HP-UX & hp Tru64 UNIX®
Side-by-Side Comparison of
TruCluster Server
and
MC/ ServiceGuard

Jan Mark Holzer
jmh@hp.com
Senior Member of Technical Staff
Enterprise UNIX® Division

April 7th, 2003

http://www.tru64.org/~jmh/decus2003/
agenda

- product roadmaps and history
- product design and principles
- high availability solutions
- ha solution architecture
- selected cluster subsystems
agenda

• product roadmaps and history
• product design and principles
• high availability solutions
• ha solution architecture
• selected cluster subsystems
• TruCluster Server technology moving forward into HP-UX/ServiceGuard
• Transition from MC/ServiceGuard today to HP-UX fused with TruCluster technology

Actual company and product names mentioned herein are the trademarks of their respective owners.
TruCluster & MC/ServiceGuard release history and roadmaps
hp TruCluster Server themes

**V4.* Cluster Products**
- **Available Server**
  - HA Applications
    - introduced in spring 1994
    - HA applications
    - failover scripts
    - AdvFS, LSM, HW RAID, standard networks
- **Production Server**
  - HA & Scalable Applications
    - introduced in spring 1996
    - Cluster Disk Service
    - Synchronization Service
    - Membership Service
    - Memory Channel

**V5.* Cluster**
- **TruCluster Server**
  - SSI Storage and System Management
    - introduced in summer 1999
    - cluster file system
    - cluster alias
    - application availability
    - clusterwide events
    - additional interconnects

**Future Technologies**
- **SSI Server**
  - SSI Process Management
    - SSI process management
    - enhanced load balancing
    - process migration
TruCluster V5.x release history

**V1.x** based products (starting in early 90’s)
- initially simple failover product (DECsafe)
- evolved to TruCluster V1.x product suite (Production Server, Available Server and MEMORY CHANNEL cluster products)

**V5.0** (Steel) — Q3CY99
- limited release

**V5.0A** (Zinc) — Q2CY00
- general release, included FC support

**V5.1** (Zulu) — Q3CY00
- GS80/GS160/GS320 platform support, CFS enhancements

**V5.1A** (Yankee) — Q3CY01
- ES45 platform support, increased low-end cluster support

**V5.1B** (Wildcat.Alpha) — H2/CY02
- New Generation HP AlphaServer platform support, CFS load balancing, CAA performance improvements
  TCR version coincides with base OS version
## ServiceGuard release history

<table>
<thead>
<tr>
<th>OS</th>
<th>SG</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.01</td>
<td>A.10.03</td>
<td>06/95 4 Member, Basic functionality (restart of apps), SAM GUI</td>
</tr>
<tr>
<td>10.10</td>
<td>A.10.04</td>
<td>12/95 8 Node Support, RS323 Link for heartbeat, Security Enhancements</td>
</tr>
<tr>
<td></td>
<td>A.10.05</td>
<td>02/96 End of Support March 31st 2002</td>
</tr>
<tr>
<td>10.20</td>
<td>A.10.06</td>
<td>06/96 End of Support December 31st 2001</td>
</tr>
<tr>
<td></td>
<td>A.10.10</td>
<td>10/97 Online administration, EMS Support</td>
</tr>
<tr>
<td></td>
<td>A.10.11</td>
<td>12/98 Support for EMS V3.0 framework and monitors</td>
</tr>
<tr>
<td></td>
<td>A.10.12</td>
<td>12/99 Platforms Release, misc bug fixes</td>
</tr>
<tr>
<td>10.30</td>
<td>A.10.08</td>
<td>N/A Only for specific customers</td>
</tr>
<tr>
<td>11.Xx</td>
<td>A.11.01</td>
<td>02/98 11.01 OS Support, Mixed Cluster Support. Up to 50 IP Aliases</td>
</tr>
<tr>
<td></td>
<td>A.11.03</td>
<td>08/98 Rotating Standby Feature, New Failover Policies. Up to 200 IP Aliases</td>
</tr>
<tr>
<td></td>
<td>A.11.04</td>
<td>12/98 Up to 16 nodes. New cmquerycl options. EMS Support</td>
</tr>
<tr>
<td></td>
<td>A.11.05</td>
<td>02/99 Advanced Tape Services</td>
</tr>
<tr>
<td></td>
<td>A.11.07</td>
<td>06/99 Support for APA (Auto Port Aggregation)</td>
</tr>
<tr>
<td></td>
<td>A.11.08</td>
<td>09/99</td>
</tr>
<tr>
<td></td>
<td>A.11.09</td>
<td>12/99 Platform Release. Limited Support for VxVM</td>
</tr>
<tr>
<td></td>
<td>A.11.12</td>
<td>12/00 Not supported on HP-UX 11i. Initial release for SG Manager</td>
</tr>
<tr>
<td></td>
<td>A.11.13</td>
<td>09/01 Platform Release. Full support for VxVM/CVM on 11i. Max pkg 60</td>
</tr>
<tr>
<td></td>
<td>A.11.14</td>
<td>03/02 Platform Release. Quorum Server. Parallel fsck/VG. Max pkgs to 200</td>
</tr>
</tbody>
</table>
UNIX® operating system roadmap

(HP-UX and HP Tru64 UNIX)

2002
11i v1.5
HP-UX
1. RAS
2. Internet Web
3. Directory/Security
4. Scalability
5. Manageability

2003
11i v1.6
Planning and Pilots
11i v2
Begin the migration
1. Migration tools
   - Phase I
2. Common System Mgmt

Itanium® Phase In
11i v3
Scalability/ Mgmt
1. TruCluster technology
2. AdvFS (16 TB)
3. Migration tools
   - Phase II
4. Common System Mgmt

2004
11i v4
1. Self tuning, self adapting vertically and horizontally

HP-Tru64 UNIX

V5.1B
1. Scalability
2. System Mgmt

V5.Vail
1. Common System Mgmt

V5. Utah

* Sales thru at least 2006, support thru at least 2011

customer value—investment protection and a better HP-UX
hp Tru64 UNIX® roadmap details

<table>
<thead>
<tr>
<th>Today</th>
<th>02</th>
<th>03</th>
<th>04</th>
<th>05</th>
</tr>
</thead>
<tbody>
<tr>
<td>V5.1A</td>
<td>V5.1B</td>
<td>V.&quot;Vail&quot;</td>
<td>V.&quot;Utah&quot;</td>
<td>V5.next</td>
</tr>
</tbody>
</table>
| • ES45 Support
• Link Aggregation (network trunking)
• OLAR Enhancements
  • CPU Hotswap
  • Mixed CPU
• Faster Patching
• Workload Mgt (Aurema)
• UNIX98 Branded | • EV7 infrastructure (up to 32p)
• Big Pages for HPTC
• Enhanced storage SAN support
• Production-level IPV6/IPSEC
• Linux® affinity enhancements (OpenOffice) | • Support up to 64p SMP
• Continued leadership storage SAN support
• Enhanced resiliency/maintenance functions
• New hardware updates | • Ev79 system rollout support
• HP-UX compatibility (migration support)
• Maintenance | • Tru64 UNIX updates releases on Alpha at least through 2006
• Support at least until 2011 |

- Best technology of Tru64 UNIX® will be integrated in HP-UX
- HP-UX compatibility tools for Tru64 UNIX
- Mixed (HP-UX and Tru64 UNIX) environment management tools
- Maintenance and support of Tru64 UNIX at least until 2011

All dates are estimates and subject to change.
<table>
<thead>
<tr>
<th>V5.1A</th>
<th>V5.1B</th>
<th>V.&quot;Vail&quot;</th>
<th>V.&quot;Utah&quot;</th>
<th>V5.next</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLSM for root</td>
<td>faster patch/upgrade installations</td>
<td>support up to 64p SMP</td>
<td>SSI enhancements</td>
<td>Tru64 UNIX update releases on Alpha at least through 2006</td>
</tr>
<tr>
<td>LAN as Cluster Interconnect</td>
<td>ongoing new platform support</td>
<td>&gt; 8-node cluster support</td>
<td>HP-UX compatibility (migration support)</td>
<td>support at least until 2011</td>
</tr>
<tr>
<td>cache directed reads</td>
<td>CFS load balancing support/feedback</td>
<td>continuing new platform/options support</td>
<td>new platform and option support as needed</td>
<td></td>
</tr>
<tr>
<td>CFS quotas</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- best technology of TruCluster will be integrated in HP-UX
- mixed (HP-UX and Tru64 UNIX) environment management tools
- maintenance and support of Tru64 UNIX and TruCluster, based on customer demand, at least until 2011

All dates are estimates and subject to change.
# HP-UX and Cluster Roadmap

**HP-UX**
- 11i v1.5
  - PA-RISC
- 11i v1
  - PA-RISC
- 11i v1.6
  - Itanium 1
- 11i v1.6
  - Itanium 2

**“RAS, Scalability for IA and PA”**
- 11i v2
- 11i v3
  - SSI Clustering built-in
    - TruCluster
    - AdvFS (16 TB)
- 11i v4
  - “Self tuning, self adapting”

**MC/Serviceguard**
- A11.14
  - New features for HP-UX 11i v1 and v2
- A11.15
- A11.31
  - TruCluster functionality (CAA, CFS, MIB, DRD, etc.)
- A11.40
  - MC/Serviceguard installed customers choose when to upgrade

**Subject to change**
- 2002
- 2003
- 2004
- 2005
HP-UX Clusters Roadmap

2004

Just Built In

TruCluster technology
(core infrastructure CAA, DRD, CFS...)

PLUS:
• 32 nodes
• high speed/low latency interconnect
• Advanced File System

2005+

Flexible Computing

• Integrate and extend SSI clustering technology with
  WLM, PRM, UDC
• SSI extensions to include
  process mgmt
• More nodes
• Enhanced Disaster Tolerance
• File system enhancements

Clusters are the System
architectures

- shared nothing
  - independent nodes
  - no data sharing
  - each node must be managed independently

- shared storage connectivity and access
  - sharing of resources such as data and storage
    - locking needs to be used to coordinate shared data access
  - every member can access the same set of data/storage

- SSI (single system image)
  - cluster manages like a single system
  - most management tasks can be issued from any member
architectures

• both products provide highly available clustering solutions
  – applications can be active/standby or active/active

• TruCluster Server implements a shared everything architecture (SSI)
  – single system disk and root filesystem
  – clients and sysadmins see the cluster as one ‘big’ node

• MC/ServiceGuard implements a shared storage model
  – only one member can access a storage device at a time
    – special rules for Oracle9i RAC and Oracle8i OPS configurations
  – local storage for system disks and “private” storage
MC/ServiceGuard configuration

achieve no single point of failure configuration by using base operating system and ServiceGuard features

shared storage

cluster interconnect

private disks

system disk

private disks

system disk

i/o multi-pathing

APA

www.decus.de
HP TruCluster Server clusters manage as a single system

Separate Systems

Manage as many Complex

Single System View

- Simplified management
- Multiple nodes for increased scalability
- Redundant nodes for availability

Paradigm Shift: Single System Personality

Trust the inventors of clusters to make UNIX clusters simple

Increased Availability

Decreased Management Costs
TruCluster Server

increased scalability
scales out easily — add a	node in minutes

simplified high
availability
load balance applications

single system
management
manage any system and
storage from anywhere as
one with a cluster file
system (CFS)
TruCluster cluster file system to ease management
cluster file system, shared root, single system image

The cluster file system is shared by all cluster nodes
All nodes see the same data
The cluster manages as a single system, even as you add nodes
product portfolio and solutions

high availability products

campuswide cluster

disaster tolerance
hp clustering solutions

broad range of offerings available

- both platforms offer local HA solutions and have common goals
- easy to implement and maintain
- mature products that have been shipping and deployed in the market for many years
- MC/ServiceGuard offers multiple DT solutions
- TruCluster can be configured in similar configurations by using third-party products and HP Services
  - Oracle redo log shipping and/or storage replication (DRM, SRDF)
- hp supports the use of the HP XP disk array with Continuous Access and the EMC Symmetrix disk array with SRDF with the MetroCluster disaster tolerant solution
  - hp XP Arrays will be supported by Tru64 UNIX® and TruCluster Server
ServiceGuard disaster-tolerant solutions

Flexibility, functionality

- System failure, Certain SW bugs, Certain human error
- Fire, Building damage
- Power outage, Tornado, Fire, Localized flooding
- Hurricane, Earthquake, Flood

Distance

- Local MC/SG Cluster: single cluster, unattended failover, same data center, systems up to 100km apart
- MetroCluster: single cluster, unattended failover, same city/region, hp Continuous Access or EMC SRDF, data centers up to 100km apart
- Continental Clusters: separate clusters, "push-button" failover, data sites at unlimited distance

System failure, Certain SW bugs, Certain human error, Fire, Building damage, Power outage, Tornado, Fire, Localized flooding, Hurricane, Earthquake, Flood
TruCluster Server disaster-tolerant solutions

flexibility, functionality

- System failure, Certain SW bugs, Certain human error
- Fire
- Building damage
- Power outage, Tornado, Fire, Localized flooding
- Hurricane, Earthquake, Flood

distance
HA solution
architecture

hardware requirements
and configuration

I/O infrastructures and
storage access

file system access and
availability
hardware requirements and configuration

- MC/ServiceGuard supports both the PA RISC and Itanium® platforms
  - as of MC/ServiceGuard 11.14.01, HP-UX 11i V1.6 (11.22) is supported
- TruCluster Server supports all AlphaServer models
- you can configure a cluster using standard hardware
  - no need for a specialized cluster interconnect
    - LAN used for heartbeat (MC/ServiceGuard) and cluster interconnect (TruCluster V5.1A)
  - high-speed cluster interconnects are available
    - HyperFabric used for message passing for MC/ServiceGuard Oracle9i RAC
      - still need LAN for heartbeat
    - Memory Channel for TruCluster
  - SCSI and Fibre Channel storage are supported
    - large number of supported devices and adapters
MC/ ServiceGuard
I/O access and infrastructure

- shared storage model for data disks
- I/O multipath through base OS options (licensed)
  - AutoPath VA and XP
  - Veritas VxVM DMP
  - LVM PVlinks (no license req.)
  - StorageWorks SecurePath (future)
- file systems are mounted on individual members
  - cannot be shared amongst member systems
  - can use NFS to cross mount
- each node has its own system disk
  - support for multiple concurrent versions
  - rolling upgrade support

TruCluster Server
I/O access and infrastructure

- shared storage mode for all data and disks
- I/O multipath through base OS and TCR
  - integrated into operating system
  - no third-party or layered product used
  - use of DRD (device request dispatcher) within the cluster
- all file systems are shared amongst all members
  - option to enable member-specific access
  - create member-specific files and directories using CDSL
- single-system disk and directory tree
  - support for multiple concurrent versions (up to two)
  - rolling upgrades supported
I/ O infrastructure

- **TruCluster Server** can reroute I/ O requests within the cluster transparent to the application
  - uses DRD (device request dispatcher)
  - nodes with no direct connectivity can access other members' storage (such as CD-ROMs and tapes)
  - can be used for file system I/ O and raw device access

- **MC/ServiceGuard** and **TruCluster** both provide a clusterized volume manager
  - Single volume manager in TruCluster Server
    - CLSM (optional)
  - Multiple VM with MC/ServiceGuard
    - shared read/ write (for raw devices)
      - CVM
      - SLVM
    - exclusive read/ write
      - VxVM
      - LVM

[Image of the slide]
### hw mgr(8)

```
jmh@janix# hw mgr -view devices -cluster
```

<table>
<thead>
<tr>
<th>HWID</th>
<th>DSF Name</th>
<th>Model</th>
<th>Location</th>
<th>Member/Host</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>/dev/disk/floppy1c</td>
<td>3.5in</td>
<td>fdi0-unit-0</td>
<td>janix.zk3.dec.com</td>
</tr>
<tr>
<td>107</td>
<td>/dev/disk/cdrom1c</td>
<td>RRD46</td>
<td>bus-0-targ-5-lun-0</td>
<td>janix.zk3.dec.com</td>
</tr>
<tr>
<td>108</td>
<td>/dev/disk/dsk25c</td>
<td>RZ1CB-CA</td>
<td>bus-1-targ-0-lun-0</td>
<td>janix.zk3.dec.com</td>
</tr>
<tr>
<td>109</td>
<td>/dev/disk/dsk26c</td>
<td>RZ1CB-CA</td>
<td>bus-1-targ-1-lun-0</td>
<td>janix.zk3.dec.com</td>
</tr>
<tr>
<td>19</td>
<td>/dev/disk/floppy0c</td>
<td>3.5in</td>
<td>fdi0-unit-0</td>
<td>oddjob.zk3.dec.com</td>
</tr>
<tr>
<td>24</td>
<td>/dev/disk/cdrom0c</td>
<td>RRD46</td>
<td>bus-0-targ-5-lun-0</td>
<td>oddjob.zk3.dec.com</td>
</tr>
<tr>
<td>25</td>
<td>/dev/disk/dsk0c</td>
<td>RZ1CB-CA</td>
<td>bus-1-targ-0-lun-0</td>
<td>oddjob.zk3.dec.com</td>
</tr>
<tr>
<td>26</td>
<td>/dev/disk/dsk1c</td>
<td>RZ1CB-CA</td>
<td>bus-1-targ-1-lun-0</td>
<td>oddjob.zk3.dec.com</td>
</tr>
<tr>
<td>27</td>
<td>/dev/disk/dsk2c</td>
<td>HSZ50-AX</td>
<td>bus-2-targ-1-lun-0</td>
<td>oddjob.zk3.dec.com</td>
</tr>
<tr>
<td>27</td>
<td>/dev/disk/dsk2c</td>
<td>HSZ50-AX</td>
<td>bus-2-targ-1-lun-0</td>
<td>janix.zk3.dec.com</td>
</tr>
<tr>
<td>28</td>
<td>/dev/disk/dsk3c</td>
<td>HSZ50-AX</td>
<td>bus-2-targ-1-lun-1</td>
<td>oddjob.zk3.dec.com</td>
</tr>
<tr>
<td>28</td>
<td>/dev/disk/dsk3c</td>
<td>HSZ50-AX</td>
<td>bus-2-targ-1-lun-1</td>
<td>janix.zk3.dec.com</td>
</tr>
</tbody>
</table>

[...]

| 48   | /dev/disk/dsk23c   | SYMMETRIX IDENTIFIER=1002 | janix.zk3.dec.com |
| 48   | /dev/disk/dsk23c   | SYMMETRIX IDENTIFIER=1002 | oddjob.zk3.dec.com |

Locally attached devices on `janix`:

- `99: /dev/disk/floppy1c`
- `107: /dev/disk/cdrom1c`
- `108: /dev/disk/dsk25c`
- `109: /dev/disk/dsk26c`

Locally attached devices on `oddjob`:

- `19: /dev/disk/floppy0c`
- `24: /dev/disk/cdrom0c`
- `25: /dev/disk/dsk0c`
- `26: /dev/disk/dsk1c`

Shared devices:

- `27: /dev/disk/dsk2c`
- `28: /dev/disk/dsk3c`
- `48: /dev/disk/dsk23c`
device request dispatcher
DRD

- provides transparent, highly available access to all disks and tapes in the cluster
  - device names are consistent using Tru64 UNIX® V5 naming mechanism
- used for all disk access in a cluster
- LSM, file systems, databases, applications
- supports block and character devices
- provides direct access and served I/O models
- `drdmg` command for status and management
- no equivalent in MC/ServiceGuard today
DRD direct access and served i/o

- Host A
- Host B
- Host C
- Host D

- Disk 1
- Disk 2
- Disk 3
- Disk 4
- Disk 5
- Disk 6

- Array controller A
- Array controller B

- Direct Access
- Cluster interconnect
- Served i/o
- Network

- Bus A
- Bus B
DRD I/O for locally attached device

I/O for nonlocal devices is served
logical storage/ volume managers

MC/ ServiceGuard
- cluster support is available through CVM or SLVM
  - separately licensed
- shared access to raw volumes
  - used by Oracle9i RAC and Oracle8i OOPS
- raw device only support
- nodes must enable a given disk group for clusterwide access
- same management commands as on a standalone system
- sometimes you must enter a command multiple times (once on each member)

TruCluster Server
- cluster support is built into LSM
  - CLSM code enabled if running in a cluster
  - license for mirroring and GUI
- shared access to LSM volumes
- raw device and file systems
- configuration changes can be made from any cluster member
- fully symmetric design
- same management interface as LSM on a single system
CLSM and CVM in a cluster environment

all I/O for raw device access is issued directly to the storage subsystem
cluster subsystem overview

file system access and availability
application availability and integration
cluster networking and client access
cluster internal services
file system access in a cluster

MC/ServiceGuard

- an individual file system is mounted by one member in the cluster
  - no clusterwide mounts and file system
- support for VxFS, HFS, NFS, CDFS and lofs
- NFS support for client and server access
  - a cluster node can be an NFS client to another member in the same cluster
  - allows sharing of file systems

TruCluster Server

- cluster file system (CFS) mounts each file system clusterwide
  - each member can see and (potentially) access every mounted file system
  - additional options to restrict access to the CFS “server”
- support for AdvFS, UFS, NFS, MFS, and DvDFS
- cache coherent clusterwide
  - including mmap()’d files
- NFS support for client and server
MC/ServiceGuard local file system access

- access to local file systems only through one member
- can use NFS for sharing
MC/ ServiceGuard local file system access

access to a local file system using NFS for the other member in the cluster

network/NFS interconnect

cluster interconnect

storage interconnect

/home node A

I/O

/home node B

I/O

/home node C

I/O
TruCluster cluster file system access

- access to file systems using CFS
- for ‘normal’ file system I/O, CFS client/server concept
- for AdvFS directIO access, no CFS client/server concept used
- added CCR feature in 5.1A to bypass client/server model for certain I/O patterns
CFS file system access
node recovery

- transparent transition of CFS server
- no interruption for file system I/O

![Diagram of CFS file system access and node recovery]

/home
node A
C R S
D F

/home
node B
CFS Clin
DRD

/home
node C
CFS Srv
DRD

hub

cluster interconnect

storage interconnect
• both clustering solutions provide a comprehensive failover framework
  – MC/ServiceGuard and cm (cluster monitor)
  – TruCluster Server uses CAA (cluster application availability)

• applications can be integrated through scripts and APIs

• extensive collection of example scripts and third-party application integration (MC/ServiceGuard)

• failure recovery either on the local node or by moving the application to another (functional) node in the cluster

• can have dependencies between applications to force ordered startup/shutdown/relocation
  – integrated in TruCluster Server via CAA
  – separate scripting toolkit with MC/ServiceGuard
**MC/ServiceGuard application packages**

- easy to configure framework
- flexible and easy to manage application resources
- up to 16 node clusters
- up to 200 application packages

Processes:
- app_Process_1
- app_Process_2
- middleware_1
- middleware_2

Disks:
- lvol_data1
- lvol_data2

Network:
- IP 16.141.8.115

Member A

pkg A
pkg B
pkg C
TruCluster CAA application resources

- easy to configure and extensible framework
- central management for applications in a TruCluster
- up to 8 node clusters
- hundreds of application resources

application name
placement policy

required resources
optional resources
application processes
subnet connectivity

member A

resource A

resource B

resource C
application failover frameworks

MC/ ServiceGuard
- packages
- assign IP address to package
- storage configuration kept in ServiceGuard configuration files
- provides a CLI (/usr/sbin/cm*) and graphical user interface
- extensive cluster toolkit in ECM T (Enterprise Cluster Master Toolkit)
  - databases and applications
  - NFS
- integrated SAP R/3 solution with SGeSAP Toolkit
  - optionally with Sommersault MC/ SGeRAC
- Oracle9i RAC on ServiceGuard

TruCluster CAA
- resources
- can use cluster alias instead of IP alias (no CAA management needed)
- no storage associated with CAA resources (ease of management)
  - storage failover/ access part of base OS and TruCluster
- CLI (/usr/sbin/caa_*) and GUI through SysMan station
- example scripts for many applications provided as part of the CAA framework
- Oracle9i RAC integration with hp DButility
cluster administration

- cluster installation and addition of cluster members
- both solutions provide a framework to manage the cluster and its services
- can run on multiple platforms
  - Linux, UNIX® (HP-UX, Tru64)
  - Windows®
  - web-based interface available
- TruCluster extends single system commands to cluster
  - all management actions can be launched from SysMan
- administration frameworks
  - ServiceGuard Manager
  - System Management Station
previous methods:
• manage as many
• poor scalability
• high availability with complexity

TruCluster Server
• increased scalability, high availability, and simplified management
• manage any system and storage, anywhere as one
• scale out easily — add a node in minutes
• simplified high availability
other clustering solutions approach it this way...

A – install UNIX® & patches
install apps & patches
configure storage

B – install UNIX & patches
install apps & patches
configure storage

C – install UNIX & patches
install apps & patches
configure storage

D – install UNIX & patches
install apps & patches
configure storage

E – install UNIX & patches
install apps & patches
configure storage

F – install UNIX & patches
install apps & patches
configure storage

G – install UNIX & patches
install apps & patches
configure storage
TruCluster Server works this way...

- install OS & patches
- configure storage
- execute clu_create
- install apps & patches
- execute clu_add_member for each node

A
B
C
D
E
F
G
H

/ root
/ usr
/ apps

controllers
controllers

fc switches

cluster interconnect

shared storage
cluster file system
Serviceguard Manager example
Sysman Station example

Contains a list of the monitored subsystems and displays their status.

Events:
named: named is in the FAILED state. Please use the Event Viewer to obtain specific information.

Press the right mouse button or select this object and press the tools menu to perform management operations.
<table>
<thead>
<tr>
<th>Command</th>
<th>CLI Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmviewcl</td>
<td>caa_stat</td>
<td>Provides status on the current state of the cluster members and services</td>
</tr>
<tr>
<td>cmmodpkg</td>
<td>caa_profile</td>
<td>Manages resource/package attributes</td>
</tr>
<tr>
<td>cmmakepkg</td>
<td>caa_register</td>
<td>Registers a resource/package</td>
</tr>
<tr>
<td>cmrunpkg</td>
<td>caa_start</td>
<td>Starts a resource/package</td>
</tr>
<tr>
<td>cmhaltpkg</td>
<td>caa_relocate</td>
<td>Relocates an application</td>
</tr>
<tr>
<td>cmhaltpkg</td>
<td>caa_unregister</td>
<td>Removes an application from CAA control</td>
</tr>
</tbody>
</table>
cluster networking

common goals

• provide highly available client network access

• cluster interconnect for heartbeat and intercluster communication
cluster interconnects

**MC/ServiceGuard**
- LAN for heartbeat and data/client access
  - transparent failover using APA
- high-speed messaging interconnect using **HyperFabric**
  - up to 2.4 GB/s bandwidth
  - can be shared between clusters
  - transparent failover
  - fiber and copper hardware
  - used in Oracle9i RAC and Oracle8i OPS configurations
  - still require LAN for heartbeat
- HMP for high-speed message passing
  - no adapter failover with HMP in HyperFabric

**TruCluster Server**
- LAN for cluster communication and client access
  - transparent failover through NetRAIN
- High-speed interconnect using **Memory Channel**
  - 100 MB/s
  - very low latency
  - transparent failover built in
  - specialized API
  - fiber and copper hardware
- RDG for high-speed message passing
  - failover built-in via CI
Both products provide network failover for clients. They use IP aliases assigned to interfaces. IP alias moves with application. Can monitor for connectivity to IP address/subnet. In ServiceGuard, part of the package configuration process. In TruCluster Server, use cluster alias or IP alias as part of the resource definition. Cluster alias can be configured:
- as virtual IP address
- to provide support for single and multi-instance services
multiple cluster aliases

Cluster Alias A (default alias)

Cluster Alias B

Node X

Node Y

Node Z

Client

Client 2

Multi-instance service

Single-instance service
cluster internal services

**MC/ ServiceGuard**
- membership management
- LAN-based heartbeat
- TOC for hanging nodes
  - safety timer to crash hung nodes to prevent data corruption
  - updates to `/dev/kepd` through `cmclid`

**TruCluster Server**
- CNX (connection manager)
- ICS (intercluster communication subsystem)
- KGS (kernel group services)
- DLM (distributed lock manager)
cluster daemons
(not a complete list)

MC/ServiceGuard daemons
- **cmclld**
  - the *main* ServiceGuard daemon
- **cmclconfd**
  - used to configure and start the cluster
- **cmsrvassistld**
  - minor assist to cmclconfd
- **cmlvmd**
  - allows “exclusive” VG activation
- **cmui**
  - SAM enhancements for SvcGd configuration
- **OpenView Network Node Manager**

TruCluster daemons
- **caad**
  - the *main* CAA (application services) daemon
- **aliasd/gated**
  - central daemons used by cluster networking
- **[icssvr_]* ICS**
  - ICS threads
- **cfsd (as of 5.1B)**
  - daemon providing load balancing information for CFS
- **vold**
  - LSM daemon
- **smsd/esmd/smauth/evmd**
  - daemons for the system management framework
- **clu_mibs**
  - cluster SNMP daemon
cluster integration with workload management

- unique to MC/ServiceGuard
- currently no integration between Aurema/ARM tech and TruCluster Server
  - can cluster hardware partitions
HA for partitioned systems

**HyperPlex**
- hard partitions with multiple nodes

**nPartitions**
- hard partitions within a node

**Virtual Partitions**
- within a hard partition

**PRM with psets**
- resource partitions within a single OS image

---

- OS image with HW isolation
- OS image with HW isolation
- OS image with HW isolation
- OS image with SW isolation
- OS image with SW isolation
- OS image with SW isolation

---

**hp-ux wlm**
- automatic goal-based resource allocation through set SLOs

---

- Isolation
  - highest degree of separation

- Flexibility
  - highest degree of dynamic capabilities

---

- Application 1 with guaranteed compute resources
- Application 2 with guaranteed compute resources
- Application n with guaranteed compute resources

---

Based on CPUs or percentages
workload manager (WLM) and process resource manager (PRM) in an MC/ServiceGuard cluster

- service-level management solutions that can be easily integrated into an MC/ServiceGuard environment to provide management of computing resources (CPU, memory and I/O utilization) according to desired SLOs both before and after application failover among nodes
- applications can be assigned maximum resources when they run on their “preferred” server
- after a failover to a different node, the resource utilization for the existing and new applications can be dynamically reset to:
  - favor important applications
  - restrict resource utilization of less important applications
TruCluster technology into hp-ux
hp-ux 11i

(D.H. Brown 2002)

- hp-ux 11i is ranked #1 in all five categories
- Tru64 UNIX is ranked #1 in scalability and systems management

#1 scalability
#1 reliability, availability, and serviceability
#1 systems management
#1 internet and web application services
#1 directory and security services
- strong leadership
- high-end scalability
- mission-critical availability
- manageability
- workload management
- security
enhancing hp-ux with Tru64 UNIX® technologies

current plans include:

• TruCluster Server software
• Advanced File System (AdvFS)
• Other select technologies
the transition from MC/ServiceGuard today to HP-UX fused with TruCluster technology
what must I do as I prepare to embark on this path?

continue your standard planning incorporating this new information.
MC/SG Oracle9i® Real Application Cluster edition today

**Public Network**

**Cluster Interconnect**

- **Node 1**
  - Oracle executables
  - system root, /usr, and /var
  - Oracle data files (raw)

- **Node 2**
  - Oracle executables
  - system root, /usr, and /var

- **Node 3**
  - Oracle executables
  - system root, /usr, and /var

- **Node n**
  - Oracle executables
  - system root, /usr, and /var

The diagram illustrates the architecture and components of a MC/SG Oracle9i® Real Application Cluster edition.
TruCluster Server
Oracle9i RAC today

the “vision” for 11.31

public network

cluster interconnect

node 1

member specific “boot” area (/vmunix, swap, ...)

node 2

cluster common root, /usr, & /var

node 3

Oracle executables

node n

Oracle data files (AdvFS file system)
HP-UX MC/ ServiceGuard based customer

• MC/ ServiceGuard packages will be able to migrate
  • potential for migration tools
• potential to reuse underlying storage infrastructure
• if application runs on MC/ ServiceGuard (or even a standalone HP-UX system) it WILL run on an HP-UX/ TruCluster system
• opportunity to consider new/ different storage architecture
  • no local storage needed
  • may want to consider consolidating applications and servers
  • drastic improvement for manageability and scalability
Tru64 UNIX®/ TruCluster based customer

- TruCluster customers will be able to migrate existing CAA applications
- potential to reuse underlying TruCluster storage infrastructure
- opportunity to consider new technologies
  - integrated workload management (such as WLM, PRM, and vPARs)
  - Tru64 to HP-UX transition tools will be available
    - script compatibility
    - warning/flagging of Tru64 specifics
helpful resources

TruCluster website
http://www.tru64unix.compaq.com/cluster/

MC/ServiceGuard website
http://www.hp.com/products1/unix/highavailability/

HP-UX/ServiceGuard docs
http://docs.hp.com

Tru64 UNIX®/TruCluster documentation
http://www.tru64unix.compaq.com/docs/

Tru64 UNIX® information site
http://www.tru64.org/