13 – Query Processing

1. Issues of Query Execution

   A. Problem
      Dependencies and Branch Prediction :
      known as pipelining hazard
      branchless implementation is good, since it doesn't have branching hazard

   B. Excessive Instructions

      i. DBMS should support different data types
         So, DBMS should check the type of data before perform some operation on that.

2. Processing Models

   A. Iterator Model

      i. == Volcano , Pipeline Model

      ii. All query plan operator implements "next" function

      iii. Each operator emits output of single tuple

      iv. Top-Down approach of query processing

   B. Materialization Model

      i. All operator processes all input and flush output all

      ii. Can push down selection operator to reduce scanning input

      iii. Good for OLTP, but not good for OLAP

   C. Vectorized / Batch Model

      i. All query plan operator implements "next" function

      ii. Each operator emits output of some batch of tuples
         (hybrid of iterator model and materialization model)

      iii. Ideal for OLAP queries

      iv. Allow to use SIMD instruction to process on batch of tuples

3. Parallel Execution

    A. Inter query paralellism

        i. Allow multiple queries to execute simultaneously

    B. Intra-query parallelism

        i. Allow multiple operator in same query to execute simultaneously

        ii. Decompose operator into independent operator instances to perform same operation on different data
        after performing operation complete, merge the result set of operator instances to make a total result set of original query plan operator