Oracle auf HP/Violin – Wirklich ein Exadata Killer?

Manfred Drozd
Benchware AG
CH-8800 Thalwil

Schlüsselworte

Exadata, HP/Violin, Data Warehouse, OLTP, Oracle 11, Flash Memory.

Einleitung

Violin Memory (http://www.violin-memory.com/) hat sich als einer der führenden Flash Memory Hersteller etabliert. In einer Partnerschaft mit HP werden Proliant Server mit dieser neuen Technologie kombiniert und für Oracle Plattformen mit höchsten Leistungsanforderungen angeboten. In den Medien wurde dieses System verschiedentlich als Exadata Killer tituliert. Wir haben ein solches System einem systematischen Benchmark unterzogen und neben den Performanceeigenschaften auch weitere wichtige Unterschiede zur Exadata untersucht.

Unser Benchmark Verfahren für Oracle Plattformen

Unser Benchmark Verfahren erlaubt eine schnelle, repräsentative und herstellerneutrale Bewertung des Preis-/Leistungsverhaltens von Oracle Plattformen unter Berücksichtigung der Oracle Lizenz- und Supportkosten. Das Verfahren wird sowohl bei der Evaluation einzelner Komponenten oder kompletter Systeme als auch bei Health Checks zur Identifikation von Performance Engpässen eingesetzt. Die detaillierten Benchmark Reports zur Oracle Exadata und zur HP/Violin Plattform findet man unter http://www.benchware.ch/benchmarks.

Beim Benchmark wird die Performance der einzelnen Komponenten systematisch analysiert und bewertet. Dabei wird die Oracle Datenbank als Load Generator genutzt, um repräsentative Ergebnisse zu erzielen:

- Die CPU Leistung hat nicht nur für einzelne Datenbank Operationen eine enorme Bedeutung, sondern auch für die Oracle Lizenzierung, selbst bei einem ULA.
- Die Server Leistung spielt vor allem bei *in-memory* SQL Operationen, wie sie im OLTP Umfeld und bei der real-time Datenanalyse vorkommen, eine bedeutende Rolle. Einige Software Hersteller setzen mittlerweile komplett auf *in-memory* Datenbanktechnik.
- Die verschiedenen Storage Zugriffsprofile wie *random read/write* und *sequential read/write* werden vermessen. Aber nicht nur der Durchsatz, sondern auch das Servicezeit Verhalten ist von hoher Bedeutung.
- Zuletzt werden die Durchsätze bei typischen Datenbank Operationen wie Laden, Scannen, Aggregieren und OLTP Transaktionen vermessen.

Wir verwenden verständliche Key Performance Metriken zur Leistungsbeschreibung, die direkt in den Projekten für eine Kapazitätsplanung angewandt werden können.

Engineered Systems

Die Exadata ist tatsächlich ein komplett vorkonfiguriertes System, dass innerhalb weniger Stunden beim Kunden funktionstüchtig ist und vom ersten Tag maximale Performance liefert. Wir kennen bislang keinen Mitbewerber, der eine solche Qualität in die Standardisierung von Oracle Plattformen eingebracht hat. Alle Mitbewerber arbeiten überwiegend mit Referenz Architekturen, die aber mehr oder weniger einer manuellen Konfiguration mit allen Nachteilen (Aufwand, Abweichungen vom Standard, Fehleranfälligkeit) ähneln. Auch für die HP/Violin Konfiguration gibt es einige nicht öffentlich zugängliche Dokumente, die einen optimalen Setup vom Betriebssystem (I/O Scheduler, Queueing, verschiedene Kernel Parameter), dem *block alignment* von LUNs und ASM (Parameter für die Größe von Disk Sektoren und der allocation unit) beschreiben. Viele diese Informationen können aber auch auf der Webseite eines Violin (und Ex-Oracle) Mitarbeiters gefunden werden (http://www.flashdba.com/). Der Support für ein echtes *engineered system* ist für den Hersteller einfacher, da er die Kundenkonfiguration exakt kennt. In diesem Punkt sehen wir einen Vorteil der Exadata.

Vendor Lock-in

Beide Systeme arbeiten mit Standard Komponenten. Violin Memory bietet ein einfaches, aber komfortables und verständliches GUI für die Administration und Konfiguration seiner Storage Systeme an. Ansonsten ergeben sich aus betrieblicher Sicht keine Unterschiede zu anderen Storage Systemen. Oracle bietet mit HCC (*hybrid columnar compression*) eine einzige optionale Funktion an, die nur auf der Exadata zur Verfügung steht und Code Änderungen bei DDL Statements zur Folge hat. Wir sehen darin kein *vendor lock-in*, da eine Exadata Datenbank jederzeit schnell auf eine andere Plattform migriert werden kann. Der Aufwand entspricht der einer Standard Datenbank Plattform Migration. Beide Systeme bedeuten kein *vendor lock-in*.

CPU- und Server

Beide Hersteller verwenden schnelle x86 Prozessoren, sodass ein Höchstmaß an Leistung pro Prozessor Core (und damit Oracle Lizenzgebühr) gegeben ist. Die Exadata X2 bietet nur 2 Servertypen an: die X2-2 mit 2 *sockets* (je 6 *cores*) und 96 GByte RAM und die X2-8 mit 8 *sockets* (je 10 *cores*) und 2 TByte RAM. Die Hauptspeicher Kapazität des X2-2 Servers kann auf 144 GByte RAM ausgebaut werden. Bei HP kann aus einem breiten Spektrum von Proliant Servern ausgewählt werden: Server mit 2, 4 oder 8 *sockets* und pro *socket* 4, 6, 8 oder 10 *cores*. Hauptspeicherkapazitäten zwischen 16 GByte und 4 TByte. Diese feineren Granulate sind für eine Optimierung der Oracle Lizenz- und Supportkosten von großem Vorteil, da die Oracle Lizenzkosten die Hardwarekosten häufig um Faktoren übertreffen. Hier muss der Kunde zwischen (extremer) Standardisierung und (extremer) Flexibilität entscheiden. Wir sehen hier einen Vorteil der HP/Violin Lösung, da die Ausbaustufen der Exadata doch in recht groben Granulaten erfolgt, die mit hohen Investitionen (Hardware und Software) verbunden sind.

Hochverfügbarkeit

Die Exadata wird standardmäßig mit einem Real Application Cluster (RAC) ausgeliefert und bietet damit eine *built in* Hochverfügbarkeit an. Bei der HP/Violin Lösung muss der Cluster selbst implementiert werden. Auch wenn der RAC Installationsprozess in den letzten Jahren stark verbessert und vereinfacht wurde, wird doch immer noch spezielles *know how* verlangt. Nicht umsonst gibt es knapp 800 Seiten dicke Bücher zur RAC Installation (z.B. Dyke, Shaw, Bach: *Pro Oracle Database 11g Rac on Linux*). Bei diesem Kriterium sehen wir einen Vorteil bei der Exadata.

Disaster Recovery

Für die Exadata wird *Data Guard* oder auch *Golden Gate* als *Disaster Recovery* Lösung angeboten. Das gleiche gilt für die HP/Violin Lösung. Violin bietet keinerlei Funktionalität für die Spiegelung von Storage Systemen zwischen zwei Standorten. Als zusätzliche Möglichkeit bietet die HP/Violin Lösung noch ein *host based mirroring* via ASM an. Beide Systeme bieten damit eine ähnliche Lösung bezüglich *Disaster Recovery*. Verglichen mit der Funktionalität klassischer Storage Systeme der Hersteller HDS oder EMC, bietet sowohl die Exadata als auch die HP/Violin ein eher bescheidenes Niveau an Funktionalität.

Skalierbarkeit der Applikation: Shared-Memory versus Shared-Disk

Die beiden Plattformen unterstützen unterschiedliche Ansätze für die Skalierung von Applikationen. Die Exadata setzt bei den X2-2 Systemen auf eine *shared-disk* Architektur mit dem Oracle *Real Application Cluster*. Für sehr große Systeme bietet Oracle die Exadata X2-8 mit 2 großen *shared-memory* Servern an. Bei HP/Violin setzt man bezüglich Skalierbarkeit auf die *shared-memory* Architektur und optional auf die *shared-disk* Architektur. Die Anwendungsprogrammierer bevorzugen den *shared-memory* Ansatz zur Skalierbarkeit von Applikationen, da sie nur auf eine hohe Parallelisierung ihrer Anwendung achten müssen. Die *shared-disk* Architektur ist in diesem Punkt etwas komplizierter: neben einer hohen Parallelisierung ist es notwendig, dass diese Parallelisierung auch absolut konfliktfrei ist. Konflikte in einer *shared-disk* Architektur sind um Zehnerpotenzen teurer als in einer *shared-memory* Architektur. Vorteil HP/Violin.

Wartung

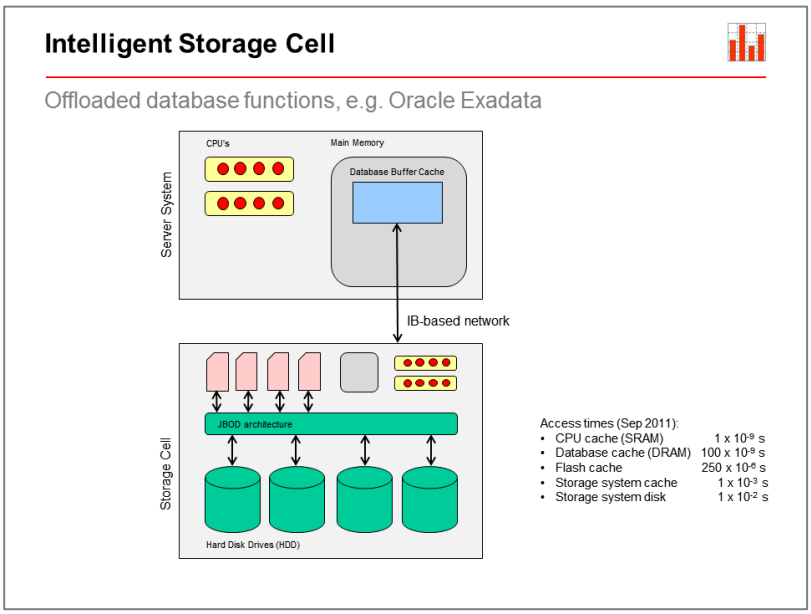
Für die Exadata wird ein *patch* für das gesamte System geliefert. Oracle hat ja mit der Übernahme von Sun Microsystems Kontrolle über sämtliche Software Komponenten inklusive Treiber. Bei der HP/Violin Lösung wird pro Komponente ein *patch* eingespielt, die zeitlich nicht abgestimmt eintreffen. Dies bedeutet für die HP/Violin Lösung mehr Wartungsfenster und mehr Wartungsarbeit. Vorteil Exadata.

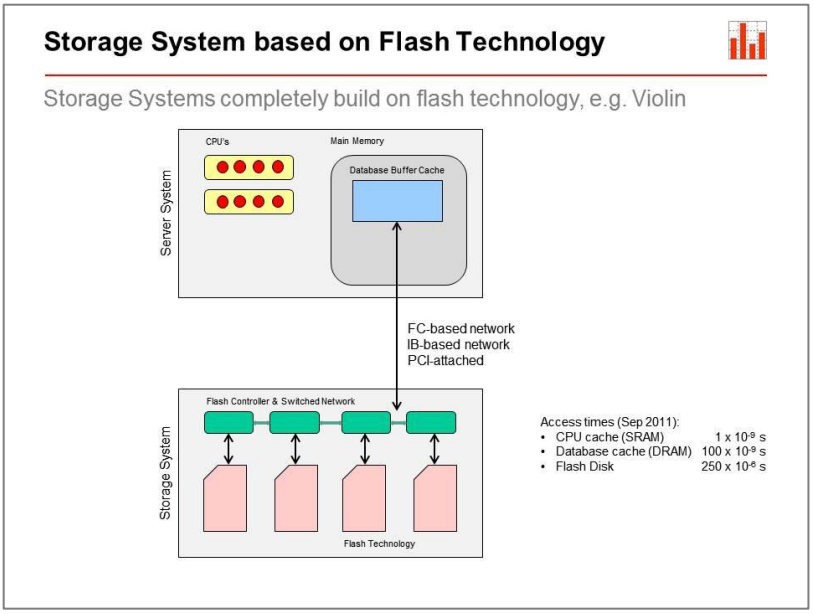
Architektur für den Einsatz von Flash Technologie

Wir sehen heute bei den Herstellern unterschiedliche Architekturen, um Flash Technologien in Oracle Plattformen zu integrieren. Dabei sind verschiedene Aspekte zu beachten, wie Volatilität der Daten, Manageability, Kapazität, Performance (Durchsatz und Latenzzeit), Kosten für Kapazität und Performance, und die Fähigkeit von mehreren Servern auf den Flash Storage zuzugreifen.

Die Exadata verfügt über mehrere intelligente Storage Cells. Diese Storage Cells können bestimmte Aufgaben des Datenbank Servers selbständig übernehmen (*offloading*). Die *offloading* Funktionen sind extrem effizient und bieten enorme Performancevorteile gegenüber anderen Lösungen. Leider können sie nur unter bestimmten Voraussetzungen genutzt werden. Die Storage Cells umfassen jeweils 12 *hard disk drives* in einer JBOD Konfiguration mit Flash Karten als *write through cache*. Dieser *cell flash cache* wird entweder von Oracle automatisch verwaltet oder Datenbankobjekte werden durch den DBA gezielt in diesem *cell flash cache* abgelegt. Lesende Operationen aus dem *cell flash cache* liefern herausragende Performancezahlen. Der schwerwiegendste Nachteil der Exadata gegenüber der HP/Violin Lösung liegt im *write through cache*, der die Anzahl *random writes* begrenzt. Pro *storage cell* können ca. 5'500 *random writes* (*capacity optimized storage cell*) oder 10'800 *random writes* (*performance optimized storage cell*) verarbeitet werden. Die Exadata kann aber so konfiguriert werden, dass Teile oder der gesamte *cell flash cache* als *grid disks* benutzt werden

können. Die heute von Oracle verwendete Flash Technologie hat aber noch nicht das *write cliff* Problem gelöst. Dies ist der Grund dafür, dass Oracle REDO Logfiles sowohl auf *grid disks* als auch auf *hard disk drives* ablegt.





Bei der HP/Violin Lösung werden alle Tier-1 Daten ausnahmslos auf Flash Technology abgelegt. Jeder Datenzugriff arbeitet daher immer mit optimaler Performance. Die Latenzzeit für lesende und schreibende Operationen liegt im Bereich weniger Millisekunden. Der erreichte Durchsatz hängt von der Anzahl Memory Arrays ab. Pro VMA 3205 werden mit 8 KByte Datenbankblöcken ca. 95'000 IOPS lesend und 70'000 IOPS schreibend erreicht. Weitere Storage Tiers mit konventioneller Disk Technologie können einfach hinzugefügt werden, wobei die Zuordnung von Daten zu Storage Tiers manuell erfolgen muss.

Fazit: wenn die Exadata die Intelligenz ihrer *storage cells* nutzen kann, ist sie jedem Mitbewerber deutlich überlegen. Wenn eine Applikation dagegen ein hohes Mutationsvolumen bewältigen muss, bei dem hohe *random write* Raten benötigt werden, ist die HP/Violin Lösung überlegen.

Komprimierungstechnologie

Die Exadata bietet mit HCC eine außergewöhnlich effiziente Komprimierungstechnologie an. HCC steht aber nur auf der Exadata zur Verfügung. Auf allen anderen Plattformen können nur die konventionellen Oracle Komprimierungsverfahren (zeilenorientiert) verwendet werden. Für die Komprimierungstechnologie liegen nun konkrete Kundenerfahrungen vor. Im besten Fall konnte für eine denormalisierte Tabelle mit einigen hundert Spalten eine Komprimierung von Faktor 60 (!) erreicht werden. Vorteil Exadata.

Support verschiedener Oracle Versionen

Auf der Exadata ist offiziell nur Oracle 11.2 unterstützt. Eine HP/Violin Lösung unterstützt prinzipiell auch ältere Linux und Oracle Versionen und bietet daher etwas mehr Flexibilität im Life Cycle von Betriebssystem- und Oracle Versionen. Vorteil HP/Violin.

Zusammenfassung

Beide Systeme bieten eine innovative Architektur mit herausragenden Performance Eigenschaften. Es hängt sehr stark von der konkreten Kundensituation und seinen Bedürfnissen ab, welches System besser geeignet ist. Tendenziell kann die Exadata ihre Stärken besser im Data Warehouse Umfeld ausspielen, während die HP/Violin Lösung im OLTP Umfeld mit garantierten I/O Servicezeiten und hohem I/O Durchsatz glänzt.

In unserem Vortrag möchten wir neben den hier genannten Aspekten auf die konkreten Performancezahlen und das Preis-/Leistungsverhältnis eingehen.

Kontaktadresse:

Manfred Drozd
Benchware AG
Seestrasse 18
CH-8800 Thalwil

Telefon: +41 44 722 16 16
Fax: +41 44 722 16 18
E-Mail manfred.drozd@benchware.ch
Internet: www.benchware.ch