CONTENTS

PREFACE vii

1

THE PROBLEM 1
1.1 Transactions 1
1.2 Recoverability 6
1.3 Serializability 11
1.4 Database System Model 17

2

SERIALIZABILITY THEORY 25
2.1 Histories 25
2.2 Serializable Histories 30
2.3 The Serializability Theorem 32
2.4 Recoverable Histories 34
2.5 Operations Beyond Reads and Writes 37
2.6 View Equivalence 38
## Contents

### 3  
**Two Phase Locking**  47

3.1 Aggressive and Conservative Schedulers  47  
3.2 Basic Two Phase Locking  49  
3.3 Correctness of Basic Two Phase Locking*  53  
3.4 Deadlocks  56  
3.5 Variations of Two Phase Locking  58  
3.6 Implementation Issues  60  
3.7 The Phantom Problem  64  
3.8 Locking Additional Operations  67  
3.9 Multigranularity Locking  69  
3.10 Distributed Two Phase Locking  77  
3.11 Distributed Deadlocks  79  
3.12 Locking Performance  87  
3.13 Tree Locking  95  

### 4  
**Non-Locking Schedulers**  113

4.1 Introduction  113  
4.2 Timestamp Ordering (TO)  114  
4.3 Serialization Graph Testing (SG1)  121  
4.4 Certifiers  128  
4.5 Integrated Schedulers  132  

### 5  
**Multiversion Concurrency Control**  143

5.1 Introduction  143  
5.2 Multiversion Serializability Theory*  146  
5.3 Multiversion Timestamp Ordering  153  
5.4 Multiversion Two Phase Locking  156  
5.5 A Multiversion Mixed Method  160
6 CENTRALIZED RECOVERY 167
6.1 Failures 167
6.2 Data Manager Architecture 169
6.3 The Recovery Manager 174
6.4 The Undo/Redo Algorithm 180
6.5 The Undo/No-Redo Algorithm 196
6.6 The No-Undo/Redo Algorithm 198
6.7 The No-Undo/No-Redo Algorithm 201
6.8 Media Failures 206

7 DISTRIBUTED RECOVERY 217
7.1 Introduction 217
7.2 Failures in a Distributed System 218
7.3 Atomic Commitment 222
7.4 The Two Phase Commit Protocol 226
7.5 The Three Phase Commit Protocol 240

8 REPLICATED DATA 265
8.1 Introduction 265
8.2 System Architecture 268
8.3 Serializability Theory for Replicated Data 271
8.4 A Graph Characterization of 1SR Histories 275
8.5 Atomicity of Failures and Recoveries 277
8.6 An Available Copies Algorithm 281
8.7 Directory-oriented Available Copies 289
8.8 Communication Failures 294
8.9 The Quorum Consensus Algorithm 298
8.10 The Virtual Partition Algorithm 304

APPENDIX 313
GLOSSARY 321
BIBLIOGRAPHY 339
INDEX 363