

Die Zukunft von OLAP in Zeiten (fast) unlimitierter Hardware-Leistung und Frontend-Funktionalität

Marc Bastien, ORACLE Deutschland B.V. & Co. KG

Dieser Artikel beschäftigt sich mit der Untersuchung, wo OLAP-Technologien in heutigen Zeiten noch sinnvoll einsetzbar sind. Sind sie es überhaupt? Wo sind die Grenzen von traditionellem OLAP und wo jene der Modellierung in modernen BI-Werkzeugen? Kann (fast) unbegrenzte Hardwareleistung OLAP überflüssig machen oder den Nutzen von OLAP sogar verstärken?

Der Aspekt der reinen Performance-Steigerung und Frontend-Nutzung bei der Nutzung von OLAP wird an dieser Stelle ausgeblendet. OLAP-Technologie war und ist für die Berechnung von Szenarien, alle Arten von Planungen, Verteilungsalgorithmen, Zeitreihen sowie für statistische Methoden und einfaches Data Mining bis heute eine wichtige Basis, die viele Hersteller zu Recht nutzen.

OLAP – Gründe, Nutzen und Kosten in der Vergangenheit

Um die Bedeutung von OLAP in vielen Unternehmen zu erkennen, muss man auch die Vergangenheit berücksichtigen, in der diese Architekturen zuhauf beschafft und in Betrieb genommen worden sind: vor oftmals fünf bis zehn Jahren, als sich Data Warehouses (DWH) durchgesetzt haben, die Abfrage der nun zentral gespeicherten Daten sich jedoch langsam und unflexibel gestaltete. Zu dieser Zeit konnten aber die meisten DWHs nicht die Anforderungen an flexible („Ad-Hoc“) Analysen befriedigen. Oft konnten aus den DWHs nur mittels Programmierung oder mit Tools wie „Oracle Reports“ (oder entsprechenden Werkzeugen anderer Hersteller) Listen und Standardberichte erzeugt oder mit Tools wie „Oracle Discoverer“ (oder entsprechenden Werkzeugen anderer Hersteller) neue Abfragen ohne SQL-Kenntnisse erstellt werden. Letztere durchaus schon mit deutlichem Vorteil gegenüber den Werkzeugen für das Reporting, aber je nach Aufbau der abzufragenden Datenbank hatten die-

se Abfragen auch gern Laufzeiten, die eher in Minuten oder Stunden gemessen wurden als in Sekunden.

So wurde in vielen Unternehmen OLAP eingeführt: Einerseits versprach man sich dadurch die Flexibilität der typischen OLAP-Frontends mit „Drag&Drop“, „Slice&Dice“, „Drill“ etc., um nur einige Schlagwörter zu nennen. Andererseits lieferten die OLAP-Datenhaltungen schon damals ohne Zutun Abfragegeschwindigkeiten, die selbst gut optimierte relationale Datenbanksysteme nur schwer erreichten.

Die Einführung der OLAP-Technologie zur Performance-Steigerung hat dann unterschiedliche Wirkung gezeigt. Zum einen für den Endanwender, der nun ein flexibles Frontend nutzen konnte, in dem die wichtigsten Geschäftsdaten des Unternehmens für diesen leicht nachvollziehbar modelliert waren, und der auf fast immer gute Performance setzen konnte. Ein Frontend, mit dem jeder schon nach kurzem Training umgehen konnte und das nach intensivem Training umfangreiche, komplexe Logik und Modelle, inklusive farbenfroher Formatierung, ermöglichte.

Zum anderen gingen diese Tools mit einer anderen Form der Datenspeicherung einher: Daten wurden in „Cubes“ außerhalb der relationalen Datenbank gespeichert, was in EDV-Abteilungen nicht immer für Freude sorgte. Je nach Technologie wurde für jede Geschäftsanforderung ein neuer Cube angelegt (man kennt Unternehmen mit Hunderten von „Cubes“), deren Datenvolumen zusammen ein Mehrfaches des ursprünglichen DWH erreichten.

Zusammenfassend kann man sagen, dass es eine Zeit gab, in der diese Architektur richtig und wichtig war: Die relationale Technologie war nicht für diese Analysen bereit, es gab keine oder wenige BI-Tools, die auf relationalen Datenbanken die gewünschte Funktionalität boten, was OLAP-Tools und OLAP-Datenspeicherung per se enthielten. Die Vorteile, ja die Notwendigkeit der Funktionalität überstiegen den Aufwand bei Weitem beziehungsweise waren schlicht anders nicht erreichbar.

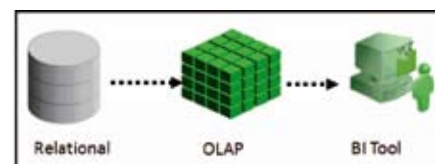


Abbildung 1: klassische DWH-OLAP-BI Schichten

OLAP, BI und DWH heute

Die klassische OLAP-Technologie wird heute von zwei Seiten bedrängt: Auf der einen Seite entwickeln die unterschiedlichen BI-Tool-Hersteller (nicht zuletzt Oracle selbst nach der Freigabe von Oracle BI EE 11g) BI-Technologien, die denen der OLAP-Tools in nichts oder nur wenig nachstehen. „Drag&Drop“ und „Drill“ sind heute keine Attribute mehr, die reinen OLAP- (oder, noch schlimmer, Client-Server-) Tools vorbehalten sind. Moderne BI-Tools arbeiten auf einer Metadaten-Schicht, die, je nach Hersteller, von der darunterliegenden Datenspeicherung unabhängig sind und „Cubes“ auf Da-

ten simulieren, die aus unterschiedlichen Quellen, aus unterschiedlichen Technologien (wie Excel, XML, Text, aus relationalen oder multi-dimensionalen Datenbanken, sogar direkt aus operativen Systemen wie Oracle Applications, PeopleSoft, SAP oder anderen) stammen.

Auf der anderen Seite werden von den Herstellern der relationalen Datenbanken immer mehr Anstrengungen unternommen, um performante Abfragen direkt auf der RDBMS sicherzustellen: relationale analytische Funktionen, OLAP- und Data-Mining-Engine in der Datenbank, besondere Indizes und weitere Optimierungen. Alles mit dem Hintergrund, dass nicht mehr die Daten zu den Algorithmen bewegt werden, sondern die Algorithmen zu den Daten. Einen anderen Ansatz verfolgt Oracle mit der Database Machine Exadata: Die fertig konfigurierte „Appliance“ stellt die notwendige Leistung zur Verfügung, um große Datenmengen in kurzer Zeit zu durchsuchen.

Der Nutzen einer multidimensionalen Datenhaltung darf aus dieser Sicht bezweifelt werden. Warum sollten Daten in eine zusätzliche Speicherungsform überführt werden, wenn diese ebenfalls „OLAP-like“ direkt in der vorliegenden Form analysiert werden kann? Diese Behauptung soll anhand der neuen Funktionalität von Oracle BI EE 11g und der Performance der Database Machine untersucht werden.

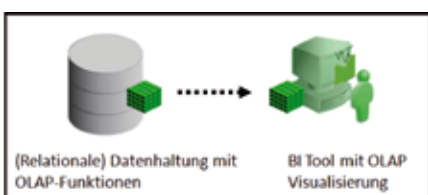


Abbildung 2: Datenhaltung mit OLAP-Funktionen

Abwägungen für eine tragfähige, aber schlanke BI/DWH-Architektur

Wie Abbildung 2 zeigt, müssen einige der OLAP-Funktionalitäten durch das BI-Tool übernommen werden, damit eine benutzerfreundliche, nutzbare und genutzte Plattform ohne speziell

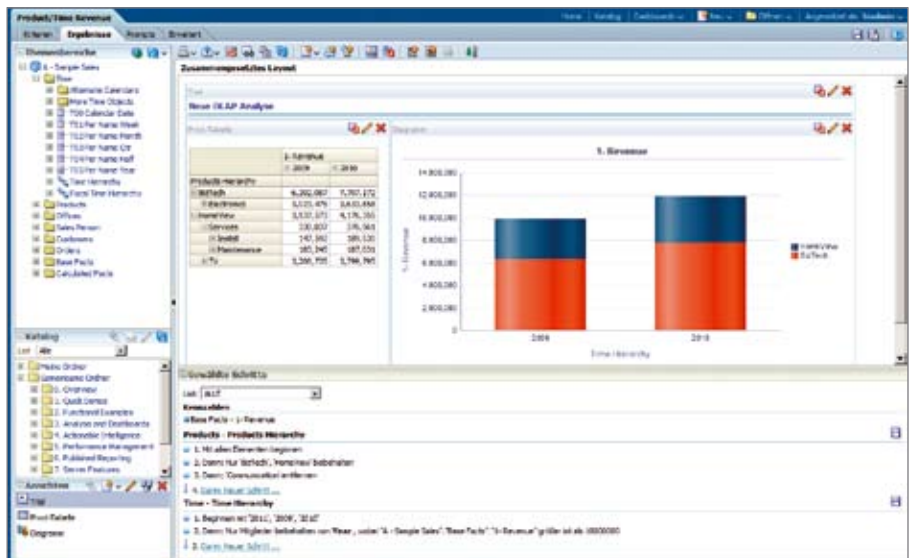


Abbildung 3: OLAP-Analyse mit Oracle BI EE 11g

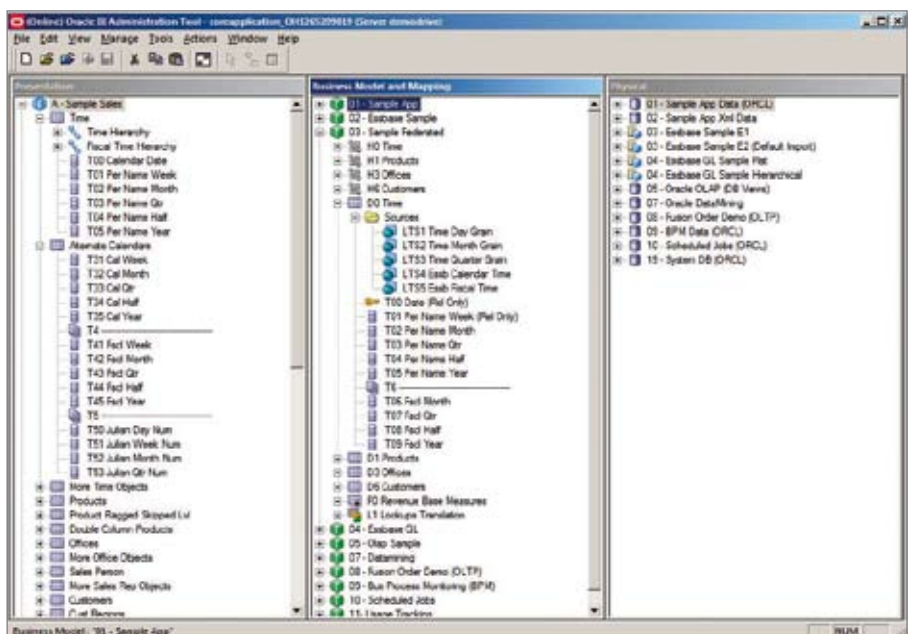


Abbildung 4: OLAP-Modellierung mit Oracle BI EE 11g

le OLAP-Technologien angeboten werden kann:

1. Was macht OLAP aus der Sicht eines Endanwenders denn so attraktiv? Es ist der intuitive Umgang mit dem Tool und den Daten. Und genau dieser Punkt ist es, den Oracle mit der Version 11g des BI-Tools adressiert: Da werden Spalten per „Drag&Drop“ vertauscht, Hierarchien mit „Drill up“ und „Drill down“ auf- und wieder zugeklappt. In Pivot-Tabellen werden Daten quer

und längs analysiert, Geschäftsgrafiken erstellt, eigene Gruppierungen erzeugt, ja sogar Daten auf Landkarten präsentiert. Dafür allein müssen keine Daten in einem speziellen Format gespeichert werden. Fairerweise muss man zugeben, dass auch BI-Tools anderer Hersteller über diese Funktionen verfügen.

2. Der nächste Punkt, den es zu betrachten gilt, ist die Modellierung des Geschäftsmodells. Im OLAP war und ist die Fähigkeit, ein Datenmodell bereitzustellen, das dem

Endanwender sein Geschäft in einem für ihn verständlichen Modell präsentiert, eines der mächtigsten Einsatzargumente überhaupt. Kein BI-System hatte in der Vergangenheit Erfolg, wenn es rein technisch die Spalten des darunterliegenden DWH ohne eine Geschäftsmodellierung präsentierte! Moderne BI-Tools verfügen deshalb über eine Metadatenschicht, die die Fähigkeit einer mehrdimensionalen Modellierung auf beliebigen Vorsystemen bietet (manche BI-Tools auch Vorkomplett-übergreifend). In Oracle BI EE kann heute eine unternehmensweite, fachliche Sicht auf und über allen Datenquellen hinweg definiert werden. Diese fachliche Sicht erreicht den Endanwender wie ein OLAP-Würfel, nur ohne die Notwendigkeit einer weiteren Datenspeicherung. Es muss aber auch betrachtet werden, dass die rein logische Datenmodellierung spätestens ein Ende findet, wenn umfangreiche Integrationsleistung erbracht werden muss oder wenn die fachlichen Regeln eine Komplexität erreichen, die im Online-Zugriff ohne die notwendige Unterstützung in der Datenhaltung nicht mehr sinnvoll abgefragt werden können. Auch aus diesem Grunde ist es ab einer gewissen Komplexität nützlich, wenn die Datenhaltung die notwendige Performance nicht nur in Form von schnellem Datendurchsatz, sondern auch in Form von unterstützten Funktionen (Logik zu den Daten) und in Form der Modellierung bereitstellt. Optimal ist es sicher, wenn letztere dann mit wenig Aufwand direkt in das BI-Tool übernommen werden kann, um nur noch Ergänzungen vornehmen zu müssen.

3. Nachdem zwei bedeutende Aspekte der reinen Präsentation für den Endanwender besprochen wurden, hier der vielleicht wichtigste: Kein BI-System wird seinen vollen Vorteil ausspielen, wenn die Endanwender mit der Antwortgeschwindigkeit nicht zufrieden sind. Die Mächtigkeit der Ad-hoc- oder OLAP-Analyse liegt in dem schnellen Erkennen von Zusammenhängen, nicht nur

in der technischen Machbarkeit eines „Drill“. Antwortzeiten unter fünf Sekunden sind unbedingt anzustreben, besser noch schneller. Und hier liegt das Problem vieler BI-Tools ohne eigene Datenspeicherung: Sie müssen sich auf die Geschwindigkeit des Vorkomplett verlassen. Insbesondere bei komplexer Geschäftsmodellierung, wie sie etwa in Oracle BI EE möglich ist, muss man sich die Frage stellen, ob das, was technisch möglich, auch sinnvoll ist: Die Abfragen gegen die Datenquelle werden beliebig komplex und somit wird es auch schwerer, eine gleichbleibende Antwortgeschwindigkeit für den Endanwender sicherzustellen. Hier schließt sich auch der Kreis in Bezug auf die Modellierung. Ist diese im Vorkomplett entsprechend vorbereitet, durchgeführt und eventuell sogar durchgerechnet, werden auch die Abfragen einfacher und damit schneller. Oracle BI EE geht hier bei Bedarf zusätzlich den umgekehrten Weg: Wenn im (relationalen) Vorkomplett die Performance nicht ausreichend ist, kann aus der Modellierung, wie sie in BI durchgeführt wurde, eine passende Persistenz-Schicht in unterschiedlichen Technologien (relationales Star-Schema, Oracle OLAP Analytic Workspace oder Essbase Cube) erzeugt werden, um beste Performance sicherzustellen. Natürlich gibt es hier zwischen den Anbietern der BI-Tools eklatante Unterschiede in der Qualität der generierten Abfragen (sowohl SQL als auch MDX bei den Tools, die es betrifft): An dieser Stelle soll dieses nicht diskutiert werden, aber es liegt auf der Hand, dass man insbesondere von den Tools, die von Datenbank-Herstellern entwickelt und vertrieben werden, wohl erwarten kann, dass sie gut auf die betreffenden Datenbanken abgestimmt sind.

Zusammenfassend für das BI kann man damit sagen, dass sich das BI-Tool auf die Darstellung und Modellierung der Endbenutzersicht konzentriert, aber seine Performance größtenteils aus dem Vorkomplett zieht.

Überlegung für die Datenhaltung

Wie in Abbildung 2 auf der linken Seite gezeigt, muss die Datenhaltung ihren Teil der OLAP-Anforderungen abdecken, ansonsten kann das BI-Tool seine volle Leistungsfähigkeit nicht ausspielen. Dazu kommen die Anforderungen, die auch schon beim BI-Tool diskutiert wurden: Performance und Modellierung. Beides kann man nicht voneinander getrennt betrachten, da, spätestens im Produktivbetrieb, die Modellierung einen entscheidenden Einfluss auf die Performance hat, insbesondere, wenn es um die fachliche (Weiter-)Entwicklung des DWH und die Kopplung an das BI-Tool geht:

- *Einfluss der Modellierung auf die Entwicklung*

Natürlich kann man argumentieren, die fachliche Modellierung würde nun komplett in der Metadatenschicht des BI-Tools durchgeführt werden. Untersuchungen zeigen jedoch Folgendes:

- Nicht alle Anforderungen (z.B. an Integration) lassen sich überhaupt zur Laufzeit realisieren.
- Eine beliebige Datenspeicherung, selbst auf einer performanten Plattform, wird auch beliebig langsam werden, wenn zu komplexe Abfragen an das falsche Datenmodell gestellt werden. Es gibt populäre Beispiele dafür.

Aus diesem Grund ist es weiterhin eine gute Option, die fachliche Modellierung bereits in der Datenhaltung vorzunehmen. Die Oracle-Datenbank bietet dafür gleich mehrere Möglichkeiten:

- * *Die klassische Modellierung eines Star- oder Snowflake-Schema*

Dem BI-Tool sollte es heute egal sein, in welcher Form die Modellierung vorliegt. Sollte diese mit den Oracle-ETL-Werkzeugen erstellt worden sein, können die Modelle in die Metadatenschicht von Oracle BI EE übernommen werden.

- * *Die Modellierung eines Analytic Workspace (AW) innerhalb der Oracle-Datenbank*

AW sind Oracle OLAP Cubes innerhalb der Oracle-Datenbank; mit ihnen ist eine fachliche Modellierung einschließlich komplexer Kennzahlen heute immer noch einfacher als mit rein relationalen Mitteln. Auf AW wird auch per SQL zugegriffen, bei der Definition von Cubes werden automatisch „Cube Views“ angelegt, auf die jedes relationale Werkzeug zugreifen kann. Diese Cube Views enthalten die fachliche Modellierung, ähnlich wie bei einem Star-Schema.

- **Modellierung und Performance**

Wie oben schon beschrieben, hat die Modellierung einen entscheidenden Einfluss auf die Performance. Zusammen mit den Möglichkeiten der Database Machine ergeben sich Schlussfolgerungen, die eventuell anderen Argumenten zuwiderlaufen. Bei vorhandener Performance (= Database Machine oder anderweitig optimierte Datenbank) wird die Leistungsfähigkeit der Maschine den Mangel an analytischer Intelligenz wettmachen:

- **Modellierung von Star- oder anderen relationalen OLAP-Modellen**

Je weniger analytische Intelligenz benötigt wird oder je mehr Leistung diesen Mangel wettmacht, desto mehr kann dieser klassische Ansatz der richtige sein. Wenn die Modellierung außerdem in einem Oracle-ETL-Werkzeug erfolgt, kann diese an Oracle BI EE weitergegeben werden. Aktuelle Kundenpräsentationen

weisen für einzelne Abfragen beim Einsatz der Database Machine Performancevorteile von über 500 Prozent aus.

- **Multidimensionale Modellierung in der relationalen Datenbank (Oracle OLAP, Analytic Workspaces):**

Diese Technologie bietet eine höhere analytische Intelligenz durch die native Modellierung in Dimensionen und Kennzahlen. Wenn die „Analytic Workspaces“ durch das entsprechende Oracle Tool („Analytic Workspace Manager“, AWM) erstellt werden, kann das Modell in Oracle BI EE importiert werden. Diese Art der Vorgehensweise wird durch seine Technologie auch bei leistungsschwächerer Plattform performant sein, in der Kombination mit der Database Machine ergibt sich allerdings durch die Integration in die Datenbank auch hier ein enormer Performance-Gewinn. Tests ergaben eine deutliche (etwa doppelt so schnelle) Steigerung der Performance beim Laden einzelner Cubes, beim Lesen eine noch stärkere Verbesserung (rund achtmal so schnell). Dabei ist bemerkenswert, dass alle OLAP-Funktionen voll unterstützt werden, also auch die für SQL eher komplizierten Abfragen einfach bleiben, nur schneller werden.

Fazit

Es gibt heute für die reine Performance- und Frontend-Betrachtung von OLAP keine technische oder fachliche Notwendigkeit mehr für separa-

te OLAP- oder „in-Memory“-Technologien:

- Die Frontend-Anforderungen lassen sich inzwischen durch die führenden BI-Tools am Markt auf allen möglichen Datenquellen realisieren.
- Die Modellierung sollte zwischen BI-Tool und Datenhaltung aufgeteilt werden:
 - Eine Modellierung rein im BI-Tool wird auch auf performanten Plattformen zu komplexen und damit exponentiell langsameren Abfragen führen. „Database Machine“ wird diese Grenze verschieben.
 - Eine Modellierung sollte bereits in der Datenhaltung erfolgen:
 - * Klassische Ansätze (wie Star-Schema) sind weiterhin gültig, die Performance einer Database Machine wird die fehlende analytische Intelligenz in Maßen ausgleichen.
 - * OLAP in der Datenbank vereinfacht die Modellierung, bietet mehr analytische Intelligenz und profitiert zusätzlich von der Leistungsfähigkeit der Database Machine.
 - * Die Modellierungen aus der Datenhaltung sollten in die Modellierung des BI-Tools übernommen werden können.
 - * Mit zunehmender Performance der Datenbank-Plattform wird die Komplexität der fachlichen Modellierung die Technologie beeinflussen.

Kontakt:

Marc Bastien

marc.bastien@oracle.com

Impressum

Herausgeber:

DOAG Deutsche ORACLE-Anwendergruppe e.V.
Tempelhofer Weg 64, 12347 Berlin
Tel.: 0700 11 36 24 38
www.doag.org

Verlag:

DOAG Dienstleistungen GmbH
Fried Saacke, Geschäftsführer
info@doag-dienstleistungen.de

Chefredakteur (VisDP):

Wolfgang Taschner,
redaktion@doag.org

Chefin von Dienst (CvD):

Carmen Al-Youssef,
office@doag.org

Titel, Gestaltung und Satz:

Claudia Wagner,
DOAG Dienstleistungen GmbH

Anzeigen:

CrossMarketteam Ralf Rutkat, Doris Budwill
www.crossmarketteam.de

Mediadaten und Preise finden Sie unter:
www.doag.org/publikationen/

Druck:

adame Advertising and Media
GmbH Berlin
www.adame.de