

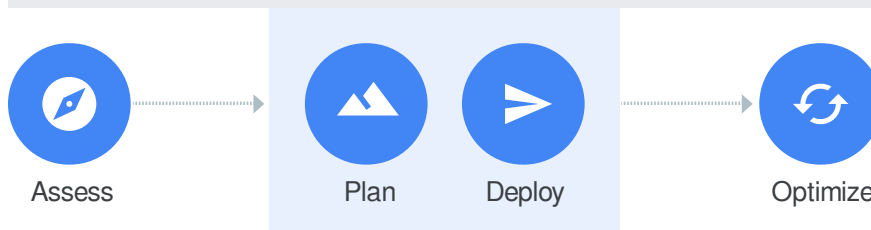
For many customers, the first step in adopting a Google Cloud product is getting their data into Google Cloud. This document explores that process, from planning a data transfer to using best practices in implementing a plan.

Transferring large datasets involves building the right team, planning early, and testing your transfer plan before implementing it in a production environment. Although these steps can take as much time as the transfer itself, such preparations can help minimize disruption to your business operations during the transfer.

This article is part of a series:

- [Migration to Google Cloud: Getting started](/solutions/migration-to-gcp-getting-started) (/solutions/migration-to-gcp-getting-started)
- [Migration to Google Cloud: Assessing and discovering your workloads](/solutions/migration-to-gcp-assessing-and-discovering-your-workloads) (/solutions/migration-to-gcp-assessing-and-discovering-your-workloads)
- [Migration to Google Cloud: Building your foundation](/solutions/migration-to-google-cloud-building-your-foundation) (/solutions/migration-to-google-cloud-building-your-foundation)
- Migration to Google Cloud: Transferring your large datasets (this document)
- [Migration to Google Cloud: Deploying your workloads](/solutions/migration-to-gcp-deploying-your-workloads) (/solutions/migration-to-gcp-deploying-your-workloads)
- [Migration to Google Cloud: Migrating from manual deployments to automated, containerized deployments](/solutions/migration-to-google-cloud-automated-containerized-deployments) (/solutions/migration-to-google-cloud-automated-containerized-deployments)

The following diagram illustrates the path of your migration journey.



The deployment phase is the third phase in your migration to Google Cloud, where you design a deployment process for your workloads.

This document is useful if you're planning a migration from an on-premises environment, from a private hosting environment, from another cloud provider to Google Cloud, or if you're evaluating the

opportunity to migrate and want to explore what it might look like.

For the purposes of this document, data transfer is the process of moving data without transforming it, for example, moving files as they are into objects.

It's tempting to think of data transfer as one giant FTP session, where you put your files in one side and wait for them to come out the other side. However, in most enterprise environments, the transfer process involves many factors such as the following:

- Devising a transfer plan that accounts for administrative time, including time to decide on a transfer option, get approvals, and deal with unanticipated issues.
- Coordinating people in your organization, such as the team that executes the transfer, personnel who approve the tools and architecture, and business stakeholders who are concerned with the value and disruptions that moving data can bring.
- Choosing the right transfer tool based on your resources, cost, time, and other project considerations.
- Overcoming data transfer challenges, including "speed of light" issues (insufficient bandwidth), moving datasets that are in active use, protecting and monitoring the data while it's in flight, and ensuring the data is transferred successfully.

This document aims to help you get started on a successful transfer initiative.

The following list includes resources for other types of data transfer projects not covered in this document:

- If you need to transform your data (such as combining rows, joining datasets, or filtering out personal identifiable information), you should consider [an extract, transform, and load \(ETL\) solution](#) (/blog/products/gcp/designing-etl-architecture-for-a-cloud-native-data-warehouse-on-google-cloud-platform) that can deposit data into a Google Cloud data warehouse. For an example of this architecture, see this [Dataflow tutorial](#) (/solutions/performing-etl-from-relational-database-into-bigquery).

- If you need to migrate a database and related apps (for example, to lift and shift a database app), you might look at documentation for [Cloud Spanner](#) (/spanner/docs/migrating-postgres-spanner), solutions for [PostgreSQL](#) (/solutions/migrating-postgresql-to-gcp) and [MySQL](#) (/solutions/migrating-mysql-to-cloudsql-using-automated-migration-workflow-tutorial), and other documentation about your database type.
- If you need to move a virtual machine (VM) instance, consider using Google's VM migration product, [Migrate for Compute Engine](#) (/migrate/compute-engine/).

Planning a transfer typically requires personnel with the following roles and responsibilities:

- **Enabling resources needed for a transfer:** Storage, IT, and network admins, an executive sponsor, and other advisors (for example, a Google Account team or integration partners)
- **Approving the transfer decision:** Data owners or governors (for internal policies on who is allowed to transfer what data), legal advisors (for data-related regulations), and a security admin (for internal policies on how data access is protected)
- **Executing the transfer:** A team lead, a project manager (for executing and tracking the project), an engineering team, and on-site receiving and shipping (to receive appliance hardware)

It's crucial to identify who owns the preceding responsibilities for your transfer project and to include them in planning and decision meetings when appropriate. Poor organizational planning is often the cause of failed transfer initiatives.

Gathering project requirements and input from these stakeholders can be challenging, but making a plan and establishing clear roles and responsibilities pays off. You can't be expected to know all the details of your data. Assembling a team gives you greater insight into the needs of the business. It's a best practice to identify potential issues before you invest time, money, and resources to complete the transfers.

When you design a transfer plan, we recommend that you first collect requirements for your data transfer and then decide on a transfer option. To collect requirements, you can use the following process:

## 1. Identify what datasets you need to move.

- Select tools like [Data Catalog](/data-catalog/) (/data-catalog/) to organize your data into logical groupings that are moved and used together.
- Work with teams within your organization to validate or update these groupings.

## 2. Identify what datasets you *can* move.

- Consider whether regulatory, security, or other factors prohibit some datasets from being transferred.
- If you need to transform some of your data before you move it (for example, to remove sensitive data or reorganize your data), consider using a data integration product like [Dataflow](/dataflow/) (/dataflow/) or [Cloud Data Fusion](/data-fusion/) (/data-fusion/), or a workflow orchestration product like [Cloud Composer](/composer/?hl=vi) (/composer/?hl=vi).

## 3. For datasets that are movable, determine where to transfer each dataset.

- [Record which storage option you select](/storage-options/) (/storage-options/) to store your data. Typically, the target storage system on Google Cloud is [Cloud Storage](/storage/docs/) (/storage/docs/). Even if you need more complex solutions after your applications are up and running, Cloud Storage is a scalable and durable storage option. For more information, see [Best practices for Cloud Storage](/storage/docs/best-practices/) (/storage/docs/best-practices).
- Understand what data access policies must be maintained after migration.
- Determine if you need to store this data in specific regions.
- Plan how to structure this data at the destination. For instance, will it be the same as the source or different?
- Determine if you need to transfer data on an ongoing basis.

## 4. For datasets that are movable, determine what resources are available to move them.

- Time: When does the transfer need to be completed?
- Cost: What is the budget available for the team and transfer costs?
- People: Who is available to execute the transfer?
- Bandwidth (for online transfers): How much of your currently available bandwidth for Google Cloud can be allocated for a transfer, and for what period of time?

Before you evaluate and select transfer options in the next phase of planning, we recommend that you assess whether any part of your IT model can be improved, such as data governance, organization, and security.

Many members of the transfer team might be granted new roles in your Google Cloud organization as part of your data transfer project. Data transfer planning is a great time to review your Cloud Identity and Access Management (Cloud IAM) permissions and [best practices for using Cloud IAM securely](/iam/docs/using-iam-securely) (/iam/docs/using-iam-securely). These issues can affect how you [grant access to](/storage/docs/best-practices#security) (/storage/docs/best-practices#security) your storage. For example, you might place strict limits on write access to data that has been archived for regulatory reasons, but you might allow many users and applications to write data to your test environment.

How you structure your data on Google Cloud depends on how you plan to use Google Cloud. Storing your data in the same Cloud project where you run your application is a simple approach, but it might not be optimal from a management perspective. Some of your developers might not have privilege to view the production data. In that case, a developer could develop code on sample data, while a privileged service account could access production data. Thus, you might want to keep your entire production dataset in a separate Cloud project, and then use a service account to allow access to the data from each application project.

Google Cloud is organized around projects. Projects can be grouped into folders, and folders can be grouped under your organization. Roles are established at the project level and the access permissions are added to these roles at the Cloud Storage bucket levels. This structure aligns with the permissions structure of other [object store providers](/docs/compare/aws/storage) (/docs/compare/aws/storage).

For more information on how to structure a Google Cloud organization, see [Best practices for enterprise organizations](/docs/enterprise/best-practices-for-enterprise-organizations) (/docs/enterprise/best-practices-for-enterprise-organizations).

To evaluate your data transfer options, the transfer team needs to consider several factors, including the following:

- Cost
- Time
- Offline versus online transfer options
- Transfer tools and technologies

- Security

Most of the costs associated with transferring data include the following:

- Networking costs
  - Ingress to Cloud Storage is free. However, if you're hosting your data on a public cloud provider, you can expect to pay an egress charge (<https://aws.amazon.com/blogs/aws/aws-data-transfer-prices-reduced/>) and potentially storage costs (for example, read operations) for transferring your data. This charge applies for data coming from Google or another cloud provider.
  - If your data is hosted in a private data center that you operate, you might also incur added costs for setting up more bandwidth to Google Cloud.
- Storage and operation costs for Cloud Storage during and after the transfer of data
- Product costs (for example, a Transfer Appliance)
- Personnel costs for assembling your team and acquiring logistical support

Few things in computing highlight the hardware limitations of networks as transferring large amounts of data. Typically you can transfer 1 GB in eight seconds over a 1 Gbps network. If you scale that up to a huge dataset (for example, 100 TB), the transfer time is 12 days. Transferring huge datasets can test the limits of your infrastructure and potentially cause problems for your business.

You can use the following calculator to understand how much time a transfer might take, given the size of the dataset you're moving and the bandwidth available for the transfer. A certain percentage of management time is factored into the calculations.

You might not want to transfer large datasets out of your company network during peak work hours. If the transfer overloads the network, nobody else will be able to get necessary or mission-critical

work completed. For this reason, the transfer team needs to consider the factor of time.

After the data is transferred to Cloud Storage, you can use a number of technologies to process the new files as they arrive, such as [Dataflow](#) (/solutions/processing-logs-at-scale-using-dataflow).

How you increase network bandwidth depends on how you connect to Google Cloud.

In a cloud-to-cloud transfer between Google Cloud and other cloud providers, Google provisions the connection between cloud vendor data centers, requiring no setup from you.

If you're transferring data between your private data center and Google Cloud, there are three main approaches:

- A public internet connection by using a public API
- Direct Peering by using a public API
- Cloud Interconnect by using a private API

When evaluating these approaches, it's helpful to consider your long-term connectivity needs. You might conclude that it's cost prohibitive to acquire bandwidth solely for transfer purposes, but when factoring in long-term use of Google Cloud and the network needs across your organization, the investment might be worthwhile.

When you use a public internet connection, network throughput is less predictable because you're limited by your internet service provider's (ISP) capacity and routing. The ISP might also offer a limited Service Level Agreement (SLA) or none at all. However, these connections offer relatively low costs, and with Google's extensive peering arrangements, your ISP might route you onto Google's global network within a few network hops.

We recommend that you check with your security admin on whether your company policy forbids moving some datasets over the public internet. Also check whether the public internet connection is used for your production traffic. Large-scale data transfers might negatively impact the production network.

To access the Google network with fewer network hops than with a public internet connection, you can use [Direct Peering](#) (/interconnect/direct-peering). By using Direct Peering, you can exchange internet

traffic between your network and Google's [Edge Points of Presence](#)

(<https://peering.google.com/#/infrastructure>) (PoPs), which means your data does not use the public internet. Doing so also reduces the number of hops between your network and Google's network.

[Peering with Google's network](#) ([https://isp.google.com/partner\\_request/?data.request\\_type=Peering](https://isp.google.com/partner_request/?data.request_type=Peering)) requires you to set up a registered [Autonomous System \(AS\) Number](#)

(<https://www.iana.org/assignments/as-numbers/as-numbers.xhtml>), connect to Google using an internet exchange, and provide an around-the-clock contact with your network operations center.

[Cloud Interconnect](#) (</interconnect/docs>) offers a direct connection to Google Cloud through Google or one of the Cloud Interconnect service providers. This service helps prevent your data from going on the public internet and can provide a more consistent throughput for large data transfers. Typically, Cloud Interconnect provides SLAs for network availability and performance of their network. [Contact a service provider](#) (</interconnect/docs/concepts/service-providers>) directly to learn more. Cloud Interconnect also supports private addressing, RFC 1918, so that the cloud effectively becomes an extension of your private data center without the need for public IP addresses or NATs.

A critical decision is whether to use an offline or online process for your data transfer. That is, you must choose between transferring over a network, whether it's a dedicated interconnect or the public internet, or transferring by using storage hardware.

To help with this decision, we provide a [transfer calculator](#) (#time) to help you estimate the time and cost differences between these two options. The following chart also shows some transfer speeds for various dataset sizes and bandwidths. A certain amount of management overhead is built into these calculations.





As noted earlier, you might need to consider whether the cost to achieve lower latencies for your data transfer (such as acquiring network bandwidth) is offset by the value of that investment to your organization.

Google offers several tools and technologies to help you perform a data transfer.

Choosing a transfer option depends on your use case, as the following table shows.

Where you're moving data from	Scenario	Suggested products
Another cloud provider (for example, Amazon Web Services or Microsoft Azure) to Google Cloud	—	<a href="/storage-transfer/docs/overview">Storage Transfer Service (/storage-transfer/docs/overview)</a>
Cloud Storage to Cloud Storage (two different buckets)	—	<a href="/storage-transfer/docs/overview">Storage Transfer Service (/storage-transfer/docs/overview)</a>

Where you're moving data from	Scenario	Suggested products
Your private data center to Google Cloud	Enough bandwidth to meet your project deadline for less than a few TB of data	<a href="/storage/docs/gsutil">gsutil</a> (/storage/docs/gsutil)
Your private data center to Google Cloud	Enough bandwidth to meet your project deadline for more than a few TB of on-premises data	<a href="/storage-transfer/docs/overview">Storage Transfer Service</a> (/storage-transfer/docs/overview)
Your private data center to Google Cloud	Not enough bandwidth to meet your project deadline	<a href="/transfer-appliance/">Transfer Appliance</a> (/transfer-appliance/)

The [gsutil tool](/storage/docs/gsutil) (/storage/docs/gsutil) is the standard tool for small- to medium-sized transfers (less than a few TB) over a typical enterprise-scale network, from a private data center to Google Cloud. We recommend that you include `gsutil` in your default path when you use [Cloud Shell](/shell/docs/features) (/shell/docs/features). It's also available by default when you install the [Cloud SDK](/sdk/docs/) (/sdk/docs/). It's a reliable tool that provides all the basic features you need to manage your [Cloud Storage](/storage/docs/) (/storage/docs/) instances, including copying your data to and from the local file system and Cloud Storage. It can also move and rename objects and perform real-time incremental syncs, like `rsync`, to a Cloud Storage bucket.

`gsutil` is especially useful in the following scenarios:

- Your transfers need to be executed on an as-needed basis, or during command-line sessions by your users.
- You're transferring only a few files or very large files, or both.
- You're consuming the output of a program (streaming output to Cloud Storage).
- You need to watch a directory with a moderate number of files and sync any updates with very low latencies.

The basics of getting started with `gsutil` are to [create a Cloud Storage bucket](/storage/docs/creating-buckets#storage-create-bucket-gsutil) (/storage/docs/creating-buckets#storage-create-bucket-gsutil) and [copy data](/storage/docs/uploading-objects) (/storage/docs/uploading-objects) to that bucket. For transfers of larger datasets, there are two things to consider:

- For multi-threaded transfers, use `gsutil -m`.

Several files are processed in parallel, increasing your transfer speeds.

- For a single large file, use [Composite transfers](#) (/storage/docs/gsutil/commands/cp#parallel-composite-uploads).

This method breaks large files into smaller chunks to increase transfer speed. Chunks are transferred and validated in parallel, sending all data to Google. Once the chunks arrive at Google, they are combined (referred to as *compositing*) to form a single object. Compositing can result in early deletion fees for objects stored in Cloud Storage Coldline and Cloud Storage Nearline, so it's not recommended for use with these types of objects.

This feature has some drawbacks, including that each piece (not the entire object) is individually checksummed, and composition of cold storage classes results in early retrieval penalties. For more information, read about [parallel composite uploads](#) (/storage/docs/gsutil/commands/cp#parallel-composite-uploads).

Like `gsutil`, [Storage Transfer Service for on-premises data](#) (/storage-transfer/docs/on-prem-overview) (in beta) enables transfers from network file system (NFS) storage to Cloud Storage. Although `gsutil` can support small transfer sizes (up to a few TB), Storage Transfer Service for on-premises data is designed for large-scale transfers (up to petabytes of data, billions of files). It supports full copies or incremental copies, and it works on all transfer options listed earlier in [Deciding among Google's transfer options](#) (#transfer-options). It also has a simple, managed graphical user interface; even non-technically savvy users (after setup) can use it to move data.

Storage Transfer Service for on-premises data is especially useful in the following scenarios:

- You have sufficient available bandwidth to move the data volumes (see the [Google Cloud Data Transfer Calculator](#) (#time)).
- You support a large base of internal users who might find a command-line tool like `gsutil` challenging to use.
- You need robust error-reporting and a record of all files and objects that are moved.
- You need to limit the impact of transfers on other workloads in your data center (this product can stay under a user-specified bandwidth limit).
- You want to run recurring transfers on a schedule.

You set up Storage Transfer Service for on-premises data by installing on-premises software (known as *agents*) onto computers in your data center. These agents are in Docker containers, which makes it easier to run many of them or orchestrate them through Kubernetes.

After setup is finished, users can initiate transfers in the Google Cloud Console by providing a source directory, destination bucket, and time or schedule. Storage Transfer Service recursively crawls subdirectories and files in the source directory and creates objects with a corresponding name in Cloud Storage (the object `/dir/foo/file.txt` becomes an object in the destination bucket named `/dir/foo/file.txt`). Storage Transfer Service automatically re-attempts a transfer when it encounters any transient errors. While the transfers are running, you can monitor how many files are moved and the overall transfer speed, and you can view error samples.

When the transfer is finished, a tab-delimited file (TSV) is generated with a full record of all files touched and any error messages received. Agents are fault tolerant, so if an agent goes down, the transfer continues with the remaining agents. Agents are also self-updating and self-healing, so you don't have to worry about patching the latest versions or restarting the process if it goes down because of an unanticipated issue.

Things to consider when using Storage Transfer Service:

- **Use an identical agent setup on every machine.** All agents should see the same Network File System (NFS) mounts in the same way (same relative paths). This setup is a requirement for the product to function.
- **More agents results in more speed.** Because transfers are automatically parallelized across all agents, we recommend that you deploy many agents so that you use your available bandwidth.
- **Bandwidth caps can protect your workloads.** Your other workloads might be using your data center bandwidth, so set a bandwidth cap to prevent transfers from impacting your SLAs.
- **Plan time for reviewing errors.** Large transfers can often result in errors requiring review. Storage Transfer Service lets you see a sample of the errors encountered directly in the Cloud Console. If needed, you can load the full record of all transfer errors to BigQuery to check on files or evaluate errors that remained even after retries. These errors might be caused by running apps that were writing to the source while the transfer occurred, or the errors might reveal an issue that requires troubleshooting (for example, permissions error).
- **Set up Stackdriver Monitoring for long-running transfers.** Storage Transfer Service lets Monitoring monitor agent health and throughput, so you can set alerts that notify you when agents are down or need attention. Acting on agent failures is important for transfers that take several days or weeks, so that you avoid significant slowdowns or interruptions that can delay your project timeline.

For large-scale transfers (especially transfers with limited network bandwidth), [Transfer Appliance](/transfer-appliance/) (/transfer-appliance/) is an excellent option, especially when a fast network connection is unavailable

and it's too costly to acquire more bandwidth.

Transfer Appliance is especially useful in the following scenarios:

- Your data center is in a remote location with limited or no access to bandwidth.
- Bandwidth is available, but cannot be acquired in time to meet your deadline.
- You have access to logistical resources to receive and connect appliances to your network.

With this option, consider the following:

- Transfer Appliance requires that you're able to receive and ship back the Google-owned hardware.
- Depending on your internet connection, the latency for transferring data into Google Cloud is typically higher with Transfer Appliance than online.
- Transfer Appliance is available only in certain countries.

The two main criteria to consider with Transfer Appliance are cost and speed. With reasonable network connectivity (for example, 1 Gbps), transferring 100 TB of data online takes over 10 days to complete. If this rate is acceptable, an online transfer is likely a good solution for your needs. If you only have a 100 Mbps connection (or worse from a remote location), the same transfer takes over 100 days. At this point, it's worth considering an offline-transfer option such as Transfer Appliance.

Acquiring a Transfer Appliance is straightforward. In the Cloud Console, you request a [Transfer Appliance](#) (/transfer-appliance/), indicate how much data you have, and then Google ships one or more appliances to your requested location. You're given a number of days to transfer your data to the appliance ("data capture") and ship it back to Google.

The expected turnaround time for a network appliance to be shipped, loaded with your data, shipped back, and rehydrated on Google Cloud is 50 days. If your online transfer timeframe is calculated to be substantially more than this timeframe, consider Transfer Appliance. The total cost for the 480 TB device process is less than \$3,000.

[Storage Transfer Service](#) (/storage-transfer/docs/overview) is a fully managed, highly scalable service to automate transfers from other public clouds into Cloud Storage. It supports transfers into Cloud Storage from Amazon S3 and HTTP.

For Amazon S3, you can supply an access key and an S3 bucket with optional filters for S3 objects to select, and then you copy the S3 objects to any Cloud Storage bucket. The service also supports daily

copies of any modified objects. The service doesn't currently support data transfers to Amazon S3.

For HTTP, you can give Storage Transfer Service a list of public URLs in [a specified format](#) (/storage-transfer/docs/create-url-list). This approach requires that you write a script providing the size of each file in bytes, along with a Base64-encoded MD5 hash of the file contents. Sometimes the file size and hash are available from the source website. If not, you need local access to the files, in which case, it might be easier to use `gsutil`, as described earlier.

If you have a transfer in place, Storage Transfer Service is a great way to get data and keep it, particularly when transferring from another public cloud.

If you would like to move data from another cloud not supported by Storage Transfer Service, you can use `gsutil` from a hosted VM instance.

For many Google Cloud users, security is their primary focus, and there are different levels of security available. A few aspects of security to consider include protecting data at rest (authorization and access to the source and destination storage system), protecting data while in transit, and protecting access to the transfer product. The following table outlines these aspects of security by product.

Product	Data at rest	Data in transit	Access to transfer product
<a href="#">Transfer Appliance</a> (/transfer-appliance/docs/2.0/faq)	All data is encrypted at rest.	Data is protected with keys managed by the customer.	Anyone can order an appliance, but to use it they need access to the data source.
<a href="#">gsutil</a> (/storage/docs/gsutil)	Access keys required to access Cloud Storage, which is encrypted at rest.	Data is sent over HTTPS and encrypted in transit.	Anyone can download and run <code>gsutil</code> . They must have permissions to buckets and local files in order to move data.
<a href="#">Storage Transfer Service</a> (/storage-transfer/docs/on-prem-overview) for on-premises data	Access keys required to access Cloud Storage, which is encrypted at rest. The agent process can access local files as OS permissions allow.	Data is sent over HTTPS and encrypted in transit.	You must have object editor permissions to access Cloud Storage buckets.

Product	Data at rest	Data in transit	Access to transfer product
<a href="/storage-transfer/docs/">Storage Transfer Service</a>	Access keys required for non-Google cloud resources (for example, Amazon S3). Access keys are required to access Cloud Storage, which is encrypted at rest.	Data is sent over HTTPS and encrypted in object editor transit.	You must have Cloud IAM permissions for the service and account to access the source and permissions for any Cloud Storage buckets.

To achieve baseline security enhancements, online transfers to Google Cloud using [gsutil](/storage/docs/gsutil) (</storage/docs/gsutil>) are accomplished over HTTPS, data is encrypted in transit, and all data in Cloud Storage is, by default, encrypted at rest. For information on more sophisticated security-related schemes, see [Security and privacy considerations](/storage/docs/gsutil/addlhelp/SecurityandPrivacyConsiderations) (</storage/docs/gsutil/addlhelp/SecurityandPrivacyConsiderations>). If you use [Transfer Appliance](/transfer-appliance/docs/2.0/faq) (</transfer-appliance/docs/2.0/faq>), security keys that you control can help protect your data. Generally, we recommend that you engage your security team to ensure that your transfer plan meets your company and regulatory requirements.

For advanced network-level optimization or ongoing data transfer workflows, you might want to use more advanced tools. For information about more advanced tools, visit [Google partners](/storage/partners/) (</storage/partners/>).

The following links highlight some of the many options (listed here in alphabetical order):

- [Aspera On Demand for Google](https://asperasoft.com/fileadmin/media/Asperasoft.com/Product_Briefs/Google_AoD_AspiraPB.pdf) ([https://asperasoft.com/fileadmin/media/Asperasoft.com/Product\\_Briefs/Google\\_AoD\\_AspiraPB.pdf](https://asperasoft.com/fileadmin/media/Asperasoft.com/Product_Briefs/Google_AoD_AspiraPB.pdf)) is based on Aspera's patented protocol and suitable for large-scale workflows. It's available on demand as a subscription license model.
- [Bitspeed](http://www.bitspeed.com/) (<http://www.bitspeed.com/>) offers optimized file transfer protocol suitable for transferring large files or large number of files. These solutions are available as physical and virtual appliances, which you can plug into existing networks and file systems.
- [Cloud FastPath by Tervela](https://www.cloudfastpath.com/) (<https://www.cloudfastpath.com/>) can be used to build a managed data stream into and out of Google Cloud. For details, see [Using Cloud FastPath to create data streams](/solutions/partners/using-cloud-fastpath-to-create-data-streams) (</solutions/partners/using-cloud-fastpath-to-create-data-streams>).
- [Komprise](https://www.komprise.com/) (<https://www.komprise.com/>) can be used to analyze data across on-premises storage to identify cold data and move it to Cloud Storage. For details, see [Using Komprise to archive cold data to Cloud Storage](/solutions/partners/komprise-archiving-cold-data) (</solutions/partners/komprise-archiving-cold-data>).

- **Signiant** (<https://www.signiant.com/>) offers **Media Shuttle** (<https://www.signiant.com/products/media-shuttle/?cc=homepage>) as a software-as-a-service (SaaS) solution to transfer any file to or from anywhere. Signiant also offers **Flight** (<https://www.signiant.com/products/flight/?cc=homepage>) as an autoscaling utility based on a highly optimized protocol, and **Manager+Agents** (<https://www.signiant.com/products/signiant-manager/?cc=homepage>) as an automation tool for large-scale transfers across geographically dispersed locations.

For a large transfer, or a transfer with significant dependencies, it's important to understand how to operate your transfer product. Customers typically go through the following steps:

1. **Pricing and ROI estimation.** This step provides many options to aid in decision making.
2. **Functional testing.** In this step, you confirm that the product can be successfully set up and that network connectivity (where applicable) is working. You also test that you can move a representative sample of your data (including accompanying non-transfer steps, like moving a VM instance) to the destination.

You can usually do this step before allocating all resources such as transfer machines or bandwidth. The goals of this step include the following:

- Confirm that you can install and operate the transfer.
  - Surface potential project-stopping issues that block data movement (for example, network routes) or your operations (for example, training needed on a non-transfer step).
3. **Performance testing.** In this step, you run a transfer on a large sample of your data (typically 3&ndash;5%) after production resources are allocated to do the following:
    - Confirm that you can consume all allocated resources and can achieve getting the speeds you expect.
    - Surface and fix bottlenecks (for example, slow source storage system).

To help ensure the integrity of your data during a transfer, we recommend taking the following precautions:



- Enable versioning and backup on your destination to limit the damage of accidental deletes.
- Validate your data before removing the source data.

For large-scale data transfers (with petabytes of data and billions of files), a baseline latent error rate of the underlying source storage system as low as 0.0001% still results in a data loss of thousands of files and gigabytes. Typically, applications running at the source are already tolerant of these errors, in which case, extra validation isn't necessary. In some exceptional scenarios (for example, long-term archive), more validation is necessary before it's considered safe to delete data from the source.

Depending on the requirements of your application, we recommend that you run some data integrity tests after the transfer is complete to ensure that the application continues to work as intended. Many transfer products have built-in data integrity checks. However, depending on your risk profile, you might want to do an extra set of checks on the data and the apps reading that data before you delete data from the source. For instance, you might want to confirm whether a checksum that you recorded and computed independently matches the data written at the destination, or confirm that a dataset used by the application transferred successfully.

Google Cloud offers various options and resources for you to find the necessary help and support to best use Google Cloud services:

- **Self-service resources** (/solutions/migration-to-gcp-getting-started#self-service\_resources). If you don't need dedicated support, you have various options that you can use at your own pace.
- **Technology partners**  
([https://cloud.withgoogle.com/partners/?partnerTypes=SPECIALIZATION\\_PARTNER&sort-type=RELEVANCE&specializations=CLOUD\\_MIGRATION\\_SPECIALIZATION](https://cloud.withgoogle.com/partners/?partnerTypes=SPECIALIZATION_PARTNER&sort-type=RELEVANCE&specializations=CLOUD_MIGRATION_SPECIALIZATION))  
. Google Cloud has partnered with multiple companies to help you use our products and services.
- **Google Cloud professional services** (/consulting/). Our professional services can help you get the most out of your investment in Google Cloud.

There are more resources to help migrate workloads to Google Cloud in the [Google Cloud Migration Center](/solutions/migration-center/) (/solutions/migration-center/).

For more information about these resources, see the [finding help section](/solutions/migration-to-gcp-getting-started#finding_help) (/solutions/migration-to-gcp-getting-started#finding\_help) of [Migration to Google Cloud: Getting started](/solutions/migration-to-gcp-getting-started) (/solutions/migration-to-gcp-getting-started).

- If you have questions about devising a transfer plan or about a specific use case, you can contact [Google Cloud Support](/support-hub/) (/support-hub/), or contact your Google Account team directly.
- To get started on your transfer, we provide the following guides:
  - For general data migration strategies: [Migrating a monolithic application to microservices on Google Kubernetes Engine](/solutions/migrating-a-monolithic-app-to-microservices-gke) (/solutions/migrating-a-monolithic-app-to-microservices-gke)
  - For an offline transfer: [Transfer Appliance](/transfer-appliance/docs/2.0/how-to) (/transfer-appliance/docs/2.0/how-to)
  - For an online transfer from a public cloud: [Storage Transfer Service](/storage-transfer/docs/how-to) (/storage-transfer/docs/how-to)
- Try out other Google Cloud features for yourself. Have a look at our [tutorials](/docs/tutorials) (/docs/tutorials).