Data in Azure

Memi Lavi www.memilavi.com

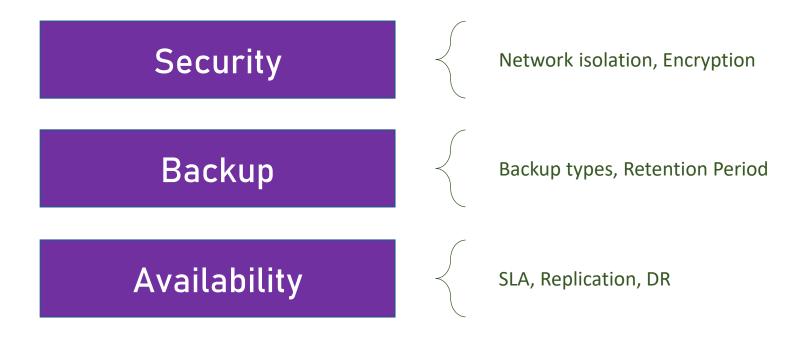


Data in Azure

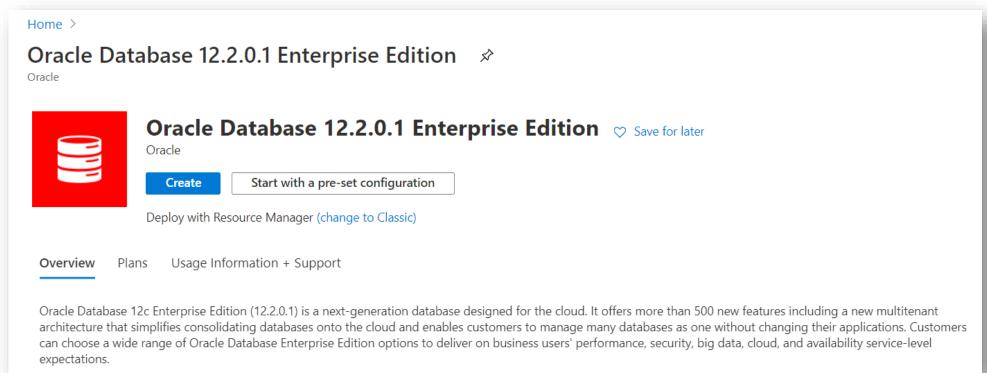
- Azure provides many data solutions as cloud services
- Relational databases, NoSQL databases, object stores
- Fully managed services
- Can be part of Azure app or completely independent
- Various pricing models
- Always better than unmanaged solutions

Major Database Features

What to look for when selecting a database?



- Azure VM can be setup with database software
- There are ready-made VMs in the marketplace



- Pros of Database on VM:
 - Full flexibility
 - Full control

- Cons of Database on VM:
 - You have to take care of everything:
 - SLA
 - Updates
 - Availability
 - Security
 - Backups
 - And more...

In this section we'll discuss only managed databases

Azure SQL

- Managed SQL Server on Azure
- Works like any other SQL Server using the same tools
- Great compatibility with on-prem SQL Server
 - Depends on the exact Azure SQL Flavor
- Offers built-in security, backups, availability and more
- Flexible pricing models

Azure SQL Flavors

Azure SQL Database

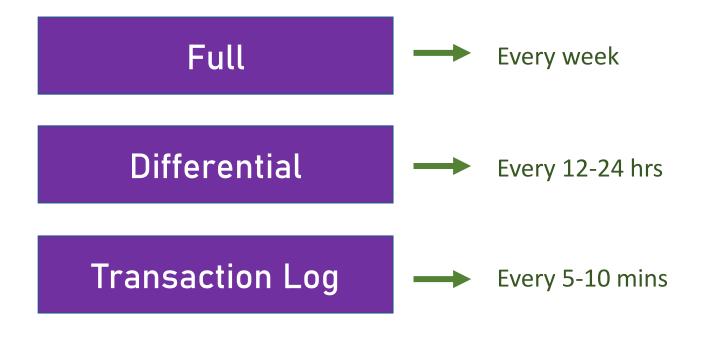
Elastic Pool

Managed Instance

- Managed SQL Server on Azure
- Single database on a single server
- Automatic backups, updates, scaling
- Good compatibility with on-prem SQL Server
 - Not all features are supported

- Security:
 - IP firewall rules
 - Service Endpoints
 - SQL & Azure AD Authentication
 - Secure communication (TLS)
 - Data encrypted by default (TDE)

Backup:



- Retention Period:
 - Regular backup: 7-35 days (default is 7)
 - Long term backup: up to 10 years

- Availability:
 - Backup is stored in a geo-redundant storage
 - Active geo-replication
 - SLA: 99.9% 99.995%, depends on tier and redundancy

Compute Tiers:

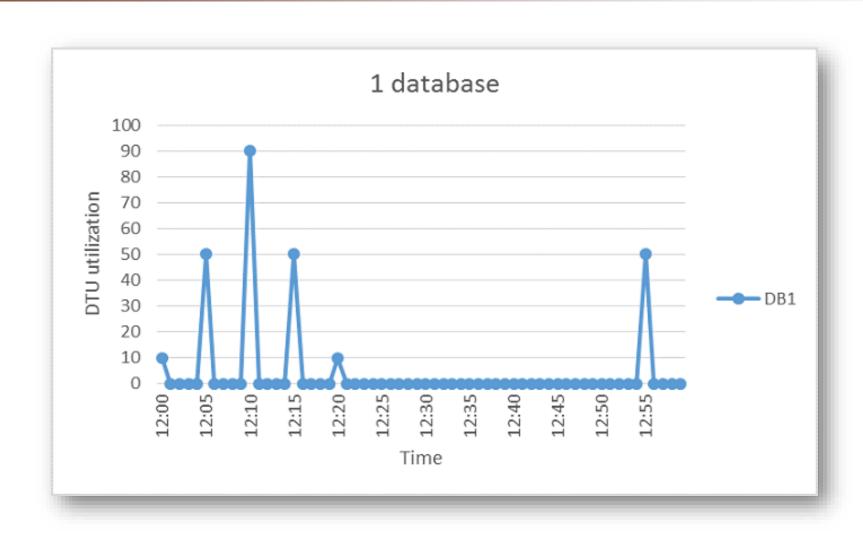
Provisioned

- Pay for allocated resources regardless of actual use
- Can be reserved

Serverless

- Pay for actual use vCore + RAM / second
- Automatically paused when inactive (pay just for storage)
- Slight delay when warming up
- Can't be reserved

- Based on Azure SQL
- Allows storing multiple databases on single server
- Great for databases with low average utilization and infrequent spikes





- Cost effective
- Purchase the compute resources you need, not the database

Managed Instance

- Closer to the on-prem SQL Server
- Near 100% compatible with on-prem SQL
- Can be deployed to VNet
- Business model close to on-prem one

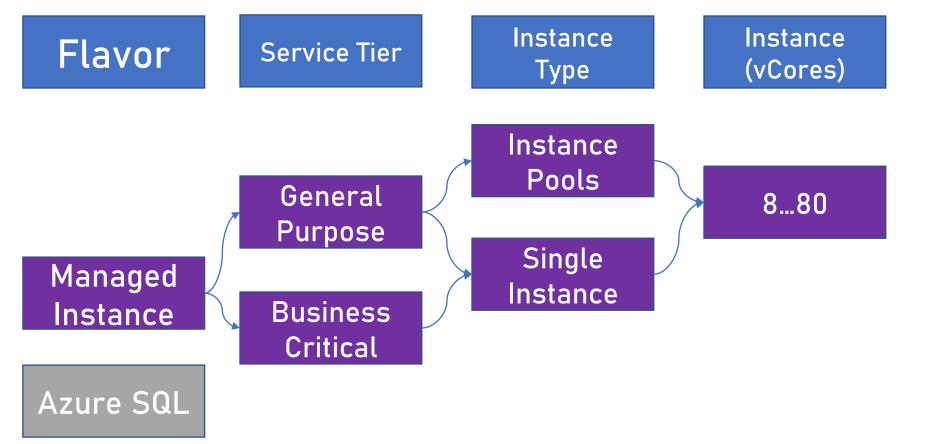
Managed Instance

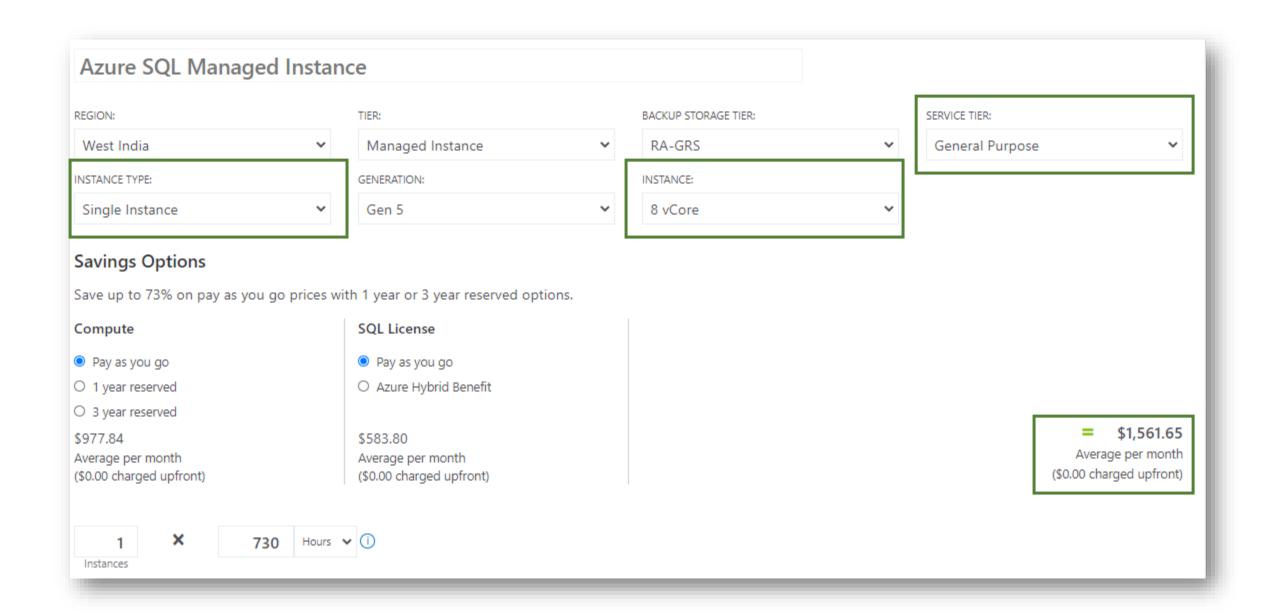
- Main differences:
 - No active geo-replication
 - SLA: 99.99%
 - Supports built-in functions
 - Runs CLR code
 - No auto scaling & tuning

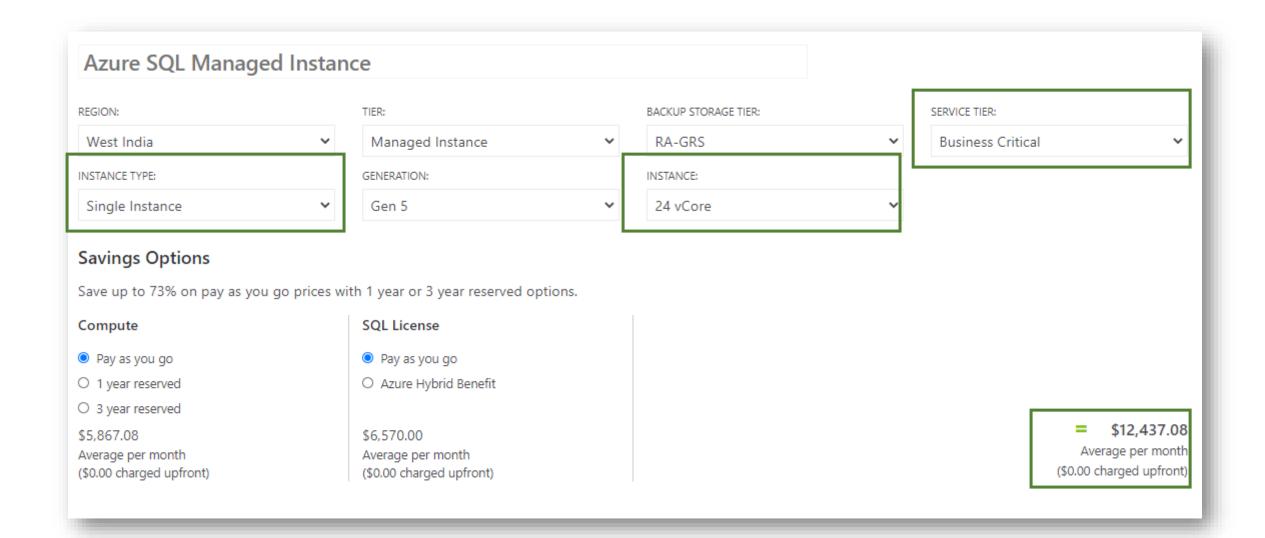
- No availability zone
- No serverless tier
- No Hyperscale

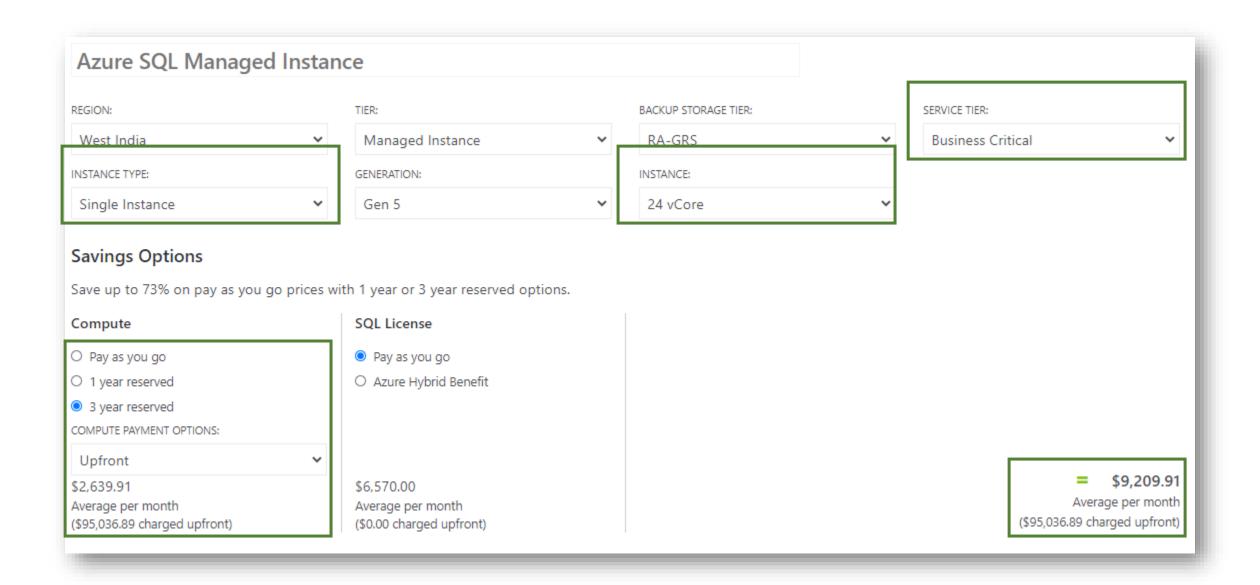
Azure SQL Pricing

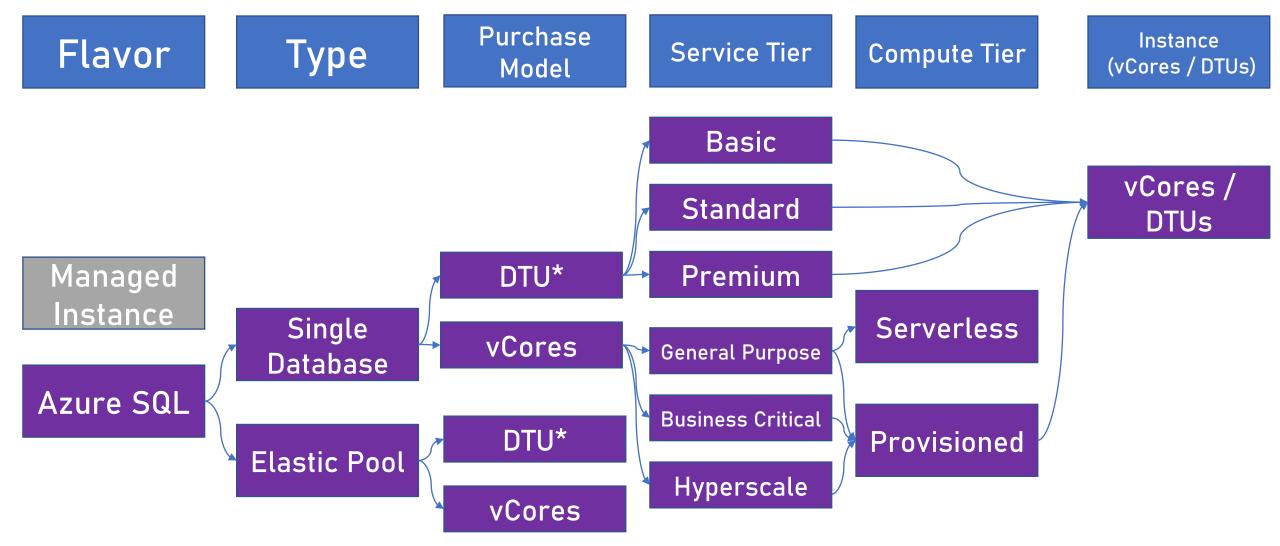










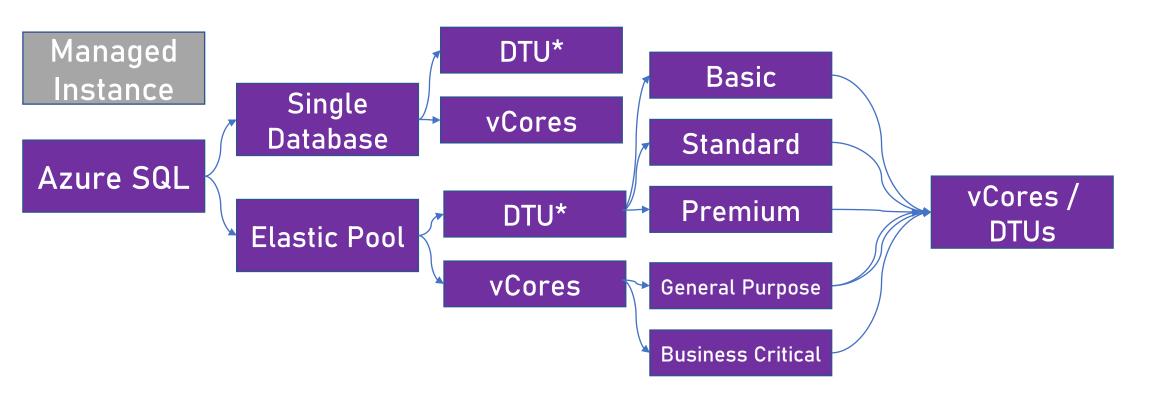


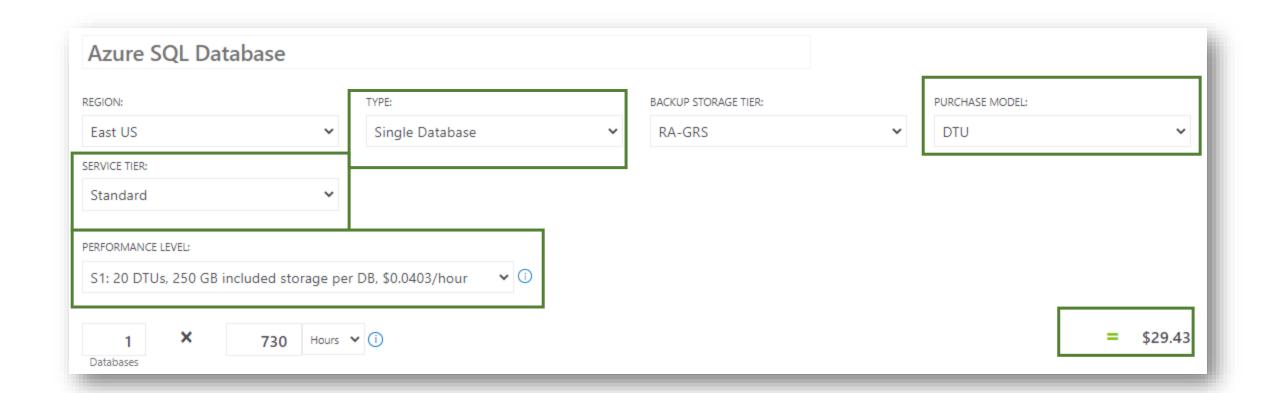
Type

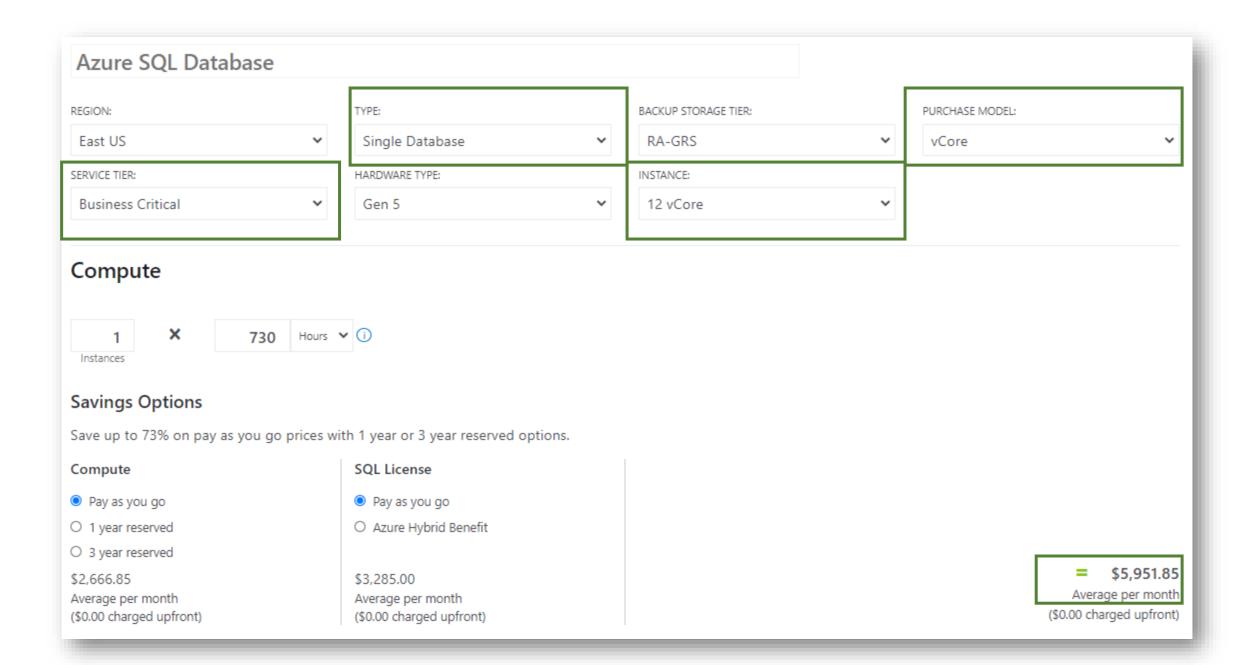
Purchase Model

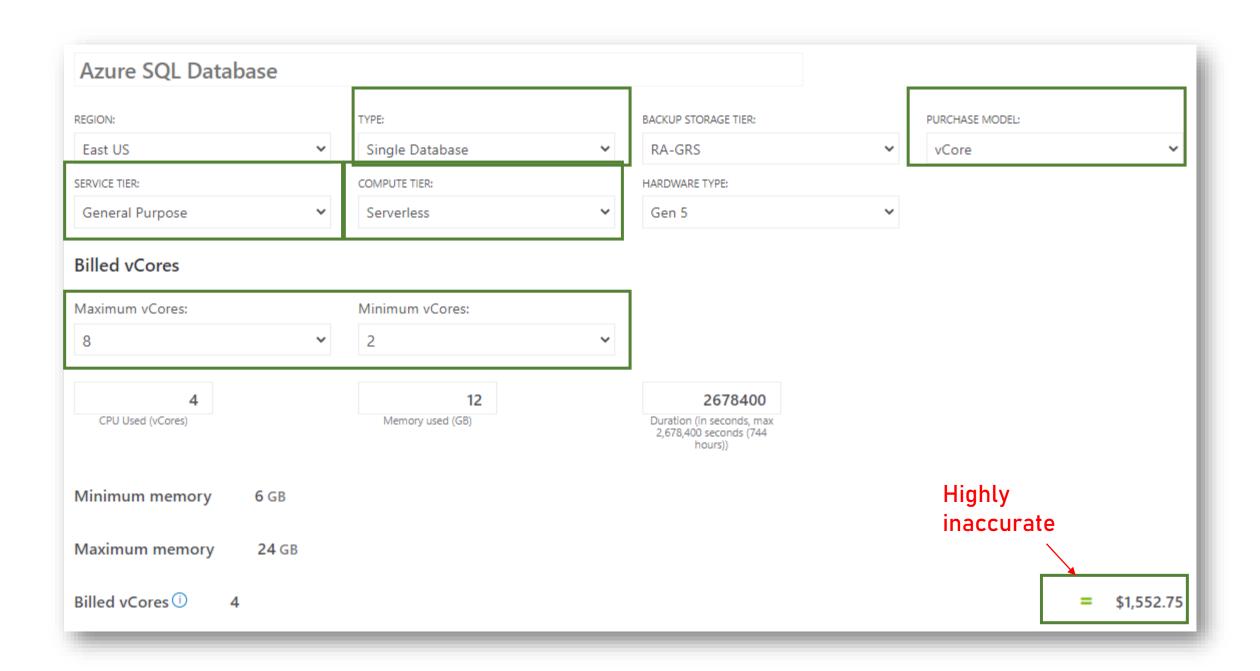
Service Tier

Instance (vCores / DTUs)









Which Azure SQL to Choose?

Are you migrating an on-prem SQL?

Managed Instance

Do you need multiple, mostly low-utilization DBs?

→ Elastic Pool

All other cases

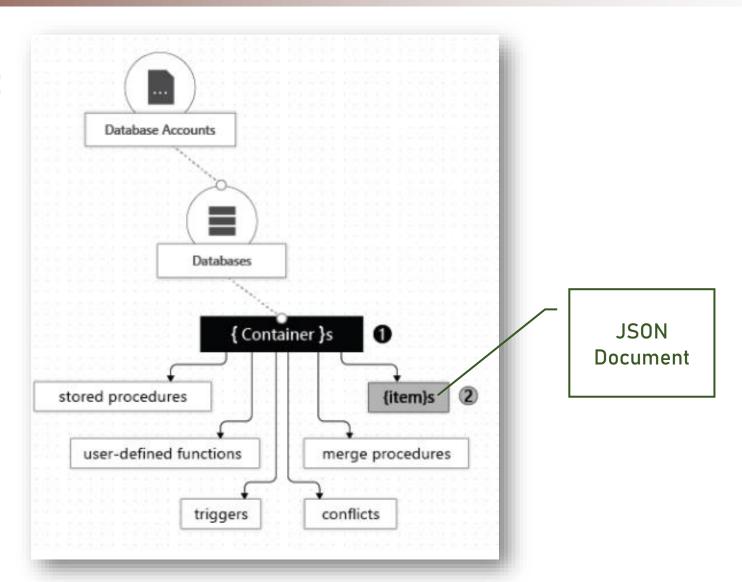
→ Azure SQL

Cosmos DB

- Fully managed NoSQL database
- Amazing performance <10ms for 99% of operations
- Globally distributed
- Fully automatic management updates, scaling, fixes etc.
- Multiple APIS:
 - SQL, Mongo, Gremlin, Azure Table, Cassandra (per account)

Cosmos DB

Hierarchical:



Cosmos DB Availability

- Can be distributed across many regions (configurable)
- API automatically picks the closest one
- When using write replication SLA is 99.999% (!)
- Managed automatically, no code changes required

Cosmos DB Backup

- Full backup every 1-24 hours (default is 4)
- Retention period 20-30 days (default is 30)

Cosmos DB Security

- IP firewall rules
- Service Endpoints
- Private Endpoints
- Azure AD Authentication
- Secure communication (TLS)
- Data encrypted by default

Cosmos DB Partitions

- Data items are divided to partitions
- Logical group of items based on a specific property
- Example: In a cars database, the Model can be a partition property



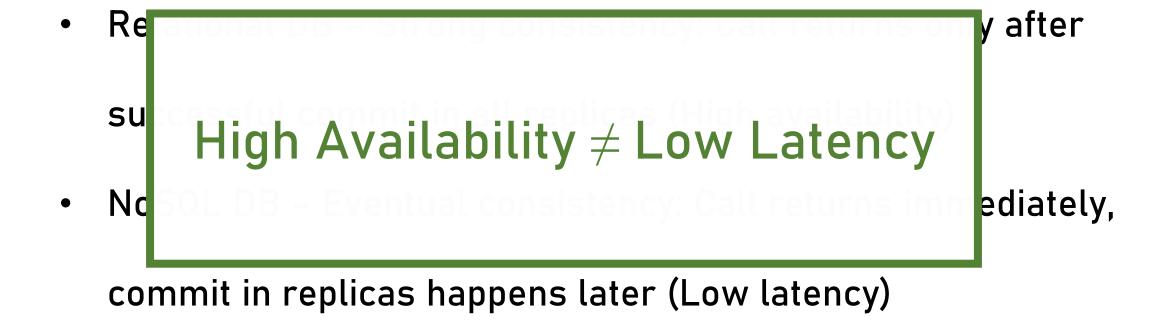
Cosmos DB Partitions

- Partitions are the basic scale unit in Cosmos DB
- Distribution and scale are per partitions
- Make sure items are divided as evenly as possible
- It's extremely important to select the right partition property
- Cannot be modified

Recording of L7S5 from the

Architecture course

Traditionally:



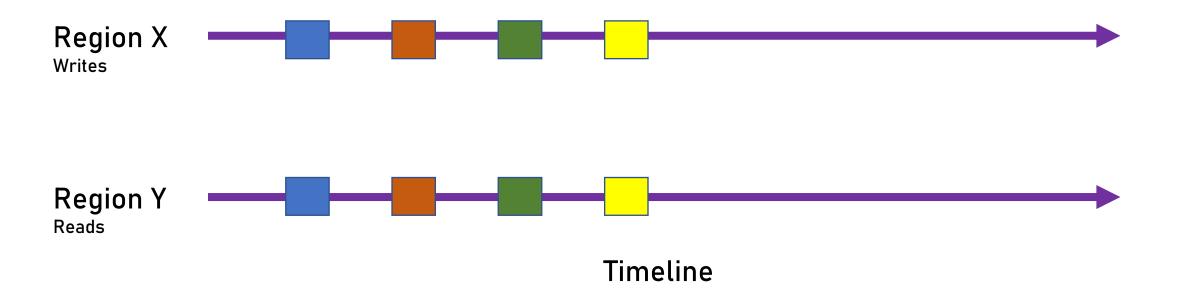
- Cosmos DB offer five consistency levels:
 - Strong (<= As in regular relational DB)
 - Bounded Staleness
 - Session
 - Consistent Prefix
 - Eventual (<= As in regular NoSQL DB)

- The basic question with consistency is:
 - If region X updates an item, and region Y reads this item, which

version will it get?

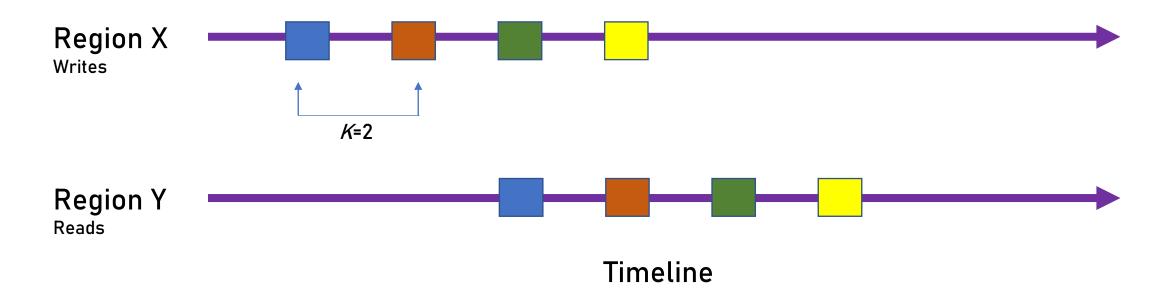
Strong

- Region Y will get the last version of the item updated in region X
- Used for mission critical data



Bounded Staleness

- Region Y will lag behind region X by K versions or Ttime
- Keeps the order of the versions
- Used for low write latency and when order is important

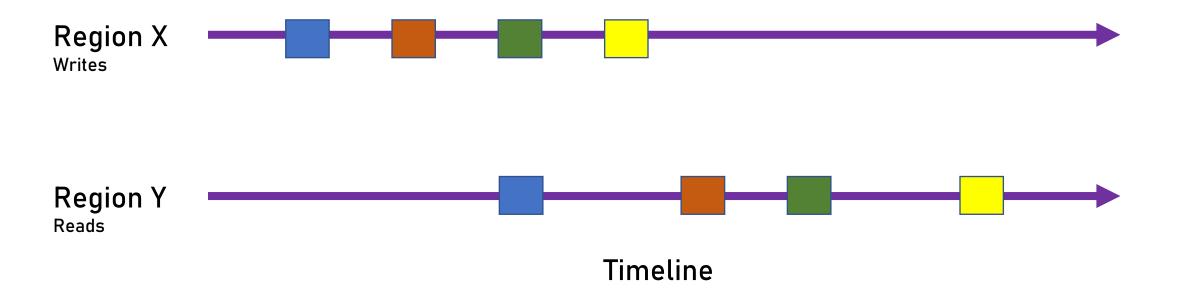


Session

- In a client session Strong consistency
- Other clients Consistent Prefix (sometimes Eventual)

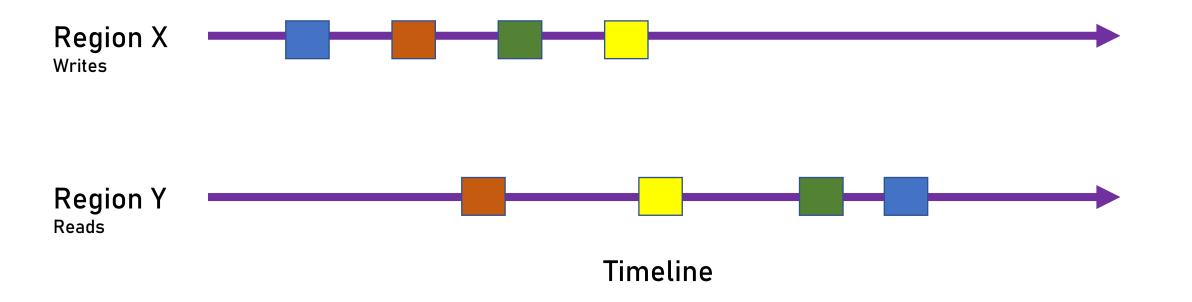
Consistent Prefix

- Keeps the order of the versions
- No guarantee of the lag size (as opposed to Bounded Context)
- Used for low write latency and when reads are infrequent



Eventual

- No order guarantee
- No guarantee of the lag size (as opposed to Bounded Context)
- Used for count of Retweets, Likes, etc.



- Configured at the account level
- Can be relaxed on the request level

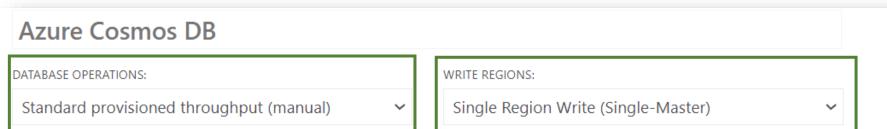
- Based on RU/s
 - Request Unit per Second
- 1 RU = Read item of size 1KB
 - Read = Get the item by its ID, not by query

results

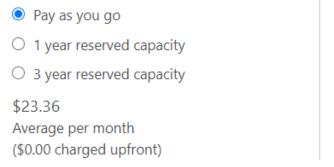
- 1 RU/s = Read 1 item of 1KB in 1 sec
- 400 RU/s = Read 400 items of 1KB in 1 sec
- Update, delete, insert, query more than 1 RU
- You can see the actual RU consumed in the response header of the

- Pricing based on:
 - Operations type: Provisioned, Auto Scale, Serverless
 - Write Regions
 - No. of provisioned RU/s

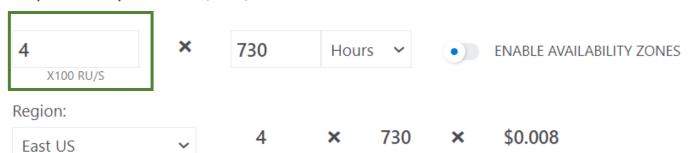
- Database Operations:
 - Provisioned Predefined number of RU/s, can be changed manually later. Offers reserved capacity up to 65% discount
 - Auto Scale Set the maximum RU/s, Cosmos scales up to this number. Good for unpredictable loads
 - Serverless Pay for what you use. In Preview, no SLA



Save up to 65% on pay as you go prices with 1-year or 3-year reserved capacity for Azure Cosmos DB.



Request units per second (RU/s)

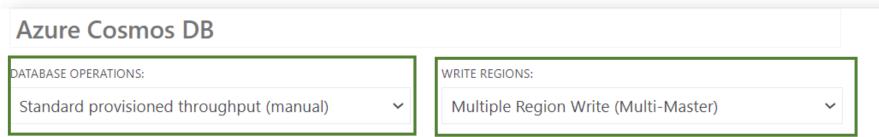


Hours

Per hour

x100 RU/s

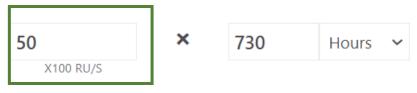
\$23.36



Save up to 65% on pay as you go prices with 1-year or 3-year reserved capacity for Azure Cosmos DB.



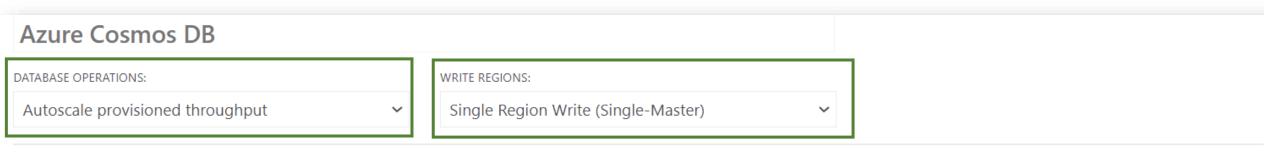
Request units per second (RU/s)



Region:

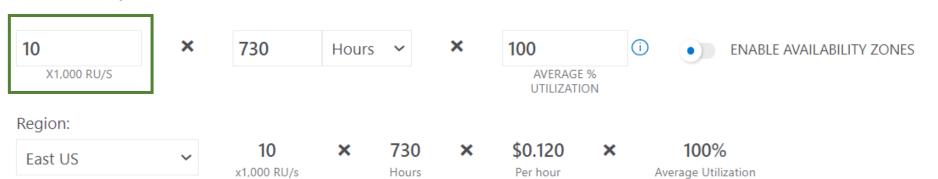


\$584.00



Pay as you go\$876.00Average per month(\$0.00 charged upfront)

Maximum Request Units Per Second (RU/s)



\$876.00

- Managed MySQL on Azure
- Works like any other MySQL database using the same tools
- Great compatibility with on-prem MySQL database
- Offers built-in security, backups, availability and more

- Security:
 - IP firewall rules
 - Service Endpoints
 - Private Endpoints
 - Regular & Azure AD Authentication
 - Secure communication (TLS)
 - Data encrypted by default

- Backup:
 - Depends on Service Tier:
 - Basic Full backup: daily
 - General Purpose up to 4GB:
 - Full backup: once a week
 - Differential backup: twice a day
 - Transaction log backup: every 5 minutes

- Backup:
 - Depends on Service Tier:
 - General Purpose up to 16GB:
 - Full backup: Once created
 - Differential backup: once a day
 - Transaction log backup: every 5 minutes

- Retention Period:
 - 7-35 days (default is 7)
 - No native long term retention support

- Availability:
 - Backup is stored in a geo-redundant storage
 - In General Purpose and Memory Optimized tiers
 - SLA: 99.99%

Azure MySQL Pricing

- Pricing based on:
 - Tier:
 - Basic Require light compute and I/O performance (ie. dev)
 - General Purpose Most business workloads
 - Memory Optimized Require in-memory performance
 - Compute (no. of vCores)

Azure MySQL Pricing

Flexible Server deployment – currently in preview, not recommended

Reservations exist

Azure Database for MySQL



DEPLOYMENT OPTION:

Single Server



COMPUTE:

Gen 5, 1 vCore, \$0.0398/hour ➤

Savings Options

Save up to 51% on pay as you go prices with the 1 year reserved option.

1 year reserved option is not available for your instance selection.

3 year reserved option is not available for your instance selection.

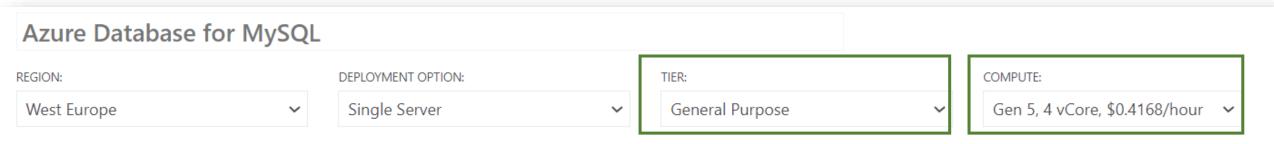
- O Pay as you go
- 1 year reserved
- 3 year reserved

\$29.05

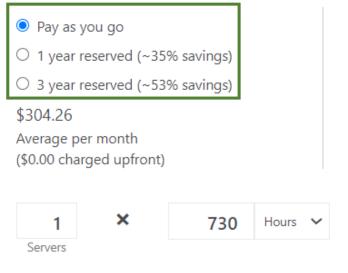
Average per month (\$0.00 charged upfront)







Save up to 51% on pay as you go prices with the 1 year reserved option.





- Managed PostgreSQL on Azure
- Works like any other PostgreSQL database using the same tools
- Great compatibility with on-prem PostgreSQL database
- Includes Hyperscale deployment
- Offers built-in security, backups, availability and more

- Security:
 - IP firewall rules
 - Service Endpoints
 - Private Endpoints
 - Regular & Azure AD Authentication
 - Secure communication (TLS)
 - Data encrypted by default

- Backup:
 - Depends on storage size:
 - up to 4GB:
 - Full backup: once a week
 - Differential backup: twice a day
 - Transaction log backup: every 5 minutes

- Backup:
 - Depends on storage size:
 - up to 16GB:
 - Full backup: Once created
 - Differential backup: thrice a day
 - Transaction log backup: every 5 minutes

- Retention Period:
 - 7-35 days (default is 7)
 - No native long term retention support

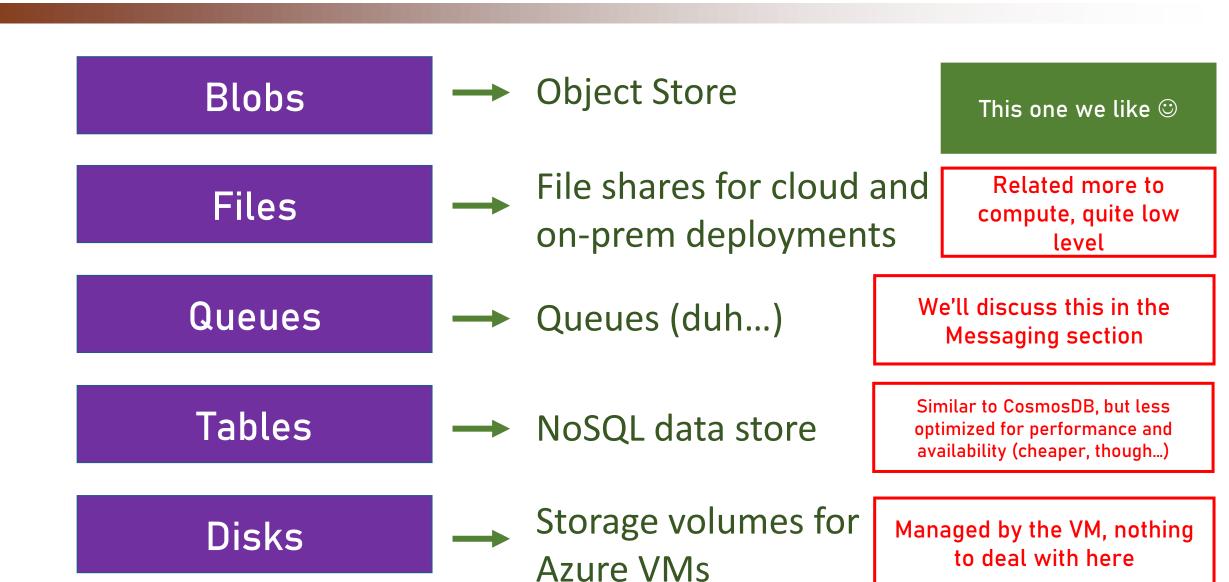
Azure PostgreSQL

- Availability:
 - Backup is stored in a geo-redundant storage
 - In General Purpose and Memory Optimized tiers
 - SLA: 99.99%

Azure Storage

- Object store
- Massively scalable
- Accessible via HTTP or HTTPS
- Client libraries for almost every language
- Durable and highly available

Azure Storage Types



Azure Blobs Storage

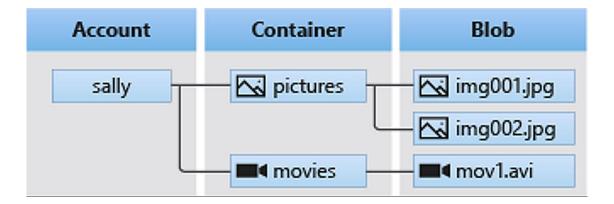
- Object store (Blob = Binary Large Object)
- Great for files, videos, documents, large texts etc.
- Up to 4.77TB per file (!), 190TB in preview (!!)
- Extremely cost effective
- Massively scalable
- Great availability options
- Extremely easy to use
- Usually used in conjunction with SQL / NoSQL database

Azure Blobs Storage

- Security:
 - IP firewall rules
 - Service Endpoints
 - Private Endpoints
 - Shared Access Signatures
 - Access Keys & Azure AD Authentication
 - Secure communication (TLS)
 - Data encrypted by default

Azure Blobs Storage

Structure:



Azure Blobs Storage Redundancy

6 options:

LRS

Locally Redundant Storage
Data is synchronously copied 3
times within the same zone

GZRS

Geo-Zone Redundant Storage Data is synchronously copied to 3 zones in the Region, and then copied asynchronously to paired Region.

Data in the secondary Region is accessible only after Failover process

ZRS

Zone Redundant Storage
Data is synchronously copied to
3 zones in the Region

RA-GRS

Read Access-Geo Redundant Storage
Data is synchronously copied 3 times within
the same zone, and then copied
asynchronously to paired Region.
There's a read-access to the data in the
secondary Region

GRS

Geo Redundant Storage

Data is synchronously copied 3 times within the same zone, and then copied asynchronously to paired Region.

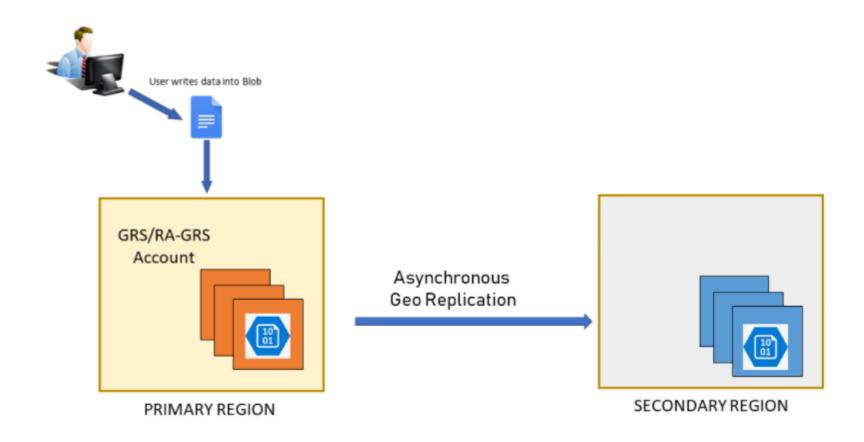
Data in the secondary Region is accessible only after Failover process

RA-GZRS

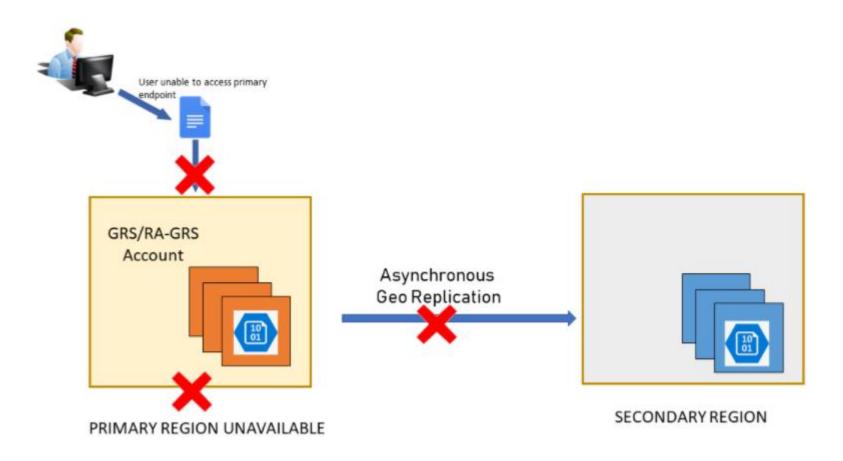
Read Access-Geo-Zone Redundant Storage

Data is synchronously copied to 3 zones within the same Region, and then copied asynchronously to paired Region.
There's a read-access to the data in the secondary Region

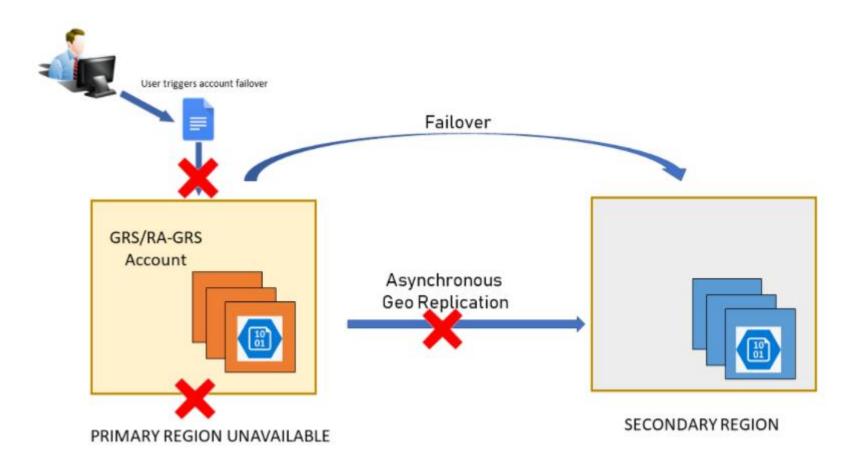
Normal circumstances:



Primary Region fails:



Failover to the secondary Region:



- Can be initiated via the:
 - Portal
 - Azure CLI
 - PowerShell

Azure Blobs Storage Tiers

Blobs are uploaded to one of three tiers:

Hot

- Data that's accessed frequently
- Best SLA (99.9%)
- Highest storage costs
- Lowest access costs
- Examples:
 - Photos to display
 - Documents to show

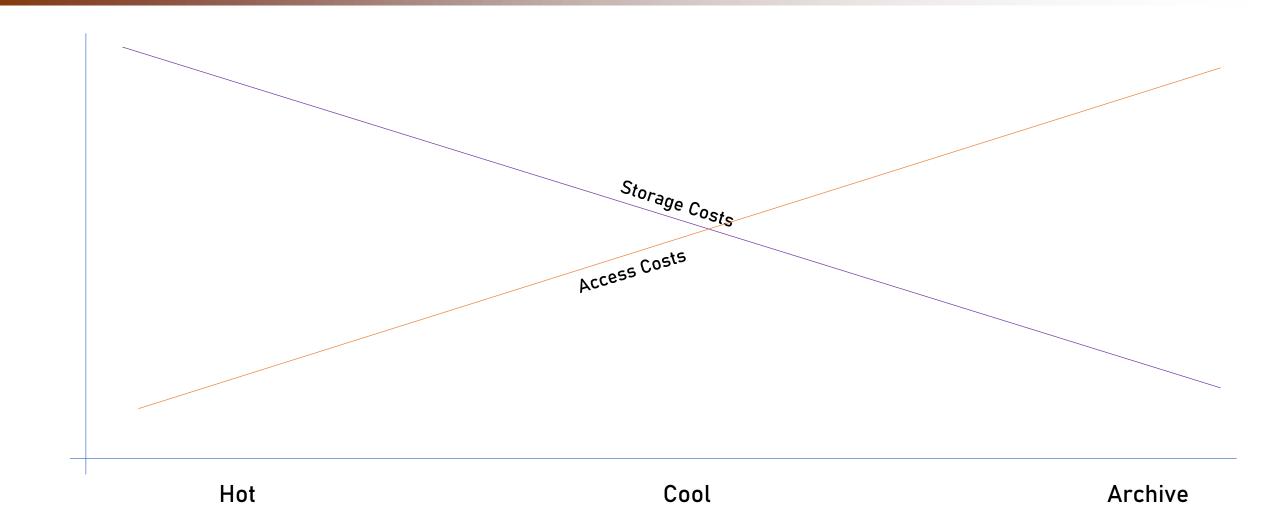
Cool

- Data that's accessed infrequently
- Slightly lower SLA (99%)
- Lower storage costs
- Higher access costs
- Must be stored for at least 30 days (or early deletion fees applied)
- Examples:
 - Short term backup
 - Data for future processing

Archive

- Data for archival
- Stored offline, no SLA
- Can take hours to retrieve
- Lowest storage costs
- Highest access costs
- Must be stored for at least 180 days (or early deletion fees applied)

Azure Blobs Storage Tiers

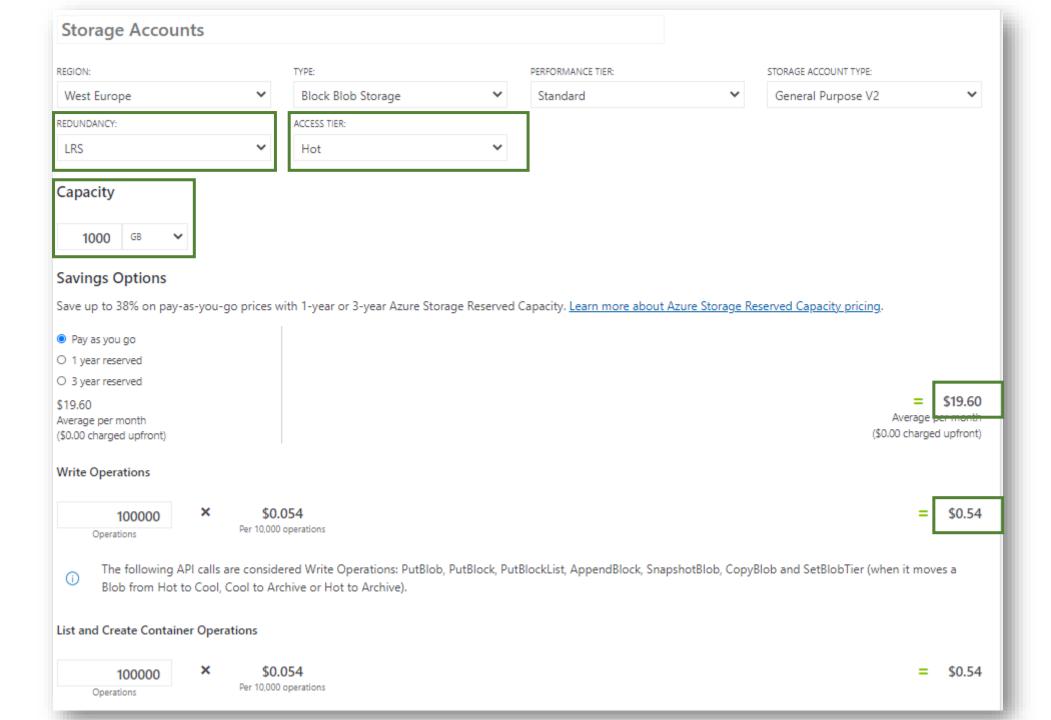


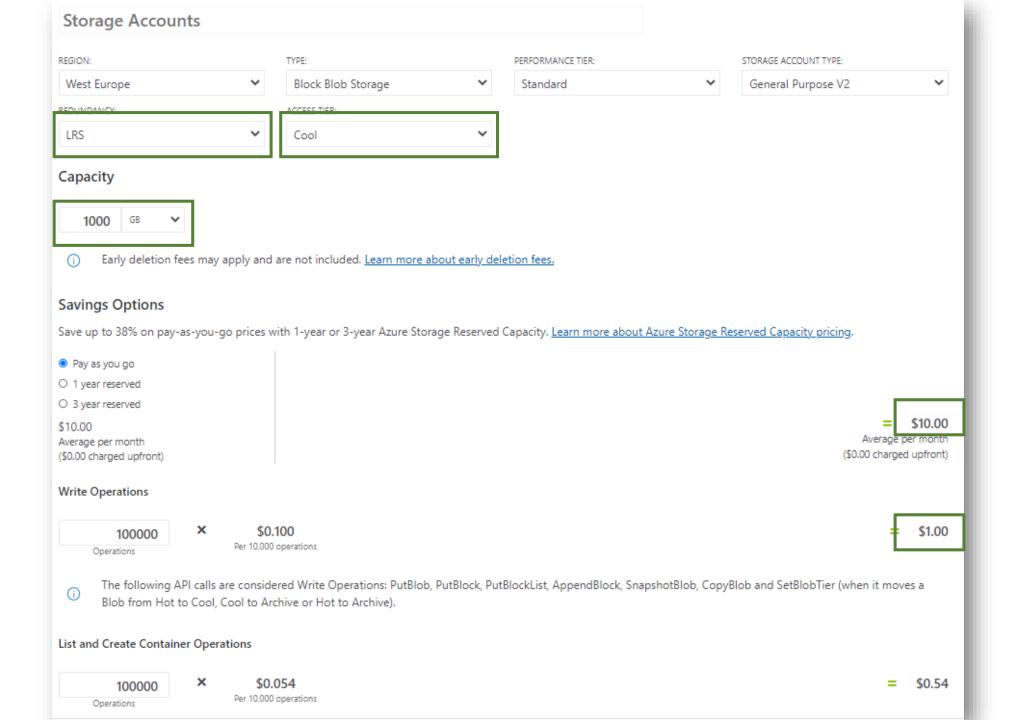
Azure Blobs Storage Tiers

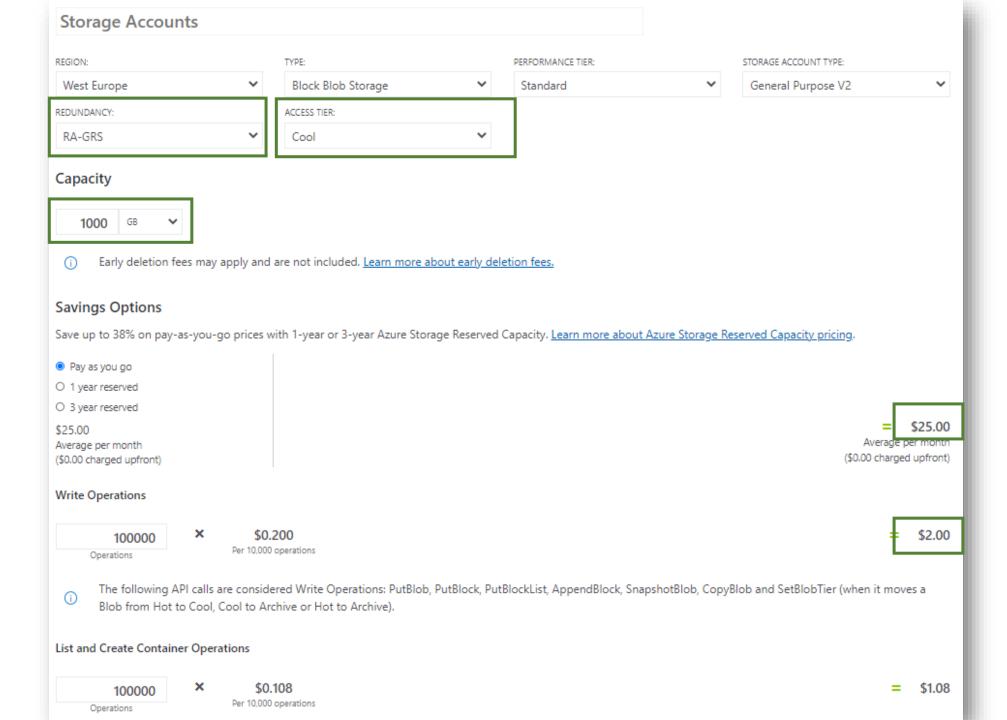
- Retrieval time is the same in Hot and Cool tiers
- Archive tier does not support ZRS, GRS and RA-GRS redundancy
- Using RA-G(Z)RS, SLA improves to 99.99% (Hot) and 99.9% (Cool)
- Tier is set at account level, can be modified per blob
- Moving between tiers can be automated by lifecycle rules

Azure Blobs Storage Pricing

- Based on:
 - Redundancy option
 - Access tier
 - Capacity







Azure Redis

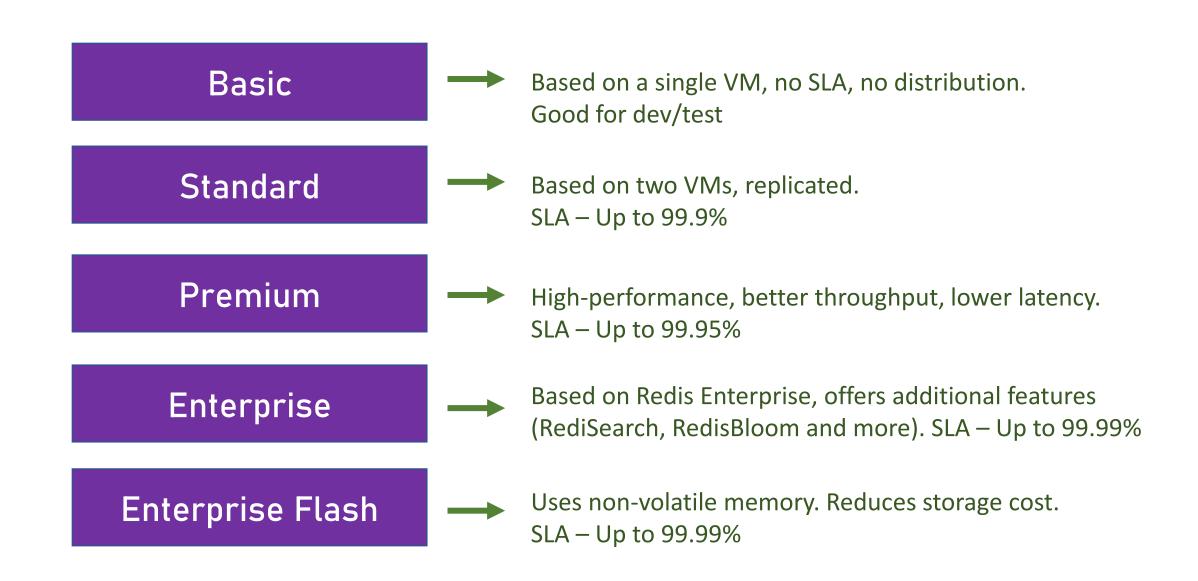
- Managed Redis on Azure
- Provides lightning-fast in-memory, distributed cache
- Great for short-lived, frequently accessed data
 - ie. shopping cart, stock quotes
- Fully compatible with OSS Redis (community edition) and

Enterprise Redis - depends on service tiers

Azure Redis

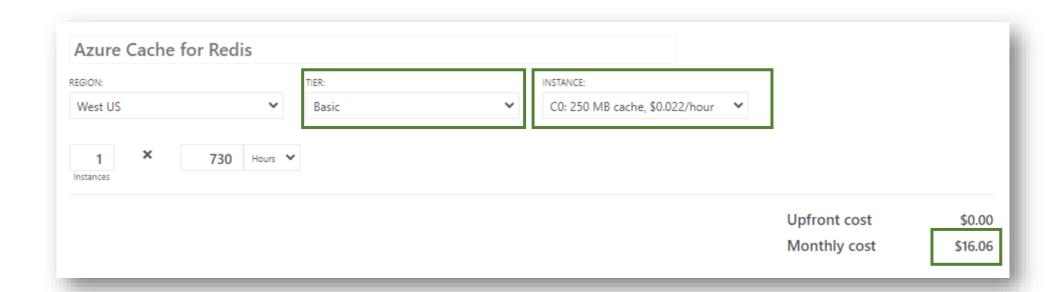
- Security:
 - IP firewall rules
 - Service Endpoints
 - Private Endpoints
 - Secure communication (TLS)

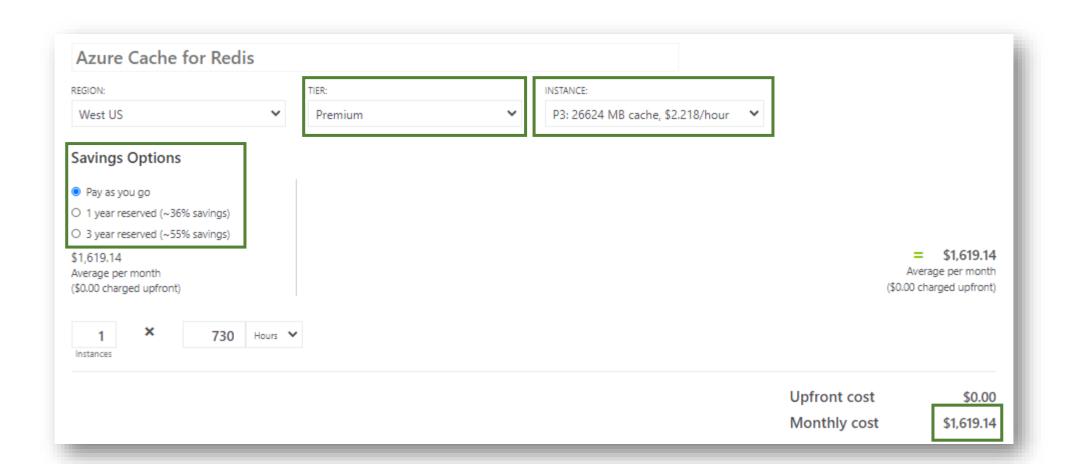
Azure Redis Service Tiers



Azure Redis Pricing

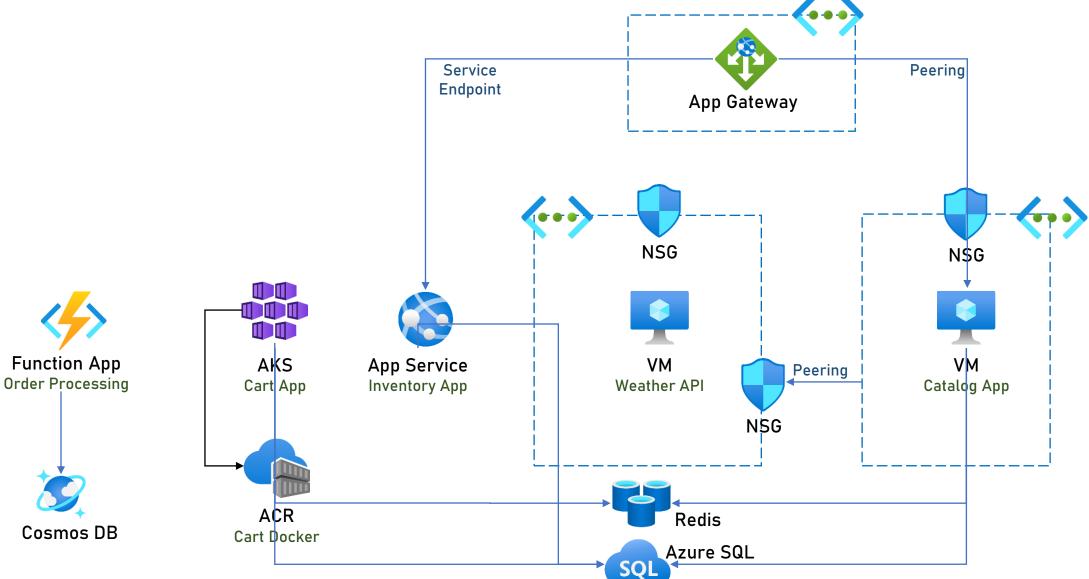
- Based on:
 - Tier
 - Memory





ReadIt!

Cloud Architecture



Are we done?

- Nope ☺
 - The Orders Function is publicly available and synchronous
 - The inventory page is open for everyone
 - We don't really know how the app is functioning
 - The website is not redundant what happens if the whole region goes down?

Selecting Data Store Solution

Data Type	Used For	Examples	Options in Azure
Relational	Structured data	Items in store, demographic data	Azure SQL, MySQL, PostgreSQL
NoSQL	Semi-structured data	Reviews, Log records, when flexibility is required	Cosmos DB (with SQL, Mongo, Azure Table API)
Graph	Data representing relationships	Family tree	Cosmos DB (with Gremlin API)
Blob	Files, videos, docs	Items' photos	Azure Blob Storage