

Michel Centi, Manfred Drozd, Andreas Zallmann, In&Out AG

Oracle Performance auf Raid-5 und Raid-10

RAID-10 vs. RAID-5

Keine Performance Steigerung bei Verwendung von RAID-10 gegenüber RAID-5.

RAID-10 liefert nur beim Schreiben von nicht sequentiellen und kleinen Datenblöcken bessere Leistungskennzahlen. Da diese Operation in Oracle asynchron ist, wird die Gesamtperformance nicht wesentlich verbessert.

In&Out AG

Kennzahlen

Die In&Out AG ist ein herstellernerut-rales und unabhängiges Consulting und Engineering Unternehmen und hat 2008 mit 34 Mitarbeitern 8.0 Mio. CHF Umsatz erzielt.

Dienstleistungen

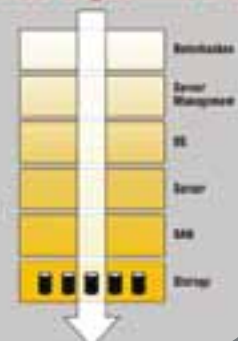


Mit den drei Geschäftsfeldern Informations-Sicherheit, Business Services und Plattformen deckt In&Out eine breite Palette von Dienstleistungen ab.

Beim Betrieb von IT Plattformen ist es unerlässlich, dass die einzelnen Technologien optimal aufeinander abgestimmt sind

Die Fachkompetenz von In&Out deckt alle Komponenten ab.

Wir haben detaillierte Kenntnisse der Technologien und Produkte von führenden Herstellern, verfügen über jahrelange Praxiserfahrung im Zusammenhang mit Migrationen und Konsolidierungen, kennen die Produkte im täglichen Einsatz – und somit deren Vor- und Nachteile – und haben breite Benchmark-Erfahrungen.



Bei Plattform Projekten ist eine übergreifende und durchgängige Sichtweise der verschiedenen Plattform-Layer vom Storage über Server bis hin zu Datenbanken oder Applikationen eine besondere Stärke der In&Out AG.

Im Speziellen setzt sich der Bereich Plattformen mit Optimierungen von High-End Systemen auseinander. In diesem Fall war die In&Out als externer Dienstleister verantwortlich für die Konzeption, Durchführung und Auswertung des Benchmarks.

Eingesetzte Tools

Für die Vermessung von Plattformen setzt In&Out zwei eigens entwickelte Produkte ein: IOgen und OraBench.

IOgen®

IOgen ist ein IO-Lastgenerator, welcher es ermöglicht, vordefinierte IO-Profile auf unterschiedlichsten Plattformen automatisiert ablaufen zu lassen. Die damit erzielten reproduzierbaren und vergleichbaren Messresultate können in nachfolgenden Tests (z.B. Oracle) als Grundlage für den IO-Durchsatz verwendet werden.

Folgende typische Key-Performance Indikatoren (KPI) von Plattformen werden durch IOgen ermittelt:

- Speed: Anzahl IO Operationen pro Sekunde (IOPS)
- Throughput: Übertragene MB pro Sekunde
- IO-Servicezeit (SVT): Übertragungsdauer pro IO

OraBench®

Mit der OraBench Suite lassen sich die Key-Performance Indikatoren einer Oracle-Datenbank auf einfache Weise eruieren. Analog zu IOgen besteht OraBench aus vordefinierten Tests (ca. 50 Stück), die automatisch ablaufen:

- PL/SQL execution
- Data load
- Data scan (sequential)
- Data select & update (random)
- Data aggregates

So kann das Performance-Verhalten einer Oracle-Datenbank inklusive der zugrundeliegenden Plattform im Detail gemessen werden. OraBench gelangt häufig bei folgenden Aufgabenstellungen zum Einsatz:

- Performance-Test einer neu implementierten Datenbank für die Produktion.
- Performance-Optimierung bestehender Datenbank Instanzen.

- Preis/Performance Vergleiche auf unterschiedlichen Plattformen, Plattform-Komponenten oder Konfigurationen.

OraBench liefert stabile, konsistente und reproduzierbare Performance-Zahlen. Diese helfen System-Architekten oder Applikations-Entwicklern z.B. folgende Fragen zu beantworten

- Wie viele Rows können innerhalb einer gewissen Zeit in die Datenbank geladen werden?
- Wie viele gleichzeitige Benutzer können durch die Plattform bedient werden?
- Welches ist der optimale Grad der Parallelisierung für die unterschiedlichen Datenbank-Operationen?

Oracle IO-Verhalten

In Oracle sind verschiedene typische IO Muster mit spezifischen Charakteristiken zu beobachten.

Random Reads (Oracle Select)

Normale Selects (z.B. über Indices) führen in der Regel zum Lesen von einzelnen DB Blöcken von 8 KB¹. Für diese Art von Lesevorgängen ist die Servicezeit der Disks für die Oracle Performance entscheidend, da der Prozess erst weiterarbeiten kann, wenn der angeforderte Datenblock zur Verfügung steht. Wenn viele Selects gleichzeitig stattfinden, ist die Gesamtbandbreite der Plattform relevant.

Sequential Reads (Multiblock IO)

Bei bestimmten Lesevorgängen, zum Beispiel Full Table Scans, geht Oracle dazu über, mehrere Datenbank Blöcke in einem IO gleichzeitig zu lesen. Die Anzahl der gleichzeitig gelesenen Blöcke wird über den Oracle Multiblock IO Parameter gesteuert. Üblicherweise werden in Oracle sequentielle Multiblock IOs von 1 MB ausgeführt. Hier ist die Servicezeit und die Gesamtbandbreite relevant.

¹ Die Oracle Blockgröße ist konfigurierbar zwischen 2 KB und 32 KB. 8 KB ist eine typische Blockgröße für OLTP-Systeme.

² ASM (Automatic Storage Management) ermöglicht ab Oracle 10g die Verwaltung der Disks innerhalb der Datenbank selbst.

Random Writes (Oracle Updates)

Random Writes finden beim Aktualisieren von Datenbank Blöcken statt. Diese sind normalerweise asynchron, d.h. Oracle wartet nicht auf das Schreiben der Blöcke. Hier ist die Gesamtbandbreite des Storage Systems von Interesse, damit die veränderten Datenblöcke beim Herunterschreiben auf den Storage nicht angestaut werden.

Sequential Writes (Oracle Redo)

Sequential Writes finden in Oracle in erster Linie auf den Online Redo-Logfiles und bei deren Archivierung statt. Die erste Operation ist synchron (Garantie der DB-Konsistenz). Somit ist die Servicezeit relevant für die Performance. Je nach Transaktionsvolumen und Änderungsrate kann auch hier die Gesamtbandbreite des Storage Systems zum Tragen kommen.

Caching

Caching findet auf dem Storage System und durch Oracle statt. Beim Einsatz eines Dateisystems mit Direct-IO oder mit ASM² findet kein zusätzliches Caching durch das Betriebssystem statt.

Als Lesecache wird die Oracle System Global Area (SGA) genutzt, da diese ist in der Regel grösser als der von der Datenbank nutzbare Storagecache ist. Beim Lesen «trifft» man den Block entweder in der SGA oder muss ihn von der physischen Disk holen. Als Schreibcache für die Schreibvorgänge wird der Storagecache genutzt, d.h. der IO wird vom Storage bestätigt, sobald dieser in den Cache geschrieben wurde. Dessen Servicezeit ist relevant für die synchronen Writes. Das Storage Backend muss jedoch in der Lage sein, synchrone und asynchrone Writes mit genügender Bandbreite auf die physischen Disks herunterzuschreiben (destaging).

Test szenarien

Grundlegend wurden die folgenden beiden Oracle-Test szenarien ausgewertet und miteinander verglichen.

- Oracle auf RAID-5 Storage
- Oracle auf RAID-10 Storage

Benchmark-Setup

Hardware

Server

- FSC Primergy RX600 S4
- 4 x Intel-Xeon Quadcore, 2.93 GHz, 128 GB RAM
- Suse Linux 10 EE 64-bit SP2
- 2 x 4 Gbit FC-HBA

Storage

- EMC CLARiiON CX3-40
- 2 x 4 Gbit Controller
- 2 Storage-Prozessoren mit je 4 GB RAM
- 20 Fibre-Channel Disks, 10k RPM, 300 GB

CLARiiON RAID-5 Setup

- 4 RAID-Gruppen mit je 4 Datendisks, 1 Parity-Disk³
- Kapazität: ca. 4.8 TB

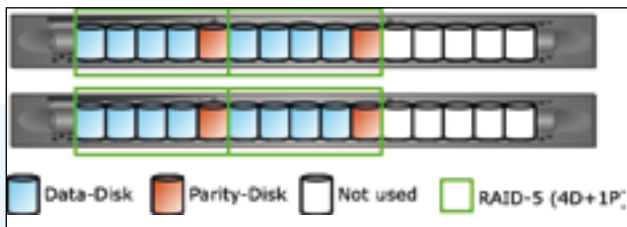


Abbildung 1
CLARiiON CX3:
RAID-5 Setup

Software

Oracle Datenbank

- Version 10.2.0.4 EE 64-bit
- 8 KB Blocksize
- 64 GB SGA, 16 GB PGA

IOgen

- Version 2.2.0
- Random & Sequential IOs
- Blockgrößen: 1 KB, 8 KB, 1024 KB
- Lesen & Schreiben

OraBench

- Version 6.8
- Large Database Configuration
256 GB
- 128 Mio. Rows, 32 Partitions à
0.85 GB

CLARiiON RAID-10 Setup

- 5 RAID-Gruppen mit je 2 gespiegelten Datendisks
- Kapazität: ca. 3.0 TB

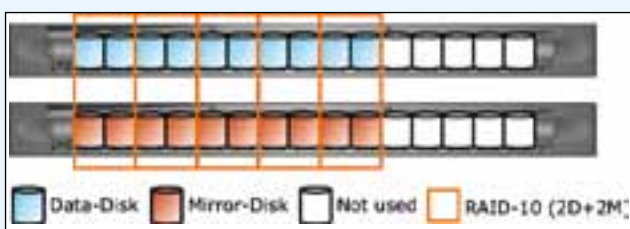


Abbildung 2
CLARiiON CX3:
RAID-10 Setup

³ In der RAID-5 Implementation ist die Parity nicht fix auf einer Disk, sondern wird aus Performance-Gründen über alle Disks verteilt.

RAID-10 vs. RAID-5

Um den Unterschied zwischen RAID-10 (Striped Mirror) und RAID-5 Konfigurationen messen zu können, wurde die gesamte Oracle-Datenbank (Datendateien, Online Redo-Logfiles und Archive Logs) auf dem jeweiligen RAID-Typ installiert.

Beide Setups wurden zu Vergleichszwecken mit der gleichen Anzahl Disks konfiguriert. Ebenso kam in beiden Fällen Oracle ASM zum Einsatz.

Die genaue Konfiguration wurde in Kapitel 4 erläutert.

Storage-Performance Random Read

Beim Lesen von 8 KB Blöcken ist die Leistung bei RAID-5 und RAID-10 relativ ähnlich. Beide Konfigurationen benötigen genau einen Lesezugriff auf eine Harddisk. Der Durchsatz wird durch die Anzahl der Disks bestimmt. Da beide Konfigurationen über je 20 Disks verfügen, die alle für das Lesen benutzt werden, zeigt sich hier kein wesentlicher Unterschied.

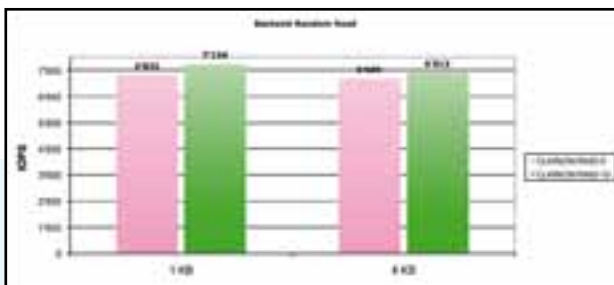


Abbildung 3 – RAID-5/10 Random Read Performance

Random Write

Beim Schreiben von 8 KB Blöcken hat RAID-5 einen wesentlichen Nachteil gegenüber RAID-10. Beim gespiegelten RAID-10 sind für jeden Block zwei Diskoperationen erforderlich, bei RAID-5 werden vier Diskoperationen benötigt, sofern kein Full-Stripe geschrieben werden kann. Zunächst müssen der überschriebene Block und die Parity gelesen werden, dann wird aus diesen und dem neuen Blockinhalt die neue Parity berechnet und schliesslich werden der neue Block und die neue Parity auf Disk

geschrieben. Bei diesem Vorgang wird vom sogenannten RAID-5 Write-Penalty gesprochen.

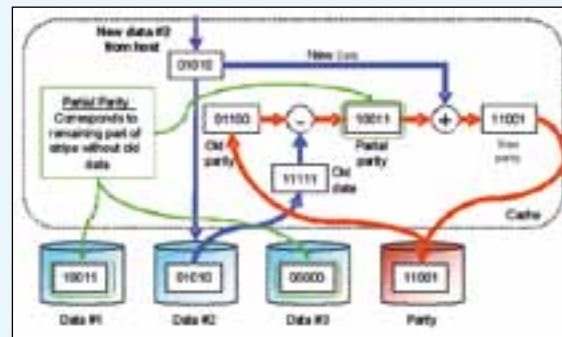


Abbildung 4 – RAID-5 Write Penalty, Quelle: HDS

Da bei RAID-5 vier Diskoperationen und bei RAID-10 nur 2 Diskoperationen pro Random Write erfolgen, ist die Bandbreite auf RAID-10 mit 5'297 IOPS etwa doppelt so hoch wie auf RAID-5 mit 2'600 IOPS.

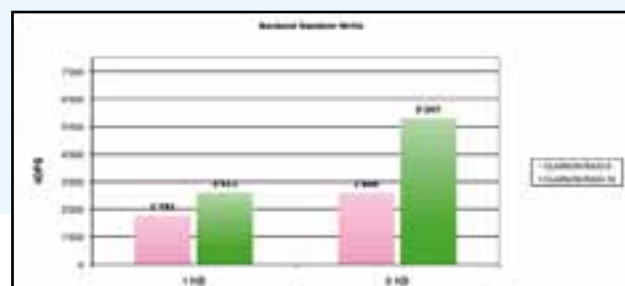


Abbildung 5 – RAID-5 / 10 Random Write Performance

Sequential Read

Ähnlich wie beim Random Read zeigen sich auch beim Sequential Read keine signifikanten Performanceunterschiede. Dies gilt für kleine und grosse Blockgrößen. Selbstverständlich lassen sich mit zunehmender Blockgrösse mehr Daten in MB/s transferieren.

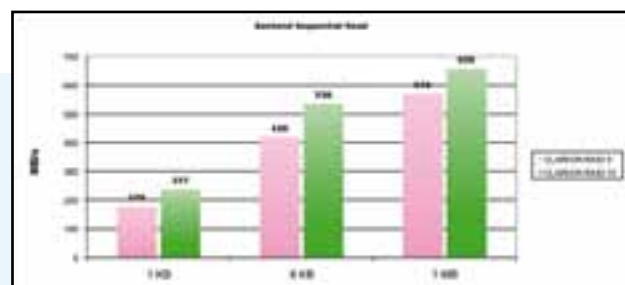


Abbildung 6 – RAID-5/10 Sequential Read Performance

Sequential Write

Interessant stellt sich die Situation beim Vergleich von Sequential Writes dar. Im Gegensatz zu Random Writes kommt in dieser Situation der RAID-5 Write Penalty nicht mehr zum Tragen, da das Stagesystem nun jeweils ganze Stripes von Daten erhält. Daraus kann direkt die (neue) Parity berechnet und der Stripe als Ganzes geschrieben werden.

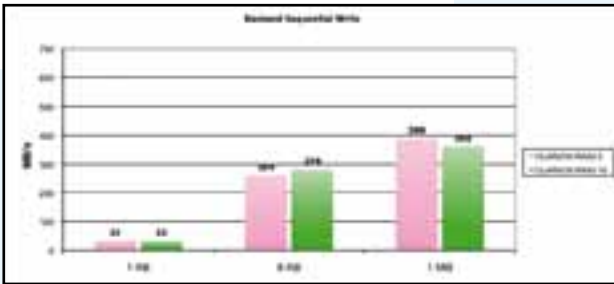


Abbildung 7 – RAID-5/10 Sequential Write Performance

Die Abbildung 7 zeigt demzufolge eine vergleichbare Sequential Write Performance über alle Blockgrößen. Aus diesem Grund werden auch sequentielle IOs wie Oracle Redo-Logfiles auf RAID-5 mit vergleichbarer Performance abgewickelt.

Fazit Storage

RAID-10 hat im Vergleich zu RAID-5 in genau einem Szenario Vorteile, nämlich für Random Writes mit relativ geringen Blockgrößen. Die Leseperformance ist in beiden Konfigurationen nur abhängig von der Anzahl der Disks und bei sequentiellen Writes oder Writes mit grossen Blockgrößen sind keine signifikanten Performanceunterschiede erkennbar.

Oracle-Performance

Mit OraBench wurde nach dem reinen IO-Test überprüft, wie sich die Unterschiede in einer Oracle Datenbank auswirken. Aus der Vielzahl der getesteten Szenarien wurden folgende typische Beispiele ausgewählt.

T311: Data Load (conventional)

250'000 Rows/Process, 2 Rows/commit
 Beim Laden von Daten in eine Tabelle, zeigen beide Konfigurationen eine absolut identische Performance. Bei diesem Test kommt es vor allem auf die Redo-Performance an, die mit 15 MB/s bei beiden Setups identisch ist. Die Anzahl der Writes bewegt sich für RAID-5 noch nicht im kritischen Bereich.

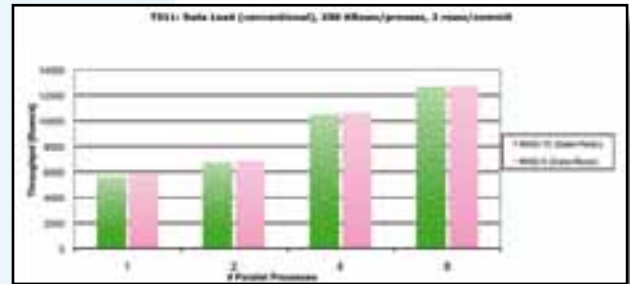


Abbildung 8 – RAID-5 und RAID-10 Oracle Data Load

T616: Data Select (random)

Primary-Key, 128 Mio. Rows, 32 Partitionen à 0.85 GB
 Bei Random Reads zeigen beide Konfigurationen in Oracle erwartungsgemäss die gleiche Performance, da die Leseperformance auf dem Stagesystem für RAID-5 und RAID-10 nahezu identisch ist.

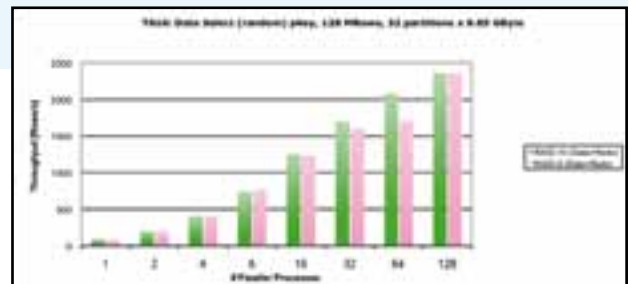


Abbildung 9 – Oracle Data Select auf RAID-5 & RAID-10

T716: Data Update (random)

Primary-Key, 128 Mio. Rows,
32 Partitionen à 0.85 GB

Interessant ist die Auswirkung der deutlich besseren Random Write Performance von RAID-10 in Oracle. Reine Random Writes kommen in Oracle jedoch gar nicht vor, jeder Block wird zunächst gelesen, bevor er geändert und wieder geschrieben wird, und parallel finden sequentielle Writes auf den Redologs statt. Demzufolge fällt der reale Performancevorteil bei Oracle Updates relativ ernüchternd aus und beträgt nur ungefähr 10%.

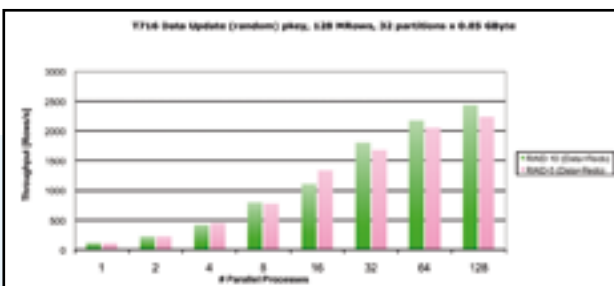


Abbildung 10 – RAID-5 und RAID-10 Oracle Data Update

Fazit Oracle

Analog zu den IO-Ergebnissen ist in einem RAID-10 Setup in fast allen Szenarien keine Performanceverbesserung in Oracle zu erzielen. Dies gilt insbesondere für die sequentiellen IOs der Redologs. Lediglich bei Random Writes zeigt sich ein geringer Performancevorteil von ca. 10% bei maximaler Parallelität. Die deutlich höhere Random Write Performance von RAID-10 kommt in Oracle aufgrund der gleichzeitig stattfindenden Reads und Sequential Writes nur in geringen Umfang zum Tragen.

Diesem geringen Vorteil stehen höhere Kosten entgegen. Da RAID-10 nur 2/3 der Kapazität aufweist (beim hier verwendeten Setup mit 4 Datendisks und 1 Paritydisk) sind die Kosten für die gleiche Kapazität ca. 50% höher. Deshalb ist in der Regel von einem Einsatz von RAID-10 für Oracle Datenbanken abzuraten. ■

Contact

In&Out AG

Michel Centi
E-Mail:
michel.centi@inout.ch

Manfred Drozd
E-Mail:
manfred.drozd@inout.ch

Andreas Zallmann
E-Mail:
andreas.zallmann@inout.ch

SMS

Leading Partners Embrace Oracle® Database 11g Release 2

Oracle Database 11g Release 2 Enables Partners to Deliver Fast, Reliable, Secure and Scalable

Leading partners worldwide are voicing their support for the latest generation of database innovation – Oracle® Database 11g Release 2 – now available today (read related press release).

Partners participating in the Oracle Database 11g Release 2 beta program as well as partner training and enablement programs include: Bentley Systems, CA, CSC (NYSE: CSC), Capgemini, Century Consultants, DataHeaven-Korea,

Escalate Retail, EMC, ESRI, Fujitsu America, Fundtech, HP and EDS, an HP Company, Intergraph, IFS, Jesta I.S. Inc., Kapow Technologies, KOLON I'Networks, Lieberman Software Corporation, McKesson Provider Technologies, MicroStrategy, Milletech Systems, Inc., NCS-Singapore, NetApp, Neusoft, Open Link Financial, Inc., OpSource, PricewaterhouseCoopers, Quest Software, Red Rock Consulting AU, S&I Systems Singapore, SAS, Siemens AG, Sinosoft, Spectrum K12 School Solutions Inc., SPSS Inc., Sterling Commerce, SunGard Energy Solutions, Symyx Technologies, Tata Consultancy Services, TEMENOS

UK, Teranode Corporation, Thales, Thermo Fisher Scientific, Thomson Reuters OpenCalais Initiative, TradeBeam, TUSC, Ufida, VelQuest Corporation, Ventyx, Voltaire, Where 2 Get It Inc. and Xactly Corporation amongst many others.

With Oracle Database 11g Release 2, partners help enable their customers to achieve a higher quality of service, while reducing the cost and complexity of information management by offering a fault-tolerant, high performance and scalable grid infrastructure that supports their business applications.