



Übung zur Vorlesung *Einsatz und Realisierung von Datenbanken im SoSe25*

Alice Rey, Maximilian Reif, Tobias Goetz (i3erdb@in.tum.de)

<http://db.in.tum.de/teaching/ss25/impldb/>

Blatt Nr. 13

Hausaufgabe 1

Führen Sie die folgenden Abfragen in der Spark-Shell aus. Als Grundlage für die Abfragen dient das TPC-H Schema. Laden Sie dazu die TPC-H Daten wie in der Vorlesung gezeigt in die Spark-Shell.

- (a) Laden Sie die `region.tbl` Datei als DataFrame Objekt in die Spark-Shell.
- (b) Ermitteln Sie die Namen aller Regionen.
- (c) Ermitteln Sie die Zahl der Länder die nicht in Europa liegen.
- (d) Ermitteln Sie die größte Bestellung aus dem Jahr 1996.
- (e) Ermitteln Sie welcher europäische Kunde im Jahr 1996 am meisten Geld ausgegeben hat.
- (f) Ermitteln Sie welche Unternehmen keine Kunden in Europa haben.

Hausaufgabe 2

Führen Sie die folgenden Abfragen in der Spark-Shell aus. Als Grundlage für die Abfragen dient das TPC-H Schema. Laden Sie dazu die TPC-H Daten wie in der Vorlesung gezeigt in die Spark-Shell.

- (a) Ermitteln Sie pro Marktsegment die Anzahl der Bestellungen in 1997.
- (b) Ermitteln Sie die Zahl der Kunden und Lieferanten pro Land.
- (c) Ermitteln Sie die Stückzahlen der verschiedenen Bauteile in Deutschland.
- (d) Ermitteln Sie, welche Kunden kein *goldenrod lavender spring chocolate lace* bestellt haben.