Energy Efficient Servers

Blueprints for Data Center Optimization

Corey Gough
Ian Steiner
Winston Saunders

Apress open

Contents

About the Authors xv

About the Technical Reviewers xvii

Contributing Authors xix

Acknowledgments xxi

Chapter 1: Why Data Center Efficiency Matters 1

An Industry's Call to Action 2

Data Center Infrastructure Energy Use 3

Energy Proportional Server Efficiency 5

Regulatory Environment 6

Measuring Energy Efficiency 8

SPECPower 8

High Performance Computing Efficiency 14

Energy Efficiency and Cost 16

Summary 20

Chapter 2: CPU Power Management 21

Server CPU Architecture/Design 21

CPU Architecture Building Blocks 22

CPU Design Building Blocks 31

Intel Server Processors 32

Introduction to Power 33
CPU Power Breakdown 34
Frequency, Voltage, and Temperature Interactions 36
Power-Saving Techniques 38
Turn It Off 38
Turn It Down 39
Power-Saving Strategies 39
CPU Power and Performance States 41
C-States 43
P-States 49
T-States 56
S-States and G-States 57
S0ix 57
Running Average Power Limit (RAPL) 58
CPU Thermal Management 66
CPU Power Management Summary 69
Summary 70
Chapter 3: Memory and I/O Power Management 71
System Memory 71
Memory Architecture Basics 71
Devices and Ranks 72
Memory Error Correction (ECC) 74
Memory Capacity 74
Device Power Characteristics 75
DDR3 vs. DDR4 76
RDIMMs, UDIMMs, SODIMMs, and LRDIMMs 77
Memory Channel Interleave and Imbalanced Memory Configurations 78

Power and Performance States 79
CKE Power Savings 79
Self-Refresh 81
Voltage/Frequency 82
DDR Thermal Management 83
Monitoring Temperature 83
Memory Throttling 83
CPU DDRIO 84
Workload Behavior 85
Memory Reliability Features 85
CPU I/Os 85
CPU Interconnect (QPI) 85
PCIe 87
Summary 90
Chapter 4: Platform Power Management 93
Platform Overview 93
Common Platform Components 93
Integration 95
Platform Manageability 97
CPU Sockets 97
Node Controllers 98
Memory Risers and Memory Buffer Chips 99
Server Chipsets 100
PCH and Platform Power Management 102
PCH Power Management 104
PCIe in Chipsets 105
PCH Thermal Management 105

Networking 105
Ambient Temperature, TDP, and Thermal Management 106
Attached Media 108
LAN Power Management Features 109
USB 111
Link Power States 111
Link Frequency/Voltage 112
Storage 112
Storage Servers and Power Management 113
HDDs and SDDs 114
SATA and SAS Drive Power Management 114
Frequency/Voltage 116
NVMe Drive Power Management 117
Power Delivery 118
Overview of Power Delivery 118
Power Converter Basics 122
Power Conversion Losses 128
Thermal Management 137
System Considerations 141
Component Thermal Management Features 143
Platform Thermal Management 145
Fan Speed Control and Design 149
Summary 151
Chapter 5: BIOS and Management Firmware 153
BIOS Firmware 154
Microcode Update 155
Advanced Configuration and Power Interface 156
Setup Utility 162

Management Firmware 163
Node Manager Capabilities 163
IPMI 168
ACPI Power Metering Objects 171
Summary 171
Chapter 6: Operating Systems 173
Operating Systems 174
C-state Control 174
C-state Policy 176
P-state Control 178
P-state Policy 183
T-state Control 187
Global Power Policy 187
Process Scheduling 188
Memory Management 192
Device Drivers 193
Virtualization 195
Power State Control 195
Consolidation 198
VM Migration 199
Comparison of Operating Environments 201
Microsoft Windows Server (including Hyper-V) 201
Linux Distributions (including KVM) 203
VMWare ESX 206
Summary 207

Chapter 7: Monitoring 209

Hardware Monitoring 209

Fixed Counters 210

Core Performance Monitors 210

Uncore Performance Monitors 211

Status Snapshots 213

Counter Access and Counter Constraints 214

Events and Metrics 214

Management Controller Monitoring 230

Component Power Sensors 230

Synthetic Sensors 231

Sensors and Events 231

Software Monitoring 235

Utilization and Processor Time 236

Processor Power State Requests 240

Scheduler, Processes, and Threads 243

Interrupts 244

Memory 245

I/O 247

Tools 249

Health Checks 249

Hardware Monitoring Tools 254

Operating System Monitoring Tools 265

Summary 268

Chapter 8: Characterization and Optimization 269

Workloads 271

Identifying Suitable Workloads 272

Workload Types 273

System Characterization 277

Steady State vs. Non-Steady State 277

Data Collection 278

Methodology 280

Analysis 281

Optimization 285

CPU Power Management 285

Memory 298

NIC 300

Storage 301

Thermal Management 302

Optimization at a Glance 303

Summary 306

Chapter 9: Data Center Management 307

Data Center Management and Power Distribution 307

Data Center Facilities 307

Power Infrastructure 308

Cooling Infrastructure 312

Simplified Total Cost Models of Cost and Compute Infrastructure 314

Performance per Watt per Dollar 316

Summary 317

Appendix A: Technology and Terms 319

Index 327