

# Bachelor-Thesis

Untersuchung des ETL-Tools Tensei-Data und dessen Einsatz im  
Prozess der Datenmigration bei der SIV.AG

Eingereicht am: 8. März 2016

von: Martin Pohl  
geboren am 13. April 1981  
in Rostock

---

## **Aufgabenstellung**

Einfügen der ausgegebenen Aufgabenstellung der Bachelor-Thesis. Der Titel der Arbeit wird bei deutschsprachigen Titeln in der englischen Fassung wiederholt.

---

## **Zusammenfassung**

Hier sollte auf max. 1/2 bis 3/4 Seiten eine Zusammenfassung erstellt werden.

- Motivation, Einordnung, Umfeld und Abgrenzung der Arbeit
- wesentliche Schwerpunkte und Ergebnisse der Arbeit

## **Abstract**

English Version.

## **Danksagung**

An dieser Stelle möchte ich mich zunächst bei meiner Betreuerin an der Hochschule Wismar Prof. Dr.-Ing Antje Raab-Düsterhöft für ihre Unterstützung bedanken.

Diese Arbeit ist am Standort Roggentin der SIV.AG entstanden. Dafür möchte ich mich besonders bei meinem zweiten Betreuer, Detlef Herold, bedanken. Meinen Dank möchte ich auch allen anderen Mitarbeitern der SIV.AG aussprechen, die stets ein offenes Ohr für meine Fragen hatten.

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>6</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>7</b>
2.1	SIV.AG . . . . .	7
2.2	kVASy® . . . . .	7
2.3	Datenbankgrundlagen . . . . .	8
2.3.1	Elemente einer relationalen Datenbank . . . . .	9
2.3.2	Beziehungen . . . . .	10
2.3.3	Schema und Ausprägung . . . . .	10
2.4	Datenintegration/Datenmigration . . . . .	11
2.5	Transformationen . . . . .	11
2.6	ETL-Prozess . . . . .	11
2.6.1	Extraktion . . . . .	12
2.6.2	Transformation . . . . .	12
2.6.3	Laden . . . . .	12
2.7	Qualitätskontrolle/Protokollierung . . . . .	12
2.8	Kriterien für eine Softwaremigration . . . . .	12
2.8.1	Ziele einer Migration . . . . .	13
2.8.2	Vorgehensweise/Migrationsplanung . . . . .	13
2.8.3	Wirtschaftliche Aspekte . . . . .	13
2.8.4	Qualitative Aspekte . . . . .	13
2.8.5	Organisatorische Aspekte . . . . .	15
2.8.6	Anforderung an die Datenmigration . . . . .	15
<b>3</b>	<b>Ist-Stand-Analyse</b>	<b>16</b>
3.1	Datenmigrationsprozess . . . . .	16
3.2	Transformationen . . . . .	19
3.3	Qualitätskontrolle/Protokollierung . . . . .	19
3.4	Ablaufsteuerung . . . . .	19
<b>4</b>	<b>ETL-Tool Tensei-Data</b>	<b>20</b>
4.1	Allgemeine Grundlagen . . . . .	20
4.2	Grafische Benutzeroberfläche . . . . .	22
4.3	Migrationsprozess . . . . .	22
4.4	Transformationen . . . . .	22
4.5	Qualitätskontrolle/Protokollierung . . . . .	22
4.6	Ablaufsteuerung . . . . .	22
<b>5</b>	<b>Implementierung und Test</b>	<b>23</b>

<b>6</b>	<b>Bewertung, Zusammenfassung und Ausblick</b>	<b>24</b>
6.1	Bewertung . . . . .	24
6.2	Ausblick . . . . .	24
6.3	Zusammenfassung . . . . .	24
	<b>Literaturverzeichnis</b>	<b>25</b>
	<b>Abbildungsverzeichnis</b>	<b>27</b>
	<b>Tabellenverzeichnis</b>	<b>28</b>
	<b>Abkürzungsverzeichnis</b>	<b>29</b>
<b>A</b>	<b>Auflistung von Quellcode und ähnliches</b>	<b>30</b>
	<b>Selbstständigkeitserklärung</b>	<b>35</b>

## 1 Einleitung

max. 2–3 Seiten

- Einführung in die Thematik und das wissenschaftlich-technische Umfeld
- Einordnung in das Wissenschaftsgebiet / tangierte Gebiete
- Analyse der Aufgabenstellung (Problemerkfassung)
- Realisierungsumfeld und Randbedingungen

## 2 Theoretische Grundlagen

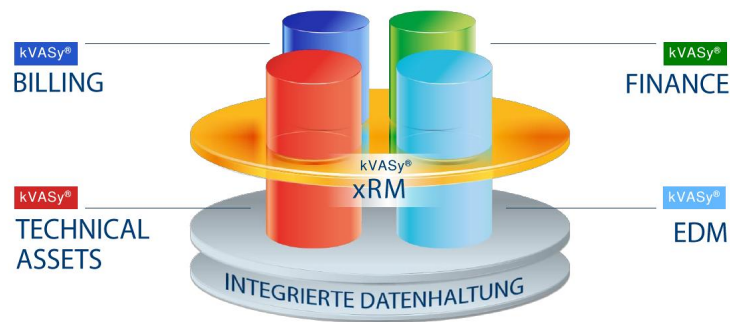
In diesem Kapitel werden zunächst die Grundlagen und Begrifflichkeiten vorgestellt. Weiterhin werden die Kriterien für eine erfolgreiche Migrationen erläutert.

### 2.1 SIV.AG

Die SIV.AG ist ein Anbieter für ganzheitliche Lösungen im Bereich der deutschen und internationalen Energie- und Wasserwirtschaft. Das Kernstück des Unternehmens ist das Softwareprodukt kVASy<sup>®</sup>, ein vollständig integriertes Enterprise Resource Planning (ERP)-System, welches insbesondere auf die Anforderungen und Prozesse der Energie- und Wasserwirtschaft ausgerichtet ist. Das Unternehmen wurde 1990 durch Jörg Sinnig als Software- und Beratungshaus gegründet und im Februar 2016 an die Harris Computer Corporation, einer Tochtergesellschaft der Constellation Software Inc. verkauft. Zum gegenwärtigen Zeitpunkt beschäftigt die SIV.AG mehr als 400 Mitarbeiter. Ihr Leistungsportfolio reicht von der Beratung und Analyse über die Implementierung und Bereitstellung der IT-Systeme bis hin zur Datenmigration, Schulung und Pflege. [SIV13]

### 2.2 kVASy<sup>®</sup>

Das zuvor bereits erwähnte kVASy<sup>®</sup> ist ein webbasiertes ERP-System und bildet das Aushängeschild der SIV.AG. Unter einem ERP-System ist eine integrierte unternehmensweite Anwendung zu verstehen, die zur Koordination wichtiger interner Prozesse eingesetzt wird [CPS10, Seite 482]. Zu der Produktfamilie von kVASy<sup>®</sup> gehören die Module Finance (Buchhaltung), Billing (Abrechnung) und EDM (Vertragsverwaltung), Technical Assets (Anlagenmanagement) und xRM (Kundenbeziehungsmanagement) (vgl. Bild 2.1). Auf der Grundlage einer zentralen Datenbasis erfolgt die Kommunikation der Module und Funktionalitäten schnittstellenfrei [SIV13, Seite15]. Die Umsetzung dessen erfolgt durch das Datenbank-Management-System (DBMS) Database 11g Release 2 von der Firma Oracle. Für die Umsetzung der Applikationslogik wurde Procedural Language/Structured Query Language (PL/SQL) verwendet, eine proprietäre Programmiersprache der Firma Oracle [Wika].



**Bild 2.1:** Produktfamilie von kVASy®[SIV16]

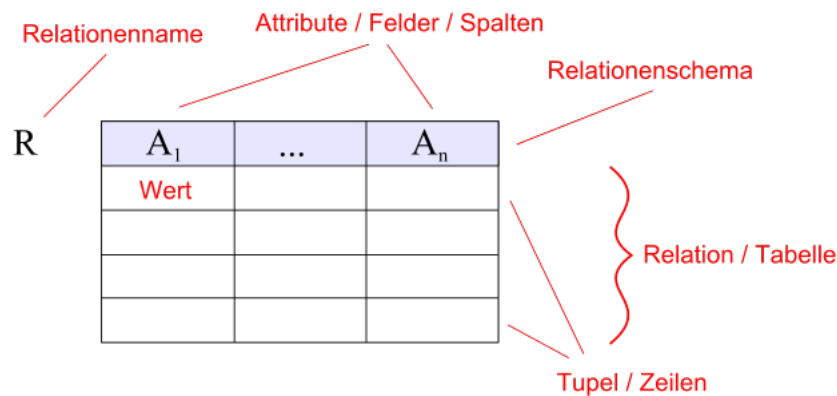
## 2.3 Datenbankgrundlagen

Indirekt wird bei Datenbanken von dem weit verbreiteten Relational Database Management System (RDBMS) ausgegangen. Dieses Datenbankmodell ist eins der Populärsten RDBMS [IT]. Ein Datenbanksystem besteht Zu den Aufgaben eines Datenbanksystems gehört es die Daten möglichst effizient und dauerhaft zu speichern. Zudem soll es eine Vereinfachung der Datenverwaltung ermöglichen. Das relationale Datenmodell ist so konzipiert das es eine mengenorientierte Datenverarbeitung erlaubt. Dabei können die Daten aus mehreren Tabellen durch Join-Verknüpfungen in einer großen Tabelle abgebildet werden. Durch die Festlegung von Spalten und Where-Klauseln kann die Menge weiter bearbeitet und somit in der Höhe und Breite eingeschränkt werden. Im Gegensatz zu einem datensatzorientiertem Vorgehen wird die Tabelle, Zeile um Zeile abgerufen und nach bestimmten Kriterien geprüft. Nachdem die Kriterien erfüllt sind werden die Daten weiterverarbeitet. [KE11][Seite 71] Somit wird die mengenorientierte Datenverarbeitung durch die Datenbanksprache SQL ausgeführt und die datensatzorientierte Verarbeitung beispielsweise durch eine Programmiersprache C++ oder Java. Die relationale Datenbank von Oracle die in der Bachelorarbeit verwendet wird, bietet die Möglichkeit eine Programmiersprache, wie etwa PL/SQL die eigens von Oracle entwickelt worden ist einzusetzen. Durch PL/SQL ist es somit Möglich Java-Programme oder C++-Programme in einem PL/SQL-Programm auszuführen.



### 2.3.1 Elemente einer relationalen Datenbank

Zu den Elementen gehören Tabellen, Spalten und Zeilen (auch Tupel genannt). In der Regel besteht eine Datenbank aus mindestens einer, meistens aber aus mehreren **Tabellen** (Relationen). Die wiederum Attribute oder auch Felder beinhalten, die sogenannten Spalten. Eine **Spalte** besitzt einen Namen sowie einen Datentyp. Zu den drei wesentlichen Datentypen zählen Number (Zahl), Varchar (Text) und Date (Datum), diese enthalten wiederum noch weitere Varianten [KE11][Seite 112]. Die Bedeutung eines Datentyps ist zum einen für ein Datenbanksystem, sowie für den Anwender wichtig. Bei einem Datenbanksystem ist eine Optimierung des physikalischen Datenspeichers möglich. Beim Anwender dient der Datentypen zur Überprüfung der Eingabe, die er getätigt hat. Sollte der vorgegebene Datentyp dem nicht Entsprechen würde es in so einem Fall zu einer Fehlerausgabe führen. Damit kann der Anwender darauf reagieren und die Eingabe korrigieren. Die **Zeilen** (Tupel) bezeichnen die eigentlichen Datensätze. Eine Zeile beinhaltet für jede Spalte einen Wert. Sollte kein Wert angegeben sein wird die Spalte mit dem Wert NULL befüllt. NULL ist somit als Wert zusehen und steht nur für die Abwesenheit eines Wertes. Die Abbildung 2.2 zeigt noch einmal eine Tabelle mit den Begrifflichkeiten die im Bereich der relationalen Datenbanken zum Einsatz kommen.



**Bild 2.2:** Begriffe einer relationaler Datenbanken [Wikb]

Zu den Vorteilen einer relationalen Datenbank gehören neben der Mehrbenutzerfähigkeit und Transaktionen auch Datensicherheit und Datenintegrität. Eine Transaktion bezeichnet eine Folge von Datenverarbeitungsbefehlen (lesen, verändern, einfügen, löschen), welches den Datenbestand in einen konsistenten Zustand überführt. Eine Transaktion sollte daher entweder vollständig und fehlerfrei oder gar nicht ausgeführt werden. [KE11][Seite 285] Zur Datensicherheit zählt sowohl der Schutz vor

Datenverlust, als auch der unerlaubte Zugriff durch Dritte. Die Datenintegrität wird durch sogenannte Constraints erreicht. Sie definieren Bedingungen, die beim Einfügen, Ändern und Löschen von Datensätzen in der Datenbank erfüllt werden müssen. [KE11][Seite 159-161]

### 2.3.2 Beziehungen

In einem relationalen Datenbanksystem werden die Beziehungen zwischen Daten, mit Hilfe von Schlüsselspalten realisiert. Diese Schlüsselspalten werden durch Constraints gebildet. Dazu gehören PRIMARY KEY, FOREIGN KEY und UNIQUE [KE11][Seite 161]. Der PRIMARY KEY und der FOREIGN KEY sollen etwas näher erläutert werden. Der PRIMARY KEY bestimmt eine oder mehrere Spalten als Primärschlüssel. Die Werte dieser Schlüsselspalten sind immer eindeutig und dürfen nicht NULL sein. Zudem ist eine Primärschlüssel nur einmal für eine Tabelle definierbar. Der FOREIGN KEY dient als Verknüpfung zwischen Tabellen. Dafür wird der Primärschlüssel aus der ersten Tabelle in die zweite Tabelle als Fremdschlüssel eingetragen. Der Fremdschlüssel enthält somit den gleichen Wert wie der Primärschlüssel. Des Weiteren kann der Fremdschlüssel in einer Tabelle einmal, keinmal oder mehrmals vorkommen.

### 2.3.3 Schema und Ausprägung

Es gibt gewisse Unterschiede zwischen einem Datenbankschema und einer Datenbankausprägung. Das Datenbankschema beschreibt nicht die einzelnen Datenobjekte, sondern es legt die Struktur der abspeicherbaren Datenobjekte fest. Somit handelt es sich bei einem Datenbankschema um Daten, die andere Daten beschreiben diese können auch als Metadaten aufgefasst werden. Im Gegensatz dazu ist bei der Datenbankausprägung, die einem ständigen Wandel unterliegt, der gegenwärtig gespeicherte Zustand zu verstehen. Somit bildet das Datenbankschema die Basis der Datenbankausprägung. Die beiden Begriffe werden im Zusammenhang mit einer Datenbank auch als intensionalen (Schema) und extensionalen (Ausprägung) Ebene bezeichnet. Die Änderung eines Datenbankschemas wird normalerweise sehr selten vorgenommen, wenn dies doch passiert spricht man von einer *Schemaevolution*. Zu Beachten ist dabei, dass eine Schemaänderung schwerwiegenden Folgen haben kann. Grund dafür ist das Aufweisen einer inkonsistenten Struktur der bereits abgespeicherten Datenobjekte. [KE11][Seite 24]

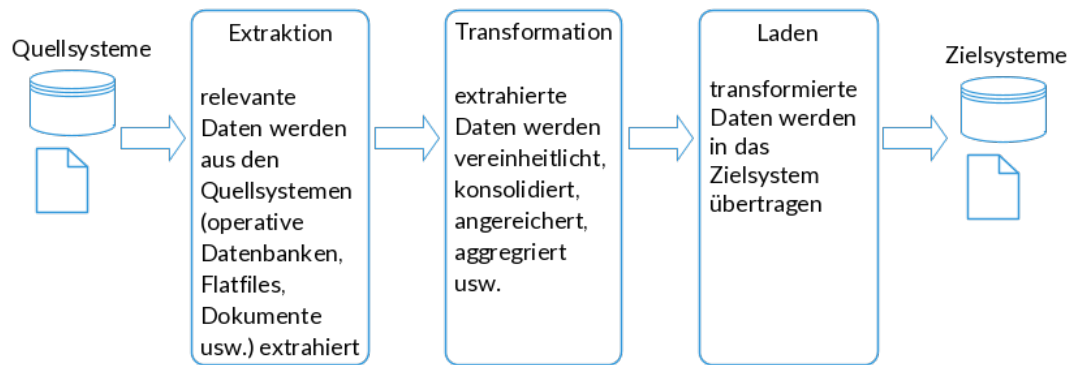
## 2.4 Datenintegration/Datenmigration

Der Begriff Migration stammt von dem lateinischen Begriff migratio und bedeutet (Aus)wanderung, d. h. die Übernahme von Daten in ein anderes System, um sie dort wieder im vollem Ausmaß zu nutzen [Dud]. Die Datenmigration oder auch Datenübernahmen genannt, verläuft von einem Quellsystem hin zu einem Zielsystem. Dabei beinhaltet das Quellsystem, die zu migrierenden Daten und das Zielsystem die importierten Daten. In dieser Arbeit wird das in Kapitel 2.2 auf Seite 7 beschriebene kVASy<sup>®</sup>, als Zielsystem verwendet.

## 2.5 Transformationen

## 2.6 ETL-Prozess

Die Abkürzung ETL setzt sich aus den drei Wörtern Extraktion, Transformation und Laden zusammen. Sie wird oft mit dem Begriff Data Warehouse in Zusammenhang gebracht. Dabei dient ETL dazu, unterschiedliche Bereiche Buchhaltung, Abrechnung, Vertragsverwaltung usw. in eine zentrale Datenbank (Data Warehouse) zusammenzuführen. Ziel ist es z. B. eine umfassende Sicht auf den Kunden zu bekommen. Aufbauend darauf können komplexe Analysen entwickelt werden, um eine erfolgreiche und wohl überlegte Geschäftsentscheidung zu treffen. Bei der Abfolge eines ETL-Prozesses werden zunächst Daten aus einer Datenquelle extrahiert. Daraufhin werden diese mit Hilfe von Transformationsregeln homogenisiert. Schließlich werden im letzten Schritt die angereicherten oder bereinigten Daten in das Ziel geladen. Die Bandbreite für den Einsatz eines ETL-Prozesses erstreckt sich über das Datenqualitätsmanagement sowie der Replikation und Synchronisation von Daten bis hin zu der Migration von Daten aus Altsystemen. Letzteres ist für die Bachelorarbeit der wichtigste Aspekt und wird dahin gehend genauer untersucht. [Ros13]



**Bild 2.3:** Allgemeiner ETL-Prozess (eig. Abb. nach [Ros13][Seite 38])

### 2.6.1 Extraktion

### 2.6.2 Transformation

### 2.6.3 Laden

## 2.7 Qualitätskontrolle/Protokollierung

## 2.8 Kriterien für eine Softwaremigration

Ziel dieser Bachelor-Thesis ist es den Einsatz des ETL-Tools im Prozess der Datenmigration zu untersuchen. Für eine Softwaremigration ist eine Vielzahl an Kriterien zu beachten. Diese allgemeinen Kriterien werden als Leitfaden für die Migration von Software durch die Beauftragte der Bundesregierung für Informationstechnik bereitgestellt. Im Migrationsleitfaden sind die Kriterien für eine erfolgreiche Migration wie folgt unterteilt [Inf12]:

- Ziele einer Migration
- Vorgehensweise/Migrationsplanung
- Strategische Aspekte
- Rechtliche Aspekte
- Wirtschaftliche Aspekte
- Qualitative Aspekte
- Aspekte des Systembetriebs
- Organisatorische Aspekte

- Sicherheitsaspekte

In den nachfolgenden Abschnitten werden die relevanten Aspekte für diese Arbeit näher erläutert.

### **2.8.1 Ziele einer Migration**

Bevor eine Softwaremigration umgesetzt wird, ist es zunächst wichtig die Ziele zu klären. Diese entstehen aus erkannten Schwachstellen oder einer notwendigen Erweiterung der Bestandssoftware. Die häufigsten Ziele sind beispielsweise [Inf12][Seite 21]:

- ein verbesserter Anwendernutzen
- das Herstellen eines rechtlich notwendigen Zustands
- die Behebung von Fehlern
- die Erweiterung des Funktionsumfangs
- eine verbesserte Integration in die vorhandenen Softwaresysteme
- die Erhöhung der Produktivität
- die bessere Nutzung vorhandener Ressourcen

Diese Migrationsziele können gewichtet und priorisiert werden. Damit bilden sie eine Grundlage für die Auswahl der Software. Des Weiteren kann auf diese Weise überprüft werden, ob die zugeordneten Ziele bei einer Softwaremigration erfüllt sind.

### **2.8.2 Vorgehensweise/Migrationsplanung**

### **2.8.3 Wirtschaftliche Aspekte**

### **2.8.4 Qualitative Aspekte**

In der Regel ist eine Untersuchung der qualitativen Aspekte bei einer Migration auf etablierte Industriestandards nicht erforderlich. Bei solchen Standards kann eine hinreichende Softwarequalität angenommen werden hinsichtlich einer hohen Akzeptanz und Verbreitung. Bei nicht etablierten Standards wird die ISO-Norm ISO/IEC 25010 für eine Reihe von Qualitätsmerkmalen verwendet [ISO11]. Die ISO-Norm unterteilt die qualitativen Aspekte in die folgenden Kriterien [Inf12][Seite 34-36]:

- Funktionale Eignung

- Leistungsfähigkeit
- Kompatibilität
- Benutzbarkeit
- Zuverlässigkeit
- Sicherheit
- Wartbarkeit
- Portabilität

Zu berücksichtigen sind auch die nachfolgenden drei Kriterien die nicht in der ISO-Norm enthalten sind [Inf12][Seite 36-37]:

- Dokumentationsgüte
- Konfigurierbarkeit
- Sonstige Qualitätsmerkmale

In der Bachelorarbeit sollen die nachfolgenden Punkte verwendet werden um den qualitativen Aspekt beurteilen zu können.

### **Funktionale Eignung**

Bei diesem Aspekt wird die Funktionalität einer Software auf Vollständigkeit und Korrektheit geprüft. Die Funktionale Eignung ist bestanden, wenn die Funktionen der Software fehlerfrei und ohne unnötigen Zwischenschritte abläuft. [Inf12][Seite 34]

### **Leistungsfähigkeit**

Die Anforderung an die Leistungsfähigkeit einer Software besteht darin, diese nach bestimmten Leistungsparametern (Minima und/oder Maxima) zu überprüfen. Beispiele für eine Analyse wären im Bereich der Antwortzeiten, Prozessorauslastung, Anzahl paralleler Benutzer usw. zu untersuchen. Zur Unterstützung gibt es entsprechende Analysewerkzeuge. Des Weiteren sollte das zu testende Systems unter Vollast betrieben werden um ein aussagekräftiges Ergebnis zu erhalten. [Inf12][Seite 34]

## Kompatibilität

Bei der Kompatibilität (Compatibility) wird überprüft, ob eine parallele Ausführung der zu untersuchenden Software mit bereits installierten Programmen auf dem System möglich ist. Wenn bei den verwendeten Programmen keine Kompatibilität besteht kann es zu einer fehlerhaften Ausführung kommen, sowie eine Verschlechterung der Leistungsfähigkeit bei einer zeitgleichen Nutzung von Hardware und Software. [Inf12][Seite 35]

## Benutzbarkeit

Die Beurteilung der Benutzbarkeit (Usability) einer Software lässt sich nicht objektiv untersuchen, um dennoch eine Benutzbarkeit zu untersuchen werden die folgenden Punkte herangezogen. [Inf12][Seite 35]

- Verwendung der Software und bereitgestellten Funktionalitäten lassen sich leicht erkennen
- eine intuitive Bedienung ohne großen Aufwand auch für einen fachfremden Anwender
- Unterstützung der Software um Fehler durch den Benutzer zu vermeiden, beispielsweise durch eine Vorgabe der Pflichtfelder oder Wertebereiche
- die Benutzerschnittstelle der Software vermittelt einen ästhetischen Design (Farben, Schriftarten, Positionierung von Feldern usw.)
- Bedienung durch Behinderte Menschen

## Wartbarkeit

Unter Wartbarkeit (Maintainability) ist der Aufwand von Korrektur und Erweiterung der Software zu verstehen. Dabei ist zu analysieren ob die Software Modular aufgebaut ist damit einzelne Module separat angepasst werden können. Des Weiteren zeichnet sich eine gute Wartbarkeit dadurch aus, dass die Ursache eines Fehlers sowie die Auswirkung einer Änderung leicht ermittelt werden kann. [Inf12][Seite 35-36]

### 2.8.5 Organisatorische Aspekte

### 2.8.6 Anforderung an die Datenmigration

### 3 Ist-Stand-Analyse

In diesem Kapitel soll der Datenübernahmeprozess der SIV.AG analysiert werden. Um einen Einblick über den Ablauf zu erhalten wurden mehrere Bezugsquellen herangezogen. Zu den zählten die interne Dokumentation „Handbuch Datenmigration“ (nur für den internen Gebrauch), Beispieldaten und die dazu gehörigen SQL-Skripte aus vergangenen Datenmigrationsprojekten sowie Befragungen der Mitarbeiter aus der Abteilung Datenmigration.

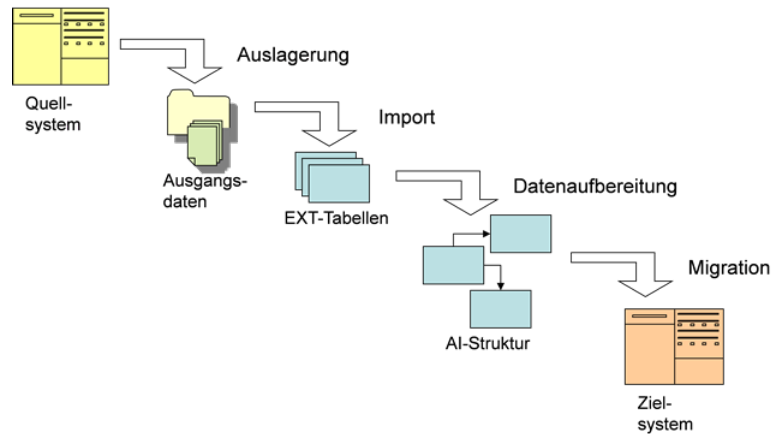
#### 3.1 Datenmigrationsprozess

Gründe für eine Datenmigration durch SIV.AG können vielfältig sein. Zum einen kann der Kunde ein eigenentwickeltes ERP-System haben, welches nicht mehr den Anforderungen entspricht, bezogen auf gesetzliche oder organisatorische Aufgaben. Auf der anderen Seiten existiert noch kein ERP-System und es wird eins benötigt um diese Aufgaben umzusetzen. Des Weiteren könnte ein Kunde eine Standardsoftware z. B. von SAP, Schleupen oder Wilken besitzen und diese ablösen. Gründe dafür wären beispielsweise Kosten zu sparen durch ein preiswertes ERP-System oder eine Fusion zweier Unternehmen mit Unterschiedlichen Systemen von dem nur noch kVASy<sup>®</sup> verwendet wird. Gleichzeitig zur Datenmigration erfolgt eine Einführung in die kVASy<sup>®</sup>-Software. Da kVASy<sup>®</sup> eine Standardsoftware ist wird eine Anpassung der Kundenwünsche nur in bestimmten Fälle vorgenommen. Im Normalfall wird das Zielsystem kVASy<sup>®</sup> nicht an den Daten des Kunden ausgerichtet, sondern es werden die Daten so transformiert das sie im Zielsystem verwendet werden können. In der Abbildung 3.1 wird ein allgemeiner Ablauf der Datenmigration durch die SIV.AG dargestellt.

Anhand der folgenden Teilschritte soll der Ablauf der Datenmigration, wie er durch die SIV.AG umgesetzt wird erläutert werden [SIV11]:

- Auslieferung und Laden der Testdaten
- Datenanalyse auf Basis von Testdaten
- Erstellung individueller Spezifikationen





**Bild 3.1:** Migrationsverfahren durch die SIV.AG

- Definition der Übernahmeroutinen
- Durchführung der Testdatenkonvertierung
- Kontrolle der Testdaten durch Funktionstests
- Spezifikation der Echtdatenmigration
- Durchführung der Echtdatenübernahme
- Kontrolle und Bearbeitung der Echtdaten
- Abnahme des Konvertierungsverfahren

Wie in Kapitel 444 beschrieben wurde, gibt es für die Umsetzung einer Datenmigration mehrere Ansätze. Um dies zu bewerkstelligen verfolgt die SIV.AG den Big-Bang-Ansatz.

### Auslieferung und Laden der Testdaten

Bevor es zu einer Datenanalyse kommt müssen die Testdaten des Kunden vorliegen. Die Lieferung der Testdaten, nach Möglichkeit die gesamten Nutzdaten, erfolgt entweder durch den Kunden selber oder ein externes Unternehmen wird durch ihn beauftragt. In besonderen Fällen bietet die SIV.AG seine Hilfestellung in Form einer Beratung oder einer technischen Unterstützung an. Die Originaldaten müssen in einer elektronischen Form vorliegen wie z. B. Excel-Tabellen, Listen, Comma-separated values (CSV)-Dateien oder Datenbankexports, obendrein muss eine Beschreibung der Daten mitgeliefert werden. Ausschlaggebend ist das die Daten in ein CSV-Format konvertierbar sind, um sie in eine Oracle-Datenbank zu importieren.

Ausschlaggebend für den Einsatz des CSV-Formats sind deren Eigenschaften zu den zählen unter anderem:

- in der ersten Zeile der CSV-Datei stehen die Spaltennamen
- genau ein Datensatz pro Zeile
- die Daten sind durch ein Semikolon im Datensatz getrennt

Für das Einlesen der Test- und Echtdatenübernahme können folgende Datenträger verwendet werden CD-ROM, DVD, USB-Festplatte, FTP, HTTPS und SSL. Die Datenübernahme in die Oracle-Datenbank wird durch den SQL\*Loader realisiert. Der SQL\*Loaders benötigt für die Ausführung 2 Dateien. Zum einen die Datei mit den Originaldaten sowie einer Steuerdatei, welche den Aufbau und die Besonderheiten von Datendateien beschreibt. Die Steuerdatei kann eine einfache Textdatei sein. In dem Fall der Datenübernahme durch die SIV.AG wird als Steuerdatei eine Datei mit der Endung .td verwendet. Im Anhang A.1 ist ein Beispiel einer solchen table description (TD) Datei zusehen. Zusätzlich erzeugt der SQL\*Loader eine Bad-Datei sowie eine Log-Datei. Die Bad-Datei ist eine Fehler-Datei, in der eine Auflistung der erzeugten Fehler dargestellt werden. Ein Fehler könnte beispielsweise sein, dass der Datentyp nicht korrekt ist, mit dem Feld des zu übernehmenden Datensatzes. Die Log-Datei (auch Ergebnisprotokolldatei genannt) enthält das Protokoll über sämtliche Aktionen die bei der Datenübernahme durchgeführt worden sind z. B. Anzahl der übernommen Datensätze. Für die Übersichtlichkeit werden die Originaltabellen nicht eins zu eins übernommen, sondern es werden die Originaltabellennamen mit dem Präfix **EXT\_** versehen, bevor sie in die Datenbank geschoben werden. Ein wichtiger Aspekt dabei ist die Überprüfung der Originaldaten der Kunden und der Aufbereiteten Originaldaten in Form des **EXT\_** Präfixes, da dieses nicht automatisch überwacht wird. Diese Sache ist genau zu kontrollieren, damit keine Daten verloren oder geändert werden. Bekannte Probleme sind Umlautkonvertierung, Fehlen des letzten Datensatzes oder Abschneiden von Werten. Deshalb muss die Anzahl der Aufbereiteten Originaldaten in jedem Fall mit den Originaldaten korrespondieren. Zu dem müssen noch gewisse Voraussetzungen gegeben sein. Zum einen das Aufsetzen eines Testsystems durch den Kunden, welches für den Zeitraum der Migration nicht veränderbar ist. Die zweite Voraussetzung ist ein Gastzugang mit Leserechten auf das Testsystem.

## **Datenanalyse auf Basis von Testdaten**

Nach dem der komplette Dump (Speicherauszug) vorliegt kann die Analyse durch einen Consultant (Berater) durchgeführt werden. Die Seite des Zielsystems braucht nicht analysiert werden, da es bereits bekannt ist. Es sei denn es sollte ein Patch oder Update von kVASy<sup>®</sup> erfolgen, in dem z. B. ein neues Pflichtfeld eingeführt oder gelöscht wird. Die Herausforderung bei einer Datenmigration ist die Datenstruktur des Quellsystems herauszufinden, bzw. wie sie untereinander zusammenhängt. Ziel- und Quellsystem bilden zwar die gleichen Bereiche in der Energie- und Wasserwirtschaft ab, dennoch sind sie divergent. Die beiden System besitzen Felder mit Namen, Zähler, Adressen, Kundennummer, etc., aber trotzdem sind sie nicht homogen, weil beide System unabhängig von einander entwickelt worden sind. Um dem Consultant seine Arbeit zu erleichtern, wird zu den Originaldaten eine Dokumentation verlangt, um einen detaillierten Zusammenhang der Originaldaten zu bekommen. Die Übergabe eine Dokumentation ist, aber nicht als Pflicht anzusehen. Gründe dafür wären beispielsweise, dass der Kunde gar keinen Einblick in seine Daten hat, weil er sich mit ihnen nicht ausführlich auseinander gesetzt hat oder die Daten werden durch ein externes Unternehmen verwaltet. Der Grund bei einem externen Unternehmen ist auf seine Geheimhaltung der Datenbankstruktur zurückzuführen. In solchen Fällen wird eine Analyse um einiges erschwert und es kommt unweigerlich zu einem Mehraufwand.

### **3.2 Transformationen**

### **3.3 Qualitätskontrolle/Protokollierung**

### **3.4 Ablaufsteuerung**

## 4 ETL-Tool Tensei-Data

Das in dieser Bachelorarbeit zu untersuchende **ETL-Tool Tensei-Data** ist von dem Unternehmen Wegtam UG entwickelt worden. Im Jahr 2012 wurde das Unternehmen durch Andre Schütz, Jens Jahnke und Frank Thiessenhusen gegründet, mit seiner Niederlassung in Bentwisch bei Rostock. Zu den Hauptaufgaben von Wegtam UG gehören die Software und Produktentwicklung, sowie die IT Technologieberatung, IT Projektentwicklung und IT Online Vermarktung. Der Firmenname entstammt aus der altnordischen Mythologie und wird dort Wörtlich übersetzt als „Der Wanderer“. Neben dem besagtem Tensei-Data führt das Unternehmen noch den Wegtam Suchagent. Für private Anwender dient es, als eine freie und anonyme Metasuchmaschine. Im Bereich der Firmenkunden ermöglicht die Wegtam Suchmaschinentechologie eine Suche über verteilte Datenquellen. [Wegd]

### 4.1 Allgemeine Grundlagen

Das Produkt Tensei-Data ist in zwei verschiedenen **Lizenzmodellen** erhältlich. Das erste Lizenzmodell ist die gemeinsam genutzte Lizenz. Sie erstreckt sich von der kostenlosen, die 14 Tage genutzt werden kann, über die Small Business, Professional und Enterprise Version mit einem festgelegten Zeitraum von einem Jahr. Die zweite Variante ist eine individuelle Lizenz, mit ihr ist es möglich Anpassung auf die jeweiligen Bedürfnisse zu gewährleisten. Angenommen die nötige Lizenz wurde erworben, erhält man die aktuellste Software-Version von Tensei-Data 1.7. [Wegb] Die nachstehende Tabelle **4.1** enthält eine Beschreibung der Komponenten die in der Lizenz für das Tool Tensei-Data von Bedeutung sind. Der Preis und Umfang einer Lizenz richtet sich, dann nach der Anzahl der eingesetzten Komponenten.

**Tabelle 4.1:** Lizenzinhalt

Begriff	Informelle Bedeutung
Nutzer	Jemand der sich am System anmeldet und Transformationen bearbeitet und starten kann
Agent	Führt eine Transformation durch des Weiteren begrenzt die Anzahl der Agenten die Anzahl der gleichzeitig durchführbaren Transformationen
Transformationskonfiguration	Ist eine Beschreibung des Arbeitsablaufs für einen Agenten (Datenquellen, Rezepte, Datenziele, etc.)
Cronjob	Führt zu einem gewünschten Zeitpunkt eine Transformationskonfiguration aus
Trigger	Mit Hilfe eines Signals stößt er eine Transformationskonfiguration an
Updates	Beinhaltet grundsätzlich alle Updates und Upgrades innerhalb der Laufzeit
Rezepte	Beschreibt eine Transformation zwischen verschiedenen Formaten via Data Format and Semantics Description Language (DFASDL)

Generell ergeben sich in allen Bereichen **Anwendungsfälle**. Die Lösungen von Tensei-Data finden ihren Einsatz bei Kleinstproblemen bis hin zu Datenbanken oder Data-Warehouse Szenarien. Daher ist es wirtschaftlich vor allem für kleine und mittlere Unternehmen interessant. [Wega] Einige Beispiele für den Tensei-Data Einsatz sind:

- Anbinden von Kundensystemen bzw. Integration von Kundendaten in das eigene Zielsystem
- Webseitenumzug von Content-Management-System (CMS)
- Zusammenführung von verschiedensten Datenquellen in firmeneigene Data-Warehouse-Systeme
- Ablösung, Wechsel oder Upgrade von Softwaresystemen (z.B. Buchhaltung, Lager, Customer-Relationship-Management (CRM), ERP, Email-Clients etc.)
- Anpassung der Daten während der Integration/Migration zur Säuberung von z.B. Kundendaten

- Extrahieren von Daten aus einer Datenbank in eine CSV Datei, Laden von Daten aus Dateien in eine Datenbank, etc.

Mit der Anschaffung von einer Software ist es meist nicht getan. Entscheidend ist die sogenannte Pflege der Software, welches durch das zentrale Element dem **Support** realisiert wird. Im Folgenden werden einige Beispiele aufgeführt, die Wegtam als Supportleistung für Tensei-Data bereitstellt. [Wegc]

- Software-Einführung von Tensei-Data für die Unterstützung des intuitiven Frontend und der umfangreichen Dokumentation
- Individuelle und anwendungsbezogene Trainings für die Mitarbeiter um mit dem System effektiv umzugehen
- Die Konfiguration und komplette Einrichtung des Systems, sowie eine individuelle Anpassung (z.B. spezifischer Transformatoren)

## 4.2 Grafische Benutzeroberfläche

Die Bedienung des Tensei-Data Systems erfolgt über eine grafische Benutzeroberfläche (Tensei Server Gui), die mit Hilfe eines Browsers realisiert wird. Als Startseite für den Überblick dient das Dashboard (vgl. Bild A.2). Es zeigt die konfigurierten Transformationskonfigurationen und die verfügbaren Agenten. Darüber erfolgt eine kontinuierliche Aktualisierung der automatisch oder manuell ausgeführten Transformationskonfigurationen. Im oberen Bereich des Fensters befindet sich ein Dropdown-Menü, mit den Punkten Resources, Cookbooks, Service und Statistics. Die wiederum enthalten bis auf Cookbooks noch Unterpunkte.

## 4.3 Migrationsprozess

## 4.4 Transformationen

## 4.5 Qualitätskontrolle/Protokollierung

## 4.6 Ablaufsteuerung

## 5 Implementierung und Test

## **6 Bewertung, Zusammenfassung und Ausblick**

### **6.1 Bewertung**

### **6.2 Ausblick**

### **6.3 Zusammenfassung**



## Literaturverzeichnis

- [CPS10] C.LAUDON, Kenneth ; P.LAUDON, Jane ; SCHODER, Detlef: *Wirtschaftsinformatik*. Pearson Studium, 2010
- [Dud] DUDEN: *Migration*. – <https://de.wikipedia.org/wiki/Migration> Abrufdatum: 16.02.2016
- [Inf12] INFORMATIONSTECHNIK, Die Beauftragte der Bundesregierung für: *Migrationsleitfaden - Leitfaden für die Migration von Software*. [http://www.cio.bund.de/SharedDocs/Publikationen/DE/Architekturen-und-Standards/migrationsleitfaden\\_4\\_0\\_download.pdf?\\_\\_blob=publicationFile](http://www.cio.bund.de/SharedDocs/Publikationen/DE/Architekturen-und-Standards/migrationsleitfaden_4_0_download.pdf?__blob=publicationFile). [http://www.cio.bund.de/SharedDocs/Publikationen/DE/Architekturen-und-Standards/migrationsleitfaden\\_4\\_0\\_download.pdf?\\_\\_blob=publicationFile](http://www.cio.bund.de/SharedDocs/Publikationen/DE/Architekturen-und-Standards/migrationsleitfaden_4_0_download.pdf?__blob=publicationFile). Version: 2012. – Abrufdatum: 08.03.2016
- [ISO11] ISO: *ISO/IEC 25010:2011*. [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=35733](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35733). [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=35733](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=35733). Version: 2011. – Abrufdatum: 07.03.2016
- [IT] IT solid: *DB-Engines*. – <http://db-engines.com/de/ranking> Abrufdatum: 17.02.2016
- [KE11] KEMPER, Alfons ; EICKLER, André: *Datenbanksysteme Eine Einführung*. Oldenbourg Verlag München, 2011
- [Ros13] ROSSAK, Ines: *Datenintegration*. Carl Hanser Verlag, 2013
- [SIV11] SIV.AG: *SIV.AG Lösungsanbieter für die Energie- und Wasserwirtschaft*. <https://www.yumpu.com/de/document/view/5767240/losungsanbieter-fur-die-energie-und-wasserwirtschaft-sivag>, 2011. – Abrufdatum: 01.03.2016
- [SIV13] SIV.AG: *SIV.AG Imagebroschüre*. [https://www.siv.de/tl\\_files/SIV/downloads/broschueren/SIV.AG\\_Imagebroschuere.pdf](https://www.siv.de/tl_files/SIV/downloads/broschueren/SIV.AG_Imagebroschuere.pdf).

- [https://www.siv.de/tl\\_files/SIV/downloads/broschueren/SIV.AG\\_Imagebroschuere.pdf](https://www.siv.de/tl_files/SIV/downloads/broschueren/SIV.AG_Imagebroschuere.pdf). Version: 2013. – Abrufdatum: 01.02.2016
- [SIV16] SIV.AG: *SIV.AG Produktportfolio*. [http://www.siv.de/tl\\_files/SIV/downloads/broschueren/SIV.AG\\_Produktportfolio.pdf](http://www.siv.de/tl_files/SIV/downloads/broschueren/SIV.AG_Produktportfolio.pdf). [http://www.siv.de/tl\\_files/SIV/downloads/broschueren/SIV.AG\\_Produktportfolio.pdf](http://www.siv.de/tl_files/SIV/downloads/broschueren/SIV.AG_Produktportfolio.pdf). Version: 2016. – Abrufdatum: 21.02.2016
- [Wega] WEGTAM: *Anwendungsfälle*. – <http://www.tensei-data.com/de/use-cases> Abrufdatum: 23.02.2016
- [Wegb] WEGTAM: *Lizenzen*. – <http://www.tensei-data.com/de/licenses> Abrufdatum: 23.02.2016
- [Wegc] WEGTAM: *Support*. – <http://www.tensei-data.com/de/support> Abrufdatum: 23.02.2016
- [Wegd] WEGTAM: *Unternehmensprofil*. – <http://donar.messe.de/exhibitor/cebit/2014/T816887/unternehmensprofil-ger-277823.pdf> Abrufdatum: 23.02.2016
- [Wika] WIKIPEDIA: *PL/SQL* — *Wikipedia, Die freie Enzyklopädie*. – <https://de.wikipedia.org/wiki/PL/SQL> Abrufdatum: 09.02.2016
- [Wikb] WIKIPEDIA: *Relationale Datenbank* — *Wikipedia, Die freie Enzyklopädie*. – [https://de.wikipedia.org/wiki/Relationale\\_Datenbank](https://de.wikipedia.org/wiki/Relationale_Datenbank) Abrufdatum: 21.02.2016

## Abbildungsverzeichnis

2.1	Produktfamilie von kVASy®[SIV16] . . . . .	8
2.2	Begriffe einer relationaler Datenbanken [Wikb] . . . . .	9
2.3	Allgemeiner ETL-Prozess (eig. Abb. nach [Ros13][Seite 38]) . . . . .	12
3.1	Migrationsverfahren durch die SIV.AG . . . . .	17
A.1	dt-Datei Beispiel (eigener Screenshot) . . . . .	30
A.2	Dashboard Tensei-Data (eigener Screenshot) . . . . .	31
A.3	DFASDL Tensei-Data (eigener Screenshot) . . . . .	32
A.4	Cronjob Tensei-Data (eigener Screenshot) . . . . .	33
A.5	Trigger Tensei-Data (eigener Screenshot) . . . . .	34

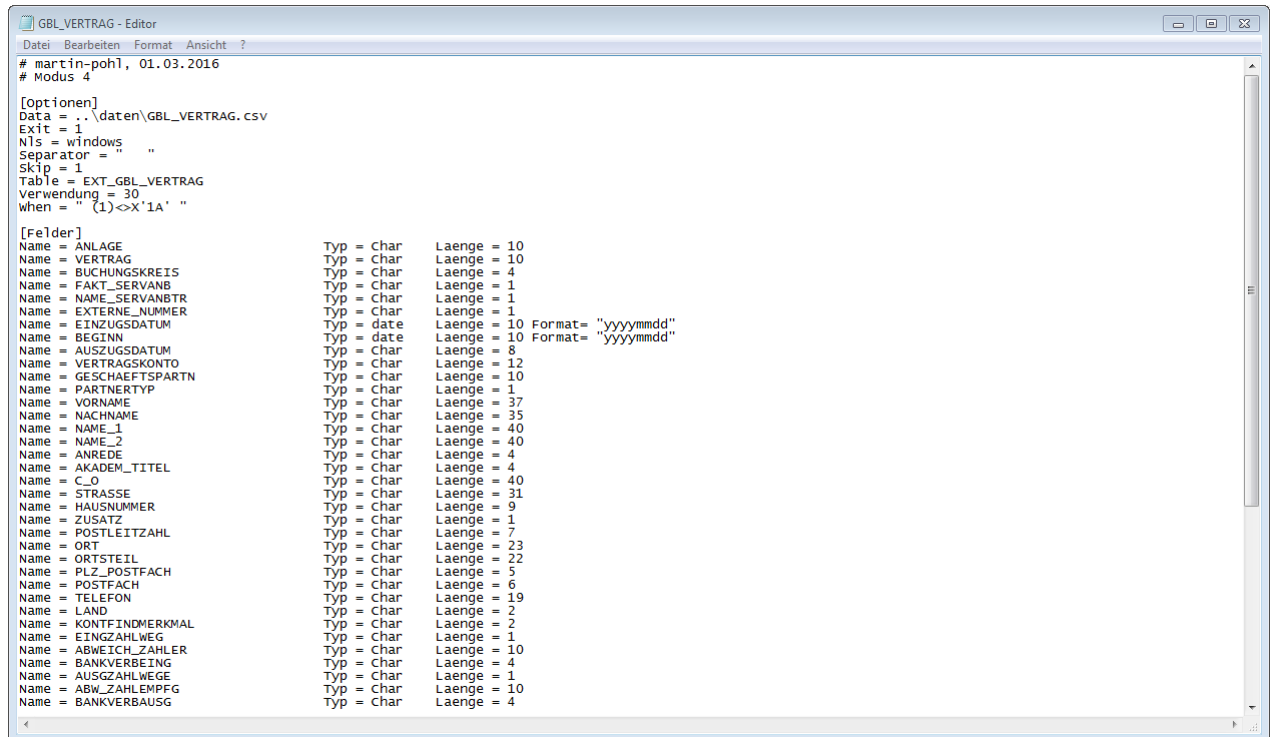
## Tabellenverzeichnis

4.1	Lizenzzinhalt . . . . .	21
-----	-------------------------	----

## Abkürzungsverzeichnis

DBMS	Datenbank-Management-System
ERP	Enterprise Resource Planning
ETL	Extraction, Transformation, Loading
PL/SQL	Procedural Language/Structured Query Language
RDBMS	Relational Database Management System

## A Auflistung von Quellcode und ähnliches



```

GBL_VERTRAG - Editor
Datei Bearbeiten Format Ansicht ?
# martin-pohl, 01.03.2016
# Modus 4

[Optionen]
Data = ..\daten\GBL_VERTRAG.csv
Exit = 1
Nls = windows
Separator = " "
Skip = 1
Table = EXT_GBL_VERTRAG
Verwendung = 30
When = "(1)<X'1A' "

[Felder]
Name = ANLAGE                Typ = Char    Laenge = 10
Name = VERTRAG               Typ = Char    Laenge = 10
Name = BUCHUNGSKREIS         Typ = Char    Laenge = 4
Name = FAKT_SERVANB          Typ = Char    Laenge = 1
Name = NAME_SERVANBTR        Typ = Char    Laenge = 1
Name = EXTERNE_NUMMER        Typ = Char    Laenge = 1
Name = EINZUGSDATUM          Typ = date    Laenge = 10 Format= "yyyymmdd"
Name = BEGINN                Typ = date    Laenge = 10 Format= "yyyymmdd"
Name = AUSZUGSDATUM          Typ = Char    Laenge = 8
Name = VERTRAGSKONTO         Typ = Char    Laenge = 12
Name = GESCHAFTSPARTN        Typ = Char    Laenge = 10
Name = PARTNERTYP            Typ = Char    Laenge = 1
Name = VORNAME               Typ = Char    Laenge = 37
Name = NACHNAME              Typ = Char    Laenge = 35
Name = NAME_1                Typ = Char    Laenge = 40
Name = NAME_2                Typ = Char    Laenge = 40
Name = ANREDE                Typ = Char    Laenge = 4
Name = AKADEM_TITEL          Typ = Char    Laenge = 4
Name = C_O                   Typ = Char    Laenge = 40
Name = STRASSE               Typ = Char    Laenge = 31
Name = HAUSNUMMER            Typ = Char    Laenge = 9
Name = ZUSATZ                Typ = Char    Laenge = 1
Name = POSTLEITZAHL          Typ = Char    Laenge = 7
Name = ORT                   Typ = Char    Laenge = 23
Name = ORSTEIL               Typ = Char    Laenge = 22
Name = PLZ_POSTFACH          Typ = Char    Laenge = 5
Name = POSTFACH              Typ = Char    Laenge = 6
Name = TELEFON               Typ = Char    Laenge = 19
Name = LAND                  Typ = Char    Laenge = 2
Name = KONTFINDMERKMAL       Typ = Char    Laenge = 2
Name = EINGZAHLWEG           Typ = Char    Laenge = 1
Name = ABWEICH_ZAHLER        Typ = Char    Laenge = 10
Name = BANKVERBEING          Typ = Char    Laenge = 4
Name = AUSGZAHLWEG           Typ = Char    Laenge = 1
Name = ABW_ZAHLEMPFG         Typ = Char    Laenge = 10
Name = BANKVERBAUSG          Typ = Char    Laenge = 4

```

**Bild A.1:** dt-Datei Beispiel (eigener Screenshot)

Resources ▾
Cookbooks
Services ▾
Statistics ▾
Hello, martin
Help
Administration ▾
Account ▾

Dashboard
Services
Statistics

Transformation configuration

1 ai-adressen-fuellen

F: 23/02/16 10:08:02
Status: Completed
Runtime: 0h 1m 29s

2 ai-kunden-fuellen

F: 23/02/16 09:44:19
Status: Completed
Runtime: 0h 1m 29s

F: 24/02/16 10:14:34
Status: Completed
Runtime: 0h 0m 5s

F: 24/02/16 10:14:18
Status: Completed
Runtime: 0h 0m 8s

Queue of scheduled transformations (Entries: 0)

Agents

Stan (1 nodes with 9,235 G...

F
L
U
C

Status: Connected

Configuration: -

Agent: Parser: Idle Processor: Idle

Bild A.2: Dashboard Tensei-Data (eigener Screenshot)

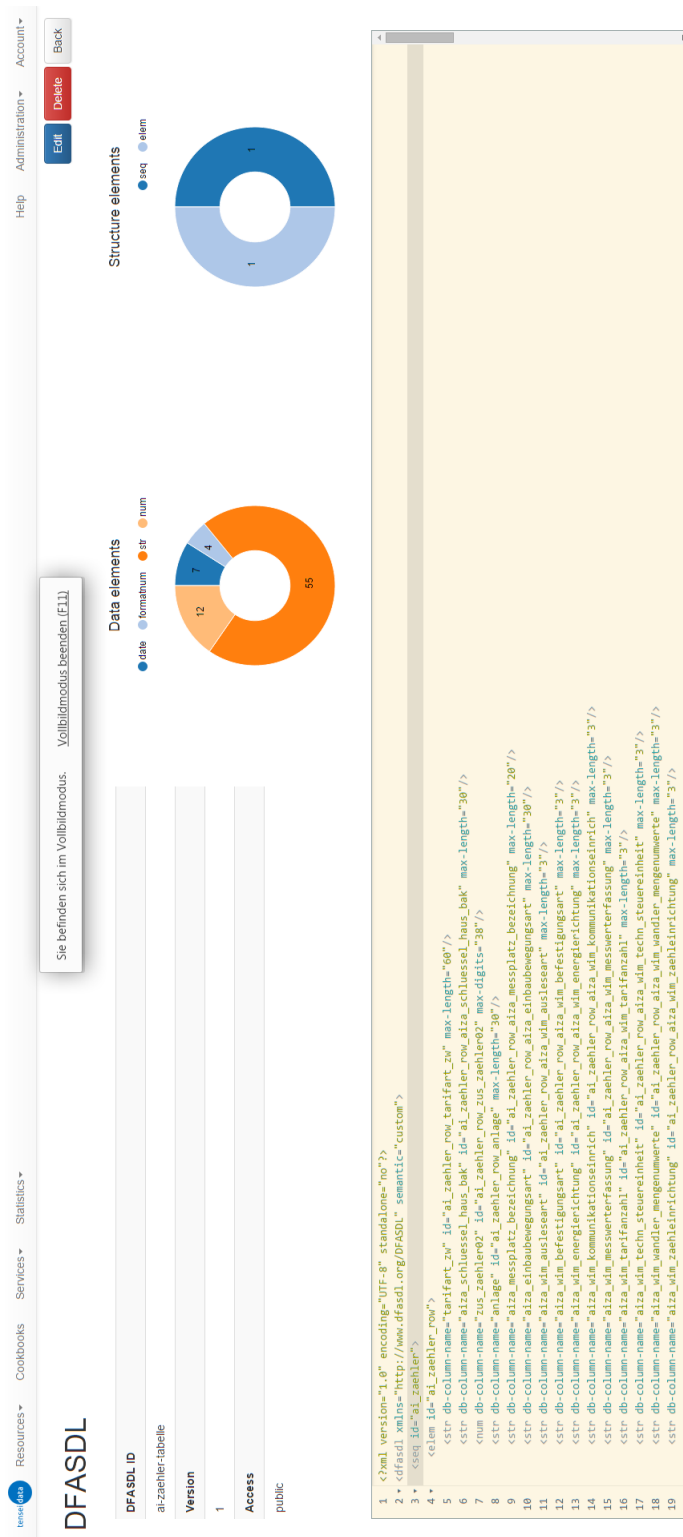


Bild A.3: DFASDL Tensei-Data (eigener Screenshot)



**Cronjob-4**

**Transformation configuration**

al-kunden-tuellen

The selected transformation configuration will be executed when the Cron is running.

**Format**

0 0 10 ? \* 1-5

The format is comprised of 6 required fields and one optional field separated by white space.

☒ Definition

**Description (optional)**

You can add an optional description.

**Cron activated**

Deactivated

If the cron is activated, it is automatically executed at the specified time value of the format.

**Access**

public

Select the access restriction for the Connection information

**Bild A.4:** Cronjob Tensei-Data (eigener Screenshot)

**Trigger-1**

**Transformation configuration**

al-adressen-quellen

The selected transformation configuration will be executed when the Trigger is running.

Select the source type for the trigger.

☐ Apache Camel URI ☒ Completion of another transformation configuration

**Transformation configuration**

al-kunden-quellen

If the transformation configuration you select here is completed then the Trigger will be run.

**The description of the trigger (optional)**

You can add an optional description for the trigger.

**Trigger activated**

Activated

If the trigger is activated, it is automatically executed when the endpoint URI sends the signal.

**Access**

public

Select the access restriction for the Connection Information

**Save** **Cancel**

**Bild A.5:** Trigger Tensei-Data (eigener Screenshot)

## Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die hier vorliegende Arbeit selbstständig, ohne unerlaubte fremde Hilfe und nur unter Verwendung der aufgeführten Hilfsmittel angefertigt habe.

Ort, Datum

Unterschrift

**Thesen**  
**Bachelor-Thesis**  
**Untersuchung des ETL-Tools Tensei-Data und dessen Einsatz im Prozess**  
**der Datenmigration bei der SIV.AG**

Eingereicht am: 8. März 2016

von: Martin Pohl  
geboren am 13. April 1981  
in Rostock

Betreuer: Prof. Dr.-Ing. Antje Raab-Düsterhöft  
2. Prüfer: Betriebswirt (VWA) Detlef Herold

1. kurze Stichpunktartige Auflistung Diskussionswürdiger Punkte der Diplomarbeit
2. hier eine weitere These