12 – Query Execution 1

1. Processing Models

   A. Iterator model
      == Volcano model == Pipeline model

      i. Each query plan operator implements “next function”
         (maybe it is like stream or thunk that I learned at PL class)

      ii. Easy to control output
         But, some query cannot use this model (like ORDER BY, subqueries, etc)

   B. Materialization model

      i. Each operator processes input and output all at once

      ii. Good for OLTP : queries only access a small number of tuples at a time
         Bad for OLAP : too large results

   C. Vectorized/Batch model

      i. Each query plan operator implements “next function” like iterator model

      ii. Output batch of records like Materialization model

      iii. Good for OLAP : reduces invocations for operator

2. Access Methods
    A. Sequential Scan
        i. Optimization
            1. prefetching, buffer bypass, parallelization
            2. zone maps
               pre-compute aggregates for the attribute values
               use these values to check whether page is needed or not
            3. late materializations
            4. heap clustering
               make index to be clustered

    B. Index Scan
        i. Type of index varies on data and query's property
        ii. Index scan page sorting
            1. Sort needed page based on page id to minimize buffer miss rate

    C. Multi-Index / Bitmap Scan
        i. Use each predicate to make some groups,
           check intersection to find actual result.

3. Expression Evaluation
    A. WHERE clause is represented as an expression tree
    B. Expression tree is flexible but slow

4. What is it?
    A. How late materialization work?