《可扩展并行计算--技术、结构与编程》

图书基本信息

书名：《可扩展并行计算--技术、结构与编程》

13位ISBN编号：9787111071761

10位ISBN编号：711107176X

出版时间：1999-05

出版社：机械工业出版社

作者：黄铠,等

页数：802

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu111.com

《可扩展并行计算--技术、结构与编程》

作者简介

About the Authors

Kai Hwang presently holds a Chair Professor of Computer Engineering at the Universityof Hong Kong (HKU), while taking a research leave from the University of Southem Califomia (USC). He has been engaged in higher education and computer research for 26 years, after eaming the Ph.D. in Electrical Engineering and Computer Science from the University ofCalifornia at Berkeley. His work on this book started at the USC and was mostly completed at the HKU.

An lEEE Feilow,he has published extensively in the areas ofcomputer architecture, digital arithmetic, parallel processing, and distributed computing. He is the founding Editor-in-Chiefof the Jownal ofParallelandDistributed Computing. He has chaired the intemational confeaeoces: lCPP86, ARITH-7, IPP96, ICAPP 96, and HPCA-4 in 1998. He has received several achicvement awards for butstanding research and academic contributions to the field of parallel computing.

He has lectured worldwide and perrtormed consulting and advisory work for US Nationai Academy of SciencesMIT Lincoln Laboratory, IBM Fishkill, TriTech in Singapore, Pujitsu and ETL in Japan, GMD in Germany, CERN Schoolof Computing, and Academia Sinica in Chiria. Presently, he leads a research group at HKU in developing aan ATM-based multicomputer cluster for high-performance computing and distributed multimedia, Intranet, ahd Intemet appliCations.

Zhiwei Xu is a Professor and ChiefArchitect at the National Center for Intelligent Computing Systems (NClC), Chinese Academy ofSciences, Beijing, China. He is also a Honorary Research Fellow ofHKU. He received a Ph.D. in Computer Engineering from the University of Southem Califomia. He participated in the STAP/MPP benchmark projects led by Dr. Hwang at USC and HKU during the past 5 years.

He has taught at the Rutgers University and New Yotk Polytechnfc University. He has published in theareas of parallet languages, pipelined vector processmg, and benchmard evaluation ofmassively parallel processors. Presently, he leads a design group at NClC in building a series of cluster-based superservers in China. His current research interest lies mainly in network-based cluster computing and the software environments for parallel programming.

《可扩展并行计算--技术、结构与编程》

书籍目录

Table of Contents
About the Authors Iv
Foreword xv
Preface xvI
Guide to Instructors/Readers xIx
Part l Scalability and Clustering
Chapter l Scalable Computer Platforms and Models
l.l Evolution ofComputer Architecture
l.l.l Computer Generations
l.l.2 Scalable Computer Architectures
l.l.3 Converging System Architectures
l.2 Dimensions ofScalability
l.2.l Resource Scalability
1.2.2 Application Scalability
1.2.3 Technology Scalability
1.3 Parallel Computer Models
l.3.l Semantic Attributes
1.3.2 Performance Attributes
l.3.3 Abstract Machine Model
l.3.4 Physical Machine Model
1.4 BasicConceptsofClustering
l.4.l Cluster Characteristics
1.4.2 Architectural Comparisons
l.4.3 BenefitsandDifficultiesofChisters
1.5 Scalable Design Principles
l.5.l Principle of Independence
l.5.2 Principle ofBalanced Design
l.5.3 Design for Scalability
1.6 Bibliographic Notes and Problems
Chapter 2 Basics of Parallel Programming
2.1 Parallel Programming Overview
2.l.l Why Is Parallel Programming Difficult?
2.1.2 Parallel Programming Environments
2.l.3 Parallel Programming Approaches
2.2 Processes, Tasks, and Threads
2.2.1 DefinitionsofanAbstractProcess
2.2.2 Execution Mode
2.2.3 Address Space
2.2.4 Process Context
2.2.5 Process Descriptor
2.2.6 Process Control
2.2.7 Variations ofProcess
2.3 Parallelism Issues
2.3.1 Homogeneity in Processes
2.3.2 Static versus Dynamic Parallelism
2.3.3 Process Grouping.
2.3.4 Allocation Issues

2.4 Interaction/Communication Issues
2.4.1 Interaction Operations
2.4.2 Interaction Modes
2.4.3 Interaction Pattems
2.4.4 Cooperative versus Competitive Interactions
2.5 Semantic Issues in ParaUel Programs
2.5.1 Program Tennination
2.5.2 Determinacy ofPrograms
2.6 Bibliographic Notes and Problems
Chapter 3 Performance Metrics and Benchmarks
3.1 System and Applicatioo Benchmarks
3.1.1 Micro Benchmarks
3.1.2 Parallel Computing Benchmarks
3.1.3 Business and TPC Benchmarks
3.1.4 SPEC Benchmark Family
3.2 Perfonnance versus Cost
3.2.1 Execution Time and Throughput
3.2.2 Utilization and Cost-Effectiveness
3.3 Basic Performance Metrics
3.3.1 Workload and Speed Metrics
3.3.2 Caveats in Sequential Performance
3.4 PerfonnanceofParallelComputers
3.4.1 Computatiomal Characteristics
3.4.2 Parallelism and Interaction Overheaas
3.4.3 Overhead Quantification
3.5 Performance of Parallel Programs
3.5.1 Performance Metrics
3.5.2 Available Parallelism in Benchmarks
3.6 Scalability and Speedup Analysis
3.6.1 Amdahl's Law: Fixed Problem Size
3.6.2 Gustafson's Law: FixedTime
3.6.3 Sun and Ni's Eaw: Memory Bounding
3.6.4 Isoperformance Models
3.7 Bibliographic Notes-aod Problems
Part II Enabling Technologies
Chapter 4 Microprocessors as Building Blocks
4.1 System Development Trends
4.l.l Advances in Hardware
4.1.2 Advances in Software
4.l.3 Advances in Applications
4.2 PrinciplesofProcessorDesign
4.2.1 BasicsoflnstructionPipeline
4.2.2 From ClSC to RlSC and Beyond
4.2.3 Architectural Enhancement Approaches
4.3 Microprocessor Architecture Families
4.3.1 Major Architecture Familiei
4.3.2 Superscalar versus Superpipelined Processors
4.3.3 Embedded Microprocessors
4.4 Case Studies of Microprocessors

4.4.l Digital's Alpha 21 164 Microprocessor
4.4.2 Intel Pentium Pro Processor
4.5 Post-RlSC, Multimedia, and VLlW
4.5.1 Post-RlSC Processor Features
4.5.2 Multimedia Extensions
4.5.3 TheVLlWArchitecture
4.6 The Future of Microprocessors
4.6.l Hardware Trends and Physical Limits
4.6.2 Future Workloads and Challenges ,
4.6.3 Future Microprocessor Architectures
4.7 Bibliographic Notes and Problems
Chapter 5 Distributed Memory and Latency Tolerance
5.1 Hierarchical Memory Technology
5.l.l Characteristics of Storage Devices
5.1.2 Memory Hierarchy Properties
5.l.3 Memory Capacity Planning
5.2 Cache Cohereoce Protocob
5.2.1 Cache Coherency Problem
5.2.2 Snoopy Coherency Protocols
5.2.3 The MESl Snoopy Protocol
5.3 Shared-Memory Consistency
5.3.1 Memory Event Ordering
5.3.2 Memory Consistency Models
5.3.3 Relaxed Memory Models
5.4 Distributed Cache/Memory Architecture
5.4.l NORMA, NUMA, COMA, and DSM Models
5.4.2 Directory-Based Coherency Protocol
5.4.3 The Stanford Dash Multiprocessor
5.4.4 Directory-Based Protocol in Dash
5.5 Latency Tolerance Techniques 250
5.5.1 Latency Avoidance, Reduction, and Hiding
5.5.2 Distributed Coherent Caches
5.5.3 Data Prefetching Strategies
5.5.4 Effects of Relaxed Memory Consistency
5.6 Multithreaded Latency Hiding
5.6.1 Multithreaded Processor Model
5.6.2 Context-Switehing Policies
5.6.3 Combining Latency Hiding Mechanisms
5.7 Bibliographic Notes and Problems
Chapter 6 System Interconnects and Gigabit Networks
6.1 Basics of Interconnection Network
6.1.1 Interconnection Environmnents
6.1.2 Networik Components
6.1.3 Network Characteristics
6.1.4 Network Performance Metrics
6.2 Network Topologies and Properties
6.2.1 Topological and Functional Properties
6.2.2 Routing Schemes and Functions
6.2.3 Networidng Topologies

6.3 Buses, Crossbar, aod Multistage Switehes
6.3.1 Multiprocessor Buses
6.3.2 Crossbar Switches
6.3.3 Multistage Interconnection Networks
6.3.4 Comparison of Switched Interconnects
6.4 Gigabit Network Technologies
6.4.1 Fiber Channel and FDDI Rings
6.4.2 Fast Ethemet and Gigabit Ethemel
6.4.3 Myrinet for SAN/LAN Construction
6.4.4 HiPPI and SuperHiPPI
6.5 ATM Switches and Networks
6.5.1 ATM Technology
6.5.2 ATMNetworkInterfaces
6.5.3 Four Layers of ATM Architecture
6.5.4 ATM Intemetwork Connectivity
6.6 Scalable Coherence Interfaee
6.6.1 SCI Interconmects
6.6.2 Implementation Issues
6.6.3 SCI Coherence Protocol
6.7 ComparisoD of Network Technologies
6.7.1 Standard Networks and Perspectives
6.7.2 Network Performance arid Applications
6.8 Bibliographic Notes and Problems
Chapter 7 Threading, Synchronization, and Communication
7.1 Software Multithreading
7.1.1 TheThreadConcept
7.1.2 Threads Management
7.1.3 Thread Synchronization
7.2 Synchronization Mechanisms
7.2.l Atomicity versus Mutual Exclusion
7.2.2 High-Level Synchronization Constructs
7.2.3 Low-Level Synchronization Primitiyes
7.2.4 Fast Locking Mechanisms
7.3 The TCP/lP Communication Protocol Suite
7.3.l Features of The TCP/IP Suite
7.3.2 UDP.TCP.andlP
7.3.3 The Sockets Interface
7.4 Fast and Efficient Coramunication
7.4.l Key Problems in Communication
7.4.2 The Log P Communication Model
7.4.3 Low-Level Communications Support
7.4.4 Communication Algorithms
7.5 Bibliographic Notes and Problems
Part lll Systems Architecture
Chapter 8 Symmetric and CC-NUMA Multiprocessors
8.1 SMP and CC-NUMA Technology
8.l.l Multiprocessor Architecture
8.1.2 Commercial SMP Servers
8.1.3 ThelntelSHVServerBoara

8.2 Sun Ultra Enterprise lOOOO System
8.2.l The Ultra E- l 0000 Architecture
8.2.2 System Board Architecture
8.2.3 Scalability and Availability Support
8.2.4 Dynamic Domains and Performance
8.3 HP/Convex Exemplar X-Class
8.3.l The Exemplar X System Architecture
8.3.2 Exemplar Software Environment
8.4 The Sequent NUMA-Q 2000
8.4.l The NUMA-Q 2000 Architecture
8.4.2 Software Environment ofNUMA-Q
8.4.3 PerformanceoftheNUMA-Q
8.5 The SGl/Cray Origin 2000 Superserver
8.5.l Design Goals of Origin 2000 Series
8.5.2 The Origin 2000 Architecture
8.5.3 The Cellular IRIX Environment
8.5.4 PerformanceoftheOrigin2000
8.6 Comparison ofCC-NUMA Architectures
8.7 Bibliographic Notes and Problems
Chapter 9 Support of Clustering and Availability
9.1 Challenges in Clustering
9.1.1 Classification of Clusters
9.1.2 Cluster Architectures
9.1.3 Cluster Design Issues
9.2 Availability Support for Clusteriog
9.2.1 The Availability Concept
9.2.2 Availability Techniques
9.2.3 Checkpointing and Failure Recbvery
9.3 Support for Single System Image
9.3.1 Single System Image Layers
9.3.2 Single Entry and Single File Hierarchy
9.3.3 Single 1/0, Networking, and Memory Space
9.4 Single System Image in Solaris MC
9.4.1 Global File System
9.4.2 Global Process Management
9.4.3 Single 1/O System Imnage
9.5 Job Management in Clusters
9.5.1 Job Management System
9.5.2 Survey of Job Management Systems
9.5.3 Load-Sharing Facility (LSF)
9.6 Bibliographk Notes and ProNems
Chapter 10 Clusters of Servers and Workstations
10.1 Cluster Products and Research Projects
10.1.1 Supporting Trend ofCluster Products
10.1.2 ClusterofSMPServers
10.1.3 ClusterResearchProjects
10.2 Microsoft Wolfpack for NT Clusters
10.2.1 Microsoft Wolfpack Configurations
10.2.2 Hot Standby Multiserver Clusters

10.2.3 Active Availability Clusters
10.2.4 Fault-Tolerant Multiserver Cluster
10.3 The IBM SP System
10.3.1 Design Goals and Strategies
10.3.2 The SP2 System Architecture
10.3.3 1/o and Intemetworking.
10.3.4 The SP System Software
10.3.5 The SP2 and Beyond
10.4 The Digital TruCIuster
10.4.1 The TmCluster Architecture
10.4.2 The Memory Channel Interconnect
10.4.3 Programming the TruCluster
10.4.4 The TruCluster System Software
10.5 The Berkeley NOW Project
10.5.1 Active Messages for Fast Communication
10.5.2 GLUnix for Global Resource Management
10.5.3 ThexFSServerlessNetworkFileSystem
10.6 TreadMarks: A Software-lmplemented DSM Cluster
10.6.1 Boundary Conditions
l0.6.2 User Interface for DSM
l0.6.3 Implementation Issues
l0.7 Bibliographic Notes and Problems
Chapter ll MPP Architecture and Performance
ll.l An Overview of MPP Technology
ll.l.l MPP Characteristics and Issues
ll.l.2 MPP Systems - An Overview
ll.2 The Cray T3E System
ll.2.l The System Architecture of T3E
ll.2.2 The System Software in T3E
11.3 New Generation of ASCl/MPPs
ll.3.l ASCl Scalable Design Strategy
ll.3.2 Hardware and Software Requirements
ll.3.3 Contracted ASCI/MPP Platforms
11.4 Intel/Sandia ASCl Option Red
ll.4.l The Option Red Architecture
ll.4.2 Option Red System Software
11.5 Parallel NAS Benchmark Results
ll.5.l The NAS Parallel Benchmarks
ll.5.2 Superstep Structure and Granulanty
ll.5.3 Memory, VO, and Communications
11.6 MPl and STAP Benchmark Results
ll.6.l MPl Performance Measurements
ll.6.2 MPl Latency and Aggregate Bandwidth
ll.6.3 STAP Benchmark Evaluation of MPPs
ll.6.4 MPP Architectural Implications
11.7 Bibliographic Notes and Problems
Part IV Part IV Parallel Programming
Chapter 12 Parallel Paradigms and Programming Models
12.1 Paradigms and Programmability

12.1.1 Algorithmic Paradigms
12.1.2 Programmability Issues
12.1.3 Parallel Programming Examples
12.2 Parallel Programming Models
12.2.1 Implicit Parallelism
12.2.2 Explicit Parallel Modeis
l2.2.3 ComparisonofFourModels
l2.2.4 Other Parallel Programming Models
12.3 Shared-Memory ProgrammiBg
12.3.1 The ANSI X3H5 Shared-Memory Model
12.3.2 ThePOSIX Threads(Pthreads)Model
12.3.3 The OpenMP Standard
12.3.4 TheSGIPowerCModel
12.3.5 Cll: A Structured Parallel C Language
12.4 Bibliographic Notes and Problems
Chapter 13 Message-Passing Programmmg
13.1 The Message-Passing Paradigm
13.1.1 Message-Passing Libraries
13.1.2 Message-Passing Modes
13.2 Message-Passing Interface (MPI)
13.2.1 MPIMessages
13.2.2 Message Envelope in MPI
13.2.3 Point-to-Point Communications
13.2.4 Collective MPI Communications
13.2.5 The MP1-2 Extensions
13.3 Parallel Virtual Machine (PVM)
13.3.1 Virtual Machine Construction
13.3.2 Process Management in PVM
13.3.3 Communication with PVM
13.4 Bibliographic Notes and Problems
Chapter 14 Data-ParalleI Programming
14.1 The Data-Parallel Model
14.2 The Fortran 90 Approach
14.2.1 Parallel Array Operations
14.2.2 Intrinsic Functions in Fortran 90
14.3 High-Performance Fortran
14.3.1 Support for Data Parallelism
14.3.2 DataMappinginHPF
14.3.3 SummaryofFortran90andHPF
14.4 Other Data-Parallel Approaches
14.4.1 Fortran 95andFortran200l
14.4.2 ThepC++andNeslApproaches
14.5 Bibliographic Notes and Problems
Bibliography
Web Resources List
Subject Index
Author Index

《可扩展并行计算--技术、结构与编程》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu111.com