

# DATA WAREHOUSE AND BUSINESS INTELLIGENCE PROJECT

**Name:** Jiyad Khan

**Roll Number:** 19I-1771

**Section:** BS DS (N)

**Course’s Instructor:** Muhammad Asif Naeem

**Due Date:** 2nd December, 2022

•

# Project Overview

METRO is one of the biggest superstores chains in Pakistan. The stores has thousands of customers and therefore it is important for the store to online analyse the shopping behaviour of their customers. Based on that the store can optimise their selling strategies e.g. giving promotions on different products.

# LEARNINGS FROM PROJECT

1. I learned Java for the first time.
2. I got an in-depth view of how a Data Warehouse works
3. I learned why Data Warehouses are essential in ourworldtoday

# MESH JOIN

The MESHJOIN Algorithm has been introduced by Polyzotis in 2008 to make a generic join operation in the transformation phase of ETL.

The main components of MESHJOIN are:

1. The disk-buffer which will be an array and used to load the disk partitions in memory.
2. The hash table stores the customers’ transactions.
3. The queue keep the record of all the customers’ transactions in memory with their respective arrival times

**STEP BY STEP IMPLEMENTATION OF MESHJOIN ALGORITHM:**

1. Read 50 tuples will be read, via TRANSACTIONS table as an input data into the hash table, join attributes in queue.
2. Next MasterData will be loaded into the disk buffer.
3. Cross matching tuples from disk buffer to the hash table.
4. If the tuple matches, add the attributes into the transaction tuple, calculate TOTAL\_SALE using QUANTITY and PRICE attributes; as it does not exists in TOTAL\_SALES.
5. The transaction tuple will have all the new attributes. Loading the tuple into Warehouse we will check whether we have info in the dimension then we will update both fact and dimension tables.
6. We will remove all joins withteh values of partition of the queue once we have completed the lookup of the disk buffers.

**=>**Repeat ALL the steps until you load all the data from TRANSACTIONS table to DW.

# Shortcomings of Mesh-Joins

Shortcoming of the meshjoin are really not negligible at here;

1. If the data keeps on coming to the stream once a time will come where the data stream buffer gets full.
2. Memory problem is great issue created while the usage of meshjoin.

# STAR SCHEMA

