Supplementary Exam : Database Systems (INFO20003_2020_SM2)

Started: Dec 22 at 10:00

Quiz Instructions

Academic Integrity Declaration

By commencing and/or submitting this assessment I agree that I have read and understood the **University's policy on academic integrity. (https://academicintegrity.unimelb.edu.au/#online-exams)**

I also agree that:

1. Unless paragraph 2 applies, the work I submit will be original and solely my own work (cheating);
2. I will not seek or receive any assistance from any other person (collusion) except where the work is for a designated collaborative task, in which case the individual contributions will be indicated; and,
3. I will not use any sources without proper acknowledgment or referencing (plagiarism).
4. Where the work I submit is a computer program or code, I will ensure that:
   a. any code I have copied is clearly noted by identifying the source of that code at the start of the program or in a header file or, that comments inline identify the start and end of the copied code; and
   b. any modifications to code sourced from elsewhere will be commented upon to show the nature of the modification.

Notes:

Exam Support Links

Support DURING the exam is available here (for first 45 minutes of exam): ***Exam support Open here (opens in new tab)***

Otherwise, please send an email to:

Assignment Project Exam Help

renata.borovica@unimelb.edu.au (mailto:renata.borovica@unimelb.edu.au) and **colton.carner@unimelb.edu.au (mailto:colton.carner@unimelb.edu.au)** and **farah.khan@unimelb.edu.au (mailto:farah.khan@unimelb.edu.au)**

https://powcoder.com

Uploading

[see the file upload question below]

Add WeChat powcoder

Images

Many questions make use of images. Above each image there is an 'image link' which opens the image in a new tab / downloads it (depending on browser) in case the image is not loading or if you'd like to just have it open in a separate window while working.

Section breakdown (total 120 marks)

1. Relational Database Modelling and Implementation (23 marks)
2. SQL and Relational Algebra (24 marks)
3. Query Processing and Optimisation (28 marks)
4. Normalisation (10 marks)
5. Data Warehousing (12 marks)
6. Database Concepts [Transactions, Capacity Planning, Distributed DBs, NoSQL] (23 marks)

Question 1	0 pts

File Upload

There are several questions which require an answer to be attached as a scan/screenshot. Instead of scanning and uploading a document for each question, you may scan all of the working for these questions together into one image/document and attach it below if you prefer.

The following questions should be included:

- Modelling

- Relational Algebra [2 x questions]

Upload | Choose a File |

Section 1: Relational Database Modelling and Implementation (25 marks)

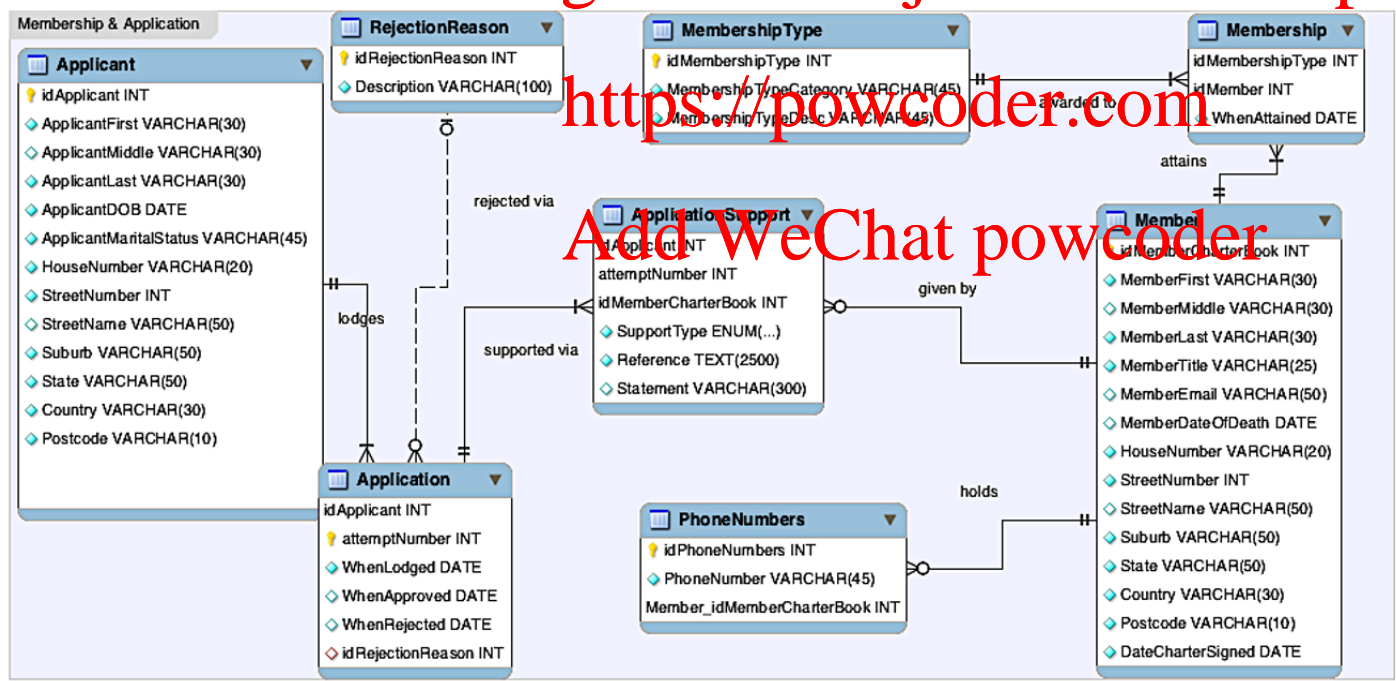**Question 2**   15 pts

Modelling

Below is a part of Physical ER model created by most of you for The Royal Society in Assignment 1.

(link to image: **examModellingSchema1.png**)

The Royal Society is interested in an extension to their database to store some additional relevant data. Your job is to extend the solution given above using the following information:

The Royal society keeps catalogues containing publications or other digital objects associated with its members. Publications can include journal papers or books. 'Other digital objects' may be videos, audios, pictures (paintings, prints, drawings, photographs), unpublished manuscripts and historical letters of correspondence, or other some other type of object. Provisional members of the Royal Society are eligible to receive full access to the publications but not to other digital objects unless these objects are public in which case everyone can view these resources.

The Royal Society maintains a publication list for all restricted and full members of the society that contains every journal paper they have ever published or any book they have authored. A paper or a book are identified by a unique publication ID and can be authored by more than one member of the society. Members may have collaborated with non-member scientists and published with them. In such cases, the database does not store any additional information about non-member authors except their names. Each paper may have a Digital Object Identifier (DOI). The database also saves Journal International Standard Serial Number (ISSN), journal name, paper title, authors list, publication date of the paper, journal impact factor at the time of publication and volume number. The impact factor of journal may vary depending on the year of publication, and as such we wish to keep track of impact factor variations throughout years. For a given book, the details recorded are: title of the book, price, publication date, unique International Standard Book Number (ISBN), authors' names, 500 words long description (that usually appears at the back of the book), language and genre.

For other digital objects, the Royal Society generates a unique catalogue ID for each object when it is entered in the database. The database keeps track of the number of times a particular object is accessed. Some of these resources are downloadable. If that is the case, the number of downloads is also updated regularly to keep up to date metrics about each object.

Draw a **physical model using Crow's foot notation** for the extension. You do *not* need to draw out the solution given in the above image (from A1), you can simply draw the additional extension tables, and redraw any tables from the above solution which you modify and/or form relationships with as part of your extension (no need to repeat the existing attributes from these tables).

Be sure to write down any assumptions you make.

When drawing the model, you can indicate if something is a PFK, PK, FK, NOT NULL (NN), Null, etc. with a simple notation such as:

(Image link: **exampleDrawing.png**)
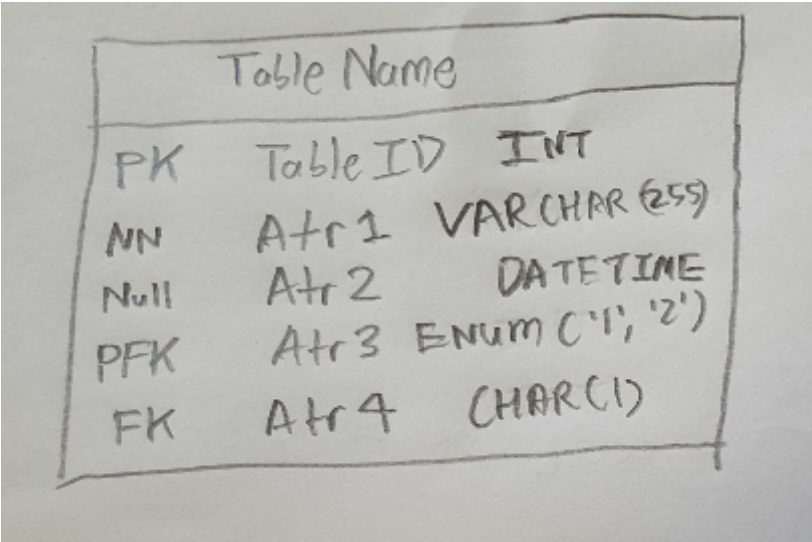
Remember you may upload the scan of this question below, or combine and scan all handwritten questions together (see top of the exam, question 1)

Upload   Choose a File

Question 3	8 pts

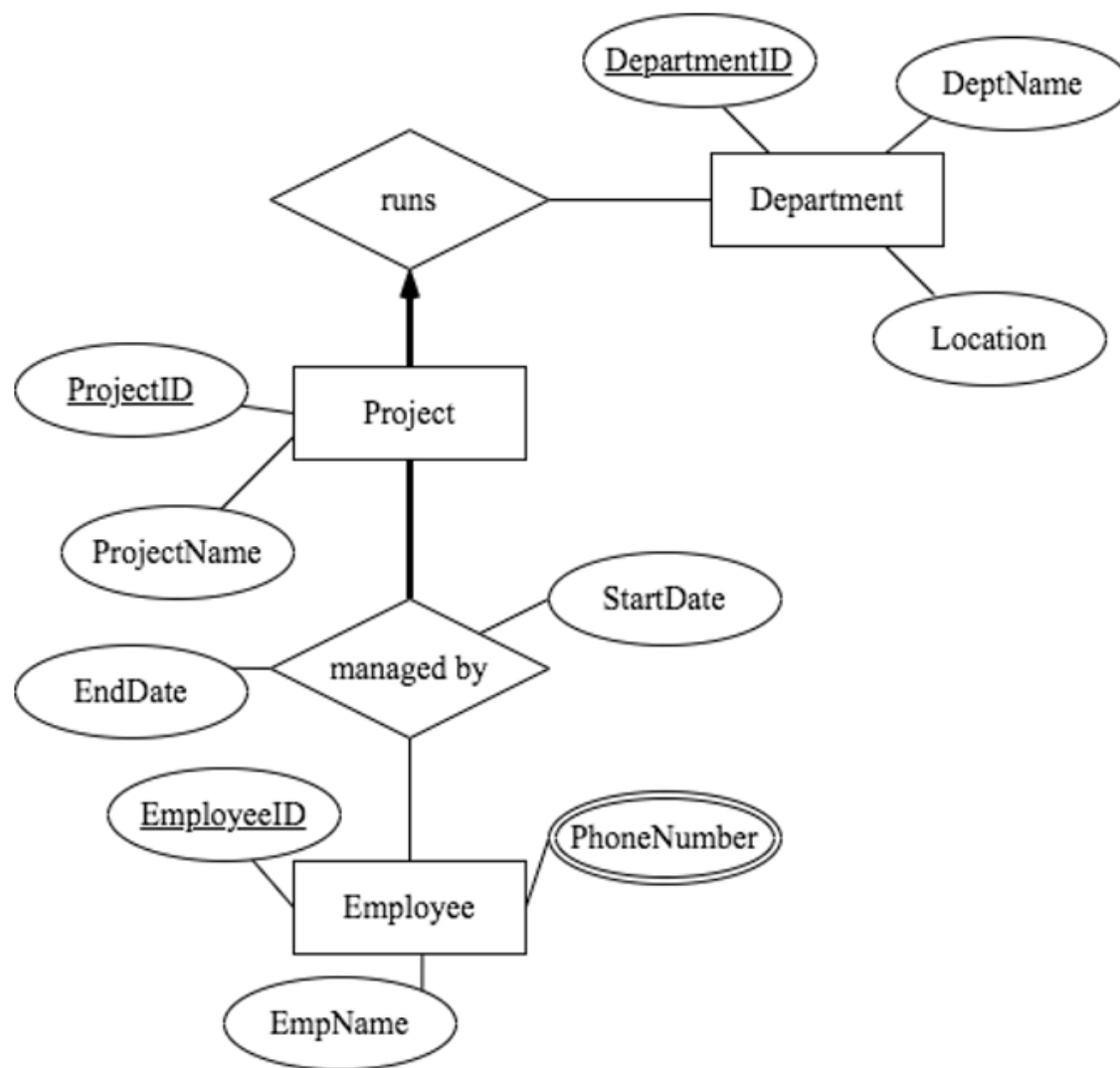## DDL

In the following conceptual model, employees can manage projects run by a department. Each department can run multiple projects, but a project is run by a single department. Overall, each project must be managed by at least one employee and can have more employees involved for managing the project during its different stages.

(Image link: **DDL.png**)

Write SQL statements to *create* the tables for the data model shown above. Be sure to specify primary and foreign keys. You *do not* need to specify whether the fields are NULL/NOT NULL. Where the data type is not obvious, feel free to choose an appropriate data type.

Assignment Project Exam Help

| Edit | View | Insert | Format | Tools | Table |

| 12pt | Paragraph | **B** | *I* | U | A | T |

https://powcoder.com

Add WeChat powcoder

p   0 words   </> ↗ ⋮

Section 2: SQL (DML) and Relational Algebra (24 marks)

SQL: Unichat

You and a group of fellow undergrads have created the UniChat startup company. The company's goal is to run a popular online discussion forum for students at the University of Melbourne. *Please note that the following data model and the description is*

similar to Assignment 2.

The system has two kinds of users: students and lecturers. Most user attributes are straightforward personal information and are listed in the ER diagram below. "Karma" is a score awarded for being active in the system and posting popular posts.

Discussions are organized into forums, each of which is about a particular topic. Individual forums can be opened and closed by lecturers. When a forum is closed it is not deleted, but simply marked as closed and hidden from users. Users can post a text message into any open forum and can comment on existing posts. Comments are stored in the same table as posts and are connected to their parent posts via a unary relationship. Users can also click a "Like" button which is attached to every post or comment; this is recorded so it can be represented to other users. Because posts and comments are the same entity, questions that refer to "posts" mean "posts and/or comments" unless we specify "top level posts". Note that in the model, "top level posts" have a FK to the forum they are posted in, but comments have NULL for forum FK (since they're not a "forum" post).

Students (but not lecturers) can add each other to friend-lists, which are used to influence the ordering of posts and the other website behaviour. When one student sends a friend-request to another, the latter can reject or accept the friendship. If the latter accepts, the pair are now friends. Note that friendship is a symmetric relation: if A is a friend of B, then B is a friend of A, whether A sent the friend request to B or vice-versa. Once a pair of users are friends, either may later unfriend the other, in which case the friendship ends for both of them. Each time a friend-request is sent, rejected, accepted or unfriended; this is recorded in the FriendOf table by adding a timestamp in the appropriate column. We calculate who a given student's friends are by querying this table: a user's "friends" means their current friends, not those where the friendship has ended or not begun.

The unichat MySQL relational database schema is shown below:

(Image link: **unichat.png**)
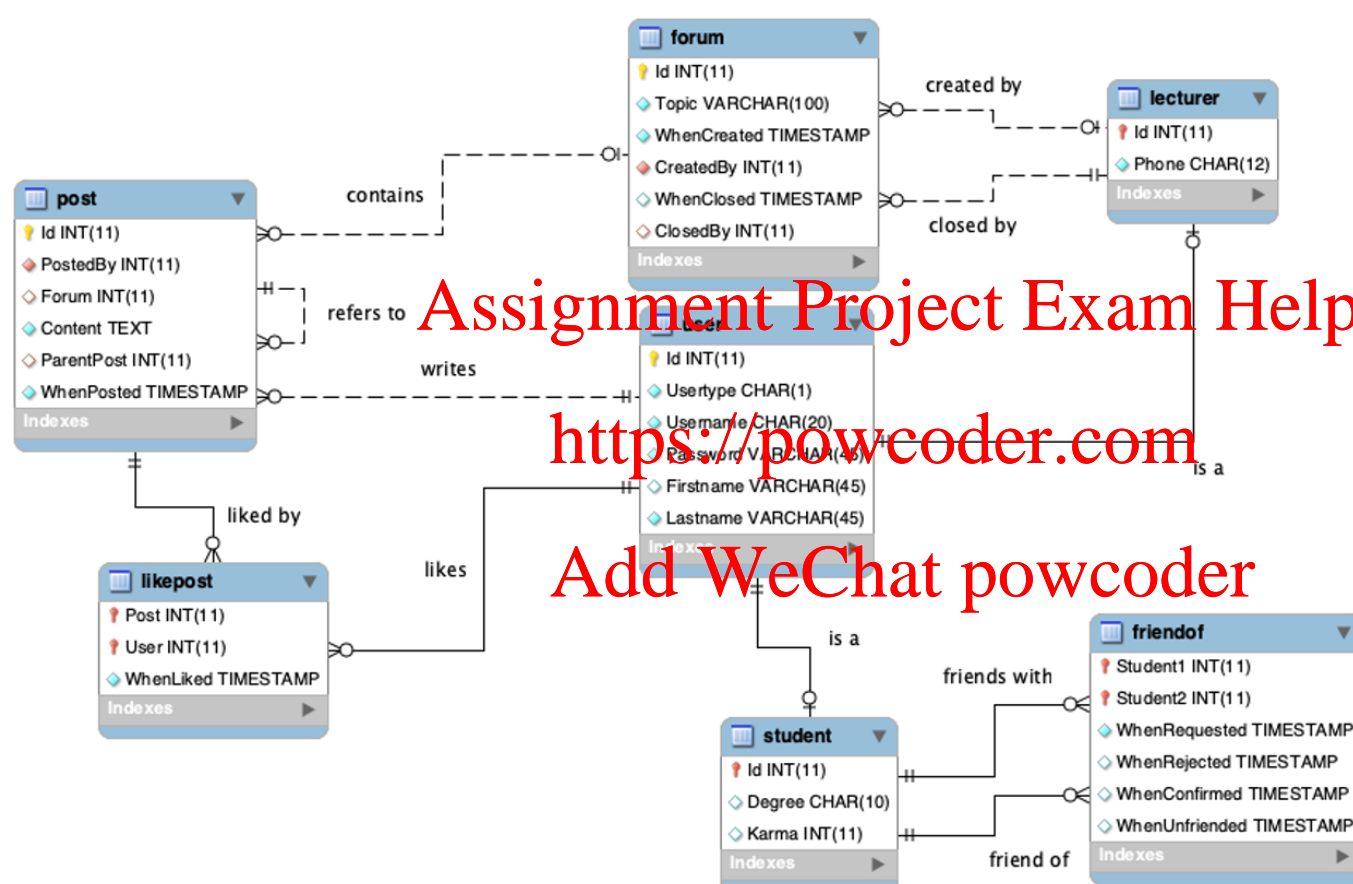
Write a **single** SQL SELECT statement for each of the following questions (QA-D). DO NOT USE VIEWS or VARIABLES to answer questions. Subqueries are allowed.

Question 4   2 pts

SQL Question A

Find the lecturers who have never created or closed a forum.

Display as (Lecturer Id, Lecturer username).

Edit   View   Insert   Format   Tools   Table

12pt   Paragraph | **B**   *I*   U   A   ✎   T² | ⋮

p   ⌨ ⓰ | 0 words | </> ↗ ⋮

Question 5   3 pts

SQL Question B

For every forum, find the total number of likes which top-level posts made by lecturers other than the forum's creator have received.

Display as (forum ID, forum topic, total number of likes from non-creating lecturers).

Present the results in descending order of the number of likes.

Edit   View   Insert   Format   Tools   Table

12pt ∨   Paragraph ∨   │   B **Assignment Project Exam Help**

https://powcoder.com

Add WeChat powcoder

p   ⌨ ⓰ | 0 words | </> ↗ ⋮

Question 6   6 pts

SQL Question C

We'll consider the 'popularity of a database technology in 2020' (referred to as *popularity*) to be the sum of the number of top-level posts which were posted in a forum that *includes the name of that database technology in the forum topic*, and the post was made in 2020.

As an example, imagine that the database was populated with the following forums and corresponding number of top-level posts:

Forum Topic	Number of top level posts in 2020
Professional MongoDB Questions	22
The MySQL Masters	8

My First MySQL	12
MongoDB Madness	6

Then in this case, the correct result for *popularity* of 'MySQL' is 20 (12+8) and the *popularity* of 'MongoDB' is 28 (6+22).

Find the *popularities* of the following technologies: "MySQL" and "MongoDB".

Display as (technology name, popularity of the database technology in 2020).

Edit   View   Insert   Format   Tools   Table

12pt ⌄   Paragraph ⌄ | **B**   *I*   U̲   A ⌄   ✏ ⌄   T² ⌄ | ⋮

p   0 words   </>   ↗   ⋮

Question 7   8 pts

SQL Question D

List the IDs of all students who are not friends with someone doing a degree of type 'MIT'. Only consider friendships which are 'confirmed'.

Display as (Student ID)

Edit   View   Insert   Format   Tools   Table

12pt ⌄   Paragraph ⌄ | **B**   *I*   U̲   A ⌄   ✏ ⌄   T² ⌄ | ⋮

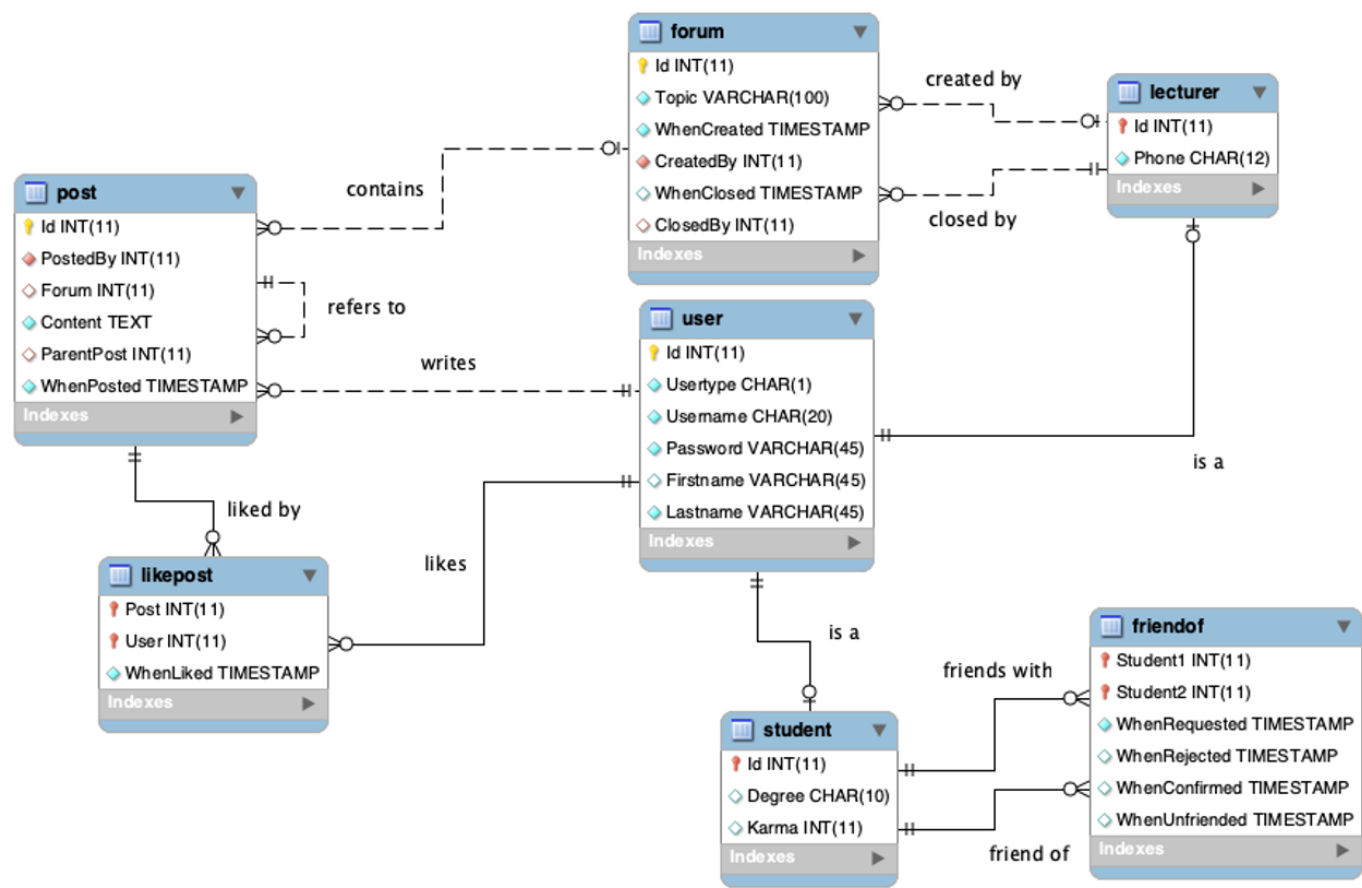p   0 words   </>   ↗   ⋮

Relational Algebra: Unichat

The unichat model is again provided here for reference:

(Image link: **unichat.png**)

Please write a **relational algebra** expression for each of the following questions using this model. This query should be uploaded as an image.

Remember you may upload the scan of these questions below, or combine and scan all handwritten questions together (see top of the exam, question 1).

Question 8	2 pts

RA Question A

Display the topics for all forums created in 2020 by lecturer with id '2'.

Upload   Choose a File

Question 9	3 pts

RA Question B

Display the content of every top-level post made in forums that are closed, for the forums that are created by lecturer with username "MaxF"

Upload   Choose a File

Section 3: Query Processing and Optimisation (28 marks)

Single Relation Plans

Consider a relation called Books that stores information about ordered books from a bookstore. Imagine that the relation Books consists of 5000 data pages and each page stores 100 tuples. Imagine that the *totalprice* attribute can take any value between 0 and 10000 ([0,10000]), and imagine that *amount* can take any value between 1 and 100, specifically [1,100]. Suppose that the following SQL query is executed frequently using the given relation:

```
SELECT genre

FROM Books

WHERE totalprice < 4000 AND amount = 20;
```

Answer the following questions A-C.

Question 10   **3 pts**

Single Relation Question A

Compute the estimated result size for the query, and the reduction factor of each filter.

RF(totalprice)	[Choose] ∨
RF(amount)	[Choose] ∨
Result Size	[Choose] ∨

Question 11   **3 pts**

Single Relation Question B

Compute the estimated cost of plan alternatives and *state which plan is the best* assuming that an *unclustered B+ tree* index on (*amount*) is (the only index) available. Suppose there are 100 index pages.

Give the lowest (estimated) cost after considering all access methods available.

[   ]

Joins

Consider two relations called Patient and Diagnosis. Imagine that the relation Patient has 1000 pages and the relation Diagnosis has 5000 pages. Consider the following SQL statement:

```
SELECT *

  FROM Patient INNER JOIN Diagnosis

    ON Patient.patientID = Diagnosis.patientID

WHERE Diagnosis.test = 'GDD';
```

There are 202 buffer pages available in memory. Both relations are stored as simple heap files. Neither relation has any indexes built on it.

Question 12
4 pts

Joins Question A

Evaluate the following join types using the number of disk I/O's as the cost. Consider Patient as the outer relation in all alternatives. Assume that sorting can be performed in two passes for both relations. All selections are performed on-the-fly after the join.

Block-oriented Nested Loop Join	[Choose] ⌄
Nested Loop Join	[Choose] ⌄
Hash Join	[Choose] ⌄
Sort Merge Join	[Choose] ⌄

Question 13
2 pts

Joins Question B

With relation to Question A above, which join would be the optimizer's choice?

○ Block Nested Loop Join

○ Hash Join

○ Sort-Merge Join

○ Nested Loop Join

Query Optimisation

Consider the following relational schema and SQL query. The schema captures information about patients, their diagnostic test referrals, and the test details.

```
Patient (PatientID, firstname, lastname, address, phonenumber)

Prescription (RefID, datefilled, doctorname, clinicname, PatientID, MedicineID)
```

```
Medicine (MedicineID, name, cost, description, type)
```

Consider the following query:

```
SELECT *

  FROM Patient P, Prescription as Pres, Medicine as M

  WHERE P.PatientID = Pres.PatientID

  AND Pres.MedicineID =   Medicine.MedicineID

  AND M.cost > '1000' AND Pres.clinicname = 'Carlton Medical';
```

The system's statistics indicate that there are 500 registered clinics, and cost of medicines range from 0 to 5000 ([0,5000]) since some medicines may be offered free of cost. There is a total of 10,000 Patients and 20,000 prescriptions and 2,000 medicines in the database. Each relation fits 100 tuples in a page. Suppose there exists a clustered B+ tree index on Medicine.cost and another one on Patient.PatientID, both of size 10 pages.

Compute the cost of the plans shown below. Assume that sorting of any relation (if required) can be done in 2 passes. NLJ is a Page-oriented Nested Loops Join. Assume that 100 tuples of a resulting join between Patient and Prescription fit in a page. Similarly, 100 tuples of a resulting join between Prescription and Medicine fit in a page. If selection over filtering predicates is not marked in the plan, assume it will happen on-the-fly after all joins are performed, as the last operation in the plan.

Question 14	Assignment Project Exam Help	6 pts

Query Optimisation Question A

https://powcoder.com

Compute the following results for the below Plan 1:

Add WeChat powcoder

1. The result size of the child join in pages (marked as 'A' in the diagram)
2. The cost of the child join in I/Os (marked as 'B' in the diagram)
3. The cost of the parent join in I/Os (marked as 'C' in the diagram)

(Image link: **annotatedPlanA.png**)
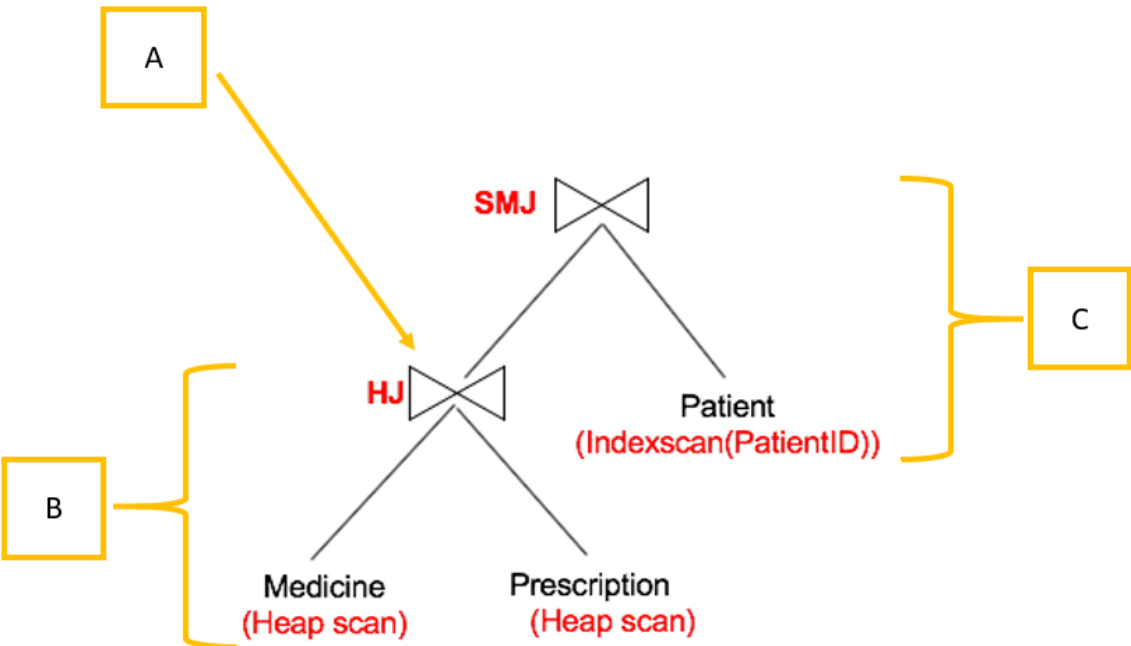
The result size of the child join in pages (marked as 'A' in the diagram).	[Choose] ∨
The cost of the child join in I/Os (marked as 'B' in the diagram)	[Choose] ∨

The cost of the parent join in I/Os (marked as 'C' in the diagram)   [Choose] ⌄

Question 15   10 pts

Query Optimisation Question B

Compute the following results for the below Plan 2:

1. The result size of the selection on cost in pages (marked as 'A' in the diagram)
2. The result size of the child join in pages (marked as 'B' in the diagram)
3. The cost of the selection on cost in I/Os (marked as 'C' in the diagram)
4. The cost of the child join in I/Os (marked as 'D' in the diagram)
5. The cost of the parent join in I/Os (marked as 'E' in the diagram)

(Image Link: **annotatedPlanB-1.png**)
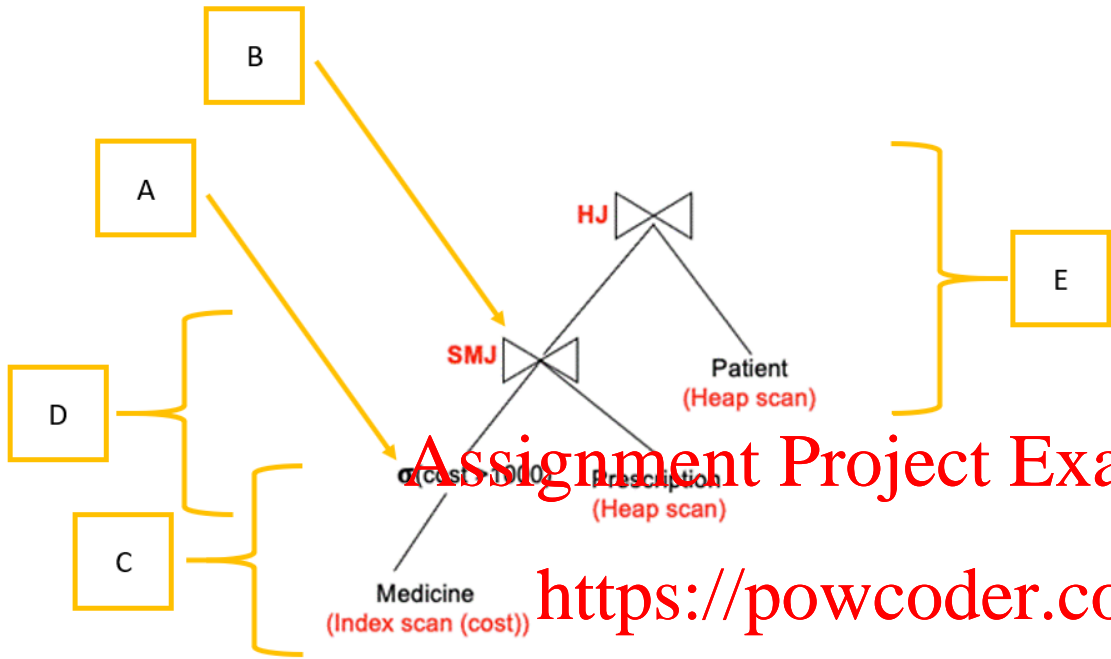
The result size of the selection on cost in pages (marked as 'A' in the diagram)   [Choose] ⌄

The result size of the child join in pages (marked as 'B' in the diagram)   [Choose] ⌄

The cost of the selection on cost in I/Os (marked as 'C' in the diagram)   [Choose] ⌄

The cost of the child join in I/Os (marked as 'D' in the diagram)   [Choose] ⌄

The cost of the parent join in I/Os (marked as 'E' in the diagram)   [Choose] ⌄

Section 4: Normalisation (10 marks)

Normalisation

The table below contains a grade report with details about the students, subjects, campus and instructors:

GradeReport(StudentID, StudentName, PhoneNumbers, CampusAddress, Major, SubjectID, SubjectTitle, Instructor, Location, Grade)

StudentID	SName	PhoneNumbers	CampusAddress	Major	SubjectID	SubjectTitle	Instructor	Location	Grade
612345	Robinson	+61400890210, +61431112345	708 Parkville	CIS	ISYS90051	Ethics	Alex	B203	A
612345	Robinson	+61400890210, +61431112345	708 Parkville	CIS	ISYS10001	Software	Parsons	B317	B
998904	Baker	+61451902341	333 Carlton	Med	ISYS90051	Ethics	Alex	B203	C
998904	Baker	+61451902341	333 Carlton	Med	BTECH034	Microbio	Miller	B223	B
998904	Baker	+61451902341	333 Carlton	Med	BTECH002	Cellbio	Benedict	B554	A

The pair (StudentID, SubjectID) is the candidate key for this relation. The following functional dependencies hold for this relation:

StudentID ->SName, PhoneNumbers, CampusAddress, Major

StudentID, SubjectID -> Grade

SubjectID -> SubjectTitle, Instructor, Location

Instructor-> Location

Answer the below questions using this relation's schema.

Question 16   1 pts

Normalisation Question A

In what normal form is this relation?

○ Not in 1st Normal Form

○ 1st Normal Form

○ 2nd Normal Form

○ 3rd Normal Form

Question 17   1 pts

Normalisation Question B

Which of the following types anomalies could arise in the schema given above?

☐ Insertion Anomaly

☐ Deletion Anomaly

☐ Update Anomaly

| **Question 18** | **8 pts** |

Normalisation Question C

Normalize the GradeReport relation to the 3rd Normal Form (3NF). For each step explicitly identify which normal form is violated, **briefly explain why,** and give the schema normalised to that form. Write the normalised tables in a textual format. Here's an example step:

```
...

Not in NF x since y. Normalised to NF x is:

 TableName (id(PK), Column, ForeignKeyColumn(FK))
 AnotherTable (id(PK), Column, AnotherColumn(PFK))
...
```

Edit   View   Insert   Format   Tools   Table

12pt ∨   Paragraph ∨ | **B**   *I*   U̲   A ∨   🖉 ∨   T² ∨ | ⋮

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

p   0 words   </> ↗ ⋮
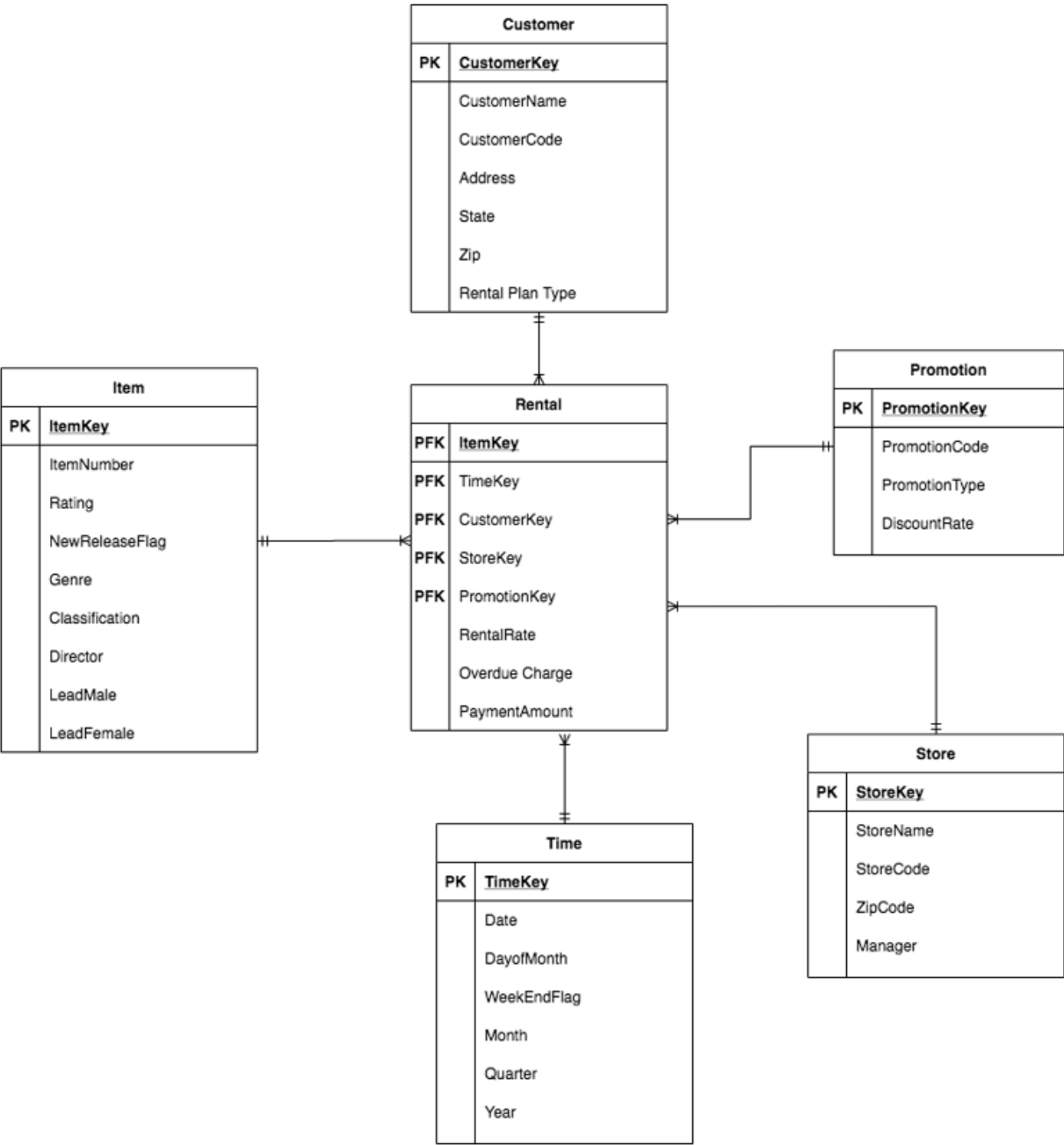
Section 5: Data Warehousing (12 marks)

Data Warehousing

The following figure shows a data model of a data warehouse for a company focused on **"Movie Rentals"** :

(Image link: **datawarehousing.png**)

Customer

PK	CustomerKey
	CustomerName
	CustomerCode
	Address
	State
	Zip
	Rental Plan Type

Item

PK	ItemKey
	ItemNumber
	Rating
	NewReleaseFlag
	Genre
	Classification
	Director
	LeadMale
	LeadFemale

Rental

PFK	ItemKey
PFK	TimeKey
PFK	CustomerKey
PFK	StoreKey
PFK	PromotionKey
	RentalRate
	Overdue Charge
	PaymentAmount

Promotion

PK	PromotionKey
	PromotionCode
	PromotionType
	DiscountRate

Store

PK	StoreKey
	StoreName
	StoreCode
	ZipCode
	Manager

Time

PK	TimeKey
	Date
	DayofMonth
	WeekEndFlag
	Month
	Quarter
	Year

Answer the questions below in relation to this model.

Assignment Project Exam Help

https://powcoder.com

Question 19

Add WeChat powcoder

1 pts

Data Warehousing Question A

Is the given model a star schema?

- ○ True
- ○ False

Question 20

3 pts

Data Warehousing Question B

Match the aspects/components of the model to their definition.

Rental Rate	[Choose] ⌄
Promotion	[Choose] ⌄
Rental	[Choose] ⌄

Movie Rentals	[Choose] ⌄
Overdue Charge	[Choose] ⌄

Question 21   **6 pts**

Data Warehousing Question C

Select which of these queries _**could**_ be answered using the above model

What was the total sum of the payments from rental of movies with rating of 3 or greater over the last month?	[Choose] ⌄
Rank promotions by the sum of rental rates they generated for rentals which occurred on any store located on 'Flinders Street'	[Choose] ⌄
For stores located in zip code 3000, which hour of the day brings in the most rentalRates?	[Choose] ⌄
Which state was the customer from who had the largest overdue charge of all time?	[Choose] ⌄
Find the percentage of transactions which used a promotion with a discountRate > 10%, grouped by the state of the customers	[Choose] ⌄
In 2000, were more 'new release' movies rented on DVD or VHS?	[Choose] ⌄

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Question 22   **2 pts**

Data Warehousing Question D

Write an SQL statement to answer the following:

For each year, display the total sum for 'paymentAmount' generated from rentals of a movie with rating greater than 8. Only display years where the total payment amount is more than 1000.

Display as (Year, sum of paymentAmount)

Edit   View   Insert   Format   Tools   Table

12pt ⌄   Paragraph ⌄ | **B**   _I_   U   A ⌄   ✎ ⌄   T² ⌄ | ⋮

p   0 words

Section 6: Database Concepts (23 marks)

[Transactions, Capacity Planning, Distributed DBs, NoSQL]

Transactions

Consider the following relation in some database:

```
Student (studentID, firstname, lastname, DOB, marks)
```

Two transactions are being run on this relation concurrently:

Transaction T1	Transaction T2
SELECT SUM(marks) FROM Student	UPDATE Student SET marks = marks + 10 WHERE studentID = "334523"
	UPDATE Student SET marks = marks - 10 WHERE studentID = "334522"
	COMMIT;
COMMIT;	

Below is one possible interleaving of these two transactions:

(Image Link: **transactionTable-1.png**)

TIME	TRANSACTION	ACTION	VALUE	TOTAL
1	T1	Read marks for studentID = "111222"	50	50
2	T1	Read marks for studentID = "333456"	40	90
3	T2	Read marks for studentID = "334523"	80	
4	T2	UPDATE marks: marks = 80 + 10		
5	T2	Write marks for studentID = "334523"	90	
6	T1	Read marks for studentID = "334523"	90	180
7	T1	Read marks for studentID = "334522"	65	245
8	T2	Read marks for studentID = "334522"	65	
9	T2	UPDATE marks: marks = 65 - 10		
10	T2	Write marks for studentID = "334522"	55	
11	T2	*****COMMIT *****		
12	T1	Read marks for studentID = "987567"	55	300
13	T1	*****COMMIT *****		

Answer the below questions with respect to this timeline

Question 23   3 pts

Transactions Question A

What kind of problem can arise when these two transactions run in parallel on the table student as shown below? Briefly describe why this is an issue.

Edit   View   Insert   Format   Tools   Table

12pt ⌄   Paragraph ⌄   **B**   *I*   U   A ⌄   ✎ ⌄   T² ⌄   ⋮

p   ⌨   ⊕ | 0 words | </>   ↗   ⋮

Question 24   1 pts

Transactions Question B

What is the <u>correct</u> value for the total sum of marks?

Capacity Planning

A company sells 5000 products. An analyst has determined that given that there are 10,000 customers, each customer places on average 5 orders per month, and each order contains on average 10 products.

(Image Link: **CapacityPlan.png**)
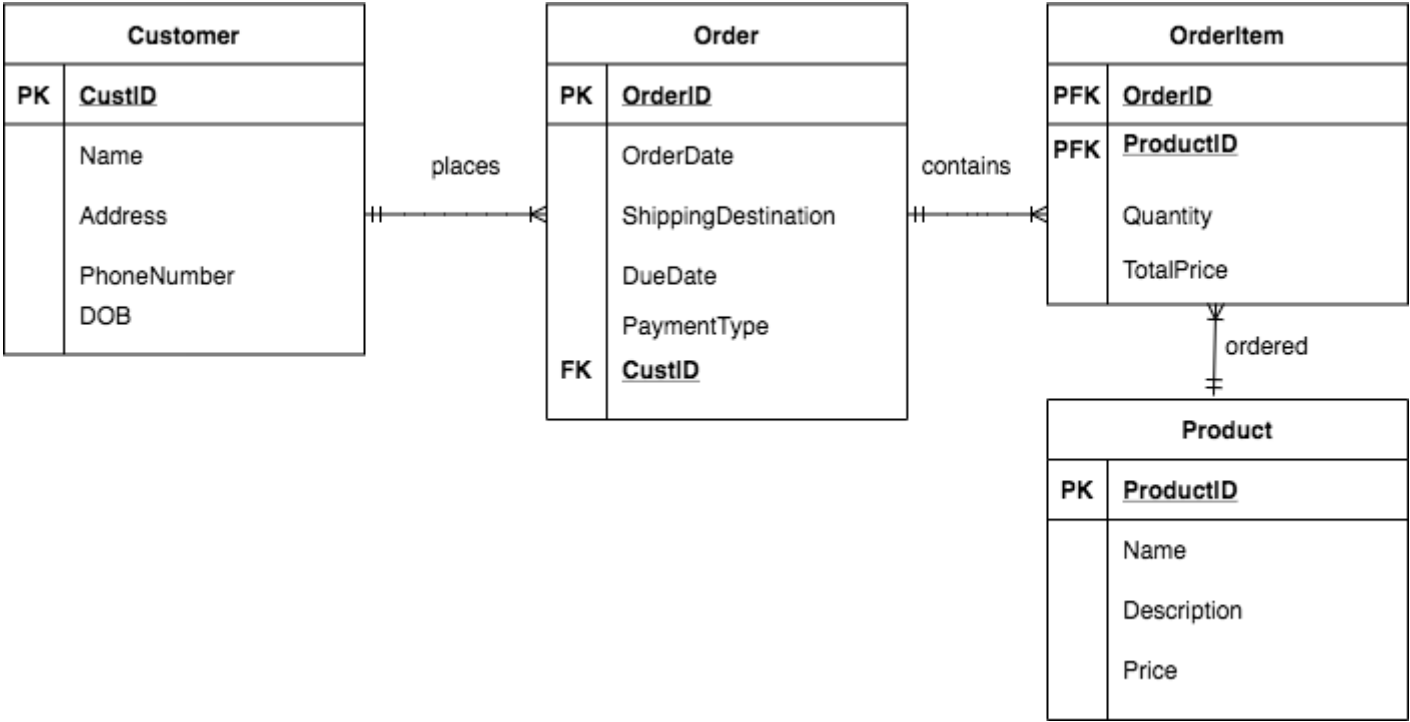
Using the MySQL information about data type storage and information from the data dictionary, the analyst has determined the following average row lengths of each table (number of rows is at go-live):

Table	Number of rows	Average row length
Customer	10,000	180 bytes
Product	5,000	160 bytes
Order	0	170 bytes
OrderItem	0	64 bytes

Assume that the number of customers and number of products does not change from year to year and that the order average remains the same as determined by the analyst.

Capacity Planning Question A

When database use begins (year 0), calculate the size of the following tables in bytes:

Question 25   **0.5 pts**

A.1

Customer Table

[   ]

Question 26   **0.5 pts**

A.2

Product Table

[   ]

Question 27   **0.5 pts**

A.3

Order Table

[   ]

Question 28   **0.5 pts**

A.4

OrderItem Table

[   ]

Capacity Planning Question B

After **five** years of database use, calculate the size of the following tables in bytes:

Question 29   **0.5 pts**

B.1

Customer Table

[   ]

Question 30   **0.5 pts**

B.2

Product Table

[   ]

Question 31   **0.5 pts**

B.3

Order Table

```
┌─────────────────────────────┐
│                             │
└─────────────────────────────┘
```

Question 32   **0.5 pts**

B.4

OrderItem Table

```
┌─────────────────────────────┐
│                             │
└─────────────────────────────┘
```

Question 33   **3 pts**

Capacity Planning Question C

The company rents a cloud server with a 15GB SSD partition, dedicated entirely to store MySQL tables for this DB.

How many years of operation could this drive accommodate?

Note that 1GB = 1024MB, 1MB = 1024 kB, 1kB = 1024 bytes. In case of a float answer, round DOWN to the closest integer (e.g. 5.67 years becomes 5 years).

Assignment Project Exam Help

```
┌─────────────────────────────┐
│                             │
└─────────────────────────────┘
```

https://powcoder.com

Add WeChat powcoder

Distributed Databases

Answer the following two questions

Question 34   **2 pts**

Distributed Databases Question A

As the business grew over time for Paradise hotel, there are multiple hotels in the same city and overall spread all over the globe. The management at Paradise Headquarters is interested in upgrading the database associated with the hotel room booking system from centralised to distributed structure. Paradise hotels in Europe specifically need to access the special rates for tourism-rich cities and upgrade status for qualifying customers. Hotels in Asia tend to access different information, to provide special discounts and additional amenities to regular customers.

Which of the following type of distributed database would be best suited for this case out of:

○ Vertically partitioned distributed database

○ Fully replicated distributed database

○ Horizontally partitioned database

Question 35 **3 pts**

Distributed Databases Question B

Why do we have communication overhead in a replicated distributed database?

| Edit | View | Insert | Format | Tools | Table |

| 12pt ⌄ | Paragraph ⌄ | **B** | *I* | U̲ | A̲ ⌄ | ✎ ⌄ | T² ⌄ | ⋮ |

p   ⌨ 🧍 | 0 words | </> ↗ ⋮

NoSQL (7 Marks)

Answer the following two questions

Question 36 **4 pts**

NoSQL Question A

Different informational and review-oriented websites store information about the published and upcoming books. Given the following information about the nature and purpose of the data stored, identify which type of the NoSQL database would be best suited for storing that website's data.

Select from the four main types of the NoSQL databases.

A. Website A provides a unique URL to download a given book and also stores the official cover page which might be PNG or PDF format.

B. Website B is more informational, and it stores the title of the book, publisher(s), authors(s), publication date(s), current Goodreads score and ISBN. This information for each book is stored as a JSON file. In the future website may allow to add supporting information if available (e.g. keywords, small bio etc.). Usually, all of the information for a book is accessed together.

C. Website C provides analytics to study the impact the authors of a given book have on its success. This website keeps track of the relationship between books and the authors to identify which book shares a number of authors.

D. Website D is focused on reviews hence it stores information about the reviewer's name, the book's name and 250 words long review for the book reviewed. These reviews are sometimes uploaded as hand-written notes, and sometimes inserted as a text. The website also allows the reviews to be submitted as an audio or video file.

Website A   [Choose] ⌄

Website B

[Choose] ⌄

Website C

[Choose] ⌄

Website D

[Choose] ⌄

Question 37   **3 pts**

NoSQL Question B

Do you think Instagram prefers eventual consistency (i.e. BASE) or conventional consistency (i.e. ACID) in its distributed databases? Explain your response in 2-3 sentences.

Edit   View   Insert   Format   Tools   Table

12pt ⌄   Paragraph ⌄ | **B**   *I*   U͟   A⌄   ✎⌄   T²⌄ | ⋮

<div style="color:red; text-align:center">

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

</div>

p   ⌨   ⓧ | 0 words | </>   ↗   ⋮

End of Exam

Remember to upload your answers to modelling + relational algebra at the top of the exam (question 1) if you've chosen to scan all questions together.

Not saved   Submit Quiz