



#37 PARMA 2023

# SQL Server PaaS

Gianluca Hotz



# Who am I?



- Gianluca Hotz | @glhotz | ghotz@ugiss.org
- Independent Consultant
  - 25+ years on SQL Server (from 4.21 back in 1996)
  - Database modeling & development, sizing & administration, modernization (upgrades & migrations), performance tuning, security
- Community
  - 25 years Microsoft [MVP](#) SQL Server/Data Platform (since 1998)
  - VMware Experts SQL Server
  - Founder and president [UGISS](#) (ex «PASS Chapter»)





UNIVERSITÀ DEGLI STUDI DI PARMA

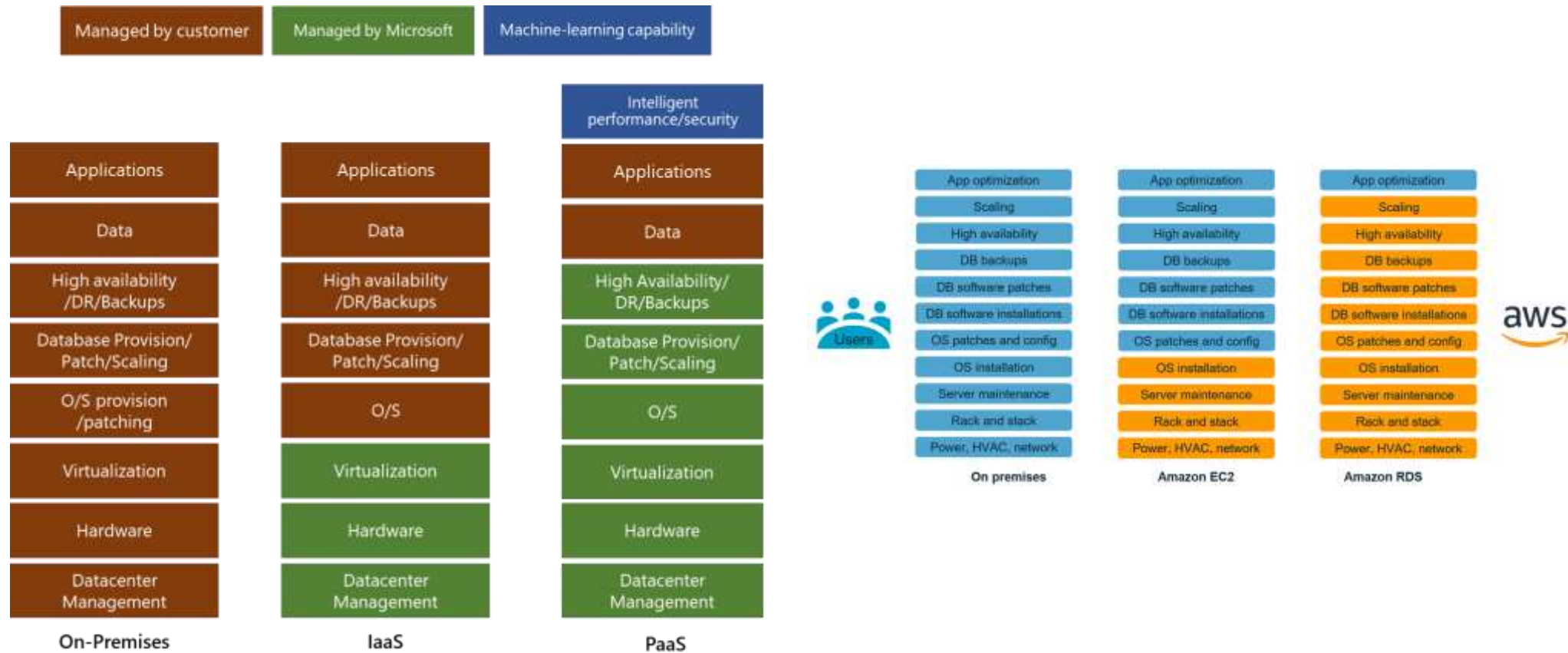


Bi Factory

DATA KNOWLEDGE ADVISOR



# «On-premises, IaaS & PaaS»



# «Azure & AWS IaaS»

## Virtual Machine in Azure

- Ambienti Windows/Linux, agnostico rispetto a «engine» che si possono installare
- Marketplace con immagini per i più diffusi (anche non relazionali)
- [Azure Dedicated Host](#)

## Diversi servizi per «Container»

- [Azure Kubernetes Service \(AKS\)](#)
- [Azure Red Hat OpenShift](#)
- [Azure Container Instances](#)

## Altri scenari

- [Azure VMware Solution](#)

## Istanze EC2 in AWS

- Ambienti Windows/Linux, agnostico rispetto a «engine» che si possono installare
- Marketplace con immagini per i più diffusi (anche non relazionali)
- [Amazon EC2 Dedicated Hosts](#)

## Diversi servizi per «Container»

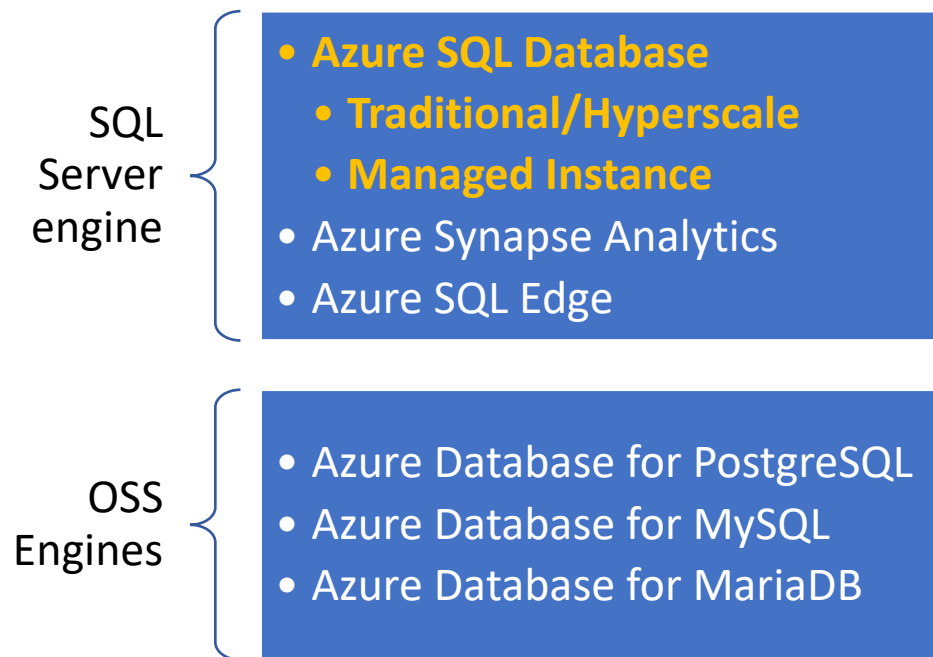
- [Amazon Elastic Kubernetes Service \(EKS\)](#)
- [Amazon Elastic Container Service \(ECS\)](#)
- [Red Hat OpenShift Service on AWS](#)

## Altri scenari

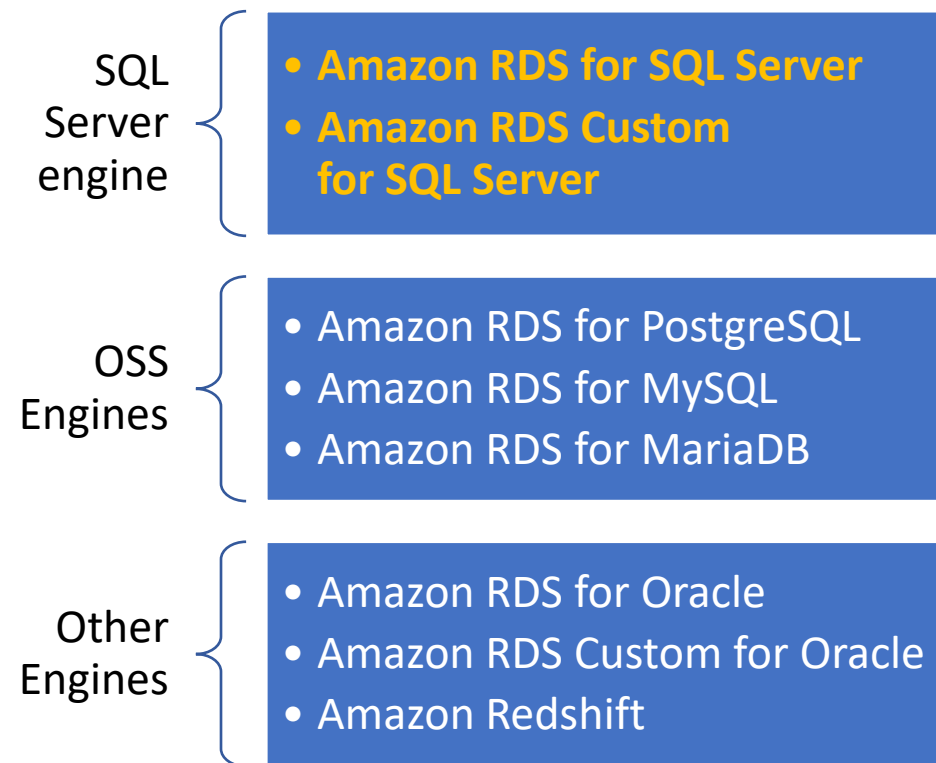
- [VMware Cloud on AWS for SQL Server](#)

# «PaaS Relational Data Services»

## Microsoft Azure



## Amazon AWS



# Deployment Azure SQL Database



## Azure SQL Database

### Single

Modello a singolo database con prestazioni carico di lavoro prevedibili

#### Best for

applicazioni che richiedono risorse garantite a livello di database

### Elastic Pool

Modello molteplici database con risorse condivise, ottimizzato per maggiore efficienza applicazioni multi-tenant

#### Best for

Applicazioni in modalità SaaS con molteplici database che possono condividere risorse

### Managed Instance

Modello a istanza con elevata compatibilità con SQL Server IaaS mantenendo i vantaggi di un PaaS

#### Best for

modernizzazione su larga scala con minori sforzi e attriti

Service Tiers

General Purpose/Standard

Business Critical/Premium

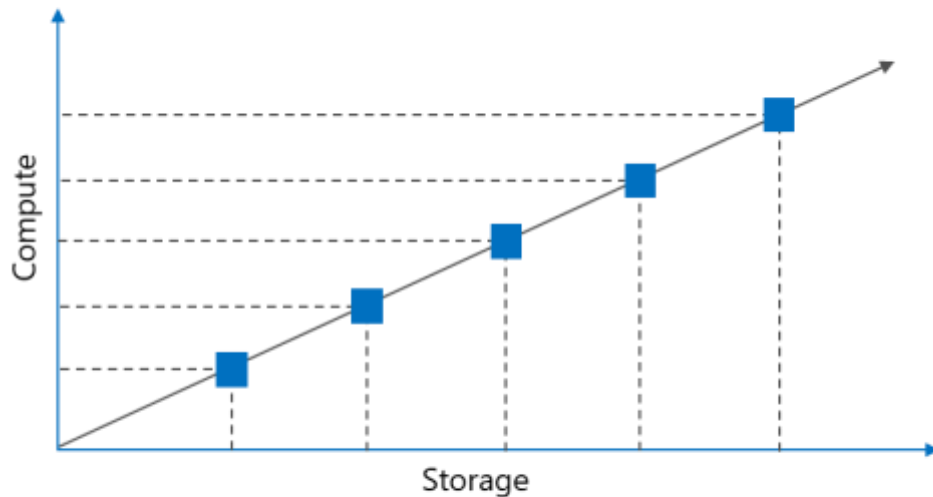
Hyperscale

Serverless

# Modelli di acquisto

## Modello DTU

Semplice, preconfigurato

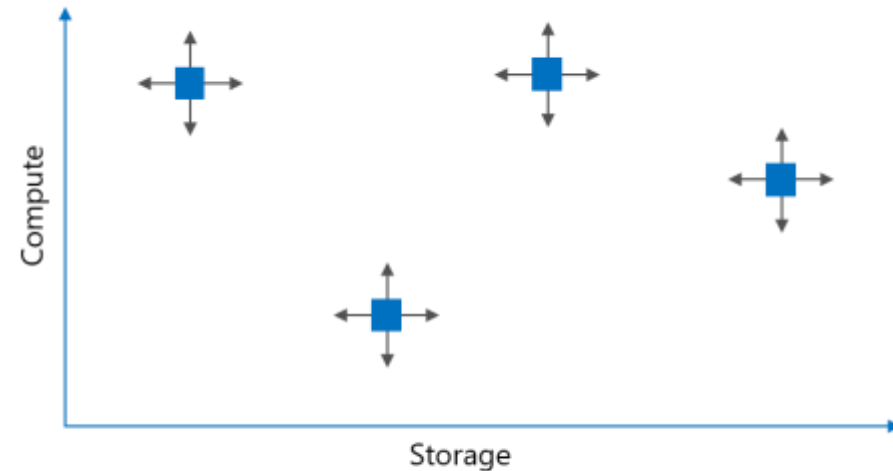


DTU sono unità pre-confezionate che rappresentano potenza computazionale  
Pensate per carico di lavoro con prestazioni predicibili ma limitate in termini di flessibilità e opzioni  
Dimensionamento basato su DTU offre semplicità di scelta

<https://learn.microsoft.com/en-us/azure/azure-sql/database/service-tiers-dtu>

## Modello vCore

Scalabilità indipendente

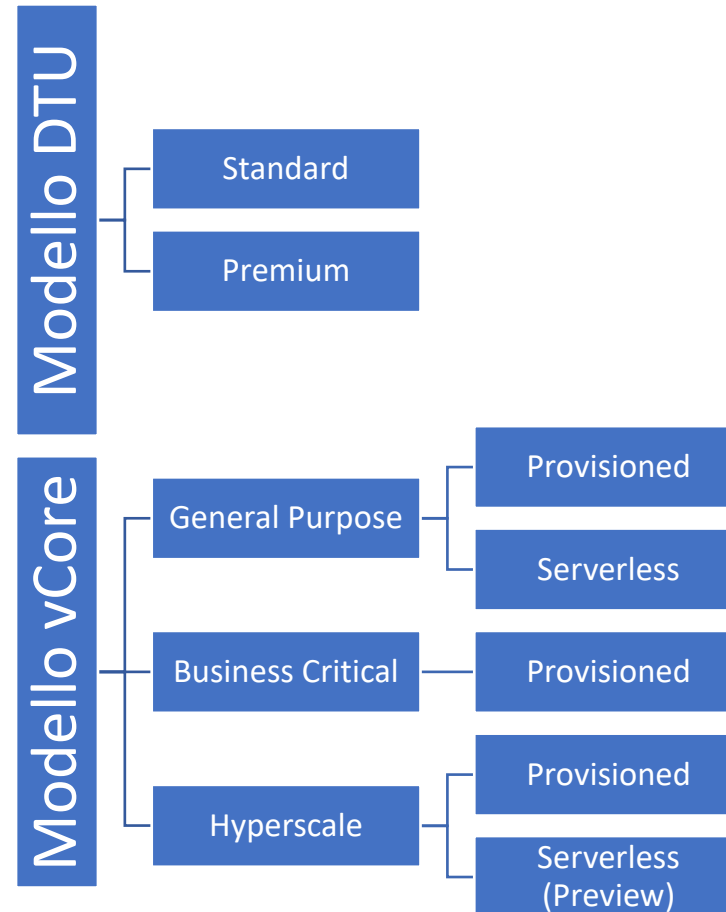


Modello che permette di scegliere risorse «Compute» e «Storage» in modo indipendente  
Permette anche di usare «Azure Hybrid Benefit for SQL Server» per risparmiare  
Offre flessibilità, controllo e trasparenza

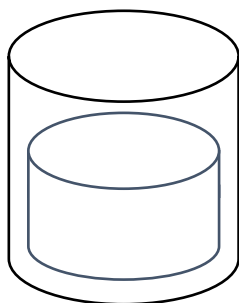
<https://learn.microsoft.com/en-us/azure/azure-sql/database/service-tiers-sql-database-vcore>



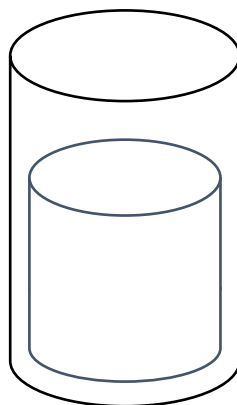
# Livelli di servizio («Service Tiers»)



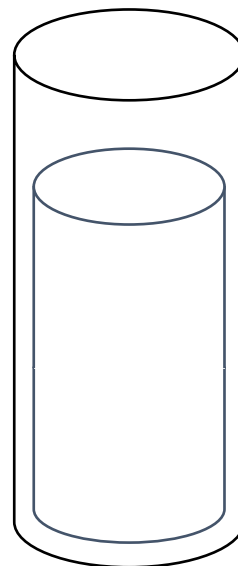
# Risorse preallocate e scalabilità



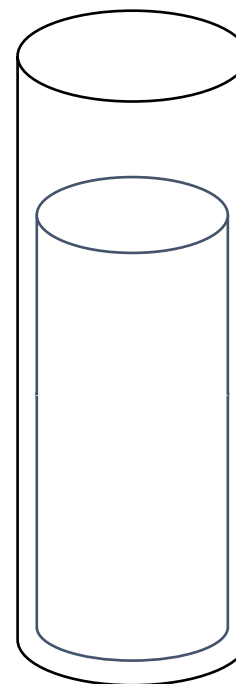
S0



S1

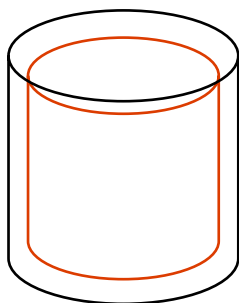


S2

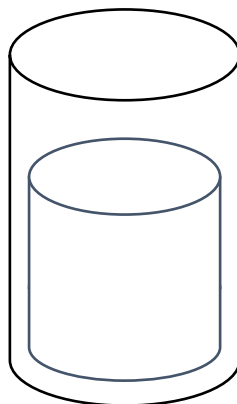


S3

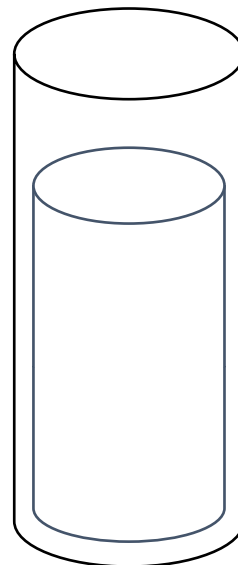
# Risorse preallocate e scalabilità



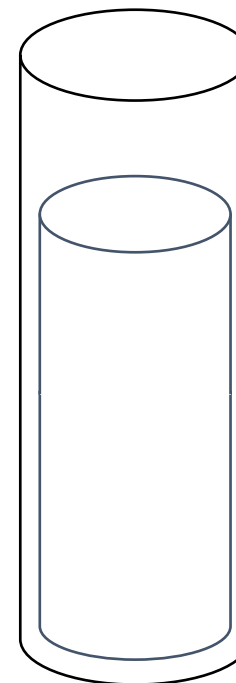
S0



S1



S2




S3

# «Scale up» e «Scale down»

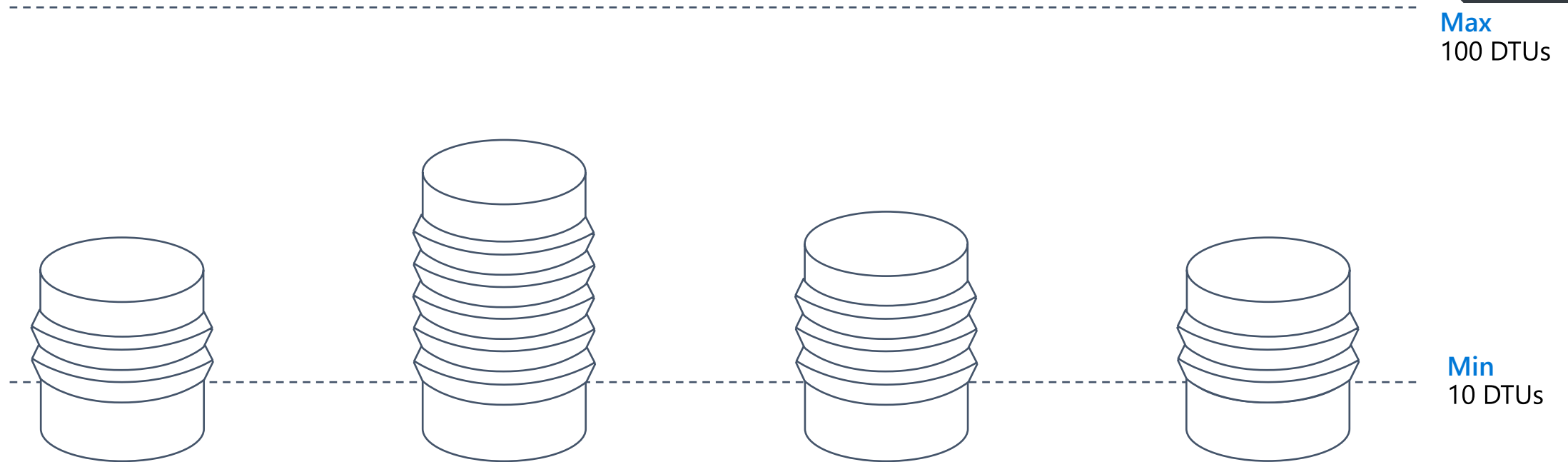
- Modifica Livello di servizio
  - «Service Tier/Performance Level»
  - «Hyperscale» e «Serverless» casi speciali
- Operazione «Online»
  - Copia o «attach/detach»
  - Database sempre disponibile, interruzione durante «switch»
  - Durata dipende dalla dimensione del database
- Attenzione a caratteristiche usate
  - Es. dimensione inferiore a massimo spazio disponibile
  - Es. fatturazione spazio eccedente quello incluso nel tier



# «Elastic Database Pools»

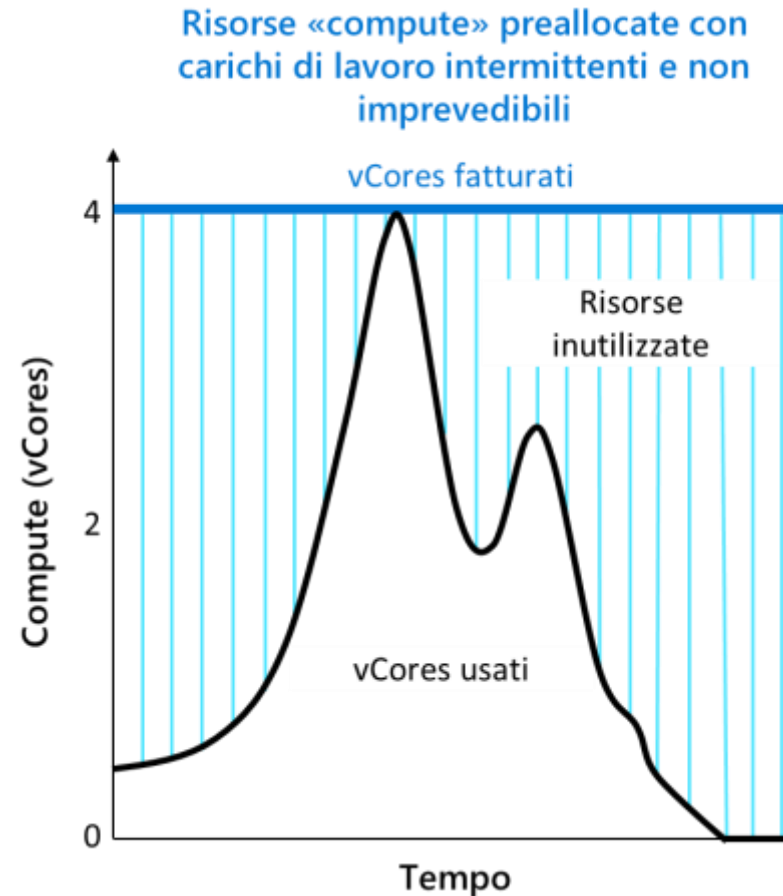
- Pool di risorse condivise da più database
  - «Auto-Scale» database entro range definiti
  - Database aggiunti/rimossi a caldo
- 

# «Elastic Database Pools»



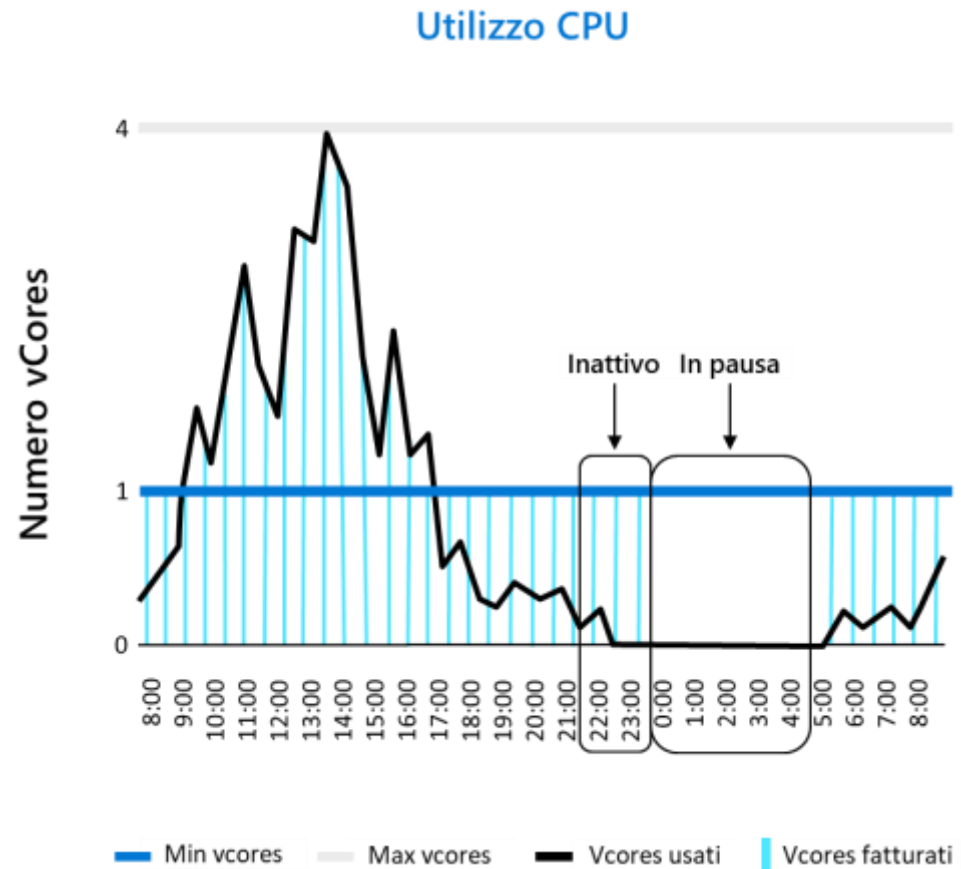
# Inefficienza preallocazione CPU

- Preallocazione
  - efficace con pattern prevedibili
- Inefficienze
  - Sovrallocazione in periodi di utilizzo minore
  - Sottoallocazione per ridurre costi
  - Tempo speso a gestirla



# «Serverless»

- Min/Max vCore in base a SLO
  - Es. 0,5-4; 0,75-6; 1-8; 10-80
- Fatturazione al secondo
- «Autopause Delay»
  - Costo zero (solo CPU & RAM)
- Al momento solo
  - General Purpose (vCore)
  - «Hyperscale» in preview
  - Gen 5 Hardware







# Database in pausa



- «Autopausing»
  - Numero di sessioni = 0, CPU = 0 per carico di lavoro del pool
  - Minimo 1 ora, massimo 7 giorni, incrementi da 10 minuti
  - Può essere disabilitato
- «Autoresuming»
  - In generale a fronte di un login (ma anche altre operazioni)
  - Connessione riceve errore, logica di «retry» obbligatoria..
  - Latenza: generalmente 1 minuto

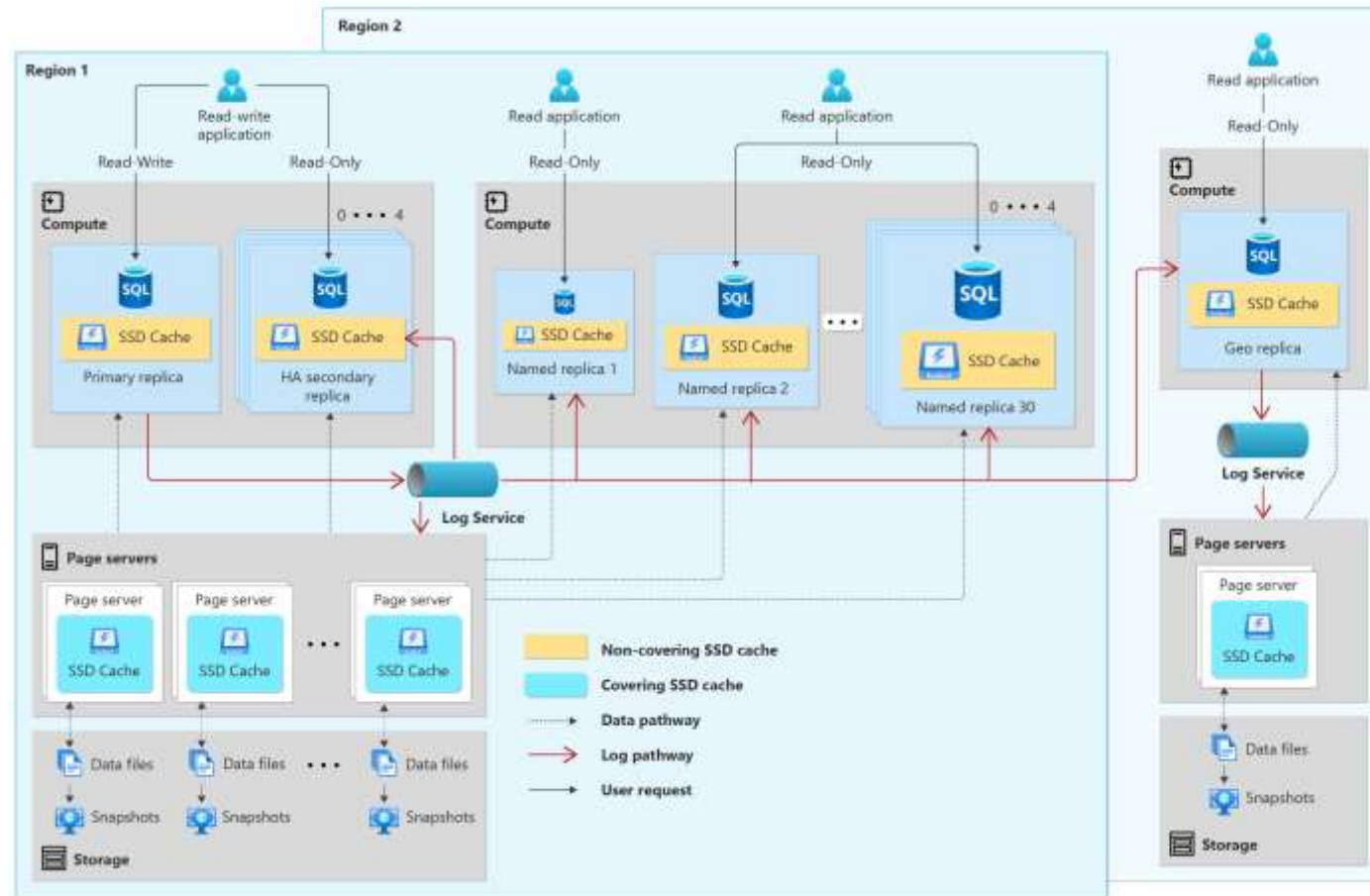


# Dettagli sulle prestazioni...



- Reattività ridimensionamento
  - Raramente risorse non disponibili... necessario «load balacing»
    - fino a qualche minuto con chiusura connessioni
- Gestione della memoria
  - Cache reclamata più frequentemente, può impattare
  - Mai sotto dimensione specifica per numero minimo vCores

# Architettura «Hyperscale»

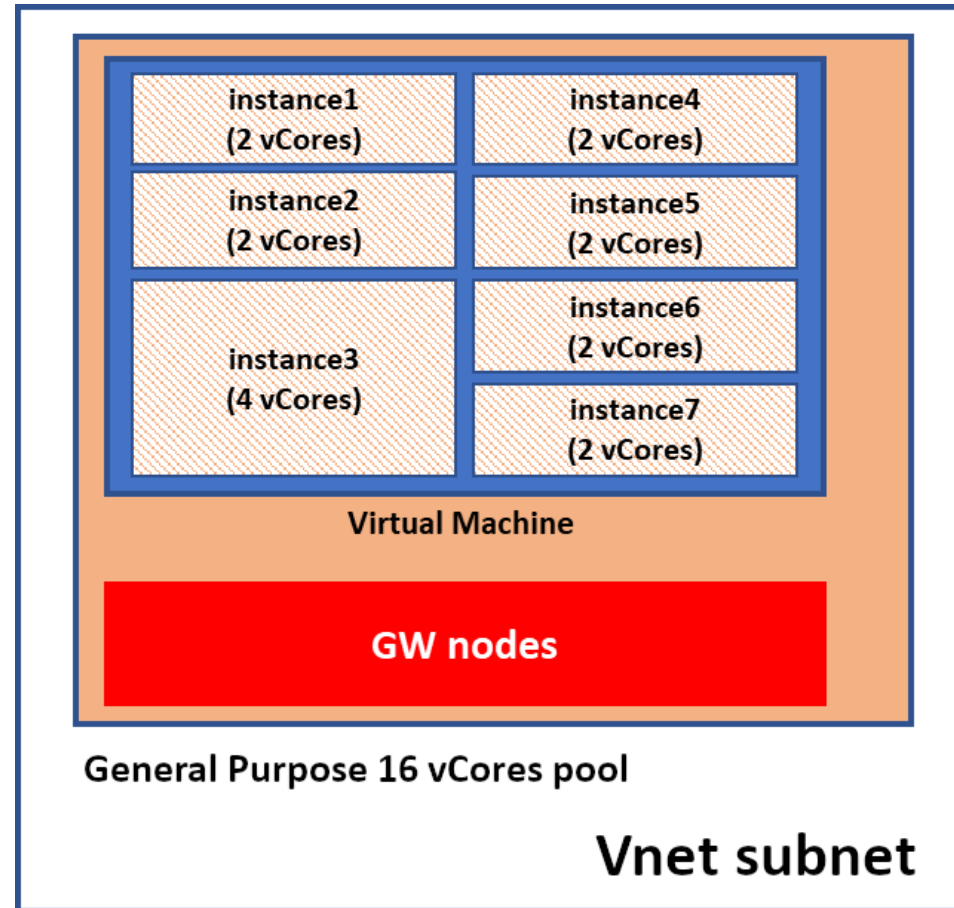


# SQL Database Managed Instance

- Opzione di Deployment di SQL Database
  - Rappresenta ed espone l'intera istanza di SQL Server
  - Molto differente «Azure SQL Database Logical Server»
- Maggiore copertura funzionalità «On-Premise»
  - Es. SQL Server Agent, Backup/Restore, CDC, CLR, Broker ecc.
  - <https://docs.microsoft.com/en-us/azure/azure-sql/database/features-comparison>
- Mantiene tutti i vantaggi di un «PaaS»
  - Completamente gestita
  - «Scale up/down» semplice e «rapido»
  - HA/DR inclusi

# Managed Instance Instance Pools

- Più MI in 1 Virtual Machine
- Possibili istanze con 2 vCore
- «[Public Preview](#)»



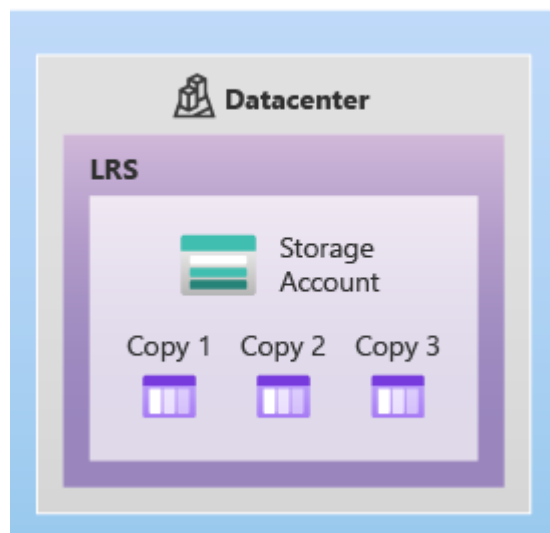
# Azure SQL Database Backups

- [Backup](#) automatici (frequenza e tipo)
  - FULL ogni settimana, DIFF ogni 12-24 ore, LOG ogni ~ 10 minuti
  - «Retention» 7-35 giorni (default 7)
  - «[Long-Term Retention](#)» 10 anni ma solo FULL (RPO 1 settimana)
- [Ridondanza](#) «Storage Backup»
  - LRS, ZRS, GRS, GZRS
- Ripristino «[Point in Time](#)»
  - Database esistente o cancellato stessa regione (10 minuti RPO)
  - «Geo-restore» in altra regione (RPO fino a 1 ora, no PIT)

# Ridondanza Backupset «LRS/ZRS»

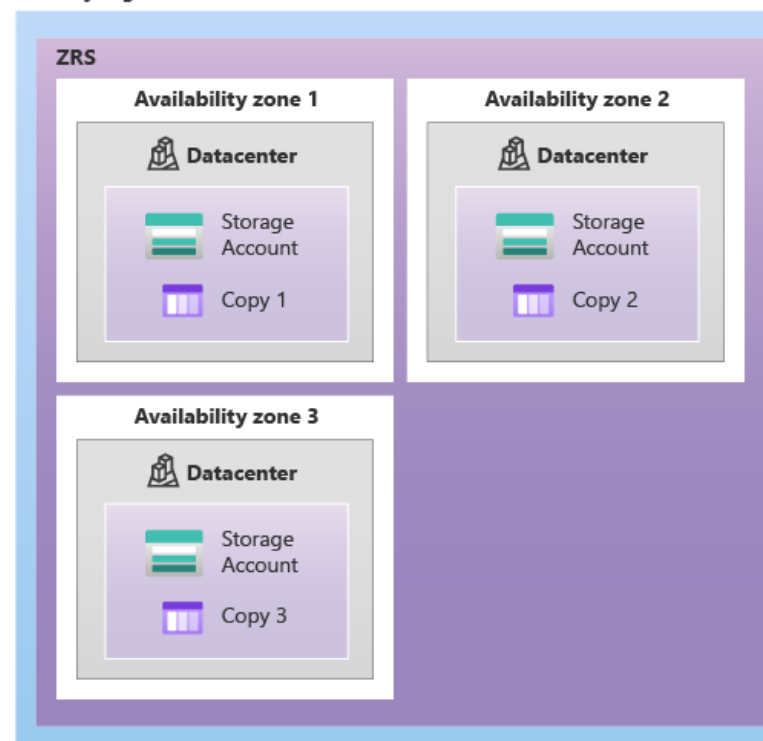
## Locally Redundant Storage

Primary region



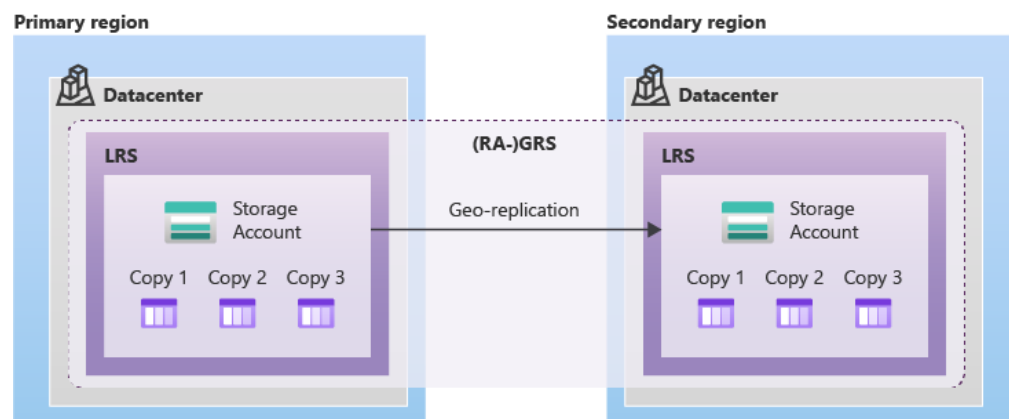
## Zone-redundant storage

Primary region

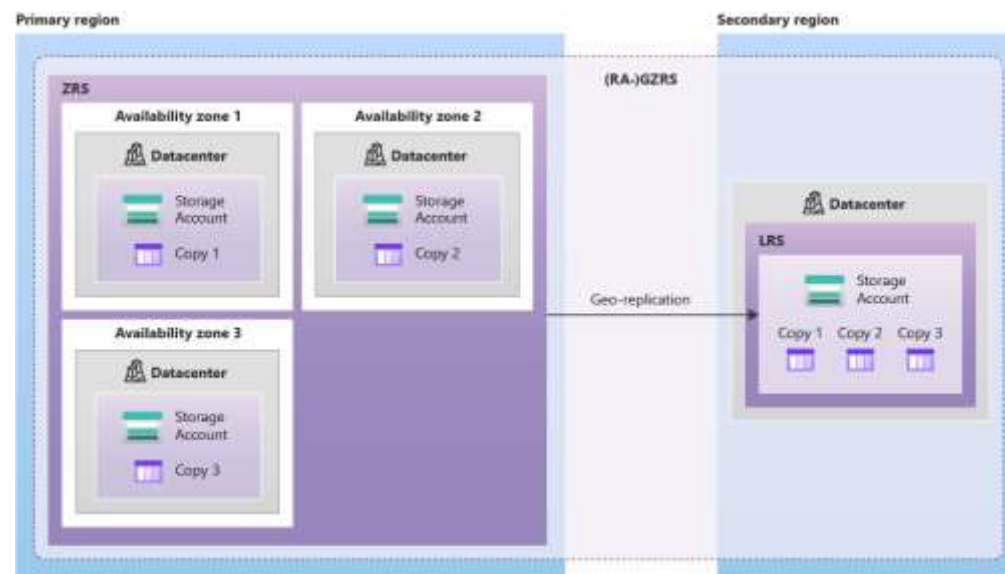


# Ridondanza Backupset «GRS/GZRS»

## Geo-redundant storage



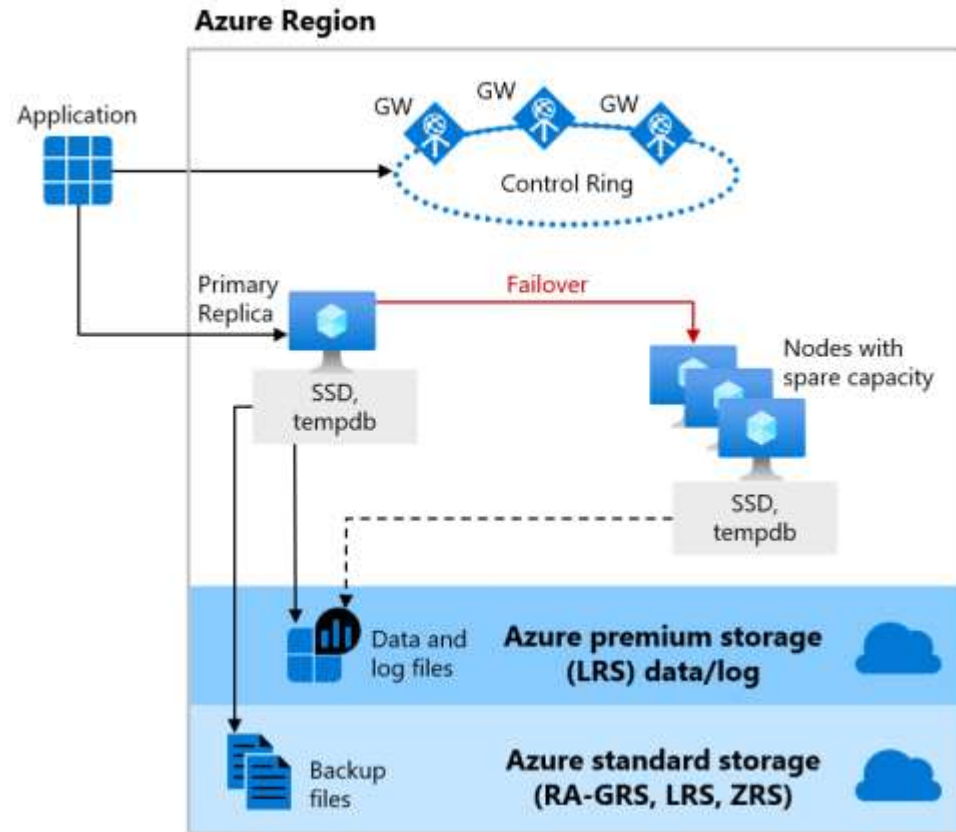
## Geo-zone-redundant storage





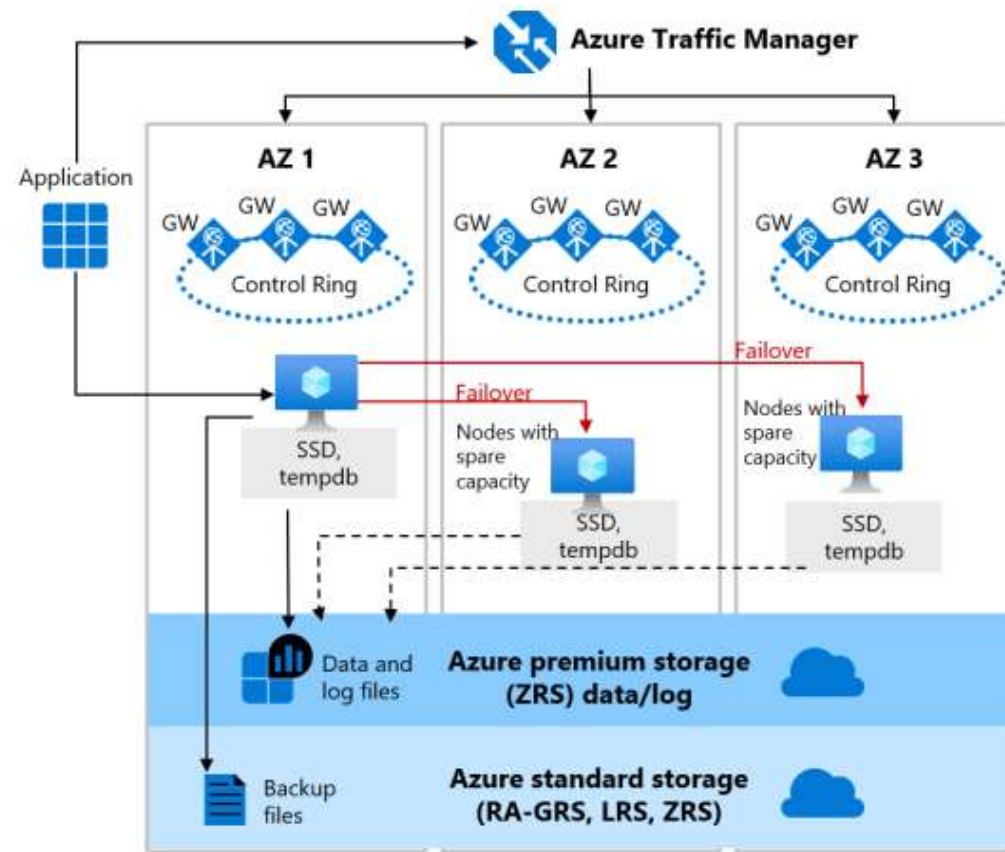
# HA «Standard»

- Livelli di servizio
  - Basic, Standard, General Purpose
- Disaccoppiamento
  - Nodi «compute»
    - Gestiti da Azure Service Fabric
    - Sempre disponibili di scorta
  - «Azure Storage» LRS
    - Disponibilità/ridondanza «built-in»
- Potenziale degrado prestazioni
  - Tempo di transizione
  - Cache fredda



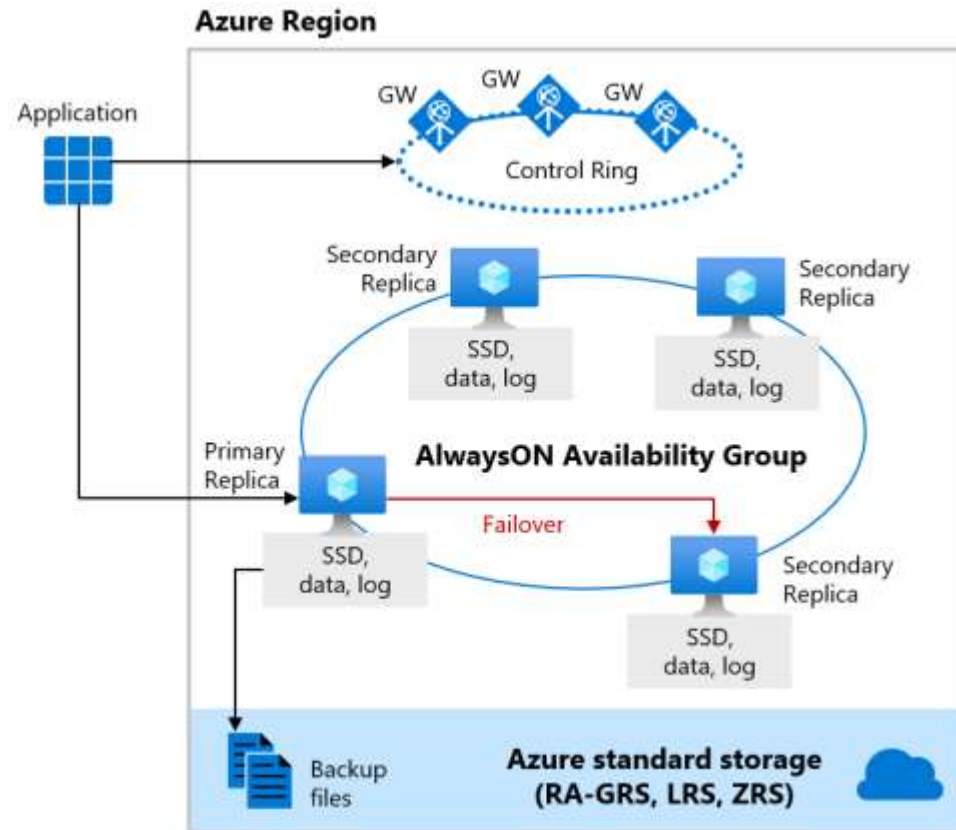
# HA «Standard Zone-Redundant»

- Livelli di servizio
  - General Purpose
- Simile a modello «Standard»
  - Nodi «compute» in più AZ
  - «Azure Storage» ZRS



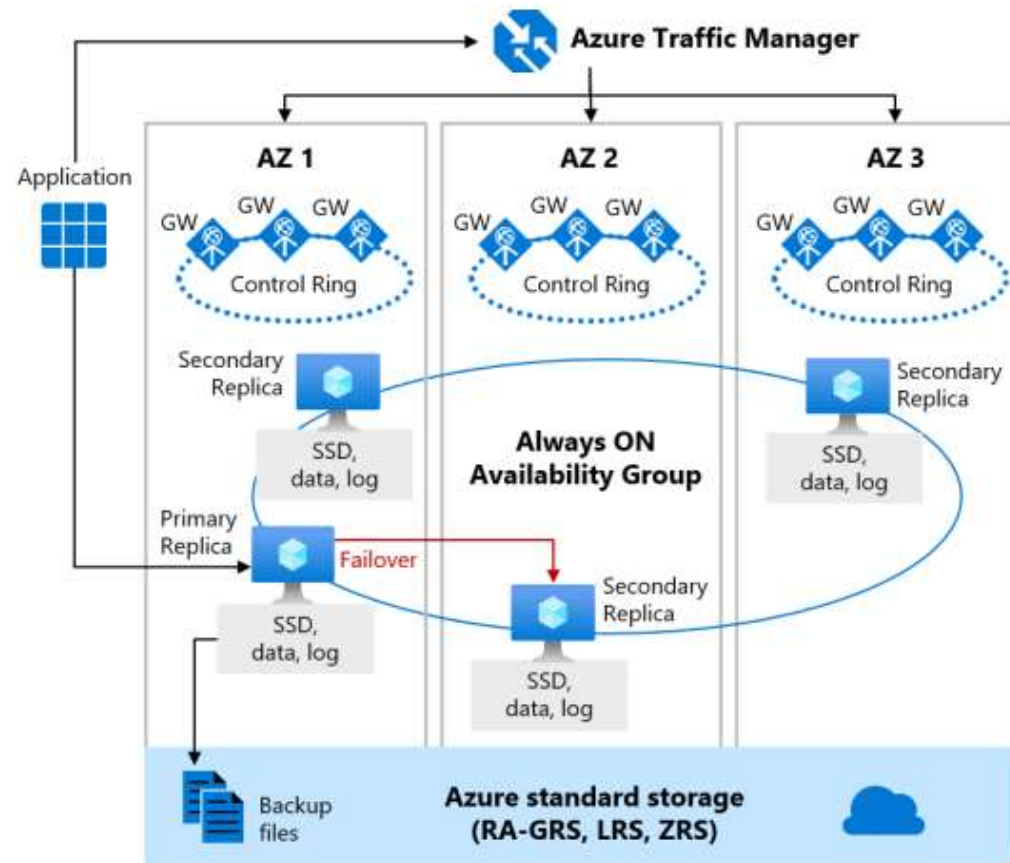
# HA «Premium»

- Livelli di servizio
  - Premium, Business Critical
- Replica «Availability Groups»
  - 1 replica primaria
  - Fino a 3 repliche secondarie
  - 1 in «[Read Scale-Out](#)»



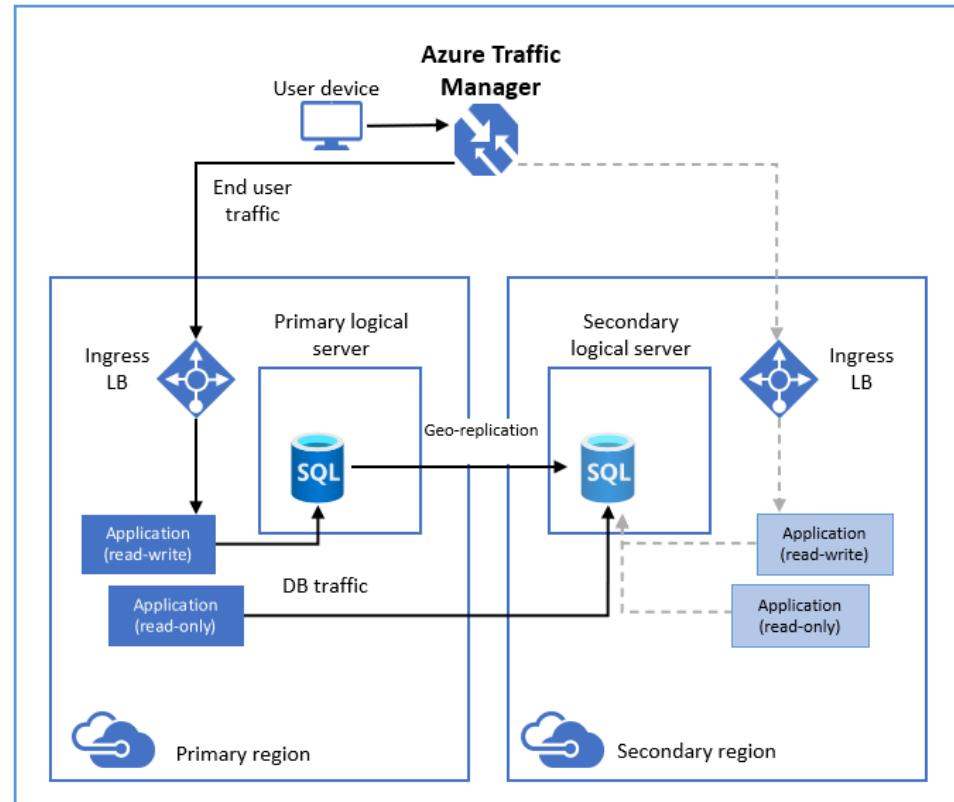
# HA «Premium Zone-Redundant»

- Livelli di servizio
  - Premium, Business Critical
- Simile a modello «Premium»
  - Repliche in più AZ
- Limiti
  - Maggiore latenza per OLTP
  - ~~No~~ Si «Managed Instance»
  - Gen5 HW con Business Critical



# DR «Active geo-replication»

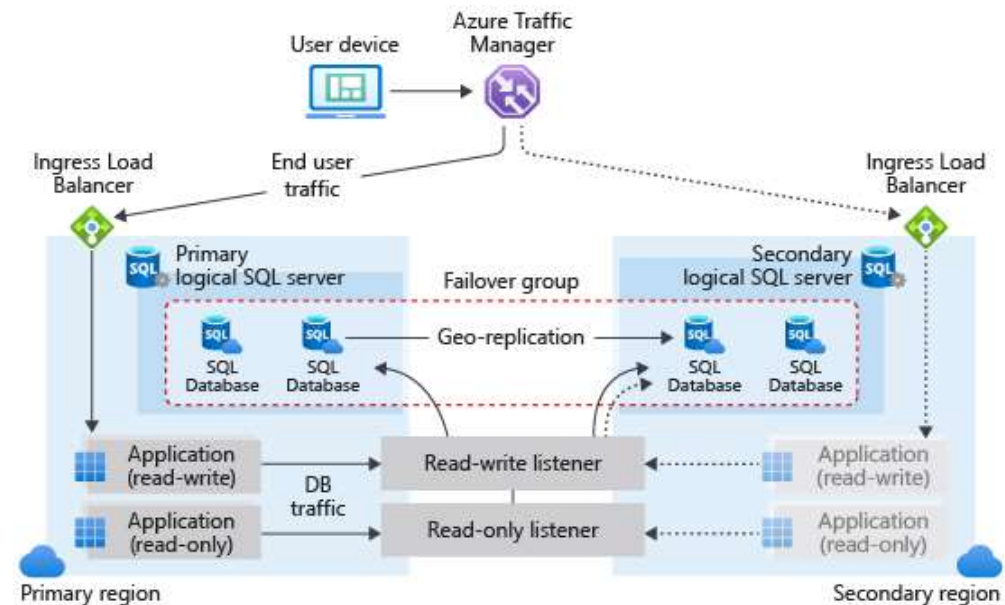
- DR «Cross-Region»
- Basato su «AlwaysOn AG»
  - Asincrono
  - Failover manuale
    - Pianificato o forzato
    - Endpoint cambia!
  - Fino a 4 repliche secondarie
  - Repliche in sola lettura
- Managed Instance
  - Non supportate



-----> Indicates end user traffic after failover to secondary region

# DR «Auto-Failover Groups»

- DR «Cross-Region»
- Basato su «AlwaysOn AG»
  - Asincrono ma...
  - Failover automatico
    - Endpoint non cambia!
    - Possibile perdita di dati!
- Managed Instance [supportate](#)





# Amazon RDS for SQL Server



- Ambiente PaaS completamente gestito
  - Licenza SQL Server inclusa nel prezzo
  - Backup automatici (point-in-time snapshot)
  - «Storage Encryption»
  - «Enhanced Monitoring»
  - Auto-aggiornamento versioni minori
  - Finestra di manutenzione (non solo per auto-aggiornamento)
- Supporto «Multi-AZ»
  - Database Mirroring o AlwaysOn Availability Groups in base a build/edizione
  - «Read Replica» (solo Edizione Enterprise, usa AlwaysOn AG)



# Licensing



- Licenza inclusa
- Versioni
  - 2014-2019
  - Solo alcune «minor build» supportate
- Edizioni
  - SQL Server Express Edition
  - SQL Server Web Edition
  - SQL Server Standard Edition
  - SQL Server Enterprise Edition
  - Developer Edition? Va scaricata e installata su istanza EC2



# Backup Snapshot

- Intera istanza tramite Snapshot a livello di storage
  - Automatici o manuali (manuali max 100 per regione)
  - Retention 0 – 35 giorni, default 7 (da console, 1 da API/CLI)
  - Possibilità di replicare copie in altre regioni
- Ripristino
  - Ripristino creando una nuova istanza
  - Ripristino a qualunque punto del periodo di retention
  - Possibile scostamento fino a 1 secondo tra diversi database...
  - Transazione cross-database potrebbero essere inconsistenti!!
- Billing
  - Solo spazio che eccede la quantità di storage allocato per istanza
  - Es. 100 GiB allocati, 50 GiB snapshot manuali, 75 GiB automatici, 25 pagati



# Backup Nativi SQL Server



- Opzione che deve essere attivata
- Backupset in storage S3 (stessa regione dell'istanza)
- Backup solo FULL e DIFFerenziale
  - Attenzione a Snapshot che cambiano base differenziale
- Restore anche dei LOG (es. migrazione)
- Non pensato per sostituire Snapshot (e DR cross-region)



# Multi-AZ Deployment



- Copia sincrona in altra AZ per alta disponibilità
- Scelta automatica in base a build/edizione
  - Always On Availability Groups
    - Versioni 2016 SP2 CU3+, 2017 CU13+ edizione Enterprise
    - Versione 2019+ edizioni Standard/Enterprise
  - Database Mirroring (DBM)
    - Versioni 2012-2017 edizioni Standard/Enterprise
- Diversa AZ ma nella stessa «Region»
- Login e permessi replicati automaticamente (ma non job e altro)
- Solo una copia di standby



# Multi-AZ Read-Replica



- Fino a 5 repliche secondarie asincrone in sola lettura
- SQL Server 2016 SP2 CU3+ solo Enterprise Edition
- Replica primaria Multi-AZ con Always On Availability Group



# Altre componenti «managed»



- [SQL Server Analysis Services](#) (solo Tabular)
- [SQL Server Integration Services](#)
- [SQL Server Reporting Services](#)
- [Microsoft Distributed Transaction Coordinator](#)

# Amazon RDS Custom for SQL Server

- Servizio gestito RDS con
  - accesso all'istanza EC2 sottostante
  - configurazione personalizzabile
  - con alcuni limiti
- Decision Matrix
  - Tra EC2, RDS e RDS Custom



# Azure Database Fleet Manager



# Fleet Manager Demo







#37 PARMA 2023

Grazie!!!

