
Business Intelligence

Praktikum 1: Der ETL-Prozess

Wintersemester 2025/2026

Prof. Dr. Arnim Malcherek

Prof. Dr. Susanne Hoffmeister

Lernziele:

- Sie sind in der Lage, für einen einfachen Usecase ein Data Warehouse in 3. Normalform zu modellieren und mit einer gängigen Datenbanksoftware umzusetzen.
- Sie sind in der Lage, mit dem Werkzeug Bonobo einen einfachen ETL-Prozess in ein Data Warehouse zu gestalten.

Vorarbeiten: Installieren Sie sich die Software Bonobo über `pip install bonobo` und die MySQL Workbench. Alternativ zur MySQL Workbench verwenden Sie den PostgreSQL Datenbankserver des Fachbereichs (siehe Moodle für Nutzungsdetails).

Inhaltlicher Rahmen:

Wir verwenden die fiktive Firma Global Bike aus dem Schulungsmaterial der SAP SE als inhaltlichen Rahmen. Sie finden eine kurze Beschreibung zur Firma auf Moodle. Dort finden Sie auch die zugehörigen Daten in Form eines einzelnen (UTF-8 kodierten) csv-Files.

Aufgabe 1

- (a) Verschaffen Sie sich mit Hilfe der Kurzpräsentation auf Moodle einen Überblick über die Firma Global Bike.
- (b) Machen Sie sich mit der Struktur der Daten auf Moodle vertraut, und erstellen Sie anschließend ein ER-Modell für die Daten für ein einfaches Data Warehouse in 3. Normalform. Achten Sie dabei auch auf sinnvolle Fremdschlüsselbeziehungen.
Als Software für die Modellierung stehen zur Auswahl:

- mySQL Workbench auf Ihrem Laptop (Download unter <https://www.mysql.com/products/workbench/>)
- SAP Power Designer auf Horizon Client

Sie können auch andere Software verwenden (z.B. Microsoft Visio), sofern ein lesbares, sinnvolles ER-Diagramm entsteht.

Laden Sie die ER-Diagramme in Ihrem Gruppenfolder auf Moodle hoch.

Aufgabe 2

- (a) Implementieren Sie das Modell mit Hilfe der MySQL-Workbench (z.B. über Menü: *Database -> Forward engineer* oder manuell) oder PostgreSQL in einem eigenen Datenbankschema.
- (b) Machen Sie sich mit der Software Bonobo vertraut (siehe Tutorials und Doku auf <https://www.bonobo-project.org>).
- (c) Laden Sie die Daten aus den csv-Dateien mit Hilfe eines Bonobo ETL-Prozesses in Ihr Data Warehouse. Für die Datenbankintegration können Sie SQLAlchemy nutzen. Es ist aber nicht zwingend notwendig.
- (d) In den Daten sind einige Fehler vorhanden. Versuchen Sie, diese mit Hilfe Ihres ETL-Prozesses zu finden und gegebenenfalls automatisch (falls möglich) oder manuell zu beheben.
- (e) Laden Sie Ihren Bonobo Code, einen SQL-Dump Ihres Data Warehouses und eine Liste der gefundenen fehlerhaften Datensätze in Ihrem Moodle Folder hoch.