

## Accepted Manuscript

Practical algorithms for execution engine selection in data flows

Georgia Kougka, Anastasios Gounaris, Kostas Tsichlas

PII: S0167-739X(14)00245-3

DOI: <http://dx.doi.org/10.1016/j.future.2014.11.011>

Reference: FUTURE 2667

To appear in: *Future Generation Computer Systems*

Received date: 8 May 2014

Revised date: 25 September 2014

Accepted date: 10 November 2014

Please cite this article as: G. Kougka, A. Gounaris, K. Tsichlas, Practical algorithms for execution engine selection in data flows, *Future Generation Computer Systems* (2014), <http://dx.doi.org/10.1016/j.future.2014.11.011>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Practical Algorithms for Execution Engine Selection in Data Flows

Georgia Kougka

*Aristotle University of Thessaloniki*

Anastasios Gounaris

*Aristotle University of Thessaloniki*

Kostas Tsichlas

*Aristotle University of Thessaloniki*

---

## Abstract

Data-intensive flows are increasingly encountered in various settings, including business intelligence and scientific scenarios. At the same time, flow technology is evolving. Instead of resorting to monolithic solutions, current approaches tend to employ multiple execution engines, such as Hadoop clusters, traditional DBMSs, and stand-alone tools. We target the problem of allocating flow activities to specific heterogeneous and interdependent execution engines while minimizing the flow execution cost. To date, the state-of-the-art is limited to simple heuristics. Although the problem is intractable, we propose practical anytime solutions that are capable of outperforming those simple heuristics and yielding allocation plans in seconds even when optimizing large flows on ordinary machines. Moreover, we prove the NP-hardness of the problem in the generic case and we propose an exact polynomial solution for a specific form of flows, namely, linear flows. We thoroughly evaluate our solutions in both real-world and flows synthetic, and the results show the superiority of our solutions. Especially in real-world scenarios, we can decrease execution time up to more than 3 times.

---

*Email addresses:* `georkoug@csd.auth.gr` (Georgia Kougka),  
`gounaria@csd.auth.gr` (Anastasios Gounaris), `tsichlas@csd.auth.gr` (Kostas Tsichlas)

Download English Version:

<https://daneshyari.com/en/article/6873558>

Download Persian Version:

<https://daneshyari.com/article/6873558>

[Daneshyari.com](https://daneshyari.com)