I.

Please answer these questions:

1. What is database? What is DBMS?
   1. Collection of related data and its metadata in a structured format.
   2. DBMS = Database management system. Software to allow interaction with the database. Manages how people interact w/the data.
2. What are the problems of file systems?
   1. Physical systems often duplicate data. Inconsistent upkeep if a record changes and needs to be changed in several places. They’re rigid and hard to update. Take up lots of space and require lots of maintenance.
3. What are the pros and cons of hierarchical database?
   1. Easy-to-understand concept (tree model). Centralizes data, reduces redundancy.
   2. Doesn’t represent the full complexity of the data – interactions other than hierarchical ones (no M:N relationships), and finding any data point requires knowing the storage path to get to it.
4. Give an example of a network database.
   1. Customers and sellers on a platform like eBay or Craigslist (you can buy and/or sell, and can have many customers and many sellers interacting in a network)
5. What are the pros and cons of relational database?
   1. Tables are flexible, logical structure to represent data.
   2. They’re structurally independent – doesn’t matter how they are designed, there is no connection to how data is accessed by user. You can do ad hoc queries with SQL
   3. Requires good software – more complex. Can easily design poorly because they are easy to use
6. What are the major types of NoSQL databases? What are the pros and cons of NoSQL databases?
   1. Key-value store
   2. Document store
   3. Graph
   4. Triple store
   5. Pros/cons: Can be used in real-time for data that is continuously growing. Can be scaled elastically, open-source, easy to use. Not much security, no indexing, immature tech, and non standardized.

II.

Given the database Student-Course, we have three tables:

* S(SNO,SNAME,AGE,SEX,SDEPT)
* SC(SNO,CNO,GRADE)
* C(CNO,CNAME,CDEPT,TNAME)

Note: The table “S” shows the id number (SNO), name (SNAME), age (AGE), gender (SEX), and the department (SDEPT) of the students. The table “SC” shows the id number of the students (SNO) and the courses they enrolled (CNO) as well as the grade they get in the courses (GRADE). The table “C” shows the id number (CNO), name (CNAME), department (CDEPT), and the name of the teacher (TNAME) of the courses.

Using this database, please write eight relational algebras below (the 1st and the 7th answers are provided as an example). You can type the relational algebras in Microsoft Word or you can write them on a paper and photo them. Both are fine.

1. The id number and the name of the courses that John Smith teaches.

 ∏CNO,CNAME(σTNAME=‘John Smith’(C))

**Explanation: σ=select, ∏=projection**

**First you should choose those record (row) with TNAME=’John Smith’ from table C. In this way, all of the columns will be selected. Since you only need two of them, you can them use “projection” to select some of the columns from all columns. Here you just need to select CNO and CNAME.**

1. The id number and the name of the male students who are 23 years old or more.

∏SNO,SNAME(σAGE>=23 ^ SEX=’Male’(S))

1. The names of the courses and their teachers that student S3 enrolled. Note that S3 is the id number of the student.

∏CNAME,TNAME(σSNO=’S3’(C|x|SC))

1. The names of the female students who at least enrolled one course taught by John Smith.

∏SNAME(σSEX=’Female’ ^ TNAME=’John Smith’(S|x|C|x|SC))

1. The id number of the courses that Linda Wild did not enroll.

∏­CNO(C) - ∏CNO(σSNAME=’Linda Wild’(SC|x|S))

1. The id number of the students who have at least enrolled two courses.

∏SNO(σ1=4 ^ 2!=5 (SC|x|SC))

1. The id number and the name of the courses enrolled by all of the students in our database.

**∏CNO,CNAME,SNO( C |×| SC ) %****[[1]](https://iu.instructure.com/courses/1774377/assignments/8760215?module_item_id=17869805" \l "_ftn1) ∏SNO(S)**

**Explanation: For these “all XXX” question, we can use % (divide) to solve it. In order to know which course(s) is enrolled by all students, you should first have a new table containing who (including SNO) enrolled which course (including CNO, CNAME), so first we join C and SC, and then select useful columns, CNO, CNAME, and SNO.**

**According to the professor’s slide, if we want to have a row subset, we should use “%”. To use “%”, we need a divider, which is a new table containing all of the student number (just a one-column table), so we do like ∏SNO(S). Finally, we do the “%”.**

1. The id number of the students who have enrolled the course taught by John Smith.

**∏SNO (**σTNAME=’John **Smith’) (C|x|SC))**

[[1]](https://iu.instructure.com/courses/1774377/assignments/8760215?module_item_id=17869805" \l "_ftnref1) You can also use ÷, they are the same.