

# Einführung in Datenbanken

## — Übungsblatt 11 (GROUP BY, HAVING) —

Ihre Lösungen laden Sie bitte in die Übungsplattform in StudIP hoch ([StudIP-Eintrag der Vorlesung], Reiter „LTI-Tool“, dann auf „Anwendung starten“).

**Einsendeschluss ist Montag, der 15.01.2024, 18<sup>00</sup>.**

### Datenbank:

Wählen Sie im Adminer das Schema „komponist\_public“:

```
[https://dbs.informatik.uni-halle.de/edb?pgsql=db&
    username=student_gast&db=postgres&ns=komponist_public]
```

Dies ist eine Datenbank, in der ich vor langer Zeit meine klassischen Musik-CDs erfasst habe.

- **KOMPONIST**(KNR, NAME, VORNAME°, GEBOREN°, GESTORBEN°)
- **STUECK**(SNR, KNR°→KOMPONIST, TITEL, TONART°, OPUS°)
- **CD**(CDNR, NAME, HERSTELLER°, ANZ\_CDS°, GESAMTSPIELZEIT°)
- **AUFNAHME**(CDNR→CD, SNR→STUECK, ORCHESTER°, LEITUNG°)
- **SOLIST**((CDNR, SNR)→AUFNAHME, NAME, INSTRUMENT°)

Verwenden Sie in Ihren Anfragen nur Informationen, die in der Aufgabenstellung erwähnt sind. Sie dürfen z.B. nicht Komponisten-Nummern in der Datenbank nachschlagen und dann in die Anfrage einsetzen. Ihre Anfrage muss mit beliebigen DB-Zuständen funktionieren.

Bedenken Sie, dass Sie in der Klausur nur 10–15 Minuten pro Anfrage haben. Testen Sie, wie lange Sie brauchen.

### Hinweis (zu ORDER BY):

Für Hausaufgaben und die Klausur gilt: Falls die Aufgabenstellung keine Sortierung vorschreibt, gibt es keinen Punktabzug, wenn Sie trotzdem eine sinnvolle Sortierung durchführen (das könnte z.B. den Vergleich mit der erwarteten Antwort erleichtern).

Falls die `ORDER BY`-Klausel aber eindeutig mangelhaft ist, ist ein Punktabzug nicht ausgeschlossen. Ein Punktabzug wäre z.B. möglich, wenn die Anfrage beweisbar in allen Datenbank-Zuständen nur ein einziges Ausgabetupel liefert. Dann wäre die Sortierung ja nicht sinnvoll.

Man kann sich auf den Standpunkt stellen, dass alle Anfragen ein sinnvoll sortiertes Ergebnis liefern sollten. Das würde auch plötzliche Änderungen der Tupel-Reihenfolge bei einem Update des DBMS ausschließen (vielleicht hat der alte Optimierer `GROUP BY` mit Sortierung implementiert, der neue mit einer Hashtabelle zum Finden der Gruppen-Werte). In dieser Vorlesung wird eine nicht explizit verlangte Sortierung also als Frage des persönlichen Stils angesehen.

In der Klausur werden verlangte Sortierungen nicht selten vergessen (mit Punktabzug). Wenn Sie sich also daran gewöhnen, dass jede Anfrage (die mehr als eine Zeile liefert) ein `ORDER BY` enthalten sollte, könnte Ihnen das Fehler in der Klausur ersparen (wenn eine Sortierung verlangt ist, müssen die Sortierkriterien natürlich auch passen).

### Aufgabe 1 (4 Punkte)

Geben Sie Name und Vorname von allen Komponisten aus, von denen es mindestens 5 Stücke in der Datenbank gibt. Geben Sie die Anzahl der Stücke mit aus, nennen Sie die Spalte `anz_stuecke` und sortieren Sie die Ausgabe danach absteigend (größte Anzahl zuerst). Bei gleicher Anzahl Stücke sortieren Sie nach dem Nachnamen des Komponisten, wenn der auch gleich ist, nach dem Vornamen.

Stücke mit unbekanntem Komponisten (d.h. mit einem Nullwert in der Spalte `knr`) können Sie natürlich ignorieren. Sie können davon ausgehen, dass Name und Vorname zusammen einen Alternativschlüssel der Komponisten-Tabelle bilden.

Die erwartete Antwort ist:

name	vorname	anz_stuecke
Händel	Georg Friedrich	25
Telemann	Georg Philipp	15
Prokofiev	Serge	13
Bach	Johann Sebastian	10
Monteverdi	Claudio	10
Biber	Heinrich Ignaz Franz	9
Mozart	Wolfgang Amadeus	9
Vivaldi	Antonio	8
Mozart	Leopold	6
Beethoven	Ludwig van	5

## Aufgabe 2 (4 Punkte)

Von welchen Orchestern gibt es Aufnahmen auf mindestens zwei CDs? Gemeint sind dabei die CD-Packs, die durch eine `CDNR` identifiziert werden (eine Doppel-CD hat nur eine `CDNR` und soll in dieser Rechnung auch nur als eine „CD“ zählen).

Natürlich soll ein Nullwert als Orchester (z.B. bei Solo-Stücken) ausgeschlossen sein.

Geben Sie den Namen des Orchesters und die Anzahl der CD-Packs aus. Sortieren Sie nach der Anzahl CD-Packs (größte zuerst) und innerhalb einer Anzahl alphabetisch nach dem Orchester-Namen.

Die erwartete Ausgabe ist:

orchester	anz_cds
Academy of St.Martin-in-the-Fields	3
London Festival Orchestra	3
Concentus musicus Wien	2
Die Zagreber Solisten	2
London Philharmonic Orchestra	2
Radio-Sinfonieorchester Stuttgart	2
Südwest-Studioorchester	2

7 Datensätze

## Aufgabe 3 (4 Punkte)

Was ist die minimale, maximale und durchschnittliche Anzahl Stücke pro CD?

Sie müssen also die Anzahl Stücke pro `cdnr` bestimmen, und jeweils durch `anz_cds` teilen (z.B. haben Doppel-CDs nur eine `cdnr`), und anschließend den Durchschnitt über diesen Anzahlen bilden.

Runden Sie die Zahlen auf eine Nachkommestelle (geht mit der Funktion `ROUND` mit der Anzahl Nachkommastellen als zweitem Argument). Nennen Sie die Spalten „Min“, „Max“ und „Durchschnitt“ — auch mit dieser Groß-/Kleinschreibung.

Die erwartete Antwort ist:

Min	Max	Durchschnitt
0.5	23.0	3.8

1 Datensatz

### Aufgabe 4 (4 Punkte)

Geben Sie für jeden Komponisten, von dem mindestens drei Stücke mit bekannter Tonart (nicht Null) in der Datenbank stehen, aus, wie viele Stücke es sind, und was der Prozentsatz in Dur-Tonarten und was der Prozentsatz in Moll-Tonarten ist. Den Komponisten identifizieren Sie durch Name und Vorname und ordnen die Ausgabe nach dem Namen, bei gleichem Namen nach dem Vornamen. Dur-Tonarten sind solche, die in „-dur“ enden, und Moll-Tonarten solche, die in „-moll“ enden. Alle vorkommenden Tonarten (die nicht Null sind) enden auf einen dieser beiden Suffixe.

Bei dieser Aufgabe sollen nur Stücke mit definierter Tonart betrachtet werden (auch in den Prozent-Berechnungen). Geben Sie die Prozente gerundet ohne Nachkommastellen aus (nutzen Sie `ROUND`, und zwar die Variante mit nur einem Parameter).

Achten Sie darauf, dass `COUNT` bei PostgreSQL einen Integer-Typ liefert (`bigint`), und PostgreSQL die Integer-Division verwendet, wenn beide Argumente einen Integer-Typ haben (es wird dann also abgerundet). Sie könnten z.B. die Anzahl Dur/Moll-Stücke mit `100.0` multiplizieren, bevor Sie durch die Gesamtanzahl dividieren.

Die erwartete Antwort ist:

name	vorname	anz	prozent_dur	prozent_moll
Bach	Johann Sebastian	10	70	30
Händel	Georg Friedrich	21	62	38
Mozart	Leopold	6	100	0
Mozart	Wolfgang Amadeus	5	80	20
Prokofiev	Serge	4	75	25
Schubert	Franz	3	67	33
Telemann	Georg Philipp	15	87	13
Vivaldi	Antonio	8	50	50
Wolf-Ferrari	Ermanno	4	100	0

9 Datensätze

### Aufgabe 5 (4 Punkte)

Erstellen Sie eine Liste aller vorkommenden Instrumente in alphabetischer Reihenfolge mit einer fortlaufenden ID (beginnend bei 1).

Sie können die ID berechnen, indem Sie zählen, wie viele Instrumente in der alphabetischen Reihenfolge vor dem aktuellen Instrument kommen (oder gleich sind).

Sortieren Sie die Ausgabe nach der so berechneten ID.

Die erwartete Antwort ist:

ID	Instrument
1	Bagpipes
2	Bariton
3	Bass
4	Cello
5	Drehleier
6	English Horn
7	Erzaehler
8	Horn
9	Oboe
10	Piano
11	Sopran
12	Tenor
13	Trumpet
14	Viola
15	Violine

15 Datensätze