

Computation and Storage in the Cloud

Understanding the Trade-Offs

Dong Yuan and Yun Yang

*Centre for Computing and Engineering Software Systems,
Faculty of Information and Communication Technologies,
Swinburne University of Technology,
Hawthorn, Melbourne, Australia*

Jinjun Chen

*Centre for Innovation in IT Services and Applications,
Faculty of Engineering and Information Technology,
University of Technology,
Sydney, Australia*



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Contents

Acknowledgements	ix
About the Authors	xi
Preface	xiii
1 Introduction	1
1.1 Scientific Applications in the Cloud	1
1.2 Key Issues of This Research	3
1.3 Overview of This Book	3
2 Literature Review	5
2.1 Data Management of Scientific Applications in Traditional Distributed Systems	5
2.1.1 Data Management in Grid	6
2.1.2 Data Management in Grid Workflows	8
2.1.3 Data Management in Other Distributed Systems	9
2.2 Cost-Effectiveness of Scientific Applications in the Cloud	10
2.2.1 Cost-Effectiveness of Deploying Scientific Applications in the Cloud	10
2.2.2 Trade-Off Between Computation and Storage in the Cloud	11
2.3 Data Provenance in Scientific Applications	12
2.4 Summary	12
3 Motivating Example and Research Issues	15
3.1 Motivating Example	15
3.2 Problem Analysis	17
3.2.1 Requirements and Challenges of Deploying Scientific Applications in the Cloud	17
3.2.2 Bandwidth Cost of Deploying Scientific Applications in the Cloud	18
3.3 Research Issues	19
3.3.1 Cost Model for Data Set Storage in the Cloud	19
3.3.2 Minimum Cost Benchmarking Approaches	20
3.3.3 Cost-Effective Storage Strategies	20
3.4 Summary	21
4 Cost Model of Data Set Storage in the Cloud	23
4.1 Classification of Application Data in the Cloud	23
4.2 Data Provenance and DDG	23

4.3	Data Set Storage Cost Model in the Cloud	25
4.4	Summary	27
5	Minimum Cost Benchmarking Approaches	29
5.1	Static On-Demand Minimum Cost Benchmarking Approach	30
5.1.1	CTT-SP Algorithm for Linear DDG	30
5.1.2	Minimum Cost Benchmarking Algorithm for DDG with One Block	32
5.1.2.1	Constructing CTT for DDG with One Block	33
5.1.2.2	Setting Weights to Different Types of Edges	34
5.1.2.3	Steps of Finding MCSS for DDG with One Sub-Branch in One Block	36
5.1.3	Minimum Cost Benchmarking Algorithm for General DDG	38
5.1.3.1	General CTT-SP Algorithm for Different Situations	38
5.1.3.2	Pseudo-Code of General CTT-SP Algorithm	39
5.2	Dynamic On-the-Fly Minimum Cost Benchmarking Approach	43
5.2.1	PSS for a DDG_LS	44
5.2.1.1	Different MCSSs of a DDG_LS in a Solution Space	44
5.2.1.2	Range of MCSSs' Cost Rates for a DDG_LS	45
5.2.1.3	Distribution of MCSSs in the PSS of a DDG_LS	47
5.2.2	Algorithms for Calculating PSS of a DDG_LS	50
5.2.3	PSS for a General DDG (or DDG Segment)	53
5.2.3.1	Three-Dimensional PSS of DDG Segment with Two Branches	54
5.2.3.2	High-Dimensional PSS of a General DDG	56
5.2.4	Dynamic On-the-Fly Minimum Cost Benchmarking	58
5.2.4.1	Minimum Cost Benchmarking by Merging and Saving PSSs in a Hierarchy	58
5.2.4.2	Updating of the Minimum Cost Benchmark on the Fly	61
5.3	Summary	64
6	Cost-Effective Data Set Storage Strategies	65
6.1	Data-Accessing Delay and Users' Preferences in Storage Strategies	65
6.2	Cost-Rate-Based Storage Strategy	66
6.2.1	Algorithms for the Strategy	67
6.2.1.1	Algorithm for Deciding Newly Generated Data Sets' Storage Status	67
6.2.1.2	Algorithm for Deciding Stored Data Sets' Storage Status Due to Usage Frequencies Change	68

6.2.1.3	Algorithm for Deciding Regenerated Data Sets’ Storage Status	68
6.2.2	Cost-Effectiveness Analysis	69
6.3	Local-Optimisation-Based Storage Strategy	69
6.3.1	Algorithms and Rules for the Strategy	70
6.3.1.1	Enhanced CTT-SP Algorithm for Linear DDG	70
6.3.1.2	Rules in the Strategy	72
6.3.2	Cost-Effectiveness Analysis	73
6.4	Summary	74
7	Experiments and Evaluations	75
7.1	Experiment Environment	75
7.2	Evaluation of Minimum Cost Benchmarking Approaches	75
7.2.1	Cost-Effectiveness Evaluation of the Minimum Cost Benchmark	76
7.2.2	Efficiency Evaluation of Two Benchmarking Approaches	77
7.3	Evaluation of Cost-Effective Storage Strategies	82
7.3.1	Cost-Effectiveness of Two Storage Strategies	82
7.3.2	Efficiency Evaluation of Two Storage Strategies	84
7.4	Case Study of Pulsar Searching Application	86
7.4.1	Utilisation of Minimum Cost Benchmarking Approaches	86
7.4.2	Utilisation of Cost-Effective Storage Strategies	87
7.5	Summary	90
8	Conclusions and Contributions	91
8.1	Summary of This Book	91
8.2	Key Contributions of This Book	92
	Appendix A: Notation Index	95
	Appendix B: Proofs of Theorems, Lemmas and Corollaries	97
	Appendix C: Method of Calculating and λ Based an Users’ Extra Budget	107
	Bibliography	109