

Database Archiving

How to Keep Lots of Data for a Very Long Time

Jack E. Olson

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

ELSEVIER.

Morgan Kaufmann Publishers is an imprint of Elsevier

**MORGAN
KAUFMANN**

Contents

Preface	xiii
Acknowledgments	xvii

PART 1 Archiving Basics

CHAPTER 1 Database Archiving Overview	3
1.1 A Definition of Database Archiving	4
1.2 Forms of Data Archiving	6
1.3 The Data Lifeline	12
1.4 Types of Data Objects	15
1.5 Data Retention Requirements Versus Data Archives	16
1.6 The Database Archives and Other Database Types	19
CHAPTER 2 The Business Case for Database Archiving	21
2.1 Why Database Archiving Is a Problem Today	21
2.2 Implications of Not Keeping Data	25
2.3 Data Volume Issues	27
2.4 Change Management	30
2.5 Current Database Archiving Practices	31
CHAPTER 3 Generic Archiving Methodology	33
3.1 The Methodology	33
3.2 Define Motivation for Archiving I	35
3.3 Identify Objects to Archive	39
3.4 Determine When to Put Objects in the Archive	42
3.5 Determine How Long to Keep Objects in the Archive	44
3.6 Determine What to Do with Discarded Objects	41
3.7 Determine Who Needs Access to Archives and How	41
3.8 Determine the Form of Archive Objects	48
3.9 Determine Where the Archive Will Be Kept	50
3.10 Determine Operational Processes Needed	57
3.11 Determine Necessary Administrative Processes	59
3.12 Determine Required Change Processes	59
CHAPTER 4 Components of a Database Archiving System	63
4.1 The Database Archiving Organization	63
4.2 Archive Application Data Gathering	64
4.3 Archive Application Design	65

4.4	Archive Data Extraction	65
4.5	Archive Data Management	66
4.6	Archive Access	66
4.7	Archive Administration	67

PART 2 Establishing a Database Archiving Project

CHAPTER 5	Origins of a Database Archiving Application	71
5.1	Problems That Lead to Database Archiving Solutions	72
5.2	Recognizing Problems	73
5.3	Assignment of Problems for Initial Study	75
5.4.	Initial Problem Study Components	75
5.5	Determining the Basic Strategy for the Archiving Application	78
5.6	The Application Strategy Chart	83
CHAPTER, 6	Resources Needed	85
6.1	People	85
6.2	Authority	87
6.3	Education	88
6.4	Repository	88
6.5	Archive Server	89
6.6	Software Tools	89
6.7	Disk Storage	90
CHAPTER 7	Locating Data	93
7.1	Inventorying Data	93
7.2	Picking the Archivist's Data	99
7.3	Documenting Data Sources	100
CHAPTER 8	Locating Metadata	103
8.1	Metadata Definitions	103
8.2	Where to Find Metadata	109
8.3	Selecting a Version of the Metadata	113
8.4	Classifying Data	114
8.5	Documenting Metadata	117
8.6	Keeping Up with Changes	118
CHAPTER 9	Data and Metadata Validation	121
9.1	Matching Data to Metadata	121
9.2	Assessing Data Quality	123

9.3 Assessing Metadata Quality	124
9.4 Validating Data Classification	125
9.5 Documenting Validation Activities	125
9.6 Repeating Validation Activities	126

PART 3 Designing Database Archiving Applications

CHAPTER 10 Designing for Archive Independence	129
10.1 Independence from Application Programs	129
10.2 Independence from DBMS	132
10.3 Independence from Systems	133
10.4 Independence from Data Formats	133
CHAPTER 11 Modeling Archive Data	137
11.1 The Source Data Model	137
11.2 The Target Data Model	142
11.3 Model Representations	147
CHAPTER 12 Setting Archive Policies	149
12.1 Extract Policies	149
12.2 Archive Storage Policies	154
12.3 Archive Discard Policies	157
12.4 Validation and Approval	159
CHAPTER 13 Changes to Data Structures and Policies	161.
13.1 Archiving System Strategies for Handling Metadata Changes	161
13.2 Metadata Change Categories	163
13.3 Changes to the Archive Data Model	170
13.4 Managing the Metadata Change Process	171
13.5 Changes to Archive Policies	172
[^] 13.6 Maintaining an Audit Trail of Changes	176

PART 4 Database Archiving Application Software

CHAPTER 14 The "Archive Data Store	181
14.1 Archive Database Choices	182
14.2 Important Features	185
14.3 False Features	193
14.4 How Choices Stack Up	195

CHAPTER 15	The Archive Data Extraction Component	197
15.1	The Archive Extractor Model ,	198
15.2	Extractor Implementation Approaches	200
15.3	Execution Options	203
15.4	Integrity Considerations	205
15.5	Other Considerations	209
CHAPTER 16	The Archive Discard Component	215
16.1	Data Structure Considerations	215
16.2	Implementation Form	219
16.3	Integrity Considerations	220
16.4	Operational Considerations	222
16.5	Audit Trail	225
CHAPTER 17	The Archive Access Component	229
17.1	Direct Programming Access	229
17.2	Access Through Generic Query and Search Tools	231
17.3	Access Through Report Generators and BI Tools	231
17.4	Selective Data UNLOAD	232
17.5	Accessing Original Data	234
17.6	Metadata Services	236
17.7	Other Access Considerations	236
PART 5	Administration of the Database Archive	
<hr/>		
CHAPTER 18	Ongoing Auditing and Testing	243
18.1	Responsibility for Ongoing Auditing and Testing	244
18.2	Auditing Activities *	244
18.3	Testing Activities	248
18.4	Frequency of Audits and Tests	250
CHAPTER 19	Managing the Archive Over Time	251
19.1	Managing the Archive Database	251
19.2	Managing the Archive Repository	253
19.3	Using Hosted Solutions	255
19.4	Managing Archive Users	256
CHAPTER 20	Nonoperational Sources of Data	261
20.1	Retired Applications	261
20.2	Data from Mergers and Acquisitions	263
20.3	e-Discovery Applications	264
20.4	Business Intelligence Data	266
20.5	Logs, Audit Trails, and Other Miscellaneous Stuff	267