

R&D Projekt

Semesterendbericht

**Bezeichnung**: Erstellung einer JRZ Demodatenbank

**Projektschlüssel:** RD16-03

**Betreuer**: DI Eduard Hirsch, DI Fabian Knirsch, BSc

**Kurzbeschreibung**: Konvertierung verschiedener Smartmeter Messdaten und Ablage in einer gemeinsamen Datenbank mit rollenbasiertem Zugriff.

**Beteiligte** **Firma**: Salzburg AG

**Studenten**: Isdor Reimar Klammer, BSc.

Maximilian Unterrainer, BSc.

Christopher Wieland, BSc.

Puch/Salzburg, 19.02.2017

Inhaltsverzeichnis

[1 Einleitung 1](#_Toc475303424)

[1.1 Motivation 1](#_Toc475303425)

[1.2 Umgebung 2](#_Toc475303426)

[2 Recherche 3](#_Toc475303427)

[2.1 ER Modell 3](#_Toc475303428)

[2.2 Analyse von Datenbankanforderungen 3](#_Toc475303429)

[2.2.1 Analyse bestehendes Datenmodell 4](#_Toc475303430)

[2.2.2 Performanceanalyse 4](#_Toc475303431)

[2.3 Analyse vorhandener Software 4](#_Toc475303432)

[2.4 Rollendefinition 5](#_Toc475303433)

[3 Umsetzung 6](#_Toc475303434)

[3.1 Systemarchitektur 6](#_Toc475303435)

[3.2 ER Modell 6](#_Toc475303436)

[3.3 Analyse der Datenmodelle 7](#_Toc475303437)

[3.3.1 Hadoop Tests 7](#_Toc475303438)

[3.3.2 MySQL Tests 7](#_Toc475303439)

[3.3.3 Testdaten 7](#_Toc475303440)

[3.4 Rollendefinition 8](#_Toc475303441)

[4 Ergebnisse 9](#_Toc475303442)

[4.1 Systemarchitektur 9](#_Toc475303443)

[4.2 ER Modell 11](#_Toc475303444)

[4.2.1 Analyse bestehendes Datenmodell 11](#_Toc475303445)

[4.2.2 Anforderungen von Energieversorgern und Netzbetreibern 12](#_Toc475303446)

[4.2.3 Analyse der Usecases 13](#_Toc475303447)

[4.2.4 Weitere Datenmodelle 13](#_Toc475303448)

[4.2.5 Rechtliche Rahmenbedingungen und daraus resultierende Einschränkungen der möglichen Messwerte 13](#_Toc475303449)

[4.2.6 Einbeziehung zusätzlicher Domänen 14](#_Toc475303450)

[4.2.7 Das COSEM Modell 14](#_Toc475303451)

[4.2.8 ER-Modell Festlegung 15](#_Toc475303452)

[4.3 Analyse möglicher Datensätze 17](#_Toc475303453)

[4.3.1 MySQL Datenbank 17](#_Toc475303454)

[4.3.2 Hadoop Datenbank 19](#_Toc475303455)

[4.3.3 Zusammenfassung der Datenbankanalyse 19](#_Toc475303456)

[4.4 Analyse vorhandener Software 20](#_Toc475303457)

[4.5 Rollendefinition 21](#_Toc475303458)

[5 Weitere Schritte 23](#_Toc475303459)

[A Literaturverzeichnis 24](#_Toc475303460)

[B Anhang 26](#_Toc475303461)

[B.1 SQL Messungen 26](#_Toc475303462)

Abkürzungsverzeichnis

AMCS Advanced Meter Communication System

COSEM Companion Specification for Energy Metering

DAVID-VO Datenformat- und Verbrauchsinformationsdarstellungs Verordnung

DLSM Device Language Messaging Specification

ETSI European Telecommunication Standards Institute

JRZ Josef Ressel Zentrum für Anwenderorientierte Smart Grid Privacy, Sicherheit und Steuerung

GDPR EU General Data Protection Regulation

IMA-VO Intelligente Messgeräte-AnforderungsVO 2011

MMS Manufacturing Messaging Specification

OBIS Object identification system, entsprechend der EN 62056-01

OSGP Open Smart Grid Protocol

RBAC Role Based Access Control

SML Smart Message Language

SOAP Simple Object Access Protocol

DBM Datenbankmodell

Abbildungsverzeichnis

Abbildung 1: Abfrage Durchschnitt pro Tag 7

Abbildung 2: Komponentenmodell SmartValAPI 11

Abbildung 3: ER-Modell Entitäten 15

Abbildung 4: Datenbankbeziehungen 16

Abbildung 5: Dauer des Datenimports 18

Abbildung 6: Dauer Berechnung des Durchschnittsverbrauchs 19

Abbildung 4: Rollendefinition - UseCase Diagramm 21

Tabellenverzeichnis

Tabelle 1: Datenformat CSV 7

Tabelle 2: Dauer des Datenimports 17

Tabelle 3: Dauer Berechnung des Durchschnittsverbrauchs 18

# Einleitung

Durch die Verabschiedung der Richtlinie 2009/72/EC [1] sind die Mitgliedsstaaten der EU aufgefordert, deren Inhalte in nationales Recht umzusetzen. Thema dieser Richtlinie ist es, die vorhandenen analogen Stromzähler durch digitale Smart Meter zu ersetzen. Mit der flächendeckenden Installation stehen sowohl den Netzbetreibern als auch den Energieproduzenten und den Verbrauchern Möglichkeiten das Netz op [2]timal zu nützen, Energie zu günstigen Preisen zu erwerben und Energieverschwendung zu verringern. Um diese Vorteile zu nützen, ist Kommunikation bezüglich des aktuellen Verbrauchs, der Netzbelastung und der im Netz vorhandenen Energie notwendig.

Über Kommunikationsprotokolle tauschen Verteilstationen, Energieeinspeiser und Smart Meter beim Endkunden Daten bezüglich des Verbrauchs aus. Der Preis für diese Vorteile ist die notwendige, zumindest teilweise Offenlegung des Energieverbrauchs des Endkunden.

Im Spannungsfeld von Schutz der Privatsphäre einerseits, und maschineller Messdatenauswertung im Rahmen des Erlaubten andererseits sollen die Ergebnisse dieses Projekts für Komfortverbesserung sorgen.

## Motivation

Es existieren unterschiedliche Strukturen und Protokolle für die Übertragung und Speicherung von Messdatenwerten. Beispiele dafür sind:

* DLSM/COSEM [2]
* SML [3]
* MMS und SOAP Abbildungen der IEC 61850 [Referenz]

Weitere Unterschiede bestehen in der Granularität der Daten, die Smart Meter Modelle zur Verfügung stellen, ebenso die Frequenz, mit der Werte ausgelesen werden können, ist unterschiedlich. All diese Unterschiede erschweren eine vergleichende Auswertung von Messdaten aus unterschiedlichen Quellen.

Dieses Projekt verfolgt vier Hauptziele:

* Schaffung einer Datenbank zur gemeinsamen, strukturierten Ablage von Messdaten
* Importmöglichkeit unterschiedlicher Messdatenformate
* Einfache Verwaltung des Zugriffs über eine rollenbasierte Zugriffskontrolle
* Geregelter Zugriff auf Messdaten über eine Schnittstelle (API)

Mit der Umsetzung des Projektes soll eine programmtechnische Auswertung der Messdaten vereinfacht, ermöglicht und geregelt erfolgen können.

## Umgebung

Die Umsetzung des Projektes erfolgt unter zu Hilfenahme von bereits im Umfeld der Fachhochschule Salzburg, beziehungsweise des Josef Ressel Zentrums für Anwenderorientierte Smart Grid Privacy, Sicherheit und Steuerung durchgeführten Projekte. Im Detail sind dies:

* Bestehendes Datenmodell: Ausgehend von „Christians Datenmodell“ (Details siehe Abschnitt ER Modell festlegen) wird die Eignung für großen Mengen von Messdaten evaluiert und steht zur Verfügung und stellt die Ausgangsbasis für die Untersuchung eines
* Importmodule (BAC1 Gruppe Oberluggauer und Co): parallel zur Projektumsetzung wird ein Programmpaket erstellt, welches unterschiedliche, in Dateien vorliegende, Datenformate in die Datenbank importiert.
* RBAC System (BAC1 Gruppe – OpenTC): stellt ein Softwarepaket zur Verfügung über das der rollenbasierte Zugriffsschutz realisiert wird.

Weitere verwendete Softwarepakete werden im Abschnitt Systemarchitektur erwähnt.

# Recherche

Die folgenden Recherchen sind notwendiger Teil des Projektes für die Informationsbeschaffung sowie der Prävention von eventuell doppelt gemachter Arbeit.

## ER Modell

Um ein geeignetes Datenbankmodell zur Verfügung stellen zu können, müssen vorhandene Datenbankmodelle genauer betrachtet werden.

Das Datenmodell zu „Christians Datenbank“ wird analysiert und auf Erweiterungsnotwendigkeiten untersucht. Die Analyse erstreckt sich auf die fachlichen Anforderungen durch das JRZ als Auftraggeber und berücksichtigt des Weiteren die möglichen Bedürfnisse von Energieversorgern und Netzbetreibern.

Zusätzlich ergibt sich die Notwendigkeit von technischen Erweiterungen durch die Verwaltung von Zugriffsrollen und Gruppen und Besitzern der Meterdaten. Des Weiteren wird Wert auf die Kompatibilität zu bestehenden Anwendungen gelegt.

## Analyse von Datenbankanforderungen

Es gibt herauszufinden, welches Datenbankmodell (DBM) für die Umsetzung des Projektes ideal ist. Um ein geeignetes DBM zu finden, wurden verschiedene Typen wie SQL, NoSQL und Hadoop genauer betrachtet.

Für die Analyse wird folgendermaßen Vorgegangen:

* Analyse des bestehenden Datenbankmodells („Christians Datenbank“)
* Performanceanalyse der zu testenden DBM sowie
* Einlesen in relevante Produkte

Beim Meeting mit der Projektbetreuung am 27.1.2017 wurde der Hinweis gegeben, dass Hadoop für dieses Projekt interessant sein könnte und wir uns das näher ansehen sollten. Zudem wurde erwähnt, dass es von Hortonworks eine Sandbox gibt, auf der ein fertig konfiguriertes Hadoop System mit unterschiedlichsten Tools verfügbar ist.

Ein großer Teil dieses Arbeitspaketes bestand darin, sich in Hadoop einzuarbeiten, Tutorials durchzumachen und erste Erfahrungen mit Big Data Systemen zu machen.

Bei den Gesprächen mit der Projektbetreuung hat sich allerdings auch herausgestellt, dass ein Weiterverwenden des bestehenden Datenmodells [Christians ER Modell] wünschenswert ist, da es bereits einiges an Software dafür gibt.

Des Weiteren spricht dafür, dass es eine BAC1 Gruppe gibt, welche sich mit dem Datenimport von frei verfügbaren Smartmeter Datensätzen beschäftigt hat. Der aktuelle Stand dieser Arbeit ist nach Aussage der Gruppe soweit, dass Daten normiert importiert werden können, die Performance allerdings noch nicht optimal ist. Das ist zum aktuellen Zeitpunkt noch nicht verifiziert, allerdings gehen wir davon aus, dass dieses Projekt nach Performanceoptimierung einsatzbereit ist. Da die verwendetet Programmiersprache allerdings C# ist, ist auf jeden Fall eine Portierung notwendig, da wir uns für das Projekt auf Java geeinigt haben.

### Analyse bestehendes Datenmodell

Bei Analyse des ER Modells stellt sich heraus, dass nur in der Tabelle ‚meter\_data‘ wirklich große Datenmengen vorhanden sind und performancekritische Abfragen ausgeführt werden. Daher konzentriert sich die erste Analyse ausschließlich auf diese Tabelle beziehungsweise eine Teil-Version davon.

### Performanceanalyse

Für erste Tests wurde ein Teil der REDD ‚low\_freq‘ Daten verwendet. Diese wurde in verschiedene Datenbanken importiert und es wurden darauf Abfragen ausgeführt. Ziel dieser Analyse war es ein Gefühl zu bekommen, wie sich die Performance mit Zunahme an Daten verhält und ob eine SQL Datenbank überhaupt in Frage kommen kann.

## Analyse vorhandener Software

Da das Projekt ein Teilprojekt eines Größeren ist, galt es festzustellen, ob schon bestehende Software für das Projekt verwendet werden kann. Somit wird redundante Arbeit vermieden und der Fokus auf die noch benötigen Dinge gesetzt.

## Rollendefinition

Mit SmartValAPI wird der Zugriff auf sensible Daten verwaltet, daher ist die Schutz des Zugriffs unerlässlich. Auf Aspekte der Datensicherheit wie physischer Zugang zum Datenbank beziehungsweise Applikationsserver geht der Abschnitt „Installation“ näher ein, dieser Abschnitt beleuchtet den Zugriff über Rollen und legt die Rollendefinition fest.

Die Kernaufgaben sind wie folgt:

* Rollen identifizieren und definieren.
* Unterschiede in fachlichen und technischen Anforderungen in notwendigen Rollen überleiten.
* Verbindung zum Code, beziehungsweise Funktionen herstellen.
* Rechtliche Umgebung einbeziehen.

# Umsetzung

In dieser Sektion wird beschrieben, wie bisher die Pakete Systemarchitektur, ER-Modell und die Rollendefinition umgesetzt wird.

## Systemarchitektur

Um die Systemumgebung festzulegen, und vor allem die Software passgenau in die Softwarelandschaft des JRZ einfügen zu können, werden die bestehenden System wie zum Beispiel Smart Viz, die in Entstehung befindlichen Zugriffsysteme (BAC1 Gruppe – OpenTC) und Importprogramme (BAC1 Gruppe Oberluggauer und Co). Was war mit Open-Nes? In die Überlegungen werden weitrs etwaige Kosten für Lizenzen und andererseits Sicherheitsaspekte einbezogen [4]. Funktionen nach Aufruf und Zusammengehörigkeit gruppieren.

## ER Modell

Beim Festschreiben des geeigneten Datenbankmodells wurde wie unterhalb beschrieben vorgegangen:

* Recherche nach Datenmodellen, die bereits abseits von „Christians Datenbank“ im Einsatz sind, wie zum Beispiel COSEM.
* Feststellen der Wertemenge die SmartMeter zur Verfügung stellen und herausarbeiten welche davon gespeichert werden.
* Kontaktaufnahme mit österreichischen Energieversorgern bezüglich der Werte von Interesse.
* Analyse der Usecases von Österreichs Energie, ergibt vor allem Daten bezüglich der Steuerung des Smart Meter als solches, und nicht der inhaltlichen Bedeutung der übertragenen Messdaten.
* Erhebung der rechtlichen Rahmenbedingungen und sich daraus ergebende Einschränkungen.
* Prüfung auf Verwendbarkeit des Datenmodells in weiteren Domänen wie zum Beispiel: Gas, Wärme und Wasser.
* Festschreiben des Datenmodells.

## Analyse der Datenmodelle

### Hadoop Tests

Für die Tests wurde eine virtuelle Maschine mit der ‚Hortonworks Hadoop Sandbox‘ aufgesetzt. Der Maschine wurden alle Cores des Hosts[[1]](#footnote-1) sowie 8GB Arbeitsspeicher zur Verfügung gestellt.

### MySQL Tests

Für die Tests wurde ein MySQL Server auf dem o.g. Testsystem aufgesetzt. Im Gegensatz zu den Hadoop Tests allerdings direkt auf dem Host-Betriebssystem.

### Testdaten

Als Testdaten wurden REDD Daten eines Hauses importiert. Die Abfrage, die darauf abgesetzt wurde, ergibt den Durchschnittsverbrauch pro Tag pro Kanal.

Die Daten liegen in CSV Dateien mit folgendem Format vor.

|  |  |
| --- | --- |
| timestamp | power |
| int | double |

Tabelle 1: Datenformat CSV

Der Timestamp lässt sich mit FROM\_UNIXTIME(timestamp) in ein Datum umwandeln womit gerechnet werden kann.

Folgende Abfrage liefert den Durchschnittsverbrauch gruppiert nach Kanal, Monat und Tag. Die Dauer ist unser Performance Index.



Abbildung 1: Abfrage Durchschnitt pro Tag

Es wurden Schritt für Schritt Kanäle hinzugefügt um einen wachsenden Datensatz zu simulieren.

## Rollendefinition

Untersuchen der rechtlichen Rahmenbedingungen:

die GDPR und das ElWOG §84 legt die Rahmenbedingungen für die Erfassung, die Weiterleitung und die Speicherung von Messdaten fest. Abs. (1) regelt die Erfassung von Viertelstundenwerten und den Zeitraum der Speicherung. Des Weiteren wird die Weitergabe des Tagesverbrauchs an den Netzbetreiber geregelt.

Die Weitergabe der Viertelstundenwerte ist nur bei ausdrücklicher Zustimmung des Endverbrauchers möglich (Abs. (2)), in begründeten lokalen Einzelfällen zur Aufrechterhaltung eines sicheren Netzbetriebes ist die Weitergabe von Viertelstundenwerten ohne ausdrückliche Erlaubnis möglich. Außerdem ist in der DAVID-VO §4 der Zugriff für vom Endverbraucher bevollmächtigte Dritte auf die Daten im Kundenportal des Endverbrauchers geregelt.

Überlegungen bezüglich der Administration führen zu einer Trennung von denjenigen Personen die Messdaten in die Datenbank importieren und jenen die die Administration des Gesamtsystems überhaben. Durch die Vergabe von beiden Rollen an eine Person ist die gesamte Administration durch eine Person, insbesondere in kleinen Installationen, möglich.

# Ergebnisse

Bisherige Ergebnisse in Bezug Systemarchitektur, ER- Modell, Analyse vorhandener Datensätze sowie vorhandener Software, Recherche noch benötigter Software beziehungsweise Funktion und momentaner Rollendefinition werden nun vorgestellt.

## Systemarchitektur

* Hardware: für den Betrieb ist keine explizite Hardware vonnöten, vom JRZ wurde eine virtuelle Maschine im Bladecenter zur Verfügung gestellt.
* Betriebssystem: auf Grund der größeren Erfahrung der Entwickler mit der Administration erfolgt die Implementierung auf einem Windows System (Windows Server 2012 R2),
* Programmiersprache: bestehende Software wurde im Umfeld des JRZ in Java implementiert, ebenso werden laufenden Projekte in Java programmiert, die Erfahrung der Umsetzenden reicht aus, um die Anforderungen umzusetzen. Im Sinne einer einfachen Übergabe, Weiterführung und Wartung fällt die Entscheidung, dieses Projekt in Java zu implementieren.
* Datenbank: neben unterschiedlichen RDBMS (Oracle Database Server, MySQL) wurden NoSQL Datenbankensysteme untersucht (MongoDB, Hadoop). Im Sinne der Integration in die bestehende Softwarelandschaft wird MySQL eingesetzt. Ziel des Projektes ist eine Integrationsdatenbank, was dazu führt, dass sehr große Datenmengen verwaltet werden können müssen, und der Zugriff in akzeptabler Zeit erfolgt. MySQL unterstützt lokale (partitioning) und verteilte (sharding) Fragmentierung. Nach dem „Guide to Scaling Web Databases with MySQL Cluster“ [5] kann der zu erwartenden Menge von Schreibzugriffen (~500k/d) durch Verteilung auf einzelne Nodes begegnet werden. Sollte ein Cluster mehrerer Nodes zu Performanceeinbußen führen können einem Cluster einfach weitere Nodes hinzugefügt werden, eine Anpassung der Applikation ist in diesem Fall nicht notwendig, über die Administration des MySQL Clusters wird das Partitionierung transparent für die Anwendung durchgeführt. Details dazu sind im Abschnitt „Datenbankzugriff“ erläutert.
* Datenbankdesigntool: untersucht wurden Oracle SQL Developer Data Modeler 4.1.5 und die MySQL Workbench 6.3.9. Für den graphischen Entwurf des Entity-Relation-Modells bieten beide Tools Unterstützung. Modeler wie Workbench generieren aus dem ER-Modell sowohl die graphische Übersicht der Tabellen und Schlüssel, als auch Generation der Skriptdateien zur Anlage der Tabellen, Indizes, Einschränkungen bezüglich referentieller Integrität und die automatische Vergabe eindeutiger Schlüssel. Da als Datenbanksystem MySQL eingesetzt wird, liegt es nahe das Designtool vom gleichen Hersteller einzusetzen und damit Kompatibilitätsprobleme beziehungsweise Nachbearbeitungen zu vermeiden.
* Schnittstellen zu BAC1 (Autorisierung): der bisherige Entwurf des API ist in einer generischen Form implementiert, sodass einerseits ein Austausch des Autorisierungsmoduls möglich ist (zum Beispiel in einem ersten Schritt realisiert als Benutzername/Passwort Zugang). Eine Abhängigkeit zum Datenmodell wurde identifiziert: die Form der Benutzerkennung des externen Rollensystems muss in den Tabellen untergebracht werden könne (Details siehe Abschnitt Datenmodell).
* Schnittstellen zu BAC1 (Importmodule): gemeinsame Basis für die API und Importmodule ist das Datenmodell (Details zu notwendigen Erweiterungen siehe im Abschnitt Datenmodell), programmtechnisch wurden keine Überschneidungen identifiziert.
* Die Anbindung und die Veröffentlichung der Schnittstellen: <wir haben noch nicht festgelegt wie das API aufgerufen werden kann> Webservice? Datenstrukturen?
* Aus den Anforderungen ergeben sich folgende Komponenten, deren Abhängigkeiten untereinander und Verbindungen zueinander, die im UML Diagramm (Abbildung 1) dargestellt werden:



Abbildung 2: Komponentenmodell SmartValAPI

Die Details zu den Komponenten und die exportierten Methoden werden im Abschnitt „API Funktionen“ beschrieben. Der Übersicht halber wird die Komponente „Logger“, die jede der angeführten Komponenten verwendet, nicht in der Übersicht dargestellt.

## ER Modell

Hier werden die Ergebnisse des zu verwendenden Datenbankmodells gezeigt.

### Analyse bestehendes Datenmodell

(„Christians Datenmodell“) Messdaten werden als Tupel in einer Tabelle (meter\_data) abgelegt, je Messzeitpunkt werden folgende Werte, sofern vom Smart Meter zur Verfügung gestellt, gespeichert.

* Nutzdaten (Momentanwerte):   
  je Phase: aktuelle Leistung, aktueller Stromverbrauch (sofern vom Smart Meter übertragen in dieser Granularität zur Verfügung gestellt, sonst als Einzelwert in Phase1). 4 Werte (count\_register1 – count\_register4) die abhängig vom Smart Meter (meter\_type) belegt werden (Details dazu im Abschnitt „Importprogramme“),   
  Gesamtwert: Spannung, Frequenz, kumulierte Werte: Verbrauch kWh.
* Verwaltungsdaten (zur Identifikation):  
  meter\_id des Smart Meters: Fremdschlüssel zu meter\_management,  
  data\_id: eindeutiger Schlüssel des Messdaten-Tupels,  
  timestamp: Erstellungszeitpunkt zu dem die Nutzdaten aufgezeichnet werden.

### Anforderungen von Energieversorgern und Netzbetreibern

Anfragen an Ebner Strom GmbH, Energie AG, Energie Steiermark, EVN AG und Netz Niederösterreich GmbH, Linz AG, Salzburg Netz GmbH, TINETZ – Tiroler Netze GmbH, Vorarlberger Energienetze GmbH und Wien Energie ergeben:

* Verweis auf das ElWOG: die erheb baren Daten sind klar definiert eine darüberhinausgehende Erfassung von Messwerten ist nicht zulässig.
* Verweis auf die laufende Ausschreibung der Smart Meter Geräte, hier werden die Messwerte in Pflicht- und optionale Werte eingeteilt.
* Die im Datenmodell vorhandene Netzfrequenz ist auf Grund der Rückmeldungen nicht auswertungsrelevant auf Basis einzelner Smart Meter nicht interessant.
* Keine Rückmeldung, beziehungsweise nur telefonisch, wobei sich aus letzterer keine weiteren Datenfelder ergeben.
* Hervorzuheben sind die Antworten von Salzburg Netz GmbH und Vorarlberger Energienetze GmbH, bezüglich der Messwerte steht für den Netzbetreiber, neben den bereits vorhandenen Werten stehen der Winkel zwischen Spannung und Strom, und die daraus resultierende Blindleistung im Fokus.

Im Gegenzug dazu kann auf die Frequenz verzichtet werden, dieses Datenfeld verbleibt auf Kompatibilitätsgründen im ERM.

### Analyse der Usecases

[6] und des Lastenheftes [7] von Österreichs Energie. Die Usescases beschäftigen sich vor allem mit Daten bezüglich der Steuerung des Smart Meter als solches und nur am Rande mit der inhaltlichen Bedeutung der übertragenen Messdaten.

Neben den, in meter\_data bereits vorhandenen Datenfeldern, bietet das Lastenheft optional die Möglichkeit der Auslesung der Blindleistung (I.-IV. Quadrant), diese Werte werden übernommen, beziehungsweise sofern geliefert in meter\_data abgelegt.

Weitere Datenfelder, wie sie zum Beispiel das ETSI in der Definition des OSGP [8] vorschlägt, werden nicht in die Menge der gespeicherten Daten aufgenommen, da diese weder von den Energieversorgern noch von den Netzbetreibern gewünscht werden (Rückmeldungen, Lastenheft).

### Weitere Datenmodelle

Fusco et al. [9] schlagen einen dualen Betrieb von RDBMS und einer NoSQL Datenverwaltung vor. Die Messwertedatenpakete werden in unterschiedlicher Granularität, zum Beispiel Rohdaten, Messdaten aggregiert nach Smartmeter, Zeitraum und vorverarbeitet zum Beispiel Durchschnittsverbrauch über einen bestimmten Zeitraum abgelegt.

Ziel dieses Ansatzes ist es auch Vorhersagen über den zukünftigen Verbrauch zu tätigen. Jene Daten die Messwerte betreffen werden in einer separaten Komponente verwaltet, dadurch soll ermöglicht werden die Datenhaltung mittelfristig ersetzen zu können (Details siehe Abschnitt Systemarchitektur).

Ein dezentraler Ansatz, wie zum Beispiel das COUGAR Sensornetzwerk [Referenz Cougar], als Alternative zu einer zentralen Datenbank bietet zwar den Vorteil, einen zentralen Angriffspunkt zu vermeiden, Messdaten hingegen ausschließlich ad hoc auszulesen widerspricht den Regelungen des ElWOG und scheidet damit aus.

### Rechtliche Rahmenbedingungen und daraus resultierende Einschränkungen der möglichen Messwerte

Die Erfassung, Übertragung und Speicherung von Smart Meter Messdaten wird in vier Richtlinien geregelt:

* ElWOG
* GDPR
* IMA-VO
* DAVID-VO

Geregelt werden einerseits Mindestanforderungen an Smart Meter, andererseits die Inhalte und die Frequenzen, mit denen die Werte ausgelesen werden dürfen. Im ElWOG werden in §84 dem Verbraucher die Daten bezüglich des „Verbrauchs der über ein intelligentes Messgerät gemessen wird“ zeitnah zur Verfügung zu stellen. Es erfolgt keine genauere Definition, welche Daten das im Detail sind, lediglich die Frequenzen, mit denen ausgelesen wird, werden festgelegt, Details dazu, siehe Abschnitt Rollendefinitionen.

Einer der Hauptgründe für die Reglementierung ist der Schutz der Privatsphäre [10]. Bezüglich der auslesbaren Daten legt die [7] jene Daten fest, die von einem Smartmeter übertragen werden müssen. Die IMA-VO und DAVID-VO befassen sich mit der Einführung der der intelligenten Zähler und der Weitergabe der ausgelesenen Daten und haben keinen Einfluss auf das Datenmodell.

### Einbeziehung zusätzlicher Domänen

Aktuell wurde die die Richtlinie 2006/3 2/EG in Österreich für die Datenerfassung und Kommunikation von Messgeräten für elektrische Energie umgesetzt, derzeit gibt es in Österreich keine äquivalenten Grundlagen für Gas und Wärme und Wasser. Technisch ist eine Erweiterung um jene Felder, die nach der rechtlichen Festlegung erfasst werden sollen, problemlos möglich, daher gehen wir in der Umsetzung dieses Projekts darauf nicht weiter ein.

### Das COSEM Modell

[2] versucht hier einen Standard zu etablieren. Nachdem die Salzburg AG als Partner des Projektes in den Ausschreibungen für Smart Meter die Kommunikation über DLSM/COSEM als Muss-Kriterium festlegt, wird in der Folge dieses Protokoll in Verbindung mit den Anforderungen des Lastenheftes von Österreichs Energie [7] als Ausgangspunkt für die Festlegung der Datenbank herangezogen.

### ER-Modell Festlegung

Ausgangsbasis ist „Christians Datenbank“, bei der Erweiterung wurde auf die Kompatibilität zu bestehenden Programmen geachtet, um diese Applikationen ohne Anpassung auch weiterhin gegen die gleiche Datenbank betreiben zu können. Folgendes Diagramm gibt einen Überblick über die Entitäten:



Abbildung 3: ER-Modell Entitäten

Aus dem existierenden Modell werden die bestehenden Entitäten übernommen, für die zusätzlichen Anforderungen ergeben sich folgende Erweiterungen:

Zusätzliche Entitäten:

* userAdmin: die Benutzerverwaltung wird als eigene Komponente eingebunden, um die Kopplung lose ausführen zu können wird im System lediglich der Schlüssel (der Benutzername, eine LDAP ID, ein etwaiges anders identifizierendes Merkmal) hinterlegt.

Zusätzliche Beziehungen:

* istSystembenutzer (userAdmin – customer): optionale Erweiterung zum customer, ermöglicht die Ablage einer externen Benutzerkennung.
* berät\_oder\_verwaltet (customer – meter\_management): verbindet m customer mit n meter\_management, Zweck ist die Abbildung von Verbinungen wie zu, Beispiel: Netzbetreiber versorgt Meter, Energieberater berät Eigentümer von Meter.
* ist\_Kunde\_von: ermöglicht die hierachische Verbindung von Kunden, zum Beispiel Energeversorger mit Kunden.

Zusätzliche Attribute:

* in meter\_data: reactive\_P1, reactive\_P2, reative\_P3: Blindleistungsanteil aufgeteilt nach Phase, sofern nur gesamt übermittelt in P1, wenn nicht ausgelesen: 0.
* phi\_P1, phi\_P2, phi\_P3: Phasenwinkel von Strom und Spannung, je Phase, wenn ausgelesen,

Aus den Typen der Beziehungen ergeben sich folgende Tabellen:



Abbildung 4: Datenbankbeziehungen

Alternative für meter\_data: da die Tabelle das maximale Set an Daten abbilden kann entstehen eventuell einige Tupel mit Null-Werten, sofern Smart Meter nicht alle Felder auslesen können. Die Alternative besteht in der Definition einer Tabelle die je Tupel einen Messwerttyp (bedingt eine weitere Definitionstabelle für die zugelassenen Werttypen) und einen Messwert.

Der Messwerttyp ist Teil des primären Schlüssels. Die Vorteile der Vermeidung von Nullwerten und der flexiblen Einführung neuer Werttypen, zum Beispiel auch Gas, Wasser oder Wärme, werden durch die Vervielfachung der Einträge einerseits und des Verlustes der Kompatibilität mit den bestehenden Softwarepaketen erkauft.

Formate der Datenfelder:

Aus der Tabellendefinition können die Formate der jeweiligen Felder ermittelt werden, inhaltlich wird festgelegt, dass in **meter\_management:**  meterId entsprechend der OBIS Identifikation befüllt wird (Details siehe Abschnitt „API Funktionen“).

## Analyse möglicher Datensätze

Es haben sich bei den Tests folgende Ergebnisse errechnen lassen:

### MySQL Datenbank

Zuerst wird die Dauer des Datenimports gemessen.

|  |  |  |
| --- | --- | --- |
| **Import Dauer** | | |
| **Vorhanden** | **Hinzugefügt** | **Dauer [s]** |
| - | 1.561.660 | 27 |
| 1.561.660 | 1.561.660 | 28 |
| 3.123.320 | 745.878 | 10 |
| 3.869.198 | 745.878 | 11 |
| 4.615.076 | 745.878 | 10 |
| 5.360.954 | 3.729.390 | 62 |
| 9.090.344 |  |  |

Tabelle 2: Dauer des Datenimports

Abbildung 5: Dauer des Datenimports

Hierbei ist zu sehen, dass die Dauer unabhängig von den bereits vorhandenen Datensätzen linear abhängig zu der Menge der Importierten Datensätze ist.

Als nächsten Schritt wird die Dauer des Query aus *Abbildung 1: Abfrage Durchschnitt pro Tag*gemessen.

|  |  |
| --- | --- |
| **Berechnung des Durchschnitts** | |
| **Anzahl Zeilen** | **Dauer** |
| 1.561.660 | 6 |
| 3.123.320 | 16 |
| 3.869.198 | 19 |
| 4.615.076 | 23 |
| 5.360.954 | 27 |
| 9.090.344 | 42 |

Tabelle 3: Dauer Berechnung des Durchschnittsverbrauchs

Abbildung 6: Dauer Berechnung des Durchschnittsverbrauchs

Abbildung 6 zeigt, dass die auch die Dauer der Abfrage linear abhängig zur Anzahl der vorhandenen Datensätze ist.

Abschließend wurden Tag und Monat aus dem Timestamp fix in die Tabelle geschrieben und mit einem Index versehen. Dadurch konnte die Berechnungsdauer bei ~ 10 Millionen Datensätzen von 42 Sekunden auf 8 reduziert werden.

### Hadoop Datenbank

Erste Tests haben gezeigt, dass die Abfragen mit HIVE annähernd gleich funktionieren wie mit SQL, die Performance allerdings auf dem Testsystem schlechter ist. Aufgrund der zufriedenstellenden Ergebnisse mit SQL wurde auf ebenso detaillierte Tests verzichtet. Die Performance lässt sich durch Hinzufügen von mehreren Nodes nahezu beliebig steigern.

### Zusammenfassung der Datenbankanalyse

Beim Import der Datensätze fällt auf, dass die Anzahl der bereits vorhandenen Daten keine Auswirkung auf die Dauer des Vorgangs hat, was uns entgegenkommt.

Die Dauer einer Berechnung des Durchschnitts nimmt proportional mit der Anzahl der Datensätze zu. Der REDD Datensatz mit niedriger Auflösung beinhaltet ungefähr 52 Millionen Datensätze.

Hier würde die Abfrage hochgerechnet 210 Sekunden dauern. Es hat sich aber gezeigt, dass sich durch die geschickte Verwendung von Indizes, die Performance des getesteten MySQL System sehr stark steigern lässt. Aus diesem Grund, und auch um die Kompatibilität zu den bestehenden Applikationen zu wahren, wird für dieses Projekt bis auf weiteres ein Relationales Datenbanksystem verwendet.

Falls sich im weiteren Verlauf des Projekts herausstellen sollte, dass eine SQL Datenbank nicht ausreichend ist, kann eine Hybrid Lösung angestrebt werden.

Hier wäre vorstellbar, dass die ganze Meterverwaltung weiterhin in einer SQL Datenbank verbleibt, die Messdaten allerdings in Hadoop abgespeichert werden.

## Analyse vorhandener Software

Für die Analyse bereits vorhandener Software wurde folgendermaßen vorgegangen:

* Abfragen der einzelnen Gruppen, welche Projekte verfolgt werden und welche Dokumente oder Projekte denen zur Verfügung stehen.
* Einschätzen, ob eines dieser Programme oder Projekte verwendet werden kann.
* Einholung der Dokumentation und des Projektes bei Interesse an einem Projekt.

Abfragen der einzelnen Gruppen, welche Projekte verfolgt werden und welche Dokumente oder Projekte denen zur Verfügung stehen.

* + Eine BAC Gruppe hat ein vielversprechendes Projekt (OpenTC). Hierbei handelt es sich um eine Software, die Rollenbasiertes Zugreifen auf Dokumente zulässt.
  + Eine weitere Gruppe wurde ausgeforscht, welche interessante Dokumentation hat. (BAC 1 Gruppe Oberluggauer und Co)

Einschätzen, ob eines dieser Programme oder Projekte verwendet werden kann:

* + Das BAC Projekt OpenTC könnte für unseren rollenbasierten Zugriff auf die Datenbank umgeschrieben werden.
  + Informationen der weiteren Gruppe (BAC1 Gruppe Oberluggauer und Co).

Einholung der Dokumentation und des Projektes bei Interesse an einem Projekt:

* + Die Dokumentation sowie das ganze Projekt konnte organisiert werden für Recherche.
  + Die Informationen der weiteren Gruppe. (BAC1 Gruppe Oberluggauer und Co)

## Rollendefinition

Folgende Rollen wurden identifiziert und mit den im Use-Case-Diagramm angeführten Anwendungsfällen verbunden, die in Klammern angeführten Attribute geben die grundlegende Ansiedlung der Personen, die die Rollen bekleiden.

Administrator (technisch): Stammdatenwartung (Meter Anlage und abhängige Daten),

Energieversorger (fachlich): jener, der den Verbraucher mit Energie versorgt.

Energieberater (fachlich): über die ausdrückliche Zustimmung des Eigentümers befugte Person.

Eigentümer (fachlich): derjenige dessen Verbrauch mit dem Smartmeter gemessen wird.

Importeur (technisch): Meterdaten in die Datenbank importieren, das Smart Meter muss bereits im System bestehen.



Abbildung 4: Rollendefinition - UseCase Diagramm

# Weitere Schritte

# Literaturverzeichnis

|  |  |
| --- | --- |
| [1] | D. E. P. u. d. R. d. E. Union, „Richtlinie 2009/72/EG des Europäischen Parlaments und des Rates vom 13. Juli 2009,“ 13 Juli 2009. [Online]. Available: http: //publications.europa.eu/resource/celex/32009R0713. [Zugriff am 19 Februar 2017]. |
| [2] | D. Craemer, K. a. Deconinck und Geert, „Analysis of state-of-the-art smart metering communication standards,“ in *Proceedings of the 5th young researchers symposium*, 2010. |
| [3] | B. f. S. i. d. IT, „BSI TR-03109 Technische Vorgaben für intelligente Messsysteme und deren sicherer Betrieb,“ 18 März 2013. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR03109/TR-03109-1\_Anlage\_Feinspezifikation\_Drahtgebundene\_LMN-Schnittstelle\_Teilb.pdf?\_\_blob=publicationFile. [Zugriff am 19 Februar 2017]. |
| [4] | *IT-Grundschutz-Profil für Open-Source-Software (GSProOSS),* 2010. |
| [5] | Oracle, „Guide to Scaling Web Databases with MySQL Cluster,“ 1970 Januar 01. [Online]. Available: https://www.mysql.de/why-mysql/white-papers/guide-to-scaling-web-databases-with-mysql-cluster/. [Zugriff am 19 Februar 2017]. |
| [6] | O. energie, „Smart Metering Use-Cases,“ 14 Dezember 2015. [Online]. Available: http://oesterreichsenergie.at/branche/stromnetze/smart-meter-use-cases.html?file=files/oesterreichsenergie.at/Downloads%20Netze/Smart%20Meter/Oesterreich%20Use%20Cases%20Smart%20Metering\_14122015\_Version\_1-1.pdf. [Zugriff am 19 Februar 2017]. |
| [7] | O. energie, „Lastenheft Smart Meter,“ 01 Juli 2013. [Online]. Available: http://oesterreichsenergie.at/branche/stromnetze/lastenheft-smart-meter.html?file=files/oesterreichsenergie.at/Downloads%20Netze/Smart%20Meter/Lastenheft\_SmartMeter\_1\_0.pdf. [Zugriff am 19 Februar 2017]. |
| [8] | ETSI, „Open Smart Grid Protocol (OSGP),“ 01 Januar 2012. [Online]. Available: http://www.etsi.org/deliver/etsi\_gs/OSG/001\_099/001/01.01.01\_60/gs\_osg001v010101p.pdf. [Zugriff am 19 Februar 2017]. |
| [9] | F. Francesco, „Data Management System for Energy Analytics and its Application to Forecasting,“ in *EDBT/ICDT Workshops*, Bordeaux, 2016. |
| [10] | S. W. M. Lisovich, „Privacy concerns in upcoming residental and commercial demand-response systems,“ in *Proc. of the Clemson University Power Systems Converence*, Clemson, SC, 2008. |
| [11] | E. P. u. Rat, „Energieeffizenz und Energiedienstleistungen und zur Aufhebung der Richtlinie 93/76/EWG des Rates,“ 05 April 2006. [Online]. Available: eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A32006L0032. [Zugriff am 19 Februar 2017]. |
| [12] | P. J. G. P. S. Bonnet, „Towards sensor database systems,“ in *International Converence on Mobile Data Management*, Berlin, 2001. |

# Anhang

## SQL Messungen

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Messung | Anzahl Zeilen | Affected | Query | DauerNoIdx | Dauer Idx |
| 1 | 1.561.660 | 23 | A | 1,172 | 1,218 |
| 2 | 1.561.660 | 23 | A | 1,297 | 1,235 |
| 3 | 1.561.660 | 23 | A | 1,25 | 1,312 |
| 4 | 1.561.660 | 23 | A | 1,266 | 1,265 |
| 5 | 2.307.538 | 745.878 | B | 52,203 | 188 |
| 6 | 2.307.538 | 745.878 | C | 32,86 | 420 |
| 7 | 2.307.538 | 58 | A | 4,265 | 4,453 |
| 8 | 2.307.538 | 58 | A | 4,172 | 4,703 |
| 9 | 2.307.538 | 58 | A | 4,218 | 6,937 |
| 10 | 2.307.538 | 58 | A | 5,468 | 7,031 |
| 11 | 3.869.198 | 1.561.660 | B | 82 |  |
| 12 | 3.869.198 | 1.561.660 | C | 76 |  |
| 13 | 3.869.198 | 81 | A | 11,359 |  |
| 14 | 3.869.198 | 81 | A | 11,203 |  |
| 15 | 3.869.198 | 81 | A | 12,625 |  |
| 16 | 3.869.198 | 81 | A | 10,891 |  |
| 17 | 4.615.076 | 745.878 | B | 51,203 |  |
| 18 | 4.615.076 | 745.878 | C | 44,609 |  |
| 19 | 4.615.076 |  | A | 15,063 |  |
| 20 | 4.615.076 |  | A | 15,89 |  |
| 21 | 4.615.076 |  | A | 14,86 |  |
| 22 | 4.615.076 |  | A | 14,985 |  |
|  |  |  |  |  |  |
|  | Anzahl | Dauer |  |  |  |
|  | 1.561.660 | 1,24 |  |  |  |
|  | 2.307.538 | 4,53 |  |  |  |
|  | 3.869.198 | 11,52 |  |  |  |
|  | 4.615.076 | 15,2 |  |  |  |

|  |  |
| --- | --- |
| Query ID | Query |
| A | select avg(power) as power, day, month, meterId from redd group by  meterId asc, month asc, day asc; |
| B | update redd set datetime=FROM\_UNIXTIME(timestamp)  where meterId is null; |
| C | update redd set day=day(datetime), month=month(datetime),  meterId = 3 where meterId is null; |

1. Intel Core i5-4690K @ 3.5Ghz [↑](#footnote-ref-1)