

Models and Algorithms for High-Performance Data Management and Mining on Computational Grids and Clouds

Alfredo Cuzzocrea

Published online: 26 October 2014
© Springer Science+Business Media Dordrecht 2014

High-performance data management and mining are well-known resource- and time-consuming activities that have attracted a great deal of interest from the research community. High-performance data management can be reasonably intended as a intermediate step of high-performance data mining activities over *large-scale amounts of data*, while still keeping unaltered the primary and self-contained focus of achieving effectiveness and efficiency in these task themselves. These topics are now of very-high interest, due to the emerging trends falling under the terms “*Big Data*” and “*Cloud Infrastructures*”. There exists a wide range of application scenarios where high-performance data management and mining play a critical role. Among these, we recall: prediction of natural disasters, analysis of massive sensor and stream data, scientific computing and e-science, fraud detection, business intelligence, cloud intelligence, and so forth.

In the application scenarios above, the usage of *computational Grids and Clouds*, which is a well-understood and mature paradigm for computational intelligence whose merits are not only limited to high-performance capabilities, but also they convey to novel amenities falling in the wide range of e-science and scientific computing. All in one, coupling high-performance computational Grids and Clouds with

data management and mining activities pursues the goal of creating a successful environment for data management and mining over large-scale amounts of data, which, beyond to the major goal of achieving high-performance during the execution of these activities, embeds a plethora of innovative features, such as: (i) providing scientists with the capabilities of allowing flexibility during the creation of scientific workflows; (ii) enriching data management and mining with relevant benefits, such as automation of analysis processes and dimensionality reduction; (iii) allowing for the capability of obtaining well-defined and well-structured data management and mining activities over large-scale data; (iv) providing problem-solving environments where analysts can re-produce the typical algorithm editing process, with meaningful abstractions on primitives and black-box components.

In this scientific context, this special issue on “*Models and Algorithms for High-Performance Data Management and Mining on Computational Grids and Clouds*” of *Journal of Grid Computing* contains three papers, which have gone through two rigorous review rounds before being accepted for final inclusion. Some of the contributions of this special issue have been invited for submission as the best papers of the *10th LNCS International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2010)*, held in Busan, Korea, during May 21-23, 2010, led by the Editor.

As a quality result, this special issue provides high-quality contributions in the context of models

A. Cuzzocrea (✉)
ICAR-CNR and University of Calabria, Via P. Bucci, 41C,
Rende, 87036 Cosenza, Italy
e-mail: cuzzocrea@si.deis.unical.it

and algorithms for high-performance data management and mining on computational Grids and Clouds, by emphasizing both theoretical as well as practical aspects of this so-interesting yet not completely-explored scientific area, which has also important and well-understood implications in different-but-related disciplines like bio-informatics, genomic computing, e-science, analytics over large-scale data repositories, graph databases, data integration tools, and so forth. The contributions of this special issue focuses the attention on a variety of topics falling in the context of high-performance distributed data management and mining from fundamentals (e.g., scheduling algorithms with resource constrained requirements) to applications (e.g., large-scale scientific applications). In the following, we provide a summary of papers contained in this special issue.

The first paper, titled “*Adaptable Scheduling Algorithm for Grids with Resource Redeployment Capability*”, by Chih-Hsuan Hsu, Cho-Chin Lin and Tsan-sheng Hsu, focuses the attention on *scheduling algorithms over resource-constrained distributed environments*, and proposes two distinct algorithms, called AMOF and AMOSF. Both of these algorithms consider availability variation as well as resource heterogeneity while scheduling an ongoing *workflow application* on the Grid. Authors also provide experimental assessment and analysis that AMOF and AMOSF outperform the well-known scheduling algorithms GS and HEFT in most of the cases.

The second paper, titled “*QoS-based Task Group Deployment on Grid by Learning the Performance Data*”, by Nithiapidary Muthuvelu, Ian Chai, Eswaran Chikkannan and Rajkumar Buyya, moves the attention on *QoS aspects of task groups to be deployed over distributed environments*. Authors show how increasing the task success rate in economic Grids by *optimally mapping the tasks to the resources prior to the batch deployment*. The task-resource mapping (*Advance QoS Planning*) is thus decided based on QoS requirement and by mining the historical performance data of the application tasks using a genetic algorithm. The mapping is then used to assist in creating the task groups. Practical experiments are conducted to validate the proposed method and suggestions are given to implement the method in a Cloud environment as well as to process real-time tasks.

Finally, the third paper, titled “*User Interaction and Data Management for Large Scale Grid Applications*”, by Alessandro Costantini, Osvaldo Gervasi, Fabiana Zollo and Luca Caprini, introduces a novel model that combines the *X3DMMS application environment with the G3CPie execution framework*, thus enabling the user to perform large-scale computations on distributed computing environments. This approach facilitates the management and the preparation of the data required to determine input files for *DL_POLY*, a popular *Molecular Dynamics (MD) package* used for studying molecular systems. In the so-obtained computational environment, researchers can define the initial configuration of the molecular system in a intuitive way, making use of the X3DMMS virtual reality environment, and prepares the related MD-package-oriented input files. After this phase, the required computations can be executed by using the G3CPie workflow environment. Authors also provide the experimental validation of the proposed model via a suitable case study.

The Editor would like to express his sincere gratitude to the Editor-In-Chief of *Journal of Grid Computing*, Prof. Peter Kacsuk, for accepting his proposal of a special issue focused on models and algorithms for high-performance data management and mining on computational Grids and Clouds, and for assisting him whenever required. The Editor would also like to thank all the reviewers who have worked within a tight schedule and whose detailed and constructive feedbacks to authors have contributed to substantial improvement in the quality of final papers.

Alfredo Cuzzocrea is currently a Senior Researcher at the Institute of High Performance Computing and Networking of the Italian National Research Council, Italy, and an Adjunct Professor at the University of Calabria, Italy. He holds 35 Visiting Professor positions worldwide (Europe, USA, Asia, Australia). He serves as Springer Fellow Editor. He serves as Elsevier Ambassador. He holds several roles in international scientific societies, steering committees for international conferences, and international panels, some of them having directional responsibility. He served as Panel Leader and Moderator in international conferences. He served as Invited Speaker in several international conferences worldwide (Europe, USA, Asia). He is member of scientific

boards of several PhD programs worldwide (Europe, Asia, Australia). He serves as Editor for the Springer series *Communications in Computer and Information Science*. He covers a large number of roles in international journals, such as Editor-In-Chief, Associate Editor, Special Issue Editor (including JCSS, IS, KAIS, FGCS, DKE, INS, BigData Research), and international conferences, such as General Chair, Program Chair, Workshop Chair, Local Chair, Liaison Chair and Publicity Chair (including CSE, ODBASE, DaWaK, DOLAP, ICA3PP, ICEIS, APWeb, SSTDM, IDEAS, IDEAL). He edited more than 30 international books and conference proceedings. His current research

interests include multidimensional data modeling and querying, data stream modeling and querying, data warehousing and OLAP, OLAM, Grid and P2P computing, privacy and security of very large databases and OLAP data cubes, models and algorithms for managing uncertain and imprecise information and knowledge, models and algorithms for high-performance distributed computing and architectures. He is author or co-author of more than 330 papers in international conferences (including EDBT, CIKM, SSDBM, MDM, DaWaK, DOLAP), international journals (including JCSS, IS, KAIS, DKE, INS) and international books (mostly edited by Springer).