Druid: Open Source Infrastructure for Exploratory Analytics

Mr. Fangjin Yang
Metamarkets

Abstract:
This talk will focus on Druid, open source infrastructure to ingest and analyze transactional data. Druid is designed for low latency ingestion and aggregation of log data, and has a distributed, column-oriented, shared-nothing architecture. We will cover the limitations with existing relational database management systems and NoSQL key/value stores that promoted Druid’s development, the architecture of the system, and where Druid sits in the data engineering space. We will also discuss the emergence of a "real-time" analytics stack at major technology companies and data focused startups, and what that stack looks at Google, Facebook, Yahoo, LinkedIn, Amazon, and in the open source world.

About the Speaker:
Fangjin is one of the main contributors to the Druid project and one first developers at Metamarkets, a San Francisco based data startup. He's deeply interested in databases, both in industry and academia. Fangjin previously worked on diagnostic optimization algorithms at Cisco Systems. He holds a BASc in Electrical Engineering and a MASc in Computer Engineering from the University of Waterloo, Canada.