



HIGH AVAILABILITY CLUSTER
Technology White Paper

Rev. 16-04-06

Index

Introduction	4
Benefits of HA Cluster	5
High-availability architecture	6
Cluster Events.....	7
Hardware recommendations.....	8
Performance recommendations.....	9
Limitations.....	11
Example Configurations.....	12

Copyright © 2015 euroNAS GmbH. All Rights Reserved.

euroNAS believes that the information in this publication is accurate of its publication date. The information in this publication is provided “as it is”. euroNAS GmbH makes no warranties of any kind with respect to the information in this publication. euroNAS GmbH specifically disclaims implied warranties or fitness for a particular purpose.

Microsoft, VMware, ESX, Xen and all other trademarks are registered trademarks and the property of their respective owners.

Introduction

This white paper provides an overview of the functionality and explains technologies around euroNAS high availability product.

In recent years, demand for high availability storage has seen explosive growth. Companies have become more and more reliant on the content in the data storage. The reliability and availability of data storage needed to be dramatically increased to prevent the severe consequences that may result from server failure. Mission critical storage systems are required to achieve greatest possible availability and fastest possible recovery from catastrophic disaster.

More than ever it is important to keep the business continuity even on total server failure.

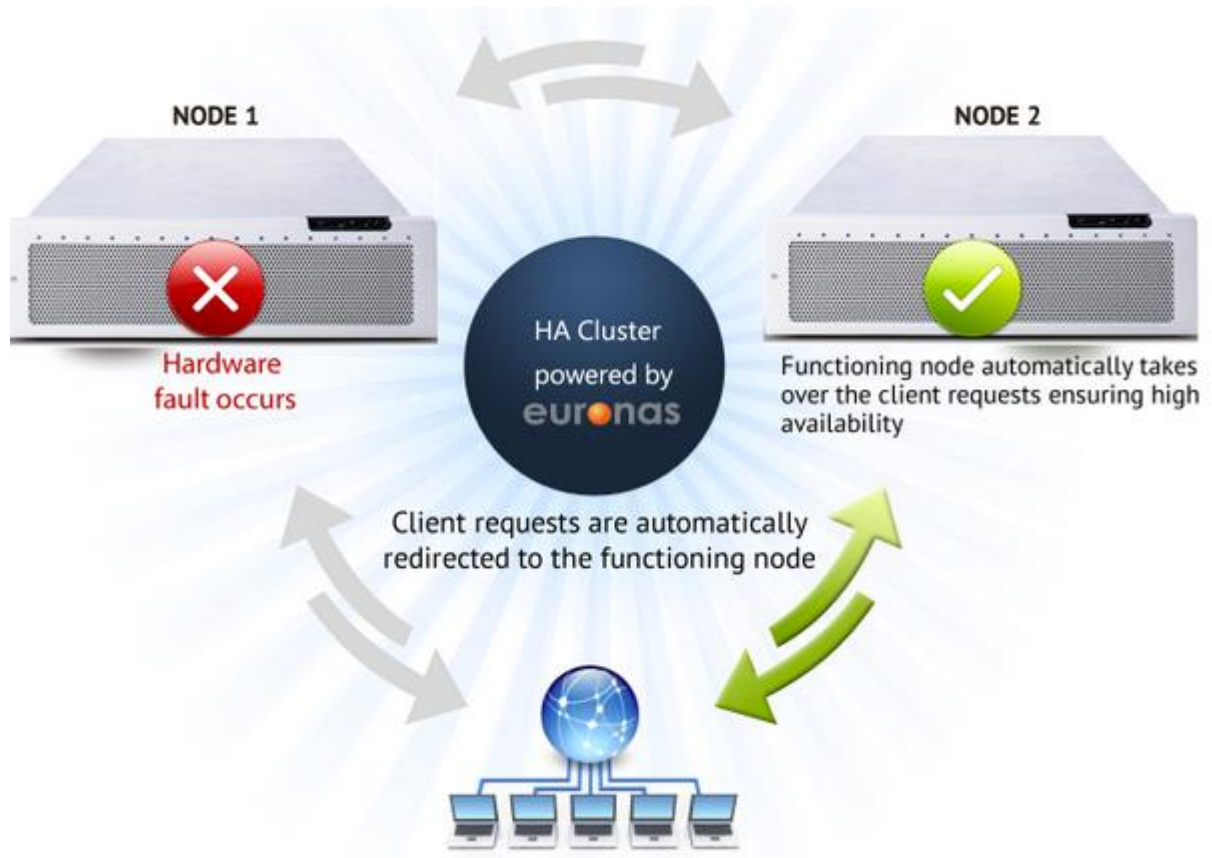
euroNAS HA Cluster is used to provide critical shared data on a network, and as system and data drives for physical and virtual production servers and critical workstations. Data is being mirrored in real time between both servers.

If you're running a virtualized environment or simply trying to meet the demands of the 24-hour business cycle, you cannot afford downtime.

By selecting euroNAS storage software you are not only getting highest possible redundancy in case of server failure but also hardware tolerant operating system that runs on any hardware.

Legacy storage manufacturers often use proprietary hardware and an operating system which will only run on the hardware in question. This makes it difficult to expand in the future and forces you to buy new units. euroNAS will grow with the needs of your company.

Also recovery of total server failure is much more difficult if proprietary hardware is in use. It is not tied to proprietary hardware – this makes it a highly flexible and scalable solution.



Benefits of HA Cluster

Higher availability

When using standard NAS server your network applications depend on this single point of failure. When using HA Cluster your network applications will continue to work - no matter which server fails, clients can still access their data. Standard NAS Servers can provide many redundant features such as raid, network port failover etc. but they don't protect you from total server failure. Even with proper backup policy it takes a while and manual intervention to get server and data back online.

With HA Cluster, should a hardware or software fault occur, it is detected through intelligent features and the software automatically and immediately moves storage requests to the other server mirrored in the cluster.

Data Replication

euroNAS advanced technology creates a cluster of two mirrored servers with real time, continuous data replication and synchronization. Both servers contain identical data securing redundancy in case of server failure.

Redundant services

euroNAS HA Cluster offers wide range of services that enable you the maximum flexibility. All this services can be used simultaneously. Currently following services are supported

- SMB/CIFS/AFP
- Active Directory authentication (for CIFS and AFP)
- NFS
- iSCSI (including persistent reservation iSCSI Lun and Target services)

Higher performance through smart resource allocation (active/active)

Powerful cluster management functionality enables you to define which resource should run on which server. This way both servers are used in most efficient way providing highest level of performance at the same time. Should one server fail, services will still automatically move to the other server. When the failed server is back online, its services will automatically move back.

This feature is not dependent on services. You can have for example iSCSI Targets split over both servers or run iSCSI Targets on one server and CIFS/AFP services on other.

All resources created on the same cluster drive will run together on the same node.

Simple Management interface

HA Cluster is perfect for IT Professionals. You are not only gaining a high available storage solution with continuous replication but also an efficient and simple to use management interface. Whole server and cluster configuration can be done within 10 minutes.

High-availability architecture

In order to achieve high-availability HA Cluster consists of following parts

Cluster node

Cluster node represents individual server that is member of the cluster. All nodes are equal – from each node you can monitor and configure cluster services.

Cluster drive

Cluster drive represents a mirror of 2 drives on each individual server.

This 2 drives can be a single disk, software or raid array. They are mirrored in real-time. On this drives shares and iSCSI Targets are installed.

This drives don't necessary need to use full size of the available space. It is possible to create multiple cluster drives on the same disk or RAID array. They can be then split across the servers.

For example :

On each individual server there is a RAID array defined of 16 TB. On cluster drive creation you can create 2 separate cluster drives of 8 TB size. Each 8 TB cluster drive can be defined to run on one of the servers. All resources defined on this cluster drive (iSCSI, NFS, AFP/CIFS) will run on this server. This way the load between the servers will be balanced and you will achieve best possible performance.

Cluster resource

Cluster resource represents network share or iSCSI Target within the cluster. This can be SMB/CIFS & AFP, NFS share or iSCSI Target. Each resource has its own size and is accessible through the ip address defined for the cluster drive.

Corporate network

Corporate network is the network used by network clients for accessing cluster resources

Replication network

Internal network used for heartbeat and data replication between the servers

Cluster Access IP Address

Cluster Access IP Address is the address that enables clients to access shares and iSCSI Targets. The greatest advantage is that this address is shared by both HA cluster systems. If any of these servers fail – the cluster IP address will move to the running server.

- Cluster Shares and iSCSI Targets will not be available for client connections until you create an Access IP Address for the Cluster Drive
- Access IP Address can only point to one cluster drive
- The IP Address must be in the Corporate Network

Cluster access IP address is defined per cluster drive. All resources on this drive are available through this IP address.

Network Test IP

The purpose of this IP(s) is to prevent storage desynchronization, also known as “split brain” situations. Such situation may occur when both nodes are running, but one or both are disconnected from the network.

HA Cluster nodes will ping this address constantly and check if still is able to reach the network. It is recommended to define more than just one IP address.

IP address on any pingable device on your network that is constantly on, router, gateway, printer, mail server etc. can be used as network test IP.

In one of the nodes is unable to ping this IP it will automatically know that it cannot reach the network and push all services to other node.

If a node cannot reach all defined test IP addresses it will shut down services to make sure that data remains consistent.

Cluster Events

Within cluster there are several types of events that can occur. Depending on type transfer of services can take different time. Initial time for servers to realize that other server has failed is around 30-40 seconds. After this resources are moved pretty quickly.

Estimated time to move around 10 resources on another server is around 50-60 seconds. To move around 30 resources usually takes about 70-80 seconds.

These times may vary depending on the performance of the hardware and the size of the storages.

Failover

Failover occurs when server detects server fault and automatically moves all resource on another server. On failover services are moved faster on another server. One example would be total crash of one of the servers.

Switchover

Switchover happens when resource is manually moved from one server to another. This is usually done when node preference settings are changed or in case of system maintenance of one of the nodes. Moving services takes little bit longer than on failover.

Hardware recommendations

Installation Disk

For best performance the euroNAS cluster software should be installed on an SSD with a minimum size of 8 GB. Installation disk is separate from data.

If using hardware RAID controller, it is recommended to create 2 RAID array – one smaller for the installation and other large for data

Data Disk

Since installation disk is exclusively used for the OS, for data at least one additional disk is needed.

Data disk can be simple disk, software RAID array or disks managed by the hardware RAID controller.

System RAM

The server should have at least 8 GB of RAM

Networking

There must be one or more reliable IP Addresses on the Corporate Network that both servers can reach for proper operation of the High-Availability/Custer features

An example would be the router which is on the Corporate Network or connects the servers to the Corporate Network

For best performance and high availability, the server should have 4 individual Network Interface Cards with a minimum bandwidth of 1 Gigabit (10 Gigabit recommended). They should be configured as two separate bonded networks. One of the bonds should be connected to the Corporate Network. The other bond should be connected directly to the other server on a private Replication Network

If the servers are in the same location, they should be cabled directly to each other without any switches between them if possible.

For top performance the Replication Network should be 10 Gigabit

If dual-port cards are used each port on the card should be assigned to a separate bond so that the failure of one card does not break the entire bond

Card 1 Port 1 -> bond0 (Corporate Network)

Card 1 Port 2 -> bond1 (Replication Network)

Card 2 Port 1 -> bond0 (Corporate Network)

Card 2 Port 2 -> bond1 (Replication Network)

Quad-port cards should be avoided

DHCP should not be used unless static leases are provided

Using DHCP without static leases will cause the cluster to fail on IP address change

Performance recommendations

The bottleneck of HA Cluster is the replication network. The performance of the whole HA Cluster depends on the replication speed between the nodes. Therefore, we recommend 10 Gigabit connections between the nodes.

For the best disk performance, we recommend use of hardware RAID Controllers.

euroNAS HA Cluster is hardware independent however we still recommend using identical hardware. Due to server replication in real time, the performance will depend on the slowest link. Using slower disks on one server will slow the server with faster disks as well.

Limitations

HA Cluster detects automatically many possible fail scenarios and moves resources automatically. There are some scenarios where HA cluster is unable to detect the problem.

Replication network failure – corporate network available from both nodes

In this case server can reach network test IP from both nodes so both nodes think that they are online and available. On the other hand, replication network is unavailable. In order to prevent possible data loss server replication and resources are halted until replication network becomes available again. In order to prevent this, we recommend network card teaming (port failover)

Powering down master node during sync process

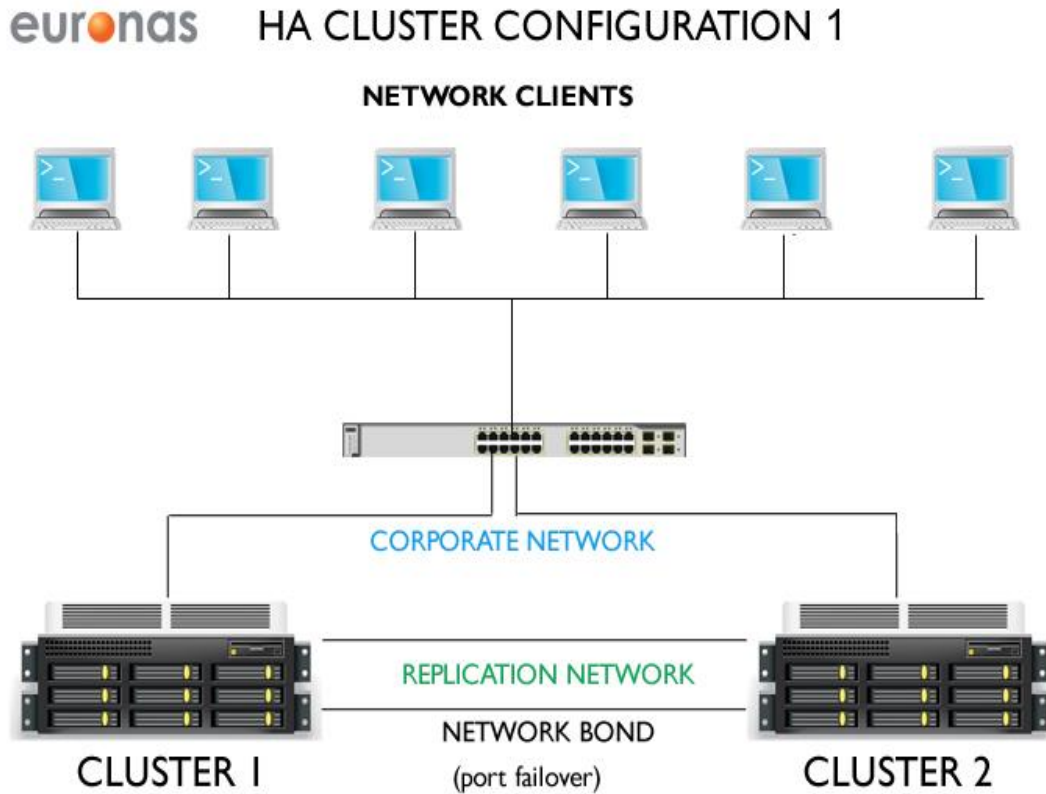
During the sync process the node with more current data must not be powered off. In this case server replication service will stop the drives in order to prevent data corruption.

Only server with less current data can be powered off. This does apply if both nodes are in sync. In this case either of the server can be powered off or rebooted.

Port-Failover (active-backup) should be used for configurations without a switch

If replication network is set in switchless configuration only network bond that should be used is port-failover (active-backup) because it is the only one that does not require any switch.

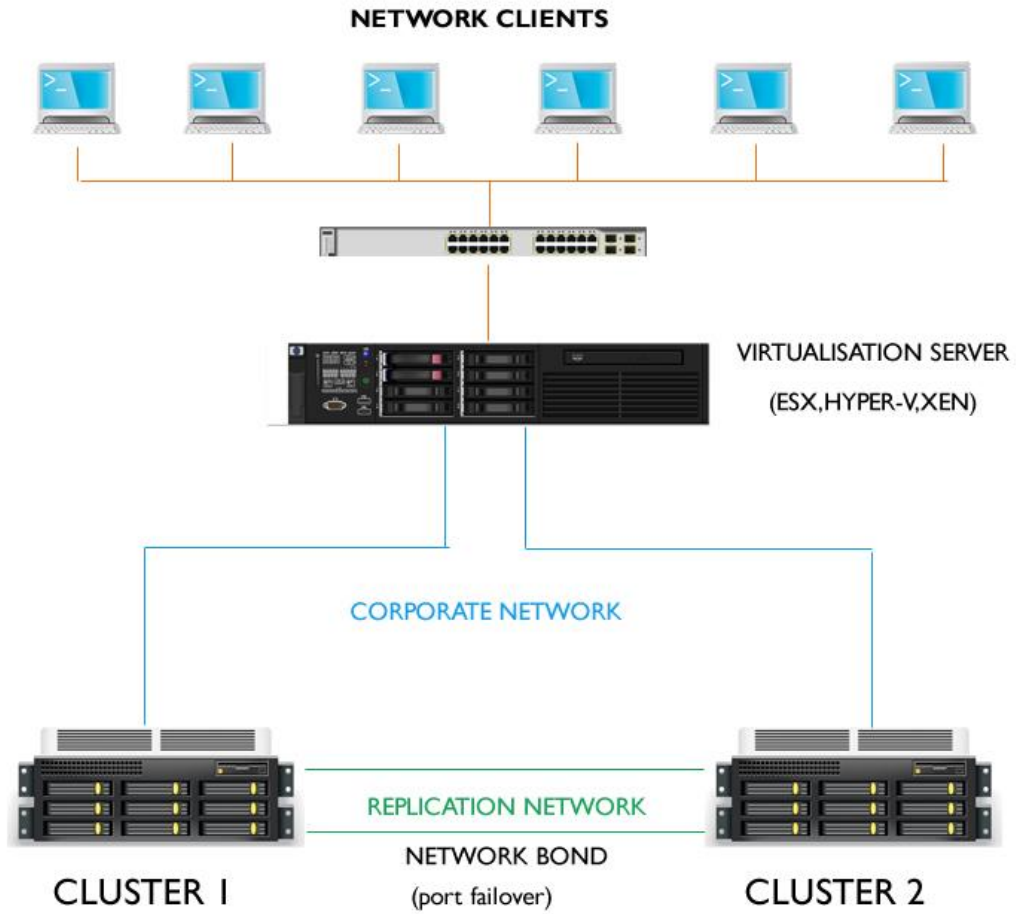
Example Configurations



In this configuration both cluster nodes are connected directly without a switch using port-failover. Clients can access data on the server by using corporate network. Should one of the servers fail other server node will automatically take over the services.

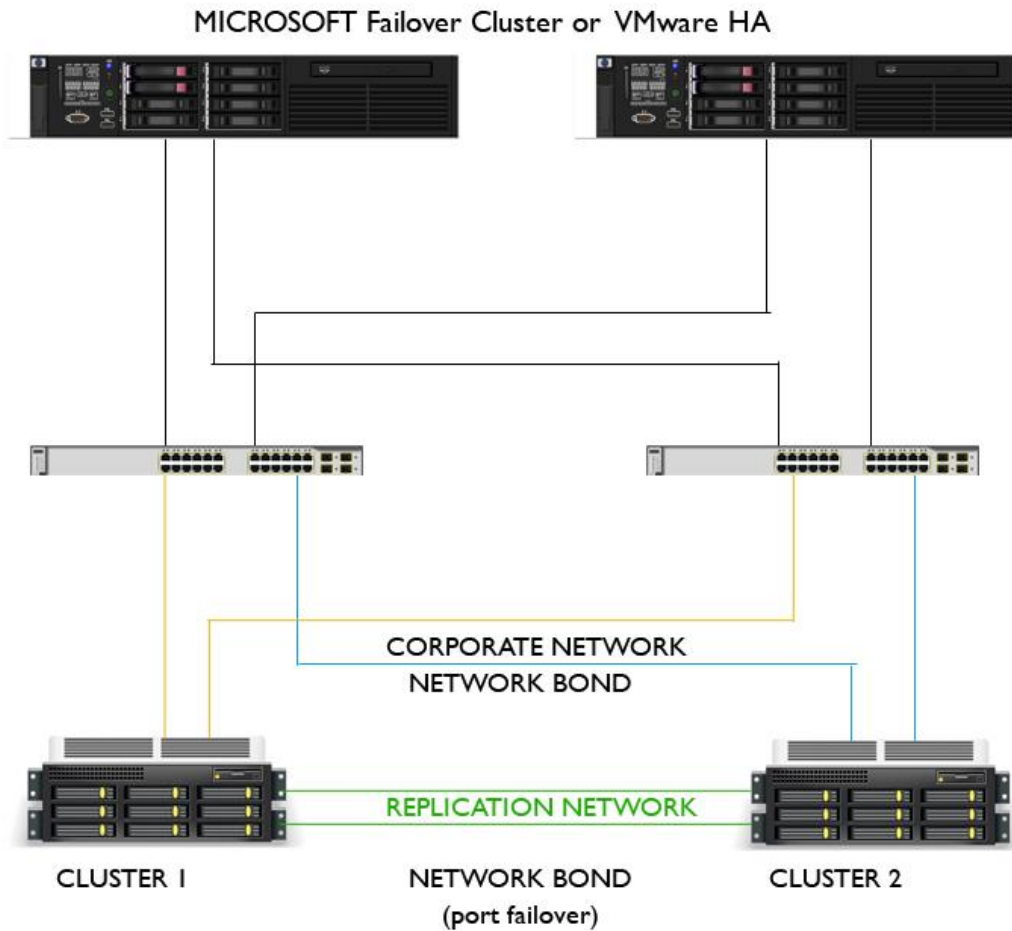
This is typical environment for use as a fileserver.

euroNAS HA CLUSTER CONFIGURATION 2



Configuration 2 shows possible connection on to a virtualization server. This can be for example VMware ESX, Microsoft Hyper-V or Citrix Xen. This configuration is done completely without a switch eliminating single point of failure and also reducing the overall costs for an additional 10 Gigabit switch. Should one of the servers fail another switch will take over.

euroNAS HA CLUSTER CONFIGURATION 3



Configuration 3 shows possible connection on to a high availability server such as Microsoft Failover Cluster or VMWare HA. In this configuration all connections are crossed across the switches eliminating single point of failure.

There are many ways of configuring the servers in proper way. Our support and sales team will be glad to assist you on the best configuration for your system.