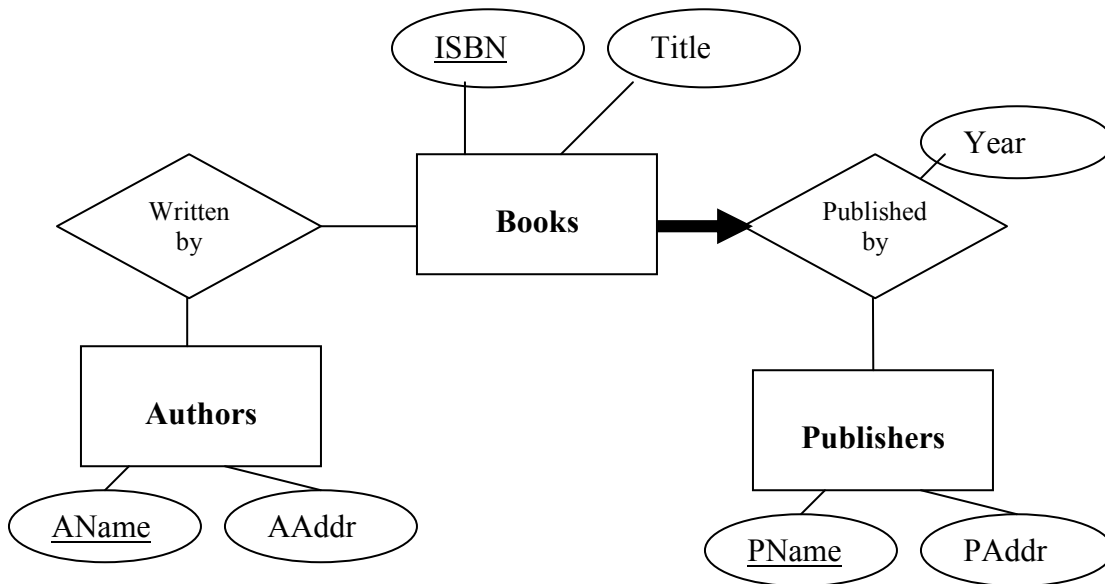


Practice Problem Set: Database Design and Transactions

Here are some sample problems on the material covered since Midterm 2 – these are intended as study questions for the final. Answers are at the end.

1. Write SQL statements to generate the relational tables required to represent the following ER diagram. You do not need to worry about normalization, but you do need to show all “not null”, primary key, and foreign key constraints:



2. Consider relation $R = (A, B, C, D, E, F)$ and the set of functional dependencies:

- $AB \rightarrow C$
- $BC \rightarrow A$
- $BC \rightarrow D$
- $D \rightarrow E$
- $CF \rightarrow B$

List the Candidate Keys of R (there are two).

3. Using Armstrong’s Axioms, show that the F.D.s in Question 2 imply that $AB \rightarrow E$. For each step be sure to indicate which axioms and other F.D.s you are using.

4. Is relation R (from Question 2) in BCNF? If yes, say why. If not, then decompose R into two or more relations that are all in BCNF. Your decomposition must be lossless, but does not need to preserve dependencies.
5. Consider the following schedule:
 r1(A), w1(A), r1(B), r2(A), w3(B), r2(C), r4(B), r4(D), w4(D), r2(D)
 Is the schedule serializable? If so, show an equivalent serial transaction order. If not, describe why not. (note: “r1(A)” means transaction 1 read item A).

For questions 6-8, consider the execution of the ARIES recovery algorithm given the following log (assume a checkpoint is completed before LSN 0 and the Dirty Page Table (DPT) and Transaction Tables for that checkpoint are empty):

LSN	Log Record
10	Update: T1 writes P1
20	Update: T2 writes P3
30	T1 commit
40	Update: T3 writes P4
50	Update: T2 writes P1
60	T1 end
70	Update: T3 writes P2
80	T2 abort
X - crash, restart	

6. During Analysis: a) What log records are read? b) What are the contents of the Dirty Page Table (DPT) and the transaction table at the end of the analysis stage?
7. During Redo: a) What log records are read? b) What data pages are read? c) What operations are redone (assuming no updates made it out to disk before the crash)?
8. During Undo: a) What log records are read? b) What operations are undone? c) Show any new log records that are written (for CLRs, be sure to show the undoNextLSN)?

In addition – it is important to make sure that you understand the Write-Ahead-Log (WAL) protocol (ARIES depends on this for correctness) and how it enables the flexible buffer management policy of “Steal/No Force”. See the lecture notes and the book for more info on this.

ANSWERS TO THE PRACTICE QUESTIONS

1 ER to Relational

```
CREATE TABLE Authors(  
  Aname VARCHAR,  
  AAddr VARCHAR,  
  PRIMARY KEY (Aname)  
)
```

```
CREATE TABLE Publishers(  
  Pname VARCHAR,  
  PAddr VARCHAR,  
  PRIMARY KEY (Pname)
```

```
CREATE TABLE Books(  
  ISBN VARCHAR,  
  Title VARCHAR,  
  Year INT,  
  PName VARCHAR NOT NULL,  
  PRIMARY KEY (ISBN)  
  FOREIGN KEY (PName) REFERENCES Publishers)
```

```
CREATE TABLE WrittenBy(  
  AName VARCHAR,  
  ISBN VARCHAR,  
  PRIMARY KEY (AName, ISBN),  
  FOREIGN KEY (AName) REFERENCES Authors,  
  FOREIGN KEY (ISBN) REFERENCES Books)
```

Note that only four tables are needed, as the "Published_by" relationship can be embedded in the "Books" table due to the one-to-many constraint.

Note that for grading purposes, the data types don't really matter, as long as they are reasonable.

2) Candidate Keys

CF and ABF

3) Functional Dependencies

1) AB->C given

- 2) $AB \rightarrow BC$ augmentation of (1) with B
- 3) $BC \rightarrow D$ given
- 4) $AB \rightarrow D$ transitivity of (2) and (3)
- 5) $D \rightarrow E$ given
- 6) $AB \rightarrow E$ transitivity of (4) and (5)

(there are other correct ways to get it, too).

4) BCNF

R is not in BCNF. Any one of $AB \rightarrow C$, $BC \rightarrow A$, $BC \rightarrow D$, $D \rightarrow E$ is sufficient to show this as none of those left-hand sides are keys, and none of them are trivial.

Decompose into: ABC ABDF and DE

You get this by first decomposing into ABC and ABDEF (lossless since AB is a key for ABC)

Then decomposing ABDEF into ABDF and DE (lossless since D is a key for DE).

(there are likely other correct decompositions as well, but you need at least three relations)

5) Concurrency Control

Yes, it is conflict serializable (no cycles in the dependency graph), and is equivalent to the order: T1, T3, T4, T2

6) Recovery - Analysis

a) All records since last checkpoint, which means all the shown entries, are read.

b)

Transaction Table

Trans	Last LSN	stat
T3	70	running
T2	80	aborting

Dirty Page Table

PageID	recLSN
P1	10
P3	20
P4	40
P2	70

7) Recovery - Redo

a) Since the smallest recLSN is 10, REDO starts there.

b) All pages in the dirty page table: P1, P3, P4, P2 are read from disk.

c)

Assuming no updates made it out to disk, all updates are redone, so:

10, 20, 40, 50, and 70

8) Recovery - Undo

a) starting from 80 and 70, we will read: 80, 70, 50, 40, and 20

b) undone operations will be: 70, 50, 40, and 20

c) New log records are (LSNs aren't important except that they must be increasing):

90 CLR T3 LSN = 70; undoNextLSN = 40

100 CLR T2 LSN = 50; undoNextLSN = 20

110 CLR T3 LSN = 40; undoNextLSN = null

120 T3 end

130 CLR T2 LSN = 20; undoNextLSN = null

140 T2 end