Algebraic representation for imperative programs

Dr. Karthik Ramachandra (Microsoft Research Lab)

Databases support declarative SQL as well as imperative functions to express data processing tasks. While the evaluation of declarative SQL has received a lot of attention resulting in highly sophisticated techniques, the evaluation of imperative programs has remained naive and highly inefficient. Imperative programs offer several benefits over SQL and hence are often preferred and widely used. But unfortunately, their abysmal performance discourages, and even prohibits their use in many situations. In this talk, I will describe a novel technique to optimize imperative functions in relational databases. We develop a relational algebraic representation for imperative programs which forms the core of this technique. This algebraic representation enables several algebraic simplifications and transformations. Moreover, it is amenable to cost-based optimization, and therefore results in efficient, set-oriented, parallel plans as opposed to inefficient, iterative, serial execution of functions. I will also present an experimental evaluation that demonstrates performance improvements of up to multiple orders of magnitude on real workloads.