

Business News

DOAG Zeitschrift für die Anwender von Oracle Business- und BI-Lösungen



Business Intelligence

Reporting-Engine

Ein Einblick in die neue
OBIEE 12c

Seite 17

Einfach bedienbar

Managed Self Service BI
mit ODI und OBI

Seite 11

ODI 12c

Data Warehouse
from Scratch

Seite 22

Wir haben die smarten Lösungen

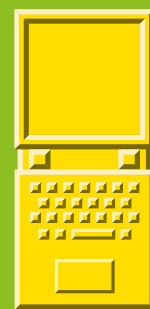
der Zukunft auf Lager

Erleben Sie die Weltneuheiten und Vordenker der Intralogistikbranche zuerst auf der

CeMAT 2016

31. Mai – 3. Juni 2016 ■ Hannover ■ Germany
cemat.de

Leitthema:
smart
supply chain
solutions



Deutsche Messe

World leading trade fair for intralogistics & supply chain management

CeMAT



Michael Klose
DOAG-Vorstand und Leiter DOAG
Business Intelligence Community

Liebe Leserinnen und Leser,

Sie wundern sich sicher, dass wir als Titelbild für den Schwerpunkt „Business Intelligence“ eine Wahrsagerin mit Glaskugel gewählt haben. In Wikipedia findet man zum Thema „Wahrsagen“: „Im Unterschied zu Prognostikern, die sich auf normale, für jeden grundsätzlich einsichtige Kausalzusammenhänge berufen, beanspruchen Wahrsager, ein den Unkundigen verborgenes Wissen über okkulte Zusammenhänge zu besitzen, das ihnen den Blick in die Zukunft ermögliche.“

Verglichen mit der Business-Intelligence-Welt stellt der klassische Analyst oder Report-Konsument sozusagen den Prognostiker dar. Dieser verlässt sich auf Key-Performance-Indikatoren, Detail-Analysen per Drill-Down oder Ad-hoc-Auswertungen.

In den letzten Jahren hat sich der BI-Markt durch immer leistungsfähigere Hardware wesentlich verändert. Waren vor zehn Jahren In-Memory-Lösungen oder Big-Data-Applikationen noch vollkommen unbekannt, werden sie mittlerweile in fast jedem Unternehmen in unterschiedlichen Ausbaustufen eingesetzt. Dadurch vollzieht sich ein Wandel in den Auswertewerkzeugen hin zu visueller Analyse. Die menschlichen Fähigkeiten des Erkennens von verborgenen Zusammenhängen und Auffälligkeiten lassen sich dadurch optimal nutzen.

Wir bieten Ihnen in dieser Ausgabe interessante Artikel zur Daten-Integration, zum klassischen Reporting bis hin zu visuellen Analysen und Data-Mining-Funktionalitäten. Die Wahrsagerin wird jetzt nicht mehr gebraucht?!

Ihr

Impressum

DOAG Business News wird von der DOAG Deutsche Oracle-Anwendergruppe e.V. (Tempelhofer Weg 64, 12347 Berlin, www.doag.org), herausgegeben. Es ist das User-Magazin rund um die Applikations-Produkte der Oracle Corp., USA, im Raum Deutschland, Österreich und Schweiz. Es ist unabhängig von Oracle und vertritt weder direkt noch indirekt deren wirtschaftliche Interessen. Vielmehr vertritt es die Interessen der Anwender an den Themen rund um die Oracle-Produkte, fördert den Wissensaustausch zwischen den Lesern und informiert über neue Produkte und Technologien.

DOAG Business News wird verlegt von der DOAG Dienstleistungen GmbH, Tempelhofer Weg 64, 12347 Berlin, Deutschland, gesetzlich vertreten durch den Geschäftsführer Fried Saacke, deren Unternehmensgegenstand Vereinsmanagement, Veranstaltungsorganisation und Publishing ist.

Die DOAG Deutsche Oracle-Anwendergruppe e.V. hält 100 Prozent der Stammeinlage der DOAG Dienstleistungen GmbH. Die DOAG Deutsche Oracle-Anwendergruppe e.V. wird gesetzlich durch den Vorstand vertreten; Vorsitzender: Dr. Dietmar Neugebauer. Die DOAG Deutsche Oracle-Anwendergruppe e.V. informiert kompetent über alle Oracle-Themen, setzt sich für die Interessen der Mitglieder ein und führen einen konstruktiv-kritischen Dialog mit Oracle.

Redaktion:

Sitz: DOAG Dienstleistungen GmbH (Anschrift s.o.)

Chefredakteur (ViSdP): Wolfgang Taschner

Kontakt: redaktion@doag.org

Weitere Redakteure (in alphabetischer Reihenfolge): Julia Bartzik, Mylène Diacquenod, Marius Fiedler, Marina Fischer, Michael Klose, Jan Peterskovsky, Fried Saacke, Dr. Frank Schönthaler

Titel, Gestaltung und Satz:

Elena Rankova, DOAG Dienstleistungen GmbH (Anschrift s.o.)

Fotonachweis:

Titel: © WavebreakmediaMicro/ fotolia.com

Foto S. 5: © leedsn/ 123rf.com

Foto S. 11: © Maxim Evseev/ 123rf.com

Foto S. 22: © enjoylife25/ 123rf.com

Anzeigen:

Simone Fischer, DOAG Dienstleistungen GmbH (verantwortlich, Anschrift s.o.)

Kontakt: anzeigen@doag.org

Mediadaten und Preise unter: www.doag.org/go/mediadaten

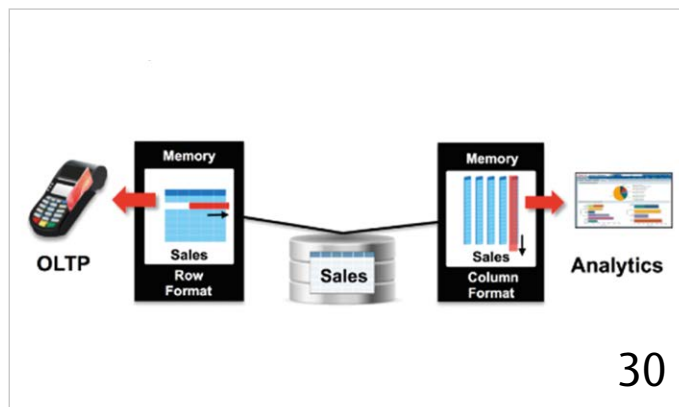
Druck:

adame Advertising and Media GmbH, Berlin, www.adame.de

Alle Rechte vorbehalten. Jegliche Vervielfältigung oder Weiterverbreitung in jedem Medium als Ganzes oder in Teilen bedarf der schriftlichen Zustimmung des Verlags. Die Informationen und Angaben in dieser Publikation wurden nach bestem Wissen und Gewissen recherchiert. Die Nutzung dieser Informationen und Angaben geschieht allein auf eigene Verantwortung. Eine Haftung für die Richtigkeit der Informationen und Angaben, insbesondere für die Anwendbarkeit im Einzelfall, wird nicht übernommen. Meinungen stellen die Ansichten der jeweiligen Autoren dar und geben nicht notwendigerweise die Ansicht der Herausgeber wieder.



Managed Self Service BI mit ODI und OBI



Business Intelligence In-Memory

3 Editorial	14 Business Intelligence und Visual Analyzer <i>Annett Thurm-Meyer</i>	27 Im BI Publisher steckt mehr als erwartet <i>Frank Effenberger und Janine Lehmann</i>
3 Impressum	17 OBIEE 12c: Ein Einblick in die neue Reporting-Engine von Oracle <i>David Michel</i>	30 Business Intelligence In-Memory <i>Annett Thurm-Meyer</i>
5 Business Intelligence: neue Technologien, neue Anwendungsbereiche <i>Ulrich Kramer</i>	21 DOAG 2016 Logistik – erstmals im Rahmen der CeMAT <i>Jan Peterskovsky</i>	33 Oracle Predictive Queries: Der einfache Weg zu neuen Erkenntnissen <i>Reinhard Mense und Marco Nätlitz</i>
9 Mit GOBIA zu zielgerichteten Business-Intelligence-Architekturen <i>David Fekete und Gottfried Vossen</i>	22 Data Warehouse from Scratch – mit ODI 12c <i>Bernhard Rosenberger</i>	36 Einführung in Data-Mining mit analytischen Funktionen und R <i>Vladimir Poliakov</i>
11 Managed Self Service BI mit ODI und OBI <i>Michael Klose</i>		



Data Warehouse auf der grünen Wiese aufgebaut

Unsere Inserenten

CeMat www.cemat.de	U 2	Proadvise www.proadvise.de	U 4	Promatis software GmbH www.promatis.de	S. 15
E-3 Magazin(B4Bmedia.net) www.b4bmedia.net	U 3	areto consulting GmbH www.areto-consulting.de	S. 7		



Business Intelligence: neue Technologien, neue Anwendungsbereiche

Ulrich Kramer, areto consulting gmbh

Business Intelligence wird auch im Jahr 2016 kräftige Wachstumsraten verzeichnen und einen seit Jahren anhaltenden Trend fortsetzen. Vor dem Hintergrund (teilweise) gesättigter Märkte, eines immensen Preiskampfes, stetigen Kostendrucks und der Notwendigkeit, immer neue Nischen erfolgreich zu entdecken und zu erschließen, werden BI-Lösungen als probates Mittel betrachtet, diese Probleme zu meistern.

Laut „BI Trend Monitor 2016“ des Marktforschungsunternehmens BARC stehen vor allem die Themen „Data Discovery and Visualization“ sowie „Data Quality und Master Data Management“ ganz oben auf der Prioritätenliste von CIOs und IT-Verantwortlichen. Den dritten Platz auf der Liste dieser Begehrlichkeiten nimmt „Self-Service BI“ ein. Erst mit einem Abstand folgen analytische Datenbanken und die agile Entwicklung im BI-Sektor.

„Big Data“, „BI Cloud Computing“ oder „Predictive Analytics“ werden zwar gemeinsam als „wichtig“ erachtet, doch liegen sie deutlich hinter den genannten Themen. Dies klingt zunächst überraschend, denn in den einschlägigen Computerzeitungen, Blogs oder in der öffentlichen Wahrnehmung dominieren diese Technologien eindeutig und werden als neue Allzweckwaffen für bessere Kundenbeziehungen, neue Marktpotenziale und weiteres Umsatzwachstum angesehen – doch bisher allerdings mit mäßigem Erfolg

und eher verhaltener Kundennachfrage. Auf diesen vermeintlichen Widerspruch hingewiesen, geben viele Kunden an, noch nicht die richtigen Anwendungsfälle entdeckt zu haben oder nicht über geeignetes Personal zu verfügen, um solche möglichen Projekte auch konsequent umzusetzen.

Einer der Wachstumstreiber in den nächsten Jahren dürfte das „Internet of Things“ sein. Immer mehr Maschinen, Geräte und Komponenten sowohl im privaten Sektor als auch im gewerblichen Bereich werden mit Sensoren ausgestattet, die kontinuierlich Daten sammeln und diese (entweder automatisch oder bei Bedarf) an zentrale Einheiten übermitteln. Auf den Seiten des Branchenverbandes Bitkom (siehe „www.bitkom.org“) finden sich zum Beispiel im Leitfaden „Big Data und Geschäftsmodell-Innovationen in der Praxis: 40+ Beispiele“ zahlreiche Anwenderberichte von Unternehmen, die diese Technologien bereits erfolgreich ein-

gesetzt haben. Nachfolgend eine Auswahl schon verwirklichter Projekte:

- Die Firma ThyssenKrupp Elevator hat ein intelligentes Überwachungssystem für Aufzüge in Betrieb. Die in den mehr als einer Millionen Aufzügen implementierten Sensoren liefern kontinuierlich Daten in Echtzeit über den Betriebszustand, sodass die Techniker notwendige Wartungsarbeiten und Reparaturen durchführen können, bevor es zu Störungen oder Pannen kommt.
- Zahlreiche Automobilhersteller statten ihre Fahrzeuge mit Sensoren aus, die bei Unfällen eine Meldung inklusive GPS-Daten an ein Call-Center übermitteln, um schnell Hilfe am Unfallort zu organisieren.
- Im privaten Sektor – zum Beispiel bei der Versorgung alter oder kranker Menschen – helfen sensorgestützte Systeme („Ambient

Assisted Living“) bei der Überwachung von Vitalwerten und weiteren Funktionen, um ein besseres Leben im Alter durch Einsatz moderner Technologie zu ermöglichen.

In den nächsten Jahren werden noch viele Einsatzszenarien aus dem Konsumenten-Bereich folgen, seien es sogenannte „Wearables“, also Smart Watches oder kleine Computer, die am Körper oder an der Kleidung getragen werden, Lösungen aus dem Umfeld von Smart Home oder solche Dienste, die helfen, Aufgaben im Alltag zu übernehmen oder sie zu erleichtern.

Immense Datenmengen analysieren

Die meisten dieser Anwendungen erzeugen aufgrund des teilweise millionenfachen Einsatzes extrem hohe Datenmengen, die mit hoher Geschwindigkeit zu verarbeiten und zu analysieren sind. Mit dem Wachstum dieses Datenbestands steigen logischerweise auch das Potenzial für neue Erkenntnisse und die Anforderung, diese Informationen gezielt zu untersuchen sowie aus den Ergebnissen neue Kampagnen oder neue Produkte und Dienstleistungen abzuleiten. Dabei werden zunehmend In-Memory-Datenbanken (IMDB) eingesetzt, die für solche Einsätze optimiert und teilweise auch einfacher zu administrieren sind als konventionelle Datenbanken.

Die reduzierte Anzahl der Optimierungstechniken bei den In-Memory-Datenbanken – zahlreiche Einstellungen werden automatisch durchgeführt beziehungsweise sind obsolet geworden – verringert auch den Entwicklungsaufwand. Technische Optimierungen beim Einsatz von In-Memory-Datenbanken lassen sich um bis zu 90 Prozent reduzieren und schaffen damit mehr Raum für das Wesentliche, nämlich die Umsetzung fachlicher Anforderungen. Die In-Memory-Option der Oracle-Datenbank erlaubt hier ebenfalls einen hohen Automatisierungsgrad, indem die effizienten automatischen Komprimierungsalgorithmen den Hauptspeicher optimal ausnutzen und auf analytische Indizes vollständig verzichtet werden kann.

Mit Einführung von In-Memory-Datenbanken im Unternehmen wachsen die Möglichkeiten, Informationen schnell, gezielt und nach immer neuen Kriterien auszuwerten, um so Erkenntnisse für mögliche zukünftige Entwicklungen zu erhalten. Predictive Analytics macht sich diese Technologie zunutze und verheißt die Erschließung vieler neuer Anwendungsfälle, etwa im Umfeld der Wartung von Anlagen und Maschinen, in der Be-

handlung von Krankheiten, in der Bevorratung und Disposition von Waren oder in der Planung und Lenkung von Verkehrsströmen.

Anwender erwarten schnelle Ergebnisse

Die verbesserten Möglichkeiten zur Analyse bis hin zu Echtzeit-Untersuchungen werden durch immer größere Gruppen von Anwendern in den jeweiligen Fachbereichen der Unternehmen genutzt. Dies setzt gleichzeitig voraus, dass die Programme intuitiver als zuvor bedient werden können, damit ein „Self-Service BI“ nicht nur mit den Unternehmens-Richtlinien in Einklang steht, sondern auch relevante und korrekte Ergebnisse für die Entscheidungsfindung bereitstellt.

Im Zeitalter knapper Personal-Ressourcen wird die Fähigkeit zur schnellen, eigenständigen Analyse zum Wettbewerbsfaktor und Alleinstellungsmerkmal. Die Anwender können nicht auf die Unterstützung des oftmals knappen IT-Personals warten und greifen zur Selbsthilfe in Form eines exzessiven Einsatzes von Microsoft Excel oder zum Aufbau einer Schatten- oder Parallel-IT durch Einführung von partiellen Lösungen und deren Realisierung durch externe Dienstleister. Self-Service BI versucht dieser Entwicklung wieder entgegenzuwirken und verspricht nicht nur eine höhere Eigenständigkeit der Fachanwender und des Managements, sondern eben auch eine Entlastung der IT-Mitarbeiter. Diese müssen sich zunehmend spezialisieren und auf die neuen und gestiegenen Anforderungen der Fachbereiche reagieren: So entstehen neue Berufsbilder, die irgendwo zwischen IT und den einzelnen Abteilungen angesiedelt sind: Im Umfeld von Business Intelligence werden vor allem „Data Scientists“ gebraucht, laut den Autoren Davenport und Patil vom Harvard Business Review „the sexiest job of the 21st century“, die als kleine Allroundgenies sowohl über fachliches Know-how als auch über ein solides Wissen in der Informationstechnologie verfügen müssen. Sie müssen überdies eine Menge über statistische Methoden wie auch Statistik wissen und über hohe Kommunikationsfähigkeiten und rhetorisches Geschick verfügen, um mit Managern und Entwicklern neue Konzepte zu erarbeiten.

Cloud BI entwickelt sich nur langsam

Mit der Einführung von Self-Service BI könnte auch ein unkomplizierter und schneller Zugang zu den Daten verbunden sein, so wie er durch Konzepte wie „Cloud Computing“ und „Software as a Service“ ermöglicht wird. Doch beträchtliche Sicherheitsbeden-

ken der IT-Verantwortlichen, die weiterhin die alleinige und vollständige Kontrolle über die Daten besitzen möchten, stehen der großen Verbreitung von Cloud BI im Wege. Den zweifellos vorhandenen Vorteilen wie beispielsweise der einfachen Verteilung von Anwendungen, der uneingeschränkten Skalierbarkeit, der umfassenden Unterstützung vieler (mobiler) Endgeräte oder dem geringen Administrations- und Entwicklungsaufwand stehen erhebliche Vorbehalte in Bezug auf Datenschutz und Datensicherheit entgegen. Folglich werten nur wenige IT-Manager im BI-Trend-Monitor 2016 dieses Thema als „wichtig“. Lediglich bei jenen Unternehmen, die bereits heute schon ihre Geschäftsdaten in der Cloud speichern, ist eine Bereitschaft zur externen Speicherung dieser kritischen Informationen gegeben. Dies sind laut einer Pressemitteilung von Oracle aus dem Juni des Vorjahres immerhin 70 Millionen Anwender. Mehr als 33 Milliarden Transaktionen pro Tag werden über die 19 weltweit verteilten Rechenzentren abgewickelt. In Deutschland unterhält Oracle zwei Rechenzentren – eines in Frankfurt am Main und eines in München.

Durch die Integration von polystrukturierten Daten aus sozialen Netzwerken oder anderen Formaten abseits von klassischen ERP-/CRM- und SCM-Systemen steigt die Komplexität der BI-Landschaft. Die Bildung von Data Lakes als unternehmensweites Repository für alle Daten-Formate und -Strukturen ermöglicht die Speicherung aller Informationen, ungeachtet der Frage, wann und auf welche Weise diese Daten ausgewertet werden sollen oder ob aus ihnen ein Nutzen generiert werden kann. Das Herzstück des Data Lake ist typischerweise ein Hadoop-Cluster, das beliebige Datenformate in großer Menge verarbeiten kann. Die meisten Data-Lake-Systeme nutzen die ELT-Methodik (Extract, Load, Transform) für das Sammeln und Integrieren der Daten, die über das Hadoop Distributed File System (HDFS) mit MapReduce, Spark oder ähnlichen Datenverarbeitungs-Frameworks extrahiert und verarbeitet werden können.

Fazit

Neue Technologien, Architekturkonzepte und Anwendungsbereiche werden vor allem die Analyse der rasant steigenden Datenbestände vorantreiben und Business Intelligence wachsende Bedeutung und steigende Umsätze bescheren. Mit der zunehmenden Verbreitung der neuen Lösungen wird auch die Hemmschwelle gerade bei Fachanwen-

Hier entsteht ein neuer Einblick.

Die Daten in Ihrem Unternehmen bieten hohes Potenzial für neue Erkenntnisse, wenn man sie gezielt und systematisch untersucht.

Predictive Analytics hilft bei der Entdeckung von verborgenen Beziehungen und ermöglicht sichere Prognosen für zukünftige Ergebnisse.

Unser Angebot für Sie

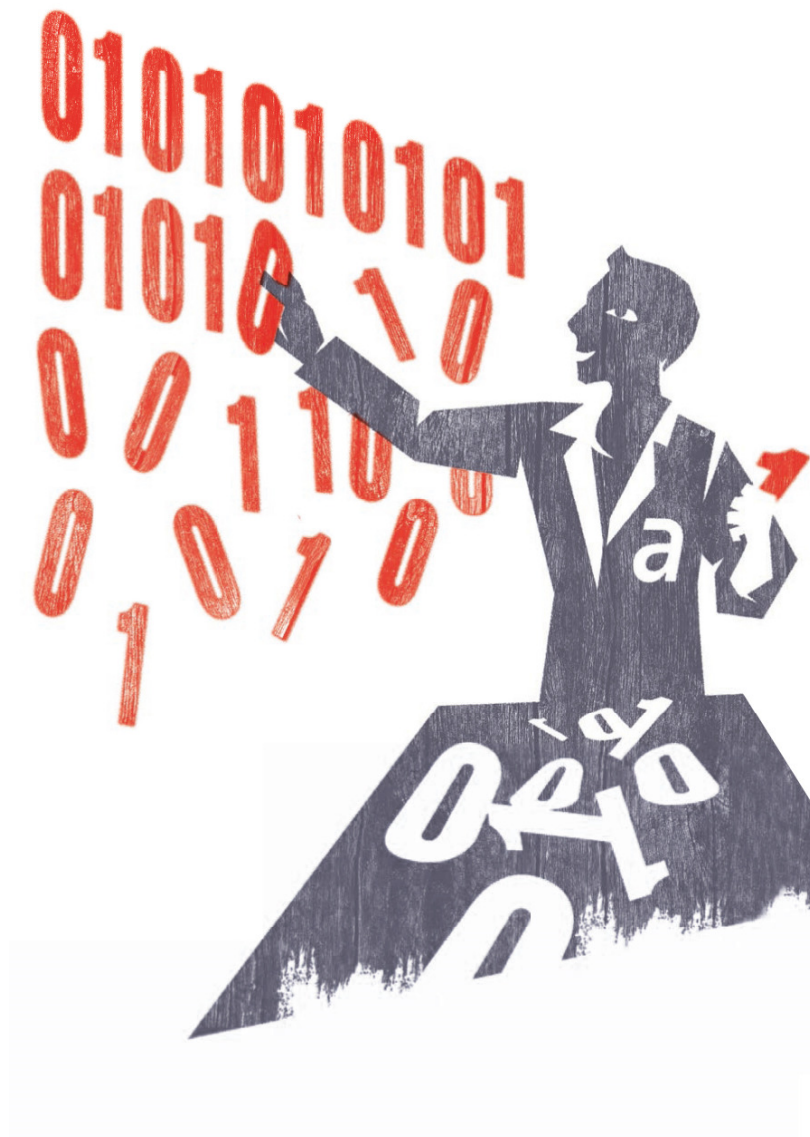
Prüfen Sie Ihr Geschäftspotenzial und vereinbaren Sie mit uns eine fallbezogene Datenanalyse.

Mehr unter www.neue-einblicke.info

Erfahren Sie mehr über Predictive Analytics und erleben Sie unsere Workshops, Vorträge und Live-Demos.

DOAG BI

8./9. Juni 2016 in Bonn



den sinken, sich systematisch mit den neuen Möglichkeiten auseinanderzusetzen und geeignete Anwendungsfälle zu identifizie-

ren. Dabei wird so manche (traditionelle) Firma eine neue Ausrichtung erfahren – hin zu einem datengetriebenen Unternehmen.

Ulrich Kramer
ulrich.kramer@areto-consulting.de

Oracle und Predictive Analytics

Oracle R Enterprise (ORE) ist ein zentraler Bestandteil der Oracle-Advanced-Analytics-Option der Oracle Database (Enterprise Edition), die das Produkt architektonisch so erweitert, das performante statistische Analysen integriert in der Datenbank ausgeführt werden können. Technisch besteht ORE aus drei Komponenten:

- **ORE Transparency Layer (TL)**
Eine Sammlung von R-Paketen, die alle grundlegenden Datentypen von

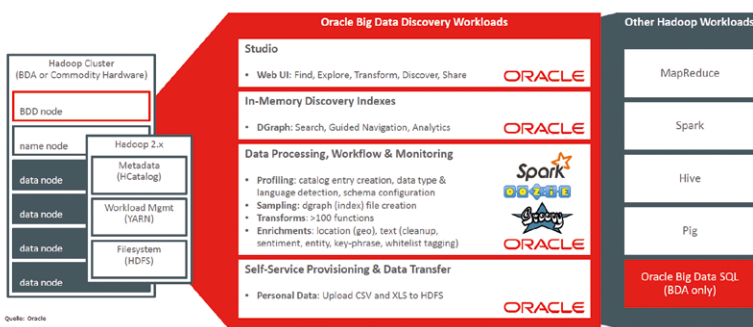
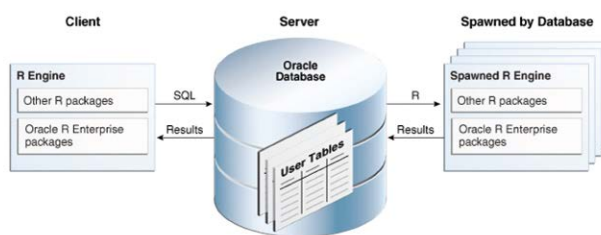
R auf äquivalente Datentypen der Datenbank mappen. Durch den TL arbeitet der Benutzer, obwohl er R-Code auf seinem Laptop eingibt, direkt auf der Datenbank. Dies ermöglicht die Arbeit mit großen Datenmengen und komplexen Berechnungen.

- **Die Statistics Engine (SE)**
Eine Sammlung von statistischen Funktionen, die als R-Paket in der R-Instanz geladen werden und mit

den Datentypen aus dem TL zusammenarbeiten. Das Besondere der SE ist, dass die statistischen Funktionen direkt in der Datenbank ausgeführt werden und ebenfalls keine Ressourcen des Client-Computers belasten. So sind beispielsweise lineare Modelle direkt auf der Datenbank verfügbar.

- **Funktion der Embedded R**
Die Datenbank übernimmt die Kontrolle der Verarbeitung der R-Skripte, die der Benutzer mit TL und SE an die Datenbank überträgt. ORE startet dafür nach Bedarf R-Instanzen, sodass zum Beispiel auch die parallele Ausführung komplexer Aufgaben möglich ist. Durch Embedded R ist kein Datentransfer aus der Datenbank erforderlich, da die R-Skripte direkt die Tabellen, Views oder andere Objekte der Datenbank als Ressource verwenden.

Architektur von Oracle R Enterprise



Oracle und Big Data Analytics

Oracle hat zahlreiche Produkte und Lösungen für Big Data im Angebot. Oracle Big Data Discovery ermöglicht es dem Anwender, intuitiv und mit visueller Unterstützung umfassende Analysen auf Hadoop-Daten durchzuführen. Das Produkt stellt eine grafische Oberfläche bereit, um

die Rohdaten in Hadoop zu finden und statistische Korrelationen zwischen Datenattributen und Kombinationen aufzudecken. Auch ohne Programmierkenntnisse können Anwender Daten aufbereiten und sofort auswerten sowie mögliche Ergebnisse wieder in Hadoop zurückschreiben

oder mit anderen Nutzern teilen. Mit dem Cloud Service, der im Herbst 2015 auf der Oracle OpenWorld vorgestellt wurde, ist das Produkt auch in der Cloud nutzbar. Neben Big Data Discovery bietet Oracle weitere Lösungen zur Analyse oder Integration von Massendaten an:

- **Oracle GoldenGate for Big Data**
Hadoop-basierte Technologie zum Streamen unstrukturierter Daten aus transaktionalen Systemen in Echtzeit in Big-Data-Systeme
- **Oracle Big Data SQL**
Erweiterung der Oracle SQL auf Hadoop und NoSQL
- **NoSQL Database**
Eine nicht-relationale Datenbank und Teil der Oracle Big Data Appliance

Mit GOBIA zu zielgerichteten Business-Intelligence-Architekturen

David Fekete und Gottfried Vossen, Universität Münster

Traditionelle Data-Warehouse-Architekturen [2,4], bestehend von unten nach oben etwa aus Datenquell-Ebene, Staging Area, Data-Storage sowie Anwendungs-Ebene, werden heutzutage durch eine Vielzahl neuartiger Big-Data-Tools wie Apache-Hadoop-Packages oder NoSQL-Datenbanken [3] infrage gestellt. Die neuartigen Tools machen es möglich, Daten zu verarbeiten und zu analysieren, die vorher nicht effizient oder überhaupt nicht betrachtet werden konnten.

Zu Big Data [4,5] zählen nicht nur viel größere Datenmengen, sondern auch vielfältigere Daten (wie Bilder, Videos, Social-Media-Beiträge) und eine schnellere Datenverarbeitung (beispielsweise Sensordaten aus dem Auto oder Smart Home). Nicht selten überfordern die Herausforderungen um Big Data die eigenen organisatorischen Analyse-Kapazitäten eines Unternehmens. Verschärft wird das Problem durch die Vielzahl von verfügbaren Big-Data-Toolsets. Insbesondere Hadoop-Distributionen wie MapR oder Hortonworks enthalten sehr viele Tools für stark heterogene Einsatzszenarien (wie Spark für die Datenstrom-Analyse oder MapReduce für parallele Verarbeitung). Dadurch wird die Auswahl des passenden Technologie-Mix für die eigene Business-Intelligence-Architektur eine viel komplexere Aufgabe als in der Vergangenheit. Auch stellt sich bei der traditionellen Data-Warehouse-Referenz-Architektur die Frage, wie sich diese in Zukunft verändern muss oder sollte, um auch Big-Data-Fragestellungen angehen zu können.

Data Warehouse + Big Data?

Diverse Ansätze, die versuchen, Data Warehouses (DWH) mit Big-Data-Tools zu kombinieren (etwa indem diese für die ETL-Schicht eingesetzt werden), bilden lediglich ausgewählte Anwendungsfälle ab, sodass daraus keine allgemeingültige Architektur abgeleitet werden kann. Bei allgemeineren Ansätzen, die die klassische DWH-Schichten-Architektur mit Big-Data-Tools erweitern [4], bleibt die Herausforderung, den richtigen Technologie-Mix zu wählen, denn diese Tools bilden sehr viele Technologien und

mögliche Szenarien ab. Verstärkt wird dies dadurch, dass viele Tools für mehrere Zwecke eingesetzt werden können und deren Rolle in der eigenen Architektur nicht von vornherein klar ist. Beispielsweise lässt sich MapReduce als generisches Tool sowohl für Datenvorverarbeitung im ETL-Prozess als auch für Analysen einsetzen.

Referenz-Architektur

Ein Ansatz zum Umgang mit dieser Herausforderung ist die von der Universität Münster entwickelte GOBIA-Methode („Goal-oriented Business Intelligence Architectures“, zu Deutsch „zielorientierte BI-Architekturen“, siehe *Abbildung 1*). Diese beinhaltet eine Technologie-unabhängige Referenz-Architektur („GOBIA.REF“), die auf dem Grundprinzip einer Schichten-Architektur basiert, einem bekannten Konzept der Informatik [1]. Ergänzt wird sie durch einen Entwicklungsprozess („GOBIA.DEV“), der sicherstellen soll, dass die Big-Data-/BI-Funktionalität und damit letztlich auch die verwendete Technologie zu den Zielen und Anforderungen einer Organisation an ihr jeweiliges BI-System passend gewählt wird. Damit soll der Kernaspekt von BI eingehalten werden, dass BI-Systeme wie ein DWH an der Unternehmung an den für sie relevanten Use Cases und deren Zielen ausgerichtet sind.

Bei der GOBIA-Methode entsteht zunächst eine konzeptionelle BI-Architektur. In einem zweiten Schritt erfolgt die Auswahl der Technologien, basierend auf der konzeptionellen BI-Architektur, für die angestrebte angepasste Architektur. Diese bildet die Blaupause für das physisch zu implementierende BI-System.

„GOBIA.REF“ ist als einfache Schichten-Architektur entworfen, da sich die übrigen Architekturen im Wesentlichen auf dieses Grundkonzept zurückführen lassen. Um die Brücke zwischen diversen Technologien zu schlagen, gibt „GOBIA.REF“ keine Technologien vor und fokussiert sich stattdessen auf BI-Funktionalität (wie „Kunden in Segmente clustern“ oder „Produkt-Sentiments ermitteln“) und die Daten, die dafür erforderlich sind.

Development-Methode

In „GOBIA.DEV“ sollen ausgehend von konkreten Geschäftszielen an das BI-System über Business Cases zunächst BI-Anforderungen abgeleitet werden. Da die Formulierung von Business Cases und Geschäftszielen in vielen Unternehmen zum Tagesgeschäft gehört, macht „GOBIA.DEV“ keine Vorgaben zu deren Herleitung.

Die BI-Anforderungen sollen insbesondere funktionale („Welche Funktionen soll das System bieten?“) und informationelle („Welche Informationen sollen vom BI-System an wen bereitgestellt werden?“) Anforderungen abdecken. Ergänzt werden diese beispielsweise durch rechtliche, organisationale und sonstige projektübliche Anforderungen. Eine BI-Anforderung könnte beispielsweise „Bereitstellung einer Analyse zum Kundenverhalten, um sein Verhalten auf Eigenschaften möglicherweise für ihn interessanter Produkte abzubilden“ sein.

Im weiteren Verlauf werden BI-Anforderungen, BI-Funktionalitäten und BI-Daten iterativ aufeinander abgestimmt (sogenanntes „Co-Alignment“). Ausgehend von den BI-Anfor-

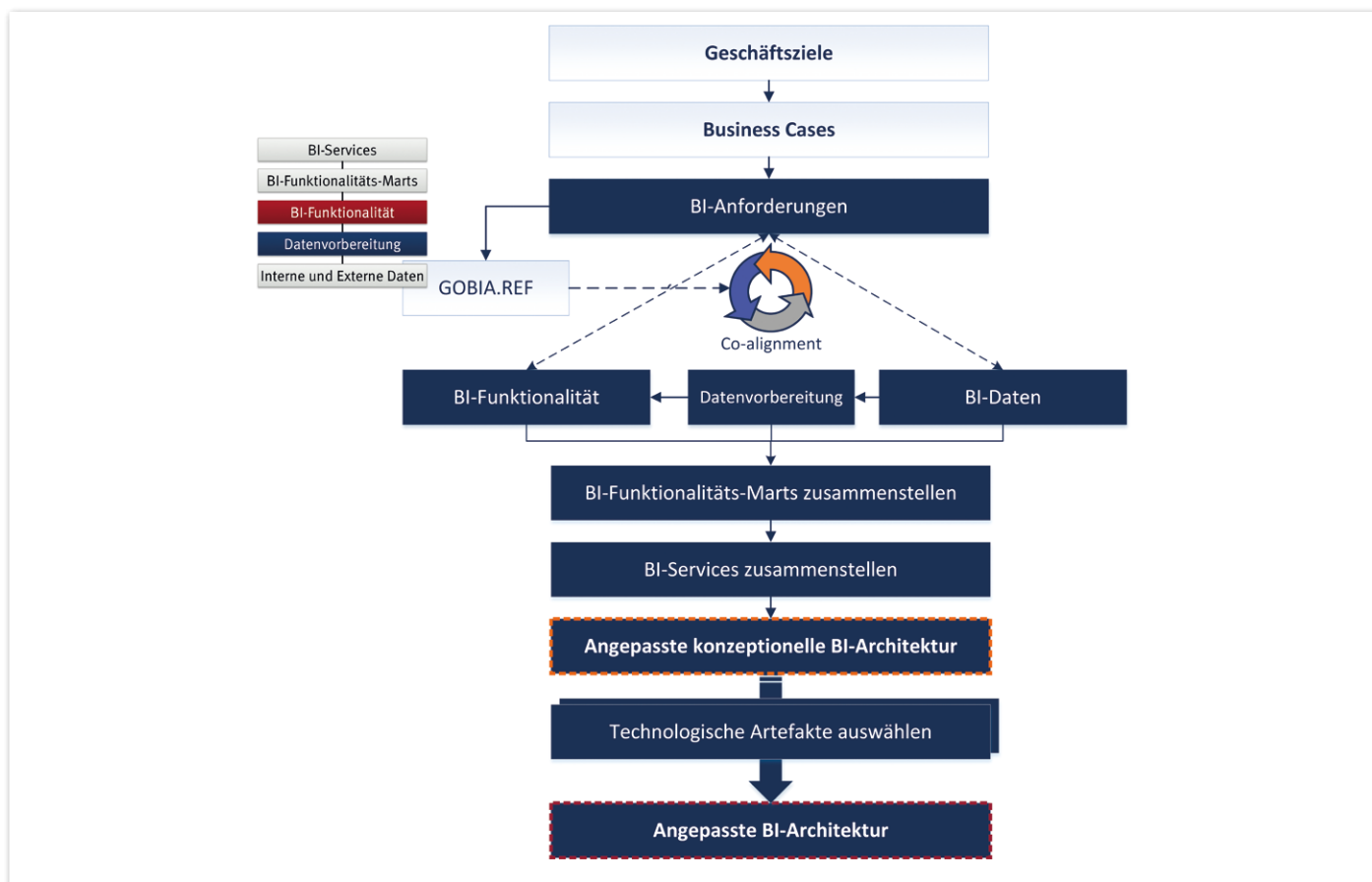


Abbildung 1: Die GOBIA-Methode

derungen werden die nötigen BI-Funktionalitäten und die passenden BI-Daten bestimmt, die diese Anforderungen umsetzen können. Beispielsweise könnten die BI-Funktionalitäten „Kunden nach Verhalten clustern“ und „Kundencluster auf Produkteigenschaften mappen“ die obigen Anforderungen erfüllen. Die BI-Funktionalitäten konsumieren dabei die bereitzustellenden Daten. Um die Eigenschaften der gewählten BI-Daten festzuhalten, können die aus der Big-Data-Diskussion bekannten „V“s (wie Volume, Variety oder Velocity) eingesetzt werden.

Ein Rückfluss zu den BI-Anforderungen ist ebenfalls vorgesehen. Wenn etwa die obige Anforderung mangels Daten nicht erfüllt wird, lässt sich die Anforderung anpassen. Sollten sich durch bestehende Daten unerkannte Potenziale ergeben, können auch ambitioniertere Anforderungen formuliert werden. Die ermittelten Elemente werden den Schichten von „GOBIA.REF“ zugewiesen.

Für die Zukunft ist vorstellbar, dass domänenspezifische Varianten von „GOBIA.REF“ in den Prozess einfließen; so könnte für die Finanz-Industrie ein spezieller Baukasten, mit üblichen Fraud-Detection-Algorithmen als BI-Funktionalitäten, bereitgestellt werden.

Falls erforderlich, lassen sich Datenvorbereitungsschritte dazu wählen. Wenn beispielsweise die zugrunde liegenden Kundendaten für die BI-Funktionalität „Kunden nach Kaufverhalten clustern“ nicht die erforderliche Qualität aufweisen, können im Sinne des bekannten ETL-Prozesses qualitätssteigernde Maßnahmen erfolgen.

Im Anschluss wird gemäß den Anforderungen die passende Bereitstellungsform der BI-Funktionalität bestimmt, etwa nach Datenausschnitt wie bei Data Marts im Data Warehouse oder prozessbasiert. Damit werden letztlich die BI-Services aufgespannt, die später durch konkrete Applikationen umgesetzt werden. Damit ist die konzeptionelle Architektur als Zwischenschritt der GOBIA-Methode vollständig.

Fazit

Die GOBIA-Methode soll mit der strukturierten Erfassung von BI-Funktionalitäten und BI-Daten sowie deren Zusammenhänge die anschließende Technologieauswahl erleichtern. Die aktive Forschung der Universität Münster zur GOBIA-Methode konzentriert sich derzeit auf die Transformation dieser konzeptionellen BI-Architektur mit ihren einzelnen Ele-

menten in eine mit den geeigneten Technologien gefüllte, individuelle BI-Architektur.

Literatur

- [1] Abeck, S., Lockemann, P.C., Schiller, J., Seitz, J.: Verteilte Informationssysteme: Integration von Datenübertragungstechnik und Datenbanktechnik. dpunkt, Heidelberg (2003)
- [2] Bauer, A., Günzel, H.: Data Warehouse Systeme. dpunkt, Heidelberg, 3rd edn. (2009)
- [3] Lechtenböcker, J., G. Vossen: NoSQL, NewSQL, MapReduce und Hadoop; Kapitel 10 in: P. Gluchowski, P. Chamoni (Hrsg.): Analytische Informationssysteme — Business Intelligence-Technologien und -Anwendungen, 5. Auflage, Springer-Verlag, Berlin 2016, 205–223
- [4] Vossen, G.: Big data as the new enabler in business and other intelligence. Vietnam Journal of Computer Science 1(1), 3-14 (Feb 2014)
- [5] Vossen, G.: Big Data: Daten sammeln, aggregieren, analysieren, nutzen; in T. Schwarz (Hrsg.): Big Data im Marketing — Chancen und Möglichkeiten für eine effektive Kundensprache; Haufe-Verlag, Freiburg, 2015, 35–54

David Fekete
david.fekete@wi.uni-muenster.de

Gottfried Vossen
g.v@www.de



Managed Self Service BI mit ODI und OBI

Michael Klose, CGI Deutschland Ltd. & Co. KG, Vorstand DOAG BI Community

In den letzten Jahren sind die Anforderungen an die Flexibilität von BI-Systemen durch Fachanwender ständig gestiegen. Dadurch erhielten viele Anbieter von Spezialanwendungen Zugang zu den Fachbereichen, um dort eigenständige Produkte für Self Service BI zu platzieren. Oftmals sind diese nicht Bestandteil der Unternehmensstrategie und führen zu einem großen Portfolio an eingesetzten Produkten in einem Unternehmen.

Die Spezialanwendungen zeichnen sich durch eine einfache Bedienbarkeit aus und erfordern keine oder nur wenige IT-Kenntnisse. Klassische BI-Tool-Anbieter wie Oracle, IBM, SAP oder MicroStrategy haben auf diesen Trend reagiert und das eigene Produkt-Portfolio dahingehend erweitert. Oracle ermöglicht mit der aktuellen Business-Intelligence-Version sowohl klassisches Reporting als auch Visual Analytics innerhalb eines Produkts. Um die Daten möglichst schnell Fachbereichen für Auswertungen über das Data Warehouse bereitstellen zu können, unterstützt der Oracle Data Integrator den Managed-Self-Service-BI-Ansatz.

Anforderungen an Self Service BI

In den Fachbereichen besteht schon lange der Wunsch, schnell und selbstständig Ad-hoc-Berichte, Dashboards und Analysen auf Data-Warehouse-Datenbeständen erstellen zu können. Problematisch oder hinderlich sind dabei häufig etablierte Prozesse im IT-Bereich, die es oft nicht erlauben, neue Datenbestände schnell in das (unternehmensweite) Data Warehouse zu integrieren und für Auswertungszwecke bereitzustellen. Zudem existieren meist Release-Zyklen durch Abhängigkeiten von Quellsystemen sowie fest definierte Testschritte wie Systemintegrations- und Endbenutzer-Tests.

Standardberichte und Analysen können gut durch die IT bereitgestellt werden, doch wie verhält es sich im Bereich „Visual Analytics“? Hier möchten Fachanwender die Datenbestände selbstständig erkunden, um unbekannte Zusammenhänge zu erkennen und daraus neue, verfeinerte Analysen ohne Wartezeit auf die IT umzusetzen.

In vielen Fällen sind auch die im Data Warehouse bereitgestellten Daten nicht ausreichend. Dies kann beispielsweise der Fall sein, wenn externe Daten von Drittanbietern (wie GfK für Marktforschungszwecke) mit den bestehenden Daten verknüpft werden müssen, um etwa neue

Marketing-Kampagnen zu erstellen. Hier ist es zwingend notwendig, die Daten ad hoc zusammenführen zu können. Gerade die Integration eigener Daten(quellen) kann es notwendig machen, bestehende Datenmodelle dynamisch anzupassen, wie es bei der Aufbereitung und Ergänzung von Hierarchien der Fall sein könnte. Auch Daten, die zum Beispiel ergänzende Informationen für einen Kunden liefern, müssen an die bestehenden Tabellen angehängt werden.

Die aus Berichten und Visual Analytics gewonnenen Erkenntnisse sollten über eine gemeinsame Collaboration-Plattform geteilt werden können. Dies trägt erheblich zur Reduzierung des Gesamtaufwands bei, da keine Auswertungen doppelt erstellt werden müssen. Zudem kann eine Analyse als Basis für andere Fachbereiche verwendet werden, wenn beispielsweise eine Kundenbasis für eine spezielle Kampagne bereits vorselektiert wurde.

Eine immer größere Bedeutung erlangt auch der Einsatz mobiler Endgeräte wie Tablets, um Ergebnisse zu präsentieren oder auch Analysen zu erstellen. Dies ermöglicht eine Erstellung am Arbeitsplatz-Rechner und die Fortsetzung oder Präsentation der Ergebnisse in einem Workshop oder Meeting.

Herausforderungen von Self Service BI

Viele Applikationen erfüllen die beschriebenen Anforderungen an Self-Service-BI-Systeme. Allerdings ergeben sich daraus auch Anforderungen, die bedient werden müssen. So sind für die Erstellung von Berichten und Analysen Prozesse im Unternehmen einzuhalten. Dies betrifft beispielsweise die Produktivsetzung eines neuen Release, das die Phasen „Entwicklung“, „Integrationstest“ und „Endbenutzertest“ durchlaufen muss. Sofern im Unternehmen ein Business Intelligence Competence Center (BICC) oder eine ähnliche Organisationseinheit existiert, ist es notwendig, dass Kennzahlen dort zur Vereinheitlichung abgestimmt und genehmigt werden.

Im Bereich „Visual Analytics“ gibt es eine Vielzahl von Anbietern, die meist nur diesen speziellen Bereich abdecken und somit eher als Insellösungen gelten. Oft sind diese auf Arbeitsplatz-Rechnern und nicht zentralisiert auf einem dafür vorgesehenen und ausreichend mit Arbeitsspeicher und CPU-Leistung versehenen Server installiert.

Daten-Integration wird bei Self-Service-BI-Systemen häufig durch die Fachbereiche

durchgeführt. Hier stellt sich allerdings die Frage, wie und durch wen garantiert werden kann, dass die Daten korrekt sind. Nicht-automatisierte Datenlieferungen wie spezielle Ad-hoc-Extrakte aus Quellsystemen können falsche oder nicht vollständige Daten liefern, die zu irreführenden oder falschen Analyse-Ergebnissen führen.

Auch unvollständige oder manuell bearbeitete Daten können Datenqualitäts-Probleme hervorrufen. Hier muss überlegt werden, was beispielsweise mit falschen oder veralteten Adressdaten geschieht oder wie mit im falschen Format manuell eingetragenen Datumswerten oder Zahlen mit falschem Dezimal- und Tausender-Trenner umgegangen wird.

Wertet man die Daten auf lokalen Arbeitsplatz-Rechnern aus, stellt das Datenvolumen oft eine weitere Herausforderung dar. Zwar sind aktuelle Rechner durchaus mit 8,16 oder auch 32 GB Speicher zu bekommen, doch wird dies schnell durch die komplette Haltung der Daten „In-Memory“ verbraucht.

Im Bereich der Daten-Modellierung besteht die Herausforderung, neue (externe) Daten mit bestehenden Daten verknüpfen zu müssen. Ein gutes Beispiel dafür ist die Anreicherung von Kundendaten im Marketing. Hier stellt sich die Frage, wie diese verknüpft werden, da Schlüsselwerte sicherlich nicht übereinstimmen. Namens- und Wohnorts-Gleichheit ergeben noch keine hundertprozentig sichere Zuordnung. Oftmals sind hier komplexe Business-Regelwerke notwendig, um die Daten zu verknüpfen.

Werden Self-Service-BI-Systeme als Insellösungen auf Arbeitsplatz-Rechnern eingesetzt, ist eine Zusammenarbeit mit anderen Mitarbeitern oft schwierig beziehungsweise Daten oder Ergebnisse werden dann häufig per E-Mail versendet. Dies ist allerdings in Bezug auf Security-Anforderungen höchst riskant, im Unternehmen verboten oder durch entsprechende Mailserver-Filter gar nicht möglich.

Gleiches gilt bei der mobilen Verwendung der Daten. Einige Self-Service-BI-Applikationen besitzen Apps für Tablets, um die Daten ebenfalls zu exportieren und zu speichern. Was passiert im Falle eines Diebstahls? Reicht hier die Code-Eingabe als Sicherungsmechanismus aus?

Managed Self Service BI als Königsweg

Wenn man die Anforderungen und Herausforderungen gegenüberstellt, kristallisieren sich konträre Aspekte zwischen Fachbereich

und IT heraus. Während die IT auf Prozesse und Qualität bedacht ist, fordern Fachbereiche schnelle, flexible Lösungen, die teilweise auch nur für kurze Zeiträume (temporär) benötigt werden. Die Anforderungen an Managed Self Service BI, um obige Punkte für Fachbereich und IT gleichermaßen zu erfüllen, sind:

- *Schnell*
 - Einfache Datenmodellierung ist gewährleistet
 - Keine beziehungsweise wenig Business-Logik ist zu implementieren
 - Schnelle Entwicklung der Ladeprozesse und Daten-Integration
- *Flexibel*
 - Variables, leicht erweiterbares Datenmodell
 - Einfache Einbindung neuer Datenquellen und Systeme
- *Temporär*
 - Spielweise für den Fachbereich
 - Kurzfristige Spezialauswertungen (wie Data Mining) möglich
 - Einmalige Daten-Analyse ohne dauerhafte Integration in das DWH
- *Prozesse*
 - Einhaltung des Deployment-Verfahrens durch DEV, SIT, UAT, PROD
 - Sicherstellung des Betriebs
 - Einbindung in die Security-Landschaft
- *Qualität*
 - Nur Integration von verifizierten Daten
 - Vollständige Prüfung auf Integrität

Man greift diese Anforderungen auf und versucht, eine geeignete Maßnahme zur Realisierung zu identifizieren. Im Mittelpunkt eines Data Warehouse steht die konsistente und historisierte Speicherung der Daten in einem vereinheitlichten Datenmodell. In der Vergangenheit wurde das Core Data Warehouse häufig mit einem (nahezu) Dritte-Normalform-Datenmodell erstellt. Durch die Redundanz-Freiheit entstehen viele Abhängigkeiten zwischen den Tabellen, was ebenfalls bedingt, dass verschiedenste Ladeprozesse implementiert werden müssen. All dies steigert den Komplexitätsgrad, reduziert die Entwicklungsgeschwindigkeit und erhöht den Testaufwand.

Gerade im Bereich der Datenmodellierung und Methodik hat sich in den letzten fünf Jahren mehr und mehr die Data-Vault-Technik durchgesetzt. Durch die geringe Anzahl verschiedener Tabellen-Typen (Hub, Link, Satellite) ist es möglich, die Ladeprozesse auf diese Anforderungen zu begrenzen. Dies führt letztendlich dazu, dass nur drei verschiedene Arten nach einem einheitlichen Muster entwickelt werden müssen.

Was bedeutet das nun im Kontext von „Managed Self Service BI“, bezogen auf die Anforderungen und Herausforderungen von IT und Fachbereichen? Durch einen Generator-Ansatz müssen Ladeprozesse (Mappings) nicht mehr manuell codiert, sondern können durch Konfiguration generiert werden. Dies ermöglicht die schnelle Entwicklung von Ladeprozessen und die leichte Einbindung neuer Datenquellen. Zudem muss der Generator einmalig intensiv getestet werden, die generierten Mappings allerdings nicht mehr, da hier dann keine Fehler mehr auftreten können. Das erlaubt ein schnelles Deployment durch die verschiedenen Umgebungen, ohne viel Aufwand und Zeit in Tests investieren zu müssen.

Die Nachteile des stark normalisierten und starren Dritte-Normalform-Datenmodells lassen sich mit einem Raw-Data-Vault-Ansatz umgehen. Raw Data Vault ermöglicht eine einfache Datenmodellierung (es gibt nur drei verschiedene Tabellen-Typen), ist sehr variabel und leicht erweiterbar. Zudem werden keine Transformationen oder Business-Logik auf die Daten angewendet. Details zu Data Vault stehen in diversen Fachbüchern.

Nachdem die Daten in das Core Data Warehouse in Raw-Data-Vault-Strukturen geladen wurden, stehen sogenannte „Raw Data Marts“ für Fachanwender bereit. Diese dienen sowohl als Spielwiese für Spezialauswertungen als auch zum Kennenlernen der Daten und Analysen im Visual-Analytics-Bereich. Oftmals ist es in diesem Umfeld leichter, mit Fachanwendern zukünftige dauerhafte Reports zu spezifizieren.

Umsetzung mit ODI und OBI

In den vorherigen Kapiteln wurden ausgiebig Anforderungen, Herausforderungen und ein möglicher Lösungsweg dargelegt. Nachfolgend ist ein möglicher Implementierungsweg mit Oracle-Produkten aufgezeigt und außerdem, welche Voraussetzungen dafür notwendig sind:

- *Datenmodellierung*
Die Stage-Tabellen im Data-Warehouse entsprechen „1:1“ den Tabellen aus dem Quellsystem, angereichert um technische Attribute wie Ladezeitpunkt und Schlüssel. Im Core Data Warehouse wird das Raw-Data-Vault-Modell nach strikten Namens-Konventionen erstellt. Dies hat zur Folge, dass keinerlei Typ-Konvertierungen wie „CHAR“ nach „NUMBER“ oder „CHAR“ nach „DATE“ durchgeführt werden. Business-Regeln wie Berechnungen oder Data Cleansing werden ebenfalls nicht umgesetzt.
- *Mapping Generierung*
Durch die strikten Konventionen bei der Daten-Modellierung und der Data-Vault-Methode ist es möglich, die Ladeprozesse mit keiner oder nur geringer Konfiguration zu generieren. Stage-Ladeprozesse können ohne weitere Konfiguration durch den Oracle Data Integrator unter Einsatz des API und Implementierung entsprechender Generatoren in Groovy automatisiert erstellt werden. Gleiches gilt für die Core-Ladeprozesse in die Raw-Data-Vault-Strukturen. Da hier im Bereich von Links nicht alles durch Namens-Konventionen abbildbar ist, müssen für Sonderfälle Tabellen und Attributs-Zuordnungen für den Generator erstellt werden.
- *„Star“-View-Generierung*
Raw-Data-Vault-Strukturen sind für einen Fachbereich durch die fremd wirkenden IDs und Tabellen-Konstrukte schwer verständlich und nicht für den direkten Zugriff geeignet. Allerdings ist es wiederum möglich, über das Data Dictionary automatisiert Views auf Basis der Links zu generieren. Letztendlich sind Links und die angebotenen Hubs nichts anderes als „Mini-Star“-Schemata. Wichtig ist auch, dass hier keine Typ-Konvertierungen oder Business-Regeln implementiert werden.
- *Quick Präsentation*
Die „Mini-Star“-Schemata, die durch die „Star“-View-Generierung entstehen, bilden die Basis für den Aufbau und die Integration in die Oracle-BI-Metadaten-Schicht. Um Daten schnell auswertbar bereitstellen zu können, sollte auch hier auf Business-Logik verzichtet werden. Da die Daten bereits im Star-Schema vor-

liegen, können sie einfach in die Metadaten integriert werden. Dies kann manuell oder auch hier durch den Einsatz eines Generators geschehen. Lediglich ein logisches Verbinden der Stars (Links) erfordert gegebenenfalls fachliches/technisches Know-how. Bis eine vollständige Integration in das Data Warehouse erfolgt, können die Präsentationen in einem getrennten Spielbereich bereitgestellt werden.

Fazit

Bei Self-Service-BI-Applikationen verwenden Fachanwender häufig die angelieferten Rohdaten, um selbst Bereinigungen und Zusammenführungen durchzuführen sowie Kalkulationen zu erstellen. Gleiches wird durch den oben beschriebenen Ansatz ebenfalls gewährleistet.

Durch den gewählten Managed-Self-Service-BI-Ansatz werden die Ladeprozesse nicht mehr manuell implementiert, sondern vom Datenmodell generiert. Sind Änderungen beispielsweise an Tabellen-Strukturen notwendig oder kommen neue Systeme und Tabellen hinzu, sind diese schnell umsetzbar.

Im Falle der Generierung sind kurze Release-Zyklen realisierbar, was auch im Data-Warehouse-Bereich den agilen Projektansatz unterstützt. Ein Einsatz von Insel-Lösungen für Analyse-Zwecke ist durch die Integration in die Enterprise-BI-Plattform nicht mehr notwendig. Ein wichtiger, nicht zu vernachlässigender Aspekt ist auch die vollständige Gewährleistung von Security- und Datenschutz-Anforderungen, da die Daten auf zentralen Servern und nicht auf lokalen Arbeitsplatz-Rechnern oder Mobile Devices liegen.

Michael Klose
michael.klose@cgi.com

Business Intelligence und Visual Analyzer

Annett Thurm-Meyer, ORACLE Deutschland B.V. & Co. KG

Oracle zählt zu den führenden Anbietern von Business Intelligence und trägt kontinuierlich zur Weiterentwicklung dieses Themas auf dem Markt bei. Dabei verfolgt man seit Jahren sehr erfolgreich die Strategie, alle Lösungen zu integrieren. Dies senkt operative Kosten, erleichtert Data Governance und beschleunigt Prozesse. Ein vorherrschender Trend ist der Plattform-Shift von On-Premise- hin zu Cloud-basierten Lösungen, den Oracle mit viel Energie umgesetzt hat. Er berücksichtigt auch bereits den nächsten Trend, den Self-Service für Endanwender. Ganz neu, und Gegenstand dieses Artikels, ist das Thema „Visual Analytics“ im Kontext der Oracles-Business-Intelligence-Strategie.

Kaum ein IT-Segment unterliegt einem so schnellen Innovationswandel wie das Thema „Business Intelligence“. „Big Data“ hat gerade Einzug in die führenden Köpfe erfolgreicher Unternehmen erhalten und während noch um den Standard von Big-Data-Datenhaltung und -analyse auf dem Markt gerungen wird, gehen die Entwicklungen schon weiter in Richtung „Internet of Things“. Parallel dazu findet in zunehmendem Maße ein Plattformwandel von On-Premise-Lösungen, also von Unternehmen betriebenen und verantworteten IT-Lösungen, hin zu Cloud-Lösungen statt, die von IT-Anbietern betrieben und verantwortet werden.

Das derzeitige Oracle-Business-Intelligence-Angebot umfasst neben der Business Intelligence Suite Extended Edition mit vielen zusätzlichen Modulen wie der führenden multidimensionalen Datenbank Essbase, den mobilen Komponenten BI Mobile HD und Mobile App Designer, dem Strategy and Scorecard Management, fertigen Bereichs- und Branchen-BI-Applikationen, auch ein Werkzeug für das Big Data Discovery und, ganz neu, den Oracle Visual Analyzer. Dabei wird Oracle sämtliche Entwicklungen sowohl On-Premise als auch zeitnah in der Cloud unterstützen.

Unternehmen haben die Wahl, ob sie selber für den Betrieb und die Datenhaltung

verantwortlich zeichnen wollen und hierfür auch die entsprechenden Budgets zur Verfügung stellen oder ob sie auf die bewährten Lösungen in der Cloud umsteigen und die Investitionsausgaben (Capital Expenditures = CAPEX) in reine Betriebskosten (Operational Expenditures = OPEX) umwandeln. Die zuvor für IT-Investitionen und Betrieb genutzten Budgets können dann für Innovationen freigesetzt werden, die einen echten Wettbewerbsvorteil bringen.

Analysing Data Visually

Viele werden sich an dieser Stelle fragen, worin der große Hype der Daten-Visualisierung

Oracle Business Intelligence Cloud Service

- Dashboards and Visual Data Discovery
- Self-service Reporting
- Data Mashup across Managed and User Data
- Access on Desktop or Mobile
- A Cloud based Oracle DB
- Best-in-class Cloud Infrastructure

Abbildung 1: Beispiel für Oracle Visual Analyzer in der Cloud

besteht. Längst verfügen alle Business-Intelligence-Werkzeuge über eine grafische Engine, die Daten bestmöglich aus den Tabellen in eine ansprechende und leicht zu interpretierende Grafik überführt. Im Falle von Oracle gibt es hier sogar einen Vorschlagsmechanismus, der weniger erfahrenen Anwendern eine optimierte Form der Darstellung für den gewählten Datenausschnitt vorschlägt.

„Visual Analysing“ ist jedoch eine komplett neue Art und Weise, sich einem Thema zu nähern, ohne dabei über Kenntnisse zu den Daten verfügen zu müssen oder ein Data Scientist zu sein. Visual Analysing ist die Königsdisziplin, wenn es um intuitive Analyse geht. Anwender denken über ein Problem nach und geben dank dem ASK-Feature lediglich das sie interessierende Schlagwort wie „Umsatz“ ein. Die Software liefert im Ergebnis eine beschriebene Liste möglicher Ergebnisse. Der Anwender wählt das Resultat und erhält sofort die passende Visualisierung. In diesem Fall eine kaum aussagekräftige Kennzahl.

Im nächsten Schritt können das Jahr, die Produkte oder eine mögliche Region ausgewählt werden. Schon liefert der Visual Analyzer die jeweils optimierte Form der Visualisierung. Ergebnisse lassen sich durch das Highlighten auf einen bestimmten Raum filtern. Das Ziehen weiterer Attribute auf Felder wie „Farbe“ oder „Kategorie“ differenziert das Ergebnis weiter. Durch das Hinzufügen oder Tauschen von Attributen kommen immer neue Einblicke auf das Problem bis hin zur Erkenntnis, welche Daten die gestellte Frage erklären.

Der Visual Analyzer erlaubt es Anwendern, mit Daten zu experimentieren und dazu aus einer Vielzahl von unterschiedlichen Optionen für die Art und Weise der Visualisierung zu wählen, ohne dabei einem üblichen analytischen Prozess folgen zu müssen. Dadurch ist es ein Einfaches, Korrelationen, Muster oder Trends zu entdecken – ohne die Anwendung komplexer statistischer Methoden. Natürlich erhebt der Visual Analyzer nicht den Anspruch, einem Data Scientist gerecht zu werden, erlaubt jedoch Endanwendern, auf anderem Wege zu ähnlichen Ergebnissen zu kommen (siehe Abbildung 1).

Integration mit Business Intelligence

Das Ergebnis, also der erklärende Datenausschnitt, kann dann zur weiteren Verarbeitung in Business Intelligence übernommen werden. Dort lassen sich die Daten vielfältig weiter nutzen, beispielsweise in Form von

Grafik, Tabelle oder Ticker, und auf einem Dashboard für wiederkehrende Entscheidungen kommunizieren. Es ist auch möglich, sie als Grundlage für Pixel-perfekte Berichte zu nutzen und die Ergebnisse etwa per E-Mail zu verschicken.

Die weitere Verarbeitung mit dem Mobile App Designer kann die Storytelling-Funktionalität des Visual Analyzer sinnvoll ergänzen. Mit wenigen Mausklicks können Analysen mit erklärenden Texten und weiteren Informationen zu einer Applikation zusammengestellt und verteilt oder zur Verfügung gestellt werden. Hierfür ist es erforderlich, den Visual Analyzer als integrierte Komponente eines Business-Intelligence-Toolsets zu verwenden.

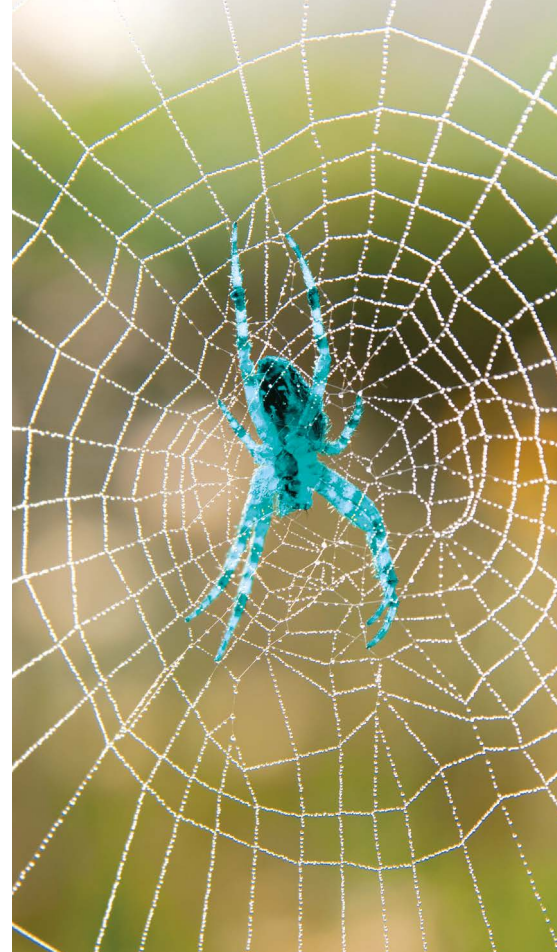
Die Integration von Visual Analyzer in die Business Intelligence Suite Extended Edition arbeitet auf Metadaten-Ebene. Der Visual Analyzer setzt auf den Metadaten des BI-Servers auf. Diese Integration unterstützt auch die Zusammenarbeit von Mitarbeitern durch den kontrollierten und systemgestützten Austausch von Erkenntnissen.

Eine Stand-alone-Lösung kann diese weitere Nutzung und Unterstützung nicht liefern. Genau hier liegt ein wesentlicher Vorteil von Visual Analyzer im Vergleich zu anderen Anbietern, die sich nicht nahtlos in die unternehmensweite Business-Intelligence-Strategie einfügen (siehe Abbildung 2).

Nutzung von Visual Analysing im Unternehmen

Es stellt sich die Frage nach typischen Anforderungen, die den Einsatz eines solchen Werkzeugs rechtfertigen. Zunächst ist anzumerken, dass Anwendungsbeispiele nicht speziellen Industrien zuzuweisen sind. Es geht eher um funktionale Anforderungen. So gibt es Mitarbeiter in Unternehmen, die sich in periodischen Abständen mit der Erfolgskontrolle beschäftigen. Dafür stehen ihnen gesicherte Daten aus einem Data Warehouse zur Verfügung. Es geht dabei zumeist um allgemeine Erfolgskennzahlen zum Thema „Finanzen“ oder zur operativen Effizienz.

Daneben gibt es Mitarbeiter, die sich mit Fragestellungen beschäftigen, die in der Regel nicht mit Data-Warehouse-Daten allein zu erklären sind. Es gilt, private Daten in Form von beispielsweise Meinungsumfragen oder Trends mit in die Problem-Analyse zu integrieren. Zur Bewältigung dieser eher inhaltlich getriebenen Fragen werden



Exzellente Baupläne für die Digitale Ökonomie!

Dafür steht PROMATIS als Geschäftsprozess-Spezialist mit mehr als 20 Jahren Erfahrung im Markt. Gepaart mit profundem Oracle Know-how schaffen wir für unsere Kunden die Digitale Transformation:

- Oracle SaaS für ERP, SCM, EPM, CX, HCM
- Oracle E-Business Suite und Hyperion
- Oracle Fusion Middleware (PaaS)
- Internet of Things und Industrie 4.0

Vertrauen Sie unserer Expertise als einer der erfahrensten Oracle Platinum Partner – ausgezeichnet mit dem Oracle Partner Excellence Award 2015.

PROMATIS



PROMATIS Gruppe
Tel.: +49 7243 2179-0
www.promatis.de · info@promatis.de
Ettlingen/Baden · Hamburg
Graz · Zürich

Oracle Business Intelligence 12c mit Visual Analyzer

IT friendly:
Governance, Singular semantic, Enterprise ready



Business user friendly:
Ease of Use, Agile, Self-service

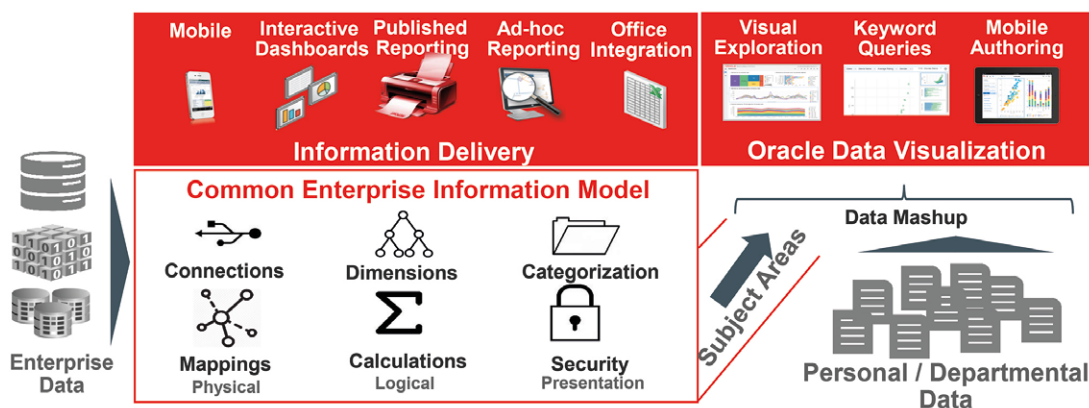


Abbildung 2: Integration mit Oracle Suite EE und Visual Analyzer

neue Daten als Excel-Dateien hinzugefügt. Diese können mit bestehenden, von der IT zur Verfügung gestellten Daten (Geschäftsbereichen) verschnitten werden. Natürlich lassen sich im letzteren Fall auch diese Daten aktualisieren. Außerdem ist ein Teilen der Daten mit anderen Nutzern möglich und schließlich können Daten auch wieder entfernt werden.

Natürlich ist es nicht immer erforderlich, sich auf private Daten zu beziehen. Es kann auch sinnvoll sein, die klassischen Data-Warehouse-Daten mit dem Konzept des Visual Analyzing, also mit dem Stellen von Fragen, aufzubereiten und gegebenenfalls ganz neue Erkenntnisse aus bereits hundertfach analysierten Daten zu erhalten.

Der sinnvolle Einsatz von Visual Analyzer im Self-Service

Wie bereits erwähnt, ist es nicht immer ausreichend, bestehende Daten zu analysieren. Ein weiterer Trend im Umfeld von Business Intelligence ist das Thema „Self-Service“. Mitarbeiter stehen unter Zeitdruck und sind nicht mehr gewillt, auf die Umsetzung analytischer Anforderungen durch die IT-Abteilung zu warten, die den Fachbereichen weitere Daten nach vorhergehender Prüfung und gegebenenfalls Anwendung von Data-Cleansing-Prozessen zur Verfügung stellt. Daten müssen sofort zur Verfügung stehen; bestenfalls mit den Unternehmensdaten

verknüpft und ergänzt (Stichwort: „Data-Blending/Data Mashup“).

Führende Analysten sagen, dass Self-Service Gefahren birgt, wenn man gesicherte Unternehmensdaten um private Daten erweitert. Dazu zählen inkonsistente, korrupte und qualitativ schlechte Daten. Entscheidungen, die auf solchen Daten beruhen, können leicht falsch sein. Somit ist verständlich, dass die Verantwortung für das Thema „Data Governance“, die bisher von der IT-Abteilung allein getragen wurde, im Falle von Daten-Blending und Data Mashup durch Endanwender auch von diesen mit verantwortet werden muss. Teilt man also diese Verantwortung, ist es wichtig, ein Tool zur Verfügung zu stellen, das Anwendern den Prozess des Data Mashup weitestgehend automatisiert abnimmt und wenig Spielraum für Fehl-Interpretationen von Daten-Definitionen, Metadaten und Datenmodellen lässt.

Die Modellierung sollte weitestgehend vom System übernommen werden, was beim Visual Analyzer vollständig gegeben ist. Er bietet sowohl das Hochladen von privaten, also nicht von der IT bereitgestellten Daten als auch die Verknüpfung von privaten Daten mit gesicherten Unternehmensdaten.

Eine weitere Herausforderung im Umfeld von Self-Service ist die Sicherheit für sensible Daten, die im Fall von Insel-Lösungen, die

sich vollständig in der Verantwortung von Endanwendern befinden, infrage zu stellen ist. Visual Analyzer greift über den Web-Logic Server auf die gleiche Sicherheitsstufe zu und stellt somit alle Daten unter die gleiche Zugriffsberechtigung.

Fazit

Viele Fachabteilungen in Unternehmen haben Anforderungen an Business Intelligence, die mit klassischen Bordmitteln der umgesetzten Business-Intelligence-Strategie nicht umzusetzen sind. Sie adressieren eine neue Dimension der Flexibilität der zugrunde liegenden Daten, aber auch der analytischen Möglichkeiten. Außerdem werden Zeitfenster immer knapper und datenbasierte Antworten auf drängende Fragen erfolgskritischer. Visual Analyzer ist die Antwort, mit der Unternehmen allen Anforderungen nach mehr Flexibilität und Vielfalt in einer kontrollierten und integrierten Form begegnen können. Dabei ist es unerheblich, ob die Business-Intelligence-Strategie On-Premise, in der Cloud oder in einer hybriden Version verfolgt wird.

Annett Thurm-Meyer
annett.thurm-meyer@oracle.com

OBIEE 12c: Ein Einblick in die neue Reporting-Engine von Oracle

David Michel, Apps Associated

Auf der letzten Oracle OpenWorld wurde als Nachfolger auf das Release 11g die neue Oracle Business Intelligence Enterprise Edition (OBIEE) vorgestellt. Die Version 12c enthält sowohl auf Seiten der Architektur, mit Auswirkung auf Installation und Konfiguration, als auch im Bereich der neuen Features einige Änderungen und Neuerungen. Dieser Artikel gibt einen ersten Einblick.

Bevor eine erste Installation der BI Suite in Release 12c vorgenommen werden kann, muss zwangsläufig JDK 8 auf dem Zielsystem vorhanden sein oder bereitgestellt werden. Ältere JDK-Versionen werden für OBIEE 12c nicht unterstützt.

Eine erste größere Veränderung innerhalb der Installationsroutine betrifft die Fusion-Middleware-Komponenten wie den WebLogic Server. Wo Release 11g noch die Möglichkeiten geboten hatte, eine komplette Installation aller benötigten Komponenten durchzuführen, muss in Release 12c eine getrennte Installation dieser Tools erfolgen. Nachdem diese bereitgestellt sind, kann die Installation („Software Only“) der eigentlichen OBIEE 12c Suite vollzogen werden.

Wie auch schon in seiner Vorgängerversion legt OBIEE 12c Meta-Informationen in der Datenbank ab. Die dazu benötigten Repositories sind mithilfe des Repository Creation Utility (RCU) anzulegen. Anders als in Release 11g muss dazu das RCU nicht gesondert heruntergeladen und installiert werden – es ist nun fester Bestandteil der OBIEE-12c-Installation.

Eine weitere Vereinfachung betrifft die Vergabe der Ports, die für die unterschiedlichen Komponenten innerhalb der BI Suite zur Verfügung stehen. In 12c lässt sich ein Port Range angeben, in dem die Komponenten anschließend installiert und konfiguriert werden (siehe *Abbildung 1*). Dies erleichtert die Administration ungemein, wenn aufgrund von Firewall-Einschränkungen oder Policies nur bestimmte Ports benutzt werden dürfen. Eine aufwändige Anpassung der Konfiguration innerhalb von Post-Installation-Aktivitäten entfällt somit.

Keine großen Veränderungen gibt es hingegen innerhalb der Development-Tools. Diese stehen nach wie vor als Windows-Anwendungen zur Verfügung und dienen der Entwicklung und Pflege des Systems (wie Administration Tool, Catalog Manager).

Eine der wohl häufigsten Aktivitäten, die in der OBIEE-11g-Systemlandschaft nach Veränderungen am Web Catalog oder nach Deployments anfiel, war die Aktualisierung der Global Unique Identifiers (GUID). Jeder interne BI-User verfügte in der Vergangenheit neben einem eindeutigen Username auch über eine solche GUID, über die er eindeutig innerhalb des Systems identifiziert

sowie mit Inhalten (Reports, Dashboards) und Rechten verknüpft war. Im Release 12c kommen diese IDs nicht mehr vor, stattdessen wird nun der Username verwendet. Die Aktualisierung der GUIDs entfällt somit im aktuellen Release.

Nach wie vor können die einzelnen Dienste der BI Suite über Skripte („cmd“ beziehungsweise „sh“) gestartet beziehungsweise gestoppt werden. Jeder Dienst dokumentiert die ausgeführten Tätigkeiten in entsprechenden Log-Files. Lediglich die Pfade, die zu den Skripten oder Log-Files führen, haben sich teilweise geändert. Start- und Stopp-Skripte befinden sich zum Beispiel unter „[ORACLE](#)

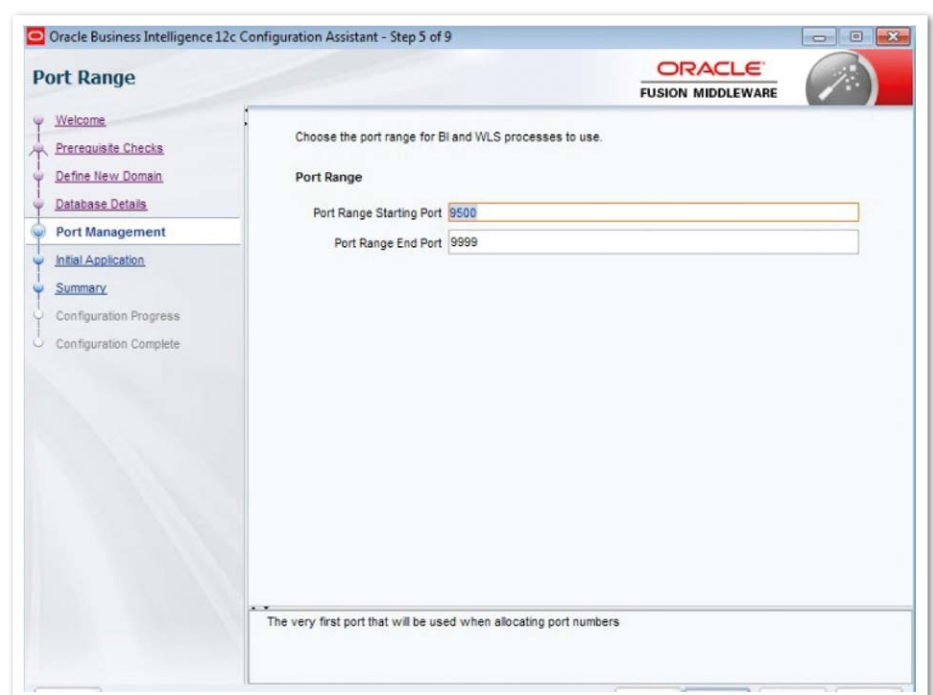


Abbildung 1: Port Management innerhalb der OBIEE-12c-Installation

HOME]/user_projects/domains/bi/bitools/bin". Eine kleine Änderung gibt es dagegen bei der Benennung von Repository und Web Catalog. Diese heißen jetzt „Datamodel“ beziehungsweise „Content“ und stehen unter „[ORACLE_HOME]/user_projects/domains/bi/bidata/service_instances/ssi/metadata“ (siehe Abbildung 2).

Die wichtigsten Änderungen in der Architektur

Eine der sicherlich größten und bedeutungsvollsten Änderungen von OBIEE 12c

liegt in der System-Architektur. Mit dem aktuellen Release ist es erstmals möglich, die BI-Systeme in „Multi Tenancy“ zu betreiben. Diese Architektur ist vor allem aus dem Bereich des Cloud Computing bekannt, bei der nicht für jeden Anwender beziehungsweise Kunden eine dedizierte Infrastruktur bereitgestellt wird, sondern alle Nutzer auf der gleichen Plattform arbeiten (siehe Abbildung 3). Dementsprechend wird im Zusammenhang mit OBIEE 12c auch vom „BICS-Model“ gesprochen. „BICS“ bezeichnet in diesem Zusammenhang die Oracle

BI Cloud Services, also eine Cloud-Lösung, die diese Architektur bereits anwendet. Umgemünzt auf die „On-Premise“-Lösung OBIEE 12c bedeutet dies, dass Oracle nun die Möglichkeit bereitstellt, mehrere Anwendungen parallel auf einem BI-System zu betreiben. Anwendungsseitig wird dieses Verfahren auch „Pluggable BI“ genannt.

Wie Abbildung 4 zeigt, bedeutet dies, dass mit einer OBIEE-12c-Installation mehrere BI-Anwendungen betrieben und mehrere Service-Instanzen pro Anwendung zur Verfügung gestellt werden können. Theoretisch könnte man somit beispielsweise das Entwicklungs-, das Test- und das Produktiv-System des BI-Servers auf einer Hardware betreiben. Ein anderer Anwendungsfall wäre, eine BI-Finanzlösung parallel zu einer BI-HR-Lösung zu betreiben, die jederzeit voneinander getrennt werden können. In Release 11g war dies nur möglich, wenn alle BI-Inhalte in einem zentralen Repository gehalten wurden. Ein einfaches Hinzufügen oder Entfernen bestimmter Inhalte war ohne Weiteres nicht möglich.

Eine weitere Neuerung auf Seiten der Architektur stellt das sogenannte „BI Application Archive“ dar; ein Werkzeug, um BI-Applikationen zwischen unterschiedlichen Umgebungen zu transportieren. Das mit diesem Tool erzeugte Archiv ist eine einzelne Datei und verfügt über die Endung „.bar“. Dementsprechend wird es auch „BAR-File“ genannt und beinhaltet typischerweise das Datamodel („Repository“), den Content („Web Catalog“) und die Security („Roles“). Es kann darüber hinaus auch mit eigenen Inhalten wie anwendungsspezifischen „css“-Informationen ergänzt werden. Ein Deployment kann daher mithilfe dieses Archivs in einem Schritt erfolgen; es ist nicht mehr nötig, zuerst das Repository zu veröffentlichen und anschließend den Web Catalog hochzuladen.

Obendrein ist es auch möglich, Abhängigkeiten zu definieren. Ein Application-Archiv kann von einem anderen abhängig sein. Abgeleitet aus dieser Tatsache ist also auch ein Deployment von mehreren BI-Applikationen auf eine BI-Domain realisierbar. Das BAR-File kann außerdem beim Erstellen der Domain (Installation) angegeben werden, um die Inhalte direkt nach Fertigstellung bereitzustellen (siehe Abbildung 5).

Ebenfalls neu ist das „Baseline Validation Tool“. Es ermöglicht, ein „Basis-Datenset“, auch „Baseline“ genannt, zu erstellen. Das Resultat kann dann mit einem ande-

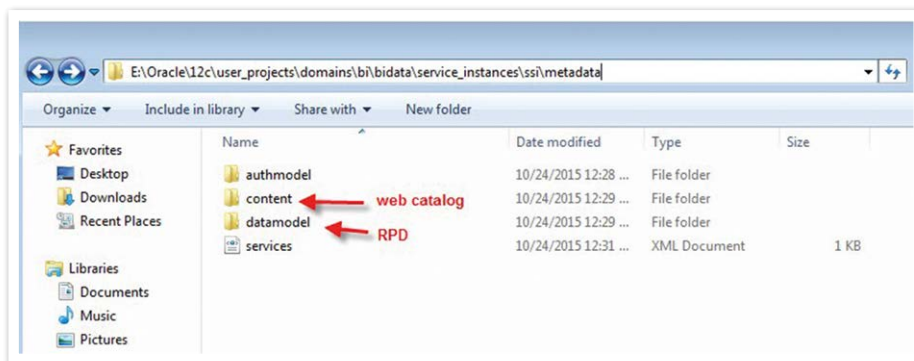


Abbildung 2: OBIEE 12c Metadaten

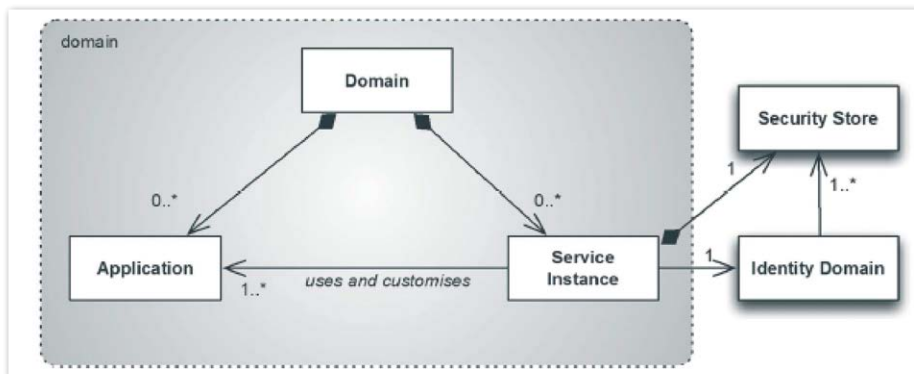


Abbildung 3: Multi Tenancy beziehungsweise BICS-Model

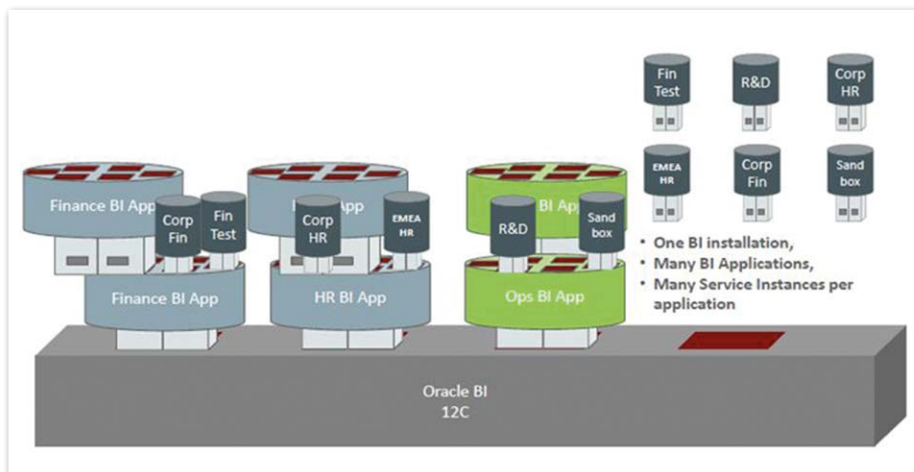


Abbildung 4: Pluggable BI

ren Basis-Data-Set verglichen werden. Die zu vergleichenden Data-Sets können dabei vom gleichen oder von verschiedenen Systemen stammen. Dieses Tool unterstützt das Regression Testing, indem Veränderungen zweier Datenstände verglichen und ausgewertet werden, und steigert auch das Vertrauen im Falle eines Upgrades oder einer Migration. Es kann im aktuellen Release 12c ebenso wie im Release 11g verwendet werden. *Table 1* und *Abbildung 6* zeigen, welche Aktivitäten dieses Werkzeug unterstützt.

Der Upgrade-Pfad von OBIEE 11g zu 12c
 Nachdem die neue OBIEE-12c-Umgebung zur Verfügung steht, kann mit dem eigentlichen Upgrade-Prozess begonnen werden. Dazu wird auf die zuvor beschriebenen Werkzeuge zurückgegriffen. Wie in der *Abbildung 7* dargestellt, werden zuerst die BI-Utilities auf die OBIEE-11g-Umgebung kopiert. Dort wird im Anschluss mithilfe des Baseline-Validation-Tools ein Basis-Data-Set der 11g-Umgebung erzeugt. Danach kann ein BI Application Archive (BAR-File) angelegt, auf die neue 12c-Umgebung im-

portiert und dort in die BI-Domain integriert werden.

Anschließend wird erneut ein Basis-Data-Set mit dem Baseline-Validation-Tool erstellt. Danach werden beide Data-Sets (von der OBIEE-11g- und der OBIEE-12c-Umgebung) miteinander verglichen. Die eventuell aufgetretenen Fehler beziehungsweise Anmerkungen des Baseline-Validation-Tool-Summary müssen im Nachgang korrigiert werden. Ist auch dieser Schritt vollzogen, kann die neue OBIEE-12c-Umgebung produktiv zum Einsatz kommen.

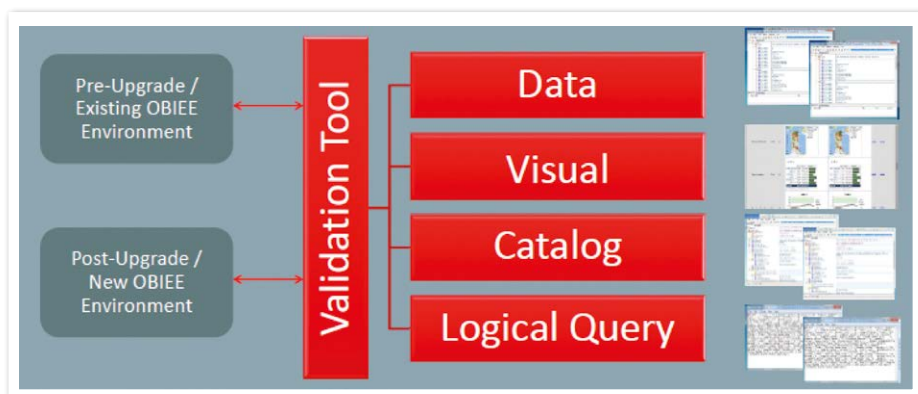


Abbildung 6: Funktionsweise Baseline-Validation-Tool

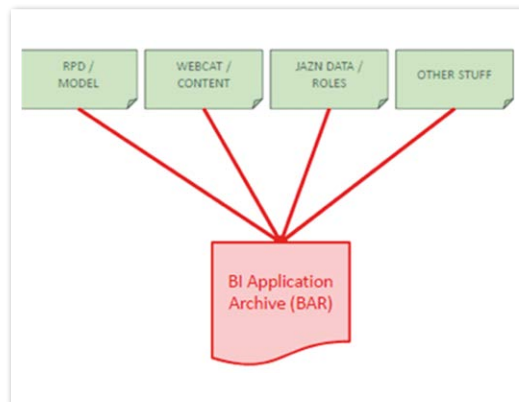


Abbildung 5: Aufbau des BAR-File

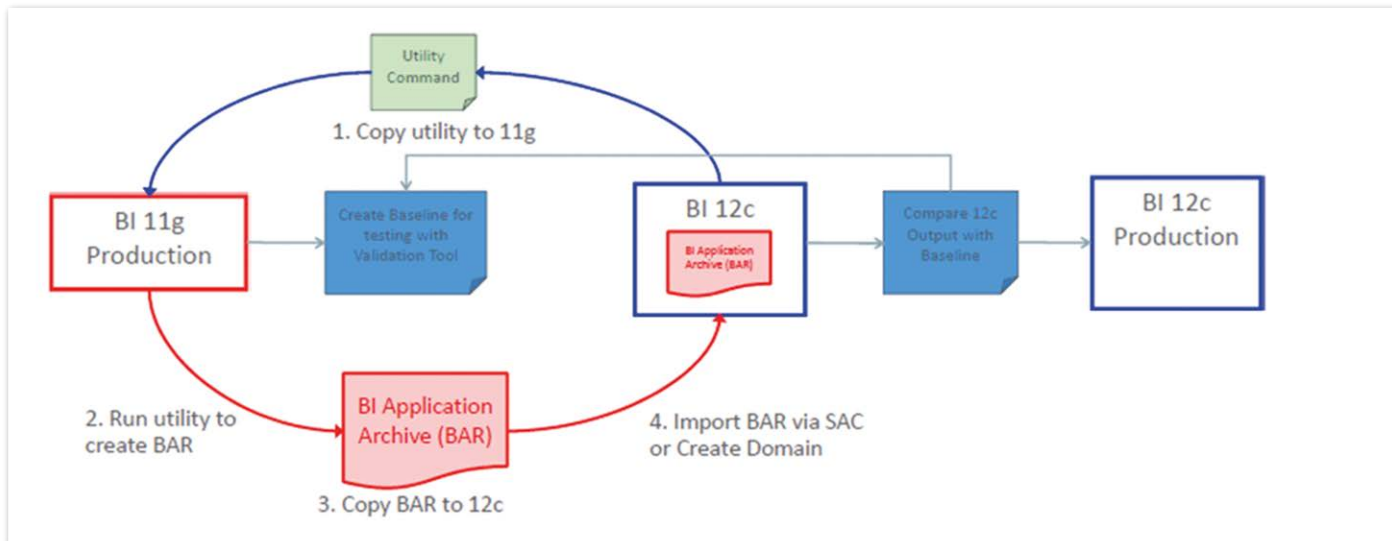


Abbildung 7: Vorgehen für das OBIEE-Upgrade (Version 11g zu 12c)

Funktion	Erläuterung
Data Verification	Vergleicht Resultate unterschiedlicher „Basis-Data-Sets“
Visual Verification	Erfasst Screenshots von Reports und vergleicht, ob diese gleich aussehen
BI Catalog Verification	Objekte und Metadaten werden mit dem Zielsystem verglichen
Logical Query Verification	Vergleicht logische Query-Definitionen mit Zielsystem
Performance Verification	Prüft Performance-Profil auf mögliche Verschlechterung

Table 1

Die wichtigsten neuen Features

Für Oracle-BI-Cloud-Anwender bereits ein alter Bekannter, rückt die OBIEE-Komponente „Visual Analyzer“ für On-Premise-Anwender zum ersten Mal mit dem Release 12c in den Vordergrund. Dabei handelt es sich um eine neue und zusätzliche Möglichkeit, um Self Service Analytics betreiben zu können. Das bedeutet, der Anwender kann – ähnlich wie in der Komponente „Analytics“ – eigene Auswertungen durch Klick- beziehungsweise „Drag & Drop“-Funktionalitäten erstellen. Dabei bietet der Visual Analyzer nicht nur die Möglichkeit, Auswertungen auf vordefinierten Data Sets („Subject Areas“) zu betreiben, sondern es können auch Data Mashups, etwa durch Einbinden von Flat Files, benutzerfreundlich, schnell und einfach betrieben werden.

Das Einbinden dieser Daten durch die IT ist hierbei nicht erforderlich, sondern kann direkt über den Browser durch den Anwender selbst erfolgen. Durch dieses sinnvolle Modul hat Oracle seine BI Suite gezielt ergänzt, um auch die Bereiche „Data Explo-

ration“ und „Data Discovery“ innerhalb des gleichen Produkts abzubilden. Nicht nur innerhalb des Visual Analyzer (siehe Abbildung 8) ist das Thema „Data Mashup“ gezielt verstärkt worden. Auch in der bereits bekannten Komponente „Analytics“ findet sich die-

ses Feature im Release 12c wieder. OBIEE 12c erlaubt es nämlich, mit logischem SQL direkt Daten aus externen Dateien aufzurufen (siehe Abbildung 9).

Wer zum Erstellen von Analysen über die im Standard vorhandenen Möglichkeiten hi-

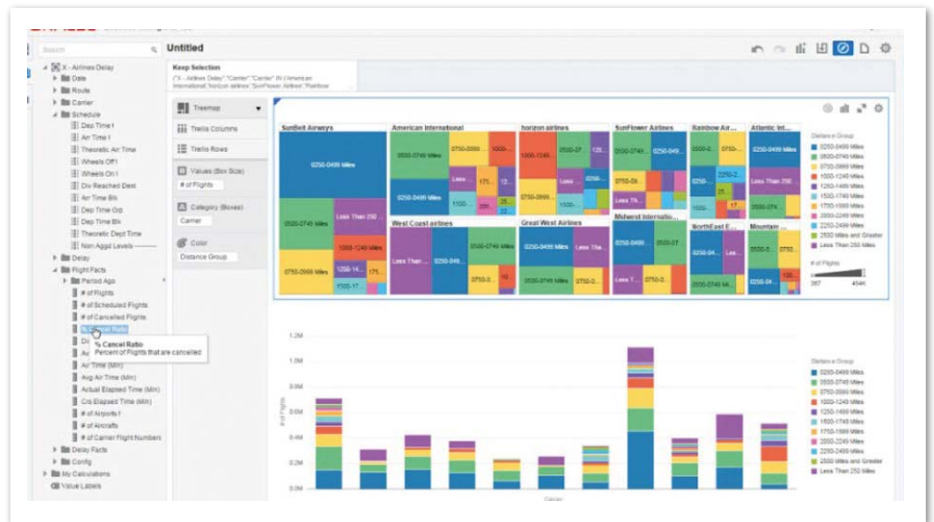


Abbildung 8: Anwendungsbeispiel mit dem Visual Analyzer

SELECT revenue, product, brand, XSA('My File.xls').score
FROM ...

- Directly access external data in analyses
 - User uploaded files
 - Any source
- Just-in-time modeling
 - First class measures and attributes
 - Mashup across enterprise and user-defined data
 - Private or shared access
- OBIEE embedded data store for user uploaded files
- Transparent caching in a database

Abbildung 9: Zugriff auf externe Dateien aus logischem SQL

```

SELECT
    Product.CategoryID,
    Day.cal_year, Sales.Revenue,
    TRENDLINE(Sales.Revenue, (day.cal_year), 'LINEAR')
FROM
    SubjectArea;

EVALUATE_R('<path to script>', Options, Output
    Column, Logical Column Expr1, Logical Column
    Expr2 ...)
```

Abbildung 10: Advanced Analytics in logischem SQL

naus möchte, kann das Feature „Advanced Analytics“ verwenden (siehe Abbildung 10). OBIEE 12c erlaubt es nun, diese Operationen direkt in der BI-Engine durchzuführen; eine externe Engine muss nicht mehr zur Verfügung stehen. Somit lassen sich die zu verwendenden Kommandos direkt im logischen SQL einbinden.

Fazit

Schon bei der Installation von OBIEE 12c lassen sich erste Unterschiede zum Vorgänger-Release 11g erkennen. Sie werden insbe-

sondere in der Architektur weiter verstärkt. Obwohl dies einige Veränderungen für Administratoren mit sich bringt, überwiegen die Vorteile. Eine der größten Neuerungen dürfte somit das parallele Betreiben mehrerer BI-Anwendungen auf einem BI-System sein. Auch die diversen Abgleichmechanismen, die das Baseline-Validation-Tool bereitstellt, vereinfachen vor allem den Umgang mit der Oracle-BI-Lösung.

Innerhalb der eigentlichen Webanwendung kann vor allem die neue Komponente „Visual Analyzer“ durch Benutzerfreundlich-

keit und einfache Bedienbarkeit überzeugen. Alles in allem kann somit von einer gelungenen und kompletten Produkt-Veröffentlichung gesprochen werden. Quellen: Oracle Workshops und Materialien im Rahmen des OBIEE-12c-Beta-Programms.

David Michel

david.michel@appsassociates.com

DOAG 2016 Logistik – erstmals im Rahmen der CeMAT

Jan Peterskovsky, DOAG Online

Unter dem Leitthema „Smart Supply Chain Solutions“ der diesjährigen CeMAT fügt sich auch die DOAG 2016 Logistik in den Rahmen der Weltleitmesse für Intralogistik und Supply Chain Management ein. Die DOAG 2016 Logistik findet am 2. Juni 2016 im Convention Center der Deutschen Messe in Hannover statt und widmet sich neben Smart Supply Chain Solutions speziell den Technologien für das Internet of Things.

In zwei parallelen Streams dreht sich alles um die Herausforderungen der Logistik im Zeitalter der Digitalisierung, sowie um aktuelle Systeme und Lösungen. Unter anderem Augmented Reality, cyber-physische Sys-

teme und Warehouse Management in der Cloud. In verschiedenen Praxis-Workshops lernen die Teilnehmer aktuelle Systeme aus erster Hand kennen und tauschen sich aktiv über die Entwicklungen der Branche aus. So thematisiert etwa Kai Hussong von promatis software richtungsweisende Methoden für das Management kollaborativer Geschäftsprozesse im Digital Workplace. Leonid Poliakov, Ubimax GmbH, demonstriert anhand konkreter praktischer Beispiele, welchen Mehrwert moderne Wearable Computing-Technologie schon heute liefern kann.

Die Referenten sind allesamt namhafte Experten aus Wirtschaft und Wissenschaft. Die

Opening-Keynote hält Prof. Dr. Franz Valle, Gründer und Gesellschafter der Unternehmensberatung Vallée und Partner sowie erster Stiftungsprofessor der Fachhochschule Münster, zum Thema „Herausforderungen Logistik 2025“. Prof. Dr. Heike Simmet, Professorin für Betriebswirtschaftslehre an der Hochschule Bremerhaven, zeigt Erfolgspotenziale der Sharing Economy als Business-Konzept in der Logistik auf. Ein anschließendes Get-Together bietet den passenden Rahmen für entspanntes Networking und fachlichen Austausch. Weitere Informationen zur DOAG 2016 Logistik gibt es unter „<http://logistik.doag.org>“.

CeMAT – die Weltleitmesse für Intralogistik und Supply Chain Management

Logistik und IT verschmelzen zunehmend. Diesem Trend wird die CeMAT, die Weltleitmesse für Intralogistik und Supply Chain Management, seit über zehn Jahren gerecht. Beim letzten Mal kamen rund 48.000 Besucher nach Hannover – Tendenz steigend. In diesem Jahr steht die Messe unter dem Leitthema „Smart Supply Chain Solutions“ – ein Schwerpunkt liegt auf der IT. Denn ausgereifte Lösungen sind inzwischen die Schnittstel-

le zwischen allen Bereichen der hochautomatisierten Logistik. Die CeMAT gliedert sich in fünf Ausstellungsbereiche: „Empack und Label&Print“, „Move & Lift“, „Store & Load“, „Manage & Service“ und „Logistics IT“. Auf letzterem dreht sich alles um Logistik 4.0: intelligente Lösungen für vernetzte Behälter und Transportsysteme, sich selbst steuernde Prozesse und intelligente Lager. Im Fokus stehen dabei unter anderem effiziente

Logistiksoftware mit Cloud-Anbindung und integrierte Systeme für Automatisierungsprozesse. Für IT-Experten gibt es mit intelligenten Algorithmen, Apps zur Prozesssteuerung und anderen innovativen Technologien auch in den anderen Ausstellungsbereichen viel zu entdecken. Die CeMAT findet vom 31. Mai bis 3. Juni 2016 in der Deutschen Messe Hannover statt. Weitere Informationen unter „<http://www.cemat.de>“.

Data Warehouse from Scratch – mit ODI 12c

Bernhard Rosenberger, MT AG Frankfurt

Oracle hat mit dem Data Integrator 12c (ODI) ein vielversprechendes ETL-Werkzeug als OWB-Nachfolger im Portfolio. Kann der ODI den Anforderungen an ein auf der grünen Wiese zu errichtendes DWH genügen? Und wird er in der aktuellen Version 12c die Nachfolge-Versprechungen gegenüber dem guten alten Warehouse Builder (OWB) erfüllen? Ein Projektbericht.

Der Kunde des Autors, die GFKL Financial Services AG, ist einer der führenden Finanzdienstleister für Forderungsmanagement. Das Unternehmen ist aufgrund einer großen Vielfalt zu verwaltender Einzeltransaktionen auf effizient strukturierte Datenhaltung angewiesen, um eine kundenorientierte, schlanke Abwicklung seines Geschäfts sicherstellen zu können. Ganz oben auf der Agenda des Kunden stand die Modernisierung des Berichtswesens. Dieses basierte auf einer in die Jahre gekommenen, historisch gewachsenen heterogenen ETL-Landschaft. Folgende Anforderungen wurden an ein Nachfolgesystem gestellt:

- Zentral organisierte Datenbereitstellung
- Vereinheitlichte Semantik
- Feingranulare Datenablage
- Offene Architektur
- Zukunftsfähige Technologie
- Tägliche Beladung sicherstellen
- Wiederaufsetzbarkeit von ETL-Schritten

Um die IT-Infrastruktur auf lange Sicht strategisch auszubauen und eine flexible Erweiterbarkeit zu ermöglichen, hat sich die GFKL dazu entschieden, ein Data Warehouse als Basis-Plattform für ein unternehmensweites Berichtswesen neu zu entwickeln. Eine der Herausforderungen des Projekts bestand darin, die aus dem breit gefächerten Leistungs-Portfolio des Unternehmens sich ergebende Datenbasis in geeigneter Form vereinheitlicht, aber auch erweiterbar zusammenzuführen.

Tool-Auswahl und Architektur-Entscheidungen

Wegen der im Hause fundiert vorhandenen technischen Kenntnisse und auch der Chance, „auf der grünen Wiese“ beginnen zu können, fiel die Entscheidung, ein DWH mit der aktuellen Datenbank-Version 12c, dem ODI 12c als ETL-Werkzeug, dem SQL Developer und Apex zur Stammdaten-Pflege komplett neu aufzubauen. Zur Unterstützung in Ar-

chitekturfragen und bei der Realisierung wurde die MT AG mit hinzugezogen. Ein gemeinsam durchgeführter Proof of Concept mit BI-Tool-Anbietern destillierte MicroStrategy als Werkzeug der Wahl heraus.

Die fachliche Herausforderung bestand in der semantischen Daten-Integration unterschiedlich ausgerichteter Tochtergesellschaften mit dem Ziel, ein zukunftsfähiges, vereinheitlichtes Berichtswesen mit Dashboards und KPIs für die Geschäftsleitung aufzubauen.

Ziel-Architektur

Die Staging Area unterteilt sich dabei in die klassische „1:1“-Datenübernahme aus den Vorsystemen und einen historisierten Cleansing-Bereich, in dem Datentyp-Konvertierungen und Umschlüsselungen umgesetzt werden (*siehe Abbildung 1*). Die Historisierung der Daten erfolgte nach der von Ralf Kimball als „SCD2“ beschriebenen Methode (Gültigkeit von Datensätzen durch Zeit-Intervall).

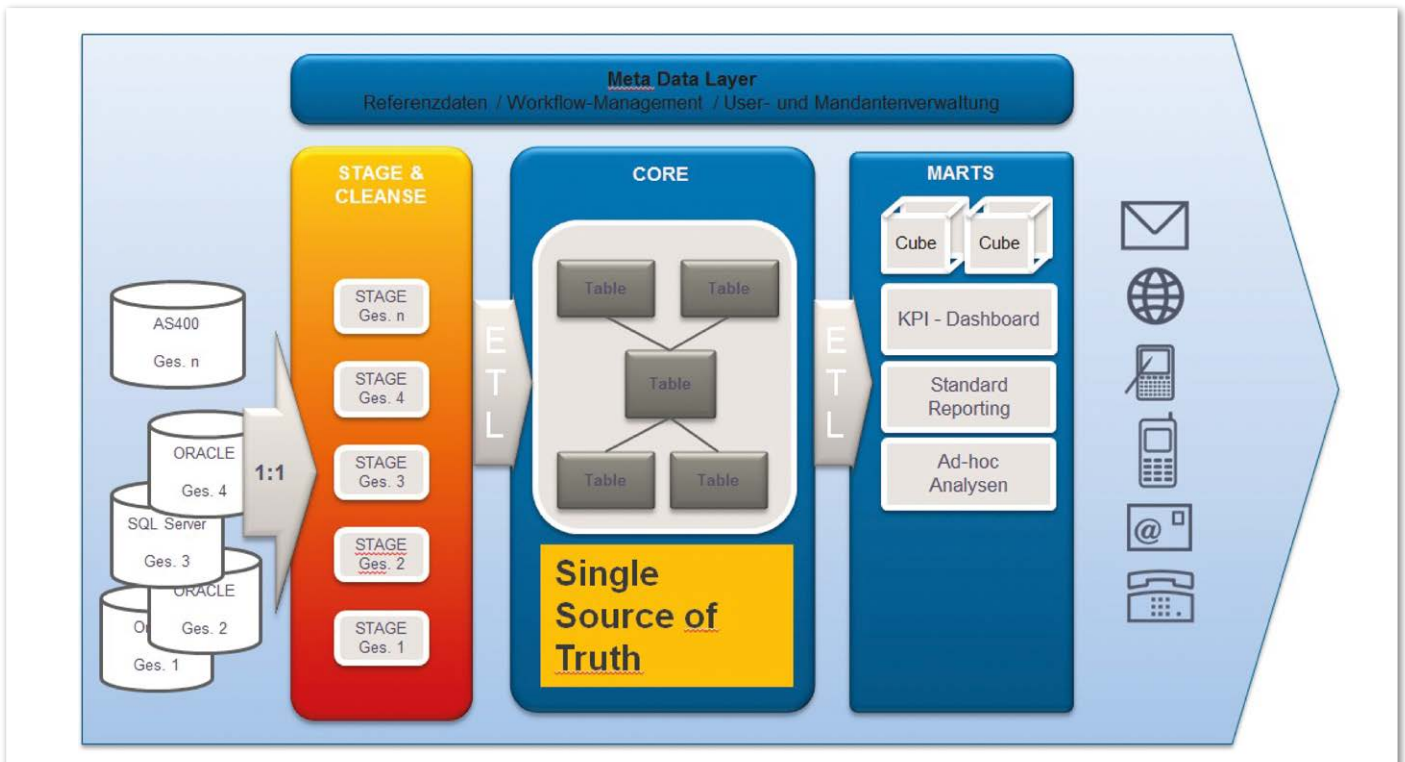


Abbildung 1: DWH Ziel-Architektur

Im Core werden die Daten basierend auf den Ergebnissen der fachlichen Analyse historisiert, strukturiert und integriert gespeichert; sie dienen als Ausgangsbasis zur Erstellung verschiedener Datamarts, die für die jeweiligen Berichte in die Abteilungen und die Geschäftsleitung bereitgestellt werden müssen. Die in ROLAP-Technologie erstellten Datamarts sind als Star- und Snowflake-Konstrukte zur optimierten Verwendung durch das eingesetzte BI-Tool MicroStrategy aufgebaut.

Das Projekt bot die einmalige Chance, die Praxistauglichkeit des ODI 12c jenseits von Labor-Bedingungen eingehend auf Herz und Nieren zu prüfen. Die dabei gemachten Erfahrungen sind nachfolgend näher dargestellt und abschließend im Vergleich zum Vorgänger-Tool OWB bewertet.

Projekt-Methodik

Das Projekt war gekennzeichnet von kurzen Entwicklungsphasen mit klar definierten Zielen. Das Team setzte sich aus Mitgliedern der Fachbereiche, fachlichen Analysten und Entwicklern zusammen. Tägliche Stand-ups und ein offenes Kommunikationsklima förderten den Informationsaustausch besonders auch auf informeller Ebene und ermöglichten ein schnelles Erkennen und Reagieren auf ungeplante Handlungsbedarf. Bei aller Agilität gaben definierte Etap-

penziele dem Team die notwendige Orientierung für ein zielgerichtetes Gelingen des Projekts. Grob skizziert waren die Projektphasen unterteilt in:

- Fachliche Analyse der Geschäftsprozesse und Datenstrukturen
- DWH-Neubau mit semantischer Daten-Integration und Ableitung erster KPI-Reports
- Historisierung des DWH mithilfe der Knowledge-Module
- Einführung des BI-Tools (MicroStrategy), Erstellung von Reports
- Weitere Phasen im Kontext eines regulären Lifecycle-Managements

Installation und Konfiguration der ODI-12c-Umgebung

Da das Projekt auf der grünen Wiese gestartet wurde, gab es zu Beginn Überlegungen, wie der ODI sinnvoll aufgesetzt werden kann. Wie soll die Trennung von Entwicklung und Produktion aussehen, welche Repositories sind erforderlich?

Welche ODI-Agents sollen zum Einsatz kommen? Neben dem bewährten Stand-alone-Agent, seit 11c ein JEE-Agent (Betrieb auf WebLogic Server), steht mit der aktuellen 12c ein Collocated-Agent (Betrieb auf Cloud Control) zur Auswahl. Die Entscheidung fiel zugunsten der schlanken und auch stabilen Lösung mit dem Stand-alone-Agent,

der auch ohne schwergewichtiges Verwaltungswerkzeug auskommt. Mehrere mögliche Repository-Konfigurationen waren zu bewerten, um im Umfeld der GFKL ein Master-Repository mit vier Work-Repositories als sinnvollste Konfiguration zu identifizieren.

Aufbau der Staging Area – Automatisierung mit Groovy

Zum Aufbau des Staging mussten vier Vor-systeme mit jeweils fünfundvierzig Tabellen angebunden werden. Um die unvermeidliche Fleißarbeit bei der Implementierung gleichartiger Mappings zu vermeiden, kam die an Java erinnernde Groovy Scripting Language erfolgreich zum Einsatz. Es reichte aus, eine Beladungsstrategie einmalig (etwa Datapump-Load oder Delta-Abzug) zu skripten, um daraus ODI-Ladestrecken für jede beliebige Tabelle automatisch generieren zu können. Das Vorgehen reduzierte die manuellen Aufwände erheblich und darüber hinaus die Fehlerquote bei der Stage-Implementierung.

In ODI Studio 12c ist erstmals eine rudimentäre Groovy-Runtime-Umgebung eingebunden. Listing 1 zeigt ein Snippet, das die intuitive Erlernbarkeit der Skriptsprache erahnen lässt. Der Code generiert in einem Mapping zwei Eingangsverbindungen eines ODI-Set-Operators. Zum schnellen Einstieg in das Thema „Groovy“ gibt es den

```

...
icps = comp_set.getInputConnectorPoints() //2 inp. Conn.points
i = icps.iterator()
icp1 = i.next() // First input connector point
icp2 = i.next() // Second input connector point
comp_1.connectTo(icp1)
comp_2.connectTo(icp2)
...

```

Listing 1

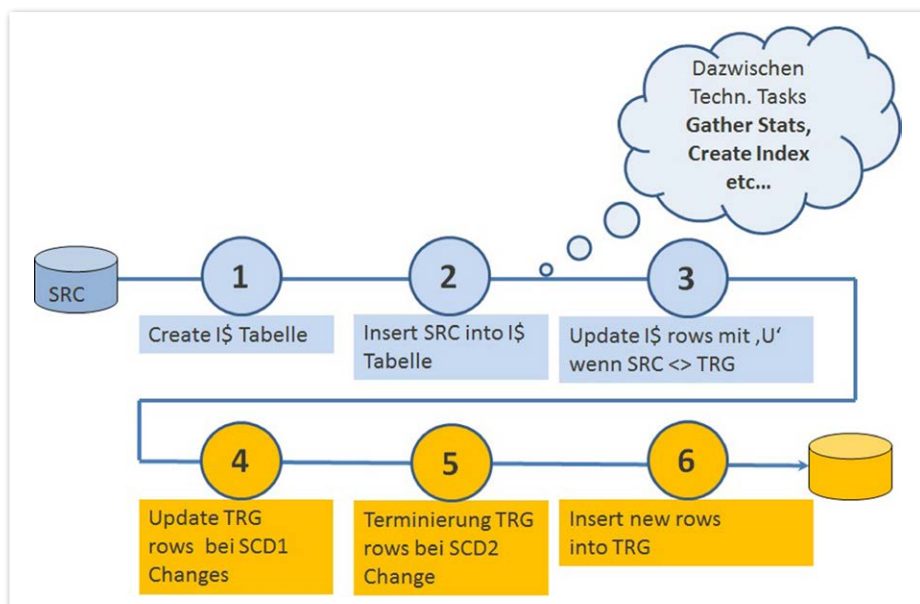


Abbildung 2: SCD2-Knowledge-Modul Kern-Algorithmus

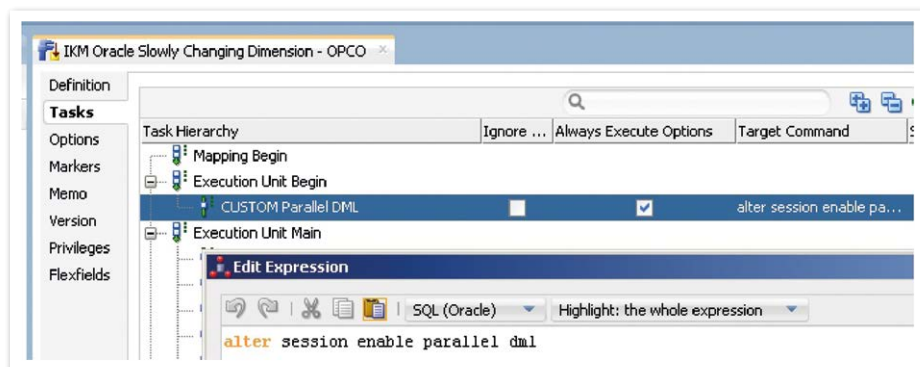


Abbildung 3: KM Custom Task hinzufügen

Oracle-Blog „ODI 12c – Mapping SDK the ins and outs“ [1].

ODI-Mappings im Core

Im Core waren teilweise komplexe Transformationen durchzuführen. Der ODI unterstützt in 12c erstmals neue Operatoren, die zusammen mit den im ODI altbekannten „Set“, „Join“ und „Filter“ eine Realisierung von Transformationen als Mappings

á la OWB ermöglichen. Die neu hinzugekommenen Operatoren „Aggregator“, „Dataset“, „Distinct“, „Expression“, „Lookup“, „Sort“ und „Split“ basieren auf einem neuen Konzept von sogenannten „Component-Style-Knowledge-Modulen“, die zur Laufzeit entsprechend generierte Code-Artefakte beisteuern.

Die komplexeren Mappings wurden aus Zeitgründen zu Beginn basierend auf Da-

tenbank-Views als Datenquelle realisiert, in denen die komplexe SQL-Logik gekapselt ist. Erste Versuche haben gezeigt, dass sich mithilfe der neuen ODI-Operatoren komplexe Mappings ebenso gut realisieren lassen wie mit dem OWB. Damit sind zukünftige View-Kapselungen obsolet. Im ODI-Mapping wären darüber hinaus Objekt-Abhängigkeiten für den Entwickler direkt sichtbar und zusätzlich eine durchgängige Lineage- und Impact-Analyse durch den ODI möglich, die mit dem View-Ansatz unterbrochen wird.

Um die Entwicklungsarbeiten beschleunigt zum Ziel bringen zu können (Quick-Win), ist man beim gekapselten View-Ansatz geblieben, da die View-Darstellung aus den Ergebnissen der Business-Analyse gut abgeleitet werden konnte. Ein Reengineering des Mappings wurde bewusst auf einen späteren Zeitpunkt verschoben.

Knowledge-Module

Als herausragendes Architektur-Feature des ODI stellten sich die Knowledge-Module (KM) heraus. Durch den Einsatz des out of the box mitgelieferten KM für die SCD2-Historisierung konnte man diese Funktionalität im Cleansing und Core enorm schnell prototypisch realisieren. Es genügte, mit Groovy-Skripting ein einfaches Mapping von Quelle zu Ziel zu generieren, diesem das gewünschte KM zuzuweisen und eine SCD-Attribut-Deklaration für Tabellenspalten im Model Editor vorzunehmen. Abschließend lassen sich eventuell noch im Mapping die spezifischen Parameter des KM anpassen. Oracle hat in 12c die Bedienbarkeit des Model-Editors verbessert. Nur ein durchaus umschiffbarer GUI-Bug in der Dropdown-Liste für „SCD Behaviour“ fiel negativ auf.

Erste Tests lieferten so in Kürze brauchbare Ergebnisse. Um die Historisierung an die Projektbedürfnisse anzupassen, wurden folgende zu ergreifende Tuning-Maßnahmen identifiziert:

- Im Quellsystem gelöschte Datensätze sind im historisierten Ziel zu terminieren
- Die Parallelisierung ist datenbankseitig zu aktivieren, da nicht als Default-Einstellung gegeben („alter session enable parallel DML“)
- Intermediär vom KM erzeugte Hilfstabellen sind zu partitionieren
- „Direct Path Loads“ sind durch Hints zu aktivieren

Aufgrund des flexiblen wie mächtigen KM-Konzepts waren diese Forderungen durch Anpassungen und Erweiterungen zügig umsetzbar, wie nachfolgend skizziert wird.

Das SCD2-Integration-KM „IKM Oracle Slowly Changing Dimension“ sieht out of the box fünfundzwanzig konfigurierbare Execution Unit Tasks vor, die von außen über Parameter (Options) steuerbar sind.

Die prinzipielle Funktionsweise lässt sich anhand der sechs wichtigsten Tasks illustrieren (siehe Abbildung 2). Alle weiteren, hier nicht dargestellten neunzehn Tasks sind für technische Schritte wie das Berechnen von Statistiken, das Erzeugen beziehungsweise Abräumen von Hilfstabellen, Indizes etc. notwendig.

Jeder Task im Knowledge-Modul wiederum besteht aus einer Art Skriptcode-Schablone, die zur Laufzeit mit den Metadaten aus dem ODI Repository substituiert wird, um lauffähigen Code zu generieren. Die vier aufgelisteten Anforderungen sind also durch entsprechende Anpassung des Skript-Codes im KM und, wenn nötig, durch das Hinzufügen neuer Mapping-Parameter lösbar.

Nachfolgende Screenshots illustrieren, wie im KM-Editor Tasks für das Terminieren

gelöschter Sätze und für das Ein- und Ausschalten der Parallelisierung hinzugefügt werden. Hinzu kamen neue Parameter (im Editor „Options“ genannt), um diese Tasks aus jedem Mapping heraus spezifisch steuern zu können (siehe Abbildungen 3 – 5).

Aus dem originalen KM abgeleiteter Code hatte eine sofortige Auswirkung auf alle darauf umkonfigurierten SCD-Mappings. Lediglich eine individuelle Anpassung der Mapping-spezifischen Option ist im Bedarfsfall bei gewünschten Abweichungen vom voreingestellten „DEFAULT“-Wert noch durchzuführen. Alle beschriebenen Maßnahmen halfen, die Laufzeiten der Mappings um ein Vielfaches zu reduzieren. Weiterführende Informationen zum Thema stehen im Blog „Maßgeschneidertes und performantes ETL durch ODI-Knowledge-Modul-Modifikationen“ [2].

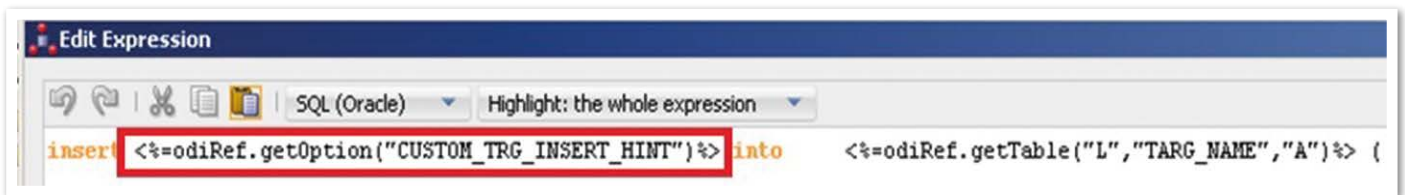


Abbildung 4: KM Optimizer Hint per Option hinzufügen

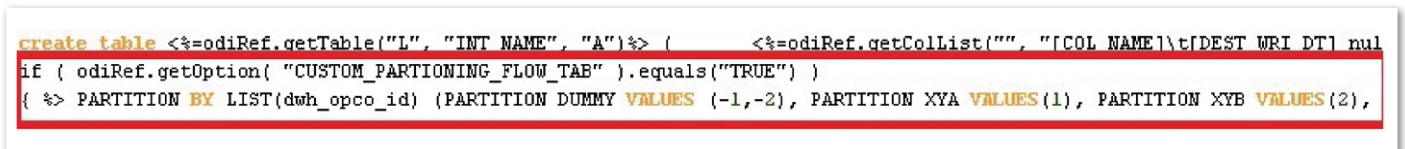


Abbildung 5: KM Customized Partitioning

Steps Hierarchy	Ena...	Scenario/Variable	Restart
root_step	<input checked="" type="checkbox"/>		Restart from failure
PROC_INSERT_DWHBELADUNGE	<input checked="" type="checkbox"/>	PROC_INSERT_DWHBELADU...	Restart from new session
REFRESH_VARIABLES	<input checked="" type="checkbox"/>	REFRESH_VARIABLES Versio...	Restart from new session
Landing	<input checked="" type="checkbox"/>		Restart all children
Cleansing	<input checked="" type="checkbox"/>		Restart all children
Seriiell	<input checked="" type="checkbox"/>		Restart from failure
Seriiell	<input checked="" type="checkbox"/>		Restart from failure
SCN_PKG_Delta2ALL_CL	<input checked="" type="checkbox"/>	SCN_PKG_Delta2ALL_CLN_V...	Restart from new session
SCN_PKG_Delta2ALL_CL	<input checked="" type="checkbox"/>	SCN_PKG_Delta2ALL_CLN_V...	Restart from new session
SCN_PKG_Delta2ALL_CL	<input checked="" type="checkbox"/>	SCN_PKG_Delta2ALL_CLN_V...	Restart from new session
Seriiell	<input checked="" type="checkbox"/>		Restart from failure
SCN_PKG_Delta2ALL_CL	<input checked="" type="checkbox"/>	SCN_PKG_Delta2ALL_CLN_E...	Restart from new session
Seriiell	<input checked="" type="checkbox"/>		Restart from failure
Aufbau Core	<input checked="" type="checkbox"/>		Restart from failure

Abbildung 6: Load Plan – Vierfach-Parallelisierung im Cleansing

Packages und LOAD Plans

Dank Packages und LOAD Plans, die seit ODI 11g eingeführt worden sind, kann auf den Einsatz eines Drittanbieter-Tools zur Jobsteuerung verzichtet werden. Die Orchestrierung des Ladevorgangs im ODI wurde mit den ODI-Elementen „Globale Variablen“, „Packages“, „Procedures“ und „Load Plans“ realisiert. Packages erwiesen sich als nützliches Strukturierungselement, um zusammengehörige Abläufe transparent zu kapseln. Dabei konnte eine sinnvoll parallelisierte Anordnung von Prozessen zu einer insgesamt deutlich optimierten Auslastung der Rechner-Ressourcen und zu reduzierten Ladezeiten beitragen (siehe Abbildung 6).

Apex zum Reporting und zur Metadaten-Pflege

Zur Vereinheitlichung der zum Teil semantisch unterschiedlich interpretierbaren Daten der Vorkomponenten war eine Bereitstellung von Lookup-Tabellen erforderlich, um eine vereinheitlichte Darstellung im CORE ablegen zu können. Die Pflege der fachlichen Codes und deren vereinheitlichte Übersetzung wurden mithilfe einer Apex-Anwendung realisiert und dem verantwortlichen Fachbereich zu Verfügung gestellt.

Um einen geregelten Zugriffsschutz zu erreichen, wurde zur Authentifizierung und Autorisierung in Apex der LDAP-Service des Unternehmens angebunden. Durch Einsatz von Oracle Virtual Private Database ist eine hohe integrierte Datensicherheit für die unterschiedlichen Zugriffsgruppen aus den Fachbereichen sichergestellt.

Als Interimswerkzeug vor Verfügbarkeit des BI-Tools MicroStrategy wurde Apex sogar zur Erstellung von KPI-Reports eingesetzt. Es waren dabei mit vertretbarem Aufwand Berichte realisierbar, die fest vordefinierte Drill Downs ermöglichen.

Aufgetretene Probleme

Konnte der Einsatz neuester Oracle-Technologien zu einem reibungslosen Projektfortschritt beitragen? Die wesentlichen Projektziele wurden im Rahmen des selbst gesteckten Zeitrahmens trotz einiger Schwierigkeiten erreicht. Der eine oder andere Service Request musste gestellt werden und so waren auch neue Versions-Einspielungen in Datenbank und ODI erforderlich, um die Probleme zu beheben.

Es ist positiv hervorzuheben, dass es mit dem ODI keine projektkritischen Komplikationen gab. Im Vergleich zu den OWB-Ver-

sionssprüngen aus der Vergangenheit hat sich ODI 12c in Anbetracht der teils vielfältigen Erweiterungen gegenüber der 11er-Version gut bewähren können – und hat die Erwartungen übertroffen.

Fazit

Wie sind die im Projekt gemachten Erfahrungen im Vergleich zum OWB einzuordnen? In Summe wurde der OWB im Rahmen des Projekts nicht wirklich vermisst. Im Gegenteil, der ODI-komplexe Mechanismus, wie SCD 2, kann elegant in den Knowledge-Modulen kapseln und hilft damit, die Entwicklungszeit zu reduzieren. Eine Bewertung im Einzelnen:

- *Komplexe Mappings mit dem ODI? Geht das?*
Dem ODI ist in 12c aufgrund neu spendierter Component-Style-Knowledge-Module die Fähigkeit gegeben worden, Mappings ähnlich zu strukturieren, wie es im OWB möglich war. Das ist wie im OWB eine Voraussetzung für vollständige Lineage- und Impact-Analysen. Außerdem erzeugt der ODI SQL-Kommandos und keine PL/SQL-Packages wie der OWB. Die Darstellung einzelner SQL-Steps im Runtime Repository erlaubt dabei einfacheres Monitoring.
- *Automatisierte Mapping-Generierung mit Groovy*
Der Einsatz der Groovy Scripting Language ermöglichte eine automatisierte Generierung von Mappings, wie es im OWB mit der Tcl-Sprache OMB*Plus möglich ist. Das an Java erinnernde Groovy überraschte mit einfacherer Handhabbarkeit als die gewöhnungsbedürftige TCL-Syntax und so konnten alle Erwartungen erfüllt und teilweise auch übertroffen werden.
- *Einsatz von Load Plans*
Mit den Elementen „Variablen“, „Procedures“ und „Packages“ lassen sich komplexe Load Plans implementieren. Parallele Prozess-Verarbeitungen sind möglich und so blieben im Projekt keine Wünsche im Vergleich zur Workflow-Technik des OWB offen. Gelegentliche Repository-Korruptionen beim Speichern von Load Plans trübten das Bild allerdings ein wenig.
- *Performance mit dem ODI*
Ein direkter Vergleich ist hier schwer zu treffen. Beide Tools generieren Code, der

auf der Ziel-Datenbank ausgeführt wird. Die KMs haben den Vorteil, dass der generierte Code direkt beeinflussbar ist. So kann dieser bei Bedarf angepasst und so bei Performance-Problemen zielgerichtet reagiert werden. Zu erwähnen sei hier der Umstand, dass die neu eingeführten Component-Style-KMs nicht anpassbar sind. Man darf gespannt sein, ob zukünftige ODI-Versionen hierzu Konfigurationsmöglichkeiten eröffnen.

- *Historisierung out of the box?*
Wie schnell eine SCD2-Historisierung mit KM realisierbar ist, hat positiv überrascht. Diese Technologie hat großes Potenzial, um Komplexität zu reduzieren und damit die Produktivität bei der ETL-Entwicklung zu steigern. Dass die SCD2-Terminierung von gelöschten Quelldatensätzen out of the box nicht als Bestandteil des KM mitgeliefert wird, konnte man durch Erweiterung des KM wieder kompensieren.
- *Metadaten-Repository*
Die Laufzeit-Informationen des ODI-Repository, die über den Operator „Tabulator“ abgefragt werden können, liefern ausreichend Informationen, um Fehler rasch zu identifizieren und zu beheben. Im Gegensatz zum OWB war es im Projekt nicht erforderlich, zusätzliche „Self-made“-Repository-Queries bereitzustellen. Gelegentlich auftretende Repository-Korruptionen, deren Ursache vor der Verfügbarkeit geeigneter Patches nicht geklärt waren, veranlassten das Team, einen Metadaten-Mart aufzubauen, um Runtime-Informationen zu persistieren.

Quellenverzeichnis

- [1] ODI 12c - Mapping SDK the ins and outs: https://blogs.oracle.com/dataintegration/entry/odi_12c_mapping_sdk_the
- [2] Maßgeschneidertes & performantes ETL durch ODI-Knowledge-Modul-Modifikationen: <https://blog.mt-ag.com/index.php/2016/02/11/odi-km-modifikationen>

Bernhard Rosenberger
bernhard.rosenberger@mt-ag.com

Im BI Publisher steckt mehr als erwartet

Frank Effenberger und Janine Lehmann, Robotron Datenbank-Software GmbH

Das Reporting nimmt eine bedeutende Rolle bei der Prozess- und Entscheidungs-Unterstützung in datengetriebenen Unternehmen ein. Der Oracle Business Intelligence (BI) Publisher bietet die Möglichkeit, komplexe Berichtsszenarien über RTF-Templates abzubilden. Dessen Schnittstellen ermöglichen zusätzlich eine Integration beispielsweise in Apex-Anwendungen, um die automatische Generierung von Dokumentationen in anderen Projekten zu unterstützen. Dieser Artikel zeigt diese Integrationsmöglichkeit anhand eines Praxisbeispiels.

Aufgrund der stetig steigenden Datenmengen in Unternehmen und der damit verbundenen Anforderungen an die Datenauswertung ist das Reporting beziehungsweise Berichtswesen in Bezug auf Effizienz und Qualität eine entscheidende Komponente. Ein besonderes Augenmerk der Endanwender liegt auf der Formular- und Dokumenten-Erstellung. Das Layout der Berichte und die zugrunde liegende Datenkomplexität bilden dabei zwei wichtige Aspekte.

Im Layout bestehen die Herausforderungen in der geforderten Einheitlichkeit, Übersichtlichkeit und Exaktheit von Berichten, da diese beispielsweise als Vorlagen oder zur Weitergabe für unterschiedliche Interessengruppen verwendet werden. Bei Dokumenten ist weiterhin auf eine allgemeingültige Struktur von Berichten hinsichtlich Inhaltsverzeichnis, Seitenangaben, Verweisen oder Diagrammen zu achten.

Aus Datensicht muss auf die Datenqualität, Verfügbarkeit und die richtige Modellierung geachtet werden. Dies ist ein wesentlicher Punkt für die zu erstellenden umfangreichen Dokumente oder detaillierten Formulare. Damit sind Layouts und die

steigende Datenkomplexität die wesentlichen Herausforderungen für ein Berichtstool.

Abbildung 1 zeigt, dass der BI Publisher die Bewältigung dieser Herausforderungen unterstützt. An dieser Stelle sei angemerkt, dass der Aspekt der Sprache beziehungsweise Übersetzung von Berichten in diesem Beitrag nicht adressiert, aber bei international zugänglichen Berichten durch den BI Publisher unterstützt wird. Die dargestellte konsequente Trennung von Layout und Daten ermöglicht selbst komplexe Reports.

Komplexe Reports

Der BI Publisher bietet vielseitige Möglichkeiten der Berichterstellung. Um einen Bericht zu erzeugen, ist im ersten Schritt ein Datenmodell zu erstellen, um die individuellen Auswertungsanforderungen abzubilden. Als Datenquelle kann jede JDBC-fähige Datenbank dienen, etwa eine Oracle-Datenbank. Im zweiten Schritt werden SQL-Abfragen erstellt, die für die Darstellung der Daten themenbezogen oder hierarchisch aufgebaut sind. Im dritten Schritt ist ein Layout für den Bericht zu generieren, das sowohl die Datenfelder als auch die Formatierung des Berichts abbildet.

Mit dem im BI Publisher integrierten, benutzerfreundlichen und durch einen Wizard gesteuerten Layout-Editor können die Daten in der Web-Oberfläche ausgegeben und Nutzer-Interaktionen ermöglicht werden.

Die Praxis der vergangenen Jahre zeigte, dass die Web-Oberfläche des BI Publisher für einfache Berichte zwar geeignet, jedoch für die gewünschten komplexen Berichtsszenarien unzureichend ist. Dies ist mit dem Fehlen von benötigten Funktionalitäten zu begründen. Deshalb ist es empfehlenswert, das BI-Publisher-Add-in für Microsoft Word zu verwenden. Dazu muss zunächst der Template Builder installiert werden, mit dessen Hilfe detaillierte Templates im Rich-Text-Format (RTF) möglich sind. Die Verwendung des BI Publisher als strategische Reporting-Lösung im Rahmen spezifischer Anforderungen bietet mit den RTF-Templates eine hohe Flexibilität im Berichterstellungsprozess.

Vor der Erstellung eines Berichts mit dem BI-Publisher-Add-in erfolgt zunächst die umfangreiche Datenmodellierung, die als Grundlage für unterschiedliche Berichte dient, wodurch sich der Aufwand des weiteren Berichterstellungsprozesses verringert. Somit kann, wie im vorherigen Kapitel beschrieben, ein Datenmodell für mehrere Layouts verwendet werden. Man sollte bereits die wesentlichen Berechnungen und Zusammenhänge im Datenmodell abbilden, um die Komplexität bei der Layouterstellung zu minimieren. Für die Erstellung aufwändiger RTF-Templates sind zusätzlich Word-Mittel, Funktionen und Variablen nutzbar. Zur Vereinfachung der Layouterstellung sind Formatvorlagen geeignet. In mehrseitigen Dokumenten kann beispielsweise die Übersichtlichkeit durch eine Referenz der Seitenzahlen im Inhaltsverzeichnis gewährleistet und Datenhierarchien können über Grup-

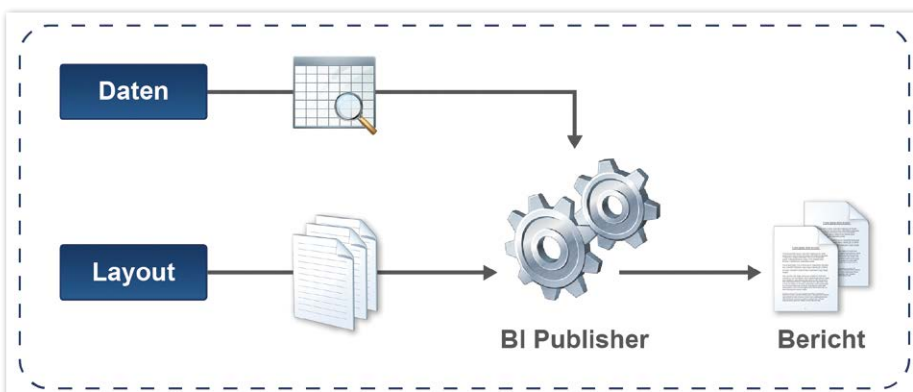


Abbildung 1: Vereinfachte Darstellung der Funktionsweise des BI Publisher

pierungen abgebildet werden. Zudem besteht die Möglichkeit, wiederkehrende und parametergestützte Bereiche eines Berichts in „Sections“ einzuteilen (siehe Abbildung 2).

Als hilfreiches Mittel bei der RTF-Erstellung dienen SQL- und XSL-Funktionen, die in BI-Publisher-Tags eingebunden sind. Als Beschreibungssprache kommen dabei XSL-Formatting Objects (XSL-FO) und XSL-Transformation (XSLT) des BI Publisher im Hintergrund zum Einsatz. Damit ist die Berichterstellung für Endanwender ohne weitreichende XML-Kenntnisse möglich.

Bei der späteren Ausführung des Berichts übersetzt der BI Publisher die Tags der Datenfelder in die standardmäßigen SQL- beziehungsweise XSL-Codes. XSLT lässt sich jedoch auch ohne Tags anwenden. Dies ist beispielsweise über XSL-Templates zu realisieren, die man in RTF-Templates referenziert. Zudem lassen sich auch einzelne Sektionen eines Berichts in andere Templates auslagern, um für weitere zu erstellende Berichte wiederverwendbar zu sein. Somit reduziert sich auch hierbei der Aufwand der Berichtsgenerierung. Darüber hinaus lassen sich Erweiterungen der Darstellungsmöglichkeiten von Diagrammen in Berichten oder die Verwendung von Variablen nur mit XSLT umsetzen. Dadurch ist jedoch ein breites Spektrum an Visualisierungsmöglichkei-

ten möglich, um eine schnellere Umsetzung von Kundenanforderungen zu erzielen.

Durch die Diversität an Ausgabeformaten lassen sich vielfältige Berichte generieren (siehe Abbildung 3). Hierbei kommen die Output-Formate PDF, HTML, Excel, PowerPoint, RTF, TXT oder CSV zum Einsatz. Für die erstellten Dokumente existieren viele Weiterleitungsmöglichkeiten. In diesem Kontext lassen sich beispielsweise Drucker, E-Mail- oder FTP-Server im BI Publisher einbinden und konfigurieren. Als Möglichkeit der umfangreichen Datenauswertung ist zudem die Integrierbarkeit der Dokumente auf Dashboards der Oracle Business Intelligence Suite zu erwähnen. Damit ergibt sich eine Interoperabilität, die die Ad-hoc-Analysen der BI Suite mit den komplexen Dokumenten des BI Publisher parametergesteuert vereint. Zudem ist die Einbindung in andere Anwendungen hervorzuheben, um Berichte entsprechend zu präsentieren. Damit all diese Funktionalitäten auch außerhalb des BI Publisher verwendet werden können, existieren mehrere Schnittstellen. Nachfolgend ist exemplarisch die Webservice-Schnittstelle näher erläutert.

Webservice-Schnittstelle

Der BI Publisher kann über eine Webservice-Schnittstelle von heterogenen Anwendungen

verwendet werden. Der Vorteil der plattform-, programmiersprachen- und protokollunabhängigen Schnittstelle erlaubt dabei eine Vielzahl von Integrationsszenarien auf Basis XML-basierter Nachrichten (SOAP). Zur Nutzung der Webservice-Schnittstelle muss lediglich ein passender Webservice-Request an den BI Publisher gesendet und die Webservice-Response vom empfangenden Programm ausgewertet werden. Es existieren folgende Services, die verwendet werden können:

- *ReportService*
Stellt Methoden bereit, um mit BI-Publisher-Berichtsobjekten zu interagieren
- *ScheduleService*
Stellt Methoden bereit, um BI-Publisher-Scheduler-Aufgaben auszuführen
- *CatalogService*
Stellt Methoden bereit, um BI-Publisher-Katalogobjekte zu verwalten
- *SecurityService*
Stellt Methoden bereit, um Sicherheitsaufgaben zu lösen

Dabei ist es auch möglich, dass die Generierung eines Dokuments anhand geänderter Daten in der Datenbank stattfindet und so ein aktuelles Dokument auf Anfrage hin erzeugt wird (ReportService). Dies ist möglich, weil der BI Publisher zwischen den Daten und der Visualisierung trennt; so kann ein Layout fest hinterlegt, die Daten jedoch dynamisch angepasst und verändert werden.

Einbindung in Apex

Es gibt bereits eine BI-Publisher-Integration in Apex. Dieser Beitrag und die folgenden Ausführungen adressieren jedoch explizit die Webservice-Schnittstelle des BI Publisher, ohne näher auf die Integrationslösung einzugehen. Dadurch wird die Funktionsweise und einfache Verwendung auch für andere Anwendungen verdeutlicht. Eine Apex-Anwendung kann die Webservice-Schnittstelle des BI Publisher verwenden, um etwa auf die Funktionalitäten der Dokumenten-Generierung zurückzugreifen. So kann für den Endanwender ein benötigtes Dokument erstellt werden, ohne dass dieser die eigentliche Anwendung verlassen und beispielsweise den BI Publisher öffnen und benutzen muss.

Die Apex-Anwendung benötigt dabei eine PL/SQL-Funktion, die den Webservice-Request (konkret: Report-Request) für ein

```

<?fo.page-number?> von <?fo.page-number-citation xdofo:lastpage-joinseq?>
<?xdoxslt.sysdate('DD.MM.YYYY')?>

<?for-each@section:BEREICHE?> INHALTSVERZEICHNIS
<?if.count(xdoxslt.distinct_values(BEZ_GESAMT))=1?>


|                  | UEB   | ANZ_TEILBEREICHE | SEITEN   |
|------------------|-------|------------------|----------|
| <b>BEREICH_1</b> | UEB_A | ANZAHL_A         | C3 - 3EC |
|                  | UEB_B | ANZAHL_B         | C3 - 9EC |
| <b>BEREICH_2</b> | UEB_D | ANZAHL_D         | C6 - 6EC |
|                  | UEB_E | ANZAHL_E         | C6 - 6EC |


<?end if?><?end for-each?>

```

Abbildung 2: RTF-Template mit SQL- und XSL-Funktionen, Sections und Variablen

Seite 1 von 1
22.01.2016

Überblick pro Jahr

	Überschrift	Teilbereiche	Seitenzahlen
Aufgabengebiet 1	Gebiet A	2	2 - 3
	Gebiet B	2	3 - 9
Aufgabengebiet 2	Gebiet D	8	9 - 34
	Gebiet E	5	34 - 56

Abbildung 3: Beispielhaftes Ergebnis eines Reports mit RTF-Template aus Abbildung 2

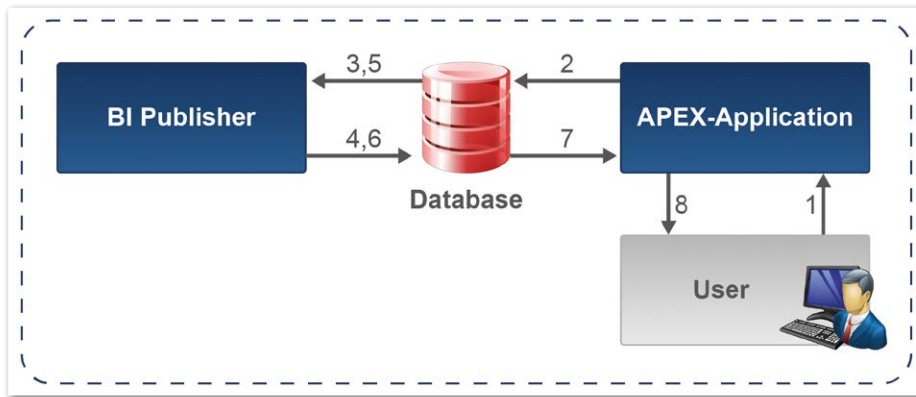


Abbildung 4: Ausführungssicht auf die Architektur für den Apex-Dokumentation-Generator

bestimmtes Dokument an den BI Publisher sendet und als Response eine binäre Datei erhält. Der Header der Datei in der Response muss dann auf das beauftragte Format gesetzt werden (etwa „application/pdf“ bei einem PDF-Dokument), um die Datei zum lesbaren Download bereitzustellen. Die Definition des Headers hängt dabei von der Art des angeforderten Dokuments ab. Das Zusammenwirken des BI Publisher und eines Apex-Projekts wird nun anhand eines konkreten und sich im Einsatz befindlichen Beispiels erklärt.

Automatisierte Generierung von Dokumentationen

In einem Software-Projekt, das auch nach der initialen Produktübergabe weiterhin gewartet, entwickelt und damit vom Auftragnehmer unterstützt wird, finden in Absprache mit dem Auftraggeber regelmäßig Änderungen statt. Dadurch sind ursprünglich erstellte Anwender-Handbücher und Software-Dokumentationen meist schnell veraltet und haben nur noch eine bedingte Gültigkeit.

In Apex-Projekten existiert die Möglichkeit, Hilfetexte zu den jeweiligen Seiteninhalten zu generieren und zusätzlich in einer Datenbank abzulegen. Die Idee des Apex-Dokumentation-Generators liegt darin, diese Hilfetexte zu nutzen, um eine aktuelle und valide Anwender-Dokumentation für den Endbenutzer bereitzustellen. Zeitgleich kann dieses Werkzeug auch während der verschiedenen Entwicklungsphasen und Meilensteine vor der ersten Übergabe des Produkts genutzt werden, um intern die Verständlichkeit und Einheitlichkeit der Hilfetexte zu kontrollieren. Dadurch findet eine enge Kopplung der Anwender-Dokumentation an die Entwicklung statt und Änderungen müssen nur an einer zentralen Stelle anstatt an zwei Orten (Hilfetext

und das Dokument) gepflegt werden. Dies ist vor allem bei umfangreichen Software-Projekten von Vorteil.

Abbildung 4 zeigt die Architektur der Anwendung aus Sicht der Ausführung, bestehend aus einer Oracle-Datenbank mit installiertem Apex und dem BI Publisher. Die angegebenen Zahlen stellen die Schritte dar, die eine aktuell gültige Erstellung einer Anwendungsdokumentation erklären.

Schritt 1 stellt eine Dokumentationsanfrage des Nutzers dar, die in der Regel durch eine Schaltfläche in der Apex-Anwendung realisiert ist. Die Betätigung der Schaltfläche veranlasst die Apex-Anwendung dazu, intern PL/SQL-Funktionen und Prozeduren aufzurufen (Schritt 2). Der nächste Schritt beschreibt den innerhalb einer Funktion angestoßenen Webservice-Request (mit der „Make_Request-Funktion“ des Package „Apex_Web_Service“) an den BI Publisher, mit dem die aktuelle Anwenderdokumentation als PDF bereitgestellt werden soll.

Der BI Publisher führt im vierten Schritt eine SQL-Query gegen die Datenbank aus und bekommt im fünften Schritt die aktuellen Hilfetexte in einer zuvor bereitgestellten Datenstruktur als Ergebnis der Abfrage zurück. Das PDF wird mit den erhaltenen Daten nun finalisiert und als Webservice-Response im sechsten Schritt zurück an die Datenbank-Funktion gegeben. Durch die Prozeduren und Funktionen wird nun von der Web-Oberfläche ein Dialog zum Download für das PDF bereitgestellt. Der User hat nun die Möglichkeit, die aktuelle Anwender-Dokumentation herunterzuladen und auf seinem Arbeitsgerät zu speichern.

Aus Sicht der Daten-Modellierung existieren drei Tabellen und zwei Views, die als Daten-Grundlage für den BI Publisher dienen. Die Tabelle „Kapitel“ stellt eine hierarchische Definition der verschiedenen

Kapitel der Anwender-Dokumentation dar. Im BI Publisher stellt die Tabelle „KAPITEL_CBP“ eine View auf die Tabelle „Kapitel“ mit einem „Connect By Prior“-Befehl (CBP) dar. Da der BI Publisher den CBP nicht alleine ausführen kann, ist dieser Umweg über die View notwendig.

Jedes Kapitel hat eine individuelle Zuordnung zu verschiedenen Inhalten. Die Beziehung ist über die Tabellen „KAPITEL_ZUORDNUNG“ und „INHALT“ dargestellt. Dieses generische Vorgehen erlaubt es, an jeder Position der Kapitel dynamische Inhalte in Form von Grafiken oder Texten zu hinterlegen. Dabei wird auch HTML- und CSS-Code berücksichtigt, der zum Beispiel in der Apex-Anwendung verwendet wird, um ein einheitliches Layout bereitzustellen. Um dies umzusetzen, besteht die Tabelle „INHALT“ aus einer View, in der Inhalte der Texte mit einer CSS der Apex-Anwendung verknüpft werden, damit Tabellen-Inhalte wie in der Anwendung dargestellt werden.

Auf Seite des RTF-Templates ist es aufgrund der dynamischen Erstellung der Anwender-Dokumentation empfehlenswert, das Deckblatt individuell je nach Kunde zu gestalten und das Template entsprechend anzupassen. Die Funktionalität des Inhaltsverzeichnis muss mithilfe des BI-Publisher-Add-Ins zusätzlich eingebaut werden, ebenso wie automatisierte Referenzen in den Kopf- und Fußzeilen.

Fazit

Der BI Publisher bietet eine Vielzahl von Funktionen, damit selbst komplexe Berichtsszenarien abgebildet werden können. Da für diese Szenarien die Web-Oberfläche des BI Publisher nicht ausreichend ist, muss auf RTF-Templates zurückgegriffen werden. Die Webservice-Schnittstelle des BI Publisher ermöglicht es, diese Funktionalitäten an andere heterogene Anwendungslandschaften zu koppeln. So sind beispielsweise automatische Dokumentations-Generierungen in Apex-Projekten möglich. Dies wurde anhand eines beim Kunden eingesetzten Projekts zur Erstellung eines stets aktuellen Anwenderhandbuchs demonstriert.

Frank Effenberger
frank.effenberger@robotron.de

Janine Lehmann
janine.lehmann@robotron.de

Business Intelligence In-Memory

Annett Thurm-Meyer ORACLE Deutschland B.V. & Co. KG

Business-Anforderungen werden komplexer, Zeitfenster kürzer, Ressourcen knapper. Der Schlüssel zu erfolgreichem Business Intelligence ist die zeitnahe Gewinnung von Einblicken und Erkenntnissen, die entscheidungsrelevante Informationen liefern.

Performance ist neben den funktionalen Innovationen die Top-Anforderung im Umfeld von Business Intelligence. Während komplexe Berichte und Analysen vor geraumer Zeit noch Stunden, wenn nicht Tage in Anspruch genommen haben, ist heute die Verarbeitung derselben Berichte und Analysen In-Memory, also im Hauptspeicher, in kurzer Zeit möglich. Damit kann Business Intelligence auch als Entscheidungsunterstützung für operative Prozesse, ja sogar near Realtime genutzt werden. Der Artikel zeigt, wie Business Intelligence mit In-Memory-Technologie zusammenarbeitet.

Wie Business Intelligence In-Memory funktioniert

Die für Business Intelligence genutzte In-Memory-Funktionalität bezieht sich bei Oracle hauptsächlich auf ein Datenbank-Management-System, das auch den Arbeitsspeicher eines Servers als Datenspeicher nutzt. Herkömmliche Systeme halten die Daten auf Festplatten. Dabei kostet der ständige Lesezugriff auf die Festplatte Zeit, die auch durch schnellere Platten nicht beliebig beschleunigt werden kann.

Die Zugriffsgeschwindigkeiten und Algorithmen für den Zugriff auf Daten im Arbeitsspeicher sind deutlich einfacher und damit performanter. Durch die gesunkenen Preise für Arbeitsspeicher ist es heute sinnvoll möglich, Daten im Arbeitsspeicher zu halten. Reicht ein Arbeitsspeicher nicht aus, können herkömmliche Technologien wie Grid Computing dazu verwendet werden, die Datenhaltung auf die Arbeitsspeicher mehrerer Server zu verteilen.

Die traditionelle relationale Oracle-Datenbank hat die Daten in Zeilen gespeichert. Jeder neue Datenbank-Eintrag wird als weitere Zeile abgespeichert. Jede Zeile ist dann wiederum in mehreren Spalten gegliedert, die jeweils Attribute beinhalten.

Diese Art der Datenhaltung ist perfekt für die Speicherung von operationalen Daten (OLTP) geeignet. Neu ist die Datenhaltung in Spalten („Columns“). Sie speichert jedes Attribut einer Transaktion oder eines Datenbank-Eintrags in einer separaten Spaltenstruktur.

Das Spaltenformat ist ideal für Analysen, weil es ein schnelleres Auslesen der Daten erlaubt, wenn nur einige Daten und nicht ganze Zeilen aus einem großen Datenbestand ausgelesen werden müssen. Logisch ist der Umkehrschluss, dass sich eine spaltenbasierte Haltung von Daten nicht für DML („einfügen“, „update“ oder „löschen“) eignet. Die In-Memory-Option der Oracle-Datenbank ermöglicht, das Beste aus beiden Welten zu nutzen und damit Oracle Business Intelligence optional zu unterstützen (siehe Abbildung 1).

Sobald eine In-Memory-Speicherung für Tabellen, Materialized Views oder auch nur ausgewählte Spalten aktiviert ist, erzeugt

die Oracle-Datenbank automatisch eine In-Memory-Kopie der Daten und hält diese synchronisiert.

Diese In-Memory-Daten stehen dann für hoch performante Analysen und Berichte zur Verfügung. Dabei ist es nicht erforderlich, die analytische Applikation anzupassen. Der Optimizer der Datenbank entscheidet, aus welchem Speicher die Daten schnellstmöglich geliefert werden können, um die Anfrage des BI-Servers zu bedienen.

Exalytics – die Business Intelligence In-Memory Machine

Oracle hat das Thema „Business Intelligence In-Memory“ mit extrem leistungsstarker Hardware zu einem Engineered System gebündelt, der Exalytics-Maschine. Dies hat neben der besseren Performance auch noch niedrigere Total Costs of Ownership zur Folge, da das komplette System zusammen gewartet und gepatcht wird. Daneben wird auch noch das operative Risiko

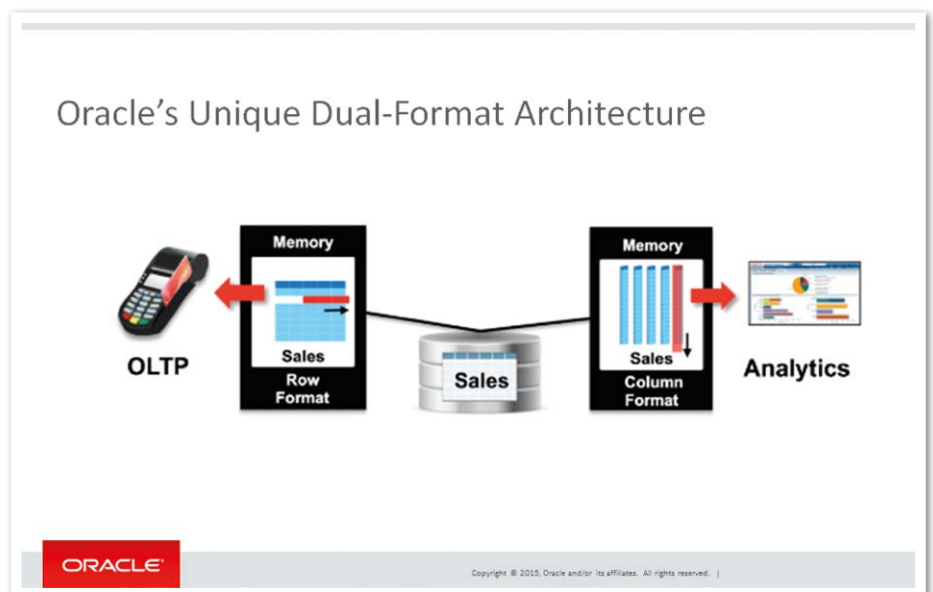


Abbildung 1: Dual-Format Architecture

deutlich gesenkt, da alle Komponenten aufeinander abgestimmt sind.

Auf diesem für die analytischen Anforderungen optimierten System kann nun neben der eigentlichen BI-Software (Business Intelligence Suite Foundation Edition) auch die Datenbank 12c mit der Database-In-Memory-Option installiert sein. Außerdem kann Oracle auch für die Enterprise-Performance-Management-Software verwendet werden. Die Exalytics kann somit nicht nur alle analytischen, sondern auch alle Anforderungen rund um das Thema „Enterprise Performance Management“ bedienen. Dazu zählen Themen wie „Ertragsmanagement“, „Umsatzmanagement“, „Nachfrage-Forecast“, „Inventarmanagement“, „Preisoptimierung“, „Profitabilitätsmanagement“ und „rollierende Forecasts“. Die Exalytics-Maschine kann wahlweise mit weiteren Engineered Systems kombiniert werden. Exadata ist für die Haltung von großen Datenbanken und Exalogic für den Betrieb der Middleware optimiert. Es kann Designgründe dafür geben, dass eine Kombination dieser verschiedenen Engineered Systems bedeutende Vorteile für das Unternehmen bringt.

Die Nutzen-Perspektive für den Anwender

Für den Fachanwender ändert sich mit Exalytics nichts an der gewohnten Arbeitsumgebung (identische Software), das BI-System ist jedoch deutlich schneller als zuvor. Die Benutzer-Akzeptanz verbessert sich stark, denn eine hohe Analyse-Performance ist das wichtigste Akzeptanz-Kriterium eines BI-Systems, wenn Informationsrelevanz und Datenqualität als gegeben gelten. Geschwindigkeit hat allerdings per se noch keinen Wert. Konkreter geschäftlicher Nutzen aufgrund der Performance entsteht aber bei folgenden Anforderungen aus der BI-Praxis:

- Bisher dauerte eine Ad-hoc-Analyse eine halbe Stunde und länger. Nun benötigt sie nur noch einen Bruchteil der Zeit, es sind also zum einen schneller Ergebnisse da, zum anderen besteht für den Fachanwender nun die Möglichkeit, tiefergehend in die Analyse einzusteigen und so neue, zusätzliche Erkenntnisse zu gewinnen.
- Bislang war die Analyse-Verknüpfung so komplex, dass selbst nach einem Tag

Laufzeit die Berechnung noch nicht beendet war. Der Prozess musste abgebrochen werden. Alternativ war diese Analyse schrittweise manuell durchzuführen, indem Einzel-Ergebnisse ermittelt und miteinander verbunden wurden. Wenn eine solche komplexe Analyse mit der Exalytics nun zuverlässig in beispielsweise zwei Stunden ohne jeglichen manuellen Aufwand durchläuft, hat das einen hohen Geschäftswert. Mit der Exalytics sind Analysen durchführbar, die vorher aufgrund ihrer Komplexität nicht umzusetzen waren.

- Das BI-System skaliert aufgrund der Hardware-Ausstattung der Exalytics sehr gut und kann deutlich mehr internen Benutzern zur Verfügung gestellt werden, was die Transparenz im Unternehmen verbessert.
- Erst durch die sekundenschnelle Analyse-Antwort werden komplett neue Geschäftsmodelle im Internet abbildbar, zum Beispiel indem in einem Webshop eine Online-Analyse dem Kunden direkte Berechnungsergebnisse bezogen auf seine aktuelle Auswahl zeigt: Entweder ist diese Analyse sofort da (Realtime) oder das ganze Angebot ergibt keinen Sinn, da selbst fünf Sekunden Wartezeit im eCommerce-Umfeld aus Kundensicht unakzeptabel sind.
- BI-Angebote in der Cloud nehmen zu, flankiert von klar formulierten Service Level Agreements (SLA). BI-Cloud-Anbieter können sich mit der Exalytics Wettbewerbsvorteile verschaffen, da sie sehr schnelle Performance zusagen können.
- Im Big-Data-Kontext sind Realtime-Filterungen der eingehenden Massendaten wichtig, um schnell die relevanten Daten zu identifizieren und diese direkt weiter zu analysieren. Hier ist höchste Performance im Datenstrom ein Muss.

Die Nutzen-Perspektive für die IT

In Data-Warehouse-Projekten ist die Optimierung der Analyse-Performance ein Dauerthema: Aus Anwendersicht stehen die Analyse-Ergebnisse nie schnell genug zur Verfügung. Exalytics bietet in vielen Fällen einen erheblichen Performance-Schub – ohne oder mit nur minimalem personellen Aufwand. Exalytics ist somit

für das IT-Projektmanagement interessant, wenn die hundert Personentage, die für die DWH-Performance-Optimierung (zusätzliche Materialized Views, Test von OLAP-Würfeln etc.) eingeplant waren, nun entfallen können. Bestehende Oracle-BI-Kunden haben zusätzlich den Vorteil, dass sie mit ihrer bestehenden Oracle-BI-Umgebung nahtlos auf die Exalytics-Plattform umziehen können. Das offerierte „All-in-one-patching“ zum Projektstart der Exalytics und die quartalsweisen Updates bedeuten eine Risiko- und Aufwands-Minimierung im BI-Betrieb.

Monitoring und Überwachung der Exalytics-Maschine mit allen Server-Komponenten erfolgen mittels Cloud Control (von der Hardware bis zum Dashboard). Hier zeigt sich der Vorteil eines optimierten, ausbalancierten Gesamtsystems, die Komplexität wird wesentlich leichter beherrschbar; zwanzig bis vierzig Prozent Aufwandsersparnis gegenüber traditioneller Vorgehensweise sind nachweisbar.

Hauptmerkmale von In-Memory Business Intelligence

Die Exalytics-In-Memory-Technologie erlaubt Terabytes von Daten direkt im Arbeitsspeicher, um sie dort schnell zu berechnen und zu analysieren. Darauf basierend wird In-Memory Business Intelligence dann durch In-Memory für Exalytics und In-Memory Essbase bereitgestellt. Diese Datenmanagement-Systeme können in unterschiedlichen Ausprägungen, je nach Datenvolumen und -nutzung, durch vier unterschiedliche Techniken umgesetzt werden. Im Ergebnis erwarten Arbeitsgruppen, Abteilungen oder ganze Unternehmen Business Intelligence mit extremer Performance, um ihre Anforderungen bestmöglich zu unterstützen:

- *In-Memory Datenreplikation*
Viele unternehmensweite Business-Intelligence-Implementierungen können inklusive vordefinierten Oracle-Business-Intelligence-Applikationen komplett In-Memory gehalten werden. Dazu werden die 12c-Datenbank mit der In-Memory-Option direkt auf der Exalytics installiert und alle Daten des Data Warehouse im Speicher vorgehalten. Dieser Mechanismus erlaubt die In-Memory-Verarbeitung für das gesamte Spektrum aller Business-Intelligence-Analysen und weist den höchsten Grad der Flexibilität auf.

Oracle Exalytics X Series

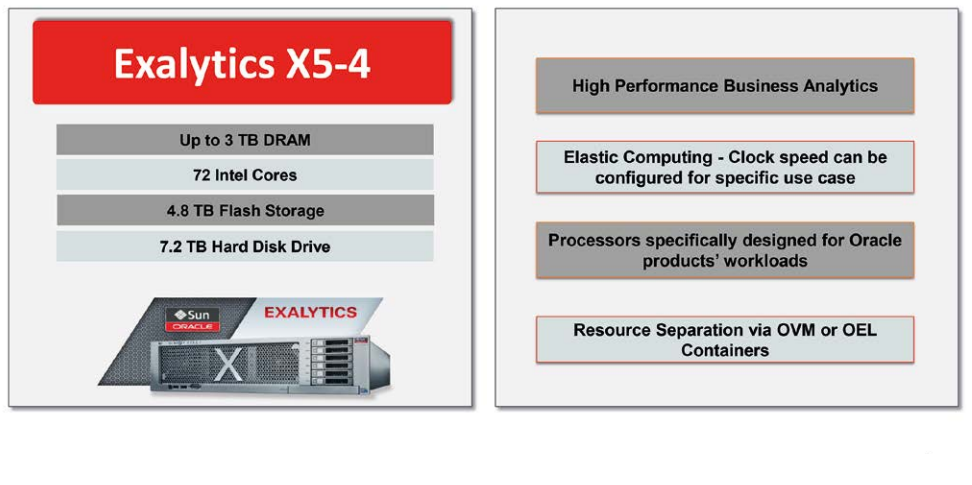


Abbildung 2: Oracle Exalytics X5-4 Maschine

- In-Memory Adaptive Data Marts**
 Die meisten Business-Intelligence-Applikationen haben Auslastungsmuster, die auf die Sammlung von spezifischen, besonders relevanten Daten aus dem Unternehmens-Data-Warehouse fokussieren. In diesem Fall ist es sinnvoll, aus diesen relevanten Daten einen Data Mart In-Memory zu definieren. Diese Vorgehensweise kann Performance-Steigerungen um den Faktor zwanzig realisieren.
- In-Memory Intelligent Result Cache**
 Exalytics Result Cache ist ein vollständig wiederverwendbarer In-Memory Cache, der mit Ergebnissen von bereits getätigten logischen Abfragen quasi automatisch befüllt wird. Jedes Abfrage-Ergebnis wird im Ergebnis-Cache gehalten und bedient ähnliche Abfragen, auch wenn nur eine Untermenge der gesamten Datenmenge benötigt wird. Zur Umsetzung einer optimalen Abfragebeschleunigung stellt Exalytics Werkzeuge zur Verfügung, um die Nutzung von Tabellen zu analysieren, die Notwendigkeit zur Vorab-Befüllung von Ergebnis-Caches zu identifizieren und automatisch auszuführen. Gerade die Vorab-Befüllung von Ergebnis-Caches („Cache Seeding“) erlaubt optimales Antwortverhalten für Abfragen zur Laufzeit.
- In-Memory Cubes**
 Essbase liefert mit seiner In-Memory-Optimierung für Exalytics eine weitere Opti-

on zur Beschleunigung von Abfragen auf festgelegten Geschäftsbereichen (relevante Ausschnitte vom Gesamt-Datenmodell, die speziellen analytischen Anforderungen gerecht werden). Der Unterschied zu anderen Formen von In-Memory-Techniken ist der Umstand, dass hier auch zurückgeschrieben werden kann. Der BI Server für Exalytics baut diese Cubes auf, indem Daten über die semantische Schicht extrahiert werden. Notwendig und hilfreich sind diese multidimensionalen Cubes hauptsächlich für Advanced Scenario Modelling und „What If“-Analysen.

Es gibt also unterschiedliche Design-Möglichkeiten, Daten optimiert für Business Intelligence zur Verfügung zu stellen. Dabei unterstützen die Werkzeuge der Exalytics-Maschine und automatisieren Prozesse weitestgehend.

Beispiel für eine Konfiguration

Eine Exalytics X5-4 verfügt über die Fähigkeit, Business Intelligence durch In-Memory-Technologie zu beschleunigen und Daten in In-Memory Caches zu speichern. Außerdem kann eine Oracle Essbase als multidimensionale Datenbank gemäß ihrer Spezifikation beschleunigt werden, da diese ebenfalls für die Nutzung auf der Oracle Exalytics In-Memory Machine optimiert ist. Hinzu kommt die Fähigkeit, In-Memory zu cachen. Ferner kann auch die Pixel-perfekte Berichtskomponente BI Publisher auf der Exalytics-Maschine beschleunigt werden.

Die Exalytics Maschine unterstützt Business Intelligence mit 72 Cores, 2,6 GHz Intel Xeon Prozessor E7-8895 v3, speziell für Oracle hergestellt, mit bis zu 3 TB Speicher, 4,8 TB Solid-State Drive NVM Express (NVMe) Flash und 7,2 TB Festplatten-Laufwerken. Die Exalytics X5-4 liefert die perfekte Balance von Computing Power, Speicher sowie I/O Footprint (siehe Abbildung 2).

Zu der Maschine gehören im Kontext von Business Intelligence folgende Softwarepakete:

- Business Intelligence Suite Foundation wEdition**
 Dieses Softwarepaket liefert umfassende und vollständige Business-Intelligence-Unterstützung inklusive Unternehmensreporting, Dashboards, Ad-hoc-Analysen, Essbase (multi-dimensionaler OLAP Server), Scorecarding und Advanced Analytics auf einer integrierten Plattform.
- Database-12c-In-Memory-Option (optional)**
 Bei diesem Ansatz werden alle zu analysierenden Daten direkt auf der Exalytics im Speicher gehalten, um optimale Performance zu gewährleisten.

Fazit

Oracle Business Intelligence In-Memory ist im Kern die Fähigkeit des BI Servers, Daten aus einer spaltenbasierten Datenhaltung zu lesen. Wirkliche Effizienz-Steigerungen sind aber nur über die Nutzung eines Engineered Systems, der Exalytics-Maschine, sinnvoll zu realisieren, da hier diverse wichtige Mechanismen und Techniken zur Verfügung stehen, die, je nach Datenvolumen und Präferenz, die Daten optimal und weitestgehend automatisch aufbereiten und auch synchronisiert halten. Dieses Arbeitersparnis ist mit dem Einsatz von Business Intelligence und eines In-Memory-Datenbank-Management-Systems allein nicht zu realisieren.

Annett Thurm-Meyer
 annett-thurm-meyer@oracle.com

Oracle Predictive Queries: Der einfache Weg zu neuen Erkenntnissen

Reinhard Mense und Marco Nätlitz, areto consulting gmbh

Mit den in der Version 12c enthaltenen „Predictive Queries“ (PQ) erhalten die Anwender die Möglichkeit, explorative Analysen direkt in der Oracle-Datenbank auszuführen, und somit einen einfachen und kostengünstigen Einstieg in die Daten-Analyse. Auf Knopfdruck lassen sich die Daten nach bestimmten Merkmalen untersuchen, um Erkenntnisse über bisher verborgene Chancen und Potenziale zu erhalten. Anhand eines einfachen Beispiels werden die Voraussetzungen und Vorgehensmodelle für den Einsatz von PQ gezeigt sowie Grenzen und Unterschiede zu anderen Methoden und Tools erklärt.

Selten waren sich die führenden Marktforschungs-Unternehmen so einig: Predictive Analytics (PA) wird unisono als der Treiber für mehr Produktivität, bessere Kundenbeziehungen und höhere Umsätze erachtet. PA verheißt effizientere Maschinen-Auslastungen, geringere Ausfälle beziehungsweise Stillstand-Zeiten und eine systematische Identifizierung und Erschließung neuer Marktpotenziale. So ist es kaum verwunderlich, dass nach einer Studie von Pierre Audoin Consultants immerhin 30 Prozent aller Unternehmen Ausgaben für PA-Software in den nächsten beiden Jahren geplant haben. Obschon die Quote hoch erscheint, fragt man sich doch: Warum nur eine Minderheit, warum nur 30 Prozent, was ist mit dem Rest?

PA ist keine Methodik, die sich quasi von allein erschließt und im Handumdrehen oder auf magische Weise revolutionäre Ergebnisse erbringt. Die Analyse sollte vielmehr durch speziell ausgebildete Mitarbeiter vorgenommen werden, die nicht nur über ein tiefgründiges Verständnis der mathematischen Grundlagen verfügen, sondern auch betriebswirtschaftliche Fragestellungen nachvollziehen können. Die Rolle solcher Data Scientists ist ebenso anspruchsvoll wie auch gefragt: Die beiden Autoren T. H. Davenport und D. J. Patil haben diese Aufgabe in den Harvard Business News gar als „the sexiest job of the 21st century bezeichnet.“ Und dieser hohe Anspruch an das Personal erklärt wiederum zumindest teilweise die bisher geübte Zurückhaltung der Unternehmen bei

der Nutzung dieser Technologie. Es fehlen Know-how und eben das geeignete Personal. Viele Betriebe scheuen überdies den hohen Implementierungsaufwand sowie die Komplexität der Anwendungen und befürchten eine zu geringe Datenqualität, wie die Analysten von Pierre Audoin Consultants in ihrer Studie noch herausfanden.

Gerade für jene Unternehmen, die PA bisher eher skeptisch und abwartend betrachtet haben, bieten sich mit den PQ der Version 12c der Oracle-Datenbank die Möglichkeit, PA schnell und ohne großen Aufwand zu testen.

Predictive Queries – Ad-hoc-Analysen der Daten

PQ sind ein neues Feature der Datenbank-Version 12c und erweitern den Funktionsumfang von SQL. Die Anwendung ist recht einfach: Der Datenbank-Programmierer schreibt ein SQL-Script und wendet es auf die Daten an. Dies geschieht in einem Schritt und erfordert keine tiefgreifenden Kenntnisse der internen Abläufe in der Datenbank. Nach Ausgabe der Ergebnisse wird das durch die Datenbank errechnete Modell wieder gelöscht, was auch als „flüchtige“ oder „temporäre“ Modellierung bezeichnet wird. Bei der Analyse wird auf die eingebauten Data-Mining-Algorithmen zugegriffen, ohne dass der Anwender jedoch das vollständige Verständnis aller notwendigen Voreinstellungen besitzen oder das sonst übliche Feintuning der Algorithmen vornehmen muss. Dies schränkt die Bandbreite der Nutzung sicherlich ein,

erlaubt aber auch dem unerfahrenen Anwender erste Gehversuche mit PA, sodass er sich stärker auf die Daten und deren Ergebnisse als auf komplexe Vorarbeiten und Programmierungen konzentrieren muss.

PQ sind in der Lage, die Daten zu partitionieren und die vorher definierten Analysefunktionen und Queries auf diese Teilmengen anzuwenden. Dies geschieht automatisch, ohne dass es weiterer Eingriffe des Programmierers bedarf. Der Ablauf ist dabei wie folgt: Im Zentrum steht die Prozedur „PREDICTION“, die in ein „SELECT“-Statement eingebunden werden kann. Der Prozedur „PREDICTION“ kann nun als Parameter übergeben werden, welche Spalte prognostiziert werden soll und welche Spalten zur Modell-Entwicklung herangezogen werden sollen. Gleichzeitig kann mit „PARTITION BY“ angegeben werden, welche Spalten eine Partitionierung innerhalb der Daten definieren.

Enthält eine Kunden-Tabelle etwa eine Spalte „Geschlecht“ oder „Bundesland“, so könnte die Modell-Entwicklung direkt anhand der partitionierten Daten nach Geschlecht beziehungsweise Bundesland erfolgen. Das Ergebnis der Prognose ist das Result-Set des ausgeführten SQL-Statements. Mit dem so erhaltenen Ergebnis lässt sich auf einfache Weise überprüfen, wie aussagekräftig und nutzbar die Daten wirklich sind. Dazu bietet sich folgende Vorgehensweise an: Zur Validierung der Ergebnisse werden historische Daten in zwei unterschiedlichen Zeiträumen herangezogen. Auf die Werte des ersten, früheren Zeitfens-

Beispiel: Oracle Predictive Queries

Wie kann der Preis eines Weins aus Bordeaux ermittelt werden? Die beliebten und häufig teuren Kennerweine werden unter anderem durch die Verkostung von Experten preislich eingestuft. Alternativ müssten die Winzer entlang der französischen Küste über Wetterbedingungen während der Ernte- und Reifezeit befragt werden, denn die kostbarsten Trauben entwickeln einen besonderen Geschmack während heißer und trockener Sommer.

Der Ökonom und Weinliebhaber Prof. Dr. Orley Ashenfelter verließ sich nicht länger auf die menschliche Intuition der Experten und entwickelte ein Predictive-Analytics-Modell, das die Preise von Bordeaux-Wein vorhersagen kann. Dieses Modell (siehe „<http://www.ft.com/intl/cms/s/0/1e9cb152-5824-11dc-8c65-0000779fd2ac.html>“) sorgte für Aufruhr in der Weinbranche, denn plötzlich konnte der Weinhändler die Preise der Weine ermit-

teln, bevor diese die nötige Reifezeit erreicht hatten (siehe „<https://www.betterment.com/resources/investment-strategy/behavioral-finance-investing-strategy/human-vs-algorithm-investing-a-lesson-from-wine-country>“). Ashenfelter nutzte eine Datenbasis von fünf- und zwanzig Bordeaux-Weinen aus den Jahrgängen 1952 bis 1987. Die abhängige Variable – das, was prognostiziert werden soll – ist der typische Preis (P) im Rahmen von Weinauktionen aus den Jahren 1990 und 1991.

Pro Jahrgang ermittelte Ashenfelter ebenfalls die unabhängigen Variablen „Durchschnittstemperatur während der Reifezeit“ (DTMP), „Regen während der Ernte“ (REGE) und „Regen im Winter“ (REGW). Der logarithmierte Preis und das Alter der Weine sind in *Abbildung 1* eingetragen. Anhand der eingezeichneten Geraden lässt sich bereits ein linearer Zusammenhang erkennen. Das prädikative Modell von Ashenfelter setzt lineare Regression über die Variablen DTMP, REGE und REGW ein, um die abhängige Variable P, den Preis, zu prognostizieren (siehe „<http://www.liquidasset.com/orley.htm>“).

Mithilfe von Predictive Queries ist das Modell von Ashenfelter schnell nachgebaut. Benötigt wird eine Tabelle „Wein“, die die fünf- und zwanzig Jahrgänge inklusive Preis und der Eigenschaften DTMP, REGE und REGW als Spalten enthält. Die Funktion „prediction“ kann nun das lineare Modell für die Vorhersage zukünftiger Weinpreise erstellen (siehe *Listing 1*).

Die Datenbank erstellt nun automatisch ein Modell, das mithilfe der linearen Regression die abhängige Variable „Preis“ anhand der unabhängigen Variablen „DTMP“, „REGE“ und „REGW“ vorhersagen kann. Werden nun weitere Jahrgänge in die Tabelle eingefügt, für die der Preis bisher unbekannt ist, jedoch „DTMP“, „REGE“ und „REGW“ bekannt sind, dann prognostiziert die oben dargestellte Query die Preise in der Spalte „PREIS_VORHERSAGE“. Das so nachgestellte Modell nach Ashenfelter erreicht eine Prognose-Genauigkeit von circa 83 Prozent gegenüber Testdaten. Zu gleichen Ergebnissen und Preisvorhersagen gelangt in unserem Test auch die Statistikumgebung R, die unter Data Scientist die Plattform der Wahl darstellt (siehe „<https://www.r-project.org/>“).

Das Beispiel veranschaulicht die Vor- und Nachteile der Predictive Queries, die ab Version 12c in der Datenbank zur Verfügung stehen. Mit der Funktion „prediction“ sind schnell Predictive-Analytics-Modelle gebaut, die erste Ergebnisse und Analysen von Daten erlauben. Sie können ohne zusätzliche Statistik-Software und Detail-Kenntnisse im Bereich der Predictive Analytics eingesetzt werden. Die Modelle sind allerdings nicht parametrisierbar, weshalb sie nicht besser an den Datensatz angepasst werden können. Auch können die Modelle nicht gespeichert werden, denn die Datenbank hält diese nur während der Ausführung der Query im Arbeitsspeicher vor.

```
select
  jahr
  , dtmp
  , rege
  , regw
  , preis
  , prediction(
for preis using dtmp, rege, regw
) over ( ) preis_vorhersage
from
  wein;
```

Listing 1

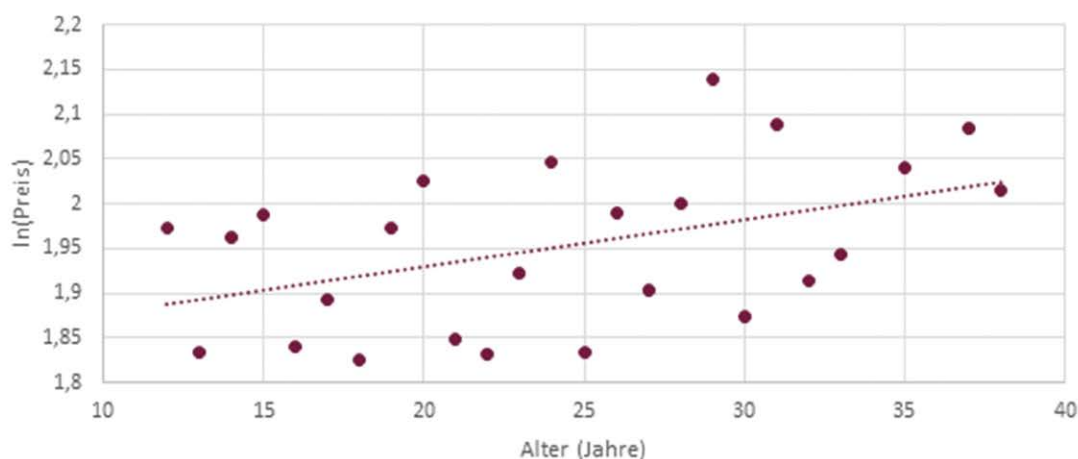


Abbildung 1: Preis (logarithmiert) gegen Alter in Jahren

ters kann der Datenbank-Entwickler zugreifen, die Schätzdaten wie auch die Ist-Daten der nachfolgenden Periode können hingegen nicht eingesehen werden.

Für eben diesen Zeitraum soll nun mithilfe der PQ eine Prognose erstellt und mit den tatsächlich erreichten sowie mit den durch die Experten vorhergesagten Werten verglichen werden. Hier zeigt sich dann, wie groß die Abweichungen sind und welche Erkenntnisse und Mehrwerte für das Unternehmen entstehen könnten.

Am Ende dieses kurzen und einfachen Prozesses stehen sich drei Werteräume gegenüber: Die tatsächlich erreichten Zahlen, die durch die Experten vormals prognostizierten Werte und eben jene Daten, die durch den Einsatz von PQ errechnet wurden. Ordnet man den Abweichungen von den Ist-Werten nun entsprechende Geldwerte zu oder quantifiziert diese anderweitig, so zeigt sich die Wirtschaftlichkeit der angewandten Methoden und somit die fallbezogene Wirksamkeit von PA. Dieses bewährte Vorgehensmodell erfüllt gleich mehrere Funktionen:

- Quantifizierung des Nutzens und Bestimmung des Return on Investment
- Nachweis der Wirksamkeit der PA-Methoden im Vergleich zu konventionellen Verfahren
- Einfache Validierung möglicher Anwendungsfälle
- Schaffen des notwendigen Vertrauens und Akzeptanz von PQ beziehungsweise Predictive Analytics im Allgemeinen

Sollte das Resultat nicht überzeugen, so muss dies nicht das Ende aller Bemühungen sein: Bessere, da auch stärker beeinflussbare Ergebnisse können durch den Einsatz der Advanced Analytics Option (AAO) erzielt werden. Die AAO erweitert die Datenbank um die zwei Komponenten R Enterprise und Data Mining. Oracle stellt eine umfassende Plattform zur Verfügung, die über alle nötigen Werkzeuge verfügt und Unternehmen in die Lage versetzt, maßgeschneiderte Analysen und Modelle im PA-Umfeld zu entwickeln. Denn mit R Enterprise verheiratet Oracle die mächtige und verbreitete Statistikumgebung R mit den bewährten Features der Oracle-Datenbank. Etwaige Analysen können so in R programmiert und durch die Datenbank integriert, parallelisiert sowie skalierbar berechnet werden. Darüber hinaus können mit Oracle Data Mining Work-

flows erstellt werden, die in der Datenbank integrierte Data-Mining- und PA-Algorithmen auf Daten anwenden. Diese Workflows können komfortabel mit einer grafischen Benutzeroberfläche innerhalb des SQL Developer entwickelt sein. Somit lassen sich mit AAO komplexe und umfassende Probleme im PA-Kontext effizient bewältigen.

Ein weiteres Unterscheidungsmerkmal der beiden Lösungsansätze betrifft das notwendige Fachpersonal. Während die Erprobung der Methoden und damit verbundene erste Schritte auch durch klassische Datenbank-Entwickler durchgeführt werden können, empfiehlt es sich beim Einsatz der AAO, auf Data Scientists oder vergleichbar ausgebildete Mitarbeiter zurückzugreifen. Diese sind nicht nur in der Lage, an den notwendigen Stellschrauben zur Verfeinerung der Ergebnisse zu drehen, sondern bringen auch das notwendige Verständnis für die mathematisch-statistischen Grundlagen und auch für die zu untersuchenden betriebswirtschaftlichen Fragestellungen mit. Spätestens mit dem Einsatz der AAO ist ein systematisches und methodisches Vorgehen unabdingbar. Dazu eignet sich besonders der Standard „Cross Industry Standard Process for Data Mining“ (CRISP-DM). Dieses bereits im Jahr 1999 veröffentlichte Rahmenwerk beschreibt den gesamten Ablauf in sechs Phasen:

- Business Understanding
- Data Understanding
- Data Preparation
- Modelling
- Evaluation
- Deployment

Advanced Analytics – die nächste Stufe benötigt Fachleute

Dieser Artikel betrachtet ausschließlich die Phasen „Data Preparation“ und „Modelling“, da diese ja durch PQ überwiegend übernommen werden. Im Mittelpunkt der Data Preparation steht die Definition eines finalen Algorithmus für die spätere Modell-Entwicklung. Spätestens zu diesem Zeitpunkt müssen Ausreißer, Messfehler oder fehlende Variablen sowie andere Unzulänglichkeiten ermittelt und bereinigt worden sein. Die mithilfe des eingesetzten Algorithmus gewonnenen Datensätze werden nach zuvor festgelegten Kriterien gefiltert, aggregiert und transformiert. Gegebenenfalls müssen die Daten noch durch weitere Informationen angereichert werden, um weitere Aspekte in die Analyse einfließen zu lassen.

Die Modelling-Phase beschreibt die Auswahl einer geeigneten PA-Methode zur Untersuchung der Aufgabenstellung. Auch hier sind verschiedene Teilprozesse zu durchlaufen, bis ein Modell entsteht, das zunächst mit Test- und Trainingsdaten und anschließend mit Daten aus den produktiven Systemen gespeist wird:

- Entwickeln eines Test-Designs mit Definition der Qualitätskennzahlen und Aufteilungen der Daten in eine geeignete Test- und Trainingsmenge
- Auswahl und Erprobung der Algorithmen und deren Dokumentation
- Erstellung des Modells und einer Dokumentation mit Begründung der Auswahl für diesen Typus

Schlussendlich erfolgt der Test der Modelle, der unter Umständen die Neuauswahl eines besser geeigneten Algorithmus erfordert oder auch zu Änderungen der Teilschritte aus den Vorphasen führen kann.

Fazit

Bereits dieser kurze Abriss zeigt, dass eine systematische und auf Nutzen-Maximierung ausgerichtete Beschäftigung mit dem Thema „Predictive Analytics“ keine spontane Angelegenheit sein kann, sondern einen Rahmen benötigt, der bei vielen Firmen nur bedingt gegeben ist. Insofern verstehen sich die PQ als ein guter Einstieg, um „sanft“ und ohne hohen Aufwand an das Thema herangeführt zu werden. Die nächsten Schritte erfolgen dann eher mit einem klassischen PA-Tool, mit den richtigen Fragenstellungen und von ausgebildeten Mitarbeitern.

Die weitere Entwicklung von PQ dürfte spannend werden: Welche zusätzlichen Funktionen zur Daten-Analyse und zur Verfeinerung möglicher Ergebnisse wird Oracle zukünftig in den Standardumfang der Datenbank einbauen und wie werden diese Möglichkeiten durch die Anwender angenommen? Der Trend zur Automatisierung von BI- und Analyse-Funktionen wird sich fortsetzen. Inwieweit Predictive Queries das Zeug zu „Business Analytics for the Masses“ haben werden, bleibt abzuwarten.

*Reinhard Mense
reinhard.mense@areto-consulting.de*

*Marco Nätlitz
marco.naetlitz@areto-consulting.de*

Einführung in Data-Mining mit analytischen Funktionen und R

Vladimir Poliakov, Nürnberg

Durch die riesige Auswahl an Paketen für die Daten-Analyse, Statistik und Visualisierungen ist die R-Software mittlerweile zum Standardwerkzeug für Daten-Auswertungen geworden. Als Open-Source-Projekt ist R eine starke Konkurrenz zu den kommerziellen Produkten. Dank der vielen Schnittstellen kann R die Daten aus verschiedenen Datenquellen lesen, etwa aus den CSV-Dateien oder aus einer Oracle-Datenbank. Dabei sind die analytischen Funktionen von Oracle ein mächtiges Tool für die Vorbereitung der Daten zur Analyse via R.

Laut Wikipedia [1] versteht man unter dem Begriff „Data-Mining“ (Daten-Bergbau) die systematische Anwendung statistischer Methoden auf eine Datenbasis mit dem Ziel, neue Trends zu erkennen. Die Methoden dieses Verfahrens kommen aus der Statistik, dem maschinellen Lernen sowie der klassischen Mustererkennung und wurden teilweise bereits vor Jahrzehnten entwickelt. Dieses „Data-Mining“-Analyseverfahren wird oft im Handel benutzt, aber im Prinzip kann man diese Methoden überall einsetzen, weil sie von der Herkunft der Daten unabhängig sind. Die Aufgaben des Data-Minings sind:

- Ausreißer-Erkennung (Abweichungs-Analyse)
- Cluster-Analyse
- Klassifikation
- Assoziations-Analyse
- Regressions-Analyse
- Ein typischer Data-Mining-Prozess sieht dabei folgendermaßen aus:
- Frage für die Data-Mining-Analyse formulieren
- Daten vorbereiten (finden, bereinigen und ins Data-Mining-Tool laden)
- Die Verteilung der Daten analysieren
- Die Variablen auswählen und dann das Datamodell mit den Trainingsdaten bilden
- Ergebnisse der Berechnung überprüfen, interpretieren und auf die Testdaten anwenden

Mögliche Einsatzbereiche für das Data-Mining-Verfahren sind:

- *Bankwesen*
Kreditwürdigkeit, Betrugserkennung

- *Handel*
Warenkorb-Analyse
- *Industrie*
Optimierung von Produktions- und Fertigungsprozessen (Optimierung der Stahlproduktion etc.)
- *Energieversorgung*
Effizienz der Anlagen (etwa bei Windanlagen)
- *IT*
Kapazitäts-Management, Ausfall der Server-Komponenten (wie Festplatten) abhängig von den Temperaturen im RZ
- *Medizintechnik, Technik allgemein*
prognostizierbare Ausfälle der Geräte
- *Sonstiges*
Vorhersage im Sportevent

Die letzte Behauptung klingt zunächst unrealistisch, aber das lässt sich dank vieler moderner Analyse-Tools relativ schnell überprüfen. Und das Thema ist eigentlich nicht neu, es gibt bereits sogar ein wissenschaftliches Buch über Statistik im Fußball [2]. Für diesen Artikel wurden die Daten zweier Eishockey-Ligen (NHL und DEL) aus den öffentlichen Datenquellen übernommen und analysiert. Dabei könnte die Data-Mining-Frage lauten: „Ist die Anzahl der Tore im Spiel von der Stärke der Mannschaften abhängig?“ Da die Mannschaftsstärke durch die Anzahl der Tore beziehungsweise der Gegentore bestimmt werden kann, wird die Data-Mining-Frage jedoch letztendlich so lauten: „Ist die Anzahl der Tore im Spiel von der vor dem Spiel herrschenden Tordifferenz beziehungsweise von der Gegentor-Differenz beider Gegner abhän-

gig?“ Ist die Frage festgelegt und somit das Ziel der Analyse gesetzt, kann man mit der Datenvorbereitung (siehe oben) beginnen.

Das ist einer der wichtigsten Teile des Data-Mining-Analyse-Prozesses, weil diese

Vorbereitung bis zu 80 Prozent der Zeit des gesamten Data-Mining-Analyse-Prozesses einnehmen kann. Häufig wird diese Phase als „ETL“ (Extract, Transform, Load) bezeichnet. Während dieser Phase werden die Daten aus den verschiedenen Datenquellen (operative Datenbanken, soziale Netzwerke, Log-Dateien etc.) geholt, bereinigt (beispielsweise ein Geburtsdatum liegt in der Zukunft oder die Geschlechtsdaten fehlen, obwohl die Anrede bekannt ist), zum einheitlichen Format transformiert (etwa das Datum wird in „DD.MM.YYYY“-Form abgespeichert oder die personenbezogenen Daten werden anonymisiert) und für die weitere Bearbeitung im Zielsystem (Data Warehouse, Hadoop Filesystem als Data Lake etc.) abgelegt. Ist die ETL-Phase abgeschlossen, lassen sich die Inputdaten in Analytical Records [3] zusammenfassen (etwa in Form einer View), falls das während der ETL-Phase nicht gemacht wurde.

Der Analytical Record ist ein Input, der von jedem Data-Mining-Tool (egal ob R oder ein anderes Tool) erwartet wird. Dabei stellen jede Zeile einen Fall, eine Spalte den vorherzusagenden Wert und die anderen Spalten die Eigenschaften des Falles, anders gesagt die Prädiktoren, darx.

Es handelt sich dabei um einige Vorbereitungsprozesse. Oft sind das die aufwändigen Manipulationen mit den operativen oder historischen Daten (Gruppieren, Summieren

```
LAG(GOALS,1,0) OVER (PARTITION BY TEAM_ID ORDER BY MATCHDATE, TEAM_ID) PREV_GOALS
SUM(PREV_GOALS) OVER (PARTITION BY TEAM_ID ORDER BY MATCHDATE, TEAM_ID) SUM_PREV_GOALS
```

MATCHDATE	TEAM_ID	TEAM	GOALS	AGAINST_GOALS	POINTS	PREV_GOALS	PREV_AGAINST_GOALS	PREV_POINTS	SUM_PREV_GOALS	SUM_PREV_AGAINST_GOALS
12.09.2014 00:00:00	1	Adler Mannheim	5	2	3	0	0	0	0	0
12.09.2014 00:00:00	2	Augsburger Panther	4	1	3	0	0	0	0	0
12.09.2014 00:00:00	3	Düsseldorfer EG	0	7	0	0	0	0	0	0
12.09.2014 00:00:00	4	ERC München	6	3	3	0	0	0	0	0
12.09.2014 00:00:00	5	ERC Ingolstadt	2	5	0	0	0	0	0	0
12.09.2014 00:00:00	7	Eisbären Berlin	1	4	0	0	0	0	0	0
12.09.2014 00:00:00	5	Grizzly Adams Wolfsburg	7	0	3	0	0	0	0	0
12.09.2014 00:00:00	8	Hamburg Freezers	3	6	0	0	0	0	0	0
12.09.2014 00:00:00	14	Ice Tigers Nürnberg	3	1	3	0	0	0	0	0
12.09.2014 00:00:00	9	Iserlohn Roosters	4	2	2	0	0	0	0	0
12.09.2014 00:00:00	10	Krefeld Pinguine	2	4	0	0	0	0	0	0
12.09.2014 00:00:00	11	Kölner Haie	2	3	1	0	0	0	0	0
12.09.2014 00:00:00	12	SERC Wild Wings	1	3	0	0	0	0	0	0
12.09.2014 00:00:00	13	Straubing Tigers	3	2	2	0	0	0	0	0
14.09.2014 00:00:00	1	Adler Mannheim	2	5	0	5	2	3	5	2
14.09.2014 00:00:00	2	Augsburger Panther	4	2	3	4	1	3	4	1
14.09.2014 00:00:00	3	Düsseldorfer EG	4	1	3	0	7	0	0	7
14.09.2014 00:00:00	4	ERC München	2	0	3	6	3	3	6	3
14.09.2014 00:00:00	6	ERC Ingolstadt	7	4	0	2	5	0	7	4
14.09.2014 00:00:00	7	Eisbären Berlin	5	1	3	1	4	0	7	1
14.09.2014 00:00:00	5	Grizzly Adams Wolfsburg	4	1	3	7	0	0	7	0
14.09.2014 00:00:00	8	Hamburg Freezers	1	4	0	3	6	0	3	6
14.09.2014 00:00:00	14	Ice Tigers Nürnberg	3	2	2	3	1	3	3	1
14.09.2014 00:00:00	9	Iserlohn Roosters	2	3	1	4	2	3	4	2
14.09.2014 00:00:00	10	Krefeld Pinguine	5	2	3	2	4	0	2	4
14.09.2014 00:00:00	11	Kölner Haie	1	4	0	2	3	1	2	3
14.09.2014 00:00:00	12	SERC Wild Wings	0	7	0	1	3	0	1	3
14.09.2014 00:00:00	13	Straubing Tigers	1	5	0	3	2	2	3	2
19.09.2014 00:00:00	1	Adler Mannheim	4	3	3	2	5	0	7	3

Abbildung 1: Die analytischen Funktionen im Einsatz

```
FUNCTION_NAME([Argumente]) OVER ([analytische Klausel])
```

Listing 1

No.	Variable	Data Type	Input	Target	Risk	Ident	Ignore	Weight	Comment
1	NR	Ident	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 364
2	HOMETEAM	Categoric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 14
3	VISITORTEAM	Categoric	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	Unique: 14
4	SUM_TOTALSCORE_REGULAR_TIME	Numeric	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Unique: 13

Abbildung 2: Definieren der Variablen in Rattle

etc.), die mithilfe der analytischen Funktionen (in allen Oracle Database Editions verfügbar) performant in der Analytical-Record-View zusammengefasst sind. Listing 1 zeigt schematisch dargestellt die generelle Syntax des Aufrufs einer analytischen Funktion [4].

Die analytischen Funktionen fassen wie die Gruppen-Funktionen die Ergebnismenge zusammen, aber es wird für jede Zeile der Ergebnismenge einzeln ein Wert ausgegeben. Das Ausgabe-Format einer analytischen Funktion entspricht in Prinzip dem Analytical Record. Je nachdem können viele Funktionen als Gruppenfunktionen oder als analytischen Funktionen ausgeführt werden (siehe Abbildung 1).

Ist die Analytical-Record-View vorbereitet, können die Daten in R geladen werden. R ist eine freie Programmiersprache, die für die Datenanalyse 1992 an der Universität Auckland in Neuseeland entwickelt wurde. Es ist eine Skriptsprache und wird nicht kompiliert, sondern in der R-Console interpretiert. Das ist jedoch kein K.o.-Kriterium. Es gibt bereits genug grafische Benutzeroberflächen, wie R-Studio, die einen Einstieg in die R-Welt ermöglichen. R-Studio kann die Daten zur Analyse importieren und hat einen Editor, die R-Console, und vieles mehr.

Die R-Tools (R-Studio, R Commander, R Data Miner etc.) präferieren per Default „comma separated values“-Files (CSV-Datei),

können aber über Schnittstellen die Daten direkt aus der Oracle-Datenbank lesen. Das ist keine große Kunst, denn dafür sind lediglich der Oracle-Instant-Client sowie zusätzliche R-Bibliotheken je nach Zugriffsart (ROracle, RODBC, RJDBC etc.) erforderlich.

Der letzte Schritt vor dem Modell-Aufbau ist die Analyse der Häufigkeitsverteilung, weil viele Modelle die Normalverteilung der Daten voraussetzen. Es gibt verschiedene Techniken und Verfahren in der Statistik zur Beschreibung der Verteilung der Daten [5], die im Rahmen dieses Artikels nicht näher beleuchtet werden. Es wird für das generalisierte lineare Modell (GLM) angenommen, dass die Daten der Poisson-Verteilung [2] unterliegen. Diese Hypothese ist der Grundstein für den Modellaufbau im R Data Miner.

R Data Miner oder kurz „Rattle“ ist ein R-Paket, das die grafische Benutzeroberfläche hat und speziell für die Data-Mining-Analyse geschrieben ist. Es kann die Daten nicht nur aus den CSV-Dateien lesen, sondern auch aus Excel-Tabellen, ODBC-Quellen, R-Datasets und anderen Formaten. Wichtig ist, dass jede Aktion in Rattle explizit mit der Schaltfläche „Execute“ bestätigt werden soll. Sind die Daten ins Tool geladen, kann man die Variablen identifizieren und mit dem Modellaufbau beginnen (siehe Abbildung 2). Wie bereits erwähnt wurde, wird für das generalisierte lineare Modell (GLM) angenommen, dass die Daten der Poisson-Verteilung [2] unterliegen. Entsprechend dieser Aussage wird in Rattle das Modell aufgebaut (siehe Abbildung 3).

Wie man sieht, sind die Input-Variablen für das Modell nicht signifikant, was bedeu-

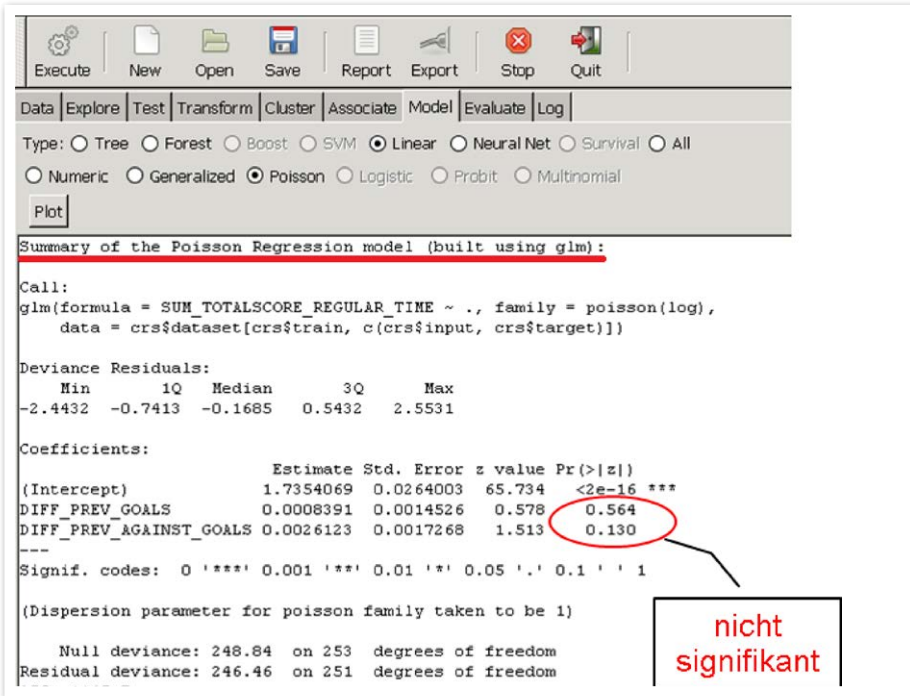


Abbildung 3: Das GLM für die Anzahl der Tore pro Spiel in der DEL Saison 2014/2015

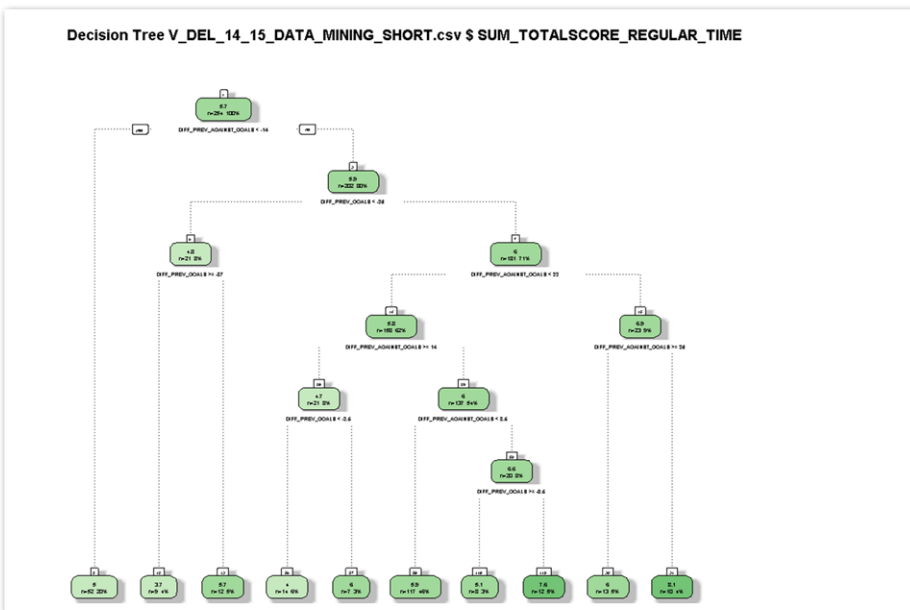


Abbildung 4: Das Entscheidungsbaum-Modell für die Anzahl der Tore pro Spiel in der DEL-Saison 2014/15

tet, dass es in der Deutschen Eishockey-Liga in der Saison 2014/15 keine Abhängigkeiten zwischen der Anzahl der Tore im Spiel und der Tordifferenz beziehungsweise der Gegentor-Differenz der Mannschaften vor dem Spiel gibt. Die Situation in der National Hockey League ist in der gleichen Saison anders [6]. Wahrscheinlich liegt es daran, dass in der DEL meistens nur freitags und sonntags gespielt wird, dagegen finden in der NHL die Spiele fast täglich statt und so langsam setzt sich das Gesetz der großen Zahlen durch.

Wenn die Input-Variablen für das Modell nicht signifikant sind, bedeutet das letztendlich, dass es einfach nicht die lineare Abhängigkeit gibt, die die Menschen so mögen. In diesem Fall können die Daten auf die anderen Modelle angewendet werden. Zum Beispiel auf das Entscheidungsbaum-Modell (Decision Tree Model), das Rattle defaultmäßig bietet. Dieses Modell ist leichter zu interpretieren und außerdem ist es robuster gegenüber Ausreißern. Jeder Zweig des Entscheidungsbaum-Modells präsentiert die

zugrunde liegende Bedingung und das prozentuale Auftreten des Ereignisses. Die Entscheidungsregeln können entweder in Textformat im Rattle-Output oder graphisch im R Studio (siehe Abbildung 4) dargestellt werden.

Im Endeffekt sollen die bereits geleisteten Schritte auch leicht reproduziert werden. Das ist ohne großen Aufwand möglich, weil Rattle alle Aktionen im Log-Output protokolliert. Diese Log-Ausgabe ist nichts anderes als eine Reihe von R-Befehlen, die in der Console ausgeführt werden. Wenn der Data Scientist oder der Daten-Bergbauer mit den Ergebnissen seiner Arbeit zufrieden ist, kann alles als R-Skript abgespeichert und zum späteren Zeitpunkt mit den Test-Daten ausgeführt werden.

Fazit

Das Data-Mining-Verfahren hilft beim Untersuchen verschiedener Beziehungen im Datenbestand, was mit der Standard-Auswertung (Statistik und/oder OLAP) praktisch nicht möglich ist, da diese Verfahren nur bestimmte und vorher festgelegte Fragen beantworten. Die analytischen Funktionen helfen dabei, die Daten für die Analyse vorzubereiten. Auch für diejenigen, die mit dem Data-Mining nichts zu tun haben, lohnt es sich, diese Funktionen kennenzulernen. Sie können gute Dienste im täglichen SQL-Leben leisten.

Dieser Artikel ist eine kleine Einführung in den Daten-Bergbau. Wie bereits am Anfang erwähnt, kann das Data-Mining-Verfahren in den verschiedensten Bereichen eingesetzt werden. Man muss aber tiefer in die Statistik, in die Verteilung der Daten einsteigen und sich mit der Signifikanz der Modelle auseinandersetzen. Die Ergebnisse der Analyse können dann verblüffend sein.

Referenzen

- [1] Data-Mining in Wikipedia: <https://de.wikipedia.org/wiki/Data-Mining>
- [2] Heuer, Andreas: Der perfekte Tipp – Statistik des Fußballspiels (Erlebnis Wissenschaft) 1. Auflage September 2012, Wiley-VCH, Weinheim Verlag ISBN 978-3-527-33103-1
- [3] Zeitschrift iX Developer Big Data 2015 – Analytics Design Patterns
- [4] Analytische Funktionen: <http://www.munisoft.de/tipps/sql/sql-allgemein/analytische-funktionen.htm>
- [5] Dormann, Carsten F.: Parametrische Statistik – Verteilungen, maximum likelihood und GLM in R. Springer Spektrum, 2013. ISBN 978-3-642-32785-6
- [6] Poliakov, Vladimir: Auf der Suche nach dem perfekten Tipp: <http://heise.de/-2776160>

Vladimir Poliakov
v.poliakov@gmx.net

Was wäre die **Meinung**
der **SAP-Community**
ohne **Veröffentlichung**
im **E-3 Magazin?**



Informations- und Bildungsarbeit von und für die SAP-Community.
Das monatliche E-3 Magazin erscheint auf Papier, Apple iOS, Android und www.e-3.de



www.e-3.de

SAP® ist eine eingetragene Marke der SAP AG in Deutschland und in den anderen Ländern weltweit. Foto: Barbaros, Shutterstock.com

Facebook.com/e3magazin

Sind Sie Up-to-date?

ORACLE BI PUBLISHER REPORTING

standardisiert | automatisiert | wirtschaftlich

**“Über das Ziel hinausschießen
ist ebenso schlimm
wie nicht ans Ziel kommen.”**

Konfuzius

Informationen sind der Schlüssel
für ein wettbewerbsfähiges
und erfolgreiches Unternehmen.

BI Reports liefern jederzeit die
entscheidungsrelevanten Informationen.

Einfache Zusammenführung
unterschiedlicher Datenquellen
in einem Bericht.

Fehlerreduktion durch automatisierte
Reporterstellung in gewohnten
Formaten (xls, doc, pdf), standardisiert
in Ihren eigenen Templates.



ADF REPORTING

interaktiv | integrierbar | mobil

**“Es wächst zusammen,
was zusammen gehört!”**

Willy Brandt

Übersichtliche und grafische
Aufbereitung Ihrer Daten.
Analysieren Sie per Drilldown-Funktion
effektiv und zuverlässig.

Integrieren Sie Ihre Berichte nahtlos
als Dashboard z.B. in Ihrem Intranet

Dateien in interaktiven Reports
per Drag & Drop bearbeiten.
Intuitiv und überall verfügbar
am PC oder Mobil-Device.

ORACLE Gold
Partner

Specialized
Primavera P6 Enterprise
Project Portfolio Management

proadvise GmbH
Brühlweg 4
73663 Berglen

Tel. +49 700 776238473
Fax +49 7195 970885
info@proadvise.com

proadvise
living your projects
www.proadvise.com