

Lab 6 - Azure Storage

Blob, Queue, Table Storage

Materiały na kolokwium - Część 2

7 grudnia 2025

Spis treści

1 Azure Storage Account

1.1 Podstawy

Co to jest Storage Account?

Storage Account to namespace dla wszystkich usług storage w Azure.

Zawiera 4 usługi:

- **Blob Storage** - object storage (pliki binarne)
- **Queue Storage** - message queue
- **Table Storage** - NoSQL key-value store
- **File Storage** - SMB file shares (network drives)

Unikalna nazwa:

- Globalnie unikalna w całym Azure
- 3-24 znaki
- Tylko małe litery i cyfry
- Przykład: mystorageaccount123

Endpoints:

- Blob: <https://{{account}}.blob.core.windows.net>
- Queue: <https://{{account}}.queue.core.windows.net>
- Table: <https://{{account}}.table.core.windows.net>
- File: <https://{{account}}.file.core.windows.net>

1.2 Typy Kont Storage

General-purpose v2 (GPv2)

Charakterystyka:

- Rekomendowany typ (current standard)
- Obsługuje wszystkie usługi storage
- Obsługuje wszystkie access tiers (Hot, Cool, Cold, Archive)
- Lowest prices
- Best dla większości scenariuszy

Use case: Default choice dla nowych projektów

Premium Storage

Charakterystyka:

- SSD-based storage
- Wyższa wydajność (niższa latencja)
- Wyższe IOPS
- Droższe

3 typy Premium:

- **Premium Block Blobs** - dla block i append blobs
- **Premium File Shares** - dla Azure Files
- **Premium Page Blobs** - dla page blobs (VHD)

Use case:

- High-transaction workloads
- Low-latency requirements
- Database files
- VMs (Premium Page Blobs)

1.3 Replication (Redundancja)

LRS - Locally Redundant Storage

Charakterystyka:

- 3 synchroniczne kopie w jednym datacenter
- Najtańsza opcja
- 11 nines durability (99.99999999% w roku)

Ochrona przed:

- Disk failure
- Rack failure
- Hardware problems

NIE chroni przed:

- Datacenter failure (fire, flood, earthquake)
- Regional disaster

Use case:

- Dev/test environments
- Dane które można odtworzyć
- Gdy nie potrzebujesz high availability
- Optymalizacja kosztów

ZRS - Zone Redundant Storage

Charakterystyka:

- 3 synchroniczne kopie w 3 różnych availability zones
- 12 nines durability (99. 999999999%)
- Wyższy koszt niż LRS

Availability Zones:

- Fizycznie oddzielone datacenter w tym samym regionie
- Niezależne zasilanie, chłodzenie, networking
- Odległość: kilometry od siebie

Ochrona przed:

- Wszystko co LRS
- Datacenter failure w regionie

NIE chroni przed:

- Regional disaster (cały region down)

Use case:

- Production workloads
- High availability w regionie
- Gdy potrzebujesz consistency i wysoką dostępność

GRS - Geo Redundant Storage

Charakterystyka:

- LRS w primary region + asynchroniczna kopia do secondary region
- 16 nines durability (99. 99999999999999%)
- Secondary region: setki kilometrów od primary
- Paired regions (z góry określone przez Azure)

Asynchronous replication:

- Opóźnienie: zwykle < 15 minut
- RPO (Recovery Point Objective): 15 min data loss możliwy
- Automatic replication

Secondary region access:

- Default: read-only (nie możesz czytać)
- Po failover: becomes primary (read-write)
- RA-GRS: Read-Access GRS (możesz czytać ze secondary)

Ochrona przed:

- Wszystko co LRS
- Regional disaster
- Catastrophic failure całego regionu

Use case:

- Business-critical data
- Disaster recovery requirements
- Compliance (data residency in multiple regions)

GZRS - Geo-Zone Redundant Storage

Charakterystyka:

- ZRS w primary region + GRS (async copy do secondary)
- Najwyższa dostępność i durability
- 16 nines durability
- Najdroższe

Kombinacja:

- ZRS protection w primary region
- GRS protection dla disaster recovery
- Best of both worlds

Use case:

- Maximum consistency, availability, durability
- Mission-critical applications
- Financial services, healthcare
- Gdy nie możesz stracić danych

Ważne!

Paired Regions:

Azure ma z góry określone pary regionów dla GRS:

- West Europe North Europe
- East US West US
- Southeast Asia East Asia

NIE możesz wybrać custom secondary region!

Failover:

- Automatic failover: NIE (musisz ręcznie zainicjować)
- Manual failover: TAK (w Azure Portal)
- Czas: 1 godzina (RTO - Recovery Time Objective)

2 Blob Storage

2.1 Podstawy

Blob Storage - Object Storage

Definicja: Massively scalable object storage dla unstructured data

Unstructured data:

- Dane bez określonego formatu/schematu
- Binary data
- Przykłady: obrazy, wideo, audio, dokumenty, logi

Zastosowanie:

- Serving images/documents do przeglądarki
- Storing files for distributed access
- Streaming video i audio
- Backup i disaster recovery
- Archiwizacja danych
- Data for analysis (big data)
- Static website hosting

Skala:

- Exabytes of data (praktycznie unlimited)
- Miliony żądań na sekundę
- Global distribution

2.2 Hierarchia Blob Storage

Hierarchia

3-poziomowa struktura:

Storage Account → Container → Blob

Storage Account:

- Top-level namespace
- Unikalna nazwa globalnie
- Authentication i authorization level

Container:

- Logiczne grupowanie blobów (jak folder, ale flat)
- Nazwa: 3-63 znaki, lowercase, cyfry, dash
- Public access level (Private / Blob / Container)
- Unlimited liczba blobów
- Unlimited liczba containerów w account

Blob:

- Actual file/object
- Max size zależy od typu (do 4. 75 TB dla block blobs)
- Unique URL: <https://{{account}}.blob.core.windows.net/{{container}}/{{blob}}>

Virtual directories:

- Blob Storage jest FLAT (nie ma prawdziwych folderów)
- Możesz używać "/" w nazwie blobu: `folder1/folder2/file.txt`
- Azure Portal pokazuje to jako foldery (wirtualne)

2.3 Typy Blobów

Block Blobs

Charakterystyka:

- Najbardziej popularny typ
- Składa się z bloków (chunks)
- Max 50,000 bloków
- Max rozmiar bloku: 4000 MiB
- Max total size: 4.75 TB

Uploading:

- Można uploadować bloki równolegle (paralelizacja)
- Po upload wszystkich bloków → commit block list
- Atomic operation (albo wszystkie bloki albo żadne)
- Resumable uploads (jeśli fail, można wznowić)

Optimized for:

- Uploading large files efficiently
- Streaming scenarios
- Sequential reading

Use case:

- Obrazy, wideo, audio
- Dokumenty (PDF, Word, Excel)
- Backups
- Application data files
- 95% przypadków użycia

Append Blobs

Charakterystyka:

- Zoptymalizowane pod operacje append (dopisywanie)
- Składa się z bloków (append blocks)
- Max 50,000 bloków
- Max rozmiar bloku: 4 MiB
- Max total size: 195 GB

Operations:

- Append - TAK (add data at the end)
- Update existing blocks - NIE
- Delete blocks - NIE
- Read - TAK

Optimized for:

- Append-only scenarios
- Data nie jest modyfikowany po zapisie

Use case:

- Application logs (ciągłe dopisywanie)
- Audit logs
- Streaming data capture
- Event data
- Telemetry data

Page Blobs

Charakterystyka:

- Collection of 512-byte pages
- Zoptymalizowane pod random read/write
- Max size: 8 TB
- Efficient random access

Operations:

- Read/write arbitrary 512-byte pages
- Random access patterns
- Low latency

Optimized for:

- Frequent random read/write operations
- Low-latency access

Use case:

- Virtual Hard Disks (VHD) dla Azure VMs
- Database files
- Random access data
- Azure Premium Storage disks

Ważne!

Page Blobs są głównie używane wewnętrznie przez Azure dla VM disks.
Rzadko używane bezpośrednio przez developerów.

2.4 Access Tiers

Hot Tier - Czesty Dostęp

Charakterystyka:

- Dla aktywnie używanych danych
- Najwyższe storage cost
- Najniższe access cost (read/write)
- Default tier dla nowych blobów
- Lowest latency

Pricing:

- Storage: najdroższe (\$\$/GB/month)
- Access: najtańsze (operacje read/write)

Use case:

- Active data w użyciu
- Dane często odczytywane/modyfikowane
- Staging area dla ETL
- Production databases
- Aplikacje wymagające low latency

Kiedy używać:

- Dostęp do danych > 1 raz dziennie
- Short-term storage

Cool Tier - Rzadki Dostęp

Charakterystyka:

- Dla rzadko używanych danych
- Niższe storage cost niż Hot
- Wyższe access cost niż Hot
- Minimum storage duration: 30 dni
- Early deletion fee (jeśli < 30 dni)

Pricing trade-off:

- Płacisz mniej za storage
- Płacisz więcej za każdy access (read/write/list)

Use case:

- Short-term backup
- Older data rarely accessed
- Disaster recovery datasets
- Media content (stare wideo)
- Data analytics datasets (after processing)

Kiedy używać:

- Dostęp do danych 1 raz na miesiąc
- Storage period > 30 dni

Ważne!

Early deletion charge: jeśli usuniesz blob przed 30 dniami, płacisz jakbyś go trzymał przez pełne 30 dni!

Cold Tier - Bardzo Rzadki Dostęp

Charakterystyka:

- Dla bardzo rzadko używanych danych
- Jeszcze niższe storage cost
- Jeszcze wyższe access cost
- Minimum storage duration: 90 dni
- Early deletion fee (jeśli < 90 dni)

Pricing:

- Storage: tańsze niż Cool
- Access: droższe niż Cool

Use case:

- Long-term backup
- Compliance data (must keep)
- Archival data z occasional access

Kiedy używać:

- Dostęp raz na kwartał lub rzadziej
- Storage period > 90 dni

Archive Tier - Archiwum

Charakterystyka:

- Offline storage (dane niedostępne bezpośrednio)
- Najniższe storage cost
- Najwyższe access cost + rehydration cost
- Minimum storage duration: 180 dni
- Early deletion fee (jeśli < 180 dni)
- Rehydration required przed odczytem

Rehydration:

- Proces przenoszenia blobu z Archive do Hot/Cool
- Standard priority: do 15 godzin
- High priority: < 1 godzina (dla blobów < 10 GB)
- Koszt rehydration + destination tier storage

2 metody rehydration:

1. **Copy to Hot/Cool** - oryginalny zostaje w Archive
2. **Change tier** - blob przechodzi do Hot/Cool

Use case:

- Long-term archival (lata)
- Compliance data (legal requirements)
- Tape replacement
- Backups rarely/never restored
- Historical data

Kiedy używać:

- Dostęp bardzo rzadki (raz na rok lub nigdy)
- Storage period > 180 dni
- Można czekać godziny na dostęp

Ważne!

Archive = OFFLINE storage!
NIE możesz bezpośrednio czytać/modyfikować archived blobs.
Musisz najpierw REHYDRATE (kopiować do Hot/Cool).

2.5 Zmiana Access Tier

Tier Changes

Lifecycle Management:

- Możesz ustawić automatyczne reguły
- Przykład: po 30 dniach → Cool, po 90 dniach → Archive
- Bazuje na last modified time lub creation time
- Filters: prefix, blob type

Manual tier change:

- Możesz zmienić tier ręcznie (Portal, API, CLI)
- Hot Cool Cold: natychmiastowe
- → Archive: natychmiastowe
- Archive →: wymaga rehydration

Account-level default tier:

- Hot lub Cool (nie może być Archive)
- Wszystkie nowe blobs domyślnie w tym tier
- Można override per-blob

3 Queue Storage

3.1 Podstawy

Queue Storage - Message Queue

Definicja: Simple, reliable message queue service dla asynchronous communication

Architektura:

Producer → Queue → Consumer

Zastosowanie:

- Decoupling components
- Asynchronous processing
- Load leveling (buffering bursts)
- Work distribution
- Task scheduling

Benefits:

- Resilience - jeśli consumer down, messages czekają
- Scalability - multiple consumers
- Flexibility - producer i consumer niezależne

3.2 Charakterystyka

Queue Storage - Specyfikacja

Message properties:

- Max message size: 64 KB
- Base64 encoded (faktycznie 48 KB użytecznego payloadu)
- Text format (zwykle JSON)