

Klausur Datenbanken
Wintersemester 2015/2016
Prof. Dr. Wolfgang May
2. März 2016, 14-16 Uhr
Bearbeitungszeit: 90 Minuten

Vorname:

Nachname:

Matrikelnummer:

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	Schätzung für "4"
Aufgabe 1 (ER-Modell)	18	14
Aufgabe 2 (Transformation in das Relationale Modell)	18	15
Aufgabe 3 (SQL und Relationale Algebra)	40	22
Aufgabe 4 (Verschiedenes)	14	6
Summe	90	57

Note:

Themenstellung: Gemälde in Museen

Alle Klausuraufgaben basieren auf einem gemeinsamen “Auftrag”: In der Klausur soll eine Datenbank über Maler und Gemälde entworfen werden.

1. Maler haben –wie alle Personen– einen Namen (meistens Vorname+Nachname; es gibt keine zwei Personen mit demselben Namen). Ausserdem ist ihr Geburtsdatum, ihr Todesdatum (falls Sie nicht mehr leben), und ihr Geburtsort gespeichert.

Über jeden Maler ist ausserdem gespeichert, wo er in welchem Zeitraum gelebt hat (wobei nur die Jahreszahlen berücksichtigt werden).

2. Zu jedem Ort/Stadt, der in der Datenbank vorkommt, muss das Land angegeben werden, da es mehrere gleichnamige Städte in verschiedenen Ländern geben kann. Innerhalb eines Landes ist der Stadtname eindeutig (ggf. durch Ergänzung wie “Frankfurt (Oder)”).

Paul Cézanne wurde am 19. Januar 1839 in *Aix-en-Provence* in *Frankreich* geboren, wo er auch bis zu seinem Tod am 22. Oktober 1906 lebte.

Vincent van Gogh wurde am 30. März 1853 in *Groot-Zundert* in den *Niederlanden* geboren und ist am 29. Juli 1890 in *Paris* (*Frankreich*) gestorben. Unter anderem hat er von 1885-1887 in *Paris* und 1888 bis 1889 in *Arles* (*Frankreich*) gelebt.

Neo Rauch wurde am 18. April 1960 in *Leipzig* (*Deutschland*) geboren, wo er auch zur Zeit noch lebt.

3. Zu jeder Stadt ist gespeichert, in welcher (geografischen) Landschaft sie liegt. Es gibt keine zwei Landschaften mit demselben Namen. Jede Landschaft liegt in genau einem Land.

Aix-en-Provence und *Arles* liegen in der *Provence*, die komplett in Frankreich liegt.

4. Kunstmuseen haben einen (eindeutigen) Namen und befinden sich in einer Stadt.

Das *Musée d’Orsay* befindet sich in *Paris*.

5. Zu jedem Gemälde ist der Titel gespeichert, und von welchem Maler es gemalt wurde (mehrere Maler können Gemälde mit demselben Titel gemalt haben). Für jedes Bild ist ausserdem gespeichert, wann es gemalt wurde (nur das Jahr), und ggf. in welchem Museum es ausgestellt ist (manche Bilder sind nicht in einem Museum, sondern in Privatbesitz oder verschollen).

Vincent van Gogh malte 1885 das Bild “*Die Kartoffeleesser*”, das sich im *Van Gogh Museum* in *Amsterdam* (*Niederlande*) befindet.

1888 malte er “*Der Maler auf dem Weg nach Tarascon*”, das verschwunden ist (vermutlich im 2. Weltkrieg vernichtet).

6. Auf einigen Bildern sind eine oder mehrere (bekannte) Personen, z.B. Maler (ggf. der Maler selber), aber auch andere bekannte Personen dargestellt. Dieses soll auch – sinnvoll abfragbar, siehe Aufgabe 2 – gespeichert werden:

1870 hat *Paul Cezanne* ein Bild “*Paul Alexis liest Emile Zola vor*” gemalt, auf dem die beiden Schriftsteller *Paul Alexis* und *Emile Zola* dargestellt sind. Das Bild befindet sich im *São Paulo Museum of Art* (in *São Paulo, Brasilien*).

Henri de Toulouse-Lautrec hat 1887 ein Bild mit dem Titel "Vincent van Gogh" gemalt, auf dem ebendieser dargestellt ist. Dieses Bild befindet sich im Van Gogh Museum in Amsterdam.

Vincent van Gogh malte 1887 das Bild "Selbstportrait mit Strohhut", das in der National Gallery of Art (NGA) in Washington (USA) zu sehen ist.

7. Einige Bilder stellen eine bestimmte Landschaft dar; auch dies soll in der Datenbank gespeichert werden:

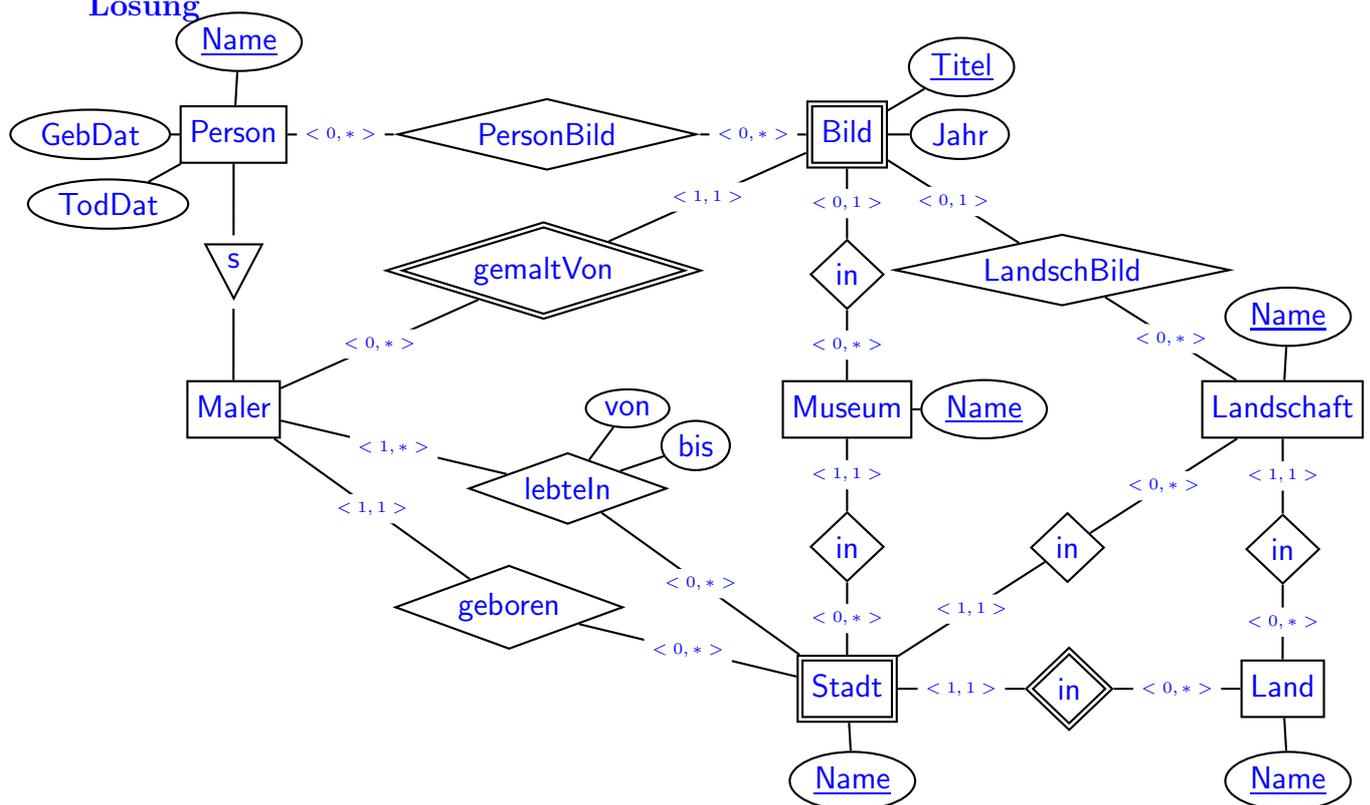
Vincent van Gogh malte 1888 das Bild "Sternennacht über der Rhone", das sich im Musée d'Orsay in Paris befindet. Auf dem Bild ist die Landschaft in der Provence zu sehen.

Paul Cezanne malte 1890 das Bild "Montagne Sainte-Victoire von Gardanne aus gesehen", das ebenfalls die Landschaft in der Provence darstellt, und in der NGA in Washington hängt.

Aufgabe 1 (ER-Modell [18 Punkte])

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

Lösung



- Abgebildete Personen sind "einfach nur" Personen, von denen man das Geburtsdatum nicht weiß, und auch nicht weiss wann sie wo lebten.
- Ob man das Geburtsdatum und -ort bei Person oder bei Maler anlegt, ist egal. Einerseits hat jede Person eines (auch wenn man es nicht weiss), andererseits wird es nur für Maler gespeichert.

- Auf einem Bild kann höchstens eine Landschaft, aber mehrere Personen sein.
- Das Attribut "Jahr" kann man auch bei der Beziehung "gemaltVon" anstatt bei "Bild" anlegen.
- Mögliche Variante: Da in der Aufgabenstellung zugesichert wird, dass jede Landschaft in genau einem Land liegt, kann man Stadt als schwachen Entitätstyp durch die Beziehung zu einer "Landschaft" anstatt zu einem "Land" identifizieren.

In dieser Anwendung spielen die Landschaften eine größere Rolle als das politische Konzept eines Landes. Man spart sich damit einige Joins in den Anfragen.

Aufgabe 2 (Transformation in das Relationale Modell [18 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, Schlüssel 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.

Lösung

Maler				
<u>Name</u>	<u>GebDat</u>	<u>TodDat</u>	<u>GebOrt</u>	<u>GebLand</u>
Paul Cézanne	19.1.1839	22.10.1906	Aix-en-Provence	F
Vincent van Gogh	30.3.1853	29.7.1890	Groot-Zundert	NL
Neo Rauch	18.4.1960	null	Leipzig	D
:	:	:	:	:

lebteIn				
<u>Maler</u>	<u>Stadt</u>	<u>Land</u>	<u>von</u>	<u>bis</u>
Paul Cézanne	Aix-en-Provence	F	1839	1906
Vincent van Gogh	Paris	F	1885	1887
Vincent van Gogh	Arles	F	1888	1889
Vincent van Gogh	Paris	F	1890	1890
Neo Rauch	Leipzig	D	1960	null
:	:	:	:	:

Bild					
<u>Titel</u>	<u>Maler</u>	<u>Jahr</u>	<u>Museum</u>	<u>Landschaft</u>	
Die Kartoffelesser	V. van Gogh	1885	Van Gogh Museum	null	
Sternennacht über der Rhone	V. van Gogh	1888	Musée d'Orsay	Provence	
Der Maler auf dem Weg nach T.	V. van Gogh	1888	null	null	
Montagne Sainte-Victoire v.G.	Paul Cezanne	1890	National Gallery of Art	Provence	
Paul Alexis liest Emile Zola vor	Paul Cezanne	1870	São Paulo M. of Art	null	
:	:	:	:	:	

PersonAufBild		
<u>Bild</u>	<u>Maler</u>	<u>Person</u>
Paul Alexis liest Emile Zola vor	Paul Cezanne	Paul Alexis
Paul Alexis liest Emile Zola vor	Paul Cezanne	Emile Zola
Vincent van Gogh	Henri de Toulouse-Lautrec	Vincent van Gogh
Selbstportrait mit Strohhut	Vincent van Gogh	Vincent van Gogh
:	:	:

Museum		
<u>Name</u>	<u>Stadt</u>	<u>Land</u>
Musée d'Orsay	Paris	F
Van Gogh Museum	Amsterdam	NL
:	:	:

Stadt		
<u>Name</u>	<u>Land</u>	<u>Landschaft</u>
Aix-en-P.	F	Provence
Arles	F	Provence
:	:	:

Landschaft	
<u>Name</u>	<u>Land</u>
Provence	F
Harz	D
:	:

- Personen, die keine Maler sind (und über die man nichts weiss) müssen nicht unbedingt gespeichert sein. Dann darf PersonAufBild.Person kein Fremdschlüssel sein.
Alternative: Tabelle "Person" anstelle "Maler", alle Personen dort mit Namen ablegen.
Vorteil: man kann das Geburtsdatum eintragen, wenn man möchte.
Nachteil: man kann nicht mehr einfach zwischen Malern und sonstigen Personen unterscheiden (Zusatzspalte "istMaler" einfügen, oder man muss es jedesmal per Anfrage, ob ein Bild gespeichert ist, herausfinden).
 - Wenn eine Landschaft in mehreren Ländern liegen könnte (was realistischer ist), müsste man in der Tabelle Landschaft(Name, Land) beide Attribute als Schlüssel auszeichnen.
 - Zu der Tabelle für die Beziehung "lebtIn" kann auch (Maler, Stadt, von) gemeinsam den Primärschlüssel bilden, wenn man mehrere Ortswechsel in einem Jahr modellieren will.
(Maler, von, bis) kann nicht Schlüssel sein, da "bis" null ist, wenn der Maler noch an diesem Ort lebt.
 - Eine Tabelle Land(Name) ist nicht notwendig, kann aber angelegt und als Fremdschlüssel verwendet werden, um z.B. Schreibfehler zu vermeiden.
Durch die dann automatisch erfolgende Indexierung hat man evtl. sogar Laufzeitvorteile.
- b) Geben Sie die CREATE TABLE-Statements für diejenige Tabelle, in der gespeichert ist, in welchen Museen sich die Bilder befinden, so vollständig wie möglich an (6 P).

Lösung

```
CREATE TABLE Bild
( Titel VARCHAR2(100),
  Maler VARCHAR2(50) REFERENCES Maler(name),
  Jahr NUMBER NOT NULL CHECK (Jahr < SYSDATE),
  Museum VARCHAR2(50) REFERENCES Museum(name),
  Landschaft VARCHAR2(20) REFERENCES Landschaft (name),
  CONSTRAINT BildKey PRIMARY KEY (Titel, Maler))
```

Basis 3P + Constraints:
 1/2P
 1/2+1/2P
 1/2P
 1/2P
 1P
 6.5P erreichbar.

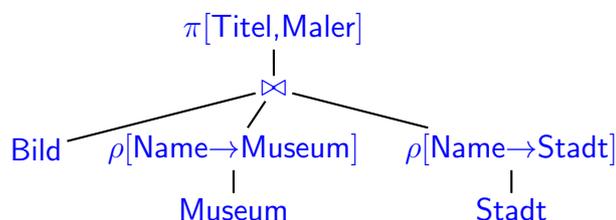
Aufgabe 3 (SQL und Relationale Algebra [40 Punkte])

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, der die Titel und Maler aller Gemälde ausgibt, die sich in einem Museum in derjenigen Landschaft befinden, die auf dem Gemälde dargestellt ist. (2+2 P)

Lösung

```
SELECT Bild.Titel, Bild.Maler
FROM Bild, Museum, Stadt
WHERE Bild.Museum = Museum.Name
      AND Museum.Stadt = Stadt.Name
      AND Museum.Land = Stadt.Land
      AND Bild.Landschaft = Stadt.Landschaft
```

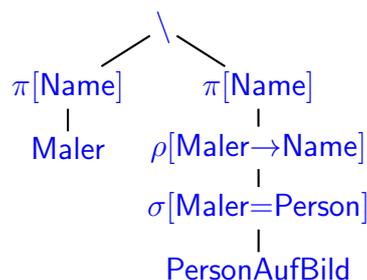


- b) Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, der die Namen aller Maler ausgibt, von denen kein Selbstbildnis in der Datenbank gespeichert ist. (2+2 P)

Lösung

```
SELECT name
FROM Maler
WHERE Maler NOT IN
      (Select name      -- oder Person
       FROM PersonAufBild
       WHERE Maler = person)

( SELECT name
  FROM Maler )
MINUS
      (Select name      -- oder Person
       FROM PersonAufBild
       WHERE Maler = person)
```



- c) Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, die alle Tripel (Malername, Geburtsdatum, Landschaft) angibt, so dass der Maler in (einer oder mehreren Städten in) dieser Landschaft einige Zeit gelebt hat, aber nie ein Bild gemalt hat, das diese Landschaft darstellt. (3+4 P)

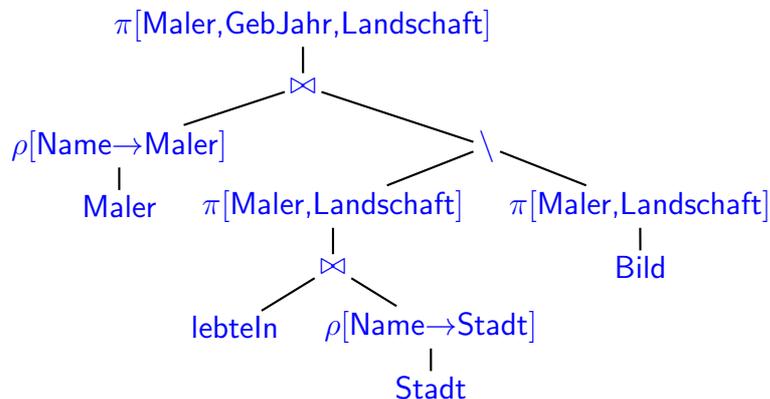
Lösung

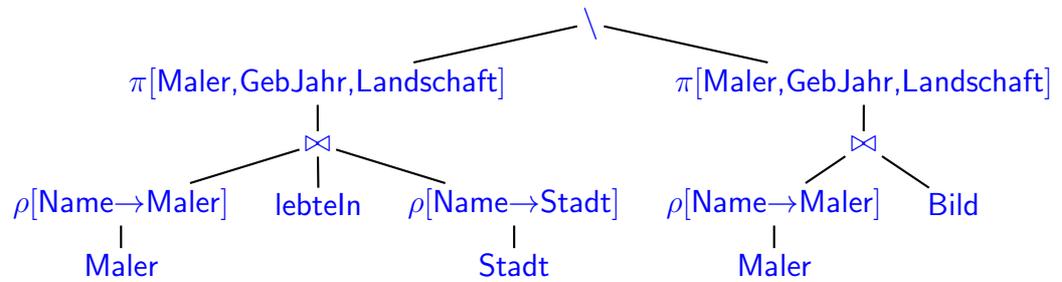
```
SELECT DISTINCT Maler.name, GebJahr, Landschaft
FROM Maler, lebteIn, Stadt
WHERE lebteIn.Maler = Maler.name
  AND lebteIn.Stadt=Stadt.Name AND lebteIn.Land=Stadt.Land
  AND NOT EXISTS
    (SELECT *
     FROM Bild
     WHERE Bild.Maler = Maler.name
       AND Bild.Landschaft = Stadt.Landschaft)
```

```
SELECT DISTINCT Maler.name, GebJahr, Landschaft
FROM Maler, lebteIn, Stadt
WHERE lebteIn.Maler = Maler.name
  AND lebteIn.Stadt=Stadt.Name AND lebteIn.Land=Stadt.Land
  AND Landschaft NOT IN
    (SELECT Landschaft
     FROM Bild
     WHERE Bild.Maler = Maler.name
       AND Bild.Landschaft = Stadt.Landschaft)
```

```
(SELECT Maler.name, GebJahr, Landschaft
 FROM Maler, lebteIn, Stadt
 WHERE lebteIn.Maler = Maler.name
   AND lebteIn.Stadt=Stadt.Name AND lebteIn.Land=Stadt.Land)
MINUS -- eliminiert Duplikate automatisch
(SELECT Maler.name, GebJahr, Landschaft
 FROM Maler, Bild
 WHERE Bild.Maler = Maler.name)
```

DISTINCT ist notwendig, falls ein Maler zwei Zeitabschnitte derselben Landschaft verbraucht hat.





- d) Geben Sie eine SQL-Anfrage an, die für jede Stadt ausgibt, wieviele Bilder, die zwischen 1870 und 1949 gemalt wurden, sich in den Museen dieser Stadt befinden. Es sollen nur Städte ausgegeben werden, in denen sich mindestens 10 solche Bilder befinden. (3 P)

Lösung

```
SELECT Stadt, Land, count(*)
FROM Museum, Bild
WHERE Museum.name = Bild.Museum
      AND Jahr >= 1870 and Jahr <= 1949
GROUP BY Stadt, Land
HAVING count(*) >= 10
```

- e) Wäre Aufgabenteil (d) einfacher oder schwieriger, wenn die Bedingung “Es sollen nur Städte ausgegeben werden, in denen sich mindestens 10 solche Bilder befinden” fehlen würde? Begründen Sie Ihre Antwort. (3 P)

Lösung Sie wäre schwieriger (und etwas unsinniger): Was ist mit Städten, in denen sich keine solchen Bilder befinden? Wenn in einer Stadt kein Museum ist, wäre es sinnvoll, sie wegzulassen; wenn sie aber Museen hat, sollte man wohl eine “0” als Anzahl ausgeben. Das müsste über ein UNION drangebaut werden:

```
... UNION
SELECT Name, Land, 0
FROM Stadt
WHERE (Name, Land) IN (Select Stadt, Land FROM Museum)
      AND NOT EXISTS
      (SELECT * FROM Bild, Museum
       WHERE Museum.name = Bild.Museum
         AND Jahr >= 1870 and Jahr <= 1949)
```

Mit einem Outer-Join geht es auch:

```
SELECT Stadt, Land, count(bld.Titel)
FROM Museum
      OUTER JOIN
      (SELECT * FROM Bild
       WHERE Jahr >= 1870 and Jahr <= 1949) bld
ON Museum.name = bld.Museum
GROUP BY Stadt, Land
```

(Hier ist für Museen, in denen kein Bild aus diesem Zeitraum hängt, ein Tupel (–museum–, –NULL–) dabei, das zum count der (non-null) Bild-Titel nichts beiträgt, und damit für die Stadt eine 0-Anzahl liefert.

- f) Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, die/der die Namen aller Museen ausgibt, die von jedem Maler, der ein Bild, das die Landschaft der *Provence* darstellt, gemalt hat, ein solches Bild (d.h., eines, auf dem auch die *Provence* abgebildet ist) besitzen. (4+4 P)

Lösung

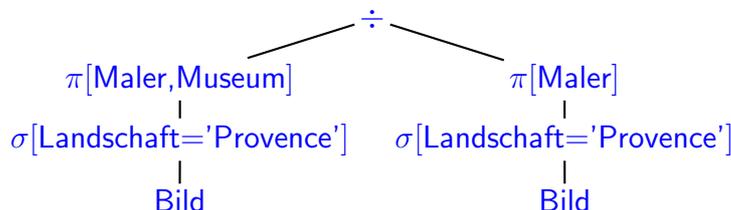
```

SELECT m.name
FROM Museum m
WHERE NOT EXISTS
  (SELECT *           -- Provence-Bild
   FROM Bild b1
   WHERE Landschaft='Provence'
   AND NOT EXISTS    -- Provence-Bild von demselben Maler in diesem Museum
     (SELECT *
      FROM Bild b2
      WHERE Landschaft='Provence'
        AND b2.Maler = b1.Maler
        AND b2.Museum = m.name ))

(SELECT m1.name FROM Museum)
MINUS
(SELECT name
 FROM ( ( SELECT m2.name, b.Maler
          FROM Museum m2, Bild b      -- kartesisches Produkt
          WHERE b.Landschaft = 'Provence' )
 MINUS
 ( SELECT Museum AS name, Maler
   FROM Bild
   WHERE Landschaft = 'Provence'))))

```

Dieser Ausdruck entspricht genau der Formel fuer die Division auf den Vorlesungsfolien



- g) Geben Sie eine SQL-Anfrage **und** einen Algebra-Ausdruck oder -Baum an, die die Titel aller Bilder und den Namen des jeweiligen Malers ausgeben, auf denen eine Landschaft **und** eine oder mehrere bekannte Personen dargestellt ist. (2+2 P)

Lösung

```

SELECT titel, maler
FROM Bild
WHERE landschaft IS NOT NULL
      AND (titel, maler) IN (SELECT bild, maler
                             FROM PersonAufBild)

oder

SELECT DISTINCT titel, maler
FROM Bild, PersonAufBild p
WHERE bild.landschaft IS NOT NULL
      AND Bild.titel = p.bild AND bild.maler = p.maler

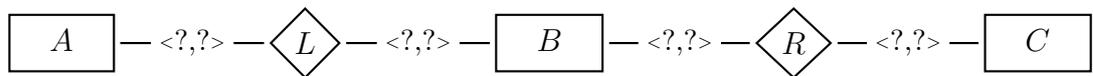
```

DISTINCT ist notwendig, falls auf einem Bild zwei Personen abgebildet sind.



- h) Ein bisschen Theorie (kann auf Basis des vorherigen Aufgabenteils (g) auch real veranschaulicht werden):

Seien $L(A, B)$ und $R(B, C)$ zwei Tabellen, die aus zwei Beziehungen entstanden sind, d.h.



Betrachten Sie den folgenden Ausdruck der relationalen Algebra:

$$(L \bowtie R) \setminus (L \bowtie R)$$

- Geben Sie das Format dieses Ausdrucks an (d.h., welche Spalten hat er (1 P))

Lösung Man erinnert sich, was das Ergebnis des Left Outer Joins ist: Das Inner Join, und dazu alle Werte “links” (L), die “rechts” (R) kein “Gegenstück” finden. Format ist also (A, B, C) (das kann man auch wissen, ohne zu wissen, was \bowtie tut, da das Ergebnisformat von $(L \bowtie R)$ schon (A, B, C) ist, und die linke Seite und das Ergebnis des “ \setminus ” auch dasselbe Format haben müssen).

- Charakterisieren Sie die Menge der Ergebnistupel (textuell oder formal (3 P)).

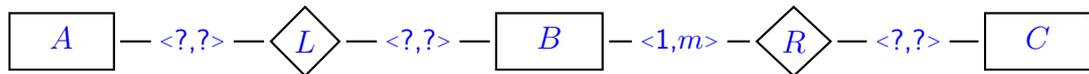
Lösung Das Ergebnis sind alle Tupel (a, b, null) , so dass $L(a, b)$ gilt, und es aber kein c gibt, so dass $R(b, c)$ gilt:

$$\{ \mu \in \text{ Tup}(A, B, C) : \mu[C] = \text{null} \text{ und } (\mu[A], \mu[B]) \in L \text{ und es gibt kein } c, \text{ so dass } (\mu[B], c) \in R \}$$

In Bezug auf Aufgabenteil (c) ist $B = \text{Bild}$ und $(L \bowtie R) \setminus (L \bowtie R)$ alle Bilder, auf denen eine Landschaft, aber keine bekannten Personen zu sehen sind.

- Geben Sie eine (sinnvolle) Bedingung an ein oder mehrere der Fragezeichen in dem obigen kleinen ER-Diagramm an, unter der das Ergebnis garantiert in allen erlaubten Datenbankzuständen leer sein muss (mit kurzer Begründung) ? (3 P)

Lösung Wenn jedes B, das die Beziehung L eingeht (also im Left Outer Join als B-Wert enthalten ist), die Beziehung R auch mindestens einmal eingehen muss (also im Right Outer Join als B-Wert enthalten ist) ist das Ergebnis leer, d.h.



wobei $m \in \{1, 2, 3, \dots\} \cup \{*\}$ ist.

In Bezug auf Aufgabenteil (g) entspräche dies der Situation, dass auf jedem Bild mindestens eine bekannte Person zu sehen ist.

Leer wäre das Ergebnis auch schon, wenn man fordern würde, dass auf jedem Bild, auf dem eine Landschaft zu sehen ist, auch mindestens eine bekannte Person zu sehen ist; diesen Zusammenhang könnte man aber im ER-Diagramm nicht durch die Kardinalitäten ausdrücken (in UML ist es möglich, wenn man OCL-Constraints dazunimmt).

Eine (nicht sinnvolle) hinreichende Bedingung wäre, dass L leer sein muss, d.h., dass zu L irgendwo eine Maximalkardinalität 0 steht.

Aufgabe 4 (Verschiedenes [14 Punkte])

- a) In welchen Spalten treten in Ihrer Datenbank Nullwerte auf? Welche (wahrscheinlich unterschiedliche) Bedeutung repräsentieren sie in der Realität? (4 P.)

Lösung

- falls es eine Tabelle Person gibt, in der alle Personen abgelegt sind: Person.GebDat: ist NULL; wenn es unbekannt ist.
- Person.TodDat/Maler.TodDat: ist bei allgemeinen Personen NULL, wenn es nicht bekannt ist; bei Malern ist es genau dann NULL, wenn der Maler noch lebt (im Beispiel *Neo Rauch*).
- lebtIn.bis: ist null, wenn der Maler noch dort wohnt (im Beispiel *Neo Rauch* in Leipzig).
- Bild.Museum: ist NULL, wenn das Bild in Privatbesitz oder verschwunden ist.
- Falls die abgebildete Landschaft mit in der Tabelle "Bild" abgelegt ist: ist NULL, wenn entweder keine Landschaft abgebildet ist, oder die abgebildete Landschaft nicht relevant ist.

- b) In der *National Gallery of Art (NGA)* wird zunehmend der Platz knapp. Deswegen wird umstrukturiert:

- Eine *National Gallery of Modern Painting (NGMP)* wird gegründet, die sich ebenfalls in Washington befindet.
- Alle Gemälde aus der *National Gallery of Art*, die zwischen 1870 und 1949 entstanden sind, werden dorthin abgegeben.

Geben Sie die entsprechenden SQL-Befehle an (3 P).

Lösung

```
INSERT INTO Museum VALUES('National Gallery of Modern Painting',  
                             'Washington', 'USA');           (1 P.)
```

```
UPDATE Bild  
SET Museum = 'National Gallery of Modern Painting'  
WHERE Museum = 'National Gallery of Art'  
AND Jahr BETWEEN 1870 AND 1949;           (2 P.)
```

- c) Die Transaktion in Aufgabenteil (b) betrifft einen "größeren" Datenbestand (die Datenbank enthält ein paar hunderttausend Einträge, und es sind evtl. mehrere hundert Bilder davon betroffen). Da lohnt sich schon ein Blick auf die Auswertung und Effizienz. (4 P)

- Skizzieren Sie, wie das Statement algorithmisch auf der in Aufgabe 2 erstellten Datenbank abläuft;
- Welche internen, datenstrukturbasierten Hilfsmittel der Datenbank können dabei die Effizienz erhöhen (Hinweis: eines ist sicher da, weitere kann der Admin angelegt/konfiguriert haben)? Beschreiben Sie kurz deren Nutzung.

Lösung Der Vollständigkeit halber zuerst: das Tupel für die *National Gallery of Modern Painting* wird angelegt, um als Fremdschlüssel für die dort einzutragenden Bilder verfügbar zu sein. (man erinnert sich: auf Schlüsseln und Fremdschlüsseln gibt es Indexe.)

Grob: Eine Schleife läuft über die durch die WHERE-Bedingung spezifizierten Tupel der Tabelle "Bild".

Sequentiell über alle Einträge der gesamten Tabelle "Bild" zu laufen und zu prüfen, ob sie die WHERE-Bedingung erfüllen wäre sehr ineffizient.

- Wegen des Fremdschlüssels auf Museum.Name ist darüber auf jeden Fall ein Index angelegt (Baum, Hash, oder eine Bitmap über Hash zugreifbar - es ist ja eine überschaubare Anzahl von Museen).

Damit muss man auf jeden Fall *höchstens* alle Bilder in der NGA betrachten (und noch nach den betroffenen Jahren selektieren).

- Für Museum.Jahr kann auch ein Index, hier sinnvollerweise ein Baumindex, existieren.

(kann der Admin mit `CREATE index bildjahr ON Bild(jahr)` machen. Damit kann man die beiden Index-Ergebnismengen schneiden, und muss nur noch auf diese Tupel zugreifen.

- Wenn die Tabelle zusätzlich noch geordnet (und damit auch gruppiert) gespeichert ist, reduziert das die Anzahl der Seitenzugriffe, da man mehrere (bis viele) gesuchte Bilder auf derselben Seite hat.

- * geordnet nach Museum: nur ein paar Seitenzugriffe notwendig (nur die Seiten, wo irgendwas aus der NGA drauf ist),

- * geordnet nach Jahr: nur die betreffenden Jahre (aber alle Museen),

- * geordnet nach (Museum, Jahr) oder (Jahr, Museum): noch besser;

- * geordnet nach Malern: besser als nichts. bringt überall da etwas, wo von einem Maler mehrere Bilder in der NGA auf der selben Seite sind.

- * im Gegensatz zu den meisten Datenbanken ist eine geordnete Speicherung nach Jahren oder Malern sogar mit wenig Verschnitt machbar, da sich dort bei den meisten nichts mehr ändert.

(gilt analog auch für clusterbasierte Speicherung nach Maler oder Museum).

- d) Während der Transaktion in Aufgabenteil (b) stürzt die DB softwaremäßig ab. Einige Bilder "sind" schon in die NGMP umgetragen worden, andere sind noch in der NGA.

Skizzieren Sie, wie die Datenbank nach einem Neustart in einen zulässigen Zustand zurückfindet. (3 P)

Lösung

- Da es nur ein Softwareabsturz war, und damit der aktuelle halbfertige Zustand im Speicher steht (im einzelnen kann dies noch davon abhängen, was zum Absturz nur im Cache geändert war, und was schon in den Hintergrundspeicher übertragen wurde), und alle Einzelaktionen im Logfile stehen, genügt es, die (und alle anderen) nicht-committete Transaktion zurückzurollen.

- Alternative: Backup der letzten Nacht einspielen, Logfile nehmen, alle committeten Transaktionen nachspielen.