

Name :- _____

Roll Number:- _____

Mid-Semester Exam, CSE656, Monsoon 2024

Marks: 30

Time: 1 Hour

Note:

- A Google Form link will be shared that has 30 MCQs. **Each Question may have more than one correct answer and carries one mark. There is no partial marking.** You will get one mark if all the correct options of the questions are selected.
- The Question number is in alphanumeric form. Please see the question detail from the hard copy given to you and answer the question on the google form by matching alphanumeric question string.
- NOTE We will collect the IP address from the Institute IT team of your m/c when you open the Google form link. If we find more than one IP addresses for the same link, a case of plagiarism will be created and forwarded to the DAC.
- Make sure only one tab is opened on your favorite browser for this Google link. All other tabs should be closed. If we find any other tab opened, you will be given ZERO marks in this mid-sem exam.
- Calculator is not allowed to use. So keep away your mobile phone.

A 'Charming Bank' (CB) has several branches in a city which can be managed by a person. It offers retail banking products/services to customers. The 'Multiply Financial Services' (MFS) offers different investment and insurance related services and products to their high net worth individuals, and has one office per city. The 'Zero Brokerage Financial Ltd' (ZERO) offers an online platform where the clients can trade the equities with zero brokerage. The ZERO partners with the 'Charming Bank' for their customers' banking needs, that is, linking the customers trading account with the bank account (for these customer, the CustomerID is prefix with '00' in the following Customer_Accounts table). Thus, the CB has two categorizes of customers, namely, their own customers and the customers of ZERO. A customer can have accounts in all CB, MFS and ZERO.

The CB management board decides to expand the business by offering services and products related to the financial sectors, and therefore decides to acquire the MFS so that the CB can reach out to broader market to acquire new clients and also do cross-offering services and products to the all customers. To achieve this objective, the CB builds an information integration system, for the following applications, which integrates data from newly acquired MFS into the CB system to provide a unified view of customer accounts, transactions, and branch information. The applications are:

- **Unified Customer View:** Consolidate customer profiles across banks.
- **Transaction History:** Merge transaction records from different banks.
- **Branch Information:** Integrate and standardize branch details from acquired banks.
- **Customer Financial Overview:** Provide a comprehensive financial summary for each customer.
- **Fraud Detection:** Aggregate transaction data for detecting fraudulent activities.

Taking a simpler view of the above problem, the schemas of the local databases and the mediated schemas are given below.

The CB database schemas are (The primary key is each schema is underlined):

Customer_Accounts(CustomerID, FullName, AccountNumber, BranchCity, Balance)

Transactions(TransactionID, AccountNumber, BranchID, CustomerID, Product_Name, Transaction_Date, Amount, Trans_Type)

Branches(BranchID, BranchCity, Managed_By)

The MFS database schemas are (The primary key is each schema is underlined):

Customer_Profile (ClientNumber, Name, AccountCode, OfficeCity, Invst_Amount)

Products_Transactions (TransNum, AccountCode, OfficeCode, ClientNumber, ProductName, Date, Amount, BuyORSell)

- Offices (OfficeCode, OfficeCity, ManagerName)

The ZERO database schema is (The primary key is the schema is underlined):

Customer_Profile (CB_BankAccountID, GuestNumber, Guest_Name, Guest_City, Portfolio_Value)

Mediated Schemas

Unified_Customers(CustomerID, FullName, AccountNumber, BranchCity, Balance)

Unified_Transactions(TransactionID, AccountNumber, BranchID, CustomerID, ProductName, TransactionDate, Amount, TransactionType)

Unified_Branches(BranchID, BranchCity, Branch_Managed_By)

Materialized Views Schemas

Customer_Financial_Overview_View (FullName, AccountNumber, Total_Net_Worth)

Branch_Transaction_Summary (BranchName, TotalTransactions, TotalAmount)

The data instances of the following relational tables are:

CB: Customer_Accounts

CustomerID	FullName	AccountNumber	Balance	BranchCity
1	Ramesh Jaju	235	100000	New Delhi
2	Deba Jyoti	324	340000	New Delhi
003	Venkata R	346	560000	Lucknow
4	Ratan Bihari	321	5000	Patna

MFS: Customer_Profile

ClientNumber	Name	AccountCode	OfficeCity	Invst_Amount
C11	Rakesh Jaoo	S2353	New Delhi	5000000
C12	Deva Jyothi	S2373	New Delhi	9400000
C13	Venkat Ramaswamy	S4321	Kochi	9600000
C14	Rathan Vihaari	S5462	Bengaluru	1000000

- **789T12.** Which attributes should be used as Blocking and Matching for entity matching in CB.Customer_Accounts and MFS.Customer_Profile tables (assume the nomenclature of attributes in both schemas is same)?
- **189B32.** What is the primary function of the `Unified_Branches` table in the mediated schema?
- **184BN3.** Which of the following techniques (or in combination of) can be used for schema matching?
- **56B78G.** When the above mediated schemas are designed, what aspects should be considered?
- **41A627.** Which SQL statement creates a table to store unified customer data? *From Customer Profile*
- **31B789.** What is the main purpose of mapping schemas of **Customer_Accounts** and **Customer_Profiles** in the integration process?
- **718C25.** In the given above schema of **Branch_Transaction_Summary**, how many Dimension Tables (DT) and Fact Tables (FT) would be required to create a materialized view that aggregates transaction amounts and total number of transactions by branch name?
- **41G678.** What is the main advantage of **Customer_Financial_Overview_View** and **Branch_Transaction_Summary** materialized views in the integrated system?
- **11987T.** In the above schema integration, which approach is best for handling synonyms in field names (e.g., `FullName` vs. `Name`, `BranchID` vs `OfficeCode`)?
- **978V45.** In the above data instances, which method is commonly used for Name string matching to align fields with similar names?
- **789B62.** In the above CB database schemas, how can schema mapping help in dealing with inconsistencies in data formats across integrated sources?
- **892B67.** What would be the super key of **Unified_Branches** relation?
- **782B79.** Which of the following are correct mappings for integrating CB.Banches and MFS.Offices schemas?
- **89G674.** If we need to find those customers who are having a/cs in both CB and MFS, what method would you use?
- **89UT31.** Let's say the Customer_Accounts table has 100 million records and Customer_Profile table has 10 million records. What challenges you may encounter in for Entity Matching task?
- **278BH2.** What challenges might arise in schema mapping when integrating customer data?
- **4BH23I.** Which methods can be used to handle variations in names during entity matching?
- **6NH241.** How can domain-specific knowledge assist in schema matching?
- **289GF5.** Which methods can be used to handle variations in names during entity matching?

- **89BH26.** What would be the Jaro Score for matching 'Deba Jyoti' and 'Deva Jyothi' strings in the above CB.Customer_Accounts and MFS.Customer_Profile relations? The formula to compute the Jaro score is $\text{jaro}(x,y) = 1 / 3[c/|x| + c/|y| + (c - t/2)/c]$. Please do not consider space as a character in the both strings.
- **91245G.** We know that each Branch belongs to a District, a District belongs to a state. This information can be captured in a schema **Branch-Info**(BranchName, District, State). Which SQL statement will analyse the total business done by all branches in a district and by in a State?
- **613B78.** In the above mediated schemas, how is 'AccountNumber' used to integrate transactions?
- **56BG17.** What is a primary benefit of using '**Customer_Financial_Overview**' materialized views for reporting?**
- **2BHU81.** Which of the following algorithms may be used for 'FullName' and 'Name' string similarity measurement?
- **7BHU29.** How can missing data affect the entity matching process?
- **256BN0.** Which techniques are useful for schema matching when integrating data from heterogeneous sources?
- **6NTUB1.** In evaluating an entity resolution process, which of the following are true regarding precision and recall?
- **89GH27.** Which strategy would be better to match the 'FullName' attribute values of **Customer_Accounts** with the 'Name' attribute values of **Customer_Profile** data strings?
- **289BG8.** To design the mediated schemas of unified customers, transactions and branches, which design methodology would be more applicable?
- **6HU219.** What will be the most suitable architecture for the above information integration system?