

Contents

1. Introduction	1
I. Experimentation	9
2. Background and Related Work	11
2.1. Background	12
2.1.1. Modern Distributed Database Systems	12
2.1.2. Cloud Deployment and Management Automation	14
2.2. Related Work	17
2.2.1. Performance of Cloud-based Distributed Systems	17
2.2.2. Elastic Scalability of Cloud-based Distributed Systems	21
2.2.3. Availability of Cloud-based Distributed Systems	25
2.2.4. Consistency of Cloud-based Distributed Systems	26
2.2.5. Cost of Cloud-based Distributed Systems	27
2.2.6. Experiment Automation for Distributed Systems	27
3. Reproduced Experiments with Cloud-based Distributed Database Systems	35
3.1. Experiment Reproduction	36
3.2. Hotspot Performance Experiments	36
3.2.1. Related Hotspot Experiments by Trushkowsky et al.	37
3.2.2. Sherpa Performance under Hotspot Workloads	37
3.2.3. Cassandra Performance under Hotspot Workloads	40
3.2.4. Discussion and Comparison of Results	41
3.3. Scalability Experiments	42
3.3.1. Reproduction of Experiments	43
3.3.2. Discussion and Comparison of Results	43
3.4. Elasticity Experiments	44
3.4.1. Reproduction of Experiments	45
3.4.2. Discussion and Comparison of Results	47
3.5. Summary and Discussion	48

4. Experiments with Cloud-based Distributed Systems and Applications	51
4.1. Cassandra-Hadoop Cluster System Experiments	51
4.1.1. Cassoop Architecture	52
4.1.2. Experiment Setup	53
4.1.3. Selected Results	55
4.1.4. Discussion of Results	56
4.2. Web Application System Experiments	56
4.2.1. Web Application Architecture	57
4.2.2. Experiment Setup	57
4.2.3. Experiment Results and Discussion	58
4.3. E-Mail Application System Experiments	59
4.3.1. E-Mail Application Architecture	59
4.3.2. Experiment Setup	60
4.3.3. Experiment Results and Discussion	62
5. Automating Experiments with Cloud-based Distributed Systems	63
5.1. Experiment Automation System Design	64
5.1.1. Experiment Client	65
5.1.2. Experiment Controller	66
5.1.3. Infrastructure Controller	67
5.1.4. Software Controller	68
5.1.5. Benchmarking Controller	69
5.1.6. Measurement Controller	70
5.2. Experiment Automation System Implementations	70
5.2.1. Elastic Lab	70
5.2.2. Elastic Lab 2.0	77
5.2.3. Elastic Lab for Continuous Experimentation	83
5.3. Comparison of Implementations	86
5.3.1. Architecture Design Alternatives	87
5.3.2. Client Alternatives	89
5.3.3. Infrastructure Controller Design Alternatives	89
5.3.4. Software Controller Design Alternatives	90
5.3.5. Benchmarking Controller	92
5.4. Summary and Discussion	92
5.4.1. Generic vs. Specific Implementations	92
5.4.2. Benefits of Experiment Automation	94
5.4.3. Limitations of Experiment Automation	95

II. Trade-offs	97
6. Background and Related Work	99
6.1. Background	99
6.1.1. A Priori Methods	100
6.1.2. Interactive Methods	102
6.2. Related Work	103
7. Trade-off Evaluation Method	111
7.1. Experiment-Driven Trade-off Evaluation Method (ETEM)	111
7.1.1. Roles	111
7.1.2. Four Phases of the Method	113
7.1.3. Trade-Off Notation	114
7.2. Instantiations of ETEM	117
7.2.1. Trade-Off Refinement Method	117
7.2.2. Experiment-Driven Multi-Objective Optimization Method	117
8. Trade-offs in Cloud-based Distributed Systems	125
8.1. Scaling Trade-offs	125
8.1.1. Vertical versus Horizontal Scale-up	125
8.1.2. Scaling Speed versus Performance Stability	128
8.2. Replication Trade-offs	130
8.2.1. Experiments and Regression Models	131
8.2.2. Solving the MOP & Iterations	136
8.2.3. Availability Evaluation	140
8.3. Summary and Discussion	140
9. Conclusion	145
Appendix	151
A. Experiment Concepts and Methods	153
A.1. Experiment Concepts and Approaches	153
A.1.1. Experiment Terminology	153
A.1.2. Experiment Setup	154
A.1.3. Experiment Styles	155
A.2. Response Surface Methodology (RSM)	156
A.2.1. General Multiple Regression Models	156
A.2.2. Experiment Designs	157

Contents

B. Experiment Setups	159
B.1. Workloads	159
B.1.1. Hotspot Workload Modeling	159
B.1.2. Workload overview	161
B.2. Measurements	163
B.2.1. Sherpa Performance under Hotspot Workloads	163
B.2.2. Cassandra Performance under Hotspot Workloads	166
B.2.3. Cassandra and HBase Scalability	167
B.2.4. Selected Results	169
B.3. Elasticity Experiments	171
B.3.1. Measurements	172
B.4. Cassandra-Hadoop Cluster System Experiments	173
B.4.1. Experiment Setup	173
B.4.2. Selected Results	173