

Table of contents

I	Foundations	1
1	Introduction	3
1.1	Problem Statement	5
1.2	Contributions	7
1.2.1	Serverless Data Processing Tensions	7
1.2.2	(Re-)Design of Serverless Data Processing Systems	8
1.2.3	Application Platform Co-Design	8
1.3	Published Work	9
1.4	Organization	11
2	Background	13
2.1	Data Processing Platforms	14
2.1.1	Analytics Engines	15
2.1.2	Execution and Development Model	15
2.1.3	Performance Assessment Methods	16
2.2	Data Processing Applications	17
2.2.1	Stream Processing	18
2.2.2	Batch Processing	19
2.2.3	Exploratory Ad-Hoc Data Processing	19
2.3	Serverless Computing Platforms	20
2.3.1	Function-as-a-Service Platforms	21
2.3.2	Existing Serverless Platforms	23
2.3.3	Development Trends	25
2.4	Serverless Data Processing	25
2.5	Summary	27
II	Serverless Data Processing	29
3	Serverless Data Processing Survey	33
3.1	Search Methodology	34
3.2	Survey Results	35
3.2.1	Applications and Frameworks	37
3.2.2	Workloads and Use-Cases	40
3.2.3	Assessment Methods	41
3.3	Synthesis	44
4	Characterizing Serverless Data Processing Tensions	47
4.1	Serverless Data Processing Architecture	47
4.1.1	Architecture Overview	48
4.1.2	User Defined Functionality	49
4.1.3	Driver	49
4.1.4	Serverless Computing Platform	50

4.1.5	Storage Systems	51
4.2	Architecture Tensions in Serverless Data Processing	52
4.3	Modelling Serverless Data Processing Quality	53
4.3.1	Application Qualities	54
4.3.2	Platform Qualities	57
5	Measurement Methods for Serverless Data Processing Co-Design	61
5.1	System Assessment Design	61
5.1.1	Application Assessments	62
5.1.2	Platform Assessments	63
5.1.3	Metrics	65
5.2	Co-Design Measurement Methods Implementation	68
5.2.1	Experiment Environments	69
5.2.2	Result Interpretation	71
III	CREW: Design and Implementation	75
6	Re-Designing Serverless Applications and Platforms	79
6.1	Orchestration	81
6.1.1	General Workaround Strategies	81
6.1.2	Invocation Generation	83
6.1.3	Observability	88
6.1.4	Evaluation	90
6.1.5	Combining Designs	99
6.2	State Management	100
6.2.1	General Workaround Strategies	101
6.2.2	Intermediate Data Transport	102
6.2.3	Programming Model	104
6.2.4	Evaluation	107
6.2.5	Design Selection	111
6.3	Computation	112
6.3.1	General Workaround Strategies	112
6.3.2	Deployment and Sizing	112
6.3.3	Serverless Executions	115
6.3.4	Evaluation	116
6.3.5	Design Selection	120
6.4	Discussion and Summary	121
7	Implementation and Evaluation	123
7.1	CREW Implementation	123
7.1.1	Execution Steps	125
7.1.2	Processing Steps	128

7.1.3	Open Issues	129
7.2	Evaluation	130
7.2.1	Workload & Protocol	130
7.2.2	Results	131
7.2.3	Analysis	136
7.3	Discussion and Summary	137
IV	Application Platform Co-Design	139
8	Application Platform Co-Design	143
8.1	The Co-Design Phenomenon	143
8.2	The Co-Design Cycle	145
8.3	Implications	147
9	Methods for Application Platform Co-Design	149
9.1	Experiment-Driven Application Platform Co-Design Method	149
9.1.1	Goals	150
9.1.2	Roles	150
9.1.3	Steps	153
9.2	Managing Design Tensions	156
9.2.1	Impact Dimensions	157
9.2.2	Trade-off Analysis	159
9.2.3	Observations	160
9.3	Conclusion and Discussion	161
V	Fin.	163
10	Conclusion	165
10.1	Summary	165
10.2	Outlook	169
	List of figures	171
	List of tables	173
	References	175
	Appendices	187
I	Source Code	187
II	Configurations	193
III	Serverless Data Processing Survey	195