SUSE® Linux Enterprise High Availability

Kai Dupke
Senior Product Manager
SUSE Linux Enterprise Server
kdupke@suse.com

Lars Marowsky-Brée
Distinguished Engineer
Architect Storage / HA
lmb@suse.com

SUSE®
We adapt. You succeed.

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

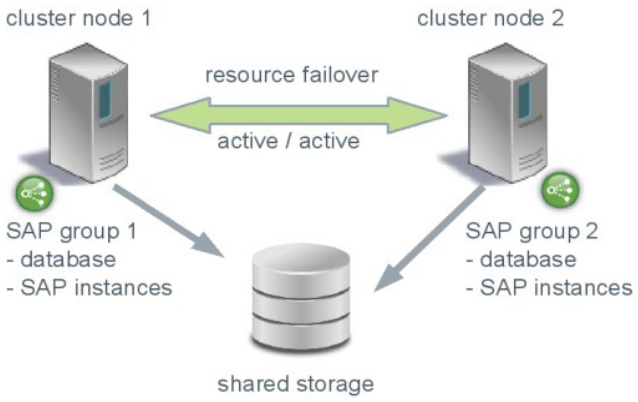
SUSE® Linux Enterprise High Availability

- 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

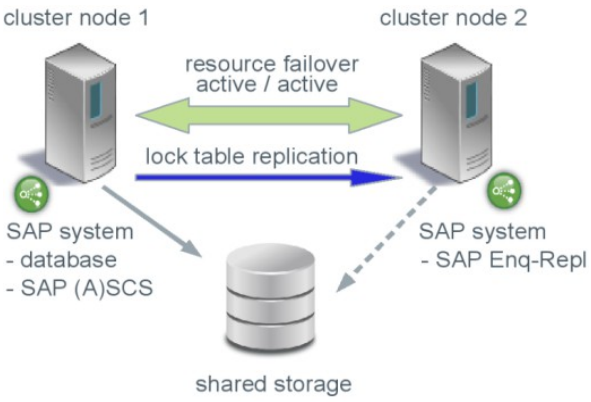
Sample Use Cases - SAP

SUSE® Linux Enterprise High Availability

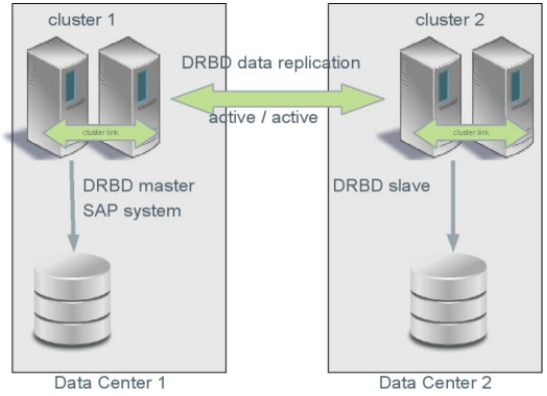
Simple Stack



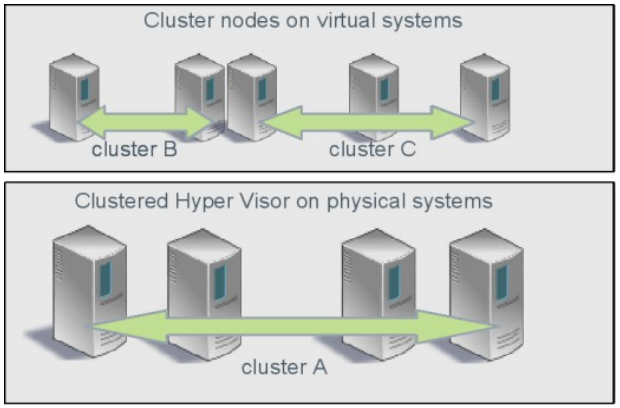
Enqueue Replication



DRBD Data Sync



HA in Virtual Environments

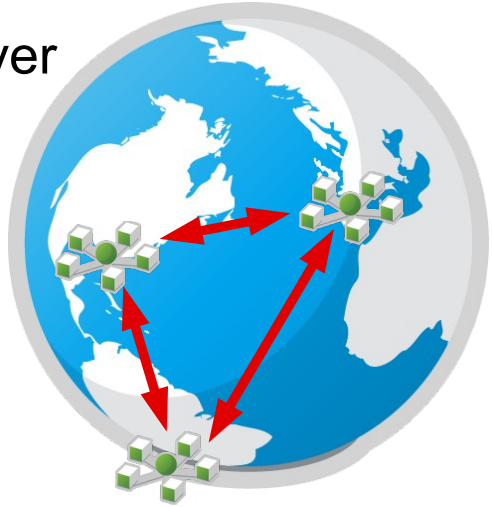


Local & Geo Cluster

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



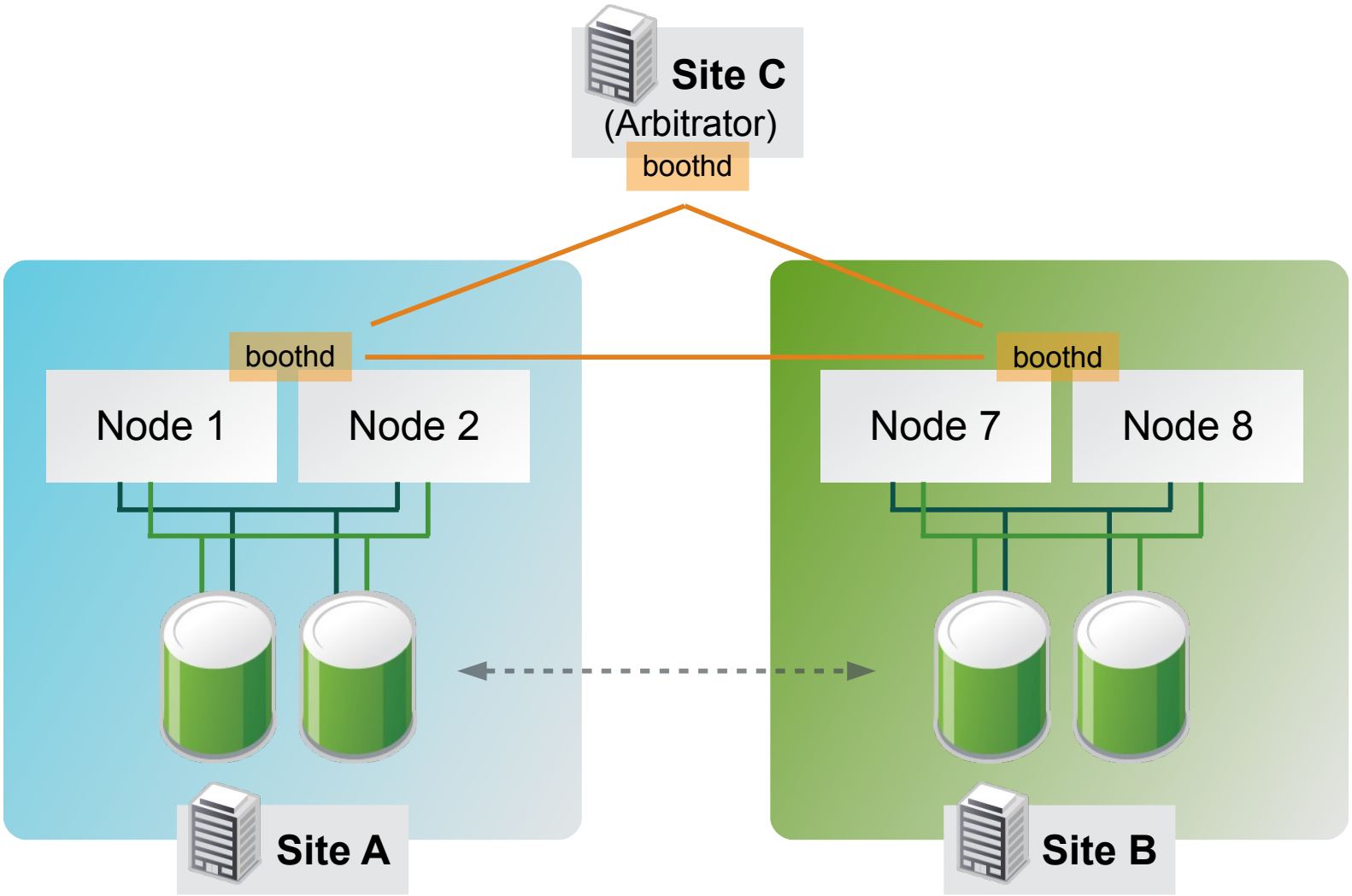
Geo Cluster – From Local to Geo

SUSE® Linux Enterprise High Availability

- Local cluster
  - Negligible network latency
  - Typically synchronous concurrent storage access
- Metro area (stretched) cluster
  - Network latency <15ms (~20mls)
  - 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)

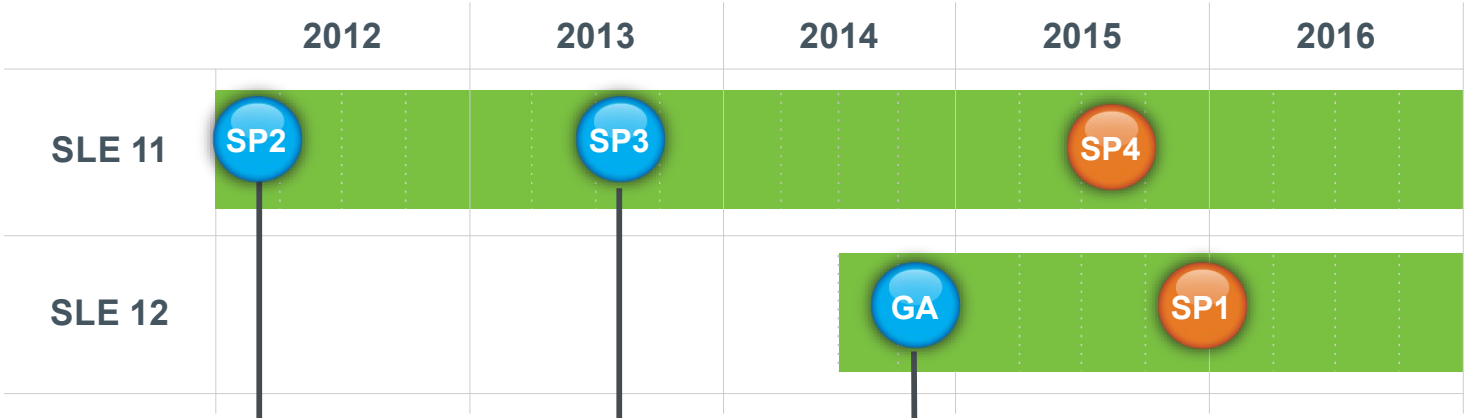
Version 12 – Key Features

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



SLE HA 11 SP2

- Geo Cluster
- History Explorer
- Setup Tools
- Improved Web Frontend

SLE HA 11 SP3

- User Interface improvements
- Remote Monitoring

SLE HA 12

- Off-line History-Explorer
- SCSI Reservation
- GEO
  - Standard stack
  - multi-tenancy
  - IP relocation

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

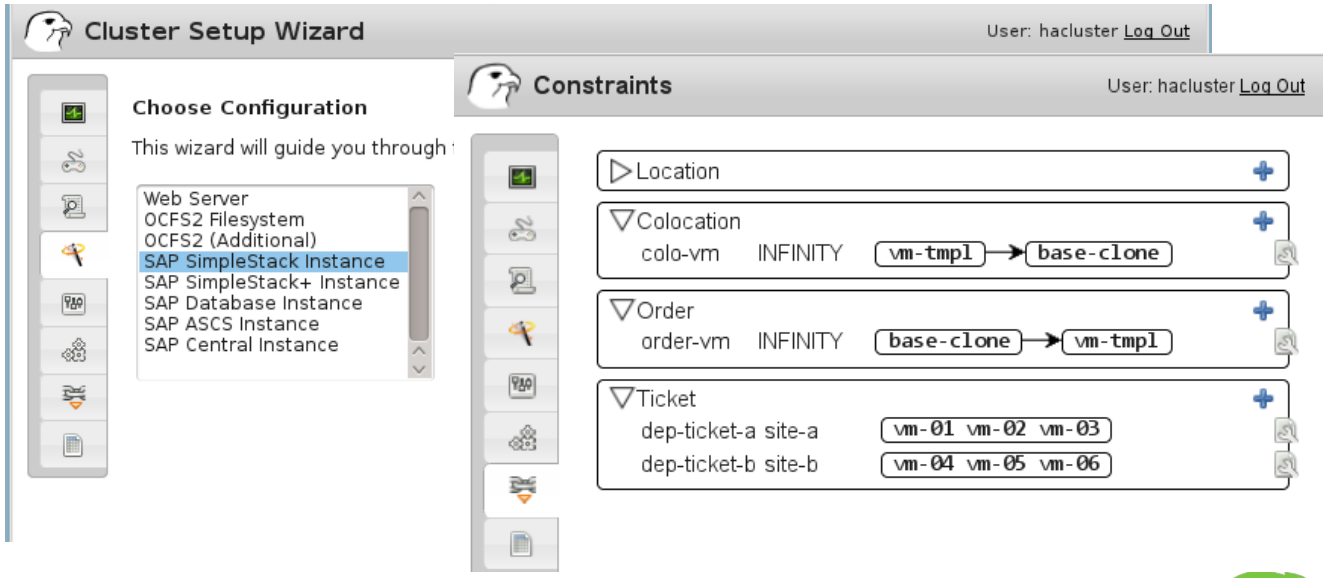
Setup & Management

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 192.168.2.1
  Options 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

SUSE

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	pe-input-295	sles11sp3-1	Details (full log)	Graph (xml)	} Diff
2013-02-09 21:09:41	pe-input-293	sles11sp3-1	Details (full log)	Graph (xml)	
2013-02-09 21:09:32	pe-input-291	sles11sp3-1	Details (full log)	Graph (xml)	} Diff
		p3-1	Details (full log)	Graph (xml)	

pe-input-295 (sles11sp3-1)

```
Feb  9 15:00:13 sles11sp3-1 pengine[2862]:     info: LogActions: Leave   s:1    (Ma
Feb  9 15:00:13 sles11sp3-1 pengine[2862]:     info: LogActions: Leave   cc:0   (St
Feb  9 15:00:13 sles11sp3-1 pengine[2862]:     info: LogActions: Leave   cc:1   (St
Feb  9 15:00:13 sles11sp3-1 pengine[2862]:   notice: LogActions: Recover www    (St
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 transi
Feb  9 15:00:13 sles11sp3-1 crmd[2863]:     info: do_te_invoke: Processing graph 48
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_acti
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_acti
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 attrd[3305]:   notice: attrd_ais_dispatch: Update relay
Feb  9 15:00:13 sles11sp3-0 attrd[3305]:   notice: attrd_trigger_update: Sending fl
Feb  9 15:00:13 sles11sp3-0 attrd[3305]:   notice: attrd_perform_update: Sent updat
Feb  9 15:00:13 sles11sp3-0 attrd[3305]:   notice: attrd_ais_dispatch: Update relay
Feb  9 15:00:13 sles11sp3-0 attrd[3305]:   notice: attrd_trigger_update: Sending fl
Feb  9 15:00:13 sles11sp3-0 attrd[3305]:   notice: attrd_perform_update: Sent updat
Feb  9 15:00:13 sles11sp3-0 lrmd[3304]:     info: cancel_recurring_action: Cancelli
Feb  9 15:00:13 sles11sp3-0 lrmd[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 lrmd[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 lrmd[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 lrmd[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

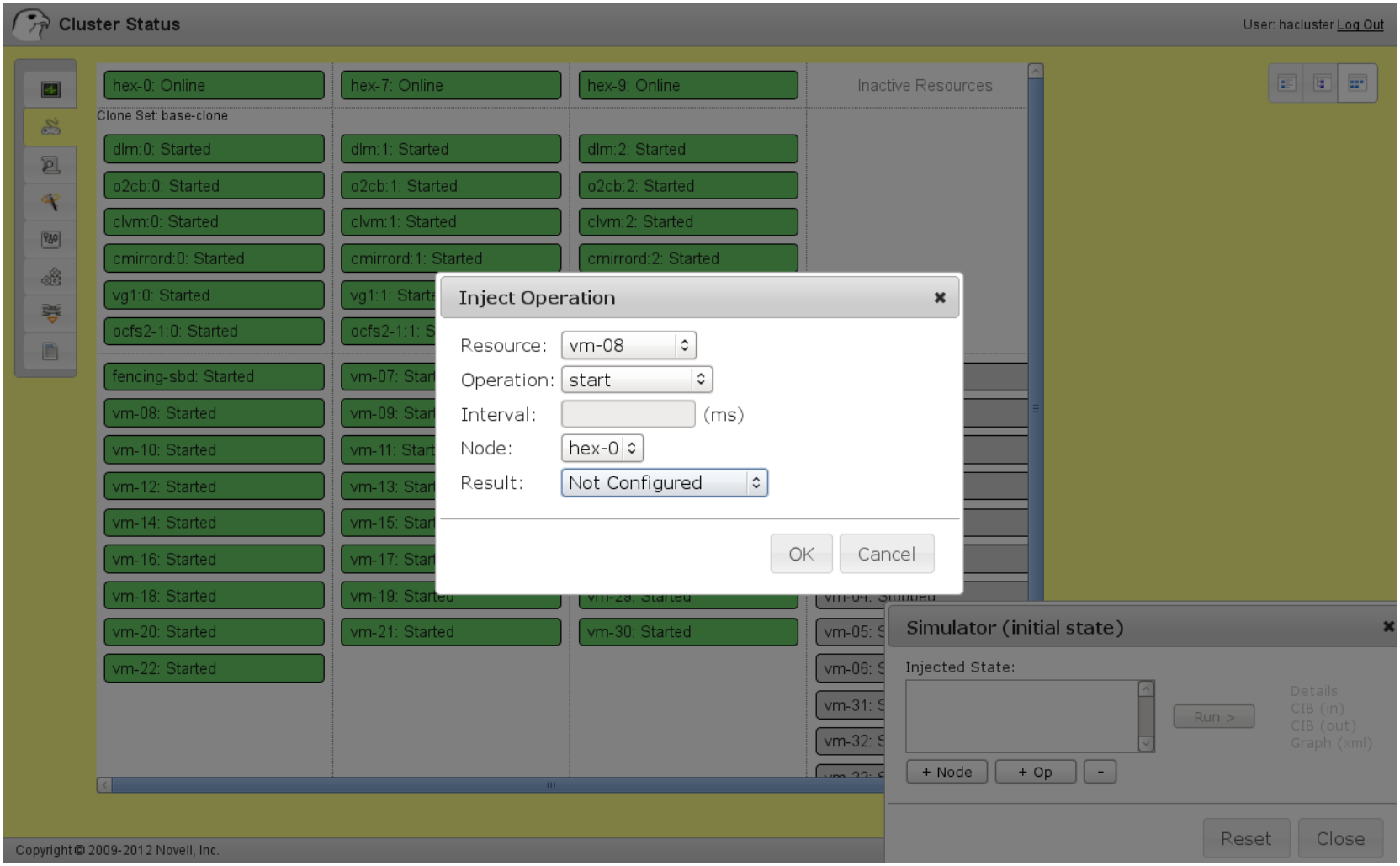
```
289                                                          291
f 1 Online: [ sles11sp3-0 sles11sp3-1 ]                      f 1 Online: [ sles11sp3-0 sles11sp3-1 ]
  2                                                            2
  3 xxx    (ocf::pacemaker:Dummy): Started sles11sp3-0         3 xxx    (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
n 9 www    (ocf::heartbeat:apache):        Started sles11sp3-0 n 9 www    (ocf::heartbeat:apache):
  >                                                            >FAILED
 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-
  >s=Timed Out): unknown error                                 >s=Timed Out): unknown error
t 13    www_asyncmon_0 (node=sles11sp3-0, call=0, rc=1, status=ct 13    www_monitor_10000 (node=sles11sp
  >omplete): unknown error                                     >tus=complete): not running
```

Legends

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

Service Pack 2 – Cluster Simulator

SUSE® Linux Enterprise High Availability Extension



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

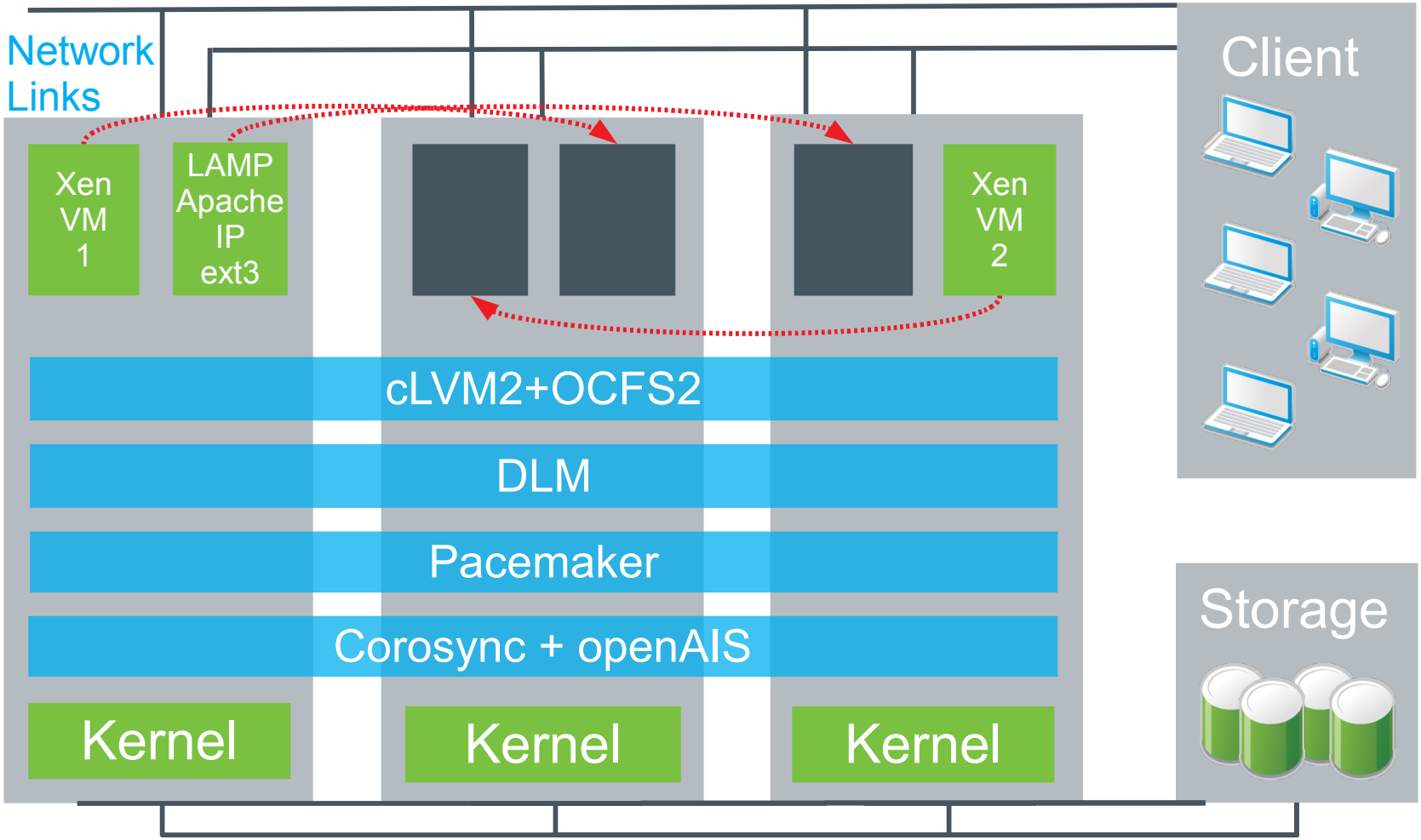
SUSE® Linux Enterprise High Availability

- 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

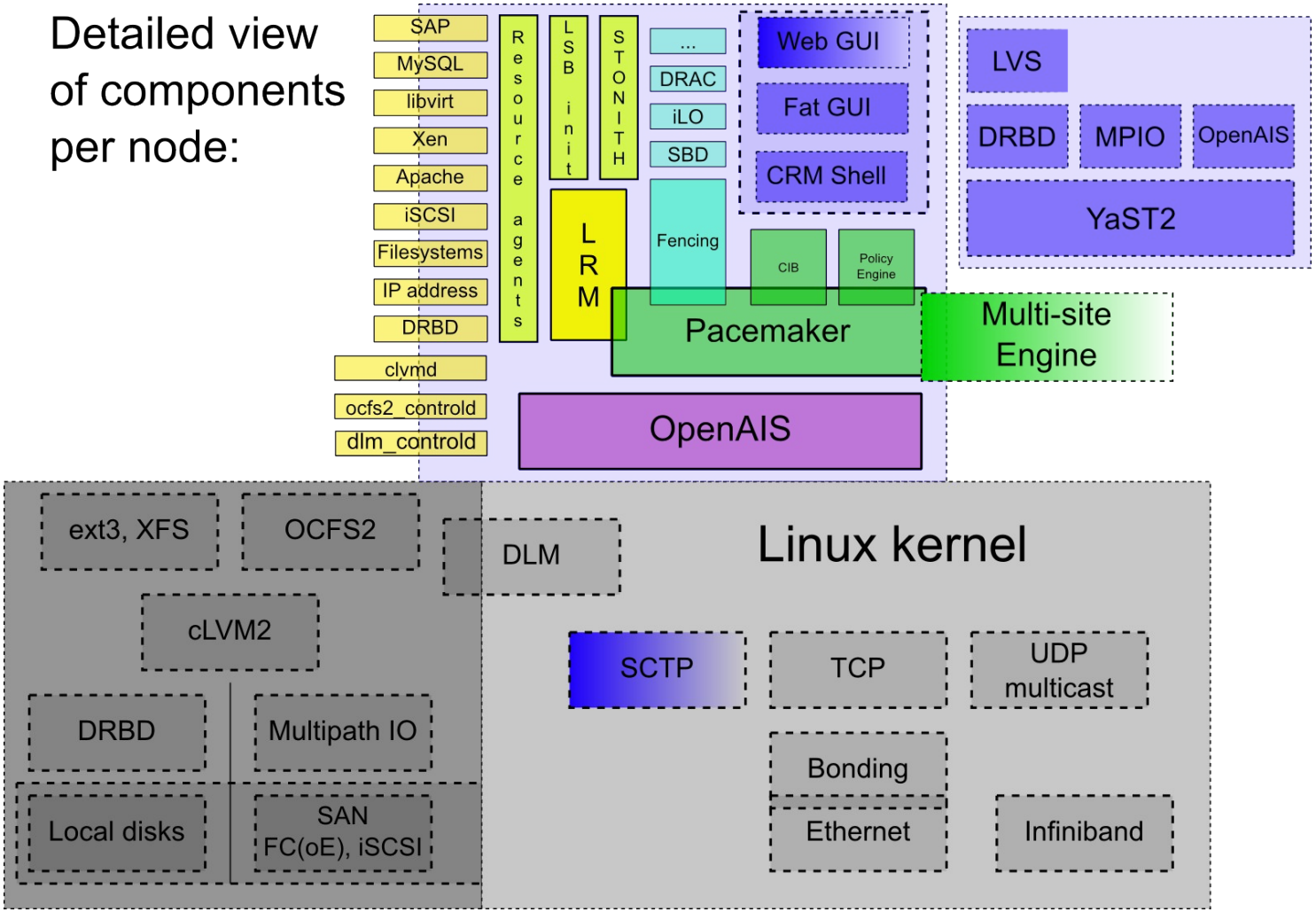
SUSE® Linux Enterprise High Availability Extension

- 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

Thank you.


SUSE
We adapt. You succeed.