

---

---

# Contents

---

---

<b>FOREWORD</b>	<b>xi</b>
<b>PREFACE</b>	<b>xv</b>
<b>1 THE CAPACITY PLANNING PROBLEM</b>	<b>1</b>
1.1 Introduction	1
1.2 The Bank Problem	2
1.3 Some Capacity Planning Situations	3
1.3.1 <i>Case 1: Increasing the Number of ATM Terminals,</i>	<i>4</i>
1.3.2 <i>Case 2: Increase in Transaction Volume,</i>	<i>4</i>
1.3.3 <i>Case 3: New Applications,</i>	<i>7</i>
1.4 The Capacity Planning Concept	7
1.5 Why is Capacity Planning Important?	10
1.6 Some Common Mistakes in Capacity Planning	11
1.7 The Bank Problem Revisited	13
1.8 Concluding Remarks	17

- 1.9 Exercises 17
- Bibliography 18

## **2 A CAPACITY PLANNING METHODOLOGY**

19

- 2.1 Introduction 19
- 2.2 The Retail Company Problem 20
- 2.3 Methodology Overview 23
- 2.4 Understanding the Current Environment 25
  - 2.4.1 *The Computing Environment Structure, 25*
  - 2.4.2 *What Is Capacity? 26*
  - 2.4.3 *Time Windows, 30*
  - 2.4.4 *Service Levels, 32*
  - 2.4.5 *Overall Description, 35*
- 2.5 The Problem of Workload Characterization 37
  - 2.5.1 *First Approach to the Problem, 38*
  - 2.5.2 *Simple Example, 40*
  - 2.5.3 *Workload Model, 43*
- 2.6 Workload Characterization Methodology 46
  - 2.6.1 *Identification of the Basic Component, 46*
  - 2.6.2 *Choice of the Characterizing Parameters, 46*
  - 2.6.3 *Data Collection, 47*
  - 2.6.4 *Partitioning the Workload, 47*
  - 2.6.5 *Calculating the Class Parameters, 51*
- 2.7 Workload Forecasting 60
  - 2.7.1 *From Business Plans to Computer Demands, 61*
  - 2.7.2 *Forecasting Techniques, 63*
- 2.8 Performance Prediction 68
  - 2.8.1 *Rules of Thumb, 70*
  - 2.8.2 *Trend Analysis, 71*
  - 2.8.3 *Benchmarks, 72*
- 2.9 Concluding Remarks 73
- 2.10 Exercises 73
- Bibliography 74

## **3 PERFORMANCE MODELS**

76

- 3.1 Introduction 76
- 3.2 Building a Performance Model of a Simple Computer System 77

- 3.3 Performance Model of an Interactive System 84
- 3.4 Performance Model of a Batch System 88
- 3.5 Multiple-Class Performance Models 89
  - 3.5.1 *Aggregation of Classes, 90*
  - 3.5.2 *Priorities, 93*
  - 3.5.3 *Shared Domains, 95*
  - 3.5.4 *Summary of Parameter Types for Multiclass Performance Models, 97*
- 3.6 Baseline Model and Modification Analysis 97
  - 3.6.1 *Baseline Model, 98*
  - 3.6.2 *Using a Database Management System, 98*
  - 3.6.3 *Using a DBMS and an Optimizing Compiler, 99*
  - 3.6.4 *Increasing Buffer Pool Size, 100*
  - 3.6.5 *Using the Log Option for Crash Recovery, 101*
  - 3.6.6 *Summary of Modification Analysis Results, 101*
- 3.7 Performance Model of a Client-Server System 102
- 3.8 Multiprocessing 106
- 3.9 Types of Models 108
- 3.10 Concluding Remarks 109
- 3.11 Exercises 110
  - Bibliography 112

**4 INTUITIVE SOLUTIONS FOR A SINGLE-PROCESSOR COMPUTER SYSTEM PERFORMANCE MODEL**

- 4.1 Introduction 113
- 4.2 Alternative View of the Modeling Paradigm 113
- 4.3 Simple Server System I 117
  - 4.3.1 *Model Solution, 118*
  - 4.3.2 *Little's Result, 123*
  - 4.3.3 *Model Predictions, 126*
- 4.4 Simple Server System II 128
- 4.5 Generalized Birth–Death Models 131
- 4.6 Alternative Approach: Operational Analysis 133
- 4.7 Concluding Remarks 136
- 4.8 Exercises 136
  - Bibliography 138

<b>5</b>	<b>EFFICIENT SOLUTIONS FOR COMPUTER SYSTEM MODELS</b>	<b>140</b>
5.1	Introduction	140
5.2	First-Principles Approach Review	141
5.3	Basic Central Server System	143
	5.3.1 <i>State Description Specification</i> ,	145
	5.3.2 <i>State Enumeration</i> ,	145
	5.3.3 <i>Transition Rate Specification</i> ,	147
	5.3.4 <i>Balance Equation Formulation</i> ,	147
	5.3.5 <i>Balance Equation Solution</i> ,	149
	5.3.6 <i>Performance Metric Derivation</i> ,	153
5.4	Simple Convolution	155
5.5	Simple Mean Value Analysis	159
5.6	Simple Decomposition/Aggregation	163
5.7	Simple Bounding Techniques	167
5.8	Concluding Remarks	171
5.9	Exercises	173
	Bibliography	174
<b>6</b>	<b>MULTIPLE-CLASS MODELS</b>	<b>176</b>
6.1	Introduction	176
6.2	Simple Two-Class Model	177
6.3	Notation and Assumptions	180
6.4	Closed Models	182
	6.4.1 <i>Exact Solution Algorithm</i> ,	185
	6.4.2 <i>Approximate Solution Algorithms</i> ,	189
6.5	Open Models	192
	6.5.1 <i>Analysis of Multiclass Open Models</i> ,	193
	6.5.2 <i>Open Model Example</i> ,	195
6.6	Mixed Models	199
6.7	Concluding Remarks	202
6.8	Exercises	202
	Bibliography	203
<b>7</b>	<b>PERFORMANCE OF CLIENT-SERVER ARCHITECTURES</b>	<b>205</b>
7.1	Introduction	205
7.2	The Telemarketing Company Problem	206

- 7.3 Performance Considerations in the CS Paradigm 207
- 7.4 Workload Characterization for CS Architectures 209
- 7.5 Solving the Telemarketing Company Problem 210
  - 7.5.1 *Birth–Death Model*, 213
  - 7.5.2 *CS Model*, 215
- 7.6 Categories of CS Performance Models 219
  - 7.6.1 *Single-Request Clients and Homogeneous Workloads*, 220
  - 7.6.2 *Single-Request Clients and Heterogeneous Workloads*, 221
  - 7.6.3 *Multiple-Request Clients and Homogeneous Workloads*, 221
  - 7.6.4 *Multiple-Request Clients and Heterogeneous Workloads*, 221
- 7.7 Multiclass MVA with Load-Dependent Devices 222
- 7.8 Multiclass Open Queueing Networks with Load-Dependent Servers 226
- 7.9 Concluding Remarks 230
- 7.10 Exercises 230
  - Bibliography 232

## 8 MODELS OF PRACTICAL COMPUTER SYSTEMS

234

- 8.1 Introduction 234
- 8.2 Modeling Memory Queueing 235
  - 8.2.1 *Flow-Equivalent Method*, 236
  - 8.2.2 *Single Class*, 237
  - 8.2.3 *Multiple Classes*, 240
- 8.3 Modeling Disk Subsystems 243
  - 8.3.1 *I/O Subsystem Architecture*, 244
  - 8.3.2 *I/O Service Time*, 246
  - 8.3.3 *Estimating Delays within the I/O Subsystem*, 247
  - 8.3.4 *Modeling I/O Contention in Closed Systems*, 252
- 8.4 Modeling CPU Scheduling Priorities 254
  - 8.4.1 *Two-Priority Example*, 254
  - 8.4.2 *SWIC Priority Algorithm*, 257
- 8.5 Modeling Paging Activity 259

- 8.6 Modeling Multiprocessing 262
  - 8.6.1 *Tightly Coupled Multiprocessors*, 263
  - 8.6.2 *Loosely Coupled Multiprocessors*, 264
- 8.7 Concluding Remarks 266
- 8.8 Exercises 266
  - Bibliography 268

## 9 OBTAINING INPUT PARAMETERS

270

- 9.1 Introduction 270
- 9.2 Measurement Techniques 272
  - 9.2.1 *Monitor Tools*, 274
  - 9.2.2 *Monitoring Modes*, 277
- 9.3 Input Parameter Description 278
  - 9.3.1 *Workload Intensity*, 279
  - 9.3.2 *Service Demands*, 280
  - 9.3.3 *Overhead Representation*, 287
- 9.4 Parameter Estimation 287
  - 9.4.1 *Steps in Obtaining Input Parameters*, 288
  - 9.4.2 *Arrival Rate*, 288
  - 9.4.3 *Multiprogramming Level*, 289
  - 9.4.4 *Number of Active Terminals and Think Time*, 290
  - 9.4.5 *CPU Service Demand*, 291
  - 9.4.6 *I/O Service Demands*, 294
- 9.5 Order Processing System Example 297
  - 9.5.1 *Average Arrival Rate*, 298
  - 9.5.2 *CPU Demand*, 298
  - 9.5.3 *Disk Demands*, 299
- 9.6 Concluding Remarks 301
- 9.7 Exercises 302
  - Bibliography 302

## 10 MODEL CALIBRATION AND VALIDATION

304

- 10.1 Introduction 304
  - 10.1.1 *Review of the Basic Modeling Paradigm*, 305
- 10.2 Calibration Example 305
- 10.3 Types of Calibrations 309

- 10.4 Effect of Calibrations 311
  - 10.4.1 *Percent and Absolute Difference*, 312
  - 10.4.2 *Calibration on the Mean Multiprogramming Level*, 313
  - 10.4.3 *Multiprogramming Level Distribution*, 314
  - 10.4.4 *Multiclass Calibration*, 315
  - 10.4.5 *Absolute Demand Calibration*, 315
  - 10.4.6 *Calibration on the Bottleneck Device*, 316
  - 10.4.7 *Ghost Server Calibration*, 316
- 10.5 Case Study of a Theoretical System 316
  - 10.5.1 *Calibrated Models*, 318
  - 10.5.2 *Prediction Tasks*, 321
- 10.6 Rules of Thumb 333
- 10.7 A More Formal View 334
- 10.8 Concluding Remarks 336
- 10.9 Exercises 336
- Bibliography 339

**11 PERFORMANCE OF NEW APPLICATIONS 340**

- 11.1 Introduction 340
- 11.2 The Credit Card Problem 342
- 11.3 Sizing the Credit Card Application 344
- 11.4 Techniques for SPE 352
  - 11.4.1 *Software Development Life Cycle and SPE*, 352
  - 11.4.2 *Obtaining Data for SPE*, 354
- 11.5 Environment for SPE 356
- 11.6 Credit Card Problem Revisited 360
- 11.7 Concluding Remarks 365
- 11.8 Exercises 366
- Bibliography 368

**12 CONCLUDING REMARKS 369**

**APPENDIX A. GLOSSARY OF TERMS 373**

**APPENDIX B. GLOSSARY OF NOTATION 382**

**APPENDIX C. FORMULAS 386**

**APPENDIX D. QUICK MANUAL FOR Q*Solver*/1 393**

**INDEX 398**