

PERSONAL_ID	NAME	Geselle	Lehrling
1	Mark	null	Verkaeuffer
2	Pete	Lagerist	null

Table 1: Zieldarstellung der Daten

1. Welche Datenfehler können bestehen und wie kann man sie finden? Welche Fehler bleiben bei welchem Verfahren unberücksichtigt?
2. Ermitteln von Ausreißer-Werten: Gegeben seien die Datenwerte aus Tabelle 3. Ermitteln Sie die Ausgleichsgerade für die gegebenen Werte. Wie können anhand der ermittelten Geraden Ausreisser bestimmt werden?
3. Erweiterung der Programmieraufgabe zur Schema-Integration. Diese Aufgabe ist Voraussetzung für den Schein. Implementieren Sie ein *Java* oder *C++*-Programm!

In dieser Aufgabe soll das Wissen über Schema-Integration vertieft werden.

Führen Sie das prepare SQL-Skript in der Datenbank aus.¹ Überführen Sie die Daten der ersten beiden Tabellen (**Bierladen1**, **Bierladen2**) in das Schema der dritten Tabelle **IntegratedBierladen**. Verlassen Sie sich nicht nur auf SQL, sondern schreiben Sie ein JAVA oder C++ Programm, welches die Transformation der Daten vornimmt.

Als Hilfestellung ist auf der Webseite eine kleine Beispielanwendung "HelloJDBC.java" gegeben. Für seine Anwendung muss der aktuelle Oracle-Treiber (ojdbc14-1.jar) in die Variable "CLASSPATH" aufgenommen werden.²

Die **target**-Tabelle und die mit ihr verbundene Datendarstellung ist in Tabelle 1 dargestellt.

Die Integration von Schemata kann für die eingefügten Daten von Mark und Pete Duplikate verursachen. Fügen Sie zu dem integrierten Schema den Datensatz mit der Personal_ID '5' hinzu (siehe Tabelle 2). Implementieren Sie einen Ansatz, um die Duplikate zu erkennen und zu löschen. (Der Datensatz der Personal_ID '5' solle als Duplikat erkannt und gelöscht werden.)

Diskutieren Sie ihre Ergebnisse. Kann dieser Ansatz parametrisiert werden?

¹http://www.witi.cs.uni-magdeburg.de/iti_db/lehre/dw/dwt1112/ueb04/ueb04_prepare.sql

²Von der Vorlesungsseite herunterzuladen.

PERSONAL_ID	NAME	Geselle	Lehrling
1	Mark	null	Verkaefer
2	Pete	Lagerist	null
5	Pete	Lagerist	null

Table 2: Zieltabelle

X	Y
1	1
2	10
3	6
4	4
5	3

Table 3: Regressionswerte