Managing the Data Warehouse

W.H. Inmon
J.D. Welch
Katherine L. Glassey

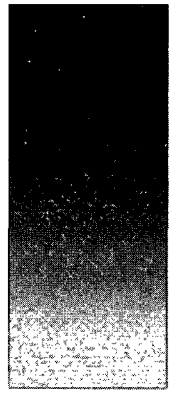
Technische Hochschule Darmstadt
FACHBEREICH INFORMATIK
BIBLIOTHEK
Inventar-Nr.: M96-03427
Sachgebiete: H-2
Standort: 1997

WILEY COMPUTER PUBLISHING

John Wiley & Sons, Inc.
New York • Chichester • Weinheim
Toronto • Singapore • Brisbane

Fachbereichsbibliothek Informatik
TU Darmstadt

59380370



Contents

	Dedication	**x**
	Preface	**xi**
1	**Monitoring Data Warehouse Activity**	**1**
	Monitoring Warehouse Activity	2
	What Data in the Data Warehouse Is Being Used?	4
	Who Is Using the Data Warehouse?	5
	What Is the Response Time?	6
	What Kind of Activities Are Being Submitted?	6
	Occurrences of Activity	7
	Granularity of the Activity Monitor	10
	Chargeback and Activity Monitoring	11
	An Activity Monitor—Why Not Any Other Monitor?	12
	Differences Between DSS and OLTP	14
	Measuring DSS Parameters of Activity	16
	What the Monitor Reports	17
	Administering the Data Warehouse Activity Monitor	20
	Summary	21
2	**Monitoring Data Warehouse Data**	**23**
	Monitor Differences	23
	From the DSS Analyst Perspective	25
	Basic Functions	26

Data Profiles 27
Thresholds of Data 30
Data Quality 32
Other Uses of a Data Warehouse Data Monitor 33
The Challenge of Managing Large Volumes of Data 35
Sectors 36
Administering the Data Warehouse Data Monitor 40
Complementary Nature of Monitoring Software 40
Summary 41

3 Managing Security in the Data Warehouse 43
Security as an Issue 43
Recognizing the Need for Security 44
Using DBMS for Security 46
A Fundamental Level of Security 47
States of Presentation to the Unauthorized User 50
Encryption Possibilities 50
Multiple Encryption Algorithms 52
The Encryption Algorithm 53
Encryption on a Worldwide Basis 54
Decrypting the Data 54
Managing the Decryption Keys 55
Workstation or End-User ID Authorization Keys 56
Decryption and Software Packages 57
Clearing the Terminal 58
The Overhead of Encryption and Decryption 59
Limiting DBMS Processing 60
Administering Security 62
Summary 63

4 Administering the Data Warehouse Data Model 65
The Data Warehouse Data Model 65
Where the Model Applies 68
Data-Mart Data Models 69
Interfaces 70
Summary 74

5 Managing Metadata in the Data Warehouse 75
The Metadata Infrastructure 75
Other Goals for the Metadata Infrastructure 77
Repository 78

Everybody Do Your Own Thing 78
A Unified Approach to Metadata Infrastructure 79
Components of the Architected Environment 79
Metadata—The Composition 82
Distributing Metadata across the Architected Environment 83
Metadata in the Operational Environment 84
ODS Shared Metadata 87
ODS DSS Analyst Shared and Local Metadata 88
Data Warehouse Shared Metadata 90
Data Warehouse DSS Analyst Shared and Local Metadata 91
Data Mart Shared Metadata 92
Data Mart DSS Analyst Shared and Local Metadata 93
Different Types of Decisions Made in Different Places 95
Justifying Metadata 95
Data Warehouse and Metadata 98
Context and Content of Historical Data 99
Versioning and Metadata 102
Metadata at the Workstation 102
Data Mart and Metadata 104
The Metadata Infrastructure 105
The Role of the Metadata Manager 106
Distributed Metadata 109
Summary 110

6 The User-Centric Data Warehouse 113
How Do Data Warehouses Differ from Other Systems? 115
Historical Information Access Process—An Illustration 116
The Ideal Environment for End-User Data Analysis and Decision Support 120
What Are the Hallmarks of a Successful Data Warehouse System? 122
What Are the Unique User Needs in the Data Warehouse? 124
A Transportation Analogy 125
What Is the Result? 127

7 Selecting End-User Access Tools 129
Including End-User Access Products 130
How to Get Started 134
Four Representative Examples 137
Company A 137
Company B 139

Company C 140
Company D 141
A Framework for "Best Practices Tool Selection" 142
Stage 1: First Cut 143
Stage 2: Request for Information 147
Stage 3: Hands-On Technical Evaluation 151
Stage 4: User Evaluation 153
Final Observations 155

8 Managing the Integration and Transformation Interface 161
Satisfying Parameters 161
Legacy DBMS Technologies 162
The Data Warehouse DBMS Technologies 163
Efficient Access of Legacy Data 164
Data Conversions 164
Operating System Conversions 165
Hardware Architecture Conversions 166
Converting Internal Structures 167
Resequencing Data 167
Application Conversions 168
Choosing from Multiple Sources 169
Supplying Default Values 170
Selecting Data Efficiently 171
Summarizing and Aggregating Data 173
Adding an Element of Time 173
Reformatting Date Fields 174
Capturing Metadata Interactively 175
Building the Interface Manually or by Automation 175
The Integration and Transformation Interface and the DWA 179
Summary 180

9 OLAP and the Data Warehouse Environment 183
Definition of OLAP 183
Architecture-Based OLAP Rules 184
The Source of OLAP Data 191
Differences between OLAP and Organizationally Structured Data Warehouse Data 192
OLAP with No Organizationally Structured Data 196
Explorers, Farmers, and Tourists 197
OLAP and the Workload 199

Mysteries of the Hypercube Revealed 201
Drill-Down Processing 205
Summary 207

10 Managing OLAP in the Data Warehouse Environment 209
Data Modeling for OLAP 209
Providing OLAP Capabilities in the Data Warehouse 213
The Star Schema and ROLAP 219
The Order of the Building Components 222
Data Acquisition for OLAP 222
External Data and OLAP 225
Indexing OLAP in the Data Warehouse Environment 226
Metadata for the OLAP Environment 228
Platforms and OLAP 231
Summary 235

11 Managed Redundancy in the Data Warehouse Environment 237
Types of Managed Redundancy 237
Redundancy between the Different Levels of the Data Warehouse Architecture 238
Redundancy Due to Denormalized Data 241
Redundancy Due to Data Artifacting 244
Redundancy Due to Different Granularity of the Same Data in the Same Level 247
Redundancy Due to Differences in Time-Variant Criteria 249
Summary 251

12 Data Warehouse Team Roles and Responsibilities 253
Data Warehouse Management Roles 255
Data Warehouse Administrator 255
Data Warehouse Organizational Change Manager 257
Database Administrator 259
Metadata Manager 261
Analysis and Design of the Data Warehouse Environment 262
Business Requirements Analysts 264
Data Warehouse Architect 267
Development of Data Warehouse Processes 270
Data Acquisition Developers 270
Data Access Developers 271
Data Warehouse Maintenance Developers 274
Supporting IS Roles 275

IS Executive Sponsor 276
Data Quality Analyst 276
Summary 280

13 Data Warehouse Staffing Requirements 283
Initial Development Staffing 283
Data Warehouse Administrator 285
Data Warehouse Organizational Change Manager 286
Database Administrator 286
Metadata Manager 286
Business Requirements Analysts 287
Data Warehouse Architect 287
Data Acquisition Developers 287
Data Access Developers 288
Data Quality Analysts 289
Data Warehouse Management Staffing 289
Data Warehouse Administrator 290
Data Warehouse Organizational Change Manager 291
Database Administrators 291
Metadata Manager 291
Business Requirements Analysts 292
Data Warehouse Architect 293
Data Acquisitions Developers 293
Data Access Developers 294
Data Quality Analysts 294
Data Warehouse Testers 294
Factors Affecting Growth in Data Warehouse Staffing 295
Summary 295

14 Data Warehouse End-User Roles and Responsibilities 297
Support Roles 297
Iteration Sponsors 297
Subject Matter Experts 299
User Support Technician 301
User Types 302
Unlimited Ad Hoc Access Users 302
Limited Ad Hoc Access Users 302
Predefined Application Users 305
Data Warehouse Initial End-User Staffing 307
Subject Matter Experts 307
User Support Technician 307

Ongoing Data Warehouse User Roles Staffing 308
Subject Matter Experts 309
User Support Technicians 309
Summary 310

15 Data Warehouse Project Principles 311
Setting Project Objectives 311
Determine the Type of Organization 311
Determine Project Size and Starting Point 314
Small is Beautiful 314
The Project Team 315
Database Designer 316
Data Archaeologists 316
Systems Programmers 317
Users and Trainers 317
Data Warehouse Administrator 317
Anticipating User Needs 317
Who Visits the Warehouse? 318
Implementation Objectives for Tool Selection 320
Data Design and Architecture Guidelines 323
System Architecture Considerations 323
Database Design for the Data Warehouse 324
Design Rules for Data Warehouse Databases 325
Star Schema or Dimensional Modeling 327
Data Considerations 329
Project Delivery Tactics 333
First Things First 334
Marketing the Project 334
Adopt a Customer-Focused Orientation 336
Deliver It Well 337
Summary 341
Write Your Own Rules 343
Anticipate Change 344
Ensure High Performance 344
Unlock User Creativity 345

Appendix PRISM Solutions Tech Topics 347

Recommended Reading 350

Glossary 357

Index 379