

Big Data Management and Apache Flink: Key Challenges and (Some) Solutions

[Abstract]

Volker Markl
TU Berlin and DFKI, Germany
volker.markl@tu-berlin.de

ABSTRACT

The shortage of qualified data scientists is effectively limiting Big Data from fully realizing its potential to deliver *insight* and provide *value* for scientists, business analysts, and society as a whole. In order to remedy this situation, we believe that novel technologies that draw on the concepts of declarative languages, query optimization, automatic parallelization and hardware adaptation are necessary. In this talk, we will discuss several aspects of our research in this area, including results in how to optimize iterative data flow programs, optimistic fault-tolerance, and steps toward a deep language embedding of advanced data analysis programs. We will also discuss how our research activities have led to Apache Flink, an open-source big data analytics system, which by now has become a major data processing engine in the Apache Big Data Stack, used in a variety of applications by academia and industry.

About the Author

Volker Markl is a Full Professor and Chair of the Database Systems and Information Management (DIMA) group at the Technische Universität Berlin (TU Berlin), director of the research group "Intelligent Analysis of Massive Data" at the German Research Center for Artificial Intelligence (DFKI), and speaker of the Berlin Big Data Center (BBDC). Earlier in his career, Dr. Markl led a research group at FORWISS, the Bavarian Research Center for Knowledge-based Systems in Munich, Germany, and was a Research Staff member & Project Leader at the IBM Almaden Research Center in San Jose, California, USA. Dr. Markl has published numerous research papers on indexing, query optimization, lightweight information integration, and scalable data processing. He holds 19 patents, has transferred technology into several commercial products, and advises several companies and startups. He has been speaker and principal investigator of the Stratosphere research project that resulted in the "Apache Flink" big data analytics system. Dr. Markl currently serves as the secretary of the VLDB Endowment and was elected as one of Germany's leading "digital minds" (Digitale Köpfe) by the German Informatics Society (GI). Volker Markl and his team earned an ACM SIGMOD Research Highlight Award 2016 for their work on implicit parallelism through deep language embedding.