Python - Introduction to ETL

Course code: PYTHON_ETL

ETL stands for Extract-Transform-Load - database functions that are integrated into one tool that extracts data from one database and places it in another database. Extract is the process of reading data from a database. Transform is the process of converting the acquired data from its original form to the form in which it is expected to be placed in another database. Load is the process of writing data to the target database. ETL is used to migrate data from one database to another to create data markets (DataMart) and data warehouses (DataWarehouse) and also to convert a database from one format or type to another.The course is intended for every adept but also for an experienced data engineer who is interested in using the Python programming language for data acquisition, processing and preparation in the future. The training is conducted extremely intensively, with practical examples from the real world, and therefore knowledge of the Python programming language is recommended at least at intermediate and higher level (PYTHON_INTRO and PYTHON_ADV) and also at least at intermediate level knowledge of relational databases.

Requirements per participant
- Knowledge of Python programming at the PYTHON_INTRO and PYTHON_ADV course level

Teaching methods
- Expert explanation with practical examples, exercises on computers.

Study materials
- Online presentation of the subject matter and exercises.

Course syllabus
- Introduction (Python, ETL, Data Engineering)
- Data sources (CSV, JSON, RESTApi, MySQL, Postgresql, Oracle 12c, NoSQL)
- Data targets (WareHouse, DataMart)
- Pandas
- ETLAlchemy
- Bonobo
- Luigi
- Airflow
- Talend
- Distributed systems (SQOOP, PIG and SPARK / Impala vs Presto /, HIVE)
- Conclusion