School of Engineering and Computer Science

SWEN 304/439 **Database System Engineering**

Assignment 1

The objective of this assignment is to test your understanding of database foundations, basic terms, and the relational data model the entity relational model. It is worth 15% of your final grade. The assignment is marked out of 100.

The assignment is due on **Friday, 1 April, 23:59 pm**. Please submit your assignment in **pdf** via the submission system.

Question 1 [20 marks]

The TOP500 project lists the 500 most powerful non-distributed computer systems in the world (also called supercomputers). Suppose we use a relational database to manage the current and future data of this project. For this purpose, we use a relation schema with the attribute set {Performance, Name, Manufacturer, Country, Year}.

The following table shows a portion of the current instance of the SUPERCOMPUTERS relation schema that stores data for some supercomputers. Note that the performance in the table is measured in petaFLOPS.

SUPERCOMPUTERS

Performance	Name	Manufacturer	Country	Year
442010	Fugaku	Fujitsu	Japan	2020
148600	Summit	IBM	United States	2018
94640	Sierra	IBM	United States	2018
93015	Sunway	NRCPC	China	2016
64590	Perlmutter	HPE	United States	2021
63460	Selene	Nvidia	United States	2020
61445	Tianhe-2A	NUDT	China	2013
44120	JUWELS	Atos	Germany	2020
35450	HPC5	Dell EMC	Italy	2020
30050	Voyager-EUS2	Microsoft	United States	2021

a) [8 marks] For every set of attributes (that is, for every subset of the attribute set) decide whether you can deduce that it is *not* a candidate key, assuming the shown instance is legal. Justify your answer.

Answer:

{Performance}: Yes. It is unique, minimal and not null.

{Name}: Yes. It is unique, minimal and not null.

{Manufacture}: No. Two **IBM** in it which is not unique.

{Country}: No. Five **United States** and two **China** in it which is not unique.

{Year}: No. Four **2020** and two **2018** and two **2021** in it which is not unique.

{Manufacture, Country}: No. Two **IBM** and **United States** in it which is not unique.

{Manufacture, Year}: No. Two **IBM** and **2018** in it which is not unique.

{Manufacture, Country, Year}: No. Two **IBM**, **United States** and **2018** in it which is not unique.

b) [4 marks] For every remaining set of attributes (that is, for every set not ruled out as a candidate key in part **a)**), discuss whether you consider it a suitable candidate key? Justify your answer.

Answer:

{Performance}: Is not a suitable candidate key. Because it might have two supercomputers with the same performance.

{Name}: Is a suitable candidate key. Because it would not have two supercomputers with the same name. Even they manufactured by the same manufacturer in the same year (eg.

148600	Summit	IBM	United States	2018
94640	Sierra	IBM	United States	2018

). Also if the manufacturer manufacture a new version of the supercomputer they will put a version number to it (eg.

61445	Tianhe-2A	NUDT	China	2013

35450	HPC5	Dell EMC	Italy	2020
30050	Voyager-EUS2	Microsoft	United States	2021

).

[2 marks] Which of the candidate keys identified in part **b)** would you choose as the primary key? Justify your answer

Answer:{Name}. Because it is the only candidate key that would not have duplicate values.

c) [2 mark] Add a new tuple for a computer into the SUPERCOMPUTERS relation. How would you check that the primary key identified in part **c)** is still valid?

Answer: If the **Name** attribute of the new tuple is not null and does not exist in the **SUPERCOMPUTERS** relation.

d) [2 mark] Create a relation that shows for each country in the table above the country and the capital, i.e., use a relation schema with attribute set {Country, Capital}. How many records are in your relation?

Answer: There are 5 records in my relation. There are 5 countries in the table above which are Japan, United States, China, Germany, Italy.

e) [2 mark]. Consider a relation schema with attribute set {Manufacturer, City} and assume that both attributes have a domain with ten values each. What would be the maximum number of records in an instance of this relation schema?

Answer: 10^2.

Question 2 [10 marks]

Suppose your software company has developed a relational database for the grocery store "Fruits and more". The underlying database schema contains the following relation schemas:

- COMPANY (Cid: STRING, Name: STRING, Location: STRING) with primary key {Cid}
- FRUITS (Fid: STRING, Name: STRING, Cid: STRING, InStock: INTEGER, Price: INTEGER) with primary key {Fid, Cid} and foreign key Cid ⊆ COMPANY[Cid]

Below you find instances of these two relation schemas:

FRUITS

Fid	Name	Cid	InStock	Price
557	Apple	23XY	50	21
85520	Pear	A15F	0	78
63311	Pear	FVT35	211	49
36773	Kiwi	23XY	50	21
36773	Kiwi	FVT35	29	22

COMPANY

Cid	Name	Location
23XY	GreatFruits	Wellington
FVT35	Yummy	Wellington
F15A	GreatFruits	Levin
A15F	BetterFruits	Lower Hutt
5AB32	NiceFruits	*null*

Your tasks are as follows.

a) [5 marks] Decide which of the following tuples can be added or removed, respectively. *Justify your answers!*

1. Insert tuple (*null*, 'Tasty', 'Wellington') into COMPANY

   Answer: Reject. Inserting a null primary key could violate not null constraints.

2. Insert tuple ('FVT35', 'SweetFruits', 'Porirua') into COMPANY

   Answer: Reject. The key value "KVT35" exists in the COMPANY could violate uniqueness constraints.

3. Delete tuple ('A15F', 'BetterFruits', *null*) from COMPANY

   Answer: Reject. This tuple doesn't match any instances in COMPANY.

4. Delete tuple ('23XY', 'GreatFruits', 'Wellington') from COMPANY

   Answer: Reject. This tuple is referred to by foreign key "23XY" in FRUITS then deleting it will make it inconsistent.

5. Insert tuple ('XYZ4', 'Wellington', 'Yummy') into COMPANY

Answer: Reject. The foreign key “XYZ4” does not match any tuple in FRUITS. Referential Integrity constraints.

b) [5 marks] Decide which of the following tuples can be added or removed, respectively. *Justify your answers!*

1. Delete tuple (‘36773’, ‘Kiwi’, ‘23XY’, 50, 21) from FRUITS

Answer: Success. There is another tuple which contains a foreign key “23XY” in FRUITS. Therefore it will not make it inconsistent. This deletion doesn’t violate any constraints.

2. Insert tuple (‘55555’, *null*, ‘F15A’, 2, 99) into FRUITS

Answer: Success. The primary keys “55555” and “F15A” are unique in FRUITS and the foreign key “F15A” exists in COMPANY. Therefore this insertion doesn’t violate any constraints.

3. Insert tuple (‘54556’, ‘Lemon’, ‘FV35’, 20, 43) into FRUITS

Answer: Reject. The foreign key “FV35” does not exist in COMPANY. Referential Integrity constraints.

4. Insert tuple (‘53557’, ‘Apple’, ‘5AB32’, 500, 1) into FRUITS

Answer: Success. The primary keys “53557” and “5AB32” are unique in FRUITS and the foreign key “5AB32” exists in COMPANY. Therefore this insertion doesn’t violate any constraints.

5. Delete tuple (‘46557’, ‘Apple’, ‘23XY’, 1, 21) from FRUITS

Answer: Reject. The Fid “46557” does not exist in FRUITS. This tuple doesn’t match any instances in FRUITS.

Question 3 [20 marks]

The Wellington Foreign Trade Office needs to translate hundreds of documents every day. To ensure professional translation in a timely manner the office cooperates with several translation agencies and expert translators in New Zealand. The processing of the data about translations as well as the checking of deadlines and quality requirements is time consuming and error prone if this is done manually on paper.

Therefore, the office wants to build a new database to record all relevant data that is needed for processing and checking translations. Suppose the following relation schemas have been proposed to be belong to the database schema for the new database.

- TranslationAgency (AgencyNumber, Name) with primary key {AgencyNumber}
- Translator (Name, Phone, Field, IRDNumber) with unknown primary key
- IsExpert (Name, Language) with primary key {Name, Language}
- TranslationOrder (OrderNumber, OrderDate, PageNumber, Budget, FromLanguage, ToLanguage, Deadline) with primary key {OrderNumber}
- Assignment (Agent, OrderNumber, Part, Language, Name) with primary key {OrderNumber, Part}

The following additional constraints are known:

1. Each translator has a unique IRD number, a unique phone and a unique name.
2. For each translator, the IRDNumber must be specified, while Field may be left blank (if not known).
3. Translators can be experts in up to four languages.
4. An agent can assign a translation order to multiple translators who can be distinguished by the assigned part of the order.

Your tasks are as follows:

a) **[3 marks]** For the relation schema Translator, identify all suitable candidate keys. Explain your answer. Which candidate key would you choose as the primary key? Justify your answer.

Answer:

{Name}: Is not null, unique and minimal. Therefore it is a suitable candidate key.

{Phone}: Is not null, unique and minimal. Therefore it is a suitable candidate key.

{Field}: Is not a suitable candidate key. Because Field may be left blank if not known.

{IRDNumber}: Is not null, unique and minimal. Therefore it is a suitable candidate key.

I would choose {Name} as the primary key. Because Name is the most commonly used in these relation schemas.

b) **[5 marks]** For each of the relation schemas, identify all suitable foreign keys (if there are any). Explain your answer.

Answer:

TranslationAgency doesn't have any foreign keys. Because it doesn't connect to any other relation schemas by any attributes in it.

Translators doesn't have any foreign keys. {Name} is connected to the IsExpert relation schema but {Name} is not the primary key of IsExpert. Therefore it can not be the foreign key.

IsExpert with the foreign key {Name}. Name is connected to the Translator relation schema and {Name} is the primary key of Translator. Therefore it could be the foreign key.

TranslationOrder doesn't have any foreign keys.

Assignment (Agent, OrderNumber, Part, Language, Name) with the foreign key{OrderNumber, Agent, Language, Name}. Because the OrderNumber connects Assignment and TranslationOrder, also the OrderNumber is the primary key of the TranslationOrder relation schema.

Agent connects Assignment and TranslationAgency, Agent is the primary key of TranslationAgency.

Name and Language connects Assignment and IsExpert, Name and Language are the primary keys of IsExpert.

Name connects Assignment and Translator, Name is the primary key of Translator.

c) [2 marks] For each of the relation schemas, decide which attributes must be declared as not null. Explain your answer.

Answer:

- TranslationAgency: The primary key {AgencyNumber} must be declared as not null.
- Translator: The primary key {Name} must be declared as not null.
- IsExpert: The primary key {Name, Language} must be declared as not null.
- TranslationOrder: The primary key {OrderNumber} must be declared as not null.
- Assignment: The primary key {OrderNumber, Part} must be declared as not null.

Because the primary key can not be null but the foreign key may contain null value.

d) [5 marks] Assume, the translator with name 'Peter Pan' in the Translator relation retires. When deleting the record of this translator from the Translator relation, all the assignments made to him should not be lost. How would you ensure this requirement? Explain your answer.

Answer: Let all the assignments made to 'Peter Pan' should not be lost. Can set these assignments Name attribute to null then they would not be lost.

e) [5 marks] Assume, a translation order with order number '42' in the TranslationOrder relation is cancelled. Suppose, however, that already some assignments have been made

to translate parts of this translation order. When deleting the record of the translation order from the TranslationOrder relation, then all the assignments should be deleted, too. How can this requirement be ensured? Explain your answer.

Answer: Because {OrderNumber} is referred to by foreign keys in Assignment, then deleting ‘42’ in the TranslationOrder relation will make it in Assignment relation inconsistent. Delete ‘42’ from the TranslationOrder relation, and all their translation order records will refer to nothing. Therefore it can delete tuples in Assignment relation that refer to the order number ‘42’ to make the deleting consistent.

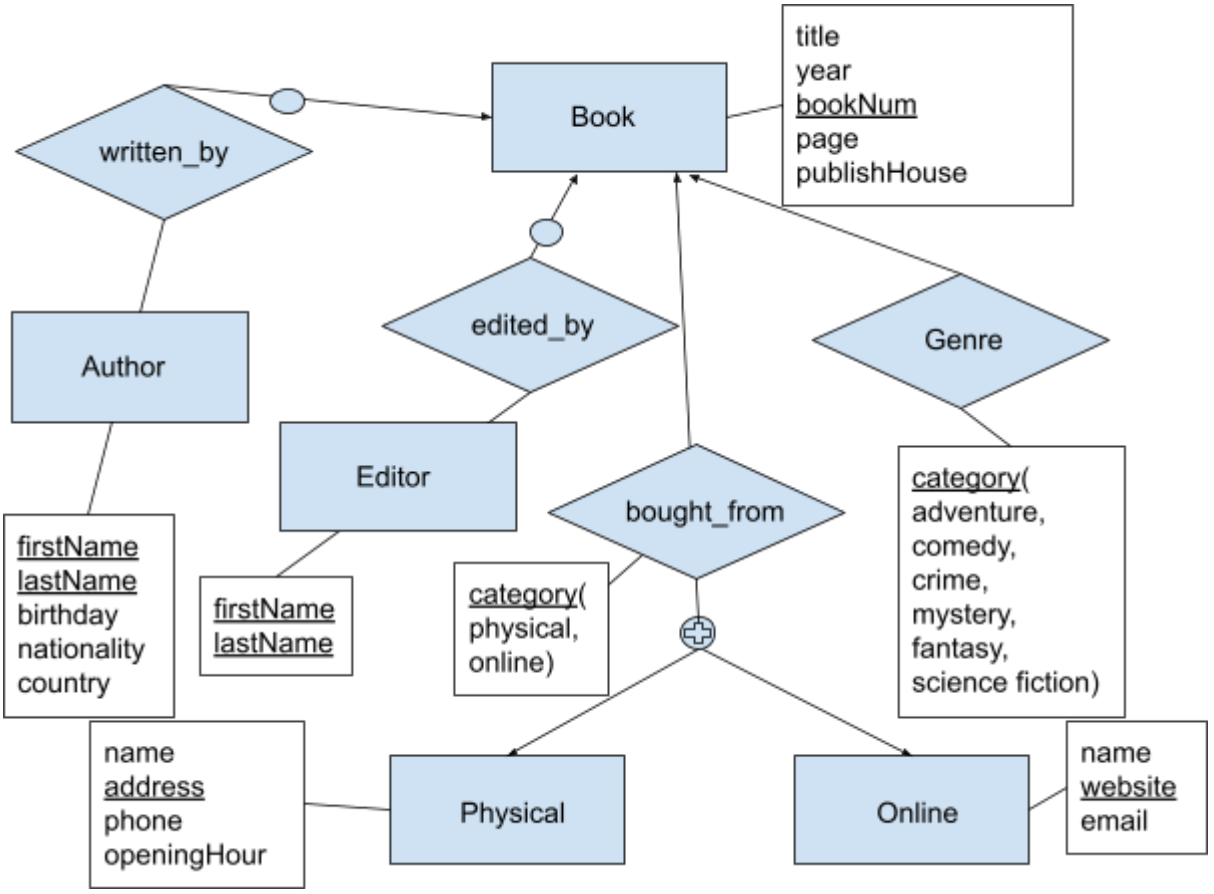
Question 4 [30 marks]

You are asked to design a new database for your grandma's collection of books. A book has a title, a release year, a unique international standard book number, a number of pages and was published by a certain publishing house.

A book can have one or more authors. The authors of a book are writers. A writer has a first name, a last name, a birthday, a nationality, and a home country. A book can have one or more editors. An editor has a first name, a last name. Editors oversee the emergence of a book from the first manuscript to the print-ready form. A book without authors has at least one editor.

Furthermore, your grandma buys books at certain bookstores which are either physical ones or online ones. Physical bookstores have a name, an address, a contact phone number and opening hours, while online bookstores have a name, a website and a contact email address. There are different genres such as adventure, comedy, crime, mystery, fantasy or science fiction. For every book at most two genres should be recorded in the database.

a) [24 marks] Draw an extended ER diagram for the database above. Write down the corresponding extended ER schema, including declarations of all the entity types (showing attributes and keys) and relationship types (showing components, attributes and keys).



Entity Types

- Book

  attributes: title, yearPublish, bookNum, pageNum, publishHouse

  primary key: bookNum

- Physical

  attributes: name, address, phone, openingHour

  primary key: address

- Online

  attributes: name, website, email

  primary key: website

- Author

  attributes: firstName, lastName, birth, nationality, country

  primary key: lastName, firstName

- Editor:

  attributes: firstName, lastName

  primary key: lastName, firstName

Relationship Types(order 1)

- written_by

  components: Book, Author

  attributes: none

  primary key: Book, Author

- edited_by

  components: Book, Editor

  attributes: none

  primary key: Book, Editor

- Genre

components: Book

attributes: category

primary key: Book, category

- bought_from

  components: Book, Physical, Online

  attributes: category

  primary key: Book, category

RDM

- Book

  attributes: title, yearPublish, bookNum, pageNum, publishHouse

  primary key: bookNum

- Physical

  attributes: name, address, phone, openingHour

  primary key: address

- Online

  attributes: name, website, email

  primary key: website

- Author

  attributes: firstName, lastName, birth, nationality, country

  primary key: firstName, lastName

- Editor

  attributes: firstName, lastName

  primary key: firstName, lastName

- written_by

  attributes: bookNum, firstName, lastName

  primary key: bookNum, firstName, lastName

foreign keys: [bookNum] ⊆ Book[bookNum] ,

[firstName, lastName] ⊆ Author[firstName, lastName]

- Editted_by

  attributes: bookNum, firstName, lastName

  primary key: bookNum, firstName, lastName

  foreign keys: [bookNum] ⊆ Book[bookNum]

  [firstName, lastName] ⊆ Editor[firstName, lastName]

- Genre

  attributes: bookNum, category

  primary key: bookNum, category

  foreign keys: [bookNum] ⊆ Book[bookNum]

- bought_from

  attributes: bookNum, category

  primary key: bookNum, category

  foreign keys: [bookNum] ⊆ Book[bookNum]

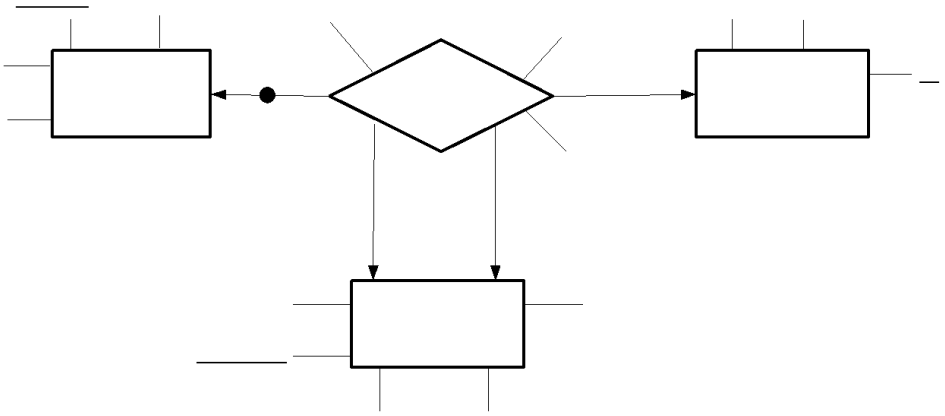
b) [6 marks] There may be information, requirements or integrity constraints that you are not able to represent in your diagram. Give three examples of integrity constraints that have not been represented in your diagram.

***Remark*:** Whenever you feel that information is missing in the problem description above, add an assumption and make your assumption explicit. In practice, you would consult the domain experts or potential users for clarification.

Answer: This diagram doesn't show the connection between Author and Editor so it can not represent a book without authors has at least one editor.

Question 5 [20 marks]

Consider the following extended ER diagram:



a) [5 marks] Present the extended ER schema of the extended ER diagram above.

Entity Types:

- Item
  - attributes: itemID, value, itemDescription, weight
  - primary key: itemID
- Courier
  - attributes: id, name, phone
  - primary key: id
- Customer
  - attributes: name, phone, address, postcode, country
  - primary key: address

Relationship Types (order 1):

- Shipment
  - components: Item, Courier, from: Customer, to: Customer
  - attributes: dateSent, dateReceived, status
  - primary key: Item, Courier, from: Customer, to: Customer

b) [10 marks] Transform your extended ER schema into a relational database schema. In particular, list all the relation schemas in your relational database schema. For each relation schema, list all attributes, the primary key, the NOT NULL constraints, and the foreign keys.

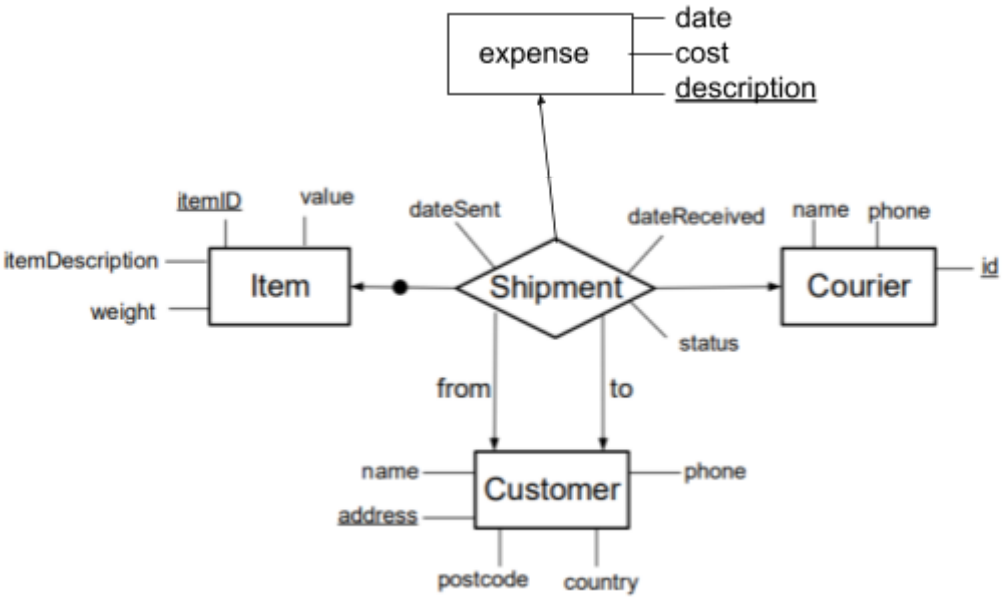
- Item
  - attributes: itemID, value, itemDescription, weight
  - primary key: itemID
  - Not Null: itemID
  - foreign keys: none
- Courier
  - attributes: id, name, phone
  - primary key: id
  - Not Null: id

    - foreign keys: none
- Customer
    - attributes: name, phone, address, postcode, country
    - primary key: address
    - Not Null: address
    - foreign keys: none
- Shipment
    - attributes: itemID, id, fromCustomer_address, toCustomer_address, dateSent, dateReceived, status
    - primary key: itemID
    - Not Null: itemID
    - foreign keys: [itemID] ⊆ Item[itemID],
      [id] ⊆ Courier[id],
      [fromCustomer_address] ⊆ Customer[address],
      [toCustomer] ⊆ Customer[address]

c) [5 marks] We also want to record information about expenses related to shipments. Each related expense has a date, a cost, and a description. The related expenses of a given shipment can be uniquely identified by the description.

Enhance the given extended ER diagram to reflect this additional information. Present the extended ER diagram with your proposed enhancements. Justify your answer.

Answer:



****