The Serverless Application Analytics Framework: Reproducible Performance for Performance Modeling

Robert Cordingly
School of Engineering and Technology
University of Washington
Tacoma WA USA
rcording@uw.edu

Wen Shu
School of Engineering and Technology
University of Washington
Tacoma WA USA
shuwen12@uw.edu

Wes J. Lloyd
School of Engineering and Technology
University of Washington
Tacoma WA USA
wlloyd@uw.edu

TABLE I. PERFORMANCE MODEL RESULTS SCNMT2 WORKLOADS

RAM (tenants)	256MB (13)	512MB (6)	1024MB (3)	2048MB (1)
	Avg runtime (ms), % Error, RMSE, deg of freedom			
CPU v2→v3	18173, -.5%, 199, 127	8189, -1.3%, 221, 58	3741, 0.11%, 36.0, 28	2580, -0.4%, 18.2, 155
CPU v2→v4	16532, -.1%, 224, 151	8396, -0.3%, 127, 414	4218, 0.09%, 102, , 406	2447, -0.3%, 14.4, 265
CPU v3→v4	17524, -.4%, 181, 127	7844, -.08%, 55.7, 58	3575, 0.24%, 43.6, 28	2487, -.54%, 18.5, 155

CPU (v2,3,4)	CPU v2	CPU v3	CPU v4
	Avg runtime (ms), % Err, RMSE, deg of freedom		
256(13)→512(6)	10324, 0.28%, 199, 414	8189, -0.13%, 48.0, 58	7722, 0.076%, 41.1, 151
256(13) →1024(3)	5252, -0.64%, 124, 406	3731, -0.44%, 22.1, 28	3894, -0.0049%, 33.0, 151
256(13)→2048(1)	2408, -0.38%, 19.0, 265	2578, -0.04%, 12.2, 127	2346, -0.33%, 17.3, 151

TABLE II. PERFORMANCE MODEL RESULTS SCMT2 WORKLOADS

RAM (tenants)	256MB (13)	512MB (6)	1024MB (3)	2048MB (1)
	Avg runtime (ms), % Error, RMSE, deg of freedom			
CPU v2→v3	51474, -.7%, 859, 127	23440,-.23%, 286, 58	10239,-.55%, 118, 28	6450, -.63%, 115, 157
CPU v2→v4	53962,-.38%, 487, 151	27281, .35%, 2912, 414	13373,-1.8%, 1516, 406	7240, 1.84%, 954, 266
CPU v3→v4	57203,-.49%, 494, 127	25074,.042%, 283, 58	11090,.431%, 174, 28	7011, 4.66%, 900, 157

CPU (v2,3,4)	CPU v2	CPU v3	CPU v4
	Avg runtime (ms), % Err, RMSE, deg of freedom		
256(13)→512(6)	48053, -.370%, 6648, 414	23439, -0.557%, 265, 58	24936, -0.49%, 235, 151
256(13) →1024(3)	21088, -.032%, 2832, 406	10239, 0.555%, 106, 28	12241, -1.02%, 281, 151
256(13)→2048(1)	8116, 1.65%, 1138, 266	6505, -1.42%, 135, 127	6872, -0.69%, 280, 151

TABLE III. PERFORMANCE MODEL RESULTS SCSMT2 WORKLOADS

RAM (tenants)	256MB (13)	512MB (6)	1024MB (3)	2048MB (1)
	Avg runtime (ms), % Error, RMSE, deg of freedom			
CPU v2→v3	41820,-.59%, 1273, 1414	20840,.312%, 728, 652	10174,0.54%, 373, 325	4699, .205%, 321, 1739
CPU v2→v4	44162,-.65%, 2037, 1613	20829,4.53%, 2875, 4572	9831, 4.30%, 1290, 4478	5052, 3.16%, 732, 2938
CPU v3→v4	46539, .24%, 763, 1414	22212,-.56%, 403, 652	10585,-.09%, 318, 325	5173,-.388%, 507, 1739

CPU (v2,3,4)	CPU v2	CPU v3	CPU v4
	Avg runtime (ms), % Err, RMSE, deg of freedom		
256(13)→512(6)	35216, 0.965%, 8389, 4572	20840, -0.39%, 355, 652	19711, 0.0476%, 725, 1613
256(13) →1024(3)	14645, 0.472%, 2625, 4478	10173, -1.02%, 568, 325	9049, 1.212%, 838, 1613
256(13)→2048(1)	5004, 3.97%, 1164, 2938	4526, 0.401%, 594, 1414	4639, 3.38%, 1110, 1613

TABLE IV. PERFORMANCE MODEL RESULTS SCNMT2 AWS→IBM

RAM (tenants)	256MB (13 / 32)	512MB (6 / 16)	1024MB (3 / 8)	2048MB (1 / 4)
	Avg runtime (ms), % Error, RMSE, deg of freedom			
AWS v2 → IBM 2.0 GHz	n/a	17666,-4.3%, 1196, 142	8013,-3.37%, 665, 236	4340,-2.38%, 559, 139
AWS v2→ IBM 2.1 GHz	n/a	16231,-4.1%, 1991, 369	7794,-1.37%, 674, 395	3639,-0.70%, 186, 265
AWS v2→ IBM 2.2GHz	30124,-4.49%, 1752, 171	13960, .13%, 1176, 99	6611. 1.40%, 287, 49	2920,-.166%, 41.49, 42
AWS v2→ IBM 2.6 GHz	24864, -2.7%, 2298, 200	12437,-5.5%, 2357, 85	6641, 3.76%, 633, 70	2549, 3.47%, 187, 24

Deleted: *Abstract*— Next generation software built for the cloud has recently embraced serverless computing platforms that leverage containers and microservices to form resilient, loosely coupled systems that are observable, easy to manage, and extend. Serverless architectures enable decomposing software into independent components packaged and run using isolated containers or microVMs. This decomposition approach enables application hosting using very fine-grained cloud infrastructure enabling cost savings as deployments are billed granularly for resource use. Adoption of serverless platforms promise reduced hosting costs while achieving high availability, fault tolerance, and dynamic elasticity. These benefits are offset by pricing obfuscation, as performance variance from CPU heterogeneity, multitenancy, and provision variation obscure the true cost of hosting applications with serverless platforms. Where determining hosting costs for traditional VM-based application deployments simply involves accounting for the number of VMs and their uptime, predicting hosting costs for serverless applications can be far more complex. To address these challenges, in this paper we introduce the Serverless Application Analytics Framework (SAAF) that supports profiling FaaS workload performance, resource utilization, and infrastructure to enable accurate performance predictions. We demonstrate the use of Linux CPU time accounting principles and multiple regression to estimate function runtime with <~2% error across heterogeneous CPUs, alternate memory settings, and to alternate FaaS platforms.

Keywords— Serverless Computing, Function-as-a-Service, Performance Evaluation, Performance Modeling, Resource Contention, Multitenancy ... [1]

Deleted: <#>SAAF RUNTIME, RESOURCE UTILIZATION, AND CONFIGURATION METRICS: 'Δ' INDICATES RAW AND DELTA VERSIONS ARE PROVIDED, '*' INDICATES FIELD GENERATED BY THE FAAS RUNNER ... [2]

Formatted: Font: 7 pt

Formatted: table head, Left, Indent: First line: 0", Space Before: 0 pt

Deleted: TO PRODUCE FAAS WORKLOAD RUNTIME PREDICTIONS WE PROFILED OUR WORKLOADS ON A SINGLE SOURCE PLATFORM, AND GENERATED REGRESSION MODELS TO CONVERT CPUUSR AND CPUIDLE TO A VARIETY OF TARGET ... [3]

Deleted: <#>EXPERIMENTAL RESULTS
<#>To evaluate our research questions, we profiled the workloads described in Table 2 on source and target platforms described in Table 3. We deployed AWS Lambda ... [4]

Formatted: Font: 8 pt

Formatted: Font: 8 pt, Do not check spelling or grammar

Formatted: Font: 8 pt, Do not check spelling or grammar

Formatted: Font: 8 pt, Do not check spelling or grammar

Formatted: Font: 8 pt, Do not check spelling or grammar

Formatted: Indent: First line: 0.2"

Page 1: [1] Deleted	Robert W. Cordingly	11/21/19 10:01:00 PM
Page 1: [2] Deleted	Robert W. Cordingly	11/21/19 10:02:00 PM
Page 1: [3] Deleted	Robert W. Cordingly	11/21/19 10:01:00 PM
Page 1: [4] Deleted	Robert W. Cordingly	11/21/19 10:01:00 PM