Failover Zonen unter Oracle Solaris Cluster 4.0 – Was ist neu?

Hartmut Streppel, Detlef Drewanz
Oracle Deutschland B.V. & Co. KG
München, Berlin

Schlüsselworte
Oracle Solaris Zones, Oracle Solaris Container, Failover Zones, Flying Container, Oracle Solaris Cluster

Einleitung
Oracle Solaris Container sind die Standardmethode, um Anwendungen in einer Solaris Umgebung zu virtualisieren und zu konsolidieren. Die Möglichkeit, auf einfache Weise Ressourcen zu teilen, zu kontrollieren und administrative Rechte zu delegieren und damit sichere und gekapselte Ablaufumgebungen zu implementieren, haben zu einer breiten Akzeptanz bei Solaris Nutzern geführt.

Einer der wesentlichen Vorteile von Containern ist die Möglichkeit, sie zwischen Solaris-Instanzen hin- und her zu schieben.

Werden Container unter der Kontrolle von Oracle Solaris Cluster genutzt und damit hochverfügbar gemacht, sind keinerlei Änderungen im Container notwendig. Der Anwender und der Administrator eines Containers merkt keinen Unterschied. Ein Cluster-Agent (HA Container) überwacht die Funktionsfähigkeit des Containers und ist im Fehlerfall in der Lage, ihn auf einen anderen Clusterknoten zu schwenken.
Der folgende Text ist ein Update der beiden Aufsätze „Minimale Downtime beim Patchen von Failover (Flying) Containern“ und „Flying Container mit Oracle Solaris Container und Oracle Solaris Cluster“, die auf der DOAG Konferenz 2011 in Nürnberg präsentiert wurden. Er beschreibt die Veränderungen beim Arbeiten mit Zonen unter Oracle Solaris 11 und Oracle Solaris Cluster 4.0.

Ein grundlegendes Verständnis der Konzepte von Zonen allgemein und Failover Zonen unter Clusterkontrolle im Speziellen ist für das Verständnis des Texts notwendig.

Zonen in Oracle Solaris 11
Die komplette Integration von Zonen mit anderen Solaris Technologien macht sie noch einfacher und sicherer benutzbar.

Zonen Installation
Die zur Installation der Zone benutzten Softwarepakete werden nicht aus der globalen Zone bezogen, sondern aus einem extern bereitgestellten Repository entnommen. Dabei wird ein Gruppenpaket `solaris-small-server` aus dem konfigurierten Repository Server installiert. Die Liste der installierten Pakete ist nur minimal von den installierten Paketen der globalen Zone abhängig.

Ein neues ZFS “dataset” wird bei der Installation einer Zone automatisch angelegt. Dies ist notwendig, um Updates mit Hilfe von Boot Environments zu ermöglichen.

Ein sog. “anet” Netzwerk Device wird automatisch angelegt, um die Verwaltung der Netzschnittstellen in einer Zone zu vereinfachen. So werden die Interfaces in der Zone mit net0, net1,

etc. bezeichnet, unabhängig davon, um welches benutzte physikalische Interface es sich handelt oder ob ein virtuelles Interface (VNIC) benutzt wird.

Die Verwendung der Publisher Informationen aus der globalen Zone vereinfacht die Arbeit mit Repositories.

Wie unter Solaris 10 müssen auch Solaris 11 Zonen in ihren wesentlichen Bestandteilen konsistent mit der globalen Zone sein. Damit wird gewährleistet, dass die zentralen Schnittstellen immer 100% funktionieren. Ein Update der Zonen erfolgt zusammen mit der globalen Zone. Dazu wird das Incorporation-Paket *entire* verwendet. Dieses Paket enthält keine Pakete, sondern optionale Abhängigkeiten zu allen Paketen, die zu einem Release oder Service Repository Update (SRU) gehören. Optionale Abhängigkeit bedeutet in diesem Falle: wenn ein Paket installiert ist, muss es eine bestimmte Version haben. Ist es nicht installiert, wird es nicht angefordert. Welche Pakete installiert werden, ist durch das Gruppenpaket *solaris-small-server* abgebildet. Mit Hilfe der Metainformationen von *entire* wird über die installierten oder zu installierenden Versionen von Paketen die Konsistenz der Pakete und Libraries zwischen den Zonen sichergestellt.

```
[laptop:hs] pkg list entire
NAME (PUBLISHER)                          VERSION                    IFO
entire                                0.5.11-0.175.0.10.0.5.0     i--
```

```
[laptop:hs] pkg info entire
Name: entire
Summary: entire incorporation including Support Repository Update
(Oracle Solaris 11 11/11 SRU 10.5).
Description: This package constrains system package versions to the
same build.  WARNING: Proper system update and correct package
selection depend on the presence of this incorporation. Removing
this package will result in an unsupported system.  For more
information see https://support.oracle.com/CSP/main/article?
cmd=show&type=NOT&doctype=REFERENCE&id=1372094.1.
Category: Meta Packages/Incorporations
State: Installed
Publisher: solaris
Version: 0.5.11 (Oracle Solaris 11 SRU 10.5)
Build Release: 5.11
Branch: 0.175.0.10.0.5.0
Packaging Date: August  3, 2012 06:26:27 PM
Size: 5.45 kB
FMRI: pkg://solaris/entire@0.5.11,5.11-
0.175.0.10.0.5.0:20120803T182627Z
```

Auch wenn wir hier nicht auf alle Ausgaben eingehen, werden einige Dinge sofort offensichtlich:

- dieses System ist mit Oracle Solaris 11 11/11 SRU 10.5 installiert, was sich sowohl aus dem

Kommentar als auch dem Versionsstring, genauer: dem Fault Management Resource Identifier FMRI des Pakets ergibt;

- das Paket entire ist extrem wichtig!

Weitere Informationen über den Inhalt des Paketes sind mit `pkg contents -m entire` ersichtlich.

Boot Environments der globalen Zone

Eine der wesentlichen Eigenschaften von Oracle Solaris 11 sind Boot Environments (BE). Diese ermöglichen den Update eines Systems im laufenden Betrieb, aber auch das Anlegen von Fallback Umgebungen für den Fall, dass ein Update ein Problem haben. Bei kritischen Updates, vor allem, wenn Kernel-Komponenten betroffen sind, werden solche Fallback BEs automatisch angelegt. BEs haben einen „Universal Unique Identifier" (UUID), den man mit „`beadm list -H`" anschauen kann. Den gleichen UUID findet man auch als Property des ZFS Root Dateisystems.

```
[laptop:hs] beadm list -H solaris-95
solaris-95;b1800ceb-7669-6d1c-8096-be53d10589e9;NR;/;18095406080;...
[laptop:hs] zfs get all rpool/ROOT/solaris-95 | fgrep uuid
rpool/ROOT/solaris-95  org.opensolaris.libbe:uuid  b1800ceb-7669-
6d1c-8096-be53d10589e9  local
```

Die unterstrichenen Zeichenketten sind die UUIDs. Sie sind, solange keine Failover Zonen verwendet werden, zwar für die Konsistenz des Systems wichtig, aber ein reines Implementierungsdetail.

Boot Environments in lokalen Zonen

Auch Zonen besitzen Boot Environments. Diese sind über die UUIDs ihrer Root-Dateisysteme mit der globalen Zone verlinkt. Damit wird verhindert, dass eine Zone mit einem Boot Environment gebootet wird, das nicht zum BE der globalen Zone passt..

Die Informationen über die UUID sind ebenfalls mit Hilfe von „`beadm list -H`" und „`zfs get all <zonerootFS>`" auslesbar. Allerdings heißt die ZFS Property in der Zone nicht uuid, sondern parentbe.

```
[zone-test:root] …..
[laptop-zone:hs] zfs get all rpool/ROOT/solaris | fgrep parentbe
rpool/ROOT/solaris  org.opensolaris.libbe:parentbe  b1800ceb-7669-
6d1c-8096-be53d10589e9  local
```

Natürlich können BEs in nicht-globalen Zonen vom Zonenadministrator für eigene Verwaltungszwecke genutzt werden. So kann z.B. ein gesamtes Bootenvironment inklusive aller Zonen-Bootenvironments (ZBE) an einen Pfad für Kontroll- oder Backup-Zwecke gemountet werden. Das System stellt aber sicher, dass solche unabhängig erstellten BEs immer mit der globalen Zone synchronisiert sind. Es ist nicht möglich, in einer nicht-globalen Zone andere Stände von Systemkomponenten zu betreiben als in der globalen Zone.

Dies deutet schon auf eine noch zu beschreibende Thematik beim Betrieb von Failover Zonen hin: Das aktuelle BE einer Failover Zone muss zu dem aktiven BE der globalen Zone auf allen Clusterknoten passen.

Updaten von Zonen

Ein genereller Update eines Systems wird mit dem Befehl `pkg update` initiiert. Während des Updates eines Systems, werden die Versionsnummern der installierten Pakete mit den

Versionsnummern in dem im Repository neu vorliegenden `entire` Paket abgeglichen, das mit einem SRU oder Release zusammen ausgeliefert wird. Entsprechende neue Versionen von Paketen werden in das neue Bootenvironment nachinstalliert. Der eigentliche Download wird auf nur die tatsächlich veränderten Dateien minimiert, so dass nicht komplette Pakete, sondern in der Regel nur Teile davon heruntergeladen und installiert werden.

Bei der Update-Prozedur werden alle Zonen, die im Zustand „`installed`“ sind, ebenfalls auf den neuesten Stand gebracht. Zu diesem Zweck werden sowohl in der globalen Zone als auch in allen lokalen Zonen in der Regel neue BEs mit ZFS-Mitteln angelegt und die Änderungen in diese neuen BEs eingepflegt. All diese neuen BEs besitzen wieder einheitliche UUIDs. Als letzte Aktion werden die neuen BEs aktiviert, so dass sie beim nächsten Reboot genutzt werden.

Failover Zonen unter Kontrolle von Oracle Solaris 4.0

Oracle Solaris Cluster 4.0 (OSC) ist die Clusterversion für Oracle Solaris 11. Failover Zonen sind weiterhin unterstützt und werden ähnlich aufgesetzt wie unter OSC3.x. Es gibt allerdings zwei neue, zusätzliche Schritte.

Zunächst müssen die UUIDs der aktiven BEs, d.h. der aktiven Root-Dateisysteme die identische UUID haben. Dies erfolgt mit Hilfe von

```
zfs set org.opensolaris.libbe:uuid=<uuid> rpool/ROOT/<rootFS>
```

Dann muss in jeder Failover Zone, die unter Clusterkontrolle betrieben wird, eine zusätzliche Eigenschaft in ihrer Zonenkonfiguration gesetzt werden. Diese erlaubt es anderen Komponenten, eindeutig festzustellen, ob eine Zone unter Clusterkontrolle, also eine Failover Zone ist, oder nicht. Diese zusätzliche Eigenschaft wird beim Anlegen der Zone mit `zonecfg` gesetzt:

```
add attr; set name=osc-ha-zone; set type=boolean; set value=true;
end;
```

Exkurs: Zonencluster und Zone Nodes

Zonencluster sind weiterhin Bestandteil von Oracle Solaris Cluster 4.0. Die Installation von Zonenclustern und deren Konfiguration geschieht wie bisher mit den bekannten Kommandos `clzonecluster`. Bei der Konfiguration eines Zonenclusters wird in allen Zonen ebenfalls eine eigene Property namens `cluster` angelegt und auf `true` gesetzt. Der Brand dieser Zonen ist „solaris“.

Das Feature, hochverfügbare Dienste unter Clusterkontrolle einfach durch Eintragen von Zonennamen in die Nodelist Property einer Ressourcegruppe in Zonen laufen zu lassen (Zone Nodes), ist nicht mehr Bestandteil von Oracle Solaris Cluster 4.0. Stattdessen sollten Zonencluster verwendet werden, die einen größeren und einfacher zu bedienenden Funktionsumfang haben. Außerdem kann bei der Verwendung von Zonenclustern die Administration von Cluster-Ressourcen, die innerhalb des Zonenclusters bekannt sind, an den Zonenadministrator übergeben werden (management delegation). Dies war mit den sog. „zone nodes“ nicht möglich.

Updaten von Failover Zonen

Unter Solaris 10 mussten relative komplexe Prozeduren befolgt werden, wenn Systeme mit Failover Zonen gepatcht wurden. Besonders aufwändig wurde es, wenn dies im laufenden Betrieb mit Hilfe von Live Upgrade geschah. Die Prozedur ist unter Oracle Solaris 11 sehr leicht geworden. Kurz und

nicht ganz vollständig beschrieben ergeben sich die folgenden Schritte:

- Update des Knoten A, auf dem die Failover Zone nicht läuft; ein neues BE wird erstellt mit einer neuen UUID; reboot des Systems in das neue BE
- Update des anderen Knoten B, auf dem die Zone aktiv ist; auch hier wird ein neues BE erstellt mit einer neuen UUID; dieses BE wird noch nicht gebootet.
- Übertragen der neuen UUID des RootFS von Knoten B auf das RootFS des Knoten A
- Verschieben der Failover Zone auf den Knoten A, der schon mit neuer Software läuft;
- reboot von Knoten B ins neue BE

Sind auf den Systemen noch weitere lokale Zonen aktiv, müssen auch die UUIDs deren BEs angepasst werden.

Durch die Integration des Updates von Failover Zonen in die Updateprozesse der globalen Zone werden alle notwendigen Schritte, vor allem die Umwandlungen der Boot Environments automatisch und zuverlässig durchgeführt.

Der Update eines mit Oracle Solaris Cluster 4.0 betriebenen Solaris Knotens geschieht mit Hilfe des „`scinstall -u update <-b bename>`“ Komandos. Dieses sorgt nicht nur für die Installation der neuen Cluster-, sondern auch für die der neuen Solaris 11 Pakete.

Zusammenfassung

Oracle Solaris Container, die unter Solaris 11 nur noch Zonen genannt werden, sind die wichtigste Virtualisierungsmethode in Oracle Solaris. Sehr viele Oracle Solaris Anwender nutzen diese Technologie, um große, bisher getrennt laufende Anwendungsumgebungen zu konsolidieren. Die Integration von Zonen mit Oracle Solaris Cluster und dem HA Zonen Agenten bringt zusätzlich Hochverfügbarkeit, ohne dass in den Zonen oder an den in den Zonen laufenden Anwendungen irgendetwas modifiziert werden muss. Oracle Solaris 11 und Oracle Solaris Cluster 4.0 liefern eine wesentlich verbesserte Integration von Zonen mit anderen Teilen des Betriebssystems. Dies bewirkt sowohl eine erhöhte Sicherheit beim Betrieb als auch eine extreme Vereinfachung beim Updaten von Failover Zonen und macht damit den Einsatz von Solaris Zonen zu einem sog. „Nobrainer“.

Referenzen

- Oracle Solaris Cluster Data Service for Oracle Solaris Zones Guide - http://docs.oracle.com/cd/E23623_01/html/E26828/
- Oracle Solaris Cluster Upgrade Guide - http://docs.oracle.com/cd/E23623_01/html/E24617
- Oracle Solaris Administration: Oracle Solaris Zones, Oracle Solaris 10 Zones, and Resource Management - http://docs.oracle.com/cd/E23824_01/html/821-1460
- Dokumentation zum Thema Update:
  - Oracle Solaris Cluster Upgrade Guide, Kapitel: Upgrading Zones Managed by HA for Oracle Solaris Zones - http://docs.oracle.com/cd/E23623_01/html/E24617/glpnn.html
  - Adding and Updating Oracle Solaris 11 Software Packages - http://docs.oracle.com/cd/E23824_01/html/E21802

- How to Get Started Creating Oracle Solaris Zones in Oracle Solaris 11 - http://www.oracle.com/technetwork/articles/servers-storage-admin/o11-092-s11-zones-intro-524494.html

Kontaktadresse:
Hartmut Streppel
ORACLE Deutschland B.V. & Co. KG
Riesstrasse 25
D-80992 München

Telefon: +49 (0) 89-1430 2588
Fax: +49 (0) 89-1430 1150
E-Mail Hartmut.Streppel@oracle.com
Internet: www.oracle.com

Detlef Drewanz
ORACLE Deutschland B.V. & Co. KG
Komturstrasse 18a
D-12099 Berlin

Telefon: +49 (0) 331-200 7341
E-Mail Detlef.Drewanz@oracle.com
Internet: www.oracle.com