JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

VDCF - Virtual Datacenter Cloud Framework

for the Solaris™ Operating System

Sun/Solaris Cluster Guide

Version 2.3
4. September 2018

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

Contents

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

1 Introduction

This documentation describes the Sun/Solaris Cluster Feature of the Virtual Datacenter Cloud Framework (VDCF) for the Solaris Operating System, Version 2.3 and explains how to use this feature.

See:

VDCF – Administration for detailed information about this product, the commands and arguments.

1.1 Overview

The Sun/Solaris Cluster products supports the failover of Solaris 10 Zones between Cluster Nodes. Using the VDCF Sun/Solaris Cluster Feature enables you to create Datasets and VDCF vServers (Zones) including Sun/Solaris Cluster configuration needed to use the automatic failover feature of Sun/Solaris Cluster.

Since VDCF Sun/Solaris Cluster Version 2.0 it is supported to integrate LDoms (VDCF GDoms) which run on Solaris 11 Control Domains into Solaris Cluster Version 4.1, 4.2, 4.3 and 4.4. This automated integration makes sure your LDoms are started on a second Cluster Node, if your primary Node fails.

1.2 Installation

The JSvdcf-suncluster package requires the JSvdcf-base package Version 7.1.0 or later.
On your Cluster Nodes the JSvdcf-client package 7.1.0 or later is required.

a) sparc platform

```
cd </cdrom/cdrom0>/vdcf/sparc

pkgadd -d ./JSvdcf-suncluster_<version>_sparc.pkg
```

b) i386 platform

```
cd </cdrom/cdrom0>/vdcf/i386

pkgadd -d ./JSvdcf-suncluster_<version>_i386.pkg
```

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

2 Sun/Solaris Cluster Installation and Configuration

2.1 Overview

Currently the Sun/Solaris Cluster Software must be installed manually on a Node, which was previously installed using the VDCF framework. The VDCF Administration Guide contains information how to install a Node using the VDCF framework.

2.2 Patch Requirements (Sun Cluster 3.2 only)

Sun Cluster requires a /globaldevices Filesysteme with 512MB, which must be configured in the /var/opt/jomasoft/vdcf/conf/<node>_partitioning.cfg before installing the Node.

VDCF requires to install the following or later Sun Cluster Patches

126106-13 / Synopsis: Sun Cluster 3.2: CORE patch for Solaris 10
Date: Apr/25/2008

2.3 Required Cluster Features

For Zones failover the Feature “Sun Cluster HA for Solaris Container” must be installed on the Compute Nodes.

For LDoms failover the Feature “HA Ldom” must be installed on the Compute Nodes.
The required package is named “ha-cluster/data-service/ha-ldom”

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

3 VDCF Sun/Solaris Cluster - Configuration

3.1 VDCF Sun/Solaris Cluster – Configuration Variables

You may override the Default Cluster Resource prefixes with your own Naming Conventions by adding the following variables in /var/opt/jomasoft/vdcf/conf/customize.cfg

export SC_POSTFIX_RESGROUP=rg

3.1.1 vServer/Zone Failover

export SC_POSTFIX_DATASET=hasp
export SC_POSTFIX_ZONEPATH=hasp_zonepath
export SC_POSTFIX_ZONEBOOT=sczbt

For the Sun Zone Boot Resource (_sczbt) timeouts may optionally be configured:

export SC_START_TIMEOUT=600 In seconds, set per default by VDCF
export SC_STOP_TIMEOUT=300 In seconds, set per default by VDCF
export SC_PROBE_TIMEOUT=120 to overwrite the 30 seconds Sun Cluster default.

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

3.1.2 LDom Failover

export SC_POSTFIX_LDOM=ldom

For the LDom Resource (_ldom) timeouts may optionally be configured:

export SC_LDOM_START_TIMEOUT=300	In seconds, set per default by VDCF
export SC_LDOM_STOP_TIMEOUT=500	In seconds, set per default by VDCF
export SC_LDOM_PROBE_TIMEOUT=60	In seconds, set per default by VDCF to overwrite the 30 seconds Solaris Cluster default
export SC_LDOM_RETRY_COUNT=0	to overwrite default Retry_count of 2

To enable additional probing features:

Ping/Console probing (see details in Chapter 3.5 Ping/Console Login Probing)

SC_LDOM_PROBE_LEVEL
SC_LDOM_PROBE_USER

Zpool Failmode (see details in Chapter 3.6 Zpool Failmode)

SC_LDOM_RPOOL_FAILMODE
SC_LDOM_ZPOOL_FAILMODE

IPMP-Monitor: (see details in Chapter 3.7 IPMP-Monitor)

/var/opt/jomasoft/vdcf/conf/ipmpmon.cfg

New Feature since VDCF Cluster Version 2.3

SC_LDOM_LOAD_FACTORS	Default=FALSE if set to TRUE the Load Factor Feature is enabled

The required Resources (RAM and CPU) for the Clustered GDom is automatically configured at gdom -c commit, to make sure the Cluster migrates the GDom to the Cluster CDom with enough free resources. This features makes GDom Failover more efficient in Multi Cluster Node Environments.

You can double check the settings on the Cluster using

```
/usr/cluster/bin/clrg show -v -p load_factors
```

3.2 Cluster Resource Definitions

3.2.1 vServer/Zone Failover

VDCF defines the following Cluster Resources in the Cluster for each vServer/Zone.

- ResourceGroup <vserver>_rg
- HAStoragePlus Resource for Datasets <dataset>_hasp
- HAStoragePlus Resource for zonepath <vserver>_hasp_zonepath
  - VDCF Zone Preboot Resource <vserver>_preboot
    - Sun Zone Boot Resource <vserver>_sczbt
      - VDCF Zone Postboot Resource <vserver>_postboot

3.2.2 LDom Failover

VDCF defines the following Cluster Resources in the Cluster for each LDom.

- ResourceGroup <ldom>_rg
- VDCF LDom Preboot Resource <ldom>_preboot
  - LDom Resource <ldom>_ldom
    - VDCF LDom Postboot Resource <ldom>_postboot

3.3 VDCF vServer/Zone Resources

A) VDCF Zone Preboot Resource
This Resource currently adds routes to the Node, which are required by the vServer.

B) VDCF Zone Postboot Resource
This Resource updates the current vServer location in the VDCF Configuration Repository.

3.4 VDCF LDom Resources

A) VDCF LDom Preboot Resource
This Resource updates the current LDom location (CDom) and State in the VDCF Configuration Repository.

B) VDCF LDom Postboot Resource
This Resource updates the current LDom location (CDom) and State in the VDCF Configuration Repository.

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

3.5 VDCF LDom Extended Probing (Ping/Console Login)

The Solaris Cluster LDom Resource only tests if the LDom is running. VDCF offers an additional probing functionality to test if a LDom is really available. To enable this feature you need to define two config vars in customize.cfg:

```
export SC_LDOM_PROBE_USER="vdcfcons:700:700:/export/home/vdcfcons"
export SC_LDOM_PROBE_LEVEL="PC"
```

This configuration is only applied while creating/installing a LDom!

After installation of the LDom the following tests are executed on the CDom:

a) Ping of the LDoms hostname (configurable re-tries, 5 seconds of timeout)

only if ping test has failed:

b) LDom console login test:
   1) if console check is okay (probe okay)
   2) if console in-use (probe okay)
   3) if console on ok prompt (configurable retries, before failing)
   4) if console check fails for other reasons (configurable retries before failing)

If console test fails the probing returns rc 201 and the LDom is switched by the Cluster to another Cluster node.

Variables to customize the probing behavior (with default settings):

- SC_LDOM_PING_RETRY=2	No. of Ping retries before Testing Console Login
- SC_LDOM_CONSOLE_OK_RETRY=2	No. of retries when getting an OK prompt
- SC_LDOM_CONSOLE_FAILED_RETRY=2	Number of retries when getting a login failure
- SC_LDOM_CONSOLE_FAILED_IGNORE=FALSE	Set to TRUE to disable login failure detection

Probing-Results are written to the logfile
`/var/opt/jomasoft/vdcf/log/vdcf_ldom_SC_probe.log` on the CDom.

3.6 LDom Zpool failmode

Zpools have a failmode property to define the OS behavior when a zpool fails. If this property is set to panic the OS will panic on zpool failures. Clustered LDoms will get switched over to another Cluster Node. To enable this failmode feature you may set these settings in customize.cfg:

```
export SC_LDOM_RPOOL_FAILMODE="panic"
export SC_LDOM_ZPOOL_FAILMODE="panic"
```

You may set it on all Zpools or just on the rpool. This property is only set while creating/installing a LDom or a zpool dataset!

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

3.7 LDom IPMP Monitor

IPMP groups in clustered LDom can be monitored to make the Cluster switch over, if one of the IPMP group does fail completely. The IPMP monitor does not have a connection into the Cluster framework, since it works standalone. When a IPMP group fails the IPMP monitor does execute the defined shutdown command, which makes the VDCF LDom Extended Probing detect the LDOM is down and a switch over will be initiated.

For the IPMP monitor there is one central configuration file on the VDCF management server where you define the IPMP groups to be monitored. The IPMP monitor will be automatically enabled during installation of a clustered LDom, if the config file exists and at least one IPMP group in the list is found.

Here is an example of the config file:

```
$ cat /var/opt/jomasoft/vdcf/conf/ipmpmon.cfg

# Configuration file for IPMP Monitoring

# IPMP groups to check
# Use IPMPMON_GROUPS to define groups
IPMPMON_GROUPS="management0,public0"
# Mode to select INCLUDE/EXCLUDE
IPMPMON_MODE="INCLUDE"

# Shutdown command to execute, when finding a failed group
IPMPMON_SHUTDOWN_COMMAND="init 0"
# Optional halt after init 0 and sleep
# IPMPMON_HALT_AFTER_SECONDS=180

# Time (sec) to wait until next check of all groups
IPMPMON_PROBE_TIME="60"

# Amount of additional checks for a failed group, before shutdown
IPMPMON_RETRY_COUNT="6"

# Time (sec) to wait between additional checks
IPMPMON_RETRY_TIME="30"
```

The IPMP Monitor offers 2 Modes

1. IPMPMON_MODE=”INCLUDE”
All IPMP groups are monitored **which match a Name** in the IPMPMON_GROUPS configuration

2. IPMPMON_MODE=”EXCLUDE”
All IPMP groups are monitored **which do not match a Name** in the IPMPMON_GROUPS configuration

VDCF delivers a Template configuration file which can be copied as follows to be activated
`cp /opt/jomasoft/vdcf/conf/ipmpmon.cfg_templ /var/opt/jomasoft/vdcf/conf/ipmpmon.cfg`

A copy of the central configuration file will be copied to the clustered LDOM and is stored here:
`/etc/vdcfbuild/ipmpmon.cfg`

You can find log information of the IPMP monitor in the systems log file /var/adm/messages or the SMF log file /var/svc/log/site-vdcf_ipmpmon:default.log

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

4 VDCF Sun/Solaris Cluster - Administration

After the Sun/Solaris Cluster Software is installed and configured on a Node, run the command

```
node -c update name=<node>
```

This operation updates the VDCF Configuration Repository about the Cluster and the Cluster Nodes.

Check the Node using

```
node -c show name=<node>
```

4.1 vServer Administration

If you create a Dataset or vServer on such configured Nodes, the Cluster is configured accordingly.

You may boot the vServer using `vserver -c boot, reboot or shutdown.`

To migrate the vServer to another Node use the VDCF command:

```
vserver -c migrate name=<vserver> node=<newnode>
```

or use the Sun Cluster command on a Cluster Node:

```
/usr/cluster/bin/clresourcegroup switch -n <newnode> <vserver>_rg
```

4.2 LDom Administration

If you create a LDom (VDCF GDom) on such configured Nodes, the Cluster is configured accordingly.

You may boot the GDom using `gdom -c boot, reboot or shutdown.`

To migrate the Guest Domain (GDom) to another Control Domain in the same Cluster use the VDCF command:

```
gdom -c migrate name=<guest domain> cdom=<newcdom>
```

or use the Solaris Cluster command on a Cluster Node:

```
/usr/cluster/bin/clresourcegroup switch -n <newcdom> <ldom>_rg
```

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

5 Patch Management

This chapter is only relevant if you use Failover Zones.

PatchSet's of Type STANDARD may be installed without any special requirements in the Sun Cluster environment.

Special handling is required for NON_STANDARD patches, which require installation in Single User Mode. Installation of patches in a Sun Cluster environment using Zones requires a complete shutdown of Sun Cluster.

Use the following sequence of commands to patch your cluster efficiently:

1. Make sure the vServer's are distributed evenly on the Cluster Nodes.

Use

`vserver -c show` to verify the vServer locations

and

`vserver -c migrate` to change the vServer locations.

2. Verify the Patch Target. It should be applied to all Cluster Nodes.

```
patchadm -c show_target name=xy
```

3. Shutdown the Sun Cluster / Reboot Node in Non-Cluster-Mode

Execute on one Cluster Node: `/usr/cluster/bin/cluster shutdown`

Boot the Nodes in Non-Cluster-Mode: `OK> boot -x`

4. Patch the Nodes using VDCF

```
patchadm -c install target=XY reboot
```

VDCF will mount all required filesystems for the vServer's on the Nodes and boot the vServer's into Single User Mode. The Patches are then applied on the Nodes in parallel. After Patch Installation the Nodes stay in Single-User State. Once patching is complete manually reboot all Cluster Nodes back into cluster mode.

JomaSoft GmbH
Falkensteinstr. 54a
9000 St. Gallen
Switzerland

6 Open Issues / Restrictions

- A VDCF vServer/Zone doesn't switch if a Network problem occurs.