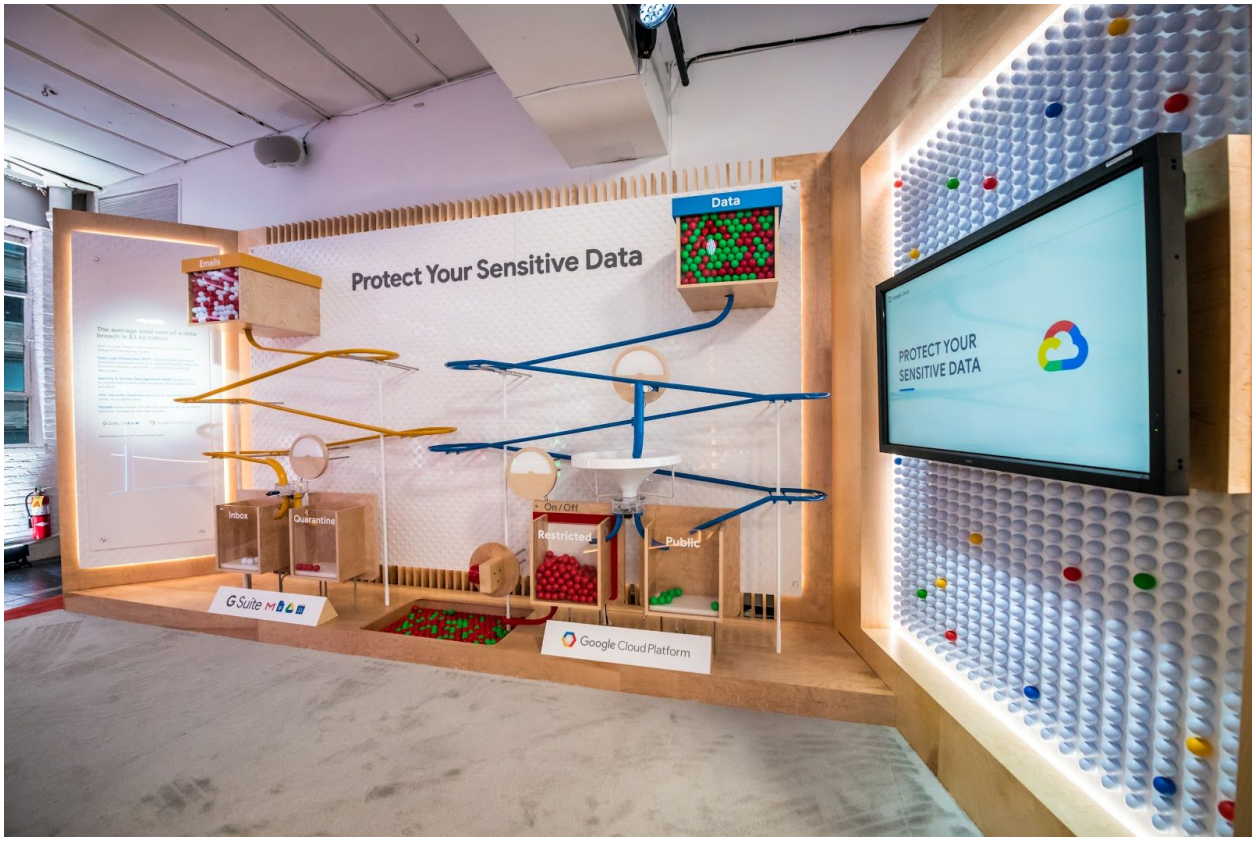
Data Management Solutions in the Cloud

Accessing, sharing, and processing data that's new to Google Cloud Platform (GCP).



For more information visit cloud.google.com/solutions/data-management

Contents

1. Introduction

The National Institutes of Health (NIH) established The Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) initiative to provide biomedical researchers with access to advanced, cost-effective, cloud-based computational infrastructure, tools, and services. Through STRIDES, researchers can take advantage of emerging data management methodologies, technological expertise, computational platforms, and tools to support cutting-edge experimentation and innovation. NIH has partnered with Google Cloud to support the STRIDES initiative through cloud services. In support of STRIDES, we've developed sets of playbooks to help enable researchers to build healthcare and life sciences solutions on Google Cloud Platform (GCP).

The goal of this playbook is to aid researchers as they transition historically on-premise datasets, workloads and pipelines to GCP. This playbook will provide researchers with methods for managing data lifecycles, accessing public datasets, leveraging cloud APIs and API gateways, managing data costs, and sharing and visualizing data. Additionally, this playbook will outline training and digital resources to help upskill and enable researchers to build on Google Cloud, while highlighting the appropriate products and services to use when architecting on GCP.

2. Learning

Generally, cloud adopters fall under one of three categories:

Cloud Novice	Cloud Ready	Cloud Native
Little to no understanding of the cloud	Familiar with the cloud, some experience	Lots of cloud experience, expert-level knowledge

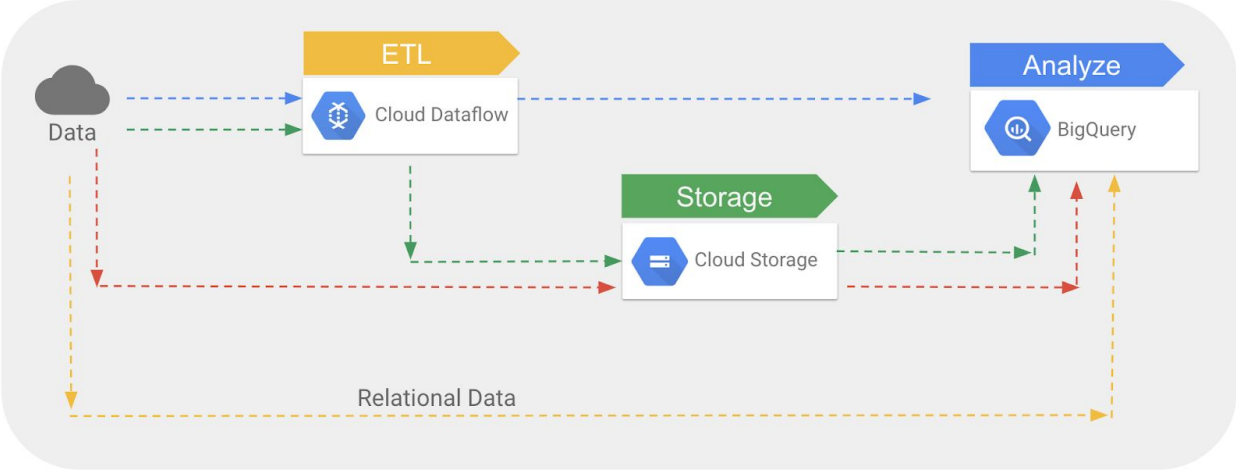
Understanding this broad spectrum of experience levels, we've highlighted key training resources to help upskill researchers on Google Cloud. Additionally, Google offers on-site, instructor-led training to enable large groups of participants across your organization.

Cloud Novice		
Video: Welcome to GCP Video: Intro to GCP for Students Video: GCP 101 Video: GCP Essentials	Documentation: GCP Conceptual Overview Documentation: About GCP Services Documentation: GCP Development & Admin Tools	Documentation: All GCP Products & Services Virtual Course: GCP Fundamentals - Core Infrastructure Virtual Lab: GCP Essentials

Cloud Ready		
Documentation: All GCP Products & Services Documentation: GCP for Data Center Professionals Documentation: GCP for AWS Professionals Documentation: GCP for Azure Professionals Documentation: GCP for OpenStack Users Video: Data Discovery in Google Cloud Video: Managing Encryption of Data in the Cloud	Virtual Course: Data Analytics with Google Cloud Virtual Course: Essential Cloud Infrastructure - Foundation Virtual Course: Essential Cloud Infrastructure - Core Services Virtual Course: Data Analytics with Google Cloud Virtual Course: Essential Cloud Infrastructure - Foundation	Virtual Course: Essential Cloud Infrastructure - Core Services Virtual Course: From Data to Insights with GCP Virtual Course: Exploring and Preparing your Data with BigQuery Virtual Lab: Cloud Architecture Virtual Lab: Cloud Engineering VIrtual Lab: Cloud Development

Cloud Native		
Documentation: All GCP Products & Services Documentation: GCP Solutions Documentation: Big data analytics Video: Tools for Migrating Your Databases to Google Cloud Video: Data Warehousing with BigQuery: Best Practices Video: Easily Prepare Data for Analysis with GCP Video: Analyzing Big Data in less time with Google BigQuery Video: Choosing your storage and databases on GCP	Video: Sensitive data management for collaborative research clouds Video: understanding and managing metadata Virtual Course: Architecting with Google Cloud Platform Virtual Course: Big Data and Machine Learning Fundamentals Virtual Course: Leveraging Unstructured Data with Cloud Dataproc on Google Cloud Platform Virtual Course: Serverless Data Analysis with Google	BigQuery and Cloud Dataflow Virtual Course: Serverless Machine Learning with Tensorflow on Google Cloud Platform Virtual Course: Building Resilient Streaming Systems on Google Cloud Platform Virtual Lab: Data Engineering Virtual Lab: Data Science on GCP Virtual Lab: Data Science on GCP - Machine Learning Virtual Lab: Google Cloud Solutions II - Data and Machine Learning

3. Data Lifecycle Management



With Google Cloud services, you can manage data throughout its lifecycle, from ingestion and storage to processing and visualization.

Ingest raw stream, batch or application data into GCP with products such as Google App Engine, Google Compute Engine, Google Kubernetes Engine, Cloud Pub/Sub, Cloud Transfer Service or Transfer Appliance.

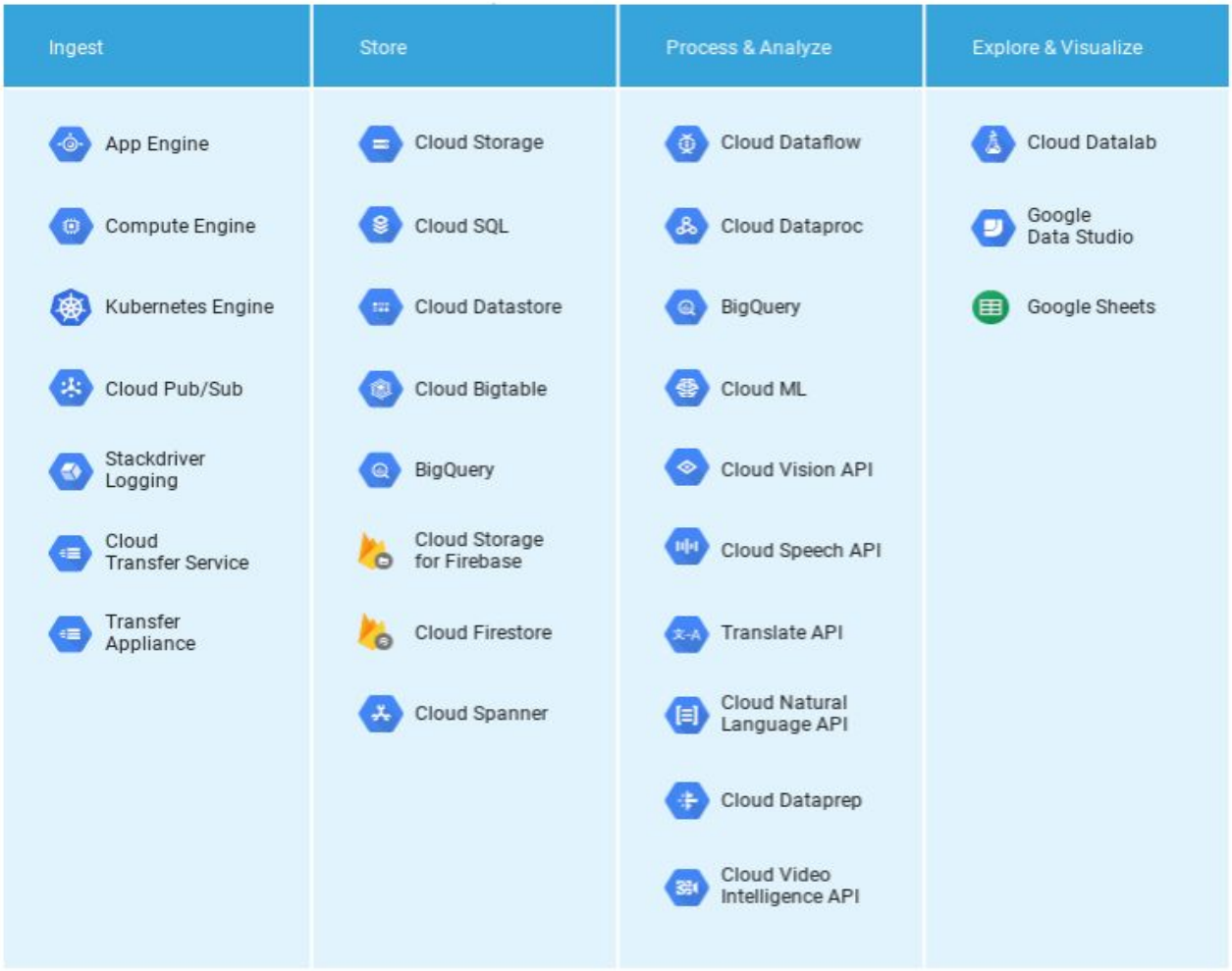
Store retrieved data for easy access and processing using Google Cloud Storage, Cloud SQL, Cloud Datastore, Cloud Bigtable, Cloud Firestore, Cloud Storage for Firebase, or Cloud Spanner.

Process and analyze data with tools like Cloud Dataflow, Cloud Dataproc, Cloud ML, Vision API, Speech API, NLP API, Cloud Dataprep, or Video Intelligence API.

Archive under-used data for future disaster recovery using Archival Cloud Storage, where Nearline, Coldline, and Archive offer ultra low-cost, highly-durable, highly available archival storage.

Visualize data using Cloud Datalab, Data Studio, or Google Sheets.

Learn more about how data management solutions and data life-cycle tools on GCP.



4. Data Management Solutions

Google Cloud offers a wide range of tools, solutions, and best practices for accessing, processing, and sharing data in the cloud. In this section we'll outline common use cases and mechanisms for data management in GCP, allowing researchers and end users to build rich collaboration ecosystems.

4.1 Public Datasets

Google Cloud Public Datasets facilitate access to high-demand public datasets, making it easy for you to conduct research and uncover new insights in the cloud. By analyzing these datasets hosted in BigQuery and Cloud Storage, you can seamlessly experience the full value of Google Cloud with ease.

▶ What are Cloud Public Datasets

Google Cloud public datasets provide a playground for those new to big data and data analysis and offer a powerful data repository of more than 100 public datasets from different industry verticals, allowing you to join these datasets with your own, to produce new insights. We provide **free storage for all public datasets** and customers ***can access up to 1TB of data/month at no cost.***

▶ How can researchers access public datasets

Which method you choose to access public data depends on how you want to work with the data. When accessing public data via the Google Cloud Console, you must authenticate with Google.

By contrast, accessing public data with gsutil or a Cloud Storage API link does not require authentication. These methods are suited for general-purpose links to publicly shared data.

▶ Benefits of using public datasets

Seamlessly access and analyze data in the cloud - Google Cloud public datasets simplify the process of getting started with analysis because all your data is in one platform and can be accessed instantly.

Sharing and hosting your data publicly on GCP will benefit your research in that it brings your data a much greater visibility and impact when other researchers can access and incorporate it into their projects. By making your data available for other researchers to use, you enable your datasets to be cited similarly to other research publication types (such as articles or books), thereby opening up more opportunities to gain academic credit for your work.

The motivations behind publishing data may range for a desire to make research more accessible, to enable citability of datasets. On the other hand, Google hosts public data sets for free through Google's Public Datasets Program, thus hosting your petabyte scale data on Google Public Dataset is the most cost-efficient way to share your data.

The larger scientific community also benefits: sharing data encourages multiple perspectives, helps to identify errors, discourages fraud, is useful for training new researchers, and increases efficient use of funding and population resources by avoiding duplicate data collection.

With Google Cloud Public Datasets, you can access the same products and resources enterprises use to run their businesses. Query data directly in BigQuery and leverage its blazing fast speeds, querying capacity and easy to use, familiar interface. You can also access ML-ready datasets leveraging GCP's machine learning capabilities such as Auto ML, Vision API and BigQuery ML (BQML) to gain additional insights for your research.

4.2 Data Storage

Cloud Storage provides worldwide, highly durable object storage that scales to exabytes of data. You can access data instantly from any storage class, integrate storage into your applications with a single unified API, and easily optimize price and performance.

▶ Overview of Storage Classes

Storage classes determine the availability and pricing model that apply to the data you store in Cloud Storage. The storage class you set for an object affects the object's availability and pricing model.

The purpose of Storage classes is to provide you with a more cost-efficient way to store your research data. You can classify your more hot and active research data from the archival ones, and apply different storage costs to them. The following tables summarize the cost for primary storage classes offered by Cloud Storage and their cost-performance. See class descriptions for a complete discussion.

As illustrated in the graphs, Standard Storage costs the most from just the Storage cost perspective, because it defaults to store your data multi-regionally, and it is optimized for performance and high-frequency access. On the other hand, there are no restrictions like data retrieval fee or minimum storage duration apply. You might want to store the research data that involves immediate analysis or application in Standard.

US (multi-region)

Standard Storage (per GB per Month)	Nearline Storage (per GB per Month)	Coldline Storage (per GB per Month)	Archive Storage (per GB per Month)
$0.026	$0.010	$0.007	$0.004

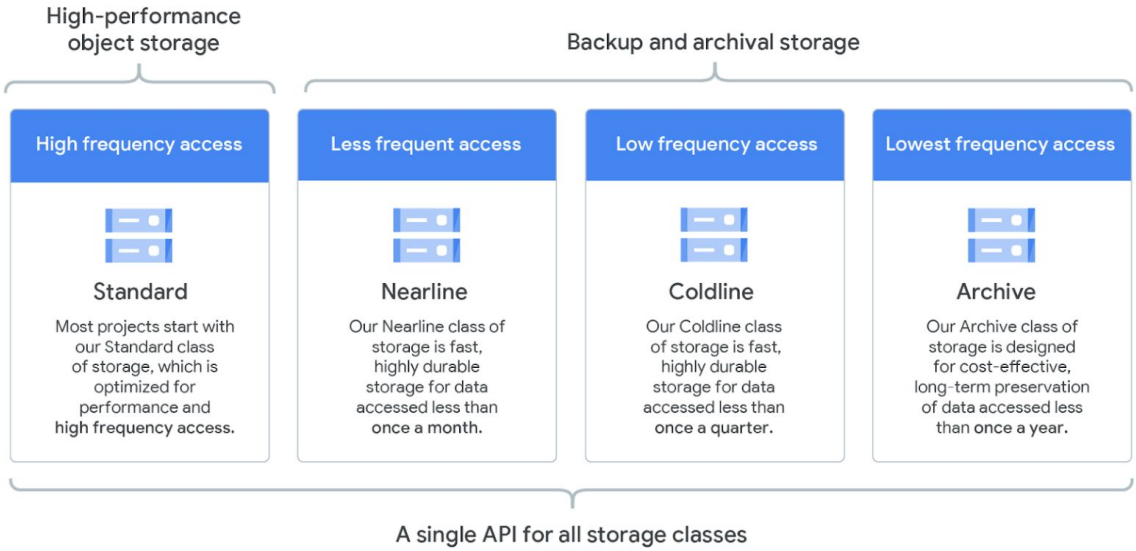
	Nearline Storage	Coldline Storage	Archive Storage
Data retrieval	$0.01 per GB	$0.02 per GB	$0.05 per GB
Minimum storage duration	30 days	90 days	365 days

You might want to consider **storing the less frequently-accessed** data in Nearline, Coldline and Archive Storage, because they have much cheaper storage costs. For example, the large volume patient records and medical images related to a genomic project that you already accomplished. However, there is a data retrieval fee and minimum storage duration related to those cheaper storage options.

Before choosing the storage class, have a rough idea in mind about your usage, retrieval and duration of needs for that data, then do the math to calculate which storage option fits best with your purpose!

The following aspects apply to **all storage classes**:

- Unlimited storage with no minimum object size.
- Worldwide accessibility and worldwide storage locations.
- Low latency (time to first byte typically tens of milliseconds).
- High durability (99.999999999% annual durability).
- Geo-redundancy if the data is stored in a multi-region or dual-region.
- A uniform experience with Cloud Storage features, security, tools, and APIs.



▶ Breakdown of cloud storage costs

Cloud Storage pricing is based on the following components:

- Data storage: storing data in buckets.
- Network usage: accessing and moving data in buckets.
- Operations usage: performing actions within Cloud Storage.
- Retrieval and early deletion fees: applicable for data stored in the Nearline Storage, Coldline Storage, and Archive Storage classes.

▶ GCS pricing example

The following example shows a simple scenario that might apply if you are hosting a PetaByte-scale dataset in Cloud Storage. The data storage amount is the average amount of data in your bucket over the course of the month. Suppose you have the following storage usage pattern in a given month:

Pricing Category	Type of Usage	Amount
Data storage	Standard Storage in a multi-region	5 PB
Network	Egress to the Americas and EMEA	1 PB
Operations	Class A operations (object adds, bucket and object listings)	100,000 operations
Operations	Class B operations (object gets, retrieving bucket and object metadata)	500,000 operations

Your bill for the month is calculated as follows:

Pricing Category	Type of Usage	Cost
Data storage	5 PB Standard Storage * $2.6 * 10^4 per PB	$ 130,000
Network	1 PB egress * $1.2 * 10^5 per PB	$ 120,000
Operations	10,000,000 Class A operations * $5 per 1,000,000 operations	$ 50
Operations	50,000,000 Class B operations * $0.4 per 1,000,000 operations	$ 20
Total		$ 250,070

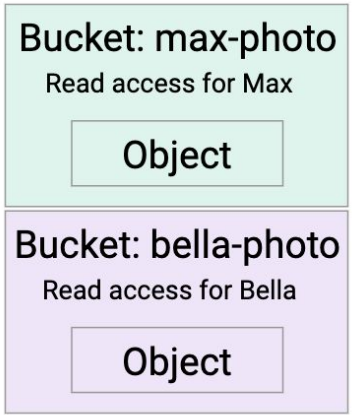
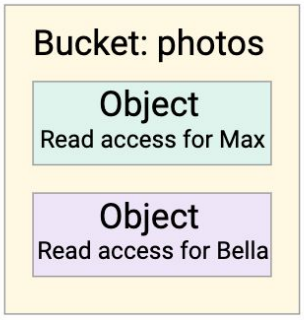
For storage usage that includes multiple storage classes as well as bandwidth consumption that spans multiple tiers, check out this detailed pricing example.

▶ Controlling access to storage

Cloud Storage is implemented through Cloud Storage Buckets. You can think of buckets as a hard drive on-prem. You control who has access to your Cloud Storage buckets and objects and what level of access they have, using Cloud IAM or ACLs, details explained below. When you create a bucket, you should decide whether you want to apply permissions using uniform or fine-grained access.

- Uniform (recommended): Uniform bucket-level access allows you to use Cloud Identity and Access Management (Cloud IAM) alone to manage permissions. Cloud IAM applies permissions to all the objects contained inside the bucket.
- Fine-grained: The fine-grained option enables you to use Cloud IAM and Access Control Lists (ACLs) together to manage permissions. With ACL, you can specify access and apply permissions at both the bucket level and per individual object.

If you have objects that contain sensitive data, we recommend storing that data in a bucket with uniform access enabled to streamline permissions. For example:

Recommended	Not recommended
Access control: uniform	Access control: fine-grained
	
This setup has a lower chance of data exposure. Adding permissions at the bucket level ensures that Max and Bella cannot see each other's data, even if new files are added to the buckets.	This setup has a higher chance of data exposure. If you do not set object permissions correctly, Max and Bella may be able to see each other's photos, as well as new files added to the bucket.

ACLs control permissioning only for Cloud Storage resources and have limited permission options, but allow you to grant permissions per individual objects. You most likely want to use ACLs for the following use cases:

- Customize access to individual objects within a bucket.
- Migrate data from Amazon S3.

▶ Implementing object lifecycles to reduce storage costs long-term

Using Object Lifecycle is beneficial for researchers because it allows you to manage data cost in a flexible way. For example, if you have a petabyte-scale collection of past patients' medical images. If you are constantly adding new images, the old ones may only be involved in research for a period of time before they become obsolete. In that case, you may not want to spend as much in storing the obsolete ones because you are not retrieving and running analytics on them as often. Thus moving them to a more cost-efficient storage class comes as a reasonable move. You can categorize and store images in buckets, and assign a lifecycle management configuration to every bucket. The configuration contains a set of rules which apply to current and future objects in the bucket.

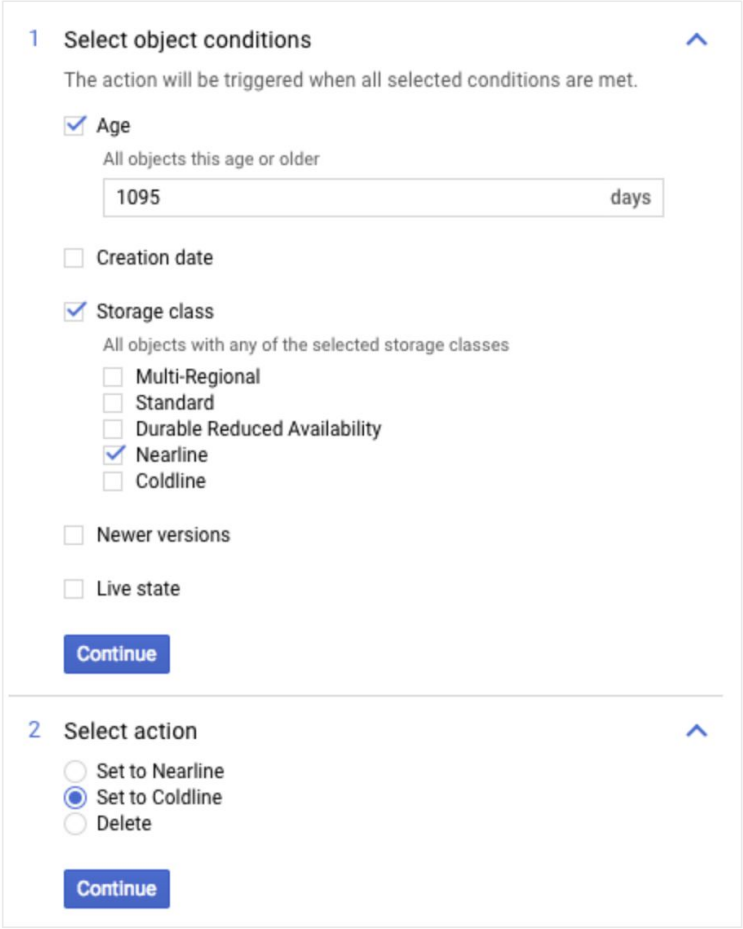
The following actions are supported for a lifecycle rule:

- Delete: Delete objects. Unless In buckets with object versioning enabled, once an object is deleted, it cannot be undeleted.
- SetStorageClass: Change the storage class of objects.

Here are some example use cases:

- Downgrade the storage class of objects older than 365 days to Coldline Storage.
- Delete objects created before January 1, 2013.
- Keep only the 3 most recent versions of each object in a bucket with versioning enabled.

To learn how to enable Object Lifecycle Management, and for examples of lifecycle policies, see Managing Lifecycles.



4.3 Data Discovery

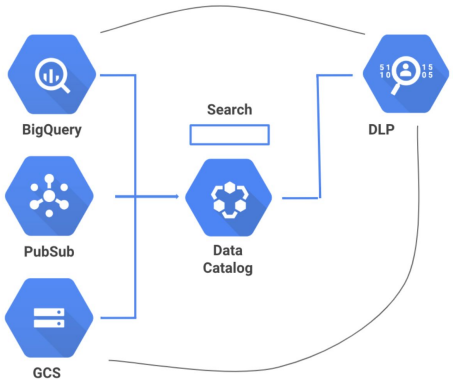
Big Data solutions on GCP help you efficiently capture, process, and analyze data at petabyte scale. Google Cloud Platform's fully managed, proven, end-to-end data analytics products remove the operational complexities of data analytics with a serverless approach to unlock important discovery quickly and efficiently.

► Managing data assets at scale with Cloud Data Catalog

Managing data assets can be strenuous without the right tools. Data Catalog provides a centralized place where organizations can find, curate and describe their data assets. It is a fully managed, scalable metadata management service in Google Cloud's Data Analytics family of products. Data Catalog can catalog the native metadata on data assets from the following Google Cloud storage system sources:

- BigQuery datasets, tables, and views
- Pub/Sub topics

With Data Catalog, you can:



- Determine the set of metadata attributes that you will capture in order to fulfill your business or regulatory needs.

- Create templates/tags to capture this business metadata. Data Catalog supports five data types that you can combine to create rich tags: double, string, boolean, datetime, and enum.

- Discover an asset through a robust search functionality that uses both technical and business metadata to return relevant results. For example, you can apply filters like data asset type (tag template, dataset, datastream, fileset or table) keywords and tags when you try to search for the dataset of demographic distribution of a virus infected town in a specific month.

- Use the Data Catalog API to record the relevant lineage metadata on each relevant step of your data pipeline.

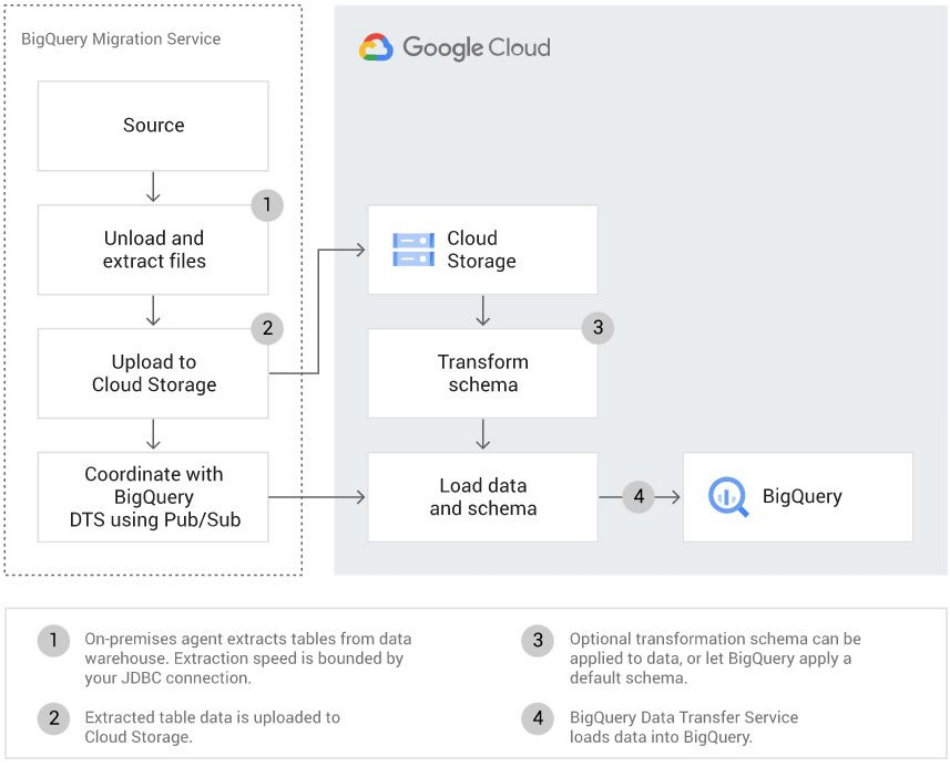
The following diagram shows a sample customer table (cust_tbl) and several business metadata tags attached to it and to its columns:

▶ BigQuery: Jump-start data analysis and uncover meaningful insights

BigQuery is Google's fully managed, petabyte scale, low cost analytics data warehouse. BigQuery is NoOps—there is no infrastructure to manage and you don't need a database administrator—so you can focus on analyzing data to find meaningful insights, use familiar SQL, and take advantage of our pay-as-you-go model.

To transform your research at ease, BigQuery enables you to:

- Quickly analyze gigabytes to petabytes of data using ANSI SQL at blazing-fast speeds, with zero operational overhead
- Efficiently run analytics at scale with a 26%-34% lower three-year TCO than cloud data warehouse alternatives
- Get seamlessly democratize insights with a trusted and more secure platform that scales with your needs



At the Stanford Center for Genomics and Personalized Medicine (SCGPM), researchers using GCP and BigQuery can now run hundreds of genomes through a variant analysis pipeline and get query results quickly. Mike Snyder, director of SCGPM, notes, "We're entering an era where people are working with thousands or tens of thousands or even million genome projects, and you're never going to do that on a local cluster very easily. Cloud computing is where the field is going."

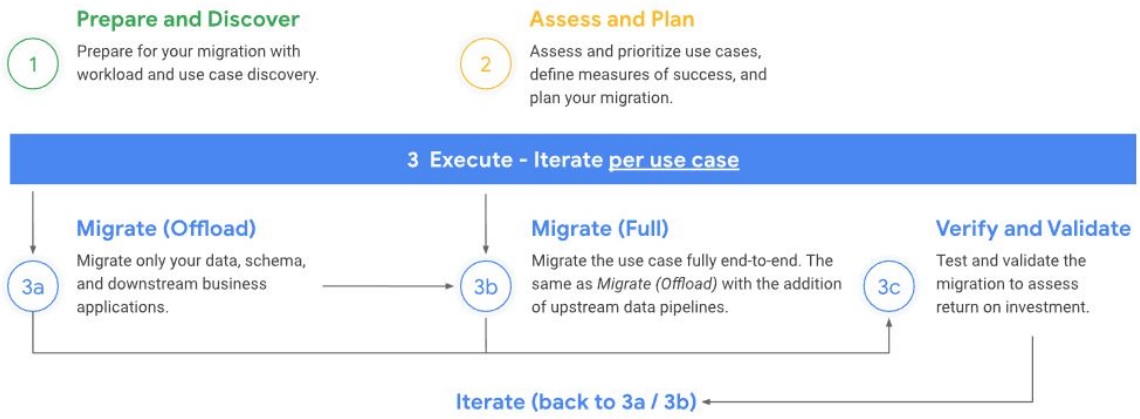
To transit from on-prem to BigQuery, here are our advice:

For scientific researchers who are new to Google Cloud, one of the most common use cases is migrating your data warehouses to BigQuery. Undertaking a migration can be a complex and lengthy endeavor. Therefore, we recommend adhering to a framework to organize and structure the migration work in phases:

1. Prepare and discover: Prepare for your migration with workload and use case discovery.
2. Assess and plan: Assess and prioritize use cases, define measures of success, and plan your migration.

3. Execute: Iterate the following steps for each use case:
   a. Migrate (offload): Migrate only your data, schema, and downstream research applications.
   b. Migrate (full): Alternatively, migrate the use case fully end-to-end. The same as Migrate (offload), with the addition of the upstream data pipelines.
   c. Verify and validate: Test and validate the migration to assess return on investment.

The following diagram illustrates the recommended framework and shows how the different phases are connected:



▶ Enabling advanced insights using Cloud Life Sciences

Cloud Life Sciences is a suite of services and tools for processing, analyzing, and annotating genomics and biomedical data at scale.It also enables advanced insights and operational workflows using highly scalable and compliant infrastructure. Cloud Life Sciences includes features such as the Cloud Life Sciences API and extract-transform-load (ETL) tools, and more.

Key capabilities:

- **Analyzing variants**
  When you export your projects from Cloud Life Sciences into BigQuery, you can analyze variants within a table. Here is an example showing how to compute the ratio of transitions to transversions in SNPs in each chromosome for each sample.

- **Running joins**
  Using BigQuery, you can run a JOIN query on variants with data described by genomic region intervals, or overlaps. This page shows how to use a complex JOIN query to take a list of gene names and do the following:
  - Find the rare SNPs overlapping the genes
  - Find 100,000 base pairs on either side of a gene for the whole genome samples

There are three queries presented, each of which demonstrates how BigQuery scales over different sizes of genomic data:

- ✓ Querying an inline table
- ✓ Querying a materialized table with specific genes
- ✓ Querying a materialized table with 250 random genes

- **Running variant transforms**
  Variant Transforms is an open-source tool used with Cloud Life Sciences. It is based on Apache Beam and uses Dataflow.

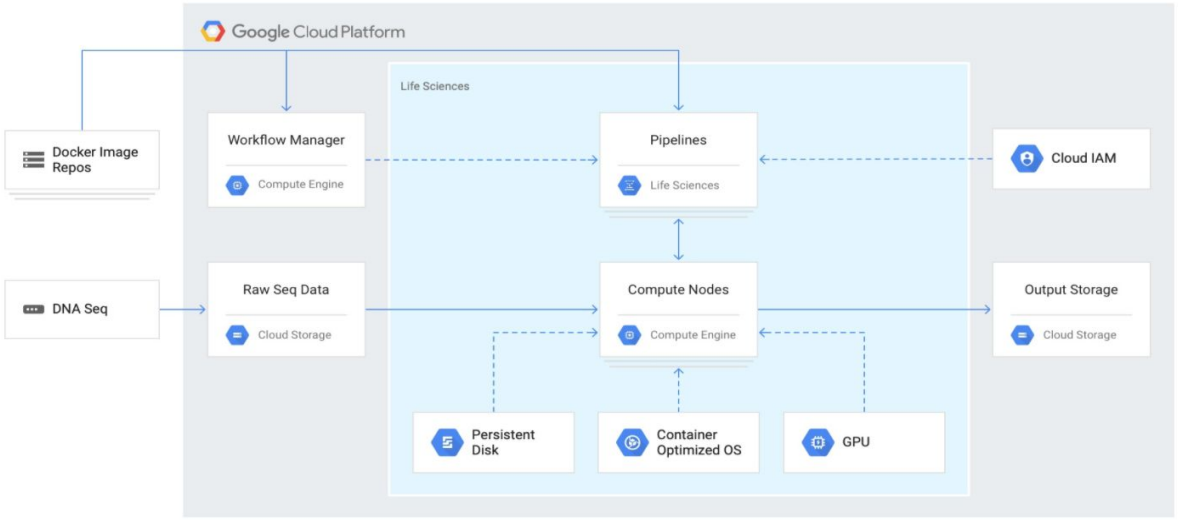
  Using the tool allows you to transform and load hundreds of thousands of files, millions of samples, and billions of records in a scalable manner. The tool also has a preprocessor which you can use to validate VCF files and identify inconsistencies.

  The typical workflow for using the tool consists of the following steps:

  1. Storing raw VCF files in Cloud Storage.
  2. Using the Variant Transforms tool to load the VCF files from Cloud Storage into BigQuery.

  You can then use BigQuery to analyze the variants. You should familiarize yourself with the BigQuery variants schema for information on how the tool loads VCF files into BigQuery tables.

Cloud Life Sciences architecture diagram



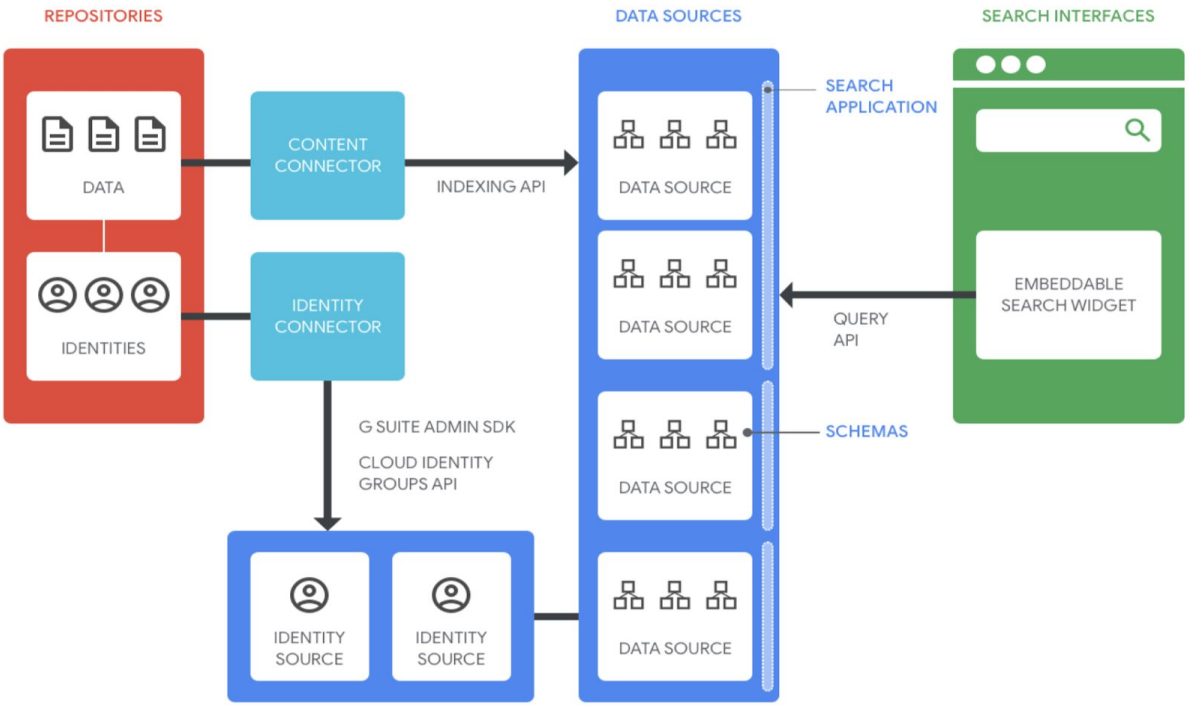
▶ Cloud Search: the best of Google Search for your research projects

Google Cloud Search allows researchers on a project to search and retrieve information, such as internal documents, database fields, and CRM data, from the project's internal data repositories.

Here are some use cases that might be solved by Google Cloud Search:

- Researchers need a way to find project or lab policies, documents, and content authorized by other researchers.
- Researches need to find internal information about lab projects
- Researchers want to view the status of all BigQuery jobs on a particular subject.
- Researchers want a definition for a project-specific term

Architecture overview: The figure below shows all key components of a Google Cloud Search implementation. For a detailed explanation of the definitions of the most important terms in the figure, please refer here.



Leveraging Cloud Search Connectors to give end users access to indexed data:

By default, Google Cloud Search indexes all of your G Suite data. You can also create your own custom program, called a connector, to index data stored in a third-party repository.

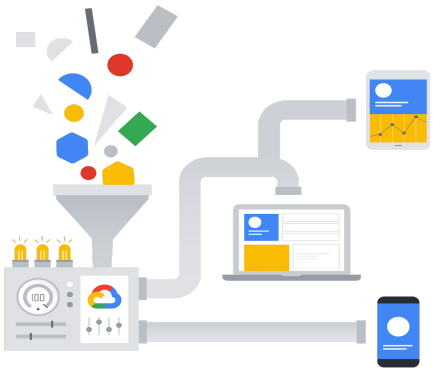
A connector can be a separate program, a script that runs in its own process, or an add-on to your repository.

There are two types of connectors: content connectors and identity connectors. Content connectors are used to traverse a repository and index the data so that Google Cloud Search can effectively search that data. Identity connectors are used to map your enterprise's identities and group rosters to the Google accounts and groups used by Google Cloud Search. These mappings facilitate setting ACLs and search quality hints during indexing.

Several connectors have been built by Google and its partners. For a list of pre-built connectors, refer to the Cloud Search connector directory.

4.4 External Data Access

Google Cloud provides researchers with the ability to share data in a customized way, within or across institutions. There are sharing and collaborating solutions for each and every data storage option that their projects are based on. Otherwise, researchers could build APIs that interact with their services and data to grant external access. With either option implemented correctly, STRIDES members will be able to collaborate with each other, share data across projects and avoid getting extra bills.



▶ Regulated Access to Google Cloud Storage (GCS) Data

Name	Retention policy	Requester Pays	Encryption
my-d...		ON	Google-managed key

Researchers can request view / edit access to a certain bucket using GCS request endpoints, which supports API calls or browser downloads with the correct Cloud IAM permission. Data owners could control who pays the resulting charges of the operation by implementing requester pays. For example, if you are hosting a terabyte-scale DICOM de-identified medical images on GCS, you would want to implement Requester Pays on the buckets where you store those images to prevent people from using your project as a way to get bulk downloads for free. For a detailed explanation of the logistics and set-up on GCP, please refer to this six-minute tutorial.

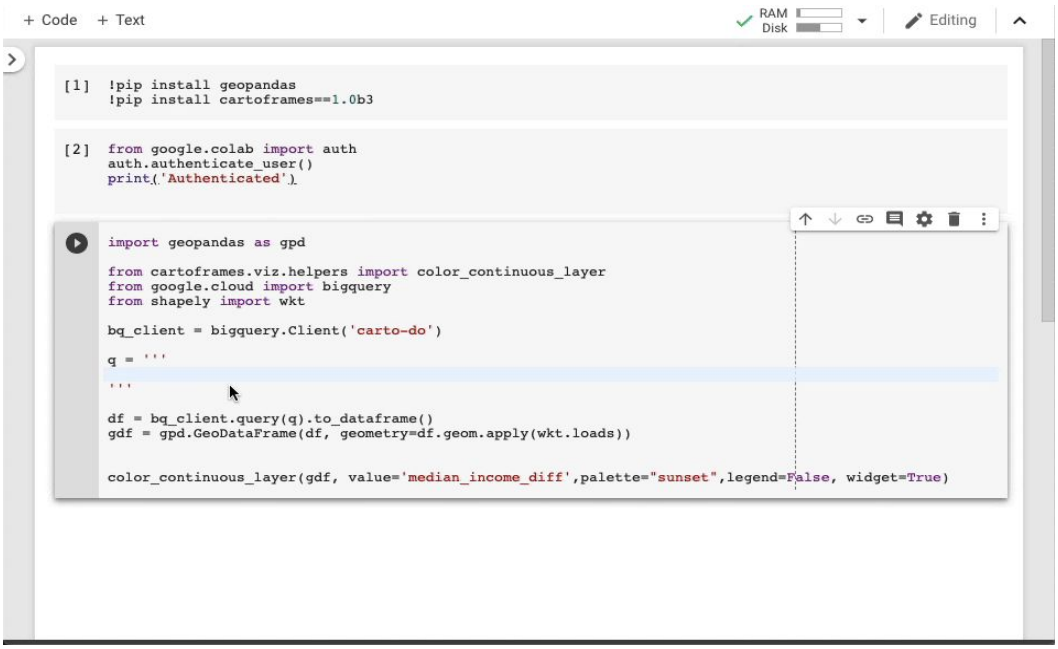
How can researchers share their data as a public dataset

You can share any of your datasets with the public by changing the dataset's access controls to allow access by "All Authenticated Users". For more information about setting dataset access controls, see Controlling access to datasets.
Pros and cons:

There are pros and cons to sharing your data as a public dataset. When you publish a dataset and it gains more visibility from other researchers, you are more likely to get collaboration offers and open-source contributors. On the other hand, exposing your dataset publicly could also cause unwanted egress charge to your billing account, as other users could get bulk downloads for free.

When you share a dataset with the public:

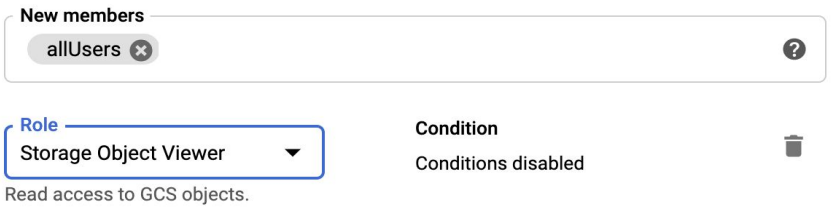
- Storage charges are incurred by the billing account attached to the project that contains the publicly-shared dataset.
- Query charges are incurred by the billing account attached to the project where the query jobs are run.
- For more information, see How charges are billed.



Public Access to GCS Data

When data owners make a bucket or specific objects they own readable to everyone on the public internet, users can view them without the IAM permission from the owner's project. One can make data public through GCP console, command line tools (gsutil) or REST apis.

When users want to access public data in Google Cloud, they also have the above options depending on how they want to work with the data.



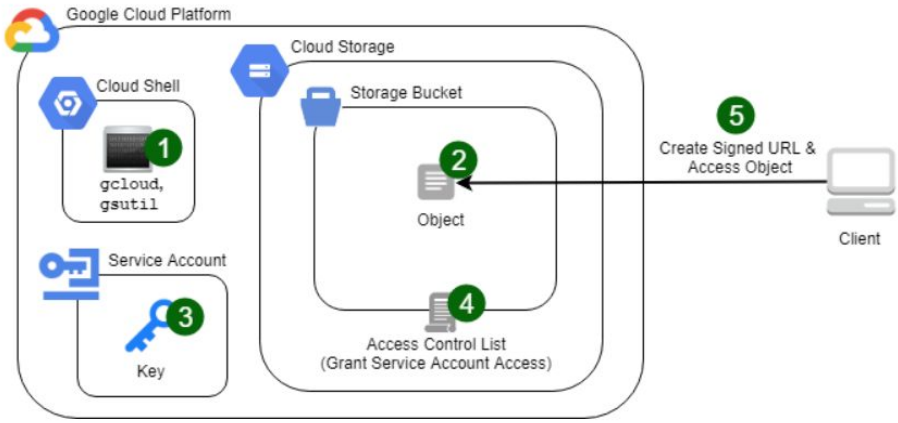
When you are sharing data publicly, where are multiple ways you can do it in a cost-effective way. You can always enable Requester Pays (see the last section for details and demo), set up quotas, or have a combination of both. Namely, allowing data viewers a certain amount of free quota before enabling Requester Pays. Here are the demonstration for the later two methods:

- Set up Quotas
  There are two main ways to view your current quota limits in the Google Cloud Console:
  - Using the Quotas page, which gives you a list of all your project's quota usage and limits.
  - Using the console, which gives you quota information for a particular API, including resource usage over time.

  You can find out how to monitor your quota usage and how to set quota alerts in Monitoring quota metrics.
- Combining Quotas with Requester Pays
  - By the time that this guide is published, Google account team is actively interacting with GCS product team to push forward this feature request, as it is mentioned multiple times during office hours. You should be expecting to see updates on our progress before long.
  - For setting up Requester Pays alone, please refer to the section " Regulated Access to Google Cloud Storage (GCS) Data".

Sharing Data with Gsutil

When operating on big data, you will find gsutil a powerful tool to work with GCS buckets/objects easily and robustly. You can seamlessly transfer your data from on-prem to Cloud, configure your GCS buckets, make copies of data and share them with the public internet.



Giving end users access to viewable data with Cloud Search

When you allow end users to view data, you might want to customize their viewing experience so that they could find information easily. Try building a search widget or creating a custom search interface that interacts with a search application. Both ways serve as good start allowing end users to find data within your project's storage.

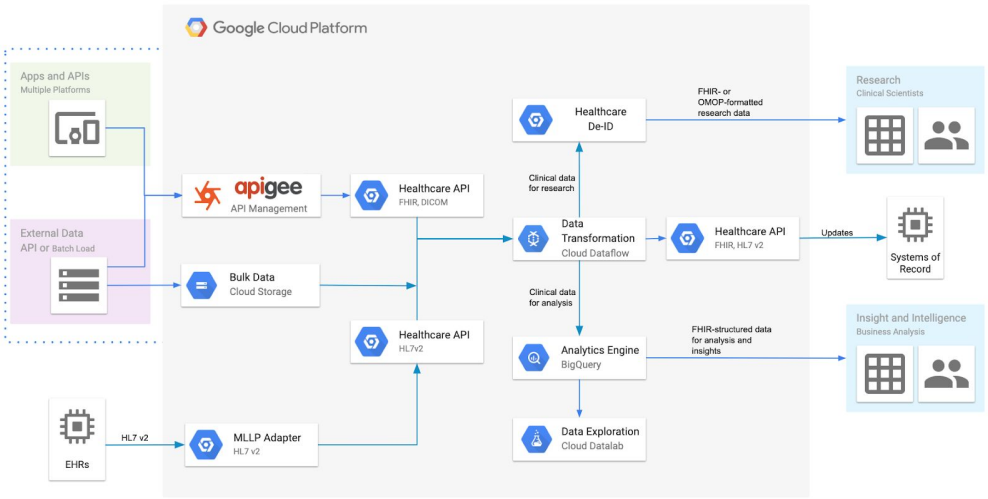
Leveraging Cloud Healthcare API

There is an increasing number of researchers using Cloud Healthcare API. For end users to have a seamless experience sharing standards-based data through Healthcare without an outstanding egress charge to the data owner, you can implement one of these solutions:

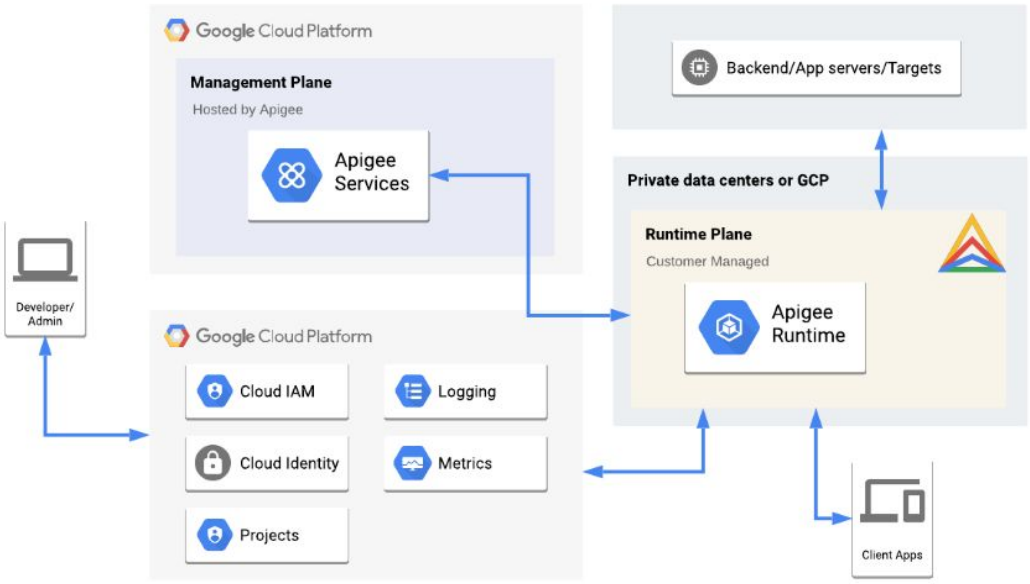
- Access all the data through Healthcare API and set quotas for external read. Build the rest of your app as an interface on top.
- Use Apigee to enforce quota limits, but all Healthcare API charges will go to the project that owns the data. Check out the description of Apigee's quota control capabilities. If you decide that Apigee is helpful for your specific scenarios, check out this tutorial for getting Apigee up and running on top of the Cloud Healthcare API.

Building APIs that interact with your services and data

With Apigee, you can build API proxies—RESTful, HTTP-based APIs that interact with your services. API proxies give you the full power of Apigee's API platform to secure API calls, throttle traffic, mediate messages, analyze API traffic data and so on.



Researchers can [design APIs] that interact with their services and data. You can also [expose or publish APIs] to grant external access to services and data. Check out [steps for publishing APIs] to make your APIs available to consume by developers.



5. Appendix

- Data & Database Management
- BigQuery public datasets
- Requester Pays | Cloud Storage
- Design APIs | Apigee
- Publish APIs | Apigee
- Publishing overview
- Introduction to Google Cloud Search
- Migrating data warehouses to BigQuery: Data governance
- Data Catalog overview | Data Catalog Documentation
- Data Catalog
- Understanding data with machine learning
- Solve with Google Cloud