Draw the ER diagram for the database.

Express the following queries using relational algebra (if possible):
(c) Find the graduate students who work only in the Database lab or in the Security lab.
(d) Find the graduate students who work on at least two projects in the Database lab.

Express the following queries using relational calculus (if possible):
(e) For each project, find the graduate student with the greatest number of hours and the graduate students with the smallest number of hours.
(f) Find the graduate students who work (in total) more than the average (the average also considers the total number of hours – in all the projects- that graduate students work).

4. Query Evaluation (Exercise 14.4 in the textbook) (30 points)
Consider the join $R \bowtie_{A_B=S_a} S$, given the following information about the relations to be joined. The cost metric is the number of page I/Os unless otherwise noted, and the cost of writing out the result should be uniformly ignored.

- Relation $R$ contains 10,000 tuples and has 10 tuples per page.
- Relation $S$ contains 2000 tuples and also has 10 tuples per page.
- Attribute $b$ of relation $S$ is the primary key for $S$.
- Both relations are stored as simple heap files.
- Neither relation has any indexes built on it.
- 52 buffer pages are available.

1. What is the cost of joining $R$ and $S$ using a page-oriented simple nested loops join? What is the minimum number of buffer pages required for this cost to remain unchanged?
2. What is the cost of joining $R$ and $S$ using a block nested loops join? What is the minimum number of buffer pages required for this cost to remain unchanged?
3. What is the cost of joining $R$ and $S$ using a sort-merge join? What is the minimum number of buffer pages required for this cost to remain unchanged?
4. What is the cost of joining $R$ and $S$ using a hash join? What is the minimum number of buffer pages required for this cost to remain unchanged?
5. What would be the lowest possible I/O cost for joining $R$ and $S$ using any join algorithm, and how much buffer space would be needed to achieve this cost? Explain briefly (no more than 8 lines).

5. Query Evaluation (15 points)
Consider the block nested loop join algorithm discussed in class. To join tables R and S, with B buffers, we reserve B-2 buffers for the outer relation (say R), one for S, and one for the output. This works out well under the assumption that each page I/O costs the same. However, this is not the case in practice. Consider an I/O cost model that has a latency of 50 msec and a transfer time of 5 msec per page (buffer). Under this model, is it best to use the same division of buffers between R, S, and output? You may assume that R has M pages, S has N pages, and the output is P pages long. Evaluate the cost of the join assuming that B1 buffers are used for R, B2 for S, B3 for the output. Based on this cost, how should you divide the buffers?