Monitoring Microsoft AppFabric Caching

eG Innovations Product Documentation

www.eginnovations.com

Total Performance Visibility

Table of Contents

Table of Figures

Chapter 1: Introduction

Microsoft AppFabric for Windows Server Caching features use a cluster of servers that communicate with each other to form a single, unified application cache system. As a distributed cache system, all cache operations are abstracted to a single point of reference, referred to as the cache cluster. In other words, your client applications can work with a single logical unit of cache in the cluster regardless of how many computers make up the cache cluster.

The primary components of the physical architecture consist of the cache server, the cache host Windows service, the cache cluster, the Windows PowerShell-based cache administration tool, the cluster configuration storage location, and the cache client.

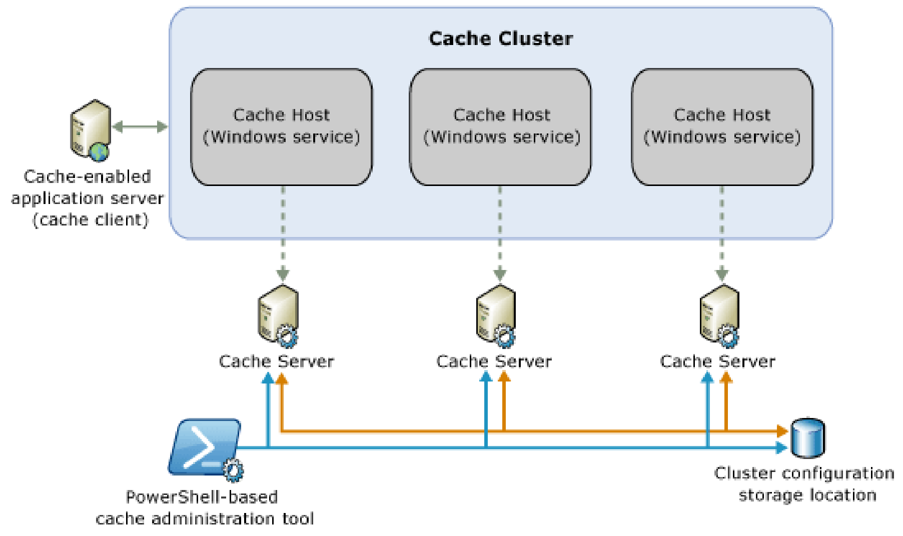

Cache Cluster
Cache Host (Windows service)
Cache Host (Windows service)
Cache Host (Windows service)
Cache-enabled application server (cache client)
Cache Server
Cache Server
Cache Server
PowerShell-based cache administration tool
Cluster configuration storage location

Figure 1.1: The Microsoft AppFabric Caching architecture

The AppFabric Caching Service is a Windows service that runs on one or more servers. Each server that runs the Caching Service is referred to as a cache server. For each cache server, only one instance of the Caching Service can be installed.

The cache cluster is a collection of one or more instances of the Caching Service working together in the form of a ring to store and distribute data. Data is stored in memory to minimize response times for data requests. The operations of the cache cluster are managed by a role, named the cluster management role. The primary responsibility of the cluster management role includes:

- Keeping the cache cluster running at all times.
- Monitoring the availability of all cache hosts in the cache cluster.

- Helping cache hosts join the cache cluster.

Any application server that is running a cache-enabled application may be loosely referred to as the cache client. For an application to be cache-enabled, it must use the AppFabric Caching assemblies.

Cache is a key ingredient in the design and delivery of a wide variety of applications ranging from single- user embedded systems to large, multi- server, multi- user installations (such as those required by banks, hospitals, etc.) providing essential services to end-users.

This dependence on Cache for performing business-critical tasks and for developing mission-critical applications could only mean that even a wafer- thin deviation in its performance could cause an enterprise to lose millions. Database administrators are thus faced with the daunting task of ensuring the 24x7 availability of the Microsoft AppFabric Caching feature and the optimal performance of all its components. eG Enterprise helps administrators in this task.

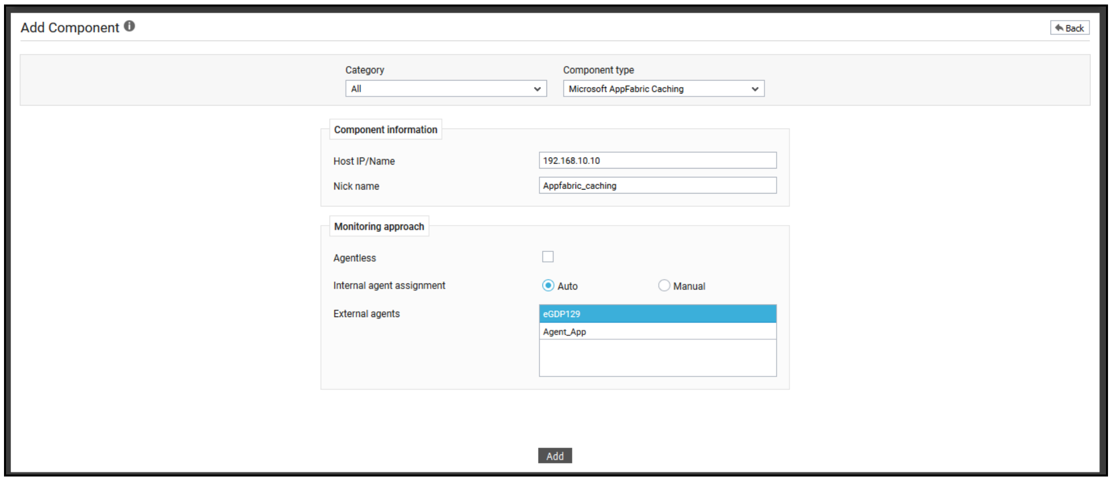
Chapter 2: How to Monitor Microsoft AppFabric Caching Using eG Enterprise

eG Enterprise monitors the Microsoft AppFabric Caching in both agent based and agentless manners.

2.1 Managing the Microsoft AppFabric Caching

eG Enterprise cannot automatically discover the Microsoft AppFabric Caching. You need to manually add the component for monitoring. To manage a Microsoft AppFabric Caching component, do the following:

1. Log into the eG administrative interface.
2. Add the component manually using the **COMPONENTS** page (Infrastructure -> Components -> Add/Modify) of the eG administrative interface. Remember that components manually added are managed automatically.
3. In the **COMPONENTS** page, select *Microsoft AppFabric Caching* as the **Component type**. Then, click the **Add New Component** button. This will invoke Figure 2.1.



Category: All
Component type: Microsoft AppFabric Caching

Component information
Host IP/Name: 192.168.10.10
Nick name: Appfabric_caching

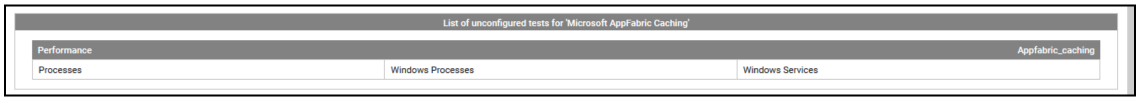
Monitoring approach
Agentless: ☐
Internal agent assignment: [✓ Auto] Manual
External agents: eGDP129, Agent_App

Figure 2.1: Adding a Microsoft AppFabric Caching component

4. Specify the **Host IP/Name** and **Nick name** of the Microsoft AppFabric Caching in Figure 2.1. Then, click the **Add** button to add the component for monitoring.

2.2 Configuring the tests

1. The tests pertaining to Microsoft AppFabric Caching will be configured, by default. When you attempt to signout of the eG administrative interface, a list of unconfigured tests listing the tests requiring manual configuration, will appear (see Figure 2.2).



List of unconfigured tests for 'Microsoft AppFabric Caching'

Performance		Appfabric_caching
Processes	Windows Processes	Windows Services

Figure 2.2: List of unconfigured tests for Microsoft AppFabric Caching

2. The **Processes, Windows Processes and Windows Services** tests require manual configuration. To know the details on configuring these tests, refer to the *Monitoring Unix and Windows Servers* document.

3. Finally, signout of the eG administrative interface.

Chapter 3: Monitoring Microsoft AppFabric Caching

eG Enterprise offers a specialized monitoring model for the Microsoft AppFabric Caching (see Figure 3.1) that monitors the cache 24 x 7 and proactively alerts administrators of probable issues in its operations, so that issues are trapped very early and resolved before its too late.

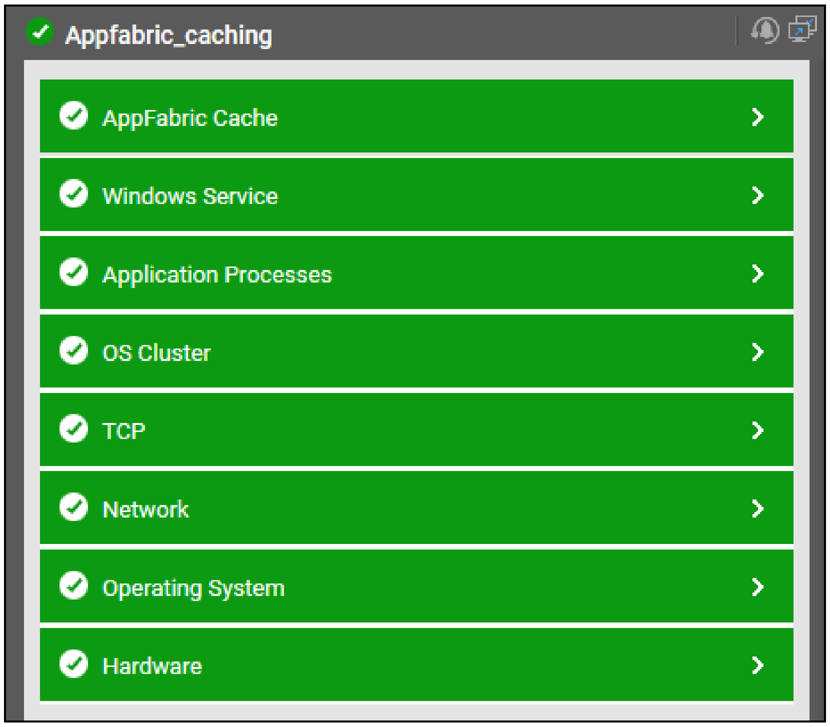


Figure 3.1: Layer model of the Microsoft AppFabric Caching

Each of the layers depicted by the hierarchical model above, is associated with a wide variety of tests that measures the performance of the Microsoft AppFabric Caching. The performance metrics reported by these tests shed light on the following:

- How many requests were not serviced by each cache?
- What is the total size of the cached data in each cache?
- What is the total size of the primary cached data in each cache?
- What is the total size of the secondary cached data in each cache?
- How many objects were stored in each cache?
- How many read requests were received by each cache since the start of the cache service?

- How many write requests were received by each cache since the start of the cache service?
- How many GetAndLock requests were received per second by each cache?
- How many GetAnd Lock requests were successful on each cache?
- What is the average quorum response time on the primary host?
- What is the average time spent to get response from the secondary servers?
- How many expired objects and evicted objects are available in the primary host?
- What is the total size of the primary cached data?
- How many retry operation exceptions were performed on the primary host since the start of the cache service?
- How many Get requests were received from all clients on the primary host since the start of the cache service?
- How many read/write requests were received from all clients since the start of the cache service?
- How many objects were stored in the host?
- How many times the replication operation was retried on the secondary cache server?

Since the **Network** layer has been dealt in *Monitoring Cisco Routers* document and the remaining layers except the **AppFabric Cache** layer have been dealt extensively in *Monitoring Unix and Windows Servers* document, the forthcoming section will deal elaborately on the AppFabric Cache layer.

3.1 The AppFabric Cache Layer

The **AppFabric Cache** layer evaluates how well the cache processes requests, the size of the data, whether cache misses are more, the request serving capability of the host, the number of times replication was retried on the secondary cache etc.

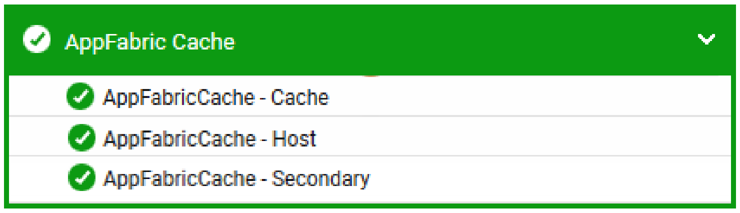


Figure 3.2: The tests associated with the AppFabric Cache layer

3.1.1 AppFabricCache - Cache Test

A named cache, also referred to as a cache, is a configurable unit of in-memory storage that all applications use to store data in the distributed cache. The AppFabric Caching Service is a Windows service that runs on one or more servers. Each server that runs the Caching Service is referred to as a cache server. For each cache server, only one instance of the Caching Service can be installed. The cache cluster is a collection of one or more instances of the Caching Service working together in the form of a ring to store and distribute data. Data is stored in memory to minimize response times for data requests.

This test auto-discovers each cache on the target Microsoft AppFabric Caching and for each cache, monitors the requests, checks how well the cache processes the requests, and reveals whether cache misses are more. The size of the data retrieved from the cache and the request processing ability of the cache is also reported.

Target of the test : A Microsoft AppFabric Caching

Agent deploying the test : An internal/remote agent

Outputs of the test : One set of results for each cache on the target Microsoft AppFabric Caching server being monitored

Configurable parameters for the test

Parameter	Description
Test Period	How often should the test be executed.
Host	The IP address of the host for which this test is to be configured.
Port	Specify the port at which the specified Host listens. By default, this is *NULL*.

Measurements made by the test

Measurement	Description	Measurement Unit	Interpretation
Cache misses	Indicates the number of requests that were not served from this cache.	Number	Ideally, the value of this measure should be 0. A very high value is a cause for concern as it indicates that the cache is poorly utilized.
Percentage of cache	Indicates the percentage of	Percent	A low value is desired for this

Measurement	Description	Measurement Unit	Interpretation
misses	requests that were not served from this cache.		measure.
Cache misses per sec	Indicates the rate at which the requests to this cache was not served successfully.	Requests/sec	
Data size	Indicates the total size of the cached data in this cache.	MB	The size of the cached data does not include cache overhead.
Primary data size	Represents the total size of primary cached data in this cache, not including cache overhead. Indicates the current primary memory data usage of this cache.	MB	
Secondary data size	Represents the total size of secondary cached data in this cache, not including cache overhead. Indicates the current secondary memory data usage of this cache.	MB	
Object count	Indicates the total number of objects stored in this cache.	Number	
Client requests	Indicates the total number of client requests including API calls to this cache.	Number	
Client requests per sec	Indicates the number of client requests per second, including API calls to this cache.	Requests/sec	
Read requests	Indicates the number of read requests received from the clients by this	Number	

Measurement	Description	Measurement Unit	Interpretation
	cache since the start of the cache service.		
Read requests per sec	Indicates the number of read requests received per second from the clients by this cache since the start of the cache service.	Number	
Objects returned	Indicates the number of objects returned by read requests received by this cache.	Number	
Objects returned per sec	Indicates the number of objects returned by read requests per second by this cache.	Number	
Write operations	Indicates the number of writes requests received by this cache since the start of the cache service.	Number	
Write operations per sec	Indicates the number of write requests received by this cache per second since the start of the cache service.	Number	
Get and lock requests	Indicates the total number of GetAndLock requests received by this cache since the start of the cache service.	Number	
Get and lock requests rate	Indicates the total number of GetAndLock requests per second received by this cache since the start of the cache service.	Number	
Successful get and lock requests	Indicates the number GetAndLock requests that	Number	

Measurement	Description	Measurement Unit	Interpretation
	were successful since start of cache service on this cache.		
Successful get and lock requests rate	Indicates the number of GetAndLock requests that were successful since the start of the cache service on this cache.	Number	

3.1.2 AppFabricCache - Host Test

This test monitors the Microsoft AppFabric Caching and reports host-level performance metrics. The request serving ability and the size of the data of the primary cache server or the host cache is monitored and reported. Using this test, administrators are alerted to poor responsiveness (if any) of the primary cache so that corrective measures can be quickly initiated.

Target of the test : A Microsoft AppFabric Caching

Agent deploying the test : An internal/remote agent

Outputs of the test : One set of results for the target Microsoft AppFabric Caching server being monitored

Configurable parameters for the test

Parameter	Description
Test Period	How often should the test be executed.
Host	The IP address of the host for which this test is to be configured.
Port	Specify the port at which the specified Host listens. By default, this is *NULL*.

Measurements made by the test

Measurement	Description	Measurement Unit	Interpretation
Avg quorum response time	Indicates the amount of time spent by write operations in replication.	Secs	

Measurement	Description	Measurement Unit	Interpretation
Avg secondary response time	Indicates the average time spent to get response from all the secondary servers.	Secs	A low value is desired for this measure.
Percentage of cache miss	Indicates the percentage of unsuccessful cache requests to the total number of requests since the start of the cache service.	Percent	
Cache misses	Indicates the total number of unsuccessful cache requests since the start of the cache service.	Number	Ideally, the value of this measure should be zero. A sudden/gradual increase in the value of this measure indicates that the requests are not serviced from the cache.
Cache misses per sec	Indicates the total number of unsuccessful cache requests per second since start of cache service.	Number	
Data size	Indicates the total size of cached data in the cache, not including cache overhead.	MB	
Evicted objects	Indicates the number of evicted objects since the start of the cache service.	Number	
Eviction runs	Indicates the number of eviction runs since the start of the cache service.	Number	
Expired objects	Indicates the number of expired objects since the start of the cache service.	Number	
Memory evicted	Indicates the amount of memory freed from cache since the start of the cache service.	Number	
Primary data size	Indicates the total size of	MB	

Measurement	Description	Measurement Unit	Interpretation
	primary cached data in the cache, not including cache overhead.		
Secondary data size	Indicates the total size of secondary cached data in the cache, not including cache overhead.	MB	
Failure exceptions	Indicates the number of exceptions that are being thrown by the server since start of cache service.	Number	
Failure exceptions per sec	Indicates the number of exceptions per second that are being thrown by the server since start of cache service.	Number	
Retry exception	Indicates the total number of retry operation exceptions since the start of the cache service.	Number	
Retry exception per sec	Indicates the total number of retry operation exceptions per second since the start of the cache service.	Number	
Client requests	Indicates the total number of client requests, including all API calls.	Number	
Client requests per sec	Indicates the total number of client requests per second, including all API calls.	Number	
Get misses	Indicates the number of Get misses from all clients since the start of the cache service.	Number	

Measurement	Description	Measurement Unit	Interpretation
Get misses per sec	Indicates the number of Get misses per second from all clients since the start of the cache service.	Number	
Get requests	Indicates the number of Get requests received from all clients since the service was started.	Number	
Get requests sec	Indicates the number of Get requests received per second from all clients since the service was started.	Number	
Get and lock requests	Indicates the total number of GetAndLock requests since the start of the cache service.	Number	
Get and lock requests per sec	Indicates the total number of GetAndLock requests received per second since the start of the cache service.	Number	
Successful get and lock requests	Indicates the number of GetandLock requests that were successful since the start of the cache service.	Number	A High value is desired for this measure.
Successful get and lock requests per sec	Indicates the number of GetandLock requests that were received successfully per second since the start of the cache service.	Number	
Read requests	Indicates the number of read requests (Bulk Get, Get and Enumeration) received from all clients since the start of the cache	Number	

Measurement	Description	Measurement Unit	Interpretation
	service.		
Read requests per sec	Indicates the number of read requests (Bulk Get, Get and Enumeration) received per second from all clients since the start of the cache service.	Number	
Write operations	Indicates the number of write requests since the start of the cache service.	Number	
Write operations per sec	Indicates the number of write requests per second since the start of the cache service.	Number	
Requests served	Indicates the number of request served and responses sent by the server since the start of the cache service.	Number	A high value is desired for this measure.
Requests served per sec	Indicates the number of request served and responses sent by the server per second since the start of the cache service.	Number	
Object count	Indicates the total number of objects stored in the host.	Number	
Objects returned	Indicates the number of objects returned by client read requests.	Number	
Objects returned per sec	Indicates the number of objects returned by client read requests per second.	Number	
Notification delivered	Indicates the number of notifications delivered to	Number	

Measurement	Description	Measurement Unit	Interpretation
	clients.		
Notification delivered per sec	Indicates the number of notifications delivered per second to clients.	Number	
Notification poll requests	Indicates the total number of poll request from client since the start of the cache service.	Number	
Notification poll requests per sec	Indicates the total number of poll request per second from client since the start of the cache service.	Number	

3.1.3 AppFabricCache - Secondary Test

This test reports the number of times replication operation was retried on the secondary cache server available on the target Microsoft AppFabric Caching server.

Target of the test : A Microsoft AppFabric Caching

Agent deploying the test : An internal/remote agent

Outputs of the test : One set of results for the target Microsoft AppFabric Caching server being monitored

Configurable parameters for the test

Parameter	Description
Test Period	How often should the test be executed.
Host	The IP address of the host for which this test is to be configured.
Port	Specify the port at which the specified Host listens. By default, this is *NULL*.

Measurements made by the test

Measurement	Description	Measurement Unit	Interpretation
Replication retries	Indicates the number of times the replication operation was retried on the secondary cache server.	Number	A low value is desired for this measure. If the value of this measure increases suddenly/gradually, then, it indicates that the response from the secondary cache server is becoming poor. This would invariably affect the time taken to complete the replication operation. Administrators should therefore, carefully analyze the real reason behind a large volume of replication operation retries and rectify problems before end users start complaining.

About eG Innovations

eG Innovations provides intelligent performance management solutions that automate and dramatically accelerate the discovery, diagnosis, and resolution of IT performance issues in on-premises, cloud and hybrid environments. Where traditional monitoring tools often fail to provide insight into the performance drivers of business services and user experience, eG Innovations provides total performance visibility across every layer and every tier of the IT infrastructure that supports the business service chain. From desktops to applications, from servers to network and storage, from virtualization to cloud, eG Innovations helps companies proactively discover, instantly diagnose, and rapidly resolve even the most challenging performance and user experience issues.

eG Innovations is dedicated to helping businesses across the globe transform IT service delivery into a competitive advantage and a center for productivity, growth and profit. Many of the world's largest businesses use eG Enterprise to enhance IT service performance, increase operational efficiency, ensure IT effectiveness and deliver on the ROI promise of transformational IT investments across physical, virtual and cloud environments.

To learn more visit www.eginnovations.com.

Contact Us

For support queries, email support@eginnovations.com.

To contact eG Innovations sales team, email sales@eginnovations.com.

Copyright © 2020 eG Innovations Inc. All rights reserved.

This document may not be reproduced by any means nor modified, decompiled, disassembled, published or distributed, in whole or in part, or translated to any electronic medium or other means without the prior written consent of eG Innovations. eG Innovations makes no warranty of any kind with regard to the software and documentation, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The information contained in this document is subject to change without notice.

All right, title, and interest in and to the software and documentation are and shall remain the exclusive property of eG Innovations. All trademarks, marked and not marked, are the property of their respective owners. Specifications subject to change without notice.