



Sparkübungsblatt

Aufgabe 1 (TPC-H in der Spark-Shell)

Führen Sie die Abfragen in der Spark-Shell aus. Als Grundlage für die Abfragen dient das TPC-H Schema.

1. Laden Sie die `region.tbl` Datei als DataFrame Objekt in die Spark-Shell.
2. Ermitteln Sie die Namen aller Regionen.
3. Ermitteln Sie die Zahl der Länder die nicht in Europa liegen.
4. Ermitteln Sie die größte Bestellung aus dem Jahr 1996.
5. Ermitteln Sie welcher europäische Kunde im Jahr 1996 am meisten Geld ausgegeben hat.
6. Ermitteln Sie welche Unternehmen keine Kunden in Europa haben.