RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF EGINEERING

DATA MINING & WAREHOUSEING
LECTURE-25

MR. DHIRENDRA
ASSISTANT PROFESSOR
RAMA UNIVERSITY

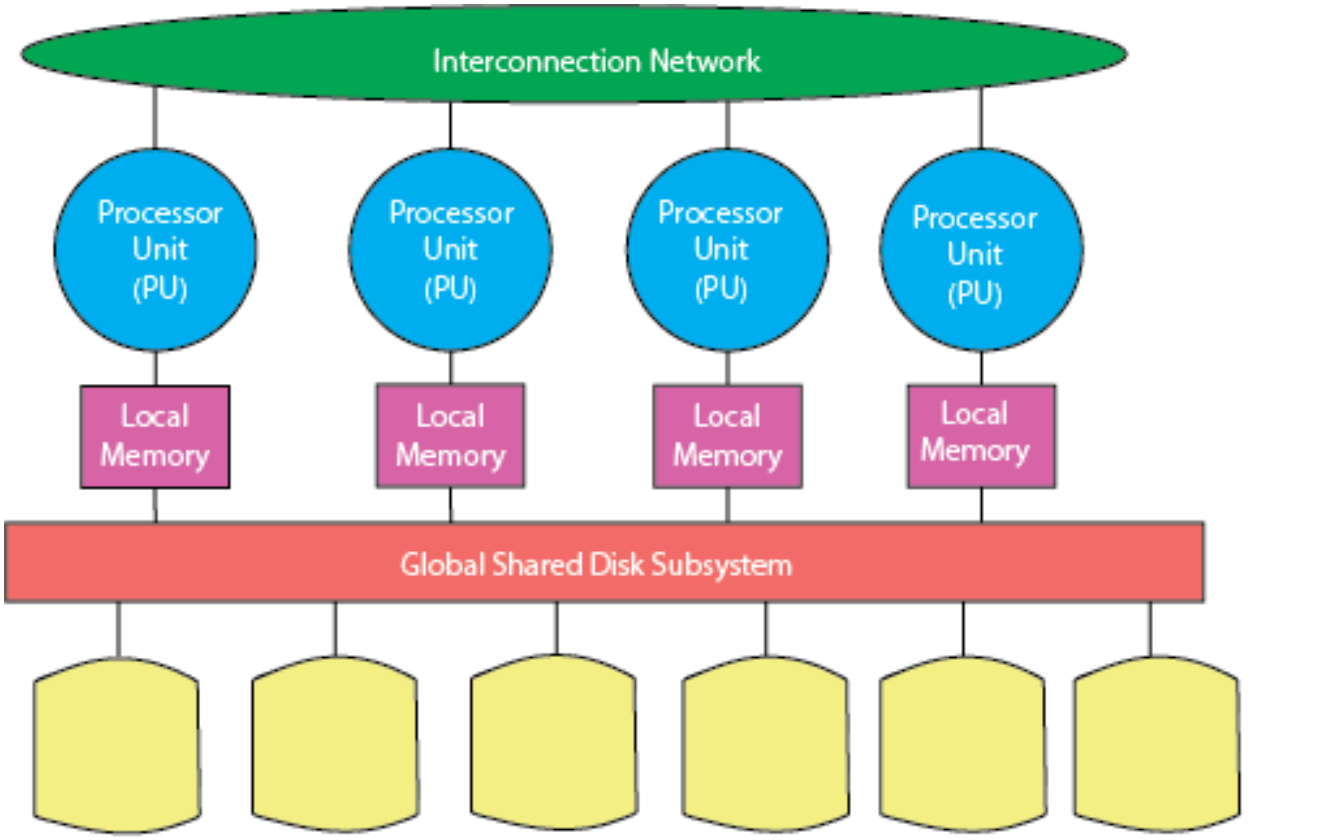
OUTLINE

- SHARED DISK ARCHITECTURE
- SHARED-MEMORY ARCHITECTURE
- SHARED-NOTHING ARCHITECTURE
- MCQ
- REFERENCES

Types of Database Parallelism

- Shared-disk architecture implements a concept of shared ownership of the entire database between RDBMS servers, each of which is running on a node of a distributed memory system.
- Each RDBMS server can read, write, update, and delete information from the same shared database, which would need the system to implement a form of a distributed lock manager (DLM).
- DLM components can be found in hardware, the operating system, and separate software layer, all depending on the system vendor.
- On the positive side, shared-disk architectures can reduce performance bottlenecks resulting from data skew (uneven distribution of data), and can significantly increase system availability.
- The shared-disk distributed memory design eliminates the memory access bottleneck typically of large SMP systems and helps reduce DBMS dependency on data partitioning.

Shared Disk Architecture



Distributed memory shared-disk architecture

Shared-Memory Architecture

•Shared-memory or shared-everything style is the traditional approach of implementing an RDBMS on SMP hardware.

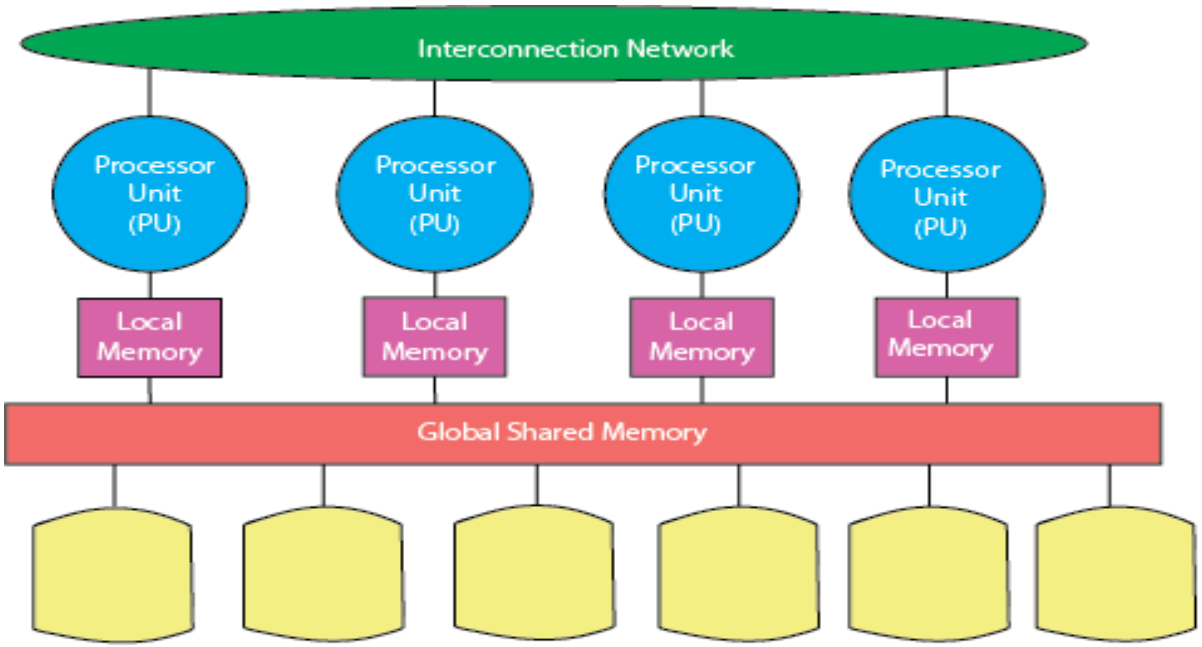
•It is relatively simple to implement and has been very successful up to the point where it runs into the scalability limitations of the shared-everything architecture.

•The key point of this technique is that a single RDBMS server can probably apply all processors, access all memory, and access the entire database, thus providing the client with a consistent single system image.

•In shared-memory SMP systems, the DBMS considers that the multiple database components executing SQL statements communicate with each other by exchanging messages and information via the shared memory.

•All processors have access to all data, which is partitioned across local disks.

Shared-Memory Architecture

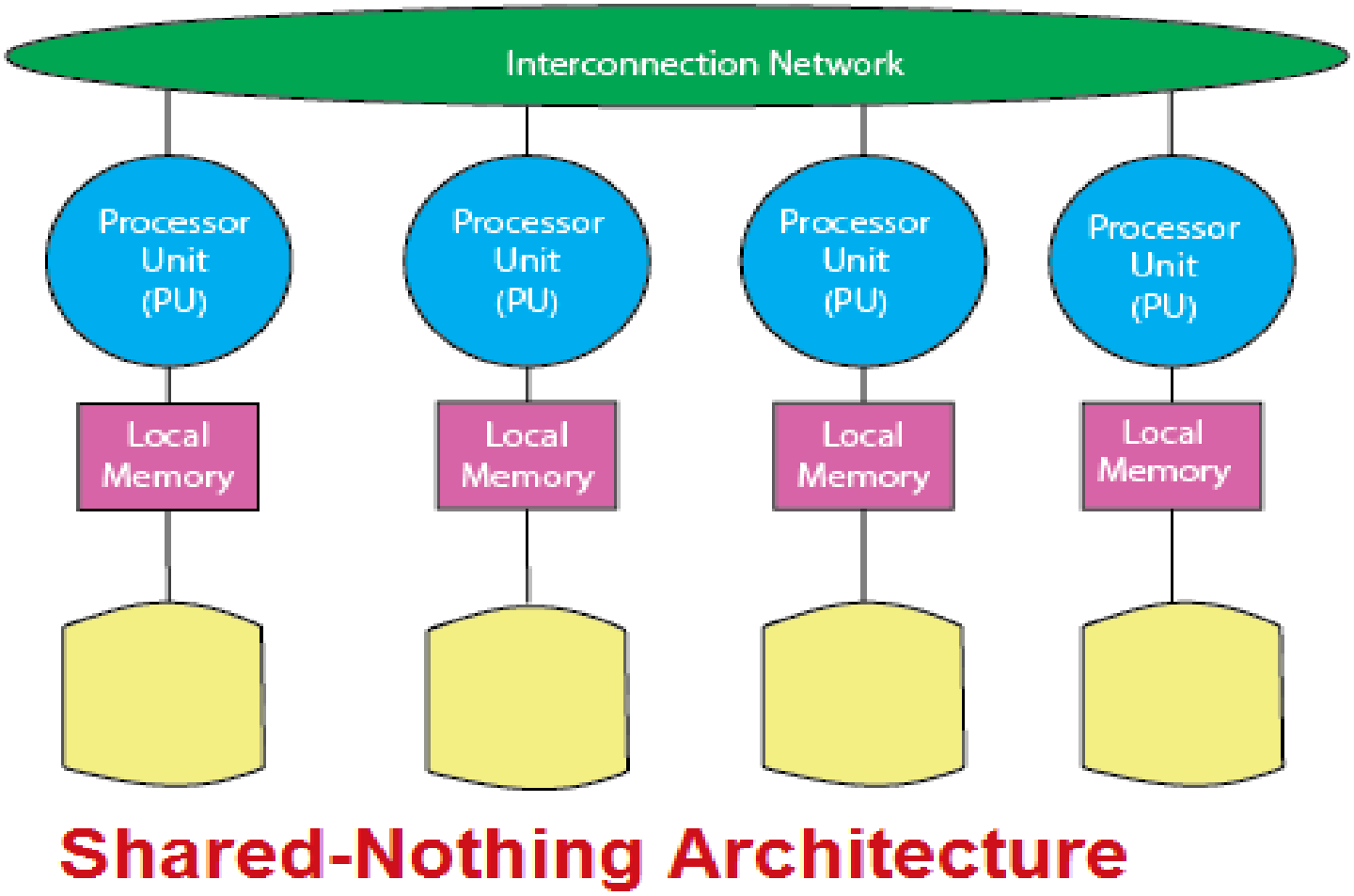


Shared-Memory Architecture

Shared-Nothing Architecture

- In a shared-nothing distributed memory environment, the data is partitioned across all disks, and the DBMS is "partitioned" across multiple co-servers, each of which resides on individual nodes of the parallel system and has an ownership of its disk and thus its database partition.
- A shared-nothing RDBMS parallelizes the execution of a SQL query across multiple processing nodes.
- Each processor has its memory and disk and communicates with other processors by exchanging messages and data over the interconnection network.
- This architecture is optimized specifically for the MPP and cluster systems.
- The shared-nothing architectures offer near-linear scalability. The number of processor nodes is limited only by the hardware platform limitations (and budgetary constraints), and each node itself can be a powerful SMP system.

Shared-Nothing Architecture



Multiple Choice Question

1. Various visualization techniques are used in _____ step of KDD.
a) selection
b) transformaion
c) data mining.
d) interpretation.

2. Extreme values that occur infrequently are called as _____.
a) outliers
b) rare values.
c) dimensionality reduction.
d) All of the above.

3. Box plot and scatter diagram techniques are _____.
a) Graphical
b) Geometric
c) Icon-based.
d) Pixel-based.

4. _____ is used to proceed from very specific knowledge to more general information.
a) Induction
b) Compression.
c) Approximation.
d) Substitution.

5. Describing some characteristics of a set of data by a general model is viewed as _____
a) Induction
b) Compression
c) Approximation
d) Summarization

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, Denniss 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.