

TUT8155

# Best Practices: Linux High Availability with VMware Virtual Machines

**Jeff Lindholm**







**SUSE**

Sr. Systems Engineer

JLindholm@suse.com



# Agenda

-  SUSE® Linux Enterprise High Availability Extension 12
-  OS Level clustering use case
-  VMware configuration best practices
-  SUSE Linux Enterprise High Availability Extension / Linux Clustering in VMware
-  SUSE Linux Enterprise Server - High Availability Cluster Demo
-  Question / Answers

# Challenge

SUSE® Linux Enterprise High Availability Extension

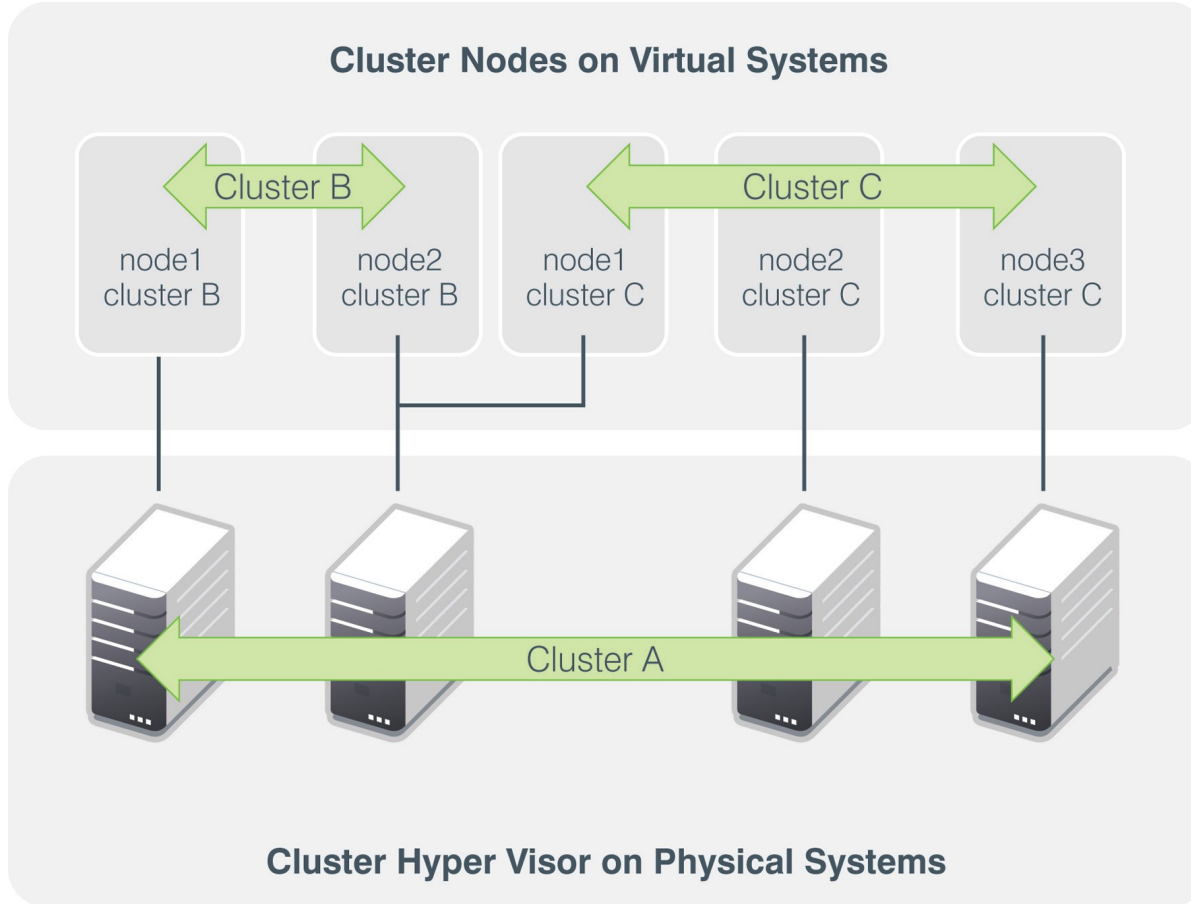
## Murphy's Law is Universal

- Faults will occur
  - Hardware crash, flood, fire, power outage, earthquake?
- Can you afford a service outage or worse, loss of data?
  - You might afford a five second blip, but can you afford a longer outage?
- How much does downtime cost?

**Can you afford low availability systems?**

# Use Case: Linux Clustering

SUSE® Linux Enterprise High Availability Extension



# Version 12 – Key Features

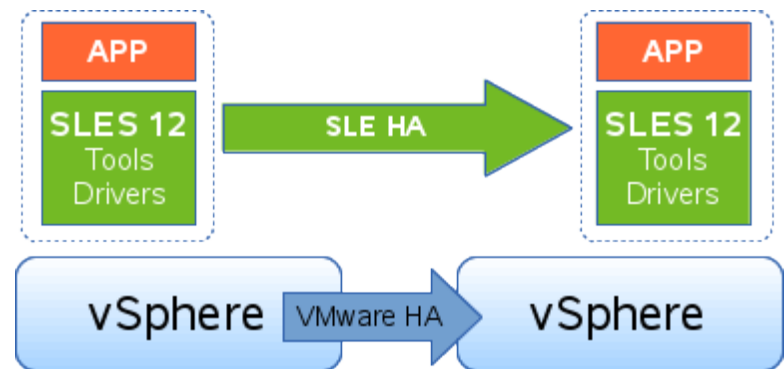
SUSE® Linux Enterprise High Availability Extension

- Major code refresh to latest upstream versions
- Pacemaker
  - Object tagging
  - Significant CIB performance
- Cluster Shell:
  - Health evaluation
  - Improved error reporting and syntax
  - Support corosync configuration
- hawk
  - Improved wizards
  - History explorer
- Geo extension
  - Improved algorithm
  - Per-site attributes in CIB
  - DNS-based IP fail-over
- GFS2 now supported in r/w mode
- New, additional fence-agents



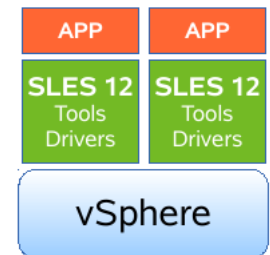
# SUSE Linux Enterprise High Availability Extension + VMware

- **SUSE Linux Enterprise High Availability Extension** complements VMware host-level HA solution for mission critical applications
- Features
  - Application level HA protects active memory contents
  - Scripts for monitoring open source services (eg, Apache, MySQL, NFS, PostgreSQL, Tomcat, KVM, Xen) and 3<sup>rd</sup> party applications (eg, SAP, Oracle, IBM DB2, WebSphere)
  - Policy-driven cluster resource manager
  - Cluster-aware file system and volume management
  - Continuous data replication
  - User-centric management tools



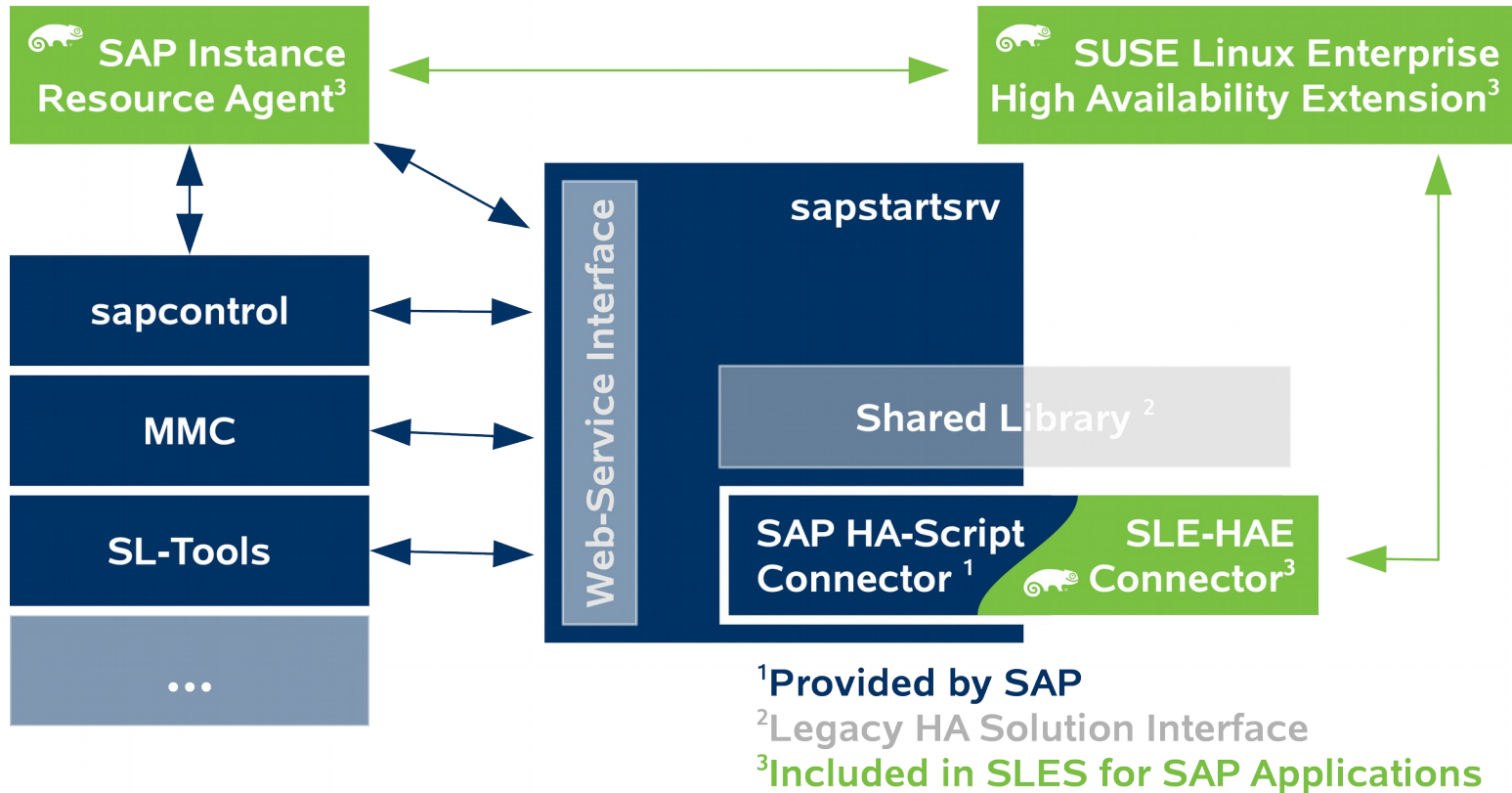
# Optimized vSphere Guest Performance

- VMware tools and drivers integrated with SUSE Linux Enterprise Server 12 for best out-of-the-box experience
  - **open-vm-tools**: eliminates the need to separately install VMware Tools and reduces operational expenses and virtual machine downtime
  - **vmware\_balloon**: physical memory management driver
  - **vmw\_vmci**, **vmw\_vsock**: provide for fast and efficient communications between guest virtual machines and hypervisors
  - **vmxnet3**: next generation of a paravirtualized NIC designed for performance
  - **vmw\_pvscsi**: driver for paravirtualized SCSI device which improves disk performance
  - **vmwgfx**: kernel driver for 3D graphics
- Fully supported by VMware via L3 support agreement



# Example: SAP HA Cluster Interface

Interfaces to integrate our HA solution in SAP

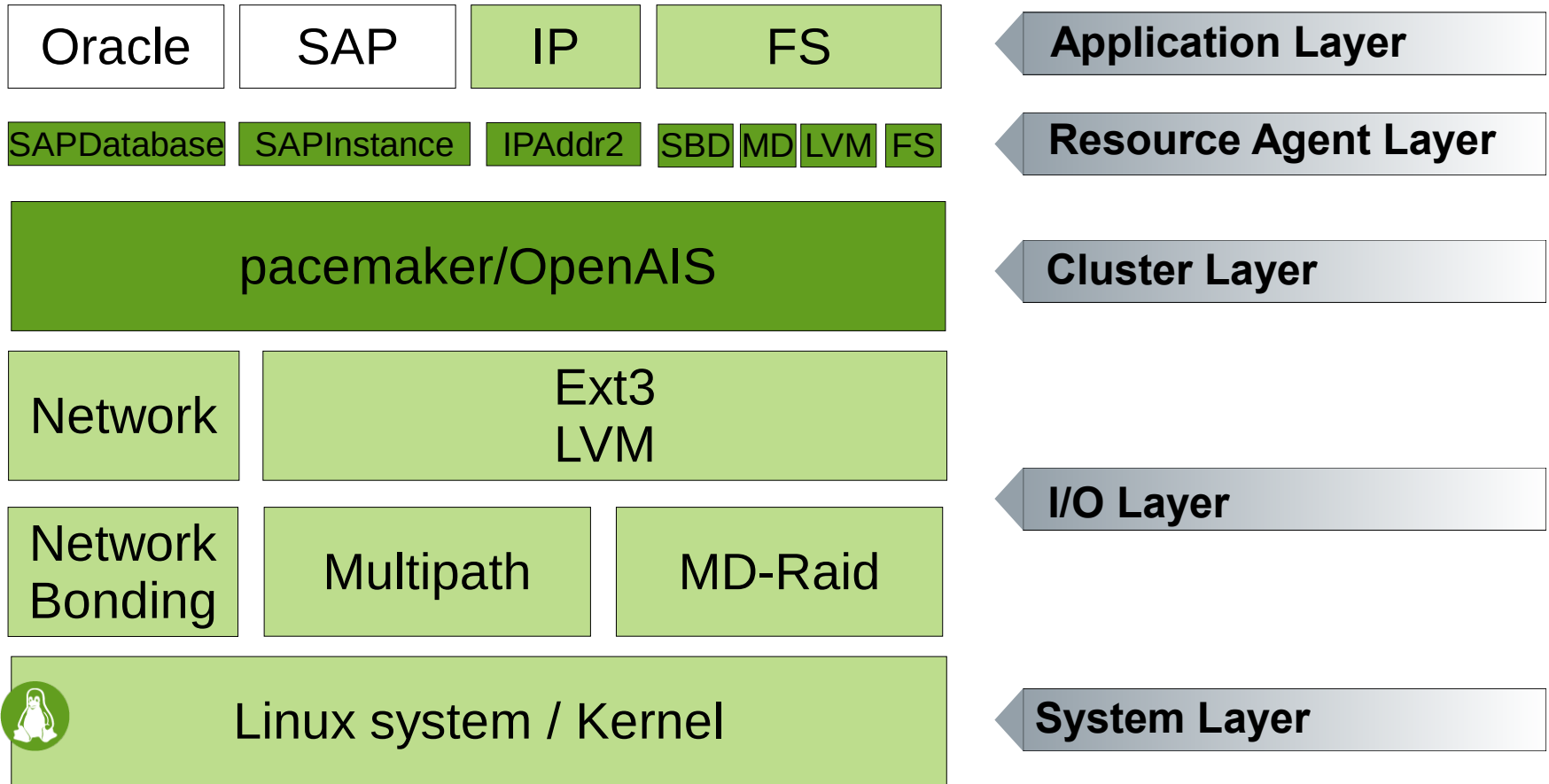


<http://scn.sap.com/docs/DOC-25453>





# Example: HA Stack for SAP



# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

- \* Both SLE HA Nodes running on ESX server 1
- \* ESX Server 3 is powered down

DB  
OS

APP  
SCS  
OS

APP  
OS

APP  
OS

APP  
OS

VMware HA and DRS Cluster

VMware ESX

VMware ESX

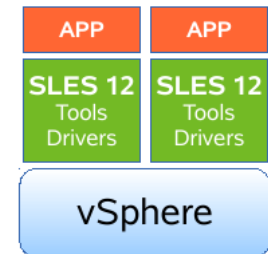
(VMware ESX)



# SUSE Linux Enterprise 11/12 VMware Virtual Machines Configuration Best Practices

# Optimized vSphere Guest Performance

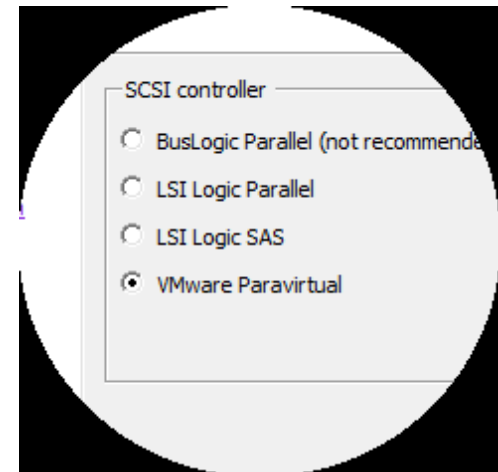
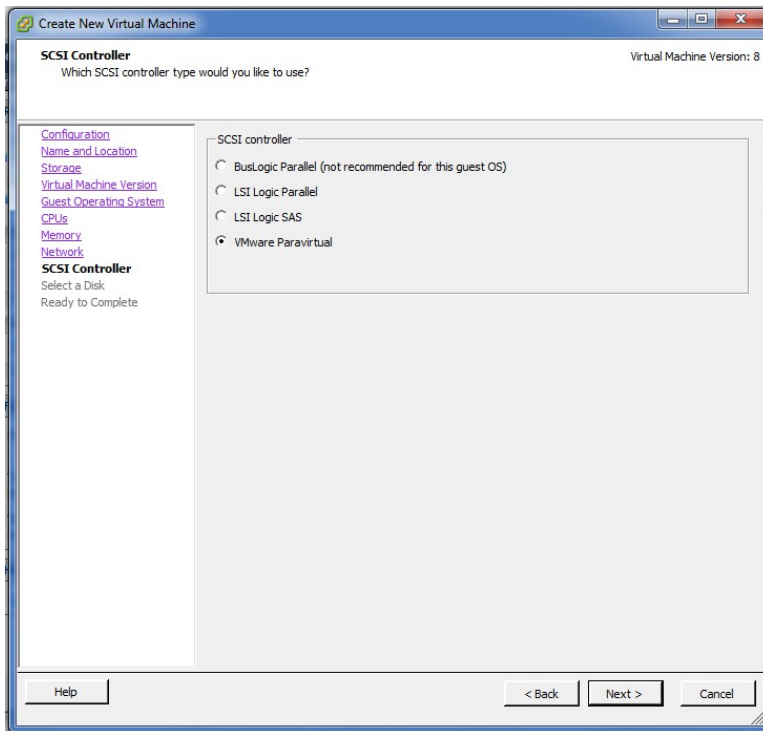
- VMware tools and drivers integrated with SUSE Linux Enterprise Server 12 for best out-of-the-box experience
  - **open-vm-tools**: eliminates the need to separately install VMware Tools and reduces operational expenses and virtual machine downtime
  - **vmware\_balloon**: physical memory management driver
  - **vmw\_vmci**, **vmw\_vsock**: provide for fast and efficient communications between guest virtual machines and hypervisors
  - **vmxnet3**: next generation of a paravirtualized NIC designed for performance
  - **vmw\_pvscsi**: driver for paravirtualized SCSI device which improves disk performance
  - **vmwgfx**: kernel driver for 3D graphics
- Fully supported by VMware via L3 support agreement



# Virtual Disk Configuration

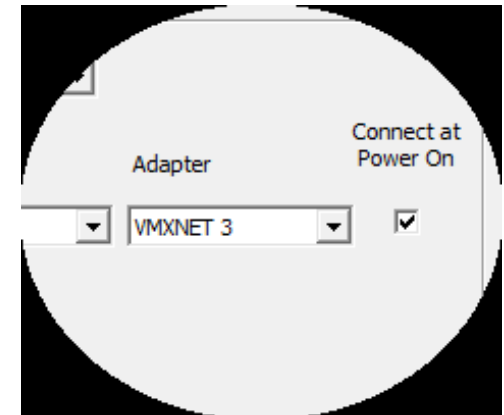
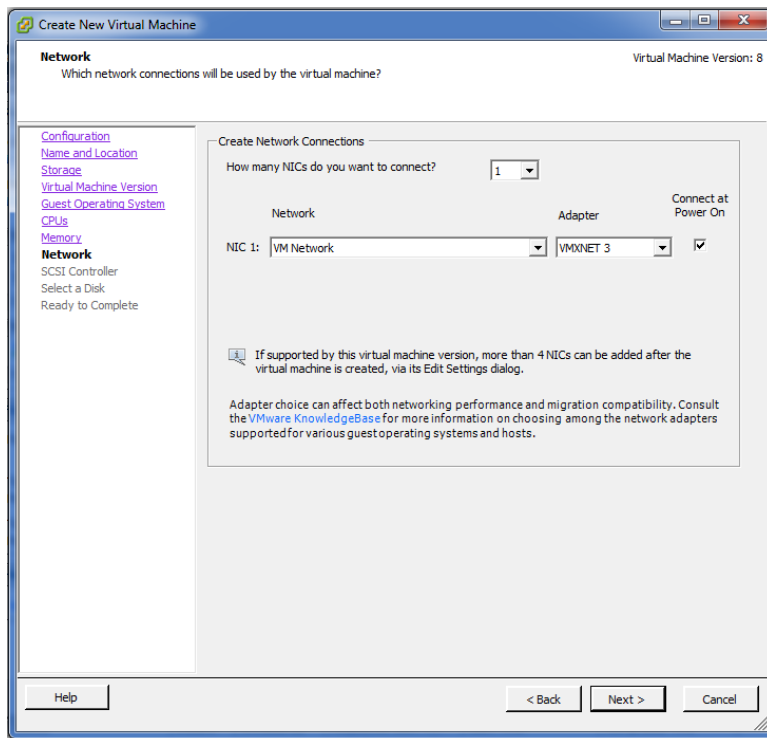
VMware Para-virtual SCSI drivers (vmw\_pvscsi) are included with SUSE Linux Enterprise Server 11 and 12

Para-virtual SCSI drivers are recommended when using SAN datastore configurations



# Virtual Network Configuration

VMware vmxnet3 network drivers are default, recommended, and built into both SUSE Linux Enterprise Server 11 and 12



# VMware / SUSE Linux Enterprise High Availability Best Practices

# Clustering with VMware

- SUSE Linux Enterprise High Availability Extension on VMware is supported by SUSE
- Fencing is accomplished by Stonith Block Device (SBD)
- Unicast heartbeat configuration is recommended for two node configurations
- Mixed physical and virtual cluster nodes are supported
  
- Shared Storage using SCSI Raw Device Maps to VM
  - Or -
- VMFS Datastore with simultaneous write protection disabled



# The Dos and Don'ts

Things you should consider



Keep cluster configuration simple



Use SBD for node fencing (STONITH)



Define and perform tests for all failure scenarios



Follow our best practices

# The Dos and Don'ts

Things you should avoid



Build Cluster cluster without node fencing (STONITH)



Go live without tests planned and done



Go live without proper operations manual



Cluster resource (like SBD and STONITH) timings shorter than SAN timings

# Considerations for SBD / Shared Storage on VMware ESXi datastores

- Disable Simultaneous write protection for shared disk devices: (multi-writer flag)
  - <http://kb.vmware.com/kb/1034165>
- Enable by-id disk presentation inside the virtual machine:
  - Add `disk.EnableUUID = "TRUE"` to cluster node `.vmx` config files
- Enable `softdog` module for SBD operation in `boot.local` prior to initial cluster setup / installation: (each node)
  - `echo 'modprobe softdog' >> /etc/init.d/boot.local`

# Multi-writer Flag Supported and Unsupported Actions or Features:

Actions or Features	Supported	Unsupported	Notes
Power on, off, restart virtual machine	√		
Suspend VM		x	
Hot add virtual disks	√		Only to existing adapters
Hot remove devices	√		
Hot extend virtual disk		x	
Connect and disconnect devices	√		
Snapshots		x	Virtual backup solutions leverage snapshots via the vStorage APIs; for example, VMware Data Recovery, vSphere Data Protection. These are also not supported.
Snapshots of VMs with independent-persistent disks	√		Supported in vSphere 5.1 update2 and later versions
Cloning		x	
Storage vMotion		x	
Changed Block Tracking (CBT)		x	
vSphere Flash Read Cache (vFRC)		x	Stale writes can lead to data loss and/or corruption
vMotion *		x	

\* = Migration of disks in multi-writer mode is supported only for Oracle RAC clusters. For more information, see the [Oracle Databases on VMware vSphere® 5 RAC Workload Characterization Study \(VMware VMFS\)](#) guide.

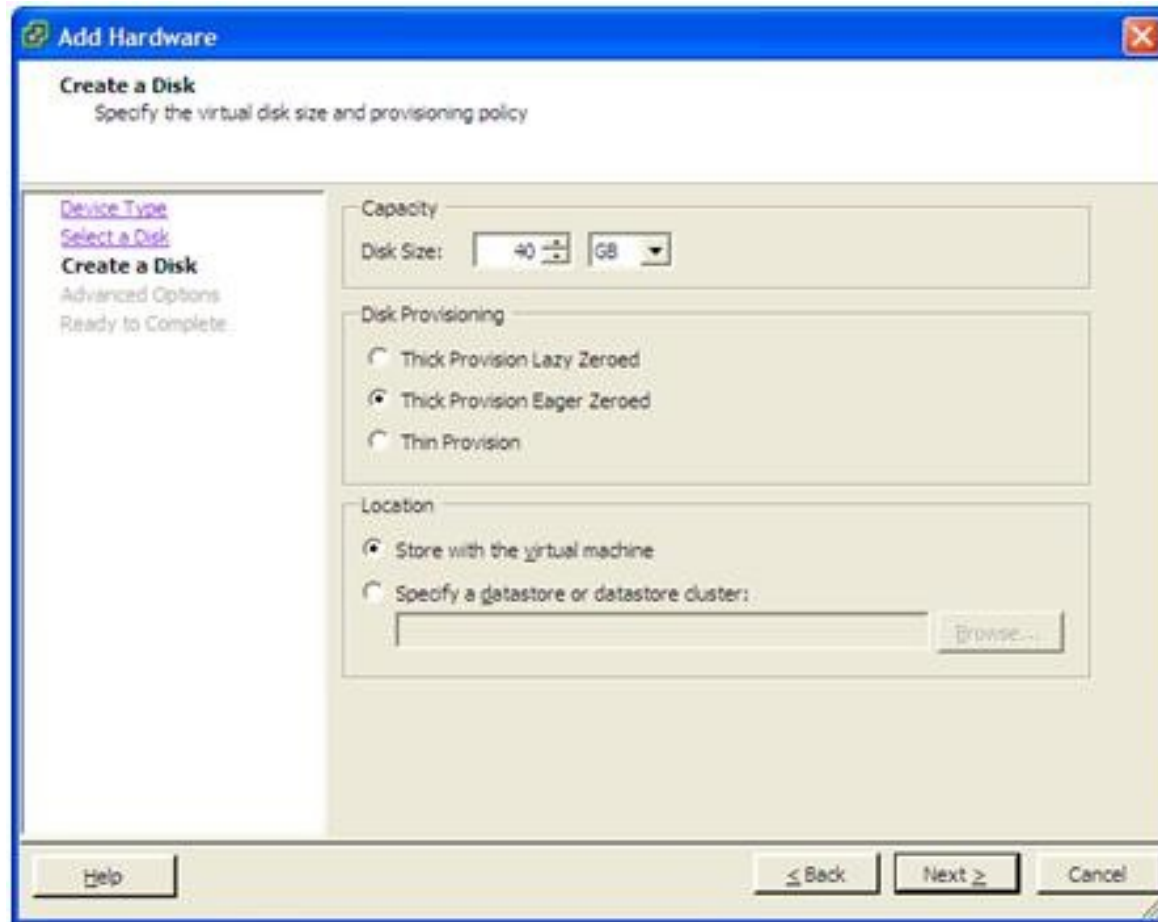
- <http://kb.vmware.com/kb/1034165>

# Other multi-writer Limitations

- When using the multi-writer mode, the virtual disk must be eager zeroed thick; it cannot be zeroed thick or thin provisioned. For more information, see A virtual machine fails to power on with the error: Thin/TBZ disks cannot be opened in multiwriter mode. VMware ESX cannot open the virtual disk for clustering. (1033570). <http://kb.vmware.com/kb/1033570>
- Sharing is limited to 8 ESXi/ESX hosts with VMFS-3 (vSphere 4.x) and VMFS-5 (vSphere 5.x) in multi-writer mode. On ESXi 5.x with VMFS-5, you can still share the virtual disks with 32 hosts for read-only access (that is, for View, linked clone, and fast provisioning use cases)
- Hot adding a virtual disk removes Multi-Writer Flag. For more information, see Hot adding a virtual disk in ESXi 5.5 removes the multi-writer flag (2078540). <http://kb.vmware.com/kb/2078540>

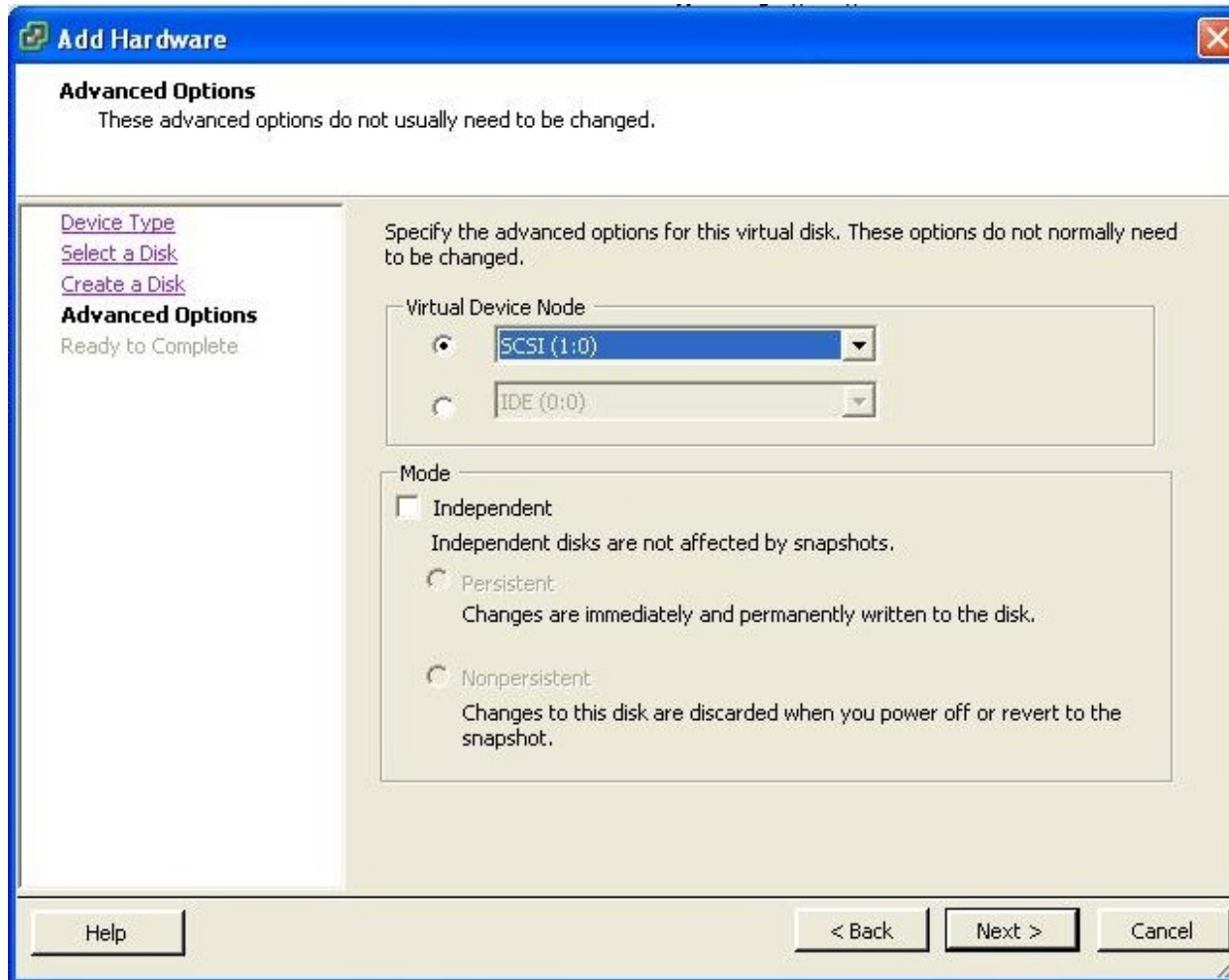
# Multi-Writer Configuration

Thick Provisioned Eager Zeroed Disk is Required



# Multi-Writer Configuration

Take note of the virtual device node setting SCSI (1:0)



# Multi-Writer Configuration

Add the multi-writer setting for each virtual disk that you want to share. For example, to share four disks, the configuration file entries look like this:

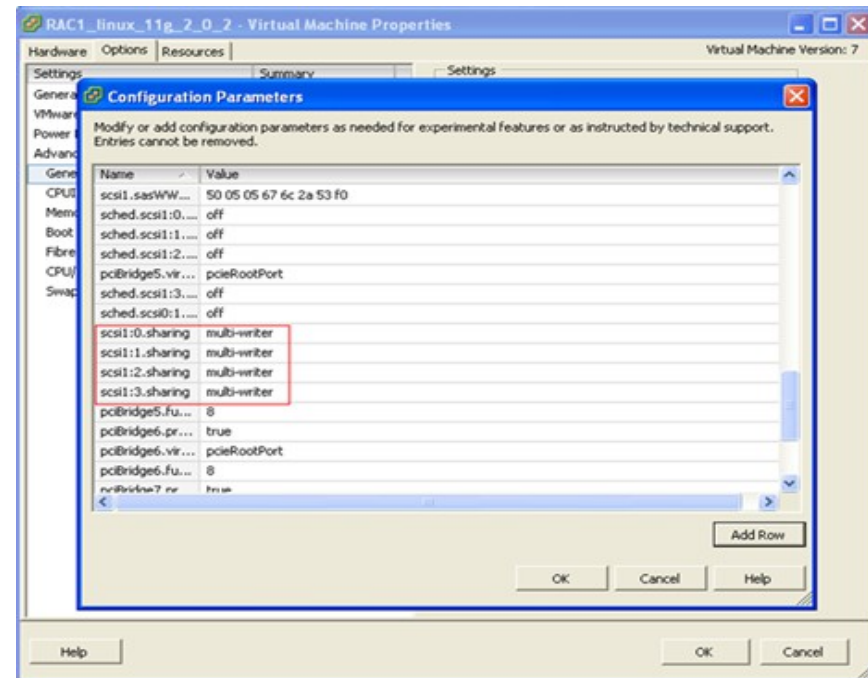
```
scsi1:0.sharing = "multi-writer"
```

```
scsi1:1.sharing = "multi-writer"
```

```
scsi1:2.sharing = "multi-writer"
```

```
scsi1:3.sharing = "multi-writer"
```

**Edit the vmx configuration file or change the configuration parameters using the vSphere client or web administration interface...**





# Disabling Simultaneous Write Protection on VMware ESXi

- Cluster ready storage configuration and disk management are **REQUIRED** to avoid multiple nodes concurrently mounting shared storage on boot
- To disable auto-activation of cluster / shared storage volumes on boot – disable boot.lvm and/or edit /etc/sysconfig/lvm to specify what LVM volume groups are activated at boot vs. activated by the cluster software
- Optional – The OCFS2 file system includes a distributed lock manager and will safely allow multiple cluster nodes to concurrently block mount shared storage (Max 32 nodes are supported by SUSE, Max 8 nodes RW limitation per VMware)

# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

- \* Both SLE HA Nodes running on ESX server 1
- \* ESX Server 3 is powered down

DB  
OS

APP  
SCS  
OS

APP  
OS

APP  
OS

APP  
OS

VMware HA and DRS Cluster

VMware ESX

VMware ESX

(VMware ESX)



# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

\* VM is migrated to ESX server 2 without  
\* SLE HA cluster interference

DB  
OS

APP  
SCS  
OS



APP  
SCS  
OS

APP  
OS

APP  
OS

APP  
OS

VMware HA and DRS Cluster

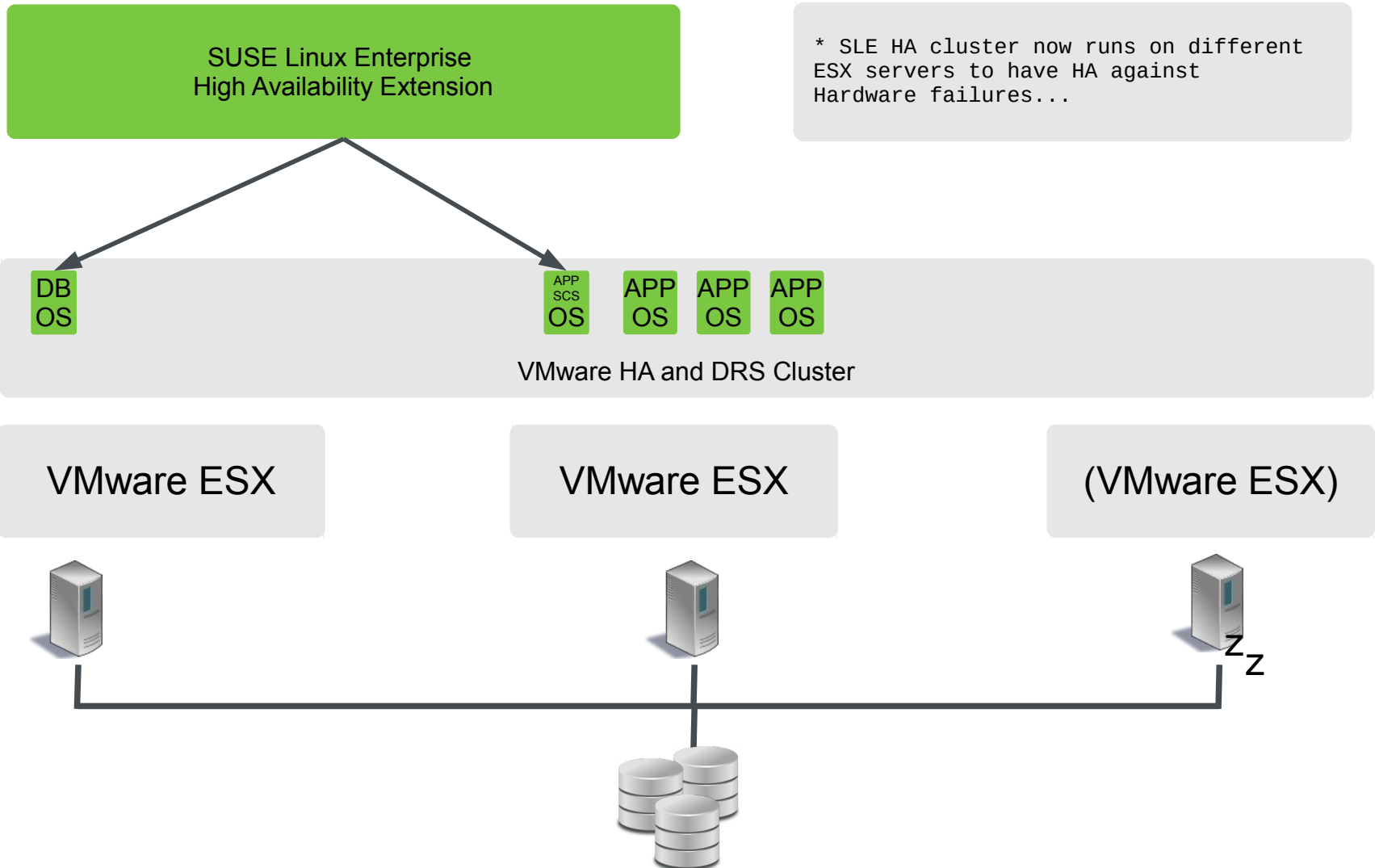
VMware ESX

VMware ESX

(VMware ESX)



# VMware HA and SUSE Linux Enterprise High Availability Extension



# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

\* SLE HA cluster now runs on different  
ESX servers to have HA against  
Hardware failures...

DB  
OS

APP  
SCS  
OS

APP  
OS

APP  
OS

APP  
OS

VMware HA and DRS Cluster

VMware ESX

VMware ESX

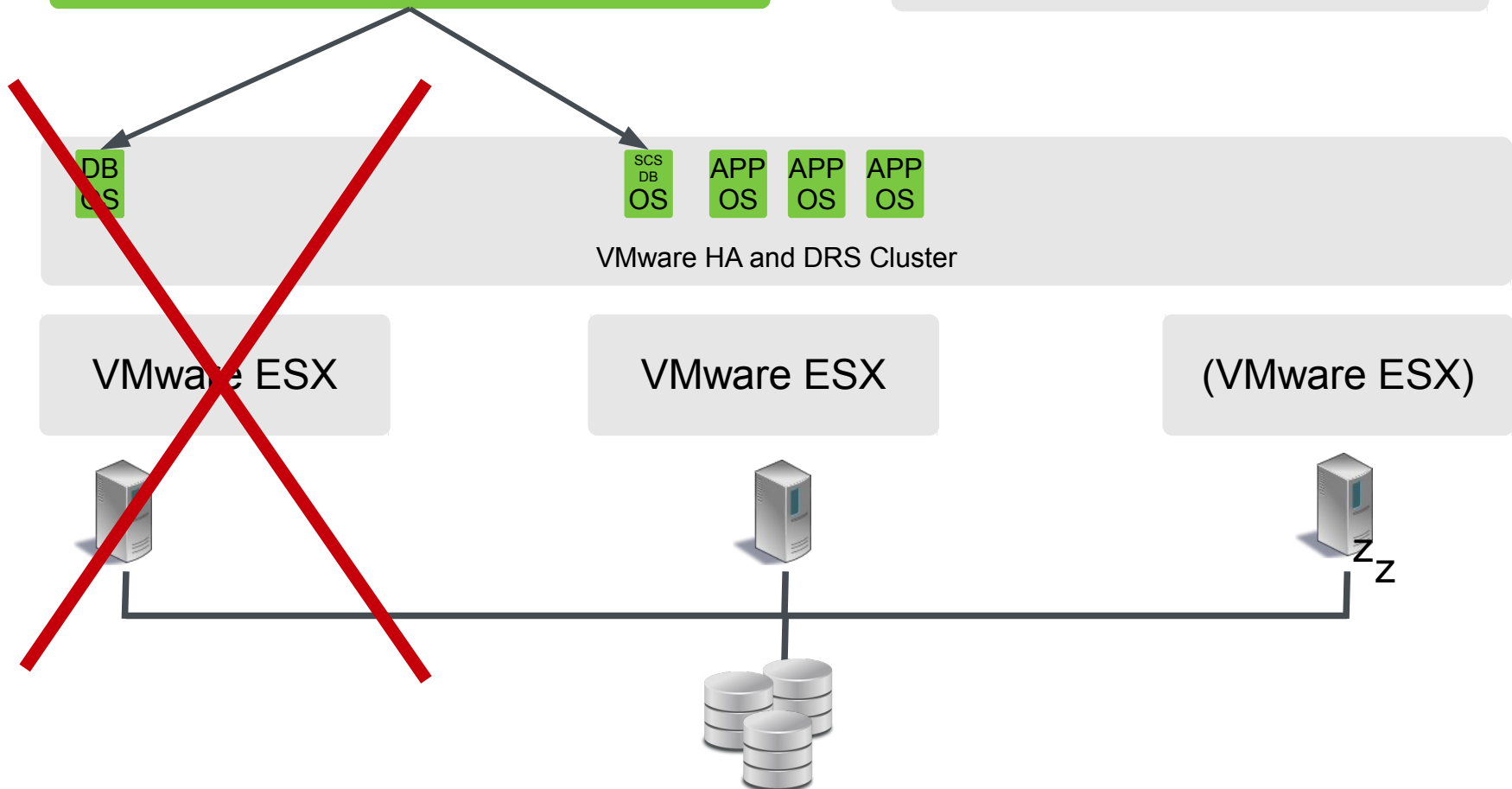
(VMware ESX)



# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

\* ... This was just in time, because  
Unfortunately a ESX hardware system fails  
\* SLE HA migrates the Database and  
optionally shutdown an Application Server



# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

- \* ESX server 1 is now in hardware Maint.
- \* VMware DPM powers up ESX server 3
- \* Failed Virtual Machines get started by VMware HA

SCS  
DB  
OS

APP  
OS

APP  
OS

APP  
OS

OS

VMware HA and DRS Cluster

(VMware ESX)

VMware ESX

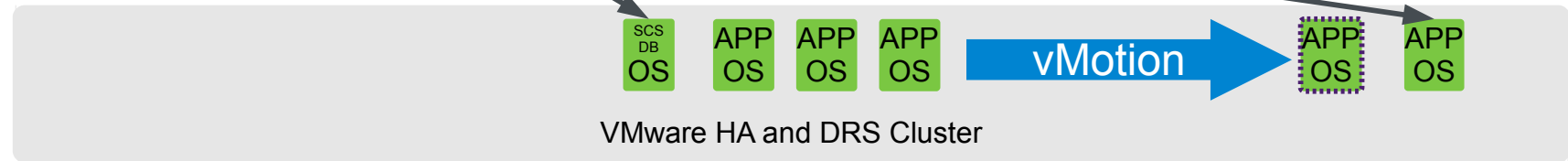
VMware ESX



# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

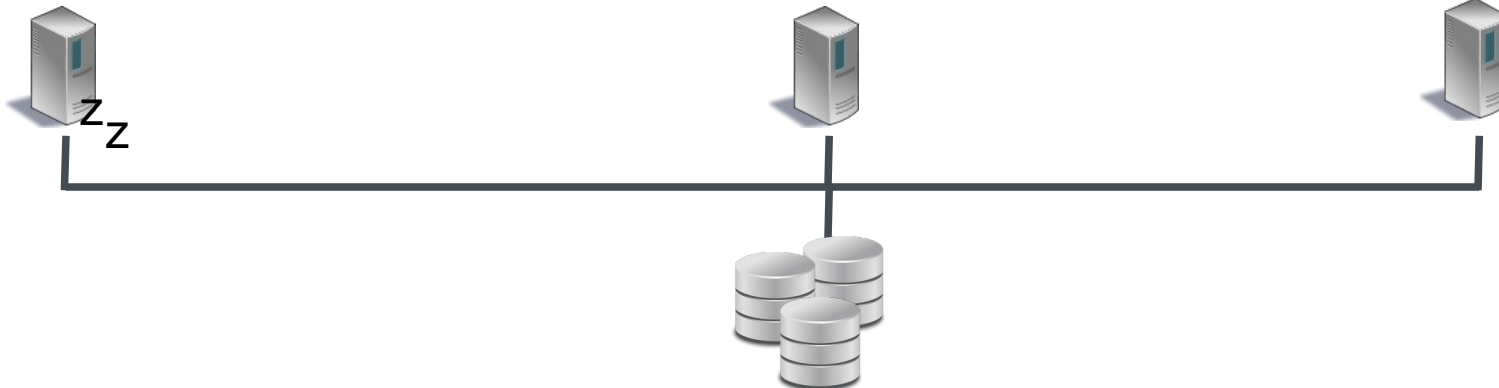
\* One of the virtual machines with an SAP application server is migrated to ESX server 3  
\* SLE HA starts the SAP application Server on the cluster node



(VMware ESX)

VMware ESX

VMware ESX





# VMware HA and SUSE Linux Enterprise High Availability Extension

SUSE Linux Enterprise  
High Availability Extension

\* Migration is ready with complete  
business continuity

SCS  
DB  
OS

APP  
OS

APP  
OS

APP  
OS

APP  
OS

APP  
OS

VMware HA and DRS Cluster

(VMware ESX)

VMware ESX

VMware ESX



# How Do We Do This?

Let's take a closer look.....



# Why Invest in SUSE with VMware ?

- Alliance partnership for 10+ years
- Joint certification and support
- Integrated VMware tools and drivers
- Supported in VMware public cloud
- Supported for OpenStack private clouds
- Recommended for SAP virtualized on VMware
- SUSE Linux Enterprise High Availability Extension complements VMware HA for mission-critical virtualized environments



# Start Now

- Visit the SUSE-VMware Alliance website at <https://www.suse.com/partners/alliance-partners/vmware/>
  - Solution briefs
  - White papers
  - Case studies
- Download SUSE Linux Enterprise Server: <https://www.suse.com/products/server/eval.html>
- Download SUSE Linux Enterprise High Availability Extension: <https://www.suse.com/products/highavailability/>
- Contact SUSE sales





**Corporate Headquarters**  
Maxfeldstrasse 5  
90409 Nuremberg  
Germany

+49 911 740 53 0 (Worldwide)  
[www.suse.com](http://www.suse.com)

Join us on:  
[www.opensuse.org](http://www.opensuse.org)

## **Unpublished Work of SUSE. All Rights Reserved.**

This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE. Access to this work is restricted to SUSE 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 SUSE. 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. SUSE 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 SUSE products remains at the sole discretion of SUSE. Further, SUSE 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 SUSE 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.

