

1. Record Nr.	UNINA9911019503303321
Autore	Majumdar Shikharesh
Titolo	Resource Management on Distributed Systems : Principles and Techniques
Pubbl/distr/stampa	Newark : , : John Wiley & Sons, Incorporated, , 2024 ©2025
ISBN	9781119912958 1119912954 9781119912965 1119912962 9781119912941 1119912946
Edizione	[1st ed.]
Descrizione fisica	1 online resource (323 pages)
Disciplina	004.36
Soggetti	Resource allocation Parallel processing (Electronic computers)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di contenuto	Cover -- Title Page -- Copyright -- Contents -- About the Author -- Preface -- Acknowledgments -- Chapter 1 Introduction -- 1.1 Introduction to Distributed and Parallel Computing -- 1.2 Types of Computing Environments -- 1.3 Units of Computation -- 1.3.1 Process -- 1.3.2 Threads -- 1.3.3 Resource Management Operations -- 1.4 Principles Underlying Resource Management -- 1.4.1 Principle 1: Use Knowledge of Application/Workload Characteristics -- 1.4.2 Principle 2: Monitor and Adjust -- 1.4.3 Principle 3: Use Knowledge of System Characteristics -- 1.4.4 Principle 4: Perform Load Balancing -- 1.4.5 Principle 5: Static Versus Dynamic -- 1.5 Evolution of Distributed Systems -- 1.5.1 Nodes Communicating via Remote Procedure Calls (RPC) -- 1.5.2 Distributed Object Based Computing Systems -- 1.5.3 Parallel and Cluster Computing -- 1.5.4 ServiceOriented Architecture -- 1.5.5 Grids -- 1.5.6 Clouds -- 1.5.7 Edge Computing -- 1.5.8 Smart Facilities -- 1.6 Summary -- 1.6.1 Book Components -- References -- Chapter 2 Characterization of Parallelism in Applications

-- 2.1 Introduction -- 2.2 The Precedence Graph Model -- 2.3 Graph Based Characteristics -- 2.3.1 Parallelism Profile -- 2.3.2 Shape -- 2.4 SinglePoint Characteristics -- 2.4.1 Maximum Parallelism -- 2.4.2 Fraction of Sequential Work -- 2.4.3 Average Parallelism -- 2.5 Performance Metrics -- 2.5.1 Completion Time -- 2.5.1.1 Processor Sharing Scheduling -- 2.5.2 Speedup -- 2.5.3 Efficiency -- 2.5.4 Completion TimeEfficiency Profile -- 2.5.4.1 The Knee -- 2.6 Impact of Parallelism Characteristics on Performance -- 2.6.1 Impact of Fraction of Sequential work -- 2.6.2 Impact of Average Parallelism -- 2.6.2.1 Bounds Based on Multiple Characteristics -- 2.7 Energy Performance TradeOff -- 2.7.1 Relationship Between Energy Consumption and Speedup. -- 2.7.2 Relationship Between Energy and Fraction of Sequential Work -- 2.7.3 Relationship Between Energy and Average Parallelism -- 2.8 Summary -- Exercises -- References -- Chapter 3 Resource Management Techniques for Distributed Computing Systems -- 3.1 Resource Allocation -- 3.1.1 Graham's Anomaly -- 3.1.2 The Impact of Processor Allocation on Performance -- 3.1.3 Optimal Allocation Techniques -- 3.1.3.1 Hu's Algorithm -- 3.1.4 Heuristic Techniques for Processor Allocation -- 3.1.4.1 Largest Processing Time First -- 3.1.4.2 The Multifit Technique -- 3.2 Task/Process Scheduling -- 3.2.1 Scheduling of Ready Tasks on a Single Processor -- 3.2.1.1 Scheduling Tasks with SLAs -- 3.2.1.2 Performance Analysis -- 3.3 Grid Scheduling with Deadlines -- 3.3.1 Performance Analysis -- 3.4 Scheduling on Client-Server Systems -- 3.4.1 Performance Analysis -- 3.4.2 Software Bottlenecks -- 3.5 Summary -- Exercises -- References -- Chapter 4 Resource Management on Systems Subjected to Uncertainties Associated with Workload and System Parameters -- 4.1 Introduction -- 4.2 Handling Errors Associated with User Estimates of Job Execution Times -- 4.2.1 Overestimation of Job Execution Times -- 4.2.1.1 Schedule Exceptions Manager -- 4.2.1.2 Prescheduling Engine -- 4.3 Underestimation of Job Execution Times -- 4.3.1 Performance Analysis -- 4.4 Handling Uncertainties Associated with the Local Scheduling Policy -- 4.5 Any Schedulability Criterion -- 4.5.1 Application of AS Criterion: An Example -- 4.6 Matchmaking in the Dark: AS CriterionBased Matchmaking -- 4.6.1 Hybrid Matchmaking -- 4.6.1.1 Independent -- 4.6.1.2 Combined -- 4.6.2 Performance Comparison of Independent and Combined Matchmaking -- 4.7 Soft Advance Reservation Requests -- 4.7.1 Computation of System Generated Estimate of Job Execution Time -- 4.7.2 Performance Evaluation -- 4.8 Summary -- Exercises. -- References -- Chapter 5 Resource AutoScaling -- 5.1 Introduction -- 5.1.1 An Example System -- 5.2 Request Characteristics -- 5.3 Horizontal AutoScaling -- 5.3.1 Reactive AutoScaling -- 5.3.1.1 Reactive AutoScaling Algorithm -- 5.3.2 Performance Analysis of the Reactive AutoScaling Algorithm -- 5.3.3 Proactive AutoScaling -- 5.3.3.1 The Proactive AutoScaling Algorithm -- 5.3.4 Performance Analysis of the Proactive AutoScaling Algorithm -- 5.4 Hybrid Auto Scaling -- 5.4.1 Performance Analysis of the Hybrid AutoScaling Algorithm -- 5.4.2 Comparison with Pure Reactive and Proactive Approaches -- 5.5 Summary -- Exercises -- References -- Chapter 6 Resource Management for Systems Running MapReduce Jobs -- 6.1 Introduction -- 6.2 MapReduce -- 6.3 Resource Management Techniques for MapReduce Job Requests to be Satisfied on a Best Effort Basis -- 6.4 Resource Management Techniques for MapReduce Job Requests with Service Level Agreements -- 6.4.1 The BudgetBased MapReduce Resource Management Technique -- 6.4.1.1 HighLevel Description -- 6.4.2 Performance Analysis -- 6.5 The Constraint

ProgrammingBased MapReduce Resource Management Technique --
 6.5.1 The MRCPRM Algorithm -- 6.5.2 Performance Analysis for MRCP
 RM -- 6.6 Errors Associated with User Estimates of Task Execution
 times -- 6.6.1 The Prescheduling ErrorHandling Technique -- 6.6.1.1
 Performance Analysis -- 6.6.2 The Runtime ErrorHandling Technique
 -- 6.6.2.1 Performance Analysis -- 6.7 Summary -- Exercises --
 References -- Chapter 7 Energy Aware Resource Management -- 7.1
 Introduction -- 7.1.1 Dynamic Voltage Frequency Scaling -- 7.2 DVFS
 Based Resource Management Techniques -- 7.3 The EAMRRM
 Algorithm -- 7.3.1 Performance Analysis of EAMRRM -- 7.3.1.1 The
 Impact of Arrival Rate of Jobs -- 7.3.1.2 The Effect of Map Task
 Execution Times -- 7.3.1.3 The Impact of Number of Resources.
 7.4 Configurable Resource Manager for Processing a Batch of
 MapReduce Jobs -- 7.4.1 Constraint Program for the Configurable
 Resource Manager -- 7.5 Performance Analysis of CRM -- 7.5.1 Effect
 of Batch Completion Time Bound -- 7.5.2 Impact of the Missed
 Deadline Ratio Bound -- 7.5.3 The Impact of Deadline Multiplier -- 7.6
 Reducing the Number of Active Servers -- 7.7 Summary -- Exercises --
 References -- Chapter 8 Streaming Data and Complex Event Processing
 -- 8.1 Introduction -- 8.2 Management of Streaming Data -- 8.3
 Dynamic PriorityBased Scheduling -- 8.3.1 The Spark Streaming Data
 Processing Platform -- 8.3.1.1 The Spark Streaming System -- 8.4
 DataDriven Priority Scheduler (DDPS) -- 8.4.1 Algorithm 1 -- 8.4.2
 Algorithm 2 -- 8.4.3 Algorithm 3 -- 8.4.4 Performance Analysis -- 8.5
 Multitennant Systems -- 8.5.1 Apache Storm -- 8.5.2 Resource
 Management on a Multitenant Storm Cluster -- 8.5.2.1 Isolation
 Scheduler -- 8.5.2.2 Static PriorityBased Scheduler (SPS) -- 8.5.2.3
 Dynamic Priority Scheduler -- 8.5.2.4 Performance Analysis -- 8.6
 Complex Event Processing -- 8.6.1 CQL Query and CEP -- 8.6.2 CEP
 Architecture -- 8.6.3 Example CEP Systems -- 8.6.4 CEP Platforms --
 8.6.4.1 Apache Siddhi -- 8.7 Remote Patient Monitoring System --
 8.7.1 The SCEP System -- 8.7.2 The Mobile CEP (MCEP) System -- 8.7.3
 Performance Analysis -- 8.7.3.1 Remaining Battery Life -- 8.7.3.2 CEP
 Latency -- 8.8 Summary -- Exercises -- References -- Chapter 9 Data
 Indexing and Filtering Techniques for Big Data Systems -- 9.1
 Introduction -- 9.2 Harnessing Big Data -- 9.3 Data Indexing -- 9.4
 Inverted Index -- 9.4.1 Enhancements to the Inverted Index Technique
 -- 9.5 GraphBased Indexing -- 9.6 Boolean AND Queries -- 9.7
 Performance Analysis -- 9.7.1 Search Latency -- 9.7.1.1 The Impact of
 Dcount on Search Latency -- 9.7.1.2 Impact of SKcount on Search
 Latency.
 9.7.2 Indexing Overhead -- 9.7.2.1 Indexing Time -- 9.7.2.2 Memory
 Usage -- 9.8 Data Filtering -- 9.8.1 Processing of Large Volumes of
 Data -- 9.9 Parallel Processing Platforms -- 9.9.1 Spark Architecture --
 9.10 Motivations for Data Reduction -- 9.10.1 Reducing Data Volume
 -- 9.11 Data Filtering -- 9.11.1 Basic Approach for Data Filtering --
 9.11.2 The Filtering Algorithm -- 9.11.3 Search Method -- 9.12
 Performance Analysis -- 9.12.1 Node Parallelism vs. Core Parallelism --
 9.13 Streaming Data -- 9.13.1 System Performance -- 9.14 Handling
 User Preferences Comprising Keywords Connected by Boolean
 Operators -- 9.14.1 Performance Analysis -- 9.15 Summary --
 Exercises -- References -- Chapter 10 SensorBased Systems -- 10.1
 Introduction -- 10.1.1 Architecture for CloudBased Smart Facilities
 Management -- 10.2 Middleware Services -- 10.3 SensorBased Bridge
 Management -- 10.4 Research Collaboration Platform for Management
 of SensorBased Smart Facilities -- 10.5 Resource Management on
 Wireless Sensor Networks -- 10.6 Scheduling on WSNs -- 10.6.1
 Knowledge Free Algorithms -- 10.6.2 KnowledgeBased Algorithms --

10.6.2.1 Description of the KnowledgeBased Scheduling Algorithms --
10.6.3 Performance of Scheduling Algorithms -- 10.6.3.1 Performance
of Algorithms That Use a Single Characteristic -- 10.6.3.2 Performance
of Algorithms That Use the Knowledge of Both Application and System
Characteristics -- 10.7 Sensor Allocation -- 10.7.1 Dynamic Allocation
Algorithms -- 10.7.1.1 KnowledgeFree Algorithm -- 10.7.1.2
KnowledgeBased Algorithms -- 10.7.2 Static Allocation Algorithms --
10.7.2.1 KnowledgeFree Allocation Algorithms -- 10.7.2.2 Allocation
Algorithms Based on the Knowledge of Applications -- 10.7.3
Performance of Sensor Allocation Algorithms -- 10.7.3.1 Performance
of Dynamic Allocation Algorithms.
10.7.3.2 Performance of Static Allocation Algorithms.

Sommario/riassunto

This book focuses on resource management techniques for distributed systems, which are essential for efficiently utilizing computational resources ranging from clusters to clouds. It explores principles and algorithms necessary for managing resources in parallel and distributed systems, addressing key topics such as process and thread management, grid and cloud computing, and innovative approaches like edge computing. The volume is structured into two parts, with the first discussing the evolution of distributed systems and the second highlighting specific resource management challenges such as allocation, scheduling, auto-scaling, and handling uncertainties in workload. Targeted at students, researchers, and industry professionals, the book aims to provide a comprehensive understanding of resource management to optimize performance and energy consumption in distributed computing environments.
