The Road to High Availability
A solution built on open source

Dejan Muhamedagic
Senior Software Engineer
SUSE Labs, dejan@suse.de
Availability, reliability and redundancy

\[ \text{Availability} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}} \]

- Maximize MTBF and minimize MTTR
- Recovery should be automatic
- Hardware and software should be robust and reliable
- Replaceable parts should be redundant
- Single Points Of Failure (SPOF) should be eliminated
Basic idea

A. Robertson (http://linux-ha.org)
Cluster terminology and concepts

• Nodes
  - Nodes are computers and computers are nodes

• Resources
  - Resources are applications

• Quorum (n/2 + 1)
  - There is always exactly one partition which has quorum

• Fencing
  - Fencing establishes the state of a node or a resource with certainty

• Split brain
  - Cluster splits into two or more partitions due to communication failure
An example of a failover cluster
A failover cluster (cont'd)
Life is hard
Two-node clusters

• No quorum
  - Two node clusters do not have quorum
  - Unless there is a tiebreaker (quorum disk or quorum server)

• Fencing replaces quorum
  - Since there is no quorum, a node has to make sure that the other node is down

• Possibility of a death match
  - Both nodes start shooting each other
How many nodes?

- Depends on the application
- Prefer *odd* number of nodes
Applications in clusters

• Know your applications
  – Make sure that they can handle a failover gracefully
  – Test applications under stress

• Resource agents
  – The tool which handles applications
  – They work autonomously
  – Use the supplied resource agents instead of LSB init scripts

• Take care of timeouts
  – Avoid false positives
  – Be generous rather than stingy
Cluster stack w/pacemaker

A. Beekhof (http://clusterlabs.org)
## Cluster elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primitive</td>
<td>Represents a resource. A basic building block.</td>
</tr>
<tr>
<td>Group</td>
<td>A set of primitives which run together and must be started in certain order.</td>
</tr>
<tr>
<td>Clone</td>
<td>Makes several copies of a resource.</td>
</tr>
<tr>
<td>Master</td>
<td>Multi-state resources.</td>
</tr>
<tr>
<td>Collocation</td>
<td>Resource relation in terms of placement.</td>
</tr>
<tr>
<td>Order</td>
<td>Resource relation in terms of action order.</td>
</tr>
<tr>
<td>Location</td>
<td>Node preference.</td>
</tr>
<tr>
<td>Property</td>
<td>Cluster property.</td>
</tr>
</tbody>
</table>
Sample pacemaker/CRM cluster
Sample cluster (crm shell)

```
node xen-b
node xen-c
primitive ap0 ocf:heartbeat:apache \
  params configfile="/mnt/drbd/conf/httpd.conf" \
  op monitor interval="30s" timeout="30s"
primitive drbd0 ocf:heartbeat:drbd \
  params drbd_resource="r0"
primitive fs0 ocf:heartbeat:Filesystem \
  params device="/dev/drbd0" directory="/mnt/drbd" fstype="ext3"
primitive ip0 ocf:heartbeat:IPaddr \
  params ip="10.2.13.100"
primitive st0 stonith:external/xen0 \
  params dom0="rondo" hostlist="xen-b xen-c" \
  op monitor interval="2h" timeout="60s"
group webservice ip0 ap0
ms ms-drbd0 drbd0 \
  meta notify="true"
clone c-fence st0
colocation fs0-with-drbd0 inf: fs0:Started ms-drbd0:Master
colocation webservice-with-fs0 inf: webservice fs0
order drbd0-before-fs0 inf: ms-drbd0:promote fs0:start
order fs0-before-webservice inf: fs0 webservice
property $id="cib-bootstrap-options" \
  no-quorum-policy="ignore"
crm(live)configure# 
```
Sample cluster (GUI)
Unpublished Work of Novell, Inc. All Rights Reserved.

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