

# SUSE® Linux Enterprise High Availability

**Kai Dupke**

Senior Product Manager  
SUSE Linux Enterprise Server  
[kdupke@suse.com](mailto:kdupke@suse.com)

**Lars Marowsky-Brée**

Distinguished Engineer  
Architect Storage / HA  
[lmbr@suse.com](mailto:lmbr@suse.com)



# Topics

SUSE® Linux Enterprise High Availability

Overview

Use Cases

Roadmap

Features

Backup

# Challenge

SUSE® Linux Enterprise High Availability

## Murphy's Law is Universal

- Faults will occur
  - Hardware crash, flood, fire, power outage, earthquake?
- Service outage and 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?**

# Benefits

SUSE® Linux Enterprise High Availability



Quickly and easily install, configure and manage clustered Linux servers



Ensure continuous access to your mission-critical systems and data



Transparent to Virtualization – nodes can be virtual or physical, or mixed!



Meet your Service Level Agreements



Increase service availability

# Features

SUSE® Linux Enterprise High Availability

- Service Availability 24/7
- Free Resource Agents
- Cluster File System
- Clustered Samba
- Virtualization Ready
- Network Load-Balancer
- Node Recovery
- Data Replication
- Unlimited Geo Clustering
- Broad Platform Support

# Leadership

SUSE® Linux Enterprise High Availability

- Long history track record
- Up-to-date Open Source High Availability stack
- Geo cluster support
- Superior Cluster File System
- Integrated Data Replication
- Full System z support
- Deep OS integration
- Ready for Virtualization

# Competition

## SUSE® Linux Enterprise High Availability

Competitive Point	SUSE Linux Enterprise High Availability Extension	Red Hat	Symantec VCS
Open Source based	Yes	Yes	No
Geo Extension	Yes	No	Yes
Supports virtualization	Hybrid physical, virtual clusters, protects guests and guest apps; supports KVM, Xen, VMware	KVM, apps within guest, clusters physical, virtual servers	VMware ESX server, protects apps in guests
OS integrated tools	Yes	Yes	No
Free tools and resource agents	Yes	No (extra for Load Balancer, Clustered Samba, and SAP Resource Agent)	No (extra charged)
Platform Support	x86, x86_64, Itanium, IBM POWER, IBM System z	Only on x86 and x86_64	Only on x86, x86_64
Major Version Upgrade	Yes	No	No
Rolling Update	Yes	No	No
Cluster File System	OCFS2, GFS2	No (extra charged for GFS2)	No (extra charged)
Data Replication	Yes	No	No (extra charged)
Node Recovery included	Yes	No	No (extra charged)
Cost	\$\$	\$\$\$	\$\$\$\$



# Use Cases



# Key Use Cases

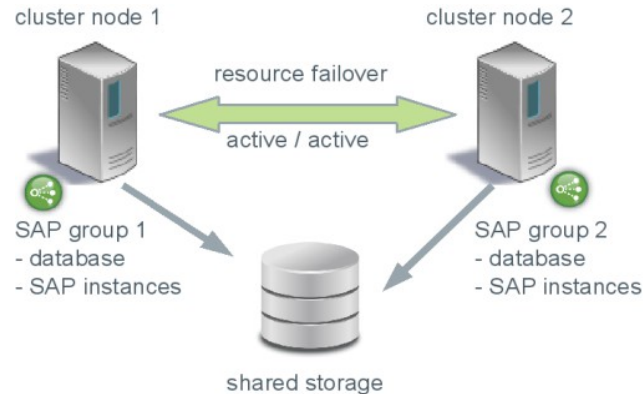
SUSE® Linux Enterprise High Availability

- High availability for mission-critical services
- Active/active services
  - OCFS2, Databases, Samba File Servers
- Active/passive service fail-over
  - Traditional databases, SAP setups, regular services
- Private Cloud
  - HA, automation and orchestration for managed VMs
- High availability across guests
  - Fine granular monitoring and HA on top of virtualization
- All Topologies
  - Local, Metro, and Geographical area clusters

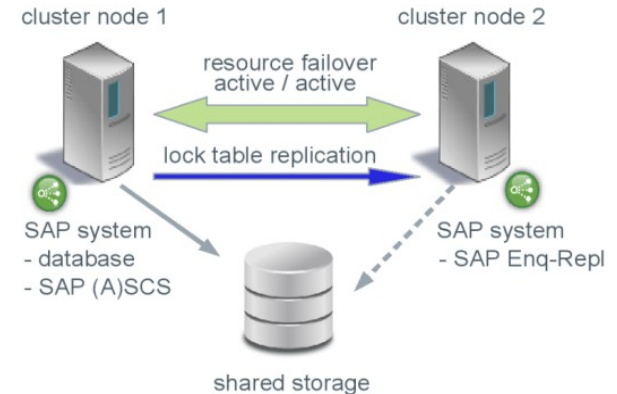
# Sample Use Cases - SAP

## SUSE® Linux Enterprise High Availability

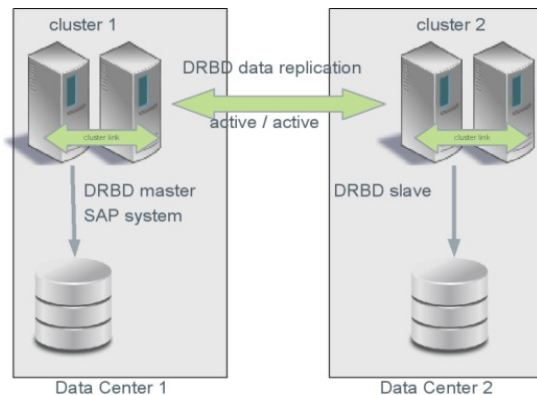
### Simple Stack



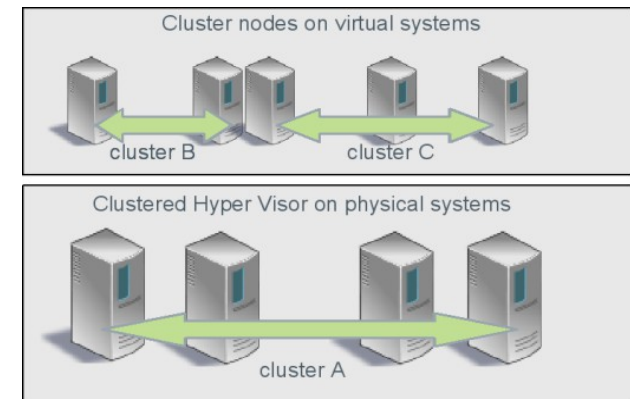
### Enqueue Replication



### DRBD Data Sync



### HA in Virtual Environments

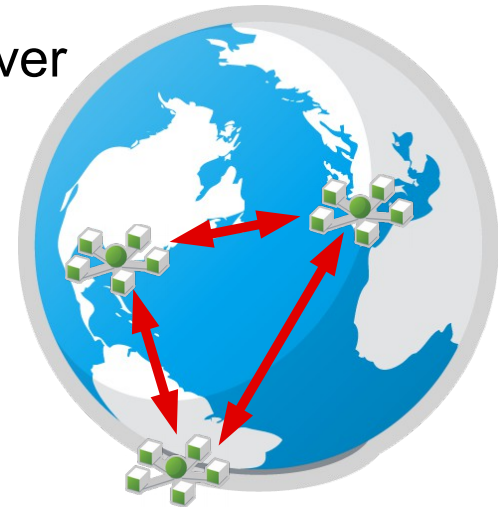


# Local & Geo Cluster

# Geo Cluster – Overview

SUSE® Linux Enterprise High Availability

- Cluster fail-over between different locations
  - Provide disaster resilience in case of site failure
  - Each site is a self-contained, autonomous cluster
  - Support manual and automatic switch-/fail-over
- Extends Metro Cluster capabilities
  - No distance limit between data centers
  - No unified storage / network needed
- Storage replicated as active / passive
  - Leverage SUSE included data replication (DRBD)
  - Integrate third-party solutions via scripts



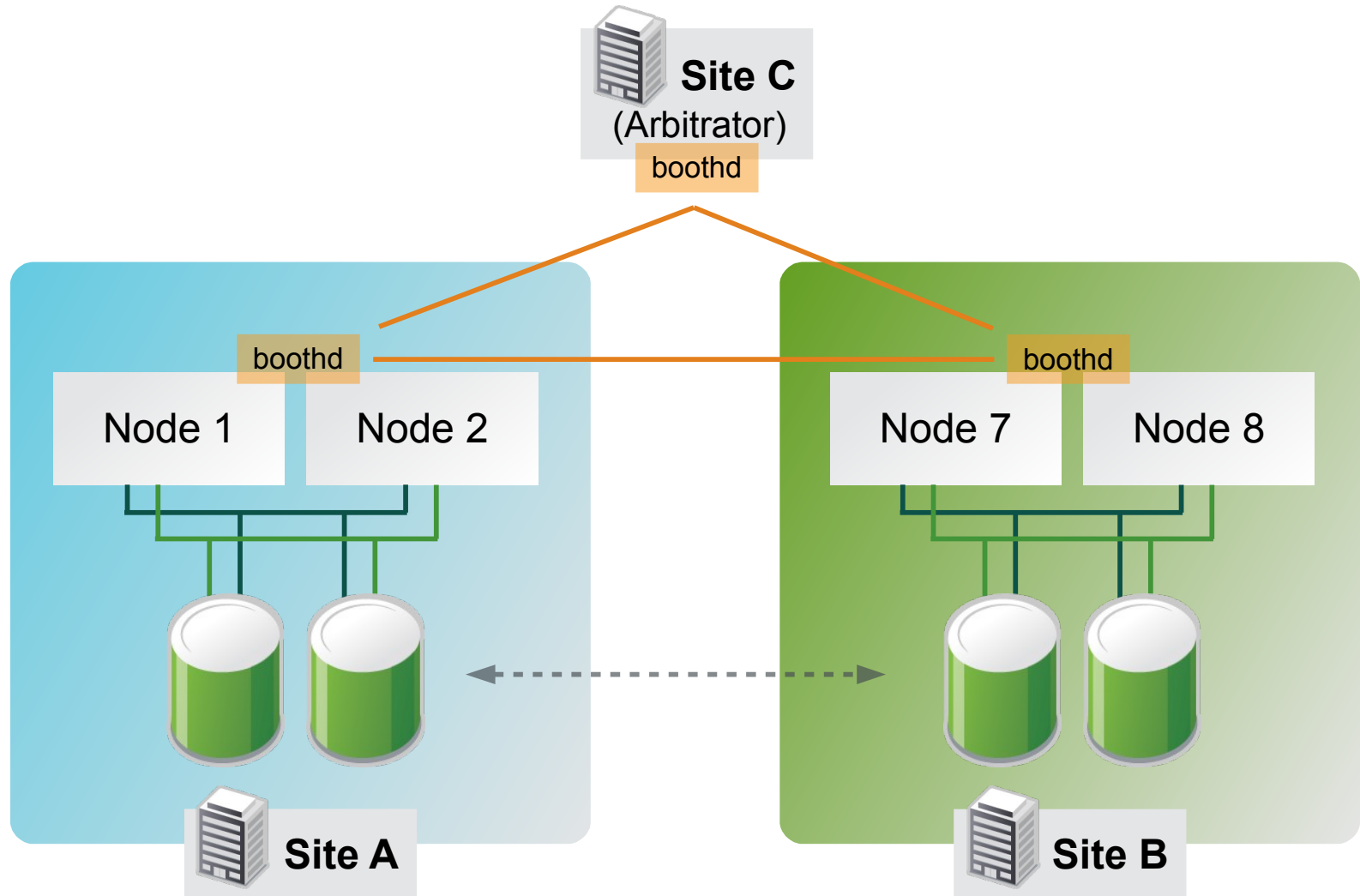
# Geo Cluster – From Local to Geo

SUSE® Linux Enterprise High Availability

- Local cluster
  - Negligible network latency
  - Typically synchronous concurrent storage access
- Metro area (stretched) cluster
  - Network latency <15ms (~20ms)
  - Unified / redundant network between sites
  - Usually some form of replication at the storage level
- Geo clustering
  - High network latency, limited bandwidth
  - Asynchronous storage replication

# Geo Cluster – Setup

SUSE® Linux Enterprise High Availability





# SUSE Linux Enterprise High Availability 12 & Roadmap



# New Features and Improvements

SUSE® Linux Enterprise High Availability Extension

- History Explorer
  - Off-line support
- Fence Agents update
  - SCSI handling
- Administration
  - Cluster health evaluation
  - crmsh improvements
  - New config options
- Node Recovery
  - Updated rear
- Load Balancer
  - HAproxy
- Cluster File System
  - OCFS2 performance improvements
  - GFS2
- Geo Clustering
  - Multi tenancy arbitrator
  - IP relocation (DNS based)

# Version 12 – Key Features

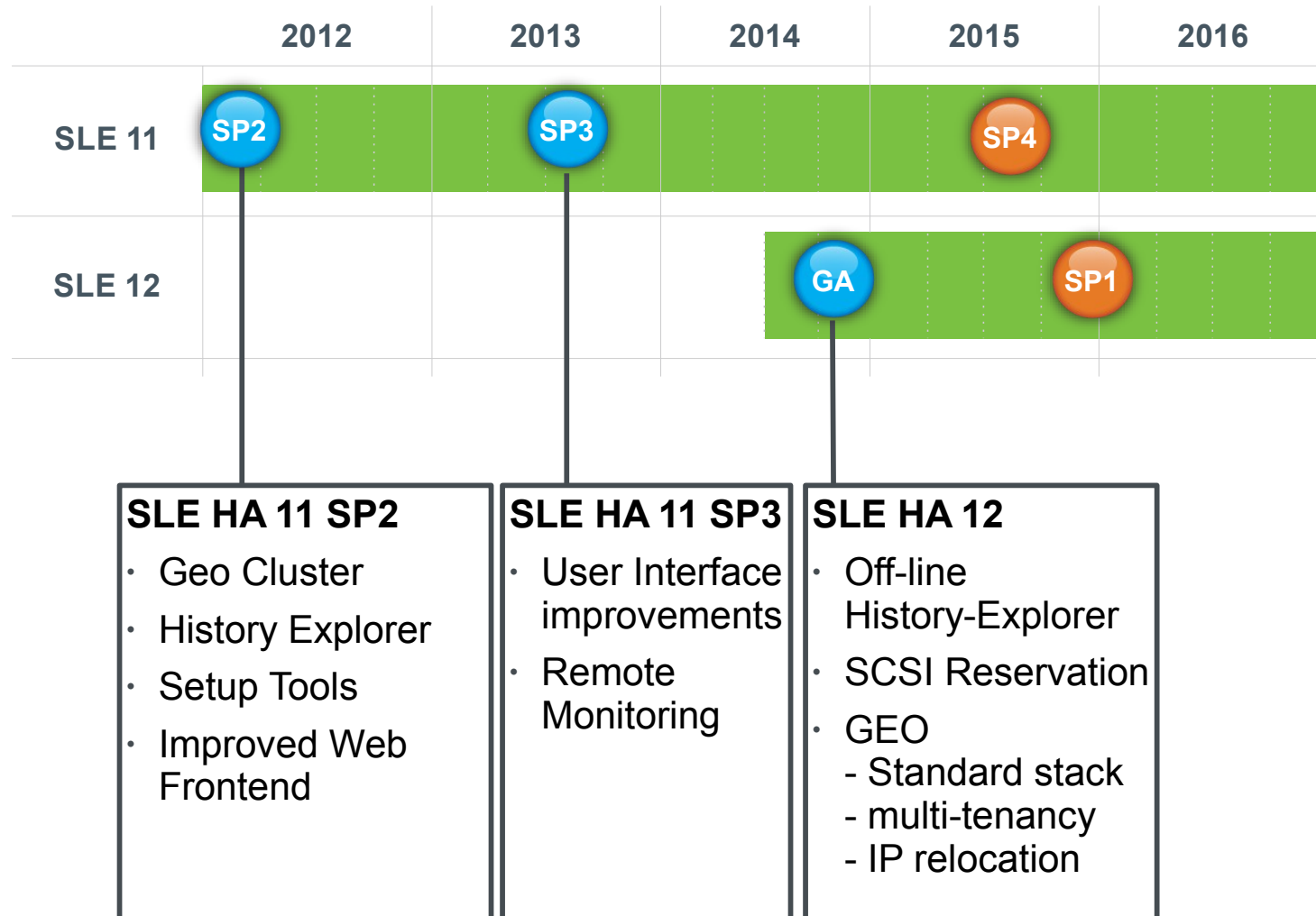
SUSE® Linux Enterprise High Availability Extension

- 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



# Roadmap

## SUSE® Linux Enterprise High Availability



# Areas to Look Into

## SUSE® Linux Enterprise High Availability

- **Failure will occur**
  - What outage is tolerable – 0s, 1s, 1min, 1hour, 1day?
- **Virtualization and Cloud**
  - Is re-{booting,deploying} a guest sufficient?
  - Install HA components in the guests?
- **Service Monitoring**
  - In depth monitoring, 'system as one' or remote monitoring?
- **Local, Metro, Geo...**
  - What is the next cluster scenario?



# Summary

SUSE® Linux Enterprise High Availability

## Fighting Murphy's Law

- Service failover at **any** distance – from local to geo
- Up to **99.9999%** availability
- Rolling updates for less **planned** downtime
- **Easy** setup, administration, management
- **Virtualization** agnostic
- **Leading** open source High Availability
- **On par** with proprietary products

**When will you start?**

Features



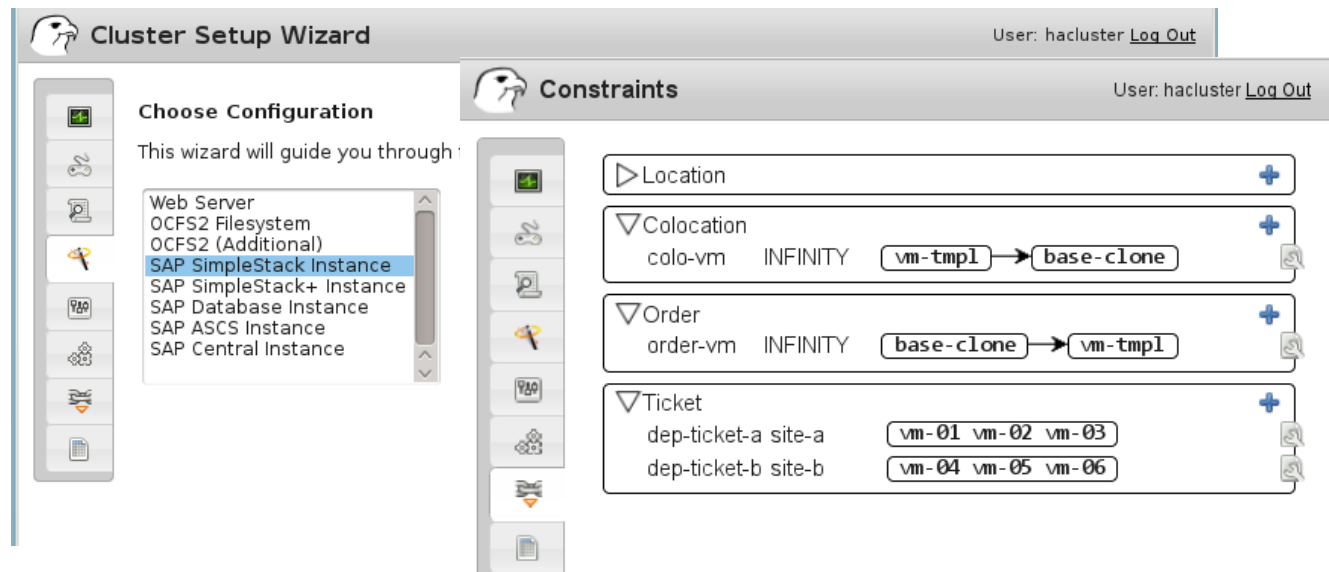
# Setup & Management




# Easy Setup – Bootstrap & Wizards






SUSE® Linux Enterprise High Availability

- Bootstrapping a cluster is really easy
  - node1 # sleha-init -i bond0 -t ocfs2 -p /dev/sdb
  - node[2...N] # sleha-join -c 192.168.2.1Options are optional
- Connect to the web console for cluster management & wizards



# Usability - hawk

 Cluster Status



Failed op: node=sles11sp3-1, resource=www, call-id=101, operation=monitor, rc-code=7

### Summary

#### Cluster Configuration

STONITH Enabled:	true
No Quorum Policy:	ignore
Symmetric Cluster:	true
Resource Stickiness:	0

#### Tickets

✓ Granted:	1
⊘ Revoked:	1

#### 2 nodes configured

▶ Online:	1
⏸ Standby:	1

#### 9 resources configured

▶ Started:	4
■ Stopped:	5

### Details

▶ xxx: Started: sles11sp3-1

▶ www: Started: sles11sp3-1

▶ dummy: Started: sles11sp3-1

▶ d2: Started: sles11sp3-1

www

#### Attributes

target-role: Started

#### sles11sp3-0

Fail Count: 0

#### sles11sp3-1

Fail Count: 1  
Last Failure: Mon Feb 11 2013 16:03:11 GMT+1100 (EST)

Close

# Command Line

## SUSE® Linux Enterprise High Availability

```
rsc_template vm-tmpl ocf:heartbeat:Xen \  
    meta allow-migrate="true" target-role="Started" \  
    utilization memory="256" cpu="2" \  
    op monitor interval="5" timeout="60" \  
    start timeout="60"  
primitive vm-02 @vm-tmpl \  
    params xmfile="/var/lib/xen/images/xm.vm-01" name="vm-01"  
primitive fencing-sbd stonith:external/sbd  
  
colocation colo-vm inf: vm-tmpl base-clone  
order order-vm inf: base-clone vm-tmpl  
  
rsc_ticket dep-ticket-a site-a: base-clone loss-policy=fence  
  
property $id="cib-bootstrap-options" \  
    enable-acl="false" \  
    migration-limit="2"  
role observer \  
    write meta:vm-01:target-role \  
    read cib  
user lmb \  
    role:observer
```

# Administration

# Remote Monitoring

- Remote monitoring of resources
  - no HA components needed
  - re-use of Nagios plugins
- Improved handling of virtual guests
  - monitor virtual services from the hypervisor
  - improve protection of VMs as cluster workload
  - guests remain unaltered – monitoring is external
- Extends pacemaker to include the concept of “container” resources

# History Explorer



History Explorer

Start Time: 2013-02-09 14:49



End Time: 2013-02-09 15:09



Display

Time	PE Input	Node			
2013-02-09 21:10:29	<a href="#">pe-input-295</a>	sles11sp3-1	<a href="#">Details (full log)</a>	<a href="#">Graph (xml)</a>	} <a href="#">Diff</a>
2013-02-09 21:09:41	<a href="#">pe-input-293</a>	sles11sp3-1	<a href="#">Details (full log)</a>	<a href="#">Graph (xml)</a>	
2013-02-09 21:09:32	<a href="#">pe-input-291</a>	sles11sp3-1	<a href="#">Details (full log)</a>	<a href="#">Graph (xml)</a>	} <a href="#">Diff</a>

pe-input-295 (sles11sp3-1)

```
Feb 9 15:00:13 sles11sp3-1 pengine[2862]: info: LogActions: Leave s:1 (M
Feb 9 15:00:13 sles11sp3-1 pengine[2862]: info: LogActions: Leave cc:0 (S
Feb 9 15:00:13 sles11sp3-1 pengine[2862]: info: LogActions: Leave cc:1 (S
Feb 9 15:00:13 sles11sp3-1 pengine[2862]: notice: LogActions: Recover www (S
Feb 9 15:00:13 sles11sp3-1 pengine[2862]: notice: process_pe_message: Calculated
Feb 9 15:00:13 sles11sp3-1 crmd[2863]: info: do_state_transition: State trans
Feb 9 15:00:13 sles11sp3-1 crmd[2863]: info: do_te_invoke: Processing graph 4
Feb 9 15:00:13 sles11sp3-1 crmd[2863]: info: te_rsc_command: Initiating action
Feb 9 15:00:13 sles11sp3-1 crmd[2863]: info: create_operation update: cib_act
Feb 9 15:00:13 sles11sp3-1 crmd[2863]: info: te_rsc_command: Initiating action
Feb 9 15:00:13 sles11sp3-1 crmd[2863]: info: create_operation update: cib_act
Feb 9 15:00:13 sles11sp3-0 apache(www)[7861]: INFO: apache not running
Feb 9 15:00:13 sles11sp3-0 crmd[3307]: notice: process_lrm_event: LRM operation
Feb 9 15:00:13 sles11sp3-0 attd[3305]: notice: attd_ais_dispatch: Update relay
Feb 9 15:00:13 sles11sp3-0 attd[3305]: notice: attd_trigger_update: Sending f
Feb 9 15:00:13 sles11sp3-0 attd[3305]: notice: attd_perform_update: Sent upda
Feb 9 15:00:13 sles11sp3-0 attd[3305]: notice: attd_ais_dispatch: Update relay
Feb 9 15:00:13 sles11sp3-0 attd[3305]: notice: attd_trigger_update: Sending f
Feb 9 15:00:13 sles11sp3-0 attd[3305]: notice: attd_perform_update: Sent upda
Feb 9 15:00:13 sles11sp3-0 lrm[3304]: info: cancel_recurring_action: Cancell
Feb 9 15:00:13 sles11sp3-0 lrm[3304]: info: log_execute: executing - rsc:www
Feb 9 15:00:13 sles11sp3-0 crmd[3307]: info: process_lrm_event: LRM operation
Feb 9 15:00:13 sles11sp3-0 apache(www)[8027]: INFO: apache is not running.
Feb 9 15:00:13 sles11sp3-0 lrm[3304]: info: log_finished: finished - rsc:www
Feb 9 15:00:13 sles11sp3-0 crmd[3307]: notice: process_lrm_event: LRM operation
Feb 9 15:00:13 sles11sp3-0 lrm[3304]: info: log_execute: executing - rsc:www
Feb 9 15:00:13 sles11sp3-0 apache(www)[8190]: INFO: apache not running
Feb 9 15:00:13 sles11sp3-0 apache(www)[8190]: INFO: waiting for apache /etc/apache
Feb 9 15:00:14 sles11sp3-0 lrm[3304]: info: log_finished: finished - rsc:www
Feb 9 15:00:15 sles11sp3-0 crmd[3307]: info: services os action execute: Manag
```

pe-input-289: pe-input-291

289	291
1 Online: [ sles11sp3-0 sles11sp3-1 ]	1 Online: [ sles11sp3-0 sles11sp3-1 ]
2	2
3 xxx (ocf::pacemaker:Dummy): Started sles11sp3-0	3 xxx (ocf::pacemaker:Dummy): Star
4 Master/Slave Set: ms [s]	4 Master/Slave Set: ms [s]
5 Masters: [ sles11sp3-0 ]	5 Masters: [ sles11sp3-0 ]
6 Slaves: [ sles11sp3-1 ]	6 Slaves: [ sles11sp3-1 ]
7 Clone Set: c [cc]	7 Clone Set: c [cc]
8 Started: [ sles11sp3-0 sles11sp3-1 ]	8 Started: [ sles11sp3-0 sles11sp
9 www (ocf::heartbeat:apache): Started sles11sp3-0	9 www (ocf::heartbeat:apache):
10	10
11 Failed actions:	11 Failed actions:
12 d_monitor_10000 (node=sles11sp3-1, call=533, rc=1, statu	12 d_monitor_10000 (node=sles11sp3-
13 www_asyncmon_0 (node=sles11sp3-0, call=0, rc=1, status=c	13 www_monitor_10000 (node=sles11sp
14	14

Legends

Colors	Links
Added	(f)first change
Changed	(n)ext change
Deleted	(t)op



# Service Pack 2 – Cluster Simulator

SUSE® Linux Enterprise High Availability Extension

The screenshot displays the 'Cluster Status' web interface. At the top, it shows 'User: hacluster' and a 'Log Out' link. The main area is a grid of resource status boxes. A 'Clone Set: base-clone' is indicated. Resources include hex-0 through hex-9 (Online), dlm-0 through dlm-2 (Started), o2cb-0 through o2cb-2 (Started), clvm-0 through clvm-2 (Started), cmirrord-0 through cmirrord-2 (Started), vg1-0 through vg1-1 (Started), ocfs2-1-0 through ocfs2-1-1 (Started), fencing-sbd (Started), and various vm-07 through vm-30 (Started). An 'Inject Operation' dialog box is open in the center, with fields for Resource (vm-08), Operation (start), Interval (empty), Node (hex-0), and Result (Not Configured). It has OK and Cancel buttons. In the bottom right, a 'Simulator (initial state)' panel shows an 'Injected State' text area, a 'Run >' button, and '+ Node', '+ Op', and '-' buttons. It also has 'Reset' and 'Close' buttons. The footer shows 'Copyright © 2009-2012 Novell, Inc.'

Cluster Status

User: hacluster [Log Out](#)

Clone Set: base-clone

hex-0: Online hex-7: Online hex-9: Online Inactive Resources

dlm-0: Started dlm-1: Started dlm-2: Started

o2cb-0: Started o2cb-1: Started o2cb-2: Started

clvm-0: Started clvm-1: Started clvm-2: Started

cmirrord-0: Started cmirrord-1: Started cmirrord-2: Started

vg1-0: Started vg1-1: Started

ocfs2-1-0: Started ocfs2-1-1: Started

fencing-sbd: Started vm-07: Started vm-09: Started

vm-08: Started vm-10: Started vm-11: Started

vm-12: Started vm-13: Started vm-15: Started

vm-14: Started vm-16: Started vm-17: Started

vm-18: Started vm-19: Started vm-20: Started

vm-21: Started vm-22: Started vm-23: Started

vm-24: Started vm-25: Started vm-26: Started

vm-27: Started vm-28: Started vm-29: Started

vm-30: Started vm-31: Started vm-32: Started

vm-33: Started

**Inject Operation**

Resource: vm-08

Operation: start

Interval: (ms)

Node: hex-0

Result: Not Configured

OK Cancel

**Simulator (initial state)**

Injected State:

Run >

+ Node + Op -

Details  
CIB (in)  
CIB (out)  
Graph (xml)

Reset Close

Copyright © 2009-2012 Novell, Inc.



# SUSE High Availability 12 New Features



Backup



Delivery

# High Availability Extension – Delivery

SUSE® Linux Enterprise High Availability

- Extension to SUSE Linux Enterprise Server
- Releases synchronized with base server product
- Annual subscriptions for x86 and AMD64&Intel64
- Included free of charge with Itanium, IBM Power, and IBM System z subscriptions
- Separate Geo Cluster option available for AMD64&Intel64 and IBM System z
- Support level inherited from the underlying SUSE Linux Enterprise Server subscription
- Free trial available



# Geo Cluster – Delivery

SUSE® Linux Enterprise High Availability

- Additional option for the SUSE Linux Enterprise High Availability Extension
  - Extends the subscription for the High Availability Extension and the SUSE Linux Enterprise Server
- Available for AMD64&Intel64 and IBM System z
- Support level inherited from the underlying SUSE Linux Enterprise Server subscription

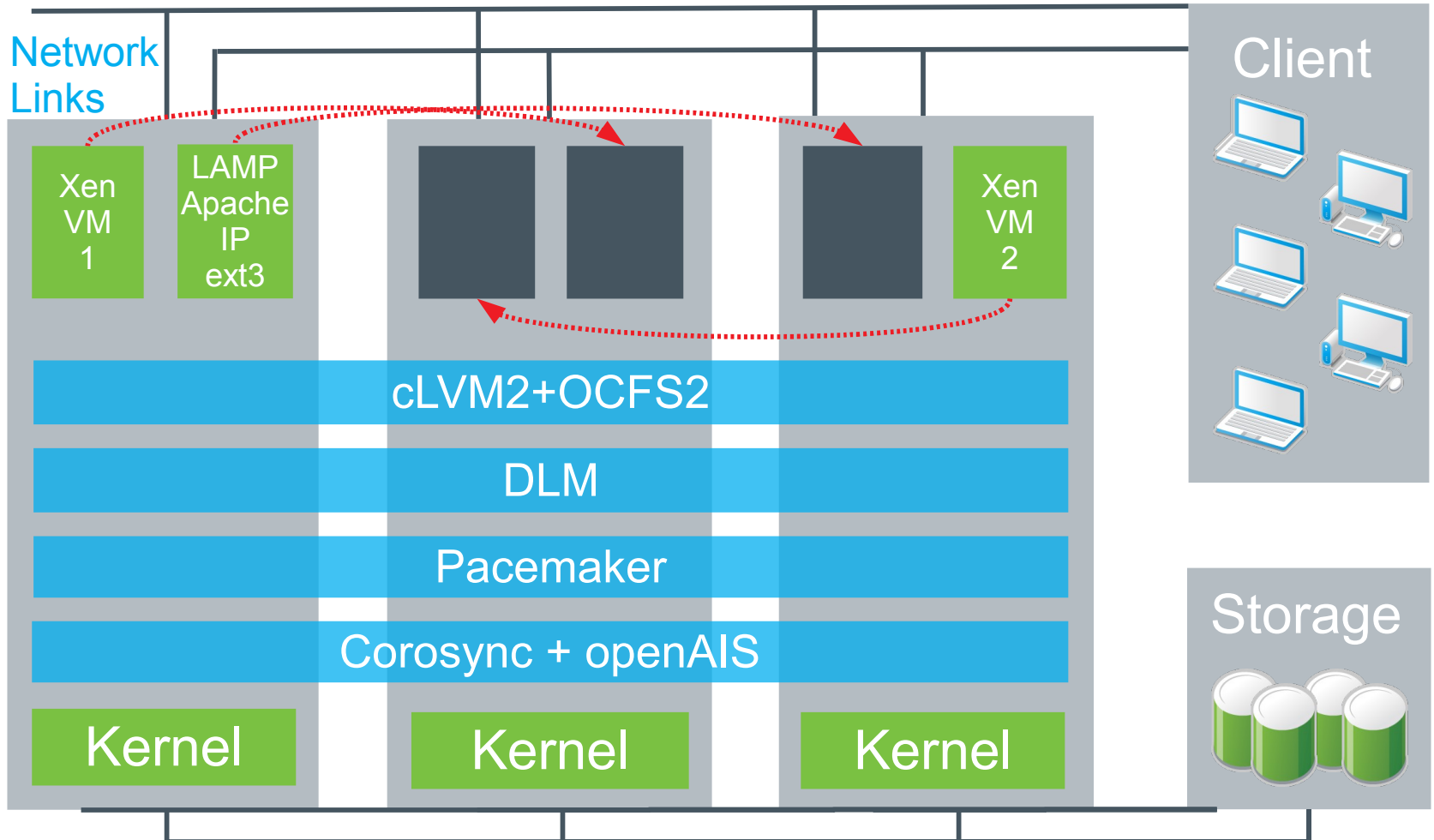


Architecture



# Cluster Example

SUSE® Linux Enterprise High Availability Extension



# Linux High Availability Stack

SUSE® Linux Enterprise High Availability Extension

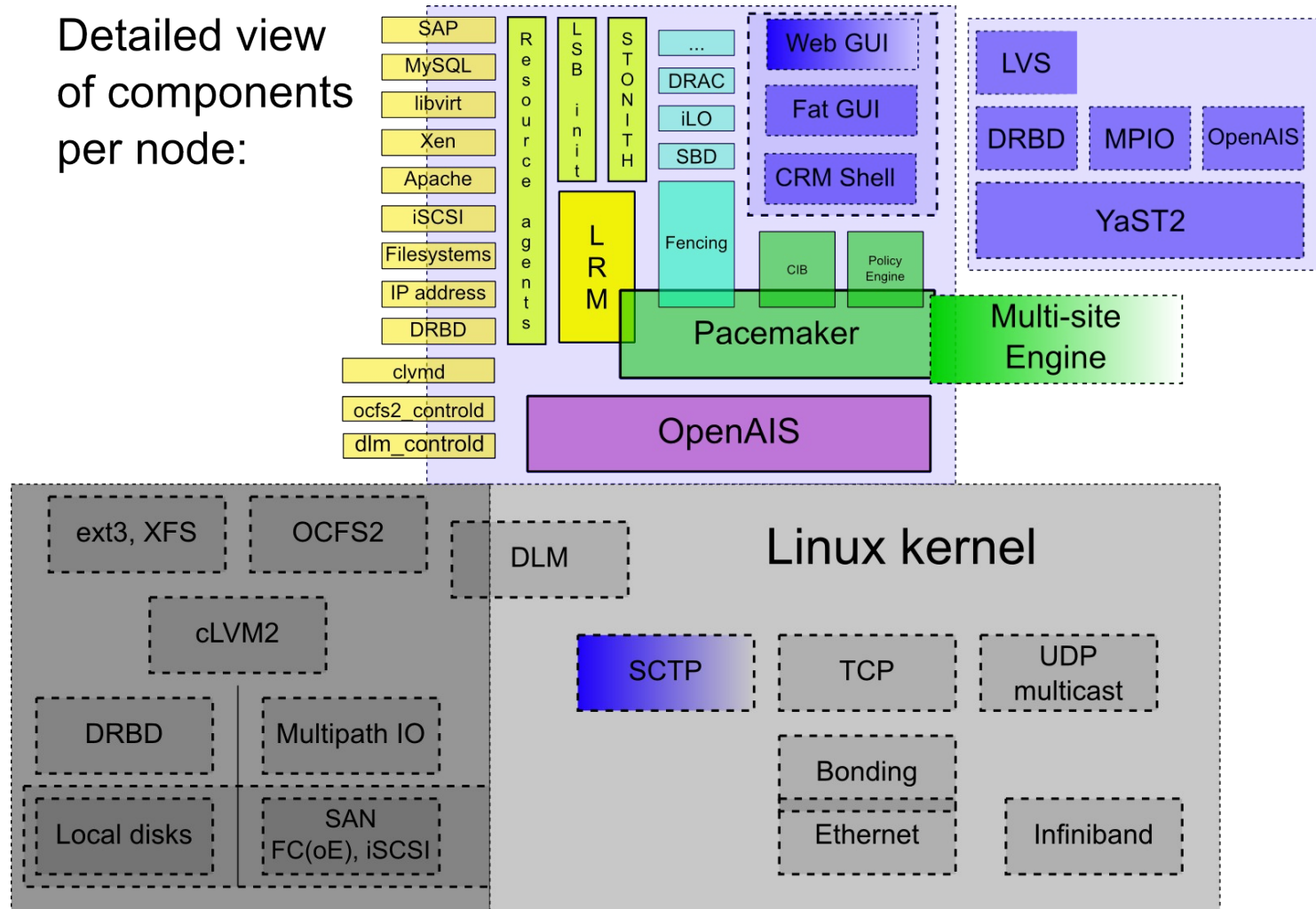
- The stack includes:
  - resource-agents – manage and monitor availability of services
  - stonith – IO fencing support (also Xen and VMware VMs)
  - corosync and OpenAIS – cluster infrastructure
  - Pacemaker – cluster resource manager
  - CRM GUI – graphical interface for cluster resource and dependencies editing
  - hawk – Web console for cluster monitoring and administration
  - CLI – improved command line to interact with the CIB: editing, prepare multiple changes - commit once, syntax validation, etc.



# Detailed Architecture

## SUSE® Linux Enterprise High Availability Extension

Detailed view  
of components  
per node:



Learn more

[www.suse.com/products/highavailability](http://www.suse.com/products/highavailability)

Thank you.





## **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.

