

Join Algorithms: From External Memory to the BSP

Ke Yi

Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China
yike@ust.hk

Abstract

Database systems have been traditionally disk-based, which had motivated the extensive study on external memory (EM) algorithms. However, as RAMs continue to get larger and cheaper, modern distributed data systems are increasingly adopting a main memory based, shared-nothing architecture, exemplified by systems like Spark and Flink. These systems can be abstracted by the BSP model (with variants like the MPC model and the MapReduce model), and there has been a strong revived interest in designing BSP algorithms for handling large amounts of data.

With hard disks starting to fade away from the picture, EM algorithms may now seem less relevant. However, we observe that many of the recently developed join algorithms under the BSP model have a high degree of resemblance with their counterparts in the EM model. In this talk, I will present some recent results on join algorithms in the EM and BSP model, examine their relationships, and discuss a general theoretical framework for converting EM algorithms to the BSP.

2012 ACM Subject Classification Theory of computation → Data structures and algorithms for data management

Keywords and phrases External memory model, BSP, join algorithms

Digital Object Identifier 10.4230/LIPIcs.ICDT.2018.2

Category Invited Talk



© Ke Yi;

licensed under Creative Commons License CC-BY

21st International Conference on Database Theory (ICDT 2018).

Editors: Benny Kimelfeld and Yael Amerdamer; Article No. 2; pp. 2:1–2:1

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany