CHALMERS UNIVERSITY OF TECHNOLOGY
Department of Computer Science and Engineering

Examination in Databases, TDA357/DIT620

Thursday 16 April 2009, 08:30-12:30

Examiner:	Graham Kemp (telephone 772 5411, room 6475 EDIT) The examiner will visit the exam room at 09:30 and 11:30.
Results:	Will be published by the end of April at the latest.
Exam review:	see course web page for time and place http://www.cs.chalmers.se/Cs/Grundutb/Kurser/dbas/DbasHT2008/index.cgi
Grades:	Grades for Chalmers students (TDA357) are normally determined as follows: ≥ 48 for grade 5; ≥ 36 for grade 4; ≥ 24 for grade 3. Grades for GU students (DIT620) are normally determined as follows: ≥ 42 for grade VG; ≥ 24 for grade G.
Help material:	One A4 sheet with hand-written notes. You may write on both sides of that sheet. That sheet must be handed in with your answers to the exam questions. English language dictionaries are allowed.

Specific instructions:

- Please answer in English where possible. You may clarify your answers in Swedish if you are not confident you have expressed yourself correctly in English.
- Begin the answer to each question on a new page.
- Write clearly; unreadable = wrong!
- Fewer points are given for unnecessarily complicated solutions.
- Indicate clearly if you make any assumptions that are not given in the question.
- Write the page number and question number on every page.

Question 1. Consider the following domain description.

10 p

> The items sold by a chain of furniture stores are identified by the item's model name and colour. Each store in the chain is in a different city. Within a particular store, the place where an item can be found is given by a section name and a shelf number (e.g. section "A", shelf "4"). While most items are sold in a single package, some items are so large that they are sold in two or more packages. Each package has a part number (e.g. if an item is sold in two packages, these packages will be identified by part numbers "1" and "2"). Each package also has a length, width, height and weight. Each purchase made from a store consists of one or more items and may include several instances of the same item (e.g. 4 chairs with the same model name and colour). The owners of the chain of furniture stores want to record which purchases were made from each store. To help with this, each purchase is given a unique purchase reference number. When recording a purchase, the quantity of each item bought must be recorded, and also the purchase date. In the case of purchases made using a credit card, the credit card number and credit card type must also be recorded.

a) Draw an E-R diagram that correctly models this domain. (5p)

b) Translate this E-R diagram into a set of relations, clearly marking all references and keys. (5p)

Question 2. 11 p

a) Suppose we have relation $R(A, B, C, D)$ and functional dependencies $BD \rightarrow A$, $C \rightarrow D$.

i) By considering the closures of all subsets of attributes, find **all** non-trivial FDs, superkeys and keys. State which FDs violate BCNF. (4p)

ii) Decompose relation R to BCNF in two different ways:

— by first decomposing on $BD \rightarrow A$

— by first decomposing on $C \rightarrow D$.

Show each step in the normalisation process, and at each step indicate which functional dependency is being used. (3p)

b) In the following relation, column **teacher** contains the name(s) of the teachers who teach the course whose code is in column **course**. Column **author** contains the names of the authors of the course textbook.

teacher	**author**	**course**
Kemp	Garcia-Molina	TDA357
Kemp	Ullman	TDA357
Kemp	Widom	TDA357
Broberg	Garcia-Molina	TDA357
Broberg	Ullman	TDA357
Broberg	Widom	TDA357
Holmström	Tardos	TIN092
Holmström	Kleinberg	TIN092

State which multi-valued dependencies hold for this relation.

Suppose tuple (Kemp, Tardos, TIN092) is added to this relation.
Which other tuple(s) must be added to the relation?

Decompose this relation to 4th normal form, and show the rows in the resulting relations.

(4p)

Question 3. Suppose relation Employees is as follows:
5 p

name	branch	salary
Andersson	3	20000
Johnsson	3	25000
Larsson	3	32000
Persson	2	28000
Svensson	2	35000

and relation Branches is as follows:

number	city
1	Stockholm
2	Paris
3	London
4	Berlin
5	Rome

a) Write two different relational algebra expressions that evaluate to:

name	city
Larsson	London
Svensson	Paris

i) one of these relational algebra expressions must use the Cartesian product operator.

ii) one of these relational algebra expressions must **not** use the Cartesian product operator.

(3p)

b) Write a relational algebra expression that computes the average salary at each branch, and sorts the results in increasing order of the average salary.

(2p)

Question 4. 8 p

Consider the following relation that contains summary information about the sales made by a company in different countries each month:

$Sales(item, quantity, country, year, month)$

Assume that the month is represented by an integer (1-12).

a) Write an SQL query that finds the total quantity of item "p001" that was sold in the first half (January to June) of 2008. (2p)

b) Write an SQL query that finds those items that had higher sales in Norway than in Denmark in March 2009. (2p)

c) Create a view $V(item, quantity, country, year)$ which contains the total amount of sales on each item in each country in each year. (2p)

d) Write an SQL query that finds those countries where the sales of item "p001" has been higher December than in July in every year for which there are sales records. (2p)

Question 5. 10 p

A database system for managing information about papers submitted to a conference has the following relations:

Papers(refNo, title)
PaperAuthors(refNo, email, name, institute)
Reviewers(refNo, email)
ProgrammeCommittee(email, name, institute)

Each paper submitted to the conference is assigned a unique reference number. Each paper can have several authors, and each author may be an author of more than one paper. Members of the programme committee are each allocated several papers to review, and each paper will be reviewed by several different reviewers. Author names and reviewer names might not be unique, however email addresses are guaranteed to be unique. The value stored in the *institute* columns is the name of the institute where the author or programme committee member works.

a) i) Suggest keys and references for these relations.
Write SQL statements that create these relations with constraints in a DBMS.
(4p)

ii) Motivate the update and delete policies that you choose for foreign key in your answer to part (i).
(2p)

b) A member of the programme committee might be an author of one or more papers that are submitted to the conference. Write an assertion that checks that no reviewer is allocated a paper to review where they are one of the authors.
(2p)

c) Define a view, *Conflicts(refNo, email)*, that lists conflicts of interest that occur when a programme committee member works at the same institute as one of the paper's authors. In this view, *refNo* is a paper reference number and *email* is the email address of a programme committee member.
(2p)

Question 6. 4 p

Suppose we have relation $Accounts(accNo, custNo, balance)$, and that this relation contains the following tuples:

accNo	custNo	balance
A001	C001	500
A002	C001	500

Suppose that customer C001 transfers 100 from account A001 to A002 (transaction T1) and withdraws 200 account A002 (transaction T2). Suppose also that a bank official runs a query to find the total amount in customer C001's accounts (transaction T3). These transactions consist of the following steps:

T1:

$Step\ T1_A:\ get\ balance\ of\ A001\ into\ X$
$Step\ T1_B:\ set\ balance\ of\ A001\ to\ X - 100$
$Step\ T1_C:\ get\ balance\ of\ A002\ into\ Y$
$Step\ T1_D:\ set\ balance\ of\ A002\ to\ Y + 100$

T2:

$Step\ T2_A:\ get\ balance\ of\ A002\ into\ Y$
$Step\ T2_B:\ set\ balance\ of\ A002\ to\ Y - 200$

T3:

$Step\ T3_A:\ get\ balance\ of\ A001\ into\ V$
$Step\ T3_B:\ get\ balance\ of\ A002\ into\ W$
$Step\ T3_C:\ print\ V + W$

a) What are the possible outcomes if all three transactions are run as serializable transactions? Explain your answer. (2p)

b) What are the possible outcomes if T1 and T3 are run as serializable transactions, but T2 is not? Explain your answer. (2p)

Question 7. 6 p

a) Suppose relation R contains the following rows:

a1	d	80
a1	k	20
a1	r	30
a2	m	60
a4	g	90
a5	d	60
a6	m	40
a7	c	80
a8	k	60
a8	s	30

i) Draw a picture that shows a *dense primary index* on the first column of relation R. (2p)

ii) Draw a picture that shows a *secondary index* on one of the other columns of relation R. (2p)

b) Give two reasons why creating an index on a particular column in a particular relation could result in worse overall performance. (2p)

Question 8. Consider the following piece of XML.
6 p

```
<?xml version="1.0" standalone="no">

<Universities>
 <City name="Gothenburg">
  <University name="CTH" type="Technical" />
  <University name="GU"  type="Comprehensive" />
 </City>
 <City name="Stockholm">
  <University name="KTH" type="Technical" />
  <University name="SU"  type="Comprehensive" />
  <University name="KI"  type="Medical" />
 </City>
 <City name="Uppsala">
  <University name="UU"  type="Comprehensive" />
 </City>
</Universities>
```

a) Write a Document Type Definition (DTD) for this XML example. (2p)

b) Write XPath expressions that find:
   - the names of all universities;
   - all University elements for the technical universities. (2p)

c) Write an XQuery expression that gives the city name and university name of the technical universities, i.e.

```
<Result>Gothenburg: CTH</Result>
<Result>Stockholm: KTH</Result>
```

(2p)

CHALMERS UNIVERSITY OF TECHNOLOGY
Department of Computer Science and Engineering

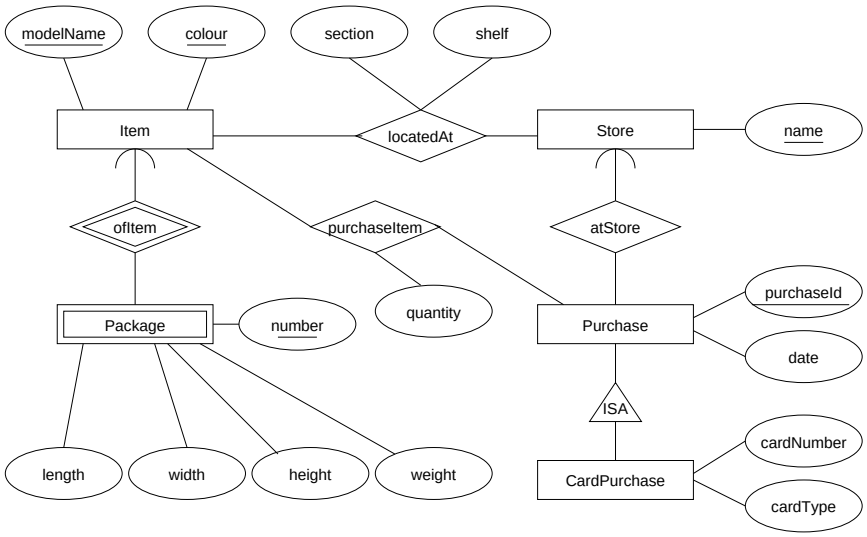
Examination in Databases, TDA357/DIT620

Thursday 16 April 2009, 08:30-12:30

Solutions

Updated 2011-12-12

Question 1. a) E-R diagram:
10 p



b) $Items(\underline{modelName}, \underline{colour})$

$Packages(\underline{modelName}, \underline{colour}, \underline{number}, length, width, height, weight)$
$(modelName, colour) \rightarrow Items.(modelName, colour)$

$Stores(\underline{name})$

$LocatedAt(\underline{modelName}, \underline{colour}, \underline{store}, section, shelf)$
$(modelName, colour) \rightarrow Items.(modelName, colour)$
$store \rightarrow Stores.name$

$Purchases(\underline{purchaseId}, date, atStore)$
$atStore \rightarrow Stores.name$

$PurchaseItems(\underline{purchaseId}, \underline{modelName}, \underline{colour}, quantity)$
$purchaseId \rightarrow Purchases.purchaseId$
$(modelName, colour) \rightarrow Items.(modelName, colour)$

$CardPurchases(\underline{purchaseId}, cardNumber, cardType)$
$purchaseId \rightarrow Purchases.purchaseId$

Question 2. 11 p

a) i) After considering the closures of all subsets of attributes, we find the following additional non-trivial FDs:

$AC \rightarrow D$
$BC \rightarrow A$
$BC \rightarrow D$
$ABC \rightarrow D$
$BCD \rightarrow A$

Superkeys are: BC, ABC, BCD, ABCD. There is one keys: BC.
FDs that violate BCNF:

$BD \rightarrow A$
$C \rightarrow D$
$AC \rightarrow D$

ii) — By first decomposing on $BD \rightarrow A$, we get $R_1(A, B, D)$ and $R_2(B, C, D)$. R_2 is not in BCNF, so we must decompose further.

— By first decomposing on $C \rightarrow D$, we get $R_1(C, D)$ and $R_2(A, B, C)$.

b) MVDs:

$course \twoheadrightarrow teacher$
$course \twoheadrightarrow author$

(Kemp,Kleinberg,TIN092) must also be added.

teacher	**course**
Kemp	TDA357
Broberg	TDA357
Holmström	TIN092
Kemp	TIN092

author	**course**
Garcia-Molina	TDA357
Ullman	TDA357
Widom	TDA357
Tardos	TIN092
Kleinberg	TIN092

Question 3. 5 p

a) i) $\pi_{name,city}(\sigma_{salary>30000 AND branch=number}(Employees \times Branches))$

ii) $\pi_{name,city}(\sigma_{salary>30000}(Employees) \bowtie_{branch=number} Branches)$

b) $\tau_{avSal}(\gamma_{branch,AVG(salary)\rightarrow avSal}(Employees)$

Question 4. 8 p

a)
```
SELECT  SUM(quantity)
FROM    Sales
WHERE   item='p001' AND year=2008 AND month<7
```

b)
```
SELECT  N.item
FROM    Sales N, Sales D
WHERE   N.item = D.item
        N.country = 'Norway'
        AND D.country = 'Denmark'
        AND N.Year = 2009
        AND D.Year = 2009
        AND N.Month = 3
        AND D.Month = 3
        AND N.quantity > D.quantity
```

c)
```
CREATE VIEW V AS
  SELECT        item, SUM(quantity) AS quantity, country, year
  FROM          Sales
  GROUP BY      item, country, year
```

d)
```
SELECT  DISTINCT C.country
FROM    Sales C
WHERE   NOT EXISTS (
        SELECT  *
        FROM    Sales J, Sales D
        WHERE   J.item = 'p001'
                AND D.item = 'p001'
                AND J.country = C.country
                AND D.country = C.country
                AND J.quantity <= D.quantity )
```

Question 5. a) i) 10 p

$Papers(\underline{refNo}, title)$
$PaperAuthors(\underline{refNo}, \underline{email}, name, institute)$
$refNo \rightarrow Papers.refNo$
$Reviewers(\underline{refNo}, \underline{email})$
$refNo \rightarrow Papers.refNo$
$email \rightarrow ProgrammeCommittee.email$
$ProgrammeCommittee(\underline{email}, name, institute)$

```
CREATE TABLE Papers (
    refNo            INT PRIMARY KEY,
    title            VARCHAR(100)
);

CREATE TABLE PaperAuthors (
    refNo            INT,
    email            VARCHAR(50),
    name             VARCHAR(30),
    institute        VARCHAR(30),
    PRIMARY KEY (refNo, email)
    FOREIGN KEY refNo REFERENCES Papers(refNo)
        ON DELETE CASCADE
        ON UPDATE CASCADE
};

CREATE TABLE ProgrammeCommittee (
    email            VARCHAR(50) PRIMARY KEY,
    institute        VARCHAR(30)
);

CREATE TABLE Reviewers (
    refNo            INT,
    email            VARCHAR(50),
    PRIMARY KEY (refNo, email)
    FOREIGN KEY refNo REFERENCES Papers(refNo)
        ON DELETE CASCADE
        ON UPDATE CASCADE
    FOREIGN KEY email REFERENCES ProgrammeCommittee(email)
        ON DELETE CASCADE
        ON UPDATE CASCADE
);
```

ii) Here are some suggestions, but some other policies will be accepted if these are well motivated.
If a paper's reference number is updated, then we want to propagate that change to the referencing relations. If a paper is deleted from the Papers relation, then information about the paper's authors should be deleted from the PaperAuthors relation, and records about reviewers assigned to review that paper should be deleted from the Reviewers relation.
If a programme committee member's e-mail address is updated, then we want to propagate that change to the Reviewers relation. If a programme committee member is deleted from the database, then we want all records about that programme committee member to be deleted from the Reviewers relation.

b)
```
CREATE ASSERTION NotReviewingOwnPaper CHECK
  ( NOT EXISTS (
        SELECT *
        FROM   Reviewers R, PaperAuthors A
        WHERE  A.refNo = R.refNo
               AND A.email = R.email
  ) );
```

c)
```
CREATE VIEW Conflicts AS
  SELECT  refNo, M.email as email
  FROM    ProgrammeCommittee M, PaperAuthors A
  WHERE   M.institute = A.institute
```

Question 6. 4 p

a) There are 6 possible orderings of transactions T1, T2 and T3, and all result in a balance of 400 in both accounts. In those cases where T3 executes before T2, the value printed is 1000. In other cases, the value printed is 800.

b) All of the outcomes in part (a) are still possible. Additionally, T2 may see data written by T1 as T1 writes it. So T2 can get value 500 or 600 into Y.

Question 7. 6 p

a) See lecture slides for "Lecture 12, Indexes" at http://www.cs.chalmers.se/Cs/Grundutb/Kurser/dbas/DbasHT2008/index.cgi/Lecture_Notes (in particular: page 3, slide 4; page 4, slide 1)

b) Various answers will be accepted. For example, consider the effect of using an index if the relation is so small that it fits in a single disc block. Also, consider the case where there are many more insert operations than data access requests. Solutions should discuss the disc blocks that might need to be transferred between disc and main memory in these cases (or any other case that is suggested in answer to this question).

Question 8. 6 p

a)
```
<!DOCTYPE Universities [

<!ELEMENT Universities (City*)>
<!ELEMENT City (University*)>
<!ELEMENT University EMPTY>

<!ATTLIST City
  name  ID      #REQUIRED >

<!ATTLIST University
  name  ID      #REQUIRED
  type  CDATA   #REQUIRED >
] >
```

b)
```
/Universities/City/University/@name

/Universities/City/University[@type="Technical"]
```

c)
```
FOR     $c IN /Universities/City
FOR     $u IN $c/University
WHERE   $u/@type = "Technical"
RETURN  <Result>{$c/@name}: {$u/@name}</Result>
```