Data management and visualization

Started on	Tuesday, 9 February 2021, 12:57 PM
State	Finished
Completed on	Tuesday, 9 February 2021, 12:57 PM
Time taken	9 secs
Grade	**0.00** out of 31.00 (**0**%)

Question 1

Not answered

Marked out of 1.50

In Datawarehouse analysis, the aggregation window defined at the physical level in extended SQL:

- (a) can be specified only on a single sort key
- (b) is based on a physical structure (e.g. an index)
- (c) is specified with the range clause
- (d) is defined by counting the rows
- (e) is appropriate for sequence data with gaps and sparse data

Risposta errata.

The correct answer is: is defined by counting the rows

Question 2

Not answered

Marked out of 1.00

In NoSQL design, the extended reference pattern has the advantage of:

- (a) reducing the overall number of documents in a collection
- (b) reducing data denormalization
- (c) reducing the CPU workload for frequent computations
- (d) reducing the join operations
- (e) reducing the reference to document extensions
- (f) reducing future technical debt
- (g) reducing document complexity

Risposta errata.

The correct answer is: reducing the join operations

Question 3

Not answered

Marked out of 1.00

Select the right configuration of a MongoDB replica set:

- (a) 2 secondary nodes, 1 arbiter node
- (b) 1 secondary node, 1 arbiter node
- (c) 1 primary node, 2 secondary nodes , 2 arbiter nodes
- (d) 2 primary nodes, 2 secondary nodes, 1 arbiter node
- (e) 2 primary nodes, 1 secondary node
- (f) 2 primary nodes, 2 secondary nodes, 2 arbiter nodes

Risposta errata.

The correct answer is: 1 primary node, 2 secondary nodes , 2 arbiter nodes

Question 4

Not answered

Marked out of 1.50

Which one of the following examples is **NOT** related to a Gestalt principle?

- (a) the bars representing smaller values are shorter
- (b) the points of a data series are connected
- (c) the color of the legend is similar to the color of the elements of the graph
- (d) the direct labeling technique improves the readability of the visualization
- (e) the points of a group are enclosed by a fine line

Risposta errata.

The correct answer is: the bars representing smaller values are shorter

Question 5

Not answered

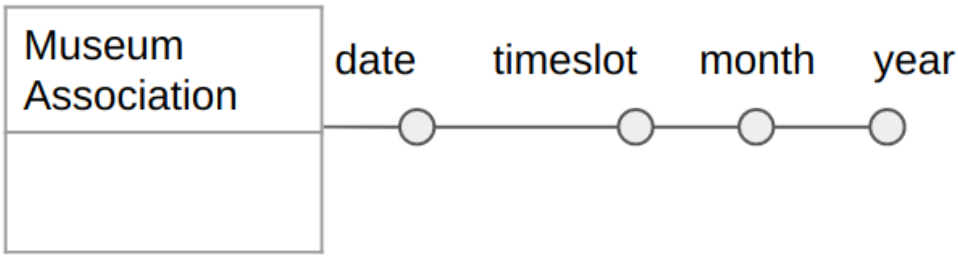
Marked out of 0.50

Data analysts of the National Association of Italian Museums are interested in analyzing the average revenue per ticket.
In particular, they would like the analyses to address the following features.

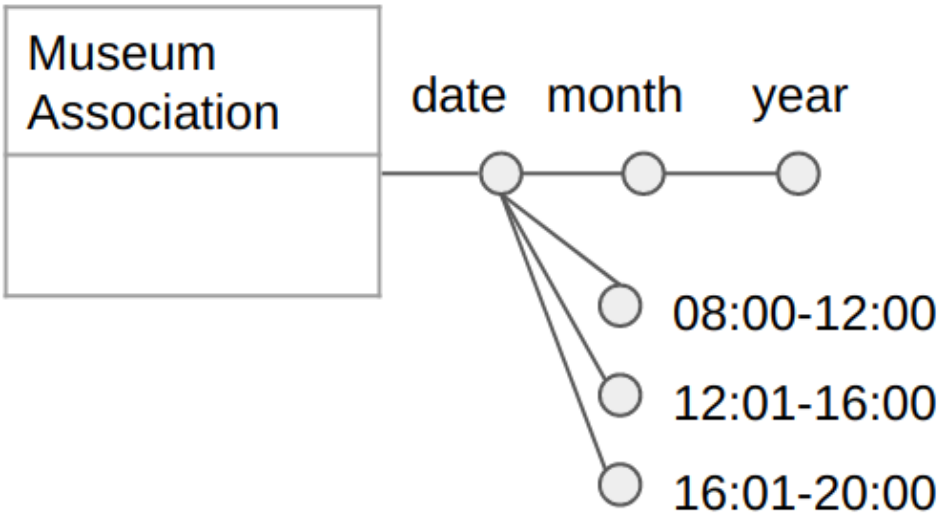
- Museums are analyzed according to their city and region. A museum has a unique name, and it is located in a specific city. The same city can host different museums.
- A museum may have some additional services available for its public. The systems records which services are available for each museum.
  Examples of additional services are “guided tours”, “audioguides”, “wardrobe”, “cafè”, “Wi-Fi”. The number of possible additional services is large and growing, hence the full list is not known a priori.
- The tickets sold by each museum are recorded. There are 4 different types of tickets: “Full price”, “Reduced-student” (for students from 18 to 24 years old), “Reduced-junior” (for young people less than 18 years old), and “Reduced-senior” (for people over 70 years old).
- The analyses must be carried out considering the date, month and year, and the time slot of the ticket emission. The time slot is stored in 3 ranges of 4-hour blocks (08:00-12:00, 12:01-16:00, 16:01-20:00).

Choose the correct conceptual schema from the proposed ones to properly define the time dimension according to the given specifications (at most one answer is correct).

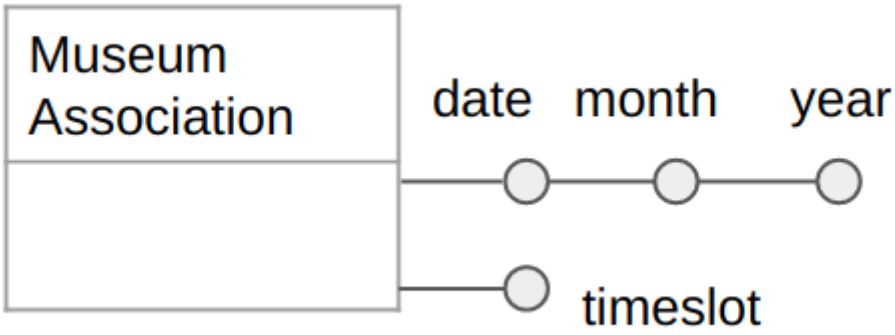
(a)



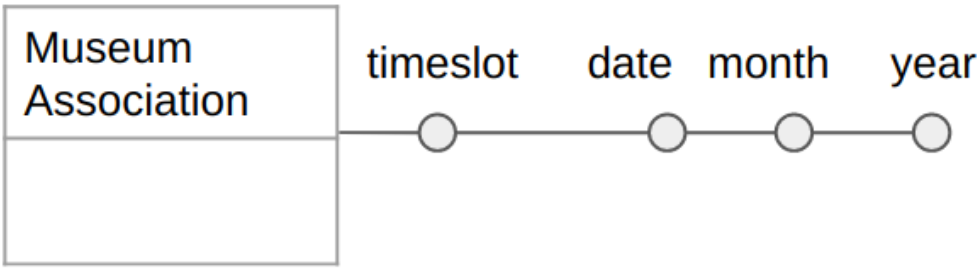
(b)



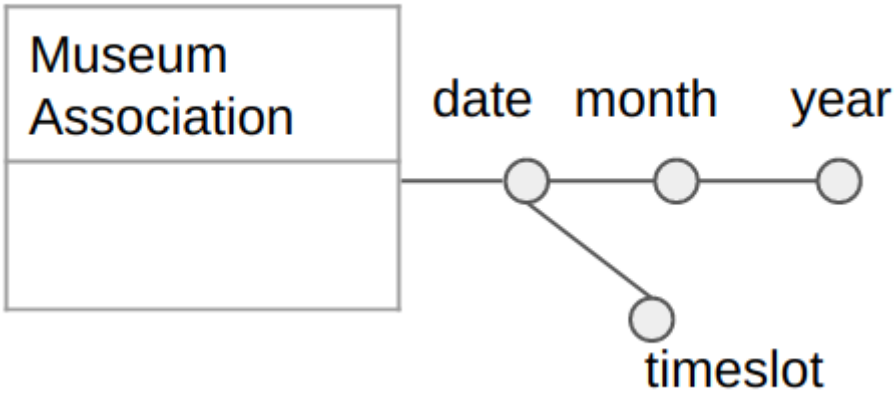
(c)



(d)

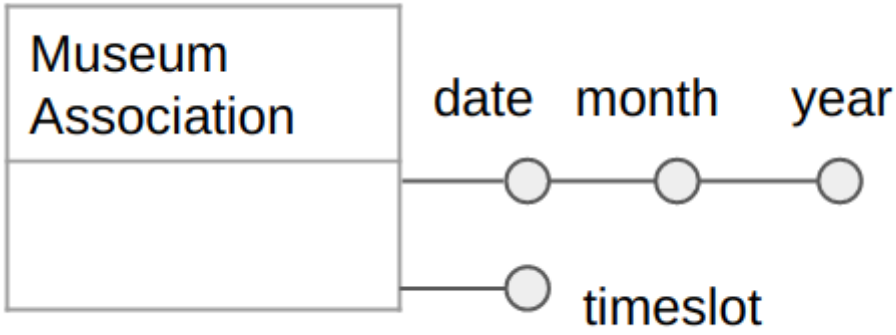


(e)



Risposta errata.

The correct answer is:



Question 6

Not answered

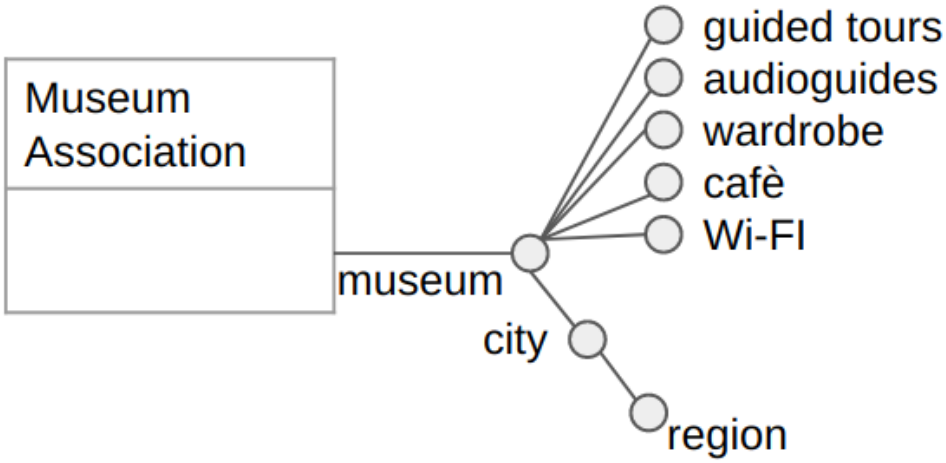
Marked out of 0.50

Data analysts of the National Association of Italian Museums are interested in analyzing the average revenue per ticket.
In particular, they would like the analyses to address the following features.

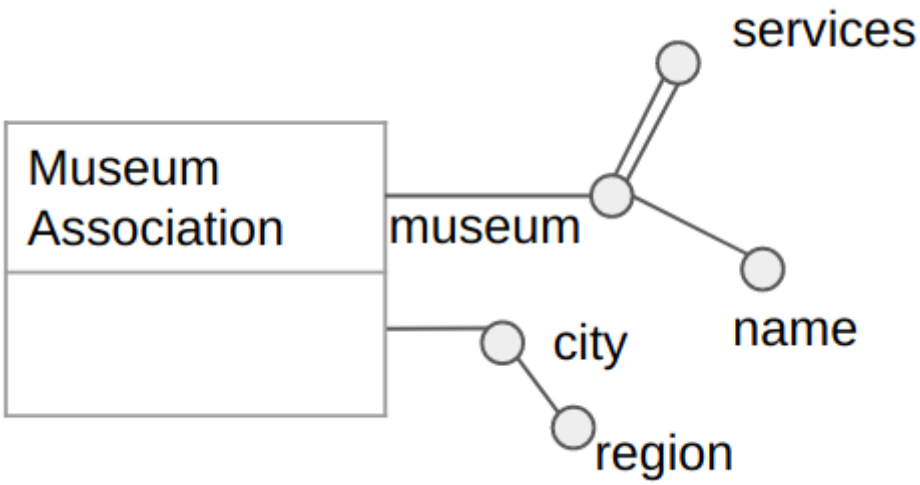
- Museums are analyzed according to their city and region. A museum has a unique name, and it is located in a specific city. The same city can host different museums.
- A museum may have some additional services available for its public. The systems records which services are available for each museum.
  Examples of additional services are “guided tours”, “audioguides”, “wardrobe”, “cafè”, “Wi-Fi”. The number of possible additional services is large and growing, hence the full list is not known a priori.
- The tickets sold by each museum are recorded. There are 4 different types of tickets: “Full price”, “Reduced-student” (for students from 18 to 24 years old), “Reduced-junior” (for young people less than 18 years old), and “Reduced-senior” (for people over 70 years old).
- The analyses must be carried out considering the date, month and year, and the time slot of the ticket emission. The time slot is stored in 3 ranges of 4-hour blocks (08:00-12:00, 12:01-16:00, 16:01-20:00).

Choose the correct conceptual schema from the proposed ones to properly define the characteristics of museum analytics according to the given specifications (at most one answer is correct).

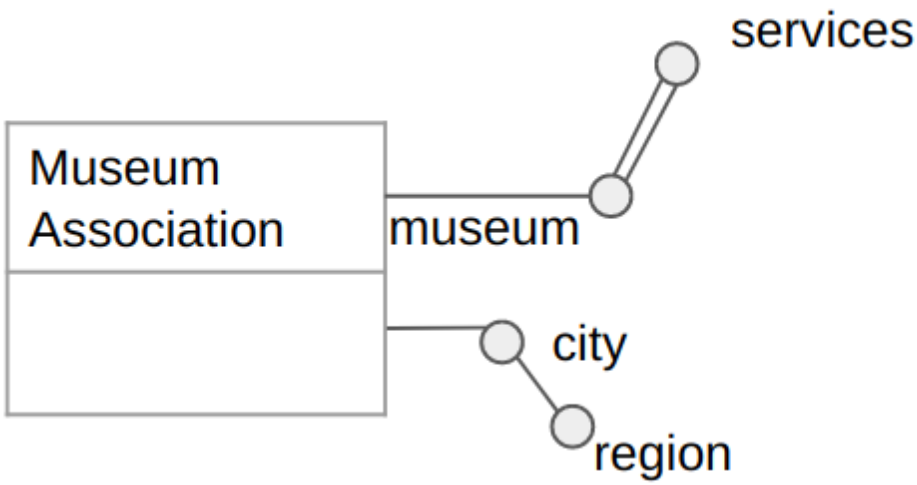
(a)



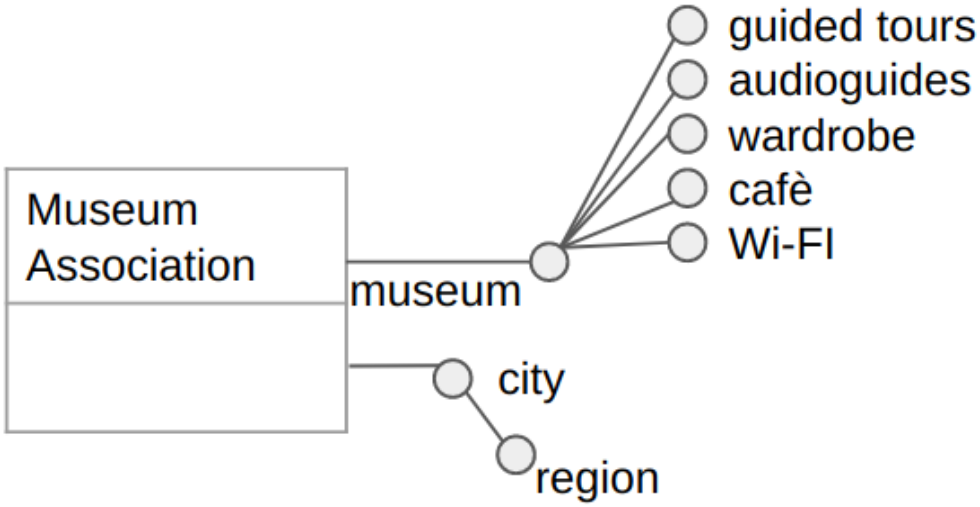
(b)



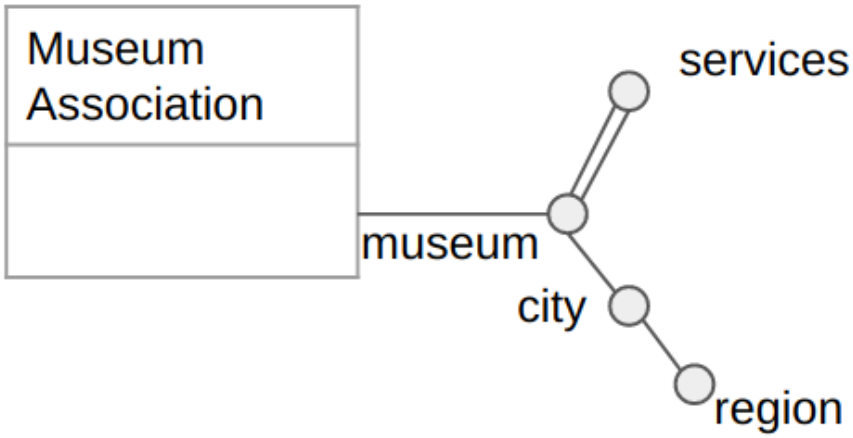
(c)



(d)

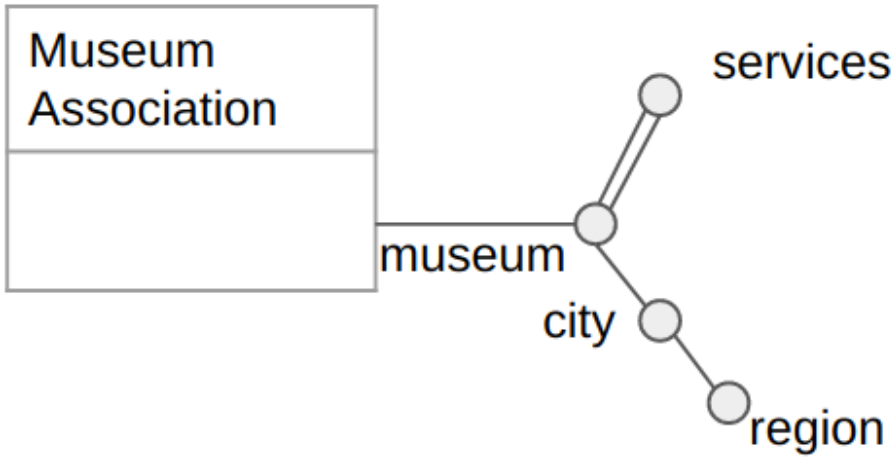


(e)



Risposta errata.

The correct answer is:



Question 7

Not answered

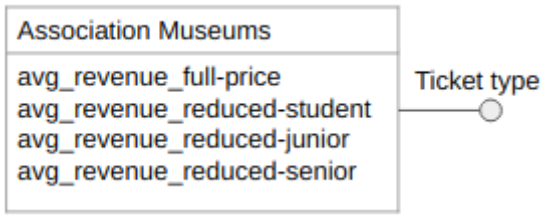
Marked out of 1.00

Data analysts of the National Association of Italian Museums are interested in analyzing the average revenue per ticket.
In particular, they would like the analyses to address the following features.

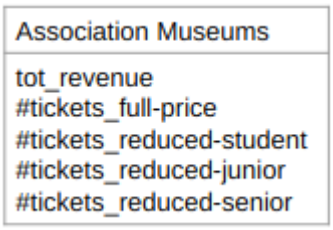
- Museums are analyzed according to their city and region. A museum has a unique name, and it is located in a specific city. The same city can host different museums.
- A museum may have some additional services available for its public. The systems records which services are available for each museum.
  Examples of additional services are “guided tours”, “audioguides”, “wardrobe”, “cafè”, “Wi-Fi”. The number of possible additional services is large and growing, hence the full list is not known a priori.
- The tickets sold by each museum are recorded. There are 4 different types of tickets: “Full price”, “Reduced-student” (for students from 18 to 24 years old), “Reduced-junior” (for young people less than 18 years old), and “Reduced-senior” (for people over 70 years old).
- The analyses must be carried out considering the date, month and year, and the time slot of the ticket emission. The time slot is stored in 3 ranges of 4-hour blocks (08:00-12:00, 12:01-16:00, 16:01-20:00).

Choose the best solution for the ticket information and measures in the conceptual schema design among those proposed (at most one answer is correct).

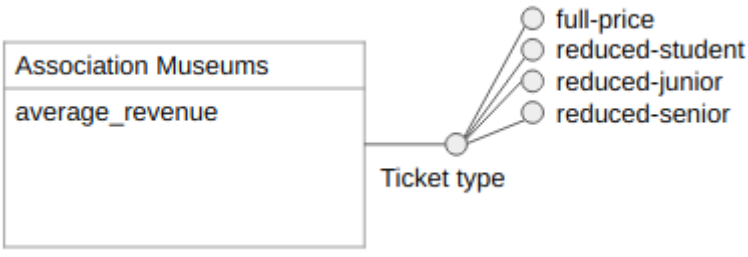
(a)



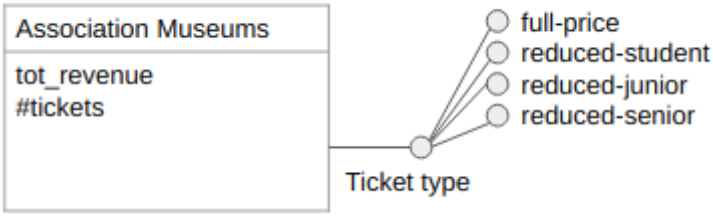
(b)



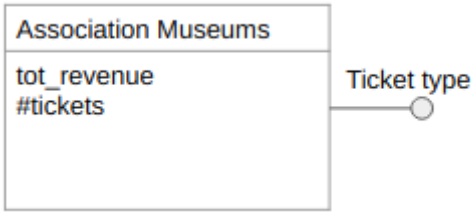
(c)



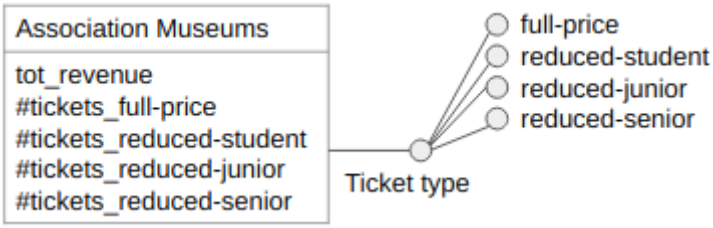
(d)



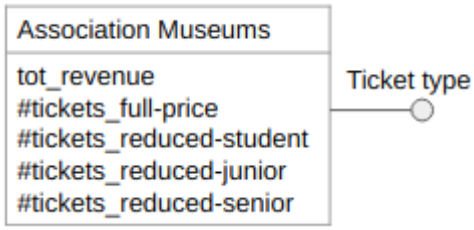
(e)



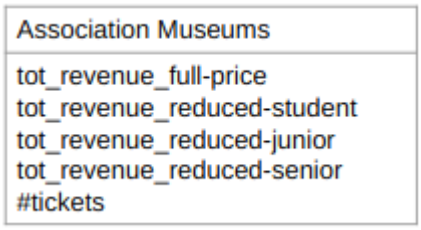
(f)



(g)



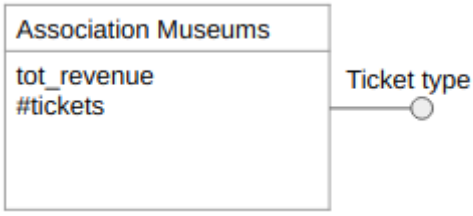
(h)



(i)

Risposta errata.

The correct answer is:

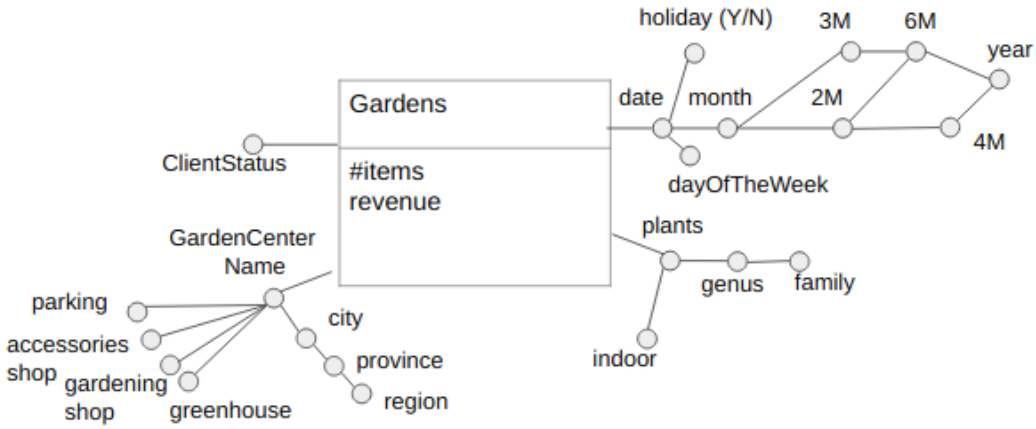


Question 8

Not answered

Marked out of 2.00

Given the following conceptual schema:



- Each garden center has a unique name. A garden center can have 0 or more services. There are 4 available services: “parking”, “accessories shop”, “gardening shop” and “greenhouse”.
- The cardinality of “ClientStatus” is 3, and it can be “1” for “Silver”, “2” for “Gold” and “3” for “Platinum”.
- A plant can be either an indoor or an outdoor plant. The genus and family of the plant are stored.

Write the logical design of the conceptual DW schema indicated in the picture.
Write each table on a new line.
Use the **bold** or the underline for identifying primary-key attributes.

Gardens(TimeId, GardenCenterId, PlantId, ClientStatus, #items, revenue)
Time(TimeId, date, month, 2M, 3M, 4M, 6M, year, dayOfTheWeek, holiday)
GardenCenter(GardenCenterId, GardenCenter, city, province, region, greenhouse, accessoriesShop, gardenShop, parking)
Plant(PlantId, plant, genus, family, indoor)

Question 9

Not answered

Marked out of 4.00

```
MusicStreaming(TimeId, SongId, PlatformId,
NumberOfStreamings, NumberOfLikes)
Time(TimeId, date, month, 2M, 3M, 6M, year, dayOfTheWeek)
Song(SongId, Song, album, classic, indie, pop, .., rock)
UserLocation(UserLocationId, province, region, country)
```

For each song and month, compute the following metrics:

- the total number of streamings
- the cumulative total number of streamings since the beginning of the year
- assign a rank to each song, separately for each album, based on the monthly number of streamings (rank 1st the most streamed song of the album for each month)

Write the requested SQL query.

```
SELECT month, song,
    SUM(NumberOfStreamings),
      SUM(SUM(NumberOfStreamings)) OVER (PARTITION BY songId, year
ORDER BY month ROWS UNBOUNDED PRECEDING)
      RANK() OVER (PARTITION BY album, month ORDER BY
SUM(NumberOfStreamings) DESC),

FROM Song S, Time T, MusicStreaming MS

WHERE S.SongId=MS.SongId AND T.Timeid=MS.Timeid

GROUP BY song, songId, month, year, album
```

Question 10

Not answered

Marked out of 4.00

```
MusicStreaming(TimeId, SongId, PlatformId,
NumberOfStreamings, NumberOfLikes)
Time(TimeId, date, month, 2M, 3M, 6M, year, dayOfTheWeek)
Song(SongId, song, album, classic, indie, pop, .., rock)
UserLocation(UserLocationId, province, region, country)
```

Separately for each song and province of the user, compute the following metrics:

- the average number of monthly likes
- the percentage of the number of likes with respect to the total number of likes received by users in the same country
- the number of likes of the album in the user province

Write the requested SQL query.

```
SELECT province, song,
    SUM(NumberOfLikes)/ COUNT(DISTINCT month),
      100*SUM(NumberOfLikes)/SUM(SUM(NumberOfLikes)) OVER (PARTITION
BY songId, country),
      SUM(SUM(NumberOfLikes)) OVER (PARTITION BY album, province)

FROM Song S, MusicStreaming MS, UserLocation Lm, Time T

WHERE S.SongId=MS.SongId AND L.UserLocationId=MS.PlatformId AND
T.TimeId=MS.TimeId

GROUP BY song, songId, province, country, album
```

Question 11

Not answered

Marked out of 2.00

Given the following document structure:

```
{
"address":
    {"building":"768",
    "coord":[-73.9685872,40.7679509],
    "street":"Madison Avenue",
    "zipcode":"10065",
    "borough":"Manhattan",
    "city": "New York"},
"sold_items": ["Smartphones", "PC", "TV"],
"reviews":[
    {"date": {date:{"2019-11-05"}}, "score":10, "description": "Lorem ipsum"},
    {"date": {date("2020-02-21"}}, "score":8, "description": "Lorem ipsum"}
    ],
"name":"Elettronic-store"
}
```

Select all the shops located in Rome that sell smartphones or TV and received at least one review with a score greater than 8. Show only the name, the street and the building.

```
db.shops.find({sold_items:{$in: ["Electronics", "Home"]},
            "address.city":"Rome",
            "reviews.score":{$gt:8} },
            {_id:0, name: 1, "address.street":1,"address.building":1 })
```

Question 12

Not answered

Marked out of 3.00

Given the following document structure:

```
{
"name":"Electrostore",
"address":
    {"building":"A1",
     "street":"via Torino",
     "zipcode":"12345",
     "borough":"Campidoglio",
     "city": "Rome"},
"sold_items": ["Smartphone", "PC", "TV"],
"reviews":[
    {"date": "2019-11-05", "score":10, "description": "Lorem ipsum"},
    {"date": "2020-02-21", "score":7, "description": "Lorem ipsum"}
    ]
}
```

For each city, compute the average and the maximum review score.

Show only the first 10 cities with the highest number of reviews.

```
db.collection.aggregate([
{$unwind: "$reviews"},
{$group: {
        _id: "$address.city",
        'countReviews':   {$sum: 1},
        'maxReviewScore': {$max: '$reviews.score'},
        'avgReviewScore': {$avg: '$reviews.score'}
}},
{$sort:
        {countReviews: -1}
},
{$limit: 10}
])
```

Question 13

Not answered

Marked out of 4.00

Design a MongoDB database to manage a warehouse for parcel delivery according to the following requirements.

Customers of the parcel delivery service are citizens identified by their social security number. They can be senders or recipients of delivered parcels. They are characterized by their name, surname, email address, a telephone number, and by different addresses, one for each type, e.g., one billing address, one home address, one work address, etc. Each address consists of street name, street number, postal code, city, province, and country.

Parcels are characterized by a unique barcode and their physical dimensions (specifically: width, height, depth, and weight). All widths, heights, and depths are always expressed in meters. All weights are always expressed in kilograms.

The recipient and the sender information required to deliver each parcel must be always available when accessing the data of a parcel. Recipient and sender information required to deliver a parcel consists of: full name, street name, street number, postal code, city, province, and country. For instance, a recipient information can be: Mario Rossi, corso Duca degli Abruzzi, 24, 10129, Torino, Torino, Italy.

The parcel warehouse is divided into different areas. Each area is identified by a unique code, e.g., 'area_51' and consists of different lines. Each line is identified by unique code, e.g., 'line_12', and hosts several racks. Each rack is identified by unique code, e.g., 'rack_33', and is made up of shelves. Each parcel is placed on a specific shelf of the warehouse, identified by a unique code, e.g., 'shelf_99'. The database is required to track the location of each parcel within the warehouse.

Given a parcel, the database must be designed to efficiently provide its full location, from the shelf, up to the area, through the rack and line.

Given a customer, the database must be designed to efficiently provide all her/his parcels as a sender, and all his/her parcels as a recipient.

Write a sample document for each collection of the database. Explicitly indicate the design patterns used.

Parcel

```
{
    _id: <string>, // barcode
    dimensions: { // also 1st-level attribute is fine
        width: <number>,
        height: <number>,
        depth: <number>,
        weight: <number>
    }
    recipient: {
        _id: <string>, // SSN, _id of the customer
        street: <string>,
        civic_number: <string>,
        zip_code: <string>,
        city: <string>,
        province: <string>
    },
    sender:{
        _id: <string>,// SSN, _id of the customer
        street: <string>,
        civic_number: <string>,
        zip_code: <string>,
        city: <string>,
        province: <string>
    },
    father_pos: <string>, // code of area/lane/rack/shelf
    locations: [
        <string> // code of area/lane/rack/shelf
    ]
}
```

Tree pattern for the position. The full list of tree-pattern ancestors is required. The parent ancestor of the tree-pattern is optional.

No collection for the areas, since no data are tracked except their code.

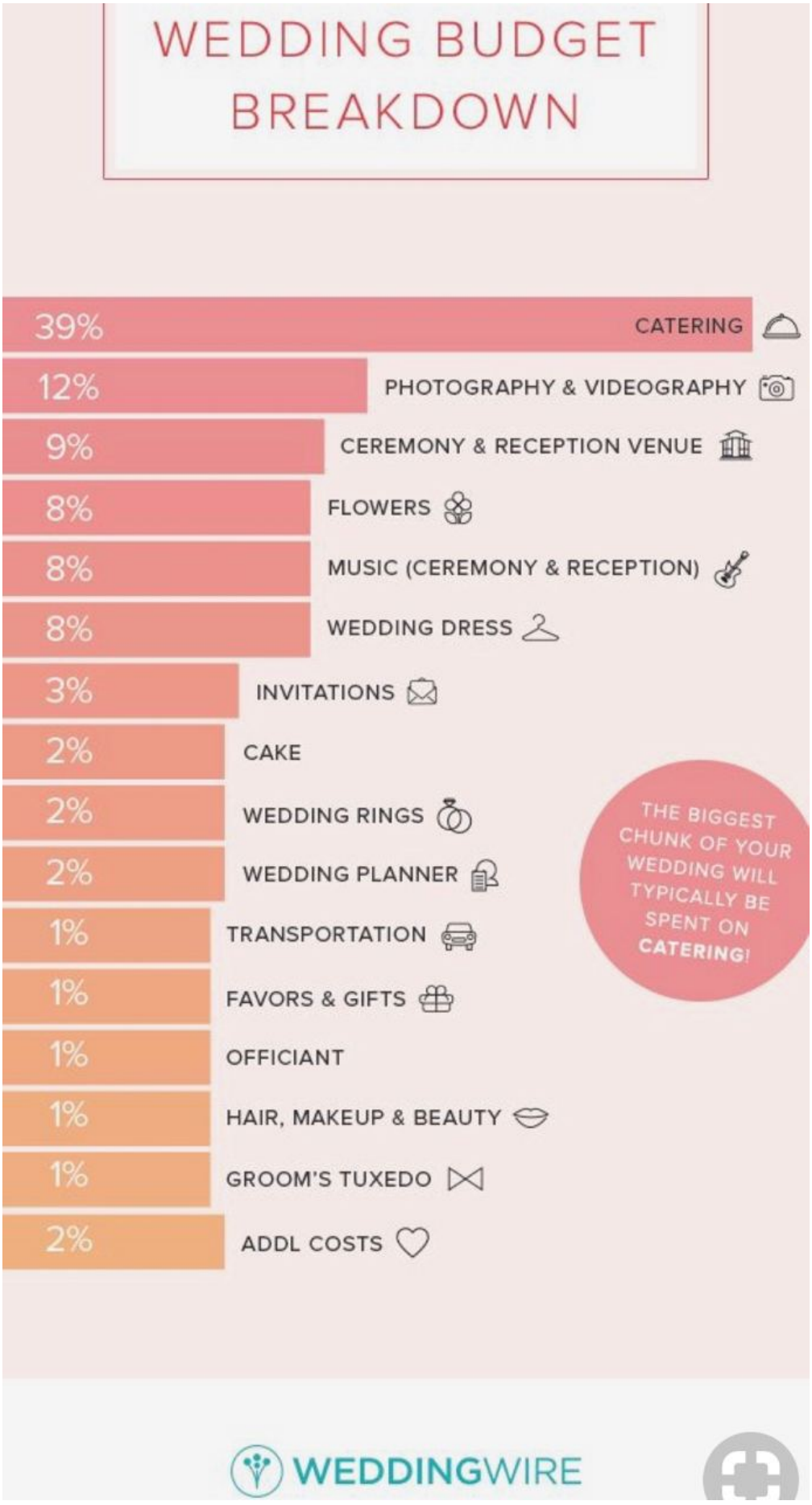
Extended reference pattern for recipient and sender address information. The recipient and sender _ids are required to look up all parcels of a given customer.

Customers

```
{
    _id: <string>, // fiscal code
    name: <string>,
    surname: <string>,
    email: <string>,
    tel: <string>,
    addresses:
        {
            home:
                {
                 street_name: <string>,
                 street_num: <string>,
                 postal_code: <int>,
                 city: <string>,
                 province: <string>,
                 country: <string>
                },
            'billing':
                {
                 street_name: <string>,
                 street_num: <string>,
                 postal_code: <int>,
                 city: <string>,
                 province: <string>,
                 country: <string>
                },
            work: {...}
        },
}
```

Attribute pattern (optional) for the addresses attribute.

Information



Analyze the above graph reporting the average breakdown of wedding costs. According to their website, WeddingWire is "the largest and most trusted global marketplace connecting engaged couples with local wedding professionals". WeddingWire published these data on a blog post dated December 2020: "We surveyed thousands of couples around the country in our WeddingWire Newlywed Report to share their wedding budgets with us".

Question 14

Not answered

Marked out of 0.25

Question

Is there a clearly defined question addressed by the visualization? Write it down.

Question 15

Not answered

Marked out of 1.25

Data

Is the data quality appropriate? Identify the inadequate characteristics and explain.

Question 16

Not answered

Marked out of 0.75

Visual Proportionality

Are the values encoded in a uniformly proportional way?

Question 17

Not answered

Marked out of 0.75

Visual Utility

All the elements in the graph convey useful information?

Question 18

Not answered

Marked out of 0.50

Visual Clarity

Are the data in the graph clearly identifiable and understandable (properly described)?

Question 19

Not answered

Marked out of 0.25

Design data

Design the visualization based on the following data structure (to be completed).

Question 20

Not answered

Marked out of 1.25

Design schema & Sketch

Fill in the required schema elements; formulas can be used if required. Then describe in words the design proposal.

Question 21

Not answered

Not graded

This is a blank question to be used as your personal notepad during the exam.

Anything written here will NOT be evaluated.