



Azure SQL Managed Instance Fundamentals

Module 2



Learning Units covered in this Module

- Lesson 1: Introduction to Managed Instances
- Lesson 2: Service Tiers and Business Continuity
- Lesson 3: Managed Instances Best Practices
- Lesson 4: Managed Instance Performance Concepts

Lesson 1: Introduction to Managed Instances

Azure SQL managed instance eases cloud migration



Customer challenge

I want to migrate to the cloud, remove management overhead, but I need instance-scoped features (Service Broker, SQL Server Agent, CLR...)



Solution

Managed instance combines leading security features with SQL Server compatibility and business model designed for on-premises customers

Key features

- Single instance or instance pool
- SQL Server surface area (vast majority)
- Native virtual network support
- Fully managed service
- On-premise identities enabled with Entra ID and AD Connect

Azure differentiators

- Near zero downtime migration using log shipping
- Fully managed business continuity with failover groups
- Projected return on investment of 212 percent over three years¹
- The best of SQL Server with the benefits of a managed service



Komatsu easily migrated 1.5 TBs of data thanks to near complete compatibility with SQL Server, plus 49% cost reduction and 25-30% performance gains.

1. Forrester Consulting. The Total Economic Impact™ of Microsoft Azure SQL Database Managed Instance.

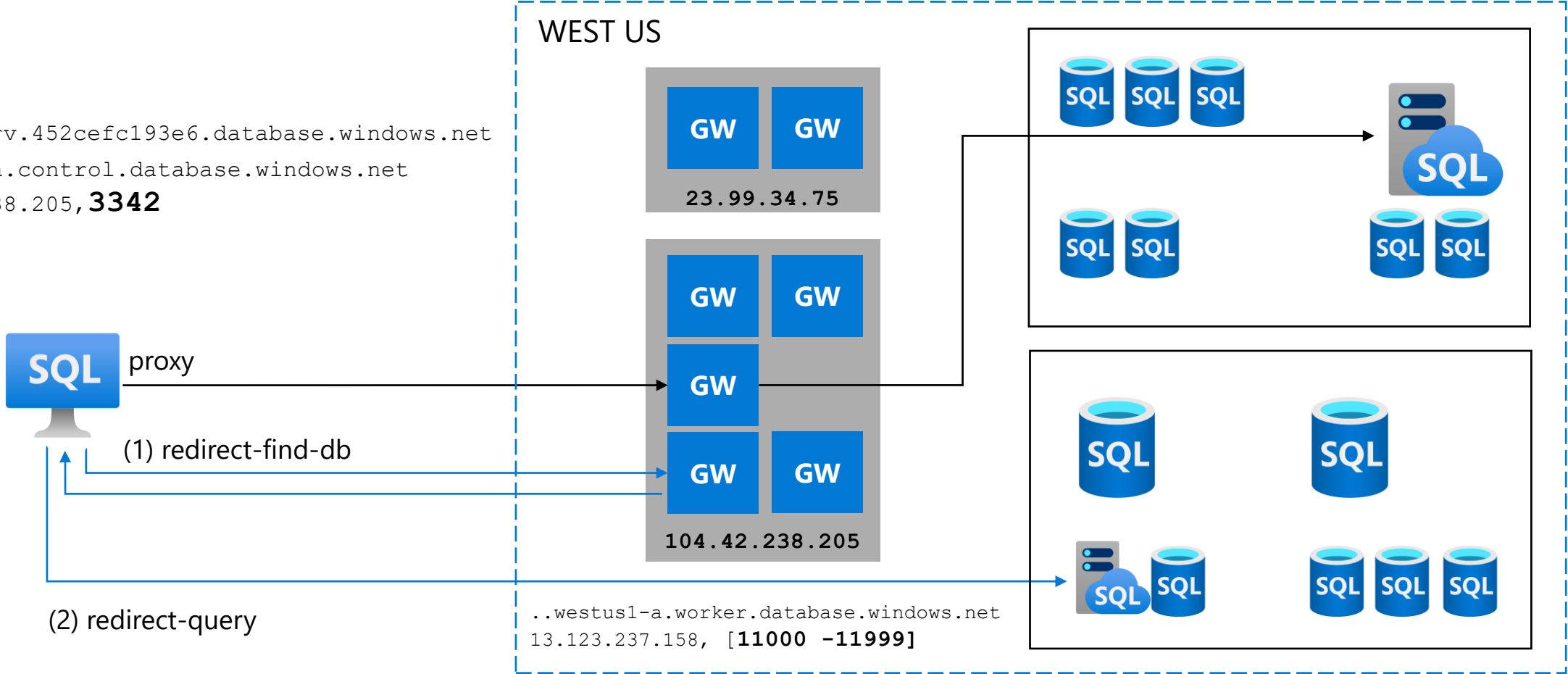
Space management

- Service Level Objective (SLO) determines maximum possible size.
- You choose max storage for instance (up to maximum possible size)
- Databases created as model default size (100Mb/8Mb and configurable)
- You can alter size and # files but not physical location.
- Msg 1105 for database or Msg 1133 for max instance storage
- General Purpose Remote Storage performance can be affected by data/log file size.



SQL MI connectivity architecture

Mysqldbdrv.452cefc193e6.database.windows.net
westus1-a.control.database.windows.net
104.42.238.205, **3342**



Deploy MI

Basics

- Subscription
- Resource group
- MI name
- Region
- Compute and storage
- Service tier
- vCores
- Storage size
- MI admin login
- Password

Network connectivity

- Virtual Network
- Connection type
- Public endpoint
- TLS version

Additional settings

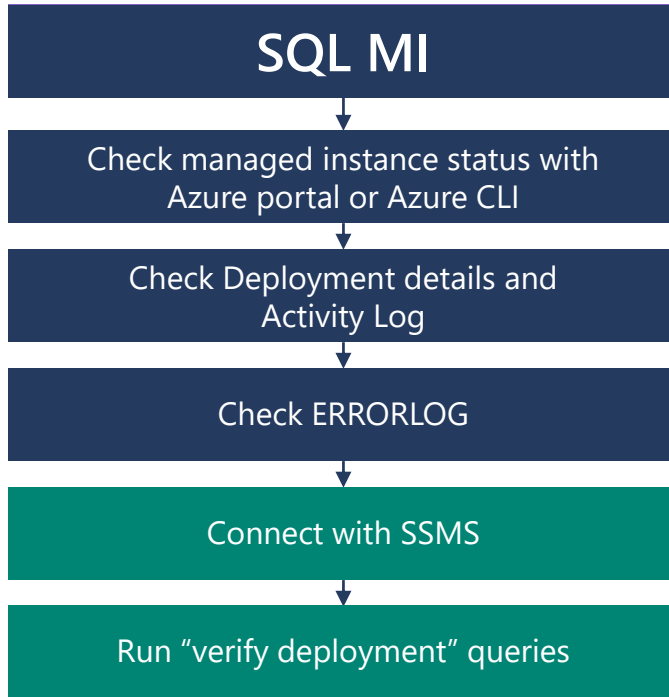
- Server Collation
- Time zone
- Geo-Repl secondary

Tags

Review & create

- Terms and Privacy

Verify Deployment



```
SELECT @@VERSION
SELECT * FROM sys.databases
SELECT * FROM sys.objects
SELECT * FROM sys.dm_os_schedulers
SELECT * FROM sys.dm_os_sys_info
SELECT * FROM sys.dm_os_process_memory
SELECT * FROM sys.dm_exec_requests
SELECT SERVERPROPERTY('EngineEdition')
SELECT * FROM sys.dm_user_db_resource_governance
SELECT * FROM sys.dm_instance_resource_governance
SELECT * FROM sys.dm_os_job_object
```



Demonstration

Deploy Azure SQL MI

- Demonstrate the deployment options




Questions?



Lesson 2: Service Tiers and Business Continuity

Azure SQL Managed Instance Service Tiers

Service tier


Select from the latest vCore service tiers available for Azure SQL Managed Instance including General Purpose and Business Critical. [Learn more](#) 

Service tier ⓘ

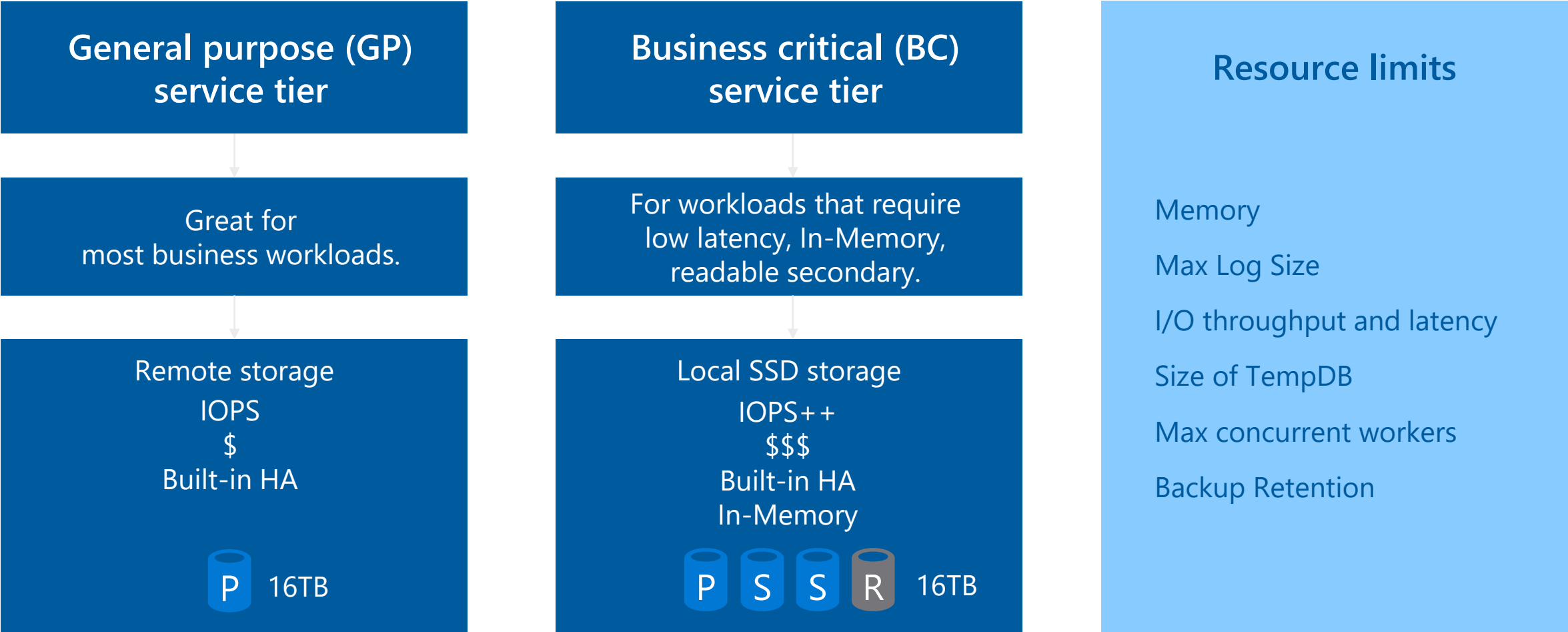
- ☒ General Purpose (4-80 vCores, 32 GB-16 TB storage capacity, Fast storage) - for most production workloads
- ☐ Business Critical (4-80 vCores, 32 GB-4 TB storage capacity, Super fast storage) - for IO-intensive and compute-intensive workloads

Next-gen General Purpose (preview)

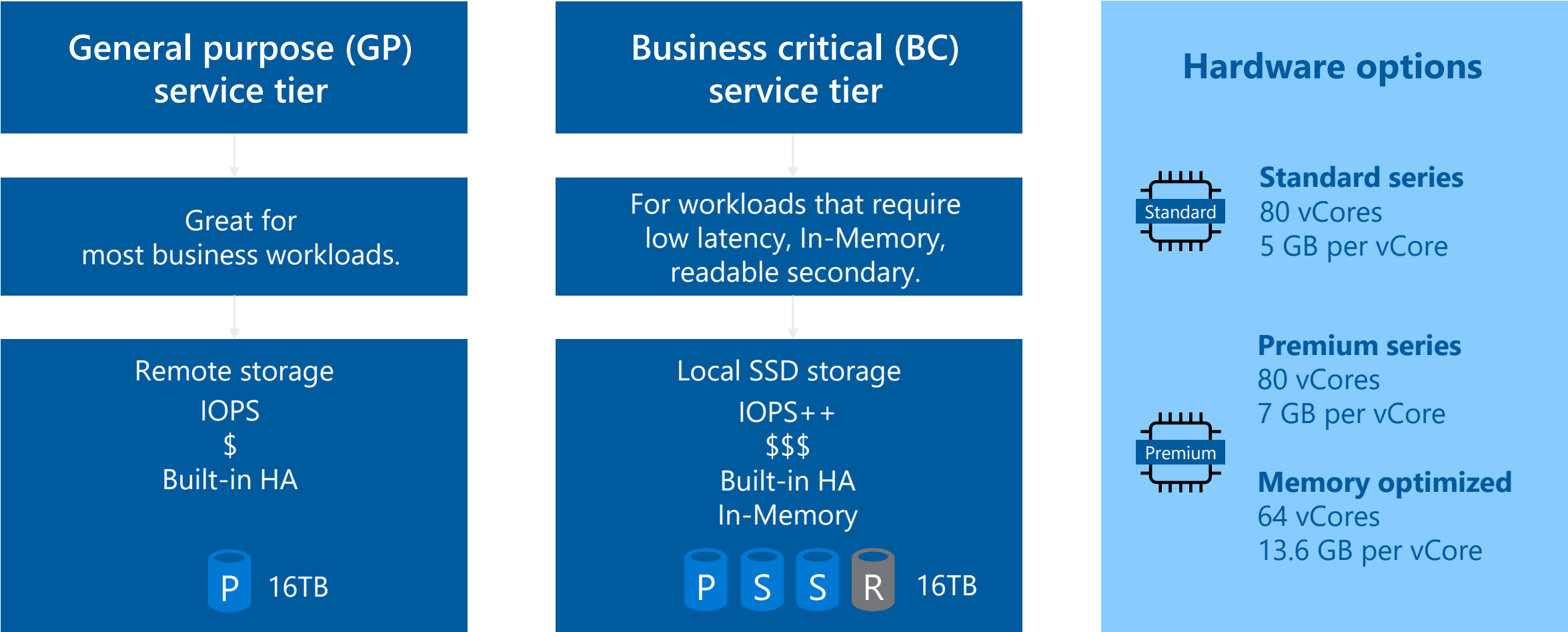
Enabled **Disabled**

i Next-gen General Purpose tier is currently in preview. By using this set of preview features, you confirm that you agree that your use of these features is subject to the preview terms in the agreement under which you obtained Microsoft Azure Services. [Learn more](#) 

Azure SQL Managed Instance Service Tiers




Azure SQL Managed Instance Service Tiers



Next-gen General Purpose (Preview)

Announced March 2024


Select from the latest vCore service tiers available for Azure SQL Managed Instance including General Purpose and Business Critical. [Learn more](#) 

Service tier ⓘ

- ☒ General Purpose (4-80 vCores, 32 GB-16 TB storage capacity, Fast storage) - for most production workloads
- ☐ Business Critical (4-80 vCores, 32 GB-4 TB storage capacity, Super fast storage) - for IO-intensive and compute-intensive workloads

Next-gen General Purpose (preview)

Enabled **Disabled**

i Next-gen General Purpose tier is currently in preview. By using this set of preview features, you confirm that you agree that your use of these features is subject to the preview terms in the agreement under which you obtained Microsoft Azure Services. [Learn more](#) 

Next-gen General Purpose (Preview)

Announced March 2024

Designed for businesses with higher performance requirements while offering the same baseline cost as the General Purpose service tier

Significant upgrades to performance, scalability, and resource flexibility over the General Purpose service tier

Uses managed disks instead of page blobs, which drastically improve storage performance metrics

3 free IOPS for every GB of reserved storage

Support of up to 500 databases per instance, and a max storage size of 32 TB

SQL Managed Instance Hardware Generations

Compute Hardware

Configure compute hardware that will run your Azure SQL Managed Instance. [Learn more](#)

Hardware generation ⓘ

☒ Standard-series (Gen 5) - Intel Broadwell, 5,1 GB RAM/vCore

☐ Premium-series - Intel Ice Lake, 7 GB RAM/vCore, up to 560 GB

☐ Premium-series - memory optimized - Intel Ice Lake, 13,6 GB RAM/vCore, up to 870,4 GB

vCores ⓘ


4 8 16 24 32 40 64 80 8

Storage in GB ⓘ

256 256

Zone redundancy ⓘ

Enabled Disabled

 Zone redundancy is not supported in General Purpose service tier in this region. Please select Business Critical or select another region. [Learn more](#)

✓ All zone redundancy requirements

SQL Managed Instance Hardware Generations

Standard-series (Gen 5)	Premium-series	Premium-series Mem Optimized
<ul style="list-style-type: none">• CPU: Intel Broadwell, Skylake and Cascade Lake, 2.3-2.5 GHz• vCore range: 4 – 80• Memory / vCore: 5.1 GB• Max instance memory: 408 GB• Max instance storage<ul style="list-style-type: none">• General Purpose: 16 TB• Business Critical: 4 TB	<ul style="list-style-type: none">• CPU: Latest 3rd Gen Intel 8370C (Ice Lake), 2.8 GHz• vCore range: 4 – 80• Memory / vCore: 7 GB• Max instance memory: 560 GB• Max instance storage<ul style="list-style-type: none">• General Purpose: 16 TB• Business Critical: 5.5 TB	<ul style="list-style-type: none">• CPU: Latest 3rd Gen Intel 8370C (Ice Lake), 2.8 GHz• vCore range: 4 - 64• Memory / vCore: 13.6 GB• Max instance memory: 870 GB• Max instance storage<ul style="list-style-type: none">• General Purpose: 16 TB• Business Critical: 16 TB

General Purpose High Availability

Behaves like Failover Cluster Instance

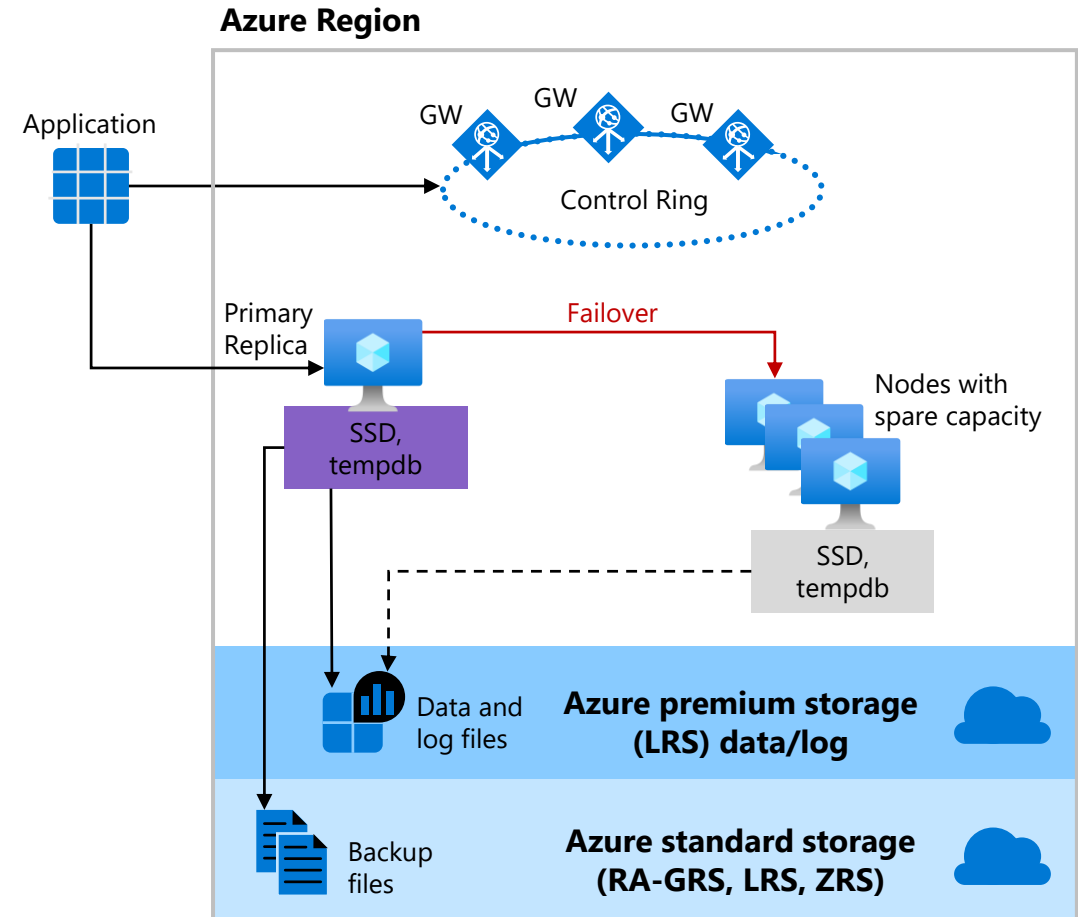
Remote storage provides data redundancy within a datacenter

Backup files are in a different location with geo-redundancy

Failover decisions based on SQL and Service Fabric

Recovery time depends on spare capacity

Connectivity redirection built-in



Business Critical High Availability

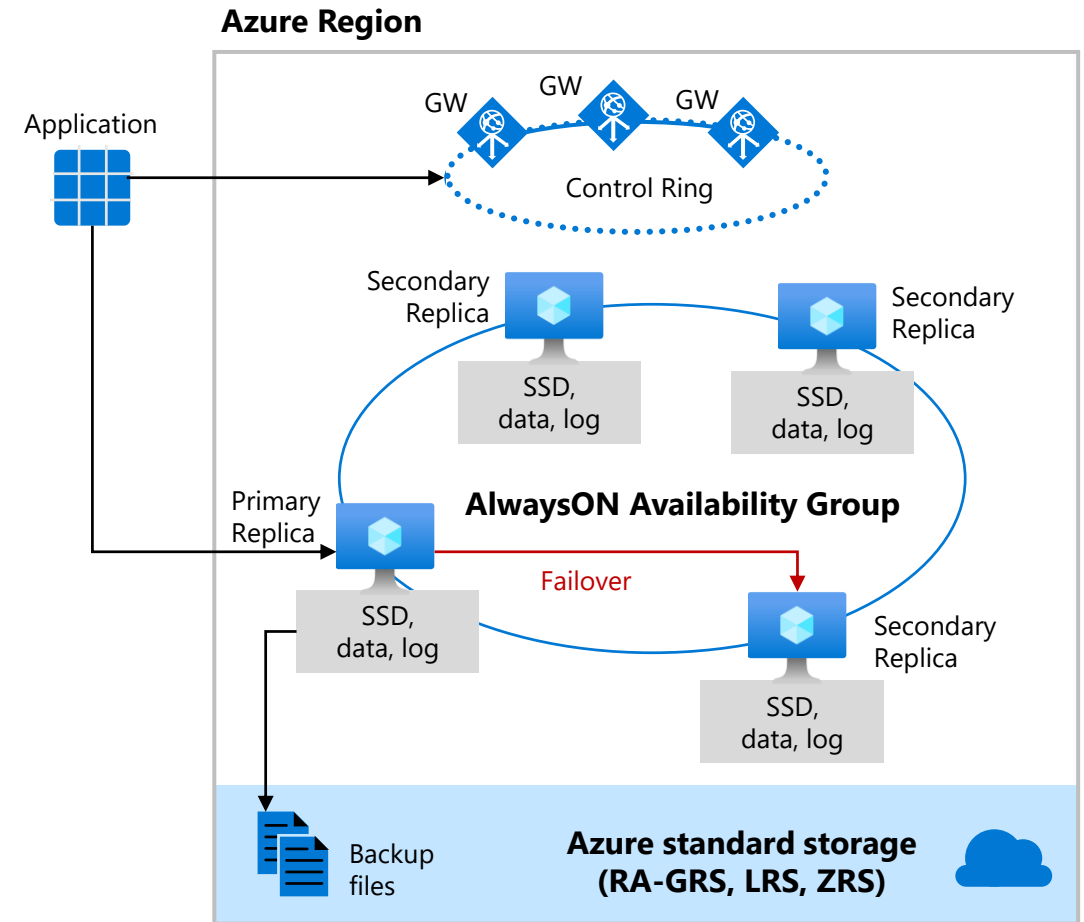
High availability is achieved by replicating both compute and storage to additional nodes.

High availability is implemented using a technology like SQL Server Always On Availability Groups.

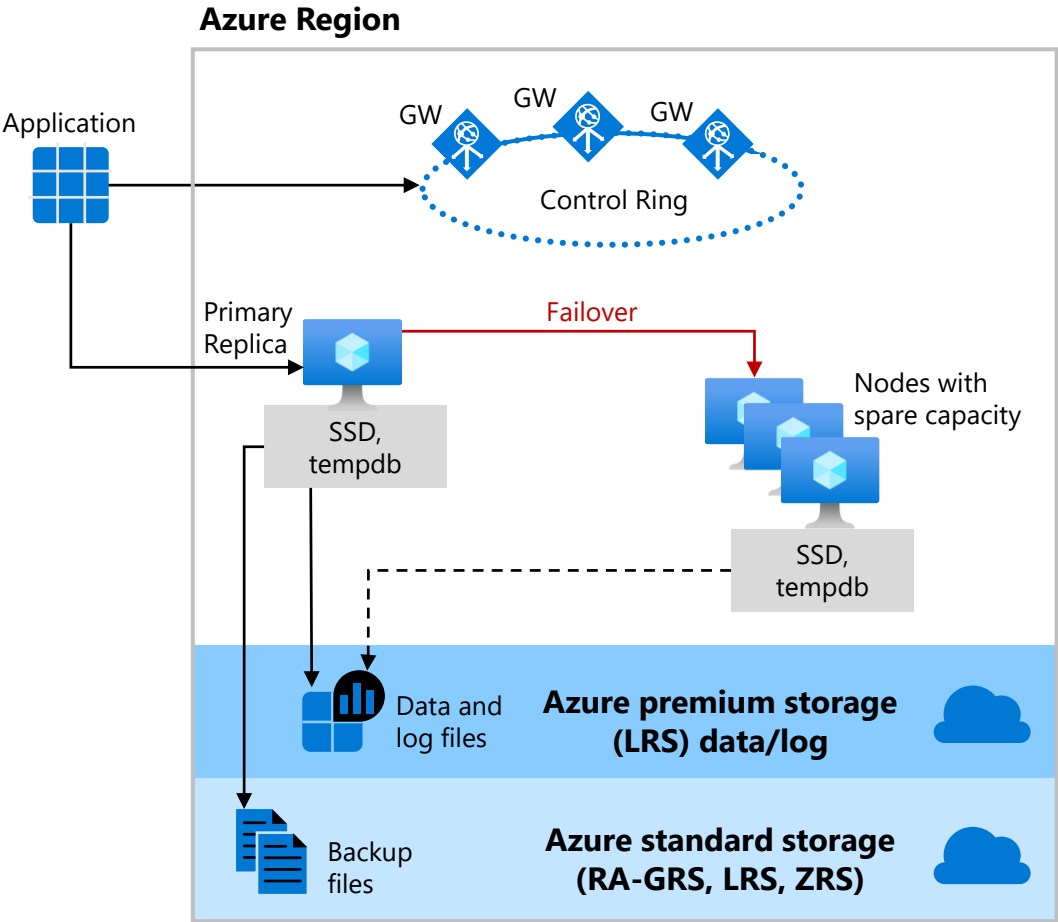
The cluster includes a single primary replica for read-write workloads, and up to three secondary replicas (compute and storage) containing copies of data.

The failover is initiated by the Azure Service Fabric.

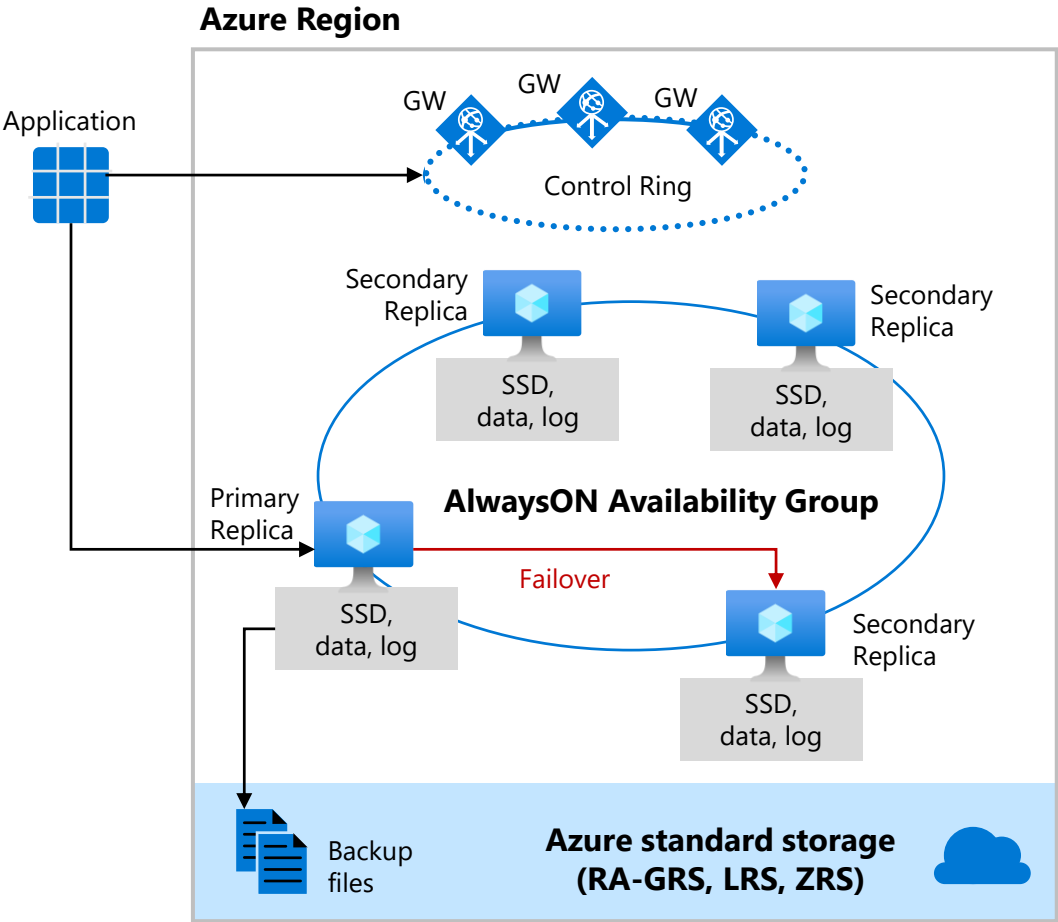
As an extra benefit, the premium availability model includes Read Scale-Out feature.



Side by Side comparison



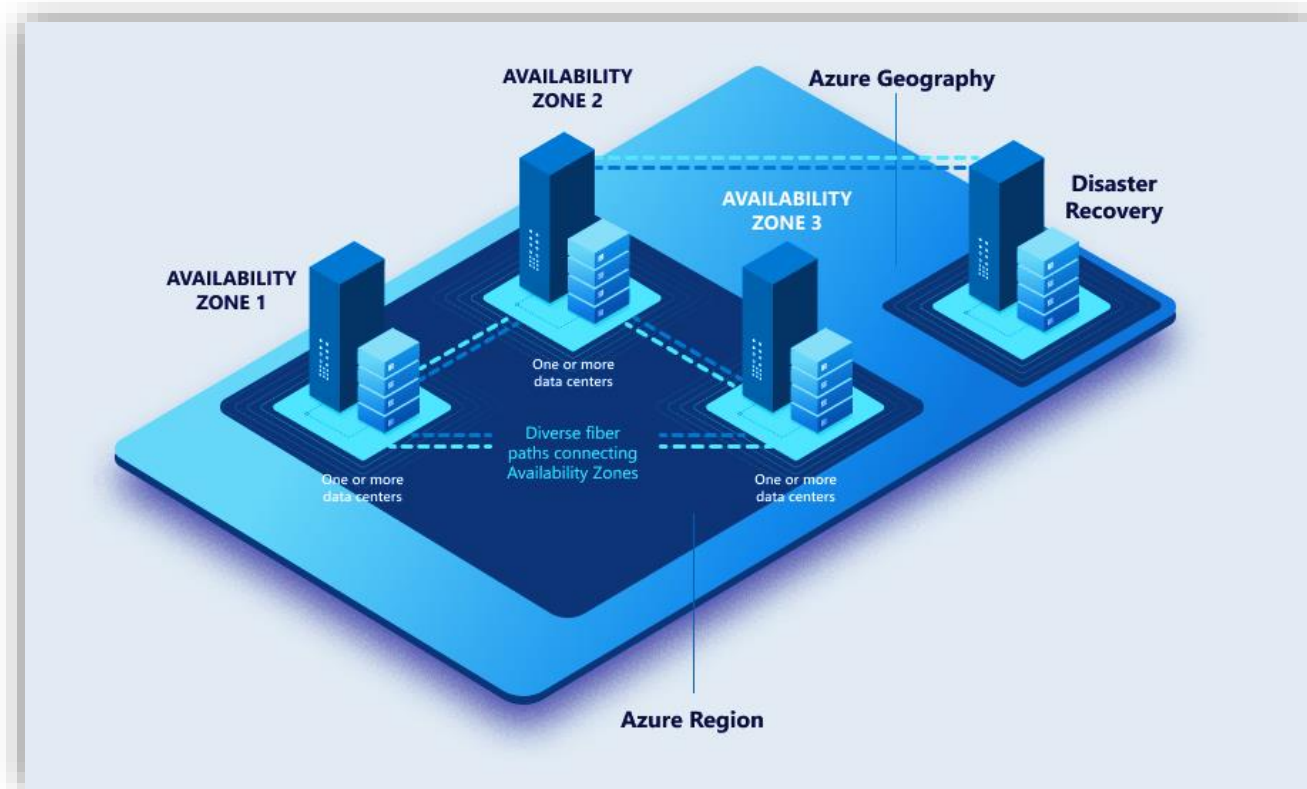
General Purpose (GP) service tier



Business Critical (BC) service tier

Backup storage redundancy

To enable high durability of backups several ways of replication are offered on instance creation.



The backups can be all located within

1. LRS: The same building (Local)
2. ZRS: Same region, different buildings (Zone)
3. GRS: Across paired regions (Geo)
4. GZRS: Different buildings AND paired regions (Geo-Zone)

Zone redundant configuration

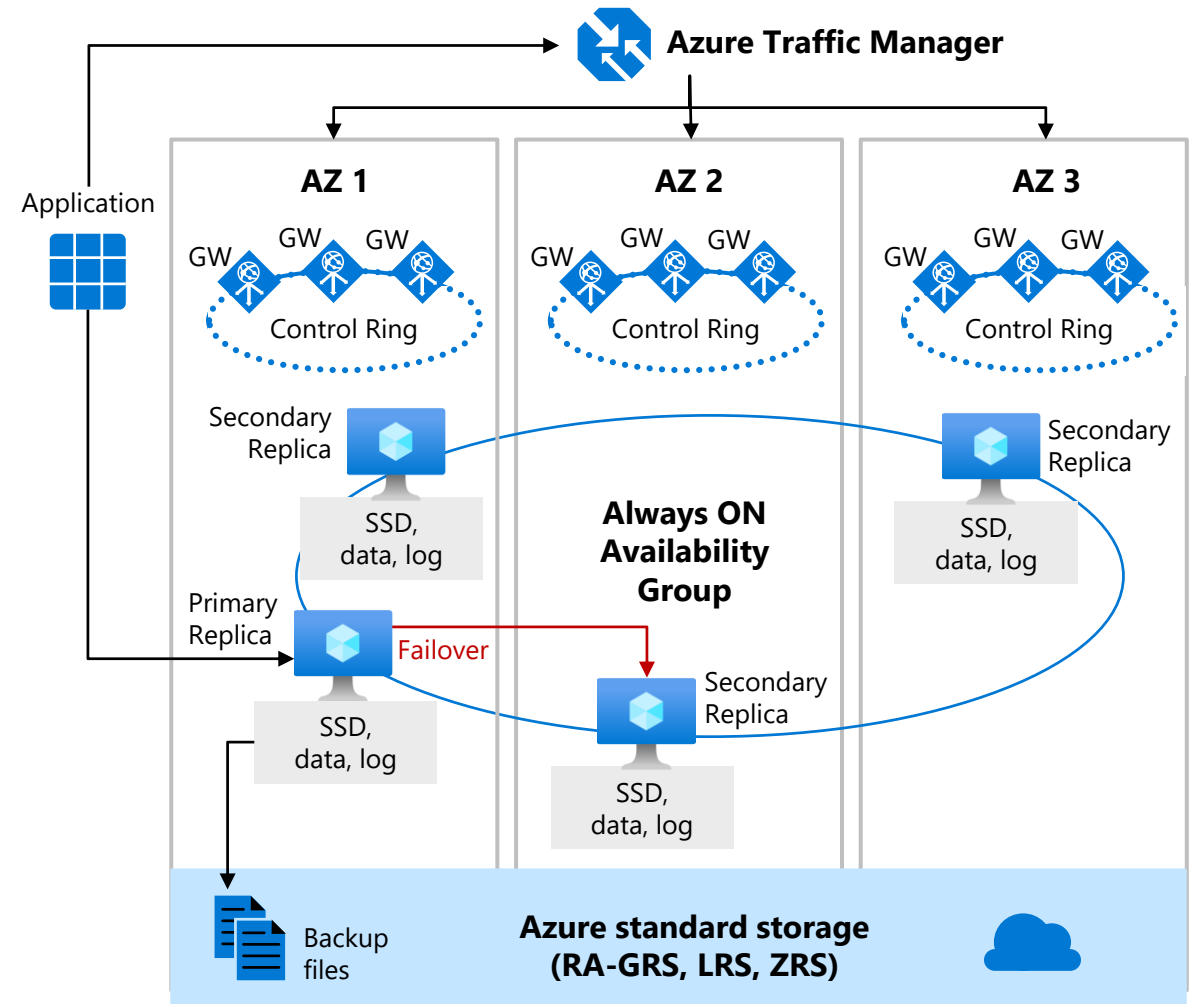
By default, the cluster of nodes for the premium availability model is created in the same datacenter.

SQL Database can place different replicas of the Business-Critical database to different availability zones in the same region.

The routing is controlled by Azure Traffic Manager (ATM).

The zone redundant databases have replicas in different datacenters with some distance between them, the increased network latency may impact the performance.

Zone redundant configurations are currently only supported in the Premium or Business Critical tiers



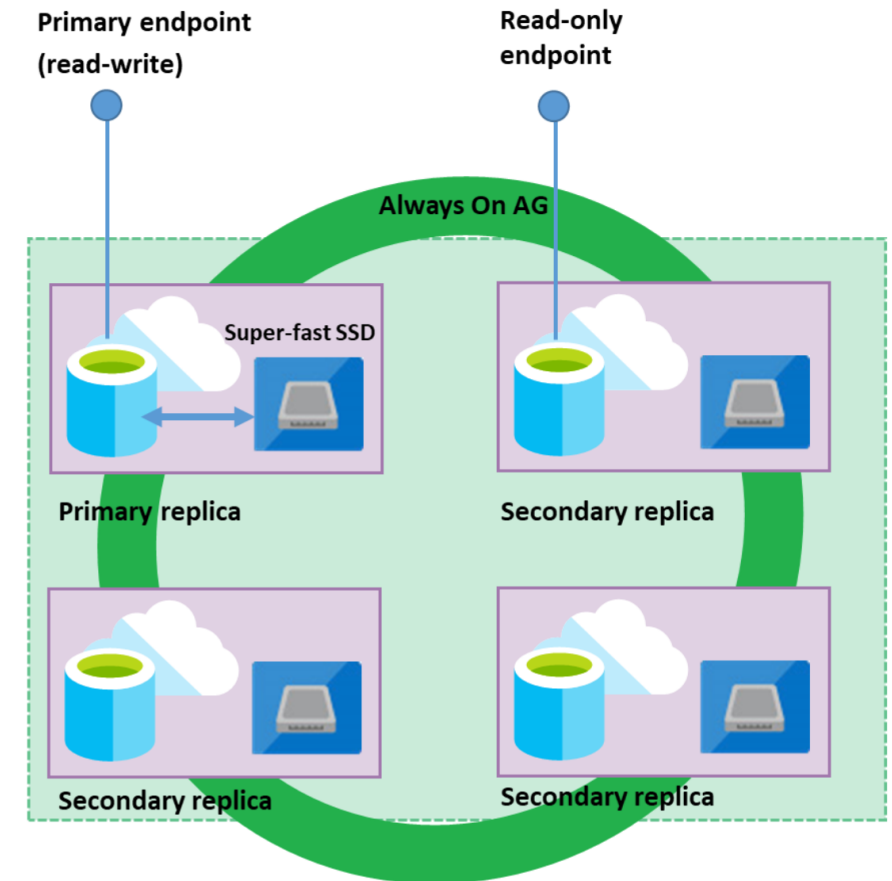
Read Scale-Out

Each database in the Premium and Business Critical service tiers is automatically provisioned with several AlwaysON replicas to support the availability SLA. These replicas are provisioned with the same compute size as the read-write.

Read Scale-Out redirects the read-only client connections to one of the ready-only replicas available instead of sharing the read-write replica.

Effectively isolate the read-only workload from the main read-write workload and doubles the compute capacity of the database or elastic pool at no additional charge.

This is ideal to scale-out for complex analytical workloads without affecting the primary OLTP workload.



Service Level Agreement (SLA)

Service tier	Single zone SLA	Multiple zones SLA
Basic, Standard, General Purpose	99.99%	N/A
Premium, Business critical	99.99%	99.995%

Business continuity	Service tier	SLA
Recovery point objective (RPO)	Business critical with Geo-DR	5 sec
Recovery Time Objective (RTO)	Business critical with Geo-DR	30 sec

[SLA for Azure SQL Database](#)

[SLA for Azure SQL Managed Instance](#)

Active geo-replication vs auto-failover groups

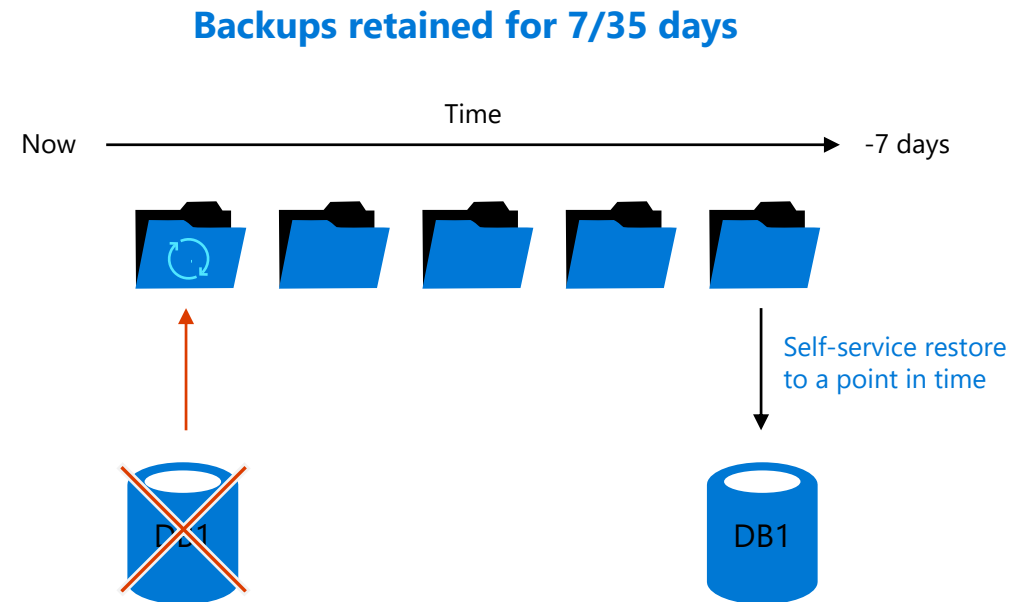
	Geo-replication (Database)	Auto-failover groups (Server)
Automatic failover	No	Yes
Fail over multiple databases simultaneously	No	Yes
Update connection string after failover	Yes	No
Managed instance supported	No	Yes
Can be in same region as primary	Yes	No
Multiple replicas	Yes	No
Supports read-scale	Yes	Yes

*Geo-Replication is not currently offered for Managed Instances

Backup and restore

Auto backups and Point in Time Restore (PITR)

- Full Database backup once a week
- Differential Backups every 12-24 hours
- Log Backups every 5-10 minutes
- Backup files on Azure storage with RA-GRS replicated
 - Can optionally select LRS or ZRS
- Backup Integrity checks
- Restore to new database
- Long-term retention (up to 10 years) of backups
- Geo-restore of databases if primary region down
- Restore backups of deleted databases



[Automatic, geo-redundant backups - Azure SQL Managed Instance | Microsoft Learn](#)

Setting Backup Policies

Home > Resource groups > jdSQLRG > jdsqlazure

jdsqlazure | Backups ☆ ...

SQL server

Search

Data management

- Backups
- Deleted databases
- Failover groups
- Import/Export history

Security

- Networking
- Microsoft Defender for Cloud
- Transparent Data Encryption
- Identity
- Auditing

Intelligent Performance

Available backups **Retention policies**

Configure and manage your automated backup retention policies. Long-term retention policies enable you to keep full backups for up to 10 years.

Search for a database

Databases in the Basic tier are limited to a 7 day retention policy.

Database	PITR	Differential backup frequency
jdsqldb	7 Days	24 Hours

Configure policies

SQL server

Point-in-time-restore

Specify how long you want to keep your point-in-time backups. [Learn more](#)

How many days would you like PITR backups to be kept? 7

Differential backup frequency

Specify how often you want differential backups to be taken. [Learn more](#)

Take a differential backup every:

24 Hours

Long-term retention

Specify how long you want to keep your long-term retention backups. You may choose to keep yearly backups for up to 10 years. [Learn more](#)

Weekly LTR Backups

Keep weekly backups for:

6 Week(s)

Database availability and consistency

Availability

- You cannot set OFFLINE and EMERGENCY
- RESTRICTED_USER access allowed
- Dedicated Admin Connection (DAC) allowed
- Accelerated Database Recovery on by default

Consistency

- Multiple copies of data and backups
- Users can execute DBCC CHECKDB (no repair)
- Database CHECKSUM on by default
- Auto Page Repair when possible
- Data integrity error alert monitoring
- Backup and restore integrity checks
- "lost write" and "stale read" detection
- Repair without notification if no impact
- Proactive notification to customers

Accelerated Database Recovery (ADR)



- How it works
 - Uses a Persisted Version Store (PVS)
 - Independent of locking and isolation levels
 - Rollback faster than you can react
 - Undo recovery faster than you can look it up
 - Transaction log truncation not tied to active transactions



- Key questions
 - Does it require more space?
 - Will it affect performance?
 - Will I still see versions in tempdb?
 - How does it work with HA?



- Read the [paper](#)
- John's [video](#)

Lesson 3: Managed Instances Best Practices

Objectives

After completing this learning, you will be able to:

- Learn best practices for Azure SQL MI
- Understand how to use BPCheck tool to assess Azure SQL MI
- Understand how to use SQL Assessment API tool to assess Azure SQL MI



Performance Best Practices for Azure SQL MI

TempDB

- Always kept on local SSD drive
- For DB, # files scales with vCores
- You get 12 files with MI
- MIXED_PAGE_ALLOCATION IS OFF
- AUTOGROW_ALL_FILES is ON
- TempDB Metadata Optimization not yet supported

Database

- Only full recovery mode is supported
- Minimal logging for bulk operations not possible

Files and Filegroups

- MI supports adding files and sizes but not physical placement
- # files and file size can be used to tune I/O performance

AdHoc Workload

- Optimize for AdHoc workload is allowed

MAXDOP

- Database level
- Instance level
- Query level
- Workload Group level in Resource Governor

Resource Governor

- Achieve predictable performance by defining
 - Resource pools
 - Workload groups
 - Classifier function

BPCheck Tool

TSQL utility to check best practices

Compatible with SQL Server versions 2008 and later, Azure VM and Managed instance

Does not support Azure SQL Database

Reports on Azure SQL MI configuration, system resources, and various statistics

Requires the `@allow_xpcmdshell = 0` since `xp_cmdshell` is not supported on Managed Instance

Demonstration

Running BPCheck

- Create report by running BPCheck against Azure SQL MI



Using SQL Assessment API in Azure Data Studio

Provides user interface for evaluating Azure SQL MI and databases for best practices

PowerShell cmdlet: Invoke-SqlAssessment

Install the SQL Server Assessment extension from the Azure Marketplace.

Results can be exported to a table.

Demonstration

Use Azure Data Studio to invoke assessment:

- How to invoke the SQL Assessment API against Azure SQL MI using SQL Assessment Extension



POC

Assess Azure SQL MI with Azure Data Studio Extension

- Install SQL Assessment extension in Azure Data Studio and run the assessment to analyze Azure SQL MI

Questions?



Lesson 4: Managed Instance Performance Concepts

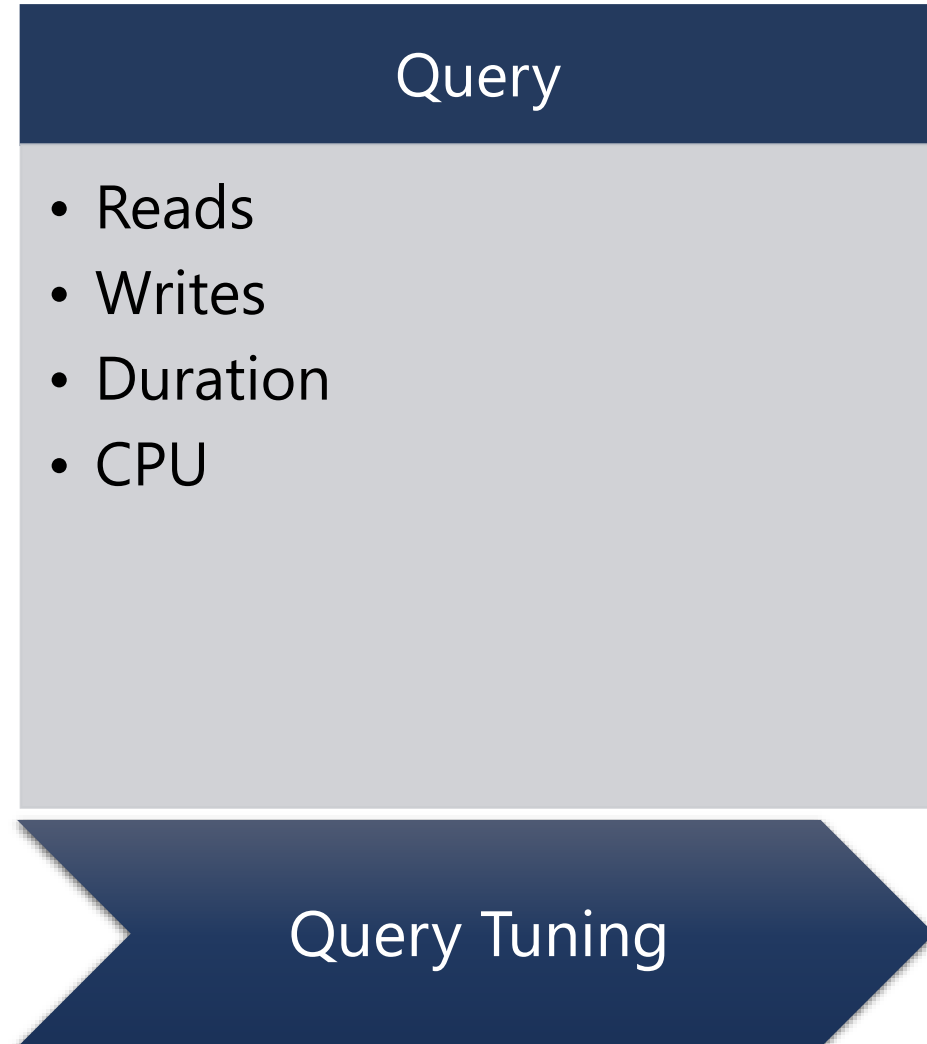
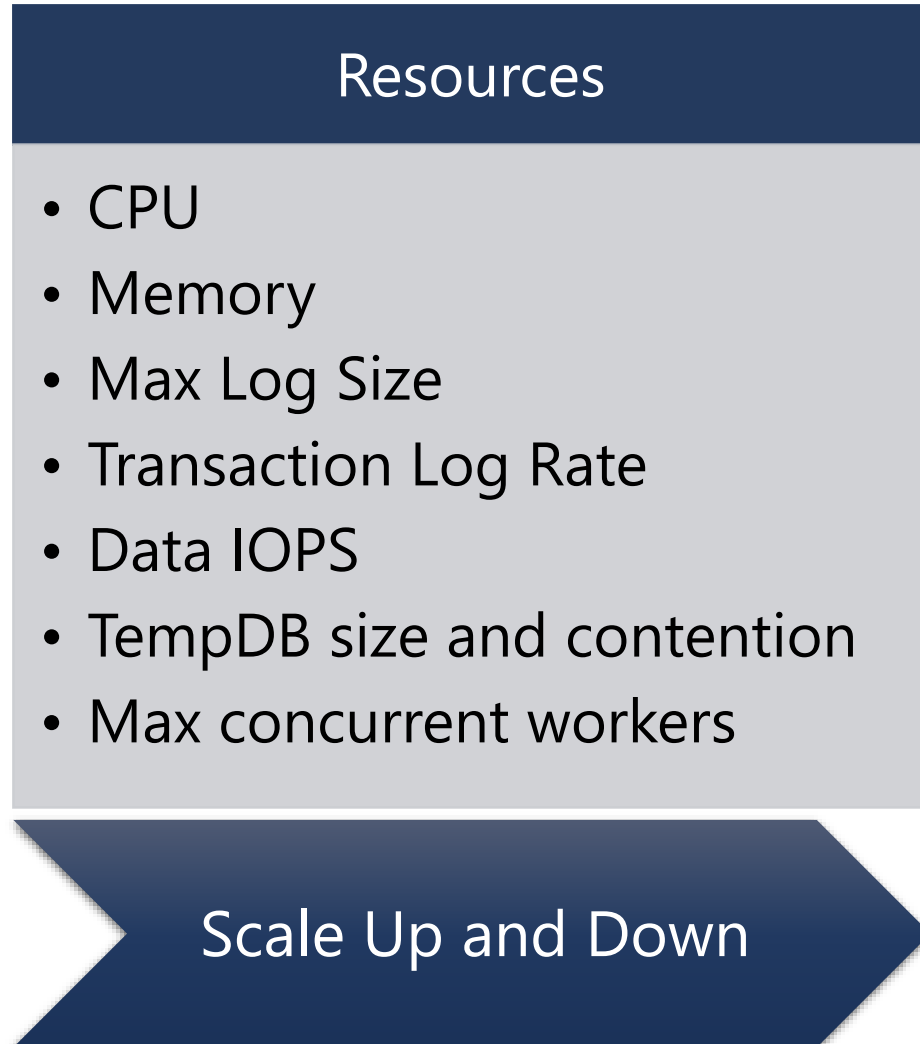
Objectives

After completing this learning, you will be able to:


- Learn the areas that can cause performance issues
- Understand query status and wait statistics



Resource limit considerations



Accelerating and Tuning Performance

Scaling CPU Capacity	I/O performance	Increasing Memory or Workers	Improving Application Latency
<ul style="list-style-type: none">• Portal, T-SQL, PowerShell, or REST• Downtime (can be very fast) but scaling can take a few hours	<ul style="list-style-type: none">• IOPS restrictions may show up in the form of longer I/O waits• Scale up vCores or move to Business Critical or change # files or file size• Consider multi-statement transactions for log I/O latency 	<ul style="list-style-type: none">• Scale up vCores for more memory• Max worker threads	<ul style="list-style-type: none">• Use redirect connections and not proxy• Optimize “chatty” applications

Disks Type	P10	P15	P20	P30	P40	P50
Max size	128 GB	256 GB	512 GB	1 TB	2 TB	4 TB
IOPS	500	1100	2300	5000	7500	7500
Throughput (MB/second)	100	125	150	200	250	250

Tune your SQL queries just like you would with SQL Server

- Indexes
- Batching
- Stored Procs
- Parameterization
- Process Result Sets

Baselining

- Create baseline
 - Create baseline for individual workloads
- Update baseline
 - When changing service level or workload
 - After any significant change in schema or code

SET STATISTICS IO, TIME

```
SET STATISTICS IO ON
GO
SET STATISTICS TIME ON
SELECT SOH.SalesOrderID, SOH.CustomerID,
OrderQty, UnitPrice, P.Name
FROM Sales.SalesOrderHeader AS SOH
JOIN Sales.SalesOrderDetail AS SOD
ON SOH.SalesOrderID = SOD.SalesOrderID
JOIN Production.Product AS P
ON P.ProductID = SOD.ProductID
SET STATISTICS IO, TIME OFF
```

Used to identify physical reads and logical reads for a query

```
(121317 rows affected)
Table 'Workfile'. Scan count 0, logical reads 0, physical reads 0, page server r
Table 'Worktable'. Scan count 0, logical reads 0, physical reads 0, page server
Table 'SalesOrderDetail'. Scan count 1, logical reads 428, physical reads 0, pag
Table 'Product'. Scan count 1, logical reads 15, physical reads 0, page server r
Table 'SalesOrderHeader'. Scan count 1, logical reads 57, physical reads 0, page

SQL Server Execution Times:
    CPU time = 94 ms,  elapsed time = 1653 ms.
```

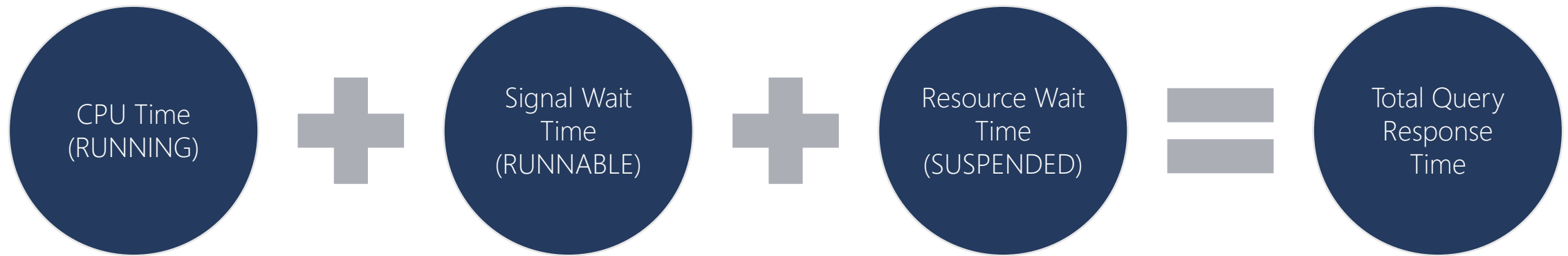
Demonstration

SET STATISTICS



Task Execution Model

- The full cycle between the several task states, for how many times it needs to cycle, is what we experience as the total query response time.



Task Execution Model

Status: Running

session_id 51	Running
---------------	---------



Runnable Queue (Signal Waits)
Status: Runnable

session_id 51	Runnable
session_id 64	Runnable
session_id 87	Runnable
session_id 52	Runnable
session_id 56	Runnable

Wait Queue (Resource Waits)
Status: Suspended

session_id 73	LCK_M_S
session_id 59	NETWORKIO
session_id 56	Runnable
session_id 55	RESOURCE_SEMAPHORE
session_id 60	IO_Completion

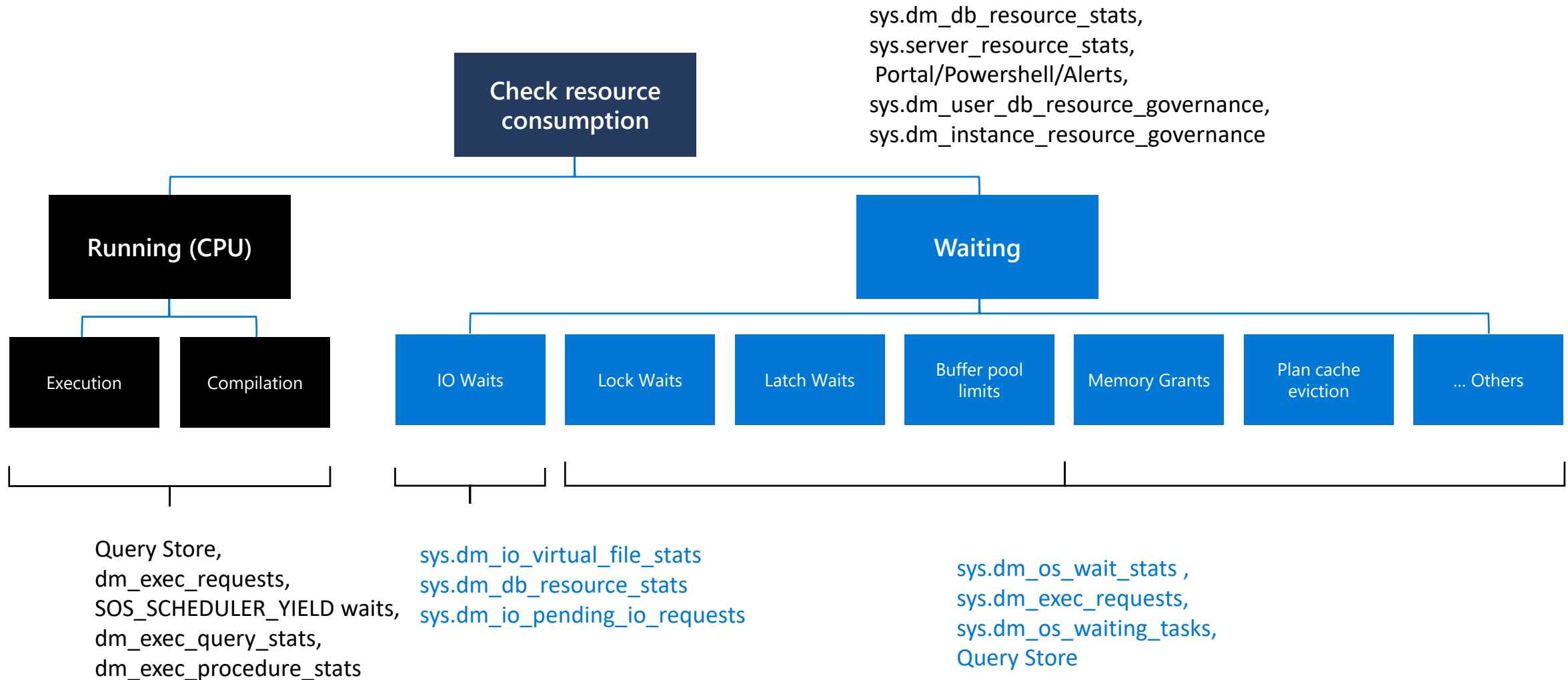


SPID56 moved to the bottom of the Runnable queue.

Waits and Queues

- Troubleshooting methodology to identify resource bottlenecks.
- Azure SQL MI maintains the state in which workers wait.
- Waits are represented by the wait statistics.
- Queues measure system resources and utilization.

It is just SQL: Running or waiting



Azure SQL MI Specific WAITS*

Log governance

- INSTANCE_LOG_GOVERNOR or PAGEIOLATCH wait for Azure SQL Managed Instance General Purpose
- HADR_THROTTLE_LOG_RATE wait for Business Critical and Geo-Replication/Failover Groups latencies

Worker limits

- SQL Server worker threads and THREADPOOL waits

Business critical HADR waits

- HADR_SYNC_COMMIT
- HADR_DATABASE_FLOW_CONTROL
- HADR_THROTTLE_LOG_RATE_SEND_RECV

* Review [Azure SQL Managed Instance resource limits](#)

Demonstration

Monitoring Waits



