Communication Cost in Parallel Query Processing

Dan Suciu, University of Washington

ABSTRACT

Fix a full, conjunctive query, and consider the following problem: what is the amount of communication required to compute the query in parallel, on p servers, over a large database instance? We define the Massively Parallel Communication (MPC) model, where the computation proceeds in rounds consisting of local computations followed by a global reshuffling of the data. Servers have unlimited computational power and are allowed to exchange any data, the only cost parameters are the number of rounds and the maximum amount of communication per server. I will describe tight bounds on the amount of communication for the case of a single round and data without skew, then discuss extensions to skewed data and multiround.

This is joint work with Paul Beame and Paris Koutris

Short Bio

Dan Suciu is a Professor in Computer Science at the University of Washington. He received his Ph.D. from the University of Pennsylvania in 1995, was a principal member of the technical staff at AT&T Labs and joined the University of Washington in 2000. Suciu is conducting research in data management, with an emphasis on topics related to Big Data and data sharing, such as probabilistic data, data pricing, parallel data processing, data security. He is a co-author of two books Data on the Web: from Relations to Semistructured Data and XML, 1999, and Probabilistic Databases, 2011. He is a Fellow of the ACM, holds twelve US patents, received the best paper award in SIGMOD 2000 and ICDT 2013, the ACM PODS Alberto Mendelzon Test of Time Award in 2010 and in 2012, the 10 Year Most Influential Paper Award in ICDE 2013, the VLDB Ten Year Best Paper Award in 2014, and is a recipient of the NSF Career Award and of an Alfred P. Sloan Fellowship. Suciu serves on the VLDB Board of Trustees, and is an associate editor for the VLDB Journal, ACM TWEB, and Information Systems and is a past associate editor for ACM TODS and ACM TOIS. Suciu's PhD students Gerome Miklau and Christopher Re received the ACM SIGMOD Best Dissertation Award in 2006 and 2010 respectively, and Nilesh Dalvi was a runner up in 2008.