



Übung zur Vorlesung *Grundlagen: Datenbanken* im WS20/21

Christoph Anneser, Josef Schmeißer, Moritz Sichert, Lukas Vogel (gdb@in.tum.de)

<https://db.in.tum.de/teaching/ws2021/grundlagen/>

Blatt Nr. Z3

In den folgenden Zusatzblättern wird der gesamte Prozess von der Modellierung eines Datenbankschemas, der Extraktion und Bereinigung der Daten, sowie der Erstellung eines geeigneten Datenbankschemas mittels SQL-DDL Statements und dem anschließenden Import der Daten in die Datenbank anhand des Beispiels der Fußball-Weltmeisterschaften der Jahre 1930 bis einschließlich 2014 erklärt.

Alle in diesem Blatt behandelten datenbankspezifischen Aspekte sind selbstverständlicherweise auch *klausurrelevant*.

Hausaufgabe 1

Nach der ER-Modellierung und der Erstellung des Datenbankschemas folgt nun die ETL-Phase. ETL steht hierbei für *Extraction, Transform* und *Load*. In diesem häufig auch als Parsen bezeichneten Schritt extrahieren wir beispielsweise aus der denormalisierten CSV-Datei `WorldCupPlayers.csv` die eindeutigen Spielernamen, Startaufstellungen, Trainer und Events. Die im ER-Modell berücksichtigten Abhängigkeiten (z.B. sind Events eindeutig einem Spieler und einem Spiel zugeordnet) müssen während des Parsens erhalten bleiben.

Nach der Extraction- und Transform-Phase (die wir bereits für Sie übernommen haben) liegen nun die normalisierten Daten vor, die Sie von Moodle unter <https://www.moodle.tum.de/mod/resource/view.php?id=1438092> herunterladen können. Entpacken Sie die in der zip-Datei enthaltenen `tbl`- und `sql`-Dateien. Für jede in unserem Schema vorhandene Relation existiert eine entsprechende `tbl`-Datei, die die normalisierten Daten enthält. Die mitgelieferte `sql`-Datei enthält alle Befehle, um die Relationen anzulegen und mittels `\Copy`-Anweisungen die Daten in die Datenbank einzufügen (Load-Phase).

Legen Sie nun eine neue Datenbank in Postgres an:

```
<nutzernamen>@maschine:$ sudo -u postgres psql
postgres=# CREATE DATABASE worldcups;
postgres=# GRANT ALL ON DATABASE worldcups to <nutzernamen>;
```

Die `psql`-Umgebung kann mit `\q` wieder verlassen werden. Wechseln Sie nun in das Verzeichnis mit den entpackten `tbl`- und `sql`-Dateien und starten Sie den Import der Daten mit folgendem Befehl:

```
<nutzernamen>@maschine:$ psql worldcups < load.sql
```

Sie können nun auf der Datenbank Anfragen ausführen.

Zum Einstieg die folgenden Fragestellungen:

1. Welche Spieler haben die meisten Tore erzielt? Zusatzaufgabe: Geben Sie auch den Zeitraum an, in dem die Tore erzielt worden sind.

2. Geben Sie für jedes Team die drei Spieler mit den meisten erzielten WM-Toren aus (sofern überhaupt drei Spieler existieren).
3. Geben Sie die ewige Torschützenliste der deutschen Nationalmannschaft an. Beachten Sie, dass es vor der Wiedervereinigung 1990 andere Team-Identifizierer für Ost- und Westdeutschland gab.
4. Geben Sie für jede Weltmeisterschaft den Sieger, den Zweitplatzierten und den Drittplatzierten aus.
5. Geben Sie für jede Nation an, wie viele Weltmeisterschaften sie gewonnen hat und sortieren Sie absteigend nach Anzahl der gewonnenen Titel.
6. Welcher Torwart hat die weißeste Weste von allen? Die Weste eines Torwarts wird mit jedem gegentorlosen Spiel, in dem er in der Startaufstellung stand, weißer.
7. Welcher Spieler hat in den meisten direkt aufeinanderfolgenden Spielen seines Teams in der Startaufstellung gestanden. Erstellen Sie ein `prepared`-Statement, das diesen Spieler für ein bestimmtes Team ermitteln soll. Ihr Statement soll den Teamnamen als Parameter verwenden.
8. Geben Sie für die WM 2014 an, wie viele Einsatzminuten jeder einzelne Spieler auf dem Platz stand.
9. Geben Sie für jeden Spieler an, wie viele Weltmeisterschaften er gewonnen hat.
10. Torquote: Welcher Spieler braucht wie viele Spiele für ein Tor? Sortieren Sie aufsteigend nach der Torquote und geben Sie die Zahl der Tore und Einsatzminuten an.

Hinweise:

- Die WM 2018 ist – aus deutscher Sicht auch aus guten Gründen – nicht verfügbar.
- Die bereitgestellten Daten enthalten zum Teil nur Nachnamen oder unbekannte Symbole – Anfragen können daher eventuell nicht korrekte Ergebnisse liefern.