

Assignment 2

Due Wed, Feb 28 at 11:59pm

General Instructions

- Feel free to talk to other members of the class in doing the homework. You should, however, write down your solutions yourself. *List the names of everyone you worked with at the top of your submission.*
- Keep your solutions brief and clear.
- Please use Piazza if you have questions about the homework but do not post answers. Feel free to use private posts or come to the office hours.

Homework Submission

- We DO NOT accept late homework submissions.
- We will be using Compass for collecting the homework assignments. Please submit your answers via [Compass](#). Hard copies are not accepted.
- Contact the TAs if you are having technical difficulties in submitting the assignment; attempt to submit well in advance of the due date/time.
- The homework must be submitted in **pdf** format. Scanned handwritten and/or hand-drawn pictures in your documents won't be accepted.
- Please do not zip the answer document (PDF) so that the graders can read it directly on Compass. You need to submit one answer document, named as **hw2_netid.pdf**.
- Please see the [assignments](#) page for more details. In particular, we will be announcing clarifications, if any, on [this page](#).

1 Short Questions (20 pts)

Provide a short answer (3-4 sentences at most) for each of the following questions. You may use figures if necessary.

1. [2.5] Suppose we have a relation R as given below, representing the exam statistics for course CS411. First project relation R to GPA (i.e., eliminate SID and Name) and then calculate the average GPA, under the set-model and the bag-model respectively. Which model is preferred in this example and why?

SID	Name	GPA
1	James	3
2	Charles	4
3	Doris	4
4	Ada	4

2. [5] Consider a relation $R(A, B, C)$. You may assume there are no null values or duplicates in R . If the result of $\sigma_{Y \neq V}(\rho_{R(X,Y,Z)}R \bowtie \rho_{R(X,V,W)}R)$ is always guaranteed to be empty, then what property of R can you infer? (Hint: think functional dependencies.)
3. [2.5] Consider any relation R that never contains more than one tuple. Is it true that R must be in Boyce-Codd Normal Form (BCNF)? Justify your answer.
4. [4] Consider a relation $R(A, B, C, D, E)$ with dependencies $AB \rightarrow CD$, $C \rightarrow AB$, $D \rightarrow AE$, list all minimal keys for R . Also, state whether the relation R is in 3NF **with reasoning**.
5. [6] Two sets of functional dependencies (FD's) F and F' are equivalent if all FD's in F' follow from the ones in F , and all FD's in F follow from the ones in F' . Consider the following three sets of functional dependencies:
 - $F1 = A \rightarrow C, B \rightarrow A$,
 - $F2 = B \rightarrow AC$
 - $F3 = AB \rightarrow C, B \rightarrow A$
 - (a) Are $F1$ and $F2$ equivalent? Justify your answer.
 - (b) Are $F1$ and $F3$ equivalent? Justify your answer.
 - (c) Are $F2$ and $F3$ equivalent? Justify your answer.

2 Relational algebra to English (15 pts)

Consider a relation *Works* (name, company, salary) with no duplicates. Consider the following relational algebra expression, written in linear notation.

$$P1(salary) = \pi_{salary}(\sigma_{company="IBM"}(Works))$$

$$P2(salary) = \pi_{salary}(\rho_{T1(s)}(P1) \bowtie_{s > salary} P1)$$

$$P3(salary) = P1 - P2$$

$$Answer(name) = \pi_{name}(Works \bowtie_{salary > s} \rho_{T2(s)}(P3))$$

State in English what is computed as the final answer briefly. Long-winded answers will be deducted points. For partial credit, explain what $P1$, $P2$ and $P3$ contain.

3 English to Relational Algebra (20 pts)

Consider the following relational database schema that describes information about students and their courses. A course is uniquely identified by its CODE (e.g., “CS411”), and a student is uniquely identified by his or her SID.

Course(CODE, units, time, room) // all courses

Student(SID, name, level) // all students, level can be “grad” or “undergrad”

Taking(SID, CODE) // current enrollment information

Write a relational algebra expression to list the information (i.e., CODE, units, time, room) of courses that are currently offered but have no graduate students enrolled.

4 Data to functional dependency (20 pts)

Consider a relation $R(A, B, C)$, satisfying some functional dependency. Two instances of R are given as below:

A	B	C
2	3	1
2	2	4

A	B	C
2	2	1
3	3	2
4	2	1

Based on R 's schema, enumerate all possible completely nontrivial functional dependencies (FDs) with only a single attribute on the right-hand side. Then, based on the instances above, for each FD you listed, label whether it:

H: Definitely holds in R .

NH: Definitely does not hold in R .

CD: Cannot be determined from the information given whether or not it holds in R .

5 Normalization (25 Points)

Consider the following relational schema for a chain store:

Sale(clerk, store, city, date, dish, size)

// a clerk sold a dish on a particular day at a given store in a city

Menu(dish, size, price)

// prices and available size for the dish

Make the following assumptions:

- Each clerk works in one store.
- Each store is in one city.
- The price of a dish is different for different sizes. The store has standardized prices: the same sized dish cannot be sold to two persons at two different prices.

1. Specify a set of completely nontrivial functional dependencies for relations Sale and Menu that encodes the assumptions described above and no additional assumptions.
2. Based on your functional dependencies in part (1), specify all minimal keys for relations Sale and Menu.
3. Are the schemas of Sale and Menu in Boyce-Codd Normal Form (BCNF) according to your answers to (1) and (2)? If not, give a decomposition into BCNF. If yes, justify your answer.
4. Now add the following assumption:

- Each city has at most one store and each store has only one clerk.

Specify additional functional dependencies to take these new assumptions into account.

5. Based on your functional dependencies for parts (1) and (4) together, specify all minimal keys for relation Sale.
6. Are the schemas of Sale and Menu in 3NF according to your answers to (1), (4) and (5)? If not, give a decomposition into 3NF. If yes, justify your answer.