

Backup und Recovery PoC auf der Recovery Appliance

Frank Schneede, Attila Mester
Oracle Deutschland B. V. & Co. KG

Schlüsselworte

ZDLRA, Recovery Appliance, Backup, Recovery, PoC

Einleitung

Die Recovery Appliance ist schon seit Oktober 2014 auf dem Markt und erweitert die Oracle Familie der Engineered Systems um eine Lösung, die die Datensicherung von Oracle Datenbanken revolutioniert, und das nicht nur in grösseren Rechenzentren. Das neue Sicherungskonzept stieß auf großes Interesse bei den Kunden und mittlerweile gibt es nicht nur Erfahrungen aus der Durchführung von PoCs, sondern auch aus produktiven Einsätzen der Recovery Appliance. In diesem Vortrag wird die Funktionsweise und die wichtigsten Eigenschaften dieser Appliance kurz vorgestellt. Weiterhin wird ein Überblick über die mit der Recovery Appliance erreichten Testergebnisse gegeben.

Recovery Appliance – Ein neuer Ansatz für Backup & Recovery

Eines der wichtigsten Ziele bei der Entwicklung der Maschine war, dass die Wiederherstellung der gesicherten Datenbanken auch beim Ausfall des kompletten Produktivsystems ohne Datenverlust möglich sein sollte. Das kann nur durch ein Continuous Data Protection (CDP) System erreicht werden, in dem alle Änderungen der Datenbank in Real-Time gesichert werden. Ausserdem haben heutige Backuplösungen das Problem, dass diese die Datenbank während des Backups sehr stark belasteten. Auch dies wird mit Hilfe der neuen Backupstrategie zur Recovery Appliance erheblich verbessert und durch die neue Architektur kann die Recovery Appliance mit einem System hunderte von Oracle Datenbanken im Rechenzentrum effizient sichern.

Einstieg mit Base Rack Konfiguration und wachsen

- **Base Rack** (entspricht Exadata Quarter Rack High Capacity)

– **2x Compute Server** mit High Speed Connectivity *

- 8 x 10 Gb Ethernet Ports pro Rack
- 12 x 40 Gb InfiniBand Ports pro Rack
- Optional: 4 x 16Gb Fibre Channel Ports für Tape-Anschluss

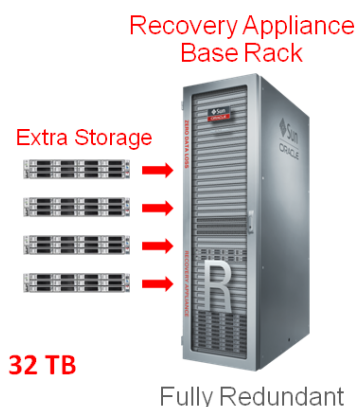
– **3x Storage Server** *

- 12 x **8TB** High Capacity Disks (je Storage Server)
- **94 TB** nutzbare Backup Kapazität



- **Jeder Storage Server** erweitert nutzbare Backup Kapazität um **32 TB**

- **Full Rack: 2x Compute Server, 18x Storage Server**
– **580 TB** nutzbare Backup Kapazität



* Siehe: [Data Sheet mit aktueller Konfiguration](#)

Abb. 1: Recovery Appliance Hardware

Die hier vorgestellte Recovery Appliance räumt mit den Nachteilen herkömmlicher Sicherungsverfahren auf, da sie gerade deren Defizite adressiert:

- Belastung der zu sichernden Systeme durch Backup-Agenten
- Lange Wiederherstellungszeiten durch Zurückspielen eines Level 0 Backups und mehrerer Level 1 Backups
- Hohe Systembelastung bei einer "incrementally updated backup strategy" durch ständiges Recovery
- Deduplizierungsalgorithmen funktionieren bei Sicherung von Datenbanksystemen in der Regel nur bei Komplettsicherungen
- Validierung der Backup Daten auf Wiederherstellbarkeit auf Storage Ebene fehlt
- Möglicher datenverlust bei Restore, da bei fehlerhaften Online Redologs eine Wiederherstellung der Datenbanken nur bis zum letzten archivierten Redolog möglich ist

Funktionsweise der Recovery Appliance im Rechenzentrum

Die Recovery Appliance nutzt RMAN, um die Daten der zu sichernden Datenbanken auf die Recovery Appliance zu übertragen. Das geschieht über ein Recovery Appliance spezifisches SBT Interface im RMAN. Mehr Änderungen auf der Client Seite sind nicht notwendig. Die entsprechenden SBT Libraries gibt es dabei schon bereits für eine 10.2.0.5 Datenbank, aber selbstverständlich auch für alle neueren Versionen.

Es sei erwähnt, dass die Recovery Appliance auch auf bereits vorhandene plattenbasierte Backups der Datenbank zurückgreifen kann. Dieser "Poll" Mechanismus transferiert die Backup Daten, wie diese z.B. schon in der Fast Recovery Area (FRA) enthalten sind, zur Recovery Appliance. Dieser "Poll" Prozess kann auch in regelmäßigen Abständen konfiguriert werden.

Die Methode, wie Backups von der Datenbank auf die Recovery Appliance erstellt werden, heißt "Delta Push". Dabei wird initial ein Level 0 Full Backup erzeugt, danach nur noch differenziell-inkrementelle Backups. Mit dieser "Incremental-Forever" Strategie wird erreicht, dass die Leistungsfähigkeit der Datenbank während des Backups nur minimal beeinträchtigt wird. Damit alle Änderungen der Datenbank zeitnah von der Recovery Appliance gespeichert werden können, kann das Delta Push Verfahren optional für den Real-Time Redo Transport konfiguriert werden. Der zur Zeit asynchrone Redo-Transport kommt direkt aus dem Redolog Buffer der Datenbank und erlaubt somit im Falle eines totalen Ausfalls des primären Systems, das System bis zur beinahe letzten Transaktion zu recovern. Diese Funktion kennt man auch von Data Guard Umgebungen. Genau wie in 12c Data Guard Umgebungen lässt sich damit auch ein echter Zero Data Loss Modus realisieren, indem man eine Far Sync Instance einsetzt. Im Gegensatz zu Data Guard unterstützt die Recovery Appliance aber auch die Redo Informationen von Betriebssystem Plattformen mit unterschiedlichen Endian Strukturen. Die Redo Streams werden in der Appliance archiviert, was das Primärsystem ebenfalls entlastet, da beim inkrementellen Backup nicht mehr unbedingt die archivierten Redo-Logs nochmals gesichert werden müssen.

Die empfangenen Backup Daten werden in der Recovery Appliance im sogenannten Delta Store gespeichert. Der Delta Store ist das Herz der Recovery Appliance, in dem er für alle Operationen zuständig ist, die für die Aufbereitung und Speicherung der Backup Daten notwendig sind. Nachdem die Appliance die Backup Daten empfangen hat, validiert und speichert sie diese in Pools in komprimierten Blöcken. Die De-Duplizierung der Backup Daten findet nicht in der Appliance,

sondern bereits auf der Client Seite durch RMAN statt, weil ein inkrementelles Backup die redundanten Blöcke beim Lesen schon eliminiert. Nach jedem inkrementellen Backup wird aus den gespeicherten Blöcken in der Recovery Appliance eine virtuelle Full Repräsentation des Backups erzeugt, was auch so in den Recovery Katalog, der sich ebenfalls auch der Appliance befindet, eingetragen wird. Aus der Perspektive des Datenbankbenutzers ist ein solches Backup nicht von einem nicht-virtuellem Full Backup unterscheidbar und auch in den Recovery Zeiten mit einem Full Backup vergleichbar. Der Restore Prozess von einem Full Backup ist wesentlich schneller als bei einer Full/Incremental Strategie. Man gewinnt also Geschwindigkeit beim Restore, obwohl nur der Plattenplatz für inkrementelle Backups beansprucht wird. Damit man sicher sein kann, dass die virtuellen Backups immer wiederherstellbar sind, werden im Hintergrund periodisch automatische Validierungsjobs ausgeführt. Alle diese Prozesse werden von der Appliance autonom, mit der Rechenkraft der Appliance ausgeführt. Die produktiven Datenbanken werden somit von Datensicherungsaktivitäten weitgehend entlastet.

Platz effizient nutzen mit "Virtual" Full Backups

Keine Full Backups mehr: Incremental-Forever Architektur

Das Diagramm zeigt den Datenfluss von 'Gesicherte Datenbanken' (links) zu einer Recovery Appliance (Mitte). Rechts davon ist der 'Delta Store' dargestellt, der die Backup-Struktur visualisiert. Er enthält:

- Tag 0 Full**: Ein vollständiges Backup (blau markiert).
- Tag 1 Incr**: Ein inkrementelles Backup (rot markiert).
- Tag N Incr**: Ein weiteres inkrementelles Backup (rot markiert).
- Tag 1 Virtual Full**: Ein virtuelles Full Backup, das die Daten von Tag 0 und Tag 1 repräsentiert (rot umrandet).
- Tag N Virtual Full**: Ein weiteres virtuelles Full Backup, das die Daten von Tag 0 und Tag N repräsentiert (grün umrandet).

- Nach initialem Full-Backup werden tägliche *Virtual Full DB Backups* aus den inkrementellen Backups erzeugt
 - Pointer-basierte Darstellung des Full Backups zum Zeitpunkt des inkrementellen Backups
 - Virtual Backups haben etwa 10x effizientere Platznutzung
 - Halten einer langen Backup Historie mit geringstem Speicherverbrauch
 - "Time Machine" für die Datenbank

ORACLE

DOAG Konferenz 2016, Nürnberg Copyright © 2015 Oracle and/or its affiliates. All rights reserved. | Frank Schneede & Attila Meister, Oracle Deutschland 19

Abb. 2: Delta Store mit virtuellen Full Backups

Der Recovery Prozess ist ebenfalls denkbar einfach. Wenn eine Datenbank oder Teile davon reconvert werden sollen, verwendet RMAN den Katalog, der auf der Recovery Appliance installiert ist. Die Zeitangabe für eine PiT Recovery ist im Rahmen der spezifizierten Datenvorhaltung (Retention) beliebig möglich. Die Recovery Appliance liefert dann das nächst gelegene Full Backup, die notwendigen Archive-Logs und die empfangenen Redo Streams, die hier den Online Redo-Logs der Datenbank entsprechen, sollten diese auf dem Primärsystem nicht mehr verfügbar sein. Somit ist eine "Zero Data Loss Recovery" der Datenbank möglich. Es sei hier erwähnt, dass nicht alle Datenbanken zwingend mit Redo Transport gesichert werden müssen. In diesem Fall sollte bei der RMAN Sicherung der Datenbank dann auch archivierte Redo-Logs mit gesichert werden, wie bei herkömmlichen Backups. Die virtuellen Full Backups stehen in diesem Fall aber auch für ein Recovery zur Verfügung.

Die Recovery Appliance kann im Rechenzentrum sinnvoll ergänzt werden. Wenn weitere Backup Kopien gewünscht werden, kann eine entfernte Appliance alle gesicherten Daten aus der ersten Appliance durch Replikation zeitnah empfangen. Die zwei Backup Kopien werden sowohl in der ersten wie auch in der zweiten Recovery Appliance katalogisiert und stehen somit transparent für RMAN Restores zur Verfügung.

Die Recovery Appliance ist standardmässig mit dem Backup Tool Oracle Secure Backup vorkonfiguriert, um Backup Kopien auf Bandlaufwerken zu erstellen. Die Erstellung der bis zu vier Bandkopien wird durch die Appliance autonom, ohne Belastung der zu sichernden Primärsysteme erledigt. Die Bandkopien können unterschiedliche Retention Zeiten haben. Damit kann eine kostengünstige Archivierung erreicht werden. Die Bandbibliotheken werden über Fiber Channel mit der Appliance direkt verbunden. Die Möglichkeit zur Verwendung einer bestehenden Tape Backup Infrastruktur wie z.B. NetBackup besteht ebenfalls, der fremde Media Server muss hier jedoch ausserhalb der Appliance installiert werden. Auf der Appliance selber wird dann lediglich der Backup Agent installiert.

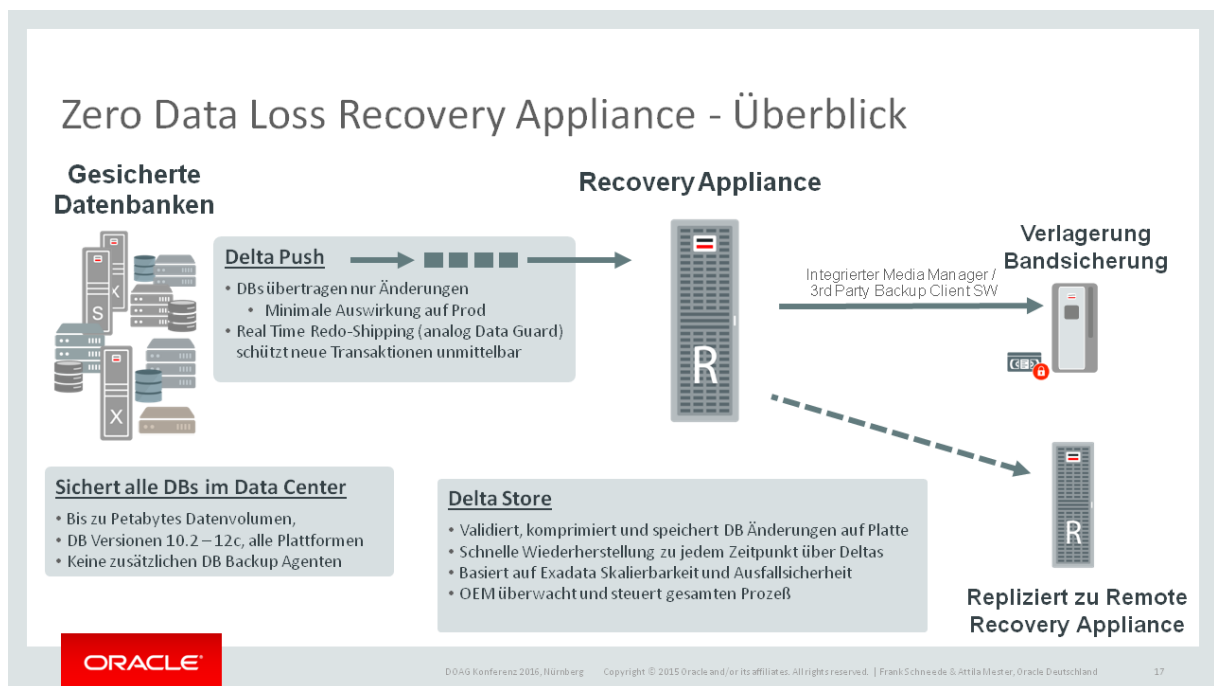


Abb. 3: Recovery Appliance im Rechenzentrum

Die Steuerung der gesamten Oracle Datenbank Backup Umgebung geschieht über den Oracle Enterprise Manager Cloud Control 12c (OEM CC), der mit einem speziellen Recovery Appliance Plugin ergänzt wird. Der OEM CC ist die einfach zu bedienende grafische Oberfläche, die für die Steuerung und Überwachung aller Engineered Systems strategisch wichtig ist. Das Management der Backup Umgebung über den OEM CC umfasst die Parametrisierung der RMAN Backups der zu sichernden Datenbanken, die Konfiguration der Recovery Appliance selbst und die Verwaltung der Replikas sowie der Bandkopien, insofern diese mit OSB erstellt wurden. Es stehen auch umfangreiche Monitoring- und Reporting Funktionen zur Verfügung, um die gesamte Infrastruktur verwalten zu können.

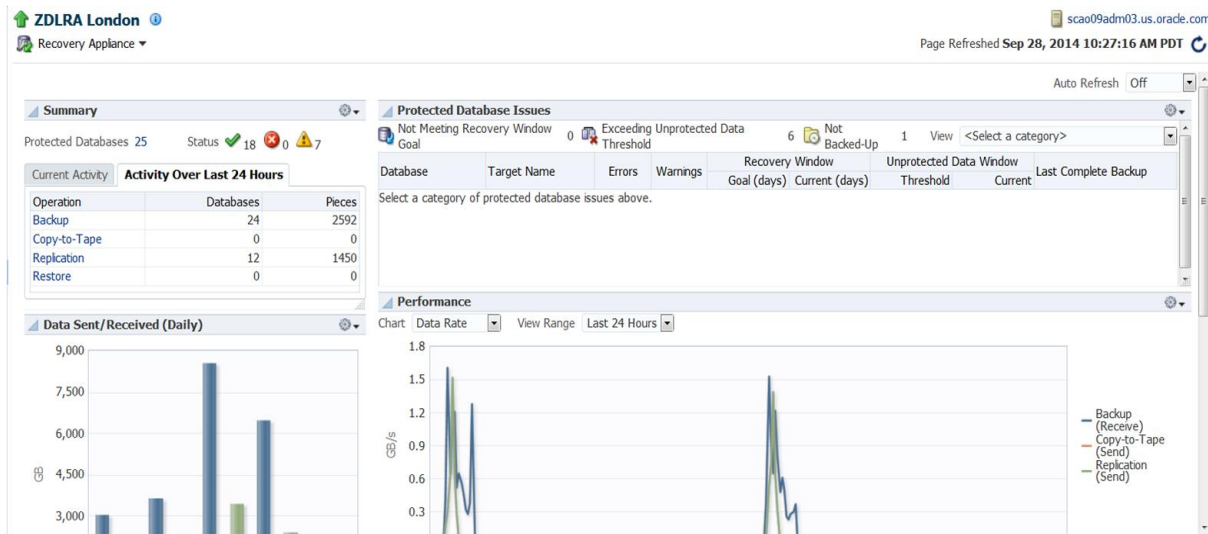


Abb. 4: Grafische Benutzeroberfläche der Recovery Appliance im OEM CC12c

Die Recovery Appliance ermöglicht den Aufbau einer privaten Backup Cloud für die gesamte Oracle Datenbank Infrastruktur eines Rechenzentrums, um somit "Backup as a Service" anzubieten. Um das Backup von hunderten von Oracle Datenbanken effizient zu verwalten, werden Protection Policies gebildet und die Datenbanken diesen zugeordnet. Zu den typischen Attributen der Protection Policies gehören die verschiedenen Recovery Zeitfenster (RTO) für Disk bzw. Tape und Einstellungen für die Replikationen. Die als Replika genutzten Recovery Appliances können eigene Policies enthalten, da diese in den jeweiligen Appliances als Metadaten abgespeichert sind. Rollenaufteilungen für mehrere Administratoren erlauben zudem die Bildung von Management Inseln, wo die jeweiligen Administratoren nur die ihnen zugewiesenen Datenbank Backups sehen und verwalten können.

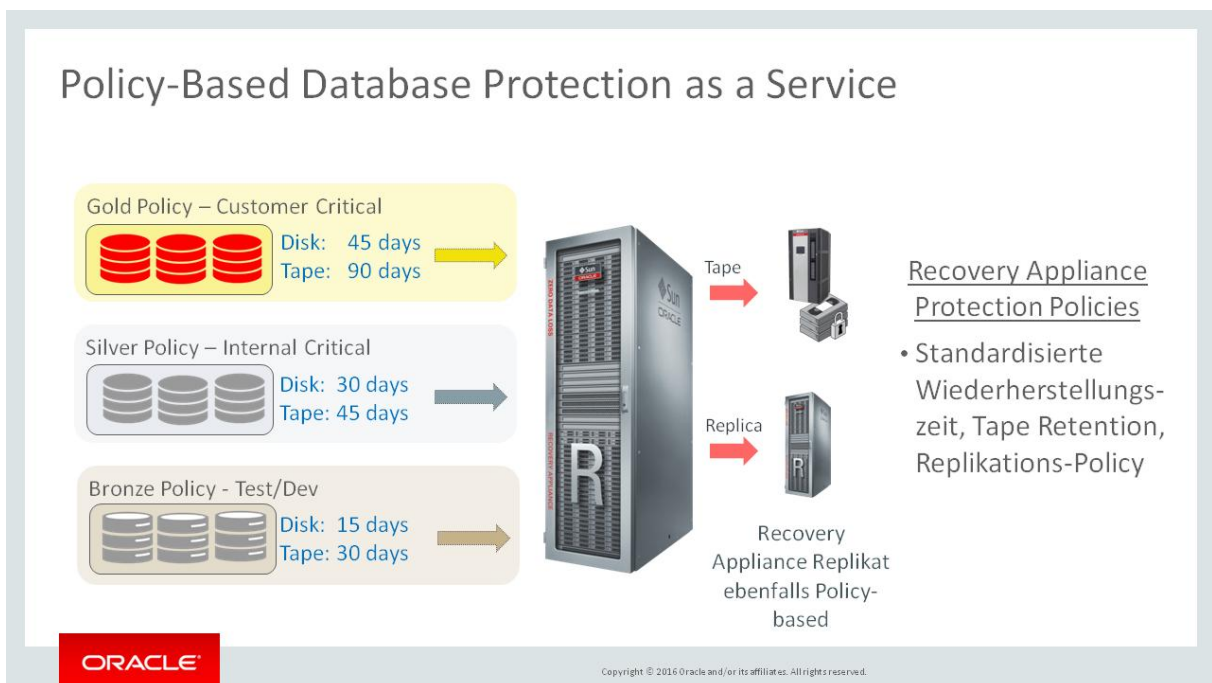


Abb. 5: Beispiel Protection Policies

Konfigurationen und Skalierbarkeit, Kapazität und Performance

Die Recovery Appliance Hardware entspricht einer Exadata mit High Capacity Disks. Die kleinste bestellbare Einheit besteht aus zwei Compute Servern und drei Storage Servern. Diese sind über redundante InfiniBand Switches miteinander verbunden. Die Netzwerkverbindungen nach aussen bestehen aus acht 10 GbE Ports pro Rack, über diese Anschlüsse werden die Backups, Recoveries und Replikation zu anderen Appliances durchgeführt. Die Storage Server enthalten jeweils 12 x 8 TB Disks. Die Basis Einheit mit 3 Storage Servern hat eine nutzbare Kapazität von 94 TB. Die Platten der Storage Server werden, genau wie bei Exadata, mit dem Automatic Storage Management gespiegelt. Dabei werden die Backup Daten einfach gespiegelt abgespeichert, die Metadaten und Katalog Daten sogar mit dreifacher Spiegelung. Vom Basis Rack auf ein Full Rack kommt man durch schrittweisen, bedarfsgesteuerten Ausbau durch einzelne Storage Server.

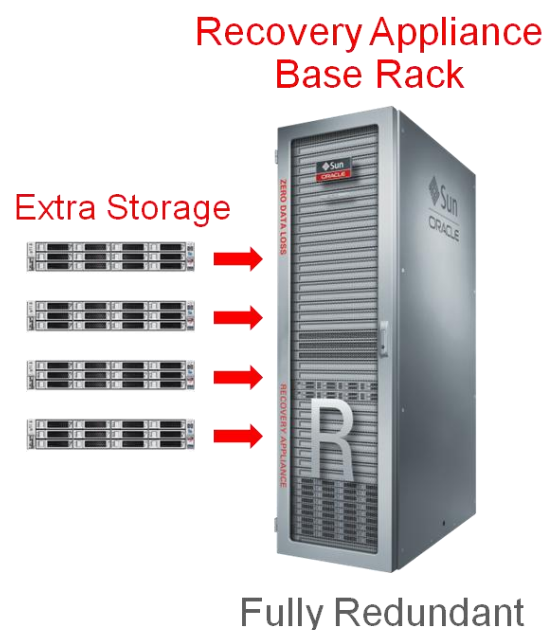


Abbildung 6: Base Rack und Kapazitätserweiterung

Ein Recovery Appliance Full Rack kann bis zu 18 Storage Server aufnehmen, das entspricht einer nutzbaren Kapazität von 580 TB. Die ersten drei Storage Server verfügen über weniger nutzbare Kapazität für Backup Daten als die Erweiterungen, da dort die zusätzlich Metadaten und der RMAN Recovery Katalog untergebracht sind. Die erreichbare tatsächliche (sog. Ingest) Backup- oder Recovery Performance liegt bei ca. 12 TB/h für ein voll ausgebautes Rack. Wenn höhere Kapazitäten und Performance erforderlich sind, können insgesamt 18 Racks miteinander zu einer Einheit zusammengeschaltet werden. Dabei erhöhen sich die Kapazität und der Durchsatz linear.

Als Sizing Richtlinie gilt, dass die Recovery Appliance für 2 Wochen Backup Datenvorhaltung ca. 1 bis 2-mal die Kapazität der zu sichernden Datenbanken haben sollte. Natürlich spielen für die genauere Berechnung die tägliche Änderungsrate, das Redo-Log Aufkommen und die Komprimierbarkeit der Datenbank ebenfalls eine Rolle.

Proof of Concept zur Verifizierung der Performance in großen Umgebungen

Seit der Einführung der Recovery Appliance (RA) im Jahre 2014 wurden zahlreiche PoCs sowohl in Kundenumgebungen wie auch unter Labor Bedingungen durchgeführt. Das hier beschriebene PoC wurde bei einem großen Finanzinstitut in den USA, der mehrere tausend Datenbanken im Einsatz hat, durchgeführt. Die ursprüngliche Umgebung konnte die SLAs hinsichtlich Backup Koordination, Belastung der Produktivumgebung und Recovery Performance nicht mehr erfüllen.

Zum Test wurden 200 Datenbanken mit einem Gesamtvolumen von ca. 80 TB in mehreren Exadata Umgebungen und mit diskreten x86 Servern eingerichtet, sie sollten gleichzeitige Sicherungen auf eine einzige Recovery Appliance durchführen. Die Backups wurden im Hintergrund durch die RA weiter auf Tape Libraries kopiert um zusätzliche Kopien der Daten mit langen Vorhaltezeiten zu bekommen.

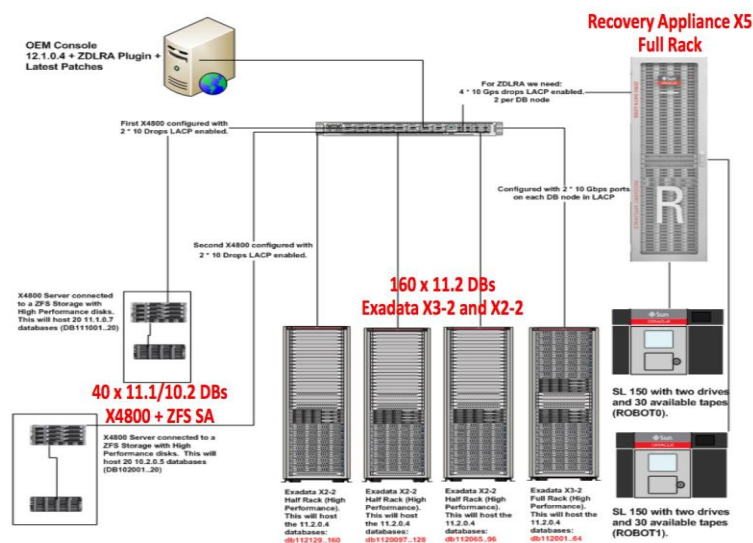


Abbildung 7: PoC Aufbau zum Verifizieren der Performance

Folgende Testszenerien und Anforderungen wurden untersucht:

- Test #1 - Komplette Initial Level 0 Backups von allen 200 Datenbanken innerhalb 24 Stunden
- Test #2 - Vollsicherungen aller Datenbanken aus der RA auf Tape Libraries in max. 7 Tagen kopieren
- Test #3 - Bei 160 Datenbanken Sicherungen < 5 Sekunden RPO erreichen durch Einschalten der Real-time Redo Shipping Option bei 12% tägliche Änderungsrate der Datenbanken
- Test #4 - Komplette Level 1 (inkremental) Backups von allen 200 Datenbanken innerhalb 8 Stunden von allen 200 Datenbanken innerhalb 24 Stunden
- Test #5 - Restore von 2 Datenbanken innerhalb 8 Stunden, während Tests #3 und #4 parallel dazu liefen.

Die Testergebnisse haben die Anforderungen bei allen fünf Szenarien um ein Vielfaches übertroffen. Ein weiteres PoC wurde zur Verifizierung der hohen Restore Durchsatzraten bei sehr großen Datenbanken durchgeführt. Traditionelle Backup Systeme, welche auf de-dup Disk Systeme oder auf Copy-on-Write Disk Storage schreiben, werden bei zunehmender Vorhaltezeit von Backup Versionen immer langsamer, wenn ältere Backups gelesen werden müssen. Bei der Recovery Appliance tritt dieser Effekt wegen der intelligenten Re-Organisation der Backup Daten und der sehr hohen Gesamtleistung des Systems nicht auf. Die Tests haben gezeigt, dass auch bis zu 50 Tage alte virtuelle Full Backups gleich schnell restored werden können.

Lizensierung und Preisgestaltung

Der Preis einer Recovery Appliance setzt sich aus dem Hardware- und der Software Preis zusammen. Die Hardware Preise entsprechen etwa denen der Exadata Maschinen. Obwohl die Funktionen der Recovery Appliance durch die installierten Oracle RAC Datenbanken realisiert sind, fallen keine zusätzlichen Datenbank Lizenzen an, nur von der Kapazität abhängige, spezielle Recovery Appliance Software Lizenzen werden als Software Preis gerechnet.

Fazit

Oracle hat bereits mit Data Guard eine Zero Data Loss Absicherung und mit dem ZFS Backup Appliance ein performantes Backup Target für RMAN Sicherungen. Beide Lösungen kommen aber hauptsächlich als einzelne Lösungen für die besonders wichtigen Datenbanken vor. In Umgebungen, in denen für alle Datenbanken ein Sicherungskonzept gesucht wird oder für viele Datenbanken Zero Data Loss Recovery gewünscht wird und bist jetzt Komplexität und Kosten von Data Guard Architekturen die Implementierung verhindert haben, ist Recovery Appliance jetzt eine Alternative.

Backup Konsolidierung oder Insourcing von Oracle Datenbanken in grossen Rechenzentren ist sicherlich auch ein passendes Anwendungsfeld, dort spielt die Skalierbarkeit und die einfache Administration mit Mandantenfähigkeit eine wichtige Rolle.

Weiterführende Informationen

<https://www.oracle.com/engineered-systems/zero-data-loss-recovery-appliance/index.html>

<http://www.oracle.com/technetwork/database/availability/recovery-appliance-ds-2297776.pdf>

<http://www.oracle.com/technetwork/database/availability/recovery-appliance-twp-2297777.pdf>

http://docs.oracle.com/cd/E55822_01/index.htm

Kontaktadresse:

Frank Schneede
Oracle Deutschland B. V. & Co. KG
Thurnithstr. 2
D-30519 Hannover

Telefon: +49 (0) 511 95787 - 250
E-Mail frank.schneede@oracle.com
Internet: www.oracle.com

Attila Mester
Oracle Deutschland B. V. & Co. KG
Schiffbauergasse 14
D-14467 Potsdam

Telefon: +49 (0) 172 8125947
E-Mail attila.m.mester@oracle.com
Internet: www.oracle.com