New situation: They are looking to now move not only new customers to the cloud but all of their existing customers. We need to make sure that each of the new types of Mo Vid customer can fit in your proposed architecture. If they do not, explain how you will change and or scale the architecture design to fit all the types of Mo Vid customers

Modified video play Solution:

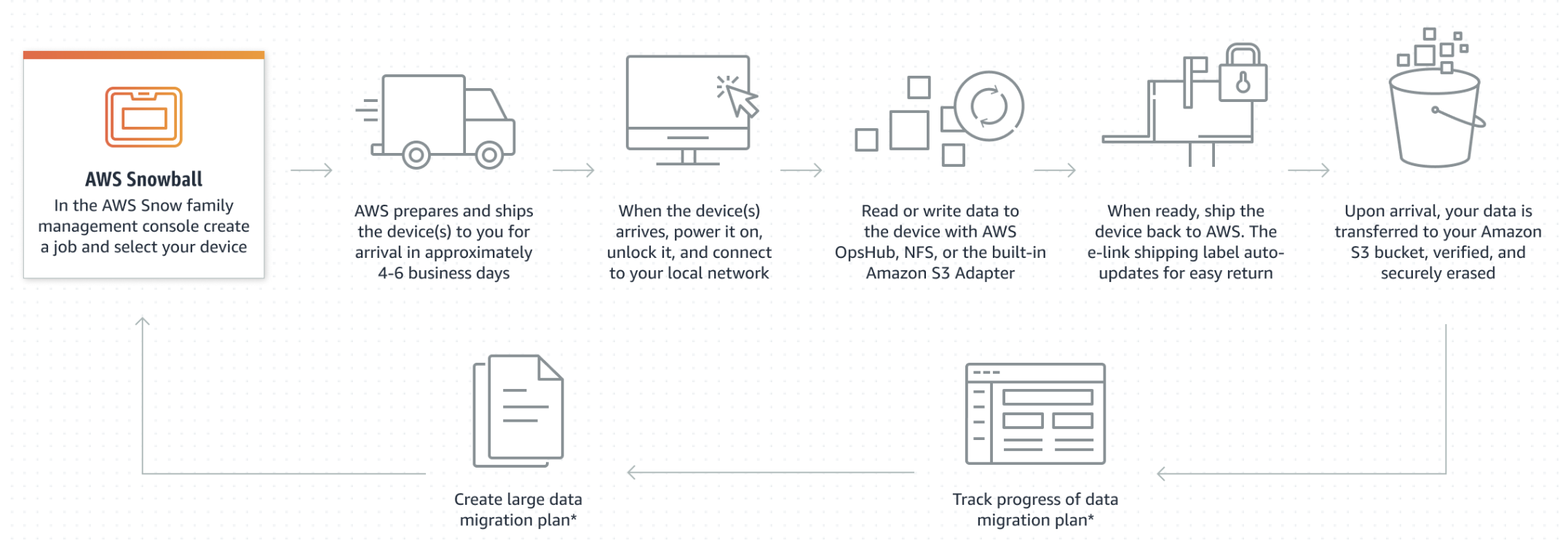
Step1

transfer data

Now we not only need to be able to store video data, but also help the original Large , Medium and Small 3 types of customers to migrate data to our new platform.

For large customer: They could use AWS Snowball. For datasets less than 10PB or distributed in multiple locations, you should use Snowball. In addition, you should evaluate the amount of available bandwidth in your network backbone. If you have a high speed backbone with hundreds of Gb/s of spare throughput, then you can use Snowmobile to migrate the large datasets all at once. If you have limited bandwidth on your backbone, you should consider using multiple Snowballs to migrate the data incrementally

In the AWS Snow Family console, select your preferred device, either Snowball Edge Compute Optimized or Snowball Edge Storage Optimized. Create a job with an Amazon S3 bucket, select Amazon Simple Notification Service (Amazon SNS) for tracking, and configure options like Amazon EC2 AMIs and a GPU. AWS prepares and ships the device to you, and you receive it in approximately 4-6 days. Once the device arrives, power it up and use AWS OpsHub to unlock it. Connect to your LAN. Use AWS OpsHub to manage the device, transfer data, or launch EC2 instances. When done, shut down and return the device to AWS. The shipping label automatically appears on the E Ink screen. When the device arrives at the AWS Region, any data stored in your on-board bucket(s) is moved to your S3 bucket and verified in about the same time it took you to load the device. All data is then securely erased from the device, and it is sanitized of any customer information.



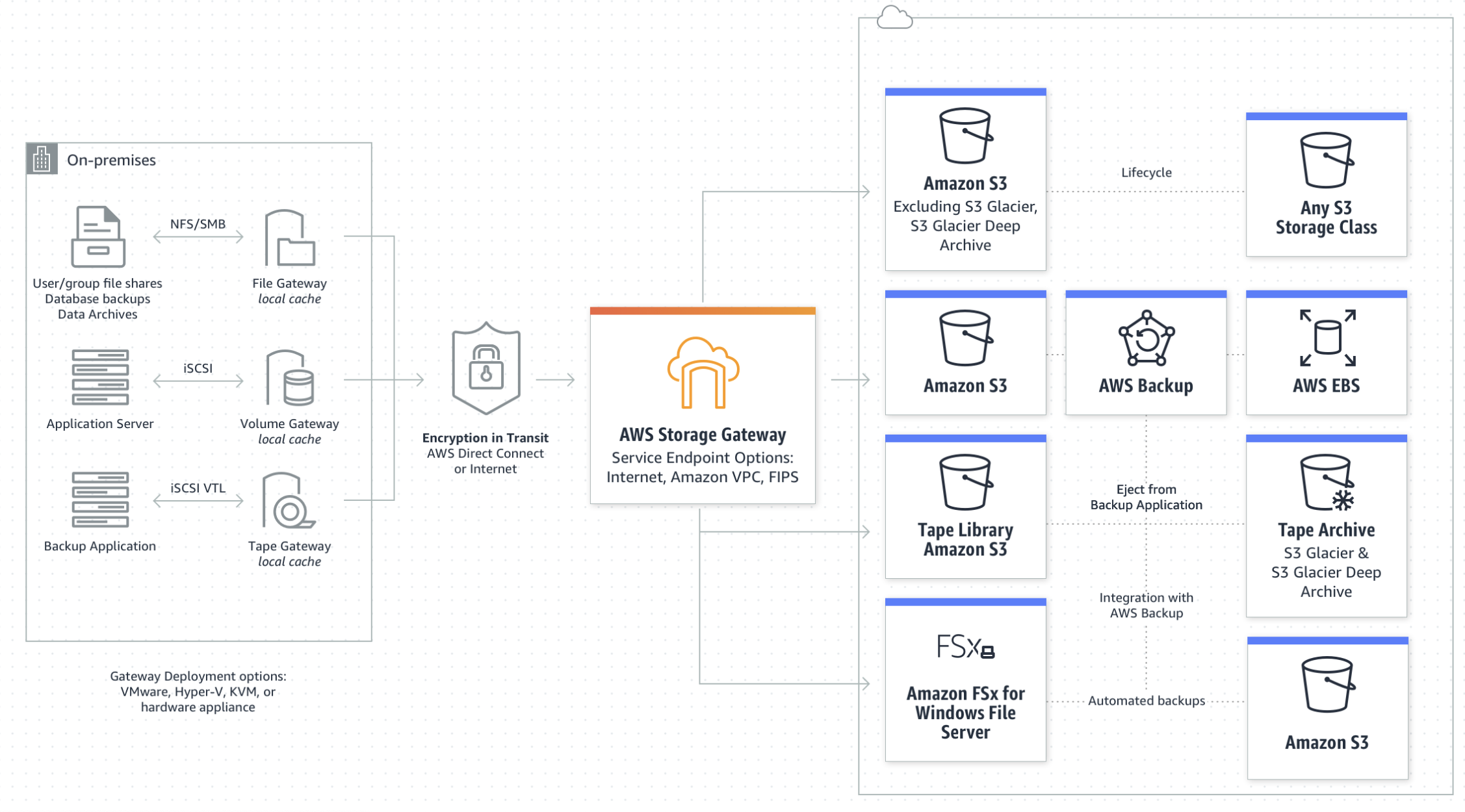
(Reference: <https://aws.amazon.com/cn/snowball/>)

As we are told: “Total video storage is greater than 60TB less than 150TB”, so We do not use snowmobile as: its dataset is not big enough, snowmobile is not suitable .When our data is over 10PB, we will consider it (reference: <https://aws.amazon.com/snowmobile/> )

For Medium and Small types customer: we use AWS Storage Gateway.

The service provides three different types of gateways – File Gateway, Tape Gateway, and Volume Gateway.File Gateway file data is stored in Amazon S3 as durable objects using Amazon S3 File Gateway or in fully managed file shares using Amazon FSx File Gateway.

Tape Gateway virtual tape library (VTL) configuration seamlessly integrates with your existing backup software for cost effective tape replacement in Amazon S3 and long term archival in S3 Glacier and S3 Glacier Deep Archive.Volume Gateway stores or caches block volumes locally, with point-in-time backups as EBS snapshots. These snapshots may be recovered in the cloud.



(Reference: https://aws.amazon.com/storagegateway)

Step2

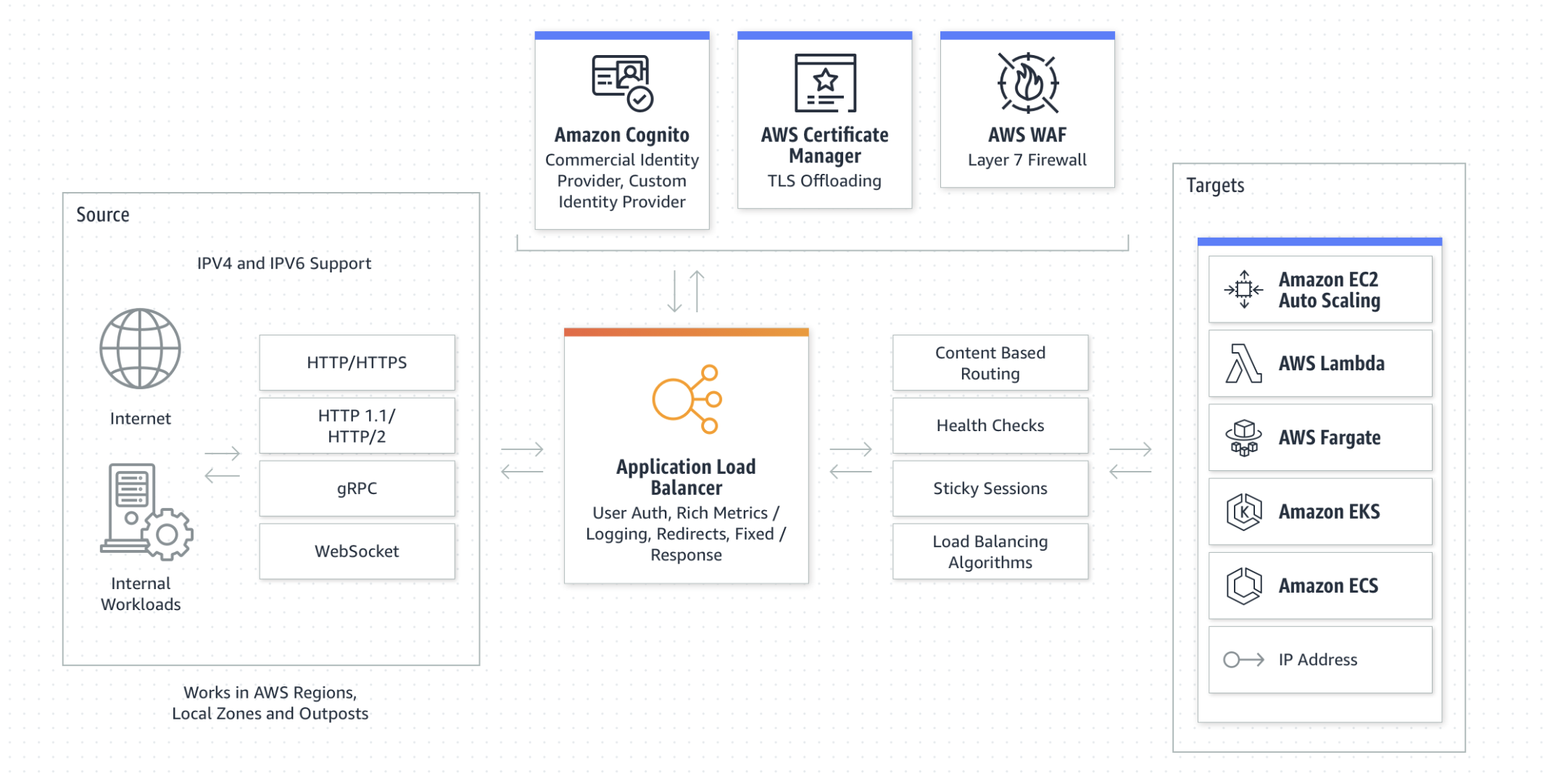
Increase robustness

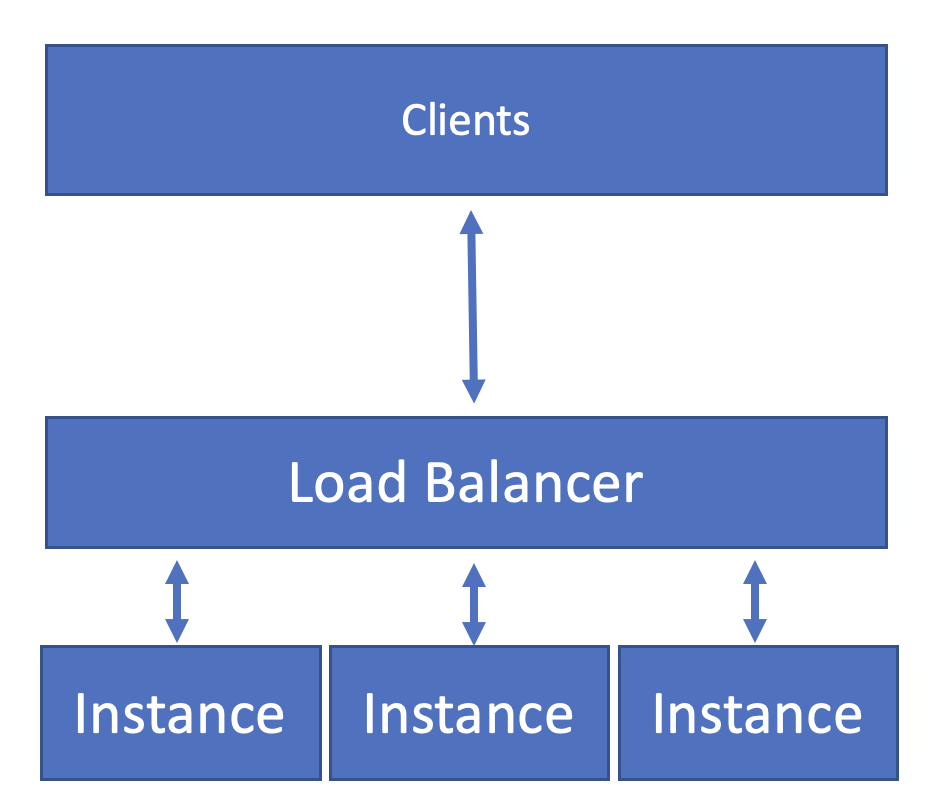
As all our customers are now on cloud, It greatly improves the requirements for service operation and maintenance robustness and automated maintenance.

We will consider to use ELB of our web systems part: user account login/register/comments/role managements/Authorization management for video playback and CRUD.

Elastic Load Balancing (ELB) automatically distributes incoming application traffic across multiple targets and virtual appliances in one or more Availability Zones (AZs).

It can realize automatic operation and maintenance management, automatically discover, identify, manage, and allocate disks, intelligently schedule applications and data according to affinity, and automatically monitor disk status and give timely warnings. Second, high-availability data support. Our video load balancing uses cross-node replicas to synchronize data to achieve high availability. When a problem occurs, it will automatically schedule the application to the high-availability data node to ensure the continuity of the application. Third, with a variety of data volume types, our video load balancing aggregates HDD, SSD, and NVMe disks to provide non-low-latency, high-throughput data services. Fourth, flexible and dynamic linear expansion can be achieved. Our video load balancing can dynamically expand capacity according to the size of the cluster to flexibly meet the data persistence requirements of applications.





(Reference: scaling.pptx and <https://aws.amazon.com/cn/elasticloadbalancing/> )

Step3

Support multiple situations:

Video playback and return visits, the application scenarios of various user operations mainly include the following three categories

1. Adapt to high-availability architecture middleware

Kafka, ElasticSearch, Redis, etc., such middleware applications have their own high-availability architecture, and at the same time have high requirements for IO access to data. The LVM-based single-copy local data volume provided by our overall solution can well meet their requirements.

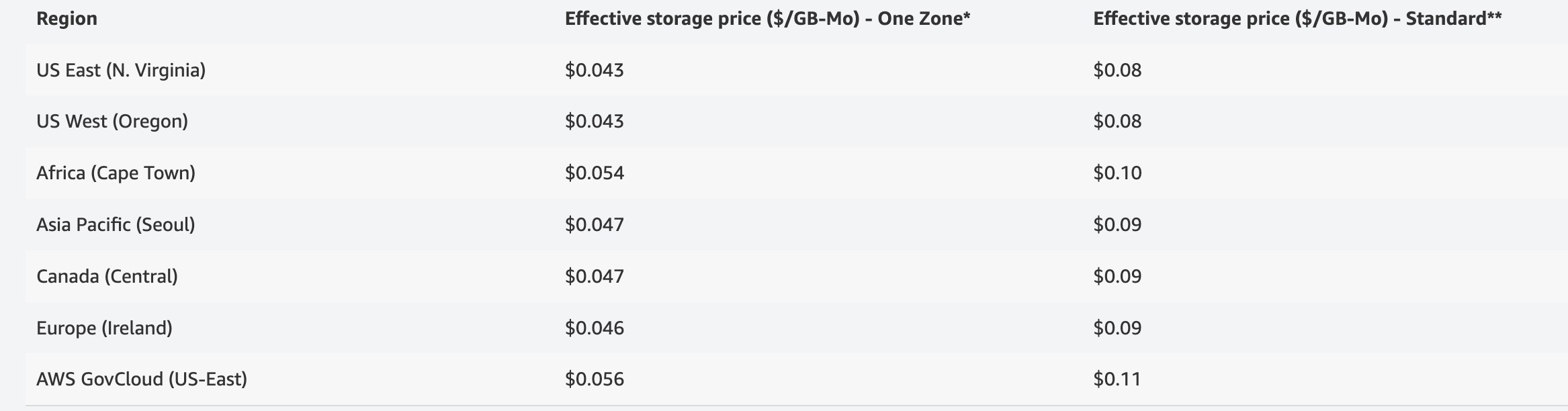
2. Provide highly available data volumes for applications

OLTP databases such as MySQL/MSSqlserver require the underlying storage to provide highly available data storage, which can quickly recover data when a problem occurs, and also require high-performance data access. The dual-copy high-availability data volume provided by our overall solution can well meet such needs.

3. Automated operation and maintenance of traditional storage software

We are considering use AWS EFS or MinIO, Ceph and other storage software need to use the disk on the K8s node. PVC/PV can be used to automatically use the single-copy local volume of our overall solution through the CSI driver to quickly respond to the deployment, expansion, and migration proposed by the business system. and other needs, realize automatic operation and maintenance based on K8s. Scale our video file system automatically as files are added, removed, and burst to higher throughput levels when necessary.

Amazon EFS file systems can automatically scale from gigabytes to petabytes of data without needing to provision storage. Tens, hundreds, or even thousands of compute instances can access an Amazon EFS file system at the same time, and Amazon EFS provides consistent performance to each compute instance. Amazon EFS is designed to be highly durable and highly available. With Amazon EFS, there is no minimum fee or setup costs, and you pay only for what you use.(reference: scaling.pptx and https://aws.amazon.com/efs/)

If we have more budgets , we could use ifs; if not , we can construct a similar file systems use open source systems like(Cph/K8s) on AWS. Following is price of Los:

Conclusion

In order to prevent server-sprawl, we consider to manager servers more automatically and accurately.

To automatic more actions: we use software-defined infrastructure, the top layer is a high-performance distributed file system, which is also the core of our overall solution. This user-mode file system is completely self-developed, fully compatible with the POSIX standard, and realizes zero data copy between user mode and kernel mode. Fully exploit the performance of high-speed media such as NVMe SSD/SCM/PM. he second layer is the data service layer. In addition to providing services between local storage of different nodes, between local storage and external storage, and between heterogeneous external storage, it can also provide services for single copy and multiple copies, strong consistency and weak consistency. Provide storage solutions separately. The third layer is the storage management layer, which is responsible for interacting with different storage devices, breaking device barriers and data islands through a unified data format, and allowing data and services to flow freely across heterogeneous devices. In addition, the system can also schedule storage devices through the CSI interface.

To manage more accurately: we use more Virtualize servers not real machines. Individual servers can also be consolidated and configured as standalone server workloads on powerful servers. Excessive reliance on virtualization brings its own virtual sprawl issues, so virtual servers should be provisioned carefully

Careful management of IT assets (reference clould\_capacity.pptx) and the use of capacity planning tools can help stop server sprawl before it begins. These tools can help track existing servers and determine when servers are running below capacity or when demand for servers has peaked and server expansion is required

######################## 分割线 ###########################

你自己写的部分也加上：

·       **Backup methodology for the video files**

o   Must be able to recover videos for 60 days after they are deleted by a user in the Mo Vid web app.

When the user chooses to delete a file, our technology does not actually physically delete the file from storage, but instead marks the file as 'deleted'. And the system will mark the time of deletion of this file. In this way, we can calculate the difference between today's date and the date the file was deleted[1]. The file is still on the server and is still owned by this user, but its status label has changed to 'Deleted'.

When a user clicks to view their deleted file information, we will query all videos with the 'deleted' tag belonging to the user. The system will display the information of these files to the user. Of course, we only show videos that have been deleted within 60 days. When the current time is more than 60 days since a video was deleted, we will actually delete the video on the server and release the memory occupied by the video. and disk location. When users see their deleted videos, and these videos are less than 60 days away from the date of deletion, the user can click the 'Restore' button, so that we will change the status of this video from 'deleted' to ' normal'. In this way, the user can use the video normally again.

·       **Backup methodology for the Mo Vid SQL data**

o   Must be able to restore any database changes for 30 days.

An inherent advantage of modern cloud-based technology is recovery speed and data reliability[2-3]. Backup methodology for the Mo Vid SQL data is a very important methodology in cloud computing. The approach our system takes is:

1. Use more reliable SQL statements to reduce the possibility of exceptions such as 'deadlock' and increase the ability to process information under concurrent conditions.

2. Use a backup of the database. The key implementation of this technology is the rollback of the database, which means that we can roll back the database information to any time (within 30 days). With this technique we can restore any database changes for 30 days.

3. Use disaster recovery techniques[3]. In order to roll back data quickly, data recovery usually does not include historical data. This shows that our system is capable of disaster recovery. That is to say, when an abnormal situation occurs, we need to allow the business to recover as quickly as possible. This means that the current environment only needs the current data, which will get the business back up and running quickly. Historical data can be rolled back slowly after business recovery. This means that cloud-based backup systems may need to regularly back up current data[4].

·      **TCO of one year of costs for running the proposed cloud setup for 10 customers.**

o   Each customer will have 250 GB of Videos.

The economics of cloud computing technology is also an important aspect that we need to consider. Our plan is that Each customer will have 250 GB of Videos. This obviously requires a lot of storage resources, which are the rarest [5]. Therefore, it will also cost us a lot of money. However, the cloud computing industry is developing rapidly, and the future prospects are very good [6]. And cloud computing customers are willing to pay for cloud services [5]. Therefore, such a plan is feasible and successful.

TCO (Total Cost of Ownership) includes the cost of product purchase to later use and maintenance. This is an evaluation standard that companies often use. If Each customer will have 250 GB of Videos, 10 customers will need 2500 GB of storage. And we also need to take into account the cost of aging, maintenance, and renewal of equipment. Now the cost of renting a server with 2500GB of storage space varies in different countries. But overall, if we use CPFS (Cloud Parallel File Storage), the total cost is about tens of thousands of dollars a year. Such a cost is acceptable to us.

#################### 新增分割线 ###########################

Here is a detailed plan for implementing the proposed architecture following structure1.docx:

Use Azure Media Services to manage and encode the video files. This will provide scalable and secure storage for the video files, and enable you to stream and transcode the videos on demand.

If you have a vast video repository and decide to make a change in the future, let's say you want to change your watermark on them, you will have to re-process all the videos. That would mean creating a new config, running it on all the existing videos, and ensuring it gets completed. That's a lot of development effort and additional spending on cloud storage

（infer: <https://imagekit.io/blog/video-streaming-azure-blob-storage-imagekit/>)

Use Azure Blob Storage to store the encoded video files. This will provide highly available and cost-effective storage for the videos, and allow you to access the files from anywhere using the Azure Blob Storage API. We can use Azure Data Lake Storage for big data analytics.

Use Azure Premium Blob Storage to store the original video files. This will provide faster access to the files and higher durability, as well as features such as data tiering and tiered storage.

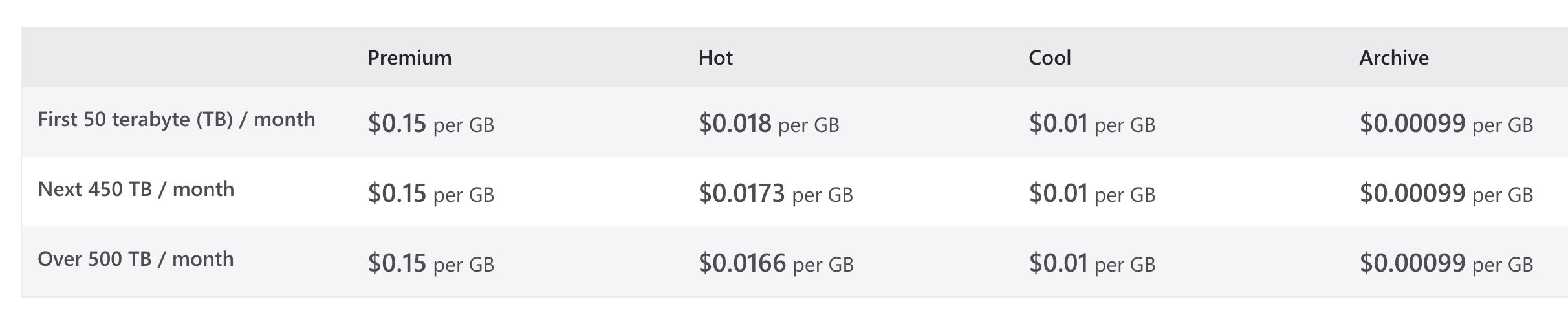
Because inbound is not counting, so the blog storage part is:

From cost https://azure.microsoft.com/en-us/pricing/details/storage/blobs/:

|  |  |
| --- | --- |
| **First 50 terabyte (TB) / month** | **$0.15** per GB |

For **small** 1T~10TB,

0.15 x 1000 - 0.15 x 1000x10 x 25(customers)

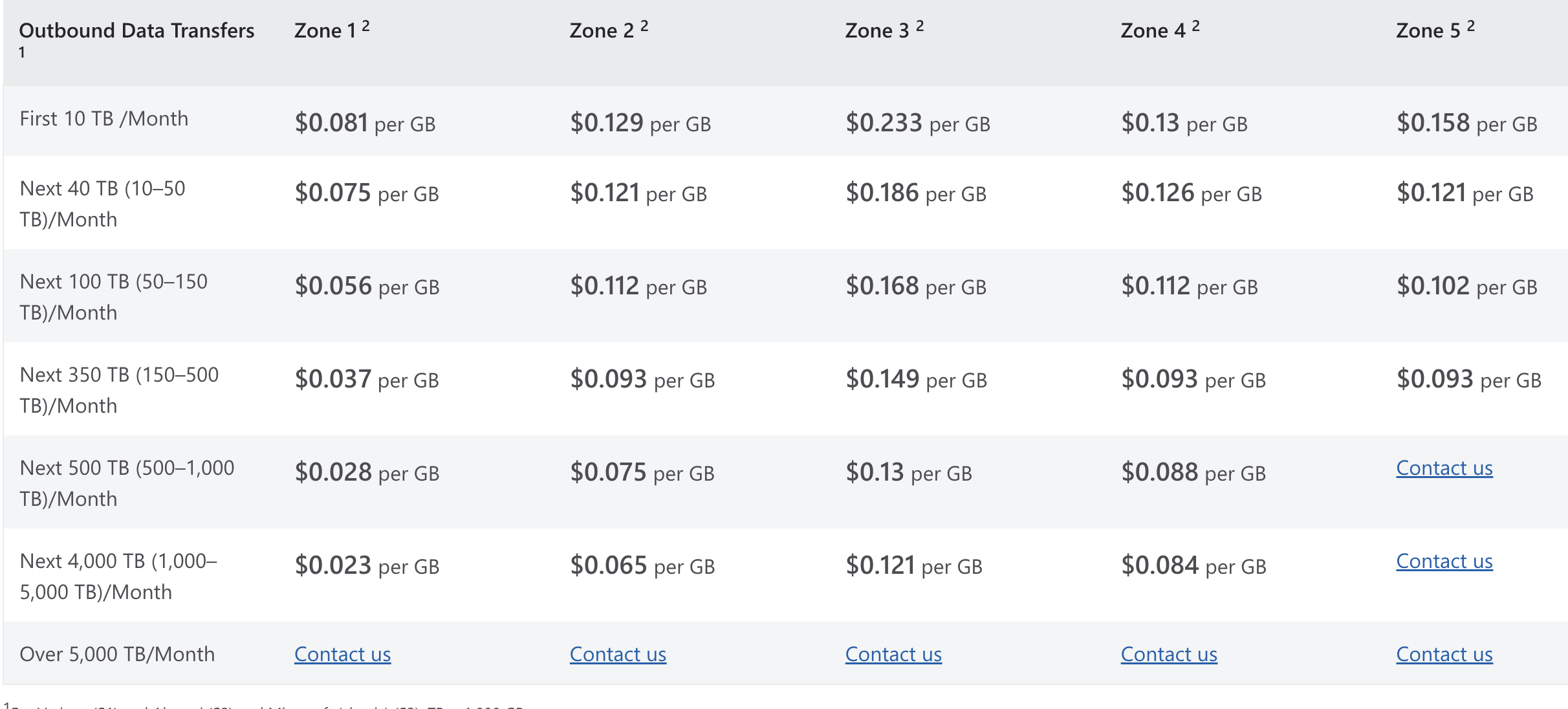
For **medium**: 10Tb~60TB

0.15x1000x10 ~ 0.15x1000x60 $ x 10(customers)

For **large**: 60Tb~160TB

0.15x1000x60 ~ 0.15x1000x160 $ x 3(customers)

注释：我写的只prenuim的部分，如果需要计算hot存储，还需要根据表中再算一遍，hot和cool估计也是需要一样算费用的

Use Azure Content Delivery Network (CDN) to distribute the video files to users across the globe. This will provide high-speed delivery of the videos to users, reducing buffering and providing a better viewing experience.

For **small**

**1000 x 0.081 \* 25**

For **medium**

**1000 x25 x 0.081\*10**

For **large**

**1000 x45x 0.081\*3**

Use Azure Virtual Machines to host the video transcoder and the web application. This will provide the compute resources and flexibility to run the transcoder and web app, and connect to the storage and media services.

Azure Elastic Container Service (ECS) is a cloud-based container orchestration service that allows users to easily deploy, manage, and scale containerized applications. It can be used to host websites and web applications by running the necessary web server software in a container. ECS offers a number of advantages over traditional infrastructure for hosting websites, including scalability, flexibility, and easy management. It is a cost-effective and efficient way to host and manage web applications in the cloud.

(

Small:

Azure ECS  Standard\_A1\_v2 type,  size 64G, memory 8G, cpu 4core, 13dollar/month \* 25

Medium：

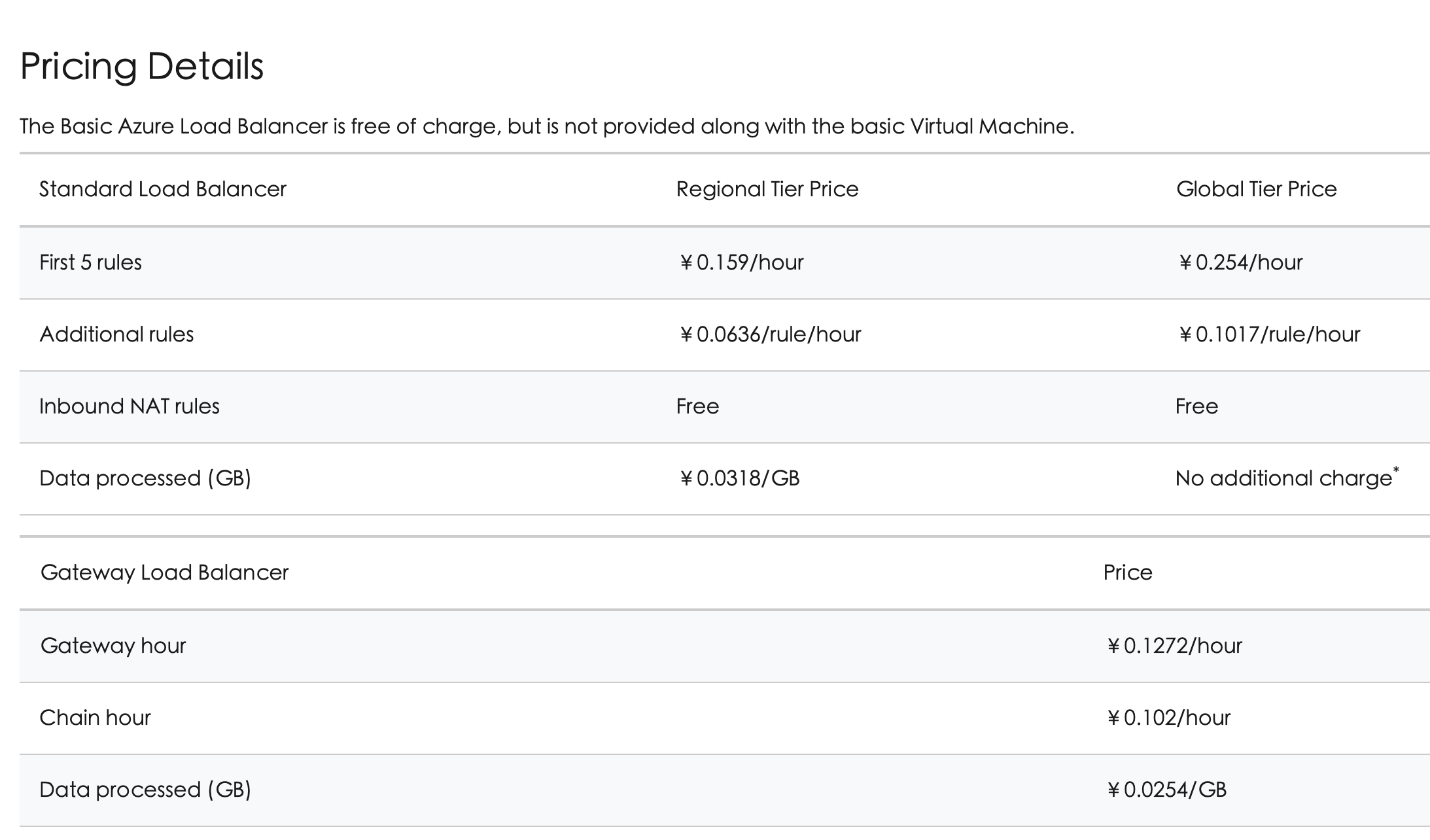
Azure ECS  Standard\_A1\_v1 type,  size 1T, memory 32G, cpu 8core, 30dollar/month \* 10

Large:

Azure ECS  super type,  size 1T, memory 64G, cpu 16core, 50-60dollar/month \* 3

这个微软需要预约才能看，我是根据aws估计的和作业1来的

)

Use Azure Load Balancer to distribute incoming traffic to the transcoder and web app across multiple virtual machines, providing high availability and scalability.

Load balancer is not need for small

For medium: 0.159 \* 24\*30 \*10(amount) + 0.0318\*1000\*60 \* 10

For large :0.159\*24\*3 + 0.0318\*1000\*150\*3

Use Azure Autoscale to automatically adjust the number of virtual machines based on the workload, providing cost-effective scaling and high availability

cost use：https://azure.microsoft.com/en-us/pricing/calculator/