World's Fastest Memcached Database on AWS

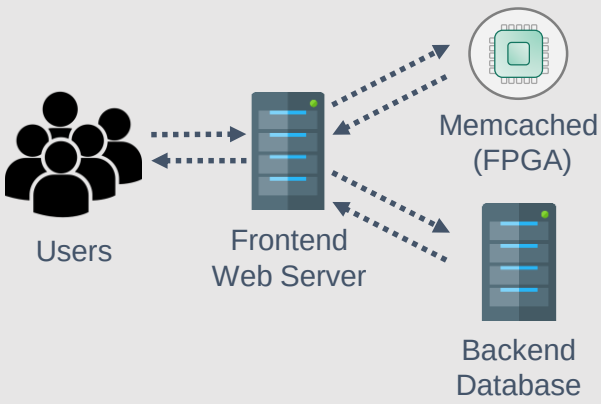
Accelerated using FPGA reconfigurable hardware

- 9X higher requests per second, over 11M requests/second
- 9X lower latency, less than 0.3ms response time
- 10X lower total cost of ownership
- Drop-in replacement for existing Memcached server

INTRODUCTION

Memcached is a high-performance in-memory object caching system, used by Facebook, Flicker, Wikipedia, and other high-traffic websites. Memcached acts as a caching layer between web servers and databases to decrease server response times.

FPGA compute instances are now being deployed in datacenters to accelerate network-centric workloads. LegUp Computing offers a cloud-deployed Memcached using AWS EC2 F1 (FPGA) instances. With a single F1 instance, LegUp's Memcached server achieves over 11M ops/sec, a 9× improvement over AWS ElastiCache, an AWS-managed CPU Memcached service. Our FPGA-accelerated Memcached server responds to network requests in under 300µs, a 9× improvement in latency over ElastiCache.



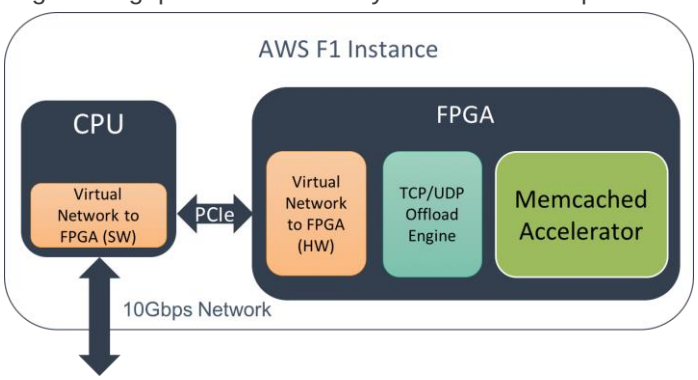
PRODUCT OVERVIEW

LegUp provides the world's fastest cloud-hosted Memcached on AWS using Amazon EC2 F1 (FPGA) instances.
Our FPGA Memcached server is compatible with standard Memcached client APIs meaning no software changes.
Simply launch LegUp's Memcached instance on AWS and connect your web server.
We also offer an on-premise solution for the datacenter with Xilinx's Alveo FPGA boards.

SOLUTION OVERVIEW

Memcached Architecture

We use a 10Gbps TCP/IP network stack on the FPGA for high throughput and low latency Memcached responses.



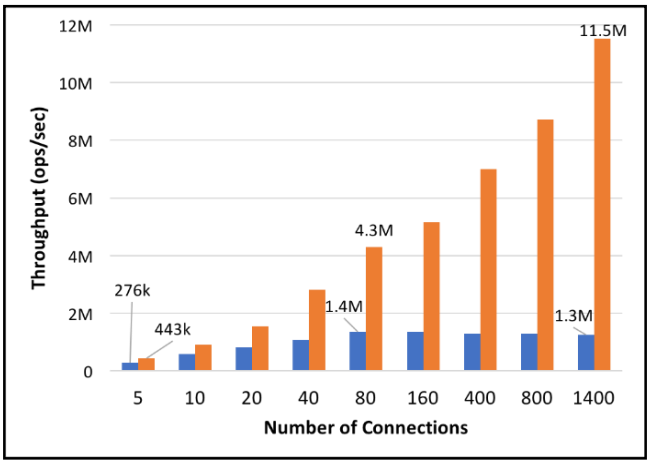
Lower Total Cost of Ownership

10% lower EC2 instance cost vs. comparable RAM size ElastiCache instance. With 9× higher ops/sec, LegUp's Memcached offers 10× better throughput/$.

AWS Instance	RAM	Cost
cache.r4.4xlarge	101GB	$1.82/hour
f1.2xlarge	122GB	$1.65/hour

9× Higher Throughput vs. AWS ElastiCache

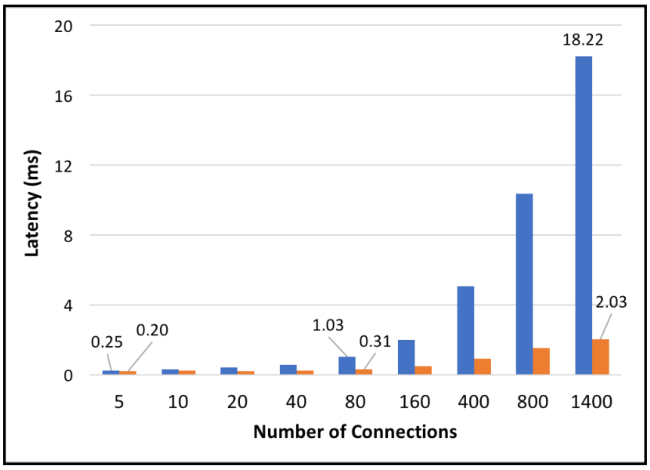
Benchmark tool: memtier_benchmark with 100B data size
Pipelining (batching) of 16, 1:1 get:set request ratio



AWS ElastiCache
FPGA Memcached

9× Lower Latency vs. AWS ElastiCache

Round-trip time including network latency
FPGA server latency is consistently <0.3ms



AWS ElastiCache
FPGA Memcached

EXTENSIBILITY & CONCLUSION

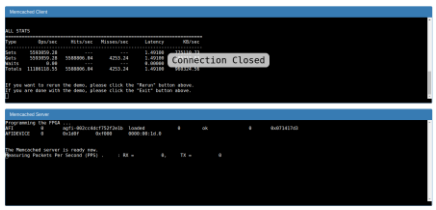
The FPGA running the Memcached server is only 20% utilized, leaving plenty of room to add user-defined functions. With LegUp's high-level synthesis technology, you can easily add additional custom computing logic to the network stack and Memcached pipeline by programming in C/C++ while keeping the same Memcached throughput and similar latency. Examples include: compression, encryption, and data analytics. The CPU is also completely free to be used. LegUp's FPGA-powered Memcached server offers the best throughput and lowest latency for customers who need faster response times and lower total cost of ownership.

TAKE THE NEXT STEP

Try our live demo by visiting:
http://www.legupcomputing.com/main/memcached_demo

- Spins up two AWS EC2 instances in the same placement group:
  F1 Memcached server and a client server
- Runs memtier_benchmark to measure requests/second and latency.

We'd love to hear from you at: info@legupcomputing.com



ABOUT LEGUP

LegUp Computing provides a platform that enables software developers to program, deploy, scale, and manage FPGA devices for accelerating high-performance applications. The LegUp platform offers a novel programming model that allows threaded C/C++ software to target FPGA devices connected to high-bandwidth networks for efficient processing of low-latency database, real-time analytics, and machine learning workloads. The company was founded by Dr. Andrew Canis, Dr. Jongsok Choi, Ruolong Lian, and Professor Jason Anderson in 2015 to commercialize the award-winning open-source LegUp high-level synthesis tool from the University of Toronto.

LegUp Computing Inc.
www.LegUpComputing.com
info@legupcomputing.com
+1 (647) 834-6654

88 College Street
Toronto, Ontario
Canada
M5G 1L4