

Klausur Datenbanken
Wintersemester 2013/2014
Prof. Dr. Wolfgang May
29. Januar 2014, 14-16 Uhr
Bearbeitungszeit: 90 Minuten

Vorname:

Nachname:

Matrikelnummer:

Studiengang:

Bei der Klausur sind **keine Hilfsmittel** (Skripten, Taschenrechner, etc.) erlaubt. Handies müssen ausgeschaltet sein. Papier wird gestellt. Benutzen Sie nur die **ausgeteilten**, zusammengehefteten **Blätter** für Ihre Antworten. Schreiben Sie mit blauem/schwarzem Kugelschreiber, Füller, etc.; Bleistift ist nicht erlaubt.

Zum **Bestehen** der Klausur sind **45** Punkte hinreichend.

- meine Note soll mit Matrikelnummer so bald wie möglich auf der Vorlesungs-Webseite veröffentlicht werden.
- meine Note soll nicht veröffentlicht werden; ich erfahre sie dann aus FlexNever oder beim zuständigen Prüfungsamt.

	Max. Punkte	Erreichte Punkte
Aufgabe 1 (ER-Modell)	16	
Aufgabe 2 (Transformation in das Relationale Modell)	19	
Aufgabe 3 (SQL und Relationale Algebra)	40	
Aufgabe 4 (Verschiedenes)	15	
Summe	90	

Note:

Themenstellung: NSA-Telefongesprächsdatenbank

Alle Klausuraufgaben basieren auf einem gemeinsamen “Auftrag”: In der Klausur soll eine Datenbank zur Speicherung von Daten über Mobiltelefongespräche (und Aufenthaltsorte) entworfen werden.

1. Handy-Telefonnummern bestehen aus einer Landesvorwahl und einer Nummer, z.B. 0049-12345 (Handynetzvorwahl und Endnummer werden nicht unterschieden; es werden keine Festnetztelefongespräche berücksichtigt.)
Zu jedem Handy ist der Hersteller bekannt. (In diesem Szenario wird angenommen, dass man SIM-Karten nicht in ein anderes Handy umstecken kann; hier geht es nur darum, dass das Handy ein Attribut hat).
2. Jedes Land hat eine eigene Landesvorwahl:
 - 0049 ist *Deutschland*, 0093 ist *Afghanistan*.
3. Zu jedem Land ist die Region gespeichert. *Deutschland* liegt in *Mitteleuropa*, *Afghanistan* liegt im *Mittleren Osten*.
4. Von manchen Handynummern ist zumindest teilweise bekannt, von welchen Personen sie benutzt werden. Eine Handynummer kann von mehreren Personen benutzt werden (bzw. benutzt worden sein). Eine Person kann mehrere Handynummern benutzen.
5. Personen haben einen (Vor+Nach)Namen.
 - Die Handynummer 0049-10000 wird/wurde von *Angela Merkel*, *Ronald Pofalla* und *Peter Tauber* benutzt (und vielleicht noch von weiteren Personen). Es ist ein *Nokia*-Handy.
 - Die Handynummer 0049-12345 wird/wurde von *Angela Merkel* und *Joachim Sauer* benutzt. Es ist ein *Apple*-Handy.
6. Handies können geortet werden, wenn sie etwas senden oder empfangen. Für jede Ortung sind Tag, Uhrzeit, sowie die geographischen Koordinaten gespeichert.
7. Koordinaten werden als Längen- und Breitengrad (mit 2 Nachkommastellen) angegeben.
 - Das Handy 0049-10000 wurde am 20.12.2013 um 09:00 an den Koordinaten (13.37, 52.52) geortet.
8. Jedem solchen Koordinatenpaar ist das Land zugeordnet, in dem es sich befindet (Annahme: bei Grenzgebieten wird nur ein Land gespeichert; bei Koordinaten in internationalen Gewässern ist kein Land zugeordnet) sowie optional die Stadt (falls dort eine Stadt ist; ebenfalls maximal eine).
 - Das Paar (13.37, 52.52) befindet sich in *Berlin* in *Deutschland*.
9. Zu jedem abgehörten Gespräch ist der Zeitpunkt (Datum und Ortszeit des Anrufers), die anrufende Handynummer und die angerufene Handynummer, sowie eine Referenz auf eine Audiodatei gespeichert.
 - Am 27.12.2013 wurde von der Telefonnummer 0049-45454 um 12:00 (Ortszeit; das Handy befand sich an den Koordinaten (9.72, 52.16) in Hannover) die Nummer 0049-98765 angerufen; die Audio-Datei liegt unter 271213-4945454-1200.mp3.

- Das dabei angerufene Handy 0049-98765 befand sich zu diesem Zeitpunkt an den Koordinaten (55.30, 25.27) in den *Vereinigten Arabischen Emiraten*, in der Stadt *Dubai*, wo es gerade 15:00 Ortszeit war.
- Das Handy mit der Nummer 0093-10101 (von dem bekannt ist, dass es von *Izmir Übüil ben Müsli* benutzt wird) wurde am 27.12.2013 um 16:00 (Ortszeit) ebenfalls an den Koordinaten (55.30, 25.27) geortet, als von ihm die Handynummer 0093-121212 angerufen wurde; die Audio-Datei liegt unter 271213-9310101-1600.mp3.
- Das dabei angerufene Handy mit der Nummer 0093-121212 wurde am 27.12.2013 um 21:00 (Ortszeit) an den Koordinaten (69.10,31.60) geortet.
Dieses Koordinatenpaar befindet sich in *Afghanistan*; dort ist keine Stadt.

Hier ist auch gleich Platz für Aufgabe 1:

Name:

MatNr.:

Aufgabe 1 (ER-Modell [16 Punkte])

Entwickeln Sie ein ER-Modell für das Szenario. Geben Sie darin die Schlüsselattribute sowie die Beziehungskardinalitäten an.

Aufgabe 2 (Transformation in das Relationale Modell [19 Punkte])

- a) **Lösen Sie diesen Aufgabenteil auf dem *letzten* Blatt und trennen dieses ab (und geben es am Ende mit ab!).** Dann haben Sie dieses Blatt separat zugreifbar um später damit die Aufgaben 2b, 3 und 4 (SQL, Relationale Algebra+SQL, Diverses) zu lösen.

Geben Sie an, welche Tabellen (mit Attributen, Schlüsseln etc.) Ihre Datenbank enthält (keine SQL CREATE TABLE-Statements, sondern einfach grafisch). (12 P)

Markieren Sie dabei auch Schlüssel (durch unterstreichen) und Fremdschlüssel (durch überstreichen).

Geben Sie die Tabellen mit jeweils mindestens zwei Beispieldupeln (z.B. denen, die sich aus dem Aufgabentext ergeben, und weiteren erfundenen) an.

- b) Geben Sie die CREATE TABLE-Statements für diejenige Tabelle, in denen die Informationen über die Gesprächsdaten abgespeichert sind, so vollständig wie möglich an (verwenden Sie u.a. die Datentypen DATE und TIME; 7 P).

Aufgabe 3 (SQL und Relationale Algebra [40 Punkte])

Verwenden Sie für diese Aufgabe die von Ihnen entworfene relationale Datenbasis. Keine der Antworten soll Duplikate enthalten.

- a) Geben Sie **eine SQL-Anfrage und einen Algebra-Ausdruck oder -Baum** an, die die Namen aller Personen ausgibt, die ein Handy benutzen, das am 28.1.2014 im Bereich der Stadt *Göttingen* geortet wurde. (3+3 P)
- b) Geben Sie eine **SQL-Anfrage** an, die alle deutschen Telefonnummern ausgibt, von denen mindestens 3 Anrufe durchgeführt wurden, bei denen sich der Angerufene zu diesem Zeitpunkt in *Nahost* oder im *Mittleren Osten* aufhielt. (4 P)
- c) Geben eine **eine SQL-Anfrage und einen Algebra-Ausdruck oder -Baum** an, die die Namen aller Länder ausgibt, zu deren Landeswahlbereich im Jahr 2013 keine Gespräche von Telefonen (als Anrufende), die von *Angela Merkel* benutzt werden, abgehört worden sind. (4+5 P)
- d) Geben Sie eine **eine SQL-Anfrage und einen Algebra-Ausdruck oder -Baum** an, der alle diejenigen Handynummern (Vorwahl und Nummer) ausgibt, in jedem Land des mittleren Ostens mindestens einmal geortet wurden. (5+5 P)
- e) Geben Sie **eine SQL-Anfrage und einen Algebra-Ausdruck oder -Baum** an, der die Nummern aller Handies, und wenn möglich auch die Namen ihrer Benutzer, ausgibt, die von einem Handy, das *Izmir Übü ben Müsli* benutzt, am 27.12.2013 angerufen wurden. (4+4 P)
- f) **Etwas Theorie:** Gegeben sind zwei beliebige Relationen $R(\bar{X})$ und $S(\bar{Y})$ mit $\bar{Y} \subseteq \bar{X}$. Wenn S nur ein Tupel enthält, wie kann der Ausdruck $R \div S$ vereinfacht werden? (3 P)

Name:

MatNr.:

Aufgabe 4 (Verschiedenes [15 Punkte])

- a) Die automatische Erfassung von Gesprächen stellt am 29.1.2014 um 14:15 ein Gespräch von 0049-12345 (das bekannte Merkelsche Partei-Handy) mit 0041-99999 (ein *Huawei*-Handy, über das bisher nichts gespeichert ist) fest und legt den Mitschnitt in der Datei 290114-4912345-1415.mp3 ab.

Leider funktioniert die geographische Ortung gerade nicht, so dass keine entsprechenden Angaben zur Verfügung stehen.

Geben Sie an, welche(s) INSERT-Statement(s) auf Ihrer Datenbank ausgeführt werden (3 P).

- b) Nehmen Sie an, das Land *Tahiti* ist in den Ergebnissen zu Aufgabe 3c) **nicht** enthalten. Kann man daraus schliessen, dass *Angela Merkel* in 2013 nie ein Handy, dessen Landesvorwahl diejenige von *Tahiti* ist, angerufen hat? Begründen Sie Ihre Aussage (3 P).

- c) Interne Auswertung/Indexe: Betrachten Sie Ihre SQL-Anfrage zu Aufgabe 3a). Nehmen Sie folgende interne Speicherung an:

- Es sind Baumindexe auf allen Schlüsseln und allen Fremdschlüsseln vorhanden.
- Die Inhalte aller Tabellen, die Spalten *Datum* und *Zeit* enthalten, sind nach (*Datum*, *Zeit*) geordnet abgelegt.
- Für eine gegebene Stadt erhalten Sie in $O(\log n)$ über einen Index das Tupel mit den Koordinaten.

Beschreiben Sie *kurz*, wie Ihre Anfrage zu Aufgabe 3a) einigermaßen effizient ausgewertet werden kann. (5 P)

- d) Geben Sie für die beiden typischen potentiellen Transaktionsfehlersituationen *Lost Update* und *Dirty Read* an, ob sie beim alltäglichen Betrieb der obigen Datenbank prinzipiell auftreten könnten, wenn man kein Transaktionsmanagement benutzen würde. Falls ja, skizzieren Sie eine solche Situation; falls nein, begründen Sie Ihre Aussage (4 P)

Name:

MatNr.:

Name:

MatNr.:

[Trennen Sie dieses Blatt am besten vor Beginn der Bearbeitung ab]

Lösen Sie hier Aufgabe 2a