ITMO University
Erasmus Mundus Master PERCCOM

Course author: Professor Andrey Shevel

Done by: Al-Hussein Hameed Jasim

OpenStack disk storage VS Amazon disk storage
Cluster, grid and cloud computing systems – June 2017

Table of contents

Abstract

This report represents a summary of the information that I collected about the cloud storages in OpenStack and AWS for the seminars on Cluster, Grid and Cloud computing systems. The aim of this report is to show the types of the cloud storages, differences between them and some recommendations for better use.

1. Introduction

Nowadays, companies are tending to use cloud storages to store their data, maintain them and get easy access when needed. As a definition, Cloud storage is a service model in which data is maintained, managed, backed up remotely and made available to users over a network (typically the Internet). Users generally pay for their cloud data storage on a per-consumption, monthly rate. Cloud storages are divided into three types, private, public and hybrid cloud storages. Moreover, there are two different of services, open source cloud storages and company owned ones.

1.1 Private Cloud Storage

Private cloud storage is exactly what the name says. This system is designed for one person or company that is specific to the customer's needs. These types of cloud storage come in two formats: on-premise and externally hosted. Both work well, but primarily for businesses, not individuals, unless the customer is running a smaller home-based company. The customer has more administrative control and can design the system to what he/she wants it to accomplish in the way of business needs. In private cloud storage, the storage provider has infrastructure in the enterprise's data center that is typically managed by the storage provider. Private cloud storage helps resolve the potential for security and performance concerns while still offering the advantages of cloud storage.

1.2 Public Cloud Storage

This is a cloud service that requires little administrative controls and can be accessed online by any anyone the customer authorizes. The customer gets the same security, but don't need to maintain the system as much as he/she would with a private cloud. It doesn't need a rigid integration with the business needs or private storage concerns. Public cloud storage is where the enterprise and storage service provider are separate and there aren't any cloud resources stored in the enterprise's data center. The cloud storage provider fully manages the enterprise's public cloud storage.

3

1.3 Hybrid Cloud Storage

Hybrid clouds offer a combination of private and public clouds. Features could be customized and there is a possibility to insert the applications that meet the needs, as well as the resources. The most important data can be kept on a private cloud, while the less important data can be stored on a public cloud and accessed by a host of people remotely. It is possible to store data in an efficient storage environment, which saves time and money. Hybrid cloud storage is a combination of public and private cloud storage where some critical data resides in the enterprise's private cloud while other data is stored and accessible from a public cloud storage provider.

1.4 Open source cloud storage

A major advantage of using open source software such as Openstack to build an open source cloud storage environment is that users can modify the source code to meet the needs of their individual businesses. Open source also prevents the need to purchase expensive licensing models and tools and vendor lock-in. Other benefits include the ability to use commodity, off the shelf hardware to build the cloud.

Some of the drawbacks to open source cloud storage include finding and paying for the technical expertise that's necessary to modify the source code. Open source cloud storage users have to forgo the comfort of a cloud storage service-level agreement (Cloud storage SLA).

2. Block storage and Object storage

Cloud storages are classified into two parts, Black storages and Object storages.

2.1 Block storage

Block storage is more like the hard drive in the computer in use and function and typically uses to provide virtual machines with an attached storage volume (virtual hard drive) Snapshots and backups of virtual servers. Block storage devices are accessible as volumes and accessed directly by the operating system, they can perform well for a variety of use cases. Good examples for block storage use cases are structured database storage, random read/write loads, and virtual machine file system (VMFS) volumes.

4

2.2 Object storage

Object storage (also referred to as object-based storage) is a general term that refers to the way in which we organize and work with units of storage. Object storage could be Static Web content, data backups and archival images, and multimedia (videos, pictures, or music) files are best stored as objects. It has its own API and supports other APIs as well. Every object contains three things:

1- The data itself, photo, document, etc.

2- An expandable amount of metadata. Defined by whoever creates the object storage.

3- A globally unique identifier which represents an address given to the object in order for the object to be found over a distributed system.

Data stored as objects have already approached the exabyte scale representing trillions of objects. Companies like Amazon (with S3) provide object storage via its public cloud platform at massive scale, while object storage can be implemented in the company data center using technology like OpenStack's Swift.

2.3 Differences between Object and Block storages

Obviously, object and block storage have very different characteristics, making them suited to storing different types of data.

With block storage, files are split into evenly sized blocks of data, each with its own address but with no additional information (metadata) to provide more contexts for what that block of data is. Block storage exists in the majority of enterprise workloads; it has a wide variety of uses.
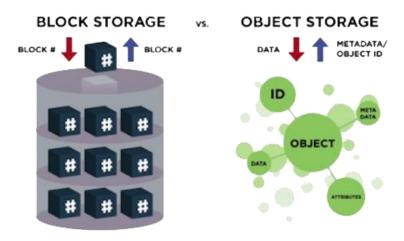

Figure1: Block storage VS Object storage

Object storage, by contrast, doesn't split files up into raw blocks of data. Instead, entire clumps of data are stored in, an object that contains the data, metadata, and the unique identifier. There is no limit on the type or amount of metadata, which makes object storage powerful and customizable. Metadata can include anything from the security classification of the file within the object to the importance of the application associated with the information. Table 1, shows a comparison between block and object storages in different context.

	Block storage	Object storage
Use	Add persistent storage to a VM	Store data
Accessed through	Operating system kernel	REST API
Accessible from	Within a VM	Anywhere via HTTP(S)
Managed by	Cinder	Swift
Persists until	Deleted by user	Deleted by user
Sizing determined by	Specified by user in initial request	Amount of available physical storage
Example of typical usage	1 TB disk	10s of TBs of dataset storage

Table 1: Comparison between Block and Object storages

3. Openstack

OpenStack is a set of software tools for building and managing cloud computing platforms for public and private clouds. Backed by some of the biggest companies in software development and hosting. OpenStack is managed by the Openstack foundation, a non-profit that oversees both development and community-building around the project. Openstack considered as a cloud toolkit and it's open source software.

OpenStack lets users deploy virtual machines and other instances that handle different tasks for managing a cloud environment on the fly. It makes horizontal scaling easy, which means that tasks that benefit from running concurrently can easily serve more or fewer users on the fly by just spinning up more instances.

In order to try OpenStack, "TryStack" lets the user tests the applications in a sandbox environment to better understand how OpenStack works and whether it is the right solution.

Currently, OpenStack storage comprises block and object storage options while an OpenStack file system project is in the works. There are, however, still specifications and compatibility concerns to be aware.

3.1 Block storage in Openstack "Cinder"

Cinder is a block storage component. This more traditional way of accessing files might be important in scenarios in which data access speed is the most important consideration. The Cinder interface specifies a number of discrete functions, including basic functionality such as create volume, delete volume and attach. There are also more advanced functions that can extend volumes, take snapshots, clone and create volumes from a VM image. Synchronous replication used to provide a high level of availability.

The physical storage media, whether disks or solid-state drives, can be located within or directly attached to the Cinder server nodes, or they can be located in external storage systems from third-party vendors. Many storage array suppliers now provide Cinder block device support. These include EMC, Hitachi Data Systems, HP, IBM and NetApp. Most suppliers provide support for iSCSI, with some including Fibre Channel and NFS connectivity.

3.2 Object storage in Openstack "Swift"

Swift is a storage system for objects and files. Swift was designed to be a highly scalable, eventually consistent object store and so it is suited to storing large volumes of data, such as images, media or even files. Rather than the traditional idea of a referring to files by their location on a disk drive, developers can instead refer to a unique identifier referring to the file or piece of information and let OpenStack decide where to store this information. This makes scaling easy, as developers don't have the worry about the capacity on a single system behind the software. It also allows the system, rather than the developer, to worry about how best to make sure that data is backed up in case of the failure of a machine or network connection.

In swift, objects are physically stored on "object servers", one of a number of entities that form a "ring" that also includes proxy servers, container servers and account servers. Objects are stored and retrieved from Swift using simple commands such as PUT or GET, based on the HTTP protocol, also known as a RESTful API. Data resiliency is managed within Swift using the concept of zones. Swift uses the idea of eventual consistency to implement data resiliency which is more risky than using block

7

storage. The downside to eventual consistency is that there is no guarantee that a read request returns the most recent version of the data.

4. Amazon Web Services (AWS)

Amazon Web Services (AWS) offers a broad set of global compute, storage, database, analytics, application, and deployment services that help organizations move faster, lower IT costs, and scale applications

4.1 Block storage in AWS "EBS"

Elastic Block Store (EBS) is a cloud service provided by Amazon. EBS supports a number of advanced storage features, including snapshotting and cloning. EBS is specifically meant for EC2 (Elastic Computing Cloud) instances and is not accessible unless mounted to one. A limitation of EBS is its inability to be used by multiple instances at once. Once it is mounted by an instance, no other instance can use it. EBS can accommodate a smaller amount of data than S3. With EBS, customer pays only for what he uses.

Amazon EBS is recommended when data must be quickly accessible and requires long-term persistence. EBS volumes are particularly well-suited for use as the primary storage for file systems, databases, or for any applications that require fine granular updates and access to raw, unformatted, block-level storage. Regarding the security, EBS volumes could be used as encrypted volumes. For EC2 pricing, to store 10GB there is a need to create a 20GB volume to hold the 10GB file, because resizing disks is not a painless operation. We have to pay for the full size of the EBS volume (20GB) rather than just the file itself. EBS costs $0.05/GB/month, so $1.00/month in EBS storage.

4.2 Object storage in AWS "S3"

Amazon Simple Storage Service (Amazon S3) is object storage with a simple web service interface to store and retrieve any amount of data from anywhere on the web.

Customers use S3 as primary storage for cloud-native applications; as a bulk repository, or data lake for analytics; as a target for backup, recovery, disaster recovery; and with serverless computing. S3 typically experiences write delays while EBS does not

Amazon S3 allows the storage and access of any amount of data at any time, from anywhere on the web, and is designed for high durability. It offers the Standard-

Infrequent Access tier, which has the same durability but at a reduced cost with a 30-day storage minimum. This is ideal for less active data that needs immediate access.

4.3 Archival storage for infrequently accessed data

For data that the customer accesses infrequently, and for which retrieval times of several hours are suitable, customer should choose a service with very high durability and availability at a very low cost.

Amazon Glacier is a low-cost service that provides high-durability storage for archiving and backup for as little as $0.01/GB/month.

5. Recommendations

a- In making the decision to use a particular platform, system architects need to weigh the risks vs. the cost of using OpenStack "free" solutions (that still need hardware) or benefiting from features offered within dedicated hardware.

b- It's better to avoid using object storage for transactional data, especially because of the eventual consistency model outlined previously.

c- As Swift uses the HTTP protocol, it would be perfectly practical to use third-party storage solutions for object storage within OpenStack.

d- OpenStack implementations are usually built around scale-out server configurations, so Fibre Channel is not the best choice of protocol.

URLs

1. http://www.cloudstoragebest.com/cloud-storage-types/

2. http://searchcloudcomputing.techtarget.com/definition/public-cloud

3. https://opensource.com/resources/what-is-openstack

4. https://wiki.openstack.org/wiki/Cinder

5. http://www.computerweekly.com/feature/OpenStack-storage-Cinder-and-Swift-explained

6. http://searchcloudstorage.techtarget.com/essentialguide/Working-with-OpenStack-storage-Tips-on-Cinder-Swift-and-the-cloud

7. http://kernelmastery.com/aws-vs-openstack/

8. http://www.differencebetween.net/technology/internet/difference-between-amazon-s3-and-amazon-ebs/