

EINFÜHRUNG IN DATENBANKEN

SOMMERSEMESTER 2009

1 Konzeptuelle Modellierung (8 Punkte)

Die folgende Beschreibung skizziert die Anforderungen an ein System zur Verwaltung von Fluggesellschaften, deren Mitarbeiter sowie deren Flugzeug- Flotte.

- Eine Fluggesellschaft ist eindeutig durch ihr Kürzel identifiziert. Ferner besitzt Sie einen Namen.
- Fluggesellschaften können mit anderen Fluggesellschaften alliiert sein (vergl. z.B. die StarAlliance von Lufthansa, Swiss Air ...). Eine Fluggesellschaft kann mehrere solcher Partner haben und selbst Partner verschiedener Gesellschaften sein.
- Eine Fluggesellschaft besitzt Flugzeuge. Ein Flugzeug kann nicht gleichzeitig mehreren Fluggesellschaften gehören. Das Datum, seit dem ein Flugzeug in Besitz einer Gesellschaft ist, ist von Bedeutung.
- Ein Flugzeug wird eindeutig durch seine Kennung identifiziert. Ferner sind wir am Alter des Flugzeuges interessiert.
- Neben dem aktuellen Besitzer sollen alle Vorbesitzer (Fluggesellschaften) eines Flugzeuges gespeichert werden. Ferner soll zugreifbar sein, wie lange sich das Flugzeug im Besitz einer Fluggesellschaft befunden hat und welche Flugmeilenleistung es bei der Gesellschaft erbracht hat.
- Ein Flugzeug ist von einem bestimmten Typ. Dieser Typ ist (eindeutig) beschrieben durch den Hersteller (z.B. Airbus) und die Modell-Bezeichnung (z.B. A380).
- Man kann zwischen Frachtflugzeug-Typen und Passagierflugzeug-Typen (und keinen anderen) unterscheiden. Für Frachtflugzeug-Typen wollen wir die Ladekapazität speichern, für Passagierflugzeug-Typen sind wir an der Anzahl der Sitzplätze interessiert.
- Eine Fluggesellschaft beschäftigt Mitarbeiter.
- Ein Mitarbeiter besitzt einen Namen, möglicherweise mehrere Telefonnummern, eine Adresse (die wiederum aus Straße, PLZ und Ort besteht) und eine Personal-Nummer, die ihn innerhalb seiner Firma eindeutig identifiziert.
- Für verschiedene Mitarbeitergruppen wollen wir Zusatzinformationen ablegen:
 - Für Leitende Angestellte muss die Höhe der Gewinnbeteiligung erfasst werden. Ferner sollen die Mitarbeiter, für die der Angestellte verantwortlich ist, zugreifbar sein. Ein Mitarbeiter hat höchstens einen Leitenden Angestellten als Vorgesetzten.
 - Ein Pilot besitzt Lizenzen für (möglicherweise mehrere) Flugzeug-Typen.
 - Die Körpergröße von FlugbegleiterInnen ist für die Einsatzplanung relevant.
- Darüber hinaus arbeiten natürlich weitere Mitarbeitergruppen (wie z.B. Sekretärinnen, etc.) bei einer Fluggesellschaft. Für diese wollen wir jedoch keine zusätzlichen Informationen erfassen.
- Die Funktion des Leitenden Angestellten ist bei Fluggesellschaften nicht der Verwaltung vorbehalten: Es ist somit möglich, dass auch ein Pilot diese Stellung erreicht.

a) Entwickeln Sie ein ER-Diagramm, das das obige Szenario modelliert. Verwenden Sie die 1:n-Notation für Kardinalitäts-Restriktionen (6 Punkte)

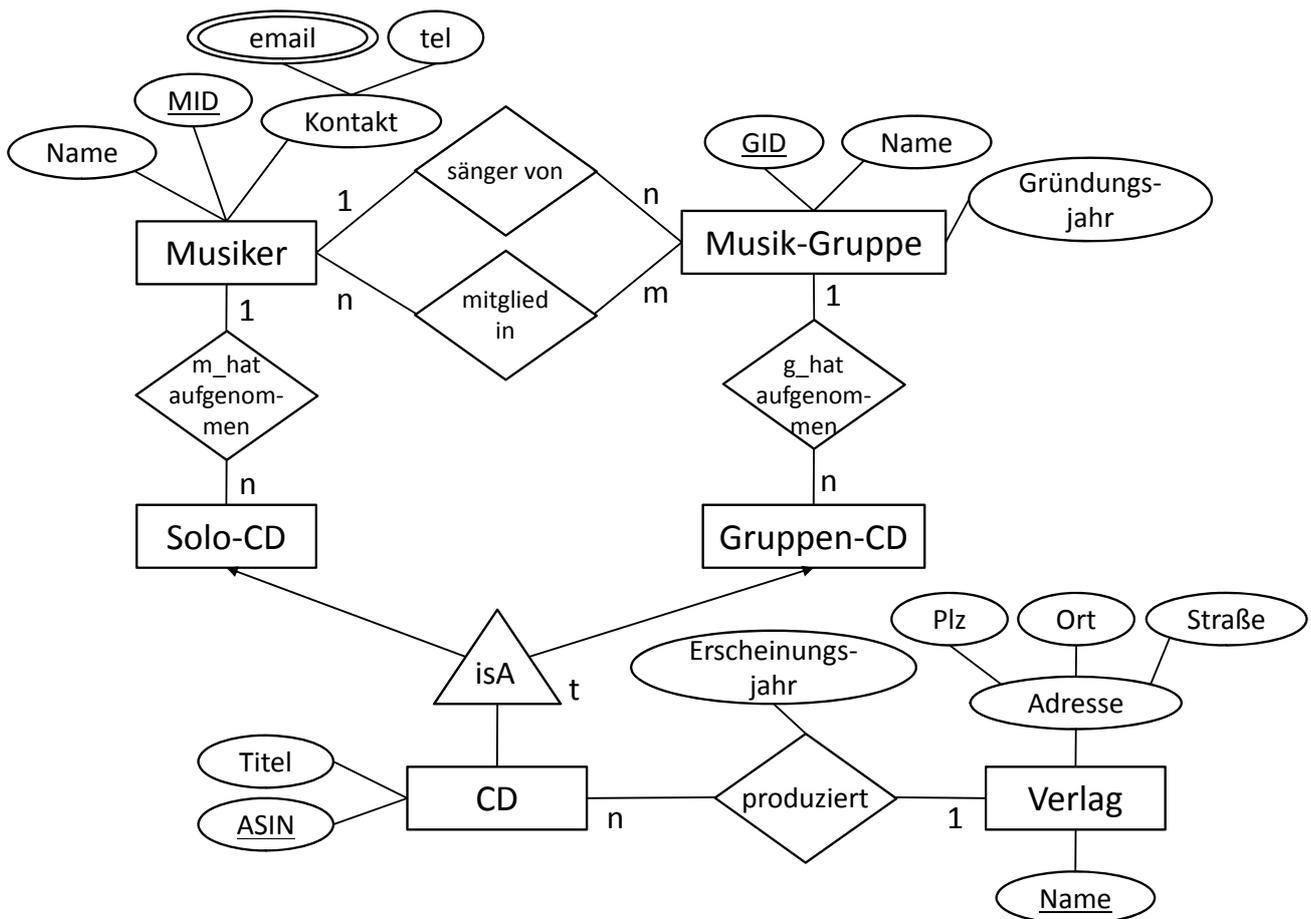
b) Fluggesellschaften bedienen mit bestimmten Flugzeug-Typen Flugstrecken.

Betrachten Sie das folgende Szenario:

- Eine Flugstrecke wird von verschiedenen Gesellschaften (möglicherweise mit dem gleichen Flugzeugtyp) bedient.
- Eine Fluggesellschaft bedient eine Flugstrecke immer mit dem gleichen Flugzeugtyp.
- Eine Fluggesellschaft bedient mehrere Flugstrecken. Möglicherweise werden auf verschiedenen Strecken einer Fluggesellschaft die gleichen Flugzeugtypen eingesetzt.

Erstellen Sie ein ER-Diagramm für diese ternäre Beziehung. Geben Sie Kardinalitätsrestriktionen in (min, max)- und 1:n-Notation an. Welche Notation ist in diesem Szenario ausdrucksstärker und warum? (2 Punkte)

2 Relationale Datenbanken (17 Punkte)



a) Bilden Sie dieses Diagramm auf ein relationales Datenbankschema ab. Verwenden Sie die formale Notation, wie sie in Vorlesung und Übung vorgestellt wurde einschließlich intra- und interrelationaler Abhängigkeiten. (5 Punkte)

- b) Drücken Sie das folgende Informationsbedürfnis im Tupelkalkül aus:
Gesucht sind die Sänger, die eine Solo-CD veröffentlicht haben, die von einem Aachener Verlag produziert wird. (3 Punkte)
- c) Drücken Sie das folgende Informationsbedürfnis durch einen passenden Ausdruck in Relationaler Algebra aus:
Gesucht sind die Namen aller Musik-Gruppen, die eine CD beim Verlag "Rockstar Records" produzieren lassen und ausschließlich aus Künstlern bestehen, die keine Solo-CD aufgenommen haben. (3 Punkte)
- d) Formulieren Sie die folgenden Anfragen in SQL:
- 1.) Gesucht sind die Titel von Solo-CDs, sowie Name und Ort des Verlags, der die CDs produziert, absteigend sortiert nach Erscheinungsjahr. Es sollen nur CDs des Sängers "Keith Caputo" ausgegeben werden. (3 Punkte)
 - 2.) Gesucht sind die Namen von Verlagen und Musik-Gruppen zusammen mit der Anzahl der CDs dieser Gruppen, die vor 2006 produziert wurden. Berücksichtigen Sie nur Bands, deren Sänger keine Solo-CD veröffentlicht hat. (3 Punkte)

3 Funktionale Abhängigkeiten und Normalformen (7 Punkte)

Betrachten Sie das folgende Relationenschema:

$$R = (U, F) = (\{A, B, C, D, E, G, H, I\}, \{CD \rightarrow EG, BC \rightarrow A, H \rightarrow BC, HI \rightarrow D, D \rightarrow H\})$$

- a) Identifizieren Sie alle Schlüsselkandidaten von R , zeigen Sie, dass es sich um Schlüsselkandidaten handelt und beweisen Sie, dass es keine anderen Schlüsselkandidaten gibt. (3 Punkte)
- b) In welcher Normalform ist R ? Warum? (1 Punkt)
- c) Wenden Sie den Dekompositions-Algorithmus zur Herleitung eines Schemas in BCNF an. (3 Punkte)

4 Basis (3 Punkte)

Gegeben sei die folgende Menge von funktionalen Abhängigkeiten für die Universalrelation

$$U = \{A, B, C, D, E, G\}; F = \{A \rightarrow DG, B \rightarrow E, C \rightarrow B, AC \rightarrow DE, G \rightarrow ABE\}$$

Berechnen Sie $BASIS(F)$, d.h. führen Sie den ersten Schritt des Synthese- Algorithmus' durch.

5 Entscheidungsfragen (4 Punkte)

Entscheiden Sie für jede der folgenden Aussagen, ob sie wahr oder falsch ist. Eine Begründung Ihrer Entscheidung brauchen Sie nicht anzugeben.

Es gilt folgendes Bewertungsschema:

Für jede Aussage: korrekte Antwort: 0,5 Pkt., falsche Antwort: -0,5 Pkt., keine Antwort: 0 Pkt.

Sie können in dieser gesamten Aufgabe nicht weniger als 0 Punkte erreichen.

- a) Bei binären Relationship-Typen ist die (min, max)-Notation für Kardinalitätsrestriktionen präziser und damit ausdrucksstärker als die 1:n- Notation.
- b) Bei der Anwendung des in der Vorlesung vorgestellten Dekompositionsalgorithmus' gehen stets funktionale Abhängigkeiten verloren.

- c) Die vertikale Fragmentierung beim Entwurf verteilter Datenbanken entspricht einer Projektion.
- d) Ein serialisierbarer Schedule heißt auch seriell.
- e) Strikte Schedules vermeiden Dirty-Read-Situationen.
- f) Zwei Datenoperationen in einem Schedule stehen in Konflikt wenn sie zu verschiedenen Transaktionen gehören und auf dem selben Objekt arbeiten.
- g) Im Falle eines Systemfehlers führt der Recovery Manager ein REDO für alle Transaktionen durch, die zum Zeitpunkt des Fehlers noch aktiv waren.
- h) Ein valides XML Dokument ist immer einem DTD oder einem XML Schema zugeordnet.

6 Serialisierbarkeit und Nebenläufigkeit (4 Punkte)

Betrachten Sie die folgenden Schedules:

$$s_1 = r_1(x)r_2(y)w_1(y)w_3(x)c_3r_1(z)w_1(x)c_1r_2(x)w_2(x)c_2$$

$$s_2 = r_1(x)w_3(x)r_2(y)w_1(y)r_1(z)w_1(x)c_1c_3r_2(x)w_2(x)c_2$$

- a) Zeigen oder widerlegen Sie, dass diese beiden Schedules konfliktäquivalent sind. (1 Punkt)
- b) Sind die Schedules konfliktserialisierbar? Begründen Sie Ihre Antwort. (1 Punkt)
- c) Entscheiden Sie für die gegebenen Schedules, in welcher der Klassen RC, ACA oder ST sie sich befinden. Begründen Sie ihre Entscheidung. (2 Punkte)

7 XPath und XQuery (7 Punkte)

Betrachten Sie das folgende XML-Dokument als ein Beispiel eines (wahrscheinlich größeren) XML-Dokumentes gleicher Struktur, das als eine Art Publikationsdatenbank fungiert und Informationen über wissenschaftliche Veröffentlichungen bereitstellt.

```
<pubdb>
  <papier jahr="2007">
    <konferenz>VLDB</konferenz>
    <titel>Model Management and Schema Mappings</titel>
    <authoren>
      <person personId="4" name="Phil Bernstein">
        <email>bernstein@microsoft.com</email>
      </person>
      <person personId="9" name="Howard Ho">
        <email>ho@ibm.com</email>
      </person>
    </authoren>
  </papier>
  <papier jahr="2000">
    <konferenz>SIGMOD</konferenz>
    <titel>A Vision for Management of Complex Models</titel>
    <authoren>
```

```
<person personId="4" name="Phil Bernstein">
  <email>bernstein@microsoft.com</email>
</person>
<person personId="2" name="Alon Halevy">
  <email>halevy@google.com</email>
</person>
<person personId="3" name="Rachel Pottinger">
  <email>pottinger@cs.washington.edu</email>
</person>
</autoren>
</papier>
...
</pubdb>
```

- a) Adressieren Sie durch Angabe entsprechender XPath-Ausdrücke die wie folgt beschriebenen XML-Knoten
- 1.) Alle Titel von Publikationen die im Jahr 2000 veröffentlicht wurden. (2 Punkte)
 - 2.) Alle Publikationen an denen "Alon Halevy" als Author beteiligt ist. (2 Punkte)
- b) Formulieren Sie eine XML-Query-Anfrage (XQuery), die alle Veröffentlichungen auf Konferenzen aus dem Jahr 2007 mit ihrem Titel, dem Erstauthor sowie Konferenz und Jahr aufführt. Der Erstauthor ist der erste in der Liste aller Autoren eines Papiers. Verwenden Sie das folgende Ausgabeformat: (3 Punkte)

```
<publikationen>
  <papier konferenz="VLDB" jahr="2007">
    <titel wert="Model Management and Schema Mappings" />
    <erstauthor>Phil Bernstein</erstauthor>
  </papier>
  ...
</publikationen>
```