Google Cloud Whitepaper
September 2018

Data deletion on Google Cloud Platform

Google Cloud

Table of contents

Overview **3**

CIO-level summary 3

Introduction 4

Data storage and replication 5

Secure and effective data deletion **6**

Data deletion pipeline 6

Stage 1 - Deletion request 6

Stage 2 - Soft deletion 7

Stage 3 - Logical deletion from active systems 8

Stage 4 - Expiration from backup systems 9

Deletion timeline 10

Ensuring safe and secure media sanitization **11**

Overview

CIO-level summary

1. Google takes a principled approach to the storage and deletion of Customer Data. Google Cloud Platform is engineered to achieve a high degree of speed, availability, durability, and consistency, and the design of systems optimized for these performance attributes must be balanced carefully with the need to achieve timely data deletion.

2. When you delete your Customer Data, Google's deletion pipeline begins by confirming the deletion request and eliminating the data iteratively from application and storage layers, from both active and backup storage systems. This process is described generally in Google's statement on deletion and retention.

3. Logical deletion occurs in phases, beginning with marking the data for deletion in active storage systems immediately and isolating the data from ordinary processing at the application layer. Successive compaction and mark-and-sweep deletion cycles in Google's storage layers serve to overwrite the deleted data over time. Cryptographic erasure is also used to render the deleted data unrecoverable. Finally, backup systems containing snapshots of Google's active systems are retired on a standard cycle.

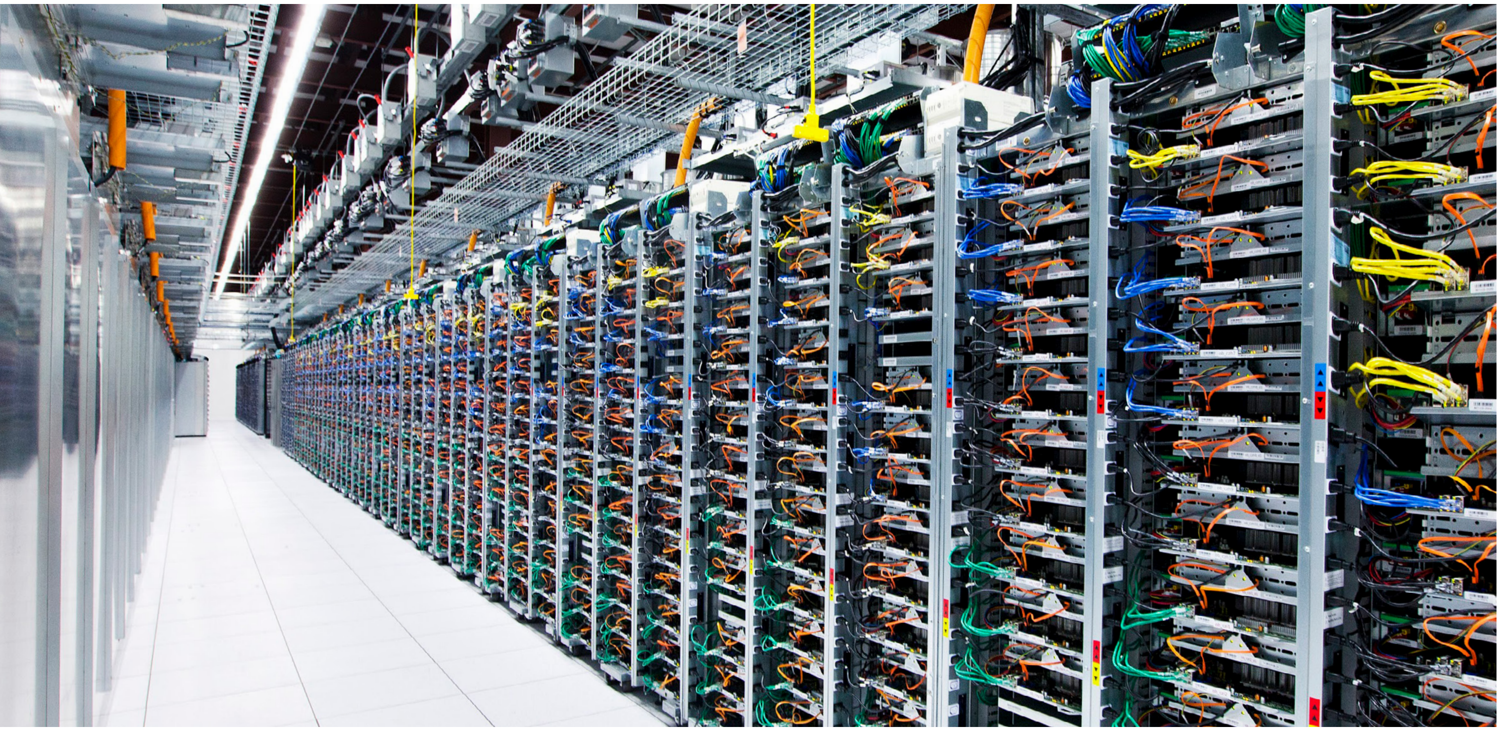
4. Deletion from application and storage layers may occur immediately depending on how storage of the data has been configured and the timing of ongoing deletion cycles in the relevant storage layers and data centers. Deletion from active systems typically completes within about two months of the deletion request. Finally, Customer Data is removed from Google's long-term backup systems, which preserve snapshots of Google systems for up to six months (180 days) to guard against natural disasters and catastrophic events.

Introduction

This document gives you an overview of the secure process that occurs when you delete your Customer Data (as defined in the Google Cloud Platform Terms of Service) stored in Google Cloud Platform. Ensuring safe deletion of Customer Data at the end of its life cycle is a basic aspect of working with data on any computing platform.

Working with data in any cloud platform that commits to high levels of availability, speed and accessibility from any location, and durability against data loss or disasters requires technical innovation to achieve prompt deletion at scale. Google, as an early player in engineering storage platforms for products that process trillions upon trillions of data elements, brings more than a decade of industry experience to bear on optimizing high performance storage systems for this task.

This whitepaper will start with an overview of how Customer Data is stored in Google Cloud Platform. Next, we will describe Google's deletion pipeline and the period of time it generally takes to complete deletion at each stage. Finally, we will describe how we prevent any reconstruction of data stored in our platform through a secure hardware decommissioning and sanitization process.



Data storage and replication

Our description of how Google Cloud Platform deletes Customer Data necessarily begins with a brief overview of how data storage works within Google's infrastructure. Google Cloud Platform offers storage services, such as Cloud Bigtable and Cloud Spanner. Most Google Cloud Platform applications and services access Google's storage systems indirectly via these Cloud storage services or through other internal storage services used by Google.

Google Cloud Platform is designed to provide low latency, highly available, scalable, and durable solutions. Data replication is critical to achieve these key performance goals. Redundant copies of Customer Data could be stored locally and regionally and even globally, depending on your configuration and the demands of customer projects. Actions taken on data in Google Cloud Platform may be simultaneously replicated in multiple data centers, so that Customer Data is highly available. When performance-impacting changes occur in the hardware, software, or network environment, Customer Data is automatically shifted from one system or facility to another, subject to customers' configuration settings, so that customer projects continue performing at scale and without interruption.

"Google Cloud Platform is designed to provide low latency, highly available, scalable, and durable solutions. Data replication is critical to achieve these key performance goals."

At the physical storage level, Customer Data is stored at rest in two types of systems: active storage systems and backup storage systems. These two types of systems process data differently. Active storage systems are Google Cloud Platform's production servers running Google's application and storage layers. Active systems are mass arrays of disks and drives used to write new data as well as store and retrieve data in multiple replicated copies. Active storage systems are optimized to perform live read / write operations on Customer Data at speed and scale.

Google's backup storage systems house full and incremental copies of Google's active systems for a defined period of time to help Google recover data and systems in the event of a catastrophic outage or disaster. Unlike active systems, backup systems are designed to receive periodic snapshots of Google systems and backup copies are retired after a limited window of time as new backup copies are made.

"Google's backup storage systems house full and incremental copies of Google's active systems for a defined period of time to help Google recover data and systems in the event of a catastrophic outage or disaster."

Throughout the storage systems described above, Customer Data is encrypted when stored at rest. The details of Google's encryption techniques are discussed in greater detail in Google's Cloud Security Whitepapers. Encryption of data at rest occurs at the application and storage layers, on both active and backup storage media.

Secure and effective data deletion

Data deletion pipeline

Once Customer Data is stored in Google Cloud Platform, our systems are designed to store the data securely until it completes the stages of Google's data deletion pipeline. This section describes this process in detail.

Stage 1 - Deletion request

The deletion of Customer Data begins when the customer initiates a deletion request. Generally, a deletion request is directed to a specific resource, a Google Cloud Platform project, or the customer's Google account. Deletion requests may be handled in different ways depending on the scope of the customer's request:

Resource deletion: Individual resources containing Customer Data, such as Google Cloud Storage buckets, can be deleted in a number of ways from the Cloud Console or via API. For example, customers may issue a remove bucket or rm -r command to delete a storage bucket through the command line or customers may select a storage bucket and delete it from the Cloud Storage Browser.

Project deletion: As a Google Cloud Platform project owner, you can shut down a project. Deleting a project acts as a bulk deletion request for all resources tied to the corresponding project_number.

Account deletion: When you delete your Google account, it deletes all Google Cloud Platform projects that are solely owned by you. Note that when there are multiple owners for a project, the project is not deleted until all owners are removed from the project or delete their Google accounts. This ensures that Google Cloud Platform projects continue so long as they have an owner.

While deletion requests are designed primarily to be used by customers to manage their data, Google may issue deletion requests automatically, for instance when a customer terminates their relationship with Google.

Stage 2 - Soft deletion

Soft deletion is the natural point in the process to provide a brief internal staging and recovery period to ensure that there is time to recover any data that has been marked for deletion by accident or error. Individual Google Cloud Platform products may adopt and configure such a defined recovery period before the data is deleted from the underlying storage systems so long as it fits within Google's overall deletion timeline.

To illustrate, when projects are deleted, Google Cloud Platform first identifies the unique project_number, then it broadcasts a suspension signal to the Google Cloud Platform products containing that project_number, for example App Engine and Cloud Bigtable. In this case, App Engine will immediately suspend operations keyed to that project_number and relevant tables in Cloud Bigtable will enter an internal recovery period for up to 30 days. At the end of the recovery period, Google Cloud Platform broadcasts a signal to the same products to begin logical deletion of resources tied to the unique project_number. Then Google waits (and, when necessary, rebroadcasts the signal) to collect an acknowledgement signal (ACK) from the applicable products to complete project deletion.

When a Google account is closed, Google Cloud Platform may impose an internal recovery period up to 30 days, depending on past account activity. Once that grace period expires, a signal containing the deleted billing account user_id is broadcasted to Google products and Google Cloud Platform resources tied solely to that user_id are marked for deletion.

"Soft deletion is the natural point in the process to provide a brief internal staging and recovery period to ensure that there is time to recover any data that has been marked for deletion by accident or error."

Stage 3 - Logical deletion from active systems

After the data is marked for deletion and any recovery period has expired, the data is deleted successively from Google's active and backup storage systems. On active systems, data is deleted in two ways.

In all Cloud products under Compute, Storage & Databases, and Big Data except Google Cloud Storage, copies of the deleted data are marked as available storage and overwritten over time. In an active storage system, like Cloud Bigtable, deleted data is stored as entries within a massive structured table. Compacting existing tables to overwrite deleted data can be expensive, as it requires re-writing tables of existing (non-deleted) data, so mark-and-sweep garbage collection and major compaction events are scheduled to occur at regular intervals to reclaim storage space and overwrite deleted data.

In Google Cloud Storage, Customer Data is also deleted through cryptographic erasure. This is an industry standard technique that renders data unreadable by deleting the encryption keys needed to decrypt that data. One advantage of using cryptographic erasure, whether it involves Google-supplied or customer-supplied encryption keys, is that logical deletion can be completed even before all deleted blocks of that data are overwritten in Google Cloud Platform's active and backup storage systems.

> "In Google Cloud Storage, customer data is also deleted through cryptographic erasure. This is an industry standard technique that renders data unreadable by deleting the encryption keys needed to decrypt that data."

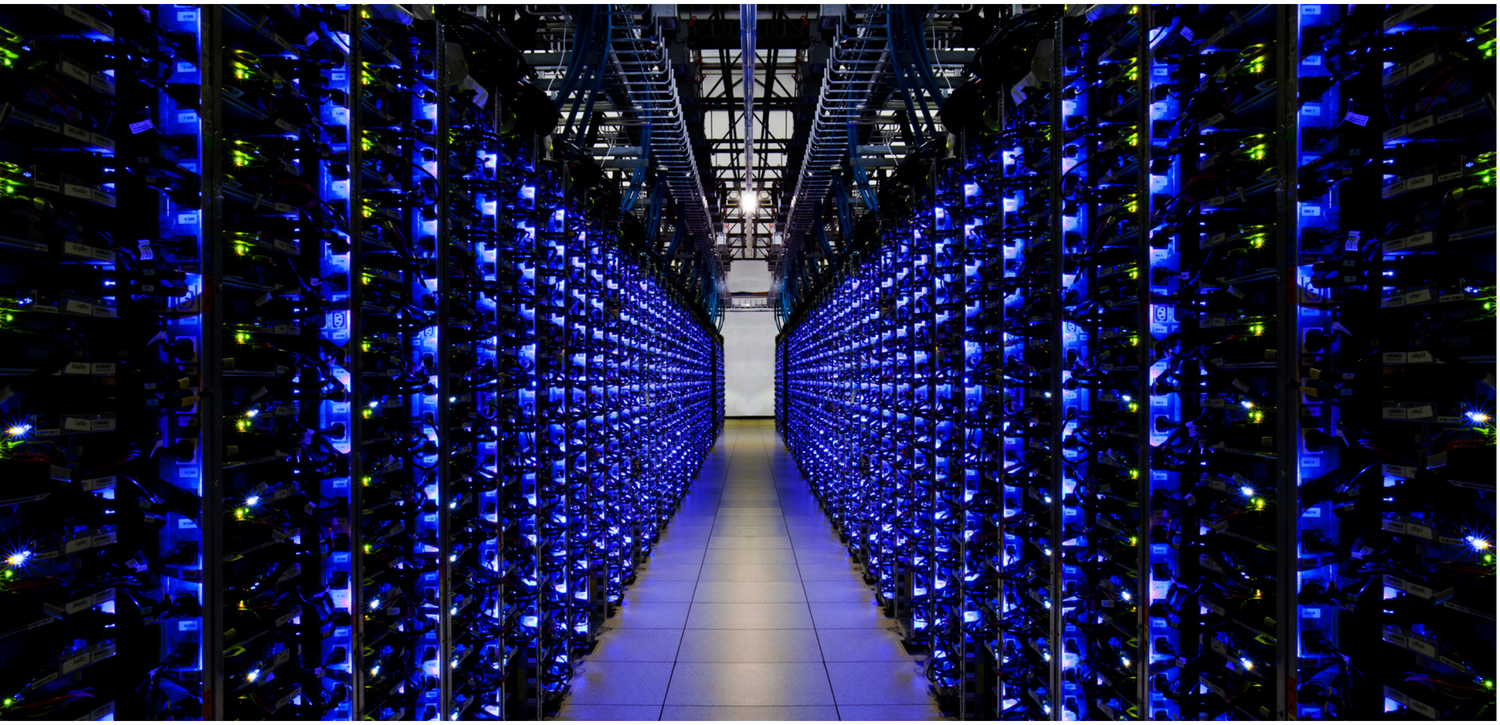
Stage 4 - Expiration from backup systems

Similar to deletion from Google's active systems, deleted data is eliminated from backup systems using both overwriting and cryptographic techniques. In the case of backup systems, however, Customer Data is typically stored within large aggregate snapshots of active systems that are retained for static periods of time to ensure business continuity in the event of a disaster (e.g., an outage affecting an entire data center), when the time and expense of restoring a system entirely from backup systems may become necessary. Consistent with reasonable business continuity practices, full and incremental snapshots of active systems are made on a daily, weekly, and monthly cycles and retired after a predefined period of time to make room for the newest snapshots.

When a backup is retired, it is marked as available space and overwritten as new daily / weekly / monthly backups are performed.

Note that any reasonable backup cycle imposes a pre-defined delay in propagating a data deletion request through backup systems. When Customer Data is deleted from active systems, it is no longer copied into backup systems. Backups performed prior to deletion are expired regularly based on the pre-defined backup cycle.

Finally, cryptographic erasure of the deleted data may occur before the backup containing Customer Data has expired. Without the encryption key used to encrypt specific customer data, the Customer Data will be unrecoverable even during its remaining lifespan on Google's backup systems.



Deletion timeline

Google Cloud Platform is engineered to achieve a high degree of speed, availability, durability, and consistency, and the design of systems optimized for these performance attributes must be balanced carefully with the need to achieve timely data deletion. Google Cloud Platform commits to delete Customer Data within a maximum period of about six months (180 days). This commitment incorporates the stages of Google's deletion pipeline described above, including:

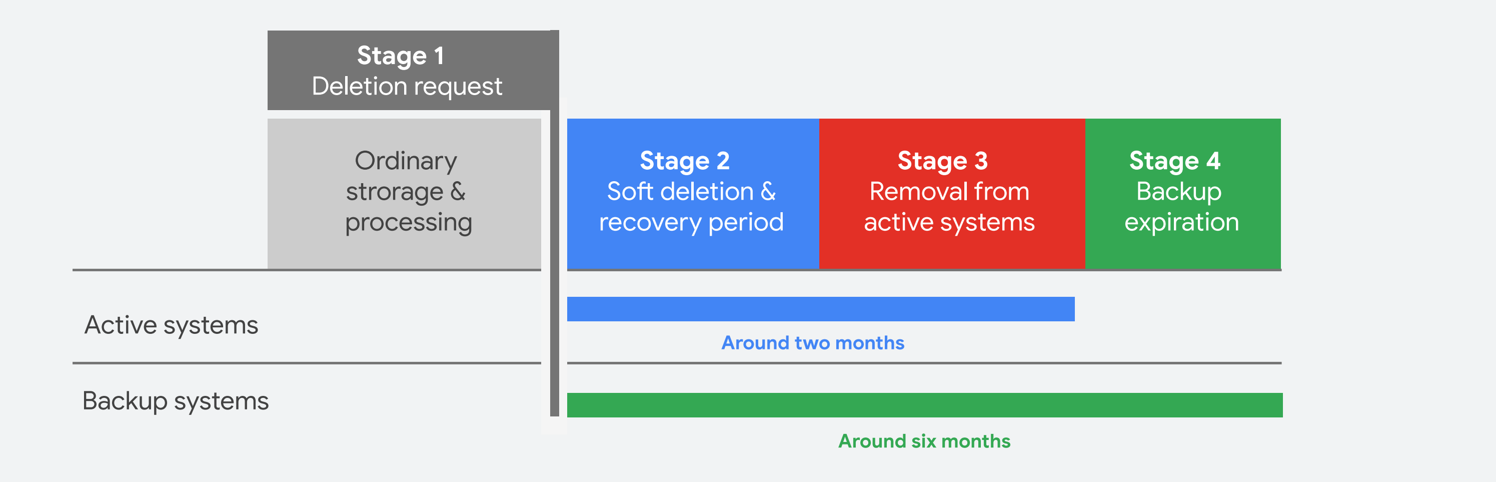
Stage 2 - Once the deletion request is made, data is typically marked for deletion immediately and our goal is to perform this step within a maximum period of 24 hours. After the data is marked for deletion, an internal recovery period of up to 30 days may apply depending on the service or deletion request.

Stage 3 - The time needed to complete garbage collection tasks and achieve logical deletion from active systems. These processes may occur immediately after the deletion request is received, depending on the level of data replication and the timing of ongoing garbage collection cycles. From deletion request, it generally takes about two months to delete data from active systems, which is typically enough time to complete two major garbage collection cycles and ensure that logical deletion is completed.

Stage 4 - Google backup cycle is designed to expire deleted data within data center backups within six months of the deletion request. Deletion may occur sooner depending on the level of data replication and the timing of Google's ongoing backup cycles.

> "From deletion request, it generally takes about two months to delete data from active systems, which is typically enough time to complete two major garbage collection cycles and ensure that logical deletion is completed."

Figure 1: The stages of Google Cloud Platform's deletion pipeline



Ensuring safe and secure media sanitization

In addition to Google Cloud Platform's deletion pipeline, a disciplined media sanitization program enhances the security of the deletion process by preventing forensic or laboratory attacks on the physical storage media once it has reached the end of its life cycle.

Google meticulously tracks the location and status of all storage equipment within our data centers, through acquisition, installation, retirement, and destruction, via barcodes and asset tags that are tracked in Google's asset database. Various techniques such as biometric identification, metal detection, cameras, vehicle barriers, and laser-based intrusion detection systems are used to prevent equipment from leaving the data center floor without authorization. Learn more in the Google Infrastructure Security Design Overview.

"Various techniques such as biometric identification, metal detection, cameras, vehicle barriers, and laser-based intrusion detection systems are used to prevent equipment from leaving the data center floor without authorization."

Physical storage media may be decommissioned for a range of reasons. If a component fails to pass a performance test at any point during its life cycle, it is removed from inventory and retired. Google also upgrades obsolete hardware to improve processing speed and energy efficiency, or increase storage capacity. Whether hardware is decommissioned due to failure, upgrade, or any other reason, storage media is decommissioned using appropriate safeguards. Google hard drives use technologies like full disk encryption (FDE) and drive locking to protect data at rest during decommission. When a hard drive is retired, authorized individuals verify that the disk is erased by overwriting the drive with zeros and performing a multi-step verification process to ensure the drive contains no data.

If the storage media cannot be erased for any reason, it is stored securely until it can be physically destroyed. Depending on available equipment, we either crush and deform the drive or shred the drive into small pieces. In either case, the disk is recycled at a secure facility, ensuring that no one will be able to read data on retired Google disks. Each data center adheres to a strict disposal policy and uses the techniques described to achieve compliance with NIST SP 800-88 Revision 1 “Guidelines for Media Sanitization” and DoD 5220.22-M “National Industrial Security Program Operating Manual.”

"Each data center adheres to a strict disposal policy and uses the techniques described to achieve compliance with NIST SP 800-88 Revision 1 “Guidelines for Media Sanitization” and DoD 5220.22-M “National Industrial Security Program Operating Manual."