

Contents

Preface	vii
1 Innovations for Personalized Medicine	1
Hasso Plattner, Matthieu-P. Schapranow and Franziska Häger	
1.1 Requirements for Personalized Medicine	1
1.1.1 Researchers	3
1.1.2 Clinicians	5
1.1.3 Patients	6
1.2 Interdisciplinary Teams	6
1.3 Trends in Hardware	9
1.4 In-memory Technology Building Blocks	13
1.4.1 Combined Column and Row Store	14
1.4.2 Complete History	14
1.4.3 Lightweight Compression	14
1.4.4 Partitioning	15
1.4.5 Multi-core and Parallelization	16
1.4.6 Active and Passive Data Store	16
1.4.7 Reduction of Layers	17
1.5 High-performance In-memory Genome Platform	17
1.5.1 Application Layer with Micro Applications	19
1.5.2 Platform Layer	21
1.5.3 Data Layer	22
1.6 Structure of the Work	22
1.7 References	25

Part I Data Processing in Personalized Medicine

2	Modeling Genome Data Processing Pipelines	31
	Marie Schäffer	
2.1	Introduction	31
2.2	Related Work	32
2.3	Modeling of Genome Data Processing Pipelines	35
2.3.1	Requirements Engineering	36
2.3.2	Modeling of Execution Semantics	38
2.3.3	Machine Readable Model Representation	43
2.4	Application Example	46
2.4.1	Pipeline Configuration User Interface	46
2.4.2	Data Format for Pipelines	48
2.5	Evaluation and Discussion	50
2.6	Conclusion and Outlook	52
2.7	References	53
3	Scheduling and Execution of Genome Data Processing Pipelines ..	55
	Cornelius Bock	
3.1	Introduction	55
3.2	Related Work	56
3.3	Method	59
3.3.1	Requirements of the Execution Environment	59
3.3.2	In-memory Database as Scheduler	60
3.3.3	Real-time Scheduling	62
3.4	Application Example	65
3.4.1	Architecture	65
3.4.2	Application of Design Patterns	69
3.4.3	Implementations of Scheduling Policies	70
3.5	Evaluation and Discussion	73
3.6	Conclusion and Outlook	73
3.7	References	74
4	Exchanging Medical Knowledge	75
	Franz Liedke	
4.1	Introduction	75
4.2	Related Work	76
4.2.1	In-memory Databases	76
4.2.2	Sharing Medical Knowledge	77
4.2.3	Requirements	79
4.3	Application Example	80
4.3.1	Use Case	80
4.3.2	System Requirements	80
4.3.3	Installing Applications	83
4.3.4	Configuring Applications	85
4.3.5	Cohort Analysis	86

- 4.4 Benchmarks 87
 - 4.4.1 Method 87
 - 4.4.2 Results 90
- 4.5 Evaluation and Discussion 90
 - 4.5.1 Predicate Scan 92
 - 4.5.2 Aggregation 93
 - 4.5.3 Join 94
- 4.6 Conclusion and Outlook 96
- 4.7 References 99

- 5 Billing Processes in Personalized Medicine 101**
 - Joseph Bethge
 - 5.1 Introduction 101
 - 5.2 Related Work 103
 - 5.3 Requirements Engineering 104
 - 5.3.1 Entity Definition 104
 - 5.3.2 Free Price System 106
 - 5.3.3 Calculation of Usage Fees 107
 - 5.3.4 In-memory Database Technology 108
 - 5.4 Application Example 110
 - 5.4.1 Database Schema 110
 - 5.4.2 Database Functionality 112
 - 5.4.3 User Interface 113
 - 5.4.4 Billing Process 114
 - 5.5 Benchmarks 115
 - 5.6 Evaluation 117
 - 5.6.1 Impact of Transaction Log Size 118
 - 5.6.2 Impact of Data Partitioning 119
 - 5.7 Conclusion and Outlook 120
 - 5.8 Appendix 121
 - 5.9 References 122

Part II Real-Time Data Analysis in Personalized Medicine

6	Real-time Analysis of Patient Cohorts	127
	Ricarda Schüler	
6.1	Introduction	127
6.2	Related Work	128
6.2.1	Cohort Analysis	128
6.2.2	K-Means Clustering	130
6.2.3	Hierarchical Clustering	131
6.2.4	In-memory Technology Building Blocks	132
6.3	Application Example	133
6.3.1	Architecture	133
6.3.2	Application	133
6.4	Benchmarks	137
6.4.1	Benchmark Data	137
6.4.2	Benchmarks for In-memory Technology	139
6.4.3	Benchmarks for R	140
6.4.4	Impact of Selected Variables	142
6.4.5	Test Procedure and Technical Environment	142
6.5	Results and Discussion	143
6.5.1	Data Size	144
6.5.2	Number of Genes	147
6.6	Conclusion and Outlook	149
6.7	References	151
7	Ad-hoc Analysis of Genetic Pathways	153
	Dominik Müller	
7.1	Introduction	153
7.2	Related Work	154
7.2.1	Pathway Analysis	156
7.2.2	Existing Analysis Possibilities	157
7.2.3	Storing a Graph	158
7.2.4	NoSQL	159
7.2.5	In-memory Database Technology	160
7.3	Creating an Integrated Pathway Database	161
7.4	Application Example	165
7.5	Benchmarks	167
7.5.1	Occurrences of Analyzed Genes	169
7.5.2	Genes in Input Set	169
7.5.3	Integrated Pathways	170
7.6	Results and Discussion	170
7.6.1	Occurrences of Analyzed Genes	170
7.6.2	Genes in Input Set	171
7.6.3	Integrated Pathways	172
7.7	Conclusion and Outlook	174

7.8	Appendix	175
7.9	References	178
8	Combined Search in Structured and Unstructured Medical Data . .	181
	David Heller	
8.1	Introduction	181
8.2	Related Work	182
8.3	In-memory Database Text Analysis Features	185
8.4	Application Example	186
	8.4.1 Customized Biomedical Dictionaries	187
	8.4.2 Customized Extraction Rules	188
	8.4.3 Post-processing of Text Analysis Results	190
	8.4.4 Trial Filtering	191
	8.4.5 User Interface	192
8.5	Benchmarks	193
	8.5.1 Genes	195
	8.5.2 Partitioning the Text Analysis Result Table	196
	8.5.3 Indexed Trials	196
8.6	Discussion	197
8.7	Conclusion and Outlook	201
8.8	Appendix	202
8.9	References	204
9	Real-time Collaboration in the Course of Personalized Medicine .	207
	Hasso Plattner and Matthieu-P. Schapranow	
9.1	Real-time Combination of Oncology Data	207
	9.1.1 Holistic Patient View	209
	9.1.2 Search in Structured and Unstructured Data	210
	9.1.3 Real-time Analysis of Patient Cohorts	211
9.2	Building Research Hypotheses	211
9.3	Pharmaceutical Feedback Loop	212
9.4	Federal Bureau of Statistics	213
9.5	Health Insurance Companies	214
9.6	Tumor Board of the Future	214
9.7	References	216
	List of Abbreviations	217
	Index	221



<http://www.springer.com/978-3-319-03034-0>

High-Performance In-Memory Genome Data Analysis
How In-Memory Database Technology Accelerates
Personalized Medicine

Plattner, H.; Schapranow, M.-P. (Eds.)

2014, XXI, 223 p. 78 illus., Hardcover

ISBN: 978-3-319-03034-0