1. This exam refers to the following relational database. If you have any questions about this database, please ask me now. Misinterpreting the schema is no excuse for an incorrect answer.

Student information at Baylor University is stored in a Students table in the database. The real database contains far more information, but for our purposes, we will just consider the attributes student_id, name and class (Freshman, Sophomore, etc.). The student_id is an integer (4 bytes) and the primary key. There is an index on that attribute. The other attributes are varchar(194) and char(2) respectively.

Baylor also offers classes for students to take. The information about classes is stored in the Courses table with attributes course_id, name and professor. The course_id is an integer (4 bytes), primary key and indexed ($B^+$-tree). The other attributes are varchar(248).

The Takes table shows which students are taking which course and the grade they made. The table has student_id, course_id and grade as attributes. The student_id is a foreign key for Students, the course_id is a foreign key for Courses, and grade is a char(2). There are no keys or indexes.

You may assume there are 10,000 students, 50,000 takes rows and 2,000 courses. There are 500 unique professors in the Courses table. Every course has at least one student, and every students is registered for at least one course. For simplicity, assume each page can hold 10,000 bytes of data.
Consider the query:

```
SELECT S.name
FROM Students S, Takes T, Courses C
WHERE S.sid = T.sid AND T.cid = C.cid and C.professor = 'Speegle';
```

Assume 100 pages in memory are available to perform this query. Describe how a DBMS would efficiently perform this query. Compute the I/O cost for efficiently performing this query.
2. Consider the SQL statement used to make initial grade assignments for students in this class:

```
UPDATE Takes
SET grade = 'A'
WHERE course_id IN
    (SELECT course_id FROM Courses WHERE professor = 'Speegle');
```

(a) Assume multigranularity locking is used. What are the lock requests generated by this statement? (12)

(b) Estimate the number of locks required under 2PL. (4)

(c) If Hamerly also assigns an A to all of his students at exactly the same time, will there be a conflict under multiple granularity locking? If so, where? Under 2PL? If so, where? (4)
(d) Assuming ARES is the recovery protocol followed and AUTO COMMIT is on. Furthermore, assume only the students in the front row are in this class. Show the log records generated for this class.

(e) Assume the DBMS fails after writing the first log record to disk. Show the log after full recovery.

3. Assume TIDs are 6 bytes long. What is the branching factor of the index for Students?
4. Describe the fragment and replicate join algorithm for the join of students and takes on a parallel databases with 25 processors. Try to perform the join as efficiently as possible.

5. Define the acronym ACID.
6. Show an example of a serializable, recoverable schedule that is not cascadeless. Prove your answer correct.

7. Define the acronym BASE
8. Reproduce the 2PC protocol. Indicate the commit point. Describe the most significant failure case. (10)

9. Give one advantage and one disadvantage of multiversion concurrency control. (4)