

Distributed XML Processing

M. Tamer Özsu

University of Waterloo
Cheriton School of Computer Science
tozsu@cs.uwaterloo.ca

Abstract. XML is commonly used to store data and to exchange it between a variety of systems. While centralized querying of XML data is increasingly well understood, the same is not true in a scenario where the data is spread across multiple nodes in a distributed system. Since the size of XML data collections are increasing along with the heavy workloads that need to be evaluated on top of these collections, scaling a centralized solution is becoming increasingly difficult. A common method for addressing this issue is to distribute the data and parallelize query execution. This is well understood in relational databases, but the issues are more complicated in the case of XML data due to the complexity of the data representation and the flexibility of the schema definition. In this talk, I will introduce our new project to systematically study distributed XML processing issues. The talk will focus on data fragmentation and localization issues.

This is joint work with Patrick Kling.