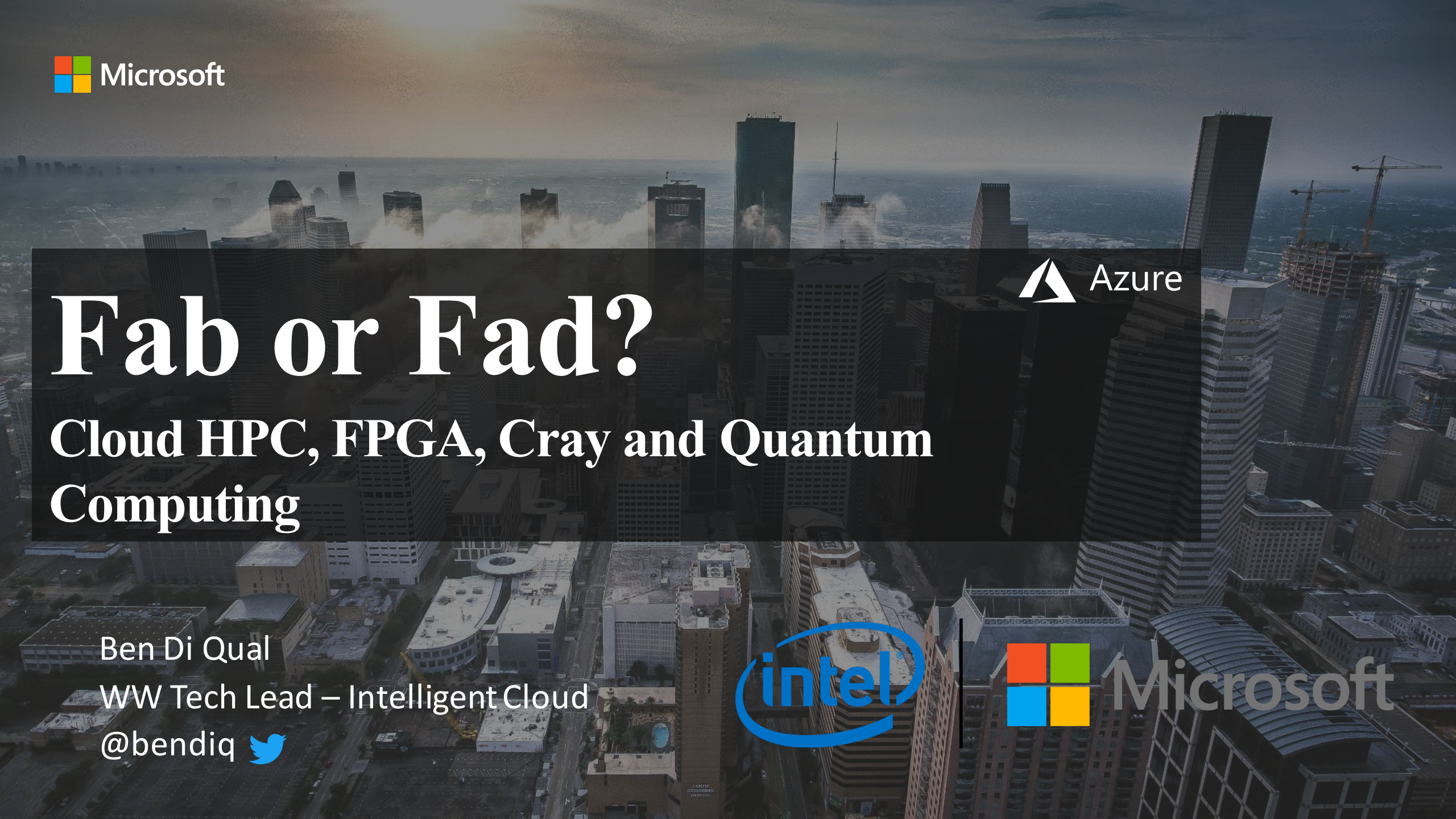

Microsoft
Azure
Fab or Fad?
Cloud HPC, FPGA, Cray and Quantum Computing
Ben Di Qual
WW Tech Lead – Intelligent Cloud
@bendiq
intel
Microsoft


Who is using the cloud today?
Who runs ALL their simulations in the cloud?!

Really? True HPC on the cloud?

From a 3rd party, independent research published on 9, February, 2017 :

Comparative benchmarking of cloud computing vendors with High Performance Linpack

Mohammad Mohammadi, Timur Bazhirov, Exabyte Inc., San Francisco, California 94103, USA

"We found **Microsoft Azure to deliver the best results**, and demonstrated that the performance per single computing core on public cloud to be **comparable to modern traditional supercomputing systems**. Based on our findings we suggest that the concept **of high performance computing in the cloud is ready for a widespread adoption** and can provide a viable and **cost-efficient alternative to capital-intensive onpremises** hardware deployments."

The media's reaction was quick:

"**The actual TOP500 run across the entire machine achieved a Linpack yield of more than 64 percent of peak**, which is fairly typically of an HPC cluster with a high-performance network. The **Azure H16 in this test had a 67 percent Linpack yield**."
-Top500.org

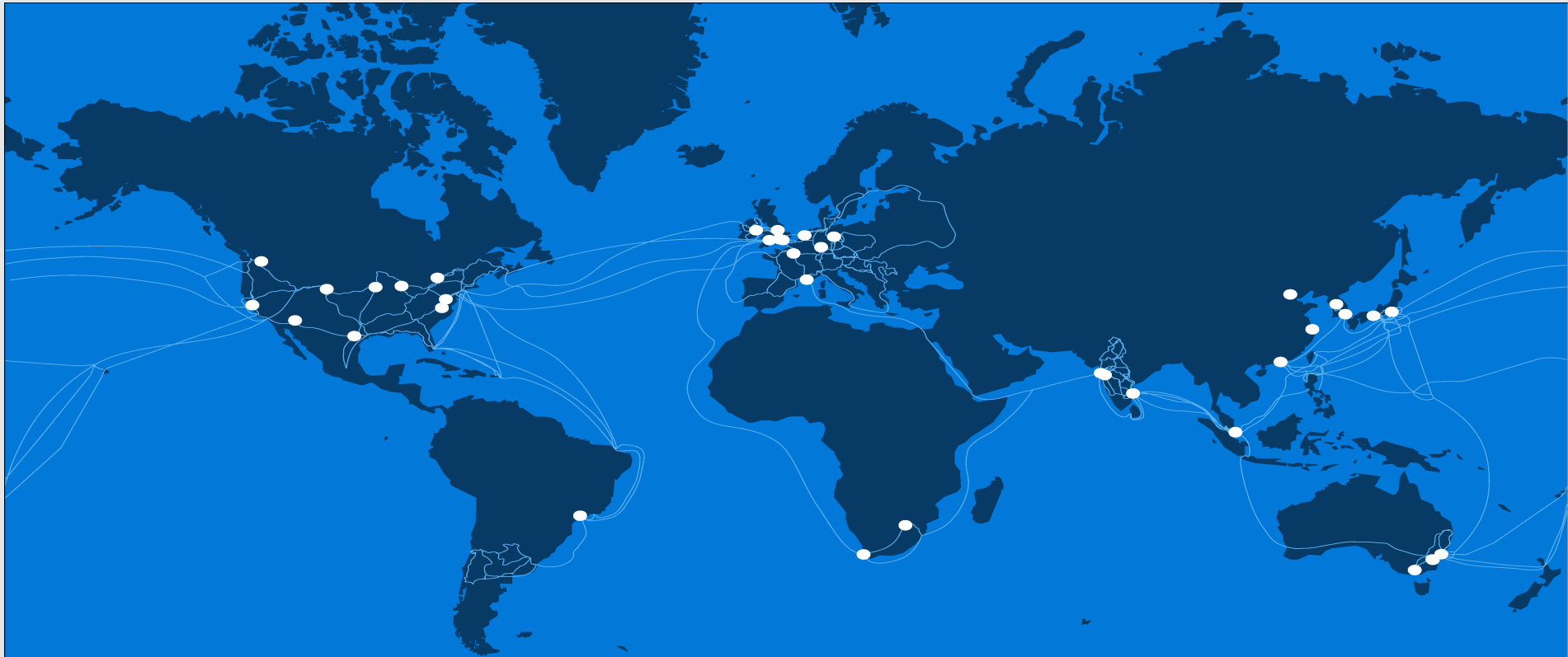
erformed best, with
ode [...] The **Edison**
to 32 nodes."
-The Register

-The Stack

ranked su
-Fudzilla

Growth

- 76% revenue growth year/year
- Double compute utilization year/year
- 46 regions operational, 8 announced, expanding fast
  - 4500 peering locations
  - 130+ edge sites



Azure Scale Momentum

1 TRILLION
Cosmos DB Rus/day

500% YoY growth

13 MILLION
AAD organizations

>16 TRILLION
Event Hubs events/week

>305 TRILLION
Storage objects

50% of VMs
are Linux VMs

1 TRILLION
Azure DB requests/day

https://cloudblogs.microsoft.com/quantum/

Microsoft | Microsoft 365 | Azure | Office 365 | Dynamics 365 | SQL | Windows 10

Microsoft Quantum

Blog Posts

Simulators are key for developing real-world quantum solutions >

March 4, 2019

By Microsoft Quantum Team

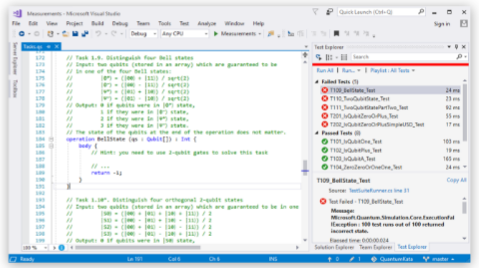
To develop, test, debug and optimize quantum programs, we can work around this hurdle using quantum

Announcing the Microsoft Quantum Network >

February 28, 2019

By Microsoft Quantum Team

Creating a scalable quantum computer will require the collective effort of many skilled and diverse



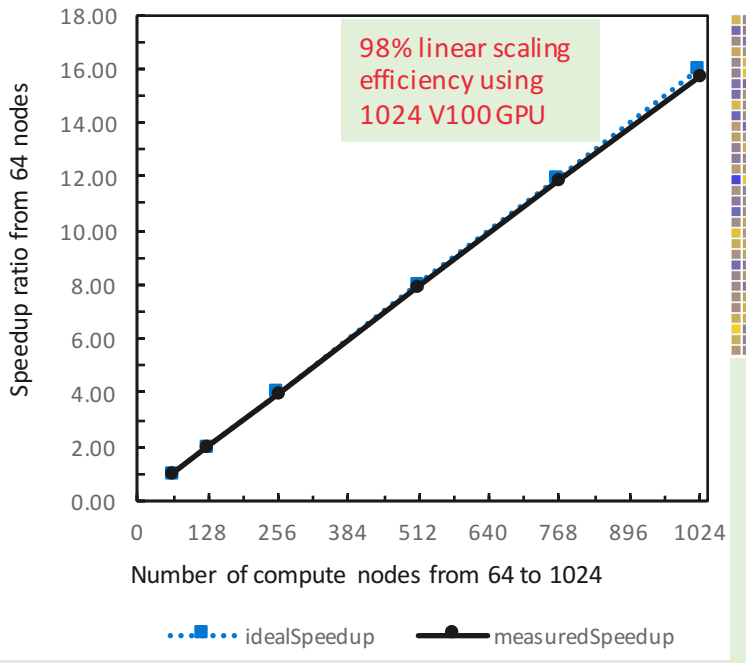
Test your quantum programming skills in the Microsoft Q# Coding Contest – Winter 2019 >

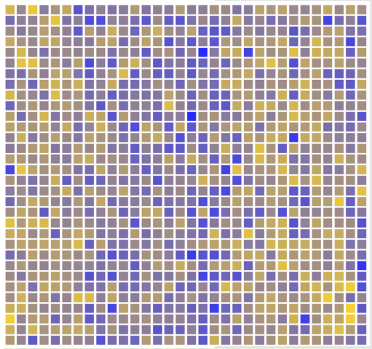
February 7, 2019

By Microsoft Quantum Team

Petaflop Cloud Supercomputing for Terapixel Visualization of a Digital Twin

Newcastle University

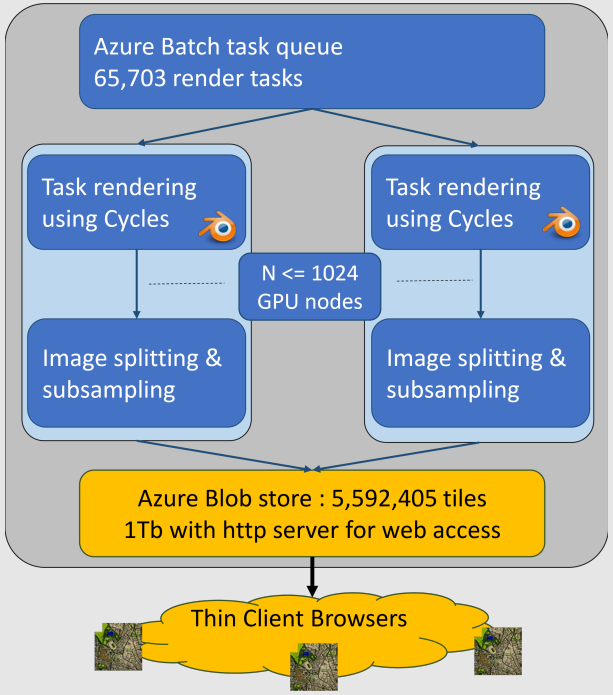




Each tile is one NVIDIA V100 in the 1024 GPU run.

Colours are total task time:

- higher than median
- lower than median



Visualization results:

the first terapixel visualization of an urban digital twin, photo-realistic rendering using Blender Cycles path tracing.
tiled image output format makes supercomputer visualization outputs accessible on low cost, thin client devices.
rendering time reduced from an estimated 34 days on one V100 GPU to 48 min. on 1024 V100 GPU.

Visual supercomputing results:

efficiently deployed a 1024 GPU, 14 PFlop cloud supercomputer: faster than any public GPU HPC system in the UK.
cost for one performance scaling graph: <£20,000, providing access to a > £10 million supercomputer.
Azure upgrade (from K80 to V100) during the project, transparently benefited from a 3x performance increase.
enables experiments with future scaled performance approximately 20 years ahead of current workstation systems.

Nick Holliman[1], Manu Antony[1], James Charleton[2], Stephen Dowsland[1], Phil James[1], Mark Turner[1]
[1]Newcastle University, [2]Northumbria University

The Alan Turing Institute
Microsoft Azure
urban observatory newcastle
Newcastle City Council
National Innovation Centre for Data
SIEMENS
Northumbria University NEWCASTLE
The Bill Douglas CINEMA MUSEUM
EPSRC Engineering and Physical Sciences Research Council
UKCRIC

Local Simulator

- Simulate a 30 qubit quantum computer
- Integrated into Visual Studio and VS Code
- Full debugging support

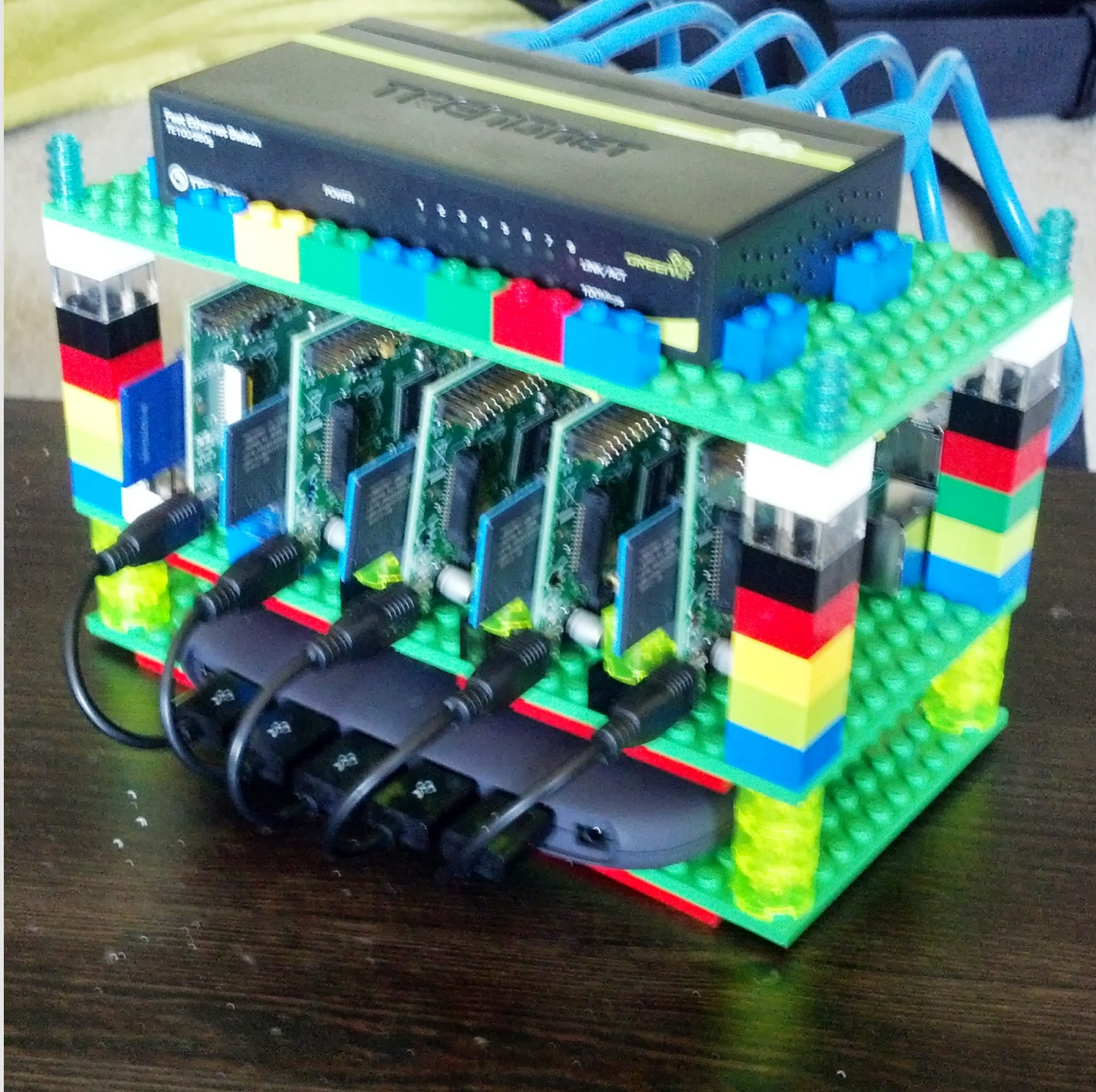
Azure Simulator

- Available for quantum solutions needing over 40 qubit simulation

What are the challenges we see in the cloud?

- Storage
  - Cost
  - Scale
  - Multi Protocol
  - Management
- 20k+ core MPI jobs
- Management
  - Cost
  - Governance and Security
  - Skills ramp ups

Ways to deal with cloud limitations…



Cloud + Cray – Rivers of Data – Not Islands of Compute

Combining Azure + Cray capabilities: agility, scale, price performance

Networking

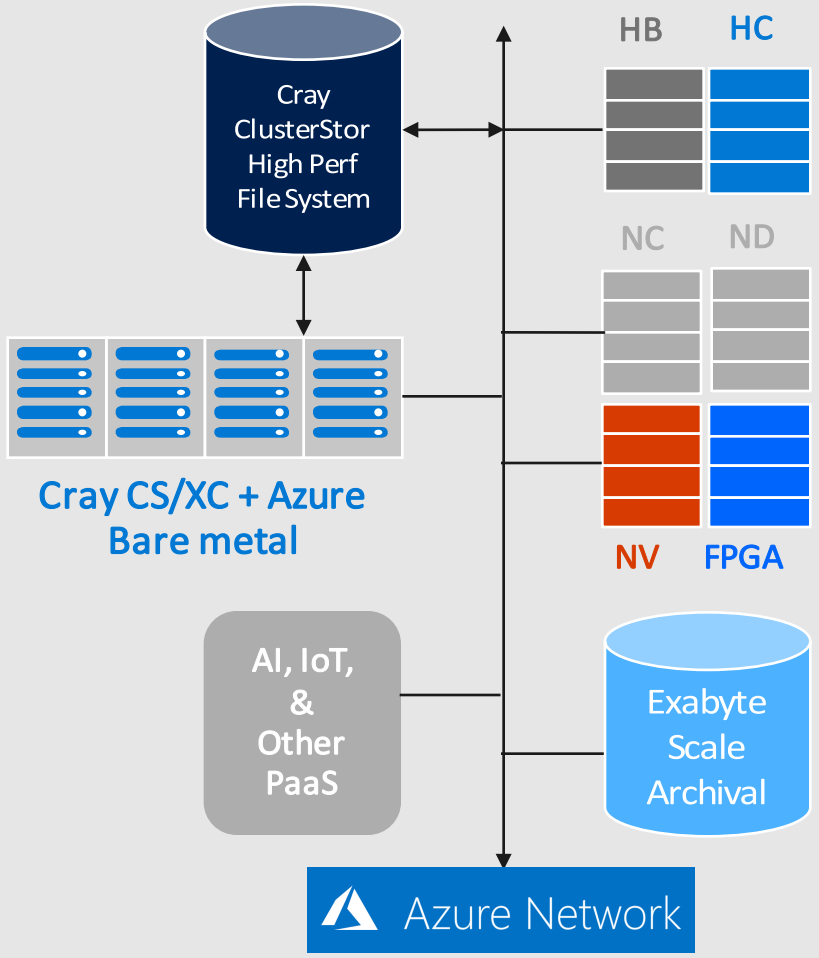
- A common, secured, and global fabric for the data-driven energy enterprise
- Azure = Exabit global network
- RDMA roadmap and NG Cray network architecture

Storage

- Move enormous data at high-speed to powerful, heterogenous, and collaborative compute environments
- Object-store scale out ≠ filesystem scale-up
- ClusterStor solutions optimized for your workflows
- buying power for innovative technologies to brings down cost per stored or moved bit

Compute

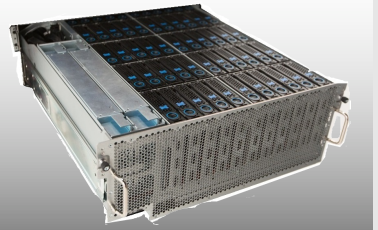
- Help customers innovate and reduce risk, despite End of Moore's Law
- Azure deploying new hardware every year and 13 megawatts every 30 days
- Multiple silicon bets -DRAM v. HBM; AMD, Intel, ARM, Nvidia, FPGA, GraphCore, Quantum
- Fast technology deployment schedule to help HPC buyer maximize benefit



Where is Cloud storage going?

- Limitless* storage
- Multi protocol to same data set
  - REST API**
  - HDFS**
  - SMB
  - NFS
- NVMe and Ultra low latency/High IO performance
- Automatic data placement – Ultra/High/Standard/Archive

Azure Storage Hardware Innovations

2010 - 2012	2012 - 2014	2014 - 2016	2016 - 2018	2018 +
Gen 1 / Gen 2	Generation 3	Generation 4	Generation 5	Generation 6
17 Regions	24 Regions	43 Regions	48 Regions	54 Regions
				
Standard Storage HDD	Cost Optimized Storage	Premium Storage SSD	Archive Storage	Microsoft HW
Purpose built Cloud Storage Hardware	40G Network Hot-swap HDD Dedicated Backend Network	SSD Caching FPGA	NVME SSD Multi Stream SSD SMR HDD RDMA	Secure Platform Hot-swap SSD EDSFF – L SSD Split-Actuator HDD

High Throughput Block Blob

Block Blob write throughput of 100Gbps+

Up from 60 MB/s

Instantaneous write throughput

No ramp up

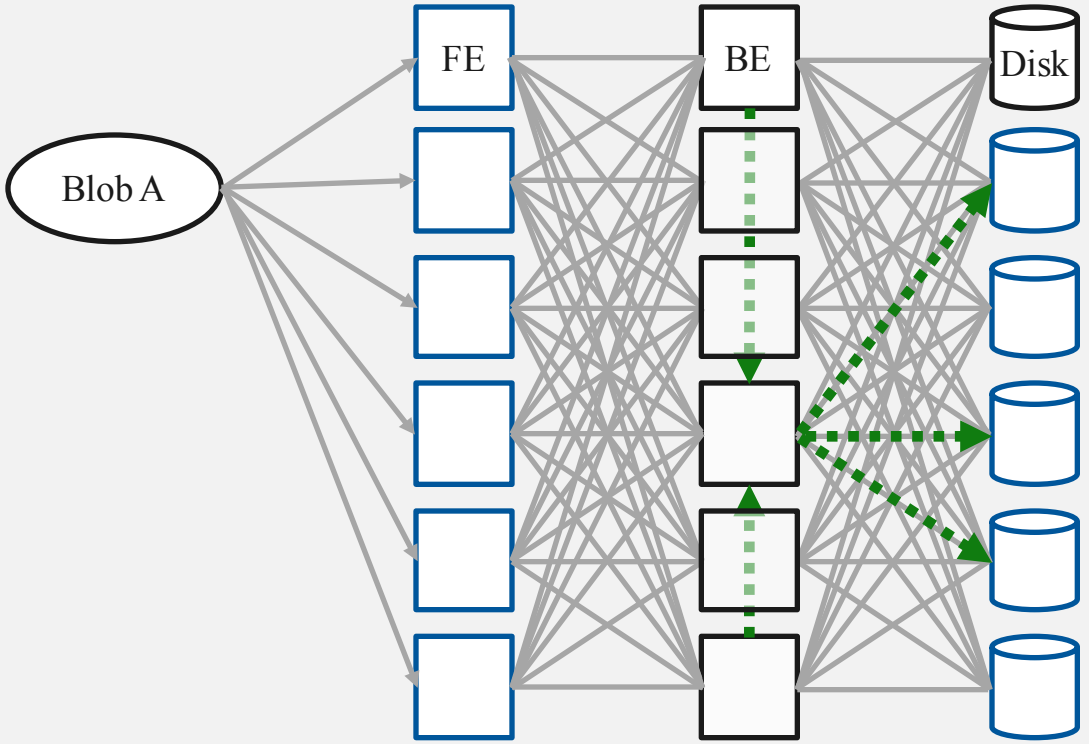
No API changes or client-side changes needed

Ingress throughput become automatically available for block blobs

Active with Put Block or Put Blob operations

>4MB for standard blob

>256KB for premium blob



Where next?

- Learn how to use the cloud https://docs.microsoft.com/en-us/learn/
- Recipes for HPC on the cloud https://github.com/az-cat/az-hpcapps
- R Studio with a cloud back end https://github.com/Azure/doAzureParallel
- Run container images and experiments https://github.com/Azure/batch-shipyard
- Run any ML framework on the cloud https://github.com/Azure/BatchAI

Questions?