Objectives

◆ In this session, you will learn to:

◆ Appreciate the need for denormalization

Understanding Denormalization

- Sometimes, you have to join multiple tables to get a simple output.

- This affects the performance of a query.

- In such cases, a degree of redundancy is introduced either by adding extra columns or extra tables.

Definition of Denormalization

- The intentional introduction of redundancy in a table in order to improve performance is called denormalization.
- Consider the tables, ORDERS and PRODUCTS, as shown in the following diagram.

ORDERNO	PRODUCTID	QTY
101	P1	2
102	P3	1
103	P1	1
104	P2	3
105	P2	2

PRODUCTID	PRODUCT DESC	COST
P1	Superheroes	20
P2	Toy Guns	10
P3	Soft Toys	12

Definition of Denormalization (Contd.)

- If a shopkeeper wants to calculate the total sales of his shop, he can use the following query:

  ```
  SELECT SUM((cost*qty)+(0.10*cost*qty))
   FROM Orders JOIN Products
   ON Orders.ProductId = Products.ProductId
  ```

- There is a join and computation involved in the preceding query.

- Therefore, if there are thousands of rows, the server will take a lot of time to process the query and return the result.

Definition of Denormalization (Contd.)

◆ To speed up the processing of the query, store the cost of each order along with the tax, as shown in the following diagram.

ORDERS

ORDERNO	PRODUCTID	QTY	PRODUCTCOST	TAX	ORDERCOST
101	P1	2	40	8	88
102	P3	1	12	1.2	13.2
103	P1	1	20	2	22
104	P2	3	30	9	99
105	P2	2	20	4	44

Definition of Denormalization (Contd.)

- To find out the total sales made, you can now issue the following query:

  ```
  SELECT SUM (OrderCost) FROM Orders
  ```

- The decision to denormalize will result in a compromise between performance and data consistency.

- Denormalization also increases disk space utilization.

Just a minute

◆ In the reporting system, the total amount paid to a contract recruiter is often required. The required result can be calculated using the following two tables.

CONTRACT_RECRUITER	PAYMENT
CCONTRACTRECRUITERCODE	CSOURCECODE
CNAME	MAMOUNT
VADDRESS	CCHEQUENO
CCITY	DDATE
SIPERCENTAGECHARGE	

Since the size of the PAYMENT table is large, the join between the two tables takes time to generate the required output. It is, therefore, necessary to improve the performance of this query.

Just a minute (Contd.)

◆ Solution:

◆ Denormalize the tables. Add a column called MTOTALPAID to the following CONTRACT_RECRUITER table.

CONTRACT_RECRUITER
CCONTRACTRECRUITERCODE
CNAME
VADDRESS
CCITY
SIPERCENTAGECHARGE
MTOTALPAID

Practice Exercises

- Exercise 1:
  - New Heights is a training institute that provides courses on various non-technical subjects, such as personality improvement and foreign languages. Xuan, the database designer at the institute, has made the following relation to represent the data about students, batches, and modules covered in the batches:
    - STUD-ID: Student's id (unique)
    - NAME: Name of student
    - BATCH-NO: Batch number (one student can belong to only one batch)
    - SLOT: Time and day during which the batch of students attends the class
    - MODULE: Module or subject (one batch will do several modules)
    - MARKS: Marks obtained in a module test
  - Xuan now needs to simplify the above relation by normalizing it.

◆ Exercise 2:

◆ Each time the salary slip for an employee is generated, the referral bonus (if present) has to be calculated and printed in the salary slip. The following tables are used for solving the preceding query.

EMPLOYEE
cEmployeeCode
vFirstName
vLastName
cCandidateCode
vAddress
cCity
cZip
cCountryCode

MONTHLY_SALARY
cEmployeeCode
mMonthlySalary
dPayDate

EMPLOYEE_REFERRALS
cEmployeeReferrralNo
cEmployeeCode
cCandidateCode

Practice Exercises (Contd.)

EMPLOYEE
cPhone
vQualification
dBirthDate
cSex
cCurrentPosition
cDesignation
cEmailId
cDepartmentCode
cRegion
ImPhoto
vSkill
dJoiningDate
dResignationDate
cSociaLSecurityNo

However, the table structures are large, it is necessary to improve the performance of this query by modifying the table structures. Identify how to increase the performance of queries.

Summary

- In this session, you learned that:
  - The intentional introduction of redundancy in a table in order to improve performance is called denormalization.
  - The decision to denormalize results in a compromise between performance and data integrity.
  - Denormalization increases disk space utilization.