

Exam of *Decision Support Databases*, 27/11/2020

Deliver all sheets within 2h from start time. It is forbidden to consult any material during the test.

1. **(Mandatory)** Let us consider the following datawarehouse, without null values, about train tickets (primary keys are underlined):

Route(RPK, Km, UnitPrice)
Traveler(TPk, TName, TCity)
Tickets(Month, RFk, TFk, Qty)

- (a) **(1 point)** Give an SQL query to find the total quantity of tickets and the total revenue, by TCity.
 - (b) **(1 point)** Give a logical query plan for the SQL query, and the type of the result.
 - (c) **(1 point)** Give an example of a *Foreign Column Join Index* on TCity. Must provide sample relations (4-5 rows each) for the schema above and the index.
 - (d) **(2 points)** Give a physical query plan for the SQL query at (a) assuming an index on TPk.
2. **(5 points)** Consider the datawarehouse of Exercise 1. A traveler is defined to be a *commuter*, if she travels at least 80% of her trips in at most 2 routes. Give an analytic SQL query to find for every TCity the number of commuters.

3. An automotive company produces motor bikes. Bikes are sold to dealers, which in turns sold them to final customers. A bike may have been sold to a dealer, but not yet re-sold to a final customer. Such bikes are called *on-sale*. Dealers sell a bike to final customers at a higher price than the one paid by the dealer to the company. The difference is the profit of the dealer.

Dealers have a name, and a location (city, state, country), and an age-of-business (number of years since they started business). Customers have a name, an address, and an age.

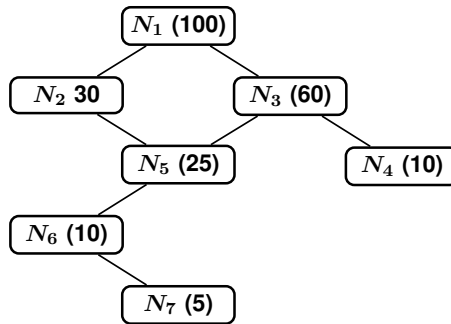
Managers are interested in the following business questions:

- (a) For the year 2018, the total revenue of bikes sold from the company to the dealers by dealer country.
- (b) For customers of age in [20,30], the total quantity of bikes sold to customers by customer gender and country.
- (c) For French dealers, the total profit of dealers by dealer city.
- (d) The number of bikes on-sales by age of business of the dealer.

With respect to the above business scenario, answer the following questions:

- (a) **(5 points)** Design a conceptual schema for the data mart to support the business questions. Your schema should at least be able to satisfy the above mentioned analysis requirements. You may motivate other suitable attributes for the dimensions. Specify the fact granularity, and for each measure, if it is additive, semi-additive or non-additive.
- (b) **(4 points)** Give a logical data mart design. Specify how dimensional attributes are updated.
- (c) **(1 points)** Write an SQL query to answer one business questions of your choice.

4. (5 points) Let us consider the following lattice of possible candidate views to materialize. The numbers associated with the nodes represent the view size, measured in terms of the number of tuples in the view.



Select 2 views to materialize, different from N_1 , with the greedy algorithm HRU.

5. (5 points) Consider the data mart of phone calls made by customers of a telephone company, without null values:

Customers(PkCust, CustName, Age)
 Calls(FkCust, Day, Month, Year, Duration, Charge)

and the query:

Q: **SELECT** Age, SUM(Charge) **AS** SC
FROM Calls, Customers
WHERE FkCust = PkCust **AND** Year = 2018 **AND** Age **BETWEEN** 20 **AND** 30
GROUP BY Age;

Show how to rewrite (if it is possible at all) the query Q using the view V

V: **SELECT** FkCust, Year, SUM(Charge) **AS** SC
FROM Calls
GROUP BY FkCust, Year;