Oracle Datenbank-Serverkonsolidierung mit Linux on System z – ein Erfahrungsbericht

Siegfried Langer
IBM Deutschland Research & Development GmbH
Böblingen

Schlüsselworte

Oracle DB, Konsolidierung, Virtualisierung, Optimierung, Performance Tuning, Linux, IBM Linux on System z, Mainframe, z/VM Hypervisor, Best Practices.

Einleitung

Zahlreiche Kunden nutzen die Vorteile der Konsolidierung einzelner Oracle-Datenbankserver. Dies beinhaltet Einsparungen bei Software-Lizenzkosten und operationalen Kosten, wie Servicepersonal, Netzwerk, Strom, Kühlung und Stellfläche. Die hohe Integration und Virtualisierung ermöglicht ein zentralisiertes Management mit reduziertem Risiko und geringerem Verwaltungsaufwand.

Zentralisierte Datensicherung, Hochverfügbarkeit und Maßnahmen für den Katastrophenfall sind einfacher zu implementieren. Cloudkonzepte mit hoher Flexibilität und schneller Aktivierung neuer (virtueller) Server können wirkungsvoll umgesetzt werden.

Allerdings stellt eine hoch-virtualisierte Systemumgebung gewisse Anforderungen an das Systemmanagement, um optimale Leistung zu erhalten und in einem dynamischen Umfeld auf Dauer sicherzustellen. Zahlreiche Aspekte sind von genereller Natur und gelten allgemein für Servervirtualisierung. Darüber hinaus bieten Linux on System z zusammen mit dem Hypervisor z/VM spezielle Funktionalitäten für die Konsolidierung zahlreicher Oracle-Datenbankserver.

Basierend auf praktischen Erfahrungen werden grundsätzliche Überlegungen zu hoch-virtualisierten Umgebungen, "Best Practices", Fallstricke und Beispiele mit Linux on System z und z/VM behandelt.

In einer konkreten Kundensituation werden Migrations- und Tuningmaßnahmen und deren Auswirkung auf den Durchsatz untersucht. Dies umfasst sowohl technische Aspekte, als auch planerische Überlegungen.

Virtualisierung – die Voraussetzung für eine optimierte und konsolidierte Systemumgebung

Einzelne physische Server müssen für Lastspitzen ausgelegt werden. Die erforderliche Leistung muss die erwartete Spitzenauslastung und zukünftige Wachstumsreserven berücksichtigen, auch wenn diese nur kurzzeitig auftreten. Aufgrund der relativ geringen Hardwarekosten und der Verfügbarkeit von Multi-Core-Servern stellt dies meist kein unmittelbares Problem dar.

Wenn man aber die Software-Lizenzkosten, die sich typischerweise an der Anzahl der Prozessorkerne (Cores) bemessen, einbezieht, dann ergibt sich ein anderes Bild. Die Konsolidierung vieler, teils nur wenig ausgelasteter Prozessoren auf einen hoch-virtualisierten Server, führt zu einer wesentlich besseren Nutzung der Ressourcen und ermöglicht erhebliche Einsparungen bei den Software-Lizenzkosten. Darüber hinaus ergeben sich teils erhebliche Einsparungspotentiale bei den operativen Kosten (Strom, Kühlung, Stellfläche, Netzwerk, Servicepersonal). Ein zentralisiertes Management reduziert den Verwaltungsaufwand und erlaubt zentralisierte Datensicherung, bessere Vorsorge für den Katastrophenfall, Hochverfügbarkeit und die Nutzung von Cloudkonzepten mit hoher Flexibilität und schneller Aktivierung neuer Server.

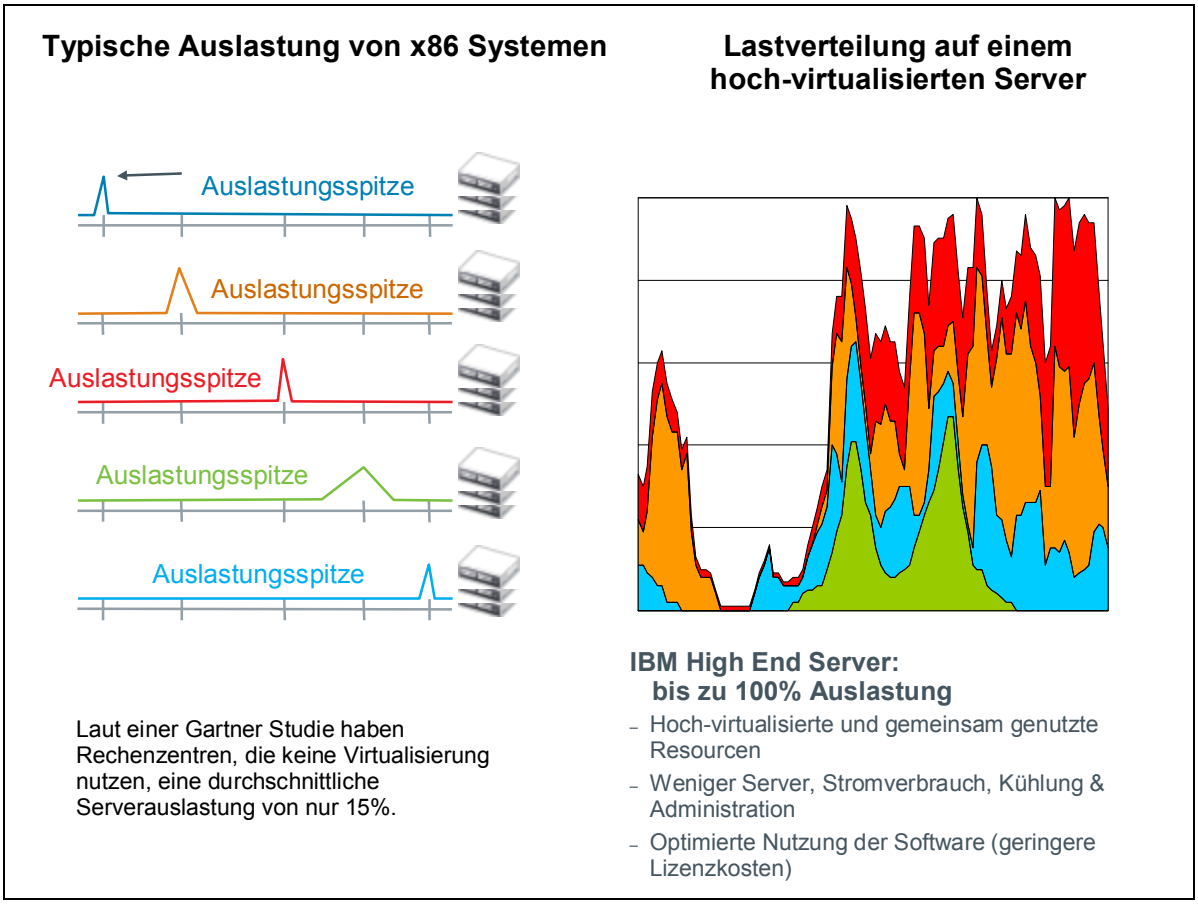

Abb. 1: Das Grundprinzip - Konsolidierung physischer Server in eine hoch-virtualisierte Umgebung

Anforderungen an eine hoch-virtualisierte Umgebung

Die Virtualisierung zahlreicher Oracle Datenbankserver stellt allerdings hohe Anforderungen an die Zielumgebung und damit an den Hypervisor, der diese Virtualisierung bereitstellen muss: effektive und sichere Verwaltung der gemeinsam genutzten (shared) Ressourcen, wie Prozessoren, interne und externe Speichermedien, Netzwerkverbindungen, und die wesentlich höhere Auslastung des Systems, das auch bei 100% noch Stabilität und Durchsatz bieten muss.

Ein System, das diesen Herausforderungen hervorragend gewachsen ist, ist der Mainframe oder besser dessen heutige Version der zEnterprise mit dem Hypervisor z/VM und Linux on System z.

Am Beispiel realer Kundensituationen wird die Architektur für die Konsolidierung zahlreicher Oracle Datenbankserver auf zEnterprise dargestellt und es werden spezifische Tuningmaßnahmen zur Optimierung exemplarisch dargestellt.

Das Projekt

Das Gesamtprojekt umfasst die Konsolidierung von annähernd 200 Oracle Datenbanken (ca. 150 x86 Blade-Server) auf IBM zEnterprise z196 unter Linux (RHEL 5). Der größte Anteil sind Oracle 10*g* Datenbanken, sowie einige 11*g* DBs. Die Größe der einzelnen Datenbanken variiert von wenigen Gigabytes bis zu mehreren Terabytes und umfasst unterschiedliche Anwendungsbereiche. Abbildung 2 vermittelt einen Überblick über die Zielkonfiguration. Die beschriebenen Tuningmaßnahmen betreffen die erste Phase des Projektes mit etwa 50 Datenbanken im produktiven Betrieb.

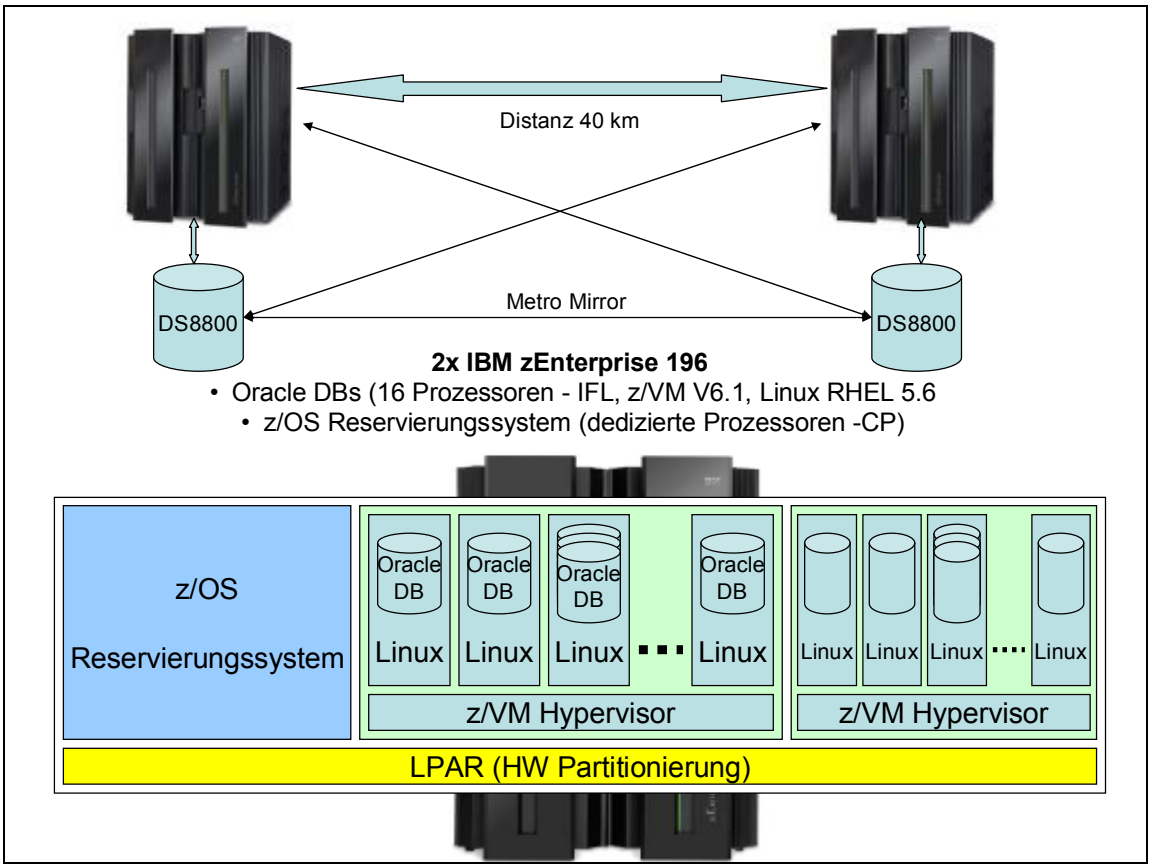

Abb. 2: Die Zielkonfiguration

Ressourcenverwaltung in virtualisierten Systemen

Bei einzelnen physischen Servern werden alle Ressourcen des Systems vom Betriebssystem dieses Servers verwaltet. Alle Prozessoren, der Speicher, die Anschlüsse für externe Verbindungen und Netzwerke sind exklusiv. Linux, zum Beispiel, verwendet eine aggressive Caching-Strategie: freier Speicher wird als Puffer verwendet. Was in einem physischen Server sinnvoll ist, kann in einer hoch-virtualisierten Umgebung den Durchsatz negativ beeinflussen. Der Hypervisor muss konkurrierende Ressourcenanforderungen verwalten. Zu groß dimensionierte Gäste führen zu mehr Aufwand oder machen es schwer, eine faire Zuteilung zwischen den anfordernden Gästen zu ermöglichen. Am Beispiel der Speicherverwaltung soll dies im Folgenden näher erläutert werden.

Der Hypervisor z/VM erlaubt es, den einzelnen Gastsystemen mehr Speicher, als physisch vorhanden, zuzuweisen (memory overcommitment). Dies ist eine sehr praktische Funktion bei vielen Gästen, die nur geringe oder seltene Anforderungen stellen (z.B. Test- oder Entwicklungssysteme), da dieser Speicher den aktiven Prozessen zur Verfügung gestellt werden kann.

Überdimensionierte Gäste binden wertvolle Ressourcen, die der Hypervisor im Gesamtsystem mühsam suchen müsste. Außerdem ist die Zeit, die für das Verschieben von gigabytegroßen Speicherinhalten notwendig ist, unter Durchsatzgesichtspunkten nicht zu vernachlässigen. Für Linux on System z Gäste gilt daher, dass diese Linux-Gäste nur so groß dimensioniert werden sollten, wie für eine gute Funktionalität notwendig. Zusätzliches Caching im Linuxgast, insbesondere I/O Caching mit laufenden Updates bindet wertvollen Speicher, der anderen Prozessen nicht zur Verfügung steht. Da der Hypervisor in erster Linie nach dem „least recently used“ Algorithmus vorgeht, wird dieser Cachespeicher nicht angefasst. Die empfohlene Vorgehensweise ist hier, direct I/O zu verwenden und die Linuxgäste speicherseitig so zu dimensionieren, dass das individuelle Linux im Normalbetrieb

gerade noch nicht auf den externen Speicher auslagert (swapped). Dadurch wird sichergestellt, dass z/VM das Gesamtsystem optimal mit Ressourcen versorgen kann.

Ähnliche Überlegungen gelten auch für die Zuordnung physischer und virtueller Prozessoren (CPU). Auch hier führt eine Überdimensionierung zu mehr Verwaltungsaufwand und damit Overhead im Hypervisor. Zusätzlich besteht das Risiko, dass sehr CPU-aktive Prozesse das Gesamtsystem dominieren und andere Gästen nur noch unzureichend Service geben können, da sie nicht mehr die benötigten Prozessor-Zeitscheiben zugewiesen bekommen.

Tuning Maßnahmen

In diesem Kundenprojekt wurden etwa 50 Oracle Datenbankserver unter Linux von einer x86 Blade-Umgebung mit unterschiedlichen Konfigurationen nach Linux on System z und z/VM migriert. Mittels einer speziellen Migrationssoftware wurde der Datenbankinhalt kopiert, was wesentlich kürzere Zeiten gegenüber der Export/Import-Funktion ermöglicht.

Bei der Dimensionierung der Zielumgebung wurde die Anzahl der virtuellen CPUs pro Server anfänglich mit maximal 4 vCPUs festgelegt[1]. z/VM ermöglicht es, diese Zahl bei Bedarf dynamisch zu erhöhen. Die anfängliche Speichergröße wurde nach folgender Formel angenähert:

```
Linux-Speichergröße = SGA + PGA + 512MB für Linuxkernel
                      (+ 256MB bis 512MB für ASM, wenn benutzt)
```

System z Linux unterstützt „direct I/O“ und „asynchronous I/O“ für Plattenspeicher. Für 10gR2 und höher ist die Empfehlung den Parameter `filesystemio_options` auf `setall` zu setzen, da Oracle sonst beide Methoden benutzt. Für Systeme mit virtualisierten logischen Datenbank-Files vermeidet dies das Zwischenspeichern von I/O Daten im Datenbank-Puffer von Oracle und im Filesystem-Cache von Linux und nutzt direct I/O. Das Übergehen des Linux Page-Cache mittels direct I/O kann den Durchsatz bis zu Faktor zwei verbessern.

Nach einiger Zeit in Produktion beklagten einige Anwender lange Laufzeiten bei spezifischen Auswerteprogrammen. Nachdem die Gesamtspeicherzuordnung für den z/VM Hypervisor erhöht worden war und sichergestellt wurde, dass die Bandbreite zum Plattenspeichersystem DS8800 keinen Engpass darstellte, führten folgende Einstellungen zu teils signifikanten Verbesserungen.

- Multipath setup „round robin“ wurde auf `rr_min_io=1` gesetzt. Der Linux default Wert ist 1000. Dadurch wird hohe Parallelität der Plattenzugriffe sichergestellt.
- Um den Einfluss des Loggens auf den aktiven Datendurchsatz zu minimieren, wurde der Log-Buffer stark vergrößert: `log_buffers=104,857,600`
- Anhand einer spezifischen Businessanalyse-Anwendung wurde der Einfluss von Oracle Optimizerhints bei den lang laufenden SQL Queries untersucht. Die Optimizerhints „FULL“ (forciert Table scans statt Index access) und „PARALLEL“ (forciert die Parallelisierung der Query) führten in diesem konkreten Beispiel zu einer enormen Beschleunigung:
    - FULL: Beschleunigung um Faktor 4
    - PARALLEL: Beschleunigung um Faktor 8
    - FULL und PARALLEL: Beschleunigung um Faktor 12

[1] Die einzelnen Prozessoren (IFL = Integrated Facility for Linux) der zEnterprise 196 verfügen über eine gegenüber x86 höheren Leistung und sind mit 5,2GHz getaktet.

Die Verwendung von Optimizerhints bei spezifischen SQL-Befehlen, statt globaler Parameter, ermöglicht es, spezifischen Einfluss auf Queries mit „large table scans“ zu nehmen ohne das Performanceverhalten der (gleichzeitigen) interaktiven Abfragen zu verändern.

Zusammenfassung

Die Konsolidierung einzelner Oracle DB Server kann zu erheblichen Kosteneinsparungen führen. Neben Einsparungen bei den Lizenzkosten besteht ein hohes Potential bei den operativen Kosten.

Die Migration einzelner individueller Server in eine hoch-virtualisierte Serverumgebung bedingt einige Anpassungen, um optimalen Durchsatz und maximalen Nutzen zu erzielen. Die Konsolidierung zahlreicher Oracle DB Server auf Linux on System z mit dem z/VM Hypervisor führt zu einer erheblichen Reduzierung der Anzahl der benötigten Prozessorkerne (Cores), insbesondere wenn wenig oder nicht durchgehend voll belastete Server, wie auch Test- und Entwicklungssysteme virtualisiert werden. Aufgrund der besonderen System z Architektur mit ihren Stärken im Input/Output-Durchsatzverhalten können bestimmte Optimierungsmaßnahmen zu stark unterschiedlichen Ergebnissen im Vergleich zu x86-basierenden Servern führen.

Transaktionsorientierte Datenbankzugriffe (über Index) und Businessanalyse-Anwendungen (typischerweise mit large table scans) stellen unterschiedliche Anforderungen an die Server. Praktische Erfahrungen in einer Produktionsumgebung haben gezeigt, dass es durch geeignetes Tuning möglich ist, diesen Doppelanforderungen zu genügen. Wenn die Anwendungen, die die Abfragen der Datenbank steuern, einbezogen werden, kann weiteres Optimierungspotential genutzt werden. Dies muss nicht zwingend einen Eingriff in die Businesslogik bedingen.

Kontaktadresse:

Siegfried Langer

IBM Deutschland Research & Development GmbH

Schönaicher Strasse 220

D-71032 Böblingen

Telefon: +49 (0) 7031-16 4228

Fax: +49 (0) 7031-16 3456

E-Mail: Siegfried.Langer@de.ibm.com

Internet: www.ibm.com

Weitere Informationen:

IBM Redbooks:

Experiences with Oracle Solutions on Linux for IBM System z

http://www.redbooks.ibm.com/abstracts/sg247634.html?Open

Installing Oracle 11gR2 RAC on Linux on System z

http://www.redbooks.ibm.com/abstracts/redp4788.html?Open