



Institut für Informatik - Lehrstuhl Datenbanken

Einführung in Datenbanken

Wintersemester 22/23 - Prof. Dr. Stefan Brass, PD Dr. Alexander Hinneburg

Übung 9: SQL-Anfragen Aggregation

Abgabe bis Mo. 9.1.2023, 18:00 Uhr,

Übungsplattform: Studip → Einführung in Datenbanken → Übungsplattform

Aufgabe 9.1:

10 Punkte

Loggen Sie sich in die PostgreSQL-Datenbank über StudIP ein und wählen Sie das Schema `empdept_public`. Die kleine, übersichtliche Spieldatenbank hat die Tabellen `emp` für Employee (Angestellter) und `dept` für Department (Abteilung). Die Typen der Attribute sind hinter den Namen dargestellt.

- `Emp(Empno: Numeric(4), Ename°: Varchar(10), Job°: Varchar(9), Mgr°: Numeric(4) → Emp(Empno), Hiredate°: Date, Sal°: Numeric(7, 2), Comm°: Numeric(7, 2), Deptno°: Numeric(2) → Dept)`
- `Dept(Deptno: Numeric(2), Dname°: Varchar(14), Loc°: Varchar(13))`

`Ename` steht für den Namen des Angestellten, `Job` bezeichnet die Arbeit des Angestellten, `Mgr` zeigt auf die `Empno` des Vorgesetzten, `Hiredate` steht für Einstellungsdatum, `Sal` für Gehalt, `Comm` für Provision und `Deptno` zeigt auf die Abteilung, wo der Angestellte arbeitet. `Dname` steht für Abteilungsnamen und `Loc` für den Ort der Abteilung.

Geben Sie für die folgenden Anfragen entsprechende SQL-Statements an. Für jede Anfrage soll nur ein SQL-Statement angegeben werden. Falls die Tutoren eine SQL-Anfrage nicht schnell verstehen können, kann es Punktabzug geben. Sie können das durch gute Formatierung und hilfreiche Kommentare (wo nötig) vermeiden. Beachten Sie aber, dass die Tutoren SQL natürlich sehr gut kennen. Vermeiden Sie triviale Kommentare. SQL-Statements mit Syntax-Fehlern werden mit Null Punkten bewertet. Geben Sie jede Teilaufgabe als separate Text-Datei mit der Endung `.sql` ab (kein Word). Unter Windows können Sie Text-Dateien mit Notepad editieren. Das Programm Notepad++ (<http://notepad-plus-plus.org>) bietet mehr Funktionen.

Beachten Sie, dass die SQL Anfragen keine Duplikate ausgeben sollen, d.h. verwenden Sie `DISTINCT`, wenn es notwendig ist. Für vergessenes oder überflüssiges `DISTINCT` werden Punkte abgezogen.

- a) Geben Sie die Anzahl der Angestellten (`COUNT_EMP`) und das durchschnittliche Gehalt für die Angestellten (`AVG_SAL`) jeder Abteilung (`DEPTNO`, `DNAME`) aus. Falls eine Abteilung keine Angestellten hat, soll sie einfach nicht ausgegeben werden.

- b) Geben Sie für jeden Job (JOB) die Anzahl (COUNT_EMP) und das durchschnittliche Gehalt aller Angestellten (AVG_SAL) aus, die diesen Job machen und in Abteilungen in Dallas oder Chicago arbeiten.
- c) Geben Sie die DEPTNO, den Namen (DNAME) und das Durchschnittsgehalt der Abteilung mit dem maximalen durchschnittlichen Gehalt (MAX_AVG_SAL) aus. Nutzen Sie nicht Order by und Limit.

Falls eine Abteilung keine Angestellten hat, soll sie einfach nicht in die Berechnungen einbezogen werden.

- d) Welche Angestellten (EMPNO, ENAME, SAL, DEPTNO, AVG_SAL) verdienen weniger als die Hälfte des Durchschnittsgehaltes ihrer Abteilung?
- e) Geben Sie für alle Abteilungen die Nummer und den Namen (DEPTNO, DNAME) zusammen mit dem größten Gehalt (MAX_SAL) aus, das ein Angestellter in dieser Abteilung verdient. Filtern Sie Abteilungen vor der Ausgabe, so dass nur die Abteilungen ausgegeben werden, deren größtes Gehalt kleiner dem Maximalgehalt eines Mitarbeiters in einer Abteilung mit der Bezeichnung 'OPERATIONS' ist.

Weil in dem Schema empdept_public kein Mitarbeiter in der Abteilung OPERATIONS arbeitet, muss in diesem DB-Zustand das leere Ergebnis rauskommen. Um die Funktion Ihrer Anfrage besser zu sehen, können Sie zusätzlich das Schema empdept_public_2 nutzen. Dieser DB-Zustand enthält zwei Abteilungen mit dem Name OPERATIONS.

- f), **Bonusaufgabe** Geben Sie den Gini-Index der Einkommen (GINI_SAL) für jede Abteilung (DEPTNO, DNAME) absteigend geordnet aus. Wenn der Gini-Index für zwei Abteilungen gleich ist, soll nach Abteilungsnummer aufsteigend sortiert werden. Der Gini-Index misst den Grad der Ungleichheit der Einkommen für eine gegebene Menge von Personen. Seien x_1, \dots, x_n die Einkommen von n Personen, dann ist der Gini-Index G definiert als:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2 \sum_{i=1}^n \sum_{j=1}^n x_j}$$

$$= \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n \sum_{i=1}^n x_i}$$

https://en.wikipedia.org/wiki/Gini_coefficient#Definition

Falls eine Abteilung keine Angestellten hat, soll sie einfach nicht in die Berechnungen einbezogen werden.