

# **LDom Performance optimieren**

**Marcel Hofstetter  
JomaSoft GmbH  
St. Gallen / Schweiz**

## **Schlüsselworte**

Virtualisierung, SPARC, T4, T5, LDom, Oracle VM Server for SPARC, VDCF

## **Einleitung**

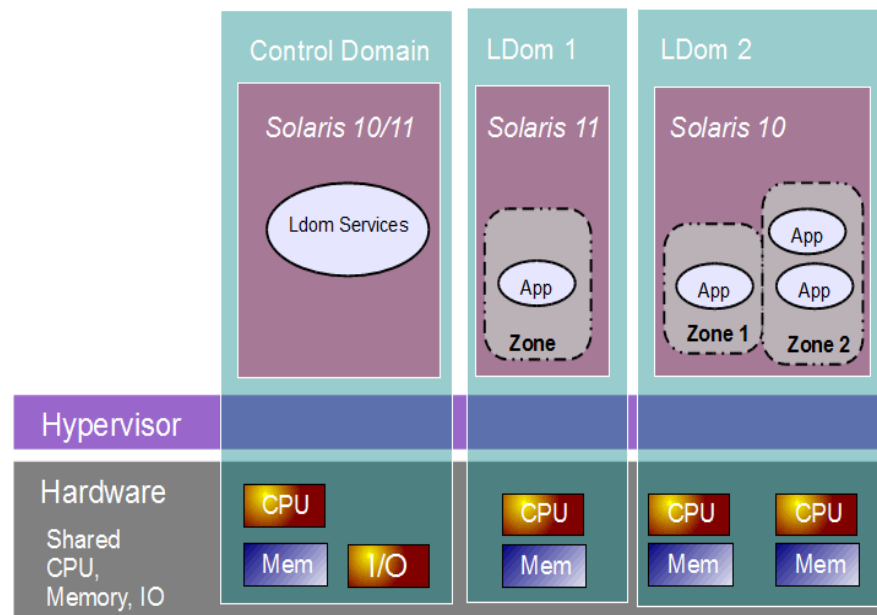
Die aktuellen Oracle SPARC T4/T5-Server sind ideale Plattformen um Applikationen unter Solaris 11 zu konsolidieren und alte Server abzulösen. Diese (CM)T-Systeme verfügen über die von der Hardware unterstützte „Logical Domains“ Technologie, mit welcher man mehrere, voneinander unabhängige Solaris Instanzen auf demselben Server betreiben kann. Dieser Vortrag enthält eine kurze Einführung in die LDom Technologie und fokussiert anschließend auf das Thema Performance. Welche Technologien bringen Vorteile und wo liegen die Einschränkungen damit. Welche „Best Practices“ bereits von unserem Management Produkt VDCF (Virtual Datacenter Control Framework) unterstützt werden, zeigen wir zum Abschluss des Vortrages.

## **JomaSoft GmbH**

Die JomaSoft wurde als Software Unternehmen im Jahr 2000 gegründet. Als Oracle Gold Partner sind wir insbesondere für Oracle Solaris 11, SPARC T4 und T5 Server spezialisiert. Wir bieten Software-Entwicklung, Consulting, Implementation und Administration im Bereich Solaris. Abgerundet wird unser Angebot durch das Produkt VDCF. Ein Framework, welches die Installation, das Management und Disaster Recovery von Solaris Zonen und LDom vereinfacht und automatisiert. Dieses Framework wird bei zahlreichen Kunden in Europa seit mehr als 7 Jahren produktiv eingesetzt.

## LDoms (Oracle VM Server for SPARC)

"Logical Domains" sind wie die Bezeichnung andeutet logische Domänen, welche vom Hypervisor der SPARC-T-Systeme unterstützt sind. Voraussetzung für diese Technologie ist ein Oracle SPARC Server der T-Serie (CMT System). Die LDom Software ist im Solaris 11 integriert und kann für Solaris 10 kostenlos bezogen werden. In jeder logischen Domäne (LDom) kann eine unabhängige Solaris Instanz installiert werden. Via Control Domain werden die LDoms verwaltet. Die Control Domain stellt virtuelle Devices und Services bereit, welche von den LDoms für Zugriff auf Storage und Netzwerk genutzt werden. Ressourcen wie CPU und Memory werden den LDoms fix zugewiesen, können aber auch zur Laufzeit verändert werden. Seit LDOM Version 2.1 können LDoms ohne Unterbrechung zwischen Systemen migriert werden (Live Migration).



### Control Domain (primary)

Die Control Domain ist die Solaris Instanz, welche direkt auf der Hardware läuft. Der LDom Manager in der Control Domain ist zuständig für die Konfiguration der LDoms. Die Control Domain benötigt selbst Hardware Ressourcen, damit sie ihre Arbeit optimal durchführen kann. Die empfohlene Start-Konfiguration besteht aus 2 CPU Cores und 8 GB RAM. Wer neuere LDom Features verwenden möchte (Wer will das nicht?) installiert die Control Domain mit Solaris 11.1 oder neuer. Ein erwähnenswertes Feature, welches unter Solaris 10 nicht verfügbar ist, ist die Console History für LDoms.

### **Virtuelle Devices**

Die Control Domain kann den LDom (oder Guest Domains) für I/O Zugriffe virtuelle Devices zur Verfügung stellen. Damit erhalten die LDom virtuelle Disks (vdisk) und virtuelle Netzwerk Interfaces (vnet). Die I/O Requests werden dann via Control Domain verarbeitet. Es liegt auf der Hand, dass diese Verarbeitung via Control Domain etwas Zeit in Anspruch nimmt und daher nicht die allerbeste Performance liefert. Für Applikationen mit durchschnittlichem Anspruch an die I/O Performance sind virtuelle Devices aber auf jeden Fall ausreichend. Für Applikationen, welche die bestmögliche I/O Performance benötigen, müssen Technologien verwendet werden, welche weniger Overhead verursachen und näher an der Hardware sind.

Der Einsatz von virtuellen Devices ist die einfachste und flexibelste Art der LDom Konfiguration. Daher empfehlen wir primär diese Konfiguration und nur wenn die Performance unbefriedigend ist, auf andere Konfigurationen auszuweichen. Das JomaSoft VDCF Management Framework setzt per Default ebenfalls auf diese virtuellen Devices.

### **Live Migration**

Nur LDom, welche ausschließlich virtuelle Devices verwenden, können ohne Unterbruch von einem Server auf einen anderen migriert werden. Nur durch die virtuellen Devices sind die LDom von der Hardware entkoppelt. Zudem ist es teilweise notwendig und auf jeden Fall zu empfehlen die SPARC Server auf demselben Firmware Level zu betreiben. Live Migration kann nützlich sein, um Downtime von Applikationen zu vermeiden. Um beispielsweise die Firmware eines Server anzuheben. Software Lizenzbedingungen können teilweise die Nutzung von Live Migration beschränken oder verbieten, bzw. die Lizenzierung verteuert sich, wenn Live Migration benutzt werden soll.

### **Virtual Functions / SR-IOV**

Neuere I/O Karten (Ethernet und FibreChannel) enthalten virtuelle Funktionen (VF), welche für LDom die direkte, performante Nutzung der I/O Ports ermöglichen. Dabei kann der Overhead mit der Verarbeitung via Control Domain eliminiert werden. LDom welche solche VF benutzen, sind von der Hardware abhängig und können deshalb nicht Live auf einen anderen Server migriert werden. VF sind von den älteren T-Systemen (T1,T2) nicht unterstützt. Die Anzahl VF pro Port ist beschränkt (7-63).

### **Direct I/O (DIO) und root I/O Domains**

Die höchste Performance wird erreicht, wenn die Hardware direkt verwendet wird. Dazu stehen zwei Möglichkeiten offen. Direkte Anbindung eines kompletten PCI Bus an eine LDom, welche damit zur root I/O Domain wird. Oder die Konfiguration einer PCI Karte an eine LDom. Die Anzahl PCI Slots eines Servers begrenzt die Anzahl LDom, welche damit aufgebaut werden können. LDom mit solcher direkter Hardwareanbindung sind damit an die Control Domain gebunden. Die LDom Technologie wird hier weniger zur flexiblen Virtualisierung angewendet, eher zur Aufspaltung von großen Servern in mehrere „statische“ Domains.

### **Disk Performance generell**

Auch für Logical Domains gelten die gängigen Empfehlungen für gute Disk Performance. Zusätzliches Memory für die LDom kann das Caching Verhalten verbessern. Verwenden sie mehrere Disks um die Anzahl parallele Requests zum Storage zu erhöhen (Striping). Redo Logs auf separate Disk legen. Bei der Verwendung von SAN Storage und ZFS, kann die Erhöhung des Kernel Parameter `zfs_vdev_max_pending` (Default: 10) die Performance zudem verbessern.

## Virtualisieren mit VDCF

Als wir 2005 begannen mit Solaris Zonen zu arbeiten, war uns schnell bewusst, dass dynamische Virtualisierung ohne zentrales Werkzeug im Chaos enden würde. Deshalb implementierten wir ein kleines Framework, welches inzwischen ziemlich umfangreich wurde und auch Logical Domains gut unterstützt.

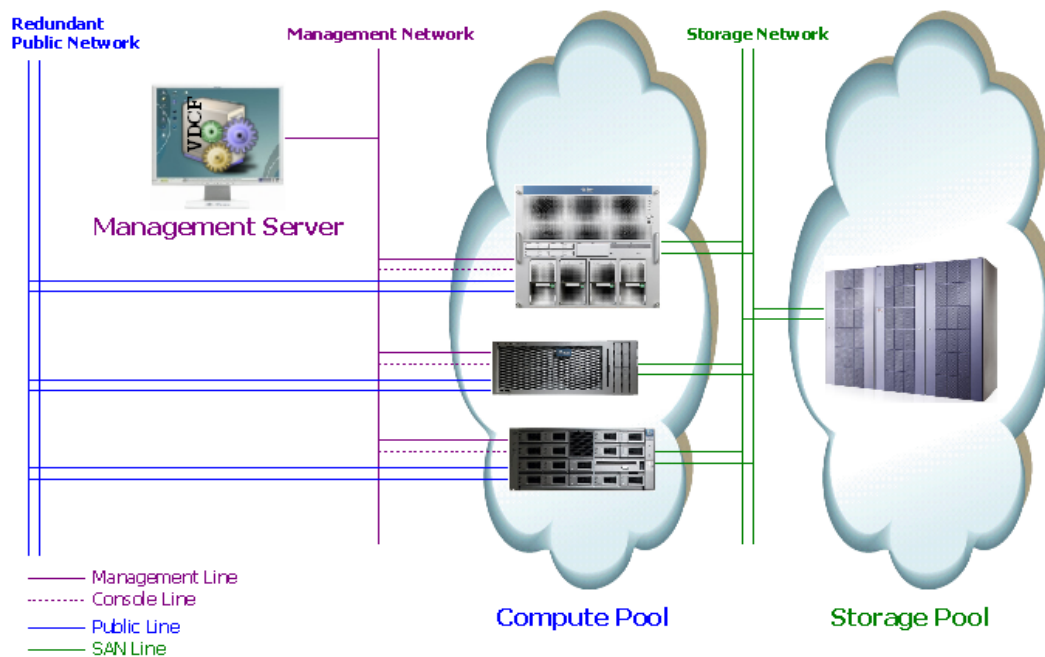
Folgende Features sind in der aktuellen Version 5 für Solaris 10 und Solaris 11 enthalten:

- Installation von physischen Servern
- Automatisierte Software Deployments (File-Copy, Scripts und Packages)
- Patch Management (Installation und Vergleiche)
- Package Management (Analyse, Suche und Vergleiche)
- Zonen Installation, Management und Migration
- LDom Installation, Management und Migration
- Überwachung von Hardware-Komponenten und Ressourcen-Auslastung
- Unterstützung von Veritas Filesystem, Veritas Cluster und Solaris Cluster

Mit VDCF werden derzeit bei zahlreichen Kunden ca. 3000 produktive Zonen verwaltet.

## Architektur

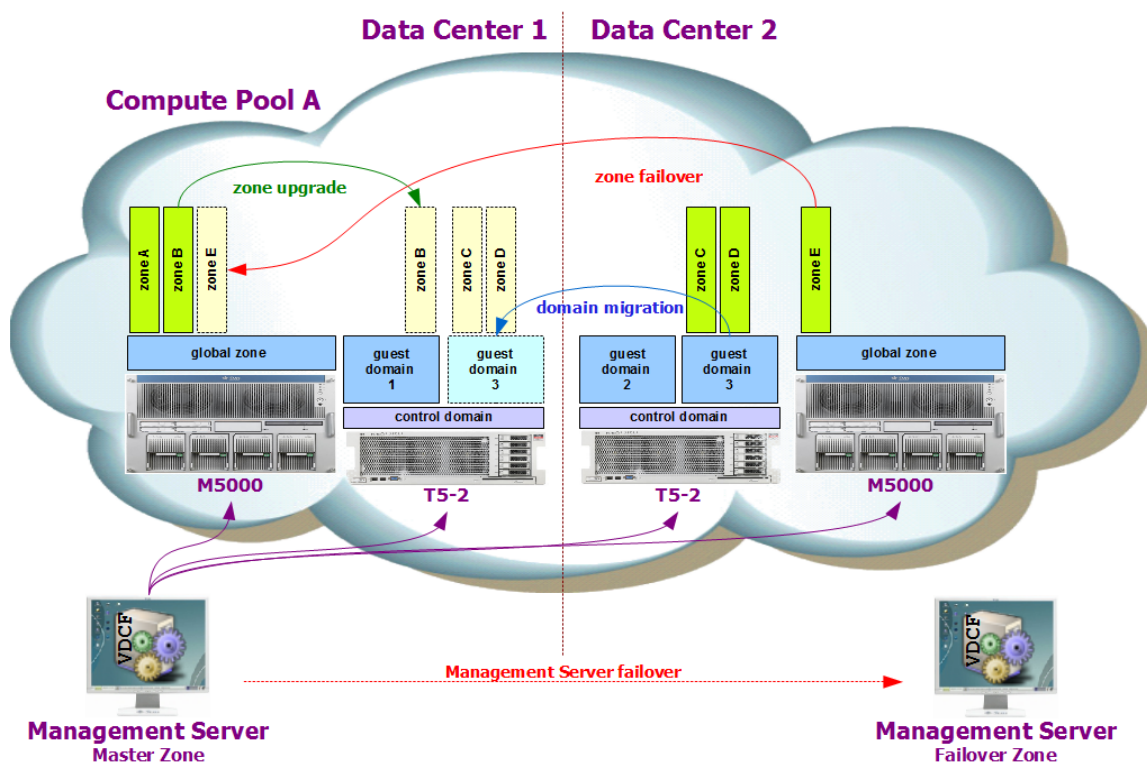
Das Virtual Datacenter Control Framework (VDCF) basiert auf einer Standard Architektur mit zentralem Management Server (mit Datenbank) und Solaris Systemen angeschlossen an einem zentralen Storage. Auf den lokalen Disks der physischen Systeme wird nur Solaris als Laufzeitumgebung installiert. Alle Daten liegen auf einem zentralen StoragePool (SAN). In einem ComputePool werden gleichartige Systeme zusammengefasst. Beispielsweise Systeme im selben Netzwerk mit Zugriff auf dieselben SAN-LUNs. Zonen- und LDom-Migrationen sind nur innerhalb desselben ComputePool möglich.



VDCF unterstützt Zonen und LDomS für Solaris 10 und Solaris 11. Die primären Ziele sind **Einfachheit, Standardisierung und Verfügbarkeit**.

Mittels einer Handvoll intuitiver Befehle können virtuelle Solaris Umgebungen auch von unerfahrenen Solaris Administratoren erstellt und betrieben werden. VDCF baut auf den existierenden Solaris Technologien auf. Virtuelle Server können innerhalb von Minuten anstatt Stunden in einer standardisierten Form bereitgestellt werden. VDCF erlaubt das manuelle und automatische Auslösen von virtuellen Server Migrationen. Dies führt zu verbesserter Verfügbarkeit und schnellem Disaster Recovery.

Die folgende Abbildung zeigt die Möglichkeiten mit VDCF. Mit VDCF wird es zum Kinderspiel vServer nach einem Hardware Ausfall auf ein Failover System zu migrieren. Wenn eine Guest Domain (LDom) Solaris Zonen enthält, werden alle Aktionen ausgeführt um auch diese Zonen vollautomatisiert auf ein zweites System zu migrieren.



Die gesamte VDCF Produkt Dokumentation ist öffentlich und eine frei verfügbare Test-Version "VDCF Free Edition" ist auf unserer Website ebenfalls zu finden: <http://www.jomasoft.ch/vdcf>

**Kontaktadresse:**

Marcel Hofstetter  
JomaSoft GmbH  
Falkensteinstrasse 54a  
CH-9000 St. Gallen



Telefon: +41 (0)71-288 92 11  
Fax: +41 (0)71-288 92 12  
E-Mail: [hofstetter@jomasoft.ch](mailto:hofstetter@jomasoft.ch)  
Internet: <http://www.jomasoft.ch>