

# VDCF - Virtual Datacenter Control Framework for the Solaris™ Operating System

## White Paper

Version 4.2  
December 2011

Copyright © 2005-2011 JomaSoft GmbH  
All rights reserved.

## 1 Introduction

This documentation describes the Virtual Datacenter Control Framework (VDCF) for the Solaris Operating System, Version 4.2.

### 1.1 Overview

Virtualization is an approach to IT that pools and shares resources so that utilization is optimized and supply automatically meets demand. The case for Virtualization is compelling: industry analysts estimate that the average utilization rate of a typical IT data-center's resources is between 15 and 20 percent.

With Virtualization, IT resources dynamically and automatically flow toward business demand, driving up utilization rates and aligning IT closely with business needs.

Pooling and sharing are at the heart of Virtualization. The logical functions of server, storage, network and software resources are separated from their physical constraints to create pooled resources. Business processes can share the same physical infrastructure. As a result, resources linked with one function, such as ERP, can be dynamically allocated to another, such as CRM, to handle peaks in demand. IT services can also be provided as a utility model, on a pay-per-use basis.

Virtualization is more than the implementation of technology. It's a new way of thinking about the IT infrastructure. To manage the environment as a whole, IT processes must be standardized and people educated on how to deliver service levels across a shared infrastructure.

#### 1.1.1 Consolidation

In many data centers, a small number of servers carry the bulk of the workload, while others run vastly under utilized, consuming your energy, time and resources.

Therefore a growing number of users have become interested in improving the utilization of their compute resources through consolidation and aggregation. Consolidation is already common concept in mainframe environments, where technology to support running multiple applications and even operating systems on the same hardware has been in development since the late 1960's. Such technology is now becoming an important differentiator in other markets (such as Unix/Linux servers), both at the low end (virtual web hosting) and high end (traditional data center server consolidation).

Virtualization technologies can help you achieve full asset utilization by identifying under performing assets and enabling asset consolidation. Consolidation means fewer assets to own and manage which in turn lowers the asset TCO.

#### 1.1.2 Virtualization

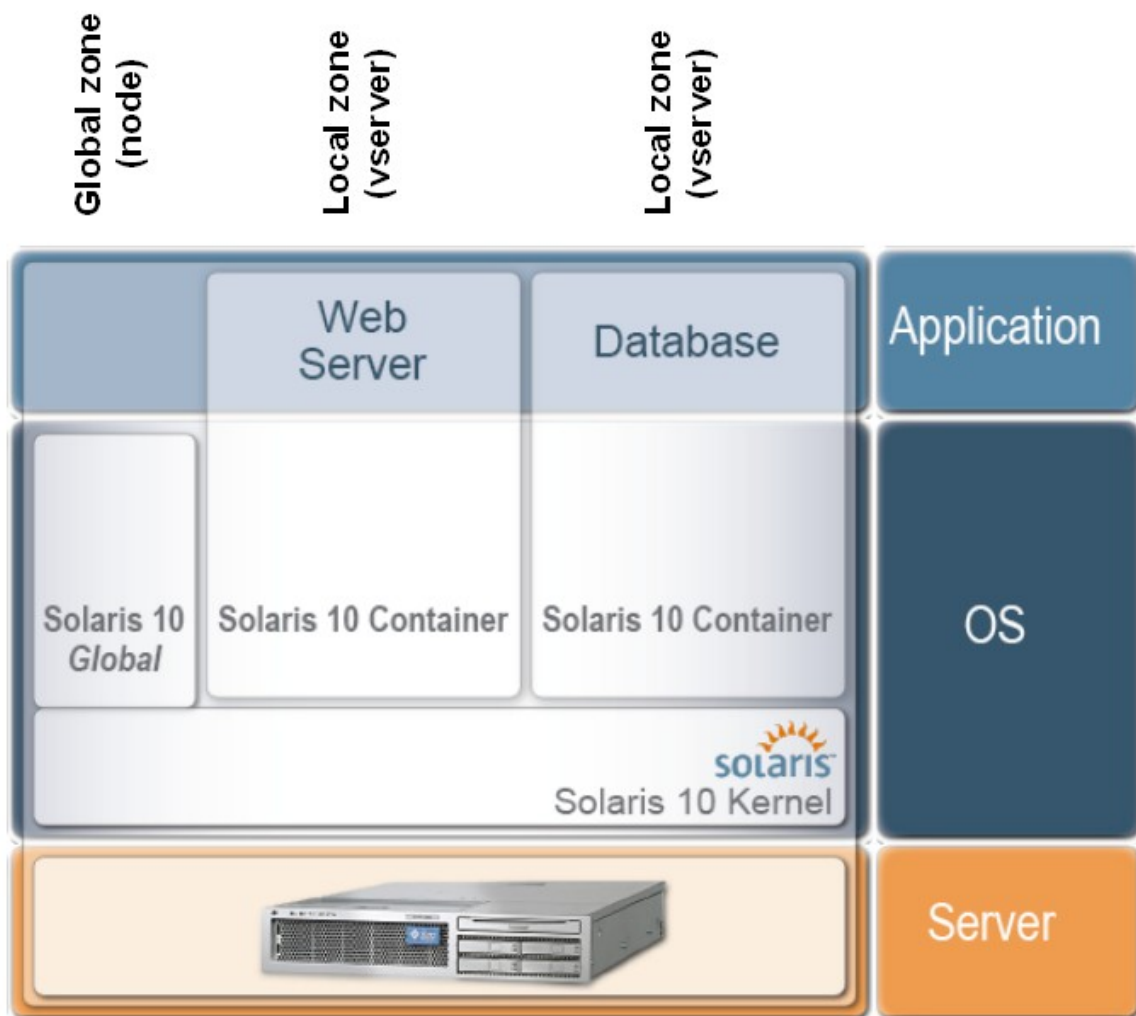
In computing terms, Virtualization is the creation of many digital abstractions that represent a real physical object.

So, in terms of servers, a virtual server may look like a single physical server to the end users and administrators. Each virtual server will be operate oblivious to the fact that it is sharing compute resources with other virtual servers. Virtual servers continue to provide the many benefits of their physical counterparts, only in a greatly reduced physical package.

Virtualization of the infrastructure addresses one of the most burning problems of today's data centers. It solves the dependencies between the numerous technology layers and creates transparency and flexibility. Resources will be administered in pools which are flexible to use and utilize.

### 1.1.3 Solaris Containers

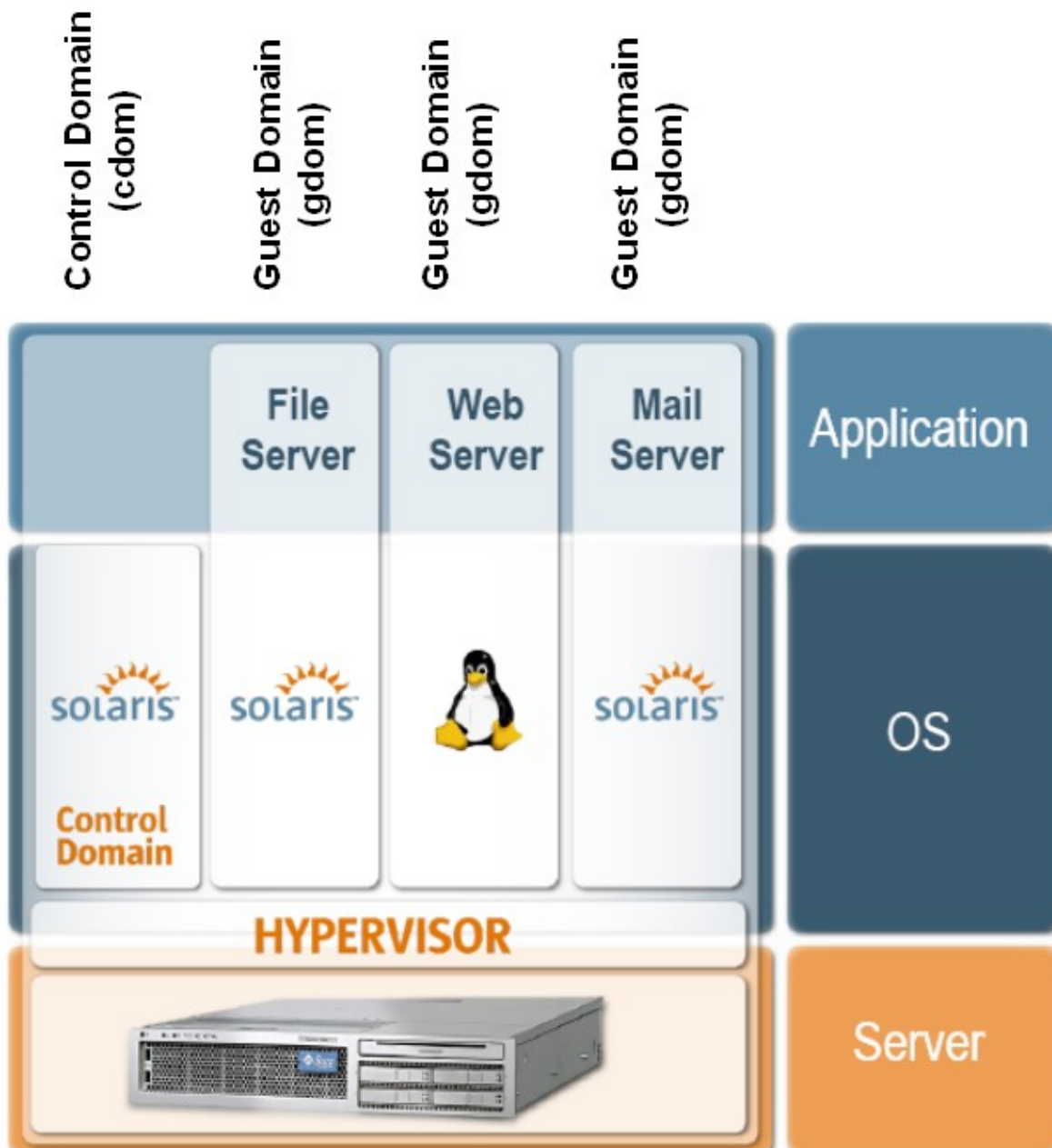
The VDCF vServer product builds on top of a Virtualization technology called Solaris Containers. A Solaris Container is logical abstraction of a Solaris application environment that can also reduce the overhead of administering multiple operating system instances. Each application running in a Container is isolated from what is happening in other Containers that may potentially be running within the same physical system. From an applications point of view, a Container looks exactly like a standard Solaris Operating Environment.



### 1.1.4 Solaris Logical Domains

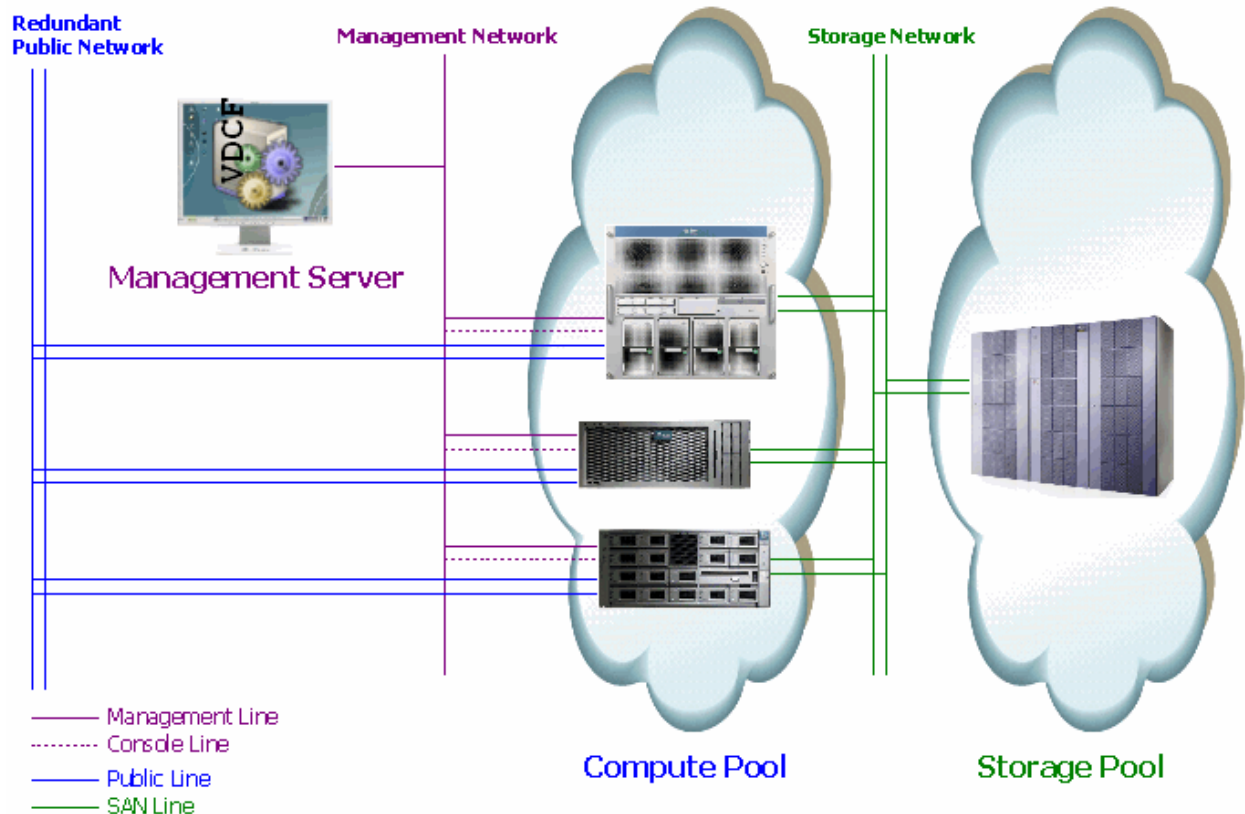
The VDCF LDom product is based on another Oracle Virtualization technology called Oracle VM Server for SPARC (previously called Sun Logical Domains). A logical domain (LDM) is a full virtual machine that runs an independent operating system instance and contains virtualized CPU, memory, storage, console, and cryptographic devices. Within the logical domains architecture, the Hypervisor is a small firmware layer that provides a stable, virtualized machine architecture to which an operating system can be written. As such, each logical domain is completely isolated and may run different Solaris Operating Systems. On each LDM server there is one control domain which controls and servers the Guest Domains. Guest Domains may contain Solaris Containers. From an applications point of view, a Guest Domain looks like a standard Solaris Operating Environment.

VDCF manages both, the control domain (cdom) and the guest domains (gdom).



### 1.1.5 Datacenter Architecture

Successful consolidation always relies on a standardized environment. VDCF follows a standard data center blueprint as a base to its architecture and design.



In the diagram above we show the generic data centers architecture complete with a management server, compute and storage pool. It also highlights the typical connections between the different entities. It separates management traffic from public and other data traffic. Data access is handled by the SAN and its associated fabrics and storage devices. A management server serves as a single point of control for the entire infrastructure.

#### Management Server

This system is the central operation cockpit to manage the Compute Server Pools. At a minimum it hosts the VDCF software but might be used for other system management products as well. The management server also serves as secure gateway to the Compute Pool infrastructure. It controls the access to the Compute Servers management interfaces and consoles.

#### Compute Pools

Services and applications run on virtual servers. Virtual servers are hosted on compute resources - servers - out of the compute pools.

#### Storage Pool

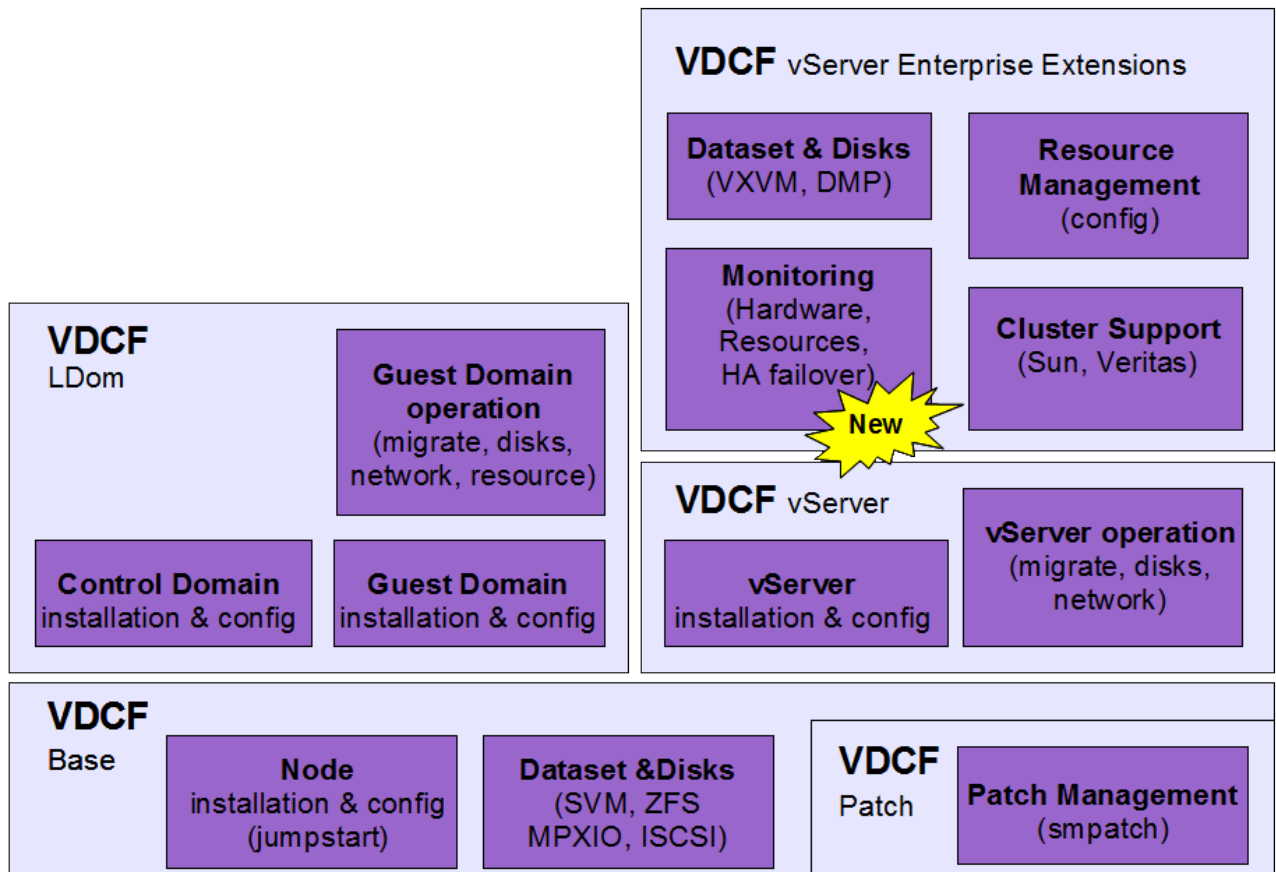
Stateful data like a virtual servers root and data filesystems are stored on SAN storage. The SAN storage serves LUN's to the compute pool. These LUN's must be visible on all or at least a subset of the physical servers. The accessibility of these LUN's on multiple physical servers is what enables VDCF to control the compute pool Virtualization.

### 1.1.6 Virtual Datacenter Control Framework (VDCF)

VDCF is a platform management framework for the Solaris Operating System. VDCF allows you to run a virtualized data center using Solaris 10 Containers and/or Logical Domains controlled by a centralized management server.

The VDCF product family consists of a common base framework and two separately available products. The VDCF vServer product is used to manage Solaris 10 Containers (Zones). Deployment of Logical Domains available on Oracle CMT Server is provided by the VDCF LDom product. To use both Virtualization technologies both products may be combined to get the most flexibility and benefit.

With VDCF, JomaSoft offers a tool to simply and effectively operate your Solaris 10 based virtual data center. On a central management server you create definitions and configurations, which are stored in the Configuration Repository. This information is then used by VDCF to populate physical servers with a Solaris build from which virtual servers or logical domains are created.



### 1.1.7 VDCF terminology

In order to facilitate the virtualized environment created and managed by VDCF, a specific terminology is applied. This terminology strictly separates the physical servers (global zone or control domain) or nodes from guest domains and virtual servers (non-global zone).

**Node:** The physical servers hardware plus the Solaris used as a global zone or as a control domain (CDOM).

A physical server is strictly used as a carrier for virtual servers and guest domains. A Node must remain stateless and might be re-installed at any time. VDCF is responsible for installing and tracking the currently active build of a particular node.

**LDom:** Control Domain (CDom) and Guest Domains (GDom).

The Control Domain is managing and serving the Guest Domains installed on the same physical server hardware. Guest Domains may be used as a node to carry vServers. VDCF is responsible for installing and tracking the currently active build of a particular Guest Domain.

**vServer:** The Solaris non-global zone.

The vServer is responsible for running the business applications. All state assigned to a particular application is contained within a vServer and its associated storage. A vServer either runs on a node directly or inside of a GDom. A vServer is built on top of at least one dataset which in turn hosts at least one filesystem carrying the configuration, programs and data for it.

**Dataset:** A storage abstraction used to manage LUN's in the volume-manager hierarchies.

A Dataset abstracts and standardizes the storage handling for vServers. Datasets use volume manager technology to create the required quality of service by grouping LUN's into different RAID (0,1,0+1) constructs. By default datasets are available in two different implementations. One uses Solaris Volume Manager (SVM) technology while the other implements on top of ZFS.

VDCF is installed in the global zone of a Solaris 10 server called the management server. In a highly available VDCF environment it may be installed in a non-global zone. From this server you install and operate your Nodes, vServers and GDoms.

The modular structure of the management server and the VDCF software makes it possible to flexibly adapt to individual customer's requirements. Extensions to the basic functionality can be simply realized by the means of adjustment and addition of individual modules.

## 1.2 Supported Environments

Currently the following System Environments are supported:

- Management Server Oracle SPARC Server and x86 Server
- Compute Node/Server Oracle SPARC Server and x86 Server
- Solaris Operating System Solaris 10 Update1 (1/06) up to Update 10 (8/11)
- Logical Domains LDoms 1.1/1.2/1.3/2.0/2.1
- Branded Zones solaris8, solaris9
- Volume Manager Solaris Volume Manager (SVM), ZFS
- Filesystem Solaris UFS, lofs, ZFS
- SAN / iSCSI Storage and HBA's compatible to  
SUN StorEdge SAN 4.4.x / Multipathing using STMS/MPXIO  
iSCSI Targets compatible to Solaris iSCSI Initiator
- Terminal Server Blackbox, Cyclades, IOLAN
- System Controller SC/ALOM, RSC, SSC, 15K, XSCF, ALOMCMT, ILOM, ILOMx86
- Network Link aggregation, IPMP and tagged VLAN for LDoms and vServer  
vServer exclusive ip-stack

For VDCF vServer Enterprise Customers the following Extensions are available:

- Resource Management Administration of vServer Resource settings
- Monitoring Hardware and Resource Monitoring, HA/Automated Failover
- Veritas Dataset Volume Manager: VXVM, Filesystem: vxfs
- Sun Cluster Integration of vServers in Sun Cluster
- Veritas Cluster Integration of vServers in Veritas Cluster

Other environments may only need small enhancements. Send us your request !

## 2 VDCF Features – Base

### 2.1 Builds and FLASH archives

A Build is a well defined set of Software packages used for provisioning a node with its required software. A Build contains only what is needed by the infrastructure and tends to be as small as possible. The process of assembling a Build out of a standard OS distribution is known as *minimization*. It allows for a lightweight and therefore more secure and faster OS install.

Builds also form the base for a standardized environment. All installations are done using a particular Build version. All systems installed with the same Build are known as having exactly the same OS software level. To facilitate this Build-Concept, VDCF uses the Solaris FLASH technology for provisioning a node. A FLASH archive therefore represents a specific Build. Multiple Build-Versions are kept in different FLASH Archives.

### 2.2 System Configurations

System Configurations are used to configure and customize your Nodes and Virtual Servers. The Configurations are stored in the Configuration Repository. They are applied automatically to the target systems during installation or afterward using the server configuration execution command.

You define the Base and Server Configurations before you install your systems. Base Configurations contain the configuration values. The Base Configurations are the reusable building blocks used to define the configuration information. The Connection between target systems and the Base Configurations is done during the addition of Server Configurations.

### 2.3 Node Discover & Configure

Before a physical server can be installed and operated using VDCF, it is necessary to determine the devices of the server. Information about the existing hardware is discovered, e.g. local disks, SAN LUN's, network interfaces, this is then loaded into the VDCF Configuration Repository.

For every Standard Platform (Solaris Server Model) a profile is created, where you define which disks and network interfaces are to be used. A Compute Node needs at least one root disk and optional, but recommended, a root mirror and a management network interface and IP address. Available network types are `management`, `public` and `backup`. Per type of network one IPMP Group is configurable.

### 2.4 Node Install or Register

When installing a Node, the first step is to assign an existing Build to the Node. The Solaris installation takes place on the local disks of the Node under control of ZFS or the Solaris Volume Manager (SVM). After the Solaris installation the System configurations are applied to the node, including the Remote Execution Environment (SSH). Finally the Node is registered in the Configuration Repository.

If a Node is defined as a CDom all necessary steps are taken to configure it as a control domain as defined.

Node Registration is the alternative solution to integrate existing Nodes into VDCF.

## 2.5 Node Operations

Once the node has been installed with a particular Build, it registers its presence within the framework. Remember that the node – the physical server – only acts as a carrier for the vServer. It manages the environment needed by vServers. However, because the node performs critical operations on behalf of the management server, it should normally not be required to log into the physical server. All day to day operational tasks on the nodes should be performed using the `VDCF node` command.

Available operations are: `boot`, `reboot`, `shutdown`

## 2.6 Node Evacuation

Node Evacuation is a new feature available since VDCF 4.0. The vServers of a Node are stopped and then distributed to the other available Nodes in the same compute pool. The Evacuation logic takes vServer categories and priorities into account and is based on their resource usage values. Using this Evacuation feature an overload of a Node is avoided.

## 2.7 Patch Management

Nodes and vServers can be patched by using the patch management functionality. Patches are rolled out in so called Patch-Sets. A Patch-Set contains a number of patches selectable through automatic system analysis or through manual selection. Patch-Sets might also be baselined, which allows inclusion of patches of a certain age. This prevents one from installing patches that are later withdrawn.

Patch-Sets are applied to a number of nodes by the means of Patch-Targets. A Patch-Target combines one or more nodes and one or more Patch-Sets. Installing a Patch-Target starts the installation of all patches listed in all selected Patch-Sets. If one or more of the patches being installed requires a reboot (non-standard patch), the target systems are taken down to single user mode for installation of the patches. After installation a reboot will be issued. If all patches being installed are standard patches (no reboot is required), the target will be applied in multi-user mode with no reboot. Installation of Patch-Sets will be launched in parallel on all specified nodes.

After installation the status of all installed Patch-Sets is tracked to show whether all patches in a set have been applied successfully and if any patch applications have failed.

Patch Management has also been enhanced to deal with Cluster Patching issues.

## 3 VDCF Features – LDom Product

### 3.1 GDom Definition & Installation

On top of a Node defined and installed as a CDom a number of GDom's can be installed. VDCF is used to define resources like LUN's used as a root device for the domain, network interfaces and IP addresses and the Solaris build used to install the domain. The GDom install command then installs the domain according to its definitions.

### 3.2 GDom Migration (detach, attach, migrate)

At installation time of a new GDom the performance of the physical node should be matched to the requirements. But after a while the requirements of the Applications may change or perhaps some maintenance work may be required on the physical node.

The VDCF framework reduces impact of such issues by being able to migrate a GDom including vServers from one node to another. The target Node must be configured with LDom Software to make use of this feature. Additionally both nodes must have access to the Disks the GDom and vServer are using. VDCF offers two types of migration: Live and Cold.

Live means that the GDom remains running while the migration takes place. There is no downtime for applications and vServers inside this GDom. Such live migrations are supported, if both control domains contain the same CPUs, run Solaris 10 Update 9 or later and LDOM software version 2.1 or later.

Cold migrations are executed by shutting down the GDom, detach it from the current CDom, attach to the target CDom and finally reboot it.

### 3.3 GDom Disaster Recovery

If one of your physical nodes goes out of service, you have two choices to recover the GDom's which were running on the physical node.

#### a) Reinstall the physical node

After a fresh install of your physical node each GDom is still in the "ACTIVATED" state. You have to detach the GDom's first and then re-attach them to the node.

#### b) Migrate the GDom to another existing node

Because your node isn't accessible at this time, you must execute a forced detach. This operation only updates the configuration repository. Using the attach operation the GDom is then migrated to another node. The old node however has the GDom's configured. Therefore to avoid conflicts you should not try to boot your old node again.

## 4 VDCF Features – vServer Product

### 4.1 vServer Definition & Installation

The first step in creating a new vServer sees it defined on the management server in the Configuration Repository. Every vServer requires at least one Dataset. On each Dataset file systems must also be created, which are of the type root, data or lofs. A minimum of one network interface configuration is also required, this includes a definition that distinguishes the IP address network type (management, public, backup). The Framework performs the allocation of network interfaces of the Node.

After completion of the configuration, the vServer is deployed to the Node by the "Commit" operation. At this time all the configured resources (Datasets, Filesystems, Networks) are created and the vServer is installed. The Solaris OS of the vServer is configured during the first system boot. A vServer can be installed directly on a Node or on a GDom.

### 4.2 vServer Operations

Apart from the run level functionality (boot, reboot and shutdown) the framework offers the possibility to modify the datasets, file systems and network interfaces. The administrator decides whether the changes are made in real-time on the running vServer or at the next reboot.

Datasets: Adding and Removing  
Filesystems: Adding, Growing, Renaming, Removing  
Network: Adding and Removing

### 4.3 vServer Migration (detach, attach, migrate)

At installation time of a new vServer the performance of the target node should be matched to the requirements. But after a while the requirements of the vServer may change or perhaps some maintenance work may be required on the node.

The VDCF framework reduces impact of such issues by being able to migrate a vServer and its application from one node to another. A vServer can also be migrated from a Node to a GDom and vice versa. The nodes usually must be installed with the same build and have the same patch level to make use of this feature. Starting with Solaris 10 10/08 it is possible to upgrade a vServer while it is attaching to its new Node or GDom. Additionally the new target Node or GDom must have access to the Datasets the vServer uses.

### 4.4 vServer Disaster Recovery

If one of your compute nodes goes out of service, you have two choices to recover the vServers which were running on the compute node.

a) Reinstall the compute node

After a fresh install of your compute node each vServer is still in the "ACTIVATED" state. You have to detach the vServer first and then re-attach them to the node.

b) Migrate the vServer to another existing node

Because your node isn't accessible at this time, you must execute a forced detach. This operation only updates the configuration repository. Using the attach operation the vServer is then migrated to another node. The old node however still references the Datasets and has the vServers configured. Therefore to avoid conflicts you should not try to boot your old node.

## 5 VDCF Features – vServer Enterprise Extensions

### 5.1 vServer Resource Management

Once a vServer has been deployed on a particular node, it might be required to control the amount of resources dedicated to it. As of Solaris 10 8/07 (Update 4) it is possible to define Resource Controls for non-global zones. The full set of these Resource Controls can be centrally managed using VDCF. The basic functionality has been enhanced to allow for Utility Based Computing models. To do so, all Nodes are rated based on their performance. A Nodes relative performance is expressed in Base Units. A vServer in turn gets a fraction of its driving Nodes Base Unit assignment. This assignment is used to configure the Fair Share Scheduler (FSS) with the required number of shares for a particular vServer. Migrating a vServer from one Node to another with different performance characteristics, will automatically adjust its FSS settings to make sure the vServer gets the resources it is entitled to.

### 5.2 Hardware, Resource and High-Availability Monitoring

VDCF Monitoring consists of a hardware (physical nodes), a resource and a High-Availability monitoring component.

Hardware monitoring is used to periodically check the state of the physical nodes (hardware and OS). This state is extracted from the information provided by the system controller of all nodes defined in the VDCF repository. If the monitor encounters a hardware fault alarming by email or a customized shell script is provided.

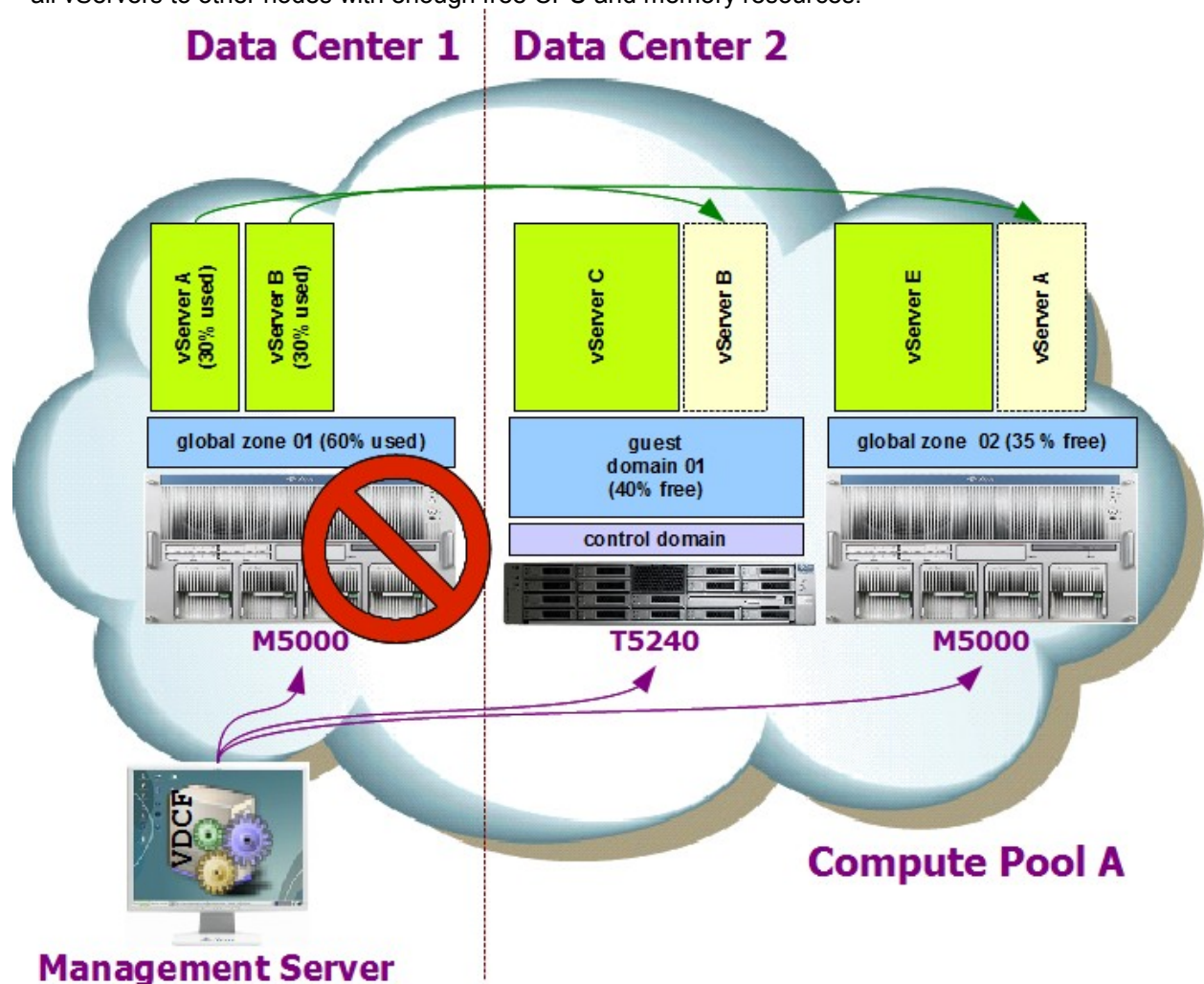
Resource Monitoring is used to collect systems resource (CPU and memory) usage of Nodes and vServer. The data collected is stored within the VDCF repository and it's possible to see the resource usage per node/vServer in a current (hourly,daily) or aggregated (monthly,yearly) view.

The High-Availability Monitoring is used to monitor the health of nodes. It is based on a VDCF service which runs on the Nodes and sends keep-alive messages to the VDCF management server. If the keep-alive messages are missing the system controller is check to detect a node failure. A complementary network probing may be activated.

### 5.3 High-Availability - automated failover

Despite the ability to migrate vServers from one Node to another, it may be required to automate this process for better availability. To achieve these HA requirements VDCF has a build-in HA monitoring component.

If a failed node is detected the node may be stopped and/or the node evacuation logic is called to failover all vServers to other nodes with enough free CPU and memory resources.



### 5.4 High-Availability integration into Cluster Software

Despite the ability to migrate vServers from one Node to another, it may be required to automate this process for better availability. To achieve these HA requirements VDCF integrates with existing Cluster Technology. In such a scenario, VDCF is still the central point of management. It is still used to create, install and control Nodes and vServers. However, the fact that the hosting Node for a particular vServer is a Cluster Node, invokes special treatment for that vServer. VDCF will hand over the monitoring and migration tasks to the Cluster Framework. This allows for automatic failover within seconds if the cluster detects a failure condition. Manual failover might be triggered using VDCF or the respective Cluster interfaces.

Currently VDCF integrates with either Sun Cluster 3.2/3.3 or Veritas Cluster 5.0

## 6 VDCF customers

These companies or institutions are using VDCF (among others):



Alpiq Holding AG  
Bahnhofquai 12  
4601 Olten

Lucerne University of  
Applied Sciences and Arts

**HOCHSCHULE  
LUZERN**

Lucerne University of Applied Sciences and Arts  
Enterprise Lab  
Technikumstrasse 21  
6048 Horw

**rtc.ch**

IT-Outsourcing & Banking Software

RTC Real-Time Center AG  
Schwarzburgstr. 160  
3097 Liebefeld



Swisscom IT Services AG  
Schochengasse 6  
9000 St. Gallen



Wegelin & Co. (Privatbankiers)  
Am Bohl 1  
9001 St. Gallen



**Zürcher  
Kantonalbank**

Zürcher Kantonalbank  
8005 Zürich



## 7 About JomaSoft GmbH

JomaSoft GmbH is a IT company founded in 2000 and we are located in St.Gallen (Switzerland).

We specialize in software engineering for Java and in Virtualization of Unix environments.

Our in-depth technology knowledge is born out of the good partnerships with Oracle and Sun Microsystems.

### Contact

Would you like to learn more about JomaSoft?  
Send us an eMail

<http://www.jomasoft.ch>  
[info@jomasoft.ch](mailto:info@jomasoft.ch)

## 8 More about VDCF

See our product page for more details about VDCF <http://www.jomasoft.ch/products/VDCF/docs>

If you are interested in VDCF, we recommend a VDCF Proof of Concept (POC) installation in your environment.

Our POC package includes a trial license for two months and two consulting days On-Site for VDCF installation and education.