Reference Design and How-To
For

High Availability 2-Node XenServer Pool Provides Full Functionality with Live Migration Without External Shared Storage for HA-Lizard

Version 1.2.11 – Initial Public Release

IMPORTANT

##
!!! iSCSI-HA and HA-Lizard are free software: you can redistribute it and/or modify
!!! it under the terms of the GNU General Public License as published by
!!! the Free Software Foundation, either version 3 of the License, or
!!! (at your option) any later version.
!!!
!!! iSCSI-HA and HA-Lizard are distributed in the hope that it will be useful,
!!! but WITHOUT ANY WARRANTY; without even the implied warranty of
!!! MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
!!! GNU General Public License for more details.
!!!
!!! You should have received a copy of the GNU General Public License
!!! along with iSCSI-HA and HA-Lizard. If not, see <http://www.gnu.org/licenses/>.
##

Table of Contents

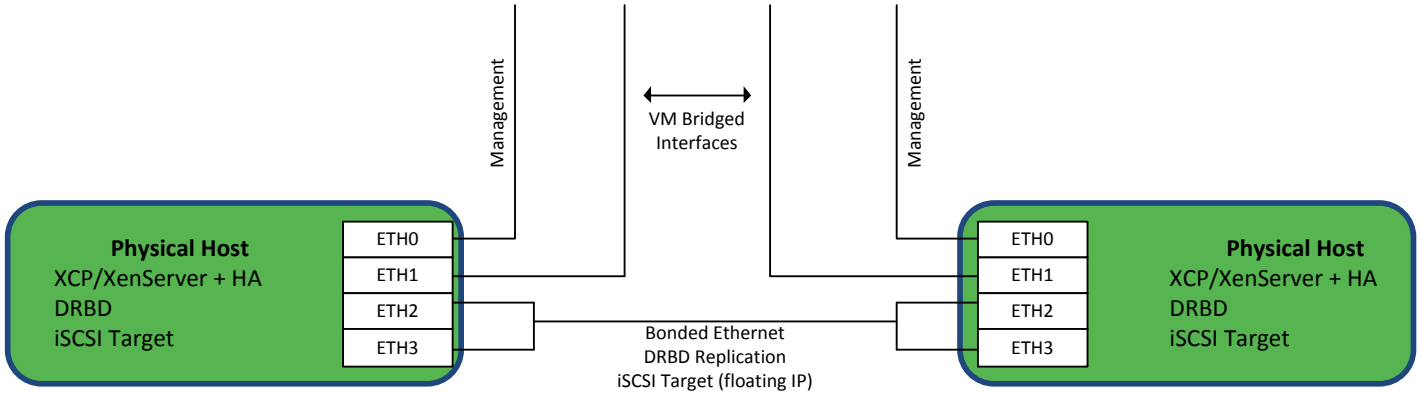
1. iSCSI-HA Add-on for XCP and XenServer

- **Purpose**

iSCSI-HA is an add-on module for two node pools utilizing Xen Cloud Platform (XCP) or XenServer virtualization environments. It is intended to build highly available two-node clusters with local storage without limiting pool advanced functionality such as live migration.

Generally, this is achieved with separate iSCSI/SAN and Xen clusters which require a minimum of four physical servers and a pair of redundant Ethernet switches to reach an adequate level of fault tolerance. This may not be the most efficient use of hardware for small cluster applications. The goal of iSCSI-HA is to provide a simple framework for building compact, highly available pools utilizing XenServer or Xen Cloud Platform with just two physical hosts.

iSCSI-HA requires DRBD for block replication of storage and an iSCSI target framework such as TGT. A sample pool design/diagram is shown below for a highly available two node pool.



In this example, the DRBD and iSCSI interface is provided via direct attachment (no Ethernet switches) on a bonded Ethernet link. This approach greatly eliminates the possibility of a split brain scenario since there are no networking devices interconnecting the hosts in any way. Additionally, utilizing a bonded Ethernet link further eliminates the possibility of communication interruption between the hosts.

The iSCSI-HA add-on does not make any decisions or employ any logic relating to cluster management and the roles of the hosts. It relies on an external HA tool such as open source HA-Lizard which supports HA in a 2 node environment with fast switching of roles. The iSCSI-HA add-on relies on the external HA logic to ensure that a pool Master is always available. Based on this, iSCSI-HA will assign a single shared/floating IP address to the Master and promote DRBD resources to follow the floating IP. The slave host will be in a demoted state

at all times. In the event of a change of pool roles, iSCSI-HA will automatically detect the new roles and promote the new Master as the iSCSI target and demote the former Master to slave/standby mode.

- **Requirements**

- XCP version 1.6 or XenServer version 6.1
- DRBD Version 8.3.x
- iSCSI target – TGT
- Pool HA – Open Source HA-Lizard

iSCSI HA features provided:

- Support for shared/floating virtual IP address to export iSCSI LUNs
- Automatic promoting of DRBD resources on pool Master
- Automatic demoting of DRBD resources on pool Slave
- Management of iSCSI service
- Extensive Logging capabilities to system log file
- Email alerting
- Dynamic iSCSI target selection auto-selects roles
- No changes to existing pool configuration required. All logic is external.
- **Minimal dependencies** – does not compromise pool stability or introduce complex SW packages. Designed to work with the resident packages on a standard XCP/XenServer host with the addition of DRBD and TGT.

Development is well tested and based on Xen Cloud Platform (XCP) version 1.6, XenServer 6.1 and DRBD 8.3

2. Create a 2-Node Highly Available Cluster

- **Assumptions**

Server Hardware

Start with two identical servers with 4 LAN interfaces and two disk partitions. This How-To is based on HP DL-360 servers with HW RAID 1+0 and four disks.

- Disks 1+2 create the first RAID 1+0 array and will be used to install XCP/XenServer
- Disks 3+4 create the second RAID 1+0 array and will be used as the iSCSI backing store

Ethernet Switch

A managed Ethernet switch is used to connect the server management interfaces. The switch management IP must be reachable by the management interfaces of the servers.

Required Software

- iSCSI-HA version 1.2x (or later)
- XenServer 6.1 or Xen Cloud Platform (XCP) 1.6
- Logic to ensure there is always a pool master.
  HA-Lizard 1.41.4 or newer
- DRBD 8.3x for iSCSI replication
- TGT iSCSI target

IMPORTANT – Unless otherwise specified – all steps should be performed on both hosts

- **IP Addresses and Host Names**

The following IP addresses and host names are used in this How-To. Adapt the settings presented to match your environment.

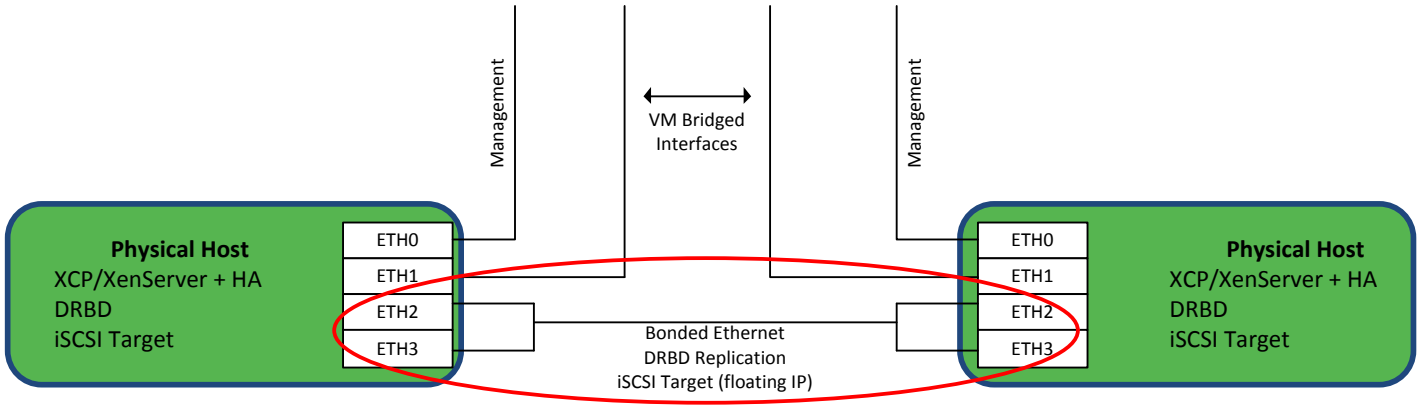
Node 1 hostname	**XCP1**
Node 2 hostname	XCP2
Node 1 XenServer Management IP	192.168.1.241
Node 2 XenServer Management IP	192.168.1.242
Node 1 DRBD/iSCSI IP Address	10.10.10.1
Node 2 DRBD/iSCSI IP Address	10.10.10.2
Shared (floating) iSCSI Address	10.10.10.3
IP Address of Ethernet Switch on Management Network or some other reliable IP that is accessed by traversing the management network	192.168.1.253
iSCSI/DRBD Backing Device	/dev/cciss/c0d1
DRBD Resource Name	iscsi1
DRBD Local Resource	/dev/drbd1

- **Step 1 – Install XCP/XenServer**

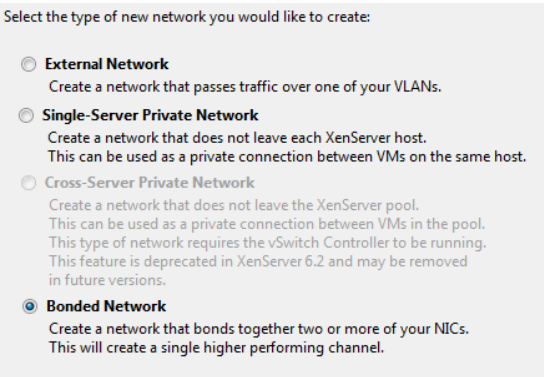
- Install XCP or XenServer on two identical hosts with a minimum four Ethernet interfaces and 2 Disks or Disk partitions. Install XenServer/XCP on the first partition - leave the second disk/partition to be used later as iSCSI storage.
- When installing – select Ethernet 0 as the management interface for each host.
- Connect to one of the hosts with XenCenter and create a new pool with the two hosts.

- **Step 2 – Create a Bonded Interface and Setup Pool Networking**

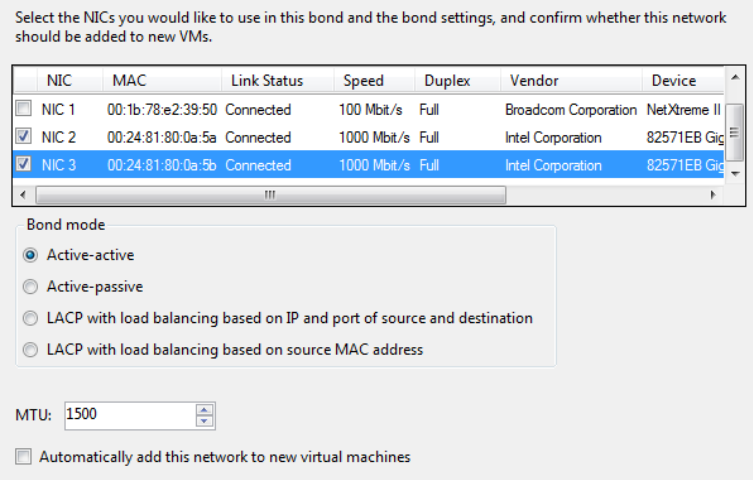
The third and fourth network interfaces (“NIC2 and NIC3”) will be used to create a bonded network as depicted below. This will serve as both the replication link and the iSCSI interface.



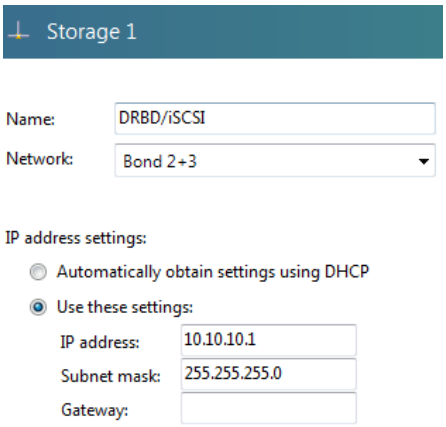
- From XenCenter – Select the “Networking” tab for the pool and “Add Network”
- Select “Bonded Network”



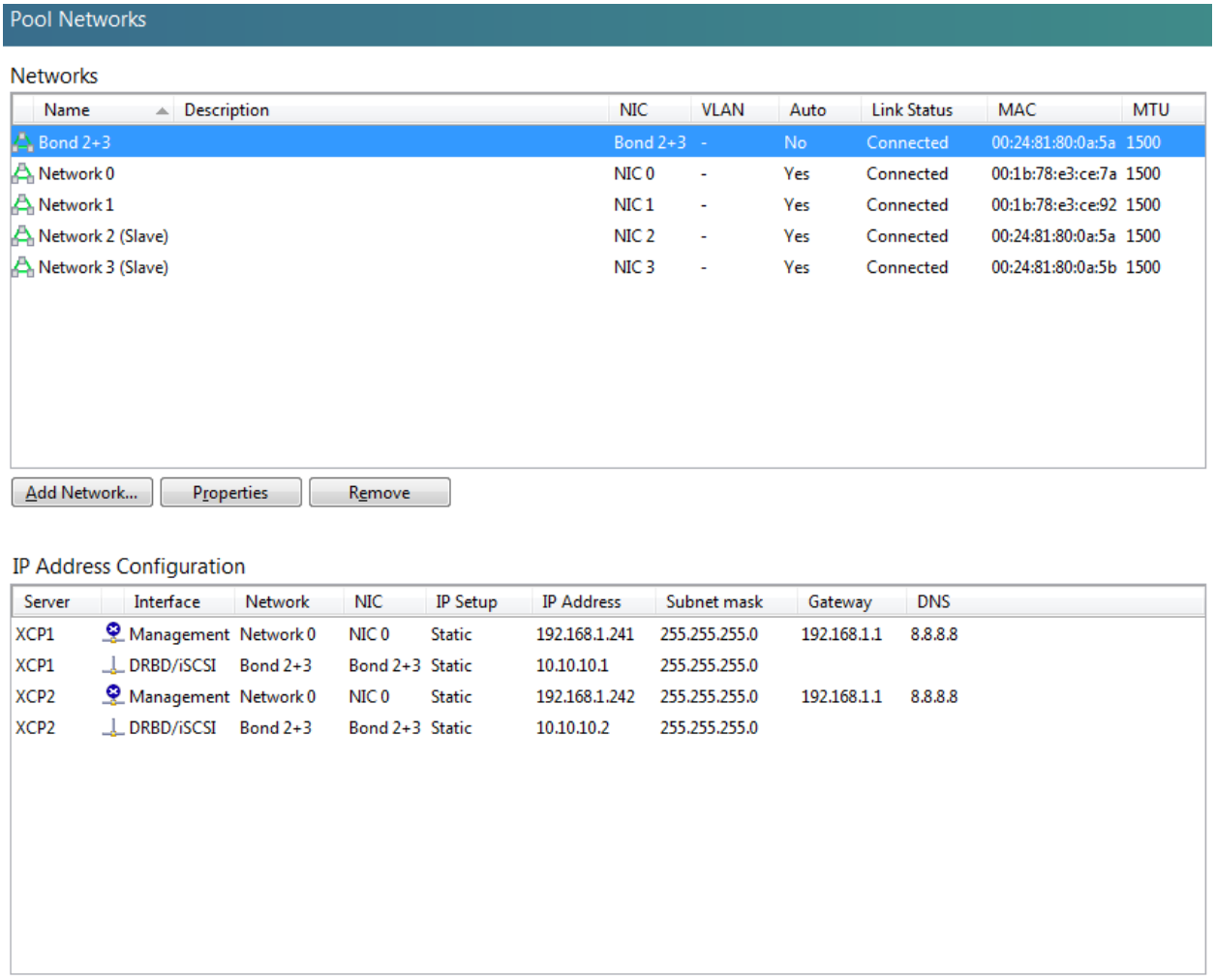
- Select NIC 2 and NIC 3 and then “Finish”



- From within XenCenter – assign an IP address for the bond on each host. In this example we use 10.10.10.1 and 10.10.10.2. Since the replication/iSCSI network is completely closed, it is not necessary to configure a default gateway for the interface.



- The final pool network configuration should look something like this.



- Check each of the network properties and ensure that “Automatically add this network to new virtual machines” is only selected for Network 1 (assuming you will not use the management interface for VM interfaces).

- Update firewall files to allow DRBD and iSCSI network traffic. The following line can be added to the iptables firewall script just above the “REJECT” line.
  “-A RH-Firewall-1-INPUT -s 10.10.10.0/24 -j ACCEPT”

  ```
  vi /etc/sysconfig/iptables
  ```
  Insert “-A RH-Firewall-1-INPUT -s 10.10.10.0/24 -j ACCEPT” save/exit – then restart FW

```
service iptables restart
```

```
# Firewall configuration written by system-config-securitylevel
# Manual customization of this file is not recommended.
*filter
:INPUT ACCEPT [0:0]
:FORWARD ACCEPT [0:0]
:OUTPUT ACCEPT [0:0]
:RH-Firewall-1-INPUT - [0:0]
-A INPUT -j RH-Firewall-1-INPUT
-A FORWARD -j RH-Firewall-1-INPUT
-A RH-Firewall-1-INPUT -i lo -j ACCEPT
-A RH-Firewall-1-INPUT -p icmp --icmp-type any -j ACCEPT
-A RH-Firewall-1-INPUT -p 50 -j ACCEPT
-A RH-Firewall-1-INPUT -p 51 -j ACCEPT
-A RH-Firewall-1-INPUT -p udp --dport 5353 -d 224.0.0.251 -j ACCEPT
-A RH-Firewall-1-INPUT -p udp -m udp --dport 631 -j ACCEPT
-A RH-Firewall-1-INPUT -p tcp -m tcp --dport 631 -j ACCEPT
-A RH-Firewall-1-INPUT -p udp -m udp --dport 67 --in-interface xenapi -j ACCEPT
-A RH-Firewall-1-INPUT -m state --state ESTABLISHED,RELATED -j ACCEPT
-A RH-Firewall-1-INPUT -m state --state NEW -m udp -p udp --dport 694 -j ACCEPT
-A RH-Firewall-1-INPUT -m state --state NEW -m tcp -p tcp --dport 22 -j ACCEPT
-A RH-Firewall-1-INPUT -m state --state NEW -m tcp -p tcp --dport 80 -j ACCEPT
-A RH-Firewall-1-INPUT -m state --state NEW -m tcp -p tcp --dport 443 -j ACCEPT
-A RH-Firewall-1-INPUT -s 10.10.10.0/24 -j ACCEPT
-A RH-Firewall-1-INPUT -j REJECT --reject-with icmp-host-prohibited
COMMIT
```

- **Step 3 – Install DRBD**

- DRBD RPMs are included in the iscsi-ha source package. These can be used only if the kernel version of your XCP/XenServer installation exactly matches the version used to build the RPMs. If using a different kernel version, it will be necessary to build RPMs from source for your kernel. The included RPMs will work with XCP 1.6 or XenServer 6.1.

Installing With Included RPMs

- Extract the iscsi-ha source package in a temporary location
  `tar –zxvf iscsi-ha*.tgz`

- Move into the RPM directory
  `cd /iscsi-ha*/RPM`

- Make a note of the kernel version used to prepare the provided RPMs
  `ls -l`

- Check Your Kernel Version
  `uname –a`

- Install the required RPMs for the matching kernel (adapt below to match your version)
  `rpm -ivh drbd-utils-8.3.15-1.i386.rpm`
  `rpm -ivh drbd-km-2.6.32.43_0.4.1.xs1.6.10.734.170748xen-8.3.15-1.i386.rpm`
  `rpm -ivh drbd-bash-completion-8.3.15-1.i386.rpm`

```
rpm -ivh drbd-heartbeat-8.3.15-1.i386.rpm
rpm -ivh drbd-pacemaker-8.3.15-1.i386.rpm
rpm -ivh drbd-udev-8.3.15-1.i386.rpm
rpm -ivh drbd-xen-8.3.15-1.i386.rpm
rpm -ivh drbd-8.3.15-1.i386.rpm
```

Installing from Source (skip this step if using the included DRBD RPMs)

- Install the required packages for building DRBD from source
  ```
  yum --enablerepo=base install gcc flex rpm-build redhat-rpm-config make
  libxslt –y
  ```

- Get the kernel-xen-devel rpm from XCP (or XenServer) binpkg.iso image and install
  ```
  rpm -ivh kernel-xen-devel-2.6.32.43-0.4.1.xs1.6.10.734.170748.i686.rpm
  ```

- Create RPMs
  ```
  mkdir /drbd/
  cd /drbd/
  wget http://oss.linbit.com/drbd/8.3/drbd-8.3.15-1.tar.gz
  tar zxvf drbd-8.3.15-1.tar.gz
  cd drbd-8.3.15-1
  ./configure --prefix=/usr --localstatedir=/var --sysconfdir=/etc --with-km
  make tgz drbd.spec drbd-km.spec
  cp drbd*.tar.gz `rpm -E %_sourcedir`
  rpmbuild -bb drbd.spec
  rpmbuild -bb drbd-km.spec
  cd /usr/src/redhat/RPMS/i386/
  ```

- Install the required RPMs
  ```
  rpm -ivh drbd-utils-8.3.15-1.i386.rpm
  rpm -ivh drbd-km-2.6.32.43_0.4.1.xs1.6.10.734.170748xen-8.3.15-1.i386.rpm
  rpm -ivh drbd-bash-completion-8.3.15-1.i386.rpm
  rpm -ivh drbd-heartbeat-8.3.15-1.i386.rpm
  rpm -ivh drbd-pacemaker-8.3.15-1.i386.rpm
  rpm -ivh drbd-udev-8.3.15-1.i386.rpm
  rpm -ivh drbd-xen-8.3.15-1.i386.rpm
  rpm -ivh drbd-8.3.15-1.i386.rpm
  ```

- **Step 4 - Install HA-Lizard**

- Copy the source tarball into a temporary location (ex. /tmp/)

- Extract its contents and move into the extracted folder
  tar -zxvf ha-lizard-1.6.4.tgz*
- Move into the “scripts” folder
  cd ha-lizard-1.6.4/scripts*
- Run the installer
  ./install
- Temporarily stop the service
  `service ha-lizard stop -w`

The installer will check if sendmail and mailx packages are installed on the server. These are only required for email alerts. Skip the installation of these packages if email alerting is not required.

The installer will install the default pool parameter set in the XAPI database. This step is only required on a single host.

Once the installer is completed, HA and watchdog services will be started. Although these services are running, HA is disabled for the pool by default. HA can then be enabled via the command line tool <ha-cfg> to enable HA once installation has been completed on all hosts within a pool.

- **Step 5 – Install iSCSI-HA**

- Copy the source tarball into a temporary location (ex. /tmp/)
- Extract its contents and move into the extracted folder
  tar -zxvf iscsi-ha-<version>.tgz
- Move into the “scripts” folder
  cd iscsi-ha-<version>/scripts
- Run the installer
  ./install
- Temporarily stop the service
  `service iscsi-ha stop -w`

The installer will check if sendmail and mailx packages are installed on the server. These are only required for email alerts. Skip the installation of these packages if email alerting is not desired.

Once the installer is completed, iSCSI-HA and watchdog services will be started.

- **Step 6 – Install iSCSI Target**

```
yum --enablerepo=base install scsi-target-utils
```

- **Step 7 – Initialize Packages**

Identify and Export iSCSI Backing Store

- Use fdisk to find the name of the disk partition to be used for the iSCSI backing store. In this example, the device path is "/dev/cciss/c0d1"
  `fdisk -l`

- DRBD will create a resource replicating this block device (/dev/cciss/c0d1). We will use DRBD resource name "/dev/drbd1" in the iSCSI configuration. Modify the iSCSI target configuration file to export "/dev/drbd1". (if using multiple partitions, this can be adapted to suit your environment). Set the scsi_id and scsi_sn to suit your needs. These can be omitted if desired.

  `vi /etc/tgt/targets.conf` (or use your preferred editor)

- Add this section to the configuration file

```
<target iqn.2013-05.com.yourdomain:yourhost>
        backing-store /dev/drbd1
        scsi_id 0000000000
        scsi_sn 0000000001
        lun 10
</target>
```

Configure DRBD

- Backup current DRBD configuration file in case you need it.
  `mv /etc/drbd.conf /etc/drbd.conf.backup`

- Create/Edit new /etc/drbd.conf with the settings below (adapt hostname, disk and IP addresses to your environment)
  `vi /etc/drbd.conf`

- Insert the following configuration parameters

```
global { usage-count no; }
common { syncer { rate 100M; } }
resource iscsi1 {
        protocol C;
        net {
                after-sb-0pri discard-zero-changes;
                after-sb-1pri consensus;
                cram-hmac-alg sha1;
                shared-secret "PUTyourSECREThere";
        }
  on XCP1 {
    device    /dev/drbd1;
    disk      /dev/cciss/c0d1;
    address   10.10.10.1:7789;
    meta-disk internal;
  }
  on XCP2 {
    device    /dev/drbd1;
    disk      /dev/cciss/c0d1;
    address   10.10.10.2:7789;
    meta-disk internal;
  }

}
```

- Initialize the Disks

```
dd if=/dev/zero bs=1M count=1 of=/dev/cciss/c0d1
drbdadm create-md iscsi1
```

- Start the DRBD service

```
service drbd start

drbdadm attach iscsi1
drbdadm syncer iscsi1
drbdadm connect iscsi1
```

** **ON PRIMARY DATA SOURCE ONLY ** :**

```
drbdadm -- --overwrite-data-of-peer primary iscsi1
```

- Edit /etc/lvm/lvm.conf and update filter to look something like this to reject reading LVM headers locally.

```
vi /etc/lvm/lvm.conf
```

Update filter to (restrict local backing device and drbd device – adjust to your environment)->

"filter = ["r|/dev/xvd.|", "r|/dev/VG_Xen.*/*|", "r|/dev/cciss/c0d1|", "r|/dev/drbd*|"]"
Set -> "write_cache_state=0"

Configure iSCSI-HA

- Edit /etc/iscsi-ha/iscsi-ha.conf
  `vi /etc/iscsi-ha/iscsi-ha.conf`

- Make the following configuration changes and save.

```
DRBD_RESOURCES=iscsi1
ISCSI_TARGET_SERVICE=/etc/init.d/tgtd
DRBD_VIRTUAL_IP=10.10.10.3
DRBD_VIRTUAL_MASK=255.255.255.0
DRBD_INTERFACE=xapi0
(if unsure of the DRBD interface try "ip addr show | grep -B 2 10.10.10"
where 10.10.10 are the first 3 octets of the bond IP)
MONITOR_MAX_STARTS=5
MONITOR_DELAY=10
MONITOR_KILLALL=1
MONITOR_SCANRATE=5
ENABLE_LOGGING=1
MAIL_ON=1
MAIL_SUBJECT="SYSTEM ALERT - FROM HOST: $HOSTNAME"
MAIL_FROM="root@localhost"
MAIL_TO='YOUR EMAIL ADDRESS HERE'
```

Configure HA-Lizard

HA-Lizard can be completely configured from the command line. This can be done on either of the two hosts as changes are globally set for all hosts within the pool. The following settings are ideal for use with iSCSI-HA which requires fast detection of host failures and switching of roles. Fencing should be used. The configuration below uses POOL fencing which removes a failed host from the pool, but will not power-off an unresponsive host. ILO or custom fencing can be used if required. Since this design does not allow primary/primary support for DRBD, there is a low likelihood of data corruption should the pool become split. Additionally, the DRBD/iSCSI link is a directly connected Ethernet bond between the two hosts with no switches in between. iSCSI-HA logic utilizes this link to determine which host should act as the iSCSI storage, further reducing the possibility of a split pool. (the below settings assume you are starting with default HA-Lizard settings.. only the following changes from default are required).

```
ha-cfg set FENCE_ENABLED 1
ha-cfg set FENCE_HEURISTICS_IPS 192.168.1.253
ha-cfg set FENCE_MIN_HOSTS 2
ha-cfg set FENCE_QUORUM_REQUIRED 1
```

```
ha-cfg set FENCE_USE_IP_HEURISTICS 1
ha-cfg set MAIL_TO <your alert email address>
ha-cfg set MONITOR_DELAY 15
ha-cfg set MONITOR_MAX_STARTS 20
ha-cfg set XAPI_COUNT 2
ha-cfg set XAPI_DELAY 10
```

The final HA-Lizard configuration should look like the example below. Use “ha-cfg get” to view the configuration.

```
DISABLED_VAPPS=()                          MAIL_FROM="root@localhost"
ENABLE_LOGGING=1                           MAIL_ON=1
FENCE_ACTION=stop                          MAIL_SUBJECT="SYSTEM_ALERT-FROM_HOST:$HOSTNAME"
FENCE_ENABLED=1                            MAIL_TO=yourmail@somedomain.com
FENCE_FILE_LOC=/etc/ha-lizard/fence        MONITOR_DELAY=15
FENCE_HA_ONFAIL=1                          MONITOR_KILLALL=1
FENCE_HEURISTICS_IPS=192.168.1.253         MONITOR_MAX_STARTS=20
FENCE_HOST_FORGET=0                        MONITOR_SCANRATE=10
FENCE_IPADDRESS=                           OP_MODE=2
FENCE_METHOD=POOL                          PROMOTE_SLAVE=1
FENCE_MIN_HOSTS=2                          SLAVE_HA=1
FENCE_PASSWD=                              SLAVE_VM_STAT=0
FENCE_QUORUM_REQUIRED=1                    XAPI_COUNT=2
FENCE_REBOOT_LONE_HOST=0                   XAPI_DELAY=10
FENCE_USE_IP_HEURISTICS=1                  XC_FIELD_NAME='ha-lizard-enabled'
GLOBAL_VM_HA=1                             XE_TIMEOUT=10
```

- **Step 8 – Start Services**

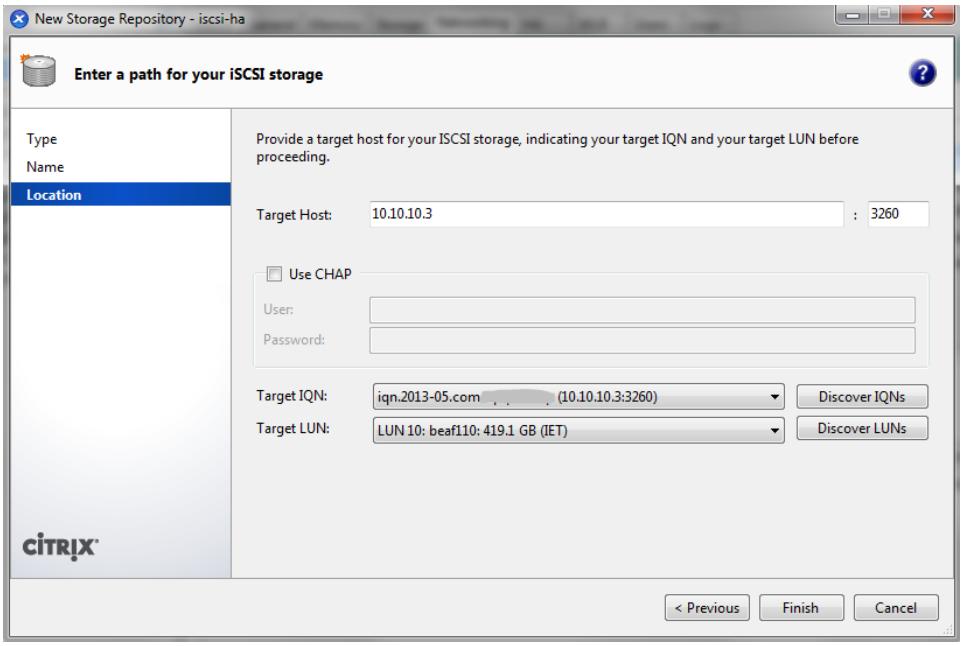
- HA-Lizard should already be running – check with “service ha-lizard-status -w”. Start the service if it is not running
  `service ha-lizard start –w`

- Enable HA from either of the hosts in the pool
  `ha-cfg status` (type “yes” when prompted)

- Start iSCSI-HA
  ```
  chkconfig iscsi-ha on
  service iscsi-ha start –w
  ```

- **Step 9 – Create a new SR**

From within XenCenter, create a new storage repository of type – iSCSI with a target of the shared/floating IP of 10.10.10.3


New Storage Repository - iscsi-ha
Enter a path for your iSCSI storage
Type
Name
Location
Provide a target host for your ISCSI storage, indicating your target IQN and your target LUN before proceeding.
Target Host:
10.10.10.3
:
3260
Use CHAP
User:
Password:
Target IQN:
iqn.2013-05.com (10.10.10.3:3260)
Discover IQNs
Target LUN:
LUN 10: beaf110: 419.1 GB (IET)
Discover LUNs
CITRIX
< Previous
Finish
Cancel

3. Managing the 2-Node Highly Available Cluster

With the installation steps completed, the cluster can now be used to create new VMs. All disk writes will be duplicated across the bonded Ethernet link from the master to the slave, thus each logical volume created by XenServer will be duplicated across the hosts. Read/Write access to the storage will only be on a single IP address which is managed by iSCSI-HA, regardless of which host is the current pool master. The overall design of the compact 2-node cluster is intended to be more resilient than traditional architectures with dual RAID levels and fewer network elements.

- Each host employs HW RAID (if configured to do so)
- Storage is duplicated with SW RAID (DRBD)
- Recovery and failover is simplified in a two node architecture
- Low likelihood of a split pool with mechanisms to prevent the possibility
- Modern server architecture can be very dense allowing for a robust, high capacity virtualization environment with minimal HW components.

- **iscsi-cfg CLI Tool**

A command line tool is provided as part of the iscsi-ha package. The tool can be called with:

```
iscsi-cfg
```

A sample output with command line monitoring arguments is shown below:

```
iSCSI-HA Monitoring Tool: Add-on for HA-Lizard: XenServer/XCP High Availability
Usage: iscsi-cfg <action>

Available actions:
<log>:            Watch iSCSI-HA log file output in real time
<get>:            Lists all iSCSI-HA configuration parameters
<status>:         Displays the iSCSI-HA operational status
```

System Logging

A live view of the system logs generated by iSCSI-HA is available by invoking:

```
iscsi-ha log
```

Viewing Configuration Parameters

A listing of configuation paramters for the local host is available by invoking:

```
iscsi-ga get
```

Viewing iSCSI-HA Status

The iSCSI-HA service is responsible for managing:

- DRBD Running State
- DRBD Resource State (primary/secondary)
- iSCSI Target (TGT) Running State
- Floating IP Address

The status of each of these can be viewed in real time by invoking:

```
iscsi-ha status
```

The information displayed will be relative to the role of the host within the XenServer pool. Below examples show the output from both the pool Master and Slave nodes.

```
********************************************************************************
iSCSI-HA Status: Version: 1.2.11  iscsi-ha (pid  1145 1141) is running...
Last Updated: Sun Jul 14 20:54:36 EDT 2013

HOST ROLE:               MASTER
DRBD ROLE:               iscsi1=Primary
DRBD CONNECTION:         iscsi1 in Connected state
ISCSI TARGET:            tgtd (pid 10093 10091) is running...
VIRTUAL IP:              10.10.10.3 is local
********************************************************************************
Control + C to exit
```

```
********************************************************************************
iSCSI-HA Status: Version: 1.2.11  iscsi-ha (pid  19960 19953) is running...
Last Updated: Sun Jul 14 20:55:31 EDT 2013

HOST ROLE:                SLAVE
VIRTUAL IP:             10.10.10.3 is not local
ISCSI TARGET:           tgtd is stopped
DRBD ROLE:              iscsi1=Secondary
********************************************************************************
```

The status will not be displayed if the iSCSI-HA service is stopped or a system failure is preventing it from running correctly. In this case, the status output will look like the below example.

```
Sun Jul 14 20:59:30 EDT 2013
############################################################
## iSCSI-ha Status cache is more than 33 seconds old
## Tolerance = 15 seconds. Check status
############################################################

Possible causes: 1) iSCSI-HA is actively failing over to standby host
                    this can take up to 60 seconds or more depending on settings

                 2) Service is not running. try 'service iscsi-ha status -w' to check running status
                    If the service is not running, try 'service iscsi-ha start -w' to start the service

Control + C to exit
```

- **Dealing with DRBD Split Brain**

Certain reboot scenarios and unexpected host restarts can cause DRBD to detect split brain which will prevent the DRBD resources from synchronizing. This scenario will likely not impact the operation of the pool since the Master node will still manage DRBD locally and ensure it is in the Primary state. Recovery from split brain should be handled cautiously as storage for one of the nodes will need to be overwritten. The following steps offer a general guideline for recovery.

- First, ensure that the XenServer pool is not split. Although highly unlikely with the 2-node architecture, it is possible that both hosts have entered into the Master role. If this has occurred and VMs on both hosts are not running in duplicate, then it will be necessary to manually merge the known good Logical Volumes from both hosts onto the new Master. Once done, the Master nodes storage can be re-synchronized with the Slave. All HA processes should be stopped during this operation.
- If it is clear which node holds the current data, the following steps should clear up the DRBD split brain.
    - On the node that is to lose its data by synchronizing with the good node (where iscsi1 is the DRBD resource name):
      ```
      drbdadm secondary iscsi1
      drbdadm -- --discard-my-data connect iscsi1
      ```
    - On the host that is the survivor with known good data:
      ```
      drbdadm connect iscsi1
      ```

Miscellaneous

- **Managing Services**

DRBD

It is best to ensure that the DRBD service is automatically started on each host when the system boots. If not already set to do so, this can be easily accomplished with:

```
chkconfig drbd on
```

However, if the DRBD service is found to be not running, iSCSI-HA will start it automatically. iSCSI-HA also acts as a watchdog for DRBD ensuring that the service is always on.

TGTD

The TGTD iSCSI target should be managed by iSCSI-HA as the service should only be running on one of the hosts. For proper operation it is necessary to instruct the host not to automatically start TGTD. This can be done with:

```
chkconfig tgtd off
```

- **Dependencies and Compatibility**

When installing iSCSI-HA onto a default Centos based DomO (XCP or XenServer), all the required tools needed to run iSCSI-HA are resident on the system with the exception of:

- DRBD – version 8.3 required. RPMs are provided in /etc/iscsi-ha/RPM/
- TGT iSCSI Target (can be installed with “yum –enablerepo=base install scsi-target-utils”)

Package is compatible with XCP version 1.6 and XenServer version 6.1. Prior releases may work but have not been tested.

For custom DomO installations, ensure the following tools are available:

xapi and xe toolstack
/bin/cat
/bin/awk
/bin/echo
/sbin/drbdadm
/bin/logger
/sbin/ifconfig
hostname

/bin/mail
/sbin/ip
/sbin/arping

- **Important Considerations**

- iSCSI-HA requires that a node within the 2-node pool ***Always*** is the pool master. If a pool failure results in a situation with no master, the iSCSI target will be unavailable and VMs cannot operate. To ensure that a master is always available HA logic should be employed in the pool.
- In the event that there is no pool master, manual intervention is required to expose the iSCSI target.
- iSCSI-HA only supports 2-node pools. It can be adapted to larger pools with some work.

- **Security and Ports**

- iscsi port 3260 used as the listen port for the iscsi target
- ICMP (ping) is used to check whether the virtual IP is live
- DRBD – ensure that the port numbers specified in drbd.conf are open

- **Support**

- Post a question on the support forum
  http://www.halizard.com/index.php/forum
- Contact the project sponsor for paid support options
  http://www.pulsesupply.com