Thesis Defense

The AFIT of Today is the Air Force of Tomorrow.

CLOUD BENCHMARK TESTING OF CASSANDRA ON RASPBERRY PI FOR INTERNET OF THINGS CAPABILITY

Daniel Richardson, Capt, USAF
17M-GE

Overview

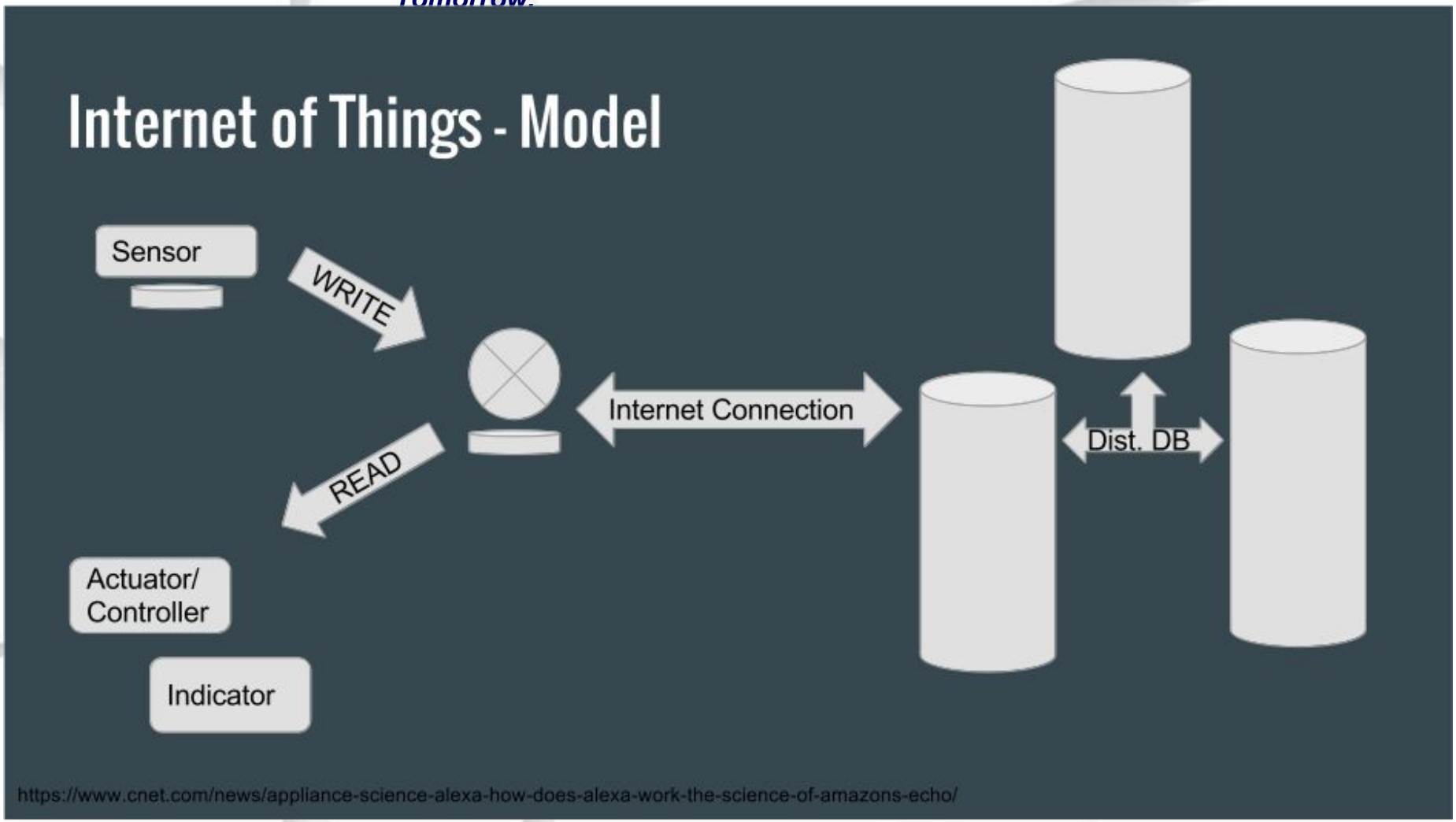
- Motivation
- Problem Statement
- Contributions
- Background and Related Works
- Experiments
- Results
- Conclusions

Motivation

- Application Space
- Traditional v. Thicker Client Model

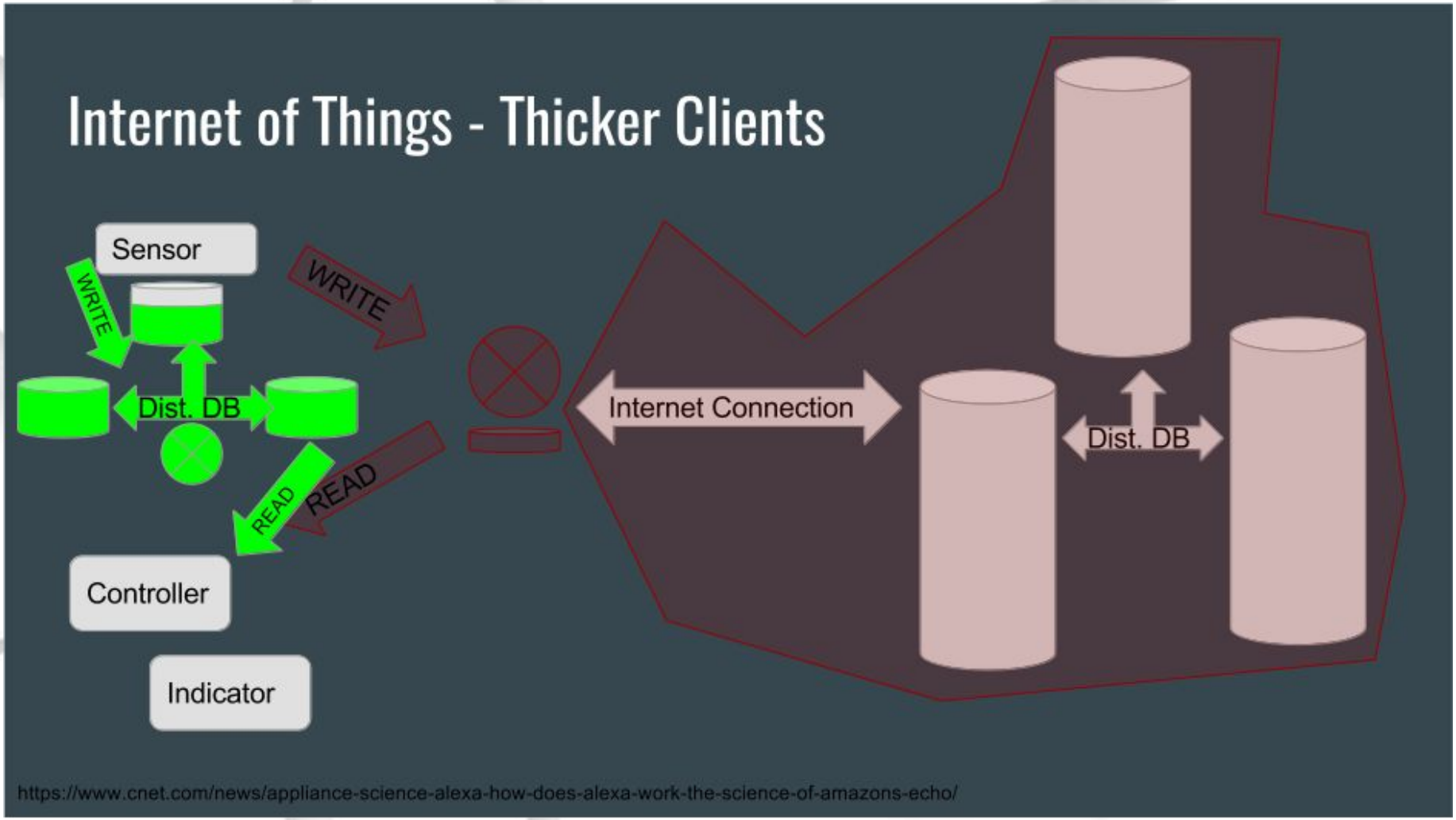
Motivation

The AFIT of Today is the Air Force of Tomorrow.



Motivation

The AFIT of Today is the Air Force of Tomorrow.



Problem Statement

- This research seeks to characterize, if any, conditions for feasible operation of distributed database technology on limited hardware.

Contributions

- Framework for Evaluation
- Insight into Scalability for Both a Wired and Wireless Configurations
- Performance Comparison between physical devices and virtual devices

Research Objectives

- Characterization
- Feasibility

Research Questions

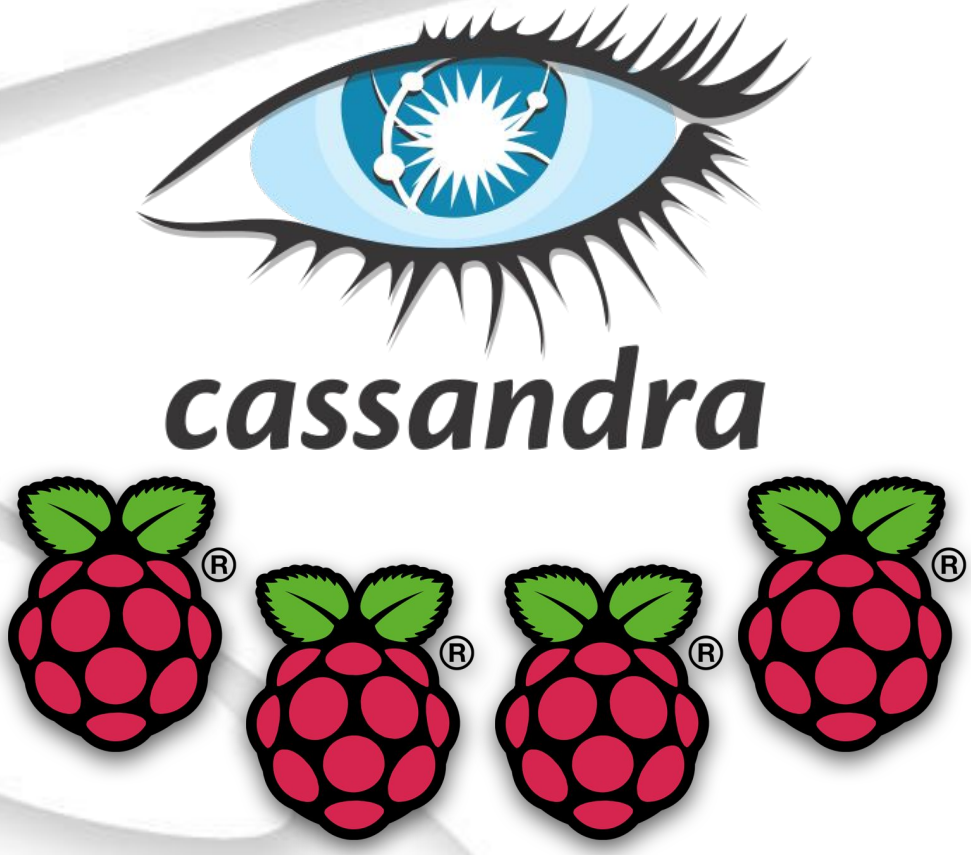
- Effect Characterization: Timing and Scalability
  - Variation in RAM
  - Wired vs. Wireless
  - Hardware vs. Virtual

Related Work

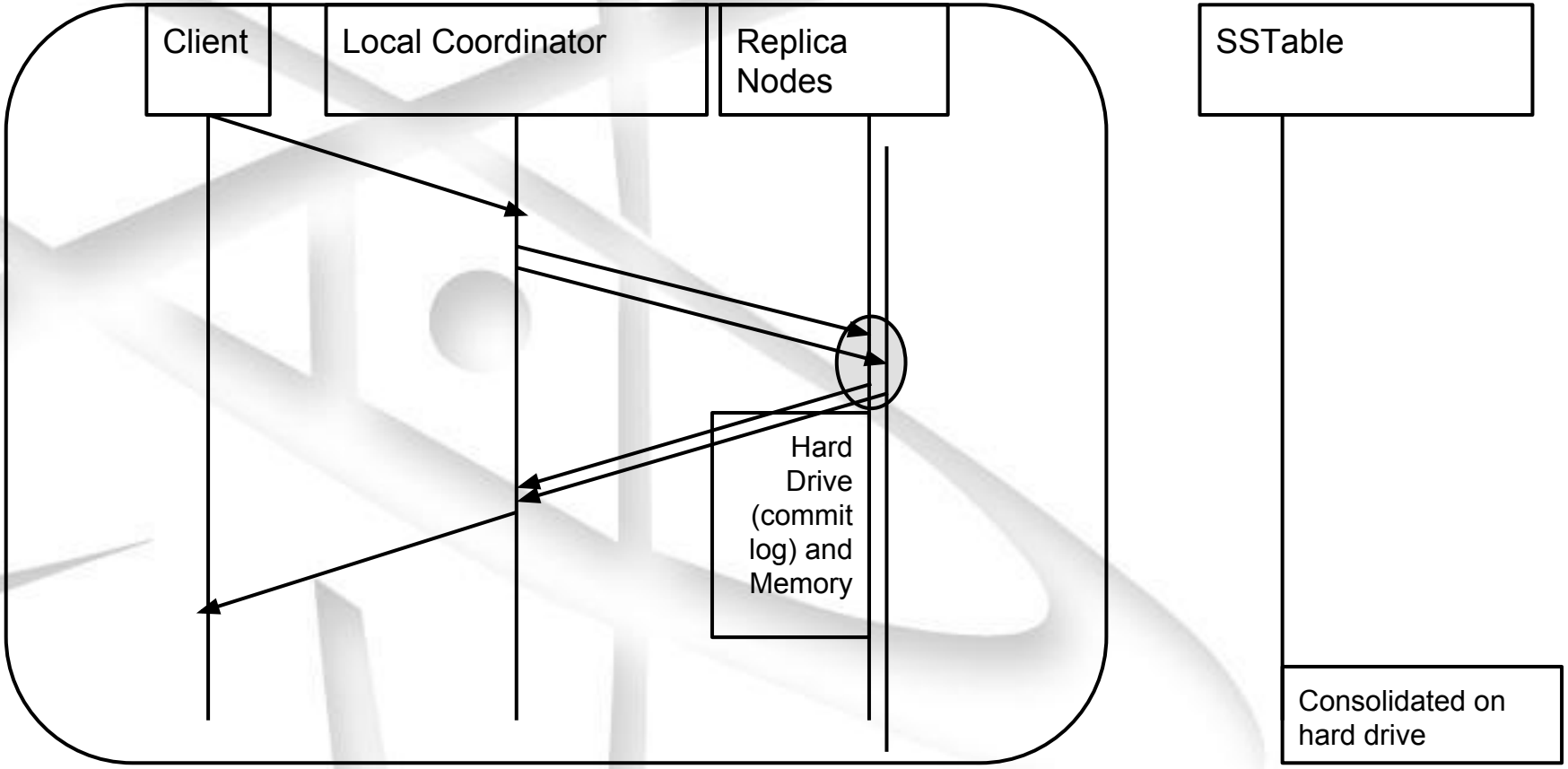
- Cooper et al. [1]
  - Initial Presentation YCSB
  - Surveys Many Different Databases Optimally Tuned
- Abramova et al.
  - Same Database: Cassandra
  - Expands on Configurations
  - Expands on Workload
- Waddington and Lin
  - Specific Workload for IoT
  - Specific, Custom Database

Background

- Cassandra and other databases
- Raspberry Pi and other Hardware
- Yahoo! Cloud System Benchmark



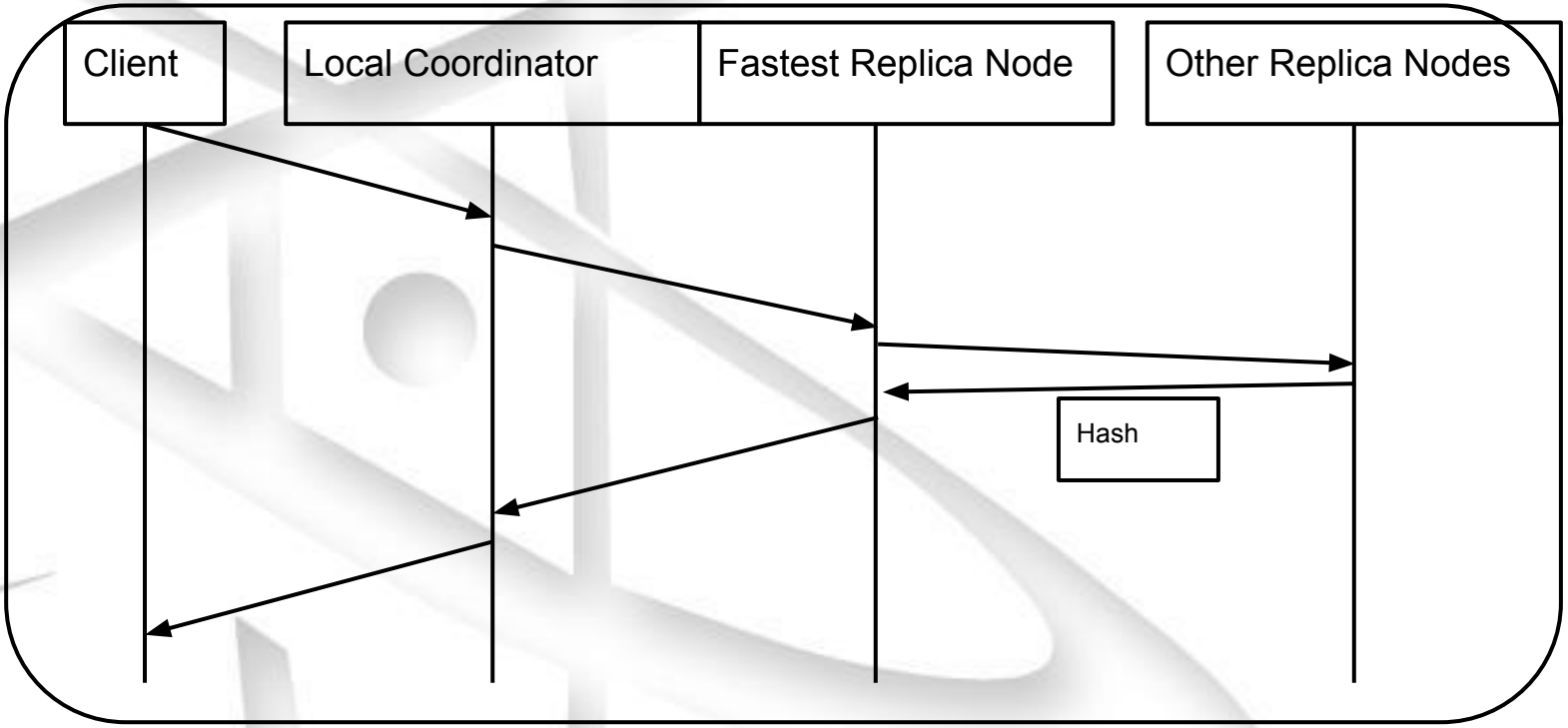
Write Path -- What Counts as a Write in Stress Testing



Adapted from https://wiki.apache.org/cassandra/WritePathForUsers

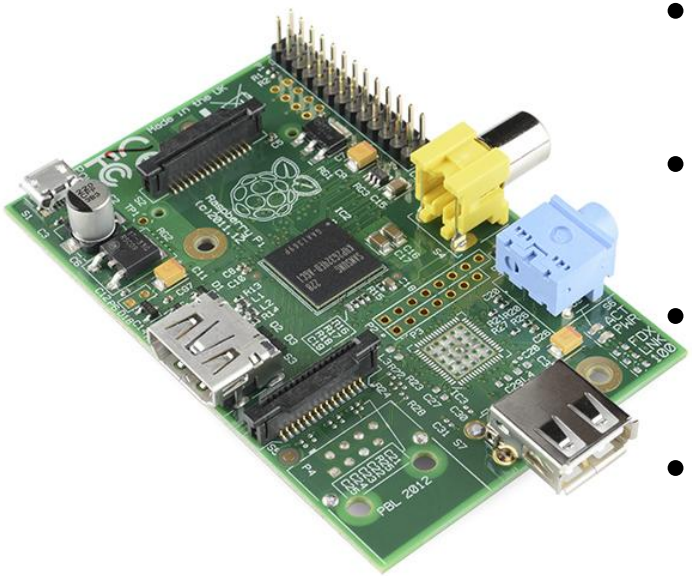
Background

Read Path - Nominal



Adapted from http://wiki.apache.org/cassandra/ReadPathForUsers

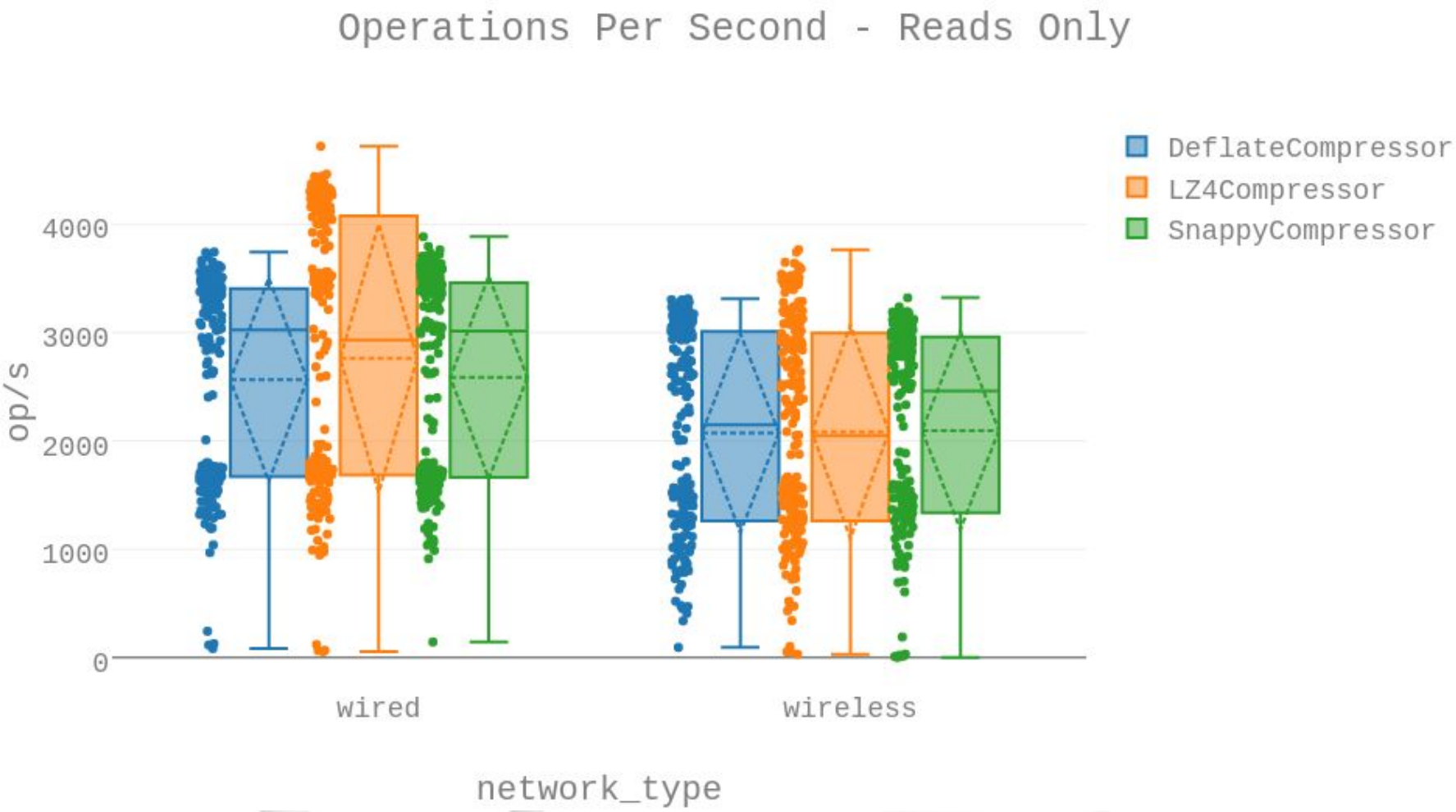
Raspberry Pi Series



- Multiple Models: 0, 1, 2, 3; A, B, +
- ARM Processor
- 1 GB RAM
- Designed for education
- Website https://www.raspberrypi.org/help/

- Example image courtesy of https://en.wikipedia.org/wiki/Raspberry_Pi

cassandra-stress

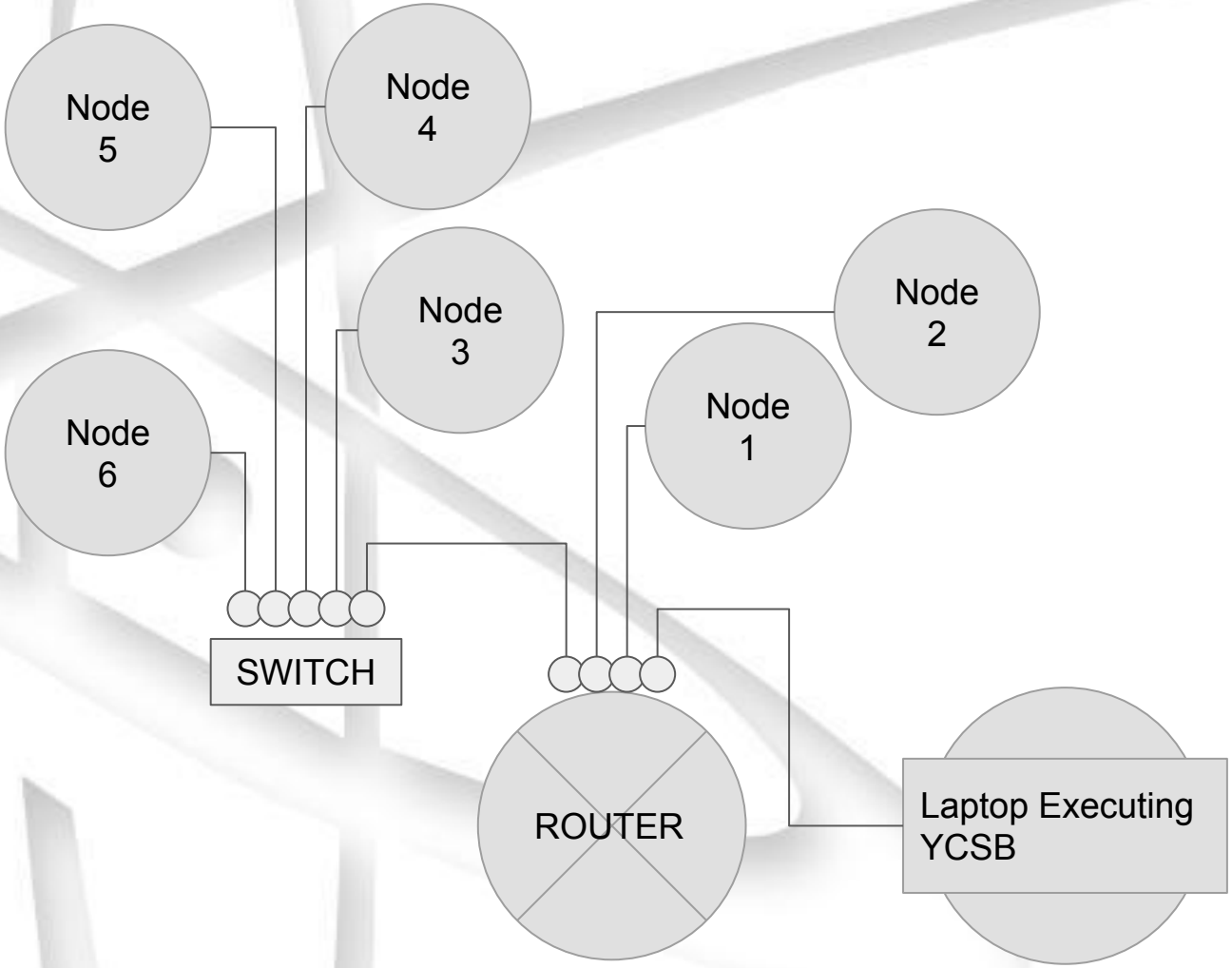


Experimental Setup

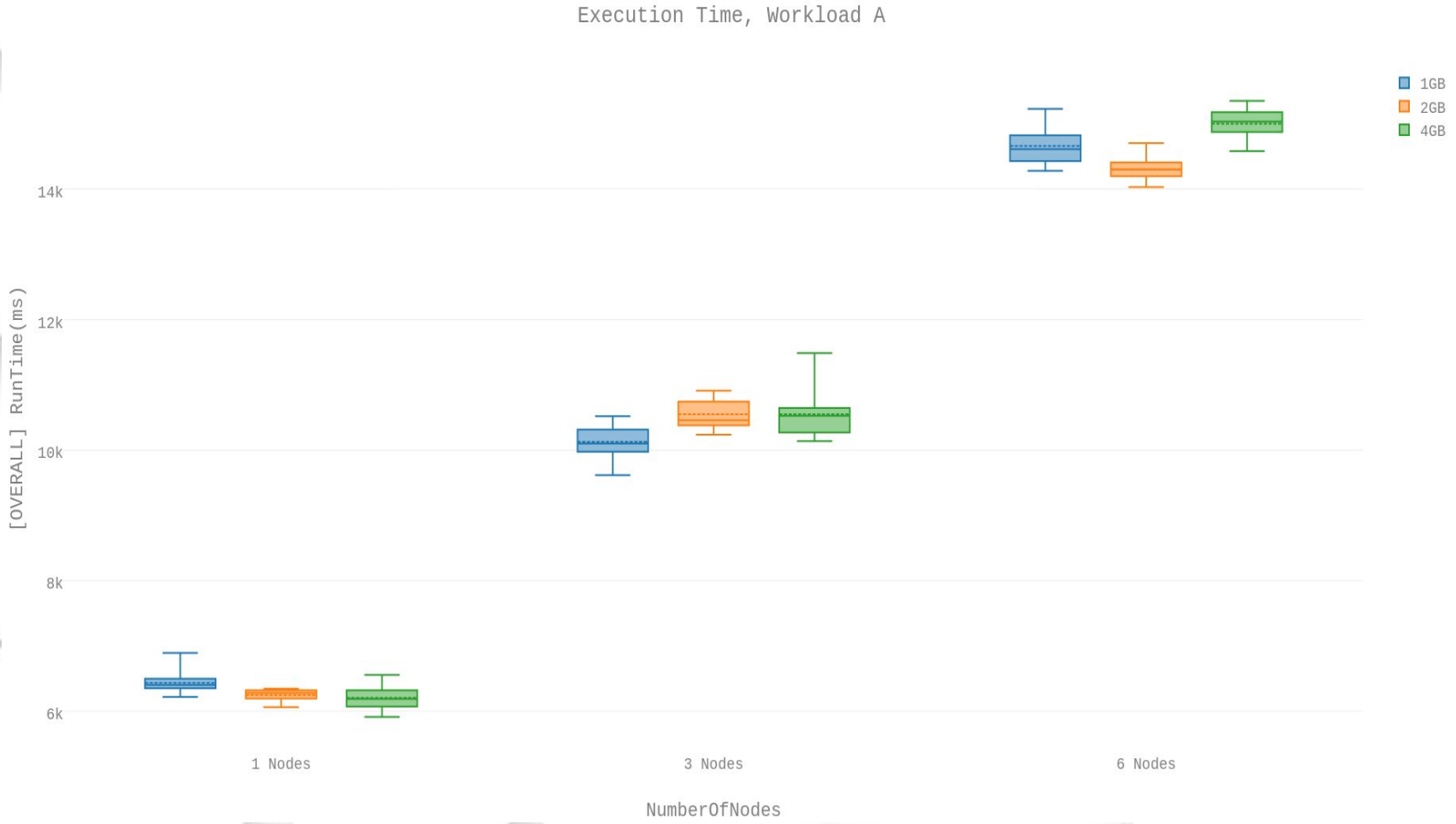
Workload	Read	Update	Scan	Insert
A	0.50	0.50	0.00	0.00
C	1.00	0.00	0.00	0.00
E	0.00	0.00	0.95	0.05
I	0.01	0.00	0.00	0.99

Communication	Platform	RAM
Nodal	Virtual Machine	1 GB
Nodal	Virtual Machine	2 GB
Nodal	Virtual Machine	4 GB
Ethernet LAN	Raspberry Pi	1 GB
802.11 LAN	Raspberry Pi	1 GB

Experimental Setup (cont)



Experimental Results

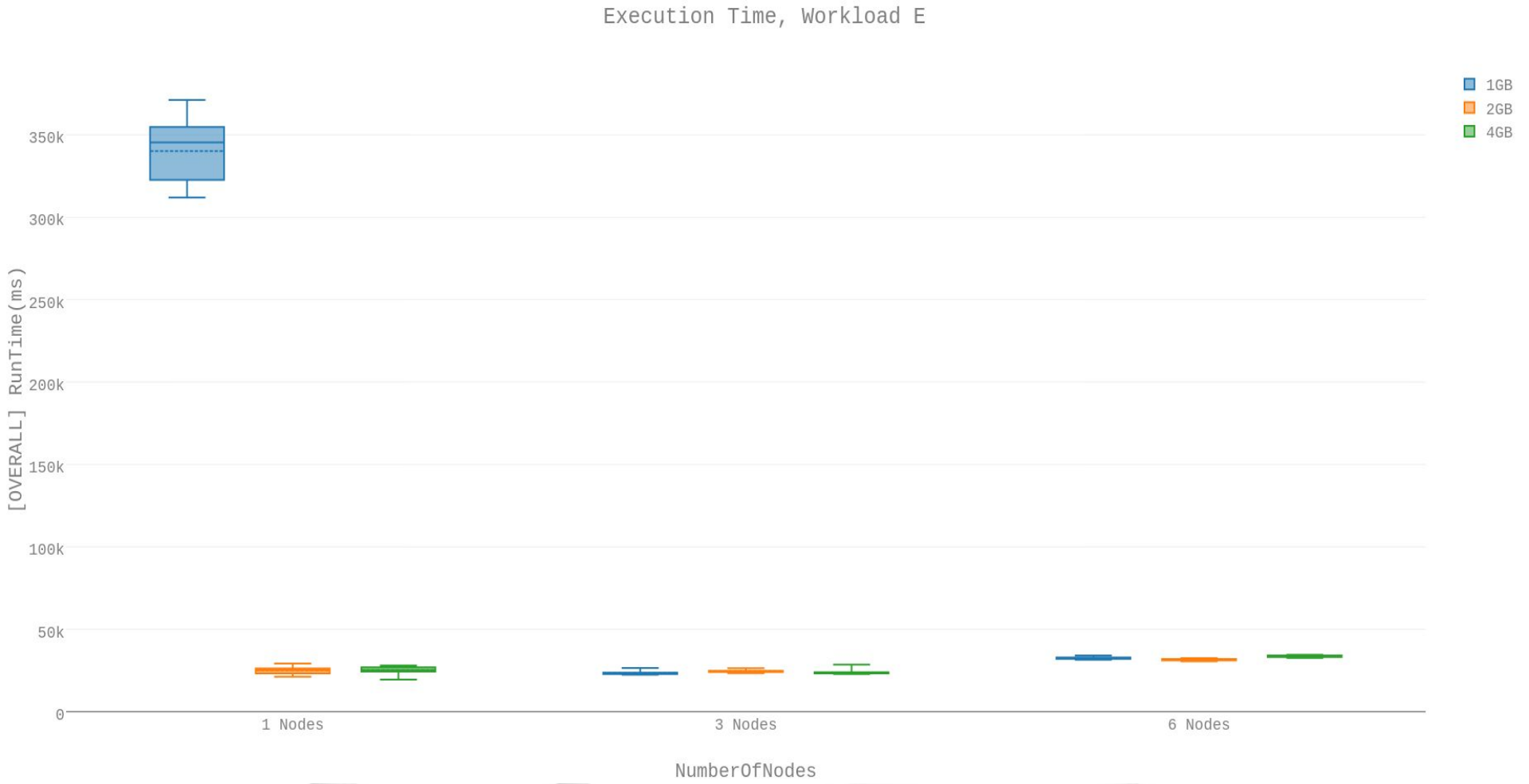


Linear Regression (A)

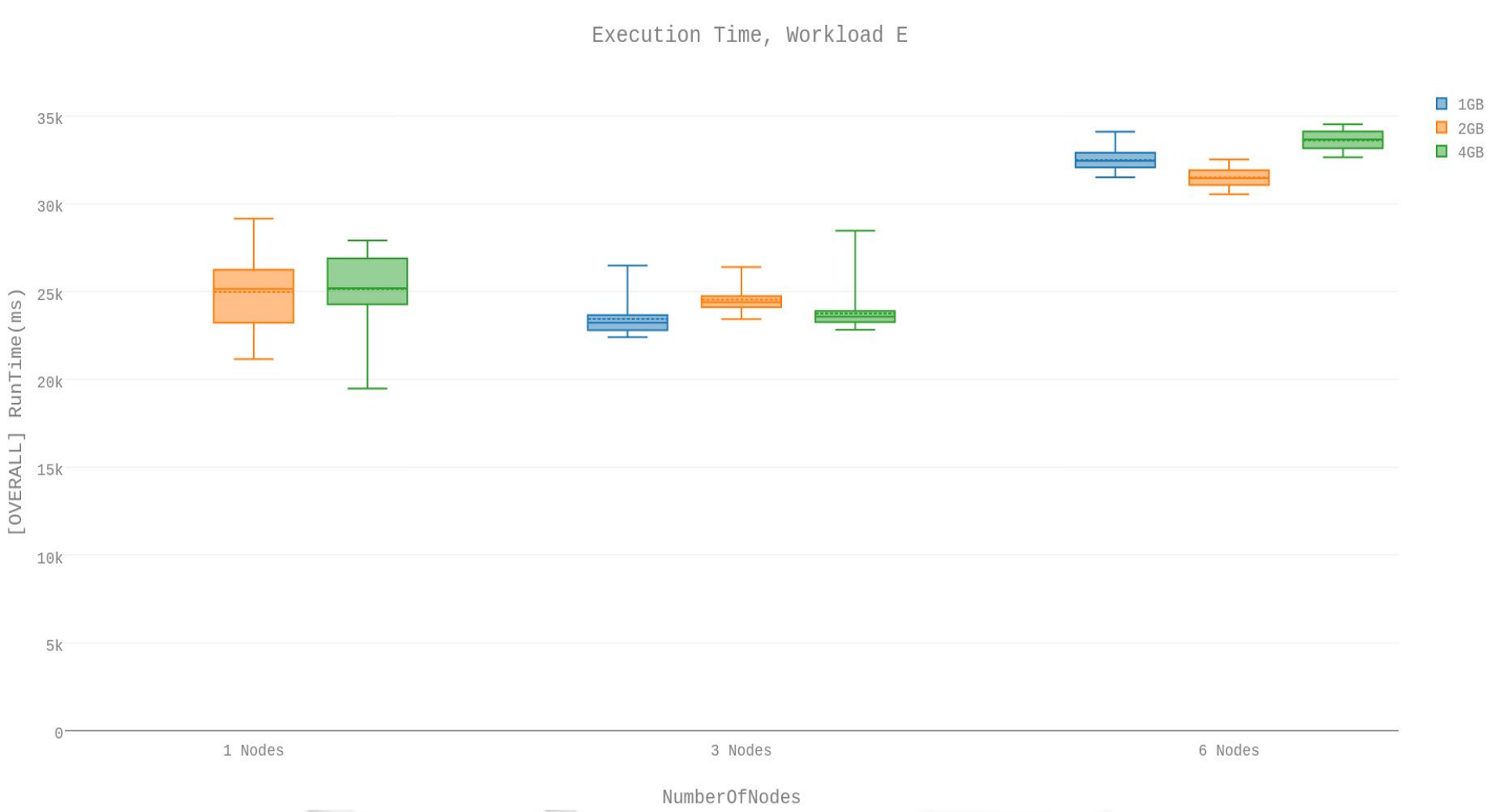
Cluster Size (nodes)	Slope (ms per GB RAM)	Intercept (ms)	r-Value	p-Value	Standard Error
1	-68.8	6.45e+03	-0.509	2.08e-05	14.9
3	120	1.01e+04	0.458	0.000162	29.9
6	147	1.43e+04	0.514	1.67e-05	31.5

Experimental Results

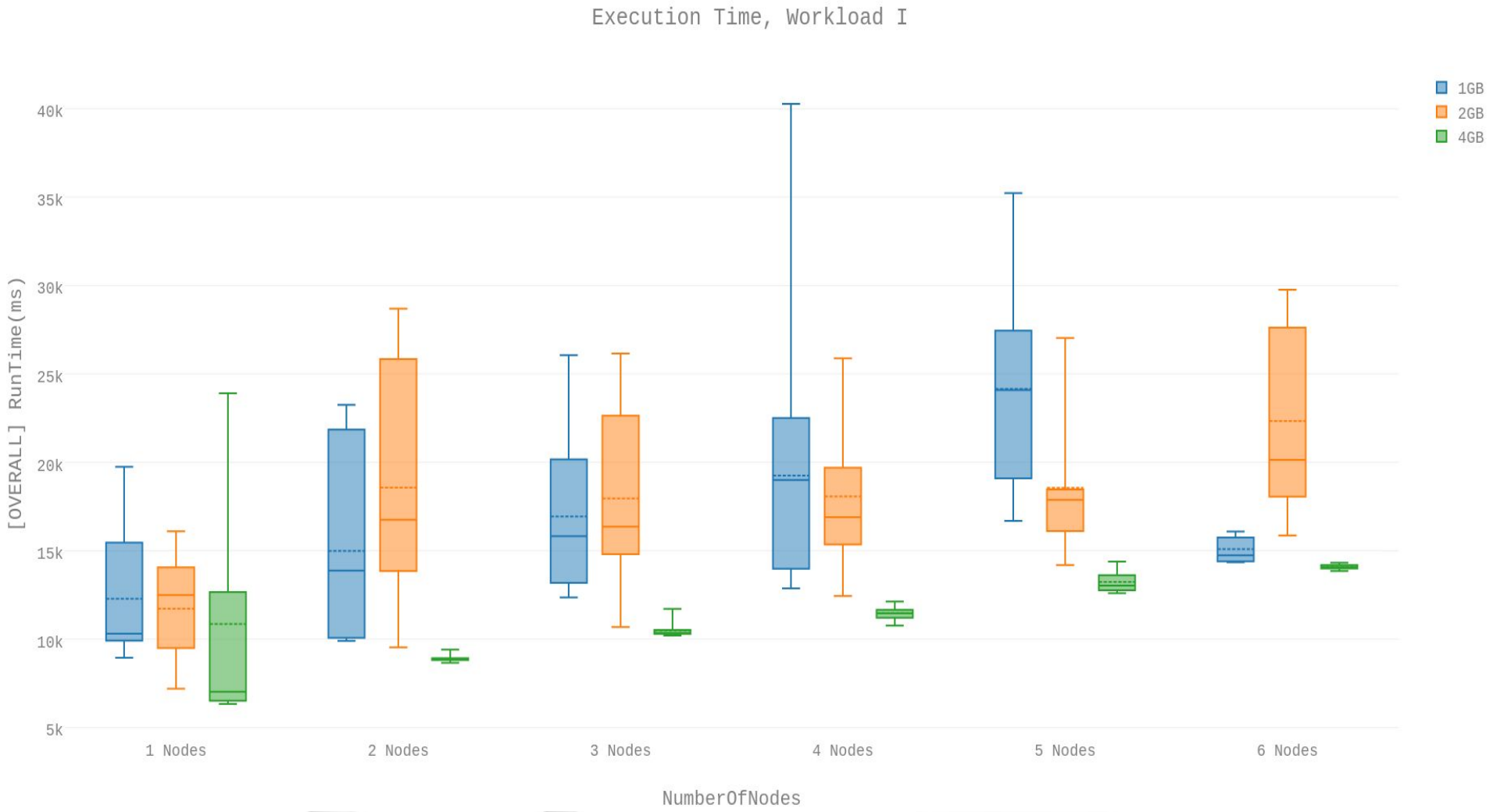
Experimental Results



Experimental Results



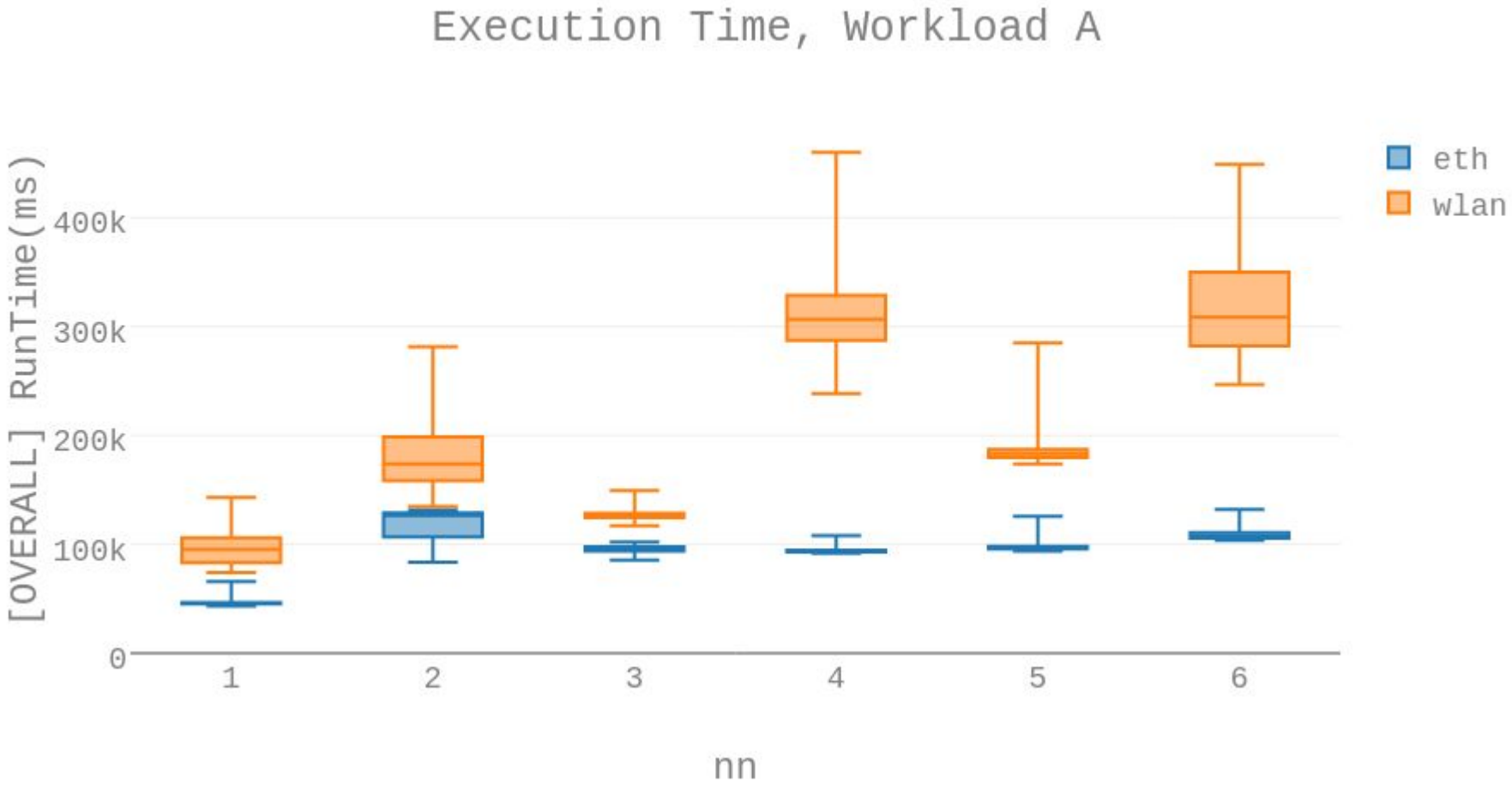
Experimental Results



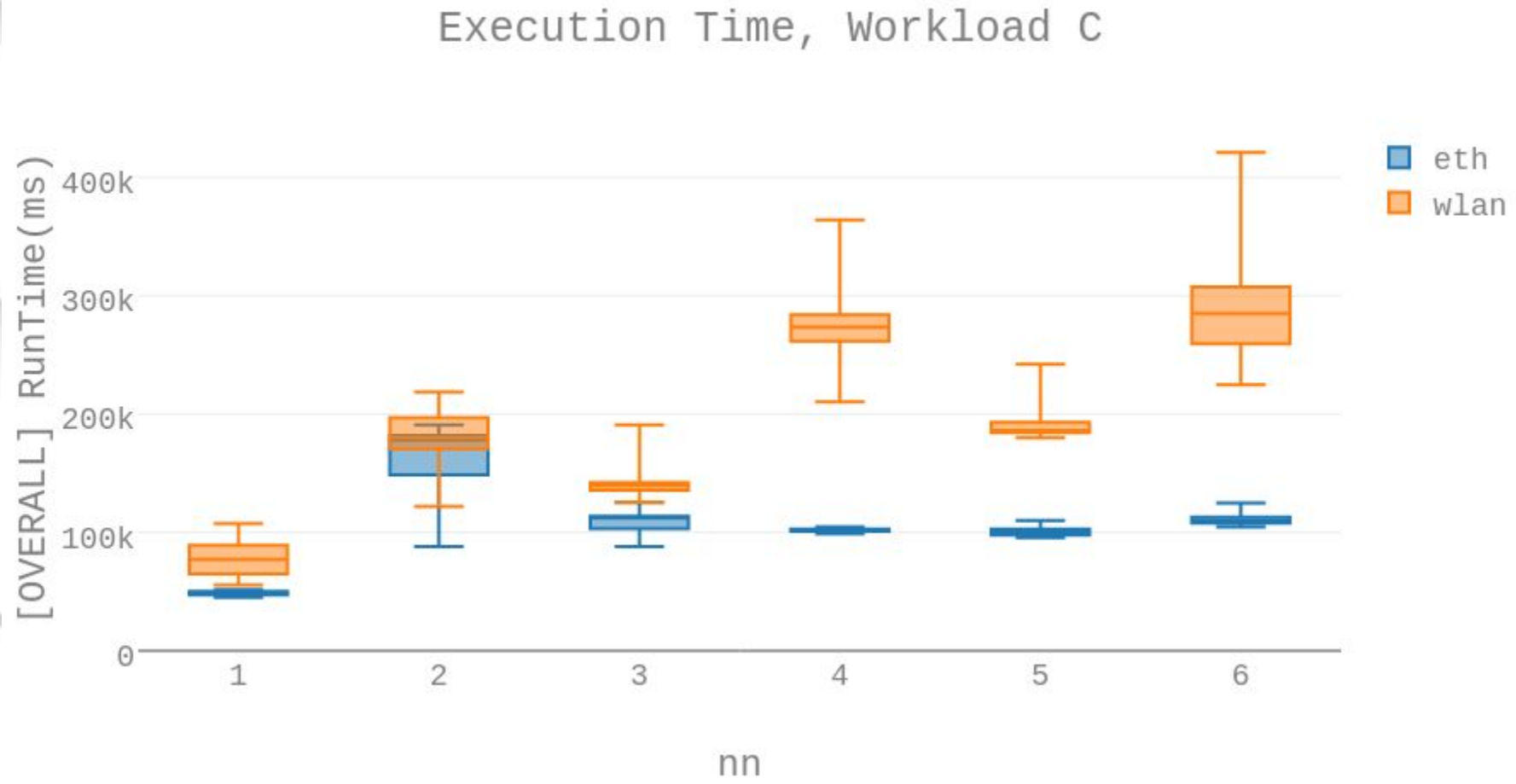
Experimental Results

- These results fail to suggest a linear prediction of performance based on RAM.

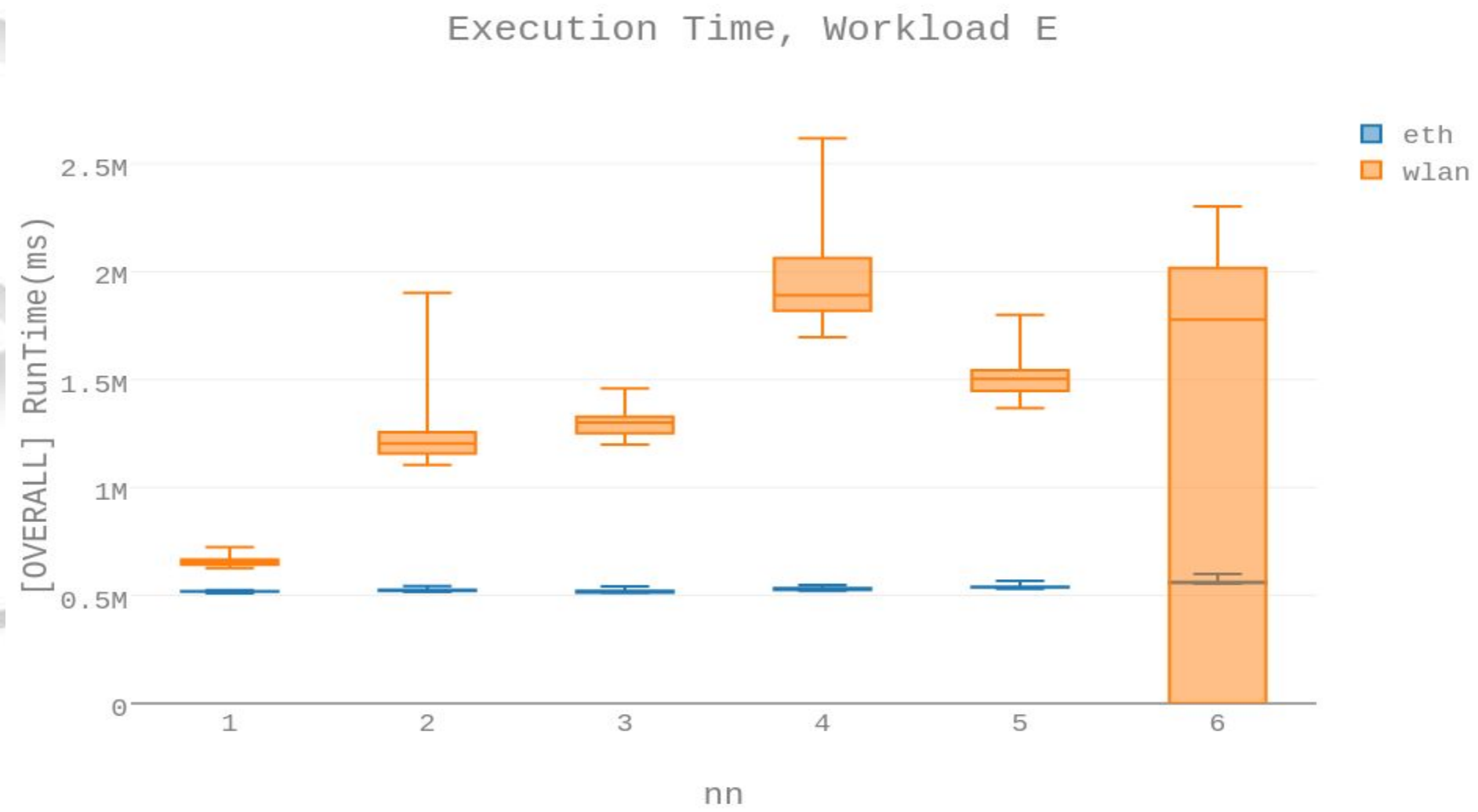
Experimental Results



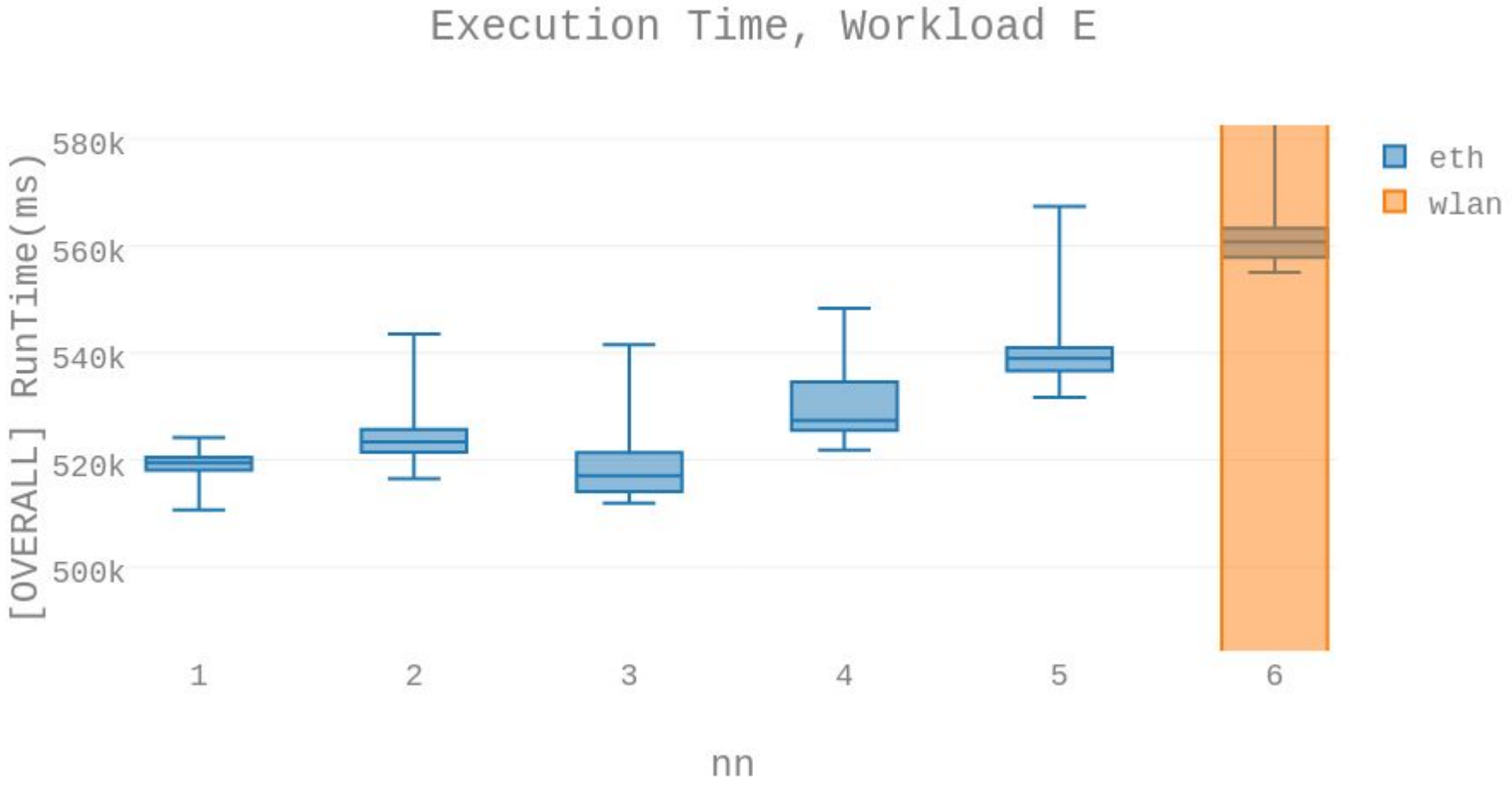
Experimental Results



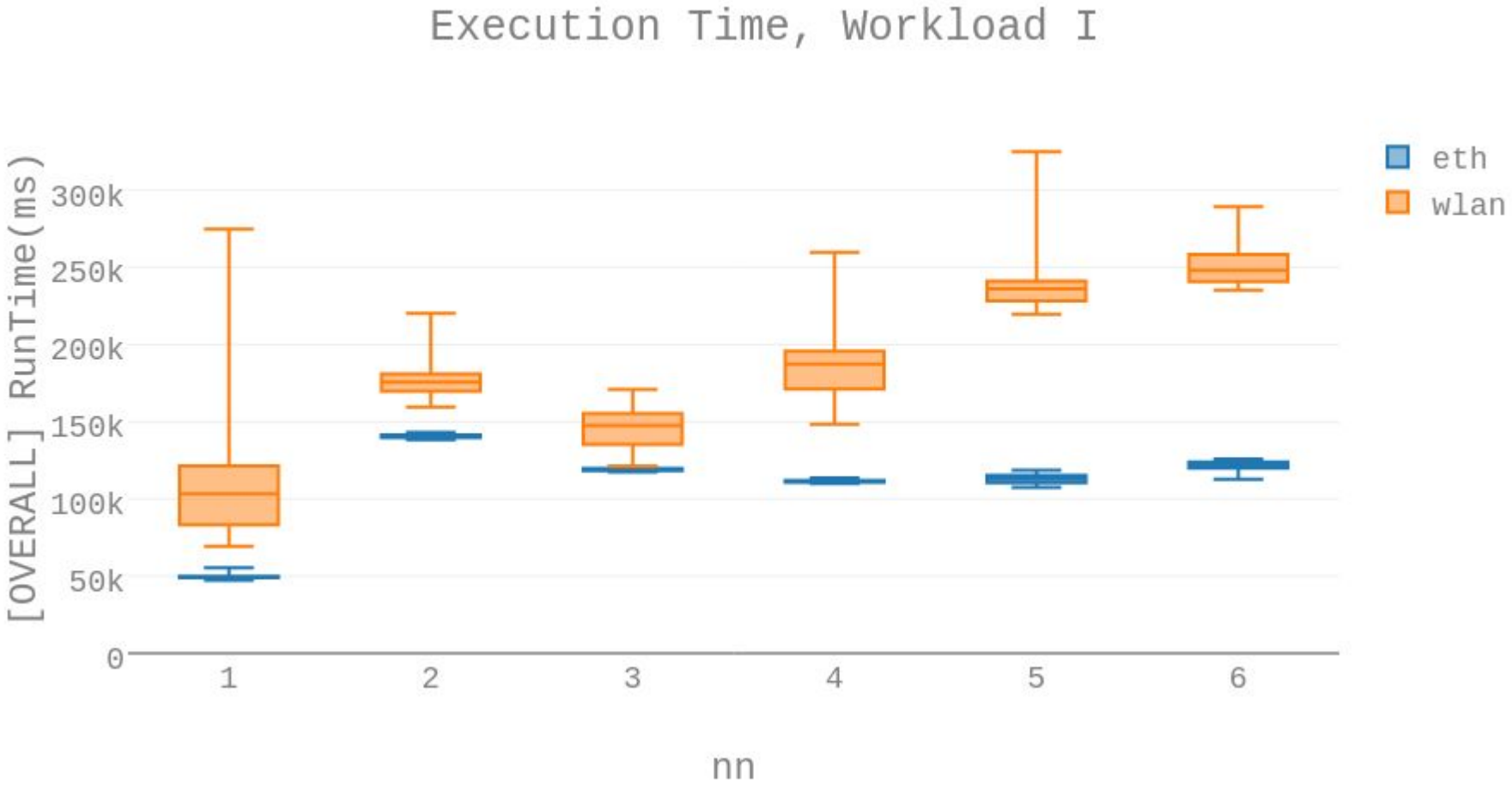
Experimental Results



Experimental Results



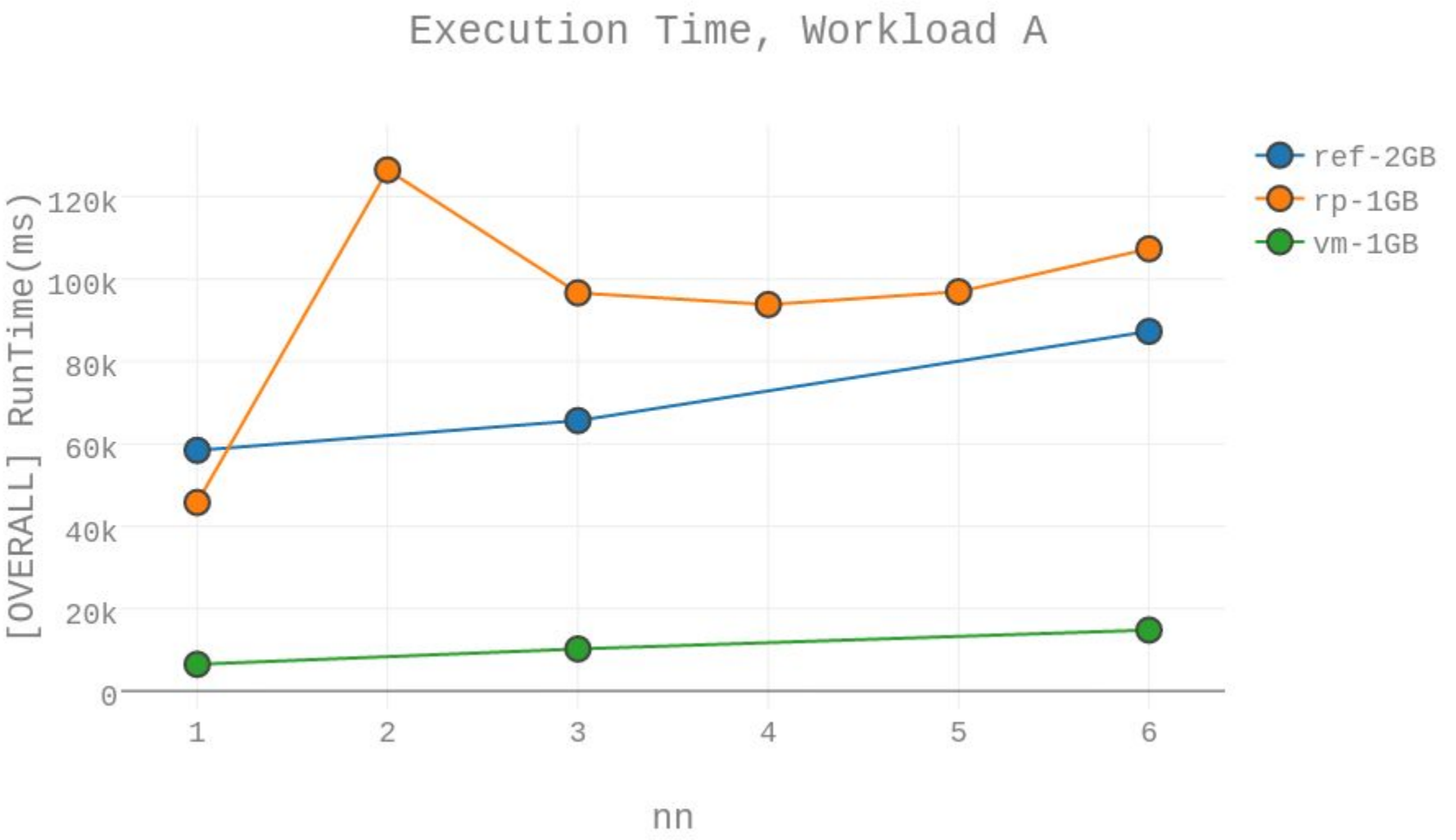
Experimental Results



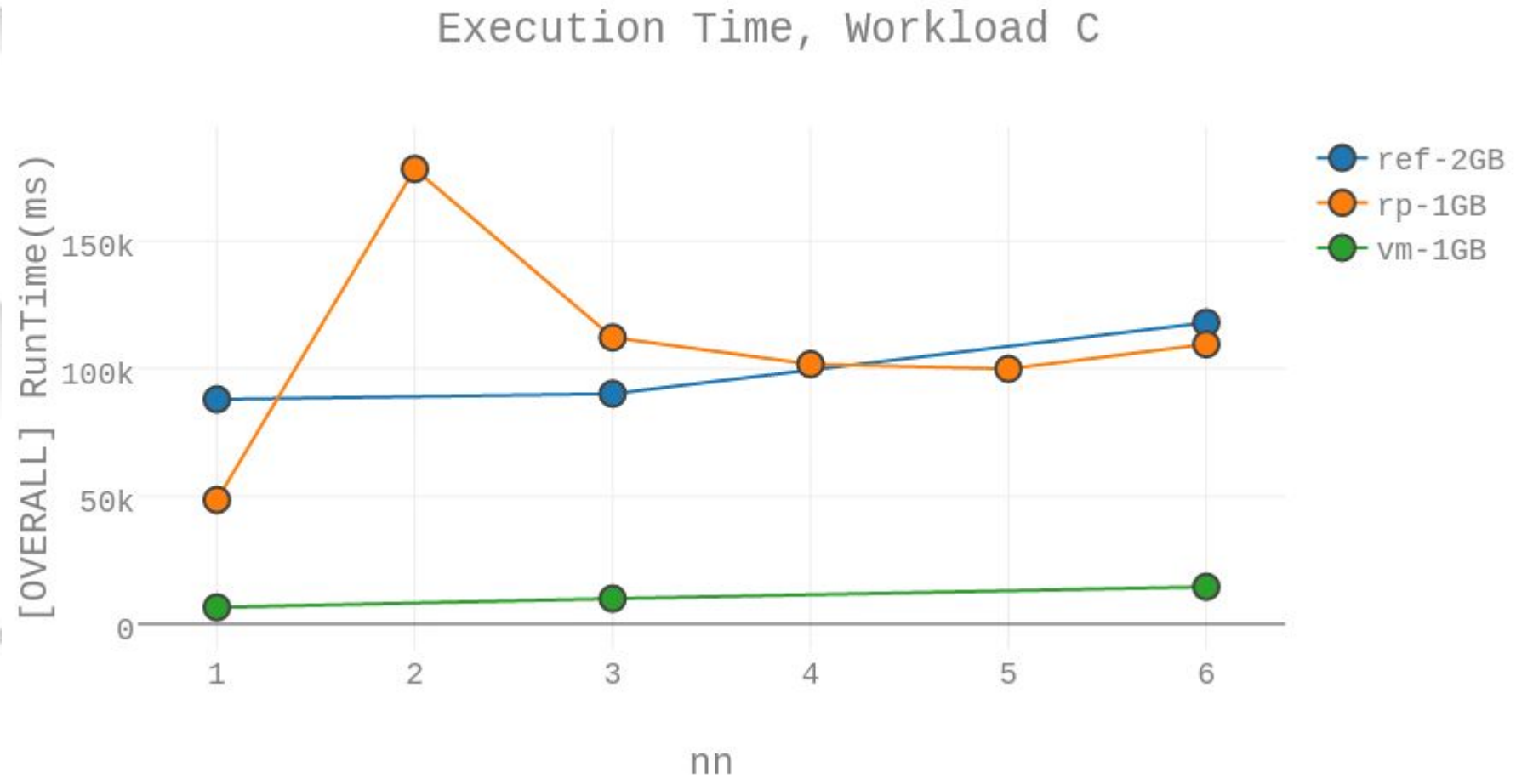
Experimental Results

- These results seem to suggest Cassandra's scalability may a bit threatened by wireless. Additional testing may be of value to see if this trend continues or if something could ameliorate it, such as employing the request-to-send.

Experimental Results



Experimental Results



Experimental Results

Experimental Results

Absolute Differentials

Number of Nodes	1-Node Cluster	3-Node Cluster	6-Node Cluster	OVERALL
Count	21	21	21	63
Mean (ms)	1.28e+04	3.16e+04	1.95e+04	2.13e+04
Standard Deviation (ms)	1.12e+03	1.65e+03	1.78e+03	7.96e+03
Minimum (ms)	1.03e+04	2.81e+04	1.64e+04	1.03e+04
25% (ms)	1.23e+04	3.07e+04	1.83e+04	1.35e+04
Median (ms)	1.28e+04	3.14e+04	1.94e+04	1.94e+04
75% (ms)	1.35e+04	3.22e+04	2.02e+04	3.06e+04
Maximum (ms)	1.52e+04	3.49e+04	2.37e+04	3.49e+04

Conclusions

- Available RAM
  - Results and interpretation fail to suggest any utilitarian linear model of performance.
- Workload
  - Results suggest that workload can make a difference in performance, and such differences were accentuated with hardware changes.
- Scalability
  - Results suggest reasonable scalability of wired and virtual clusters over cluster size... wireless less so and results suggest the utility of additional experimentation.
- Raspberry Pis versus Virtual Machines
  - Results suggest confirmation of a cost in execution time.
  - Results also suggest more experimentation could lead to a correction factor for simulated applications.

Future Work

- Varying Database Size
- Wireless Configurations (request-to-send, maximum transmission unit)
- Varying Hardware, not just the Raspberry Pi
- Testing Larger Clusters
- Varying Thread Count

Future Work

- WiFiPi Prototype
  - Sniff traffic with scapy
  - Filter out probe requests
  - Extract SSIDs (plaintext), IP Addresses
  - Append to Distributed Database

Sources

[1] Cooper, B. F., Silberstein, A., Tam, E., Ramakrishnan, R., & Sears, R. (2010). *Benchmarking cloud serving systems with YCSB. Proceedings of the 1st ACM Symposium on Cloud Computing - SoCC '10, 143–154. http://doi.org/10.1145/1807128.1807152*

[2] Abramova, V., Bernardino, J., & Furtado, P. (2014). *Testing Cloud Benchmark Scalability with Cassandra. 2014 IEEE World Congress on Services, 434–441. http://doi.org/10.1109/SERVICES.2014.81*

[3] Waddington, D. G., & Lin, C. (2016). *A Fast Lightweight Time-Series Store for IoT Data.*

[4] Lourenco, J. R., Abramova, V., Cabral, B., Bernardino, J., Carreiro, P., & Vieira, M. (2015). No SQL in Practice: A Write-Heavy Enterprise Application. *Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015*, 584–591. http://doi.org/10.1109/BigDataCongress.2015.90

[5] Abramova, V., & Bernardino, J. (2013). NoSQL databases: MongoDB vs cassandra. *Proceedings of the International C* Conference on Computer Science and Software Engineering, ACM 2013*, 14–22. http://doi.org/10.1145/2494444.2494447

End