# Parallel Computing in Database Query Processing
# Optimizing Performance through Parallelism

## INTRODUCTION

Parallel query optimization involves devising a strategy for executing database queries efficiently using parallel hardware. The specific approach varies based on the parallelism capabilities of the hardware, but typically involves dividing the data among multiple processors, known as partitioning. Parallel query processing refers to the conversion of abstract queries into execution plans that can be effectively run in parallel on a multi-processor computer system.

### Query Processing in DBMS
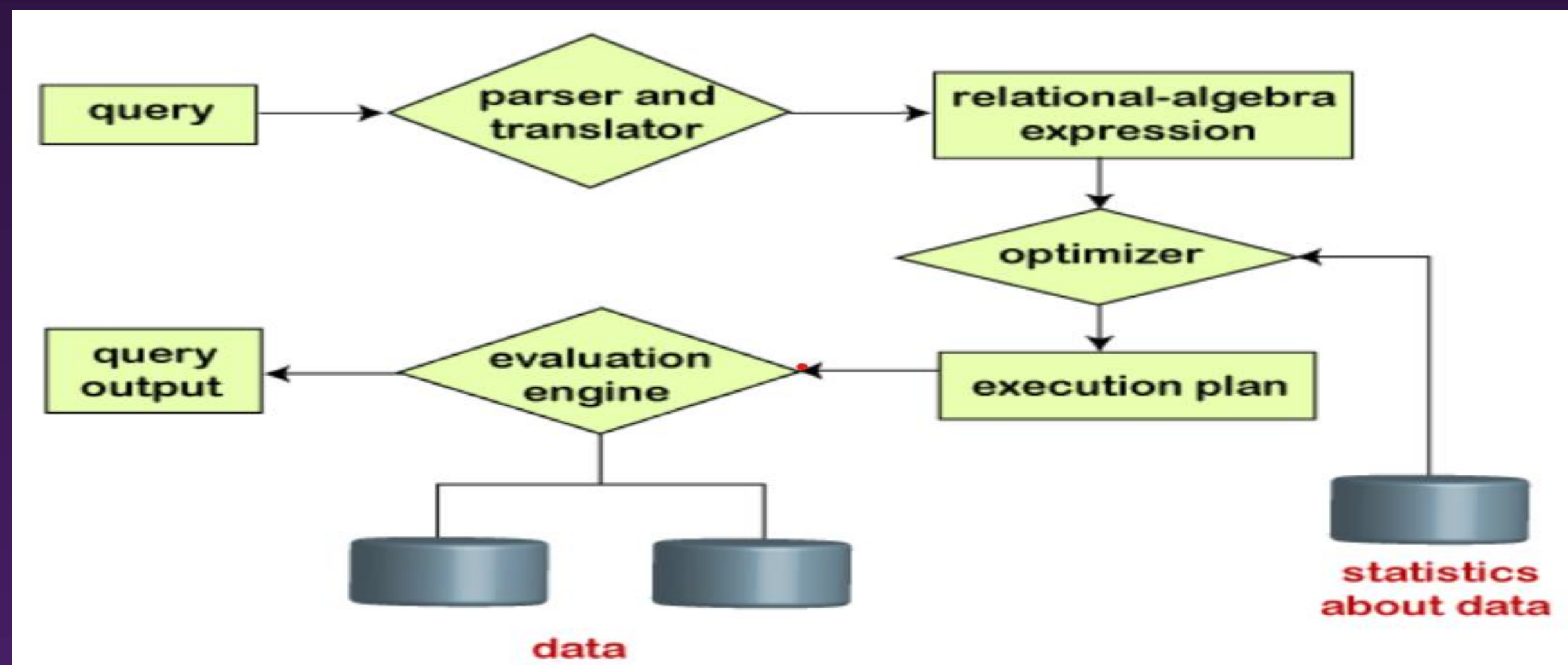


## Database Query Processsing

- **What is a Query in DBMS**? A query is a request sent to the database to fetch data. It involves specifying conditions, and the database retrieves matching data if it exists.

- **What is query optimization:** It involves determining the most efficient way to execute a query by considering factors like the database structure, data, indexes, statistics, and available system resources. The goal of query optimization is to reduce the expense associated with executing the query, which may include factors such as time, disk input/output, memory usage, or network traffic.

- **How to improve query performance?** : To enhance the efficiency of your query execution plan, employing strategies and recommended approaches can optimize both your query and database system. Common methods include indexing, partitioning, caching, and tuning.
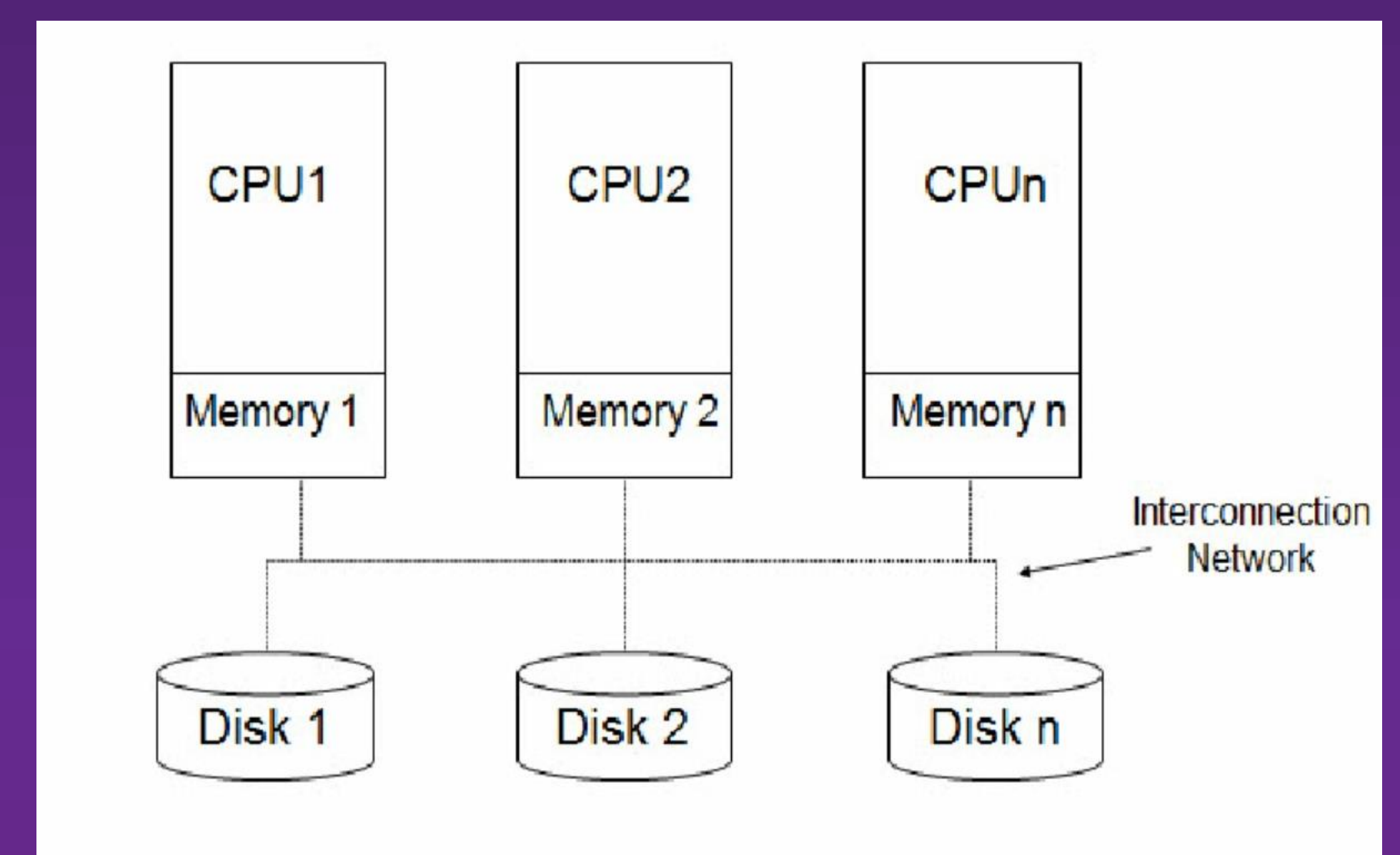
## Advanced Databases: Parallel Databases

### 1. SHARED MEMORY MACHINE
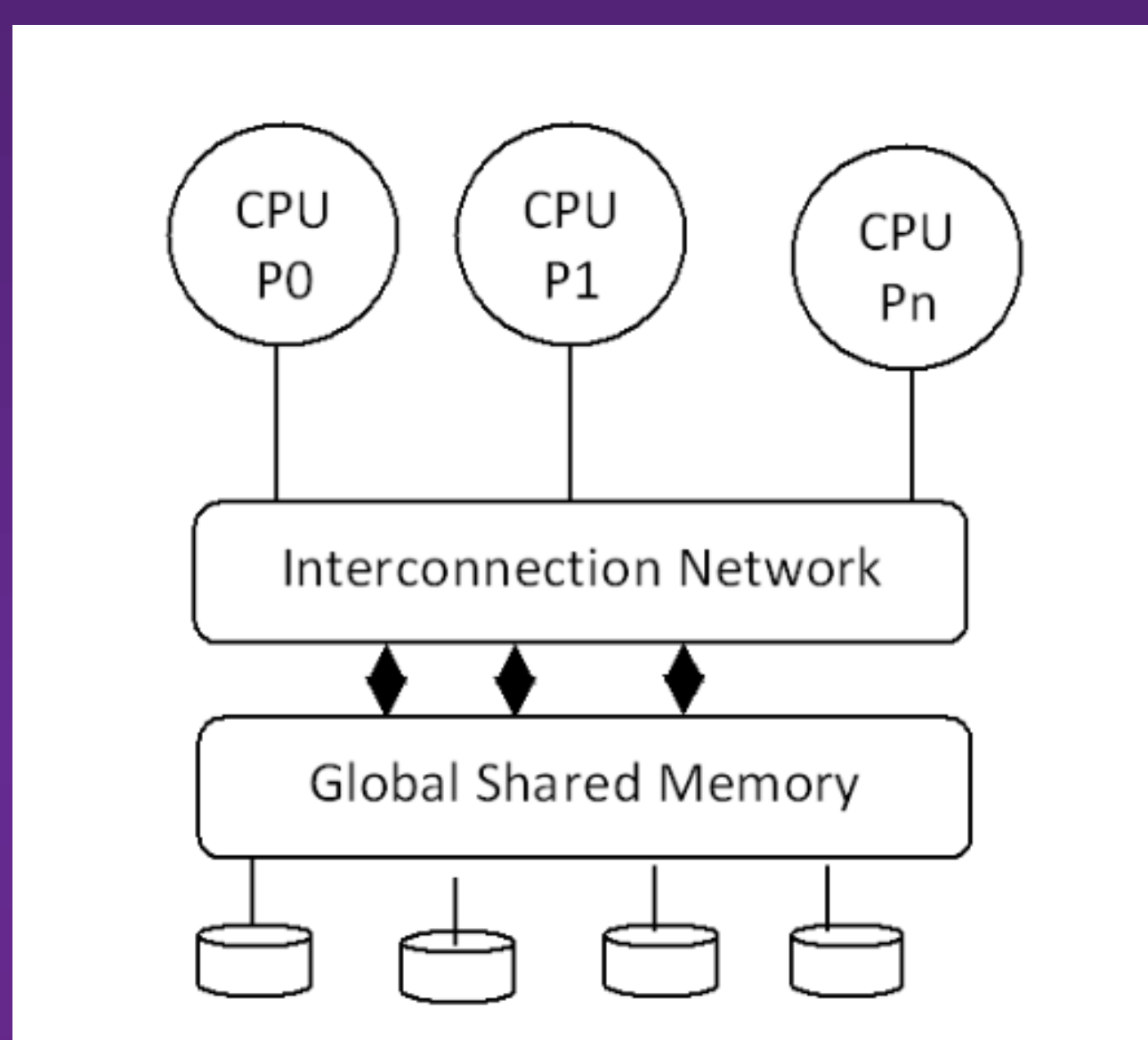


### 2. SHARED NOTHING MACHINE
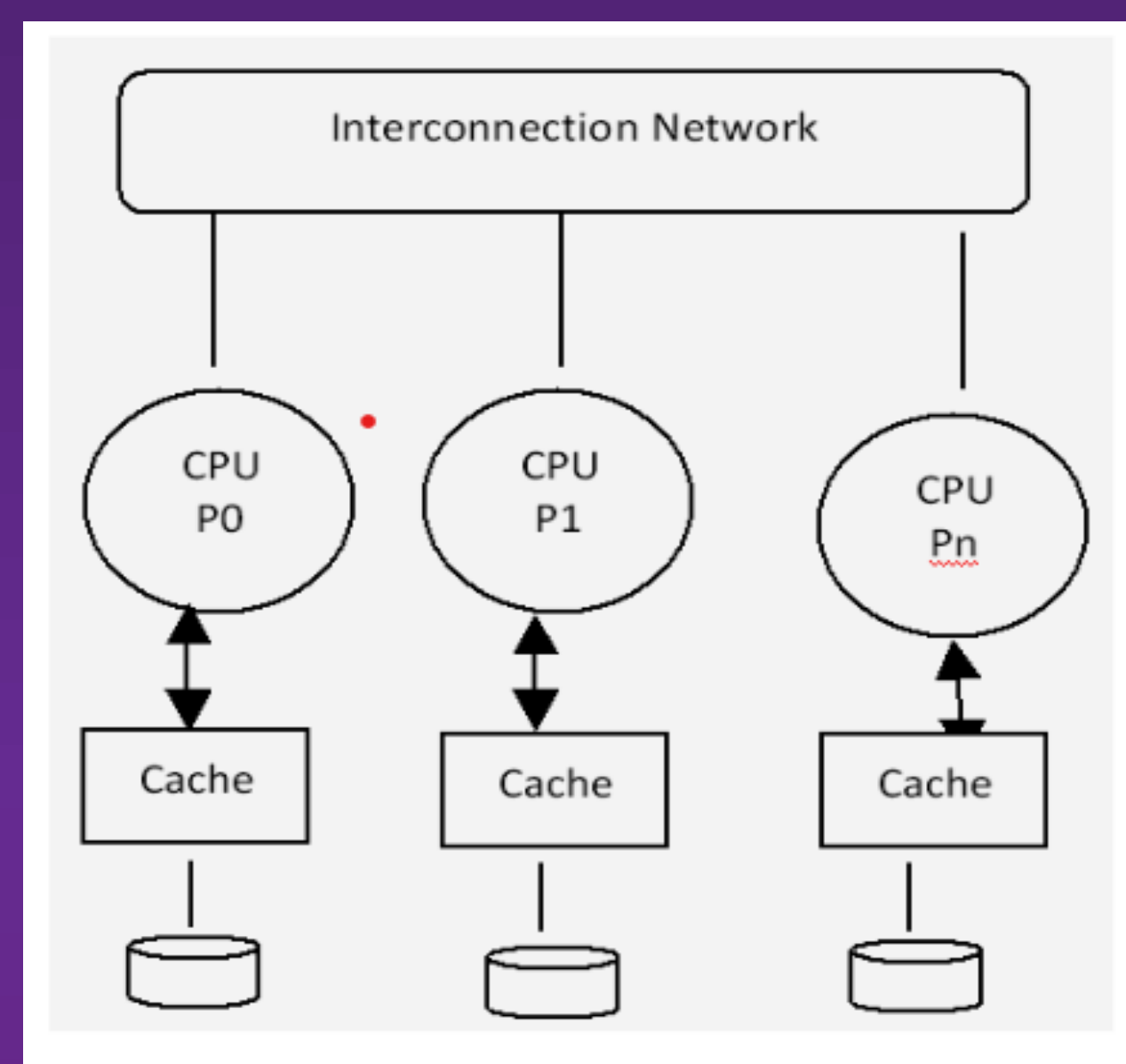


### 3. SHARED DISK ARCHITECTURE



## Partitioning Strategies

1) The last scheme we will talk about is called **round robin partitioning**. In this scheme we go record by record and assign each record to the next machine

2) In **hash partitioning** scheme, each record is hashed and is sent to a machine matches that hash value. This means that all like values will be assigned to the same machine (i.e. if value 4 goes to machine 1 then all of the 4s must go to that machine), but it makes no guarantees about where close values go.

3) **Range partitioning** is a type of relational database partitioning wherein the partition is based on a predefined range for a specific data field such as uniquely numbered IDs, dates or simple values like currency
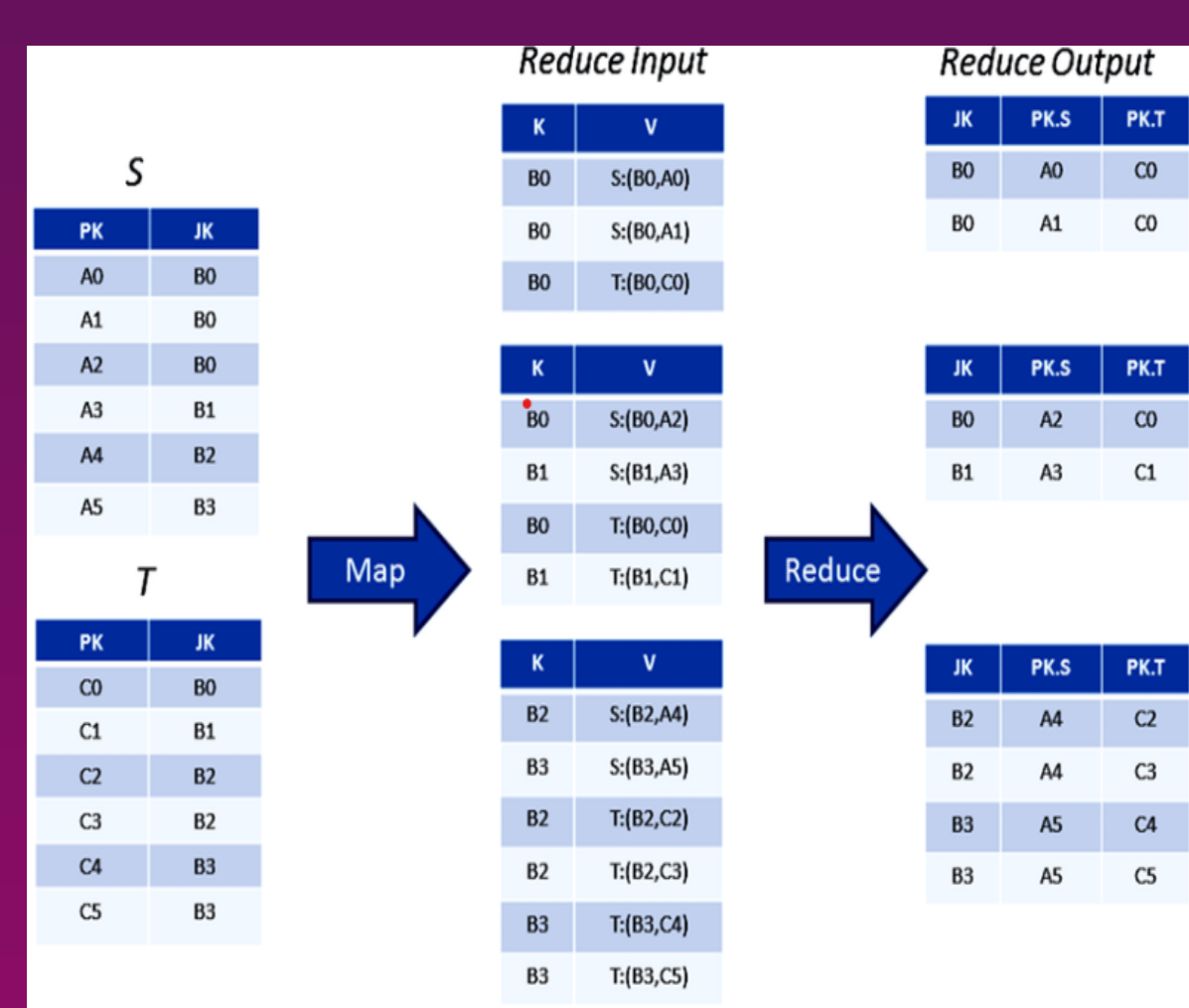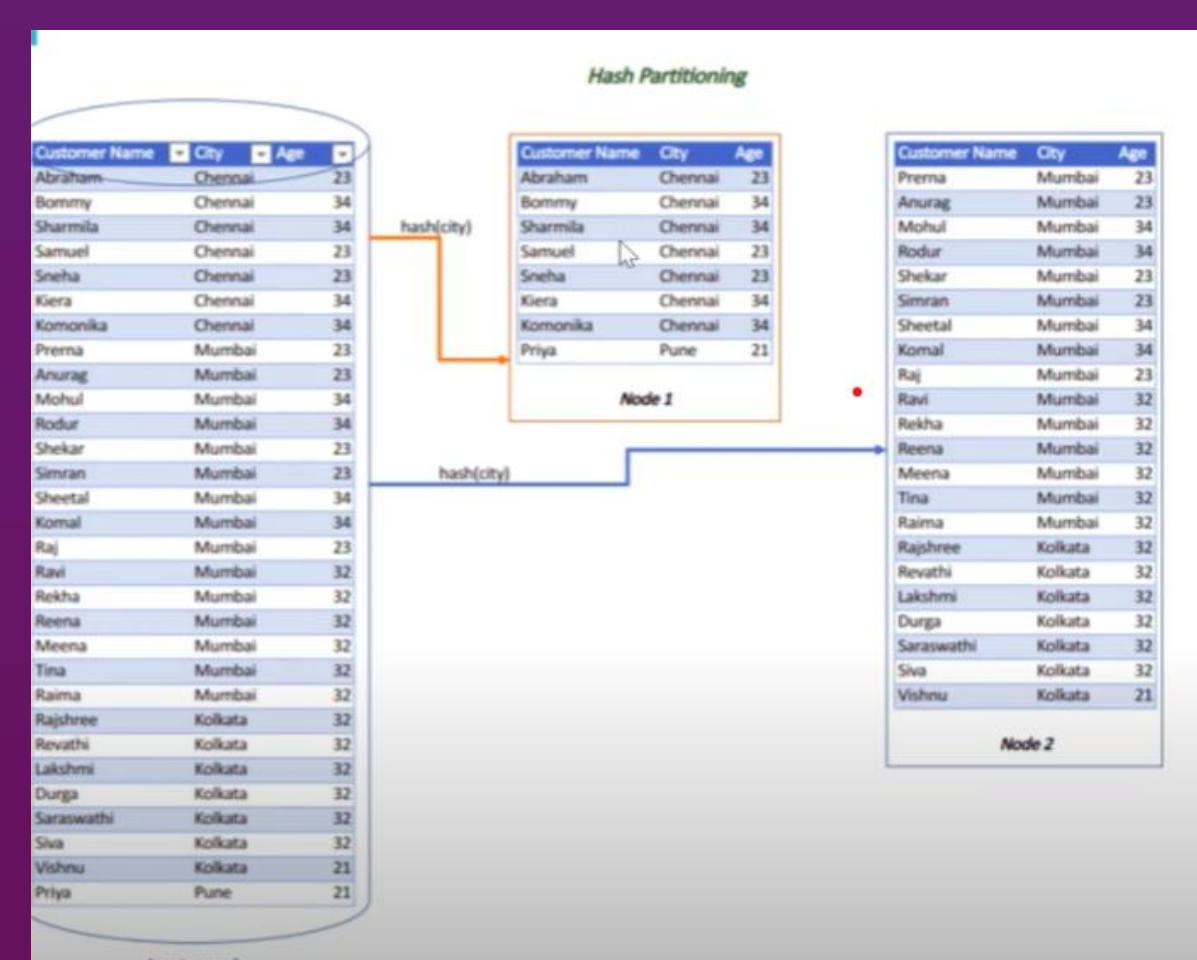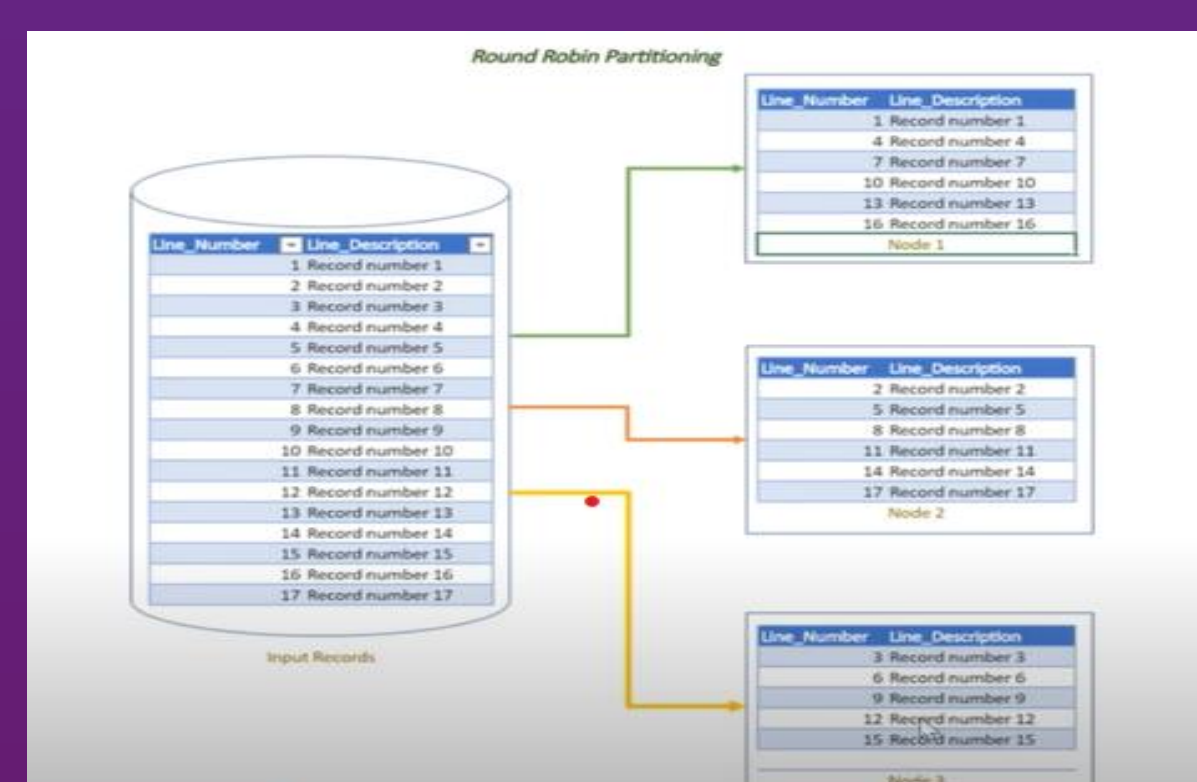




## Implementation in SQL Server

- SQL Server can utilize parallelism to speed up query execution.

- Parallelism divides query processing among multiple workers, akin to "many hands make light work."

- This approach reduces the overall execution time by processing different parts of the query simultaneously.

- Operations like Stream Aggregate benefit from parallelism, especially for tasks such as counting rows or performing aggregations on large datasets.

- Parallel query execution plans feature icons denoting operations involving multiple workers, such as small yellow arrows.

- Each worker handles a distinct segment of the query concurrently, and their partial results are combined to produce the final result.

- Parallel query processing offers significant performance improvements for resource-intensive queries on large datasets.
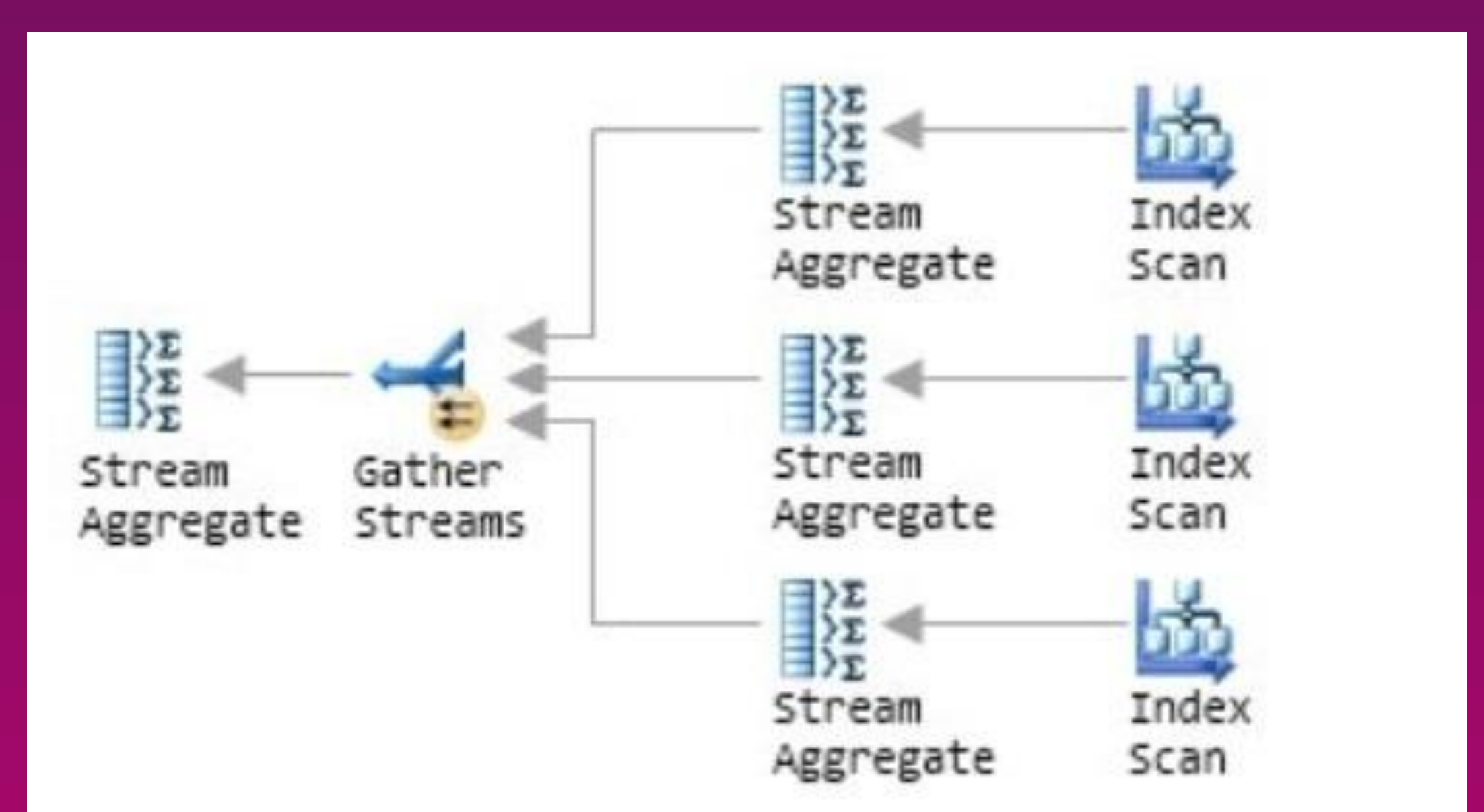
Submitted By,
Anas Ansari (22010219), Medini Patil (22010059)
Division: C
Class: BTech (2023-24)