# Examples of Midterm Questions

The questions bellow serve as examples to give you an idea of the format of the midterm. They do not cover all the material you will be examined on in the midterm.

Information

The following **incomplete** ER diagram represents crafting recipes in a game.

Each craft requires a certain level of proficiency in one job, and is uniquely identified by the corresponding job together with the craft_number. A craft always costs at least one item (but can cost multiple different items) and always produces exactly one item. An item is not necessarily used in some craft nor produced by a craft, and it can be used in multiple crafts but there can be at most one craft that produces a given item.
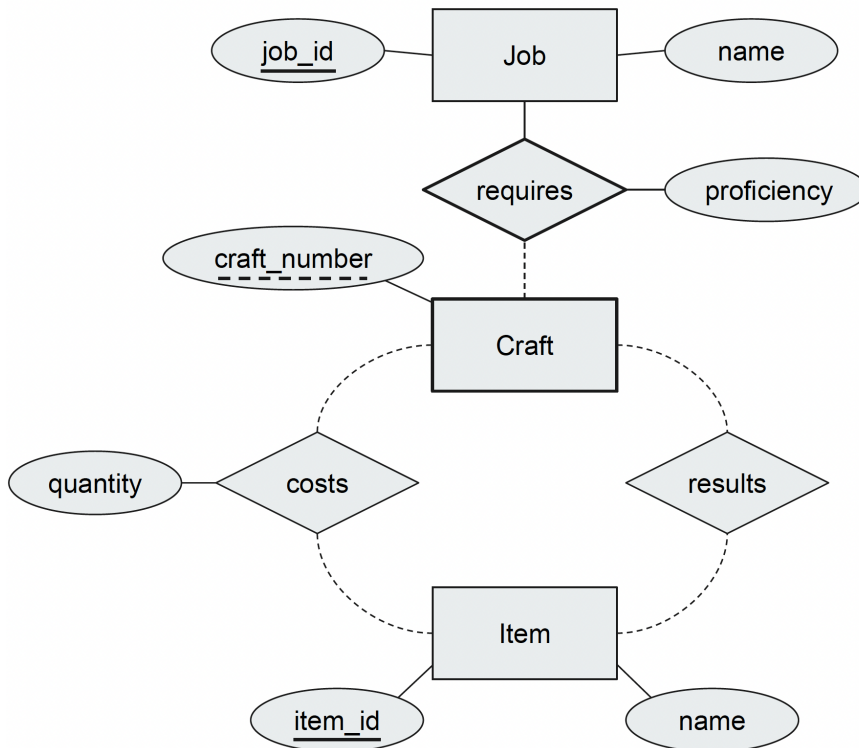
Fill in the gaps in the sentences. Words can be reused multiple times or not be used at all.

The line between "Craft" and "requires" should be: | a bold arrow |

The line between "Craft" and "costs" should be: | a bold line |

The line between "Craft" and "results" should be: | a bold arrow |

The line between "Item" and "costs" should be: | a normal line |

The line between "Item" and "results" should be: | a normal arrow |

| a normal line | | a bold line | | a normal arrow | | a bold arrow |

Information

Consider the following table.

building	floor	room	lab_id	seats
BC	2	22	22	1
BC	2	40	22	2
BC	2	60	35	1
CE	1	3	NULL	200
CM	1	3	NULL	190
INR	3	12	48	3
INR	0	12	NULL	13

Fill in the gaps in the sentences. Words can be reused multiple times or not be used at all.
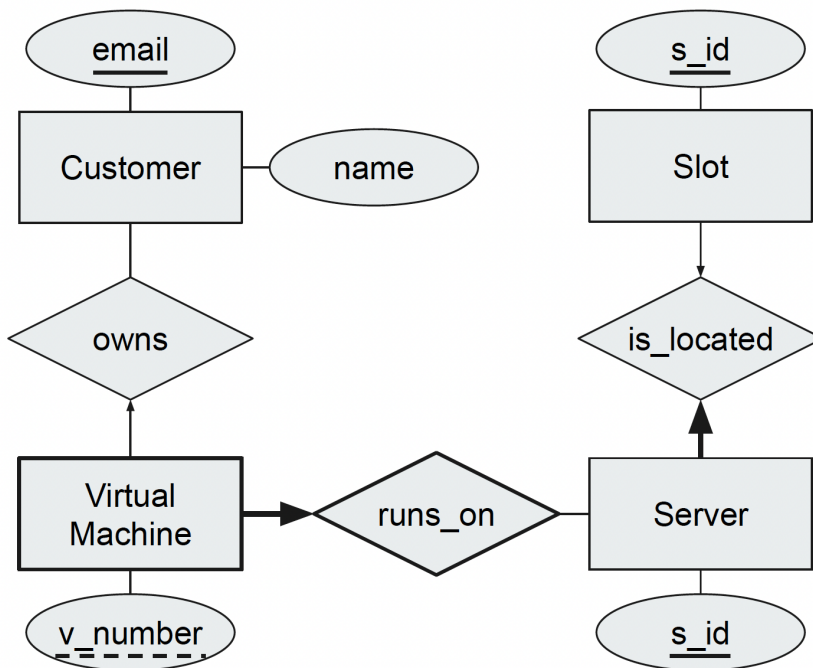
The set of attributes (floor, room) is part of a candidate key .

The set of attributes (building, floor, room) is a candidate key .

The set of attributes (building, floor, room, seats) is a superkey .

The lab_id attribute cannot be part of a candidate key because it is nullable , but it could be (part of) a foreign key .

| a candidate key | a superkey | part of a candidate key | a foreign key |

| unique | | not unique | not null |

Consider the following ER model.

Question 3

Complete

Mark 2.00 out of 2.00

Fill in the gaps in the sentences. Words can be reused multiple times or not be used at all.

The relationship "owns" between the "Virtual Machine" and the "Customer" is a | 1-N | relationship.

The relationship "is_located" between the "Server" and the "Slot" is a | 1-1 | relationship.

| 1-1 | | 1-N | | N-M |

Question 4

Complete

Mark 3.00 out of 3.00

Fill in the gaps in the sentences. Words can be reused multiple times or not be used at all.

The participation of the "Server" entity in the "is_located" relationship is | total | .

The participation of the "Server" entity in the "runs_on" relationship is | partial | .

The participation of the "Virtual Machine" in the "owns" relationship is | partial | .

| total | | partial |

Question 5

Complete

Mark 2.00 out of 2.00

Fill in the gaps in the sentences. Words can be reused multiple times or not be used at all.

The | Virtual Machine | entity is a weak entity, while the | Server | entity is its owner entity.

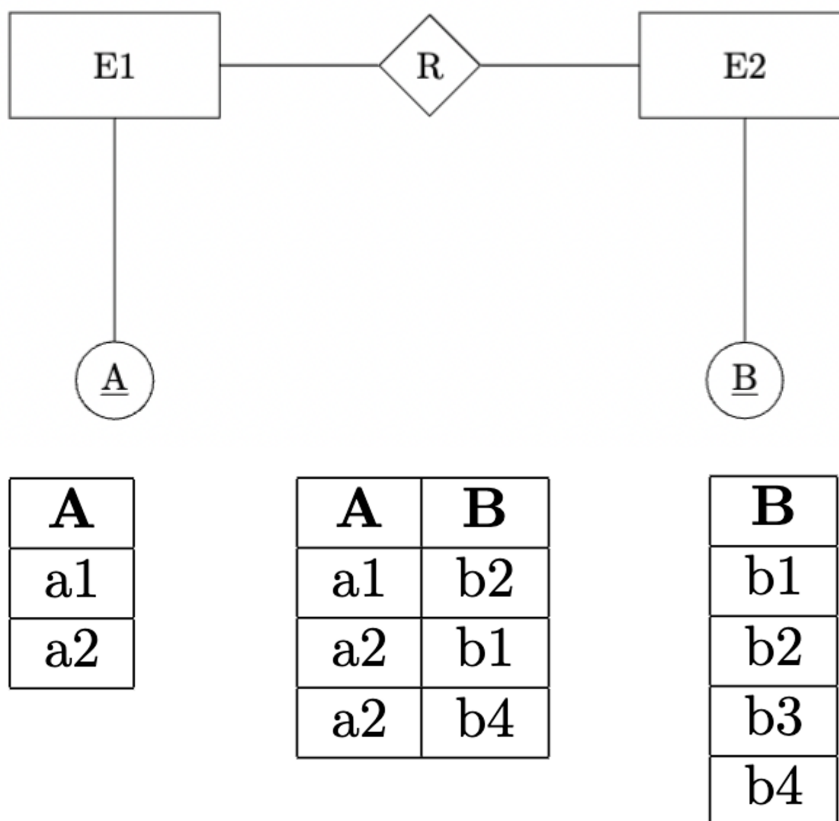| Customer | | Virtual Machine | | Server | | Slot |

Consider the following table. Assume that the values given in the figure are finite (the table will not be updated and no values will be added/removed).

A	B	C	D
a1	b1	c1	d1
a1	b2	c2	d1
a2	b3	c1	d2
a1	b4	c2	d3

Which of the following key(s) or set of key(s) could serve as primary keys? You may select multiple options.

- ☐ a. A
- ☑ b. B
- ☐ c. C
- ☐ d. D
- ☐ e. A, D
- ☐ f. A, C
- ☑ g. C, D

Consider the ER model of the first figure with two entities, E1 and E2, and the relationship R between them. We have translated the ER model to the relations shown in the second figure.

NOTE: there are *no* bold lines or arrows in the ER model to indicate key constraints or relationship cardinalities. ***This is because the ER diagram is intentionally incomplete, not because there are no key constraints.***

Select all the possible ways in which entity E1 could be connected to relationship R in the E-R model so that there are no constraint violations with the data of the example:
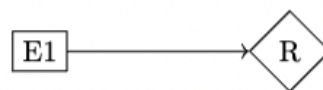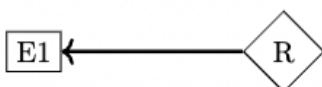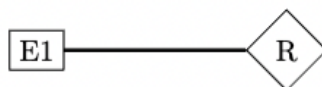
a.

b. ✓

c.

d.

e.
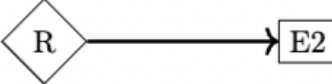

f. ✓

Select all the possible ways in which entity E2 could be connected to relationship R in the E-R model so that there are no constraint violations with the data of the example:

a.

R ————▶ E2
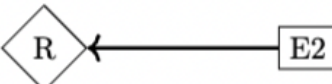
b.

R ———— E2

☑ c.

R ◀———— E2

d.

R ———▶ E2

e.

R ◀———— E2

☑ f.

R ———— E2

Information

Consider a brand that owns several hotels. You are given the task of designing a database for this brand. You may want to write down the ER model to solve this question.

- Every hotel has a unique hotel-id, a name, an address, and several rooms.
- Every room has a room-id that is unique within the hotel. There might be rooms that have the same room-id if they are in different hotels. For every room, we also want to store the maximum number of guests it can accommodate.
- For every customer, we keep track of their unique customer-id, their name, and their credit card.
- Every time a customer books a room, we want to store the corresponding room and customer, the day of the check-in and the day of the check-out. Hotels allow customers to book rooms specifying only the check-in date but not the check-out date.

Question 9

Complete

Mark 1.50 out of 1.50

Please select the attributes for the table:

HOTEL(...)

- ☐ a. check-in-date
- ☐ b. credit-card
- ☑ c. hotel_id
- ☐ d. customer-id
- ☑ e. address
- ☐ f. max-guests
- ☑ g. name
- ☐ h. check-out-date
- ☐ i. room-id

Question 10

Complete

Mark 1.50 out of 1.50

Please select the attributes for the table:

CUSTOMER(...)

- ☐ a. max-guests
- ☐ b. check-in-date
- ☑ c. credit-card
- ☐ d. hotel_id
- ☐ e. check-out-date
- ☑ f. customer_id
- ☑ g. name
- ☐ h. room-id
- ☐ i. address

Please select the attributes for the table:

ROOM(...)

- ☐ a. check-in-date
- ☑ b. max-guests
- ☐ c. credit-card
- ☐ d. check-out-date
- ☐ e. customer-id
- ☐ f. address
- ☑ g. hotel-id
- ☐ h. name
- ☑ i. room-id

Please select the attributes for the table:

BOOKING(...)

- ☐ a. max-guests
- ☐ b. credit-card
- ☑ c. hotel-id
- ☑ d. customer-id
- ☑ e. room-id
- ☑ f. check-in-date
- ☑ g. check-out-date
- ☐ h. name
- ☐ i. address

Please select the primary key(s) for the table:

BOOKING(...)

- ☑ a. hotel-id
- ☐ b. credit-card
- ☐ c. customer-id
- ☑ d. room-id
- ☐ e. name
- ☐ f. address
- ☑ g. check-in-date
- ☐ h. check-out-date
- ☐ i. max-guests

In the database of the brand of hotels, we want to additionally store for every hotel its set of owners (if any). An individual can be an owner of more than one hotel. How many extra tables do we need in order to store this information?

- ○ a. 0
- ○ b. 1
- ◉ c. 2
- ○ d. 3

Information

MCQ: Relational Algebra

Consider two **non-empty** relations $R\langle A, B, C\rangle$ and $S\langle D, E\rangle$ where all the attributes are of the same type. For the following pairs of relational algebra expressions, choose the best option that compares their results under the usual **set semantics**.(Hint: Evaluate both expressions for simple cases where R and S contain one or two tuples first to eliminate some options.)

Expression 1: $\sigma_{(A=a) \wedge (B=b)} (R) \times \sigma_{(D=d)} (S)$

Expression 2: $\sigma_{(A=a) \wedge (D=d)} (R \times S)$

- a. The two expressions produce the same result in all cases.

- ◉ b. The two expressions produce the same result only for specifc instances of R and S.

- c. The two expressions produce the same result some times and different results some other times when evaluated multiple times on the same instances of R and S.

- d. The two expressions always produce different results.

Expression 1: $\sigma_{(A=a)}(R) \times \sigma_{(D=d)} (S)$

Expression 2: $\sigma_{(A=a) \vee (D=d)} (R \times S)$

- a. The two expressions produce the same result in all cases.

- ◉ b. The two expressions produce the same result only for specifc instances of R and S.

- c. The two expressions produce the same result some times and different results some other times when evaluated multiple times on the same instances of R and S.

- d. The two expressions always produce different results.

Expression 1: $\pi_A (R) \times \pi_D (S)$

Expression 2: $\pi_{(A,D)} (R \times S)$

- ◉ a. The two expressions produce the same result in all cases.

- b. The two expressions produce the same result only for specifc instances of R and S.

- c. The two expressions produce the same result some times and different results some other times when evaluated multiple times on the same instances of R and S.

- d. The two expressions always produce different results.

Expression 1: $\pi_A(R) \times \rho_{(E \to A)}\pi_E(S)$
Expression 2: $\pi_A(R \times S)$

- a. The two expressions produce the same result in all cases.
- b. The two expressions produce the same result only for specifc instances of R and S.
- c. The two expressions produce the same result some times and different results some other times when evaluated multiple times on the same instances of R and S.
- ◉ d. The two expressions always produce different results.

Expression 1: $\pi_A(R) \times R$
Expression 2: $\pi_A(R \times R)$

- a. The two expressions produce the same result in all cases.
- b. The two expressions produce the same result only for specifc instances of R and S.
- c. The two expressions produce the same result some times and different results some other times when evaluated multiple times on the same instances of R and S.
- ◉ d. The two expressions always produce different results.

Information

Consider a database with the following schema:

Person (name, age, gender)
Visits (name, restaurant)
Eats (name, burger)
Serves (restaurant, burger, price)

For questions 20+21, drag and drop some of the given options to the boxes to write the queries in relational algebra.

For questions 22-25, select if the statement is true or false.

Note: The given schema is relevant only for questions 20+21

Find all restaurants visited by at least one person under the age of 18.:

$\pi_{restaurant}$ (($\sigma_{age<18}$ Persons) ⋈ Visits)

Eats | $\pi_{restaurant}$ | | $\pi_{age<18}$ | Visits | Serves | $\sigma_{age>18}$ | Persons | $\sigma_{age<18}$ | $\sigma_{restaurant}$

Find the names of all people who visit only restaurants serving at least one burger they eat:

π_{name} (Person) – π_{name} (Visits – $\pi_{name, restaurant}$ (Eats ⋈ Serves)

)

Eats | σ_{name} | $\pi_{name, restaurant}$ | $\sigma_{name, restaurant}$ | Serves | π_{name} | $\sigma_{restaurant}$

Person | $\pi_{restaurant}$ | π_{name} | σ_{name} | Visits

The following SQL statement is syntactically correct:

SELECT city, state, COUNT(*) as num_customers

FROM customers

GROUP BY city, state

WHERE num_customers < 10

ORDER BY city;

Select one:

○ True

● False

Question 23

Complete

Mark 1.25 out of 1.25

In SQL SELECT, the default is that duplicates are not eliminated

Select one:

◉ True

○ False

Question 24

Complete

Mark 1.25 out of 1.25

When using the ORDER BY clause, it always appears as the last clause in a SELECT statement.

Select one:

○ True

◉ False

Question 25

Complete

Mark 1.25 out of 1.25

A relation can have only one foreign key, and it must reference the primary key of the same table.

Select one:

○ True

◉ False

Question 26

Complete

Mark 1.00 out of 1.00

Choose the correct options after reading the following two statements.
(A) A primary index is often a clustered index.

◉ True ○ False

(B) It is always possible to have two clustered indices on two different columns in a table.

○ True ◉ False

Choose the correct options after reading the following two statements.
(A) Fixed-length records store field offsets at the beginning of the record.

○ True ◉ False

(B) When storing fields in length-data pairs, variable-length records allow direct access to fields.

○ True ◉ False

Choose the correct options after reading the following two statements.
(A) Inserting a record has a cheaper I/O cost on Sorted Files compared to Heap File in the most typical scenarios.

○ True ◉ False

(B) Equality search has a cheaper I/O cost on Sorted Files compared to Heap file in the most typical scenarios.

◉ True ○ False

Choose the correct options after reading the following two statements.
(A) A sparse index stores one data entry for every search-key value.

○ True ◉ False

(B) A sparse index takes up less disk space than a dense index.

◉ True ○ False

File Organization

Migroslin, a long-standing figure in the food review scene, has decided to digitize its 200 years of restaurant reviews across Switzerland in order to create its own mobile app. Their design of the reviews table is as follows, where the primary key of the table is the **reviewID**:

```
REVIEW(reviewID, restaurant, canton, reviewer, year, rating)
```

The size of each data record in the table is 200B. The size breakdown per attribute is as follows: **reviewID** is 8B, **restaurant** is 64B, **canton** is 4B, **reviewer** is 100B, **year** is 8B, and **rating** is 16B (total of 200B per record).

The disk page size is 8192B: each page can store only a whole/integer number of records, and every page contains a 44B header and 48B footer of metadata.

Migroslin keeps a strict tradition of only having exactly 500 reviews done per year (for each of the past 200 years, up to 2022 -- 2023 data is not yet available), from a pool where exactly 50 different food critics/reviewers are selected, from each exactly ten reviews are done. Critics can provide reviews for multiple years, but *no critic is allowed to make more than 40 reviews in total*.

Data is input year by year, so reviews in a given year are recorded together in a contiguous way -- but multiple employees record them in parallel, so they are not sorted on the year unless explicitly mentioned (for example, all 2016 reviews are next to each other, but they might come after 2018 reviews).

Unless explicitly noted, NSM is used as the page layout, there are no subpages, and nothing resides in the main memory - meaning that disk I/Os are necessary. I/O cost indicates the number of disk page accesses.

Answer the questions using only the final result numbers, without variables or formulas. If asked for an explanation, briefly explain your reasoning (variables, formulas, and short text explanation are OK in that case).

Question 30

Complete

Mark 1.00 out of 1.00

How many records fit on a single page? Enter the final result.

Answer: 40

Question 31

Complete

Mark 1.00 out of 1.00

What is the size of the **REVIEW** file, expressed by the number of disk pages? Enter the final result.

Answer: 2500

Question 32

Not answered

Not graded

Briefly explain your answer to the previous question. You can provide a formula or any text explanation.

Answer: [   ]

Information

[Updating Rating] Answer the following questions (marked with **[Updating Rating]**) assuming the following scenario.

The **REVIEW** file is implemented by Migroslin as a heap file. One day, an employee found out that the reviews made by Gordon Aligro were using the rating scale 0-5 instead of the usual standard 1-6. In order to fix the database, they need to run the following query

```
update REVIEW
set rating = rating + 1
where reviewer = 'Gordon Aligro'
```

We are interested in finding the <u>maximum I/O cost</u> of this query.

Question 33

Complete

Mark 1.50 out of 1.50

[Updating Rating] The I/O cost of searching for the qualifying records to update is (enter the final value only):

Answer: [2500]

Question 34

Complete

Mark 1.00 out of 1.00

[Updating Rating] The maximum number of records that needs to be updated is:

Answer: [40]

Question 35

Complete

Mark 1.00 out of 1.00

[Updating Rating] The maximum I/O cost of the update query is:

Answer: [2540]

Question 36

Complete

Mark 3.00 out of 3.00

[Updating Rating] If an index on the column reviewer existed, which index would be the most suitable for performing such updates frequently? You are not allowed to modify the heap file.

Data Structure | Hash

Format | Unclustered

Density | Dense

Question 37

Not answered

Not graded

Briefly explain your answer to the previous question. You can provide a formula or any text explanation.

Answer:

Information

[Finding the Stars] Answer the following questions (marked with **[Finding the Stars]**) assuming the following scenario.

The **REVIEW** file stored by Migroslin is now sorted by the **year** column. As part of the mobile app, we want to find the best performing restaurants (the Migroslin stars) for each year. To be a Star, a restaurant should have a rating of at least 5. For example, to find the stars of 2022, we can use the following query:

```
select unique restaurant
from REVIEW
where year = 2022 and rating >= 5
```

Use binary search for finding the first page that satisfies the condition, and write the final result values to the following questions.

Remember that $\log_2(1024) = 10$, i.e. $2^{10}=1024$, $\log_2(2048) = 11$, and so on; and that we can read only full pages (so if we need to read 10.2 pages, we will read 11).

Consider only the best/minimal case for packing together the records with the year 2022 into pages. For example, if the records for one year take up 2.5 pages, assume that this takes 3 I/Os (while technically, the records for a year may start anywhere on the disk page and require more than 3 I/Os to read 2.5 pages of this example. If needed for reads, use the simplification that the years start from the beginning of a page).

Question 38

Complete

Mark 1.50 out of 1.50

[Finding the Stars] What is the I/O cost of a binary search, to find the first page with the year 2022? (answer with the final value only, not with a formula)

Answer: 12

Question 39

Not answered

Not graded

Briefly explain your answer to the previous question. You can provide a formula or any text explanation.

Answer: [   ]

Question 40

Complete

Mark 1.00 out of 1.00

[Finding the Stars] How many pages will reviews for the year 2022 take? Consider the best/minimal case for packing together the records with the year 2022 into pages. We can read only full pages. Answer with the final page number value.

Answer: [13]

Question 41

Not answered

Not graded

Briefly explain your answer to the previous question. You can provide a formula or any text explanation.

Answer: [   ]

Question 42

Complete

Mark 1.00 out of 1.00

[Finding the Stars] Overall, the total cost of the query is: (answer with the final value)

Answer: [25]

Question 43

Complete

Mark 3.00 out of 3.00

[Finding the Stars] You are asked to move the sorted structure back to a heap file, but replace it with an index. You want the app responsive to this query (with varying year inputs) thousands of years into the future! What is the best possible index, given the following choices?

Index column [Keep on `year`]

Format [Clustered]

Density [Sparse]

Question 44

Not answered

Not graded

Briefly explain your answer to the previous question. You can provide a formula or any text explanation.

Answer: [   ]

Information

[Vaudoise Ratings] Answer the following questions (marked with **[Vaudoise Ratings]**) assuming the following scenario.

The **REVIEW** file is stored on disk, while Migroslin implements an *unclustered* B+ tree index on the **canton** column. The intent was to let users find out how well the restaurants in their cantons are doing. Consider the following query, which returns the average rating of all reviews of restaurants in Vaud -- whose restaurants are the target of 20% of all reviews in the database:

```
select avg(rating)
from REVIEW
where canton = 'VD'
```

Assume that accessing the index itself does not incur any I/O cost (it is memory resident) - the only I/O cost comes from accessing the file pages.

Question 45

Complete

Mark 3.00 out of 3.00

[Vaudoise Ratings] What is the <u>worst-case</u> I/O cost of the query? Enter the final value only.

Answer: [20000]

Question 46

Complete

Mark 3.00 out of 3.00

[Vaudoise Ratings] Mobile app users really want to find Migroslin stars (the restaurants with a rating of at least 5) in their respective canton. You want to modify the current canton index so that the "star in canton" query is as efficient as possible. What is the best index among the possible options below? A compound index is represented as `columnA > columnB`.

Index column   [Move to `canton` > `rating`]

Data Structure [Keep B-Tree]

Density   [Dense]

For example, the "star in Vaud" query can be given as

```
select restaurant
from REVIEW
where rating >= 5 and canton = 'VD'
```

Question 47

Not answered

Not graded

Briefly explain your answer to the previous question. You can provide a formula or any text explanation.

Answer:

Question 48

Complete

Mark 3.00 out of 3.00

Choose the correct options after reading the statements.
(A) Column store (DSM) layout is very beneficial when **all** columns of a relation have to be retrieved.

○ True   ● False

(B) PAX layout supports only fixed-length records.

○ True   ● False

Information

Page Layouts

Alphabet Books is a big publishing company that keeps the information about their books in the following table, where the primary key is BookID:

Book(BookID, author, title, genre, year, price)

The size of each data record in the table is 200B. The size breakdown per attribute is as follows: BookID is 8B, author is 64B, title is 100B, genre is 4B, year is 8B, and price is 16B (for a total of 200B per record).

The records are organized into 4096B pages, each page can store only a whole/integer number of records, and every page contains a 48B header and 48B footer of metadata. The data is sorted on the primary key.

The company was founded in 1974, and since then, they have been publishing 400 books per year. Book information about the year's publishings is only recorded at the end of each year, and 2013 is the last year recorded in the table (40 years in total).

Answer the questions using only the final result numbers, without variables, units, or formulas. If asked for an explanation, briefly write your procedure (variables, formulas, and short text explanation are OK).

Question 49

Complete

Mark 4.00 out of 4.00

If using NSM (N-ary Storage Model), how many tuples can each page hold? **Answer:** [20] (2 points)

If using NSM, how many pages in total does the file need (data over 40 years)? **Answer:** [800] (2 points)

Question 50

Not answered

Not graded

(Optional) Briefly explain your answer to the question above. You can use free text or formulas to explain.

Answer: [   ]

Question 51

Complete

Mark 6.00 out of 6.00

If using DSM, for every sub-relation (attribute), calculate how many values each page holds and the total number of pages required for the sub-relation. (0.5 points for each answer)

(Reminder, the page size is 4096 bytes, with 48 bytes of header and 48 bytes of footer metadata per page)

Sub-Relation	Number of values per page	Number of pages
BookID	500	32
author	62	259
title	40	400
genre	1000	16
year	500	32
price	250	64

Question 52

Not answered

Not graded

(Optional) Briefly explain your answer to the question above. You can use free text or formulas to explain.

Answer: [   ]

Question 53

Complete

Mark 7.00 out of 7.00

If using PAX, calculate each mini page's **size in bytes**. (1 point for each answer)

Is there an equal number of tuples inside mini-pages of a given page? [Yes]

Enter the sizes in the table below (without the unit, just the number, the unit is considered to be *bytes*).

Header	48
BookID	160
author	1280
title	2000
genre	80
year	160
price	320
Footer	48

Question 54

Not answered

Not graded

(Optional) Briefly explain your answer to the question above. You can use free text or formulas to explain.

Answer: [   ]