

# BI in der Cloud: Lohnt sich das?

**Jan Schreiber**  
**Loopback.ORG GmbH**  
**Hamburg/Frankfurt/Berlin**

## **Schlüsselworte**

DWH; BI; Cloud; Architektur; Migration

## **Einleitung**

Welche Cloud-Angebote für DWH und BI gibt es in der Oracle Welt und wie kann man sie nutzen? OTBI, BICS, Database (Exadata) Cloud Service, PaaS und IaaS - wie hängen diese Angebote zusammen, wie kann man sie nutzen, welche benötigt man wirklich und für welche Anwendungsfälle lohnt es sich, die Cloud Angebote einer konventionellen On-Premises-Lösung vorzuziehen?

## **Ausgangssituation**

Anfang des Jahres durften wir ein SAP-BW-Ablöse-Projekt einer größeren Klassifikationsgesellschaft mit Sitz in Oslo und Hamburg begleiten. Das Unternehmen war zuvor aus einem deutschen und einem norwegischen Stamm hervorgegangen. Die deutsche Sparte brachte eine etablierte SAP-BW-Umgebung mit, in der das Reporting- und Business-Intelligence-Geschäft erfolgreich betrieben wurde. Im Rahmen der Konsolidierung sollte die BI-Architektur nun aber international vereinheitlicht werden. Als Technologie sollte das in Norwegen bereits erfolgreich eingeführte Oracle BI verwendet werden.

Das Unternehmen fährt eine explizite Cloud-Strategie. Agilität, also schnelle Bereitstellung, kombiniert mit einem kleinen Fußabdruck in Sachen Administrationskosten und Hardware-Bedarf sind die wesentlichen Treiber dafür, ein neues Projekt, wenn möglich mit Cloud-Lösungen, zu entwerfen. Auch die transparente Skalierbarkeit spielt hier eine Rolle: Lösungen sollen als Prototyp entworfen werden können, um anschließend in Betrieb genommen und unternehmensweit ausgerollt zu werden, ohne dass ein neues technisches Architekturkonzept erstellt und implementiert werden muss.

Oracle-Cloud-Lösungen, insbesondere HCM, die Human Resources Lösung von Oracle, waren in einem der beiden Unternehmensteile bereits vorhanden. Der Hersteller hatte den IT-Verantwortlichen des Kunden auch bereits die Oracle BI Applications vorgestellt, eigenständige, schnell installierbare Mini-Lösungen, mit denen sich sehr schnell Einsichten aus den HR-Daten gewinnen lassen sollten. Für den Umzug der bisherigen HCM On-Premise Installation in die Oracle HCM Cloud gab es bereits eine Lösungsskizze. Und für die Gewinnung der wichtigsten Erkenntnisse aus den bisherigen Reports über BI Apps existierten erste Ansätze.

Im Rahmen der weiteren Konsolidierung und der Gedanken, die sich der Kunde zur Ablösung der bisherigen SAP-BW-Berichte machte, kamen nun aber eine Fülle von Anforderungen dazu. Außerdem reifte die Erkenntnis, dass noch lange nicht alle Anforderungen an ein zukünftiges BI-Reporting überhaupt erfasst waren. Im Haus existierte nämlich noch kein vollständiger Katalog der SAP-BW-Anwendungen, insbesondere waren die Nutzergruppen und die Relevanz der bestehenden Berichte nicht immer bekannt.

## **Transaktionales BI**

Als "DWH-Baukasten" sind die Oracles BI Apps (OBIA) ein sehr interessantes Angebot, um schnell Daten aus einer Anwendung mit den Methoden der Business Intelligence aufzubereiten und so darzustellen, wie man es von Oracles BI-Lösungen gewöhnt ist - inklusive Dashboards und interaktiven Analysen. Für Oracle Apps, wie zum Beispiel HCM, existieren fertige Lösungen, die die in der Regel gewünschten Ergebnisse bereits fertig aufbereitet präsentieren. Selbst die fachliche Logik,

also das Wissen darum, welche für die Analyse notwendigen Zahlen in welchen Tabellen der Anwendung verborgen sind, wird bereits mitgeliefert. So spart sich der Kunde die eigene Entwicklung der gesamten Wertschöpfungskette des klassischen Data Warehouse: Quellsystem-Analyse, Datenmodellierung, ETL und Erstellung von Dashboards und Reports.

Neben OBIA existiert eine weitere Produktreihe namens Oracle Transactional Business Intelligence (OTBI) - transaktional, weil die BI-Elemente direkt aus den operativen VORSYSTEMEN, den transaktionalen Anwendungen gespeist werden. Wenn schnelle Analysen direkt aus den Oracle Applications benötigt werden, ist OTBI eine interessante Lösung: schnell, schlank und trotzdem ist OTBI dem klassischen BI im OBIEE ebenbürtig.

Im Rahmen der weiteren Anforderungsanalyse wurde jedoch deutlich, dass die Möglichkeiten von OBIA und OTBI für die vollständige Lösung wahrscheinlich nicht ausreichen würden. Im Rahmen der Konsolidierung war nun doch das Bedürfnis nach unternehmensweiten Zahlen offenbar geworden - insbesondere nach Analysen, die unterschiedliche Unternehmensbereiche zueinander ins Verhältnis setzen konnten. Diese Erweiterung um zusätzliche Kennzahlen schien weder mit OBIA noch mit OTBI belastbar umsetzbar zu sein.

### **BI Cloud Services**

Neben OBIA und OTBI unterhält Oracle ein weiteres Cloud-basiertes Angebot: Business Intelligence Cloud Services. Dies ist eine Alternative zur Installation des Oracle Business Intelligence (OBIEE) Servers im eigenen Haus. Im Gegensatz zum OTBI werden hier keine "Convenience-Rezepte" verarbeitet, sondern es steht ein Großteil der Funktionalitäten des OBIEE zur Verfügung, das mit einem cloud-basierten Datenbank-Schema verbunden werden kann, in dem die Daten dimensional aggregiert werden können.

Zur Anpassung an die Cloud-Umgebung hat Oracle vor allem die Funktionalität des Admin Tools in eine neue Cloud-Oberfläche integriert. Das traditionelle Admin Tool läuft auf einem Windows-PC. Hier wird das Repository erstellt: den physischen Datenbankobjekten werden logische Entitäten gegenübergestellt, die beispielsweise Tabellen durch Joins über entsprechende gemeinsame Attribute verbinden. Schließlich wird das für den Benutzer sichtbare "Universum" an geschäftlichen Relationen definiert. Diese Funktionalität wird mit Hilfe des neuen Cloud Admin Tools weitgehend vom Windows Fat Client und seiner in die Jahre gekommenen Optik befreit. Allerdings ist der Funktionsumfang auch nicht identisch. Die vollständige Definition der logischen und physischen Schicht muss auch in BICS weiterhin im Admin Tool erfolgen - kann dann aber "in die Cloud" hochgeladen werden, dieser Ansatz wird auch als "lift and shift" bestehender OBIEE-Repositories bezeichnet.

Wie ein traditionelles BI-Tool muss auch BICS mit den darzubietenden Daten bestückt werden. In der physischen Schicht des OBIEE könnte man natürlich transaktionale Daten der VORSYSTEME direkt bezeichnen. Nun ist es aber so, dass die Data Warehouse Community ihre Referenzarchitektur nicht ohne Grund entworfen hat: Performance, Konformität und Konsistenz sind nur einige Schlagwörter, die eine Architektur bezeichnen, die sich im Laufe der Jahre als zwar nicht ganz ohne Aufwand an Modellierung, ETL und Speicherplatz realisieren lässt, die aber eben doch in den allermeisten Fällen auf längere Sicht Zeit und Aufwand spart und die entscheidenden Ergebnisse realistisch liefern kann.

### **DWH Platform as a Service**

Ein klassisches, vollständiges Data Warehouse kann Daten aus verschiedenen Quellen annehmen, zwischenspeichern und in ein gemeinsames Datenmodell integrieren, welches über die Zeit historisiert - oft weiter zurückreichend als die Quellsysteme - und konsolidiert, also Redundanz und Widerspruch zwischen Daten verschiedener Quellen vermeidet. Des Weiteren kann es die konsolidierten Daten - in der Regel als dimensionales Modell - für die Auswertung mit BI-Tools vorbereiten. Um das zu erreichen, werden in der Regel verschiedene Schichten angelegt: Die Stage, der Core und ein oder mehrere Data Marts. Die Daten werden auseinandergenommen, meist komplett transformiert und in

das dimensionale Modell geladen. Dafür werden eine Datenbank zur Ablage der jeweiligen Datenschichten und ein Extract-Transform-Load (ETL)-Tool benötigt.

Selbstverständlich bietet Oracle ein umfangreiches Datenbank-Lineup in ihrer Cloud an. Der Hersteller fährt eine klar kommunizierte Cloud-First-Strategie, neue Datenbank-Versionen erscheinen mittlerweile zuerst in der Cloud. Oracle Database Cloud Service (DBCS) ist in verschiedenen Varianten buchbar: Database as a Service (DBaaS), Database Schema Service (DBSS) und Database Exadata Cloud Service (ExaCS).

Für die Datenbank, in der ETL ablaufen und die eine vollständige DWH-Schichten-Implementation beherbergen soll, wurden verschiedene Szenarien evaluiert. DBCS, Exadata Cloud Service und eine eigenständige Installation in AWS. DBCS Exadata Express Cloud Service wurde wegen der Volumenbegrenzung nicht näher berücksichtigt. Bei genauerer Betrachtung wurde schnell deutlich, dass die Oracle Cloud Angebote nicht nur einfacher einzurichten sind als eine eigene Installation in einer generischen Public-Cloud-Umgebung, sondern auch wesentlich leistungsfähiger und damit auch günstiger sind. Insbesondere haben andere Cloud-Angebote meist den Nachteil, dass die Storage-Bandbreite nicht für eine DWH-Installation ausgerichtet ist, und die CPUs dann dementsprechend nicht ausgelastet sind. Hier bringt die Abstimmung zwischen Hard- und Software in der Oracle Cloud den entscheidenden Vorteil. Die Entscheidung zwischen Exadata Cloud Service und DBCS ist letzten Endes von der gewünschten Performance abhängig.

Eine on-premise Exadata-Installation birgt höhere Anfangskosten als das Bereitstellen des Exadata Cloud Service. Betrachtet man die Aufwendungen auf fünf Jahre, werden beide Varianten etwa gleich teuer, was Hardware und Lizenzen (nach Preisliste) betrifft. Werden alle Kosten berücksichtigt (Aufbau, Installation, Kühlung), ist das Cloud-Angebot günstiger.

Wie einfach und vor allem praktikabel ist es tatsächlich, DBCS als DWH-Datenbank mit eigener Modellierung zu verwenden? Denn dass Stage, Core und Datamart Layer eigenständig modelliert werden sollten, war mittlerweile klar. Zu groß waren die Vorteile, die sich aus einem applikations- und bereichsübergreifenden Enterprise Datawarehouse (EDWH) ergeben könnten. Aus den konsolidierten Daten sollten eigene, womöglich maschinell unterstützte Entscheidungen getroffen werden, außerdem sollten die aufbereiteten Daten des DWH und die Analysen der BI auch den Kunden angeboten werden können.

Mittlerweile stand eine Architekturvorstellung des EDWH im Raum, die eine persistierte Stage, einen Core Layer im Data Vault Modell und mehrere Data Marts umfasste. Als ETL bzw. ELT-Tools sollte der Oracle Data Integrator (ODI) zum Einsatz kommen, da hiermit auf bestehende Rezepte inklusive einer weitgehenden Automatisierung zurückgegriffen werden konnte, die den ursprünglich gewünschten agilen Charakter des Projektes unterstützte. Wie also erfolgt das ETL in der Cloud?

Es existieren verschiedene Wege, Daten aus Quellsystemen in ein DWH auf Basis der Cloud-Datenbank zu transportieren. DBCS liefert von Haus aus eine SQL-Developer-Anbindung, eine APEX-Maske, ein REST- und ein SOAP-Interface. Für Datenlieferungen von Vorsystemen sind darüber hinaus die Lieferung von Data-Pump-Daten über SCP sowie die direkte ETL-Anbindung des ODI über eine JDBC-Schnittstelle und entsprechende ODI-Knowledge-Module verfügbar.

### **Infrastructure as a Service**

An dieser Stelle wurde klar, dass Oracles Cloud-Strategie dem Geschäftsmodell vieler Anbieter in der Cloud ähnelte: zunächst wurden Angebote mit überschaubarer Komplexität in die Cloud verlagert, Produkte, die weitgehend uniform abbildbar sind und die mit möglichst wenig Anpassungen bei einer möglichst großen Kundenzahl erfolgreich einsetzbar sein könnten. Das trifft auf die BI Apps zu, auf fertige Rezepte für OTBI-Bausteine oder für eine Standard-Datenbank in der Cloud. Sollte jedoch ein komplexes EDWH in der Cloud betrieben werden, kamen die Software-as-a-Service (SaaS, zum Beispiel HCM in der Cloud) und Plattform-as-a-Service-Angebote (PaaS, zum Beispiel BICS) an ihre Grenzen. Denn eine "ODI Cloud" war zum Projektzeitraum nicht verfügbar.

Spätestens, wenn ein DWH mit mehreren selbstmodellierten Schichten, verschiedenen - auch noch nicht in der Cloud vorliegenden - Quellen und einer ETL aufgebaut werden soll, bewegt der Kunde

sich gezwungenermaßen in Richtung Infrastructure-as-a-Service (IaaS) und betreibt die Infrastruktur in der Cloud mit dem Betriebssystem als Grenzschicht. Letzten Endes hat man so lediglich eigene, virtuelle Maschinen bekommen, auf denen die DWH-Komponenten selbst installiert und betrieben werden müssen. In diesem Markt ist die Compute Cloud von Oracle vergleichbar, wenn nicht austauschbar, mit dem anderen Anbieter wie z.B. AWS. Es könnten einige der ursprünglichen Ziele realisiert werden, der Kunde müsste zwar keine eigene Hardware vorhalten und pflegen, aber der Aufwand für die Installation und Wartung der Software bliebe ihm nicht erspart.

### **Das Endresultat ist ein Gemischtwarenladen**

Der Prototyp der neuen BI-Landschaft besteht aus den Applikationen, die, wo verfügbar, in der Cloud verwendet werden: HCM, UCM, Field Service, Finance. Das DWH ist als Oracle Exadata Cloud Service geplant. ETL wird mit dem ODI ausgeführt, der als Infrastrukturkomponente in der Oracle Compute Cloud installiert ist. Die ODI Studio Clients laufen lokal oder auf einem Terminal Server in der Compute Cloud. Modelliert wird mit dem SQL Developer Data Modeler, der lokal installiert ist. Die Versionskontrolle erfolgt über ein in der Compute Cloud beheimatetes svn oder git. Die Datenbanken sind in einer ExaData-Cloud-Service-Quarter-Rack-Umgebung untergebracht, und als BI-Komponente dient BI Analytics Cloud Service. Daten aus bestehenden On-Premise-Systemen werden über einen ODI-Agenten geladen und teilweise in der Storage Cloud abgelegt.

Das größte Problem ist, dass selbst die Oracle Cloud Apps nicht immer über benutzbare Schnittstellen zum ODI oder zu DBCS verfügen, sondern darauf ausgelegt sind, Exporte als Dateien, die per Hand aus einer GUI erzeugt werden sollen, auszuführen. Werden solche Applikationen On-Premise betrieben, existiert meist die Möglichkeit, ETL direkt aus dem Datenmodell der Anwendung zu betreiben - bei den Cloud Apps ist die Datensinke der Anwendung ein geschlossenes System, in das nicht mehr hineingegriffen werden kann.

### **Fazit**

Die Anfang 2017 bereitstehenden Oracle-Cloud-BI-Angebote sind in erster Linie dort erfolgreich einsetzbar, wo es um schnelle, abgrenzbare Lösungen geht. Ein Enterprise Datawarehouse (EDWH) ist aber in den meisten Fällen kein einfach abgrenzbares Projekt. Der Gedanke, die komplette Business Intelligence in die Cloud zu verlagern, ist für Entscheider sehr attraktiv, aber nicht immer zufriedenstellend umsetzbar. Hier muss selbst konzipiert, entwickelt und installiert werden. Ob die Komplexität der Schnittstellen der unterschiedlichen Cloud-Produkte untereinander am Ende des Tages einfacher zu handhaben ist als technische Schnittstellen von on-premise betriebenen Systemen, bleibt individuell auszuloten. Die Gesamtsumme an Komplexität, sowohl technischer als auch vertraglicher Art, verringert sich durch den Cloud-Einsatz aber sicher nicht.

Und es gibt noch einen anderen Aspekt: Ein strategischer Vorteil eines EDWH ist, einen konsolidierten und vollständigen Datenpool jederzeit im Zugriff zu haben. Erfolgreiche Unternehmen möchten sich natürlich auf ihr Kerngeschäft konzentrieren und keine Unsummen für Anschaffung und Betrieb im Zweifel schnell veralteter IT ausgeben. Aber die Verteilung der eigenen Daten auf eine Vielzahl von gemieteten Cloud-Anwendungen birgt die Gefahr, eben diesen Vorteil aufzugeben. Es droht eine Zersiedelung des Datenbestandes, in der die Übersicht und vor allem auch die Hoheit über die eigenen Daten verloren gehen kann. Hier gilt es, den Konsolidierungsgedanken eines wohlmodellierten Data Warehouses nicht aus dem Blick zu verlieren.

### **Kontaktadresse:**

Jan Schreiber  
Loopback.ORG GmbH  
Hopfenstraße 26  
D-20539 Hamburg

Telefon: +49 (0) 40 2263236-10  
E-Mail: [js@loopback.org](mailto:js@loopback.org)  
Internet: <http://www.loopback.org/>  
Twitter: @JanLoopback