

Master Thesis

Develop and build a Data Warehouse for structured, semi-structured and unstructured data

Keywords: Data Science, Data Warehouse, Data Pipeline, ETL, Data Integration

This topic is formulated in accordance to Data Management phase in *SmartHelm* project. As part of the project, we obtain different categories of data such as Order Management data (structured data), Weather data, Geo Information data (semi-structured data) and EEG sensor data (unstructured data). The heterogeneous data in different formats (e.g. CSV, XML, JSON) should be first extracted from the source and transformed into a uniform format and second loaded in the Data Warehouse (DWH) by using various data processing tools. Also, suitable e.g. ETL, ELT methods should be chosen in the best suitable way for in-house data storage. Data Integration plays a key role before storing the data in the DWH and should be also considered.

Aim: The goal of this research will be first to find out suitable ETL or ELT approaches, DWH techniques for storing structured, semi-structured and unstructured data as well as the use of Data Integration techniques and second to implement the findings as a DWH. Summary of topics and procedures:

- Data Warehousing
- ETL, ELT
- Data Integration
- Scheme, Model etc. design
- Storing the data from SmartHelm project and prepare a data catalogue

The developed artifact will be used as a data management tool within the project framework to store the collected data from various sources.

Note: The thesis can be written in German or English. There are no restrictions.

Falls Sie Interesse an einer Bearbeitung des Themas haben, melden Sie sich bitte bei:

M. Sc. Harish Moturu (harish.moturu@uni-oldenburg.de)

M. Sc. Christian Janßen (christian.janssen@uni-oldenburg.de)

DEPARTMENT FÜR INFORMATIK

ABTEILUNG
WIRTSCHAFTSINFORMATIK I
VERY LARGE BUSINESS
APPLICATIONS

PROF. DR. JORGE MARX GÓMEZ

MITARBEITER
M.SC. CHRISTIAN JANßEN

TELEFONDURCHWAHL
(0441) 7 98 – 4494

FAX
(0441) 7 98 – 4472

EMAIL
christian.janssen@uni-oldenburg.de

GEBÄUDE A4
Uhlhornsweg 84 – Raum A4 3-318

OLDENBURG
30.06.2020



VERY LARGE
BUSINESS
APPLICATIONS

Carl von Ossietzky
Universität Oldenburg

POSTANSCHRIFT
D-26111 Oldenburg

PAKETANSCHRIFT
Ammerländer Heerstraße 114 - 118
D-26129 Oldenburg

TELEFONZENTRALE
(0441) 7 98 – 0

BANKVERBINDUNG
Landessparkasse zu Oldenburg
Kto. Nr.: 1 988 112
BLZ: 280 501 00
BIC: BRLADE21LZO