Querying Big, Dynamic, Distributed Data

Minos Garofalakis
Technical University of Crete
Chania, Greece
minos@softnet.tuc.gr
http://www.softnet.tuc.gr/~minos/

ABSTRACT

Effective Big Data analytics pose several difficult challenges for modern data management architectures. One key such challenge arises from the naturally streaming nature of big data, which mandates efficient algorithms for querying and analyzing massive, continuous data streams (that is, data that is seen only once and in a fixed order) with limited memory and CPU-time resources. Such streams arise naturally in emerging large-scale event monitoring applications; for instance, network-operations monitoring in large ISPs, where usage information from numerous sites needs to be continuously collected and analyzed for interesting trends. In addition to memory- and time-efficiency concerns, the inherently distributed nature of such applications also raises important communication-efficiency issues, making it critical to carefully optimize the use of the underlying network infrastructure. In this talk, we introduce the distributed data streaming model, and discuss recent work on tracking complex queries over massive distributed streams, as well as new research directions in this space.

Categories and Subject Descriptors

H.2.8 Database Applications; C.2.4 Distributed Systems; I.1.2 Algorithms.

Keywords

Big data analytics; Distributed data streams; Continuous queries; approximate query processing; Sketch data summaries

BIO

Minos Garofalakis received the Diploma degree in Computer Engineering and Informatics from the University of Patras, Greece in 1992, and the M.Sc. and Ph.D. degrees in Computer Science from the University of Wisconsin-Madison in 1994 and 1998, respectively. He worked as a Member of Technical Staff at Bell Labs, Lucent Technologies in Murray Hill, NJ (1998-2005), as a Senior Researcher at Intel Research Berkeley in Berkeley, CA (2005-2007), and as a Principal Research Scientist at Yahoo! Research in Santa Clara, CA (2007-2008). In parallel, he also held an Adjunct Associate Professor position at the EECS Department of the University of California, Berkeley (2006-2008. As of October 2008, he is a Professor of Computer Science at the School of Electronic and Computer Engineering of the Technical University of Crete, and the Director of the Software Technology and Network Applications Laboratory (SoftNet). Prof. Garofalakis' research focuses on Big Data analytics, spanning areas such as database systems, data streams, data synopses and approximate query processing, probabilistic databases, and data mining. His work has resulted in over 120 published scientific papers in these areas, and 35 US Patent filings (27 patents issued) for companies such as Lucent, Yahoo!, and AT&T. GoogleScholar gives over 9000 citations to his work, and an h-index value of 50. Prof. Garofalakis is an ACM Distinguished Scientist (2011), and a recipient of the IEEE ICDE Best Paper Award (2009), the Bell Labs President's Gold Award (2004), and the Bell Labs Teamwork Award (2003).