

## Supporting High-level Abstractions through XML Technology

Xiaogang Li and Gagan Agrawal

To appear at the *16th Workshop on Languages and Compilers for Parallel Computing (LCPC03)*, College Station, TX, 2-4 October 2003

### Abstract

Development of applications that process large scientific datasets is often complicated by complex and specialized data storage formats. In this paper, we describe the use of XML technologies for supporting high-level programming methodologies for processing scientific datasets. We show how XML Schemas can be used to give a high-level abstraction of a dataset to an application developer. A corresponding *low-level* Schema describes the actual layout of data and is used by the compiler for code generation. The compiler needs a systematic way for translating the high-level code to a low-level code. Then, it needs to transform the generated low-level code to achieve high locality and efficient execution. This paper describes our approach to these two problems. By using Active Data Repository as the underlying runtime system, we offer an XML based front-end for storing, retrieving, and processing flat-file based scientific datasets in a cluster environment.