

Big Data Analytics 2013



Invited talk

Mergeable Summaries

Ke Yi, *Hong Kong University of Science and Technology*

This talk is about the mergeability of data summaries. Informally speaking, mergeability requires that, given two summaries on two data sets, there is a way to merge the two summaries into a single summary on the two data sets combined together, while preserving the error and size guarantees. This property means that the summaries can be merged in a way akin to other algebraic operators such as sum and max, which is especially useful for computing summaries on massive distributed data. Several data summaries are trivially mergeable by construction, most notably all the sketches that are linear functions of the data sets. But some other fundamental ones like those for heavy hitters and quantiles, are not (known to be) mergeable. In this talk, we will see how these summaries are indeed mergeable or can be made mergeable after appropriate modifications.