REPORT Project 0

Sara Djambazovska

In this first part of the Database Systems Project, I implemented the six basic operators, in a volcano-style tuple-at-a-time operators: ***Aggregate, Filter, Join, Project, Scan*** and ***Sort*** in their respective *.scala* classes in the package ch.epfl.dias.cs422.rel.early.volcano.

The implementations are quite straight forward.

For Project, given the evaluator, the projection is obtained by only passing the input tuples through it.

I did a small optimization in Scan, where I stored the values from the scanned (scannable) RowStore into an indexed sequence, and only accessed that storage in the calls to next().

Filter is quite simply implemented as well, with next() only returning the next input tuple that satisfies the predicate.

Aggregate is implemented by using the suggested algorithm in the class description, with a few storage optimizations - using a map to store the groupBy result, and doing the whole processing in the open() method only, so that any call to next is very efficient.

All the processing (sorting) in the Sort class is done in the open() method as well, which fills up a sorted sequence given the Field Collations.

To have better performance, I used a HashJoin implementation for the Join operator, using an internal hash table.

The performance for the Query Test with my implementation exceeded the baseline implementation, with a speedup of around 1.06 times.