



UNIVERSITÄT HEIDELBERG UND HOCHSCHULE HEILBRONN

Fakultät für Informatik

## Diplomarbeit

Einsatz von OLAP in der Messtechnik der T-Mobile

---

vorgelegt von: Xin Li  
eingereicht am: 01. November. 2010  
Studiengang: Medizinische Informatik  
Studienrichtung: Informationssysteme des  
Gesundheitswesens

Referent: Prof. Dr. Gerhard Peter  
Korreferent: Prof. Dr. Heinrich Kray

## **Danksagung**

An dieser Stelle möchte ich allen danken, die zum Abschluss dieser Arbeit beigetragen haben. Mein ganz besonderer Dank geht an Prof. Dr. Gerhard Peter für seine wissenschaftliche Betreuung, Verbesserungsvorschläge, Anregungen und Kritik. Ein grosses Dankeschön an Prof. Dr. Heinrich Kray für die Übernahme des Korreferents und Anregungen.

Der praktische Teil dieser Arbeit entstand in Kooperation mit der Firma T-Mobile GmbH in Stuttgart. An dieser Stelle möchte ich mich herzlich bei Herrn Roland Hackenberg und den Kollegen von der gesamten Abteilung für die gute Zusammenarbeit bedanken.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Aufgabenstellung . . . . .	1
1.2	Aufbau der Arbeit . . . . .	2
<b>2</b>	<b>Business Intelligence</b>	<b>4</b>
2.1	Einführung von BI . . . . .	4
2.2	Anwendungsbereich für BI . . . . .	5
2.2.1	Datenbank und Data Mining . . . . .	5
2.2.2	Unterschiede zwischen OLAP und Data Mining . . . . .	6
2.2.3	OLAP im Data Warehouse . . . . .	6
2.3	Transformationsprozess - ETL (Extraktion, Transformation, Laden) . . . . .	7
2.3.1	Filterung . . . . .	8
2.3.2	Harmonisierung . . . . .	8
2.3.3	Aggregation . . . . .	9
2.3.4	Anreicherung . . . . .	10
<b>3</b>	<b>Grundlage von OLAP</b>	<b>11</b>
3.1	Definition von OLAP . . . . .	11
3.2	OLAP-Spezifikation . . . . .	12
3.2.1	Die Grundregeln von Codd . . . . .	12
3.2.2	FASMI . . . . .	13
3.3	Das multidimensionale Datenmodell . . . . .	14
3.3.1	Grundidee . . . . .	14
3.3.2	Fakten . . . . .	15
3.3.3	Kennzahl . . . . .	15
3.3.4	Dimension . . . . .	16
3.3.5	Hierarchien . . . . .	17
3.3.6	Würfel . . . . .	17
3.4	OLTP und OLAP im Vergleich . . . . .	19
3.5	Das relationale Schema . . . . .	20
3.5.1	Star-Schema . . . . .	20
3.5.2	Snowflake-Schema . . . . .	22

3.6	Funktionen des OLAPs . . . . .	23
3.7	ROLAP, MOLAP, DOLAP und HOLAP . . . . .	25
<b>4</b>	<b>Vergleich von OLAP - Tools</b>	<b>27</b>
4.1	BI Konzept . . . . .	27
4.2	Jaspersoft . . . . .	28
4.2.1	JasperServer . . . . .	28
4.2.2	Installation . . . . .	29
4.2.3	Jasper Analysis . . . . .	30
4.2.4	Jasper ETL . . . . .	30
4.2.5	iReport . . . . .	31
4.3	Pentaho . . . . .	32
4.3.1	Allgemeines . . . . .	32
4.3.2	Installation . . . . .	34
4.3.3	Pentaho Analysis Services (Mondrian) . . . . .	34
4.3.4	PentahoETL (Kettle) . . . . .	38
4.3.5	Pentaho Report Designer . . . . .	40
4.3.6	Pentaho Enterprise Edition . . . . .	41
4.4	Palo Suite . . . . .	42
4.4.1	Palo OLAP Server . . . . .	42
4.4.2	Palo Web . . . . .	42
4.4.3	Palo ETL . . . . .	44
4.4.4	Palo for Excel . . . . .	45
4.4.5	Palo Supervision Server . . . . .	46
<b>5</b>	<b>Implementierung des multidimensionalen Datenmodells bei T-Mobile</b>	<b>47</b>
5.1	Aufbau der OLAP-Datenbank . . . . .	47
5.1.1	Erstellung der Datenbank . . . . .	47
5.1.2	Testdaten . . . . .	50
5.2	Framework . . . . .	50
5.2.1	Verbindung mit der Datenbank . . . . .	50
5.2.2	Schemabrowser JasperAnalysis Schema Workbench . . . . .	51
5.2.3	Benutzeroberfläche JasperAnalysis Schema Workbench . . . . .	52
5.2.4	Integration mit JasperAnalysis User's Guide . . . . .	54
5.2.5	Visualisierung mit Analysis View . . . . .	56
5.2.6	Implementierung . . . . .	58
5.2.6.1	MDX und OLAP . . . . .	58
5.2.6.2	Übersetzung der Benutzeraktionen in MDX . . . . .	59
<b>6</b>	<b>Bewertung der BI Standardlösungen im Zusammenhang mit T-Mobile</b>	<b>60</b>

---

6.1	Bewertung der BI-Funktionalität . . . . .	60
6.2	Ergebnisse der Bewertung . . . . .	62
<b>7</b>	<b>Zusammenfassung</b>	<b>65</b>
<b>8</b>	<b>Rückblick</b>	<b>66</b>
	<b>Literaturverzeichnis</b>	<b>68</b>
	<b>Abbildungsverzeichnis</b>	<b>70</b>
	<b>Tabellenverzeichnis</b>	<b>72</b>
	<b>Abkürzungsverzeichnis</b>	<b>73</b>
	<b>Glossar</b>	<b>74</b>

# 1 Einleitung

Heutzutage ist es für Unternehmen immer wichtiger, dass man in kürzester Zeit die richtigen Informationen erhalten kann, um Entscheidungsprozesse optimal bilden zu können. Die Informationstechnologie lässt sich zurzeit in der Welt weit und breit nutzen. Sie kann nicht nur den Daten sammeln und austauschen, sondern auch kann die Informationen systematisch und rapid verarbeiten, verwalten, analysieren und interpretieren. Die wachsende Datenmenge und Komplexität erfordert immer stärker Analyseinstrumente gewinnen an Bedeutung.

In großen Unternehmen Umgebungen stoßen einfache Abfrage-Werkzeuge für den Datenzugriff schnell an die Grenzen ihre Leistungsfähigkeit und der Performance-Gewinn wird auch zum zentralen Thema. Aus diesen Gründen wird das Unternehmen T-Mobile den Analysewerkzeuge, den sogenannten OLAP (Online Analytical Processing)-Tools, einsetzen, um die Daten in sogenannten mehrdimensionalen Tabellenstrukturen flexibel und schnell analysieren bzw. auswerten zu können. Ein weiterer Punkt warum das Analysieren der Call Daten interessant sein könnte ist das Überwachen von Fehlern. Durch diese Technologie kann die Lösungsqualität hinsichtlich der zu verarbeitenden Komplexität erhöht werden. Als Grundlage für die OLAP Anfragen wurde zunächst ein Data Warehouse aufgebaut, in dem die Daten aus dem Quellsystem integriert und strukturiert bereitgestellt werden.

Des weiteren ist zu erstellen, in wie weit der Einsatz von Reporting-Tools und zusätzlicher OLAP-Werkzeuge, um die gewünschten Berichtsfunktionalitäten zu ermöglichen. Diese Funktionalitäten sollen Schließlich in geeigneter Weise umgesetzt und bereitgestellt werden.

Dabei sind die Anforderungen der Mitarbeiter und damit der Benutzer des Systems und die Rahmenbedingungen der Firma T-Mobile, z.B. bezüglich Auswahl der Software, zu berücksichtigen.

## 1.1 Aufgabenstellung

Ziel dieser Arbeit ist es durch OLAP Anfragen Informationen über Call-Daten von Datenänderungen zwischen Informationssystemen zu gewinnen. Anhand dieser Auswertungen wird versucht, ein konkretes Modell zu entwickeln. Der Schwerpunkt dieser Arbeit liegt in der Untersuchung der Implementierungsmöglichkeiten und der Modellentwicklung für mögliche

Anwendung.

Auf Basis einer Anforderung-Analyse wird ein beispielhaftes Modell einer OLAP-Datenbank für Kunden-Call-Informationen erstellt. Das Modell ist mit den Open-Source-Produkten JasperServer, Pentaho und Palo umzusetzen und damit zu vergleichen. Dies soll aufzeigen, inwiefern eine Umsetzung der geforderten Aufgabenstellungen erfüllt werden können.

In der Implementierung des Datenmodells wird eine OLAP-Datenbank entworfen und mit Beispieldatensätzen gefüllt. Die geforderten Analysen und Auswertungen werden mit diesen Datensätzen erstellt. Diese Informationen können weitergearbeitet werden. Man kann für die verschiedenen Hierarchieebenen unterschiedliche Ansichten und Zugriffsrechte einrichten.

Als Grundlage für die OLAP-Anfragen wurde zunächst ein Data Warehouse aufgebaut, in dem die Daten aus dem Log integriert und strukturiert bereitgestellt werden. Um jedoch die Daten aus dem Log ins Data Warehouse zu laden, sind ETL-Prozesse erforderlich. Zum Schluss werden die implementierten OLAP-Anfragen auf dem Data Warehouse ausgeführt. Dadurch werden Informationen gewonnen, die genutzt werden können, um zum Beispiel Optimierungspotenziale zu ermitteln.

Unternehmen benötigen Zugang zu Echtzeitdaten, um ihre Geschäftsziele zu erreichen sowie um hohe Service-Level für ihre Kunden aufrechtzuerhalten. Heute verlangen Anwender, dass Daten nach einer Änderung sofort zur Verfügung stehen. Für sie besteht die Herausforderung darin, Daten über diese Architekturen in Echtzeit zu replizieren und zu synchronisieren.

## 1.2 Aufbau der Arbeit

Um den Kontext, in dem OLAP eingesetzt werden soll, zu verdeutlichen, werden im **Kapitel 2** zunächst die Klärung und die Anwendung von Business Intelligence (BI) und Data Warehouse vorgenommen.

Darauf aufbauend werden grundlegende Begriffe aus dem Bereich OLAP im **Kapitel 3** beschrieben. Es wird eine Übersicht über die verschiedenen Ansätze, Arten und Modellierungstechnik des OLAP gegeben.

Im **Kapitel 4** werden einige vorhandene Visualisierungszeuge (Jasper, Pentaho und Palo) mit Interfaces getestet und verglichen.

Im **Kapitel 5** steht die Umsetzung eines konkreten OLAP Modells im Vordergrund. (Der Schwerpunkt liegt nicht in der eindeutigen Definition und Festlegung notwendiger Unternehmenskennzahlen, sondern vielmehr in der Untersuchung der Visualisierungsmöglichkeiten und der Modellentwicklung für mögliche Firmenanwendung.)

Für den Aufbau der OLAP-Datenbank und Implementierung von Anfragen wurde beispielhaft mit Werkzeugen von Jasper umgesetzt. (Zunächst wird eine OLAP-Datenbank entworfen und mit Datensätzen gefüllt.) Mit dieser Datenbasis werden dann die geforderten Analysen und Auswertungen erstellt. Diese Informationen werden zur Weiterbearbeitung bereitgestellt. In der technischen Umsetzung wird, soweit möglich, versucht die Anforderungen durch Konfiguration und Parametrisierung von Jasper zu erfüllen. Dies soll aufzeigen, inwieweit eine Umsetzung der geforderten Aufgabenstellungen ohne tiefere Programmierkenntnisse erfüllt werden können.

Zum Schluss werden die verwendeten Produkte im Hinblick auf die Anforderungen im **Kapitel 6** bewertet. Es werden die Vorzüge dargestellt und die Schwierigkeiten, die sich bei der Umsetzung ergaben, herausgearbeitet.

Die Zusammenfassung bildet den Abschluss der Arbeit im **Kapitel 7** und bewertet, ob die verwendeten Produkte für die Umsetzung der geforderten BI-Lösung bei T-Mobile geeignet sind.



## 2 Business Intelligence

Dieses Kapitel erläutert die Einführung für OLAP und ihre Anwendung in **Business Intelligence** (BI). Außerdem wird die Grundlagen des Data Warehouse für BI Tools erläutert. Anschließend werden die einzelnen Komponenten des Data Warehouses erläutert.

### 2.1 Einführung von BI

Business Intelligence ist die entscheidungsorientierte Sammlung, Aufbereitung und Darstellung geschäftsrelevanter Informationen [Sch06].

In dieser Arbeit wird diese Definition der Geschäftsdatenanalyse verwendet. Die wesentlichen Aspekte, die daraus hervorgehen, sind [Ker07]

- Entscheidungsorientierung  
BI dient dazu, Entscheidungsgrundlagen zu verbessern.
- Datensammlung  
BI beinhaltet die Steuerung der Zugriffe auf Datenbanken.
- Datenaufbereitung  
Dies ist das eigentliche Kernstück. Es werden aus Rohdaten Informationen mit Hilfe von mathematischen Verfahren und Analysen gewonnen.
- Informationsdarstellung  
BI stellt nicht nur Datensammlung und Datenauswertungen zur Verfügung, sondern bildet die richtigen Informationen in der richtigen Struktur für die unterschiedlichen Benutzergruppen ab.
- Geschäftsrelevante Informationen  
Da das Speichern von Daten sehr aufwändig und teuer ist, werden in BI Lösungen so viele geschäftsrelevante Daten wie nötig, aber so wenig wie möglich verwendet.

BI-Technologie stellt jedem Anwender die richtigen Menge und Darstellungsart von Information zur Verfügung. Zusammenfassend hat BI zum Ziel, Entscheidungsgrundlagen zu verbes-

sern, die Transparenz der Unternehmenshandlungen zu erhöhen und Zusammenhänge von isolierten Informationen aufzuzeigen [Sch06].

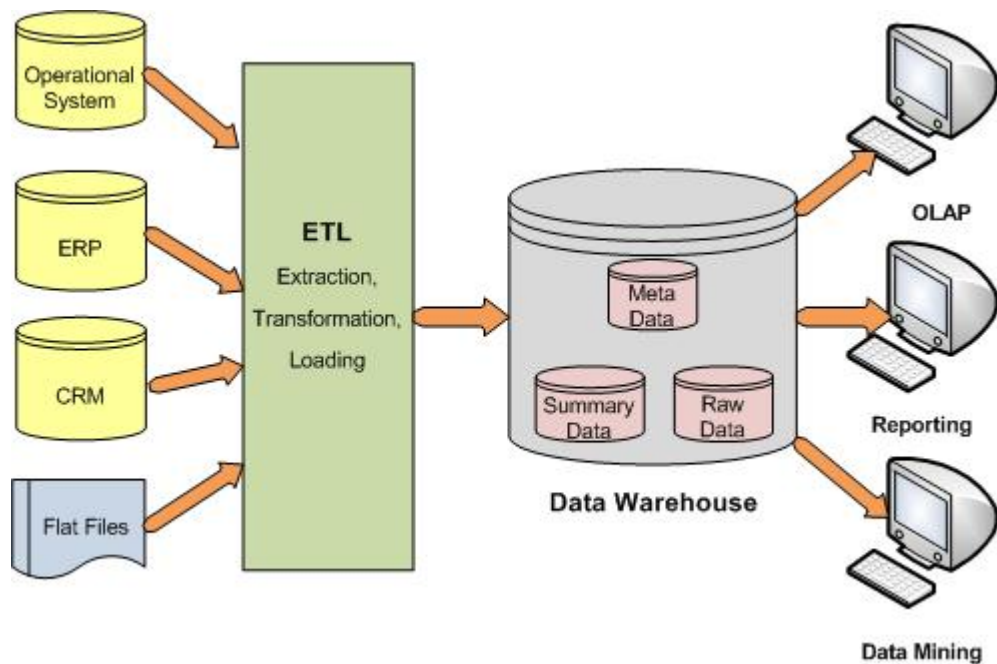


Abbildung 2.1: Die Business Intelligence-Plattform

Es ist notwendig, dass OLAP im Unternehmen eingesetzt werden, um die Daten schnell und effizient verarbeiten, analysieren und auswerten zu können. Unter BI kann man OLAP gut verstehen, der eine Entscheidungsfindung unterstützt. In der Arbeit tauchen weitere wichtige Begriffe aus dem Bereich BI auf, die im Folgenden kurz erklärt werden sollen. Eine detaillierte Erläuterung von OLAP erfolgt in den Kapitel 3.

## 2.2 Anwendungsbereich für BI

Business Intelligence im analyseorientierten Sinne umfasst sämtliche Anwendungen, bei denen der Entscheidungsvorbereiter direkt mit dem System arbeitet, d. h. einen unmittelbaren Zugriff auf eine Benutzungsoberfläche mit interaktiven Funktionen besitzt. Hierzu gehören neben OLAP auch Data Mining sowie Data Warehouse [Car04].

### 2.2.1 Datenbank und Data Mining

Eine wesentliche Voraussetzung zum erfolgreichen Einsatz von BI ist eine gut strukturierte Datenbank, in der alle benötigten Daten und Informationen zur Ermittlung der Kennzah-

len abgelegt sind [Ker07]. Datenbank erfordert ausreichende Datenbestände, die sich aus mehreren Feldern zusammensetzen können. Und die Datenqualität ist auch wichtig, so dass richtige Verfahren sinnvoll durchgeführt werden können. Datenbank besteht aus mehreren Tabellen, die aus ein oder mehreren Datensätze bilden. Eine neue Tabelle kann aus vorhandenen Tabellen erzeugen. Tabellen können mit anderen Tabellen in Relation gebracht werden. Bei der Datenbank kann es durch SQL-Sprache (Structured Query Language) abgefragt und verändert.

Der Begriff „Data Mining“ bezeichnet Techniken und Methoden, die aus bestehenden Datenbeständen neue, bis dahin unbekannte Informationen erzeugen. Mittlerweile sind eine Menge von Einsatzgebieten bekannt, wo Data Mining Verfahren genutzt werden können. Deshalb haben sich verschiedenen Klassifikationen etabliert, wie z.B. Cluster-Analysen, Entscheidungsbäume, Faktorenanalysen oder Neuronale Netze. Hinter all diesen Verfahren stecken mathematische Algorithmen die komplexe Zusammenhänge aufdecken können [Sch06].

### 2.2.2 Unterschiede zwischen OLAP und Data Mining

Es gibt eine Reihe von Unterschieden zwischen OLAP und Data Mining, die im Folgenden erläutert werden:

- OLAP-Tools bieten neben den Standardanfragen auch Funktionen wie zum Beispiel Slice und Dice, die nötig sind, um komplexere Anfragen zu stellen und Vergleiche zu erstellen. Data-Mining-Tools gehen nach Meinung von Griffin weiter: Sie bieten Informationen an, von denen der Benutzer, noch nicht wusste, dass er danach suchte.
- OLAP liefert statistische Belege für die Hypothesen eines Anwenders; Data Mining liefert selbständig neue Zusammenhänge aus den Daten heraus.
- OLAP verwendet multidimensionale Strukturen um Analysen durchführen zu können. Bei Data Mining sind multidimensionale Strukturen nicht zwingend erforderlich.

### 2.2.3 OLAP im Data Warehouse

Data Warehouses (DWH) sind von den operativen Datenbeständen getrennte, logisch zentralisierte dispositive Datenhaltungssysteme [Car04]. Ein Data Warehouse ist eine zentrale Stelle, an der alle Daten gesammelt werden können.

Die Daten, auf die über OLAP-Funktionalität zugegriffen wird, stammen üblicherweise aus einem sogenannten Data Warehouse. In einem Data Warehouse geht es ausschließlich um die Analyse der gesamten Daten.

---

Ein Data Warehouse wird von Inmon durch folgende Merkmale charakterisiert [Gla97]:

- Orientierung an unternehmensbestimmenden Sachverhalten. Dies kann vereinfacht auch als Entscheidungsrelevanz bezeichnet werden. Man konzentriert sich auf die Daten. Operative Prozesse, d.h. die Entstehung der Daten, sind nicht von Bedeutung, sondern es zählt ausschließlich die Verwendung.
- Zeitraumbezug in Abgrenzung zum Zeitpunktbezug operativer Systeme. Das Herausfiltern von Trends wird besonders betont. Der Datenhorizont beträgt daher häufig bis zu zehn Jahre. In Abhängigkeit des Alters werden dabei unterschiedliche Verdichtungsgrade gespeichert.
- Struktur- und Formatvereinheitlichung durch konsistente Bezeichner und einheitliche Strukturen.
- Nicht-Volatilität, was bedeutet, dass die Daten, abgesehen von wenigen Korrekturen, nicht mehr geändert werden. Das Data Warehouse gestattet somit nur lesenden Zugriff.

Die Informationen in einem Data Warehouse werden verwendet, um nützliche Wissen zu exportieren, das wiederum genutzt werden kann, um Entscheidungen treffen zu. Um Wissen aus einem Data Warehouse zu exportieren sind Business Intelligence Tools wie zum Beispiel OLAP erforderlich.

Data Warehouse und OLAP können auch innerhalb des Messtechnik in T-Mobile eingesetzt werden. Der Prozess der Messtechnik von Call-Datenänderungen zwischen Informationssystemen kann mit Hilfe eines DWHs und OLAP analysiert werden, um beispielsweise anschließend Optimierung durchzuführen.

## 2.3 Transformationsprozess - ETL (Extraktion, Transformation, Laden)

Der Begriff ETL-Prozess wird aus den englischen Bezeichnungen der Teilschritte „Extraktion“, „Transformation“ und „Laden“ definiert.

- **Extraktion** ist verantwortlich für das Lesen der Daten aus dem ERP-System, einer Datenbank, einer Datei oder aus anderen bestehenden Quellsystemen in einen gemeinsamen Datenspeicher. Diese Extraktion kann sofort, periodisch, ereignisgesteuert oder auf Anfrage eines Anwenders, z.B. des Data-Warehouse-Administrators, geschehen (vgl. Monitoring-Strategien).
- **Transformation** ist verantwortlich für Weiterverarbeitung der extrahierten Daten aus

der Quelle für das Load in das Ende Ziel. Das Ziel dabei ist die Überführung aller Daten in ein einheitliches Format (Datenmigration), d.h. die Anpassung der Datentypen, der Datumformate, der Maßeinheiten, der Kodierungen, der Kombination bzw. der Separierung von Attributwerten etc.

- **Laden** ist verantwortlich für das Load der transformierten Daten in eine zentrale Datensammelstelle (Data Warehouse).

Der Transformationsprozess umfasst alle Aktivitäten zur Umwandlung der operativen Daten in betriebswirtschaftlich interpretierbare Daten. Er setzt sich aus den Teilprozessen der Filterung, der Harmonisierung, der Aggregation sowie der Anreicherung zusammen [Car04].

### 2.3.1 Filterung

Die erste Schicht der Transformation stellt die Filterung dar. Mit ihrer Hilfe werden die für das Data Warehouse benötigten Daten, die meist aus heterogenen unternehmensinternen und -externen Quellen stammen, selektiert, zwischengespeichert und von Mängeln befreit. Die Filterung unterteilt sich aus diesem Grunde in die beiden Phasen Extraktion und Bereinigung. Im Rahmen der Extraktion werden die unternehmensexternen und insbesondere die operativen unternehmensinternen Daten in speziell hierfür vorgesehene Arbeitsbereiche (**staging areas**) des Data Warehouses eingestellt. Die Bereinigung dient der Befreiung der extrahierten Daten sowohl von syntaktischen als auch von semantischen Mängeln [Car04].

### 2.3.2 Harmonisierung

Die zweite Schicht der Transformation ist die Harmonisierung. Sie bietet die Möglichkeit, die gefilterten Daten zu verknüpfen. Die syntaktische und betriebswirtschaftliche Harmonisierung werden untersucht.

**Syntaktische Harmonisierung** Die operativen und externen Daten weisen meist eine hohe Heterogenität auf und müssen syntaktisch harmonisiert werden. Mit Hilfe dieser Transformationsregeln werden Schlüsseldisharmonien in den Extrakten bereinigt und die Probleme unterschiedlich kodierter Daten sowie die Schwierigkeiten bei der Verwendung von Synonymen und Homonymen gelöst [Car04].

- *Schlüsseldisharmonien* basieren auf Unverträglichkeiten der Primärschlüssel in den extrahierten und bereinigten Daten und entstehen durch die Verwendung unterschiedlicher Zugriffsschlüssel in der operativen Datenhaltung. Künstliche Primärschlüssel werden generiert und die Primärschlüssel der operativen Systeme als Fremdschlüssel mit-

führt.

- Unter *Unterschiedlich kodierten Daten* werden Daten verstanden, die über identische Attributnamen und eine identische Bedeutung verfügen, jedoch unterschiedliche Domänen bzw. Wertebereiche aufweisen. Die Lösung dieses Problems liegt in der eindeutigen Wahl einer Domäne und der Verwendung entsprechender Zuordnungs- bzw. Mapping-Tabellen.
- Bei *Synonymen* handelt es sich um Attribute, die zwar unterschiedliche Namen besitzen, jedoch dieselbe Bedeutung und dieselbe Domäne aufweisen. Die Differenzen können durch eine Festlegung der Attributbezeichnungen und eine Überführung der anderen Attributbezeichnungen bereinigt werden.
- *Homonyme* weisen zwar denselben Attributnamen auf, besitzen jedoch unterschiedliche Bedeutungen. Daher können sie über neu zu vergebende Attributnamen unterschieden werden.

**Betriebswirtschaftliche Harmonisierung** Für eine optimale Überführung der operativen Daten in entscheidungsunterstützende Datenbestände muss die betriebswirtschaftliche Harmonisierung durchgeführt werden. Sie gleicht die betriebswirtschaftlichen Kennziffern ab und legt die gewünschte Granularität (Detaillierungsgrad) der Daten fest [Anc10a].

- Abgleichung der betriebswirtschaftlichen Kennziffern: Die Daten werden betriebswirtschaftlich vereinheitlicht. Daten wie Währung müssen nach gebiets- und ressortspezifischen Kriterien in homogene Werte transformiert werden.
- Festlegung der gewünschten Granularität der dispositiven Daten: Je niedriger die Datengranularität, desto detaillierter sind die Daten. Im Management werden niedrige **Granularitäten** bevorzugt, weil die Analysen dementsprechend detaillierter betrachtet werden können. Zur Festlegung des gewünschten Detaillierungsgrades sind weitere Transformationsregeln nötig.

### 2.3.3 Aggregation

Die **Aggregation** stellt die dritte Schicht der Transformation dar. Sie dient der Erweiterung der gefilterten und harmonisierten Daten um Verdichtungsstruktur. Die Hauptaufgabe der Aggregationsschicht liegt in der Aggregation (Verdichtung, Ansammlung) der Daten. Alle Einzeldaten müssen über Aggregationsalgorithmen zusammengefasst werden. Um bestimmte betriebswirtschaftliche Kennzahlen festzustellen, ist es nötig, werden Vorsummierungen auf dieser Ebene durchgeführt und in den Datenbestand übernommen. Im Rahmen der Aggrega-

tion werden Dimensionshierarchietabellen entwickelt. Alle Einzelelemente sind im Element wie „Gesamt“ oder „Alle“ von solchen Hierarchietabellen aggregiert.

### 2.3.4 Anreicherung

Nach erfolgreicher Aggregation der Daten folgt die Anreicherungsschicht. In dieser Phase werden die betriebswirtschaftlichen Kennzahlen berechnet und integriert.

Dieser Prozess ist wichtig für Informationen, die von mehreren Endanwendern benutzt werden. Dabei werden verschiedene Kennzahlen für verschiedene Anwender errechnet. Beispielsweise ist für einen Produktmanager der wöchentliche Deckungsbeitrag auf Produktebene wichtig, ein Filialleiter hat dagegen mehr Interesse an den jährlichen Deckungsbeiträgen auf Filialebene [Anc10a].

Nach den Transformationsprozessen werden die Daten aus Quellsystemen extrahiert, transportiert, integriert und in das Zielsystem geladen. Jetzt stehen die Daten für analytische Zwecke zur Verfügung.

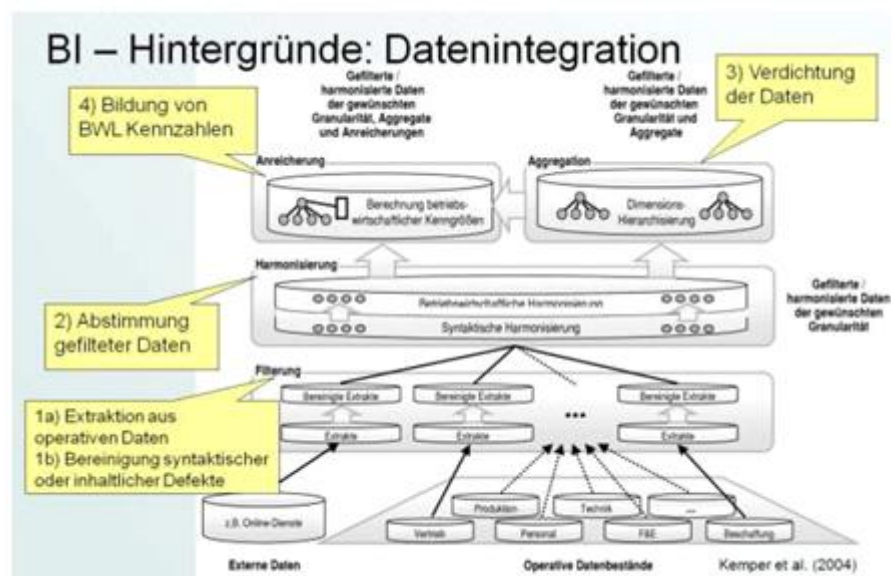


Abbildung 2.2: Normierung der Daten für die spätere betriebswirtschaftlich einheitliche Darstellung

## 3 Grundlage von OLAP

Nach Erläuterung der Beziehung zwischen OLAP und Business Intelligence wird auf die grundlegenden Begriffe des OLAPs eingegangen. Es wird eine Übersicht über die verschiedenen Ansätze, Arten und Modellierungstechniken des OLAPs gegeben.

### 3.1 Definition von OLAP

OLAP ist die Abkürzung für Online Analytical Processing und eine Datenbank-Technologie. Dahinter steht das Versprechen, einer großen Anwenderanzahl einen schnellen, direkten und interaktiven Zugriff auf analysegerechte Datenbestände zu ermöglichen. OLAP sind als Analyse- und Planungswerkzeug auf die Verarbeitung mittlerer und größerer Datenmengen ausgerichtet. OLAP sind auch eine Technologie, die für die Benutzer die Daten interaktiv abfragen, explorieren und analysieren realisieren kann. Der Einsatz von OLAP eignet sich für die verschiedensten Bereiche der Unternehmung. Verschiedene Anwendungsfelder sind [Kar00]:

- Database Marketing
- Konsolidierungslösungen
- Vertriebscontrolling
- Planung/Budgetierung
- Performance Measurement/Balanced Scorecard
- Activity Based Costing
- Risk Management
- Wertbasierte Unternehmenssteuerung



## 3.2 OLAP-Spezifikation

### 3.2.1 Die Grundregeln von Codd

Ein höherer Bekanntheitsgrad mehrdimensionaler Werkzeuge ist durch eine Veröffentlichung von Codd im Jahr 1993 erreicht worden. Er hat den Begriff OLAP geprägt, indem er 12 Grundregeln für eine adäquate Unterstützung von Analyseaufgaben formulierte. 1995 hat Codd die Regeln auf 18 erweitert. Die Regeln lassen sich in vier Kategorien unterteilen [Kar00]:

Basiseigenschaften:

1. Mehrdimensionale konzeptionelle Sichtweise.
2. Intuitive Datenbearbeitung
3. Zugriffsmöglichkeit
4. Batch- und Online-Durchgriff
5. Unterstützung verschiedener Analysemodelle
6. Client-Server-Architektur
7. Transparenz
8. Multi-User-Unterstützung

Spezielle Eigenschaften:

9. Trennung denormalisierter Daten
10. Getrennte Speicherung von OLAP- und Basisdaten
11. Unterscheidung zwischen Null- und fehlenden Werten
12. Behandlung von fehlenden Werten

Berichtseigenschaften:

13. Flexible Berichterstellung
14. Stabile Antwortzeit
15. Dynamische Handhabung dünnbesiedelter Matrizen

Dimensionen:

16. Generische Dimensionen
17. Unbegrenzte Dimensionen und Verdichtungsstufen
18. Uneingeschränkte Operationen über mehrere Dimensionen

### 3.2.2 FASMI

Da diese Regeln einer sehr theoretischen Basis bedürfen und schwer zu behalten sind, wurde die praktische FASMI-Definition entwickelt. Die FASMI-Definition beschreibt OLAP in fünf Schlüsselwörtern: Fast Analysis of Shared Multidimensional Information.

FASMI steht für [Kar00]:

- Fast - Einfache Abfragen sollten in wenigen Sekunden bereitgestellt werden. Generell sollten auch komplexe Abfragen nicht länger als 20 Sekunden benötigen, da damit leicht der Analysepfad durchbrochen wird.
- Analysis - Ein OLAP-System muss in der Lage sein, die benötigte Logik abzubilden. Somit muss der Anwender ohne die Verwendung von Programmierlogik in der Lage sein, auch komplexere Ableitungen zu hinterlegen.
- Shared - Ein OLAP-System sollte mehrbenutzerfähig sein und über entsprechende Zugriffsschutzmechanismen verfügen. Auch Locking-Mechanismen für den Schreibzugriff sollten vorhanden sein.
- Multidimensional - Die Multidimensionalität ist das Schlüsselkriterium von OLAP-Datenbanken, da über Hierarchisierungen und deren Kombinationen die logische Sichtweise von Organisationen und deren Verfahren am besten präsentiert und analysiert werden kann. Zusammenhängende Einheiten von Hierarchien werden in Dimensionen zusammengefasst, von denen mehrere in einer OLAP-Datenbank vorhanden sein und bei der Abfrage miteinander kombiniert werden können [Nil98].
- Information - Es müssen alle Basisdaten und abgeleiteten Daten verfügbar sein. Messpunkt sind die Eingabedaten, nicht der benötigte Speicherbedarf.

### 3.3 Das multidimensionale Datenmodell

Der Import-Calls verwaltet Call-Connection-Statistiken, um zur Connection-Zahlen von Calls z.B. je `opc`<sup>1</sup> oder pro Typ zu analysieren. In diesem Kapitel werden am Fallbeispiel die Grundbegriffe von multidimensionalen Datenmodellen eingeführt.

#### 3.3.1 Grundidee

Ein multidimensionales Datenmodell, das der OLAP-Datenbank zugrunde liegt, wird über Fakten, Kennzahlen, Dimensionen, Würfel und Schema charakterisiert.

Die Fakten sind die zu analysierten Informationen und werden durch mehreren Dimensionen dargestellt. Die Dimensionen können auch die Hierarchieebenen enthalten, die für den Anwender sehr effizient zu bearbeiten sind. Man kann die Werte als Kennzahl bezeichnen, auf die mit Hilfe von den Dimensionen zugegriffen werden soll. Unter Roll-Up oder Drill-Down können die Daten auf die nächsthöhere oder nächstniedrige Dimensionsebene abgebildet werden. Die Daten werden auch durch Verringern oder Erhöhen der Dimension auf dem Würfel analysiert.

In Abbildung Datenwürfel 3.1 mit Connection wird ein Ausschnitt der OLAP-Datenbank, beruhend auf den Daten aus Tabelle Ausschnitt der Tabelle mit Connection, abgebildet. Dargestellt sind Connection mit den Eigenschaften DPC, Typ und OPC,. Die dunkel markierte Zelle stellt die Anzahl der Connection, hier beispielsweise fünf, die Stuttgart 12053 (DPC) anstreben, aus 1408 (OPC) stammen und sich im MOC, (Typ) befinden dar.

OPC	DPC	Typ				
		MOC	MOC	MOC	MTC	MTC
1408	12047	3	-	5	1	2
	12048	2	3	7	3	-
	12053	-	1	5	3	-
	9740	6	-	2	2	1
	9813	-	-	1	4	3
1976	12047	14	8	-	-	2
...	...	...	...	...	...	...
...	...	...	...	...	...	...

Tabelle 3.1: Ausschnitt der Tabelle mit Connection

<sup>1</sup>Eine ausführlichere Beschreibung dieser Eigenschaften findet im Kapitel 5.1.1 statt.

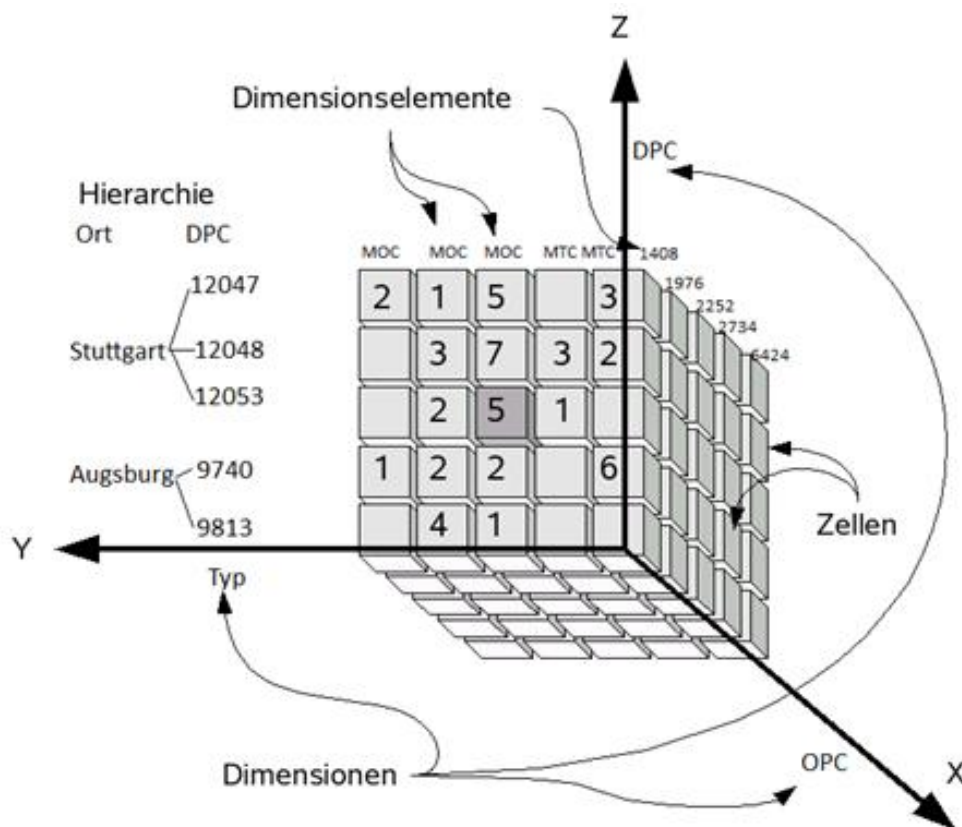


Abbildung 3.1: Beispiel Datenwürfel

### 3.3.2 Fakten

Fakten in einem Würfel sind die tiefste Informationsebene und werden über alle Dimensionen hinweg aggregiert. Fakten bezeichnen als quantitative Daten, numerische Werte zu repräsentieren.

Für die drei Dimensionen der Abbildung Datenwürfel 3.1 mit Connection kann die dunkel markierte Zelle in Funktionsschreibweise ausgedrückt werden: Call-Connection- Statistik(1408, 12053, MOC) = 5. Da jede Dimension auf die tiefsten Granularitätsstufe jeweils fünf Instanzen enthält, ist die Anzahl der Fakten:  $5 \times 5 \times 5 = 125$ .

### 3.3.3 Kennzahl

Der numerische Aspekt der zu analysierenden Fakten wird als **Kennzahl** bezeichnet. Im Fallbeispiel vom Import-Calls werden Connection-Zahlen für Call in Verbindung, Attmpt-Zahlen für Call in Einrichtung.

Man kann mehrere Werte zu einer Kennzahl aggregieren. Wobei können verschiedene Aggregationsfunktionen wie Summierung, Durchschnitt, Minimum oder Maximum, etc. benutzt werden. Die numerischen Werte werden dadurch in die Kennzahl vorausgesetzt. Wenn die Kennzahlen nicht-numerische Werte wie Text- oder Wahrheitswerte enthalten, ist das Aggregationsverfahren im Allgemeinen das Auszählen (Count). Wenn die Dimension geändert wird, ändert sich auch die Kennzahl.

Die speziellen Kennzahlen Connection-Zahlen und Attmpt-Zahlen in Call-Connection-Statistiken ergeben sich aus zwei unterschiedlichen Zählweisen. Die Statistik von Connection-Zahlen zählt alle verbindenden Calls, d.h. die Gesamtzahl aller verbindenden Calls zwischen einer bestimmten OPC und einer bestimmten DPC. Die Attmpt-Zahlen sind die Statistik von aller eingerichteten Calls zwischen einer bestimmten OPC und einer bestimmten DPC.

### 3.3.4 Dimension

Die **Dimension** ist die konzeptionelle Darstellung des multidimensionalen Datenmodells. Wesentliches Merkmal des multidimensionalen Datenmodells ist die Unterteilung der Daten in Fakten und Dimension. Dimension bezeichnet auch als qualitative Daten. Im Fallbeispiel vom Import-Calls Typ, opc und dpc.

Die Eigenschaften werden als Ausprägungen in einer Dimension zusammengefasst (Tag, Monat und Jahr gehören bspw. zu einer zeitlichen Dimension) und tragen so zur Strukturierung bei. Dabei entsteht eine Struktur durch:

- ein Schema
  - (Land, Region, Kontinent) bzw.
  - (Tag, Woche, Monat, Quartal, Jahr)

und

- seine Werte
  - (China, Deutschland, ..., Zypern)
  - (Nordamerika, Ostasien, ..., Zentraleuropa)
  - (Afrika, Amerika, ..., Europa) bzw.
  - (1, 2, ..., 31)
  - (1, 2, ..., 52)

- (1, 2, ..., 12)
- (1, 2, 3, 4)
- (2000, 2001, ..., 2010)

### 3.3.5 Hierarchien

Man kann die Elemente von Dimensionen hierarchisch aufbauen. Eine geografische Dimension besteht aus drei Hierarchieebenen: Kontinent - Region - Land. Durch diese verschiedenen Hierarchieebenen entsteht eine Baumstruktur.

OLAP bzw. eine multidimensionale Datenanalyse läuft im Allgemeinen darauf hinaus, mit der Perspektive verschiedener Kombinationen aus den Dimensionselementen unterschiedliche Sichten auf die Unternehmensdaten vorzunehmen. Dabei ist es für den Analyseprozess kennzeichnend, dass einzelne Sichten schnell von anderen gefolgt werden [Mic06].

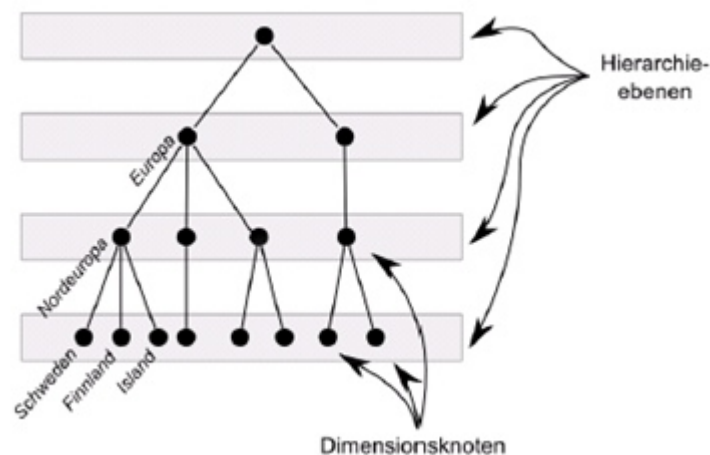


Abbildung 3.2: Dimensionshierarchie

### 3.3.6 Würfel

Der Würfel soll eine Charakteristik multidimensionaler Datenstrukturen zum Ausdruck gebracht werden. Die Achsen des Koordinatensystems bezeichnen die Dimensionen und enthalten die Dimensionselemente.

Im Fall von drei Dimensionen kann der Würfel anschaulich visualisiert werden: wie in Abbildung Datenwürfel mit Connection zu sehen, ist jede Kante des Würfels eine Dimension, die Zelle enthalten die Fakten.

Man kann die Projektionen auf die Achsen aggregieren. In Abbildung Projektionen (vgl.

Abbildung 3.3) im dreidimensionalen Datenwürfel (die Achsen entsprechen der Zuordnung von Abbildung Datenwürfel mit Connection) sind folgende Projektionen dargestellt :

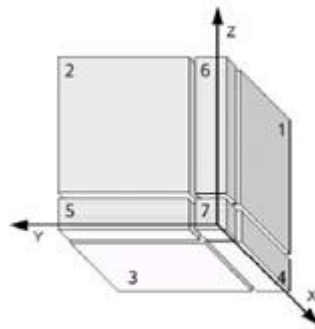


Abbildung 3.3: Projektionen im dreidimensionalen Datenwürfel

1. Projektion auf XZ- **Ebene**

- Connection-Zahl abhängig von OPC über alle Type und DPC

2. Projektion auf YZ-Ebene

- Connection-Zahl abhängig von Typ über alle OPC und DPC

3. Projektion auf XY-Ebene

- Connection-Zahl abhängig von Typ und OPC über alle DPC

4. Projektion auf X- Achse

- Connection-Zahl abhängig von OPC über alle Type und DPC

5. Projektion auf Y- Achse

- Connection-Zahl abhängig von Typ über alle OPC und DPC

6. Projektion auf Z- Achse

- Connection-Zahl abhängig von DPC über alle Type und OPC

7. Projektion auf alle Achse

- Connection-Zahl ohne Einschränkung

Im n-dimensionalen Würfel gibt es also  $2^n$  Projektionen um Daten zu verdichten. Dazu gehören die sieben dargestellten Projektionen, siehe Abbildung Projektionen im dreidimensionalen

Datenwürfel, sowie das Aggregat der einzelnen Zelle. So erhält man  $2^3 = 8$  Verdichtungen im dreidimensionalen Datenwürfel [Got00].

### 3.4 OLTP und OLAP im Vergleich

**OLTP** (Online Transaction Processing) bezeichnet ein Benutzungsparadigma von Datenbanksysteme und Geschäftsanwendungen, bei dem die Verarbeitung von Transaktionen im Vordergrund steht [Sou]. OLTP kann die Daten in einer traditionellen relationalen Datenbank speichern, vor allem grundlegende, täglich Transaktionsverarbeitung, wie Bankgeschäfte.

OLAP ist eine wichtige Anwendung von dem Data-Warehouse-System und unterstützt die Analyse von komplexen Operationen, wobei der Schwerpunkt auf den Entscheidungsprozess zu unterstützen, und bietet ein intuitives Verständnis der Abfrageergebnisse.

Die folgende Tabelle stellt die wesentlichsten Unterschiede zwischen OLAP und OLTP dar [Kur99]:

	OLTP	OLAP
Funktion	Datenverarbeitung basierend auf einem zuverlässigen Transaktionskonzept (Commit, Roll-back), Day-to-Day Business Operations	Unterstützung des Prozesses der Entscheidungsfindung
DB Design	prozeß-bzw. transaktionsorientiert	objektorientiert
Datenzugriff	100 Datensätzen (read/write)	10 Millionen Datensätzen (read/write)
Daten	operativ (Atomic Data), aktuell, flüchtig und detailliert	analytisch verdichtet
DB-Schema	normalisiert	multidimensional
Organisation der Daten	hierarchisch	anhand vorab definierter Dimensionen und Kennzahlen, multidimensionales semantisches Datenmodell
Datenhaltung	Daten werden in unterschiedlichen Systemen (verteilt, heterogene Systemlandschaften) gehalten	Daten können auf unterschiedlichen Systemen gehalten und verarbeitet werden



Plattformen	heterogen Systemlandschaften	verschiedene Datenbanken und Betriebssystem
Behandlung der Zeit	keinen Schlüssel „Zeit“, keine Zeitreihenanalyse	Dimension „Zeit“ für komplexe Analysen, Zeitreihenanalysen, Prognose und Simulation
Aktualität der Daten	nur aktuelle Daten	nur historische verdichtete Daten, auch Prognosedaten
Systemauslastung	eher konstant	viele Auslastungsspitzen
Antwortzeit verhalten	optimiert für effiziente Speicherung der Daten, gutes Update, aber nicht garantiertes Abfrageverhalten	konstantes Abfrageverhalten durch „Materialized Views“ oder Subwürfelbildung
Datenvolumen	Gigabyte	Gigabyte bis mehrere Terabyte
Anschaffungskosten (HW + SW)	moderat bis teuer	minimal bis moderat
Amortisation	langfristig	kurzfristig
Verwendung durch Anwender	sehr häufig	Analysezwecke

Tabelle 3.2: Unterschiede zwischen OLTP und OLAP

## 3.5 Das relationale Schema

Um die Mehrdimensionalität in einem relationalen System nachzubilden werden spezielle Datenbankschemen benötigt. Die Daten werden in Star- oder Snowflake-Schema modelliert.

### 3.5.1 Star-Schema

Das Star-Schema (oder Sternschema) unterscheidet zwischen zwei Arten von Tabellen [Wol98]. Im Mittelpunkt steht die Faktentabelle (enthält Bewegungsdaten) mit Spalten für jede Kennzahl sowie für die Fremdschlüssel zu jeder Dimension. Die Dimensionstabellen (mit Stammdaten) enthalten beschreibende Informationen zu den Fakten. Die Schlüssel der zur Faktentabelle beitragenden Dimensionen sind die Verbindung zwischen Fakten- und Dimensionstabellen [Got00].

Die Faktentabelle *Conrate-fact* steht im Mittelpunkt, um sie herum positionieren sich die

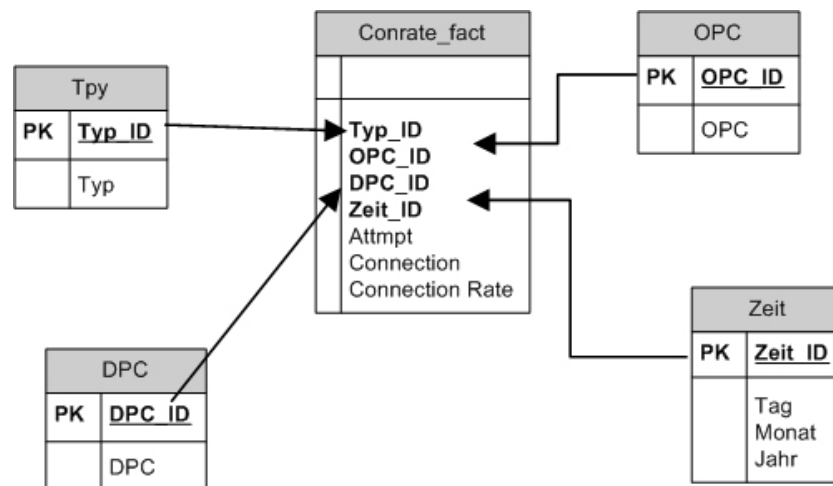


Abbildung 3.4: Call-Connection-Statistik in Star-Schema

Dimensionstabellen *Typ*, *OPC*, *DPC* und *Zeit*. Da sich dadurch ein sternförmiges Gebilde ergibt, liegt die Analogie zum Star-Schema nahe.

Die Faktentabelle in Abbildung Call-Connection-Statistik (Abbildung 3.4) in Star-Schema besitzt drei Kennzahlen *Attmpt*, *Connection* und *Connection Rate*, die die Call-Connectionzahlen darstellen.

Die Faktentabelle besitzt als Primärschlüssel einen zusammengesetzten Schlüssel aus den Primärschlüsseln der einzelnen Dimensionstabellen.

Das Star-Schema und seine Varianten zeichnen sich durch folgende Vorteile aus [Car04]:

- Einfache und daher intuitive Datenmodelle.
- Geringe Anzahl von **Join**-Operationen.
- Geringe Anzahl physischer Data-Warehouse-Tabellen.
- Geringer Aufwand im Rahmen der Data-Warehouse-Wartung.

Dem stehen auch Nachteile gegenüber:

- Verschlechtertes Antwortzeitverhalten bei sehr großen Dimensionstabellen.
- Redundanz innerhalb der Dimensionstabellen durch das mehrfache Festhalten identischer Fakten.

### 3.5.2 Snowflake-Schema

Das Snowflake-Schema ist eine Erweiterung des Star-Schemas. Bei diesem bleibt die Faktentabelle wie beim Star-Schema analog. Jedoch sind im Gegensatz zum Star-Schema differierend, denn diese enthalten nicht mehr alle Dimensionselemente, sondern nur Daten über die Dimensionshierarchien. Die Dimensionen werden dafür weiter verfeinert, indem sie klassifiziert oder normalisiert werden. Jedenfalls werden die Dimensionstabellen dabei um die Attribute erweitert, damit jede Ausprägung einer Dimension in einer eigenen Tabelle dargestellt werden kann. D.h. beim verbreiteten Snowflake-Schema werden die Daten in den Dimensionstabellen in der **Dritte Normalform** (3NF) gespeichert. Durch die Normalisierung entsteht für jede Hierarchiestufe einer Dimension eine eigene Tabelle und führt demnach zu kleineren und besser strukturierten Datenmengen. Durch diese Weiterverzweigung des Datenmodells entsteht die Form einer Schneeflocke, was diesem Schema den Namen verleiht [Sou].

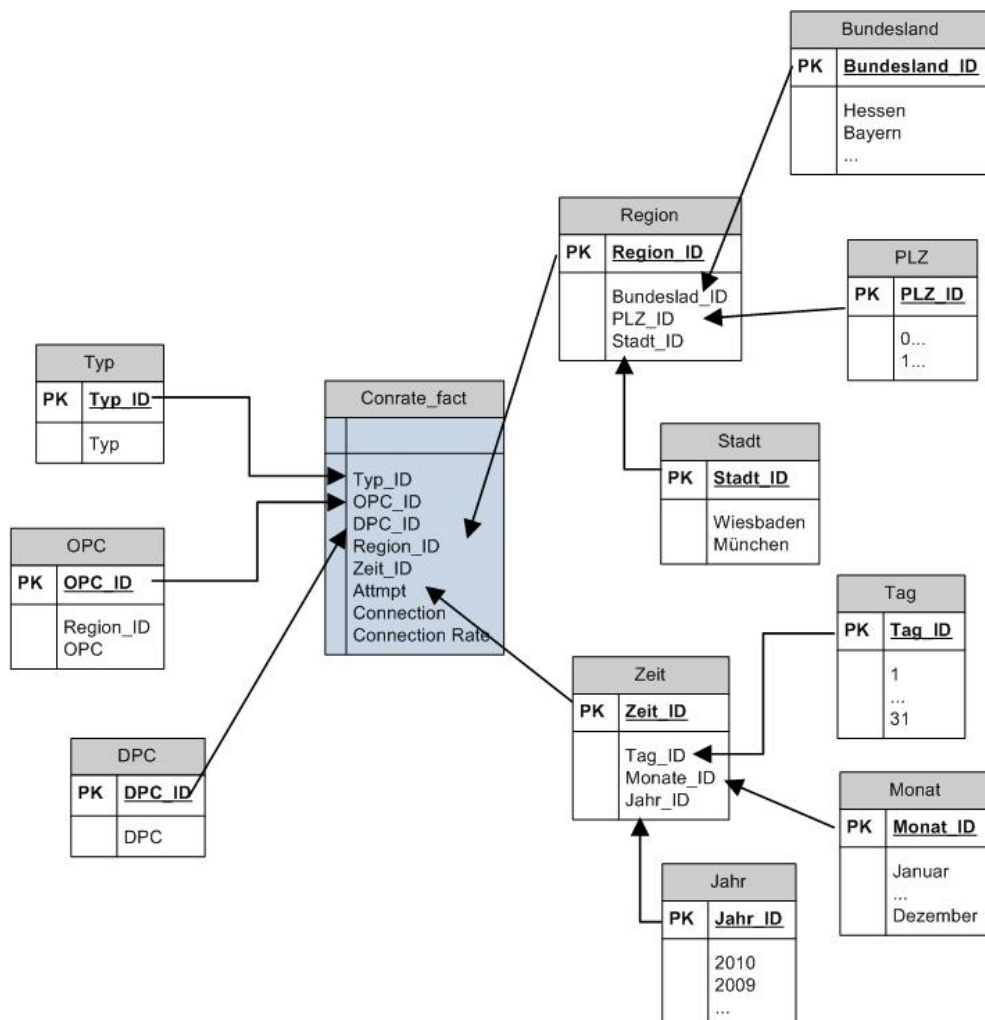


Abbildung 3.5: Snowflake-Schema

---

Die Dimensionstabellen enthalten nicht mehr alle Dimensionselemente, sondern lediglich Daten über die Dimensionshierarchien, das heißt ein Tupel für jede Hierarchieebene. Die Dimensionstabellen sind über die Schlüsselattribute sowohl mit der zentralen Faktentabelle als auch mit den Attributtabeln, welche die deskriptiven Informationen über die Dimensionselemente enthalten, verknüpft. Der Vorteil dieser Struktur liegt in teilweise deutlich kürzeren Zugriffszeiten [(Hr00)].

### 3.6 Funktionen des OLAPs

Wichtige OLAP-Funktionen sind Roll-Up, Drill-Down, Slice und Dice und Pivoting. Durch diesen Funktionen kann der Benutzer also die Daten aus relativ aggregierten Daten analysieren, um die Zusammenhänge mit den gesuchten Informationen herauszufinden. d.h. man arbeitet sich von der Übersicht zum Detail durch.

**Roll-Up** Das Roll-Up geht von der detaillierten Daten wieder auf die verdichtete Daten über, wodurch sich die Granularität verringert. Es ist sinnvoll, wenn Roll-Up der Elemente einer Stufe zu den Elementen einer weiteren entlang der Dimensionshierarchie geschieht. Die Struktur der Daten in Form von Dimensionshierarchien und der Kennzahl wird von Roll-Up benutzt.

**Drill-Down** Eine inverse Operation zu Roll-Up ist Drill-Down, das die Daten verfeinert. Beim Drill-Down kann man die Steigerung von einem bestimmten Aggregationsniveau auf die jeweils nächsttiefere und detailliertere Verdichtungsstufe. Diese Operation stellt die genauere Daten dar.

Roll-Up und Drill-Down stehen zur Verfügung, um innerhalb der Dimensionshierarchien zu navigieren.

**Slice und Dice** Diese Operation der Selektion ermöglicht die Auswahl zur Bildung von horizontalen oder vertikalen Ebenen (Slice = Scheibe) und Subwürfeln bzw. einzelnen Zellen (Dice = Würfel) am Würfel. Die Operation können mit Roll-Up und Drill-Down kombiniert werden.

Ein Slice ist in dem dreidimensionalen **Hypercube** eine Scheibe, die aus dem Datenwürfel entnommen wird. Faktisch wird dies durch eine Beschränkung einer Dimension auf einen Wert umgesetzt. Das Ergebnis ist dann eine zweidimensionale Matrix.

Ein Dice ist ein mehrdimensionaler Ausschnitt eines Hypercubes. Hierbei werden mehrere

Dimensionen jeweils durch eine Menge von Dimensionselementen eingeschränkt. Das Ergebnis ist ein neuer multidimensionaler Datenraum, der extrahiert oder weiterverarbeitet werden kann [Car04].

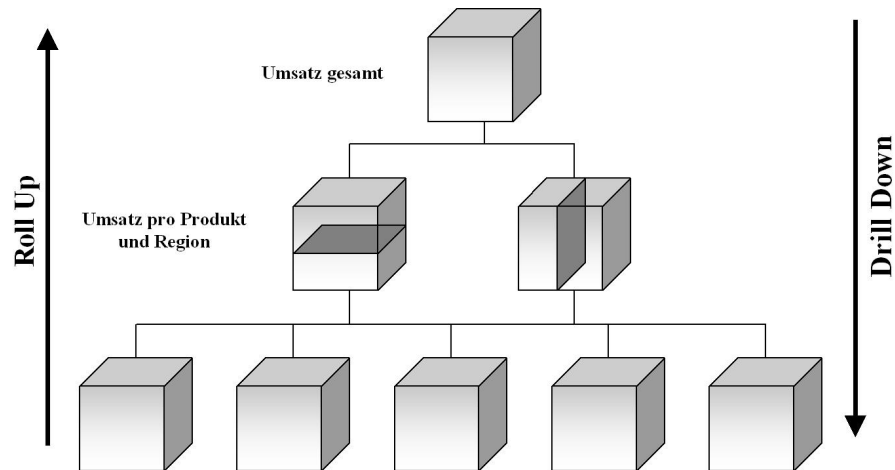


Abbildung 3.6: Roll-Up & Drill-Down

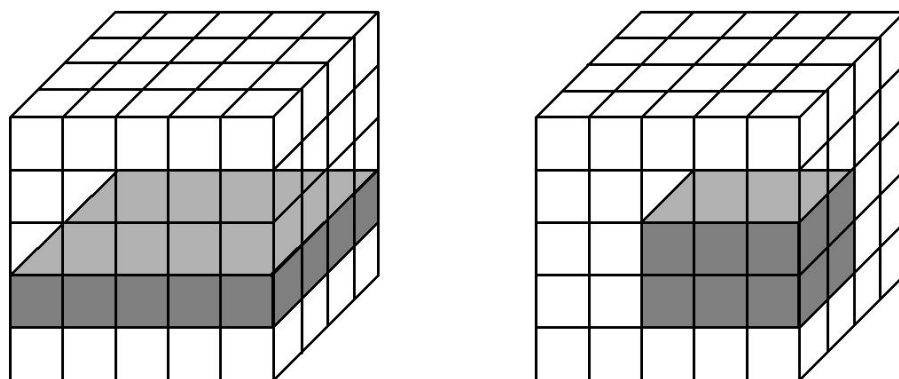


Abbildung 3.7: Slice & Dice

**Pivoting** Unter Pivoting (deutsch Pivotierung) versteht man das Drehen des Würfels durch Vertauschen von Spalten-, Zeilen- oder Seitendimensionen, wodurch eine neue Sicht auf den analysierten Datenbestand aufgebaut wird.

---

### 3.7 ROLAP, MOLAP, DOLAP und HOLAP

Für OLAP-Datenmodelle existieren die verschiedenen Architekturen von OLAP-Datenbanken. Zu den Modellierungsverfahren gehören MOLAP, ROLAP, HOLAP und DOLAP.

**ROLAP - Relationales OLAP** Die Daten, die zur Analyse benötigt sind, werden in relationalen Datenbanken gespeichert. Die zur Abfrage notwendigen SQL-Anweisungen werden zur Laufzeit erzeugt und auf die Transaktions- oder Aggregationstabellen abgesetzt. Relationales OLAP kann dadurch eine grössere Anzahl von Dimensionen und Ausprägungen verarbeiten, kann dem Benutzer aber keine verlässliche Aussage über die zu erwartenden Antwortzeiten liefern [Nil98].

**MOLAP - Multidimensionales OLAP** Die klassische OLAP-Architektur ist Multidimensionales OLAP. Durch die für mehrdimensionale Zugriffe optimierte Speicherung ist die Verarbeitungsform praktisch unschlagbar performant. Die Orthogonalität der Anwendbarkeit von Funktionen über beliebige Dimensionen kann als wesentliches Kriterium für Mehrdimensionalität angesehen werden [Kar00].

**DOLAP - Desktop OLAP** Desktop OLAP-Systeme benötigen kein Server-Backend, um OLAP-Abfragen zur Verfügung stellen zu können. Stattdessen werden alle benötigten Daten aus den zur Verfügung stehenden Datenquellen auf den Client geladen und dort multidimensional aufbereitet. Dieses Vorgehen verursacht einen hohen Netzwerkverkehr und benötigt sehr gut ausgestattete PCs, deren Wartbarkeit aufgrund von unterschiedlichen Konfigurationen stark beeinträchtigt sein kann [Nil98].

**HOLAP - Hybrides OLAP** Hybrides OLAP nutzt sowohl die Möglichkeit der physikalisch multidimensionalen wie auch der relationalen Speicherform im Star- oder Snowflake-Schema. Hierdurch können die Vorteile beider Formen miteinander kombiniert werden - hierdurch wird allerdings auch die Komplexität des Gesamtsystems erhöht und kann so eine zügige Implementierung verzögern [Nil98].

	ROLAP	MOLAP	DOLAP	HOLAP
Antwortzeit	nicht linear	linear	linear	MDB-Teil linear, sonst nicht linear
Technisch möglicher Aktualisierungszyklus	Echtzeit	Zyklisch	Echtzeit	Echtzeit
Mögliche Komplexität	Mittel	Hoch	Mittel	Hoch
Datenzugriff	SQL - nur lesend	API - lesend und schreibend	Proprietär - nur lesend	API -lesend und schreibend
Resultierender Netzwerkverkehr bei Zugriffen	Mittel	Gering	Hoch	Mittel
Praktikable Grössenbeschränkung pro Datenbank	Terabyte-Bereich	Mehrere hundert Gigabyte	Bis 100 Megabyte	Mehrere hundert Gigabyte
Bemerkungen	Datenbank-Administrator erforderlich	Kurze Implementierungszeiten	Preisgünstige Einstiegsmöglichkeit	Datenbank-Administrator erforderlich

Tabelle 3.3: Vergleich der OLAP Architektur

## 4 Vergleich von OLAP - Tools

Als Fenster zum OLAP und den enthaltenen Daten dienen den Anwendern Benutzeroberflächen und Frontends. Verschiedene Frontends kommen in einem Unternehmen in den meisten Fällen zum Einsatz, weil die Anforderungen von Anwender zu Anwender und Abteilung zu Abteilung stark variieren können.

JasperSoft, Pentaho und Palo werden hier betrachtet, da es drei der größten unternehmensbezogenen OLAP-Initiativen in der Welt sind.

Die Beschreibung konzentriert sich auf den Allgemein, die Installation, die Analyse und den Report des OLAP Werkzeugs von JasperSoft, Pentaho und Palo.

### 4.1 BI Konzept

Es sind für Unternehmen meist mit hohen Kosten verbunden, um ein Data Warehouse implementieren und um ein Business Intelligence Projekt durchführen zu können. Die Mehrzahl der Kosten ist für die Investition in geeignete Technologien. Entsprechende Produkte stehen zur Realisierung von Business Intelligence zur Verfügung. Die Produkte alleine können Business Intelligence nicht betreiben, sondern müssen mit anderen kombiniert werden, um einsetzbare Systeme zu bilden. Es ist schwierig, dass dies wegen verschiedener Probleme bildet. Die proprietäre Software lösen die Probleme, weil sie alle benötigten Komponenten für Business Intelligence beinhalten.

Heutzutage entwickeln sich Open Source Project wie Jasper und Pentaho, die beide können Business Intelligence Lösungen aus „einer Hand“ anbieten. Jedox versucht mit dem Palo eine umfassende Business Intelligence Lösung kostengünstig anzubieten.



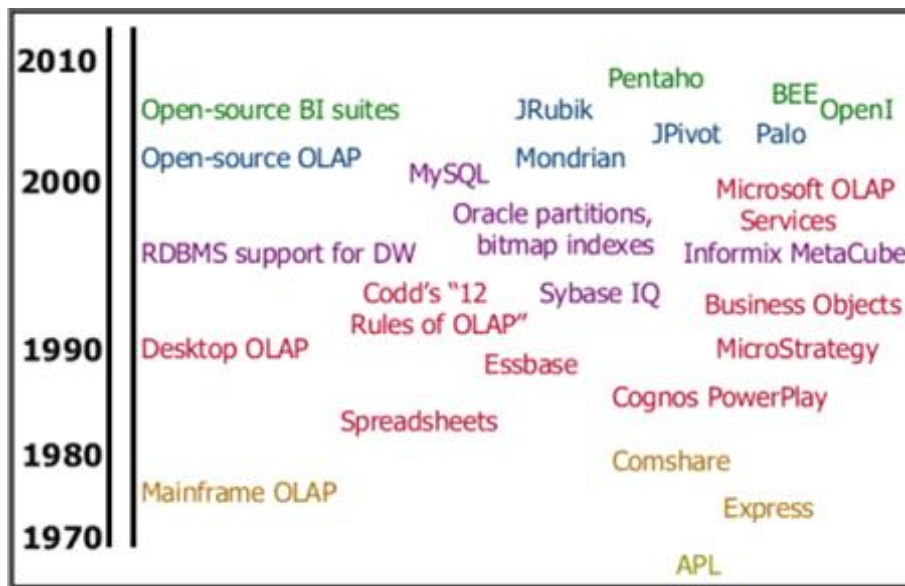


Abbildung 4.1: Historische Entwicklung im BI Bereich [Pen10]

## 4.2 Jaspersoft

Jaspersoft versucht mit dem JasperServer, dem JasperAnalysis, JasperETL und iReports eine umfassende BI-Lösung anzubieten. In dieser Arbeit werden ausschließlich die Softwareprodukte JasperServer, JasperAnalysis und JasperETL zur Realisierung des BI-Modells verwendet.

### 4.2.1 JasperServer

JasperServer ist ein hoch leistungsfähiger Berichtsserver und auch eine Benutzeroberfläche, der sich als Stand-Alone-Applikation implementieren oder mit anderen Anwendungen integrieren lässt und zahlreiche Abfrage-, Reporting- und Datenanalyse-Funktionen für Endanwender bietet [JAS10].

JasperServer baut auf interaktiven Web 2.0 AJAX-Technologien und dynamischem HTML auf. Endanwender können damit selbstständig Berichte und **Ad-hoc-Abfragen** erstellen, ohne auf die Hilfe von IT-Mitarbeitern angewiesen zu sein. Die Benutzer benötigen lediglich einen Webbrowser, um mit unternehmensspezifischen Datenansichten arbeiten zu können; Kenntnisse zu Datenbanken oder Abfragesprachen sind nicht notwendig [JAS10].



Abbildung 4.2: Übersicht JasperServer

JasperServer gibt es in folgenden Funktionen:

- Umfangreiche Visualisierung
- Integrierte Analyse
- SaaS - Fähigkeit
- Unmittelbarer Portalzugriff
- Berichtsplanung und Berichtsverteilung

#### 4.2.2 Installation

Eine Standardinstallation des JasperServers besteht aus einer MySQL-Datenbank und einem mitgelieferten Tomcat-Server. Beide Komponenten sind Optional und können auch durch bereits vorhandene Web- und Datenbankserver ersetzt werden.

Die eigentliche Jasper-Applikation kommt in Form einer Java-WebApp und liegt auch im entsprechenden Deployment-Verzeichniss des Servers. Sie besteht aus der eigentlichen Server-Applikation und den notwendigen Bibliotheken wie JasperReports (zur eigentlichen Berichterstellung), Quartz (Scheduling) und anderen Bibliotheken für Datenbankverbindungen und den diversen Ausgabeformaten.

Ein Installation ohne dieses Bundle ist ebenfalls möglich und die entsprechenden Pakete bei SourceForge(<http://sourceforge.net/>) verfügbar [NET10].

### 4.2.3 Jasper Analysis

JasperAnalysis ist die Ergänzung und die Erweiterung von den leistungsstarken JasperServer. JasperAnalysis bietet leistungsstark webbasierte OLAP-Funktionen, mit der Anwender die Auswertungen vornehmen, um Ursachen, Trends und Muster zu entdecken. JasperAnalysis ermöglicht multidimensionale Analyse der Datenbestände anhand zuvor aufgestellter Hypothesen und dient zur Aufspürung von nicht auf Anhieb sichtbaren Zusammenhängen.

JasperAnalysis dient zur Datenaufbereitung und Datenweiterverarbeitung. Hier können OLAP-Modelle erstellt werden. Eine ausführlichere Beschreibung dieses Tools findet im Kapitel 5.2 statt.

### 4.2.4 Jasper ETL

JasperETL ist ein Tool, das die bequeme Übernahme von Daten aus unterschiedlichen Datenquellen in ein Datawarehouse ermöglichen soll.

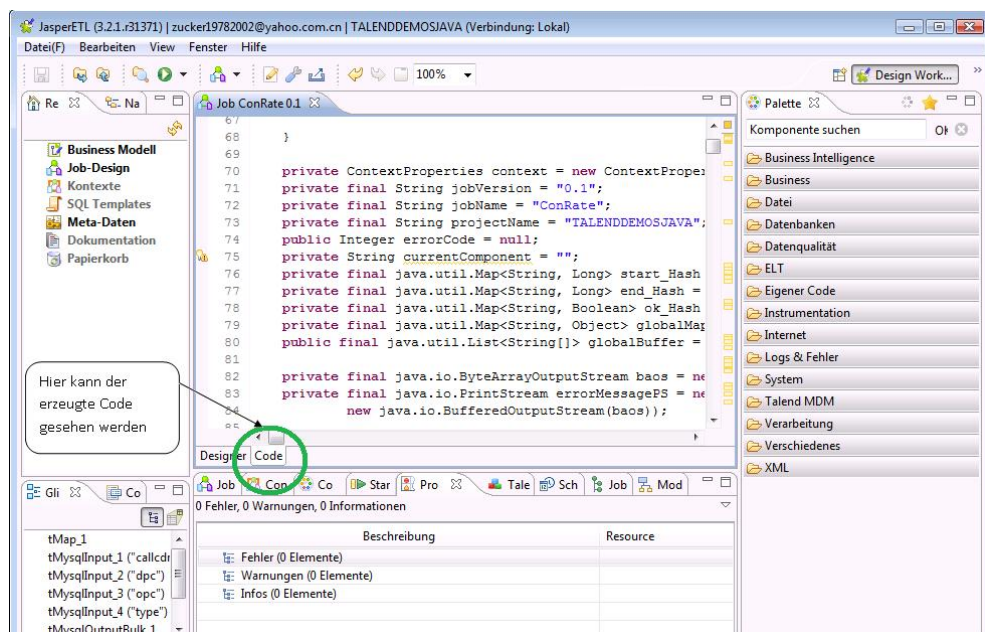


Abbildung 4.3: Übersicht JasperETL

JasperETL kann die ETL-Prozesse durch Drag und Drop auf einer Benutzeroberfläche erstellen. Die Code (Java oder Perl) können durch einen Codegenerator die grafischen erstellten Prozesse automatisch umgewandelt werden. JasperETL basiert auf einer Eclipse-Umgebung. Neue Jobs und die zugehörigen Komponenten können bei dem Repository Baum View angelegt, bearbeitet und verwaltet werden. Der Paletten View umfasst eine große Anzahl an Komponenten, z.B. die Verbindungen zu verschiedener Datenbanken oder Verwaltung von

Dateien (Ein- und Ausgabe).

JasperETL fällt dem Prozessablauf durch den großen Funktionsumfang beim ETL leicht.

### 4.2.5 iReport

Der iReport ist ein grafisches Tool. Es bieten die Möglichkeit, die Inhalte eines Berichts innerhalb eines Dokuments (JRXML) zu definieren. Außerdem können Berichte und zugehörige Elemente von einem zentralen Ort aus verwaltet werden.

Ein Report Inspektor enthält die einzelnen Komponenten des zu erstellenden Berichts, kann die Darstellung des Berichts angepasst werden, wie z.B. die Definition von Kopf- und Fußzeilen in einem Report stattfinden zu können.

Während der Erstellung der Berichte entsteht gleichzeitig automatisch einen XML-Code. Zunächst wird eine „.jrxmlDatei erzeugt, der nächste Schritt erzeugt daraus eine „jasper“-Datei, woraufhin eine „jrprint“-Datei entsteht. Jetzt kann der Bericht z.B. als .html- oder .pdf-Datei geöffnet werden [Anc10a].

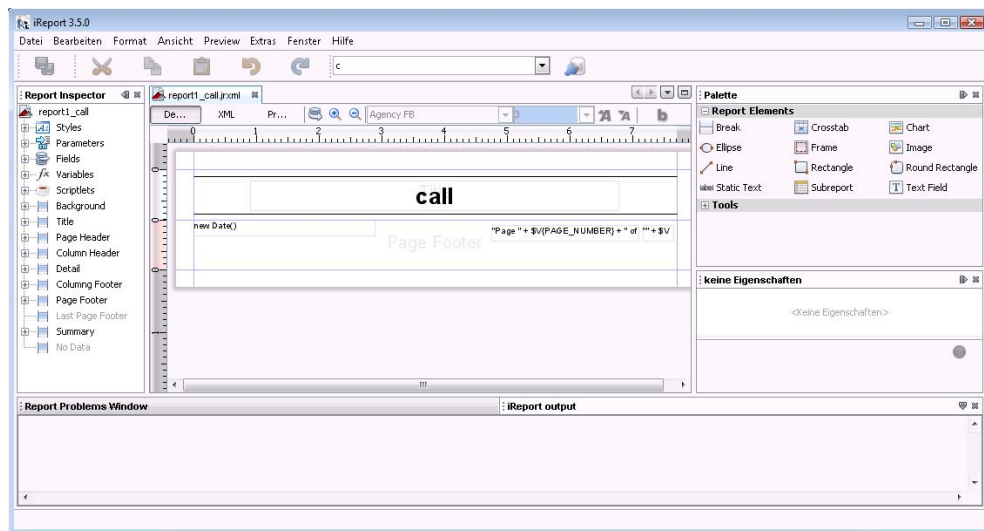


Abbildung 4.4: iReport

Man kann auch mit iReports [JAS10]:

- Berichte und Dashboards mit mehreren Tabellen, Kreuztabellen, Diagrammen und Grafiken erstellen.
- Diagramme unter Verwendung eines Drag & Drop-WYSIWYG-Designers erstellen.
- Metadatenschicht-Abfragen für SQL, MDX und Jaspersoft mittels eines grafischen

Tools für die Abfrageentwicklung erstellen.

- Berichte von einer Desktopumgebung erstellen, testen und ausführen.
- Zugriff auf jede Unternehmens- und Kundenspezifischen Datenquelle erhalten.
- Lokalisierte Berichte mit einer Berichtsdefinition definieren.
- Berichte in mehreren Formaten wie HTML, PDF und RTF vorab prüfen.

## 4.3 Pentaho

### 4.3.1 Allgemeines

Pentaho ist eine Open Source Business Intelligence Software und bietet eine BI-Plattform. Pentaho ist eine kollaborative BI-Lösung mit einer AJAX-basierten Oberfläche, die auch für End-User leicht zu bedienen ist. Über die User Konsole können alle Pentaho-User zentral auf die BI-Daten zugreifen und mit ihnen arbeiten.

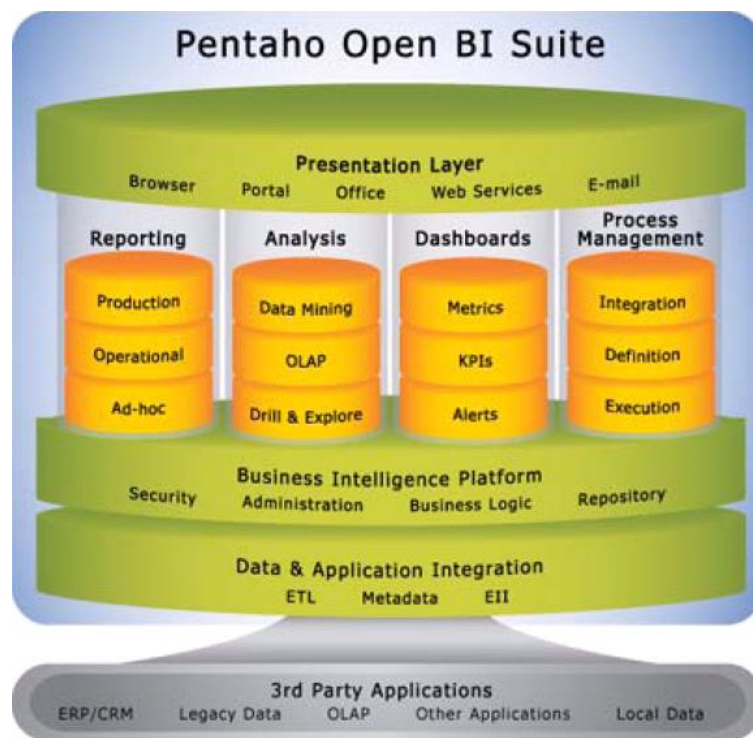


Abbildung 4.5: Aufbau der Pentaho BI Suite [Pen10]

Die Pentaho BI Suite umfasst vier Teilbereiche zur Auswertung von Geschäftsdaten und Op-

timierung von Geschäftsabläufen: Reporting, Analyse, Dashboards und Prozessmanagement.

Die BI Suite stellt ein komplettes Paket aus unterschiedlichen Software-Komponenten zur Verfügung. Das Unternehmen Pentaho agiert dabei als Sponsor und Besitzer von weiteren Open-Source-Produkten, die mit ihren Funktionalitäten in die Pentaho BI Plattform integriert sind. Alle Teilprojekte, die mit Pentaho zu tun haben, werden unter dem übergeordneten Label „Pentaho BI Project“ geführt. Hierzu gehören [Fre10]:

- Mondrian, der Open Source OLAP Server,
- JFreeReport, die Open Source Reporting Komponente,
- Kettle, als Lösung der Datenintegration (ETL),
- Weka, als Komponente für Data Mining,
- Und Pentaho selbst, als umfangreiche Open Source BI Suite.

Pentaho integriert mehrere bewährte Open Source-Projekte aus dem Bereich Business Intelligence: Mondrian, Kettle oder JFreeReport.

Pentaho bietet flexible Deployment-Möglichkeiten. So kann das Produkt als eingebettete Komponente in anderen Anwendungen, als angepasste BI-Lösung oder als vollständig integrierte „out-of-the-box“-Plattform eingesetzt werden. Wer den gesamten Funktionsumfang von Pentaho nutzen möchte, benötigt dafür einen J2EE-Applikationsserver. ETL, Reporting und viele andere Funktionen können jedoch ebenso in Standalone-Anwendungen eingebunden werden. Durch die LDAP-Anbindung und das Java Single Sign-On ist es daneben möglich, Pentaho in bestehende IT-Landschaften einzubinden [Anc10b].

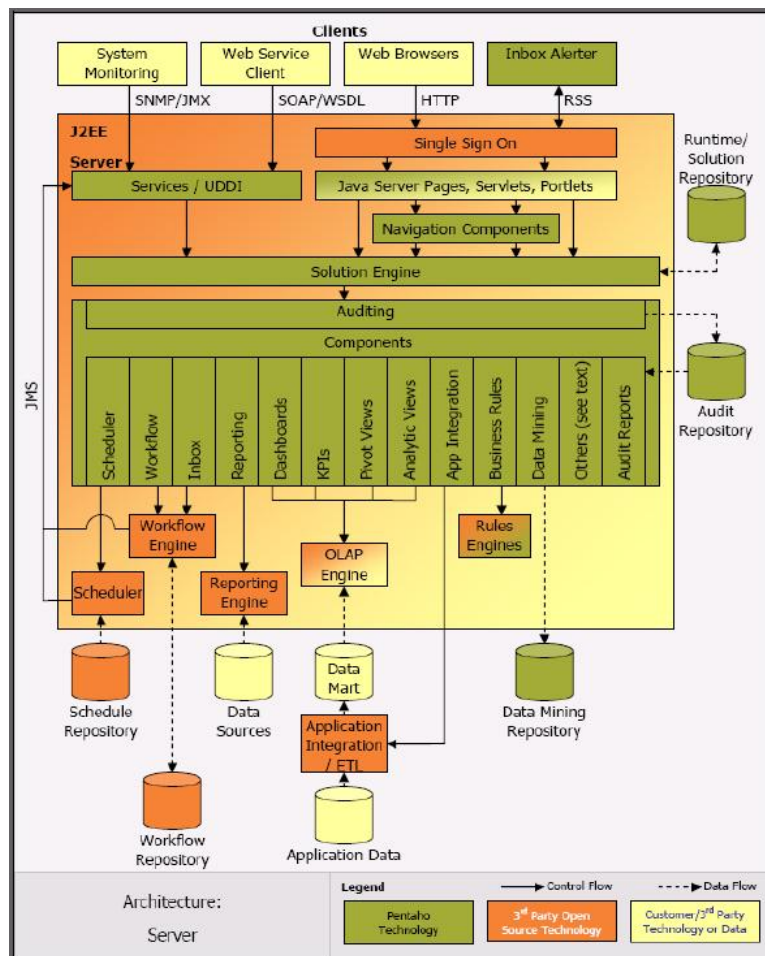


Abbildung 4.6: Architektur des Pentaho Servers [Pen10]

### 4.3.2 Installation

Der Link zum Download von Pentaho sind auf der offiziellen Pentaho Webseite unter [http://www.pentaho.com/products/try\\_bi\\_suite.php](http://www.pentaho.com/products/try_bi_suite.php) verfügbar. Die Pentaho Open BI Suite Pre-configured Installation (PCI) Version ist eine komplette Server Installation, die direkt nach dem Download einen schnellen Einstieg in die BI-Lösung von Pentaho bietet. Sie beinhaltet einen vorkonfigurierten Applikation Server, eine Datenbank und Beispiele, welche Lösungen mit der Pentaho BI Suite entworfen werden können. In der Windows-Variante der PCI Version ist zusätzlich ein Java Runtime Environment (JRE) enthalten [Pen10].

### 4.3.3 Pentaho Analysis Services (Mondrian)

Der Pentaho Analysis Services basiert auf dem Open Source Mondrian. Man kann mit der Mondrian Schema-Workbench, der Mondrian-Webanwendung und dem Aggregation Designer



für die Analyse zusammen benutzen.

**Mondrian Schema-Workbench** Die Mondrian Schema-Workbench wird entwickelt, um komplexere Schemata für OLAP zu erstellen. Ein visuelles Schema kann mit Mondrian Schema-Workbench erstellt werden und als XML-Datei abspeichern werden. Um ein Schema zu erstellen, muss ein Würfel definiert werden, der eine Fakte-Tabelle und mehrere Dimension-Tabelle enthält. MDX-Abfragen können in einem Editor ausgeführt werden.

Man kann ein neues Schema über diese Benutzeroberfläche erstellen oder ein vorhandenes Schema bearbeiten. Sobald ein neues Schema erstellt wird, muss es auf eine Datenbank zugriff werden, um die Datenbankverbindung zu konfigurieren. Per Mausklick über das Würfelsymbol wird ein Würfel generiert, können weitere Elemente, bspw. Dimensionen, Hierarchien, Levels oder Kennzahlen in diesen Würfel eingebunden werden und die zugehörige Fakte-Tabelle mit ihren Dimension-Tabelle hinzugefügt werden (Abbildung Mondrian Schema-Workbench 1). Die Elemente, aus denen der Würfel aktuell besteht, können in der Baumstruktur des Schemas bearbeitet werden (Abbildung Mondrian Schema-Workbench 2). Wenn ein Element ausgewählt wird, können die zugehörigen Attribute konfiguriert und editiert werden (Abbildung Mondrian Schema-Workbench 3). Das Schema kann in der jeweiligen BI-Plattform publizieren.

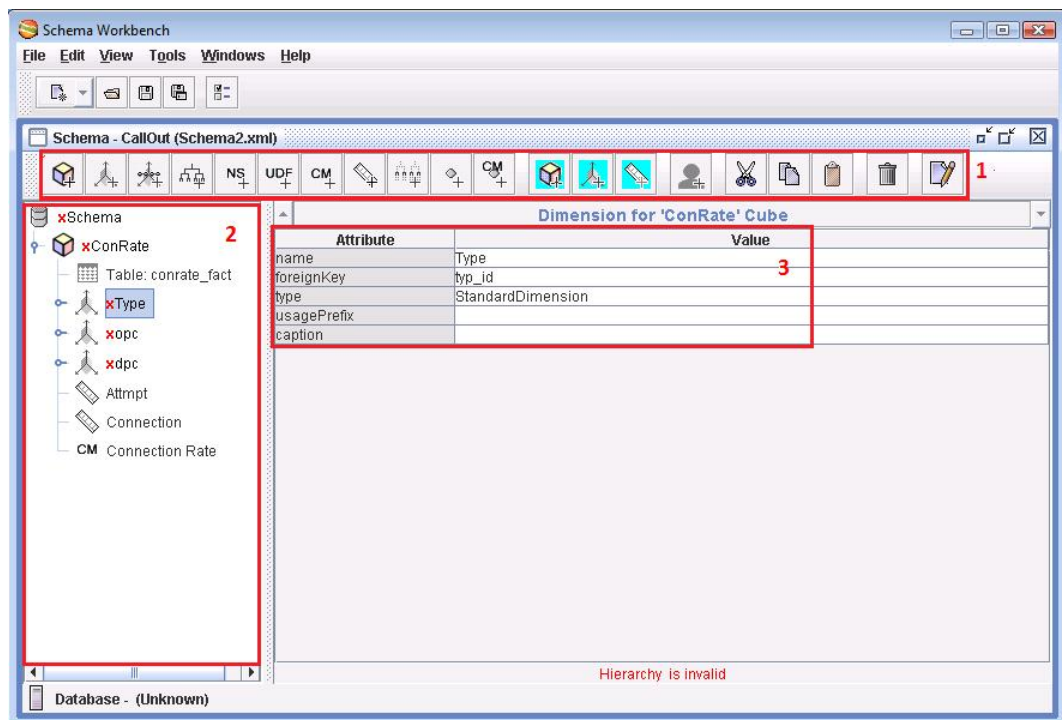


Abbildung 4.7: Mondrian Schema-Workbench



**Mondrian-Webanwendung** Für das Webfrontend für Mondrian gibt es momentan drei Anwendungen: Das von Pentaho erst kürzlich erworbene Produkt CleraView des Drittanbieters LucidEra, JPivot und die Pentaho Analysis Tools (PAT). Klassisches Webfrontend war bis vor kurzem das Open Source-Projekt JPivot. JPivot besitzt alle Funktionen zur Durchführung von OLAP-Analysen, wurde jedoch von Anwendern als umständlich und visuell wenig ansprechend wahrgenommen. Seit kurzem setzt Pentaho daher Clearview ein, allerdings nur in der Enterprise Edition. Die Community Version enthält weiterhin JPivot (siehe Abbildung 4.8), es ist aber geplant, JPivot durch das Open Source-Projekt Pentaho Analysis Tools (PAT) zu ersetzen. PAT gilt als anwenderfreundlicher und zeitgemäßer als JPivot, eine erste stabile Version soll Weihnachten 2009 veröffentlicht werden. Da PAT im Gegensatz zu JPivot noch nicht in die Pentaho BI-Plattform integriert ist, die meisten Benutzer aber bis zur Einbindung von PAT mit JPivot arbeiten dürften [Anc10a].

Die Daten werden mit JPivot über MDX-Sprachen abgefragt, um die Daten zu zugreifen. JPivot ist für grafische Aufbereitung in den Pentaho BI-Plattform zuständig. Der Benutzer kann über JPivot die wichtige OLAP Operationen wie slice und dicing, oder drill down oder roll up durchführen.

**Drill Down to Pivot Table**

		Markets			
Product	Time	+APAC	+EMEA	+Japan	+NA
+Classic Cars	+2003	115.011	691.273	120.696	587.428
	+2004	199.372	1.015.790	42.071	581.043
	+2005	97.574	384.538	18.835	237.791
+Motorcycles	+2003	60.789	141.836	16.485	178.109
	+2004	63.159	204.042	31.959	291.421
	+2005	65.870	161.260	4.176	55.020
+Planes	+2003	42.663	154.519	60.556	90.016
	+2004	57.681	209.128	49.177	202.942
	+2005	11.082	128.008		60.985
+Ships	+2003		172.428	14.156	58.238
	+2004	35.323	186.992	10.453	142.904
	+2005	3.070	67.845	8.407	48.856
+Trains	+2003	1.681	29.538	13.279	28.304
	+2004	8.226	90.973		25.551
	+2005		17.995	3.524	15.398
+Trucks and Buses	+2003	11.298	228.699	44.498	135.936
	+2004	80.634	185.421	13.349	252.572
	+2005	53.735	86.859		61.281
+Vintage Cars	+2003	111.639	263.695	22.888	281.727
	+2004	147.212	504.062	21.470	324.815
	+2005	105.688	83.324	7.979	191.727

Slicer: [(All)-All Customers] [Kennzahl-Sales]

Abbildung 4.8: JPivot

**Aggregation Designer** Der Aggregation Designer (siehe Abbildung 4.9) ist eine Benutzeroberfläche, die OLAP-Analysen erleichtert.

Der Designer erstellt Tabellen, die der Datenbank den Aggregationsprozess ersparen [Anc10a].

Der linke Teil umfasst die Dimensionen, die schon aus dem Cub ausgewählt werden, und die entsprechende Levels. Der rechte Teil zeigt die Performance des Ergebnisses. Nach dem Anlegen der Datenbankverbindungen und des schon vorhandenen Schemas werden die zugehörigen Dimensionen angezeigt. Man kann auch die entsprechenden Levels auswählen. Das aggregierte Schema wird gespeichert und danach kann im BI-Server veröffentlicht werden.

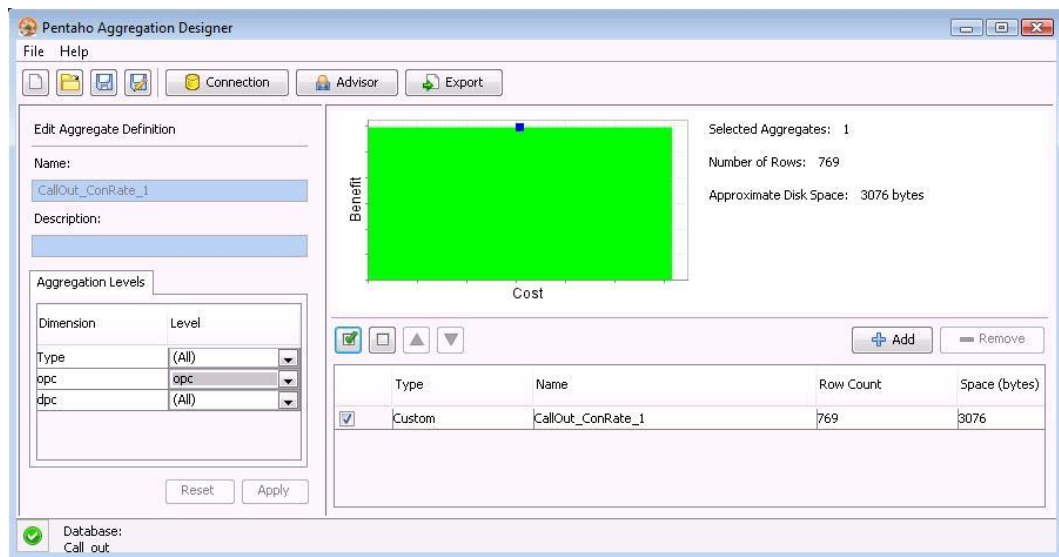


Abbildung 4.9: Aggregation Designer

#### 4.3.4 PentahoETL (Kettle)

Pentaho Data Integration ist ein ETL-Tool von Pentaho für den ETL- (Extract, Transform und Load) Prozess. Man kann die Metadaten mit dem Programm Pentaho Data Integration steuern. Die Daten können bei der Transformation extrahiert, transformiert, transportiert und geladen werden.

Die Software Pentaho Data Integration war ursprünglich unter dem Namen Kettle als eigenständiges Projekt bekannt. Die Software ist unter der Lizenz GNU Lesser General Public License (LGPL) veröffentlicht. Die Produkt Pentaho Data Integration kann auch als eigenständiges Paket losgelöst von der Business Intelligence Kohärenz und unabhängig von den anderen Softwarepaketen von Pentaho und deren BI-Plattform genutzt werden [Fre10].

Mögliche Einsatzgebiete können u.a. sein [Fre10]:

- Das Beladen eines Data-Warehouse.
- Der Ex- und Import von Informationen von Datenbanken, Excel-, Text-, CSV-, oder XML-Dateien.
- Eine Datenmigration zwischen Datenbanken.
- Die Unterstützung eines Backupprozesses für Informationen aus Datenbanken.

Die Software Pentaho Data Integration besteht aus vier Einzelapplikation [Anc10a]:

- **Spoon** ist verantwortlich für die grafische Benutzerschnittstelle.

- **Pan** ist zuständig für die Stapelverarbeitung von Transformation.
- **Kitchen** erledigt die Stapelverarbeitung von Job-Ketten.
- **Carte** hat die Aufgabe, einen RRemote-Server zu aktivieren, mit dem Spoon-Transformationen remote ausgelöst werden können.

Um den ETL-Prozess mit der Software Pentaho Data Integration zu gestalten, muss zunächst eine Datenbankverbindung konfiguriert werden. Die entsprechenden Datenbanktabellen müssen noch für das „Repository“ angelegt werden. Der Benutzer muss sich anmelden, wenn das „Repository“ erfolgreich konfiguriert wurde.

Nach der Anmeldung wird das Hauptfenster Spoon angezeigt. Darin werden die Transformationen und Jobs angelegt. Die linke Sicht des Fensters enthält die „Steps“ (einzelne Bestandteile) des Jobs bzw. der Transformation. Dort sieht man dann alle „Steps“. Der Benutzer kann auf die rechte Seite des Fensters die links ausgewählte „Steps“ einfach per Drag & Drop ziehen und auf der Seite ordnen und sequenzialisieren.

Die „Hops“ verbinden die separaten „Steps“ miteinander und geben die Richtung zwischen den „Steps“. Ein „Hop“ kann per Mausklick der mittleren Maustaste zwischen zwei „Steps“ erstellt und die Richtung zum Zielsteps festgelegt werden.

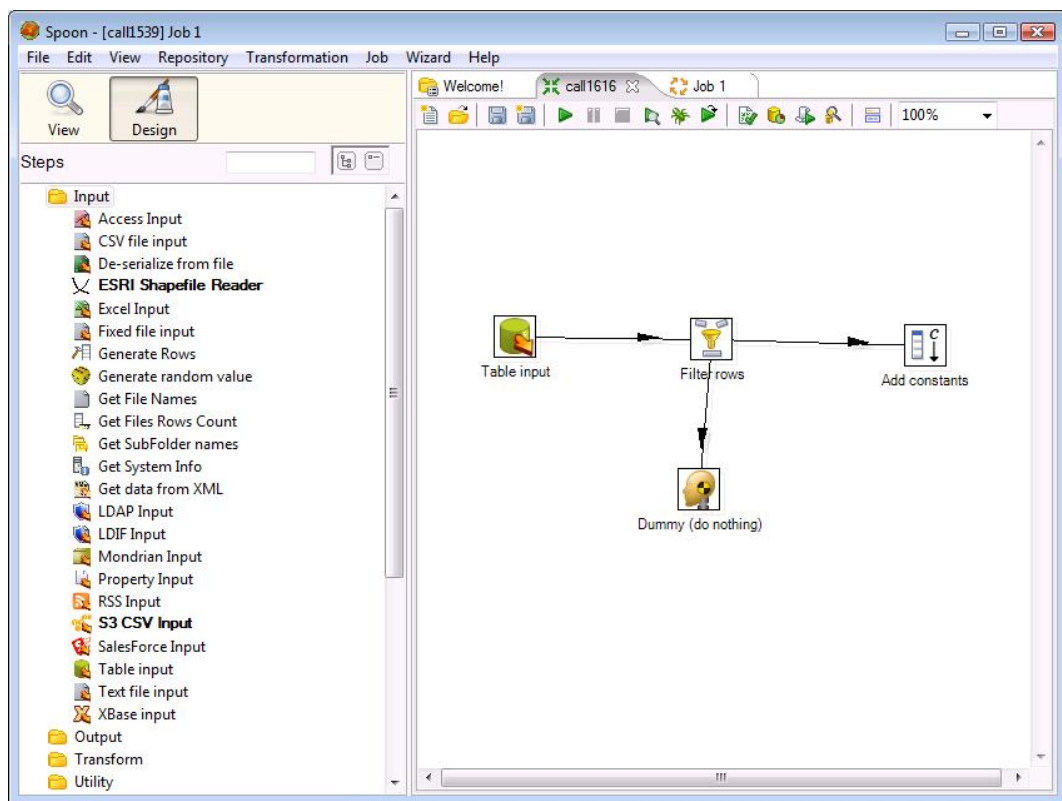


Abbildung 4.10: Pentaho Data Integration-Spoon

### 4.3.5 Pentaho Report Designer

Eine weitere Funktionalität der Pentaho BI Plattform ist das Reporting. Das Reporting wird durch das Modul JFree Report umgesetzt, das im Pentaho BI Server integriert ist. Das Projekt JFree Report als eine freie Bibliothek von Java-Klassen zum Generieren von Berichten ist seit dem 17. Januar 2006 offiziell Teil der Pentaho Corporation und wurde -wie auch Kettle- in das Produktportfolio aufgenommen. Der damalige Chef von JFree Report hat auch die Leistung des Projekts bei Pentaho übernommen. Reports, die mit JFree Report erstellt werden, basieren auf einer XML-Struktur, in der der Aufbau der Objekte, die mit dem Report angezeigt werden, definiert ist. JFree Report unterstützt die folgenden Formate: PDF, HTML, CSV, Excel und Klartext [Fre10].

Der Pentaho Report Designer ist eine grafische Benutzeroberfläche, mit der der Benutzer die Elemente des Reports per Drag & Drop auf den Report bearbeiten kann. Der Report Designer bietet die Möglichkeit, die erstellten Berichte auf dem Pentaho Server zu publizieren.

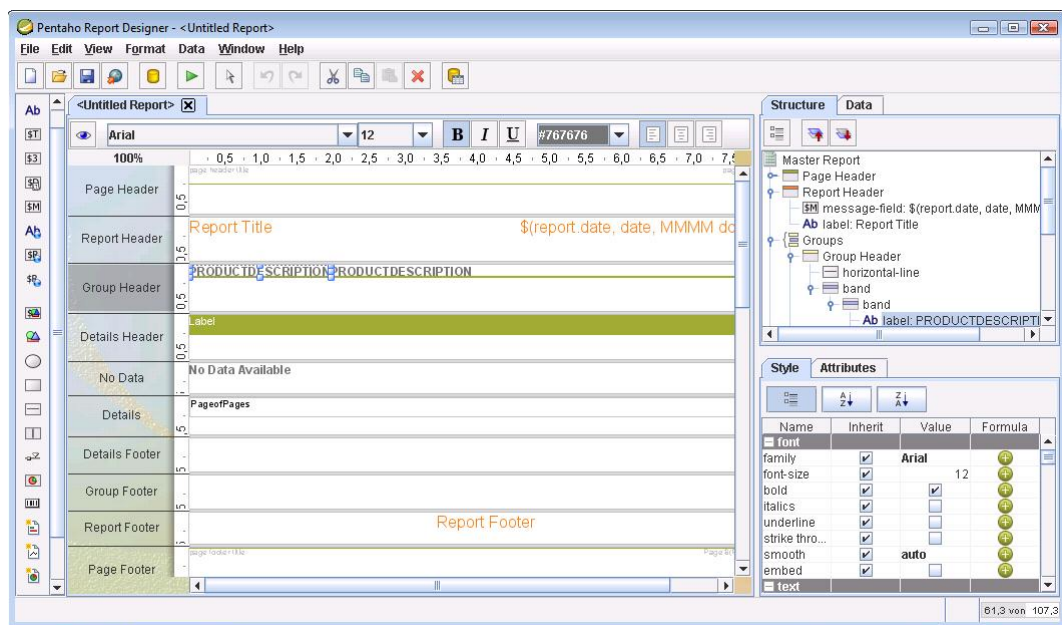


Abbildung 4.11: Pentaho Report Designer

Wie in der Abbildung Pentaho Report Designer zu sehen ist, wird das Hauptfenster des Pentaho Report Designer angezeigt. Auf der linken Seite befindet sich die „Palette“. In der „Palette“ sind alle existierenden Berichtselemente des Report Designers aufgelistet. Zentral in der Mitte des Fensters steht der Bericht, der verarbeitet werden kann. Auf der rechten Seite werden einige Elemente, die auf dem Report definiert werden müssen und in einer hierarchischen Struktur angeordnet werden, dargestellt, z.B. „Page Header“, „Report Header“, „Groups“, „Report Footer“, „Page Footer“. Der Benutzer kann die Eigenschaften des Berichts durch Layout und Design des Berichts verändern und danach den Bericht speichern oder

publizieren.

Die Struktur von Pentaho Report Designer ist unkompliziert. Der Benutzer braucht wenig Zeit, um sich einzuarbeiten. Der Report Designer steht dem Benutzer viele Auswahlmöglichkeiten an qualitativ hochwertigen Diagrammen zur Verfügung, um ansprechende Berichte zu erstellen.

### 4.3.6 Pentaho Enterprise Edition

Die Pentaho Enterprise Edition stellt ein kostenpflichtiges Komplettpaket der Pentaho BI-Suite zur Verfügung. Design-Tools besteht in der Enterprise Edition aus unterschiedlichen Software-Komponenten wie z.B. Aggregation Designer, Data Integration, Design Studio, Metadata Editor, Report Designer und Schema Workbench.

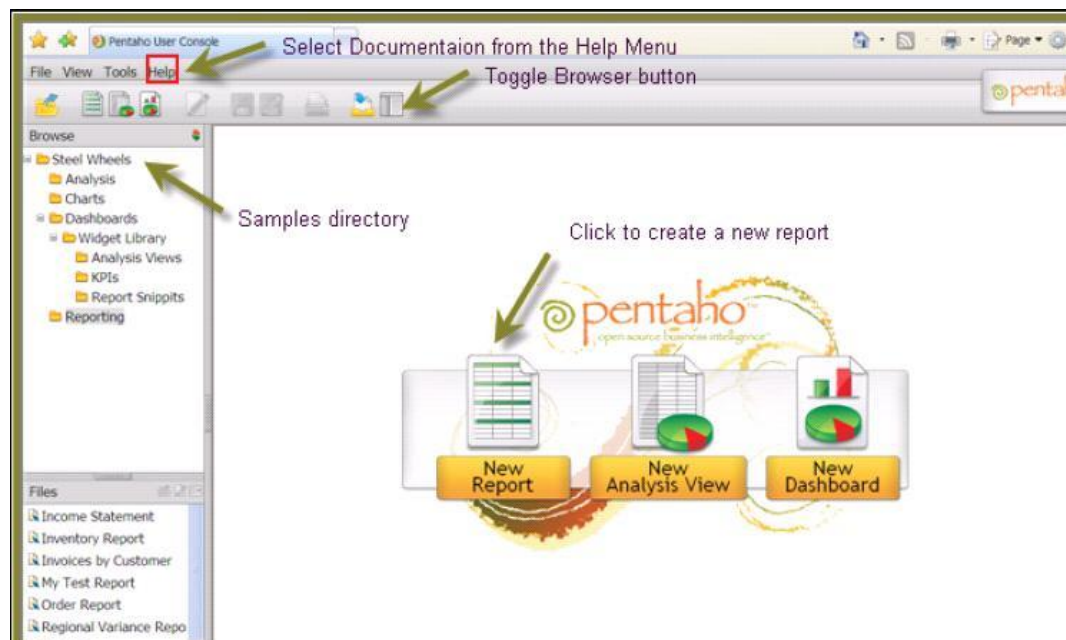


Abbildung 4.12: User Console der Pentaho Enterprise Edition

Die Community Version liefert Design-Tools nicht innerhalb eines Gesamtpaketes, sondern bietet sie einzeln zum Download an. Vergleicht man die einzelnen Komponenten der beiden Versionen, sind kaum Unterschiede zwischen den Community Design-Tools und den Enterprise Edition Design-Tools festzustellen mit Ausnahme des Dashboard Designers. Mit dem Designer kann der Benutzer anhand vorher gespeicherter Berichte Dashboards erstellen. Dafür können Filter gesetzt werden, um Abhängigkeiten zu erzeugen. Grafisch gesehen ist der Dashboard-Designer sehr übersichtlich strukturiert und enthält die wichtigsten Funktionen zur Dashboard-Generierung [Anc10a].

## 4.4 Palo Suite

Die Palo ist ein Open Source Werkzeug der Business Intelligenz. Sie bietet den Benutzern die Möglichkeiten, Berichte zu erstellen und zu analysieren. Die Palo wird in Deutschland von der Freiburger Jedox AG entwickelt. Palo gehört zur Kategorie von MOLAP<sup>1</sup>. Man kann mit Palo die Daten schnell zugreifen und bearbeiten. Jedox bietet die Community Version und die kostenpflichtigen Enterprise Edition an.

Die Palo Suite enthält folgende Komponenten:

- Palo OLAP Server
- Palo Web
- Palo ETL
- Palo for Excel

Mit der Unterstützung vom Palo ETL Server kann man die Daten aus Datenbanken importieren. Jetzt besteht nicht nur die Möglichkeit, Daten in Excel anzuzeigen und weiterzuarbeiten, sondern diese Ergebnisse auch unmittelbar in Palo Web darzustellen und zu analysieren

### 4.4.1 Palo OLAP Server

Der Palo OLAP Server ist MOLAP Server und auch das Herzstück der Palo Suite. Der Vorteil von dem Sever ist eine Zeitersparnis des Datenzugriffs. Man kann auch die Daten in Echtzeit analysieren. Palo OLAP bietet die Möglichkeit der Erweiterung, durch APIs in Java, PHP, C++ und .NET in andere Software-Umgebungen integriert zu werden.

### 4.4.2 Palo Web

Palo Web bietet eine webbasierte Oberfläche (siehe in Abbildung 4.13). Es können über die autorisierte Mitglieder die Berichte geplant, erstellt und analysiert werden. Man braucht auf Benutzer PCs nicht installiert, weil alle Komponenten 100% web fähig sind.

Palo Web enthält folgende Komponenten:

- **Palo Spreadsheet** dient zum Microsoft Excel und vereinfacht die Erstellung von Berichten.
- **Palo Pivot** kombiniert Zeilen-, Spalten-, Seiten- und Datenfeldern, um beliebige Sich-

---

<sup>1</sup>MOLAP: multidimensionales OLAP.



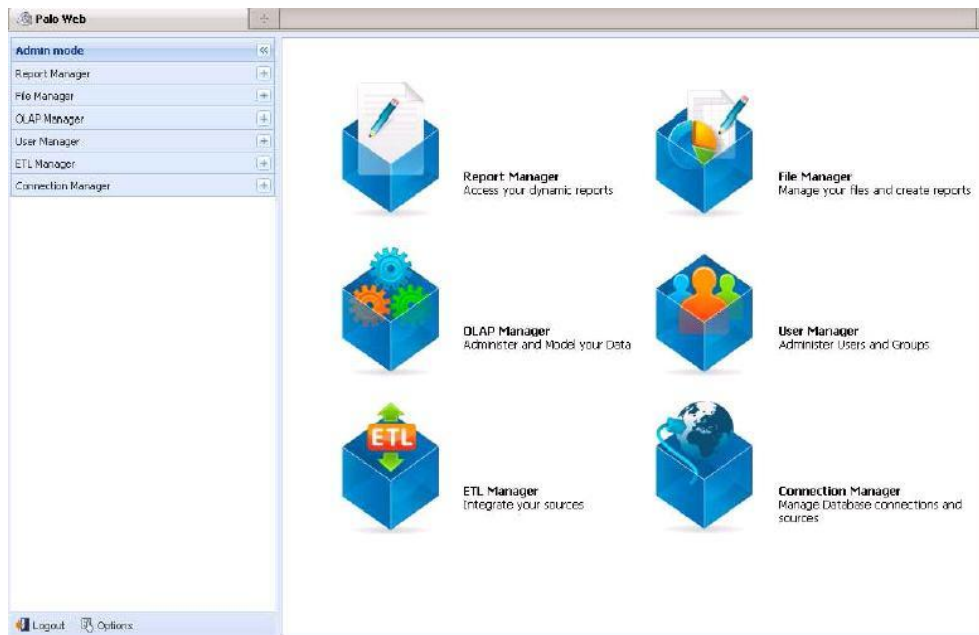


Abbildung 4.13: Palo Web

ten auf die Daten zu erzeugen. Palo Pivot bestimmt als mehrdimensionaler Filter die Werte in der Pivot-Tabelle.

- **Palo User Manager** verwaltet alle Benutzer und Gruppen.
- **Palo ETL Manager** bietet die Möglichkeit, die Daten zu extrahieren, transformieren und laden.
- **Palo File Manager** verwaltet die Dateien, die abgelegt werden.

Es gibt noch zwei Komponenten in der Palo Suite Premium Edition:

- Palo Report Manager stellt die erstellten Berichte und Diagramme dar.
- Palo Modeller ermöglicht den Aufbau und Wartung von Palo OLAP Modellen.

Zur Berichtserstellung können die Daten durch das Anlegen von Arbeitsmappen im Web Browser aufgerufen werden.

Im Vergleich zur Vorgängerversion wurde der Palo Worksheet Server 3.0 auf Basis von C++ komplett überarbeitet und heißt jetzt Palo Web. Wirklich neu an Palo Web sind die sogenannten Dyna Ranges (eine dynamische Formatierungstechnologie), mittels derer sich die Berichtsstruktur automatisch an neue Datenkonstellationen anpasst. Vorbei die Zeiten als Excel-Berichte mühsam angepasst werden mussten, weil eine neue Produktgruppe vom Betrieb eingeführt wurde. Neben den Dyna Ranges hat sich Jedox sehr viel Mühe mit der



Charting-Engine gegeben. Die Diagramm-Galerie in Palo Web ist ziemlich ausgefeilt und sogar einige MicroCharts/sparklines sind integriert [Ope10].

### 4.4.3 Palo ETL

Palo ETL ist Bestandteil der Palo Suite und ermöglicht den Datenaustausch zwischen Quell- und Zielsystemen. Mit Hilfe des Palo ETL Servers können die Daten in oder aus Datenbanken importiert und exportiert werden. Er bietet Standardschnittstellen zu den gängigen Programme und SAP ERP- (SAP nur in der kostenpflichtigen Enterprise Edition) und BW-System.

Der Palo ETL Server wird über einen Webbrowser mit dem ETL-Web-Client angezeigt und entsprechend der hierarchischen Strukturen abgebildet. Der Palo ETL Server bietet die Möglichkeit, die Zeit Hierarchien zu generieren, damit die Einrichtung der Zeit Dimension erheblich erleichtert werden kann.

The screenshot displays the Palo ETL Server interface. On the left, there is a tree view under 'ETL Projekt Explorer' containing 'ImportBiker', 'Variablen', 'Verbindungen', 'Extracts', 'Transforms', 'Loads', and 'Jobs'. The main window shows a table titled 'Überblick der ausgeführten Jobs auf Server "ETL Server"'. Below the table is a detailed log of job execution steps.

Projektname	Job ID	Jobname	Ausführungsdatum	Ausführungsergebnis
ImportBiker	4	Masterdata	Aug 3, 2009 15:55:00	successful
ImportBiker	3	Masterdata	Aug 3, 2009 15:51:16	successful
ImportBiker	2	default	Aug 3, 2009 15:42:28	successful

Log entries (partial):

```

2009-08-03 15:55:00,234 INFO [job Masterdata] (ExecutionStack.java:125) - Starting execution of job Masterdata
2009-08-03 15:55:00,234 INFO [job Masterdata] (ExecutionStack.java:130) - Parameter fallback: true
2009-08-03 15:55:00,234 INFO [job Masterdata] (ExecutionStack.java:130) - Parameter logLink: 100
2009-08-03 15:55:00,234 INFO [job Masterdata] (DimensionLoad.java:69) - Starting load Customers of Dimension Customers
2009-08-03 15:55:00,235 INFO [job Masterdata] (TableSource.java:212) - Data retrieval from extract CustomerRegions
2009-08-03 15:55:00,828 INFO [job Masterdata] (ElementLoad.java:241) - No new elements are loaded
2009-08-03 15:55:02,078 INFO [job Masterdata] (ConsolidationLoad.java:203) - New consolidations loaded: 465
2009-08-03 15:55:02,203 INFO [job Masterdata] (AttributeLoad.java:122) - New Attributes loaded: 932
2009-08-03 15:55:02,203 INFO [job Masterdata] (DimensionLoad.java:93) - Finished load Customers of Dimension Customers
2009-08-03 15:55:02,203 INFO [job Masterdata] (DimensionLoad.java:69) - Starting load ProductsAll of Dimension Products
2009-08-03 15:55:02,218 INFO [job Masterdata] (TableSource.java:212) - Data retrieval from transform Products_Join
2009-08-03 15:55:02,218 INFO [job Masterdata] (TableSource.java:212) - Data retrieval from extract Products
2009-08-03 15:55:02,236 INFO [job Masterdata] (TableSource.java:212) - Data retrieval from extract ProductCategoriesAll
2009-08-03 15:55:02,330 INFO [job Masterdata] (ElementLoad.java:241) - No new elements are loaded
2009-08-03 15:55:02,625 INFO [job Masterdata] (ConsolidationLoad.java:203) - New consolidations loaded: 336
2009-08-03 15:55:02,625 INFO [job Masterdata] (AttributeLoad.java:124) - No new attributes are loaded
2009-08-03 15:55:02,625 INFO [job Masterdata] (DimensionLoad.java:93) - Finished load ProductsAll of Dimension Products
2009-08-03 15:55:02,625 INFO [job Masterdata] (DimensionLoad.java:69) - Starting load OrderLines of Dimension OrderLines
2009-08-03 15:55:02,640 INFO [job Masterdata] (TableSource.java:212) - Data retrieval from extract OrderLines
2009-08-03 15:55:03,015 INFO [job Masterdata] (ElementLoad.java:241) - No new elements are loaded
2009-08-03 15:55:03,312 INFO [job Masterdata] (ConsolidationLoad.java:203) - New consolidations loaded: 969
2009-08-03 15:55:03,328 INFO [job Masterdata] (AttributeLoad.java:124) - No new attributes are loaded
2009-08-03 15:55:03,328 INFO [job Masterdata] (DimensionLoad.java:93) - Finished load OrderLines of Dimension OrderLines
  
```

Abbildung 4.14: Palo ETL

Um den ETL-Prozess durchführen zu können, muss zunächst eine Extraktion auf die Quellsysteme festgelegt werden. Mehrere Transforms können abgebildet werden, wie Normalisierung, Aggregation, Join, Filtern und Sortieren, usw. Mit Hilfe von Loading wird das Zielsystem angegeben.

Jobs gruppieren mehrere Loads und können durch die Verwendung von Variablen dynamisiert werden. Mit Hilfe von Palo ETL ist der Drill-Through von Palo auf die Belegdaten in den Quellsystemen durch Zwischenspeichern der erforderlichen Daten möglich. Damit können Informationen über den Ursprung der Daten auf einem hohen Detaillierungsgrad über Navigation angezeigt werden [Pal10].

#### 4.4.4 Palo for Excel

Palo for Excel besteht aus dem Palo Olap Server und dem Palo-AddIn für Excel bzw. Open-Office Calc [Wem10].

Bisher bestand die Möglichkeit, Daten aus Datenbanken in Excel über Datenverbindungen darzustellen und zu verknüpfen. Dort konnten sie weiter aufbereitet werden. Es ist komplex, um die Daten für Planungs- und Analyse Anwendungen zu bearbeiten. Jetzt bestehen die Verbesserungen durch Palo for Excel: Man kann durch Versionisierung und Benutzermanagement das berühmte Excel-Chaos verhindern.

Somit besteht jetzt nicht nur die Möglichkeit, Daten in Excel anzuzeigen und weiterzuverarbeiten, sondern diese Ergebnisse auch zu organisieren und einfach und transparent strukturiert darzustellen.

Durch diese neuen Module kann eine wesentliche Anforderung der Unternehmen erfüllt werden: Die Daten können weiterhin mit bekannten Werkzeugen (Excel) verarbeitet und aufbereitet werden.

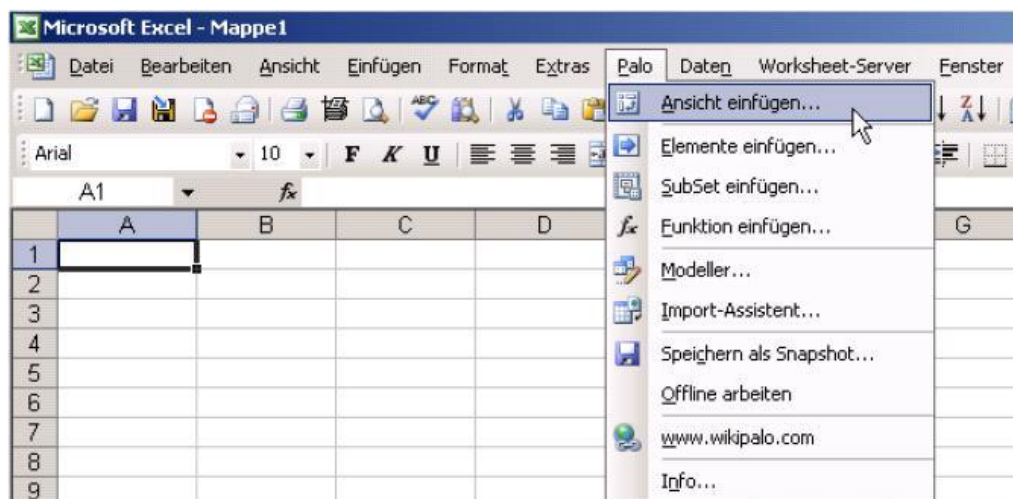


Abbildung 4.15: Palo for Excel

Nach der Installation und dem Neustart von Excel kann man einen Menüpunkt mit dem Titel „Palo“ finden, über den die Funktionalitäten auf Palo OLAP Server in Excel zugegriffen werden können.

Wie oben bereits beschrieben, ist es möglich, eine Verbindung zwischen Excel und den Palo OLAP Server zu erstellen. Über die Option Ansicht einfügen wird eine Palo Datenansicht aufgebaut. Damit können die Daten aus erstellten Würfeln in Excel angezeigt werden.

Um auf die Würfeldaten zugreifen zu können, ist es notwendig, eine Datenbankverbindung

zu erzeugen. Diese Verbindung kann über die Option Modeller verwaltet werden. Es besteht bei der Verwaltung die Möglichkeit, Datenbank anzulegen oder zu löschen.

Nach der Selektion des Palo Würfels können die zugehörigen Dimensionen und Elementen angezeigt werden. Die Dimensionen und Elementen können auch angelegt werden, um Würfel zu strukturieren.

Daneben lassen sich Daten nicht nur analysieren, sondern gegeben falls auch zurückschreiben. Palo eignet sich dadurch auch gut als Planungswerkzeug [Anc10a].

#### **4.4.5 Palo Supervision Server**

Der Palo Supervision Server ist nur in der Palo Enterprise BI Suite enthalten und dient als Kontrollinstanz für die einzelnen Abläufe bezüglich der Palo Datenbank. Das betrifft beispielsweise den Prozess „Login eines Benutzers“, d.h. wann ist der Benutzer eingeloggt, welche Änderungen hat er vorgenommen, wann hat er sich abgemeldet. Das Tool arbeitet mit PHP Codes, die als Warnsysteme fungieren, wenn definierte Grenzwerte über- bzw. unterschritten werden. Der Benutzer kann in solchen Fällen per SMS oder E-Mail kontaktiert werden. Die Anwendungsgebiete für den Einsatz des Palo Supervision Server sind Bereiche, in denen Transparenz eine wichtige Rolle spielt, z.B. im Controlling bei der Analyse bestimmter Kennzahlen wie Gewinn, Umsatz oder Liquidität [Anc10a].

## 5 Implementierung des multidimensionalen Datenmodells bei T-Mobile

Der Hauptaspekt dieses Kapitels ist die Implementation des OLAP-Werkzeugs. Die Generierung der Datenbank und die Visualisierung des OLAP-Werkzeugs JasperServer werden beschrieben. Die Implementierung des Datenmodells werden an einem konkreten Beispiel aufgezeigt.

### 5.1 Aufbau der OLAP-Datenbank

Um das erstellte multidimensionale Datenmodell von dem Unternehmen T-Mobile umzusetzen, muss eine OLAP-Datenbank als Grundlage vorhanden sein.

Das Herzstück der Lösung ist der MySQL Administrator und MySQL Query Browser, die für den Aufbau der Datenbank benötigt werden. Mit diesen Tools können die auf dem Server enthaltenen Datenbanken angezeigt, verwaltet und bearbeitet werden. Diese Tools dienen zur Erstellung von Datenbanken mit Tabellen, Views, Triggern und zur Tabellen-Manipulation.

#### 5.1.1 Erstellung der Datenbank

Die Datenbank wird erstellt, enthält sie nur Tabellen, die für die BI-Lösung erforderlich sind. Die Datenbank wird anhand der Anforderungen konzipiert und mit diesen Tabellen gefüllt. In der Datenbank handelt es sich zum Teil um importierte Daten der Tabelle `callcdr_dex_import_calls`, die in der Datenbank als Quelldaten schon enthalten ist.

Folgende Dimension- und Fakte-Tabellen werden aus der Quelldaten-Tabelle `callcdr_dex_import_calls` künstlich erzeugt und verwendet:

#### **Fakte-Tabelle:**

`conrate_fact`: Aggregation der Dimension-Tabellen, wie Gesamtsummen(Call Setup und Call Connection) und Prozentberechnungen(Rate von Call).

JasperETL steht zur Verfügung, mit dem die benötigten Daten integriert werden können. Als Datenbasis benötigt man eine vorher erstellte Fakte-Tabelle, die alle relevanten Daten enthält.

Zur Erstellung einer Fakte-Tabelle muss als erstes ein Job definieren werden (siehe Abbildung JasperETL).

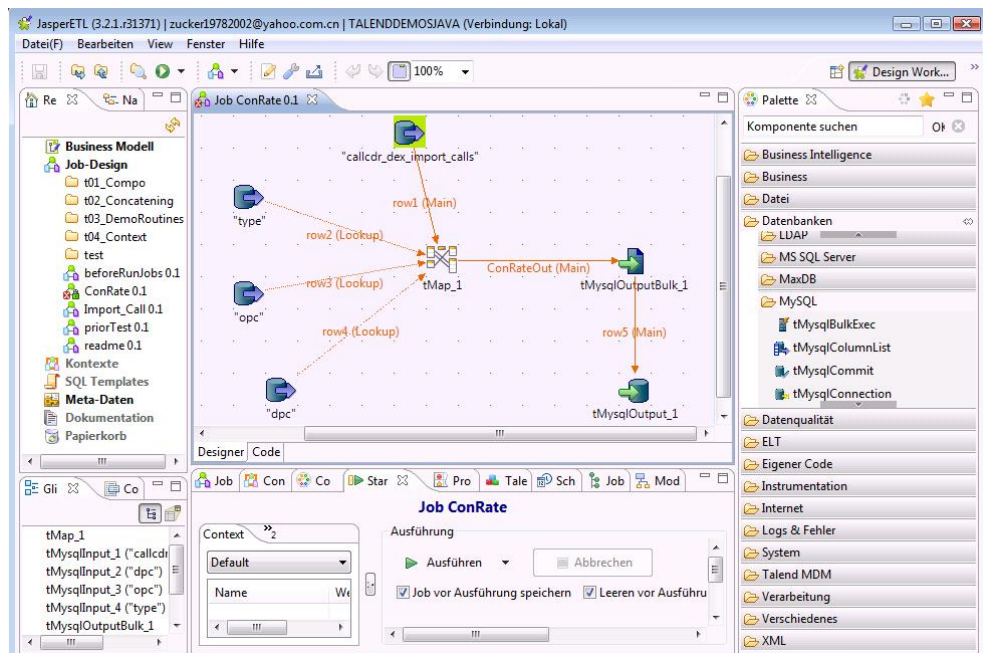


Abbildung 5.1: Jasper ETL

Dort können die benötigten Tabellen aus der vorher angefertigten Datenquellansicht ausgewählt werden. Um eine Fakte-Tabelle werden die Dimension-Tabellen gruppiert. Die Dimension- und Fakte-Tabellen sind über ein Tool (tMap) miteinander verknüpft (siehe Abbildung tMap 5.2).

Das tMap ermöglicht Joins, Spalten- oder Zeilenfilter, Transformationen und multiple Ausgaben. Die Schlüsselteile der Fakte-Tabelle verweisen auf die Primärschlüssel der Dimension-Tabellen. Dadurch können diese verknüpft werden. Die ausgewählten relevanten Daten werden in die Fakte-Tabelle (`conrate_fact`) durch ein Tool (tMysqlOutPut) importiert.

Die Erstellung einer Fakte-Tabelle erfolgt mit Hilfe vom JasperETL.

Es wird eine neue Fakte-Tabelle in der Datenbank festgelegt. Als Datenbasis dient diese Fakte-Tabelle.

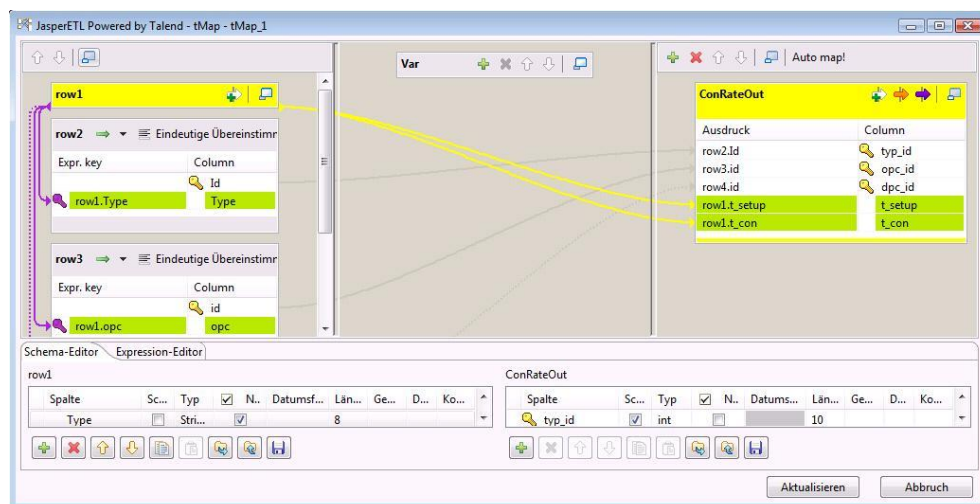


Abbildung 5.2: tMap

### Dimension-Tabelle:

**type:** Diese bildet zwei Type (MOC<sup>1</sup> und MTC<sup>2</sup>) gegenüber dem Call ab.

**opc:** Sie beschreibt OPC<sup>3</sup>.

**dpc:** Sie beschreibt DPC<sup>4</sup>.

### Kennzahl:

**Attmpt:** Bei den Attmpt wird die Gesamtzahl der eingerichteten Calls nach OPC, DPC und Zeit( $t\_setup$ ) ausgewertet.

Die Kennzahl Attmpt ist die Summe aller eingerichteten Calls, die nach Zeit( $t\_setup$ ) zwischen einer bestimmten OPC und einer bestimmten DPC berechnet wird.

**Connection:** Bei den Connection wird die Gesamtzahl der verbundenen Calls nach OPC, DPC und Zeit( $t\_con$ ) ausgewertet.

Die Kennzahl Connection ist die Summe aller verbundenen Calls, die nach Zeit( $t\_con$  nicht Null) zwischen einer bestimmten OPC und einer bestimmten DPC berechnet wird.

**Connection Rate:** Zur Berechnung der Kennzahl Connection Rate werden alle verbundenen Calls ( $t\_con$  nicht Null) aufsummiert und durch die Anzahl der eingerichteten Calls geteilt.

Berechnung der Kennzahl:  $Connection\ Rate = Connection / Attmpt$ .

<sup>1</sup>MOC(Mobile Originating Call): Ein vom Mobilfunkkunden gestartetes Gespräch.

<sup>2</sup>MTC(Mobile Terminating Call): Ein vom Mobilfunkkunden empfangenes Gespräch.

<sup>3</sup>OPC(Originating Point Code): Quelladresse in einem SS7 - Netz.

<sup>4</sup>DPC(Destinating Point Code): Zieladresse in einem SS7 - Netz.

## 5.1.2 Testdaten

Die hier verwendeten Testdaten bilden die Struktur von **Import-calls** ab und enthalten Typ zum Call, Quelladresse-(OPC) und Zieladresse(DPC)-Code, Setupzeit, Verbindungszeit usw.

## 5.2 Framework

In diesem Kapitel wird die Implementierung des OLAP-Werkzeuges JasperAnalysis Schema Workbench beschrieben. Die Darstellung der zugehörigen Hierarchien wird im Schemabrowser und in JasperAnalysis beschrieben.

### 5.2.1 Verbindung mit der Datenbank

JasperAnalysis Schema Workbench wird dazu benötigt, um das Projekt darzustellen. Dort muss als erstes eine Datenbankverbindung eingerichtet werden. Mit Hilfe dieser Datenbankverbindung können dann verschiedene Datenquellansichten mit den benötigten Tabellen der Datenbank erstellt werden. Diese Views dienen als Basis für die Erstellung des Würfels und ermöglichen unterschiedliche Sichten auf die gleiche Datenquelle.

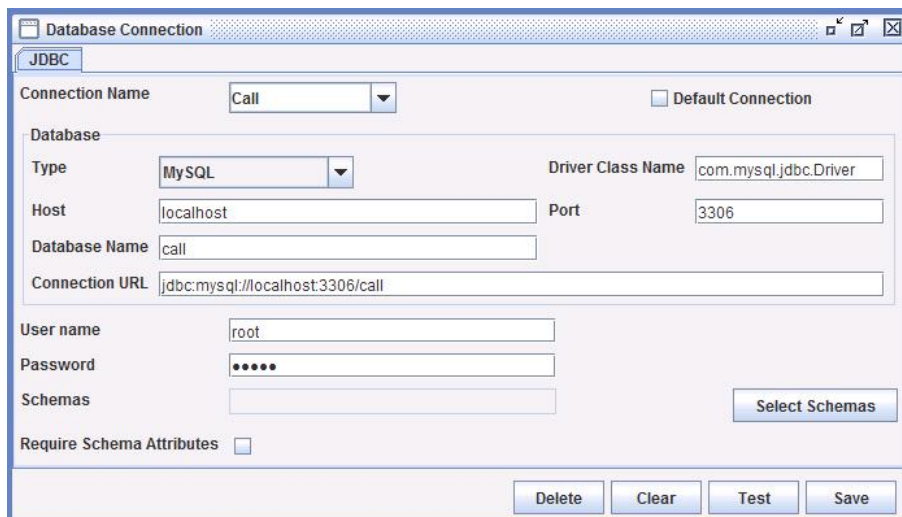


Abbildung 5.3: Datenbank Verbindung

In JasperAnalysis Schema Workbench steht ein vorgefertigtes Tool zur Verfügung, mit dem eine Verbindung zur Datenbank eingerichtet werden können.

Nach Erstellung dieser Verbindung kann Schema Explorer angezeigt werden.



## 5.2.2 Schemabrowser JasperAnalysis Schema Workbench

Ein OLAP-Schema in JasperAnalysis ist als die XML-Datei gespeichert, die eine hierarchische Struktur hat. Das Schema umfasst Komponenten wie Würfeln, zugehörigen Dimensionen, Hierarchien, Levels oder Kennzahlen. Das Schema kann in einer Baumstruktur angezeigt werden. Man kann die MDX-Abfragen in einem Editor ausführen.

Der im Projekt umgesetzte Schemabrowser (siehe Abbildung Schemabrowser 5.4) ist wie folgt aufgebaut: ein Schema wird zuerst definiert, und dann die aus der OLAP-Datenbank an die Applikation eingebundenen Würfeln sind die Ordner der obersten Stufe. Sie werden durch ein Würfelsymbol gekennzeichnet.

So kann der Benutzer des Schemabrowsers direkt in die Dimension (beispielsweise opc, dpc,...) navigieren, indem er die Hierarchien, abgebildet über die Metapher Symbol, auf- oder zuklappt. Die kompaktere Darstellung visualisiert eindeutig die Hierarchiestufen.

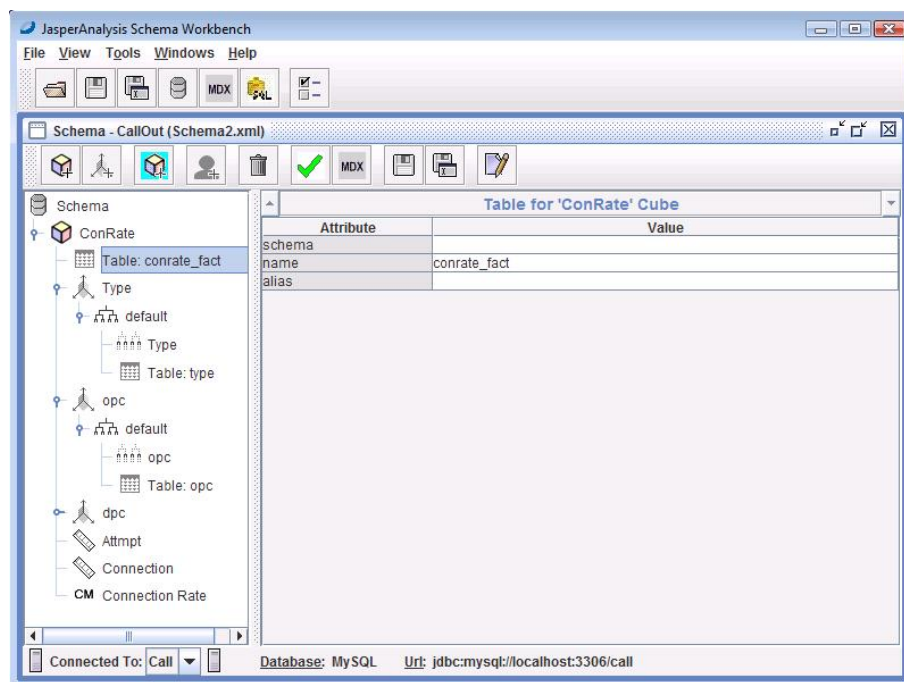


Abbildung 5.4: Schemabrowser

Die Dimensionen werden entsprechend ihrer hierarchischen Strukturen abgebildet. Dies sorgt für die Verschachtelung. Der Startknoten einer Dimension trägt die Bezeichnung der Dimension, bspw. opc oder dpc. Die granularsten Elemente der Dimensionshierarchie sind die Blattknoten, die nicht weiter zu expandieren sind. Sie besitzen zur besseren optischen Unterscheidung ein unterschiedliches Symbol. Alle Symbole zwischen Startknoten und Blattknoten sind die Stufen der Aggregationshierarchie. Die Dimension basiert auf einer Dimension-Tabelle (hier zum Beispiel: opc, dpc,...).



Zum Definieren von einer Dimension muss als erstes ein Dimensionname und Fremdschlüssel (Englisch: foreignkey) zugehöriger Dimension-Tabelle in der Datenbank vergeben werden. Dann müssen die entsprechenden Felder ausgefüllt werden. Zum Definieren von anderen Elementen (zum Beispiel: Hierarchie, Kennzahlen usw.) ist es der Definition einer Dimension ähnlich.

Der Würfel ist jeweils mit drei Farben (gelb, blau und rot) markiert, umgesetzt über einen der Dimensionsbeschriftung vorangestellten farbigen Quader. Der Würfel basiert auf einer Fakten - Tabelle (hier zum Beispiel: `conrate_fact`). Dimensionen mehrerer Würfel werden mit der linearen Markierung auf dem Dimensionssymbol versehen (siehe in Abbildung Schemabrowser: besitzt eine lineare Markierung).

Im Schemabrowser des JasperAnalysis Schema Workbenchs werden nicht nur Dimensionen, sondern auch die Kennzahlen (Englisch: measure) und Aggregationsfunktionen angezeigt. Man kann durch eine SQL - Abfrage volle Dimensionen aus diesen Dimensionstabellen verbinden.

Um die Elemente hinzuzufügen oder zu löschen werden vom Benutzer ausgewählte Funktionen durchgeführt und ausgegraut beim rechts Klicken dargestellt.

Der wesentlich kompakte Schemabrowser bietet als Navigationswerkzeug erweiterte Manipulationsmöglichkeiten.

### 5.2.3 Benutzeroberfläche JasperAnalysis Schema Workbench

JasperAnalysis Schema Workbench besitzt zwei Hauptkomponenten auf der Benutzeroberfläche (siehe Abbildung 5.5 Benutzeroberfläche): eine horizontale Menüleiste, darunter steht ein Funktionsbereich.

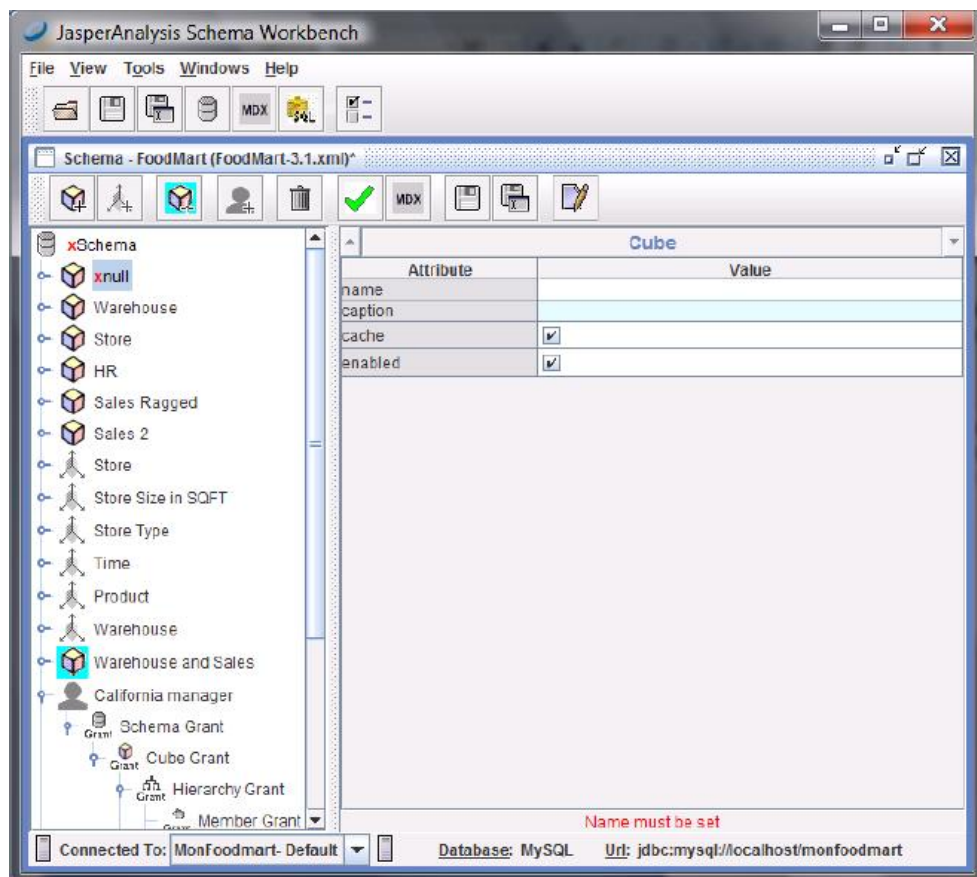


Abbildung 5.5: Benutzeroberfläche

Es bietet die folgenden Funktionen auf dem Funktionsbereich:

- **Schema Explorer:** Ein Editor integriert mit der zugrunde liegenden Datenquelle für die Validierung. In dem Schema Explorer gibt es einen Navigationsbereich (siehe Absatz 5.2.2) auf der linken Seite und Visualisierungsbereich auf der rechten Seite.
- **MDX Query:** Test durch MDX-Abfragen für das Schema in der Datenbank (siehe Abbildung 5.6 MDX Query).
- **JDBC Explorer:** das Betrachten zugrunde liegenden datenbankbasierten Struktur.

Die Benutzeroberfläche umfasst verschiedene Manipulationsmöglichkeiten, die dem Benutzer erlauben, Lösungen zu seiner Fragestellung zu finden. Die Anfragen erfolgen manipulierbar, ein reiner Sprachmodus zur Exploration der Daten ist durch MDX-Abfrage vorgesehen.

Die Buttons der Menüleiste dienen der allgemeinen Funktionsauswahl. Dazu gehört eine Datenbankfunktion, die eine Verbindung mit einer Datenbank erstellt. Über die Menüleiste erfolgt die Wahl der Abfragesprache (MDX oder SQL). Des Weiteren können die Schemen als XML-Dateien geöffnet oder gespeichert werden.

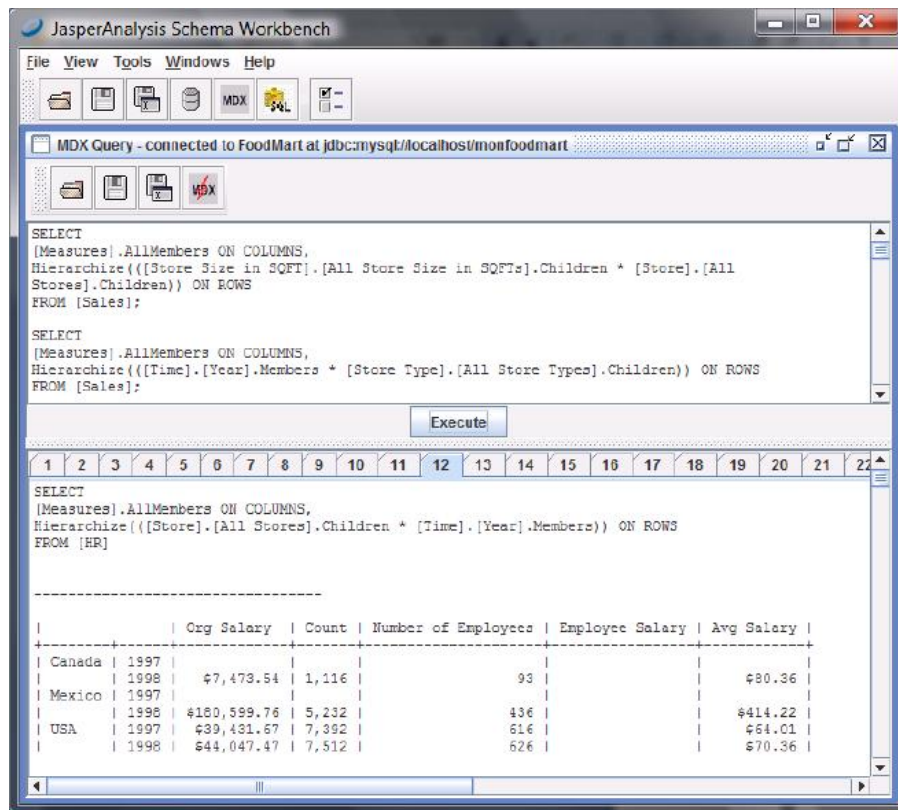


Abbildung 5.6: MDXQuery

## 5.2.4 Integration mit JasperAnalysis User's Guide

Um mit dem JasperServer Analyse erstellen zu können, muss zuerst Tomcat Server gestartet und sich unter <http://hostname:8080/jasperserver-pro> eingeloggt werden (siehe Abbildung 5.7 JasperAnalysis User's Guide Login).

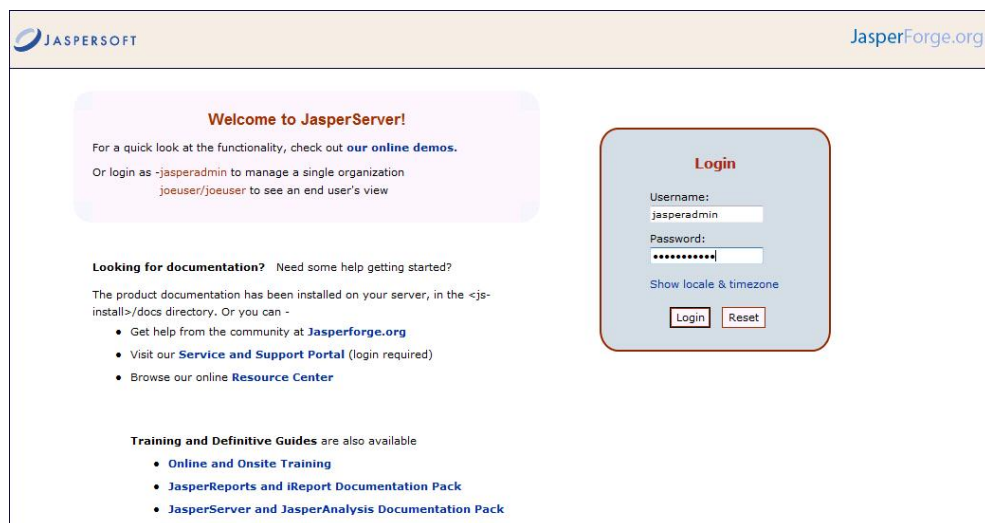


Abbildung 5.7: JasperAnalysis User Guide Login

Die Analysen können im Repository Page des JasperServers erstellt werden (siehe Abbildung 5.8 Repository Page).

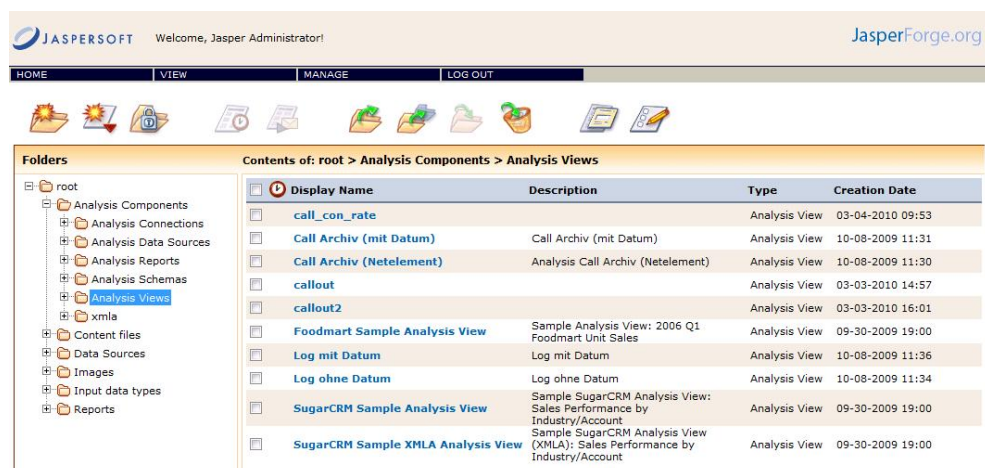


Abbildung 5.8: Repository Page

Die Repository Page bietet die Möglichkeit, Analysis Data Source, Analysis Schema, Analysis Connection, Analysis Views und Analysis Report zu erstellen. Sie sind für Anzeigen von Analyse notwendig.

Um Analyse erstellen zu können, muss zunächst eine neue Analysis Date Source, das schon im JasperAnalysis Schema Workbench erledigt wird, erstellt werden. Und dann Analysis Schema und Analysis Connection werden auch erstellt, letztlich wird Analysis Views erstellt.

## 5.2.5 Visualisierung mit Analysis View

Das Analysis View basiert auf einer Verbindung und einer MDX-Abfrage. Eine Verbindung besteht aus einem Data Source und einem Schema.

Dieser Bericht ist für die Rate der Connection-Calls erforderlich, da viele Detailinformationen enthalten sind.

Der Report `call_con_rate` zeigt alle Calls an, bei denen der Anteil der Connection-Calls über der Zahl der eingerichteten Calls liegt. Dazu werden alle eingerichteten und verbindenden Zahlen der Calls zwischen einer bestimmten OPC und einer bestimmten DPC berechnet und nach Type (MOC und MTC) aggregiert.

Dimensions						Measures		
Type	Type	opc	opc	dpc	dpc	Attmpt	Connection	Connection Rate
[-] Alle Type		[+] Alle opc		[+] Alle dpc		89.879	66.524	0,74
Alle Type	AIF/MOC	[+] Alle opc		[+] Alle dpc		50.744	39.998	0,788
	AIF/MTC	[+] Alle opc		[+] Alle dpc		39.135	26.526	0,678
		Alle opc	1408	[+] Alle dpc		30	30	1
				Alle dpc	1433	2	2	1
					1440	1	1	1
					1456	3	3	1
					1465	1	1	1
					1480	1	1	1
					1490	11	11	1
					1491	11	11	1
				1424	[+] Alle dpc	39	28	0,718
				1433	[+] Alle dpc	46	30	0,652
				1440	[+] Alle dpc	18	12	0,667
				1448	[+] Alle dpc	35	23	0,657
				1456	[+] Alle dpc	42	23	0,548
				1465	[+] Alle dpc	59	41	0,695
				1480	[+] Alle dpc	18	12	0,667

Abbildung 5.9: Analysis View

In der Startansicht wird der Bericht zusammengeklappt auf die Callkategorie angezeigt. Durch Drilldown können die zur Callkategorie gehörenden Calls mit ihren Type, OPC, DPC, eingerichteten Attmpt-, verbindenden Connection- und Connection Ratezahlen eingesehen werden. Durch das Aufklappen des +-Symbols wird die untergeordneten Ebenen angezeigt.

In diesem Report ist der **Drillthrough**-Bericht enthalten, der die Details mit der Unix-Time anzeigt.

Durch die Bindung der Ansicht des Drillthrough-Bericht an das Dimensions-Feld wird beim Aufrufen der Drillthrough-Bericht nicht angezeigt, sondern erst durch das Drücken der Zahl auf dem Measures-Feld zusätzlich zum eigentlichen Bericht dargestellt.

Der Drillthrough-Bericht auf den Typ (AIF/MTC) und OPC (1408) sieht wie folgt aus:



## call\_con\_rate




Drill Through Table for [t_setup = 30]			
Type	opc	dpc	t_setup
AIF/MTC	1408	1433	1.260.795.422,27
AIF/MTC	1408	1433	1.260.795.353,86
AIF/MTC	1408	1440	1.260.795.432,86
AIF/MTC	1408	1456	1.260.795.479,10
AIF/MTC	1408	1456	1.260.795.408,04
AIF/MTC	1408	1456	1.260.794.660,62
AIF/MTC	1408	1465	1.260.795.483,47
AIF/MTC	1408	1480	1.260.795.454,04
AIF/MTC	1408	1490	1.260.794.700,67
AIF/MTC	1408	1490	1.260.795.479,79

Page 1/3    Goto Page 1    Rows/page 10

Abbildung 5.10: Analysis View - Drillthrough-Bericht

Im Drillthrough Bericht werden entsprechende Calls mit Type, OPC, DOC und Setup Time (Unix-Time) angezeigt. Durch die im Drillthrough Bericht angezeigten Pfeile und Textfelder für die Seiteinstellung in der letzten Zeile ist eine weitere Bearbeitung durch den Benutzer möglich.

Das Analysis View bietet die Möglichkeit, Berichte nicht nur in Tabellen oder Matrix Form zu erstellen, sondern auch direkt Diagramme zu gestalten. Dazu muss in der Link von der Tool Bar das Diagramm-Symbol  geklickt werden und beim Klicken des Edit Diagramm-Symbols  der gewünschte Diagramm-Typ und die entsprechenden Einstellungen ausgewählt und gemacht werden.

Zur Weiterverteilung der Berichte ist die Funktion Exportieren des Berichts sinnvoll. Wie in Abbildung Übersicht der Diagramme von der Tool Bar sichtbar ist, besteht die Möglichkeit, den angezeigten Bericht mit *Output as Excel*  als XLS und mit *Print as PDF*  als PDF zu speichern. Außerdem kann der Bericht als PDF aus direkt gedruckt werden. Spezielle Output-Optionen, wie z.B. Report Options, Page Layout usw., müssen durch *Edit Output Options*  eingestellt werden.



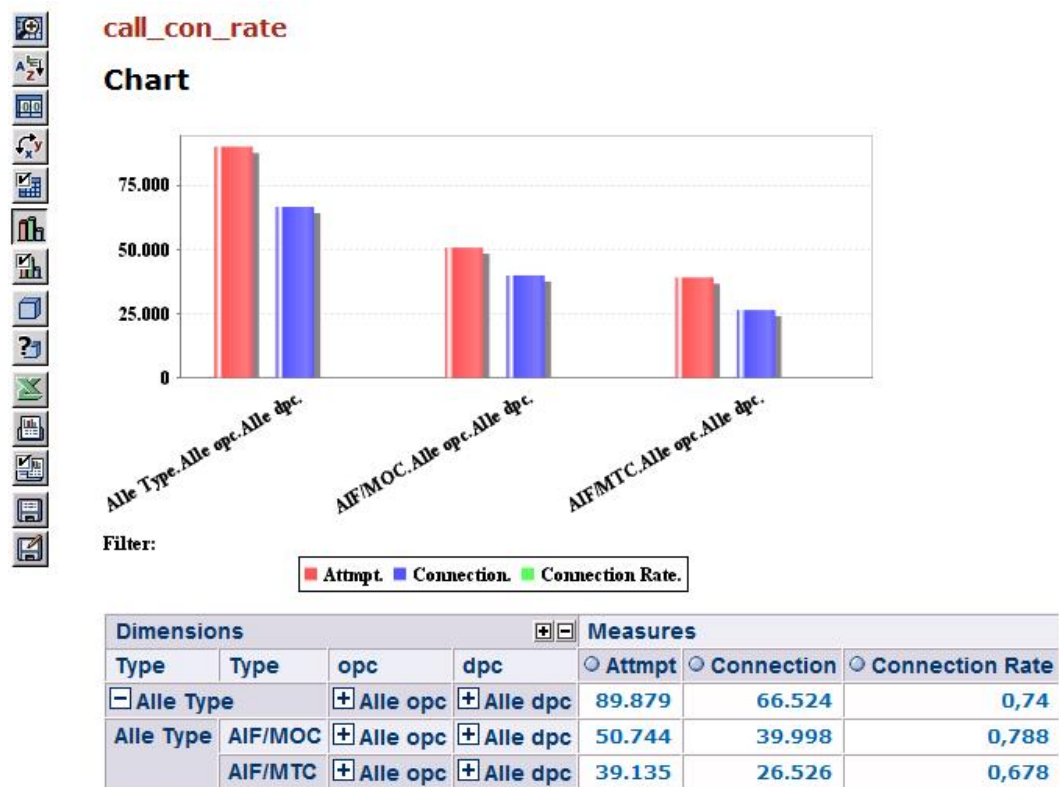


Abbildung 5.11: Übersicht der Diagramme

## 5.2.6 Implementierung

Einige Implementationsdetails leiten über zu einem Schwerpunkt des Kapitels, der Generierung der Datenbankabfragen in **MDX**.

### 5.2.6.1 MDX und OLAP

Multidimensional Expressions (MDX) ist eine Datenbanksprache für OLAP-Datenbanken. Um MDX-Anfragen zu stellen müssen die multidimensionalen Daten in relationaler Form vorliegen. Operationen wie *Slicing*, realisiert mit **where** können dann ausgeführt werden.

Sind beispielsweise die Anzahl des eingerichteten Calls (**Attmpt**) für alle *Type*, *OPC* und *DPC* gesucht, sieht eine typische MDX-Anfrage, auch Select-Anweisung oder „SWF-Block“ (für **Select**, **From**, **Where**) genannt, für die Calls wie folgt aus:

```
Select NON EMPTY {[Measures].[Attmpt]} ON COLUMNS,
NON EMPTY Hierarchize({([Type].[Alle Type], [opc].[Alle opc],
[dpc].[Alle dpc])}) ON ROWS
```

from [ConRate]

Dimensions			Measures
Type	opc	dpc	Attmpt
+ Alle Type	+ Alle opc	+ Alle dpc	89.879

Abbildung 5.12: Ergebnis der MDX-Anfrage

Dabei werden aus in der Regel einem OLAP-Würfel, englisch cube (From) eine Menge von Dimensionen und zu ihnen gehörenden Klassifikationsknoten ausgewählt (Select) und auf verschiedene Achsen der Ergebnistabelle (COLUMNS, ROWS, ...) abgebildet. Mit einem Slice (Where) kann eine Auswahl innerhalb der Fakten getroffen werden. Mit einfachen eckigen Klammern werden Zeichenketten als Namen gekennzeichnet. Geschweifte Klammern dienen der Definition von Mengen [Sou].

### 5.2.6.2 Übersetzung der Benutzeraktionen in MDX

**Select-Klausel** In MDX der Select-Klausel, die bestimmt, welche Achsendimensionen der Quelle auszugeben sind. Den Ergebnisraum werden von den Achsendimensionen beschrieben, wobei jeder Dimension bereits eine Rolle zugewiesen wird (ON COLUMNS, ON ROWS). Weiterhin werden innerhalb der Select-Klausel normalerweise eine Menge von Klassifikationsknoten (members) ausgewählt, wobei diese aus unterschiedlichen Klassifikationsebenen (levels) stammen können [Kru10].

**From-Klausel** Die From-Klausel gibt in MDX die Menge der Datenwürfel an, aus denen der Ergebniswürfel generiert werden soll. Wird mehr als ein Datenwürfel herangezogen, ist zu beachten, dass die Datenwürfel jeweils mindestens eine gemeinsame Dimension aufweisen müssen, damit eine Verbundkompatibilität gewährleistet ist [Kru10].

**Where-Klausel** Die Where-Klausel wird in MDX schließlich dazu benutzt, um den Ergebnisraum des zu Grunde liegenden Datenwürfels einzugrenzen. Dabei können in der Where-Klausel nicht nur *measures* benannt werden, sondern auch *dimensions* bzw. deren *members*. Durch die Definition einer Where-Klausel findet in MDX damit ein echtes Slicing statt [Kru10].



## 6 Bewertung der BI Standardlösungen im Zusammenhang mit T-Mobile

Die BI - Standardlösungen mit JasperSoft, Pentaho und Palo Suite werden auf Basis der Erfahrung bei der technische Umsetzung, die in Kapitel 4 und 5 beschrieben ist, beurteilt. Die Bewertungen beziehen sich auf die Möglichkeiten der Produkte, so dass aufgetretene Probleme und/oder Fehler im endgültigen Release beseitigt sein können.

### 6.1 Bewertung der BI-Funktionalität

Bei der Installation in JasperSoft-BI-Suite stehen nur einzelne Werkzeuge, wie JasperServer, JasperAnalysis, JasperETL und JasperReports (umfasst iReport), zur Verfügung. In Pentaho-BI-Suite besteht die Möglichkeit, die Pentaho-BI-Suite als Gesamtpaket zu konzipieren. Dies ist soweit komfortabel, als in der Pentaho-BI-Suite bei der Anwendung verschiedene Werkzeuge, zur Auswahl stehen. Die einzelnen Komponenten wurden auch separat eingesetzt.

Die Anmeldung der Pentaho ist anders in JasperSoft. Die Anmeldeoption in JasperSoft beschränkt sich auf eine Administratoransicht. Pentaho bietet zwei verschiedene Anmeldeoptionen, auszuwählen. Man kann als User (User Console Login) oder als Administrator (Enterprise Console Login) anmelden. Enterprise Console kann aber nur bei der Anmeldung vom Administrator in der Pentaho verwendet werden. Dies ist sehr sinnvoll, da die Verwaltung vom Administrator dadurch einfach wird. In JasperSoft gibt es keine solche Console.

Im OLAP-Analyse Bereich sind die Funktionen zwischen JasperAnalysis und Pentaho Analysis Services (Mondrian) fast gleich. In Pentaho Mondrian Schema-Workbench stehen nur mehr Faktoren als in JasperAnalysis in der horizontalen Menüliste zur Verfügung. Würfeln, Dimensionen, Hierarchien und Kennzahlen können über dieses analysierte Workbench einfach und übersichtlich dargestellt werden.

Hinzu kommt, dass bei der Anwendung Pentaho Mondrian Schema-Workbench die Bedienungsanleitung nicht alle notwendigen Punkte restlos klärt. Sie ist bei der Analyseanwendung zur Unterstützung hilfreich. Im Gegensatz dazu ist die Bedienungsanleitung in JasperAnaly-

sis ausführlich geschrieben.

Bei der Erstellung der grafischen Prozesse In JasperETL kann die Coden durch einen Codegenerator in Java oder Perl automatisch umgewandelt werden. Diese Umwandlung ist aber sehr komfortabel.

Auffällig entspricht die JasperETL dem operationalen Business Intelligence sehr. Pentaho Data Integration (PentahoETL) ist sich sehr gut in die Pentaho-BI-Suite zu integrieren.

JasperETL liefert eine umfassende Anzahl an Konnektoren, PentahoETL enthält sie weniger. In dem Platten-View der JasperETL werden die Komponenten nach Funktionsbereichen aufgebaut, aber in der PentahoETL nach Anwendungsbereichen. Zu SAP besteht eine eigene Schnittstelle in JasperETL, den Java-Connector SAP JCo. In PentahoETL existiert auch ein SAP-Plugin, ein kostenpflichtiges Plugin ProERPconn.

Der Unterschied in Palo ETL zu den ETL-Werkzeugen der beiden anderen Suiten ist, dass die Datenflussmodellierung erstellt wird. Der ETL-Server bestricht im Vergleich zu PentahoETL oder JasperETL mit einer guten Performance und bietet in der Enterprise Version einen Konnektor zur Integration von SAP-Daten an. Interessant für Perl-Anwender und ein Unterscheidungsmerkmal zu Pentaho und Palo: In JasperETL können Codeanpassungen nicht nur in Java, sondern auch in Perl vorgenommen werden [Anc10a].

Die folgende Tabelle (6.1) stellt die wesentlichsten Unterschiede zwischen JasperReports (iReport) und Pentaho Reporting dar [Fra10a]:

Produkt	Jasperreports (iReport 3.0.1)	Pentaho Reporting Engine 0.8.9.5; Designer 1.7
Gründung	2001	2004
Lizenz	L/GPL	LGPL, Mozilla Public License
Firma	Jaspersoft	Pentaho
Diagramme	JFreechart	JFreechart
Formate	HTML, PDF, CSV, XLS, RTF, SVG, ODT	HTML, PDF, CSV, XLS, RTF
Datenquellen	Eine, JDBC, Tabellen-Model, Javabeans, XML, Hibernate, CSV, OLAP, XMLA	Eine, JDBC, OLAP, XMLA
Betrachter	SWING, SWT, Servlet	SWING, Servlet
IDE	Eclipse, Netbeans	Eclipse, OpenOffice
Funktionen	Unterbericht, Kreuztabelle, Internationalisierung	Unterbericht, automatischer Zeilenumbruch
Dokumentationen	Kostenpflichtige Handbücher, Bücher (Deutsch, Englisch), große Community	Kostenlose Handbücher, Bücher zu OpenOffice Base, kleinere Community
Gesamteindruck	Kleine, gut integrierbare Lösung in Eigenentwicklungen oder in die Jaspersoft-BI-Suite, für versierte Anwender	Verbesserte Integration in die Pentaho-BI-Suite, für Endanwender geeignet

Tabelle 6.1: Bewertung der Reporting Clients

## 6.2 Ergebnisse der Bewertung

JasperETL, PentahoETL und Palo ETL haben die Vorteile [Fra10b]:

- Gute Integrierbarkeit
- Flexible Einsatz-, Anpassbarkeit
- Unabhängigkeit

Sie haben aber auch die Nachteile:

- Mangelnde Parallelisierung
- Keine Restartfähigkeit

In JasperSoft bietet *iReport* mehr Funktionen und Diagramme als der Pentaho Report Designer. Diese Reporting Tools verfügen über eine deutlich große Anzahl an Funktionen. Drill-down und Drillthrough lassen sich einfach und schnell visualisieren. Es ist unkompliziert, dass alle Berichte oder Diagramme als PDF-Dateien oder anderen Formaten (z.B. Excel, HTML, ...) gespeichert werden.

---

Als Berichtsdesign kann *iReport* voll funktional verwendet werden. *iReport* bietet die Möglichkeiten zur dynamischen Anpassung oder komplexeren Darstellung von Berichten. Gegenüber anderen Produkten ist *iReport* das älteste und reifste. Das Erstellen eines XML-Berichts ist einfach und die Laufzeitbibliothek klein.

Ein Kritikpunkt bei dem Publizieren eines Berichts von *iReport* ist die fehlende Aktualisierungsmöglichkeit. In *iReport* besteht keine Möglichkeit, die Berichte direkt auf der JasperSoft BI-Plattform publiziert zu werden. Das Programm muss erneut herunter geladen und installiert werden.

Das besondere an Pentaho Report Designer ist die sehr leichte Bedienung und die damit verbundene Möglichkeit einer schnellen Berichterstellung, so dass dieses Tool insbesondere für Laien interessant ist.

Ein klarer Nachteil von Pentaho ist, dass alle Dokumentationen nur auf Englisch vorhanden sind. Dabei können ganze Kapitel nicht auf Deutsch übersetzt werden.

Die Vorteile von Palo liegen bei der Anpassungsfähigkeit einzelner Zellen, der Feststellung von Planzahlen und der leichten Bedienung.

In der Enterprise Edition können Rollen für die angelegten Benutzer hinterlegt werden. Dadurch ist es schnell und einfach möglich, Benutzerrechte auf Würfel oder Datenbanken zu vergeben.

Das Excel Add-in bietet eine bequeme Möglichkeit, einfach über Excel auf die Daten zugreifen zu können. Dadurch wird die Integration der Daten deutlich erleichtert. Diese Anwendung ist Benutzerfreundlich.

Der ETL-Server besticht im Vergleich zu PentahoETL oder JasperETL mit einer guten Performance und bietet in der Enterprise Version einen Konnektor zur Integration von SAP-Daten an [Anc10a].

### **Einsatzgebiete**

Pentaho Reporting ist eine Weiterentwicklung von „JFreeReport“ und bietet Entwicklern zwei Programme zur Definition von Berichten: den „Report Designer“ und den „Report Wizard“. Mit Letzteren kann selbst ein Endanwender einfach in sieben Schritten einen Ad-hoc-Bericht erstellen. Diese einfache Bedienbarkeit macht sich auch die MS-Access-Alternative „Open Office Base“ zunutze, die Pentaho Reporting als Berichtskomponente verwendet. Alternativ zum Report Wizard kann ein Entwickler seine Berichte auch manuell mit dem Report Designer erstellen und über einen Server ausführen und verteilen lassen.

Für BI-Lösungen, die höhere Anforderungen bezüglich eines einheitlichen Aussehens, der dynamischen Anpassung oder komplexeren Darstellung von Berichten stellen, bietet sich Jasper-

Report an. Es ist das älteste und damit reifste der hier betrachteten Open-Source-Produkte. Dank der mit ihm erstellten einfach strukturierten XML-Berichtsdefinitionen, seiner Programmierschnittstelle und der sehr kleinen Laufzeitbibliothek lässt sich das Tool bequem in bestehende Anwendungen einsetzen. Für die Berichtserstellung bietet das kostenlose Designerwerkzeug „iReport“ gegenüber dem kostenpflichtigen Produkt „JasperAssistant“ den Vorteil, dass es wie die Reporting-Engine vom Anbieter selbst entwickelt wurde und damit eine gute Integration verspricht.

Die Verwendung des grafischen Designers iReport ist auch für Endanwender mit SQL-Kenntnissen geeignet. Die einfachste Oberfläche enthält Pentaho Reporting. Selbst ein Fachanwender wird durch einen Wizard in wenigen Schritten zur Erstellung eines einfachen Berichtes geführt. Die Auswahl eines für Endanwender geeigneten Werkzeuges hängt stark von den Anforderungen an seine Bedienbarkeit und die Anpassbarkeit beziehungsweise Komplexität der Berichte ab. Dieses Spektrum wird von allen drei Open-Source-Produkten abgedeckt und geht weit über eine Alternative für Berichte mit Microsoft Excel hinaus.

## 7 Zusammenfassung

Mit der vorliegenden Arbeit wurden neu eingeführte Begriffe übersichtlich und verständlich dargestellt und einen Überblick über die verschiedenen OLAP-Werkzeuge gegeben.

Die Funktionalität der Standardprodukten ist mit den technischen Möglichkeiten der OLAP-Werkzeuge für T-Mobile sicherlich ausreichend. Die BI-Lösungen mit OLAP-Werkzeuge werden komplex erstellt. Die Einarbeitung in die OLAP-Werkzeuge wird dadurch erleichtert, dass es fertige Installations-Packages inklusive Demodaten und Beispielanwendungen zum Download bereitgestellt werden.

Durch den modularen Aufbau aller drei der vorgestellten OLAP-Werkzeuge ist die Plattform bzgl. Individueller Anforderungen anpassungsfähig. Sie sind im OLAP Bereich in der Bedienung ähnlich, d.h. es gibt keinen großen Unterschied im OLAP Bereich. Die mit dem JasperSoft, Pentaho und Palo gebotenen Möglichkeiten werden zur Umsetzung der Anforderungen des T-Mobiles genügen.

Es bietet die Möglichkeit, im BI Bereich die einzelnen Community-Tools einzusetzen. Der wesentliche Vorteil dieser Tools ist, dass sie keine Kosten für die BI-Umsetzung verursachen. Selbst wenn man die kostenpflichtigen Enterprise Versionen auswählen würde, Bestände immer noch ein deutlicher Kostenvorteil. Dies ist ein wesentlicher Punkt, da die Unternehmen bei der Befragung die finanziellen Aspekte der Umsetzung von BI als kritisch ansahen.

Die untersuchten OLAP-Werkzeuge können den Anspruch der Nutzung der Business Intelligence erfüllen. Eine Evaluierung der Möglichkeiten der OLAP-Werkzeuge für die Umsetzung einer BI-Integration kann, hinsichtlich der einfachen Einarbeitung in die Produkte sowie der umfangreichen Integration in Standardprodukten, durchaus sinnvoll sein.

Die größten Unterschiede liegen im Bedienungskomfort der einzelnen Lösungen. So ist Pentaho sicherlich das überzeugendste BI-Komplettpaket. Sollen nur Einzelanwendungen für bestimmte BI-Bereiche eingesetzt werden, stehen JasperSoft und Palo dem bekannteren Konkurrenten aber in nichts nach [Anc10a].

## 8 Rückblick

Ich möchte an dieser Stelle auf die Erstellung der Diplomarbeit zurückblicken, dass es bei Business Intelligence um ein sehr weitgreifendes Thema geht. Ich habe in den vergangenen Monaten viel Neues gelernt und es mir bei der Bearbeitung meiner Diplomarbeit klar geworden ist. Die Herangehensweise an diese Thematik mit Hilfe von Literatur und anschließenden Interviews mit T-Mobile zeitintensiv, aber sehr sinnvoll. Ich habe mich mit einigen Grundlagen, wie beispielsweise mit Business Intelligence oder mit der Anwendung der OLAP-Werkzeuge intensiv beschäftigt. Ich konnte dadurch mein vorhandenes Grundwissen aus dem Studium vertiefen, aber auch neue Sachverhalte verstehen.

Aus der technischen Ebene her war es unabdinglich, sich intensiv mit dem Aufbau und der Struktur der OLAP-Werkzeuge zu befassen. Ohne gute Einarbeitung in die Software, sowie die darin enthaltenen Funktionen, hätten die Anforderungen nicht beispielhaft umgesetzt werden können. Man kann mit Hilfe von Literatur das Wissen über OLAP-Datenbank und OLAP erwerben, es war zur sinnvollen Anwendung der Analyse nötig.

Die Einarbeitung in die Pentaho und in die JasperSoft war etwas erschwert, da zum einen noch wenig englische Literatur in der zur Verfügung stand, und zum anderen noch einige „Bugs“ in der Software enthalten waren.

Insgesamt sind OLAP-Werkzeuge in der Handhabung sehr intuitiv, so dass die Anforderungen und die Benutzerfreundlichkeit erfüllt werden können. Durch Konzeption, Datenarchitektur, technische Architektur und Implementierung konnte ein weit reichender Einblick in die Gestaltung von BI-Lösung gewonnen werden.

Zusammenfassend finde ich, dass dem Business Intelligence und dem OLAP-Werkzeug in Zukunft eine zunehmende Aufmerksamkeit zuteilwerden wird. Obwohl OLAP schon seit ein paar Jahren bekannt ist und verschiedene Produkte mit unterschiedlichen Funktionalitäten auf dem Markt erhältlich sind, ist es noch relativ junger Sektor, in dem enormes Potential für die Zukunft steckt.

# Anhang



## Literaturverzeichnis

- [Anc10a] Ancud IT, [http://www.controllingportal.de/upload/pdf/fachartikel/software/Produktvergleich\\_Open\\_Source\\_BI.pdf](http://www.controllingportal.de/upload/pdf/fachartikel/software/Produktvergleich_Open_Source_BI.pdf). *Ein Vergleich der Open Source BI-Lösungen JasperSoft, Palo und Pentaho*, März 2010.
- [Anc10b] Ancud IT, <http://www.ancud.de/de/loesungen/pentaho/>. *Pentaho - Open Source Business Intelligence*, Januar 2010.
- [Car04] Kemper Hans-Georg; Mehanna Walid; Unger Carsten. *Business Intelligence Grundlagen und praktische Anwendungen*. Vieweg, 2004.
- [Fra10a] Pientka Frank. *Berichtssoftware - warum nicht open source*. 2010.
- [Fra10b] Pientka Frank. *Datenschaufler für java*. 2010.
- [Fre10] Norman Volkert (B.Sc.); Prof. Dr. Klaus Freyburger. *Open source business intelligence zur unterstützung eines software-validierungsprozesses*. 2010.
- [Gla97] W. H. Inmon ; J. D. Welch; K. L. Glassey. *Managing the Data Warehouse*. New York u. a. O., 1997.
- [Got00] Vossen Gottfried. *Datenmodelle, Datenbanksprachen und Datenbankmanagement-System*. Oldenbourg R. Verlag GmbH, München, 2000.
- [(Hr00] Mucksch Harry; Behme Wolfgang (Hrsg). *Das Data Warehouse-Konzept Architektur - Datenmodelle - Anwendungen Mit Erfahrungsberichten*. GABLER, 2000.
- [JAS10] *TeXnicCenter*. <http://www.jaspersoft.com/de/jasperserver>, Januar 2010. Homepage JasperServer.
- [Kar00] Oehler Karsten. *OLAP Grundlagen, Modellierung und betriebswirtschaftliche Lösungen*. Carl Hanser Verlag München Wien, 2000.
- [Ker07] Pietzko Kerstin. *Business Intelligence Grundlage, Konzepte, Umsetzung*. VDM Verlag Dr. Müller, 2007.
- [Kru10] Prof. Dr. Klaus Kruczynski. *Einführung in mdx*. 2010.

- 
- [Kur99] Andreas Kurz. *Data Warehousing Enabling Technologies*. Verlag MITP, 1999.
- [Mic06] Microsoft. *Olap - einführung und Überblick*. *Microsoft Press Deutschland*, 2006.
- [NET10] *JasperServer - Aufbau*. [http://www.netways.de/de/produkte/jasper/bi\\_suite/jasperserver/](http://www.netways.de/de/produkte/jasper/bi_suite/jasperserver/), Januar 2010.
- [Nil98] Clausen Nils. *OLAP Multidimensionale Datenbanken Produkte, Markt, Funktionsweise und Implementierung*. ADDISON-WESLEY, 1998.
- [Ope10] OpenBI, <http://openbi.info/2009/12/27/openbi-palo-bi-suite-3-0>. *Palo BI-Suite 3.0 - the Good, the Bad and the Ugly*, April 2010.
- [Pal10] Palo, <http://www.jedox.com/de/produkte/Palo-Suite/palo-etl-server.html>. *Palo ETL*, April 2010.
- [Pen10] Pentaho, <http://www.i-bi.de/extdoc/2007-10-OSBI.pdf>. *Opensource Business Intelligence zur Unterstützung eines Software-Validierungsprozesses*, Januar 2010.
- [Sch06] Holger Schrödl. *Business Intelligence mit Microsoft SQL Server 2005*. Hanser, 2006.
- [Sou] Sourceforge, <http://de.wikipedia.org/wiki/>. *Wikipedia*. Projektseite der Software TeXnicCenter, einer Entwicklungsumgebung für LaTeXdokumente unter Microsoft Windows.
- [Wem10] Raue Kristian; Krämer Matthias; Kriehn Wemer. *Erste schritte mit palo for excel commercial open source business intelligenz für performance management*. 2010.
- [Wol98] Martin (Hrsg.) Wolfgang. *Data Warehousing - Data Mining - OLAP*. International Thomson Publishing GmbH, Bonn, 1998.

# Abbildungsverzeichnis

2.1	Die Business Intelligence-Plattform . . . . .	5
2.2	Normierung der Daten für die spätere betriebswirtschaftlich einheitliche Darstellung . . . . .	10
3.1	Beispiel Datenwürfel . . . . .	15
3.2	Dimensionshierarchie . . . . .	17
3.3	Projektionen im dreidimensionalen Datenwürfel . . . . .	18
3.4	Call-Connection-Statistik in Star-Schema . . . . .	21
3.5	Snowflake-Schema . . . . .	22
3.6	Roll-Up & Drill-Down . . . . .	24
3.7	Slice & Dice . . . . .	24
4.1	Historische Entwicklung im BI Bereich [Pen10] . . . . .	28
4.2	Übersicht JasperServer . . . . .	29
4.3	Übersicht JasperETL . . . . .	30
4.4	iReport . . . . .	31
4.5	Aufbau der Pentaho BI Suite [Pen10] . . . . .	32
4.6	Architektur des Pentaho Servers [Pen10] . . . . .	34
4.7	Mondrian Schema-Workbench . . . . .	35
4.8	JPivot . . . . .	37
4.9	Aggregation Designer . . . . .	38
4.10	Pentaho Data Integration-Spoon . . . . .	39
4.11	Pentaho Report Designer . . . . .	40
4.12	User Console der Pentaho Enterprise Edition . . . . .	41
4.13	Palo Web . . . . .	43
4.14	Palo ETL . . . . .	44
4.15	Palo for Excel . . . . .	45
5.1	Jasper ETL . . . . .	48
5.2	tMap . . . . .	49
5.3	Datenbank Verbindung . . . . .	50
5.4	Schemabrowser . . . . .	51
5.5	Benutzeroberfläche . . . . .	53

---

5.6	MDXQuery . . . . .	54
5.7	JasperAnalysis User Guide Login . . . . .	55
5.8	Repository Page . . . . .	55
5.9	Analysis View . . . . .	56
5.10	Analysis View - Drillthrough-Bericht . . . . .	57
5.11	Übersicht der Diagramme . . . . .	58
5.12	Ergebnis der MDX-Anfrage . . . . .	59

# Tabellenverzeichnis

3.1	Ausschnitt der Tabelle mit Connection . . . . .	14
3.2	Unterschiede zwischen OLTP und OLAP . . . . .	20
3.3	Vergleich der OLAP Architektur . . . . .	26
6.1	Bewertung der Reporting Clients . . . . .	62

## Abkürzungsverzeichnis

AJAX .....	<b>A</b> synchronous <b>J</b> avaScript and <b>X</b> ML
API .....	<b>A</b> pplication <b>P</b> rogramming <b>I</b> nterface
BI .....	<b>B</b> usiness <b>I</b> ntelligence
DOLAP .....	<b>D</b> esktop <b>O</b> nline <b>A</b> nalytical <b>P</b> rocessing
DPC .....	<b>D</b> estinating <b>P</b> oint <b>C</b> ode
DWH .....	<b>D</b> ata <b>W</b> arehouse
ETL .....	<b>E</b> xtraktion <b>T</b> ransformation <b>L</b> aden
HOLAP .....	<b>H</b> ybrides <b>O</b> nline <b>A</b> nalytical <b>P</b> rocessing
JRE .....	<b>J</b> ava <b>R</b> untime <b>E</b> nvironment
MDX .....	<b>M</b> ultidimensional <b>E</b> xpressions
MOC .....	<b>M</b> obile <b>O</b> riginatin <b>C</b> all
MOLAP .....	<b>M</b> ultidimensionales <b>O</b> nline <b>A</b> nalytical <b>P</b> rocessing
MTC .....	<b>M</b> obile <b>T</b> erminating <b>C</b> all
OLAP .....	<b>O</b> nline <b>A</b> nalytical <b>P</b> rocessing
OLTP .....	<b>O</b> nline <b>T</b> ransaction <b>P</b> rocessing
OPC .....	<b>O</b> riginating <b>P</b> oint <b>C</b> ode
ROLAP .....	<b>R</b> elationales <b>O</b> nline <b>A</b> nalytical <b>P</b> rocessing
SQL .....	<b>S</b> tructured <b>Q</b> uery <b>L</b> anguage
XML .....	<b>X</b> xtensible <b>M</b> arkup <b>L</b> anguage

# Glossar

## **Ad-hoc-Abfragen**

Ad-hoc Abfragen sind ein flexibles Mittel, um schnell innerhalb von Transact-SQL auf Daten einer fremden Datenquelle zuzugreifen. 28

## **Aggregation**

Um eine konsistente Datenbasis zu erhalten, werden bei der Aggregation die in den Dimensionen festgelegten Berechnungsregeln ausgeführt. Dies kann z.B. Summierung der alle Produkte zu Produktgruppen, aller Produktgruppen zu Obergruppen, Obergruppen zu Gesamtbestand sein. 9

## **Business Intelligence**

Werkzeuge der aktuellen Generation von entscheidungs-unterstützten Informationssystemen, die Mitarbeiter (nicht nur Management) in Unternehmen durch Berichtswesen, Analysefunktionen und Wissensfindung (Data Mining) Entscheidungshilfe geben. 4

## **Dimension**

Eine Dimension ist die strukturierte und meistens hierarchische Zuordnung einzelner Ausprägungen zueinander. 16

## **Drillthrough**

Diese Funktionen ermöglichen die auswahlbasierte Verbindung zwischen der OLAP-Datenbank und den relationalen Datenbeständen. Beispiel: Abfrage der zu einem OLAP-Bericht gehörenden Transaktionsdaten aus dem RDBMS mit Hilfe des automatisch erzeugten SQL-Befehls. 54

## **Dritte Normalform**

Die 3. Normalform ist erfüllt, wenn die 2. Normalform erfüllt ist und die Nicht-Schlüssel-Attribute funktional unabhängig voneinander sind.

Sind A und B Attribute eines Relationstyps, so ist B funktional abhängig von A, wenn für jedes Vorkommen ein und desselben Wertes von A immer derselbe Wert von B auftreten muß. 22

## **Ebene**

Ebenen beziehen sich auf die Positionierung eines Elements innerhalb der Hierarchie

der Dimension. 17

### **Granularitäten**

Granularitäten ist der Verdichtungsgrad und beschreibt die ‘Feinheit’, in der die Fakten über die Kennzahlen beschrieben werden. 9

### **Hypercube**

Ein Hypercube (sinngemäß ein Würfel mit vielen Dimensionen) bezeichnet sich auf eine Sammlung multidimensionaler Daten. Die ‘Ecken’ des Würfels werden Dimensionen genannt, die jeweiligen Bestandteile jeder Dimension heißen Elemente. 24

### **Join**

Die Join Operation verbindet zwei Relationen (Tabellen) ähnlich wie das kartesische Produkt allerdings werden nur solche Tupel ausgewählt, die in einer gewissen Bedingung miteinander stehen, indem sie ein Selektionsprädikat erfüllen. 21

### **Kennzahl**

Als ‘Kennzahl’ bezeichnet man i. d. R. das Merkmal, das gemessen wird, Kennzahlenwert ist die im einzelnen ermittelte Zahl. 15

### **MDX**

Multidimensional Expressions (MDX) ist eine Datenbanksprache für OLAP-Datenbanken. Sie wurde von Microsoft vorangetrieben und etabliert sich als Industriestandard. MDX wurde wesentlich von Mosha Pasumansky entwickelt. Die grundlegenden Bestandteile von MDX sind Measures und Dimensions, die den Fakten und Dimensionen eines Data-Warehouse entsprechen. 56

### **OLTP**

Online Transaction Processing Systeme werden für Tagesgeschäftsaktivitäten eingesetzt, haben hohe Ansprüche an die Bearbeitungsgeschwindigkeit. Für multidimensionale Analysen sind sie nicht geeignet. 19

### **Slice und Dice**

Slice und Dice beschreibt treffend die benutzerfreundlichen Auswertungsprinzipien von OLAP-Datenbeständen, da auf einfachste Weise das analysierte Ergebnis eingeschränkt oder erweitert bzw. geklippt und gedreht werden kann. 23

### **staging areas**

Staging wird typischerweise bei der Implementierung eines ETL-Prozesses verwendet. Die Daten werden dabei zunächst aus den Datenquellen extrahiert (extract) und in der Staging-Area gesammelt und zwischengespeichert. 8