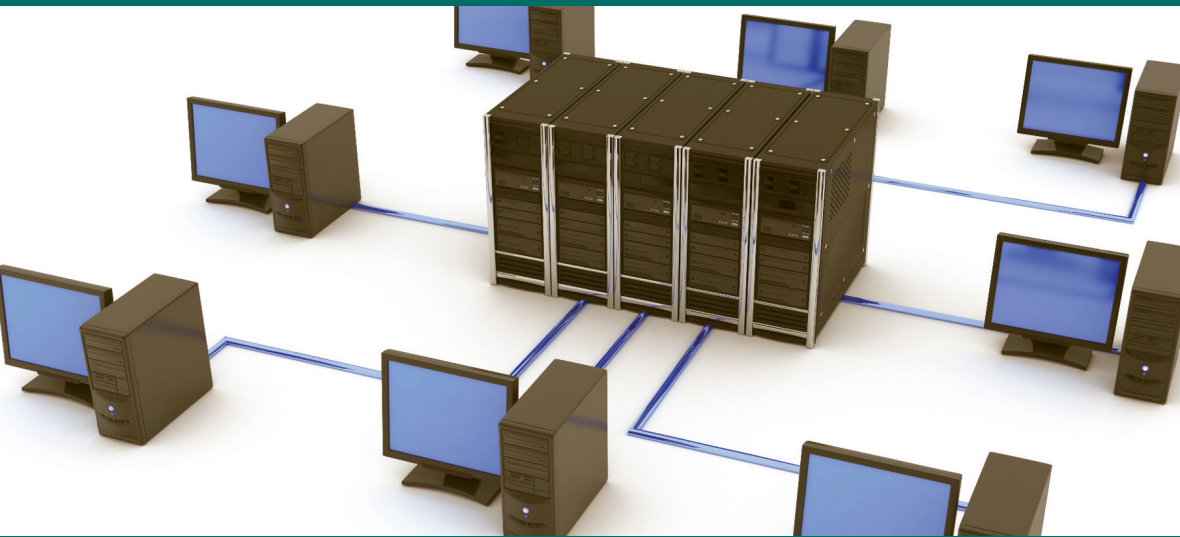


FOCUS
COMPUTER ENGINEERING SERIES



Scheduling of Large-scale Virtualized Infrastructures

Toward Cooperative Management

Flavien Quesnel

ISTE

WILEY

Scheduling of Large-scale Virtualized Infrastructures

FOCUS SERIES

Series Editor Narendra Jussien

Scheduling of Large-scale Virtualized Infrastructures

Toward Cooperative Management

Flavien Quesnel

ISTE

WILEY

First published 2014 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2014

The rights of Flavien Quesnel to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Control Number: 2014941926

British Library Cataloguing-in-Publication Data

A CIP record for this book is available from the British Library

ISSN 2051-2481 (Print)

ISSN 2051-249X (Online)

ISBN 978-1-84821-620-4



Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY

Contents

LIST OF ABBREVIATIONS	xi
INTRODUCTION	xiii
PART 1. MANAGEMENT OF DISTRIBUTED INFRASTRUCTURES	1
CHAPTER 1. DISTRIBUTED INFRASTRUCTURES BEFORE THE RISE OF VIRTUALIZATION	3
1.1. Overview of distributed infrastructures	3
1.1.1. Cluster	3
1.1.2. Data center	4
1.1.3. Grid	4
1.1.4. Volunteer computing platforms	5
1.2. Distributed infrastructure management from the software point of view	6
1.2.1. Secured connection to the infrastructure and identification of users	6
1.2.2. Submission of tasks	7
1.2.3. Scheduling of tasks	8
1.2.4. Deployment of tasks	9
1.2.5. Monitoring the infrastructure	9
1.2.6. Termination of tasks	10

1.3. Frameworks traditionally used to manage distributed infrastructures	10
1.3.1. User-space frameworks	10
1.3.2. Distributed operating systems	11
1.4. Conclusion	12
CHAPTER 2. CONTRIBUTIONS OF VIRTUALIZATION . . .	13
2.1. Introduction to virtualization	13
2.1.1. System and application virtualization	13
2.1.2. Abstractions created by hypervisors	16
2.1.3. Virtualization techniques used by hypervisors	17
2.1.4. Main functionalities provided by hypervisors	19
2.2. Virtualization and management of distributed infrastructures	22
2.2.1. Contributions of virtualization to the management of distributed infrastructures	22
2.2.2. Virtualization and cloud computing	24
2.3. Conclusion	25
CHAPTER 3. VIRTUAL INFRASTRUCTURE MANAGERS USED IN PRODUCTION	27
3.1. Overview of virtual infrastructure managers	27
3.1.1. Generalities	27
3.1.2. Classification	28
3.2. Resource organization	28
3.2.1. Computing resources	28
3.2.2. Storage resources	30
3.3. Scheduling	31
3.3.1. Scheduler architecture	31
3.3.2. Factors triggering scheduling	33
3.3.3. Scheduling policies	35
3.4. Advantages	37
3.4.1. Application programming interfaces and user interfaces	39
3.4.2. Isolation between users	39
3.4.3. Scalability	40
3.4.4. High availability and fault-tolerance	42

3.5. Limits	45
3.5.1. Scheduling	45
3.5.2. Interfaces	46
3.6. Conclusion	46
PART 2. TOWARD A COOPERATIVE AND DECENTRALIZED FRAMEWORK TO MANAGE VIRTUAL INFRASTRUCTURES	49
CHAPTER 4. COMPARATIVE STUDY BETWEEN VIRTUAL INFRASTRUCTURE MANAGERS AND DISTRIBUTED OPERATING SYSTEMS	51
4.1. Comparison in the context of a single node	52
4.1.1. Task lifecycle	52
4.1.2. Scheduling	53
4.1.3. Memory management	56
4.1.4. Summary	58
4.2. Comparison in a distributed context	59
4.2.1. Task lifecycle	59
4.2.2. Scheduling	61
4.2.3. Memory management	62
4.2.4. Summary	64
4.3. Conclusion	64
CHAPTER 5. DYNAMIC SCHEDULING OF VIRTUAL MACHINES	67
5.1. Scheduler architectures	67
5.1.1. Monitoring	68
5.1.2. Decision-making	68
5.2. Limits of a centralized approach	69
5.3. Presentation of a hierarchical approach: Snooze	70
5.3.1. Presentation	70
5.3.2. Discussion	71
5.4. Presentation of multiagent approaches	72
5.4.1. A bio-inspired algorithm for energy optimization in a self-organizing data center	72
5.4.2. Dynamic resource allocation in computing clouds through distributed multiple criteria decision analysis	73

5.4.3. Server consolidation in clouds through gossiping . .	74
5.4.4. Self-economy in cloud data centers – statistical assignment and migration of virtual machines	75
5.4.5. A distributed and collaborative dynamic load balancer for virtual machine	76
5.4.6. A case for fully decentralized dynamic virtual machine consolidation in clouds	77
5.5. Conclusion	78
 PART 3. DVMS, A COOPERATIVE AND DECENTRALIZED FRAMEWORK TO DYNAMICALLY SCHEDULE VIRTUAL MACHINES	 83
 CHAPTER 6. DVMS: A PROPOSAL TO SCHEDULE VIRTUAL MACHINES IN A COOPERATIVE AND REACTIVE WAY	 85
6.1. DVMS fundamentals	86
6.1.1. Working hypotheses	86
6.1.2. Presentation of the event processing procedure . . .	87
6.1.3. Acceleration of the ring traversal	90
6.1.4. Guarantee that a solution will be found if it exists	90
6.2. Implementation	95
6.2.1. Architecture of an agent	96
6.2.2. Leveraging the scheduling algorithms designed for entropy	98
6.3. Conclusion	99
 CHAPTER 7. EXPERIMENTAL PROTOCOL AND TESTING ENVIRONMENT	 101
7.1. Experimental protocol	101
7.1.1. Choosing a testing platform	101
7.1.2. Defining the experimental parameters	101
7.1.3. Initializing the experiment	102
7.1.4. Injecting a workload	102
7.1.5. Processing results	103
7.2. Testing framework	103

7.2.1. Configuration	103
7.2.2. Components	104
7.3. Grid'5000 test bed	106
7.3.1. Presentation	107
7.3.2. Simulations	107
7.3.3. Real experiments	108
7.4. SimGrid simulation toolkit	109
7.4.1. Presentation	109
7.4.2. Port of DVMS to SimGrid	109
7.4.3. Advantages compared to the simulations on Grid'5000	110
7.4.4. Simulations	110
7.5. Conclusion	111

CHAPTER 8. EXPERIMENTAL RESULTS AND VALIDATION OF DVMS

8.1. Simulations on Grid'5000	113
8.1.1. Consolidation	114
8.1.2. Infrastructure repair	119
8.2. Real experiments on Grid'5000	123
8.2.1. Experimental parameters	123
8.2.2. Results	125
8.3. Simulations with SimGrid	128
8.3.1. Experimental parameters	128
8.3.2. Results	129
8.4. Conclusion	130

CHAPTER 9. PERSPECTIVES AROUND DVMS

9.1. Completing the evaluations	133
9.1.1. Evaluating the amount of resources consumed by DVMS	133
9.1.2. Using real traces	134
9.1.3. Comparing DVMS with other decentralized approaches	135
9.2. Correcting the limitations	135
9.2.1. Implementing fault-tolerance	135
9.2.2. Improving event management	136

9.2.3. Taking account of links between virtual machines	138
9.3. Extending DVMS	138
9.3.1. Managing virtual machine disk images	139
9.3.2. Managing infrastructures composed of several data centers connected by means of a wide area network	139
9.3.3. Integrating DVMS into a full virtual infrastructure manager	140
9.4. Conclusion	140
CONCLUSION	141
BIBLIOGRAPHY	145
LIST OF TABLES	157
LIST OF FIGURES	159
INDEX	163

List of Abbreviations

ACO	Ant Colony Optimization
API	Application Programming Interface
BOINC	Berkeley Open Infrastructure for Network Computing
BVT	Borrowed Virtual Time scheduler
CFS	Completely Fair Scheduler
CS	Credit Scheduler
DOS	Distributed Operating System
DVMS	Distributed Virtual Machine Scheduler
EC2	Elastic Compute Cloud
EGEE	Enabling Grids for E-sciencE
EGI	European Grid Infrastructure
I/O	Input/Output
GPOS	General Purpose Operating System
IaaS	Infrastructure as a Service

IP	Internet Protocol
JRE	Java Runtime Environment
JVM	Java Virtual Machine
KSM	Kernel Shared Memory
KVM	Kernel-based Virtual Machine
LHC	Large Hadron Collider
MHz	Megahertz
MPI	Message Passing Interface
NFS	Network File System
NTP	Network Time Protocol
OSG	Open Science Grid
PaaS	Platform as a Service
SaaS	Software as a Service
SCVMM	System Center Virtual Machine Manager
URL	Uniform Resource Locator
VIM	Virtual Infrastructure Manager
VLAN	Virtual Local Area Network
VM	Virtual Machine
WLCG	Worldwide LHC Computing Grid
XSEDE	Extreme Science and Engineering Discovery Environment

Introduction

Context

Nowadays, increasing needs in computing power are satisfied by federating more and more computers (or nodes) to build distributed infrastructures.

Historically, these infrastructures have been managed by means of user-space frameworks [FOS 06, LAU 06] or distributed operating systems [MUL 90, PIK 95, LOT 05, RIL 06, COR 08].

Over the past few years, a new kind of software manager has appeared, managers that rely on system virtualization [NUR 09, SOT 09, VMW 10, VMW 11, APA 12, CIT 12, MIC 12, OPE 12, NIM 13]. System virtualization allows dissociating the software from the underlying node by encapsulating it in a virtual machine [POP 74, SMI 05]. This technology has important advantages for distributed infrastructure providers and users. It has especially favored the emergence of cloud computing, and more specifically of infrastructure as a service. In this model, raw virtual machines are provided to users, who can customize them by installing an operating system and applications.