Graph Databases

Ian Robinson, Jim Webber, and Emil Eifrem

O'REILLY

Beijing • Cambridge • Farnham • Koln • Sebastopol • Tokyo

Table of Contents

Foreword vii
Preface ix

1. Introduction 1
What Is a Graph? 1
A High-Level View of the Graph Space 4
Graph Databases 5
Graph Compute Engines 6
The Power of Graph Databases 8
Performance 8
Flexibility 8
Agility 9
Summary 9

2. Options for Storing Connected Data 11
Relational Databases Lack Relationships 11
NOSQL Databases Also Lack Relationships 14
Graph Databases Embrace Relationships 18
Summary 23

3. Data Modeling with Graphs 25
Models and Goals 25
The Property Graph Model 26
Querying Graphs: An Introduction to Cypher 27
Cypher Philosophy 27
START 29
MATCH 29
RETURN 30
Other Cypher Clauses 30

A Comparison of Relational and Graph Modeling 31
Relational Modeling in a Systems Management Domain 33
Graph Modeling in a Systems Management Domain 37
Testing the Model 38
Cross-Domain Models 40
Creating the Shakespeare Graph 44
Beginning a Query 45
Declaring Information Patterns to Find 46
Constraining Matches 47
Processing Results 48
Query Chaining 49
Common Modeling Pitfalls 50
Email Provenance Problem Domain 50
A Sensible First Iteration? 50
Second Times the Charm 53
Evolving the Domain 56
Avoiding Anti-Patterns 61
Summary 61

4. Building a Graph Database Application 63
Data Modeling 63
Describe the Model in Terms of the Application's Needs 63
Nodes for Things, Relationships for Structure 64
Fine-Grained versus Generic Relationships 65
Model Facts as Nodes 66
Represent Complex Value Types as Nodes 69
Time 70
Iterative and Incremental Development 72
Application Architecture 73
Embedded Versus Server 74
Clustering 78
Load Balancing 79
Testing 82
Test-Driven Data Model Development 83
Performance Testing 89
Capacity Planning 93
Optimization Criteria 93
Performance 94
Redundancy 97
Load 97

Summary 98

5. Graphs in the Real World 99
Why Organizations Choose Graph Databases 99
Common Use Cases 100
Social 100
Recommendations 101
Geo 102
Master Data Management 103
Network and Data Center Management 103
Authorization and Access Control (Communications) 104
Real-World Examples 105
Social Recommendations (Professional Social Network) 105
Authorization and Access Control 116
Geo (Logistics) 124
Summary 139

6. Graph Database Internals 141
Native Graph Processing 141
Native Graph Storage 144
Programmatic APIs 150
Kernel API 151
Core (or "Beans") API 151
Traversal API 152
Nonfunctional Characteristics 154
Transactions 155
Recoverability 156
Availability 157
Scale 159
Summary 162

7. Predictive Analysis with Graph Theory 163
Depth- and Breadth-First Search 163
Path-Finding with Dijkstra's Algorithm 164
The A* Algorithm 173
Graph Theory and Predictive Modeling 174
Triadic Closures 174
Structural Balance 176
Local Bridges 180
Summary 182

A. N0SQL Overview 183

Index.