

Clustering in Virtuozzo-Based Systems



(c) 1999-2008

CHAPTER 1

This document provides general information on clustering in Virtuozzo-based systems. You will learn what clustering scenarios are supported in Virtuozzo Containers 4.0 and how they can effectively be put into practice in your working environments.

The document does not deal with setting up and configuring clusters in Virtuozzo-based systems. For detailed information on these activities, please turn to the following resources:

- The **Deploying Microsoft Clusters in Virtuozzo-Based Systems** document familiarizes you with the way to use the MSCS (Microsoft Clustering Service) and NLB (Network Load Balancing) clustering technologies to create reliable and highly available clusters, including Virtuozzo Failover Clusters, in Virtuozzo-based systems.
- The **Deploying RHCS Clusters in Virtuozzo-Based Systems** document describes how to set up Virtuozzo failover and data sharing clusters on Hardware Nodes running the RHEL 5 and CentOS 5 operating systems.

Introduction

Clustering is a technology providing solutions for increased server availability and efficiency by creating reliable and highly available systems (*clusters*). Virtuozzo Containers 4.0 allows you to deploy the following types of cluster configurations in your working environments:

- *Virtuozzo Failover Cluster*: this cluster configuration can be used to provide failover support for your Virtuozzo Containers installations and to ensure that Virtuozzo vital services and Containers continue running in the event of a planned or unplanned Hardware Node downtime.
- *Container-to-Container cluster*: this cluster configuration can be used to provide high-availability and reliability for your mission-critical applications by failing over the application resources from one Container acting as a cluster node to another in the event of a hardware or software failure.
- *NLB cluster*: this cluster configuration can be used to distribute client requests among the cluster nodes according to the current network load, thus, greatly increasing your applications and services availability and performance.
- *Data sharing cluster*: this cluster configuration can be used to provide simplified manageability of your Containers and Virtuozzo Templates, to allow fast Container migration between cluster nodes, and to provide failover support for Virtuozzo mission-critical services and Containers.

The following sections dwell on each of those cluster configurations in detail helping you make a decision about what configuration best suits your needs for performance, high availability, and manageability.

Virtuozzo Failover Cluster

Using the Microsoft Windows and Red Hat clustering software, you can create a Virtuozzo Failover Cluster consisting of two or more Hardware Nodes and providing failover support for Virtuozzo mission-critical services and Containers. If one Hardware Node in the cluster fails or is taken offline as part of a planned maintenance, the services and Containers from the problem server are automatically failed over to another Node in the cluster. The following picture shows a simple cluster configuration that can be used to provide high-availability for your Virtuozzo Containers 4.0 installations:

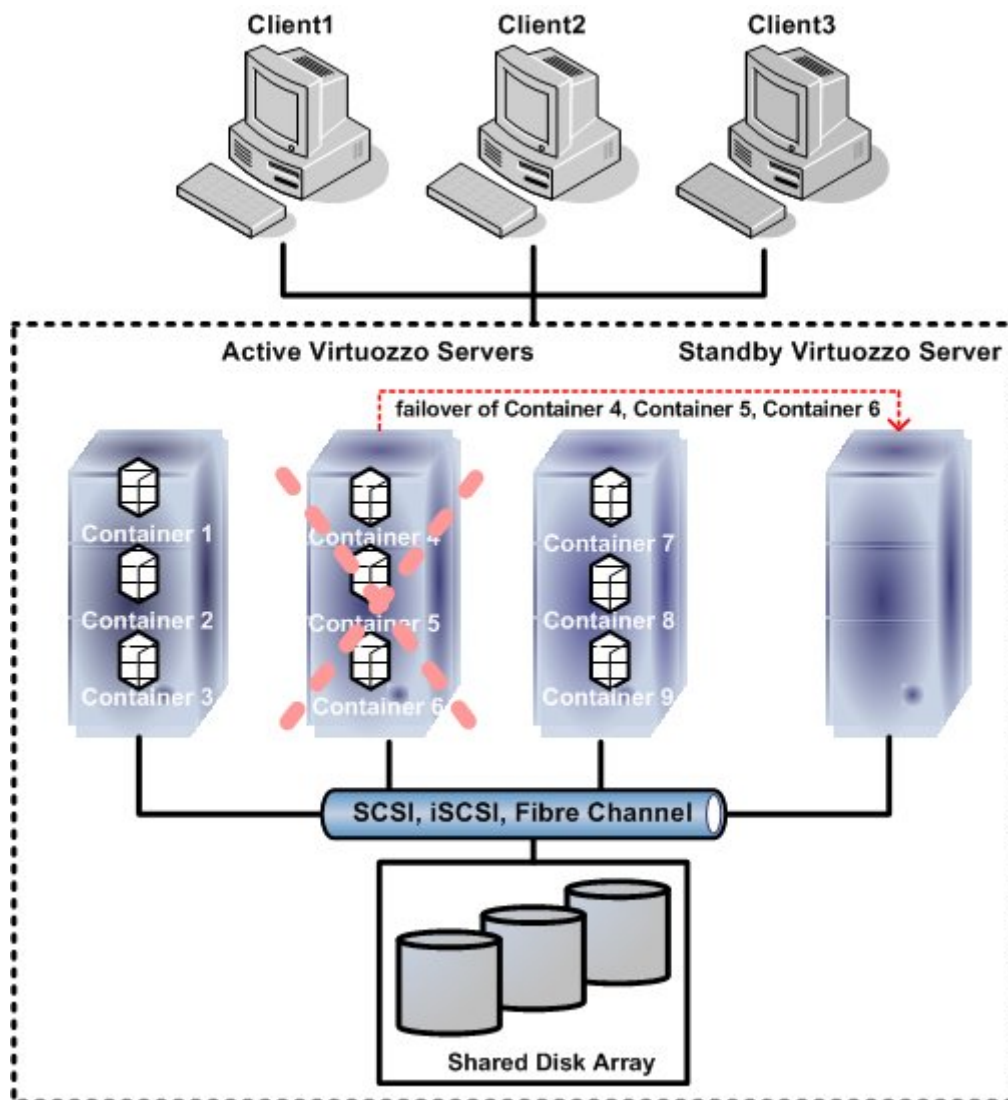


Figure 1: Failing Over Containers

In this clustering scenario, the Virtuozzo Failover Cluster comprises three active and one standby node. All nodes in the cluster have one and the same version of the Virtuozzo Containers software installed. Each of the active nodes hosts a number of Containers configured as cluster resources and capable of failing over to the standby node if the corresponding active node becomes inaccessible. For example, if the active node hosting Container4, Container5, and Container6 is taken offline as part of a planned or an unplanned event, the clustered software will detect the problem, fail over these Containers to the standby node, and start them there.

All the nodes in the cluster are connected to a reliable and fault tolerant shared disk array on a Storage Area Network (SAN). The SAN provides a high-speed communication channel (via SCSI, iSCSI, or Fibre Channel) between each node in the cluster and the shared disk array. The disk array is used to store all the cluster configuration data and Virtuozzo resources and allows access to these resources to the active nodes in the cluster.

The primary goal of implementing a Virtuozzo Failover Cluster in your working environments is to provide a high degree of availability and reliability for Virtuozzo mission-critical services and Containers by treating them as cluster resources and failing them over to the standby Hardware Node in the event of an active node failure. However, you may also think of deploying this type of cluster to provide high availability for your applications and services (both cluster-aware and cluster-unaware) by running them inside Containers residing on any of the active Hardware Nodes participating in the Virtuozzo Failover Cluster.

Note: The clustering scenario described above is supported in:

- Virtuozzo Containers 4.0 for Windows utilizing the Microsoft Cluster Service software;
 - Virtuozzo Containers 4.0 for Linux utilizing the Red Hat Cluster Suite software.
-

Container-to-Container Cluster

Deploying a Container-to-Container cluster allows you to provide high availability and reliability for your mission-critical applications (both cluster-aware and cluster-unaware). In this type of cluster, one node is active and owns all the application resources at any one time, and another node is in the standby state. If the active node fails, the application resources are failed over to the standby node which comes online.

In Virtuozzo Containers 4.0, any Container can be configured to operate as a full member in a cluster supporting virtually every clustering configuration that can be implemented for cluster-aware and cluster-unaware applications on standalone servers. Like any other standalone server, a Container can be set up to function as an active cluster node running some vital application or it can be configured to take over the failed application resources of another cluster node. For example, you can unite two Containers into a Container-to-Container cluster and make this cluster host the Active Directory application. In this case the clustering software will keep check of this application and, when detecting a failure inside the active Container, quickly assign the Active Directory application responsibility to the standby Container and start it there. At the same time, you can consolidate a Container and a standalone server to provide high availability for the Active Directory application, which is shown in the picture below:

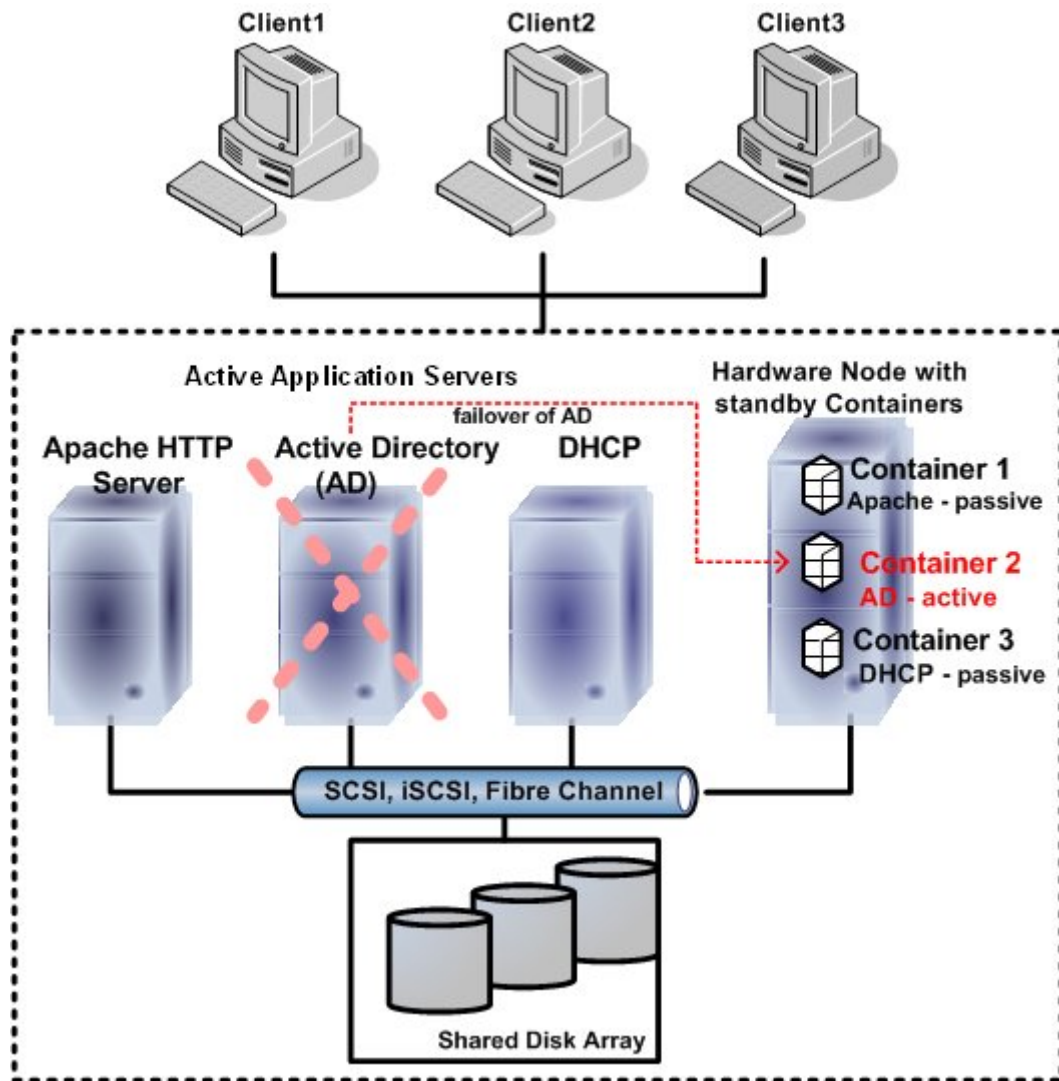


Figure 2: Failing Over Applications

In this example the cluster consists of:

- Three active nodes. These are standalone servers hosting two cluster-aware applications (*Active Directory* and *DHCP*) and one cluster-unaware application (*Apache HTTP Server*) and owning all the resources of these applications. The applications are configured to run in the cluster and can be failed over to a standby node in the cluster in the event of a hardware or software failure.
- Three standby nodes. These are Containers residing on a single Hardware Node running the Virtuozzo Containers 4.0 software. The Containers are configured to take control over the corresponding application resources (*Apache HTTP Server*, *Active Directory*, or *DHCP*) in the case of an active node failure. For example, if the *Active Directory* application becomes inaccessible on the node where it is currently hosted, its resources will be failed over to *Container 2* residing on the Hardware Node and configured as the standby node for this application.
- All nodes in the cluster communicate with each other across a local interconnect (Fibre Channel, SCSI, or iSCSI) and share a common disk array on a SAN. The shared disk array is used to store all the cluster configuration and application data.

Using the Virtuozzo Containers technology in this clustering configuration brings you the following benefits:

- Hardware cost reduction – instead of deploying a physical standby server for every active server in a Failover Cluster, you set up just one with as many Virtuozzo Containers running on it as needed. Each Container can be set up as a standby server, just like a physical machine, but will share the resources of the same host with the rest of the Container-based standby servers.
- More efficient hardware utilization – a standby server in a Failover Cluster normally spends most of the time in the standby mode. In a scenario when a standalone physical server is used, the server resources are greatly underutilized. Using Containers to create the standby nodes enables a high density of Containers on the same server and brings the hardware utilization to the optimum level. Once the active standalone server fails, the workload fails over to the passive Container, which, due to the unique resource management and scalability available with Virtuozzo Containers 4.0, can be scaled to the entire resources of the server, potentially maintaining the same level of processing on the initially standalone server.
- Maintenance cost reduction – by using the Virtuozzo Containers software, you can consolidate many physical standby servers into one or fewer servers, which will significantly reduce the costs associated with hardware and software maintenance.

Along with using physical SAN-based storage devices, Virtuozzo Containers 4.0 also allows you to create special loopback files on the Hardware Node and to configure them as shared storage devices for your clusters. However, using loopback files as cluster storage devices imposes a number of limitations on the cluster configuration. The main of these limitations are listed below:

- A cluster may consist of Containers only.
- All Containers participating in such a cluster must reside on one and the same Hardware Node, i.e. on the Node where the corresponding loopback file is located.
- All cluster and application resources are hosted on a single Hardware Node and, therefore, not protected from hardware crashes.

In spite of all these limitations 'loopback file'-based clusters are perfectly suitable for dealing with software crashes and administrative errors. They can also be used for testing and demonstrating purposes.

Note: The clustering scenario described above is supported in Virtuozzo Containers 4.0 for Windows utilizing the Microsoft Service Cluster software.

NLB Cluster

Network Load Balancing (NLB) provides a high level of reliability, availability, and scalability for your mission-critical applications and services. In Virtuozzo Containers 4.0, Containers can be combined with each other or with standalone servers to form an NLB cluster. The following picture shows a typical NLB cluster configuration that can be deployed in your working environments:

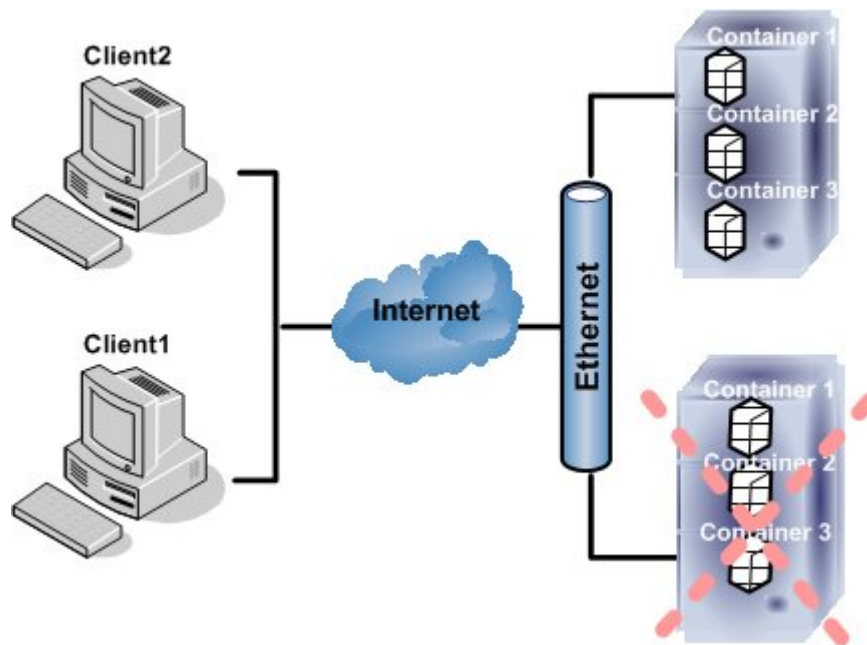


Figure 3: Deploying NLB Cluster

In this configuration the cluster consists of six Containers residing on two Hardware Nodes:

- Container 1, Container 2, and Container 3 on the first Hardware Node and
- Container 1, Container 2, and Container 3 on the second Hardware Node.

Each Container runs a separate copy of an application (e.g. Web server) and is configured to be able to respond to client requests. Network Load Balancing distributes incoming client requests across the Containers in the cluster in accordance with the current network load. This cluster configuration is defined by the following main features:

- If one Container on a certain Hardware Node fails, the load is automatically redistributed among the remaining active Containers on two Nodes. When the failed Container goes online again, it transparently rejoins the cluster and regains its share of the workload.
- If all Containers on one and the same Node fail (e.g. when the Hardware Node is taken offline as part of an unplanned event), the incoming network traffic is redirected to Container1 and Container2 on the second Node. When the Containers become functional again, they are transparently rejoined to the cluster.

- As the traffic to your network applications increases or as your applications require more server power, you can dynamically add another Node to the cluster and create three more Containers on this Node.

Note: The clustering scenario described above is supported in Virtuozzo Containers 4.0 for Windows utilizing the Microsoft Network Load Balancing software.

Data Sharing Cluster

Data sharing clusters are server farms that share storage devices on a storage area network (SAN) and share data on those storage devices. In a data sharing cluster, data can be written or read by any server to or from any file on a common shared storage device. The following picture illustrates the structure of a typical data sharing cluster:

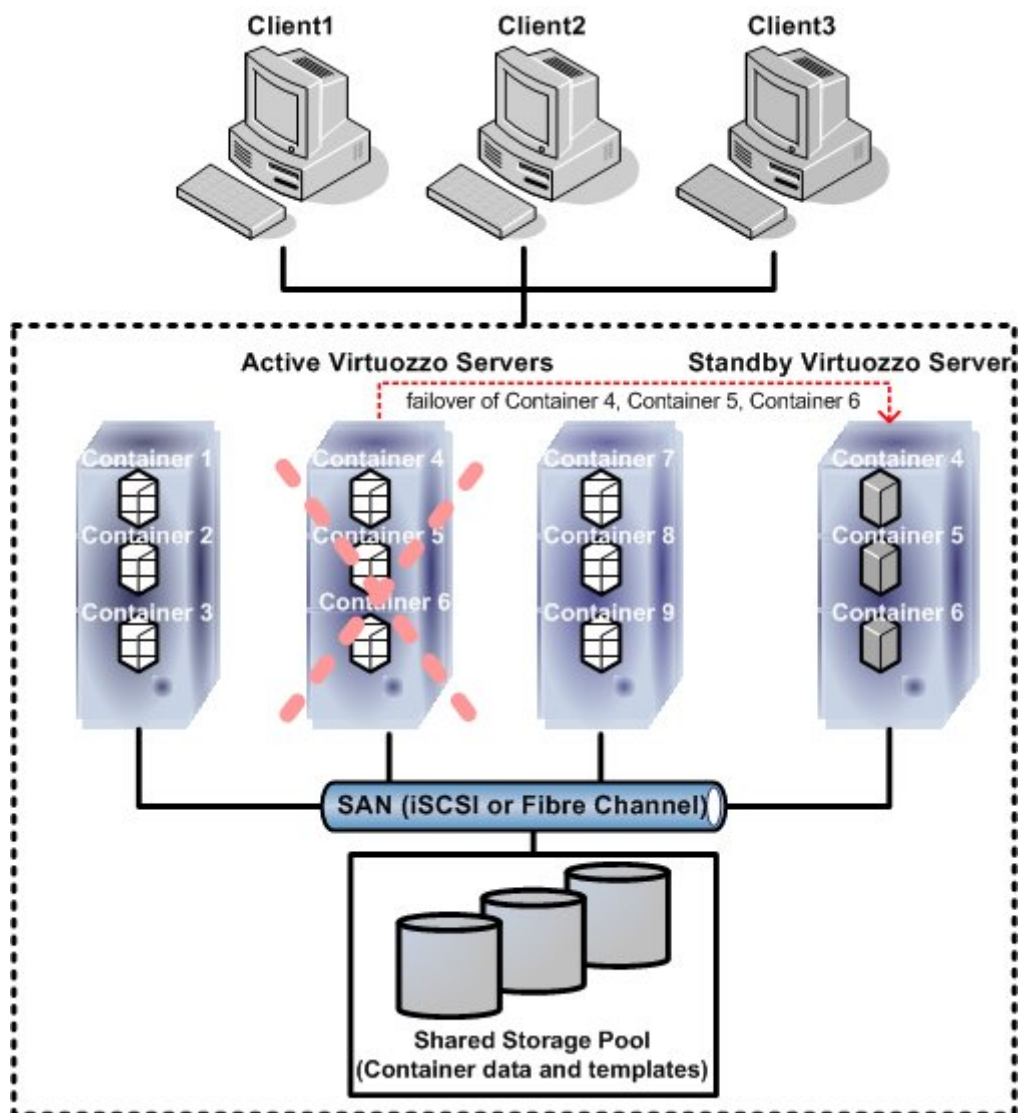


Figure 4: Deploying Data Sharing Cluster

In this example the data sharing cluster comprises four identical Hardware Nodes running the Virtuozzo Containers 4.0 software. Three Hardware Nodes in the cluster are active meaning that they are currently hosting a number of Containers, and one Hardware Node is configured as a standby server (e.g. it can be used to host the Containers from some active node when this node is taken offline for maintenance reasons).

The nodes are connected with a shared disk array on a storage area network (SAN) over a high-speed communication channel (via iSCSI or Fibre Channel). The Virtuozzo Templates and Container private data from all servers in the cluster are kept on this shared disk array and can simultaneously be accessed by any server.

Deploying a data sharing cluster made up of Virtuozzo Hardware Nodes allows you to achieve the following main goals:

- Simplify the process of managing your Containers and Virtuozzo Templates since all Containers and templates are residing on a single disk array shared by all nodes in the cluster.
- Greatly speed up the process of migrating running Containers between the cluster nodes. In fact, the migration is almost imperceptible to users since all Container data in the cluster is stored on a shared SAN storage and there is no need to move this data between the Nodes during the Container migration.
- Provide failover support for Virtuozzo vital services and Containers. Each server in the cluster is running the clustering software responsible for monitoring the health of Virtuozzo Containers installations and failing over the services and Containers from a failed node to the standby node.

Note: The clustering scenario described above is supported in Virtuozzo Containers 4.0 for Linux utilizing Red Hat Cluster Suite and Red Hat GFS.
