Einfacher Aufbau von verfügbare Architekturen mit Oracle Solaris 11

Heiko Stein
etomer GmbH
Berlin

Detlef Drewanz
Oracle Deutschland B.V. & Co. KG
Potsdam

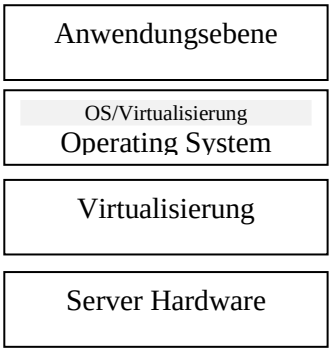
Schlüsselworte:
Oracle Solaris 11, Verfügbarkeit, Solaris Zonen, IPMP, MPXIO

Einleitung
Die Verfügbarkeit von Anwendungen auf einfachen Architekturen (Single Node, kein Cluster) kann durch Maßnahmen des Betriebssystems und der Auswahl der richtigen Hardware verbessert werden. Dazu ist eine gute Kenntnis der verwendeten Plattform-Architektur notwendig.

Dieser Vortrag erläutert ausgehend von den grundlegenden Anforderungen von Anwendungen die Möglichkeiten von Oracle Systemen beim Aufbau von verfügbaren Architeturen. Dazu zählen die Bereiche der Hardware; wie das Bussystem, die IO-Anbindung, Disks und Netzwerke, sowie die Funktionalitäten von Oracle Solaris. Für Solaris werden Funktionalitäten und Anwendungsbeispiele für das Service Management Facility (SMF), die Fault Management Architektur (FMA), MultiplexIO (MPXIO), Netzwerk-Aggregationen/Trunking/Bonding, IP-Multipathing (IPMP), Virtual Router Redundancy Protocol (VRRP) und dem Oracle Solaris 11 Load Balancer (ILB) vorgestellt und diskutiert.

Anforderungen von Anwendungen an einfache Architekturen

Die Anforderungen von Anwendungen an Architekturen können sehr unterschiedlich sein und richten sich vor allem nach SLA (Service Level Agreements). Ein wichtiger Aspekte dabei ist die Verfügbarkeit der Anwendung oder des realisierten Service. Diese Verfügbarkeit richtet sich jedoch nicht nur nach der Uptime, sondern auch der verfügbaren bzw. garantierten Leistung eines Service. Auch wenn eine zugesagte Leistung nicht in vollem Umfang erbracht wird, kann die Anwendung als nicht verfügbar betrachtet werden.



Die herkömmliche Art Verfügbarkeit zu realisieren, erfolgt oft durch den Aufbau einer redundanten Architektur, die durch eine Software (Clustering) abgesichert wird. Dabei können die Teile der redundanten Architektur auch örtlich getrennt voneinander installiert sein. Nicht immer kann oder muß jedoch eine Clustersoftware zum Einsatz kommen. Auch ohne Cluster lassen sich in Grenzen verfügbare Architekturen realisieren. Die Verfügbarkeit einer Architektur ergibt sich dabei jedoch aus dem Produkt aller beteiligten Schichten. Selbst eine ausgezeichnete Anwendungsschicht kann eine ungenügend verfügbare Architektur wieder ausgleichen, jedoch kann eine gute Architektur die Realisierung einer hohen Verfügbarkeit der Anwendung unterstützen. Das folgende Bild zeigt dies schematisch.

Bereits Betriebssysteme und Hardware verfügen über verschiedene Eigenschaften, die beim Aufbau von verfügbaren Architekturen ohne Cluster helfen können. Im folgenden werden die vier Ebenen einer genaueren Betrachtung in Bezug auf den den Aufbau von Architekturen mit Oracle SPARC-Systemen und Oracle Solaris unterzogen. Die Oracle Server Hardware und Oracle Solaris zeichnen sich hier durch eine Vielzahl von Eigenschaften aus, die bereits vorhanden sind. Zusätzlich sind die Eigenschaften von Oracle VM Server for SPARC (LDoms) und die Virtualisierungseigenschaften des Betriebssystems durch Solaris Zonen zu betrachten.

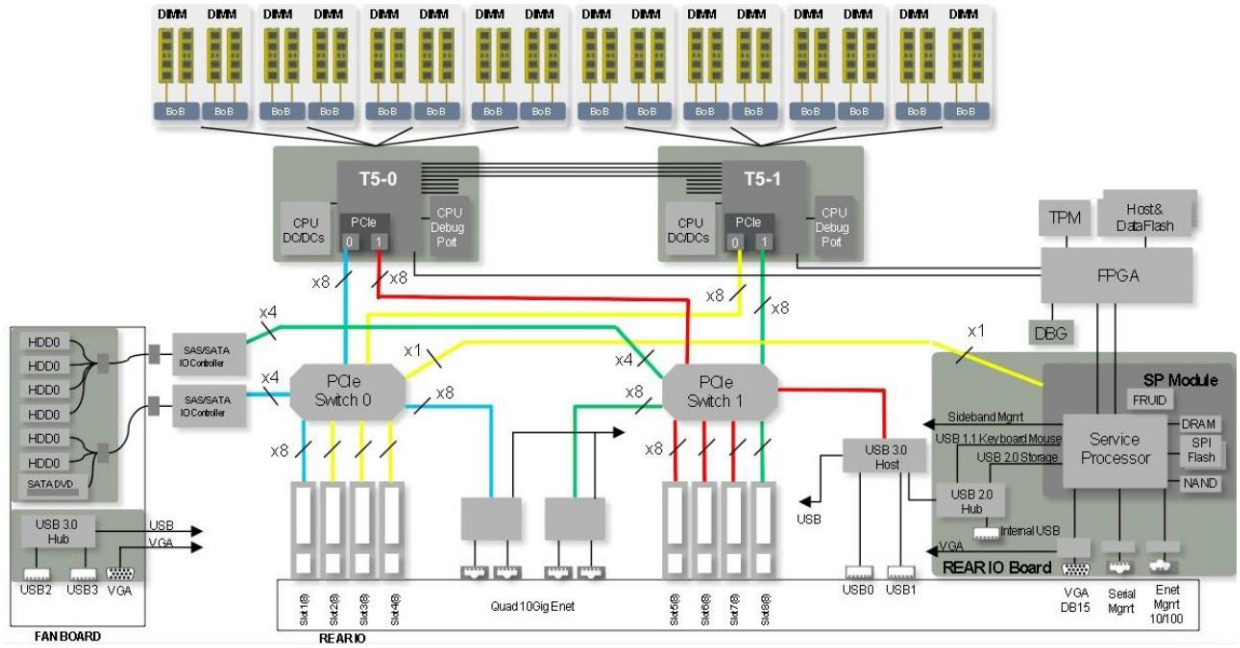
Hardware: Verfügbarkeit und Oracle SPARC Server

Vor dem Hintergrund des Aufbau von hochverfügbaren Architekturen mit immer leistungsfähigerer Software wandert die Bedeutung der benutzten Hardware immer weiter aus dem Fokus. Vielfach wird davon ausgegangen, dass die verwendete Hardware keine besonderen Eigenschaften aufweisen muss, da die Software die wichtigen Aspekte beisteuert. Wie oben jedoch beschrieben, leistet die verwendte Hardware einen wichtigen Beitrag zur Realisierung von verfügbaren Architekturen. So stellt die richtige Hardware bereits eine gewisse Redundanz auf der Hardware-Eben bereit oder stellt die erforderlich Leistungsfähigkeit zur Verfügung, um auf Anwendungsebene die geforderte Leistung zu liefern und den Ablauf von hochverfügbaren Softwarekomponenten überhaupt erst zu ermöglichen.

Aktuelle Serversysteme verfügen im Allgemeinen nur über mehrere interne Festplatten oder redundante Lüfter. Die Kenntnis der verwendeten Hardware für auch zu Konsequenzen bei der Auswahl von Erweiterungskarten und den entsprechenden Steckplätzen. Aktuelle Datenblätter und Whitepaper geben Aufschluß über weiterführende Details:

- Welche Festplatten sind über welche I/O-Kanäle mit welchen Disk-Controllern verbunden.
- Gibt es redundante I/O-Kanäle ?
- Welche Leistungsfähigkeit haben die I/O-Kanäle ?
- Welche Steckplätze nutzen welche I/O-Kanäle ?
- Über welche I/O-Kanäle laufen ggf. vorhandene interne Netzwerkports und Festplattencontroller ?

Das folgende Bild zeigt beispielhaft die Zuordnung redundanter I/O-Pfade zu IO-Slots einer SPARC T5-2 im Blockdiagramm.

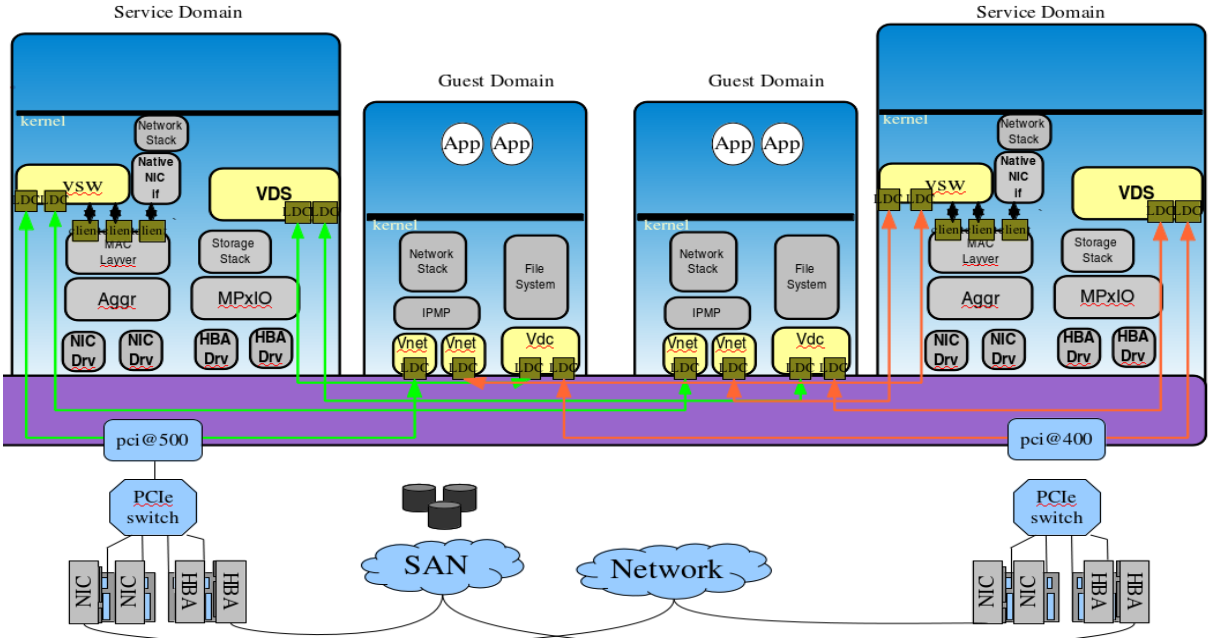


Wer z.B. nach dem obigen Bild eine redundante und performante Storage Anbindung realisieren muss, sollte die beiden benötigten FC-Controllern möglichst auf beide PCIe Switches verteilen.

Virtualisierung: Verfügbarkeit und Oracle VM Server for SPARC

Oracle VM Server for SPARC stellt Virtuelle Maschinen (oft auch LDoms genannt) als Laufzeitumgebungen für Solaris zur Verfügung. Auch hier kann es aus Gründen der Verfügbarkeit notwendig sein, über redundante IO-Pfade für Netzwerk oder Festplatten zu verfügen.

Bei der Verwendung von Direct-IO oder Root-Domains werden die entsprechenden IO-Slots der Domain zugeweisen. Guest Domains verwenden virtuelles IO, das durch Service Domains zur Verfügung gestellt wird. Werden hier redundante IO-Pfade benötigt, so sind redundante Service Domains aufzusetzen. Diese nutzen jeweils eigene IO-Slots und PCIe-Switches und stellen diese über voneinander unabhängige Service Domains den Guest Domains zur Verfügung. Die Steuerung der Alternativen IO-Pfade erfolgt im Allgemeinen in den Guest Domains selbst. Das folgende Bild zeigt schematisch den Aufbau von redundanten Service Domains.



Operating System: Oracle Solaris 11 und die Fault Management Architecture

Oracle Solaris verfügt über eine Reihe von Funktionalitäten zur Realisierung von verfügbaren Architekturen. Dabei geht es jedoch nicht nur um die Erzeugung redundanter IO-Pfade und Services, sondern auch um die Überwachung der benutzten Hardware und Services.

FMA (Fault Management Architecture) empfängt durch den Solaris Fault Manager Daten über Hardware- und Software-Fehler und diagnostiziert mögliche Ursachen. Nach der Diagnose werden ggf. fehlerhafte Komponenten abgeschaltet oder nicht weiter benutzt. Das folgende Beispiel zeigt die Benutzung der FMA Kommandos zur Fehleranzeige: `fmdump, fmadm faulty`

- Anzeige aufgetretener Fehler

```
# fmdump
TIME                 UUID                                 SUNW-MSG-ID EVENT
Aug 18 00:22:31.7928 7cdf8206-9931-4e14-ed45-bb0a34391d2b SUNOS-8000-KL Diagnosed
```

- Anzeige weiterer Fehlerdetails

```
# fmadm faulty -u 7cdf8206-9931-4e14-ed45-bb0a34391d2b
```

- Fehleranzeige im Detail

```
# fmdump -Vp -u 7cdf8206-9931-4e14-ed45-bb0a34391d2b
```

Operating System: Oracle Solaris 11 und die redundante IO-Anbindung

Eine redundante IO-Anbindung kann in Solaris über zwei Formen realisiert werden:

- Disk
- Netzwerk

Die redundante Anbindung von Disk-IO in Solaris erfolgt für FC, SAS und iSCSI (IO-Multipathing).

- Einschalten

```
# stmsboot -e
```

- Ausschalten

```
# stmsboot -d
```

- Konfigurationsdateien
  - `/kernel/drv/fp.conf; /kernel/drv/scsi_vhci.conf`
- Funktionskontrolle

```
# mpathadm list lu
# luxadm display …
```

Eine redundante Netzwerkanbindung kann in Solaris über zwei unterschiedliche Formen realisiert werden:

- Auf Datalink Ebene (IEEE 802.3ad Aggregates oder Datalink Multipathing)
  - Einschalten

```
# dladm create-aggr …
```

  - Ausschalten

```
# dladm delete-aggr …
```

  - Konfigurationsdatei
    - `/etc/dladm/*`
  - Funktionskontrolle

```
# dladm show-aggr …
# dlstat
```

- Auf IP-Stack Ebene (IP Multipathing)
  - Einschalten

```
# ipadm create-ipmp …
```

  - Ausschalten

```
# ipadm delete-ipmp …
```

  - Konfigurationsdatei
    - `/etc/ipadm/*`
  - Funktionskontrolle

```
# ipmpstat …
```

Zusätzlich können im Netzwerkumfeld verschiedene Steuerungsmechanismen zur Sicherung von Performance angewendet werden:

- CPU Ressourcen für Netzwerkinterfaces
- Bandbreitenmanagement für Interfaces
- Prioritätssteuerung für Interfaces

```
# dladm set-linkprop …
```

- Flows

```
# flowadm …
```

Datalink Multipathing und IP Multipathing realisieren redundante Netzwerkverbindungen einzelner Server. Eine weiterführende Redundanz ergibt sich durch die Konfiguration redundanter Router. Das Virtual Router Redundancy Protocol (VRRP) unterstützt bei Gruppierung von Routern im Internet zur Ermittlung redundanter Routen (`vrrpadm`).

Eine Menge von gleichartigen Diensten kann einem Client über einem Load Balancer wie ein Dienst präsentiert werden. Der Load Balancer sorgt bei Dienstanfragen für eine ausgewogene Auslastung der einzelnen Dienstanbieter und toleriert den Ausfall einzelner. Oracle Solaris stellt einen software-basierten Load Balancer zur Verfügung, der bei richtigem Einsatz zur Erhöhung der Verfügbarkeit der Gesamtarchitekturen führt, da der Ausfall einer Einzelkomponente toleriert wird (`ilbadm`).

Operating System: Oracle Solaris 11 und das Service Management Facility

Das Service Management Facility (SMF) stellt in Solaris eine Infrastruktur bereit, über die Dienste gestartet, gestoppt, überwacht und bei ungeplanter Beendigung automatisch neu gestartet werden können. Dieses wird über sogenannte Process Contracts (im Kernel realisiert) erreicht, in denen die Services gestartet und überwacht werden. Eigenschaften der Services werden über Service Properties in Service Manifesten definiert.

Wenn SMF die Ausführung und den Ablauf von Services überwacht und Services bei Ausfall automatisch neu startet, so hilft diese Funktionalität bei der Sicherstellung von Netzwerkservices. Es kann zwar keine redundante Ablaufumgebung zur Verfügung gestellt werden, jedoch wird ein lokaler Dienst oder eine Gruppe von Diensten aktiv überwacht.

- Einschalten

```
# svcadm enable …
```

- Setup eigener Service

```
# svcbundle …
```

- Ausschalten

```
# svcadm disable
```

- Konfigurationsdatei:
  - `/var/svc/manifest/*`
- Funktionstest/Status

```
# svcs …
```

Operating System: Virtualisierung mit Oracle Solaris 11

Virtualisierung auf der Betriebssystem-Ebene wird in Solaris 11 über Solaris Zones realisiert. Da es sich bei den Zonen um keine unabhängigen Solaris Instanzen handelt, werden die Treiber, der Kernel und die gesamte IO-Infrastruktur aus der globalen Zone genutzt. Zur Vereinfachung der Architektur wird dabei im Allgemeinen die IO-Redundanz durch die globale Zone realisiert. D.h. die globale Zone stellt Multipathing Geräte an die Zonen zur Verfügung, die dort benutzt werden. Lediglich dort, wo in der Zone eine Information über den Ausfall eines IO-Pfades benötigt wird (Solaris Cluster), wird die IO-Redundanz in der Zone erzeugt (Aggregate oder IP-Multipathing). Konfigurationen zum Ressourcenmanagement und zur Bandbreitensteuerung des Netzwerkes befinden sich in der Zonenkonfiguration der jeweiligen nicht-globalen Zone.

Zusammenfassung

Der Aufbau von verfügbaren Anwendungsarchitekturen mit Oracle Solaris ist kein Hexenwerk. Das Betriebssystem verfügt über eine Reihe bekannter und weniger bekannter Mechanismen zur Absicherung von I/O-Pfaden, von Diensten oder zugesagter Leistung. Die besten Mechanismen im Betriebssystem nützen jedoch nichts, wenn die verwendete Hardware-Architektur nicht bereits grundlegende Bedingungen erfüllt.

Literaturverzeichnis

Oracle Solaris 11.1 Information Library
http://docs.oracle.com/cd/E26502_01/

Kontaktadressen:

Heiko Stein
etomer GmbH
Drakestraße 60
12205 Berlin
Telefon: +49 (0) 30 33503720
Fax: +49 (0) 30 33503718
E-Mail Heiko.Stein@etomer.de
Internet: http://www.etomer.com

Detlef Drewanz
Oracle Deutschland B.V. & Co. KG
Schiffbauergasse 14
D-14467 Potsdam
Telefon: +49 (0) 331 200 7341
E-Mail Detlef.Drewanz@oracle.com
Internet: http://www.oracle.com