

Datenbank-Cloning mit Oracle ZFS Storage Appliances

Franz Haberhauer, ORACLE Deutschland B.V. & Co. KG

Mit ZFS verfügt Oracle über eine erprobte Speichertechnologie mit einem höchst effizienten „Copy on Write“-Mechanismus, der ein schnelles und platzeffizientes Cloning von Datenbanken erlaubt.

ZFS kann nicht nur unter Oracle Solaris direkt für Datenbanken auf ZFS-Dateisystemen genutzt werden, es dient in den ZFS Storage Appliances auch als internes Dateisystem. Die neuen Cloning-Features in der Oracle-Software – „DBaaS Snap Clone“ in Enterprise Manager 12c und das Cloning von Pluggable Databases in der Oracle Datenbank 12c – unterstützen direkt ein „Thin Cloning“ für Datenbanken, die auf ZFS Storage Appliances liegen. Darüber hinaus bietet Oracle mit dem Snap Management Utility eine Lösung für das Sichern, Wiederherstellen, Provisionieren und insbesondere das Cloning von Oracle-10g- und -11g-Datenbanken auf Basis der Snapshot-/Clone-Funktionalität der ZFS Storage Appliances. Diese können im Storage Area Network (SAN) genutzt werden, im Vordergrund steht aber die Nutzung als Network Attached Storage (NAS) über (d)NFS (Files) oder iSCSI (Block-Storage unter ASM).

Copy on Write mit Solaris ZFS

ZFS ist ein modernes lokales Dateisystem, das sich seit fast zehn Jahren in Solaris bewährt hat und in Solaris 11 als Root-Filesystem vorgegeben ist – nicht zuletzt wegen seiner Snapshot- und Cloning-Funktionalitäten. Diese motivieren, darauf auch Datenbanken zu betreiben. Best Practices für Datenbanken auf ZFS hat der Autor in der DOAG News vom Juli 2013 vorgestellt [1].

ZFS überschreibt Daten grundsätzlich nicht unmittelbar, sondern folgt dem Prinzip „Copy on Write“ (CoW). Dabei werden geänderte Blöcke wie neue Blöcke an eine neue Stelle auf der Platte geschrieben, was wahlfreie Schreibzugriffe in effiziente, sequenzielle Schreiboperationen umwandelt und damit den Zugriffs-Charakteristika moderner Platten entgegenkommt. Dateien werden in Baumstrukturen verwaltet, wobei sich Änderungen erst mit der Änderung des Wurzelknotens („Überblock“) materialisieren. Wird dabei der vorige Wurzelknoten nicht



freigegeben, kann er als Wurzel eines Snapshot dienen, der eine Sicht auf die bisherige Baumstruktur – und damit den bisherigen Datei-Inhalt – liefert.

In ZFS sind Snapshots „read only“. Sobald sie schreibbar sind, spricht man von Clones. Der große Vorteil dieser Technologie besteht darin, dass für Clones gegenüber dem Ausgangszustand lediglich das tatsächlich geänderte Volumen als zusätzliche Speicherkapazität benötigt wird („Thin Cloning“). Zudem sind beim Anlegen nur minimal Zeit und Ressourcen erforderlich und die Zahl der Snapshots und Clones ist nicht begrenzt.

Snapshots können an jeder Stelle in der ZFS-Datei-Hierarchie angelegt werden, beispielsweise für das Verzeichnis „/mypool/mydataset/mydir“ im Pool „mypool“ mit dem Kommando „# zfs snapshot mypool/mydataset/mydir@201404081400“. Dabei identifiziert der String nach dem „@“ den Snapshot. Über „# zfs clone /mypool/mydataset/mydir@201404081400 mypool/mydataset/clonedir“ wird daraus ein Clone unter dem Pfad „/mypool/mydataset/clonedir“ sind weitere ZFS-Datasets im Pfad eingebunden, lassen sich über die Option „-r“ synchronisiert rekursive Snapshots darüber hinweg erstellen.

ZFS Storage Appliance

Mit Solaris verfügt Oracle über ein Betriebssystem, das alle Anforderungen an ein Speichersystem abdeckt: ZFS als leistungsstarkes lokales Dateisystem; COMSTAR, eine SCSI-Target-Funktionalität für die Bereitstellung von Block-Storage über Fibre Channel (FC) im SAN oder iSCSI im IP-Netzwerk und über iSER und SRP insbesondere auch via InfiniBand, sowie NFS als Netzwerk-Dateisystem, das ja einst bei Sun entwickelt worden war. Auch CIFS ist performant direkt im Solaris-Kern implementiert.

Während früher für NAS-Filer im Hinblick auf die verfügbare Hardware dedizierte Betriebssysteme entwickelt wurden, ist heute aufgrund der Leistungsexplosion bei der Hardware der Einsatz von General-Purpose-Betriebssystemen naheliegend, um von den Entwicklungen für andere Einsatzbereiche zu profitieren – insbesondere was Hardware-Unterstützung, Treiber, Skalierbarkeit, Sicherheit etc. angeht – und Entwicklungsbudgets besser in höherwertige Funktionen investieren zu können. Die ZFS Storage Appliances [2] bringen eine Vielzahl an Funktionalitäten mit wie Kompression oder Deduplizierung, die nicht separat lizenziert werden müssen. Außerdem ist auf diesen Speichersystemen für Oracle-Datenbanken „Hybrid Columnar Compression“ einsetzbar.

Für die Administration gibt es zum einen eine webbasierte, intuitive grafische Oberfläche. Eine Besonderheit ist dabei Analytics, eine Komponente, die eine sehr fein granulierte Analyse von IO-Mustern ermöglicht, bis hin zur Visualisierung des Spektrums der Latenz einzelner Typen von E/A-Operationen oder der Offsets in Dateien [3]. Hierdurch wird die Diagnose von I/O-Problemen signifikant vereinfacht – und in einigen Fällen überhaupt erst möglich. Zum anderen ist eine Integration sowohl mit dem Enterprise Manager Ops Center [4] wie auch – zumindest was das Monitoring angeht – über ein Storage-Plug-in direkt mit dem Enterprise Mana-

ger möglich [5]. Speziell für das Cloning von Datenbanken gibt es eine weitergehende Werkzeugunterstützung, die nachfolgend vorgestellt wird.

Zunächst noch kurz zur Hardware der ZFS Storage Appliances [6]: Die aktuellen Systeme der ZS3-Familie (ZS3-2, siehe *Abbildung 1*, beziehungsweise ZS3-4) verfügen jeweils über einen einzelnen Controller oder über einen HA-Cluster aus zweien (aktiv-aktiv). Auf diesen Controllern oder Köpfen läuft die Betriebssoftware. Zur Speicherung kommen Disk Shelves mit bis zu 24 Platten zum Einsatz. In Engineered Systems werden ZFS Storage Appliances intern in der Exalogic, Exalytics oder dem SPARC SuperCluster verbaut. Sie sind aber auch als Backup-Plattform insbesondere für die Exadata populär – nicht zuletzt, da sie eine direkte InfiniBand-Anbindung erlauben [7]. Hierbei wurden bis zu 26 TB/Stunde beim Backup und 17 TB/Stunde beim Restore gemessen. Seit Anfang März 2014 ist für diesen Einsatzbereich vorkonfiguriert die ZS3 Backup Appliance ZS3-BA verfügbar [8] und seit April 2014 die Version 2.0 des Oracle Engineered Systems Backup Utility for Oracle ZFS Storage Appliance, über das in wenigen Schritten (etwa generierte RMAN-Skripte) das Backup für Exadata, ODA und SuperCluster auf ZFS Storage Appliances aufgesetzt werden kann [9].

Sind Oracle-12c-Datenbanken über dNFS auf den ZFS Storage Appliances abgelegt, kommt das Oracle Intelligent Storage



Abbildung 1: ZFS Storage Appliance ZS3-2 mit zwei Clustered Controller Heads, einem DE-24P Disk Shelf mit 24 2.5"-10.000rpm-900GB- und einem DE-24C Disk Shelf mit 24 3.5"-7200rpm-4TB-Laufwerken

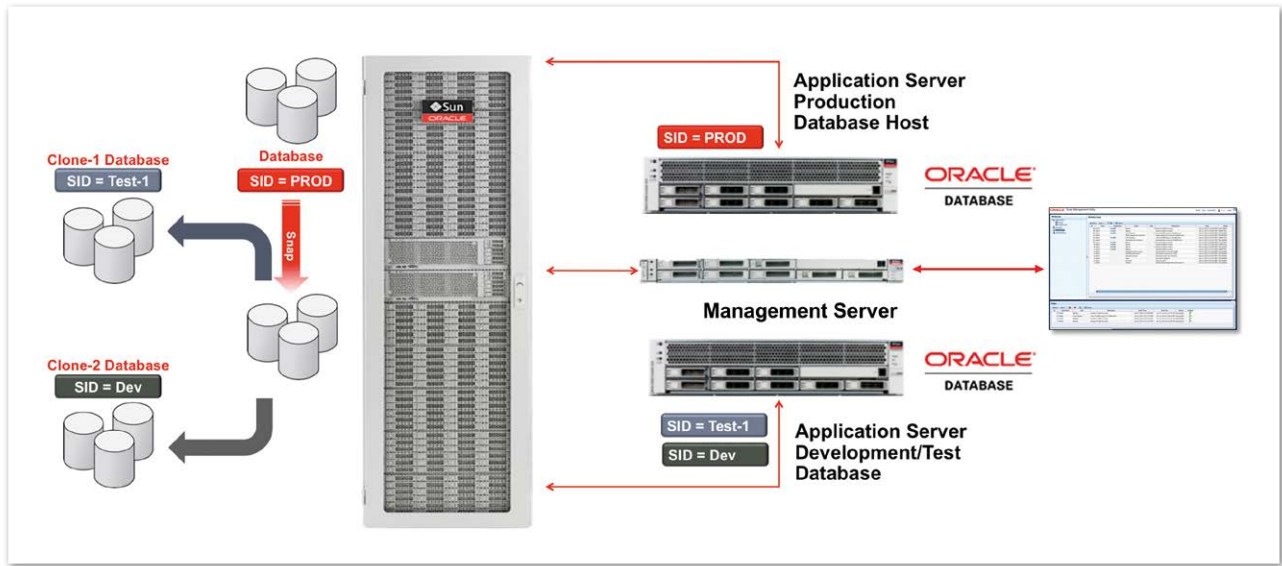


Abbildung 2: Snap Management Utility for Oracle Database: Datenbank-Clones und Management-Server können auf verschiedene Server verteilt sein

Protocol (OISP) zum Tragen, über das die Datenbank exklusiv den ZFS-SA Meta-Informationen über die Last bereitstellt, die diese für automatisches Tuning nutzen [10].

Datenbank-Cloning mit der ZFS Storage Appliance

Wie gesagt, unterstützen die neuen Cloning-Features in der Oracle-Software – „DBaaS Snap Clone“ im Enterprise Manager 12c oder das Cloning von Pluggable Databases in der Oracle Datenbank 12c – direkt ein „Thin Cloning“ für Datenbanken, die auf ZFS Storage Appliances liegen. Um eine ZFS Storage Appliance im Enterprise Manager als Storage-Server zu nutzen, werden administrative Nutzer beziehungsweise Rollen mit den entsprechenden Privilegien angelegt und die Appliance im EM registriert [11]. Danach erfolgt das Management aus dem Enterprise Manager, der mit der Appliance über „ssh“-Verbindungen kommuniziert. Auch für das Cloning von Pluggable Databases werden Username und Passwort für einen entsprechend autorisierten Nutzer (hier in einem Oracle Database TDE Keystore) hinterlegt [12] und das Cloning auf der Appliance über eine SQL-Klausel getriggert.

Während die meisten Software-Features in den ZFS Storage Appliances einschließlich Snapshots frei nutzbar sind, erfordern das Cloning von Daten auf einem System sowie die Replikation zwischen Systemen zusätzliche Lizenzen.

Beim Enterprise Manager Snap Clone, ebenfalls ein Feature des Oracle Cloud Management Pack for Oracle Database, ist eine „Restricted Use License“ für das ZFS-Storage-Appliance-Cloning enthalten. Auch beim nachfolgend vorgestellten Snap Management Utility ist das Cloning von Oracle-Datenbanken über das Tool bereits mit lizenziert.

Snap Management Utility for Database

Während die oben skizzierten neuen Funktionalitäten Enterprise Manager 12c beziehungsweise die Datenbank 12c voraussetzen, bietet das Oracle Database Snap Management Utility (SMU, siehe Abbildung 2) eine Lösung für das Sichern, Wiederherstellen, Provisionieren und insbesondere das Cloning von 10g- und 11g-Datenbanken (auch RAC) auf der Basis der Snapshot-/Clone-Funktionalität der ZFS-Storage-Appliances-Utility (SMU) [13,14]. Gegenüber manuellen Vorgehensweisen [15] bietet das SMU den Vorteil einer unterstützten Software, die gepflegt und weiterentwickelt wird und nicht zuletzt aufgrund der enthaltenen Cloning-Lizenz auch preislich attraktiv ist. Zudem verfügt sie nicht nur über ein Command-Line-Interface, sondern auch über eine Browserbasierte Nutzerschnittstelle. SMU kann für Datenbanken unter Oracle Solaris, Linux oder Windows genutzt werden, wobei Direct NFS und ASM auf iSCSI 11g erfordern.

Das Snap Management Utility ermöglicht Cloning einerseits auf Basis eines Snap-Backups einer Datenbank und andererseits auf Basis eines RMAN-Backups (siehe Abbildung 3). Beim ersterem wird auf der ZFS StorageAppliance ein Snapshot einer Datenbank angelegt – als Offline- (cold) oder als Online-Backup (hot, nur bei Datenbanken auf (d)NFS). Daraus wird ein Clone erstellt als Datenbank auf demselben Host wie die Ausgangsdatenbank oder auf einem anderen Host, der dieselbe Architektur und Datenbank-Software wie der Host der Ausgangsdatenbank haben muss. Eine Konversion zwischen einer Single-Instance- und einer RAC-Umgebung ist möglich.

Für das Cloning auf Basis eines RMAN-Backups muss dieses im „Image Copy“-Format vorliegen. Quell- und Zielhost des Datenbank-Cloning können bei dieser Variante nicht identisch sein, müssen aber denselben Softwarestand haben. Dabei wird vom RMAN-Backup ein ZFS-Clone erzeugt, der dann auf dem Zielhost eingebunden wird. SMU startet auf dem Zielhost eine temporäre Datenbank-Instanz, über die verschiedene Informationen aus dem Backup-Controllfile ermittelt werden – insbesondere der Recovery-Point (maximale SCN) und die Größe der Flash Recovery Area. Die temporäre Instanz wird dann wieder heruntergefahren, ein Parameterfile für die neue Clone-Datenbank erzeugt, diese damit gestartet und ein Recovery durchgeführt.

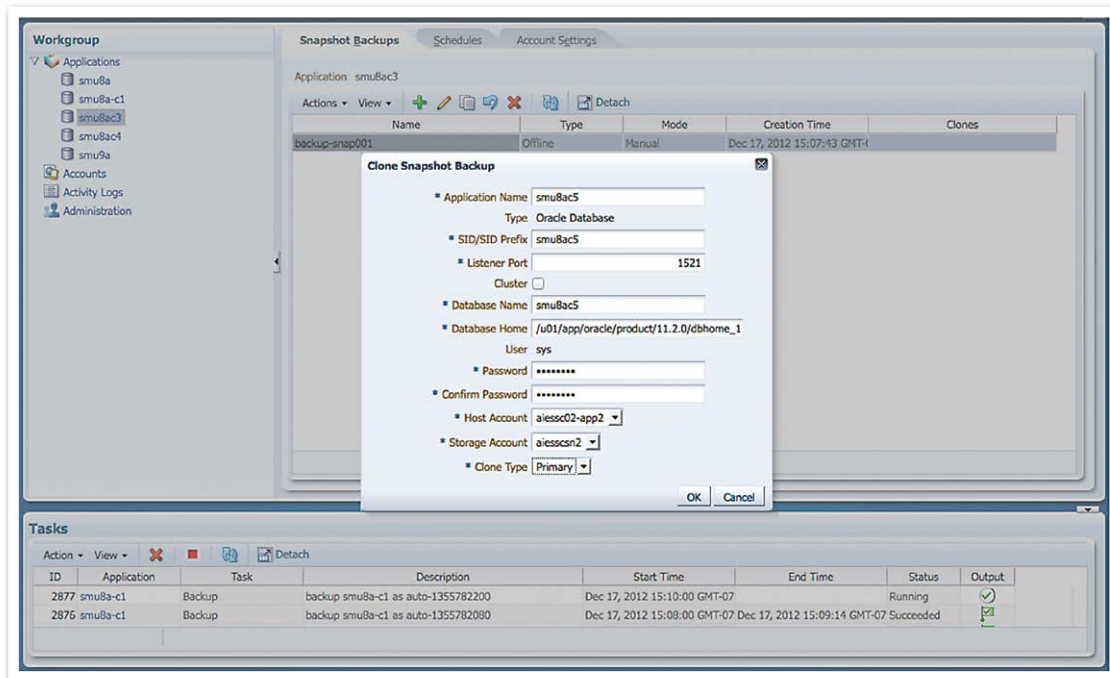


Abbildung 3: Snap Management Utility: Clone-Snapshot-Backup-Dialog

Fazit

Mit den ZFS Storage Appliances hat Oracle Speichersysteme im Produktportfolio, die aufgrund der zugrunde liegenden „Copy on Write“-Technologie hervorragend für Thin Cloning geeignet sind. Das Cloning von Datenbanken erfordert ein Zusammenspiel von Datenbank- und Storage-Administration. Mit den neuen Cloning-Funktionalitäten in der Datenbank 12c sowie im Enterprise Manager 12c lassen sich die Cloning-Features nun transparent direkt aus der Datenbank beziehungsweise dem Enterprise Manager heraus nutzen.

Mit dem Snap Management Utility steht zudem ein Werkzeug zur Verfügung, das Thin Cloning auch für ältere Datenbank-Versionen einfach nutzbar macht –demonstriert in einem YouTube, wobei zunächst Analytics für Datenbanken gezeigt wird [16]. ZFS Storage Appliances haben sich übrigens auch in der Oracle-IT und den Cloud-Datacentern durchgesetzt. Hier ist inzwischen eine Gesamtkapazität von weit über 200 Petabyte installiert [17].

Literaturhinweise

- [1] Franz Haberhauer, Best Practices für Datenbanken auf ZFS: DOAG News, Ausgabe 03/2013, S. 18-21
 [2] Architectural Overview of the Oracle ZFS Storage Appliance, An Oracle White Paper, Januar 2014: <http://www.oracle.com/technetwork/server-storage/sun-unified-storage/documentation/o14-001-architecture-overview-zfsa-2099942.pdf>

- [3] Oracle ZFS Storage Appliance Analytics Guide: http://docs.oracle.com/cd/E27998_01/html/E48490/index.html
 [4] How to Prepare a Sun ZFS Storage Appliance to Serve as a Storage Device with Oracle Enterprise Manager Ops Center 12c: <http://www.oracle.com/technetwork/articles/servers-storage-admin/howto-zfs-appl-ops12c-1840019.html>
 [5] Oracle Enterprise Manager System Monitoring Plug-in for Oracle ZFS Storage Appliance: <http://www.oracle.com/technetwork/oem/grid-control/downloads/zfs-storage-plugin-487867.html>
 [6] ZFS Storage Appliance Site u.a. mit 3D-Modellen der Hardware: <http://www.oracle.com/technetwork/server-storage/sun-unified-storage/overview>
 [7] Oracle ZFS Storage Appliance: <http://www.oracle.com/us/products/servers-storage/storage/nas/overview> und <http://www.oracle.com/us/products/servers-storage/storage/nas/zs3-ds-2008069.pdf>
 [8] Backup and Recovery for Oracle Engineered Systems with Oracle ZFS Storage ZS3-BA, Oracle Data Sheet, March 2014: <http://www.oracle.com/us/products/storage/engineered-systems-backup/ds-zs3-ba-140228-2162876.pdf>; Oracle ZFS Storage ZS3-BA: <http://www.oracle.com/us/products/servers-storage/storage/nas/zs3-ba>
 [9] Oracle Engineered Systems Backup Utility for Oracle ZFS Storage Appliance: <http://www.oracle.com/technetwork/server-storage/sun-unified-storage/downloads/zfsa-plugins-1489830.html>
 [10] Oracle Intelligent Storage Protocol, Data Sheet: <http://www.oracle.com/us/products/servers-storage/storage/nas/oracle-isp-ds-final-2139042.pdf>
 [11] Registering and Managing Storage Servers: http://docs.oracle.com/cd/E24628_01/doc.121/e28814/cloud_db_setup.htm#autoid7
 [12] Ritu Kamboj, Oracle Multitenant on SPARC Servers and Oracle Solaris, September 2013: <http://www.oracle.com/technetwork/articles/>

[servers-storage-admin/multitenant-on-sparc-solaris-2016889.html](http://www.oracle.com/technetwork/articles/servers-storage-admin/multitenant-on-sparc-solaris-2016889.html)

- [13] Snap Management Utility for Oracle Database: <http://www.oracle.com/us/products/servers-storage/storage/nas/snap>
 [14] Oracle Snap Management Utility for Oracle Database Documentation: http://docs.oracle.com/cd/E39520_01
 [15] Database Cloning using Oracle Sun ZFS Storage Appliance and Oracle Data Guard, Oracle White Paper, December 2011: <http://www.oracle.com/technetwork/database/features/availability/maa-db-clone-szfsa-172997.pdf>
 [16] Art Licht, Simplifying Database-to-Storage Management, YouTube, March 2013: <https://www.youtube.com/watch?v=w4YhPiHP88>
 [17] Sun ZFS Storage Appliance and Oracle IT, Use Cases and Benefits, Oracle White Paper, September 2012: <http://www.oracle.com/us/products/servers-storage/storage/nas/resources/zfs-saoracle-it-whitepaper-100812gc-1875031.pdf>



Franz Haberhauer
franz.haberhauer@oracle.com