

# Einsatzkriterien und Erfolgsfaktoren Oracle Database Machine / Exadata

Alfred Schlaucher, Oracle , Hamburg

## Schlüsselworte:

Database Machine (DBM), Exadata, Nutzen, Einsatzkriterien

### *Solo-Storage, -Netzwerk, -Server sind von gestern: heute ist die Integrierte Black Box*

Die Oracle Database Machine (DBM) tritt nicht von ungefähr gerade jetzt in den Markt der integrierten Datenbank-Systeme. Seit Jahren kann man eine schleichende Veränderung bei Lösungen für solche High End Datenbank-Systeme erkennen. Das Thema „Daten“ ist nach wie vor einer der herausragenden Aspekte in der IT. Durch neue Techniken ist das Sammeln von Daten einerseits leichter geworden, gleichzeitig ist es Bestandteil vieler neuer Geschäftsideen und Geschäftsprozesse. Es klingt nach einem Allgemeinplatz: Aber die dabei entstehenden Datenmengen wachsen schneller als vor Jahren vermutet. Allerdings entsteht bei Verantwortlichen ein neuer bewussterer Umgang mit dem Phänomen *Datenwachstum*, denn es bedeutet vor allem mehr Kosten.

Technologische Gegebenheiten haben in den vergangenen 20 Jahren zu einem festen Schema von Hardware-Architekturen und ganzen Hardware-Parks geführt. Die Anforderungen an Hardware haben sich in Form von Hardware-Segmenten verselbständigt, was oft auch zu organisatorischen Strukturen in den Anwenderunternehmen führte:

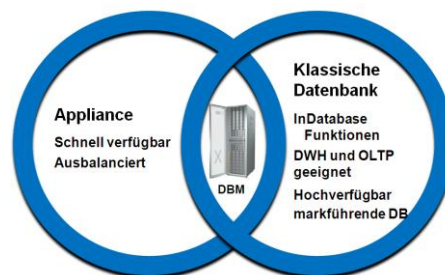
- Es entstanden zentral administrierte SAN-Landschaften. Die Speichersysteme mit Caches und Clonig-Funktionen sind heute Hochleistungssysteme.
- Server-Maschinen wurden zu virtuellen Rechnerverbänden gekoppelt.
- Unternehmensweit wurden Netzwerkverbindungen zu *öffentlichen Autobahnen* standardisiert.

Diese Entwicklung verstellt den Blick auf die neuen Herausforderungen, die große Datenmengen mit sich bringen. Derart spezialisierter und hochgezüchteter Umgang mit Storage, Servern und Netzen führt angesichts der wachsenden Herausforderungen von hohen Datenmengen zu gewaltigen Kosten, zu fehlender Flexibilität gegenüber besonderen Anforderungen und letztlich zu einer kaum noch verstehbaren Komplexität.

Andererseits wuchsen fast unbemerkt von der breiten IT-Community Lösungen, die mit diesen tradierten Gewohnheiten getrennter Verwaltung von Hardware-Einheiten brechen.

Haben Lösungen wie Sand oder Sybase IQ sich noch auf eine modifizierte Speicherstruktur beschränkt (spaltenorientiert), bieten Netezza oder Teradata komplette Hardware-Software-Pakete für Spezialaufgaben an. Wie umkämpft dieses Segment heute ist, zeigen der Kauf von Sybase durch SAP, die Integration von DataAllegro in Microsofts SQL Server 2008 oder die Bündelung der Speziallösung Greenplum mit EMC Hardware.

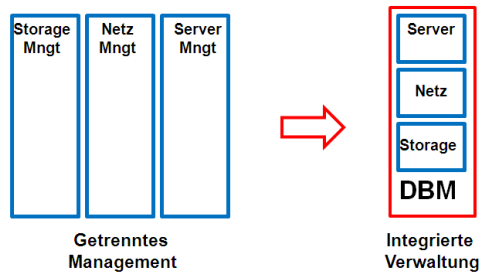
Die Exadata / DBM fügt sich zunächst in diese Reihe ein. Sie ist jedoch nicht mit einer herkömmlichen Appliance vergleichbar, denn bei der DBM treffen die Vorteile einer Appliance mit den vielen wertgeschätzten Features einer über Jahre bewährten und gut eingeführten Datenbank zusammen.



Die DBM führt IT-Verantwortliche in ein Dilemma. Auf der einen Seite tradierter segmentierter IT-Betrieb von getrenntem Storage-, Netz- und Server-Management, auf der anderen Seite eine All-In-One-Lösung, die neben bestehenden Leistungsmerkmalen auch bei Kosten und Betrieb Vorteile verspricht. Sie speichert bis zu 28 TB Nutzerdaten und liest diese mit bis zu 21 GB/Sec aus. Sie besteht aus 168 Platten, einem 8 Knoten RAC-Cluster-Serververbund und einem

leistungsfähigen intern Netzwerk mit 40 - 80 Gb/sec Netzdurchsatz. Die Datenbank-Software ist bei Auslieferung komplett installiert. Die Maschine ist bereits nach wenigen Tagen einsatzbereit.

Was sind die Einsatzbedingungen und speziellen Nutzeneffekte der DBM? Die folgende Aufstellung kann bei diesem Dilemma helfen.



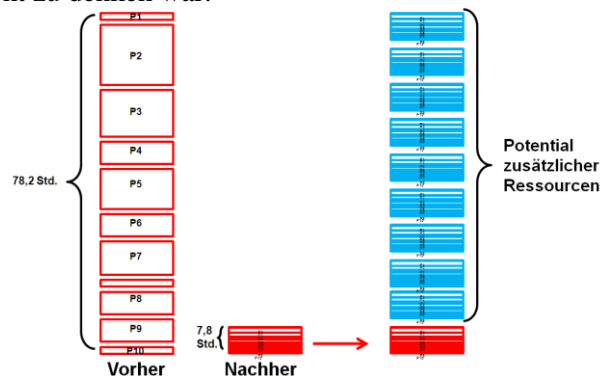
### Wozu so viel Performance?

Über Performance bei der *DBM* zu sprechen ist fast schon unfair. Denn bereits die kleinste Ausbaustufe (Quarter Rack) liefert mit 4,5 GB/Sec Leseperformance ein Vielfaches dessen, was traditionelle Systeme schaffen (oft kleiner als 1GB/Sec kaum mehr als 2 GB/sec).

Der Nutzen der extrem hohen Performance liegt dabei nicht in einer Verkürzung der Abfragezeit von vielleicht 8 auf 1 Sekunde. Der Nutzen liegt in dem Potential von fast unerschöpflichen Leistungsreserven.

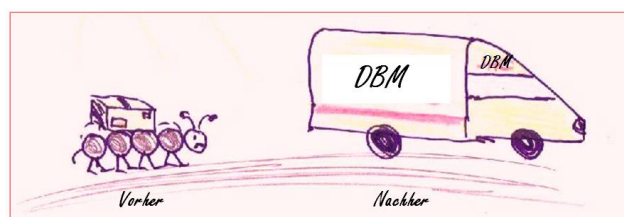
Die entscheidende Frage lautet nicht: „Wie schnell soll die Anwendung X laufen“, sondern „Wieviel zusätzliches Potential für mehr und weitere Aktivitäten ist gut für mein Unternehmen?“.

In einer Analyse fand ein Exadata-Anwender bezogen auf eine Testreihe von 10 standardisierten Auswerte-Jobs heraus, dass diese 10 Jobs auf seiner alten Hardware 78,2 Std Laufzeit benötigten. Die gleichen Jobs waren auf der *DBM* aufaddiert in 7,8 Std. fertig. Es blieben 70,4 Std. freie Zeit. Was geschieht in diesen 70,4 Stunden freier Rechenkapazität? Es ist Kapazität für zusätzliche Dinge, an die zuvor nicht zu denken war.



### Anwender werden mutig

Beobachtungen zeigen, dass Anwender mit der schnelleren *DBM* mehr Abfragen formulieren und ihre Analysen plötzlich einen Komplexitätsgrad erreichen, an den sie sich zuvor nicht heranwagten. Während in der Vergangenheit manche Abfragen Stunden benötigten oder ergebnislos abbrechen, sind plötzlich bereits nach Minuten und Sekunden Daten auf dem Bildschirm. Das macht Anwender mutig und sie formulieren neben mehr auch komplexere Abfragen. Das zeigt, dass bisherige Hardware-Lösungen mit Einschränkungen für die Anwender verbunden waren. Anwender hätte mehr Nutzen aus den Datenbanken gezogen und sie tun es jetzt mit der *DBM*.



## **Die Chance für Optimierungen**

Manche Umgebungen sind so stark ausgelastet, dass kaum Gelegenheit für Optimierungen bleibt. Ein Umbauen von Ablaufprozessen oder das Testen von neuen, effektiveren Komponenten wird durch fehlende Kapazität verhindert. Dabei sind Optimierungen angesichts geänderter Geschäftsregeln oft nötig. Mit der extrem schnellen DBM sind solche Optimierungen leicht machbar weil die Kapazität für zusätzliche Programmläufe gegeben ist.

### **Schnellere Services**

Die Verarbeitung von Daten in IT-Systemen stellt oft ein zeitlicher Hemmschuh für Verfahren und Geschäftsprozesse dar. Wenn die Durchführung von betrieblichen Maßnahmen, Marketing-Kampagnen, Bereitstellung von Erhebungsdaten, Bereitstellen von Datenservices usw. von der Rechengeschwindigkeit der Hardware abhängt, dann hat Performance direkten und messbaren Einfluss auf geschäftlichen Erfolg. Lebt ein Unternehmen von dem Verkauf solcher Services z. B. an Partnerunternehmen, dann ist eine Erhöhung der „Schlagzahl“ im Rechenzentrum bares Geld. Es gibt viele Praxisbeispiele, bei denen die Einführung der DBM zu einer Verbesserung der Service-Angebote führte.

### **Online Verfügbarkeit maximieren**

Es sind oft Verwaltungsaufgaben wie das Erstellen einer Backup-Sicherung oder das Neuaufbauen von Indexstrukturen oder das Neu-Komprimieren großer Tabellen, die zu geplanten *Downtimes* führen und Nutzer von ihren Daten ausgrenzen. Die extreme Performance der DBM erlaubt die parallele Durchführung von solchen Aktivitäten auch im laufenden Online-Betrieb.

## **Wieviel Terabyte für wieviel Geld?**

Die vorgenannten Beispiele zeigen auf, dass die Performance und das damit verbundene Potential Entscheidungskenngrößen für den Einsatz der DBM sind. Glücklicherweise liefert die DBM bereits in einem Quarter Rack mit 6 TB nicht komprimiertem Speichervolumen genügend Platz. Die Performance ist das Maß der Dinge und nicht der Speicherplatz. Es sind Konstellationen möglich, bei denen sich mehrere OLTP-Anwendungen den Storage teilen und in der Summe nur 1 TB Datenvolumen beanspruchen auch wenn 6 und mehr TB zur Verfügung stehen.

Selbst wenn man größere Volumen benötigt, muss noch nicht eine größere Ausbaustufe gewählt werden. Vorausgesetzt die Performance von 4,5 GB/Sec reicht aus, kann ein Quarter Rack durch ein neues Spalten-bezogenes Komprimierungsverfahren bis zu 20 TB Nutzdaten aufnehmen (der Begriff Nutzdaten meint Tabellen und Indexdaten ohne Spiegelung oder Verwaltungsdaten wie Temp, Rollback etc.).

Die entscheidungsrelevante Frage lautet nicht: „*Was kostet 1 TB Speicherplatz?*“ sondern „*Was kostet 1 TB Speicherplatz der mit 4,5 oder 10 oder 21 GB/sec Performance lesbar ist?*“.

## **Kalkulierbare Skalierung**

Eine häufig genannte Anforderung ist die Planbarkeit von Rechen- und Platzkapazitäten. Die technische Architektur der DBM erlaubt lineares Skalieren. Das bedeutet, dass neben dem Hinzufügen von Plattenkapazität auch die Performanceleistung proportional steigt. Die interne Verarbeitung (z. B. Interconnect zwischen den RAC-Knoten mittels Infiniband) lässt keine Latenzen für eine interne Koordination aufkommen. 2-fach eingesetzte Hardware ist auch doppelt so schnell.

## **Senken des Verwaltungsaufwands und der Betriebskosten**

### **Zwei Drittel weniger Administrationsaufwand**

Die DBM erfordert 2/3 weniger Verwaltungsaufwand als herkömmliche Lösungen. Das ist leicht nachvollziehbar, wenn man folgende Fakten betrachtet:

- *Weniger Abstimmungen:* Administrativen Tätigkeiten rund um Storage, Netzwerk, Server, Betriebssystem- und Datenbank-Software können an einer Stelle zusammengefasst werden. Erheblicher Zeitaufwand auf Abstimmungen zwischen verantwortlichen Stellen entfällt.
- *Mehr freie Wochenenden:* Die extreme Performance der Maschine verkürzt Verwaltungsarbeiten wie Re-Indizierung, Kopieren, Sichern und Komprimieren von Daten und reduziert die Arbeitszeit der Mitarbeiter. Verplante Wochenenden werden für Administratoren zur freien Zeit.

- *Weniger Indexe:* Bis zu 50% und mehr der Indexstrukturen entfallen, weil die DBM bei sequentielltem direktem Lesen auf den Tabellen z. T. schneller arbeitet als auf den Indexen. Dadurch verringert sich die Zeit für Re-Indizierung.
- *Abgestimmtes Patching:* Patches sind in der Regel durch Rolling Upgrades und die ausfallsichere Architektur der DBM schneller und unkomplizierter einzuspielen. Durch die Abstimmungen von Betriebssystem, Datenbank-Software und Firmware treten weniger Komplikationen auf.

### **Weniger Aufwand für Tuning und Konfiguration - weniger Experten**

Als Appliance ist die DBM fertig vorkonfiguriert. Es entfällt der oft Monate dauernde Aufwand für Auswahl, Sizing und Abstimmung von Storage, Netzwerk und Server. Das macht sie so unkompliziert. Es sind seltener teure Experten für Tuning und Performance-Analysen nötig.

### **Weniger Strom, Kühlung und Stellfläche im Rechenzentrum**

In modernen Rechenzentren schlagen Betriebskosten wie Strom, Kühlung aber auch Stellflächen-bezogene Umlagen für Operating besonders kräftig zu Buche. Die DBM integriert Plattenkapazität von 100 TB (und mehr) inkl. Netzwerk und 8-Knoten-RAC-Cluster auf einer Stellfläche von 0,8 m<sup>2</sup>.

### **On-The-Fly kürzere Verarbeitungszeiten ohne Umprogrammieren der Anwendung**

Die DBM geht einen riesigen Schritt in die Richtung kürzerer Aktualisierungsfrequenzen und Near-Realtime im Data Warehouse. Wichtig dabei ist: *Dies ist möglich, ohne die Verarbeitungslogik der bestehenden Anwendung zu ändern.* Viele heutige Realtime-Konzepte ignorieren leider immer noch die in einem Data Warehouse nötigen Verarbeitungszeiten für Transformationen und Aggregationen. Diese Zeit sind nun mal auch bei einem Realtime-Konzept zwingend nötig. Aber genau solche Zeiten werden durch schnelleres Laden und weniger Index-Aktualisierung mit der DBM auf bis zu ¼ reduziert (Schreibgeschwindigkeit 8TB/Std, Wegfall vieler Indexe, zusätzliche parallele Abläufe etc).

### **Ein neuer Abteilungsrechner?**

Die DBM bietet die Chance tradierte und aufwendige Betriebskonzepte über Bord zu werfen und einen deutlich schlankeren Betrieb zu implementieren. Die einfache Verwaltung hat manche Abteilung schon dazu veranlasst eine DBM in eigener Verantwortung zu betreiben. Das Wissen eines Oracle-Datenbank-Administrators reicht aus, um auch die komplette Hardware-Umgebung zu verwalten. Damit ist es auch für kleinere Unternehmen machbar und erschwinglich High-End-Storage und -Datenbank-Lösungen zu betreiben.

### **Zusammenfassend sind die Einsatzkriterien:**

- Besteht Bedarf nach mehr Abfrageperformance für DB-Benutzer?
- Müssen zusätzliche Last und Aufgabenstellungen über das bisherige Leistungsvermögen hinaus abgebildet werden?
- Sind Verwaltungs- und Betriebskosten zu senken?
- Sollen Betriebsabläufe durch weniger involvierte Stellen vereinfacht werden?
- Muss für mehr Leistung / Volumen in der Zukunft geplant werden?
- Soll bei der Gestaltung von Programm- und Job-Abläufen mehr flexible Optionen geschaffen werden?
- Soll eine höhere Online-Verfügbarkeit gewährleistet sein?
- Sollen Ladeläufe in einem DWH kürzer werden?
- Ist das zu speichernde Datenvolumen größer als 1 TB?
- Steht ein Hardware-Wechsel in den kommenden 18 Monaten an?
- Soll der Betrieb von Storage, Server und Netzwerk in die eigene Verantwortung gelegt werden?

Kontakt Daten: Alfred.Schlaucher@oracle.com