moved the ASE database file to another location or system, copy it to its original location: `/var/ase/config/asecdb`. To set the file's owner to `root`, group to `system`, and mode to `644`, enter the following commands:

```
# cd /var/ase/config
# chown root:system asecdb
# chmod 644 asecdb
```

2.9 Register One TruCluster Software License (PS, AS, MC)

Before you install a TruCluster product, you must use its Product Authorization Key (PAK) to register a product license. (TruCluster licenses are described in Section 1.2.) A PAK is included in your product kit. If you

do not have a PAK, make contact with your DIGITAL Customer Services representative.

As shown in the following table, each TruCluster product has a separate PAK:

Product	PAK
TruCluster Production Server Software	TCR-UA
TruCluster Available Server Software	ASE-OA
TruCluster MEMORY CHANNEL Software	MCA-UA

For information on installing a PAK, see the DIGITAL UNIX *Software License Management* manual, `lmf`(8), and `lmfsetup`(8).

Note

You must register the appropriate PAK before installing a TruCluster product; if no PAK is registered, the installation procedure displays the following message:

```
There are no TruCluster Software licenses installed. In order
to install a TruCluster product you must first install the appropriate
LMF PAK (TCR-UA or MCA-UA or ASE-OA).
```

3

Full Installation

This chapter explains how to perform a full installation of a TruCluster product. If you are installing for the first time, use this chapter. Before you begin the installation, complete the preinstallation tasks described in Chapter 2.

Table 3–1 summarizes the full installation tasks. It lists the tasks in order, shows the TruCluster products to which each task applies, and provides pointers to necessary information.

Note

For Production Server, a cluster member does not have to belong to an available server environment (ASE), although there must be at least one ASE in the cluster. If you are installing Production Server on a system that will not be part of an ASE, ignore the tasks related to ASE membership and services.

Table 3–1: Full Installation Tasks

Task	Product			See:
Load the TruCluster kit. The installation procedure starts automatically when you load the kit.	PS	AS	MC	Section 3.1
Specify the IP name and address of the cluster interconnect (PS), the member network interface (AS), or the adapter (MC).	PS	AS	MC	Section 2.1 and Section 3.2
Specify an ASE identifier.	PS			Section 3.3
Decide whether to run the ASE logger daemon on this system.	PS	AS		Section 3.4
Initialize the ASE database.	PS	AS		Section 3.5
Select a kernel configuration file.	PS	AS	MC	Section 3.6
Renumber shared SCSI buses.	PS	AS		Section 3.7
Build and install a new kernel.	PS	AS	MC	Section 3.8
Add host entries to `/etc/hosts` (optional for MC).	PS	AS	MC	Section 2.1 and Section 3.9

Table 3–1: Full Installation Tasks (cont.)

Task	Product			See:
Enable new distributed lock manager (DLM) interfaces and reboot the system.	PS			Section 3.10
Reboot the system.		AS	MC	Section 3.11
After all systems are installed, populate the member list in the ASE database.	PS	AS		Section 3.12
After all systems are installed, create consistent device special files for an ASE tape service.	PS	AS		Section 3.13
After all systems are installed, specify a tie-breaker disk (only for a two-system, virtual-hub cluster).	PS			Section 3.14
After all systems are installed, run `clu_ivp` to verify the installation.	PS	AS	MC	Chapter 5

3.1 Load the TruCluster Kit (PS, AS, MC)

Caution

The following procedure assumes that you are either installing a TruCluster product for the first time, or that you performed a full installation of the DIGITAL UNIX Version 4.0D operating system. In either case, your `/vmunix` kernel is compatible with the TruCluster installation.

However, if you are performing a rolling or simultaneous upgrade, and used the update installation procedure for the DIGITAL UNIX operating system, boot `/genvmunix` before loading the TruCluster kit. Booting `/genvmunix` removes the possibility of a TruCluster kernel build failure caused by unresolved symbols.

To load the TruCluster kit, follow these steps:

1. Log in as superuser.
2. Change directory to root (`cd /`).
3. Mount the device or directory containing the TruCluster product kit.
4. Enter the `setld -l` command and specify the directory where the kit is located. For example:

   # **`setld -l /TCR150/kit`**

The installation procedure starts automatically and lists the available mandatory and optional subsets associated with the license you registered. (See Table 1–4 for a description of the subsets associated with each license.) You can choose one of the following subset installation options:

- All mandatory subsets only
- All mandatory and selected optional subsets
- All mandatory and all optional subsets

DIGITAL recommends that you choose the "All mandatory and all optional subsets" option.

After you select an option, the installation procedure checks that there is sufficient file system space. After this check complete satisfactorily, the installation procedure copies the subsets onto your system. (The following directories are the default locations for the majority of installed files: `/opt/TCR`*nnn*`/`, `/usr/opt/TCR`*nnn*`/`, and `/var/opt/TCR`*nnn*`/`. The `/usr/opt/TCR`*nnn*`/sbin/clu_install` script controls most of the installation process.)

Note

You cannot install individual TruCluster product subsets. For example, the following command results in an error:

```
# setld -l /TCR150/kit/TCRASE150
TCRASE150 cannot be installed. Please do not install subsets
individually.
```

3.2 Specify a Network Interface IP Name and Address (PS, AS, MC)

The installation procedure prompts you for an Internet Protocol (IP) name and address to associate with the system's network interface. For Production Server and Available Server, this IP name is stored as the value of the `CLUSTER_NET` variable in the `/etc/rc.config` file. (See Section 2.1 for information about required IP names and addresses.)

In the following Production Server example, the MEMORY CHANNEL IP name is formed by adding the prefix `mc` to the current host name (`clu14`) in order to identify this IP name as an interface to the MEMORY CHANNEL subnet:

```
Configuring "TruCluster Configuration Software " (TCRCONF150)

Enter the IP NAME for the cluster interconnect: mcclu14
```

Note

If you make a mistake when specifying the IP name for the cluster interconnect, press Return when prompted for the IP address. The installation procedure will prompt for a new IP name.

The installation procedure reads the system's `/etc/hosts` file to determine whether an entry exists for the IP name. If an entry for the IP name exists, the installation procedure displays the entry and asks whether you want to replace the existing entry with the IP name and address you just specified. For example:

- If you want to use the existing `/etc/hosts` entry, answer **n**; the information you specified during installation is ignored.
- If you want to replace the existing `/etc/hosts` entry, answer **y**. The installation procedure then replaces the entry in the`/etc/hosts` file with the IP name and address you specified.

The installation procedure automatically configures the network interface.

3.3 Specify an ASE Identifier (PS)

Production Server lets you organize systems and shared storage devices into one or more nonoverlapping available server environments (ASEs). The following list provides basic information about ASEs and ASE IDs:

- An ASE manages a collection of systems and the shared SCSI buses to which they are connected, providing an environment in which services can be started, stopped, and automatically relocated in response to software or hardware failures.
- An ASE must have at least two member systems. A system can belong to only one ASE.
- Each ASE has a unique ASE identifier (ASE_ID), a variable stored in `/etc/rc.config` with a value of 0-63, inclusive. All members in an ASE have the same ASE_ID.
- The concept of multiple ASEs, and therefore the need for multiple ASE_IDs, applies only to Production Server.

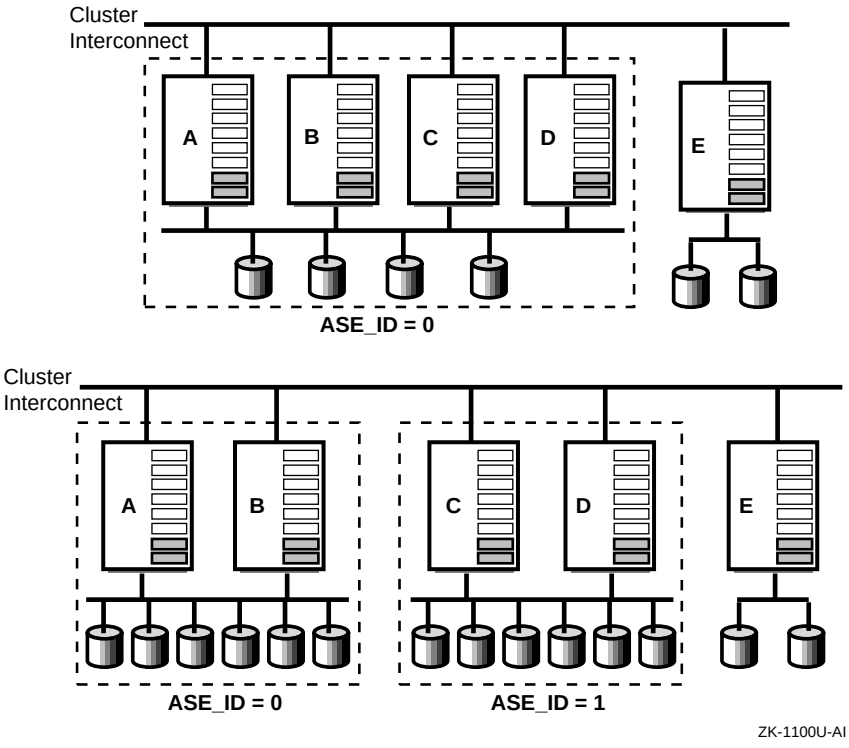
  If you install Available Server, there is only one ASE and its ASE_ID is automatically set to `ASE_ID=0` by the installation procedure.
- A system can be a member of a Production Server cluster without being a member of an ASE. If you answer `n` when prompted `Will the system be in an ASE? [y]`, the installation procedure does not put the ASE_ID variable in `/etc/rc.config` and sets `ASE="off"`.

Figure 3–1 shows two examples of ASEs within a cluster.

In the first example, systems A, B, C, and D are in one ASE. When a cluster consists of a single ASE, DIGITAL recommends that you assign the default ASE_ID value of 0 to each system. System E is a member of the cluster, but is not a member of the ASE.

In the second example, the cluster is organized into two ASEs, each with two systems. Systems A and B have an ASE_ID of 0, and systems C and D have an ASE_ID of 1. System E is a member of the cluster, but is not a member of an ASE.

Figure 3–1: Assigning ASE_IDs to Systems in a Cluster



Caution

It is critical that you correctly assign ASE_IDs because Production Server does not automatically detect conflicts in ASE membership or errors in ASE_ID assignment. After installing Production Server on all cluster members, use the `clu_ivp` utility to verify the ASE_IDs (see Chapter 5).

The installation procedure asks whether the system will be part of an ASE. Answer **y** to this prompt if you want the system to be part of an ASE.

```
Will the system be in an ASE? [y]: Return
```

Note

If you answer **n** to this prompt, the installation program does not prompt for an ASE_ID.

In the following example, the cluster contains two ASEs; the first ASE has ASE_ID=0 and the second ASE has ASE_ID=1. The system being installed will belong to the second ASE.

```
You must now enter an ASE identifier (ASE_ID) for this node.  All nodes within
the same ASE must use the same ASE_ID.  Each separate ASE must have its own
unique ASE_ID.  The range of valid ASE_ID numbers is 0 to 63 inclusive.

Enter the ASE_ID number. [0]: 1
```

3.4 Decide Whether to Run the ASE Logger Daemon on This System (PS, AS)

The installation procedure asks whether you want to run the ASE logger daemon (aselogger) on this system:

```
Do you want to run the ASE logger on this node? [n]: y
```

If you answer **y** to this prompt, the ASE logger daemon starts when the system is rebooted.

DIGITAL recommends that you run the ASE logger daemon on at least one system in each ASE. You can run the ASE logger daemon on more than one system; however, you will have virtually duplicate logs on these member systems. There is no synchronization between the logs on different systems.

If you decide to run only one ASE logger daemon in each ASE, run it on a highly available system or a standby system. If the ASE logger daemon is not running on any member system, or the system running the logger daemon goes down, messages pertaining to cluster operation are logged locally.

If you choose not to run the ASE logger daemon when you install TruCluster software, you can later set the ASELOGGER variable to start a logger daemon each time the system boots by entering the following command:

```
# rcmgr set ASELOGGER 1
```

For more information on the ASE logger daemon, see the TruCluster Software Products *Administration* manual.

3.5 Initialize the ASE Database (PS, AS)

If the installation is a first time installation, the installation procedure automatically creates and initializes a new ASE database, `/var/ase/config/asecdb`. Whenever the ASE database is initialized, the following message is displayed:

```
Initializing a new V1.5 ASE database ...
```

Section 4.2 describes the options for rolling and simultaneous upgrades when the installation procedure detects an existing ASE database.

3.6 Select a Kernel Configuration File (PS, AS, MC)

At this point in the installation process, the kernel configuration and build procedure begins. You are prompted for the name of a kernel configuration file. You can accept the default or enter the name of another configuration file. In the following example, the default configuration file, `CLU14`, is accepted:

```
The kernel will now be configured using "doconfig".

Enter the name of the kernel configuration file. [CLU14]: Return
```

After you specify the name of the kernel configuration file, the installation procedure asks whether you want to edit the file (after first saving the original configuration file with a `.bck` extension):

```
*** KERNEL CONFIGURATION AND BUILD PROCEDURE ***

Saving /sys/conf/CLU14 as /sys/conf/CLU14.bck

Do you want to edit the configuration file? (y/n) [n]: Return
```

To edit the kernel configuration file, answer **y**. Otherwise, accept the default response (**n**). If you answer **y** and the EDITOR shell environment variable is defined, `doconfig` starts that editor; otherwise, it starts `ed`.

3.7 Renumber Shared SCSI Buses (PS, AS)

Because systems can have different numbers of internal buses, a SCSI controller installed in an I/O bus slot might not have the same bus number on each member. However, within an ASE, each system must use the same

device number when referencing a shared device. The relationship between device numbers and logical bus numbers is as follows:

- Device numbers are derived from the logical bus numbers defined in each system's configuration file. If you connect a shared bus to SCSI controllers that have the same logical bus number on each system, each shared device will have the same device number on each system.
- Bus numbers are assigned to SCSI controllers during the kernel configuration process and are specified in the system configuration file. When you configure a kernel by running the `doconfig` program using the generic kernel, an algorithm is used to probe the SCSI controllers installed in the system. As the probe algorithm encounters the adapters, it assigns logical bus numbers to the SCSI controllers in sequence, starting with the number `0`.

Because an ASE depends on each member system having a consistent view of shared SCSI devices, the `doconfig` program automatically calls the ASE I/O bus renumbering utility (`/var/ase/sbin/ase_fix_config`), which lets you assign a specific bus number to each external SCSI controller installed in a system before you rebuild the kernel.

Note

The first time `ase_fix_config` is run, it saves information about the SCSI controller options in the system's kernel configuration file. Whenever you run `doconfig`, it automatically calls `ase_fix_config`. The `ase_fix_config` utility then compares its saved SCSI information with that in the current version of the configuration file. If it detects no SCSI controller differences, it exits silently. However, when you invoke `ase_fix_config` directly, it does not exit silently; you have the option to view and modify controller information.

The following sections show you how to perform these tasks:

- Selecting shared SCSI controllers
- Specifying bus numbering
- Changing references to renumbered devices in system files

3.7.1 Selecting Shared SCSI Controllers

The `ase_fix_config` utility displays the SCSI controllers that can be used for the shared buses in the ASE on the Shared I/O Bus Selection Menu. Select the SCSI controllers whose bus numbers you want to reassign. (Local bus and unsupported SCSI controllers are not given a option numbers.)

```
The ASE I/O Bus Renumbering Tool has been invoked

Select the controllers that define the shared ASE I/O buses.

        Name    Controller      Slot    Bus     Slot
     )  scsi0   psiop0          0       pci0    1
    1)  scsi1   pza0            0       pci0    6
    2)  scsi2   pza1            0       pci0    7

    q)  Quit without making changes

Enter your choices (comma or space separated): 1 2

        scsi1 pza0 0 pci0 6
        scsi2 pza1 0 pci0 7

Are the above choices correct (y|n)? [y]:  Return
```

3.7.2 Specifying Bus Numbering

After you select the SCSI controllers, you can specify a number at which to start numbering the controllers.

```
                I/O Controller Name Specification Menu

All controllers connected to an I/O bus must be named the same on all ASE
members.  Enter the controller names for all shared ASE I/O buses by assigning
them one at a time or all at once with the below options.

        Name            New Name        Controller      Slot    Bus     Slot
    1)  scsi1           scsi1           pza0            0       pci0    6
    2)  scsi2           scsi2           pza1            0       pci0    7

    f)  Assign buses starting at a given number
    v)  View non shared controllers
    s)  Show previous assignments
    q)  Quit without making any changes
    x)  Exit (done with modifications)

Enter your choice [f]:  Return

What number would you like your shared scsi controllers to start at? [16]: 4
```

Note

Bus numbers equal to or greater than the starting number are assumed to be shared buses. Bus numbers less than the starting number are assumed to be private to this system. When choosing a starting number, pick a number that is high enough to allow for expansion of private storage without having to renumber shared buses.

When you finish renumbering buses, choose option **x** to exit. The installation procedure displays the new SCSI controller information. When you confirm that the new SCSI controller configuration is correct, the

ase_fix_config utility saves the original configuration file with an .asesave extension, changes the numbers assigned to the SCSI buses in the kernel configuration file, and creates device special files.

```
Your new scsi controller configuration is:

        Name    Controller      Slot    Bus     Slot
        scsi0   psiop0          0       pci0    1
        scsi4   pza0            0       pci0    6
        scsi5   pza1            0       pci0    7

Is this ok? [y]:  Return

Moving scsi controllers:
        scsi1 to scsi4
        scsi2 to scsi5

Saving the original configuration file to CLU14.asesave

The configuration file CLU14 has been updated.

Creating necessary devices:

MAKEDEV: special file(s) for rz32:
rz32a rrz32a rz32b rrz32b rz32c rrz32c rz32d rrz32d rz32e rrz32e rz32f rrz32f
rz32g rrz32g rz32h rrz32h
MAKEDEV: special file(s) for rz33:
rz33a rrz33a rz33b rrz33b rz33c rrz33c rz33d rrz33d rz33e rrz33e rz33f rrz33f
rz33g rrz33g rz33h rrz33h
.
.
.
```

3.7.3 Changing References to Renumbered Devices in System Files

The ase_fix_config script changes the numbers assigned to the SCSI buses in the system configuration file. If any of the renumbered devices contain file systems or filesets, edit the following system files before booting the new kernel:

- For a UNIX file system (UFS), manually edit the /etc/fstab file to correct any references to the renumbered devices. Using the previous example, the original /etc/fstab file entry is as follows:

  ```
  /dev/rz26c /usr/users2 ufs rw,groupquota
  ```

  After the bus renumbering, you must change the entry to the following:

  ```
  /dev/rz34c /usr/users2 ufs rw,groupquota
  ```

- For an Advanced File System (AdvFS) fileset, you must correct the symbolic links in the/etc/fdmns/*domainname* directories to point to the new device names. For example, if /dev/rz26c is assigned to the AdvFS domain users, then the following symbolic link exists:

  ```
  /etc/fdmns/users/rz26c -> /dev/rz26c
  ```

  After the bus renumbering, you must change the link to the following:

```
/etc/fdmns/users/rz34c -> /dev/rz34c
```

If SCSI bus renumbering done during the installation causes the names of disks already given to the Logical Storage Manager (LSM) to change, you must reconfigure the LSM metadata. For disk groups other than the `rootdg` disk group, perform the following steps after rebooting the system but before using the `asemgr` utility to add volumes in the affected disk groups to ASE services.

Note

Disks that are placed in the `rootdg` disk group must be on a private SCSI bus. The commands in the following steps do not apply to the `rootdg` disk group.

1. Enter a `voldisk define` command using the new disk name for all disks whose names change due to the bus renumbering. For example, if device `rz8` becomes `rz16` and `rz9` becomes `rz17`, enter the following commands:

   ```
   # voldisk define rz16
   # voldisk define rz17g
   # voldisk define rz17h
   ```

2. Import the affected disk group:

   ```
   # voldg import dgname
   ```

3. Perform recovery for the disk group:

   ```
   # volrecover -sb
   ```

   You are now ready to add volumes contained in this disk group to services created with the `asemgr` utility.

If you do not make these changes before you boot the new kernel, the file systems or filesets will be unavailable, but the data is not lost.

3.8 Build and Install a New Kernel (PS, AS, MC)

The `doconfig` program names the new kernel `/sys/`*`filename`*`/vmunix`, where *`filename`* is the name of the configuration file you specified when you configured the cluster kernel components (see Section 3.6).

If the kernel build is successful, the name of the new kernel file is displayed as follows:

```
Working....Mon Apr 14 13:53:22 EDT 1997
Working....Mon Apr 14 13:55:22 EDT 1997
```

```
        Working....Mon Apr 14 13:57:29 EDT 1997

The new kernel is /sys/CLU14/vmunix
```

If the kernel build is not successful, see the troubleshooting information in Section 6.2.

When the kernel build is successful, the installation procedure displays a list of instructions. The following example shows the instructions for a Production Server installation:

```
The kernel build was successful. Please perform the following actions:

        o Move the new kernel to /.
        o Before rebooting make sure that the cluster interconnect IP
          addresses for all cluster members are recorded in each member's
          /etc/hosts file.
        o Reboot the system.
```

Note

The displayed instructions indicate that you reboot the system after verifying that all required host entries are in `/etc/hosts`. Depending on your installation type, follow the task steps listed in Table 3–1 or Table 4–1, and reboot at the indicated steps. The tasks are ordered to minimize the number of reboots required during installation.

The installation procedure does not automatically move the new kernel to the root directory. You can rename the new kernel or save the existing kernel before manually moving the new kernel to the root directory.

Before moving the original kernel aside and copying the new one to the root directory, use the `df` command to check that there is enough disk space for both files.

Move the new kernel to the root directory. In the following example, the old kernel is saved as `vmunix.save` and the new kernel, `/sys/CLU14/vmunix`, is moved to the root directory:

```
# cp /vmunix /vmunix.save
# mv /sys/CLU14/vmunix /
```

After you verify the proper operation of the new kernel, you can remove the old kernel (called `vmunix.save` in this example). DIGITAL recommends that you keep a kernel that does not contain cluster support (for example, `/genvmunix`).

3.9 Add Host Entries to /etc/hosts (PS, AS, MC)

To avoid relying on name servers for critical operations, make sure that each system has the IP names and addresses of all other systems that are in the cluster or ASE. If you have not yet added these names and addresses to `/etc/hosts`, see Section 2.1 and do so before booting the new kernel.

3.10 Enable New DLM Interfaces and Reboot the System (PS)

When performing a Production Server full installation or a simultaneous upgrade, you can enable new distributed lock manager (DLM) interfaces and boot the new kernel as a single operation. Once you have enabled DLM features on one system and booted that system, do not boot systems that do not have the new DLM features enabled. This section assumes that you are either installing for the first time, or that you are performing a simultaneous upgrade and that no system is booted until it is upgraded and new DLM features are enabled.

Caution

Do not run the `dlm_enable` script during a rolling upgrade. However, after you install Production Server Version 1.5 on all cluster members, you can run the `dlm_enable` script and reboot each system in turn. Section 4.4 describes how to enable new DLM interfaces following a rolling upgrade.

Before running the `dlm_enable` script on the first system, make sure the power to the MEMORY CHANNEL hub is turned on. Turning on the power to the MEMORY CHANNEL hub after a cluster member has booted does not result in the formation of a cluster.

To enable the new DLM features and reboot the system, follow these steps :

1. Run the `dlm_enable` script:

   # **`/usr/sbin/dlm_enable`**

2. The script displays:

   ```
   Would you would like to reboot the system at this time (y/n)? [n]:
   ```

   Enter **`y`**. The script then runs the `shutdown -r now` command.

During the reboot, startup messages are displayed on the console. Example A–1 shows the startup messages following a successful Production Server installation.

To check the version of the installed software, query the value of the `cluster_version` attribute in the `/etc/sysconfigtab` file by entering the following command:

```
# sysconfig -q clubase cluster_version
clubase:
cluster_version = DIGITAL UNIX TruCluster V1.5-n  (Rev. nnn); 11/10/97 17:47
```

3.11 Reboot the System (AS, MC)

To reboot the system, enter the following command:

```
# shutdown -r now
```

During the reboot, startup messages are displayed on the console. Example A–1 shows the startup messages following a successful Production Server installation.

To check the version of the installed software, query the value of the `cluster_version` attribute in the `/etc/sysconfigtab` file by entering the following command:

```
# sysconfig -q clubase cluster_version
clubase:
cluster_version = DIGITAL UNIX TruCluster V1.5-n  (Rev. nnn); 11/10/97 17:47
```

3.12 Populate the Member List in the ASE Database (PS, AS)

If this system will be part of an ASE, and you are not using an existing ASE database, the installation procedure initialized an ASE database on this system with an empty member list. If you performed a simultaneous upgrade and are reusing a saved ASE database, you must still populate the member list.

After all systems are installed, run the `asemgr` utility on one system in each ASE to populate the member list with the IP names of ASE members. On the selected member system, enter the following command to invoke the `asemgr` utility:

```
# asemgr
```

When the `asemgr` utility prompts for a list of the IP names of the ASE members, enter the name of each member's interface using the names you specified during the installation. Use a comma to separate each name. (If you specify a fully qualified IP name, the `asemgr` utility shortens the name to its unqualified alias, that is, excluding the domain portion of the name. The installation procedure automatically adds the unqualified alias, if needed, when it adds the system's interface name to the `/etc/hosts` file.)

The asemgr utility then displays the list of member names that you specified. When you confirm that the list is correct, the asemgr utility displays the ASE Main Menu. For example:

```
Enter a comma-separated list of member systems' interfaces on the cluster
interconnect within ASE 1.
Enter Members: mcclu13, mcclu14

Member List: mcclu13, mcclu14

Is this correct (y/n) [y]:  Return

               ASE Main Menu

    a)  Managing the ASE         -->
    m)  Managing ASE Services    -->
    s)  Obtaining ASE Status     -->

    x)  Exit                            ?)  Help

Enter your choice: x
```

See the TruCluster Software Products *Administration* manual for information on setting up ASE services.

3.13 Create Consistent Device Special Files for an ASE Tape Service (PS, AS)

If you intend to configure a highly available tape service in an ASE, make sure that the raw tape device special files (those files with the rmt prefix) are named consistently on all members of that ASE. If all tape devices are on the shared SCSI bus, each member will have identical device special files for them, and you need do nothing more. However, if some members also have private tape devices, the device special files for the tape devices may be given names that are different than those used by other ASE members.

This difference is caused by the bus probe code that runs as part of the base operating system installation. The code scans private buses before shared buses, creating physical tape tz device special files as it encounters tape devices. The conversion of the tz device special files to rmt device special files, which occurs at the first boot of the new kernel, follows the order of the tz device names.

To remedy this situation, you must first list the physical tape devices from each ASE member and determine which are on the shared SCSI bus. To do so, use the scu utility with the following commands:

```
# scu
scu> scan edt
Scanning all available buses, please be patient...
scu> show edt
```

You will see listings similar to those in the following example:

```
CAM Equipment Device Table (EDT) Information:

    Device: RZ28       Bus: 0, Target: 0, Lun: 0, Type: Direct Access
    Device: RZ26       Bus: 0, Target: 1, Lun: 0, Type: Direct Access
    Device: RZ26       Bus: 0, Target: 2, Lun: 0, Type: Direct Access
    Device: RRD45      Bus: 0, Target: 6, Lun: 0, Type: Read-Only Direct Access
    Device: RZ28M      Bus: 1, Target: 1, Lun: 0, Type: Direct Access
    Device: RZ28M      Bus: 1, Target: 2, Lun: 0, Type: Direct Access
    Device: DEC  N01   A10 Bus: 1, Target: 6, Lun: 7, Type: Processor
    Device: TZ887      Bus: 2, Target: 5, Lun: 0, Type: Sequential Access
    Device: TZ Media   Changer Bus: 2, Target: 5, Lun: 1, Type: Medium Changer
    Device: DEC  N01   A10 Bus: 2, Target: 6, Lun: 7, Type: Processor
```

In this example, TZ887 is the hardware name of a tape device on a shared SCSI bus. If you are not sure which SCSI buses are shared, run the ase_fix_config utility. (The doconfig program, when called by the TruCluster software installation procedure, automatically runs the ase_fix_config utility. You also run the ase_fix_config utility each time there are changes in the SCSI configuration of an ASE.) The utility provides information similar to the following:

```
Select the controllers that define the shared ASE I/O buses.

        Name    Controller      Slot    Bus     Slot
     )  scsi0   psiop0          0       pci0    1
    1)  scsi1   pza0            0       pci0    7
    2)  scsi2   pza1            0       pci0    8

    q)  Quit without making changes

Enter your choices (comma or space separated): q
```

To determine the device special file names for each tape device, enter the following command:

```
# ls -l /dev/rmt*
```

You will see listings similar to the following:

```
crw-rw-rw-   1 root     system     9,37894 Aug  1 11:24 rmt0a
crw-rw-rw-   1 root     system     9,37890 Aug  1 11:24 rmt0h
crw-rw-rw-   1 root     system     9,37888 Aug  1 11:24 rmt0l
crw-rw-rw-   1 root     system     9,37892 Aug  1 11:24 rmt0m
```

As you enter these commands on each system, make sure that the device special file name of each highly available tape device is the same on each member and that the device's major and minor number, listed in column 5 of the display, match on each member. You need compare only one of the listed files for each tape: for instance, rmt0h in this example.

On any member with a private tape device, you need to delete the existing rmt files and then rebuild them, starting with those for the highly available tape devices. After you delete the files, rebuild them as follows:

- Examine the output of the scu commands previously listed.

- Compute the tape device's unit number from its SCSI address according to the following formula:

  ```
  unit_number = ( 8 * bus_number ) + target_id
  ```

  In the example of the TZ887 device, this resolves to:

  ```
  unit_number = (8 * 2) + 5
  unit_number = 21
  ```

  (See mc(7), distributed with the SCSI CAM Layered Components kit, for additional help on this step.)

- Use the MAKEDEV script to create the device special files, supplying the device unit number obtained in the previous step. For example:

  ```
  # ./MAKEDEV tz21
  ```

3.14 Specify a Tie-Breaker Disk for Virtual Hub Mode (PS)

A two-system cluster can use a virtual hub connection between the systems to provide access to MEMORY CHANNEL hardware. However, if the MEMORY CHANNEL cable is disconnected from either system, this type of cluster configuration can partition (that is, divide into two one-system clusters, each running its own connection manager's director daemon, cnxmgrd).

Caution

If not detected, a cluster partition can result in unpredictable behavior and corruption of shared data.

In order for the connection manager to detect (and remedy) a cluster partition, you must specify at least one tie-breaker disk to which both systems have access. When a virtual hub is in use and a tie-breaker disk is defined, the cluster requires that either both systems, or one system and the disk, must be present for the cluster to run or continue to run.

Note

In virtual hub mode, the connection manager and the distributed lock manager (DLM) will not function until a tie-breaker disk is configured.

A device specified as a tie-breaker disk remains available to the service that manages it. When there is only one system remaining in a two-system cluster, the connection manager asks the ASE availability manager driver

about the reservation status of the device. When the connection manager determines that there is a partition, the remaining member system is shut down.

Note

When performing a full installation or a simultaneous upgrade, you can configure a tie-breaker disk on each system after installing the latest version of the Production Server software on both systems.

When performing a rolling upgrade, you must configure a tie-breaker disk on each system after upgrading it to the latest version of the Production Server software. (When you deinstall the old Production Server software, the CNX_DISK variable is removed from `/etc/rc.config`. If you do not configure a tie-breaker disk after upgrading the first system, it will be unable to maintain the cluster when you halt the second system.)

Use the `cnxshow` command to determine whether or not a tie-breaker disk is required and, if required, whether or not the tie-breaker disk has been configured. For example, after installing Production Server, but prior to configuring a tie-breaker disk, the `cnxshow` command displays the following information:

```
# cnxshow

Cluster View from mcclu14

Director: Unknown   Suspended: Yes

Node monitor using tie-breaking disk: Required but not defined

The node monitor and the local node specify different disks.
The local node specifies disk : Unknown

Hostname      Cluster I/F   CS_ID      Incarnation          Comm Okay  Member
------------------------------------------------------------------------------
clu13         mcclu13       0000,0000  0000000000004cf70        No       ?
clu14         mcclu14       0000,0000  00000000000305f0         No       ?
```

(See `cnxshow`(8) for more information on the `cnxshow` command.)

If your cluster uses a virtual hub, follow these steps to set up and define a tie-breaker disk:

1. Run the `clu_ivp` utility to ensure that the ASE is properly configured. (See Chapter 5 for information on the `clu_ivp` utility.)
2. On one system, use the `asemgr` utility to set up at least one ASE service that can run on either system and that includes at least one

participating disk. For example, you might set up a disk service to provide a highly available mail service. Include at least one disk in this service.

The ASE service that manages the tie-breaker disk can be a Network File System (NFS) service, a distributed raw disk (DRD) service, a disk service or a tape service (that specifies a disk).

Note

Do not restrict the automatic service placement (ASP) policy of this service.

For information on how to set up an ASE service, see the TruCluster Software Products *Administration* manual.

3. To ensure that the service is known to the ASE availability manager driver and properly configured on the other system, use the `asemgr` utility to manually relocate the service to the other system. For example, to relocate the service `ase_nfs_1` to system `mcclu13`, enter the following command:

   # **asemgr -m ase_nfs_1 mcclu13**

4. On both systems, run the `cnxset` command, specifying the name of at least one raw disk device participating in the ASE service. For example:

   # **cnxset -d /dev/rrz10c**

```
Verifying: /dev/rrz10c
     /dev/rrz10c is being watched by the local ASE
     /dev/rrz10c is an acceptable device

Local configuration updated...You must also update all other nodes
```

Note

You must specify a raw disk device name. Do not specify the name of a Logical Storage Manager (LSM) volume or an Advanced File System (AdvFS) fileset.

The `cnxset` command checks that the name you specify has these characteristics:

- Is a reference to a physical disk
- Is a character device

- Is known to the ASE availability manager driver

The cnxset command places this device name in the CNX_DISK variable in the /etc/rc.config file and notifies the connection manager that a configuration change has occurred. You can verify the setting of the CNX_DISK variable by entering the following command:

```
# rcmgr get CNX_DISK
/dev/rrz10c
```

Note

If the device you specify cannot be verified as a physical disk or as a character device, the cnxset utility adds the device name to the /etc/rc.config file but the connection manager might not function.

Before using the cnxset -d command, it is recommended that a tie-breaker disk belong to an ASE service, the ASE service is online, and the service has run on both systems (manually relocating the ASE service to the other system, if necessary, as explained in step 3, which shows how to relocate an ASE service). However, the cnxset command lets you configure a tie-breaker disk before setting up ASE services. If you do so, make sure to set up an ASE service that uses the disk as soon as possible.

In the following example, device /dev/rrz41c is unknown to the ASE availability manager driver on the current system:

```
# cnxset -d /dev/rrz41c

Verifying: /dev/rrz41c
     /dev/rrz41c is unknown to the local ASE
     Performing further checks on this device...
       /dev/rrz41c is a disk
       /dev/rrz41c will be configured anyway
     /dev/rrz41c is an acceptable device

Local configuration updated...You must also update all other nodes
```

One of the following conditions is likely to produce these messages:

- Device /dev/rrz41c has not yet been added to an ASE service.
- Device /dev/rrz41c has been added to an ASE service, but the service is off line.

In the following example, the cnxset utility is unable to open device /dev/rrz1c:

```
# cnxset -d /dev/rrz1c

Verifying: /dev/rrz1c
     /dev/rrz1c is unknown to the local ASE
```

```
    Performing further checks on this device...
      /dev/rrz1c can not be opened. Unable to verify device
      /dev/rrz1c will be configured anyway
    /dev/rrz1c is an acceptable device

Local configuration updated...You must also update all other nodes
```

One of the following conditions is likely to produce these messages:

- The ASE service in which device /dev/rrz1c participates is running on the other system; therefore, that system has a reservation on the device.
- The ASE service to which device /dev/rrz1c belongs has not run on the current system.

The cnxset utility is unable to verify a device in these cases. Since the specified device appears to be a disk (its name is of the correct form), the cnxset utility accepts the specification and configures the device as a tie-breaker disk.

You can specify up to three raw devices as tie-breaker disks. By defining multiple tie-breaker disks, you eliminate a disk as a single point of failure when the cluster is operating with only one system. For greater reliability, DIGITAL recommends that each tie-breaker disk participate in a different ASE service. In this way, you can take one of these services off line without affecting the tie-breaker capability. You must manually relocate each such ASE service before defining the tie-breaker disks.

Use a single cnxset -d command to define all tie-breaker disks. Use commas (no spaces) to separate raw device names on the cnxset command line. For example:

```
# cnxset -d /dev/rrz10c,/dev/rrz11c,/dev/rrz12c
```

Note

When two tie-breaker disks are specified, the connection manager requires that one member, plus both disks, be present. Any other policy would risk a partition involving two “one member and one tie-breaker” configurations. The two tie-breaker disk configuration introduces another risk (that is, either of the tie-breaker disks failing would bring down the cluster), so it is not very useful.

When three tie-breaker disks are specified, the connection manager requires that one member, plus two disks, be present. This is the optimal configuration for systems that require no single point of failure.

Each time you use the cnxset -d command, any previously defined tie-breaker disks are replaced with the new definitions.

To remove all tie-breaker disk definitions from the `/etc/rc.config` file, use the `cnxset -D` command.

For more information, see `cnxset`(8).

4

Upgrade Installation

The TruCluster upgrade procedures assume that you are performing an update installation of the DIGITAL UNIX operating system, and thereby preserving existing TruCluster configuration and database files. If you perform a full installation of the DIGITAL UNIX operating system, you must perform a full installation of the TruCluster product.

Note

Some versions of TruCluster products are supported on more than one version of the DIGITAL UNIX operating system. If you are updating only the operating system, see Section 4.6.

Table 4–1 summarizes the upgrade installation tasks. It lists the tasks in order, indicates the TruCluster products to which each task applies, and provides pointers to necessary information. The column headings *R* and *S* refer to rolling and simultaneous upgrades. The section titles in this chapter also use the *R* convention to indicate when a section applies only to a rolling upgrade.

Both full and upgrade TruCluster installations use the same installation script. Because both installation types share several steps, Table 4–1 often refers to sections in Chapter 3.

Table 4–1: Upgrade Installation Tasks

Task	Production Server		Available Server		See:
	R	S	R	S	
Read the introductory chapters.	X	X	X	X	Chapters 1 and 2
Read the rolling upgrade restrictions.	X		X		Section 4.1
Perform preinstallation tasks.	X	X	X	X	Table 2–1
Load the TruCluster kit. The installation procedure starts automatically when you load the kit.	X	X	X	X	Section 3.1
Specify the IP name and address (`CLUSTER_NET`) of the cluster interconnect (PS) or the member network interface (AS).	X	X	X	X	Section 2.1 and Section 3.2
Specify an ASE identifier (`ASE_ID`).	X	X			Section 3.3
Decide whether to run the ASE logger daemon on this system.	X	X	X	X	Section 3.4
Initialize the ASE database.	X	X	X	X	Section 4.2
Select a kernel configuration file.	X	X	X	X	Section 3.6
Build and install a new kernel.	X	X	X	X	Section 3.8
Add host entries to `/etc/hosts`.	X	X	X	X	Section 3.9
Enable new DLM interfaces and reboot the system.		X			Section 3.10
Reboot the system.	X		X	X	Section 3.11
Return the member to the ASE.	X		X		Section 4.3
After all systems are upgraded, if using a saved ASE database, return the member to the ASE.		X		X	Section 4.3
After all systems are upgraded, if not using a saved ASE database, populate the member list in the ASE database.		X		X	Section 3.12
Specify a tie-breaker disk (`CNX_DISK`) (only for a two-system, virtual-hub cluster).	X	X			Section 3.14
After all systems are upgraded, enable new DLM interfaces and reboot each system.	X				Section 4.4
After all systems are installed, create consistent device special files for an ASE highly available tape service.	X	X	X	X	Section 3.13
After all systems are upgraded, restart ASE services.		X		X	Section 4.3

Table 4–1: Upgrade Installation Tasks (cont.)

Task	Production Server		Available Server		See:
	R	S	R	S	
After all systems are upgraded, use `asemgr` to turn on new features.	X		X		Section 4.5
After all systems are upgraded, run `clu_ivp` to verify the installation.	X	X	X	X	Chapter 5

4.1 Rolling Upgrade Restrictions (PS-R, AS-R)

Before performing a rolling upgrade, note the following restrictions:

- As a general note, do not attempt to use new features until all members are upgraded. For example, do not enable new DLM interfaces (Section 4.4) until the entire cluster is upgraded.
- The extended UIDs/GIDs feature is disabled by default in the DIGITAL UNIX operating system. See the DIGITAL UNIX documentation for information on enabling and using this feature.
- During a rolling upgrade, you should not run the Cluster Monitor (`cmon`) utility until all cluster members are upgraded. If you must use the Cluster Monitor utility during a rolling upgrade, apply Patch TCR141-006 to all Version 1.4 (Available Server) or Version 1.4A (Available Server or Production Server) member systems before starting the rolling upgrade. Follow the installation instructions that accompany the patch.

Note

If you are installing this patch on TruCluster Available Server Software Version 1.4, copy the patched `tractd` file to `/usr/opt/TCR140/sbin/tractd`. The `TCR141` subdirectory included in the target file path in the patch's installation instructions is correct only for a Version 1.4A installation.

- During a rolling upgrade, do not create or modify an ASE service that uses the Logical Storage Manager (LSM) on a system that is running DIGITAL UNIX Version 4.0D. Creating or modifying LSM services on a system running DIGITAL UNIX Version 4.0D may result in these services not being properly relocated. After upgrading all cluster members to DIGITAL UNIX Version 4.0D, you can create or modify LSM services on any member.

- When you delete an ASE service that uses Advanced File System (AdvFS) filesets or LSM volumes, the storage configuration information is retained on the system. Therefore, you should delete AdvFS or LSM services on a system that has already been upgraded to DIGITAL UNIX Version 4.0D to avoid the possibility of overwriting the information when installing the operating system during the rolling upgrade procedure.
- During a rolling upgrade, you cannot use the audit feature on file system objects (for example, files or directories) that are located on shared devices.
- The MEMORY CHANNEL API library will not operate in a cluster in which member systems are running different versions of the TruCluster software.

  When a member boots into a Production Server cluster, an attempt is made to initialize the MEMORY CHANNEL API (because the installation procedure sets IMC_AUTO_INIT=1 in `/etc/rc.config`). This initialization is transparent when all members of the cluster are running the same version of the Production Server software. However, when a Production Server Version 1.5 system boots into a cluster where one or more members are running earlier versions, there is a MEMORY CHANNEL API version incompatibility.

  In this case, the initialization procedure generates `imc_init:` and `mcs_IMC_init:` error messages indicating a version incompatibility, and the MEMORY CHANNEL API initialization fails. Because the Production Server does not rely on the MEMORY CHANNEL API for cluster operations, you can usually ignore these messages. To initialize the MEMORY CHANNEL API after all members are upgraded to Version 1.5, run the `/usr/sbin/imc_init` command on each member.

  However, this initialization behavior has some important consequences for any highly available services that use the MEMORY CHANNEL API.

  Note

  The following restriction applies *only* to customers who have created a highly available service that uses the MEMORY CHANNEL API library routines.

  The rolling upgrade process consists of booting newly installed Version 1.5 members into an existing cluster. When each of these members is booted, its MEMORY CHANNEL API library initialization fails as noted earlier, leaving only the remaining Version 1.4A members capable of running any highly available services that use the MEMORY CHANNEL

API library routines. At some point in the rolling upgrade, you must stop these services in order to complete the upgrade of the last Version 1.4A system to Version 1.5.

To minimize the downtime of these services, perform the following procedure:

1. Use the `asemgr` utility to relocate any services using the MEMORY CHANNEL API library routines to a Version 1.4A system.
2. Upgrade all other systems to Version 1.5 of the Production Server.
3. On the remaining Version 1.4A system, stop all services that use the MEMORY CHANNEL API library.
4. Halt the Version 1.4A system.
5. Run the `/usr/sbin/imc_init` command on all Version 1.5 systems.
6. Use the `asemgr` utility on a Version 1.5 system to restart the service.
7. Upgrade the remaining Version 1.4A system to Version 1.5.

- The upgrade procedure is not reversible. To return to an earlier version of the operating system and TruCluster software after DIGITAL UNIX Version 4.0D and TruCluster software are installed, you must deinstall the TruCluster software subsets, and then reinstall the operating system and the earlier version of TruCluster software.

Note

DIGITAL recommends that you complete the rolling upgrade procedure as quickly as possible and minimize any hardware and software changes during the upgrade. For example, if you want to upgrade the member systems and also remove a member system, delete the member system before you perform the upgrade. If you want to add a member system, perform the upgrade and then add the new member.

4.2 Initialize the ASE Database (PS, AS)

If you did not save the ASE database when deinstalling the TruCluster subsets, the installation procedure automatically creates and initializes a new ASE database, `/var/ase/config/asecdb`. However, if the installation procedure detects an existing ASE database, it asks whether you want to use this database:

```
An old ASE database file has been found. Do you want to use this (y/n):
```

If you answer **n**, the installation procedure deletes the database and creates a new one.

- For a rolling upgrade, answer **n**. You want the existing ASE to repopulate this system when it is returned to the ASE.
- For a simultaneous upgrade, answer **y** if you saved the ASE database when deinstalling the TruCluster subsets and want to reuse that database. After all systems are upgraded, you must run the `asemgr` utility to reinitialize the member list before the ASE services in the saved database can be started. The following message is displayed to remind you to reinitialize the member list before starting ASE services:

  ```
  Note: You will need to run asemgr to reinitialize the member list before the
  saved ASE services can be started.
  ```

4.3 Return the System to Its ASE (PS, AS)

To return the member to the ASE, follow the steps for your type of upgrade.

For a rolling upgrade:

- Run the `asemgr` utility on an existing ASE member system and add the upgraded system to the ASE.
- If the upgraded member's name was removed from the member list in any ASE service's automatic service placement (ASP) policy when you deleted the system from the ASE, use the `asemgr` utility to stop the ASE service and add the member back to the ASP policy member list.

For a simultaneous upgrade using a saved copy of the ASE database:

- Check that all systems are upgraded.
- Run the `asemgr` utility on *one member system only* in each ASE to reinitialize the member list.
- If you changed your configuration by removing a system from the ASE, and you are reusing an existing ASE database, the `asemgr` utility checks the ASP policies of services for the deleted member name. You are asked to confirm whether or not to remove the deleted member name from the service's ASP policy. If deleting the member name results in an empty list:
  - A service with the Favored Member policy effectively becomes one with the Balanced policy.
  - A service with the Restricted to Member policy cannot be started until its ASP policy is modified to specify an existing member.
  - A service with the Balanced policy is not affected.

4.4 Enable New DLM Interfaces and Reboot (PS-R)

Complete a rolling upgrade of all systems before enabling the new distributed lock manager (DLM) interfaces. After all systems are upgraded, run the `dlm_enable` script on each system.

Caution

After the new DLM interfaces are enabled on a cluster member, the DLM will panic if any cluster member is running a version of TruCluster software prior to Version 1.5, or if a Version 1.5 member attempts to use these new interfaces to communicate with another Version 1.5 member that does not have the interfaces enabled. Once you enable the new features on one member, make sure to enable these features on all cluster members before running an application that uses the new DLM interfaces.

To enable the new DLM features and reboot, follow these steps on each system in turn:

1. Run the `dlm_enable` script:

   ```
   # /usr/sbin/dlm_enable
   ```

2. The script displays:

   ```
   Would you would like to reboot the system at this time (y/n)? [n]:
   ```

   Enter **y**. The script then runs the `shutdown -r now` command.

4.5 Enable New ASE Features (PS-R, AS-R)

Complete a rolling upgrade of all systems before enabling new ASE features. After all system are upgraded, enable new ASE features on one member in each ASE.

Note

Because there are no software checks that prevent you from enabling new features before all systems are upgraded, make sure that all systems are upgraded to the latest version of the TruCluster product before continuing.

To enable new features, run `asemgr` and enable the features by selecting the `f` menu option, Enable ASE V1.5 functionality. Once you enable new

features, this menu option is disabled (the `f` is no longer visible). The following example shows the menu option with the `f` option enabled:

```
# asemgr

        TruCluster Production Server (ASE)

                ASE Main Menu

    a)  Managing the ASE         -->
    m)  Managing ASE Services    -->
    s)  Obtaining ASE Status     -->

    x)  Exit                             ?)  Help

Enter your choice: a

                Managing the ASE

    a)  Add a member
    d)  Delete a member
    n)  Modify the network configuration
    m)  Display the status of the members
    C)  Display the configuration of the ASE database
    l)  Set the logging level
    e)  Edit the error alert script
    t)  Test the error alert script
    f)  Enable ASE V1.5 functionality

    q)  Quit (back to the Main Menu)
    x)  Exit                             ?)  Help

Enter your choice: f
```

4.6 How to Update the DIGITAL UNIX Operating System Without Upgrading the TruCluster Software (PS, AS)

If the following conditions are met, you can update the DIGITAL UNIX operating system without upgrading the installed version of the TruCluster product:

- Both versions of the operating system support the installed version of the TruCluster product.
- The newer version of the operating system supports update installation.

The DIGITAL UNIX *Release Notes* provide information on the supported update paths for the operating system and the supported versions of TruCluster products.

When you update the DIGITAL UNIX operating system on an existing TruCluster member system, but are not upgrading the TruCluster software, you do not need to deinstall the TruCluster software kit before proceeding with the base system update. Correspondingly, you do not need to reinstall the TruCluster software kit after you have completed the update of the operating system. The current ASE configuration database (`/var/ase/config/asecdb`) is automatically preserved during an update of the operating system.

Note

You must update the operating system on all cluster or ASE members. At the end of the update, all systems must be running the same version of the DIGITAL UNIX operating system.

To update the operating system underlying the TruCluster software, perform the following procedure on each member system:

1. Disable the TruCluster software by adding or editing the `clubase` stanza entry in `/etc/sysconfigtab` to include the following lines:

   ```
   clubase:
               cluster_disable=1
   ```

2. Shut down and halt the system using a command similar to the following:

   ```
   #shutdown -h +30 Please log out
   ```

3. From the console prompt, invoke the `init` command and boot `/genvmunix` to single-user mode; for example:

   ```
   >>> init
   >>> boot -fi genvmunix -fl s
   ```

4. Update the operating system. See the DIGITAL UNIX *Release Notes* and the DIGITAL UNIX *Installation Guide* for information about updating the operating system. See Section 1.3 for information about the operating system subsets required for the TruCluster product.

5. Reenable the TruCluster software by setting the value of the `cluster_disable` attribute to `0` in the `/etc/sysconfigtab` file:

   ```
   clubase:
               cluster_disable=0
   ```

6. Reboot the system.

The procedure is complete after you update the last member's operating system.

5

Verifying the Installation with the clu_ivp Utility

After all systems are installed or upgraded, use the cluster installation verification program, `clu_ivp`, to detect configuration errors. The `clu_ivp` utility can test the installation of the Production Server, Available Server, or MEMORY CHANNEL products; the type of tests performed depend on the product.

Note

Because the utility tests only the system on which it is run, you must run `clu_ivp` on each system to verify the basic configuration of a Production Server cluster or an Available Server configuration.

As part of checking the installation configuration, `clu_ivp` calls two other utilities: `cnxshow` and `drd_ivp`. See `clu_ivp`(8), `cnxshow`(8), and `drd_ivp`(8) for more detailed information on each utility. Use these utilities and the information in this chapter when verifying an installation. These utilities do not modify the system in any way. You can run them at any time to diagnose problems.

By default, the `clu_ivp` utility displays error conditions only. For example:

```
# /usr/sbin/clu_ivp
Cluster Installation Verification Procedure (IVP)
ERROR: asemgr has not been run to specify the member names.
Resolution: run 'asemgr' to specify the member names.
ERROR: drd_ivp validation tests failed.
Resolution: run 'drd_ivp -p -v -c' for more details.

Installation verification failed, 2 errors detected.
```

When an error is detected, the `clu_ivp` utility suggests corrective action. In some cases, the error reported by the `clu_ivp` utility is the symptom of another problem. Read all the error messages generated by the `clu_ivp` utility before attempting to correct problems. When the corrective action suggested by the `clu_ivp` utility does not solve the problem, examine the system's error log files and console output for additional clues.

Note

Following a first-time Production Server installation before any distributed raw disk (DRD) services are configured, you might see `drd_ivp` errors. With the exception of `drd_ivp validation tests failed` type messages, these errors are generally benign.

For more informative output, use the `clu_ivp -v` (verbose) option. In addition to reporting error conditions, the utility displays confirmation of each verification check as it is performed.

When run in verbose mode, the `clu_ivp` utility looks for configured controllers and SCSI buses that are eligible for sharing. It finds controllers that have records in the `/sys/conf/HOSTNAME` file; these records are created by booting the `/genvmunix` kernel and running the `doconfig` program. The following sample output from `clu_ivp` shows information on SCSI controllers and devices:

```
Found these supported SCSI controllers:
-----------------------------
  pza0 is scsi2
  pza1 is scsi3
  pza2 is scsi4
Scanning EDT.........
Found these devices on bus scsi2:
  device              vendor   drive            firmware
  ------              ------   -----            --------
  rrz18c              DEC      RZ26N            0616
  rrz20c              DEC      RZ28             442D
Found these devices on bus scsi3:
  device              vendor   drive            firmware
  ------              ------   -----            --------
  rrz27c              DEC      RZ28             D41C
  rrz29c              DEC      RZ26             392A
Found these devices on bus scsi4:
  device              vendor   drive            firmware
  ------              ------   -----            --------
  rrz34c              DEC      RZ28B            0006
  rrz36c              DEC      RZ26N            0616
```

If the `clu_ivp` utility cannot find any controllers, one possibility is that the `doconfig` program was not run under `/genvmunix` since those controllers were installed. Because the `clu_ivp` utility scans the shared SCSI buses directly to detect disk devices connected to these buses, there is

no need to run the doconfig program after adding a disk to a shared SCSI bus to make the disk visible to the clu_ivp utility.

To check that all members in an ASE have the same view of shared devices (same bus numbers, same device numbers), run clu_ivp -v on each system and compare the results.

You can also use the file command to verify that other systems connected to the shared bus see the same devices on the shared buses. In the following example, the output of the file command on system clu14 indicates that device rz18 is on the same bus (number 2) as on system clu13:

```
# file /dev/rrz18c
/dev/rrz18c: character special (8/34818) SCSI #2 RZ26N disk \
             #144 (SCSI ID #2) (SCSI LUN #0)
```

Note that because of device reservations, the file command reports errors after an ASE service is started on a disk if the command is invoked on a system that is not running the service.

The clu_ivp utility cannot tell whether your SCSI cabling is correct. If you have any questions about cabling requirements, see the TruCluster Software Products *Hardware Configuration* manual.

6

Troubleshooting

This chapter describes the following problems, which you might encounter during installation and suggests corrective actions:

- Setting logging levels (PS, AS, MC)
- Kernel build fails (PS, AS, MC)
- Cannot ping members across the primary network (PS, AS)
- MEMORY CHANNEL cables are crossed (PS)
- System cannot join the cluster (PS)
- ASE validation fails (PS, AS)
- The `drd_ivp` utility cannot determine ASE membership (PS)
- Inconsistent view of shared SCSI devices (PS, AS)

6.1 Setting Logging Levels (PS, AS, MC)

For Production Server and Available Server, you can set the `asemgr` logging level to `Informational`, which increases the amount of messages written to `/var/adm/syslog.dated/`*`date`*`/daemon.log`.

For Production Server and MEMORY CHANNEL, you can use the `mchan_debug` attribute in the `/etc/sysconfigtab` file to generate verbose MEMORY CHANNEL error messages. Set the attribute as shown in the following example:

```
mchan:
        mchan_debug=1
```

You must reboot the system in order for the `mchan_debug` change to take effect. The additional debug information, when included in a problem report, can help your DIGITAL service representative diagnose problems.

6.2 Kernel Build Fails (PS, AS, MC)

After prompting for configuration options, the installation procedure attempts to build a new kernel using the `doconfig` utility. If the newly configured kernel cannot be built, the installation procedure displays the following message:

```
*** WARNING ***
An error has occurred during system configuration.  A partial listing
of the error log file (./errs) follows:
.
.
.
*** NOTE ***
The customized kernel for this machine could not be successfully
created.  One possible problem could be kernel layered products
that might be incompatible with the operating system.  This
script will now automatically attempt to build a kernel using the
operating system only.
Is this ok? (y/n) [y]:
```

If the rebuild is still unsuccessful, the installation procedure displays the following message:

```
*** NOTE ***
A new kernel for this machine could not be successfully created.

Unable to build the new kernel. Please perform the following actions:

        o Run "doconfig" to build a good kernel.
        o Move the new kernel to /.
        o Before rebooting make sure that the MEMORY CHANNEL IP
          addresses for all cluster members are recorded in each member's
          /etc/hosts file.
        o Reboot the system.
```

For information on building, tuning, and debugging kernels see the DIGITAL UNIX *System Administration*, *System Configuration and Tuning*, and *Kernel Debugging* manuals.

6.3 Cannot Ping Members Across the Primary Network (PS, AS)

The primary network for Production Server is the MEMORY CHANNEL subnet; the primary network for Available Server is the network attached to the interface specified during installation. (Section 2.1 has a description of each product's primary network interface.)

If a member system does not respond to the `ping` command, do the following:

1. Check that each member's primary interface is configured `UP` and that the appropriate interface-related entries are present in that member's `/etc/rc.config` and `/etc/hosts` files.

   In the following example, host `clu14`'s primary network interface is `mc0`:

   ```
   # ifconfig mc0

   mc0: flags=863<UP,BROADCAST,NOTRAILERS,RUNNING,SIMPLEX>
        inet 10.0.0.2 netmask ffffff00 broadcast 10.0.0.255 ipmtu 8008
   ```

The interface is configured UP, and has the following NETDEV_*n* and IFCONFIG_*n* entries in the member's /etc/rc.config file:

```
# egrep "mc0|10.0.0.2" /etc/rc.config
NETDEV_1="mc0"
IFCONFIG_1="10.0.0.2 netmask 255.255.255.0"
```

The interface's host entry in /etc/hosts associates the IP address assigned to the IFCONFIG entry to the IP name assigned the CLUSTER_NET entry:

```
# rcmgr get CLUSTER_NET
mcclu14
# grep mcclu14 /etc/hosts
10.0.0.2     mclu14.abc.def.com     mcclu14
```

2. Make sure that the following entries are in each member system's /etc/hosts:
   - An entry for each member system's IP name and IP address on the cluster's primary network.
   - The IP host addresses used by critical network services such as BIND, NIS, and NTP.
   - For Production Server, the MEMORY CHANNEL IP address of the connection manager service (cluster_cnx), which must be host number 42 on the MEMORY CHANNEL subnet. (The clu_ivp utility checks for the presence of the cluster_cnx service but does not verify its IP address.)
   - For systems with more than one network interface, the IP host names and addresses used to communicate with cluster members and clients through those network interfaces. For example, a Production Server cluster has a conventional Ethernet or FDDI network in addition to its MEMORY CHANNEL subnet; an Available Server ASE often has a secondary network as a backup.

6.4 MEMORY CHANNEL Cables Are Crossed (PS, MC)

Each system in a failover-capable cluster must have identically configured MEMORY CHANNEL adapters.

For a physical hub configuration, if the primary adapter is plugged, for example, into the primary hub's linecard in slot 3, the alternate adapter must be plugged into the alternate hub's linecard in slot 3. (The slot location determines the adapter's node ID, and the node IDs must be identical among all cluster members.)

If the MEMORY CHANNEL adapters are not connected properly, the system can panic with the following message:

```
rm_check_cables: cables are crossed
```

In a two-system, virtual-hub cluster, the jumper settings determine the node IDs. A system's primary and alternate adapters must be jumpered identically (either as VH0 or VH1). See the TruCluster Software Products *Hardware Configuration* manual for information on configuring MEMORY CHANNEL adapters. See Section 3.14 for information on setting up a tie-breaker disk for virtual-hub clusters.

6.5 System Cannot Join the Cluster (PS)

If `cnxshow` indicates that a system is unable to join the cluster, perform the following checks:

1. Use the `ps ag` command to verify that the `portmap` and `cfgmgr` processes are running. These processes, while not specific to clusters, must be running in order for the cluster to operate. For example:

   ```
   # ps ag | egrep "portmap|cfgmgr" | grep -v egrep
   224 ??   I    0:04.77 /usr/sbin/portmap
   244 ??   I    0.00.01 /sbin/cfgmgr
   ```

2. Check initialization and error messages (for example, the `daemon.log` and `kern.log` files, and the `uerf` utility). See Appendix A for examples of startup, cluster formation, and cluster recovery messages.

6.6 ASE Validation Fails (PS, AS)

If either the `clu_ivp` utility or the `drd_ivp` utility (Production Server only) reports that the available server environment (ASE) validation checks failed, run the `asemgr` utility with the `-d` and `-h` options on one system in each ASE to ensure that all ASE member systems are up and running. For example:

```
# asemgr -dh
        Member Status

Member:                  Host Status:    Agent Status:
mcclu6                   UP              RUNNING
mcclu7                   UP              RUNNING
```

See `asemgr`(8) for more information on these options.

Because each Available Server installation consists of a single ASE, the following applies only to Production Server installations.

All members in an ASE must have the same ASE ID. You can use the `rcmgr get ASE_ID` command to check the ASE identifier (ASE_ID) of each system. For example:

```
# rcmgr get ASE_ID
1
```

To change a system's ASE_ID, follow these steps:

1. If DRD services are configured, delete all services on the system.
2. Shut the system down to single-user mode.
3. Set the ASE_ID value. In the following example, the `rcmgr` command is used to set the ASE_ID value to 2 to match the ASE_ID assigned to the other members in the ASE:

   ```
   # rcmgr set ASE_ID 2
   ```
4. Halt and reboot the system.
5. Add any DRD services that were deleted.

6.7 The drd_ivp Utility Cannot Determine ASE Membership (PS)

If the `drd_ivp` utility is run (either manually or as part of the `clu_ivp` utility) prior to defining the available server environment (ASE) member list, it can report that it is unable to determine ASE membership. For example:

```
#drd_ivp

Cluster Configuration Information

Hostname               ASE_ID   BSSD   BSSD    DRD    Lic
                                 Reg   Resp   Conf    Reg
------------------------------------------------------------
mcclu6                      0    Yes    Yes    Yes    Yes
mcclu7                      0    Yes    Yes    Yes    Yes

DRD configuration validation tests succeeded.
Unable to determine which nodes are in the same ASE
as node mcclu6.   Verify that node mcclu6 is up and that it
has the ASE_ID parameter in its '/etc/rc.config' file.
Verify that mcclu6 is registered as a member of an ASE.
Unable to determine which nodes are in the same ASE
as node mcclu7.   Verify that node mcclu7 is up and that it
has the ASE_ID parameter in its '/etc/rc.config' file.
Verify that mcclu7 is registered as a member of an ASE.
Failed to validate ASE_ID values.
```

Use the `asemgr` utility to populate the ASE member list. Then rerun either the `clu_ivp` utility or the `drd_ivp` utility to check that the systems are registered as members of the ASE.

The TruCluster Software Products *Administration* manual provides more information on troubleshooting the DRD subsystem.

6.8 Inconsistent View of Shared SCSI Devices (PS, AS)

If the member systems connected to a shared SCSI bus have inconsistent views of the devices on the bus (all ASE members must have identical numbers for shared buses and devices), do the following:

1. Make sure that all shared SCSI cables are connected and terminated as described in the TruCluster Software Products *Hardware Configuration* manual.
2. For systems that support the `bus_probe_algorithm` console variable, check that its value is set to `new` (see Section 2.3).
3. Verify that the shared SCSI buses are numbered equivalently on each system. As mentioned in Chapter 5, you can run the `clu_ivp` utility on each system and compare the output to check whether all system have the same view of shared SCSI buses and devices. If you discover an inconsistency, do the following on the affected system or systems:
   a. Run the `/var/ase/sbin/ase_fix_config` utility, described in Section 3.7, and adjust the bus numbering.
   b. Build a new kernel using the `doconfig -c` *`HOSTNAME`* command.
   c. Move the new kernel to `/vmunix`.
   d. Reboot the system.

A

Cluster-Related Messages in System Log Files

The following three sections show excerpts from system log files in the /var/adm/syslog.dated/*date* directories:

- Startup messages following Production Server installation (taken from kern.log)
- Formation of a new Production Server cluster (taken from daemon.log)
- Recovery of an existing Production Server cluster (taken from daemon.log)

These messages track normal cluster startup operations; therefore, in addition to providing some level of assurance that cluster formation and recovery operations are proceeding in an orderly fashion, they also provide a starting point for troubleshooting cluster-related problems.

A.1 Startup Messages Following Production Server Installation

Example A–1 shows a transcript of a portion of the startup messages displayed during a reboot of the first cluster member system after installing Production Server. This information is also sent to /var/adm/syslog.dated/*date*/kern.log. Callouts in this example highlight messages relevant to cluster installation.

Example A–1: Startup Messages Related to Cluster Installation

```
>>> boot
.
.
.
jumping to bootstrap code

Digital UNIX boot - Wed May 28 17:05:23 EDT 1997

Loading vmunix ...
.
.
.
pci0 at nexus
eisa0 at pci0
ace0 at eisa0
ace1 at eisa0
lp0 at eisa0
```

Example A–1: Startup Messages Related to Cluster Installation (cont.)

```
fdi0 at eisa0
fd0 at fdi0 unit 0
cirrus0 at eisa0
cirrus0: Cirrus Logic CL-GD5428 (SVGA) 512 Kbytes
pci2000 at pci0 slot 8
isp0 at pci2000 slot 0
isp0: QLOGIC ISP1020A
isp0: Firmware revision 5.19 (loaded by console)
scsi0 at isp0 slot 0
rz0 at scsi0 target 0 lun 0 (LID=0) (DEC     RZ28M    (C) DEC 0568) (Wide16)
rz1 at scsi0 target 1 lun 0 (LID=1) (DEC     RZ29B    (C) DEC 0007) (Wide16)
rz5 at scsi0 target 5 lun 0 (LID=2) (DEC     RRD45   (C) DEC  1645)
pza0 at pci2000 slot 1
pza0 firmware version: DEC  F01  A10
scsi1 at pza0 slot 0
rz9 at scsi1 target 1 lun 0 (LID=3) (DEC     RZ26     (C) DEC 392A)
rz10 at scsi1 target 2 lun 0 (LID=4) (DEC     RZ26     (C) DEC 392A)
processor at scsi1 target 6 lun 7 (LID=12) (DEC ASE DEC  L01  A10 TMV2) (Wide16)
pza1 at pci2000 slot 2
pza1 firmware version: DEC  L01  A10
scsi2 at pza1 slot 0
rz18 at scsi2 target 2 lun 0 (LID=13) (DEC     RZ26N    (C) DEC 0744)
rz19 at scsi2 target 3 lun 0 (LID=14) (DEC     RZ26N    (C) DEC 0616)
processor at scsi2 target 6 lun 7 (LID=22) (DEC ASE DEC  L01  A10 TMV2) (Wide16)
pza2 at pci2000 slot 3
pza2 firmware version: DEC  F01  A10
scsi3 at pza2 slot 0
pza3 at pci2000 slot 4
pza3 firmware version: DEC  F01  A10
scsi4 at pza3 slot 0
mchan0: Module revision = 33E   1
mchan0: jumpered as VH1 configuration
mchan0 at pci0 slot 11
tu0: DECchip 21040: Revision: 2.3
tu0 at pci0 slot 13
tu0: DEC TULIP (10Mbps) Ethernet Interface, hardware address: 08-00-2B-E5-F8-0A
tu0: console mode: selecting 10BaseT (UTP) port: half duplex
gpc0 at eisa0
Created FRU table binary error log packet
kernel console: ace0
dli: configured
clubase: configured   2
dlmsl: configured   3
drd: configured.   4
cnxagent: configured   5
dlm: configured.   6
memory channel thread init   7
rm_sw_init: begin MC initialization.
rm_boot_am_i_alone: entered
checking for existing memory channel nodes   8
rm_slave_init
rm_get_proto: returning vers = 1
slave unit boot phase 0: checking cables   9
slave unit boot phase 1: request data ...
slave unit boot phase 2: get lock data from all nodes
slave unit boot phase 3: update request ...
memory channel software inited - node 1 on mc0   10
rm_get_proto: returning vers = 1
ccomsub: state change detected via remote node 0
ccomsub: configured   11
```

Example A–1: Startup Messages Related to Cluster Installation (cont.)

```
mcnet: configured
memory channel - adding node 0
RM member change callback: no change in member bitmap 0x3
ADVFS: using 1153 buffers containing 9.00 megabytes of memory
starting LSM
Checking local filesystems
/sbin/ufs_fsck -p
.
.
.
Streams autopushes configured
Initializing the ASE Availability Manager   12
AM found a host at bus 1 target 6, lun 7
AM found a host at bus 2 target 6, lun 7
Configuring network
hostname: clu14.abc.def.com   13
.
.
.
/usr/sbin/drd_dma: Peer-to-peer DMA is NOT sure to work between   14
             scsi and MEMORY CHANNEL controllers
/usr/sbin/drd_dma: Peer-to-peer DMA over MEMORY CHANNEL is NOT enabled.
ONC portmap service started
Cluster member started
Starting ASE ...   15
        Initializing the ASE Availability Manager
        ASE logger started (/usr/sbin/aselogger)
        ASE agent started (/usr/sbin/aseagent)
ASE member started
.
.
.
cnxagent: Get MC information reports hubless   16
cnxagent: added node mcclu13
cnxagent: mcclu14 is now a cluster member   17
dlm_agent: resuming lock activity
.
.
.
Network Time Service started
cnxagent: resuming
.
.
.
Printer service started
The system is ready.
```

The messages highlighted in Example A–1 indicate the following:

1. The three `mchan` lines indicate that a device probe has found the MEMORY CHANNEL adapter and determined its revision number. This adapter is jumpered as VH1, indicating that it is part a virtual hub. (The message indicate whether a MEMORY CHANNEL adapter is jumpered as VH0 or VH1 (virtual hub) or connects to a MEMORY CHANNEL hub.)
2. The cluster component is initializing.
3. The Distributed Lock Manager (DLM) Session Layer component is initializing.
4. Distributed raw disk (DRD) is initializing.

5 The connection manager is initializing.

6 The DLM is initializing.

7 The general-purpose MEMORY CHANNEL thread has completed initialization.

8 The system is looking for other nodes connected to the MEMORY CHANNEL that may be either running or in the process of booting.

9 This system is the second to boot (slave) and initialize MEMORY CHANNEL code.

10 The initialization of low-level MEMORY CHANNEL software is complete.

11 The cluster communication subsystem is initializing.

12 The ASE availability manager driver is initializing. The hardware probes for shared buses and reports any active hosts found.

13 The system prints its hostname (the output from `/sbin/hostname`).

14 The `drd_dma` checks the hardware configuration to determine whether the system can use peer-to-peer DMA, and prints the result.

15 The ASE daemons are started.

16 The `cnxagent` subsystem reports that the cluster is operating in a virtual hub configuration.

17 The system is identified as a cluster member.

See the TruCluster Software Products *Administration* manual for descriptions of important messages generated by TruCluster products.

A.2 Formation of a New Cluster

Example A–2 shows messages from the `daemon.log` file related to the formation of a new Production Server cluster.

Example A–2: Log File Showing Formation of a New Cluster

```
May 17 17:49:33 mcclu5 cnxpingd: starting 1
May 17 17:49:33 mcclu5 cnxagentd: starting
May 17 17:49:34 mcclu5 cnxmond: changed alias with : /sbin/ifconfig mc0 alias 10.0.0.42
netmask 255.255.255.0 2
May 17 17:49:34 mcclu5 cnxmgrd: starting 3
.
.
.
May 17 17:50:14 mcclu5 cnxmgrd: attempting cluster recovery/formation 4
May 17 17:50:14 mcclu5 cnxmgrd: recovery, considering mcclu5
May 17 17:50:14 mcclu5 cnxmgrd: node mcclu5, cluster incarn 0, update_seq 0
May 17 17:50:14 mcclu5 cnxmgrd: node mcclu5 not a member
May 17 17:50:14 mcclu5 cnxmgrd: forming a cluster
May 17 17:50:14 mcclu5 cnxmgrd: completed cluster recovery/formation
.
.
.
May 17 17:50:18 mcclu5 cnxmgrd: starting join operation for mcclu5 5
May 17 17:50:18 mcclu5 cnxmgrd: join, getting status from mcclu5
May 17 17:50:18 mcclu5 cnxmgrd: node mcclu5, cluster incarn 0, update_seq 0
May 17 17:50:18 mcclu5 cnxmgrd: adding node mcclu5
```

Example A–2: Log File Showing Formation of a New Cluster (cont.)

```
May 17 17:50:18 mcclu5 cnxmgrd: update complete, summary follows
May 17 17:50:18 mcclu5 cnxmgrd:   members are:
May 17 17:50:18 mcclu5 cnxmgrd:    mcclu5
May 17 17:50:18 mcclu5 cnxmgrd:   timed out are:
May 17 17:50:18 mcclu5 cnxmgrd:    none
May 17 17:50:18 mcclu5 cnxmgrd: finished join operation, update_seq 2
.
.
.
May 17 17:50:18 mcclu5 xntpd[668]: xntpd version 1.3
May 17 17:51:45 mcclu5 ASE: local HSM Notice: member mcclu8 is UP [6]
.
.
.
May 17 17:51:52 mcclu5 cnxmgrd: starting join operation for mcclu8 [7]
May 17 17:51:52 mcclu5 cnxmgrd: join, getting status from mcclu8
May 17 17:51:52 mcclu5 cnxmgrd: node mcclu8, cluster incarn 0, update_seq 0
May 17 17:51:52 mcclu5 cnxmgrd: adding node mcclu8
May 17 17:51:53 mcclu5 cnxmgrd: update complete, summary follows
May 17 17:51:53 mcclu5 cnxmgrd:   members are:
May 17 17:51:53 mcclu5 cnxmgrd:    mcclu5
May 17 17:51:53 mcclu5 cnxmgrd:    mcclu8
May 17 17:51:53 mcclu5 cnxmgrd:   timed out are:
May 17 17:51:53 mcclu5 cnxmgrd:    none
May 17 17:51:53 mcclu5 cnxmgrd: finished join operation, update_seq 3
.
.
.
May 17 17:52:04 mcclu5 ASE: local Director Notice: agent on mcclu8 came ONLINE [8]
.
.
.
May 17 17:54:54 mcclu5 cnxmond: node mcclu8 timed out [9]
May 17 17:55:30 mcclu5 cnxmgrd: intend to remove node mcclu8
May 17 17:55:54 mcclu5 cnxmgrd: starting removal operation
May 17 17:55:54 mcclu5 cnxmgrd: removing node mcclu8
May 17 17:55:55 mcclu5 cnxmgrd: update complete, summary follows
May 17 17:55:55 mcclu5 cnxmgrd:   members are:
May 17 17:55:55 mcclu5 cnxmgrd:    mcclu5
May 17 17:55:55 mcclu5 cnxmgrd:   timed out are:
May 17 17:55:55 mcclu5 cnxmgrd:    none
May 17 17:55:55 mcclu5 cnxmgrd: finished removal operation, update_seq 4
```

[1] The connection manager monitor daemon, `cnxmond`, starts the ping daemon, `cnxpingd`, and agent daemon, `cnxagentd`, on system `mcclu5`.

[2] The connection manager monitor daemon (`cnxmond`) on system `mcclu5` acquires a spinlock on the MEMORY CHANNEL bus (`mc0`) and registers the `cluster_cnx` service alias (`10.0.0.42`).

[3] As a result of acquiring the spinlock and registering the `cluster_cnx` alias, the connection manager starts the director daemon (`cnxmgrd`) on system `mcclu5`.

[4] The director tries to find systems that are eligible to become cluster members. If the director finds eligible systems, these systems become members and the cluster is formed from them. In this case, no other eligible systems were found; therefore, the cluster is formed with no other members.

[5] Systems are added to the cluster as members one at a time. Since system `mcclu5` is up and running, it is considered first and added as a member.

6 System mcclu8 is detected as up and running.

7 System mcclu8 becomes a cluster member. The cluster now has two members, mcclu5 and mcclu8.

8 The ASE director detects a new agent on system mcclu8.

9 The director detects a system failure and removes the failed system from the cluster membership. The cluster now has only one member, mcclu5.

A.3 Recovery of an Existing Cluster

Example A–3 shows an excerpt from the daemon.log file related to the recover of an existing Production Server cluster.

Example A–3: Log File Showing Recovery of an Existing Cluster

```
May 17 18:18:03 mcclu5 cnxmond: changed alias with : /sbin/ifconfig mc0 alias 10.0.0.42 netmask 1
255.255.255.0
May 17 18:18:03 mcclu5 cnxmgrd: starting
May 17 18:18:43 mcclu5 cnxmond: recovery delay completed
May 17 18:18:43 mcclu5 cnxmgrd: attempting cluster recovery/formation
May 17 18:18:43 mcclu5 cnxmgrd: recovery, considering mcclu5

May 17 18:18:43 mcclu5 cnxmgrd: node mcclu5, cluster incarn 0000000000080e90, update_seq 2 2
May 17 18:18:43 mcclu5 cnxmgrd: found member of cluster incarnation 0000000000080e90
May 17 18:18:43 mcclu5 cnxmgrd: node is mcclu5, update_seq is 2
May 17 18:18:43 mcclu5 cnxmgrd: same update_seq 2, node mcclu5
May 17 18:18:43 mcclu5 cnxmgrd: using mcclu5
May 17 18:18:43 mcclu5 cnxmgrd: recovering a cluster 3
May 17 18:18:43 mcclu5 cnxmgrd: starting recovery update
May 17 18:18:43 mcclu5 cnxmgrd: update complete, summary follows
May 17 18:18:43 mcclu5 cnxmgrd:   members are:
May 17 18:18:43 mcclu5 cnxmgrd:    mcclu5
May 17 18:18:43 mcclu5 cnxmgrd:   timed out are:
May 17 18:18:43 mcclu5 cnxmgrd:    none
May 17 18:18:43 mcclu5 cnxmgrd: finished recovery update, update_seq 3
May 17 18:18:43 mcclu5 cnxmgrd: completed cluster recovery/formation
```

1 A director is selected and cluster recovery is started.

2 The director finds system mcclu5, which was a member of a cluster (note the nonzero value of cluster incarn).

3 The cluster created in Example A–2 is recovered.

B

Modifications to System Files

The following files are created or modified as a result of installing TruCluster software:

`/etc/hosts`
: The IP address and interface name associated with the primary network interface.

  For Production Server, the IP address of the `cluster_cnx` service.

`/etc/rc.config`
: Table B–1 lists the configuration variables that are either manipulated by the installation procedure or are of interest during installation.

  See the TruCluster Software Products *Administration* manual for a complete list of TruCluster `rc.config` variables.

`/etc/sysconfigtab`
: Table B–2 lists the configuration variables that are either manipulated by the installation procedure or are of interest during installation.

  See the TruCluster Software Products *Administration* manual for a complete list of TruCluster `sysconfigtab`variables.

`/sys/conf/`*`NAME`*`.list`
: This file registers the cluster software and revised SCSI bus numbers with the kernel before the installation procedure calls the `doconfig` program.

`/sys/`*`NAME`*`/vmunix`
: A new kernel is built and placed in `/sys/`*`NAME`*`/vmunix`, where *`NAME`* is the name of the configuration file specified during installation.

`/var/ase/config/config.state`
: For Production Server and Available Server, the file created and maintained by the `ase_fix_config` command to store SCSI controller information.

You must manually modify the files described in Table B–3 on each cluster member after Production Server is installed on all member systems.

Table B–1: /etc/rc.config Variables

Variable	Product			Comment
	PS	AS	MC	
ASE	X	X		ASE="on" indicates that the ASE is configured. The installation procedure adds the ASE variable and sets its value to on or off depending on user input during the installation.
ASE_ID	X	X		The system's ASE identifier; legal values are 0–63. The installation procedure adds the ASE_ID variable. For Production Server, the installation procedure assigns the user-specified value; for Available Server, the value defaults to 0.
ASELOGGER	X	X		ASELOGGER=1 indicates that the system will run an ASE logger daemon. The installation procedure adds the ASELOGGER variable, and sets its value to 0 or 1 depending on user input.
CLUSTER_NET	X	X		For PS and AS, the name of the system's primary network interface. The installation procedure adds a CLUSTER_NET entry and sets its value to the user-specified interface name.
CNX_DISK	X			The device name of the tie-breaker disk specified using the cnxset utility. See Section 3.14 for information on setting up a tie-breaker disk.
IMC_AUTO_INIT	X		X	IMC_AUTO_INIT=1 indicates that libimc will be automatically initialized whenever the system boots. This initialization involves reserving approximately 4.5 MB for the library. The default value at installation is 1.
IMC_MAX_ALLOC	X		X	Determines the maximum aggregate amount (in MB) of MEMORY CHANNEL address space that libimc can allocate for its use across the cluster. If the value of this variable differs among cluster members, the largest value specified is used. The default value at installation is 10.
IMC_MAX_RECV	X		X	Determines the maximum aggregate amount (in MB) of physical memory that libimc can map for reading MEMORY CHANNEL address space. This limit is node-specific. The default value at installation is 10.
NETDEV	X	X	X	The network device name for the system's primary PS, AS, or MC network interface. The installation procedure adds a NETDEV_*n* entry.
NUM_NETCONFIG	X	X	X	The number of configured network devices. The installation program increments the value of NUM_NETCONFIG.

Table B–1: /etc/rc.config Variables (cont.)

Variable	Product			Comment
	PS	AS	MC	
TCR_INSTALL	X	X	X	Indicates a successful installation when its value is TCR, ASE, or MCA. Indicates an unsuccessful installation when its value is BAD.
TCR_PACKAGE	X	X	X	Indicates TruCluster product installed on system: TCR for Production Server; ASE for Available Server; MCA for MEMORY CHANNEL

Table B–2: /etc/sysconfigtab Attributes

Subsystem	Attribute	Product			Comment
		PS	AS	MC	
clubase	cluster_disable	X	X	X	For Production Server and Available Server, when cluster_disable is set to 1, disables all cluster components. For MEMORY CHANNEL, when cluster_disable is set to 1, the MEMORY CHANNEL API library is not automatically initialized at boot time (even if IMC_AUTO_INIT is set to 1 in /etc/rc.config). The default value is 0.
	cluster_version	X	X	X	TruCluster product version string. Set by installation procedure.
dlm	dlm_disable_rd	X			During installation, new DLM features are deliberately disabled until turned on by the dlm_enable script, which sets dlm_disable_rd=0. See Section 3.10 and Section 4.4 for more information on this attribute. The default value is 1.
mchan	mchan_debug	X		X	Set to 1 for verbose MEMORY CHANNEL error messages. The default value is 0.
rm	rm-no-inheritance	X		X	Controls MEMORY CHANNEL object inheritance at fork(). Required in order to use libimc routines. The value is set to 1 by the installation procedure.

Table B–3: Manual Modifications to System Files

File	Modification
/etc/fdmns/*domainname*	If you renumbered shared SCSI buses, correct the symbolic links in these directories to point to the new device names.
/etc/fstab	If you renumbered shared SCSI buses, modify entries in this file to reference the new device names.
/etc/hosts	Add a host entry for each member system.
/etc/svc.conf	Make sure that the access path for the hosts database lists local first.
/sys/*HOSTNAME*/vmunix	Move the new kernel built during cluster kernel configuration to the root (/) directory.

Index

A

E

F

G

H

I

K

L

M

N

O

P

R

S

T

U

V

X

How to Order Additional Documentation

Technical Support

If you need help deciding which documentation best meets your needs, call 800-DIGITAL (800-344-4825) before placing your electronic, telephone, or direct mail order.

Electronic Orders

To place an order at the Electronic Store, dial 800-234-1998 using a modem from anywhere in the USA, Canada, or Puerto Rico. If you need assistance using the Electronic Store, call 800-DIGITAL (800-344-4825).

Telephone and Direct Mail Orders

Your Location	Call	Contact
Continental USA, Alaska, or Hawaii	800-DIGITAL	Digital Equipment Corporation P.O. Box CS2008 Nashua, New Hampshire 03061
Puerto Rico	809-754-7575	Local Digital subsidiary
Canada	800-267-6215	Digital Equipment of Canada Attn: DECdirect Operations KAO2/2 P.O. Box 13000 100 Herzberg Road Kanata, Ontario, Canada K2K 2A6
International	—	Local Digital subsidiary or approved distributor
Internal (submit an Internal Software Order Form, EN-01740-07)	—	SSB Order Processing – NQO/V19 *or* U.S. Software Supply Business Digital Equipment Corporation 10 Cotton Road Nashua, NH 03063-1260

Reader's Comments

TruCluster Software Products
Software Installation
AA-R88HA-TE

Digital welcomes your comments and suggestions on this manual. Your input will help us to write documentation that meets your needs. Please send your suggestions using one of the following methods:

- This postage-paid form
- Internet electronic mail: `readers_comment@zk3.dec.com`
- Fax: (603) 884-0120, Attn: UBPG Publications, ZKO3-3/Y32

If you are not using this form, please be sure you include the name of the document, the page number, and the product name and version.

Please rate this manual:

	Excellent	Good	Fair	Poor
Accuracy (software works as manual says)	☐	☐	☐	☐
Clarity (easy to understand)	☐	☐	☐	☐
Organization (structure of subject matter)	☐	☐	☐	☐
Figures (useful)	☐	☐	☐	☐
Examples (useful)	☐	☐	☐	☐
Index (ability to find topic)	☐	☐	☐	☐
Usability (ability to access information quickly)	☐	☐	☐	☐

Please list errors you have found in this manual:

Page	Description
_____	_____
_____	_____
_____	_____
_____	_____

Additional comments or suggestions to improve this manual:

_____
_____
_____
_____

What version of the software described by this manual are you using? _____

Name, title, department _____
Mailing address _____
Electronic mail _____
Telephone _____
Date _____

Do Not Cut or Tear − Fold Here and Tape

digital™

NO POSTAGE
NECESSARY IF
MAILED IN THE
UNITED STATES

BUSINESS REPLY MAIL
FIRST−CLASS MAIL PERMIT NO. 33 MAYNARD MA

POSTAGE WILL BE PAID BY ADDRESSEE

DIGITAL EQUIPMENT CORPORATION
UEG PUBLICATIONS MANAGER
ZKO3−3/Y32
110 SPIT BROOK RD
NASHUA NH 03062−9987

Do Not Cut or Tear − Fold Here

Cut on
Dotted
Line