



Einfach, sicher, MiniCluster

Stefan Hinker, ORACLE Deutschland B.V. & Co. KG

Neben der Oracle Cloud Machine, der Exadata X6 und der ODA X6 wurde in diesem Sommer ein weiteres Engineered System angekündigt. Diesmal nicht unter dem Motto „groß, schnell, teuer“, sondern als „klein, einfach, sicher und preiswert“. Die Rede ist vom Oracle MiniCluster S7-2. Grundlage des Systems sind die gleichzeitig angekündigten S7-Server, in denen Oracle die schon länger erwarteten „Sonoma“-CPUs auf den Markt bringt.

Der MiniCluster ist ein Engineered System wie die Exadata Database Machine oder der SuperCluster, als dessen kleinen Bruder man ihn durchaus betrachten kann. Damit ist der Anspruch klar: wenige, sauber definierte Konfigurations-Optionen, einfachste, schnelle Installation und höchste Leistung. Während Leistung natürlich bei einem eher kleinen System durch die verwendeten Komponenten vorgegeben ist, hat Oracle beim Thema „einfach“ die Messlatte ein Stück höher gelegt. Ziel war es, die Installation und den normalen Betrieb vollständig im Browser zu steuern und den Kontakt zur Kommandozeile überflüssig zu machen. Gleichzeitig sollte MiniCluster beim Thema „Sicherheit“ keine Kompromisse machen – im Jahr 2016 durchaus zeitgemäß, wenn

auch nicht selbstverständlich. Zu den Details: Man kann sich für den MiniCluster oder gegen ihn entscheiden, mehr Optionen gibt es nicht. Entscheidet man sich dafür, bekommt man Folgendes:

- Zwei S7-2-Server mit jeweils
 - Zwei SPARC-S7-CPUs mit je acht Kernen, die mit 4.27 GHz noch ein wenig höher takten als die in den T- und M-Modellen verwendete SPARC M7
 - 512 GB RAM aus sechzehn 32-GB-Modulen
 - Viermal 10 GBit Ethernet (onboard) in Base-T-Ausführung. Davon sind zwei Ports für den Interconnect zwischen beiden Systemen reserviert.
 - Viermal 10 GBit Ethernet mit Glasfaser-Anschluss
 - Sechsmal 1.2 TB interne Festplatten für Betriebssystem und lokale Daten
- Ein Plattengehäuse „DE3-24C“ mit
 - Vierzehnmal 1.6-TB-SSD zur Nutzung als Datenbank-Speicher. Oracle formatiert hier auf 1.2 TB, um die verbleibende Kapazität zur Lebensverlängerung zu nutzen. Bei Normal Redundancy stehen so ca. 7.2 TB netto als Datenbank-Speicher zur Verfügung.
 - Viermal 200 GB schreib-optimierte SSD zur Nutzung als Redo-Log-Devices
 - Sechsmal 8 TB Festplatten für sonstigen Speicherbedarf

Bei der Hardware drängen sich Parallelen zur Oracle Database Appliance (ODA) auf. Die erste auffällige Komponente sind die beiden Rechenknoten: S7-2, aufbauend auf den vor etwa einem Jahr angekündigten und nun verfügbaren „Sonoma“-CPUs. Details dazu unter [1]. Zu erwähnen ist die auch in dieser CPU vorhandene Software in Silicon-Features: Crypto-Support für fünfzehn verschiedene Algorithmen, Silicon Secured Memory sowie die „Database Acceleration Engine“ (DAX), um die In-Memory-Operationen der Datenbank signifikant zu beschleunigen.

Insgesamt erhält man zwei Knoten mit je sechzehn SPARC-S7-Kernen mit den von den T7-/M7-Systemen her bekannten Eigenschaften. Davon werden für die Infrastruktur pro Knoten zwei Kerne reserviert, sodass zweimal vierzehn Kerne für Datenbanken beziehungsweise Anwendungen zur Verfügung stehen.

Die Memory-Ausstattung erscheint auf den ersten Blick mit 512 GB pro Knoten etwas dürftig. Allerdings ist das Verhältnis von RAM zu CPU sehr gut ausbalanciert: Aus sechzehn Kernen und 512 GB RAM pro Knoten ergibt sich ein Verhältnis von 32 GB pro Kern – verglichen mit ähnlich ausgestatteten T- oder X- Servern ist das recht üppig. Eine ODA X5-2 beispielsweise kommt selbst auf 786 GB aufgerüstet nur auf ein Verhältnis von etwa 21 GB pro Kern.

Keine Kompromisse gibt es beim Speicher für die Datenbanken. Hier setzt MiniCluster ausschließlich auf Flash. Anders als bei der neuen ODA X6 wird dabei jedoch nicht auf Hochverfügbarkeit verzichtet – beide Compute-Knoten haben redundanten Zugang zu den Dual-ported-SAS-SSDs. Die ebenfalls enthaltenen 8 TB großen SAS-Festplatten sind ausschließlich als Shared-Storage für unstrukturierte Daten verwendet.

Aus dieser eigentlich recht einfachen Hardware baut Oracle tatsächlich einen kleinen MiniCluster. Genau wie beim SuperCluster gibt es Umgebungen für die Datenbank sowie für allgemeine Anwendungen. Beide lassen sich voll automatisiert anlegen und installieren – im Falle der Datenbank bis hin zur laufenden RAC-Datenbank, im Falle der allgemeinen Anwendung natürlich nur bis zum laufenden Betriebssystem. Dabei ist man dem Ziel „alles im Browser“ sehr nahe gekommen:

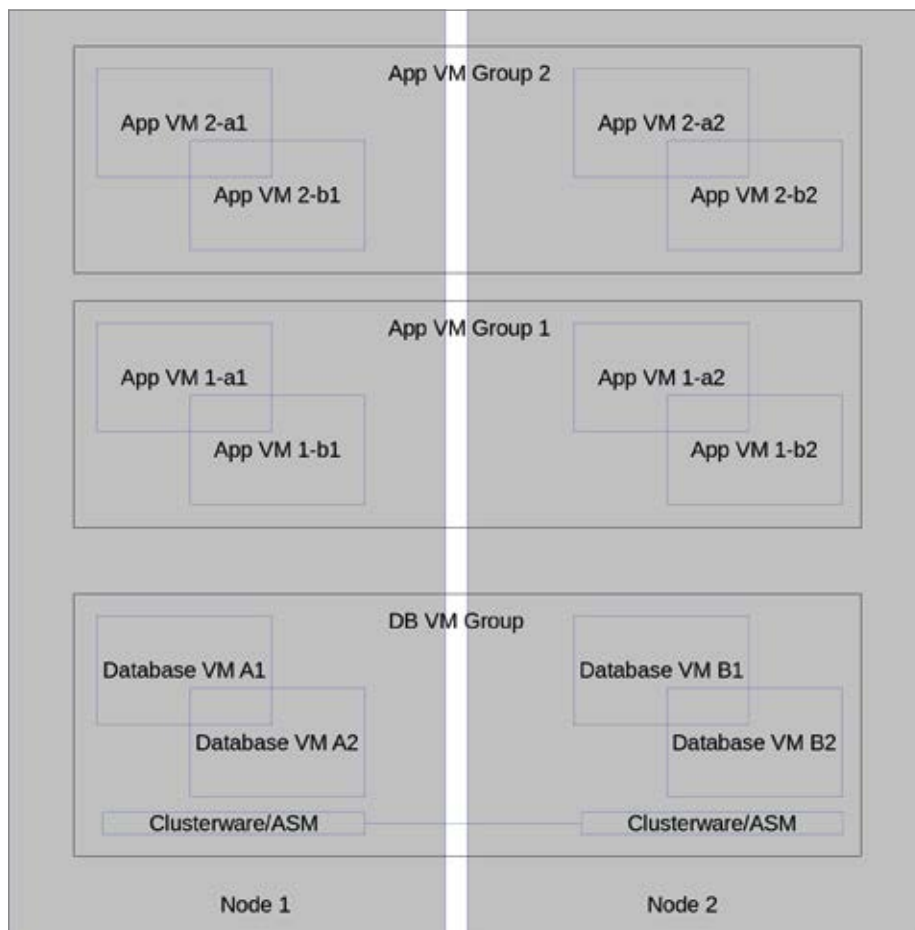


Abbildung 1: VM Gruppen

- Die initiale Konfiguration erfolgt mit einem Offline-Tool, das unter Windows, Linux, MacOS oder Solaris zum Einsatz kommen kann. Dort werden Basis-Parameter wie Hostnamen, IP-Adressen, Zeitzonen etc. abgefragt. Hier macht man auch das erste Mal Bekanntschaft mit der Sicherheits-Philosophie des Systems: Man muss sich zwischen dem „PCI-DSS Profile“ (Payment Card Data Security Standard) und „CIS Equivalent Profile“ (einem Sicherheitsprofil auf Basis der Benchmarks des Center of Internet Security) entscheiden. Das System wird dann so konfiguriert, dass die entsprechenden Security-Benchmarks weitestgehend erfüllt sind. Eine dritte Option steht nicht zur Wahl. Um Details kann und muss man sich nicht kümmern. Die gesammelten Daten werden in einer Konfigurationsdatei gespeichert.
- Diese Datei überträgt man auf das frisch ausgepackte, angeschaltete System und startet dann die Erstinstallation. Dies ist der einzige Schritt, der eine SSH-Verbindung erfordert und nicht im Browser abläuft. Es gibt ein Video, das diesen Schritt zeigt [2].
- Der Abschluss der Konfiguration wird dann jedoch bereits im Browser gesteuert und man kann den Kommandozeilen-Login getrost beiseitelegen. Das Resultat ist die fertig konfigurierte Basis-Infrastruktur:
 - Eine kleine virtuelle Maschine, die aus den sechsmal 8TB Platten ein hochverfügbares NFS liefert
 - Die SSDs sind fertig vorformatiert, um im nächsten Schritt von der Datenbank verwendet zu werden
 - Auf dem primären Rechenknoten läuft auf Port 443 der Webserver, über den das System verwaltet wird
- Nach einer einmaligen Konfiguration des Administrator-Accounts kann es losgehen mit der Erzeugung der Datenbank- und Anwendungs-Umgebungen.

In der aktuellen Software-Version erlaubt der MiniCluster den Betrieb einer Grid-Infrastructure-Umgebung, der sogenannten „DB VM Gruppe“, sowie mehre-

re Anwendungs-Umgebungen. In dieser Grid-Infrastructure-Umgebung können dabei pro Knoten bis zu vier VMs konfiguriert sein. Grob skizziert sieht das dann so aus wie im Diagramm „VM Gruppen“ in *Abbildung 1*.

In den einzelnen VMs können nun unterschiedliche Datenbanken installiert sein – es stehen 11g und 12c jeweils in den Varianten Single-Instance, RAC-One-Node und RAC zur Auswahl. (Standard Edition 2 ist in Vorbereitung). Eine VM kann dabei durchaus mehrere verschiedene Oracle-Homes und mehrere verschiedene Datenbanken enthalten, das System ist hier trotz der stark vereinfachten Installation sehr flexibel. Die Installation erfolgt dabei in mehreren Schritten. Insgesamt dauert dieser Prozess etwa ein bis zwei Stunden, die im Wesentlichen in den jeweiligen Deploy-Phasen verstreichen (*siehe Abbildung 2*).

Im Vergleich zur manuellen Installation auf selbst installierten Systemen, aber auch im Vergleich zur Installation auf ODA, SuperCluster oder Exadata Database Machine mit OEDA beziehungsweise Java-one-Command ist die Installation einer RAC-Datenbank hier wirklich einfach und vor allem dank der fast vollständigen Automatisierung zuverlässig. Es gibt natürlich weniger Freiheiten in der Konfiguration, dafür bekommt man aber eine vorab getestete, optimierte und stabile Datenbank-Installation. Das Ziel „Push-Button Simplicity“ ist hier eindeutig erreicht. Eine Kommandozeile wird tatsächlich nicht gebraucht, was den MiniCluster zur idealen Plattform auch für diejenigen macht, die bisher wenig mit Unix oder Linux zu tun hatten.

Die Anwendungs-Umgebungen sind in ganz ähnlicher Weise angelegt. Hier gibt es die Möglichkeit, mehrere Umgebungen zu konfigurieren. Die Grenze ist die Anzahl der fest zugewiesenen CPU-Kerne. Verzichtet man auf eine feste Zuweisung, teilen sich alle VMs alle CPUs, was bei zeitlich unterschiedlicher Auslastung die Gesamt-Effizienz des Systems steigern kann. Diese Zuweisung lässt sich später jederzeit ändern. Ist eine Anwendungs-VM fertig, steht ein Login mittels SSH zur Verfügung. Damit und über das optionale NFS werden dann Anwendungen installiert.

Ein ganz wesentliches Merkmal des MiniCluster ist die kompromisslose Sicherheits-Philosophie. Das System wurde so entwickelt, dass es ohne wesentli-

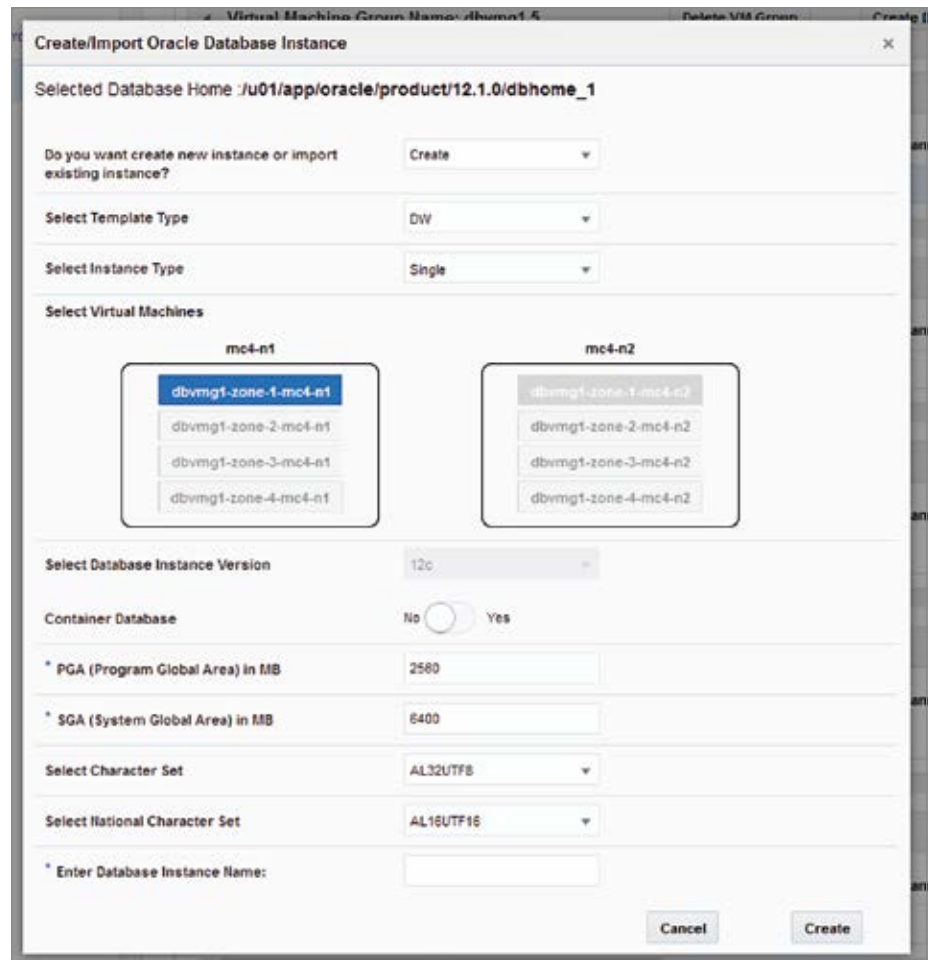


Abbildung 2: Eine Datenbank anlegen

che Nacharbeiten einen Compliance-Test für PCI-DSS oder CIS (beziehungsweise STIG) besteht. Dafür wurden sowohl die Infrastruktur als auch die VMs bei der Installation entsprechend angepasst. Zum gewählten Sicherheitsprofil kann man Compliance-Berichte in der Benutzeroberfläche anfordern und später abrufen. Sie geben eine gute Übersicht über die jeweiligen Tests und deren Ergebnisse. Selbst wenn keine 100-prozentige Compliance erreicht wird, sind die Compliance-Berichte eine gute Hilfe, um etwa bei einem Audit etwaige Abweichungen aufgrund spezifischer Anforderungen zu erklären. So kann man beispielsweise wählen, ob eine VM-Gruppe Zugriff auf das zentrale NFS bekommen soll. Hat sie diesen Zugriff, muss der entsprechende Compliance-Test natürlich fehlschlagen (Shared Storage ist in der Regel nicht mit PCI-DSS vereinbar). Während das System hier die Möglichkeit bietet, für jede VM entweder einmalig oder wiederkehrend Berichte anzufertigen, müssen eventu-

elle weitere Verbesserungen (etwa ein Passwort für das Booten des Systems) manuell vorgenommen werden. Im Vergleich mit einem entsprechenden Bericht von einem manuell installierten System fallen schnell prominente Punkte auf, bei denen MiniCluster gehärtet wurde. Einige Beispiele sind:

- Alle Non-root-Dateisysteme sowie der Swap-Bereich sind verschlüsselt
- Strenge Passwortregeln
- Größenbegrenzung für /tmp
- Keine Removable Media
- Kein DNS Multicasting
- Korrekte Konfiguration von NTP
- Diverse OS-Dienste sind ausgeschaltet (CUPS, net-snmp, r-services etc.)

Zum Schluss die Frage nach der Leistung: Wie bei Engineered Systems üblich, veröffentlicht Oracle keine Benchmarks für MiniCluster. Wenn man jedoch die beiden für die Leistung ausschlaggebenden Komponenten „IO“ und „CPU“ betrachtet,

bekommt man durchaus einen Eindruck von der Leistungsfähigkeit. Für die Datenbank-Leistung ist oft der Speicher bestimmend. Da Oracle hier ausschließlich auf Flash setzt, ist eine recht hohe Leistung zu erwarten. Ein kleines Experiment mit „CALIBRATE_IO“ nennt knapp 650k IOPS oder 8.6 GB/Sek. Das kann sich durchaus sehen lassen, insbesondere da die entsprechende Latenz offenbar zu klein für die verwendete Maßeinheit ist.

Auch im Bereich „CPU-Leistung“ wird man auf den Seiten von Oracle recht schnell fündig. Auf dem „Bestperf“-Performance-Blog [3] stehen unter anderem Werte für SPECjEnterprise 2010 und SPECjbb2015, die pro CPU-Kern einen Vorsprung von 1.8x bis 3.8x vor der Intel E5-2699v4 CPU zeigen. Die Leistung der Hardware-Unterstützung für die In-Memory-Features der Datenbank wird in zwei weiteren Beispielen anhand der „Real Cardinality Database“ gezeigt. Hier hat die SPARC-CPU die Nase um den Faktor 2 beziehungsweise sogar 7.7 vorn – pro Kern.

Auch wenn im Falle des RCDB-In-Memory-Benchmarks vermutet werden darf, dass dieser die Vorzüge der CPU und des DAX besonders hervorhebt, ist ein Leistungsvorsprung nicht von der Hand zu weisen. Da sich die Leistung pro Kern – eine EE-Lizenzierung angenommen – direkt in eine Leistung pro Lizenz umrechnen lässt, wird dann auch verständlich, warum Oracle mit hoher Leistung und Effizienz wirbt. Selbst wenn man, vorsichtig gerechnet, nur 1,5 x86-Broadwell-Kerne für einen SPARC-Kern annimmt, bezahlt sich der MiniCluster schnell durch eingesparte Lizenzen. Den Gewinn in der Administration durch die Automatisierung noch nicht eingerechnet.

Fazit

Einfache Installation und Administration, hohe und vor allem nachprüfbar sichere bereits in der Basis-Konfiguration sowie eine Leistung, die in Anbetracht der vorhandenen Hardware das Maximum aus den einzusetzenden Lizenzen herausholt, sind die Kernpunkte des MiniCluster. Das hohe Verhältnis von Memory pro Kern in Verbindung mit der Hardware-Unterstützung für die In-Memory-Opera-

tionen macht die Maschine für Decision-Support-Anwendungen interessant.

Da die Dekompression der In-Memory-Area in der CPU unterstützt wird, können in dem auf den ersten Blick kleinen Memory durchaus ansehnliche Datenmengen ohne Leistungsverlust untergebracht werden. Dabei wird Sicherheit großgeschrieben und auf Hochverfügbarkeit, wie bei einem Engineered System nicht anders zu erwarten, nicht verzichtet.

In Verbindung mit der einfachen, weitgehend automatisierten Administration ist MiniCluster damit gerade auch für diejenigen interessant, die wegen der Administration unbekannter Betriebssysteme oder der Komplexität und Größe eines SuperCluster bisher auf diese Kombination verzichtet haben.

Weiterführende Links

- [1] S7 CPU Datasheet: <http://www.oracle.com/us/products/servers-storage/sparc-s7-processor-ds-3042417.pdf>
- [2] https://otube.oracle.com/media/Oracle+MiniCluster+S7-2+Initial+Installation/0_ayuu8ecy/2346
- [3] Leistung der S7-CPU (einschl. Referenzen für im Artikel genannte Leistung): <https://blogs.oracle.com/BestPerf>
- [4] https://blogs.oracle.com/BestPerf/entry/20160629_imdb_sparc_s7_2
- [5] https://blogs.oracle.com/BestPerf/entry/20160629_rte_sparc_s7_2
- [6] https://blogs.oracle.com/BestPerf/entry/20160629_jent_sparc_s7_2
- [7] https://blogs.oracle.com/BestPerf/entry/20160629_jbb_sparc_s7_2



Stefan Hinker
stefan.hinker@oracle.com



Exzellente Baupläne für die Digitale Ökonomie!

Dafür steht PROMATIS als Geschäftsprozess-Spezialist mit mehr als 20 Jahren Erfahrung im Markt. Gepaart mit profundem Oracle Know-how schaffen wir für unsere Kunden die Digitale Transformation:

- Oracle SaaS für ERP, SCM, EPM, CX, HCM
- Oracle E-Business Suite und Hyperion
- Oracle Fusion Middleware (PaaS)
- Internet of Things und Industrie 4.0

Vertrauen Sie unserer Expertise als einer der erfahrensten Oracle Platinum Partner – ausgezeichnet mit dem EMEA Oracle Excellence Award 2016.

PROMATIS



PROMATIS Gruppe
Tel. +49 7243 2179-0
www.promatis.de
Ettlingen/Baden · Hamburg · Berlin
Wien (A) · Zürich (CH) · Denver (USA)