

---

# Contents

<b>Preface</b> .....	VII
<b>1 What Is Guerrilla Capacity Planning?</b> .....	1
1.1 Introduction .....	1
1.2 Why Management Resists Capacity Planning .....	1
1.2.1 Risk Management vs. Risk Perception.....	2
1.2.2 Instrumentation Just Causes Bugs.....	3
1.2.3 As Long as It Fails on Time .....	4
1.2.4 Capacity Management as a Homunculus.....	5
1.3 Guerrilla vs. Gorilla .....	6
1.3.1 No Compass Required .....	7
1.3.2 Modeling Is Not Like a Model Railway .....	8
1.3.3 More Like a Map Than the Metro .....	8
1.4 Tactical Planning as a Weapon .....	9
1.4.1 Scalability by Spreadsheet .....	10
1.4.2 A Lot From Little .....	11
1.4.3 Forecasting on the Fly .....	13
1.4.4 Guerrilla Guidelines .....	14
1.5 Summary.....	16
<b>2 ITIL for Guerrillas</b> .....	17
2.1 Introduction .....	17
2.2 ITIL Background .....	17
2.2.1 Business Perspective .....	19
2.2.2 Capacity Management .....	21
2.3 The Wheel of Capacity Management.....	21
2.3.1 Traditional Capacity Planning .....	21
2.3.2 Running on the Rim .....	23
2.3.3 Guerrilla Racing Wheel .....	24
2.4 Summary.....	25

<b>3</b>	<b>Damaging Digits in Capacity Calculations</b>	27
3.1	Introduction	27
3.2	Significant Digits	28
3.2.1	Accuracy	28
3.2.2	Precision	29
3.3	Sifting for SigDigs	30
3.3.1	Count by Zeros	30
3.3.2	Significance and Scale	32
3.4	Rounding Rules	32
3.4.1	Golden Rule	34
3.4.2	Sum Rule	34
3.4.3	Product Rule	34
3.5	Planning With Dollars and Sense	35
3.5.1	Cost Metric	35
3.5.2	Significant Digits	36
3.6	Expressing Errors	37
3.6.1	Absolute Error	37
3.6.2	Relative Error	37
3.6.3	Standard Deviation	37
3.6.4	Standard Error	38
3.6.5	Error Bars	38
3.6.6	Instrumentation Error	39
3.7	Interval Arithmetic	39
3.8	Summary	40
<b>4</b>	<b>Scalability—A Quantitative Approach</b>	41
4.1	Introduction	41
4.2	Fundamental Concepts of Scaling	41
4.2.1	Geometric Scaling	42
4.2.2	Allometric Scaling	43
4.2.3	Critical Size	44
4.2.4	Sizing Examples	45
4.3	Hardware Scalability	47
4.3.1	Ideal Parallelism	48
4.3.2	Amdahl’s Law	49
4.3.3	Multiuser Scaleup	52
4.3.4	Serial-Parallel Duality	55
4.3.5	Scaled Speedup	56
4.4	Universal Scalability Model	56
4.4.1	The Role of Coherency	58
4.5	Other Scalability Models	63
4.5.1	Geometric Model	63
4.5.2	Quadratic Model	63
4.5.3	Exponential Model	64
4.6	Multicores and Clusters	66

4.7	Summary	68
<b>5</b>	<b>Evaluating Scalability Parameters</b>	<b>71</b>
5.1	Introduction	71
5.2	Benchmark Measurements	72
5.2.1	The Workload	72
5.2.2	The Platform	74
5.2.3	The Procedure	75
5.3	Minimal Dataset	75
5.3.1	Interpolating Polynomial	76
5.3.2	Regression Polynomial	76
5.4	Capacity Ratios	77
5.5	Transforming the Scalability Equation	77
5.5.1	Efficiency	78
5.5.2	Deviation From Linearity	78
5.5.3	Transformation of Variables	79
5.5.4	Properties of the Regression Curve	80
5.6	Regression Analysis	82
5.6.1	Quadratic Polynomial	82
5.6.2	Parameter Mapping	83
5.6.3	Interpreting the Scalability Parameters	85
5.6.4	Error Reporting	86
5.7	Less Than a Full Deck	87
5.7.1	Sparse Even Data	88
5.7.2	Sparse Uneven Data	90
5.7.3	Missing $X(1)$ Datum	91
5.8	Summary	94
<b>6</b>	<b>Software Scalability</b>	<b>97</b>
6.1	Introduction	97
6.2	Amdahl's Law for Software	98
6.3	Universal Software Scalability	100
6.4	Concurrent Programming and Coherency	102
6.5	UNIX Multitasking Application	103
6.5.1	The Workload	103
6.5.2	The Platform	104
6.5.3	Regression Analysis	104
6.6	Windows-Based Applications	107
6.6.1	The Workload	107
6.6.2	The Platform	108
6.6.3	Regression Analysis	109
6.7	Multitier Architectures	110
6.7.1	The Workload	111
6.7.2	The Platform	111
6.7.3	Regression Analysis	112

- 6.7.4 Why It Works ..... 114
- 6.8 Classification by Workload..... 115
- 6.9 Summary ..... 116
- 7 Fundamentals of Virtualization ..... 117**
  - 7.1 Introduction ..... 117
  - 7.2 The Spectrum of Virtual Machines ..... 118
    - 7.2.1 VM Spectroscopy ..... 118
    - 7.2.2 Polling Rates and Frequency Scales ..... 119
  - 7.3 Microlevel Virtual Machines: Hyperthreading ..... 119
    - 7.3.1 Micro-VM Polling ..... 122
    - 7.3.2 Thread Execution Analysis ..... 123
    - 7.3.3 Missing MIPS Explained ..... 124
    - 7.3.4 Windows 2000 Production Server ..... 126
    - 7.3.5 Guerrilla Capacity Planning ..... 127
  - 7.4 Mesolevel Virtual Machines: Hypervisors ..... 127
    - 7.4.1 Fair-Share Scheduling ..... 129
    - 7.4.2 Meso-VM Polling ..... 132
    - 7.4.3 VMWare Share Allocation Analysis ..... 134
    - 7.4.4 J2EE WebLogic Production Application..... 135
    - 7.4.5 VMWare Scalability Analysis ..... 137
    - 7.4.6 Guerrilla Capacity Planning ..... 138
  - 7.5 Macrolevel Virtual Machines: Hypernets..... 138
    - 7.5.1 Macro-VM Polling ..... 139
    - 7.5.2 Bandwidth Scalability Analysis ..... 140
    - 7.5.3 Remote Polling Rates ..... 141
    - 7.5.4 Guerrilla Capacity Planning ..... 142
  - 7.6 Summary ..... 142
- 8 Web Site Planning ..... 143**
  - 8.1 Introduction ..... 143
  - 8.2 Analysis of Daily Traffic ..... 144
    - 8.2.1 The Camel and the Dromedary ..... 144
    - 8.2.2 Unimodal but Bicoastal ..... 146
  - 8.3 Effective Demand ..... 148
    - 8.3.1 Modeling Assumptions ..... 149
    - 8.3.2 Statistical Approach ..... 149
  - 8.4 Selecting Statistical Tools ..... 150
    - 8.4.1 Spreadsheet Programming ..... 150
    - 8.4.2 Online Support ..... 150
  - 8.5 Planning for Data Collection ..... 151
    - 8.5.1 Commercial Collectors: Use It or Lose It ..... 151
    - 8.5.2 Brewing in the Background ..... 151
  - 8.6 Short-Term Capacity Planning ..... 152
    - 8.6.1 Multivariate Regression of Daily Data..... 152

8.6.2	Automation Using Spreadsheet Macros . . . . .	153
8.7	Long-Term Capacity Planning . . . . .	155
8.7.1	Nonlinear Regression of Weekly Data . . . . .	155
8.7.2	Procurement Curves . . . . .	156
8.7.3	Estimating Server Scalability . . . . .	157
8.7.4	Calculating Capacity Gains . . . . .	158
8.7.5	Estimating the Doubling Period . . . . .	161
8.8	Summary . . . . .	162
<b>9</b>	<b>Gargantuan Computing—GRIDs and P2P</b> . . . . .	<b>165</b>
9.1	Introduction . . . . .	165
9.2	GRIDs vs. P2P . . . . .	166
9.3	Analysis of Gnutella . . . . .	167
9.4	Tree Topologies . . . . .	168
9.4.1	Binary Tree . . . . .	169
9.4.2	Rooted Tree . . . . .	169
9.4.3	Cayley Tree . . . . .	169
9.5	Hypernet Topologies . . . . .	169
9.5.1	Hypercube . . . . .	170
9.5.2	Hypertorus . . . . .	170
9.6	Capacity Metrics . . . . .	170
9.6.1	Network Diameter . . . . .	170
9.6.2	Total Nodes . . . . .	171
9.6.3	Path Length . . . . .	171
9.6.4	Internal Path Length . . . . .	171
9.6.5	Average Hop Distance . . . . .	171
9.6.6	Network Links . . . . .	172
9.6.7	Network Demand . . . . .	172
9.6.8	Peer Demand . . . . .	172
9.6.9	Bandwidth . . . . .	173
9.7	Relative Bandwidth . . . . .	173
9.7.1	Cayley Trees . . . . .	173
9.7.2	Trees and Cubes . . . . .	174
9.7.3	Cubes and Tori . . . . .	175
9.7.4	Ranked Performance . . . . .	176
9.8	Summary . . . . .	176
<b>10</b>	<b>Internet Planning</b> . . . . .	<b>179</b>
10.1	Introduction . . . . .	179
10.2	Bellcore Traces . . . . .	180
10.3	Fractals and Self-Similarity . . . . .	182
10.4	Fractals in Time . . . . .	186
10.4.1	Short-Range Dependence . . . . .	186
10.4.2	Long-Range Dependence . . . . .	188
10.5	Impact on Buffer Sizing . . . . .	190

- 10.5.1 Conventional Buffer Sizing..... 190
- 10.5.2 LRD Buffer Sizing..... 192
- 10.6 New Developments ..... 193
  - 10.6.1 Ethernet Packetization ..... 194
  - 10.6.2 LRD and Flicker Noise ..... 196
- 10.7 Summary..... 197
  
- 11 Going Guerrilla—A Case Study ..... 199**
  - 11.1 Introduction ..... 199
  - 11.2 Guerrilla Monitoring Phase ..... 199
  - 11.3 The Basic Solution ..... 201
    - 11.3.1 Implementation Details ..... 202
    - 11.3.2 Orca Output Examples ..... 203
    - 11.3.3 Round-Robin Database ..... 203
  - 11.4 Extending the Basic Solution ..... 206
    - 11.4.1 Mainframe Data Processing..... 206
    - 11.4.2 Guerrilla Planning Phase ..... 207
    - 11.4.3 Monitoring With ORCAAlerts..... 208
  - 11.5 Future Developments ..... 209
  - 11.6 Summary..... 210

**Appendix**

- A Amdahl and the Repairman ..... 213**
  - A.1 Repairman Queueing Model ..... 213
  - A.2 Amdahl’s Law for Parallel Subtasks..... 214
    - A.2.1 Single Task ..... 215
    - A.2.2 Two Subtasks..... 215
    - A.2.3 Multiple Subtasks ..... 215
  - A.3 Amdahl’s Law for Concurrent Multitasks ..... 217
  - A.4 Note On Nelson’s Approach..... 217
  
- B Mathematica Evaluation of NUMA Parameters ..... 219**
  - B.1 Mathematica Packages ..... 219
  - B.2 Import the Data ..... 219
  - B.3 Tabulate the Data ..... 220
  - B.4 Plot Normalized Data ..... 220
  - B.5 Nonlinear Regression..... 221
  - B.6 ANOVA Report ..... 221
  - B.7 Maximal CPU Configuration ..... 222
  - B.8 Plot of Regression Model ..... 222

**C Abbreviations and Units** . . . . . 223

  C.1 SI Prefixes . . . . . 223

  C.2 Time Suffixes . . . . . 223

  C.3 Capacity Suffixes . . . . . 224

**D Programs for Chapter 3** . . . . . 225

  D.1 Determine SigDigs in VBA . . . . . 225

  D.2 Determine SigDigs in Mathematica . . . . . 226

  D.3 Determine SigDigs in Perl . . . . . 227

**E Programs for Chapter 8** . . . . . 229

  E.1 Example Data Extractor in Perl . . . . . 229

  E.2 VBA Macro for Calculating  $U_{\text{eff}}$  . . . . . 231

**F The Guerrilla Manual** . . . . . 235

  F.1 Weapons of Mass Instruction . . . . . 235

  F.2 Capacity Modeling Rules of Thumb . . . . . 238

  F.3 Scalability on a Stick . . . . . 240

    F.3.1 Universal Law of Computational Scaling . . . . . 240

    F.3.2 Areas of Applicability . . . . . 241

    F.3.3 How to Use It . . . . . 241

**Bibliography** . . . . . 243

**Index** . . . . . 249



<http://www.springer.com/978-3-540-26138-4>

Guerrilla Capacity Planning

A Tactical Approach to Planning for Highly Scalable  
Applications and Services

Gunther, N.J.

2007, XX, 253 p. 108 illus., Hardcover

ISBN: 978-3-540-26138-4