Oracle Hochleistungs-Plattformen

Manfred Drozd
Benchware AG
CH-8800 Thalwil

Schlüsselworte

Performance, Benchmarking, Flash Storage, In-Memory SQL Verarbeitung.

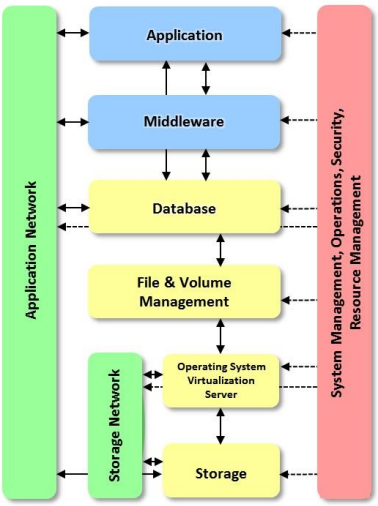
Einleitung

Neue Technologien wie Flash Speicher und große Hauptspeicherkapazitäten für In-Memory Verarbeitung werden zunehmend in Oracle Plattformen eingesetzt und bieten einen enormen Performanceschub für Oracle Applikationen.

Die In-Memory SQL Verarbeitung erlaubt die schnellst mögliche Datenbankverarbeitung. Nicht nur Oracle Times Ten, sondern auch der konventionelle Oracle Datenbankserver unterstützt mit verschiedenen Funktionen die In-Memory Verarbeitung und bietet eine herausragende Performance für alle Art von Applikationen. Unsere Benchmark Ergebnisse mit Servern verschiedener Hersteller und unterschiedlichen Prozessoren (Intel Xeon, IBM Power und Sun SPARC) werden vorgestellt.

Bei der Integration von Flash Storage werden unterschiedliche Architekturen angeboten: PCI attached Flash Storage im Datenbank Server oder extern als Flash Storage in herkömmlichen Storage Systemen. Die verschiedenen Lösungen werden qualitativ und quantitativ verglichen. In einem Benchmark vergleichen wir zwei Architekturen des Herstellers Hitachi Data Systems: PCI attached Flash Storage in einem HDS UCP Server mit FusionIO Modulen und der gleiche Server mit einer FC attached HDS HUS-VM, die komplett und ausschliesslich mit Flash Storage konfiguriert ist.

Komplexität von Oracle Plattformen



Oracle Plattformen setzen sich aus unterschiedlichen Layern zusammen. Der Systemarchitekt muss unter unzähligen Produkten verschiedener Hersteller, verschiedenen Technologien, Optionen und Konfigurationsparametern für seine Oracle Plattform die geeignete Lösung auswählen.

Wegen der hohen Komplexität solcher Plattformen kann die Performance nicht vorweg bestimmt werden und muss durch einen Benchmark überprüft werden. Am besten als qualitätssichernde Massnahme vor Inbetriebnahme einer neuen Plattform.

Die Komplexität von Oracle Plattformen ist der wesentliche Grund, warum Engineered Systems mittlerweile von allen namhaften Herstellern angeboten werden – allerdings in sehr unterschiedlichen Reifegraden.

Abbild 1: Komplexität von Oracle Plattformen

Benchmarks von Oracle Plattformen

Wir verwenden systematische Benchmarks von Oracle Plattformen, um folgende Fragestellungen zu beantworten:

- Welche Leistung liefert meine heutige Plattform und welche Leistungsreserven sind noch vorhanden? (*capacity planning*)
- Welche Leistung bieten vergleichbare Lösungen anderer Hersteller? Welche Plattform bietet das beste Preis-/Leistungsverhältnis? (*price performance ratio*)
- Entspricht die Leistungsfähigkeit einer self-engineered Plattform den Performance Anforderungen? (*health check*)
- Welchen Einfluss haben Änderungen der Plattform, z.B. ein Austausch des Storage Systems, auf die Leistungsfähigkeit der Plattform? (*quality control*)

Die beiden oberen Layer (Application, Middleware) der Plattform (siehe Abbild 1) bestimmen die Performance Anforderungen (*performance requirements*). Die unteren Layer (Database, File & Volume Management, Server und Storage) müssen die geforderte Performance erbringen (*performance delivery*).

Genau an dieser Schnittstelle setzen wir mit unseren Benchmark Tests an, um zu überprüfen, ob die Performance Anforderungen der Applikationen erfüllt werden können. In über 40 Einzeltests werden die wichtigsten Performance Kennzahlen (*key performance metrics*) für alle repräsentativen Datenbankoperationen vermessen.

Die einzelnen Komponenten wie Prozessor, Server und Storage können separat vermessen werden, wobei immer das Performanceverhalten der Komponente im Oracle Betrieb im Vordergrund steht. CPU Tests werden mit Oracle Datentypen durchgeführt, Storage Tests werden mit Oracle I/O Operationen und Oracle Tabellen durchgeführt und Server Tests mit Oracle In-Memory SQL Verarbeitung auf im Buffer Cache befindlichen Tabellen.

Dieser Ansatz hat den Vorteil, dass wir das exakte Leistungsverhalten einer Plattform im Oracle Betrieb kennen. Die vermessenen Performance Kennzahlen können in Form eines SLA den Applikationen, die auf Oracle aufsetzen, garantiert zur Verfügung gestellt werden.

Der gesamte Benchmark einer Oracle Plattform dauert nur einige Tage, um schnell und zuverlässig aussagekräftige Performance Kennzahlen zu ermitteln. Für jeden Test werden AWR oder STATSPACK Reports erstellt, um detailliert das Oracle Verhalten in Stresssituationen analysieren zu können und Hinweise für weitere Performance Verbesserungen zu erhalten.

In-Memory SQL Verarbeitung

Die In-Memory SQL Verarbeitung ist die schnellst mögliche Form der SQL Verarbeitung. Dabei sind alle Daten im Oracle Buffer Cache. Bei lesenden Transaktionen finden kaum (*block cleanout*) oder keine I/O Operationen statt.

Oracle bietet verschiedene Technologien zur Unterstützung der In-Memory Verarbeitung. So können Daten ganz gezielt in den KEEP Cache gestellt werden (bei Verwendung des Betriebssystems Linux maximal 32 GByte), unabhängig von der Zugriffshäufigkeit. Auch die parallele SQL Verarbeitung funktioniert auf Objekten, die komplett im Buffer Cache liegen. Das Leistungsverhalten von Applikationen kann sich bei intensiver Nutzung der In-Memory Technologie verändern.

Die Servicezeit von Transaktionen sinkt um Faktoren und die CPU Auslastung schnellt hoch. Die In-Memory Verarbeitung ist somit die einfachste Form der Performance Optimierung, wenn Applikationssoftware nicht geändert werden kann.

Die Kosten für Hauptspeicher sind in den letzten Jahren stetig gesunken. Für x86 Server kostet 1 TByte Hauptspeicher ca. 25'000 USD (Listenpreis bei Verwendung von 16 GByte DIMM). Selbst bei RISC Servern betragen die Kosten weniger als 55'000 USD (Listenpreis 16 GByte DIMM).

Die In-Memory Verarbeitung ist insbesondere auch bei Servern mit 2 Sockets interessant, da solche Server mit der deutlich günstigeren Oracle Standard Edition betrieben werden können.

Unsere Benchmark Tests für Server Systeme unterscheiden drei verschiedene In-Memory SQL Transaktionsprofile:

- *full table scan* – so wie er häufig in DWH Anwendungen vorkommt, um grosse Datenmengen auf bestimmte Kriterien zu durchsuchen.
- *primary key access* mit einem Treffer pro Transaktion – OLTP lookup Transaktion, um z.B. eine Kunden-, Produkt- oder Kontonummer zu suchen.
- *secondary key access* mit ca. 25 Treffern pro Transaktion – OLTP Transaktion, z.B. um die letzten 25 Kontobuchungen beim e-banking anzuschauen.

Im Vortrag werden die Benchmark Ergebnisse verschiedener Server mit unterschiedlichen Prozessoren vorgestellt (Intel x86, IBM Power 6 und 7, Sun Sparc T5).

Hier die Ergebnisse mit einem Server, der über zwei Intel Xeon E5-2690 Prozessoren (total 16 Cores) mit einer Taktrate von 2.9 GHz verwendet. Die Spalten der Tabelle haben folgende Bedeutung:

- #J Anzahl Prozesse (jobs)
- #T Anzahl Threads bei Oracle interner SQL Parallelisierung
- lread logical reads im buffer cache
- pread physicacl reads im buffer cache

Zunächst die Ergebnisse für einen In-Memory Full Table Scan:

#J	#T	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	Throughput rows/sec [rps]	Throughput txn/sec [tps]	SQL service time [s]	Buffer lread [bps]	Buffer pread [bps]	Elap time [s]
1	1	4	3	1	96	2.953E+06	2.400E+01	4.240E-02	1.285E+05	0.000E+00	127
2	1	7	6	1	93	6.303E+06	5.000E+01	3.848E-02	2.741E+05	0.000E+00	119
4	1	13	12	1	87	1.282E+07	1.030E+02	3.815E-02	5.574E+05	0.000E+00	117
8	1	25	24	0	75	2.521E+07	2.020E+02	3.878E-02	1.095E+06	0.000E+00	119
16	1	49	48	1	51	4.762E+07	3.810E+02	4.071E-02	2.065E+06	0.000E+00	126
32	1	94	94	0	6	6.885E+07	5.510E+02	5.615E-02	2.984E+06	0.000E+00	130

Tabelle 1: Performance Ergebnisse In-Memory Full Table Scan

Bei zunehmender Belastung des Systems durch eine wachsende Anzahl Prozesse (Spalte #J) steigt auch die CPU bis zu Sättigung (Spalte CPU busy).

Ein einzelner Prozess (#J=1) kann pro Sekunde über 2.9 Millionen Datensätze a 300 byte durchsuchen. 32 solcher Prozesse können bei voller Auslastung des Servers pro Sekunde knapp

70'000'000 Datensätze a 300 byte durchsuchen. Dies entspricht einer In-Memory Scan Rate von 20 Giga Byte pro Sekunde [GBps] Nutzdaten.

Hier die Ergebnisse für einen In-Memory Primary Key Access mit einer Trefferrate von einer Row pro SQL Statement bzw. Transaktion:

#J	#T	CPU busy [%]	CPU user [%]	CPU sys [%]	CPU idle [%]	Throughput rows/sec [rps]	Throughput txn/sec [tps]	SQL service time [s]	Buffer lread [bps]	Buffer pread [bps]	Elap time [s]
1	1	4	3	1	96	2.500E+04	2.500E+04	4.010E-05	7.503E+04	0.000E+00	120
2	1	7	6	1	93	4.959E+04	4.959E+04	4.006E-05	1.488E+05	0.000E+00	121
4	1	13	12	1	87	9.917E+04	9.917E+04	3.947E-05	2.975E+05	0.000E+00	121
8	1	25	23	2	75	2.000E+05	2.000E+05	3.903E-05	5.999E+05	0.000E+00	120
16	1	49	45	4	51	3.871E+05	3.871E+05	3.991E-05	1.161E+06	0.000E+00	124
32	1	95	88	8	5	5.120E+05	5.120E+05	6.113E-05	1.527E+06	0.000E+00	125

Tabelle 2: Performance Ergebnisse In-Memory Primary Key Access

Auch hier ist die Skalierbarkeit bis zur Auslastung der CPU Leistung gegeben (95% CPU Auslastung bei 32 Prozessen).

Ein einzelner Prozess (#J = 1) kann 25'000 solcher Lookp-Up Transaktionen pro Sekunde ausführen. Die Servicezeit dieser Transaktion liegt im Durchschnitt bei 40 µs. Bei voller Auslastung des Servers werden über 512'000 Transaktionen mit einer durchschnittlichen Servicezeit von 61 µs erreicht.

Für Look-Up Transaktionen hat Oracle Benchmark Ergebnisse für Times Ten publiziert. Times Ten schafft eine solche Transaktion auf dem gleichen Servertyp in 2 µs!

Flash Technologie in Oracle Plattformen

Hardware Hersteller bieten verschiedene Lösungen für die Integration von Flash Storage für Oracle Plattformen an.

PCI attached Flash Storage in einem Server bietet kürzest mögliche Servicezeiten für I/O Operationen, kann aber nicht für RAC Cluster eingesetzt werden. FC attached Flash Storage ermöglicht nicht ganz so gute I/O Servicezeiten, benötigt für einen hohe Durchsatz entsprechend viele Host Bus Adapter, bietet aber (wie gewohnt) die gesamte Palette der Storage System Funktionalität. Für die betrieblichen Prozesse ergeben sich keine Änderungen. Eine genaue Gegenüberstellung findet man unter http://www.benchware.ch/resources/

Unsere Benchmark Tests für Storage Systeme unterscheiden vier verschiedene Transaktionsprofile:

- *sequential read und write*
- *random read und write*

Beim random read Test mit einem Server, der über interne PCI attached Flash Storage Module verfügt, wird eine Höchstleistung von über 530'000 IOPS erreicht (Tabelle 3). Wohlgemerkt für 8 kByte Oracle Datenbank Blöcke, gemessen an der Schnittstelle zwischen Applikation und Datenbank.

#J	#T	CPU busy [%]	CPU sys [%]	Physical read [iops]	Physical read [bps]	Physical read [MBps]	Physical write [iops]	Physical write [bps]	Physical write [MBps]	REDO write [iops]	Hitrate db flash [%]	Hitrate exa flash [%]	Elap time [s]
1	1	3	2	32731	32729	256	9	9	0	2	0	0	51
2	1	6	4	63901	63899	499	47	56	0	5	0	0	52
4	1	12	7	121679	121676	951	141	146	1	10	0	0	55
8	1	23	13	230022	230020	1797	298	293	2	18	0	0	58
16	1	45	27	387952	387949	3031	529	510	4	31	0	0	69
32	1	78	45	510868	510866	3991	713	675	5	40	0	0	105
64	1	95	52	533028	533026	4164	741	697	6	41	0	0	202
128	1	98	51	510848	510846	3991	681	638	5	39	0	0	305

Tabelle 3: Performance Ergebnisse In-Memory Full Table Scan

Aus dem AWR Report erfahren wir, dass Oracle bei dieser hohen Belastung 2 verschiedene *system calls* verwendet:

- der wait event "db file sequential read" steht für *system calls*, die nur einen Block pro *system call* lesen
- der wait event " db file parallel read" steht für *system calls*, wo Oracle pro *system call* eine Reihe von *non-contiguous* Blöcken liest.

Top 5 Timed Foreground Events

Event	Waits	Time(s)	Avg wait (ms)	% DB time	Wait Class
DB CPU		4,451		20.9	
db file parallel read	2,686,643	2,033	1	9.6	User I/O
db file sequential read	3,079,002	1,343	0	6.3	User I/O

Tabelle 4: Oracle wait events bei > 500'000 IOPS

Die durchschnittliche Servicezeit pro single-block beträgt bei voller Last unter 500µs. Bei etwas geringerer Last (#J = 32) werden immer noch über 500'000 IOPS a 8 kByte erreicht, aber bei einer Servicezeit von weniger als 300 µs.

Bei der CPU Auslastung sehen wir auch, dass der SYS Anteil markant zunimmt, da das Volumen an *system calls* dramtisch ansteigt. Wir haben auch schon Plattformen gesehen, die bei einem solchen Volumen geradezu kollabieren (über 90% SYS Anteil).

Zusammenfassung

Die In-Memory Technologie ermöglich eine schnelle, einfache, kostengünstige und wirksame Performance Optimierung von Applikationen.

Flash Storage bietet einen Durchsatz und eine I/O Servicezeit, die bislang vollkommen unbekannt waren. Mit PCI attached Flash Storage kann die I/O Performance einzelner Datenbanken und Applikationen um Faktoren verbessert werden. Aber auch FC attached Flash Storage erreicht ein sehr hohes I/O Volumen bei fast so guten Servicezeiten.

Beide Technologien ermöglich eine deutlich bessere CPU Auslastung bis hin zur Sättigung und damit ein verbessertes Preis-/Leistungsverhältnis von Oracle Plattformen.

Kontaktadresse:

Manfred Drozd
Benchware AG
Seestrasse 18
CH-8800 Thalwil

Telefon: +41 (0) 44-722 16 16
Fax: +49 (0) 12-345 6788
E-Mail Ihre@adresse.de
Internet: www.adresse.de