CARL VON OSSIETZKY UNIVERSITÄT OLDENBURG
FAKULTÄT II
INFORMATIK, WIRTSCHAFTS- UND RECHTSWISSENSCHAFTEN

CARL VON OSSIETZKY UNIVERSITÄT OLDENBURG · 26111 OLDENBURG

DEPARTMENT FÜR INFORMATIK

ABTEILUNG
WIRTSCHAFTSINFORMATIK I
VERY LARGE BUSINESS APPLICATIONS

PROF. DR. JORGE MARX GÓMEZ

MITARBEITER
M.SC. HARISH MOTURU

TELEFONDURCHWAHL
(0441) 7 98 – 4784

FAX
(0441) 7 98 – 4472

EMAIL
harish.moturu@uni-oldenburg.de

GEBÄUDE A4
Uhlhornsweg 84 – Raum A4 3-318

OLDENBURG
01.04.2021

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
IBAN: DE 4628 0501 0000 0198 8112

Masterarbeitsthema Ausschreibung

Design and build a Data Warehouse methodology to store and integrate data from hetrogeneous data sources.

Keywords: Schema design techiniques, Data Pipeline, ETL, Data Integration.
Description: 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, navigation data (Structured data), weather data, Geo Information Data (semi-structured data category) and EEG sensor data (unstructured category). The heterogeneous data is transferred through various protcols such as REST API, general file transfer, web API services etc., now these transfered data sources obtained in various formats such as (.csv, JSON, xdf etc.,) should be extracted from the sources and transformed using various data processing tools into a uniform format then finally loaded to store in Data Warehouse (DWH). In principle through literature study, the state-of-the-art ETL should be chosen in way that it best suitable for in-house data storage rather that cloud based ETL tools. Data storage is completely done in-house database.
Data Integration plays a key role before storing the data in the main Data Storage (DWH), because stored data must be productive to utilize it for implementing data analysis as well data evaluation techniques. Therefore, within this Thesis there is scope to research the concept of Data warehousing and ETL methods in depth. In addition, can practically implement the best suitable methods for our Data Requirements and goals.

AIM: To find out and implement the suitable ETL approaches, data warehousing techniques for storing structured, Semi-structured and unstructured Data in an in-house Data Warehouse with the following aspects.

- Study about the available data sources
- design the schema
- develop data source Connectors
- deploy ETL process
- Data Integration.
- Build a model
- Prepare Data Catalogue.

The main goal from the Thesis is to build a scalable Data storage system, which can be set into application on the Data collected in the Project from various sources.
Note: The Thesis can be written either in German or English. No restrictions.
References:

1. Sahiet, Dumin, and P. D. Asanka. "ETL framework design for NoSQL databases in dataware housing." *Int. J. Res. Comput. Appl. Robot.* 3 (2015): 67-75.
2. Choudhary, Ravi Kumar. "Key organizational factors in data warehouse architecture selection." *Vivekananda Journal of Research* 1.1 (2012): 24-32.

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

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