



Quick Start Guide

Version1.5



Revision History

Date	Version	Description
10-10-2016	1.0	Initial version
19-10-2016	1.1	Update the initial Cluster Management password
16-11-2016	1.2	Update screen shots related to replace management node feature
01-06-2017	1.3	Add Time Setup and Cluster NIC Settings screen shots
29-10-2017	1.4	Add the cluster tuning page and update deployment wizard and dashboard screenshots
08-01-2018	1.5	Update node services and dashboard screen shots
16-12-2018	1.6	Update screen shots based on Release 2.2



Contents

1. Purpose	3
2. Pre-requisites	3
3. PetaSAN Overview	3
4. Planning the cluster network.....	4
5. System Installation.....	6
6. Node Deployment.....	9
6.1 Introduction	9
6.2 Deploying our first node	9
6.3 Deploying our second node	14
6.4 Deploying our third node	15
6.5 Deploying our fourth node.....	18
7. Cluster Management	19
7.1 Configuring the cluster.....	19
7.2 Creating disks	22
7.3 Storage Over-Commit	23
8. Client connection	24
9. Command line access (for power users)	28



1. Purpose

The purpose of this guide is to quickly get you up and running using PetaSAN. It is recommended to be used as the first introductory guide. It will go through the main stages, starting from cluster planning and ending with disk creation and usage from client machines. For the sake of clarity, it will leave many of the finer details to other documentation guides.

2. Pre-requisites

The demonstration cluster in this guide requires 4 machines with 2 disks and 2 network interfaces each. For demo purposes, low end hardware or running within a virtualized environment will suffice. For production requirements, please refer to the *PetaSAN Recommended Hardware Guide*. In addition, the example client presented requires the use of a separate machine running Windows.

3. PetaSAN Overview

PetaSAN is an open source Scale-Out SAN solution. Scale-Out means we can scale up the system by adding more machines, aka storage nodes, to our cluster. The scaling is for both capacity as well as performance.

PetaSAN is designed from the ground up to do one thing: provide highly available clustered iSCSI disks. In PetaSAN, an iSCSI disk can have many access paths, each identified by its virtual IP address. These IP addresses are clustered across several storage nodes. Access to the same disk is load balanced across different nodes in a symmetric Active/Active fashion. Moreover these IP addresses are not tied to particular hosts, but rather are virtual IPs that can be dynamically moved from node to node. If a node fails, its virtual IPs will be re-assigned to other nodes thus providing high availability on a path level, this is all done transparently to the client.

A PetaSAN iSCSI disk is mapped to all physical disks in the cluster, concurrent access to the disk will execute in parallel across all physical disks in the cluster, this is beneficial for applications with high concurrency requirements. For example in a clustered hypervisor setup where each node is housing many VMs and each VM is running multiple applications like database servers with concurrent transactions. There will be many simultaneous I/O threads accessing the same disk.

Internally PetaSAN uses LIO (www.linux-iscsi.org) for its iSCSI Target server, Ceph (www.ceph.com) for its storage engine and consul (www.consul.io) for cloud scale distributed resource management.

PetaSAN attempts to make the life of the system administrator much simpler by providing point and click management, yet it also allows the power user full command line access if desired.



4. Planning the cluster network

In PetaSAN, there are 5 different subnets that need to be setup:

Management Subnet: This subnet carries management traffic. It is used by the PetaSAN web applications to manage the different cluster nodes as well as by the nodes to communicate among themselves for management purposes. This is the only subnet configured with a gateway and dns to enable communication with other networks. The administrator needs to assign a fixed static IP for each node on this subnet.

iSCSI 1 Subnet: This is the first of 2 subnets dedicated to iSCSI traffic, it is here where iSCSI client initiators communicate with their iSCSI Target disks. An iSCSI Target disk can have multiple paths; each path is accessed by a virtual IP within one of the 2 iSCSI subnets. The virtual IPs are assigned to the different nodes dynamically by the PetaSAN distributed resource management system. If a node fails, its virtual IPs will be transferred equally to other functioning nodes. The administrator does not directly assign virtual IPs to nodes.

iSCSI 2 Subnet: This is the second iSCSI subnet.

Backend 1 Subnet: This is a subnet used for backend communication within the PetaSAN system. The administrator needs to assign a fixed static IP for each node on this subnet. It is not required to understand in detail what this subnet does in order to set up the PetaSAN system and get it up and running, but for the curious, it is used by LIO iSCSI Target layer to communicate with the Ceph storage engine.

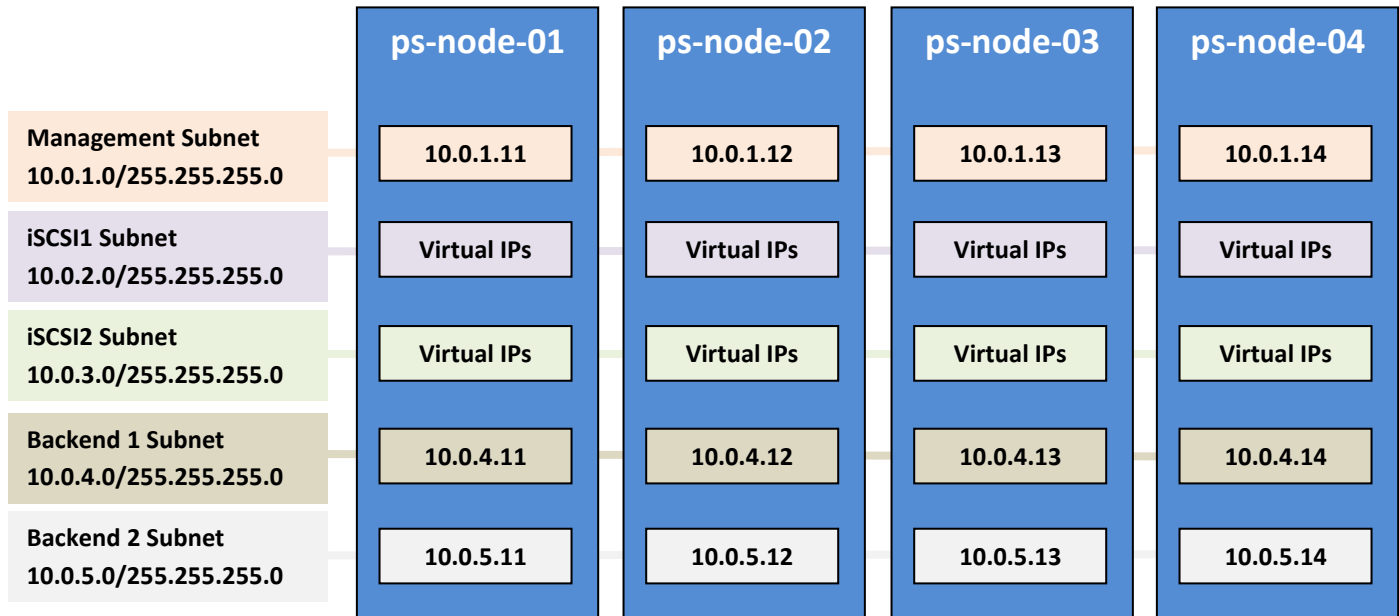
Backend 2 Subnet: This is the second subnet used for backend communication within the PetaSAN system. The administrator needs to assign a fixed static IP for each node on this subnet. It is not required to understand in detail what this subnet, but for the curious, it is used by the Ceph storage engine to replicate and move data among the different storage nodes.

PetaSAN requires the use of 2 to 4 physical network interfaces:

- For 4 interfaces, it is recommended to combine the Management and iSCSI1 subnets onto a single interface, the remaining subnets are segregated on separate interfaces. This setup provides the best performance.
- For 3 interfaces, it is recommended to combine the Management and iSCSI1 subnets onto a single interface, leave iSCSI2 on an interface by itself and combine the Backend interfaces on another.
- For 2 interfaces, this is the minimum requirement. It is recommended to combine the Management, iSCSI1 and Backend 1 on one interface, iSCSI2 and Backend 2 on the other.

➤ *Note: In PetaSAN, the number of network interfaces used and their mappings to subnets is the same for every cluster node and is defined at cluster creation time.*

For our 4 node demo setup, the different subnet definitions and node IP assignment will be as follows:



- *Note: in this demo setup, the chosen subnet mask will yield 255 different addresses per subnet. For larger deployments a more suitable mask may be needed to provide a larger address range, this is especially true for the iSCSI subnets.*

We will be using the minimum of 2 network interfaces for this demo; hence the mapping of subnets to physical interfaces will be as follows:

Subnet	Interface
Management	eth0
iSCSI1	eth0
iSCSI2	eth1
Backend 1	eth0
Backend 2	eth1

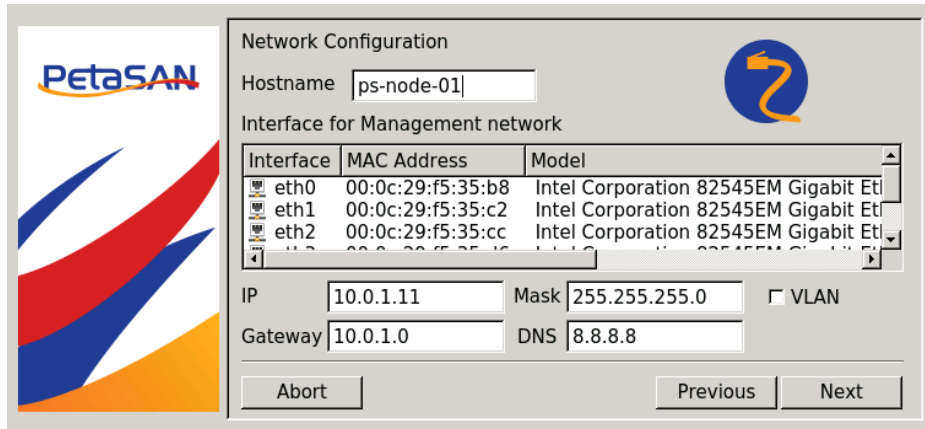
- *Note: PetaSAN is based on Ubuntu Linux hence Ethernet network interfaces are named as ethX where X is a zero based index of the interface device. eth0 means the first Ethernet interface, eth1 the second.*
- *Note: If trying PetaSAN in a virtual desktop environment, make sure mapping different subnets to a single interface is supported; else create a virtual interface for each subnet.*

5. System Installation

The installation iso can be burned on CD or burned to USB using widely available USB tools such as rufus www.rufus.org.

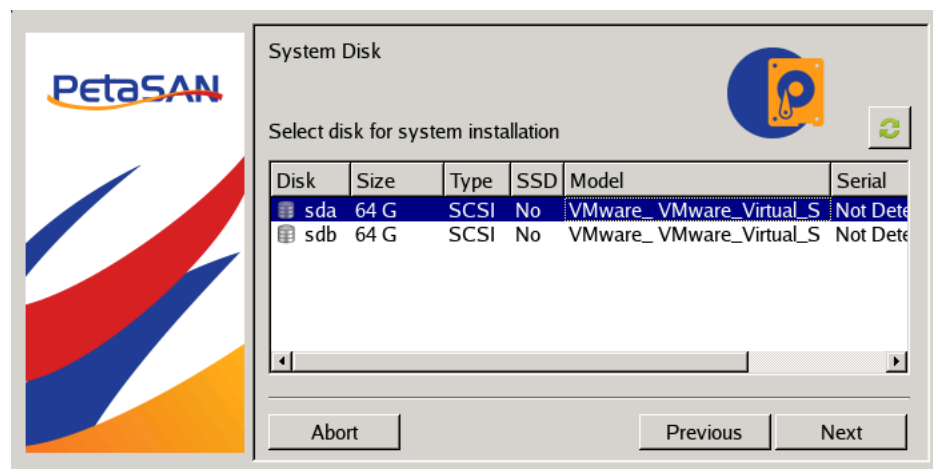
There are 3 main settings to perform during the installer, configuring the Management Network, choosing the installation disk and setting up the system time.

Network configuration



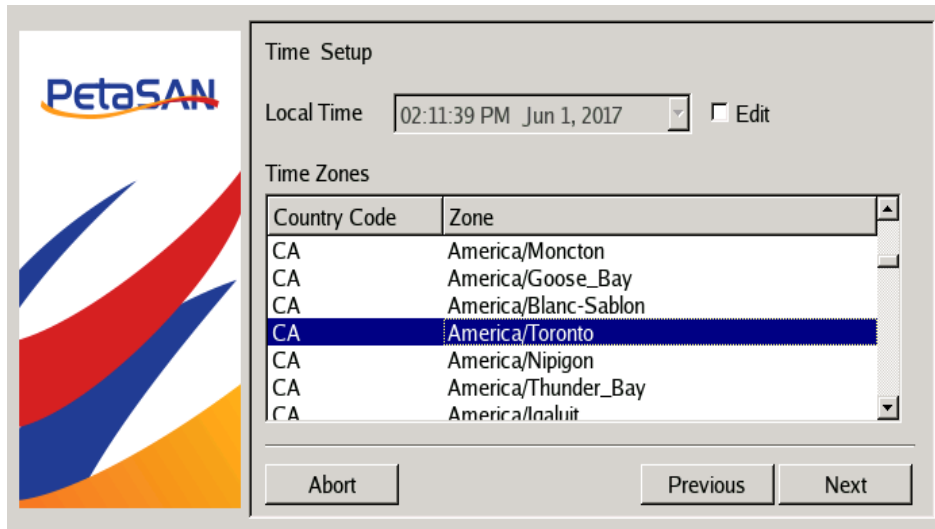
- *Note: PetaSAN does not rely on having an external DNS or gateway to function properly. Having access to an external network is useful when performing time syncs with an external time server and could be used to download/upgrade packages.*

System Disk selection

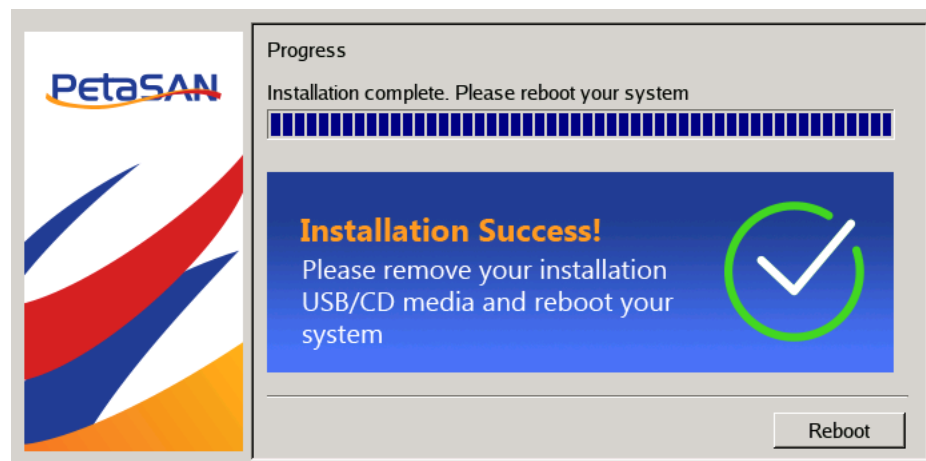


Time Setup

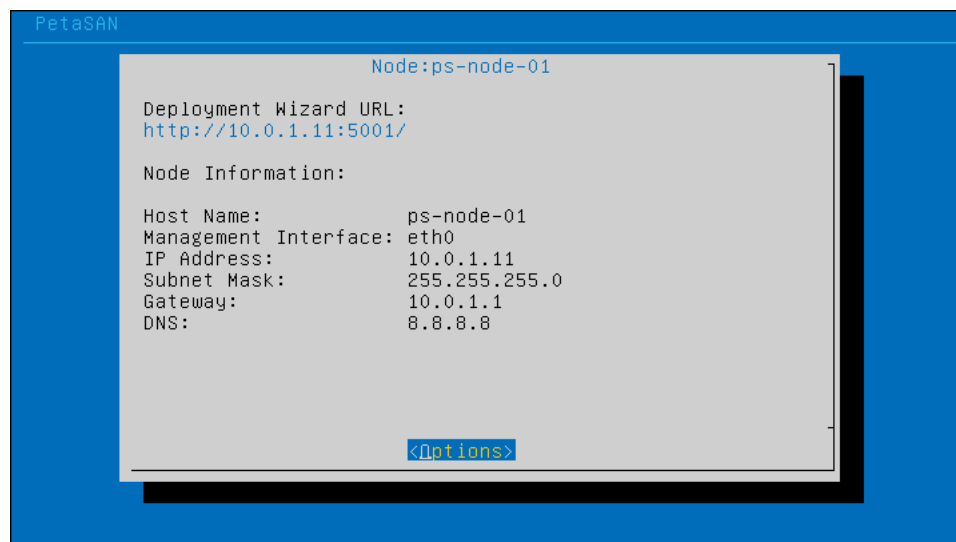
Adjusting the machine time and time zone



Upon successful completion, we need to remove the install media before rebooting



On reboot a console will be display node data as well list the url for the Deployment Wizard



We need to repeat the above steps for installing the remaining nodes:

	First node	Second node	Third node	Fourth node
Hostname	ps-node-01	ps-node-02	ps-node-03	ps-node-04
Management Interface	eth0	eth0	eth0	eth0
IP Address	10.0.1.11	10.0.1.12	10.0.1.13	10.0.1.14
Mask	255.255.255.0	255.255.255.0	255.255.255.0	255.255.255.0
Gateway	10.0.1.1	10.0.1.1	10.0.1.1	10.0.1.1
DNS	8.8.8.8	8.8.8.8	8.8.8.8	8.8.8.8
System disk	sda	sda	sda	sda



6. Node Deployment

6.1 Introduction

Once we are done with the installer, each node sits in isolation relative to other nodes and needs to be instructed what to do next, this step is referred to as “deploying” the node and is accomplished by using the dedicated web based Deployment Wizard. There are 2 initial choices for deploying a node, either make it a first seed node to form a new cluster or make it join an existing one. A PetaSAN cluster needs at least 3 nodes to be up and running with further nodes being added/joined as needed for scaling out the cluster.

Each node in the cluster can be configured to run the Local Storage Service and/or the iSCSI Target Service. The Local Storage Service serves all local physical disks, with the exception of the system disk, to the PetaSAN cluster storage pool; the iSCSI Target Service enables the node to serve the user created iSCSI disks to connecting client initiators. By default both services are checked/enabled in the Deployment Wizard but for heavy workloads it may be better to setup dedicated nodes. Please refer to the *PetaSAN Recommended Hardware Guide* for the recommended hardware requirements for each service.

In addition to the 2 above services, the first 3 nodes run additional Management and Monitoring services. This is fixed and cannot be changed by the user. The first 3 nodes in a PetaSAN cluster are privileged nodes that carry the brain of the cluster; the cluster is alive and functioning upon deployment of the third node. Because of their management functions, these nodes are referred to as Management nodes. A cluster needs to have at least 2 of the 3 Management nodes up and running else the entire cluster will not function.

6.2 Deploying our first node

The web based Deployment Wizard for a node can be accessed via the following url:

<http://management-ip-address:5001>

so for our first node, the url is:

<http://10.0.1.11:5001>

➤ *Note: The url is also displayed on the console screen of each node*



The Deployment Wizard is comprised of several steps, in the first step we need to choose between creating a new cluster or joining an existing one, in our case we need to create a new one.

1 Step 1 2 Step 2 3 Step 3 4 Step 4 5 Step 5 6 Step 6 7 Step 7 8 Step

Wizard Start

☒ Create New Cluster
☐ Join Existing Cluster
☐ Replace Management Node

Next

We then need to enter the cluster name and a cluster password

1 Step 1 2 Step 2 3 Step 3 4 Step 4 5 Step 5 6 Step 6 7 Step 7 8 Step

Cluster Identity

Cluster Name:*
Demo

Password:*

Confirm Password:*

Next



The next step is to define the Cluster NIC settings for the new cluster, you can select to use Jumbo Frames and/or NIC Bonding, for this demo purpose we will just use the default values

The next step is to define the network settings for the new cluster as per the earlier explanation in *Planning the cluster network* section.

For convenience the relevant values are as follows:

iSCSI1 Interface:	eth0
iSCSI2 Interface:	eth1
Backend 1 Interface:	eth0
Backend 1 Base IP / Mask:	10.0.4.0 / 255.255.255.0
Backend 2 Interface:	eth1
Backend 2 Base IP / Mask:	10.0.5.0 / 255.255.255.0

- *Note: When using the IP address control to input values, you can use the horizontal arrow keys to move between digits and the space bar to jump to the next octet field. Ctrl-A to select all digits.*

It is important to mention that for the iSCSI subnets, even though we define their interface mappings here, we do not define their addresses. As we well see, this is done later after the cluster is up and running and can be changed at any time during normal operation. The interface mappings and the backend network settings are defined at cluster creation time and cannot change. Note also that the Management Subnet is set earlier within the installer and its interface mapping is shown here for reference only.



The next step is Cluster Tuning

In this step select your cluster size depending on the number of disks you have (or anticipate to have) in your cluster. You also need to choose the number of replicas, which is the number of copies (including original) you would like to create when storing your data

You can also select the Tuning Template that best matches your cluster hardware, for the demo purpose we will select the Default template.

The screenshot shows a web browser window with the address bar displaying '10.0.1.11:5001'. The PetaSAN logo is in the top left corner. A progress bar at the top indicates eight steps, with 'Step 5' highlighted in blue. Below the progress bar, the title 'Cluster Tuning' is displayed. The main content area contains three dropdown menus: 'Cluster Size:*' with '15 to 50 disks' selected, 'Number of Replicas:*' with '2' selected, and 'Tuning Template:*' with 'Default' selected. The 'Tuning Template' dropdown list is expanded, showing options: 'Default', 'Generic Entry Level Hardware', 'Generic High End Hardware', 'Generic Mid Range Hardware', and 'Custom'. Below these dropdowns is a '+Show Details' link. A 'Next' button is located in the bottom right corner of the form area.



The next step is to define the backend IPs of the node itself

Backend 1 IP: 10.0.4.11

Backend 2 IP: 10.0.5.11

1 Step 1 2 Step 2 3 Step 3 4 Step 4 5 Step 5 6 Step 6 7 Step 7 8 Step

Node Node1 Network Settings

Node Backend 1 IP:
10.0.4.11

Node Backend 2 IP:
10.0.5.11

Next

The following step is to define what services the current node will run. We will keep the default selection checked for the Local Storage Service as well as the iSCSI Target Service. Note the Management and Monitoring Services is checked and cannot be changed for the first 3 cluster nodes.

By selecting the Local Storage Service you will be able to select the disks you want to add to PetaSAN. You can add disks as OSDs which are used by PetaSAN for actual data storage, or (advanced case) you can add them as journals (WAL/DB) on faster devices to speed up the operations of OSDs. If you have identical disks that you want to assign to PetaSAN, just add them as OSDs.

1 Step 1 2 Step 2 3 Step 3 4 Step 4 5 Step 5 6 Step 6 7 Step 7 8 Step 8

Node Node01 Services

☒ Management and Monitoring Services
☒ Local Storage Service
☒ iSCSI Target Service

Please select disks that will be formatted and assigned to the PetaSAN storage pool. Any existing data will be lost

	Name	Size	Type	SSD	Vendor	Model	Serial	Usage
<input checked="" type="checkbox"/>	sdb	111 GB	SATA	Yes		KINGSTON_SV300S37A120G	50026B725502CF48	<input type="radio"/> OSD <input checked="" type="radio"/> Journal
<input checked="" type="checkbox"/>	sdc	111 GB	SATA	Yes		KINGSTON_SV300S37A120G	50026B7262042DF5	<input checked="" type="radio"/> OSD <input type="radio"/> Journal

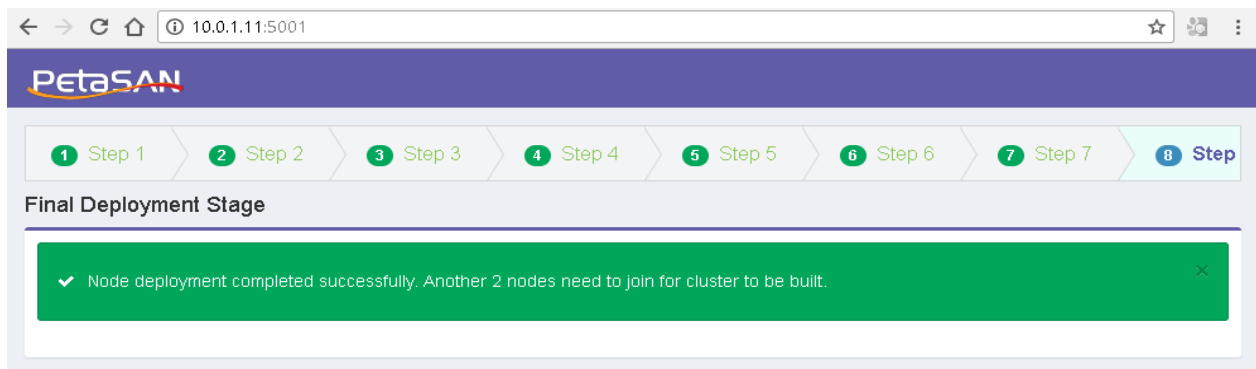
Showing 1 to 2 of 2 entries

Next

- *Note: You can skip adding disks now and do it later after building the cluster using the Management interface "Node List/Physical Disk List" but for the demo purpose we will add them here.*



On the final step, a message is displayed indicating the node deployment is complete, the new cluster metadata has been initialized successfully but the cluster will not come alive until 2 other nodes join in.

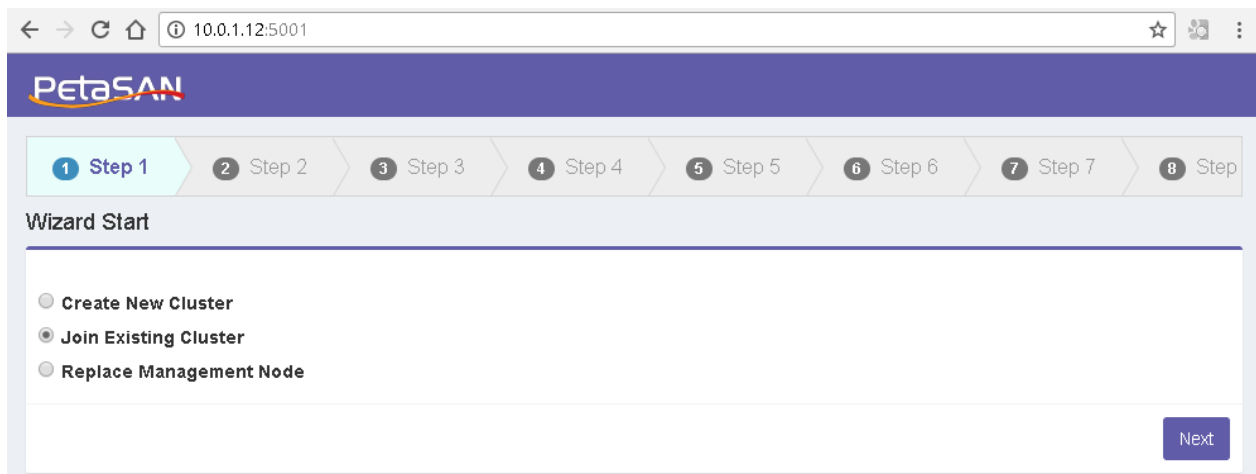


6.3 Deploying our second node

The Deployment Wizard for our second node is accessed from url:

<http://10.0.1.12:5001>

In our first step, we need to join the existing cluster:





Then we need to input the IP address of the first node, followed by the cluster password.

← → ↻ 🏠 ⓘ Not secure | 10.0.1.12:5001 ☆ 🖨 ⋮

PetaSAN

1 Step 1 2 Step 2 3 Step 3 4 Step 4 5 Step 5 6 Step 6 7 Step 7

Join Cluster

Management node IP to join:*

🖨 10.0.1.11

Cluster Password:*

.....

Next

The next step is to define the backend IPs of the node itself

Backend 1 IP: 10.0.4.12

Backend 2 IP: 10.0.5.12

Next for the services, leave the defaults as before.

Pressing next, a message will indicate successful node deployment, but the cluster is awaiting a third node to come alive.

6.4 Deploying our third node

This is similar to deployment of our previous node. The deployment url is

<http://10.0.1.13:5001>

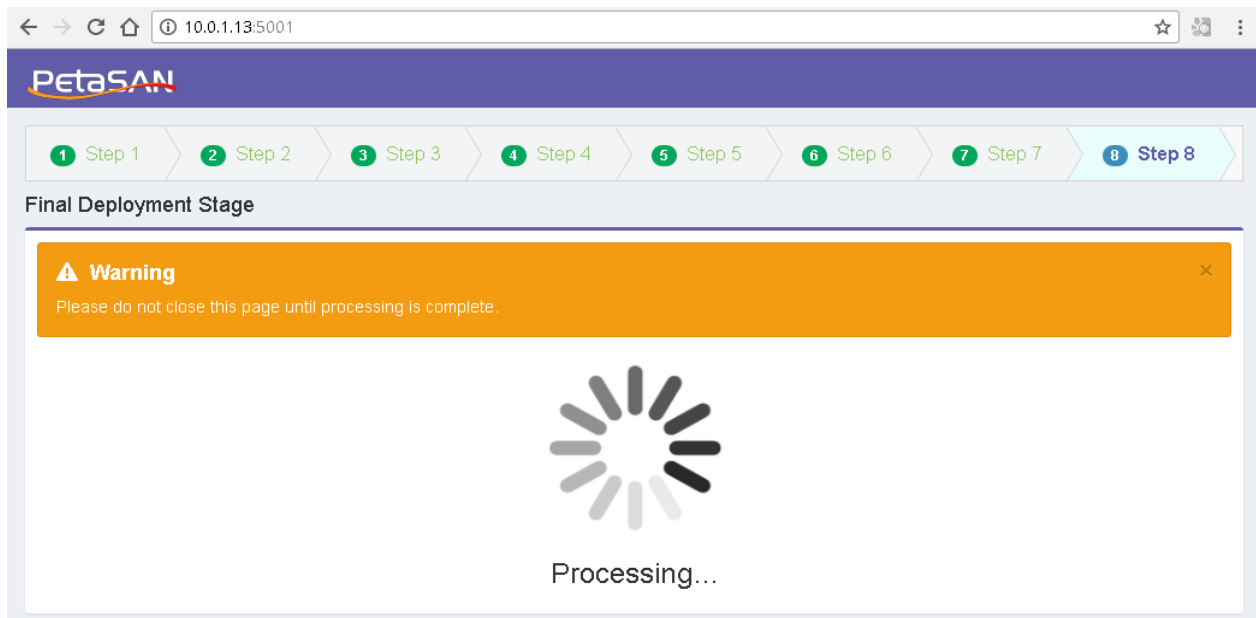
The backend IPs are:

Backend 1 IP: 10.0.4.13

Backend 2 IP: 10.0.5.13

For the services, leave the default as before.

All quite so far, but on the last step, the real action begins. The cluster node count reaches 3 and these Management nodes will start working with one another to bring the cluster alive. This will take several minutes, depending on how many disks you have. This is a good time to relax and get a cup of coffee.



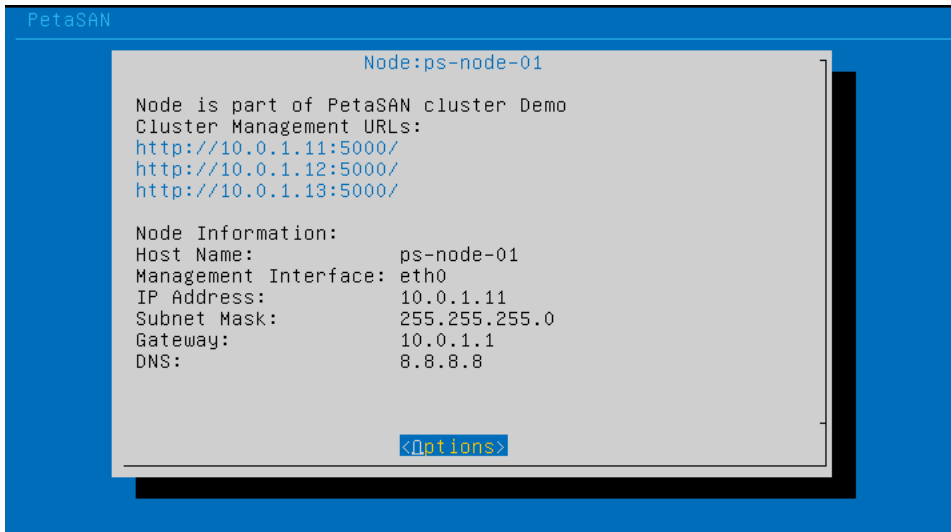
Upon successful completion, the following message will be displayed.

✓ Node deployment completed successfully. Cluster has been built and is now ready for use.

With links to the Cluster Management URLs

Congratulations! We have successfully built our first cluster.

The cluster is ready for use. The console menu for the cluster nodes will be updated to reflect the new urls for the web based Cluster Management application. Note that the cluster management urls are served by all 3 Management nodes at port 5000 and will be used to manage all nodes in the cluster; in contrast the node Deployment Wizard is at port 5001 and is specific to the node being deployed.



```
PetaSAN

Node:ps-node-01

Node is part of PetaSAN cluster Demo
Cluster Management URLs:
http://10.0.1.11:5000/
http://10.0.1.12:5000/
http://10.0.1.13:5000/

Node Information:
Host Name:      ps-node-01
Management Interface: eth0
IP Address:     10.0.1.11
Subnet Mask:    255.255.255.0
Gateway:       10.0.1.1
DNS:           8.8.8.8

<Options>
```

At this stage, we can go ahead and use the cluster now and add further nodes later, when we have a need to scale out. However in our case it may be simpler to keep with the flow and add a fourth node now before using our cluster.



6.5 Deploying our fourth node

Beyond the first 3 Management nodes, nodes are referred to as Storage nodes. They can be designated to run the Local Storage Service or the iSCSI Service or both (the default), but for heavy workloads it may be better to setup dedicated nodes for each service. Please refer to the *PetaSAN Recommended Hardware Guide* for the recommended hardware requirements for each service.

The deployment is similar to what we had done before. The deployment url is

<http://10.0.1.14:5001>

The backend IPs are:

Backend 1 IP: 10.0.4.14

Backend 2 IP: 10.0.5.14

For the services, leave the default as before.

The wizard will take several minutes to complete as it prepares the local disks and adds them to the cluster pool.

We are done with the deployment phase and without further due, let's begin using our newly created PetaSAN cluster.



7. Cluster Management

7.1 Configuring the cluster

The Management Application can be accessed at port 5000 of our Management nodes. In our case this would be

<http://10.0.1.11:5000>

Log in using the following credentials

user: **admin**
password: *password*

➤ *Note: You can change the initial password once logged in*

The image shows a login interface for PetaSAN. At the top, the PetaSAN logo is displayed. Below it, a white box contains the text "Sign in to start your session". There are two input fields: "User Name" with a user icon and "Password" with a lock icon. A blue "Sign In" button is located at the bottom right of the white box.

PetaSAN

Sign in to start your session

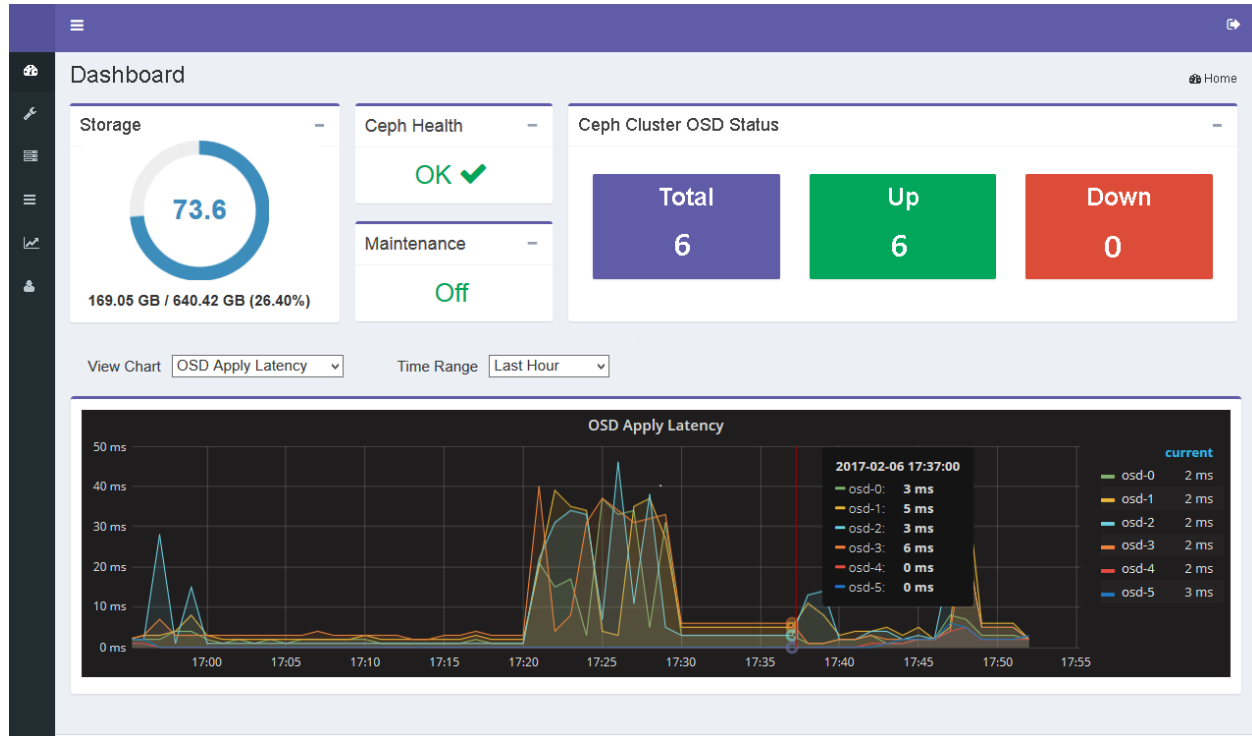
User Name

Password

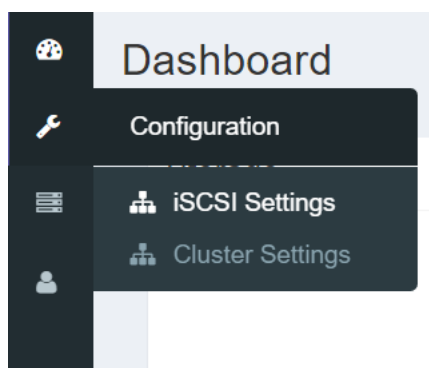
Sign In

Our home page is a dashboard showing the health of our storage engine which is based on Ceph. We can see all the OSDs we have added and they are all healthy. We also see the current IOPS and Read/Write bandwidth, which is not much since we do not have any activities right now.

We can also check if the cluster maintenance mode is “ON” or “OFF”.



Before we can start creating and using iSCSI disks, we need to configure the cluster iSCSI settings; this is accessed from the Configuration menu on the left side pane as shown:



In iSCSI settings configuration, enter the following values:

Iqn base prefix: iqn.2016-05.com.petasan

iSCSI1

Subnet Mask: 255.255.255.0

Auto IP From: 10.0.2.100

Auto IP To: 10.0.2.254

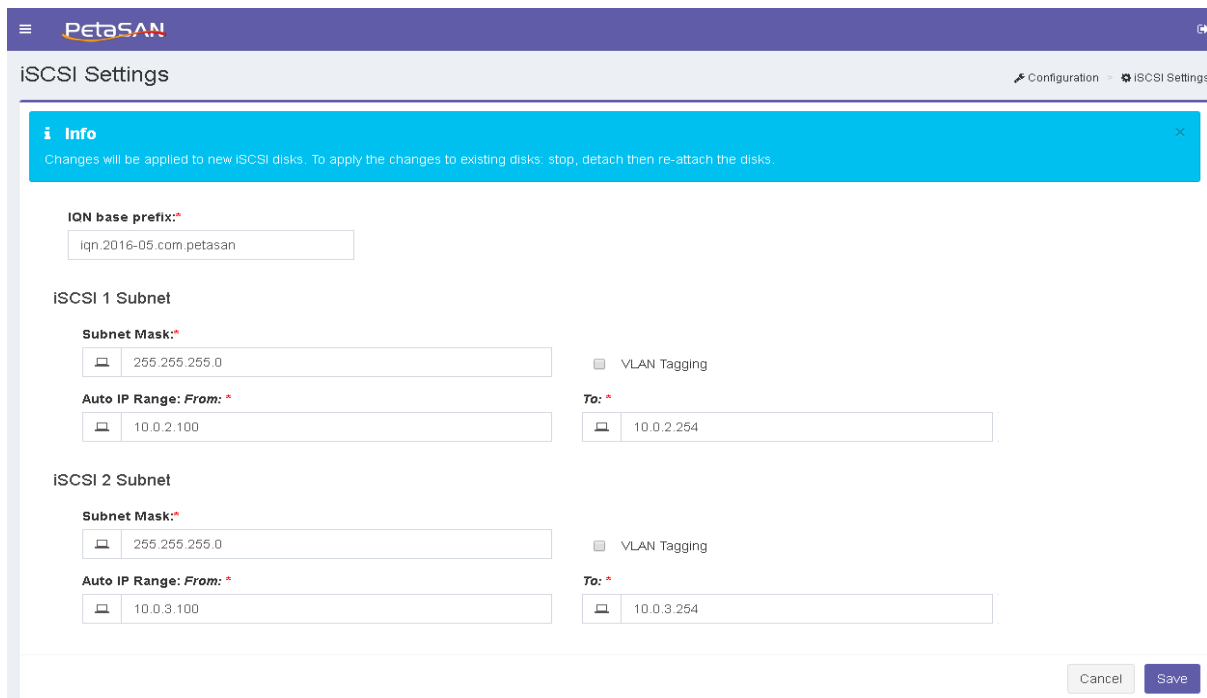
iSCSI2

Subnet Mask: 255.255.255.0

Auto IP From: 10.0.3.100

Auto IP To: 10.0.3.254

The base iqn is used as prefix for naming all created iSCSI targets. As per the *Planning the cluster network* section, we setup subnets 10.0.2.X/255.255.255.0 and 10.0.3.X/255.255.255.0 for iSCSI1 and iSCSI2 subnets respectively. We also define auto IP ranges; these are used by the PetaSAN system to optionally assign virtual IP addresses automatically to newly created iSCSI disks, in a way somewhat similar to DHCP.



iSCSI Settings

Info
Changes will be applied to new iSCSI disks. To apply the changes to existing disks, stop, detach then re-attach the disks.

IQN base prefix:
iqn.2016-05.com.petasan

iSCSI 1 Subnet

Subnet Mask: 255.255.255.0 ☐ VLAN Tagging

Auto IP Range: From: 10.0.2.100 **To:** 10.0.2.254

iSCSI 2 Subnet

Subnet Mask: 255.255.255.0 ☐ VLAN Tagging

Auto IP Range: From: 10.0.3.100 **To:** 10.0.3.254

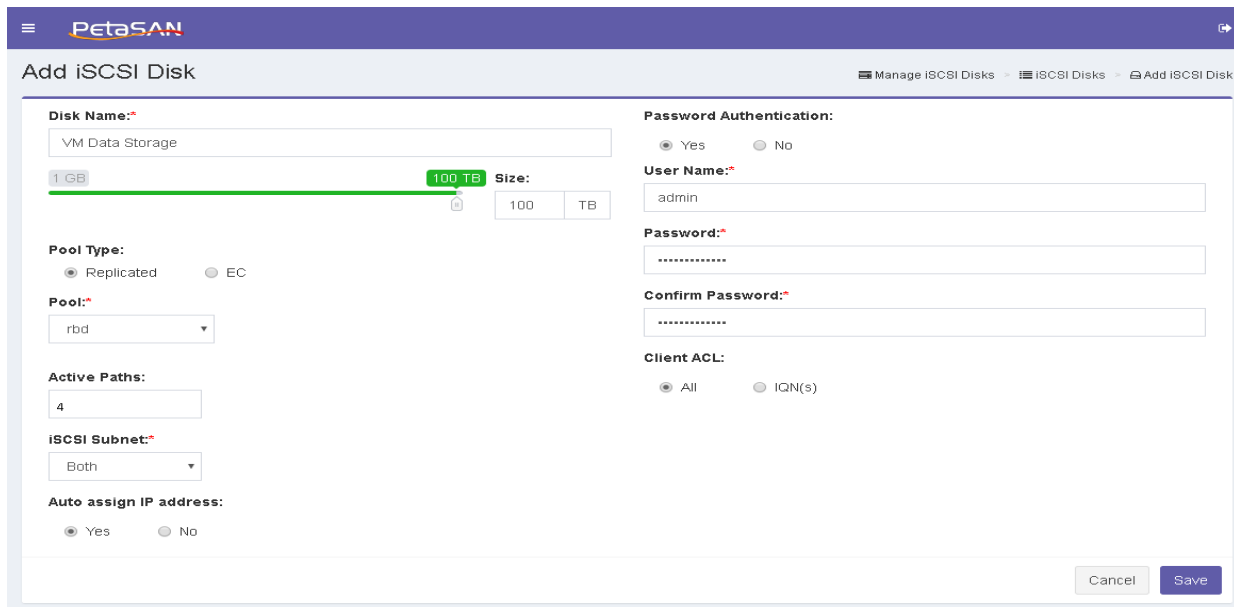
Cancel Save

- **Note:** iSCSI settings can be changed at any time during cluster operation; this will affect the IP addresses allocated to newly created disks. Disks already created will still use their existing IP settings. To re-assign new IP settings to older disks, they need to be stopped, detached and then re-started.

7.2 Creating disks

So far it has all been configuration work. We are now ready to start using PetaSAN and do what it is designed to do best, create and serve disks.

Go to the Add Disk page, and create a 100 TB disk on the rbd pool using 4 paths distributed across both iSCSI subnets:



Once created successfully, the disk will be listed on the Disk List page



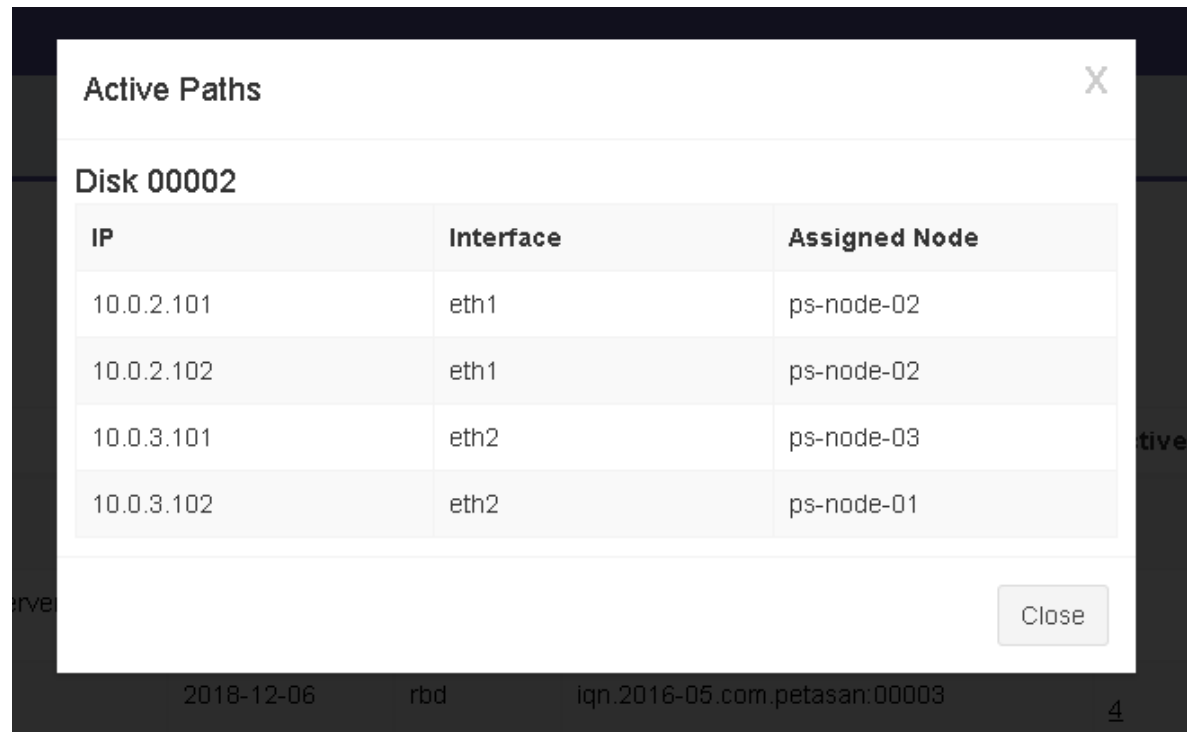
Disk Id	Size	Name	Created	Pool	IQN	Active Paths	Status	Action
00001	100 TB	BigDisk	2018-08-29	rbd	iqn.2016-05.com.petasan:00001	4	Started	

If this sounds too easy...*it is!*

The Disk List page includes a Search feature which can search and filter disks based on several criteria like disk name and size.

There are also several actions we can perform on the disk such as Start/Stop, Edit, Delete as well Attach and Detach. Detach removes iSCSI settings from disk metadata but does not remove the disk image.

To view the paths for our disk, under the Active Paths column click on the number of paths shown (in our case 4) this will show the IP addresses, the Ethernet cards used and their current node assignments.



Active Paths		
Disk 00002		
IP	Interface	Assigned Node
10.0.2.101	eth1	ps-node-02
10.0.2.102	eth1	ps-node-02
10.0.3.101	eth2	ps-node-03
10.0.3.102	eth2	ps-node-01

2018-12-06 rbd iqn.2016-05.com.petasan:00003 4

7.3 Storage Over-Commit

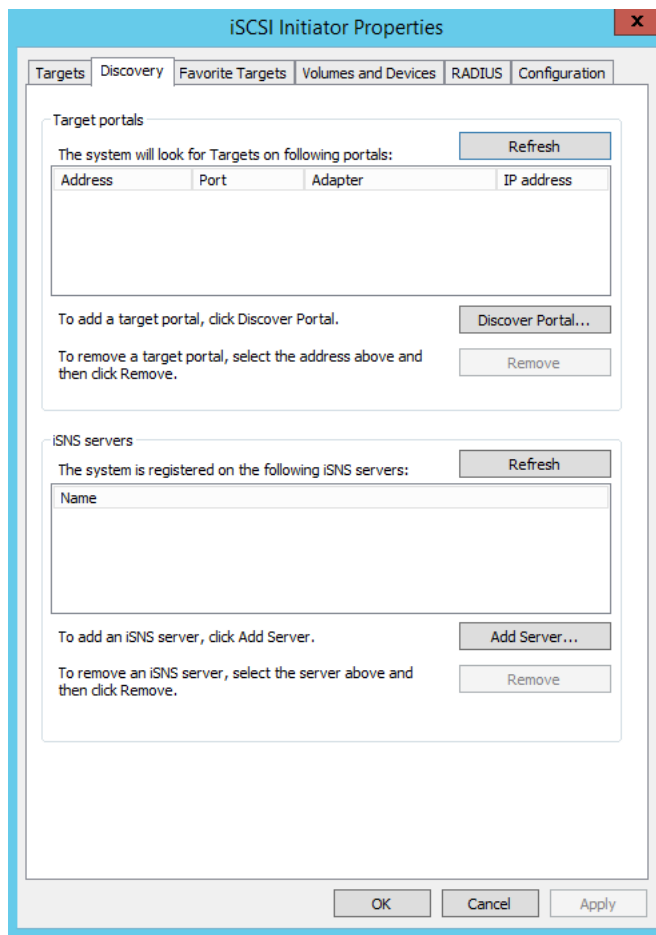
PetaSAN uses cloud technology which allows us to over-commit /thin provision storage. This means we can create iSCSI disks whose total storage exceeds the actual physical storage available in our cluster. We have done this in our previous example; we created a 100 TB iSCSI without having this storage physically available.

This is possible since we will not fill all our iSCSI disks from the very beginning, we may even never fill them at all (on some disks at least). PetaSAN dashboard shows us shows us how much total data (aggregated from all iSCSI disks) has actually been written to our physical storage and how much free storage we have left. We need to add more storage / nodes as needed if our usage approaches the physical available limit.

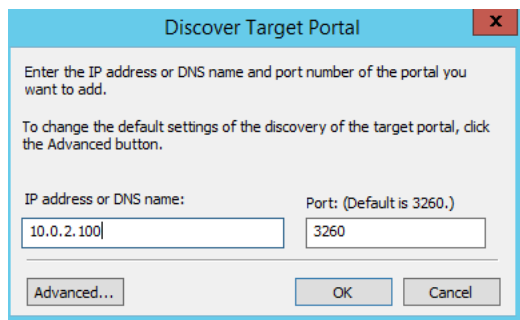
8. Client connection

Connecting iSCSI client initiators is largely dependent on the client operating system environment. In this example we will connect from a Windows client machine. To keep things simple, we will connect using a single path only. A more detailed description for connecting using MPIO and CHAP based authentication will be covered in another guide.

Open the “iSCSI Initiator Properties” select the “Discovery” tab then click on the “Discover Portal” button.



Enter the IP address of our first path: 10.0.2.100



Discover Target Portal

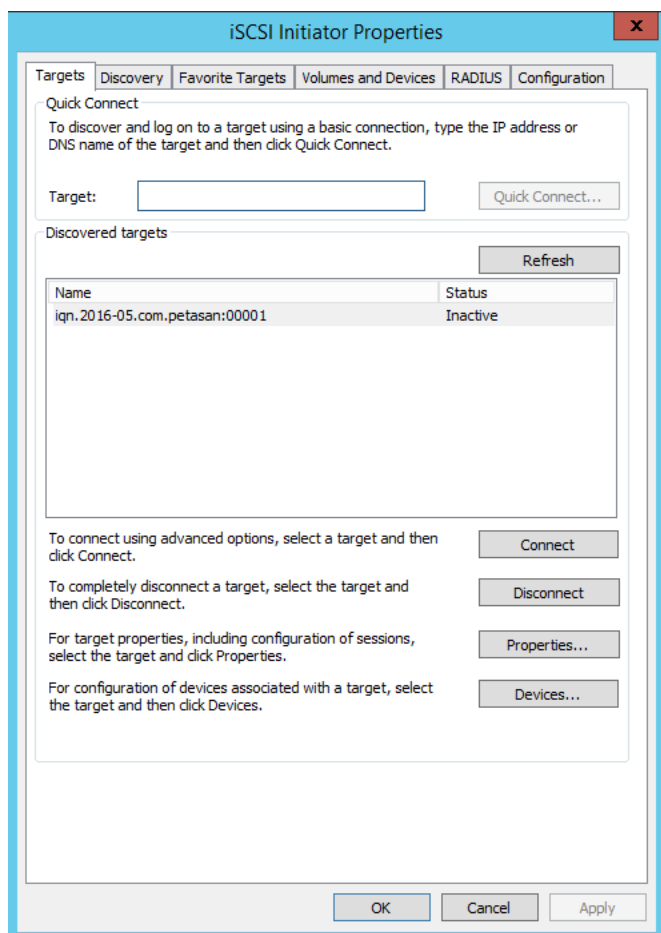
Enter the IP address or DNS name and port number of the portal you want to add.

To change the default settings of the discovery of the target portal, click the Advanced button.

IP address or DNS name: Port: (Default is 3260.)

Advanced... OK Cancel

Select the “Targets” tab then select our disk and click on the “Connect” button.



iSCSI Initiator Properties

Targets Discovery Favorite Targets Volumes and Devices RADIUS Configuration

Quick Connect

To discover and log on to a target using a basic connection, type the IP address or DNS name of the target and then click Quick Connect.

Target: Quick Connect...

Discovered targets

Refresh

Name	Status
iqn.2016-05.com.petasan:00001	Inactive

To connect using advanced options, select a target and then click Connect.

To completely disconnect a target, select the target and then click Disconnect.

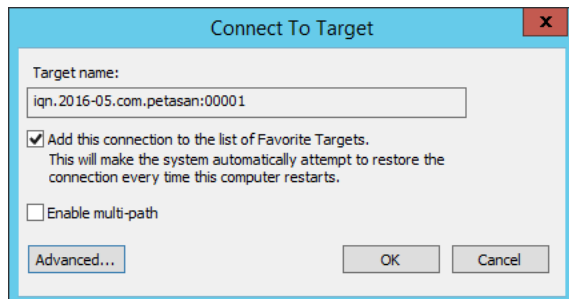
For target properties, including configuration of sessions, select the target and click Properties...

For configuration of devices associated with a target, select the target and then click Devices...

Connect Disconnect Properties... Devices...

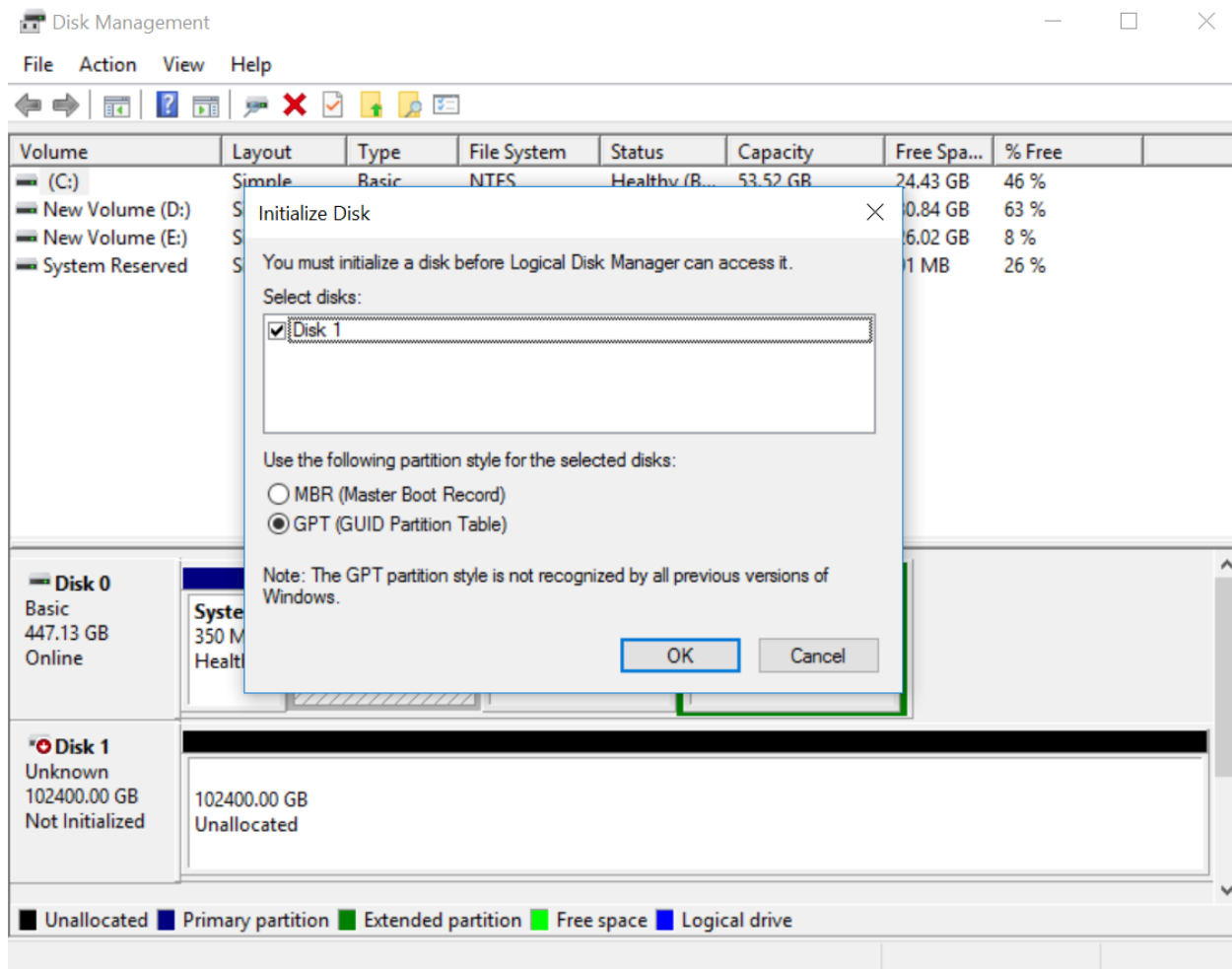
OK Cancel Apply

In the “Connect To Target” click “OK”

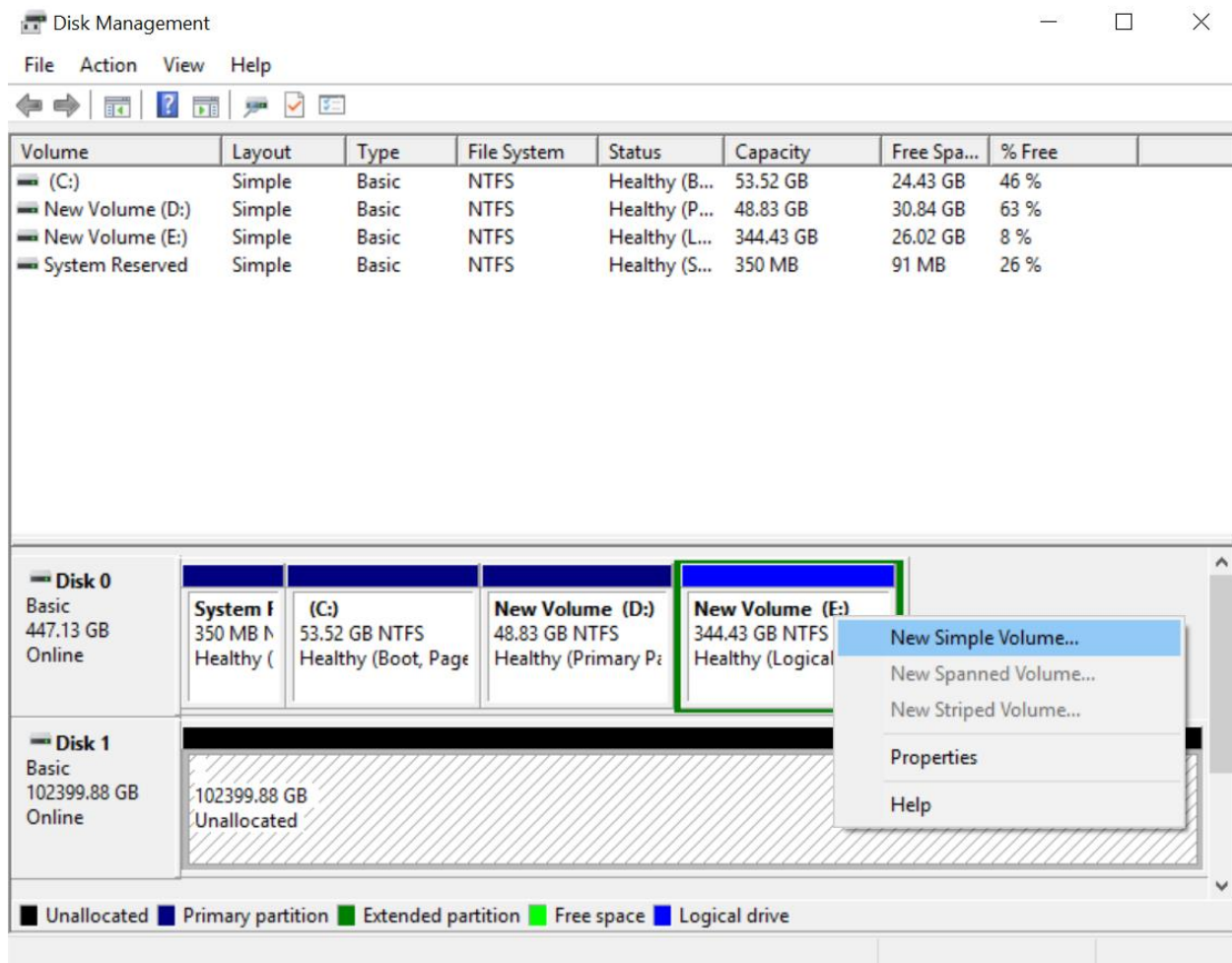


The disk is now connected, this is similar to having attached a new physical disk, we need to initialize and format it.

Open “Disk Management” and click “OK” to initialize the iSCSI disk.

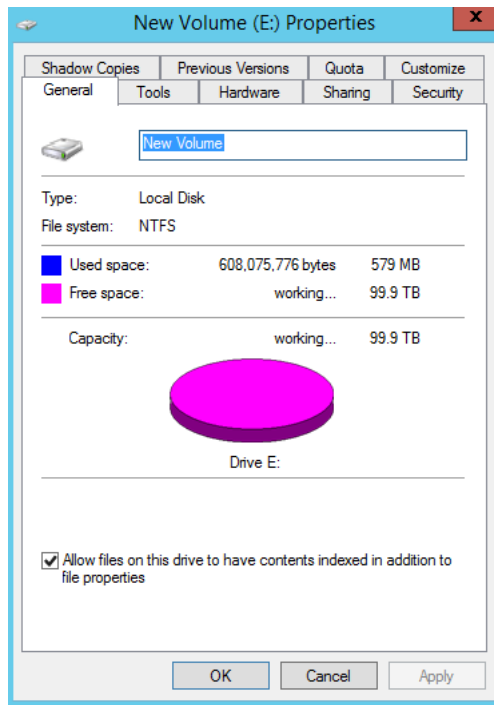


Once initialized, right click on the disk and select “New Simple Volume...” from the menu



This will open a wizard which will format the disk using NTFS and assign a drive letter, just use the default values.

Congratulations. We have formatted our 100 TB disk and is ready for use.



9. Command line access (for power users)

One major design goal of PetaSAN is to provide easy to use system management with point and click interfaces. However we have also made the system open for power users who would like to get their hands dirty with the system internals, for example to access the underlying Ceph and iSCSI layers at the command line level.

All nodes can be accessed securely using SSH with the following credentials:

username: root
password: Cluster password