Technology Report
April 2020

DCIG

Decision Points and Best Practices for Performing Cloud-to-Cloud (C2C) Data Migrations

By Jerome M Wendt

wasabi
hot cloud storage

Table of Contents

Decision Points and Best Practices for Performing Cloud-to-Cloud (C2C) Data Migrations

Executive Summary

Roughly 85% of companies already store some data with cloud providers or plan to do so in the very near future. No one disputes that point. It is the cloud's cost and complexity that has a growing number of these same organizations wringing their hands. While they love the cloud's benefits, they hoped to lower their costs and simplify IT as they adopted it. Instead, they often experience the opposite—more costs and complexity.

Trying to remedy this situation using their existing cloud provider's solutions puts them between a rock and a hard place.

- If they put all their data on a provider's single storage tier, their costs may escalate to budget-breaking levels.
- Alternatively, if they distribute data across multiple storage tiers to lower costs, their applications may break, underperform, or both.
- Further, they must make time to configure, manage, and monitor these policies.

These challenges leave them yearning for why they initially signed up for the cloud: access to simple storage services.

In response to the demands, second-generation cloud providers that focus on storage have emerged. These providers offer a single, well performing, reliable tier of cloud storage at a single low price. This approach frees organizations to host their data in the cloud without worrying about its cost or complexity.

To turn these promises of these second-generation cloud providers into a reality, organizations must take three steps:

1. Identify data to store in a cloud that gives them the most bang for the buck
2. Identify a second-generation cloud provider to host this data
3. Get their data into the new cloud

Most organizations will find their archive and backup data the most logical data to move. This data represents the greatest percentage of data, if not all of it, that they store in the cloud. It poses low levels of risk to migrate and can potentially generate significant monthly cost savings once moved.

Performing a cloud-to-cloud migration of these data types has become easier to perform in recent years. Organizations may choose from any of four available data migration options that best align with their priorities. Many organizations will find they save both time and money by changing cloud providers. The savings realized from using a second-generation cloud provider can offset whatever costs the data migration may incur.

Executive Summary

(continued)

To achieve this goal, they need to choose a second-generation cloud storage provider that meets their needs. Wasabi Technologies represents such a provider. It gives organizations access to the affordable, high performance cloud storage that they seek. It also delivers on intangibles such as a provider network and integration with archive and backup provider software.

One cannot overstate the benefits of choosing a second-generation cloud storage provider and migrating their data to it. Organizations get a return to simple storage services while gaining access to sought-after simplicity, storage cost reductions, and performance improvements.

The Cloud is the Future

Organizations will use the cloud, in whole or in part, to support their overall IT approach and strategy. If that statement describes your company, you are not alone. A June 2019 survey of individuals from nearly 600 companies found that 84 percent of them hold a similar view.[1]

The cloud's broad level of adoption and acceptance across all organizational types comes as no surprise. Many consider using the cloud an integral part of their ongoing IT strategy when they consider its benefits, which include:

- 24x7x365 availability with five nines or more of uptime
- A pay-as-you-go on-demand business model
- Access to enterprise-class hardware and software features
- Elimination of managing on-premises IT infrastructure
- Enterprise-class data centers
- Theoretically unlimited capacity

These benefits, among many others, address the challenges that organizations routinely confront when managing and supporting its on-premises IT infrastructure. Yet as more organizations store their data in the cloud, they will encounter an old set of challenges.

Cost and Complexity . . . They're Back!

Unexpected costs and complexity top the list when it comes to storing and managing data in the cloud. Many organizations store far more data in the cloud than they initially anticipated. This leads to unexpected, recurring overruns that may grow month after month as a recent study revealed.[2]

Hidden and Unexpected Cloud Costs

Organizations ideally want to know how much the cloud will cost before they get the bill. Organizations largely only understand and budget for storing data in-house and on-premises. They may fail to grasp the extra fees that API operations, data retrievals, and egress fees, among others, incur.

These hidden fees lead to organizations paying far more than they expected or budgeted. Adding to their woes, many lack a reliable means to predict or budget what their cloud storage will cost going forward. These factors make them reticent to expand their use of the cloud.

The Hidden "Gotchas" of Storage Tiering

Storing data on lower cost storage certainly has appeal, especially when used to store infrequently accessed data. The "gotchas" show up when organizations move data between storage tiers. Data transfer fees may vary significantly and become nearly impossible to accurately estimate in advance. Consider:

Data Transfer Charges

First-generation cloud service providers charge anywhere from $0.02/GB to $0.09/GB to move data. The differences in fees stem from a few factors. These include:

- The amount of data transferred
- If they transfer data within or across regions
- If they transfer data within or outside availability zones
- If they transfer data inside or outside the provider's cloud

These variables make it difficult to accurately estimate data transfer costs.

Other fees that organizations may incur include:

- **API / Data Request Fees.** First-generation cloud service providers charge for API calls such as write actions (PUT/COPY/POST) and read actions (GET). These actions can quickly incur hidden costs behind the scenes.
- **Egress Fees.** First-generation cloud providers charge up to nine cents per GB for data transferred out of the cloud or even between availability zones.
- **Data Retrieval Fees.** If organizations retrieve data from archival storage tiers such as AWS Glacier, they incur additional data retrieval fees. These fees range from one to three cents per GB.

1. Eagle, Liam. Hybrid and Multicloud Strategies Accelerate Digital Transformation. June 2019, link, pg. 9.
2. https://www.softchoice.com/about-softchoice/news-and-insights/research/state-of-cloud-readiness-study. Referenced 12/3/2019.

Storage Tiering Rarely the Answer

To lower cloud storage costs, organizations turn to using storage tiering that many cloud providers make available. However, doing so re-introduces complexity and risk back into cloud storage management.

Storage tiering requires that organizations take time to investigate and understand each storage tier. They must determine which applications support these tiers. They must create policies that handle placing data across their various storage tiers.

Even if they implement these policies, organizations receive no guarantee storage tiering will operate flawlessly. Applications and users cannot always handle the slower response times associated with some lower cost storage tiers.

Organizations rightfully want to avoid these hassles associated with storage tiering. To realize this ideal, organizations should seek out an economical, well-performing cloud storage solution that keeps cloud storage costs low and management simple.

The Second Generation of Cloud Providers

In response to this demand for a simpler cloud storage solution, a second generation of cloud providers has emerged. These second-generation providers focus on being "best of breed" for cloud infrastructure components such as compute and storage.

As part of making cloud storage easier to use and manage, second-generation cloud providers view data differently. They regard all data, regardless of its age, classification, or origin, as equally valuable. Translated, they esteem the value of the data greater than the value of the storage.

This value-centric view of data impacts the design of their cloud storage infrastructure. They place all data on a single storage tier that affords it the same level of availability, durability, performance, and reliability.

Taking this approach, these providers optimize their cloud storage offering to deliver it at an attractive price point. This makes it practical for organizations to keep all their data on a single storage tier. This, in turn, results in less complexity and reduced management in the environment.

Expected Benefits of Using a Second-generation Cloud Storage Provider

Organizations looking to adopt a second-generation cloud storage provider should set the proper expectations.

Their storage compares favorably to AWS's S3 Standard Infrequently Accessed (S3-IA) tier. The AWS S3–IA storage tier costs less than the AWS S3 Standard tier and delivers millisecond response times. Second-generation cloud storage providers typically best this offering. Their storage offers comparable or better performance than S3–IA at one-half to one-fifth its cost.

While cost is easy to measure, measuring performance becomes more subjective. Organizations will likely expect their application performance to improve once they move their data to the new platform.

In most cases, performance improvements will occur. However, these improvements will depend upon variables ranging from network connectivity to changes in the underlying storage tier. Organizations should minimally expect the second-generation provider to deliver the same performance at a lower cost.

Organizations should expect to see reduced storage management complexity and lower costs. Further, organizations do not need to worry about data egress, transfer, or retrieval fees. These providers do not charge any of these fees.

Choosing the Best Cloud-to-Cloud Data Migration Strategy

Once an organization decides to migrate its data, it must decide which cloud-to-cloud data migration option or options to pursue. While no easy button exists that magically moves data to the new cloud, proven methodologies do exist. An organization may follow these proven options that use established tools and approaches to simplify data migrations.

Organizations have more options available to them than they may realize. These options give them leeway to take both business and technology factors into consideration as they make their final choice. The four most common options for migrating their data from one cloud to another include:

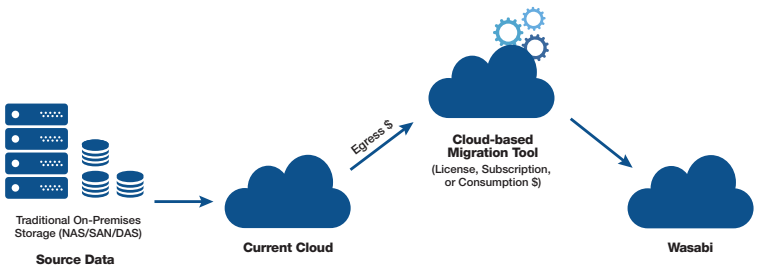
1. All-cloud Internet-based Data Migration
2. All-cloud Direct Connect Data Migration
3. Self-hosted Cloud Migration
4. Source Data Migration

The most appropriate choice depends on several factors. These include one-time and ongoing costs; performance; and, how well the option aligns with each organization's overall technical and business goals.

Option 1: All-cloud Internet-based Data Migration

Using an all-cloud internet-based data migration, an organization migrates data directly from one cloud provider to the other. This technique moves the data while keeping minimal or no data on-premises for any length of time.

OPTION 1 All-cloud Internet-based Data Migration



Source: Wasabi

Organizations will primarily find this approach appealing as they may use their existing internet connection. In the simplest scenario, an internet-based migration could equate to performing a cut-and-paste of an organization's data.

Considerations

Prior to initiating an internet-based migration, organizations need to carefully weigh the following items:

- ***Amount of data to migrate.*** The more data one migrates, the more likely it could negatively impact the performance of applications that access the internet or data in the cloud.
- ***Impact of disruptions in connectivity during the migration.*** Degradations or disruptions in internet service may stop a data migration and force an organization to re-initiate the process.
- ***Higher egress fees.*** An organization may pay up to 7x more per GB when using an internet-based migration versus direct connect.
- ***Temporary storage required on-premises.*** The amount of available on-premises storage needed to cache data being migrated may impact how quickly the data migration completes.
- ***Bandwidth of internet connection.*** The size of the pipe to the internet will impact the speed and duration of the migration.
- ***Performance.*** The data migration duration will depend upon the size of the data set and the network connection to the new cloud.

Recommendation

An internet-based migration frequently makes the most sense when an organization has smaller amounts of data to migrate (less than 100TB.) If there's more than 100TB of data to migrate, verify the organization can tolerate a longer data migration time if it will impact production applications during peak times.

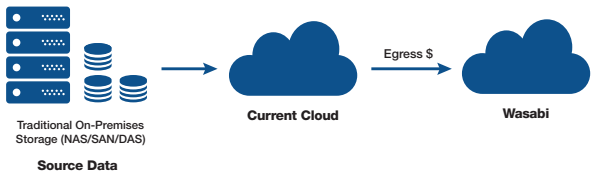
Key Questions to Ask

- How much data is there to migrate?
- Can migrations of individual large files or objects be re-started at the point where an interruption in the migration occurred?
- Can production applications sustain interrupted or delayed connections to data in the cloud?
- What are the internet connection's average loads throughout the day?

Option 2: All-cloud Direct Connect Data Migration

A direct connection often provides organizations the fastest means to migrate large amounts of data quickly between two clouds. It consists of a direct wide area network (WAN) connection between the first- and second-generation cloud providers.

OPTION 2 All-cloud Direct Connect Data Migration



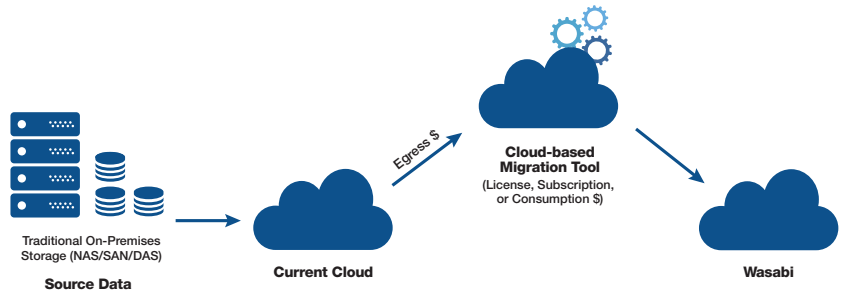
Source: Wasabi

This approach provides the following three key benefits for organizations.

- ***It reduces the time of the cloud-to-cloud data migration.*** Organizations have unfettered access to the full bandwidth of the direct connection between the two clouds.

- ***It reduces the amount of egress fees charged.*** Using a direct connection, organizations may migrate more data at a lower egress fee per GB.
- ***Experience the second-generation cloud provider's benefits more quickly.*** Organizations can more quickly realize the new cloud's benefits of lower costs, improved performance, and reduced complexity.

OPTION 3 Self-hosted Data Migration



Source: Wasabi

Considerations

While a direct connect can accelerate cloud-to-cloud data migrations, organizations should consider the following factors before implementing it.

1. ***Count how many applications access and store data in the cloud.*** The more applications that access data in the cloud, the more complicated planning the data migration becomes.
2. ***A direct connect incurs additional charges.*** Direct connect charges vary by cloud provider with charges potentially including a setup fee and an hourly or monthly service fee.
3. ***Allow time to set up a new direct connect.*** Allow at least a week to put a direct connect in place between two clouds.

Recommendation

Organizations that must quickly and cost-effectively migrate a large amount of data (100TB or more) should prioritize using a direct connection. It simplifies the planning of the data migration and expedites its execution. Archival and backup data represent the best use cases for this data migration option.

Key Questions to Ask:

- How much data is there to migrate?
- How quickly can a direct connect be set up between the two clouds?
- How many applications access and store data in the cloud?
- Which applications access and store data in the cloud?

Option 3: Self-hosted Data Migration

Using a self-hosted data migration, organizations first move data from their existing cloud to on-premises storage. Once on-premises, they then move the data to the new cloud.

Benefits

This approach resembles the all-cloud internet-based data migration option in that an organization uses its existing internet connection to move its data. However, it differs in two ways.

1. ***It breaks the data migration process up into two parts.*** An on-premises storage cache houses data coming from the source first-generation cloud for an extended time. Should any disruptions occur, the data migration process may continue as some level.
2. ***It reduces the need to restart copies of large files or objects.*** Using on-premises storage decreases the distance the file must initially travel. This improves the odds of it completing each leg of the migration.

Considerations

Prior to repatriating data on-premises as part of a cloud-to-cloud data migration, organizations need to carefully weigh the following items:

- ***May need a large amount of storage capacity on-premises.*** An organization will need to determine how much data it will migrate and how long it will retain it on-premises. The on-premises storage must be sufficiently large and possess the needed availability and performance characteristics.
- ***Setup time.*** An organization will minimally need to configure some storage on-premises to host data coming from the source cloud.

- ***Egress fees.*** The first-generation cloud provider will still charge egress fees for data coming out of its cloud.
- ***Time to complete the migration.*** All data will go over an internet connection as opposed to a direct connection. This will likely increase the time it takes to complete the migration.

Recommendation

Organizations should use this option when a direct connection is not an option. An organization must allow time for the migration to complete and confirm it will not impact production applications during their peak times.

Key Questions to Ask

- How many large files are there to migrate?
- Is direct connect an option?
- How much storage space is needed on-premises?
- What are the peak times of usage for the organization's internet connection?

Data Migration Tools: Methods and Costs

As organizations consider performing a data migration, it behooves them to consider software tools to migrate the data. These tools migrate data in one of two ways:

- **Migrates** all data from an existing cloud to new cloud for cost-savings, performance, or other reasons. Once it migrates the data, it deletes it from the old cloud.
- **Duplicates** data from the existing cloud to the new cloud. It keeps data in both locations for availability, cost, and/or performance reasons.

Data migration tools do incur fees above and beyond any fees incurred migrating out of the first-generation provider's cloud. While they incur fees, they help accelerate data migrations and make the migrations more reliable. Providers license these tools in at least two ways.

1. **Cloud-hosted.** Providers host their tools in public clouds and charge per-GB or per-TB transfer fees. These fees cover compute and bandwidth costs as well as the cost of the tool.
2. **Self-hosted.** Organizations pay a licensing fee to host the tool in their environment.

Examples of Data Migration Tools

Cloud-hosted Tools

- Acembly C2C
- CloudFuze
- Cloudsfer
- Flexify.io

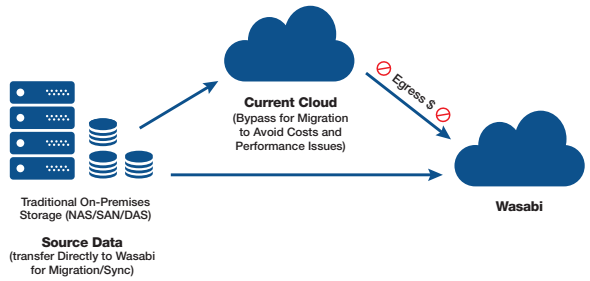
Self-hosted Tools

- AWS CLI
- NetApp CloudSync
- Nodeum
- Scality Zenko
- S3 Tools S3cmd

Option 4: Source Data Migration

A source data migration requires no formal data migration from one cloud to another. Rather, organizations simply stop storing new data with their existing first-generation cloud provider and start storing data with their new second-generation cloud provider.

OPTION 4 **Source Data Migration**



Source: Wasabi

Organizations may choose this option if they determine it costs too much to download data from their current cloud provider. In this case, organizations simply re-upload source data to the new cloud directly from their original data application or system.

Benefits

A source data migration eliminates many of the hassles associated with executing a data migration. It takes minimal time to implement. It incurs no egress fees. Organizations simply point their on-premises applications or systems at the new cloud and configure them to start storing their data on it.

Considerations

Abandoning data in the existing cloud to use the new cloud as a storage target should only be done after careful consideration.

- ***Limited use cases.*** Organizations will need to verify they can abandon all their data. Backup data may allow for this use case.
- ***Cloud storage costs still accrue.*** This option may leave data orphaned with the first-generation cloud provider until deleted. Organizations will continue to pay for data stored with the first-generation cloud provider until they delete it.
- ***Must be able to complete a full first backup of all applications.*** Doing a new first backup to the new cloud could take multiple days or longer to complete.

Recommendation

Organizations should only perform source data migrations after they authoritatively confirm they can abandon data residing in their first-generation cloud. They should realistically only consider doing a source data migration if they only use their existing cloud as a backup target.

Key Questions to Ask

- Are there any other applications beside backup that require a data migration?
- Is only backup data currently stored in the existing first-generation cloud?
- Can the organization schedule and complete an initial, full back up of all its applications?
- How quickly can the organization delete data stored with the first-generation cloud provider?

Best Practices for Performing a Cloud-to-Cloud Data Migration

Once an organization establishes which cloud-to-cloud migration strategy or strategies it intends to perform, it must put a few more pieces in place to complete the data migration. Minimally, it should take the following five steps to successfully execute upon the data migration.

Step 1—Contact and Involve All Application Owners

Application owners want and need to know if their organization intends to migrate their data to another cloud. A data migration could impact application owners in multiple ways. They will likely experience or observe changes in billing, management, and application performance during and after the migration.

Variables that Influence the Choice of a Data Migration Strategy

Multiple variables will influence the choice of the "best" cloud-to-cloud data migration strategy for any organization. While not comprehensive, these seven items will influence an organization's choice of its cloud-to-cloud data migration strategy.

Available budget. If migrating large data amounts (100TB+) it may be more economical to use a direct connection. Even though it incurs a charge, direct connect egress fees are much lower than egress fees associated with using an internet connection.

Time constraints. An organization should consider using the cutover or direct connect options if it has lots of data to move and little time to move it.

Staff skills. Staff will need the skills to successfully execute a data migration.

Data migration tools. These tools help to automate the process, optimize data flows, and recover from any disruptions in network service.

Amount of data to migrate. The amount of data to migrate may encourage organizations to migrate their data more quickly. In this way, they can more quickly lower their storage costs.

Network connection. Internet connections with ample, available bandwidth make it more feasible for organizations to this connection to do a data migration.

How and when applications access and use data. Organizations must establish when and how their applications access and use data in the cloud.

Step 2—Optimize the Network

The amount of network optimization that an organization performs largely depends upon the type of data migration that it performs. If it uses a direct connect or cutover as its data migration strategy, very little network optimization is needed. If the organization uses internet-based or repatriation to migrate data, it will need to carefully analyze and understand its current network traffic.

Step 3—Leverage Service Provider Support

Organizations will ideally engage a service provider to assist them in the migration. Many organizations do not regularly perform data migrations and may overlook or miss key steps. A service provider can help guide them through the data migration process and avoid common pitfalls.

Step 4—Monitor Migration Progress

Once the data migration begins, monitor network bandwidth consumption. If using a direct connect, maximize data throughput and keep the network pipe as full as possible. If using an internet connection, balance data migration traffic with the traffic of its other applications. If the data migration starts to impact production applications, throttle back the data migration. Conversely, during periods of low application activity, increase the data migration transfer rate.

Step 5—Data Migration Validation and Clean-up

Once an organization completes the migration of its data to the new cloud, a few more tasks remain. Validate that all applications can access and use the new cloud. "Delete", or schedule for "deletion", any data in the first-generation cloud that it no longer needs. Finally, decommission any direct connects or temporary storage used during the data migration.

A Return to Simple Storage Services

Organizations ready for a return to simple storage services should look to second-generation cloud providers to deliver them. They make implementations simple. They simplify pricing. Even hosting data on them may be relatively simple depending on the data type they first want to host and the migration method they select.

Start with Archive and Backup Data

Organizations that store large amounts of archive and backup data in the cloud should prioritize migrating these two data types first. They represent the fastest growing sets of data that organizations can most easily migrate.

Being both cost and performance sensitive, they are well-suited for migrating to and hosting with a second-generation cloud provider. These cloud providers deliver the low costs and performance that these two data types require minus the management complexity of first-generation clouds.

Wasabi Hot Cloud Storage

Wasabi Hot Storage illustrates the type of simple storage services that organizations should expect. Wasabi focuses exclusively on cloud storage that provides a single storage tier that organizations may access via its S3-compatible storage interface. Its single storage tier compares favorably to AWS's S3–IA storage tier but costs half its price.

Wasabi can help step companies through the data migration process. It partners with multiple service providers. These providers will analyze any organization's data migration needs to help them plan and accelerate the data migration process.

Further, Wasabi's relationships with multiple archive and backup providers ensure that organizations store their data in a cloud certified with their respective offerings.

The Wasabi Ball

Wasabi also offers one other data migration option for organizations. Organizations may rent a transfer appliance called a "Wasabi Ball." Designed to transfer large amounts of data, this physical device contains 100 TB of storage capacity. In instances where organizations have more than 100 TB, they may rent multiple appliances to move larger amounts of data.

Once they receive the appliance, they simply connect the Wasabi Ball to their local network. They then follow its instructions to select files and data on their local network or in the cloud that they can transfer to the appliance. Once transferred, the organization then sends the appliance back to Wasabi. Wasabi completes by the migration by ingesting the data from this appliance directly into the organization's Wasabi storage bucket.

Organizations that have internet connections of less than 1Gbps may find this option very appealing. They do not have to consume bandwidth on their existing connection to migrate data to Wasabi. It also costs only a minimal amount to rent the appliance. It is about $300 for up to two weeks within the US with fees varying outside of the US.

Second-Generation Storage Clouds—Their Time Has Come

Organizations should stop feeling trapped by the cost and complexity of their first-generation cloud storage solutions. Second-generation cloud providers such as Wasabi specialize in offering simpler, more economical, and higher performing cloud storage options.

Wasabi simplifies managing storage in the cloud while lowering costs that help offset any costs organizations may incur in migrating to its cloud. These factors, taken together, explain why providers like Wasabi stand poised to re-introduce "simple storage services" back into the cloud storage conversation. ■

About DCIG

DCIG empowers the IT industry with actionable analysis that equips individuals within organizations to conduct technology assessments. DCIG delivers informed, insightful, third party analysis and commentary on IT technology. DCIG independently develops and licenses access to DCIG Buyer's Guides. It also develops custom content in the form of blog entries, executive white papers, podcasts, competitive intelligence reports, webinars, white papers, and videos. More information is available at **www.dcig.com.**

DCIG, LLC // 7511 MADISON STREET // OMAHA NE 68127 // 844.324.4552
dcig.com