PCS

What’s new since 2017

Tomáš Jelínek

February 2020

tojeline@redhat.com
PCS? I’ve never heard of it...
PCS — Pacemaker/Corosync Configuration System

- Corosync — corosync.conf, tools, quorum
- Nodes — start, stop, add, remove
- Pacemaker — CIB, tools, metadata
- Resource and fence agents
- Distributing files (Corosync, Booth, SBD, authkeys)

- Corosync QDevice
- SBD — with shared storage or watchdog only
- Booth cluster ticket manager
Agenda

- Pcs branches
- Node names
- Corosync 3 + knet
- New features
- Retrospective
- Plans
- Q & A
PCS branches
PCS branches

PCS 0.9.x

- Corosync 2 or Corosync 1 with CMAN
- Pacemaker 1.1
- Python 2.7+ and Python 3
- Ruby 2.0+

- Maintenance mode
PCS branches

PCS 0.10.x

- Corosync 3 with knet
- Pacemaker 2
- Python 3.6+
- Ruby 2.2+

- Active development
Node names
Node names in cluster components

- **Corosync**
  - Node ID
  - One address per node per ring / link – IP or name
  - Node name – for other components

- **Pacemaker**
  - Node name – from Corosync or hostname

- **Pcs 0.9.x**
  - Corosync nodes – ring0 addresses
  - Pacemaker nodes – CIB / status names
  - Pcs nodes – pcs auth
  - Names may be different in each context
Solution

- Make node names the node identifiers
  - Pcs pushes them to cluster components

- Addresses
  - Each cluster component can have its own addresses
  - Optional
  - Pcs – node addresses default to node names
  - Corosync – node addresses default to addresses from pcs
  - Pacemaker (remote and guest) – node addresses default to addresses from pcs
Node names in pcs

[node1]# pcs host auth node1 node2 node3

[node1]# pcs host auth \
  node1 addr=10.0.0.11  \
  node2 addr=10.0.0.12:2225  \
  node3 addr=node3.example.com
Corosync 3
knet
Cluster setup – syntax

```bash
pcs cluster setup <cluster name>
 (<node name> [addr=<node address>]...)...
 [transport knet|udp|udpu
  [<transport options>]
  [link <link options>]...
  [compression <compression options>]
  [crypto <crypto options>]
 ]
 [totem <totem options>]
 [quorum <quorum options>]
 [--enable] [--start [--wait[=<n>]]]
```
Cluster setup – examples

[node1]# pcs cluster setup newcluster node1 node2 \ 
   --enable --start

[node1]# pcs cluster setup newcluster \ 
   node1 addr=10.0.1.11 addr=10.0.2.11 \ 
   addr=10.0.3.11 addr=10.0.4.11 \ 
   node2 addr=10.0.1.12 addr=10.0.2.12 \ 
   addr=10.0.3.12 addr=10.0.4.12 \ 
   transport knet \ 
   link linknumber=3 mcastport=55405 \ 
   link linknumber=1 transport=sctp

Corosync 3 + knet
Cluster setup – what it does

- Validations
- Delete old cluster config files from nodes if any
- Send pcs tokens to nodes
- Create and send corosync and pacemaker authkeys to nodes
- Create and send corosync.conf to nodes
  - Some options are set automatically (two_node, auto_tie_breaker)
- Enable and start cluster daemons
Nodes

[node1]# pcs cluster node add node3 --enable --start

[node1]# pcs cluster node add node3 \\
addr=10.0.1.13 addr=10.0.2.13 \\
addr=10.0.3.13 addr=10.0.4.13 \\
--enable --start
Links

[node1]# pcs cluster link add \
    node1=10.0.5.11 node2=10.0.5.12 node3=10.0.5.31 \ 
    options linknumber=5

[node1]# pcs cluster link update 5 \ 
    node3=10.0.5.13 \ 
    options transport=sctp

[node1]# pcs cluster link delete 5 
[node1]# pcs cluster link remove 5
New features
Safe disable

Do not disable a resource if it would have an effect on other resources

```bash
pcs resource disable <resource id>...
   [--wait[=n]]
   [--safe [--no-strict]] [--simulate]
```

```bash
pcs resource safe-disable <resource id>...
   [--wait[=n]]
   [--no-strict] [--simulate] [--force]
```
Resource relations

[node1]# pcs resource relations res-group

```
res-group
|-- inner resource(s)
|   |-- members: r1 r2
|   |   |-- r1
|   |   |-- r2
|   `-- order
|       |-- start r3-clone then start res-group
|       `-- r3-clone
|           |-- inner resource(s)
|           `-- r3
```
Retrospective
Retrospective

- Input data validations, current cluster status considered
- Reporting as many errors at once as possible
  - Previously one error per run
- Messages are more user friendly (hints)
- Reporting effects of `--force`
- Multiple operand commands (node remove)
Validations example – pcs cluster setup 1/4

- Command syntax
  - Section keywords
  - name=value pairs
  - No duplicities
  - If the syntax is not valid, pcs does not proceed

- Nodes are known to pcs (pcs host auth)

- Corosync options
  - Transport, link, crypto, compression, totem, quorum
  - Option names
  - Option values – an enum, an integer (positive, negative, zero allowed)
  - Combination of options
Validations example – pcs cluster setup 2/4

▷ Nodes and links
  - All nodes have addresses for all links
  - No duplication in node addresses and names
  - Number of links (depends on the transport)
  - Mixing IPv4 and IPv6 in one link
  - Addresses match ip_version option (if specified)
  - Addresses are resolvable

▷ Node status check
  - Reachable
  - Not in a cluster (full-stack or remote nodes, configs and services)
  - Version of cluster components
Validations example – pcs cluster setup 3/4

```bash
# pcs cluster setup newCluster \
rh81-node1 addr=10.0.0.1 addr=::1 \
rh81-node22 addr=10.0.0.2 \
transport knet x=y ip_version=ipv4 \
link linknumber=4 ping_timeout=a
```
Validations example – pcs cluster setup 4/4

Error: Host 'rh81-node22' is not known to pcs, try to authenticate the host using 'pcs host auth rh81-node22' command

Error: Address '::1' cannot be used in link '1' because the link uses IPv4 addresses

Error: All nodes must have the same number of addresses; node 'rh81-node1' has 2 addresses; node 'rh81-node22' has 1 address

Error: invalid knet transport option 'x', allowed options are: 'ip_version', 'knet_pmtud_interval', 'link_mode'

Error: Cannot set options for non-existent link '4', 2 links are defined starting with link 0

Error: 'a' is not a valid ping_timeout value, use a non-negative integer

Error: If link option 'ping_timeout' is specified, link option 'ping_interval' must be specified as well

Error: rh81-node1: Running cluster services: 'corosync', 'pacemaker', the host seems to be in a cluster already, use --force to override

Error: rh81-node1: Cluster configuration files found, the host seems to be in a cluster already, use --force to override

Error: Some nodes are already in a cluster. Enforcing this will destroy existing cluster on those nodes. You should remove the nodes from their clusters instead to keep the clusters working properly, use --force to override

Error: Errors have occurred, therefore pcs is unable to continue
Plans
Near future

- Change corosync settings in a running / existing cluster
- Pacemaker tags
- --simulate
Other plans

- New web UI
  - In progress – dashboard, resources management

- REST API
  - Working alpha for a few commands
  - Internal use, changes are expected

- Asynchronous requests

- Running commands on remote hosts
## HA Cluster Management

### Clusters

<table>
<thead>
<tr>
<th>Clusters</th>
<th>Issues</th>
<th>Nodes</th>
<th>Resources</th>
<th>Fence devices</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster-1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>cluster-2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>cluster-3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

### Node Status

<table>
<thead>
<tr>
<th>Node</th>
<th>Status</th>
<th>Quorum</th>
</tr>
</thead>
<tbody>
<tr>
<td>node-1</td>
<td>Online</td>
<td>Yes</td>
</tr>
<tr>
<td>node-2</td>
<td>Offline</td>
<td>No</td>
</tr>
<tr>
<td>node-3</td>
<td>Online</td>
<td>Yes</td>
</tr>
</tbody>
</table>

### Cluster 4

- 1 issue
- 2 nodes
- 4 resources
- 0 fence devices

⚠️ No fencing configured in the cluster
HA Cluster Management

Clusters > cluster-2

UPDATE INSTANCE ATTRIBUTES OF RESOURCE "A" SUCCESSFULLY UPDATED

FILES (ONE OR MORE) WHICH CONTAIN Extra ENVIRONMENT VARIABLES. IF YOU WANT TO PREVENT SCRIPT FROM READING THE DEFAULT FILE, SET THIS PARAMETER TO EMPT}
Thank you