The PACT Programming Model

Parallelization Contracts (PACTs)

1. MAP
2. REDUCE
3. CROSS
4. MATCH
5. COGROUP

The PACT Optimizer

Optimizing PACT Jobs

1. Generate Estimates
2. Generate Interesting Properties
3. Enumeration & Pruning

The Nephele Execution Engine

Executing PACT Jobs

Example Task: 3rd TPC-H Query

SELECT l_orderkey, o_shippriority, 
sum(l_extendedprice) as revenue 
FROM orders, lineitem 
WHERE o_orderkey = l_orderkey 
AND o_custkey IN [x] 
AND o_orderdate = [y] 
GROUP BY l_orderkey, o_shippriority

Example Task: Alternative 1

Example Task: Alternative 2

Overview & Context

The Stratosphere prototype integrates:

- The PACT Programming Model: A generalization & extension of the Map/Reduce programming model
- The PACT Optimizer: A RDBMS-inspired optimizer for schema-free parallel dataflows
- The Nephele Execution Engine: A parallel datalflow execution engine

Stratosphere is funded as DFG research unit FOR1306 with additional support from HP and IBM.