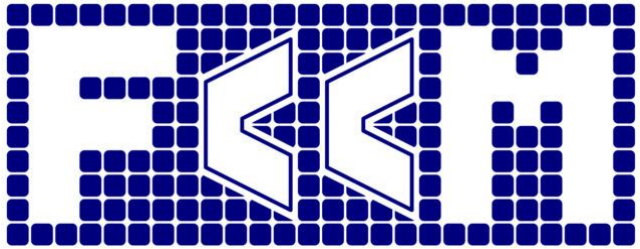
Cloud-Scale Key Value Store in FPGA

[John W. Lockwood](), PhD & CEO
[Algo-Logic Systems, Inc.]()



28th IEEE International Symposium
on Field Programmable Custom
Computing Machines (FCCM-2020)

May 6, 2020

About the Speaker

John W. Lockwood
CEO & PhD
Algo-Logic Systems, Inc.

John W. Lockwood is the founder and CEO of Algo-Logic Systems, Inc. He designs and implements networking systems in reconfigurable hardware, specifically in the areas of low latency networking, Internet security, and electronic commerce.
Previously, Prof. Lockwood managed the NetFPGA program in the Department of Electrical Engineering at Stanford University and led the Reconfigurable Network Group as a Tenured Professor in the department of Computer Science at Washington University in St. Louis.

In Collaboration With

Outline

1. Motivation for Ultra-High-Performance Key Value Store (KVS)
2. Field Programmable Gate Array (FPGA) Acceleration
3. Implemented KVS solution in a 1U Rackmount Dell Server
4. How we achieve 490M IOPs with the KVS in logic + Redis in software
5. Scaling the KVS beyond 1B IOPs in a 1U server and 40B IOPs per rack

ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Motivation for Key Value Store (KVS) in the Cloud

In-Memory KVS systems are used widely in the cloud

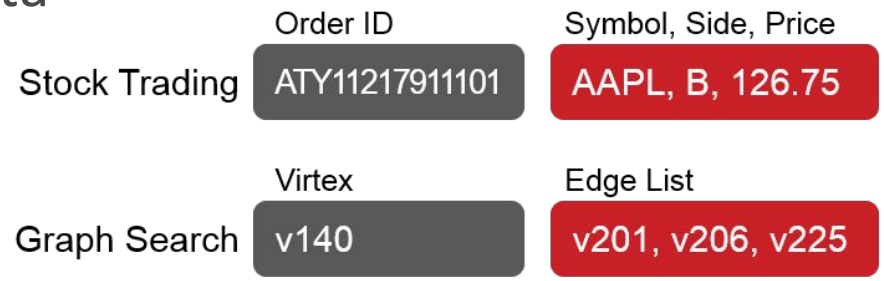
- Amazon DynamoDB
  - Used for shopping carts & active session store (profile, messages, target promotions)
  - Milliseconds of latency to retrieve small values (< 400 KB)
- Facebook RocksDB
  - Used to track the state of users, graph search, and cache for Hadoop
  - Embedded database for key-value data written in C/C++ using RAM and Flash
- Microsoft FASTER
  - *"Managing large application state easily, resiliently, and with high performance is one of the hardest problems in the cloud today"*
- Redis
  - Portable across all cloud providers and available for on-premise deployments
  - Open-source code base with professional support

ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Motivation for Fast, High-Throughput, Compact KVS

Key Value Stores can be used for Real-Time Big Data

- Market data for stocks, options, and futures
  - Tick at **nanosecond** time-scales
- Graph analytics operate on big data
  - Searches can require billions of IOPs



Space is at a Premium in Data Centers

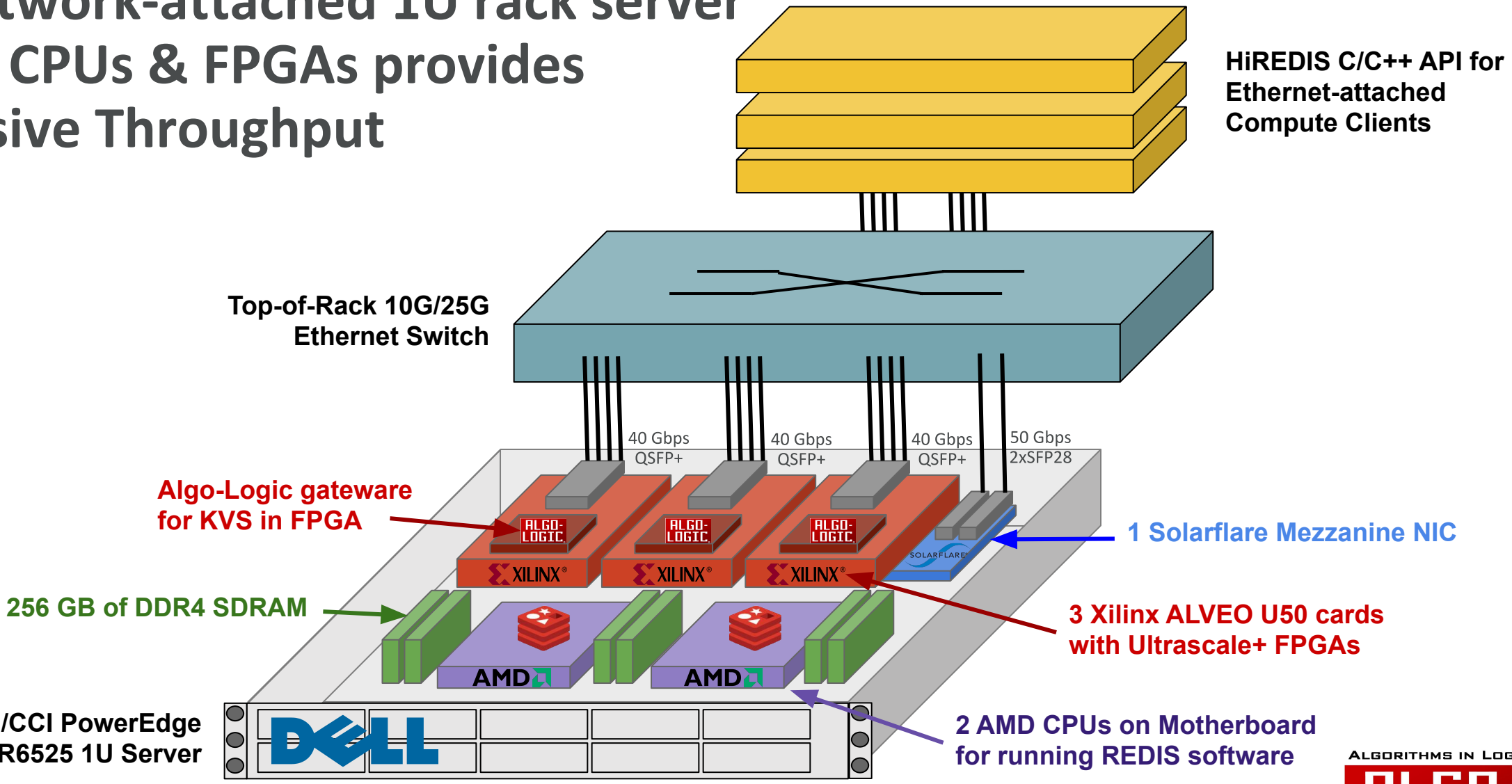
- 1U Server occupies minimum space
- 1.75” tall and 19” wide

Field Programmable Gate Arrays (FPGAs)

- Dramatically increase network throughput
- Radically reduce latency
- Fit within expansion slots of a standard Dell Server

ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

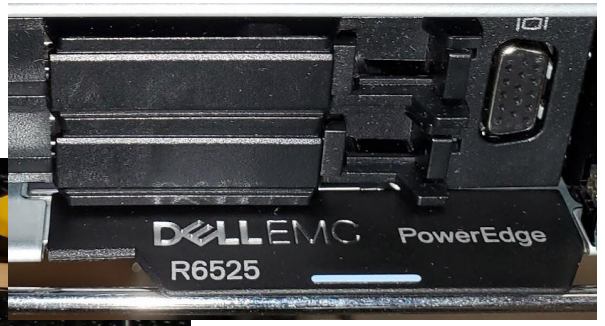
A network-attached 1U rack server with CPUs & FPGAs provides massive Throughput

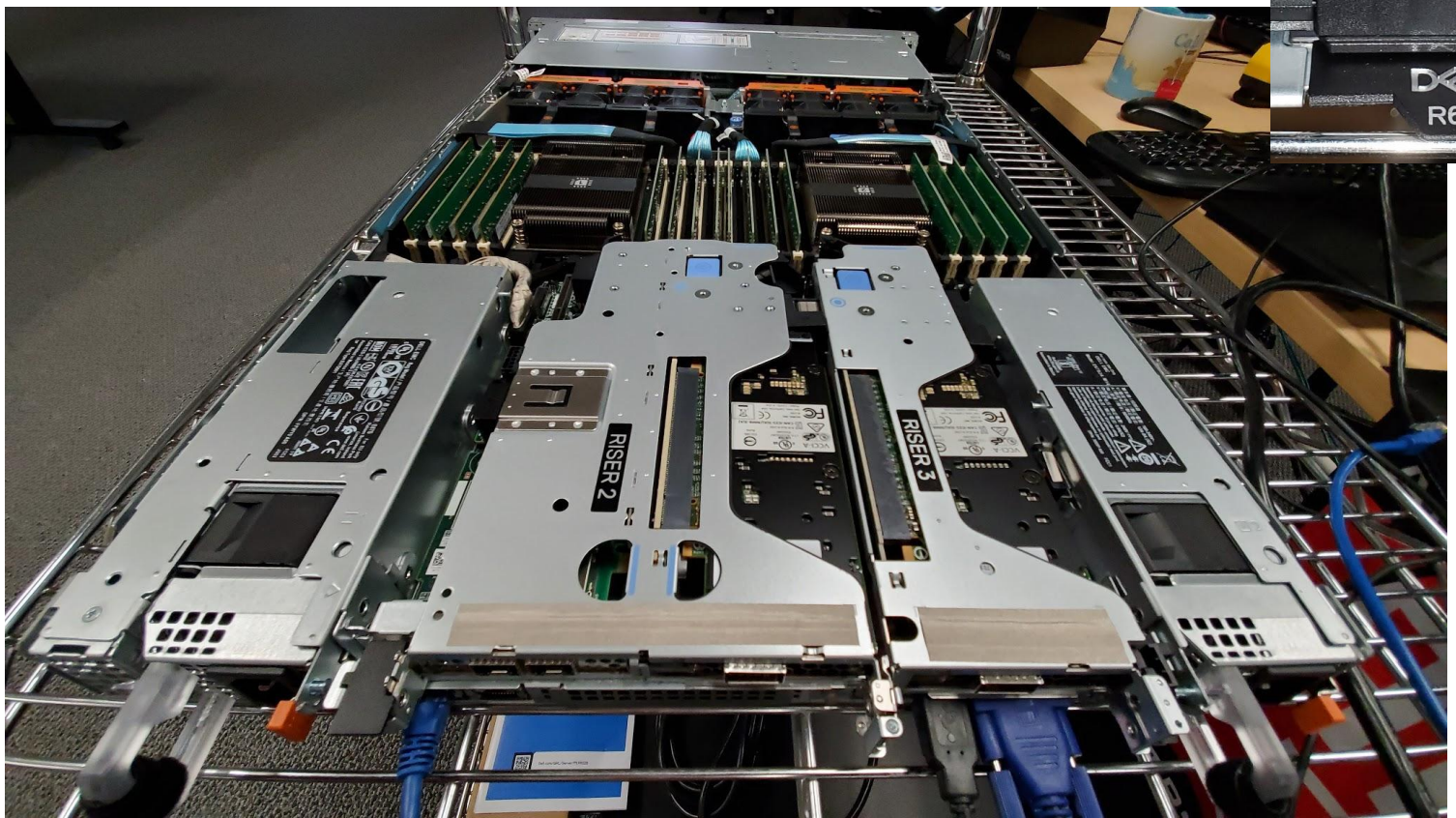


ALGORITHMS IN LOGIC
ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

Details of the Dell/CCI PowerEdge R6525 1U Rack Server

- Two AMD EPYC 7402 24-core CPUs (96-way multi-threaded)
- 256 GB of ECC DRAM using 16 DDR4 DIMMs
- Three half-height slot with Xilinx U50 FPGA cards with UltraRAM
- One Mezzanine slot with Solarflare Cloud Onload NIC





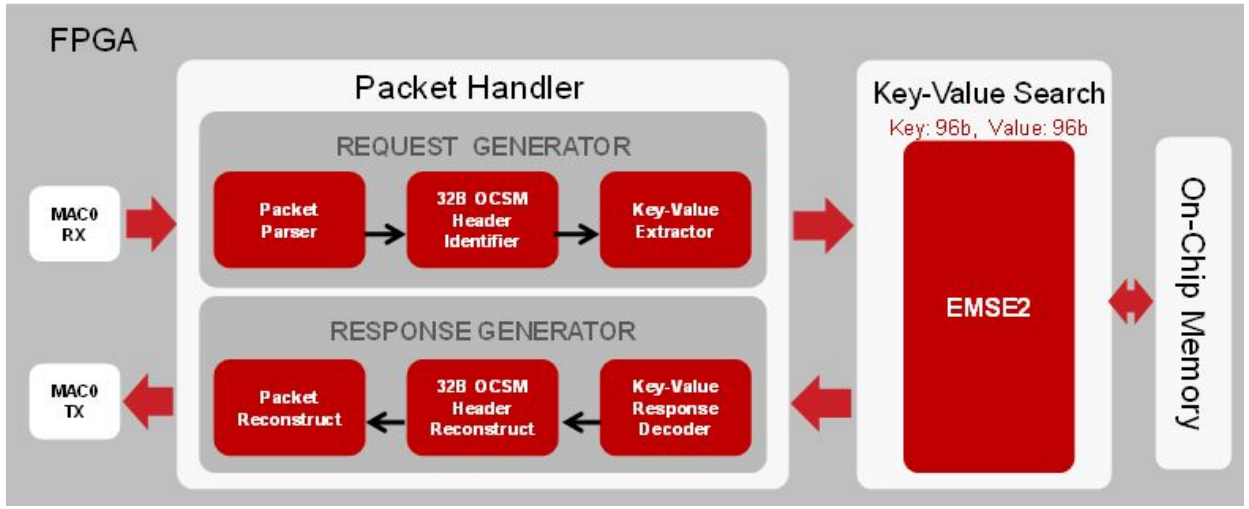
ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Software and Gateware

- Software
  - CentOS 7.7 Linux Operating System
  - SolarFlare Kernel-Bypass Cloud Onload NIC
  - REDIS 5.0.8
- Gateware
  - Algo-Logic Key Value Store (KVS)
- Client Software API
  - C/C++ API using HiRedis to REDIS software
  - C/C++ API modeled on HiRedis to KVS gateware

ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Algo-Logic's Network-Attached KVS in FPGA Logic



ALGORITHMS IN LOGIC
ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

Examples:	Key	Value
Directory	Company: Algo-Logic	Phone #: (408) 707-3740
Forwarding Tables	IP Address: 204.2.34.5	Interface : MAC Address: Eth6 : 02:33:29:F2:AB:CC
Data De-duplication	Content Hash: XYZ	Storage Block ID: 948830038411
Stock Trading	Order ID: ATY11217911101	Symbol, Side, Price: AAPL, B, 126.75
Graph Search	Virtex: v140	Edge List: v201, v206, v225

See Also: Algo-Logic GDN Search (Key Value Store)

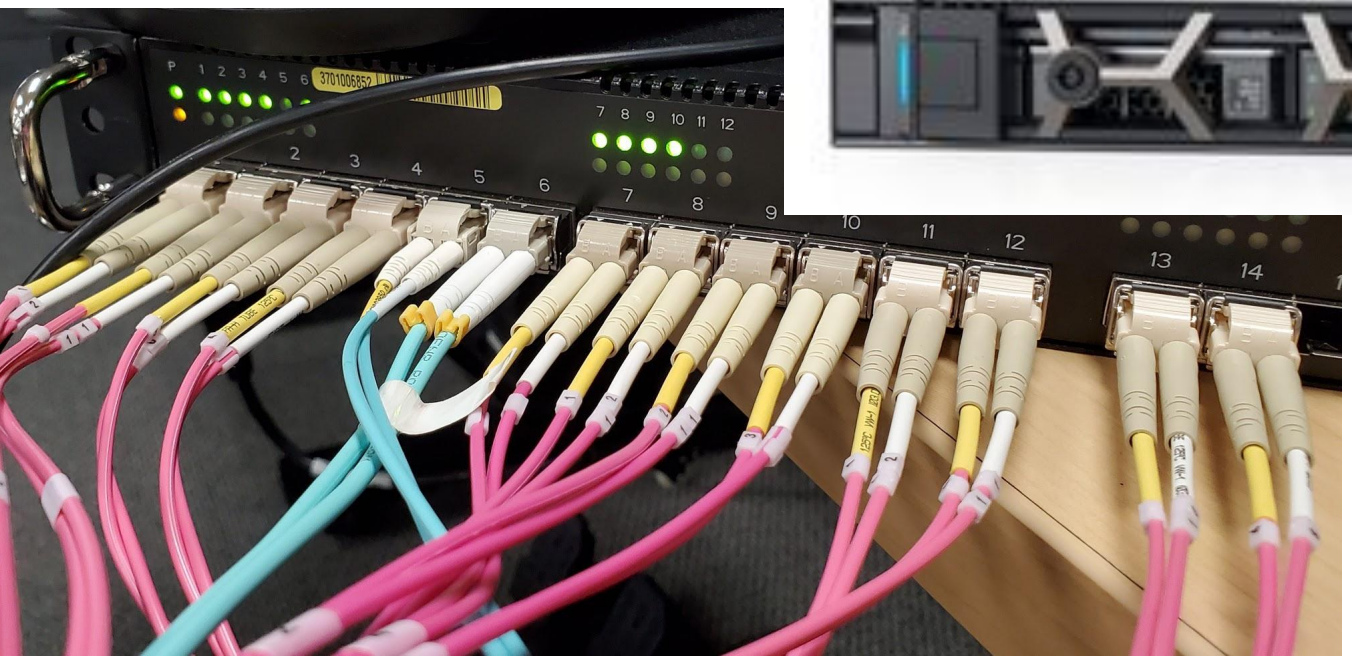
ALGORITHMS IN LOGIC
ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

Xilinx ALVEO U50 FPGA Card

- Ultrascale+ FPGA
  - 872k LookUp Tables (LUTs)
- QSFP28 Ethernet (100 Gbps)
  - Splits to 4 x SFP+ or SFP28 ports
- PCIe Form Factor
  - Half-Height
  - Single-slot

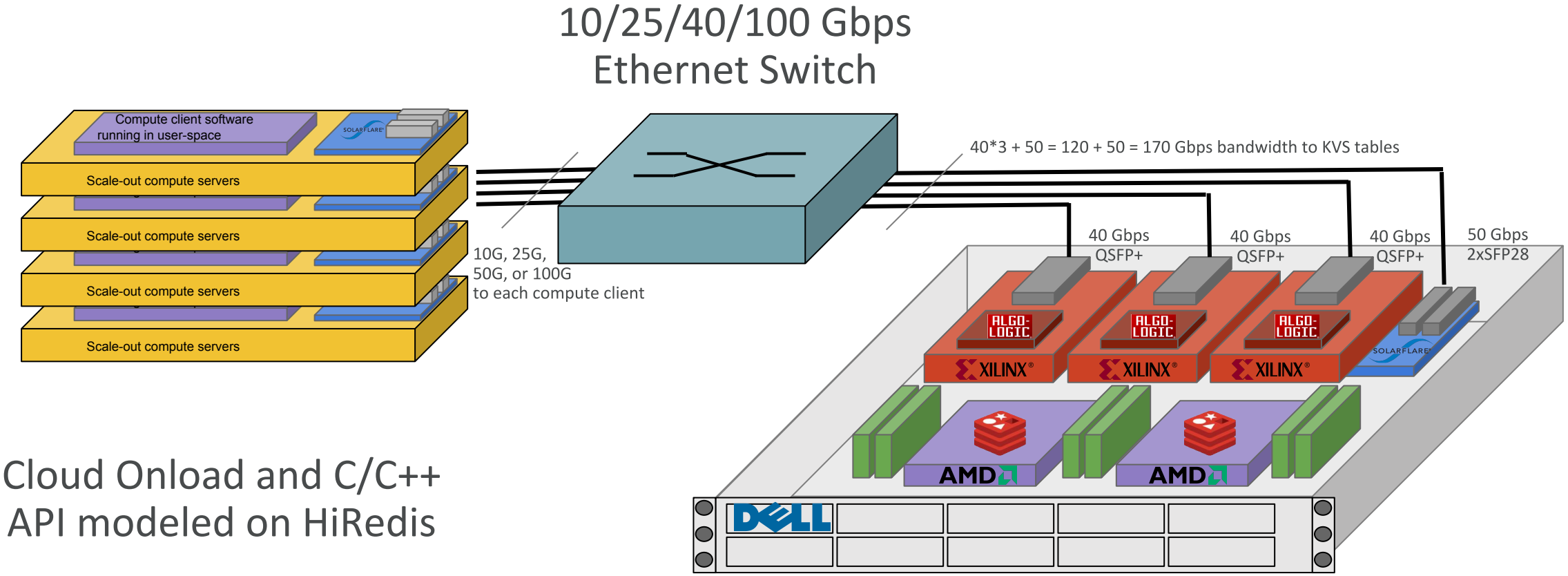
ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Patching ports from the Dell to top-of-rack switch



ALGORITHMS IN LOGIC
ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

Client Software:
HiREDIS C/C++ API



Cloud Onload and C/C++
API modeled on HiRedis

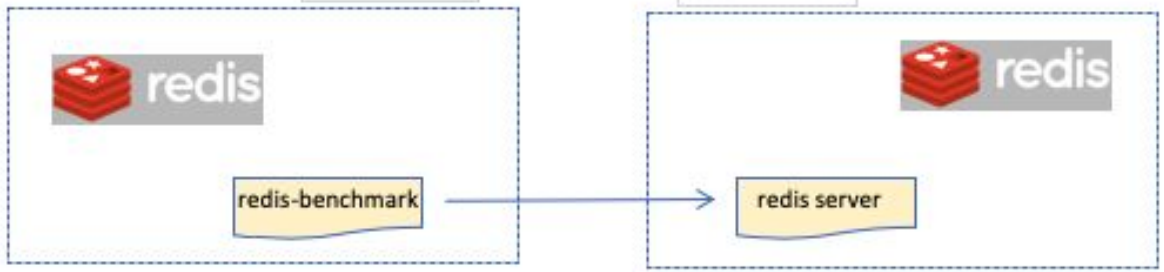
Key Value Store in 1U Server

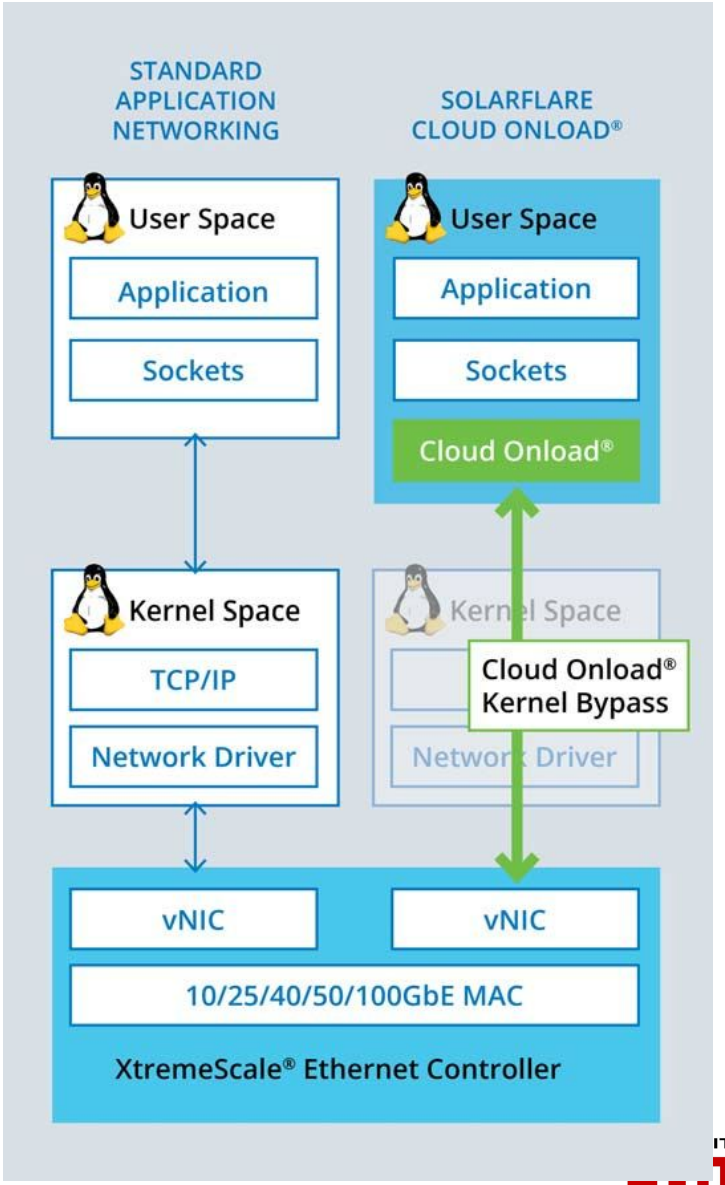
ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Optimizing Client Software

- Kernel bypass
  - Network driver & TCP/IP runs in user-space
  - Using Cloud Onload from Solarflare, a Xilinx company
  - Running on Solarflare X2 NIC

- redis-benchmark, as per:
  SF-121461-CD Solarflare Cloud Onload Redis Cookbook Issue 2



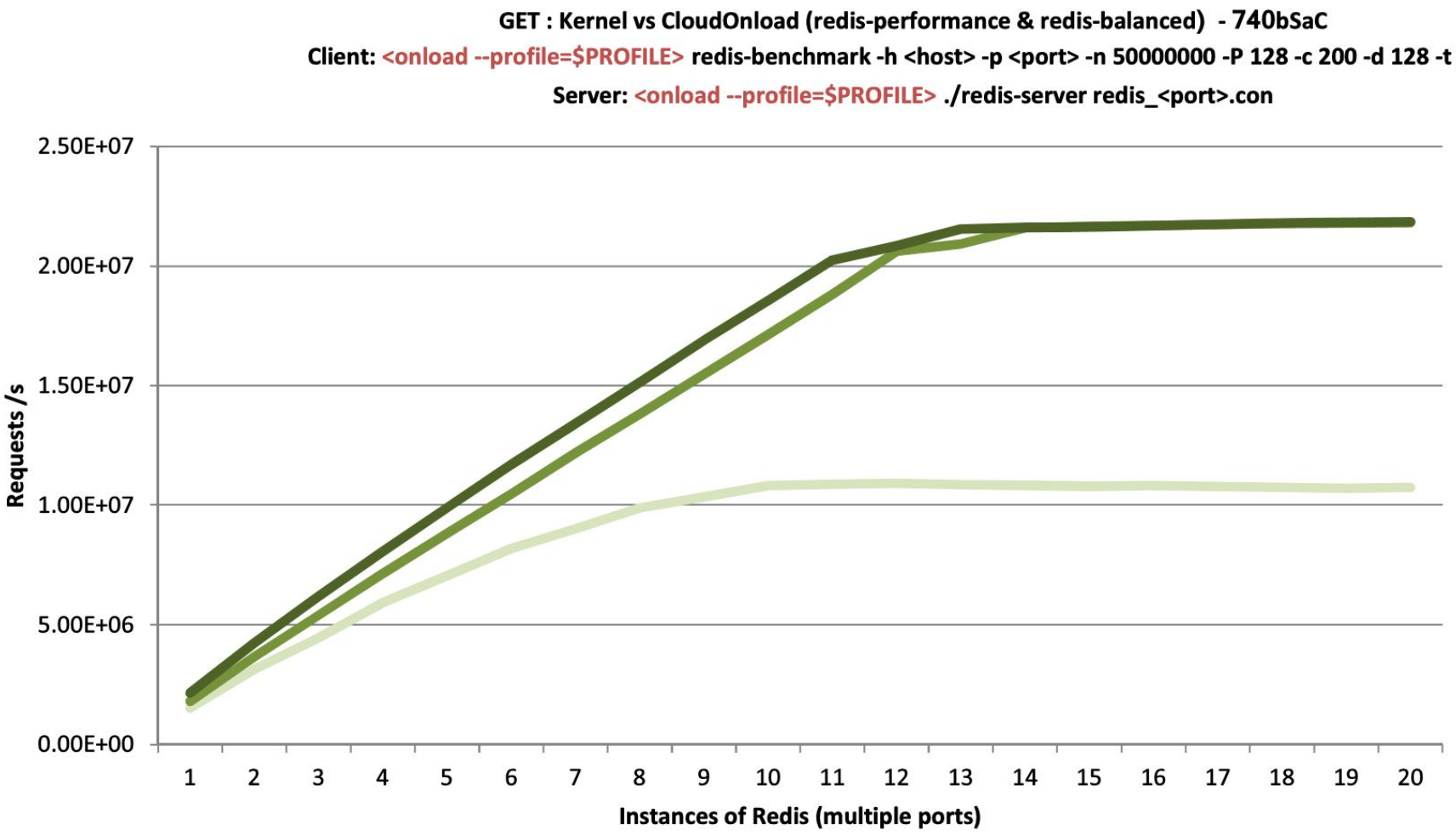


XILINX®

SOLARFLARE®

ALGORITHMS IN LOGIC
ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

Throughput of Redis in Software w/Solarflare Cloud Onload



20M IOPs per 25Gb/s link

25Gb/s Kernel

25Gb/s Onload performance

25Gb/s Onload balanced

SOLARFLARE®
A XILINX COMPANY

Clustered Multi Threading

From: SF-121461-CD Solarflare Cloud Onload Redis Cookbook Issue 2,
© 2019, Solarflare Communications, Inc.

See Also: Redis Running with Onload Sees a 100% Performance Gain

ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Throughput of Algo-Logic KVS

- Minimum-size object for Algo-Logic KVS in FPGA
  - 12 byte key, 12 byte value, 8 byte header = 32 Bytes/GET request
- Maximum-size packet = Payload + Packet header
  - Packet Payload
    - 44 GETs/Packet * 32 Bytes/GET = 1408 Bytes
  - Packet Headers: Ethernet + IP + UDP + OCSM
    - 18 + 20 + 8 + 8 = 58 Bytes/Packet Header
  - Total Packet Size = 1408 Bytes + 58 Bytes = 1466 Bytes < standard MTU

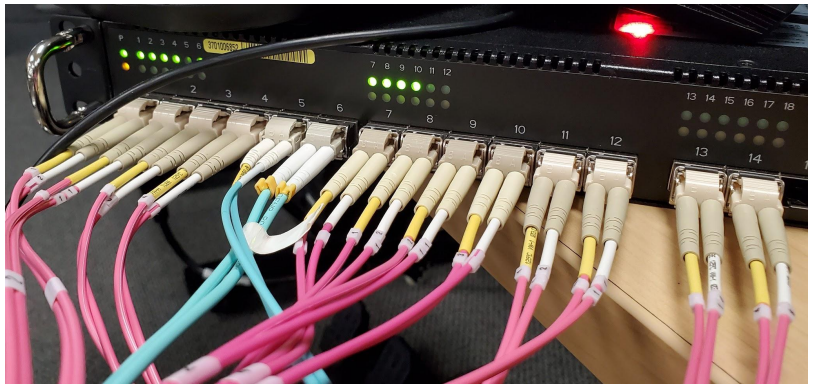
    - Amortized: 1466/44=33.3 Bytes/GET
- Throughput per port, FPGA card, 3 FPGA cards, and total
  - Each SFP+ port on U50 FPGA card
    - (10 Gbps)/((8 bits/Byte)*33.3 Bytes/GET = **37.5 M GET/s per SFP+ port**
  - Each U50 FPGA Card: 37.5M * 4 ports/card = **150M GET/s per FPGA card**
  - Each 1U Server with 3 U50 Cards FPGA cards: 150M * 3 = **450M GET/s per server**

See Also: Algo-Logic GDN Search (Key Value Store)

Key Outcomes

Total Throughput in 1U Rackmount Server

- 3*150M IOPs from FPGA Key Value Store
  - Implemented on 3 Xilinx ALVEO U50 Cards
  - Each U50 card fits in a Half-High PCIe slot.
  - Connected with 4 * 10 Gigabit Ethernet Ports



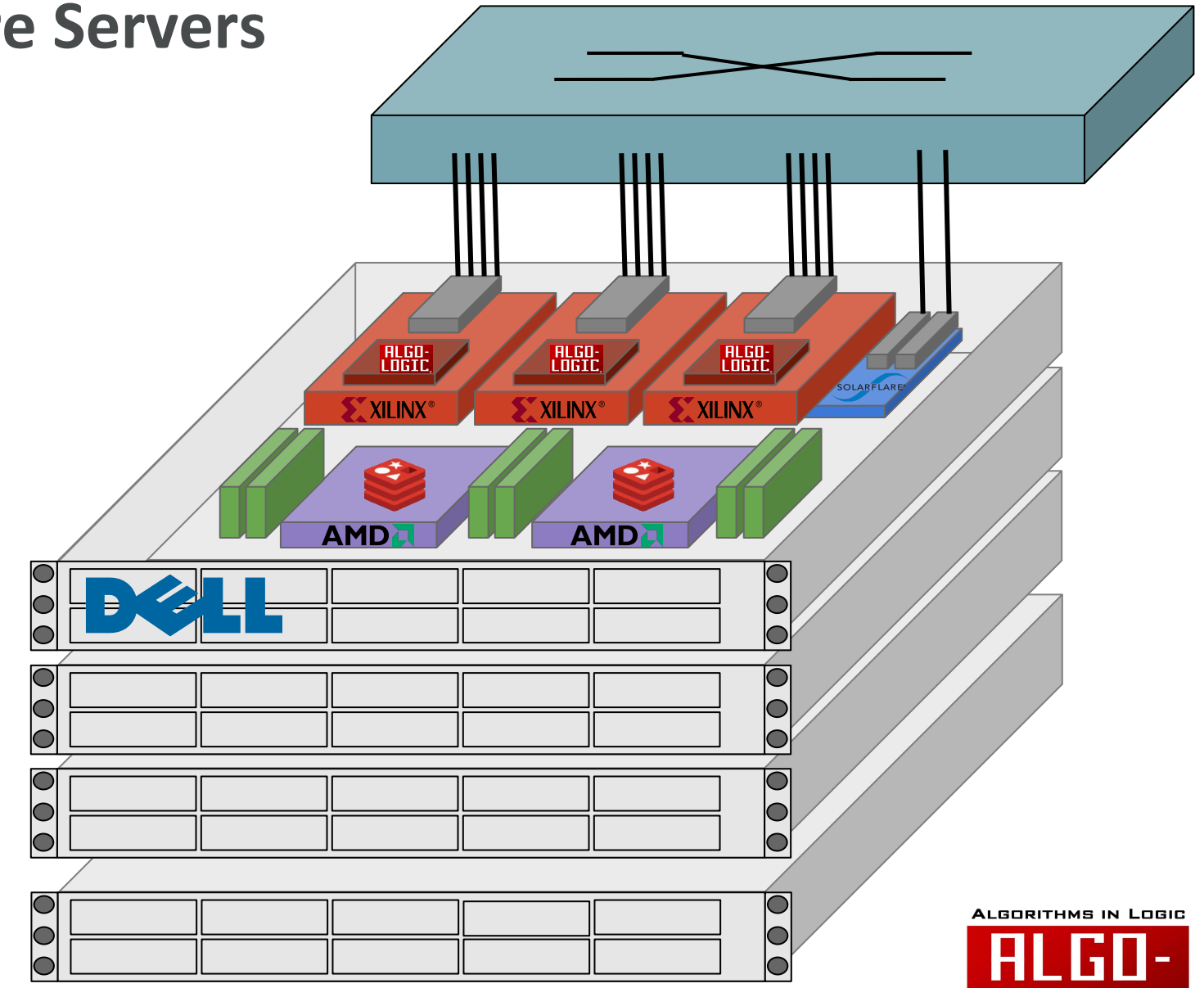
- 2*20M IOPs from Redis in Software on Dell AMD Server
  - Using Dual-port Solarflare NIC on Mezzanine card
  - Each Mezzanine card has 2 * 25 Gigabit Ethernet

- Combined
  - 1U server provides
  - 450M + 40M = 490M IOPs
  - 1.75” Tall and 19” wide

ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

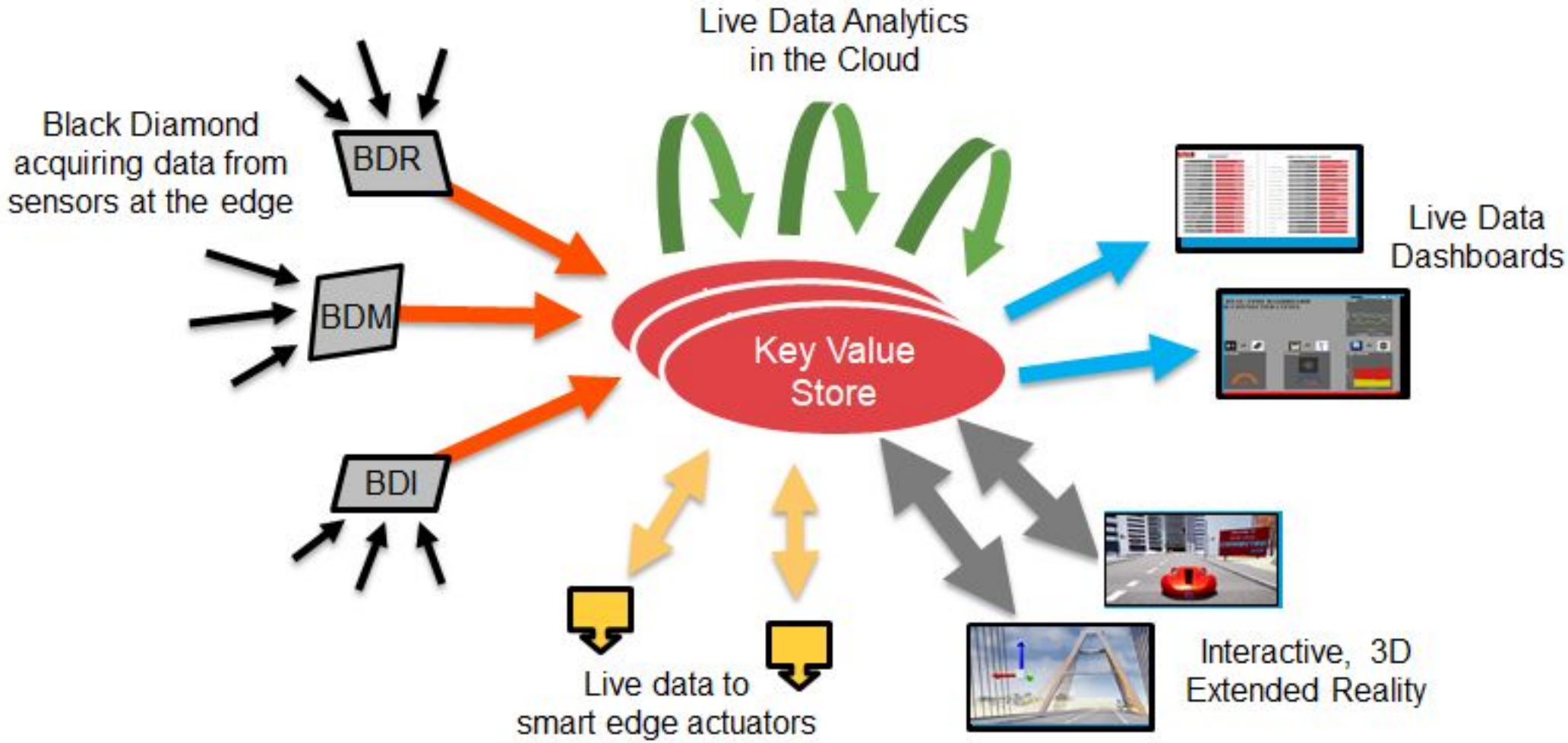
Scaling the Key Value Store Servers

- **A single 1U Rack Server**
  - 490M IOPs ~= Half Billion IOPS
  - 12*10G+ 2*25G = 170 Gbps/server
  - 2 CPUs, 3 FPGAs, 1 NIC
- **Scaling out with 2 Servers**
  - 1B IOPs in 2U of space
- **Scaling up to 25G/port on all 14 ports**
  - 12*25G + 2*25G = 350 Gbps/server
  - 1.165G IOPs in 1U of space
- **Scaling up and out to 1 Rack of 40 Servers**
  - 46B IOPs in 40U (1 rack)
  - 80 CPUs, 120 FPGAs, 40 NICs
  - Network: 350Gbps*40 = 14 Tbps/rack
- **Scaling to 1 Isle of a datacenter (64 Racks)**
  - **75B IOPs via 896 Tbps (~1 Petabit/sec)**
  - Using 2,560 servers with 5,120 CPUs, 7,680 FPGAs and 2,560 NICs, 655TB of RAM



ALGORITHMS IN LOGIC
ALGO-LOGIC
HTTP://ALGO-LOGIC.COM

Utilizing KVS for Real-Time Data Fusion & Analytics



THMS IN LOGIC
GO-LOGIC®
HTTP://ALGO-LOGIC.COM

Conclusions

1. FPGAs massively increase network-attached Key Value Store Throughput
2. Cloud Onload NICs improve client software and Redis server throughput
3. A 1U Dell server with FPGA Accelerators can handle 490M IOPs
4. KVS is being used today for sensor fusion and real-time analytics
5. Algo-Logic's KVS Solution is available today in a preconfigured Dell server or as a monthly service on-line over high-speed VPN

ALGORITHMS IN LOGIC
ALGO-LOGIC®
HTTP://ALGO-LOGIC.COM

Thank You!

http://Algo-Logic.com/kvs

redislabs
HOME OF REDIS