Impala Reading List

Introduction to (Impala-like) database systems

- Architecture of a Database System (Hellerstein, Stonebraker, Hamilton)
- Everything You Always Wanted to Know About Compiled and Vectorized Queries But Were Afraid to Ask
- Encapsulation of Parallelism in the Volcano Query Processing System (Goetz Graefe)
- SCOPE: Parallel Databases Meet MapReduce (Zhou et. al.)
- Parallel Database Systems: The Future of High Performance Database Processing (DeWitt, Gray)
- Hash joins and hash teams in Microsoft SQL Server (Graefe, Bunker, Cooper)
- An Overview of Query Optimization in Relational Systems (Chaudhuri)
- MonetDB/X100: Hyper-Pipelining Query Execution (Boncz, Zukowski, Nes)
- DB2 with BLU Acceleration: So Much More than Just a Column Store (Raman et. al.)
- Dremel: Interactive Analysis of Web-Scale Datasets (Melnik et. al.)
- The MADlib Analytics Library (Hellerstein et. al.)

Published by the Impala team

- Impala: A Modern, Open-Source SQL Engine for Hadoop - Kornacker et. al., CIDR 2015

Other Impala publications

- SQL-on-Hadoop: Full Circle Back to Shared-Nothing Database Architectures - Floratou et. al., VLDB 2014