

Performance Measurements And Comparison For Gluster FS And Azure Blob Storage

Roopali Vij

Abstract: as the world of knowledge based systems and digital knowledge sharing grows, business models involving document management and storage of large blocks of files is becoming the need of the hour. It is very essential to select the correct and efficient file system to store the files in order to ease the retrieving and addition of files. Although using cloud technologies to save such data offers flexibility, the biggest challenge is to select whether to opt for a distributed file system mounted over cloud virtual machines or use PaaS based file storages available as platforms. Here, we compare the performance of two environments both deployed on a same storage account on Azure cloud. One is GlusterFS file system mounted on a virtual machine on Azure and the other is PaaS based Azure blob storage using a website for encryption hosted on Azure.

Index Terms: Azure, Blob Storage, Cloud applications, File Systems, Gluster, PaaS, Storage Systems, Systems, Technology.

1 INTRODUCTION

In the era of new knowledge materials being digitally shared around the world everyday in large amounts, the need for maintaining large amount of data and documents in computer applications has increased tremendously. We require an efficient means to store and render a knowledge artifact that we save digitally on any infrastructure -, particularly on cloud infrastructure with the growing world of cloud computing. With the growth of internet and knowledge sharing, an inevitable need of good file systems is rising and as 'necessity is the mother of all inventions', these days market provides us with all sorts of big data solutions and cloud storage options. Most frequently, people are in doubt of deploying a file system or going for a PaaS based storage solutions. Some very good options available in market to suffice this growing need are:

1. Distributed FileSystem
 - GlusterFS
 - CEPH
2. PaaS Based Storage
 - Azure Blob
 - Amazon S3

While there are many aspects like security, performance, structuring, server side encryption etc., in this paper performance of two well known systems (GlusterFS and Azure Blob Storage) are compared to ease one of the parameters for your choice.

2 Configurations Used For Measurements

2.1 Gluster File System Configuration

Three Azure virtual machines with configuration of Centos 6.5 D2 (installed Linux as operating system) are being used as nodes for a three node Gluster file system setup. GlusterFS server encryption is used to measure the scenario for encrypted files.

2.2 Azure Blob Storage Configuration

A storage account was created and a blob storage container was used in the same geographical zone as the above VM(S) used for GlusterFS. Since, Azure Blob storage does not provide facility for server side encryption, blob encryption policy along with Azure keyVault is used to encrypt the files for the respective scenario. The code for encryption was hosted as a basic subscription website on the same azure environment.

3 A short Introduction Of Both Systems

3.1 Gluster File System

The Gluster File System (GlusterFS) is an open source distributed file system that can scale out in building-block fashion to store large amount of data. GlusterFS has a client and server component. Servers are typically deployed as storage bricks, with each server running a glusterfsd daemon to export a local file system as a volume. The glusterfs client process, which connects to servers with a custom protocol over TCP/IP, InfiniBand or Sockets Direct Protocol, creates composite virtual volumes from multiple remote servers using stackable translators. By default, files are stored whole, but striping of files across multiple remote volumes is also supported. The final volume may then be mounted by the client host using its own native protocol via the FUSE mechanism, using NFS v3 protocol using a built-in server translator, or accessed via gfapi client library. Native-protocol mounts may then be re-exported e.g. via the kernel NFSv4 server, SAMBA, or the object-based OpenStack Storage (Swift) protocol using the "UFO" (Unified File and Object) translator.

3.2 Azure Blob Storage

Azure Blob storage is a service for storing large amounts of unstructured data, such as text or binary data, that can be accessed from anywhere in the world via HTTP or HTTPS. You can use Blob storage to expose data publicly to the world, or to store application data privately. In the most general sense, the term "blob" is commonly understood to mean "Binary Large Object." Many of us are familiar with this term from its usage in database-land, where "blob data" might be data stored in our database which does not conform to an established data type as defined by the database. Such data are usually (if the database supports it) persisted as plain binary data (image files come to mind as an example). 4

- *Roopali Vij is Bachelors of Technology in Electronics and Communications from Guru Gobind Singh Indraprastha University, India, PH-918447631116.*
- *E-mail: roopali_vij@live.com*

Performance Results The performance matrix created for the two systems for same document load is given below:

4.1 Gluster File System Performance Results

Performance metrics for Gluster File System was captured by using 200 identical files of approximately 1MB each (total load of 195 MB). Notice how the remarkable performance degrades as soon as the encryption functionality is added to the files.

Data Size	No of files	Encryption	Operation	Time Elapsed (seconds)
195MB	200	ON	Write	173.16607
195MB	200	ON	Read	9.615847
195MB	200	OFF	Write	9.780412
195MB	200	OFF	Read	6.279751

4.2 Azure Blob Storage Performance Results

Again, in case of Azure blob storage performance measurement, the same files which were used to measure the performance of GlusterFS were utilized. Although the performance of retrieving and adding files to Azure blobs cannot surpass the Gluster File System on similar environment, it definitely gives a better platform if you need to encrypt your files.

Data Size	No of files	Encryption	Operation	Time Elapsed (seconds)
195MB	200	ON	Write	33.637981
195MB	200	ON	Read	40.885859
195MB	200	OFF	Write	15.629685
195MB	200	OFF	Read	19.162182

5 CONCLUSION

Performance metrics for GlusterFS as a file system on Azure cloud was compared with PaaS based Azure blob Storage. With similar data load and similar hosting environments, although GlusterFS performs better in usual scenarios but the performance of GlusterFS decreases considerably when encryption is used as compared to Azure Blob storage.

ACKNOWLEDGMENT

The author wishes to thank Mr. Sai Nath Kapoor, Associate Manager for cloud computing and IoT projects for his support.

REFERENCES

- [1] Cloud Computing: Principles and Paradigms edited by Rajkumar Buyya, James Broberg, Andrzej M. Goscinski
- [2] Cloud Computing Bible By Barrie Sosinsky
- [3] Cloud Computing: Principles, Systems and Applications edited by Nikos Antonopoulos, Lee Gillam
- [4] <http://justazure.com/azure-blob-storage-part-one-introduction/>
- [5] <http://www.tecmint.com/introduction-to-glusterfs-file-system-and-installation-on-rhelcentos-and-fedora/>
- [6] SaaS Growth Strategies - Marketing, Customer Acquisition, & Churn Reduction Consulting for SaaS providers by Lincoln Murphy