

ALPHA 5 COMPILATION

OLIVER KENNEDY

1. COMPILER WORK LOOP

The compiler maintains a work queue datastructure. As long as the work queue is non-empty, the compiler will perform the following steps on whatever expression is at the head of the work queue¹:

- (1) **Prematerialize the expression.** This stage simplifies the expression, factorizes it, and computes a list of all lifted subexpressions with non-zero deltas. Later iterations of this stage will split the factorized expression into multiple subexpressions that can have their deltas computed individually.
- (2) **Compute Deltas.** For each *stream* appearing in the expression, apply the delta operation to the expression to get a set of delta expressions.
- (3) **Materialize the delta expressions.** This involves pulling out all the relations in the expression and replacing them with externals. The lifted subexpressions identified in the prematerialization phase are used as a guide in doing so. Additionally, maps that have already been added to the work queue or compiled are re-used if appropriate.
- (4) **Update the work queue.** Newly instantiated datastructures are added to the work queue. The recently compiled datastructure is also stored.

1.1. **Prematerialization.** Simplification is the application of a set of rewrite rules that provides the following guarantees:

- (1) Values appearing in sum and product terms are merged together and constants are combined/curried.
- (2) Lifts and AggSums are un-nested as far as possible: Expressions that always evaluate to 0 or 1 (comparisons, lifts, and keyed relations) can always be pulled out of a lifted product. If all of the output variables of an expression are group-by variables of an AggSum of a product that the expression appears in, the expression can be lifted out. These variables can then be removed from the AggSum's group-by variable list.
- (3) AggSums are factorized
- (4) Unnecessary AggSums are eliminated if the aggregated expression's output variables are all GB vars in the aggsun. Also, any GB vars that are not output variables of the aggregated expression.

¹This section assumes that the head is a standard map datastructure. If not, then we might do something more interesting. We'll find out once we actually start supporting non-map datastructures.

2. DOMAIN MAINTENANCE