

Fallstudie: Datenintegration und Entwicklung einer MongoDB-Datenbank

Die Professorenschaft bittet Sie und Ihr Team, die Datenbank für das "Professoren-Informationssystem" prof-IS zu überarbeiten.

Aufgabe 1: Synchronisation des Datenmodells der prof-IS-Datenbank mit dem Data Dictionary (60 Minuten)

Datenbank laden

Die Datenbank liegt in Form einer MySQL-Datenbank vor. Um die Datenbank zu importieren, führen Sie folgende Schritte aus:

1. Legen Sie sich eine Datenbank mit dem Namen profis an.
2. Laden Sie die Datei in ein SQL Fenster in der Workbench und führen Sie das Skript aus:
`dump_latin1_german2_ci.sql`; Hinweis: diese wurde mit folgendem Befehl erstellt: `mysqldump --default-character-set=latin1 -u root -p profis > dump_latin1_german2_ci.sql`

Data Dictionary überarbeiten und MySQL-Datenbank anpassen

Die Professoren haben ein **Data Dictionary** für die Datenbank entwickelt und erwarten als erstes, dass Sie das Data Dictionary mit der Datenbank abgleichen.

Ziel ist es,

- ein "möglichst" **sauberes Ziel-Datenmodell** zu erzeugen,
- die bestehende Datenbank inklusive ihrer Daten in das Ziel-Datenmodell zu laden bzw. das Datenbankmodell anzupassen.
- Im Zweifelsfall sollen die Definitionen des Data Dictionaries gelten.

Diskutieren Sie im Team die Datentypen des Datenmodells und schlagen Sie ggf. Verbesserungen vor. Setzen Sie die Verbesserungen um.

Erweitern / ändern Sie das Data Dictionary oder passen Sie dieses an Ihre Vorstellungen an.

Aufgabe 2: Datenimport - dem Wildwuchs ein Ende

Die Professoren haben bisher ihre Daten in unterschiedlichen Formaten gespeichert:

- Einige der Professoren verwenden eine App, die die Daten als JSON-Dokumente ablegt.
- Eine andere Gruppe von Professoren verwendet Excel als Programm zur Verwaltung der Daten.
- Wiederum andere arbeiten mit Linux und speichern die Daten in CSV-Formaten.

Die Quellen liegen also als MySQL-Datenbank (diese haben Sie schon geladen), Excel, CSV und in Form von JSON vor:

- [CSV-Datei](#)
- [JSON-Datei](#)

Ziel ist es,

- die Daten in die von Ihnen verbesserte MySQL-Datenbank zu laden,
- die Daten vor dem Laden zu bereinigen

1. Überlegen Sie sich in Ihrem Team ein **Import-Strategie** für die Daten (Reihenfolge, Tools, ...)

a) Welche Integrationsplattform möchten Sie verwenden?

- Excel als Integrationsplattform mit Power Query
- Programmierung in Python
- VBA
- ...

b) Was müssen Sie beim Import beachten?

- Welche Fehler müssen Sie beim Import bereinigen und welche Arten von Fehlern können auftreten?
- Wie gehen Sie mit diesen Fehlern um?

Erstellen Sie eine Liste mit Fehlern, die wie folgt aufgebaut ist:

| Fehlernummer | Fehlerbeschreibung | Bereinigungsmöglichkeiten | Fehlerkategorie |
|--------------|-------------------------------|---------------------------|-----------------|
| F-1 | Falscher Datentyp in Spalte x | Konvertierung | Datenmodell |

Hinweis: Sie werden u.a. Fehler in den Namen der Studierenden finden.

Aufgabe 3: Transfer in MongoDB

Entwickeln Sie basierend auf dem MySQL-Datenmodell für das Professoren_Informationssystem eine MongoDB-Datenbank. Verwenden Sie dazu den MongoDB Relational Migrator (separates Tool).

Überlegen Sie sich sinnvolle Dokumentstrukturen zur Speicherung der Daten im JSON-Format.

Ein Erklärvideo finden Sie auf YouTube: [How to Migrate from Relational Databases to MongoDB | Relational Migrator 101](#)