

High Performance Datawarehouse Analyse – die Oracle OLAP Option 11g im Vergleich

Stefan Vogel
FRT Consulting
Graz, Österreich

Schlüsselworte

Oracle, OLAP, analytische Funktionen, Aggregate, Materialized Views, Cube Organized Materialized Views, Cube Refresh, Simba MDX,

Einleitung

Im folgenden Manuskript wird die Oracle OLAP Option 11g als Alternative vorgestellt, wenn es darum geht, in einem bestehenden Oracle Datawarehouse mit vergleichsweise wenig Aufwand bezogen auf Hardware- und Entwicklungsressourcen zum einen eine beträchtliche Performance-Steigerung von bestehenden Abfragen zu erhalten und zum anderen, auf Funktionalitäten eines OLAP-Werkzeuges wie z.B. analytische Funktionen zugreifen zu können. Im ersten Teil werden die verschiedenen Lösungsansätze einer solchen Aufgabenstellung miteinander verglichen, im zweiten Teil wird die Technologie der Oracle OLAP Option im Detail betrachtet. Der dritte Teil beschäftigt sich mit dem Zugriff auf die OLAP Option um dann im vierten und letzten Teil, die Oracle OLAP Option der Oracle Essbase Lösung gegenüber zu stellen.

Die Architektur

In einer typischen DWH-Architektur sind die Daten in einem Sternmodell gespeichert: die Kennzahlen liegen atomar in der zentralen Faktentabelle vor. Die abhängigen Entitäten werden als Dimensionen referenziert.

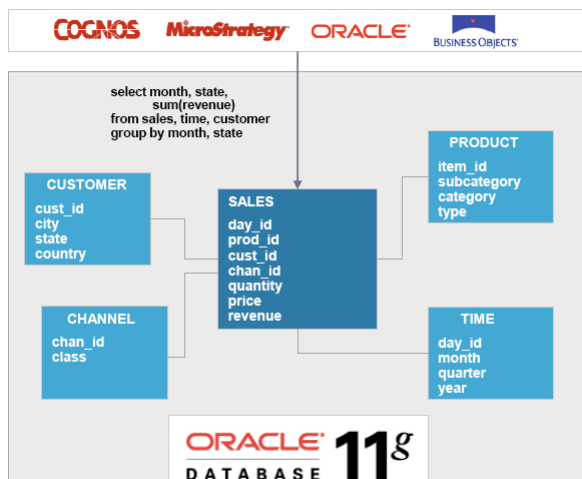


Abbildung 1 typische Architektur mit Oracle Datenbank als Data Warehouse

In den Dimensionen sind letztendlich auch Geschäftsprozesse hinterlegt, die es erlauben Informationen zu bestimmten Gruppen zusammenzufassen. Über diese Hierarchien sind Beziehungen zwischen Daten kodiert. Diese Beziehungen sind fix so dass z.B. sichergestellt ist, dass ein Kunde immer zu einem bestimmten Kundentyp zuordenbar ist oder eine Anlage immer zu einem bestimmten Anlagentyp gehört. Häufig sind auch die Organisationsstrukturen eines Unternehmens in einer Parent-Child-Hierarchie abgebildet.

Da Abfragen zumeist nicht auf der niedrigsten Detailebene erfolgen, ergeben sie sich auf der Datenbank als Summenabfragen, die die Kennzahlen über die verschiedenen Level der Dimensionen aggregieren.

Auch wird die Kennzahl in vielen Fällen in Abhängigkeit von weiteren Merkmalen berechnet; so z.B. als Abweichung zum Vormonat oder Vorjahr, oder als Soll-/Ist-Vergleich zu den hinterlegten Planzahlen oder einfach als laufende Summe des aktuellen Kalender- oder Geschäftsjahres.

Die Daten liegen jedoch auf Detailebene vor. Um zum gewünschten Ergebnis zu kommen, sind Aggregationen und unter Umständen auch Kalkulationen notwendig. Diese werden dann zur Laufzeit ausgeführt und können entsprechend ressourcenintensiv sein.

Um den Anforderungen von kurzen Antwortzeiten nachkommen zu können, gibt es verschiedene Lösungsansätze:

Vorausberechnung und Speicherung in physischen Tabellen

Vorausberechnung der verschiedenen analytischen Funktionen und Speicherung der Ergebnisse auf den verschiedenen Aggregationsebenen in separaten physischen Datenbanktabellen. Diese Methode ist sehr nahe liegend, da damit die gewünschten Daten in der gewünschten Form vorliegen. Und sie sind in derselben Datenbasis gespeichert und somit in den ETL-Prozess integriert. Der Single-Point-of-Truth bleibt erhalten. Und die bestehenden Werkzeuge können wie gewohnt per SQL-Abfragen auf die jetzt vorbereiteten Daten darauf zugreifen. Allerdings ergeben sich oftmals nicht unerhebliche Entwicklungsaufwände der neuen Aggregate, da jede Kennzahl für die verschiedenen Aggregationsebenen berechnet und damit entsprechend ausprogrammiert werden müssen. Hinzukommt, dass die Kalkulation der so neu entstandenen Aggregate ebenfalls eine gewisse Zeit in Anspruch nimmt und das kann sich bei den doch eher knapp bemessenden Zeitfenstern zur Aktualisierung der Datenbasis als hinderlich erweisen.

Verlagerung der analytischen Funktionen in den BI-Server

Eine weitere Möglichkeit ist die Kalkulation der analytischen Funktionen in den BI-Server zu verlegen. Damit ist zwar eine zeitnahe und weniger aufwendigere Entwicklung gegeben, allerdings muss die Berechnung auf den verschiedenen Aggregationsebenen immer noch zur Laufzeit erfolgen. Das ist vielleicht zu wenig, um ein Abfrageergebnis in der erforderlichen Zeit zu erhalten. Auf jeden Fall sind die Kennzahlen dann nur mehr in diesem BI-Server verfügbar; d.h. man hat sich auf dieses BI-Tool festgelegt und die Verwendung weiterer BI-Server bedeutet diese analytischen Funktionen dort ebenfalls entwickeln zu müssen. Auch kann es u.U. eine Herausforderung sein, die verschiedenen Aggregationstabellen mit den verschiedenen analytischen Funktionen zu matchen.

Verwendung eines eigenen OLAP-Servers

Um diesem Dilemma zu entgehen, kann ein eigener OLAP-Server verwendet werden. OLAP-Server sind speziell dazu geschaffen, analytische Funktionen auf unterschiedlichsten Aggregatsstufen über die verschiedenen Dimensionen voraus zu berechnen. Sie liefern auch für die komplexesten Kennzahlen zeitnahe Antworten. Somit sind Performance der analytischen Funktionen und die Verwaltung der unterschiedlichen Aggregatsstufen gelöst. Allerdings ist der OLAP-Server ein eigenes System und in der Regel auch eine eigene Technologie, die gelernt werden will. Als eigenes System bringt es eigene technische Voraussetzungen mit (Anforderungen an Hardware, Betriebssystem, etc.) und es bedarf eigener Trainings und Mitarbeiter, die diese Technologie beherrschen. Da es sich um eine andere Technologie handelt, sind unter Umständen die bestehenden Abfragetools nicht mehr zu verwenden, da die Abfragesprache jetzt MDX heißt und nicht mehr SQL. Jedenfalls ist der OLAP-Server eine separate eigenständige Datenquelle und damit gibt es keinen Single-Point-of-Truth mehr.

Oracle OLAP als integrierte Lösung

Oracle OLAP scheint in diesem Umfeld eine Alternative, die alle Herausforderungen meistert. Als integrierter Bestandteil der Oracle Datenbank nutzt Oracle OLAP dieselben Ressourcen. Die Daten werden ebenfalls in denselben Datendateien gespeichert, die auch von der Oracle Datenbank genutzt werden. Es können dieselben Sicherheitszugriffsrichtlinien verwendet werden und ist vollständig

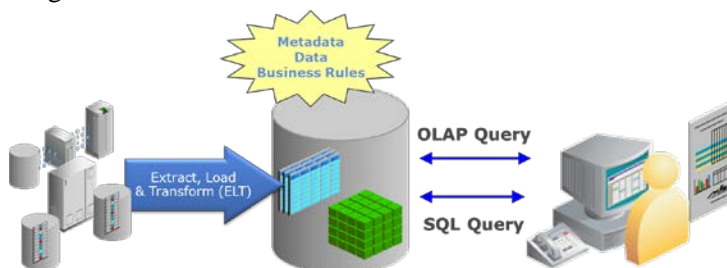


Abbildung 2 Ein integrierter Würfel auf den mittels SQL und MDX zugegriffen werden kann

kompatibel mit Real Application Cluster und Grid Computing.

Oracle OLAP bietet eine große Vielzahl von analytischen Funktionen an, die zur Vorausberechnung auf die Aggregate angewendet werden können. Die Aggregatsbildung ist sehr variabel und kann sich zwischen den verschiedenen Dimensionen unterscheiden. Die Datenmenge, die zur Vorausberechnung verwendet werden soll, kann bestimmt werden, sodass ein Optimum zwischen Zeit für die Vorausberechnung und Antwortzeit geschaffen werden kann. Auch finden verschiedene Aggregationsverfahren (Summe, Maximum, Minimum, etc.) Verwendung. Die Daten können per SQL und MDX abgefragt werden. Als integrierter Bestandteil der Oracle Datenbank kann mit der OLAP Option auch das Prinzip Single-Point-of-Truth beibehalten werden.

SQL-Abfragen auf den Würfel

Abfragen von OLAP-Cubes mit SQL ist einfach. Die Aggregationen und Rechenregeln sind im Würfel eingebettet und werden innerhalb des Würfels ausgeführt. Die Anwendung kann Kennzahlenabfragen und dimensionale Abfragen stellen, indem die relevanten Spalten in der zur Verfügung gestellten View ausgewählt werden.

Werden Dimensionen oder Würfel in der OLAP Option angelegt, stehen damit nämlich auch Views für den direkten Zugriff auf die Metastrukturen zur Verfügung. Mit diesen Views kann per SQL

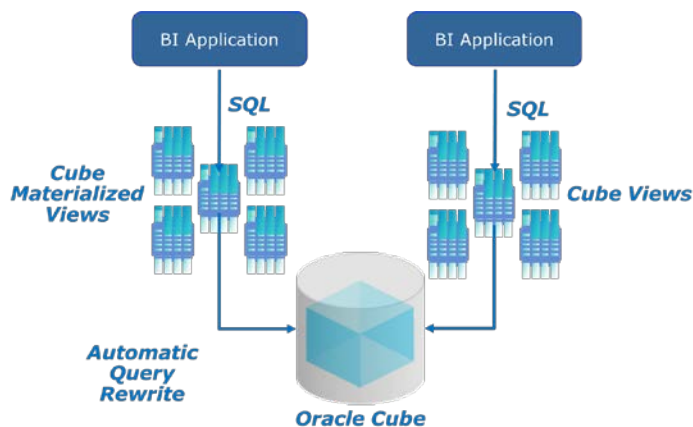


Abbildung 3 Zugriff auf den Oracle Cube mittels zur Verfügung gestellter Views

entweder auf die Spalten und/oder auf die Hierarchien zugegriffen werden. Der Zugriff auf den Würfel selbst erfolgt über eine Cube-Organized-Materialized-View. Dahinter verbirgt sich das Management von so vielen Materialized views wie es für die Aggregation der verschiedenen Kennzahlen (plan und unter Verwendung der analytischen Funktionen) notwendig ist. Wird eine Abfrage gegen den Würfel gestellt, wird über die Oracle OLAP Technologie die abgefragten Aggregationseben automatisch erkannt.

Ebenfalls ist es möglich, die Menge der vorausberechneten Kennzahlendaten zu bestimmen, um somit ein gutes

Verhältnis zwischen Rechenzeit und Antwortzeit zu erzielen.

Eine zusätzliche Funktion bietet die Möglichkeit, das Query Rewrite einzuschalten. Ist dieses aktiviert, entscheidet der Optimizer bei einer Abfrage, die zwar auf eine Faktentabelle geht, aber von den Daten und Aggregationsstufen im Würfel enthalten ist, dass das Select auf den Würfel umgeleitet wird. Das Ergebnis ist eine signifikant kürzere Antwortzeit. Damit erfahren durch den Bau von Würfeln mit Oracle OLAP bestehende Abfragen eine Performanceverbesserung, ohne dass an den Abfragen etwas geändert werden muss.

Neben der Möglichkeiten die Daten und Strukturen des Würfels mittels SQL abzufragen besteht darüber hinaus mittels einem Konnektor der Firma Simba den Würfel ebenfalls mit MDX abzufragen. Somit steht der Würfel nicht nur den SQL-Abfragetools sondern auch den OLAP-Tools zur Verfügung, die nur MDX verstehen.

Cube Materialized Views

Kernstück der Oracle OLAP Option sind die Cube Organized Materialized Views. Materialized Views werden bereits seit der Version Oracle 8i zur Verfügung gestellt. Mit ihnen gelingt es die Antwortzeiten bei Abfragen gegen die Oracle Datenbank mit wenig Aufwand signifikant zu verkürzen. Wie der Name sagt, sind MVs Views, die die selektierten Daten jedoch in der

gleichnamigen physischen Tabelle persistieren. Über ebenfalls zur Verfügung gestellte DBMS-Packages können die MVs dann aktualisiert – sprich neu gerechnet - werden. Ein besonderes Feature ist vor allem Query Rewrite. Ist diese Funktionalität eingeschaltet prüft der SQL-Optimizer anhand der Kosten ob eine Abfrage sinnvollerweise auf die MVs umgeschrieben werden kann. Der Anwender braucht dafür keine Anpassungen am bestehenden Code vorzunehmen. Voraussetzung ist also, dass die Abfrage genau die Attribute und den Aggregationslevel der MV trifft. Um möglichst viele Treffer zu erhalten kann so eine erhebliche Anzahl von MVs entstehen, die entsprechend organisiert werden müssen. Die Cube Organized Materialized Views stellen jetzt genau die Sammlung an Materialized Views zur Verfügung, die notwendig sind, damit der Würfel mit all seinen Dimensionsebenen und Hierarchien und den analytischen Funktionen und Aggregationsebenen abgefragt werden kann.

Zugriff mittels Abfragewerkzeug

Über jedes beliebige Abfragewerkzeug, das entweder SQL oder MDX versteht, können jetzt die

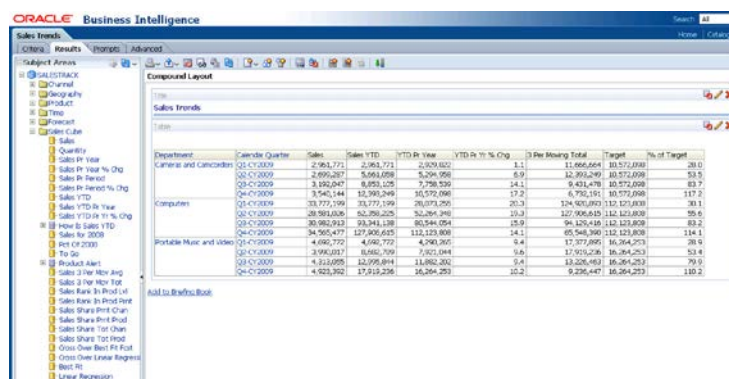


Abbildung 4 Beispiel einer SQL-Afrage mit OBIEE

Würfel in der Oracle OLAP Option abgefragt werden. Die Dimensionen und ihre Attribute als auch die Kennzahlen, hinter denen sich die oftmals komplexen analytischen Funktionen verbergen, stehen als Metadaten zur Verfügung und können von den Werkzeugen individuell selektiert werden. Mittels dem SIMBA MDX-Konnektor ist auch Excel ein mögliches Abfragewerkzeug, über das sowohl die dimensional Strukturen als auch die Kennzahlenaggregate z.B. in

Pivottabellen dargestellt werden können.

Eine Alternative zur Datenbank-integrierten OLAP Lösung: Oracle Essbase

Während die OLAP Option als in die Oracle Datenbank integrierte Lösung alle damit verbundenen Vorteile nutzt, gibt es darüber hinaus Anforderungen an eine OLAP-Lösung, die damit nicht erfüllt

Feature	Oracle OLAP 11g	Oracle Essbase
Storage Method	Multidimensional Arrays	Multidimensional Arrays
Query language	SQL, via query rewrite or SQL views MDX via Simba Plugin	MDX and XML/A
Process Type	In-database, embedded process	Stand-alone server
Product dependency	Oracle Database Enterprise Edition	Any relational database
Primary use	Enhancing SQL-based Data Warehouses	Supporting OLAP analytical applications

werden können (z.B. Planungskapazitäten, Forecast-Analysen, etc.). Für diese Fälle oder auch in dem Fall, dass nicht die Oracle Datenbank die zentrale Datenbasis ist, stellt sich Oracle Essbase als eine weitere Alternative dar. Oracle Essbase ist eine eigenständige OLAP-Engine und damit das zuvor vorgestellte dritte Szenario. Als Standalone OLAP-Server ist es ein

vollständig auf die Anforderungen von multidimensionalen Analysen, Planungen, Vorausberechnungen ausgerichtetes separates System und insofern auch eine eigene Technologie, die zwar kein SQL, dafür aber neben MDX auch XML/A versteht.

Kontaktadresse:

Stefan Vogel
FRT Consulting GmbH
Liebenauer Hauptstraße 2-6
A-8041 Graz
Österreich

Telefon: +43 (0) 316-71 12 12
Fax: +43 (0) 316-71 12 12 - 99
E-Mail: stefan.vogel@frt.at
Internet: www.frt.at