

# Exadata und Exalogic X4-2



Herbert Rossgoderer und Matthias Fuchs, ISE Information Systems Engineering GmbH

Dieser Artikel beschreibt die Entwicklung der Oracle Engineered Systems Exadata und Exalogic bis zum heutigen Tag und stellt die neuen Funktionen der X4-2-Generation vor. Zudem wird auf Aspekte wie Leistung, Installation, Kosten und Betrieb der Systeme eingegangen.

Die Exadata Database Machine, das Flaggschiff der Oracle Engineered Systems, ist bereits im Jahr 2008 als erstes System dieser Produktfamilie unter dem Namen „Exadata V1“, damals noch mit Hewlett-Packard-Hardware, auf den Markt gekommen. Ab dem Jahr 2009 waren diese Systeme dann als Exadata V2 mit Sun-Hardware ausgestattet, was bei allen Nachfolgemodellen beibehalten wurde.

Das System ist als optimale Plattform für die Oracle-Enterprise-Datenbank konzipiert, insbesondere für skalierbare Parallelverarbeitung im Real Application Cluster (RAC), bei dem I/O- und Prozessorleistung im Einklang stehen. Man spricht deshalb auch von einem optimal ausgegogenen System (well balanced system). Als Betriebssystem stehen Oracle Linux und Solaris x86 zur Wahl.

Die Exadata Database Machine wurde wie auch die spätere Exalogic von Anfang an in einem 19-Zoll-Schrank (Rack) als Gesamtsystem ausgeliefert. Man kann diese Racks komplett (Full Rack), als Halb-

Viertel- und seit der Version X3-2 auch als Achtel-Rack erwerben. Die Systeme sind voll redundant ausgelegt und können problemlos auf die nächste Größe erweitert werden. Außerdem lassen sich mehrere Schränke via InfiniBand zusammenschalten, die dann wiederum als ein integriertes System fungieren.

InfiniBand kam von Beginn an als Netzwerk-Technologie zwischen den Datenbank-Knoten und den Exadata-Storage-Servern (auch Exadata Cell Server) zum Einsatz.

Die aktuell in der Version X4-2 verwendeten InfiniBand-QDR-Verbindungen (40 Gbit/s) sind aktiv-aktiv gekoppelt und haben demnach eine Übertragungsrate von 80 Gigabit pro Sekunde pro Pfad. Um die OLTP-Performance zu verbessern, wurden die mechanischen Platten durch Flash-Speicher ergänzt und von Generation zu Generation in ihrer Kapazität erhöht. Im Zuge der Weiterentwicklungen stiegen die Anzahl der CPU-Cores sowie die Festplattenkapazität ebenfalls deutlich an. In Kombination von

	V1	V2	X2	X3	X4
<b>Release-Jahr</b>	2008	2009	2010	2012	2013
<b>Cores</b>	64	64	96	128	192
<b>Memory (GB)</b>	256	576	1152	2048	4096
<b>Storage (TB)</b>	168	336	504	504	673
<b>Flash (TB)</b>	0	5,3	5,3	22,4	44,8
<b>Connectivity (Gb/s)</b>	8	24	184	400	400

Tabelle 1: Full Rack Exadata von V1 bis X4-2

Software- und Hardware-Verbesserungen konnte so der Datendurchsatz von 8 Gigabit pro Sekunde (V1) auf 400 Gigabit pro Sekunde (X4-2) um den Faktor 50 gesteigert werden (siehe Tabelle 1).

### Exadata Software X4-2

Das für die extreme I/O-Leistungsfähigkeit entscheidende Innovationsmerkmal der Exadata Database Machine ist ein intelligenter Storage, der nur möglich ist, weil Datenbank und Storage sich gegenseitig kennen und die gleiche Sprache sprechen. Das Hauptziel des intelligenten Storage ist, den Datentransfer zwischen Speicher und Datenbank zu reduzieren. Die Verarbeitung von SQL-Statements soll weitestgehend im Storage erfolgen. Nur das Ergebnis der Abfragen wird zur Weiterverarbeitung an die Datenbank übermittelt (Smart Scan beziehungsweise Offloading).

Das Ziel, den I/O erst gar nicht entstehen zu lassen, ist über einen Storage-Index realisiert. Der weiteren I/O-Leistungssteigerung dienen die Flash-Karten in den Storage-Servern (Cell Flash Cache), die zur Lese-Optimierung (Smart Flash Cache), zur Verbesserung der Redo-Performance (Smart Flash Log) sowie als Write-Behind-Cache (Smart Flash Cache Write-Back) konfiguriert werden können. Zur weiteren Optimierung der I/O-Leistung und der Speicherkapazität wird ein spaltenorientiertes Komprimierungsverfahren für Tabellen angeboten, die sogenannte „Hybrid Columnar Compression“ (HCC), mit der sich Tabellen-Inhalte in günstigen Konstellationen um bis zu einem Faktor von 50 reduzieren lassen. Um bei der Konsolidierung vieler Datenbanken auf eine Database Machine die enorme I/O-Leistungsfähigkeit weiterhin für die kritischsten Datenbanken zu gewährleisten, kann man mit dem I/O Resource Manager eine entsprechende Priorisierung dieser Datenbanken definieren. Zusätzlich sorgt das in der X4-Generation neu hinzugekommene Network Resource Management für niedrige Antwortzeiten bei zeitkritischen Datenbank-Operationen, selbst wenn parallel dazu Reports, Batchläufe oder Backups durchgeführt werden.

### Exadata und In-Memory

Die neuen X4-2-Datenbank-Knoten haben standardmäßig 256 GB Hauptspeicher,

	X2-2	X3-2	X4-2
Cores	48	64	96
Memory (GB)	384	1024	1024
Storage TB	40	60	80
Storage SSD Read TB	2	4	6,4
Local SSD Raid 1	80	100	400

Tabelle 2: Achte-Exalogic von X2-2 bis X4-2

der auf 512 GB erweitert werden kann. Somit ist es bereits beim kleinsten System (Achte-Rack) möglich, auf 1 TB Hauptspeicher zurückzugreifen. Die Achte- und Viertel-Racks bestehen jeweils aus zwei Datenbank-Knoten und drei Storage-Servern. Dies ist die Mindestausstattung, um Redundanz und damit Ausfallsicherheit zu gewährleisten. Beim Achte-Rack ist jedoch die Hälfte der Prozessoren und Platten deaktiviert, wohingegen der Hauptspeicher voll vorhanden ist. Somit bildet bereits die Achte-Lösung eine ideale Plattform für die zukünftige In-Memory-Option der Datenbank 12c (12.1.0.2).

Für hoch performante OLTP-Anwendungen steht damit selbst bei einem Achte-Rack 1 TB Hauptspeicher (DRAM) zur Verfügung und für Echtzeit-Analysen kann auf den fast 5 TB großen Cell-Flash-Cache zurückgegriffen werden, der sich durch Hardware-Komprimierung effektiv auf das bis zu Fünffache erweitern lässt. Durch das massiv parallele Storage-Grid skaliert der I/O-Durchsatz linear über alle mechanischen Platten und Flash-Module beim Ausbau auf ein Viertel-, Halb- oder Full-Rack.

### Exalogic von X2-2 bis X4-2

Die Exalogic ist nach Exadata das zweite System im Rahmen der Engineered-Systems-Familie; später kamen Exalytics und Big Data Appliance hinzu. Exalogic ist optimiert für die Verwendung von Middleware-Applikationen. Dabei können Oracle-Applikationsserver wie WebLogic oder Tuxedo zum Einsatz kommen. Als Basis-Betriebssystem dienen Oracle Linux oder Solaris X86 – somit sind alle Arten von Anwendungen möglich, die unter diesem Betriebssystem zertifiziert sind. So entsteht ein breites Einsatzgebiet für dieses Engineered System.

Im Jahr 2011 wurde die erste Version (X2-2) der Exalogic ausgeliefert, 2013 folgte die Version X3-2, dieses Jahr die Version X4-2. Es wurde jeweils auf die aktuelle Prozessorfamilie gewechselt (Intel Xeon). Die Anzahl der Cores pro Prozessor ist von 6 über 8 auf 12 gestiegen. Tabelle 2 zeigt ein Achte-Rack im Überblick.

In der aktuellen Version X4-2 kommen auf einen Core gut 10 GB Hauptspeicher. Ein Core besteht dabei aus zwei Threads. Somit ist für jeden Thread ein Speicher von 4 GB vorhanden. Das entspricht in etwa der Größe einer Standard Java VM. Die Exalogic kann in drei verschiedenen Ausprägungen betrieben werden:

- Bare Metal
- Virtualisierung
- Mixed Bare Metal virtualisiert (neu)

In der Bare-Metal-Variante ist Oracle Linux oder Solaris X86 direkt auf den Servern installiert. Bei der Virtualisierung kommt Oracle VM zum Einsatz. Diese Variante ist optimiert für den InfiniBand Stack. Die virtuellen Maschinen (vServer) können nur mit Oracle Linux betrieben werden. Bei der Installation im Mixed Mode werden virtuelle Server mit Bare-Metal-Servern innerhalb einer Exalogic kombiniert. Somit kann man für den jeweiligen Einsatzzweck immer die beste Variante wählen.

Im Rahmen des letzten virtuellen Software-Release wird der Oracle Assembly Builder unterstützt. Somit lässt sich das gesamte System, zum Beispiel bestehend aus einem WebLogic Server, einer SOA Suite und einem Service Bus, im Rahmen eines Assembly auf der Exalogic innerhalb weniger Minuten in einem Schritt installieren. Darüber hinaus wurde die Größe des Exalogic-Control-Stacks von fünf virtuellen Maschinen auf drei reduziert. Die

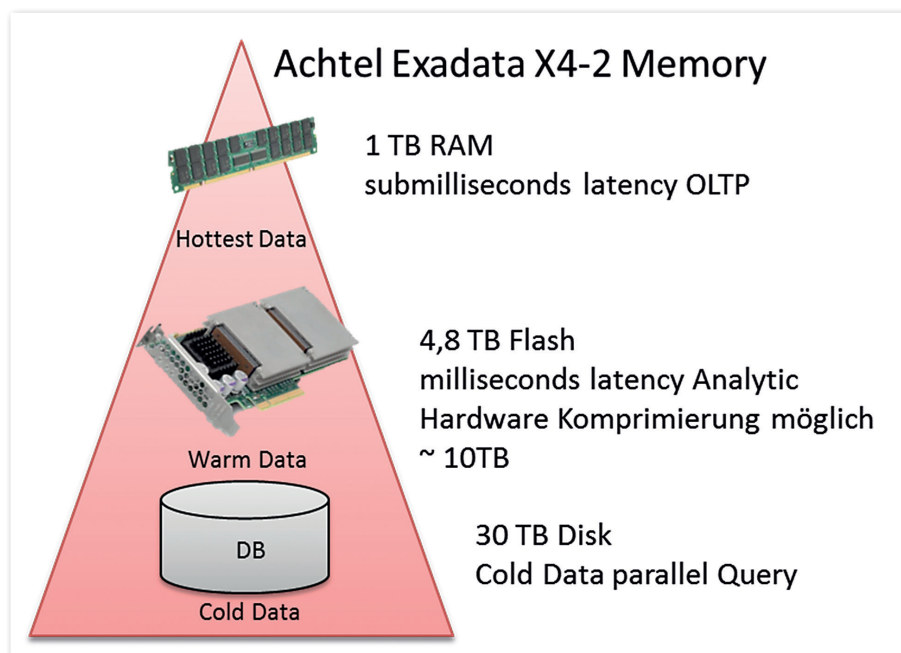


Bild 1: Memory- und Storage-Verhältnisse Achtel-Exadata X4-2 High Capacity Disks

Templates der virtuellen Maschinen wurden durch OSWatcher-Skripte erweitert; die Verwendung von Logical Volume Manager ist jetzt Standard. Die bekannten Oversub-Skripte der virtuellen CPUs, also mehr CPUs als vorhanden zuzuweisen, bleibt erhalten und der OVM-Stack ist auf die aktuelle Version 3.2 angehoben.

### Exalogic-Performance

Bei den Engineered Systems spielt der Exabus eine herausragende Rolle. Er beruht auf InfiniBand-Technologie. Diese wird im QDR-Modus eingesetzt, der eine Bandbreite in der X4-2 im active-active Mode (Exadata) von 80 GBit und deutlich niedrigere Latenzzeiten als 10 GBit Ethernet ermöglicht. Die Geschwindigkeit wird nicht nur innerhalb der einzelnen Knoten erreicht, sondern auch zwischen verschiedenen Engineered Systemen, etwa beim Zugriff auf eine Exadata. Zusätzlich zur schnellen Netzwerk-Verbindung werden speziell optimierte Treiber für den InfiniBand-Stack und die Server-Hardware verwendet, was die Performance der schnellen Exabus-Infrastruktur nochmals steigert. Hierzu zählen die Optimierungen für WebLogic, Tuxedo oder Coherence.

Neben den herausragenden Hardware-Komponenten gibt es Software, die nur im Rahmen der Exalogic verwendet werden

darf (gegebenenfalls auch auf der Database Appliance). Dazu zählt der Traffic Director, ein hochverfügbarer Software-Loadbalancer. Er kann neben den Standardprotokollen für Web-Applikationen (http/https) auch WebLogic-Protokolle wie „t3s“ oder jede TCP/IP-Connection verteilen. In Kombination mit der Möglichkeit, Daten zwischenspeichern und über den InfiniBand-Stack zu verteilen, ist das die ideale Ergänzung, um eine Exalogic einfach im Enterprise-Netzwerk zu integrieren. Es kann jeder Connect auf die Exalogic über den Traffic Director erfolgen. Alle weiteren Netzwerke, die nur intern verwendet werden, müssen im Firmennetzwerk nicht bekannt sein.

Ein entscheidender Vorteil eines Engineered Systems ist die komplette Integration in Cloud Control Enterprise Manager. Dadurch hat man alle Komponenten des Systems vom Switch über die Server und den Storage bis hin zum Software-Stack in einem Tool integriert. In Kombination mit dem Oracle Platinum Support werden die Systeme im Rahmen eines Supportvertrags kostenlos gepatcht. Im Falle eines Hardware-Defekts können automatisch Service Requests per ASR erzeugt werden. Patches werden immer über das ganze System getestet. Die Abhängigkeiten zwischen Storage, Servern und Switches ist somit immer gewährleistet.

### Die Key-Indikatoren Zeit und Kosten

Die Kosten für die Hardware einer Exadata oder Exalogic übersteigen leicht die Kosten für in etwa vergleichbare Standardserver. Die Einsparungen ergeben sich durch den schnelleren Aufbau einer kompletten Umgebung. Das Gesamtsystem ist etwa zwei Tage nach der Lieferung verfügbar, das heißt, die Hardware läuft und das Betriebssystem beziehungsweise die Datenbanksoftware sind installiert. Nach erfolgreicher Integration in Oracle Cloud Control, die je nach Kundenumgebung in einem halben Tag erledigt ist, kann mit dem Provisionieren der Software oder dem Aufsetzen der Datenbanken begonnen werden. Die Installation von Oracle-Middleware-Software wie SOA Suite oder WebCenter kann auf Basis von Templates oder Assemblies erfolgen, die Oracle zur Verfügung stellt. Hält man sich an den Enterprise Deployment Guide für Oracle Fusion Middleware, ist die Installation mit Cloud-Control-Automatisierung in wenigen Stunden erledigt.

### Das ISE Oracle Technology Center

Die Firma ISE Information Systems Engineering GmbH mit Firmensitz in Gräfenberg und Niederlassungen in Nürnberg und München betreibt das einzige Oracle-Exa-Stack-Test-Center mit Exadata, Exalogic und Exalytics in Deutschland. Es ist im modernsten 5-Sterne-Rechenzentrum Europas bei der noris network AG in Nürnberg untergebracht. Dort können Kunden die neuesten Oracle-Technologien, bestehend aus Hardware und Software, für ihre Einsatzzwecke evaluieren und die professionelle Beratung durch die ISE-Exa-Experten erfahren.

Die Firma ISE wurde am 20. Januar 2014 mit dem renommierten Oracle Excellence Award Germany für das Jahr 2013 in der Kategorie „Engineered Systems“ ausgezeichnet. Dies ist bereits die zweite Auszeichnung innerhalb von drei Jahren; bereits im Jahr 2011 hat ISE den Oracle Excellence Award Germany in der Kategorie „Exadata“ erhalten.

Der Enterprise Deployment Guide gibt vor, wie eine hochverfügbare Installation von Oracle Middleware Software erfolgen sollte. Voraussetzung ist, dass die Software Library des Cloud Control gefüllt ist. Somit kann man eine Woche nach Lieferung der Hardware mit den ersten Applikations-Deployments oder ersten Datenbank-Migrationsonstests starten.

Das Lizenzmodell von Oracle-Fusion-Middleware-Produkten für Exalogic basiert auf Trusted Partition. Dies bedeutet, es können Teile eines Servers lizenziert werden. In einer virtuellen Umgebung basiert der Core-Faktor auf virtuellen CPUs. Zwei virtuelle CPUs entsprechen einem physikalischen Core. Es muss maximal die Anzahl aller physikalischen Cores lizenziert werden.

### Fazit

Exadata und Exalogic X4-2 bieten viele größere und kleinere Neuerungen. Durch

die Erhöhung von Memory- und Flash-Speicher in der Exadata ist Oracle auf dem Weg zur In-Memory-Datenbank. Durch die zu erwartenden neuen In-Memory-Features der Version 12.1.0.3 und der X4-2-Hardware ist eine nochmalige deutliche Steigerung der Performance möglich.

Die Virtualisierung der Exalogic ist mit der neuen Hardware und den neuen Features in Version 2.0.6 eine echte Alternative. Die Performance der Bare-Metal-Installation und der Virtualisierung durch die konsequente Verwendung des InfiniBand Stacks ist ohnehin konkurrenzlos. Die betrieblichen Möglichkeiten durch die Verwendung von Assemblies und erweiterten Installations-Features über Cloud Control sind im Rahmen von Oracle-Produkten ein weiteres Alleinstellungsmerkmal.

Die DOAG veranstaltet am 14. Mai 2014 im ISE Oracle Technology Center den DOAG Exaday 2014, [www.doag.org/go/exaday](http://www.doag.org/go/exaday)



Herbert Rossgoderer  
[herbert.rossgoderer@ise-informatik.de](mailto:herbert.rossgoderer@ise-informatik.de)



Matthias Fuchs  
[matthias.fuchs@ise-informatik.de](mailto:matthias.fuchs@ise-informatik.de)