

Big Data Analytics 2013



Invited talk

Making Big Data Analytics Interactive and Real-Time

Frank McSherry, *Microsoft Research*

This talk will cover a new computational framework, differential dataflow, that generalizes standard incremental dataflow for far greater re-use of previous results when collections change. Informally, differential dataflow distinguishes between the multiple reasons a collection might change, including both loop feedback and new input data, allowing a system to re-use the most appropriate results from previously performed work when an incremental update arrives. Our implementation of differential dataflow efficiently executes queries with multiple (possibly nested) loops, while simultaneously responding with low latency to incremental changes to the inputs. We show how differential dataflow enables orders of magnitude speedups for a variety of workloads on real data, and enables new analyses previously not possible in an interactive setting.

This is joint work with Derek G. Murray, Rebecca Isaacs, and Michael Isard.