Integrating analytics with relational databases

Raasveldt, M.

Author: Raasveldt, M.
Title: Integrating analytics with relational databases
Issue Date: 2020-06-09

Four scientific propositions related to the dissertation:

1. The inaccessibility of database technology to data scientists has lead to the creation of alternatives that ignore decades of innovation in database research.
2. Database client protocols have not been changed since computers had 1MB of memory. Now that computers have 1TB of memory they are in dire need of a redesign.
3. Data management solutions need to efficiently integrate with the tools that data scientists are using, otherwise they are useless for them.
4. It is a failure of the data management community that the majority of data scientists are still using flat files to manage their data sets.

Four propositions related to the field:

1. Most people don't need big data solutions - they need fast and easy small to medium data solutions.
2. A system that performs great on a single benchmark but poorly on nine others is very useful for making a scientific publication. However, a system that performs average on all ten benchmarks is much more useful in practice.
3. It is easy to do well on benchmarks when reality is ignored.
4. How fast your benchmark runs is irrelevant if nobody can install and use your system.

General propositions about programming:

1. The widespread practice of enormous amounts of code dependencies has a very big cost. However, it is not the developer that incurs the cost, it is the user of the code.
2. Reinventing the wheel is a great idea if the current wheel is falling apart.
3. Don't attach too much value to code. Sometimes throwing away old code is a better contribution than adding new code.