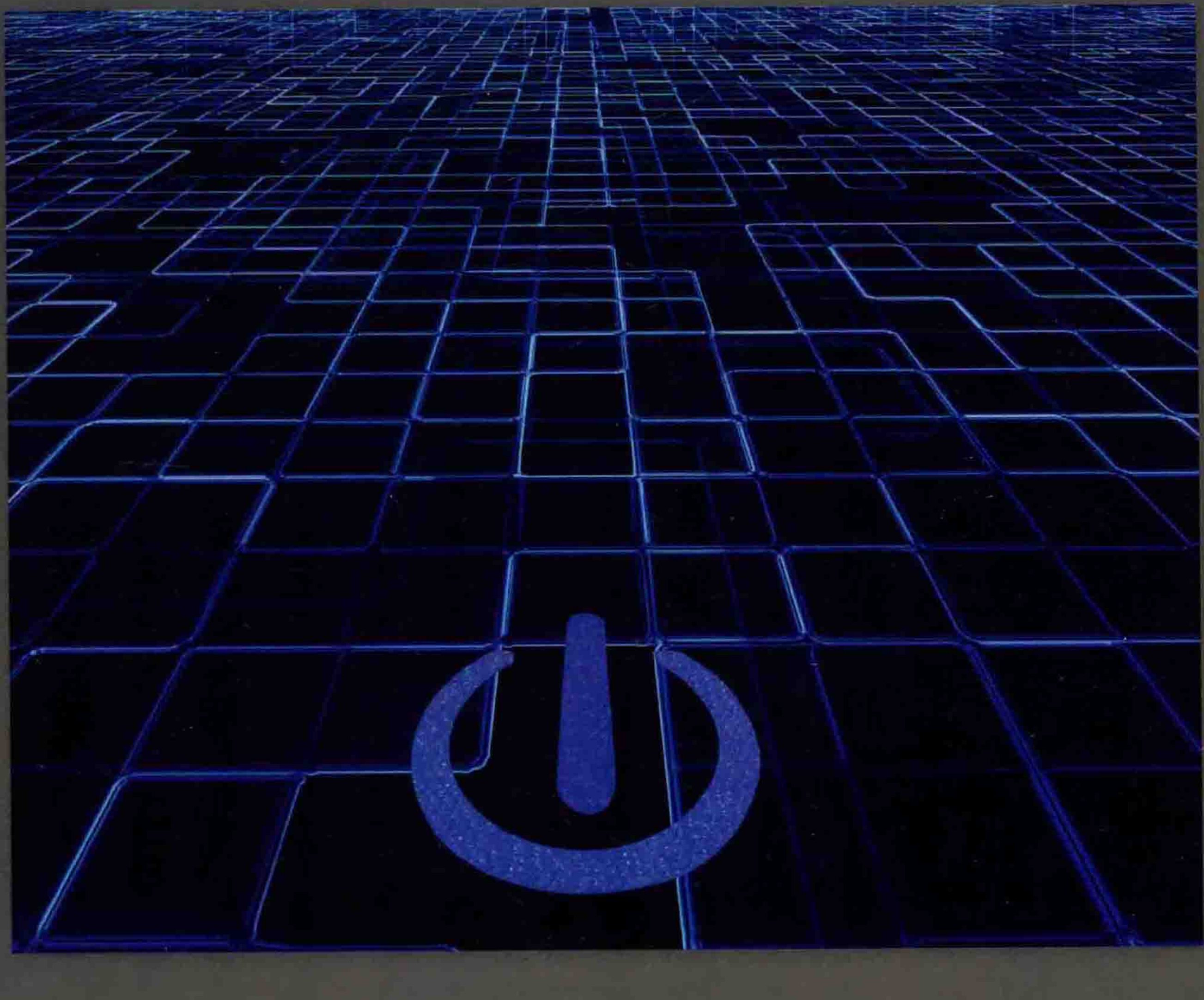
PREMIER REFERENCE SOURCE

# Computational and Data Grids:
## Principles, Applications, and Design



**Nikolaos Preve**

# Computational and Data Grids:

## Principles, Applications, and Design

Nikolaos P. Preve
*National Technical University of Athens, Greece*



Information Science
REFERENCE

| | |
|---|---|
| Senior Editorial Director: | Kristin Klinger |
| Director of Book Publications: | Julia Mosemann |
| Editorial Director: | Lindsay Johnston |
| Acquisitions Editor: | Erika Carter |
| Development Editor: | Hannah Ablebeck |
| Production Editor: | Sean Woznicki |
| Typesetters: | Adrienne Freeland, Mackenzie Snader |
| Print Coordinator: | Jamie Snavely |
| Cover Design: | Nick Newcomer |

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com



Library of Congress Cataloging-in-Publication Data

Computational and data grids: principles, applications, and design / Nikolaos Preve, editor.
p. cm.
Includes bibliographical references and index.
Summary: "This book provide relevant theoretical frameworks covering the latest empirical research findings in the area of grid computing, with a critical perspective bridging the gap between academia and the latest achievements of the computer industry"--Provided by publisher.
ISBN 978-1-61350-113-9 (hardcover) -- ISBN 978-1-61350-114-6 (ebook) -- ISBN 978-1-61350-115-3 (print & perpetual access) 1. Computational grids (Computer systems) 2. Computer networks. I. Preve, Nikolaos, 1979-
QA76.9.C58C65 2012
004'.36--dc22

2011009965

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

*To Anastasia and Peter, who have always been trying to enlighten the obscure parts of my life by giving me the strength to cope with the difficulties of life*

# List of Reviewers

Harold Enrique Castro Barrera, *Universidad de los Andes, Colombia*
Edgar Eduardo Rosales Rosero, *Universidad de los Andes, Colombia*
Mario, José Villamizar Cano, *Universidad de los Andes, Colombia*
Mehdi Sheikhalishahi, *Universita' della Calabria Rende, Italy*
Mohamed El-Refaey, *EDC, Egypt*
Bhaskar Prasad Rimal, *Kookmin University, Seoul, Korea*
Florin Pop, *University Politehnica of Bucharest, Romania*
Alexandru Costan, *University Politehnica of Bucharest, Romania*
Roberto Barbera, *University of Catania, Italy*
Cevat Şener, *Middle Eastern Technical University, Turkey*
Philipp M. Glatz, *Graz University of Technology, Austria*
Antonios Gouglidis, *University of Macedonia, Greece*
Axel Tenschert, *High Performance Computing Center Stuttgart, Germany*
Vassiliki Pouli, *National Technical University of Athens, Greece*
Salvatore Distefano, *Università di Messina, Italy*
Francis Xavier Ochieng, *Jomo Kenyatta University of Agriculture and Technology, Kenya*
Mark Anderson, *Edge Hill University, UK*
Bastian Koller, *High Performance Computing Centre Stuttgart, Germany*
Andras Micsik, *MTA SZTAKI, Hungary*

# Preface

After a lot of development efforts, computer science led us undoubtedly to technological revolutions which have been characterized by the creation of the Internet that influences the future of this science. The next revolutionary step occurred by the necessity of the creation of a new computer network, when researchers realized that the computer resources are underutilized. Researchers observed that machines spent much time idly waiting for human input increasing their cost through their underutilization. Their efforts concentrated in maximizing the utilization of computational resources, decreasing at the same time the cost of the computers. A vision for a new computer infrastructure was born at Argonne National Laboratory. The fathers of this unforeseen revolution were Foster & Kesselman (1997). They coined a new term about this new infrastructure which changed the way we think about computer science while they aimed to make computational resources available and efficient to everyone, like electricity.

The term *Grid* has defined a new scientific area of computing which combines heterogeneous, geographically dispersed computer resources that are a part of various administrative domains and cooperate in order to reach a common goal. The most significant achievement of this new emerged infrastructure is the resources sharing across various loosely coupled networks. The outcome of resources sharing combination with uniqueness characteristics, such as adaptability, applicability, flexibility, interoperability, usability, and scalability, is a grid network which provides us with vast computational and storage capabilities.

A strong basis of the definition was given by the father of the grid, Foster (2002), who defined a grid network with the following requirements. A grid system should have resources coordination that are not subject to centralized control, the usability of standard, open, general-purpose protocols and interfaces, while it delivers nontrivial Quality-of-Service (QoS). The integration between distributed and heterogeneous resources is achieved through a middleware. The usage of a grid middleware is compulsory because it acts as a mediator layer providing a consistent and homogeneous access to resources managed locally with different syntax and access methods.

Grid systems achieved with great success to integrate, virtualize, and manage resources and services within distributed, heterogeneous, and dynamic Virtual Organizations (VOs). The necessity of grid is obvious in scientific communities which demand to access vast computational resources in order to be able to run high-performance applications. For this reason scientific applications were and are the most important exploiter of grids. However, the continuous increasing demand in specific grid infrastructures from commercial organizations and scientific communities led us to categorize a grid network. There are classes which define a grid network. We have Access Grids, Bio Grids, Computational Grids, Cluster Grids, Campus Grids, Commodity Grids, Data Grids, Knowledge Grids, Science Grids, Sensor Grids, and Tera Grids. Although, a grid network must be evaluated according to the running applications, business value, and scientific results that it delivers, not its architecture (Foster, 2002).

The grid computing is a newly developed technology among similar large-scale computer implementations, such as cloud computing, distributed and parallel computing, Internet, peer-to-peer networks, and virtualization technologies. The benefits of grid technology and its most important achievements are addressed below. These achievements have comprised the basis for next generation computer systems indicating a way for building future network infrastructures.

- A user utilizes a grid infrastructure without having to investigate the underlying architecture while he is able to manage his owned resources
- A grid uses the underutilized processing power, maximizing the available resources of the system and minimizing the execution time of a large job
- Complex resources demanding scientific problems can be solved with parallel CPU and data storage capacities that grid provides
- The computational resources are allocated, aggregated, and collaborated in a grid environment besides its heterogeneity, geographical dispersion, and administrative variety.
- A grid does not share the data between users but it permits common access to them (many-to-many sharing)
- Grid removes the barriers to virtualization technologies expending its capabilities

Nowadays, grids are not a common good for scientific communities only. An increasing interest in this technology has begun from large companies and organizations which focus on grid implementations. This revolution influenced the processor industry which built multithreaded processors based on grid technology assisting to be spread faster. Thus, a need for global standardization occurred ensuring the interoperability of grid infrastructures. After seven years of life the Open Grid Forum (OGF), previously the Global Grid Forum (GGF), is beginning to produce standards that meet the needs of the community and that are being adopted by commercial and open source software providers (Smith et al. 2009). Also, we have to mention some of the most important organizations that made efforts focusing on the development of grid computing with multiple contributions on the field. These are World Wide Web Consortium (W3C), Organization for the Advancement of Structured Information Standards (OASIS), Web Service Interoperability Organization (WS-I), Distributed Management Task Force (DMTF), Internet2, Liberty Alliance, and Enterprise Grid Alliance (EGA).

Grid is a technology that is going to become prominent in the next few years, expecting a wide proliferation in its use. Grid computing infrastructures are already accessible to many users covering their needs in computer resources. Nevertheless, grids will have an ever increasing role comprising a basis in the field of scientific research. It is therefore necessary a thorough understanding of principles, designs, and applications in a grid environment. After so many innovations and achievements in various scientific areas all these years, we are still wondering and carrying opinions and thoughts that computer science is a future science. So, several achievements in computer science such as grid have opened the door for a different future of this scientific area.

## BOOK ORGANIZATION

This book is organized into three major sections containing 15 chapters and dealing respectively with principles, designs, and applications of grid computing. A brief description of each chapter follows.

**Section 1:** Principles include aspects, challenges and trends in grid computing.

**Chapter 1** presents a representative set of projects focused on providing solutions for the use of idle computing cycles aiming to provide an overview of the main implementations and research on Desktop Grids and Volunteer Computing Systems (DGVCSs). This chapter also introduces a new taxonomy model dealing with the occurred issues. A discussion aims to the evolution stages, main implementations and research on DGVCSs and through the presented analysis it succeeds in identifying the main characteristics of DGVCSs.

**Chapter 2** introduces a new buzzword computing paradigm focusing on the infrastructure. Having the proposed paradigm as a basis, the authors analyze various technologies around it in software and networking domains that are involved in complementing grid and cloud computing. Through the presentation of a new architecture which is mainly inspired by Infrastructure as a Service (IaaS) model to address grid and cloud complementarity approach, they analyze and evaluate current practices and services that are applied in these infrastructures defining new research topics that address this issue.

**Chapter 3** extends the discussion through an analysis of the key concepts of Service Oriented Architecture (SOA), grid, and cloud computing demonstrating a tight relation between these concepts in order to develop a highly scalable application system. This chapter also presents a coverage approach for concepts of Web 2.0 related to grid computing and on-demand enterprise model.

**Chapter 4** focuses on the resource heterogeneity, the size and number of tasks, the variety of policies, and the high number of constraints which are some of the main characteristics that contribute to this complexity. This chapter presents a holistic approach of the necessity and the requirements of scheduling mechanism in grid systems while it offers a critical analysis of existing methods and algorithms, scheduling policies, fault tolerance in scheduling process, scheduling models and algorithms and optimization techniques for scheduling in grid environments.

**Chapter 5** deals with grid infrastructures that produce enormous size of data which should be supported by scalable data storage and management strategies. This chapter addresses the key issues of data handling in grid environments and deals with the upcoming challenges in distributed storage systems. It also presents how existing solutions cope with these high requirements while it indicates their advantages and limitations.

**Section 2:** Designs focuses on different grid architectures and methodologies for different grid networks.

**Chapter 6** introduces the necessity of world standard platforms in order to support e-Science and foster virtual research communities. A description of the developed e-Infrastructure around several countries is follows with an outlook on the very important issue of their long term sustainability.

**Chapter 7** focuses on resource aware sensor grid middleware. This chapter investigates misconceptions in design, simulation, test and measurement that need to be overcome or be considered for successful implementations. A framework for design, simulation, and testing is developed in sensor grids. This chapter also presents an approach that implements performance optimizations and resource awareness with a minimum of negative impact from mutual side effects.

**Chapter 8** develops an access control model for grid computer environments. The authors analyze the Role Based Access Control (RBAC) and Usage Control ABC ($UCON_{ABC}$) models demonstrating how the theoretical access control models and architectures are implemented into mechanisms. They also provide a comparison between the examined access control models and mechanisms, aiming to expose the different aspects of grid access control.

**Chapter 9** examines the challenge of ontology matching in a grid environment in a scalable and high efficient way. An approach for ontology matching based on the requirements of grid architecture is introduced in this chapter while discussing and focusing on related approaches and tools.

**Section 3:** Applications dealing with emerged issues in the field of various grid implementations.

**Chapter 10** faces the issue of security in grid computing introducing an approach to security in grid environments that are built using Service Oriented Architecture (SOA) technologies. This chapter also describes in-depth the security protocols and technologies that have applied on a Web Service (WS) based grid environment.

**Chapter 11** proposes a lightweight cryptography algorithm combining the strong and highly secure asymmetric cryptography technique (RSA) with the symmetric cryptography (AES) protecting data and files in a grid environment. In this chapter the authors propose an algorithm named, Secure Storage System ($GS^3$), and it has been implemented on top of the Grid File Access Library (GFAL) of the gLite middleware in order to provide a file system service with cryptography capability and POSIX interface. A detailed description of $GS^3$ about its implementation is given based on a well developed evaluation performance.

**Chapter 12** extends the usage of grid computing in other scientific fields such as meteorology in order to predict and assess wind and solar resources. This chapter develops an approach based on utilization of remote grid computing essentially undertaking grid computing remotely by accessing the grid computers in host countries with more advanced Information Technology infrastructure.

**Chapter 13** describes the implementation of grid services and defines an approach to a development framework which would enable the creation of agile services. The authors present an alternative solution which adopts aspect-oriented programming as a core component in the framework and they achieve to develop agile services in a grid environment focusing on teleworking.

**Chapter 14** addresses the requirements of academic end users, the grid paradigm and underlines past developed technologies based on the needs of potential business end users. The authors demonstrate that the trend has changed towards the use of grid technologies within electronic business. This chapter also focuses on the rationale behind the performed developments through the presentation results of the BREIN project. Moreover, a generic solution is presented and it is applied to a variety of distinct application areas.

**Chapter 15** presents the potentialities of a new innovative Internet QoS (Quality-of-Service) architecture known as Flow-Aware Networking (FAN). Besides, the QoS provisioning for grid computing, the authors also propose a new promising QoS paradigm as a potential solution to achieve better performance of FAN architecture over DS architecture.

## BOOK OBJECTIVE

The vision of grid computing inspired many scientists to get actively involve in the field along these years developing and evolving this emerging technology. This book deals with computational and data grids. The key objective is to provide grid students, practitioners, professionals, professors and researchers with an integral vision of the topic.

The idea of writing this book came up after the increasing success and interest of scientific community on grid computing. So, this book aims to foster awareness of the essential ideas by exploring current and future developments in the grid area. Specifically, this book focuses on these areas that explore new methodologies, developments, deployments, and implementations that cover the latest research findings in the area of grid computing, making this mission even more complex. The book describes the state-of-the-art, innovative theoretical frameworks, advanced and successful implementations as well as the latest research findings in the area of grid computing.

The purpose of this book is to provide the reader with the latest research findings and new presented perspectives which are implemented in various grid implementations around the world. Moreover, it will motivate the reader to follow several different methodologies through the contents. The book delves into details of grids, guiding the reader through a collection of chapters dealing with key topics. By including in our book these characteristics, we target the book to readers who want to go deeper into this scientific field and gear students, practitioners, professionals, professors and researchers who have a basic understanding in grid computing. The reader will also have a working knowledge of how this technology is utilized in computer science and how grid computing is able to support other scientific fields. The presentation of current theories, implementations, applications and their impact on grid computing provide the reader with a strong basis for further exploration in this area. At the same time, the mixed-balanced book structure helps the reader to obtain a holistic approach of today's grid systems around the world.

The value of this book is focused on a compact coverage of grid computing technologies that are important to the reader to know today. It also aims to provide an essential knowledge, comprising the foundations for further development and more in-depth education or specific skills in the scientific area of grid computing. Everyone who reads this book should walk away at least with the terminology and basic background to understand the trends that are currently taking place. This provides the reader with a foundation upon which to build his knowledge. The book may serve both as an introduction and as a technical reference familiarizing the readers with the subject and contributing to new advances in the field.

The book attracted the interest of academia and industry around the world in the area of grid computing. Undergraduate and graduate students, researchers, professors, system designers and programmers, and IT policy makers contributed in this book who are actively involved in the field. The book received 153 full chapter submissions and each submission received two or three blind double-reviews by at least two experts and independent reviewers. As a result, 27 chapter submissions were accepted, with an acceptance rate 17.6%. In this book 15 submissions, out of 27, are included.

## REFERENCES

Foster, I. (2002). What is the grid? A three point checklist. *GRIDtoday*, *1*(6), 22–25.

Foster, I. & Kesselman, C. (1997). Argonne workshop explores construction of a national computational grid. *Parallel Computing Research Newsletter*, *5*(4).

Smith, C., Kielman, T., Newhouse, S., & Humphrey, M. (2009). The HPC basic profile and SAGA: Standardizing compute grid access in the open grid forum. *Concurrency and Computation*, *21*(8), 1053–1068. doi:10.1002/cpe.1408

# Acknowledgment

This book brought memories in my mind from my early student years in the School of Electrical and Computer Engineering of the National Technical University of Athens when my journey in this science began.

*"As you set out for Ithaca, pray that the road is long, full of adventure, full of lore...And if you find her poor, Ithaca will not have fooled you. Wise as you have become, with so much experience, you must already have understood what Ithacas mean." — Constantine P. Cavafy (1911)*

A friend of mine recited to me the above poem in a discussion on our future plans in our lives and in this science. A lot of years have passed and a lot of things have changed since then. I have begun my acknowledgment with this poem because I want to address it to the readers of this book who had or still have the same feeling as me.

Every book requires months of preparation, and this book is no exception. The credit for this book goes first and foremost to the authors of each chapter. As editor, I would like to thank them from the bottom of my heart and to express my deepest appreciation to all the authors for their participation in this project, for their excellent contributions and their continuous interest in it. Making this kind of compilation is a huge responsibility which would not have been possible without the efforts and patience of the contributors. This book is a proof that when people work towards a common goal, they cannot be affected by distances, limitations, obstacles, discriminations, circumstances, various situations, different languages and cultures.

I would like to express my appreciation to Jan Travers, Vice President of IGI Global, for the given opportunity. Special thanks also to all the staff at IGI Global for their continuous encouragement and support on this project throughout the whole process, from the inception of the initial idea to the final publication. In particular, to Kristin M. Klinger, Erica Carter, Christine Bufton, and Hannah Abelbeck. So, Christine and Hannah, please accept my warmest compliments for your excellent guidance and professional support. In IGI Global, you prove that you are an excellent publisher who knows about professionalism and cares for its people.

Finally, I would like to express my deepest gratitude to my family for the long evenings and weekends that I spent developing this book and I was not with them.

*Nikolaos P. Preve*
*National Technical University of Athens, Greece*

Section 1

# Principles of Grid Infrastructures

# Table of Contents

## Section 2
## Grid Network Designs

**Section 3**
**Applications for Grid Computing**

# Chapter 1
# Desktop Grids and Volunteer Computing Systems

**Harold Enrique Castro Barrera**
*Universidad de los Andes, Colombia*

**Edgar Eduardo Rosales Rosero**
*Universidad de los Andes, Colombia*

**Mario José Villamizar Cano**
*Universidad de los Andes, Colombia*

## ABSTRACT

*Desktop Grids and Volunteer Computing Systems (DGVCSs) are approaches of distributed systems aimed at the provision of large scale computing infrastructures to support eScience project, by taking advantage of non-dedicated resources, most of them desktop computers presenting low levels of use. Those computers are available through Internet or Intranet environments, have partial availability, are highly heterogeneous, and are part of independent administrative domains. Currently, the idle computational capabilities of millions of volunteer distributed computing resources are exploited without affecting the quality of service perceived by the end users. This chapter presents a comprehensive state of the art of DGVCSs, providing a global picture of the main solutions and research trends in DGVCSs. It will discuss the evolution of such systems by analyzing representative examples. We identify the main characteristics for DGVCSs, and we introduce a new taxonomy to categorize these projects to make their study easier.*

## INTRODUCTION

Also known as volunteer computing (Sarmentwa, 2001) or public resources (Anderson, Cobb, Korpela, Lebofsky, & Werthimer, 2002) (SETI@home, 2010), Desktop Grids and Volunteer Computing Systems (DGVCSs) are approximations of the distributed computing which seek to maximize the efficient use of partially available computing resources. This includes the non exclusive use of computing resources, while ensuring that interactive users of those shared resources do not perceive any deterioration in the quality of service. Such strategies are intended to provide a computing infrastructure at a large scale, primarily used to support the development of e-science projects,

DOI: 10.4018/978-1-61350-113-9.ch001