22 – Introduction to Distributed Databases

1. System Architecture

    A. Shared Everything

    B. Shared Memory

        i. CPUs access to global memory

        ii. Each CPU should know about other CPU's action

    C. Shared Disk

        i. Each CPU has separate memory, access to global disk

        ii. CPUs must communicate with network to know about other's current state (like update)

    D. Shared Nothing

        i. Each DBMS has separate CPU, memory, disk

        ii. DBMSs must communicate with network

        iii. Hard to increase capacity and ensure capacity, but better performance

2. Design Issues

    A. Homogenous Nodes : all nodes can do same tasks

    B. Heterogenous Nodes : nodes are assigned specific tasks

3. Partitioning Schemes
DBMS executes query fragment on each partition and combines it

   A. Naïve table partitioning

      i. Each table is one partition

      ii. Queries that doesn't communicate between tables are ideal

   B. Horizontal Partitioning

      i. Choose partitioning key and partition table using some partition technique(range partition, hash partition, etc)

      ii. Queries that searching with partitioning key is ideal

      iii. Consistent hashing

         1. Hash space is like ring

         2. Using one direction, nearest node to hash key is selected.
            It is efficient to add node and re-hash
            because just one node is affected by new node.

      iv. Logical vs. Physical

         1. Logical : All records are contained in same disk, but managed by separate nodes. (shared disk)

         2. Physical : all records are contained in different node with separate disk (shared nothing)

4. Distributed Concurrency Control

   A. Centralized coordinator :
      Coordinator determines whether to lock the partition and commit the change of the partition

   B. Decentralized Coordinator :
      Each partition has coordinator⋯

5. What is it?

   A. How Decentralized coordinator work⋯?