Active/Active Storage Cluster
the Zero-Single-Point-of-Failure setup

February 12, 2013

Janusz Bak
CTO

Introduction to Open-E DSS V7

AGENDA

- Open-E DSS V7 – Storage Operating System
- The build content
- GUI and Console
- NAS, iSCSI & FC Unified
- LVM
- User and Tasks Snapshots
- Data Replication
- Volume Replication
- High Available Cluster

Auto Failover iSCSI

- Setup Chart: http://blog.open-e.com/ping-node-explained/
- Volume Replication
- Replication Tasks settings
- iSCSI Targets Settings
- Auxiliary paths, Ping Nodes
- Failover and Failback

Auto Failover (iSCSI and NFS)

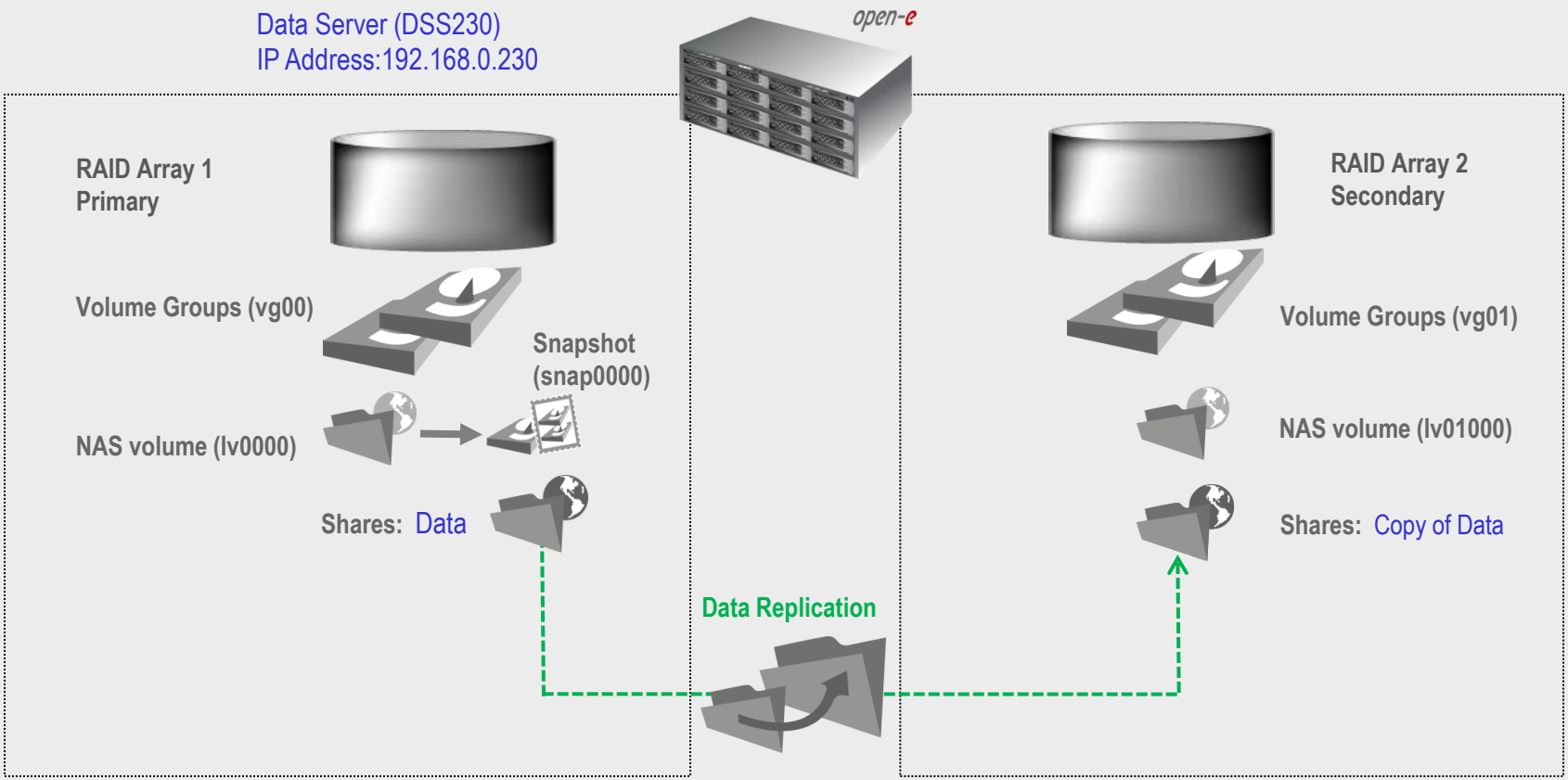
AGENDA

- Setup Chart: http://www.open-e.com/library/how-to-resources/
- How to configure snapshots for data replication
- How to configure data replication in LAN, WAN and within single system
- Data Replication deployment with one-to-one, one-to-many and many-to-one

CONFIGURE HARDWARE

Hardware Requirements

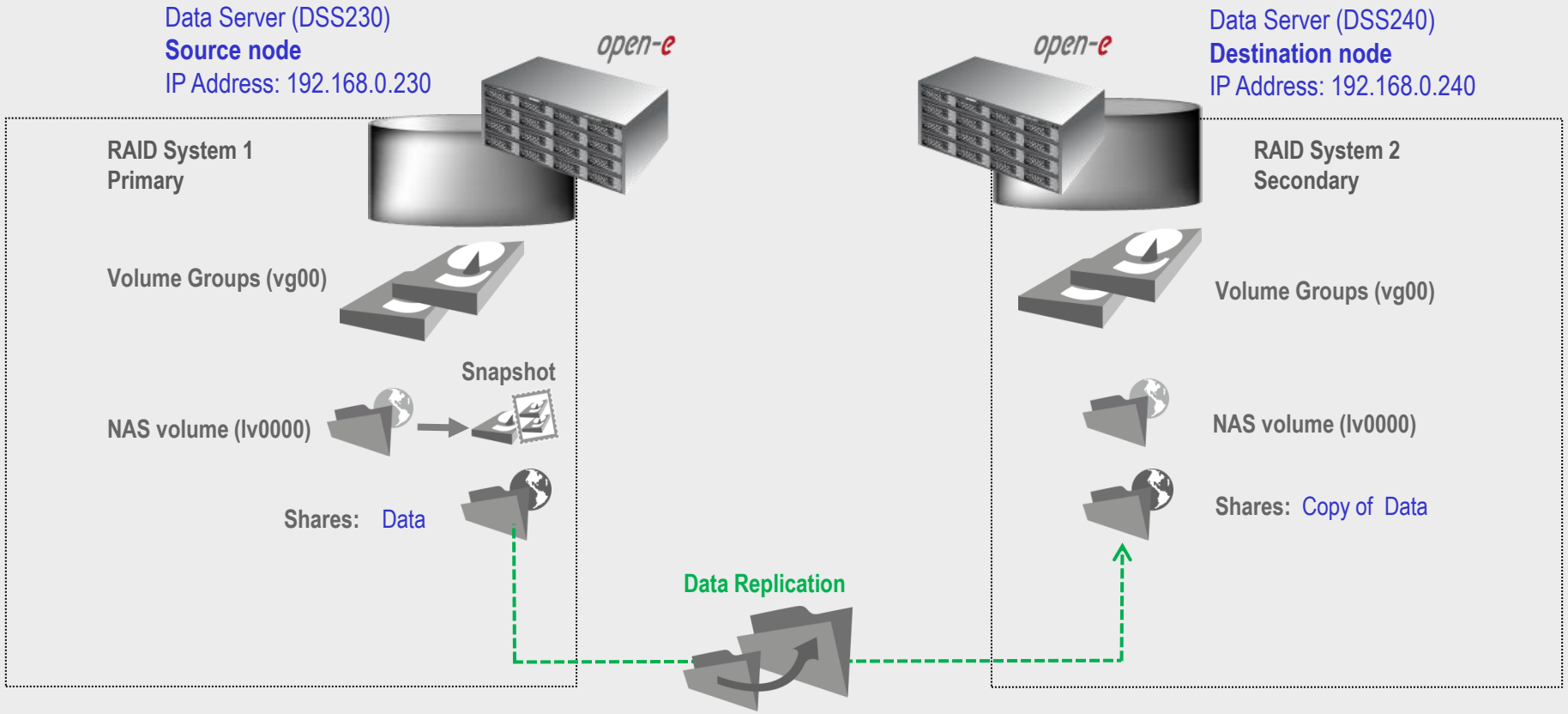
To run the data replication on Open-E DSS V7, a minimum of two RAID arrays are required on one system. Logical volumes working on RAID Array 1 must have snapshots created and enabled. An example configuration is shown below:



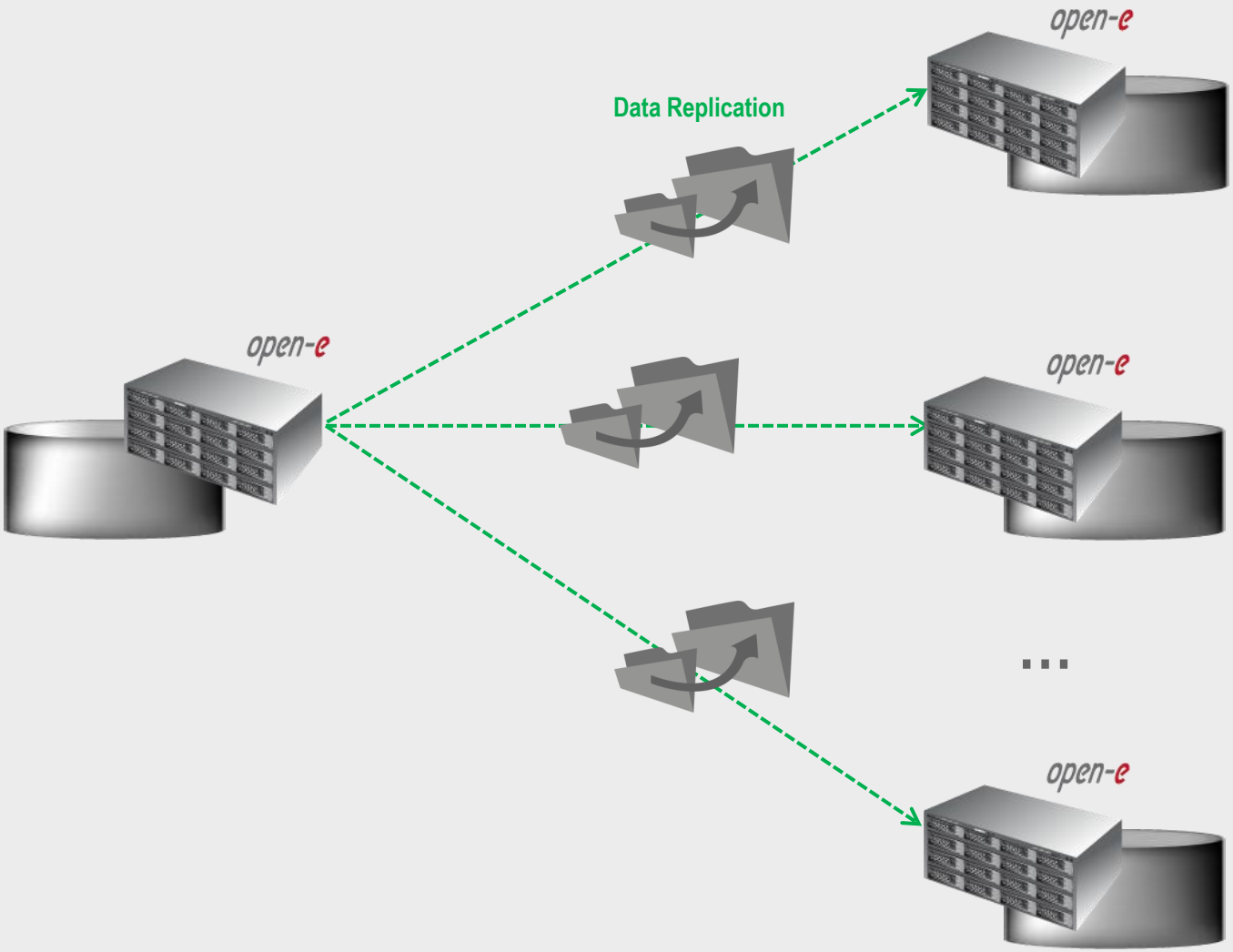
CONFIGURE HARDWARE

Hardware Requirements

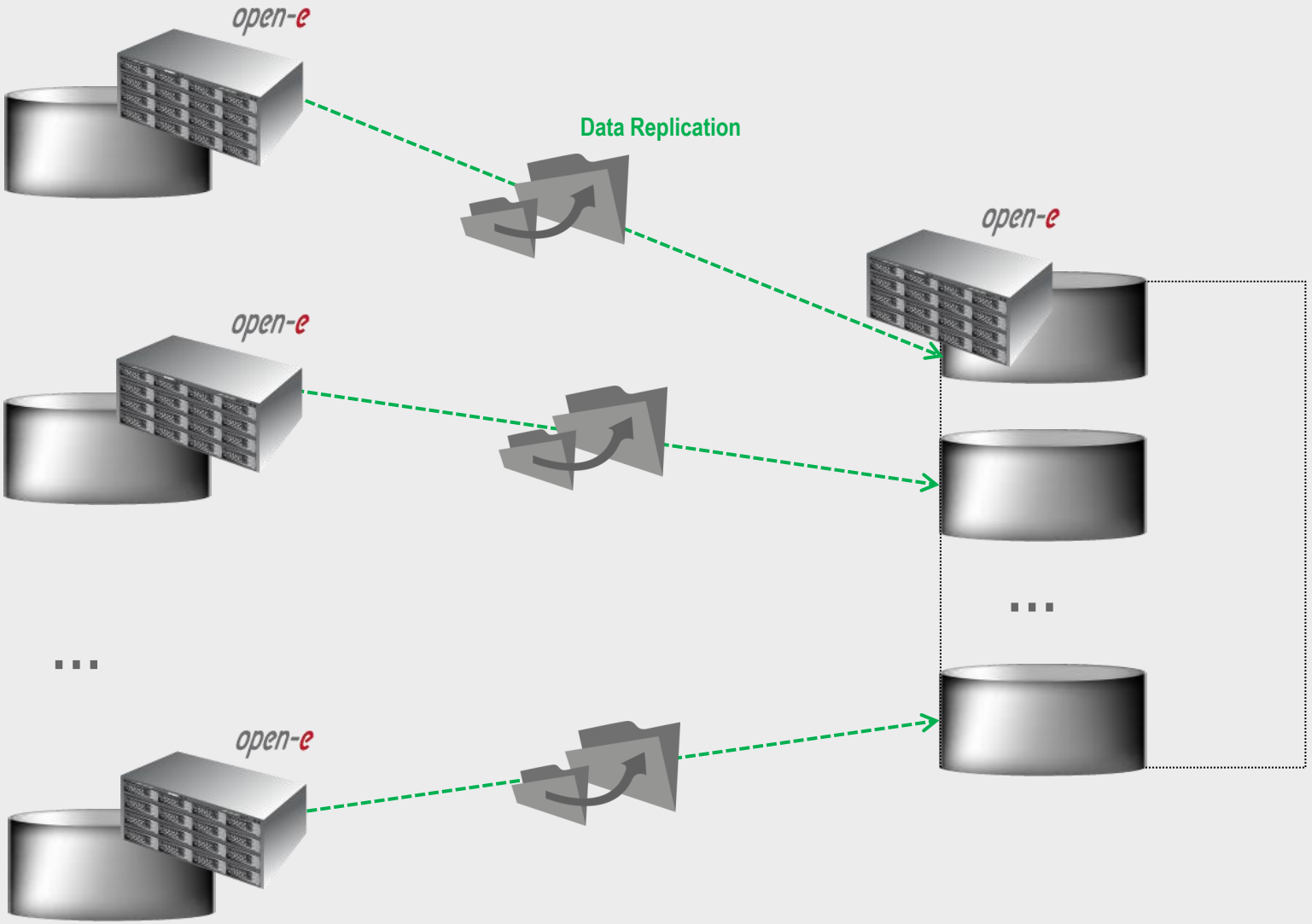
To run the data replication on Open-E DSS V7 over LAN, a minimum of two systems are required. Logical volumes working on source node must have snapshots created and enabled. Both servers are working in the Local Area Network. An example configuration is shown below:



DATA REPLICATION: ONE-TO-MANY



DATA REPLICATION: MANY-TO-ONE



MPIO with VMware and Microsoft

Setup Chart: http://blog.open-e.com/ping-node-explained/

Step-by-step: http://www.open-e.com/library/how-to-resources

Bonding vs. MPIO

Setup Chart: http://blog.open-e.com/bonding-versus-mpio-explained
http://blog.open-e.com/ping-node-explained/

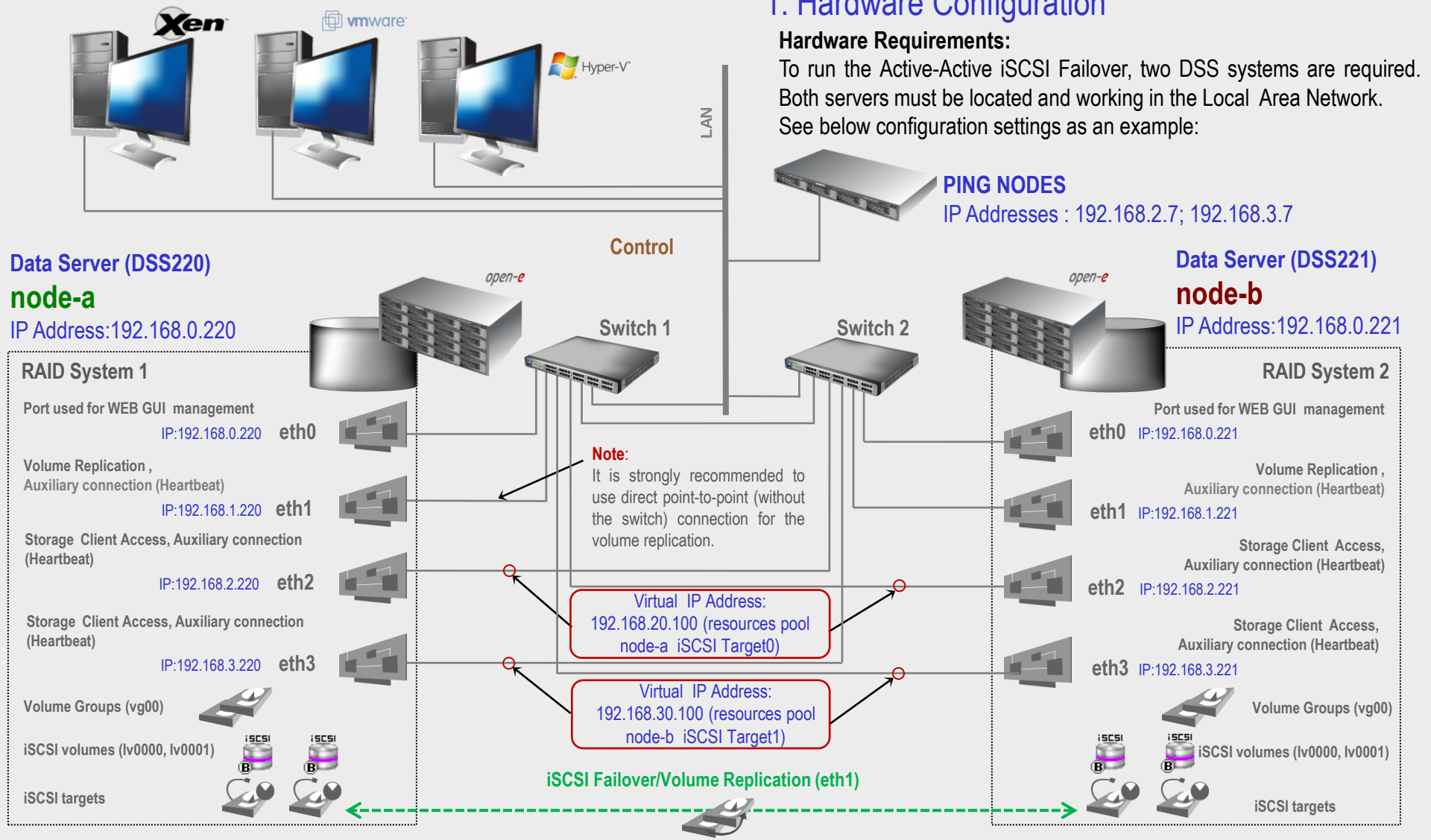
Bonding types, LACP
Multipath: target and initiator

Open-E DSS V7 Active-Active iSCSI Failover

1. Hardware Configuration

Hardware Requirements:

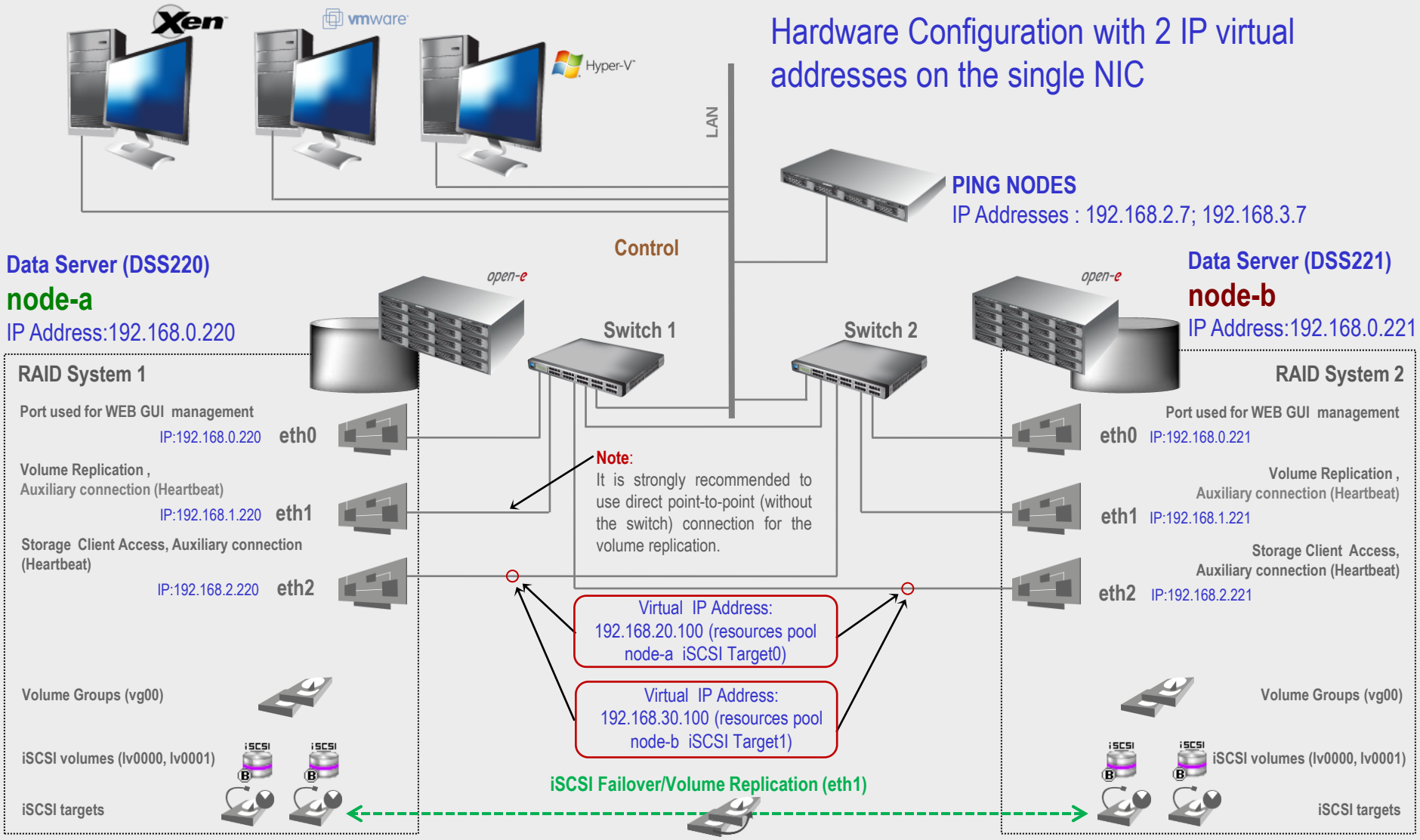
To run the Active-Active iSCSI Failover, two DSS systems are required. Both servers must be located and working in the Local Area Network. See below configuration settings as an example:



NOTE:

To prevent switching loops, it's recommended to use RSTP (802.1w) or STP (802.1d) protocol on network switches used to build A-A Failover network topology.

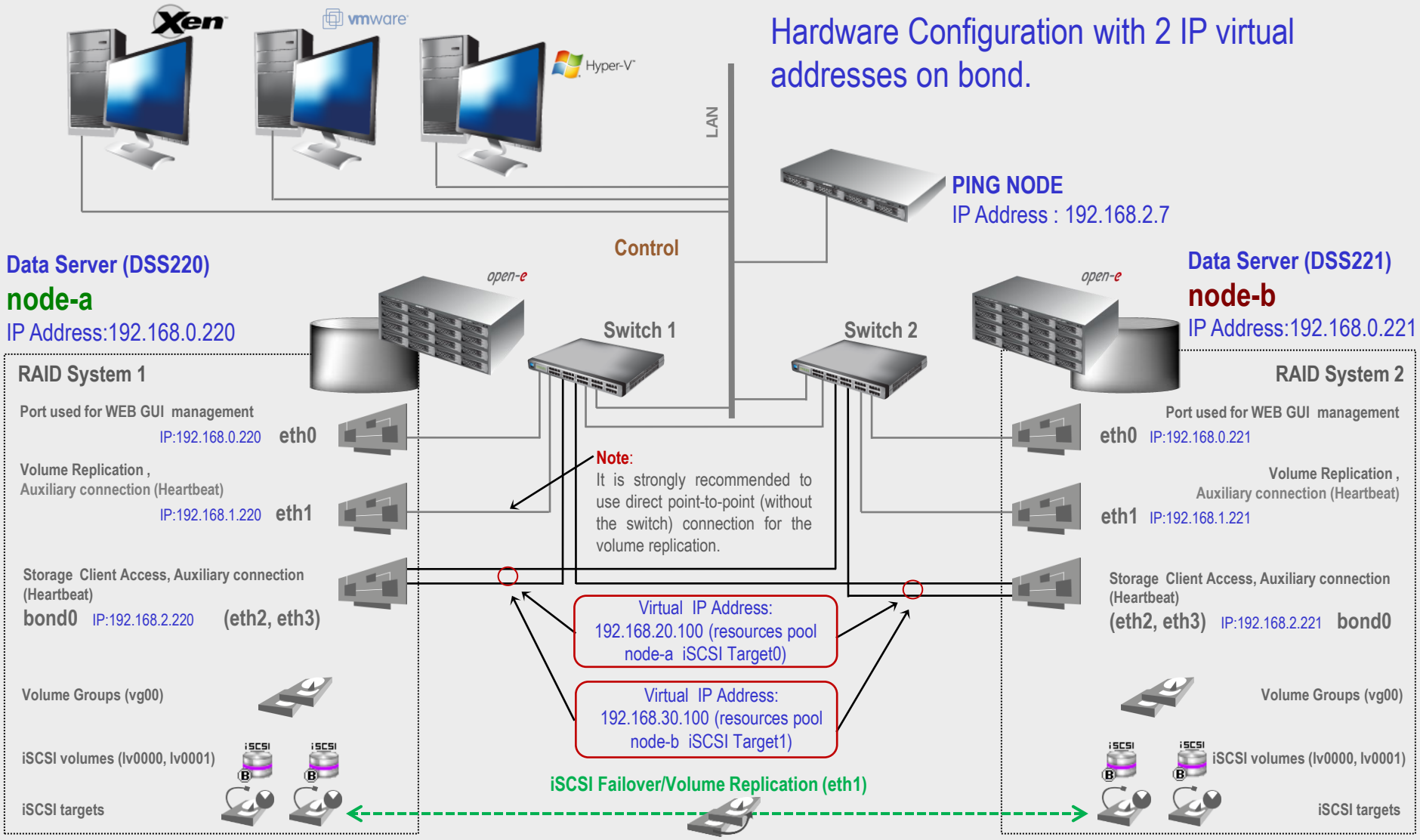
Open-E DSS V7 Active-Active iSCSI Failover



NOTE:

To prevent switching loops, it's recommended to use RSTP (802.1w) or STP (802.1d) protocol on network switches used to build A-A Failover network topology.

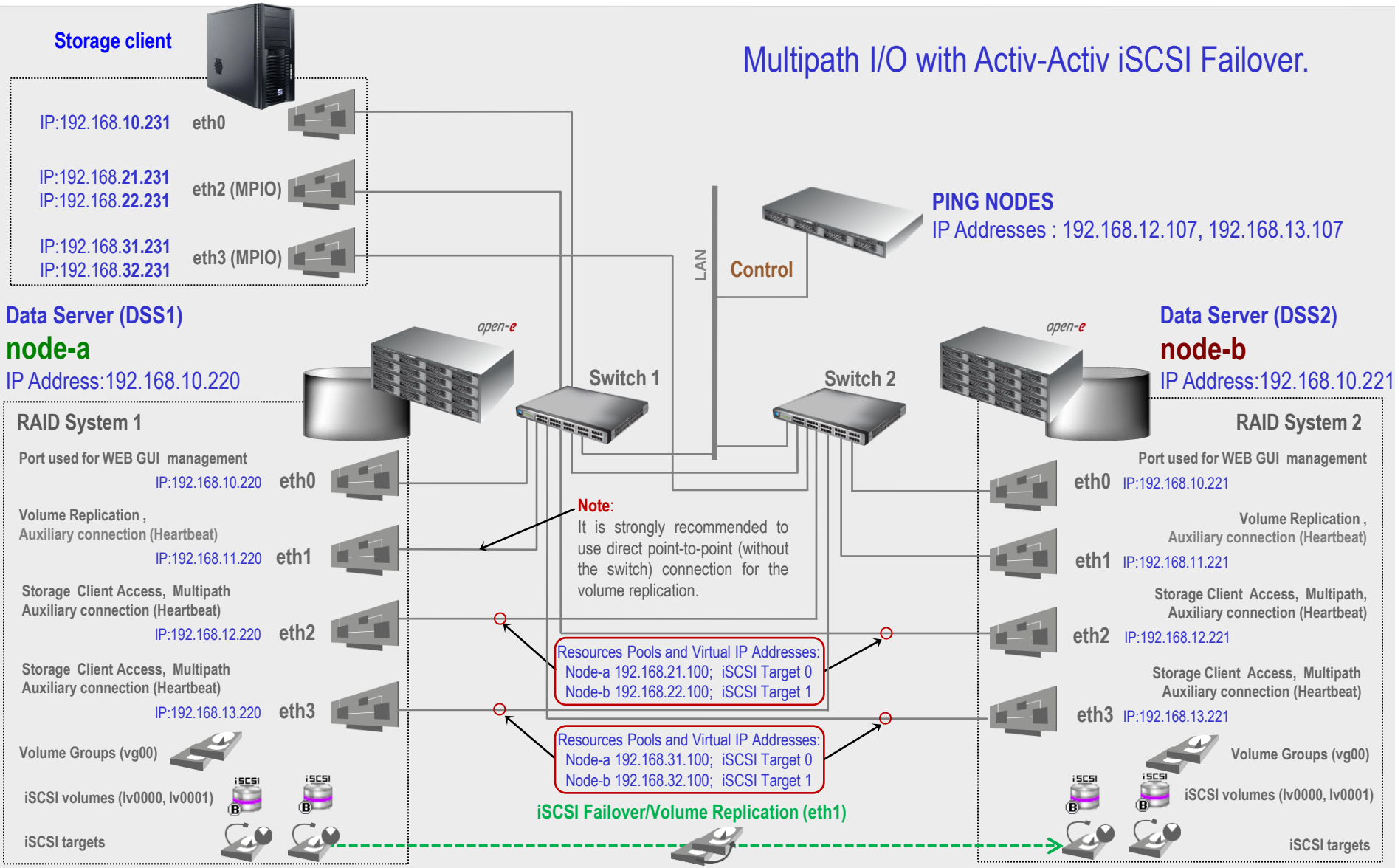
Open-E DSS V7 Active-Active iSCSI Failover



NOTE:

To prevent switching loops, it's recommended to use RSTP (802.1w) or STP (802.1d) protocol on network switches used to build A-A Failover network topology.

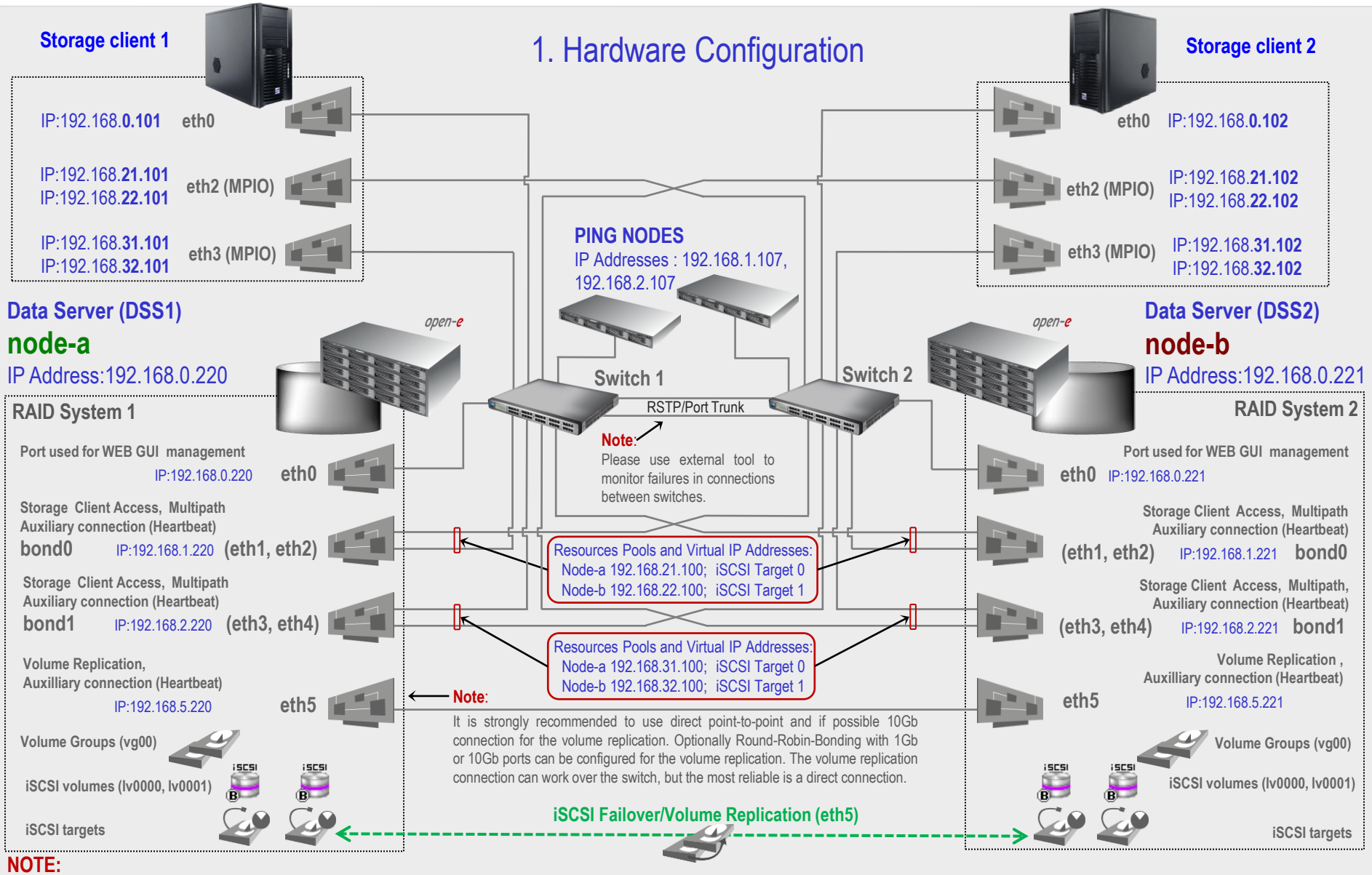
Open-E DSS V7 Active-Active iSCSI Failover



NOTE:

To prevent switching loops, it's recommended to use RSTP (802.1w) or STP (802.1d) protocol on network switches used to build A-A Failover network topology.

Open-E DSS V7 with Multipath Active-Active iSCSI Failover



NOTE:

To prevent switching loops, it's recommended to use RSTP (802.1w) or Port Trunking on network switches used to build A-A Failover network topology.

Open-E DSS V7 with Multipath Active-Active iSCSI Failover

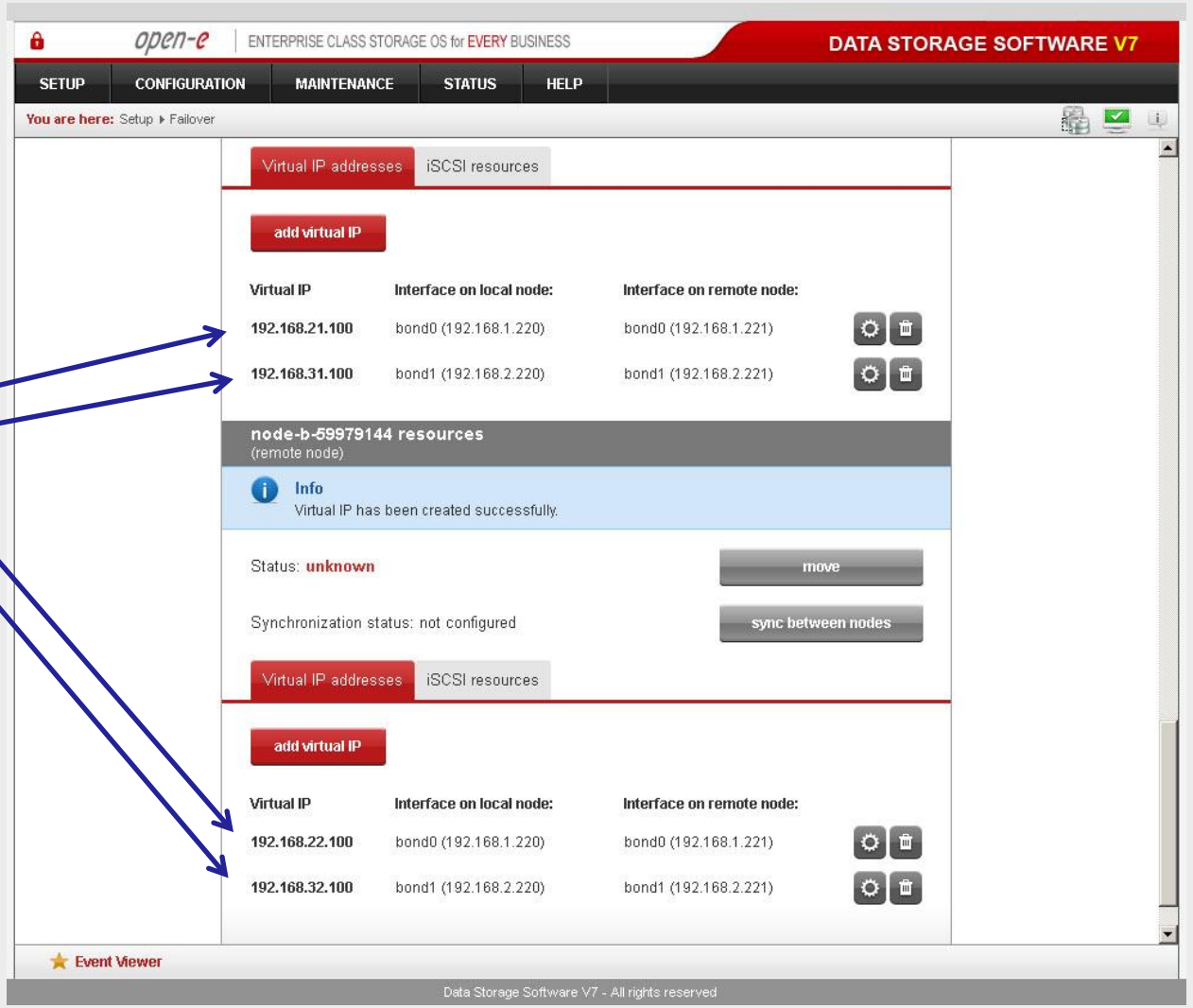
Data Server (DSS1)

node-a

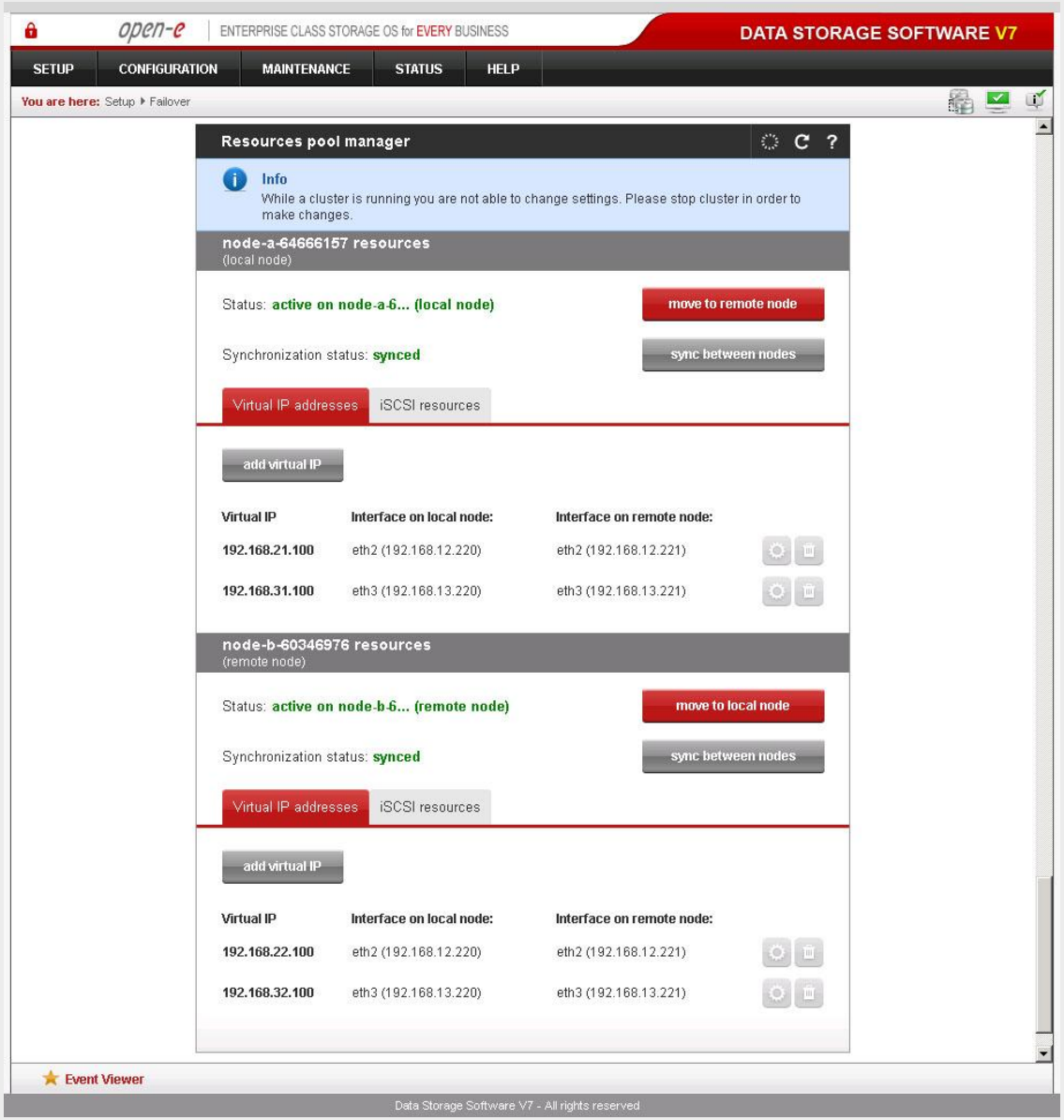
IP Address:192.168.0.220

6. Configure Failover

Now you have 4 **Virtual IP** addresses configured on two interfaces.



Open-E DSS V7 Active-Active iSCSI Failover



Storage Appliances – Best Practices

open-e

AGENDA

- How to check hardware health
- How to check hardware RAID
- What to do in case of DEGRADED RAID (as a result of a drive failure)
- How to use iSCSI and FC Volumes
- Static vs. Dynamic Target Recovery with Vmware
- How to maintain Multi-Storage Units System
- How to install Open-E Data Storage Software (DSS)
- Relation between Volume size and RAM
- Hardware quality: System Temperature, System Chassis, Vibrations, Raid Port Labeling

KEEP YOUR DATA REDUNDANT

NOTE: Never start production without any data backup plan.

- **For mission-critical (very expensive) data:**
  - Mandatory: perform periodical data backups (incremental archiving) with DSS V7 built-in backup, or with a third-party backup appliance (e.g. Backup Exec) via a built-in agent.
  - Optional: additionally run data replication periodically, with frequency according to application needs.

- **For non-critical data:**
  - Mandatory: run at least data replication periodically, with frequency according to application needs.
  - Optional: additionally perform periodical data backups, with frequency according to application needs.

NOTE: RAID arrays are NOT to be considered as a data backup. RAID does not provide real data redundancy, it merely prevents a loss of data availability in the case of a drive failure. Never use RAID 0 in a production system - instead, use redundant RAID levels, such as 1, 5, 6 or 10. A single drive failure within a RAID 0 array will inherently result in data loss, which will require restoration of the lost data from a backup.

Best Practices

CHECK HARDWARE HEALTH

1. Before starting the system in full production, create a 10GB iSCSI volume in File-I/O mode (with initialization). On a regular RAID array, a 10GB volume should be created and initialized in approximately 110 seconds.
   Optional: create a 100GB iSCSI File-I/O volume (with initialization and medium speed). On a regular RAID array, a 100GB volume should be created and initialized in approximately 15 minutes.
2. Delete the created 10GB (and/or 100GB) volume.
3. Create a new iSCSI File-I/O volume (with initialization) using ALL free space.
4. After the iSCSI volume spanning all available space has been initialized, reboot the system.
5. After the reboot, check the event viewer for errors - the event viewer must be free of any errors.
6. If event viewer is clean, delete the test iSCSI volume.
7. Create new volumes as required, and start production.
   Optional: Once system temperature becomes stable, perform measurements to check whether it remains within a range allowed for your system components.

CHECK HARDWARE RAID

IMPORTANT NOTE: Make sure that the hard disk trays are labeled correctly with their properport numbers. Errors in labeling of the disk trays may result in pulling out wrong drive. Be aware that the port count may start with 0 on some RAID controllers, and with 1 on others.

1. **Before starting the production system, create and initialize the RAID array; also, configure email notifications in the RAID controller GUI (and/or in the DSS GUI).**
2. **Create a 100GB iSCSI volume in File-I/O mode (with initialization), and during the iSCSI initialization process REMOVE a hard disk from the RAID array.**
3. **Check the event viewer for errors. There must be entries informing about the array now being degraded; however, there must be NO reported I/O errors, as the degraded state must be transparent to the OS.**
4. **Now, re-insert the drive. It is likely that partial logical volume data, as well as partial RAID metadata will still reside on the drive; in most cases, this residual partial data must be deleted before a rebuild can be started.**

DEGRADED RAID (AS A RESULT OF A DRIVE FAILURE)

1. **Before starting the production system, create and initialize the RAID array; also, configure**
2. **Run a full data backup.**
3. **Verify the backed-up data for consistency, and verify whether the data restore mechanism works.**
4. **Identify the problem source, i.e. find the erroneous hard disk. If possible, shut down the server, and make sure the serial number of the hard disk matches that reported by the RAID controller.**
5. **Replace the hard disk identified as bad with a new, unused one. If the replacement hard drive had already been used within another RAID array, make sure that any residue RAID metadata on it has been deleted via the original RAID controller.**
6. **Start RAID the rebuild.**

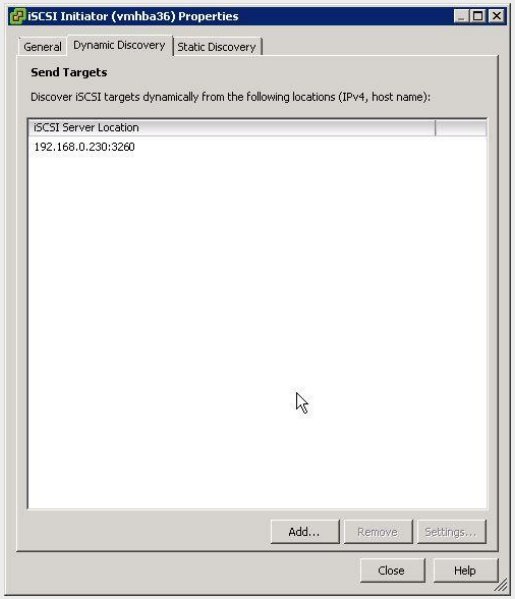
IMPORTANT NOTE: Never use hot-spare hard disks, as a hot-spare hard disk will jump in automatically, and the array will start rebuilding immediately. A RAID rebuild is a heavy-duty task, and the probability of another drive failure during this process is higher than usual; thus, it is a best practice to start the rebuild in step 5 rather than immediately.

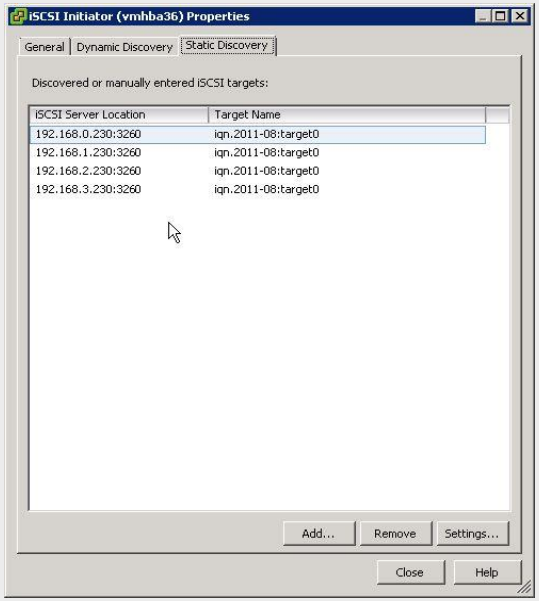
iSCSI AND FC VOLUMES

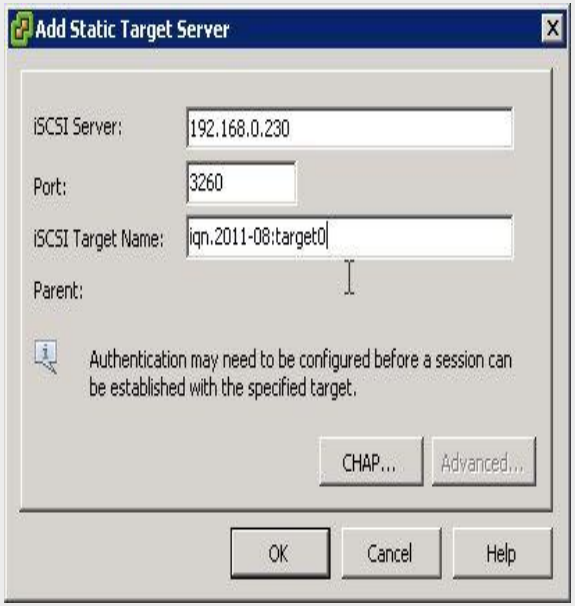
1. **iSCSI and FC volumes emulate a raw SCSI drive; in the case they will be partitioned and formatted with a regular (non-cluster) file system like NTFS, EXT3, XFS, etc., they must be used by a host exclusively. The I/O of an iSCSI target is block-based (as opposed to file-based), which means that changes made by one person will not be seen by another person working on the same target/volume.**
2. **An iSCSI/FC volume usually represents a slice of a RAID disk array, often allocated one per client. iSCSI/FC imposes no rules or restrictions on multiple computers sharing an individual volume. It leaves shared access to a single underlying file system as a task for the operating system.**

WARNING: If two or more hosts using a non-cluster file system write to the same target/volume, the file system will crash, which will more than likely result in data loss. In order to make more concurrent connections to the same target practically possible, utilization of a special SAN file system like GFS, OCSF etc. is required.

STATIC VS. DANAMIC DISCOVERY IN VMWARE







Best Practices

Multi-Storage Units System

1. Create a separate volume group for every external storage unit. This is a good practice, as such a configuration proves to be more reliable: in case one of the units has a problem, the others can continue to work.
2. If the application requires the addition of external storage units into the same volume group, make sure the connections are very reliable, as ALL of the storage will become unavailable if only one of the units is missing.

How To Install Open-E Data Storage Software V7

NOTE: Never start production without any data backup plan.

- **With hardware RAID:**

  It is recommended to create a 2GB-sized logical unit for DSS V7, and a second logical unit spanning all of the remaining space for the user data.

NOTE: RAID controllers do not support creating more than one logical unit from within the controller BIOS. For example, the HP Smart Array needs to be booted from a Smart Array Management CD in order to be able to create a RAID array with multiple logical units. Please refer to your RAID controller user manual.

- **With software RAID:**

  It is required to install DSS V7 on a separate boot media. Please use boot media like a HDD, a SATA-DOM, or an IDE-DOM. Please DO NOT use USB-DOM for production.

VOLUME SIZE AND RAM

1. **Avoid creating volumes larger than 64TB;**
2. **It is recommended to install an amount of RAM calculated in the following way:**
   - (Size of RAM to install in GB) = (Size of the largest volume in TB) / 2
   - For example: if the size of the largest volume is 32TB, the recommended amount of RAM is 16GB.

System Temperature

Generally, high temperatures will shorten the lifespan of hard disks, thus try to use hard disks with an operation temperature as low as possible.

1. BeTIP: In order to estimate the temperature levels the hard drives within your system may reach during daily operation, you can connect drives into SATA ports on the motherboard, create a NAS or iSCSI volume in DSS, and then run any random pattern test.
2. In the case your hard disks are connected to the mainboard SATA controller, you can monitor the temperature of these hard disks from within the DSS GUI (STATUS -> S.M.A.R.T.). For this functionality to be available, make sure that S.M.A.R.T has been enabled in the BIOS setup of your system's motherboard, as well as in the DSS console (press Ctrl-Alt-W in the console to enter Hardware Configuration, then navigate Functionality Options -> enable S.M.A.R.T).
3. If you are using a RAID controller, please refer to its user manual to find information on how to monitor hard disk temperatures. Some RAID controllers do support such functionality, while others don't.

System Chassis, Vibrations, Raid Port Labeling

- In order to avoid unexpected vibrations, always try to use hard disks with the same RPM spindle speed. Should you be forced to mix 7200, 10,000, and/or 15,000 RPM drives, please use drives with anti-vibration firmware, and make sure that the hard disks' and chassis' vendors declare support for utilization in such an environment (which is not to be assumed without prior verification).

- In order to avoid unexpected vibrations, always try to use hard disks with the same RPM spindle speed. Should you be forced to mix 7200, 10,000, and/or 15,000 RPM drives, please use drives with anti-vibration firmware, and make sure that the hard disks' and chassis' vendors declare support for utilization in such an environment (which is not to be assumed without prior verification).

Thank you!