

Scalable graph analytics with GRADOOP

Erhard Rahm
University of Leipzig
rahm@informatik.uni-leipzig.de

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.

References

- Junghanns, M., Petermann, A., Gomez, K., Rahm, E.: *GRADOOP - Scalable Graph Data Management and Analytics with Hadoop*. Tech. report, Univ. of Leipzig, June 2015
- Petermann, A.; Junghanns, M.; Müller, R.; Rahm, E.: *BIIIG - Enabling Business Intelligence with Integrated Instance Graphs*. Proc. 5th ICDE Workshop on Graph Data Management (GDM) 2014
- Petermann, A.; Junghanns, M.; Müller, R.; Rahm, E.: *Graph-based Data Integration and Business Intelligence with BIIIG*. Proc. of the VLDB Endowment (PVLDB), Vol. 7, No. 13, 2014
- Petermann, A.; Junghanns, M.; Müller, R.; Rahm, E.: *FoodBroker - Generating Synthetic Datasets for Graph-Based Business Analytics*. Proc. 5th Int. Workshop on Big Data Benchmarking (WBDB), 2014
- Rahm, E., Nagel, W.: *ScaDS Dresden/Leipzig - Ein serviceorientiertes Kompetenzzentrum für Big Data*. Proc. GI-Jahrestagung, LNI, 2014