

MASTERARBEIT

EINFÜHRUNG EINES DATA WAREHOUSE BEI EINEM UNTERNEHMEN IN DER SOZIALWIRTSCHAFT ZUR OPTIMIERUNG DER KUNDENDATEN

Am Beispiel Jugend am Werk Steiermark GmbH

ausgeführt am



Studiengang

Informationstechnologien und Wirtschaftsinformatik

Von: Thomas Knaller

Personenkennzeichen: 1410320028

Graz, am 3. Juli 2017

.....
Unterschrift

EHRENWÖRTLICHE ERKLÄRUNG

Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benützt und die benutzten Quellen wörtlich zitiert sowie inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

.....

Unterschrift

KURZFASSUNG

Diese Masterarbeit befasst sich mit den Themen „Business Intelligence“ und „Data Warehousing“ im Umfeld der Sozialwirtschaft in Österreich am Beispiel des Unternehmens Jugend am Werk Steiermark GmbH (JAW). Das Ziel der Arbeit ist die Erläuterung der Vorgangsweise zur Erstellung eines Data Warehouse in einer heterogenen Systemlandschaft, wie sie bei JAW existiert. Im Zuge der Arbeit werden die Fachbegriffe rund um diese Themen erläutert und die Besonderheiten eines Data Warehouse im Vergleich zu operativen Datenbankapplikationen aufgezeigt. Um die Daten von einer solchen operativen Applikation in ein Data Warehouse zu übertragen, wird ein so genannter Extract-Transform-Load-Prozess durchgeführt, mit dessen Merkmalen sich diese Arbeit ebenfalls beschäftigt. Darauf aufbauend werden die dafür technisch eingesetzten Werkzeuge analysiert sowie deren Begrifflichkeiten geklärt. Auch die gesetzlichen Rahmenbedingungen des Datenschutzes, die direkten Einfluss auf ein Data Warehouse haben, werden im Verlauf der Arbeit dargelegt. Die Regelungen des Datenschutzes widersprechen jedoch den grundlegenden Charakteristika eines Data Warehouse. Diese Konfliktsituation sowie ein möglicher Lösungsansatz werden ebenfalls diskutiert. Mithilfe des Wissens aus dieser Vorbereitung werden die Dienstleistungen des Unternehmens erfasst und deren Datenhaltung analysiert. Durch die Einteilung in strukturierte Datenbankapplikationen und unstrukturierte Dateien wird sichtbar, dass ein Data Warehouse nur mit sehr großem Aufwand tatsächlich sämtliche Unternehmensdaten beinhalten kann. Anschließend wird ein gemeinsames Datenmodell für diese Daten geschaffen, welches die Daten aus der heterogenen Applikationslandschaft beherbergen wird. Abschließend wird ein Prototyp erstellt, der den Weg der Implementierung sowie die Möglichkeiten eines Data Warehouse und der dazugehörigen Werkzeuge beschreibt.

ABSTRACT

This thesis investigates business intelligence and data warehousing within the context of the Austrian social economy. In order to be able to draw a more sophisticated picture, the author's employer Jugend am Werk Steiermark GmbH (JAW) acts as an example. The goal of the thesis is to map the implementation of a data warehouse in a heterogeneous environment, such as at JAW. The technical terms of the topics are explained and the characteristics of a data warehouse in comparison to an operational database application are pointed out. The "extract-transform-load" process that transfers data from the operational application into the data warehouse is described. The technologies used in this process are analysed and their concepts explained. Legal aspects of privacy have direct influence on such a data warehouse and are thus also covered. Privacy stands in immediate conflict with a data warehouse and this conflict, as well as a possible solution, is discussed. Building on this, the thesis continues by creating an inventory of the company's services and its data management, which are categorized as "structured" and "unstructured" data. This categorization illustrates the enormous workload to be completed before a data warehouse can house a company's entire data. Moreover, a mutual data model, which allows the data warehouse to accommodate the data of the heterogeneous application environment, is created. The thesis concludes with the creation of a prototype to show the necessary steps, as well as the possibilities of a data warehouse and its associated tools.

INHALTSVERZEICHNIS

1	EINLEITUNG	8
1.1	Aufgabenbeschreibung	8
1.2	Motivation	8
1.3	Zielsetzung	9
1.4	Forschungsfragen	9
1.5	Spezifikation	10
1.6	Dienstleistungsbereiche	11
2	BUSINESS INTELLIGENCE	13
2.1	Begriffsdefinition	13
2.2	Data Warehouse	16
2.2.1	Themenorientierung (subject orientation)	17
2.2.2	Beständigkeit (non-volatility)	17
2.2.3	Vereinheitlichung (integration)	17
2.2.4	Zeitorientierung (time variance)	18
2.3	Datenmodell	19
2.3.1	Normalformen operationaler Datenbanken	19
2.3.2	Dimensionales Modell	20
2.3.3	Slowly Changing Dimensions	20
2.3.4	Surrogate Key	21
2.3.5	Faktentabellen	21
2.4	Extract-Transform-Load	22
2.4.1	Extract	22
2.4.2	Transform	23
2.4.2.1	Datenvereinheitlichung	23
2.4.2.2	Anpassung an dimensionales Datenmodell	24
2.4.2.3	Fehlerbehebung	24
2.4.3	Load	25
2.5	OnLine Analytics Processing	26
2.5.1	Cube	26
2.5.2	Measure	27
2.5.3	Datamart	27

2.5.4	MDX.....	28
3	WERKZEUGE & UMGEBUNG	29
3.1	Werkzeuge.....	29
3.1.1	SQL Server	29
3.1.2	SQL Server Integration Services	31
3.1.3	SQL Server Analysis Services.....	31
3.1.4	SQL Server Data Tools	32
3.1.5	Excel / PowerPivot.....	33
3.1.6	Power BI	34
3.2	Datenschutz.....	34
3.2.1	Begriffe	35
3.2.2	Aufbewahrungspflichten, Löschpflichten & Data Warehouse	37
4	ANALYSEN & METHODEN.....	38
4.1	OLAP	38
4.1.1	Slicing	38
4.1.2	Dicing.....	39
4.1.3	Pivoting.....	40
4.1.4	Drill-Down / Roll-Up.....	40
4.1.5	Drill-Through & Drill-Across	41
4.1.6	Split & Merge	42
4.2	Explorative Datenanalyse.....	42
4.3	Befragung	44
5	IST-ZUSTAND.....	45
5.1	Aufstellung Datenhaltung	45
5.1.1	Gliederung nach Applikation.....	45
5.1.2	Aufstellung unstrukturierte Daten	46
5.1.3	Aufstellung Daten nach Dienstleistung.....	47
5.2	Analyse Geschäftsprozesse	48
5.2.1	Stammdaten	49
5.2.2	Eintritt.....	50

5.2.3	Änderung	50
5.2.4	Austritt.....	51
5.2.5	Monatliches Reporting	51
5.3	Analyse Datenmodelle.....	52
5.3.1	Dokumente am Fileserver (keine Datenbank).....	52
5.3.2	AABDB.....	52
5.3.3	agilion	54
5.3.4	ARGE-Datenbank	55
5.3.5	Be-FIT	55
5.3.6	eAMS	56
5.3.7	„KundInnenakt stationär“	56
5.3.8	BMD NTCS	56
5.3.9	GBPDB	59
6	DIMENSIONALES DATENMODELL	62
6.1	Dimension Zeit.....	62
6.2	Dimension Kunde	63
6.3	Dimension Kundennummer	65
6.4	Dimension Dienstleistung	66
6.5	Dimension Kostenstelle	67
6.6	Dimension Standort	67
6.7	Fakten Kundenstamm.....	69
6.8	Fakten Kundenstamm-Snapshot Stichtag	71
7	TIEFERE ANALYSE & PROTOTYP.....	73
7.1	Datenquellen BMD & „KundInnenakt stationär“	73
7.1.1	Besonderheiten.....	73
7.1.2	Datenfehler	74
7.1.2.1.	Überlappende Zeiträume	74
7.1.2.2.	Ungültige Zeitraumeinträge	76
7.1.3	Late Data Entry / Rückwirkende Daten	77
7.2	Datenquelle AABDB	77
7.2.1	Besonderheiten.....	77
7.2.2	Datenfehler	78
7.3	Prototyp	78

7.3.1	Staging Area beladen	79
7.3.2	Staging transformieren mit SSIS	80
7.3.3	Data Warehouse beladen	83
7.3.3.1.	Zeitdimension	83
7.3.3.2.	Dimensionen beladen	83
7.3.3.3.	Fakten erzeugen	84
7.3.4	Würfel bereitstellen mit SSAS	86
7.3.5	Daten analysieren	87
8	RESÜMEE	90
	ABKÜRZUNGSVERZEICHNIS.....	92
	ABBILDUNGSVERZEICHNIS	94
	TABELLENVERZEICHNIS	95
	LISTINGS 96	
	LITERATURVERZEICHNIS	97
	ANHANG 1: ERHEBUNG DIENSTLEISTUNGEN UND DATENABLAGEN	103
	ANHANG 2: SCRIPT ZUR ANLAGE DER DIMENSION ZEIT.....	118
	ANHANG 3: SCRIPT ZUR ANLAGE DER TRANSFORMATION „DIENSTLEISTUNG“ DER DATENQUELLE „BMD“.....	121
	ANHANG 4: SCRIPT ZUR ANLAGE DER TRANSFORMATION „KUNDE“ DER DATENQUELLE „BMD“ 122	
	ANHANG 5: SCRIPT ZUR BELADUNG DER DIMENSIONEN „KUNDE“ UND „KUNDENNUMMER“ 124	
	ANHANG 6: SCRIPT ZUR BELADUNG DER FAKTENTABELLE „KUNDENSTAMM“	126
	ANHANG 7: SCRIPT ZUR BELADUNG DER FAKTENTABELLE „KUNDENSTAMM-SNAPSHOT“	129

1 EINLEITUNG

Die Aufgabenstellungen von IT-Abteilungen in Unternehmen werden immer umfangreicher und gehen mehr in die Richtung des Wissensmanagements. Das bedeutet oft, dass die tägliche Arbeit im Umfeld der klassischen IT-Infrastruktur zurückgeht und die Datenablage, die Datensammlung, sowie die Datenverarbeitung in den Vordergrund rücken. Genau in diese Aufgabenstellung greift die Thematik Business Intelligence – mit vielen unterschiedlichen Unterdisziplinen, die aus Daten neues Wissen generiert. Dazu gehört insbesondere das strukturierte Sammeln und Verarbeiten von Daten aus unterschiedlichen Quellen sowie das Abspeichern in einer gemeinsamen Datenstruktur, dem so genannten Data Warehouse.

1.1 Aufgabenbeschreibung

Ausgehend von der häufigen Herausforderung des Managements der Jugend am Werk Steiermark GmbH (JAW), Informationen zu bestimmten Kunden zu beschaffen, wurde schnell klar, dass eine gemeinsame Datenablage über alle Abteilungen und Datenbanksysteme hinweg notwendig ist, um schnelle, effiziente und richtige Entscheidungen treffen zu können. Die bestehenden dezentralen operativen Applikationen in eine gemeinsame Applikation, ein allwissendes ERP-System, zu überführen, ist jedoch eine sehr weitreichende Aufgabenstellung, die gravierende Änderungen in sämtlichen, oft sehr gut funktionierenden, Geschäftsprozessen sowie einen hohen wirtschaftlichen Aufwand mit sich bringt.

1.2 Motivation

Aus dieser Aufgabenbeschreibung heraus ist es für den Autor und für das betroffene Unternehmen sehr interessant, die operativen Geschäftsprozesse darzustellen und Massendatenverarbeitung in den Fokus dieser Arbeit zu stellen. Das Auslesen von Daten aus bestehenden Datenbankapplikationen und das Finden einer homogenen Lösung für die gemeinsame Ablage ist eine herausfordernde Tätigkeit des Datenmanagements. Auch ist das Erhöhen der Reaktionsfähigkeit der Geschäftsführung eine essentielle Aufgabe der IT-Abteilung, die zu einer positiven Weiterentwicklung des Unternehmens beiträgt.

1.3 Zielsetzung

Da wie bereits erwähnt ein ERP-System eine weitaus komplexere Aufgabenstellung darstellt, als technisch die Daten aller Systeme in einem Data Warehouse (DW) zur Verfügung zu stellen und dem Management mit einer möglichst einfachen Oberfläche bereitzustellen, beschäftigt sich diese Masterarbeit mit der Einführung eines solchen DW. Der Fokus liegt hierbei auf den Stammdaten der Kunden von JAW, um im ersten Schritt die Auffindbarkeit der Daten zu erhöhen, dabei aber noch keinen computergestützten Mehrwert aus den Informationen selbst zu generieren. In dieser Ausprägung ist es mit einer Art Enterprise Search vergleichbar, welche Kundendaten in einer optimal für das Management aufbereiteten Form zur Verfügung stellt und in Zukunft auf weitere Datenarten und Auswertungen erweitert werden kann.

1.4 Forschungsfragen

Die Arbeit untersucht, wie und unter welchen Umständen es möglich ist, ein heterogenes System bestehend aus unterschiedlichen Datenquellen zu einem DW zu vereinen und damit einen Mehrwert für die Nutzer dieser Systeme erzielen zu können.

Folgende Hypothesen werden während der Mastarbeit untersucht:

Hypothese 1:

- H1: Die Daten aller Dienstleistungen aus dem Sozialbereich liegen in strukturierter und zentral zugänglicher Form vor und können in einem DW gesammelt, mit anderen Daten verknüpft und weiter ausgewertet werden.
- H0: Es gibt kaum Daten von Dienstleistungen im Sozialbereich, die in strukturierter Form und zentral zugänglich vorliegen.

Hypothese 2:

- H1: Die Analyse von langfristigen Daten hilft dabei, Trends zu erkennen, die außerhalb des Einflussbereiches des Unternehmens liegen. Ein Trend könnte beispielsweise eine Veränderung der Kundenzahlen im Asylbereich nach politischen Entscheidungen sein.
- H0: trotz der starken politischen Prägung der Sozialwirtschaft sind keine Muster in der Vergangenheit zu finden.

Hypothese 3:

- H1: Durch die Bereitstellung einer zentralen Informationsquelle kann die Geschäftsführung bessere Entscheidungen treffen.

- H0: Die Geschäftsführung ignoriert die Daten bei Entscheidungsfindungen mangels Vertrauen in die Datenqualität.

Methode

Die Operationalisierung dieser Hypothesen wird wie folgt ablaufen:

Hypothese 1 wird mittels Befragungen überprüft. Das bedeutet, dass gemeinsam mit den Verantwortlichen in den einzelnen Dienstleistungsbereichen erarbeitet wird, bei welchen Dienstleistungen welche Arten der Datenspeicherung genutzt werden. Um die H0 verwerfen zu können, wird gemessen, ob die Anzahl an Dienstleistungen mit strukturierter Datenablage dividiert durch die Anzahl an Dienstleistungen Gesamt größer gleich 75% ist. Das Ziel ist hierbei bewusst unter 100% angesetzt, da bei einigen Dienstleistungen eine rein externe Applikation zur Datenhaltung genutzt wird, auf die seitens JAW IKT kein Einfluss besteht und im Rahmen dieser Masterarbeit nur kurz erwähnt werden. Ist dies nicht der Fall, kann die H0 nicht verworfen werden und die Hypothese hat sich nicht bewahrheitet.

Für die Hypothese 2 werden die gesammelten Daten explorativ untersucht und eine These aufgestellt, welche einen Trend beschreibt und begründet. Zum Verwerfen der H0 und der damit einhergehenden Verifizierung der H1 muss diese These nachvollziehbar überprüft werden können.

Für die Hypothese 3 wird ein Prototyp skizziert, welcher der Geschäftsführung zeigen soll, was der Output des Projekts sein kann. Mithilfe der Reaktionen auf diese Prototypen wird ebenfalls der Projektoutput optimiert, sodass etwaige Wünsche berücksichtigt werden können. Diese könnten beispielsweise die Transparenz der Datenherkunft betreffen, damit eben das mangelnde, in der H0 beschriebene, Vertrauen verworfen werden kann. Die Reaktionen der Geschäftsleitung, welche mithilfe einer Befragung eingeholt werden, dienen der Erfolgsmessung dieser Hypothese.

1.5 Spezifikation

Im Rahmen der Arbeit werden Begriffe verwendet, die im allgemeinen Sprachgebrauch anders verstanden werden. Unter dem Begriff „Kunden“ werden ausschließlich natürliche Personen verstanden, welche die Sozialdienstleistungen von JAW in Anspruch nehmen. Somit ist der Begriff „Kunde“ klar getrennt und unabhängig vom Debitor, welcher die Rechnung erhält und im allgemeinen Sprachgebrauch als Kunde gilt.

1.6 Dienstleistungsbereiche

Wie in Abbildung 1 ersichtlich, gliedert sich die Jugend am Werk Steiermark GmbH im Groben in eine Geschäftsführung, deren Stabstellen Qualitätsmanagement, Facility Management und Marketing & Öffentlichkeitsarbeit, sowie in drei Geschäftsbereiche. Das Personalmanagement ist für sämtliche personellen Themenstellungen verantwortlich, Finanzmanagement & Verwaltung mit den Teams Buchhaltung, Controlling, Informations- und Kommunikationstechnologie sowie Personalverrechnung kümmert sich um sämtliche finanziellen und informationstechnologischen Aspekte und der Geschäftsbereich Dienstleistungsmanagement um die operative Tätigkeit des Unternehmens.

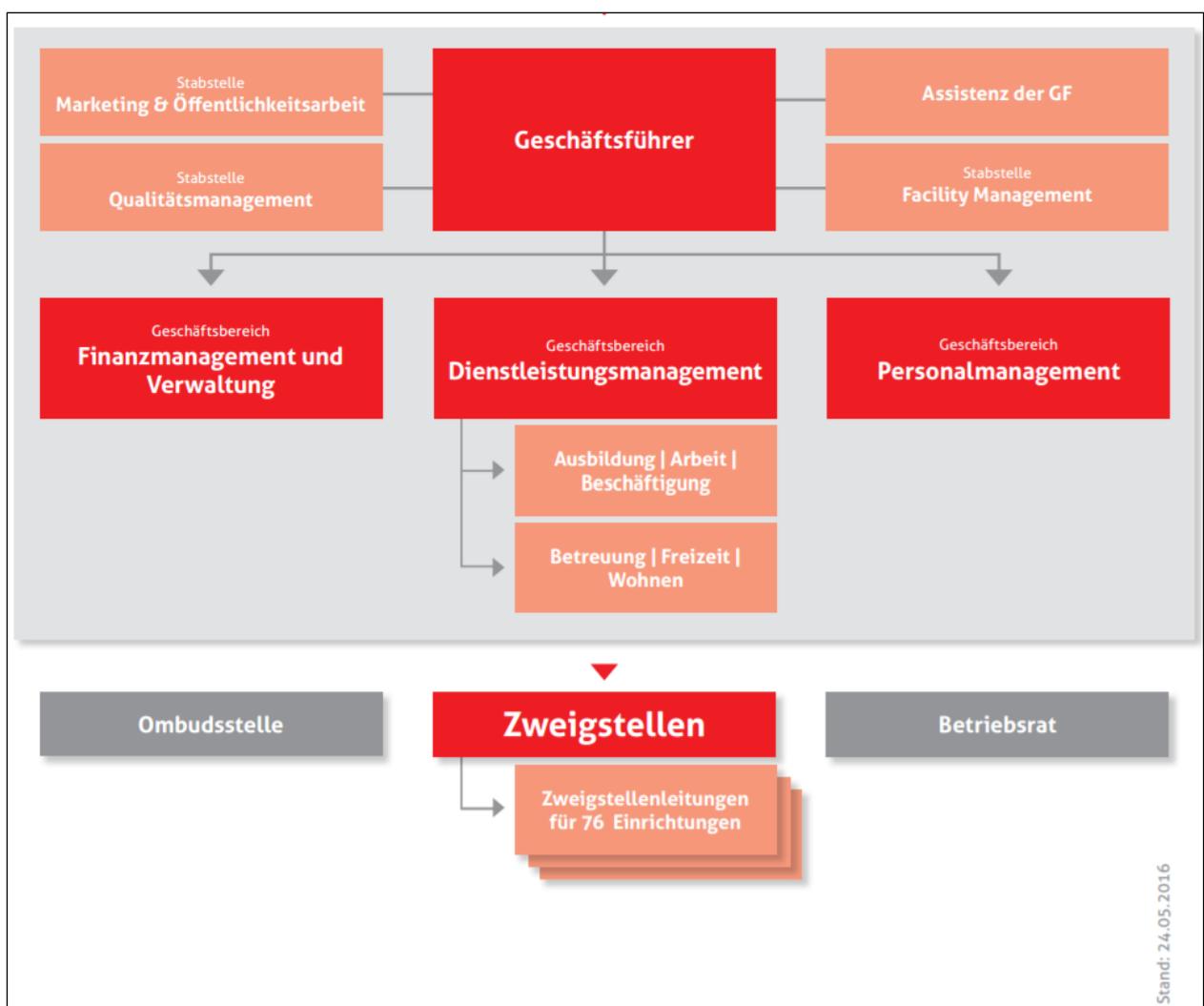


Abbildung 1 Organigramm Jugend am Werk Steiermark GmbH. Quelle: Qualitätsmanagementsystem im Intranet von JAW.

Das Dienstleistungsmanagement selbst wird in zwei Bereiche unterteilt, die von unterschiedlichen Personen geleitet werden. Der Bereich „Ausbildung | Arbeit | Beschäftigung“ (AAB) beschäftigt sich mit der täglichen Arbeit von Menschen sowie deren Heranführen an den Arbeitsmarkt. Dies wird mithilfe von Aufträgen unterschiedlichster Fördergeber realisiert, unter anderem vom Bund

Österreich in Form des Sozialministeriumservice, dem Arbeitsmarktservice, dem Land Steiermark oder auch der Stadt Graz. Der Bereich „Betreuung | Freizeit | Wohnen“ (BFW) hat die Aufgabe, Menschen in ihrer Freizeitgestaltung und bei ihrer Wohnsituation zu unterstützen. Die Dienstleistungen haben verschiedene Zielgruppen, aktuell werden Jugendliche, Menschen mit Behinderung, Langzeitarbeitslose und Asylwerber fokussiert.

Unterhalb der beiden Dienstleistungsmanager sind Zweigstellenleiter angesiedelt, welche unmittelbar mit den operativ tätigen Mitarbeitern die Interessen des Unternehmens regional vertreten. Sie sind Vorgesetzte einer Zweigstelle, welche aus mehreren unterschiedlichen Standorten und Dienstleistungsangeboten bestehen kann.

Zusätzlich werden zahlreiche Projekte mit EU-Geldern realisiert, welche zumeist den Wissensaustausch über sozialpädagogische Themen zum Inhalt haben und Kooperationen zwischen unterschiedlichen Trägern in allen EU-Ländern hervorrufen.

2 BUSINESS INTELLIGENCE

Business Intelligence (BI), Big Data, Data Warehousing (DW), OLAP, Data Mining (DM) – über all diese Begriffe stolpert man, sobald man über längere Zeit mit großen Datenmengen zu tun hat. Sei es in der IT oder im Controlling von großen Unternehmen, beim Lesen von Nachrichten aus der Politik oder bei Produktvorstellungen großer Software-Konzerne – diese Schlagwörter fallen immer häufiger. Um einen Einblick in die Welt von BI bieten zu können, wird im Zuge dieser Masterarbeit versucht, diese Begriffe zu erläutern, voneinander abzugrenzen und deren Schnittstellen zueinander zu beschreiben.

2.1 Begriffsdefinition

BI ist der aktuellste Fachbegriff für eine Themen- und Aufgabenstellung, die schon lange in der Computerwelt existiert und hat keine eindeutige Definition in einem Lexikon – es ist der interdisziplinäre Überbegriff für alle Strategien, Methoden, Technologien und Darstellungsvarianten zur Verarbeitung und Interpretation von großen Datenmengen. Diese reichen von Grundlagen der Betriebswirtschaftslehre (BWL) über Informatik und Softwareentwicklung hin zu Operations Research (OR) – der mathematischen und algorithmischen Analyse von betriebswirtschaftlichen Prozessen und Abläufen. Müller/Lenz beschreiben BI als Formel: „BI = 50% BWL/OR + 25% DM/Statistik + 25% DW“. (Müller & Lenz, 2013, S. 3) Diese Formel deckt sich mit der Übersetzung aus dem Englischen, Intelligence steht in diesem Zusammenhang schließlich für Information, Einsicht und Verständnis (Gluchowski, Gabriel, & Dittmar, 2008, S. 89): Es ist ein hohes Maß an Wissen über betriebswirtschaftliche Abläufe (BWL) und deren analytische Quantifizierung (OR) notwendig, um mit etwas Mathematik (DM/Statistik) und Technik (DW) schließlich neue Informationen und damit Wissen und Wettbewerbsvorteile zu generieren (Schön, 2012, S. 273).

BI ist ein Begriff für Management Support Systeme (MSS), also Computersysteme, die dem Management eines Unternehmens Daten liefern, auf deren Basis Entscheidungen gefällt werden können. Somit können solche Systeme auch entscheidungsunterstützende Systeme (EUS) genannt werden. Diese Systeme fällen selbst keine Entscheidungen, sie liefern lediglich Daten, auf deren Basis Fach- und Führungskräfte bessere Entscheidungen treffen können. (Lusti, 1999, S. 3)

In der Literatur wird in drei unterschiedliche Arten gegliedert, wie der Begriff BI verstanden werden kann. Das „enge“ BI-Verständnis beschränkt sich auf die Themen OLAP und Management Information Systems (MIS) sowie Executive Information Systems (EIS). (Kemper, Baars, & Mehanna, 2010, S. 3). Das enge BI-Verständnis hat also den Fokus rein auf denjenigen

Systemen und Anwendungen, welche die Fach- und Führungskräfte direkt mit Informationen zur Entscheidungsfindung versorgen, ohne dabei auf dafür notwendige Hilfssysteme oder infrastrukturelle Komponenten Rücksicht zu nehmen.

OLAP wird in Kapitel 2.5 „OnLine Analytics Processing“ detailliert erläutert.

MIS entstanden in den 1960er Jahren als erster Versuch, über mehrere Insellösungen hinweg Daten einheitlich darzustellen und dem Management zur Verfügung zu stellen. Allerdings wurde aus einem Informationsdefizit eine Informationsflut, da die Daten kaum aufbereitet oder analysiert wurden. Auch heute sind MIS noch in Unternehmen zu finden, meistens in Form von standardisierten Berichten mit nur sehr einfachen algorithmischen Auswertungen. (Gluchowski, Gabriel, & Dittmar, 2008, S. 72ff)

EIS entstanden in den 1980er Jahren als verbesserte Variante von MIS. Hierbei werden die Informationen dialog- und datenorientiert transportiert, wodurch die Akzeptanz in der Zielgruppe „Fach und -Führungskräfte“ massiv gesteigert werden konnte. EIS sind jedoch immer unternehmensspezifisch. Die Daten werden als mehrdimensionaler Würfel, also als Cube, aufbereitet und haben dadurch bereits einen hohen Verwandtschaftsgrad zu OLAP. (Gluchowski, Gabriel, & Dittmar, 2008, S. 74)

Das „analyseorientierte“ BI-Verständnis ergänzt das enge BI-Verständnis um folgende weitere Themengebiete, die nachfolgend näher beschrieben werden:

- Ad-Hoc-Reporting
- Kennzahlen / Balanced Scorecards
- Data / Text Mining
- Business Activity Monitoring
- Analytisches Customer Relationship Management (CRM)

Ad-Hoc-Reporting ist der Fachbegriff für individuell benötigte Auswertungen von Fach- und Führungskräften, die in der Regel die Expertise von bestimmten Personen erfordern, um geliefert werden zu können. Diese Personen sind in den meisten Unternehmen in einer Controlling-Abteilung angesiedelt. Im Sinne von Business Intelligence sollten die Rohdaten jedoch in einer derart verständlichen Form vorliegen, dass sie entweder sehr rasch vom Controlling oder sogar von der Fach- oder Führungskraft selbst ausgewertet werden können. (Schön, 2012, S. 58)

Kennzahlen bzw. Balacaced Scorecards werden für jedes Unternehmen individuell festgelegt und entweder mühevoll vom Controlling zusammengestellt oder ausprogrammierter Weise automatisiert zur Verfügung gestellt. Die größte Herausforderung dabei ist jedoch nicht die technische Umsetzung, sondern die unternehmensspezifische Definition der einzelnen BSC-

Elemente und der darunterliegenden Strategie, was diese Zahlen aussagen und welche Bedeutung Änderungen der einzelnen Ergebniswerte haben. (Gluchowski, Gabriel, & Dittmar, 2008, S. 223)

Data bzw. Text Mining sind Verfahren, die mithilfe von Ansätzen künstlicher Intelligenz versuchen, durch Clusterbildung, neuronale Netze, Regressionsanalysen und weitere algorithmische Analysen Informationen und Wissen aus großen Datenmengen zu generieren. (Schön, 2012, S. 254)

Das Business Activity Monitoring beschreibt Systeme und Methoden zur Messung von Geschäftsprozessen und deren Zielen bzw. der Zielerreichung. (Gluchowski, Gabriel, & Dittmar, 2008, S. 90)

Analytisches Customer Relationship Management ist das Resultat der Entwicklung, dass sich Unternehmen stärker auf die Anliegen und Reize ihrer Kunden ausrichten und das Kundenbeziehungsmanagement daher einen elementaren Bestandteil der unternehmerischen Tätigkeit darstellt. Nach jahrelangem Sammeln von unzähligen Informationen im Austausch mit den Kunden – angefangen von Webshops, über Callcenter bis hin zu persönlichen Gesprächen mit den Kunden – die alle in CRM-Systemen landen, wurde klar, dass in diesen Daten ein großes Potential zur Analyse und damit einhergehenden besseren Rücksichtnahme auf die Bedürfnisse der Kunden möglich ist. (Gluchowski, Gabriel, & Dittmar, 2008, S. 236ff)

Im Gegensatz zum engen BI-Verständnis geht es beim analytischen BI-Verständnis mehr darum, woher die Daten kommen und wie sie aufbereitet werden können, um einen tatsächlichen Mehrwert liefern zu können.

Beim „weiten“ BI-Verständnis kommen noch weitere Themen hinzu, nämlich Standard-Reporting, ETL und Data Warehouse. (Gluchowski, Gabriel, & Dittmar, 2008, S. 92)

Standard-Reporting befasst sich mit einheitlichen und regelmäßigen Berichten, wie beispielsweise Berichten aus Rechnungslegungsvorschriften wie dem UGB¹, EStG² oder aus IFRS³, sowie internen Berichten an Entscheidungsträger wie Monatsberichte, Quartalsberichte oder Soll-Ist-Vergleiche.

ETL wird in Kapitel 2.4 „Extract-Transform-Load“ und Data Warehouse in Kapitel 2.2 „Data Warehouse“ detaillierter beleuchtet.

¹ Unternehmensgesetzbuch

² Einkommensteuergesetz

³ International Financial Reporting Standards

Dies bedeutet, dass sich der weitere Verlauf dieser Masterarbeit auf diejenigen Teilgebiete konzentriert, die eher technischer Natur sind und in unmittelbarem Zusammenhang mit Data Warehouse stehen. Zunächst wird Data Warehouse selbst beschrieben, um anschließend auf dessen Befüllung mittels ETL näher einzugehen sowie die Analyse der dort vorhandenen Daten mittels OLAP zu beschreiben.

2.2 Data Warehouse

Ein Data Warehouse hat zum Ziel, Managern und Analysten Möglichkeiten zu geben, um schnellere und bessere Entscheidungen zu fällen. Dazu stellt es Datensammlungen mit folgenden Eigenschaften bereit:

- Themenorientierung (subject orientation): die Auswahl der Daten basiert auf Dingen oder Personen und nicht auf Prozessen
- Beständigkeit (non-volatility): die Daten werden dauerhaft (nichtflüchtig) gespeichert
- Vereinheitlichung (integration): es werden alle Daten, unabhängig vom Datenformat in der Quellapplikation, einheitlich gespeichert
- Zeitorientierung (time variance): Analysen sollen zeitliche Veränderungen berücksichtigen können, es wird die Dimension „Zeit“ eingeführt

(Krishnan, 2013, S. 128) (Navrade, 2008, S. 13) (Silvers, 2008, S. 12)

Ein Data Warehouse ist im Gegensatz zur Datenbank einer operativen Anwendungsapplikation keine transaktionale, sondern eine analytische Datenbank. Dies bedeutet, dass die Datenstruktur nicht auf die tägliche operative Tätigkeit und damit einhergehende Transaktionen wie beispielsweise Buchungen ausgelegt ist, sondern ein Datenmodell bereitstellt, welches analytische Verfahren effizienter zulässt (subject-oriented). (Lusti, 1999, S. 124ff)

Bei einer genaueren Betrachtung teilt sich der Begriff in die Richtungen „zentrales Data Warehouse“ (DW), „Enterprise Data Warehouse“ (EDW) und „Data Warehouse System“ (DWS). Ein DW erstreckt sich nicht unbedingt auf ein ganzes Unternehmen, sondern bildet meist nur einen Teilbereich, beispielsweise eine Anwendungsapplikation ab. Ein EDW ist eine unternehmensweite Datenbank, die sämtliche Daten in ein einheitliches Datenmodell überführt und alle Daten beinhaltet. Ein EDW kann auch der Datenlieferant für ein einzelnes DW sein. (Lusti, 1999, S. 131) DWS ist der Oberbegriff für alle an einem EDW beteiligten Prozesse, Technologien und Aufgabenstellungen. Für DWS gibt es eine Referenzarchitektur in der Literatur, die alle erforderlichen Komponenten beinhaltet. (Navrade, 2008, S. 14)

Dieses Referenzmodell sieht einen Data Warehouse Manager vor, welcher zur Steuerung des Data-Warehousing-Prozesses eingesetzt wird. (Navrade, 2008, S. 17) Dieser Prozess extrahiert die Daten aus einer Datenquelle in einen Staging-Bereich (extract), wo sie mithilfe von Datentransformationen nachbearbeitet werden (transform). Anschließend werden die Daten in das Data Warehouse geladen (load), wo sie für die Benutzerschnittstelle (z.B. OLAP) bereitgestellt werden. (Navrade, 2008, S. 15) (Gluchowski, Gabriel, & Dittmar, 2008, S. 132). Dieses Verfahren wird im Detail in Kapitel 2.4 „Extract-Transform-Load“ beschrieben.

2.2.1 Themenorientierung (subject orientation)

Da das Data Warehouse ein langfristig angelegtes Werkzeug darstellt, darf dessen Datenstruktur nicht abhängig von den Vorsystemen sein. Die Themenorientierung versucht diese Entkopplung zu schaffen, indem das Datenbankdesign sehr realitätsnahe und abteilungs-, prozess- sowie in Konzernen sogar firmenübergreifend sämtliche Informationen zu einem Thema, beispielsweise einem Kunden, zusammenfasst. (Silvers, 2008, S. 13) Dies ermöglicht den Blick auf das Gesamtbild. (Nagabhushana, 2006, S. 27)

2.2.2 Beständigkeit (non-volatility)

Das Data Warehouse muss für eine bestimmte Abfrage zu beliebigen Zeitpunkten immer dieselben Werte liefern – die Daten dürfen nie verändert, überschrieben oder gelöscht werden. Wenn es Änderungen in den Quelldaten gibt, so müssen diese nachvollziehbar in das Datawarehouse überführt werden. Vgl. dazu 2.2.4 Zeitorientierung (time variance) (Navrade, 2008, S. 30)

2.2.3 Vereinheitlichung (integration)

Durch die Vielfalt an spezialisierten Datenanwendungen in einem Unternehmen geschieht es häufig, dass Daten, die dasselbe beschreiben, in verschiedenen Systemen mit unterschiedlichen Informationsständen vorkommen. Beispielsweise kann es sein, dass ein Kunde einer Bank in der Anwendung zur Verwaltung von Krediten eine andere Kundennummer und eine andere Adresse hinterlegt hat, als in der Anwendung zur Verwaltung von Bankkonten. Der Kundenbetreuer ist dann derjenige, der über das Wissen verfügt, dass es sich trotzdem um dieselbe Person handelt, er aber in den unterschiedlichen Systemen mit unterschiedlichen Kundennummern nach derselben Person suchen muss.

Ein Data Warehouse hat die Aufgabe, Daten aus mehreren operativen Systemen zusammenzuführen und damit einhergehend genau solche Dubletten zu vermeiden. Das bedeutet, dass in der Tabelle „Kunden“ dieser Kunde nur einmal vorkommen darf und über

bestimmte Attributsätze wie beispielweise Vorname+Nachname+Sozialversicherungsnummer oder E-Mail-Adresse Dubletten erkannt werden. (Nagabhushana, 2006, S. 26)

Zusätzlich müssen alle Daten vereinheitlicht werden, wozu man in drei Arten der Vereinheitlichung unterscheidet. Die **Form** gibt an, dass dieselben Daten immer auf dieselbe Art und Weise abgespeichert werden. Beispiele dafür wären der Datentyp von Währungsfeldern, die Formatierung von Telefonnummern oder in welcher Reihenfolge beziehungsweise mit welchem Trennzeichen Vor- und Nachname in einem Feld dargestellt werden. Die **Funktion** von Daten muss ebenfalls vereinheitlicht werden. Beispielsweise verwendet die Schadensabwicklung den Code 32B um anzugeben, dass ein repariertes Teil am Rückweg zum Kunden ist, die Buchhaltung jedoch verwendet denselben Code zur Markierung von unausgeglichenen Konten. Diese Codes müssen in eine einheitliche, unternehmensweite Sprache übersetzt werden. Auch muss die **Einheit**, in der Mengen gemessen werden, vereinheitlicht werden. Im Einkauf wird der Artikel „Mineralwasser“ in Kisten gemessen, im Verkauf geht es immer um einzelne Flaschen. Diese drei Gruppen der Datenintegration müssen in unternehmensweit eindeutigen Daten resultieren. (Silvers, 2008, S. 14)

2.2.4 Zeitorientierung (time variance)

Sämtliche Daten werden in einem Data Warehouse mit einem Zeitbezug gespeichert. Dies bedeutet, dass aus einem Data Warehouse nicht nur Abfragen wie „Wie hoch war der Umsatz am Standort Kapfenberg im Jänner?“ oder „Wie hoch war der durchschnittliche monatliche Umsatz am Standort Kapfenberg?“, sondern auch Abfragen wie „Wie hoch war der durchschnittliche monatliche Umsatz am Standort Kapfenberg, während Max Mustermann die Standortleitung innehatte?“ möglich sind. Es gibt in der Realität selten Ereignisse, wie den Wechsel eines Standortleiters, die mit dem Wechsel des Kalendermonats zusammenfallen. Damit ein Analyst nicht erst die Zeiträume der einzelnen Standortleiter in Kapfenberg heraussuchen muss, um dann die Umsätze im richtigen Zeitraum abfragen zu können, speichert das Datawarehouse diese Dinge in Form von Dimensionen. (Silvers, 2008, S. 16ff) Dadurch ist es auch möglich, Trends zu erkennen und diese auf geänderte Rahmenbedingungen zurückzuführen. (Navrade, 2008, S. 14)

In jedem DW gibt es eine Zeitdimension, um einfache Operationen auf Basis des Datums durchführen zu können. Außerdem ist es damit möglich, auch ohne SQL⁴-Kenntnisse beispielsweise die Jahreszahl aus einem Datum zu extrahieren. (Kimball & Ross, 2013, S. 79f)

⁴ Structured Query Language: die Abfragesprache für relationale Datenbanksysteme (RDBMS)

Der Primärschlüssel in der Zeitdimension wird in der Regel als Ganzzahl im Format JJJJMMTT angegeben, damit dieser auch selbst aussagekräftig ist und einfache JOIN-Operationen⁵ ermöglicht. (Kimball & Ross, 2013, S. 101f)

2.3 Datenmodell

Um die Ziele eines Data Warehouse erreichen zu können, muss das Datenmodell gut durchdacht sein. Kimball/Ross beschreiben diesen Prozess als „dimensionale Modellierung.“ OnLine Transactional Processing (OLTP) Datenbankmodelle, also die Datenbankmodelle der operativen Datenanwendungen, sind darauf ausgelegt, Redundanzen zu vermeiden und mithilfe von Beziehungen unter vielen Tabellen, die Daten miteinander zu verknüpfen. Dies passiert, indem eine dreistufige Normalisierung durchgeführt wird, welche im Folgenden näher erläutert wird.

2.3.1 Normalformen operationaler Datenbanken

In der ersten Normalform wird sichergestellt, dass in einem Datenfeld nur ein atomarer, sprich ein einziger Wert vorkommt. Bei einer Tabelle, die Einkaufsdaten speichert, wäre es notwendig, pro gekauftem Artikel in einem Einkauf, eine eigene Zeile anzulegen. Dies resultiert automatisch im Kopieren von identen Werten, wie beispielsweise dem Namen des Kunden oder dem Bestelldatum. Hier kann die Kombination aus Kundennamen, Bestelldatum und Produkt der Primärschlüssel sein. In der zweiten Normalform wird sichergestellt, dass es einen eindeutigen Primärschlüssel gibt, indem dieser gegebenenfalls als laufende Nummer eingeführt wird. Durch die eindeutige Markierung jeder Zeile mit der laufenden Nummer wird erkennbar, dass das weiterkopieren von Kunden und Produkt mit dem vollen Namen eine starke Redundanz darstellt, die nicht notwendig wäre. In der dritten Normalform werden all diese nicht vollfunktional abhängigen Attribute in eigene Tabellen ausgliedert und dort mit neu eingeführten Primärschlüsseln identifiziert. Dadurch gibt es keine Redundanzen. (Schicker, 2017, S. 58ff)

In dieser dritten Normalform, auch 3NF genannt, werden in der Regel alle Daten von operativen Datenbanksystemen gespeichert. Dieses 3NF-Modell wird meistens durch ein Entity-Relationship-Diagramm dargestellt, das die einzelnen Entitäten – in der Datenbank als Tabellen ausgeführt – und deren Beziehungen untereinander abbildet. In einem Data Warehouse ist das 3NF-Modell jedoch nicht zielführend, da dies zu komplexen JOIN-Operationen in der Datenabfrage führt, was nicht sehr benutzerfreundlich ist, Abfragen langsamer macht und

⁵ JOIN-Operationen verbinden die Daten aus mehreren Tabellen, um aus den Fremdschlüsseln wieder menschenlesbare Daten zu machen.

darüber hinaus zu einer sehr komplexen Historisierung der Daten führt. Daher werden Data Warehouse Datenbanken nach dem dimensionalen Modell aufgebaut. Dieses dimensionale Modell von Data Warehouses kann ebenfalls mit ER-Diagrammen dargestellt werden, es gibt aber weitaus weniger Entitäten und Beziehungen. Es beinhaltet auch dieselben Informationen wie das normalisierte Modell, aber die Daten werden in Paketen in ein Format gebracht, das für den Benutzer eine einfache Abfrage ermöglicht und die Abfragegeschwindigkeit steigert. Das dimensionale Modell wird auch „Sternschema“ genannt, da das ER-Diagramm einem Stern ähnelt, in dessen Mitte eine Faktentabelle und um diese herum Dimensionen gespeichert werden. (Kimball & Ross, 2013, S. 7f)

2.3.2 Dimensionales Modell

Die Dimensionen sind stark an handgeschriebene oder in Excel verwaltete Tabellen erinnernde Daten, die einen hohen Grad an Redundanzen aufweisen. Sie haben eher den Charakter von Stammdaten, die sich nicht häufig ändern und enthalten beispielsweise Informationen über Mitarbeiter wie deren Name, Sozialversicherungsnummer oder Wohnort. Solche Daten ändern sich zwar nicht oft, es ist aber dennoch möglich. Um diese Änderungen dennoch abbilden zu können, gibt es den Fachbegriff „Slowly Changing Dimension“ (SCD), welcher nachfolgend erklärt wird.

2.3.3 Slowly Changing Dimensions

Man unterscheidet zwischen unterschiedlichen Strategien, wie man diese seltenen aber doch vorkommenden, Änderungen im dimensionalen Datenmodell abbilden kann, um zu jedem Zeitpunkt korrekte Auswertungen – und auch gleiche Ergebnisse für einen gleichen Zeitraum zu unterschiedlichen Abfragezeitpunkten – zu erhalten. Die entsprechende SCD-Strategie wird nicht für eine gesamte Dimension, sondern pro Dimensionsattribut festgelegt, was unterschiedliche SCD-Techniken innerhalb einer Dimension nicht selten macht. Diese unterschiedlichen Strategien werden in die folgenden sieben Typen eingeteilt:

Typ 0 behält den originalen Wert. Dies ist nützlich für Attribute, die „original“ im Namen haben und könnte beispielsweise die ursprüngliche Kreditbewertung eines Kunden beinhalten. Auch die Attribute einer Zeitdimension werden meist dem Typ 0 entsprechen, da sich diese nicht ändern.

Typ 1 überschreibt die Daten. Dies ist zwar einfach zu realisieren, bringt aber das Problem, dass ein und dieselbe Auswertung über denselben Zeitraum zu unterschiedlichen Zeitpunkten unterschiedliche Ergebnisse liefern kann; auch bei einem OLAP-Cube können dadurch zusätzliche administrative Aufgaben entstehen.

Typ 2 legt immer eine neue Zeile an. Dies ist der häufigste Typ von SCD und erfordert, dass drei weitere Spalten in die Dimension eingefügt werden, um ein einfaches Abfragen zu ermöglichen: „Gültig ab“, „Gültig bis“ und „Ist Gültig“. Die ersten beiden Spalten sind vom Typ DateTime, da sie ein Datum darstellen, das nicht direkt ein Dimensionsattribut ist – andere Felder mit dem DateTime-Datentyp referenzieren immer auf die Zeitdimension. „IstGültig“ ist ein Indikator, der angibt, ob eine Zeile die aktuell gültige Version ist. Es darf daher pro natürlichem Schlüssel immer nur eine Zeile das Attribut „IstGültig“ auf „ja“ haben.

Bei Typ 3 gibt es ein zusätzliches Attribut in der Dimension, welches eine „alternative Realität“ repräsentiert; dadurch sind „was-wäre-wenn“-Analysen sehr einfach möglich.

Die Typen 4 bis 7 sind sehr selten und für Spezialfälle oder sehr große Datenmengen konzipiert, wobei man hier von einem Ausmaß von mehreren Millionen Zeilen spricht, die sich häufig ändern.

(Kimball & Ross, 2013, S. 53ff)

2.3.4 Surrogate Key

Dadurch, dass diese SCD für einen einzelnen Datensatz möglicherweise mehrere Versionen bereitstellt, gibt es in der Dimensionstabelle mehrere Einträge zu ein und demselben Primärschlüssel der Quelldatenbank. Da ein Primärschlüssel aber eindeutig sein muss, wird der Begriff Surrogate Key oder Ersatzschlüssel eingeführt. Dieser Schlüssel ist ein im Data Warehouse frei erfundener Schlüssel, der üblicherweise eine einfache laufende Nummer ist, die bei 1 beginnt. Der ursprüngliche Primärschlüssel aus der Datenquelle wird als natürlicher Schlüssel oder natural key bezeichnet und in der Dimension mitgeführt. (Kimball & Ross, 2013, S. 46f)

2.3.5 Faktentabellen

Wie in Kapitel 2.3.1 „Normalformen operationaler Datenbanken“ beschrieben, steht im Zentrum des Sternschemas eine Faktentabelle. Diese stellt einen Geschäftsprozess, wie beispielsweise Buchungen auf einem Konto oder Änderungen in einem CRM-System, dar. Andere Faktentabellen können auch aus den Dimensionen abgeleitet werden, wie etwa die Anzahl an Mitarbeitern im Unternehmen zu einem bestimmten Zeitpunkt. Dies liegt in der Natur dieser Daten: Einerseits werden Mitarbeiterstammdaten zwar als Dimensionen gesehen, die sich möglicherweise langsam ändern (SCD), andererseits ist auch jede dieser seltenen Änderungen immer das Resultat eines Geschäftsprozesses und damit einer Transaktion – beispielsweise der Beginn eines Dienstverhältnisses oder das Einbringen der Wohnortänderung eines Dienstnehmers. (Kimball & Ross, 2013, S. 43f)

Streng genommen sind diese Transaktionen aber „faktenlose Faktentabellen“, da sie keine messbare Größe enthalten, sondern nur die Information, dass sich etwas geändert hat und was sich geändert hat. Für die einfachere Handhabung in Auswertungen ist es von Vorteil, einen Zähler (COUNTER) mitzuführen – eine Spalte, die immer nur den Wert 1 beinhaltet. Damit kann die Abfrage von einem umständlichen COUNT(DISTINCT DIMENSIONKEY), das die Anzahl an unterschiedlichen Dimensionswerten zählt, zu einem SUM(COUNTER) vereinfacht werden, das einfach diese 1-Werte aufsummiert, wenn im Zuge der Abfrage nach bestimmten Merkmalen gruppiert wird (GROUP BY). Ein reines COUNT(DIMENSIONKEY) würde bei einem GROUP BY zu einem falschen Wert führen, da es alle DIMENSIONKEY zählt, die an der Abfrage beteiligt waren und nicht nur jene, die im Zuge des GROUP BY in unterschiedliche Gruppen unterteilt wurden. (Kimball & Ross, 2013, S. 331f)

2.4 Extract-Transform-Load

ETL steht für „Extract-Transform-Load“ und beschreibt die drei Schritte, die notwendig sind, um Daten aus den Quellsystemen in das Datawarehouse zu überführen. Diese einzelnen Schritte werden auf den nächsten Seiten detailliert erläutert. (Wang, 2009, S. 573ff)

2.4.1 Extract

Beim ersten Schritt „Extract“ werden die Daten aus dem Quellsystem ausgelesen und zwischengespeichert. Dieser Schritt wird durchgeführt, um die Last für die Ausführung des nachfolgenden Schrittes „Transform“ vom produktiven Datenbankserver zu nehmen und diese stattdessen in Richtung Data Warehouse zu verfrachten. Außerdem ist dies der einzige Schritt, der sich um die Schnittstelle zu den operativen Datenquellen kümmern muss; alle weiteren Aufgaben im ETL-Prozess können dann autonom und ohne Abhängigkeiten zu den Ausgangssystemen durchgeführt werden. Diese „Staging-Area“ ist ein Konstrukt aus vielen Tabellen, deren Struktur grundsätzlich jener der Datenquelle entspricht und ermöglicht es dem nachfolgenden „Transform“, datenverändernde Operationen auf diesen Tabellen durchzuführen, wie beispielsweise das Normalisieren von Adressdaten mithilfe einer zusätzlichen externen Datenquelle. (Kimball & Ross, 2013, S. 450) Das Beibehalten der originalen Datenstruktur hilft auch dabei, die Komplexität während dem Entwickeln des Datenflusses gering zu halten sowie die Last auf die Datenquellen zu verringern. (de Ruiter, 2012)

Eine weitere essentielle Aufgabe des Extract-Prozesses ist das Herausfiltern von Änderungen an den Quelldaten. Dies dient dazu, die bewegte Datenmenge und die Ausführungsdauer zu verringern und damit eine hohe Ausführungshäufigkeit zu ermöglichen. Je nach Datenquelle ist es einfach, die geänderten Daten zu identifizieren, indem eingefügte, geänderte und gelöschte

Daten als solche markiert werden. In den meisten operativen Datenbanken wird jedoch nur der aktuelle Datenstand gespeichert und diese Veränderungen müssen auf Basis der letzten Datenextraktion berechnet werden. (Jensen, Pedersen, & Thomsen, 2010, S. 68)

Es gibt drei Varianten zur Datenextraktion aus den Quellsystemen: statisch, unmittelbar oder verzögert. Die statische Datenextraktion liest sämtliche Daten aus dem Quellsystem aus und ist insbesondere bei der ersten Beladung des Data Warehouse relevant. Die unmittelbare Datenextraktion wird entweder mithilfe von eigenen Programmierungen in der Quellenanwendung, Datenbanktriggern oder durch das Abgreifen der Transaktionsprotokolle realisiert. Datenbanktrigger sind kleine Programme, die innerhalb eines Datenbanksystems auf bestimmte Ereignisse reagieren und weitere Aktionen ausführen. Im Rahmen von ETL werden die Trigger so programmiert, sämtliche Änderungen zusätzlich an eine andere Stelle zu schreiben, damit sie vom Extract-Prozess ausgelesen werden können. Transaktionsprotokolle werden von Datenbanksystemen im Rahmen der Datensicherung mitgeschrieben und beinhalten ebenfalls sämtliche Datenänderungen. Die verzögerte Datenextraktion kann entweder mithilfe von Zeit- und Datumsstempeln auf den Datensätzen oder durch den Vergleich des aktuellen Stands des Quellsystems mit dem aktuellen Stand des Data Warehouse geschehen. (Ponniiah, 2010, S. 287-294)

2.4.2 Transform

Der zweite Schritt „Transform“ wandelt die Daten des Quellsystems in die Struktur und das Format des Data Warehouse um. Zusätzlich hat dieser Schritt die Aufgabe, Daten zu bereinigen und mit fehlerhaften Daten umzugehen. Auch das Umschlüsseln von Primärschlüssel auf Ersatzschlüssel ist eine Aufgabe des „Transform“. (vgl. 2.3.4 Surrogate Key) (Jensen, Pedersen, & Thomsen, 2010, S. 68f)

Insgesamt kann davon gesprochen werden, dass der Schritt Transform die Aufgabe hat, die Eigenschaften „integration“ und „time variance“ des DW sicherzustellen. (vgl. 2.2 Data Warehouse)

2.4.2.1. Datenvereinheitlichung

Die Integration der Daten kümmert sich um unterschiedliche Probleme mit den Quelldaten. Zunächst gilt es, Dubletten zu finden und zu eliminieren. Dazu ist es in der Regel notwendig, sich mit den Daten vertraut zu machen und Attributsätze zu finden, die eine eindeutige Identifizierung ermöglichen. Ein Attribut zur Identifizierung von Dubletten bei Personen könnte die Sozialversicherungsnummer oder die E-Mail-Adresse von Angestellten eines Unternehmens sein. (Chen, 2001, S. 78) Konkret kann die Deduplizierung mithilfe der SQL-Funktion

ROW_NUMBER() erledigt werden, indem nur die jeweils erste Zeile, die einer Gruppierung nach den zuvor gefundenen Attributsätzen entspricht, ausgewählt wird. (Aspin, 2013, S. 484)

Eine weitere Aufgabe der Vereinheitlichung ist, unterschiedliche Darstellungen desselben Werts zu koordinieren. Beispielsweise könnte das Quellsystem implizieren, dass es sich bei allen Geldbeträgen um Euro handelt und die Währung daher nicht extra abspeichern. Im Data Warehouse wird die Währung extra gespeichert, um unterschiedliche Währungen unterstützen zu können. Aufgrund von unterschiedlichen Programmierparadigmen kann es auch sein, dass in einem System die lokale Uhrzeit gespeichert wird, in einem anderen jedoch die Weltzeit (UTC). Beim Transform-Prozess müssen diese unterschiedlichen Darstellungsweisen auf einen gemeinsamen Nenner gebracht werden, um die Integration der Daten aus mehreren Quellsystemen zu ermöglichen.

2.4.2.2. Anpassung an dimensionales Datenmodell

Wie bereits in Kapitel 2.3.2 „Dimensionales Modell“ erläutert, ist die Datenstruktur im DW eine andere als in den Datenquellen. Daraus resultieren mit hoher Wahrscheinlichkeit auch andere Namen für Tabelle und Spalten, die im Zuge des „Transform“ adaptiert werden müssen. Da die Daten beim Extract in die Staging-Area kopiert wurden, ist es einfach, nun dort Datenbankviews anzulegen, die diese Transformation durchführen – die Namen der Views geben die Namen der Zieltabelle an, die neuen Spaltennamen werden ebenfalls berücksichtigt. Die zuvor erwähnten Vereinheitlichungen von Mengenangaben, Zeitstempeln und Währungsbeträgen können ebenfalls bereits in diesen Views durchgeführt werden. Dieses Mapping kann alternativ auch mithilfe des ETL-Tools (vgl. Kapitel 3.1.2 „SQL Server Integration Services“) durchgeführt werden.

Mithilfe von JOIN-Operationen kann an dieser Stelle auch die Denormalisierung für das dimensionale Datenmodell durchgeführt werden. Dadurch wird die notwendige Duplizierung der Attributwerte zunächst nur als Definition durchgeführt, deren Ergebnis sehr einfach kontrolliert und bei Fehlern adaptiert werden kann. Dies erleichtert das Entwickeln des ETL-Prozesses enorm und spart darüber hinaus Rechenzeit und Speicherplatz. Diese Datenanreicherung kann alternativ ebenfalls mithilfe des ETL-Tools abgewickelt werden.

(de Ruiter, 2012)

2.4.2.3. Fehlerbehebung

In operationalen Datenbankapplikationen gibt es oftmals unzureichende Datenvalidierungsregeln, die Fehleingaben oder fehlende Eingaben erlauben. Je nach Relevanz dieser Felder für das DW muss entschieden werden, wie mit diesen Fehlern umgegangen wird.

Häufig werden Felder, die den Mitarbeitern vor Ort als nicht sinnvoll erscheinen, bei denen jedoch die Applikation zu einer Eingabe zwingt, mit einem beliebigen Wert befüllt. Wenn dies passiert, ist zu entscheiden, ob es überhaupt Sinn macht, das Feld in das DW aufzunehmen oder nicht.

Andere Daten, die bereinigt werden müssen, sind beispielsweise Zeitraumangaben. Unter Umständen gibt es in einer Datenquelle die Möglichkeit, einzelne Datensätze mit den Feldern „Gültig Ab“ und „Gültig Bis“ zu befüllen, durch eine unvollständige Eingabevalidierung in der Applikation sind jedoch überlappende Zeiträume entstanden, die bei einer logischen Betrachtung der Datenstruktur und der Daten selbst nicht auftreten dürften – auch diese Fehler müssen im Zuge des „Transform“ identifiziert und bereinigt werden. Identifiziert werden solche überlappenden Zeiträume, indem die Tabelle mithilfe einer JOIN-Operation mit sich selbst verglichen wird. Ein überlappender Eintrag existiert dann, wenn das „Gültig Ab“ eines Eintrags A kleiner oder gleich dem „Gültig Bis“ eines anderen Eintrags B ist und zusätzlich das „Gültig Bis“ des Eintrags A größer oder gleich dem „Gültig Ab“ des Eintrags B ist. Diese überlappenden Zeiträume müssen entweder im Zuge des „Transform“ automatisch bereinigt werden, indem beispielsweise das Enddatum des vorangehenden Zeitraums auf den Tag vor den Beginn des zweiten Zeitraums gesetzt wird. Alternativ kann als Strategie gewählt werden, den ETL-Prozess abubrechen und einer für die entsprechenden Daten verantwortlichen Person ein E-Mail mit Informationen über die fehlerhaften Datensätze zu schicken. Dadurch könnten diese auch gleich im Quellsystem richtiggestellt werden.

Um etwaigen falschen Berechnungen bei Gruppierungen oder Filtern in den späteren Auswertungen vorzubeugen, ist es außerdem vorteilhaft, NULL-Werte – also leere Datenbankfelder – insbesondere bei Dimensionen durch aussagekräftige Werte zu ersetzen, beispielsweise durch den Text „unbekannt“. Dadurch ist sichergestellt, dass Auswertungen richtige Werte liefern. Außerdem erhöht dies die Klarheit für Endanwender. (Kimball & Ross, 2013, S. 48)

2.4.3 Load

Im dritten und letzten Schritt werden die Daten dann in die Struktur des Data Warehouse geladen („load“).

Wie in den Kapiteln 2.3.3 Slowly Changing Dimensions sowie 2.3.4 Surrogate Key erwähnt, ist es notwendig, die natürlichen Schlüssel der Datenquellen beizubehalten und zusätzlich einen Ersatzschlüssel einzuführen, damit auch nach der Vervielfachung des natürlichen Schlüssels eine einzelne Zeile der Dimension eindeutig identifizierbar ist und in einer Faktentabelle referenziert werden kann. Dies ist Aufgabe des „Load“-Schrittes. Dazu kann entweder ein MERGE INTO-SQL Statement herangezogen werden oder auch mit dem ETL-Tool das

entsprechende Datenmatching und Berücksichtigen der SCD durchgeführt werden. (Thorntwaite, 2008)

Nach dem Beladen kann bereits auf die Daten im DW zugegriffen werden, indem beispielsweise mithilfe von PowerPivot eine Excel-Arbeitsmappe mit dem DW verknüpft wird. Allerdings ist hier noch kein gravierender Geschwindigkeitsvorteil gegenüber Abfragen aus den einzelnen Datenquellen spürbar – für diesen Benefit wird OnLine Analytics Processing (OLAP) herangezogen.

2.5 OnLine Analytics Processing

OLAP ist eine Technologie zur Verarbeitung der potentiell sehr großen Datenmengen im Data Warehouse. Es ist darauf spezialisiert, Daten nach bestimmten Kriterien zu filtern und zu aggregieren, um KPIs zu berechnen. Die Besonderheit von OLAP im Vergleich zu traditionellen Auswertungen aus einem Anwendungsprogramm ist, dass es sehr großen Datenmengen sehr schnell verarbeiten kann. Dazu werden die Daten in einem so genannten „Cube“ zu mehrdimensionalen Würfeln zusammengefasst.

(Wang, 2009, S. 540f)

2.5.1 Cube

Streng genommen sind diese „Cubes“ keine Würfel, da sie mehr als drei Dimensionen aufweisen und heißen deshalb eigentlich „Hypercubes“. Eine Excel-Tabelle verfügt über zwei Dimensionen: Zeilen und Spalten. Beispielsweise könnte das Verkaufsdatum – also die Zeitdimension – als Zeilen, und die verkauften Stück sowie der Verkaufspreis – also die Variablen – als Spalten hinterlegt werden. Diese Daten enthalten keine Information über das verkaufte Produkt – dies stellt eine dritte Dimension dar. Nimmt man also die Information dazu, welches Produkt verkauft wurde, hat man einen Würfel – die x-Achse für die Variablen, die y-Achse für die Zeit und die z-Achse für das Produkt. An jedem Punkt im Würfel ist genau ein Wert vorhanden – nämlich der entsprechende Wert der Verkäufe eines bestimmten Produktes zu einem bestimmten Zeitpunkt, vgl. Abbildung 2. Diese Werte werden als „Measures“ bezeichnet und im Anschluss an dieses Kapitel näher erläutert.

Einen OLAP-Cube oder Hypercube kann man sehr einfach um weitere Dimensionen erweitern, beispielsweise um die Information, wo das Produkt verkauft wurde – diese 4 Dimensionen sind bereits nicht mehr visualisierbar, aber genau das Spezialgebiet von OLAP. Aus diesem Beispiel mit dem klassischen dreidimensionalen Würfel kommen auch die Begriffe für die Methoden, mit

deren Hilfe OLAP-Analysen durchgeführt werden. Dies wird in Kapitel 4.1 OLAP ausführlich erläutert. (Thomsen, 2002, S. 74)

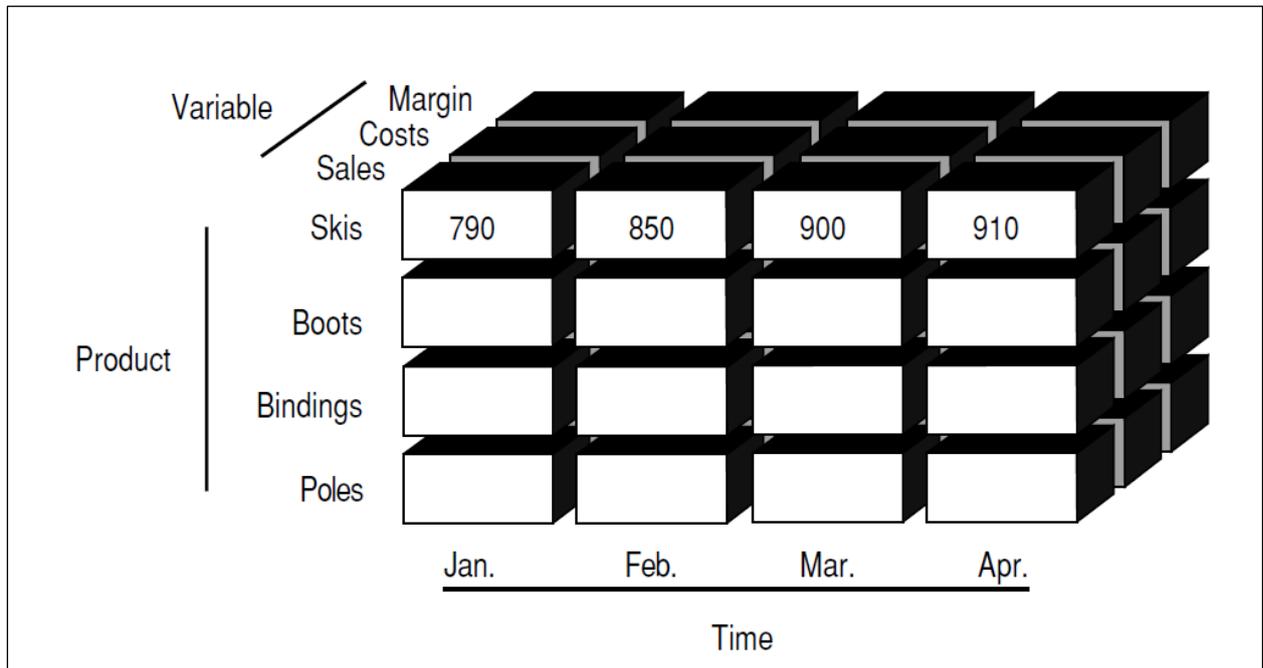


Abbildung 2 Dreidimensionaler Würfel mit Zeit, Produkten und Variablen. Quelle: (Thomsen, 2002, S. 49)

2.5.2 Measure

Measures sind Zahlenwerte in Faktentabellen, auf die Operationen wie Summe (SUM) oder die Durchschnittsberechnung (AVG) ausgeführt werden können. Dies könnten in einer Produktverkaufsstatistik die Stückzahl der verkauften Produkte oder der Verkaufspreis sein. In OLAP haben Measures insofern eine besondere Bedeutung, als dass diese Attribute im Cube als solche markiert werden. Dadurch werden die erwähnten Operationen SUM und AVG vom Cube vorberechnet und die Ergebnisse stehen sehr schnell bereit, auch bei einer großen Anzahl an Daten.

(Rainardi, 2008, S. 378)

Da in sehr großen Umgebungen das EDW sowie die Cubes sehr große Datenmengen beinhalten, von denen oftmals nur ein Ausschnitt für einen Anwendungsfall einer Abteilung relevant ist, werden diese Ausschnitte mithilfe von Datamarts gezielt zur Verfügung gestellt.

2.5.3 Datamart

Ein EDW enthält sämtliche Daten des gesamten Unternehmens. Um die richtigen Abteilungen mit den richtigen Daten mitsamt speziell auf sie zugeschnittene Aggregationen und Transformationen zu versorgen, werden Datamarts verwendet. Sie enthalten eine Kopie eines

Teils des EDW mit genau den richtigen Daten und werden als OLAP-Würfel bereitgestellt. Ein weiteres Ziel von Datamarts ist das Entlasten des Data Warehouse in großen Umgebungen. (Krishnan, 2013, S. 150) (Jensen, Pedersen, & Thomsen, 2010, S. 11)

Zur Abfrage eines Datamarts wurde eine eigene Abfragesprache namens „MDX“ entwickelt, welche die Besonderheiten dieses Datenformats berücksichtigt.

2.5.4 MDX

MDX steht für Multidimensional Expressions und ist die Abfragesprache in OLAP, die SQL stark ähnelt. Im Gegensatz zu SQL, das zwar ebenfalls für die Massenabfrage von Daten geeignet ist, liegt der Fokus bei MDX darauf, optimal mit dem Stern-Schema (vgl. 2.3.2 „Dimensionales Modell“) umzugehen. Das bedeutet, dass statt dem Abfragen einer Tabelle, die mit vielen anderen Tabellen mittels JOIN vereint wird, eine Faktentabelle abgefragt wird und die zugehörigen Dimensionen einfach verwendet werden können, da im Cube die Informationen über die Zusammenhänge der Daten hinterlegt sind. Diese MDX-Abfragen können mithilfe des entsprechenden Tools zur Bereitstellung von Cubes sehr einfach grafisch zusammengestellt werden (vgl. 3.1.4 SQL Server Data Tools). (Rainardi, 2008, S. 365, 435) Darüber hinaus ist es mit MDX sehr einfach möglich, Zeilen und Spalten dynamisch auszugeben sowie miteinander zu vertauschen. Beispielsweise können die unterschiedlichen Produkte sehr einfach als Spalten im Ergebnissatz dargestellt werden, was mit SQL sehr komplex wäre. (Gluchowski, Gabriel, & Dittmar, 2008, S. 186ff)

3 WERKZEUGE & UMGEBUNG

Als Auswahlhilfe für die Technologien in der Umsetzung des DW wurde in erster Linie der „Magic Quadrant for Data Warehouse Database Management Systems“ von Gartner herangezogen. Dort liegt Microsoft mit dem SQL Server sowie den SQL Server Data Tools und weiteren Technologien im Bereich der „leaders“ und gleichzeitig im Bereich der „visionaries“, vor Sybase aber nach Netezza, IBM, Oracle und Teradata. (Feinberg & Beyer, 2010)

Aufgrund der Nähe der JAW IKT zu Microsoft Produkten und den bereits vorhandenen Lizenzen wird im weiteren Verlauf dieser Masterarbeit auf Microsoft SQL Server und die dazugehörigen Tools gesetzt.

Das folgende Kapitel beschäftigt sich mit der Beschreibung dieser Werkzeuge sowie des Datenschutzes als essentielle Rahmenbedingung für die Erstellung eines DW.

3.1 Werkzeuge

Zur Realisierung eines DW sind viele Schritte notwendig – was auch in unterschiedlichen Softwareprodukten resultiert, die diese einzelnen Schritte realisieren. SQL Server ist der Datenbankserver von Microsoft, der einerseits die meisten operationalen Datenquellen sowie auch das Data Warehouse selbst beheimatet, andererseits stellt er auch die Grundlage für die gesamte BI-Plattform dar, d.h. die meisten nachfolgenden Werkzeuge sind Teil der SQL Server Produktpalette. Die SQL Server Integration Services (SSIS) sind die Automatisierungsplattform für den ETL-Prozess. Die SQL Server Analysis Services (SSAS) sind die Analyseplattform für OLAP-Datenwürfel. Die SQL Server Data Tools sind eine Zusammenstellung von Visual Studio Werkzeugen, um SSIS- und SSAS-Projekte zu erstellen und bereitzustellen. Microsoft Excel ist nicht mehr nur die Tabellenkalkulation, sondern auch der Business Intelligence Client von Microsoft. Power BI ist die Plattform von Microsoft, die Self-Service BI ermöglicht.

Nachfolgend werden diese Werkzeuge näher erläutert, sowie im weiteren Verlauf der Arbeit zur Erstellung eines Prototyp eingesetzt, vgl. dazu Kapitel 7.3 „Prototyp“.

3.1.1 SQL Server

SQL Server ist der Datenbankserver von Microsoft. Er wird für die Datenspeicherung vieler Applikationen eingesetzt und ist als relationales Datenbankmanagementsystem konzipiert. Als solches besteht die Struktur aus mehreren Ebenen, die hierarchisch aufgebaut sind und nachfolgend erklärt werden.

- Auf höchster Ebene steht die Datenbankinstanz; diese ist eine Datenbankserverumgebung, die Rahmenbedingungen wie den Speicherort der Datenbankdateien oder Konfigurationen wie die Speicher- oder Prozessornutzung festlegt.
- Die Datenbank ist eine Einheit innerhalb einer Datenbankinstanz, welche angibt, in welcher beziehungsweise wie viele Dateien die Daten abgespeichert werden oder wie die Datensicherung handzuhaben ist.
- Das Schema ermöglicht eine Gruppierung von tieferliegenden Objekten aus logischer oder sicherheitsrelevanter Perspektive innerhalb einer Datenbank.
- Die Tabelle ist eine Sammlung von Spalten und Zeilen, um einen bestimmten Anwendungsfall darzustellen – beispielsweise eine Entität in 3NF-Modellen bzw. eine Dimension oder eine Faktentabelle in dimensionalen Modellen.
- Eine Spalte legt fest, welche Art von Daten – also Zahlen, Textdaten, Binärdaten oder Datumsinformationen – gespeichert werden, welche Länge diese besitzen und welchen Namen das Feld trägt.
- Eine Zeile ist ein Eintrag in einer Tabelle und besitzt an verfügbaren Feldern genau jene, die durch die Spalten festgelegt werden. An dieser Stelle – in den einzelnen Feldern der Zeile – werden die konkreten Nutzdaten gespeichert.

Auf derselben Hierarchieebene wie Tabellen stehen „Views“, welche eine bestimmte „Sicht“ auf Daten in Datenbanken darstellen. Sie sind wiederverwendbare, abgespeicherte Datenbankabfragen, um die Verwendung von Abfragen zu erleichtern. Ebenfalls auf der selben Ebene befinden sich „Stored Procedures“, also „gespeicherte Prozeduren“, welche in der Microsoft SQL Sprache „Transact-SQL“ (T-SQL) geschriebene Programme innerhalb einer Datenbank darstellen. Diese bieten die Möglichkeit, komplexere Logiken und in ausgewählten Fällen auch Geschäftslogik abzubilden.

Innerhalb einer Tabelle kann man mithilfe von Indizes die Abfrage bestimmter Spalten beschleunigen oder auch eindeutige Werte erzwingen. Außerdem gibt es „Trigger“, also „Auslöser“, welche bei bestimmten Ereignissen auf einer Datenbanktabelle eine in T-SQL geschriebene Programmlogik ausführen. Ein Beispiel hierfür wäre die Validierung von Datumsbereichen, die sicherstellt, dass das „Bis-Datum“ nicht vor dem „Von-Datum“ liegt.

Zwischen mehreren Tabellen gibt es „Foreign Keys“, also „Fremdschlüssel“, welche die Beziehung zwischen diesen Tabellen beschreiben. Sie helfen einerseits bei der Weiterverarbeitung des Datenmodells in Client-Tools, damit diese die Beziehungen der Tabellen automatisch erkennen können, und andererseits stellen sie die referentielle Integrität sicher. Das

bedeutet, dass es zu einem Fremdschlüsselwert auch immer tatsächlich auch einen Eintrag in der entsprechenden „fremden“ Tabelle geben muss. Um das sicherzustellen, kann eine Einschränkung definiert werden, die dafür sorgt, dass übergeordnete Zeilen erst dann gelöscht werden können, wenn alle untergeordneten Zeilen gelöscht wurden oder der Lösch-Befehl an die untergeordneten Zeilen weitergegeben wird.

(Paragon Corporation, 2005) (Otey & Otey, 2006, S. 34ff) (Ben-Gan, et al., 2010, S. 1) (Ben-Gan, et al., 2010, S. 81) (Ben-Gan, et al., 2010, S. 145)

3.1.2 SQL Server Integration Services

SQL Server Integration Services (SSIS) ist eine umfangreiche Werkzeugsammlung in SQL Server, welche auf die Verarbeitung, Transformation und Übertragung von Daten spezialisiert ist. Im Umfeld eines DW ist SSIS im gesamten ETL-Prozess involviert. Mithilfe von SSIS werden die Daten aus unterschiedlichsten Datenquellen geladen, transformiert, angereichert und in das DW geschrieben. SSIS bietet eine hohe Flexibilität in der Konfiguration der einzelnen Aufgabenschritte und eine umfangreiche Fehlerbehandlung, was für einen produktiven Einsatz essentiell ist. Einfache SSIS-Pakete können einerseits mit dem SQL Server Management Studio (SSMS), dem Datenbankadministrationswerkzeug, erstellt werden und dienen meist administrativen Zwecken. Andererseits kommen für komplexere Aufgabenstellungen die SQL Server Data Tools (SSDT, vgl. Kapitel 3.1.4) zum Einsatz.

(Microsoft Corporation) (Rossum & Baccaro, 2015, S. 3)

SSIS ermöglicht auch „Scripting“-Tasks, mit deren Hilfe ein vollständiger C#-Code in ein SSIS-Paket eingebettet werden kann. (Rossum & Baccaro, 2015, S. 27)

3.1.3 SQL Server Analysis Services

SQL Server Analysis Services (SSAS) ist ein Modul von Microsoft SQL Server, das zur statistischen Analyse großer Datenmengen konzipiert ist. Es ist ein Produkt von Microsoft, das es ermöglicht, eine OLAP-Datenbank bereitzustellen, welche von Client-Tools abgefragt werden kann.

(Microsoft Corporation)

Bei SSAS gibt es seit SQL Server 2012 nicht nur das multidimensionale Modell, das im Hintergrund die Daten tatsächlich mehrdimensional speichert, sondern auch ein Tabular-Modell. Aus Client-Sicht sehen die bereitgestellten Daten nach wie vor wie in einem Cube aus, innerhalb von SSAS werden die Daten jedoch in Tabellenform gespeichert. Beide Formate werden mithilfe der SQL Server Data Tools (SSDT) entwickelt und bereitgestellt (vgl. Kapitel 3.1.4 „SQL Server

Data Tools“). Der gravierendste Unterschied zwischen einem mehrdimensionalen und einem tabellenbasierten SSAS-Projekt ist, dass man in SSDT entweder tatsächlich einen Cube erstellt (mehrdimensional) oder ein „Modell“ (tabellenbasiert). Es stehen aber annähernd die gleichen Funktionalitäten zur Verfügung, lediglich das Design-Tool hat einen etwas anderen Aufbau. Der Grund dahinter ist die Weiterentwicklung von Microsoft SQL Server, wodurch die tabellenbasierte Variante mit noch größeren Datenmengen umgehen kann, als das mehrdimensionale Modell. Zugunsten dieses tabellenbasierten Systems wird das mehrdimensionale Modell seitens Microsoft nicht mehr weiterentwickelt. Zusätzlich hat das Tabular-Modell die Möglichkeit, direkt auf die Quelldaten – die beispielsweise in einem DW gespeichert sind – zuzugreifen, ohne die einzelnen Tabellen erst noch verarbeiten zu müssen.

(Duncan & Guyer, 2017)

3.1.4 SQL Server Data Tools

Um SSIS- und SSAS-Projekte zu entwickeln, werden die SQL Server Data Tools (SSDT) verwendet. Diese sind eine auf Microsoft Visual Studio aufbauende Entwicklungsumgebung; auch SQL Server Datenbanken können damit entwickelt werden.

SSAS-Projekte bestehen aus Verbindungsmanagern und Paketen, wie in Abbildung 3 dargestellt. Die Verbindungsmanager stellen Verbindungen zu Datenquellen und -zielen her. Als Datenquelle können hier nicht nur sämtliche Datenbanken, für die es einen entsprechenden Treiber in Windows gibt, ausgewählt werden, auch HTTP, FTP, andere Web-Quellen, Excel-Dateien oder Text- bzw. CSV-Dateien sind möglich.

(Microsoft Corporation)

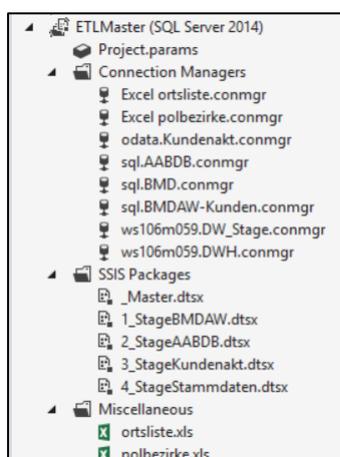


Abbildung 3 Aufbau eines SSIS-Projekts in SSDT. Quelle: eigene Darstellung (Microsoft SQL Server Data Tools)

In einem SSIS-Paket werden die Aufgaben festgelegt, die von diesem Paket durchgeführt werden sollen. Die Auswahl erfolgt mithilfe eines grafischen Design-Tools. Es kann sowohl

konfiguriert werden, dass die Aufgaben in einer bestimmten Reihenfolge nacheinander durchgeführt werden, als auch, dass die Aufgaben parallel ausgeführt werden können. Die zwei am häufigsten verwendeten Aufgaben sind „Execute SQL Task“ und „Data Flow Task“. Der „Execute SQL Task“ führt ein angegebenes T-SQL-Skript mithilfe des angegebenen Connection Managers aus. Dies ist nützlich, um beispielsweise Tabellen mithilfe von „TRUNCATE TABLE“ zu leeren oder auch Stored Procedures auszuführen. Der „Data Flow Task“ beschreibt den Datenfluss auf Attributebene von einer Datenquelle zu einem Datenziel und kann währenddessen Transformationen durchführen, mehrere Datenquellen zu einer vereinigen, Skripts ausführen und auch Fehlerbehandlungen durchführen; beispielsweise Zeilen, bei denen ein Skript fehlgeschlagen ist, in eine andere Tabelle umleiten, damit diese nicht verloren gehen können (vgl. Kapitel 3.1.2).

(Microsoft Corporation) (Rainardi, 2008, S. 191)

3.1.5 Excel / PowerPivot

Microsoft Excel ist das Tabellenkalkulationsprogramm in der Microsoft Office Produktpalette. Ursprünglich als Werkzeug zur computerunterstützten Verarbeitung von Tabellen entwickelt, ist es inzwischen zu einem Business Intelligence Werkzeug geworden. Anbindungen von externen Datenquellen, unter anderem SQL Server und Analysis Services, sind bereits seit vielen Versionen möglich. Allerdings stößt man bei großen Datenquellen schnell an die Grenzen der Berechnungskapazität für Microsoft Excel und ein Analysis Services Server mit bereitgestellten Cubes ist nicht immer vorhanden. Seit Excel 2010 steht mit PowerPivot ein mächtiges Werkzeug zur Verfügung, das mit sehr großen Datenmengen umgehen kann. Die Rohdaten sind hierbei nicht direkt in einem Excel-Arbeitsblatt verfügbar, sondern nur über den Button „PowerPivot“ erreichbar. Dort sind ebenfalls zahlreiche Datenquellen möglich. Durch das Einfügen einer PivotTable aus PowerPivot in das Excel-Arbeitsblatt können die Daten analysiert werden, was auch bei einer sehr großen Datenmenge sehr rasch vonstattengeht. Im PowerPivot-Fenster können mehrere Tabellen verwaltet werden und Beziehungen zwischen diesen Tabellen hergestellt werden. Auch das Hinzufügen neuer Spalten ist dort möglich. All dies kann dann in der PivotTable in Excel ausgewählt und ausgewertet werden. Durch den großen Anklang von PowerPivot wurde SSAS überarbeitet und das Tabular-Model eingeführt – es wird exakt gleich erstellt und hat im Hintergrund dieselbe Technologie wie PowerPivot.

(Russo & Ferrari, 2011, S. 1)

3.1.6 Power BI

Power BI ist einerseits ein Sammelbegriff für die Excel-Features PowerPivot, PowerQuery und PowerView. PowerPivot wurde bereits in Kapitel 3.1.5 erläutert. PowerQuery ist der Name für die Anbindung von Microsoft Excel an unterschiedlichste Datenquellen, die seit Excel 2013 direkt in das Programm integriert ist. PowerView ist eine interaktive Variante von PivotChart, welche eine bessere grafische Benutzerführung durch eine große Menge an Daten bietet. Andererseits ist Power BI auch die Cloud-Plattform von Microsoft, welche die rasche Visualisierung sowie den mobilen Zugriff auf die BI-Daten auch ohne Microsoft Excel ermöglicht. (Microsoft Corporation)

Damit der Cloud-Dienst Zugriff auf die Daten auf den unternehmenseigenen Servern hat, ist es notwendig, einen „On-premises data gateway“ zu installieren, der die Daten aus SSAS, einem DW oder anderen Datenquellen in der Cloud bereitstellt. (Iseminger, 2017)

3.2 Datenschutz

Da in einem DW Daten über Personen gespeichert werden, ist es wichtig, sich mit den rechtlichen Rahmenbedingungen auseinanderzusetzen, um einen gesetzeskonformen Umgang mit diesen Daten sicherzustellen.

Der Datenschutz ist in Europa ein Recht, das in einem völkerrechtlichen Vertrag von 1981 mit dem Titel „Übereinkommen zum Schutz des Menschen bei der automatischen Verarbeitung personenbezogener Daten“ (kurz: Europäische Datenschutzkonvention), den auch Österreich unterzeichnet hat, seinen Ursprung hat. Dieser Vertrag hat das Ziel, die Privatsphäre aller Menschen im europäischen Raum zu schützen, unabhängig davon, ob sie aus Europa stammen oder nicht. Hier entstand auch der Begriff „personenbezogene Daten“ für „jede Information über eine bestimmte oder bestimmbare natürliche Person („Betroffener“)“ (Bundeskanzleramt, 1981)

1995 wurde dieser Vertrag durch die EU-Richtlinie 95/46/EG „zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Datenverkehr“ ergänzt und detaillierter spezifiziert, was Datenschutz bedeutet. (Europäisches Parlament und Rat, 1995) Aus dieser Richtlinie entstand in Österreich das „Datenschutzgesetz 2000“ (DSG2000), das die österreichische Umsetzung dieser Richtlinie darstellt. Der erste Artikel trägt den Titel „Grundrecht auf Datenschutz“ – was den Status des Datenschutzes als Grundrecht in Österreich hervorhebt. Die ersten Worte des Gesetzes lauten: *„Jedermann hat, insbesondere auch im Hinblick auf die Achtung seines Privat- und Familienlebens, Anspruch auf Geheimhaltung der ihn betreffenden personenbezogenen Daten, soweit ein schutzwürdiges Interesse daran besteht“*. Das bedeutet, dass grundsätzlich jegliche personenbezogenen Daten schützenswert sind, sofern diese *„infolge ihrer allgemeinen Verfügbarkeit oder wegen ihrer mangelnden Rückführbarkeit auf den*

Betroffenen einem Geheimhaltungsanspruch nicht zugänglich sind“. Als Beispiel ist an dieser Stelle das Telefonbuch zu nennen: Alle Informationen, die in das Telefonbuch eingetragen sind, sind allgemein verfügbar – daher besteht kein schutzwürdiges Interesse. (Bundeskanzleramt, 2000)

Mit der Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates „zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG“ – auch genannt Datenschutz-Grundverordnung – werden sowohl das DSG 2000 als auch die Richtlinie 95/46/EG in ihrer aktuell bestehenden Form (Stand: 1. Juni 2017) abgelöst. Die Datenschutz-Grundverordnung (DSGVO) wurde am 27. April 2016 beschlossen und tritt mit 25. Mai 2018 in Kraft.

3.2.1 Begriffe

In der DSGVO werden einige Begriffe definiert, deren Erläuterung für das Verständnis dieser Verordnung notwendig ist und die daher im Folgenden kurz dargelegt werden.

- Personenbezogene Daten: sämtliche Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person, auch genannt betroffene Person, beziehen. Identifizierbar bedeutet, dass direkt oder indirekt durch Zuordnungen von Kennungen die Identität der betroffenen Person festgestellt werden kann.
 - Im DSG2000 bereits gleichbedeutend
- Sensible Daten: personenbezogene Daten bezüglich rassischer und ethnischer Herkunft, politischen Meinungen, religiösen oder weltanschaulichen Überzeugungen; Gewerkschaftszugehörigkeit, genetische Daten, biometrische Daten, Gesundheitsdaten, Daten zum Sexualleben oder der sexuellen Orientierung
 - Im DSG2000 bereits gleichbedeutend
- Verarbeitung: jede automationsunterstützte und nicht-automationsunterstützte Verwendung von Daten, also: erheben, erfassen, organisieren, ordnen, speichern, anpassen, verändern, auslesen, abfragen, offenlegen, verbreiten, bereitstellen, abgleichen, verknüpfen, einschränken, löschen, vernichten
 - Im DSG2000 wird stattdessen der Begriff „Verwenden von Daten“ verwendet
- Verantwortlicher: entscheidet, zu welchem Zweck mit welchen Mitteln personenbezogene Daten verarbeitet werden
 - Im DSG2000 wird stattdessen der Begriff „Auftraggeber“ verwendet

- Auftragsverarbeiter: verarbeitet im Auftrag des Verantwortlichen, z.B. externe Buchhaltung, Cloud-Dienstleister
 - Im DSG2000 wird stattdessen der Begriff „Dienstleister“ verwendet
- Einwilligung muss enthalten: alle zu verarbeitenden und zu übermittelnden Daten, Zwecke der Datenverarbeitung, Empfänger der Daten, Information über das Recht, die Einwilligung jederzeit widerrufen zu können
 - In der DSGVO neu: eine eindeutige Handlung ist abzugeben – Stillschweigen, Untätigkeit oder standardmäßig angekreuzte Kästchen gelten nicht als Einwilligung zur Datenverarbeitung (Unger, 2016)

(Wirtschaftskammer Österreich, 2016) (Bundeskanzleramt, 2000) (Europäisches Parlament und Rat, 2016)

- Datenschutzbeauftragter:
 - ist eine zu bestellende Rolle in all jenen Unternehmen, die entweder regelmäßig systematische Überwachungen von Betroffenen durchführen müssen oder deren Kerntätigkeit die umfangreiche Verarbeitung sensibler Daten oder Daten aus strafrechtlichen Verurteilungen oder Straftaten ist
 - ist Anlaufstelle für alle Mitarbeiter, die Unternehmensführung und Behörden in allen Belangen des Datenschutzes
 - darf keine Anweisung bezüglich der Ausführung dieser Tätigkeit erhalten, berichtet immer der höchsten Managementebene
 - hat es im DSG2000 nicht gegeben

(Wirtschaftskammer Österreich, 2017) (Bundeskanzleramt, 2000) (Europäisches Parlament und Rat, 2016)

- Datenschutzbehörde: eine Dienstbehörde des Bundes Österreich, die unabhängig und an keine Weisungen gebunden ist. Sie ist das Kontrollorgan für alle Belange des Datenschutzes in Österreich

(Wirtschaftskammer Österreich, 2016) (Bundeskanzleramt, 2000) (Europäisches Parlament und Rat, 2016)

3.2.2 Aufbewahrungspflichten, Löschpflichten & Data Warehouse

Aus den gesetzlichen Rahmenbedingungen zum Thema Datenschutz geht somit hervor, dass sämtliche personenbezogenen Daten nur so lange aufbewahrt werden dürfen, wie es für die tatsächliche Geschäftsbeziehung mit der Person notwendig ist.

Diese Notwendigkeit ist je nach Auftraggeber unterschiedlich definiert. Jeder Fördervertrag mit dem AMS enthält beispielsweise eine Datenschutzvereinbarung, die eine Aufbewahrung aller Teilnehmerdaten für 24 Monate nach Ende der Maßnahme vorsieht, unabhängig von Ein- und Ausstieg der Teilnehmer. Nach diesen 24 Monaten müssen die Daten allerdings vom Auftragnehmer gelöscht werden. Abgesehen davon, dass nicht vollständig geklärt ist, ob JAW als Dienstleister im Auftrag des AMS – also als Auftragsverarbeiter – agiert oder selbst als Auftraggeber – also Verantwortlicher –, stellt diese Anforderung einen Konflikt mit den Charakteristika eines Data Warehouse dar, zu denen die dauerhafte unveränderliche Speicherung der Daten zählt. (vgl. Kapitel 2.2.2 „Beständigkeit (non-volatility)“).

Somit bedarf dieser Konflikt einer Lösung – die Daten müssen zwar gelöscht, sollen aber dennoch aufbewahrt werden. Damit die Daten im Data Warehouse nicht insofern verfälscht werden, als dass sich die errechneten Kennzahlen durch die Löschung der Daten verändert, also die Beständigkeit gefährdet ist, stellt eine Herangehensweise die „Anonymisierung“ dar. Darunter versteht man, die direkte oder indirekte Identifizierbarkeit einer Person unmöglich zu machen, indem die Daten mit Pseudodaten überschrieben werden. Allerdings reicht es hierbei nicht unbedingt, die eindeutig einer Person zuordenbaren Daten wie Vorname, Nachname, Adresse, Telefonnummer, E-Mail-Adresse oder Sozialversicherungsnummer zu entfernen – unter Umständen reichen die Informationen, dass es sich bei der Person um einen männlichen 24-Jährigen aus Mureck handelt, dessen Geburtsland Syrien ist und der eine Dienstleistung von Sozialträger X beansprucht hat, um die Person doch wieder identifizieren zu können. Ginge es bei dem Datensatz beispielsweise um Graz, wäre die Chance auf eine Identifikation weitaus geringer. Das macht die Arbeit des Anonymisierens durchwegs komplex und viele Kombinationsmöglichkeiten müssen bedacht werden, damit ein Rückschluss auf die Person tatsächlich nicht mehr möglich ist.

Zusätzlich müssen etwaige Freitextfelder bedacht werden, in denen immer Daten stehen können, welche die Identifikation einer Person ermöglichen, auch wenn das Feld in der entsprechenden Applikation gar nicht dafür gedacht war, diese Information zu beherbergen.

(Wes, 2017) (Sumroy & Cousin, 2016)

4 ANALYSEN & METHODEN

Mit den in Kapitel 3.1 „Werkzeuge“ beschriebenen Tools ist es möglich, mit Daten auf vielfältige Weise zu interagieren, um sie zu analysieren. Diese Interaktionen sind produktunabhängig und daher auch mit OLAP-Tools anderer Hersteller auf dieselbe Art und Weise möglich. Im Folgenden wird näher auf die unterschiedlichen Möglichkeiten der Analyse eingegangen.

Anschließend werden die Methoden „Explorative Datenanalyse“ sowie „Befragung“ erläutert, welche für die Hypothesenüberprüfung (vgl. Kapitel 1.4 „Forschungsfragen“) verwendet werden.

4.1 OLAP

Zur Analyse einer OLAP-Datenbank steht eine Reihe von Analysemethoden zur Verfügung.

In der Literatur wird zumeist von folgenden Methoden gesprochen:

- Slicing
- Dicing
- Pivoting
- Drill-Down / Roll-Up
- Drill-Through & Drill-Across
- Split & Merge

Nachfolgend werden die einzelnen Analysemethoden erläutert und die Vorgehensweise im Rahmen ihrer Anwendung skizziert.

4.1.1 Slicing

Beim Slicing wird eine Scheibe vom Würfel „ausgeschnitten“, indem auf eine einzelne Dimension ein Filter gelegt wird. Durch den Filter ist ein kleinerer Ausschnitt der Daten sichtbar, also nur noch eine „Scheibe“ des Würfels. Umgelegt auf einen dreidimensionalen Würfel mit den Dimensionen Zeit, Produkt und Ort bedeutet das, dass beispielsweise nur noch ein bestimmter Zeitabschnitt betrachtet wird, was dem Ausschneiden dieser einen Zeitscheibe gleichkommt (vgl. Abbildung 4). Slicing ermöglicht das detaillierte Betrachten von bestimmten Aspekten des Datenwürfels. (Ballard, Data modeling techniques for data warehousing, 1998, S. 45) (Ballard, Farrell, Gupta, Mazuela, & Vohnik, 2006, S. 87)

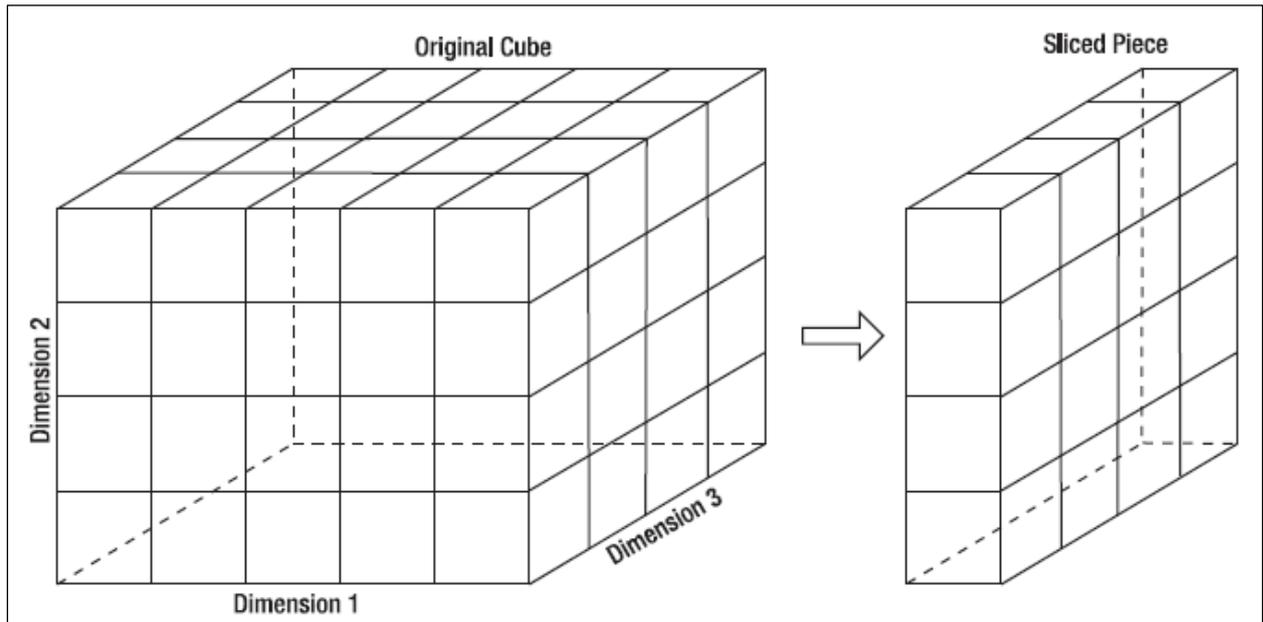


Abbildung 4 Slicing. Quelle: (Rainardi, 2008, S. 413)

4.1.2 Dicing

Dicing ist der Vorgang, einen kleineren Würfel aus einem Cube „herauszuschneiden“. Das geschieht, indem bestimmte, gefilterte Einträge von einer Dimension auf eine Achse und Einträge einer anderen Dimension auf eine andere Achse gelegt werden. Beim Beispiel Zeit-Produkt-Ort würde das bedeuten, dass aus dem Cube ein Teilwürfel mit einem bestimmten Zeitraum an mehreren Orten ausgewählt wird, was in einem kleineren Würfel resultiert (vgl. Abbildung 5). Mithilfe von Dicing können Zusammenhänge zwischen den einzelnen Dimensionen analysiert werden. (Ballard, Data modeling techniques for data warehousing, 1998, S. 45) (Ballard, Farrell, Gupta, Mazuela, & Vohnik, 2006, S. 88)

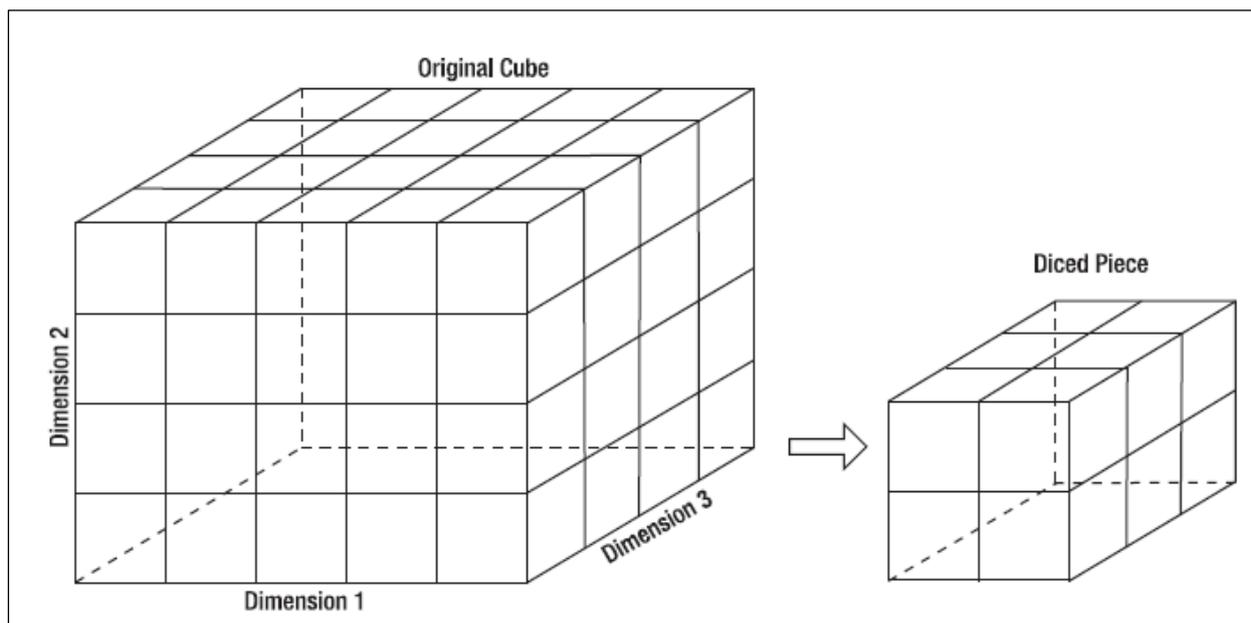


Abbildung 5 Dicing. Quelle: (Rainardi, 2008, S. 414)

4.1.3 Pivoting

Pivoting ist der Vorgang, die Spalten und Zeilen – also die Achsen – einer Auswertung miteinander zu vertauschen. Dies ermöglicht ebenfalls eine andere Sicht auf dieselben Daten, es resultiert sogar in denselben Ergebnissen – nur in anderer Reihenfolge. Je nach Art und Menge an Datensätzen der ausgewählten Dimensionen kann dieser Vorgang die Übersichtlichkeit und Verständlichkeit stark fördern.

Beispielsweise wird eine Auswertung mit den Dimensionen Zeit und Produkt bei Auswahl der Quartale als Dimensionsattribut einfacher lesbar sein, wenn die Zeit als Spalten und das Produkt als Zeilen hinzugefügt werden als umgekehrt.

(Ballard, Farrell, Gupta, Mazuela, & Vohnik, 2006, S. 90)

4.1.4 Drill-Down / Roll-Up

Ein Drill-Down ist das Navigieren in eine höhere Detailstufe, also hinunter in Richtung Rohdaten. Beispielsweise ist ein einfacher Drill-Down das Navigieren von einer jährlichen Datendarstellung hinunter zu einer monatlichen Ansicht. Es wird also die Granularität erhöht, indem von der Zeitdimension das Attribut „Monat“ angezeigt wird, das über mehr unterschiedliche Werte verfügt, als das Attribut „Jahr“.

Das Roll-Up ist genau der umgekehrte Weg: die Daten werden zusammengefasst („eine Dimension wird zusammengerollt“) und aus den einzelnen Monaten werden wieder Jahre, vgl. auch Abbildung 6.

(Ballard, Data modeling techniques for data warehousing, 1998, S. 44f)

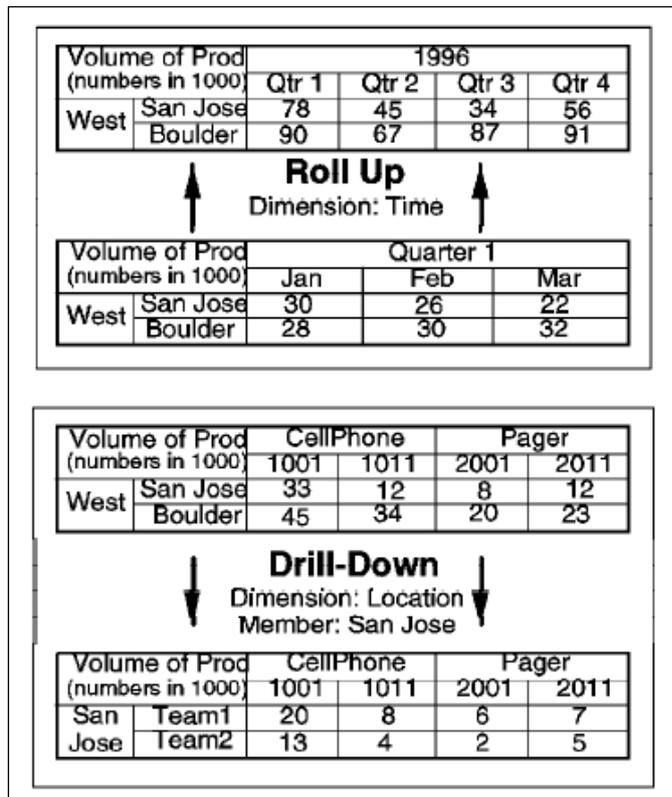


Abbildung 6 Roll-Up und Drill-Down. Quelle: (Ballard, Data modeling techniques for data warehousing, 1998, S. 45)

4.1.5 Drill-Through & Drill-Across

„Drill-Through“ und „Drill-Across“ sind Vorgänge, die eine andere Perspektive auf Daten schaffen, indem auf eine andere Datenquelle zugegriffen wird, welche dieselben Dimensionen nutzt.

Beim Drill-Through bohrt man durch die tiefste Ebene der Drill-Down-Möglichkeiten hinunter auf die originale relationale Datenquelle, also die Rohdaten des DW. Dadurch wird der höchste Detaillierungsgrad erzielt, der mit den vorhandenen Daten möglich ist.

Beim Drill-Across wird zu einem benachbarten Cube „hinübergebohrt“, beispielsweise von den Verkaufszahlen zu den Einkaufszahlen. Diejenigen Dimensionen, die in beiden Cubes vorhanden sind – beispielsweise Zeit und Ort – werden beim Wechsel in den anderen Cube automatisch ausgewählt und die Dimension Kunde, die es nur bei den Verkaufszahlen gibt, wird durch die Dimension Lieferant ersetzt.

(Kemper, Baars, & Mehanna, 2010, S. 103f)

4.1.6 Split & Merge

Ein „Split“ passiert, wenn eine weitere Dimension auf eine Achse gelegt wird, die Werte also genauer aufgegliedert werden. Beispielsweise könnte zunächst Zeit auf der einen und Produkt auf deren Achse ausgewertet werden. Durch das Hinzufügen von Ort als zusätzliche Dimension auf eine dieser Achsen werden die zuvor angezeigten Werte gesplittet, also auf mehrere Zellen aufgeteilt. Dies ermöglicht das genauere Analysieren auf Basis von mehreren Attributen, um den Detailgrad der Auswertung zu erhöhen.

„Merge“ ist der umgekehrte Weg: Werte werden wieder aggregiert, indem eine Dimension von einer der Achsen entfernt wird, der Detaillierungsgrad sinkt also.

(Kemper, Baars, & Mehanna, 2010, S. 105f)

4.2 Explorative Datenanalyse

Die Analyse von Daten kann zumeist in drei Kategorien eingeteilt werden:

- Zusammenfassung der Daten: Durch die Reduzierung des Detaillierungsgrades erhöht sich die Übersichtlichkeit und Erfassbarkeit der Daten.
- Finden versteckter Zusammenhänge: Hierbei geht es um das Identifizieren von wichtigen Fakten, Zusammenhängen, Anomalien oder Trends in den Daten, die aus der Zusammenfassung heraus nicht erkennbar sind.
- Treffen von Voraussagen: Es wird versucht, mithilfe der vergangenen Daten eine Voraussage über die zukünftige Entwicklung zu treffen.

(Myatt, 2017, S. 2ff) (Cleff, 2008, S. 8f) (Taylor & Cihon, 2004, S. 20f)

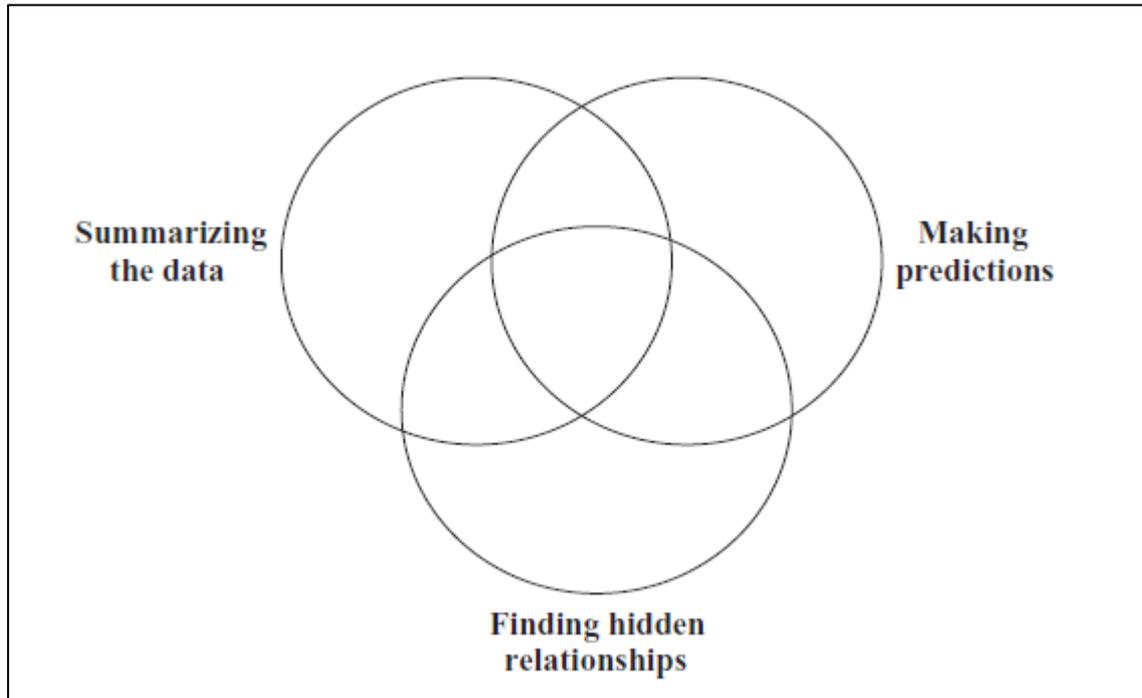


Abbildung 7 Zusammenhang der drei Kategorien von Datenanalysen. Quelle: (Myatt, 2017, S. 3)

Die explorative Datenanalyse und das Data Mining umfassen mehrere Methoden, darunter beispielsweise:

- Zusammenfassungstabellen: Rohdaten werden auf unterschiedliche Art und Weise zusammengefasst und in Tabellenform präsentiert
- Graphen: ermöglichen dem menschlichen Auge das visuelle Identifizieren von Trends und Zusammenhängen
- Beschreibende Statistik: Beschreibungen von Informationen in einer bestimmten Spalte, beispielsweise der Mittelwert oder die Extremwerte (Minima, Maxima)
- Inferenzstatistik: mathematische Analyse der Daten
- Korrelationsstatistik: Zusammenhänge in den Daten erkennen
- Suchen: bestimmte Fragen stellen, die mit den Daten beantwortet werden können
- Gruppieren: Methoden zum Organisieren von Daten in kleinere Einheiten, um Antworten auf Fragen zu finden
- Mathematische Modelle: eine mathematische Gleichung oder ein Prozess, der Voraussagen treffen kann

(Myatt, 2017, S. 3f) (Cleff, 2008, S. 3f)

4.3 Befragung

Die Befragung ist gleichzusetzen mit dem Interview. In der Literatur wird die Befragung als schriftliche Form des Interviews angesehen. Im weiteren Verlauf dieser Arbeit werden aus Gründen der Lesbarkeit diese Begriffe synonym verwendet. Es gibt mehrere Dimensionen, die eine solche Befragung oder ein Interview beschreiben und bewerten:

- Intention des Interviews: Bei der Agenda des Interviews ist zu unterscheiden, ob es sich um das Einholen von Informationen oder das Beeinflussen des Befragten handelt, um eine Erkenntnis- oder Bewusstseinsveränderung des Befragten zu provozieren.
- Standardisierung: Der Grad der Standardisierung gibt an, wie stark der Ablauf des Interviews im Vorfeld geplant wurde. Nichtgeplante Kommunikation, auch Alltagskommunikation, kann durchaus auch als wissenschaftliche Befragung herangezogen werden und gilt als qualitatives Interview. Quantitative Interviews besitzen einen hohen Grad der Standardisierung und sind für eine größere Menge an Befragten ausgelegt.
- Struktur der zu Befragenden: Handelt es sich mehrheitlich um Einzelbefragungen, so ist von qualitativen Interviews auszugehen; handelt es sich mehrheitlich um Gruppenbefragungen, ist von quantitativen Interviews auszugehen
- Form der Kommunikation: Klassischerweise kann man zwischen schriftlicher und mündlicher Kommunikation unterscheiden. Meistens werden quantitative Interviews schriftlich und qualitative Interviews mündlich durchgeführt.
- Stil der Kommunikation: Es wird zwischen „weichem“ und „hartem“ Interview unterschieden. Beim weichen Interview wird empathisch auf den Befragten eingegangen, das harte Interview hat eine starke Ähnlichkeit zu einem Verhör.

Des Weiteren gibt es unterschiedliche Interviewtypen, von denen insbesondere das diskursive Interview eine hohe Relevanz für diese Masterarbeit besitzt. Hierbei werden die Befragten „primär als Theoretiker und Experten ihrer selbst, ihrer Geschichte und ihrer Eigenheiten angesprochen“ (Stigler, 2001, S. 35). Das Ziel dieser diskursiven Interviews ist das kommunikative Validieren von Thesen des Interviewers.

(Stigler, 2001, S. 33ff) (Atteslander, S. 24f)

5 IST-ZUSTAND

Bei JAW gibt es zurzeit sehr viele unterschiedliche Dienstleistungen mit ebenso unterschiedlichen Anforderungen an deren Datenhaltung. Meist resultiert die strukturierte Datenhaltung aus den Abrechnungskonventionen und ist nicht direkt von der tatsächlichen operativen Aktivität ausgehend. Im Vorfeld zur Arbeit wurden Befragungen mit den Dienstleistungsmanagern durchgeführt, um sämtliche Dienstleistungen von JAW zu erfassen. Anschließend wurde deren Art der Datenspeicherung analysiert und zusammengefasst. (vgl. Anhang 1: Erhebung Dienstleistungen und Datenablagen).

Das Ergebnis dieser Analyse wird nachfolgend dargelegt und mit grafischen Darstellungen untermalt. Anschließend wird auf diejenigen Geschäftsprozesse eingegangen, welche mit den Stammdaten interagieren. Abschließend werden die Datenmodelle der eruierten Datenbankapplikationen analysiert und erläutert.

5.1 Aufstellung Datenhaltung

Durch die Erhebung der unterschiedlichen Dienstleistungen ist sichtbar, dass für ein und dieselbe Dienstleistung oftmals mehrere Datenbankapplikationen eingesetzt werden. Daher unterteilt die Aufstellung in „Datenbank primär“ und „Datenbank sekundär“. Diejenige Applikation, die von einem breiteren Publikum innerhalb JAW und damit häufiger genutzt wird, wodurch auch eine größere Menge an Daten vorhanden ist, wird „Datenbank primär“ genannt. Wird dahinter, möglicherweise von einem anderen Personenkreis, mit denselben Informationen in einer weiteren Datenbank gearbeitet, gilt diese als „Datenbank sekundär“. Dies gilt unabhängig davon, ob es eine automatische Schnittstelle zwischen diesen Systemen gibt oder manuelle Eingaben in das System des Auftraggebers erfolgen.

5.1.1 Gliederung nach Applikation

Die Informationen über die Dienstleistungen, also Auftraggeber und Datenbanksystem, werden nachfolgend in Tabelle 1 aufgeschlüsselt.

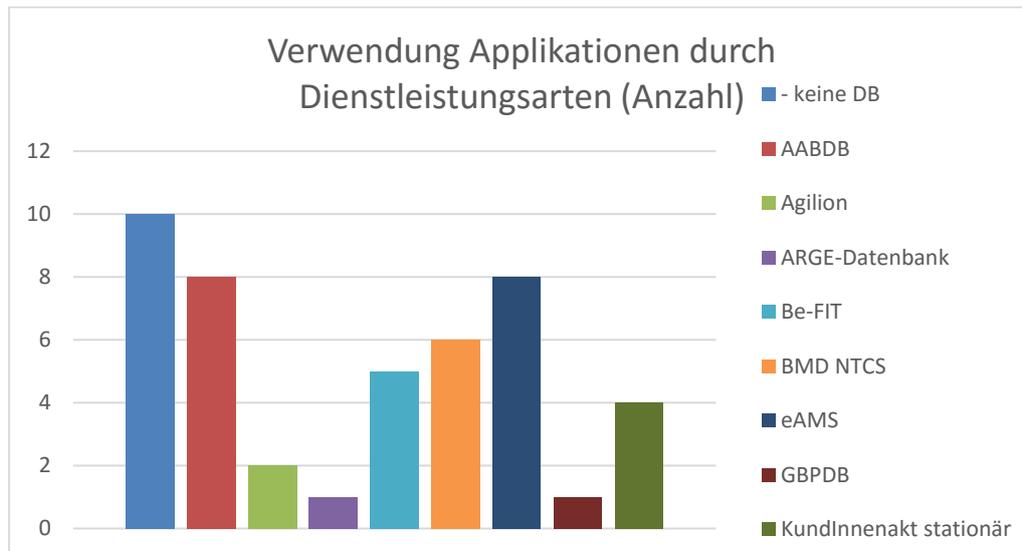


Abbildung 8 Verwendung Applikationen durch Dienstleistungsarten (Anzahl). Quelle: eigene Darstellung (Microsoft Excel)

Als Zusammenfassung ergibt sich, wie in Abbildung 8 ersichtlich, dass lediglich 10 von 45 (=22%) mit dieser Methode erfassten Dienstleistungen keine Datenbankapplikationen zur strukturierten Datenablage verwenden. Die Anwendungen „AABDB“ und „eAMS“ liegen auf Platz zwei, auf Platz drei folgt „Be-FIT“. Für die grafischen Auswertungen wurden sowohl die Spalten „Datenbank primär“ als auch „Datenbank sekundär“ herangezogen – dadurch ist die Gesamtanzahl an (nicht-)Datenbankanwendungen höher als die Anzahl der in Tabelle 1 eruierten Dienstleistungsarten.

5.1.2 Aufstellung unstrukturierte Daten

Zusätzlich ist aus Tabelle 1 abzulesen, dass die Mehrheit der Dienstleistungen unstrukturierte Office-Dokumente am Fileserver zur Datenspeicherung verwendet (vgl. Abbildung 9), sowie welchem Dienstleistungsmanagement (DLM) die entsprechende Dienstleistung zuzuordnen ist (vgl. 1.6 Dienstleistungsbereiche).

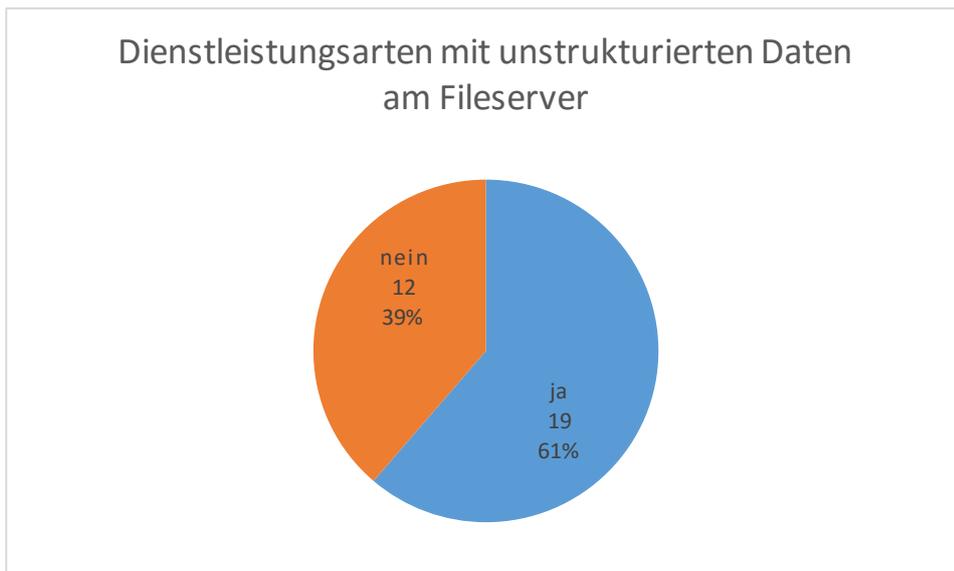


Abbildung 9 Dienstleistungsarten mit unstrukturierten Daten am Fileserver. Quelle: eigene Darstellung (Microsoft Excel)

Diese Mehrheit an Dienstleistungsarten mit unstrukturierten Daten am Fileserver wird durch Abbildung 9 visualisiert.

5.1.3 Aufstellung Daten nach Dienstleistung

Tabelle 1 stellt das Ergebnis der Befragungen der Dienstleistungsmanager dar und ist, wie zuvor erwähnt, nach Dienstleistungsbereich und Datenapplikation gegliedert. Zusätzlich werden die einzelnen Auftraggeber aufgelistet.

Dienstleistung	DLM	Auftraggeber	Datenbank primär	Datenbank sekundär	Office-Dokumente am Fileserver
Arbeitsassistentz	AAB	SOMS / NEBA	Be-FIT		ja
Jobcoaching	AAB	SOMS / NEBA	Be-FIT		ja
Überbetriebliche Lehrausbildung (ÜBA)	AAB	AMS	AABDB	eAMS	nein
Perspektivenwerkstatt	AAB	AMS	AABDB	eAMS	nein
Integrative Berufsausbildung (IBA)	AAB	AMS	AABDB	eAMS	nein
BeTrain	AAB	SOMS	Be-FIT		ja
Lehrlingscoaching	AAB	WKO	AABDB		ja
Mobile Dienstleistungen nach StKJHG-DVO	BFW	Land Steiermark	Agilion		ja
Mobile Dienstleistungen nach LEVO-StBHG	BFW	Land Steiermark	Agilion		ja
Streetwork	BFW	Land Steiermark	- keine DB		ja
Teilstationäre Dienstleistungen nach LEVO-StBHG	AAB	Land Steiermark	KundInnenakt stationär	BMD NTCS	nein

Grundversorgung	BFW	Land Steiermark	- keine DB	BMD NTCS	ja
Unbegleitete minderjährige Flüchtlinge (UMF)	BFW	Land Steiermark	- keine DB	BMD NTCS	ja
Produktionsschule	AAB	SOMS / NEBA	Be-FIT		ja
Bewerbungsunterstützung	AAB	AMS	AABDB	eAMS	nein
Jugendcoaching	AAB	SOMS / NEBA	Be-FIT		ja
Gemeinnütziges Beschäftigungsprojekt (GBP)	AAB	AMS	GBPDB	eAMS	nein
Stationäre Dienstleistungen nach LEVO-StBHG	BFW	Land Steiermark	KundInnenakt stationär	BMD NTCS	nein
Inabschluss	AAB	Land Steiermark	- keine DB		ja
Jobfit	AAB	AMS	AABDB	eAMS	nein
Frühe Hilfen	BFW	Stmk. GKK	- keine DB		ja
Neuland - 24-Stunden-Betreuung	BFW	Stadt Graz	KundInnenakt stationär	BMD NTCS	nein
Jugendzentrum	BFW	Stadt Graz	- keine DB		ja
Perspektivencheck für Jugendliche in Graz	AAB	AMS	AABDB	eAMS	nein
Jugendpark	AAB	AMS	AABDB	eAMS	nein
Stationäre Dienstleistungen nach StKJHG (DVO)	BFW	Land Steiermark	KundInnenakt stationär	BMD NTCS	nein
Housing First	BFW	Stadt Graz	- keine DB		ja
Selbsthilfekontaktstelle	BFW	Land Steiermark	- keine DB		ja
Pflegeeltern	BFW	Land Steiermark	- keine DB		ja
Flexible Hilfen	BFW	Land Steiermark	ARGE-Datenbank		ja
Begegnungszentrum Vinzenz-Muchitsch-Haus	BFW	Stadt Graz	- keine DB		ja

Tabelle 1 Tabellarische Darstellungen der Dienstleistungen und deren Datenapplikationen. Quelle: eigene Darstellung (Microsoft Excel)

5.2 Analyse Geschäftsprozesse

Um das DW so gestalten zu können, dass ein Mehrwert für das Unternehmen entsteht, gilt es zunächst zu verstehen, aus welchen Geschäftsprozessen die im DW aufbewahrten Daten entstehen und wofür sie verwendet werden. Damit kann dann ein Datenmodell erstellt werden, das genau diese Informationen analytisch verwertbar macht. (vgl. Kapitel 2.3.4 Dimensionales Modell)

5.2.1 Stammdaten

Die Stammdaten sind in ihrer Natur Daten, die sich nicht verändern sollten und die Basis für weitere geschäftliche Tätigkeiten darstellen. Allerdings müssen diese auch zu einem bestimmten Zeitpunkt in einer bestimmten Art und Weise aufgenommen werden. Sie bilden die Grundlage für das operationale Handeln des Unternehmens und legen fest, wie, wann und in welchem Ausmaß Leistungen schlussendlich zur Verrechnung kommen. Daher sind diese essentiell für den Betrieb und Fortbestand eines Unternehmens: Werden Rechnungen auf Basis falscher Stammdaten- oder auch gar nicht- gestellt kann dies schwerwiegende Folgen haben.

Je nach Datenquelle bei JAW haben die Stammdaten unterschiedlich große Auswirkungen auf die Fakturierung. In jenen Bereichen, in denen das AMS als Auftraggeber auftritt, haben die Stammdaten der operativen Tätigkeit nur mit starker Verzögerung einen Zusammenhang zur Abrechnung, welche in der Regel einmal jährlich durchgeführt wird. Bei dieser Abrechnung müssen die Teilnehmerzahlen sowie deren Anwesenheiten und auch die zugehörigen Trainer und deren Anwesenheiten sehr detailliert aufgeschlüsselt werden. Dies bedeutet zwar, dass die Aktualität der Stammdaten nicht direkt verrechnungstechnisch relevant ist, aber die laufend notwendigen Aufzeichnungen diese Stammdaten ebenfalls benötigen.

Bei den Dienstleistungen, die auf Basis von steiermärkischen Landesgesetzen durchgeführt werden, ist die Aktualität von Stammdaten weitaus relevanter, da die Rechnungen im Monatsrhythmus exakt auf Basis der erbrachten Leistungen gestellt werden müssen und deren Plausibilität und Korrektheit zeitnah überprüft wird.

Ein großes Problem derartiger Stammdaten in den üblichen Literatur-Datenbankdesigns ist, dass sich Stammdaten ändern und diese Änderung nicht nur später nachvollziehbar sein muss, sondern im täglichen Betrieb eines Unternehmens auch unabhängig vom Zeitpunkt der Eingabe gültig sein können muss – beispielsweise ist es keine Seltenheit, dass eine Bezirkshauptmannschaft die mündliche Zusage gibt, dass ein bestimmter Kunde die eine oder andere Art von Bescheid erhalten wird. Auf Basis dieser Zusage wird unter Umständen monatelang betreut, bevor der Bescheid, auf dessen Basis die Abrechnung erst möglich ist, rückwirkend eintrifft. Umgekehrt gibt es auch Fälle, bei denen bereits monatelang im Vorhinein klar sein wird, welche Art von Bescheid ab einem bestimmten Stichtag seine Gültigkeit haben wird.

Deshalb ist es von Vorteil, die Änderung an Stammdaten als Geschäftsprozess zu sehen – das Ausstellen eines Bescheides durch eine Bezirkshauptmannschaft verursacht Änderungen der Stammdaten in Systemen von JAW sowie der Art und Weise, wie die operative Dienstleistung tatsächlich durchgeführt wird – obwohl in den Systemen eigentlich nur Stammdaten angepasst werden.

Gelten die Änderungen von Stammdaten als Prozess, verfügt dieser über mehrere unterschiedliche Transaktionen, die eintreten können. Diese Transaktionen werden nun näher erläutert.

5.2.2 Eintritt

Der Eintritt ist ein Geschäftsprozess, der mehrere Schritte beinhaltet. Die meisten davon sind in der operativen Tätigkeit mit dem Kunden angesiedelt und werden im Rahmen dieser Masterarbeit nicht behandelt. Eine der Tätigkeiten ist jedoch das Anlegen von Stammdaten in einem Datenbanksystem, was sehr wohl in diesem Rahmen relevant ist. Die Stammdaten müssen einige Pflichtfelder – wie beispielsweise den Namen des Kunden – beinhalten, viele weitere Daten sind jedoch unter Umständen erst später verfügbar und werden daher nachträglich eingetragen. Diese nachträglichen Eintragungen könnten zwar als Änderung gesehen werden, allerdings ist der gesamte Datensatz, der durch den Eintritt entsteht, als eine einzelne Transaktion zu sehen, auch wenn sich dessen Entstehung über einen längeren Zeitraum erstreckt.

Wie zuvor im Beispiel erwähnt, könnte es sein, dass die Betreuung eines Kunden beginnt, obwohl noch kein Bescheid ausgestellt wurde und damit streng genommen noch kein Betreuungsvertrag entstanden ist. Die nachträgliche Eingabe dieses Bescheids in die Applikation ist jedoch keine Änderung im Sinne der Stammdaten, da der Bescheid rückwirkend zum Eintrittsdatum gültig ist und damit die Abrechnung beeinflusst. Für das DW ist also der gesamte Datensatz, unabhängig vom Zeitpunkt der Eingabe, relevant.

Zusätzlich ist es möglich, dass ein und dieselbe Person unterschiedliche Dienstleistungen von JAW bezieht und daher in unterschiedlichen Systemen oder sogar in einem einzelnen System mehrmals angelegt wird, jedoch unterschiedliche Kundennummern bekommt. Da es rein abrechnungstechnisch irrelevant ist, ob es sich um dieselbe Person handelt oder nicht, werden einfach die exakt selben Stammdaten unter neuer Kundennummer mit der Information der anderen Dienstleistung erneut angelegt.

5.2.3 Änderung

Eine Änderung findet statt, wenn sich die wirksamen Daten – wiederum unabhängig vom Zeitpunkt der Eingabe – ändern. Dies könnte, um beim Beispiel von Bescheiden aus der Landesgesetzgebung heraus zu bleiben, die Ausstellung eines neuen Bescheids sein, dessen Gültigkeit entweder an den alten Bescheid angrenzt oder ihn bereits vor Ablauf des alten Bescheids ablöst.

Eine andere klassische Änderung wäre der Wohnortwechsel des Kunden, was in einer neuen Wohnanschrift resultiert. Diese Änderung ist üblicherweise ein Fall für eine SCD (vgl Kapitel. 2.3.3

„Slowly Changing Dimensions“) und in klassischen DW führt dies lediglich in der zugehörigen Dimension zu einer Änderung. Alle danach eintretenden Fakten werden mit der neuen Version der Kundendimension verknüpft. Da sich der gesamte betrachtete Geschäftsprozess nur um die Stammdaten kümmert, erzeugen diese auch in der Faktentabelle eine neue Zeile, um die neuen Umstände in Auswertungen entsprechend darstellen zu können.

5.2.4 Austritt

Der Austritt ist klar definiert mit dem Ende des Betreuungsverhältnisses zwischen JAW und dem Kunden. Wie beim Eintritt bereits erwähnt, kann eine einzelne Person jedoch mehrere unterschiedliche Dienstleistungen beanspruchen, was den Fall möglich macht, dass eine Person nur das Betreuungsverhältnis einer bestimmten Kundennummer beendet – die Person selbst aber nach wie vor ein Kunde von JAW ist.

Da das Austrittsdatum wie alle Daten in den Applikationen unabhängig vom Zeitpunkt der Eintragung ist, ist auch dieser Eintrag in der Faktentabelle auf das Austrittsdatum selbst bezogen und nicht auf den Zeitpunkt der Eingabe.

5.2.5 Monatliches Reporting

Eine Frage, die sich das Management regelmäßig stellt, um die Fluktuation messen zu können, ist, wie viele Kunden in einem bestimmten Zeitraum ein- oder ausgetreten sind. Eine hohe Fluktuation kann bei manchen Dienstleistungsbereichen eine hohe Betreuungsqualität darstellen, da die Kunden rasch an den ersten Arbeitsmarkt vermittelt werden konnten, in anderen kann dies Probleme mit der Dienstleistung an sich oder den handelnden Personen vor Ort aufzeigen. Durch die Umsetzung sowohl des Eintritts als auch des Austritts als eine Transaktion in der Faktentabelle kann dies sehr einfach ausgewertet werden.

Doch nicht nur die Fluktuation ist eine relevante Messgröße, sondern auch die Anzahl der Kunden, aufgeschlüsselt nach Dienstleistungen, Standorten oder Kostenstellen, in denen sie betreut werden. Diese Frage ist aktuell ebenfalls nur durch das Zusammensammeln von Daten aus den vielen unterschiedlichen Datenquellen möglich und wird, um Aufwände niedrig zu halten, nicht für alle Dienstleistungen regelmäßig beantwortet. Ein DW, das diese Möglichkeit jedoch auf Knopfdruck bietet, würde die Relevanz dieser Frage bzw. dessen Beantwortung erhöhen. Dazu wird aus den statischen Stammdaten für jeden Kalendertag eine Zeile in einer Snapshot-Faktentabelle erstellt, die dann auch Auswertungen über Zeiträume hinweg ermöglicht.

5.3 Analyse Datenmodelle

Jede Datenapplikation wird nun kurz textuell beschrieben, sowie darauf eingegangen, welche Daten strukturiert innerhalb der Applikation abgelegt sind und welche möglicherweise nur außerhalb der Applikation in Form von unstrukturierten Dokumenten vorhanden sind. Dazu wird auch ein Ausschnitt des jeweiligen Datenbankmodells analysiert und dargestellt.

Die Reihenfolge, in der die Applikationen nachfolgend erläutert werden, ergibt sich aus der Legende in Abbildung 8 auf Seite 46.

5.3.1 Dokumente am Fileserver (keine Datenbank)

Wie in Kapitel 5.1 „Aufstellung Datenhaltung“ beschrieben gibt es viele Dienstleistungen, die keine Datenbank zur strukturierten Datenablage verwenden. Bei diesen Daten ist es zwar prinzipiell auch im Zuge des ETL-Prozesses (vgl. Kapitel 2.4 „Extract-Transform-Load“) möglich, sie in eine strukturierte Form zu bringen und in einem Data Warehouse abzulegen, jedoch würde dies den Umfang dieser Masterarbeit sprengen und wird daher nicht weiter behandelt.

5.3.2 AABDB

Die AABDB ist eine, unter Zuhilfenahme eines externen Dienstleisters, von JAW eigens entwickelte Datenbankapplikation, die auf die Datenhaltung von AMS-Maßnahmen mit Kurs- und Qualifizierungsteilnehmern spezialisiert ist. Der Name ist eine simple Zusammenstellung aus AAB („Ausbildung | Arbeit | Beschäftigung“) und DB („Datenbank“). Die Software ist sehr umfassend aufgebaut, um einerseits auf Knopfdruck sämtliche Formulare und Bestätigungen für die Kunden und das AMS, sowie auch statistische Informationen zu den Projekten zu erhalten. Andererseits ist auch die Abrechnung von Maßnahmenebenkosten aus der BMD NTCS Finanzbuchhaltung mitintegriert. Da sich diese Masterarbeit auf die Kundendaten konzentriert, fällt das Datenbankschema wie in Abbildung 10 ersichtlich sehr überschaubar aus. Ebenfalls ersichtlich ist, dass die Tabelle „TeilnehmerInnen“ ein Feld namens „Mit_MitarbNr_BMD“ beinhaltet. Das rührt daher, dass alle sich in der ÜBA befindlichen Teilnehmer über eine wiederum eigene BMD NTCS Datenbankinstanz (vgl. Kapitel 5.3.8 „BMD NTCS“) eine Lehre in einem Unternehmen außerhalb von JAW absolvieren. Die Teilnehmer sind dafür bei JAW als Lehrlinge angestellt und erhalten die entsprechende gesetzliche Lehrlingsentschädigung. JAW verrechnet einen Teil dieser Kosten weiter an die externen Firmen in Form von Firmenbeitragsrechnungen. Durch die Stellung als Lehrlinge kann die Personalverrechnung als führender Datenlieferant für diese Personen herangezogen werden, was einen Import aus BMD NTCS in die AABDB nützlich macht.

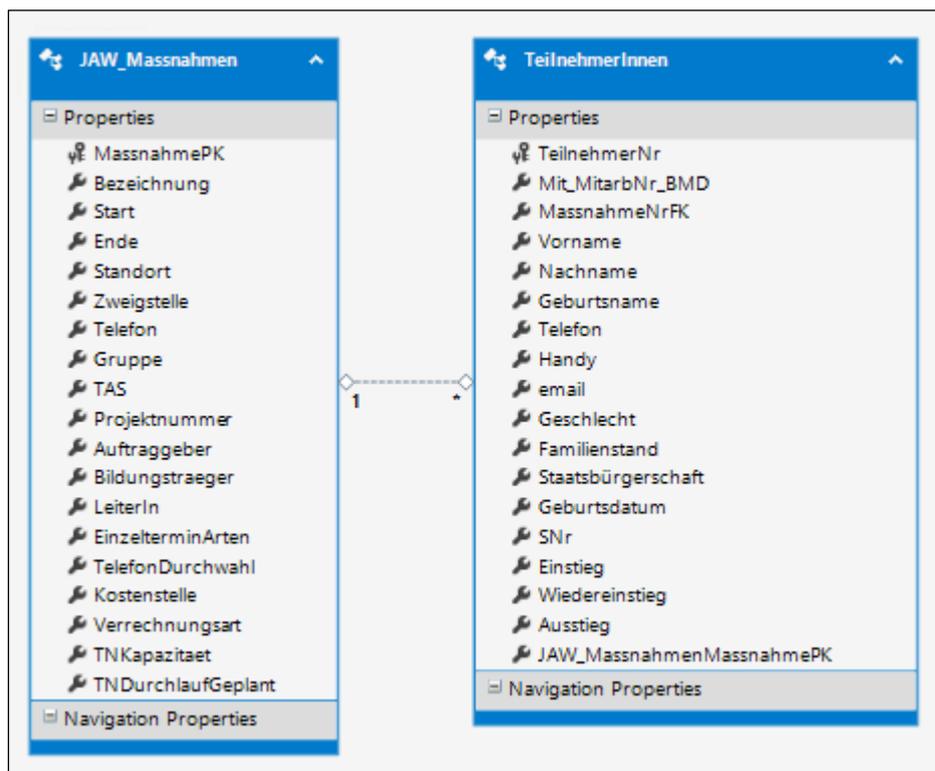


Abbildung 10 Datenstruktur AABDB. Quelle: eigene Darstellung (Microsoft Visual Studio)

Somit sind an relevanten Kundendaten strukturiert innerhalb der AABDB gespeichert:

- persönliche Daten (Name, Geschlecht, Geburtsdatum, SVNR, Anschrift)
- Dienstleistung von/bis (=AMS-Maßnahme)
- Standort bei JAW
- Kostenstelle bei JAW
- Eintritt/Austritt

Hierarchisch gesehen stellt das übergeordnete Element der Datenstruktur die Maßnahme dar, welche für jedes Maßnahmenjahr mit einer eigenen Projektnummer des AMS versehen wird. Um dies technisch abzubilden, existiert für Maßnahmen, die über mehrere Jahre hinweg ausgeführt werden, pro Jahr ein eigener Datensatz. Auf dieser Ebene werden auch Kostenstelle, Standort und viele weitere Informationen hinterlegt. Darunter werden die Kunden als eigenständige Datensätze angelegt. Sollte ein und dieselbe Person während eines Projektdurchlaufs mehrmals ein- und wieder aussteigen, so wird diese Person mehrfach angelegt, um alle Eintritts- und Austrittsdaten korrekt erfassen zu können. Welche Dienstleistung die Kunden in Anspruch nehmen, ergibt sich implizit daraus, in welcher Maßnahme sie angelegt wurden.

5.3.3 agilion

Die Applikation „agilion“ von der agilion GmbH (Homepage "agilion", kein Datum) ist die Zeit- und Leistungserfassung für die stundensatzfinanzierten mobilen Dienstleistungen aus StBHG und StKJHG. Die Mitarbeiter in der unmittelbaren Betreuung erfassen mittels Android-App während ihrer Arbeit die Beginn- und Endzeiten sowie die betreuten Kunden, gefahrenen Kilometer und vieles mehr.

Damit dies gut funktioniert, sind viele Stammdaten notwendig. Ein Auszug aus dem Datenbankmodell in Abbildung 11 zeigt die direkt mit den Kundenstammdaten in Zusammenhang stehenden Tabellen.

Es ist auch eine sehr starke Ausprägung von 3NF sichtbar – es gibt tatsächlich keine redundanten Daten in diesem Modell, es werden sogar alle Adressen in eine eigene Tabelle ADDRESSES ausgelagert. Sowohl jeder Kunde CUSTOMER, als auch jede Dienstleistungseinrichtung SERVICE_FACILITY, werden als Kontakt CONTACT angelegt, um auch hier keine Datenredundanzen zu erzeugen.

Die Zuweisungstabelle zwischen Kunden, Dienstleistungseinrichtungen und Dienstleistungsarten, CUSTOMER_SERVICE_ASSIGNMENTS, ist der Kern dieses Teils des Datenbankmodells von agilion. Hier werden die Kunden mit den Dienstleistungsarten SERVICE_CATEGORIES und den Dienstleistungseinrichtungen verknüpft. In dieser Tabelle gibt es Redundanzen, allerdings nur auf Basis von Fremdschlüsseln, um einfachere und schnellere Abfragen im täglichen Betrieb zu ermöglichen. Beispielsweise ist der Kunde mit dieser Tabelle sowohl direkt als auch über die Verträge SERVICE_CONTRACTS verknüpft.

Diese Tabelle wird bei der Überführung in das dimensionale Modell in der Mitte des Sternschemas landen, indem die Daten transformiert und als Faktentabelle verwendet werden.

Somit sind an relevanten Kundendaten strukturiert in agilion gespeichert:

- persönliche Daten (Name, Geschlecht, Geburtsdatum, SVNR, Anschrift)
- Dienstleistung von/bis
- Standort bei JAW
- Kostenstelle bei JAW
- Eintritt/Austritt

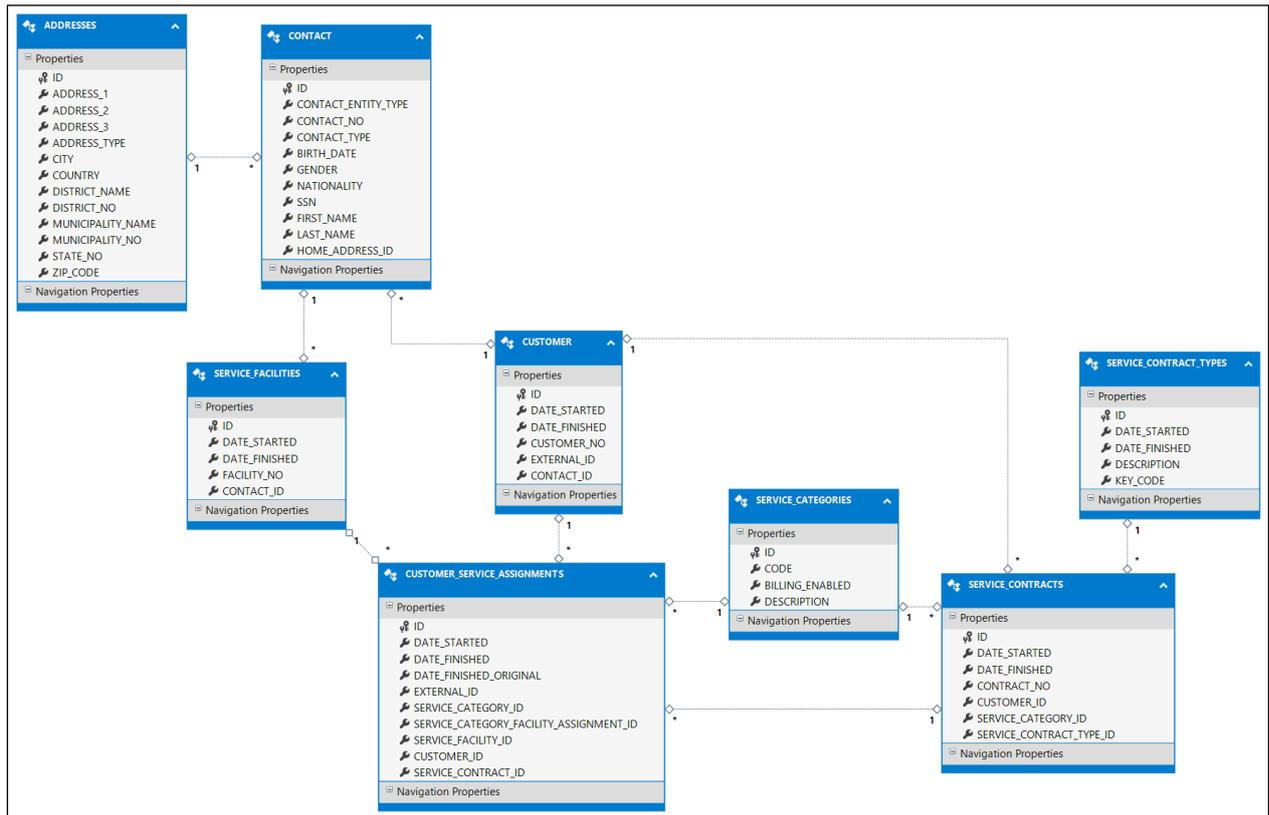


Abbildung 11 Datenstruktur agilien. Quelle: eigene Darstellung (Microsoft Visual Studio)

5.3.4 ARGE-Datenbank

An den ARGE-Datenbanken sind zusätzlich zu JAW noch weitere Unternehmen beteiligt und eine datenrechtliche Abgrenzung bzw. Abstimmung mit diesen Partnern würde den Rahmen dieser Masterarbeit sprengen. Darüber hinaus ist diese ARGE-Datenbank eine rein externe Applikation, auf die aus technischer Sicht kein Zugriff für die JAW IKT besteht.

5.3.5 Be-FIT

Be-FIT steht für Betriebs-Fach-Informationstechnologie und ist die E-Government-Applikation des Sozialministeriumservice. Darüber werden nicht nur diverse Projektanträge, Angebotseinreichungen und Finanzpläne an die Bundesbehörde kommuniziert, auch die laufende Verwaltung ist darin enthalten. Dafür ist der Teil „Monitoring berufliche Integration“ (MBI) zuständig.

Diese Daten werden ausschließlich innerhalb der E-Government-Anwendung gespeichert und vom Bund weiterverarbeitet, daher werden sie mangels technischem bzw. automatisiertem Zugriff im Zuge dieser Masterarbeit nicht weiter behandelt.

5.3.6 eAMS

Das eAMS ist das umfassende Portal des AMS, in welches nicht nur sämtliche Unternehmen mit ihren Maßnahmen, sondern auch sämtliche als arbeitslos gemeldeten Personen selbst viele Eintragungen machen. Über das eAMS wird ein Großteil der Geschäftsprozesse in Verbindung mit dem AMS abgewickelt; angefangen von der Projekteinreichung über die Finanzpläne und der täglichen Teilnehmerverwaltung bis hin zur Sicherstellung des Datenschutzes.

Als komplett externe Anwendung wird das eAMS in dieser Masterarbeit nicht weiter behandelt.

5.3.7 „KundInnenakt stationär“

Der „KundInnenakt stationär“ ist eine auf Microsoft SharePoint (Homepage "Microsoft SharePoint", kein Datum) basierende Eigenentwicklung von JAW. Das Ziel dieser Applikation ist die strukturierte Ablage sämtlicher Daten, die im Zuge der Erfüllung von stationären Dienstleistungen erzeugt oder anderweitig verarbeitet werden. In der ersten Ausbaustufe wurden damit die im QMS bestehenden Word-Formulare zur Erfassung der Kundenstammdaten abgebildet sowie an die bestehende Abrechnung mittels BMD NTCS angebunden.

Durch die starke dezentrale Verwendung dieser Applikation ist es möglich, einen größeren Datenstamm zu erzeugen. So ist beispielsweise im Gegensatz zum zugrundeliegenden BMD NTCS im „KundInnenakt stationär“ auch die Wohnadresse der Kunden hinterlegt.

5.3.8 BMD NTCS

BMD NTCS ist eine Finanzbuchhaltungs-, Warenwirtschafts, CRM-, Lohnverrechnungs- und Zeiterfassungssoftware der BMD Systemhaus GesmbH (Homepage "BMD Systemhaus GesmbH", kein Datum). Vom Hersteller selbst wird es als ERP-System angepriesen, als Endkunde stößt man sehr schnell an die Grenzen, die man bei einer ausgewachsenen ERP-Software mit kundenspezifischer Programmierung ausweiten könnte – was bei BMD NTCS nur sehr beschränkt möglich ist. Bei JAW wird BMD NTCS für die oben genannten Aufgaben verwendet.

Zur Kundenstammdatenverwaltung und der Abrechnung mittels BMD NTCS wurde eine eigene Datenbankumgebung eingerichtet, die unabhängig von der produktiven Datenbank arbeitet. Dort werden die Kunden im Mitarbeiterstamm angelegt. Dies hat den Vorteil, dass sehr viele Informationen mit Zeitbezug – als so genannte historische Datensätze – hinterlegt werden können. Damit kann beispielsweise ein Nachfolgebekleid der Bezirksverwaltungsbehörde bereits lange vor dessen Gültigkeit im System hinterlegt werden und erst zum Stichtag werden die Informationen in den produktiven Datenstamm übernommen.

Zur Abrechnung der so genannten Tagsatzleistungen – d.h. all jenen, die als „Datenbank sekundär“ BMD NTCS nutzen – ist es notwendig, die tägliche Anwesenheit der Kunden aufzuzeichnen und für eventuelle Abwesenheiten stark ausdifferenzierte Bezeichnungen zu verwenden. Deshalb werden diese „Mitarbeiter“-Datensätze in der Zeiterfassung verwendet, um pro verrechenbarem Tag eine einstündige Buchung zu hinterlegen, welche die Durchführung der Abrechnung ermöglicht. Die Verwendung der Zeiterfassung hat auch den Vorteil, dass das dahinterliegende Zeitbewertungssystem angewendet werden kann, um etwaige Abschlags-, Zuschlags oder Kontingentberechnungen auf Basis der eingetragenen Mengen durchführen zu können.

Die Bewertung der Zeitbuchungen wird anschließend in ein spezielles Datenformat exportiert, das von der Warenwirtschaft im produktiven BMD NTCS eingelesen werden kann und dann die entsprechenden Rechnungen erzeugt.

Abbildung 12 zeigt das Datenmodell der Stammdaten in BMD NTCS. Wie darin erkenntlich ist, werden einige Felder verwendet, die „Freifeld“ heißen – das bedeutet, dass diesen von JAW ein eigener Name und damit eine eigene Bedeutung zugesprochen wurde. Diese werden auszugsweise in Tabelle 2 dargelegt.

Tabelle	Feld	JAW-spezifische Bezeichnung
DVH_DIENSTVERHAELTNIS	DVH_FREIFELD4	Austrittsgrund
PER_PERSON	PER_FREIFELD4	Debitorennummer
PER_PERSON	PER_FREIFELD5	Projektnummer
PER_PERSON	PER_KOSTENSTELLE	KST
PER_PERSON	PER_PSW_FREIFELDA01_40	Eigenleistung/Stundenkontingent
PER_PERSON	PER_PSW_FREIFELDA02_100	Familienbeihilfe
PER_PERSON	PER_PSW_FREIFELDA03_255	Bescheid
PER_PERSON	PER_PSW_FREIFELDB07	Betreuungszuschlag Senioren
PER_PERSON	PER_PSW_FREIFELDN04_5	Systemerhalter
PER_PERSON	PER_PSW_FREIFELDN05_9	Betreuungszuschlag/Z-Ther
PER_PERSON	PER_PSW_FREIFELDN06_9	Leistungsart lt. Bescheid
PSW_PERSONENWERTE	PSW_FREIFELDA03_255	Bescheid
PSW_PERSONENWERTE	PSW_FREIFELDN06_9	Leistungsart
PSW_PERSONENWERTE	PSW_KOREABTEILUNG	WIPS-Standort

Tabelle 2 Übersicht "umbenannte" Felder BMD NTCS. Quelle: eigene Darstellung (Microsoft Excel)

BMD NTCS ist in der Kundenstammdatenverwaltung organisatorisch gesehen keine führende Datenbank, sondern bekommt die Daten aus anderen Systemen (vgl. dazu Tabelle 1). Technisch gesehen ist BMD NTCS jedoch das führende System, da es mangels Programmierschnittstellen nahezu vollständig händisch gepflegt werden muss – daher wird das vorgelagerte System in der technischen Schnittstelle nachrangig angebunden. Das bedeutet, dass grundsätzlich alle Daten aus BMD NTCS in Richtung führendes System, insbesondere dem „KundInnenakt stationär“, laufend überschrieben werden. Dortige Benutzereingaben führen zu einer Art

„Überarbeitungsauftrag“ für die Mitarbeiter in der Verrechnungsabteilung, damit diese die entsprechenden Änderungen in BMD NTCS durchführen. Während ein solcher Überarbeitungsauftrag bei einem Kunden offen ist, werden keine Daten zu diesem Kunden aus BMD NTCS in Richtung „KundInnenakt stationär“ überschrieben. Damit sind sowohl die organisatorischen als auch die technischen Vorrangregeln gewährleistet.

Somit sind an relevanten Kundendaten strukturiert innerhalb von BMD NTCS gespeichert:

- persönliche Daten (Name, Geschlecht, Geburtsdatum, SVNR)
- Dienstleistung von/bis (=Bescheid mit Leistungsart laut LEVO-StBHG bzw. StKJHG-DVO bzw. auch individuell mit JAW vereinbarte Leistungsarten [Grundversorgung, UMF, Neuland, IBA])
- Standort bei JAW („WIPS-Standort“ = Analog zu WIPS-Datenbank⁶)
- Kostenstelle bei JAW
- Eintritt/Austritt

Hierarchisch betrachtet beginnt die Datenstruktur bei den einzelnen Personen als Ausgangsdatensätze. Diesen Personen werden dann die Dienstleistung sowie der Standort zugewiesen.

⁶ Die WIPS-Datenbank dient lediglich der Erfassung von statistischen Daten über JAW und dessen Standorte bzw. Dienstleistungseinrichtungen für das Land Steiermark, weswegen im Rahmen dieser Masterarbeit nicht näher darauf eingegangen wird.

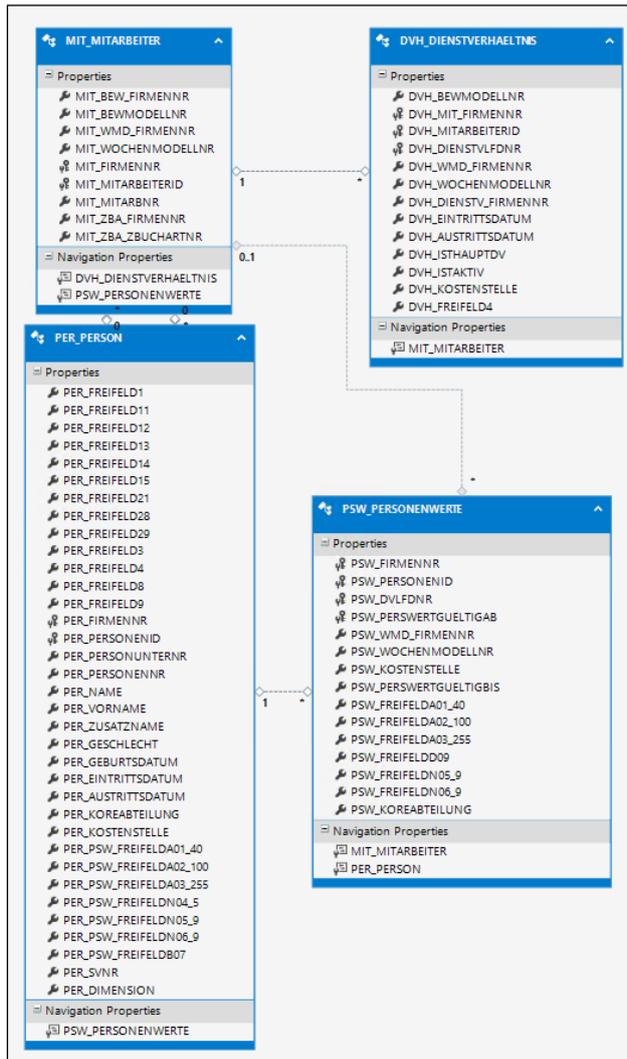


Abbildung 12 Datenmodell der Kundenstammdaten in BMD NTCS. Quelle: eigene Darstellung (Microsoft Visual Studio)

5.3.9 GBPDB

Die „GBPDB“ ist eine Eigenentwicklung von JAW unter Zuhilfenahme eines externen Auftragnehmers. Diese Applikation entstand bei einer großen Auftragsannahme von JAW, für das AMS ein „Gemeinnütziges Beschäftigungsprojekt“ (GBP) durchzuführen. Da für die Nutzung der vorhandenen Datenbankapplikationen für diese Maßnahme umfangreiche Adaptierungsarbeiten notwendig gewesen wären, wurde zur Förderung der Lieferantendiversität eine neue Applikation geschaffen, die genau die Anforderungen eines GBP erfüllt.

Die Datenbank ist mandantenfähig aufgebaut, was zur Folge hat, dass jedes angelegte Projekt komplett unabhängig von den anderen verwaltet werden kann und unterschiedliche Projektstammdaten möglich sind. In Abbildung 13 wird ein Ausschnitt des Datenbankmodells skizziert. Die scheinbar „lose“ Tabelle „Projects“ hält diese Mandantendaten, in jeder Tabelle gibt es die Spalte „ProjectId“, welche auf das entsprechende Projekt verweist.

Wie bereits erwähnt, werden in dieser Anwendung ausschließlich Daten von GBPs verwaltet, was die Klassifizierung der Dienstleistung für das DW einfach macht. Der grobe Datenaufbau richtet sich an der Realität aus, sodass in einem Projekt Transitarbeitskräfte („Employees“) angelegt werden, welche potentiell mehrere Anstellungen („Assignments“) haben können. Durch die zeitlich nach und nach gewachsenen Anforderungen an die Entwicklung ist dies jedoch nie der Fall, da die Eingabe über die Benutzeroberfläche in dieser Form nicht möglich ist. Bei einem etwaigen Wiedereintritt werden die Teilnehmer daher neu angelegt.

Somit sind an relevanten für das DW in der GBPDB vorhanden:

- persönliche Daten (Name, Geschlecht, Geburtsdatum, SVNR, Anschrift)
- Standort bei JAW
- Kostenstelle bei JAW
- Eintritt/Austritt

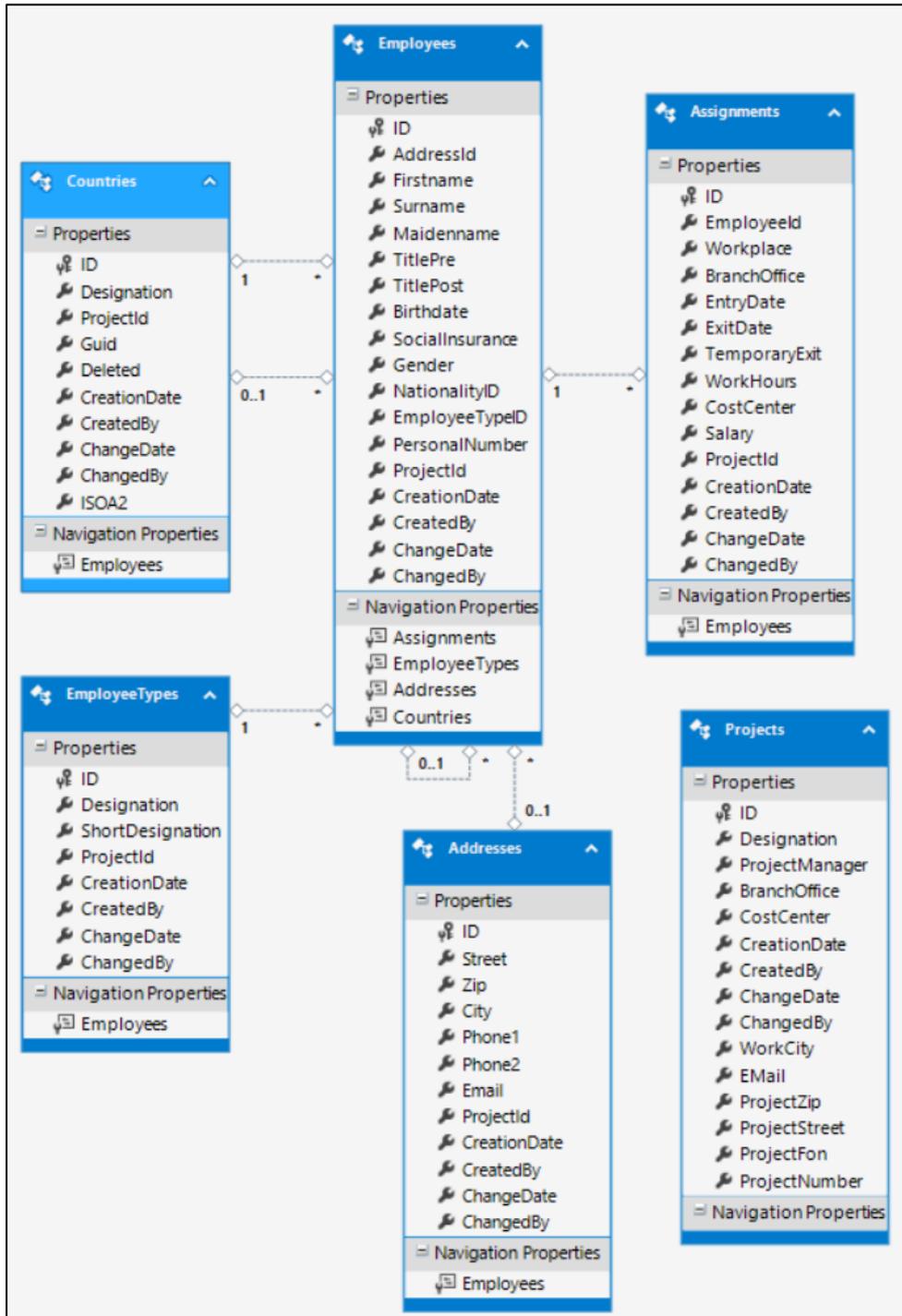


Abbildung 13 Datenbankmodell "GBPDB". Quelle: eigene Darstellung (Microsoft Visual Studio)

6 DIMENSIONALES DATENMODELL

Diese Masterarbeit fokussiert sich auf die Stammdaten der Kunden, die in unterschiedlichen 3NF-Datenmodellen vorliegen. Das zur Implementierung eines DW notwendige dimensionale Datenmodell dieser Daten wird in diesem Kapitel entwickelt.

Da es sich rein um Stammdaten handelt, wirkt es auf den ersten Blick schwierig, eine Faktentabelle (vgl. Kapitel 2.3.4 „Surrogate Key“) zu finden. Allerdings können aus den Änderungen Fakten abgeleitet werden, indem die zugrundeliegenden Geschäftsprozesse herangezogen sowie periodische Schnappschüsse der Daten erstellt werden. In weiterer Folge werden nun zunächst die Dimensionen und anschließend die daraus ableitbaren Faktentabellen erläutert.

6.1 Dimension Zeit

Die erste Dimension in jedem DW ist die Zeitdimension (vgl. 2.2.4 Zeitorientierung (time variance)). Da bei JAW das Geschäftsjahr ident mit dem Kalenderjahr ist, fällt eine Komplexitätsebene weg.

Wie in Tabelle 3 skizziert, wird bei jedem Kalendertag nicht nur das entsprechende Datum mit dem Datentyp „Date“ gespeichert, sondern auch Informationen, die auch via Formatierung oder Berechnung möglich wären. Dazu zählen das ausgeschriebene Datum, die Quartalsnummer, die Monatszahl, die Jahreszahl sowie der ausgeschriebene Wochentag. Diese Informationen ließen sich mithilfe von Formatierungsregeln auch einfach darstellen, allerdings ist es in einer Auswertung einfacher und performanter, diese Felder einfach zur Verfügung zu haben, anstatt sich erst noch um die Formatierung kümmern zu müssen. Es werden auch zusätzliche Informationen zu den einzelnen Tagen in dieser Dimension abgelegt. Die Information, ob es sich bei einem Tag um einen Wochentag handelt oder nicht, ist mit sehr einfachen Regeln feststellbar, jedoch werden diese Informationen aus Gründen der Performance und Benutzerfreundlichkeit bereits berechnet abgespeichert. Weiters wird hinterlegt, ob es sich um einen Feiertag handelt, was bei JAW wenig Komplexität aufweist, da alle Dienstnehmer in der Steiermark arbeiten, wo es einheitliche Feiertage gibt. Zusätzlich sind im Kollektivvertrag der Sozialwirtschaft Österreich (SWÖ-KV) zwei Sonderregeln verankert: Der 24. sowie der 31. Dezember gelten als Feiertage. Diese Information wird auch in dieser Dimension gespeichert, da bei JAW einheitlich der SWÖ-KV zum Einsatz kommt.

Key	Volles Datum	Datum	Wochentag	Monat	Quartal	Jahr	IstFeiertag	IstWochentag
20170102	02. Jänner 2017	02.01.2017	Montag	1	Q1	2017	0	1
20170401	01. April 2017	01.04.2017	Dienstag	4	Q2	2017	0	1
20170701	01. Juli 2017	01.07.2017	Mittwoch	7	Q3	2017	1	1
20170705	05. Juli 2017	05.07.2017	Samstag	7	Q3	2017	0	0

Tabelle 3 Datumsdimension mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel)

Diese vielen Felder eignen sich darüber hinaus für das Definieren von Attributhierarchien, um ein rasches und gezieltes Filtern in Auswertungen zu ermöglichen.

6.2 Dimension Kunde

Da es sich um Stammdaten von Kunden handelt, ist die zweite Dimension der Kunde. Dieser hat mehrere Attribute wie Name, Wohnort und Geburtsdatum. Das Ziel der Arbeit ist es, die Stammdaten aus unterschiedlichen Anwendungsapplikationen in ein EDW zu konsolidieren. Dadurch gibt es keinen natürlichen Schlüssel für die Kundendimension – dieser muss im Zuge der Beladung des Data Warehouse genauso erfunden werden, wie der in einem DW übliche Ersatzschlüssel. Dies ist nötig, um bei weiteren Beladungen sowie bei Auswertungen die einfache Identifikation einer einzelnen Person mithilfe des natürlichen Schlüssels über mehrere Versionen, also mehrere Ersatzschlüssel, hinweg zu ermöglichen, ohne viele Attribute wie Name und Sozialversicherungsnummer gemeinsam hierfür verwenden zu müssen. Das Verwenden von mehreren Attributen als „erfundenen“ natürlichen Schlüssel hätte noch größere Tücken, da sich auch Namen ändern können. Die daraus resultierende Struktur mit Beispieldaten ist in Tabelle 4 dargestellt.

Damit dieser „erfundene natürliche“ Schlüssel auch tatsächlich einen Mehrwert im Zuge von Beladung und Abgleich liefern kann, werden die unterschiedlichen Kundennummern in den verschiedenen Datenquellen – also die dortigen natürlichen Schlüssel – zusätzlich in einer Mini-Dimension „Kundennummer“ gespeichert. (vgl. Kapitel 6.3 „Dimension Kundennummer“)

Da diese Dimension „Kunden“ natürlicherweise eine SCD ist, sind auch die Spalten GültigAb, GültigBis sowie IstGültig erforderlich (vgl. Kapitel 2.3 „Datenmodell“).

Die Attribute des Kunden sind demnach:

- Key: der Surrogate Key des DW
- IdentifierKey: der „erfundene natürliche“ Schlüssel
- Vorname
- Nachname
- Geschlecht

- SVNR
- Wohnort
- PLZ
- Straße
- Geburtsdatum
- GültigAb
- GültigBis
- IstGültig

Key	IdentifierKey	Vorname	Nachname	PLZ	Geburtsdatum_FK	GültigAb	GültigBis	IstGültig
1	1	Max	Mustermann	1234	19930412	20170101	20170228	0
2	1	Max	Mustermann	4543	19930412	20170301	29991231	1
3	2	Hansi	Huber	4321	19870503	20170101	29991231	1
4	3	Maria	Musterfrau	8045	19910205	20170201	29991231	1

Tabelle 4 Aufbau der Dimension "Kunden" mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel)

Zusätzliche Attribute, die aus platztechnischen Gründen in Tabelle 4 nicht dargestellt werden, sind:

- Bezirk: der Bezirk des Wohnortes, um Auswertungen nicht nur auf Basis der adressüblichen Felder Straße, PLZ & Ort, zu ermöglichen
- Bundesland: das Bundesland des Wohnortes. Obwohl JAW nur in der Steiermark tätig ist, ist es durchaus möglich, dass Kunden auch in anderen, meist angrenzenden, Bundesländern wohnen
- Land
- Telefonnummer
- Mobilnummer
- E-Mail-Adresse
- Anrede
- Titel vorgestellt
- Titel nachgestellt
- Staatsbürgerschaft
- Geburtsland

Damit können alle Attribute eines Kunden, welche direkt die Person in der realen Welt betreffen und JAW im Rahmen der operativen Tätigkeit aufnimmt, abgespeichert werden. Alle weiteren Attribute, die nicht mehr direkt die Person, sondern die Betreuung bei JAW betreffen, werden in anderen Dimensionen oder bei den Fakten abgelegt.

6.3 Dimension Kundennummer

Die Dimension Kundennummer dient dazu, die unterschiedlichen natürlichen Schlüssel aus den Quellapplikationen vereinheitlicht zu speichern. Es wäre auch denkbar, die einzelnen Kundennummern direkt in der Dimension Kunde abzulegen, aber aufgrund von mehreren gleichzeitigen Betreuungsverhältnissen⁷ oder Wechseln innerhalb von JAW kann es zu mehreren Kundennummern für dieselbe Person in derselben Datenquelle kommen.

Damit ergibt sich die in Tabelle 5 abgebildete Struktur:

- Key: der Surrogate Key im DW
- NaturalKey: die Kundennummer in der Datenquelle
- Kunde_FK: ein Verweis auf den Surrogate Key der Dimension Kunde, um die zeitlich korrekten Daten abfragen zu können
- IdentifierKey: der „erfundene natürliche“ Schlüssel der Dimension Kunde, um eine Person eindeutig identifizieren zu können
- Datenquelle: der Name der Datenquelle, um einerseits die Beladung des DW im ETL-Prozess zu vereinfachen und andererseits auf Basis der Quellapplikation statistische Auswertungen zu ermöglichen

Key	NaturalKey	Kunde_FK	IdentifierKey	Datenquelle
1	800001	1	1	BMDAW-Kunden
2	800002	2	1	BMDAW-Kunden
3	900001	2	1	BMDAW-Kunden
4	1234	4	3	AABDB

Tabelle 5 Dimension „Kundennummer“ mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel)

⁷ Z.B. ein Kunde bezieht die Leistung TaB und TW gleichzeitig und hat damit im Abrechnungsprogramm BMD NTCS zwei unterschiedliche Kundennummern. In der Applikation „GBPDB“ muss jeder Kunde bei jedem Neueintritt neu angelegt werden, was ebenfalls zu zusätzlichen Kundennummern führt.

6.4 Dimension Dienstleistung

Die Dimension Dienstleistung beinhaltet die Namen der einzelnen Dienstleistungen, mit denen diese in den entsprechenden Datenquellen angelegt wurden. Die Daten aus der Datenquelle BMD NTCS sind Leistungen, die in steierischen Landesgesetzen verankert sind. Diese wurden dort eine entsprechende Leistungsnummer versehen, die auch im DW als NaturalKey verwendet wird. In der Datenquelle AABDB wird jede einzelne AMS-Maßnahme als Dienstleistung abgelegt, auch wenn sich diese eigentlich nur anhand der geographischen Lage unterscheiden. Um sie trotzdem im DW zusammenfassen zu können, wird noch eine Spalte „Typ“ eingeführt, über die beispielsweise alle „überbetrieblichen Lehrausbildungen“ zusammengefasst werden können. Bei den Maßnahmen des Landes Steiermark ist damit die Unterteilung in die einzelnen Gesetze möglich.

Das daraus resultierende Modell wird in Tabelle 6 dargelegt. Folgende Attribute sind darin enthalten:

- Key: der Surrogate Key des DW
- NaturalKey: der natürliche Schlüssel in der Quellapplikation
- Bezeichnung: der Name der Dienstleistung in der Quellapplikation
- BezeichnungKurz: eine Abkürzung für den Namen der Dienstleistung
- Typ: ein Typ zur Gruppierung von ähnlichen Dienstleistungen
- Auftraggeber*
- GültigAb*
- GültigBis*
- IstGültig*
- Datenquelle: der Name der Datenquelle, um einerseits die Beladung des DW im ETL-Prozess zu vereinfachen und andererseits auf Basis der Quellapplikation statistische Auswertungen zu ermöglichen

Die mit * markierten Attribute sind aus Platzgründen in Tabelle 6 nicht aufgeführt.

Key	NaturalKey	Bezeichnung	Bezeichnung Kurz	Typ	Datenquelle
1	4858	Tagesbegleitung & Förderung	B&F	BHG	BMDAW-Kunden
2	4860	Teilhabe an Beschäftigung	TaB	BHG	BMDAW-Kunden
3	34	Überbetriebliche Lehrausbildung Graz	ÜBA-G	ÜBA	AABDB
4	4076	Teilzeitbetreutes Wohnen	TBEW	BHG	BMDAW-Kunden

Tabelle 6 Dimension "Dienstleistung" mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel)

Da sich die Dienstleistungen sowie deren Auftraggeber im Laufe der Zeit verändern, ist diese Dimension ebenfalls als SCD ausgelegt.

6.5 Dimension Kostenstelle

Die Organisation von JAW ist in unterschiedliche Kostenstellen gegliedert, für die von Zweigstellenleitern vor Ort die Verantwortung getragen wird. Daher ist die Kostenstelle als Dimension ein wichtiges Auswertungskriterium. Sämtliche Kostenstellen werden von der Controlling-Abteilung in BMD NTCS angelegt und verwaltet, wodurch das Attribut „Datenquelle“ obsolet ist.

Die in Tabelle 7 ersichtliche Struktur hat folgende Attribute:

- Key: der Surrogate Key des DW
- KST: die Nummer der Kostenstelle; diese ist gleichzeitig der natürliche Schlüssel in der Datenquelle.
- Bezeichnung: der Name der Kostenstelle
- GültigAb
- GültigBis
- IstGültig

Key	KST	Bezeichnung	GültigAb_FK	GültigBis_FK	IstGültig
1	270	KomAZ Liezen	20000101	29991231	1
2	214	KomAZ Knittelfeld	20000101	29991231	1
3	457	ÜBA Graz	20000101	29991231	1
4	219	WV Knittelfeld	20000101	29991231	1

Tabelle 7 Dimension "Kostenstelle". Quelle: eigene Darstellung (Microsoft Excel)

Da sich Kostenstellen im Laufe der Zeit verändern, indem neue dazu kommen, bestehende wegfallen oder sich die Bezeichnung ändert, wird auch diese Dimension als SCD ausgeführt.

6.6 Dimension Standort

Die Organisationsstruktur von JAW erlaubt, dass sowohl ein Standort mehrere Kostenstellen beheimatet, als auch, dass eine Kostenstelle für mehrere Standorte gültig ist. Dadurch ist der Standort als eigene Dimension wichtig, um auch auf den Aspekt des Standorts hin Auswertungen zu ermöglichen – wenn gewünscht, auch über mehrere Kostenstellen hinweg.

Als örtliche Gegebenheit hat ein Standort eine Adresse, eine PLZ und einen Ort. Wie bei der Dimension „Kunde“ bietet es sich daher an, die Attribute Bezirk und Land automatisch zu ergänzen, um auch regionsbezogene Auswertungen zu ermöglichen. Alle Standorte werden einheitlich in der Datenquelle BMD NTCS angelegt, was das Attribut Datenquelle auch hier obsolet macht.

Daraus resultiert folgende Struktur:

- Key: der Surrogate Key des DW
- NaturalKey: die Standortnummer in BMD NTCS
- Bezeichnung: der JAW-interne Name des Standortes; häufig gleich dem Namen der dort überwiegend vorherrschenden Kostenstelle
- BezeichnungKurz: eine Kurzform des Namens
- Straße*
- PLZ
- Ort
- Bezirk*
- Bundesland*
- Land*
- GültigAb*
- GültigBis*
- IstGültig*

Die mit * markierten Attribute sind aus Platzgründen in Tabelle 8 nicht aufgeführt.

Key	NaturalKey	Bezeichnung	BezeichnungKurz	PLZ	Ort
1	270001	Kompetenz- & Ausbildungszentrum Liezen	KomAZ Liezen	1234	Liezen
2	214001	Kompetenz- & Ausbildungszentrum Knittelfeld	KomAZ Knittelfeld	1234	Knittelfeld
3	457001	Ausbildungszentrum Graz	AZ Graz	8045	Graz
4	219001	Wohnverbund Knittelfeld	WV Knittelfeld	8423	Knittelfeld
5	214001	Kompetenz- & Ausbildungszentrum Knittelfeld	KomAZ Knittelfeld	1234	Knittelfeld
6	219001	Wohnverbund Knittelfeld	WV Knittelfeld	8423	Knittelfeld

Tabelle 8 Dimension "Standort". Quelle: eigene Darstellung (Microsoft Excel)

Da sich auch Gebäude und deren Bezeichnung im Laufe der Zeit verändern können, wird auch diese Dimension als SCD realisiert.

6.7 Fakten Kundenstamm

Um aus mehr oder minder statischen Stammdaten eine Faktentabelle zur Auswertung zu bekommen, werden die Transaktionen, die zu diesen Daten geführt haben, rekonstruiert. Das bedeutet, dass beispielsweise eine Datenbankzeile einer Quellapplikation, die das Feld „Eintrittsdatum“ beinhaltet, für die Transaktion „Eintritt“ steht. Die meisten anderen Attribute dieser Quelldatenbankzeile werden in den zuvor beschriebenen Dimensionen gespeichert. Die zweite Transaktion für diese Faktentabelle ist „Austritt“, die aus dem Attribut „Austrittsdatum“ ableitbar ist. Als dritter Transaktionstyp ist „Änderung“ zu nennen, der zumeist eine neue Version in der einen oder anderen Dimension erzeugt. Zum Beispiel verursacht die Änderung des Wohnortes eine neue Version in der Dimension Kunde.

Die daraus resultierende Struktur, skizziert in Tabelle 9, beinhaltet folgende Attribute:

- Key: fortlaufender Schlüssel im DW
- Transaktion: die Art der Transaktion (Eintritt/Austritt/Änderung)
- Datum_FK: der Zeitpunkt, an dem die Transaktion stattgefunden hat
- Kunde_FK: Referenz auf Dimension Kunde
- Kundennummer_FK: Referenz auf Dimension Kundennummer
- Dienstleistung_FK: Referenz auf Dimension Dienstleistung
- Standort_FK*: Referenz auf Dimension Standort
- Kostenstelle_FK*: Referenz auf Dimension Kostenstelle
- Alter*: das Alter des Kunden berechnet aus der Differenz zwischen dem Datum laut Datum_FK und dem Geburtsdatum laut Kunde_FK
- Zähler*: dient zur einfacheren Zählung von Ergebnissen in Auswertungen (vgl. 2.3.5 Faktentabellen)
- Datenquelle: der Ursprung dieses Eintrags

Die mit * markierten Attribute sind aus Platzgründen in Tabelle 9 nicht aufgeführt.

Die Referenzen auf Kunde und Kundennummer erzeugen streng genommen eine starke Redundanz. Um Auswertungen sowohl in Richtung realer Person als auch in Richtung Kunde bei JAW einfach zu ermöglichen, ist es jedoch von Vorteil, beide hier abzuspeichern.

Key	Transaktion	Datum_FK	Kunde_FK	Kundennummer_FK	Dienstleistung_FK	Datenquelle
1	Eintritt	20170101	1	1	1	BMDAW-Kunden
2	Änderung	20170201	1	1	1	BMDAW-Kunden
3	Austritt	20170228	1	1	1	BMDAW-Kunden
4	Eintritt	20170301	2	2	2	BMDAW-Kunden
5	Eintritt	20170301	2	3	3	BMDAW-Kunden
6	Eintritt	20170401	4	4	4	AABDB

Tabelle 9 Faktentabelle "Kundenstamm". Quelle: eigene Darstellung (Microsoft Excel)

In der täglichen Arbeit mit den beschriebenen Datenquellen hat sich gezeigt, dass sich sämtliche Stammdaten, die ein Außenstehender als fixe Werte ansehen würde, durchaus ändern können. Dies sind insbesondere das Eintrittsdatum und das Austrittsdatum, die sich durch äußere Umstände ändern können. Daher ist es bis zu einem gewissen, näher zu definierenden, Zeitpunkt notwendig, die Faktentabelle bei jeder Beladung des DW komplett neu zu befüllen und keine inkrementellen Beladungen durchzuführen. Die Änderung des Austrittsdatums passiert beispielsweise häufig bei Teilnehmern von GBPs, die grundsätzlich mit einem befristeten Dienstverhältnis bei JAW angestellt werden, deren Befristung in einigen Fällen jedoch verlängert wird. Außerdem kann es sein, dass noch während dieses befristeten Dienstverhältnisses eine Arbeitsstelle gefunden werden kann, wodurch sich eine einvernehmliche Lösung ergibt, die vor dem „geplanten“ Austrittsdatum liegt.

Aus diesen Fakten heraus ist es möglich, folgende Fragen zu beantworten:

- Wie viele Kunden sind eingetreten?
- Wie viele Kunden sind ausgetreten?

Beide einschränkbar auf alle Dimensionen, also

- nach realen Personen, dadurch:
 - nach Herkunftsbezirk
 - nach Geburtsland
- nach Kundennummer, dadurch:
 - nach Datenquelle
- nach Zeitraum
- nach Standort, dadurch:
 - nach Bezirk der Dienstleistungserbringung
- nach Kostenstelle

- nach Dienstleistung, dadurch:
 - nach Auftraggeber
 - nach Typ

6.8 Fakten Kundenstamm-Snapshot Stichtag

Aus der zuvor beschriebenen transaktionsorientierten Faktentabelle können Snapshots abgeleitet werden, um weitere Statistiken zu erzeugen. Dazu werden diese Transaktionen in Relation zur Zeitdimension gesetzt, um die Frage „wie viele Kunden waren bei JAW?“ beantworten zu können.

Dazu sind folgende Attribute erforderlich:

- Key: fortlaufender Schlüssel im DW
- Datum_FK: Referenz auf Dimension Zeit
- Kunde_FK: Referenz auf Dimension Kunde
- Kundennummer_FK: Referenz auf Dimension Kundennummer
- Dienstleistung_FK: Referenz auf Dimension Dienstleistung
- Standort_FK: Referenz auf Dimension Standort
- Kostenstelle_FK: Referenz auf Dimension Kostenstelle
- Zähler: dient zur einfacheren Zählung von Ergebnissen in Auswertungen (vgl. 2.3.5 Faktentabellen)
- Alter: das Alter des Kunden berechnet aus der Differenz zwischen dem Datum laut Datum_FK und dem Geburtsdatum laut Kunde_FK

In dieser Tabelle gibt es dadurch eine Zeile pro:

- Kalendertag
- Realer Person
- Kundennummer
- Dienstleistung
- Standort
- Kostenstelle

Damit lässt sich die Frage auch noch nach den folgenden Kriterien verfeinern:

- nach realen Personen, dadurch:

- nach Geburtsland
- nach Kundennummer, dadurch:
 - nach Datenquelle
- nach Datum (Stichtag)
- nach Standort, dadurch:
 - nach Bezirk der Dienstleistungserbringung
- nach Kostenstelle
- nach Dienstleistung, dadurch:
 - nach Auftraggeber
 - nach Typ
- nach Alter

Das berechnete Alter ermöglicht Zielgruppenanalysen. Damit diese effizient vonstattengehen, ist die Definition einer Hierarchie auf Basis der Altersgruppen vorteilhaft. Beispielsweise könnten alle Werte kleiner als 18 als „minderjährig“ und alle anderen als „volljährig“ in die Hierarchie „Volljährigkeit“ aufgenommen werden.

Bei der Auswahl eines Stichtages ist die damit verbundene Anzahl eindeutig. Wählt man einen längeren Zeitraum, z.B. ein Monat, muss aus dieser Anzahl noch der Durchschnitt über den Zeitraum gebildet werden.

7 TIEFERE ANALYSE & PROTOTYP

Im Folgenden wird im Detail auf die Datenquellen BMD und AABDB sowie eine Analyse dieser aktuell vorhandenen Daten eingegangen, in dessen Zuge auch die Besonderheiten der Datenquellen dargelegt werden. Diese Datenquellen wurden ausgewählt, da BMD eine komplexe Datenquelle mit zusätzlichen externen Daten darstellt, wodurch die Flexibilität des Beladungsprozesses dargelegt wird. AABDB hat eine einfache Datenstruktur, die jedoch in einer hohen Anzahl an Dubletten resultiert, was somit einen weiteren essentiellen Aspekt des ETL-Prozesses darstellt.

Anschließend wird der ETL-Prozess für diese Datenquellen in Form eines Prototyps durchskizziert sowie eine Analyse der gewonnenen Daten durchgeführt.

7.1 Datenquellen BMD & „KundInnenakt stationär“

Wie in Kapitel 5.3.8 „BMD NTCS“ beschrieben, ist es bei der Datenquelle BMD ausreichend, in Hinblick auf das Ziel dieser Masterarbeit – dem Verarbeiten der Kundendaten – vier Tabellen zu betrachten. Die Besonderheiten sowie die Strukturen dieser Tabellen werden nun näher betrachtet.

7.1.1 Besonderheiten

Die Datenbank für die Kundendaten stellte ursprünglich eine Kopie der Produktivdatenbank mit den Daten der Angestellten von JAW dar. Durch diesen Umstand und dadurch, dass die Kunden als Mitarbeiter angelegt werden, sind in den entsprechenden Tabellen einige Daten vorhanden, die nicht als Kundendaten gelten und herausgefiltert werden müssen. Dies lässt sich sehr einfach realisieren, da JAW-intern definiert ist, dass alle angelegten Kunden Personalnummern größer gleich 100.000 besitzen.

In der Regel verfügen alle Kunden über eine Zeile in PER_PERSON und eine Zeile in DVH_DIENSTVERHAELTNIS mit demselben Eintritts- und Austrittsdatum, jeweils pro Dienstleistungsart und pro Organisationseinheit, in der diese Dienstleistung in Anspruch genommen wurde. Durch einen Austritt und anschließenden Wiedereintritt in die gleiche oder eine ähnliche Leistung in derselben Organisationseinheit kann es jedoch sein, dass eine PER_PERSON – und damit die Kundennummer – wiederverwendet wird, allerdings eine neue Zeile in DVH_DIENSTVERHAELTNIS entsteht. In der Tabelle PSW_PERSONENWERTE entstehen durch viele Einflüsse neue Zeilen. Nachfolgend wird eine nichtvollständige Aufzählung dieser Einflüsse dargelegt.

- Änderung der bezogenen Leistung durch neuen Bescheid
- Änderung der Kostenstelle durch organisatorische Veränderungen
- Änderung des Verantwortlichen für die Kostenstelle (Feld „Abteilungsleiter“)
- Änderung des Standorts durch operative Gegebenheiten (Feld „KORE-Abteilung“ bzw. „WIPS-Standort“)

Aufgrund der starken Abrechnungsorientierung und der daher nur darauf ausgerichteten Datenhaltung in NTCS sind viele klassische Kundendaten dort nicht erfasst, beispielsweise die Anschrift. Das liegt daran, dass die Rechnung für diese Kunden nicht an die Kunden selbst gehen, sondern an eine Bezirksverwaltungsbehörde, für die in der Finanzbuchhaltung sehr wohl sämtliche Daten gespeichert sind. Daher ist es notwendig, auch Daten aus dem „KundInnenakt stationär“ in das DW zu laden, damit die Anschrift in der Dimension Kunden befüllt werden kann. Da die dortigen Daten aber komplett auf den Daten in NTCS basieren, sind Dublettenprüfung und Datenbereinigungen nur bei NTCS notwendig, aus dem „KundInnenakt stationär“ werden lediglich zusätzliche Informationen abgerufen.

7.1.2 Datenfehler

Für den Betrieb des DW und die Korrektheit der Daten ist es essentiell, dass es keine Ungereimtheiten in den Daten gibt, wie beispielsweise Überlappungen im Gültigkeitszeiträumen. Dies lässt sich einfach überprüfen, indem man für eine Tabelle mit Gültigkeitszeiträumen eine JOIN-Operation auf sich selbst ausführt und die Zeiträume entsprechend abgleicht. Außerdem kann es passieren, dass das Eintritts- und Austrittsdatum der Tabelle DVH_DIENSTVERHAELTNIS nicht zwischen Eintritts- und Austrittsdatum der Tabelle PER_PERSON liegt, oder auch der Gültigkeitszeitraum der Einträge in PSW_PERSONENWERTE nicht zwischen Eintritts- und Austrittsdatum der Tabelle DVH_DIENSTVERHAELTNIS liegen.

In einem ETL-Produktivsystem müsste das System selbst diese Abfragen durchführen und einen Systemadministrator oder eine für die Datenquelle verantwortliche Person über etwaige Datenfehler informieren und diesen bereinigen lassen, da automatisiert nicht korrekt festgestellt werden kann, auf welche Art und Weise diese Einträge zu bereinigen sind. Im Zuge dieser Masterarbeit wurden die Datenfehler vom Autor direkt in der Datenquelle bereinigt.

7.1.2.1. Überlappende Zeiträume

Ein Zeitraum A überlappt sich dann mit einem anderen Zeitraum B, wenn das Enddatum des Zeitraums A nach dem Anfangsdatum des Zeitraums B liegt und das Anfangsdatum des

Zeitraums A vor dem Enddatum des Zeitraums B liegt. In SQL ausgedrückt und auf die Tabelle DVH_DIENSTVERHAELTNIS angewendet, ergibt sich Listing 1.

Zusätzliche angewandte Bedingungen sind: Es muss sich um dieselbe Person handeln, also DVH_MIT_FIRMENNR sowie DVH_MITARBEITERID müssen gleich sein, es darf jedoch nicht derselbe Datensatz sein, DVH_DIENSTVLFDR muss also unterschiedlich sein. Diese drei Spalten sind auch der Primärschlüssel in der Datenquelle. Außerdem wird die zuvor erwähnte Bedingung angewandt, dass die Personalnummer größer gleich 100.000 sein muss. Darüber hinaus stellen nichtvorhandene Werte beim Enddatum, also NULL in der Datenbank, eine unendliche Gültigkeit dar. Diese Regel wird mittels „Enddatum ist NULL oder“ in SQL implementiert.

```
SELECT
    A.DVH_MITARBEITERID,
    A.DVH_EINTRITTSDATUM EinA,
    A.DVH_AUSTRITTSDATUM AusA,
    B.DVH_EINTRITTSDATUM EinB,
    B.DVH_AUSTRITTSDATUM AusB
FROM BUERO.DVH_DIENSTVERHAELTNIS A, BUERO.DVH_DIENSTVERHAELTNIS B
JOIN BUERO.PER_PERSON ON PER_PERSONENID = DVH_MITARBEITERID AND PER_FIRMENNR =
DVH_MIT_FIRMENNR
WHERE
PER_PERSONENNR >= 100000
AND (A.DVH_AUSTRITTSDATUM IS NULL OR A.DVH_AUSTRITTSDATUM >= B.DVH_EINTRITTSDATUM)
AND (B.DVH_AUSTRITTSDATUM IS NULL OR A.DVH_EINTRITTSDATUM <= B.DVH_AUSTRITTSDATUM)
AND A.DVH_MIT_FIRMENNR = B.DVH_MIT_FIRMENNR
AND A.DVH_MITARBEITERID = B.DVH_MITARBEITERID
AND A.DVH_DIENSTVLFDR <> B.DVH_DIENSTVLFDR
```

Listing 1 SQL-Abfrage zur Identifizierung überlappender Zeiträume in der Tabelle DVH_DIENSTVERHAELTNIS

Zum Zeitpunkt der Verfassung dieser Arbeit ergab diese Abfrage keine Ergebnisse, es sind also keine überlappenden Einträge vorhanden.

Dieselbe Abfrage wird auf die Tabelle PSW_PERSONENWERTE umgelegt in Listing 2 dargestellt.

```
SELECT
    A.PSW_PERSONENID,
    A.PSW_PERSWERTGUELTIGAB AbA,
    A.PSW_PERSWERTGUELTIGBIS BisA,
    B.PSW_PERSWERTGUELTIGAB AbB,
    B.PSW_PERSWERTGUELTIGBIS BisB
FROM BUERO.PSW_PERSONENWERTE A, BUERO.PSW_PERSONENWERTE B
JOIN BUERO.PER_PERSON ON PER_PERSONENID = PSW_PERSONENID AND PER_FIRMENNR = PSW_FIRMENNR
WHERE
PER_PERSONENNR > 100000
AND (A.PSW_PERSWERTGUELTIGBIS IS NULL OR A.PSW_PERSWERTGUELTIGBIS >=
B.PSW_PERSWERTGUELTIGAB)
AND (B.PSW_PERSWERTGUELTIGBIS IS NULL OR A.PSW_PERSWERTGUELTIGAB <=
B.PSW_PERSWERTGUELTIGBIS)
AND A.PSW_FIRMENNR = B.PSW_FIRMENNR
AND A.PSW_PERSONENID = B.PSW_PERSONENID
AND A.PSW_DVLFDR = B.PSW_DVLFDR
```

```
AND A.PSW_PERSWERTGUELTIGAB <> B.PSW_PERSWERTGUELTIGAB
```

Listing 2 SQL-Abfrage zur Identifizierung überlappender Zeiträume in der Tabelle PSW_PERSONENWERTE

Diese Abfrage ergab zum Zeitpunkt der Abfassung dieser Arbeit 26 Ergebnisse, die vom Autor vor der weiteren Durchführung der Arbeit im Quellsystem bereinigt wurden.

Die anderen Tabellen von NTCS haben eine Struktur, die überlappende Zeiträume gar nicht zulässt.

7.1.2.2. Ungültige Zeitraumeinträge

Wie zuvor erwähnt, muss jeder Eintrag in der Tabelle DVH_DIENSTVERHAELTNIS zwischen Eintritts- und Austrittsdatum der Tabelle PER_PERSON liegen. Die SQL-Abfrage Listing 3 überprüft diese Einschränkung.

```
SELECT
    PER_PERSONENID,
    PER_EINTRITTSDATUM,
    PER_AUSTRITTSDATUM,
    DVH_EINTRITTSDATUM,
    DVH_AUSTRITTSDATUM
FROM BUERO.PER_PERSON
JOIN BUERO.DVH_DIENSTVERHAELTNIS ON PER_PERSONENID = DVH_MITARBEITERID AND PER_FIRMENNR =
    DVH_MIT_FIRMENNR
WHERE
    PER_PERSONENNR > 100000
AND (
    (PER_AUSTRITTSDATUM IS NOT NULL AND DVH_AUSTRITTSDATUM IS NULL)
    OR DVH_AUSTRITTSDATUM > PER_AUSTRITTSDATUM
    OR PER_EINTRITTSDATUM > DVH_EINTRITTSDATUM
)
```

Listing 3 SQL-Abfrage zur Identifizierung von Einträgen in DVH_DIENSTVERHAELTNIS, die außerhalb des Zeitraums von PER_PERSON liegen.

Auch hier gab es Ergebnisse, die vom Autor bereinigt wurden.

Weiters dürfen die Gültigkeitszeiträume von PSW_PERSONENWERTE nicht außerhalb des Zeitraums von DVH_DIENSTVERHAELTNIS liegen, dies wird mit der Abfrage in Listing 4 überprüft.

```
SELECT
    PER_PERSONENID,
    DVH_EINTRITTSDATUM,
    DVH_AUSTRITTSDATUM,
    PSW_PERSWERTGUELTIGAB,
    PSW_PERSWERTGUELTIGBIS
FROM BUERO.PER_PERSON
JOIN BUERO.DVH_DIENSTVERHAELTNIS ON PER_PERSONENID = DVH_MITARBEITERID AND PER_FIRMENNR =
    DVH_MIT_FIRMENNR
JOIN BUERO.PSW_PERSONENWERTE ON DVH_MITARBEITERID = PSW_PERSONENID AND DVH_MIT_FIRMENNR =
    PSW_FIRMENNR AND DVH_DIENSTVLFDR = PSW_DVLFDR
WHERE
    PER_PERSONENNR > 100000
AND (
```

```
(DVH_AUSTRITTSDATUM IS NOT NULL AND PSW_PERSWERTGUELTIGBIS IS NULL)  
OR PSW_PERSWERTGUELTIGBIS > DVH_AUSTRITTSDATUM  
OR DVH_EINTRITTSDATUM > PSW_PERSWERTGUELTIGAB  
)
```

Listing 4 SQL-Abfrage zur Identifizierung von Einträgen in PSW_PERSONENWERTE, die außerhalb des Zeitraums von DVH_DIENSTVERHAELTNIS liegen

Bei dieser Überprüfung kam es ebenfalls zu Datenfehlern, die vom Autor behoben wurden.

7.1.3 Late Data Entry / Rückwirkende Daten

Ein DW sollte in der Regel jede Änderung an Daten in Produktivsystemen, insbesondere Stammdaten, die im DW als Dimensionen abgebildet werden, nachvollziehbar historisch mitführen, damit dieselbe Auswertung zu einem späteren Zeitpunkt dasselbe Ergebnis liefert (vgl. Kapitel 2.3.3 „Slowly Changing Dimensions“).

Im Falle der Stammdaten in BMD NTCS ist es jedoch so, dass die Stammdaten selbst bereits mehr oder weniger historisiert geführt werden, da die Einträge einen Gültigkeitszeitraum besitzen. Diese historisierten Daten können sich jedoch ebenfalls ändern – beispielsweise wird das Ende von Bescheiden sobald es bekannt ist, von den zuständigen Mitarbeitern in die PSW_PERSONENWERTE eingetragen und sofort ein nachfolgender, unendlich gültiger Eintrag mit der Leistungsart „unbescheidet“ angefügt. Dieser Feldwert, der die Art der Dienstleistung angibt, ändert sich, sobald der neue Bescheid ausgestellt wurde. Diese Art der Datenhaltung ist notwendig, da es regelmäßig vorkommt, dass eine Betreuung durchgeführt wird, obwohl noch kein Bescheid vorhanden ist, dieser jedoch nachgereicht wird (vgl. Kapitel 5.2.2 „Eintritt“). Dies ändert die Vergangenheit – und somit potentiell auch Auswertungen. Allerdings ist dieser neue Bescheid in der Realität auch tatsächlich rückwirkend gültig, was das Ergebnis von entsprechenden Auswertungen verändern muss.

7.2 Datenquelle AABDB

In Kapitel 5.3.2 wurde bereits erläutert, dass für die Umsetzung des Ziels dieser Masterarbeit der „AABDB“ in lediglich zwei Tabelle relevant sind. Die „JAW_Massnahmen“ wird zur Ablage der Projektdaten sowie die Tabelle „TeilnehmerInnen“ zur Ablage der Teilnehmerinformationen herangezogen. Für die Durchführung des ETL-Prozesses sind folgende Besonderheiten in diesen Daten zu beachten.

7.2.1 Besonderheiten

In dieser Applikation werden die einzelnen Maßnahmen scharf voneinander getrennt, diese sind wie bei einer Mandantenverwaltung in großen ERP-Systemen voneinander abgeschottet. Da eine

Maßnahme immer eine Projektnummer mit einem bestimmten Projektzeitraum beim Auftraggeber hat, werden diese auch bei einer Verlängerung in der Anwendung neu angelegt. Es kommt bei diesen Verlängerungen regelmäßig vor, dass Teilnehmer übernommen werden, was in einer Kopie des Datensatzes resultiert. Zusätzlich wird für einen Teilnehmer dann ein neuer Eintrag angelegt, wenn er während des Projektzeitraums wieder eintritt. Dieses Vorgehen resultiert in einer hohen Zahl an Dubletten bei den Teilnehmern.

7.2.2 Datenfehler

Da es keine Zeitraumzuweisungen, sondern nur ein Einstiegs- sowie ein Ausstiegsdatum in der AABDB gibt, können daraus keine Datenfehler mit überlappenden Zeiträumen entstehen, wie sie bei NTCS existieren könnten. Allerdings ist die Unvollständigkeit der Daten, insbesondere in Hinblick auf die Konsolidierung der Dubletten, hier stärker ausgeprägt.

Listing 5 stellt eine Abfrage zur Identifizierung jener Datensätze dar, die keine Sozialversicherungsnummer oder kein Geburtsdatum eingetragen haben. Das Ergebnis dieser Abfrage ist, dass bei einer Datenmenge von 9267 Datensätzen 900 Stück unvollständig sind, was einem Anteil von 9,7% entspricht. Für diese Datensätze kann keine ausreichend genaue Dublettenprüfung erfolgen.

```
DECLARE @gesamt INT,@unvollstaendig INT

SELECT @unvollstaendig = count(*) FROM TeilnehmerInnen WHERE snr IS NULL OR geburtsdatum
IS NULL
SELECT @gesamt = COUNT(*) FROM TeilnehmerInnen

SELECT @gesamt AS Gesamt,@unvollstaendig AS
Unvollstaendig,1.0*@unvollstaendig/@gesamt*100 AS Prozent
```

Listing 5 SQL-Abfrage zur Identifizierung der Menge unvollständiger Datensätze

Für den weiteren Verlauf der Masterarbeit werden diese Datenfehler ignoriert und die Inkonsistenz im DW akzeptiert. In einer Produktivumgebung müssten die Daten in der „AABDB“ bereinigt werden, um korrekte Auswertungen aus dem DW zu ermöglichen.

7.3 Prototyp

In diesem Kapitel wird der ETL-Prozess für die zuvor beschriebenen Datenquellen erläutert. Zu diesem Zweck wird zunächst eine Staging Area beladen, um die Daten aus dem Quellsystem zu extrahieren. Anschließend werden die Daten in das dimensionale Modell transformiert, um diese in weiterer Folge in das DW zu laden. Nach der Durchführung des ETL-Prozesses wird ein Würfel designt und für die weitere Analyse bereitgestellt. Abschließend erfolgt ein erster Einstieg in die Analyse des damit gewonnenen Datenstandes.

7.3.1 Staging Area beladen

Zur Vorbereitung der Staging Area wird eine eigene SQL Server Datenbank mit dem Namen „DW_Staging“ angelegt. Mithilfe eines SSIS-Projektes wird diese Staging Area befüllt. In diesem SSIS-Projekt gibt es folgende Pakete:

- 0_Master: Manager zur Festlegung der Reihenfolge der weiteren Pakete
- 1_StageBMDAW
- 2_StageAABDB
- 3_StageKundenakt

Im ersten Paket „0_Master“ werden lediglich nacheinander alle folgenden SSIS-Pakete ausgeführt. Dies führt dazu, dass bei der automatischen Ausführung mittels SQL Server Agent nur ein einzelnes Paket geplant werden muss, welches alle weiteren Pakete aufruft.

Das zweite Paket „1_StageBMDAW“ sowie das dritte Paket „2_StageAABDB“ übertragen die Daten aus der BMDAW-Kunden-Datenbank bzw. aus der AABDB in die Staging Area. Dazu gibt es für jede Tabelle (vgl. Kapitel 5.3.8 „BMD NTCS“ bzw. Kapitel 5.3.2 „AABDB“) einen „Execute SQL Task“ sowie einen „Data Flow Task“. Abbildung 14 zeigt das Ergebnis für das zweite Paket.

Der „Execute SQL Task“ führt ein „Truncate“ auf die entsprechende Staging-Tabelle aus, d.h. er leer diese komplett aus, damit keine Altdaten mehr vorhanden sind, bevor die Daten erneut aus der ursprünglichen Datenquelle gelesen werden. Der „Data Flow Task“ besteht lediglich aus einer „OLE DB Source“, welche die Zeilen aus der ursprünglichen Datentabelle abfragt sowie einer „OLE DB Destination“, welche die Zeilen in die entsprechende Staging Tabelle überträgt. Im Zuge der Konfiguration der Staging Tabellen bei der „OLE DB Destination“ wird diese auch gleich aus SSDT heraus erstellt, es ist also ein impliziter Schritt. Die „OLE DB Source“ schränkt die abzufragenden Daten auch auf diejenigen ein, die tatsächlich Produktivdaten in der Datenquelle sind (vgl. Kapitel 7.1.1 „Besonderheiten“ sowie Kapitel 7.2.1 „Besonderheiten“).

Durch diese Schritte befinden die Daten in der Staging Area und können nun weiterbearbeitet werden.



Abbildung 14 SSIS-Paket "1_StageBMDAW". Quelle: eigene Darstellung (Microsoft SQL Server Data Tools)

Das vierte Paket „3_StageKundenakt“ liest die Daten aus dem „KundInnenakt stationär“ aus. Die Daten sind dort in vier SharePoint Listen mit den Namen „BHG Stammdatenblatt“, „BHG KomAZ“, „BHG Wohnen“ sowie „KJHG Stammdaten“ gespeichert. Der Ablauf ist grundsätzlich derselbe wie bei den vorigen beiden Paketen, allerdings ist die Datenquelle eine „ODATA Data Source“, da der Zugriff auf Microsoft SharePoint aus SSIS heraus auf diese Art am effizientesten möglich ist.

7.3.2 Staging transformieren mit SSIS

Nun werden für jede Datenquelle die Daten in das dimensionale Datenmodell des DW transformiert. Dazu werden die Daten mithilfe von Views in das entsprechende Format gebracht, wobei folgende Aspekte berücksichtigt werden:

- Dimensionales Datenmodell
- Datentypen der Felder
- Vereinheitlichung der Datenwerte (z.B. Straßennamen)

An dieser Stelle wird noch keinerlei Rücksicht auf das Identifizieren und Konsolidieren von Duplikaten gelegt, da dies über alle Datenquellen hinweg bei der Beladung des DW durchgeführt wird.

Die einzelnen Dimensionen werden zunächst innerhalb einer einzelnen Datenquelle in das Spaltenformat des DW transformiert. Anschließend werden Gesamtviews auf die Dimensionen erzeugt, indem die einzelnen Dimensionsviews mithilfe von UNION in eine einzige View vereinheitlicht, welche dann vom nachfolgenden „Load“ weiterverwendet werden, um die Dublettenprüfung und Abwicklung der SCDs durchzuführen.

Im dimensionalen Datenmodell sind drei der sechs Dimensionen direkt von einer bestimmten Datenquelle abhängig. Das sind die Dimensionen „Dienstleistung“, „Kunde“ und „Kundennummer“. Die anderen drei Dimensionen „Kostenstelle“, „Standort“ und „Zeit“ sind universell eingesetzte Dimensionen.

Die datenquellenunabhängige Dimension „Kostenstelle“ bezieht ihre Daten direkt aus der produktiven Datenbankinstanz von BMD NTCS, wo das Controlling die aktuelle Liste der Kostenstellen pflegt. Mithilfe eines MERGE-Statements werden diese unter Berücksichtigung der SCD in die entsprechende DW-Tabelle überführt. Dasselbe gilt für die Dimension „Standort“, wobei diese im Rahmen dieser Masterarbeit nicht konsolidiert in das DW überführt wird, sondern lediglich ein Beispieleintrag eingefügt wird, der in allen Fakten verwendet wird. Das liegt daran, dass zum Zeitpunkt der Erstellung dieser Arbeit gerade eine Datenbankapplikation bei JAW in Entwicklung ist, welche die Datenhaltung über die einzelnen Standorte vereinheitlichen wird, das Projekt aber noch nicht abgeschlossen ist. Die Dimension „Zeit“ wird in Kapitel 7.3.3 „Data Warehouse beladen“ einmalig erstellt und wird aufgrund ihrer Unveränderlichkeit nicht vollautomatisch beladen, sondern das Skript zur Befüllung händisch ausgeführt.

Nun wird auf die Besonderheiten bei der Transformation der datenquellenabhängigen Dimensionen näher eingegangen.

Die Dimension Dienstleistung der AABDB ist lediglich eine andere Darstellung der Tabelle „JAW_Massnahmen“ und wird mit einer sehr einfachen Abfrage in das dimensionale Modell überführt, vgl. Listing 6.

```
CREATE VIEW [AABDB].[DimDienstleistung]
AS
SELECT
    MassnahmePK NaturalKey,
    Bezeichnung Bezeichnung,
    Gruppe BezeichnungKurz,
    Gruppe Typ,
    'AMS' Auftraggeber,
    Start GueltigAb,
    Ende GueltigBis,
    1 IstGueltig,
    'AABDB' Datenquelle
FROM JAW_Massnahmen
```

Listing 6 SQL-Abfrage zur Erstellung der Transformations-View für die Dimension "Dienstleistung" der Datenquelle "AABDB".

Bei der Datenquelle „BMD“ gestaltet sich die Transformation ähnlich simpel, allerdings muss hier erst eine Hierarchieebene zur Unterscheidung der Dienstleistungsarten eingeführt werden. Dies ist notwendig, um unter anderem die Trennung nach den Geschäftsbereichen AAB und BFW zu ermöglichen, welche sich beide „BMD“ als Abrechnungsinstrument teilen. Diese Transformation ist in Anhang 3: Skript zur Anlage der Transformation „Dienstleistung“ der Datenquelle „BMD“

aufgelistet und geschieht, indem die verschiedenen Leistungen auf Basis ihrer Leistungsartennummer in die unterschiedlichen Gesetze unterteilt werden.

Die Transformation für die Dimension „Kunde“ gestaltet sich in beiden Datenquellen aufwändiger. Zunächst ist es notwendig, einen Datenstamm einzupflegen, der es ermöglicht, auf Basis der vorhandenen Adressdaten wie Postleitzahl und Ort auf das Bundesland sowie den Bezirk zu schließen. Die Statistik Austria veröffentlicht regelmäßig aktuelle Tabellen über die in Österreich gültigen politischen Bezirke (STATISTIK AUSTRIA, 2017) und die Verteilung von Postleitzahlen auf Gemeinden (STATISTIK AUSTRIA, 2017). Mithilfe dieser Daten ist es möglich, diese Anreicherung durchzuführen, allerdings sind die Tabellen zunächst zusammenzuführen. Dies ist möglich, da in der Ortsliste jede Gemeinde über eine sechsstellige Kennziffer verfügt, wobei die ersten drei Stellen gleichzeitig die Kennziffer des politischen Bezirks darstellen. Das Zusammenführen ist ebenfalls eine View in der Staging-Datenbank, die erstellt werden kann, nachdem die beiden Tabellen mithilfe von SSIS (analog wie in Kapitel 7.3.1 „Staging Area beladen“ beschrieben) in die Staging-Area geladen werden – lediglich die Datenquelle ist nicht „OLEDB“, sondern „Excel“.

Bei der Datenquelle „AABDB“ ist sind folgende Transformationen für die Dimension „Kunde“ notwendig:

- Bei der Sozialversicherungsnummer werden lediglich die ersten 4 Stellen abgespeichert, getrennt vom Geburtsdatum. Auch wenn dies in den meisten Fällen ausreichend ist, lautet die korrekte Schreibweise einer Sozialversicherungsnummer „NNNN MMDDJJJJ“. Mithilfe des Geburtsdatums wird diese Schreibweise erzeugt.
- Bezirk und Bundesland werden mithilfe der zuvor erwähnten Daten angereichert.

Die Dimension „Kunde“ gestaltet sich bei der Datenquelle „BMD“ bereits etwas komplexer. Zunächst wird das Geschlecht in BMD mit dem Wert „0“ für „männlich“ und 1 für „weiblich“ gespeichert. Dies wird auf den entsprechenden ausgeschriebenen Text geändert.

Weiters fehlen in BMD die Informationen zum Wohnort des Kunden, da diese nicht verrechnungsrelevant sind, weil die Rechnung an die Bezirksverwaltungsbehörde und nicht an den Kunden selbst ausgestellt wird. Deshalb sind Informationen aus dem „KundInnenakt stationär“ hinzuzuziehen, welche analog zur Vorgangsweise in Kapitel 7.3.1 „Staging Area beladen“ in die Staging Area geladen werden – hier ist als Datenquelle „OData“ mit der REST-URL von SharePoint anstatt „OLEDB“ zu wählen, ansonsten ist die Vorgangsweise dieselbe. Diese Transformation wird in Anhang 4: Script zur Anlage der Transformation „Kunde“ der Datenquelle „BMD“ aufgelistet.

Die Transformationen für die Dimension „Kundennummer“ sind bei beiden Datenquellen schwach ausgeprägt, Listing 7 zeigt als Beispiel die Transformation für die Datenquelle „BMD“.

```
CREATE VIEW BMDAW.[DimKundennummer] AS
SELECT
  PER_PERSONENNR NaturalKey,
  NULL Kunde_FK, --lookup DimKunde
  NULL IdentifierKey, -- lookup DimKunde
  'BMDAW-Kunden' Datenquelle
FROM PER_PERSONEN
```

Listing 7 SQL-View für die Transformation zur Dimension "Kundennummer" aus der Datenquelle "BMD".

Dies liegt daran, dass die meiste Logik hinsichtlich der Kundennummern bei der Dubletteneliminierung im Zuge der Beladung des DW abgebildet wird.

Anschließend werden die unterschiedlichen Views der Dimensionen noch mithilfe von UNION in entsprechende Gesamt-Views überführt. Dieses Vorgehen ermöglicht ein einfaches Abfangen von NULL-Werten, die im DW als „unbekannt“ dargestellt werden sollen. Außerdem wird die Schreibweise von Textfeldern vereinheitlicht, indem sichergestellt wird, dass der erste Buchstabe jedes Wortes in Groß- und der Rest in Kleinbuchstaben geschrieben wird.

7.3.3 Data Warehouse beladen

Bei der Beladung des Data Warehouse müssen nun SCD und Dubletten berücksichtigt werden. Für beide Problemstellungen gibt es in der grafischen Oberfläche von SSDD Werkzeuge, die jedoch in ihrer Konfiguration relativ aufwändig sind, wobei dennoch manche Features nicht realisierbar sind. Daher werden die Beladungen als Stored Procedures umgesetzt, die hauptsächlich aus einer einzelnen MERGE-Operation bestehen (vgl. Kapitel 2.4.3 „Load“).

7.3.3.1. Zeitdimension

Die Dimension „Zeit“ ist in einem DW konstant und verändert sich nicht, daher muss diese nur einmalig nach den in Kapitel 6.1 „Dimension Zeit“ beschriebenen Regeln für einen großen Zeitraum befüllt werden. Für die Beladung der Zeitdimension gibt es in SSDD die Möglichkeit, diese mithilfe eines Assistenten auf der Benutzeroberfläche automatisch zu erstellen. Diese Vorgehensweise ist jedoch in den Konfigurationsmöglichkeiten – beispielsweise hinsichtlich Feiertagen – eingeschränkt. Daher wird auf ein Script zurückgegriffen, das auf der Webseite mssqltips.com veröffentlicht wurde, und dieses an die Gegebenheiten von JAW angepasst. Das angepasste Script ist in Anhang 2: Script zur Anlage der Dimension Zeit aufgelistet.

7.3.3.2. Dimensionen beladen

Bei der Beladung der Dimensionen auf Basis der zuvor erstellten Transformationen ist insbesondere zu beachten, dass Dubletten ausgemerzt werden. Das dimensionale Datenmodell sieht vor, dass für alle Kunden die Kundennummer im Quellsystem nicht direkt in der Dimension

„Kunde“ sondern in der Dimension „Kundennummer“ gespeichert wird. Um diese Komplexität sicher realisieren zu können, wird auf ein MERGE-Statement gesetzt.

Wie in Kapitel 7.3.2 „Staging transformieren mit SSIS“ erwähnt, besitzen die Dimensionen „Kostenstelle“, „Dienstleistung“ und „Standort“ eine geringe Komplexität und können mit dem Beispiel von Thorntwaite (Thornthwaite, 2008) sehr einfach realisiert werden. Die Dimensionen „Kunde“ und „Kundennummer“ sind hingegen weitaus komplexer, da sie aufgrund der Diversität der Datenquellen keinen eindeutigen natürlichen Schlüssel besitzen. Aus diesem Grund muss die Beladung neben der Dublettenprüfung gleichzeitig die Verwaltung eines künstlichen „IdentifierKey“ erledigen. Anhang 5: Script zur Beladung der Dimensionen „Kunde“ und „Kundennummer“ realisiert diese Aufgabenstellung. Die Dublettenprüfung geschieht mithilfe des Scriptblocks `ROW_NUMBER() OVER (PARTITION BY Nachname, Vorname ORDER BY Nachname, Vorname) AS RN` sowie einem abschließendem `WHERE RN = 1`. Dies bedeutet, dass lediglich die Gleichheit von Vor- und Nachname geprüft wird, um die Eindeutigkeit zu gewährleisten. Diese Vorgangsweise wurde vom Autor gewählt, da für eine genauere Prüfung eine Vielzahl von Datensätzen in den Datenquellen nachgepflegt werden müssten, was im Zuge dieser Masterarbeit nicht möglich gewesen wäre, sondern nur auf unternehmensweite Anweisung durch die Geschäftsführung. Aufgrund des gewünschten Prototyp-Ergebnisses wurde dieser Aufwand nicht umgesetzt.

Der Rest dieses SQL-Scripts ist bis auf folgende zwei Unterschiede gleich wie die MERGE-Statements der anderen Dimensionen: Der „künstliche Schlüssel“, also der „IdentifierKey“ wird realisiert, indem der Key der Zeile in alle leeren Identifier-Key-Felder geschrieben wird. Bei der nächsten Beladung wird der IdentifierKey vom MERGE-Statement fortgeschrieben und identifiziert daher einen Kunden eindeutig. Dies wird mithilfe eines einfachen UPDATE-Statements implementiert.

Die Beladung der Dimension Kundennummer ist der abschließende Schritt der Beladung. Hierfür werden die bereits in die Dimension „Kunde“ geladenen Datensätze mit den Daten aus der Staging-Area auf Basis von Vornamen und Nachnamen zusammengefügt und die entsprechenden Keys der Kundentabelle und NaturalKeys aus den Datenquellen in die Dimension eingefügt.

Abschließend werden die erstellten Stored Procedures mithilfe eines „Execute SQL Task“ in einem SSIS-Package automatisiert ausgeführt, um das DW zu beladen.

7.3.3.3. Fakten erzeugen

Auf Basis der Rohdaten in der Staging Area sowie den nun vorhandenen Dimensionsdaten kann auf die einzelnen Transaktionen, die zu den Stammdaten geführt haben, zurückgeschlossen werden.

Dazu werden für jede Transaktionsvariante jeder Datenquelle entsprechende Datenzeilen erzeugt. In der Datenquelle „AABDB“ gibt es lediglich Informationen über einen Einstieg und einen Ausstieg, was für das DW in „Eintritt“ und „Austritt“ übersetzt wird. Es werden zunächst die Eintritte beleuchtet, was bedeutet, dass für jeden Kunden in der Rohdatentabelle „TeilnehmerInnen“ eine Zeile erzeugt wird, welche mit den Spalten der Faktentabelle „Kundenstamm“ korreliert. Als Datum der Transaktion gilt hierbei das Eintrittsdatum. Anschließend werden dieselben Daten für die Ausstiege erzeugt; hierbei gibt es lediglich zwei Unterschiede: Es wird geprüft, ob der Kunde überhaupt ein eingetragenes Ausstiegsdatum hat und wenn nicht, wird dieser ausgelassen. Der zweite Unterschied besteht darin, dass als Datum der Transaktion das Austrittsdatum gilt.

Für die Datenquelle „BMD“ gilt grundsätzlich dieselbe Vorgangsweise, allerdings gibt es hier auch den Typ „Änderung“, da BMD in den „historischen Daten“ auch die alten Stammdaten mit entsprechendem Gültigkeitsdatum mitführt. Als Basistabelle dient hier „PER_PERSONEN“.

Die Hauptaufgabe bei diesen Transformationen stellt das Nachschlagen der Surrogate Keys im DW dar, um die Faktentabelle mit den entsprechenden Fremdschlüsseln befüllen zu können.

Das gesamte Skript ist in Anhang 6: Script zur Beladung der Faktentabelle „Kundenstamm“ angeführt.

Die Faktentabelle „Kundenstamm-Snapshot“ besitzt eine höhere Komplexität. Aus den Transaktionen in der Faktentabelle „Kundenstamm“ heraus werden die Zeiträume ermittelt, zu denen jeder einzelne Kunde mit einem bestimmten Satz an Informationen – also bis zu einer Transaktion „Änderung“ – aktiv war. Dazu wird die Tabelle FaktKundenstamm mit sich selbst vereinigt. Zur Veranschaulichung hat die „linke“ FaktKundenstamm den Alias „ein“ und die „rechte“ FaktKundenstamm den Alias „aus“. Die Bedingung für die Vereinigung sind:

- Es muss sich um dieselbe Kundennummer handeln (ein.Kundennummer_FK ist gleich aus.Kundennummer_FK)
- Es muss sich um unterschiedliche Datensätze handeln (ein.Key ungleich aus.Key)
- Der „ein“-Datensatz muss ein Datum vor dem ersten „aus“-Datensatz haben (ein.Datum_FK ist kleiner oder gleich aus.Datum_FK)
- Der „ein“-Datensatz“ beschreibt entweder eine „Eintritt“- oder eine „Änderung“-Transaktion
- Der „aus“-Datensatz beschreibt entweder eine „Austritt“- oder eine „Änderung“-Transaktion
- Gruppiert nach ein.Kundennummer_FK und ein.Datum, sortiert nach ein.Datum_FK und aus.Datum_FK ist nur das erste Ergebnis relevant

Damit stehen die Gültigkeitszeiträume der einzelnen Daten fest. Zur Beladung kann nun die Dimension Zeit mit diesen Daten vereinigt werden, als Bedingung gilt der Zusammenhang des Zeitraums zwischen „ein“ und „aus“ mit dem Datum der Dimensionstabelle. Aufgrund starker Datenfehler (Daten ab 1.1.1900) und vielen unendlich gültigen Einträgen wird die Dimension Datum bei dieser Beladung auf den Zeitraum vom Start der Tagsatzabrechnung mit BMD NTCS bei JAW im Jahr 2013 bis fünf Jahre in die Zukunft des aktuellen Datums eingeschränkt.

Das gesamte Skript ist in Anhang 7: Script zur Beladung der Faktentabelle „Kundenstamm-Snapshot“ angeführt.

7.3.4 Würfel bereitstellen mit SSAS

Da nun die Daten in dimensionaler Form im DW vorhanden sind, werden diese mithilfe von SSAS in einfach verfügbarer Form zur Verfügung gestellt. Nach Erstellung des entsprechenden Projekts in SSDT wird das DW als Datenquelle hinzugefügt. Die Datenbanktabellen werden automatisch als Dimensionen erkannt und die Relationen zwischen den Dimensionen und den Faktentabellen hergestellt. Es muss nun nur noch die Zeit-Dimension als solches festgelegt werden, dann stehen bereits grundsätzliche Auswertungsmöglichkeiten zur Verfügung.

Zusätzlich werden an dieser Stelle die Hierarchien definiert, die eine logisch geordnete Filterauswahl in Client-Tools wie Excel ermöglichen. Abbildung 15 zeigt die Hierarchien „Kalender“ und „Wochen“ für die Dimension Zeit und Abbildung 16 die regionale Hierarchie der Dimension Kunde.

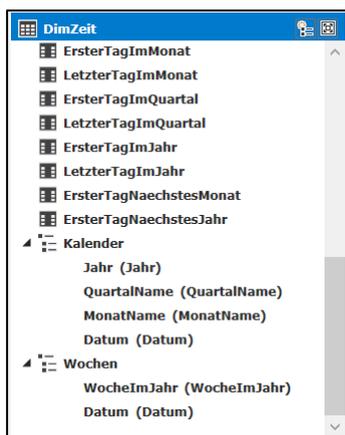


Abbildung 15 Hierarchien in der Dimension Zeit. Quelle: eigene Darstellung (Microsoft SQL Server Data Tools)

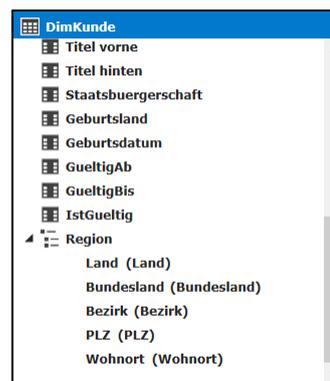


Abbildung 16 Regionale Hierarchie in der Dimension Kunde. Quelle: eigene Darstellung (Microsoft SQL Server Data Tools)

Damit auch eine Analyse auf Basis von Altersgruppen möglich ist, wird im DW eine fixe Tabelle, wie in Tabelle 10 angegeben, erstellt.

Key	Altersgruppe	Von	Bis
1	0-8	0	8

2	9-12	9	12
3	12-18	12	18
4	18-30	18	30
5	30-50	30	50
6	50+	50	9999

Tabelle 10 Tabelle Altersgruppe. Quelle: eigene Darstellung (Microsoft Excel)

Diese Tabelle wird dem Datenmodell des SSAS-Projekts ebenfalls hinzugefügt. In den Tabellen „FaktKundenstamm“ sowie „FaktKundenstamm-Snapshot“ kann somit eine neue Spalte „Altersgruppe“ hinzugefügt werden, indem eine Formel wie in Listing 8 angegeben konfiguriert wird.

```
Altersgruppe:=
CALCULATE (
  VALUES (Altersgruppen[Altersgruppe]);
  FILTER (
    DimAltersgruppen;
    FaktKundenstamm[Alter] >= Altersgruppen[von]
    && FaktKundenstamm[Alter] < Altersgruppen[bis]
  )
)
```

Listing 8 Formel für Spalte „Altersgruppe“

Die Formel sagt aus, dass die Werte der Spalte „Altersgruppe“ aus der Tabelle „Altersgruppen“ ausgegeben werden soll, wenn das Alter größer oder gleich dem Wert in der Spalte „Von“ und kleiner dem Wert in der Spalte „Bis“ ist. Dadurch, dass bei „Von“ der Operator „größer gleich“ und bei „Bis“ der Operator „kleiner“ verwendet wird, sind die Daten in Tabelle 10 nachvollziehbar. Bei der Analyse in Client-Tools kann somit auch nach Altersgruppe gruppiert und ausgewertet werden.

Dies führt zur weiteren Analyse in Microsoft Excel, welche möglich ist, sobald das SSAS-Projekt bereitgestellt wurde.

7.3.5 Daten analysieren

In Microsoft Excel ist es möglich, sich direkt mit SSAS zu verbinden. Dazu klickt man in der Registerkarte „Daten“ auf den Menüpunkt „Externe Daten abrufen“ und wählt bei „Aus anderen Quellen“ die Option „Aus Analysis Services“ aus. Nach der Eingabe des Servernamens ist noch die SSAS-Datenbank sowie der entsprechende Cube auszuwählen. Fährt man nun mit einem „PivotTable-Bericht“ fort, hat man im Auswahlfenster „PivotTable-Felder“ Zugriff auf alle Faktentabellen, Dimensionen und deren Attribute sowie auf definierte Measures.

Um nun beispielsweise den zeitlichen Verlauf der durchschnittlichen Anzahl an Kunden je Datenquelle zu analysieren, ist die PivotTable mit dem Kalender aus der Dimension Zeit auf der horizontalen x-Achse (Zeilen) und die Datenquellen aus der Dimension Kundennummer auf der vertikalen y-Achse (Spalten) zu konfigurieren, vgl. Abbildung 17.

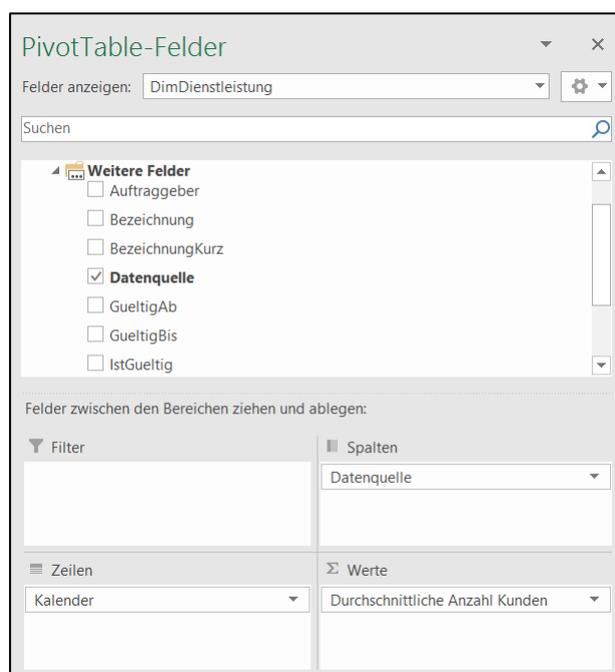


Abbildung 17 Konfiguration PivotTable zur Auswertung der durchschnittlichen Anzahl an KundInnen im zeitlichen Verlauf. Quelle: eigene Darstellung (Microsoft Excel)

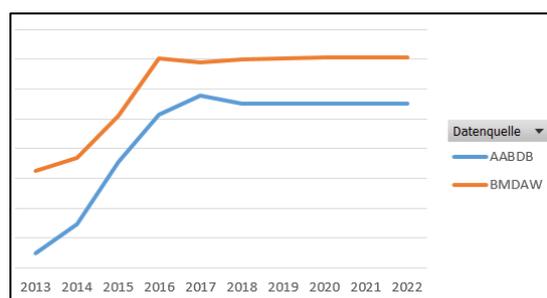


Abbildung 18 KundInnen im zeitlichen Verlauf. Quelle: eigene Darstellung (Microsoft Excel)

Fügt man nun noch ein PivotChart Liniendiagramm hinzu, werden die Zahlen, wie in Abbildung 18 demonstriert, grafisch dargestellt. Durch die Hierarchie der Zeit ist zunächst nur die Detailstufe „Jahr“ sichtbar; in der PivotTable lässt sich das Jahr jedoch ausklappen, wodurch Quartale nicht nur in Form von Zahlen in der Tabelle, sondern auch in Form von Datenpunkten in der Grafik ersichtlich werden. Es wird also ein Drill-Down mithilfe der Dimensionshierarchie „Kalender“ durchgeführt (vgl. Kapitel 4.1.4 „Drill-Down / Roll-Up“).

Die H0 von Hypothese 2 besagt, dass Ereignisse, die außerhalb des Einflussbereichs des Unternehmens liegen, keine Auswirkungen auf die Analyse der langfristigen Daten haben. Nach der Sperre der Balkanroute im März 2016 (derStandard.at, 2016) war eine stark verzögerte Auswirkung auf die Anzahl an neu eingetretenen Flüchtlingen in der Grundversorgung von JAW erkennbar, wie Abbildung 19 belegt. Die Verzögerung resultiert daraus, dass die Grundversorgung erst zum Tragen kommt, nachdem eine Erstaufnahme in Österreich passiert ist und die Asylwerber vom Verteilerzentrum ausgehend zu JAW zugewiesen werden (asylkoordination österreich) Diese Erkenntnis wird in Kapitel 8 „Resümee“ bei der Hypothesenüberprüfung weiterverwendet. Zur Erstellung von Abbildung 19 wurde in Microsoft Excel ein PivotChart erstellt, bei dem die Dimension Dienstleistung auf den Typ „Flüchtlingsbetreuung“ gefiltert wurde, das Attribut „Transaktion“ der Faktentabelle Kundenstamm auf „Eintritt“ gefiltert wurde und der Kalender auf die horizontale x-Achse gelegt und auf das Jahr

2016 gefiltert wurde. Anschließend wurde für die Werte das Measure „SummeTransaktionen“ ausgewählt.

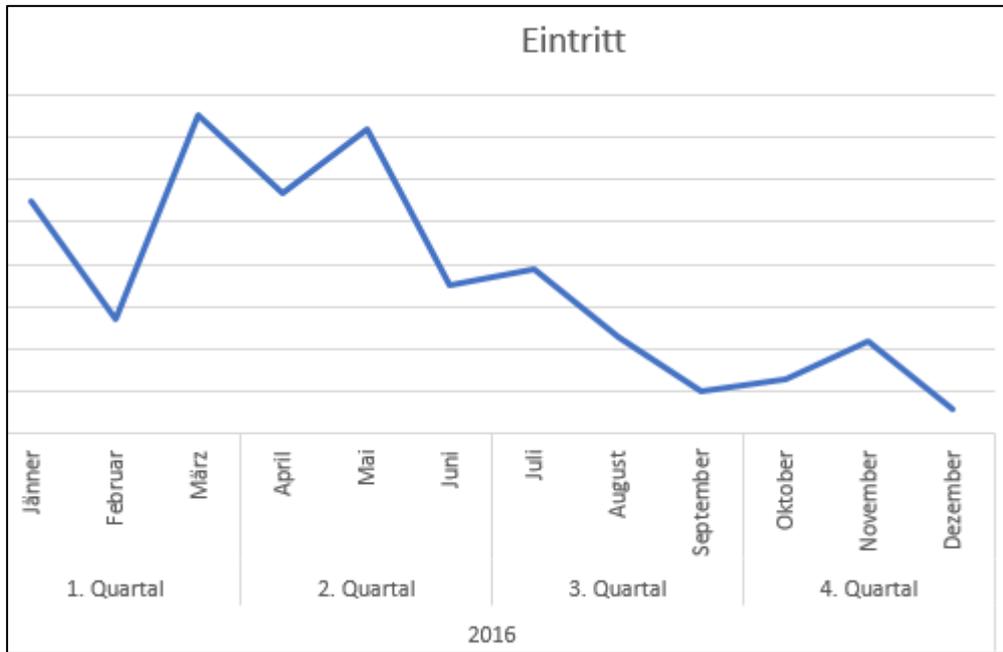


Abbildung 19 Verlauf der Anzahl an Eintritten in den Dienstleistungs-Typ "Flüchtlingsbetreuung" im Jahr 2016. Quelle: eigene Darstellung (Microsoft Excel)

8 RESÜMEE

Ziel dieser Masterarbeit war es, die Implementierung eines Data Warehouse (DW) am Beispiel der Jugend am Werk Steiermark GmbH zu planen und in weiterer Folge prototypisch umzusetzen. Dieses DW soll die Stammdaten der Kunden beinhalten und analysierbar bereitstellen, sowie eine zentrale Anlaufstelle für Fragen rund um das Thema Kundenstamm bieten. Dies konnte umgesetzt werden, indem für zwei Datenquellen der Beladungsvorgang implementiert und ein OLAP-Cube bereitgestellt wurde.

Die zentrale Fragestellung der Arbeit war das Finden einer geeigneten Herangehensweise für die Implementierung eines DW, mit der ein Mehrwert für das Unternehmen generiert werden kann. Um dies zu ermöglichen, wurden zunächst sämtliche Dienstleistungsangebote gemeinsam mit den entsprechenden Verantwortlichen analysiert.

Durch die Aufarbeitung der einzelnen Dienstleistungen und deren Datenhaltung wurde ersichtlich, dass ein Großteil des Arbeitsaufwandes im Rahmen der Implementierung eines DW aus dem Sammeln von Informationen, Abläufen und Datenarten im Unternehmen besteht. Die zweitgrößte Herausforderung stellt das Analysieren der Quelldaten, das Finden von Fehlern sowie einer geeigneten Strategie, um mit diesen Fehlern umzugehen, dar. Die tatsächliche technische Umsetzung macht einen verhältnismäßig kleinen Teil des Implementierungsprozesses aus. Die Berücksichtigung dieses Umstandes ermöglicht es, ein DW zu kreieren, welches einen tatsächlichen Mehrwert für das Unternehmen darstellt.

Die zu verwerfende H0 der Hypothese 1 besagt, dass kaum strukturierte Daten zu den Dienstleistungen im Sozialbereich existieren. In Kapitel 5.1.1 „Gliederung nach Applikation“ wurde eine Analyse der verwendeten Datenapplikationen durchgeführt und es konnte festgestellt werden, dass für lediglich 22% der Dienstleistungen keine strukturierte Datenablage vorhanden ist. Aufgrund dieses Ergebnisses lässt sich die H0 verwerfen, H1 ist demnach verifiziert.

Bei Hypothese 2 besagt die H0, dass der politische Einfluss auf die Sozialwirtschaft so groß ist, dass sich keine Muster aus den Daten herauslesen lassen. Aus den in Kapitel 7.3.5 „Daten analysieren“ ausgewerteten Daten lässt sich jedoch ableiten, dass äußere Einflüsse sehr wohl die Daten verändern – die Sperre der Balkanroute im Zuge der Flüchtlingskrise hat im Jahr 2016 zu einem starken Rückgang an Eintritten bei der Asylwerberbetreuung geführt. Auf Basis dieser Erkenntnis lässt sich die H0 verwerfen und die H1 dieser Hypothese gilt als verifiziert.

Zur Verifizierung der Hypothese 3 muss die H0 widerlegt werden, welche besagt, die Geschäftsführung würde mangels Vertrauen in die Datenqualität ein DW ignorieren. Der Prototyp aus Kapitel 7.3 „Prototyp“ fand einen guten Anklang bei der Geschäftsführung. Die Befragung ergab, dass eine weitere Entwicklung des DW durchaus im Interesse des Unternehmens liegt,

das Vertrauen in die Korrektheit der Daten ist ebenfalls gegeben. Allerdings wurde während der Erstellung dieser Masterarbeit ein Projekt für eine zentrale Stammdatenverwaltung (MDM – Master Data Management) gestartet, welches genau dieselben Daten, wie sie in dieser Masterarbeit behandelt wurden, in einer eigenen Applikation verwaltet. Im Unterschied zu einem DW werden die Daten nicht nur aus Datenquellen in das MDM übernommen, sondern dort auch weiterverarbeitet und anschließend wieder auf die Datenquellen verteilt. Ein solches MDM ist auch für ein langfristig funktionierendes DW notwendig, weshalb die Umsetzung eines DW für die operationalen Daten noch nicht beauftragt wird. Damit kann die H0 der Hypothese 3 verworfen werden und die Hypothese gilt als verifiziert.

Für die tatsächliche Durchführung eines DW-Projektes sind aufbauend auf dem in dieser Arbeit generierten Wissen weitere und tiefere Datensammel- und Analysearbeiten zu erledigen. Anschließend müssen die Datenquellen sowie die Fehlerbehandlungsstrategie weitaus genauer definiert werden, um einen langfristig stabilen und vertrauenswürdigen Datenstamm zu erhalten. Mit einer größeren Vielfalt an Daten durch mehr unterschiedliche Datenquellen steigt auch der Nutzen dieser Technologie.

Nach Abschluss dieser Masterarbeit wird JAW, wie zuvor erwähnt, ein MDM implementieren, das zunächst die mitarbeiterbezogenen Stammdaten umfassen wird. In weiterer Folge, auch in Hinblick auf die Anforderungen der DSGVO, gut über alle personenbezogenen Daten im Unternehmen Bescheid wissen zu müssen, wird das MDM dann sehr rasch auf die Kundendaten erweitert werden. Durch die vollständige allumfassende Speicherung der Unternehmens-Stammdaten im MDM werden viele alltägliche Abläufe erleichtert und beschleunigt. Aber auch das zukünftige DW wird davon profitieren: Sind Stammdaten mit hoher Datenqualität vorhanden, die über mehrere Datenquellen gehen, kann die Implementierungszeit eines vollständigen EDW stark verkürzt werden. Nach der vollständigen Umsetzung des MDM-Projekts wird JAW sich erneut der Frage widmen, ob, in welchem Umfang und wann ein DW im Unternehmen eingeführt wird. Aus aktueller Sicht scheint der Nutzen eines DW bei JAW sehr hoch und eine Implementierung wird voraussichtlich beauftragt werden. Das Projekt wird, sollten keine unvorhergesehenen Umstände auftreten, im Sommer 2018 gestartet werden.

ABKÜRZUNGSVERZEICHNIS

3NF	dritte Normalform
AAB	Ausbildung Arbeit Beschäftigung; ein Teil des DLM von JAW
AMS	Arbeitsmarktservice
ARGE	Arbeitsgemeinschaft; GesbR
bfi	Berufsförderungsinstitut Steiermark
BI	Business Intelligence
BFW	Betreuung Freizeit Wohnen; ein Teil des DLM von JAW
BWL	Betriebswirtschaftslehre
CRM	Customer Relationship Management
CSV	Comma Separated Values
DB	Datenbank
DLM	Dienstleistungsmanagement
DM	Data Mining
DSG2000	Datenschutzgesetz 2000
DSGVO	Datenschutz-Grundverordnung
DW	Data Warehouse
DWS	Data Warehouse System
EDW	Enterprise Data Warehouse
EIS	Executive Information Systems
ERP	Enterprise Resource Planning
ER	Entity Relationship
ERD	Entity Relationship Diagram
EStG	Einkommensteuergesetz
ETL	Extract-Transform-Load
EU	Europäische Union
FTP	File Transfer Protocol
GBP	Gemeinnütziges Beschäftigungsprojekt
GesbR	Gesellschaft bürgerlichen Rechts
HTTP	Hypertext Transfer Protocol
IFRS	International Financial Reporting Standards
IKT	Informations- und Kommunikationstechnologie (IT-Abteilung von JAW)
IT	Informationstechnologie
JAW	Jugend am Werk Steiermark
KPI	Key Performance Indicator
LEVO-StBHG	StBHG Leistungs- und Entgeltverordnung
MDM	Master Data Management
MIS	Management Information Systems
MSS	Management Support Systeme

NEBA	Netzwerk Berufliche Assistenz
OLAP	On-Line Analytical Processing
OLTP	On-Line Transactional Processing
OR	Operations Research
SCD	Slowly Changing Dimension
SOMS	Sozialministeriumservice
SQL.....	Structured Query Language
SSAS	SQL Server Analysis Services
SSDT	SQL Server Data Tools
SSIS.....	SQL Server Integration Services
StBHG.....	Steiermärkisches Behindertengesetz
StGKK.....	Steiermärkische Gebietskrankenkasse
StKJHG.....	Steiermärkisches Kinder- und Jugendhilfegesetz
StKJHG-DVO.....	StKJHG-Durchführungsverordnung
T-SQL	Transact-SQL
UGB	Unternehmensgesetzbuch
UMF	Unbegleitete minderjährige Flüchtlinge
UTC	Universal Time Coordinated
WIPS.....	Webbasierendes Informations- und Präsentationssystem des Sozialbereichs des Landes Steiermark
WKO	Wirtschaftskammer Österreich

ABBILDUNGSVERZEICHNIS

Abbildung 1 Organigramm Jugend am Werk Steiermark GmbH. Quelle: Qualitätsmanagementsystem im Intranet von JAW.	11
Abbildung 2 Dreidimensionaler Würfel mit Zeit, Produkten und Variablen. Quelle: (Thomsen, 2002, S. 49)	27
Abbildung 3 Aufbau eines SSIS-Projekts in SSDT. Quelle: eigene Darstellung (Microsoft SQL Server Data Tools).....	32
Abbildung 4 Slicing. Quelle: (Rainardi, 2008, S. 413)	39
Abbildung 5 Dicing. Quelle: (Rainardi, 2008, S. 414).....	40
Abbildung 6 Roll-Up und Drill-Down. Quelle: (Ballard, Data modeling techniques for data warehousing, 1998, S. 45)	41
Abbildung 7 Zusammenhang der drei Kategorien von Datenanalysen. Quelle: (Myatt, 2017, S. 3)	43
Abbildung 8 Verwendung Applikationen durch Dienstleistungsarten (Anzahl). Quelle: eigene Darstellung (Microsoft Excel).....	46
Abbildung 9 Dienstleistungsarten mit unstrukturierten Daten am Fileserver. Quelle: eigene Darstellung (Microsoft Excel).....	47
Abbildung 10 Datenstruktur AABDB. Quelle: eigene Darstellung (Microsoft Visual Studio).....	53
Abbildung 11 Datenstruktur agilion. Quelle: eigene Darstellung (Microsoft Visual Studio)	55
Abbildung 12 Datenmodell der Kundenstammdaten in BMD NTCS. Quelle: eigene Darstellung (Microsoft Visual Studio)	59
Abbildung 13 Datenbankmodell "GBPDB". Quelle: eigene Darstellung (Microsoft Visual Studio)	61
Abbildung 14 SSIS-Paket "1_StageBMDAW". Quelle: eigene Darstellung (Microsoft SQL Server Data Tools).....	80
Abbildung 15 Hierarchien in der Dimension Zeit. Quelle: eigene Darstellung (Microsoft SQL Server Data Tools).....	86
Abbildung 16 Regionale Hierarchie in der Dimension Kunde. Quelle: eigene Darstellung (Microsoft SQL Server Data Tools)	86
Abbildung 17 Konfiguration PivotTable zur Auswertung der durchschnittlichen Anzahl an KundInnen im zeitlichen Verlauf. Quelle: eigene Darstellung (Microsoft Excel).....	88
Abbildung 18 KundInnen im zeitlichen Verlauf. Quelle: eigene Darstellung (Microsoft Excel)	88
Abbildung 19 Verlauf der Anzahl an Eintritten in den Dienstleistungs-Typ "Flüchtlingsbetreuung" im Jahr 2016. Quelle: eigene Darstellung (Microsoft Excel)	89

TABELLENVERZEICHNIS

Tabelle 1 Tabellarische Darstellungen der Dienstleistungen und deren Datenapplikationen. Quelle: eigene Darstellung (Microsoft Excel).....	48
Tabelle 2 Übersicht "umbenannte" Felder BMD NTCS. Quelle: eigene Darstellung (Microsoft Excel)	57
Tabelle 3 Datumsdimension mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel).....	63
Tabelle 4 Aufbau der Dimension "Kunden" mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel)	64
Tabelle 5 Dimension „Kundennummer“ mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel)	65
Tabelle 6 Dimension "Dienstleistung" mit Beispieldaten. Quelle: eigene Darstellung (Microsoft Excel) ...	66
Tabelle 7 Dimension "Kostenstelle". Quelle: eigene Darstellung (Microsoft Excel).....	67
Tabelle 8 Dimension "Standort". Quelle: eigene Darstellung (Microsoft Excel).....	68
Tabelle 9 Faktentabelle "Kundenstamm". Quelle: eigene Darstellung (Microsoft Excel).....	70
Tabelle 10 Tabelle Altersgruppe. Quelle: eigene Darstellung (Microsoft Excel).....	87

LISTINGS

Listing 1 SQL-Abfrage zur Identifizierung überlappender Zeiträume in der Tabelle DVH_DIENSTVERHAELTNIS.....	75
Listing 2 SQL-Abfrage zur Identifizierung überlappender Zeiträume in der Tabelle PSW_PERSONENWERTE	76
Listing 3 SQL-Abfrage zur Identifizierung von Einträgen in DVH_DIENSTVERHAELTNIS, die außerhalb des Zeitraums von PER_PERSON liegen.....	76
Listing 4 SQL-Abfrage zur Identifizierung von Einträgen in PSW_PERSONENWERTE, die außerhalb des Zeitraums von DVH_DIENSTVERHAELTNIS liegen	77
Listing 5 SQL-Abfrage zur Identifizierung der Menge unvollständiger Datensätze.....	78
Listing 6 SQL-Abfrage zur Erstellung der Transformations-View für die Dimension "Dienstleistung" der Datenquelle "AABDB".....	81
Listing 7 SQL-View für die Transformation zur Dimension "Kundennummer" aus der Datenquelle "BMD".	83
Listing 8 Formel für Spalte „Altersgruppe“	87

LITERATURVERZEICHNIS

- Arbeitsmarktservice Steiermark. (April 2016). Geschäftsbericht 2015. (D. Riedl, & M. Dieter, Redakteure) Graz, Steiermark, Österreich. Abgerufen am 2. Jänner 2017 von http://www.ams.at/_docs/600_gb_2015_web.pdf
- Aspin, A. (2013). *SQL Server 2012 Data Integration Recipes*. New York: Apress.
- asylkoordination österreich. (kein Datum). Das Asylverfahren seit 2015. Abgerufen am 16. Juni 2017 von <http://www.asyl.at/adincludes/dld.php?datei=12.03.ma,asylverfahren2016.pdf>
- Atteslander, P. (kein Datum). *Methoden der empirischen Sozialforschung*.
- Ballard, C. (1998). *Data modeling techniques for data warehousing*. United States: IBM.
- Ballard, C., Farrell, D. M., Gupta, A., Mazuela, C., & Vohnik, S. (2006). *Dimensional Modeling* (1. Ausg.). United States: IBM.
- Ben-Gan, I., Sarka, D., Wolter, R., Low, G., Katibah, E., Kunen, I., & Delaney, K. (2010). *Inside Microsoft SQL Server 2008: T-SQL Programming*. Redmond, Washington: Microsoft Press.
- Bertrand, A. (20. Oktober 2015). Creating a date dimension or calendar table in SQL Server. (mssqltips.com, Hrsg.) Abgerufen am 9. Juni 2017 von <https://www.mssqltips.com/sqlservertip/4054/creating-a-date-dimension-or-calendar-table-in-sql-server/>
- Bundeskanzleramt. (28. Jänner 1981). Datenschutzübereinkommen. Abgerufen am 1. Juni 2017 von <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=1000965>
- Bundeskanzleramt. (1. Jänner 2000). Datenschutzgesetz 2000. Abgerufen am 1. Juni 2017 von <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=bundesnormen&Gesetzesnummer=10001597>
- Chen, Z. (2001). *Intelligent Data Warehousing* (1. Ausg.). Boca Raton, Fla: CRC Press.
- Cleff, T. (2008). *Deskriptive Statistik und moderne Datenanalyse*. Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Th. Gabler | GWV Fachverlage GmbH.
- de Ruiter, A. (5. Dezember 2012). Designing an ETL process with SSIS: two approaches to extracting and transforming data. Microsoft Corp. Abgerufen am 23. Mai 2017 von <https://blogs.msdn.microsoft.com/andreasderuiter/2012/12/05/designing-an-etl-process-with-ssis-two-approaches-to-extracting-and-transforming-data/>

- derStandard.at. (9. März 2016). Länder auf der Balkanroute riegeln Grenzen komplett ab. Abgerufen am 15. Juni 2017 von <http://derstandard.at/2000032554989/Balkanroute-faktisch-dicht-Slowenien-laesst-keine-Fluechtlinge-mehr-durch>
- Duncan, O., & Guyer, C. (17. März 2017). Comparing Tabular and Multidimensional Solutions (SSAS). Microsoft Corporation. Abgerufen am 15. Juni 2017 von <https://docs.microsoft.com/en-us/sql/analysis-services/comparing-tabular-and-multidimensional-solutions-ssas>
- Europäisches Parlament und Rat. (24. Oktober 1995). Richtlinie 95/46/EG. Abgerufen am 1. Juni 2017 von <http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:31995L0046>
- Europäisches Parlament und Rat. (27. April 2016). Datenschutz-Grundverordnung. Abgerufen am 1. Juni 2017 von <http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A32016R0679>
- Feinberg, D., & Beyer, M. (28. Jänner 2010). Magic Quadrant for Data Warehouse Database. Stamford, Connecticut, USA. Abgerufen am 30. Jänner 2017 von <ftp://129.35.224.15/software/kr/data/pdf/gardwdmq.pdf>
- Gluchowski, P., Gabriel, R., & Dittmar, C. (2008). *Management Support Systeme und Business Intelligence* (2. Ausg.). Berlin Heidelberg: Springer.
- Goeken, M. (2006). *Entwicklung von Data-Warehouse-Systemen* (1. Ausg.). Wiesbaden: Deutscher Universitäts-Verlag | GWV Fachverlage GmbH.
- Homepage "agilion". (kein Datum). Abgerufen am 29. Mai 2017 von <https://agilion.net>
- Homepage "BMD Systemhaus GesmbH". (kein Datum). Abgerufen am 29. Mai 2017 von <https://www.bmd.at/>
- Homepage "Microsoft SharePoint". (kein Datum). Abgerufen am 29. Mai 2017 von <https://products.office.com/de-at/sharepoint/>
- Imhoff, C., Galemme, N., & Geiger, J. (2003). *Mastering Data Warehouse Design - Relational and Dimensional Techniques*. Indianapolis, Indiana, USA: Wiley Publishing, Inc.
- Iseminger, D. (1. Juni 2017). On-premises data Gateway. Abgerufen am 16. Juni 2017 von <https://powerbi.microsoft.com/en-us/documentation/powerbi-gateway-onprem/>
- Jensen, C. S., Pedersen, T. B., & Thomsen, C. (2010). *Multidimensional Databases and Data Warehousing*. Morgan & Claypool.
- Kemper, H.-G., Baars, H., & Mehanna, W. (2010). *Business Intelligence - Grundlagen und praktische Anwendungen* (3. Ausg.). Wiesbaden: Vieweg+Teubner Verlag | Springer Fachmedien Wiesbaden GmbH.

- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit* (3. Ausg.). Indianapolis, IN: John Wiley & Sons, Inc.
- Knyrim, R., & Trieb, G. (2. Juni 2014). Das künftige EU-Datenschutzrecht - Neue Anforderungen an die unternehmerische Compliance. (L. V. KG, Hrsg.) Wien. Abgerufen am 1. Juni 2017 von http://www.preslmayr.at/tl_files/Publikationen/2014/Das%20kuenftige%20EU-Datenschutzrecht%20-%20Neue%20Anforderungen%20an%20die%20unternehmerische%20Compliance_Knyrim_Trieb.pdf
- Krishnan, K. (2013). *Data Warehousing in the Age of Big Data*. Waltham, Massachusetts, USA: Elsevier Inc.
- Kroenke, D. M., & Auer, D. J. (2014). *Database Processing - Fundamentals, Design and Implementation* (13. Ausg.). New Jersey: Pearson Education, Inc.
- Land Steiermark - Amt der Steiermärkischen Landesregierung. (24. November 2011). Ausschreibungsunterlagen für das Vergabeverfahren Sozialversicherungsrechtliche Absicherung und qualitätssichernde Hilfen/Maßnahmen zur Festigung des Pflegeverhältnisses. Abgerufen am 4. Jänner 2017 von [http://www.verwaltung.steiermark.at/cms/dokumente/11685470_74835893/70d08cda/11%2012%2014%20Vergabeverfahren%20sozialversicherungsrechtliche%20Absicherung%20\(2\).pdf](http://www.verwaltung.steiermark.at/cms/dokumente/11685470_74835893/70d08cda/11%2012%2014%20Vergabeverfahren%20sozialversicherungsrechtliche%20Absicherung%20(2).pdf)
- Land Steiermark - Amt der Steiermärkischen Landesregierung. (kein Datum). Flexible Hilfen - im Sinne des StKJHG - Case Management. Graz, Steiermark, Österreich. Abgerufen am 5. Jänner 2017 von <http://www.soziales.steiermark.at/cms/beitrag/12256878/10024850/>
- Land Steiermark - Amt der Steiermärkischen Landesregierung. (kein Datum). Grundversorgung von AsylwerberInnen. Graz, Steiermark, Österreich. Abgerufen am 3. Jänner 2017 von <http://www.soziales.steiermark.at/cms/ziel/112908349/DE>
- Land Steiermark - Amt der Steiermärkischen Landesregierung. (kein Datum). Integrative Berufsausbildung - Anlehre und Teilqualifizierung, Vorlehre. Graz, Steiermark, Österreich. Abgerufen am 2. Jänner 2017 von <http://www.soziales.steiermark.at/cms/beitrag/11980846/10024850/>
- Land Steiermark - Amt der Steiermärkischen Landesregierung. (kein Datum). Streetwork - Mobile Jugendarbeit. Graz, Steiermark, Österreich. Abgerufen am 2. Jänner 2017 von <http://www.soziales.steiermark.at/cms/beitrag/11980875/10024850/>
- Land Steiermark - Amt der Steiermärkischen Landesregierung. (kein Datum). Überbetriebliche Lehrausbildung. Graz, Steiermark, Österreich. Abgerufen am 2. Jänner 2017 von <http://www.soziales.steiermark.at/cms/beitrag/11327793/54867022/>
- Larisch, D. (2011). *Microsoft SharePoint Server 2010*. München: Carl Hanser Verlag.

- Lusti, M. (1999). *Data Warehousing und Data Mining*. Basel: Springer-Verlag Berlin Heidelberg.
- Microsoft Corporation. (kein Datum). Analysis Services. Abgerufen am 15. Juni 2017 von <https://msdn.microsoft.com/de-de/library/bb522607.aspx>
- Microsoft Corporation. (kein Datum). SQL Server Data Tools. Abgerufen am 15. Juni 2017 von [https://msdn.microsoft.com/en-us/library/hh272686\(v=vs.103\).aspx](https://msdn.microsoft.com/en-us/library/hh272686(v=vs.103).aspx)
- Microsoft Corporation. (kein Datum). SQL Server Integration Services. Abgerufen am 15. Juni 2017 von <https://msdn.microsoft.com/de-de/library/ms141026.aspx>
- Microsoft Corporation. (kein Datum). Was ist Power BI? Abgerufen am 16. Juni 2017 von <https://powerbi.microsoft.com/de-de/what-is-power-bi/>
- Müller, R. M., & Lenz, H.-J. (2013). *Business Intelligence*. Berlin Heidelberg: Springer-Verlag.
- Myatt, G. (2017). *Making Sense of Data*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Nagabhushana, S. (2006). *Data Warehousing*. New Delhi: New Age International (P) Ltd., Publishers.
- Navrade, F. (2008). *Strategische Planung mit Data-Warehouse-Systemen* (1. Ausg.). Wiesbaden: Betriebswirtschaftlicher Verlag Dr. Th. Gabler | HWV Fachverlage GmbH.
- Otey, M., & Otey, D. (2006). *Microsoft SQL Server 2005 Developer's Guide*. USA: The McGraw-Hill Companies / Osborne.
- Paragon Corporation. (30. November 2005). Choice Between Stored Procedures, Functions, Views, Triggers, Inline SQL (updated). Abgerufen am 15. Juni 2017 von <http://www.paragoncorporation.com/ArticleDetail.aspx?ArticleID=28>
- Ponniah, P. (2010). *Data warehousing fundamentals for IT professionals* (2. Ausg.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- Rad, R. (2014). *Microsoft SQL Server 2014 Business Intelligence Development Beginner's Guide*. Birmingham: Packt Publishing.
- Rainardi, V. (2008). *Building a Data Warehouse*. Berkeley, CA: Apress.
- Rossum, J., & Baccaro, R. (2015). *Extending SSIS with .NET Scripting*. Apress.
- Russo, M., & Ferrari, A. (2011). *PowerPivot for Excel 2010 - Give Your Data Meaning* (1 Ausg.). Redmond, Washington: Microsoft Press.
- Schicker, E. (2017). *Datenbanken und SQL* (5. Ausg.). Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

- Schön, D. (2012). *Planung und Reporting im Mittelstand*. Wiesbaden: Gabler Verlag | Springer Fachmedien.
- Silvers, F. (2008). *Building and Maintaining a Data Warehouse*. Boca Raton, FL: Taylor & Francis Group, LLC.
- Sozialministeriumservice. (kein Datum). Welche NEBA-Leistung kommt in Frage. Wien, Wien, Österreich. Abgerufen am 2. Jänner 2017 von <http://www.neba.at/index.php/neba/welche-neba-leistung-kommt-in-frage>
- Stadt Graz. (kein Datum). Informationen für Pflegefamilien. Graz. Abgerufen am 4. Jänner 2017 von <http://www.graz.at/cms/beitrag/10148296/3203249>
- STATISTIK AUSTRIA. (12. Juni 2017). Gemeinden. Abgerufen am 12. Juni 2017 von https://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/gemeinden/index.html
- STATISTIK AUSTRIA. (6. Juni 2017). Politische Bezirke. Abgerufen am 12. Juni 2017 von https://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/politische_bezirke/index.html
- Steiermärkische Gebietskrankenkasse. (4. Jänner 2017). *Frühe Hilfen - Gut begleitet von Anfang an!* Abgerufen am 4. Jänner 2017 von [Frühe Hilfen - Gut begleitet von Anfang an! www.stgkk.at/gutbegleitet](http://www.stgkk.at/gutbegleitet)
- Stigler, H. (14. Jänner 2001). Skriptum Methodologie.
- Sumroy, R., & Cousin, R. (Juli 2016). Personal data, anonymisation and pseudonymisation under the GDPR. (S. a. May, Hrsg.) Abgerufen am 21. Juni 2017 von <https://www.slaughterandmay.com/media/2535637/personal-data-anonymisation-and-pseudonymisation-under-the-gdpr.pdf>
- Taylor, J., & Cihon, C. (2004). *Statistical Techniques for Data Analysis* (2. Ausg.). Boca Raton London New York Washington, D.C.: CRC Press LLC.
- Thomsen, E. (2002). *OLAP Solutions* (2. Ausg.). John Wiley & Sons, Inc.
- Thorntwaite, W. (6. November 2008). Design Tip #107 Using the SQL MERGE Statement for Slowly Changing Dimension Processing. Abgerufen am 29. Mai 2017 von <http://www.kimballgroup.com/2008/11/design-tip-107-using-the-sql-merge-statement-for-slowly-changing-dimension-processing/>
- Unger, S. (10. August 2016). Änderungen durch die Datenschutz-Grundverordnung (Teil I). Wien. Abgerufen am 1. Juni 2017 von <https://www.unger->

rechtsanwaelte.at/aktuelles/detail/article/aenderungen_durch_die_datenschutz_grundverordnung
_teil_i.html

Wang, J. (2009). *Encyclopedia of Data Warehousing and Mining* (2. Ausg.). Hershey, Pennsylvania, USA: Information Science Reference (an imprint of IGI Global).

Wes, M. (25. April 2017). Looking to comply with GDPR? Here' a primer on anonymization and pseudonymization. Abgerufen am 21. Juni 2017 von <https://iapp.org/news/a/looking-to-comply-with-gdpr-heres-a-primer-on-anonymization-and-pseudonymization/>

Wirtschaftskammer Österreich. (19. Oktober 2016). EU-Datenschutz-Grundverordnung (DSGVO):. Abgerufen am 1. Juni 2017 von <https://www.wko.at/service/wirtschaftsrecht-gewerberecht/EU-Datenschutz-Grundverordnung:-Wichtige-Begriffsbestimmu.html>

Wirtschaftskammer Österreich. (14. April 2016). Rechtsdurchsetzung im Datenschutzrecht. Abgerufen am 1. Juni 2017 von https://www.wko.at/service/wirtschaftsrecht-gewerberecht/Die_Rechtsdurchsetzung_im_Datenschutzrecht.html

Wirtschaftskammer Österreich. (10. Mai 2017). EU-Datenschutz-Grundverordnung (DSGVO): Datenschutzbeauftragter. Abgerufen am 1. Juni 2017 von <https://www.wko.at/service/wirtschaftsrecht-gewerberecht/EU-Datenschutz-Grundverordnung:-Der-Datenschutzbeauftragt.html>

Wirtschaftskammer Österreich. (kein Datum). *Lehre statt Leere - Lehrlings und Betriebscoaching*. Abgerufen am 2. Jänner 2017 von www.lehre-statt-leere.at

ANHANG 1: ERHEBUNG DIENSTLEISTUNGEN UND DATENABLAGEN

Um eine unternehmensweite Informationsplattform im Sinne von Business Intelligence schaffen zu können, ist es notwendig, Informationen über die angebotenen Leistungen und den dort vorhandenen und verwendeten Informationen zu sammeln. Dazu werden nun, abgeleitet vom Organigramm, die einzelnen Zweigstellen analysiert und deren Dienstleistungsangebot aufgeschlüsselt. Anschließend werden diese Dienstleistungen hinsichtlich Dokumentation, Auftraggeber und vorhandenen Softwareapplikationen kurz beschrieben. Daraus ergibt sich abschließend, welche Applikationen für den Prototyp des Datawarehouse herangezogen werden.

Zweigstellen

Nachfolgend werden die einzelnen Zweigstellen mithilfe des Bereichs „Organigramme“ im JAW-Qualitätsmanagementsystem⁸ aufgelistet, um anschließend diese Dienstleistungen kurz beschreiben sowie auf Auftraggeber und Dokumentationssystem eingehen zu können. Als zusätzliche Datenquelle dient die Beschreibung der Dienstleistungen auf der Homepage von JAW, www.jaw.or.at (Stand: 2.Jänner 2017).

- Arbeitsassistentz – JobCoaching
 - Arbeitsassistentz
 - JobCoaching
- Ausbildungszentrum Graz
 - ÜBA
 - Perspektivenwerkstatt
 - IBA
 - BeTrain
 - Lehrlingscoaching

⁸ Das JAW-Qualitätsmanagementsystem nach ISO 9001:2008 ist für sämtliche Mitarbeiter im Intranet abrufbar.

- Beratungszentrum Voitsberg
 - Mobile Dienstleistungen nach StKJHG-DVO
 - Mobile Dienstleistungen nach LEVO-StBHG
 - Streetwork
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
- Flüchtlingsbetreuung
 - Grundversorgung
 - UMF
- Kompetenz- und Ausbildungszentrum Deutschlandsberg
 - Produktionsschule
 - Bewerbungsunterstützung
 - Jugendcoaching
 - ÜBA
 - Perspektivenwerkstatt
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
- Kompetenz- und Ausbildungszentrum Knittelfeld
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
- Kompetenz- und Ausbildungszentrum Leoben/Eisenerz
 - Produktionsschule
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
 - Mobile Dienstleistungen nach LEVO-StBHG
 - GBP
- Kompetenz- und Ausbildungszentrum Mureck
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
- Kompetenz- und Ausbildungszentrum Mürzzuschlag
 - IBA
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
 - Produktionsschule

- Kompetenz- und Ausbildungszentrum Kapfenberg & Wohnverbund Kapfenberg
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
 - Stationäre Dienstleistungen nach LEVO-StBHG
- Kompetenz- und Ausbildungszentrum Liezen & Wohnverbund Liezen
 - ÜBA
 - Perspektivenwerkstatt
 - IBA
 - Inabschluss
 - Mobile Dienstleistungen nach LEVO-StBHG
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
 - GBP
 - Jobfit
 - Stationäre Dienstleistungen nach LEVO-StBHG
- Leander
 - Mobile Dienstleistungen nach StKJHG-DVO
- mobil.betreut Region Oststeiermark
 - Mobile Dienstleistungen nach LEVO-StBHG
 - Mobile Dienstleistungen nach StKJHG-DVO
 - Frühe Hilfen
- Neuland
 - 24-Stunden-Betreuung
 - Jugendzentrum
- Perspektivencheck
 - Perspektivencheck für Jugendliche in Graz
- Perspektivenzentrum Graz
 - Produktionsschule
 - Jugendpark
 - GBP

- Soziale Dienste Süd – Jugend und Familienbegleitung (SDS)
 - Mobile Dienstleistungen nach StKJHG-DVO
- Stella
 - Mobile Dienstleistungen nach StKJHG-DVO
- Tartaruga
 - Stationäre Dienstleistungen nach StKJHG-DVO
- Verbund – Kompetenz- und Ausbildungszentren Graz Ost
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
- Verbund – Kompetenz- und Ausbildungszentren Graz West
 - Teilstationäre Dienstleistungen nach LEVO-StBHG
- Wohnverbund Graz
 - Stationäre Dienstleistungen nach LEVO-StBHG
 - Mobile Dienstleistungen nach StKJHG-DVO
- Wohnverbund & mobil.betreut Knittelfeld
 - Mobile Dienstleistungen nach LEVO-StBHG
 - Stationäre Dienstleistungen nach LEVO-StBHG
- Wohnverbund & mobil.betreut Mureck
 - Mobile Dienstleistungen nach LEVO-StBHG
 - Stationäre Dienstleistungen nach LEVO-StBHG
- Zentrum R6
 - Mobile Dienstleistungen nach StKJHG-DVO
 - Housing First
 - Selbsthilfekontaktstelle
 - Pflegeeltern-Betreuung
 - Flexible Hilfen
 - Stationäre Dienstleistungen nach StKJHG-DVO
- Zentrum SR3
 - Mobile Dienstleistungen nach StKJHG-DVO

- Jugendzentrum
- Vinzenz-Muchitsch-Haus – bespielen-betreuen-begegnen
- Zweigstelle Bruck/Mur
 - Mobile Dienstleistungen nach LEVO-StBHG
 - Stationäre Dienstleistungen nach StKJHG-DVO

Dienstleistungen

Nun werden die Dienstleistungen kurz beschrieben und Auftraggeber sowie Dokumentationssystem erwähnt. Das erste Auftreten der Dienstleistung in der Liste der Zweigstellen entscheidet über nachfolgende Reihenfolge, jede Dienstleistung wird nur einmalig beschrieben. Auf Basis dieser Beschreibung wird in der Folge eine tabellarische Darstellung dieser Informationen erstellt, mit deren Hilfe dann die auf die jeweiligen Applikationen und deren Datenstrukturen eingegangen werden kann.

Arbeitsassistentz

Die Arbeitsassistentz unterstützt Menschen mit Behinderungen mit Assistenzbedarf den gesamten Prozess von Situationsanalyse über Arbeitssuche bis hin zum Ersteinstieg in die Berufswelt. Darüber hinaus kann die Arbeitsassistentz auch zur Krisenintervention zur Sicherung eines Arbeitsplatzes herangezogen werden. (Sozialministeriumservice)

Diese Dienstleistung wird im Auftrag von NEBA, einer Marke des SOMS, durchgeführt. Die Aufzeichnungspflichten werden mithilfe des Datenbanksystems Be-FIT (Betriebs-Fach-Informationstechnologie) des Sozialministeriumservice erfüllt.

Jobcoaching

Das Jobcoaching hat eine ähnliche Aufgabenstellung wie die Arbeitsassistentz, richtet sich aber an alle Menschen mit Lernschwierigkeiten. Es wird im Zuge des Berufseinstiegsprozesses auch die Möglichkeit geschaffen, Praktika und Berufserprobungen zu absolvieren. (Sozialministeriumservice)

Diese Dienstleistung wird ebenfalls im Auftrag von NEBA, einer Marke des SOMS, durchgeführt. Die Aufzeichnungspflichten werden daher auch hier mithilfe des Datenbanksystems Be-FIT des Sozialministeriumservice erfüllt.

Überbetriebliche Lehrausbildung (ÜBA)

Die ÜBA unterstützt Jugendliche mit Problemen bei der Lehrstellensuche bei ebendieser. Auch mithilfe der großen Vernetzung von JAW werden die Jugendlichen an Lehrstellen in Firmen vermittelt. Auch während der Lehre in diesen Firmen und der damit einhergehenden Berufsschule begleitet die ÜBA die Jugendlichen und hilft sowohl bei schulischen als auch bei arbeitsrelevanten Fragestellungen. (Land Steiermark - Amt der Steiermärkischen Landesregierung)

Die gesamte Stammdatenverwaltung und laufende Dokumentation der ÜBA findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. ÜBA ist eine im Auftrag des AMS durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist.

Perspektivenwerkstatt

Bei der Perspektivenwerkstatt handelt es sich um eine Vorabklärung für Lehrstellensuchende, um die weitere Vorgangsweise in Richtung Lehrstelle, ÜBA oder IBA festzustellen. (Arbeitsmarktservice Steiermark, 2016, S. 22)

Die gesamte Stammdatenverwaltung und laufende Dokumentation der Perspektivenwerkstatt findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. ÜBA ist eine im Auftrag des AMS durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist.

Integrative Berufsausbildung (IBA)

Die IBA richtet sich an diejenigen Personen, die auch mithilfe der ÜBA keine Lehrstelle finden und bietet die Möglichkeit, durch Verlängerung der Lehrzeit zu einer Voll- oder Teilqualifizierung in einem Lehrberuf zu gelangen. (Land Steiermark - Amt der Steiermärkischen Landesregierung)

Die gesamte Stammdatenverwaltung und laufende Dokumentation der IBA findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. IBA ist eine im Auftrag des AMS durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist. Durch eine zusätzliche Beteiligung des Landes Steiermark wird auch auf die Funktionalitäten von BMD NTCS zurückgegriffen.

BeTrain

Das Projekt BeTrain (Begleitung – Training – Integration) bildet Menschen mit Behinderung in einer dualen Ausbildung zu Callcenter Agents/Customer Care Agents aus. Dazu arbeiten diese produktiv in den Telefonvermittlungen für die Jugend am Werk Steiermark GmbH und das bfi Steiermark.

Diese Dienstleistung wird im Auftrag des Sozialministeriumservice durchgeführt, die Aufzeichnungspflichten werden mithilfe des Datenbanksystems Be-FIT (Betriebs-Fach-Informationstechnologie) des Sozialministeriumservice erfüllt.

Lehrlingscoaching

Unter dem Motto „Lehre statt Leere“ steht das Lehrlingscoaching nicht nur Lehrlingen, sondern auch Lehrbetrieben mit Rat und Tat zur Seite, sollten berufliche oder inhaltliche Differenzen in einer Lehre auftreten. (Wirtschaftskammer Österreich, kein Datum)

Das Lehrlingscoaching wird im Auftrag von WKO und Sozialministeriumservice durchgeführt. Zur Dokumentation wird auch hier die von JAW eigenentwickelte Datenbankapplikation „AABDB“ herangezogen.

Mobile Dienstleistungen nach StKJHG-DVO

Auf Basis des Steiermärkischen Kinder- und Jugendhilfegesetzes (StKJHG) gibt es unterschiedliche mobile Dienstleistungen, die in dessen Durchführungsverordnung (DVO) beschrieben sind. Viele davon bietet JAW an:

- Interdisziplinäre Frühförderung und Familienbegleitung (IFF)
- Sozialpädagogische Kinder- und Jugendbetreuung (SKJB)
- Sozialpädagogische Familienbetreuung (SFB)
- Familienhilfe (FAMH)
- Krisendienst für Familien (KD-FAM)
- Psychologische Behandlung (PSYBEH)
- Psychotherapie (PSYTHER)
- Sozialbetreuung (SOZBET)

Die Daten dieser mobilen Dienstleistungen werden hauptsächlich in der Datenbankapplikation „agilion“ von agilion GmbH abgespeichert. Zusätzlich gibt es noch eine Verlaufsdocumentation und andere Fließtextdokumente am Fileserver.

Mobile Dienstleistungen nach LEVO-StBHG

Auf Basis des Steiermärkischen Behindertengesetzes (StBHG) gibt es unterschiedliche mobile Dienstleistungen, die in dessen Leistungs- und Entgeltverordnung (LEVO) beschrieben sind. Sämtliche dieser mobilen Dienstleistungen bietet JAW an:

- Interdisziplinäre Frühförderung und Familienbegleitung (IFF)
- Interdisziplinäre Sehfrühförderung und Familienbegleitung (IFF-Seh)
- Interdisziplinäre audiologische Frühförderung und Familienbegleitung (IFF-Hör)
- Wohnassistenz (ASS-W)
- Familienentlastungsdienst (FED BHG)
- Freizeitassistenz (ASS-F)
- Mobile sozialpsychiatrische Betreuung (MS-BET PSY)

Die Daten dieser mobilen Dienstleistungen werden hauptsächlich in der Datenbankapplikation „agilion“ von agilion GmbH abgespeichert. Zusätzlich gibt es noch eine Verlaufsdocumentation und andere Fließtextdokumente am Fileserver.

Streetwork

Hierbei handelt es sich um eine anonyme, mobile Form der Jugendarbeit, bei der die Mitarbeiter aktiv auf Jugendliche an offenen Plätzen, in Parks und in Lokalen zugehen und Unterstützung bei schwierigen Lebenssituationen anbieten. (Land Steiermark - Amt der Steiermärkischen Landesregierung)

Diese Dienstleistung wird im Auftrag des Landes Steiermark durchgeführt. Dadurch, dass es sich um eine anonyme Dienstleistung handelt, gibt es keine Aufzeichnungen über Kundenstammdaten.

Teilstationäre Dienstleistungen nach LEVO-StBHG

Die Teilstationären Dienstleistungen laut LEVO -StBHG sind Tagesbegleitung & Förderung (B&F BHG) sowie Teilhabe an Beschäftigung in der Arbeitswelt (TaB BHG). JAW bietet diese beide an. B&F richtet sich an Menschen mit Behinderung, die in Betrieben am Arbeitsmarkt keinen Arbeitsplatz finden können. TaB richtet sich an Menschen, die sehr wohl am Arbeitsmarkt arbeiten können – allerdings nicht als dortige Angestellte, sondern über den Sozialträger in Form von Praktika und Probearbeiten vermittelt werden, mit dem Ziel, dort eine Festanstellung zu bekommen und damit von den Sozialleistungen unabhängig zu sein.

Die Datenablage dieser Dienstleistungen ist zurzeit im Wandel. Der inhaltlich relevante Teil sowie die Aufzeichnung der laufenden, abrechnungsrelevanten Informationen passiert in der JAW-Eigenentwicklung „KundInnenakt stationär“ auf Basis Microsoft SharePoint. Von dort werden die Daten an die Finanzbuchhaltungssoftware BMD NTCS übergeben, wo die führende Stammdatenverwaltung erfolgt sowie die Abrechnung durchgeführt wird. Zusätzlich gibt es noch einzelne Dokumente am Fileserver, die in den „KundInnenakt stationär“ eingearbeitet werden.

Grundversorgung

JAW bietet für Asylwerber die Teil-Selbstversorgung an. Dazu werden diese in kleinen Wohnungen untergebracht, um gemeinsam als Familie oder in einer WG-artigen Umgebung zu leben. (Land Steiermark - Amt der Steiermärkischen Landesregierung)

Die laufende Verwaltung der Asylwerber und deren Unterkünfte wird innerhalb einer Excel-Applikation abgewickelt. Zur Abrechnung werden diese Stammdaten in die Finanzbuchhaltungssoftware BMD NTCS übergeben, wo abschließend die Abrechnung durchgeführt wird. Zusätzlich gibt es noch einzelne Dokumente am Fileserver.

Unbegleitete minderjährige Flüchtlinge (UMF)

Gemäß der Grundversorgungsvereinbarung zwischen Bund und Ländern erhalten minderjährige, die ohne gesetzliche Vertretung um Asyl ansuchen, eine besondere, ihrem Alter entsprechende Betreuung. (Land Steiermark - Amt der Steiermärkischen Landesregierung) Diese ist weitaus umfassender angelegt, die Datenhaltung bei JAW erfolgt aber analog zur Grundversorgung.

Produktionsschule

Die Produktionsschule hilft Jugendlichen dabei, in der Arbeitswelt Fuß zu fassen, indem diese ab dem ersten Tag produktiv in den Bereichen Holz, Metall oder Küche eingesetzt werden und deren Erzeugnisse anschließend auch marktüblich verkauft werden.

Die Produktionsschule ist eine Maßnahme, die von AMS und SOMS gemeinsam finanziert wird. Die Aufzeichnung erfolgt via be-FIT und eAMS.

Bewerbungsunterstützung

Diese Dienstleistung unterstützt Arbeitssuchende bei der Erstellung und Optimierung von Bewerbungsunterlagen und bereitet auch auf Vorstellungsgespräche vor.

Die gesamte Stammdatenverwaltung und laufende Dokumentation der Bewerbungsunterstützung findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. Die Bewerbungsunterstützung ist eine im Auftrag des AMS durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist.

Jugendcoaching

Das Jugendcoaching im Auftrag von NEBA und Sozialministeriumservice hilft Jugendlichen dabei, ihre beruflichen Interessen abzustecken und sie individuell an entsprechende Bereiche am Arbeitsmarkt heranzuführen. (Sozialministeriumservice)

Diese Dienstleistung wird ebenfalls im Auftrag von NEBA, einer Marke des SOMS, durchgeführt. Die Aufzeichnungspflichten werden daher auch hier mithilfe des Datenbanksystems Be-FIT des Sozialministeriumservice erfüllt.

Gemeinnütziges Beschäftigungsprojekt (GBP)

GBPs unterstützen Langzeitarbeitslose und andere arbeitsmarktpolitische Problemgruppen bei der Integration in den Arbeitsmarkt. Dazu werden diese bei JAW angestellt und arbeiten ab dem ersten Tag wie angestellt. Die Arbeitskraft dieser Personen wird, wie bei allen anderen Arbeitgebern auch, voll verkauft – sei es in Form von Mittagsmenüs, Botengängen, handwerklichen Erzeugnissen oder sozialen Dienstleistungen.

Diese Dienstleistung wird im Auftrag des AMS durchgeführt, als Datenapplikation kommt die „GBPDB“ zum Einsatz; ebenfalls eine Eigenentwicklung von JAW. Zusätzlich, wie bei allen AMS-Projekten, wird intensiv mit dem eAMS gearbeitet.

Stationäre Dienstleistungen nach LEVO-StBHG

Das Land Steiermark bietet, in der LEVO-StBHG genormt, mehrere stationäre Dienstleistungen an, die alle von JAW realisiert werden:

- Vollzeitbetreutes Wohnen für Menschen mit Behinderung (WH BHG)
- Trainingswohnung für Menschen mit Behinderung (TW BHG)
- Teilzeitbetreutes Wohnen (TBEW BHG)

Die Datenablage erfolgt analog zu Teilstationäre Dienstleistungen nach LEVO-StBHG.

Inabschluss

Inabschluss ist eine schulische Maßnahme, die den Teilnehmern sämtliches Wissen aus vier Jahren Hauptschule in einem Jahr beibringt – wodurch Erwachsene sämtlicher Altersklassen ihren Pflichtschulabschluss nachholen können.

Diese Maßnahme wird vom Land Steiermark finanziert, die Aufzeichnung erfolgt in einer speziellen Excel-Applikation.

Jobfit

Menschen ab dem 15. Lebensjahr, die beim AMS als arbeitslos vermerkt sind, aber realistische Berufsziele haben, werden dabei unterstützt, diese in die Tat umzusetzen.

Die gesamte Stammdatenverwaltung und laufende Dokumentation von Jobfit findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. Jobfit ist eine im Auftrag des AMS durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist.

Frühe Hilfen

Die „Frühen Hilfen“ unterstützen Familien in der Anfangsphase, solange das Kind noch sehr klein ist (0 bis 3 Jahre) in sämtlichen Belangen rund um das Kind. (Steiermärkische Gebietskrankenkasse, 2017)

Diese Dienstleistung wird im Auftrag der StGKK erbracht. Die Abrechnung erfolgt lediglich auf Basis der Arbeitsstunden, die von den Mitarbeitern geleistet wurden; die Aufzeichnungen über die Kunden werden am Fileserver abgelegt.

Neuland 24-Stunden-Betreuung

Neuland ist eine intensive Form der Wohnbetreuung mitsamt Tagesstruktur - eine 24-Stunden-Dienstleistung für Menschen mit schwerer Behinderung und herausforderndem Verhalten.

Die Datenablage erfolgt analog zu Teilstationäre Dienstleistungen nach LEVO-StBHG.

Jugendzentrum

JAW betreibt zwei Jugendzentren im Auftrag der Stadt Graz, die Jugendlichen einerseits einen Platz zum Verbringen ihrer Freizeit bieten soll aber andererseits auch sozialpädagogische Hilfestellungen bietet.

Finanziert werden die Jugendzentren auf Basis der dort arbeitenden Mitarbeiter, die Jugendlichen bleiben in der Regel anonym, wodurch es keine zu verwaltenden Stammdaten gibt.

Perspektivencheck Graz für Jugendliche

Der Perspektivencheck richtet sich an Jugendliche, die beim AMS gemeldet sind und auf der Suche nach einer Lehrstelle oder einem Arbeitsplatz sind. Es wird dabei geholfen, einen persönlichen Perspektivenplan zu erstellen und auf eine Lehrstelle bzw. einen Arbeitsplatz vorzubereiten. Darüber hinaus wird auch bei Lehrstellen- bzw. Arbeitsplatzsuche unterstützt.

Die gesamte Stammdatenverwaltung und laufende Dokumentation des Perspektivenchecks findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. Der Perspektivencheck für Jugendliche ist eine im Auftrag des AMS durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist.

Jugendpark

Der Jugendpark richtet sich an junge Menschen zwischen 15 und 25 Jahren, die beim AMS gemeldet sind, die bisher noch keine passende Ausbildung, Lehrstelle oder Arbeitsplatz finden konnten. Der Fokus liegt auf handwerkliche Arbeiten, die das Selbstbewusstsein und die Freude am Arbeiten schüren sollen und auch bei der Stellensuche wird unterstützt.

Die gesamte Stammdatenverwaltung und laufende Dokumentation des Jugendparks findet in der von JAW eigenentwickelten Datenbankapplikation „AABDB“ statt. Der Jugendpark ist eine im Auftrag des AMS und der Stadt Graz durchgeführte Maßnahme, wodurch auch das AMS-System „eAMS“ an der laufenden Dokumentation beteiligt ist.

Stationäre Dienstleistungen nach StKJHG-DVO

Das Land Steiermark bietet, in der StKJHG-DVO genormt, mehrere stationäre Dienstleistungen an, wovon die meisten von JAW realisiert werden:

- Sozialpädagogische Wohngemeinschaft für Kinder und Jugendliche (WG-SPÄD)
- Kriseninterventionsstelle/Krisenunterbringung (KRISE)
- Betreutes Wohnen (MOB)
- Betreute Wohngruppe (MOB-WG)
- Betreutes Wohnen in Krisensituationen (MOB-KRISE)
- Betreutes Wohnen von jugendlichen Familien (MOB-FAM)

Die Datenablage erfolgt größtenteils analog zu Teilstationäre Dienstleistungen nach LEVO-StBHG. Einzig bei der Dienstleistung Kriseninterventionsstelle/Krisenunterbringung wird noch nicht auf den „KundInnenakt stationär“ gesetzt, sondern auf eine Kombination aus Office-Dokumenten am Fileserver und BMD NTCS.

Housing First

Housing First bietet Frauen, die aus unterschiedlichsten Gründen in die Wohnungslosigkeit geraten und in einer Notschlafstelle oder einem Frauenhaus untergebracht werden die Möglichkeit, selbst wieder Fuß zu fassen, indem gemeinsam mit den Mitarbeitern eine neue Wohnung gesucht wird, sämtliche Behördenwege erledigt werden und auch praktische handwerkliche Fähigkeiten beigebracht werden. Dies alles wird von einer intensiven sozialpädagogischen Betreuung begleitet.

Housing First ist eine von der Stadt Graz beauftragte Dienstleistung. Die Daten der betreuten Frauen werden nur in Form von Office-Dokumenten am Fileserver abgelegt und nur anonymisiert an den Auftraggeber weitergegeben.

Selbsthilfekontaktstelle

Die Selbsthilfekontaktstelle ist eine Initiative von Land Steiermark und Stadt Graz, um Selbsthilfegruppen dabei zu helfen, bekannter zu werden, etwaige Behördenwege zu absolvieren, Veranstaltungen zu organisieren und auch beim Aufbau und internen Abläufen von Selbsthilfegruppen wird unterstützt. Die Tätigkeiten sind sehr transparent über die Homepage www.selbsthilfe-stmk.at ersichtlich, die Kunden sind immer Vereine bzw. Personengruppen und keine Einzelpersonen, weshalb es keine Kundenstammdaten im Sinne dieser Masterarbeit gibt.

Pflegeeltern

Gemeinsam mit a:pfl alternative:pflegefamilie gmbh stellt JAW für Pflegefamilien im Rahmen der Dienstleistung „Sozialversicherungsrechtliche Absicherung bzw. qualitätssichernde Hilfen/Maßnahmen zur Festigung des Pflegeverhältnisses“ des Landes Steiermark systematische Beratungs- und Qualitätssicherungsangebote zur Verfügung was bedeutet, dass diese ein Dienstverhältnis bei JAW eingehen, um sozialversicherungsrechtlich abgesichert zu sein. Insofern sind die Kunden keine Kunden im eigentlichen Sinn mehr, sondern Mitarbeiter und werden daher im Rahmen dieser Masterarbeit nicht mehr weiter beachtet.

(Stadt Graz) (Land Steiermark - Amt der Steiermärkischen Landesregierung, 2011)

Flexible Hilfen

Die flexiblen Hilfen sind eine besondere Auslegungsform des StKJHG, die nicht direkt in Dienstleistungen der LEVO einordbar sind, sondern mit einem gesonderten Vertrag mit dem Land Steiermark individuell zugeschnittene Sozialdienstleistungen anbieten. Die flexiblen Hilfen werden in Arbeitsgemeinschaften mit anderen Sozialträgern gemeinsam durchgeführt, um die komplette Palette an Qualifikationen und Dienstleistungsangeboten der Mitarbeiter zur Verfügung stellen zu können, welche die Kunden potentiell benötigen. (Land Steiermark - Amt der Steiermärkischen Landesregierung)

Die Verwaltung der Kundendaten wird in einer gesonderten Datenbankapplikation der jeweiligen ARGE abgewickelt. Da es sich von Natur aus nicht um reine JAW-Daten handelt, die in dieser

ARGE-Datenbank verwendet werden, werden diese von dieser Masterarbeit ausgeschlossen, um Komplikationen mit den ARGE-Partnern zu vermeiden.

Vinzenz-Muchitsch-Haus – bespielen-betreuen-begegnen

Im Auftrag der Stadt Graz betreibt JAW das Vinzenz-Muchitsch-Haus mit der Aufgabe, sozialräumliche Dienstleistungen anzubieten. Dazu werden Veranstaltungen organisiert, Räumlichkeiten zur Verfügung gestellt und Treffen von Interessensgruppen abgehalten.

Da die Kunden des Vinzenz-Muchitsch-Hauses keine natürlichen Einzelpersonen sind, wird es in dieser Masterarbeit nicht weiter behandelt.

ANHANG 2: SCRIPT ZUR ANLAGE DER DIMENSION ZEIT

```

CREATE TABLE #FixeFeiertage
(
  Monat INT,
  Tag INT,
  Name NVARCHAR(255)
)

INSERT INTO #FixeFeiertage VALUES
(1,1,'Neujahr'),
(1,6,'Heilige Drei Könige'),
(5,1,'Staatsfeiertag'),
(8,15,'Maria Himmelfahrt'),
(10,26,'Nationalfeiertag'),
(11,1,'Allerheiligen'),
(12,8,'Maria Empfängnis'),
(12,24,'Heiliger Abend (SWÖ-KV)'),
(12,25,'Weihnachten'),
(12,26,'Stephanstag'),
(12,31,'Silvester (SWÖ-KV)')
;
GO
DROP FUNCTION dbo.GetEasterHolidays
GO
CREATE FUNCTION dbo.GetEasterHolidays(@year INT)
RETURNS TABLE
WITH SCHEMABINDING
AS
RETURN
(
  WITH x AS
  (
    SELECT Datum = CONVERT(DATE, RTRIM(@year) + '0' + RTRIM([Month])
      + RIGHT('0' + RTRIM([Day]),2))
      FROM (SELECT [Month], [Day] = DaysToSunday + 28 - (31 * ([Month] / 4))
      FROM (SELECT [Month] = 3 + (DaysToSunday + 40) / 44, DaysToSunday
      FROM (SELECT DaysToSunday = paschal - ((@year + @year / 4 + paschal - 13) % 7)
      FROM (SELECT paschal = epact - (epact / 28)
      FROM (SELECT epact = (24 + 19 * (@year % 19)) % 30)
      AS epact) AS paschal) AS dts) AS m) AS d
  )
  SELECT [Datum], FeiertagName = 'Ostersonntag' FROM x
  UNION ALL SELECT DATEADD(DAY, 1,Datum), 'Ostermontag' FROM x
  UNION ALL SELECT DATEADD(DAY, 39,Datum), 'Christi Himmelfahrt' FROM x
  UNION ALL SELECT DATEADD(DAY, 50,Datum), 'Pfingstmontag' FROM x
  UNION ALL SELECT DATEADD(DAY, 60,Datum), 'Frohnleichnam' FROM x
);
GO

DECLARE @StartDate date = '1900-01-01'
DECLARE @NumberOfYears int = 500
DECLARE @EndDate date = DATEADD(YEAR,@NumberOfYears,@StartDate)

SELECT
  CONVERT(INT,FORMAT(d, 'yyyyMMdd')) [Key],
  d Datum,
  DATEPART(DAY, d) Tag,
  DATEPART(WEEKDAY, d) WochentagNr,

```

```

DATEPART(WEEK, d) WocheImJahr,
DATEPART(ISO_WEEK, d) ISOWocheImJahr,
DATEPART(MONTH, d) Monat,
DATEPART(QUARTER, d) Quartal,
DATEPART(YEAR, d) Jahr,
CONVERT(BIT,0) IstFeiertag,
CONVERT(NVARCHAR(255),NULL) Feiertag,
DATEFROMPARTS(YEAR(d), MONTH(d), 1) ErsterTagImMonat,
DATEFROMPARTS(YEAR(d), 1,1) ErsterTagImJahr
INTO #tmp
FROM (
SELECT d = DATEADD(DAY, rn - 1, @StartDate)
FROM
(
SELECT TOP (DATEDIFF(DAY, @StartDate, @EndDate)) rn = ROW_NUMBER() OVER (ORDER BY
s1.[object_id])
FROM sys.all_objects AS s1
CROSS JOIN sys.all_objects AS s2
ORDER BY s1.[object_id]
) AS x
) AS y
GO
;WITH x AS
(
SELECT d.Datum, d.Feiertag, h.FeiertagName, d.IstFeiertag
FROM #tmp AS d
CROSS APPLY dbo.GetEasterHolidays(d.Jahr) AS h
WHERE d.Datum = h.Datum
)
UPDATE x SET IstFeiertag = 1, Feiertag = FeiertagName;
;WITH x AS
(
SELECT d.Datum, d.Feiertag, h.FeiertagName, d.IstFeiertag
FROM #tmp AS d
CROSS APPLY (SELECT DATEFROMPARTS(d.Jahr,f.Monat,f.Tag) Datum, Name FeiertagName FROM
#FixeFeiertage f) AS h
WHERE d.Datum = h.Datum
)
UPDATE x SET IstFeiertag = 1, Feiertag = FeiertagName;
GO

INSERT INTO DimZeit
SELECT
[Key],
Datum,
Tag,
CASE WHEN WochentagNr BETWEEN 2 AND 6 THEN 1 ELSE 0 END IstWochentag,
FORMAT(Datum, 'dddd', 'de-AT') Wochentag,
WochentagNr,
IstFeiertag,
Feiertag,
CONVERT(TINYINT, ROW_NUMBER() OVER
(PARTITION BY ErsterTagImMonat, WochentagNr ORDER BY Datum))
WochentagNrImMonat,
DATEPART(DAYOFYEAR, Datum) TagNrImJahr,
CONVERT(TINYINT, DENSE_RANK() OVER
(PARTITION BY Jahr, Monat ORDER BY WocheImJahr)) WocheImMonat,
WocheImJahr,
ISOWocheImJahr,
Monat,
FORMAT(Datum, 'MMMM', 'de-AT') MonatName,

```

```
CONVERT(TINYINT, Quartal) Quartal,  
Quartal+'. Quartal' QuartalName,  
Jahr,  
CONVERT(INT, FORMAT(Datum, 'MMyyyy')) MMJJJJ,  
FORMAT(Datum, 'MMM yyyy', 'de-AT') MonatJahr,  
ErsterTagImMonat,  
EOMONTH(ErsterTagImMonat) LetzterTagImMonat,  
MIN(Datum) OVER (PARTITION BY Jahr, Quartal) ErsterTagImQuartal,  
MAX(Datum) OVER (PARTITION BY Jahr, Quartal) LetzterTagImQuartal,  
ErsterTagImJahr,  
MAX(Datum) OVER (PARTITION BY Jahr) LetzterTagImJahr,  
DATEADD(MONTH, 1, ErsterTagImMonat) ErsterTagNaechstesMonat,  
DATEADD(YEAR, 1, ErsterTagImJahr) ErsterTagNaechstesJahr  
FROM #tmp  
GO  
DROP TABLE #tmp  
DROP TABLE #FixeFeiertage
```

Quelle: Mischung aus eigener Entwicklung sowie von mssqltips.com. (Bertrand, 2015)

ANHANG 3: SCRIPT ZUR ANLAGE DER TRANSFORMATION „DIENSTLEISTUNG“ DER DATENQUELLE „BMD“

```
CREATE VIEW BMDAW.[DimDienstleistung]
AS
SELECT
    Nummer NaturalKey,
    Titel Bezeichnung,
    kommentar BezeichnungKurz,
CASE
WHEN Nummer IN (4002,4010,4011,4012)
OR Nummer BETWEEN 4050 AND 4068
OR Nummer BETWEEN 4080 AND 4814
OR Nummer BETWEEN 4850 AND 4862
THEN 'StBHG-TS'

WHEN Nummer IN (4003,4008)
OR Nummer BETWEEN 4069 AND 4079
OR Nummer BETWEEN 4815 AND 4819
THEN 'StBHG-Wohn'

WHEN Nummer IN (4004,4012)
OR NUMMER BETWEEN 4820 AND 4849
THEN 'StKJHG'

WHEN Nummer BETWEEN 4870 AND 4872
THEN 'Flü'

WHEN Nummer = 4874 THEN 'SbgBHG'

WHEN Nummer = 48705 THEN 'SbgFlü'

ELSE
'unbekannt - '+Nummer
END Typ,
'Gesetz' Auftraggeber,
'BMDAW' Datenquelle
FROM FVW_FELDVORGABEWERTE
```

ANHANG 4: SCRIPT ZUR ANLAGE DER TRANSFORMATION „KUNDE“ DER DATENQUELLE „BMD“

```

CREATE VIEW BMDAW.[DimKunde] AS

SELECT
    PER_PERSONENNR,
    Vorname,
    Nachname,
    Geschlecht,
    SVNR,
    Ort Wohnort,
    PLZ,
    Strasse,
    (select top 1 [politischer bezirk] from PLZBezirke where (plz = postleitzahl AND
(ort=gemeindenname OR ort=ortschaftsname)) or (CharINDEX(plz,postleitzahl) > 0 AND
(ort=gemeindenname OR ort=ortschaftsname)) OR (CharINDEX(plz,postleitzahl) > 0)
    order by case when (plz = postleitzahl AND (ort=gemeindenname OR
ort=ortschaftsname)) then 1
                    when CharINDEX(plz,postleitzahl) > 0 and (ort=gemeindenname OR
ort=ortschaftsname) then 2
                    when (CharINDEX(plz,postleitzahl) > 0) then 3
                    end
) Bezirk,
    (select top 1 Bundesland from PLZBezirke where (plz = postleitzahl AND
(ort=gemeindenname OR ort=ortschaftsname)) or (CharINDEX(plz,postleitzahl) > 0 AND
(ort=gemeindenname OR ort=ortschaftsname)) OR (CharINDEX(plz,postleitzahl) > 0)
    order by case when (plz = postleitzahl AND (ort=gemeindenname OR
ort=ortschaftsname)) then 1
                    when CharINDEX(plz,postleitzahl) > 0 and (ort=gemeindenname OR
ort=ortschaftsname) then 2
                    when (CharINDEX(plz,postleitzahl) > 0) then 3
                    end
) Bundesland,
    'Österreich' Land,
    Telefonnummer,
    Mobilnummer,
    Anrede,
    [Titel vorne],
    [Titel hinten],
    Staatsbuergerschaft,
    Geburtsland,
    CONVERT(DATE,Geburtsdatum) Geburtsdatum,
    'BMDAW' Datenquelle
FROM (
    SELECT
        PER_PERSONENNR, -- muss zu IdentifierKey werden
        PER_VORNAME Vorname,
        PER_NAME Nachname,
        CASE WHEN PER_GESCHLECHT = 0 THEN 'männlich' WHEN PER_GESCHLECHT=1 THEN 'weiblich'
ELSE 'unbekannt' END Geschlecht,
        PER_SVNR SVNR,
        ISNULL(s.Ort,j.Ort) Ort,
        ISNULL(s.PLZ,j.PLZ) PLZ,
        ISNULL(s.Straße,j.Straße) Strasse,
        ISNULL(s.Telefonnummer,j.Telefonnummer) Telefonnummer,
        ISNULL(s.Handy,j.Handy) Mobilnummer,
        CONVERT(NVARCHAR(255),NULL) Anrede,
        CONVERT(NVARCHAR(255),NULL) [Titel vorne],

```

```
    CONVERT(NVARCHAR(255),NULL) [Titel hinten],
    staaten.staat Staatsbuergerschaft,
    ISNULL(s.Geburtsort,j.Geburtsort) Geburtsland,
    ISNULL(ISNULL(s.Geburtsdatum,j.Geburtsdatum),PER_GEBURTSDATUM) Geburtsdatum
FROM
    PER_PERSONEN p
    LEFT JOIN BHGKomAZ k ON p.PER_PERSONENNR=k.KundenNr
    LEFT JOIN (SELECT ROW_NUMBER() OVER (PARTITION BY KundenNr ORDER BY Geändert
DESC) AS rn, * FROM BHGWohnen w) w ON p.PER_PERSONENNR=w.KundenNr AND w.rn=1
    LEFT JOIN BHGStammdatenblatt s ON (k.KundeId IS NOT NULL AND k.KundeId = s.ID) OR
(w.KundeId IS NOT NULL AND w.KundeId = s.ID)
    LEFT JOIN KJHGStammdaten j ON p.PER_PERSONENNR = j.KundenNr
    LEFT JOIN Staatsbuergerschaften staaten ON (s.StaatsbuergerschaftId IS NOT NULL
AND s.StaatsbuergerschaftId = staaten.ID) OR (j.StaatsbuergerschaftId IS NOT NULL AND
j.StaatsbuergerschaftId = staaten.id)
) sub
```

ANHANG 5: SCRIPT ZUR BELADUNG DER DIMENSIONEN „KUNDE“ UND „KUNDENNUMMER“

```
INSERT INTO DWH.dbo.DimKunde
(IdentifizierKey, Vorname, Nachname, Geschlecht, SVNR, Wohnort, PLZ, Strasse, Bezirk,
Bundesland, Land, Telefonnummer, Mobilnummer, Anrede, [Titel vorne], [Titel hinten],
Staatsbuergerschaft, Geburtsland, Geburtsdatum, GueltigAb, GueltigBis, IstGueltig)
SELECT
IdentifizierKey, Vorname, Nachname, Geschlecht, SVNR, Wohnort, PLZ, Strasse, Bezirk,
Bundesland, Land, Telefonnummer, Mobilnummer, Anrede, [Titel vorne], [Titel hinten],
Staatsbuergerschaft, Geburtsland, Geburtsdatum, GueltigAb, GueltigBis, IstGueltig
FROM
(
MERGE INTO DWH.dbo.DimKunde tgt
USING (
SELECT
*
FROM (
SELECT
ROW_NUMBER() OVER (PARTITION BY Nachname,Vorname ORDER BY Nachname,Vorname) RN,
n.Kunde_FK,
n.IdentifizierKey,
k.*
FROM DW_Stage.Stage.DimKunde k
LEFT JOIN DWH.dbo.DimKundennummer n ON k.Datenquelle = n.Datenquelle AND k.NaturalKey =
n.NaturalKey
) k
WHERE RN = 1
) src
ON src.Kunde_FK = tgt.[Key]
WHEN NOT MATCHED THEN
INSERT (Vorname, Nachname, Geschlecht, SVNR, Wohnort, PLZ, Strasse, Bezirk, Bundesland,
Land, Telefonnummer, Mobilnummer, Anrede, [Titel vorne], [Titel hinten],
Staatsbuergerschaft, Geburtsland, Geburtsdatum, GueltigAb, GueltigBis, IstGueltig)
VALUES
(src.Vorname, src.Nachname,src.Geschlecht,
src.SVNR,src.Wohnort,src.PLZ,src.Strasse,src.Bezirk,src.Bundesland,src.Land,src.Telefonnu
mmer,src.Mobilnummer,src.Anrede,src.[Titel vorne], src.[Titel hinten],
src.Staatsbuergerschaft,src.Geburtsland,src.Geburtsdatum,'1900-01-01',null,1)
WHEN MATCHED AND IstGueltig=1 THEN
UPDATE SET IstGueltig=0, GueltigBis=DATEADD(DAY,-1,GETDATE())
OUTPUT
$action action,
src.IdentifizierKey,
src.Vorname,
src.Nachname,
src.Geschlecht,
src.SVNR,
src.Wohnort,
src.PLZ,
src.Strasse,
src.Bezirk,
src.Bundesland,
src.Land,
src.Telefonnummer,
src.Mobilnummer,
src.Anrede,
src.[Titel vorne],
src.[Titel hinten],
```

```
src.Staatsbuergerschaft,  
src.Geburtsland,  
src.Geburtsdatum,  
getdate() GueltigAb,  
null GueltigBis,  
1 IstGueltig  
) m where action='UPDATE'  
;  
UPDATE DWH.dbo.DimKunde SET IdentifierKey=[Key] WHERE IdentifierKey IS NULL  
  
INSERT INTO DWH.dbo.DimKundennummer (NaturalKey,Kunde_FK,IdentifierKey,Datenquelle)  
SELECT  
sk.NaturalKey, k.[Key], k.IdentifierKey,sk.Datenquelle  
FROM DWH.dbo.DimKunde k  
LEFT JOIN DW_Stage.Stage.DimKunde sk ON k.Nachname=sk.Nachname AND k.Vorname=sk.Vorname  
LEFT JOIN DWH.dbo.DimKundennummer n ON k.[Key] = n.Kunde_FK  
WHERE n.[Key] IS NULL
```

ANHANG 6: SCRIPT ZUR BELADUNG DER FAKTENTABELLE „KUNDENSTAMM“

```

INSERT INTO DWH.dbo.FaktKundenstamm
(Transaktion,Datum_FK,Kunde_FK,Kundennummer_FK,Dienstleistung_FK,Standort_FK,Kostenstelle
_FK,[Alter],Zaehler,Austrittsgrund,Datenquelle)
SELECT
'Eintritt' Transaktion,
CONVERT(INT,FORMAT(bmd.DVH_EINTRITTSDATUM,'yyyyMMdd')) Datum_FK,
k.[Key] Kunde_FK,
n.[key] Kundennummer_FK,
d.[Key] Dienstleistung_FK,
1 Standort_FK,
kst.[Key] Kostenstelle_FK,
ISNULL(DATEDIFF(YEAR,k.Geburtsdatum,DVH_EINTRITTSDATUM) - CASE WHEN
DATEADD(YEAR,DATEDIFF(YEAR,k.Geburtsdatum,DVH_EINTRITTSDATUM),k.Geburtsdatum) >
DVH_EINTRITTSDATUM THEN 1 ELSE 0 END,0) [Alter],
1 Zaehler,
NULL Austrittsgrund,
'BMDAW' Datenquelle
FROM
DVH_DIENSTVERHAELTNIS bmd
JOIN PER_PERSONEN p ON bmd.DVH_MIT_FIRMENNR = p.PER_FIRMENNR AND bmd.DVH_MITARBEITERID =
p.PER_PERSONENID
JOIN PSW_PERSONENWERTE ON DVH_MIT_FIRMENNR = PSW_FIRMENNR AND PSW_DVLFDNR =
DVH_DIENSTVLFDR AND DVH_EINTRITTSDATUM = PSW_PERSWERTGUELTIGAB AND DVH_MITARBEITERID =
PSW_PERSONENID
LEFT JOIN DWH.dbo.DimKundennummer n ON p.PER_PERSONENNR = n.NaturalKey AND 'BMDAW' =
n.datenquelle
LEFT JOIN DWH.dbo.DimKunde k ON n.Kunde_FK = k.[Key]
LEFT JOIN DWH.dbo.DimDienstleistung d ON PSW_FREIFELDN06_9 = d.NaturalKey AND 'BMDAW' =
d.Datenquelle
LEFT JOIN DWH.dbo.DimKostenstelle kst ON PSW_KOSTENSTELLE = kst.KST
WHERE kst.[Key] IS NOT NULL
AND d.[Key] IS NOT NULL
UNION ALL
SELECT
'Austritt' Transaktion,
CONVERT(INT,FORMAT(bmd.DVH_AUSTRITTSDATUM,'yyyyMMdd')) Datum_FK,
k.[Key] Kunde_FK,
n.[key] Kundennummer_FK,
d.[Key] Dienstleistung_FK,
1 Standort_FK,
kst.[Key] Kostenstelle_FK,
ISNULL(DATEDIFF(YEAR,k.Geburtsdatum,DVH_AUSTRITTSDATUM) - CASE WHEN
DATEADD(YEAR,DATEDIFF(YEAR,k.Geburtsdatum,DVH_AUSTRITTSDATUM),k.Geburtsdatum) >
DVH_AUSTRITTSDATUM THEN 1 ELSE 0 END,0) [Alter],
1 Zaehler,
ISNULL(komaz.AustrittGrundValue,wohnen.AustrittsgrundValue) Austrittsgrund,
'BMDAW' Datenquelle
FROM
DVH_DIENSTVERHAELTNIS bmd
JOIN PER_PERSONEN p ON bmd.DVH_MIT_FIRMENNR = p.PER_FIRMENNR AND bmd.DVH_MITARBEITERID =
p.PER_PERSONENID
JOIN PSW_PERSONENWERTE ON DVH_MIT_FIRMENNR = PSW_FIRMENNR AND PSW_DVLFDNR =
DVH_DIENSTVLFDR AND DVH_EINTRITTSDATUM = PSW_PERSWERTGUELTIGAB AND DVH_MITARBEITERID =
PSW_PERSONENID
LEFT JOIN DW_Stage.dbo.BHGKomAZ komaz ON komaz.KundenNr = p.PER_PERSONENNR
LEFT JOIN DW_Stage.dbo.BHGWohnen wohnen ON wohnen.KundenNr = p.PER_PERSONENNR

```

```

LEFT JOIN DWH.dbo.DimKundennummer n ON p.PER_PERSONENNR = n.NaturalKey AND 'BMDAW' =
n.datenquelle
LEFT JOIN DWH.dbo.DimKunde k ON n.Kunde_FK = k.[Key]
LEFT JOIN DWH.dbo.DimDienstleistung d ON PSW_FREIFELDN06_9 = d.NaturalKey AND 'BMDAW' =
d.Datenquelle
LEFT JOIN DWH.dbo.DimKostenstelle kst ON PSW_KOSTENSTELLE = kst.KST
WHERE kst.[Key] IS NOT NULL
AND d.[Key] IS NOT NULL
AND DVH_AUSTRITTSDATUM IS NOT NULL
UNION ALL
SELECT
'Änderung' Transaktion,
CONVERT(INT,FORMAT(PSW_PERSWERTGUELTIGAB,'yyyyMMdd')) Datum_FK,
k.[Key] Kunde_FK,
n.[key] Kundennummer_FK,
d.[Key] Dienstleistung_FK,
1 Standort_FK,
kst.[Key] Kostenstelle_FK,
ISNULL(DATEDIFF(YEAR,k.Geburtsdatum,PSW_PERSWERTGUELTIGAB) - CASE WHEN
DATEADD(YEAR,DATEDIFF(YEAR,k.Geburtsdatum,PSW_PERSWERTGUELTIGAB),k.Geburtsdatum) >
PSW_PERSWERTGUELTIGAB THEN 1 ELSE 0 END,0) [Alter],
1 Zaehler,
NULL Austrittsgrund,
'BMDAW' Datenquelle
FROM
DVH_DIENSTVERHAELTNIS bmd
JOIN PER_PERSONEN p ON bmd.DVH_MIT_FIRMENNR = p.PER_FIRMENNR AND bmd.DVH_MITARBEITERID =
p.PER_PERSONENID
JOIN PSW_PERSONENWERTE ON DVH_MIT_FIRMENNR = PSW_FIRMENNR AND PSW_DVLFDNR =
DVH_DIENSTVLFDR AND DVH_EINTRITTSDATUM < PSW_PERSWERTGUELTIGAB AND DVH_MITARBEITERID =
PSW_PERSONENID
LEFT JOIN DWH.dbo.DimKundennummer n ON p.PER_PERSONENNR = n.NaturalKey AND 'BMDAW' =
n.datenquelle
LEFT JOIN DWH.dbo.DimKunde k ON n.Kunde_FK = k.[Key]
LEFT JOIN DWH.dbo.DimDienstleistung d ON PSW_FREIFELDN06_9 = d.NaturalKey AND 'BMDAW' =
d.Datenquelle
LEFT JOIN DWH.dbo.DimKostenstelle kst ON PSW_KOSTENSTELLE = kst.KST
WHERE kst.[Key] IS NOT NULL
AND d.[Key] IS NOT NULL
AND (DVH_EINTRITTSDATUM <= PSW_PERSWERTGUELTIGABIS OR PSW_PERSWERTGUELTIGABIS IS NULL)
AND DVH_EINTRITTSDATUM <= PSW_PERSWERTGUELTIGAB
UNION ALL
SELECT
'Eintritt' Transaktion,
CONVERT(INT,FORMAT(Einstieg,'yyyyMMdd')) Datum_FK,
k.[Key] Kunde_FK,
n.[key] Kundennummer_FK,
d.[Key] Dienstleistung_FK,
1 Standort_FK,
kst.[Key] Kostenstelle_FK,
ISNULL(DATEDIFF(YEAR,k.Geburtsdatum,Einstieg) - CASE WHEN
DATEADD(YEAR,DATEDIFF(YEAR,k.Geburtsdatum,Einstieg),k.Geburtsdatum) > Einstieg THEN 1
ELSE 0 END,0) [Alter],
1 Zaehler,
NULL Austrittsgrund,
'AABDB' Datenquelle
FROM
TeilnehmerInnen
join JAW_Massnahmen m ON MassnahmeNrFK=MassnahmePK
LEFT JOIN DWH.dbo.DimKundennummer n ON TeilnehmerNr = n.NaturalKey AND 'AABDB' =
n.datenquelle
LEFT JOIN DWH.dbo.DimKunde k ON n.Kunde_FK = k.[Key]

```

```
LEFT JOIN DWH.dbo.DimDienstleistung d ON MassnahmeNrFK = d.NaturalKey AND 'AABDB' =
d.Datenquelle
LEFT JOIN DWH.dbo.DimKostenstelle kst ON m.Kostenstelle = kst.KST
WHERE kst.[Key] IS NOT NULL
AND d.[Key] IS NOT NULL
AND Einstieg IS NOT NULL
UNION ALL
SELECT
'Austritt' Transaktion,
CONVERT(INT,FORMAT(Ausstieg,'yyyyMMdd')) Datum_FK,
k.[Key] Kunde_FK,
n.[key] Kundennummer_FK,
d.[Key] Dienstleistung_FK,
1 Standort_FK,
kst.[Key] Kostenstelle_FK,
ISNULL(DATEDIFF(YEAR,k.Geburtsdatum,Ausstieg) - CASE WHEN
DATEADD(YEAR,DATEDIFF(YEAR,k.Geburtsdatum,Ausstieg),k.Geburtsdatum) > Ausstieg THEN 1
ELSE 0 END,0) [Alter],
1 Zaehler,
'unbekannt' Austrittsgrund,
'AABDB' Datenquelle
FROM
TeilnehmerInnen
join JAW_Massnahmen m ON MassnahmeNrFK=MassnahmePK
LEFT JOIN DWH.dbo.DimKundennummer n ON TeilnehmerNr = n.NaturalKey AND 'AABDB' =
n.datenquelle
LEFT JOIN DWH.dbo.DimKunde k ON n.Kunde_FK = k.[Key]
LEFT JOIN DWH.dbo.DimDienstleistung d ON MassnahmeNrFK = d.NaturalKey AND 'AABDB' =
d.Datenquelle
LEFT JOIN DWH.dbo.DimKostenstelle kst ON m.Kostenstelle = kst.KST
WHERE kst.[Key] IS NOT NULL
AND d.[Key] IS NOT NULL
AND Ausstieg IS NOT NULL
```

ANHANG 7: SCRIPT ZUR BELADUNG DER FAKTENTABELLE „KUNDENSTAMM-SNAPSHOT“

```

;with tmp AS (
    SELECT
        ROW_NUMBER() OVER (PARTITION BY ein.Kundennummer_FK,ein.Datum_FK ORDER BY
ein.Datum_FK,aus.Datum_FK) RN,
        ein.Datum_FK Von,
        aus.Datum_FK Bis,
        ein.Kunde_FK,
        ein.Kundennummer_FK,
        ein.Dienstleistung_FK,
        ein.Standort_FK,
        ein.Kostenstelle_FK,
        k.Geburtsdatum
    FROM DWH.dbo.FaktKundenstamm ein
    LEFT JOIN DWH.dbo.FaktKundenstamm aus ON ein.Kundennummer_FK = aus.Kundennummer_FK
and (aus.transaktion = 'Austritt' OR aus.transaktion='Änderung') AND ein.[key]<>aus.[key]
and ein.datum_fk <= aus.datum_fk
    LEFT JOIN DWH.dbo.DimKunde k ON ein.Kunde_FK = k.[Key]
    WHERE Ein.Transaktion = 'Eintritt' OR Ein.Transaktion='Änderung'
)

INSERT INTO DWH.dbo.FaktKundenstammSnapshot (Datum_FK, Kunde_FK,
Kundennummer_FK,Dienstleistung_FK,Standort_FK,Kostenstelle_FK,Zähler,[Alter])
SELECT
z.[Key] Datum_FK,
t.Kunde_FK,
t.Kundennummer_FK,
t.Dienstleistung_FK,
t.Standort_FK,
t.Kostenstelle_FK,
1 Zähler,
ISNULL(DATEDIFF(YEAR,Geburtsdatum,Datum) - CASE WHEN
DATEADD(YEAR,DATEDIFF(YEAR,Geburtsdatum,Datum),Geburtsdatum) > Datum THEN 1 ELSE 0 END,-
1) [Alter]
FROM (SELECT * FROM DWH.dbo.DimZeit where Jahr >= 2013 AND Datum < (SELECT
LetzterTagImJahr FROM DWH.dbo.DimZeit WHERE Datum =
CONVERT(DATE,DATEADD(YEAR,5,getdate())))) z
JOIN tmp t
ON t.Von <= z.[Key] AND (t.Bis IS NULL OR t.Bis >= z.[Key])
where t.rn = 1

```