**Assignment 12**

**Swapnil Shashikant Kamate**

**Part -1**

Problems 1 and 2 are based on the following query:

SELECT EMP\_LNAME, EMP\_FNAME, EMP\_AREACODE, EMP\_SEX FROM EMPLOYEE WHERE EMP\_SEX = 'F' AND EMP\_AREACODE = '615' ORDER BY EMP\_LNAME, EMP\_FNAME;

1)What is the likely data sparsity of the EMP\_SEX column?

**Answer:**

In EMP\_SEX column, the data sparsity is low because two possible values of gender M and F. If there are more than two options as well to choose in this column.

2. What indexes should you create? Write the required SQL commands.

**Answer:**

One of the indexes can be created on Employee names (EMP\_LNAME and EMP\_FNAME) and another can be created on EMP\_AREACODE.

CREATE UNIQUE INDEX EMP\_INDEX2 ON EMPLOYEE(EMP\_LNAME,EMP\_FNAME);

CREATE UNIQUE INDEX EMP\_INDEX1 ON EMPLOYEE(EMP\_AREACODE);

3. Using Table 11.4 as an example, create two alternative access plans. Use the following assumptions:

a. There are 8,000 employees.

b. There are 4,150 female employees.

c. There are 370 employees in area code 615.

d. There are 190 female employees in area code 615.

**Answer:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Plan** | **Step** | **Operation** | **i/o Operations** | **i/o Cost** | **Resulting Set Rows** | **Total i/o Costs** |
| A | A1 | Full Table Scan  Select rows with EMP\_SEX=’F’ and EMP\_AREACODE = ‘615’ | 8000 | 8000 | 190 | 8000 |
| A | A2 | SORT | 190 | 190 | 190 | 8190 |
| B | B1 | Index Scan Range of EMP\_INDEX1 | 370 | 370 | 370 | 370 |
| B | B2 | Table access by ROW\_ID | 370 | 370 | 370 | 740 |
| B | B3 | Select Rows with EMP\_SEX = ‘F’ | 370 | 370 | 190 | 930 |
| B | B4 | SORT | 190 | 190 | 190 | 1120 |

I have used SORT function to sort the results by employee last name and employee first name. Index created in the previous question can be used i.e EMP\_INDEX1 to get employee Row ID’s. Then we filter this data with SEX =’F’ and finally sorted data with results by employee last name and first name.

Problems 4−6 are based on the following query:

SELECT EMP\_LNAME, EMP\_FNAME, EMP\_DOB, YEAR(EMP\_DOB) AS YEAR FROM EMPLOYEE WHERE YEAR(EMP\_DOB) = 1976;

4. What is the likely data sparsity of the EMP\_DOB column?

**Answer:**

It will have high sparsity as it stores all the employee birthdays.

5. Should you create an index on EMP\_DOB? Why or why not?

**Answer:**

It would be a bad idea to create an index as it uses YEAR in storing Employee DOB.

Part -2

1. Specify the minimum types of operations the database must support to perform the following operations. These operations include remote requests, remote transactions, distributed transactions, and distributed requests. At site C

Answer:

a) Remote request

b) Remote request

c) Distributed request

d) Distributed request: - UPDATE Customer and Insert statement require remote request but the entire query requires distributed request capabilities.

e) Distributed request: - It accesses two remote sites.

**At Site A:**

f) Distributed request: - One local and one remote.

g) Remote request

h) Distributed request: - Table is partitioned into two fragments

**At Site B:**

i) Remote request

j) Distributed request: - One local and one remote.

k) Distributed request: - Table is partitioned into two fragments

2)

**Answer:**

The customer table should be partitioned horizontally by state. There should be 4 rows corresponding to each state. This will help in listing the current customers by region and the new customers by region. The invoice table should be partitioned horizontally by region. This will help in getting results for queries like getting all invoices by customer and all invoices by region.