

4. Versuch: Optimierung, Datumsangaben, Rekursion, Zugriffsrechte

Aufgabe 4.1 (Optimierung Korrelation; 15 P.)

In Aufgabe 2.5 haben Sie einen Korrelationskoeffizienten berechnet. Im folgenden finden Sie die SQL-Anfrage zur Formel aus Aufgabe 2.5, der Einfachheit halber hier zur Korrelation zwischen Landesbevölkerung und -größe.

```
WITH Average AS
(
  SELECT AVG(population) population, AVG(area) area
  FROM Country
  WHERE population IS NOT NULL AND area IS NOT NULL
)
SELECT
  SUM((Country.population - Average.population)*(Country.area - Average.area)) /
  (
    COUNT(*) *
    SQRT(SUM(POWER(Country.population - Average.population,2))/COUNT(*) *
    SQRT(SUM(POWER(Country.area - Average.area, 2))/COUNT(*)
  ) r
FROM Country CROSS JOIN Average
WHERE Country.population IS NOT NULL AND Country.area IS NOT NULL;
```

Wenn Sie den Ausführungsplan dieser Anfrage untersuchen, stellen Sie fest, dass die Tabelle *Country* zweimal gelesen werden muss. Formulieren Sie die Anfrage (Formel) nun so um, dass ein einmaliges Lesen von *Country* genügt. Geben Sie die Umformungsschritte kurz an.

Hinweis: Sie müssen also die verschachtelten Aggregationen entfernen. Die Funktion CORR berechnet ebenfalls diesen Korrelationskoeffizienten. Wenn Sie den Ausführungsplan dieser Funktion ansehen, können Sie erkennen, mit welchen Aggregationen Sie auskommen können. Außerdem können Sie sich überlegen, wie die verwendeten NVL2-Ausdrücke die not-null-Bedingungen ersetzen können.

Aufgabe 4.2 (Korrelation Organisationen; 15 P.)

Berechnen Sie die Korrelation zwischen dem Pro-Kopf-Einkommen und der Mitgliedschaft in der Europäischen Union eines Landes. Berücksichtigen Sie nur europäische Länder. Die Formel dafür lautet (punkt-biseriale Korrelation):

$$r_{pb} = \frac{\bar{y}_0 - \bar{y}_1}{\sigma_y} \cdot \sqrt{\frac{n_0 n_1}{n^2}}$$

\bar{y}_0 und \bar{y}_1 sind die arithmetischen Mittelwerte der Merkmalsausprägungen (hier: Pro-Kopf-Einkommen) der beiden Gruppen, σ_y ist die Standardabweichung des Merkmals (ohne Unterscheidung nach Gruppen), ist die Anzahl Mitglieder, n_1 die Anzahl Nicht-Mitglieder und $n = n_0 + n_1$. Hier wird nicht zwischen positiven und negativen Korrelationen unterschieden.

- Schreiben Sie eine Anfrage, die den Korrelationskoeffizienten berechnet.
- Schreiben Sie die Anfrage jetzt so, dass alle benötigten Tabellen nur einmal gelesen werden. Falls Sie das schon in Teil a) geschafft haben, haben Sie hier nichts zu tun. **Hinweis:** Zur Lösung bietet sich ein äußerer Verbund und die Verwendung von NVL2, DECODE oder CASE an.
- Schreiben Sie eine Anfrage, die für jede Organisation die Korrelation zwischen der Mitgliedschaft eines Landes in dieser Organisation und dem Pro-Kopf-Einkommen eines Landes berechnet. Sortieren Sie die Ausgabe aufsteigend nach r_{pb} .

Aufgabe 4.3 (Direkter Mengenvergleich; 5 P.)

Erzeugen Sie eine Sicht, die alle Flüsse wie in *River* enthält. Zusätzlich soll die Sicht eine mengenwertige Spalte *countries* (verschachtelte Tabelle) haben, die zu jedem Fluss alle Länder (*code*) enthält, durch die der Fluss fließt. Schreiben Sie dazu eine Anfrage, die alle (ungeordneten) Paare von Flüssen ausgibt, die durch die gleichen Länder fließen sowie die Menger dieser Länder (nur *code*). **Hinweis:** Auf mengenwertigen Attributen über einen einfachen Datentyp können Mengenvergleiche direkt ausgeführt werden.

Aufgabe 4.4 (Datumsangaben; 20 P.)

Die Relation *politics* enthält das Attribut *Independence*.

- a) Formulieren Sie eine Anfrage, die alle Länder liefert, die zwischen 1300 und 1600 gegründet wurden.
- b) Ermitteln Sie das arithmetische Mittel und die Standardabweichung der Gründungsdaten der europäischen Staaten.
- c) Geben Sie für alle Länder das Gründungsdatum in einer zweispaltigen Tabelle aus, in der die erste Spalte den Country-Code enthält, die zweite das Gründungsdatum in einem Format wie in folgendem Beispiel (ohne Anführungszeichen): "18. Jan"
- d) Gruppieren Sie alle Länder nach den Wochentage ihres Gründungsdatums und berechnen Sie davon den Modalwert (den Wochentag mit den meisten Gründungen).

Aufgabe 4.5 (Fibonacci-Zahlen; 15 P.)

Berechnen Sie mit einem rekursiven WITH die Fibonacci-Zahl f_n . Eine optimale Lösung erlaubt die Angabe von n in der WHERE-Klausel der Hauptanfrage und liefert genau n und f_n zurück. **Hinweis:** Überlegen Sie sich, wie Sie die Berechnung beim gewünschten f_n anhalten können. Eine Möglichkeit beinhaltet die Verwendung von *rownum*.

Aufgabe 4.6 (Sicherheit; 15 P.)

In den Skripten `tax_employee.sql` und `tax_inspector.sql` finden Sie ein kleines Steuerprüfungsszenario. Darin gibt es zwei Benutzer, die Tabellen zu einem globalen Modell beisteuern: Ein Angestellter (`employee`) verwaltet die Steuerpflichtigen in der Tabelle `taxpayer` (verantwortlich für das Einfügen, Löschen und Ändern dieser Daten). Der Steuerprüfer (`inspector`) benötigt nur Lesezugriff auf diese Tabelle. Er verwaltet die Tabellen `employment` mit Informationen zu Arbeitgebern und `revenue` über das jährliche Einkommen der Steuerzahler. Der Angestellte hat hierauf keinen Zugriff, sondern nur auf eine Sicht mit aggregierten Werten für statistische Auswertungen.

Installieren Sie das Szenario bei zwei Mitgliedern Ihrer Gruppe (ersetzen Sie "inspector" und "employee" durch Ihre Benutzernamen) und machen Sie sich damit vertraut. Ihnen dürfte schnell eine Sicherheitslücke auffallen. Schreiben Sie ein Skript, mit dem Sie zeigen, wie sich der Angestellte unbemerkt Zugang zu vertraulichen Informationen verschaffen kann.

Aufgabe 4.7 (Erdoberfläche; 5 P.)

Wie groß ist der Anteil der Erdoberfläche, der von Meeren bedeckt ist?

Abgabe: 23.6.2010, 11h