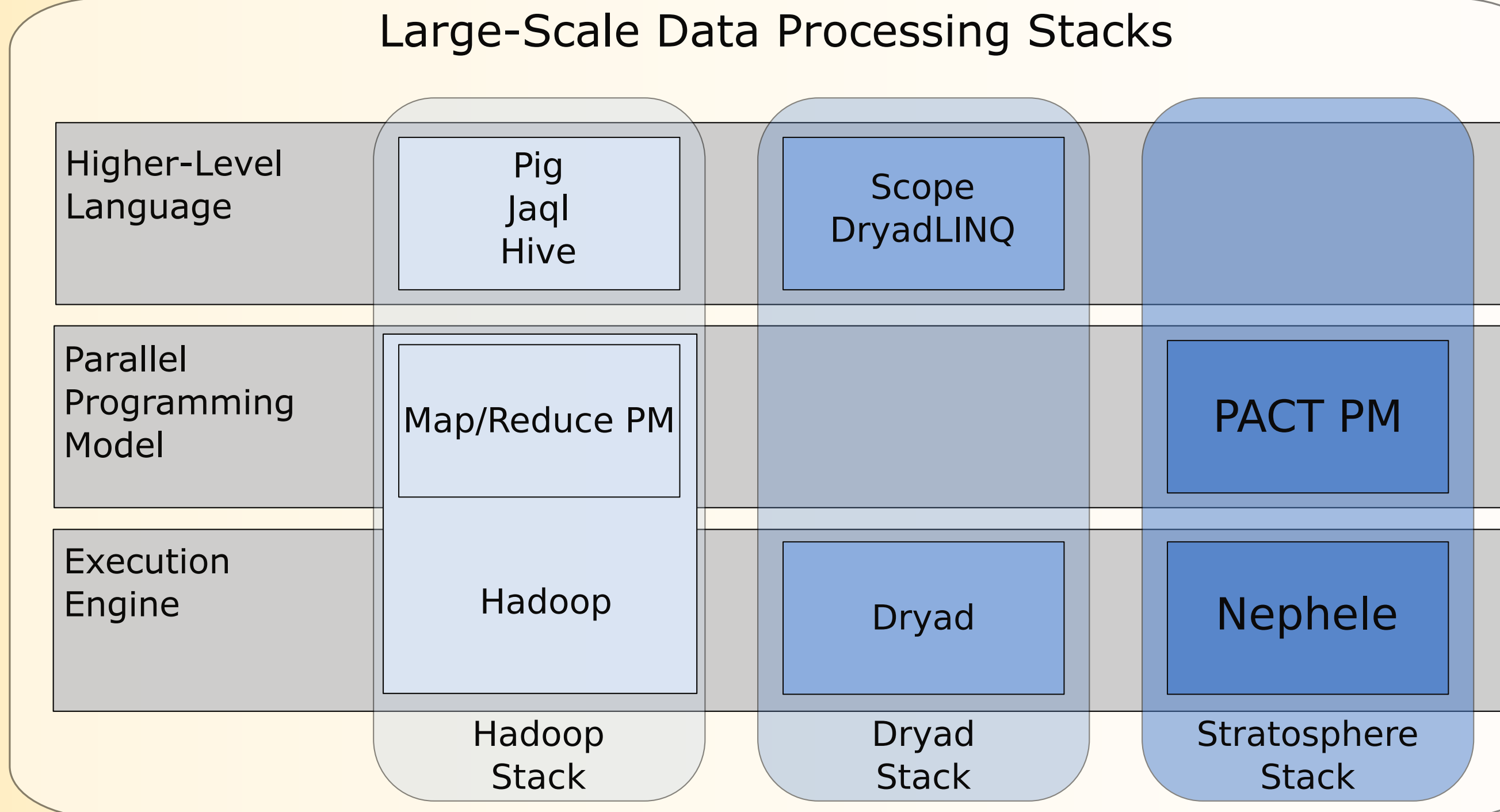


Overview & Context

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 Nephelē Execution Engine:**
A parallel dataflow execution engine

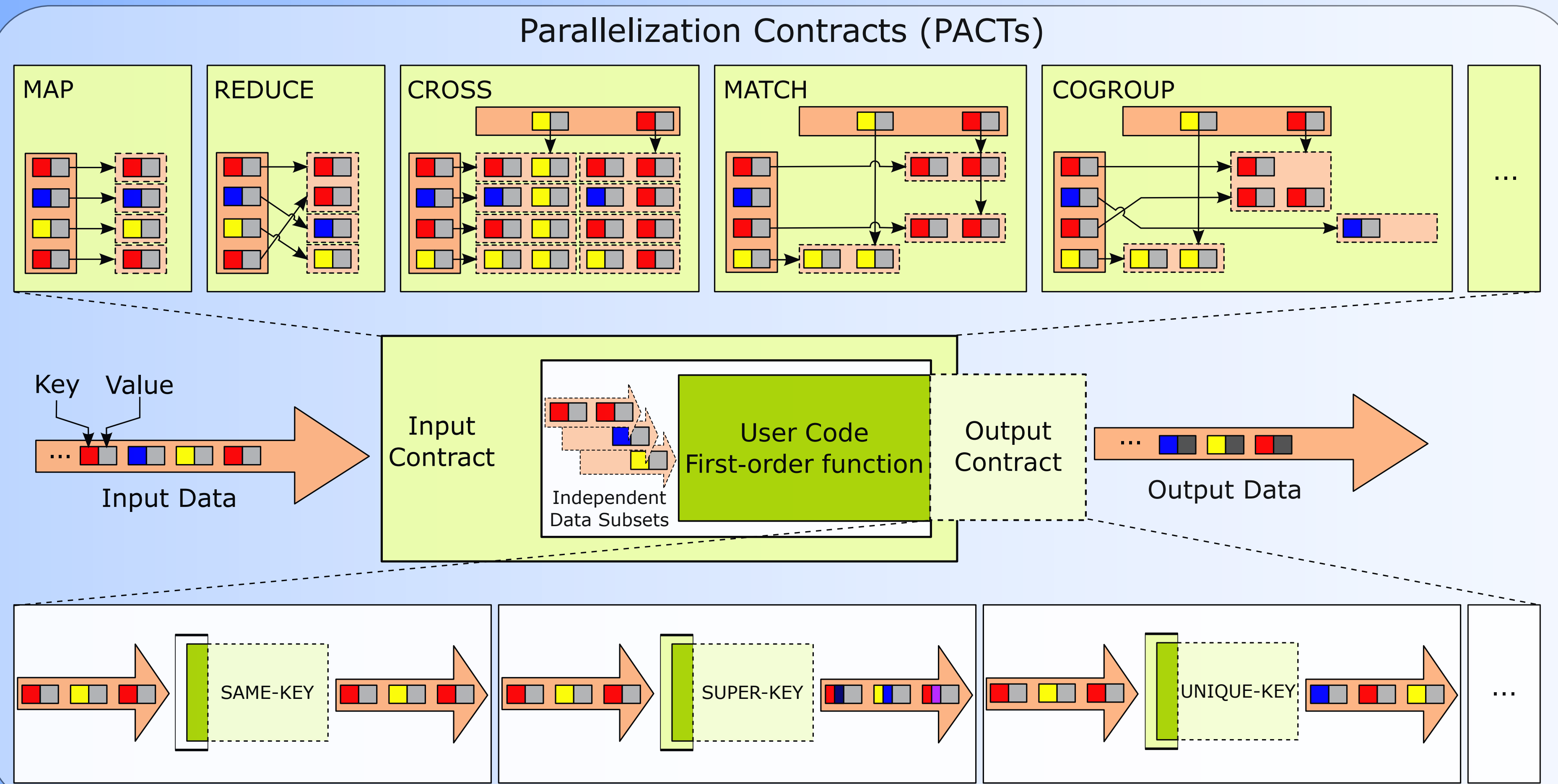


The Stratosphere Project

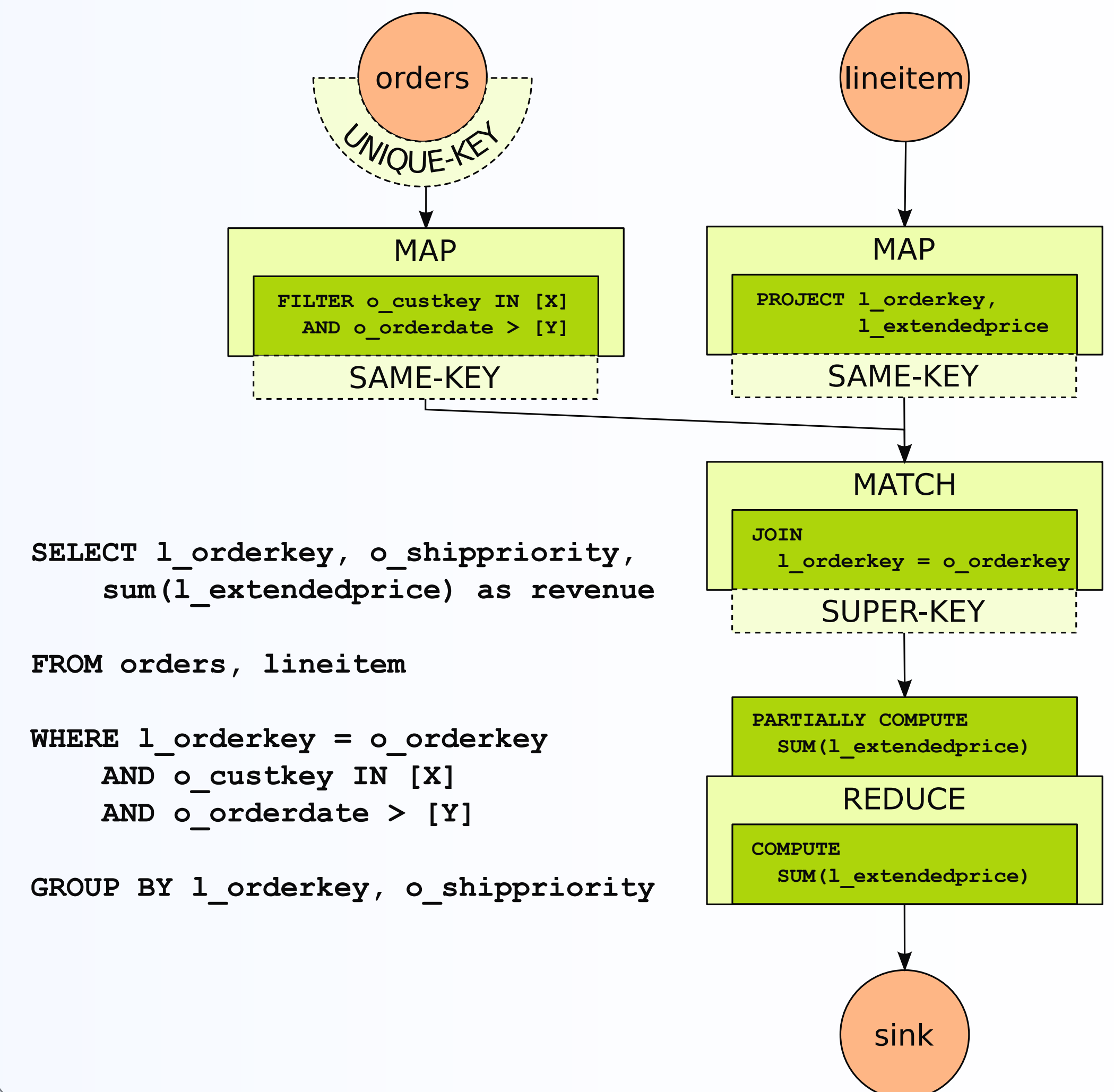
- Joint project of TU Berlin, HU Berlin, and HPI Potsdam
- Explores the power of Cloud computing for complex information management applications
- Researches and prototypes a web-scale data analytics infrastructure
- Processes textual and (semi-)structured data
- Follows a database-inspired approach

The PACT Programming Model

Writing PACT Jobs

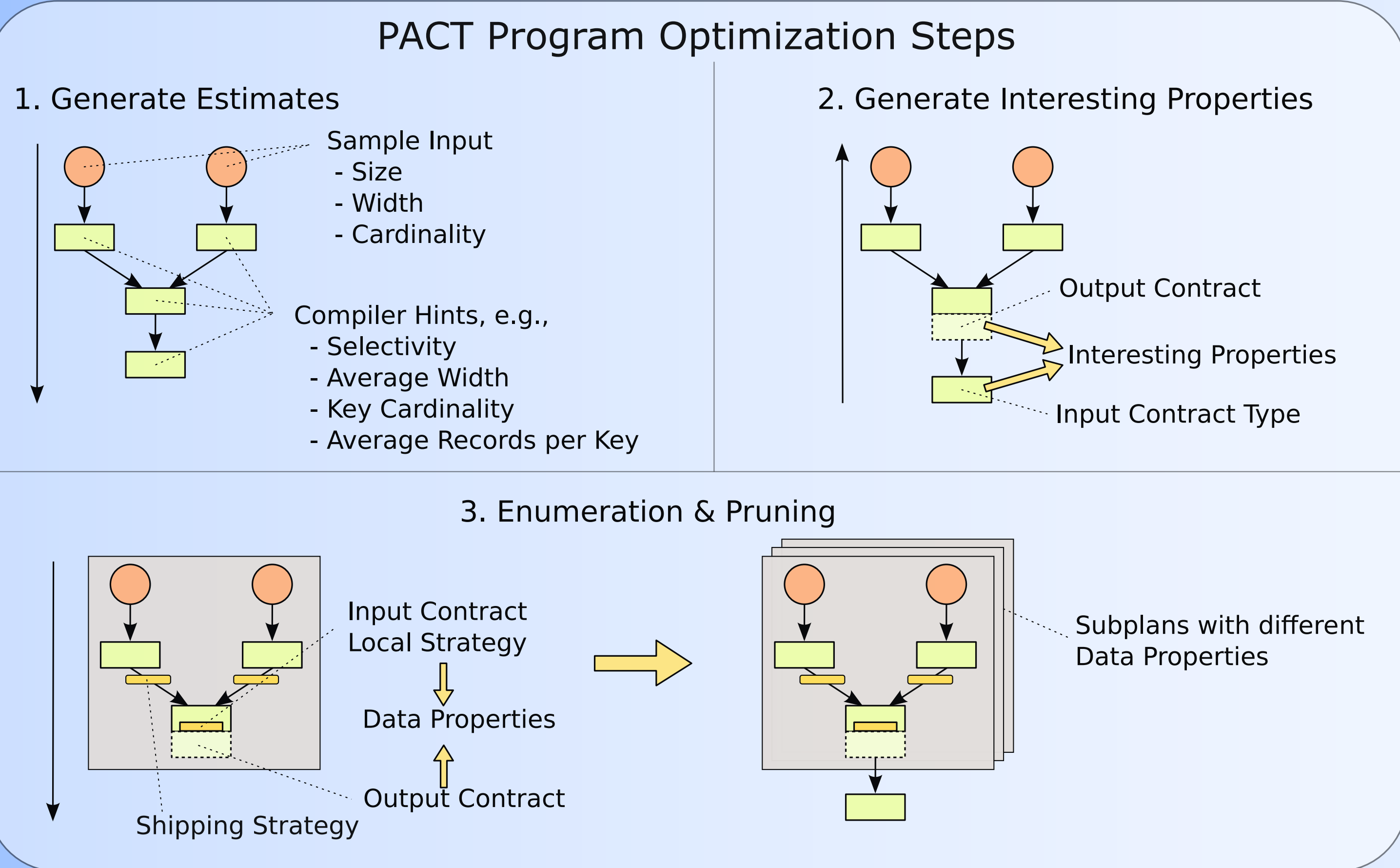


Example Task: 3rd TPC-H Query

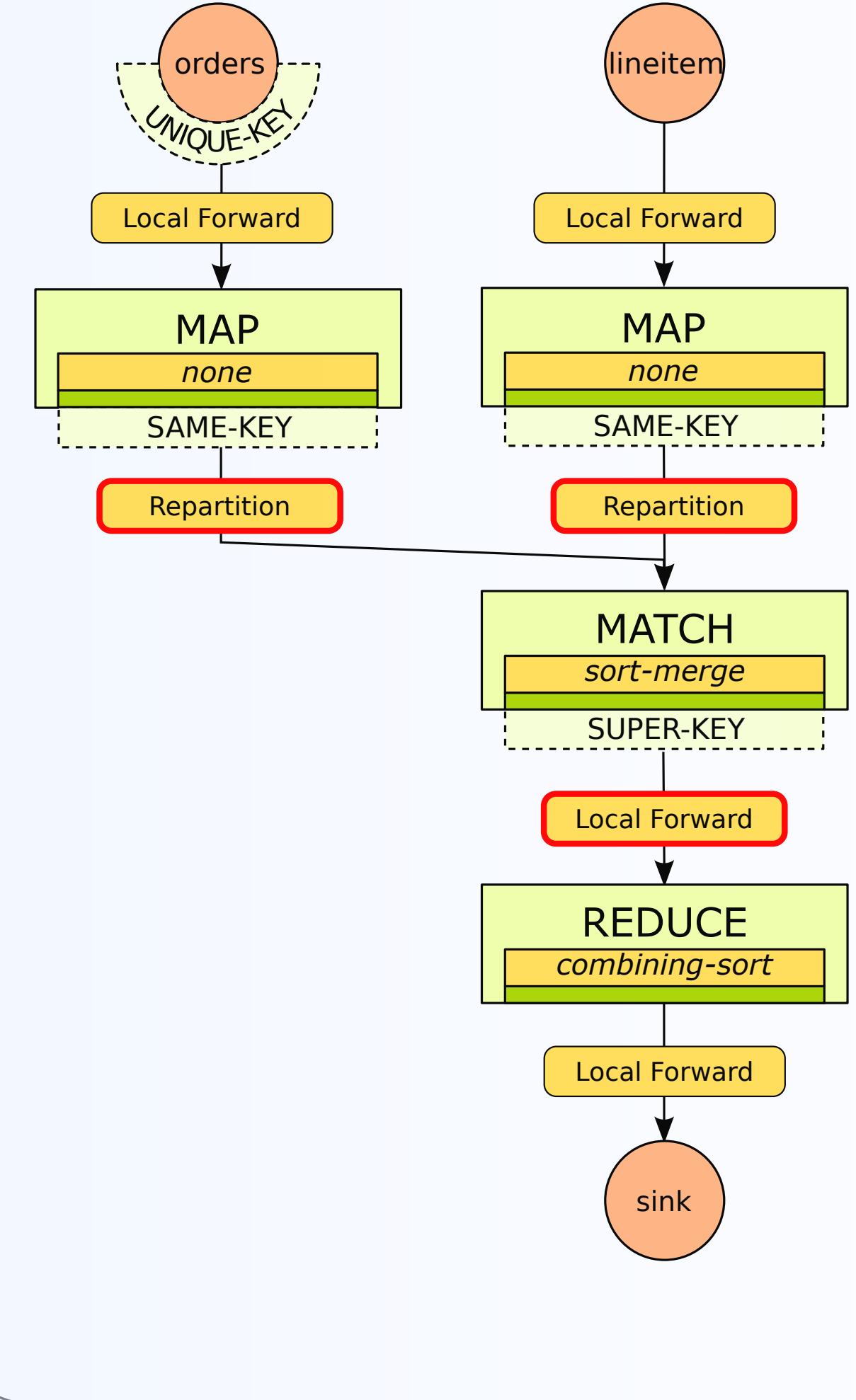


The PACT Optimizer

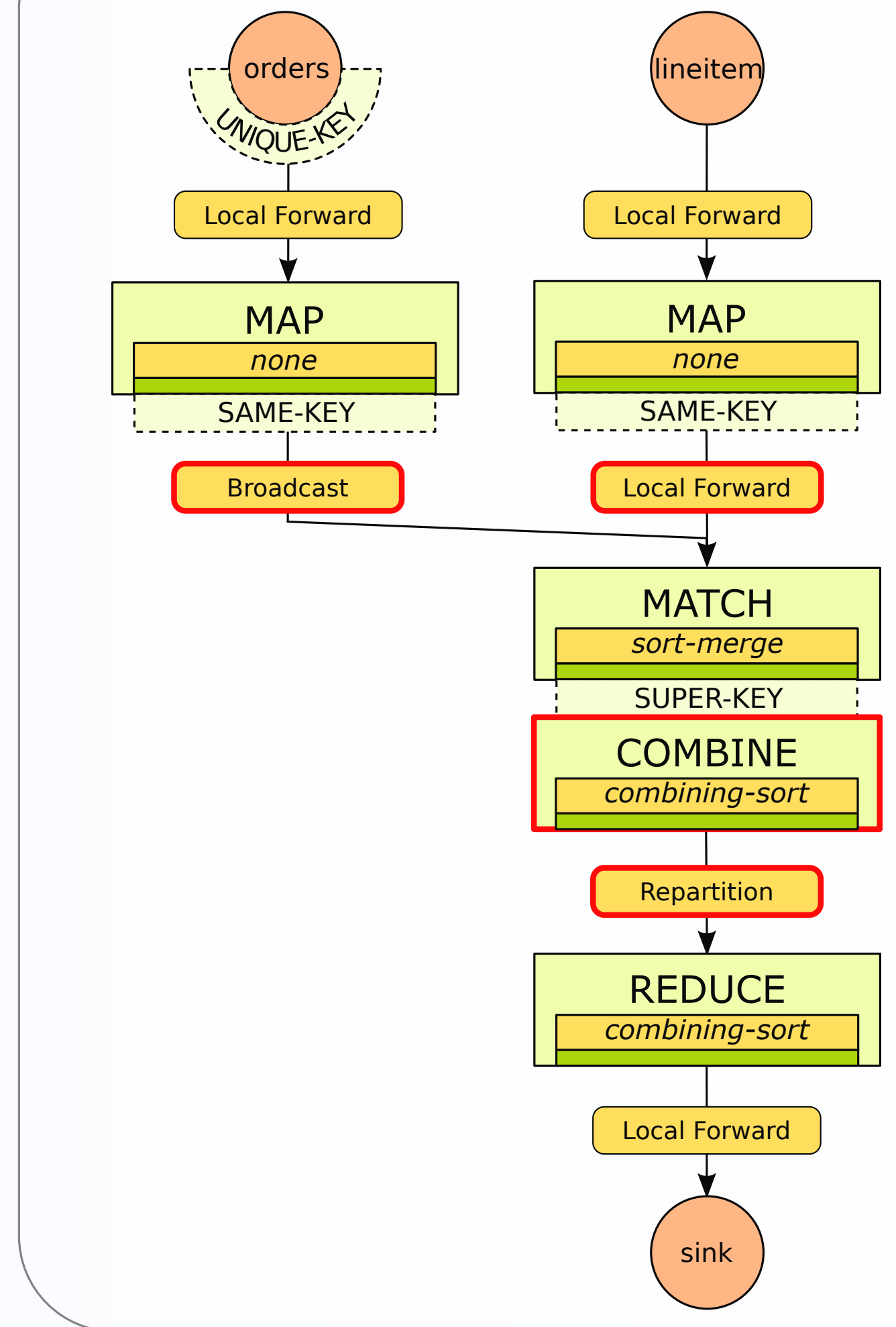
Optimizing PACT Jobs



Example Task: Alternative 1

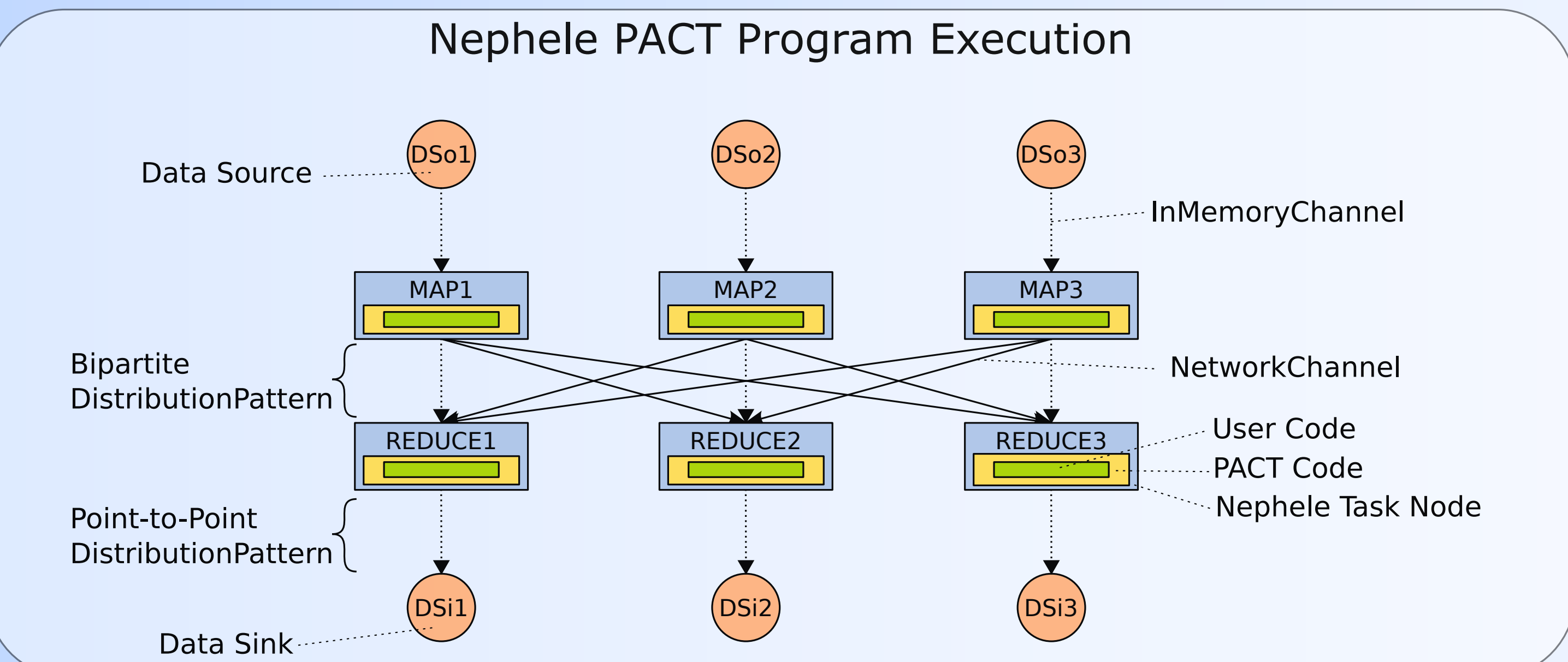


Example Task: Alternative 2

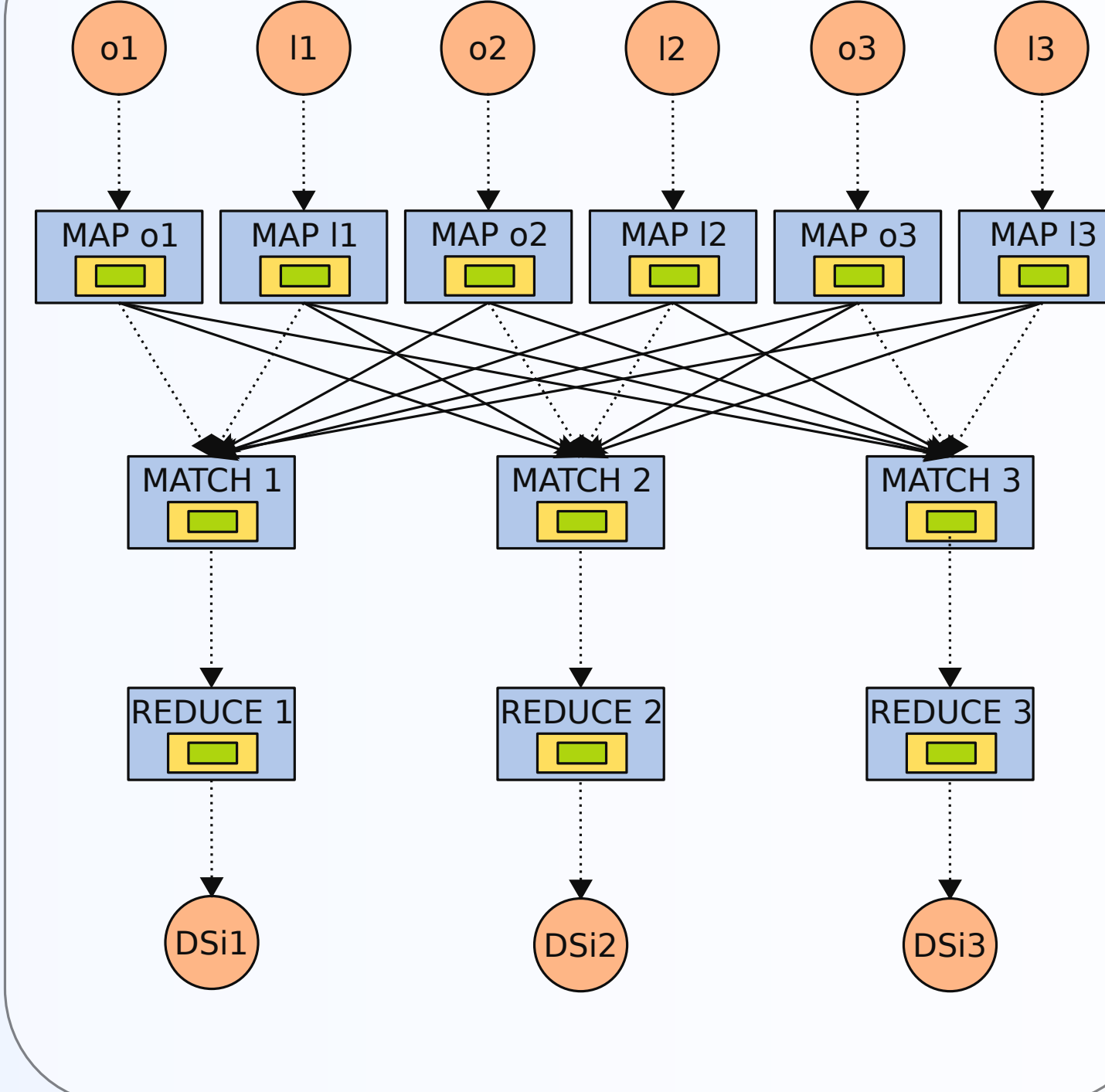


The Nephelē Execution Engine

Executing PACT Jobs



Execution Alternative 1



Execution Alternative 2

