

- Azure at a Glance
- Why NVMe?
- Issues at Scale
 - Form factors
 - Need to allow for “rot in place”
 - Need for remote debugging
 - Need for security



2 Million

miles
intra-datacenter fiber

72+

Tb per second
backbone

54

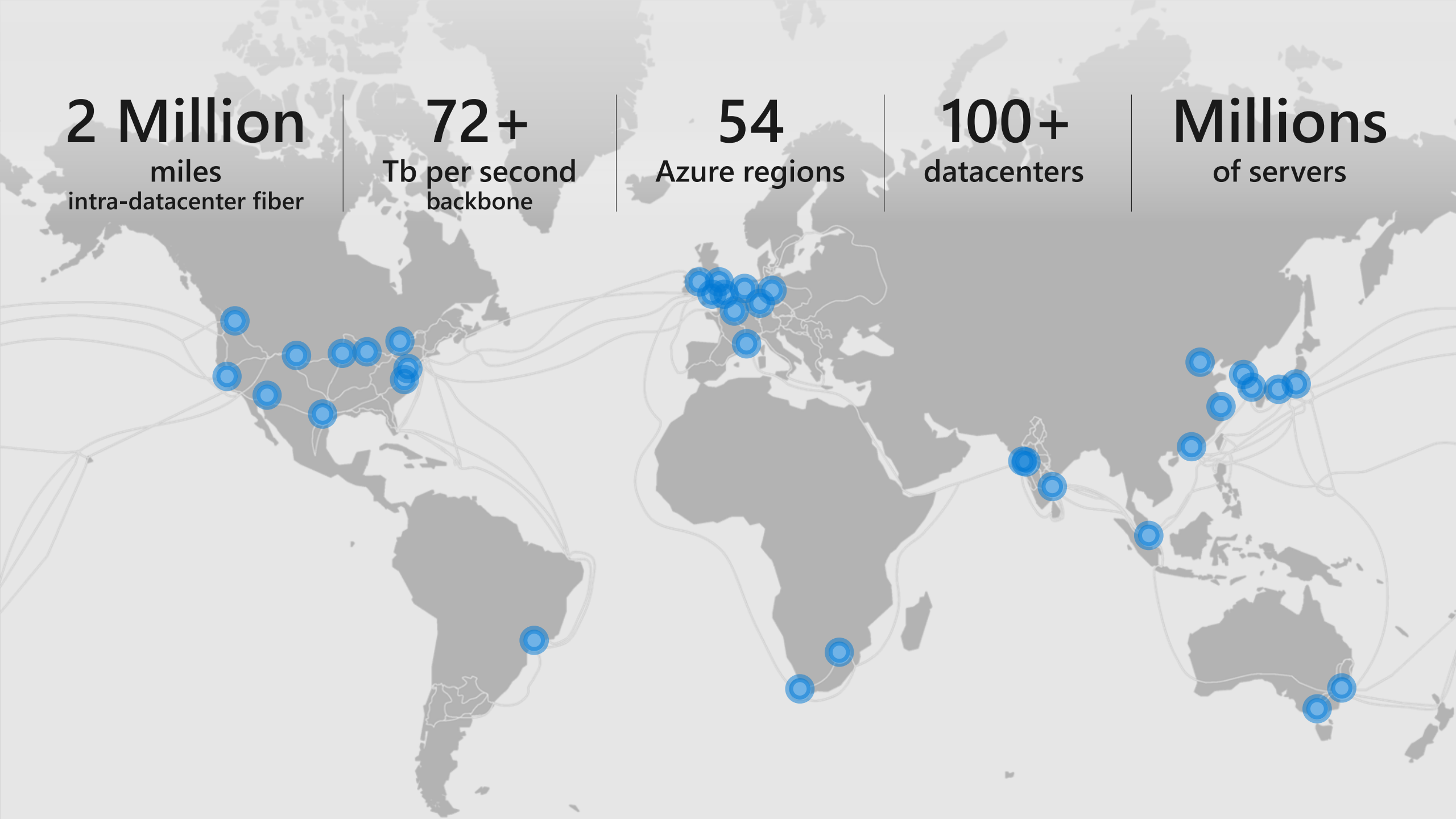
Azure regions

100+

datacenters

Millions

of servers



2+ Million

miles
intra-datacenter fiber

72+ +

Tb per second
backbone

60

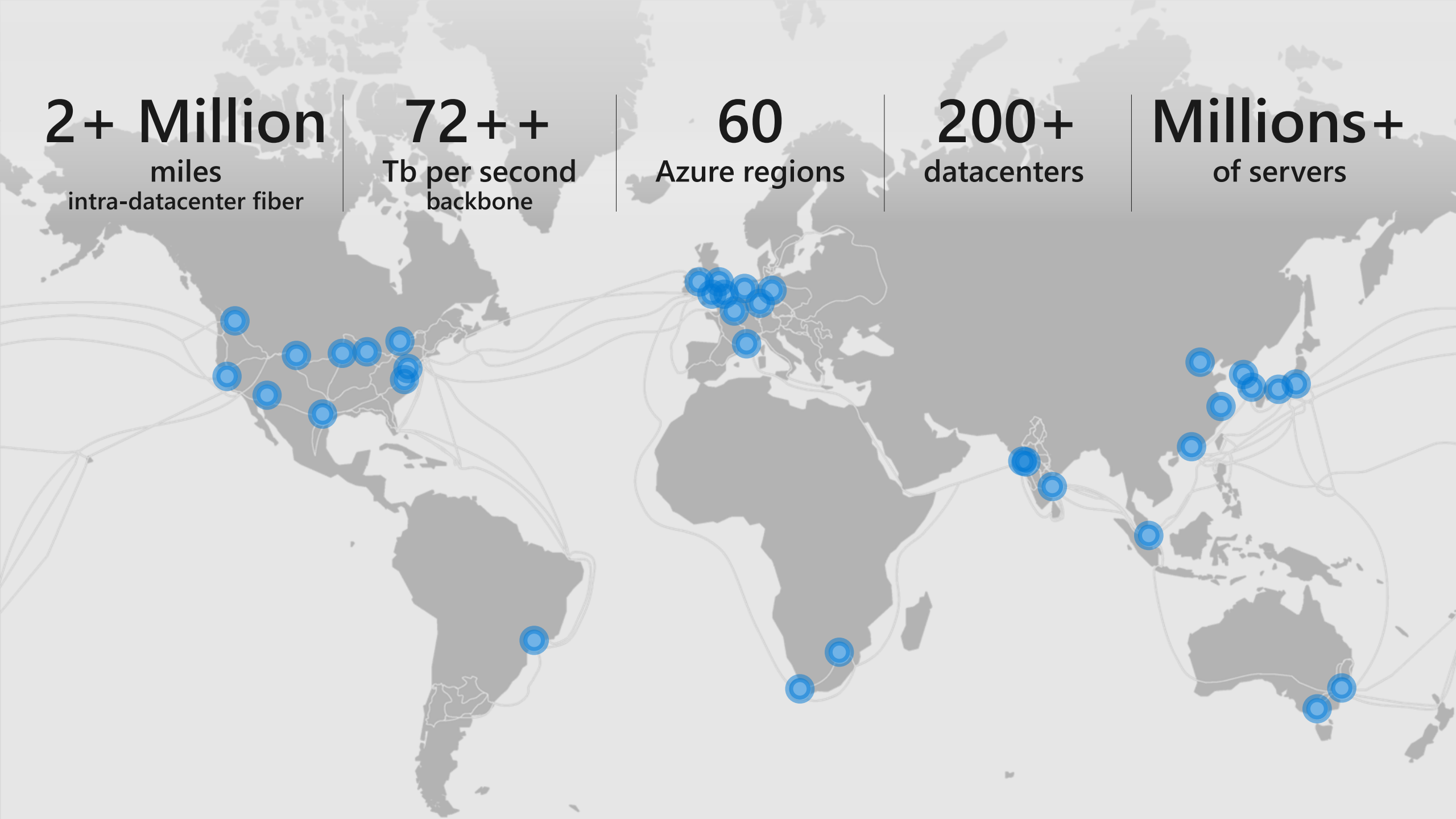
Azure regions

200+

datacenters

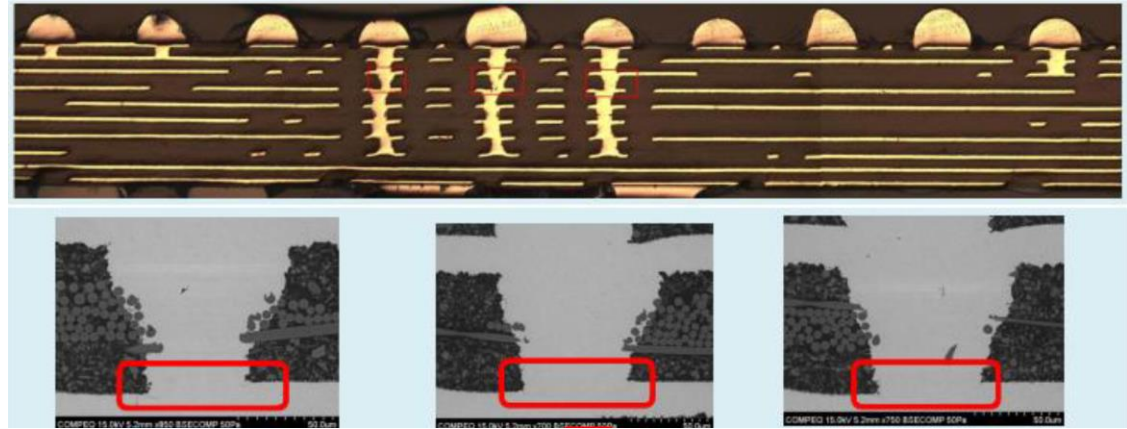
Millions+

of servers



Issues at Scale – Form Factors

- m.2 has run its course
 - Power and thermal constraints
 - Fragile PCB and connector
 - Not hot-swappable
- E1.L and E1.S are here to replace it
 - Built from the ground up for datacenter use cases
- **Good news is that they support NVMe too!**



E1.L (SFF-TA-1007)

- Density Optimized
- 318.75 x 38.4 mm
- Supports > 40W
- Up to 48 Standard NAND sites



E1.S (SFF-TA-1006)

- 111.5 x 31.5 mm
- Up to 12 Standard NAND sites
- Supports >12W



E3 (SFF-TA-1008)

- Ultra high-performance applications
- (104.9/142.2) x 78mm
- Supports up to 70W
- Up to 48 Standard NAND sites



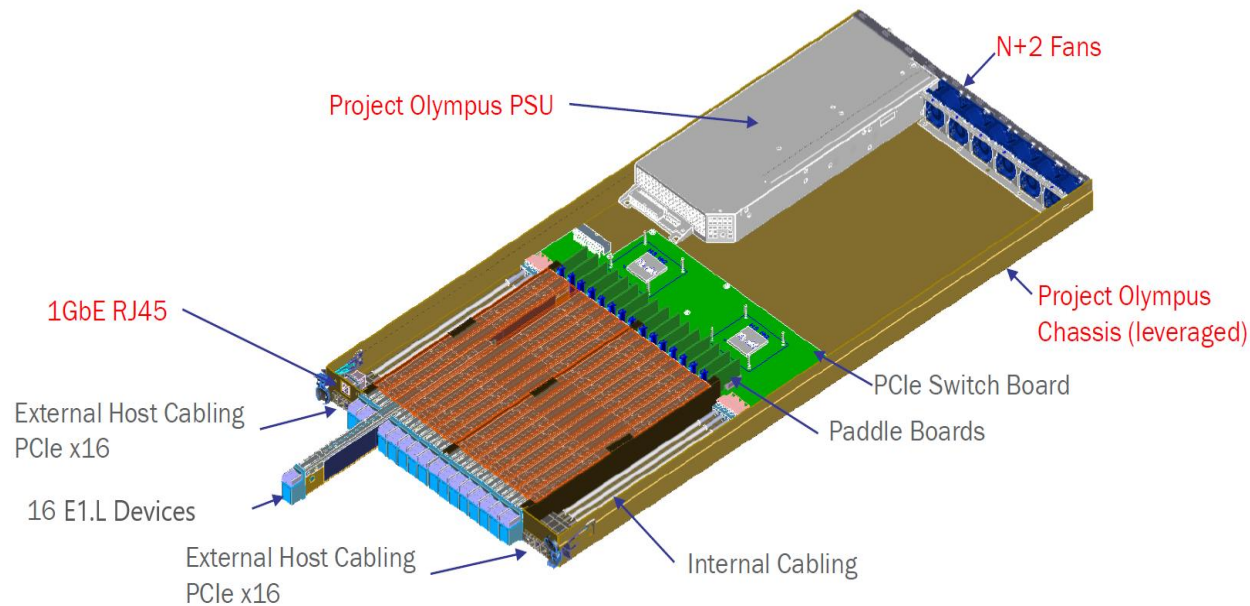
EDSFF Advantages

- Same Protocol: NVMe
- Same Interface: PCIe
- Same Connector: SFF-TA-1002
- Same Pinout and Functions

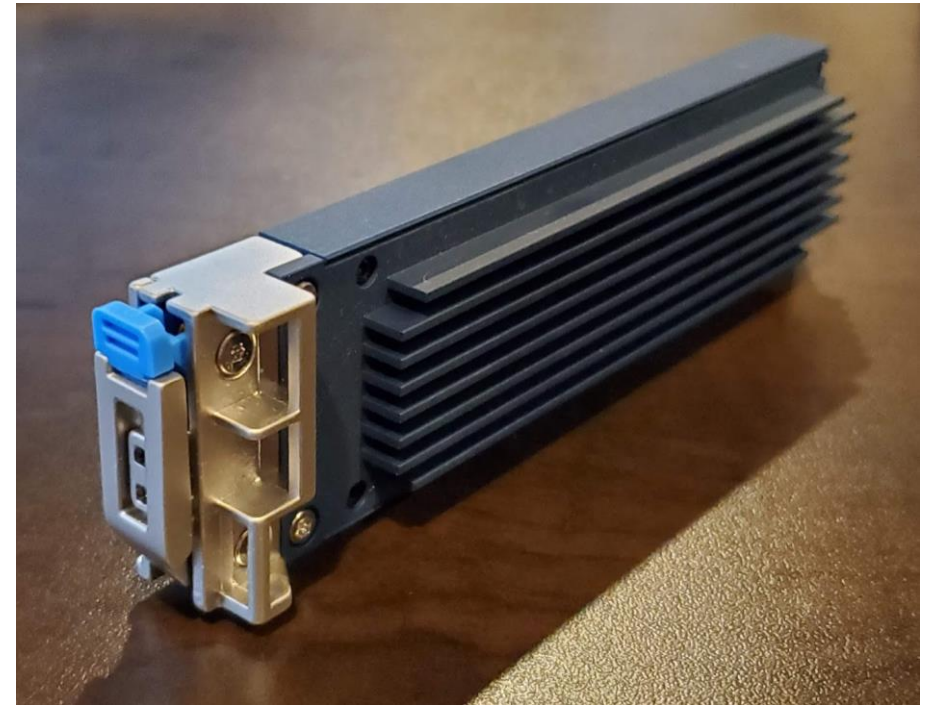
Different Usages
Same Expectations!

Issues at Scale – Form Factors

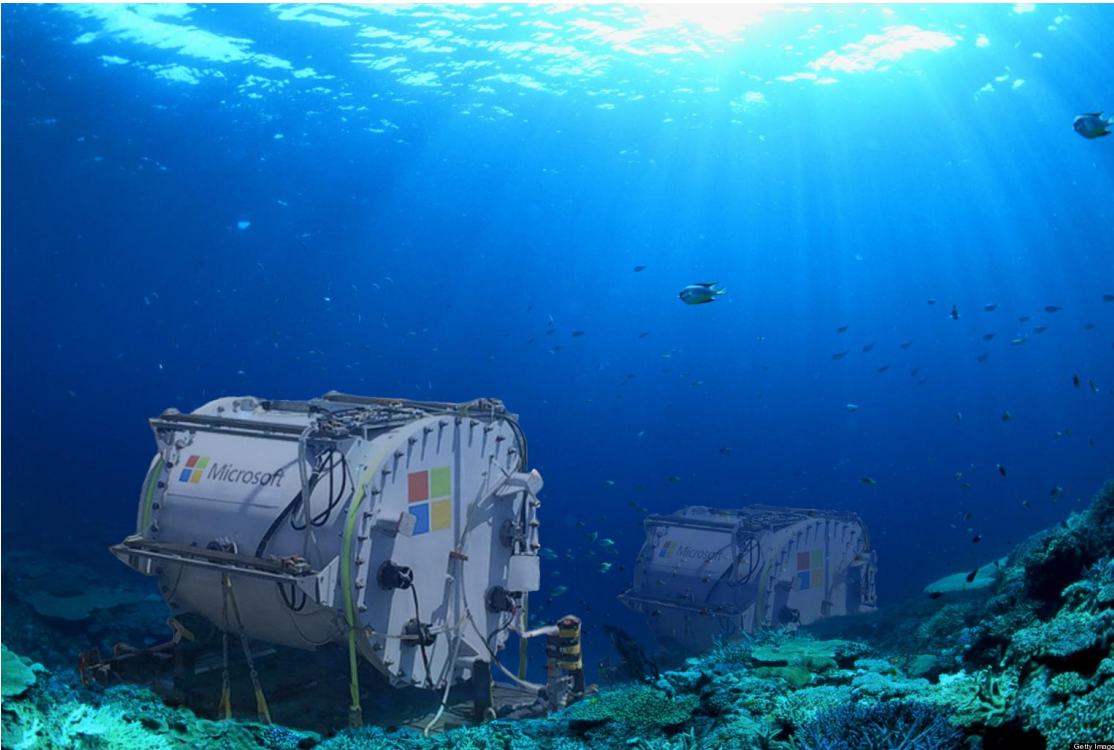
- **Widely deploying E1.L**



- **Beginning ramp of E1.S 15mm**



Issues at Scale – Need to allow for “Rot in Place”



Use the Endurance and Performance metrics for auto tiering

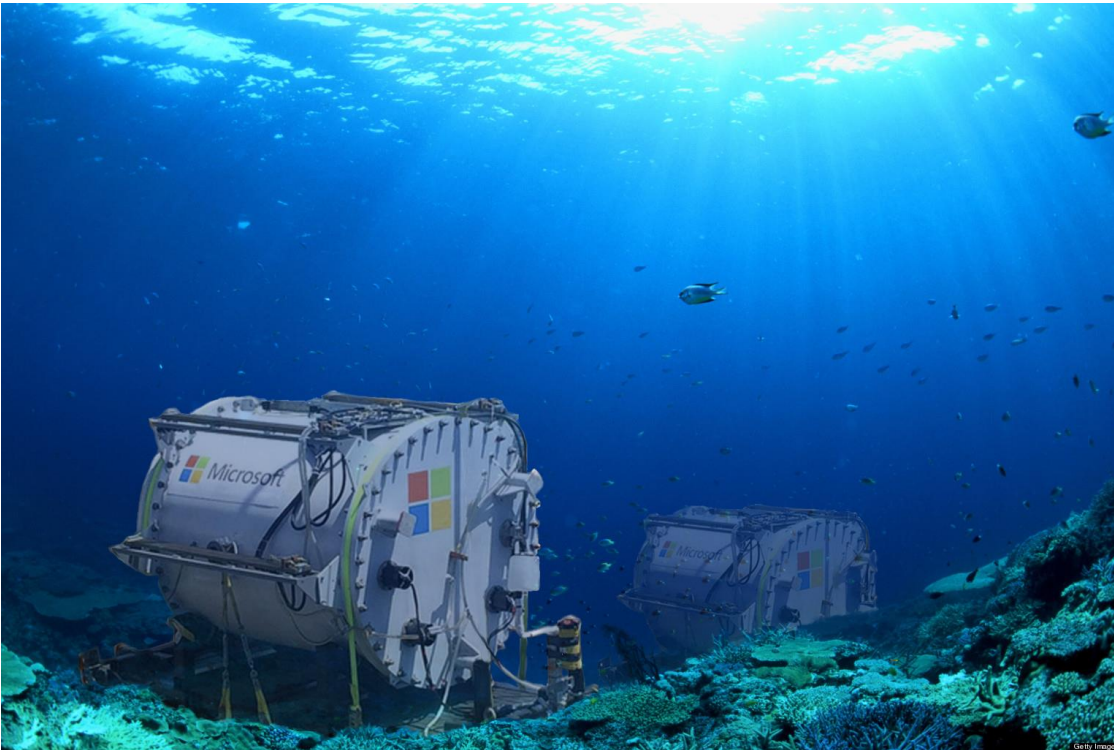
- Allows for fitting the workload to the device
- Allows for the ability to adjust the temperature of the data over time
- Allow for 5 - 7 year device service life

Zoned Name Spaces for QLC

- Reduce WAF due to large sequential writes
- Reduce DRAM due to large indirection unit
- Reduce overprovisioning due to minimal garbage collection

Issues at Scale – Need to allow for “Rot in Place”

Not a lot of progress here 😞



Use the Endurance and Performance metrics for auto tiering

- Allows for fitting the workload to the device
- Allows for the ability to adjust the temperature of the data over time
- Allow for 5-to-7-year device service life

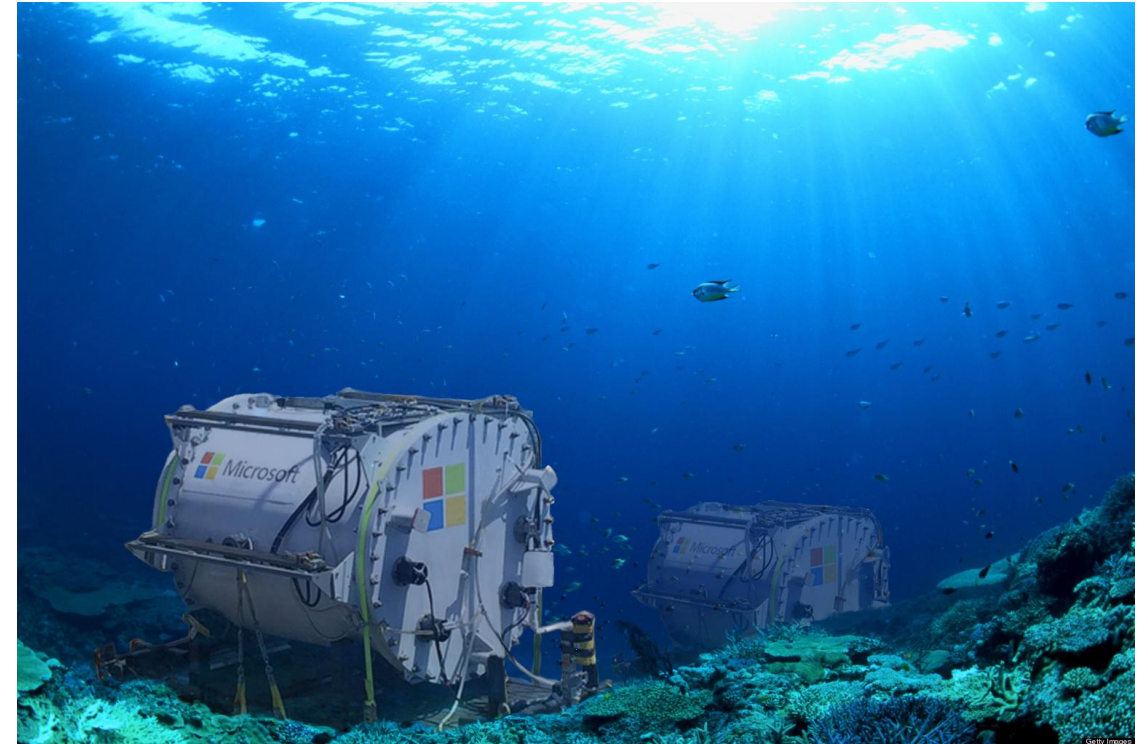
Zoned Name Spaces for QLC

- Reduce WAF due to large sequential writes
- Reduce DRAM due to large indirection unit
- Reduce overprovisioning due to minimal garbage collection

Issues at Scale – Need for Remote Debugging

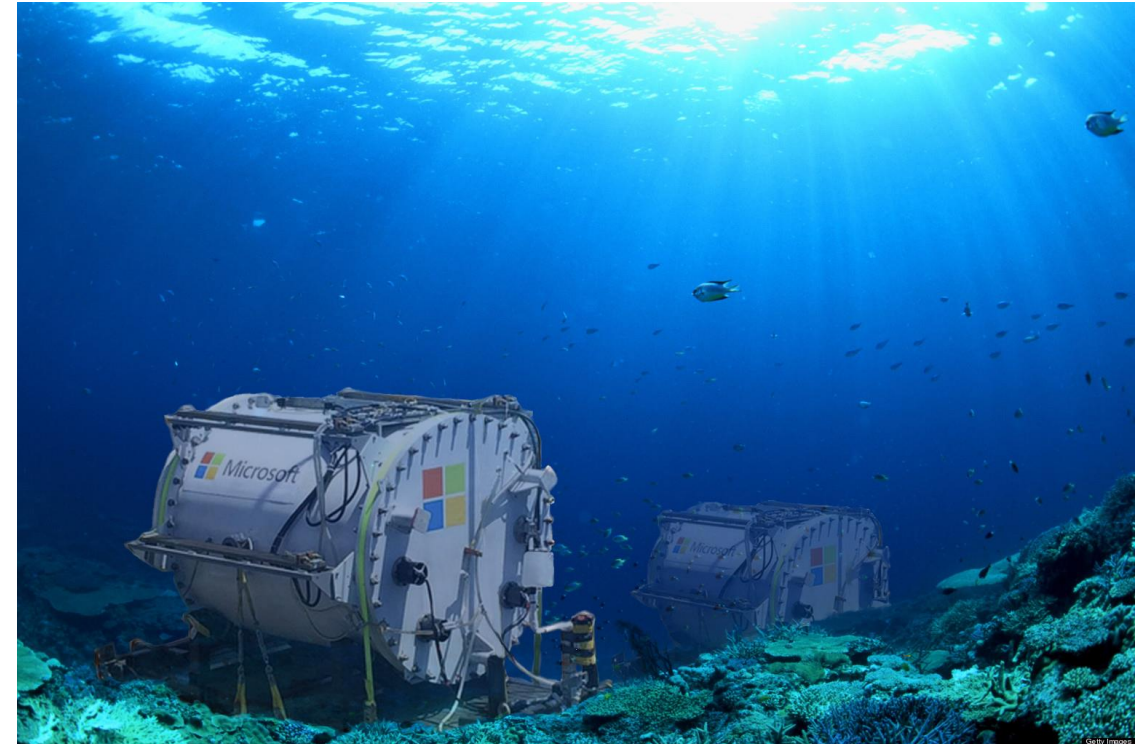
- Timestamp
 - Drive events correlated to system (BIOS and OS) events
- Telemetry
 - Host initiated - IO failures
 - Drive Initiated - Firmware panic?
- SMART
 - Both standard and vendor unique collected once an hour
 - Hey SSD IHVs. How many terabytes would you like to see?

Caveat: Any data that leaves the datacenter must be in human readable form!



Issues at Scale – Need for Remote Debugging

- Expanded C0h Log Page
 - More counters to catch problems early
- Latency Monitoring
 - Continuous monitoring to catch latency spikes
- Error Recovery
 - SSD should never brick



Issues at Scale – Need for Security



eDrive on Windows

- Opal v2 plus IEEE 1667 secure silo

Hardware Root of Trust

- Secure boot
- Signed firmware
- Cerberus

Device Hardening

- Pen and Fuzz testing
- Locking of debug ports and vendor unique commands

Issues at Scale – Need for Security



Attestation

- SPDm via SMBus, PCIe VDM, DOE

FIPS Certification

- FIPS 140-3 Level 1

Device Hardening

- 3rd Party code audits