Step-by-Step Guide

Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings (supported since version up28)

Software Version: JovianDSS ver. 1.00 up28

Last update: August 2020

Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings

The aim of this document is to demonstrate an example setup of an Advanced Metro High Availability Cluster with 2 Rings. The 2 Rings option is available as of version 1.0 up28.

The Advanced Metro High Availability Cluster is using an Ethernet link for Disk Exports instead of SAS. It can work with JBOD-less hardware so that disks are present in both cluster nodes and are mirrored via an Ethernet path.

The Advanced Metro HA Cluster uses the same Ring-Ping design for Cluster Management as the Open-E JovianDSS Standard HA Cluster. But it enables an additional functionality – the "Remote disks mirroring paths for Cluster over Ethernet" – in order to configure a special Ethernet link for disk exports.

In this example (**setup on page 5**), 2 NICs are used for iSCSI Targets only.

The bond is preferred for NFS and SMB shares but for iSCSI path redundancy MPIO is a better choice. The Open-E JovianDSS works as a Unified Storage Appliance, providing NAS and SAN (iSCSI ,NFS, SMB). All services requiring path redundancy will need to configure 2 bonds. **Please refer to other examples shown on pages 46-52.**

In this document, two pools are created with 4 (2-local + 2-remote) disks in every mirror group. In case of the other node, reboots or failure, the mirror groups are still redundant with 2 disks in every mirror.

The 4-way mirror provides limited storage efficiency of only 25%. For increased storage efficiency, a hardware RAID controller and just a simple mirror over 2 disk units with a RAID array behind can be used. Open-E JovianDSS includes built-in all RAID tools and drivers for Broadcom (LSI) and Microsemi (Adaptec). In case of Areca, a driver is included as well but the Web-GUI has to be accessed via the controller's ETH-port.

Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings

The software version 1.0 up28 supports NEW options for cluster configuration. Here is the comparison summary:

The up27 and older versions:

- Cluster bind-ring and ping nodes must be on the same Active-Backup bond and MUST go via network switch.
- If node has lost all ping nodes signals, all pools on this node will be exported. The export is done because the node is most probably not reachable and must export pools so other node can import and failover will be possible.
- If node has lost all ping nodes and ring-heartbeat, the node will be rebooted immediately. The reason is similar as in the case of all pings lost but in the case of no ping nodes and no ring signal, it is assumed the node is fully isolated and reboot will be safer for other node to failover.

Note: It is not possible to use different bonding than Active-Backup for the cluster bind-ring.

The up28 and newer versions:

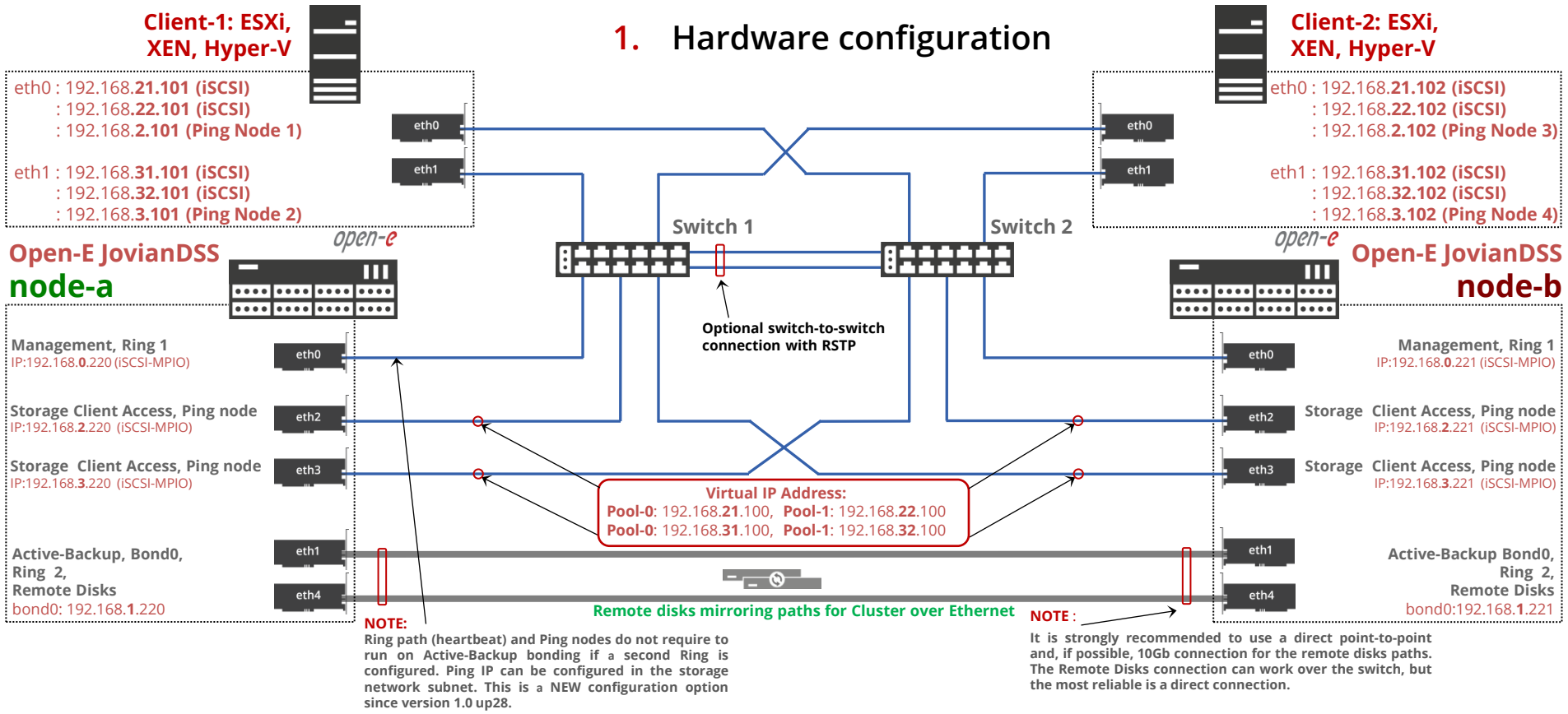
- Everything what was valid and working in the up27 will also work in the up28 and newer.
- Up to 2 rings can be configured and ping nodes to be configured on the same path of the storage paths and not on the rings network.
- There is NO obligatory active-backup bond for the cluster bind-ring. Both rings can be configured on nonbonded NIC but optionally can work on an active-backup bond as well.
- The second ring can be configured via a mirror path if Advanced Metro Cluster is used but in such case the mirror path can work on a single NIC or an active-backup bond. If a round-robin bond is configured for the mirror path, it is not passible to configure the second ring on it. In such case, the second ring can be configured on extra point-to-point single path or an active-backup bond path.
- If a node has lost all ping node signals, all pools on this node will be exported. The export is done because the node is most probably not reachable and must export pools, so the other node can import and failover will be possible (**this works the same as in the up27**).
- If a node has lost all ping nodes and a ring-heartbeat, the node will be rebooted immediately. The reason is similar as in the case of all pings lost, but in the case of no ping nodes and no ring signal, it is assumed the node is fully isolated and a reboot will be safer for other node to perform a failover (**this works same as in the up27**).

Note: MS Hyper-V cluster as a storage client via iSCSI use Persistent-Reservations synchronization which works on bind-ring only. It does NOT work via a second ring. This is why for Hyper-V cluster it is still obligatory to use an Active-Backup bond for the bind-ring path.

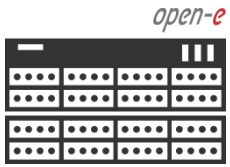
To set up an Advanced Metro HA Cluster, perform the following steps:

1. Hardware configuration
2. Network configuration
    - 2.1. Create a mirroring path bond
    - 2.2. Select a default gateway
    - 2.3. Network configuration on node-b
3. Time and date settings
4. Nodes binding
5. Adding rings
6. Ping Nodes
7. Mirroring path
8. Start the cluster service
9. Create a new Pool
    - 9.1. Add a data group
    - 9.2. Add a write log
    - 9.3. Add a read cache
10. Enter a virtual IP
11. System monitoring setup
12. Failover test

1. Hardware configuration

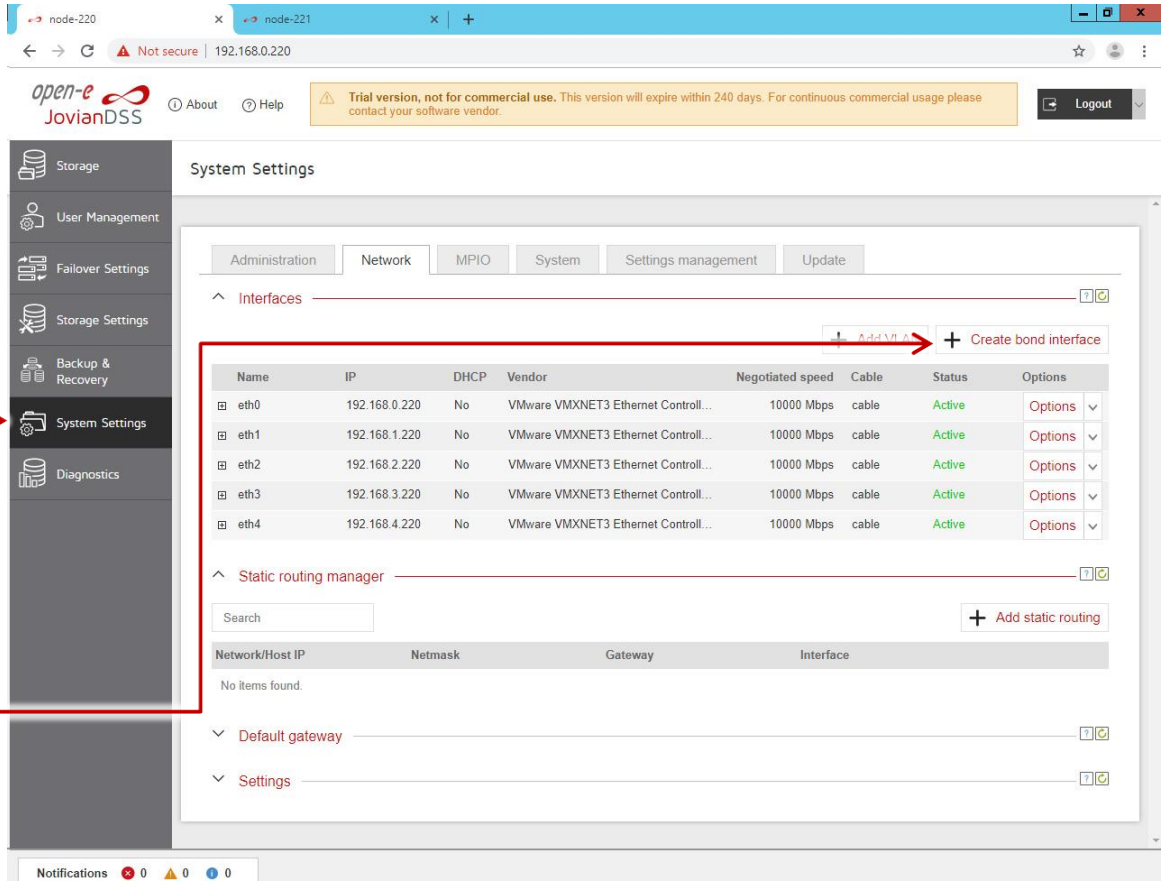


2. Network configuration

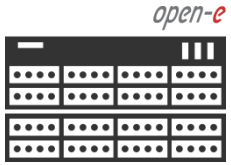


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Select **System Settings** from the main menu and next select the **Network** tab. Click the **Create bond interface** button.
This will be the bond for the mirror path.

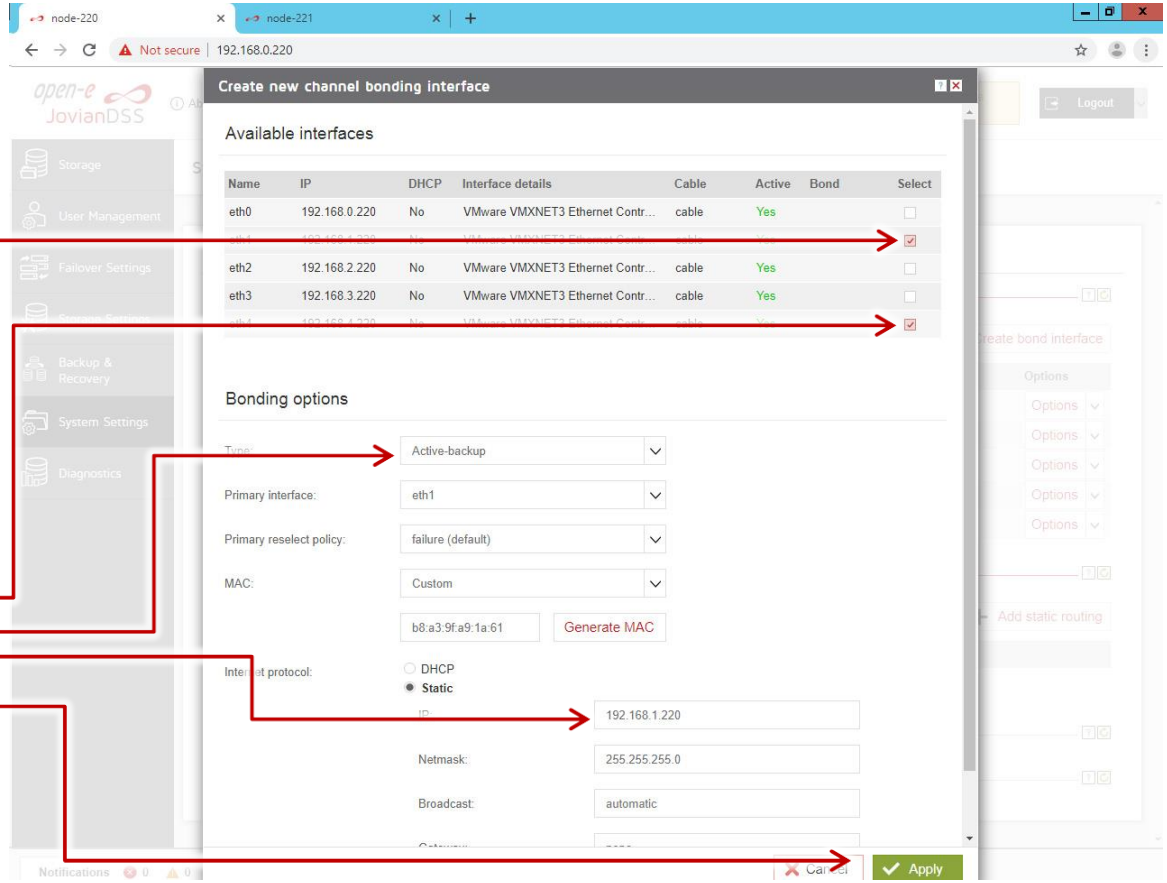


2.1. Network configuration. Create a mirroring path bond

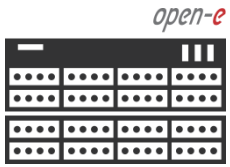


Open-E JovianDSS: **node-a**
IP Address: 192.168.**0**.220

Enter all required details of the bond and click the **Apply** button.



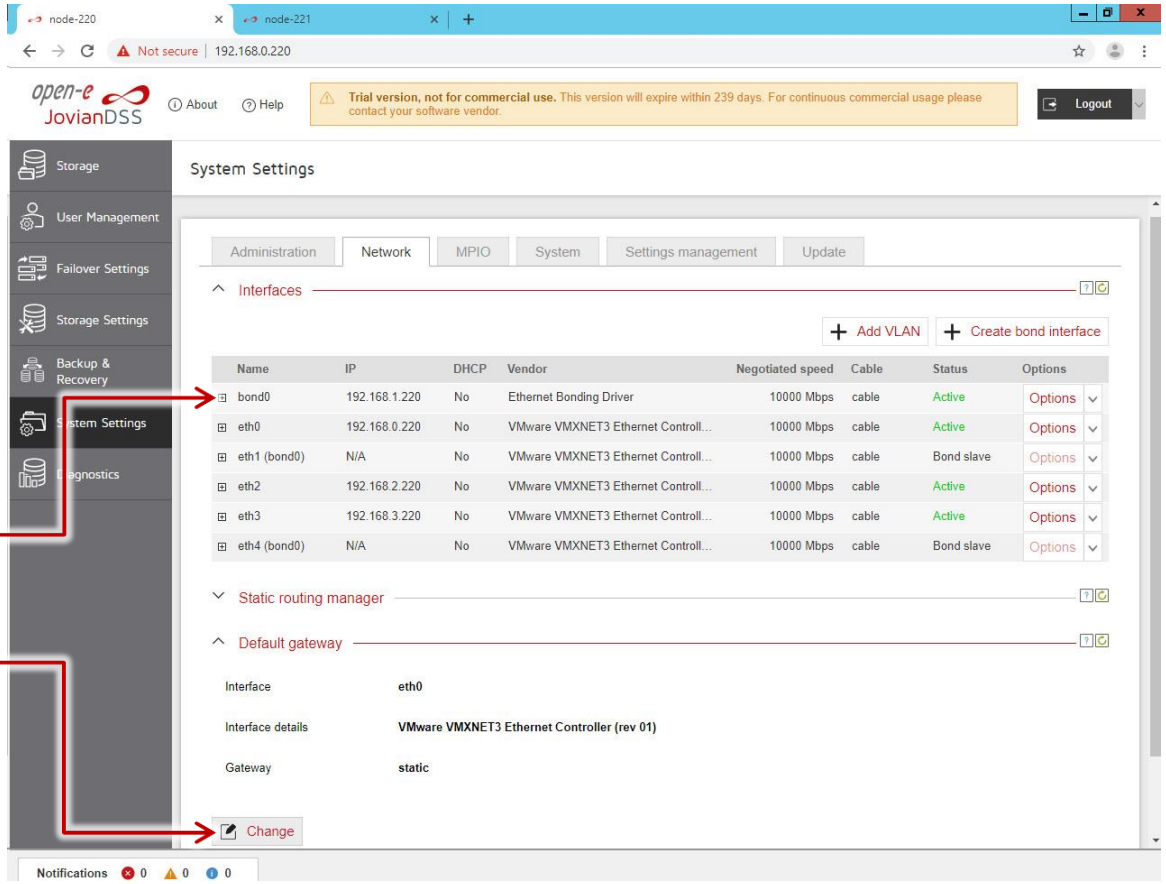
2.2. Network configuration. Select a default gateway



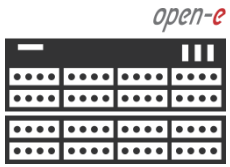
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

A bond is created properly. Overview is shown in the **Interfaces** field.

Next, in the **Default gateway** field, click the **Change** button.

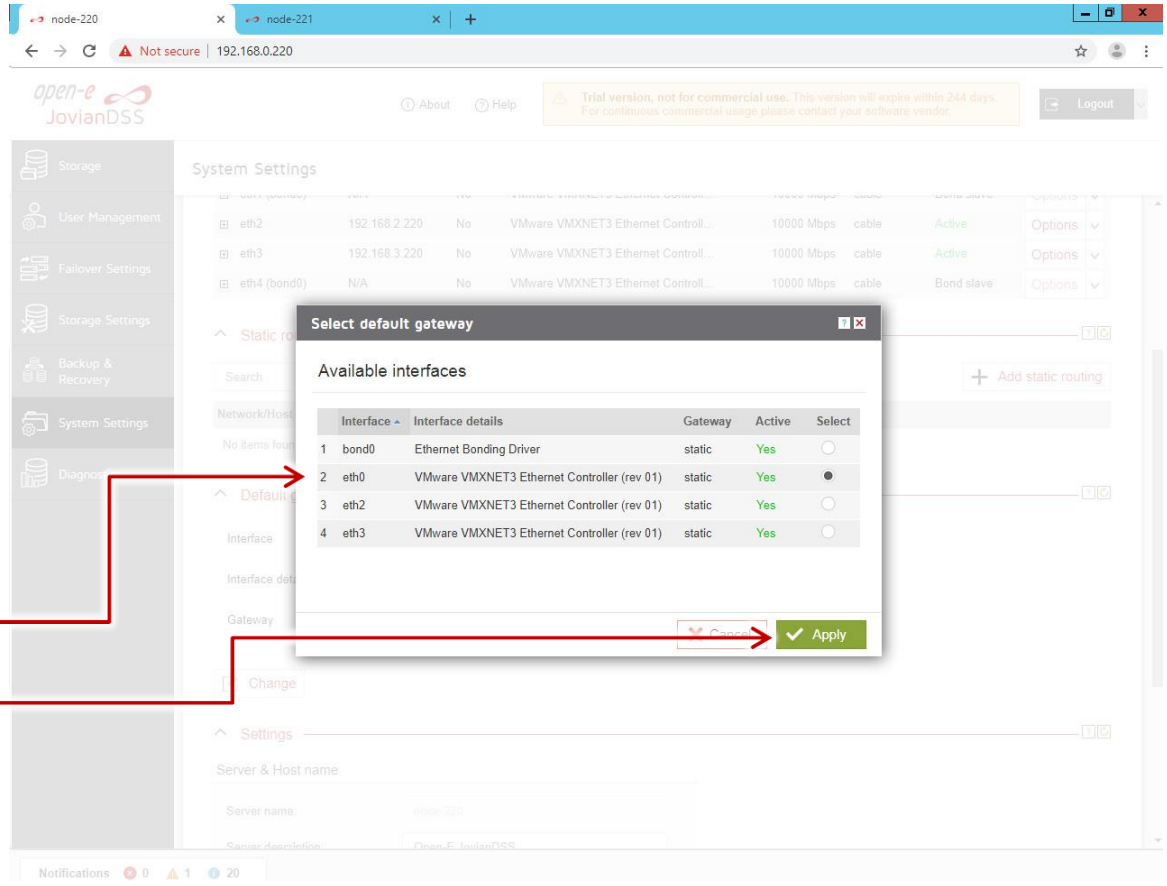


2.2. Network configuration. Select a default gateway

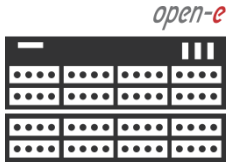


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Select a proper interface and click the **Apply** button.



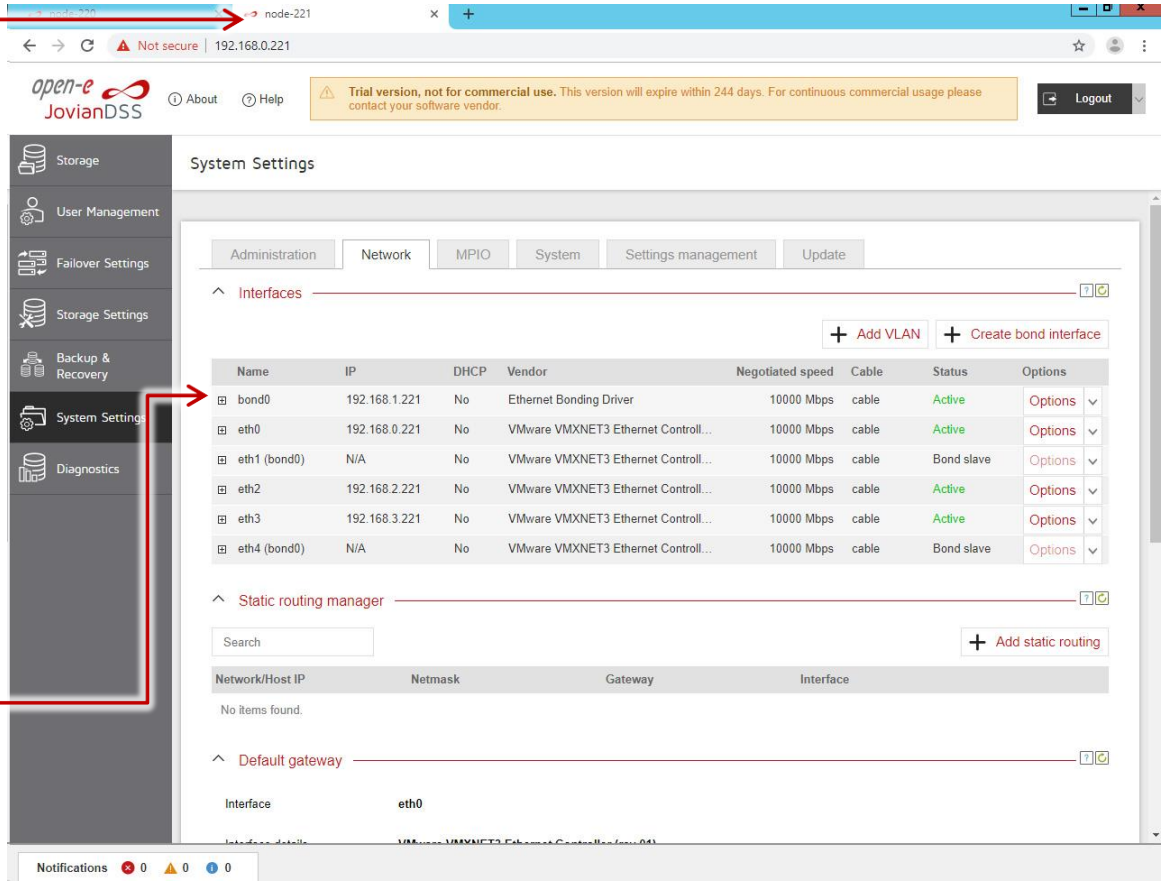
2.3. Network configuration. Network configuration on node-b



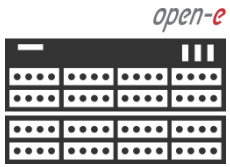
Open-E JovianDSS: **node-b**
IP Address: 192.168.0.**221**

Go to the **second cluster node** and create a bond interface accordingly.

The screenshot shows a properly created bond and a default gateway on the second node.



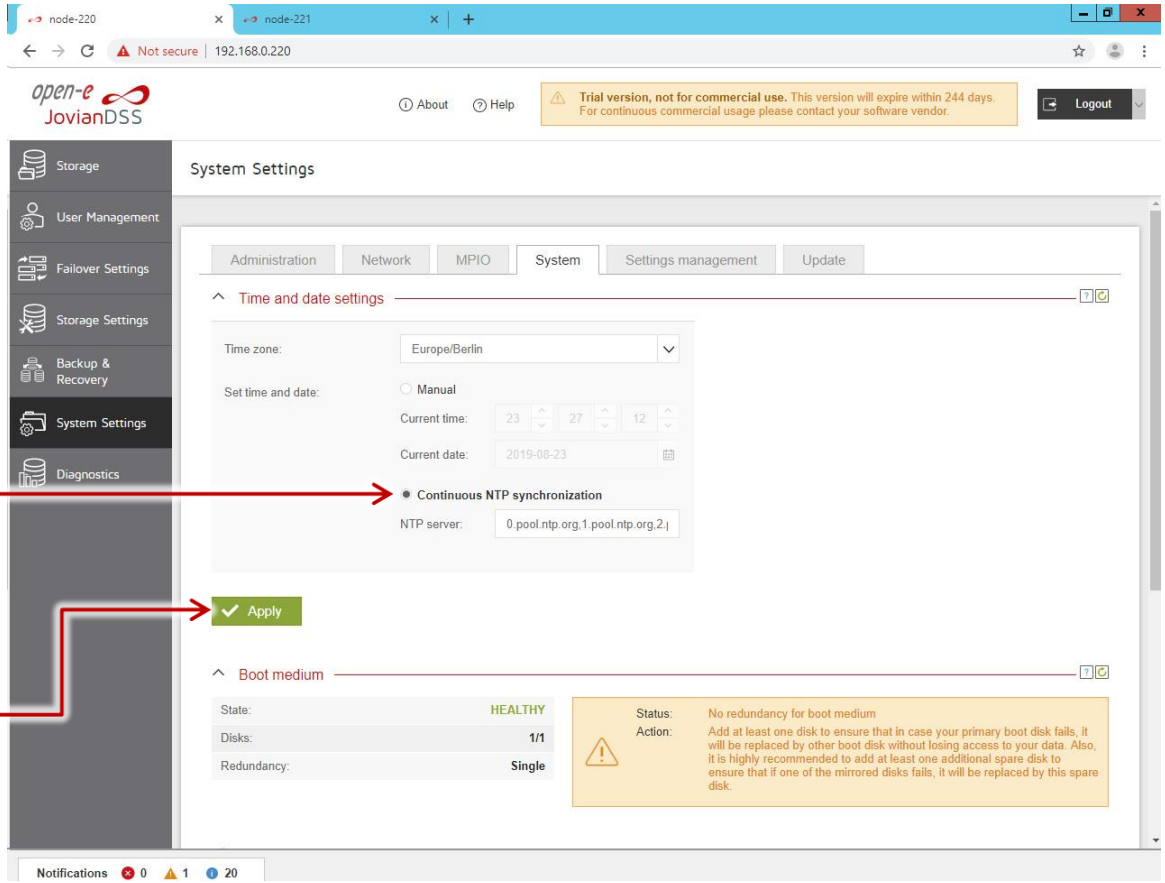
3. Time and date settings



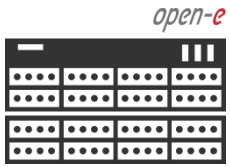
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.**220**

In the **System** tab, in **Time and date settings,** select the **Continuous NTP synchronization** and click apply.

Repeat this step for the second cluster node as well.



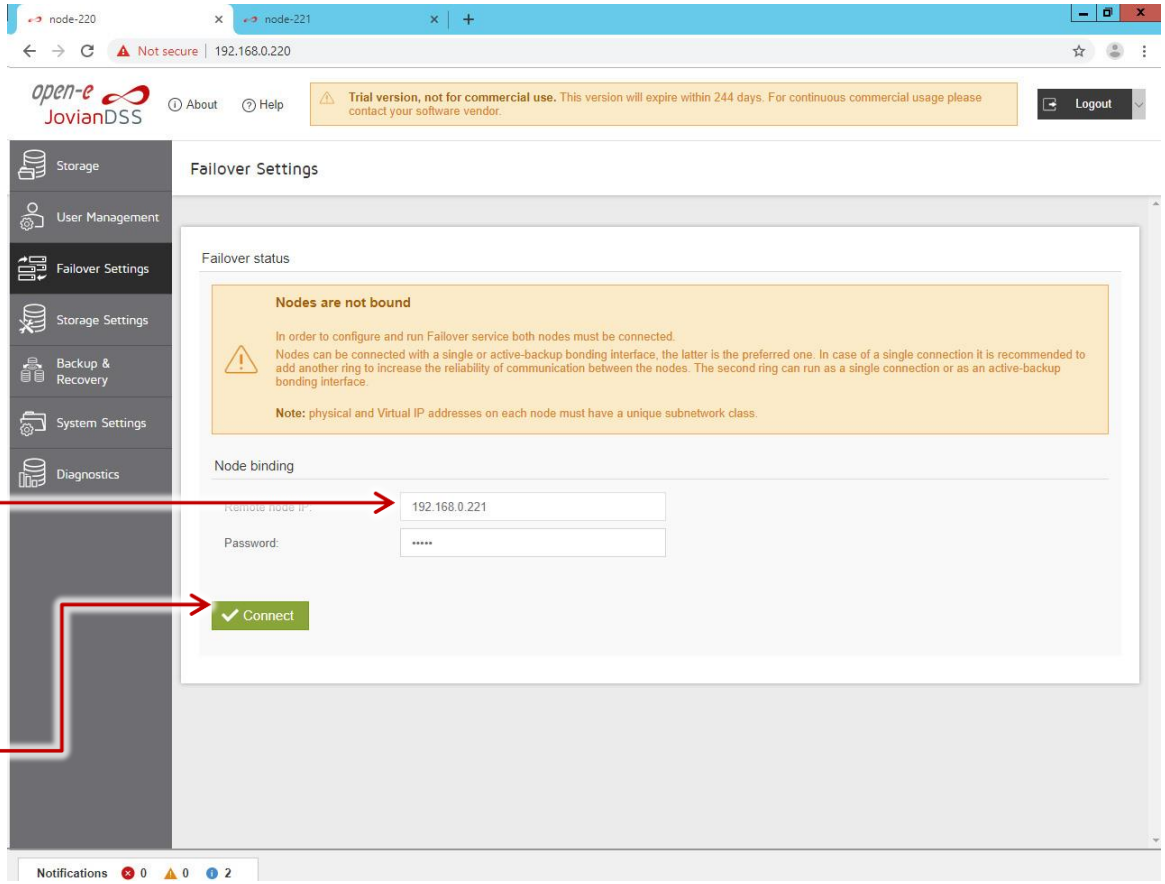
4. Nodes binding



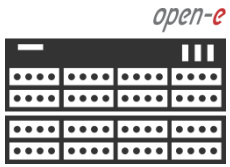
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In the main menu, select **Failover Settings** and enter the IP address of the NIC interface of the second node and enter the current administrator password (default: admin) and click the **Connect** button.

NOTE:
This is a new option. Version 1.0 up27 required to use Active-Backup bond for the cluster host binding.



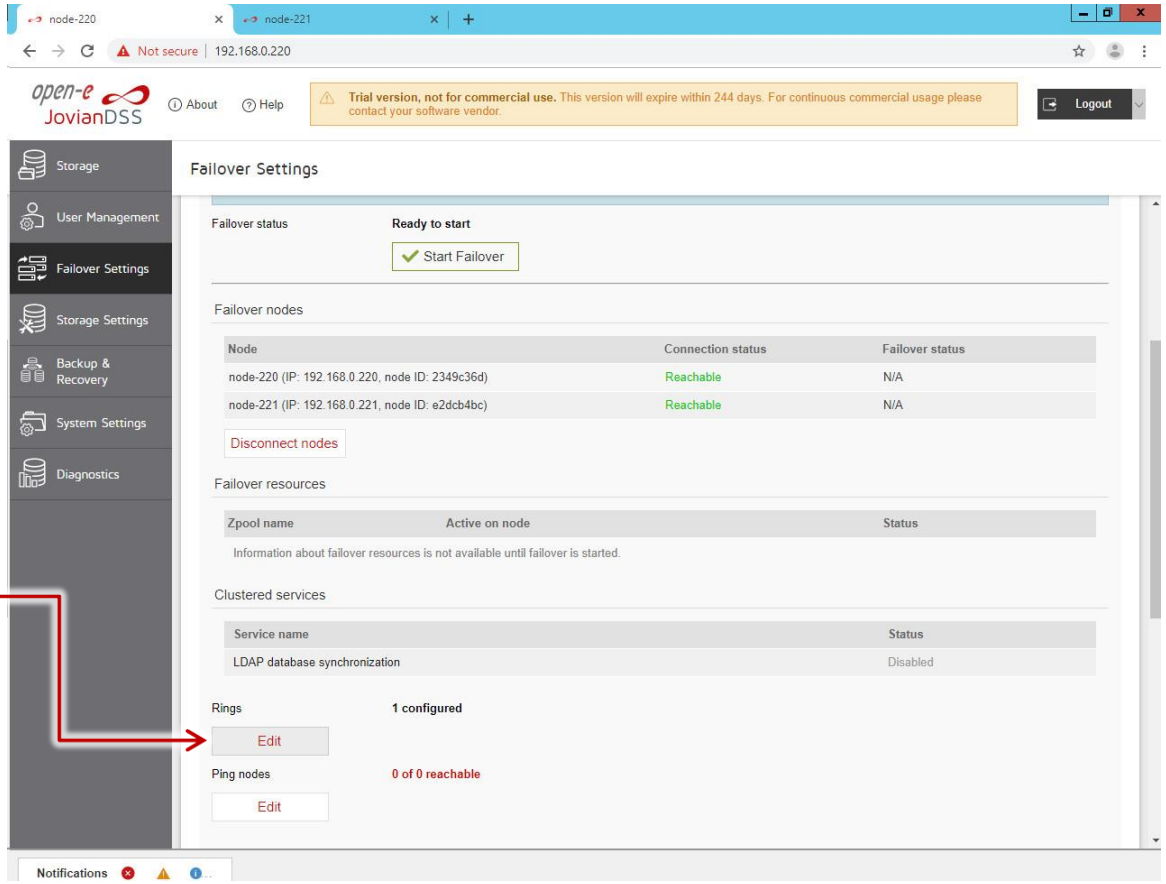
5. Adding rings



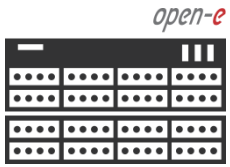
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In **Failover Settings**, click the **Edit** button in the **Rings** section and select at least two rings.

NOTE:
This is a new option. Version 1.0 up27 required to use an Active-Backup bond for the cluster host binding.



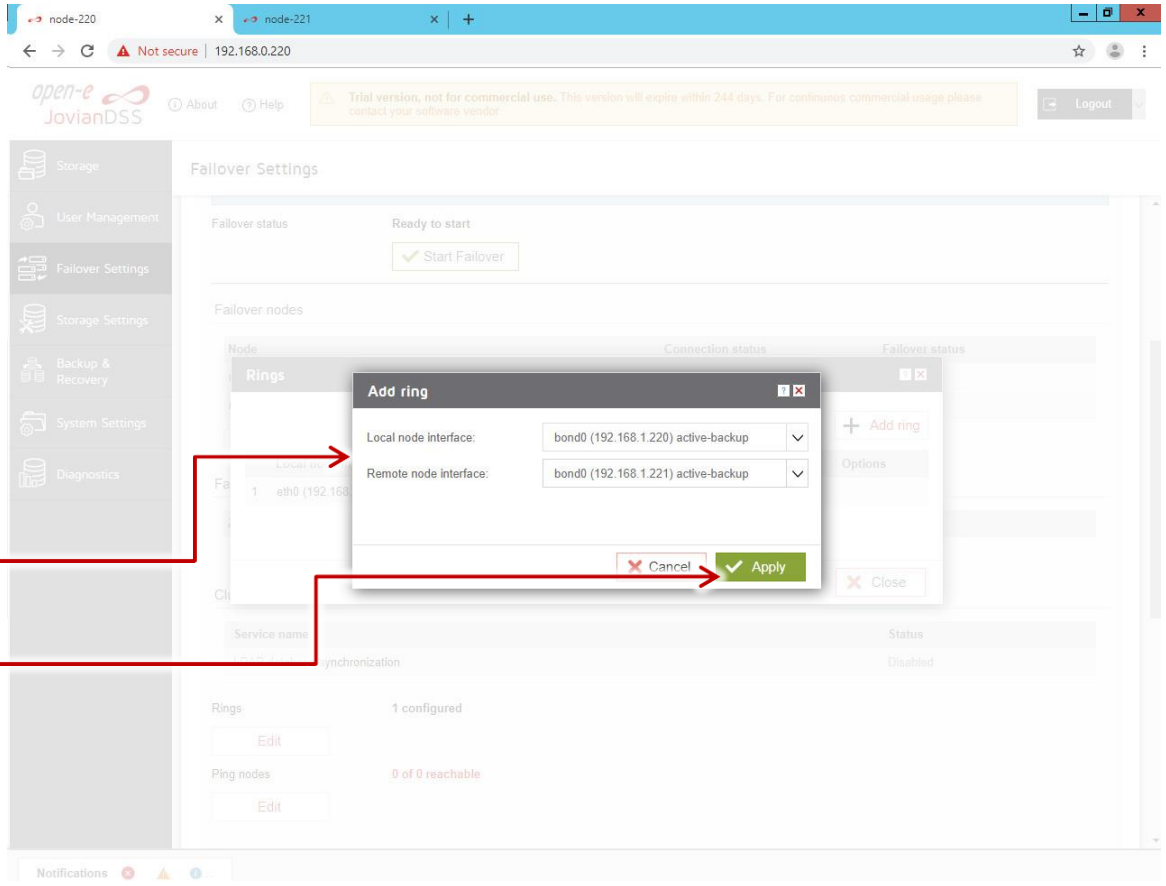
5. Adding rings



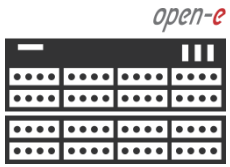
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Next, select the interfaces for the ring for local and remote nodes and click the **Apply** button.

NOTE:
The bond0 will be used for mirror path as well.



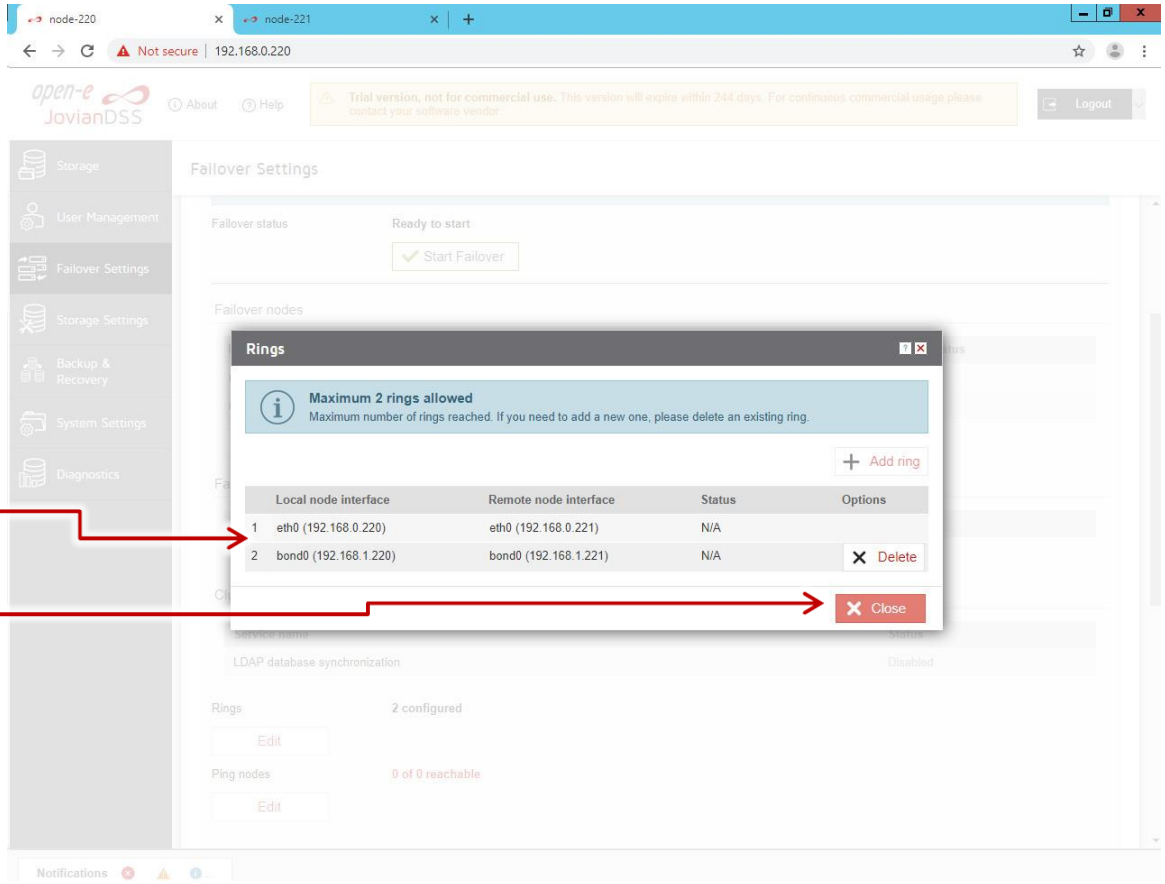
5. Adding rings



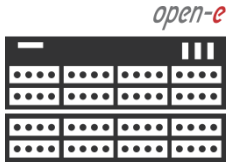
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Two rings in local and remote nodes are selected. Now, click the **Close** button.

NOTE:
Maximum number of 2 rings is allowed. If you need to add a new one, delete an existing ring.



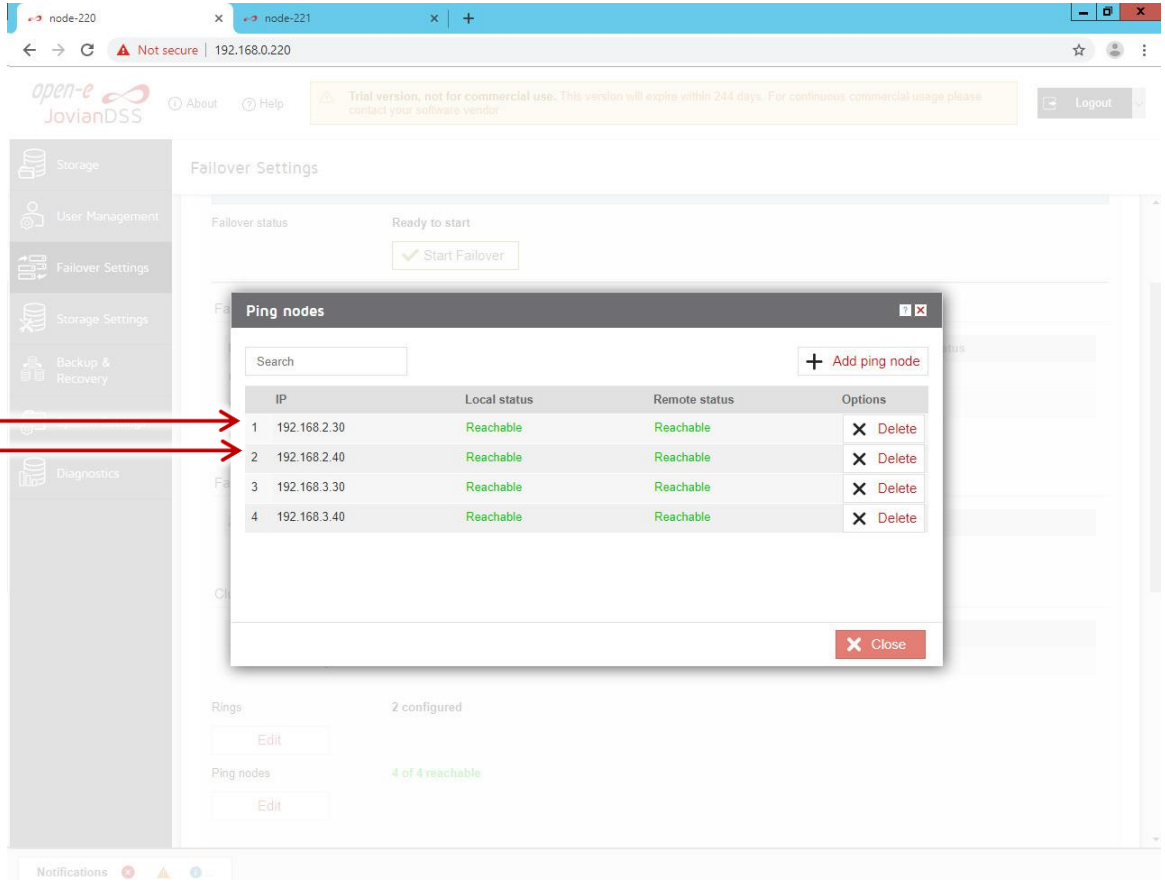
6. Ping Nodes



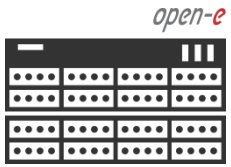
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In the Failover settings, click the **Edit** button in the **Ping nodes** section and enter at least two ping nodes.

NOTE:
It is recommended to configure more than 2 ping nodes but NOT more than 6.

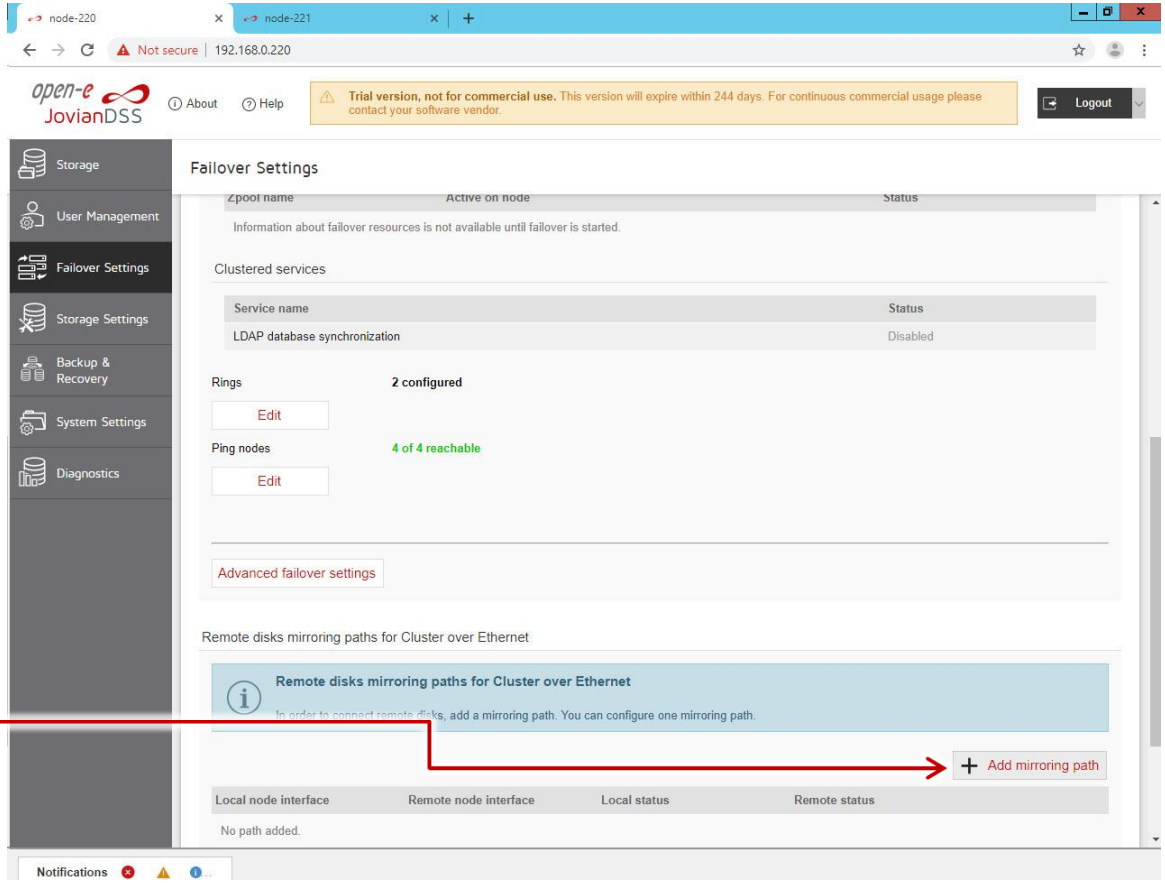


7. Mirroring path

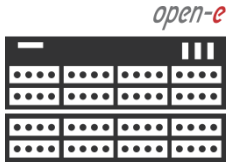


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Next, click the **Add mirroring path** button.

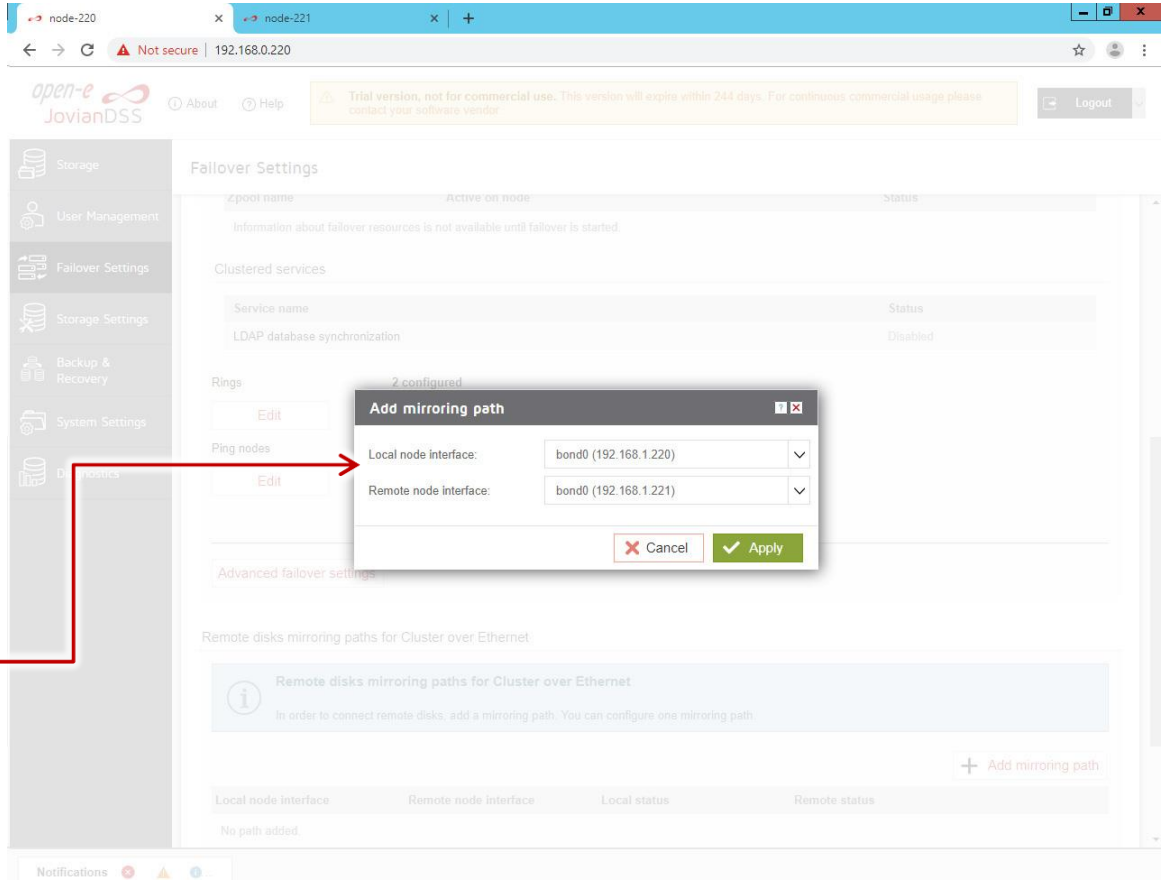


7. Mirroring path

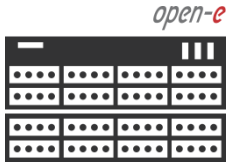


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In **Add mirroring path**, select proper interfaces and click the **Apply** button.

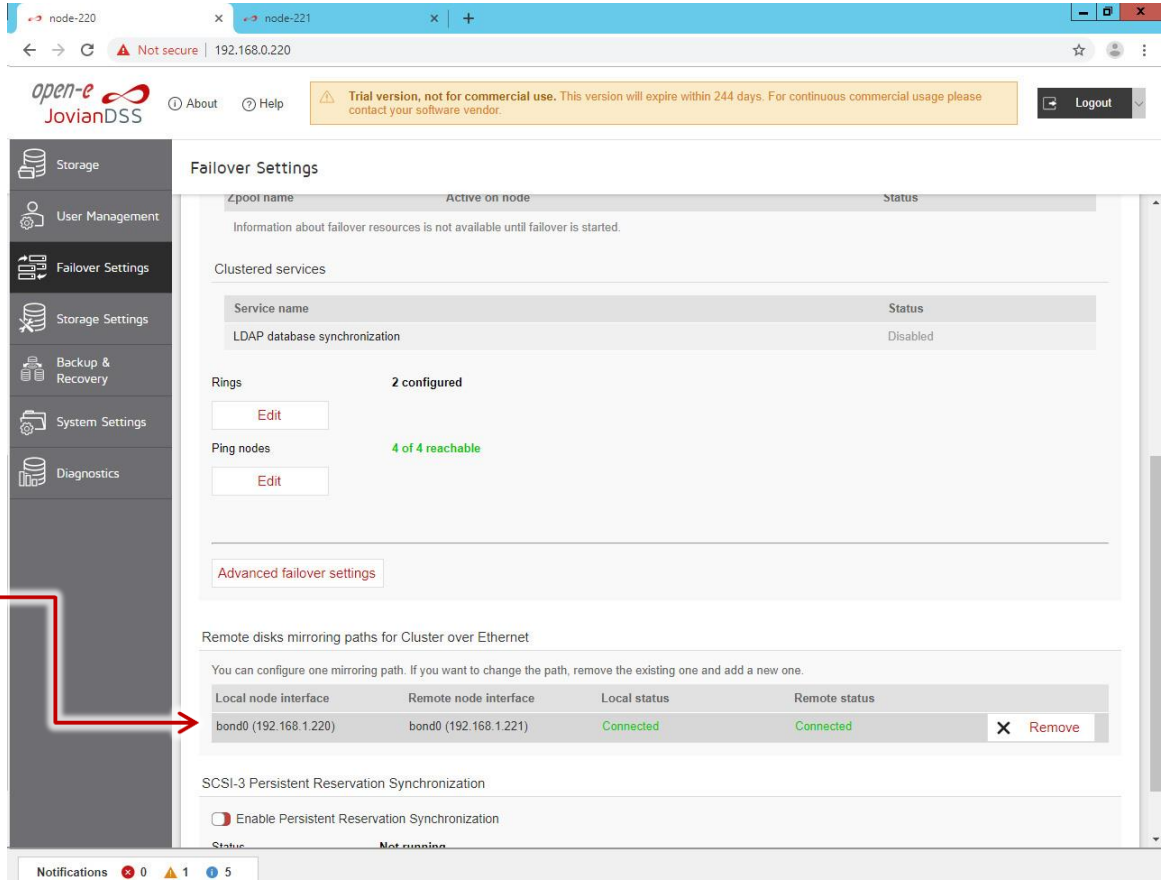


7. Mirroring path

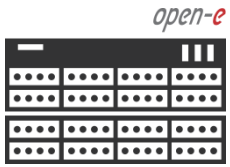


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Mirroring path shows the **Connected** status.

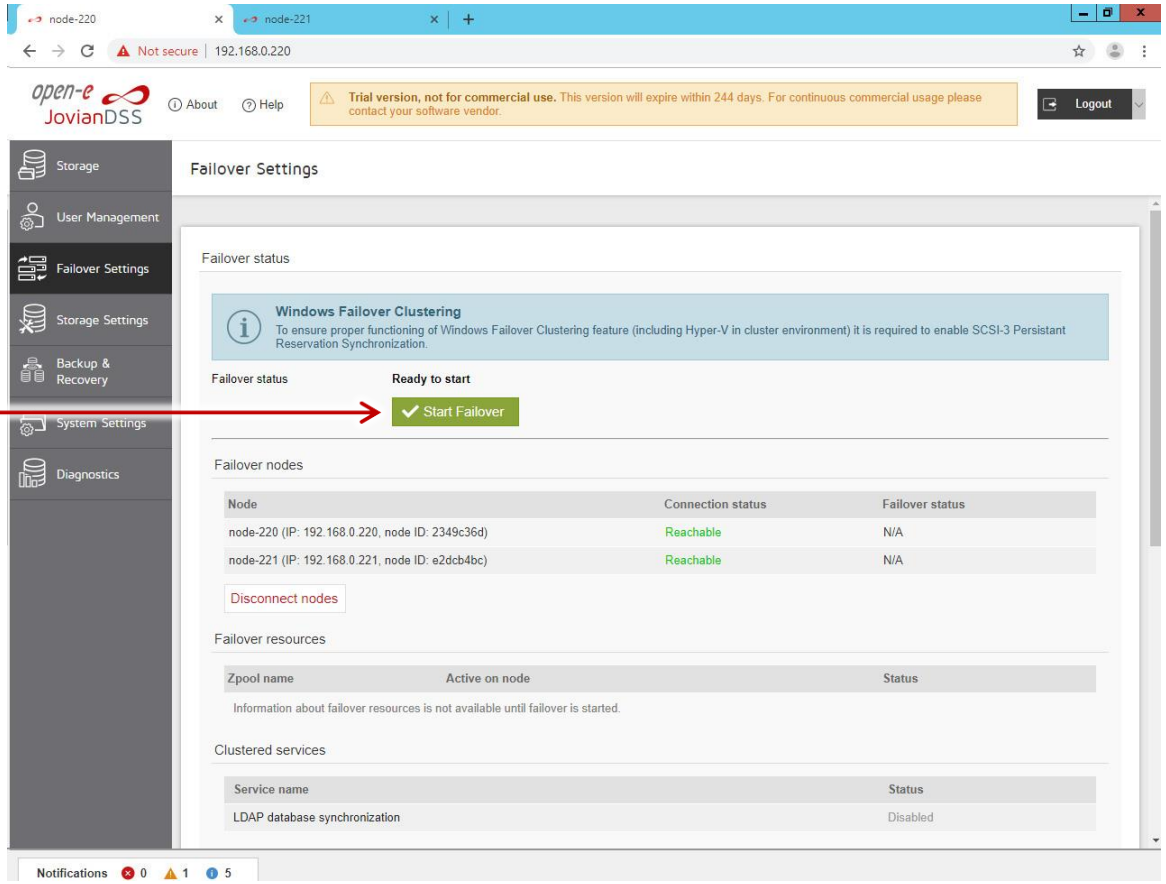


8. Start the cluster service

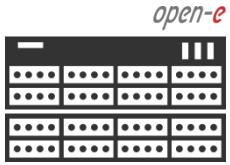


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Now, the cluster is ready to start.
In order to start the cluster services, click the **Start Failover** button.

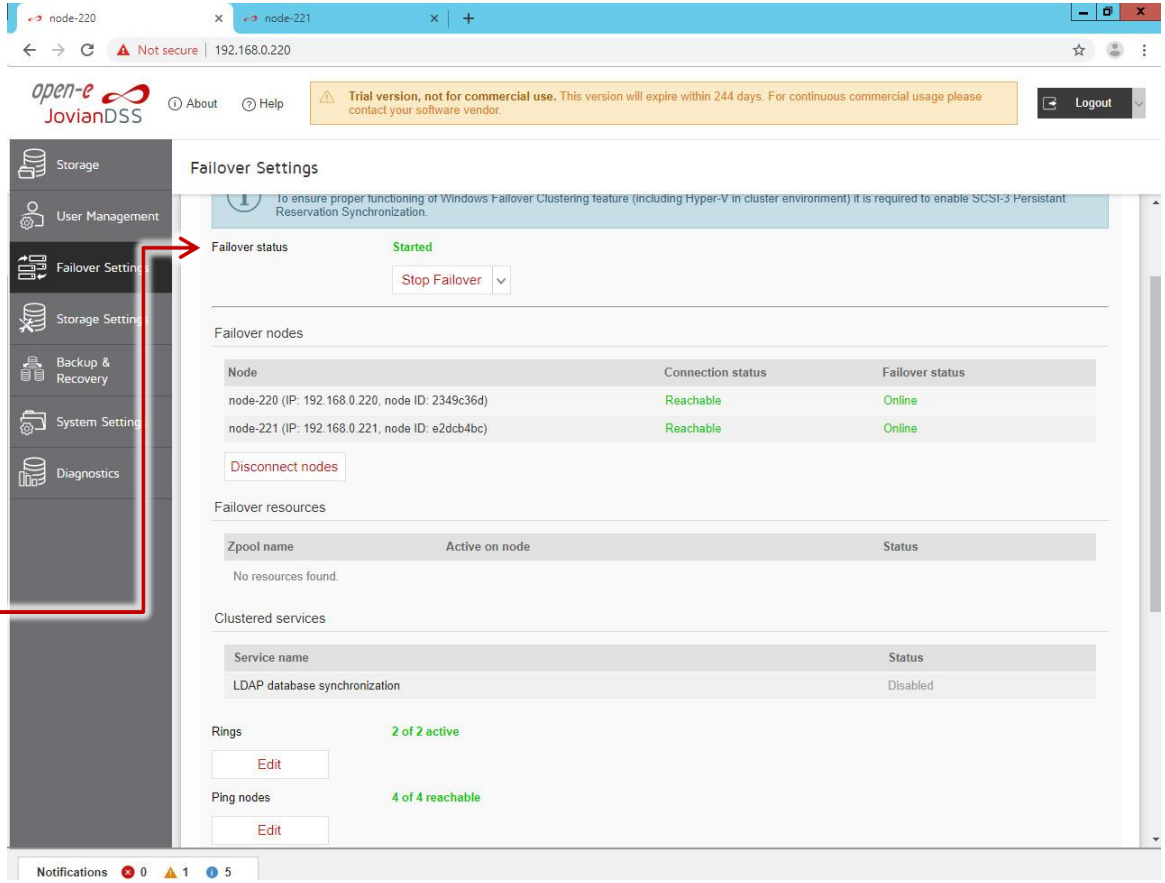


8. Start the cluster service

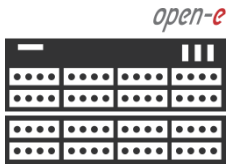


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In a short moment, the HA Cluster will be started and the status will show: **Started**.

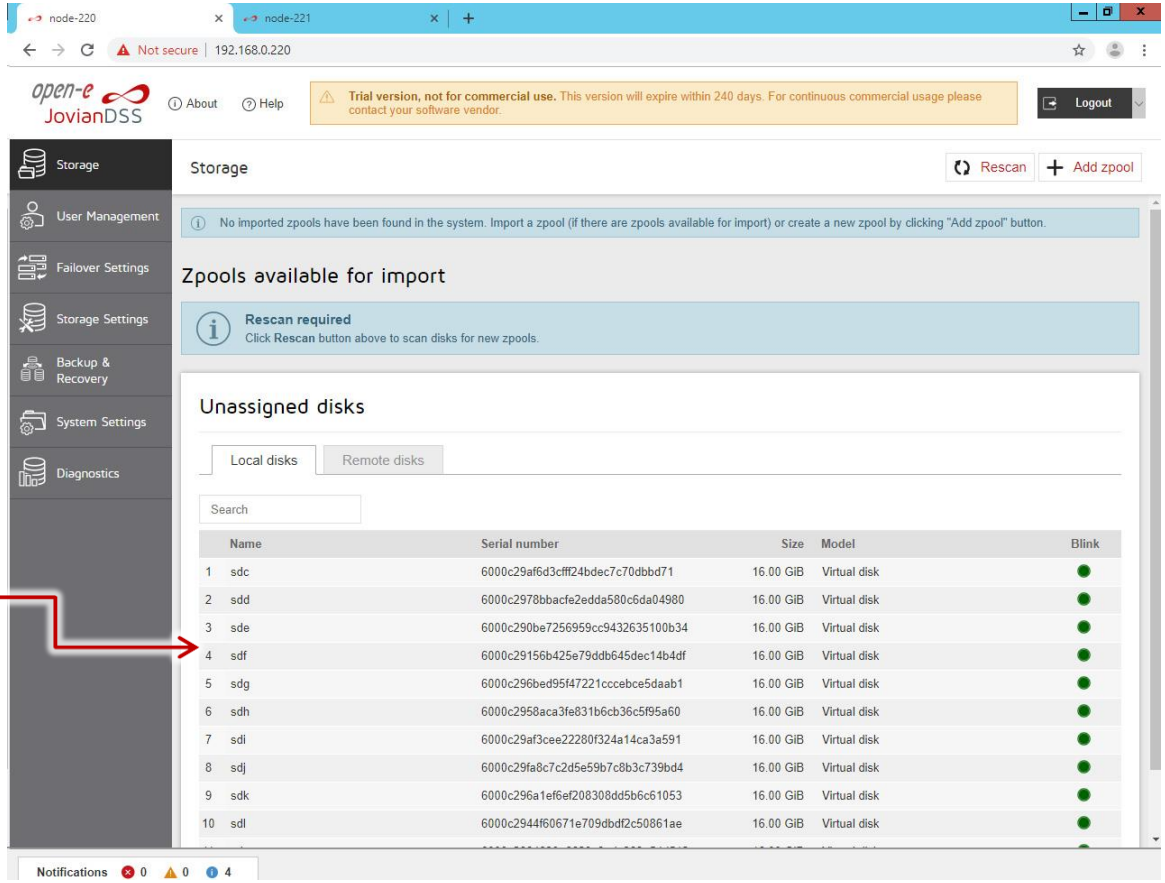


9. Create a new Pool

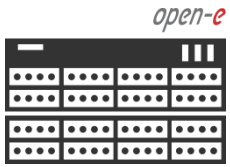


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Go to the menu **Storage**. In the **Local disks** tab, all local disks are listed.

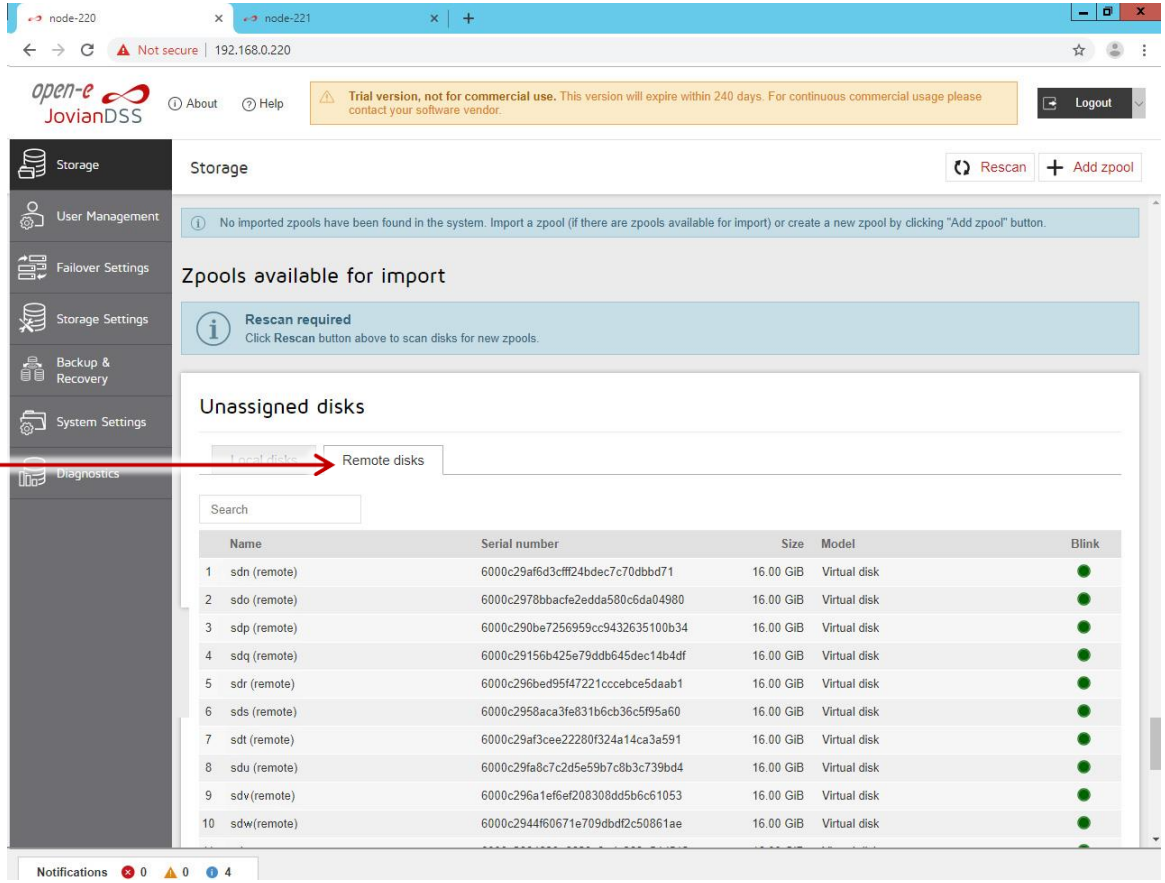


9. Create a new Pool

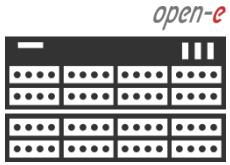


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In the **Remote disks** tab, all disks from a remote node are listed.



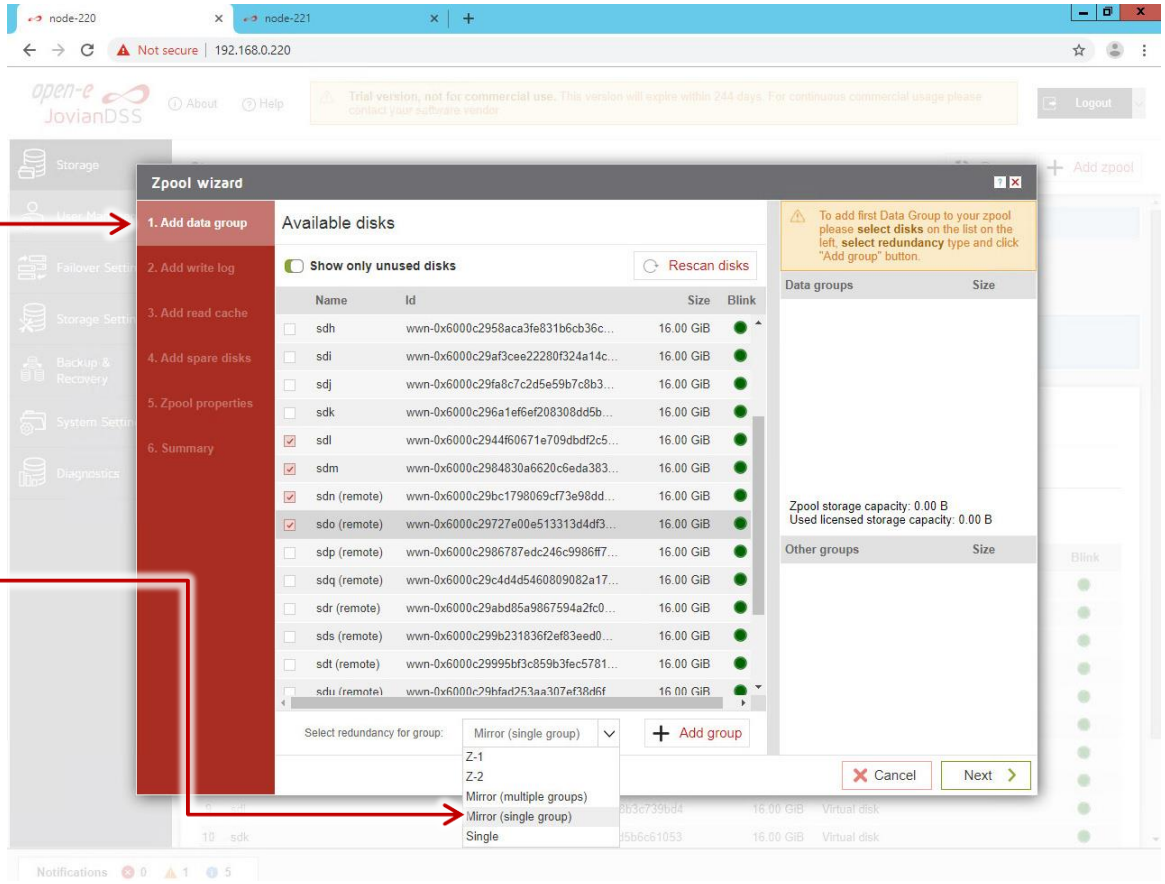
9. Create a new Pool



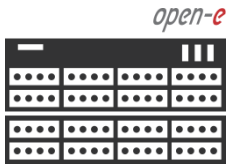
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In the Storage tab, click **Add zpool button**. Then, add data groups by selecting 4 (2 local and 2 remote) disks and select Mirror (single group) from the pull-down menu and click the Add group button, then click the **Next** button.

NOTE: It is recommended to set 4 (**2-local + 2-remote**) disks in every mirror group. In case of other node reboots or failure, the mirror groups are still redundant with 2 disks in every mirror.



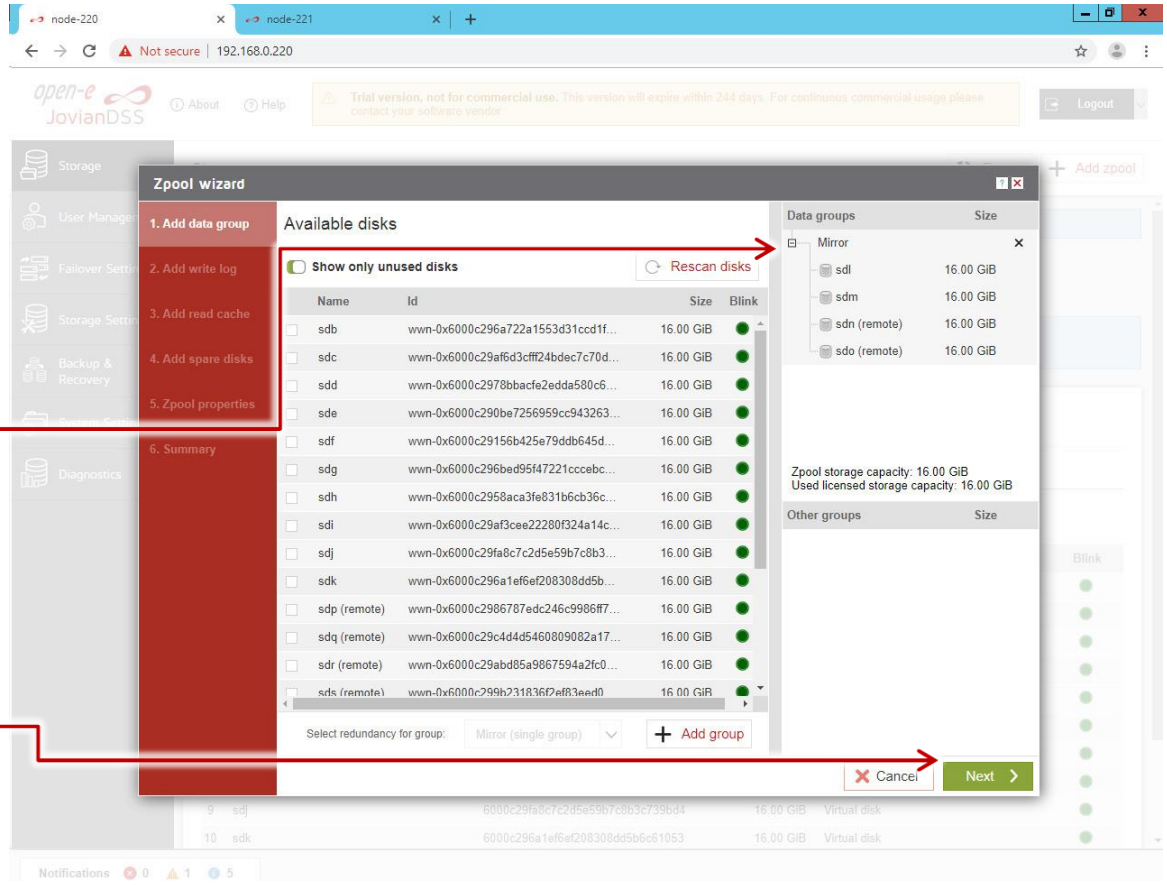
9. Create a new Pool



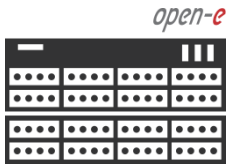
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Click the **Next** button.

NOTE:
A 4-way mirror provides limited storage efficiency of only 25%. For increased storage efficiency, a hardware RAID controller can be used with a simple mirror over 2 disk units with a RAID array behind. Open-E JovianDSS includes built-in all RAID tools and drivers for Broadcom (LSI) and Microsemi (Adaptec). In case of Areca, a driver is included as well but the WebGUI of Areca has to be accessed via the controller's ETH-port.

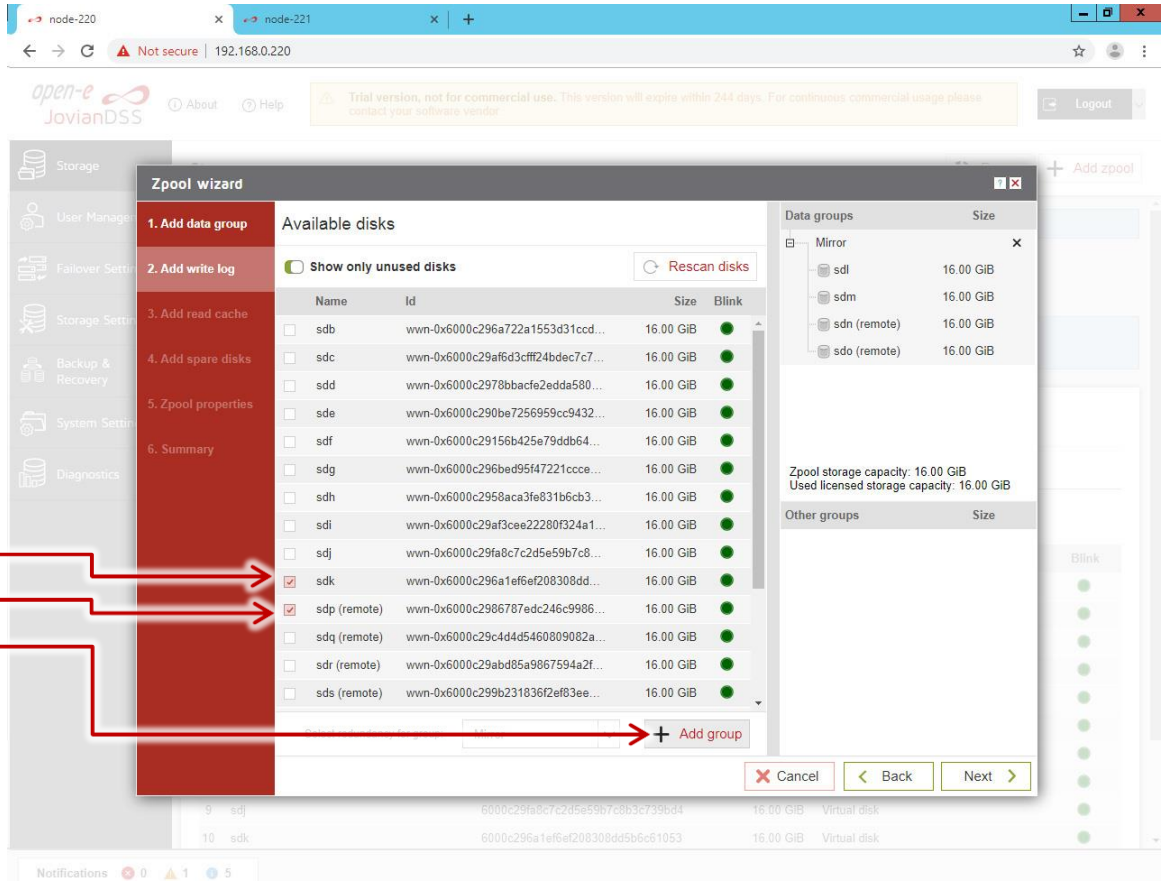


9. Create a new Pool

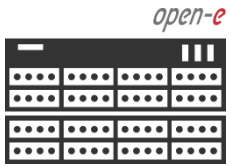


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Select 2 disks (**local + remote**) for the write log and click the **Add group** button.

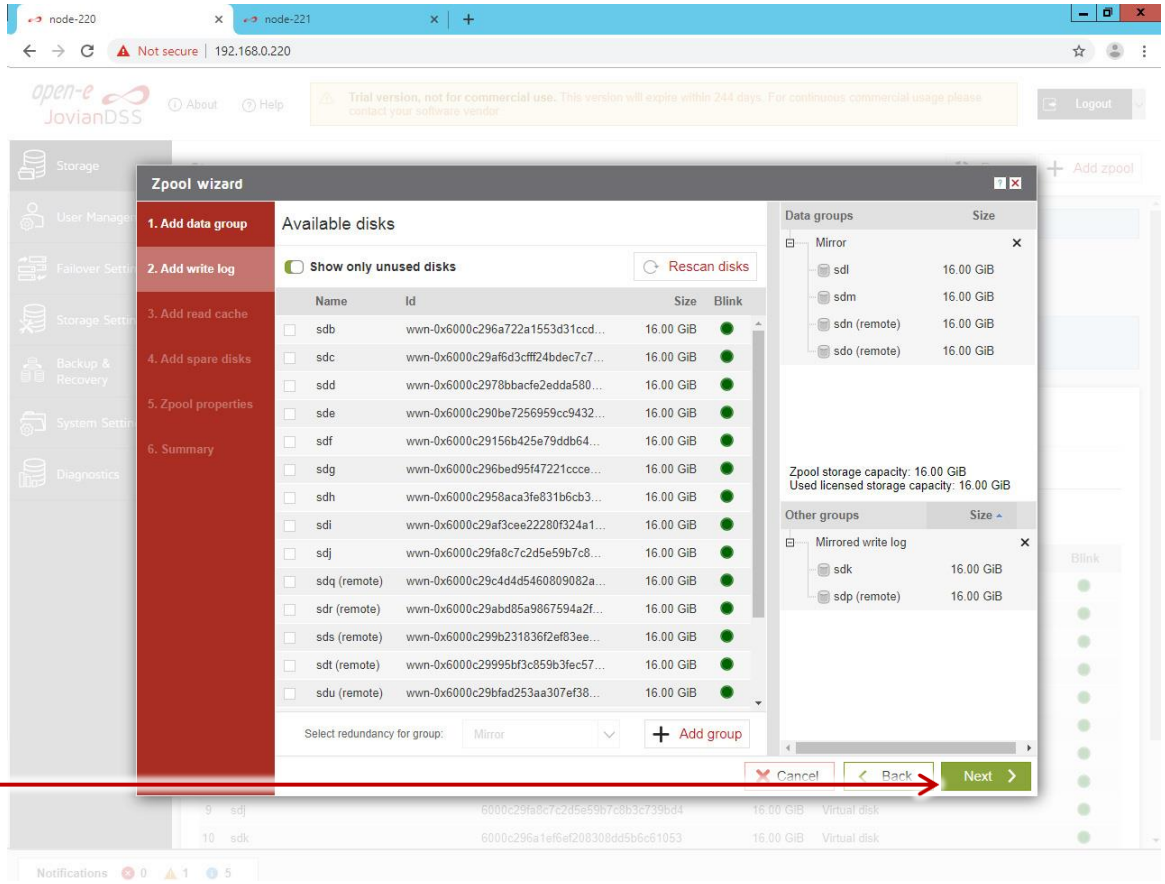


9. Create a new Pool

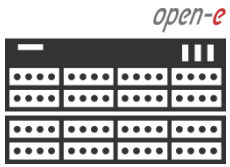


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Click the **Next** button.

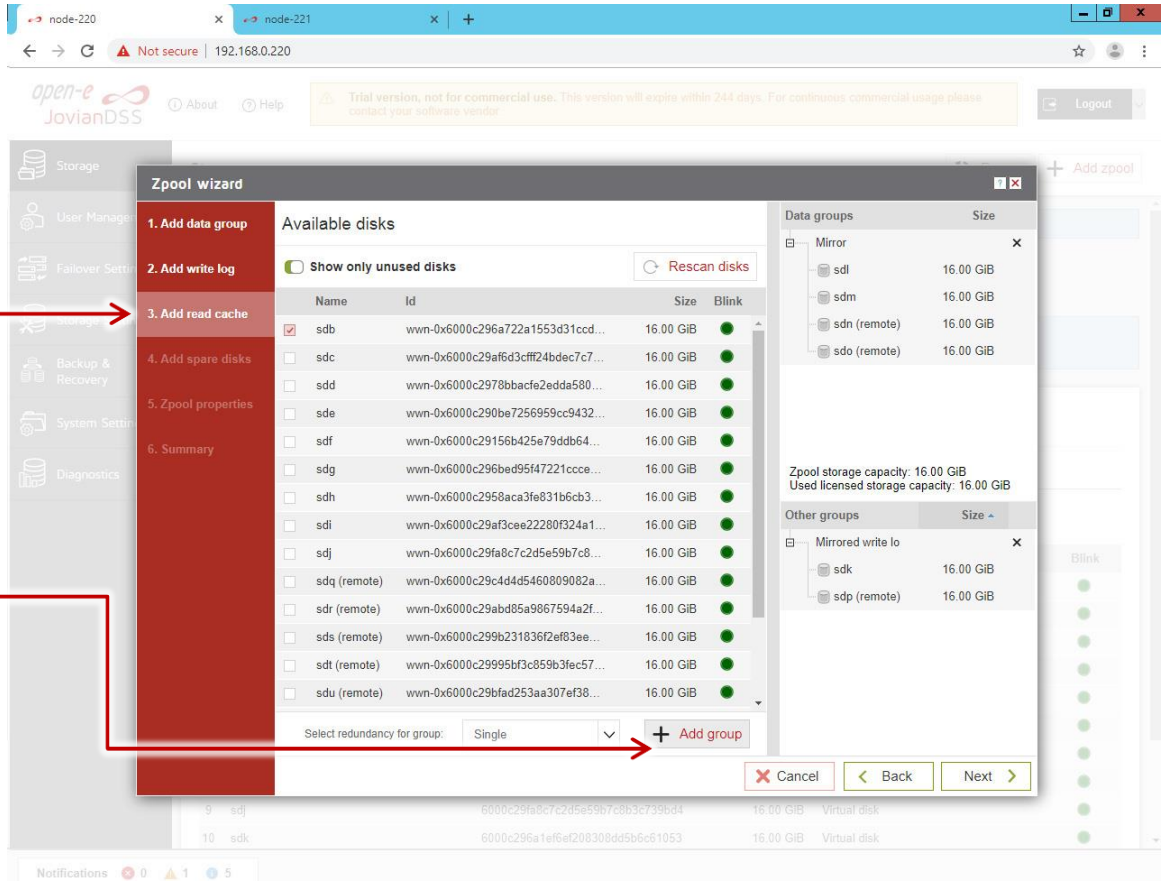


9. Create a new Pool

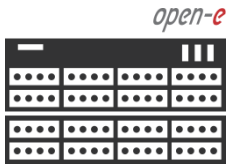


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Select a local SSD disk for the level-2 read cache and click the **Add group** button.

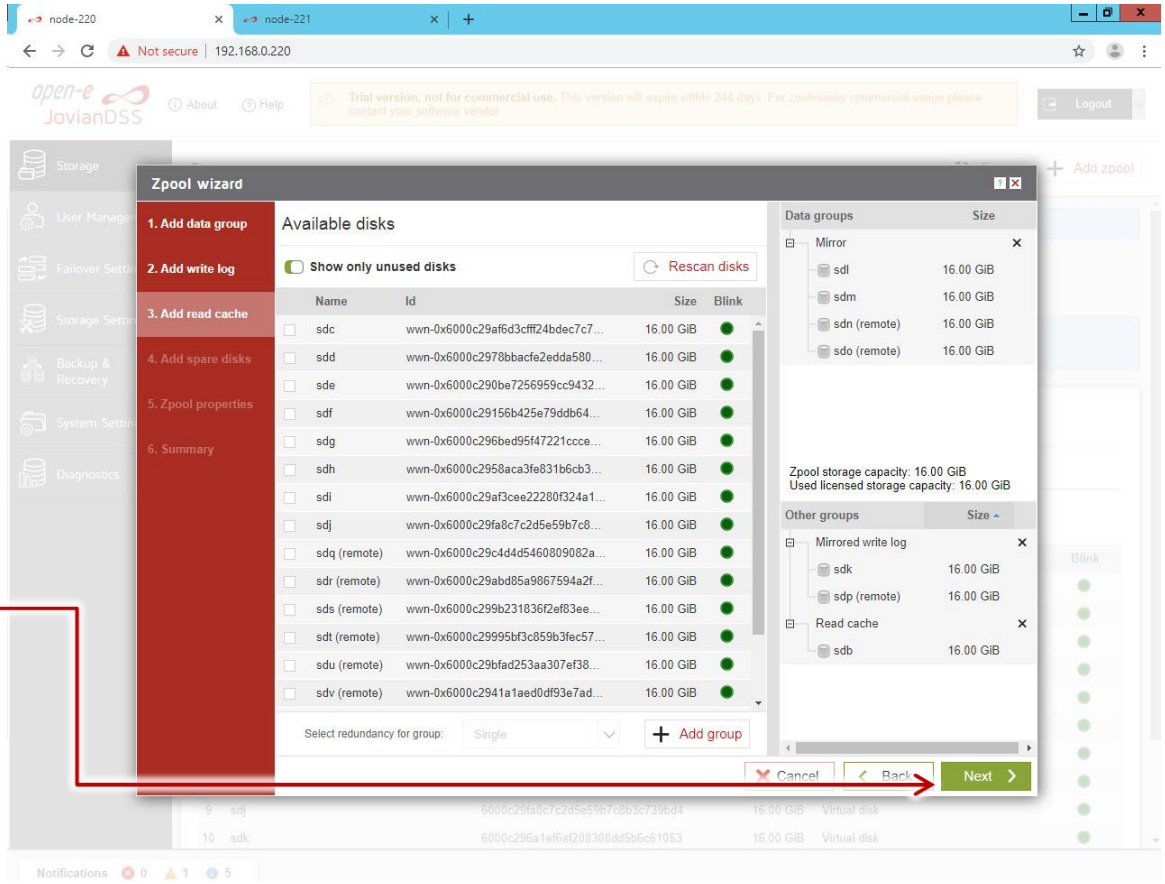


9. Create a new Pool

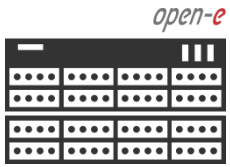


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Click the **Next** button.

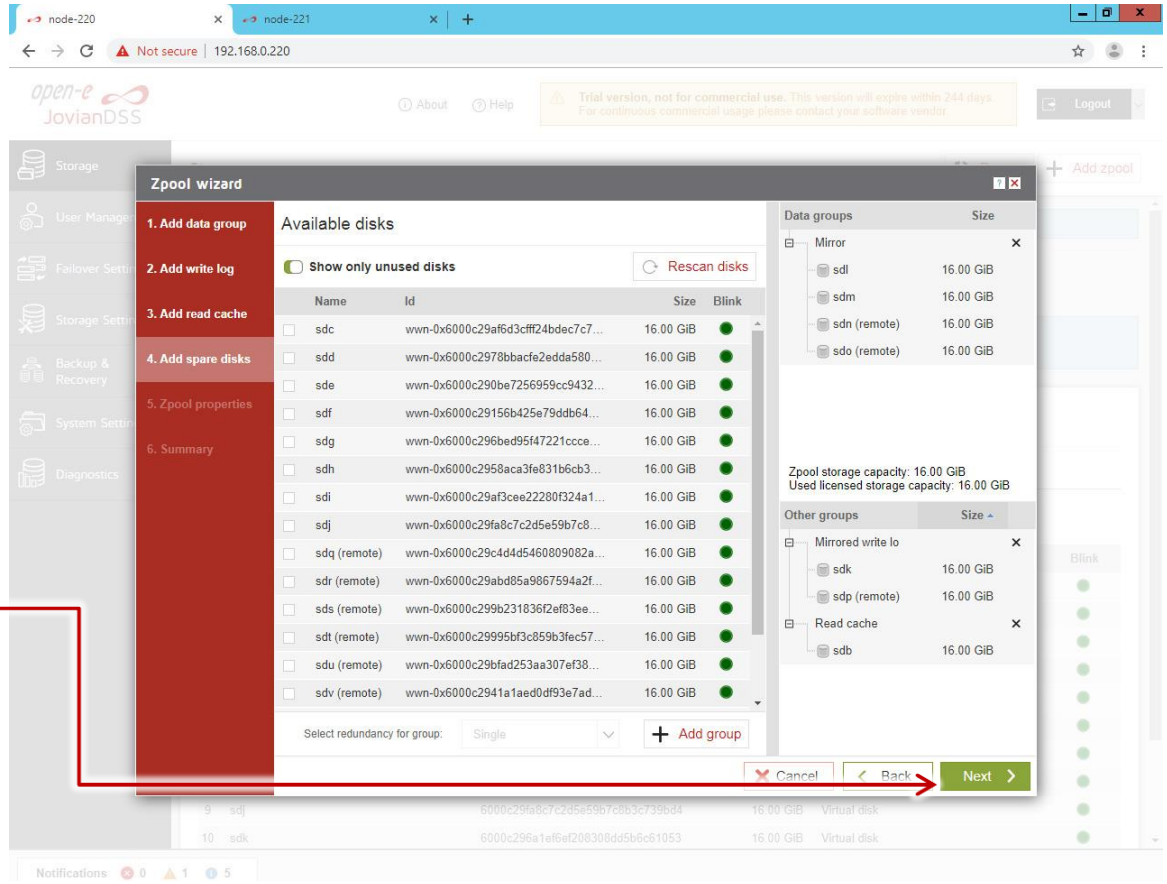


9. Create a new Pool

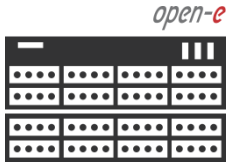


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In this setup, we skip the **Add spare disks**. You can always add spares later if you need.
Now, click the **Next** button.

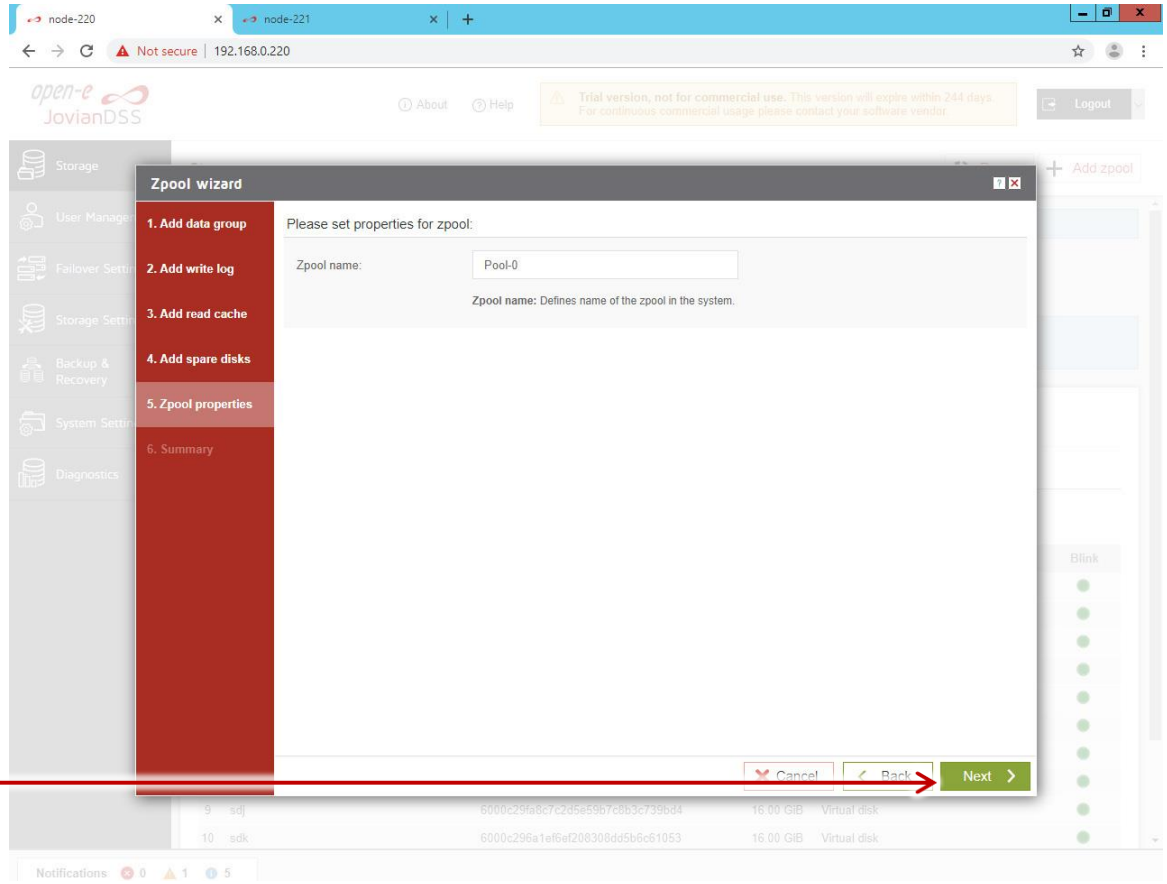


9. Create a new Pool

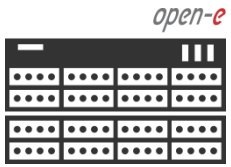


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Confirm the pool name, then click the **Next** button.

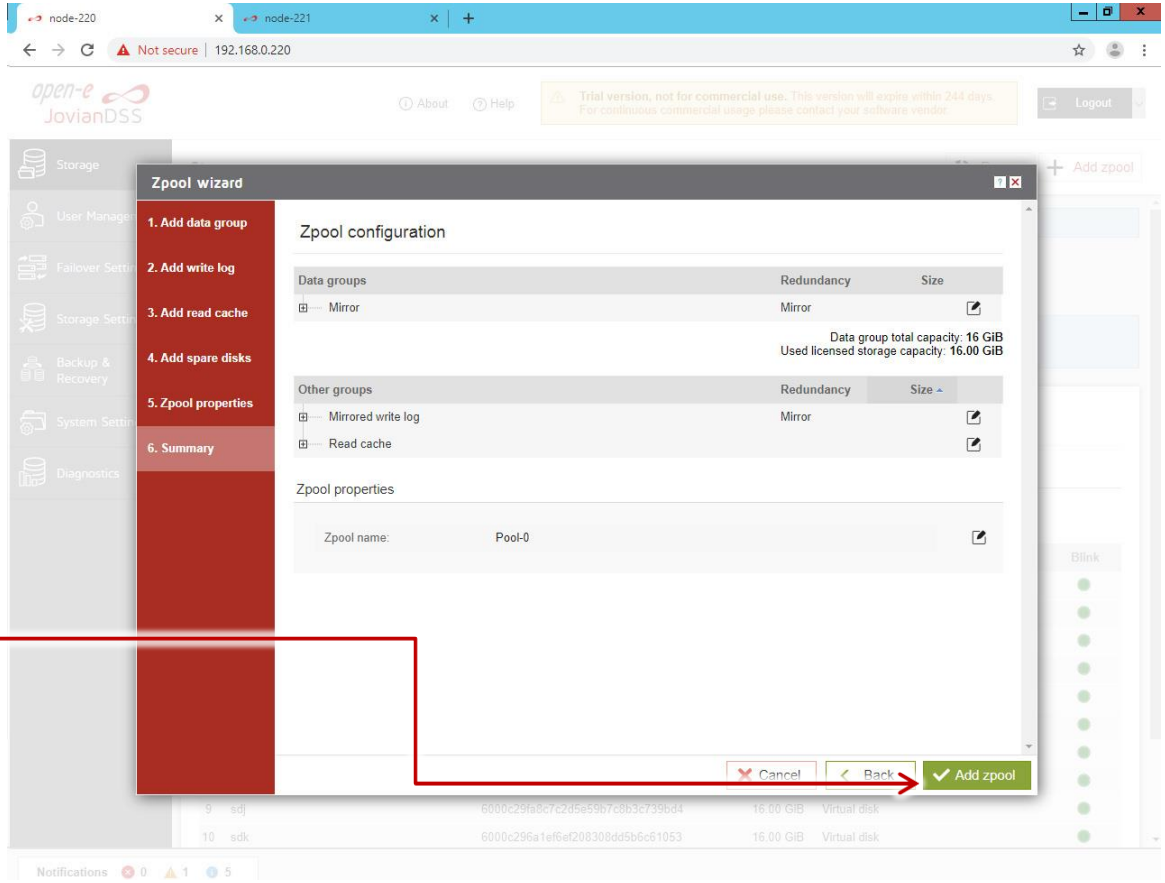


9. Create a new Pool

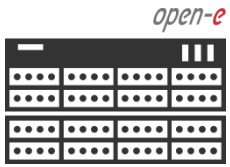


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Summary of the **Zpool wizard** step, then click the **Add zpool** button.

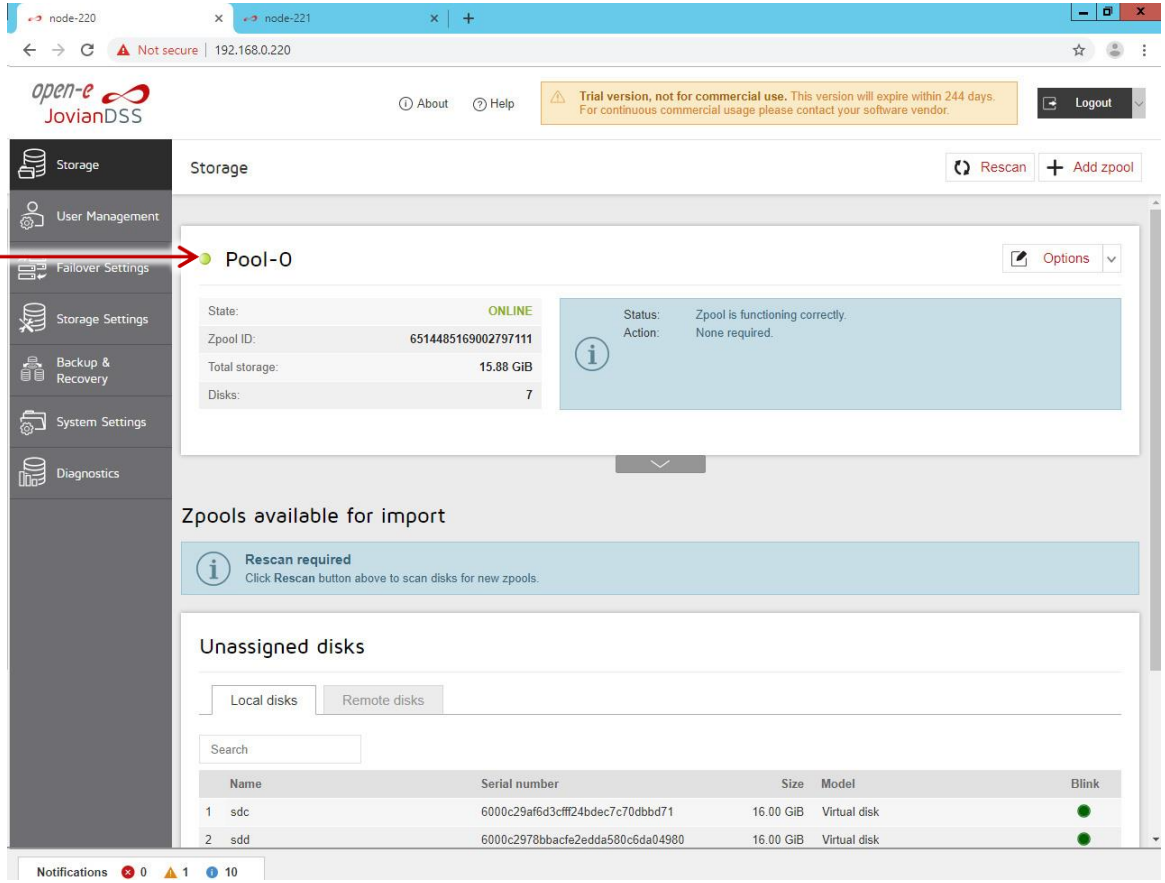


9. Create a new Pool

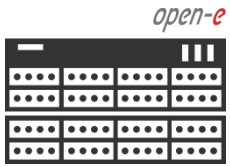


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

The **Pool-0** is now created.
Accordingly, you can configure a second pool (**Pool-1**).

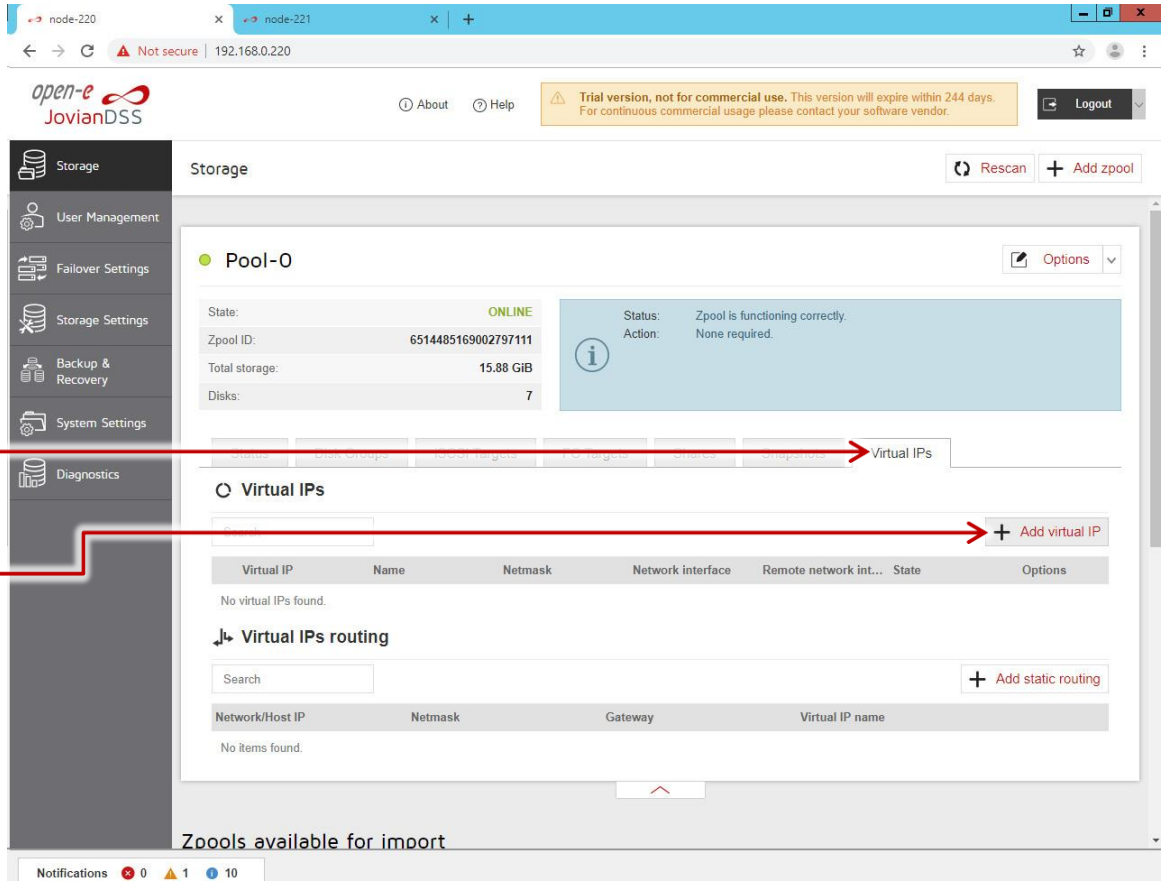


10. Enter virtual IP

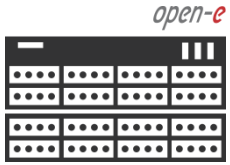


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In the **Storage** menu, select the **Virtual IPs** tab. Next, click **Add virtual IP**.

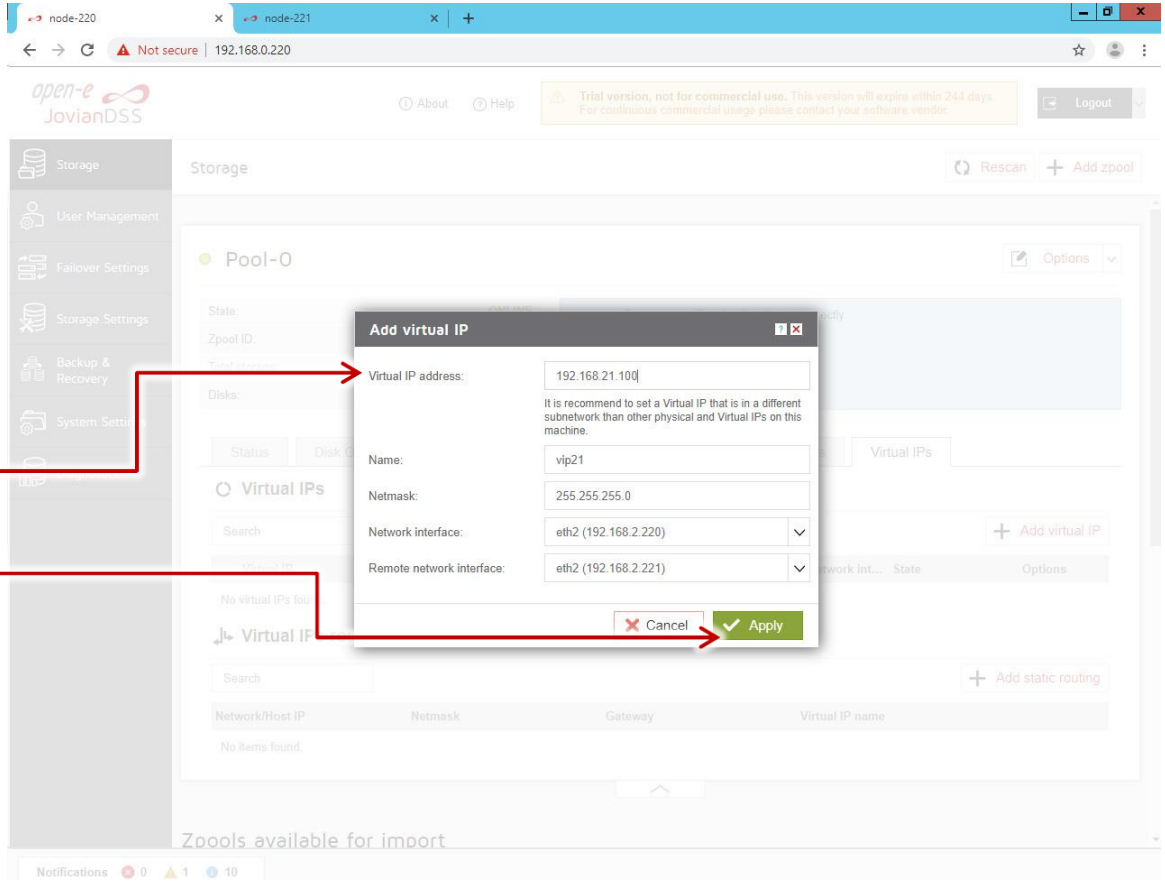


10. Enter virtual IP

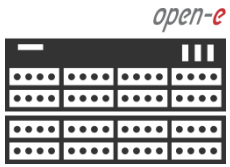


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Next, enter the virtual IP address and assign it to the required interfaces. Next, click the **Apply** button.



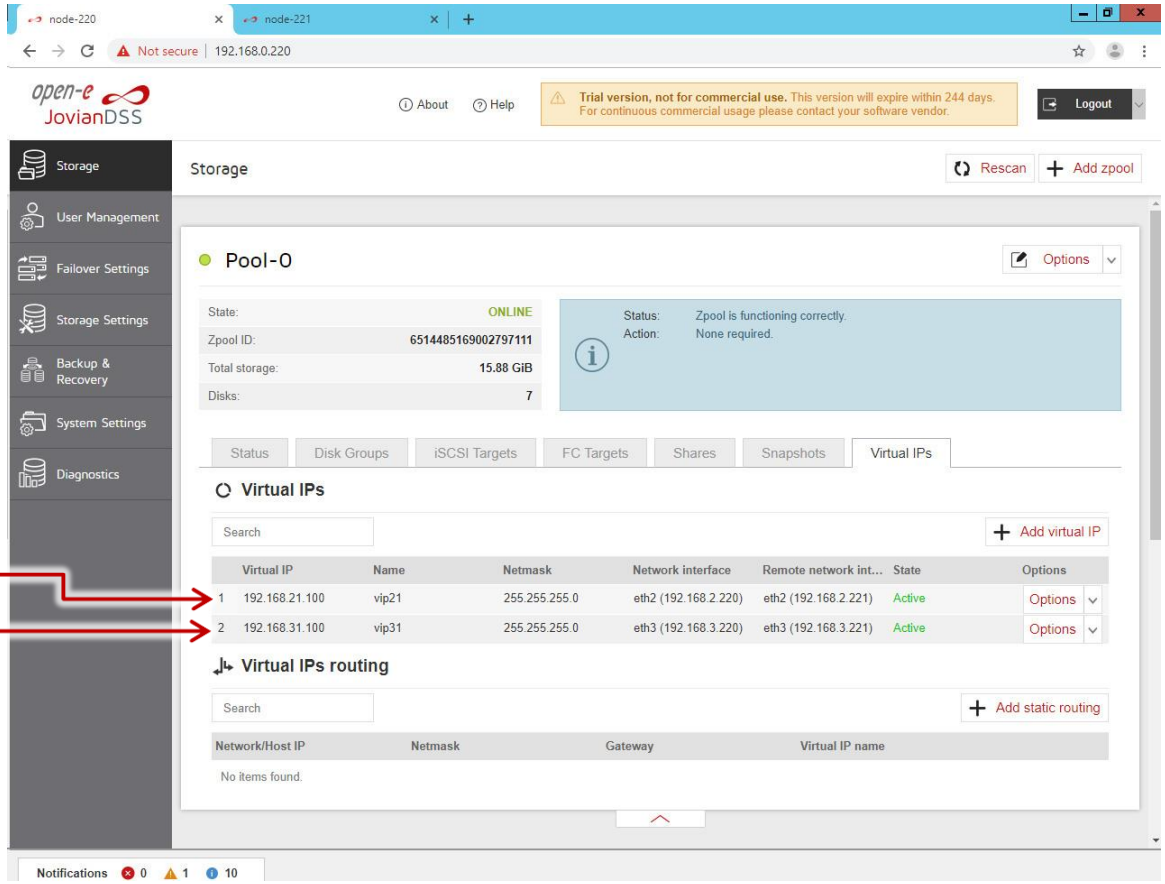
10. Enter virtual IP



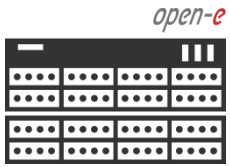
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Both virtual IP addresses are created on **Pool-0.**

Note: the VIP's are in separate networks as this is recommended.



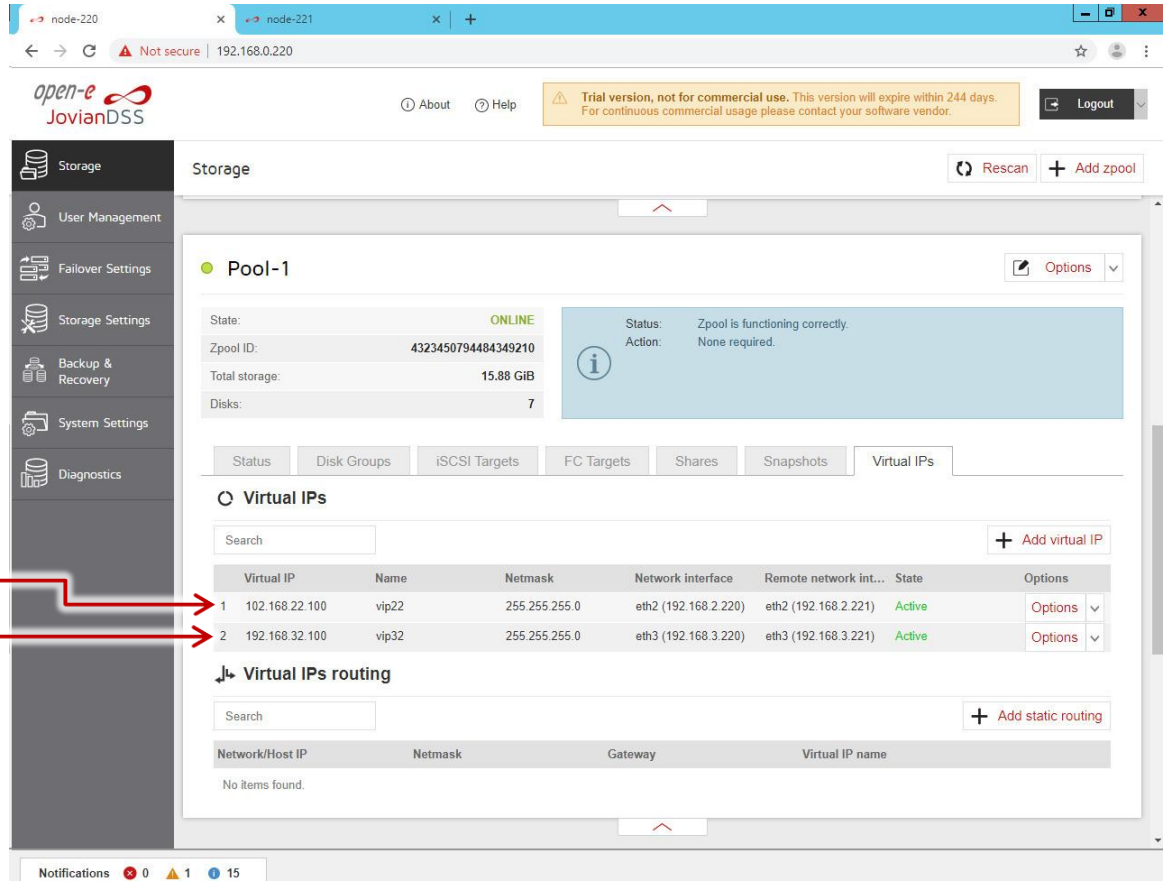
10. Enter virtual IP



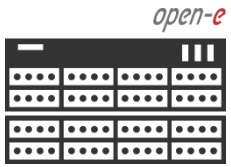
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Both virtual IP addresses are created on **Pool-1**.

Note: the VIP's are in separate networks as this is recommended.

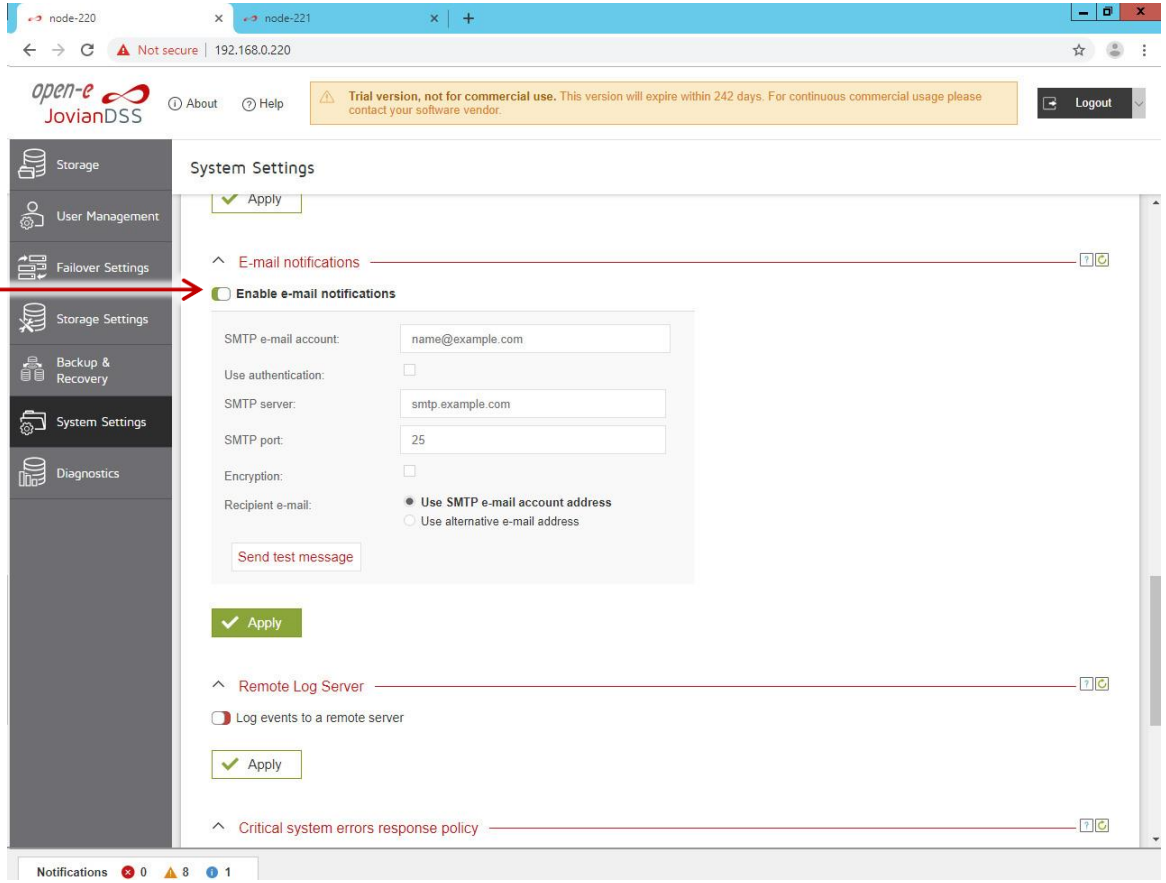


11. System monitoring setup

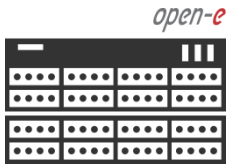


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

In the **System Settings** tab, set up the proper **E-mail notifications.**

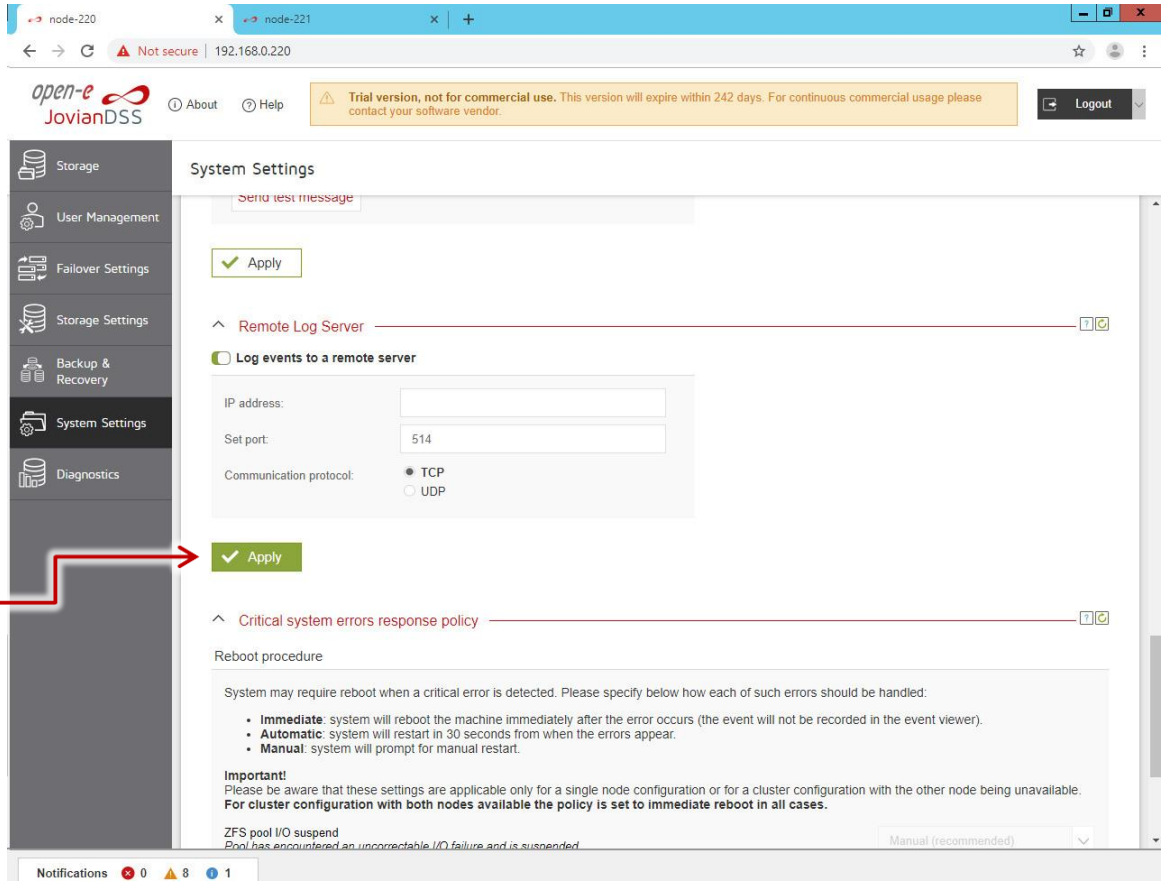


11. System monitoring setup

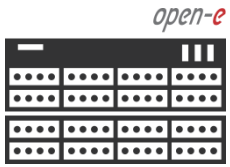


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

It is obligatory to use external monitoring software via SNMP or **Remote Log Server** or a built-in Checkmk agent.



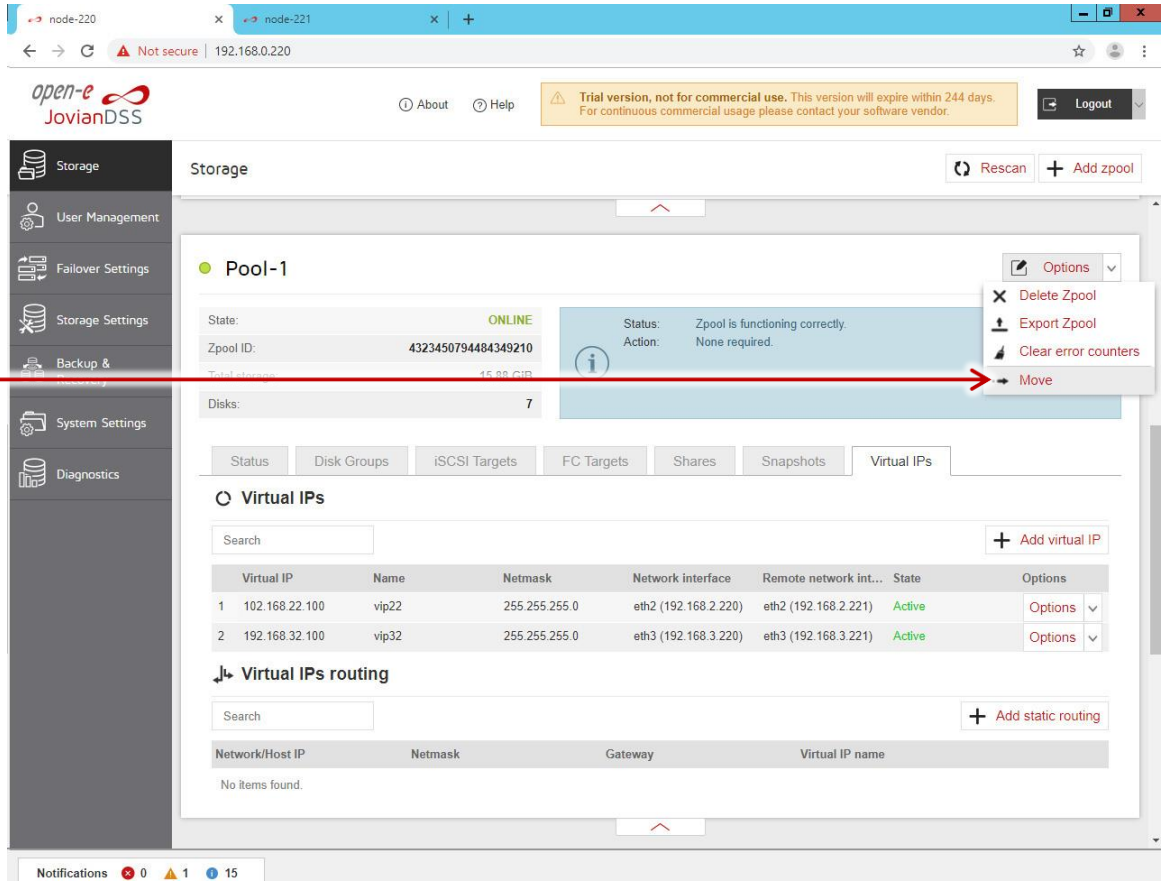
12. Failover test



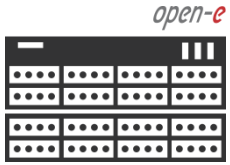
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Now, in order to test failover, select **Storage** from the main menu and in the **Options** drop-down menu, select **Move**.

The pool will be exported on the current node and will be imported on the second node.

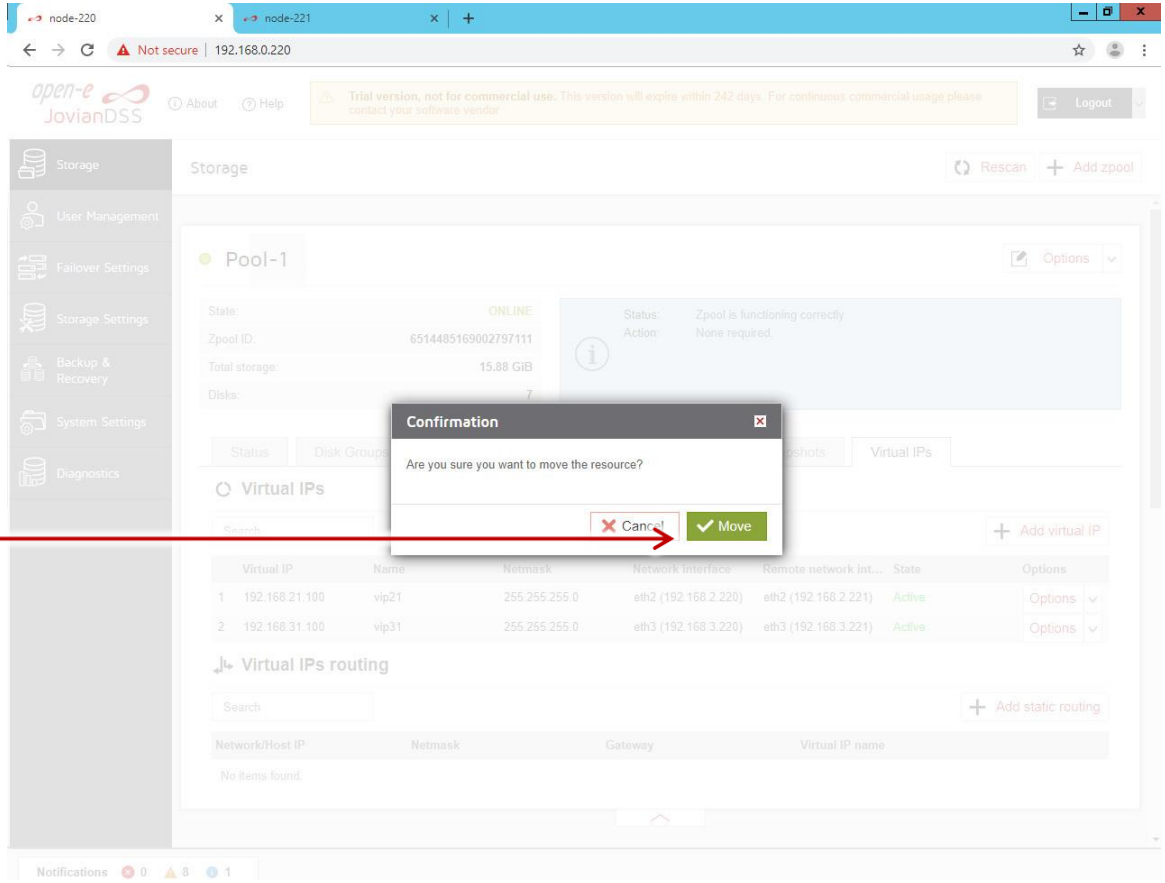


12. Failover test

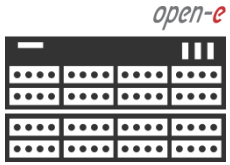


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Click the **Move** button to start the failover.



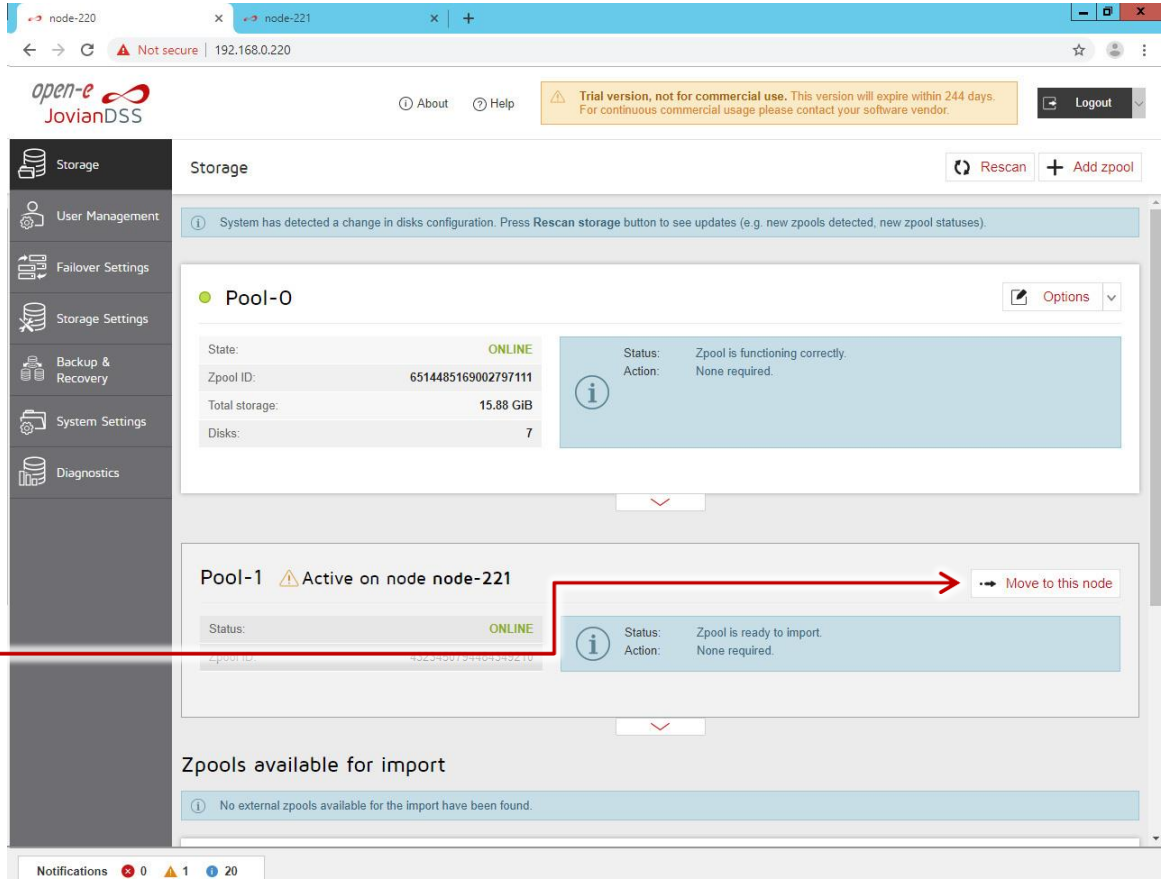
12. Failover test



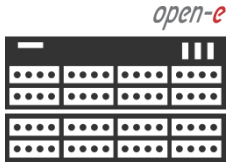
Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

The **Pool-1** was exported from Node-220 and imported on Node-221. Node-220 (node-a) GUI will show the **Pool-1** is active on node-221 (node-b).

In order to move pool activity back to node-220 (node-a), click the **Move to this node** button.

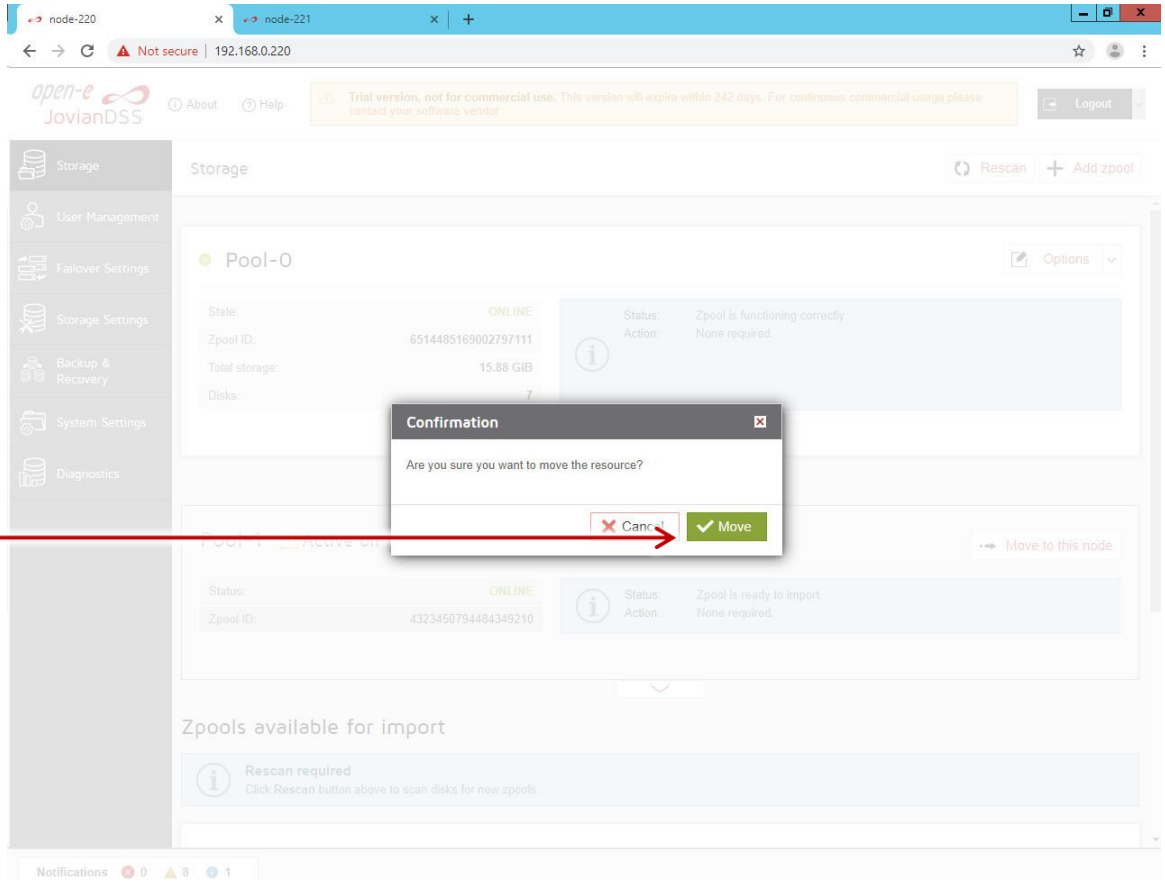


12. Failover test

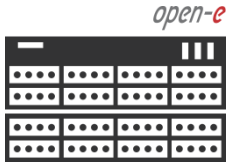


Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

Click the **Move** button to start failover. It will start exporting the pool on Node-221 (**node-b**) and next it will be importing back to Node-220 (**node-a**).

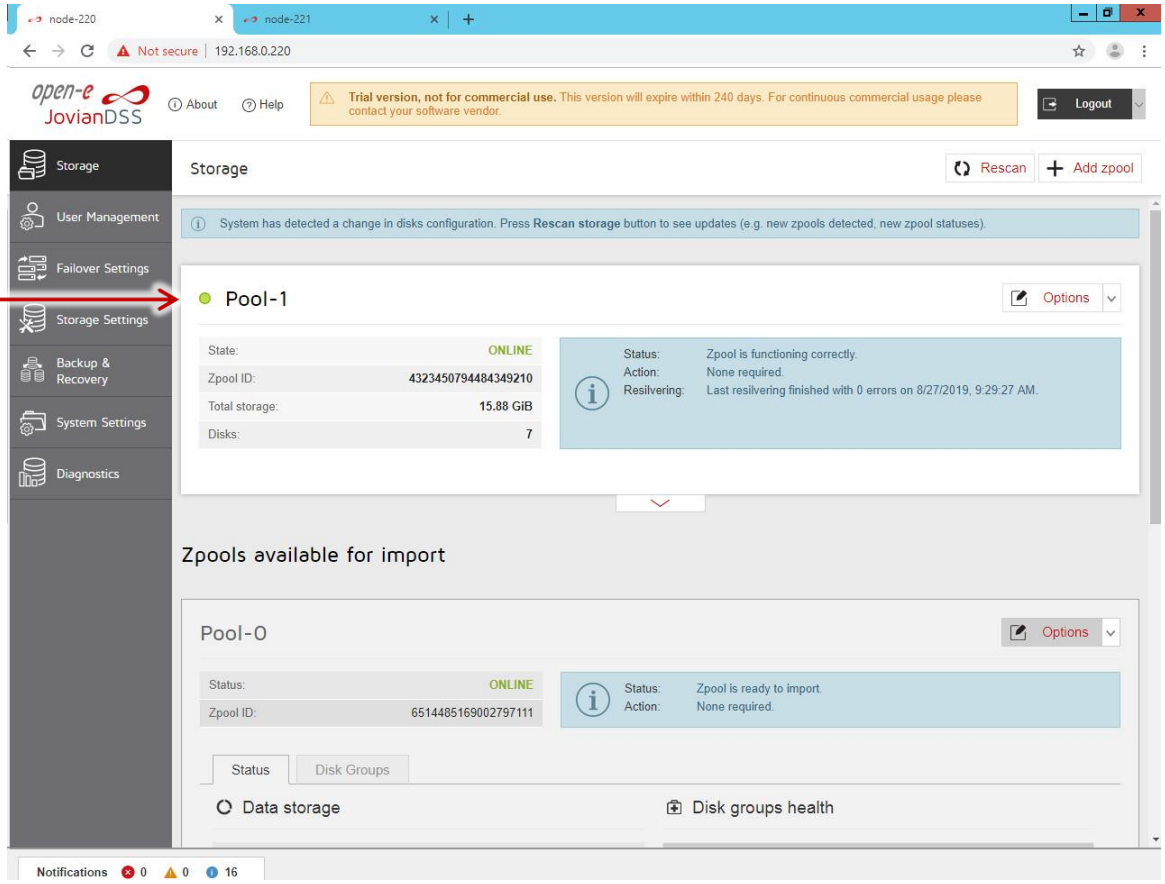


12. Failover test



Open-E JovianDSS: **node-a**
IP Address: 192.168.0.220

The failover test is completed. The **Pool-1** is active back on Node-220 (node-a)
Now, create an iSCSI target or NFS, SMB shares and connect storage clients to either of them. Once storage clients are connected, run one more failover test with a reboot of the first node and next after a successful failover, with reboot of the second node.



NOTE:

The step-by-step guide is based on a configuration from **page 5**. It uses two storage access paths and two virtual IPs per pool. This setup can be used for iSCSI with a multipath for non-zero-point-of-failure cluster.

There are plenty of possible configurations. Next examples are shown on **page 46 and 52**.

On **page 46** instead of two storage paths, there is a single bond. This setup can be used for NFS or SMB as a bond, assuring redundancy on the storage path. This setup cannot be used for iSCSI, as iSCSI requires two storage paths for redundant iSCSI multipaths.

On **page 47** instead of just two storage paths, there are two bonds. This setup can be used also for iSCSI Initiators with multipath or for mixed iSCSI/SMB/NFS environments.

Both configurations on **page 46 and 47** have redundant mirroring path. It uses a point-to-point Active-Backup bond instead of a single Ethernet connection. The Round-Robin bond cannot be used if the ring 2 is also configured over the mirror path.

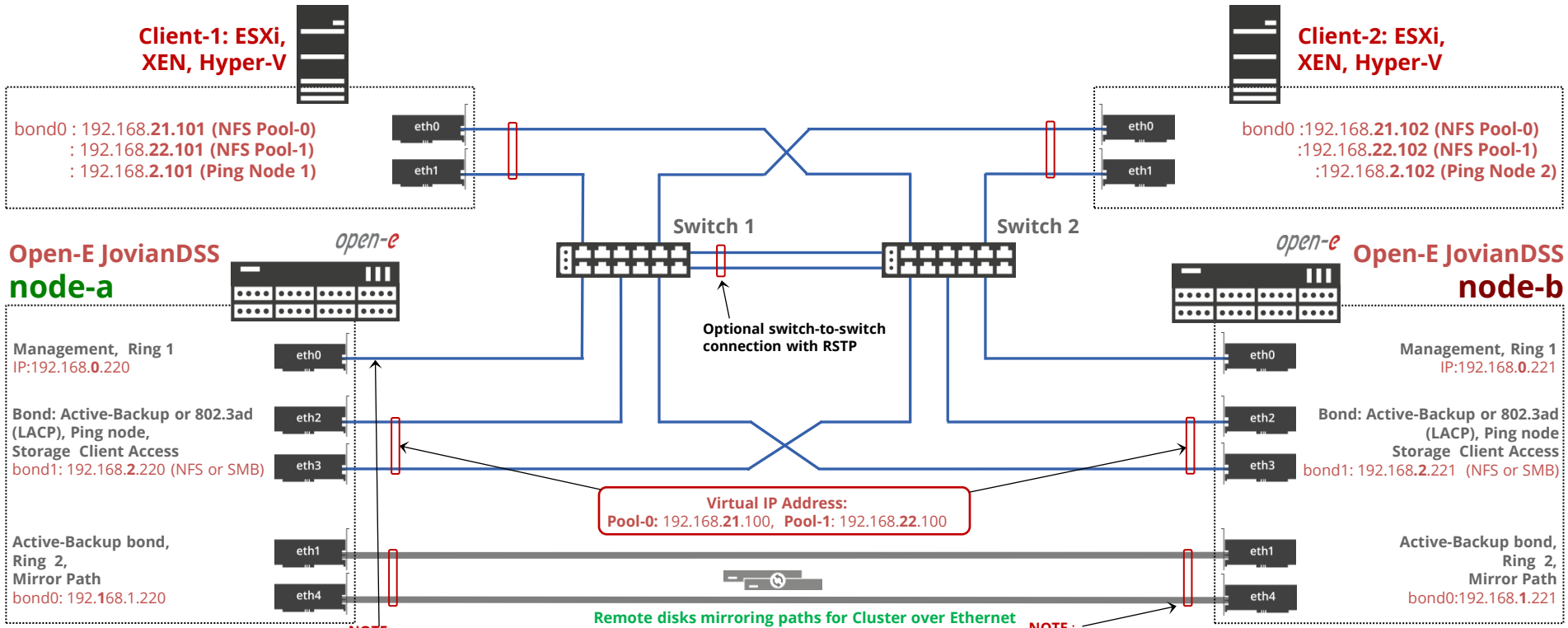
If mirror path is configured over a Round-Robin bond for better mirror performance, the second ring require extra dedicated path like shown on the configuration example on **page 48.** Here, the second ring path is configured over eth5-eth5 point-to-point path. It is strongly recommended to configure the second ring via point-to-point path as this will be switch failure independent and will not need both storage nodes to reboot in case of all Ethernet networks are down.

On **page 49** another example with 4 NICs only in every storage node. Here, the single point-to-point path is used for mirror path and second ring, 2 lines for iSCSI multipath and the first NIC pair is used for first ring, web management, and also for On- & Off-site Data Protection to another Open-E JovianDSS (not shown on the chart).

On **page 50..52** sites are connected via **limited number of connections**. On **page 50** connection between switches and direct point-to-point mirror path is available. On **page 51 and 52** only switch-to-switch connection is available. In such case, **ping nodes must be configured on one site only** so cluster will not force to import pools and split in case of lost connections between both sites. Such situation can happen if a switch failed or switch port used for site-to-site connection failed, or the cable was removed or damaged.

If the setup shown on **page 51 or 52** is used and **ping nodes are configured on both sites**, and connection between both sites is lost, both storage cluster nodes are obviously split. On both sites all pools will be imported in the degraded mode and cluster status on both nodes will be “separated”. The mirror path will show the “disconnected” status. This will be as long as the disconnected cluster status shows the “separated” mode. The reason for this behavior is to prevent auto-rejoin after restoring the connection between both sites. The re-join is impossible and the administrator must detach the lost (UNAVAILABLE) disks from mirrors on proper pools, delete wrong pools on the lost and detached disks. The cluster nodes must be disconnected using the “Disconnect” button in the GUI. After disconnect, the cluster must be re-configured, plus new and empty disks (after wrong pools are deleted) must be attached in order to re-mirror all data (GUI will show resilver running). It is required to restore the cluster after this split, **THIS KIND OF SETUP (no direct-point-to-point mirror path and switch-to-switch connection only, and ping nodes configured on both sites) IS STRONGLY NOT RECOMMENDED AND MUST BE AVOIDED.** This is why the setup on **page 51 and 52** shows ping nodes configured on one site only!

Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for NAS (NFS, SMB) storage clients.

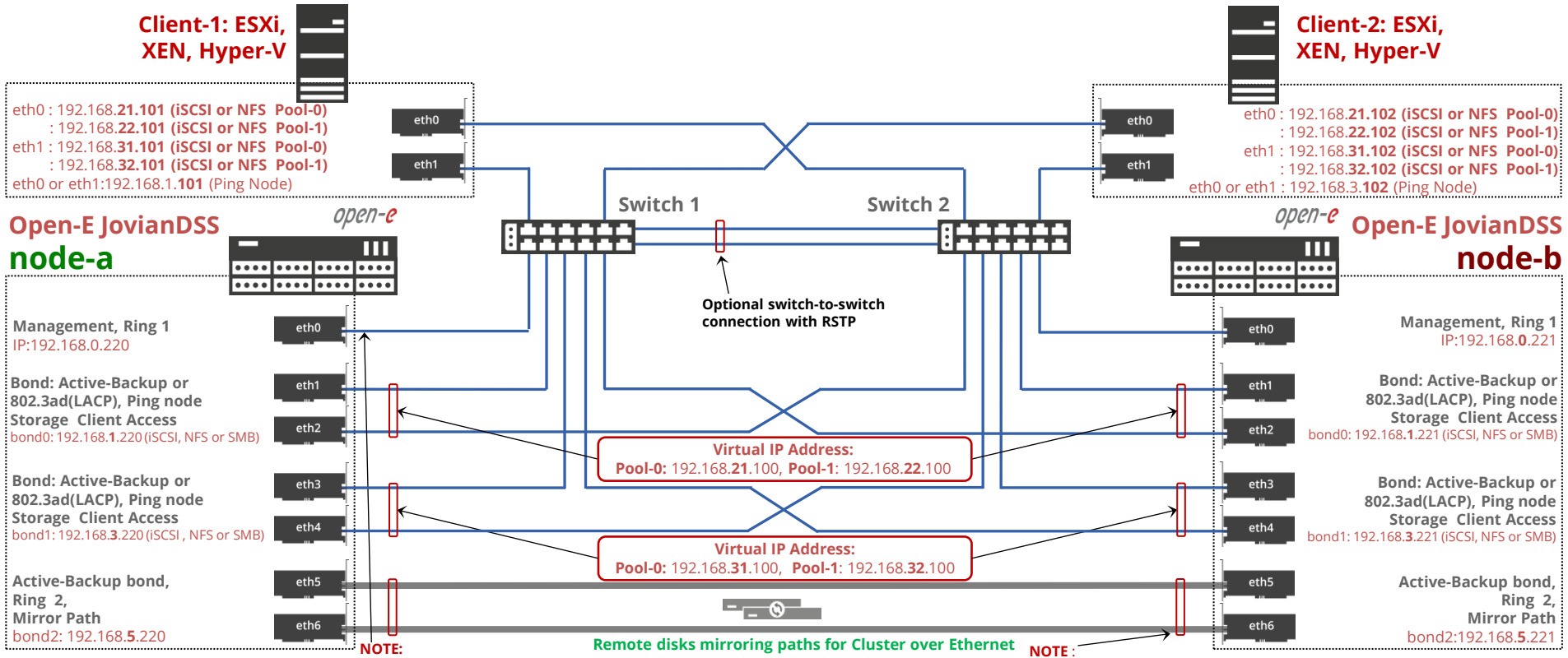


NOTE:
SMB can be used by other network clients but not for hypervisors.

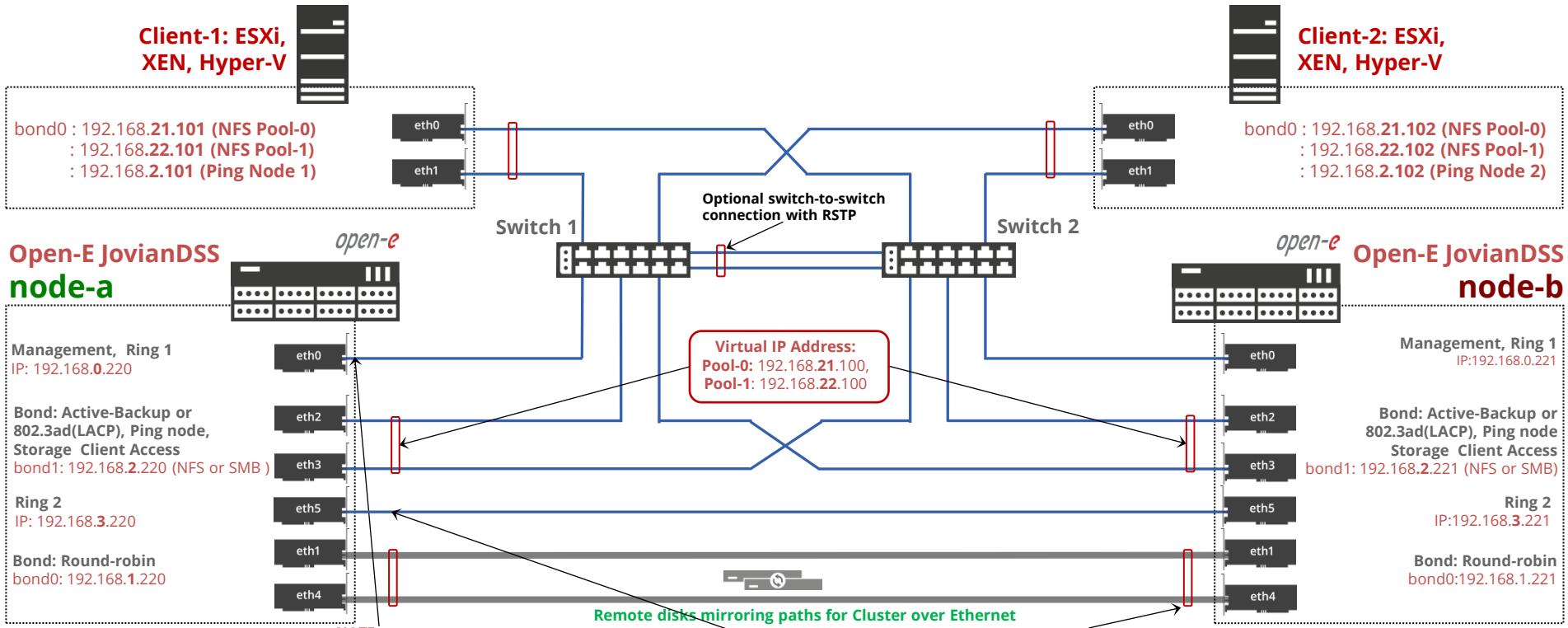
NOTE:
Cluster bind, ring path (heartbeat) and Ping nodes do not require to run on Active-Backup bonding if second Ring is configured. Ping IP can be configured in the storage network path. This is a NEW configuration option since version 1.0 up28.

NOTE :
It is strongly recommended to use direct point-to-point and, if possible, 10Gb connection for the remote disks paths. The Remote Disks connection can work over the switch but the most reliable is a direct connection.

Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for NAS (NFS, SMB) and iSCSI storage clients.



Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for NAS (NFS, SMB) storage clients.



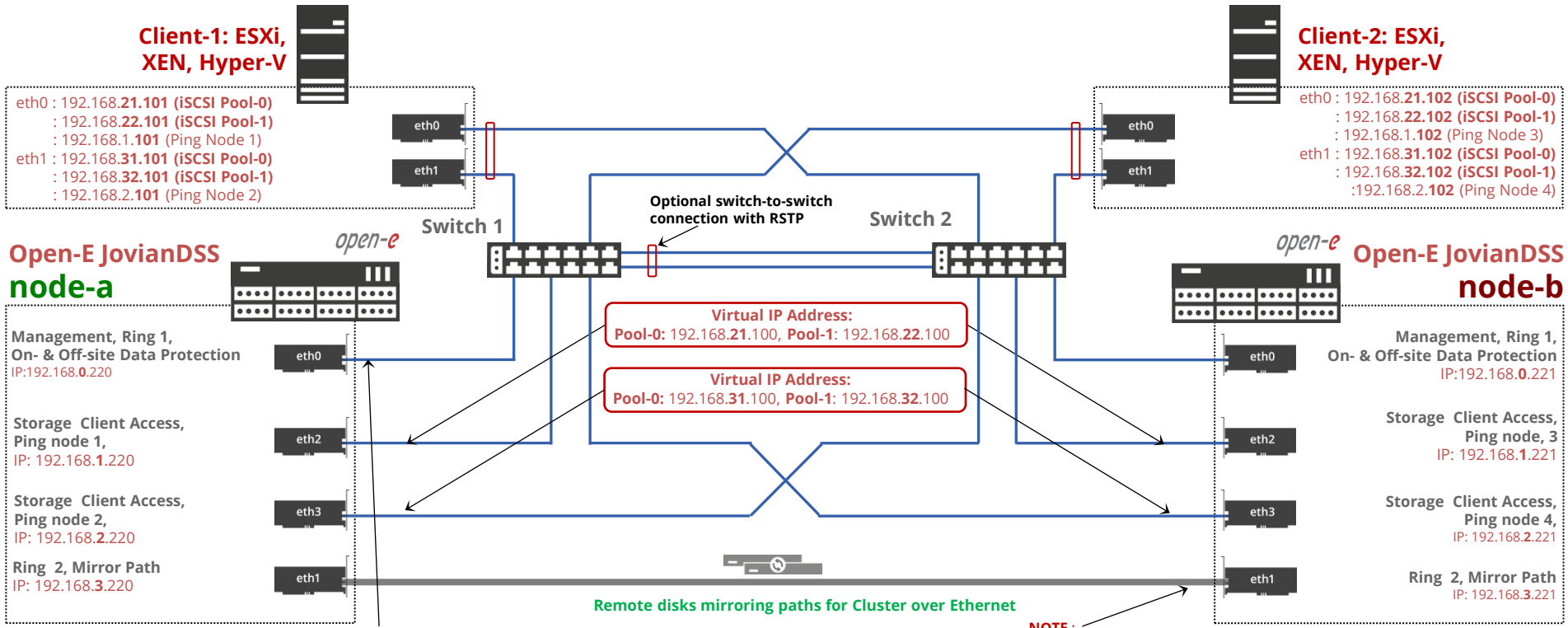
NOTE:
SMB can be used by other network clients but not for hypervisors.

NOTE:
Cluster bind, ring path (heartbeat) and Ping nodes do not require to run on Active-Backup bonding if second Ring is configured. Ping IP can be configured in the storage network path. This is a NEW configuration option since version 1.0 up28.

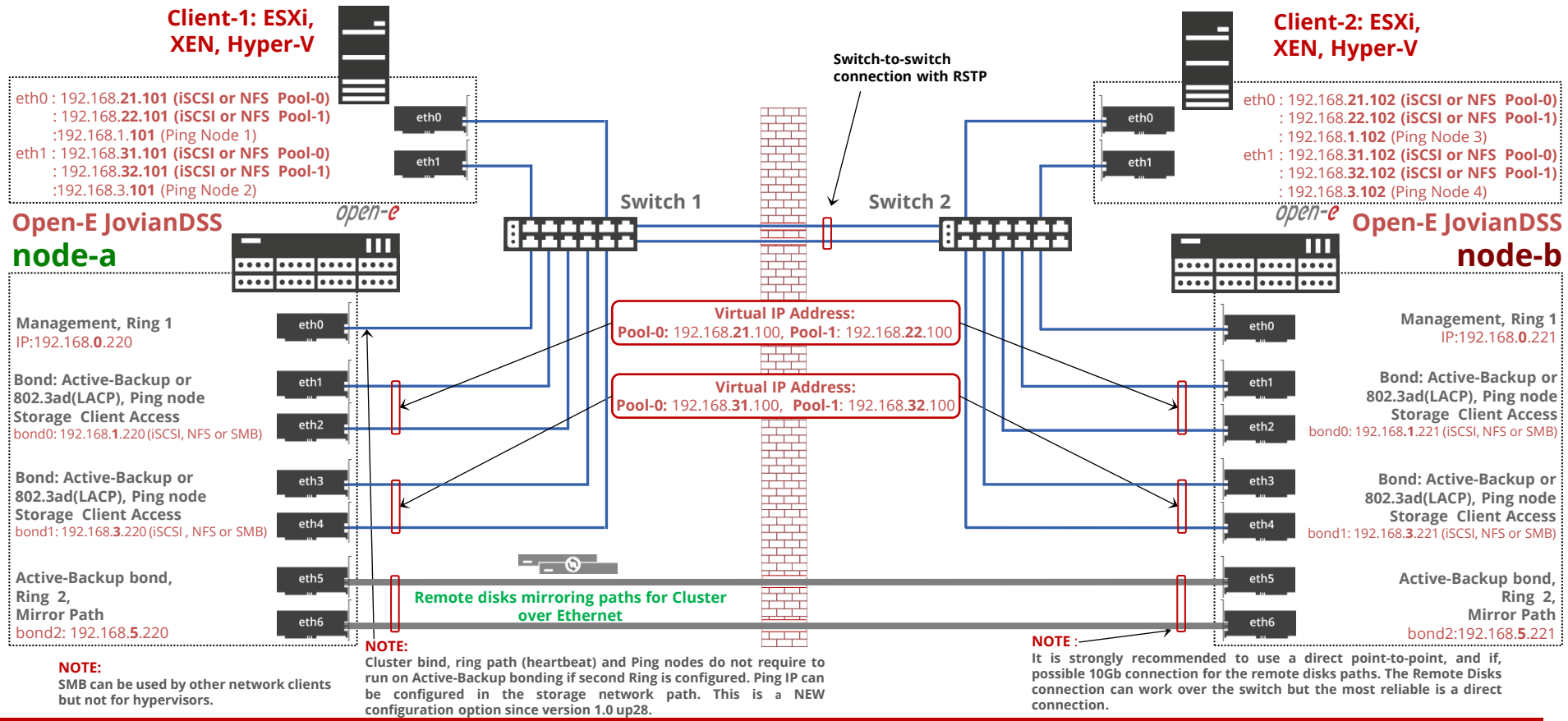
NOTE:
The ring2 can be configured on a dedicated point-to-point connection and not on the mirror-path. In such case, the mirror path can use a Round-Robin bond type.

NOTE :
It is strongly recommended to use a direct point-to-point and, if possible, 10Gb connection for the remote disks paths. The Remote Disks connection can work over the switch but the most reliable is a direct connection.

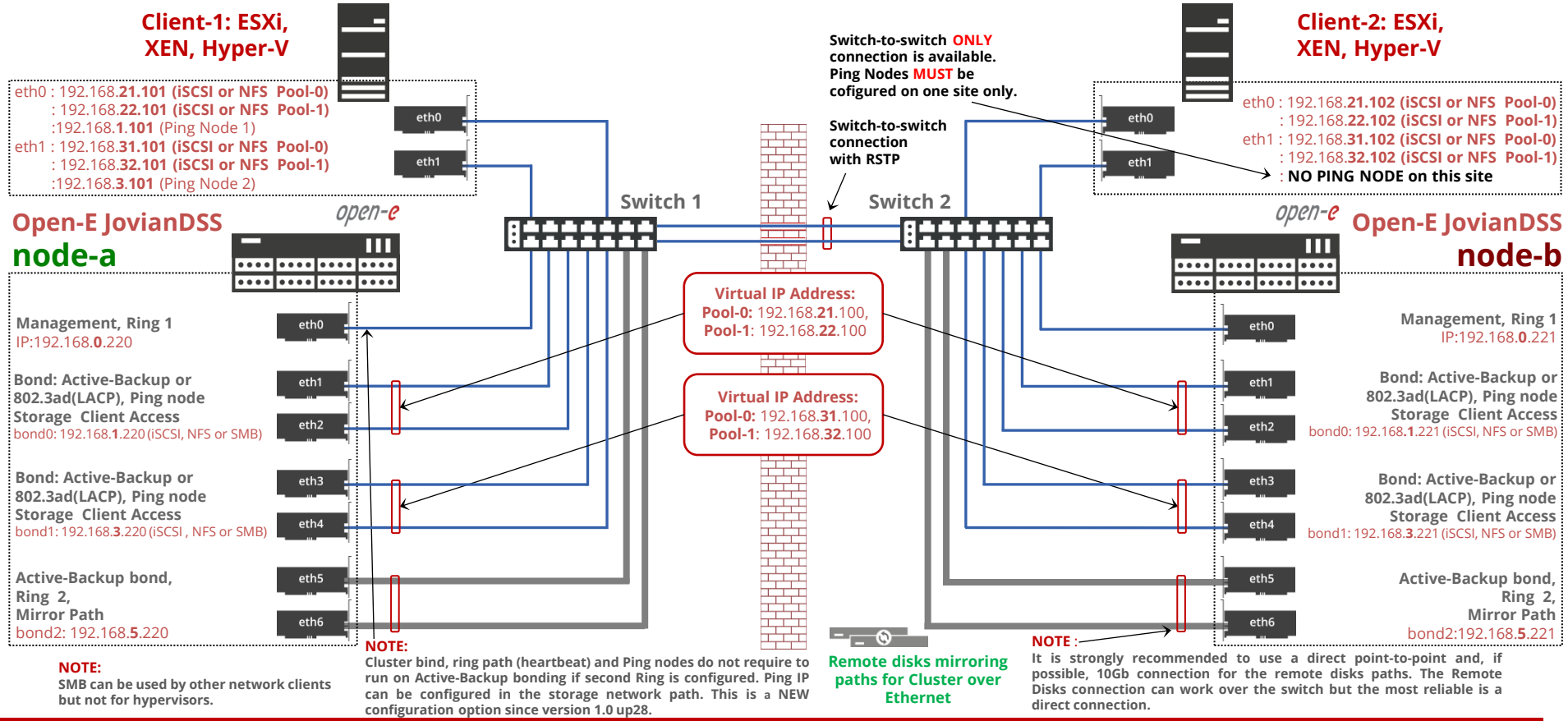
Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for iSCSI storage clients.



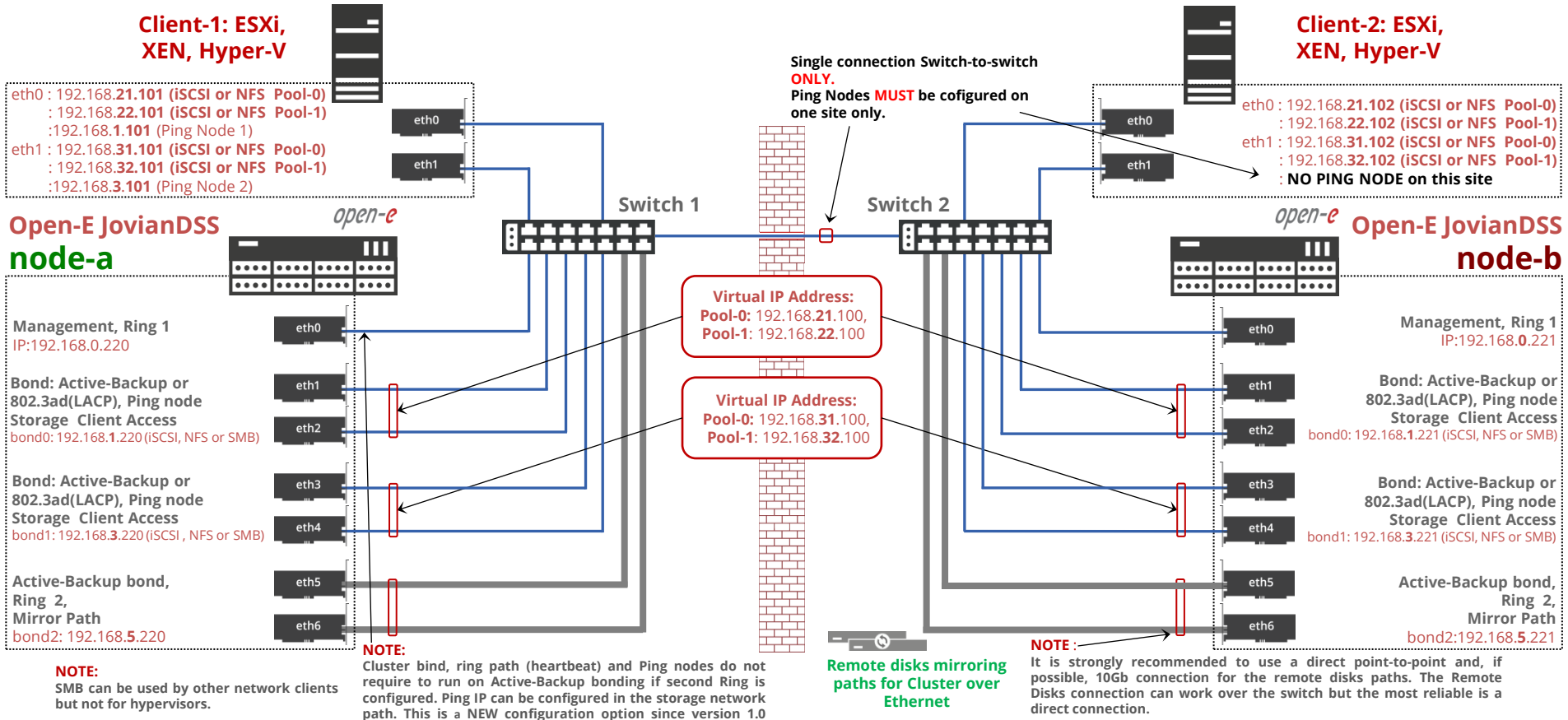
Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for NAS (NFS, SMB) and iSCSI storage clients.



Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for NAS (NFS, SMB) and iSCSI storage clients.



Open-E JovianDSS Advanced Metro High Availability Cluster with 2 Rings and bonded storage path for NAS (NFS, SMB) and iSCSI storage clients.



Copyright
(c) 2004-2020 Open-E, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written consent of Open-E, Inc.

Trademarks
The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

Disclaimer
Open-E, Inc. assumes no responsibility for errors or omissions in this document, and Open-E, Inc. does not make any commitment to update the information contained herein.

Thank You!