Whitepaper: Google Cloud Storage Nearline
Summary

Google Cloud Storage Nearline is a low cost storage class available in Google Cloud Storage. It provides the low cost usually associated with offline storage, while providing immediate access to your data through the same Google Cloud Storage APIs used for standard online storage. Google Cloud Storage Nearline is suitable for workloads traditionally filled by offline storage, such as cold storage and disaster recovery, but without the access penalties typically associated with offline storage solutions.
Online vs. Offline Storage

Online storage systems provide low latency access to the data they store. Offline storage systems do not provide low latency access, but have other advantages that make them a useful part of a larger storage system. Typically one of those advantages is lower cost per byte vs. online storage systems.

In traditional on-premises IT infrastructure, file servers provide online storage. Through a network protocol such as NFS, clients are able to access data with time-to-first-byte measured in seconds or less. In contrast, tape libraries provide offline storage. The time-to-first-byte is usually measured in minutes or hours, or perhaps even days if the data is on a tape in an offsite location and needs to be retrieved. Because of this high latency, normal network access protocols such as NFS cannot typically be used to access offline storage, so a different API or protocol is used, and different tools are often required. Despite these inconveniences, offline storage remains an important part of an overall IT infrastructure for infrequently accessed data because of its significantly lower...
cost. Cold storage and disaster recovery scenarios are often implemented using offline storage.

In cloud infrastructure, online storage is provided by object storage services such as Google Cloud Storage. These services are similar to traditional file servers, but are generally accessed via a different network protocol that is more suitable to the massive scale at which they operate.

Cloud storage services exist that provide offline storage, where the time-to-first-byte is measured in hours. Similar to tape libraries, these offline storage services are accessed using network protocols that are significantly different than online cloud storage services, and therefore require different tools. However, because they can provide storage at a significantly lower cost than online storage services, they are attractive for the same scenarios that tape libraries are used for in a traditional IT infrastructure: cold storage and disaster recovery.
A Third Option: Nearline Storage

Google Cloud Storage Nearline challenges the conventional wisdom that a service must provide only offline storage, with all the associated inconvenience, to provide storage at a price that makes it attractive for cold storage and disaster recovery purposes. Google Cloud Storage Nearline provides the convenience of online storage at the price of offline storage.

The Standard Storage class in Google Cloud Storage is optimized for high availability and low latency. This makes it ideal for scenarios such as serving content to end users, where every millisecond matters.

The Nearline Storage class in Google Cloud Storage is intended for infrequently accessed data, such as disaster recovery and cold storage scenarios. These scenarios are also more tolerant of slightly lower availability and slightly higher latency than online “serving” scenarios. The combination of a-priori knowledge that the data will be infrequently
accessed, and slightly relaxing the availability and latency expectations means that storage can be provided at a significantly lower cost per GB per month, allowing it to be offered at a price that is competitive with offline storage services, but without the associated inconveniences and long delays in accessing the data.

When it comes to the reliability of a storage system, there are two factors that are generally discussed: availability and durability. Loosely speaking, availability is the probability that you can access your data right now, and durability is the probability that you will be able to access your data eventually. Your data can be durable (i.e. not lost) while not currently being available (i.e. you can't access it right now).

Cold storage and disaster recovery scenarios can tolerate lower availability than online serving scenarios where an end user will receive an error if the data is not immediately available. However, cold storage and disaster recovery scenarios still demand high durability, and therefore data stored in the Nearline Storage class is just as durable as data stored in the Standard storage class.

Nearline: Convenience Two Ways

The previous section states that Google Cloud Storage Nearline provides the convenience of online storage at the price of offline storage. This convenience comes from two factors:

1. Your data is available as soon as you realize you need it.
2. The tools, processes, and APIs you already use to manage your online storage do not change for Nearline storage.

The benefit of having access to your data the moment you realize you need it is quite clear. For cold storage scenarios, your workflow is not interrupted, so your productivity remains high. For disaster recovery scenarios, you are back up and running hours sooner than you would be with an offline storage solution.

However, the benefit of using the exact same tools you already use today to access archived data is just as significant. No engineering effort needs to be spent learning new tools and processes. No code needs to be rewritten to use different APIs.

What does this look like in practice? You can use the gsutil command line tool to access your Nearline storage in exactly the same way you use it for your Standard storage.
Cloud Storage Browser web UI available in the Google Developers Console also works unchanged. The same authentication and authorization (ACLs) that you are already using for Standard data will work unchanged for Nearline data. Even advanced features such as Object Versioning and Access Logs work exactly the same way for Nearline storage.

**Cold Storage vs. Disaster Recovery**

Cold storage and disaster recovery are both considered “backup” or “archival” storage scenarios. In both scenarios, data is accessed infrequently after initially being written, and data reads do not require the same extremely high availability and low latency as online serving scenarios. In this way, the scenarios are very similar.

However, cold storage and disaster recovery also differ in a very important way. In a disaster recovery scenario, when you need to read the archived data, you likely need to read all of it, and you need to read it quickly so you’re back up and running as quickly as possible.

As a concrete example, perhaps your business runs servers on premises, and your plan in the case of a natural disaster, fire, or other catastrophe is to restore backups of those servers stored in Google Cloud Storage to on-demand instances in Google Compute Engine. Your time to recovery will be affected by how quickly you can read all the backups of all the affected servers, so it is important to have high throughput available from Google Cloud Storage Nearline (where your backup data is stored) and the Google Compute Engine zone where your on-demand instances are started.

In contrast, cold storage is simply archiving data that you might need someday, or are legally obligated to retain, but the chances of having to read any particular piece of it is low. When you do need to read some of it, you probably only need to read a small fraction of what you’ve archived.

Perhaps you are a media company, and you want to archive all the digital assets that went into creating a particular program. At some point in the future, while producing a future program, you want to reuse some of the graphics that were produced for the previous program. With Google Cloud Storage Nearline, you can simply use the Cloud Storage Browser, gsutil, or a third party cloud storage browsing tool to copy the relevant objects to your working area, and continue with your work. Note how in this scenario, there is great value to having immediate access to the data, and familiar tools to work with it. If you had to submit a job to an (unfamiliar) offline storage service to
retrieve your data and wait several hours for it to become available, half a work day might be wasted. Also, you may not know exactly which object you want, so being able to easily browse and view the objects is critical.

**Continental vs. Regional Storage**

Each bucket in Google Cloud Storage has a Location property which allows you to control where your data is stored. Location can be specified at two different levels of granularity: Continental or Regional.

The current options for Continental storage are United States (US), European Union (EU), and Asia (ASIA).

Regional location is more fine grained. It allows you to specify that your data should be stored in a particular Google Compute Engine region.

*Note: Regional buckets are, as of the date of this whitepaper, still an Alpha feature. Support for regional buckets using the Nearline Storage class may or may not be supported at the time you are reading this. Check [https://cloud.google.com/storage/docs/regional-buckets](https://cloud.google.com/storage/docs/regional-buckets) for the most up to date status of regional buckets, including which regions and which storage classes are available.*

For a disaster recovery scenario, where you intend to bring up replacement servers in a particular Compute Engine region, it is a good idea to store your backups in a bucket with the Location set to that region. This will provide you with the best network throughput between your Cloud Storage bucket and the Compute Engine instances you are restoring your data to, minimizing your time to recovery.

For cold storage scenarios like the media company example described earlier, you may not be restoring your data to a Compute Engine instance. In this case, your read throughput is usually limited by your connection to the Internet, so you will not notice any better throughput from a regional location vs. a continental location. In fact, choosing a continental location gives Google more flexibility on where your data can be stored, which could improve performance, particularly if you read your data from multiple places (e.g., you have offices in New York and Los Angeles).

Therefore, as a general guideline, we recommend using regional locations for disaster recovery, and continental locations for cold storage.
On Demand I/O

The speed with which you can read data from Google Cloud Storage Nearline scales with the amount of data you store. The more you store, the higher read throughput you can expect. You should expect 4 MB/s of read throughput per TB of data stored. This means that, by default, you can read every byte you have stored in Cloud Storage Nearline in about three days.

For cold storage scenarios, the default throughput provided by Cloud Storage Nearline is likely completely adequate, because you generally only need to read a small subset of your data at any particular time. However, for some disaster recovery scenarios, you may want to read a large portion of your data over a short period of time, which may not be possible with the default throughput of 4 MB/s per TB of data stored. For these situations, Google offers On Demand I/O.

Enabling On Demand I/O tells Cloud Storage Nearline that you wish to exceed the default level of throughput offered for the amount of Nearline storage that you are using. It can be enabled in the Google Developers Console, in the Settings area of Cloud Storage (i.e. under Storage -> Cloud Storage -> Settings in the left navigation bar).

Of course, providing higher than normal throughput requires more resources, and thus there is a cost associated with enabling this feature. However, at the time this whitepaper is being written, On Demand I/O is temporarily being offered for no additional charge. This may have changed by the time you are reading this whitepaper, so you should check the latest pricing for Google Cloud Storage Nearline (available at https://cloud.google.com/storage/pricing) before enabling this feature.

On Demand I/O is provided on a best effort basis. Enabling On Demand I/O does not guarantee that you will see an increase in throughput. What it does is indicate to Google that you would like the highest throughput we can provide at that point in time, and are willing to be billed for the cost of that increased throughput, as described in the previous paragraph.

Migrating to Google Cloud Storage Nearline

Google offers Cloud Storage Transfer Service that can be used to migrate petabyte scale data from other online storage services to Google Cloud Storage. For data that is already
being frequently accessed, this can be an excellent way to completely migrate to Google Cloud Storage.

Cloud Storage Transfer Service can also be used to transfer petabyte scale data between locations within Google Cloud Storage (e.g. from a US bucket to an EU bucket), as well as between storage classes (e.g. from a Standard Storage class bucket to a Nearline Storage class bucket). Storage transfers can be configured to run periodically, and can be configured to transfer objects that meet certain criteria. They can also delete the original objects after they have been transferred. For example, you could configure a rule that transfers objects from a Standard Storage class bucket to a Nearline Storage class bucket when the objects are 30 days old, and deletes the original Standard object after the transfer, thus automatically “aging out” objects from the Standard Storage class to the Nearline Storage class.

Note that offline storage services designed for disaster recovery and cold storage scenarios are built with the assumption that the data they store will be infrequently accessed, and often it is expensive to read large volumes of data from these offline storage services. The specifics will vary from service to service, but if you are currently using Amazon Glacier, we have written a whitepaper that provides some best practices for migrating your data to Google Cloud Storage Nearline.

**Conclusion**

In this paper we have examined how Google Cloud Storage Nearline is a faster, more convenient archival storage solution when compared with existing offline storage services. It is a great fit for disaster recovery and cold storage scenarios. However, viewing Google Cloud Storage Nearline as only a replacement for slower and less convenient offline storage services is potentially missing a large opportunity, because fundamentally, Google Cloud Storage Nearline is a new layer of the storage hierarchy - one that perhaps you can use in new and innovative ways.

The classic storage hierarchy used to have three main layers - main memory (DRAM), magnetic disk, and magnetic tape. Decades of research and engineering were spent optimizing where to store which bytes, and in what format, to provide the best price/performance for various applications. Then solid state disks (SSDs) appeared, adding a new layer of storage in between main memory and magnetic disk, and suddenly many of the old assumptions were no longer valid, causing a wave of innovation in how applications are built, and what is possible.
Similarly, in the cloud world, developers have become accustomed to the notion that you can have fast “online” storage, suitable for frequently accessed “hot” data, or you can have inexpensive “offline” storage, suitable for almost never accessed “cold” data, but nothing in between. Nearline storage bridges this gap, providing a fundamentally new type of storage in the cloud universe, suitable for data that won’t be accessed frequently, but where there is value in having that data immediately available.

If you’re currently paying a premium to store data online, but are accessing it less than approximately once per month, and do not require sub-second latency when you do, then moving that data to Google Cloud Storage Nearline can lower your storage cost. However, perhaps an even more interesting exercise is to examine the data you are currently storing offline, and ask if there is a way to extract more value from that data by storing it in a more immediately accessible form. Lowering storage costs is great, but realizing previously untapped value from your data could have even higher impact on your business.