

Jenga and the Art of Data-Intensive Ecosystems Maintenance

Panos Vassiliadis*

*Technische Universität Berlin, Germany

Software maintenance amounts up to 60% of the resources spent on building and operating a software system. Data-intensive ecosystems that include several software applications tightly coupled to underlying data repositories cannot escape the above rule. In such environments, the impact of evolution is three-fold: (a) syntactical (meaning, that the evolution of either the data repositories or the software that implements the operational processes of the data warehouse can lead to operational failures and crashes due to some form of syntactic incorrectness), (b) semantic (meaning that changes in a view or a data transformation module can lead to different semantics for the propagated data), and (c) performance-oriented, as different configurations of the ecosystem's components (be it data or software) lead to different performance for its operations.

As in all data-intensive ecosystems, the evolution of the software constructs of a data warehouse environment can severely affect its operations. In this talk, we focus on two aspects of the management of data warehouse evolution. On the one hand, we are interested in predicting the maintenance effort of ETL workflows. To this end, we present the findings of a case study on how a set of graph-theoretic metrics can be used for the prediction of evolution vulnerability for the components of ETL scenarios. On the other hand, we are interested in supporting the graceful evolution of the ecosystem's components and we present a method for the adaptation of ecosystems that assess the potential impact of a change and rewrite the ecosystem's components in order to adapt to the change.

Panos Vassiliadis graduated the Varvakio Experimental School in 1990 and obtained his Diploma in Electrical Engineering and his PhD from the Department of Electrical and Computer Engineering, of the National Technical University of Athens (NTUA) in 1995 and 2000, respectively. He has joined the Department of Computer Science of the University of Ioannina in 2002 and since then, He is a member of the Distributed Management of Data (DMOD) Laboratory. So far, his research has focused on Data Warehouse technology, with particular interest on issues like data warehouse metadata repositories and metadata modeling, data warehouse quality, On-Line Analytical Processing (OLAP), and Extraction-Transformation-Loading (ETL). Currently, his on-going research is also targeted towards Metadata-Rich Data-Centric Information Systems, with particular emphasis to the modeling, pattern-based design, and evolution of their underlying database infrastructure as well as to Web-Services with particular emphasis to SOA maintenance.