Fundamentals of Data Engineering
(COE848)

Mid Term Exam

Feb 2020

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	16	
2	15	
3	27	
4	20	
5	22	
Total	100	

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

a) Logical Data Model

Logical data model describes data stored in database, and the relationships among the data

b) Atomicity of updates

Ensuring that either all series of operations occur, or nothing occurs

c) Total relationship

All entities from an entity set participate in the relationship with at least one entity from another entity set

d) Set Difference Operator

A\B: All elements in A except those exit in B

Question 2: Let us consider an extended relational algebra operation, division, (*R ÷ S)*, as follows:

"The result of the **division** operation is only those tuples in *R* that are connected to ***all*** tuples in *S*"

For this operation, we assume that attributes of S (denoted as ATTR(S)) is a subset of attributes of R (denoted as (ATTR(R)). We also use (ATTR(R)-ATTR(S) to denote a set of attributes in R that does not belong to S.

Express (*R ÷ S)* using only primary relational algebra operations (Project, Select, Rename, Union, Intersection, Set Difference, Cartesian product, Join (conditional or natural), (left or right or full) Outer join). Explain your assumptions, if any.

Hint: Use the example in the next page to visualize how division works but express division for the general case and NOT for this specific example.

Response to Question 2:

A = $((\pi_{ATTR(R)-ATTR(S)}R) \times S) \setminus R$

B = $\pi_{ATTR(R)-ATTR(S)}R \setminus \pi_{ATTR(R)-ATTR(S)}$A

$R \div S$ = B $\times S$

Movie_Genre					
Id	Title	Language	RunningTime	GenreId	Title
1	2001: A Space Odyssey	English	142	1	Sci-fi
1	2001: A Space Odyssey	English	142	3	Adventure
2	Rosemary's Baby	English	136	6	Drama
2	Rosemary's Baby	English	136	7	Horror
2	Rosemary's Baby	English	136	2	Action
2	Rosemary's Baby	English	136	1	Sci-fi
2	Rosemary's Baby	English	136	3	Adventure
2	Rosemary's Baby	English	136	4	Comedy
2	Rosemary's Baby	English	136	5	Crime

Genre	
GenreId	Title
1	Sci-fi
2	Action
3	Adventure
4	Comedy
5	Crime
6	Drama
7	Horror

Movie_Genre ÷ Genre					
Id	Title	Language	RunningTime	GenreId	Title
2	Rosemary's Baby	English	136	6	Drama
2	Rosemary's Baby	English	136	7	Horror
2	Rosemary's Baby	English	136	2	Action
2	Rosemary's Baby	English	136	1	Sci-fi
2	Rosemary's Baby	English	136	3	Adventure
2	Rosemary's Baby	English	136	4	Comedy
2	Rosemary's Baby	English	136	5	Crime

Question 3: Given the following table definitions:

Enrollment (stdId, crsNumber, startDate, finalGrade)
Student (stdId, stdName, address, age)
Course (crsNumber, crsName, description)

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

a. Find the name of all students older than 30 years old.

$$\sigma_{age>30} Student$$

b. Find courses that are enrolled by students.

$$\pi_{\text{crsNumber}} \text{Enrollment}$$

c. Find the name of students who enrolled in some courses.

$$\pi_{\text{stdName}} Enrollment \bowtie Student$$

d. Find the name of students who enrolled in at lease two different courses

$A=\rho_{<EN2'(cn2/srsNumber,\ sId2/stdId)>}(\text{Enrollment}) \bowtie_{cn2 <> crsNumber \text{ and } sId2= stdId} \text{Enrollment}$

$\text{Result}=\pi_{\text{stdName}} A \bowtie Student$

e. Find the course description of all the courses that are enrolled by **some** students older than 30 years old.

$$\pi_{\text{description}}\text{Course} \bowtie \sigma_{age>30} Student \bowtie \text{Enrollment}$$

f. Find the course description of all the courses that are enrolled only by students who are older than 30 years old.

$$A=\pi_{\text{crsNumber}}\ \sigma_{age<=30} Student \bowtie \text{Enrollment}$$

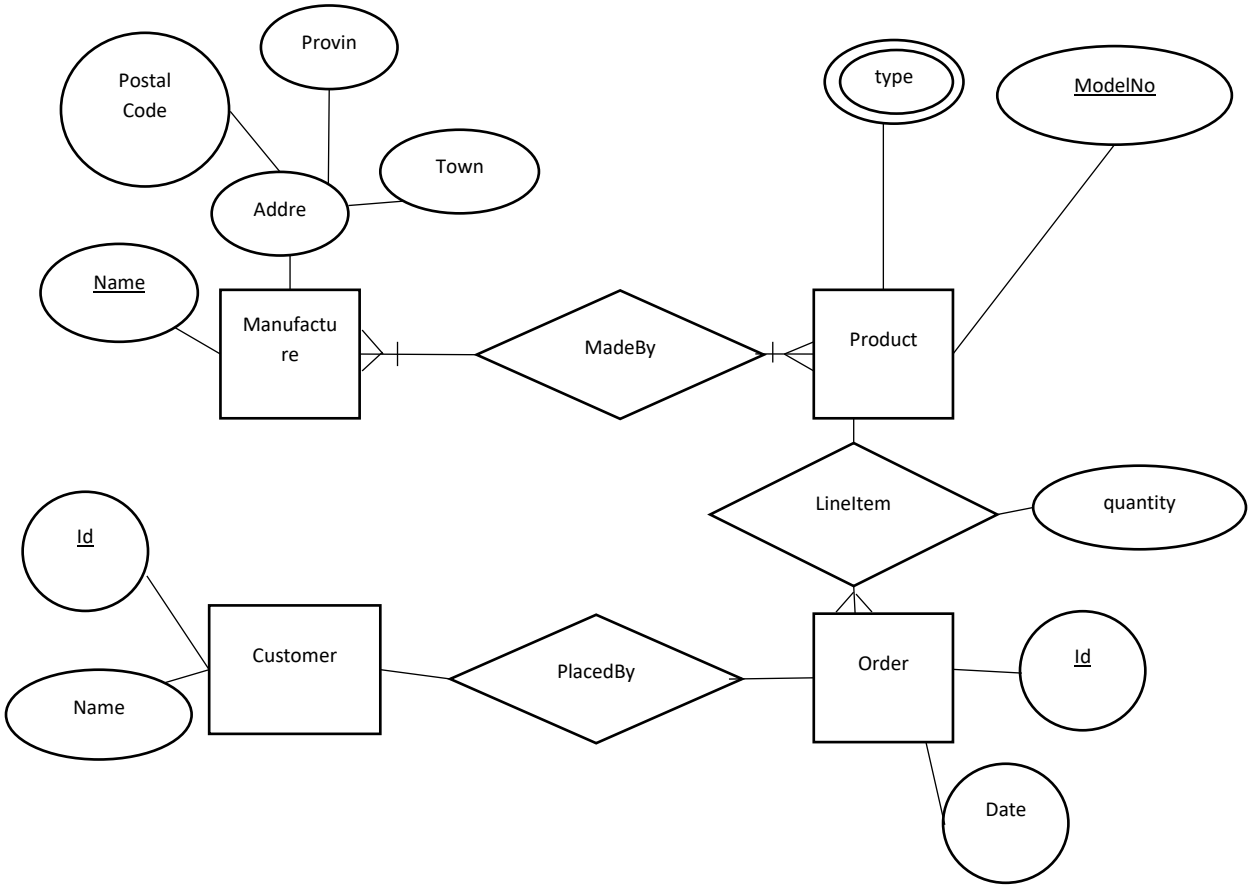
$$B=\pi_{\text{crsNumber}}\ \sigma_{age>30} Student \bowtie \text{Enrollment}$$

$$\text{Result}= \pi_{\text{description}}\text{Course} \bowtie (B \setminus A)$$

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

In this diagram, a Product can be in a number of orders but an order has only one product

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(<u>k</u>,p)



Response to Question 4

Manufacture(Name, postalCode, province, town)
Product(ModelNo)
Type(typeid, description)
typeProduct(modelNo, typeid)
MadeBy(Manufacture.Name, Product.ModelNo)

Order(Id, Date, Product.ModelNo, quantity)
Customer(Id, name)
PlacedBy(Customer.Id, Product.Id)

Question 5: A simple database for Canada post:

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

The database have the following information:

- Items number (unique), weight, dimensions, insurance amount, destination, and final delivery date
- Items are received into the Canada Post system at a single retail center
- Retail centers are characterized by their type, uniqueID, and address
- Items are shipped via one or more standard Canada Post transportation events (i.e., flights, truck deliveries).
- These transportation events are characterized by a unique scheduleNumber, a type (e.g, flight, truck), and a deliveryRoute.