Final Proposal

INFO 540 IT Architecture & Infrastructure (Fall 2022)

Group 3

Tracy Chan, Bill Chen, Gray Li, Songyun Xia, Tianyu Xie

Dec 2022

1 Introduction

1.1 Business background

We work for a mid-sized company with roughly 25,000 employees. Approximately 5 years ago, our company adopted a cloud-based software as a service (SaaS) file storage and collaboration solution in an effort to modernize and streamline collaboration across our company.

The solution has served the company well, making access to files and folders from remote locations simpler and easier. As a result, the system is widely adopted across the enterprise. Our company now has over 2 petabytes of data stored in the system across tens of thousands of shared folders with complex access rights that have developed over time.

When our company signed the original contract with the SaaS vendor, it provided unlimited data storage for a flat rate. The IT organization's existing IT budget covers the cost of the system for the entire company, so the service is provided "for free" to internal users.

Our existing contract will expire in just under 1 year, and the procurement team is reviewing the terms of the new contract. The SaaS vendor has indicated that unlimited storage is no longer an option. They have moved to a new pricing model that charges a standard rate per terabyte per year. This change will cause our company's annual cost for the service to balloon from $125k per year to over $1.5M per year.

1.2 Assumptions

The company is using a dedicated T3 line at 50Mb/s.

Currently, the company has 2 Pb of data stored on the SaaS cloud. The employees, on average, use 25 Gb of the SaaS cloud storage (625,000 Gb in total). The remaining 1.375 Pb are shared data which combined 85% archive data and 15% high-frequent used data.

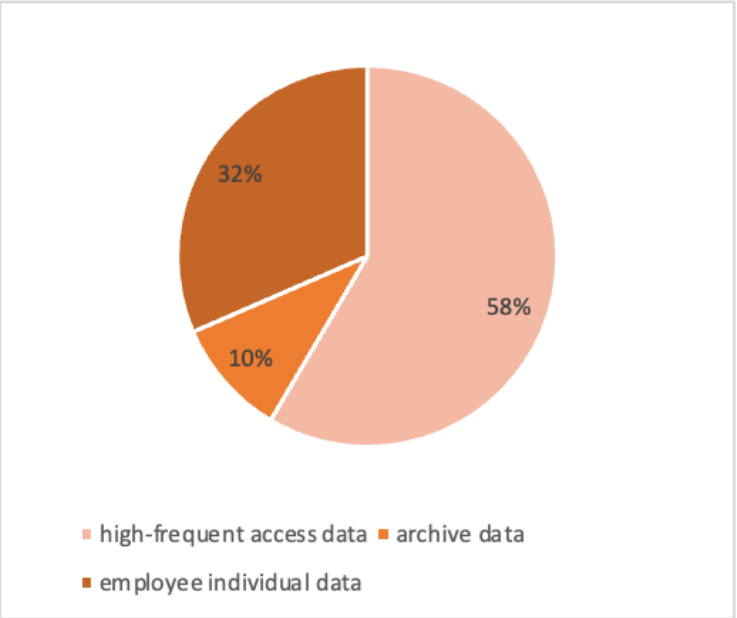

Figure 1: Proportion of 2 PB data

The company is using Gmail. Gmail provides 50 Gb of google drive storage for each account. The employees hardly use the enterprise Gmail account for storage.

1.3 Objectives

Archive storage: develop a secure, cost-effective archive storage service to house old data that is accessed infrequently.

On-premise storage: develop an on-premise storage solution to house large data sets that must be accessed frequently but are used by only a handful of individuals.

Migration tools and processes: establishing migration tools and processes for moving individual user content over to the cloud storage provided by your email vendor while maintaining access rights on the content.

2 Archive storage

After classifying the data, we first want to adopt a method to archive data. Our team identify three options including 2 could service and 1 local tape library method.

2.1 Option 1 AWS S3 Glacier Deep Archive

The Amazon S3 Glacier storage classes offer the best performance, quick retrieval options, and affordable archive storage in the cloud since they were created specifically for data archiving. It offers three archive storage classes tailored for various access patterns and retention times.

The cheapest storage for long-lived archive material accessed less than once a year and recovered asynchronously is provided by **AWS S3 Glacier Deep Archive**. S3 Glacier Deep Archive offers the lowest cost storage in the AWS cloud at pricing substantially cheaper than storing and preserving data on-premises tape or off-site archiving, at just $0.00099 per GB-month (or $1 per TB-month). It is a less expensive and simpler alternative to tape storage.

Optional choice:

We just considered the optimized choice when using AWS, but it has other choices that we also consider.

When long-lived data has to be retrieved once every three months and millisecond retrieval is required, S3 Glacier Instant Retrieval offers low-cost storage. S3 Glacier Instant Retrieval provides comparable low latency, high throughput, and durability to S3 Standard-IA at a little lower per-GB storage cost and higher per-GB retrieval cost.

For archive material accessed 1-2 times per year and recovered asynchronously, S3 Glacier Flexible Retrieval offers low-cost storage up to 10% less expensive (than S3 Glacier Instant Retrieval). For archive material that does not need immediate access but needs the flexibility to retrieve sizable data sets without incurring any fees, such as backup or disaster recovery use cases, S3 Glacier Flexible Retrieval is the best storage class. The most cost-effective retrieval solutions with access times ranging from minutes to hours are provided by S3 Glacier Flexible Retrieval, which also offers free bulk retrievals.

These three classes are all designed for 99.999999999% (11 9s) of data durability and 99.9% availability by redundantly storing data across multiple physically separated AWS Availability Zones in a given year.

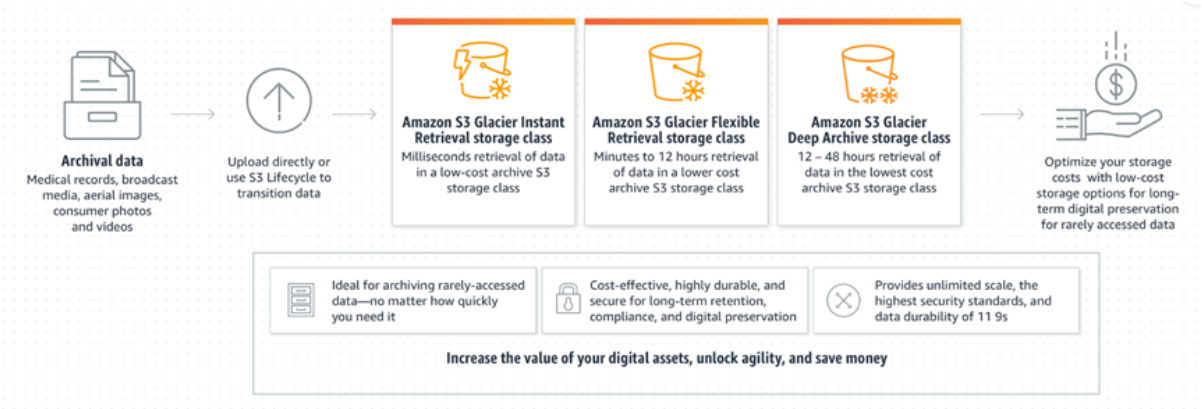

Figure 2: AWS using process

Cost:

The pricing of AWS S3 Glacier Deep Archive is 0.00099 per GB / month, and we have 1.17 PB data that needs to be archived.

So the cost of using AWS to store is $14574.787/year (1.17*1024*1024*0.00099*12)

We also need to calculate the retrieval fee for using AWS S3 Glacier Deep Archive. Here we assume that every data will be retrieved once a year(since most archive data won't be retrieved once a year, the exact cost actually is much better than the assumption). We don't require the retrieval time for archive data so we will choose Bulk retrieval. So the cost to retrieval is $3067.0848/year(1.17*1024*1024*0.0025)

	PUT, COPY, POST, LIST requests (per 1,000 requests)	GET, SELECT, and all other requests (per 1,000 requests)	Lifecycle Transition requests into (per 1,000 requests)	Data Retrieval requests (per 1,000 requests)	Data retrievals (per GB)
S3 **Glacier Deep Archive*****	$0.05	$0.0004	$0.05	See below	See below
Standard	n/a	n/a	n/a	$0.10	$0.02
Bulk	n/a	n/a	n/a	$0.025	$0.0025

Figure 3: The pricing of AWS S3 Glacier Deep Archive retrieval

The total cost is $17641.8718/year

Security for Amazon S3:

We can store our data in Amazon S3 and secure it from unauthorized access with encryption features and access management tools. S3 is the only object storage service that allows us to block public access to all of our objects at the bucket or the account level with S3 Block Public Access. S3 maintains compliance programs, such as PCI-DSS, HIPAA/HITECH, FedRAMP, EU Data Protection Directive, and FISMA, to help us meet regulatory requirements. AWS also supports numerous auditing capabilities to monitor access requests to our S3 resources:

- Block Public Access
- Object Lock
- Object Ownership
- Identity and Access Management
- Amazon Macie
- Encryption
- AWS Trusted Advisor
- AWS PrivateLink for S3
- Verify data integrity

2.2 Option 2 Google Cloud Storage

Google Cloud Archive Storage is an affordable, reliable storage option for disaster recovery, online backup, and data archiving. Our data is accessible in milliseconds rather than hours or days, as with the "coldest" storage options provided by other Cloud providers.

The availability of archive storage is lower than that of Google Standard storage. In addition to greater data access and management fees, a 365-day minimum storage period is also required for archive storage. When we intend to retrieve data less frequently than once a year, archive storage is the ideal option.

Table 1: Availability of Archive storage data

Location Type	Availability SLA	Typical monthly availability
multi-region	99.9%	99.95%
dual-region	99.9%	99.95%
region	99.0%	99.9%

Optional choice:

Coldline Storage offers exceptionally affordable, highly reliable storage for data that is accessed rarely. When somewhat poorer availability, a 90-day minimum retention period, and higher data access fees are acceptable trade-offs for cheaper at-rest storage costs, coldline storage is perfect.

Google Cloud recommended products and security capabilities to help organizations achieve a strong security posture and protections for their Google Cloud environment.

Cost:

The pricing of Google Cloud Archive Storage is 0.0012 per GB / month, and we have 1.17 PB data that needs to be archived.

So the cost of using Google Cloud to store is $17666.4084/year (1.17*1024*1024*0.0012*12).

We also need to calculate the retrieval fee for using Google Cloud. Here we assume that every data will be retrieved once a year(since most archive data won't be retrieved once a year, the exact cost actually is much better than the assumption). The pricing is $0.05 per GB, so the cost to retrieval is $61341.696/year(1.17*1024*1024*0.05)

The total cost is $79008.1044/year.

Security for Google Cloud:

Google Cloud is responsible for providing infrastructure security, which includes security through the entire information processing life cycle including hardware infrastructure, service deployment, storage services, user identity, internet communications, and operational and device security.

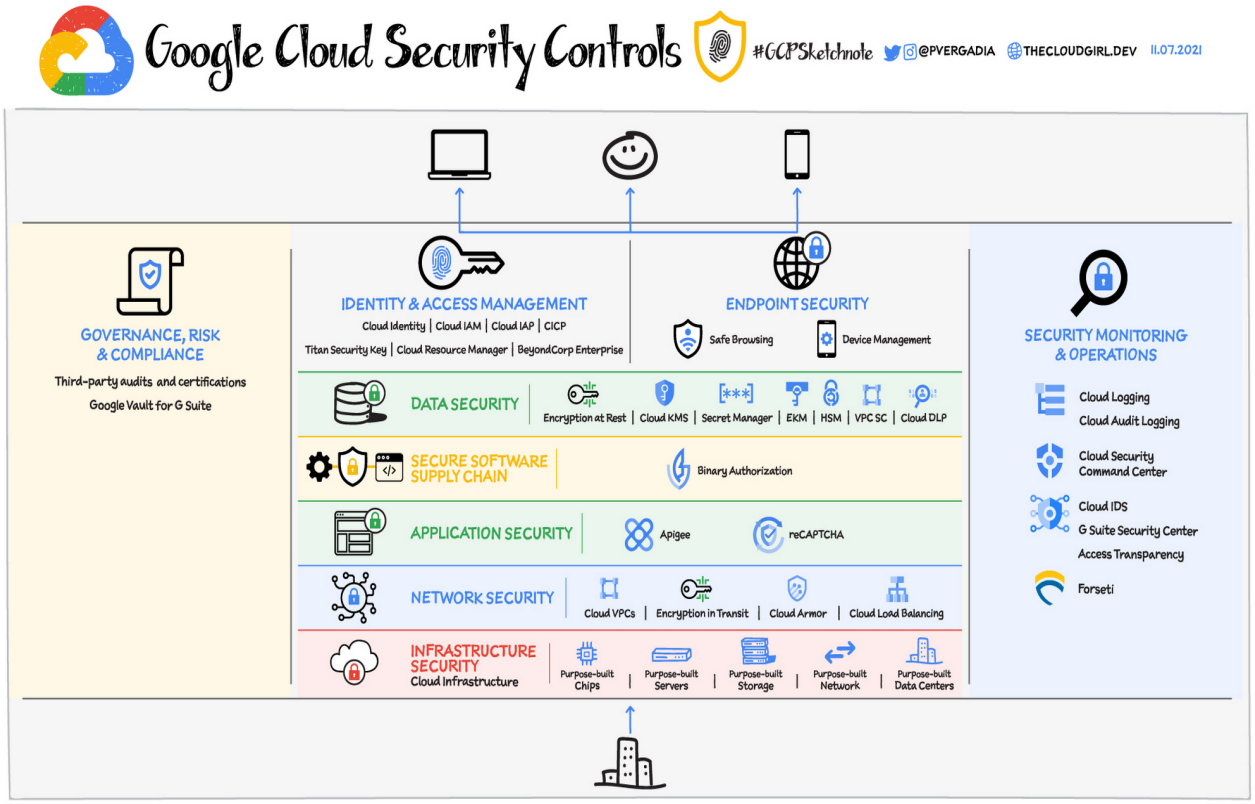

Figure 4: Security architecture of Google Cloud

2.3 Option 3 IBM TS4300 Tape Library

IBM TS4300 Tape Library is a high-density, highly scalable, and simple-to-manage solution intended to keep data safely stored for a long time while lowering the expenses associated with data center space and utilities. We can extend vertically up to seven modules with extension for Linear Tape-Open (LTO) Ultrium cartridges, drives, and redundant power supplies thanks to its modular architecture, which enables you to increase cartridge and drive capacity as necessary. The improved data protection provided by IBM TS4300 aids in meeting security and regulatory standards.

To use IBM TS4300 Tape Library, we need to consider the following:

- Tape drives: 12 TB native/30 TB compressed (Ultrium 8)
- Hot-swap components: power supplies, tape drives
- Systems management: Storage Management Initiative Specification (SMI-S) enabled
- Limited warranty: Three-year in-warranty coverage of select parts
- Weight: Base unit maximum 28 kg (61.5 lb); Expansion unit maximum 22.6 kg (49.6 lb)
- Dimensions (H x W x D) 114 mm (4.48 in.) x 446 mm (17.56 in.) x 873 mm (34.37 in.)
- base model and with 6 expansion modules
- Management software options include:

  IBM Spectrum Archive™
  IBM Spectrum Protect™

Cost:

After estimation, we need to purchase components as follows:

- LTO 8 full-height fiber channel drive
- 2.0M Mini-SAS/Mini-SAS 1X Cable *12
- 10-meter OM3 fiber Cable (LC) *12
- Ultrium 8 Data Cartridge (5-Pack) *10
- Path Failover
- LTO Library Managed Encryption
- IBM Spectrum Archive Library Edition
- IBM Spectrum Protect Suit
- SAS tape drive support (Required)
- Fiber tape drive support

The total cost is $26,770.00. However, we need to add some extra costs, including power, space rent, the cost of temperature control, etc.

Security for IBM TS4300 Tape Library:

All supported tape drives in this library support encryption.

The encryption-enabled drive contains the hardware and firmware to encrypt and decrypt host tape application data. The host application or host server provides an encryption policy and encryption keys. A drive digital certificate is installed at manufacturing time. Each drive receives a unique serial number and certificate. The T10 application might validate each drive instance by checking the drive's digital certificate.

The library provides these options.

- Encryption disabled
- Application Managed Encryption (AME)
- Library Managed Encryption (LME). LME is a built-in feature that is enabled by using a purchased license. The LME feature can be ordered from the factory or ordered as a field upgrade. To order a feature, contact your IBM Sales Representative or Business Partner. See Optional Features. For configuration details, see Configuring Library Managed Encryption.

2.4 Summary

Table 2: Archive storage plan comparison

	AWS glacier	Google Cloud	IBM TS4300
storage fee	$14574.79/year	$17666.40/year	at least $26,770
extra cost	no	no	yes
retrieval fee	$3067.08/year	$61341.70/year	none
retrieval delay	within 12 hours	within milliseconds	within 5 minutes
security	acceptable	acceptable	acceptable
availability	high	high	low
performance	high	high	middle
ease of implementation and use	easy	easy	difficult

After analysis and comparison, we decided to use AWS S3 Glacier Deep Archive.

Regarding the cost, the first is the storage fee, which we think highly of. The cost of AWS is the lowest. Mentioned that since the cost to implement a tape library is one-time, the total price will be lower after2 years. But considering the extra cost of using a tape library, this time would be much longer. (Probably 4 years)IBM TS4300 doesn't have a retrieval fee; however, AWS glacier only costs $3067.08/year, and, as we mentioned before, the actual cost may be much less than we estimate. Although the retrieval delay time of AWS is the longest, since we have little requirement for retrieval, we don't need to worry about deletion delay time.

Next, we also estimate the non-functional attributes. For security, each solution has a security policy, and the risk is acceptable. For availability, since we may be out of power or infected by other factors, we cannot promise using a tape library is as highly available as cloud storage.

For performance, since we used cloud services instead of tape library before, our IT group is more familiar with the cloud. We can cause some human error and spend more time solving it. It's easier for us to implement and use other cloud services.

As a result, among cloud storage and tape library, we prefer cloud service. Then, among several cloud vendors, we want to choose the cheapest and also secure vendor - AWS S3 Glacier Deep Archive.

3 On-premise storage

This section will cover the development of an on-premise storage system for holding sizable data sets that must be accessed regularly but are only utilized by a select few people. There are often two approaches to managing the frequently used data: building our own data storage or outsourcing them to already-existing multi-tenant data centers.

While there may be differences in the storage possibilities, the other modifications to the new storage technique are generally identical.

Access rights reorganization

The first step is to rearrange how data is currently stored. The complexity of our access privileges now comes from the earlier habit of keeping data across several shared files. Each department should have a dedicated folder with information specific to that department. Additionally, access privileges need to be managed with extra caution. We save data according to time and project, for instance.

Access to these folders should also be rigorously controlled; as a result, once a project is finished and no more changes are required, the files should no longer be available to the person who worked on them.

Employees naturally only have access to the data they need to work with. HR would be involved in this reformation as the database would be periodically updated by a team of specialists, resulting in some staff changes. People need to be educated on the value of information security, as evidenced by the fact that 2 petabytes of data are currently being kept in the system over tens of thousands of shared folders. When saving these data, employees were careless. Additionally, it is crucial to remind all staff members not to allow access to someone they shouldn't.

3.1 Option 1: DAS(Direct Attached Storage)

Digital storage immediately connected to the host computer that will access it is known as direct-attached storage (DAS).

For data IO read/write and storage maintenance management, DAS storage depends heavily on the server host operating system. Data backup and recovery need resources from the server host (including CPU, system IO, etc.). Data streams must return to the host before going to the hard disks (or libraries)

Occasionally, all needed to use this infrastructure is to connect the storage to the workstations without an internet connection. SCSI connections are typically used as the connection route between the server host and directly attached storage. The SCSI channel will become the IO bottleneck as the server CPU's processing power, storage hard disk space, and the number of hard disks in the array increase steadily. This is because the server host's SCSI ID resources are finite, as is the number of SCSI channel connections that can be made.

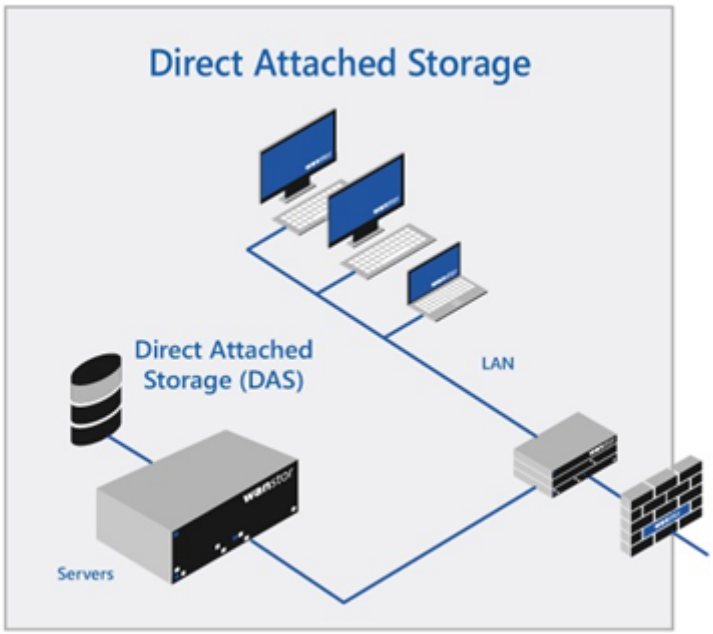

Figure 5: DAS conceptual model

The access control is simple, just do not connect those PCs which need no access right.

Cost:

One storage could carry 120 TB HDD or 4.8TB SSD. If we use Lenovo SSD, the cost is $300000 ((2000+12*700)*(150/4.8)). But if we use HDD, the cost could be $10800(2*2000+17*400).

3.2 Option 2 NAS(Network Attached Storage)

In the NAS storage structure, the storage system is directly connected to the network through the network interface and is accessed by users over the network rather than being connected to a server or client through the I/O bus.

NAS refers to the addition of storage devices to a group of computers using a common network structure (such as Ethernet). It focuses on assisting departments and working groups in addressing the demand for quickly expanding storage capacity. Users now utilize NAS to share documents, photos, movies, and other media. Some NAS manufacturers have included cloud storage as a result of the growth of cloud computing, considerably easing use by businesses and individual users.

Cost:

We will need $21800(3*5000+17*400) to purchase storage devices and $14000 to purchase Ethernet Switch. So the total cost is $35800

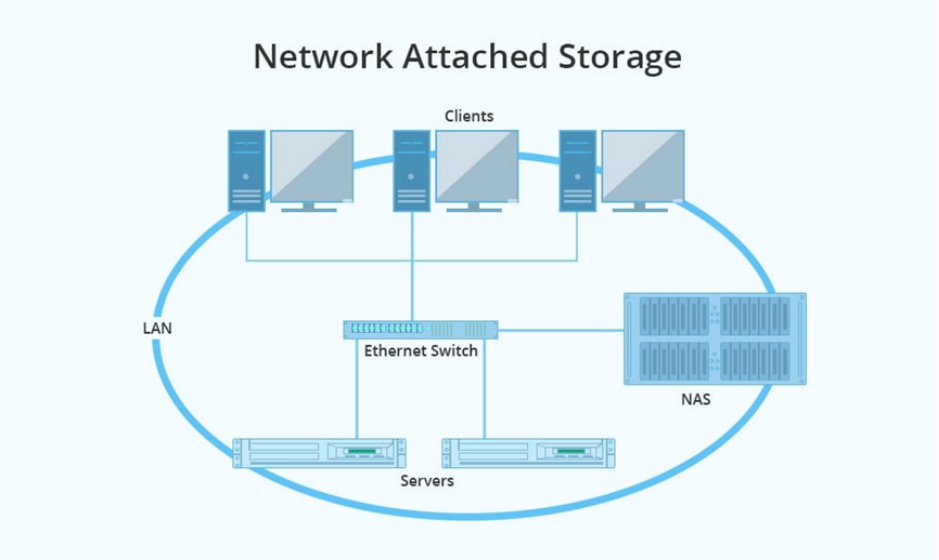

Figure 6: NAS conceptual model

3.3 Option 3 SAN(Storage Area Network)

A storage area network, or SAN, is a specialized, fast network that gives storage devices network access. Hosts, switches, storage components, and storage devices make up standard SAN configurations. These components are connected via various technologies, topologies, and protocols. SANs may connect several locations.

A SAN offers storage devices to a host to give the impression that the storage is locally attached. Adopting various virtualization techniques allows for this streamlined storage display to a host.

Storage may be shared among several hosts and managed centrally using both SAN and NAS (network-attached storage) (servers). While SAN may make use of both Ethernet and Fibre Channels, NAS is dependent on Ethernet. Additionally, SAN concentrates on high performance and low latency, whereas NAS concentrates on usability, management, scalability, and lower TCO (TCO). NAS storage controllers have file system ownership and storage partitioning, in contrast to SAN.

Cost:

Expect to pay about $100,000 for hardware for a small SAN of, say, 10 servers, half a terabyte of storage, and 16-port switches. Bigger SANs can cost well over $1 million.

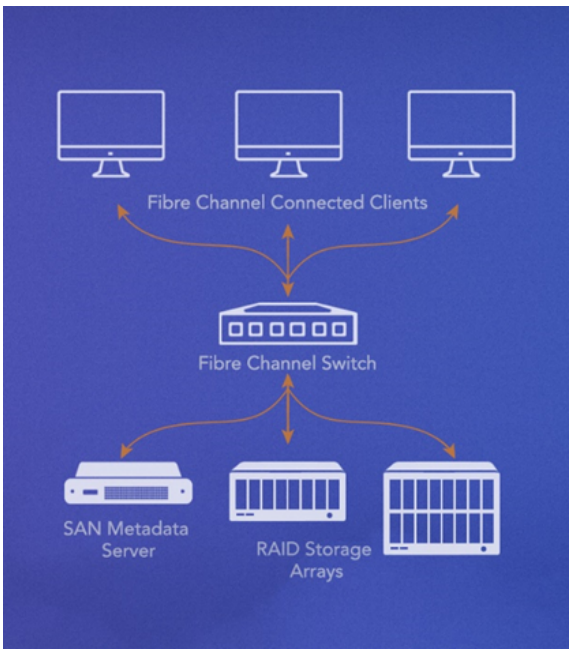

Figure 7: SAN conceptual model

3.4 Summary

The price of SAN is too high, and we do not need such great performance for now. Besides, the NAS is sufficient when only a few employees have access rights to the on-premise storage. Thus we eliminate SAN. DAS may be reasonably efficient currently, but the data size may grow in the future, and DAS has insufficient scalability, while NAS is less likely to encounter this issue.

The number of people required for setting on-premise storage is also our concern. Because of the complexity of SAN, it would need more employees compared with NAS. Taking all into consideration, we finally settled on NAS.

Table 3: on-premise storage plan comparison

	DAS	NAS	SAN
Cost	Low	Moderate	High
Setup	Easy	Moderate	Difficult
Scalability	Limited	Good	Great
Personnel required	Few	Moderate	Many
Backup	No	Local	Safe

4 Data migration plans

4.1 Moving shared data

4.1.1 Option 1: uploading data to destination data warehouse manually

The most intuitive way is to upload the data to the destination on-premises / cloud storage manually.

4.1.2 Option 2: using AWS snow to migrate data to AWS

Petabyte-scale data transfer with on-board computing and storage using AWS snow. AWS Snowball Edge is a 80TB data transmission appliance with built-in computing and storage. To transport huge volumes of data into and out of AWS, to serve as a temporary storage layer for big local datasets, or to support local workloads in off-the-grid or remote areas, users may utilize Snowball Edge.

4.1.3 Option 3: using 3rd party data pipeline

The data pipeline from the source cloud to the destination cloud storage is often set up by a 3rd party migration provider. A useful option for creating data pipelines is Dextrus. It can automate and operationalize utilizing built-in approval and version control mechanisms, and establish batch and real-time streaming data pipelines in a matter of minutes.

4.1.4 Summary

The service of AWS snow is recommended because it works well with the AWS glacier which was mentioned in the previous part. And when moving PetaByte-scale data, migrating it using hardware will be the most time-saving solution.

Table 4: Public data migration plan comparison

	Option 1: upload manually	Option 2: AWS snow	Option 3: Data pipeline
Time	high	low	high
Finance	low	low	low
Labor	high	low	low
Reputation	low	high	high
Difficulty	high	low	low

Most importantly, both uploading the data manually and using a data pipeline will need to migrate data using the internet. To discuss deeper, using a T3 data pipeline with a speed of 250MB/s to transfer 1.2 Pb data will take more than 50 days. Besides, in our case, to move 1.2Pb data with a data pipeline that runs at 1 dollar per hour will cost around $2000. AWS snow, however, will only slightly increase the cost to about $2800.

Therefore, we suggest using AWS snow to migrate our petabyte-scale archive data.

4.2 Moving individual data

4.2.1 Option 1: using the tools provided by the email service provider

Email service providers usually offer several tools or instructions to assist businesses in moving their data from another cloud to their services.

Consider Gmail as an example. The administrators in the organization can follow Google's data migration guidelines if the company is switching from its current data service provider to Gmail, with 50G of personal Google Drive space for each employee.

The administrators should carefully prepare the migration procedures and consider the volume and intricacy of the data before beginning (Google suggests moving fewer than 20 million files at a time). The IT staff would also need to conduct a data audit, noting the data that requires specific processing, such as restricted or shared with users outside the firm.

The group can then establish a file connection. When the connection has been confirmed, a scan of all the required migration data can be done. Next, establish a connection with the target. In this step, the administrator must authorize the account in the target service to configure the target connection. The admin can then configure the migration after that. The creation and management of identity mapping is the most crucial task. The permissions in the source environment are guaranteed to be reflected in the target environment via identity mapping. The migration is ready to begin after the aforementioned have been configured.

4.2.2 Option 2: using services of cloud migration provider

Cloud migration services providers are independent companies that aid organizations in navigating the challenging process of moving their data from one cloud or on-premises location to another. They are generally independent of the provider of cloud services. This allows them to suggest a cloud environment that is more specifically tailored to the needs of the organizations.

There are several third parties from which to choose. However, the businesses mentioned below are pioneers in the relocation sector.

Deloitte is the first one we suggest. In addition to being a top consulting firm, Deloitte is a leader in technology professional services. It provides managed cloud services, such as cloud migration, for AWS, Azure, and Google Cloud. Their cloud migration services include:

- Creating a migration plan.
- Putting automatic migration procedures in place.
- Getting the business ready for the new alignment.

Additionally, they provide application modernization services that aid in converting mission-critical code into a cloud-friendly language, facilitating a seamless move.

`To establish if the migration team can securely migrate traditional apps to the cloud and the best method to accomplish it, Deloitte offers automated assessment tools that examine programs and uncover dependencies. To transfer related apps simultaneously and minimize downtime, the migration team also assists enterprises in identifying dependencies. The Deloitte team migrates an organization's apps and, if required, modernizes them to guarantee that they will function in the cloud and offer all of the benefits of cloud-native applications.

4.2.3 Option 3: Using 3rd party storage migration software

For moving individual data, third-party migration solutions are also excellent options. In contrast to option 2, which entails bringing support personnel on location, migration tools are often self-guided. Customers acquire migration software licenses and carry out the data migration following the service provider's instructions.

IT can easily migrate email, files, and other workloads from various Source and Destination endpoints using a 3rd party migration tool like Bititan, which offers an automated, industry-leading SaaS solution.

4.2.4 Summary

In the comparison of these 3 choices, option 3, which uses cloud migration software, is recommended.

Table 5: Individual data migration plan comparison

	Option 1: **Email service provider**	**Option 2:** **Cloud migration provider**	**Option 3:** **Cloud migration software**
Time	high	low	low
Finance	low	high	medium
Labor	high	low	low
Reputation	-	high	high
Easiness	low	high	high

In option 3, the cost is relatively low when compared with option 2. because option 2 will send a team on-site, which increases the service fee. Also, option 3 will require very little work from the IT team compared to option 1. Most importantly, option 3 will not use the bandwidth because the software provider is responsible for setting up the data pipeline from cloud to cloud.

5 Justification

5.1 Recommendation and Explanation

After analyzing all the previously proposed solutions, we believe that a more targeted solution should be adopted to complete data storage and migration. Ultimately, we decided that AWS S3 Glacier should first be used for archival storage. AWS S3 Glacier is easier to use, more professional, and cheaper at only $0.00099 per GB per month (or $1 TB per month). And can be tailored for various access patterns and retention times. In addition, this long-term archival material that is accessed less than once a year and restored asynchronously provides the best solution for archival material that does not require immediate access but needs the flexibility to retrieve large data sets without incurring any costs.

For local storage, using NAS is the best solution. The storage system of NAS is directly connected to the network through the network interface, and the user accesses it through the network instead of connecting to the server or client through the I/O bus. This page means that it can better solve the demand for rapid storage capacity expansion than other methods. Some NAS manufacturers started offering cloud storage. And NAS storage devices can enable many users to work together and exchange data. This makes it easier to use compared to other methods, improves efficiency, can reduce the possibility of redundant personnel, and the cost is relatively low

We believe that AWS snow should be used for data migration to AWS for shared data. Compared with the other two methods, it is faster and does not occupy bandwidth. And built-in computing and storage can support off-grid or local workloads in remote areas. And it can run well in the AWS ecosystem. Such stability and cost performance are the reasons why it becomes the choice. For personal data, third-party software (Bititan) should be used. Compared with other options, third-party software has comprehensive technology and guidance and can perform data migration more quickly and accurately according to the requirements, reducing labor costs and overhead. And it easily migrates email, files, and other workloads from various source and destination endpoints. The low personnel required and the fact that it is easy to use and relatively inexpensive make it an obvious choice. Such a dynamic solution is targeted and can better save costs and human resources while providing relatively stable media Libraries and platforms can be better used by consumers and administrators.

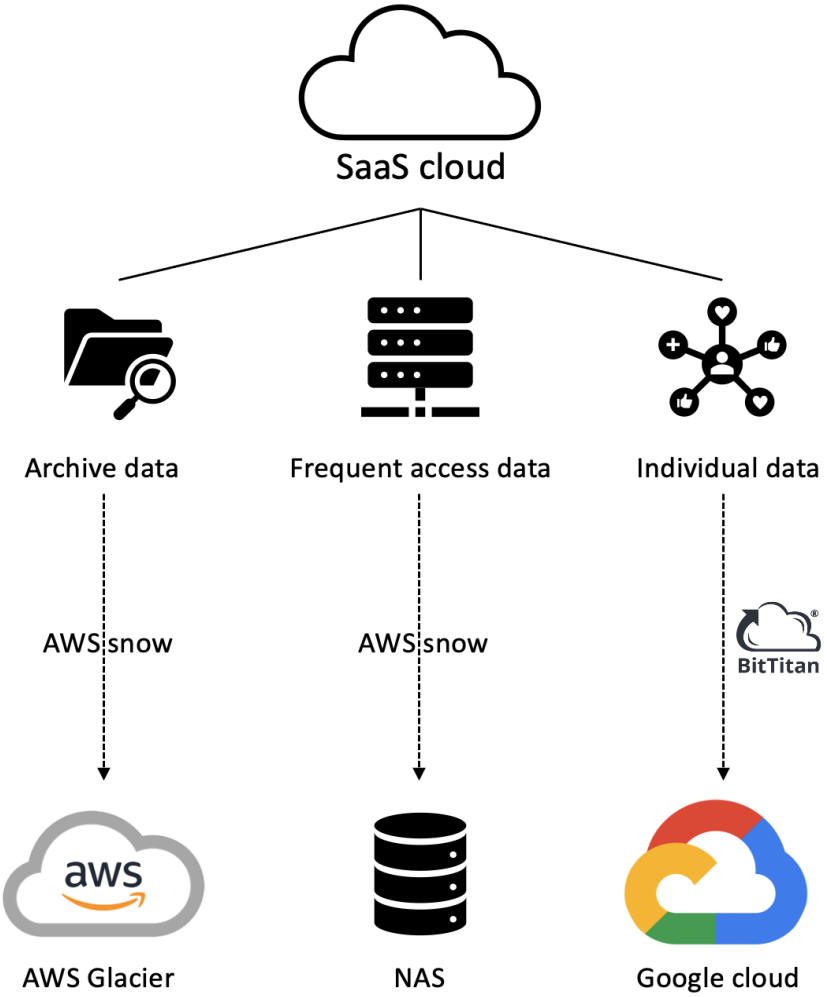

Figure 8: Process Architecture

5.2 Real-world example

Joyn is a German streaming service platform like Netflix. Joyn had more than 3 Petabytes of media archive data stored in on-premises facilities in Germany, with a significant portion of its long-term archive on local storage hardware in their data center, which it rented for a high monthly fee. To save the budget, they decided to migrate to AWS.

In the migration of such a large quantity of data, they used 40 AWS snowball devices to transfer data from their original data warehouse to the data center of AWS. The data is ultimately stored in AWS Glacier, and they can use intelligent tiering to automatically classify data according to their frequency of being used.

As a result, Joyn successfully moved 3.4 Pb data in 3 months and saved millions of data center fees every year. Besides, because of AWS's dynamic storage optimization feature, users can easily access the media library.

5.3 Business value

The company can store data more stably, allowing the company to run more smoothly. It can reduce expenses, allowing companies to spend less on data storage, and more funds can be invested in areas where they may be needed. And our recommended solution can be easy to use, can better improve work efficiency, does not require more personnel and hardware configuration, and reduces the problem of redundant personnel or redundant equipment that may be left after data transfer superior. And the stability of the selected platform can also allow the company to better guarantee the use of users and managers now and in the future. Thereby avoiding possible network security risks caused by instability and the possibility of hardware damage. Most importantly, relatively speaking, the data transfer can be completed quickly and safely at the same cost. And in the future, the company can continue to use this solution without spending more money on training possible staffing. The stability of such a solution can better help enterprises develop.

5.4 Implementation plan

5.4.1 Cost and time estimate

One-time cost

To perform the new IT infrastructure, a one-time payment of purchasing new devices and data migration fee is generated.

We will need QNAP TVS-H874 NAS devices *3, which cost 15000. Then we will purchase Dell 400-BJKY HDD * 17, which will cost 6800. Moreover, we'll need 14000 to set up ethernet switches.

The migration fee is a combination of 2 parts, the first part is the rental fee of AWS snow appliances. To meet the requirement of migrating 1.375 Pb data, 18 AWS snow appliances (Snowball Edge Storage Optimized) will be needed. Each appliance offers 80 Tb of HDD capacity and transfers data at a speed of 250 Mb/s to 400 Mb/s when using the Amazon S3 interface. It will take 2 days to fill up each appliance. The fee for renting the appliances is $30/device/day. We suggest having these devices for 5 days for the task. Thus, the cost of each AWS appliance is $150 and it will cost $2,700 for the archive data migration.

The migration of individual data will be performed through BiTitan. Each license of migrating 25Gb individual data from SaaS to Google Cloud is $15. Therefore, a budget of $375,000 will be needed for this part.

In conclusion, a one-time payment of $413,500 will be generated for the data migration and new infrastructure.

Table 6: Estimated one-time payment in establishing new IT infrastructure

content	cost per device or user	quantity	total (US dollar)

NAS	35800	1	35800
AWS snow storage devices	150	18	2700
migration software license	15	25000	375,000
			413500

Yearly cost

When we use AWS S3 Glacier Deep Archive, we need to purchase the service annually, so there is a budget of $14574.79 for storage every year. Then we assume every data will be retrieved once a year, so the cost for the retrieval fee will be $3067.08. Thus, there is a total budget of $17641.87 per year for using the AWS S3 Glacier Deep Archive.

Table 7: Estimated yearly cost

content	cost（US dollar)
AWS storage fee	14574.79
AWS retrieval fee	3067.08
	17641.87

5.4.2 Time

The data migration for employee personal data will take a maximum of 5 days to complete. The migration related to AWS snow appliances will take 15 days, because there is an estimated delivery of 5 days for the appliances to arrive and return. The tasks can be carried out at the same time, therefore, the time estimated to perform the data migration is 15 days.

Table 8: Estimated time

category	task	time (days)
task 1: individual data	migrate data using 3rd party service	5
total		5 days

task 2: public data	AWS deliver devices	5
	copy data to devices + copy new on-premise device	5
	deliver devices back to AWS warehouse	5
total		15 days
		total 15 days

5.4.3 Communication plan

Firstly, IT team and Business unit will meet to develop the plan, estimate the cost and identify the access control.

Next, team leader (or CIO) will report to the board to get the approval for service changes.

Then, we will inform users and employees for changes.

Finally, after implementation, we have to train the users and employees how to use new services.

During the above communication process, at least four emails will be sent to the employees and other related stakeholders to notify them about the up-coming data migration.

The first email will be performed as preliminary communication. This email will be sent before the migration. In this email, no specific action is required.

The second email will be sent several days before the migration begins. It will be a reminder of the timeline and a list of actions the users may need to perform or aid in the migration, like backuping data.

The third email will only be sent to the related users who are required to perform immediate actions like granting access to email accounts or turning off certain applications during the migration.

The last email will be a final communication. To be specific, this email will be sent after the migration in order to tell users that the data has been transferred to a new cloud and the instructions of using the new cloud.

5.4.4 IT infrastructure

We need to set or change some infrastructures. For application layer, we will use apps or software access to AWS and NAS. For storage layer, we will adopt storage methods like AWS cloud storage, AWS snow and NAS storage. For compute part, we need several servers. We also need buy switches and utilize our dedicated T3 line to provide networking support.

5.4.5 Personnel requirement

We assigned several tasks for the IT team, Business unit Services desk, and all employees so that we can adapt and get used to new changes quickly.

IT team needs to classify the data, migrate the data, determine access rights with the business unit and set up NAS.

The business unit needs to purchase services and determine access rights with the IT team.

The services desk needs to provide training and guidance for users and employees

All employees need to get training and create an AWS account.

6 Resources

37, 2 Ye. “CN.” *Amazon*, Amazon, https://aws.amazon.com/cn/s3/features/security/.

Engdahl, Sylvia. “Blogs.” *Amazon*, Greenhaven Press/Gale, 2008, https://aws.amazon.com/blogs/media/prmbp-joyn-readies-exclusive-content-for-audiences-amazon-s3-intelligent-tiering-and-glacier/.

“IBM TS4300 Tape Library - Overview.” *IBM*, https://www.ibm.com/products/ts4300?mhsrc=ibmsearch_a&mhq=IBM+TS4300+Tape+Library.

Migration Planning & Strategy Guide – Bittitan Help Center. https://help.bittitan.com/hc/en-us/articles/360044417394-Migration-Planning-Strategy-Guide.

Miles, Ellen. “Snowball.” *Amazon*, Scholastic Inc., 2015, https://aws.amazon.com/snowball/pricing/.

Pratt, Mary K. “San Costs.” *Computerworld*, Computerworld, 1 July 2002, https://www.computerworld.com/article/2576607/san-costs.html.

“Pricing | Cloud Storage | Google Cloud.” *Google*, Google, https://cloud.google.com/storage/pricing.

Product Documentation for IBM Guardium Data Encryption, 24 Aug. 2022, https://www.ibm.com/support/pages/product-documentation-ibm-guardium-data-encryption.

“S3.” *Amazon*, Strand Street Press, 2002, https://aws.amazon.com/s3/features/.

“S3.” *Amazon*, Strand Street Press, 2002, https://aws.amazon.com/s3/storage-classes/glacier/?nc1=h_ls.

“Storage Classes | Google Cloud.” *Google*, Google, https://cloud.google.com/storage/docs/storage-classes#coldline.