

Hasso Plattner Alexander Zeier

In-Memory Data Management

An Inflection Point for Enterprise Applications

fyA Springer

Contents

Foreword	XI
<i>By</i> <i>Prof. John L. Hennessy (Stanford University, California, USA), and</i> <i>Prof. David A. Patterson (University of California at Berkeley, USA)</i>	
Preface	XIII
Introduction	1
PART I - An Inflection Point for Enterprise Applications	5
1 Desirability, Feasibility, Viability-The Impact of In-Memory	7
1.1 Information in Real Time - Anything, Anytime, Anywhere.	7
1.1.1 Response Time at the Speed of Thought	9
1.1.2 Real-Time Analytics and Computation on the Fly.	10
1.2 The Impact of Recent Hardware Trends.	11
1.2.1 Database Management Systems for Enterprise Applications...	11
1.2.2 Main Memory Is the New Disk	14
1.2.3 From Maximizing CPU Speed to Multi-Core Processors.	15
1.2.4 Increased Bandwidth between CPU and Main Memory.	17
1.3 Reducing Cost through In-Memory Data Management	20
1.3.1 Total Cost of Ownership... ..	20
1.3.2 Cost Factors in Enterprise Systems	21
1.3.3 In-Memory Performance Boosts Cost Reduction.	22
1.4 Conclusion	23
2 Why Are Enterprise Applications So Diverse?	25
2.1 Current Enterprise Applications	25
2.2 Examples of Enterprise Applications	27
2.3 Enterprise Application Architecture	29
2.4 Data Processing in Enterprise Applications	30
2.5 Data Access Patterns in Enterprise Applications	31
2.6 Conclusion	31

3	SanssouciDB - Blueprint for an In-Memory Enterprise Database System	33
3.1	Targeting Multi-Core and Main Memory.	34
3.2	Designing an In-Memory Database System.	36
3.3	Organizing and Accessing Data in Main Memory.	37
3.4	Conclusion.	40
PART II	- SanssouciDB - A Single Source of Truth through In-Memory..	41
4	The Technical Foundations of SanssouciDB.	43
4.1	Understanding Memory Hierarchies.	43
4.1.1	Introduction to Main Memory.	44
4.1.2	Organization of the Memory Hierarchy.	47
4.1.3	Trends in Memory Hierarchies.	49
4.1.4	Memory from a Programmer's Point of View.	50
4.2	Parallel Data Processing Using Multi-Core and Across Servers.	57
4.2.1	Increasing Capacity by Adding Resources.	57
4.2.2	Parallel System Architectures.	59
4.2.3	Parallelization in Databases for Enterprise Applications.	61
4.2.4	Parallel Data Processing in-SanssouciDB.	64
4.3	Compression for Speed and Memory Consumption.	68
4.3.1	Light-Weight Compression	69
4.3.2	Heavy-Weight Compression	73
4.3.3	Data-Dependent Optimization.	73
4.3.4	Compression-Aware Query Execution.	73
4.3.5	Compression Analysis on Real Data	74
4.4	Column, Row, Hybrid - Optimizing the Data Layout.	75
4.4.1	Vertical Partitioning	75
4.4.2	Finding the Best Layout	78
4.4.3	Challenges for Hybrid Databases.	81
4.5	The Impact of Virtualization.	81
4.5.1	Virtualizing Analytical Workloads.	82
4.5.2	Data Model and Benchmarking Environment	82
4.5.3	Virtual versus Native Execution.	83
4.5.4	Response Time Degradation with Concurrent VMs.	84
4.6	Conclusion.	86
5	Organizing and Accessing Data in SanssouciDB.	89
5.1	SQL for Accessing In-Memory Data	90
5.1.1	The Role of SQL.	90
5.1.2	The Lifecycle of a Query.	91
5.1.3	Stored Procedures.	91
5.1.4	Data Organization and Indices.	91
5.2	Increasing Performance with Data Aging	92
5.2.1	Active and Passive Data	93

- 5.2.2 Implementation Considerations for an Aging Process 95
- 5.2.3 The Use Case for Horizontal Partitioning of Leads. 95
- 5.3 Efficient Retrieval of Business Objects 98
 - 5.3.1 Retrieving Business Data from a Database 98
 - 5.3.2 Object Data Guide. 99
- 1 5.4 Handling Data Changes in Read-Optimized Databases ; 100
 - 5.4.1 The Impact on SanssouciDB. 101
 - 5.4.2 The Merge Process. 103
 - 5.4.3 Improving Performance with Single Column Merge 107
- 5.5 Append, Never Delete, to Keep the History Complete. 109
 - 5.5.1 Insert-Only Implementation Strategies. 110
 - 5.5.2 Minimizing Locking through Insert-Only. 111
 - 5.5.3 The Impact on Enterprise Applications. 114
 - 1 5.5.4 Feasibility of the Insert-Only Approach 117
- 5.6 Enabling Analytics on Transactional Data. 118
 - 5.6.1 Aggregation on the Fly. 119
 - 5.6.2 Analytical Queries without a Star Schema...; ;*..... 128
- 5.7 Extending Data Layout without Downtime. 135
 - 5.7.1 Reorganization in a Row Store. 135
 - 5.7.2 On-The-Fly Addition in a Column Store. 136
- 5.8 Business Resilience through Advanced Logging Techniques. 137
 - 1 5.8.1 Recovery in Column Stores 138
 - 5.8.2 Differential Logging for Row-Oriented Databases. 140
 - 5.8.3 Providing High Availability..... 141
- 5.9 The Importance of Optimal Scheduling for Mixed Workloads. 142
 - 5.9.1 Introduction to Scheduling. 142
 - 5.9.2 Characteristics'of a Mixed Workload. 145
 - 5.9.3 Scheduling Short and Long Running Tasks. 146
- 5.10 Conclusion. 148

PART III - How In-Memory Changes the Game. 151

- 6 Application Development 153**
 - 6.1 Optimizing Application Development for SanssouciDB 153
 - 6.1.1 Application Architecture. 154
 - 6.1.2 Moving Business Logic into the Database 155
 - 6.1.3 Best Practices. 157
 - 6.2 Innovative Enterprise Applications. 158
 - 6.2.1 New Analytical Applications. 158
 - 6.2.2 Operational Processing to Simplify Daily Business. 162
 - 6.2.3 Information at Your Fingertips with Innovative User-Interfaces . 164
 - 6.3 Conclusion. 169

7 Finally, a Real Business Intelligence System Is at Hand	171
7.1 Analytics on Operational Data	171
7.1.1 Yesterday's Business Intelligence.	171
7.1.2 Today's Business Intelligence.	174
7.1.3 Drawbacks of Separating Analytics from Daily Operations	176
7.1.4 Dedicated Database Designs for Analytical Systems	178
7.1.5 Analytics and Query Languages.	180
7.1.6 Enablers for Changing Business Intelligence.	182
7.1.7 Tomorrow's Business Intelligence.	183
•7.2 How to Evaluate Databases after the Game Has Changed	185
7.2.1 Benchmarks in Enterprise Computing	185
7.2.2 Changed Benchmark Requirements for a Mixed Workload....	187
7.2.3 A New Benchmark for Daily Operations and Analytics.	188
7.3 Conclusion	192
8 Scaling SanssouciDB in the Cloud	193
8.1 What Is Cloud Computing?.	194
8.2 Types of Cloud Applications.	195
8.3 Cloud Computing from the Provider Perspective.	197
8.3.1 Multi-Tenancy.	197
8.3.2 Low-End versus High-End Hardware	201
8.3.3 Replication.	201
8.3.4 Energy Efficiency by Employing In-Memory Technology	202
8.4 Conclusion.	204
9 The In-Memory Revolution Has Begun	205
9.1 Risk-Free Transition to In-Memory Data Management	205
9.1.1 Operating In-Memory and Traditional Systems Side by Side ..	206
9.1.2 System Consolidation and Extensibility.	207
9.2 Customer Proof Points.	208
9.3 Conclusion.	209
References.	211
About the Authors.	221
Glossary.	223
Abbreviations.	231
Index.	233