Zone redundant (ZR) HA

Last updated by | Hamza Aqel | Jan 10, 2023 at 7:32 AM PST

What is ZR HA?

• Ensure data is always available with zone redundant HA

• Synchronous replication across availability zones for high resiliency

• Choose the availability zone for your database for improved connectivity

Why ZR HA?

• Non-HA Flexible server provides some form of HA – similar to Single server

• Data is stored in 3 copies • When the node crashes, it is restarted again.

• But this does not protect from AZ-level failures.

• Most mission-critical applications that require high uptime and protection from various failures, including AZ faults

• Customers choose HA to have high uptime during planned and unplanned downtimes

• There will be some performance impact and customers are generally OK with that

• Though they may have issues with % of impact

Features/Benefits of ZR HA:

• Provides high availability and automatic failovers

• Deploys primary and standby servers across availability zones

• Replicates data in synchronous mode

• Enables high uptime during planned and unplanned downtime events

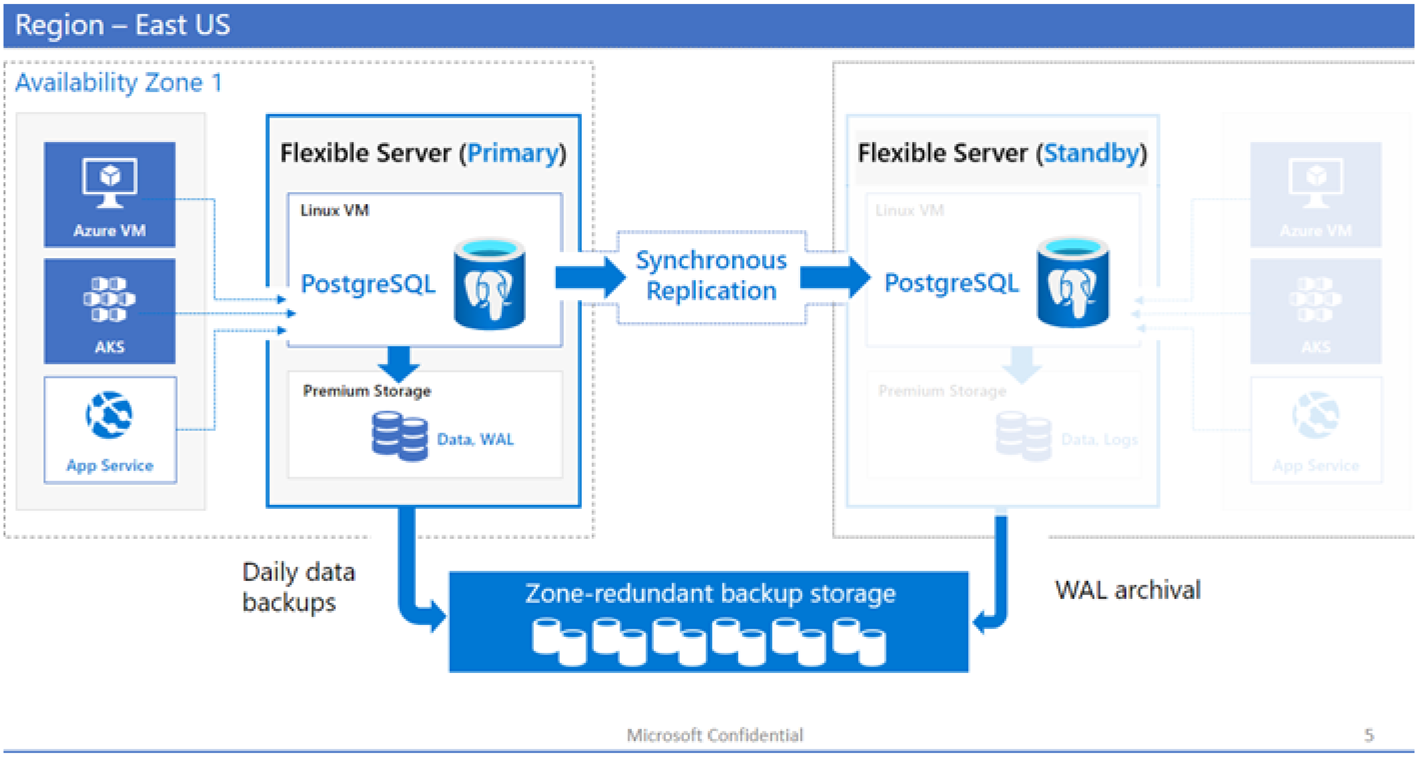
```
Planned events: Scale compute, upgrades, etc.

Unplanned events: Node failure, AZ failure, etc.
```

• Failover within 1-2 minutes

• No SLAs offered during preview

• Schedule planned events including Azure maintenance with Managed maintenance window to further reduce downtime impact

Customer view:



Region – East US

Availability Zone 1

Azure VM

AKS

App Service

Flexible Server (Primary)

Linux VM

PostgreSQL

Premium Storage

Data, WAL

Synchronous Replication

Flexible Server (Standby)

Linux VM

PostgreSQL

Premium Storage

Data, Logs

Azure VM

AKS

App Service

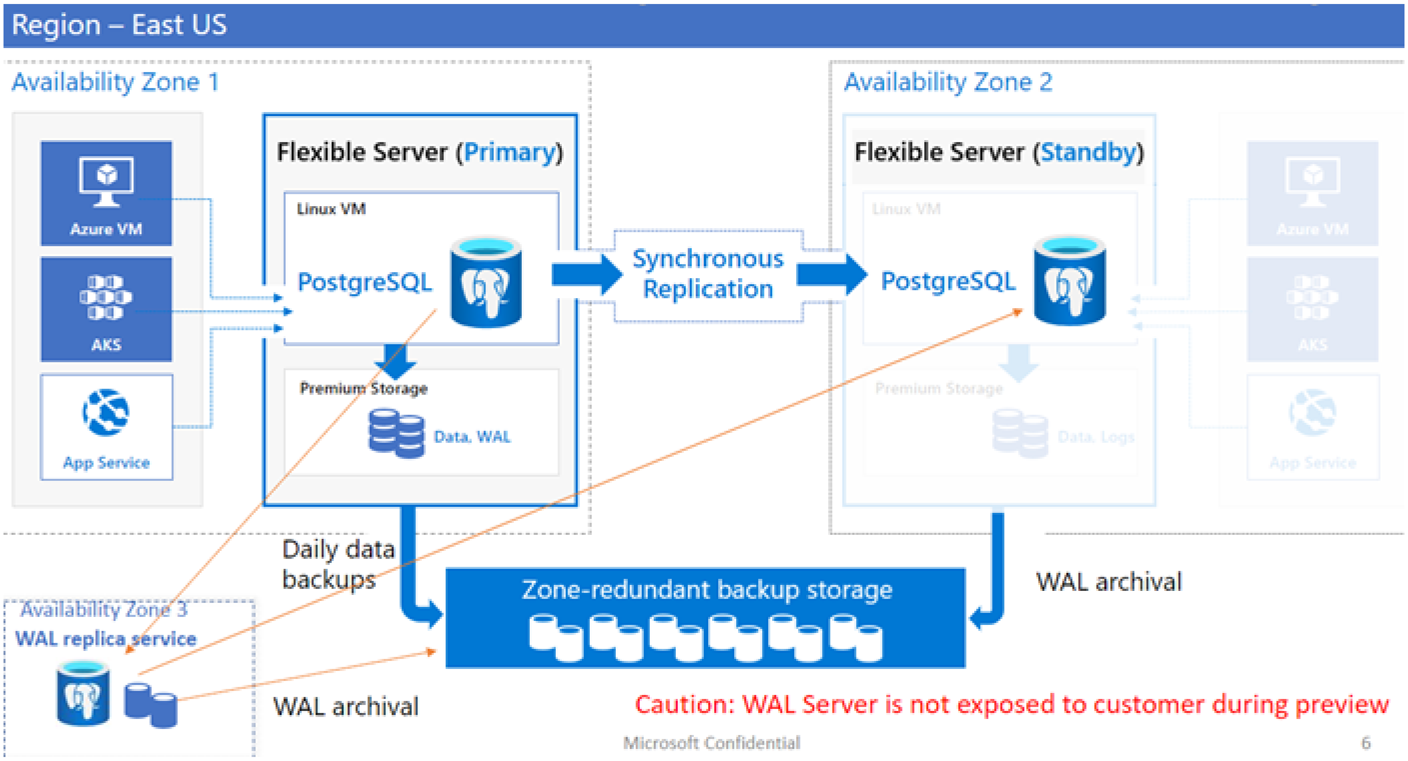
Daily data backups

Zone-redundant backup storage

WAL archival

Microsoft Confidential 5

Behind the scene view: ZR HA **MSFT confidential



Region – East US

Availability Zone 1

Azure VM

AKS

App Service

Flexible Server (Primary)

Linux VM

PostgreSQL

Premium Storage

Data, WAL

Synchronous Replication

Availability Zone 2

Flexible Server (Standby)

Linux VM

PostgreSQL

Premium Storage

Data, Logs

Azure VM

AKS

App Service

Daily data backups

Zone-redundant backup storage

WAL archival

Availability Zone 3

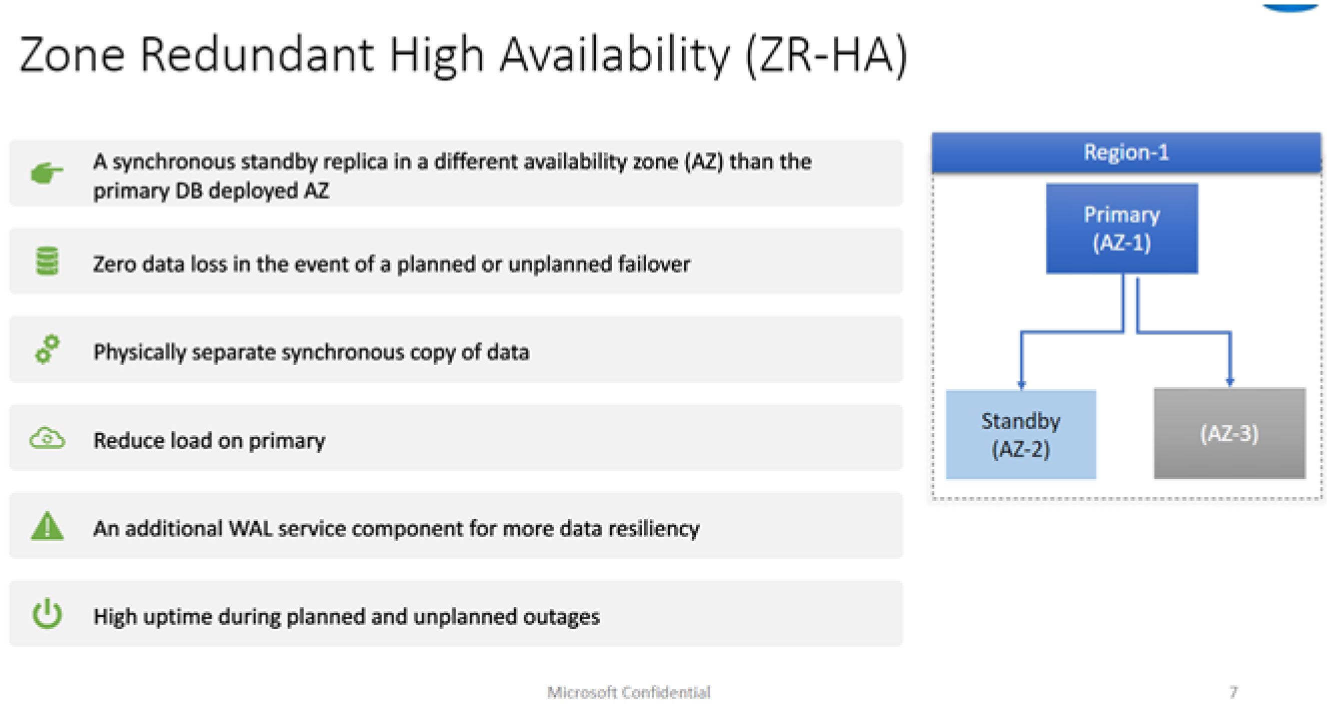
WAL replica service

WAL archival

Caution: WAL Server is not exposed to customer during preview

Microsoft Confidential 6

**MSFT confidential

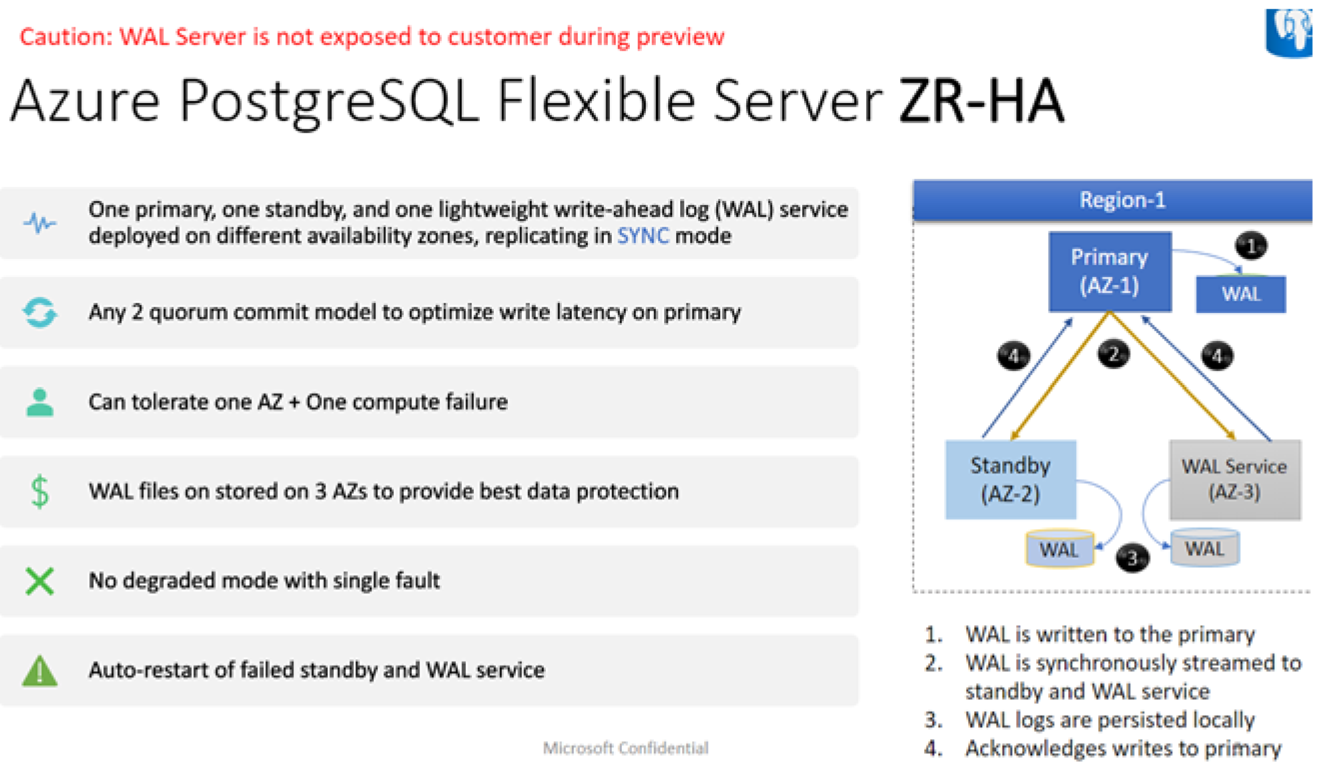


Zone Redundant High Availability (ZR-HA)

- A synchronous standby replica in a different availability zone (AZ) than the primary DB deployed AZ
- Zero data loss in the event of a planned or unplanned failover
- Physically separate synchronous copy of data
- Reduce load on primary
- An additional WAL service component for more data resiliency
- High uptime during planned and unplanned outages

Microsoft Confidential 7

**MSFT confidential



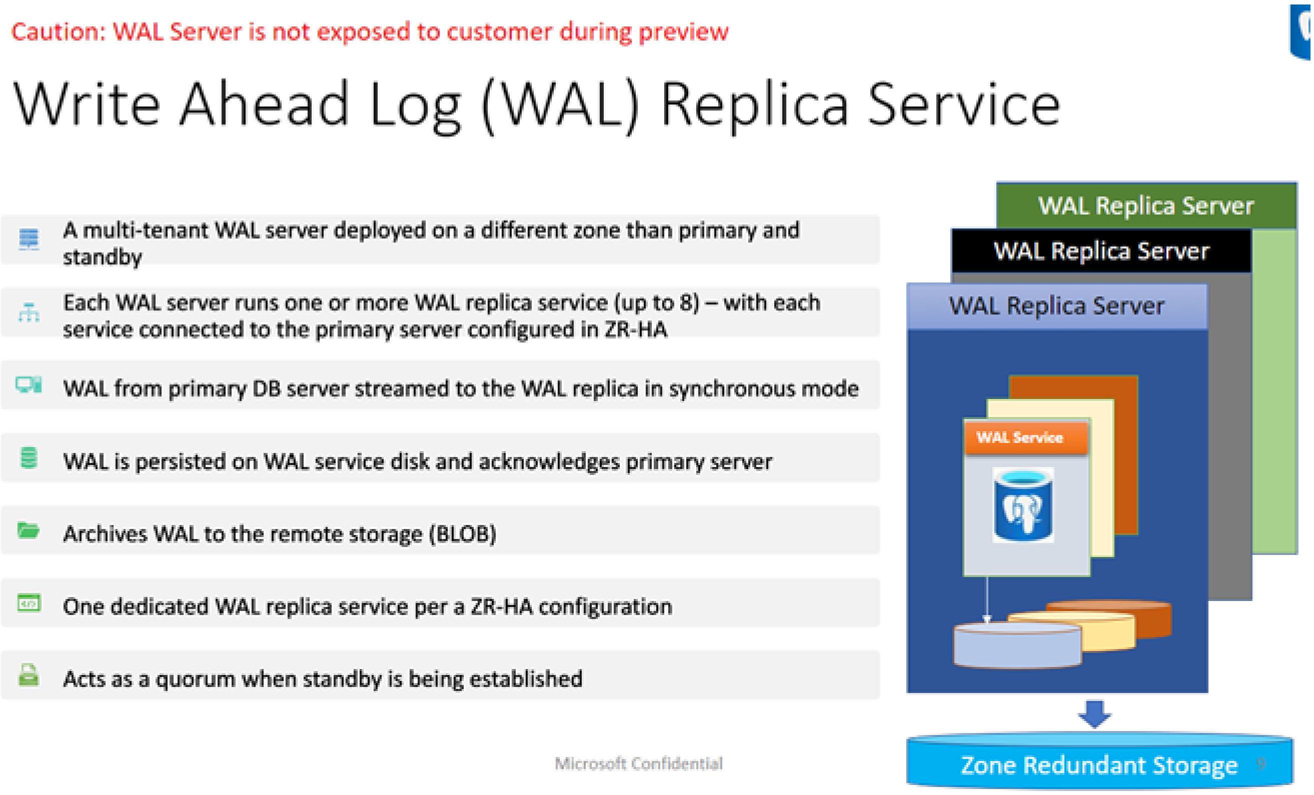
Caution: WAL Server is not exposed to customer during preview

Azure PostgreSQL Flexible Server **ZR-HA**

- One primary, one standby, and one lightweight write-ahead log (WAL) service deployed on different availability zones, replicating in SYNC mode
- Any 2 quorum commit model to optimize write latency on primary
- Can tolerate one AZ + One compute failure
- WAL files on stored on 3 AZs to provide best data protection
- No degraded mode with single fault
- Auto-restart of failed standby and WAL service

1. WAL is written to the primary
2. WAL is synchronously streamed to standby and WAL service
3. WAL logs are persisted locally
4. Acknowledges writes to primary

Microsoft Confidential



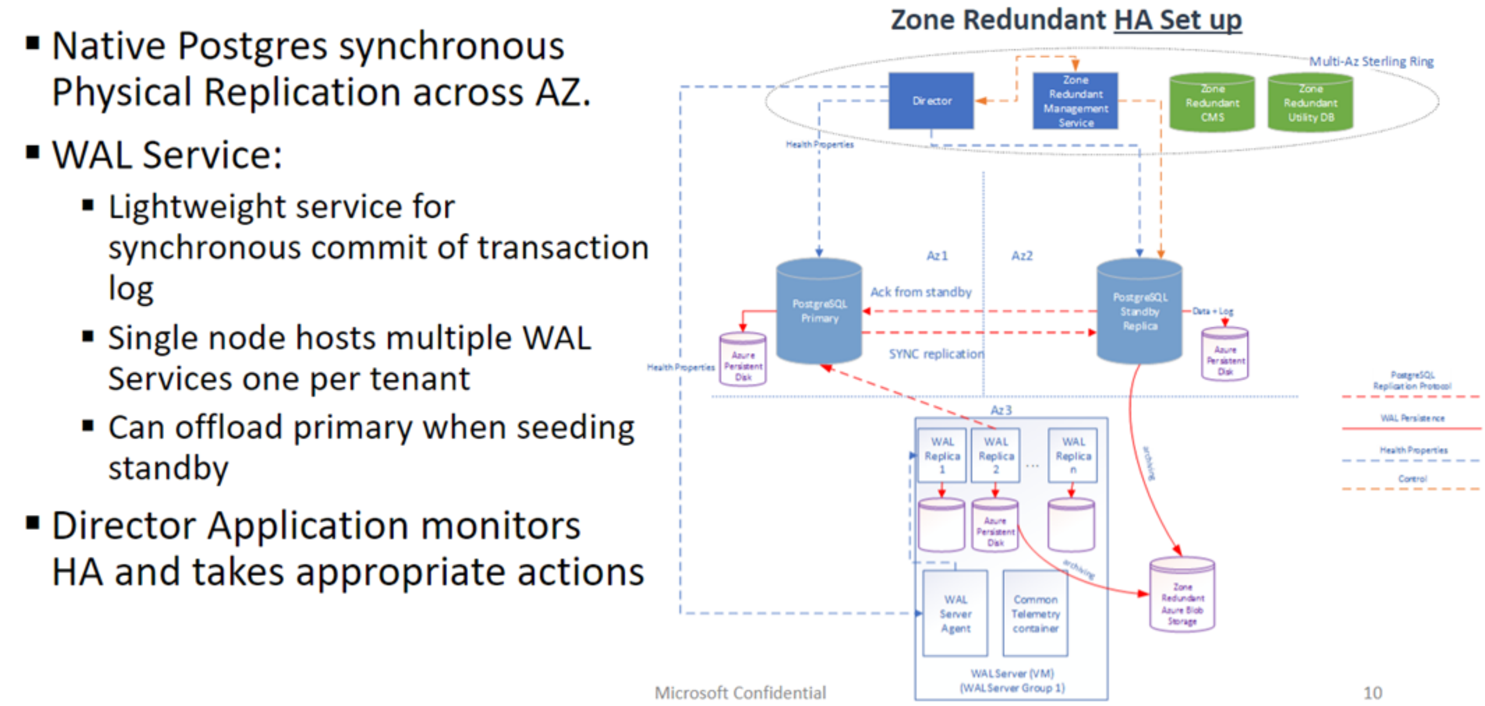
Caution: WAL Server is not exposed to customer during preview

Write Ahead Log (WAL) Replica Service

- A multi-tenant WAL server deployed on a different zone than primary and standby
- Each WAL server runs one or more WAL replica service (up to 8) – with each service connected to the primary server configured in ZR-HA
- WAL from primary DB server streamed to the WAL replica in synchronous mode
- WAL is persisted on WAL service disk and acknowledges primary server
- Archives WAL to the remote storage (BLOB)
- One dedicated WAL replica service per a ZR-HA configuration
- Acts as a quorum when standby is being established

Microsoft Confidential

Service Architecture – Data Plane of PostgreSQL **MSFT confidential

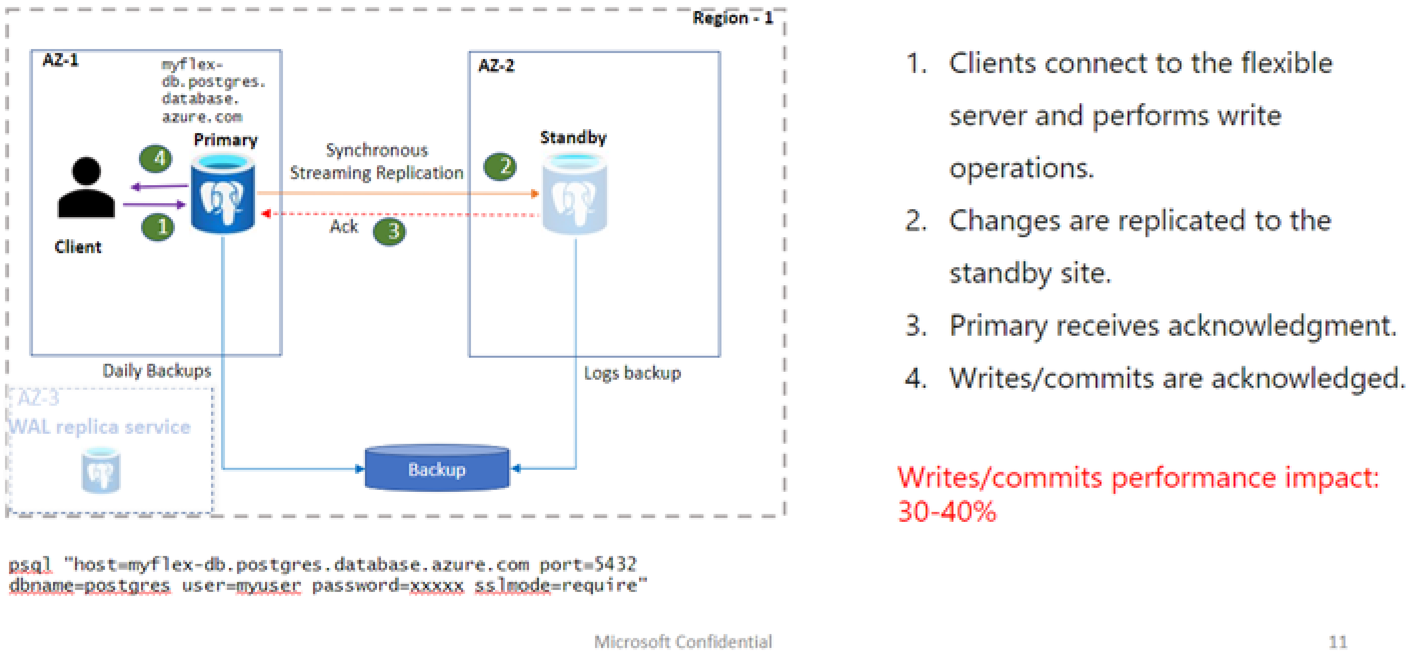


- Native Postgres synchronous Physical Replication across AZ.
- WAL Service:
  - Lightweight service for synchronous commit of transaction log
  - Single node hosts multiple WAL Services one per tenant
  - Can offload primary when seeding standby
- Director Application monitors HA and takes appropriate actions

Zone Redundant HA Set up

Microsoft Confidential 10

Steady state HA**MSFT confidential



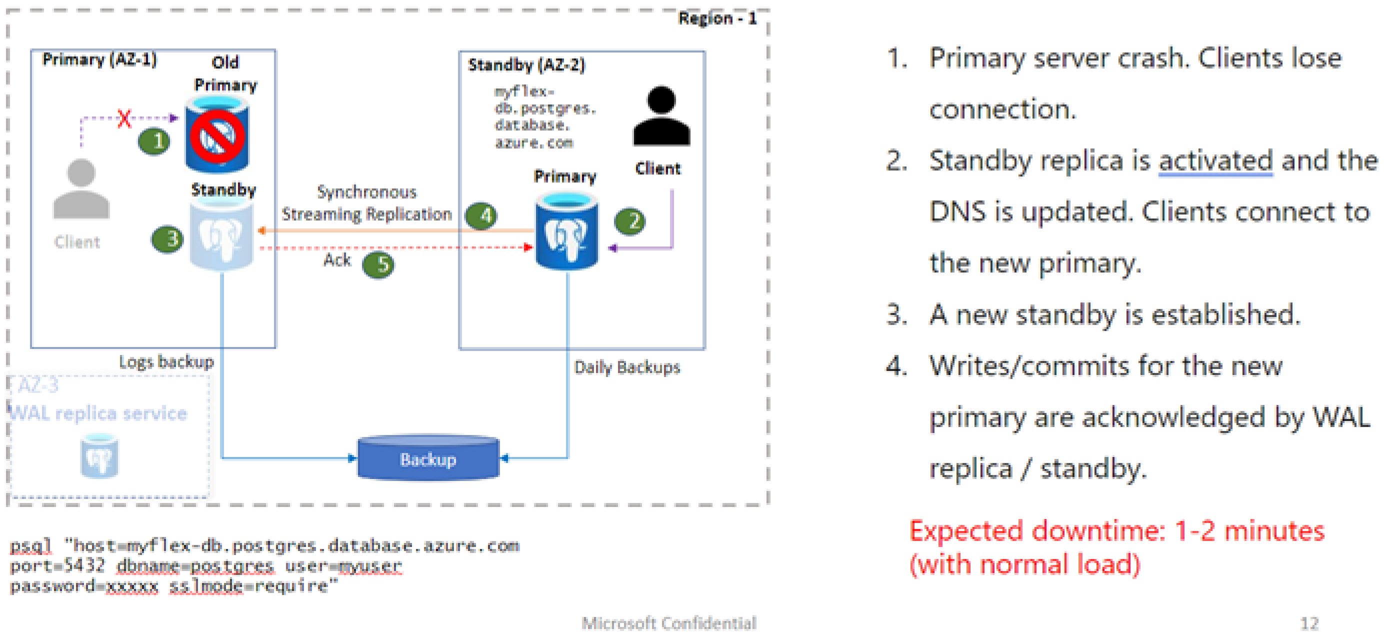
```
psql "host=myflex-db.postgres.database.azure.com port=5432
dbname=postgres user=myuser password=xxxxx sslmode=require"
```

1. Clients connect to the flexible server and performs write operations.
2. Changes are replicated to the standby site.
3. Primary receives acknowledgment.
4. Writes/commits are acknowledged.

Writes/commits performance impact: 30-40%

Microsoft Confidential 11

Failover state - Unplanned**MSFT confidential



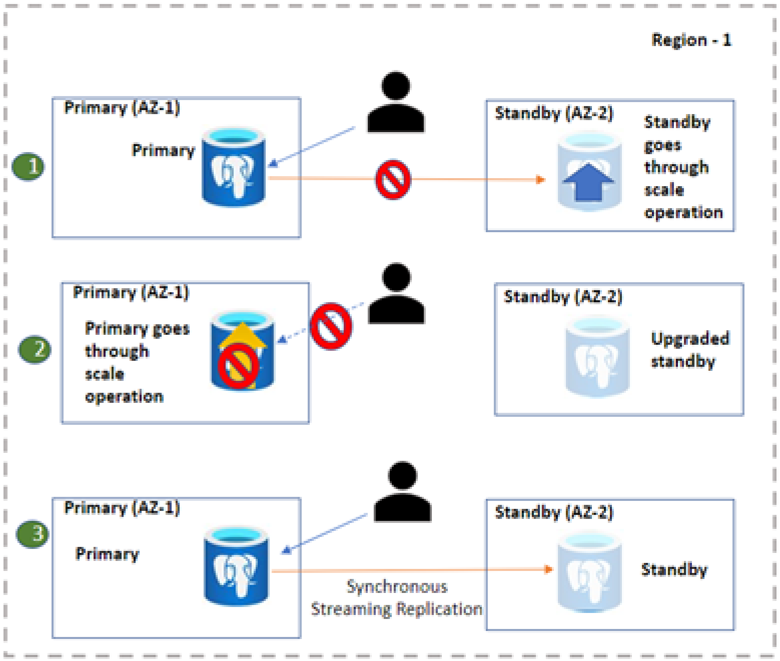
```
psql "host=myflex-db.postgres.database.azure.com
port=5432 dbname=postgres user=myuser
password=xxxxx sslmode=require"
```

1. Primary server crash. Clients lose connection.
2. Standby replica is activated and the DNS is updated. Clients connect to the new primary.
3. A new standby is established.
4. Writes/commits for the new primary are acknowledged by WAL replica / standby.

Expected downtime: 1-2 minutes (with normal load)

Microsoft Confidential 12

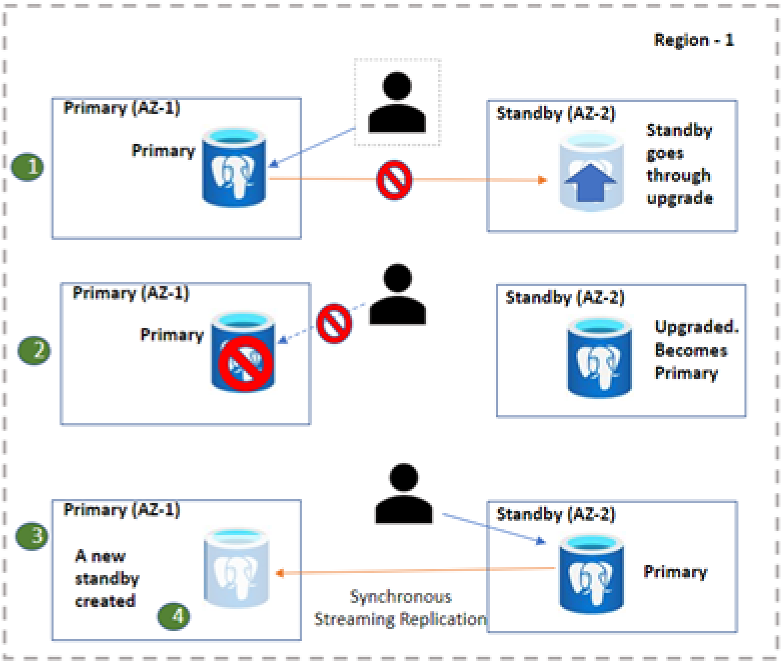
Scale Compute Operation **MSFT confidential



1. Standby first goes through the scale process. Replication is terminated. Primary is always connected to the WAL service to keep the quorum.
2. Once complete, the primary server goes through the scale operation. Client loses connectivity.
3. Once the scale operation is complete, clients can resume their operation. [2-5 minutes of downtime]

Microsoft Confidential 13

Maintenance/Minor version upgrade **MSFT confidential

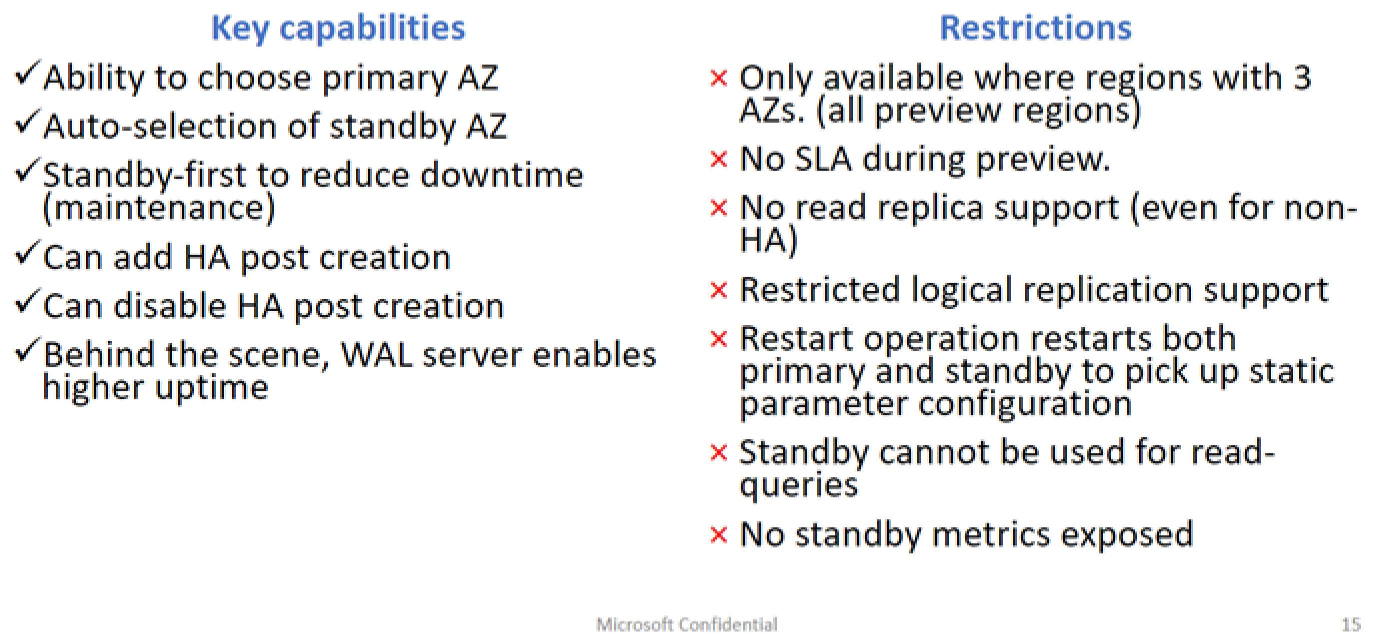


Happens during maintenance window

1. Standby first goes through the upgrade process. Replication is terminated. Primary connects to WAL service to keep quorum.
2. Once standby is upgraded, failover is initiated while the DNS is updated.
3. Clients connect to the new primary.
4. A new standby is established with the updated version

Microsoft Confidential 14

Features and Limitations in Public Preview **MSFT confidential



Key capabilities

- ✓ Ability to choose primary AZ
- ✓ Auto-selection of standby AZ
- ✓ Standby-first to reduce downtime (maintenance)
- ✓ Can add HA post creation
- ✓ Can disable HA post creation
- ✓ Behind the scene, WAL server enables higher uptime

Restrictions

- × Only available where regions with 3 AZs. (all preview regions)
- × No SLA during preview.
- × No read replica support (even for non-HA)
- × Restricted logical replication support
- × Restart operation restarts both primary and standby to pick up static parameter configuration
- × Standby cannot be used for read-queries
- × No standby metrics exposed

Microsoft Confidential 15

Potential questions / cases:

Performance

• Due to synchronous replication, customers can experience 30-40% performance hit for writes and commits

--Remote transmission from the memory + write + acknowledgement

Why scale compute/storage operation is taking longer?

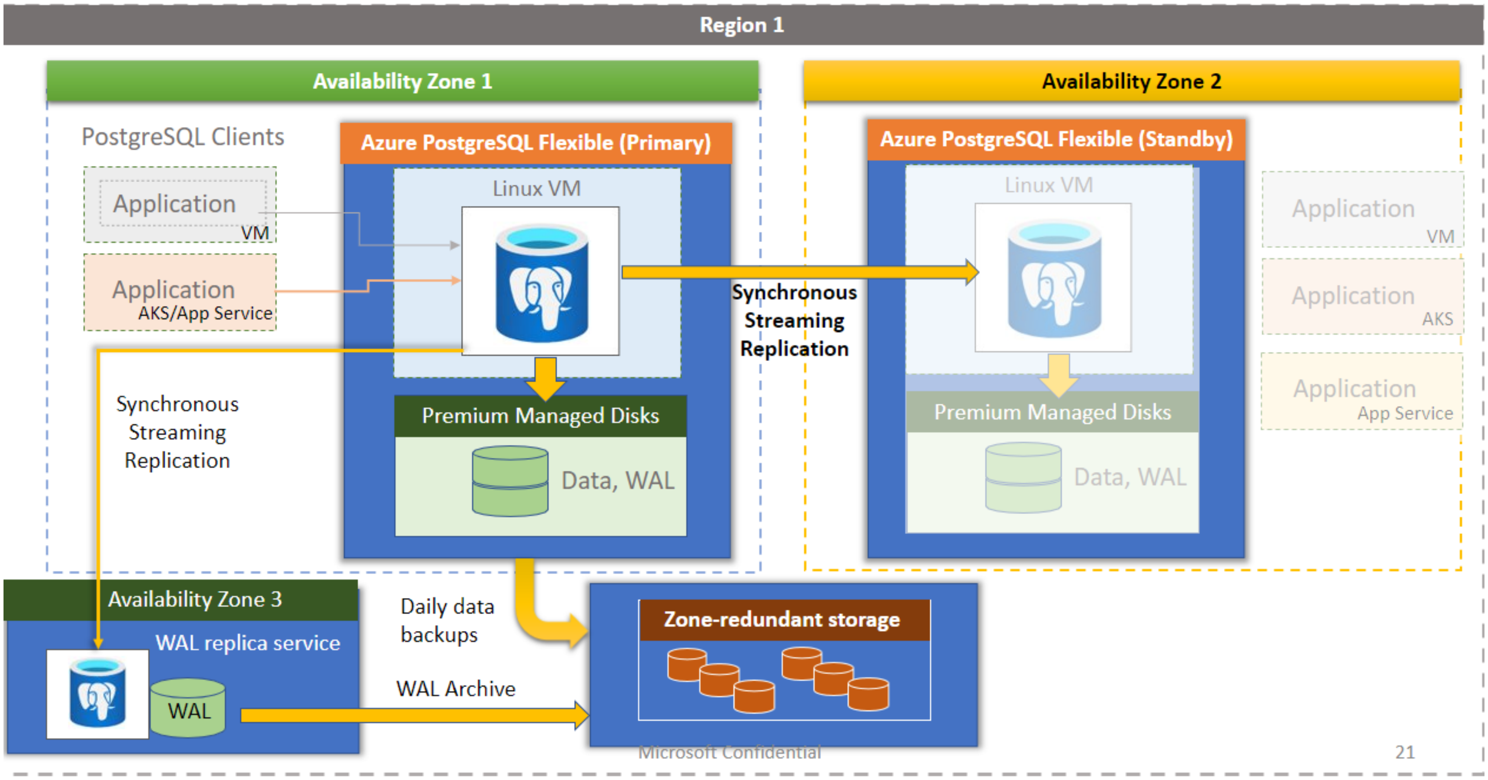
• Currently, we scale the standby first and then primary. No failover. Hence, users may experience longer downtime during primary reconfiguration. Plans to change it for GA.

How to test failover?

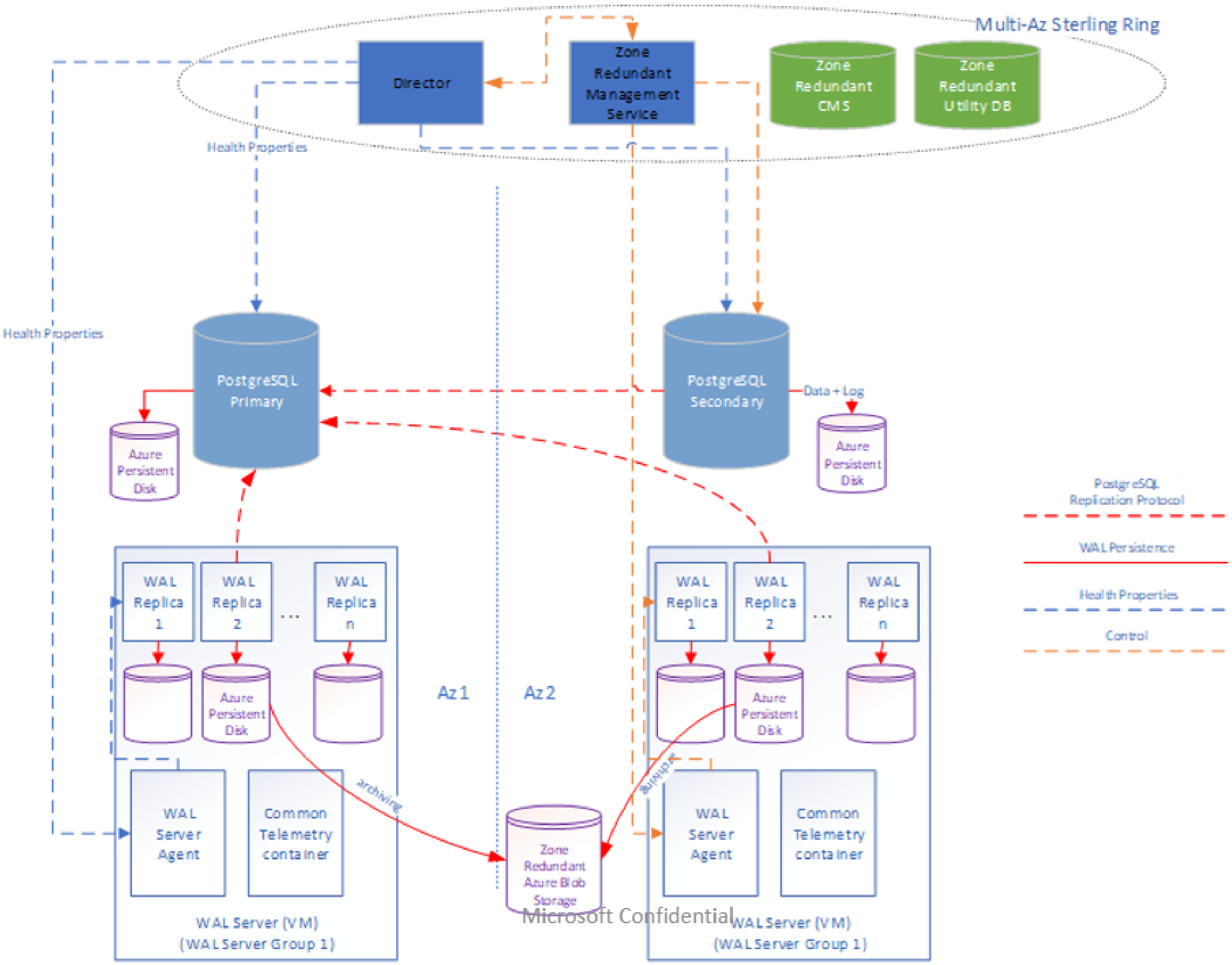
• On-demand failover is planned for GA. [Interim API shown below]

az rest -m post --header "Accept=application/json" -u "https://management.azure.com/subscriptions/<subscription>/resourceGroups/<resource group>/providers/Microsoft.DBforPostgreSQL/flexibleServers/<server name>/restart?api-version=2020-02-14-privatepreview" --body "{"RestartWithFailover": true}"

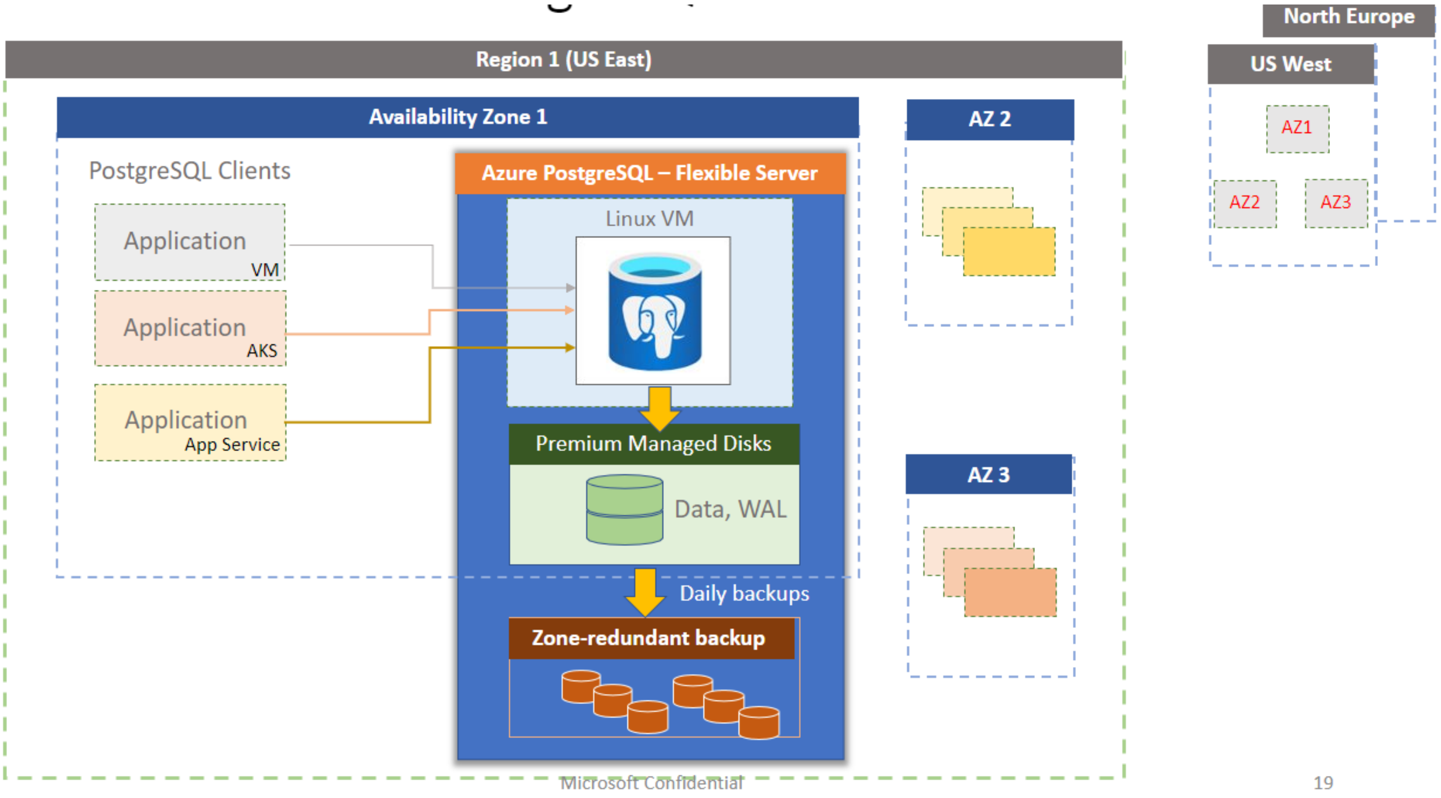
Azure PostgreSQL Flexible Server: ZR-HA Architecture **MSFT confidential



Postgres WAL Replica Service**MSFT confidential



Azure services(IAAS,Webapp,AKS) and Azure Database for PostgreSQL Flexible Server**MSFT confidential



Things to know:

• Our zone redundant HA is the same as Multi-AZ terminology that many customers use. So, if they refer multi-AZ in their case, it is our ZR HA.

• The 30% performance impact is only during heavy write workload.

Public document: https://docs.microsoft.com/en-us/azure/postgresql/flexible-server/overview#high-availability

How good have you found this content?