Exercises Week 3:
ER Model to Relations &
File Systems & File Layouts

Exercise 3.1

(a) Consider the university database from Exercise 1.2 (week 1) and the ER diagram you designed. Write SQL statements to create the corresponding relations and capture as many of the constraints as possible. If you cannot capture some constraints, explain why.

(b) Assume the following example:

- Professor “Marc Johnson” with SSN “P1”, age 48, rank 50, and research specialty “Databases”.
- Professor “Anna Pacson” with SSN “P2”, age 45, rank 50, and research specialty “Communications”.
- Professor “Paul Lonson” with SSN “P3”, age 44, rank 45, and research specialty “Smart homes”.
- Department “Computer Science” with department number 1, and main office “CAMPUS1”. The department is run by Professor “Marc Johnson”.
- Department “Electrical Engineering” with department number 2, and main office “CAMPUS2”. The department is run by Professor “Anna Pacson”.
- Department “Architecture” with department number 3, and main office “CAMPUS3”. The department is run by Professor “Paul Lonson”.
- Professor “Marc Johnson” works in the “Computer Science” department for 50% of his time, in the “Electrical Engineering” department for 25% of his time, and in the “Architecture” department for 25% of his time.
- Professor “Anna Pacson” works in the “Electrical Engineering” department for 100% of her time.
- Professor “Paul Lonson” works in the “Architecture” department for 100% of his time.
- Project with project number 1, sponsor “NSF”, starting data 01/01/2015, ending data 01/01/2020, budget $800000.
- Graduate student “John Marcson” with SSN “S1”, age 30, degree program “Postdoc”, major in the “Computer Science” department, and without a supervisor graduate student (he advises himself).
- Professor “Marc Johnson” is the principal investigator of the project with project number 1.
- There are no co-investigators for the project with project number 1. (Note: For this exercise, you can assume that a project can have no co-investigators).
- The project with project number 1 has one research assistant, the graduate student

"John Marcson". The research assistant is supervised by Professor "Marc Johnson". Based on your solution to exercise (a), write the tuples that each relation contains.

Exercise 3.2: Consider the ER diagram that you designed for the Prescriptions-R-X chain of pharmacies in Exercise 1.3 (week 1). Define relations corresponding to the entity sets and relationship sets in your design using SQL.

Exercise 3.3: Consider a relation stored as a randomly ordered file and a single unclustered index on the field *sal* of the relation. If you want to retrieve all records with *sal* > 20, is using the index always the best alternative? Explain.

Exercise 3.4: Alphabet Books is a big publishing company that keeps information about their books in the following table:

BOOK(***BookID***, author, title, genre, year, price)

The size of each record is 200B (*BookID* is 8B, *author* is 64B, *title* is 100B, *genre* is 4B, *year* is 8B and *price* is 16B). The records are organized in 4KB pages (each page stores an integer number of records, as well as 48 bytes of header metadata and 48 bytes of footer metadata). The company was founded in 1974, and since then they have been publishing 400 books per year. Book information about the year's productions is only recorded at the end of each year and 2013 is the last year recorded in the table. You are hired as the DBA. Primary key of the table is BookID and data is sorted on the primary key.

Answer the questions using only numbers (and, for complex calculations, arithmetic operators), but no variables (e.g., $log_{17}(200)$ is acceptable, but not $log_X(N)$). Unless explicitly noted, nothing is in main memory. "I/O cost" means number of disk page accesses.

Data is stored on flash devices using PAX page layout. Each page consists of 1KB **subpages** that are the unit of transfer for flash devices.

Clarification 1: One can read either the entire 4KB page or a 1KB subpage at a time starting from byte 0, byte 1024, byte 2048 or byte 3072. One cannot read starting from, for example, byte 500 to 500+1024.

Clarification 2: Notice the difference between a PAX minipage and a flash subpage. A minipage (of a PAX page) contains values from only one attribute. A subpage is a 1KB portion of the 4KB flash page.

Clarification 3: Assume that BookIDs are assigned sequentially and are correlated with years.

a. How many tuples can a page hold? How many pages are needed in total? b. Calculate the offsets (from the beginning of the page) of each minipage assuming that minipages and data inside them are stored contiguously after the metadata. Explain in which subpage each minipage is stored.
c. How much data (i.e., how many bytes) would the system read for the following query (Explain):

SELECT * FROM BOOK WHERE BookID = 512

d. How much data (i.e., how many bytes) would the system read for the following query (Explain):
SELECT AVERAGE(price) FROM BOOK WHERE year >= 2004;

e. How much data (i.e., how many bytes) would the system read for the following query (Explain):
SELECT author, title FROM BOOK WHERE year >= 1994;

Exercise 3.5: What main conclusions can you draw from the discussion of the basic file organizations discussed in Section 13.2 and 13.3 of "Database System Concepts (SEVENTH EDITION)2 by Abraham Silberschatz, Henry F. Korth, and S. Sudarshan.? Which of the five organizations would you choose for a file where the most frequent operations are as follows?

a. Search for records based on a range of field values.
b. Perform inserts and scans, where the order of records does not matter.
c. Search for a record based on a particular field value.