Google Cloud storage: A journey towards Cloud storage

Kumar Satyam[1]

[1] School of Informatics and Computing, Bloomington, IN 47408, U.S.A.
[*] Corresponding authors: ksatyam@indiana.edu

Now-a-days due to exponential growth of data, it is almost impossible to store the huge amount the data locally in our own systems or local storage boxes which are prone to data loss in case of hardware failure. With cloud storge one can save the data in cloud without any risk of data loss because of redundacy of data over cloud.This paper explains about the Google cloud storage technology which is one of cloud storage vendor.We will discuss what are the key elements of Google cloud storage and try to relate it with the general cloud storage technology.[1].

Keywords: Cloud,Google Cloud Storage, BigQuery, Buckets, Objects, I524

https://github.com/satyamsah/sp17-i524/blob/master/paper1/S17-IR-2031/report.pdf

INTRODUCTION

Google cloud storage is a unified object storage mainly targeting enterprises and developer which offer capability of both live data serving with high I/O throughput to data archiving. As shown in Fig:1, it is an RESTful online file storage web service for accessing data on Google Cloud platform [2]. It is an IAAS cloud delivery model which provides good performance and scalability using Google's cloud platform.

KEY TERMS

- Projects:

  All the data in Google Cloud storage belongs inside a project. A project consist a set of users, set of API and billing, authentication and monitoring setting for those APIs.[3]

- Buckets:

  Buckets are basic containers that hold your data. Everything that you store in Google Cloud Storage must be contained in a bucket. Buckets have bucket name. In Google cloud storage, buckets are created to store your data. A Bucket has three properties,a location is where the bucket and its contents are stored, a globally unique name and default storage class.For example, if you specify the location 'eu'(European union) when you create a bucket A, then bucket A and any objects of bucket A are stored on server of European union region.

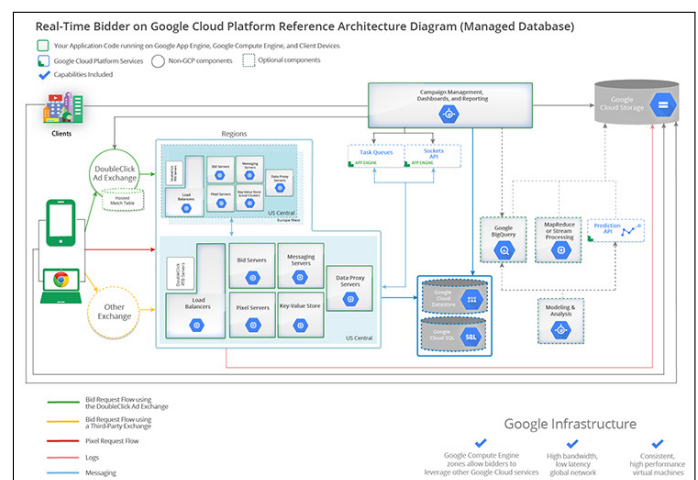## ARCHITECTURE -GOOGLE CLOUD PLATFORM

Fig. 1. Architecture of Google cloud platform using Google cloud storage as persistent layer.Image source : [3]

- Objects:

  Objects are individual pieces of data which are stored in Google cloud storage. A single object can go up to 5 TB in size. Objects have 2 components: Object data and object metadata. The object data component is usually a file that you want to store in Google cloud storage. Object name is just meta data to cloud storage. A common character in the file system is the slash (/). By slash an object name appear as they are stored in hierarchical structure . But Google cloud storage sees the object as independent objects with no hierarchy.

- Data opacity:

  The object's data component is completely opaque to Google cloud storage. It is just a chunk of data to Google Cloud storage.

- Object immutability:

  Objects are immutable, which means that an object cannot change throughout its storage lifetime. An objects storage life time is between the successful object creation and successful object deletion.

- Hierarchy:

  Google cloud storage uses a flat namespace to store objects. However, some tools show it in hierarchical manner.

- Namespace:

  There is only one google cloud storage namespace that means every bucket should have a unique name. Object name must be unique inside a bucket.

STORAGE CLASSES

In Google cloud storage bucket is the place where we can store data.As explained earlier bucket needs be a 'Storage class'. One can think this a service tiers of creating a bucket.There are 4 storage classes offered in cloud cloud storage:

- Multi-regional storage:

  This offers 99.95 % availability with geo-redundant feature. It provides maximum availability of your data even in case of large scale disruption. It is used to store data that is frequently access around the world, such as serving website content. The cost of using this storage class is $0.026 per GB per month. Till now there are three multi – region location

    - asia — Asia Pacific
    - eu — European Union
    - us — United States

- Regional storage:

  If you want high availability and redundancy within a single region you can go for "Regional" storage. This offers 99.95 % availability. The cost of using this storage class is $0.02 per GB per month. It is recommended to the users to use "Regional" storage to utilize durable reduced availability. You can move data from DRA to other storage classes by performing a storage transfer. It is best when you want to optimized latency, availability and, network bandwidth within the same location. It can still be read globally , but if your bucket is used primarily by client outside region , it is

better to switch to multi-region location.It is best for application which require high data analysis and computation.Till now there are three multi – region location :

    - asia-east1 — Eastern Asia-Pacific
    - asia-northeast1 — Northeastern Asia-Pacific
    - europe-west1 — Western Europe
    - us-central1 — Central United States
    - us-east1 — Eastern United States
    - us-west1 — Western United States

- Nearline Storage:

  It is a storage offering which is suited for the archived data which the user access less than once a month .It is low cost, highly durable storage service for infrequently accessed data. It is ideal for back-up and serving long-tail multimedia. The cost of using it is $ 0.01 per GB per month.

- Coldline storage:

  Coldline storage is the storage which is used for the data which the user access less than once a year. Typically this used in case of disaster recovery where it provides low latency to access data stored in Coldline storage ,or for the data that is archived and may or may not be needed at some future time. Cost is using this $ 0.007 per GB per month.

- Standard storage:

  When you create a bucket without specifying storage class it will create a standard bucket which can be either be a multi-regional or a regional storage based on the location you are creating the storage bucket. If you have a standard storage in a multi-regional location like 'us' then it is equivalent to multi regional storage and continue to exist in 'us' location. If the standard storage is in the regional storage like 'us-east', it will be equivalent to 'regional' storage

- Note: Each bucket has a default storage class which one can specify while creating the bucket. The objects also take the default storage unless specified otherwise. You can change the default storage class of your bucket anytime. But the objects which were created in the bucket before changing the storage class will remain in the same(earlier) storage class and will not change to new storage class. It will only affect the storage class of the objects which will be created after the change of the storage class of the bucket. One can specify the storage class of each object which is added to the bucket or can change their name afterwards using the API calls.

ACCESS CONTROL OPTIONS

This takes care of authentication and authorization mechanism. You can control who has access to your Cloud storage buckets and objects as well as what level of access they have.

- Identity and access management(IAM):

  IAM grants access to project's bucket and objects.For example, you can specify that a user has full control of all the objects in your project, but cannot create, modify, or delete any buckets in your projects. IAM gives broad control over the project but not fine-grained control over individual buckets or projects.

- Access control list:

  It is authorization mechanism of Google cloud storage in which we grant read or write access to users for individual buckets or objects.

- Signed URLs:

  This give time limited read-write access to an object through a URL the owner can generate. This can be generates using "gsutil" or by a program. The owner only need to share the url and recipient can access the object through the URL for a specific duration of time which the owner specifies.

- Signed Policy Document:

  This is a document which Specify the type,size and, other upload characteristics of data which can be uploaded and can be used by website owner to allow visitors to upload files to Google file storage

GOOGLE CLOUD PLATFORM CONSOLE

We can use Google cloud Platform Console to perform simple storage management tasks for cloud storage.It requires only a browser to work. Some of the task include: - Enable the Google Cloud storage API for a project - Creating and deleting buckets -Uploading, downloading and deleting objects - Managing ACLs for objects and buckets.

We can use other methods like using the "gsutil" command line tool or any of the client library that supports Google Cloud Storage.

Accessing Google Cloud Platform Console

Google Cloud Platform Console requires no setup or installation and one can access it directly through a browser. There are different ways to access cloud platform console depending on the type of user:

- A project member:

  In order to use Google Cloud Platform Console as a project member, the person account must be added to the list. A current project owner can give the person account privileges like owner, editor, or viewer access to the project which applies to all the buckets defined in the project.

- A user granted read or write access to a bucket:

  In this case , a project owner gives you read or write access to a bucket name.This is useful for the people who are not the project member but need to access the bucket.This URL will prompt you to authenticate with a Google account if you are not already signed in.

- A user is granted read access to an object:

  In this case the project owner configures an object as publicly viewable and sends you the URL to access the object. This URL will prompt you to authenticate with a Google account if you are not already signed in.This is different from the process of "sharing your URL publicly". In the option of "sharing your URL publicly" the URL does not need any authentication

Using the Google Cloud Platform Console to Manage Your Data

An owner can perform various actions which can be performed while accessing Google Cloud storage using console.

- To create a bucket

- Uploading data to a bucket

- Downloading data from a bucket

- Creating and using folders

- Deleting objects, folders, and buckets

- Sharing your data publicly

- Setting bucket permissions

- Setting object permissions and metadata

- Filtering objects to view

- Adding a member to a project

GOOGLE CLOUD STORAGE IN BIGDATA

Google cloud platform and google cloud storage is apt to handle massive tasks working on both the structured and unstructured way.Analytics and machine intelligence at web-scale have been used to solve data intensive tasks.Google BigQuery [4] is cloud platform's managed data warehouse that lets you economically query massive volumes. It is ideal to run Spark and Hadoop.

GOOGLE CLOUD STORAGE COMPETITORS

In the cloud platform business where all the market leader like AWS, Microsoft Azure, IBM Bluemix etc., Google cloud platform is also giving good competition.It is spreading its services by coming up with more regions and availability zones all across the world with cheap rates.

CONCLUSION

This paper has shown some of the key aspects of the cloud storage systems, how it works and how it is an integral part of the Google cloud platform,some core concepts of Google Cloud storage. It also showed how Google cloud storage performing with respect to Big data technology. And how it is performing in comparison to other vendors who are in the market.

REFERENCES

[1] Kun Liua and Long-jiang Dong, "Research on cloud data storage technology and its architecture implementation," Web Page, College of Oriental Application and Technology Beijing Union University, Beijing and Software Development Department Adobe (Beijing) Corporation, Beijing, jan 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877705811065192

[2] Wikipedia, "Google cloud storage," accessed: 02-25-2017. [Online]. Available: https://cloud.google.com/storage/docs

[3] G. Inc., "Google cloud storage," accessed: 02-25-2017. [Online]. Available: https://cloud.google.com/storage/docs

[4] ——, "Google bigquery," accessed: 02-25-2017. [Online]. Available: https://cloud.google.com/bigquery/