

Incremental Consolidation of Data-Intensive Multi-flows

Petar Jovanovic¹, Oscar Romero¹, Alkis Simitsis², and Alberto Abelló¹

¹ Universitat Politècnica de Catalunya (BarcelonaTech), Barcelona, Spain,
{petar|oromero|aabello}@essi.upc.edu,

<http://www.essi.upc.edu/dtim>

² HP Labs, Palo Alto, CA, USA,
alkis@hp.com

Abstract. Business intelligence (BI) systems depend on efficient integration of disparate and often heterogeneous data. The integration of data is governed by *data-intensive* flows and is driven by a set of information requirements. Designing such flows is in general a complex process, which due to the complexity of business environments is hard to be done manually. In this chapter, we deal with the challenge of efficient design and maintenance of *data-intensive* flows and propose an incremental approach, namely *CoAL*, for semi-automatically consolidating data-intensive flows satisfying a given set of information requirements. *CoAL* works at the logical level and consolidates data flows from either high-level information requirements or platform-specific programs. As *CoAL* integrates a new data flow, it opts for maximal reuse of existing flows and applies a customizable cost model tuned for minimizing the overall cost of a unified solution. We demonstrate the efficiency and effectiveness of our approach through an experimental evaluation using our implemented prototype.

Keywords: Business Intelligence, data-intensive flows, workflow management, data warehousing

The paper has been published in *IEEE Transactions on Knowledge and Data Engineering*, 28(5): pp. 1203-1216 (2016).

DOI: <http://dx.doi.org/10.1109/TKDE.2016.2515609>