

Exploring the High Availability Storage Infrastructure.

Tutorial 323 – Brainshare 2007

Jo De Baer

Technology Specialist

Novell - jdebaer@novell.com

Novell[®]

Agenda

- The High Availability Storage Infrastructure
 - very brief overview
- Resources
 - where to get information
- Setup example
 - remainder of the session time
- What's new in SP1 ?
 - if time permits !
- Questions & Answers
 - if time permits -> ask questions during the session !

The High Availability Storage Infrastructure

High Availability Storage Infrastructure^N

- Enterprise Volume Management (EVMS2)
 - Cluster-aware volume manager
 - Single, unified system for handling all storage management tasks
 - Unparalleled flexibility and extensibility
- Integrates Oracle Cluster File System 2 (OCFS2)
 - Symmetrical parallel cluster file system
 - Optimized for fast access to large files
 - Joint development with Oracle
- Heartbeat 2 Clustering Services
 - Powerful resource dependency model based on XML
 - Modular design with new cluster resource manager
 - Large clusters: 16 nodes tested, no inherent limit
 - Resources actively monitored for health



Business Objectives

- Open Source enterprise-quality high availability storage infrastructure to support mission-critical workloads in the data center
- Help customers to reduce cost in their data center workloads
- Increase market share by inclusion in product (no additional subscription required)
- Key ISV & IHV solution stacks certifications
- Enticing to OEMs

High Availability Storage Infrastructure Advantages

- High Availability Storage Infrastructure is part of the distribution, no extra cost
- Key certifications and industry support (IBM, Oracle, etc.)
- Broadest number of File Systems supported
 - greater flexibility, better performance of applications/services
- Technically superior integrated H/A solution
 - other Linux vendors are also moving their solution to userspace in future to follow

High Availability Storage Infrastructure Use Cases

- Cluster managed, relocatable virtual machines
 - Storage Infrastructure = Virtualized Storage Systems
 - Integrated Solution of Virtual Storage and Virtual Machines
- Oracle RAC
 - Scalable Oracle Database deployments
- Line of Business Applications
 - High Availability for Strategic Applications using Heartbeat 2 and EVMS2
 - SAP Business Applications (White papers coming)

The background of the slide is a solid green color with a pattern of diagonal stripes in varying shades of green, creating a sense of movement and depth. The stripes are oriented from the top-left to the bottom-right.

Resources

High Availability Storage Infrastructure Documentation

- <http://www.novell.com/documentation/sles10/index.html>
- Heartbeat 2
 - <http://www.linux-ha.org/HeartbeatTutorials>
 - Documentation for Heartbeat 2 is in process of being improved
- OCFS2
 - <http://oss.oracle.com/projects/ocfs/documentation/>
- EVMS2
 - http://sourceforge.net/docman/?group_id=25076
- Setup example :

http://wiki.novell.com/index.php/SUSE_Linux_Enterprise_Server#High_Availability_Storage_Infrastructure

The background of the slide is a solid green color with a pattern of diagonal stripes in a lighter shade of green, creating a textured, layered effect.

Setup example

Goal of the setup

To create a virtual machine which is kept H/A in a two node cluster. When the physical node that houses the virtual machines crashes, the virtual machine is restarted on the other physical node.

If time permits : we add a second one-node cluster inside the virtual machine which monitors the service that is offered by the virtual machine. If the service fails, this second cluster resets the virtual machine.

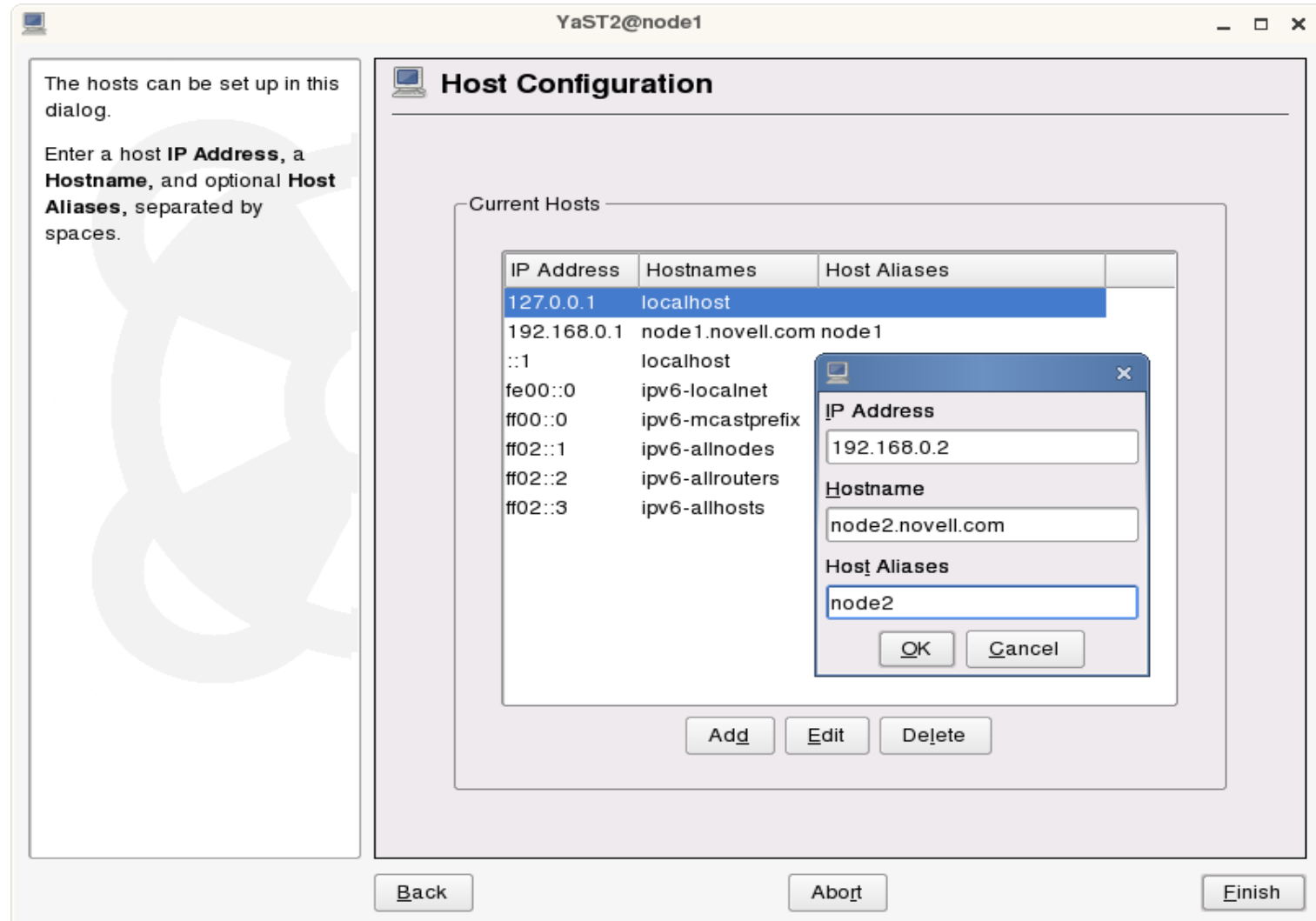
What will we do ?

- Preparation
 - NTP, hostname resolution, ...
- Configuring Heartbeat 2
 - sanity check with simple resource
- Configuring iSCSI
- Configuring OCFS2
 - setting user space managed heartbeat
 - configuring Heartbeat 2 STONITH
- Configuring Xen
 - creating a virtual machine with sync enabled loop module
- Integrating a Xen virtual machine as a cluster resource

If time permits

- Configuring a layered cluster
 - single node cluster inside the virtual machine which monitors the service that is offered by the virtual machine

Preparations



Heartbeat 2

The screenshot displays the Linux HA Management Client window. The interface is divided into several sections:

- Top Bar:** Contains tabs for 'Connection', 'Resources', and 'Nodes', along with a toolbar with icons for adding, removing, and refreshing resources.
- Left Panel:** A tree view showing the hierarchy of the HA setup:
 - linux-ha (Status: with quorum)
 - Nodes
 - node1 (Status: running)
 - node2 (Status: running(dc))
 - resource_test (Status: running on [node2])
 - Resources
 - resource_test (Status: running on [node2]) - This item is selected and highlighted in blue.
 - Constraints
 - Places
 - Orders
 - Colocations
- Right Panel:** Shows details for the selected resource 'resource_test', which is currently running on [node2].
 - Attributes:**
 - Resource ID: resource_test
 - Type: IPAddr
 - Class: ocf
 - Provider: heartbeat
 - Parameters:** A table listing configuration parameters:

name	value
description	
restart_type	ignore
resource_stickiness	0
is_managed	default
multiple_active	stop_start
- Bottom Bar:** Shows the connection status 'Connected to 127.0.0.1' and buttons for 'Apply' and 'Reset'.

iSCSI

YaST2@node1

Partition your hard disks...

This is intended for **experts**. If you are not familiar with the concepts of hard disk **partitions** and how to use them, you might want to go back and select **automatic** partitioning.

Nothing will be written to your hard disk until you confirm all your changes with the "Apply" button. Until that point, you can safely abort.

For LVM setup, using a non-LVM root device and a non-LVM swap device is recommended. Other than the root and swap devices, you should have partitions managed by LVM.

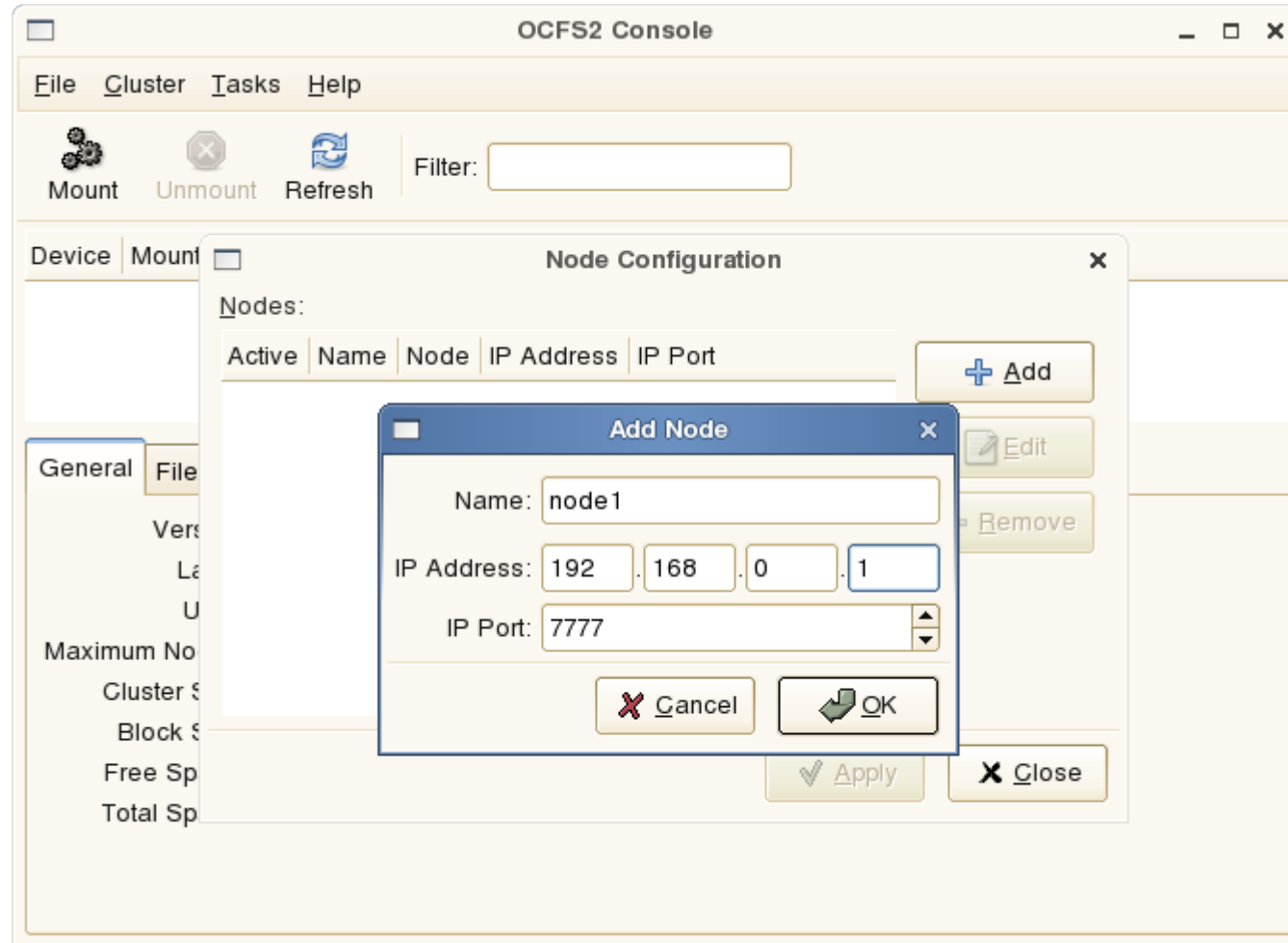
The table to the right shows the current partitions on all your hard disks.

Hard disks are designated like this

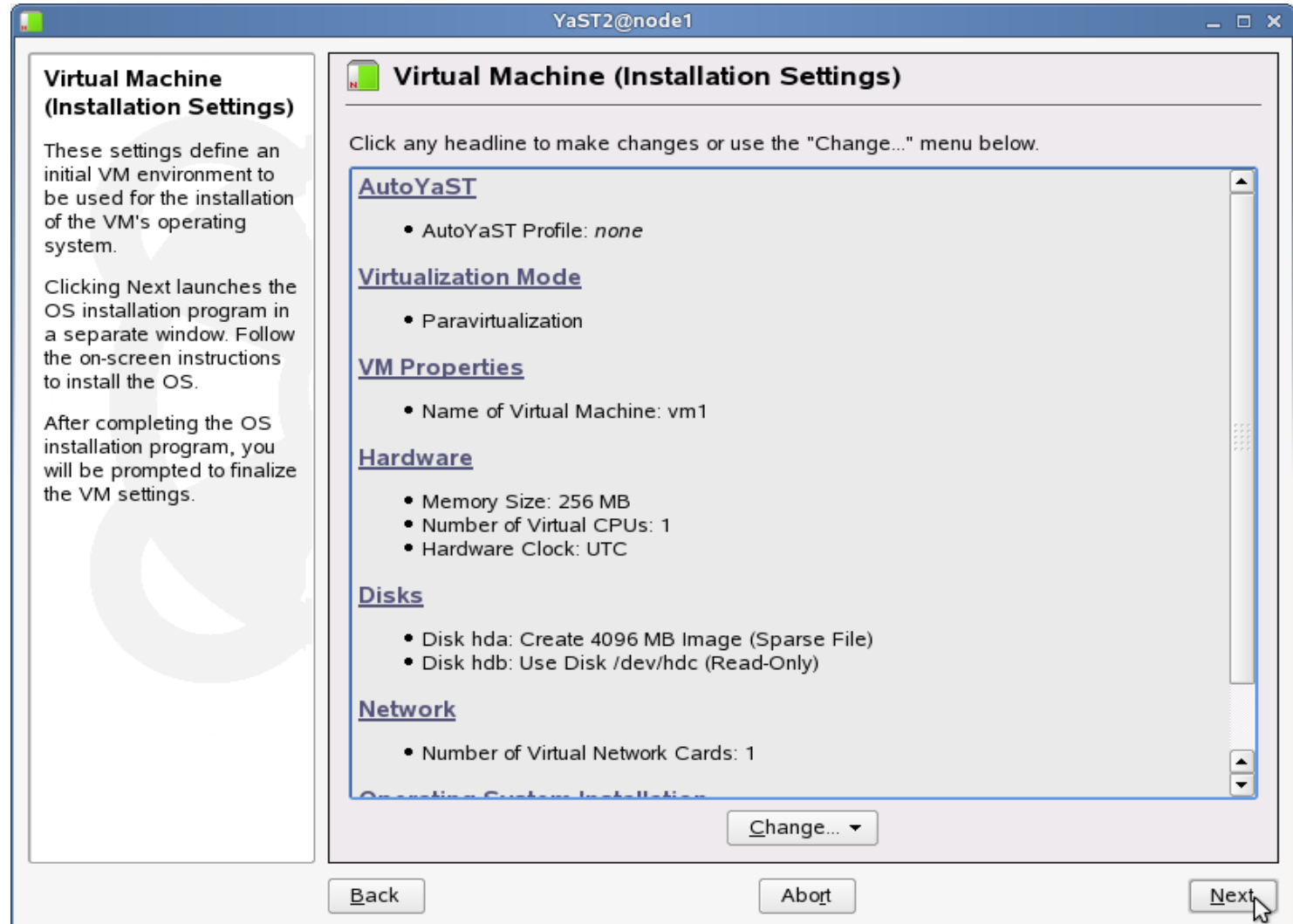
Expert Partitioner

Device	Size	F	Type	Mount	Mount By	Start	End	Us
/dev/hda	37.2 GB		HITACHI_DK23FB-40			0	4863	
/dev/hda1	1.0 GB		Linux swap	swap	K	0	130	
/dev/hda2	15.0 GB		Linux native	/	K	131	2089	
/dev/hda3	21.2 GB		Linux native			2090	4863	
/dev/sda	21.2 GB		IET-VIRTUAL-DISK			0	2773	
/dev/sda1	20.0 GB		Linux native			0	2610	
/dev/sda2	502.0 MB		Linux native			2611	2674	
/dev/sda3	776.5 MB		Linux native			2675	2773	

OCFS2



Xen



Where to find the used XML files

- <http://wiki.novell.com>
- SUSE[®] Linux Enterprise Server
 - High Availability Storage Infrastructure
 - > Exploring the High Availability Storage Infrastructure
 - > Conquering the High Availability Storage Infrastructure
 - » **Example setup files**

The background of the slide is a solid green color with a pattern of diagonal stripes in a lighter shade of green, creating a sense of motion or energy.

What's new in SP1?

SUSE® Linux Enterprise Server 10 SP1



- Values: Improved Robustness, Scalability, & Integration
- OCFS2 Update (v1.2 or later)
- EVMS2
 - fixes for cluster scalability (2.5.6 or later)
 - resource agent for shared cluster container
- Heartbeat 2 fixes (2.0.7 or later)
- General updates to EXT3 and ReiserFS
- **Heartbeat-Xen VM Migration integration**
- Additional CIM management providers

The background of the slide is a solid green color with a pattern of diagonal stripes in varying shades of green, creating a sense of movement and depth. The stripes are most prominent on the right side and fade towards the left.

Questions and Answers

Novell®

Unpublished Work of Novell, Inc. All Rights Reserved.

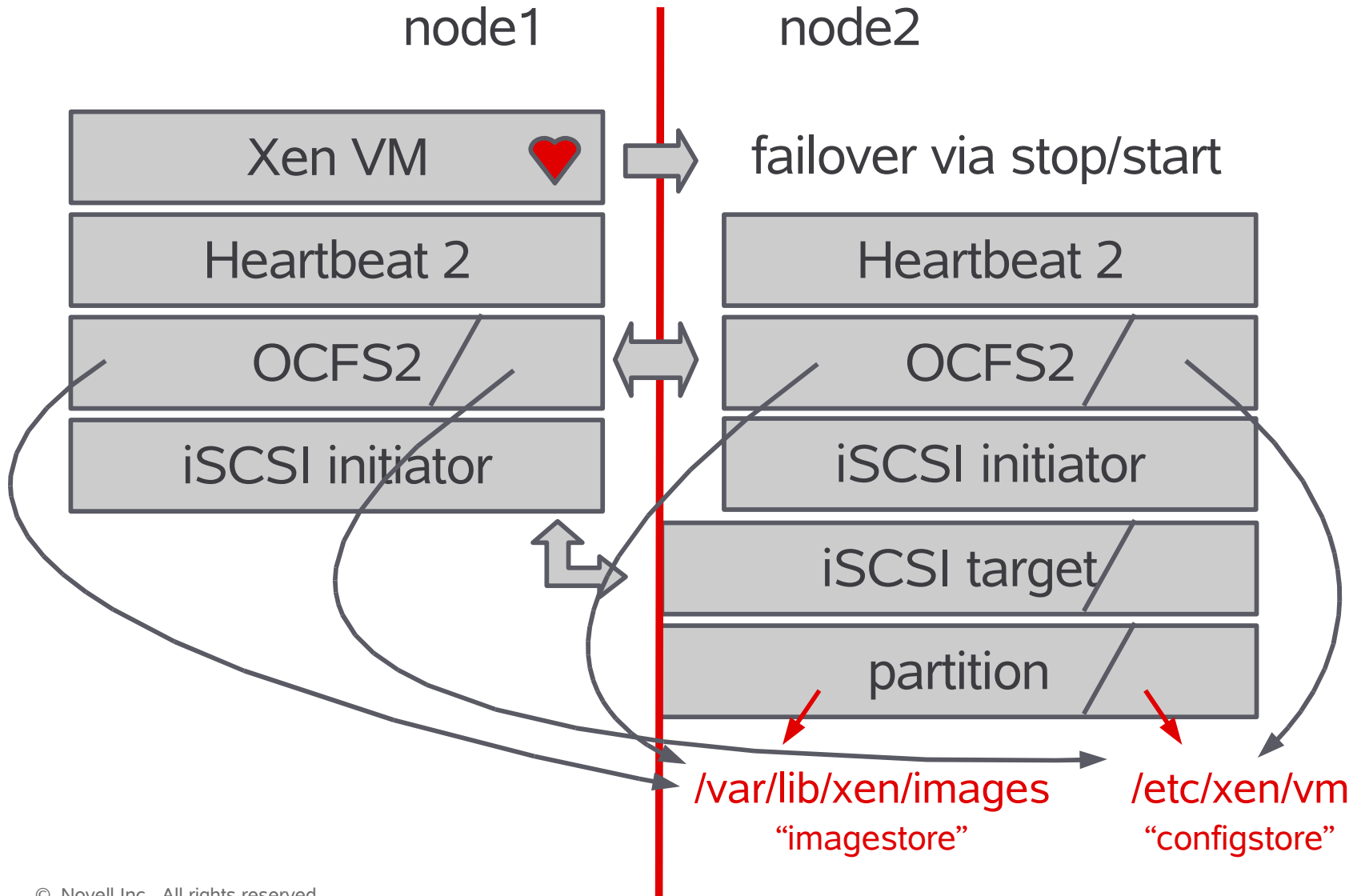
This work is an unpublished work and contains confidential, proprietary, and trade secret information of Novell, Inc. Access to this work is restricted to Novell employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of Novell, Inc. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

General Disclaimer

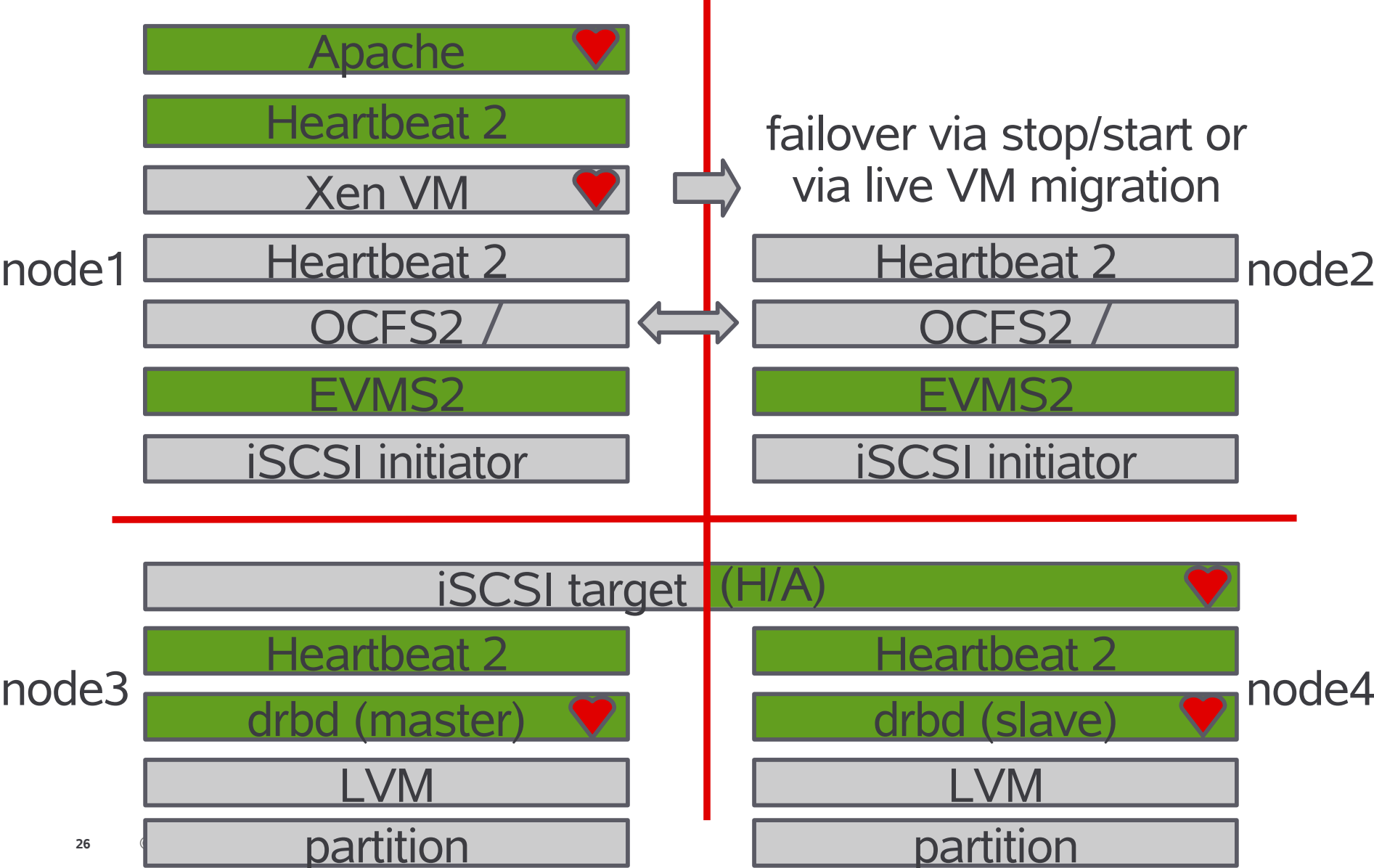
This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. Novell, Inc., makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for Novell products remains at the sole discretion of Novell. Further, Novell, Inc., reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All Novell marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners



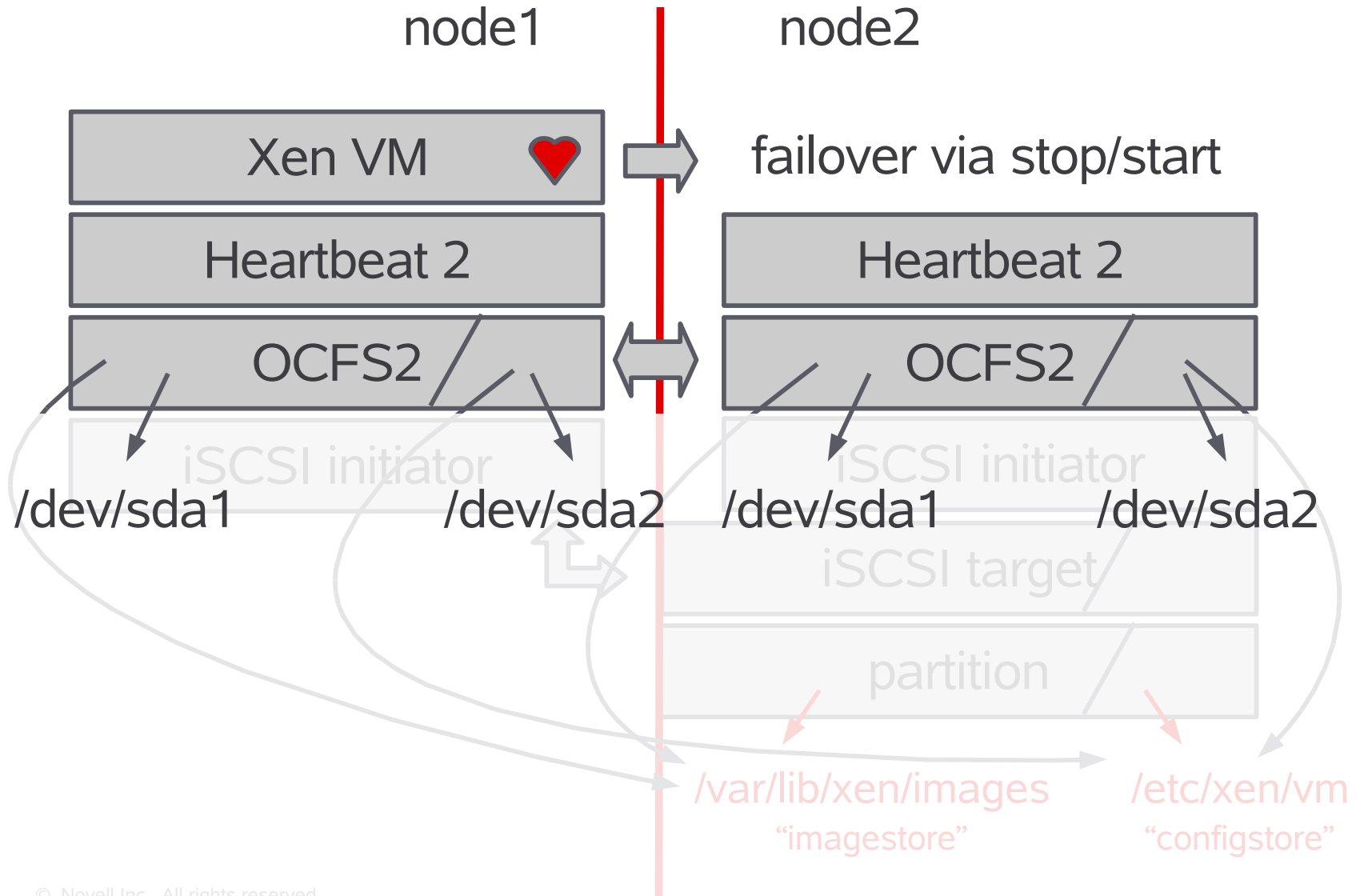
SLES 10 FCS (“Exploring the HASF”)



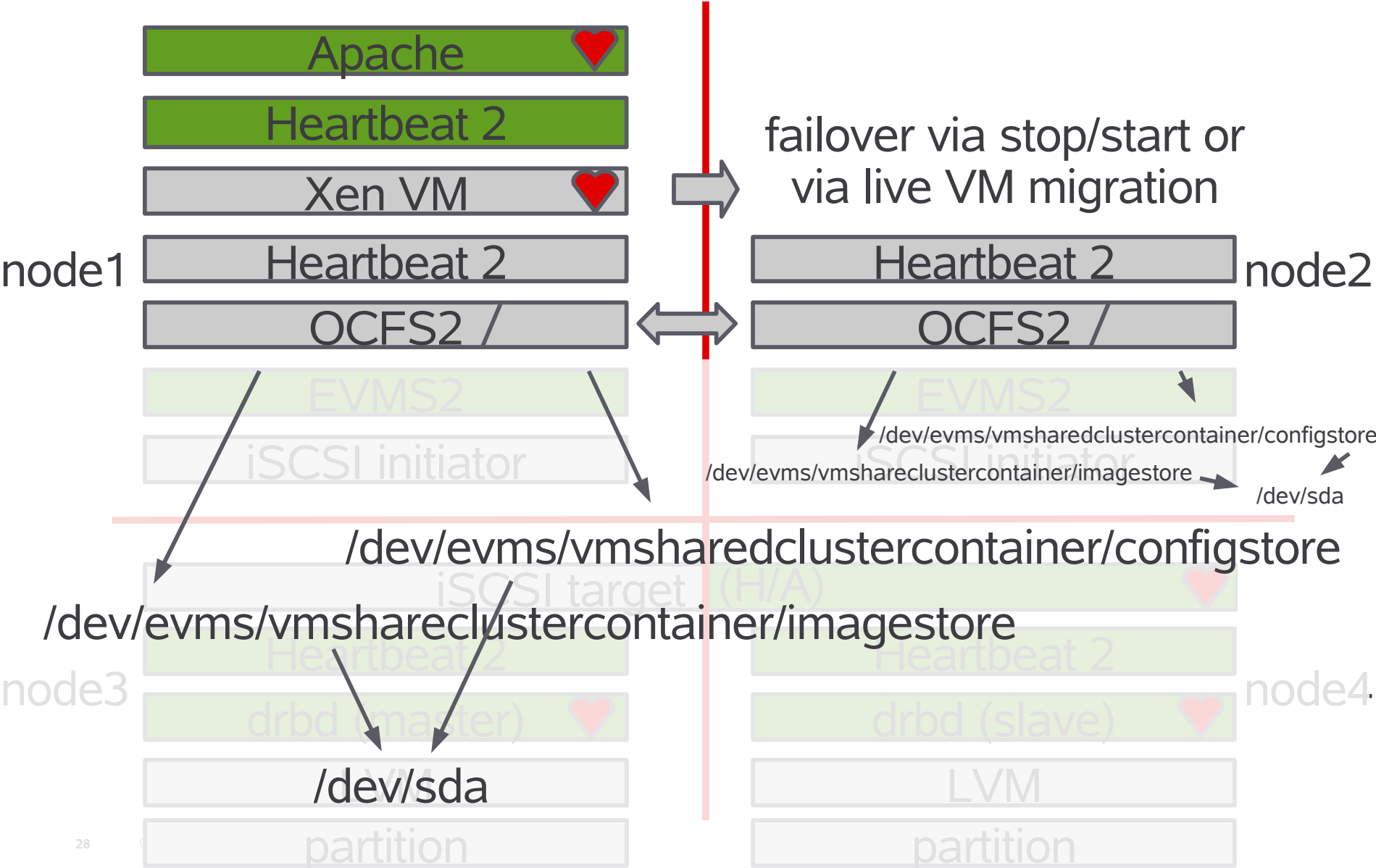
SLES 10 SP1 (“Conquering the HASI”)



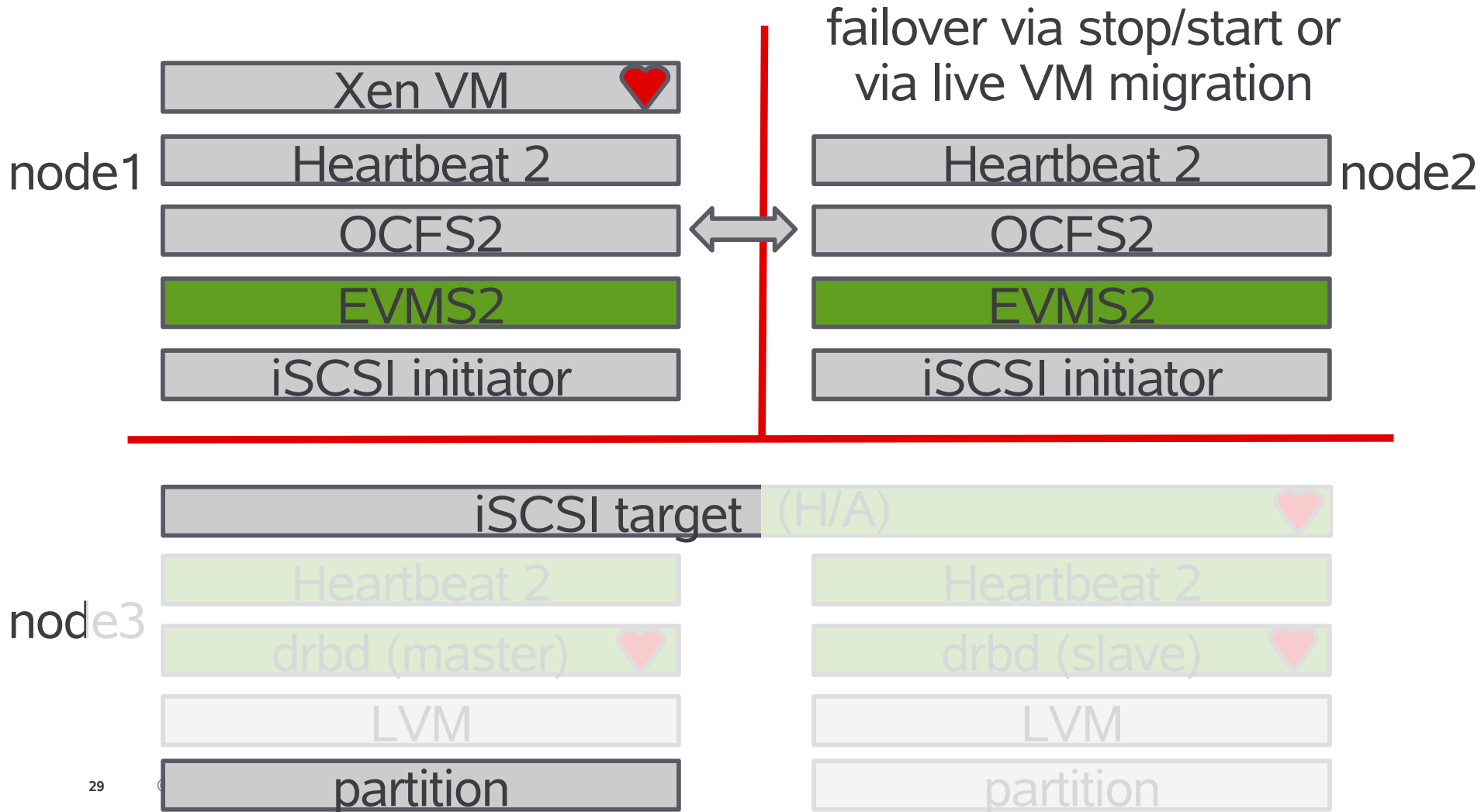
SLES 10 FCS (“Exploring the HASF”)



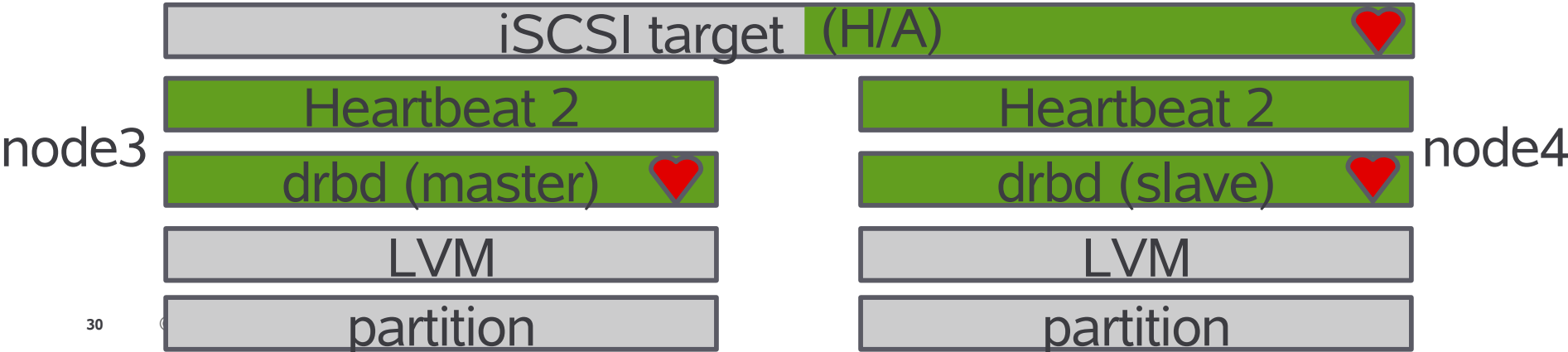
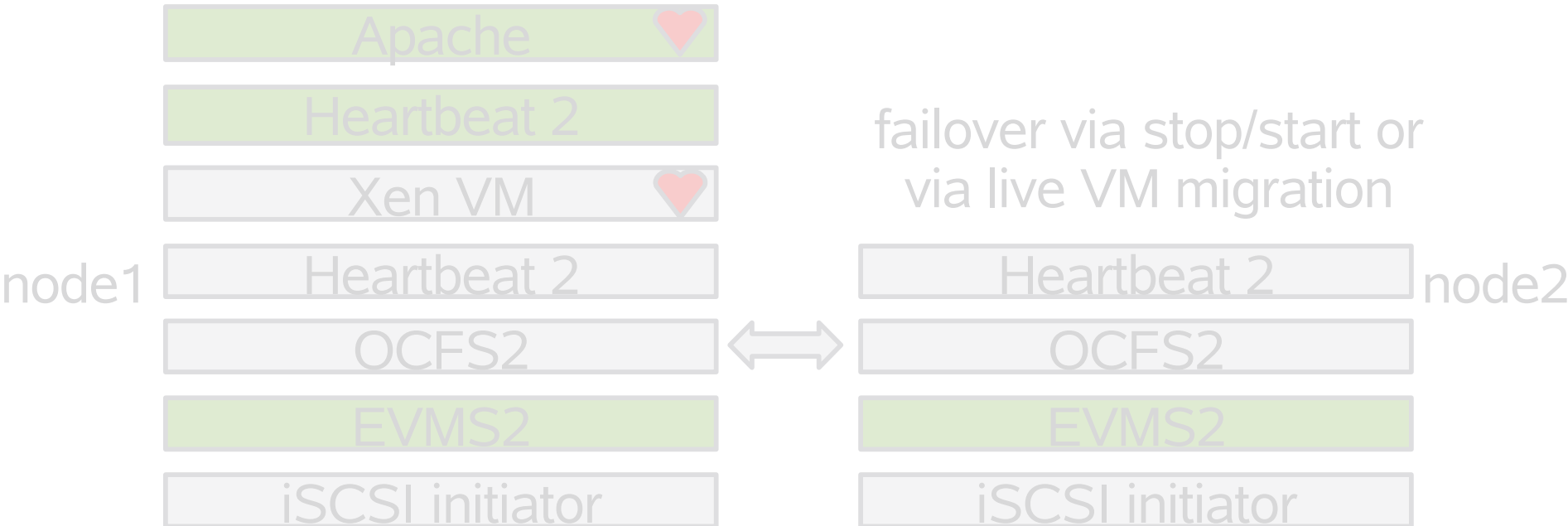
SLES 10 SP1 (“Conquering the HASI”)



TUT 323 – Brainshare 2007



ATT 320 – Brainshare 2007



cibbootstrap.xml

```
<cluster_property_set id="cibbootstrap">
  <attributes>
    <nvpair id="cibbootstrap-01" name="transition_idle_timeout" value="60"/>
    <nvpair id="cibbootstrap-04" name="stonith_enabled" value="true"/>
    <nvpair id="cibbootstrap-05" name="stonith_action" value="reboot"/>
    <nvpair id="cibbootstrap-06" name="symmetric_cluster" value="true"/>
    <nvpair id="cibbootstrap-07" name="no_quorum_policy" value="stop"/>
    <nvpair id="cibbootstrap-08" name="stop_orphan_resources" value="true"/>
    <nvpair id="cibbootstrap-09" name="stop_orphan_actions" value="true"/>
    <nvpair id="cibbootstrap-10" name="is_managed_default" value="true"/>
  </attributes>
</cluster_property_set>
```

```
# cibadmin -C -o crm_config -x ./cibbootstrap.xml
```

stonithcloneset.xml

```
<clone id="stonithcloneset" globally_unique="false">
  <instance_attributes id="stonithcloneset">
    <attributes>
      <nvpair id="stonithcloneset-01" name="clone_node_max" value="1"/>
    </attributes>
  </instance_attributes>
  <primitive id="stonithclone" class="stonith" type="external/ssh" provider="heartbeat">
    <operations>
      <op name="monitor" interval="5s" timeout="20s" prereq="nothing" id="stonithclone-op-01"/>
    </operations>
    <instance_attributes id="stonithclone">
      <attributes>
        <nvpair id="stonithclone-01" name="hostlist" value="node1,node2"/>
      </attributes>
    </instance_attributes>
  </primitive>
</clone>
```

```
# cibadmin -C -o resources -x ./stonithcloneset.xml
```

evmsdcloneset.xml

```
<clone id="evmsdcloneset" globally_unique="false">
  <instance_attributes id="evmsdcloneset">
    <attributes>
      <nvpair id="evmsdcloneset-01" name="clone_node_max" value="1"/>
    </attributes>
  </instance_attributes>
  <primitive id="evmsdclone" class="ocf" type="Evmsd" provider="heartbeat">
    <operations>
      <op name="monitor" interval="5s" timeout="20s" id="evmsdclone-op-01"/>
    </operations>
  </primitive>
</clone>
```

```
# cibadmin -C -o resources -x ./evmsdcloneset.xml
```

evmscloneset.xml

```
<clone id="evmscloneset" notify="true" globally_unique="false">
  <instance_attributes id="evmscloneset">
    <attributes>
      <nvpair id="evmscloneset-01" name="clone_node_max" value="1"/>
    </attributes>
  </instance_attributes>
  <primitive id="evmsclone" class="ocf" type="EvmsSCC" provider="heartbeat">
  </primitive>
</clone>
```

```
# cibadmin -C -o resources -x ./evmscloneset.xml
```

evms to evmsd order constraint

evmstoevmsdorderconstraint.xml

```
<rsc_order id="evmsdorderconstraints-01" from="evmscloneset" to="evmsdcloneset"/>
```

```
cibadmin -C -o constraints -x ./evmstoevmsdorderconstraint.xml
```

imagestorecloneset.xml

```
<clone id="imagestorecloneset" notify="true" globally_unique="false">
  <instance_attributes id="imagestorecloneset">
    <attributes>
      <nvpair id="imagestorecloneset-01" name="clone_node_max" value="1"/>
      <nvpair id="imagestorecloneset-02" name="target_role" value="started"/>
    </attributes>
  </instance_attributes>
  <primitive id="imagestoreclone" class="ocf" type="Filesystem" provider="heartbeat">
    <operations>
      <op name="monitor" interval="20s" timeout="60s" id="imagestoreclone-op-01"/>
      <op name="stop" timeout="60s" id="imagestoreclone-op-02"/>
    </operations>
    <instance_attributes id="imagestoreclone">
      <attributes>
        <nvpair id="imagestoreclone-01" name="device" value="/dev/evms/vmsharedclustercontainer/imagestore"/>
        <nvpair id="imagestoreclone-02" name="directory" value="/var/lib/xen/images"/>
        <nvpair id="imagestoreclone-03" name="fstype" value="ocfs2"/>
      </attributes>
    </instance_attributes>
  </primitive>
</clone>
```

cibadmin -C -o resources -x ./imagestorecloneset.xml

configstorecloneset.xml

```
<clone id="configstorecloneset" notify="true" globally_unique="false">
  <instance_attributes id="configstorecloneset">
    <attributes>
      <nvpair id="configstorecloneset-01" name="clone_node_max" value="1"/>
      <nvpair id="configstorecloneset-02" name="target_role" value="started"/>
    </attributes>
  </instance_attributes>
  <primitive id="configstoreclone" class="ocf" type="Filesystem" provider="heartbeat">
    <operations>
      <op name="monitor" interval="20s" timeout="60s" id="configstoreclone-op-01"/>
      <op name="stop" timeout="60s" id="configstoreclone-op-02"/>
    </operations>
    <instance_attributes id="configstoreclone">
      <attributes>
        <nvpair id="configstoreclone-01" name="device" value="/dev/evms/vmsharedclustercontainer/configstore"/>
        <nvpair id="configstoreclone-02" name="directory" value="/etc/xen/vm"/>
        <nvpair id="configstoreclone-03" name="fstype" value="ocfs2"/>
      </attributes>
    </instance_attributes>
  </primitive>
</clone>
```

© Novell 2011. All rights reserved. # cibadmin -C -o resources -x ./configstorecloneset.xml

*store to evms order constraints

imagestoretoevmsorderconstraint.xml

```
<rsc_order id="evmsorderconstraints-01" from="imagestorecloneset" to="evmscloneset"/>
```

```
cibadmin -C -o constraints -x ./imagestoretoevmsorderconstraint.xml
```

configstoretoevmsorderconstraint.xml

```
<rsc_order id="evmsorderconstraints-02" from="configstorecloneset" to="evmscloneset"/>
```

```
cibadmin -C -o constraints -x ./configstoretoevmsorderconstraint.xml
```

sles10.xml

```
<primitive id="sles10" class="ocf" type="Xen" provider="heartbeat">
  <operations>
    <op name="monitor" interval="10s" timeout="60s" id="xen-op-01"/>
    <op name="stop" timeout="60s" id="xen-op-02"/>
  </operations>
  <instance_attributes id="sles10_instance">
    <attributes>
      <nvpair id="xen-01" name="xmfile" value="/etc/xen/vm/sles10"/>
    </attributes>
  </instance_attributes>
  <meta_attributes id="sles10_meta">
    <attributes>
      <nvpair id="xen-02" name="allow_migrate" value="true"/>
    </attributes>
  </meta_attributes>
</primitive>
```

sles10location.xml

```
<rsc_location id="sles10_location" rsc="sles10">  
  <rule id="pref_sles10_location" score="INFINITY">  
    <expression attribute="#uname" operation="eq" value="node1"/>  
  </rule>  
</rsc_location>
```

```
# cibadmin -C -o constraints -x ./sles10location.xml
```

VM to *store order constraints

sles10toimagestoreorderconstraint.xml

```
<rsc_order id="sles10orderconstraints-01" from="sles10" to="imagestorecloneset"/>
```

```
cibadmin -C -o constraints -x ./sles10toimagestoreorderconstraint.xml
```

sles10toconfigstoreorderconstraint.xml

```
<rsc_order id="sles10orderconstraints-02" from="sles10" to="configstorecloneset"/>
```

```
cibadmin -C -o constraints -x ./sles10toimagestoreorderconstraint.xml
```