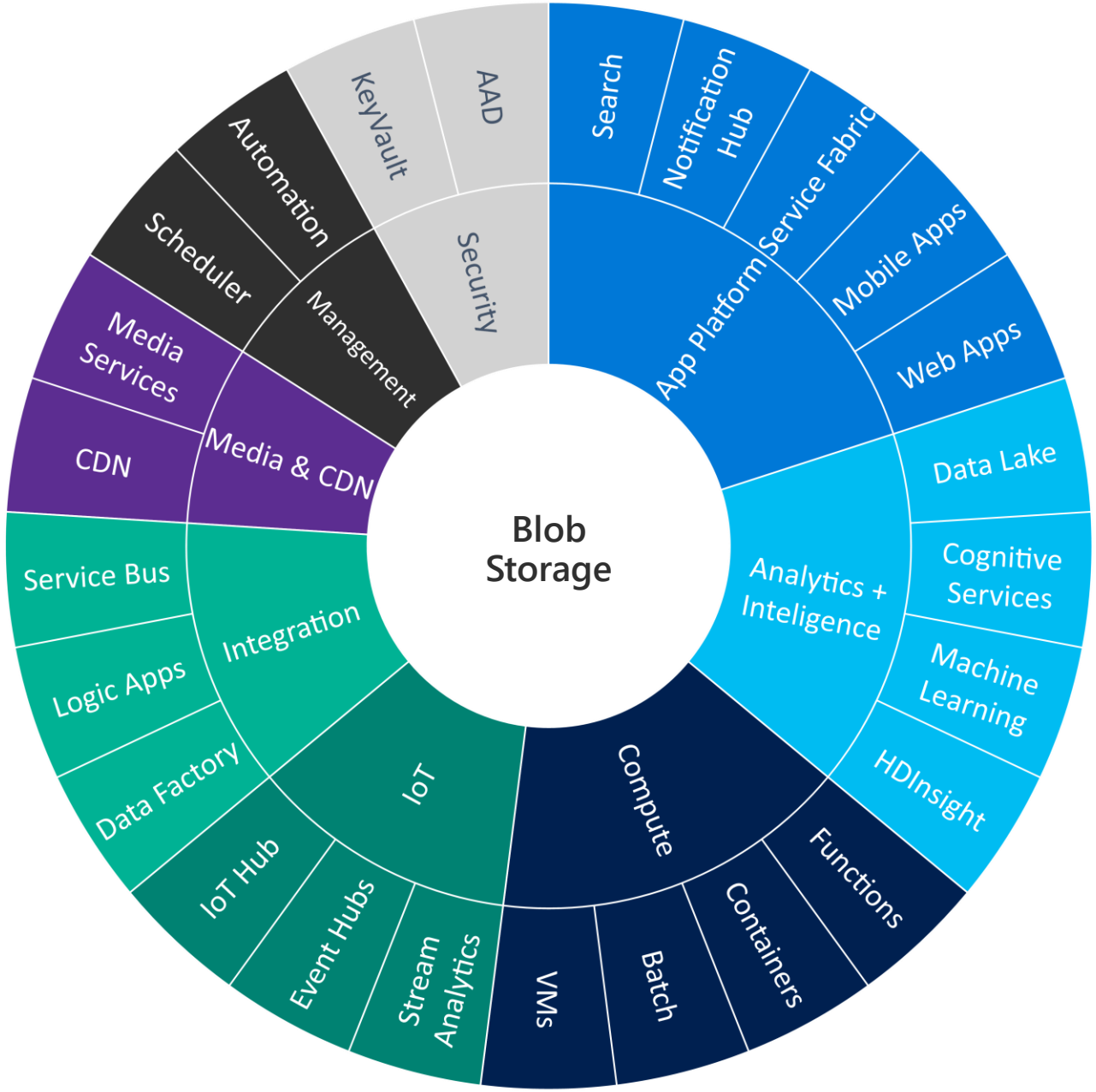
Azure Storage Essentials

Disclaimer

- Source https://docs.microsoft.com/en-us/azure
- https://channel9.msdn.com/Events/Build/2018/BRK3313
- Classic versus ARM (We will focus on ARM)


Blob Storage
App Platform
Search
Notification Hub
Service Fabric
Mobile Apps
Web Apps
Analytics + Inteligence
Data Lake
Cognitive Services
Machine Learning
HDInsight
Compute
Functions
Containers
Batch
VMs
IoT
IoT Hub
Event Hubs
Stream Analytics
Integration
Service Bus
Logic Apps
Data Factory
Media & CDN
Media Services
CDN
Management
Automation
Scheduler
Security
KeyVault
AAD

Disk Storage

Premium

Standard

Reliable, persistent, high performing storage for Virtual Machines

Object Storage

Azure Blobs

Secure, Scalable storage for unstructured data

File Storage

Azure Files

Azure NetApp Files

Lift and shift of legacy applications that require file shares to the cloud

Data Transport

Azure Import/Export

Azure DataBox

Move or migrate data into Azure

Hybrid Storage

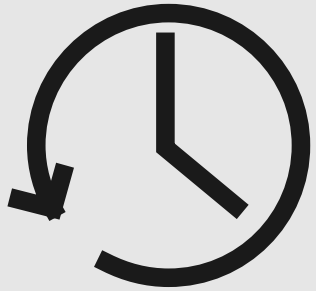
Azure StorSimple

Azure File Sync

Avere

Secure, intelligent data tiering between on-premises and cloud storage

Ultra SSD – The fastest disk in the public cloud



Low Latency

< 1 ms
Sub-millisecond

High Capacity

up to **64 TiB**
per disk

High throughput

up to **2000 MB/s**
per disk

High IOPS

up to **160000 IOPS**
per disk

Introducing Azure Data Box Edge

An AI-enabled edge computing appliance with network data transfer capabilities

Network Storage Gateway

Network data transport to Azure while retaining local access to files

Edge Compute

Use IoT edge compute modules to analyze, filter, and transform data as it moves to Azure

Cloud Managed

Easily manage your fleet from the Azure portal

Introducing Azure Data Box Gateway

A virtual network appliance that moves data in and out of Azure

Network Storage Gateway

Network data transport to Azure while retaining local access to files

Virtual Machine

Runs as a Hyper-V or VMware virtual machine on your hardware

Cloud Managed

Easily manage Data Box Gateway from the Azure portal

Storage Services

- Azure Blobs: A massively scalable object store for text and binary data.
- Azure Files: Managed file shares for cloud or on-premises deployments.
- Azure Queues: A messaging store for reliable messaging between application components.
- Azure Tables: A NoSQL store for schemaless storage of structured data (now part of Cosmos DB)

Blob Storage

- Azure Blob storage is object storage solution for the cloud.
- Blob storage is optimized for storing massive amounts of unstructured data, such as text or binary data.
- Serving images or documents directly to a browser.
- Streaming video and audio.
- Storing data for backup and restore, disaster recovery, and archiving.

Azure Files

- SMB (Server Message Block) based network file share
- Multiple VMS can share files with R and W
- URL accessible with (SAS token)
- AD auth is not supported

Queue Storage

- Azure Queue service is used to store and retrieve messages.
- Queue messages can be up to 64 KB in size,
- Queue is scalable - can contain millions of messages.
- Queues are generally used to store lists of messages to be processed asynchronously.

Table storage

- Moved to Cosmos (As part of multi model support)

Disk Storage

- Managed and unmanaged disk capabilities used by virtual machines
- OS disk is registered as a SATA drive (C:). Max Capacity of 2048 GB
- Data disks are registered as SCSI drives with a max capacity of 4095 GB

Types of Storage accounts

- Standard
  - Standard storage accounts on magnetic media
- Premium
  - Premium storage provides high-performance storage for page blobs, which are primarily used for VHD files. SSD based.

Types of storage accounts

Type of storage account	General-purpose Standard	General-purpose Premium	Blob storage, hot and cool access tiers
Services supported	Blob, File, Queue Services	Blob Service	Blob Service
Types of blobs supported	Block blobs, page blobs, and append blobs	Page blobs	Block blobs and append blobs

Securing Access to Storage

- Azure AD
  - RBAC
- Shared Access Signature (SAS)
- Public Access

Encryption

- Encryption at REST
  - SSE -Storage Service Encryption
    - Microsoft Managed Keys
    - Customer Managed Keys
      - Create a Key Vault and store your keys
      - Grant storage account access to Key Vault
      - Soft Delete and Do Not Purge (protection again ransomware)
- Client side encryption

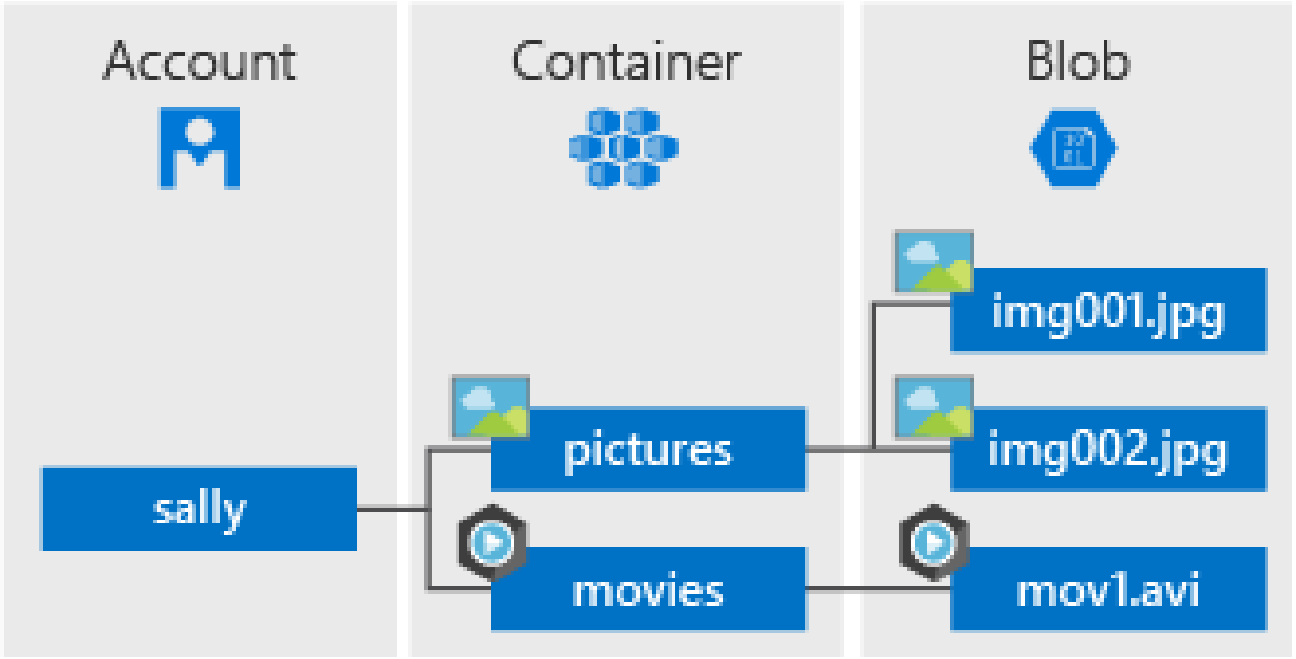
Replication

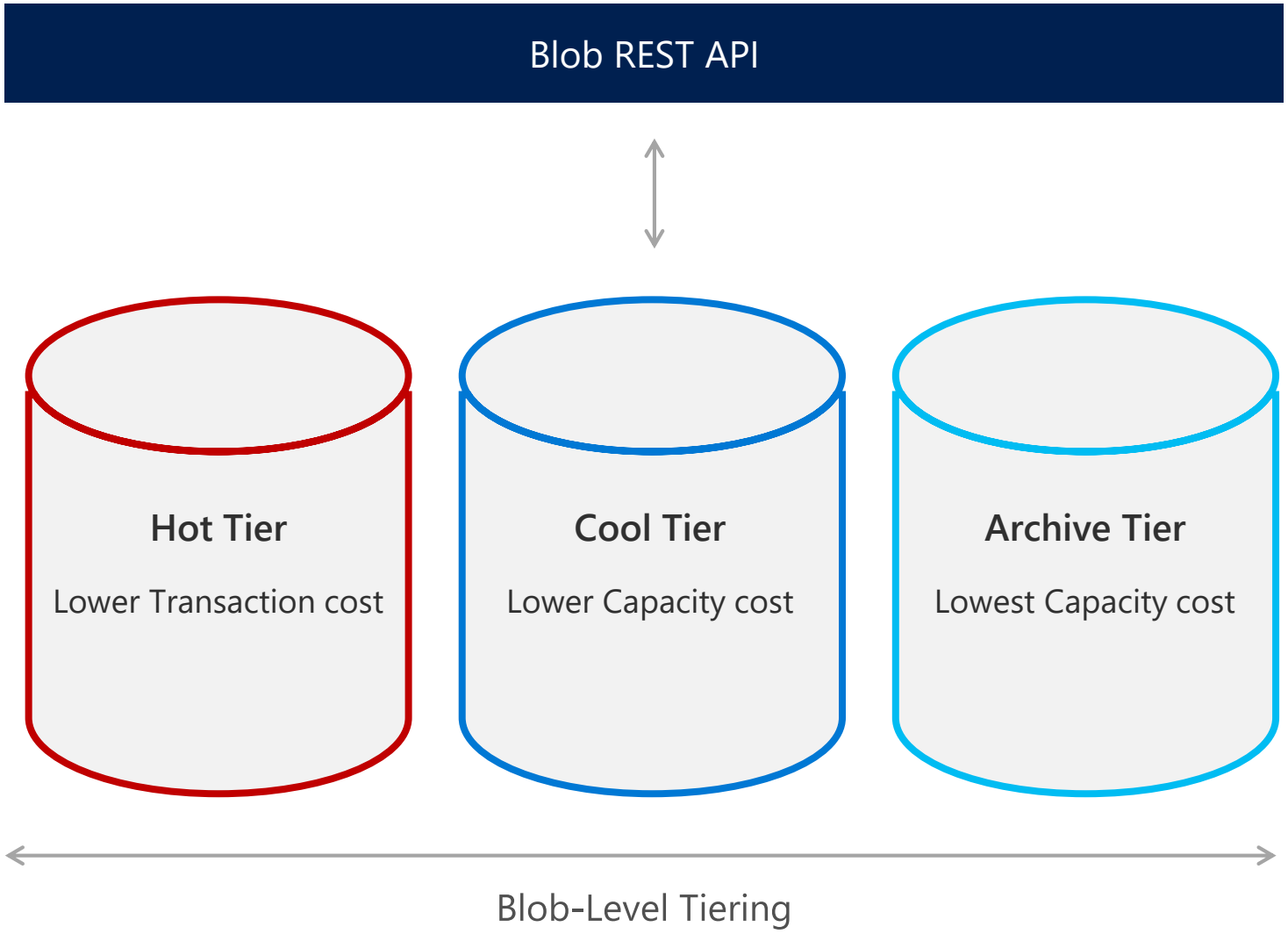
- Locally-redundant storage (LRS): The simplest, low-cost replication strategy that Azure Storage offers.
- Zone-redundant storage (ZRS): A simple option for high availability and durability.
- Geo-redundant storage (GRS): Cross-regional replication to protect against region-wide unavailability.
- Read-access geo-redundant storage (RA-GRS): Cross-regional replication with read access to the replica.

Tools

- AzCopy
- Azure Storage Explorer
- Cerebrata
- SDKs in C#, Java, Python, PHP, Ruby, C++

Blob Concepts




Blob REST API
Hot Tier
Lower Transaction cost
Cool Tier
Lower Capacity cost
Archive Tier
Lowest Capacity cost
Blob-Level Tiering

Pricing

- Storage Costs
  - depends on tier
- Data Access Costs
  - charge for reads on cool and archive tiers
- Transactions Costs
  - charged for all tiers, although higher for cool
- Geo-replication Costs
  - Only applies to RA-GRS and GRS
- Outbound Transfer Costs
  - Billing for bandwidth usage
- Changing the Storage Tier
  - Moving from hot to cool will incur writing charge for cool tier

Azure Blob

	Hot	Cool	Archive
	Frequently accessed data	Less frequently accessed data	Rarely accessed data
PER TB PER MONTH	$18.40	$10.00	$2.00
PER 10K WRITE OPERATIONS	$0.05	$0.10	$0.10
RETRIEVAL TIMES	Immediate	Immediate	Hours
USE CASE	Cloud native application data storage	Repository for server backups	Medical records archiving
EXAMPLES			

Others

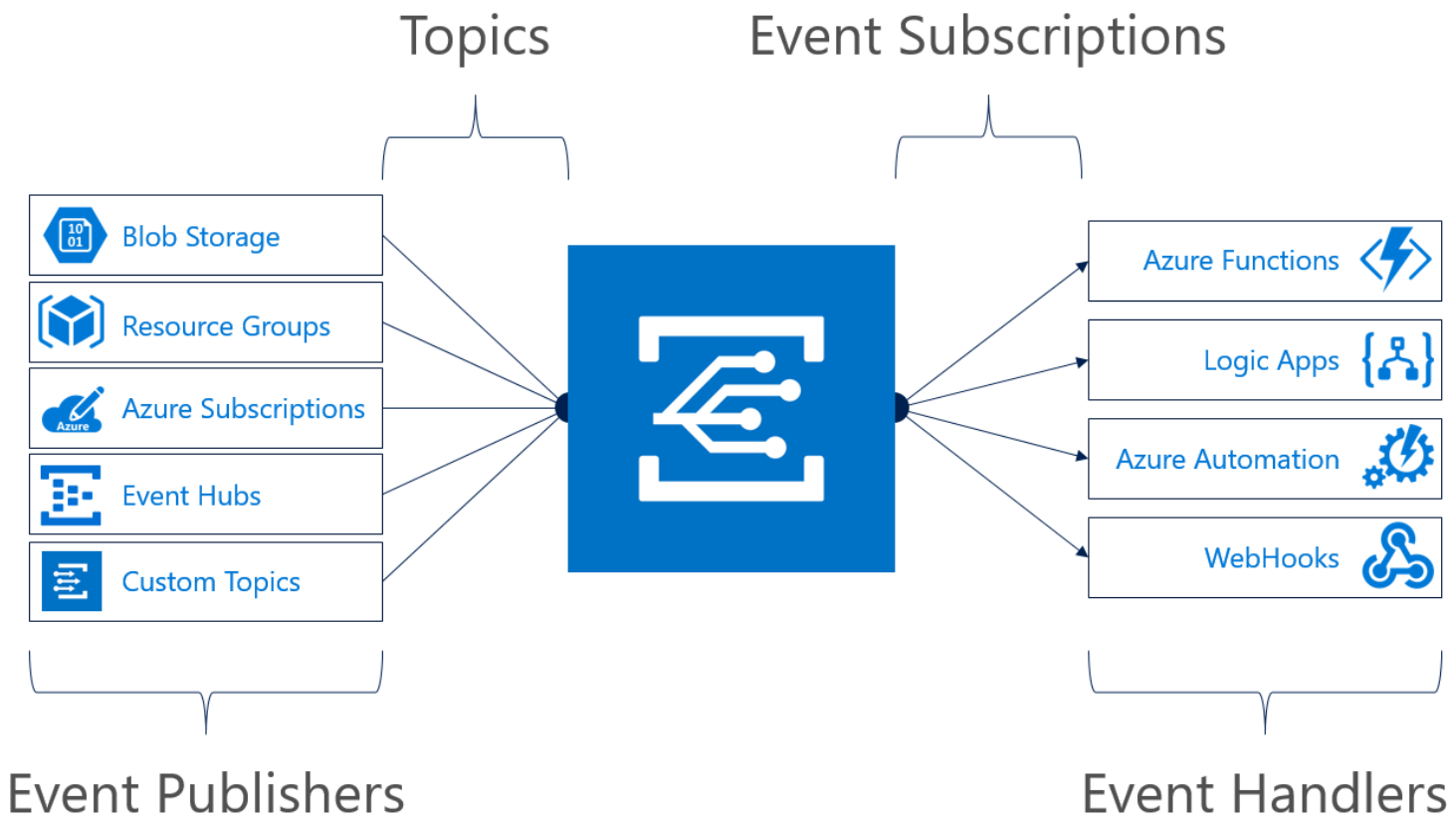
- Static Web Site Support
- Soft Delete
- Data Life Cycle Management

Scalability Targets

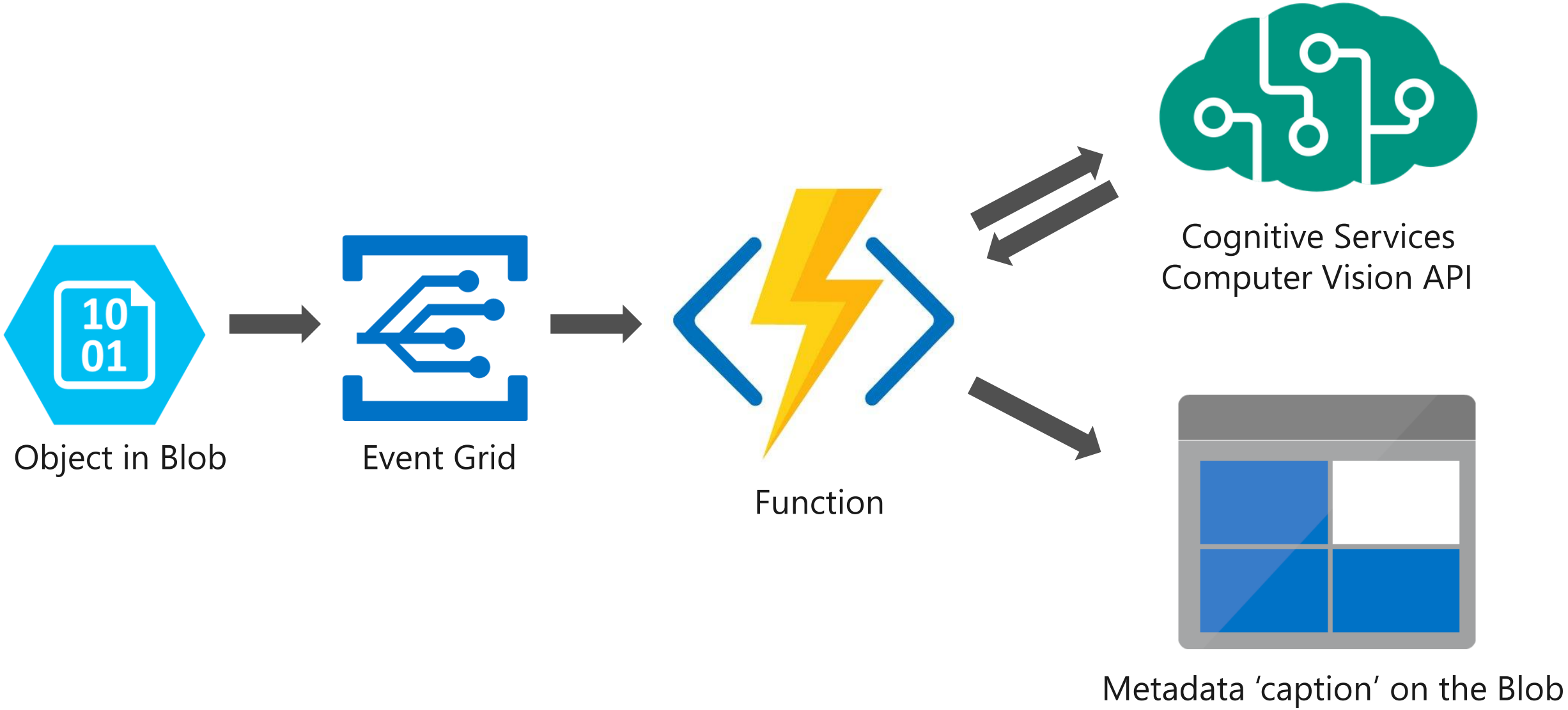
Resource	Default	New Limit
Max capacity for Blob storage accounts	500 TB	5PB (10x increase)
Max TPS/IOPS for Blob storage accounts	20K	50K (2.5x increase)
Max ingress for Blob storage accounts	5-20 Gbps (varies by region/ redundancy type)	50Gbps (up to 10x increase)
Max egress for Blob storage accounts	10-30 Gbps (varies by region/ redundancy type)	50Gbps (up to 5x increase)

Tracking changes without polling

Blob created
Blob Deleted



Serverless image processing with Cognitive Services



Concurrency

- Optimistic (check if data has changed since last read - ETAG)
- Pessimistic (Acquire a Lease)
- Azure Files Concurrency is based on the protocol
  - SMB
    - Read, Write, Shared Read, Shared Write etc.
  - REST
    - Respects the SMB mode specified earlier.

Storage Analytics

- Logging
  - Successful requests
  - Failed Requests
  - Requests using SAS
- Metrics data
  - Transaction metrics – API operations, errors
  - Capacity metrics – Amount of storage used

Move Data into Storage

- AzCopy
- Azure Import / Export Service
- Data Box

Blob Types

- Block
  - Ideal for storing text and binary files
- Append
  - Ideal for logs
- Page
  - Frequent random read / write operation – Ideal for OS and data disks. Standard and Premium
  - Azure Disks are page blocks persisted as VHD
  - Snapshot feature

DR

- VM Disks – Use the Azure Backup service
- Block blobs –Create a snapshot of each block blob, or copy the blobs to another storage account in another region
- Files – use AzCopy or Azure PowerShell to copy your files to another storage account in another region.

Storage Security

- Firewall
- VNET - Virtual Service Endpoints
- SAS (IP Filtering)

Using CDN with Storage

- Deliver content efficiently to the users
  - Better performance
  - Large scaling to better handle instantaneous loads
  - Serving users from edge locations
- Support for HTTPS for custom domain names
- Note - CDN does not respect SAS (need to use Cache response headers)

Azure Search Integrated with Storage

- Search as a Service
- Managed Infrastructure (availability, scale)
- Automated Indexing of Blob storage

Shared Access Signature

- Specify permissions read / write / delete
- Specify Source IP address where the requests will originate from.
- Specify protocol to be used e.g HTTP/HTTPS
- Types
  - Account (assigned
  - Service (Blob)
- Policy
  - Additional level of control
  - Group SAS
  - Revoke control after the SAS has been issued

Managed Disks

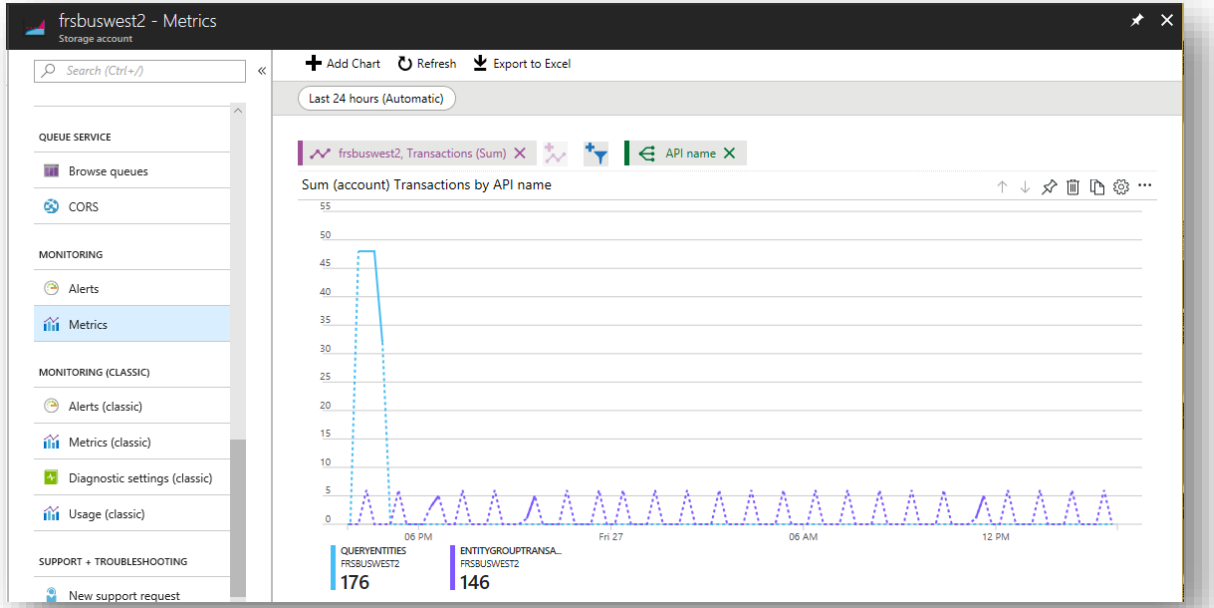
- Better reliability by ensuring that the disks are sufficiently isolated
- Handles storage for you behind the scenes
  - Previously, you had to create storage accounts to hold the disks
  - Copy custom images to multiple storage accounts
- Create 10,000 VM Disks in subscription
- Managed Disk Snapshots (different from incremental snapshots)
  - Independent of source disk
  - Can be used recreate new managed disks

Images versus Snapshot

- Image is a generalized version of a VM Disk
- Use it to create a new VM
- Snapshot is a point in time version of disk
  - Snapshots only apply to a single disk
  - What happens when your machine has multiple disks?

Storage with Azure Monitor

- Monitoring Charts and Alerts
- Ingest Analytics



Azure Data Lake

- Unlimited Storage - no limit on account size, file sizes or amount of data stored
- HDFS compatible (so is Azure Storage Blob Container)
- Tuned for Big Data Analytics

Azure Storage Durability & Availability

Strong Consistency

- 3 replicas + erasure coding

Data Integrity

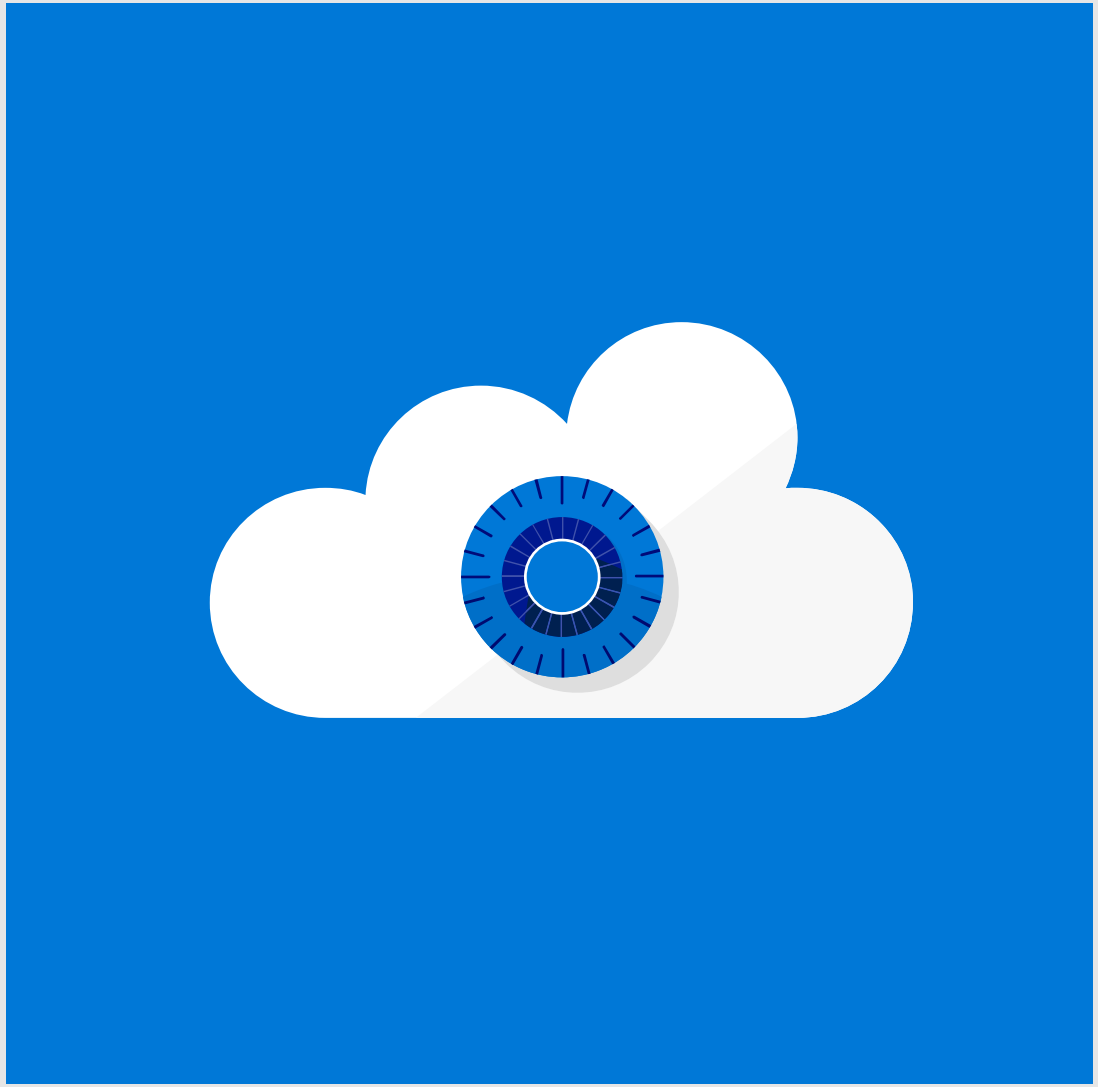
- MD5 hash on ingress/egress
- CRC checksum & “bit rot” protection

Disaster Recovery (BCDR)

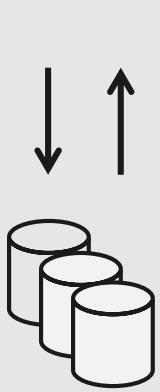
- Geo-redundant storage (GRS)

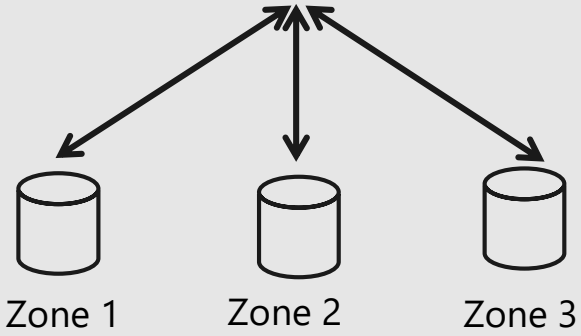
High Availability (HA)

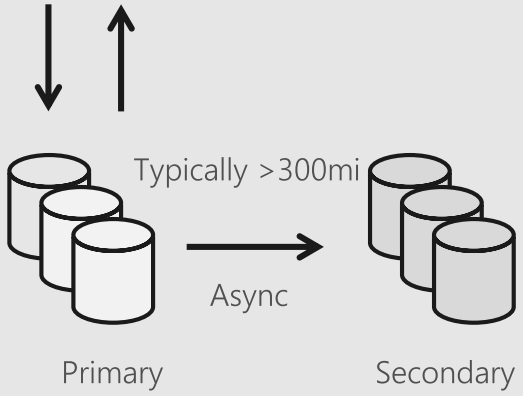
- 99.9% availability SLA
- 99.99% for reads with RA-GRS

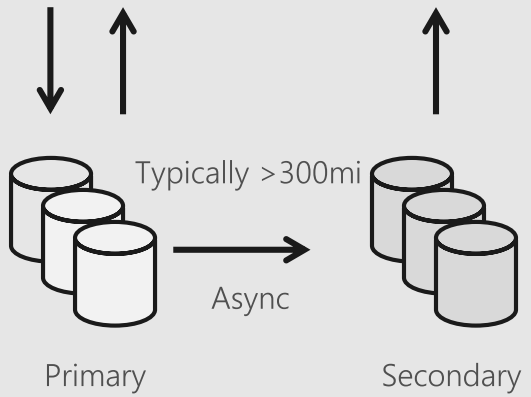


Azure Storage Replication Options









LRS

3 replicas, 1 region

Protect against disk, node, rack failures

Write is ack'd when all replicas are committed

Superior to dual-parity RAID

11 9s of durability

ZRS

3 replicas **across 3 Zones**

Protect against disk, node, rack and **zone** failures

Synchronous writes to all 3 zones

12 9s of durability

Available in 8 regions

GRS

6 replicas, 2 regions (3/region)

Protects against major regional disasters

Asynchronous to secondary

16 9s of durability

RA-GRS

GRS + Read access to secondary

Separate secondary endpoint

RPO delay to secondary can be queried

Roadmap: Customer controlled failover to GRS location
Geo-redundant ZRS Storage

Azure Backup – Checkpoint your Disks and Files

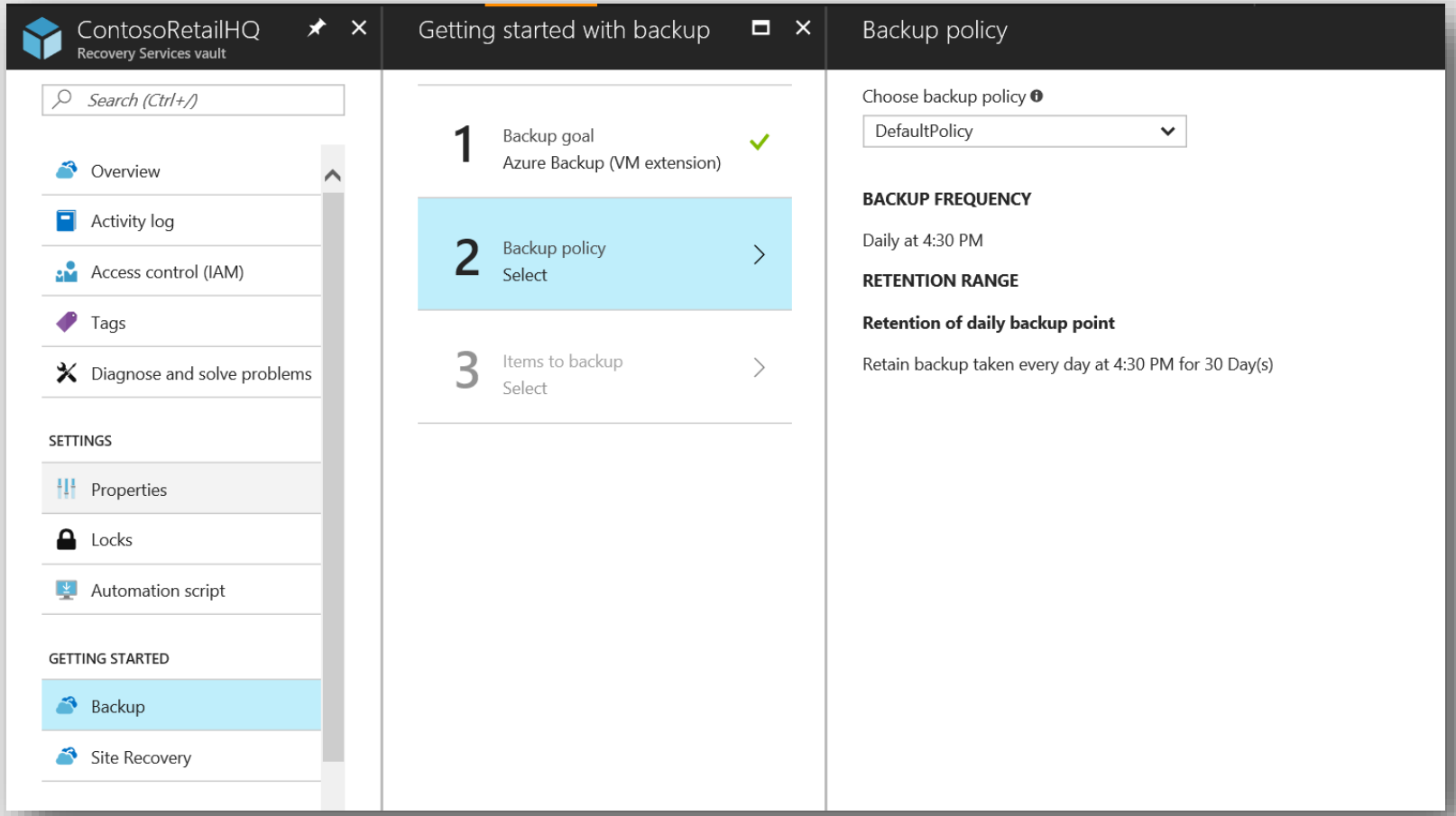
Data protection

Fully managed service - Deeply integrated with Disks and Files

Protect against ransomware and human errors

Meet compliance goals with data retention and encryption

Start backup in minutes and restore data quickly when needed



Soft Delete - Checkpointing for Blob Storage

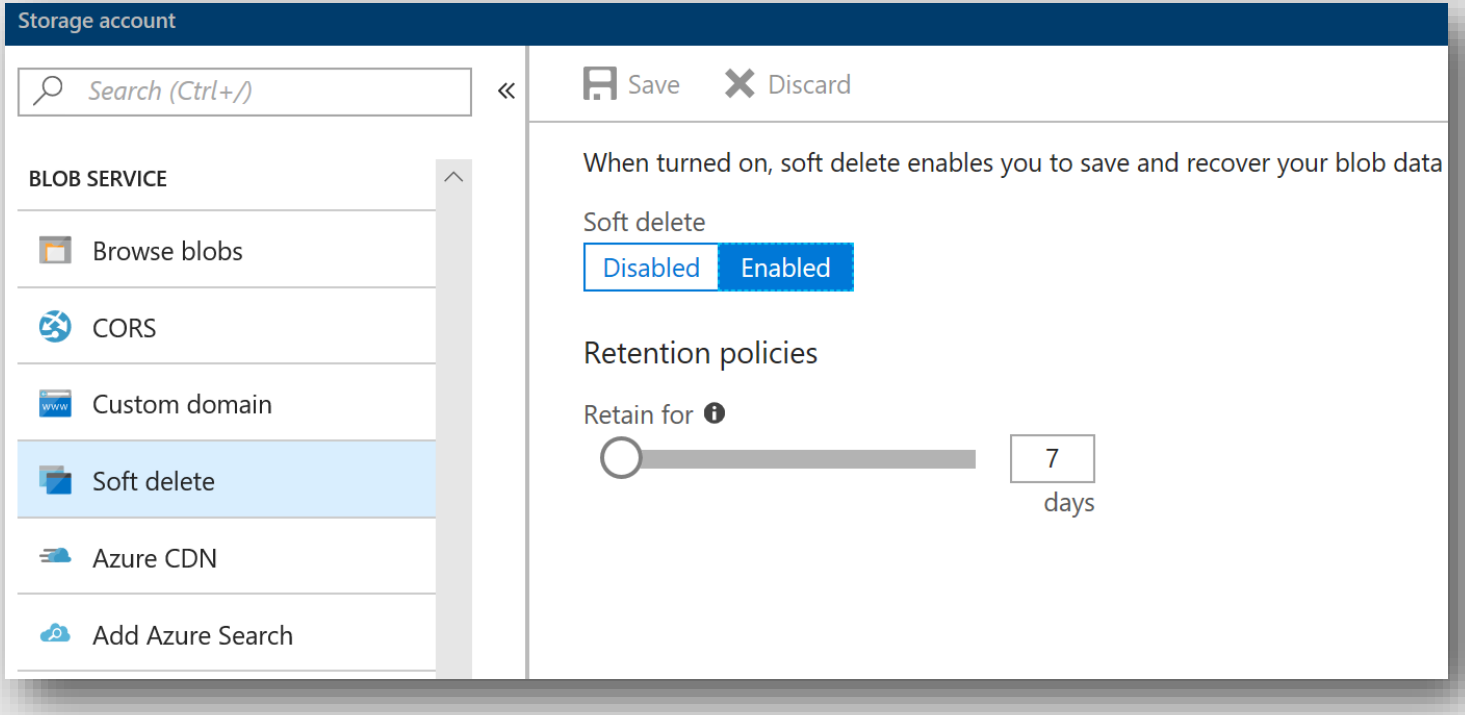
Data protection

Protects against accidental deletes and overwrites of blobs

Configurable retention period for soft deleted data up to 1 year

Overwritten blobs stored as soft deleted snapshots

Fully Backwards compatible with existing code



Security & Compliance

Encryption at Rest

- Always on with MS keys
- Customer supplied keys –

Data security

Your storage account is currently encrypted with Microsoft managed key by default. You can choose to use your own key.

☑ Use your own key

Encryption key

○ Enter key URI

◉ Select from Key Vault

---

* Key Vault

cbrookskeyvault >

---

* Encryption key

DataEncryptionKey >

'cbrookscustomkey' will be granted access to the selected key vault. Both soft delete and purge protection will be enabled on the key vault and cannot be disabled. Learn more

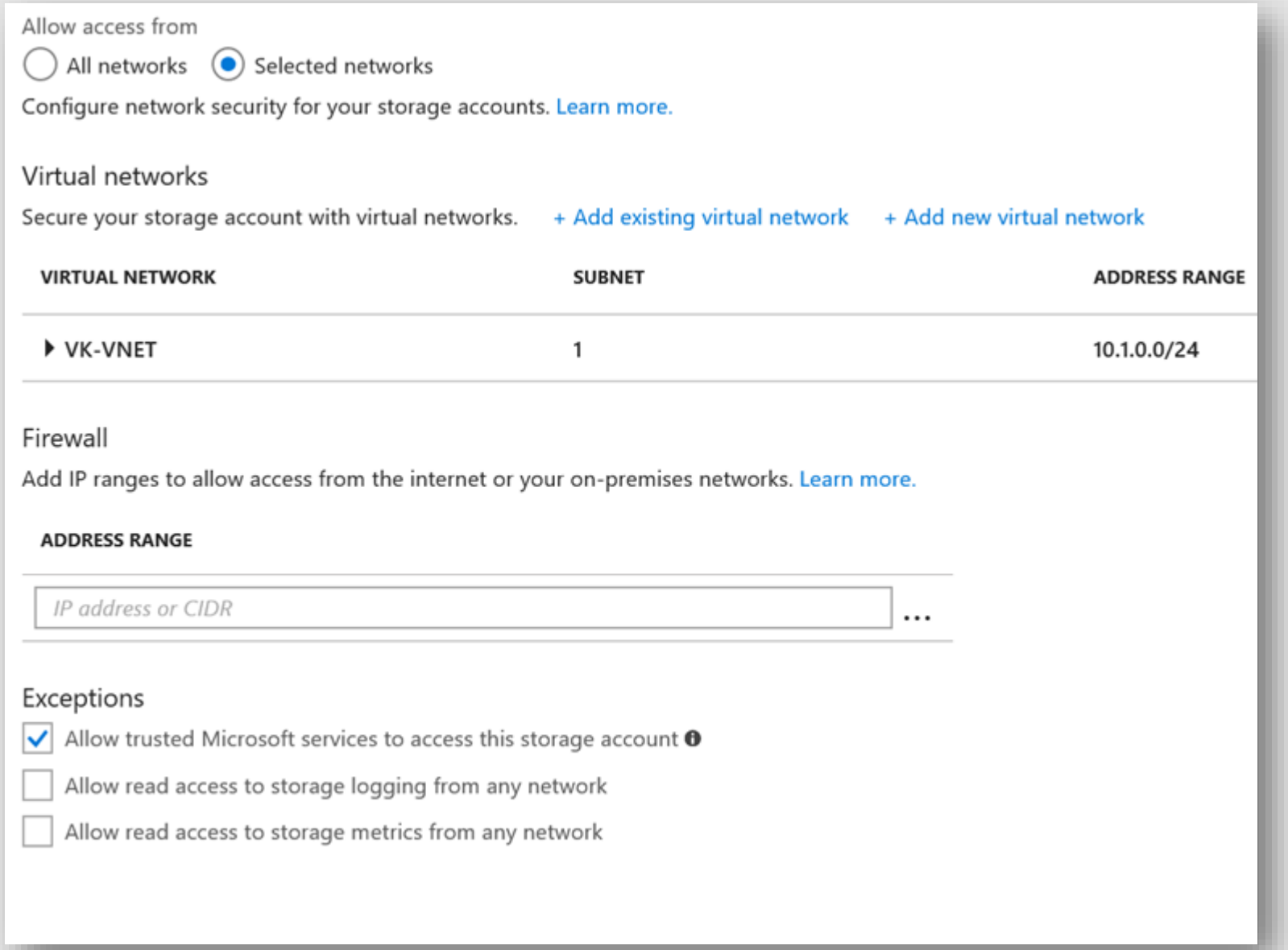
Security & Compliance

Encryption at Rest

- Always on with MS keys
- Customer supplied keys

Firewalls and Virtual Networks

- Restrict access from public internet



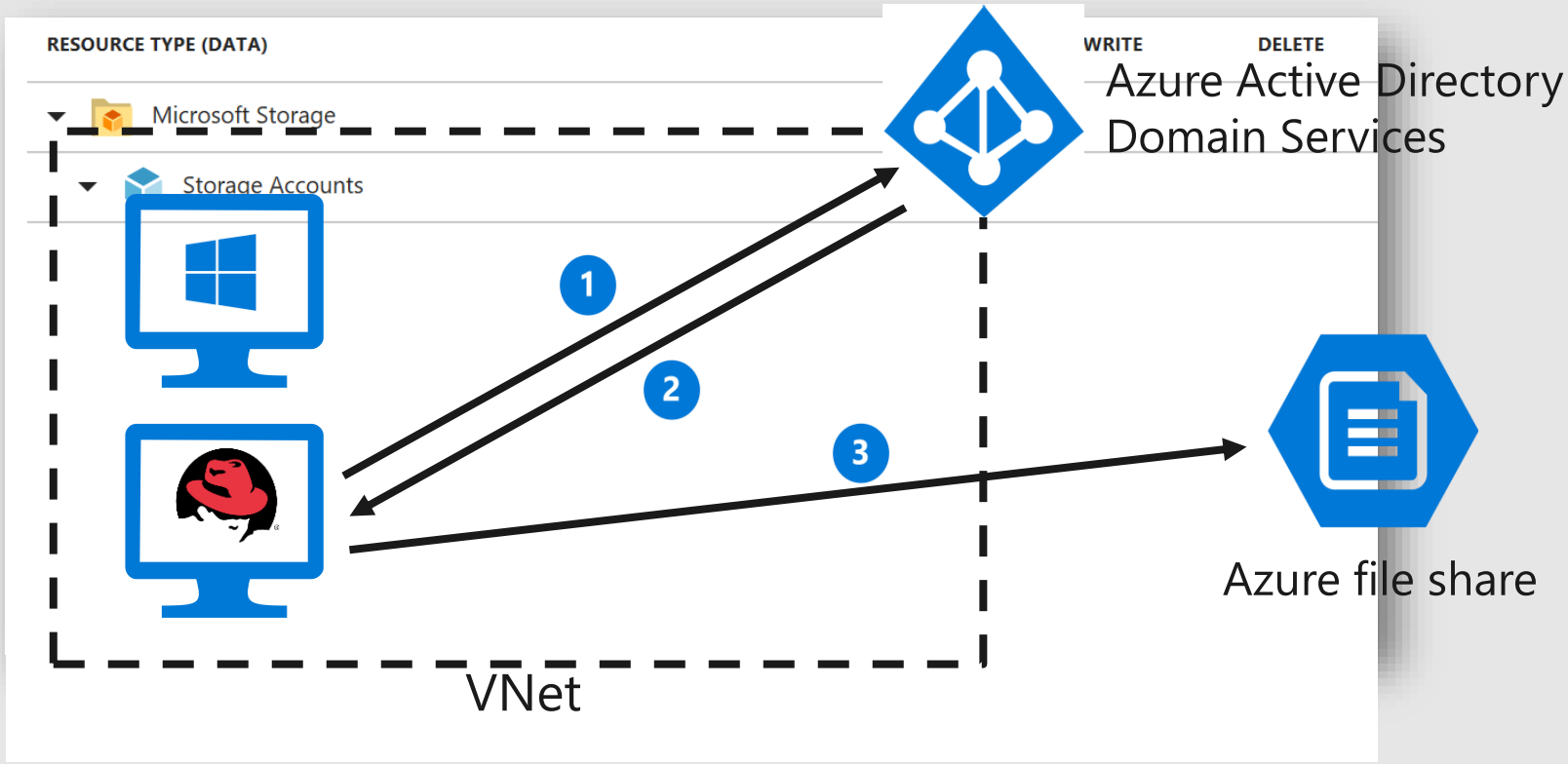
Security & Compliance

Data security

Encryption at Rest
- Always on with MS keys
- Customer supplied keys –

Firewalls and Virtual Networks
- Restrict access from public internet –

AAD, OAuth and RBAC
- AAD creds for Auth instead of Shared key/SAS
- Full RBAC at container level
- AAD integration for Files



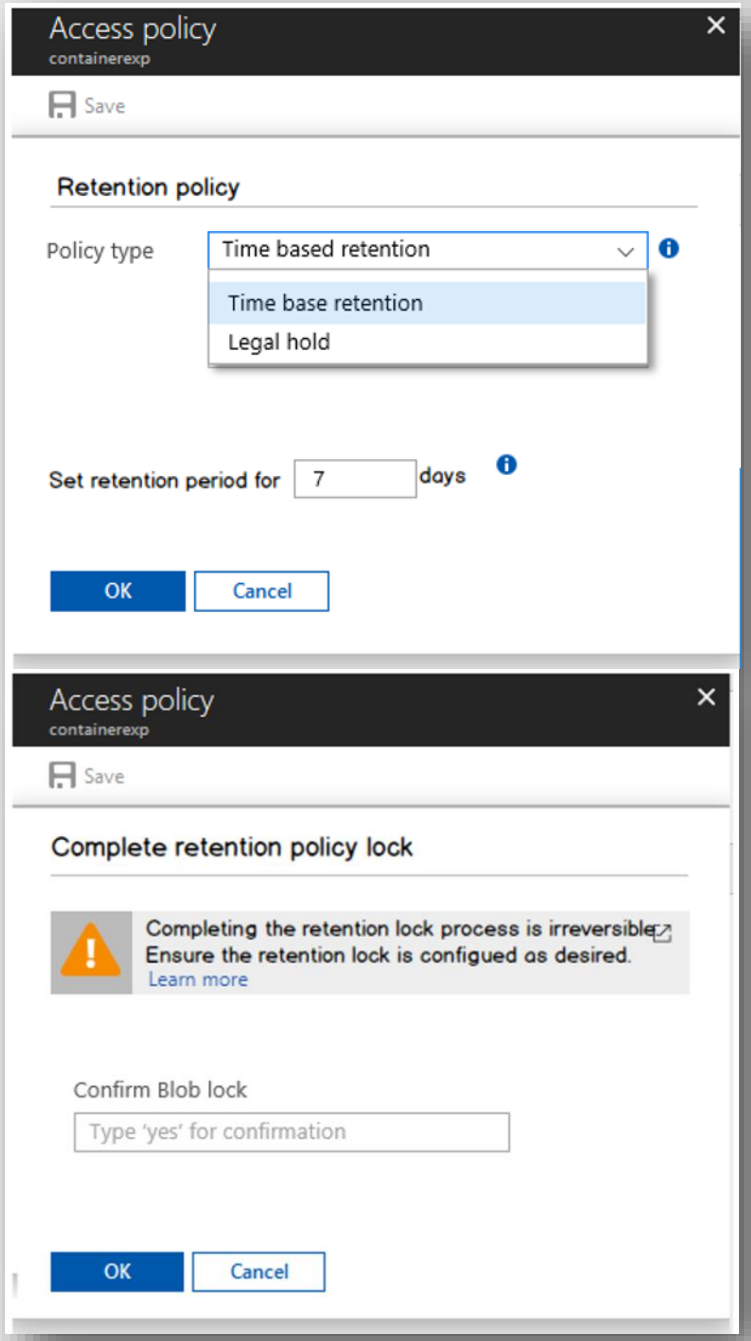
Security & Compliance

Encryption at Rest
- Always on with MS keys
- Customer supplied keys

Firewalls and Virtual Networks
- Restrict access from public internet

AAD, OAuth and RBAC
- AAD creds for Auth instead of Shared key/SAS
- Full RBAC at container level
- AAD integration for Files

Immutable Blob Storage
- SEC 17a-4(f) compliant
- Across all Storage tier



Support for time-based retention

Container level configuration

RBAC support and policy auditing

Blobs cannot be modified or deleted for N days

Support for legal holds with tags

Container level configuration

Blobs can not modified or deleted when legal hold is set

Support for all Blob tiers

Applies to hot, cool and cold data

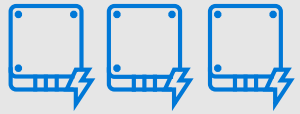
Policies retained when data is tiered

Disk Scalability prior to today

Scalability

Standard HDD

Standard SSD



Premium SSD

Single disk value

	Low-cost storage	Consistent performance	High performance
Max Size	4 TiB	4 TiB	4 TiB
Max IOPS	Up to 500	Up to 500	7,500
Max BW	Up to 60 MBps	Up to 60 MBps	250 MBps

Massive Azure Files Scalability Improvements

Scalability

Limited Preview

	Standard Files Today	Standard Perf Increase	Premium Files (SSD)
Average latency	Double digits ms		
Max IOPS	1,000 IOPS		
Max bandwidth	60 MBps	8-10X	80-100X
Max Share Size	5TB		

Premium Blob Storage

Scalability

Azure object storage backed by solid-state drives

High throughput and low latency object storage

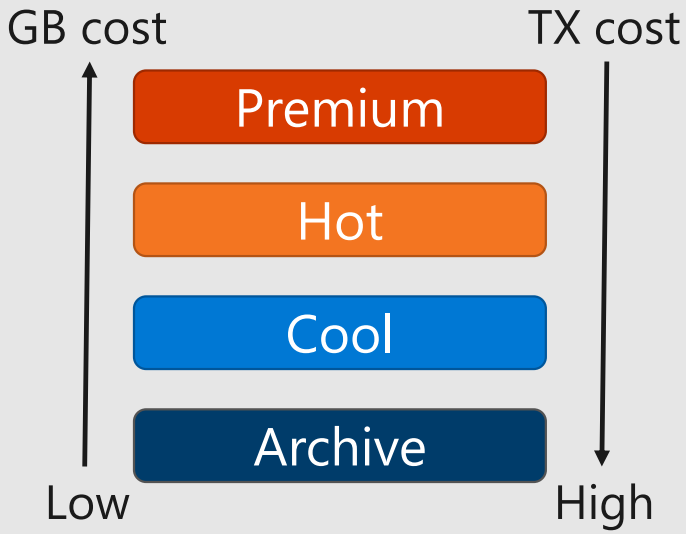
- Lower latency
- Consistent latency

Block blobs and Append blobs

Consistent API

Locally Redundant Storage

Object-level tiering to Hot is on the roadmap



Massive Azure Blob Scalability Improvements

Scalability

Limited Public Preview

	Azure Blobs Today	High Throughput Blobs	Premium Blobs
Latency	Double digit ms latency		
Account Cap	5PB		
Account IOPS	50,000		
Per Object tps	500		
Per Object Write	60MB/s		

1 Default limit is 2PB. 5PB on request.

Access Patterns – Cloud Approach

Multi modal across unified data storage layer

- Multiple protocols access the same data simultaneously – REST, HDFS, NFS, SMB,…

Simplified data management

- Single unified data layer – No silos, no data movement
- Common durability, security, scalability, etc.

Cost Effective

- Can tier cooler data to cooler tiers or archive it
- Higher utilization and only copy of the data


Azure Data Lake Storage Gen2

Access patterns

HDInsight

HORTONWORKS

cloudera

databricks

High performance with massive scale for analytics workloads

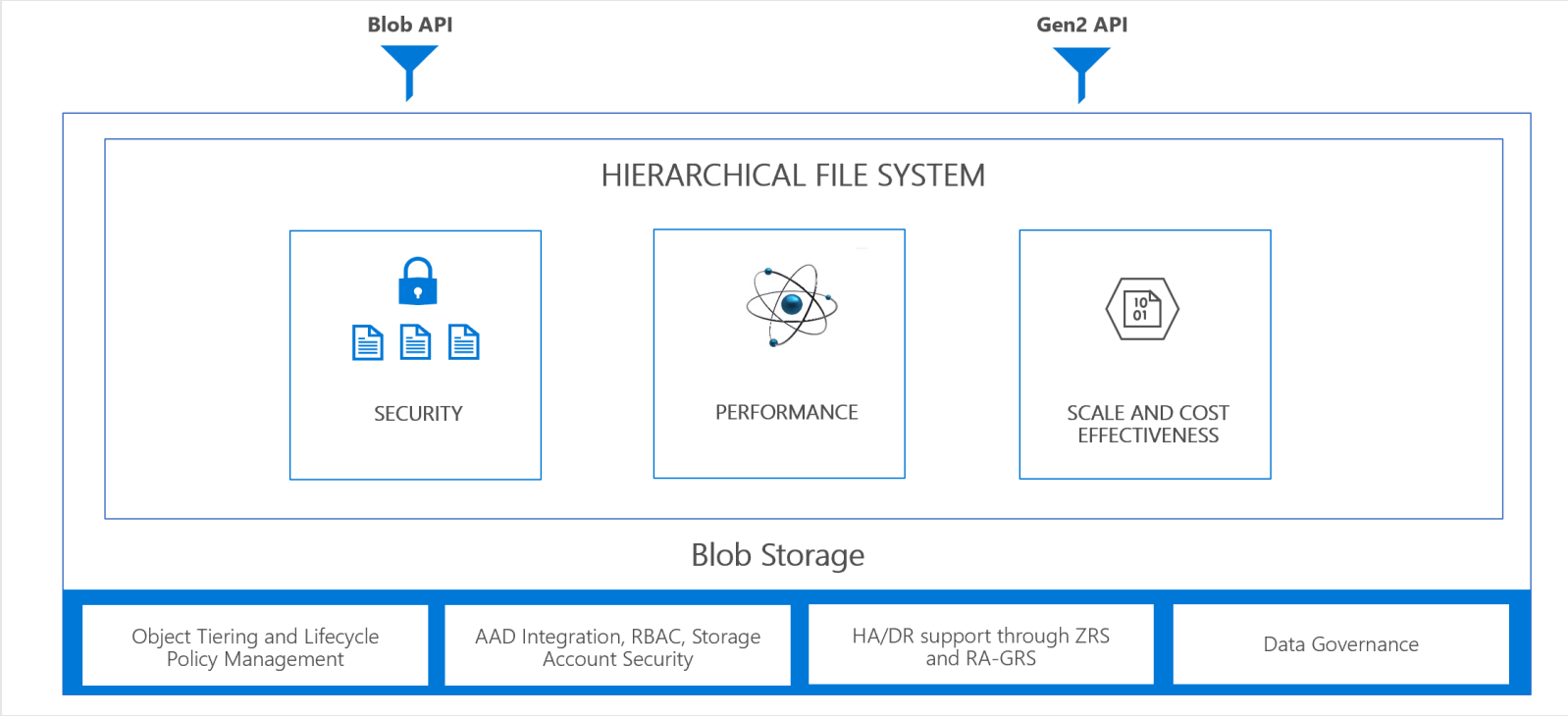
Enterprise grade security – AAD integration with fine-grained ACLs, VNet service endpoints & data encryption

Hierarchical namespace with Hadoop Filesystem compatibility

Fully integrated with all major analytics engines

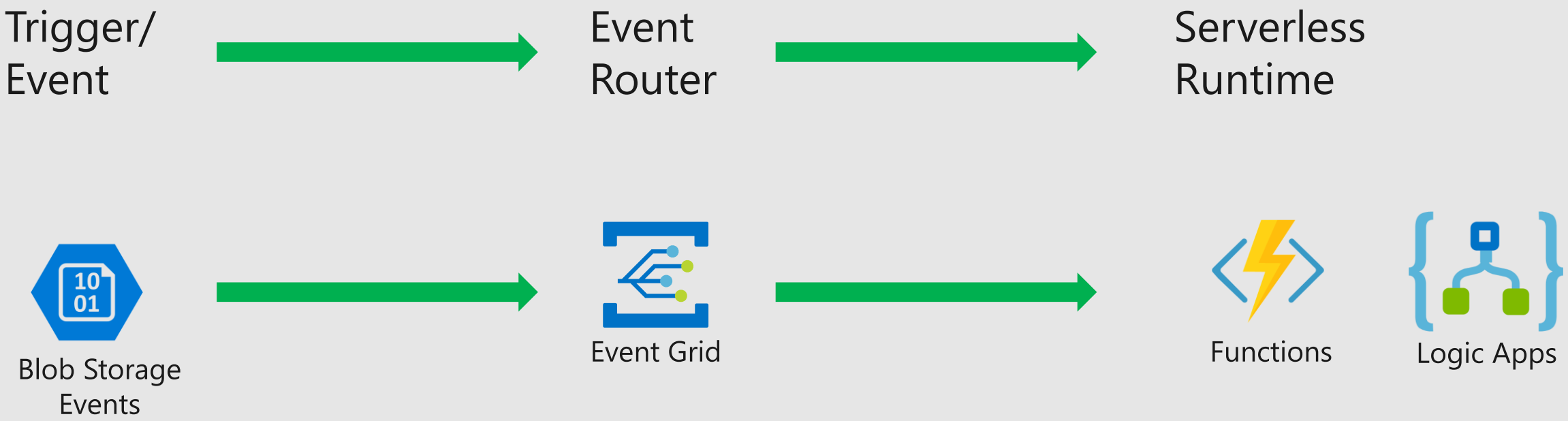
Eliminates data silos with object storage and filesystem endpoints against the same data

Built on Blob Storage - Global availability, data lifecycle policies, multiple redundancy options, lower TCO



https://azure.microsoft.com/en-us/services/storage/data-lake-storage/

Serverless Computing with Data as a trigger



Demos