Fundamentals of Data Engineering

(COE848)

Mid Term Exam

Feb 2018

This is a CLOSED BOOK exam. Textbooks, notes, laptops, calculators, personal digital assistants, cell phones, and Internet access are NOT allowed.

This is a 120-minute exam.

There are 5 questions with a total of 100 marks.

Please read each question carefully, and write your answers legibly in the space provided. You may do the questions in any order you wish, but please

USE YOUR TIME WISELY.

When you are finished, please hand in your exam paper and make sure you are **signed out**.

Good luck!

Student Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Score: \_\_\_\_\_ %

|  |  |  |
| --- | --- | --- |
| Question | Maximum Mark | Received |
| 1 | 15 |  |
| 2 | 20 |  |
| 3 | 20 |  |
| 4 | 20 |  |
| 5 | 25 |  |
| Total | 100 |  |

Question 1: Explain the following terms in 2-3 sentences:

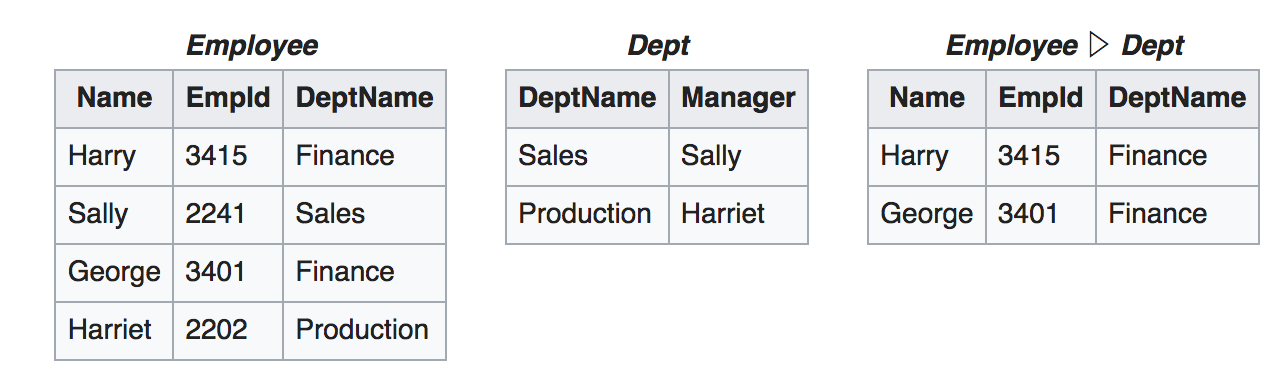
1. Declarative Language
2. Union Compatibility
3. Right Outer Join
4. Total Participation

Question 2: Let us define a new relational algebra operation, call it the anti-join operation (*R* ▷ *S)* and define it formally as follows:

“the result of the ‘anti-join’ operation is only those tuples in *R* for which there is *no* tuple in *S* that is equal on their common attribute names.”

Express *R* ▷ *S* using only primary relational algebra operations. Explain your assumptions, if any.

Hint: Use the following example to visualize how anti-join works but express anti-join for the general case and NOT for this specific example.



Question 3: Given the following table definitions:

Suppliers (sid : integer, sname : string, address : string)

Parts (pid : integer, pname : string, color : string, cost : real)

Catalog (sid : integer, pid : integer, price : real)

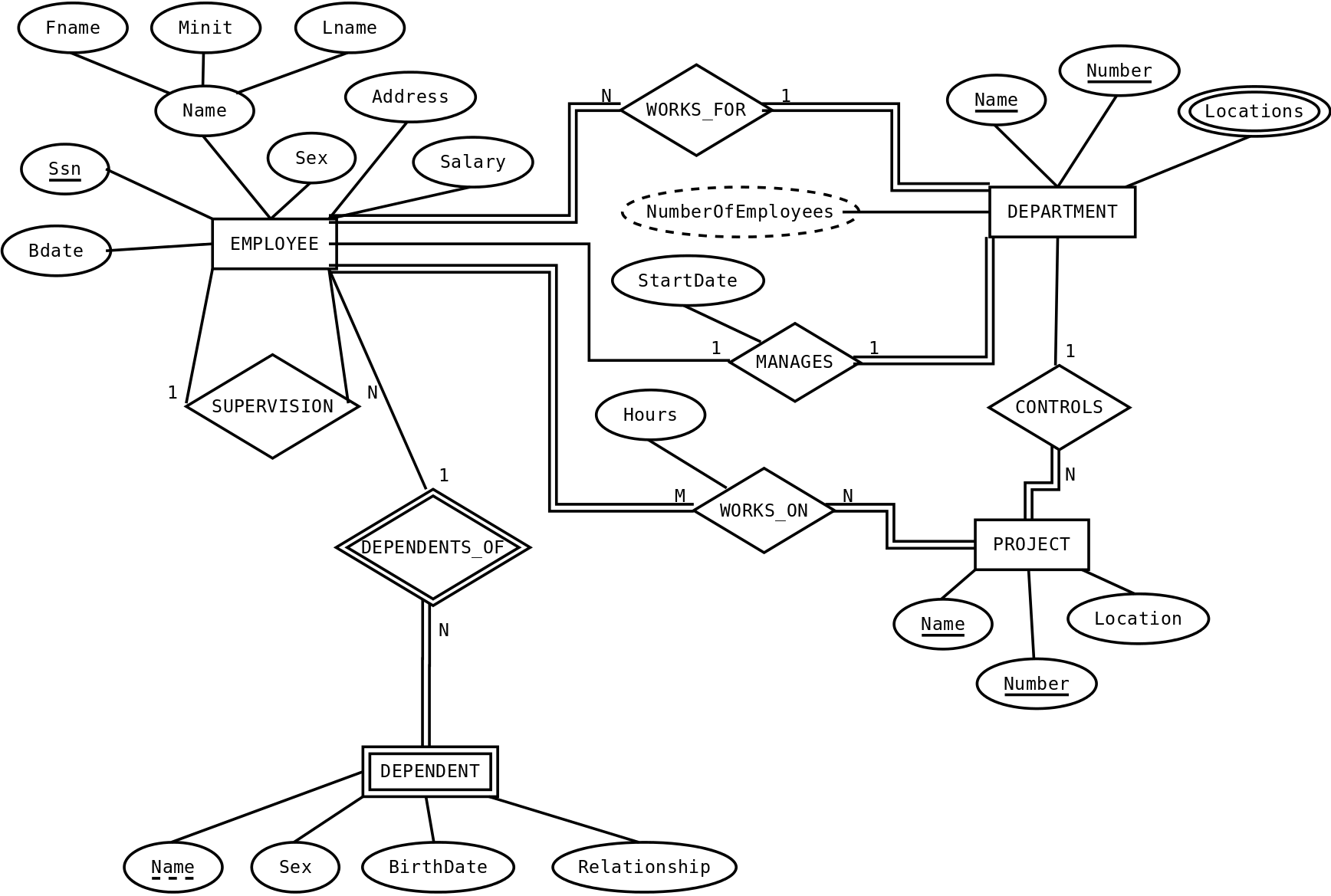
Provide appropriate relational algebra expressions for each of the following queries:

1. Find the names of those parts that are green and cost more than $100.
2. Find the name of those parts that are sold cheaper than they cost to be built.
3. Find the names of all suppliers who sell at least one part for twice as much as it cost them to build it.
4. Find the name of all those suppliers who sell exactly one part.
5. Find the name of those parts that have the highest cost for being manufactured.

Question 4: Convert the ER diagram below to a database schema. Indicate the keys for each table in your answer.

Note: For a table T with attributes k and p where k is the primary key, you can use the following notation in your answer:

T (k,p)



Question 5: Suppose you are given the following requirements for a simple database for the National Hockey League (NHL):

* the NHL has many teams,
* each team has a name, a city, a coach, a captain, and a set of players,
* each player belongs to only one team,
* each player has a name, a position (such as left wing or goalie), a skill level, and a set
* of injury records,
* a team captain is also a player,
* a game is played between two teams (referred to as host\_team and guest\_team) and has a date and a score (such as 4 to 2).

Construct a clean and concise ER diagram for the NHL database schema. List your assumptions and clearly indicate the cardinality mappings in your ER diagram.