

Name:

Klasse:

Datum:

Schuljahr: 2022/23

Daten importieren

Oft ist es notwendig, Daten von anderen Quellen zu importieren. Hat man einen großen Kundenkreis, ist es sehr aufwendig, jedes Mal bei einem neuen Kunden die PLZ bzw. den Ort einzugeben. Auf der Seite www.suche-postleitzahl.org kann man sich die Daten von Deutschland herunterladen. Die Daten stehen unter der Open Database Licence frei zur Verfügung.

```
osm_id,ags,ort,plz,landkreis,bundesland
1104550,08335001,Aach,78267,Landkreis Konstanz,Baden-Württemberg
1255910,07235001,Aach,54298,Landkreis Trier-Saarburg,Rheinland-Pfalz
62564,05334002,Aachen,52062,Städteregion Aachen,Nordrhein-Westfalen
62564,05334002,Aachen,52064,Städteregion Aachen,Nordrhein-Westfalen
62564,05334002,Aachen,52066,Städteregion Aachen,Nordrhein-Westfalen
62564,05334002,Aachen,52068,Städteregion Aachen,Nordrhein-Westfalen
62564,05334002,Aachen,52070,Städteregion Aachen,Nordrhein-Westfalen
```

Das Format entspricht nicht unserem Datenschema - es muss also schrittweise angepasst werden. Für die weiteren Schritte verwenden wir die Datenbank gml mit den schon vorhandenen Tabellen, u. a. die Tabelle ort!

```
MariaDB [gml]> show tables;
+-----+
| Tables_in_gml |
+-----+
| abteilung      |
| gehalt         |
| mitarbeiter    |
| ort            |
+-----+
```

1. Temporäre Tabelle für die CSV-Daten anlegen.

Die Struktur entspricht dabei genau der CSV-Datei. Die Spaltenbreite wurde hier großzügig bemessen. Entspricht die Spaltenreihenfolge der Reihenfolge der CSV-Datei, so muss beim Import bzgl. der Spalten nichts zusätzlich angegeben werden. Für die Tabelle ist kein PK notwendig, da diese dann später sequentiell komplett gelesen wird (full table scan).

```
CREATE TABLE csv_ort (
  osm_id    VARCHAR(20),
  ags      VARCHAR(20),
  ort      VARCHAR(50),
  plz      VARCHAR(10),
  landkreis VARCHAR(50),
  bundesland VARCHAR(50)
) ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

```
MariaDB [gml]> desc csv_ort;
+-----+-----+-----+-----+-----+-----+
| Field      | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| osm_id     | varchar(20)   | YES  |     | NULL    |       |
| ags        | varchar(20)   | YES  |     | NULL    |       |
| ort        | varchar(50)   | YES  |     | NULL    |       |
| plz        | varchar(10)   | YES  |     | NULL    |       |
| landkreis  | varchar(50)   | YES  |     | NULL    |       |
| bundesland | varchar(50)   | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
```



2. **Datenimport** (für Text-Dateien) wird über den Befehl `LOAD DATA INFILE` realisiert. Der Verzeichniseintrag, in dem die CSV-Datei abgespeichert worden ist, muss natürlich entsprechend angepasst werden!

```
LOAD DATA LOCAL INFILE 'src/csv/zuordnung_plz_ort_landkreis.csv'  
  INTO TABLE csv_ort  
  CHARACTER SET utf8  
  FIELDS TERMINATED BY ','  
  LINES TERMINATED BY '\n'  
  IGNORE 1 ROWS;
```

Die Ausgabe ist dabei:

```
Query OK, 12937 rows affected (0.212 sec)  
Records: 12937 Deleted: 0 Skipped: 0 Warnings: 0
```

Achtung:

Bei einigen Distributionen ist der Befehl `LOAD DATA LOCAL INFILE` aus Sicherheitsgründen deaktiviert.

Wenn dies bei Ihnen der Fall ist, so ändern Sie die Einstellungen. Siehe hierzu `Server prints bad error message when user executes LOAD DATA LOCAL and it is disabled`

```
The used command is not allowed with this MariaDB version
```

oder

```
The used command is not allowed because the MariaDB server or client has disabled the local infile capability
```

Die Tabelle hat folgenden Inhalt:

```
select ort,plz,landkreis,bundesland from csv_ort limit 4;
```

ort	plz	landkreis	bundesland
Aach	78267	Landkreis Konstanz	Baden-Württemberg
Aach	54298	Landkreis Trier-Saarburg	Rheinland-Pfalz
Aachen	52062	Städteregion Aachen	Nordrhein-Westfalen
Aachen	52064	Städteregion Aachen	Nordrhein-Westfalen

Da die Tabelle fast 13.000 Datensätze enthält, ist es sinnvoll, auf die Spalten `ort`, `landkreis` und `bundesland` einen Index zu legen. Der Index sorgt dafür, dass die Spalte über den Index sortiert ist. Somit wird der Zugriff deutlich schneller, wenn man z. B. nur alle Landkreise anzeigen lassen will, die in Bayern sind. Der Nachteil, der Index benötigt zusätzlich Speicher.

Informieren Sie sich über den Befehl `CREATE INDEX`.

Notiz:



Index für die Spalten anlegen:

```
CREATE OR REPLACE INDEX csv_ort_bundesland ON csv_ort (bundesland);
CREATE OR REPLACE INDEX csv_ort_landkreis ON csv_ort (landkreis);
CREATE OR REPLACE INDEX csv_ort_ort ON csv_ort (ort);
```

Mit dem Befehl `SHOW INDEX` werden die Indexe zu einer Tabelle angezeigt.

Der Anhang `\G` sorgt dafür, dass die Ausgabe vertikal erfolgt.

Wird eine Tabelle gelöscht, werden automatisch alle Indexe gelöscht.

```
MariaDB [gm1]> show index from csv_ort\G
***** 1. row *****
      Table: csv_ort
      Non_unique: 1
      Key_name: csv_ort_bundesland
      Seq_in_index: 1
      Column_name: bundesland
      Collation: A
      Cardinality: 32
      Sub_part: NULL
      Packed: NULL
      Null: YES
      Index_type: BTREE
      Comment:
Index_comment:
***** 2. row *****
      Table: csv_ort
      Non_unique: 1
      Key_name: csv_ort_landkreis
      Seq_in_index: 1
      Column_name: landkreis
      Collation: A
      Cardinality: 596
      Sub_part: NULL
      Packed: NULL
      Null: YES
      Index_type: BTREE
      Comment:
Index_comment:
***** 3. row *****
      Table: csv_ort
      Non_unique: 1
      Key_name: csv_ort_ort
      Seq_in_index: 1
      Column_name: ort
      Collation: A
      Cardinality: 13132
      Sub_part: NULL
      Packed: NULL
      Null: YES
      Index_type: BTREE
      Comment:
Index_comment:
```



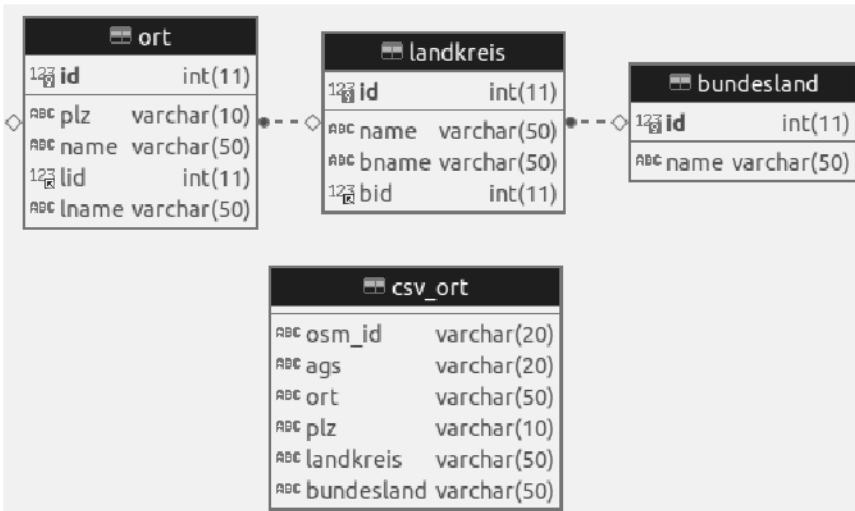
3. Tabellen ergänzen bzw. anpassen



Damit im nächsten Schritt die `bid` ermittelt werden kann, wird temporär der Bundeslandname (`bname`) mit gespeichert (genauso beim `ort`).

AUF-04-1 Erstellen Sie für die Tabellen entsprechende `create table` bzw. `alter table` Befehle und notieren Sie diese.

Notiz:



4. Bundesländer anzeigen und einfügen

Da auf die Spalte `bundesland` ein Index gelegt worden ist, wird dieser automatisch genutzt und somit ist die Ausgabe sortiert.

```
select distinct bundesland from csv_ort;
```

```
+-----+
| bundesland |
+-----+
| Baden-Württemberg |
| Bayern |
| Berlin |
| Brandenburg |
| Bremen |
| Hamburg |
| Hessen |
| Mecklenburg-Vorpommern |
| Niedersachsen |
| Nordrhein-Westfalen |
| Rheinland-Pfalz |
| Saarland |
| Sachsen |
| Sachsen-Anhalt |
| Schleswig-Holstein |
| Thüringen |
+-----+
```

```
INSERT INTO bundesland
      (name) select distinct bundesland from csv_ort;
```

```
Query OK, 16 rows affected (0.007 sec)
Records: 16 Duplicates: 0 Warnings: 0
```

```
select * from bundesland;
```

```
+-----+-----+
| id | name |
+-----+-----+
| 1 | Baden-Württemberg |
| 2 | Bayern |
| 3 | Berlin |
| 4 | Brandenburg |
| 5 | Bremen |
| 6 | Hamburg |
| 7 | Hessen |
| 8 | Mecklenburg-Vorpommern |
| 9 | Niedersachsen |
| 10 | Nordrhein-Westfalen |
| 11 | Rheinland-Pfalz |
| 12 | Saarland |
| 13 | Sachsen |
| 14 | Sachsen-Anhalt |
| 15 | Schleswig-Holstein |
| 16 | Thüringen |
+-----+-----+
```

Notiz:



5. Landkreise einfügen¹

```
INSERT INTO landkreis (name,bname)
  select
    distinct landkreis,bundesland
  from csv_ort
  where landkreis > ''
  order by landkreis;

# bid anpassen
update landkreis set bid=(select id from bundesland where bname=name);

# landkreise anzeigen
select * from landkreis limit 10;
```

id	name	bname	bid
1	Alb-Donau-Kreis	Baden-Württemberg	1
2	Altmarkkreis Salzwedel	Sachsen-Anhalt	14
3	Bodenseekreis	Baden-Württemberg	1
4	Burgenlandkreis	Sachsen-Anhalt	14
5	Donnersbergkreis	Rheinland-Pfalz	11
6	Eifelkreis Bitburg-Prüm	Rheinland-Pfalz	11
7	Ennepe-Ruhr-Kreis	Nordrhein-Westfalen	10
8	Enzkreis	Baden-Württemberg	1
9	Erzgebirgskreis	Sachsen	13
10	Hochsauerlandkreis	Nordrhein-Westfalen	10

Notiz:

¹ Achtung: kreisfreie Städte wie z.B. München haben keinen Landkreis!



6. Orte einfügen

```
# alte Orte löschen
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
TRUNCATE ort;

INSERT INTO ort (plz,name,lname)
  select
    plz, ort,landkreis
  from csv_ort
  order by ort;

# lid anpassen
update ort set lid=(select id from landkreis where lname=name);

# Orte anzeigen
select * from ort limit 10;
```

id	plz	name	lid	lname
1	78267	Aach	143	Landkreis Konstanz
2	54298	Aach	235	Landkreis Trier-Saarburg
3	52080	Aachen	286	Städteregion Aachen
4	52078	Aachen	286	Städteregion Aachen
5	52076	Aachen	286	Städteregion Aachen
6	52074	Aachen	286	Städteregion Aachen
7	52072	Aachen	286	Städteregion Aachen
8	52070	Aachen	286	Städteregion Aachen
9	52068	Aachen	286	Städteregion Aachen
10	52066	Aachen	286	Städteregion Aachen

AUF-04-2 Warum müssen vor dem Befehl TRUNCATE ort die Checks ausgeschaltet werden?

Notiz:



7. Aufräumen, Mitarbeiter anpassen und FKs setzen

```
ALTER table ort drop column lname;
ALTER table landkreis drop column bname;
DROP table csv_ort;

# Mitarbeiter: oid neu setzen
update mitarbeiter set oid=(
  select id from ort where plz='85662' AND name='Hohenbrunn')
  where id in (8,9);
update mitarbeiter set oid=(
  select id from ort where plz='85653' AND name='Aying')
  where id in (6,7);
update mitarbeiter set oid=(
  select id from ort where plz='82061' AND name='Neuried')
  where id in (4,5);
update mitarbeiter set oid=(
  select id from ort where plz='85521' AND name='Hohenbrunn')
  where id=3;
update mitarbeiter set oid=(
  select id from ort where plz='85653' AND name='Aying')
  where id=2;
update mitarbeiter set oid=(
  select id from ort where plz='85609' AND name='Aschheim')
  where id=1;

# FKs
ALTER table ort ADD CONSTRAINT fk_ort_landkreis
  Foreign key (lid) References landkreis (id);

ALTER table landkreis ADD CONSTRAINT fk_landkreis_bundesland
  Foreign key (bid) References bundesland (id);
```

Notiz:



8. Prüfen...

```
desc ort;
```

Field	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
plz	varchar(10)	YES		NULL	
name	varchar(50)	YES		NULL	
lid	int(11)	YES	MUL	NULL	

```
desc landkreis;
```

Field	Type	Null	Key	Default	Extra
id	int(11)	NO	PRI	NULL	auto_increment
name	varchar(50)	YES		NULL	
bid	int(11)	YES	MUL	NULL	

```
select m.name,m.vorname, o.plz, o.name  
from mitarbeiter m left join ort o on m.oid=o.id;
```

name	vorname	plz	name
Walker	Jonny	85609	Aschheim
Schlau	Susi	85653	Aying
Ratlos	Rudi	85521	Hohenbrunn
Keller	Brigitte	82061	Neuried
Keller	Josef	82061	Neuried
Huber	Sepp	85653	Aying
Meier	Siglinde	85653	Aying
Mair	Hans	85662	Hohenbrunn
Maier	Peter	85662	Hohenbrunn

Sie sehen, mit SQL solch einen Datenimport zu machen, ist nicht immer ganz so einfach. Verwendet man dagegen eine Programmiersprache wie Java, ist dies deutlich einfacher.

AUF-04-3 Führen Sie schrittweise den Import selber durch!

Notiz:



Problem: kreisfreie Städte

Bei dem jetzigen Datenmodell gibt es ein Problem bei kreisfreien Städten, da diese keinem Landkreis zugeordnet sind.

```
select * from ort where lid is null limit 5;
```

id	plz	name	lid
287	92224	Amberg	NULL
331	91522	Ansbach	NULL
393	63739	Aschaffenburg	NULL
394	63741	Aschaffenburg	NULL
395	63743	Aschaffenburg	NULL

AUF-04-4 Somit lässt sich nicht ermitteln, in welchem Bundesland die Stadt liegt. Überlegen Sie sich dazu eine Lösung und setzen Sie diese um.

Notiz:

