Big Data Analytics: From SQL to Hadoop and Beyond

Nicolas Spyratos*†1

¹Université Paris-Sud - Paris 11 (UP11) – Université Paris-Sud - Paris 11 – Bâtiment 300 - 91405 Orsay cedex, France

Résumé

We present a high level query language, called HiFun, for defining analytic queries over big data sets. An analytic query in HiFun is defined to be a well-formed expression of a functional algebra, whose operations combine functions to create HiFun queries (in much the same way as the operations of the relational algebra combine relations to create relational algebra queries). We show that a HiFun query can be encoded as a map-reduce job, and also as a SQL group-by query when the data set resides in a relational database. We also present a formal method for rewriting HiFun queries and defining query execution plans. As a case study, we show how the rewriting method for HiFun queries can be applied in the rewriting of map-reduce jobs and SQL group-by queries. **Short bio**: *Nicolas Spyratos* is currently professor emeritus at the University of Paris South, scientific advisor of the Japan Science and Technology agency (JST), member of the Greek National Council of Research and Innovation and Adjunct Senior Researcher at the FORTH Institute of Computer Science in Greece. His research interests include databases, big data analytics, digital libraries and conceptual modeling. He has published over 200 papers in refereed international journals and conferences and has participated in over 20 European and international research projects. He has supervised 24 doctoral theses and has been evaluator for the European programs Esprit and Esprit-Bra as well as for the National Science Foundation (NSF) and major scientific journals.

---

*Intervenant

†Auteur correspondant: Nicolas.Spyratos@lri.fr