



RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF ENGINEERING

DATA MINING & WAREHOUSEING
LECTURE-24

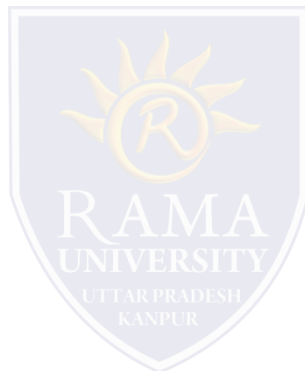
MR. DHIRENDRA

ASSISTANT PROFESSOR

RAMA UNIVERSITY

OUTLINE

- ❖ **TYPES OF DATABASE PARALLELISM**
- ❖ **INTRAQUERY PARALLELISM**
- ❖ **INTERQUERY PARALLELISM**
- ❖ **MCQ**
- ❖ **REFERENCES**



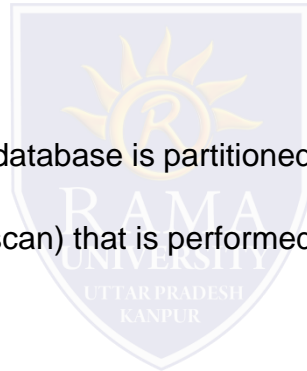
Types of Database Parallelism

Parallelism is used to support speedup, where queries are executed faster because more resources, such as processors and disks, are provided. Parallelism is also used to provide scale-up, where increasing workloads are managed without increase response-time, via an increase in the degree of parallelism.

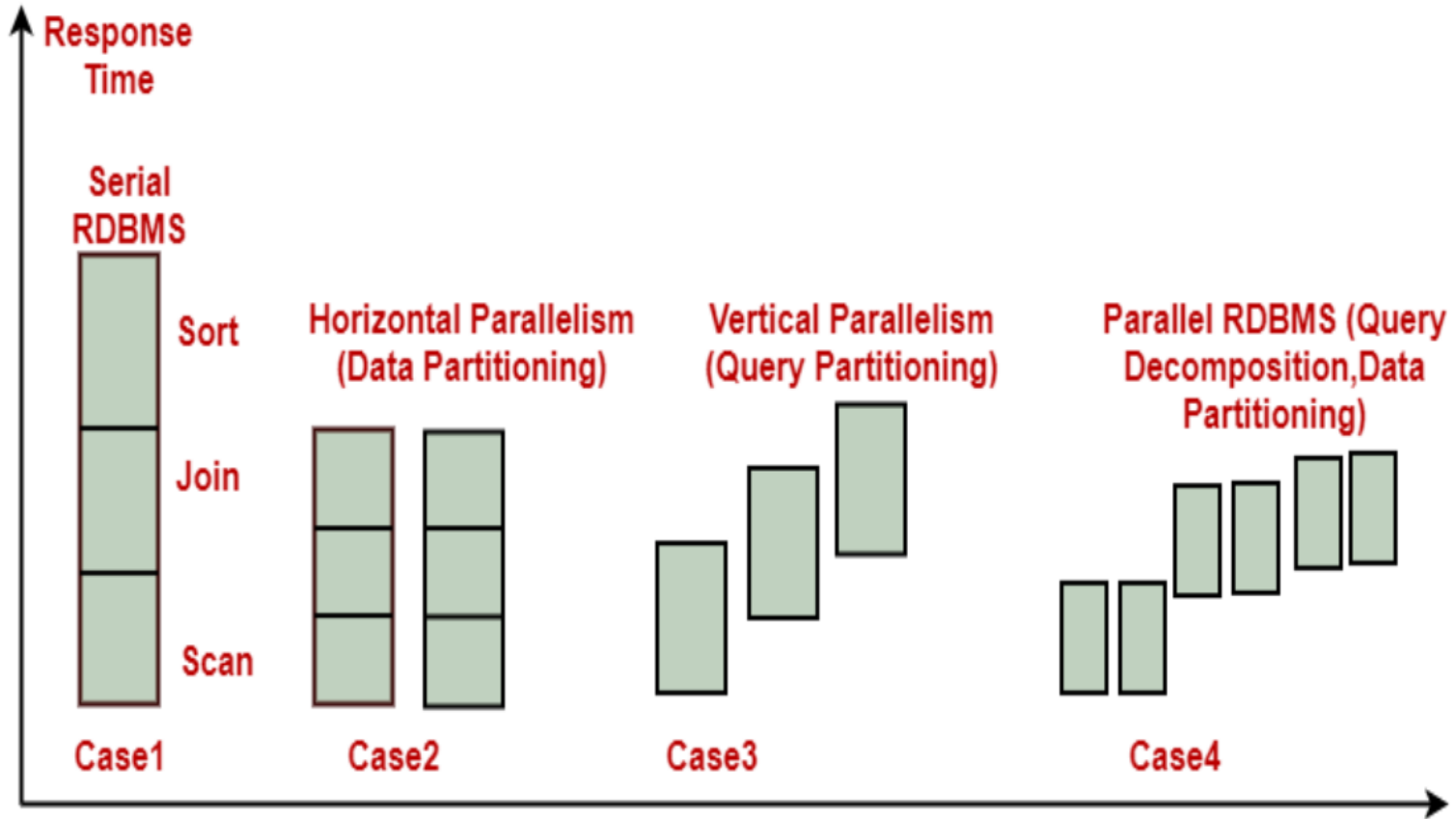
Different architectures for parallel database systems are shared-memory, shared-disk, shared-nothing, and hierarchical structures.

Horizontal Parallelism: It means that the database is partitioned across multiple disks, and parallel processing occurs within a specific task (i.e., table scan) that is performed concurrently on different processors against different sets of data.

Vertical Parallelism: It occurs among various tasks. All component query operations (i.e., scan, join, and sort) are executed in parallel in a pipelined fashion. In other words, an output from one function (e.g., join) as soon as records become available.

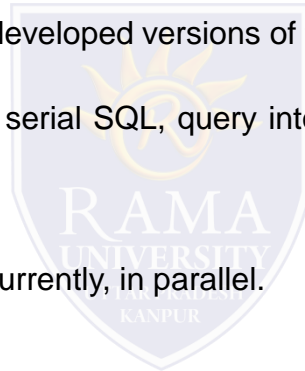


TYPES OF DATABASE PARALLELISM



Intraquery Parallelism

- Intraquery parallelism defines the execution of a single query in parallel on multiple processors and disks. Using intraquery parallelism is essential for speeding up long-running queries.
- Interquery parallelism does not help in this function since each query is run sequentially.
- To improve the situation, many DBMS vendors developed versions of their products that utilized intraquery parallelism.
- This application of parallelism decomposes the serial SQL query into lower-level operations such as scan, join, sort, and aggregation.
- These lower-level operations are executed concurrently, in parallel.



Interquery Parallelism

- In interquery parallelism, different queries or transaction execute in parallel with one another.
- This form of parallelism can increase transactions throughput. The response times of individual transactions are not faster than they would be if the transactions were run in isolation.
- Thus, the primary use of interquery parallelism is to scale up a transaction processing system to support a more significant number of transactions per second.
- Database vendors started to take advantage of parallel hardware architectures by implementing multiserver and multithreaded systems designed to handle a large number of client requests efficiently.
- This approach naturally resulted in interquery parallelism, in which different server threads (or processes) handle multiple requests at the same time.
- Interquery parallelism has been successfully implemented on SMP systems, where it increased the throughput and allowed the support of more concurrent users.

Multiple Choice Question

1. _____ is a the input to KDD.

- a) Data
- b) Information
- c) Query
- d) Process

2The output of KDD is _____.

- a) Data
- b) Information
- c) Query
- d) Useful information.

3. The KDD process consists of _____ steps.

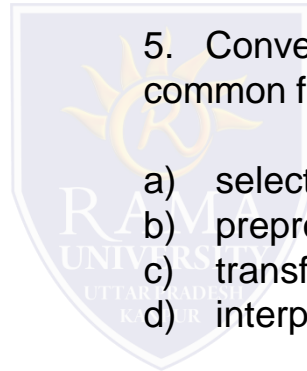
- a) three
- b) four
- c) five
- d) six

4. Treating incorrect or missing data is called as _____.

- a) selection
- b) preprocessing
- c) transformation
- d) interpretation

5. Converting data from different sources into a common format for processing is called as _____.

- a) selection
- b) preprocessing.
- c) transformation
- d) interpretation



REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm
- <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf> DATA MINING BOOK WRITTEN BY Micheline Kamber
- <https://www.javatpoint.com/three-tier-data-warehouse-architecture>
- M.H. Dunham, “ Data Mining: Introductory & Advanced Topics” Pearson Education
- Jiawei Han, Micheline Kamber, “ Data Mining Concepts & Techniques” Elsevier
- Sam Anahory, Dennis Murray,” data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, “ Pearson Education
- Mallach,” Data Warehousing System”, TMH
- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE’97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997
- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB’96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD’97
- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.
- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.