Implementing the SUSE Linux Enterprise High Availability Extension on System z

Mike Friesenegger
SUSE Technical Specialist
mikef@novell.com but soon to be mikef@suse.com

Agenda

- SUSE Linux Enterprise Server for System z
- What is a high availability (HA) cluster?
- What is required to build a HA cluster using SLES?
- Demoing the features of SLE HAE
  - Managing a cluster with the GUI and CLI
  - Resources primitives and resource groups
  - Resource Constraints
  - STONITH
  - cLVM and OCFS2
- Call to Action

SUSE Linux Enterprise Server for System z

SUSE® Linux Enterprise Server for System z
10 years on the Mainframe

http://www.novell.com/partners/ibm/mainframe/img/timeline_lores.pdf

- The first deployments on Linux for the mainframe were file and print servers.
- The first piece of software that became popular was Samba.
- The first large commercial customer for SUSE Linux Enterprise Server for S/390 was Telia, the largest telecommunications company in Sweden.
- Today, companies are running their mission-critical workloads on top of SUSE Linux Enterprise Server for System z.

Why Customers Prefer SUSE® Linux Enterprise Server for System z

The optimized version for IBM System z:

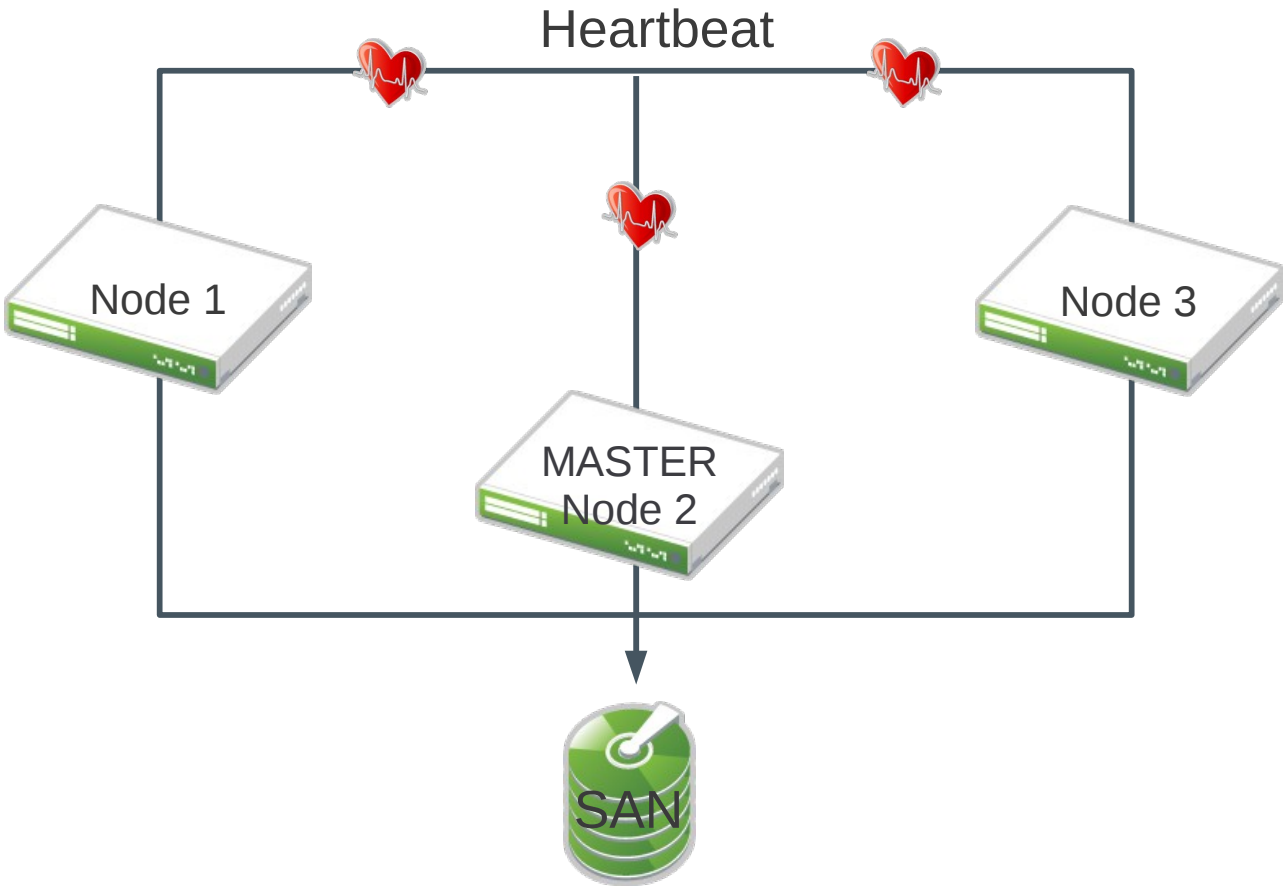
- **SUSE Linux Enterprise Server is #1**
  - in mainframe Linux market (80%+ share)
  - in SAP-on-Linux market (75% share)
  - in High Performance Computing (6 of top 10)
- **SUSE Linux Enterprise Server for System z:**
  - Fully supported by IBM – supports all benefits of the mainframe
  - 10 years of expertise (available five years ahead of competition)
  - Ideal for workload consolidation, providing major cost savings
  - New features specific to System z
  - Hosting of Subscription Management Tool on System z
  - More than 1,700 certified applications available

Differentiators - Unique tools for SUSE Linux Enterprise Server for System z

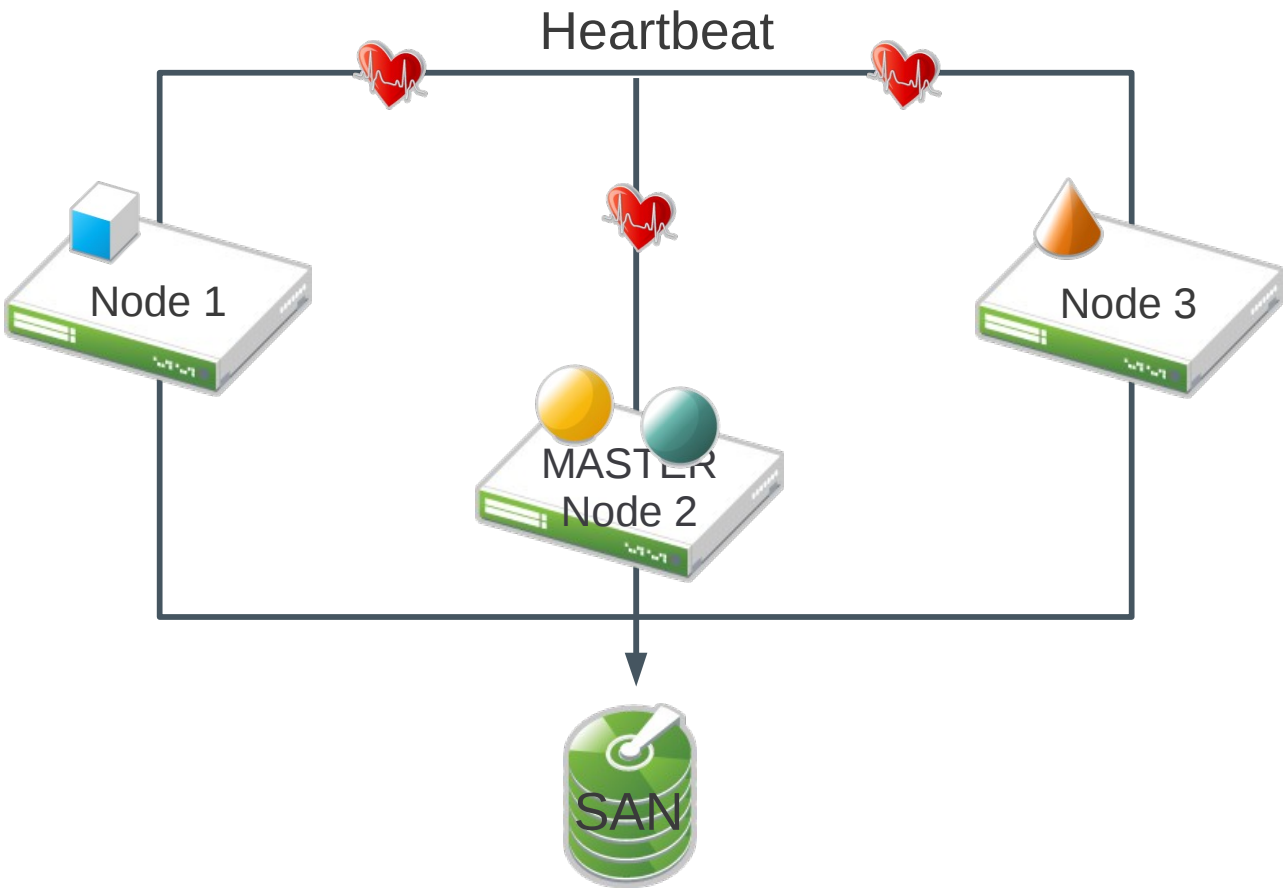
- **Yast and Integrated Systems Management**
  Configure every aspect of the server
- **Subscription Management Tool Hosting**
  Subscription and patch management made easy
- **High Availability Extension for SLES**
  Included in SLES for System z
- **.NET Applications on Linux: Mono**
  Migrate existing .NET applications to Linux without having to rewrite code
- **Starter System for System z**
  Starter System for System z is a pre-built installation server

What is a high availability (HA) cluster?

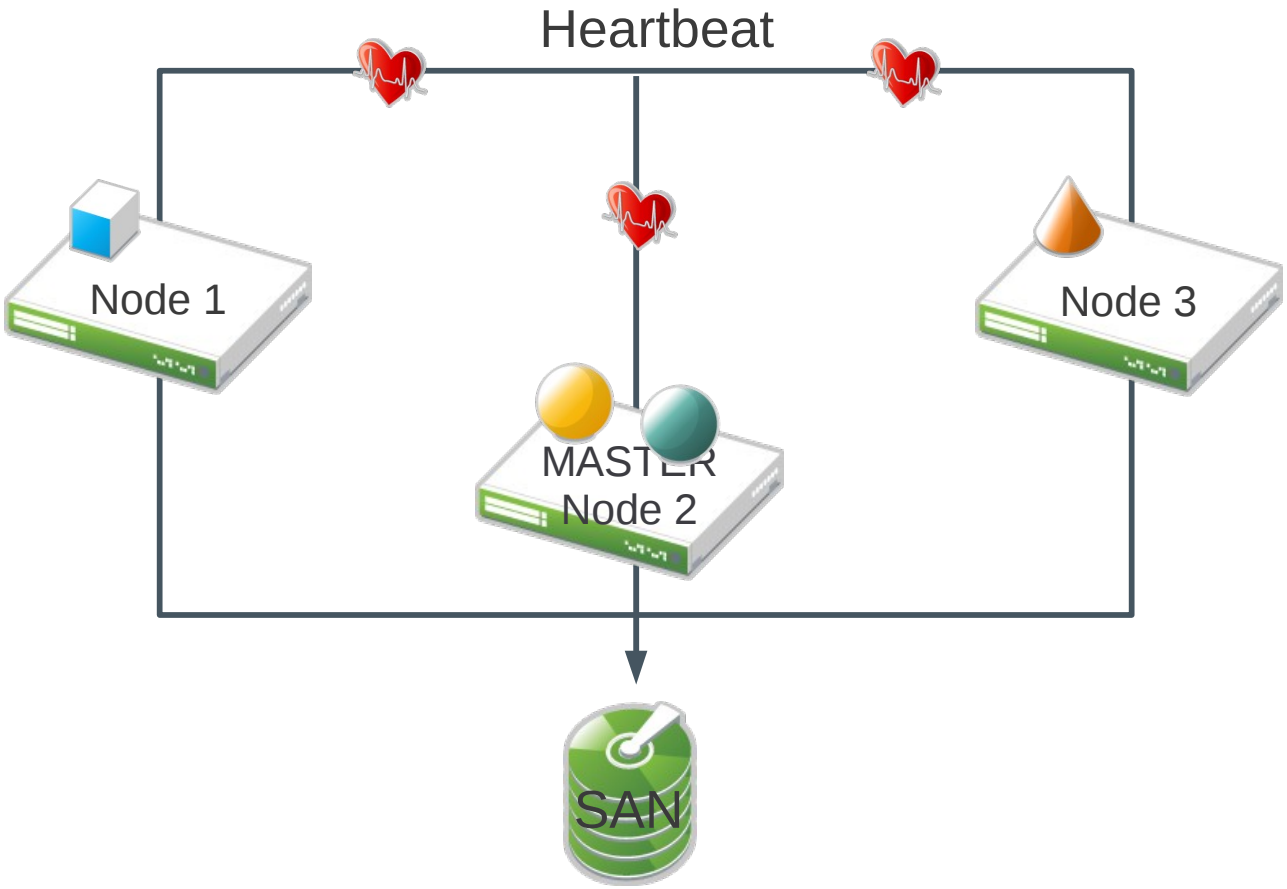
A Simple HA Cluster



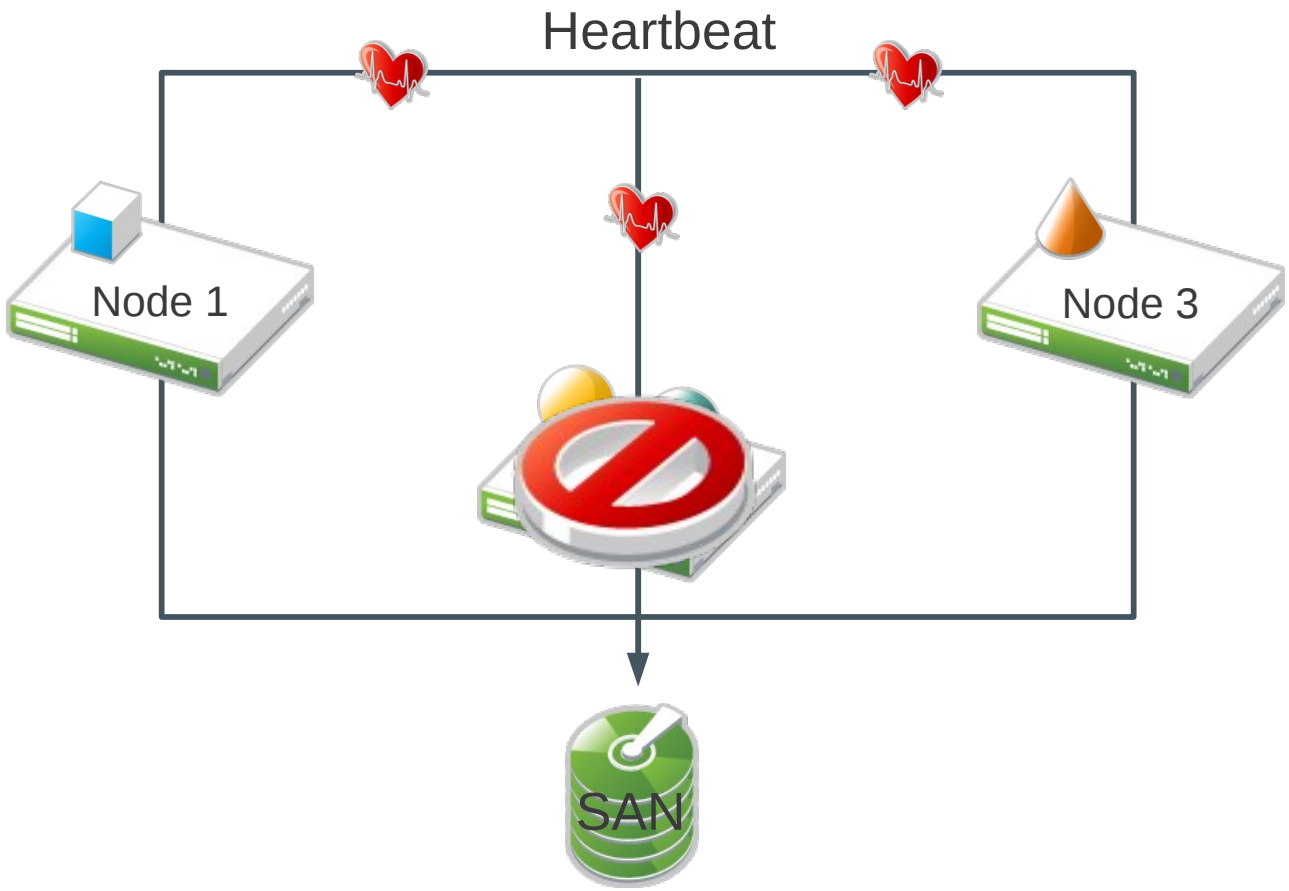
Resources Running in the Cluster



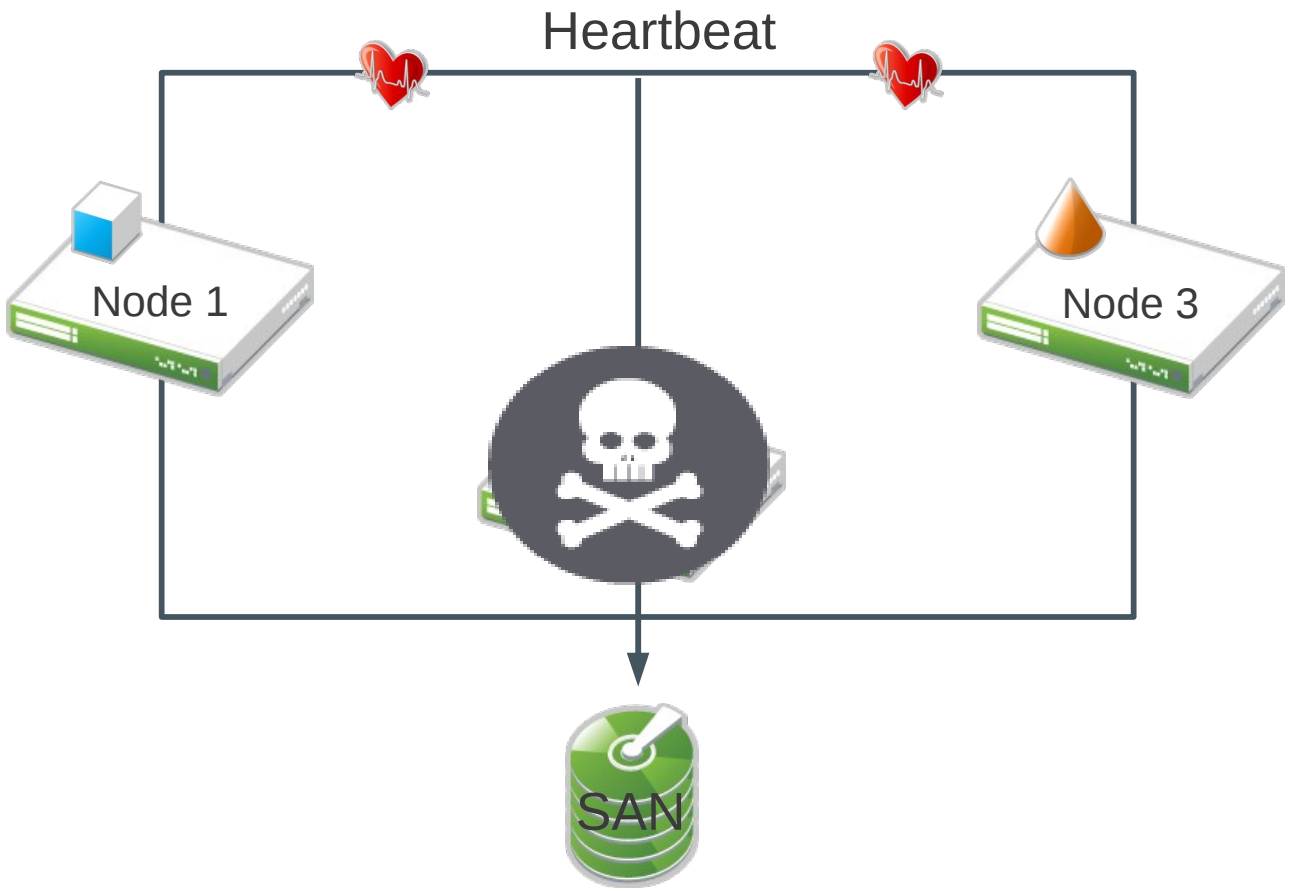
Migrating a Resource



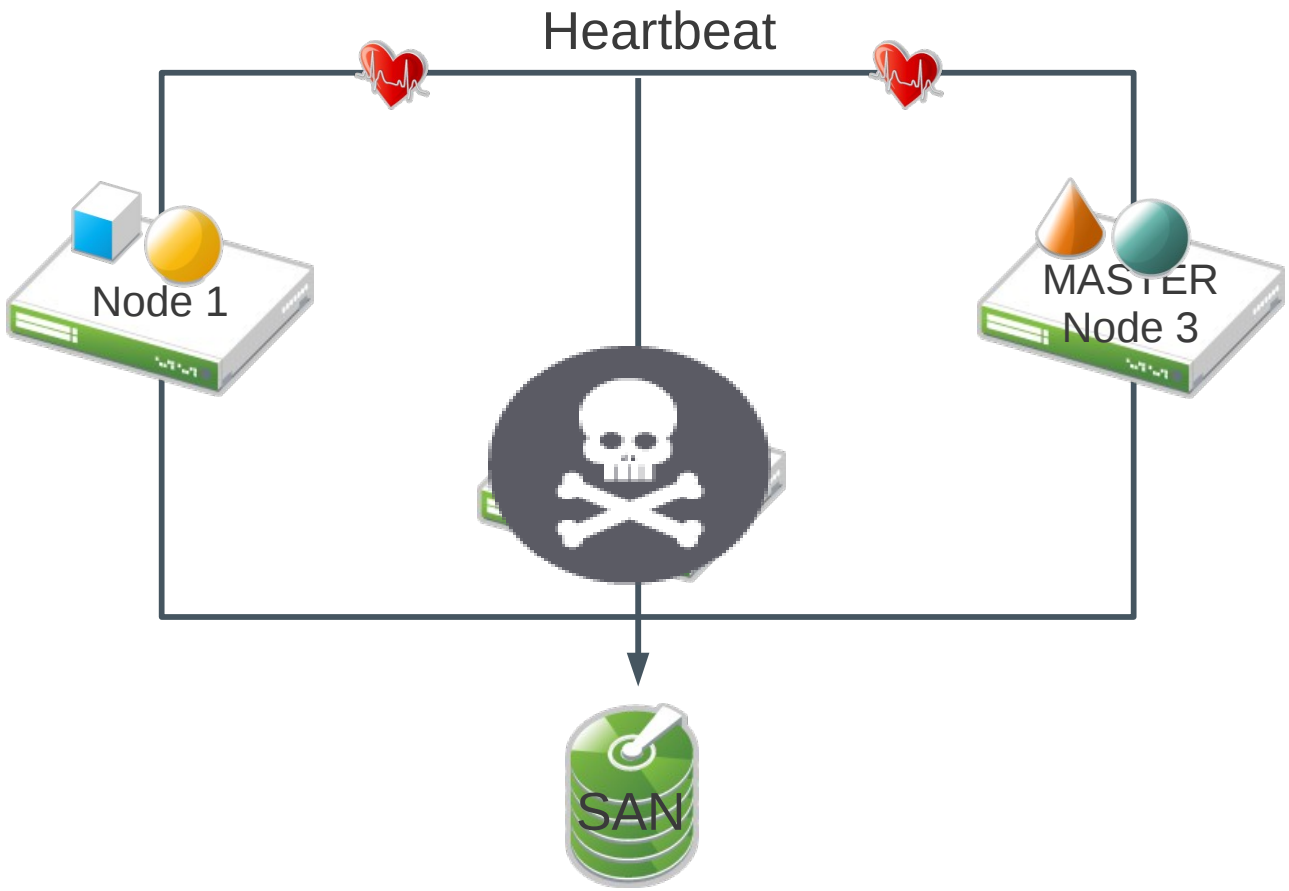
Node Failure in the Cluster



STONITH the Failed Node Out of the Cluster



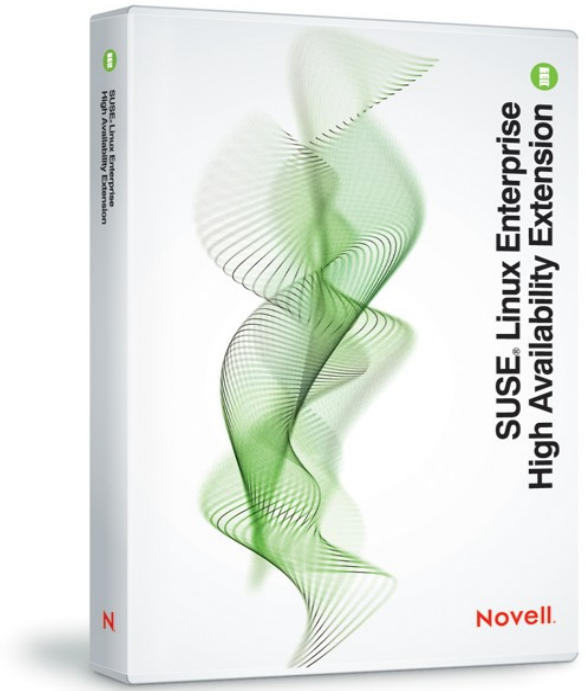
Resources brought up on other nodes the Cluster



What is required to build a HA cluster using SLES?

SUSE® Linux Enterprise High Availability Extension

- An affordable, integrated suite of robust open source clustering technologies
- Used with SUSE Linux Enterprise Server
- Implement a cluster using physical or virtual Linux servers
- Benefits
  - Cost effectively meet your service-level agreements
  - Ensure continuous access to your mission-critical systems and data
  - Maintain data integrity
  - Increase resource utilization



SUSE® Linux Enterprise High Availability Extension 11 SP1

Key Capabilities Today

- All components use open source technologies
- Flexible, policy-driven clustering solution
  - Metro area cluster up to 20 miles
  - Clustered SAMBA (CIFS)
  - Includes 82 Resource Agents like Oracle, DB2, SAP, Apache, MySQL, PostgreSQL
- Cluster aware file system and volume manager
  - OCFS2 shared-disk POSIX-compliant generic purpose cluster file system
  - Clustering extensions to the standard LVM2 toolset
- Host Based Continuous Data Replication
- Disaster Recovery Integration
- User Friendly Tools
  - No matter if you prefer the CLI or GUI

SUSE® Linux Enterprise High Availability Extension

x86 and x86_64

- Additional cost per year, per server
- Support level inherited by base SUSE Linux Enterprise Server

System z, Power, Itanium

- Bundled with base SUSE Linux Enterprise Server at no additional charge
- Support level inherited by base SUSE Linux Enterprise Server

SUSE® Linux Enterprise High Availability Extension
Key Capabilities Future

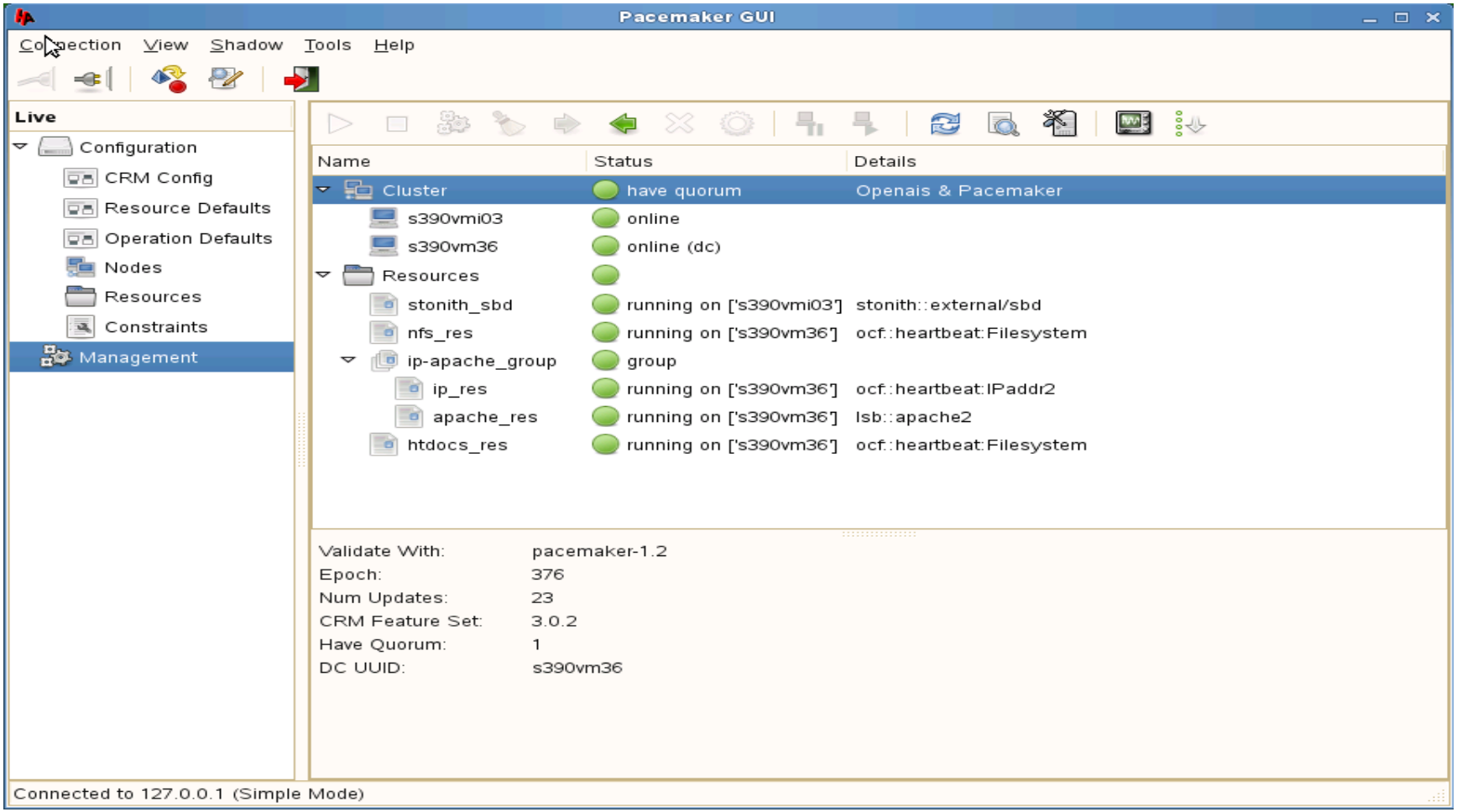
- Usability and Management
  - Full web GUI
  - Improved access control
- Ease of Use
  - Guided and automated configuration
  - Prepackages applications
  - Preloaded clusters
- Clusters Functionality
  - Wide area clusters
  - Improved data replication
  - Unix cluster stack leadership
- Backup and Disaster Recovery
  - Backup integration
  - DR automation

Demoing the features:

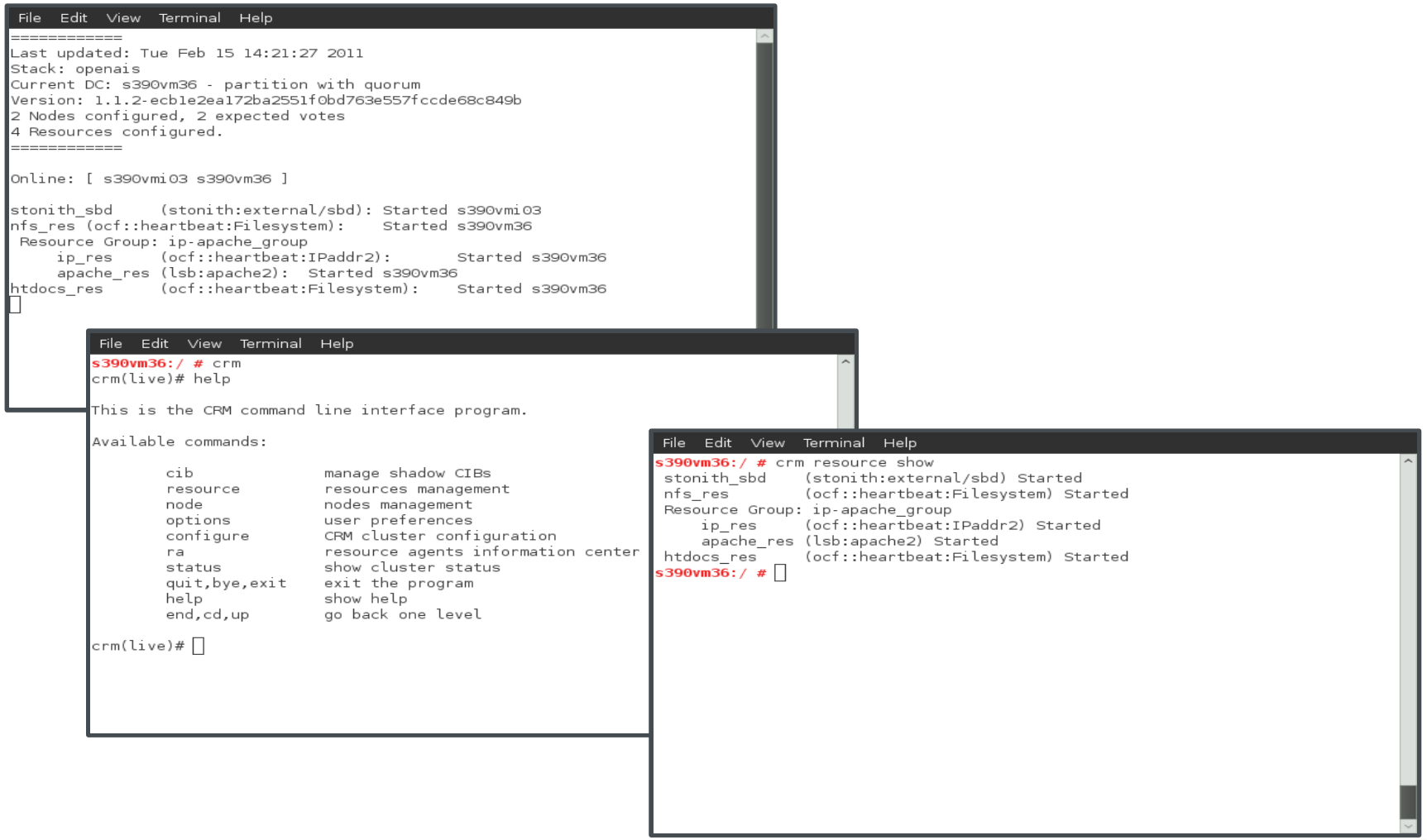
Managing a cluster with the GUI and CLI

NOTE: Screenshots are provided to help visualize the demoed features during the session.

Start the GUI with crm_gui



Use crm_mon and crm for the CLI

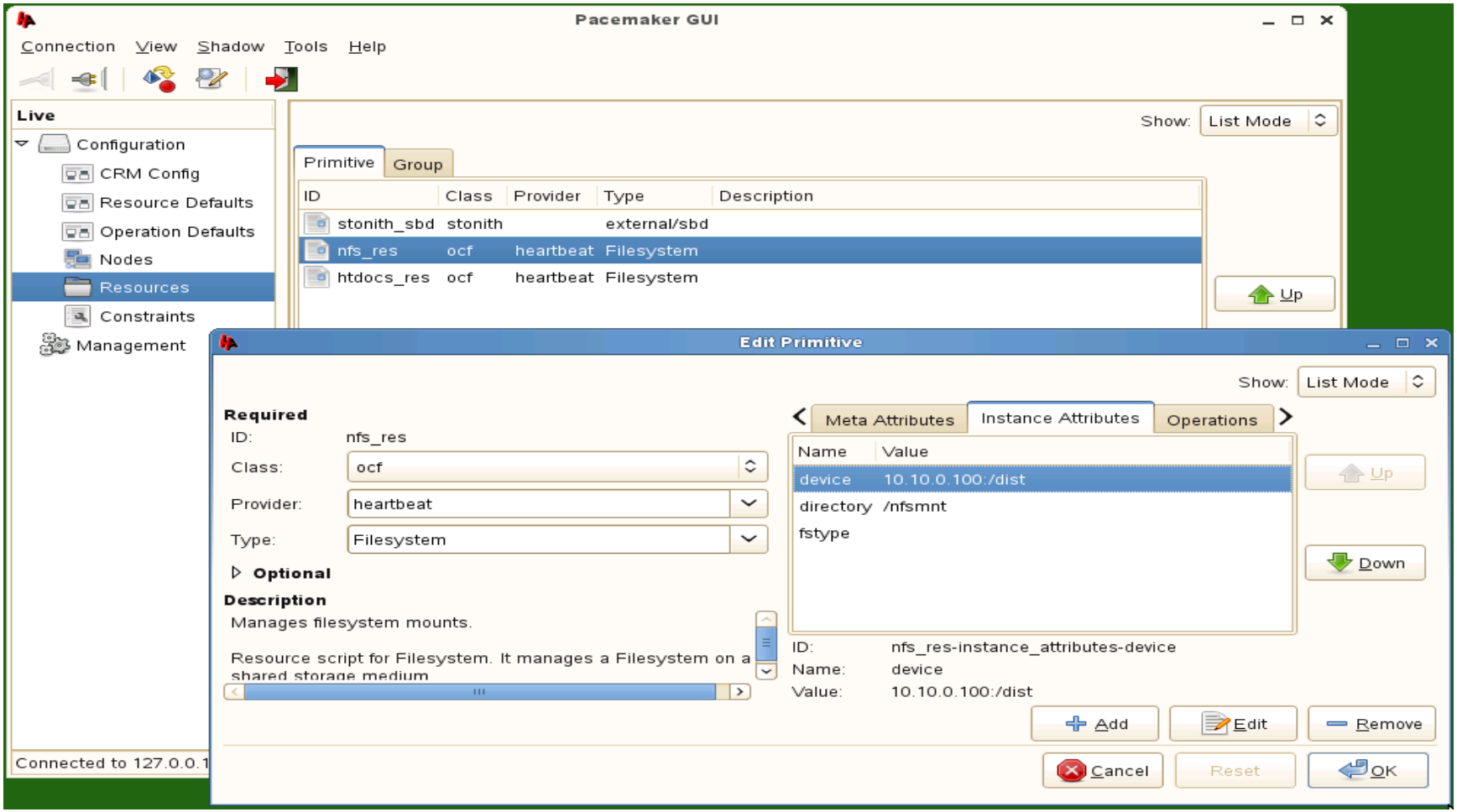


Demoing the features:

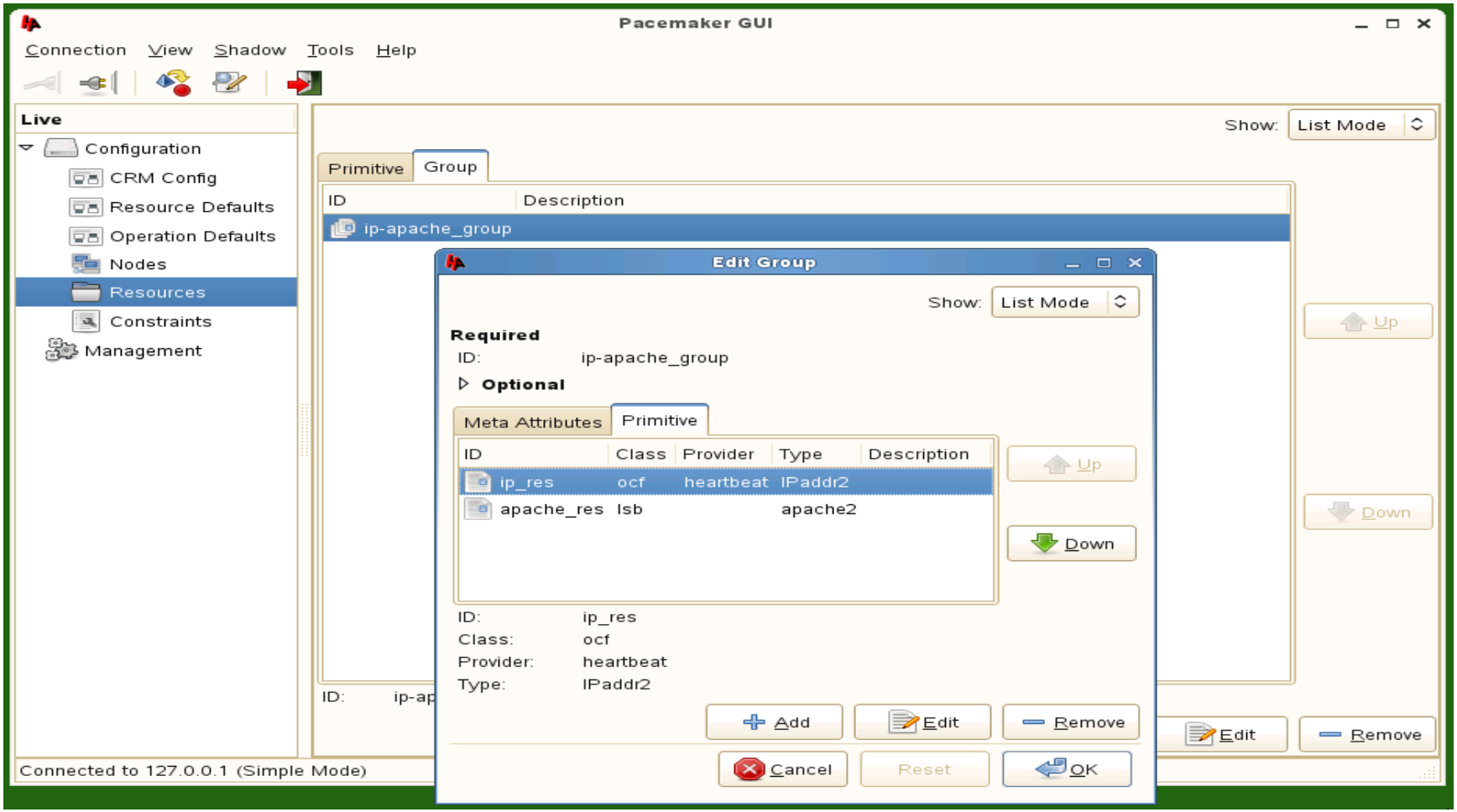
Resource primitives and resource groups

NOTE: Screenshots are provided to help visualize the demoed features during the session.

A resource primitive



A resource group

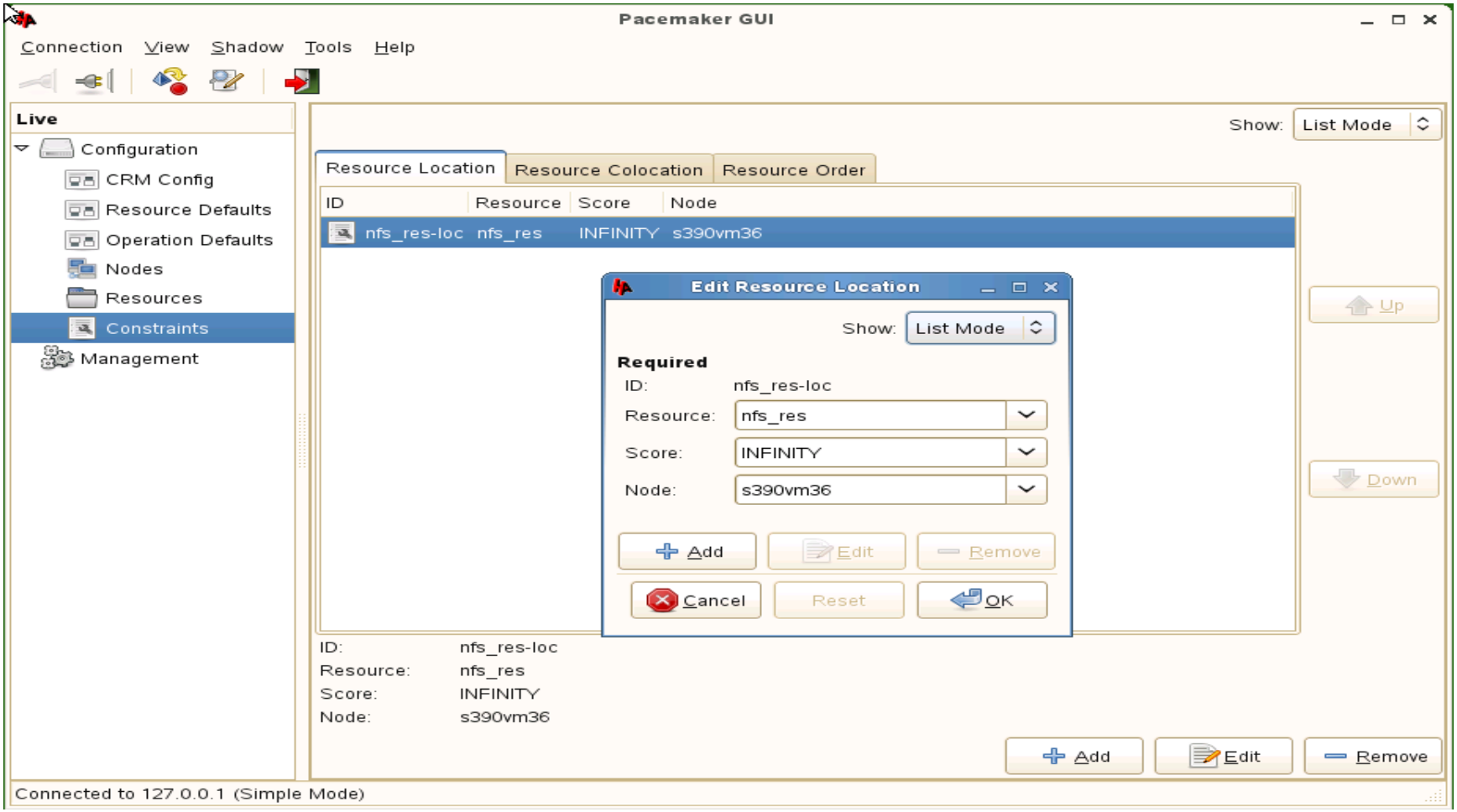


Demoing the features:

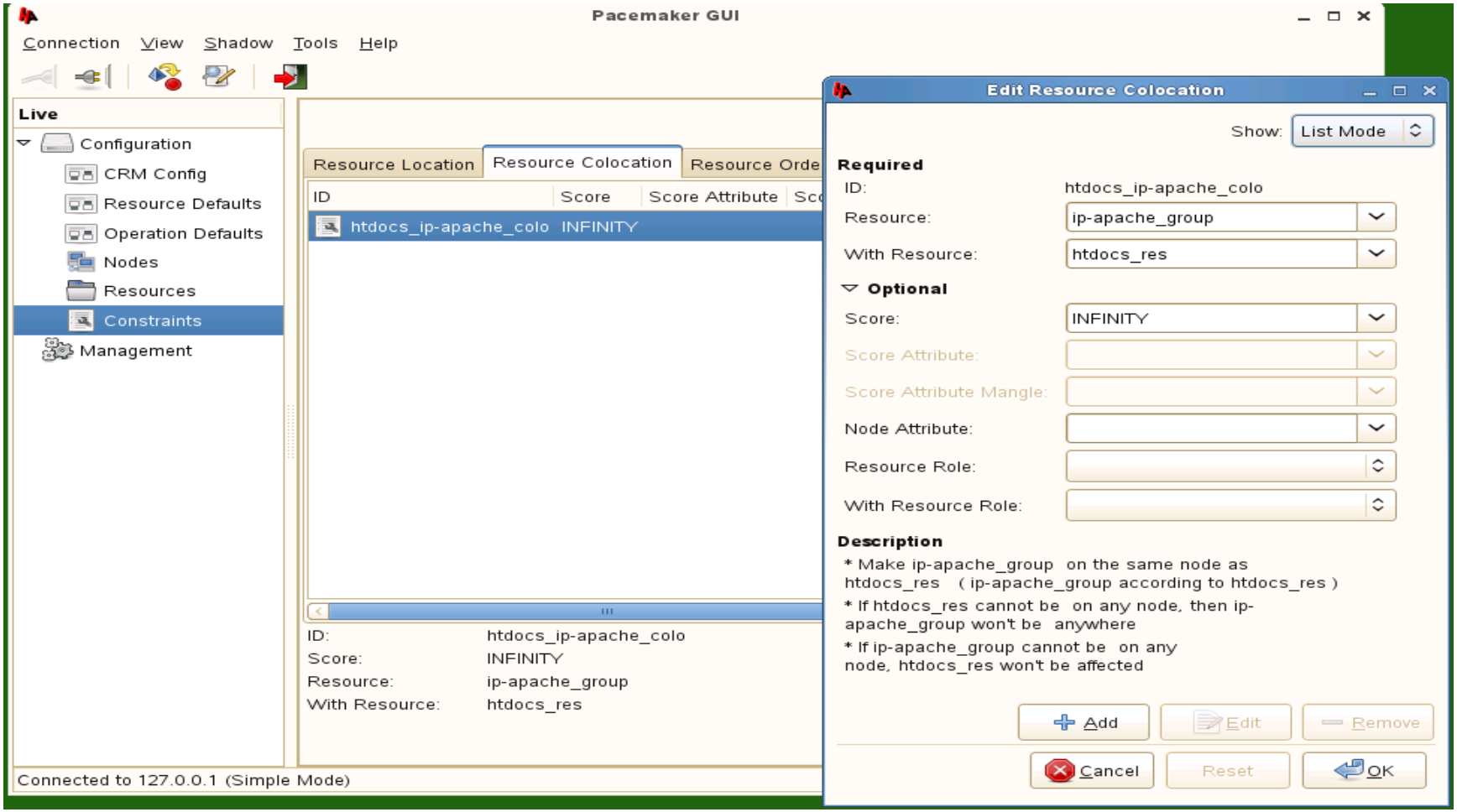
Resource constraints

NOTE: Screenshots are provided to help visualize the demoed features during the session.

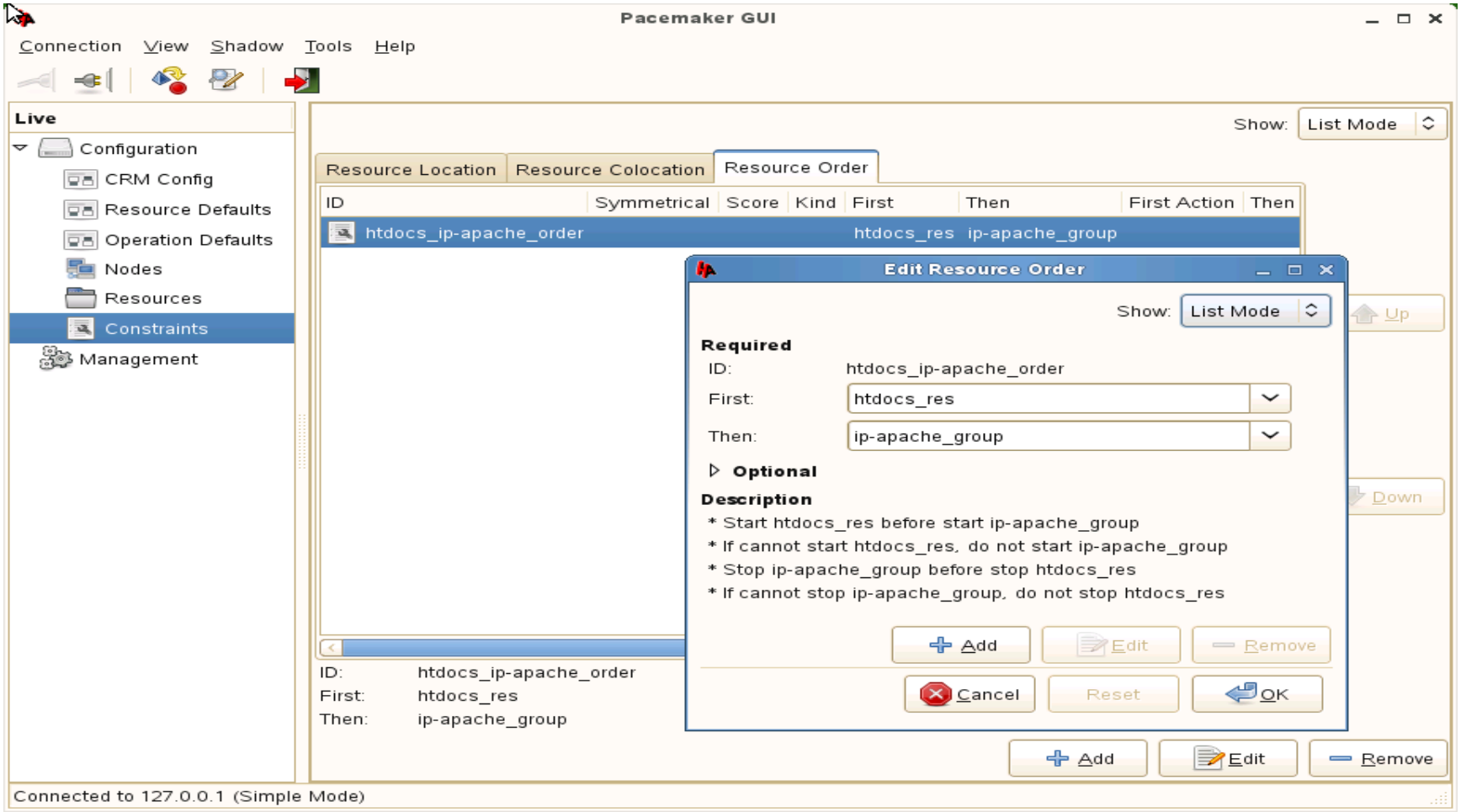
A resource location constraint



A resource colocation constraint



A resource order constraint



Demoing the features:

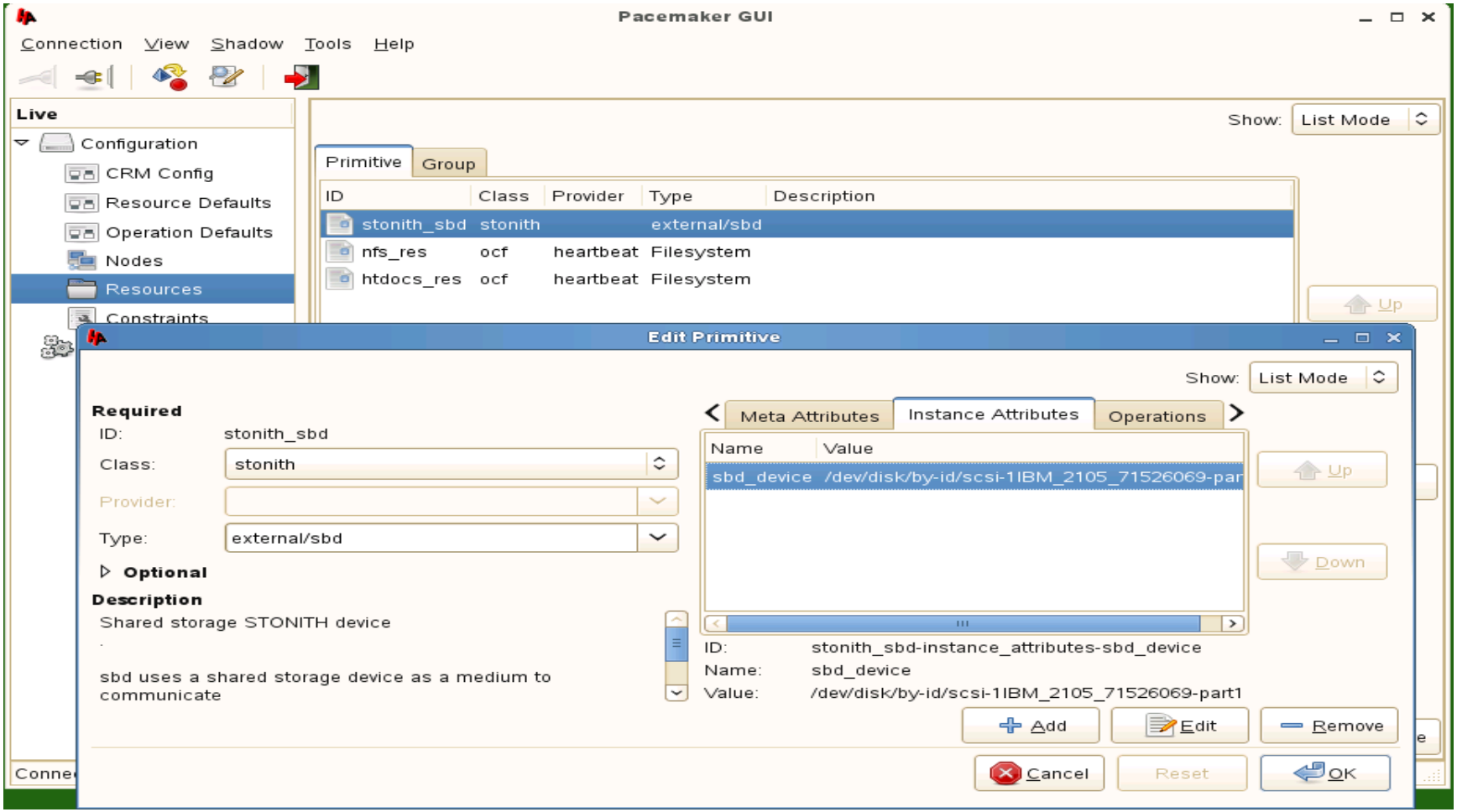
STONITH

NOTE: Screenshots are provided to help visualize the demoed features during the session.

What is STONITH?

- **S**hoot **T**he **O**ther **N**ode **I**n **T**he **H**ead
- Simple concept
  - A machine in the cluster wants to make sure another machine in the cluster is dead
  - STONITH is used to remotely power down a node in the cluster
  - Simple and reliable, albeit admittedly brutal
- Fencing is another term but not as graphic!
- Modular and extensible
  - 33 STONITH modules included in SLE11 SP1 HAE
  - Two of interest for System z: SBD and snIPL
- SLE HAE requires a STONITH device by default
  - Recommended practice to have one configured!

A Split Brain Detector (SBD) STONITH resource



Demoing the features:

cLVM and OCFS2

NOTE: Screenshots are provided to help visualize the demoed features during the session.

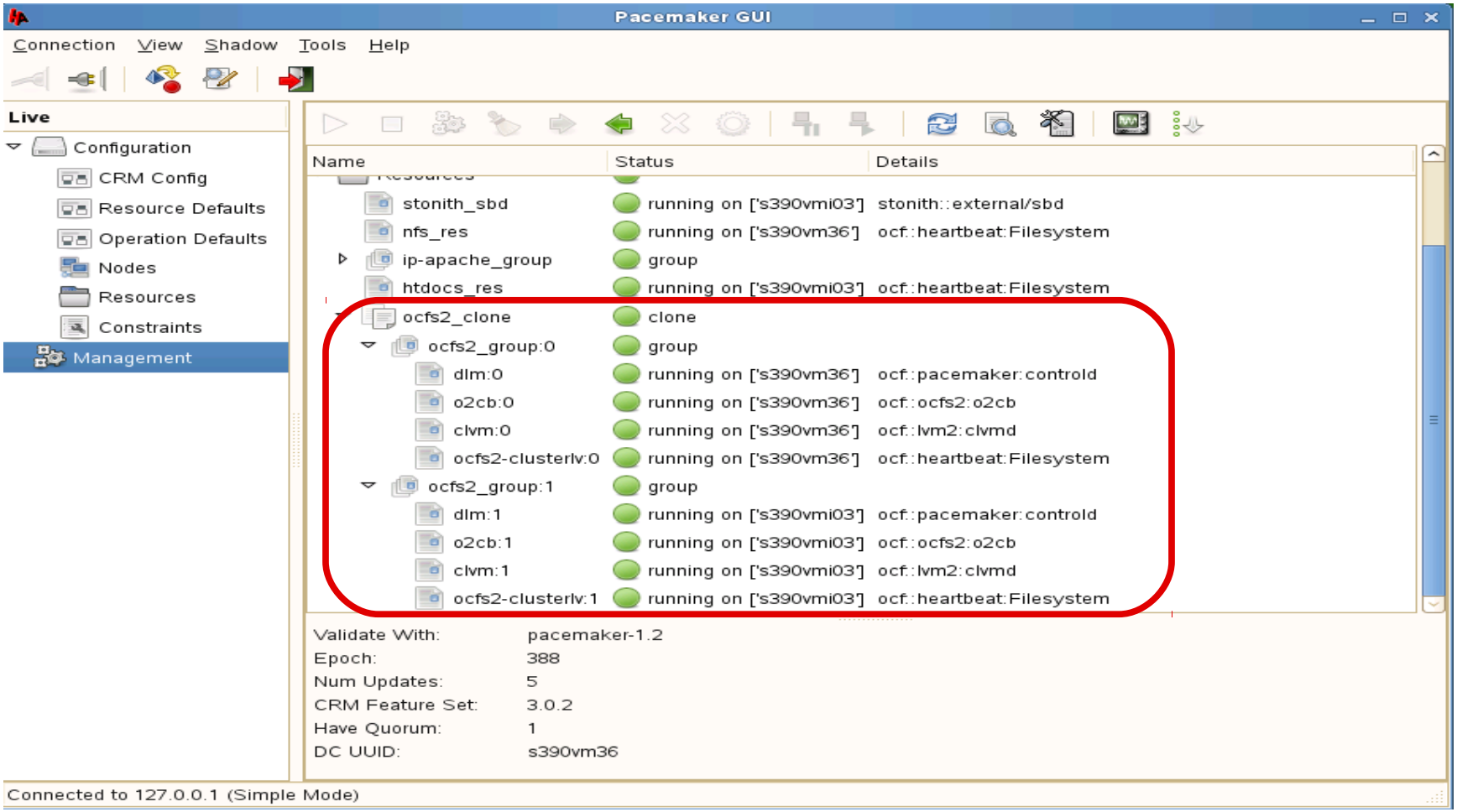
Understanding the definitions of cLVM and OCFS2 in the HA cluster

- cLVM
  - Cluster-aware logical volume manager uses the same LVM management tools to manage PVs, VGs and LVs
- OCFS2
  - Oracle Clustered File System v2
- dlm
  - Distributed Lock Manager manages locking within the cluster
- o2cb
  - OCFS2 cluster software stack
- Cloned resource
  - a resource or resource group that runs on all nodes in the cluster

Understanding the configuration of cLVM and OCFS2 in the HA cluster

- Four resource primitives in a cloned resource group (primitive names are arbitrary)
  - dlm
  - o2cb
  - clvm
  - ocfs2-clusterlv
- Resource primitive start order is important
- The last resource primitive mounts the clustered filesystem on all nodes in the cluster

The cLVM and OCFS2 configuration



The mounted OCFS2 filesystem

```
File  Edit  View  Terminal  Help
s390vmi03:~ # mount
/dev/dasda2 on / type ext3 (rw,acl,user_xattr)
proc on /proc type proc (rw)
sysfs on /sys type sysfs (rw)
debugfs on /sys/kernel/debug type debugfs (rw)
devtmpfs on /dev type devtmpfs (rw,mode=0755)
tmpfs on /dev/shm type tmpfs (rw,mode=1777)
devpts on /dev/pts type devpts (rw,mode=0620,gid=5)
fusectl on /sys/fs/fuse/connections type fusectl (rw)
securityfs on /sys/kernel/security type securityfs (rw)
gvfs-fuse-daemon on /root/.gvfs type fuse.gvfs-fuse-daemon (rw,nosuid,nodev)
/dev/sdb2 on /media/disk-10 type ext3 (rw,nosuid,nodev)
/dev/sdb2 on /srv/www/htdocs type ext3 (rw)
none on /sys/kernel/config type configfs (rw)
/dev/mapper/clustervg-clusterlv on /ocfs2mnt type ocfs2 (rw,_netdev,acl,cluster_
stack=pcmk)
s390vmi03:~ #
```

Attend the SUSE Linux High Availability Extensions Hands-on Workshop (9348 and 9494).

Monday 3 – 6pm.

Thank you.

Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
+www.suse.com

Join us on:
www.opensuse.org