

Wie kann eine virtuelle Umgebung besser verfügbar sein als eine physische Umgebung, wo doch mit der Virtualisierung auch zusätzliche Komplexität eingeführt wird? Der Artikel zeigt, dass virtuelle Umgebungen hier mehr zu bieten haben, als es zunächst scheint.

Hochverfügbar durch Virtualisierung – (wie) geht das?

Patrick Schwanke, CarajanDB GmbH

Angesichts der zahlreichen Technologien, die unter den Stichworten „Virtualisierung“ oder „Partitionierung“ laufen, konzentrieren wir uns hier auf Intel-basierte Virtualisierung mit VMware oder Oracle VM. Ziel ist die Reduzierung von Auszeiten bei Hardware-Wartungen, Hardware-Ausfall, Blue Screen oder Kernel Panic, Ausfall der Oracle-Instanz oder auch einer größeren Infrastruktur (Desaster-Fall). Im weiteren Sinne zählen auch der De-facto-Ausfall beziehungsweise die Nichterfüllung von SLAs durch schlechte Antwortzeiten hierzu, auch dabei können virtuelle Umgebungen helfen. Zum Schluss stellt der Autor Erfahrungen hinsichtlich Lizenz-Anforderungen für die unterschiedlichen Absicherungsszenarien vor.

Wartung der Hardware oder des Hypervisors

Techniken wie vMotion von VMware oder VMs Secure Live Migration von Oracle gehören heutzutage zum Standard-Repertoire virtueller Infrastrukturen. Damit können einzelne VMs ohne jede Auszeit auf einen anderen physischen Server migriert werden. Auf diese Weise kann man einen physischen Server „evakuieren“, das heißt alle VMs auf andere Server verlagern. Dann führt man die anstehenden Wartungsarbeiten durch, etwa den Einbau oder Austausch von Hardware oder auch ein Upgrade des Hypervisors. Anschließend startet man den Server durch und bevölkert ihn analog wieder mit VMs.

Im einfachsten Fall geschieht das Ent- und Bevölkern des Servers manuell. In Verbindung mit VMwares voll-

automatischem Dynamic Resource Scheduling (DRS) genügt es auch, für den betreffenden Server den sogenannten „Maintenance-Modus“ zu (de-)aktivieren. Die DRS-Logik führt dann automatisch alle notwendigen vMotion-Events durch.

Ausfall der Hardware oder des Hypervisors

Virtuelle Infrastrukturen bieten eine hochinteressante Alternative zum klassischen Failover-Cluster (active/passive-Cluster), nämlich den sogenannten „HA-Cluster“. Die Namensgebung ist hierbei einheitlich zwischen Oracle VM und VMware, wobei das letztgenannte Unternehmen diese Technik schon wesentlich länger anbietet.

Ein HA-Cluster ist eine Gruppe physischer Server, die wechselseitig füreinander einspringen können, wobei entsprechende Reservekapazitäten konfiguriert werden können und sollten. Beispielsweise sollte ein Fünf-Knoten-Cluster nicht zu mehr als 80 Prozent ausgelastet sein, da ansonsten der Ausfall eines Knotens vom Rest-Cluster nicht geschultert werden könnte.

Der Charme eines HA-Clusters liegt darin, dass mit minimalem Aufwand jede VM mit jeder darin laufenden Anwendung clusterfähig wird, ohne dass man sich mit Techniken wie MSCS, Linux-Cluster oder Oracle RAC One Node beschäftigen muss.

Der Nachteil gegenüber einem herkömmlichen Failover-Cluster ist, dass nach dem Ausfall des aktiven Knotens und der Erkennung dieser Tatsache die VM einfach auf einem anderen physischen Server gestartet wird; der passive Knoten muss also von Null auf gebootet werden. Somit handelt es sich um einen Kaltstart, die Auszeit liegt daher eher im Minuten- als im Sekundenbereich. Da die Knoten mithilfe von Heartbeats überwacht werden, erkennt der Mechanismus auch einen Ausfall des Hypervisors wie VMwares sogenannten „Purple Screen Of Death“ (PSOD, siehe Abbildung 1).

Windows- oder Linux-Ausfall innerhalb der VM

Ein VMware HA-Cluster kann auch überwachen, ob das Betriebssystem innerhalb der VM noch läuft (Stichwort

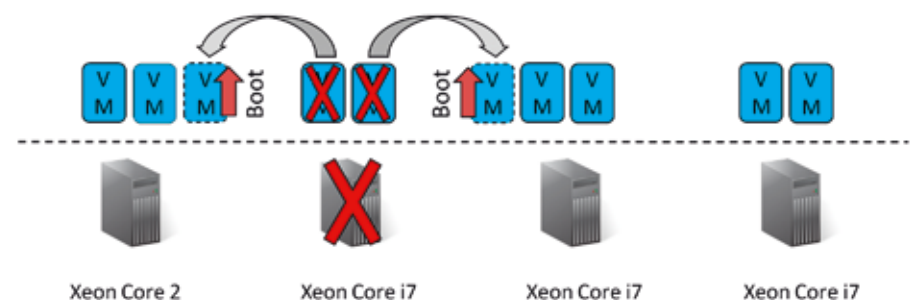


Abbildung 1: VMware HA-Cluster für Knoten- oder Hypervisor-Ausfall

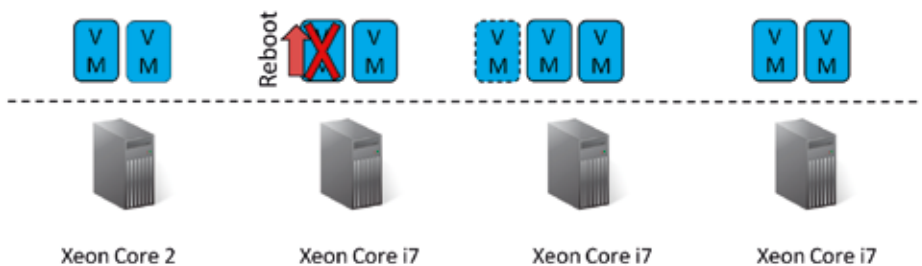


Abbildung 2: VMware HA-Cluster für den Ausfall des Gastbetriebssystems

„Virtual Machine Monitoring“). Im Falle eines Windows-Bluescreen oder eines Linux-Kernel-Panic löst der HA-Cluster einen Reboot der VM aus (siehe Abbildung 2). Nach mehreren erfolglosen Reboots kann man auch eine Alarm-Mail verschicken lassen. Auch hier sorgt ein Heartbeat dafür, solche Situationen zu erkennen, diesmal jedoch ein Heartbeat zwischen der HA-Clusterlogik und den VMware-Tools innerhalb der virtuellen Maschinen. Voraussetzung ist daher, dass – zumindest für die zu überwachenden VMs – die VMware-Tools installiert sind. Für Windows-VMs ist dies normalerweise kein Problem, für Linux-VMs wird ein GCC-Compiler benötigt, der aber ohnehin für die Installation der Oracle-Software benötigt wird.

Bevor man diese Überwachungsfunktion aktiviert, sollte man peinlich genau prüfen, ob auch wirklich in allen betreffenden VMs die VMware-Tools installiert sind und die anderen VMs aus der Überwachung ausgenommen sind. Ansonsten führt das Nicht-Vorhandensein der VMware-Tools nämlich zu fehlenden Heartbeats, sodass die VM permanent durchgestartet würde. Ähnliches kann passieren, wenn die VM extrem ausgelastet ist und die VMware-Tools keine Prozessorzeit mehr bekommen. Lässt der Heartbeat länger auf sich warten als der eingestellte Timeout, wird die VM durchgestartet. Der Timeout sollte daher nicht zu gering gewählt werden. Ein zweistelliger Sekundenbetrag ist zu empfehlen.

Ausfall der Oracle-Instanz

Hinsichtlich eines Ausfalls der Oracle-Instanz oder des Listener kann keine der Virtualisierungslösungen einen

direkten Mehrwert erbringen. Genau wie in der physischen Welt ist hier eine Überwachung notwendig, beispielsweise mithilfe der Oracle-Clusterware. Seit der Datenbank-Version 11g R2 kann diese nicht nur für den Aufbau von RAC-Datenbanken verwendet werden, sondern auch für Nicht-RAC-Datenbanken einen automatischen Neustart von Listener, Datenbank- oder ASM-Instanz veranlassen (Stichwort „Oracle Restart“). Wichtig zu wissen ist, dass ab der Datenbank-Version 11.2.0.2 auch RAC-Instanzen in virtuellen Maschinen unterstützt werden (siehe [1]). Unter [2] findet sich eine Beschreibung, wie ein RAC-Cluster auf Basis von VMware aufgesetzt wird.

Disaster Recovery

Auch wenn das Thema „Disaster Recovery“ nicht Schwerpunkt dieses Artikels ist, seien hier kurz einige Lösungen vorgestellt. VMware selbst bietet mit dem Site Recovery Manager (vSRM) und dem seit vSRM5 enthaltenen vSphere Replication eine Lösung, um virtuelle Maschinen an einen entfernten Standort zu replizieren.

Ähnliches liefern Drittanbieter-Lösungen wie vReplicator von Quest

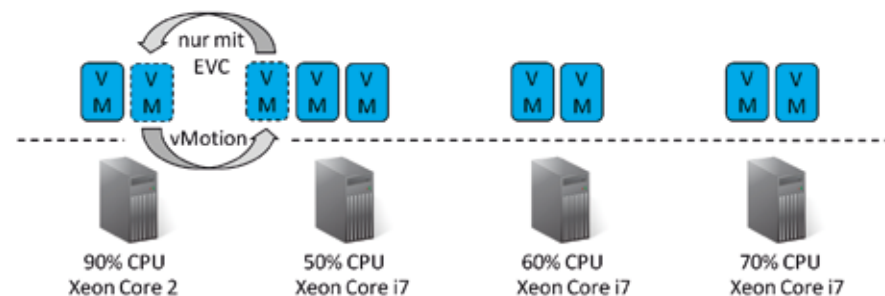


Abbildung 3: VMware DRS-Cluster für Lastverteilung

Software oder Backup & Replication von Veeam.

Auch VMware Fault Tolerance (FT) wirkt auf den ersten Blick interessant, da es eine Replikation einzelner VMs ohne spezielle Storage-Anforderungen leistet und Bestandteil der vSphere 4 Advanced Edition beziehungsweise vSphere5 Enterprise Edition ist. Hier sind jedoch zahlreiche Restriktionen zu beachten, die in [3] detailliert beschrieben sind. Insbesondere ist FT auf VMs mit einer einzigen vCPU beschränkt. Kritische Datenbanken sind dagegen oftmals mit mehr CPU-Kernen ausgestattet oder benötigen diese Option für die Zukunft. In diesen Fällen kommt FT nicht in Frage.

Grundsätzlich sollte man bei virtualisierten Oracle-Datenbanken Oracle-eigene DR-Szenarien auf Basis von Physical-Standby-Datenbanken prüfen, da diese sehr robust und einfach zu verwalten sind. Sei es per Data Guard für die Enterprise Edition oder per Skript beziehungsweise mithilfe von Werkzeugen wie Dbvisit Standby oder Libelle DBShadow für die Oracle Standard Edition (One).

Lastverteilung

Ein Aspekt, den man nicht unterschätzen sollte, ist die automatische Lastverteilung. Bei VMware läuft diese Funktion seit längerem unter der Bezeichnung „Dynamic Resource Scheduling“ (DRS, siehe Abbildung 3). Oracle VM bietet dies in der neuesten Version 3.0 unter demselben Namen an. Dabei werden mehrere physische Server zu einem DRS-Cluster zusammengeschlossen. Im Oracle-Umfeld sollte man dies nicht als Alternative

zu RAC sehen, wohl aber als Ergänzung. Ein DRS-Cluster überwacht die Auslastung der physischen Server und verschiebt bei Ressourcen-Engpässen einzelne VMs per vMotion auf einen anderen physischen Server mit geringerer Auslastung.

DRS liefert zwar keinen zusätzlichen Schutz gegen Ausfall irgendwelcher Komponenten, hilft aber, schlechte Antwortzeiten durch überlastete Infrastruktur zu vermeiden. Sobald es also nicht nur um nominelle Verfügbarkeit geht, sondern auch um die Einhaltung von SLAs und die Zusicherung einer gewünschten Service-Qualität, ist DRS eine interessante Ergänzung.

Einige Hinweise aus den Erfahrungen des Autors:

- Ein DRS-Cluster sollte möglichst von vornherein im EVC-Modus (Enhanced vMotion Compatibility) erstellt und betrieben werden. Nachträglich lässt sich EVC nur noch mit einer Auszeit für alle VMs aktivieren. Wenn sich im Laufe der Zeit unterschiedliche CPU-Generationen in der Hardware ansammeln, werden DRS und vMotion ohne EVC zu einer Einbahnstraße, das heißt vMotion ist nur zu identischen oder neueren Prozessormodellen hin möglich. Der Weg zurück auf einen älteren Prozessor ist dann verwehrt, Server mit älteren Prozessoren verweisen zunehmend. Abbildung 3 zeigt dies.
- Da Oracle-RAC nicht nur Lastverteilung, sondern auch Hochverfügbarkeit realisieren soll, dürfen bei einem DRS-Cluster die RAC-VMs einer Datenbank nicht auf demselben physischen Server laufen, da dieser ansonsten einen Single Point of Failure darstellen würde. Hierfür setzt man eine sogenannte „Anti-Affinity-Regel“ ein, die verhindert, dass DRS die beiden Instanzen auf denselben physischen Server schiebt.
- DRS kann im voll- sowie im halbautomatischen Modus betrieben werden. Im letzteren Fall werden lediglich vMotion-Empfehlungen ausgesprochen, die ein VMware-Administrator per Knopfdruck umsetzen kann. Oftmals geht dadurch

aber unnötig Zeit verloren. Der Autor empfiehlt daher den vollautomatischen Modus. Die gelegentlich anzutreffende Befürchtung, dass dann ständig unkontrolliert VMs verschoben werden, kann er nicht bestätigen. Und selbst wenn dem so wäre, wird durch vMotion-Ereignisse keinerlei Auszeit verursacht. Außerdem kann man die vMotion-Häufigkeit durch den Aggressivitätslevel des DRS-Clusters beeinflussen.

- DRS-Clustering kann mit dem bereits beschriebenen HA-Clustering kombiniert werden. Dies ist auch die übliche Vorgehensweise in der Praxis; es wird mit kombinierten HA- und DRS-Clustern gearbeitet.

Lizenzierung

Alle Lizenz-Informationen werden nach bestem Wissen des Autors gegeben und spiegeln den Stand Dezember 2011 wider. Verbindliche Informationen finden sich jeweils bei den Herstellern selbst, entsprechende Links sind nachfolgend angegeben. Hinsichtlich der erforderlichen Lizenzierung sollte man zunächst entscheiden, welche VMware-Lizenzedition benötigt wird. Abbildung 4 gibt einen Überblick über die verschiedenen Editionen von vSphere v4 und v5 im Hinblick auf die hier besprochenen Clustering-Fähigkeiten. Als weitere Merkmale sind Obergrenzen hinsichtlich CPU- und Memory-Ausstattung der physischen Server und der VMs zu

beachten. Schließlich ist auch die sogenannte „Hot-Add-Funktion“ aufgeführt, die das Ändern der CPU- und Memory-Ausstattung einer VM im laufenden Betrieb ermöglicht, was bei Ressourcen-Engpässen innerhalb der VM für eine einfache, vertikale Skalierung sorgt.

Verbindliche Lizenzinformationen finden sich auf der VMware-Webseite, etwa unter [4]. Angesichts der aufgeführten Merkmale empfiehlt der Autor in den meisten Fällen die vSphere Enterprise Edition (siehe Abbildung 4).

Oracle unterteilt die verschiedenen Virtualisierungstechniken in Hard- und Soft-Partitioning, wobei die genaue Abgrenzung in [5] beschrieben ist. Dabei zählt VMware grundsätzlich als Soft-Partitioning, es müssen also Oracle-Datenbank-Lizenzen für den gesamten vSphere-Server erworben werden, auf dem eine VM mit einer Oracle-Datenbank läuft. Wird ein HA- und/oder DRS-Cluster benutzt, sind alle vSphere-Server des Clusters zu lizenzieren. Für größere Datenbank-Umgebungen lässt sich dieser Nachteil umgehen, indem ein Oracle-Only-Cluster (oder ein Oracle-Only-vSphere-Server) aufgebaut wird, auf dem ausschließlich Oracle-Datenbanken laufen. In diesem Falle wird sowieso die komplette Hardware benutzt und lizenziert. Je geringer dagegen der Anteil an Oracle-Software, desto höher der Lizenz-Overhead.

Oracle VM wird dagegen unter bestimmten Bedingungen, nämlich mit

	v4 Standard	v4 Advanced	v4 Enterprise	v4 Ent. Plus	v5 Standard	v5 Enterprise	v5 Ent. Plus
Max RAM pro Server	256GB	256GB	256GB	1TB	No limit	No limit	No limit
Max CPU-cores pro CPU-Sockel	6	12	6	12	No limit	No limit	No limit
Max vRAM pro CPU-Sockel	No limit	No limit	No limit	No limit	32GB	64GB	96GB
Max CPU-cores pro VM	4	4	4	8	8	8	32
vMotion und HA-Cluster	Ja	Ja	Ja	Ja	Ja	Ja	Ja
vCPU & vMem Hot-Add	Nein	Ja	Ja	Ja	Nein	Ja	Ja
DRS-Cluster	Nein	Nein	Ja	Ja	Nein	Ja	Ja

Abbildung 4: VMware-Lizenzeditionen

aktiviertem CPU-Pinning, als Hard-Partitioning gewertet (siehe [6]), sodass nur die tatsächlich benutzten physikalischen CPUs lizenziert werden müssen. Bisher ging dadurch allerdings die „Live Migration“-Fähigkeit, also das Pendant zu vMotion, verloren beziehungsweise es musste auf dem neuen physikalischen Server ein erneutes CPU-Pinning durchgeführt werden. Der Autor hat bisher keine Erfahrungen darüber, ob die in Oracle VM 3 enthaltenen HA- und DRS-Funktionen mit der CPU-Pinning-Anforderung verträglich, ob also HA und DRS in einer Hard-Partitioning-Umgebung möglich sind.

Fazit

Verschiedenste Arten geplanter und ungeplanter Auszeiten können mithilfe

der beschriebenen Virtualisierungstechniken umgangen oder zumindest reduziert werden. VMware hat nach wie vor einen technologischen Vorsprung vor Oracle VM. Oracle VM hat erst mit der neuesten Version 3 vergleichbare Techniken eingebaut, hier stehen praktische Erfahrungen noch aus. Lizenztechnisch lohnt sich der Einsatz von VMware jedoch nur, wenn reine Oracle-Umgebungen aufgebaut werden, also separate vSphere-Server oder -Cluster nur für Oracle-Datenbanken.

Referenzen

- [1] Support für Oracle-Produkte auf VMWare: MOS Doc-ID 249212.1
- [2] Aufbau eines RAC-Clusters auf Basis von VMware vSphere: <http://streetkiter.wordpress.com/2011/11/08/oracle-rac-11-2-0-3-auf-vmware-esxi/>
- [3] Restriktionen bei VMware Fault Tolerance: <http://communities.vmware.com/blogs/>

- vmroyale/2009/05/18/vmware-fault-tolerance-requirements-and-limitations
- [4] VMware-Lizenzinformationen: http://www.vmware.com/files/pdf/vsphere_pricing.pdf
- [5] Oracle Hard Partitioning und Soft Partitioning: <http://www.oracle.com/us/corporate/pricing/partitioning-070609.pdf>
- [6] Hard Partitioning mit Oracle VM: <http://www.oracle.com/technetwork/topics/virtualization/ovm-hardpart-167739.pdf>

Patrick Schwanke
 CarajanDB GmbH
 patrick.schwanke@carajandb.com



IT-Consulting	Schulungen	Software-Lösungen	Oracle Lizenzen
<ul style="list-style-type: none"> › Performance Tuning <ul style="list-style-type: none"> • Oracle Datenbank Tuning • Oracle SQL + PL/SQL Tuning › Real Application Clusters › Data Guard + Fail Safe › Datenbank Management <ul style="list-style-type: none"> • Konfiguration • Backup & Recovery • Migration und Upgrade › OEM Grid Control › Oracle Security › Services <ul style="list-style-type: none"> • Remote DBA Services • Telefon-/Remotesupport <p>Nutzen Sie unsere Kompetenz für Ihre Oracle Datenbanken.</p>	<ul style="list-style-type: none"> › Oracle SQL › Oracle PL/SQL › Oracle DBA › Oracle APEX › Backup & Recovery › RMAN › Neuerungen 10g/11g › Datenbank Tuning › Datenbank Monitoring › Datenbank Security Wir bieten Ihnen öffentliche Kurse sowie Inhouse-Schulungen. 	<ul style="list-style-type: none"> › Individualsoftware <ul style="list-style-type: none"> • .NET und Visual Basic • Java › Oracle APEX › PL/SQL Unser Ziel: Individuelle Softwareentwicklung mit Fokus auf Ihre Zufriedenheit. 	<ul style="list-style-type: none"> › Oracle Datenbanken <ul style="list-style-type: none"> • Standard Edition One • Standard Edition • Enterprise Edition • Personal Edition › Oracle Produkte <ul style="list-style-type: none"> • Enterprise Manager • Oracle Tools Optimale Lizenzierung durch individuelle Beratung.

