

---

**Fachhochschule Ludwigshafen am Rhein**

**Hochschule für Wirtschaft**

Fachbereich III

Studiengang Wirtschaftsinformatik

**Bachelor-Thesis**

**Open Source Business Intelligence**

**zur Unterstützung eines Software-Validierungsprozesses**

Vorgelegt von:

**Volkert, Norman aus Mannheim**

Betreuer:

**Prof. Dr. rer. nat. Klaus Freyburger**

Erstellt bei:

**sobedi informationsmanagement gmbh**

Betreuer im Unternehmen:

**Dipl. Betriebswirt (FH) Dirk Laufer**

Bearbeitungszeit:

25.06.2007 bis 31.08.2007



---

# 1 Inhaltsverzeichnis

1	Inhaltsverzeichnis .....	2
2	Abbildungs- & Tabellenverzeichnis.....	4
2.1	Abbildungsverzeichnis .....	4
2.2	Tabellenverzeichnis .....	6
3	Einleitung .....	7
3.1	Motivation .....	7
3.2	Überblick .....	9
4	Business Intelligence & Data-Warehousing.....	10
5	Open Source: Eine neue Generation von Software .....	12
5.1	Open Source im Allgemeinen.....	12
5.2	Open Source Business Intelligence .....	21
6	Pentaho - Open Source Business Intelligence .....	24
6.1	Pentaho allgemein .....	24
6.2	Allgemeine Werkzeuge .....	28
6.2.1	Die Pentaho Business Intelligence Suite .....	28
6.2.2	Das Pentaho Design Studio .....	35
6.2.3	Installation der Pentaho BI Plattform .....	51
6.2.3.1	Installation der Pentaho Open BI Suite Pre-configured Installation ..	51
6.2.3.2	Manuelles Deployment der Pentaho BI Plattform.....	57
6.3	Der ETL-Prozess mit der Software Pentaho Data Integration .....	63
6.4	OLAP mit Mondrian.....	74
6.4.1	Mondrian Schema Workbench vs. Cube Designer.....	76
6.4.2	Parent-Child-Hierarchie .....	80
6.4.3	Anzeige eines Cubes mit JPivot .....	82
6.5	Reporting mit dem Modul Pentaho Report Designer .....	88
7	Praktischer Einsatz der Pentaho BI Suite am Beispiel der LAB-Simulator Reporting Engine.....	91
7.1	Die Unternehmen.....	91
7.2	Software-Validierung .....	91
7.3	Der LAB-Simulator .....	93
7.4	Die Reporting Engine als Modul des LAB-Simulators.....	99

---

7.5	Das Datenmodell der LAB-Simulator Reporting Engine.....	103
7.6	Der ETL-Prozess .....	108
7.7	Die Mondrian Cubes der Reporting Engine .....	127
7.8	JFree Reports .....	133
8	Fazit .....	145
9	Glossar & Abkürzungsverzeichnis .....	147
10	Literaturverzeichnis .....	150
10.1	Printmedien.....	150
10.1.1	Primär- und Sekundärliteratur .....	150
10.1.2	Pentaho Dokumentationen: .....	150
10.1.3	Firmeninterne Dokumentationen.....	151
10.2	Onlinequellen .....	151
10.2.1	Webseiten der Pentaho Corporation.....	151
10.2.2	Wikipedia - Freie Online-Enzyklopädie.....	152
10.2.3	Weitere Seiten im WWW .....	152
11	Anhang .....	154
11.1	Oracle SQL Code zum Anlegen der Datenbankobjekte.....	154
11.2	XML-Code des LAB-Simulator Cube 01 - Cycle-Process-Logic.....	165
11.3	XML-Code des LAB-Simulator Cube 02 - Cycle - Time scale - Logic .....	171
11.4	Beispiel der Struktur einer Pentaho Data Integration XML-Datei.....	178
11.5	LAB-Simulator Performance Report - CPU-Usage .....	183
11.6	PENTAHO PUBLIC LICENSE Version 1.0 .....	187
12	Eidesstattliche Erklärung.....	199

## 2 Abbildungs- & Tabellenverzeichnis

### 2.1 Abbildungsverzeichnis

Abbildungs-Nr.	Beschreibung	Seite
Abbildung 3-1:	Historische Entwicklung im BI-Bereich.....	8
Abbildung 5-1:	Copyleft-Ausprägungen.....	16
Abbildung 5-2:	Übersicht Open Source Bereiche.....	18
Abbildung 5-3:	TCO Faktoren.....	20
Abbildung 6-1:	Aufbau der Pentaho BI Suite.....	26
Abbildung 6-2:	Pentaho - Versionsvergleich Open Source vs. Subscription.....	28
Abbildung 6-3:	Architektur des Pentaho Servers.....	30
Abbildung 6-4:	Pentaho Server - Verzeichnisstruktur: Pentaho Solution Repository ...	33
Abbildung 6-5:	Pentaho Server - Verzeichnisstruktur: Pentaho Data.....	34
Abbildung 6-6:	Pentaho Server - Verzeichnisstruktur: Pentaho JBoss.....	35
Abbildung 6-7:	Pentaho Design Studio - Konfiguration Server URL.....	38
Abbildung 6-8:	Pentaho Design Studio - Pentaho Projekt anlegen.....	39
Abbildung 6-9:	Pentaho Design Studio - Tab General.....	40
Abbildung 6-10:	Pentaho Design Studio - Process View (links), Process Flow (rechts)	42
Abbildung 6-11:	Pentaho Design Studio - Process Inputs - Process Resource.....	44
Abbildung 6-12:	Pentaho Design Studio - Process Inputs - Process Input.....	45
Abbildung 6-13:	Pentaho Design Studio - Process Outputs.....	46
Abbildung 6-14:	Pentaho Design Studio - Process Action Relational.....	48
Abbildung 6-15:	Pentaho Design Studio - Tab XML Source.....	49
Abbildung 6-16:	Pentaho Design Studio - Tab Test.....	50
Abbildung 6-17:	Beispiel der Datei email_config.xml.....	55
Abbildung 6-18:	Installation des JBoss AS - Installationspakete.....	59
Abbildung 6-19:	Pentaho Data Integration - OS-Zeit-Kosten-Modell.....	64
Abbildung 6-20:	Pentaho Data Integration - Prozessablauf.....	65
Abbildung 6-21:	Pentaho Data Integration - Transformation.....	66
Abbildung 6-22:	Pentaho Data Integration - Job.....	67
Abbildung 6-23:	Pentaho Data Integration - Repositorykonfiguration.....	69
Abbildung 6-24:	Pentaho Data Integration - Repositoryanmeldung.....	69
Abbildung 6-25:	Pentaho Data Integration - Spoon.....	71

---

Abbildung 6-26: Pentaho Data Integration - Umgebungsvariable .....	72
Abbildung 6-27: Pentaho Data Integration - Unterstütze Datenbanken.....	74
Abbildung 6-28: Cubeaufbau mit dem Cube Designer .....	78
Abbildung 6-29: Mondrian Schema Workbench: Datenbankverbindung.....	79
Abbildung 6-30: Mondrian Schema Workbench - Ansicht einer Cubestruktur.....	80
Abbildung 6-31: Parent-Child-Hierarchie.....	81
Abbildung 6-32: Parent-Child-Hierarchie - Closure-Tabelle.....	82
Abbildung 6-33: JPivot - Browseransicht .....	84
Abbildung 6-34: OLAP Navigator .....	85
Abbildung 6-35: Drill Through .....	86
Abbildung 6-36: Drill Through - Ergebnis.....	86
Abbildung 6-37: Workspace des Pentaho Report Designer .....	90
Abbildung 7-1: Aufbau Generic Object Framework (GOF) .....	95
Abbildung 7-2: Module des LAB-Simulators .....	96
Abbildung 7-3: Hierarchische Logik des LAB-Simulators.....	99
Abbildung 7-4: (Hilfs-) Tabellen.....	105
Abbildung 7-5: Datenmodell: Star Schema Teil 1 .....	106
Abbildung 7-6: Datenmodell: Star Schema Teil 2 .....	108
Abbildung 7-7: ETL - Job zum Beladen des DWH.....	111
Abbildung 7-8: ETL - Loading-Tables-DT-VV,SOLV,SR,PRO,SC,SV,PRC,PS.ktr .	112
Abbildung 7-9: Step Text File Input: Konfiguration der Felder .....	113
Abbildung 7-10: Zuordnung: Felder Datenstrom → Feder Tabelle.....	114
Abbildung 7-11: Step: Insert / Update.....	115
Abbildung 7-12: Transformation: Delete-Datasets .....	116
Abbildung 7-13: Job: Execute SQL Script.....	117
Abbildung 7-14: Job: Execute a shell script.....	118
Abbildung 7-15: ETL: Update-Table-Event_Record_Check-svActivityObjectId#2...	121
Abbildung 7-16: Transformation: Loading-Table-CLOSURE_SIMULATION01 .....	121
Abbildung 7-17: Transformation: Loading-Table-DT_SIMULATION .....	122
Abbildung 7-18: Step: Filter rows .....	123
Abbildung 7-19: Transformation: Loading-Table-FT_SIM_EVENT.....	125
Abbildung 7-20: allMember einer Dimension .....	129
Abbildung 7-21: Attribute des Levels der Parent-Child-Hierarchie .....	131

Abbildung 7-22: JPivot - LAB-Simulator Cube 01 .....	132
Abbildung 7-23: JPivot - LAB-Simulator Cube 02.....	133
Abbildung 7-24: JFree Report: Auswahl des Simulation Cycle Step 1 .....	134
Abbildung 7-25: Auswahl des Simulation Cycle Step 2 .....	134
Abbildung 7-26: Action Sequence - Performance Report.....	135
Abbildung 7-27: Pentaho Design Studio - Process Action: Relational.....	136
Abbildung 7-28: Pentaho Design Studio - Process Action: Prompt/Secure Filter .....	137
Abbildung 7-29: Pentaho Design Studio - Process Action: JavaScript.....	138
Abbildung 7-30: Pentaho Design Studio - Process Action: Pentaho Report .....	140
Abbildung 7-31: Pentaho Report Designer - Definition der Parameter .....	141
Abbildung 7-32: Pentaho Report Designer - Parameterzugriff.....	142
Abbildung 7-33: CategorySetCollectorFunction & LinChartExpression .....	143

## 2.2 Tabellenverzeichnis

Tabellen-Nr.	Beschreibung	Seite
Tabelle 2-1:	Business Intelligence-Umsätze in Deutschland 2006 .....	10
Tabelle 3-1:	Software-Lizenzmodelle .....	13
Tabelle 4-1:	Vergleich Datei- vs. Datenbank-Repository .....	32
Tabelle 6-2:	Funktionsvergleich Cube Designer vs. Schema Workbench .....	78

## **3 Einleitung**

### **3.1 Motivation**

Business Intelligence (BI), Data-Warehousing: Begriffe die aus der IT nicht mehr wegzudenken sind.

Mit der wachsenden Globalisierung und den damit verbundenen Chancen, aber auch Risiken, wird es für Organisationen zunehmend wichtiger, die richtigen Entscheidungen zu treffen, um auf Marktveränderungen oder innovative Entwicklungen schnell reagieren zu können. Durch den sich immer schneller entwickelnden IT-Sektor und dem damit verbesserten Informationsaustausch in allen Businessbereichen, gewinnt der Faktor Zeit hinsichtlich des Wertes einer Information immer mehr an Gewicht. Zusätzlich zu den inhaltlichen Aspekten einer Information spielt deren Aktualität eine immer größere Rolle. Folglich müssen Entscheidungen nicht nur inhaltlich korrekt getroffen werden, sondern auch zeitlich optimal terminiert sein. Als elementares Instrument wird deshalb in vielen Unternehmen ein Data-Warehouse-gestütztes System zur Unterstützung der Entscheidungsfindung und der strategischen Planungen eingesetzt.

Die Implementierung eines Data-Warehouse und die Durchführung eines Business-Intelligence-Projekts sind für Unternehmen meist mit hohen Kosten verbunden. Ein großer Teil der Kosten entsteht durch die Investition in geeignete Technologien, z.B. ein Data-Warehouse und weiterer Komponenten, die für Business Intelligence benötigt werden.

Mit der Geburt von Open Source in den 90er Jahren des 20. Jahrhunderts war es nur eine Frage der Zeit, bis entsprechende Produkte zur Realisierung von Business Intelligence zur Verfügung standen. Nach der Jahrtausendwende erschienen erstmals einzelne Komponenten auf dem Markt.

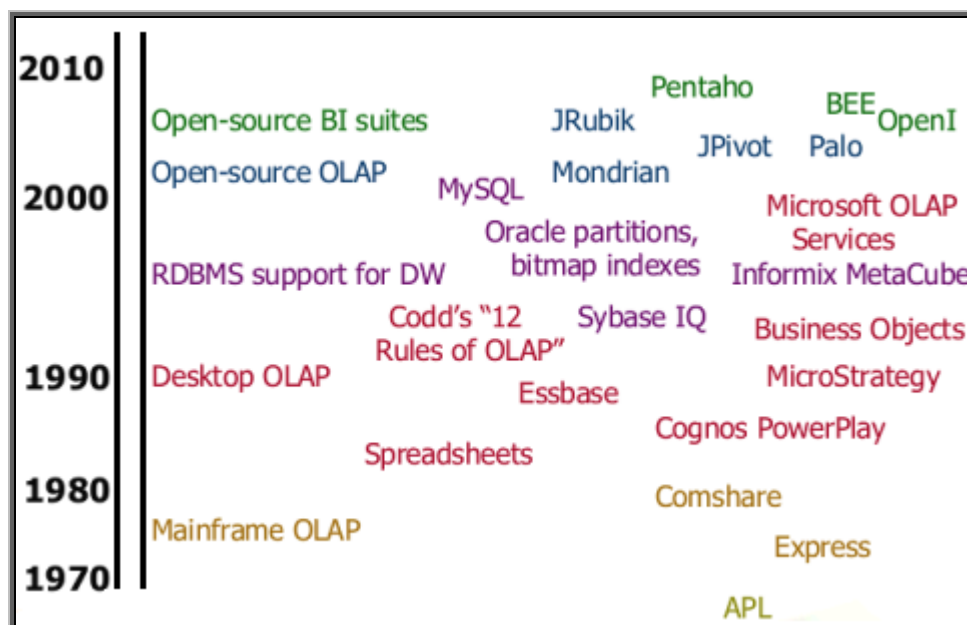


Abbildung 3-1: Historische Entwicklung im BI-Bereich

Die Produkte alleine reichten nicht aus, um Business Intelligence betreiben zu können. Sie mussten mit anderen kombiniert werden, um einsetzbare Systeme zu bilden. Aufgrund verschiedener Probleme gestaltete sich dies mitunter als schwierig. Der große Vorteil von Anbietern proprietärer Software war, dass diese Lösungen anbieten konnten, die alle benötigten Komponenten für Business Intelligence beinhalteten und somit aus „einem Guss“ waren.

Betrachtet man die Gegenwart, so entwickeln sich Open Source Projekte, wie bspw. Pentaho, JasperSoft, BEE, oder OpenI, deren Ziel es ist, ebenfalls Business Intelligence Lösungen aus „einer Hand“ anbieten zu können. Dies eröffnet viele Möglichkeiten, da die Zugangsbarrieren zu Open Source Software praktisch nicht existieren.

Aufgrund des Potentials, das in der Thematik Business Intelligence besteht ist es höchst interessant, die Möglichkeiten dieser Open Source Projekte zu erforschen und die Praxistauglichkeit zu testen.

Aufgrund des Potentials, das in den Open-Source-Produkten steckt, ist der Aufbau von Know-how höchst interessant, um die Technologien in zukünftigen Projekten einsetzen zu können.

## 3.2 Überblick

Im vorliegenden Typoskript wird zu Beginn allgemein auf Open Source eingegangen. Im Anschluss werden die bekanntesten Open-Source-Business-Intelligence-Projekte vorgestellt. Im weiteren Verlauf wird der Fokus auf die Lösung von Pentaho gerichtet. Dabei werden die einzelnen Komponenten vorgestellt und in einem Praxisfall unter Einsatz der Pentaho Open Source Business Intelligence Platform eine Lösung aufgebaut, diese dient dazu, Daten aus einem Softwarevalidierungsprozess zu analysieren.

Ziele und Inhalt des Typoskripts sind:

- Ihnen eine allgemeine Sicht auf Open Source zu vermitteln.
- Ihnen die bekannten Open-Source-Business-Intelligence-Projekte vorzustellen.
- Ihnen den Einstieg in die Welt von Pentaho zu erleichtern.
- Die Beschreibung des Aufbaus, des Ablaufs der Installation und der Komponenten der Pentaho Business Intelligence Platform aus technischer Sicht.
- Die Darstellung des praktischen Einsatzes von Pentaho in den Bereichen ETL-Prozess, Data-Warehouse mit einem Star Schema, Cubes inkl. Reports.

Kenntnisse des Lesers im Bereich Business Intelligence und im Umgang mit Datenbanken werden vorausgesetzt. Das Typoskript geht nicht im Detail auf Business Intelligence ein. Der Fokus liegt auf der Umsetzung und dem Einsatz von Open Source Business Intelligence mit Pentaho.

Der Autor gibt an vielen Stellen nützliche Hinweise, die beim Arbeiten mit Pentaho helfen „Fallen“ zu vermeiden.

Aus Vereinfachungsgründen und zur besseren Lesbarkeit wurde mitunter bei unspezifischen Darstellungen im vorliegenden Typoskript die männliche und nicht die weibliche Form gewählt. An dieser Stelle sei darauf verwiesen, dass der Autor sich ausdrücklich zu der in Artikel 3 des Grundgesetzes der Bundesrepublik Deutschland verankerten Gleichberechtigung von Mann und Frau bekennt und auch in den Formulierungen, in denen nur die männliche Form gewählt ist, Frauen immer eingeschlossen sind.