NAME: **NEPTUN CODE:**

INTRODUCTION TO DATABASE SYSTEMS
MIDTERM EXAM FALL 2014

---

1. Prove, that Armstrong axioms are sound. (5 points)

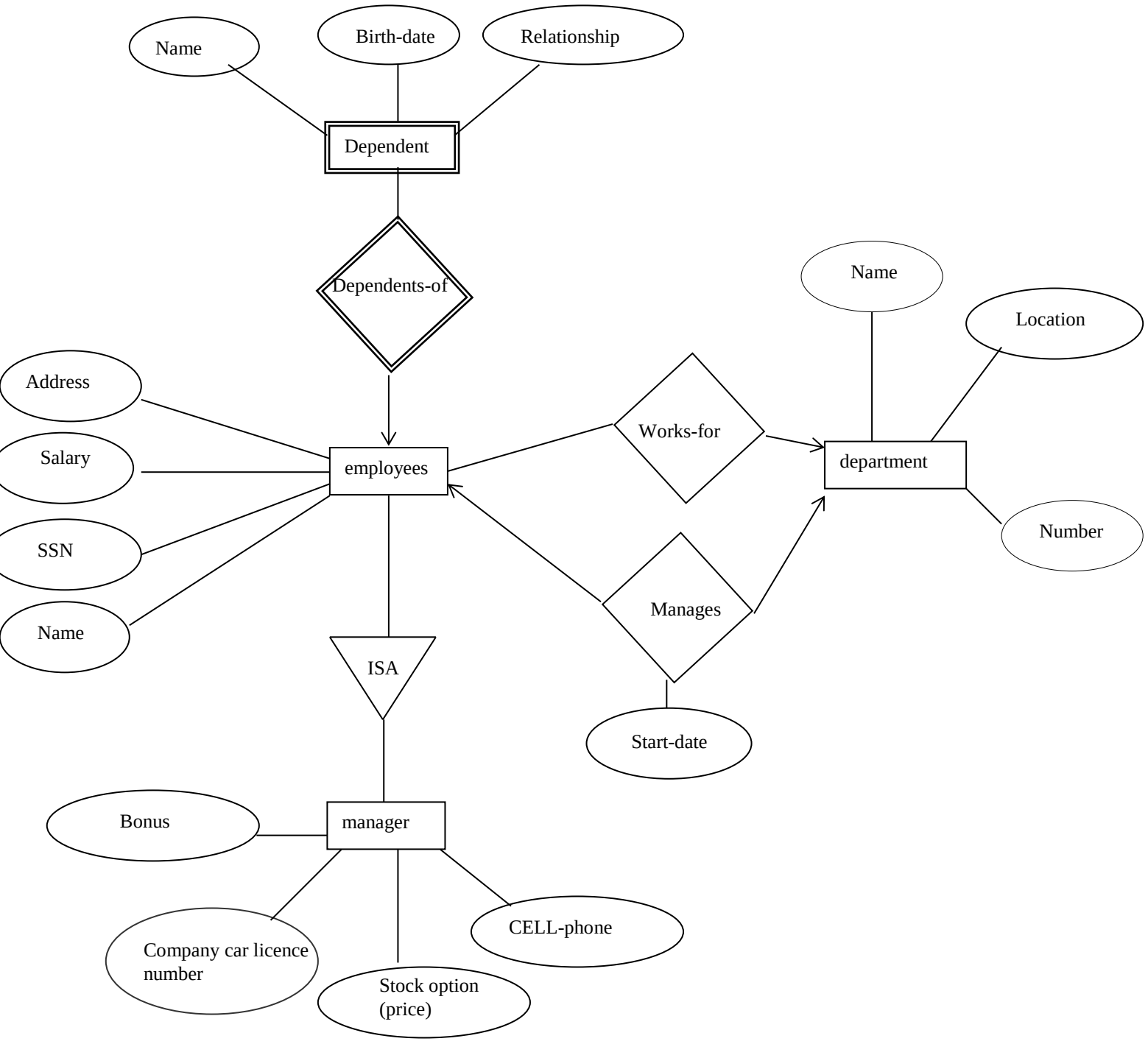
2. Give the definition of functional dependency. Give a proof that if $\alpha \rightarrow \beta$ and $\delta \rightarrow \gamma$ hold, then $\alpha\delta \rightarrow \beta\gamma$ holds. (1+2 points)

3.1 Use functional dependencies to define superkey:
(1 points)

3.2 Explain why any set of attributes containing a superkey is also a superkey.
(2 points)

3.3 What two properties must a candidate key possess?
(2 points)

4. Underline PKs in the E/R modell. (1 points)
Translate the E/R model below into relational model (4 points)
What reasonable changes would you make in this figure, according to the real world situation? (2 points).



5. Consider the attributes R with A B C D E F G which have the following functional dependencies:
Fd={AD → F, AE → G , DF → BC, E → C, G → E}

5. 1. List all candidate keys, explain:
(2 point)

5. 2. Give 3 superkeys:
(1 point)

5. 3. Give a primary key:
(1 point)

5.4 Which of the following dependencies are implied by those in ***Fd*** above? If your answer is YES, please explain. If you can explain in 2 different ways, then you might earn extra points. (1 point each)

a. ADC → B : YES/NO, because....

b. A → G: YES/NO, because....

5. 5 Give 2 elements of Fd^+ being different from the trivial dependencies. (2 point)

6. Consider the attributes R with A B C D E F G which have the following functional dependencies:
Fd={AD → F, AE → G , DF → BC, E → C, G → E}
Consider the decomposition into 3 relations: (ADF) (EC) (ABDEG).

This decomposition (IS) (IS NOT) lossless, because…… (4 points)

If the relation is not in BCNF then give a lossless BCNF decomposition. Use the algorithm we learnt.
(6 points)

7.
Given the following database tables, choose all queries in SQL or Relational Algebra (RA) as indicated. Primary keys are underlined. Note that the driver involved in a car accident may not always be the owner of the car. Assume that accident_date is of type integer, and represents a year (e.g. 1980). Year is also of type integer. We assume that a car cannot get involved in more than one accident at a certain date.

Person(SSN, name, address)
Car(license, year, model)
Accident(license, accident_date, driver, damage_amount)
Owns(SSN, license)

7.1 Find the SSN of every person, who owns a TOYOTA or a DODGE

WRITE HERE the appropriate SQL statement (1 point):

WRITE HERE the appropriate RA expression (1 point):

7.2
Find the SSN of all persons who have had all of their cars involved in an accident.

WRITE HERE the appropriate SQL statement (2 point):

WRITE HERE the appropriate RA expression (2 point):

Person(SSN, name, address)
Car(license, year, model)
Accident(license, accident_date, driver, damage_amount)
Owns(SSN, license)

7.3. Who is the driver who participated in the most costly accident? Return the driver and the amount of damage. (3 points)

7.4 Find the license number of all cars that have been involved in more than one accident. Order the output regarding the alphabetical order of the licence plates.
(DO NOT RETURN DUPLICATES) USE SQL ONLY (2 point):

7. 6 Find owners having the largest number of cars. (2 point)
RA ONLY:

8. Can you express division by basic relational algebraic operators? If yes, please write here, and explain step by step. (4 points)

9

9.1 What does it mean that Armstorng axioms are sound and complete? (2 points)

9.2What is the relationship between 1NF, 2NF, 3NF, BCNF? Draw a figure, and explain briefly. (3 points)

PROBLEM 4

10. What is normalization, why it is needed? (5 points)

11.

Step	T_1	T_2	T_3	T_4
1	READ(W)			
2		READ(X)		
3		WRITE(X)		
4				READ(V)
			WRITE(V)	
5		WRITE(Y)		
6			READ(Y)	
7				WRITE(X)
8			READ(V)	
9	WRITE(W)			
10	READ(Y)			
11			WRITE(X)	
12			WRITE(Y)	
13		READ(Z)		

Draw precedence graph for the partial schedule above. (3 points) Is this a conflict serializable schedule? (1 point) If **yes**, show an equivalent serial schedule for T_1, T_2, T_3, T_4. Explain. If **no**, argue why not. (1 point)

12. Why 2 Phase protocol is important? (4 points)

13.

Step	T_1	T_2	T_3	T_4
1	READ(W)			
2		READ(X)		
3		WRITE(X)		
4				READ(V)
5	READ(Y)			
6			WRITE(V)	
7		WRITE(Y)		
8			READ(Y)	
9				WRITE(X)
10			READ(V)	
11	WRITE(W)			
12			WRITE(X)	
13			WRITE(Y)	
14		READ(Z)		

Suppose that in the schedule above we apply timestamp protocol.

T1 has timestamp 10,
T2 has timestamp 20,
T3 has timestamp 30
T4 has timestamp 40

Which of the transactions has to be aborted? Why? Could you apply Thomas' writing rule? Why or why not? 2.2 Suppose that the transaction(s) you decided to abort in 2.1 were not present. After the succeful execution of tyransactions keeping that schedule, what are the timestamps for V and X?

(6 points)

14. Consider a database management system running transactions concurrently. The table below shows a part of the log-file. What type of logging technique is used? Describe recovery process. Give the values for each data item. (5 point)

Step	*Log Entry*
1	<T_1 **start**>
2	<T_1, *A*, 100, 200>
3	<T_2 **start**>
4	<T_2, *C*, 700, 800>
5	<T_3 **start**>
6	<T_2, *D*, 900, 300>
7	<T_3, *E*, 9, 30>
8	<T_2 **commit**>
9	< **checkpoint** *S* >
10	<T_3, *F*, 15, 20>
11	<T_3 **commit**>
12	<T_1, *H*, 75, 50>
13	<T_4 **start**>
14	<T_4, *G*, 57, 25>
15	<T_4 **commit**>
16	<T_1, *B*, 104, 204>
17	<T_1 **commit**>
18	<T_5 **start**>
19	<T_5, *I*, 33, 44>
20	<T_5, *C*, 800, 850>
21	<T_6 **start**>
22	<T_6, *D*, 300, 350>

15. Describe UNDO or REDO recovery technique. (5 points)