Unlock the Data Center

Andy.Tomlin
@qistor.com

Disrupting Key-Value

Abstract: Key-value storage underpins all large-scale web applications, consuming hundreds of thousands of servers with substantial power and cost implications. As data center trends shift from traditional compute and software-defined solutions to accelerated compute, key workloads like AI and networking are adopting hardware acceleration, but storage lags behind. This talk will address current data center challenges and argue that hardware-accelerated key-value storage offers a superior path compared to computational storage, promising enhanced efficiency and performance.

Speaker: **Andy Tomlin: CEO, Chief Architect, Founder**

Executive at Sandisk, Sandforce, WD, Samsung, Kioxia, 30yrs Storage experience. ~60 patents, many Storage, Flash and SSD products delivered to Market. Start up experienced (Sandforce exec, Agoraic founder) https://www.linkedin.com/in/andytomlin/

History of Storage Devices

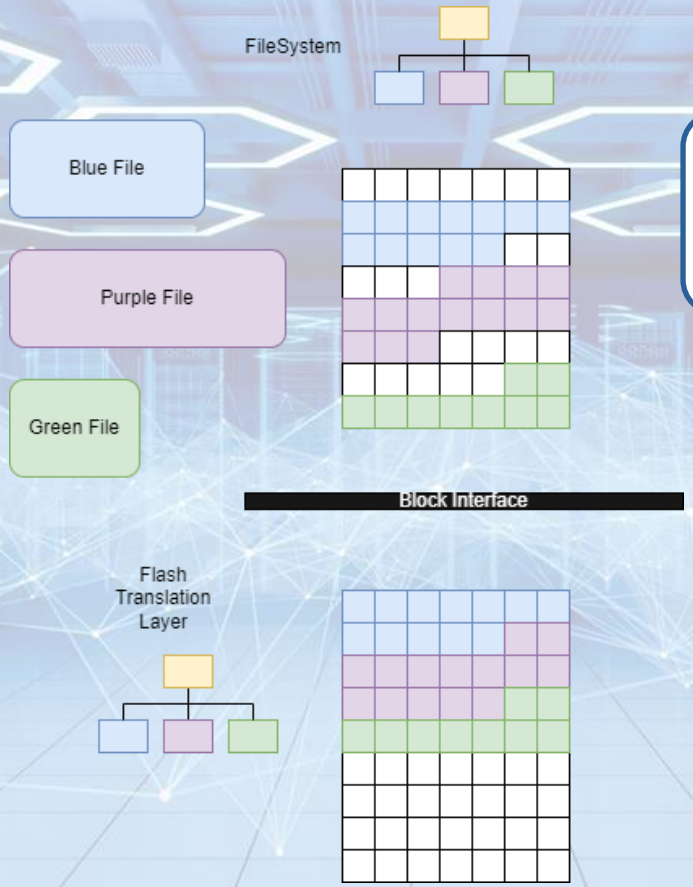
- Since the beginning of Storage, devices have used fixed-capacity blocks
- For over 50 years this has been 512 byte block size (+ some variants)
- However, the things we store are not fixed size, and are either smaller than a block or larger than a block.

Andy.Tomlin
@qistor.com

Mapping systems

The size mismatch is solved by abstractions such as FileSystems and Databases

We're really good at this!



However, these abstractions are the root of the inefficiencies

The layering of mapping systems causes read and write amplification resulting in capacity, power and performance problems

Andy.Tomlin
@qistor.com

Optimal solution

Axiomatic best solution

Perform allocation of space and location tracking in one place

Lowest Level

Perform mapping in Hardware

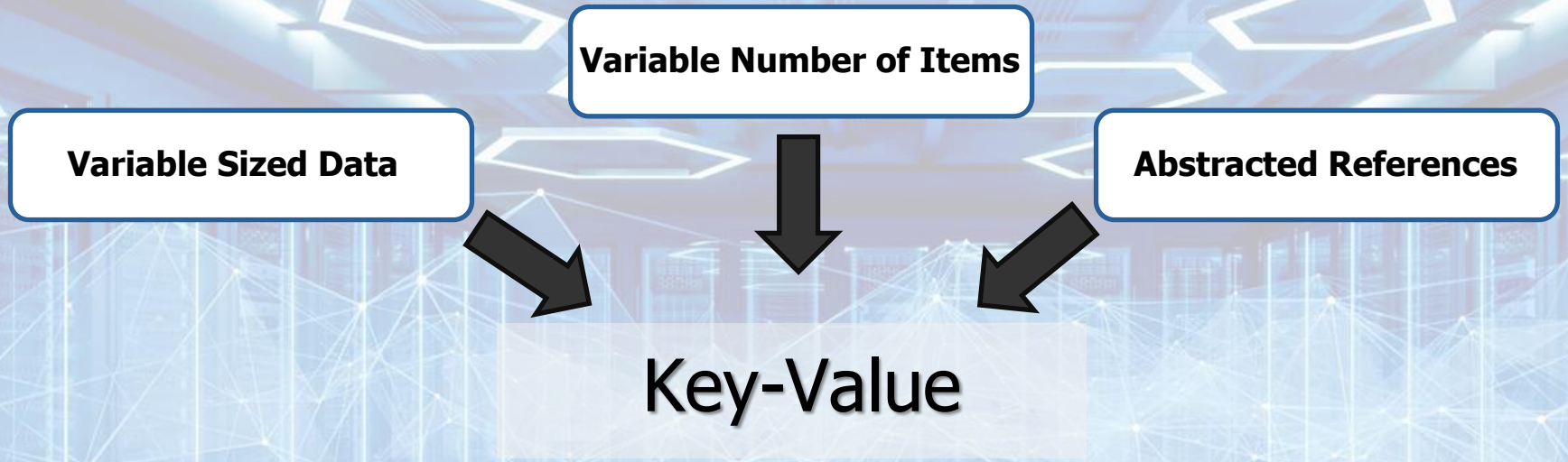
Best Performance

Does not mean eliminating higher level functionality above, just that those systems deal with abstracted locations that don't need to change

Single garbage collection
Single overprovisioning
Variable sized data
Variable number of items

Andy.Tomlin
@qistor.com

Implication of Optimal Solution



- The capability of today's HW means that this problem can be solved in HW if we have the will and creativity to do it

Andy.Tomlin
@qistor.com

A Different Storage Abstraction

What is Key-Value?

An abstract user defined reference to variable length data

Why is this useful?

An application can generate the key from other reference data without needing to tracking it

Andy.Tomlin
@qistor.com

Key-Value Storage Underpins the Web

NETFLIX

slack

Microsoft

All of Twitter's real time storage usage, for Tweets, Users, and DM's is built on top of Key-Value Storage

Most large-scale Web applications today are built on Key-Value Storage

amazon

craigslist

airbnb

Facebook infrastructure is built on top of Key-Value storage.

UBER

ebay

Google

Key-Value is fast, simple, byte addressable

$7B Market, Fastest growing DBMS segment
AI Use cases will only increase growth

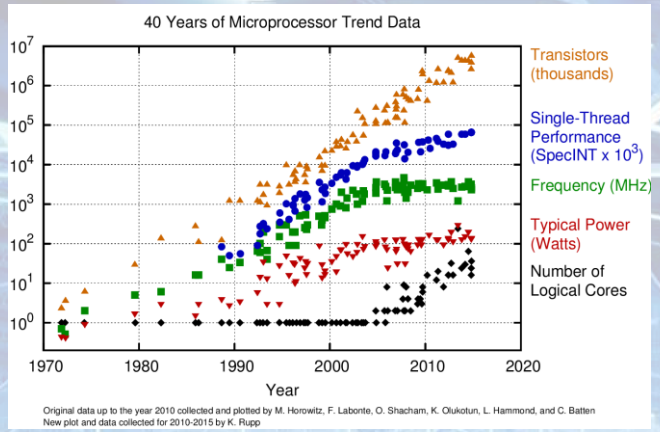
Andy.Tomlin
@qistor.com

Data Centers Must Become More Power Efficient

The problem of Data Center power consumption is penetrating mainstream media.
Data Centers will consume 8% of US power by 2030

Andy.Tomlin
@qistor.com

Data Center Scaling is a fundamental Problem



The 'Software Defined' Era is ending

We cannot just keep adding Servers to scale:

Power, cooling, floor space

Customer & Product dimensions

Customers prefer OPEX vs CAPEX

A Service is simpler for customers to try, purchase deploy, and manage.

QiStor IP & new Architecture enables game changing alternative to CPU based solutions

Scaling Methodology

Frequency → **Cores** → **Accelerated Compute**

Moore's law

Today

Accelerated Compute must replace regular Compute

Andy.Tomlin
@qistor.com

AI Accelerates Demand for Key-Value Performance

Training data, a large quantity of unstructured data consumed with high-performance

AI infrastructure, model storage

Vector Databases running on top of Key-Value stores

Direct, Key-Value store to AI engine is an exciting possibility as a method to reduce CPU training bottlenecks

The entire AI pipeline and management system solution heavily utilizes key-value

Vector Database innovations need key-value storage engines to operate efficiently

AI demand is driving up power usage, putting enormous strain on existing Data Center Infrastructure, requiring new innovations for improved infrastructure efficiency

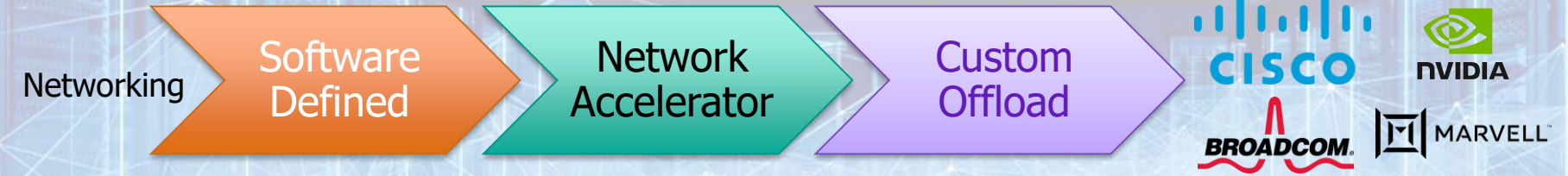
Andy.Tomlin
@qistor.com

Data Center Workloads Shifting to Hardware Offloads

AMD

Google

NVIDIA



CISCO

NVIDIA

BROADCOM

MARVELL

NVIDIA

BROADCOM

PLiOPS
EXTREME DATA PROCESSOR

KALRAY
THE POWER OF MORE

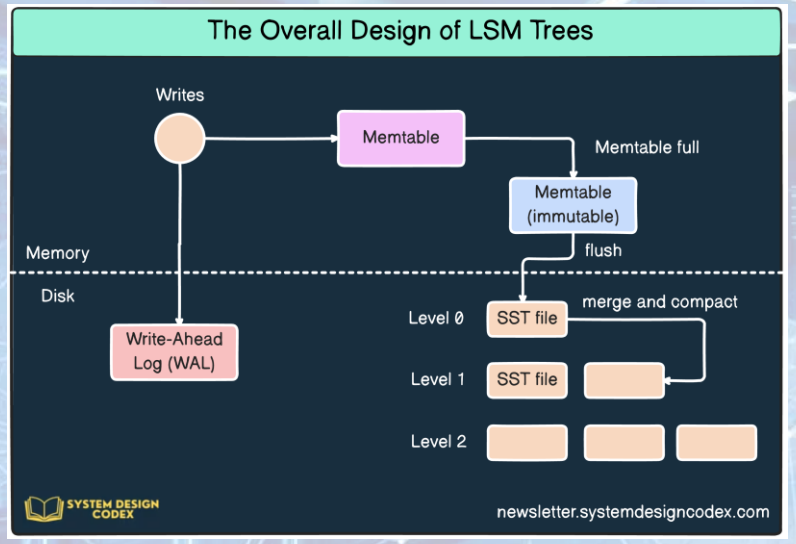
Storage is lagging vs other domains

Compression, security, data management

Andy.Tomlin
@qistor.com

Today's Software Solutions Create High Write-Amp

Log Structure Merge Tree



Sample Config

250TB System
500K read 100K write
1K value size average

Sample Solution

Nodes: 18 x i3en.24xlarge
(96 vCPUs, 768 GB RAM, 58.95 TB storage)
Disk storage: 1.03 PB
Total vCPU: 1,728, Total RAM: 13,824 GB

High write amplification due to WAL and SST level compaction
High read amplification due to levels and file size
L0 – 10MB, L1 – 100MB, L2 – 1GB …

Andy.Tomlin
@qistor.com

Criteria To Move Algorithms to Hardware

- Are the requirements well understood?
- Are they stable?
- Is there sufficient volume to enable commercialization?
- Is the problem important enough to solve?
- Will HW solution bring additional performance & power improvements?

Compression

Encryption

Repetitive Calculation

Algorithms

Andy.Tomlin
@qistor.com

Path to an Optimal Key-Value Store

Do Less Work

QiStor Algorithm

Less reads and writes by:

- No log structured merge
- Eliminate mapping layers

Use Less Flash & Better Performance

Save Money, Power and Time

10-100x more efficient in HW

Replace expensive general purpose compute

With custom QiStor IP

Green solution to scaling KV

Easy to Deploy and Consume

- Customer can try before buy
- No HW purchase
- Utilize existing Data Center infrastructure
- Scale up with demand

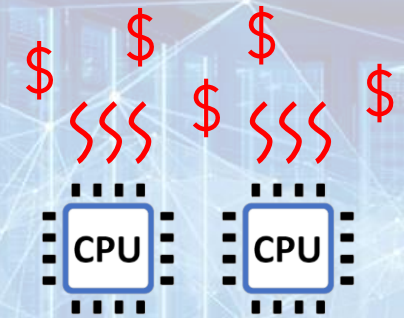
Customer Focused

Andy.Tomlin
@qistor.com

Key Value-as-a-Service

The Problem

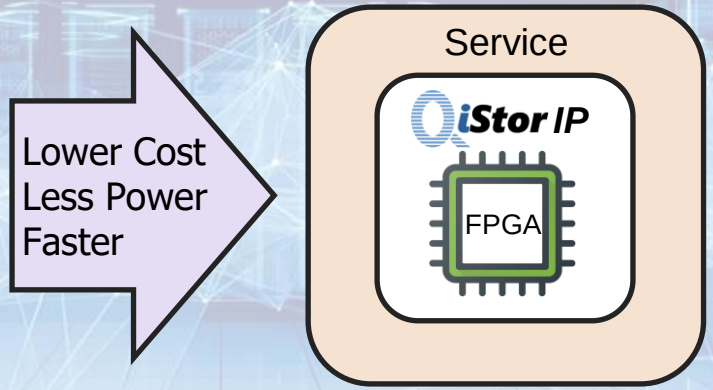
Data Center Demand is growing faster than CPU scaling



Database & Key-Value cost is driven by expensive compute

The Solution

Disruptive HW-accelerated Key-Value Service



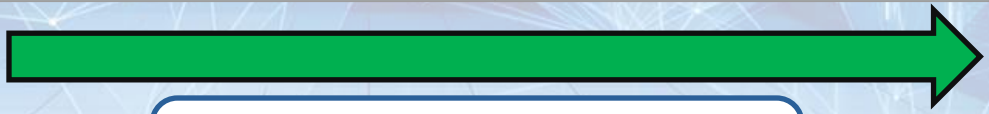
Utilize existing Cloud based FPGA infrastructure

Database Market size >$100B WW, 15% CAGR

Andy.Tomlin
@qistor.com

How Does SW LSM compare to HW acceleration?

	LSM Vendor recommendations	HW accelerated
Storage	50% utilization (100% overprovisioned) (100TB = 200TB Flash)	75% utilization (33% overprovisioned (100TB = 133TB Flash)
DRAM	30 : 1 to Storage	N/A
Servers	1 vCPU per 12K operations	1 FPGA per 750K operations, max 8 FPGA per server



Significant Flash, Server, Memory, Power, Latency, Cost reductions

Andy.Tomlin
@qistor.com

New Offload Architecture has Breakthrough Benefit

AEROSPIKE

	Per server	Total (20 servers)
Capacity	60TB	1200TB
Power[1]	2.91kW	58.2kW
Write (kops)		753
Read (kops)		3017
Latency		1ms
3yr TCO		$4.17M

https://aerospike.com/lp/y-running-operational-workloads/

Comparison of 1PB Solutions

New Architecture

QiStor

	Per server	Total (2 servers)
Capacity	Disaggregated	1000TB
Power[1]	2.91kW	5.8kW
Write (kops)[2]		1500
Read (kops)		6000
Latency		250uS
3yr TCO		~$1M

Key Takeaways

10x - Less Hardware

10x - Less Power

10x - TCO Reduction

higher performance

Better Reliability

CO_2 *savings* of 325 Metric Tons/yr

1.Based on EPA power calculator, AWS outpost Data. Just rack power, does not include cooling power savings

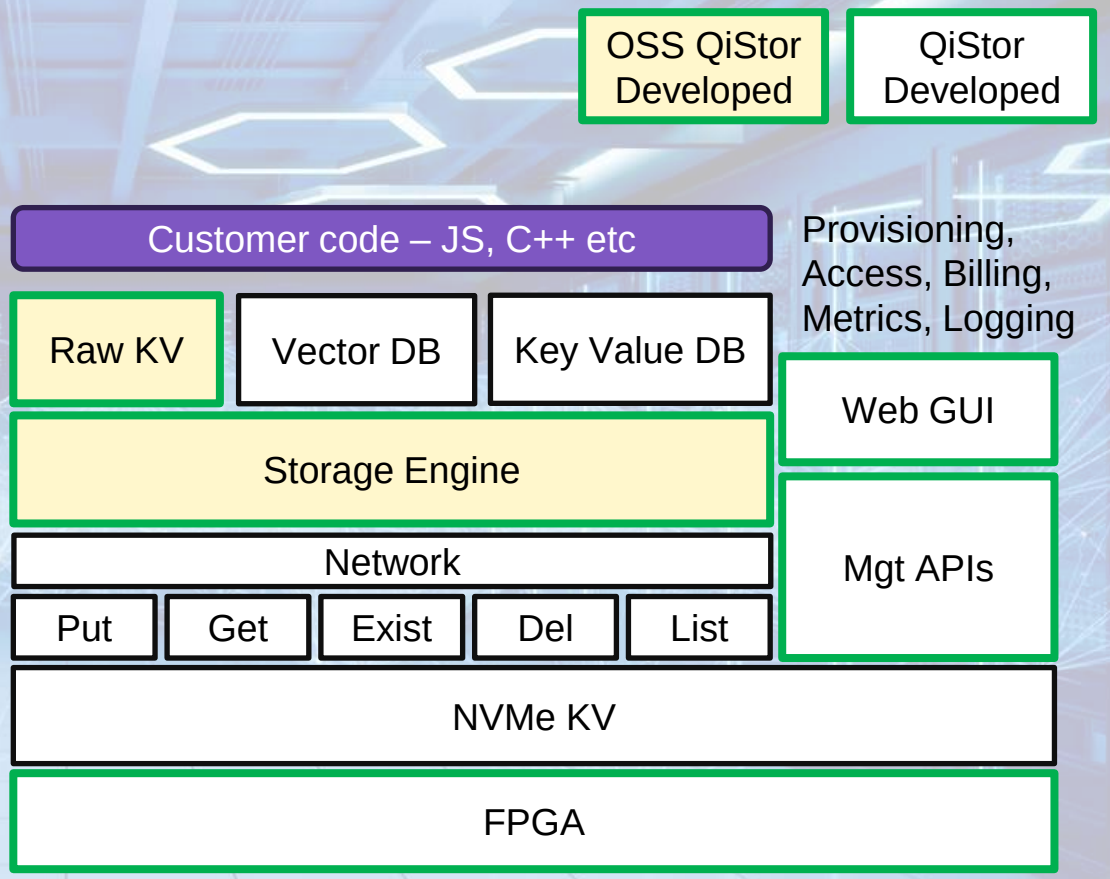
2.QiStor numbers based on model – 4 FPGA config + drive power estimate for Disaggregated

Andy.Tomlin
@qistor.com

Solution Stack

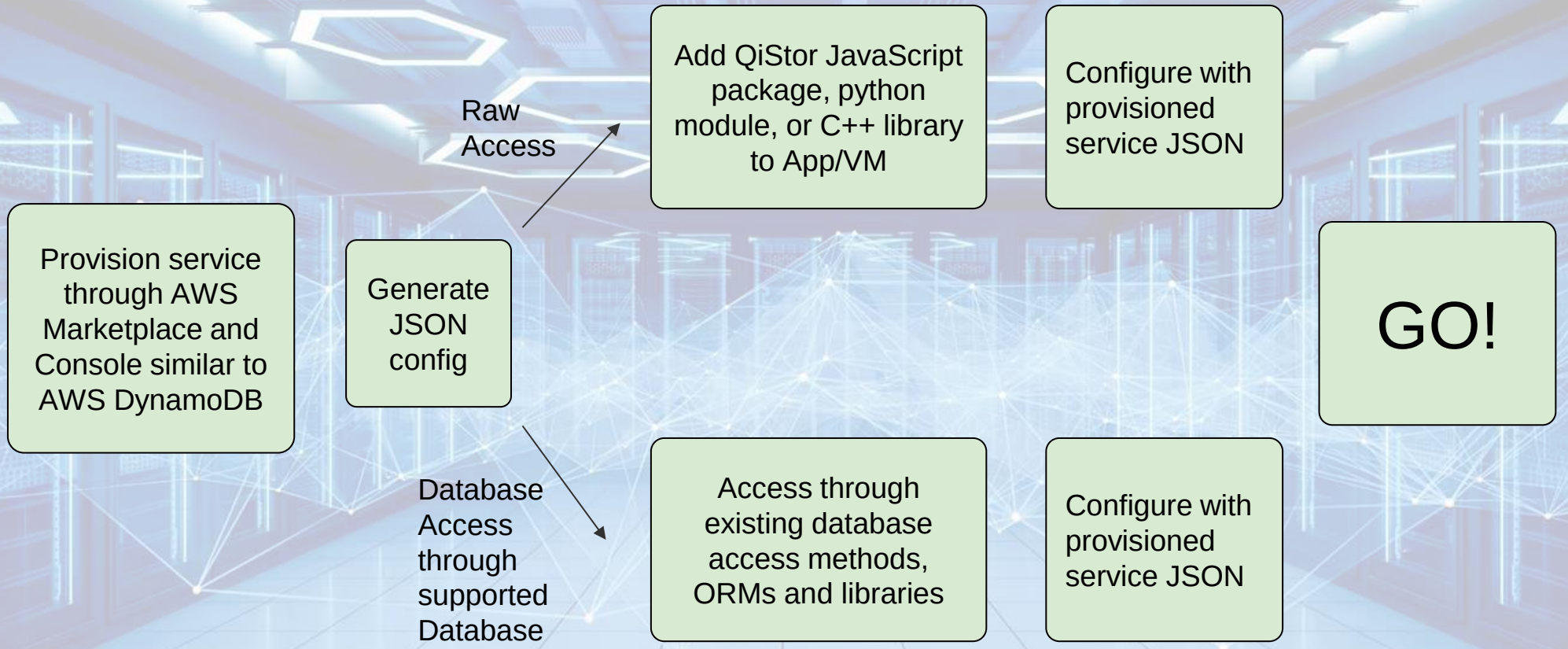
Key Goals

- Avoid proprietary interfaces
- Open-source Storage Engine
- Integrate with existing Databases and ORM's to minimize customer code changes
- NVMe KV interface via SPDK/DPDK



Andy.Tomlin
@qistor.com

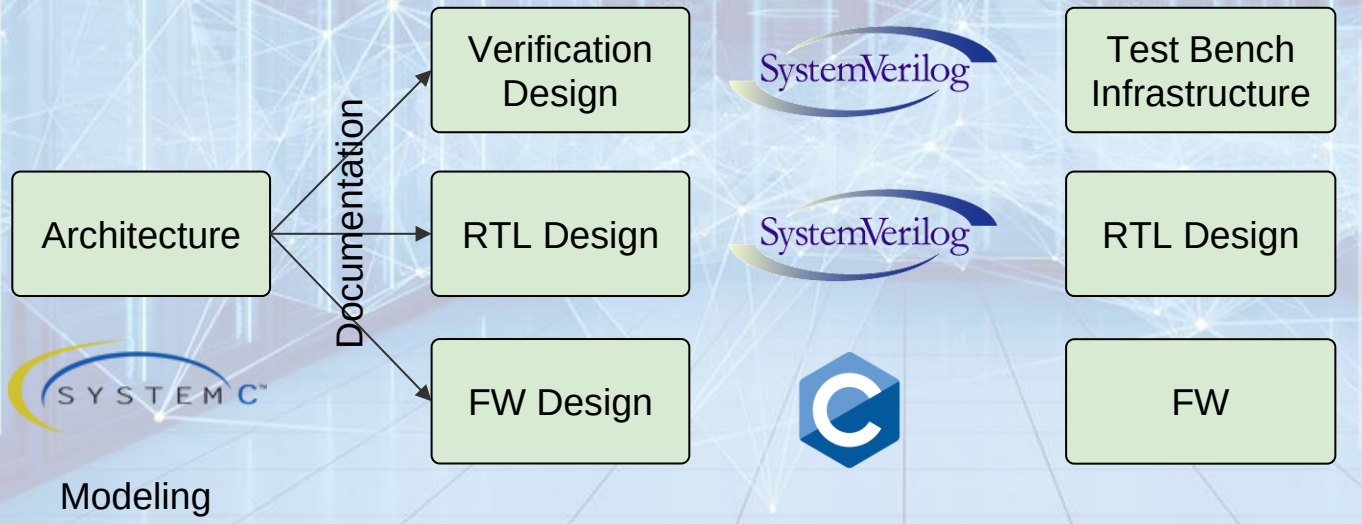
Customer Flow



Andy.Tomlin
@qistor.com

Key Technical Challenges

- Developing complex Algorithms in HW (FPGA/ASIC)
- Typically takes large team 40-100 people including Architecture, Design, Verification, Firmware



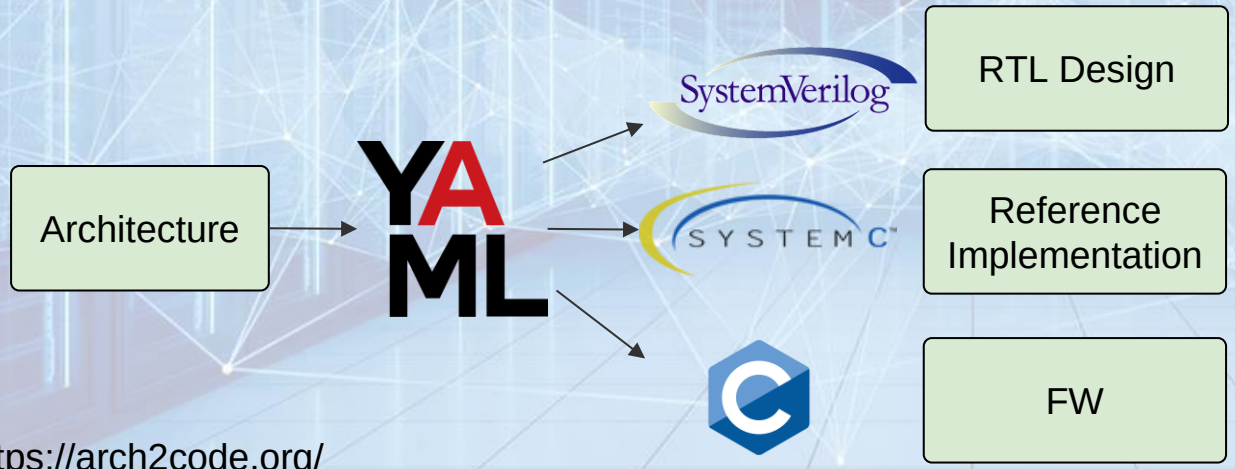
Complexity

- Communication
- Keeping everything in synch, documentation
- Change management

Andy.Tomlin
@qistor.com

Arch2Code

- Describe architecture in YAML
  - Define all design elements: Interfaces, Memories, Registers, Functional Decomposition
  - Generate all Headers, Hierarchy, Implementation plumbing
- Create complete Reference SW implementation (SystemC)
- Port implementation from SystemC to SystemVerilog



- Documentation minimized
- No Synchronization effort
- Minimal Developers
- Faster Debug

https://arch2code.org/

Andy.Tomlin
@qistor.com

Arch2Code - Testing

- SystemC Reference Implementation is also the Test Harness
- Any part of the hierarchy can be replaced with RTL
  - Reference implementation provides test environment
- Any part of the hierarchy can be replaced with Paired RTL / Model self checking block
  - Significantly faster debug of RTL as expected response is provided by reference implementation


- QiStor team has proved methodology
  - Datapath width changes done in 1 day
  - RTL Design mapped to FPGA implementation in <1 week
  - AWS-F1 FPGA Demo with NVMe KV bring-up with no HW bugs
  - Debug of RTL much faster

Andy.Tomlin
@qistor.com

Summary

Block Storage Fundamental Issues are not going away

Why HW Accelerated Database/Key-Value Service

Scaling	Moore's law no longer meeting demand growth with SW only solutions
Performance	HW Acceleration provides clear benefits
Complexity	Customers want the benefit of HW acceleration without complexity & HW purchase

Why Is It Needed Now

Infra Structure	FPGA's are available today in Data Centers at a maturity and pricepoint
Data Center Roadmap	No other solution for DC Power, Cooling & Footprint
Customer	Customers limited by cost of existing solution

About QiStor

We are tackling these tough problems today.

Come see our demo!

Fund Raising New Round Now

Andy.Tomlin
@qistor.com