High Availability for SUSE® Linux Enterprise
Roadmap, Use Cases, Current Status & Outlook

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

Lars Marowsky-Brée
Distinguished Engineer
Architect High-Availability and Storage
lmb@suse.com
Topics
SUSE® Linux Enterprise High Availability Extension

Overview
Current Status
Geo Cluster
Outlook
Challenge
SUSE® Linux Enterprise High Availability

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?
Overview
Overview
SUSE® Linux Enterprise High Availability Extension

• Most modern and complete open source solution for implementing high available Linux clusters

• A suite of robust open source technologies that is:
  - Affordable
  - Integrated
  - Virtualization agnostic

• Used with SUSE Linux Enterprise Server, it helps to:
  - Maintain business continuity
  - Protect data integrity
  - Reduce unplanned downtime for your mission-critical Linux workloads
Benefits
SUSE® Linux Enterprise High Availability Extension

- 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
Key Use Cases
SUSE® Linux Enterprise High Availability Extension

• 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 Extension

Simple Stack
- cluster node 1
  - resource failover
  - SAP group 1
    - database
    - SAP instances
  - shared storage
- cluster node 2
  - resource failover
  - SAP group 2
    - database
    - SAP instances
  - shared storage

Enqueue Replication
- cluster node 1
  - resource failover
  - SAP system
    - database
    - SAP (A)SCS
  - shared storage
- cluster node 2
  - resource failover
  - SAP system
    - SAP Enq-Repl
  - shared storage

DRBD Data Sync
- cluster 1
  - DRBD master
    - SAP system
  - DRBD data replication
  - active / active
  - Data Center 1
- cluster 2
  - DRBD slave
  - active / active
  - Data Center 2

HA in Virtual Environments
- Cluster nodes on virtual systems
  - cluster B
  - cluster C
- Clustered Hypervisor on physical systems
  - cluster A
Reference — German Air Traffic Control
SUSE® Linux Enterprise High Availability Extension

• Controls all air traffic over Germany
• Government owned
• Long term HA and SUSE customer

• See the DFS YouTube video

„SUSE Linux Enterprise plays an important role as the strategic platform”

„SUSE was quick at hand with providing qualified, in-depth answers“

“So far we have not seen a single downtime that could be tracked down to SUSE Linux itself”

— Pieter Hollants
DFS Linux Service and Competence Center

Quite obviously, SUSE’s engineers are doing an excellent job in providing a rock-solid operating system.
Current Status
Current Status
SUSE® Linux Enterprise High Availability Extension

Fighting Murphy's Law

• Service failover at any distance – from local to geo
• 99.9999% availability with the appropriate tuning
• 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?
Leadership
SUSE® Linux Enterprise High Availability Extension

- 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 Extension**

<table>
<thead>
<tr>
<th>Competitive Point</th>
<th>SUSE Linux Enterprise High Availability Extension</th>
<th>Red Hat</th>
<th>Symantec VCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requires shared storage</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Open Source based</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Geo Extension</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Supports virtualization</td>
<td>Hybrid physical, virtual clusters, protects guests and guest apps; supports KVM, Xen, VMware</td>
<td>KVM, apps within guest, clusters physical, virtual servers</td>
<td>VMware ESX server, protects apps in guests</td>
</tr>
<tr>
<td>OS integrated tools</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Free tools and resource agents</td>
<td>Yes (extra for Load Balancer, Clustered Samba, and SAP Resource Agent)</td>
<td>No</td>
<td>No (extra charged)</td>
</tr>
<tr>
<td>Platform Support</td>
<td>x86, x86_64, Itanium, IBM POWER, IBM System z</td>
<td>Only on x86 and x86_64</td>
<td>x86, x86_64</td>
</tr>
<tr>
<td>Major Version Upgrade</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Rolling Update</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Node Recovery included</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cost</td>
<td>$$</td>
<td>$$$$</td>
<td>$$$$$</td>
</tr>
</tbody>
</table>
Features
SUSE® Linux Enterprise High Availability Extension

• Service availability 24/7
  - Policy driven clustering
• Shared and Scaled data-access
  - Cluster file system
  - Clustered Samba
• Scale Network Services
  - IP load-balancing
• Virtualization Agnostic
  - Platform independent setup
• Disaster tolerance
  - Data replication via IP
  - Node recovery
• User friendly tools
  - Graphical user interface
  - Unified command line
• Free Resource Agents
• Geo Clustering
  - Cluster across unlimited distance
Recent Enhancements
Service Pack 2 – Added Features
SUSE® Linux Enterprise High Availability Extension

• Easy Installation and Set-Up
  - Cluster Bootstrap & Join
  - Templates and Wizards

• Improved Supportability
  - History Explorer
  - Log File Query Tools

• Efficient Management
  - Access Control Lists
  - Enhanced Web Console

• Improved Reliability
  - Multiple SBD devices for storage-based fencing

• Additional Capabilities
  - Joining of Clustered SAMBA to Active Directory
  - Load Balancer Connection Tracking and Replication
  - ReaR support for SUSE boot media
Service Pack 3 – **Added Features**

SUSE® Linux Enterprise High Availability Extension

- **Managing remote resources**
  - Black box external monitoring

- **Management**
  - Dashboard to manage multiple clusters
  - Display setup in a Cluster Diagram

- **Usability**
  - Improved web console

- **Supportability**
  - Cluster Simulator with configuration changes
  - Enhanced History Explorer

- **Administration**
  - Pacemaker, YaST2, & Resource Agents

- **Stack refresh** (drbd, sbd, pacemaker ...)

SUSE®
Selected Features
Easy Setup – **Bootstrap**

SUSE® Linux Enterprise High Availability Extension

- Bootstrapping a cluster is really easy:
  - node1 # sleha-init -i bond0 -t ocfs2 -p /dev/sdb
  - nodeN # sleha-join -c 192.168.2.1

- Connect to the hawk web console for cluster management
Easy Setup – Wizards
SUSE® Linux Enterprise High Availability Extension

• Connect to the hawk web console
• Start a wizard

Cluster Setup Wizard

Choose Configuration
- Web Server
- OCFS2 Filesystem
- OCFS2 (Additional)
- SAP SimpleStack Instance
- SAP SimpleStack+ Instance
- SAP Database Instance
- SAP ASCS Instance
- SAP Central Instance

Constraints

- Location
- Colocation
- Order
- Ticket
Blackbox Monitoring

- Improved handling of virtual machines as clustered services
- External monitoring of resources from hypervisor
  - No installation necessary inside the unaltered guest
- Re-uses Nagios/Icinga plugins
- Extends pacemaker to include the concept of “container” resources
hawk – Cluster Dashboard & Diagram
Usability - hawk

Cluster Status

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

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
From Local Cluster to Geo Cluster
Local & Stretched Cluster
SUSE® Linux Enterprise High Availability Extension

Clients

SLES SLE HA

SLES SLE HA

SLES SLE HA

SLES SLE HA

SLES SLE HA

SLES SLE HA
Geo Cluster – From Local to Geo
SUSE® Linux Enterprise High Availability Extension

• 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 – Overview
SUSE® Linux Enterprise High Availability Extension

• Cluster fail-over between different data center 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 Distributed Replicated Block Device (DRBD)
  - Can integrate third-party solutions via scripts
Geo Cluster – Setup
SUSE® Linux Enterprise High Availability Extension

Site A
- Node 1
- Node 2

Site C (Arbitrator)
- boothd

Site B
- Node 7
- Node 8

Site C
- Node 1
- Node 2

Node 7
- Node 8

Node 1
- Node 2

Node 7
- Node 8

Site A
- Node 1
- Node 2

Site B
- Node 7
- Node 8
Roadmap
SUSE® Linux Enterprise High Availability Extension

<table>
<thead>
<tr>
<th>Year</th>
<th>SLE 11</th>
<th>SLE 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>GA</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>SP1</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>SP2</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>SP3</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td>GA</td>
</tr>
</tbody>
</table>

**SLE HA 11 SP1**
- Metro Area Cluster
- Samba Cluster
- Web GUI
- Cluster Test Drive
- Node Recovery

**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**
- Extended Geo capabilities
- Ease of use
- Major refresh
What’s Coming in SLE HA 12
SUSE® Linux Enterprise High Availability Extension

- Re-basing the stack on SLE 12 and latest upstream
- Storage
  - drbd, clustered RAID1, SCSI reservation handling
- Usability
  - hawk (web) and crm shell
  - Further CIB and crm shell syntax enhancements
  - Documentation
  - Service wizards and templates
    - That are able to configure the host system (corosync, NFS, etc ...)
- Performance and scalability
What’s Coming in SLE HA GEO 12
Geo Clustering for SUSE® Linux Enterprise High Availability Extension

• Extend framework to a full solution stack
• Reference architecture
  – Based on LAMP stack, DRBD storage
  – IP take-over (via dynamic DNS)
  – Sync of CIB and configuration files between sites
  – Documentation
• Address multiple tenancy for overlapping environments
• Working with partners on further integration
Areas to Look Into
SUSE® Linux Enterprise High Availability Extension

• 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?
Learn More

www.suse.com/products/highavailability

Thank you.
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
SP2 Feature Details
Service Pack 2 – Hawk Enhancements
SUSE® Linux Enterprise High Availability Extension
Service Pack 2 – **Cluster Simulator**

SUSE® Linux Enterprise High Availability Extension
Service Pack 2 – crm Enhancements
SUSE® Linux Enterprise High Availability Extension

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 xmfilee="/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
Service Pack 2 – **Multiple SBD Fencing**
SUSE® Linux Enterprise High Availability Extension

- **STONITH Block Device (SBD) fencing** is recommended by SUSE
  - SBD fencing is highly reliable
  - Independent of management board (firmware, settings, etc.) thanks to hardware watchdog integration
  - Equal setup in physical and virtual environments, reducing variance in deployments

- **Multiple SBD fencing**
  - Supports redundancy in fencing channels
  - Enhanced reliability leveraging independent storage systems

- **Integration with Pacemaker & corosync status!**
SP3 Feature Details
hawk – Usability

- More intuitive handling with status icons
- Make in-flight actions (start/stop/...) visible
- Direct access to history explorer from status messages
Supportability Enhancements

- Cluster Simulator allow configuration changes
- History Explorer
  - Clean up old reports
  - Display diff for PE inputs
  - Link to transition log
Administration – Pacemaker

- Maintenance mode for a single node
- Calculate the default probe timeout
- Improved handling of group based utilization
Administration – YaST2 & Agents

• YaST2
  – Base setup for ReaR with NFS backend
  – Join AD domain with CTDB object
  – Refresh iplb module

• Resource Agents
  – nfsserver: timeout configurable
  – Oracle: improve log settings
  – SysInfo: handle disk full
cLVM2

• RAID10 for cmirrord
  - Support multiple physical volumes/devices per mirror leg
  - In case where the SAN does not allow the one PV to be grown
• Support for 'mirrored' dirty log
  - Reduce need for full array resync
Assorted Updates in SP3

- sbd now honors quorum/pacemaker health
  - Partially released in SP2 maintenance
  - UUID in metadata, better debugging mode
- DRBD updated to 8.4.3
  - Significant performance improvements for activity logging
  - Bug fixes
- corosync updated to 1.4.5
- ocfs2-tools to 1.8.2
  - Faster fsck, bug fixes
Architecture
Cluster Example
SUSE® Linux Enterprise High Availability Extension

Network Links
- Xen VM 1
- LAMP Apache IP ext3
- cLVM2+OCFS2
- DLM
- Pacemaker
- Corosync + openAIS
- Kernel
- Kernel
- Kernel

Client

Storage
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:
Thank you.

Learn More
www.suse.com/products/highavailability
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.