University of Carthage 2019-2020
ENI-Carthage Eng. School

Data Integration & Data Warehousing Exam

Instructor: Dr. Rim Moussa Group: 3 IIS Time limit:90min

First Name:..... Last Name: ….....: Student ID#:.....

✂----------

Give four challenging issues in data integration /3

.....

.....

.....

.....

Compare row-oriented stores to column-oriented stores /2

.....

.....

.....

.....

Compare MOLAP to ROLAP engines in three points /3

.....

.....

.....

Relational Data warehouse Design & Performance Tuning Exercise /12

A consortium of banks wants to develop a data warehouse for effective decision-making about their loan schemes. The banks provide loans to customers for various purposes,such as House Building Loan, Car Loan, Educational Loan, Personal Loan, New Home Loan.
The whole country is categorized into states (e.g. California, Utah...), then cities (SF, LA,NYC,...).
Each Customer is described by his/her fname,lname,date-of-birth, gender, education-level, yearly-income.
We assume that
_the customers and banks may be in different cities and different states.
_loan is disbursed to customers at different rates.
_the warehouse age is 20 years. There are 50 states, 107,000 cities over the 50 states, 10 banks, 100 millions of customers, 5 purposes of loans, 2 genders (F or M).
The data warehouse should record an entry for each disbursement of loan to customer.

a.Design a star schema for the data warehouse clearly identifying the fact table(s), dimension tables, their attributes and measures along with the primary key and foreign key relationships.

DO NOT WRITE ANYTHING HERE

✂----------

b. Let's consider the following business queries
Q1: *the average loan rate per customer age category and per customer yearly income* category.
Q2: *the average amount of loans per quarter (winter,spring,summer,autumn).*
What would you add to the Customer and Date tables table as categorized attributes ?

.....

.....

.....

c. Devise an OLAP cube with 2 measures, five different dimensions (describe each hierarchy of levels and related properties) and one degenerate dimension,

.....

.....

.....

.....

DO NOT WRITE ANYTHING HERE

First Name:..... Last Name: ….....: Student ID#:.....

✂----------

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

d. Assume a change in a customer *c* address, (c moved from one state to another)
If you change the customer address in the DW, give an example of an OLAP query which resultset will be wrong after this update

.....

.....

what would you recommend to handle the *data change capture*

.....

.....

.....

.....

e. Estimate the cardinality of the big materialized view calculating the total amount of loans disbursed bank-wise (b for bank), state and city-wise (sc), Customer's Gender-wise(g) and loan' year and quarter-wise (yq)..

.....

.....

.....

DO NOT WRITE ANYTHING HERE

✂----------

f. Estimate in bytes the size of bitmap indexes built on the following columns:

customer.gender .....

customer.city .....

loan.type .....

loan.customer-city .....

Assuming we implemented a bitmap indexes on each of customer gender, customer city, customer-yearly-income-category, and a join bitmap index on loan-customer-gender. Tell whether the engine will shuffle data blocks (d), index blocks (i) or both data and index blocks (d & i) from Hard drives to Main memory for the following queries.

Q3 number of female customers from Seattle ....

Q4 number of female and MBA educated customers, high yearly-income customers ....

Q5 show details of female, MBA educated, high yearly-income customers ….....

Q6 Calculate min, max, average amount of loans state-wise, customer educ-level wise, and customer gender wise ….....

g. You are given the following compression schemes for columnar storage: run-length encoding, bitmap encoding, dictionary encoding, frame of reference encoding, differential encoding and no compression; propose suitable compression scheme(s) for the following columns

loan.type > .....

Loan.rate > .....

Loan.amount > .....