

Swissgrid Data Analysis

Prepared by

K.S.N.R Ranasinghe
139177V

Faculty of Information Technology
University of Moratuwa, Sri Lanka

May 2017

Swissgrid Data Analysis

Prepared by

K.S.N.R Ranasinghe

139177V

Thesis submitted to the Department of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree Master of Science in Information Technology.

May 2017

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education.

Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student : K.S.N.R Ranasinghe

Student Number : 139177V

Student Signature: Date:

Name of Supervisor(s) : Dr. Lochandaka Ranathunga

Signature of Supervisor(s):..... Date:

Dedication

This Dissertation is dedicated to my loving husband, my loving mother and my darling son for always encouraging me and being by my side.

Acknowledgement

This project would not have been a success if not for the many people in my life. Their encouragement and guidance has helped me immensely to complete the project. I take this opportunity to express my gratitude to the people who have been instrumental in the success of the completion of this research.

My heartfelt gratitude goes to my supervisor Dr.Lochandaka Ranatunga not only for his guidance and mentoring to complete the project but also for the encouragement given all throughout the project.

I would also like to thank Prof.Asoka Karunananda for the valuable knowledge he gave us on writing a Thesis, which helped me in great scale to compose this thesis.

It is with heart felt gratitude that I would like to thank all the other Senior lecturers, Lecturers, Instructors, and staff members who helped me in many numerous ways to make this Project a success.

I would specially like to thank Ms. Dinesha Malwalage, Mrs.Chaya Atapattu and Mr.Kasun Atapattu for the encouragement they have given me throughout.

I would also like to thank all my other my M.Sc Information Technology Batch 07 batch mates for all the help given.

Last but not least, the research would have not been possible if not for my family. I would like to thank them for being understanding and for the encouragement they have given me throughout and for being the backbone of the success of this research.

Abstract

The swissgrid is the national power grid in Switzerland which is the largest in that area. Not only does it supply power to Switzerland it also exports and imports power from its neighboring countries. The power grid must be kept at a balance of 50Hz frequency. In order to help the operators maintain and take necessary action to maintain this frequency, monitoring the grid is vital. Currently, studies do not clearly show of any prediction models that the swissgrid uses. Hence this study is focused on assisting the operators monitor the grid and help them predict the energy consumption for the swiss control block.

In order to assist the operators, and interested parties, the grid data has been analyzed in order to derive real-time and batch analytics using the WSO2 Data Analytics Server. The real-time analytics computed based on the Siddi engine and batch analytics based on the Apache Spark engine is able to be viewed on a central dashboard powered by WSO2 Data Analytics Server. Moreover the solution also provides the ability to configure to detect any anomalies in the power grid and alarm any interested parties via SMS or E-mail.

The research goes on to find a model to predict the total energy consumption of the swiss control block. A model was successfully built in order to predict the energy consumption using the Linear regression with rolling window analysis. Using a 30 day window it was found that the model's optimal training data set is of 1.5 years worth data. The research expanded to find any co-relation between multiple factors that would affect the energy consumption. Using the season, and whether or not the date is a holiday, a model was built based on the multiple regression algorithm. This model was found to be trained better with the least Absolute Mean average for a data set of 2 years of data. The models built were tested against the predicting data set which proved to have predict the energy consumption easily.

The aim of the project is achieved by providing centrally viewable statistics and a tested model to predict the total energy consumption for the swiss control block.

Contents

Chapter 1 - Introduction	1
1.1 Background & Motivation	1
1.2 . Aim and Objectives	4
1.2. 1 Aim	4
1.2.2 Objectives	4
1.3 Tools and technology	5
1.4 Structure of the thesis	5
Chapter 2 - Literature Review	7
2.1 Introduction	7
2.2 Other's work	7
2.3 Summary	13
Chapter 3 - Using technology to analyze the Swissgrid Data	14
3.1 Introduction	14
3.2 WSO2 DAS for Real time and Batch Analytics	14
3.3 Java for preprocessing data and publishing events	16
3.4 SciKit Learn for building the ML model to predict the Consumption	16
3.5 Using the rolling window analysis of time series with Linear regression	16
3.6 Summary	17
Chapter 4 - Turning Swissgrid Data into Information	18
4.1 Introduction	18
4.2 System Design & Development	18
4.3 Language for Implementation	18
4.4 Apache Cassandra No SQL DB	18

4.5 Writing a Java client to publish events to WSO2 DAS	19
4.5 Real-time analysis using WSO2 Data Analytics Server	19
4.5.1 Reading the Data	20
4.5.2 Persisting the Data	20
4.5.3 Computing the statistics	20
4.5.4 Creating Alerts	21
4.6 Batch Analysis using WSO2 Data Analytics Server	22
4.6.1 Using Apache Spark for Batch Analytics	22
4.7 Displaying the statistics in the WSO2 DAS dashboard	23
4.8 Forecasting the end user consumption	23
4.8.1 Preprocessing the data	23
4.8.2 Using SciKit Learn to build the model	24
4.9 Summary	24
Chapter 5 - System Design	26
5.1 Introduction	26
5.2 High level design	26
5.3 WSO2 DAS high level architecture	27
5.4 The Event flow	28
5.5 The Machine Learning model design	29
5.5.1 Absolute Mean Error (MAE)	30
5.6 Summary	31
Chapter 6 - Implementation	32
6.1 Introduction	32
6.2 Real-time Analytics	32
6.2.1 Publishing data to the WSO2 DAS	32
6.2.2 Configuring the event stream	32
6.2.3 Writing the Execution Plan	33
6.2.4 Writing the output stream	34
6.3 Batch Analytics	35

6.3.1 Persisting the events to a data store	35
6.3.2 Using Spark SQL to Query the database	35
6.4 Publishing and viewing data in the WSO2 dashboard	36
6.5 Implementing the ML model	36
6.5.1 Reading the data for analysis	37
6.5.2 Using Linear Regression	37
6.5.3 Calculating the Absoluter Error Mean	38
6.5.4 Using rolling window analysis	38
6.6 Summary	38
Chapter 7 - Evaluation	39
7.1 Introduction	39
7.2 Evaluating the Preprocessed data	39
7.3 Evaluating the Machine Learning Model	39
7.3.1 Rolling Window test results.	39
7.3.2 Multiple Feature Test	45
7.4 Functional testing - Real Time and Batch analytics	49
Chapter 8 - Conclusion	53
8.1 Introduction	53
8.2 Analysis of the size of the data set Size.	53
8.2.1 Rolling window analysis with Linear Regression for one factor	53
8.2.2 Multiple Linear Regression analysis	54
8.3 Retrospect	54
8.4 Limitations	55
8.5 Future Work	55
8.6 Summary	56
References	57
<i>Appendix B - Screen shots of the WSO2 DAS configuration</i>	65

Persisting the Event	65
Creating the Spark SQL script	66
Appendix C - Data preprocessor	67
<i>Appendix D – Source for building the model to predict energy consumption – with one factor</i>	71
Appendix E – Source code for building the prediction model with multiple regression	73

List of Figures

Figure 1.1 - Swiss Electricity Market in 2010	2
Figure 2.1 - Wide Area Monitoring High Level Architecture	10
Figure 2.2 - Monthly End User Consumption	11
Figure 2.3 - Monthly Energy Consumption	11
Figure 3.1 - Using WSO2 DAS for Real time and Batch Analytics	14
Figure 3.2 - Linear Regression	17
Figure 4.1 - Real time event flow	20
Figure 4.2 - Spark Components	22
Figure 5.1 - High level design of the proposed system	26
Figure 5.2 - WSO2 DAS high level architecture	27
Figure 5.3 - Event Flow	28
Figure 5.4 - Rolling Window Analysis	30
Figure 7.1 - 1 year data training set results	40
Figure 7.2 - 1.5 years training set results	41
Figure 7.3 - 2 years data training set results	42
Figure 7.4 - 2.5 years data training set results	43
Figure 7.5 - 3 years training data set results	44
Figure 7.6 - Training data set results of 1 year	46
Figure 7.7 - 1.5 years training data set results	47
Figure 7.8 - 2 years training data set results	47
Figure 7.9 - 2.5 years training data set results	48
Figure 7.10 - 3 years training data set results	49
Figure 8.1 - Rolling windows data set analysis	53
Figure 8.2 - Multiple regression data set analysis	54

List of Tables

Table 1.1 - Swissgrid facts	2
Table 2.1 - Summary of previous studies	12
Table 7.1 - Functional test case1	50
Table 7.2 - Functional test case2	51
Table 7.3 - Functional test case3	52