



Computer Science Colloquium
October 30th @ 3:45 p.m. 228 MSB

Dr. Feng Yu, Ph.D.
Department of Computer Science & Information System
Youngstown State University, Ohio

C S2: A New Database Synopsis for Query Estimation

Abstract:

Fast and accurate estimations for complex queries are profoundly beneficial for large databases with heavy workloads. In this research, we propose a statistical summary for a database, called CS2 (Correlated Sample Synopsis), to provide rapid and accurate result size estimations for all queries with joins and arbitrary selections. Unlike the state-of-the-art techniques, CS2 does not completely rely on simple random samples, but mainly consists of correlated sample tuples that retain join relationships with less storage. We introduce a statistical technique, called reverse sample, and design a powerful estimator, called reverse estimator, to fully utilize correlated sample tuples for query estimation. We prove both theoretically and empirically that the reverse estimator is unbiased and accurate using CS2. Extensive experiments on multiple datasets show that CS2 is fast to construct and derives more accurate estimations than existing methods with the same space budget.

Bio:

Dr. Feng Yu is currently an assistant professor of Computer Science and Information System at Youngstown State University, Youngstown, OH. He received his Ph.D. in Computer Science Department at Southern Illinois University Carbondale, IL, in 2013 and his M.S. in Pure Mathematics from Shandong University, China, in 2008. Before he joined YSU, he was a research programmer at University of Illinois at Urbana-Champaign. His research interests include database query optimization, big data, cloud computing, and data mining.