CoAI: Incremental requirement-driven design and deployment of data intensive flows

Petar Jovanovic, Oscar Romero, Alberto Abelló
Universitat Politècnica de Catalunya, BarcelonaTech
petar | oromero | aabell}@essi.upc.edu

Alikis Simitsis
HP Labs, Palo Alto, CA, USA
alkis@hp.com

Problem under study
- Incrementally integrating individual data flows into a unified flow satisfying the entire set of inf. requirements
- Maximizing the reuse of the existing data flows
- Lowering overall execution time by sharing data and computation
- Considering execution costs

Defining information requirements

Challenges & Solution
1) Incremental advancement
- To guarantee semantical overlapping when comparing two operations
- We must ensure that all the predeceasing operations of both operations coincide
- Bottom-up (sources-to-target) algorithm for consolidating data flows

2) Operation reordering
- Swap, distribute/factorize, merge/split, association
- Generic equivalence rules to guarantee the equivalence of reordering transformations
- In terms of properties of data flow operations (i.e., Schema, Value, Order)

α = (S, V, O, Post)
Pre = (S_{pre}, V_{pre}, O_{pre})
Post = (S_{post-gen}, S_{post-rem}, V_{post}, O_{post})

- Check for the conflicts
  ✓ Schema
  ✓ Value
  ✓ Order

3) Operation comparison
- Generic operation comparison
- Full and partial operation matching

Further reading: