

Options available under Primary AlwaysOn AG Replica

In SQL Server, the **Primary Replica** in an **AlwaysOn Availability Group (AG)** is the server that actively handles read-write transactions for the databases in the AG. It replicates transactions to one or more **Secondary Replicas**. Various configuration options can be set for the **Primary Replica** to control failover behavior, availability modes, backup preferences, and read operations.

Here is a detailed explanation of the key options available under the **Primary AlwaysOn AG Replica** in SQL Server:

1. Failover Mode

The **failover mode** determines how and when the system will switch roles between the **Primary Replica** and a **Secondary Replica**.

Options:

- **Automatic:** Failover happens automatically in the event of a failure, but only if the secondary replica is in synchronous commit mode and the failover condition is met (e.g., a heartbeat failure between primary and secondary).
- **Manual:** Failover must be initiated by an administrator. This provides more control, particularly in disaster recovery scenarios where unplanned failovers could be undesirable.

Key Considerations:

- **Automatic failover** is often used in high availability (HA) setups where uptime is critical, and downtime must be minimized.
- **Manual failover** is more typical in disaster recovery (DR) configurations where administrators want to control the process of switching to a secondary site.

2. Availability Mode

This option controls how the **Primary Replica** sends transaction log records to **Secondary Replicas** and ensures data synchronization between the replicas.

Options:

- **Synchronous Commit:** The primary replica waits for the secondary replica to confirm that the transaction is hardened on the secondary before committing it on the primary. This mode guarantees **zero data loss** and is ideal for **high availability**.
- **Asynchronous Commit:** The primary replica does not wait for acknowledgment from the secondary replica before committing the transaction. This mode provides better performance but allows for **potential data loss** in the event of a failure, making it more suitable for **disaster recovery** across geographically distant sites.

Key Considerations:

- **Synchronous Commit** is often used within the same data center for high availability.
- **Asynchronous Commit** is used for disaster recovery scenarios where performance is a priority and some data loss is acceptable in exchange for lower latency.

3. Readable Secondary (Read-Only Routing)

This option defines whether the **Secondary Replicas** associated with the primary replica are configured to allow read-only workloads. This is particularly useful for load balancing read operations such as reporting, thereby offloading the primary replica.

Options:

- **Yes:** The secondary replicas are available for read-only queries. In addition, you can set up **Read-Only Routing**, where incoming read-intent connections are automatically routed to available secondary replicas.
- **No:** The secondary replicas cannot serve read-only workloads and are purely used for high availability or disaster recovery.

Key Considerations:

- Enabling **Readable Secondary** offloads reporting and read-intensive queries from the primary, improving performance on the primary replica.
- **Read-Only Routing Lists** can be configured to control which replicas are chosen for read-only workloads.

4. Backup Preferences

The **backup preferences** setting defines where backups of databases in the availability group should be taken.

Options:

- **Primary:** Backups should always be taken on the primary replica. This is a simple option but may add additional load to the primary server.
- **Secondary:** Backups should only be taken on the secondary replicas, ensuring that the primary replica is not burdened with the backup process.
- **Prefer Secondary:** Backups are preferred on secondary replicas, but if no secondary replica is available, the backup will be performed on the primary.
- **Any Replica:** Backups can be performed on either primary or secondary replicas, depending on availability.

Key Considerations:

- **Offloading backups to secondary replicas** reduces the load on the primary server, improving its performance for read-write operations.
- You can configure **Backup Priority** for each replica, allowing SQL Server to choose the best replica for backup based on the assigned priority.

5. Session Timeout

The **session timeout** setting defines the maximum amount of time (in seconds) that the **Primary Replica** will wait for a response from a **Secondary Replica** before considering it unresponsive. This is important for monitoring the health of connections between replicas and ensuring timely failover.

Options:

- **Default: 10 seconds**
- The value can be customized based on network latency or other factors specific to the environment.

Key Considerations:

- A shorter timeout value ensures that failover happens more quickly in the event of a problem but may cause premature failover if there is temporary network instability.
- A longer timeout value can reduce the chances of false failovers but may delay failover in actual failure scenarios.

6. Application Intent

The **application intent** setting helps in controlling the type of queries that the **Primary Replica** will handle. This is used in conjunction with **Read-Only Routing** to direct read and write workloads appropriately.

Options:

- **ReadWrite**: This indicates that the connection should be directed to the primary replica, which can handle both read and write operations.
- **ReadOnly**: This indicates that the connection is read-only and can be routed to a secondary replica that is configured to handle read-only workloads.

Key Considerations:

- When **Read-Only Routing** is enabled, connections with the **ReadOnly** intent will automatically be routed to available secondary replicas.
- This setting optimizes query performance by segregating read and write workloads between replicas.

7. Host and Port Configuration

The **Primary Replica** is identified by its **host name** (server name or IP address) and **port number**. The default port for SQL Server is **1433**, but custom ports can be used for security or network management purposes.

Options:

- **Host Name**: The DNS name or IP address of the primary replica.
- **Port Number**: The communication port used by the primary replica for client connections.

Key Considerations:

- Custom ports can be used to enhance security by obscuring the default port.
- **Host Name** must be correctly configured to ensure that client applications and secondary replicas can connect to the primary.

8. Quorum Participation

In an AlwaysOn Availability Group deployed over a **Windows Server Failover Cluster (WSFC)**, quorum ensures that the system can accurately determine the health of the replicas and make decisions such as failover. The **Primary Replica** is a voting member in the quorum.

Options:

- **Yes:** The primary replica participates in quorum voting.
- **No:** The primary replica does not participate in quorum voting.

Key Considerations:

- Proper quorum configuration ensures that failover decisions are made based on the majority of healthy nodes.
- Disabling quorum participation for certain replicas can prevent issues in multi-site or disaster recovery setups, especially where network latency might cause replicas to be falsely marked as failed.

9. Health Detection

SQL Server monitors the health of the **Primary Replica** and **Secondary Replicas** to determine if failover is necessary. The **Primary Replica** sends heartbeat signals to the secondary replicas, and their responses are used to determine health.

Key Considerations:

- **Automatic health detection** is critical for ensuring that automatic failover happens only when necessary.
- Proper network and server monitoring should be in place to avoid false positives or downtime.

10. Load Balancing for Read-Only Workloads

When multiple secondary replicas are configured as readable, SQL Server provides options for balancing read-only queries among them based on their configuration and availability.

Key Considerations:

- Read-only workload can be distributed among available replicas based on **Routing Lists** and **Replica Priority**.
- This feature helps in optimizing resource usage by ensuring that no single replica is overloaded with read-only queries.

Best Practices for Configuring the Primary AlwaysOn AG Replica

1. **Use Synchronous Commit for Mission-Critical Applications:** For high availability, ensure that the primary replica and at least one secondary replica are configured for synchronous commit with automatic failover.
2. **Enable Read-Only Routing:** To offload read-intensive workloads (e.g., reporting), configure secondary replicas as readable and use **Read-Only Routing** to automatically direct read-only traffic to secondary replicas.
3. **Offload Backups to Secondary Replicas:** Configure the **Backup Preferences** to prioritize secondary replicas for backups, reducing the load on the primary replica and improving its performance.
4. **Test Failover Scenarios Regularly:** Simulate both planned and unplanned failovers to ensure that your configuration works as expected during actual failure events.
5. **Monitor Network and Replication Latency:** Especially in multi-site configurations, monitor replication latency and tune the **Session Timeout** setting to avoid premature failovers caused by temporary network issues.
6. **Secure Host and Port Settings:** Use custom port numbers and secure the communication between replicas using encryption (SSL/TLS) to protect data in transit.

Summary of Primary Replica Options

Option	Description
Failover Mode	Specifies whether failover is automatic or manual.
Availability Mode	Defines how data is replicated to secondary replicas (synchronous or asynchronous commit).
Readable Secondary	Controls whether secondary replicas are available for read-only workloads.
Backup Preferences	Specifies where backups should occur (primary, secondary, or both).
Session Timeout	Sets the time the primary waits for a secondary replica to respond before considering it unresponsive.
Application Intent	Indicates whether the connection should be read-write or read-only.
Host and Port Configuration	Defines the host name and port used by the primary replica for client and secondary connections.
Quorum Participation	Determines whether the primary replica participates in quorum voting.
Health Detection	Monitors the health of replicas to ensure proper failover behavior.
Load Balancing	Distributes read-only workloads across multiple readable secondary replicas.

These configuration options for the **Primary AlwaysOn AG Replica** provide flexibility and control in optimizing high availability and disaster recovery environments in SQL Server. Properly configuring these settings helps ensure data protection, service continuity, and optimized performance.

<https://www.sqldbachamps.com>

Praveen Madupu - +91 98661 30093
Sr SQL Server DBA, Dubai
praveensqldba12@gmail.com

<https://www.sqldbachamps.com>