devicepros

HW Acceleration of Memcached

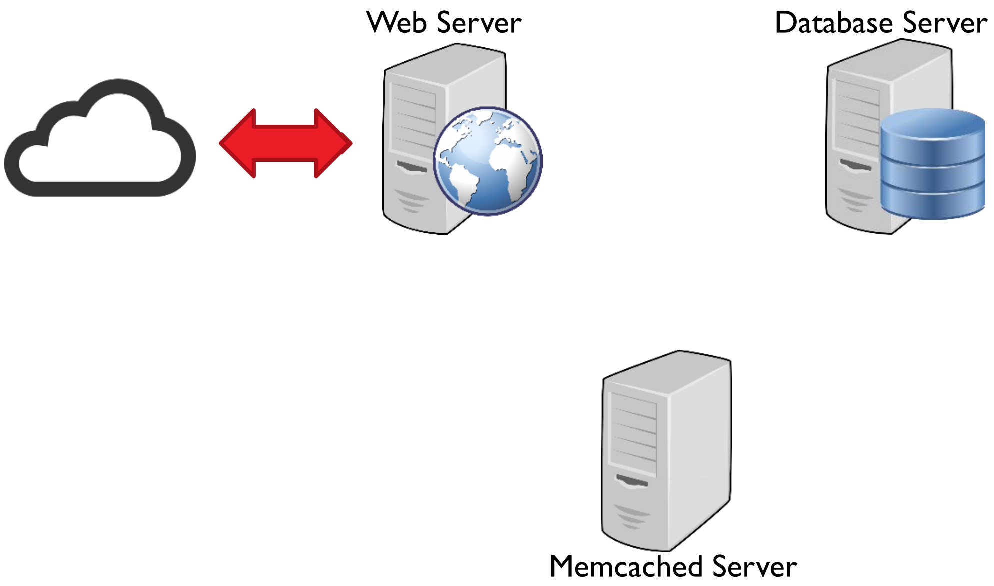
Aug 5, 2014

Derrill Sturgeon
CEO, devicepros
derrill.sturgeon@devicepros.net
(408) 504-5414

Agenda

- Quick introduction to Memcached
- Scaling Memcached as a software solution
- Performance and some use cases
- A Memcached server on FPGA

How Memcached is used

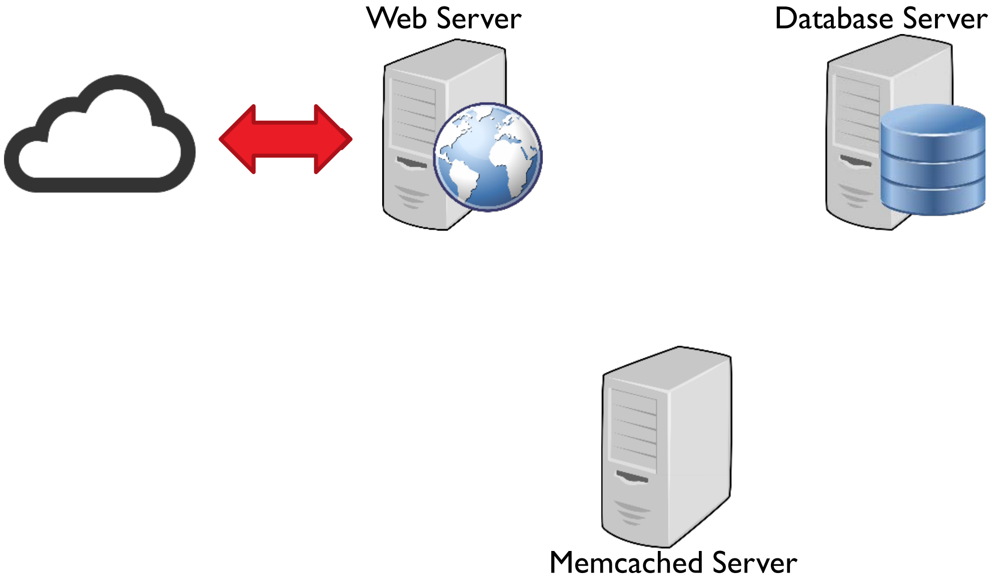


Web Server

Database Server

Memcached Server

How Memcached is used



Web Server

Database Server

Memcached Server

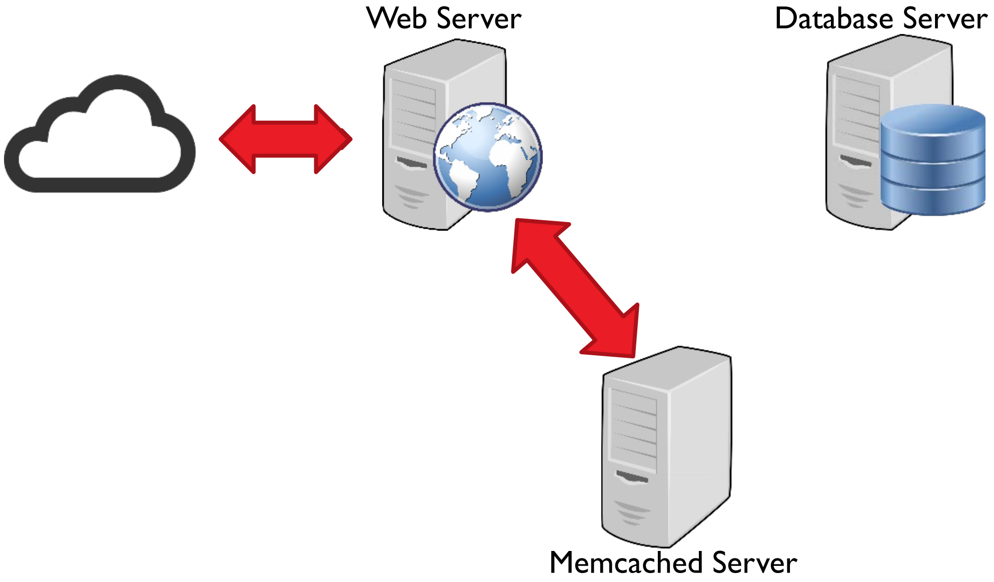
```
function get_foo(foo_id)

    foo = memcached_get("foo:" . foo_id)
    return foo if defined foo

    foo = fetch_foo_from_database(foo_id)
    memcached_set("foo:" . foo_id, foo)
    return foo

end
```

How Memcached is used



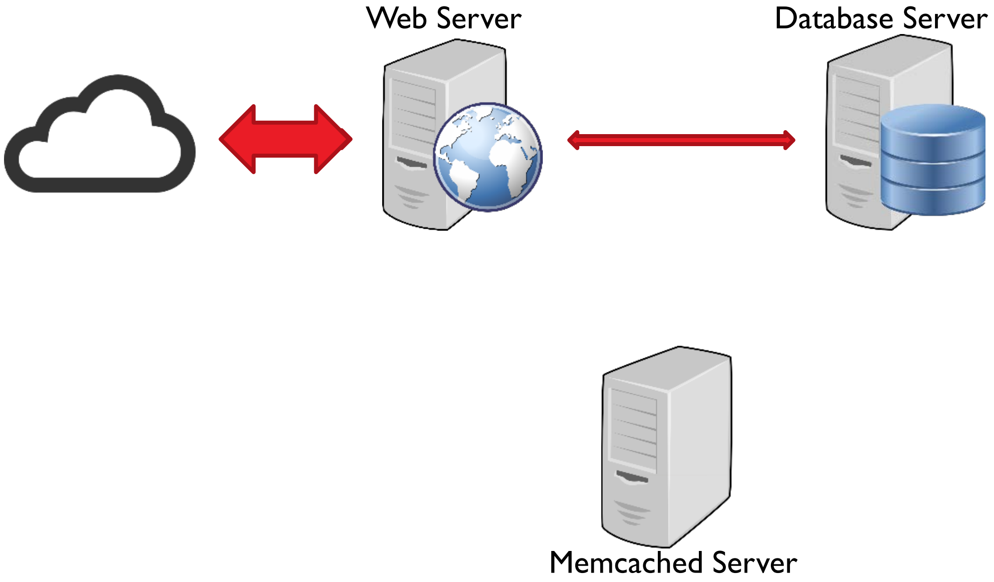
```
function get_foo(foo_id)

    foo = memcached_get("foo:" . foo_id)
    return foo if defined foo

    foo = fetch_foo_from_database(foo_id)
    memcached_set("foo:" . foo_id, foo)
    return foo

end
```

How Memcached is used



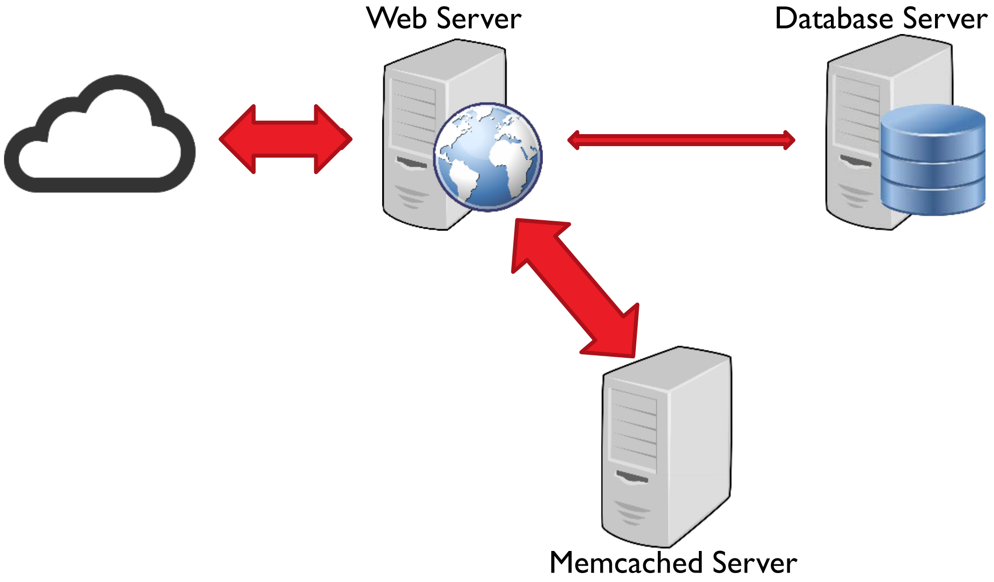
```
function get_foo(foo_id)

    foo = memcached_get("foo:" . foo_id)
    return foo if defined foo

    foo = fetch_foo_from_database(foo_id)
    memcached_set("foo:" . foo_id, foo)
    return foo

end
```

How Memcached is used



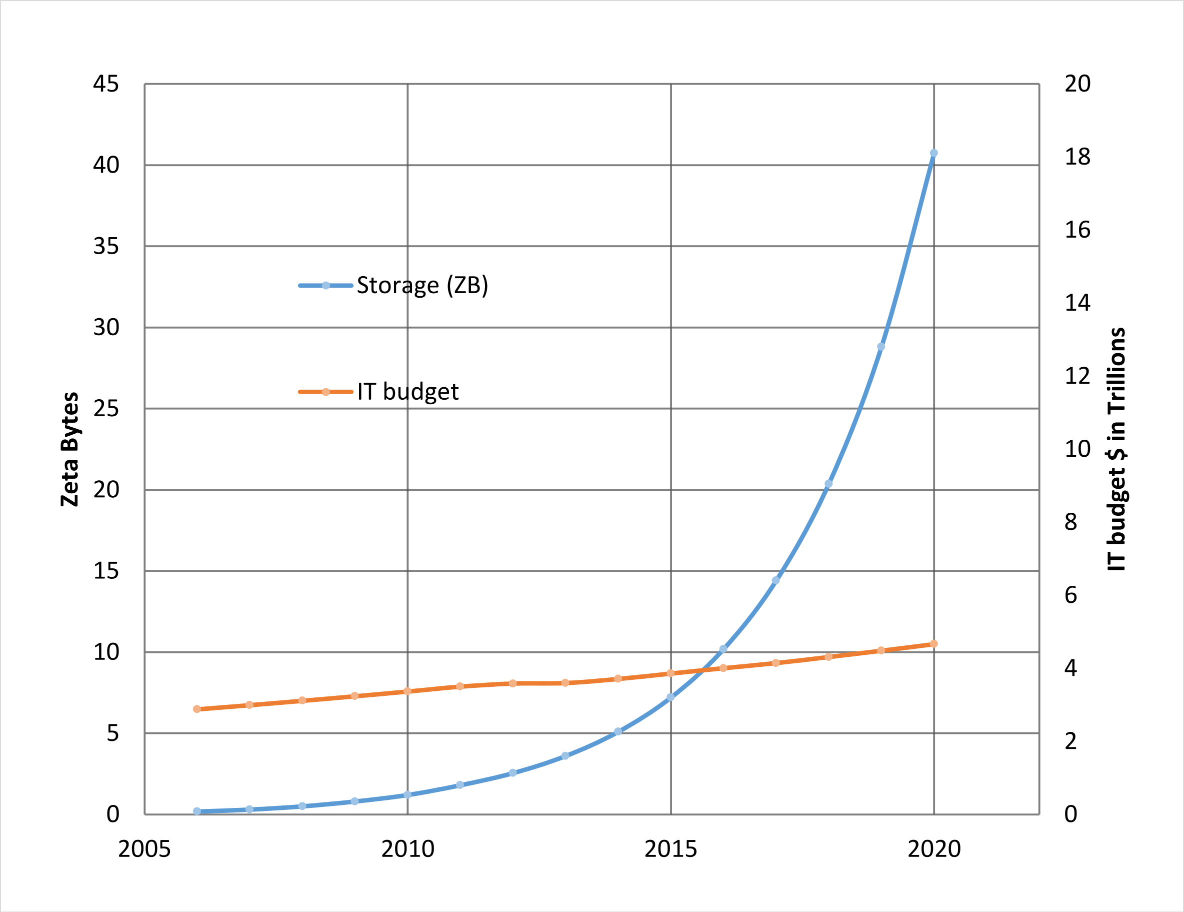
```
function get_foo(foo_id)

    foo = memcached_get("foo:" . foo_id)
    return foo if defined foo

    foo = fetch_foo_from_database(foo_id)
    memcached_set("foo:" . foo_id, foo)
    return foo

end
```

Universe of Data growth vs. IT budgets



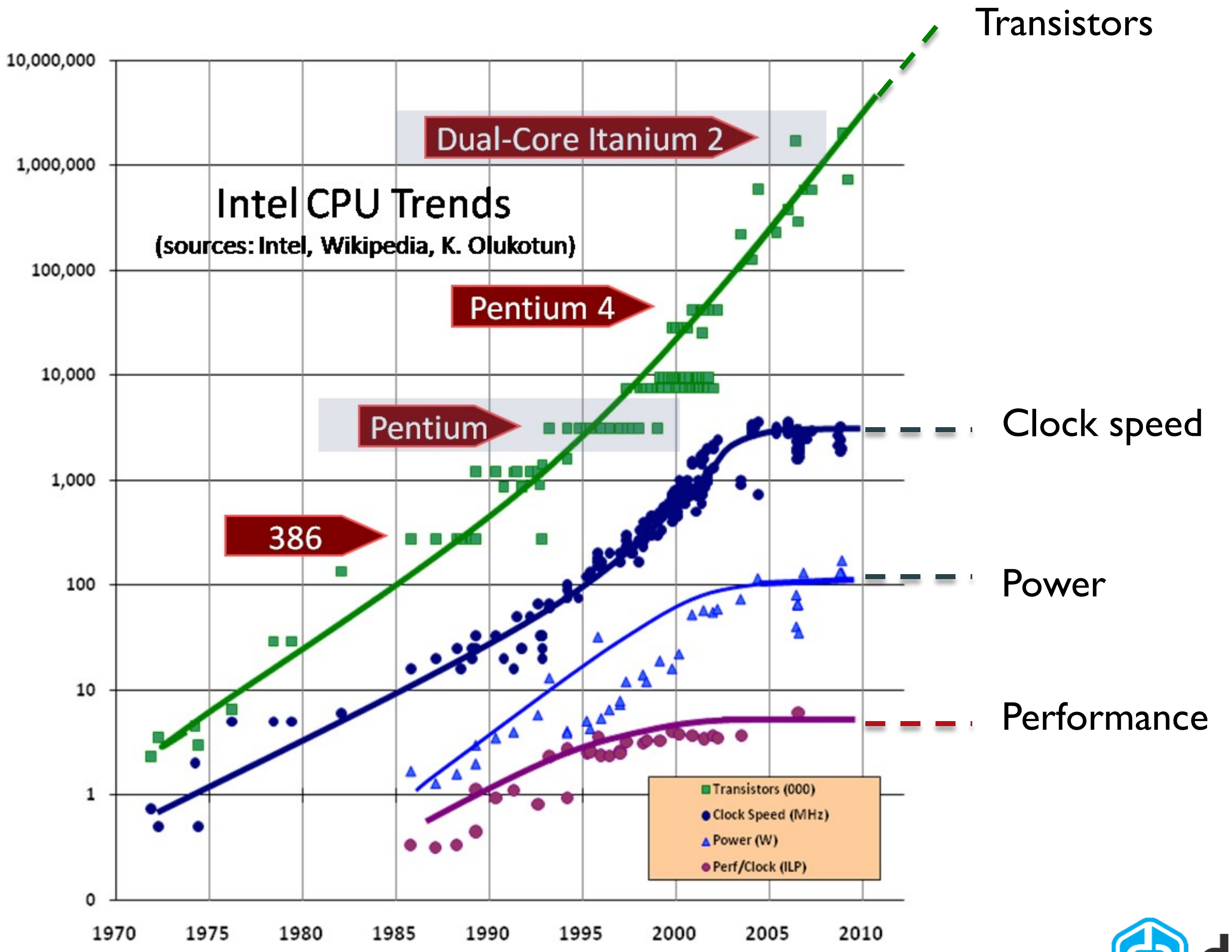
IDC Digital Universe Study

Growth by 2020

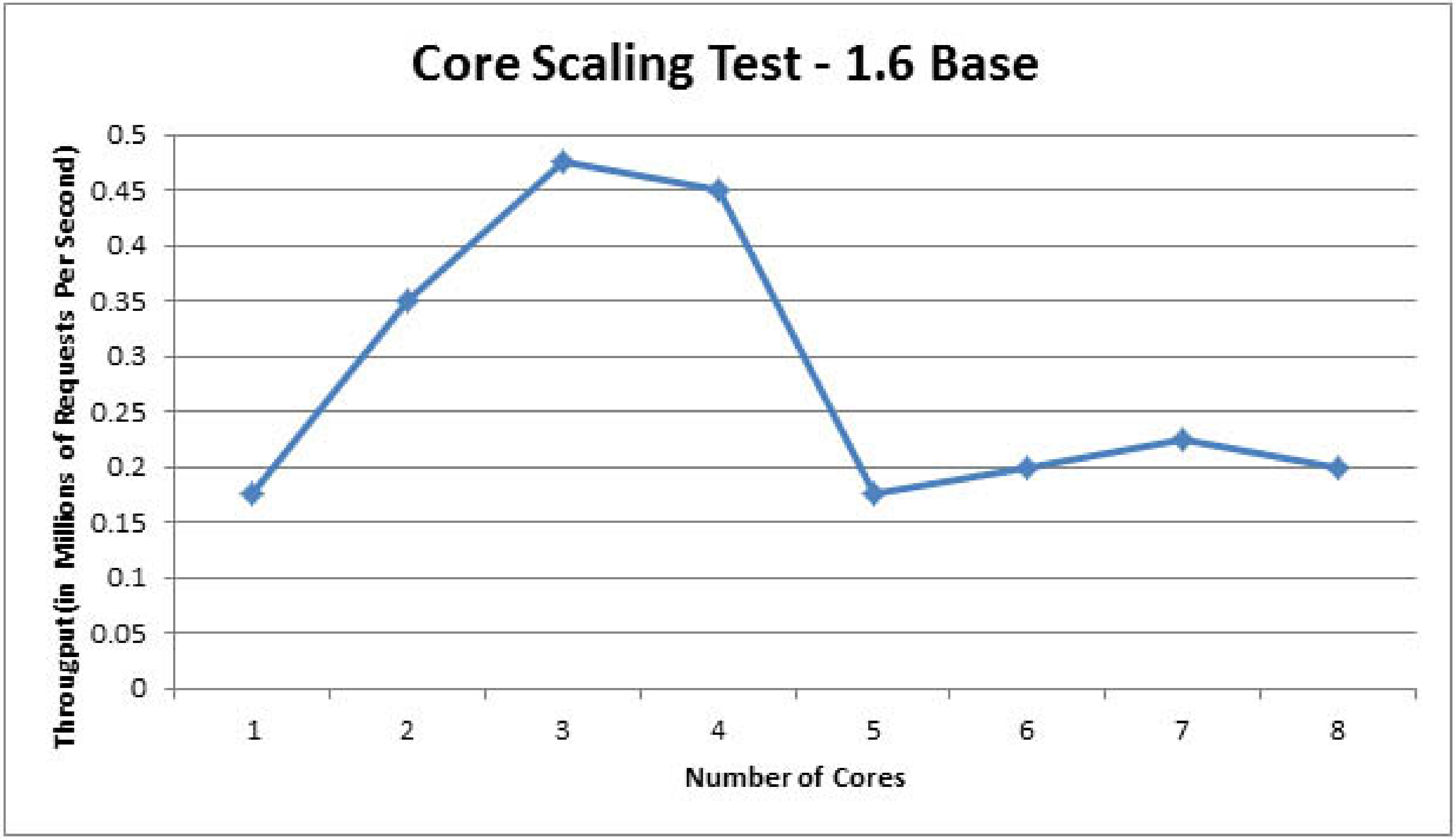
- 14X Enterprise data
- 10X Servers
- vs. 1.5X IT professionals

Memcached accounts for 10-30% of servers

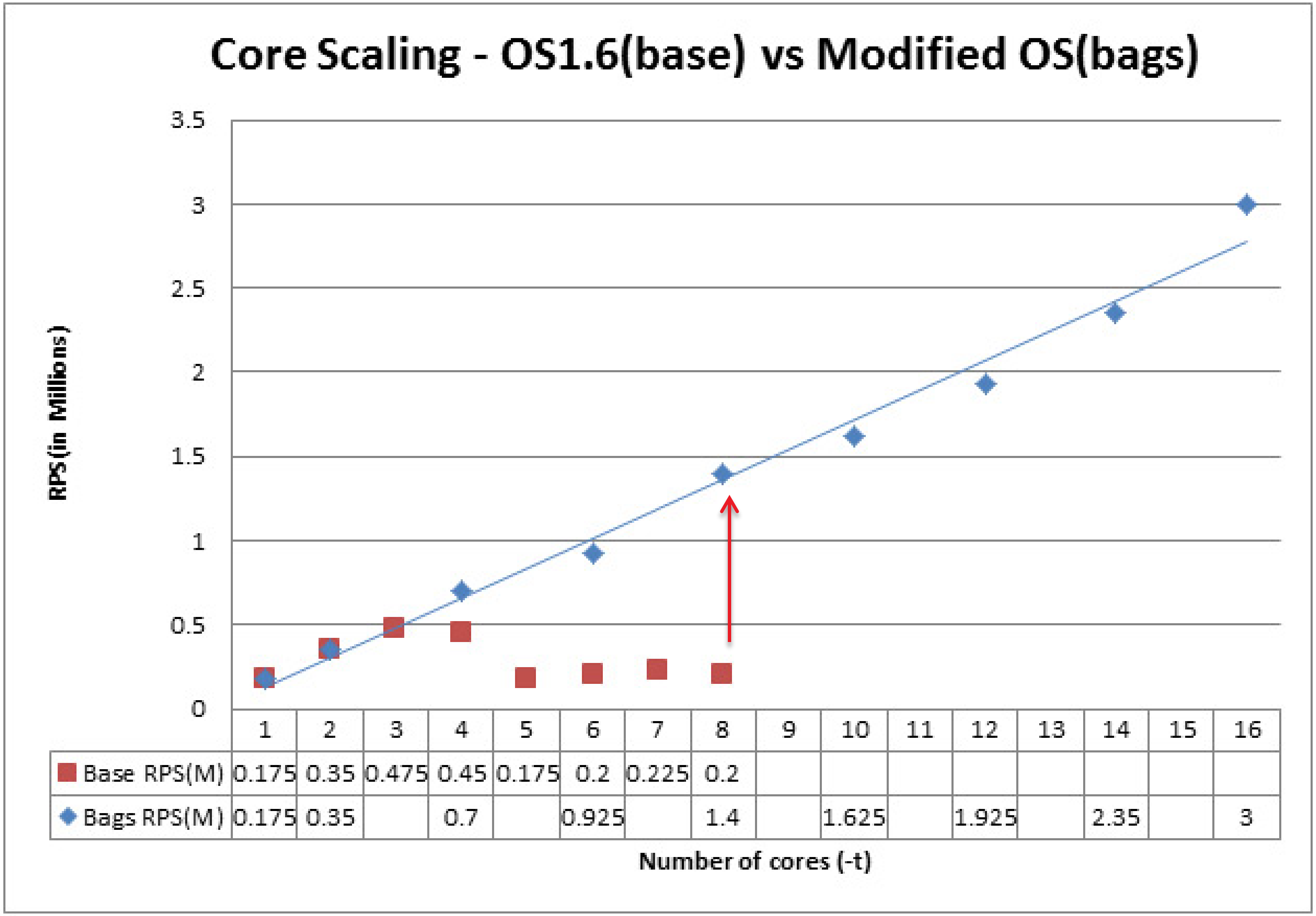
Intel CPU Trends



Intel 2012 Paper on Memcached scalability



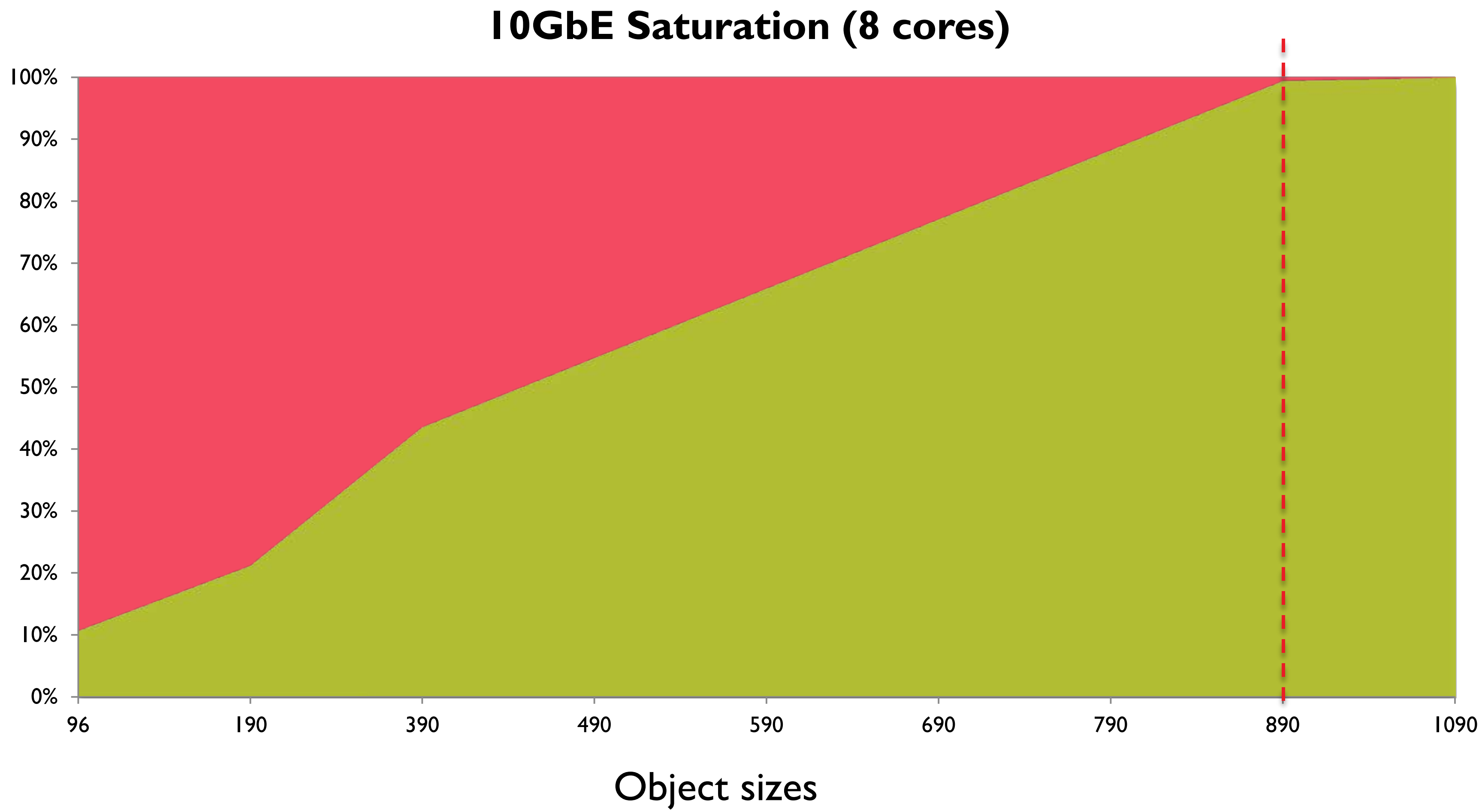
Intel 2012 Paper on Memcached scalability



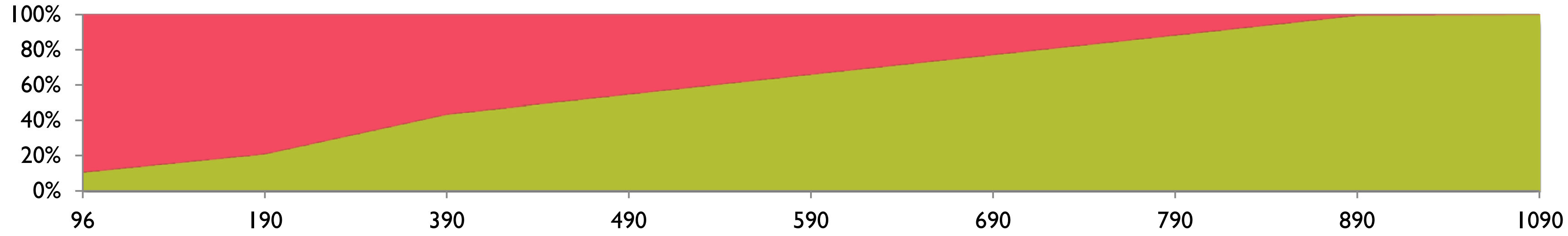
Core Scaling - OS1.6(base) vs Modified OS(bags)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
🟥 Base RPS(M)	0.175	0.35	0.475	0.45	0.175	0.2	0.225	0.2								
◆ Bags RPS(M)	0.175	0.35		0.7		0.925		1.4		1.625		1.925		2.35		3

Network not saturated for small Object sizes

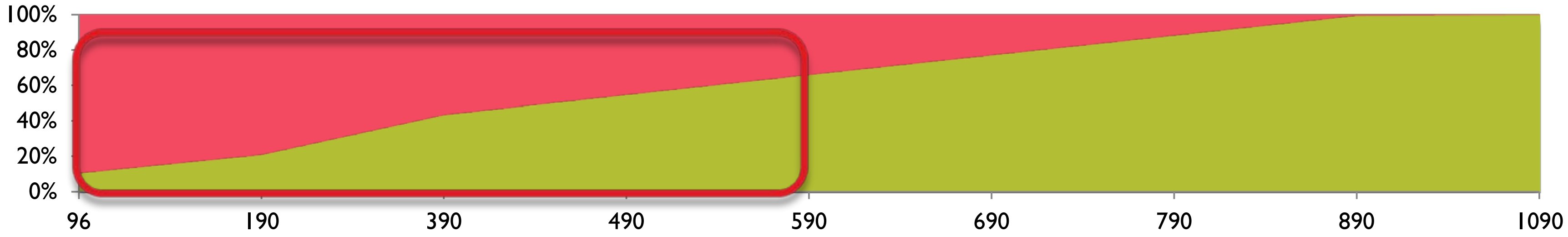


Network not saturated for small Object sizes


10GbE Saturation (8 cores)

Calculated probability of value sizes									
Value size [Bytes]	128	256	512	768	1014	2048	4096	22000	32000
Facebook: ETC	0.55	0.075	0.285	0.015	0.025	0.025	0.025	0	0
Facebook: USR	1	0	0	0	0	0	0	0	0
Facebook: APP	0.12	0	0.63	0.21	0.03	0.01	0	0	0
Facebook: VAR	0.78	0.02	0.17	0.03	0	0	0	0	0
Twitter	0	0	0	0.1	0.85	0.05	0	0	0
Wiki	0	0	0	0	0.58	0.02	0.1	0.25	0.05
Flicker	0	0	0	0	0	0	0	0.1	0.9
Youtube	0	0	0	0	0	0.75	0.11	0.11	

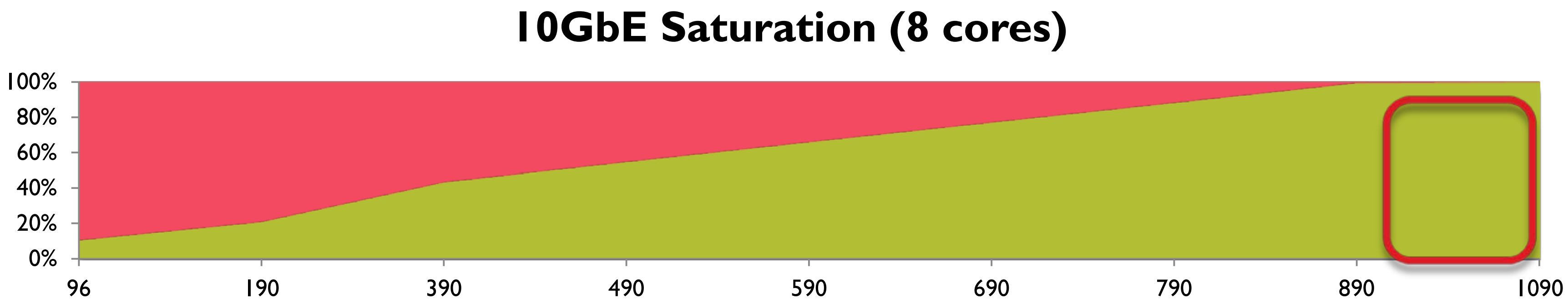
Network not saturated for small Object sizes



10GbE Saturation (8 cores)

Calculated probability of value sizes									
Value size [Bytes]	128	256	512	768	1014	2048	4096	22000	32000
Facebook: ETC	0.55	0.075	0.285	0.015	0.025	0.025	0.025	0	0
Facebook: USR	1	0	0	0	0	0	0	0	0
Facebook: APP	0.12	0	0.63	0.21	0.03	0.01	0	0	0
Facebook: VAR	0.78	0.02	0.17	0.03	0	0	0	0	0
Twitter	0	0	0	0.1	0.85	0.05	0	0	0
Wiki	0	0	0	0	0.58	0.02	0.1	0.25	0.05
Flicker	0	0	0	0	0	0	0	0.1	0.9
Youtube	0	0	0	0	0	0.75	0.11	0.11	

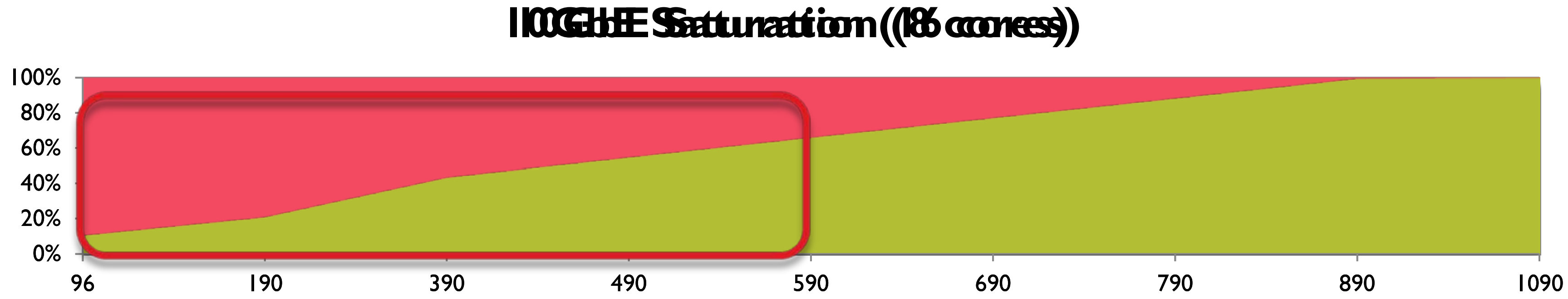
Network not saturated for small Object sizes



Calculated probability of value sizes

Value size [Bytes]	128	256	512	768	1014	2048	4096	22000	32000
Facebook: ETC	0.55	0.075	0.285	0.015	0.025	0.025	0.025	0	0
Facebook: USR	1	0	0	0	0	0	0	0	0
Facebook: APP	0.12	0	0.63	0.21	0.03	0.01	0	0	0
Facebook: VAR	0.78	0.02	0.17	0.03	0	0	0	0	0
Twitter	0	0	0	0.1	0.85	0.05	0	0	0
Wiki	0	0	0	0	0.58	0.02	0.1	0.25	0.05
Flicker	0	0	0	0	0	0	0	0.1	0.9
Youtube	0	0	0	0	0	0.75	0.11	0.11	

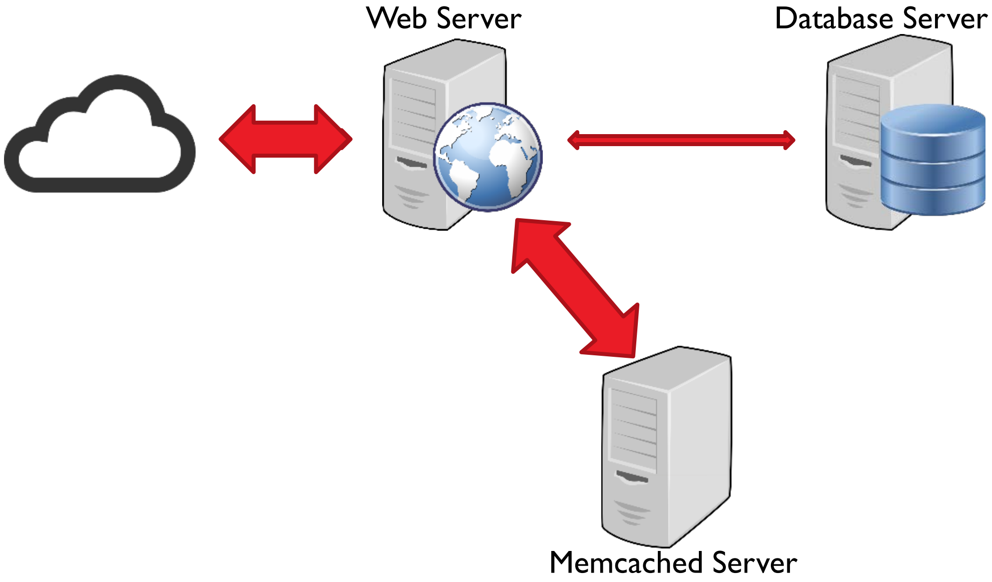
Network not saturated for small Object sizes



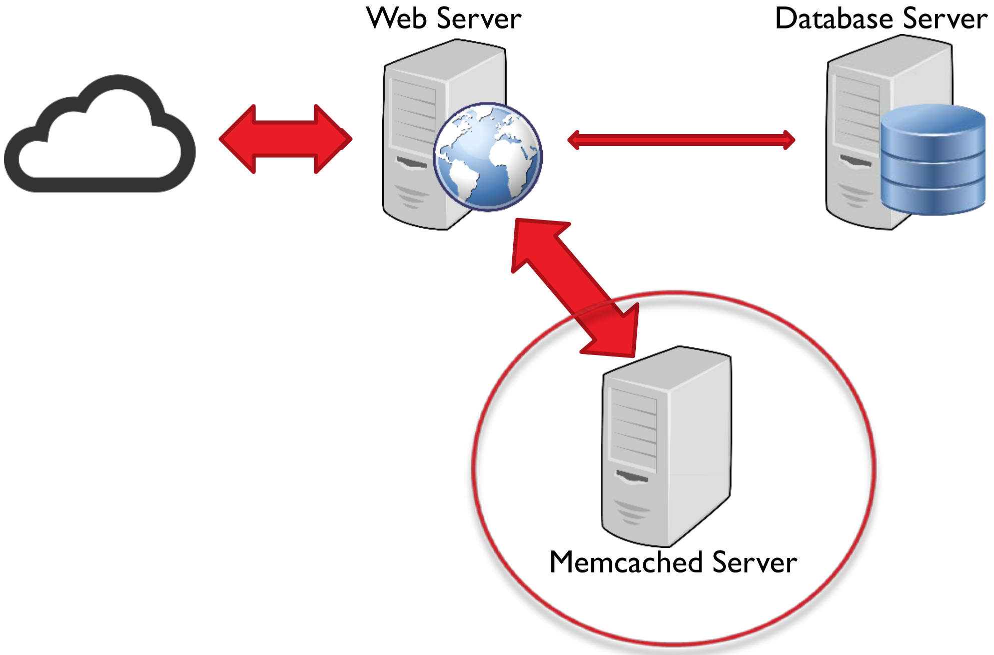
Calculated probability of value sizes									
Value size [Bytes]	128	256	512	768	1014	2048	4096	22000	32000
Facebook: ETC	0.55	0.075	0.285	0.015	0.025	0.025	0.025	0	0
Facebook: USR	1	0	0	0	0	0	0	0	0
Facebook: APP	0.12	0	0.63	0.21	0.03	0.01	0	0	0
Facebook: VAR	0.78	0.02	0.17	0.03	0	0	0	0	0
Twitter	0	0	0	0.1	0.85	0.05	0	0	0
Wiki	0	0	0	0	0.58	0.02	0.1	0.25	0.05
Flicker	0	0	0	0	0	0	0	0.1	0.9
Youtube	0	0	0	0	0	0.75	0.11	0.11	

How would this look with 16 cores?

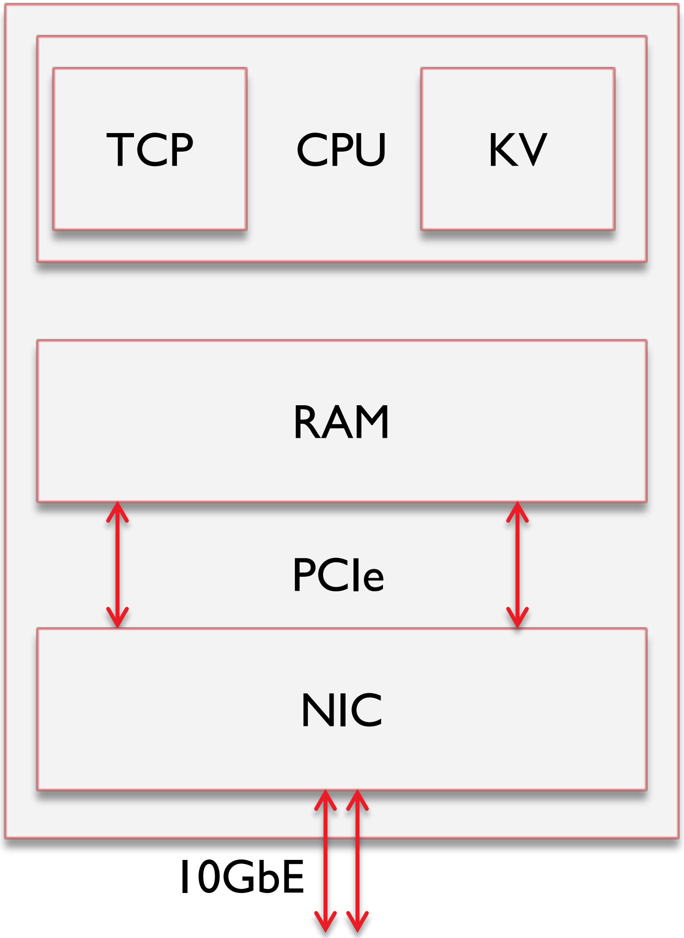
Memcached



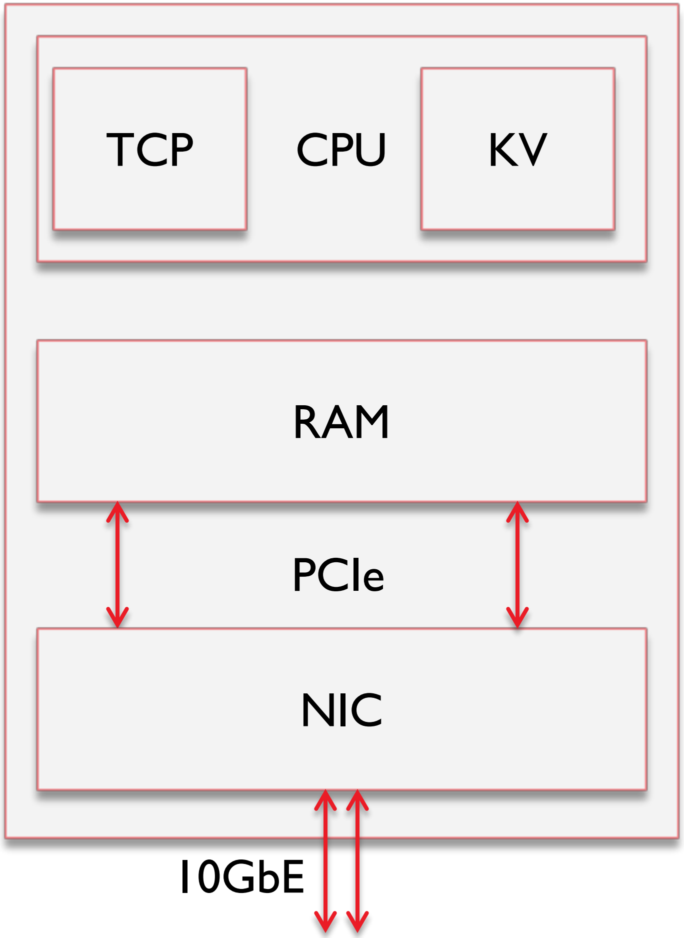
Memcached

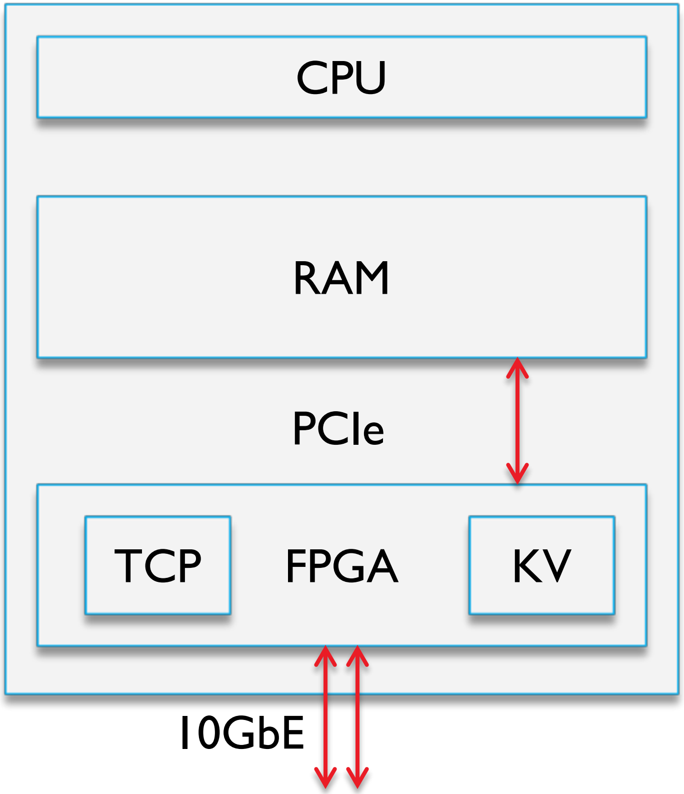


High Level View of Software Approach

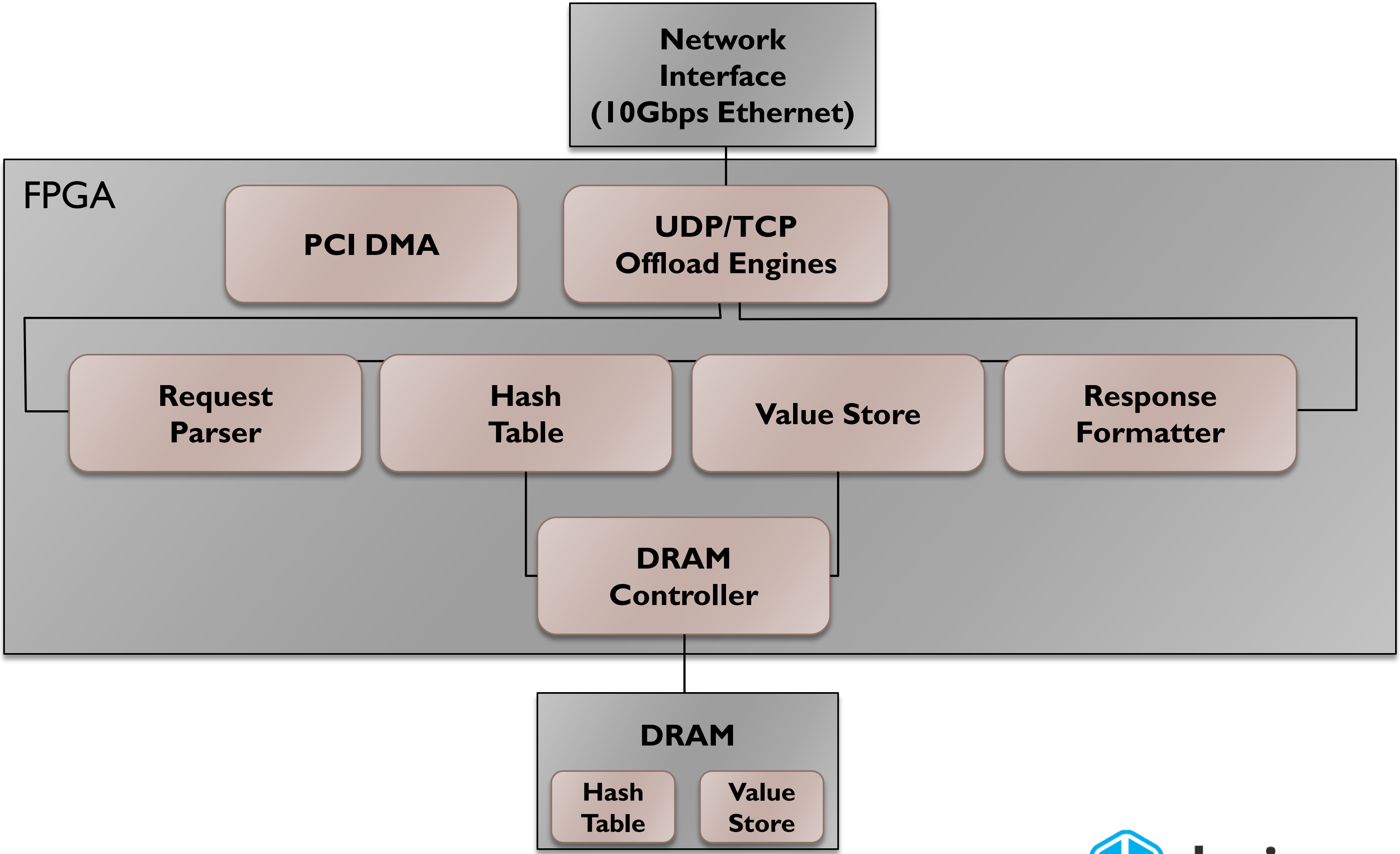


Alternative approach with FPGA



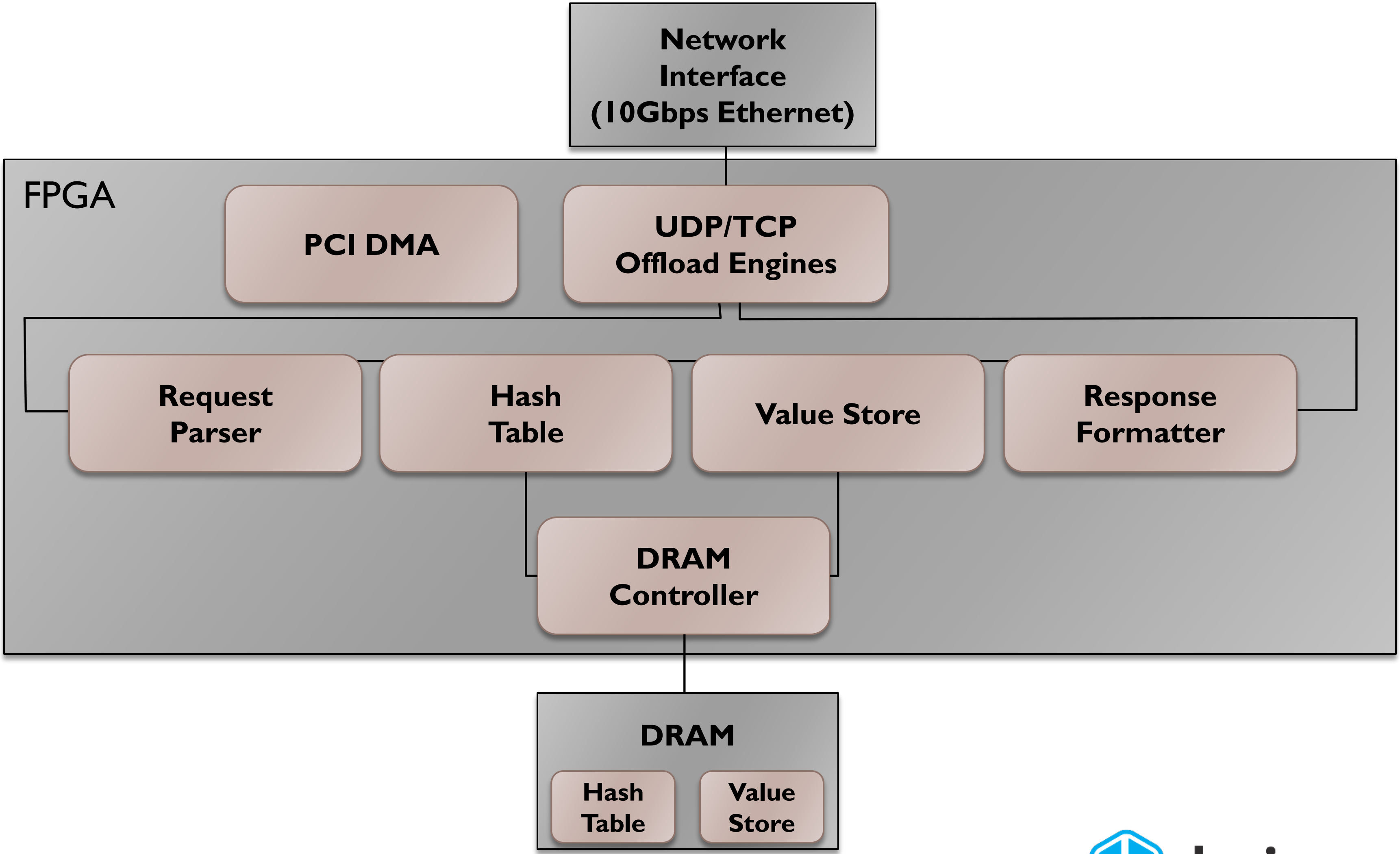


Xilinx Memcached processing pipeline



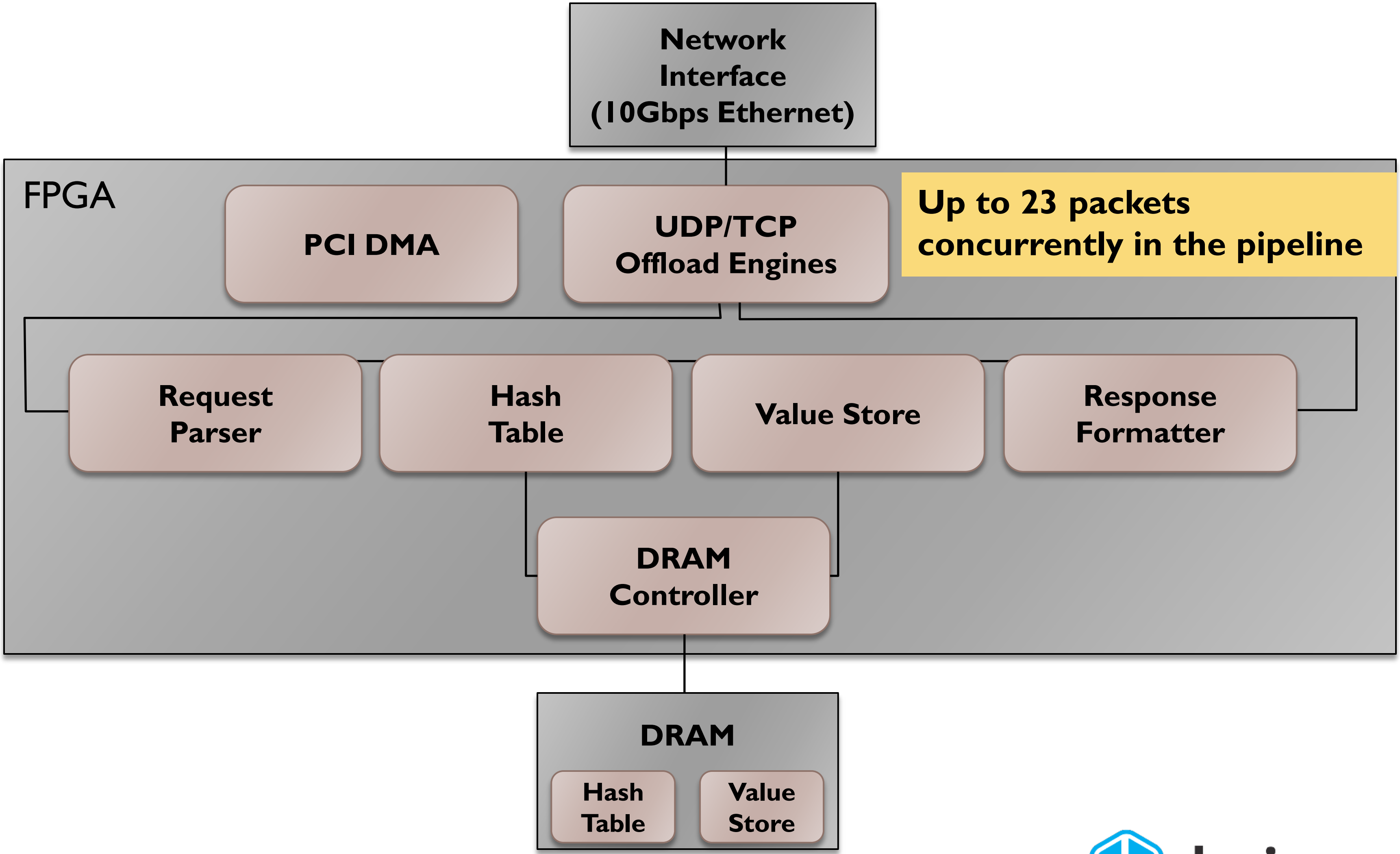
Xilinx Memcached processing pipeline

481 cycles @ 156MHz vs. 0.5-1Million cycles @ 2GHz



Xilinx Memcached processing pipeline

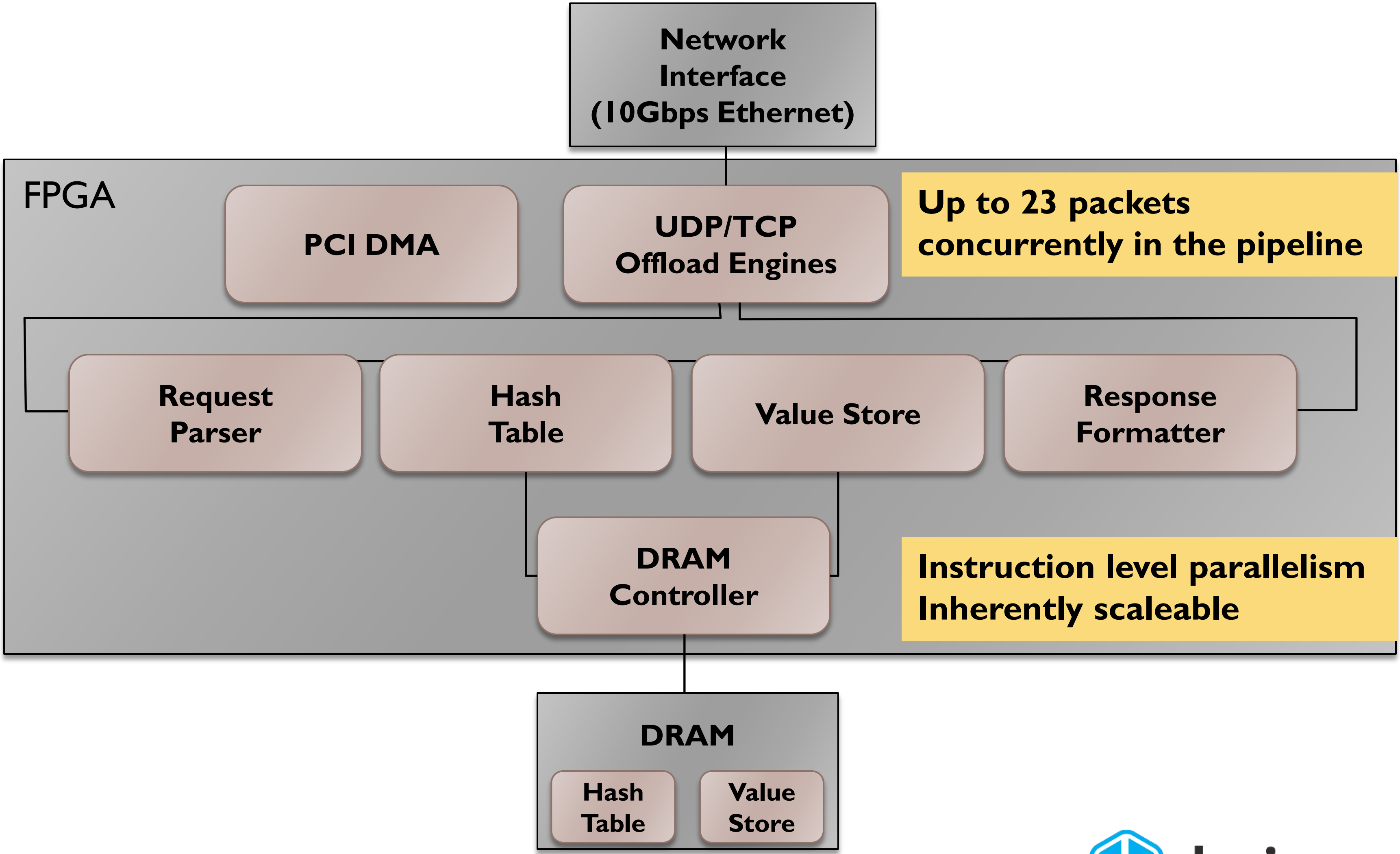
481 cycles @ 156MHz vs. 0.5-1Million cycles @ 2GHz



Up to 23 packets concurrently in the pipeline

Xilinx Memcached processing pipeline

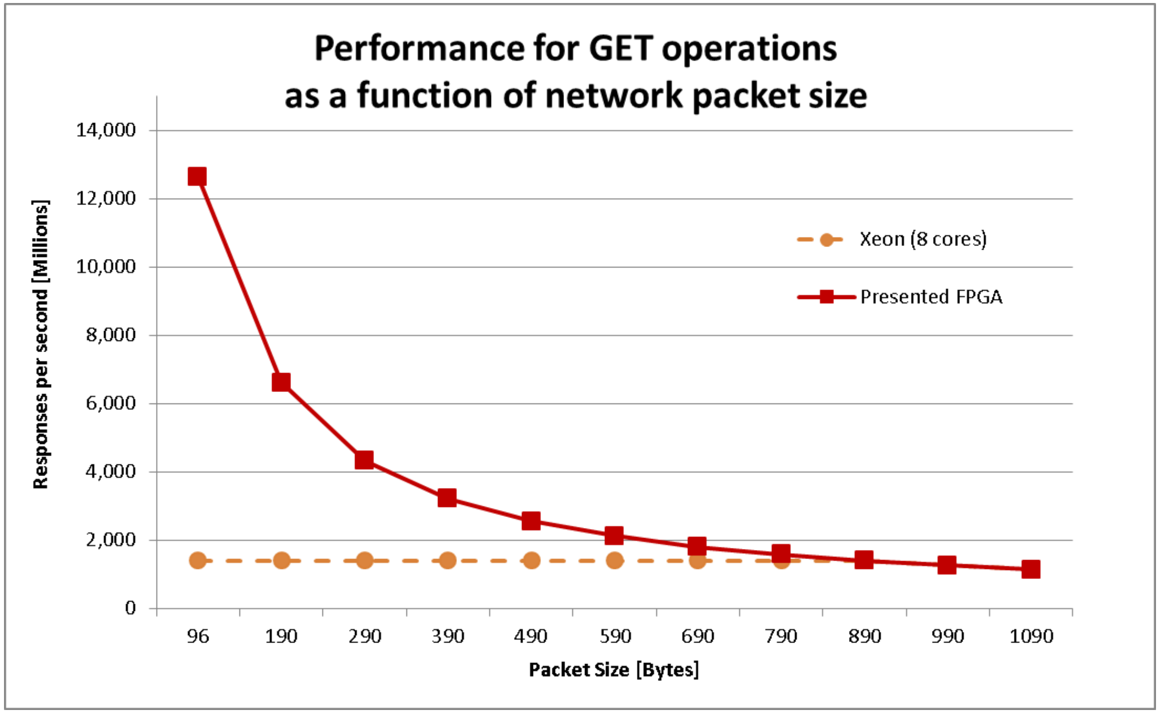
481 cycles @ 156MHz vs. 0.5-1Million cycles @ 2GHz



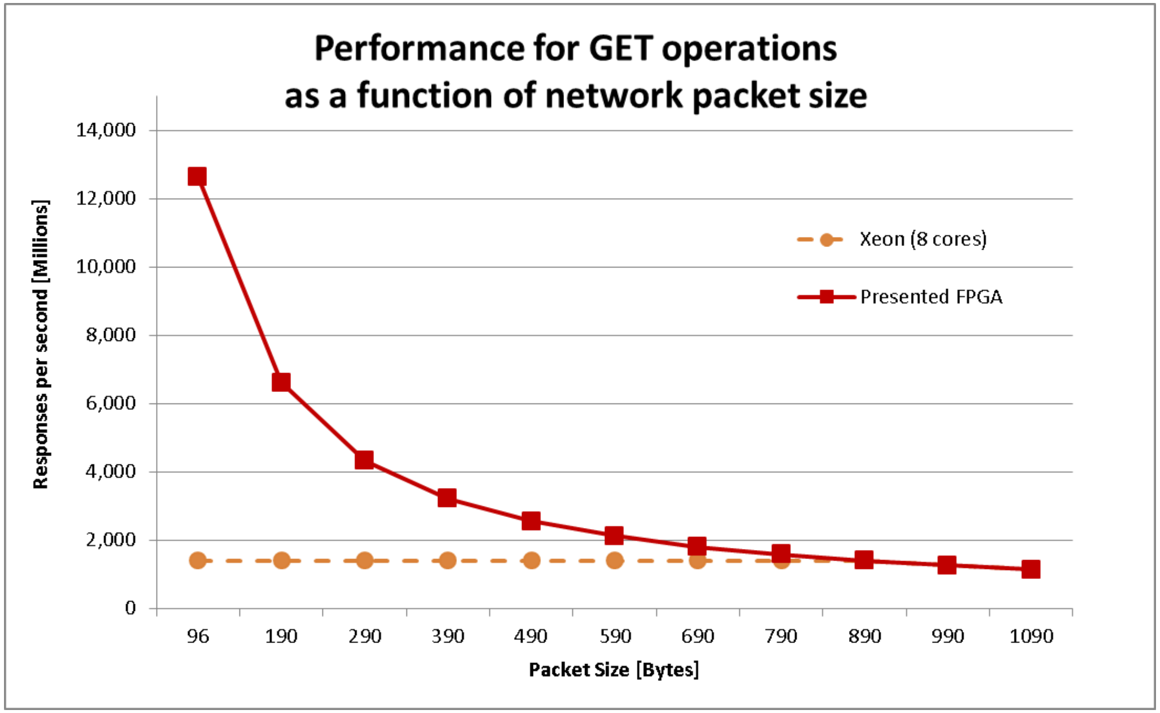
Up to 23 packets concurrently in the pipeline

Instruction level parallelism
Inherently scaleable

FPGA vs. Xeon 8 core with tuned Memcached

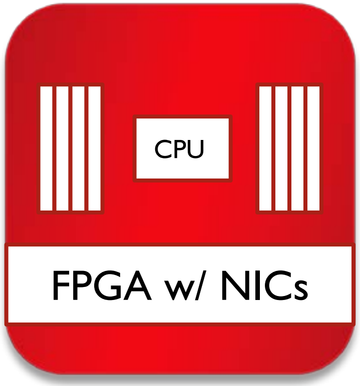


FPGA vs. Xeon 8 core with tuned Memcached

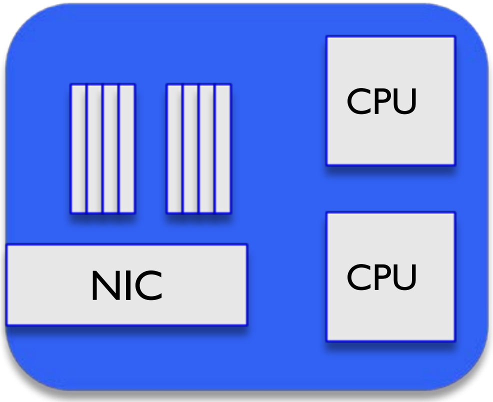


Platform	RPS [M]	Latency [us]	RPS/W [K]
Intel Xeon (8 cores)	1.34	200-300	7
FPGA (board only)	Up to 13.02	3.5-4.5	254.8
FPGA (with host)	Up to 13.02	3.5-4.5	106.7

Memcached server cost, power, performance



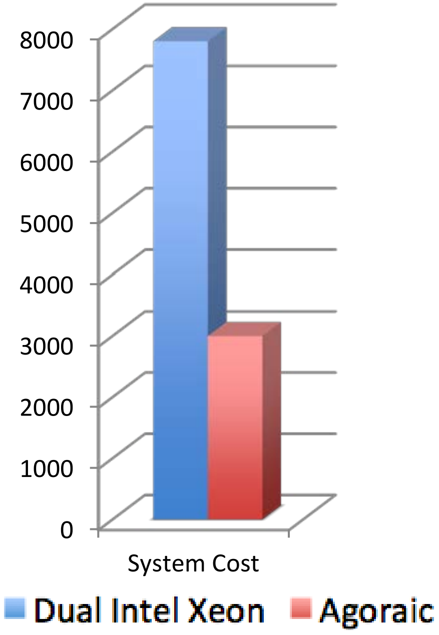
Agoraic Solution



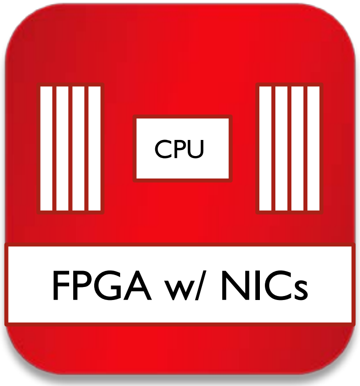
Intel Dual Xeon

devicepros

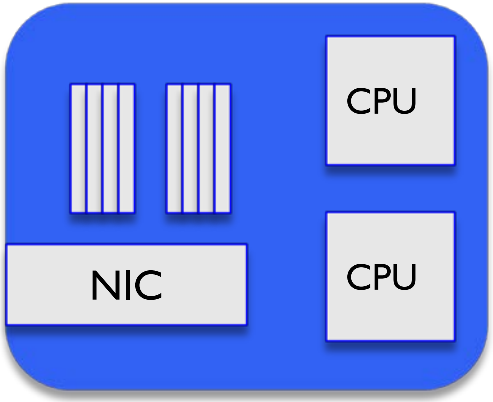
Memcached server cost, power, performance



60% lower
System Cost

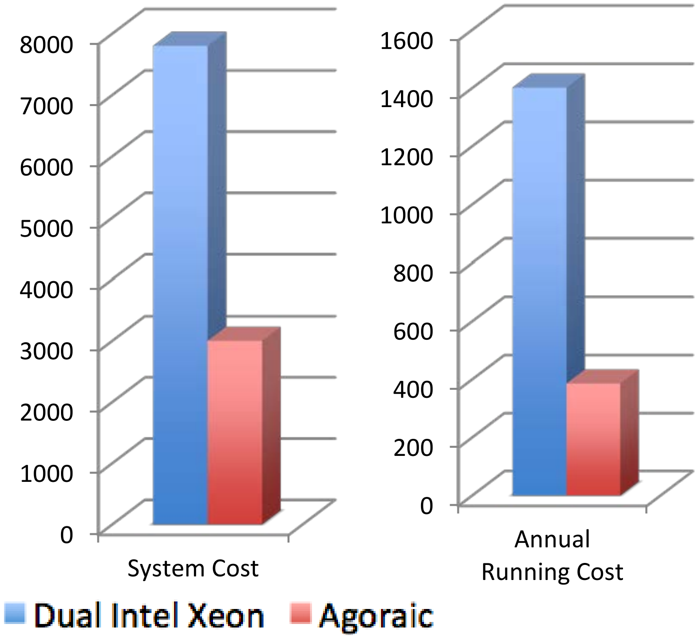


Agoraic Solution



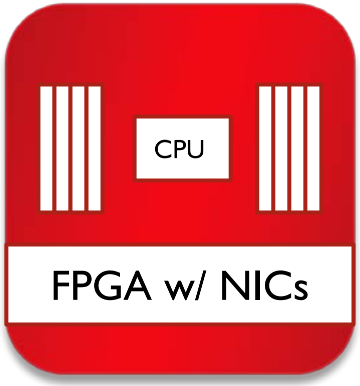
Intel Dual Xeon

Memcached server cost, power, performance

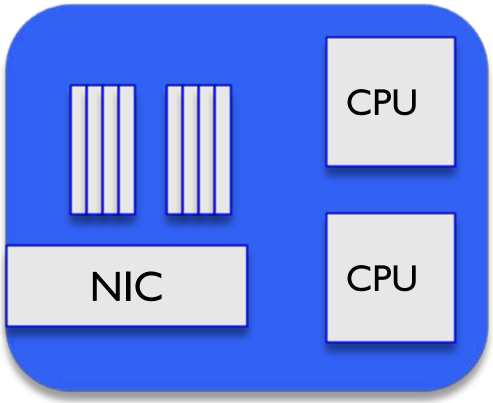


60% lower
System Cost

73% lower
Annual
Running Cost

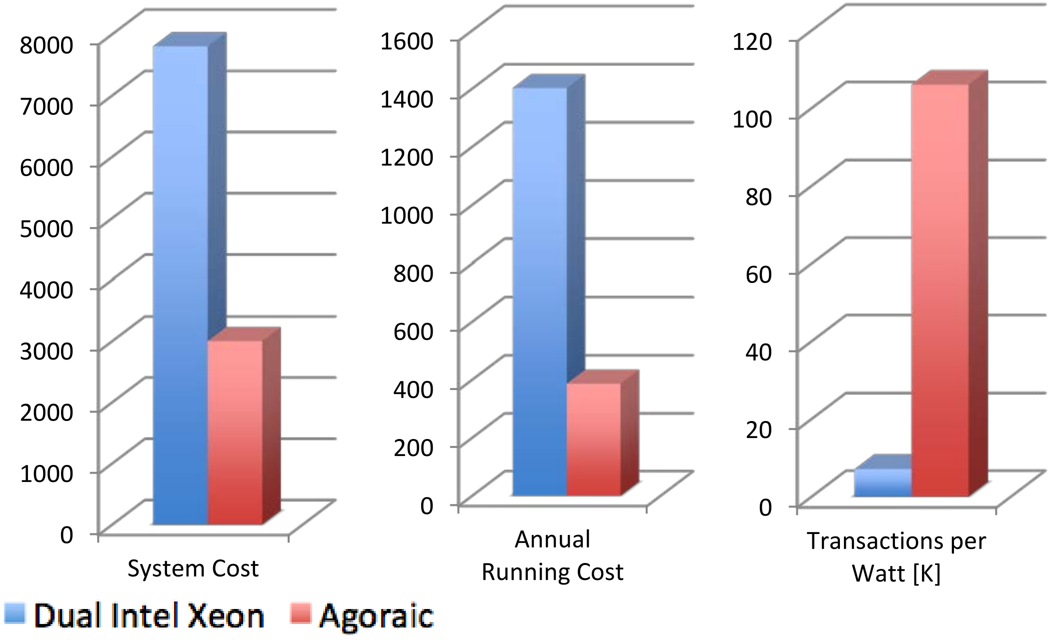


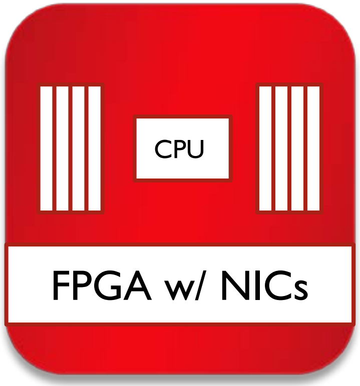
Agoraic Solution



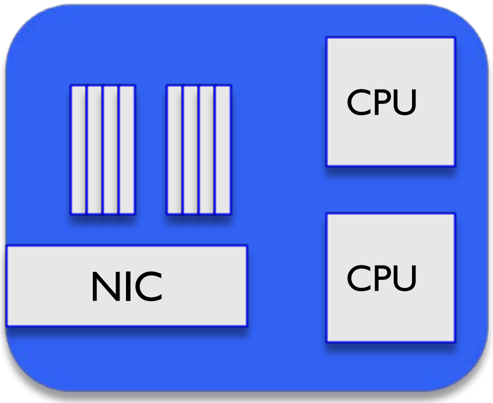
Intel Dual Xeon

Memcached server cost, power, performance





Agoraic Solution



Intel Dual Xeon

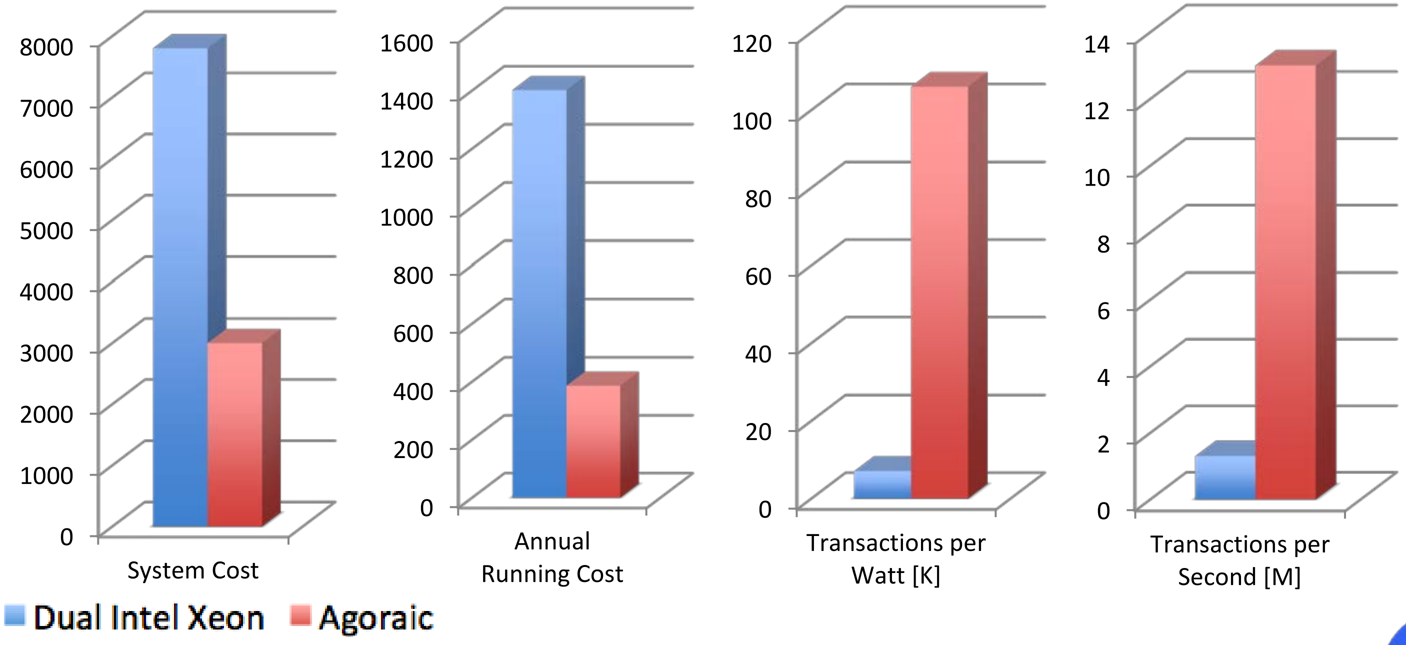
60% lower
System Cost

73% lower
Annual
Running Cost

15X
Transactions
Per Watt

devicepros

Memcached server cost, power, performance

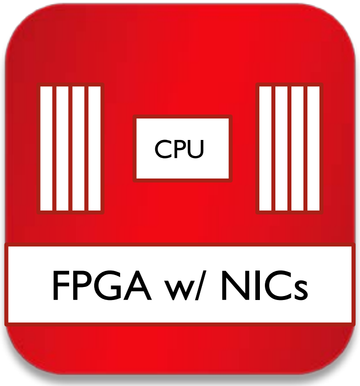


60% lower
System Cost

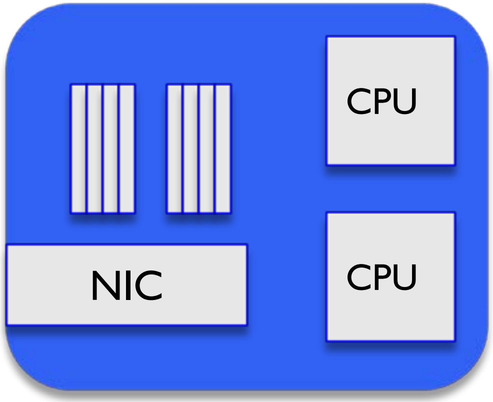
73% lower
Annual
Running Cost

15X
Transactions
Per Watt

10X
Transactions
Per Second



Agoraic Solution



Intel Dual Xeon

Scoring Memcached

Factors to consider	
Volume	Yes, 10-30% of servers 10X server growth by 2020
Low complexity, narrow focus	Yes
Stable design	Yes
Needs lower latency	Yes
Needs better throughput	Yes, but seems achievable with multiple cores as well as HW approach.
Power efficiency matters	Yes, for all objects sizes 90 – 95% power reduction

devicepros

Conclusions Memcached HW Acceleration

Favorable factors for the FPGA approach

- 7K queries / Watt vs. 100-200 queries / Watt
- Significantly better max latency and distribution.

Some Caveats

- Similar throughput looks achievable without HW Acceleration
  - ❑ Could 64 cores yield 12 million TPS? Need to put this in context with power measurement for TPS per Watt.
- The FPGA results are experimental vs. the x86 results which included a more comprehensive feature set including exception handling and support for larger numbers of TCP sessions.

devicepros

References, links

[1] Michaela Blott, Kees Vissers - Xilinx Research, Dataflow Architectures for 10Gbps Line-rate Key-value-Stores

http://www.hotchips.org/wp-content/uploads/hc_archives/hc25/HC25.50-FPGA-epub/HC25.27.510-Dataflow-Blott-Vissers-Xlinix-final_no_animation.pdf and Video: http://youtu.be/16eoLZ-wIWA

[2] ATIKOGLU, B., XU, Y., FRACHTENBERG, E., JIANG, S., AND PALECZNY, M. Workload analysis of a large-scale key-value store

[3] WIGGINS, A., AND LANGSTON, J. Enhancing the scalability of memcached. In Intel Software Network (2012).