DISTRIBUTED DATA MANAGEMENT FOR GRID COMPUTING

MICHAEL DI STEFANO

A JOHN WILEY & SONS, INC. PUBLICATION

CONTENTS

FOREWORD xv

PREFACE xvii

ACKNOWLEDGMENTS xxi

PART I AN OVERVIEW OF GRID COMPUTING

1 What is Grid Computing? 3

The Basics of Grid Computing, 3
Leveling the Playing Field of Buzzword Mania, 4
Paradigm Shift, 7
Beyond the Client/Server, 7
New Topology, 10

2 Why are Businesses Looking at Grid Computing? 13

History Repeats Itself, 13
Early Needs, 14
Artists and Engineers, 14
The Whys and Wherefores of Grid Computing, 17
Financial Factors, 17
Business Drivers, 19
Technology's Role, 19

3 Service-Oriented Architecture **21**

What is Service-Oriented Architecture (SOA)?, 21
Driving Forces Behind SOA, 23
Maturing Technology, 24
Networking, 24
Distributed Computing (Grid), 25
Resource Provisioning, 25
Web Services, 25
Business, 25
World Events, 26
Enter Basic Supply–Demand Economics, 27
Fundamental Shift in Computing, 29

4 Parallel Grid Planes **31**

Using Art to Describe Life: Grid is the Borg, 31
Grid Planes, 32
Compute Grids, 33
Data Grids, 34
Compute and Data Grids—Parallel Planes, 35
True Grid Must Include Data Management, 36
Basic Data Management Requirements, 36
Coordinating the Compute and Data Grid Planes, 36
Data Surfaces in a Data Grid Plane, 37
Evolving the Data Grid, 38

PART II DATA MANAGEMENT IN GRID COMPUTING

5 Scaling in the Grid Topology **43**

Evolution in Data Management, 43
Client/Server Evolution, 44
Grid Evolution, 44
Different Implementations of a Data Grid, 45
Level 0 Data Grids, 45
FTP in Grid, 46
Distributed Filing Systems, 47
Faster Servers, 47
Metadata Hubs and Distributed Data Integration, 48
Level 1 Data Grids, 48
Foundations, 49
Case Study: Integrasoft Grid Fabric (IGF), 51
Application Characteristics for Grid, 53

6 Traditional Data Management 59

Data Management, 59
History, 59
Features, 60
Mechanics, 60
Data Structure, 61
Access, 62
Integrity, 63
Transaction, 63
Events, 64
Backup/Recovery/Availability, 64
Security, 64
Key for Usability, 65

7 Relational Data Management as a Baseline for Understanding the Data Grid 67

Evolution of the Relational Model, 67
Parallels to Data Management in Grid Environments, 68
Analysis of the Functional Tiers, 69
Language Interface, 69
Data Management Engines, 69
Resource Management Engines, 69
Engines Determine the Type of Data Grid, 70
Data Management Features, 70

8 Foundation for Comparing Data Grids 73

Core Engine Determines Performance and Flexibility, 73
Replicated versus Distributed, 74
Centralized versus Peer-to-Peer Synchronization, 75
Access to the Data Grid, 75
User-Level APIs, 75
Spring-Based Interfaces, 76
Support for Traditional Data Management Features, 76
Support for Data Management Features Specific to Grid Computing, 76

9 Data Regionalization 79

What are Data Regions?, 80
Data Regions in Traditional Terms, 80
Data Management in a Data Grid, 84
Data Distribution Policy, 85
Data Distribution Policy Expression, 87

Data Replication Policy, 88
Data Replication Policy Expression, 89
Synchronization Policy, 90
Load-and-Store Policy, 90
Data Load Policy Expression, 93
Data Store Policy Expression, 94
Event Notification Policy, 95
Event Notification Policy Expression, 96
Quality-of-Service (QoS) Levels, 96

10 Data Synchronization **99**

Intraregion Synchronization, 100
Interregion Synchronization, 101
Synchronization Architectures, 102
Centralized Synchronization Manager, 102
Peer-to-Peer Synchronization, 103
Synchronization Patterns, 104
Synchronization Granularity, 105
Synchronization Policy Expression, 106
Synchronization Pattern Simulations, 108
Synchronization Policy as a Standard Interface, 109

11 Data Integration **111**

Enterprise Application/Information Integration (EAI/EII) in Grid, 111
Straight-Through Processing (STP), EAI, and EII, 111
EII in Grid, 116
Natural Separation of Process and Data, 118
Data Load Policy, 120
Data Store Policy, 124
Load, Store, and Synchronization, 126
Enterprise Data Grid Integration, 129

12 Data Affinity **133**

A Measurable Quantity, 134
What to Expect from Data Affinity, 135
How to Achieve Data Affinity, 135
Regionalization, Synchronization, Distribution, and Data Affinity, 135
Data Distribution is Key to Data Affinity, 137
Data Affinity and Task Routing, 139
Integration of Compute and Data Grids, 139
Examples, 141

PART III PRACTICAL APPLICATIONS OF GRID COMPUTING

13 Which Applications are Good Candidates for the Grid 145

Grid Enabling Application Characteristics, 145
Atomic Tasks, 145
Complex Data Sets, 146
Data Collection, 146
Operations, 146
Gridable Applications, 147
Compute-Intensive Applications, 147
OLAP Data Analysis, 148
Data Center Operations, 148
Compute Utility Service, 149
Use Case Presentations, 149

14 Calculation-Intensive Applications 153

Description, 153
Use Cases, 154
General Architecture, 156
Data Grid Analysis, 160

15 Data Mining and Data Warehouses 165

Description, 165
Use Cases, 166
General Architecture, 168
First Use Case, 168
Second Use Case, 170
Enter the Compute Grid, 172
Data Grid Analysis, 172
Benefits and Data Grid Specifics, 174

16 Spanning Geographic Boundary 177

Description, 177
Business Use Cases, 178
Financial Services, 178
Operations, 180
Following the Sun, 183
General Architecture, 184
Data Grid Analysis, 185
Benefits and Data Grid Specifics, 188

17 Command and Control **191**

Problem Description, 191
Solution Architecture, 192
Command and Control Without a Data Grid, 193
Command and Control with a Data Grid, 194
Observations and Comparisons, 195
Data Grid Analysis, 196
Application Spinoffs, 202

18 Web Service's Role in the SOA/SONA Evolution **203**

Definition of Web Services, 203
Description, 205
Data Management: The Keystone to Web Services, 206
Web Services, Grid Infrastructures, and SONA, 208
The Undiscovered Past, 208
The SONA Model, 210
Connecting the Dots of the Past into the Continuum of the Present, 211
Service-Oriented Network Architecture (SONA), 212
Network Computing Power Explosion, 214
Consequences of Moore's and Metcalfe's Laws, 215
Isomorphism to Evolution of Previous Systems, 215
Grid and Web Services as Manifestation of State Transition, 215
Conclusion, 215

19 The Compute Utility **217**

Overview, 218
Architecture, 220
Geographic Boundary, 221
Command-and-Control Systems, 221
Macro/Microscheduling, 223

PART IV REFERENCE MATERIAL

20 Language Interface **229**

Programmatic, 230
Query-Based, 232
XML-Based, 234

21 Basic Programming Examples **235**

HelloWorld Example, 236
Coarse Granularity, 236

Coarse Data Atom, 236
Writer Program, 237
Reader Program, 239
Fine Granularity, 240
Writer Program, 240
Reader Program, 243
Random-Number Surface Example, 245

22 Additional Reading **251**

Useful Information Sources, 251
White Papers, 252
Grid Computing, 252
GridFTP, 252
Distributed File Systems, 252
Standards Bodies, 253
Globus—Data Grid, 253
Global Grid Forum, 253
W3C, 253
Public and University Grid Efforts, 253
Scientific Research Use of Grid Computing, 254
Web Services, 254
Distributed Computing, 255
Compute Utility, 255
Service-Oriented Architectures, 256
Data Affinity, 256

23 White Paper: Natural Attraction Forces of Data Bodies within a Data Grid to Describe Efficient Data Distribution Patterns **257**

Introduction, 257
Observation, 258
Hypothesis, 259
Laws of Attraction, 259
How Does This Fit in with Data Distribution Patterns of Single Data Bodies within a Data Grid Fabric?, 260
Collision of Single Data Bodies, 261
Effects of the Data Grid on a Single Data Body, 265
Conclusions, 265

24 Glossary of Terms **267**

REFERENCES **273**

INDEX **277**