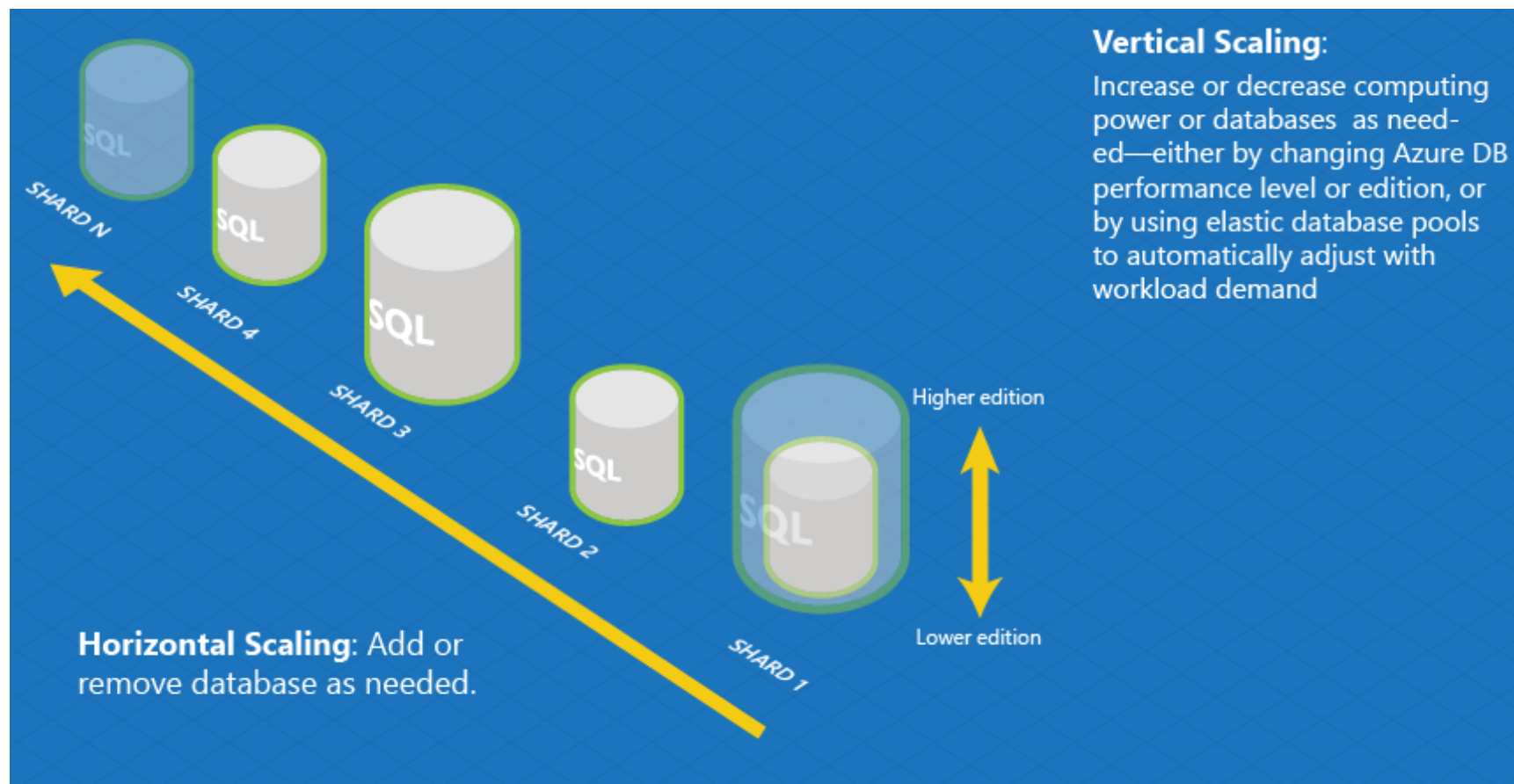# Azure SQL DB

Scaling

Prerequisite: Purchasing model and service tier

Azure SQL Database Scaling

Azure SQL Database supports two types of scaling:

- **Vertical scaling:** Scale up or down the database by adding more compute power.

- **Horizontal scaling:** Scale out or Add more databases and to shard your data into multiple database nodes.

Vertical vs Horizontal Scaling

Azure SQL Database supports two types of scaling:

- **Vertical scaling:**

  - Scale up or down the database by adding more compute power.

  - CPU Power, Memory, IO throughput, and storage

  - DTU and vCore models to scale

  - Dynamic Scalability (Note: this is not auto-scale)

  - Any change that you made will be almost instant .

- **Horizontal scaling:** Scale out or Add more databases and to shard your data

  into multiple database nodes.

Azure SQL Database Scaling

Read Scale Out

- Allows you to use the capacity of the read-only replicas for read-only queries.

- Feature is intended for the applications that include logically separated read-only workloads, such as analytics

- <u>Benefit</u>: When Secondary nodes handles heavy reports and analytical queries, primary writable node saved resources that might be used to improve the performance

- Secondary nodes is asynchronous

- Premium (DTU-based model) or in the Business Critical (vCore-based model)

- Also available in Hyperscale with secondary replica creation

Global Scale-out/Sharding

- Split your data into multiple database nodes.

- Every database shard is an independent database where you can add or remove resources as needed.

- Application may access only the shard associated to that region without affecting other shards.

- Why?

  - Data or transaction throughput exceed the capabilities of individual database

  - Tenants may require physical isolation

  - Different sections of a database may need to reside in different geographies for compliance, performance, or geopolitical reasons.
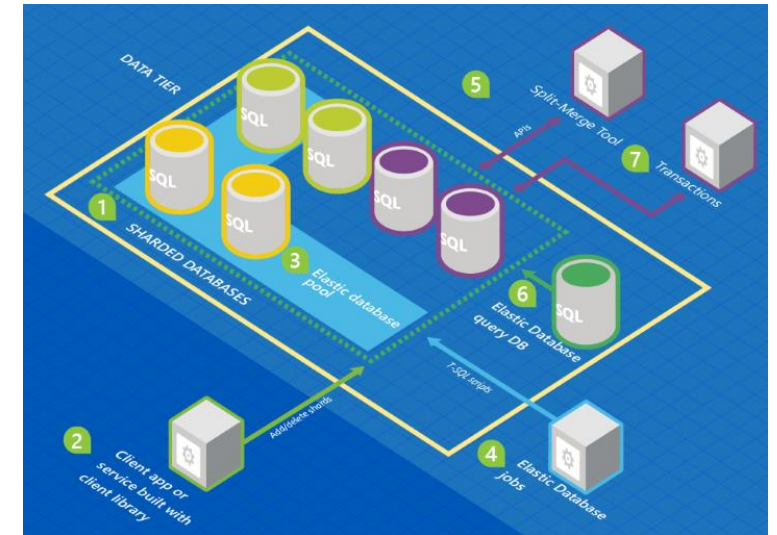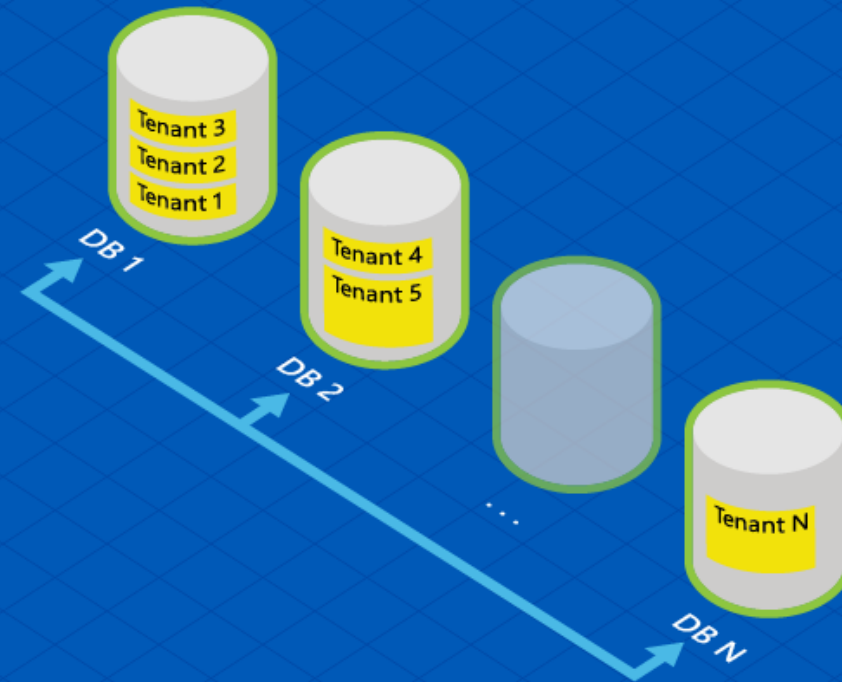
Single Tenancy vs Multi Tenancy

Azure SQL Database Scaling

Change Service Tier

- From Standard/General Purpose to Premium/Business Critical.

- In Standard/General Purpose - Data stored on Azure premium disks

- In Premium/Business Critical - Data stored on local SSD

Azure SQL DW | Scaling

Azure SQL Warehouse Scaling

Scaling in SQL Data Warehouse

- SQL DW allows us to scale or pause at will

- Scale up during heavy demand

- Pause to cut cost

- DWUs are CPU, memory, and I/O bundled into units of compute scale

- Increasing DWU's

    - increase query performance

    - Also increase maximum number of concurrent queries and concurrent sl
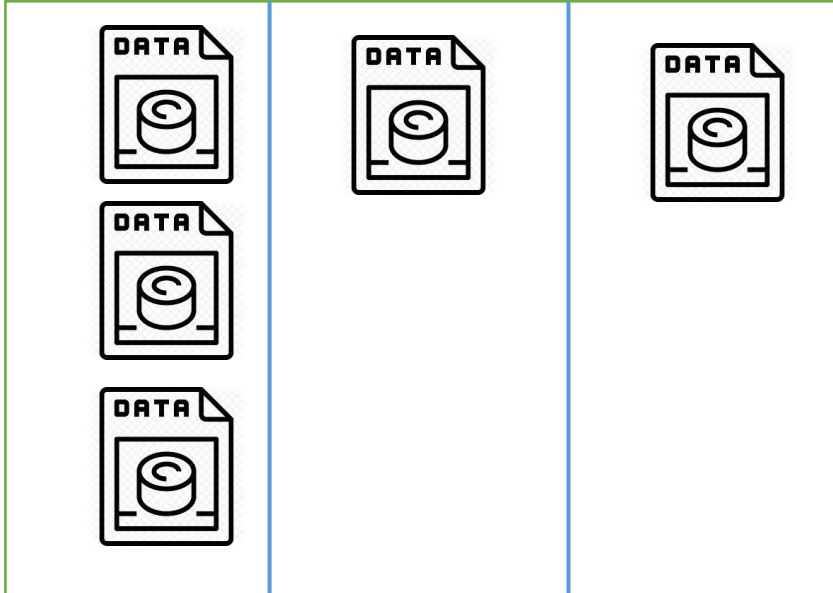
- Modify with GUI, PowerShell, or TSQL

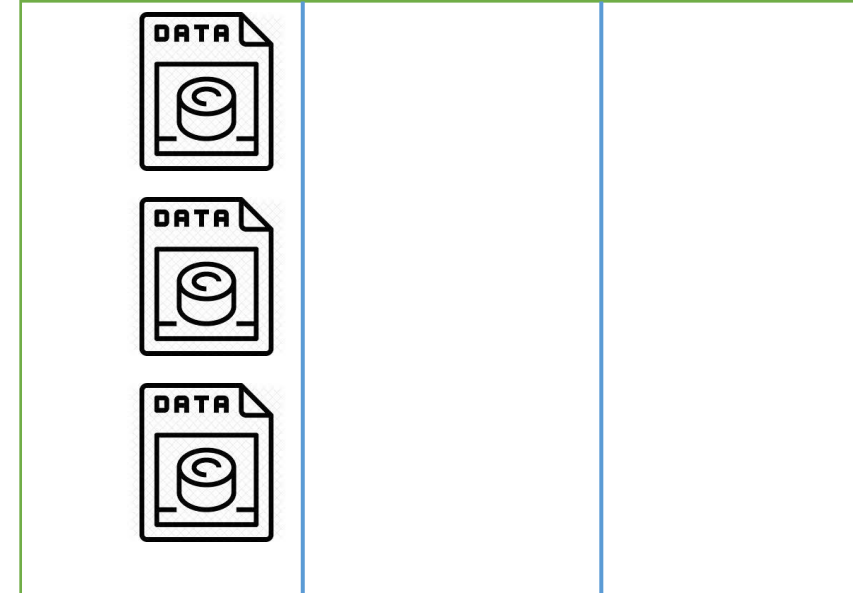# Azure SQL DB – HA and DR Options

Region A

Region B (Read)

Storage Clusters
Each cluster is physically separate in what's called an availability zone, with its own separate utilities and networking.
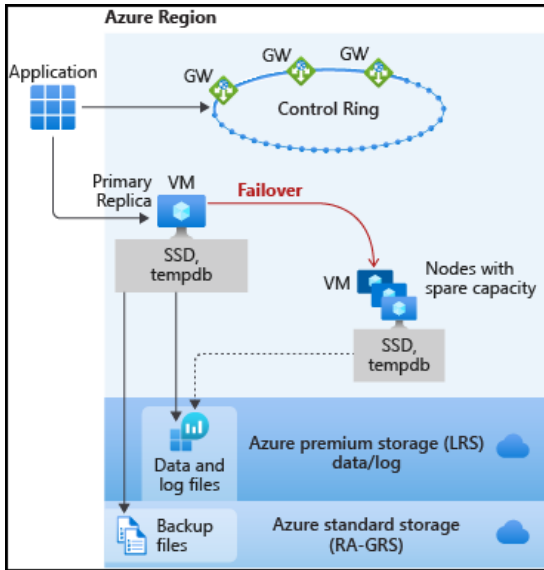
Region B
Hundreds of miles away from the primary region to prevent data loss in the event of a natural disaster.

Azure SQL DB – High Availability Architecture

**Standard availability model**

- Separation of compute and storage

- Performance degradation during maintenance activities

- Budget-oriented business applications

- Basic, Standard, and General Purpose service tier availability
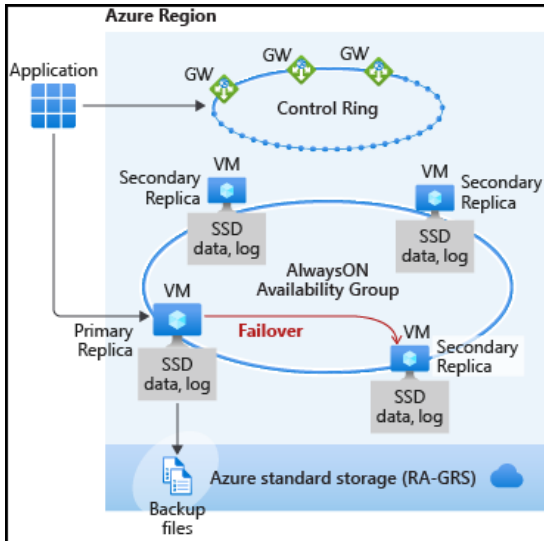
Premium availability model

- Integrates compute resources and storage (locally attached SSD) on a single node

- Replicating both compute and storage to additional nodes creating a three to four-node cluster.

- Data is synchronized to at least one secondary replica before committing each transaction.

- Targets mission critical applications with high IO performance and high transaction rate

- Guarantees minimal performance impact during maintenance activities.

- Premium and Business Critical service tier availability

- Read Scale-Out feature

Azure SQL DB –Zone Redundancy
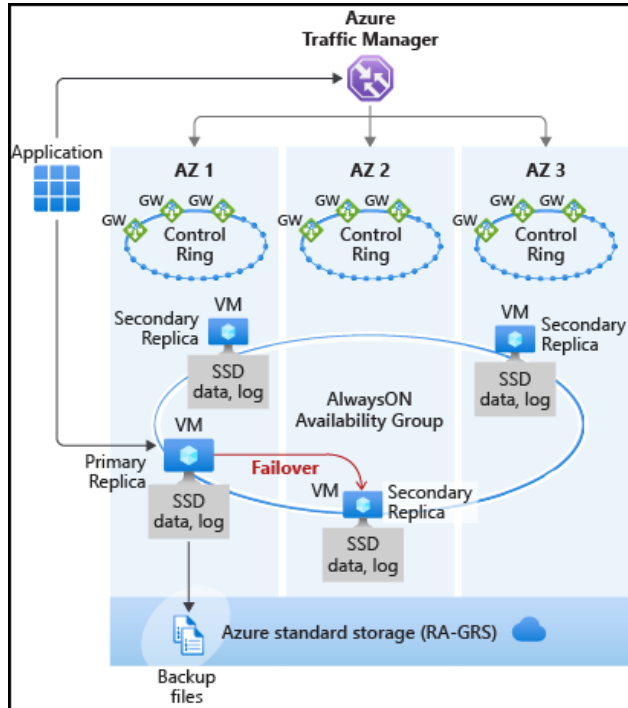
Zone redundant configuration

- Place different replicas of the Business Critical database to different availability zones in the same region

- Does not create additional database redundancy

- Enable it at no extra cost

- Supported in the Premium and Business Critical service tiers

- Not available in SQL Managed Instance.

- Increased network latency may increase the commit time and thus impact the performance of some OLTP workloads.

Azure SQL DB – Geo Replication
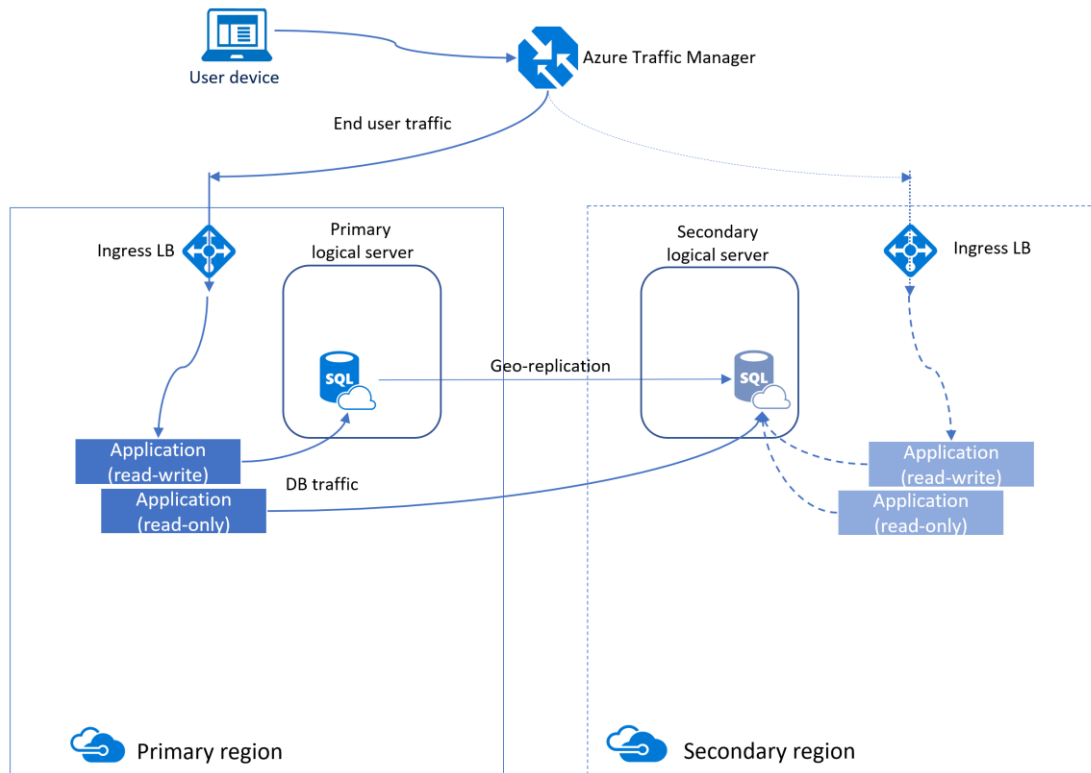
- **Create a readable secondary database in the same region or cross-region**
- **Use cases:**
  - Can failover to the secondary database in case of an outage
  - Migrate a database from one server to another server in the same or cross region with minimal downtime.
- We can create up to four secondaries for each primary database.
- **Data Loss:**
  - Uses the Always-on feature to replicate committed transactions to the secondary database asynchronously.
  - May lag the primary database at any point in time.
- **Manual - Forced Failover**
  - This will make your secondary database immediately online and start accepting connections. Forced failover may result in data loss.

- **Issues**:
  - Supports only manual failover
  - End-point connection must be changed in the application after the failover
  - Must have the same firewall rules and the logins to run applications successfully without any discrepancies

Compare geo-replication with failover groups
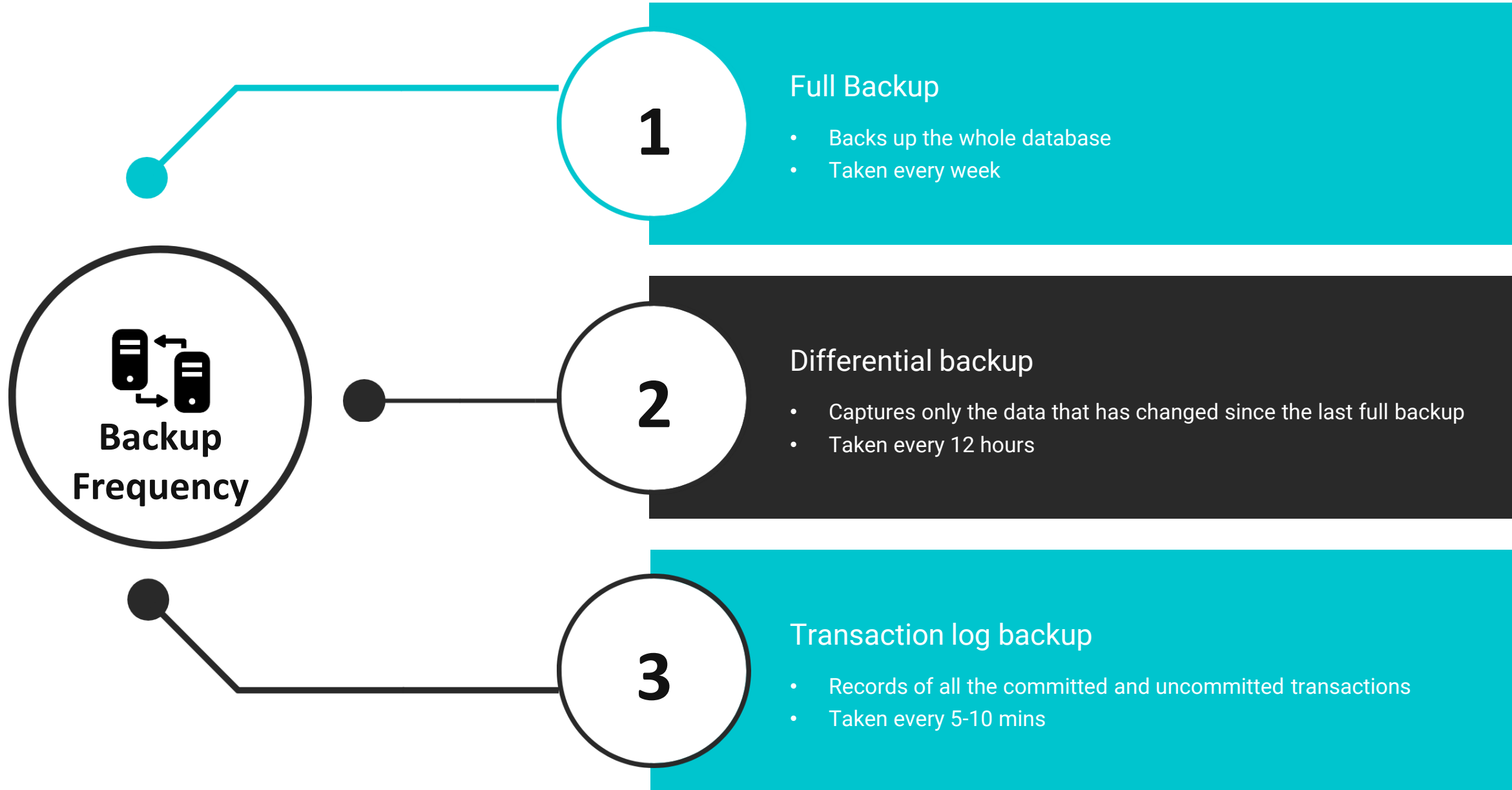
Auto-failover groups simplify the deployment and usage of geo-replication and add the additional capabilities as described in the following table:

	Geo-replication	Failover groups
Automatic failover	No	Yes
Fail over multiple databases simultaneously	No	Yes
User must update connection string after failover	Yes	No
SQL Managed Instance support	No	Yes
Can be in same region as primary	Yes	No
Multiple replicas	Yes	No
Supports read-scale	Yes	Yes

Backup Frequency

1 — Full Backup
- Backs up the whole database
- Taken every week

2 — Differential backup
- Captures only the data that has changed since the last full backup
- Taken every 12 hours

3 — Transaction log backup
- Records of all the committed and uncommitted transactions
- Taken every 5-10 mins

Storage cost and Security of Backup files

Storage Cost

Backup files are copied to RA-GRS standard blob storage by default to paired region

Security

Backup are automatically encrypted at rest using TDE, blob storage is also protected

Backup Retention Period

Backup storage redundancy

- Point in time restore (PITR) - 7-35 days

- Long-term retention - Up to 10 years

Long term retention (LTR)

- One or more long term retention periods to your database to meet regulatory, compliance or other business purposes

- Full backups can be taken up to 10 years

- Stored in RA-GRS blob storage

- Any change of the LTR policy applies to the future backups

LTR and Managed Instance

- LTR is not yet available for databases in Managed Instances

- You can use SQL Agent jobs to schedule copy only database backups as an alternative to LTR beyond 35 days

- These backups can be kept in the Azure blob storage
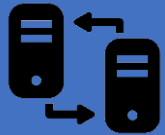
Backup Restore

Backup usage

- Point-in-time restore of existing database

- Point-in-time restore of deleted database

- Geo-restore

- Restore from long-term backup

- *If you delete an Azure Logical SQL Server, all elastic pools and databases that belong to that logical server are also deleted and cannot be restored*

Restore Time is impacted By

- Size of database & Compute size of the database

- Number of transaction logs and Amount of activity

- Network bandwidth if the restore is to a different region

- Concurrent restore requests being processed region

Azure SQL DW Backup and Restore

➤ Snapshots of your data warehouse are taken throughout the day creating restore points

➤ These restore points are available for 7 days

    ➤ Retention period of 7 days cannot be changed

➤ SQL pool supports an 8 hour recovery point objective (RPO)

➤ Replicated to paired region once a day

➤ You can also take user-defined snapshots

    ➤ Retention period 7 days cannot be changed

    ➤ 42 max restore point possible

    ➤ Can be created using PowerShell or portal

➤ When you drop a SQL pool, a final snapshot is created and saved for seven days

    ➤ SQL Pool should <u>not be</u> in paused state