

# Inside Azure Datacenter Architecture

Markus Klein, Microsoft Deutschland GmbH  
Roland Meier, Freelancer





# Inside Azure Datacenters

[Microsoft Datacenter Tour -  
YouTube](#)





# Azure Global Infrastructure



# Azure physical infrastructure

## Geography

- Discrete market with two or more regions
- Meets data residency and compliance requirements
- Fault-tolerant to protect from complete region failure

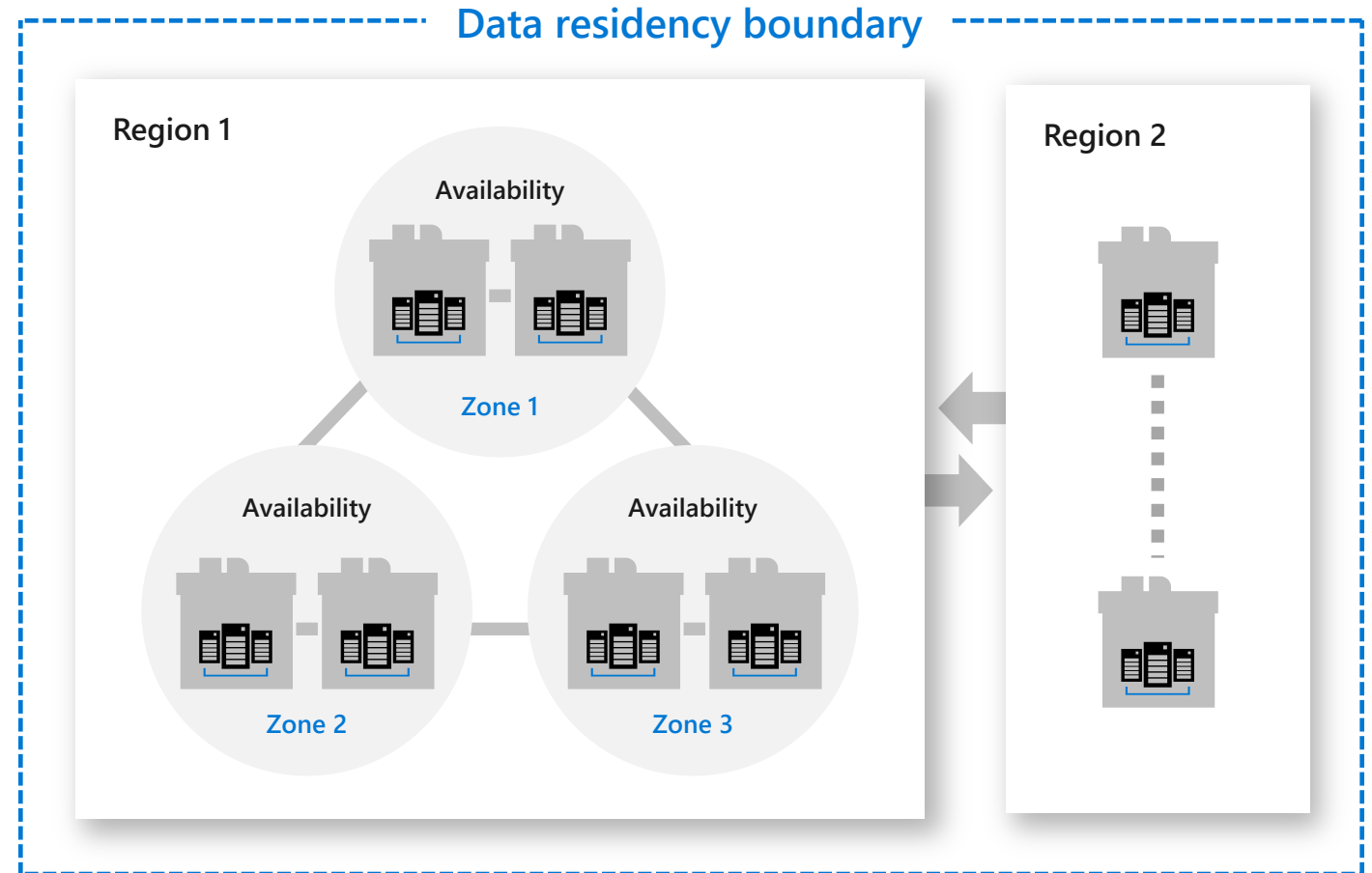
## Region

- Set of datacenters within a metropolitan area
- Network latency perimeter <2ms

## Availability Zones

- Unique physical locations within an Azure region
- Each zone is made up of one or more DCs
- Independent power, cooling and networking
- Inter-AZ network latency <2ms
- Fault-tolerant to protect from datacenter failure

## Geography





Dublin

DB3

DB4





# Dublin

DB5

DB3

DB4





# Dublin

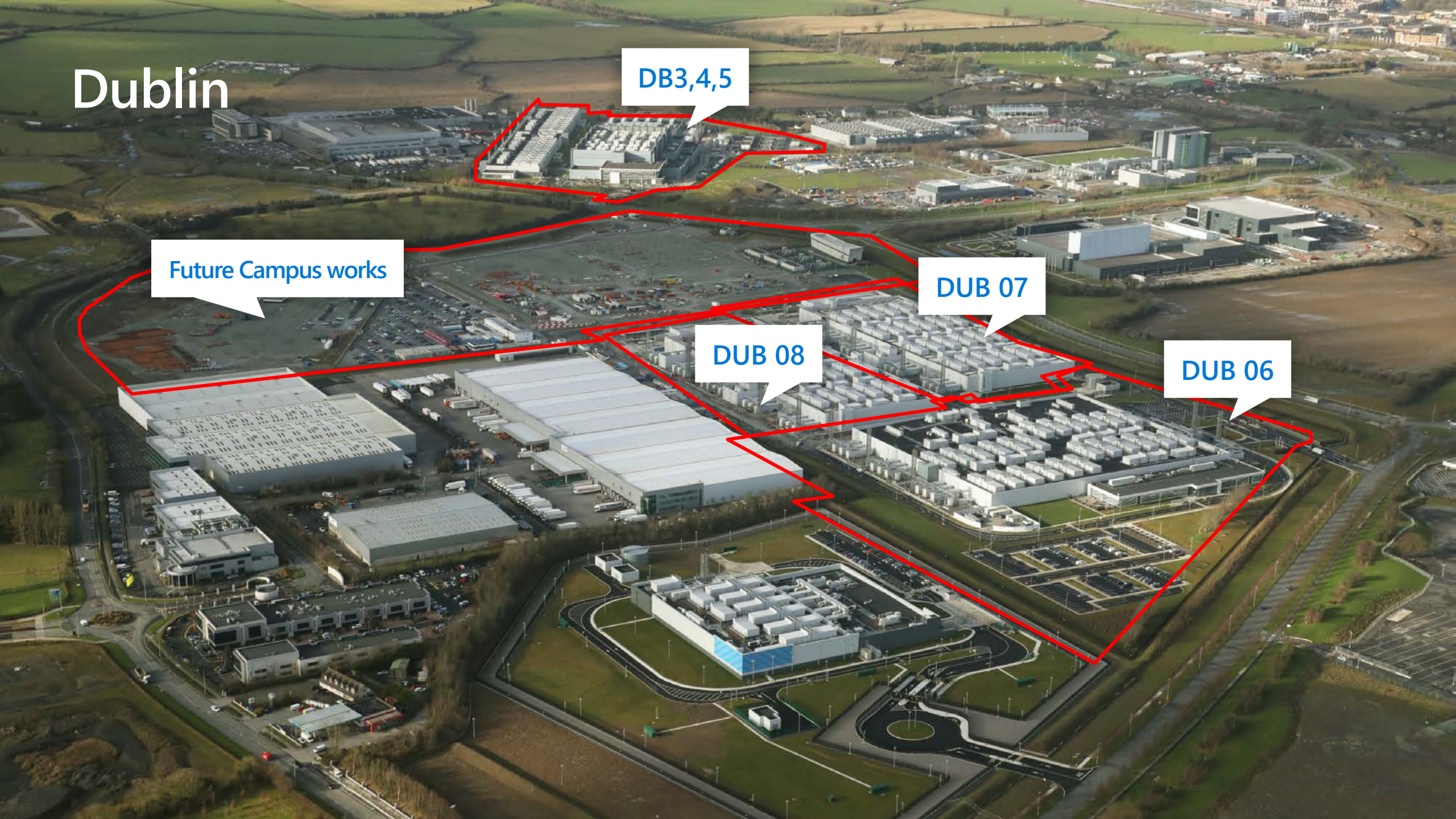
DB3,4,5

Future Campus works

DUB 07

DUB 08

DUB 06





# Singapore







# Azure data center reliability

**>99.999%**

Datacenter availability  
across our fleet of more  
than 100 datacenters and  
millions of servers.

**2.6B**

Datacenter Power &  
Cooling IoT points per day

---

**55%**

Reduction in Human-  
Caused datacenter power  
and cooling incidents

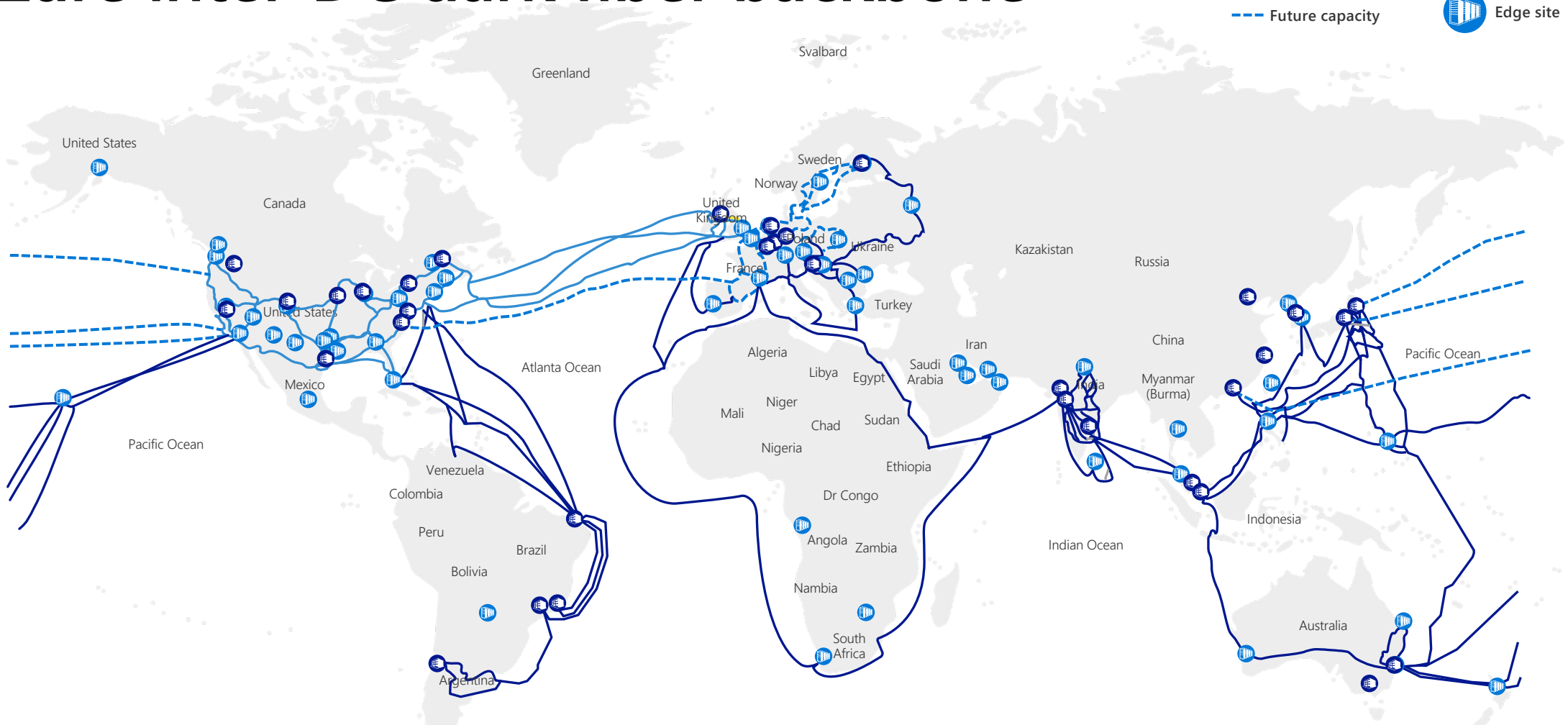
---

**2800**

Average monthly datacenter  
events correlated against  
Azure customer experience



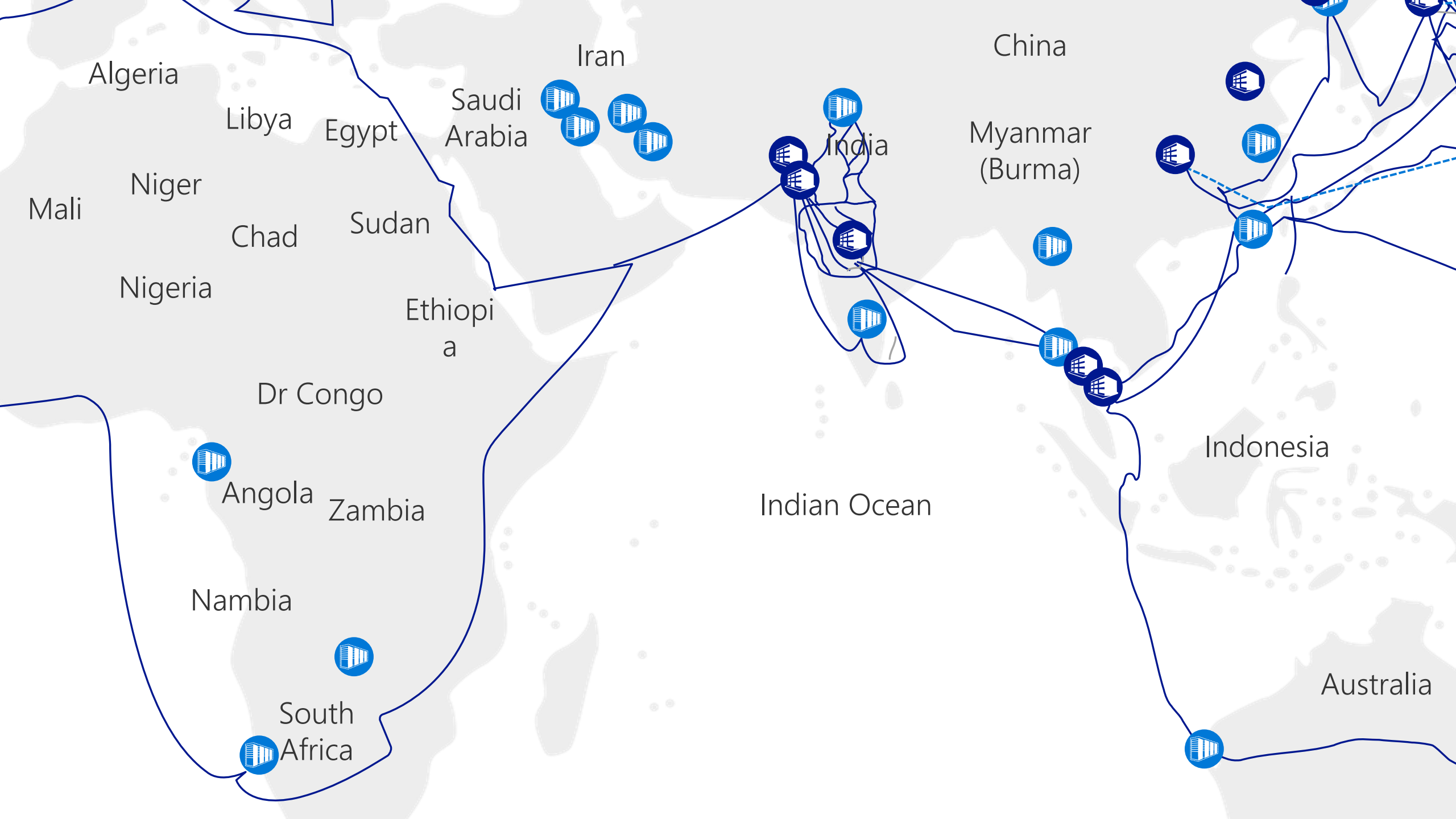
# Azure inter-DC dark fiber backbone



100k+ miles of fiber and subsea cable

150+ edge sites

## Insert text here





# Azure Regional Networks

## Edge

Connects Region to Internet and Enterprise peers

## Regional Network Gateway

Connects Regions to Regions, DC to DC

Contiguous geographical area  
~100km in diameter

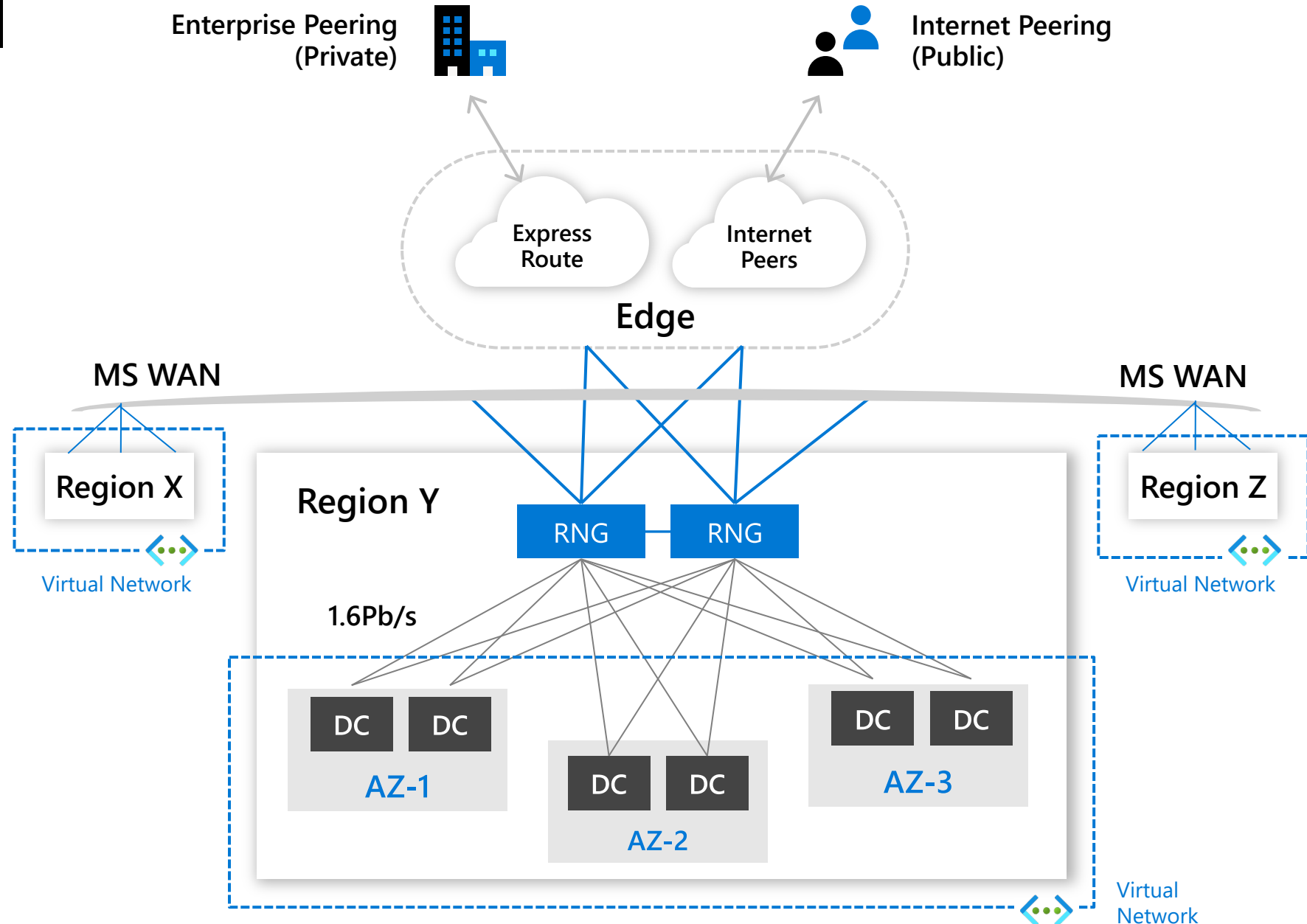
T-shirt sized (28 to 528MW)

## Data Centers

DC to DC latency – 1.8ms

AZ to AZ latency – 1.0ms

Within DC – 100 microseconds



# Microsoft is committed to operating environmentally responsible datacenters



Industry leader in renewable energy



Energy- and water-efficient datacenters



Carbon neutral since 2012



LEED Gold and zero-waste certification for our latest datacenters



More energy and carbon efficient than traditional enterprise datacenters





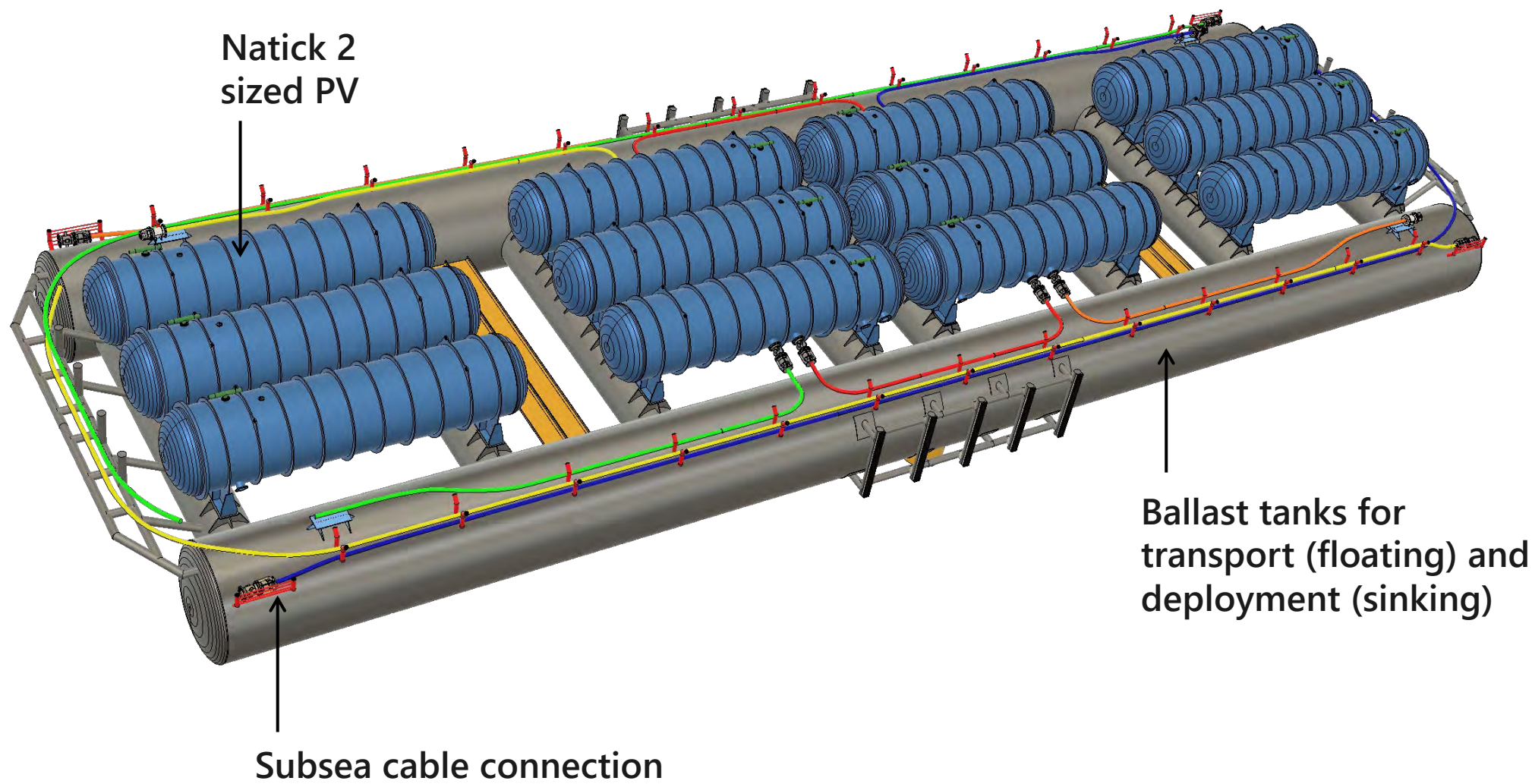
# Project Natick















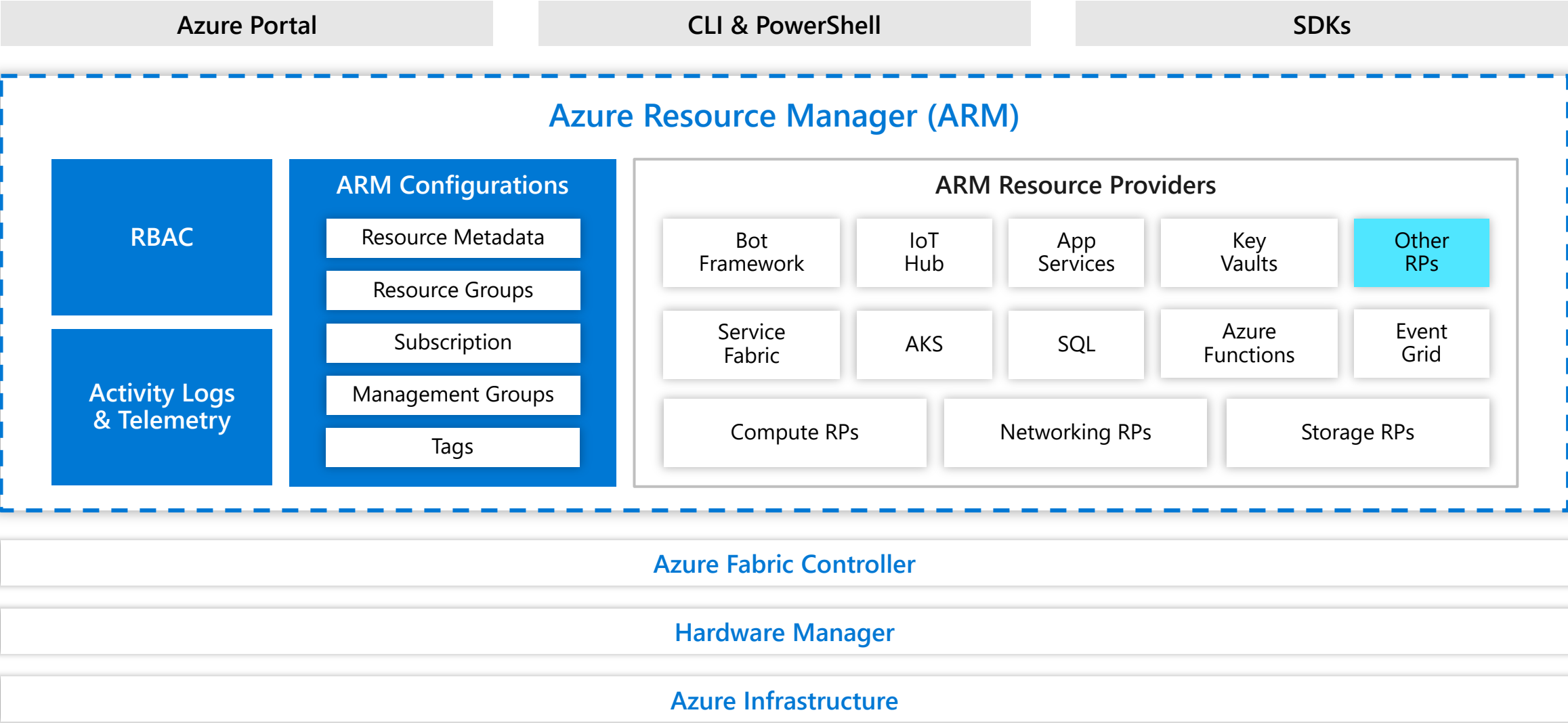




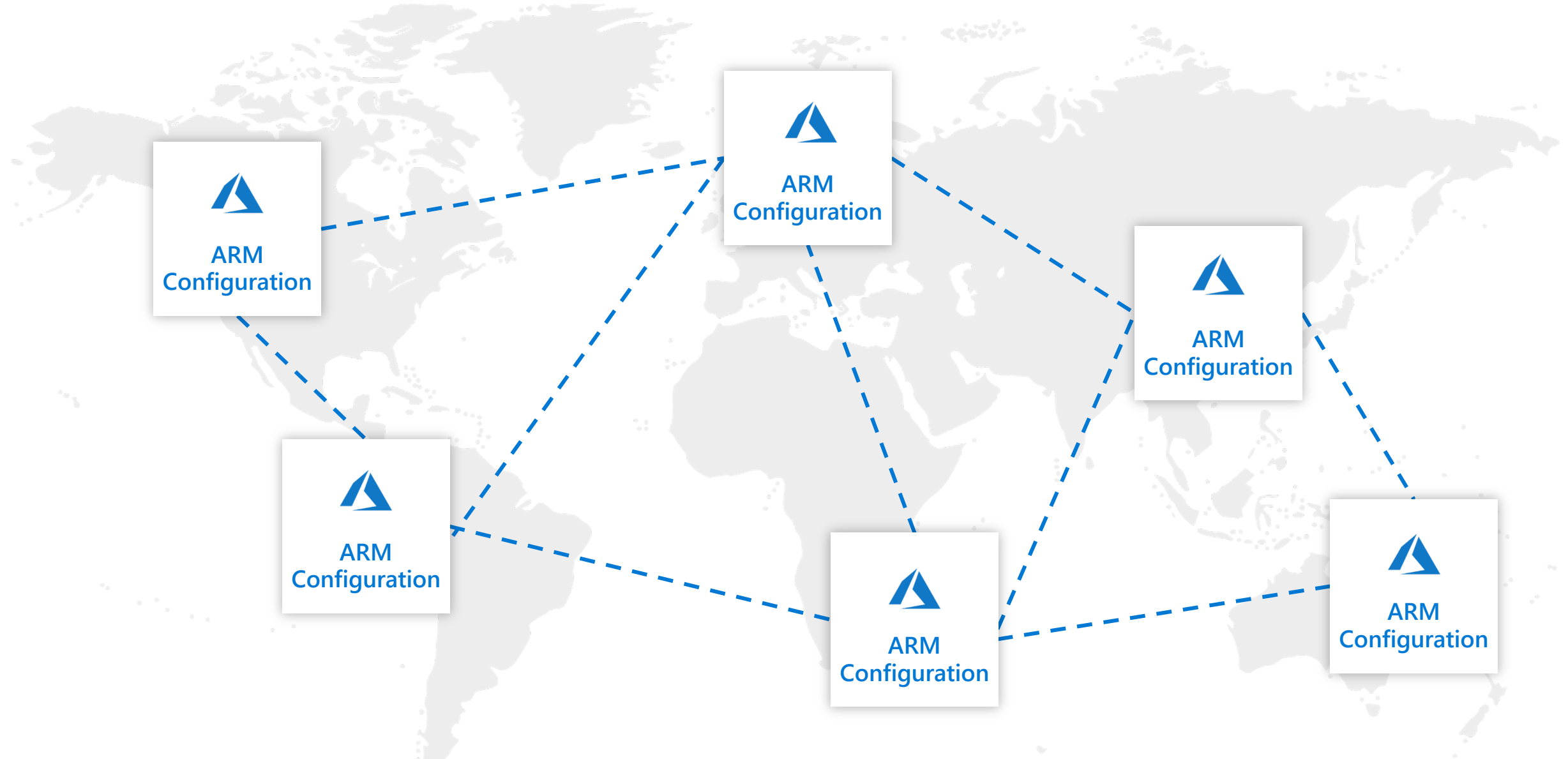
# Inside Azure Compute and Applications



# Azure Architecture

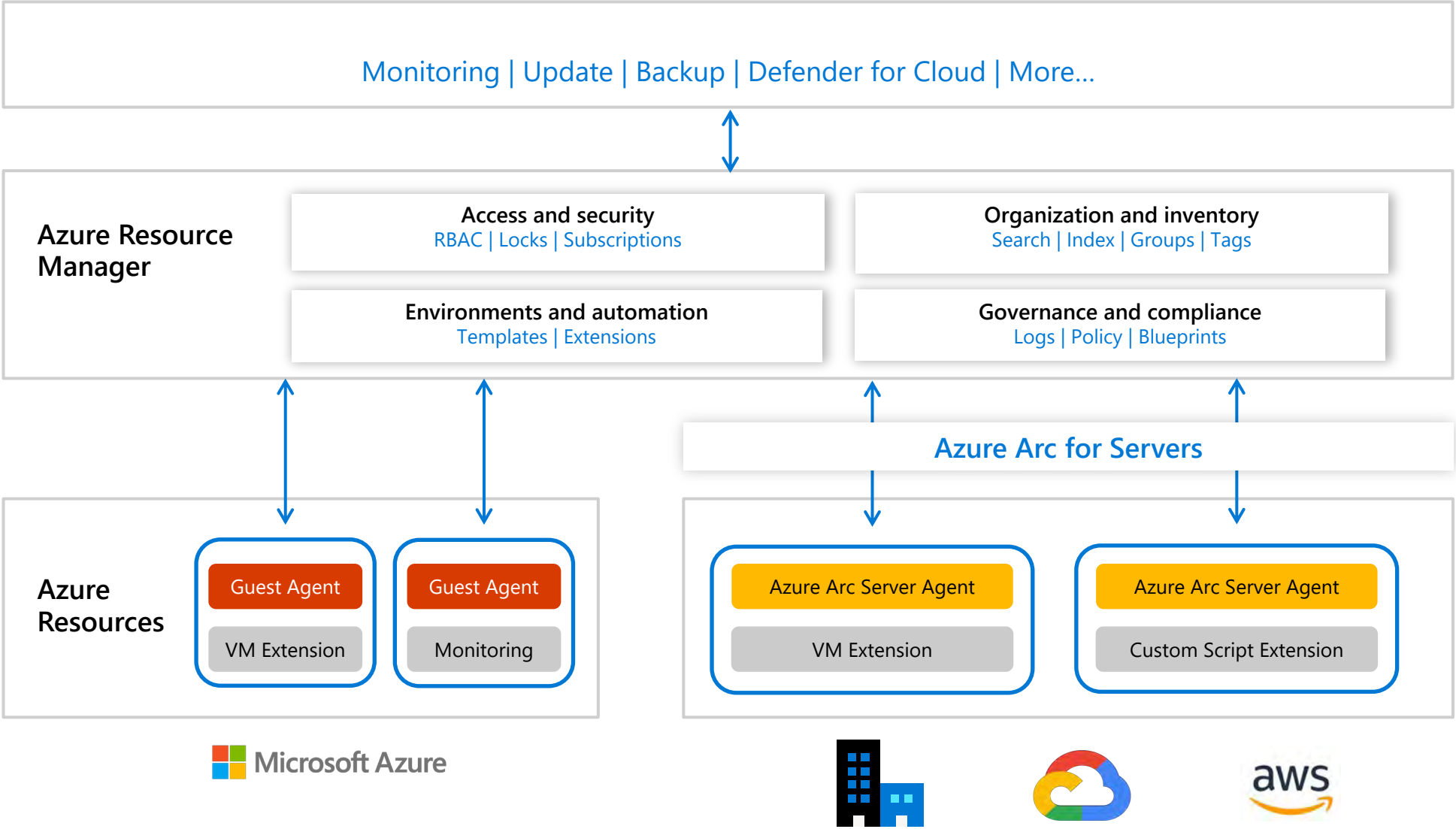


# Globally available



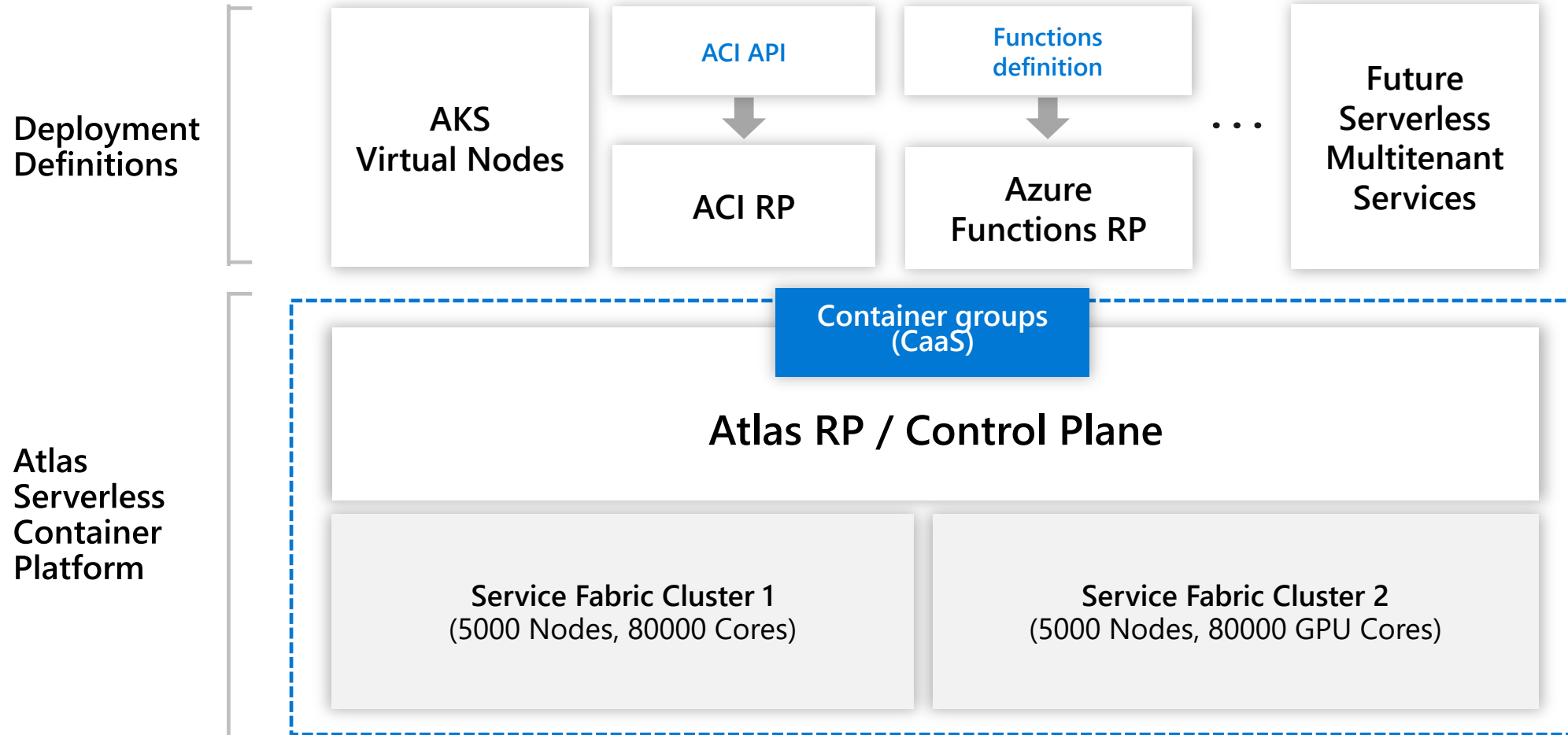


# Server management – Azure Arc



# Atlas

Multitenant, serverless containers platform for container-based Azure services





# Project Teleport



SMB mounted pre-expanded layers from Azure Container Registry to Teleport Hosts

Any custom image, to any serverless host, at 90% of the startup time of locally cached images

|                                | 2K    | 200MB | 2GB    | 5GB    |
|--------------------------------|-------|-------|--------|--------|
| Dedicated VM                   | 1.8s  | 12.7s | 83.9s  | 412.8s |
| Azure Container Instance (ACI) | 25.3s | 66.4s | 188.1s | 522.4s |
| Project Teleport               | 2.8s  | 3.3s  | 4.1s   | 7.6s   |

# Azure Confidential Computing

## Confidential Computing Vision

Data is fully in control of the customer

Azure has no access to customer data

Code accessing customer data is authorized by the customer

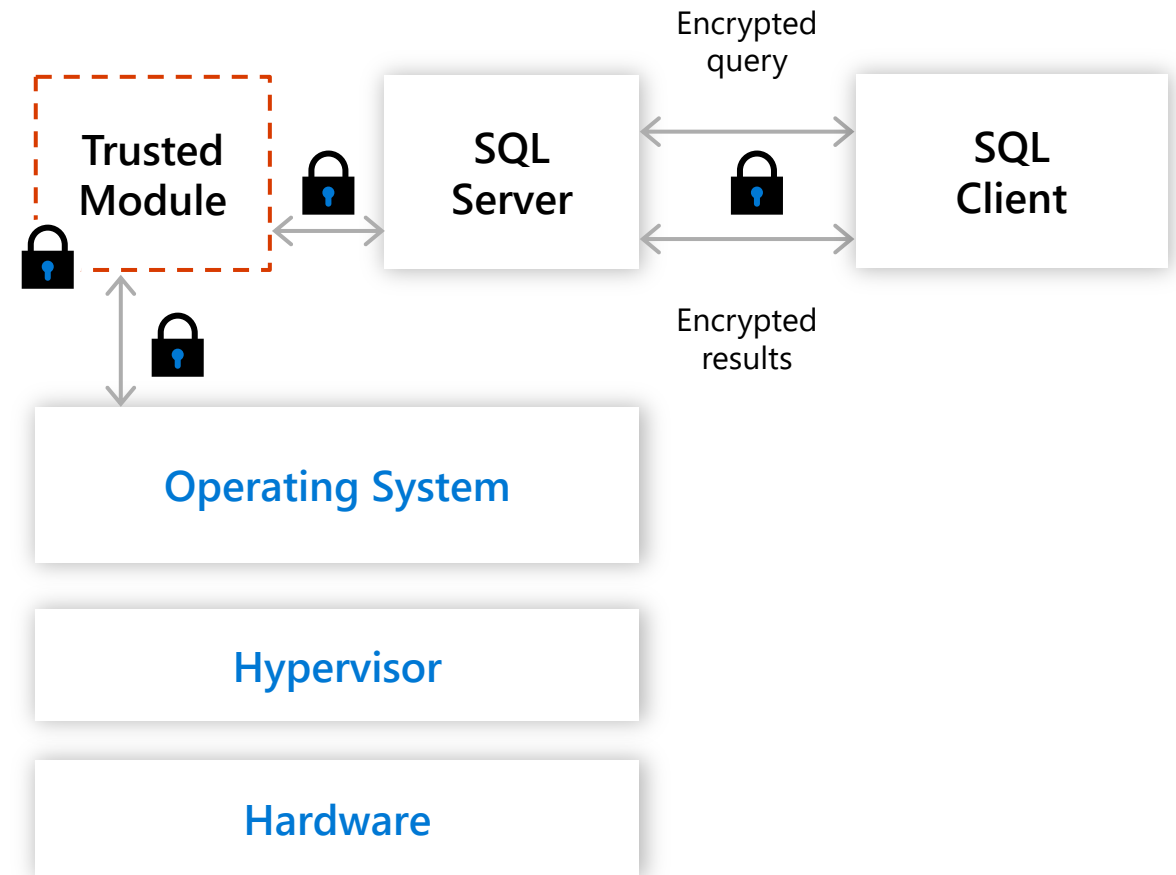
## Confidential Computing Approach

Working with silicon partners to enable confidential computing

Tools to deploy, manage and develop TEE applications

Services to support TEE attestation

Confidential PaaS and SaaS services





# Confidential Computing Inference Service Architecture

## ONNX Runtime ported to OpenEnclave

Confidential execution of ONNX models in Intel SGX® enclaves

## New security services:

HSM key storage on Azure Key Vault

Federated authentication of hardware enclaves by Azure Attestation Service

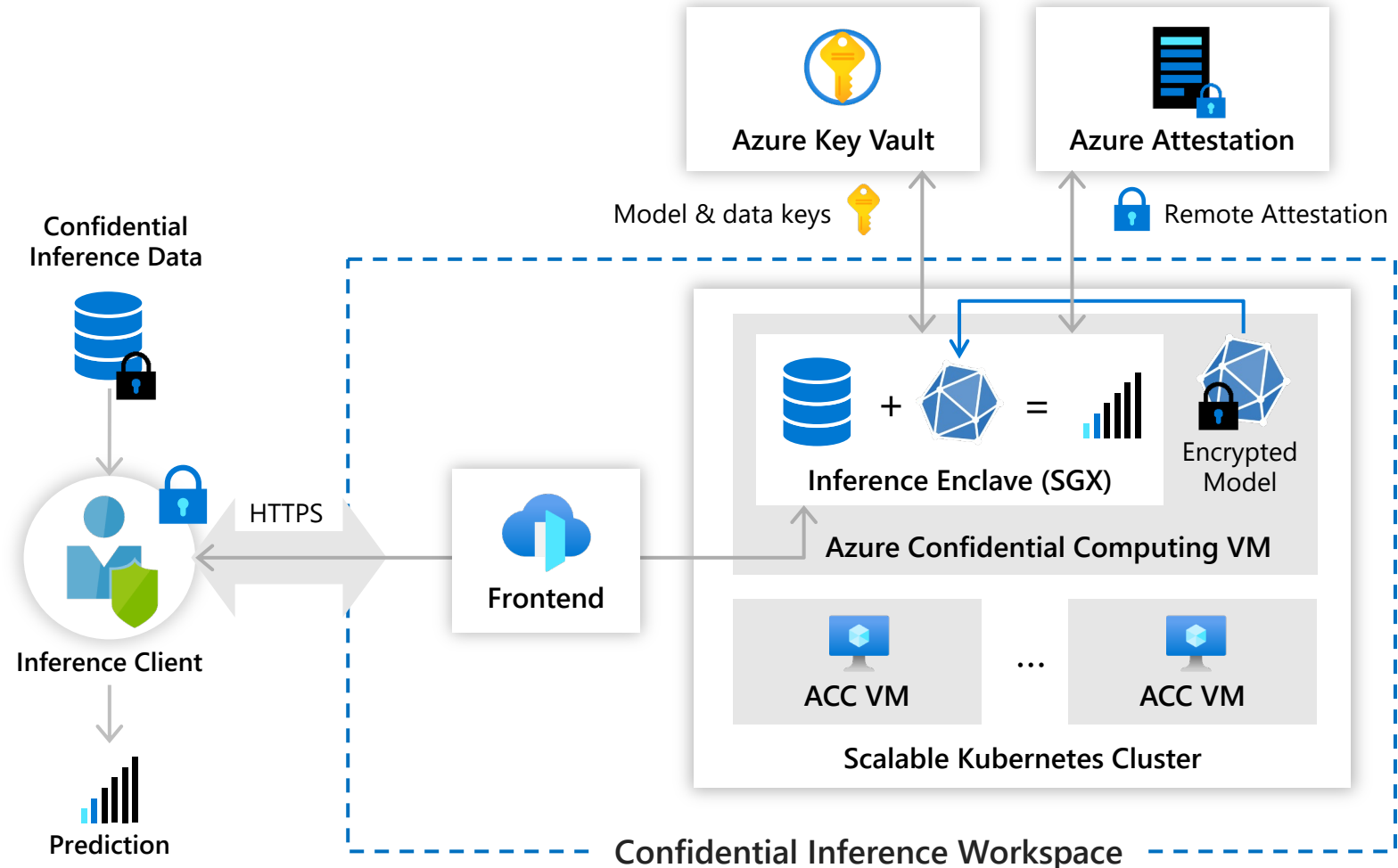
## New protocols:

Secure key export from Key Vault to an authorized hardware enclave

Tunnelled, server-stateless encryption of inference request and response

## Verifiable confidentiality:

Service users can check hardware evidence that proves only authorized code can access inference data



# Inside Azure Networking





# Azure Networking overview

## DC hardware

- SmartNIC/FPGA
- SONiC

## Services

- Virtual Networks
- Load Balancing
- VPN Services
- Firewall
- DDoS Protection
- DNS and Traffic Management

## Intra-region

- DC Networks
- Regional Networks
- Optical Modules

## WAN backbone

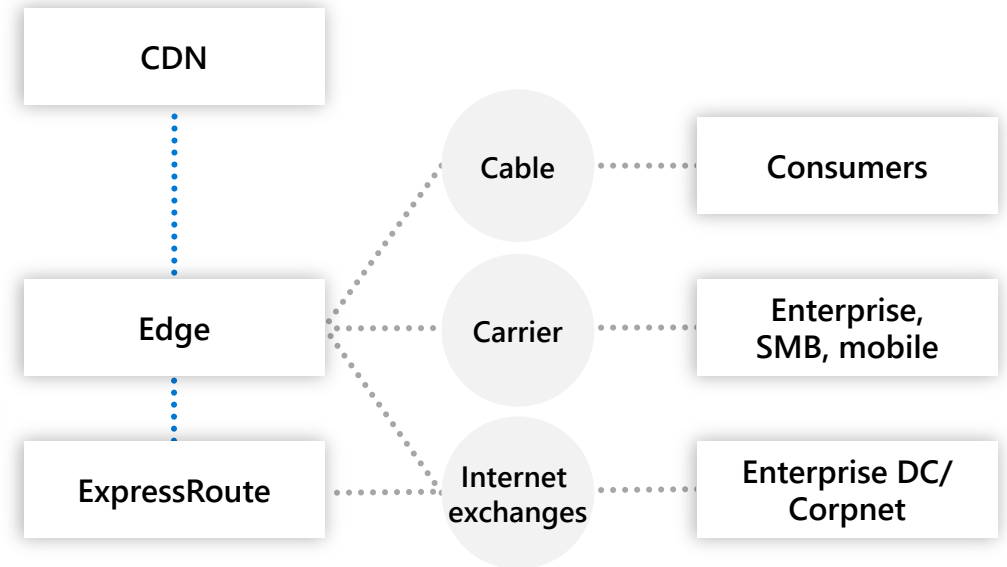
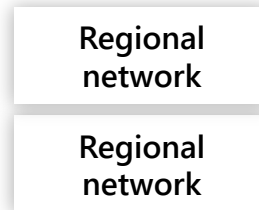
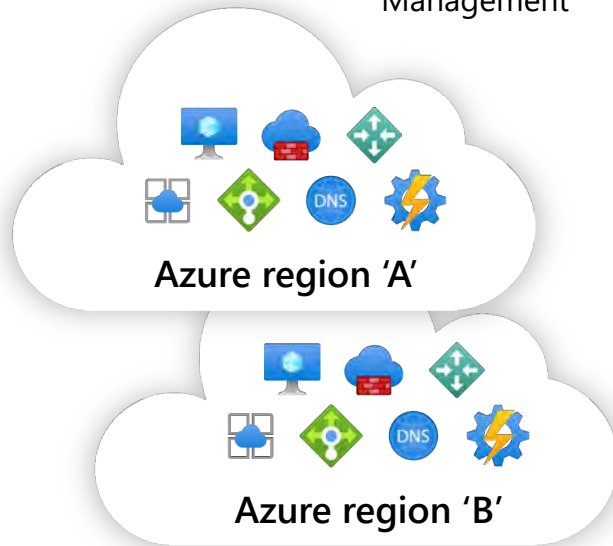
- Software WAN
- Subsea Cables
- Terrestrial Fiber
- National Clouds

## CDN

- Acceleration for applications and content

## Last mile

- E2E monitoring (Network Watcher, Network Performance Monitoring)



## Edge and ExpressRoute

- Internet Peering
- ExpressRoute

# Private Link

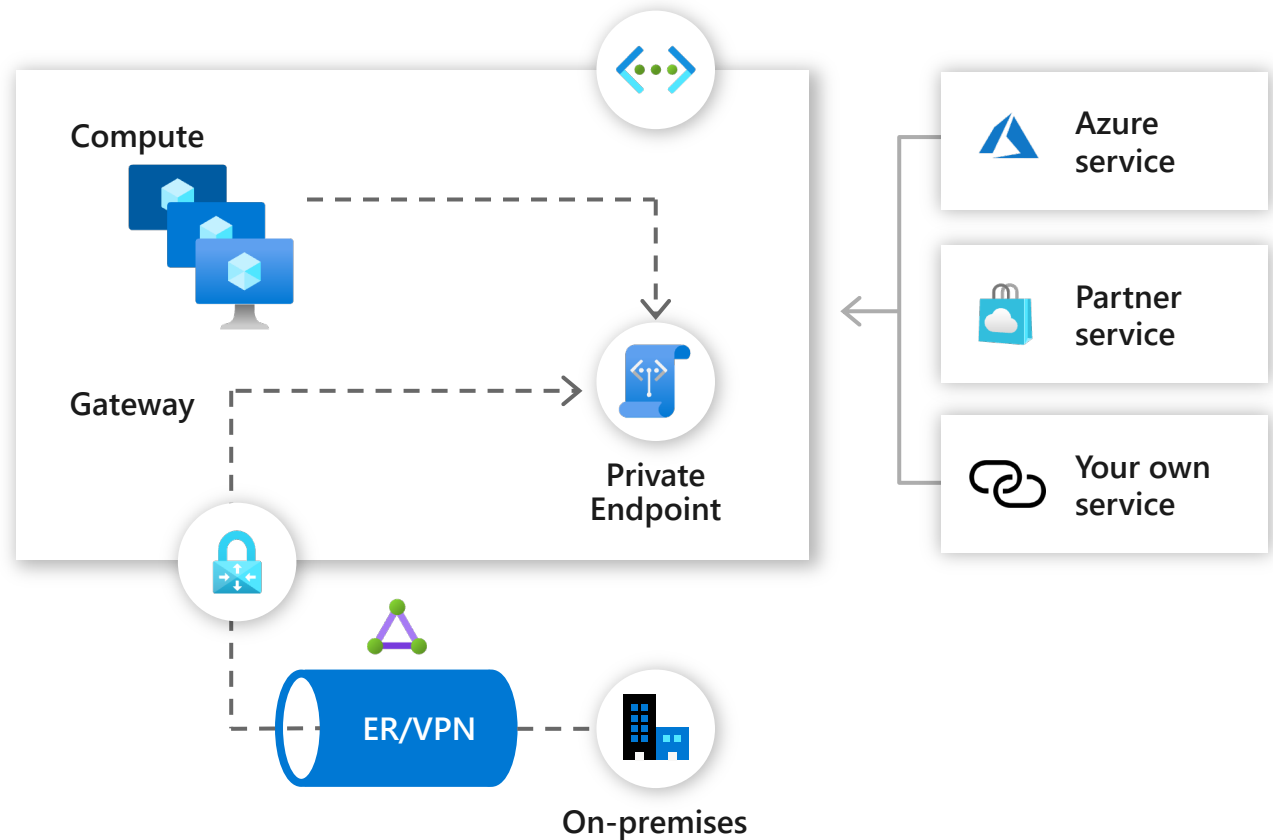
Highly secure and private connectivity solution for Azure Platform

Consistent experience across Azure services, partner services and your own services

Simplified networking

- No Internet gateway, NAT devices, public IP, ER or VPN
- Predicable IP addresses for PaaS resources
- Access from peered on-prem networks privately

Simplified security

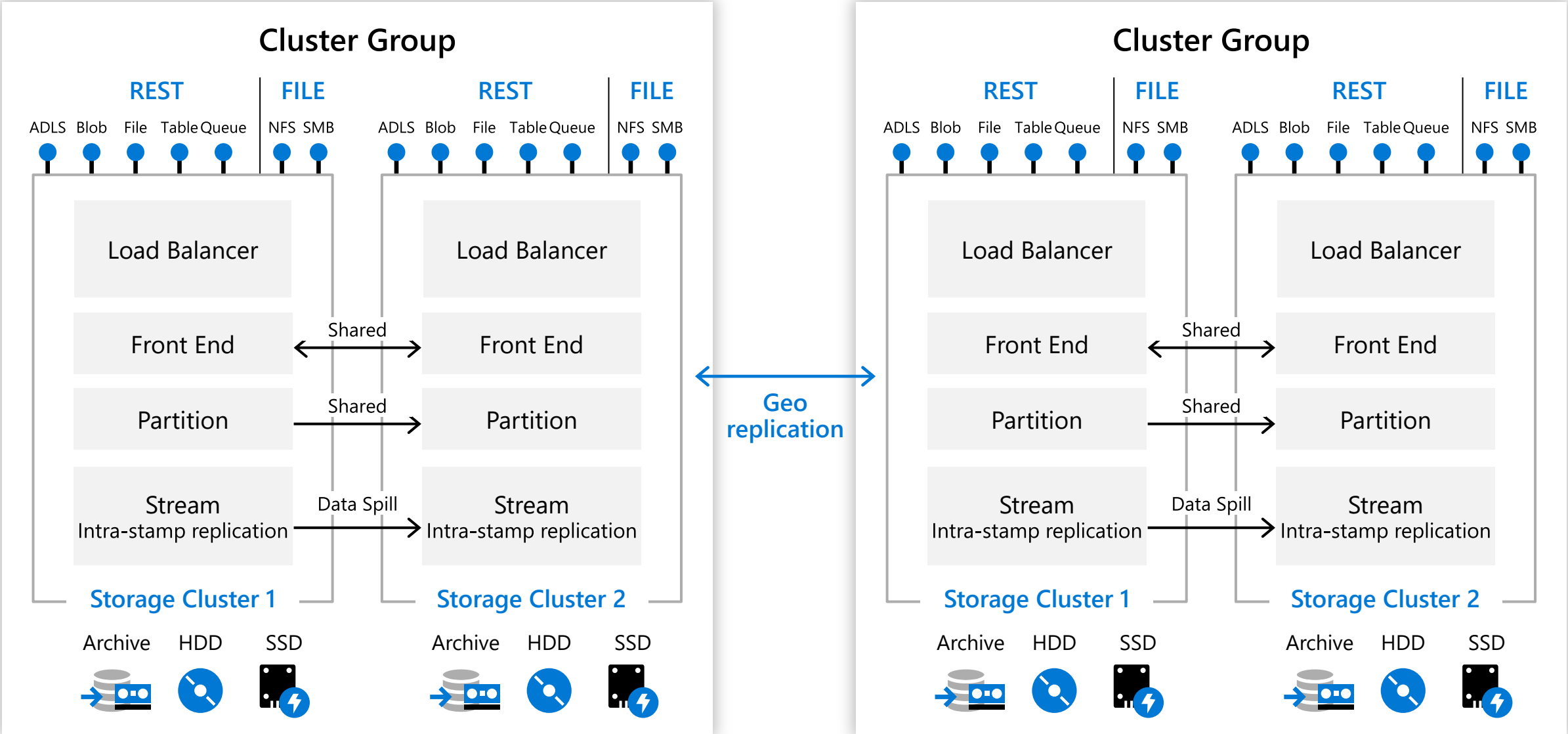




# Inside Azure Storage

A photograph of a data center aisle. In the foreground, there are blurred server racks with blue and red cables. In the middle ground, a technician wearing a headset and a dark shirt is standing on a green step ladder, working on a server rack. The background shows more server racks and yellow overhead lighting. The text "Inside Azure Storage" is overlaid on the left side of the image.

# Azure Storage architecture





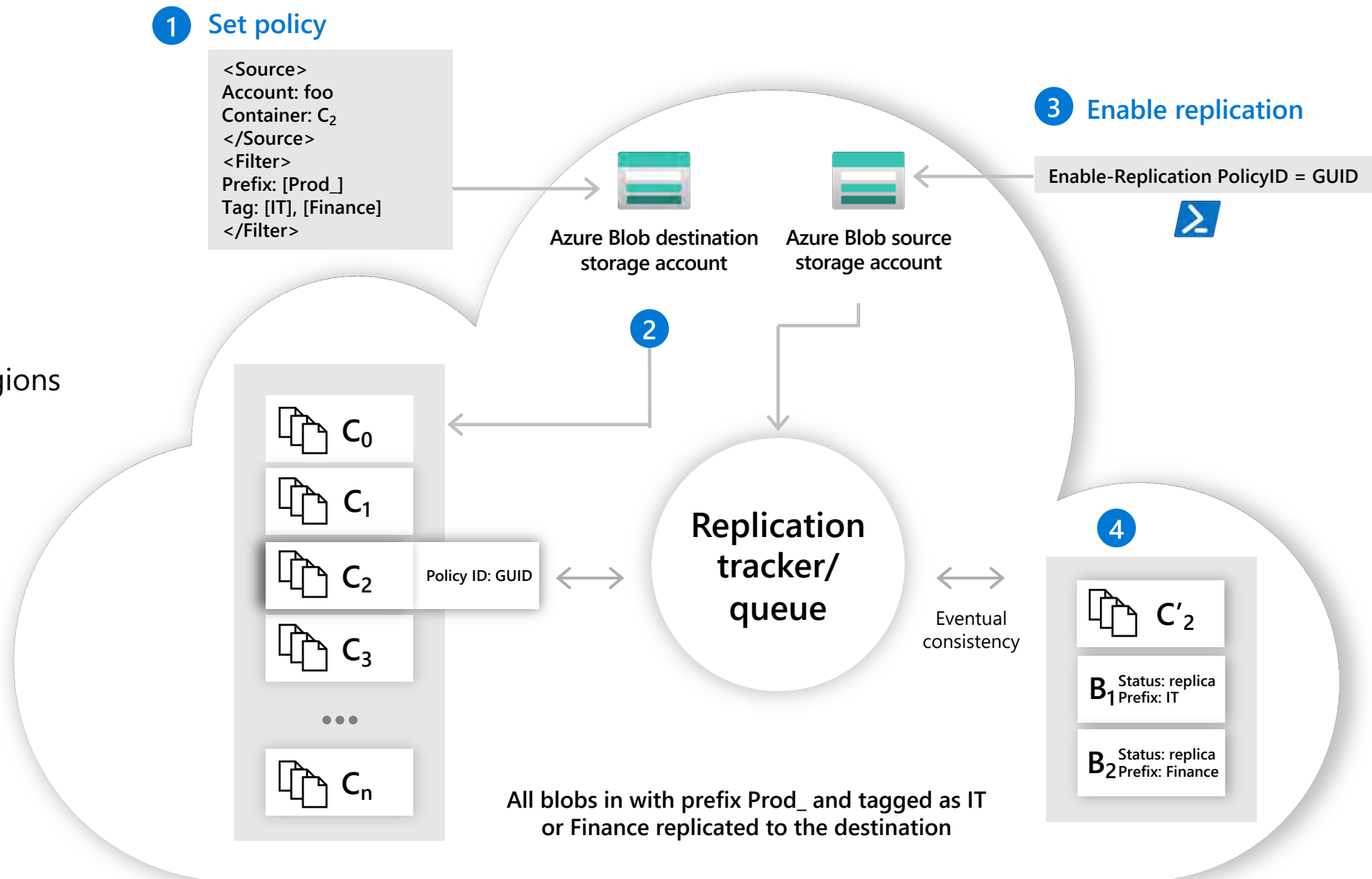
# Object Replication Service

Flexible replication at the scope of choice to the regions of choice for block blobs

Minimize latency for your applications or create a low-cost backup solution and more

No additional cost

Requires versioning and change feed to be enabled



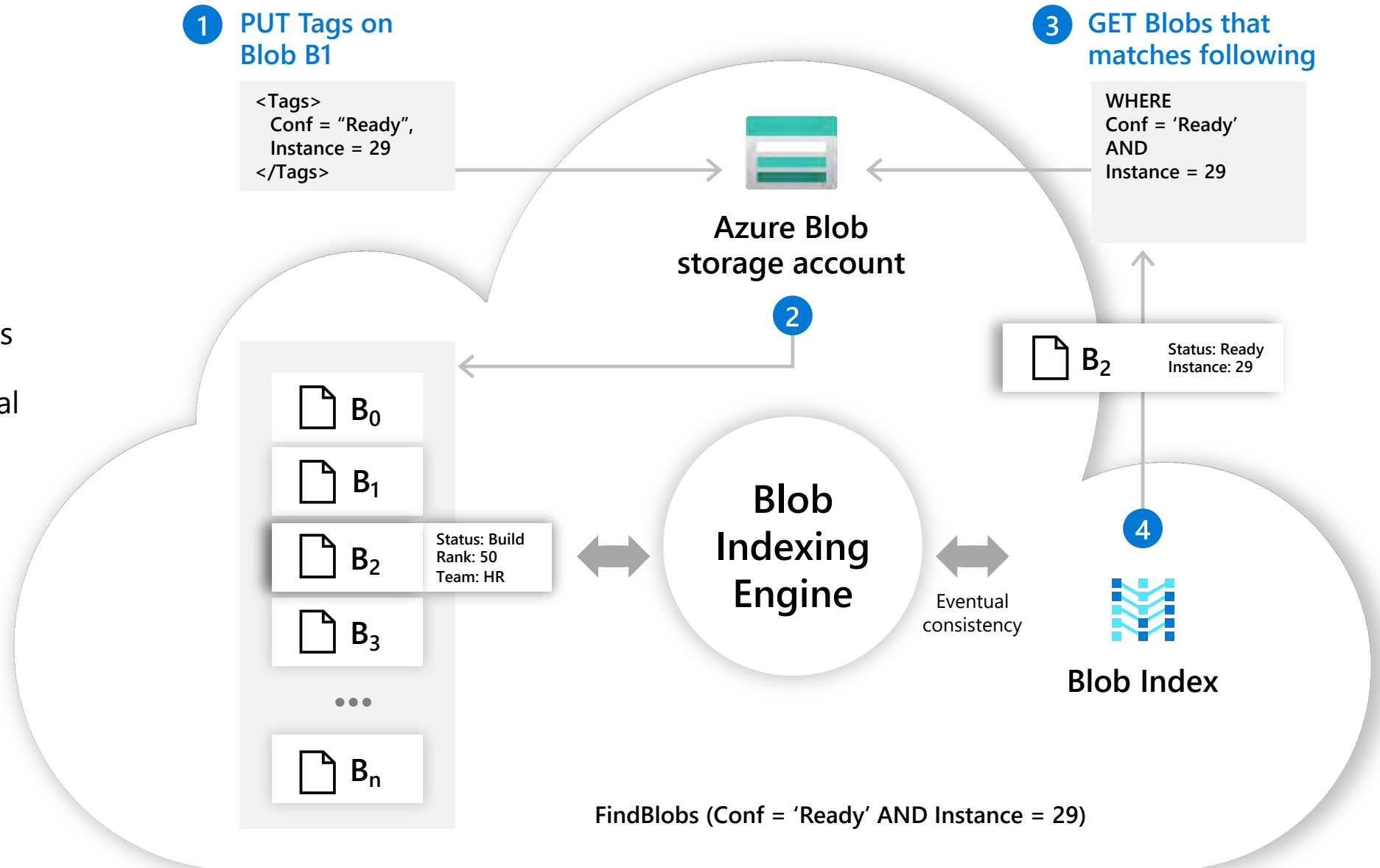
# Azure Blob Index

Multi-dimensional search for blobs that satisfy some key-value conditions

Blobs will support a special kind of key-value sub resource which will be auto indexed

This special index (Blob Index) powers the FindBlobs API

A FindBlobs lookup over millions of blobs can return results in seconds





# Azure Blob Quick Query

What if blob storage understood schema?

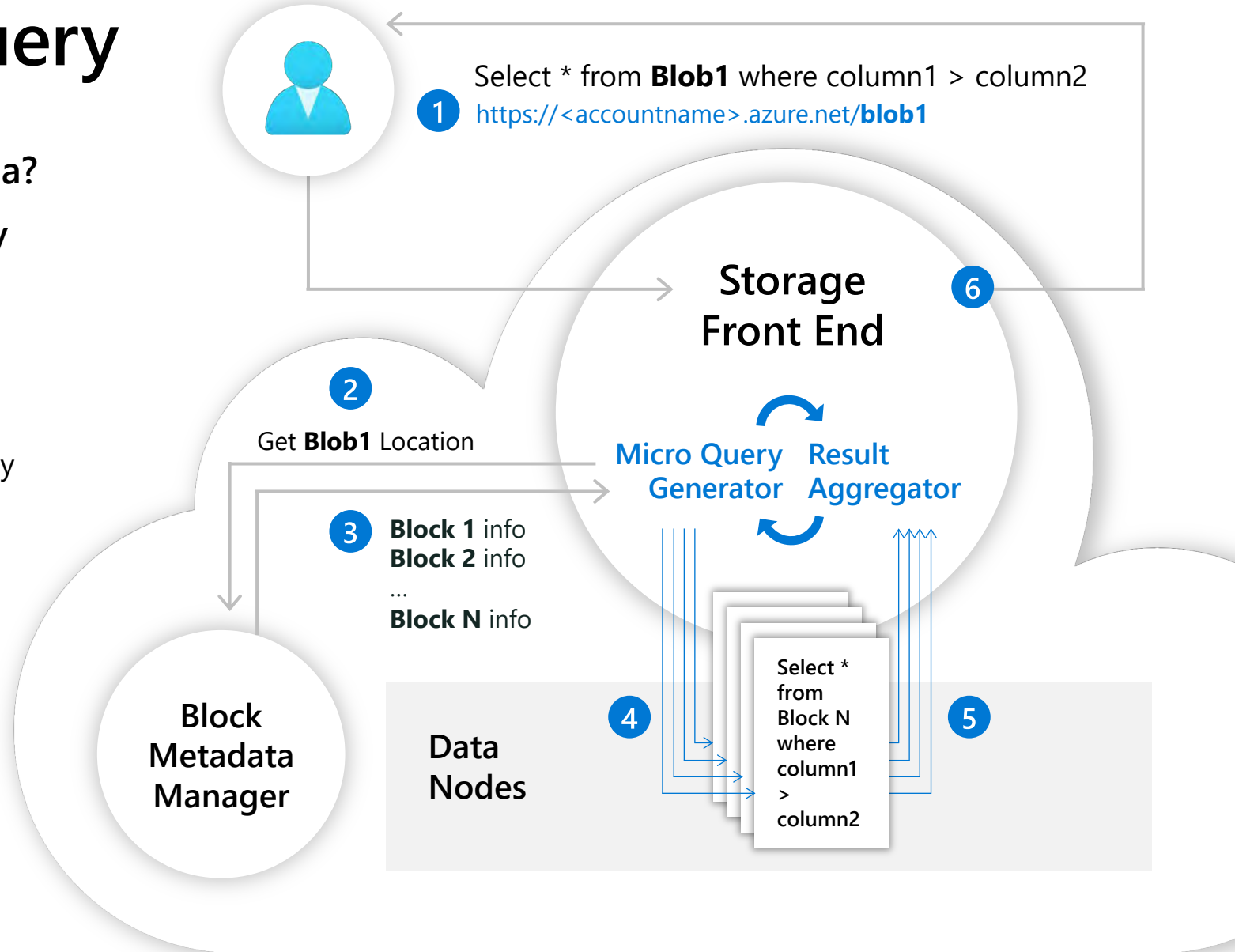
What if you could write a SQL like query to retrieve and filter your data?

## Available options today:

- Run custom code on compute node to parse/filter
- Or upload from Blob to SQL Azure then run a query
- Or spin up Hadoop Cluster (HDI) to read Blob data and filter

## Quick Query:

- ✓ SELECT empid (string), age (int) FROM blob1 WHERE zipcode = 98067
- ✓ Serverless, elegant, simple, cost effective
- ✓ CSV, JSON
- ✓ CSV\_Split (splits CSV files into regions with full records)
- ✓ Archival data





# Inside Azure Servers





# Azure servers: General purpose

↕ RAM    ↔ Cores



## Gen 2

|            |                    |
|------------|--------------------|
| Processor  | 2 x 6 Core 2.1 GHz |
| Memory     | 32 GiB             |
| Hard Drive | 6 x 500 GB         |
| SSD        | None               |
| NIC        | 1 Gb/s             |

## Godzilla

|            |                     |
|------------|---------------------|
| Processor  | 2 x 16 Core 2.0 GHz |
| Memory     | 512 GiB             |
| Hard Drive | None                |
| SSD        | 9 x 800 GB          |
| NIC        | 40 Gb/s             |

## Intel Gen 6

|            |   |
|------------|---|
| Processor  | 2 x Skylake 24 Core 2.7GHz              |
| Memory     | 768GiB DDR4                             |
| Hard Drive | None                                    |
| SSD        | 4 x 960 GB M.2 SSDs and 1 x 960 GB SATA |
| NIC        | 40 Gb/s + FPGA                          |

## Optimized Gen 6

|            |                          |
|------------|--------------------------|
| Processor  | 2 x 24 core Skylake Lake |
| Memory     | 192 GB DDR4              |
| Hard Drive | None                     |
| SSD        | 5 x 960 GB M.2 NVMe      |
| NIC        | 40 Gb/s + FPGA           |

## AMD Gen 6

|            |                     |
|------------|---------------------|
| Processor  | 2 x 32 core Naples  |
| Memory     | 512 GB DDR4         |
| Hard Drive | None                |
| SSD        | 7 x 960 GB M.2 NVMe |
| NIC        | 50 Gb/s + FPGA      |

## Intel Gen 7

|            |                          |
|------------|--------------------------|
| Processor  | 2 x 26 core Cascade Lake |
| Memory     | 576 GB DDR4              |
| Hard Drive | None                     |
| SSD        | 7 x 960 GB M.2 NVMe      |
| NIC        | 50 Gb/s + FPGA           |

## AMD Gen 7

|            |                     |
|------------|---------------------|
| Processor  | 2 x 32 core Rome    |
| Memory     | 768 GB DDR4         |
| Hard Drive | None                |
| SSD        | 7 x 960 GB M.2 NVMe |
| NIC        | 50 Gb/s + FPGA      |

## Beast

|            |                                |
|------------|--------------------------------|
| Processor  | 4 x 18 Core 2.5 GHz            |
| Memory     | 4096 GiB                       |
| Hard Drive | None                           |
| SSD        | 4 x 2 TB NVMe, 1 x 960 GB SATA |
| NIC        | 40 Gb/s                        |

## Beast v2

|            |                                |
|------------|--------------------------------|
| Processor  | 8 x 28 Core 2.5 GHz            |
| Memory     | 12 TiB                         |
| Hard Drive | None                           |
| SSD        | 4 x 2 TB NVMe, 1 x 960 GB SATA |
| NIC        | 50 Gb/s                        |

# Azure servers: General purpose

↕ RAM    ↔ Cores

3x  
Beast



0.00000000533  
Beast v2s



Gen 2

|            |                    |
|------------|--------------------|
| Processor  | 2 x 6 Core 2.1 GHz |
| Memory     | 32 GiB             |
| Hard Drive | 6 x 500 GB         |
| SSD        | None               |
| NIC        | 1 Gb/s             |



Godzilla

|            |                     |
|------------|---------------------|
| Processor  | 2 x 16 Core 2.0 GHz |
| Memory     | 512 GiB             |
| Hard Drive | None                |
| SSD        | 9 x 800 GB          |
| NIC        | 40 Gb/s             |



Gen 6

|            |   |
|------------|---|
| Processor  | 2 x Skylake 24 Core 2.7GHz              |
| Memory     | 768GiB DDR4                             |
| Hard Drive | None                                    |
| SSD        | 4 x 960 GB M.2 SSDs and 1 x 960 GB SATA |
| NIC        | 40 Gb/s + FPGA                          |



Optimized Gen 6

|            |                          |
|------------|--------------------------|
| Processor  | 2 x 24 core Skylake Lake |
| Memory     | 192 GB DDR4              |
| Hard Drive | None                     |
| SSD        | 4 x 960 GB M.2 NVMe      |
| NIC        | 40 Gb/s + FPGA           |



Gen 7

|            |                          |
|------------|--------------------------|
| Processor  | 2 x 26 core Cascade Lake |
| Memory     | 576 GB DDR4              |
| Hard Drive | None                     |
| SSD        | 7 x 960 GB M.2 NVMe      |
| NIC        | 50 Gb/s + FPGA           |



Beast

|            |                                |
|------------|--------------------------------|
| Processor  | 4 x 18 Core 2.5 GHz            |
| Memory     | 4096 GiB                       |
| Hard Drive | None                           |
| SSD        | 4 x 2 TB NVMe, 1 x 960 GB SATA |
| NIC        | 40 Gb/s                        |



Beast v2

|            |                                |
|------------|--------------------------------|
| Processor  | 8 x 28 Core 2.5 GHz            |
| Memory     | 12 TiB                         |
| Hard Drive | None                           |
| SSD        | 4 x 2 TB NVMe, 1 x 960 GB SATA |
| NIC        | 50 Gb/s                        |

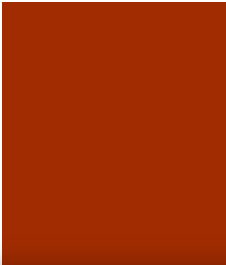


Azure Sphere

|           |                          |
|-----------|--------------------------|
| Processor | 2 x M4 Core @ 200 MHz    |
| Memory    | 64KB RAM                 |
| WiFi      | 2.4/5.0 GHz 802.11 b/g/n |



# Azure servers: Special purpose



| HPC        |                        |
|------------|------------------------|
| Processor  | 2 x 12 Core 2.4 GHz    |
| Memory     | 128 GiB                |
| Hard Drive | 5 x 1 TB               |
| SSD        | None                   |
| NIC        | 10 Gb/s IP, 40 Gb/s IB |



| HB         |                                 |
|------------|---------------------------------|
| Processor  | 2 x 32 Core 2.5 GHz             |
| Memory     | 240 GiB                         |
| Hard Drive | None                            |
| SSD        | 2 x 960 GB NVMe                 |
| NIC        | 50 Gb/s Ethernet, 100 Gb EDR IB |



| HBv2       |   |
|------------|---|
| Processor  | 2 x 64 Core                             |
| Memory     | 480 GiB                                 |
| Hard Drive | None                                    |
| SSD        | 2 x 960 GB NVMe                         |
| NIC        | 50 Gb/s Ethernet, 200 Gb HDR Infiniband |



| ND         |                     |
|------------|---------------------|
| Processor  | 2 x 14 Core 2.6 GHz |
| Memory     | 448 GB              |
| Hard Drive | None                |
| SSD        | 3 x 960 GB NVMe     |
| NIC        | FDR Infiniband      |
| GPU        | NVIDIA P40          |



| NDv2       |                             |
|------------|-----------------------------|
| Processor  | 2 x Skylake 24 Core 2.7 GHz |
| Memory     | 768 GiB                     |
| Hard Drive | None                        |
| SSD        | 6 x 960 GiB NVMe            |
| NIC        | 40 Gb/s                     |
| GPU        | 8 GPU with NVLink           |



| Lv2        |                    |
|------------|--------------------|
| Processor  | 2 x 32 core 2.0GHz |
| Memory     | 1 TiB              |
| Hard Drive | None               |
| SSD        | 12 x 2 TB NVMe     |
| NIC        | 40 Gb/s            |

# SCUTI-O

## Azure Exclusive Cloud Enterprise-class storage device:

The world's fastest SSD; 8us latency and 9GB/s bandwidth at the application level.

Writes as fast as Reads; Reads IOPs at 2.35M and Writes IOPs at 1.95M

A single NVMe drive equipped with 128 I/O queue pair, it provides 1 queue pair per logical processor to 1 processor in 128VP VM.

Constant performance: ultra-performance 3D XPoint media provides superior endurance over NAND.

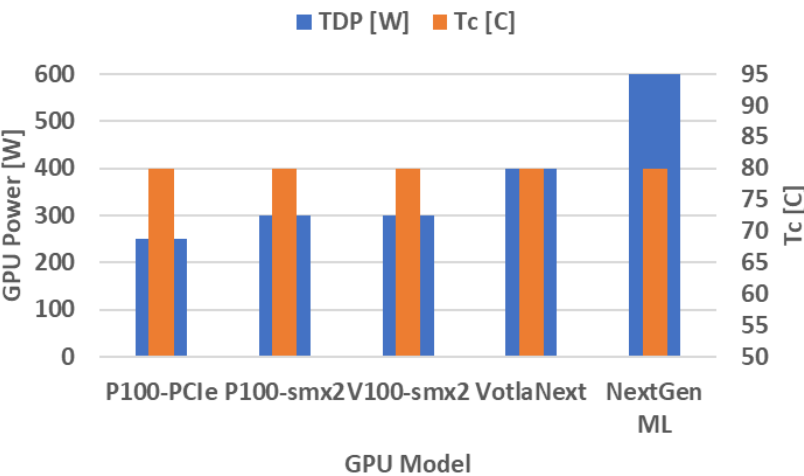
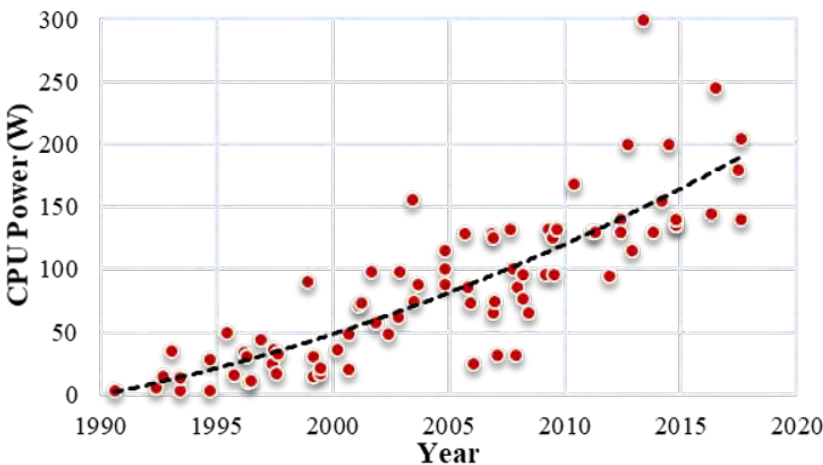
Hardware Built-in Quality of Services for multiple VM scaling.





# Liquid Cooling

## CPU trends



## GPU trends



MSFT G50 Expansion 4kW



Nvidia DGX-2, 10kW

# Liquid Cooling

Micro-channel Cold Plates



One phase immersion



Air Cooled Olympus



Two phase immersion





# Pre-Provisioned Service

## Improved Windows VM Deployment Performance

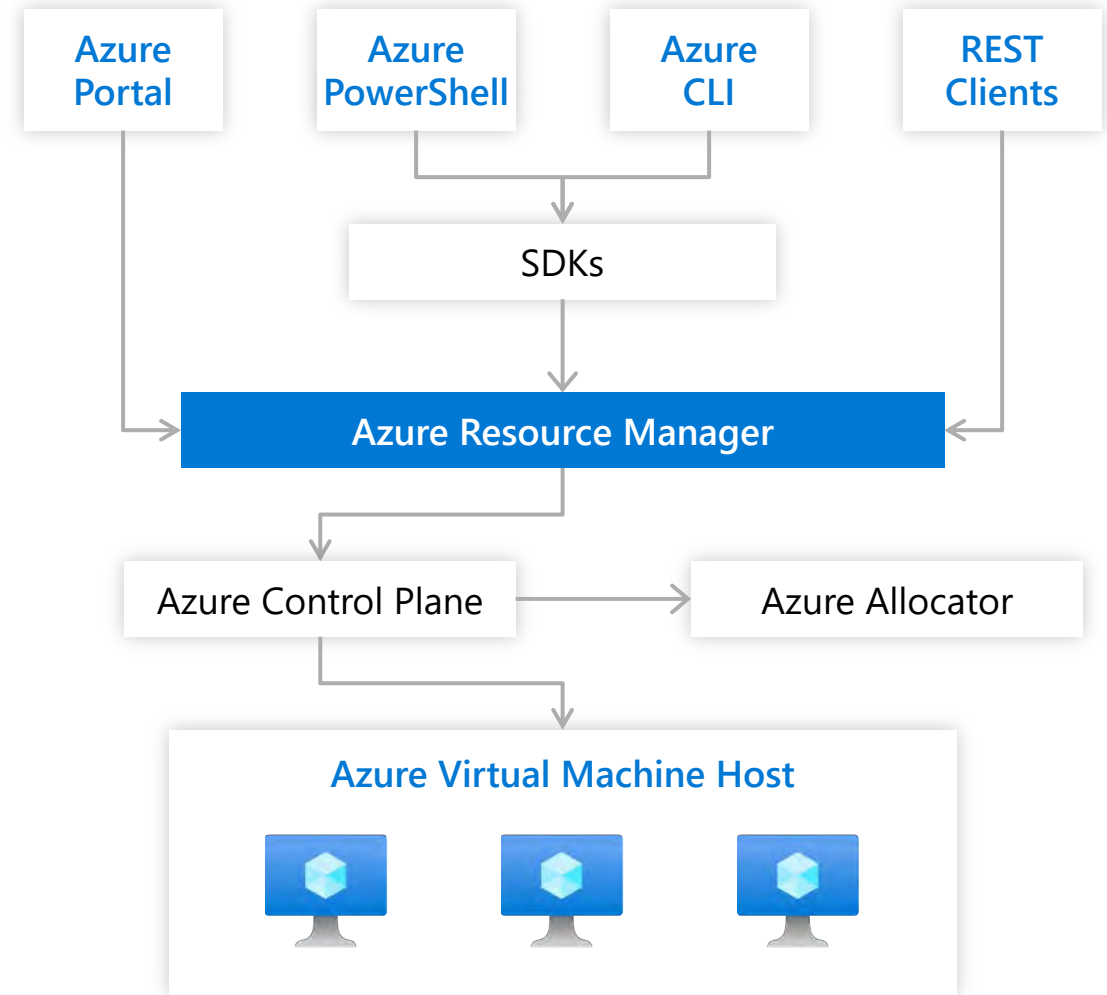
**Predict:** Usage per customer & across Azure using AI/ML

**Pre-Provision:** Prepare VMs ahead of time and keep them in a "dormant" state

**Energize:** Able to go from "dormant" to "active" in seconds

---

Up to 80% latency improvements expected for Windows VMs; VM ready < ~60s avg





# Thank you