* Column-store systems completely vertically partition a database into a collection of individual columns that are stored separately, which enable queries to read just the attributes they need, rather than having to read entire rows from disk and discard unneeded attributes once they are in memory.
* Late materialization (late tuple reconstruction) refers to delaying the joining of columns into wider tuples. A select operator scans a single column at a time with a tight for-loop, resulting in cache and CPU friendly patterns.
* The advantages of late materialization: selection and aggregation operators tend to render the construction of some tuples unnecessary (avoid the overhead of constructing it altogether), cache performance is improved when operating directly on column data, the vectorized optimizations described above have a higher impact on performance for fixed-length attributes.
* A complete sort is not necessary to reduce random access performance overhead in value extraction of join output, since most storage media are divided into contiguous blocks of storage, and random access within a block is significantly cheaper than random access across blocks. Therefore, the database does not need to completely sort the position list before using it to extract values from columns; rather, it just needs to be partitioned into the blocks on storage in which those positions can be found.
* The flexibility of vertices message passing interface is that, a vertex can send messages to any other vertex that it can keep track of, not just its neighbors, which allows many of these systems to implement efficient parallel algorithms that perform pointer jumping like in their PRAM algorithm counterparts
* We consider single-machine out-of-core systems as performing synchronous execution, since each iteration needs to stream the whole graph on disk(s); but this is different from Pregel’s model where a vertex always receives a value from the last iteration.
* In a single-machine out-of-core systems, a vertex may be accessing the value of another vertex that is written to disk when processing a previous shard in the same iteration, and thus the total number of iterations can be less than that of a Pregel algorithm
* Single-machine out-of-core systems may not be used for algorithms running for many iterations but only a small number of vertices compute in each iteration, since the entire graph needs to be streamed again and again.