| **National University of Computer and Emerging Sciences, Lahore Campus** | | | | | |
| --- | --- | --- | --- | --- | --- |
| final design | **Course:** | **Data Warehousing and Business Intelligence** | **Course Code:** | **DS3003** | |
| **Program:** | **BS (Data Science)** | **Semester:** | **Fall 2023** | |
| **Date:** | **Thu 14-Sep-2023** | **Total Marks:** | **10** | |
|  |  |  |  | |
| **Section:** | **BDS-5A** | **Max. Time:** | **10 min.** | |
| **Quiz:** | **1 (PDM)** |  |  | |
| \* | | | | |

**Q.** Consider the following normalized data structure:

SALES (saleId, storeId, saleDate, …)

SALES\_DETAIL (transactionId, saleId, itemId, itemQty, …)

Assume there are two million sales and twenty million sales details. Record length of both tables is same i.e., 100 bytes and each column of both tables (including PK/FK columns) is of same size i.e., 10 bytes.

Query: SELECT \* FROM sales S JOIN sales\_detail D ON S.saleId = D.saleId;

You are required to improve the performance of the above query using pre-join de-normalization technique. Show your de-normalized data structure schema and evaluate the increase in additional storage cost (in %age or MB) for the de-normalized data structure.

**Ans:**

**Normalizes Structure:**

**Sales: 2m \* 100 = 200m**

**Sales\_detail: 20m \* 100 = 2000m**

**Total cost: 2200m**

**Denormalized Structure:**

**20m \* 190 = 3800m**

**Increase in additional storage cost: ((3800 – 2200)/2200)\*100 = 73%**

**or**

**Increase in additional storage cost: 3800 – 2200= 1600m/1024\*1024= 1526 MB**