

Scalable graph analytics with GRADOOP

[Abstract]

Erhard Rahm
University of Leipzig
Augustusplatz 10
Leipzig, Germany
rahm@informatik.uni-leipzig.de

ABSTRACT

Many Big Data applications in business and science require the management and analysis of huge amounts of graph data. Previous approaches for graph analytics such as graph databases and parallel graph processing systems (e.g., Pregel) either lack sufficient scalability or flexibility and expressiveness. We are therefore developing a new end-to-end approach for graph data management and analysis at the Big Data center of excellence ScaDS Dresden/Leipzig. The system is called Gradoop (Graph analytics on Hadoop). Gradoop is designed around the so-called Extended Property Graph Data Model (EPGM) which supports semantically rich, schema-free graph data within many distinct graphs. A set of high-level operators is provided for analyzing both single graphs and sets of graphs. The operators are usable within a domain-specific language to define and run data integration workflows (for integrating heterogeneous source data into the Gradoop graph store) as well as analysis workflows. The Gradoop data store is currently utilizing HBase for distributed storage of graph data in Hadoop clusters. An initial version of Gradoop is operational and has been used for analyzing graph data for business intelligence and social network analysis.

About the Author

Erhard Rahm is full professor for databases at the computer science institute of the University of Leipzig. His current research focusses on Big Data and data integration. He has authored several books and more than 200 peer-reviewed journal and conference publications. His research on data integration and schema matching has been awarded several times, in particular with the renowned 10-year best-paper award of the conference series VLDB (Very Large Databases) and the Influential Paper Award of the conference series ICDE (Int. conf. on Data Engineering). Prof. Rahm is one of the two scientific coordinators of the new German center of excellence on Big Data ScaDS (competence center for SCALable Data services and Solutions) Dresden/Leipzig that started its operation in Oct. 2014.