**44-560 Advanced Topics in Database Systems**

**Assignment 05: Database Performance Tuning and Query Optimization**

The problems on this worksheet use the following database.



The above ER diagram is the assumed database design of a pharmaceutical company like (CVS, Walgreens, etc.) representing Point-of-Sale transactions at a store by a customer buying medicine. Assuming the customer table contains the registered customers for the company, there will hundreds of thousands of transactions occurring in an hour across all the stores in the USA. The data management design must be optimized to make the transactions fast and efficient. There are two access plans given below where it varies in the approach of the execution of the same query. You will need to calculate the results/rows for that particular access plan and compare it with the other.

Assume that

* The Medicine table has 70,000 rows.
* The Customer table has 10,000 rows.
* The Transaction table has 3,000,000 rows.
* The TransactionType table has 10 rows.
* The transaction date '09-08-2016' shows up in 500 rows of the Transaction table.
* Medicine ‘Paracetamol’ shows up in 30000 rows of the Transaction table.
* Customer ‘111342’ shows up in 500 rows of the Transaction table
* Transaction type ‘CASH-HALF’ shows up in 800 rows of the Transaction table.
* Transaction type ‘INSURANCE-FULL’ shows up in 900 rows of the Transaction table.

1. Consider the following query:

SELECT Customer.custID, Customer.firstName, Customer.lastName, Transaction.storeId, Transaction.date, Transaction.transactionType

FROM Customer, Transaction

WHERE Customer.custID = Transaction.custID

AND Transaction.date = '09-08-2016';

Below are two access plans for executing this query. Fill in the missing information in the last four columns.

Plan A:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Step** | **Operation** | **I/O Operations** | **I/O Cost** | **Resulting Rows** | **Total I/O Cost** |
| A1 | Cartesian product (Customer, Transaction) | 10,000 +3,000,000 | 3,010,000 | 30,000,000,000 | 3,010,000 |
| A2 | Select rows from A1 with matching custID | 30,000,000,000 | 30,000,000,000 | 3,000,000 | 30,003,010,000 |
| A3 | Select rows from A2 with Transaction.date = '09-08-2016' | 3,000,000 | 3,000,000 | 500 | 30,006,010,000 |

Plan B:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Step** | **Operation** | **I/O Operations** | **I/O Cost** | **Resulting Rows** | **Total I/O Cost** |
| B1 | Select rows from Transaction with Transaction.date = '09-08-2016' | 3,000,000 | 3,000,000 | 500 | 3,000,000 |
| B2 | Cartesian product (B1, Customer) | 10,000+500 | 10,500 | 5,000,000 | 3,010,500 |
| B3 | Select rows in B2 with matching custID | 5,000,000 | 5,000,000 | 500 | 8,010,500 |

1. Which plan is better between the two plans in question 1? Explain your reasoning.

Answer: Plan is better since Plan B has less Total I/O cost compared to Plan A

1. Consider the following query:

SELECT Transaction.transactionType, Transaction.date, Medicine.medicineName,

FROM Transaction, Medicine

WHERE Transaction.medicineID = Medicine.medicineID

AND Transaction.transactionType = ‘CASH-HALF’;

Below are two access plans for executing this query. Fill in the missing information in the last four columns.

Plan A:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Step | Operation | I/O Operations | I/O Cost | Resulting Rows | Total I/O Cost |
| A1 | Cartesian product (Transaction, Medicine) | 3,000,000+70,000 | 3,070,000 | 210,000,000,000 | 3,070,000 |
| A2 | Select rows from A1 with matching medicineID’s | 210,000,000,000 | 210,000,000,000 | 3,000,000 | 210,003,070,000 |
| A3 | Select rows from A2 with transactionType = ‘CASH-HALF’ | 3,000,000 | 3,000,000 | 800 | 210,006,070,000 |

Plan B:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Step | Operation | I/O Operations | I/O Cost | Resulting Rows | Total I/O Cost |
| B1 | Select rows from Transaction with transactionType = ‘CASH-HALF’ | 3,000,000 | 3,000,000 | 800 | 3,000,000 |
| B2 | Cartesian product (B1, Medicine) | 800+70,000 | 70,800 | 56,000,000 | 3,070,800 |
| B3 | Select rows in B2 with matching medicineID’s | 56,000,000 | 56,000,000 | 800 | 59,070,800 |

1. Which plan is better between the two plans in question 3? Explain your reasoning.

Answer: Plan is better since Plan B has less Total I/O cost compared to Plan A

**Please submit this word document in submission without zipping the file with the naming convention as LastName\_Assignment05.docx.**