Table of Contents

Foreword ..... **ix**

Preface ..... **xi**

I Introduction to Redundant Disk Array Architecture **1**

1 A Case for Redundant Arrays of Inexpensive Disks (RAID) **3**
David Patterson, Garth Gibson, and Randy Katz

2 Disk System Architectures for High Performance Computing **15**
Randy H. katz, and Garth A. Gibson and D. Patterson

3 The Performance of Parity Placements in Disk Arrays **35**
Edward K. Lee and Randy H. Katz

4 A Performance Comparison of RAID-5 and Log-Structured Arrays **55**
Jai Menon

II Advanced Disk Array Architectures **65**

5 Parity Logging Overcoming the Small Write Problem in Redundant Disk Arrays **67**
Daniel Stodolsky, Garth Gibson, and Mark Holland

6 Distributed RAID - A New Multiple Copy Algorithm **81**
Michael Stonebraker and Gerhard A. Schloss

7 The HP AutoRAID Hierarchical Storage System **90**
John Wilkes, Richard Golding, Carl Staelin, and Tim Sullivan

8 Scalable Distributed Log Structured Arrays **107**
Witold Litwin and Jai Menon

9 Comparison of Sparing Alternatives for Disk Arrays **117**
Jai Menon

10 Destage Algorithms for Disk Arrays with Non-Volatile Caches **129**
Anujan Varma and Quinn Jacobson

III Fault Tolerance Issues in Disk Arrays **147**

11 Failure Correction Techniques for Large Disk Arrays **149**
Garth A. Gibson, Lisa Hellerstein, Richard M. Karp, Randy H. Katz and David A. Patterson

12 Tolerating Multiple Failures in RAID Architectures with Optimal Storage and Uniform Declustering **161**
Guillermo A. Alvarez, Walter A. Burkhard, and Flaviu Cristian

13 Parity Declustering for Continuous Operation in Redundant Disk Arrays **173**
Mark Holland and Garth A. Gibson

14 The EVENODD Code and its Generalization **187**
Mario Blaum, Jim Brady, Jehoshua Bruck, Jai Menon, and Alexander Vardy

IV Caching and Prefetching **209**

15 RAPID-Cache — A Reliable and Inexpensive Write Cache for Disk I/O Systems **211**
Yiming Hu, Qing Yang, and Tycho Nightingale

16 Informed Prefetching and Caching **224**
R. Hugo Patterson, Garth A. Gibson, Eka Ginting, Daniel Stodolsky, and Jim Zelenka

17 Practical Prefetching Techniques for Multiprocessor File Systems **246**
David Kotz and Carla Schlatter Ellis

18 Design Issues of a Cooperative Cache with no Coherence Problems **260**
Toni Cortes, Sergi Girona and Jesús Labarta

19 Collective Buffering: Improving Parallel I/O Performance **272**
Bill Nitzberg and Virginia Lo

V Parallel File Systems **283**

20 The Vesta Parallel File System **285**
Peter F. Corbett and Dror G. Feitelson

21 The Zebra Striped Network File System **309**
John H. Hartman and John K. Ousterhout

22 *PPFS*: A High Performance Portable Parallel File System **330**
James V. Huber, Jr., Christopher L. Elford, Daniel A. Reed, Andrew A. Chien, and David S. Blumenthal

23 The Global File System **344**
Steven R. Soltis, Thomas M. Ruwart, Grant M. Erickson, Kenneth W. Preslan, and Matthew T. O'Keefe

24 Serverless Network File Systems **364**
Thomas E. Anderson, Michael D. Dahlin, Jeanna M. Neefe Matthews, David A. Patteson, Drew S. Roselli, and Randolph Y. Wang

VI Parallel I/O Systems **387**

25 Parallel I/O Subsystems in Massively Parallel Supercomputers **389**
Dror G. Feitelson, Peter F. Corbett, Sandra Johnson Baylor, and Yarsun Hsu

26 RAID-II: A High-Bandwidth Network File Server **408**
Ann L. Chervenak, Ken Shirriff, John H. Hartman, Ethan L. Miller, Srinivasan Seshan, Randy H. Katz, Ken Lutz, David A. Patterson, Edward K. Lee, Peter M. Chen, and Garth A. Gibson

27 Petal: Distributed Virtual Disks **420**
Edward K. Lee and Chandramohan A. Thekkath

28 A Cost-Effective, High-Bandwidth Storage Architecture 431
Garth A. Gibson, David F. Nagle, Khalil Amiri, Jeff Butler, Fay W. Chang, Howard Gobioff, Charles Hardin, Erik Riedel, David Rochberg, and Jim Zelenka

29 RAID-x: A New Distributed Disk Array for I/O-Centric Cluster Computing 445
Kai Hwang, Hai Jin and Roy S. C. Ho

30 Designing a Self-Maintaining Storage System 453
Satoshi Asami, Nisha Talagala, and David A. Patterson

31 Modeling and Evaluation of Fibre Channel Storage Area Networks 464
Xavier Molero, Federico Silla, Vicente Santonja and JoséDuato

VII Parallel I/O Programming Paradigms 475

32 Overview of the MPI-IO Parallel I/O Interface 477
Peter Corbett, Dror Feitelson, Sam Fineberg, Yarsun Hsu, Bill Nitzberg, Jean-Pierre Prost, Marc Snir, Bernard Traversat, and Parkson Wong

33 Disk Resident Arrays: An Array-Oriented I/O Library for Out-of-Core Computations 488
Ian Foster and Jarek Nieplocha

34 Active Disks: Programming Model, Algorithms and Evaluation 499
Anurag Acharya, Mustafa Uysal, and Joel Saltz

35 Disk-directed I/O for MIMD Multiprocessors 513
David Kotz

VIII Parallel I/O Applications and Environments 537

36 Applications-Driven Parallel I/O 539
Nicholas P. Galbreath, William D. Gropp, and David M. Levine

37 Comparing Multimedia Storage Architectures 548
Benoit A. Gennart and Roger D. Hersch

38 High Availability in Clustered Multimedia Servers 555
Renu Tewari, Daniel M. Dias, Rajat Mukherjee, and Harrick M. Vin

39 An Architecture for a Scalable High-Performance Digital Library 566
R. Grossman, X. Qin, W. Xu, H. Hulen, and T. Tyler

40 I/O Requirements of Scientific Applications: An Evolutionary View 576
Evgenia Smirni, Ruth A. Aydt, Andrew A. Chien, Daniel A. Reed

41 Mitra: A Scalable Continuous Media Server 595
Shahram Ghandeharizadeh, Roger Zimmermann, Weifeng Shi, Reza Rejaie, Douglas J. Ierardi, and Ta-Wei Li

IX Emerging Technologies and Future Trends 615

42 An Introduction to the InfiniBand Architecture 617
Gregory F. Pfister

43 XML, Hyper-media, and Fortran I/O 633
Dror G. Feitelson and Tomer Klainer

44 I/O Programming Paradigms: Past and Future 645
Mahmut Taylan Kandemir and Alok Choudhary

45 Scientific Applications using Parallel I/O **655**
Ron Oldfield and David Kotz

Index **667**