No clear heuristic for JIT code generation

Just In Time compilation is a key technology in column stores

Infinite Possibilities

One solution does not fit all

No clear conclusion given the various setups

Design Space for SELECT: > 1 Strategies

4 Loops Strategy

1. inter = select(R.a, max1)
2. inter = select_fetch(R.b, max2, inter)
3. inter = select_fetch(R.c, max3, inter)
4. dest = select_fetch(R.d, max4, inter)

1 Loop Strategy

dest = select(R.a, max1, R.b, max2, R.c, max3, R.d, max4)

branching
vectorization
loop unrolling

One loop does not fit all

With Branches

Without Branches

One loop fits all, but … what about robustness?

Using separate loops for each predicate results in scanning fewer cache lines when the first filter selects very few tuples.

For low selectivity and a single loop we are reading data that is not going to be used in the branchless case.

Future Work

Disjunctive Selects
Alternate Distributions
More complex queries
Correlations