

Avoiding the Cost of Confusion: SQL Server Failover Cluster Instances versus Basic Availability Group on Standard Edition

Written by Edwin M Sarmiento,
a Microsoft Data Platform MVP and Microsoft Certified Master

Introduction

With the arrival of SQL Server 2016, customers enthusiastically welcomed the introduction of Basic Availability Groups in Standard Edition - a feature that is set to replace the now deprecated Database Mirroring feature. This provides an opportunity to enable cost-effective high availability and disaster recovery solution for mission-critical databases without the need for an expensive Enterprise Edition license. However, this also introduces confusion among customers who now have to make the decision to choose between implementing Failover Cluster Instances (FCI) or Basic Availability Groups. This confusion can lead to unnecessary hardware, deployment and operations costs that can make the licensing cost seem irrelevant. This paper discusses the difference between FCI and Basic Availability Groups in SQL Server 2016 Standard Edition to serve as a guide to choosing the right solution.

The Side Effects of SQL Server Core-based Licensing

When SQL Server 2012 was released, licensing was made available on a per-core basis. This was unlike in previous versions of SQL Server where licensing was made available on a per-physical processor/socket. Prior to SQL Server 2012, the goal was to buy processors with as many cores as possible in order to maximize the overall processor performance per processor license. This practice became disadvantageous with the core-based licensing model, particularly with SQL Server Enterprise Edition.

Core-based licensing required a minimum of four (4) core licenses per processor which are sold in packs of two (2), where each SKU covers two processors.

Using the core-based licensing model, if you have a dual-core, dual-processor machine - a total of four (4) cores - to run SQL Server 2012 and higher versions, you would need to purchase four (4) SQL Server core license SKUs to cover eight (8) cores despite only having a total of four (4) cores.

SQL Server 2014 introduced yet another change in licensing: the right to install and run an SQL Server instance as a passive failover has become a Software Assurance benefit. If you are running a log shipping secondary, a database mirroring partner, a standby node in an FCI or a secondary replica in an Availability Group, you must purchase Software Assurance to get them licensed.

These changes in the SQL Server licensing model have created a major trend in data center management to minimize cost: an increase in virtualized and consolidated environments. Customers provision a high-end physical server, having processors with as many cores as possible, and license all of the cores with an SQL Server Enterprise

Edition license plus Software Assurance. An unlimited number of virtual machines running SQL Server Enterprise Edition can be deployed on the physical host.

While virtualization and consolidation somehow reduced SQL Server licensing costs, it increased management overhead and bypassed the need for more granular recovery objectives and service level agreements. Every SQL Server virtual machine deployed on the physical host was treated the same regardless of recovery objectives and service level agreements. And while the underlying physical host can be protected by means of failover clustering, the SQL Server virtual machines are not. A powered down virtual machine due to a blue screen can be offline for an extended period of time.

The Drive Towards Enterprise Edition

It was also in SQL Server 2012 when the Availability Groups feature was introduced and made available in Enterprise Edition only. At the same time, Microsoft announced that Database Mirroring will be removed in a future version of SQL Server. Prior to SQL Server 2012, high availability was commonly provided by either using an FCI or Database Mirroring - both of which were available in Standard Edition.

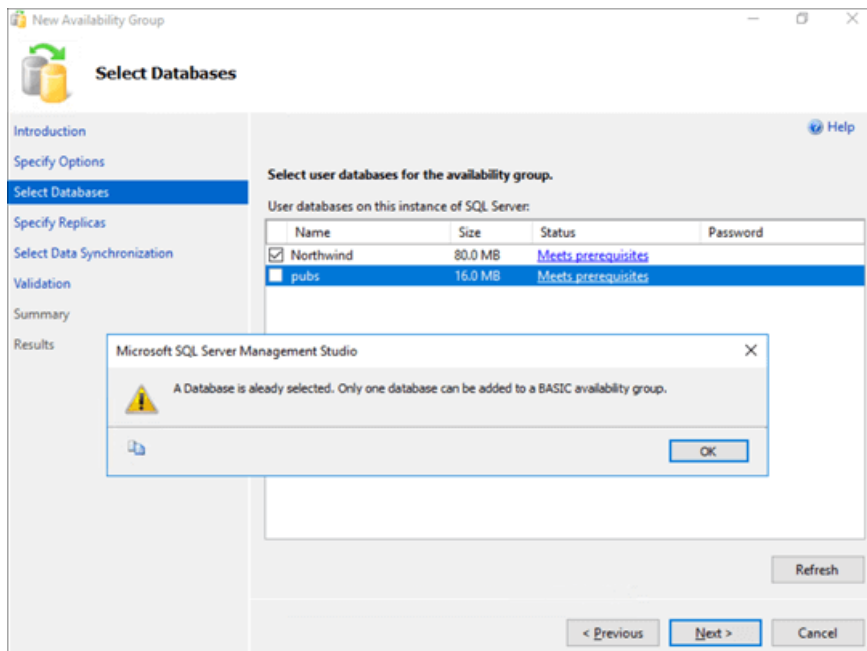
Customers running an FCI continued to maintain their existing infrastructure and upgraded to later versions as appropriate. However, those running Database Mirroring were faced with the challenge of upgrading especially when the principal and the mirror SQL Server instances are running Standard Edition. The dilemma: either postpone the upgrade and run the risk of being in an unsupported configuration or make the leap and upgrade to Enterprise Edition.

These changes in licensing and new features made it very clear that Microsoft really does want customers to move towards Enterprise Edition.

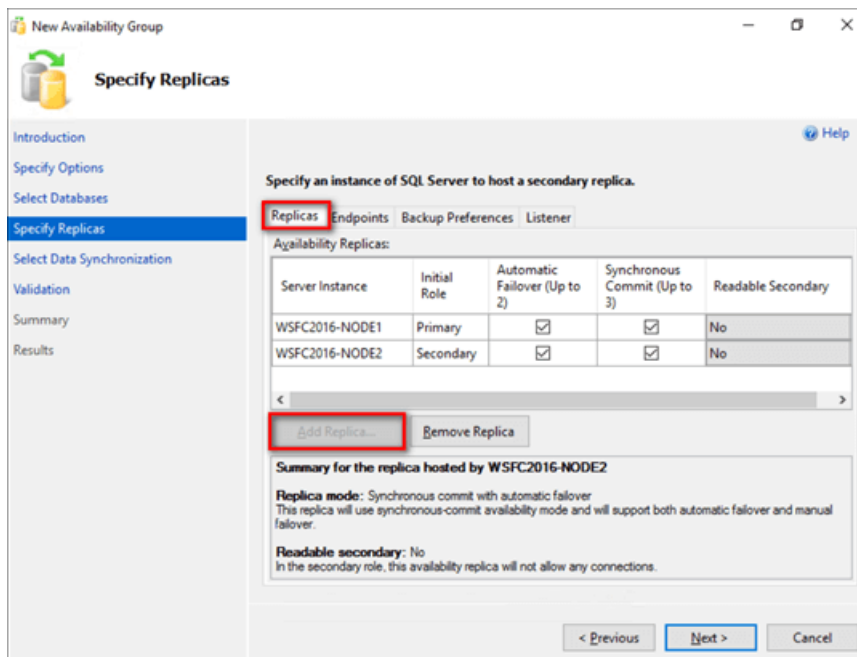
Basic Availability Groups in SQL Server 2016 Standard Edition

SQL Server 2016 introduced the ability to run Basic Availability Groups in Standard Edition. This made Basic Availability Group a truly viable replacement for customers who want to upgrade their Database Mirroring configuration to SQL Server 2016. But unlike its Enterprise Edition counterpart, there are some limitations.

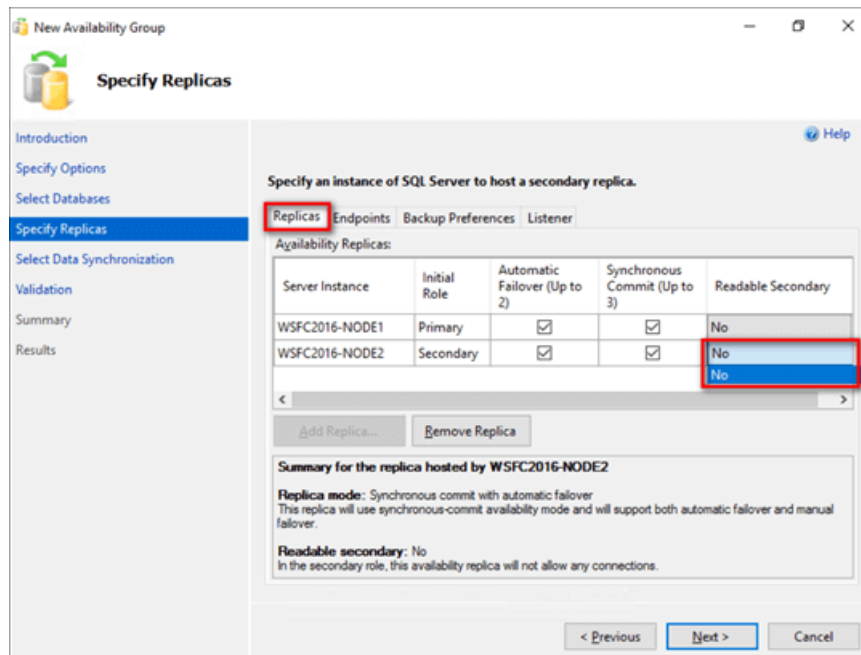
- **You can only join one database per Basic Availability Group.** While you can have an unlimited number of Availability Groups within a replica as your hardware resources permit, you can only have one database in a Basic Availability Group. This becomes a challenge when dealing with applications that require multiple databases like SharePoint. You will get this warning message should you choose more than one database in the **Select Databases** page of the **New Availability Group Wizard**.



- **You can only have two (2) Availability Group replicas - one primary and one secondary.** Traditional Availability Groups in Enterprise Edition allow you to have up to four (4) secondary replicas with SQL Server 2012 and eight (8) secondary replicas with SQL Server 2014 and higher. The increased number of secondary replicas allows you to have both high availability and disaster recovery solutions within the same infrastructure. With Basic Availability Groups, you are limited to two (2) replicas. The **Add Replica** button in the **Specify Replicas** page of the **New Availability Group Wizard** will automatically get disabled after adding the second replica.



- **You cannot enable the secondary replica as a readable secondary.** This is the same as Database Mirroring in Standard Edition. If you want to enable the secondary replica (or mirror partner in Database Mirroring) for reporting workloads, you need Enterprise Edition. **No** is the only option when you click on the drop-down list under the **Readable Secondary** column in the **Specify Replicas** page of the **New Availability Group Wizard**.



- **You cannot run backups on the secondary replica.**
- **You cannot upgrade Basic Availability Groups to the traditional Availability Groups in Enterprise Edition.** If you need to upgrade your SQL Server licenses from Standard to Enterprise Edition, you have to remove the Basic Availability Group first and recreate it after the upgrade.

While there are limitations, there are also advantages when you compare them with Database Mirroring in Standard Edition.

- **Allows asynchronous commit mode.** Database Mirroring in Standard Edition only allows for synchronous commit mode which significantly affects performance on the principal database. Asynchronous commit mode enhances performance on the primary replica. However, this comes with the risk of potential data loss.
- **Leverages listener name for automatic application redirection.** Database Mirroring relies on modifying the application connection string to automatically redirect client

applications to the mirror server upon failover. If the client application does not support the Failover Partner connection string parameter, you need to create a process that depends on DNS aliases for client application redirection after a failover. An Availability Group listener name is a Windows Server Failover Clustering resource that resides in the same cluster resource group as the Availability Group. The listener name moves together with the Availability Group. Client applications simply need to use the listener name and they will always be redirected to the current primary replica.

- **AlwaysOn_health Extended Events Session.** The AlwaysOn_health Extended Events session was introduced in SQL Server 2012 to assist database administrators with root-cause diagnosis when troubleshooting Availability Groups. With Database Mirroring, you had to rely on a combination of the SQL Server error logs and the Windows Event logs to find out what caused a Database Mirroring configuration to go offline or failover unexpectedly.

Comparing FCI and Basic Availability Groups in SQL Server 2016 Standard Edition

Understanding the benefits and limitations of Basic Availability Groups can give you enough information to decide whether you would like to still use Database Mirroring regardless of its deprecation status or upgrade to Enterprise Edition to take advantage of all the Availability Group features. However, both options still put you at a disadvantage - the former being at the risk of using an unmaintained feature, the latter being more expensive.

Another high availability feature exists in SQL Server 2016 Standard Edition: Failover Cluster Instances (FCI). This feature is not new and has been around since SQL Server 6.5. Prior to SQL Server 2005, FCI was only available in Enterprise Edition. With Standard Edition, there is a limit of two (2) nodes for an FCI - similar to Basic Availability Groups.

There are a lot of similarities between FCI and Basic Availability Groups. Highlighted below are the differences between the two that can help you make the right decision in choosing the appropriate high availability solution. While the comparisons apply to traditional Availability Groups as well, the focus will be on Basic Availability Groups.

Instance-level protection versus database-level protection

An SQL Server instance refers to the entire SQL Server installation with its own system databases, user databases, logins, memory and CPU allocations, etc. Each SQL Server instance is its own independent installation and can be assigned resources such as CPU, memory, disk, and network. A database is an object that resides within an instance.

Understanding the difference between an instance and a database can help make decisions in implementing the appropriate high availability solution in terms of operational efficiency.

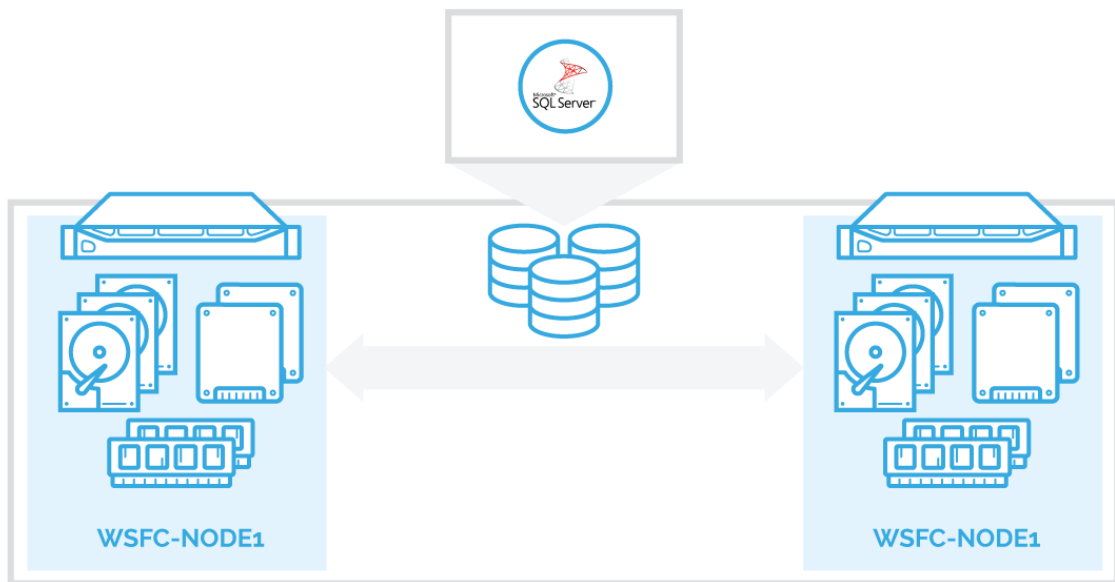
In an FCI, the entire instance is protected. If the primary node becomes unavailable, the entire instance is moved to the standby node. This takes care of the SQL Server logins, SQL Server Agent jobs, custom objects stored in the system databases, certificates, Extended Events sessions, and more. These instance-level objects are stored in the system databases which are physically stored in shared storage.

In an Availability Group - be it the traditional or basic - only the databases in the group are protected. System databases cannot be added to an Availability Group - only user databases are allowed. It is the database administrator's responsibility to replicate system objects like SQL Server logins, SQL Server Agent jobs, certificates, etc. on all Availability Group replicas. If all the dependent system objects are not replicated on all replicas, the database may end up becoming inaccessible to the application; as in the case of missing SQL Server logins or partially functional as in the case of missing certificates for Always Encrypted (SQL Server 2016 Service Pack 1 made Always Encrypted available in Standard Edition).

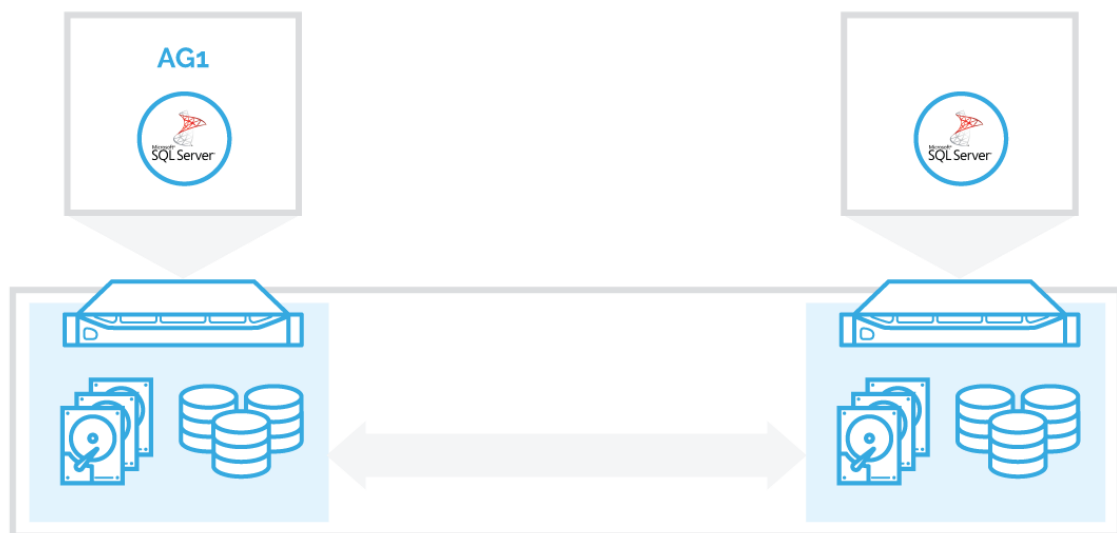
If you are more concerned with instance-level protection to minimize possible human error during change management processes, then, an FCI is the way to go. There is a possibility of increased human error with Basic Availability Groups since it is the database administrator's responsibility to replicate all of the changes to all of the replicas.

Local storage versus shared storage

An FCI requires some form of shared storage. The shared storage is accessible to all of the nodes in the failover cluster but only the current primary node has ownership at any given point in time. The system and user databases are stored on the shared storage. When a failover occurs, ownership of the shared storage moves from the current primary to the standby, making the databases available to the new primary node. From a capacity point-of-view, you only need to provision disk space based on the sizes of the databases. However, from an availability point-of-view, the shared storage becomes a single point of failure. The FCI will remain offline if the shared storage becomes unavailable, regardless of the number of nodes in the failover cluster.



An Availability Group does not require shared storage. Each replica has its own local storage independent of the Availability Group. If the primary replica becomes unavailable and a failover occurs, any secondary replica can take over without having to rely on the availability of the system and user databases from the primary replica. From a capacity point-of-view, you need to provision disk space based on the sizes of the databases and the number of secondary replicas, significantly increasing the cost per gigabyte. But from an availability point-of-view, because the Availability Group replicas do not rely on a single storage source, the SQL Server instance hosting the replicas remains online regardless of what happens to the primary replica. Each Availability Group replica has its own copy of the system databases, thus, it is always online regardless of whether it is functioning as a primary or a secondary replica. This is also the reason why failover is much faster in an Availability Group than it is in an FCI.



If you don't have a dedicated team of storage engineers who are responsible for a shared storage subsystem, then, Basic Availability Group is the right solution. The complexities of a shared storage subsystem like a storage area network (SAN) require the proper administration to guarantee high availability and resiliency because the storage becomes a single point of failure.

Single copy versus multiple copies

You may have certain disaster recovery requirements to have multiple copies of your databases. If this is the case, then, Availability Group is your solution.

Because of the shared storage requirement, an FCI will only have a single copy of the databases. If you need multiple copies of your databases, you will need to implement either log shipping, database mirroring or Availability Group together with the FCI. This requires additional cost in terms of licensing and administrative overhead. Not so with Availability Groups. Each replica has its own copy of the databases, making it both a high availability and disaster recovery solution that meets the “multiple copies of the database” requirement. If the storage subsystem on the primary replica becomes unavailable, failing over to the secondary replica addresses high availability and the secondary replica’s copy of the databases addresses disaster recovery.

More Licensing Beyond the Usual

Just because both FCI and Basic Availability Groups are a Standard Edition feature doesn’t mean you no longer have to worry about licensing once purchased. In both 2-node FCI and Basic Availability Groups, a license is required to run on the primary server. The standby server is only covered by a license if it comes with Software Assurance.

Assume that you will only run a single instance FCI in a 2-node failover cluster and two replicas in a Basic Availability Group. One might think that a single Standard Edition license with Software Assurance will cover both servers. This is true in the case of an FCI. Since you only have a single FCI, it can only be available on one of the failover cluster nodes at any given point in time. You cannot run the same FCI on all of the failover cluster nodes at the same time. In an FCI, the standby node acts as a true standby server.

In an Availability Group, since all of the replicas are online and available at any given point in time, there is a possibility of running multiple Availability Groups simultaneously on any of the replicas. This is specifically true in Basic Availability Groups where you will have multiple Availability Groups, one per database. Because an Availability Group exists as a resource group in a failover cluster, it can run on any of the available replicas. This does not become a licensing issue if you only have one Basic Availability Group. However, to maximize IT investments, there will be more than one database running on a SQL Server instance acting as a replica in a Basic Availability Group. Thus, requiring more than one Basic Availability Group.

In almost all cases, the failover cluster will move all of the Basic Availability Groups together when a failover occurs. However, there will be rare cases where one Basic Availability Group might end up in a different replica than the rest, for example, initiating a manual failover of one of the Basic Availability Groups or a failover policy that automatically moves a Basic Availability Group from one replica to another. When this happens, you need to define a process on how to keep all of

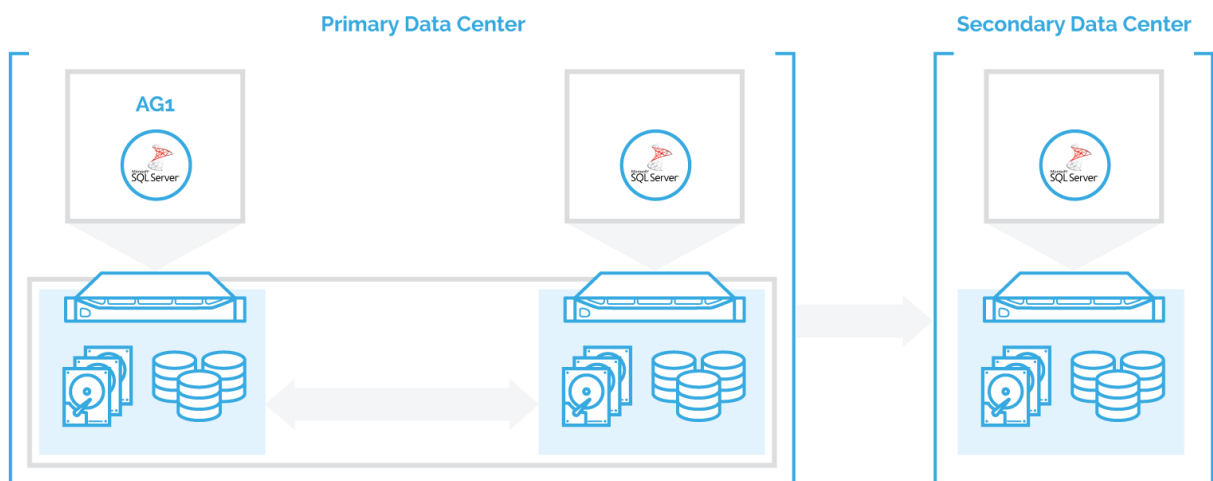
the Basic Availability Groups running on the same replica to avoid being in violation of your licensing agreement.

Below is a summary of the comparison between Basic Availability Groups and FCI in SQL Server 2016 Standard Edition.

	Basic Availability Group	Failover Cluster Instance
Instance-level protection	x	✓
Database-level protection	✓	✓
Local storage	✓	x
Shared storage	x	✓
Single copy of the database	x	✓
Multiple copies of the database	✓	x
No licensing violation	x	✓

Combining FCI and Basic Availability Groups in SQL Server 2016 Standard Edition

There could be cases where you would need instance-level protection, multiple copies of your databases and a guaranteed true standby server while minimizing licensing costs. Instead of implementing just an FCI or a Basic Availability Group, why not have both? The reference diagram below describes this architecture.



All three (3) servers are nodes in the same failover cluster. A 2-node SQL Server FCI with shared storage runs on WSFC-NODE1 and WSFC-NODE2 while WSFC-NODE3 with local storage runs a standalone SQL Server instance. AG1 is a Basic Availability Group that has two (2) replicas - the FCI and the standalone instance. The FCI provides instance-level protection for local high availability, guaranteeing a true standby server, while the Basic Availability Group provides the extra copy of the databases for disaster recovery purposes. A failover of the FCI from WSFC-NODE1 to WSFC-NODE2 and vice versa will take the Basic Availability Group offline. However, you are guaranteed that all of the Basic Availability Groups will only run on either WSFC-NODE1 or WSFC-NODE2. That's because you lose the ability to automatically failover Basic Availability Groups when an FCI is configured as one of the replicas. You will have to manually force the Basic Availability Groups to run on WSFC-NODE3.

And while you may need an extra Standard Edition license on top of Software Assurance to cover WSFC-NODE3, it is still relatively cheaper than an Enterprise Edition license.

Summary

Basic Availability Groups is a welcome addition to the list of high availability and disaster recovery features in SQL Server 2016 Standard Edition. But with all of the features available, choosing the right solution can become daunting and confusing. Understanding the difference between an FCI and Basic Availability Groups in SQL Server 2016 Standard Edition can help guide you in making decisions when implementing the appropriate high availability and disaster recovery solution for your SQL Server databases.