

Contents

<i>Preface</i>	<i>page</i>	<i>xiii</i>
1 Introduction		1
1.1 The Importance of Workloads		2
1.2 Types of Workloads		5
1.2.1 Workloads in Different Domains		6
1.2.2 Dynamic vs. Static Workloads		7
1.2.3 Benchmarks		8
1.3 Workload Modeling		10
1.3.1 What It Is		10
1.3.2 Why Do It?		11
1.3.3 How It Is Done		16
1.4 Roadmap		21
2 Workload Data		22
2.1 Data Sources		22
2.1.1 Using Available Logs		22
2.1.2 Active Data Collection		29
2.2 Data Usability		34
2.2.1 Representativeness		34
2.2.2 Stationarity		42
2.3 Data Filtering and Cleaning		46
2.3.1 Noise and Errors		47
2.3.2 Multiclass Workloads		50
2.3.3 Anomalous Behavior and Robots		52
2.3.4 Workload Flurries and Flash Crowds		56
2.3.5 Identifying Noise and Anomalies		60
2.4 Educated Guessing		63
2.5 Sharing Data		64
2.5.1 Data Formats		66

2.5.2	Data Volume	70
2.5.3	Privacy	71
3	Statistical Distributions	73
3.1	Describing a Distribution	75
3.1.1	Histograms, pdfs, and CDFs	76
3.1.2	Central Tendency	86
3.1.3	Dispersion	90
3.1.4	Moments and Order Statistics	95
3.1.5	Focus on Skew	98
3.2	Some Specific Distributions	101
3.2.1	The Exponential Distribution	102
3.2.2	Phase-Type Distributions	106
3.2.3	The Hyper-Exponential Distribution	107
3.2.4	The Erlang Distribution	109
3.2.5	The Hyper-Erlang Distribution	111
3.2.6	Other Phase-Type Distributions	112
3.2.7	The Normal Distribution	114
3.2.8	The Lognormal Distribution	115
3.2.9	The Gamma Distribution	118
3.2.10	The Weibull Distribution	119
3.2.11	The Pareto Distribution	120
3.2.12	The Zipf Distribution	123
3.2.13	Do It Yourself	128
4	Fitting Distributions to Data	130
4.1	Approaches to Fitting Distributions	131
4.2	Parameter Estimation for a Single Distribution	131
4.2.1	Justification	132
4.2.2	The Method of Moments	133
4.2.3	The Maximum Likelihood Method	134
4.2.4	Estimation for Specific Distributions	136
4.2.5	Sensitivity to Outliers	138
4.2.6	Variations in Shape	140
4.3	Parameter Estimation for a Mixture of Distributions	141
4.3.1	Examples of Mixtures	141
4.3.2	The Expectation-Maximization Algorithm	142
4.4	Re-Creating the Shape of a Distribution	145
4.4.1	Using an Empirical Distribution	145
4.4.2	Modal Distributions	148
4.4.3	Constructing a Hyper-Exponential Tail	151
4.5	Tests for Goodness of Fit	155
4.5.1	Using Q-Q Plots	155
4.5.2	The Kolmogorov-Smirnov Test	159
4.5.3	The Anderson-Darling Test	160
4.5.4	The χ^2 Method	161
4.6	Software Packages for Distribution Fitting	161

5 Heavy Tails	163
5.1 The Definition of Heavy Tails	164
5.1.1 Power-Law Tails	164
5.1.2 Properties of Power Laws	165
5.1.3 Alternative Definitions	172
5.2 The Importance of Heavy Tails	175
5.2.1 Conditional Expectation	175
5.2.2 Mass-Count Disparity	178
5.3 Testing for Heavy Tails	185
5.4 Modeling Heavy Tails	193
5.4.1 Estimating the Parameters of a Power-Law Tail	193
5.4.2 Generalization and Extrapolation	200
5.4.3 Generative Models	206
6 Correlations in Workloads	213
6.1 Types of Correlation	213
6.2 Spatial and Temporal Locality	215
6.2.1 Definitions	215
6.2.2 Statistical Measures of Locality	217
6.2.3 The Stack Distance and Temporal Locality	218
6.2.4 Working Sets and Spatial Locality	222
6.2.5 Measuring Skewed Distributions and Popularity	223
6.2.6 Modeling Locality	223
6.2.7 System Effects on Locality	231
6.3 Locality of Sampling	231
6.3.1 Examples and Visualization	232
6.3.2 Quantification	235
6.3.3 Properties	240
6.3.4 Importance	240
6.3.5 Modeling	242
6.4 Cross-Correlation	245
6.4.1 Joint Distributions and Scatterplots	245
6.4.2 The Correlation Coefficient and Linear Regression	249
6.4.3 Distributional Correlation	258
6.4.4 Modeling Correlations by Clustering	262
6.4.5 Modeling Correlations with Distributions	269
6.4.6 Dynamic Workloads vs. Snapshots	270
6.5 Correlation with Time	272
6.5.1 Periodicity and the Diurnal Cycle	273
6.5.2 Trends	280
7 Self-Similarity and Long-Range Dependence	283
7.1 Poisson Arrivals	284
7.1.1 The Poisson Process	284

7.1.2	Nonhomogeneous Poisson Process	285
7.1.3	Batch Arrivals	285
7.2	The Phenomenon of Self-Similarity	287
7.2.1	Examples of Self-Similarity	287
7.2.2	Self-Similarity and Long-Range Dependence	289
7.2.3	The Importance of Self-Similarity	291
7.2.4	Focus on Scaling	292
7.3	Mathematical Definitions	294
7.3.1	Data Manipulations	295
7.3.2	Exact Self-Similarity	297
7.3.3	Focus on the Covariance	299
7.3.4	Long-Range Dependence	300
7.3.5	Asymptotic Second-Order Self-Similarity	303
7.3.6	The Hurst Parameter and Random Walks	306
7.4	Measuring Self-Similarity	309
7.4.1	Testing for a Poisson Process	309
7.4.2	The Rescaled Range Method	310
7.4.3	The Variance Time Method	316
7.4.4	Measuring Long-Range Dependence Directly	318
7.4.5	Using Wavelets and Logscale Diagrams	319
7.4.6	Spectral Methods: The Periodogram and Whittle Estimator	325
7.4.7	Comparison of Results	337
7.4.8	Validation	338
7.4.9	Software for Analyzing Self-Similarity	339
7.5	Modeling Self-Similarity	340
7.5.1	Classical Long-Range Dependent Models	340
7.5.2	Multiscale Wavelet-Based Construction	345
7.5.3	Bias Models	347
7.5.4	The M/G/ ∞ Queueing Model	350
7.5.5	Merged On-Off Processes	353
7.6	More Complex Scaling Behavior	356
8	Hierarchical Generative Models	357
8.1	Locality of Sampling and Users	358
8.2	Hierarchical Workload Models	359
8.2.1	Hidden Markov Models	359
8.2.2	Motivation for User-Based Models	362
8.2.3	The Three-Level User-Based Model	366
8.2.4	Other Hierarchical Models	367
8.3	User-Based Modeling	370
8.3.1	Modeling the User Population	370
8.3.2	Modeling User Sessions	374
8.3.3	Modeling User Activity within Sessions	383
8.3.4	User Resampling	389
8.4	Performance Feedback	391

9 Case Studies	399
9.1 Human User Behavior	399
9.1.1 Sessions and Job Arrivals	400
9.1.2 Interactivity and Think Times	401
9.1.3 Daily Activity Cycle	403
9.1.4 Patience	405
9.1.5 Mobility	406
9.1.6 Runtime Estimates	407
9.2 Desktop and Workstation Workloads	410
9.2.1 Process Runtimes	410
9.2.2 Application Behavior	412
9.2.3 Multimedia Applications and Games	420
9.2.4 Benchmark Suites vs. Workload Models	421
9.2.5 Predictability	423
9.2.6 Operating Systems	424
9.2.7 Virtualization Workloads	425
9.3 File System and Storage Workloads	426
9.3.1 The Distribution of File Sizes	426
9.3.2 File System Access Patterns	429
9.3.3 Feedback	431
9.3.4 I/O Operations and Disk Layout	432
9.3.5 Parallel File Systems	434
9.4 Network Traffic and the Web	434
9.4.1 Internet Traffic	434
9.4.2 Email	441
9.4.3 Web Server Load	441
9.4.4 User Sessions	448
9.4.5 E-Commerce	449
9.4.6 Search Engines	451
9.4.7 Media and Streaming	457
9.4.8 Peer-to-Peer File Sharing	459
9.4.9 Online Gaming	460
9.4.10 Web Applications and Web 2.0	461
9.4.11 User Types	462
9.4.12 Feedback	463
9.4.13 Malicious Traffic	464
9.5 Data-Centric Workloads	465
9.5.1 Database Systems	465
9.5.2 Information Retrieval	468
9.5.3 Big Data	469
9.6 Parallel Jobs	473
9.6.1 Arrivals	474
9.6.2 Rigid Jobs	475
9.6.3 Speedup	480
9.6.4 Parallel Program Behavior	482

9.6.5 Load Manipulation and System Size	485
9.6.6 Grid Workloads	488
10 Summary and Outlook	490
Appendix: Data Sources	495
<i>Bibliography</i>	501
<i>Index</i>	541