1. Record Nr. UNINA9910819467403321

Autore	Taniar David
Titolo	High-performance parallel database processing and grid databases / / David Taniar, Clement H.C. Leung, Wenny Rahayu, Sushant Goel
Pubbl/distr/stampa	Hoboken, NJ, : J. Wiley, c2008
ISBN	1-281-83125-5 9786611831257 0-470-39136-7 0-470-39135-9
Edizione	[1st ed.]
Descrizione fisica	1 online resource (574 p.)
Collana	Wiley Series on Parallel and Distributed Computing ; ; v.67
Altri autori (Persone)	LeungClement H. C RahayuJohanna Wenny GoelSushant
Disciplina	004.35
Soggetti	High performance computing Parallel processing (Electronic computers) Computational grids (Computer systems)
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references.
Nota di contenuto	High-Performance Parallel Database Processing and Grid Databases; Contents; Preface; Part I Introduction; 1. Introduction; 1.1. A Brief Overview: Parallel Databases and Grid Databases; 1.2. Parallel Query Processing: Motivations; 1.3. Parallel Query Processing: Objectives; 1.3.1. Speed Up; 1.3.2. Scale Up; 1.3.3. Parallel Obstacles; 1.4. Forms of Parallelism; 1.4.1. Interquery Parallelism; 1.4.2. Intraquery Parallelism; 1.4.3. Intraoperation Parallelism; 1.4.4. Interoperation Parallelism; 1.4.5. Mixed Parallelism-A More Practical Solution; 1.5. Parallel Database Architectures 1.5.1. Shared-Memory and Shared-Disk Architectures1.5.2. Shared-Nothing Architecture; 1.5.3. Shared-Something Architecture; 1.5.4. Interconnection Networks; 1.6. Grid Database Architecture; 1.7. Structure of this Book; 1.8. Summary; 1.9. Bibliographical Notes; 1.10. Exercises; 2. Analytical Models; 2.1. Cost Models; 2.2. Cost Notations; 2.2.1. Data Parameters; 2.2.2. Systems Parameters; 2.2.3. Query Parameters; 2.2.4. Time Unit Costs; 2.2.5. Communication Costs; 2.3.

Skew Model; 2.4. Basic Operations in Parallel Databases; 2.4.1. Disk Operations; 2.4.2. Main Memory Operations
2.4.3. Data Computation and Data Distribution2.5. Summary; 2.6. Bibliographical Notes; 2.7. Exercises; Part II Basic Query Parallelism; 3. Parallel Search; 3.1. Search Queries; 3.1.1. Exact-Match Search; 3.1.2. Range Search Query; 3.1.3. Multiattribute Search Query; 3.2. Data Partitioning; 3.2.1. Basic Data Partitioning; 3.2.2. Complex Data Partitioning; 3.3. Search Algorithms; 3.3.1. Serial Search Algorithms; 3.3.2. Parallel Search Algorithms; 3.4. Summary; 3.5. Bibliographical Notes; 3.6. Exercises; 4. Parallel Sort and GroupBy; 4.1. Sorting, Duplicate Removal, and Aggregate Queries
4.1.1. Sorting and Duplicate Removal4.1.2. Scalar Aggregate; 4.1.3. GroupBy; 4.2. Serial External Sorting Method; 4.3. Algorithms for Parallel External Sort; 4.3.1. Parallel Merge-All Sort; 4.3.2. Parallel Binary-Merge Sort; 4.3.3. Parallel Redistribution Binary-Merge Sort; 4.3.4. Parallel Redistribution Merge-All Sort; 4.3.5. Parallel Partitioned Sort; 4.4. Parallel Algorithms for GroupBy Queries; 4.4.1. Traditional Methods (Merge-All and Hierarchical Merging); 4.4.2. Two-Phase Method; 4.4.3. Redistribution Method; 4.5. Cost Models for Parallel Sort
4.5.1. Cost Models for Serial External Merge-Sort4.5.2. Cost Models for Parallel Merge-All Sort; 4.5.3. Cost Models for Parallel Binary-Merge Sort; 4.5.4. Cost Models for Parallel Redistribution Binary-Merge Sort; 4.5.5. Cost Models for Parallel Redistribution Merge-All Sort; 4.5.6. Cost Models for Parallel Partitioned Sort; 4.6. Cost Models for Parallel GroupBy; 4.6.1. Cost Models for Parallel Two-Phase Method; 4.6.2. Cost Models for Parallel Redistribution Method; 4.7. Summary; 4.8. Bibliographical Notes; 4.9. Exercises; 5. Parallel Join; 5.1. Join Operations; 5.2. Serial Join Algorithms
5.2.1. Nested-Loop Join Algorithm

Sommario/riassunto

The latest techniques and principles of parallel and grid database processing The growth in grid databases, coupled with the utility of parallel query processing, presents an important opportunity to understand and utilize high-performance parallel database processing within a major database management system (DBMS). This important new book provides readers with a fundamental understanding of parallelism in data-intensive applications, and demonstrates how to develop faster capabilities to support them. It presents a balanced treatment of the theoretical and practical aspects of high-