13 – Query Execution 2

1. Parallel vs. Distributed
    A. Parallel : Close resources, high speed communication, cheap cost
    B. Distributed : far resources, slow communication, high cost
2. Process Models
    A. Process per DBMS Worker
        i. Relies on OS scheduler
        ii. Process crash doesn't matter for entire system
        iii. Use shared memory
    B. Process Pool
        i. Worker use any free process in a pool
        ii. Relies on OS scheduler and shared memory
        iii. Bad for CPU cache locality
    C. Thread per DBMS Worker
        i. DBMS scheduling
        ii. Thread crash may matter for entire system
3. Execution Parallelism
    A. Inter-query : different queries are executed at same time
        i. Good for improving performance of entire DBMS
        ii. Easy for read-only queries, hard for update queries
    B. Intra-query : execute operations of a single query at same time
        i. Good for improving performance of single query
        ii. Implementation
            1. Intra-operator : decompose operation, process subsets in parallel
            2. Inter-operator : pipeline operation
            3. Bushy : combination of inter and intra operator

4. I/O Parallelism
    A. Split DBMS to multiple devices
    B. Multi-Disk Parallelism
        i. Same as RAID (learned from OS class)
    C. Partitioning
        i. Split table into some segments
        ii. Horizontal (row partitioning) / Vertical (column partitioning)