SEMINARI DI AVVIAMENTO AL LAVORO

Corsi di Laurea e Laurea Magistrale in Informatica

Nell'ambito dei Seminari per l'avviamento al Lavoro organizzati dal Corso di Laurea Magistrale in Informatica Lunedì 30 Novembre 2015, alle ore 14:00, nell' aula 7 del DMI si svolgerà seminario dal titolo

The Data Management Entity: A Simple Abstraction to Facilitate Big Data Interoperability

Prof. Damianos Chatziantoniou
Department of Management Science and Technology
Athens University of Economics and Business (AUEB)

Abstract

Today's big data era is described by intense variety in data management systems, query languages and programming paradigms. Each system targets well a specific application area, reinforcing the belief that the era of one-size fits all has gone for good. Interoperability, systems' connectivity, federation and high-level data models become once again the core of research initiatives. In this talk we present a layered architecture to support interoperability among different data management systems, generalized under the term *data management entities* (DMEs). DMEs range from JVMs running java programs to Hadoop systems employing complex MapReduce jobs to traditional RDBMS running SQL queries to stream engines and CEP scripts. The top layer consists of a universe of DMEs, communicating through a well defined http-like protocol: a DME *transparently* invokes another DME's data manipulation task, regardless task's nature. Communicating DMEs share/operate on a shared data object, a key-value set (KVS) - just a set of key-value pairs - which exists in the layer below and is referenced through a unique (internet-wide) address via a well-defined API. This layer serves as the transient common memory space for communicating DMEs and consists of *globally addressable* KVSs, organized in domains, sub-domains, etc. In a way, this approach constitutes a form of remote procedure call *by reference* (the KVS is the common reference). We argue that this architecture allows the construction of high level query languages and cost-based distributed query processing engines, involving completely heterogeneous data manipulation tasks. For example, we show that MapReduce evaluation algorithm and distributed relational query processing are just instances of the proposed architecture. We also claim that it can easily facilitate the end-to-end processing in big data applications, an established goal in the research agenda set by the Beckman report.