



High Availability

Using readily available open source products to enhance Insight Server

This document assumes advanced knowledge on the part of the systems administrator with the abilities to setup and configure the additional software used. This includes installation and editing of configuration files needed to provide the HA solution. General terms of use are used to allow the administrator to configure the systems as allowed by the company's policy.

Before you begin, request an additional key for the server and WebClient from Bynari. Each key is tied to the MAC address and will need to be unique to each server.

Definition:

High-Availability means that a service will continue to be available in the event of hardware failure. This can be completed on a small scale using freely available software called DRBD and Heartbeat, and two servers. The two servers will each have a hard drive partition set aside for DRBD. The partition will be silently replicated from one server to the other in real time, much like RAID-1. Services on the primary server will use the DRBD partition for all settings and persistent storage. The secondary server will maintain a copy of the DRBD partition in real time but will not mount it. Heartbeat will run on both servers and trigger a changeover of services if there is a problem. The secondary server will take over the IP address, DRBD partition, and services of the primary server, keeping the services available until the problem can be corrected.

Requirements:

- Two servers running Linux. They need not be the same distribution but does make operation easier.
- DRBD 0.7.x - packaged for most distributions. If you have not customized the kernel which comes with the standard package, you may not need to do anything but install the package and configure it. If you are running a custom kernel, you may have to build DRBD from the source, because the kernel module will have to be built. You will need the kernel source and configuration that matches your running kernel.
- Heartbeat 1.x - also packaged for most distributions.



Resources:

- <http://www.drbd.org>
- [DRBD quick-start guide](#)
- <http://www.linux-ha.org>
- [DRBD in a Heartbeat \(Linux Journal article\)](#)

Quick DRBD Setup:

Install the package for your distribution, or compile from source. The two servers should have the same version installed to ensure compatibility. These instructions work with version 0.7.x, but not with 8.

After installation, ensure that the kernel module loads cleanly and loads automatically at system startup. This is important to the recovery process. If ServerA goes down, ServerB will then become the primary and continue writing data to the DRBD controlled file system. When ServerA eventually comes back up, the DRBD volume will have to be attached in secondary mode and resynchronized to the partition on ServerB. If the kernel module and startup script are properly configured to start at boot time, this step will be automatically performed.

Shrink the file system you want to replicate, but leave the disk partition alone. This gives DRBD space to store its metadata. It only needs 128M per 4TB or so. Make sure that the DRBD space is not bigger than your free space available. If it is larger than your free space, YOU WILL DESTROY DATA.

The EXT2 file system can be shrunk with a tool called `resize2fs`. This does not work with an EXT3 file system, so you have to remove the journal and convert the file system to EXT2 first.

The following assumes /dev/sda3 is mounted on /opt:

```
# /etc/init.d/insightserver stop
# umount /opt
# fdisk -s /dev/sda3
# fsck -n /dev/sda3
# tune2fs -O ^has_journal /dev/sda3
```



Check the file system for errors, do the resize, convert back to EXT3:

```
# e2fsck -f /dev/sda3
# resize2fs /dev/sda3 6000M (was 6300M, shrink only a small amount)
# tune2fs -j /dev/sda3
```

Configure DRBD:

Example DRBD Config:

```
resource insight {
    protocol        C;
    #incon-degr-cmd "echo '!DRBD! pri on incon-degr' | wall ; sleep 60 ";
    startup { wfc-timeout 0; degr-wfc-timeout 120; }
    disk { on-io-error detach; }
    syncer {
        rate        4M;
        #group       0;
    }
    on mail1.example.net {
        device       /dev/DRBD0;
        disk         /dev/sda3;
        address      192.168.30.30:7791;
        meta-disk    internal;
    }
    on mail2.example.net{
        device       /dev/DRBD0;
        disk         /dev/sda3;
        address      192.168.30.31:7791;
        meta-disk    internal;
    }
}
```

Load the DRBD kernel module and attach to the resource:

```
# modprobe DRBD
# DRBDadm attach {resourcename}
```

To verify that the load command was successful, type below:

```
# DRBDadm state {resourcename}
```

If successful, you get the message:
"Secondary/Unknown"

Both servers will be configured the exact same way with the same size partition and file system. Each server will show as a secondary device because it is not connected to the other system.



Select the machine containing your Insight server installation as your primary device. At the moment, it is configured as an inconsistent secondary, so it needs to be set as a primary:

```
# DRBDadm -- --do-what-I-say primary {resourcename}
```

Then connect the two together by starting the connection on each machine:

```
# DRBDadm connect {resourcename}
```

You can monitor the sync progress by using this command:

```
# watch cat /proc/DRBD
```

You can now mount the file system on the primary machine and run Insight Server. Everything that is stored on Insight server will be replicated, on the fly, to the secondary.

Quick Heartbeat setup:

Heartbeat configuration is hard to explain, but simple to show as an example. There are some files that need explanation:

Go to directory, `/etc/ha.d/`

Below are examples of files, *ha.cf* and *haresources*

ha.cf has many options covered in the Heartbeat documentation.

Example of *ha.cf*:

```
debugfile /var/log/ha-debug
logfile    /var/log/ha-log
logfacility local0
keepalive 2
deadtime 30
warntime 10
initdead 120
udpport    694
bcast eth0
auto_failback off
node mail1.example.net
node mail2.example.net
respawn hacluster /usr/lib/heartbeat/ipfail
```



In the example above, pay particular interest to the "node" option, where the two participating server nodes are named.

Example of haresources file:

The configuration information in this file **must be identical on both servers** and may have one or more lines like this:

```
mail1.example.net IPaddr::192.168.30.32/24 DRBD insightserver
```

The above line specifies the primary server, same IP address to be shared by both systems, and a list of services to start and stop.

In directory /etc/ha.d/resource.d/:

DRBD and insightserver, are customized script files to be used to customize the events that take place when a server changes status. Note that *DRBD and insightserver* are mentioned above in *haresources*. They are listed in the order, in which they need to be executed at startup, and are executed in reverse when the server goes down.

Example DRBD script:

```
#!/bin/bash

case "$1" in
  start)
    echo Starting DRBD resources...
    DRBDadm primary all
    DRBDadm connect all
    mount /dev/DRBD0 /opt
    ;;
  stop)
    echo Stopping DRBD resources...
    umount /opt
    DRBDadm secondary all
    ;;
  status)
    exit 3
    ;;
esac
```

Example insightserver script:

```
#!/bin/bash

# Wrapper script for /etc/init.d/insightserver, to give LSB-compliant
# responses back to Heartbeat 2. If using Heartbeat 1.x, this is
# unnecessary.
```



```
case $1 in
  start)
    /etc/init.d/insightserver start
    exit 0
  ;;
  stop)
    /etc/init.d/insightserver stop
    exit 0
  ;;
  status)
    if ! -d /opt/insight? ; then
      echo OK - Insight Server is not running because the DRBD
      volume is apparently not mounted.
      exit 3
    else
      if /etc/init.d/insightserver status ; then
        exit 0
      else
        exit 3
      fi
    fi
  ;;
esac
```

To test the configuration, unplug the network cable, and this will simulate a failure. Once the test is complete, read the below instructions:

The Failover worked, now how do I get the primary to take over again?

At this stage, the primary server has gone down and the secondary server has successfully taken over all services. Congratulations! You now have a fault tolerant mail server.

The system is configured as "Active/Passive", meaning it will remain running on the secondary server until the configuration is manually switched back to the primary. This allows the DRBD volume on the downed primary server to be connected and resynchronized with the volume running on the active secondary server. To make the switch, sign on the root shell on the primary system and:

Verify that the kernel module is loaded:

```
# lsmod | grep DRBD
DRBD      147796  3
```

Verify that the DRBD partition is attached and that the resync has completed:



```
# DRBDadm state insight
"Secondary/Primary"
```

This output means that the local partition is attached as Secondary and is syncing to the remote Primary.

```
# cat /proc/DRBD
version: 0.7.23 (api:79/proto:74)
SVN Revision: 2686 build by root@mail1.example.net, 2007-01-29 14:00:04
 0: cs:Connected st:Secondary/Primary ld:Consistent
    ns:78316 nr:0 dw:158852 dr:197646 al:25 bm:198 lo:0 pe:0 ua:0 ap:0
```

Look for the word, "Consistent". If you see "Inconsistent" and a progress indicator, the resync hasn't finished yet. Wait until it completes before going any further. When it has finished, you need only restart Heartbeat on the secondary server. It will release the resources and notify the primary that it is safe to take over. On the secondary server:

```
# /etc/init.d/heartbeat restart
```

Watch the logs on each machine, and you'll see services come down on one machine and up on the other.

YOU HAVE COMPLETED THE CONFIGURATION!