CSE 444 Quiz 1 and 2
Winter 2019
February 12, 2019

- Please read all instructions carefully.

- Write your name and UW student number below.

- This quiz is meant to be a combination of two smaller quizzes and not a comprehensive exam. If you seem to finish very early, that is not a bad sign.

- Please write your answers in the boxed space provided on the exam.

- Your answers do not need to take up all the available space, and often the best answers are concise.

- This is a closed-book quiz. No electronic devices are allowed, including cell phones used merely as watches. Silence your cell phones and place them in your bag.

Relational algebra operators:

Union ∪	Difference —	Selection σ	Projection π	Join ⋈
Rename ρ	Duplicate elimination δ	Grouping and aggregation ɣ	Sorting τ	

By writing your name below, you certify that you have not received any unpermitted aid for this exam, and that you will not disclose the contents of the exam to anyone in the class who has not taken it.

NAME: _____

STUDENT NUMBER: _____

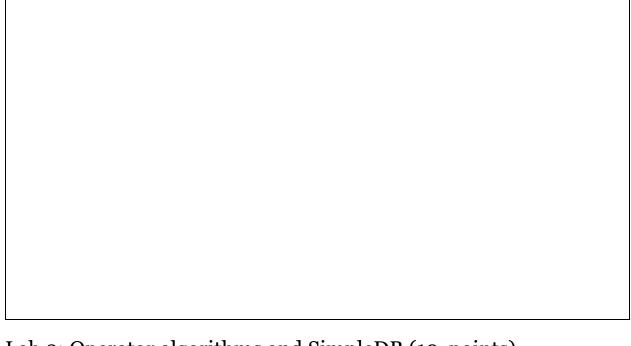# Lab 1: Storage management and SimpleDB (10 points)

a) (5 points)   Different designs are possible to store data on disk in a relational DBMS. In this problem we will consider what happens if we change part of the storage design of SimpleDB. In lab 1, our design is that each relation is stored in a separate file; each file consists of pages; and each page holds a set of tuples.

For this problem we will only change how the tuples are stored within a page. Instead of storing all the fields of a tuple together, we will now first store the values of attribute 1 for all tuples together, then the values of attribute 2, then the values of attribute 3, etc. We do this inside each page.

Explain what components of SimpleDB we need to modify to support this new design and why. You do not need to propose new algorithms or implementation details, just the outline of what you would change.

b) (5 points)

A software engineer decides to optimize the Buffer Pool Manager implementation of a DBMS. Because index pages are frequently accessed, the software engineer changes the implementation such that index pages are never flushed from the Buffer Pool. How will this change affect the performance of the system? Assume there are many tables and many different queries over those tables from multiple users.

Lab 2: Operator algorithms and SimpleDB (10 points)

We will work with the following schema for a database of movies

    CREATE TABLE Movie
        (mid Int,
        title String,
        genre String,
        year Int);

Say we are going to run the following query in SimpleDB, on the relation Movie.

    DELETE
    FROM Movie
    WHERE genre = 'comedy' AND
            year = 2018;   (typo in original: previously read '2018' as a String instead of
Int)

a) (5 points) Draw the physical query plan that will be executed by your implementation of SimpleDB. This should be a relational algebra tree with annotations for the physical operators. Example annotations are "sequential scan", "index lookup", "on the fly"...

b) (5 points) Assume our buffer pool can hold at most 10 pages, and the size of the Movie relation is 100 pages.
What will the contents of the buffer pool be after the first call of next() to the Delete operator at the top of your query plan? This answer will depend on your choice of implementation of the buffer pool in SimpleDB lab 2.