#Three

Date:
Monday Sept 16th, 2019

Topics

- Course Content Overview
- Cloud Storage
- Hands-on Lab #2: S3
- Content Delivery Network
- Hands-on Lab #3: CloudFront
- Next Steps

Course Content Overview

Course Content Overview

We are here! 

Week	Date	Topics, Readings, Assignments, Deadlines	Due Date
1	6:00 PM PT, 8/26	Course Logistics & projects Introduction to Cloud Technologies	
2	6:00 PM PT, 09/02	Labor Day Campus Closed – No Lecture	
3	6:00 PM PT, 09/09	Fundamentals	Homework #1 Due
4	6:00 PM PT, 09/16	Storage Content Delivery Network	Team Formation Due
5	6:00 PM PT, 09/23	Compute, Serverless	
6	6:00 PM PT, 09/30	Databases, Migrating Data to Cloud	Homework #2 Due
7	6:00 PM PT, 10/07	Big Data, Data Streaming	Quiz #1 Due
8	6:00 PM PT, 10/14	MIDTERM EXAM (Close book, Close notes). Bring student ID	
9	6:00 PM PT, 10/21	Artificial Intelligence I	Project #1 Due:
10	6:00 PM PT, 10/28	Artificial Intelligence II	
11	6:00 PM PT, 11/04	Internet of Things (IoT)	Project #2: Design Due
12	6:00 PM PT, 11/11	Veterans Day Campus Closed – No Lecture	Homework #3 Due
13	6:00 PM PT, 11/18	Cloud Security	Project #2: Component I Due
14	6:00 PM PT, 11/25	Cloud Management	Project #2: Component II Due
15	6:00 PM PT, 12/02	Project presentation & discussion I	
16	6:00 PM PT, 12/09	Project presentation & discussion II	Quiz #2 Due
17	6:00 PM PT, 12/16	FINAL EXAM Thu, Dec 16 (close book, close notes). Bring student ID	

Cloud Storage

Introduction: Why Cloud Storage

Compelling Economics

Pay as you go

No upfront investment
No commitment

No risky capacity planning

No need to provision for redundancy or overhead

Easy to Use

Self service administration

SDKs for simple integration

Reduce risk

Durable and Secure

Avoid risks of physical media handling

Speed, Agility, Scale

Reduce time to market

Focus on your business, not your infrastructure

Amazon S3

Highly durable object storage for all types of data

Internet-scale storage
Grow without limits

Built-in redundancy
Designed for 99.999999999% durability

Low price per GB per month
No commitment
No up-front cost

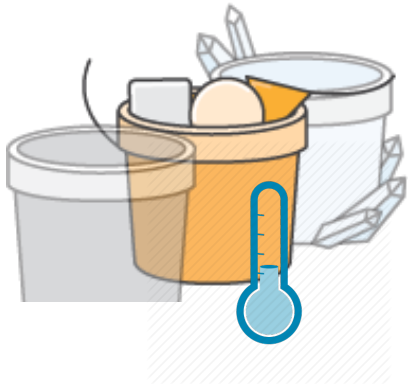
Benefit from AWS's massive security investments

Amazon S3 key features

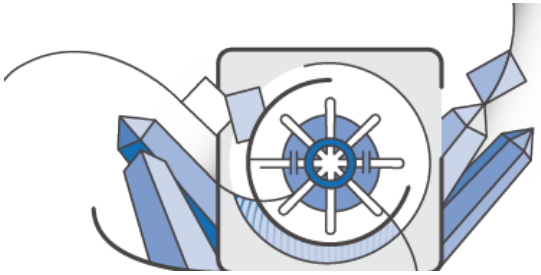
- **Management Console**
  - Bucket management
  - Monitoring spend
  - Managing lifecycle
- **Easy Integration**
  - AWS SDKs simplify programming
  - REST and SOAP APIs
- **Data Management**
  - Lifecycle management
  - Cost control
- **Data protection**
  - Versioning
  - Multi-factor delete
  - Encryption
  - Flexible Access Control Mechanisms
  - Time-limited access to object
  - Audit Logs

Choice of storage classes on S3

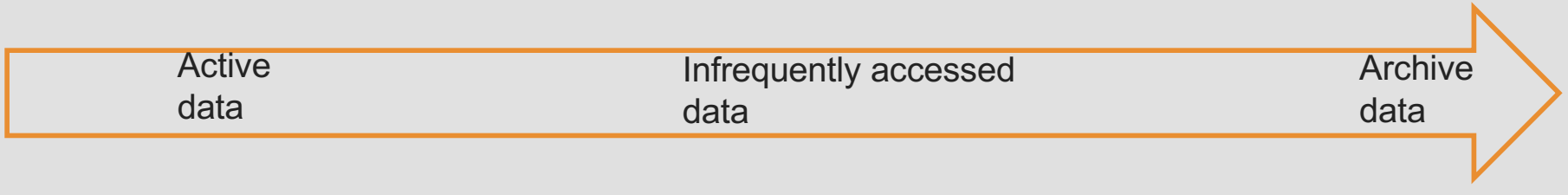
Standard



Standard - Infrequent Access, One Zone - IA



Amazon Glacier, Deep Archive



https://aws.amazon.com/about-aws/whats-new/2018/04/announcing-s3-one-zone-infrequent-access-a-new-amazon-s3-storage-class/

Object Storage Service Use Cases

Data tiering using S3 Life Cycle Policies

- S3 Standard
  - Web Apps
  - Big Data Analytics
  - Small objects and temporary scratch space
- IA,One Zone IA
  - File sync and share
  - Active Archive
  - Enterprise backup
  - Media transcoding
  - Geo-redundancy/DR
- Glacier, Deep Archive
  - Deep/offline archives
  - WORM-compliant data

Use cases for Standard-Infrequent Access

- File sync and share and consumer file storage
- Backup and archive and disaster recovery
- Long-retained data

Standard - Infrequent Access storage

Durable

Designed for 11 9s of durability

Available

Designed for **99.9**% availability

High performance

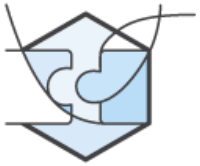
Same as Standard storage

Easy to use

- No impact on user experience
- Simple REST API

Secure

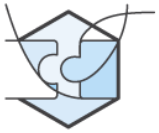
- Bucket policies
- AWS Identity and Access Management (IAM) policies
- Many encryption options



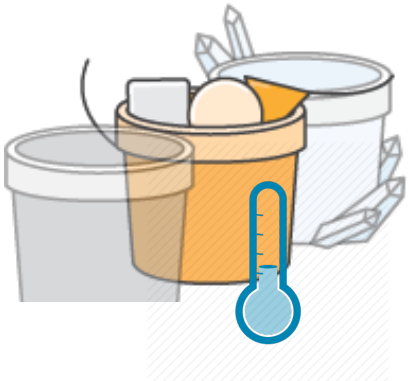
Integrated

- Lifecycle management
- Versioning
- Event notifications
- Metrics

Standard - Infrequent Access storage



Integrated: Lifecycle management

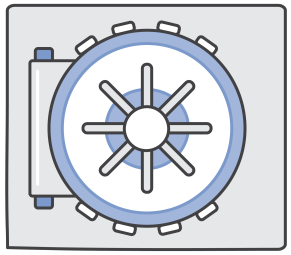


Standard - Infrequent Access

- Directly PUT to Standard - IA
- Transition Standard to Standard - IA
- Transition Standard - IA to Amazon Glacier storage
- Expiration lifecycle policy
- Versioning support

Amazon Glacier

- Archival storage for infrequently accessed data



Amazon Glacier is optimized for infrequent retrieval

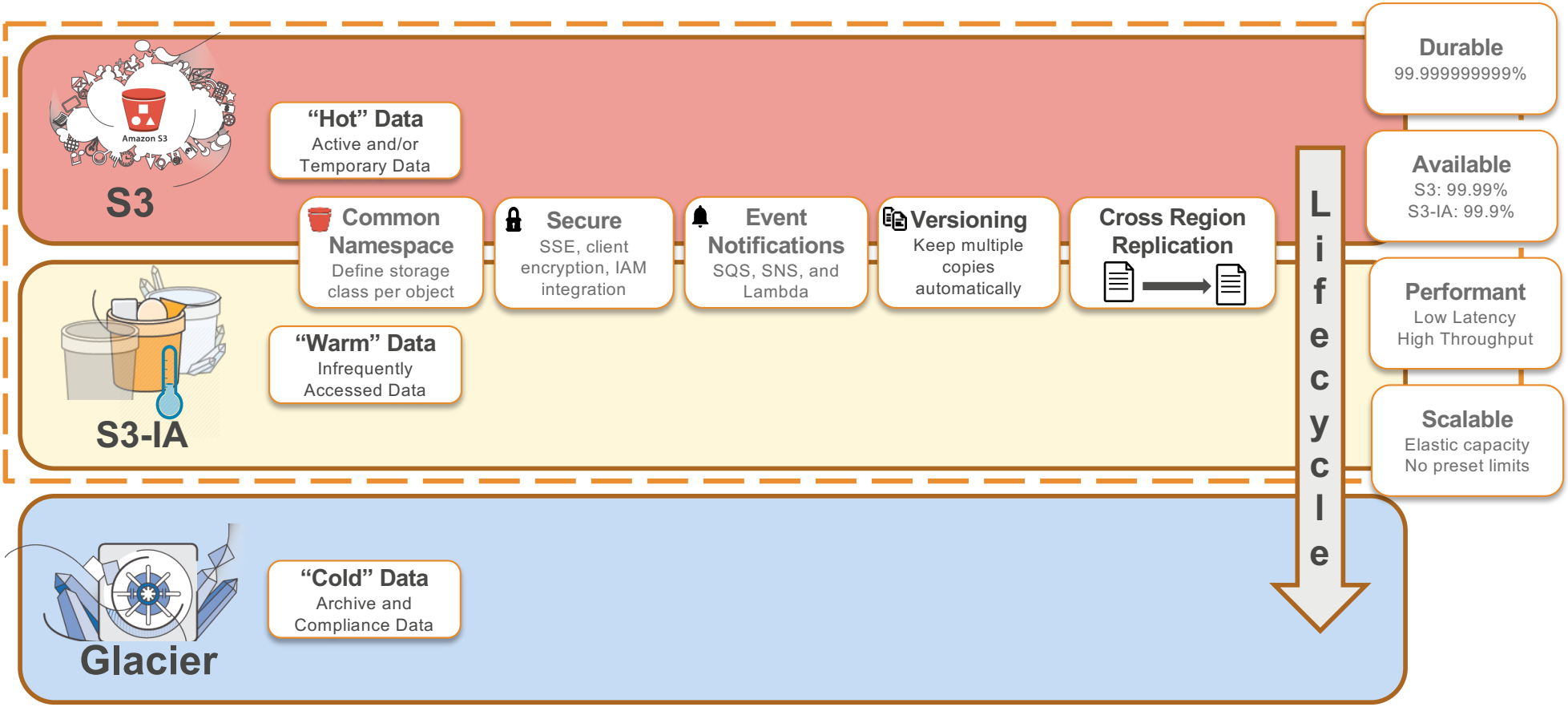
Even lower cost than Amazon S3; Same high durability

Stop managing physical media

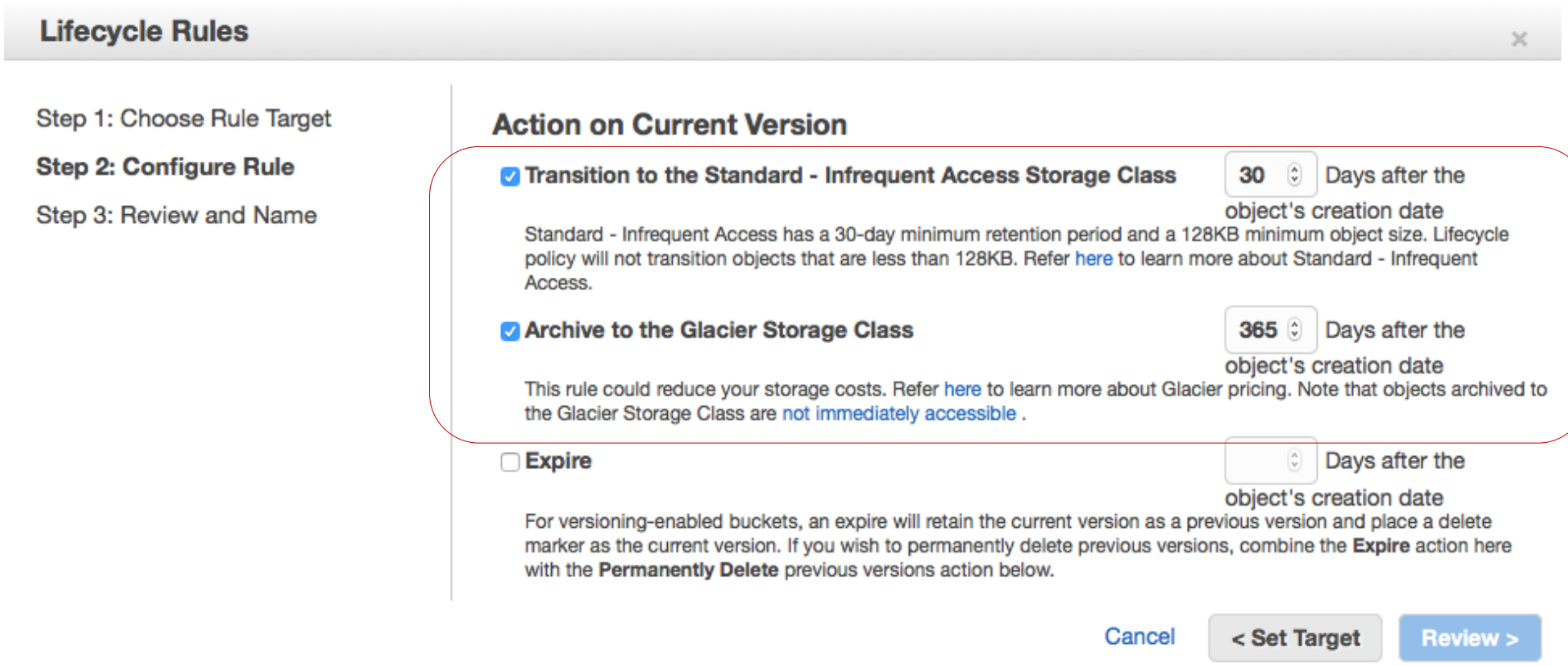
Amazon Glacier key features

- **Optimized for infrequent access**
  Typical retrieval times of 3-5 hours
- **Archives**
  Durable stored block of data with unique ID
  Unlimited number of archives
- **Vaults**
  Container for storing archives
- **Data inventory**
  Archive size, creation data, and description
- **Management console**
  Create and delete vaults
- **AWS SDK**
  Java, .NET, PHP and Python SDKs simplify programming
  REST API
- **Access control**
  AWS IAM secure access control
- **Integrated lifecycle management with S3**
  Automates archiving of S3

Storage Tiered To Your Requirements



Transition older objects to Standard - IA



Standard - Infrequent Access storage

Lifecycle policy

```
<LifecycleConfiguration>
 <Rule>
   <ID>sample-rule</ID>
   <Prefix>documents/</Prefix>
   <Status>Enabled</Status>
   <Transition>
     <Days>30</Days>
     <StorageClass>STANDARD-IA</StorageClass>
   </Transition>
   <Transition>
     <Days>365</Days>
     <StorageClass>GLACIER</StorageClass>
   </Transition>
 </Rule>
</LifecycleConfiguration>
```

Standard Storage -> Standard - IA

Standard - Infrequent Access storage

Lifecycle policy

```
<LifecycleConfiguration>
 <Rule>
   <ID>sample-rule</ID>
   <Prefix>documents/</Prefix>
   <Status>Enabled</Status>
   <Transition>
     <Days>30</Days>
     <StorageClass>STANDARD-IA</StorageClass>
   </Transition>
   <Transition>
     <Days>365</Days>
     <StorageClass>GLACIER</StorageClass>
   </Transition>
 </Rule>
</LifecycleConfiguration>
```

Standard Storage -> Standard - IA

Standard - IA Storage -> Amazon Glacier

Restricting access by IP addresses

Bucket policy with IPv4

```
{ "Version": "2012-10-17",
  "Id": "S3PolicyId1",
  "Statement": [
    { "Sid": "IPAllow",
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:*",
      "Resource": "arn:aws:s3:::examplebucket/*",
      "Condition": {
          "IpAddress": {"aws:SourceIp": "54.240.143.0/24"}
          "NotIpAddress": {"aws:SourceIp": "54.240.143.188/32"} } } ] }
```

Updating bucket policy with IPv6

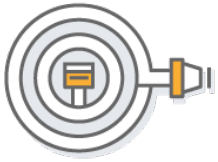
```
{ "Version": "2012-10-17",
  "Id": "S3PolicyId1",
  "Statement": [
    { "Sid": "IPAllow",
      "Effect": "Allow",
      "Principal": "*",
      "Action": "s3:*",
      "Resource": "arn:aws:s3:::examplebucket/*",
      "Condition": {
          "IpAddress": "aws:SourceIp":
                  [ "54.240.143.0/24", "2001:DB8:1234:5678::/64" ]}
          "NotIpAddress": {"aws:SourceIp":
                  ["54.240.143.128/30", "2001:DB8:1234:5678:ABCD::/80”]}}]}
```

Data ingestion into S3

Data ingestion into AWS storage services

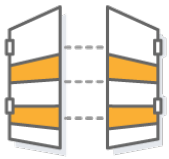
AWS Import/Export Snowball
Accelerate PBs with AWS-provided appliances

- 80 TB model, global availability
- For Exabyte scale, use snowmobile



Amazon Kinesis Firehose

- Ingest device streams directly into AWS data stores



AWS Storage Gateway

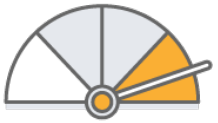
- Instant hybrid cloud
- Up to 120 MB/s cloud upload rate (4x improvement), and

AWS Direct Connect

- Co-location (COLO) to AWS

ISV Connectors

- CommVault
- Veritas
- etcetera



Amazon S3 Transfer Acceleration

- Move data up to 300% faster using AWS's private network

What is AWS Snowball? Petabyte-scale data transport

Ruggedized case
"8.5G impact"

80 TB
10 GE network

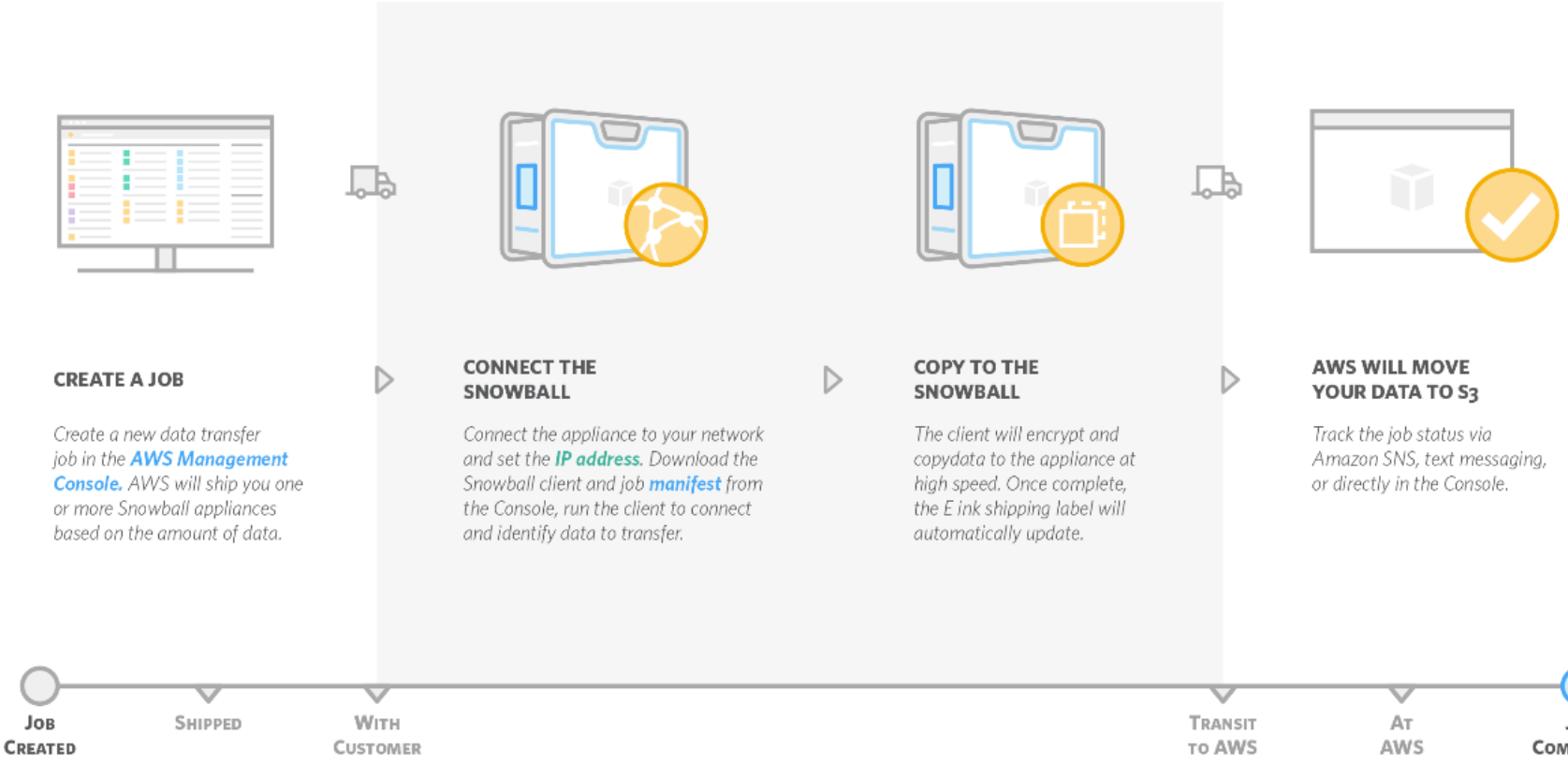
E-ink shipping label

Rain- and dust-resistant

Tamper-resistant case and electronics

All data encrypted end-to-end

How it works


CREATE A JOB
Create a new data transfer job in the AWS Management Console. AWS will ship you one or more Snowball appliances based on the amount of data.
CONNECT THE SNOWBALL
Connect the appliance to your network and set the IP address. Download the Snowball client and job manifest from the Console, run the client to connect and identify data to transfer.
COPY TO THE SNOWBALL
The client will encrypt and copydata to the appliance at high speed. Once complete, the E ink shipping label will automatically update.
AWS WILL MOVE YOUR DATA TO S3
Track the job status via Amazon SNS, text messaging, or directly in the Console.
JOB CREATED
SHIPPED
WITH CUSTOMER
TRANSIT TO AWS
AT AWS
JOB COMPLETED

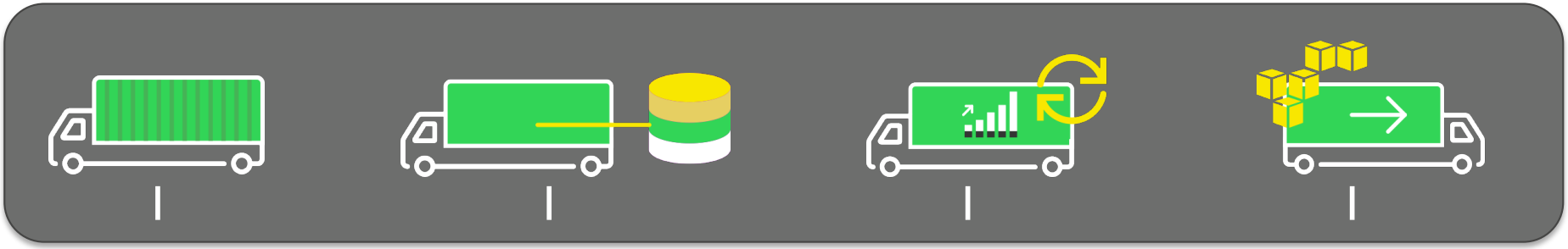
What about Exabyte Scale?

1 Exabyte = 1000 PB

https://youtu.be/8vQmTZTq7nw

What is AWS Snowmobile? Exabyte-scale data transport

AWS Snowmobile: 100PB Container



45-foot long rugged container & truck

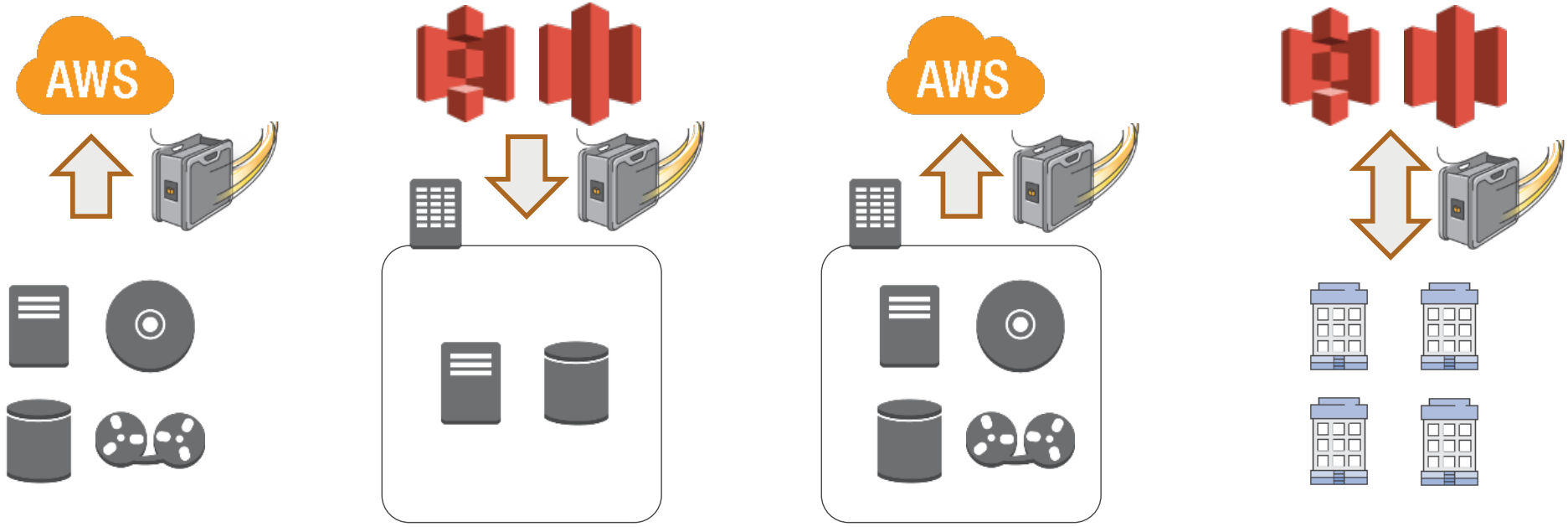
Connect to your datacenter with fiber cable

Fill ‘er Up!

Transports Data To AWS

https://aws.amazon.com/snowmobile/

Use cases: AWS Import/Export Snowball



Cloud Migration

Disaster Recovery

Data Center Decommission

Content Distribution

S3 Transfer Acceleration

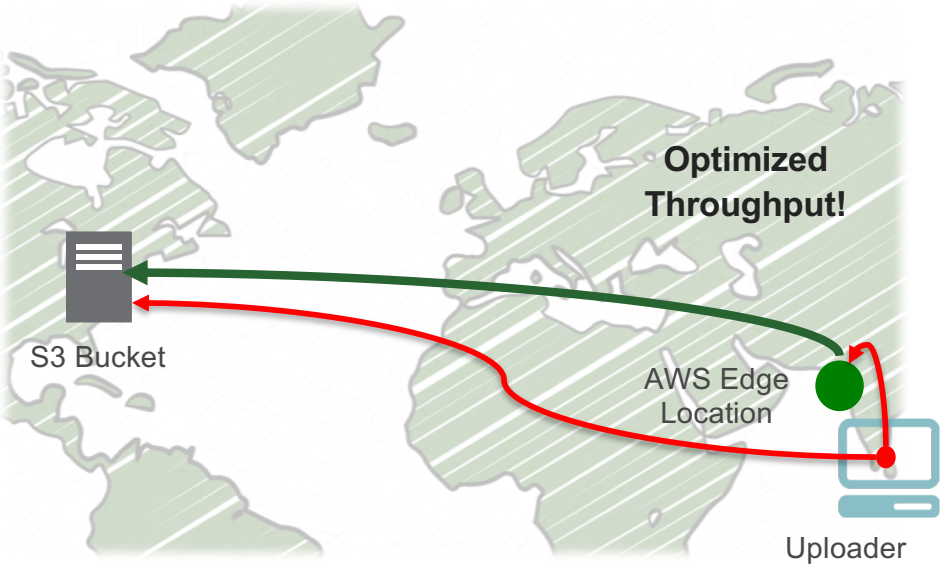
Typically 50%-400% faster

Change your endpoint, not your code

No firewall exceptions

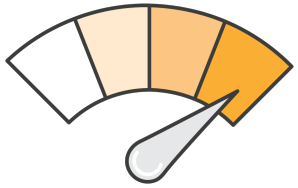
No client software required

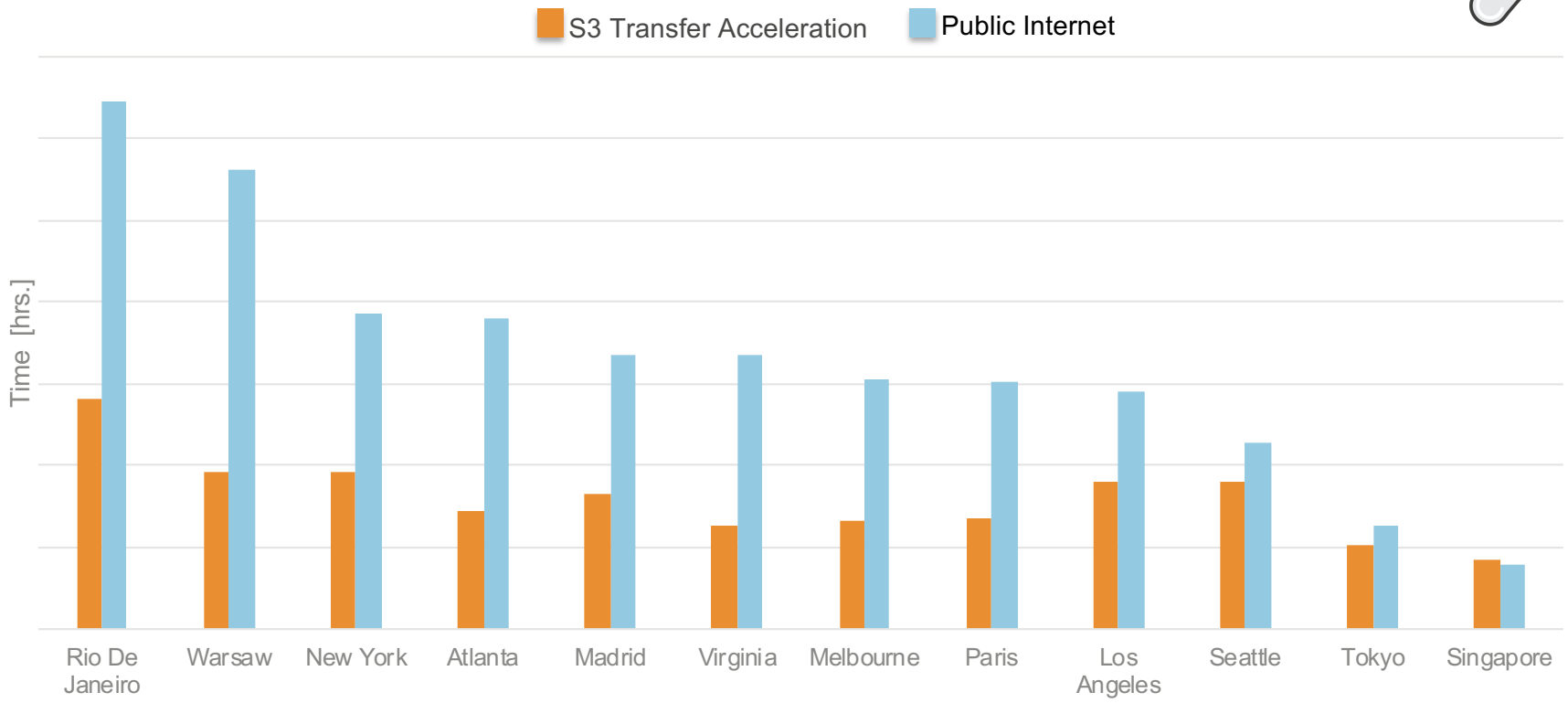
190 Points of Presence (179 Edge Locations and 11 Regional Edge Caches) in 72 cities across 33 countries



https://aws.amazon.com/cloudfront/features/

How fast is S3 Transfer Acceleration





500 GB upload from these edge locations to a bucket in Singapore

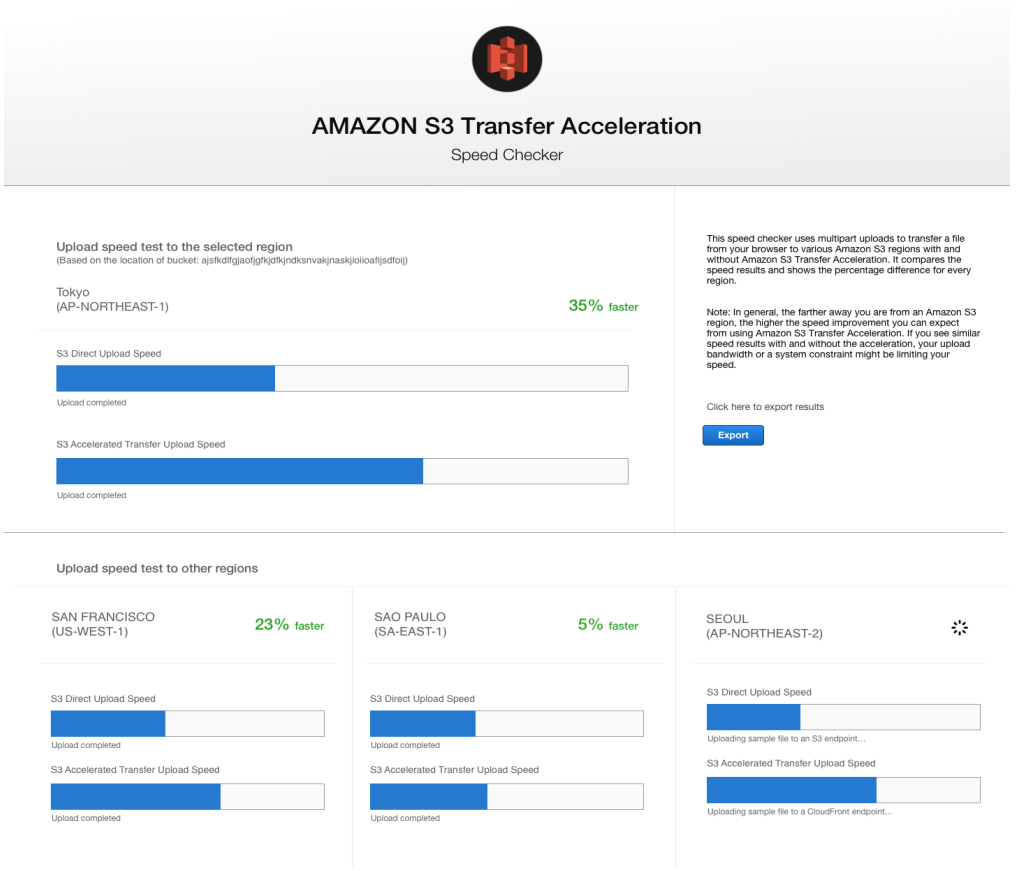
How to start using S3 TA?

1. Enable S3 Transfer Acceleration on your S3 bucket.

2. Update your endpoint to <bucketname>.**s3accelerate**.amazonaws.com.

How to check performance gains?

[Click Here!]



When to use Snowball vs. Transfer Acceleration

AWS Snowball

- Large, infrequent uploads
- 5–10 day tolerance
- Supports large data transfers, from TBs to PBs

Transfer Acceleration

- Recurring, frequent uploads
- GBs or TBs of upload from distributed locations
- Long geographic distances
- Supports accelerating transfers into and out of S3, using AWS edge locations

What is AWS Storage Gateway?

Service connecting an on-premises software appliance with cloud-based storage

Works with your existing applications

Secure and durable storage in AWS

Low-latency for frequently used data

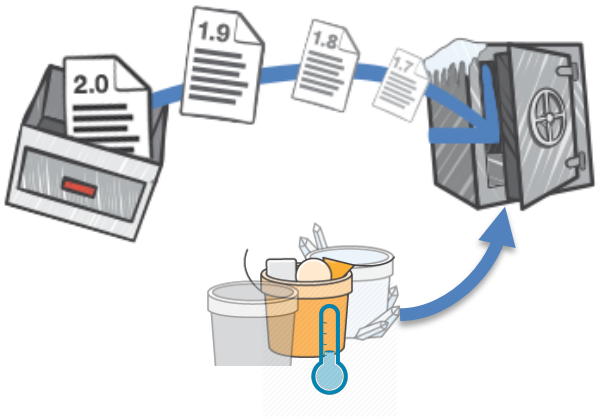
Scalable and cost-effective on-premises storage - $$ per gateway per month + S3/Glacier storage fees

Best Practice #1: Use versioning

Versioning

- Protects from accidental overwrites and deletes
- New version with every upload
- Easy retrieval of deleted objects and roll back to previous versions

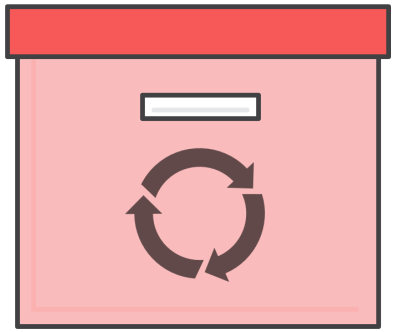
Best Practice #2: Use lifecycle policies



Lifecycle policies

- Automatic tiering and cost controls
- Includes two possible actions:
  - Transition: archives to Standard - IA or Amazon Glacier based on object age you specified
  - Expiration: deletes objects after specified time
- Actions can be combined
- Set policies at the bucket or prefix level
- Set policies for current version or non-current versions

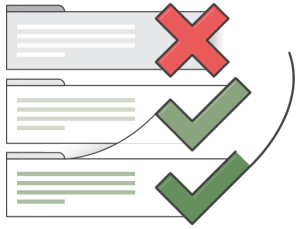
Versioning + lifecycle policies



Versioning → Recycle bin

Lifecycle policies → Automatic cleaning

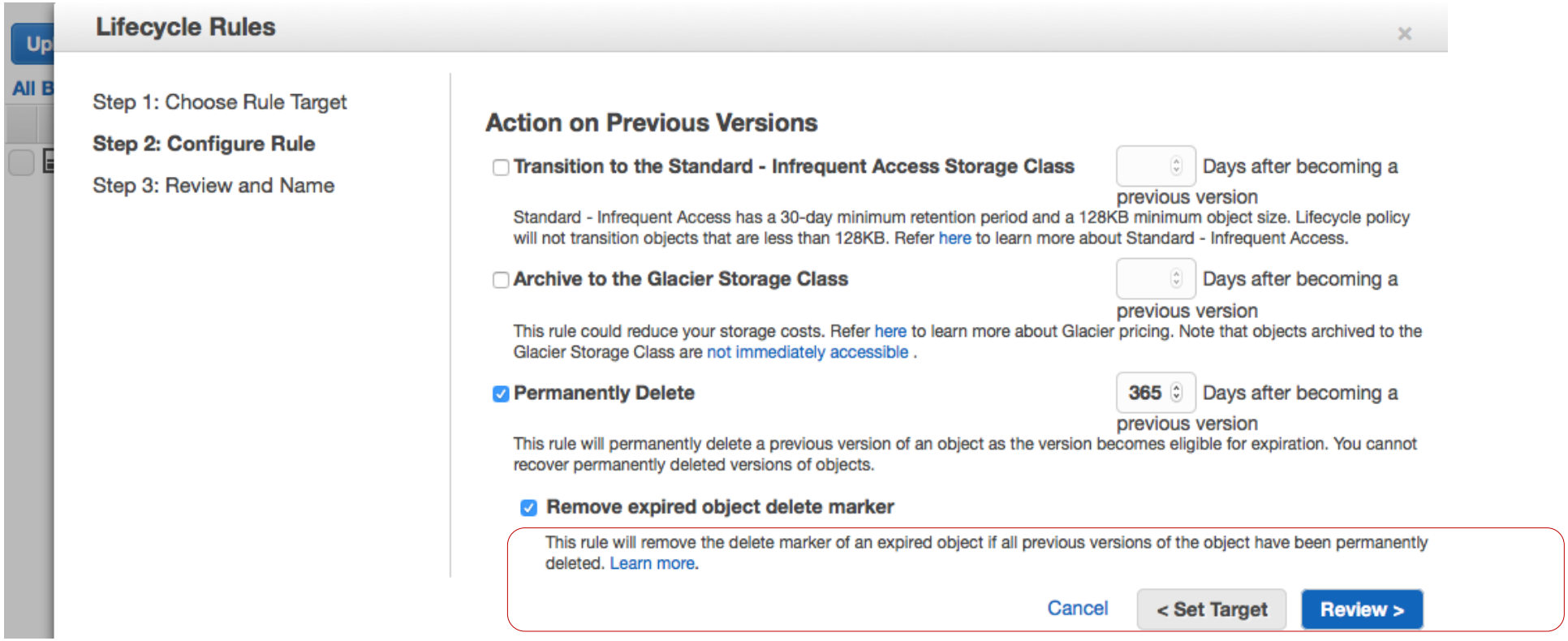
Expired object delete marker policy



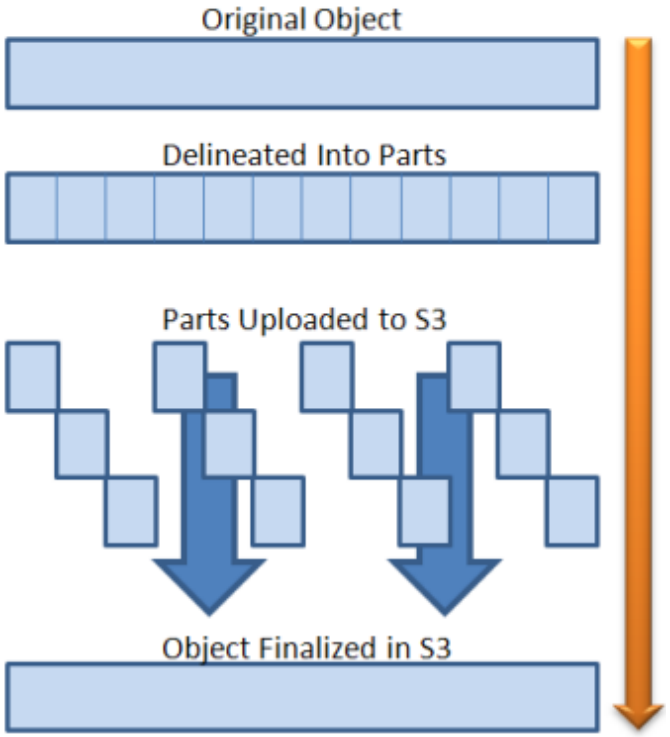
Expired object delete marker

- Deleting a versioned object makes a delete marker the current version of the object
- Removing expired object delete marker can improve list performance
- Lifecycle policy automatically removes the current version delete marker when previous versions of the object no longer exist

Enable policy with the console



Best Practice #3: Parallelizing PUTs with multipart uploads



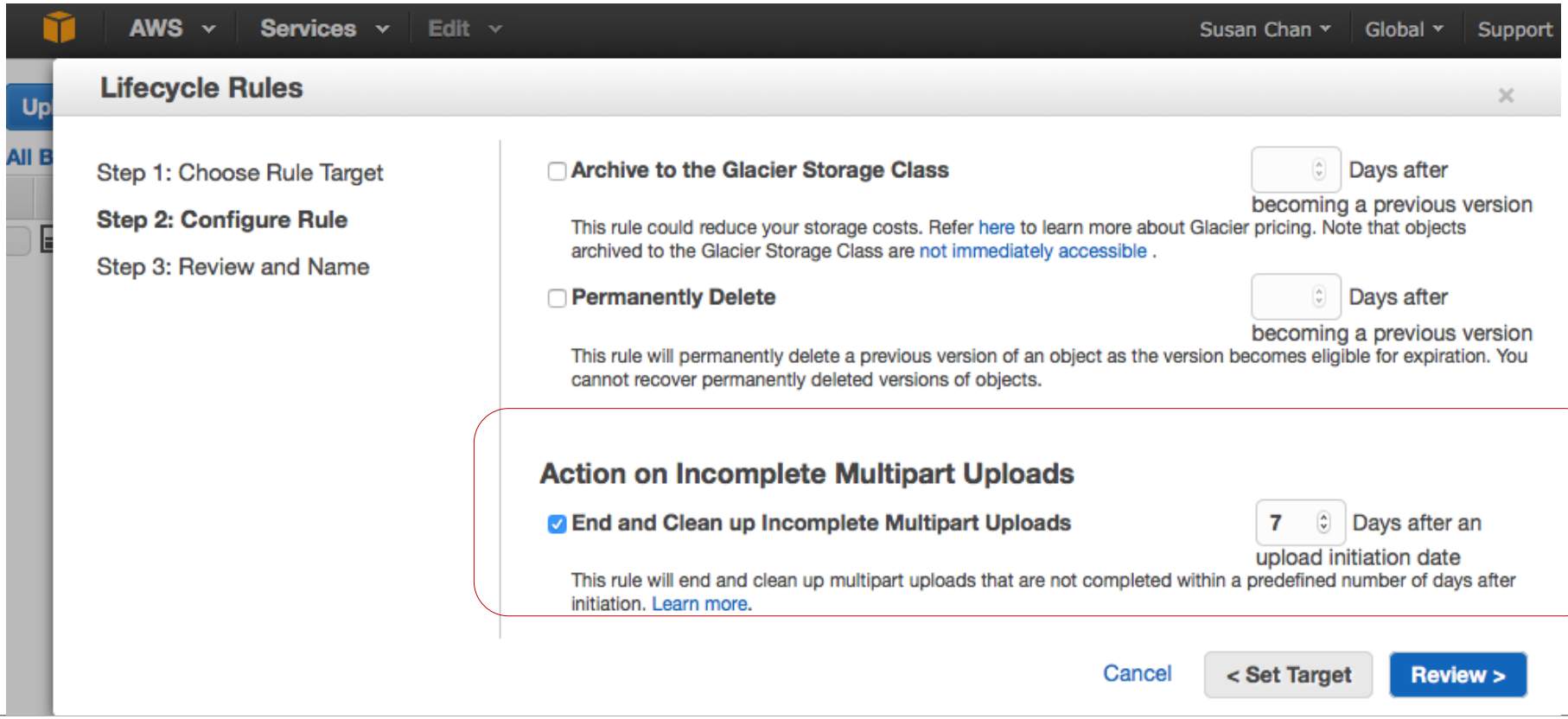
- Increase aggregate throughput by parallelizing PUTs on high-bandwidth networks
  - Move the bottleneck to the network, where it belongs
- Increase resiliency to network errors; fewer large restarts on error-prone networks

Best Practice #4: Incomplete multipart upload expiration policy

Incomplete multipart upload expiration

- Partial upload does incur storage charges
- Set a lifecycle policy to automatically make incomplete multipart uploads expire after a predefined number of days

Best Practice #5: Enable policy with the AWS Management Console



Or enable a policy with the API

Example lifecycle policy

```
<LifecycleConfiguration>
    <Rule>
        <ID>sample-rule</ID>
    <Prefix>MyKeyPrefix/</Prefix>
        <Status>rule-status</Status>
    <AbortIncompleteMultipartUpload>
    <DaysAfterInitiation>7</DaysAfterInitiation>
    </AbortIncompleteMultipartUpload>
    </Rule>
</LifecycleConfiguration>
```

Best Practice #6: Restrict deletes

- Bucket policies can restrict deletes
- For additional security, enable MFA (multi-factor authentication) delete, which requires additional authentication to:
  - Change the versioning state of your bucket
  - Permanently delete an object version
- MFA delete requires both your security credentials and a code from an approved authentication device

Best Practice #7: Distribute key names

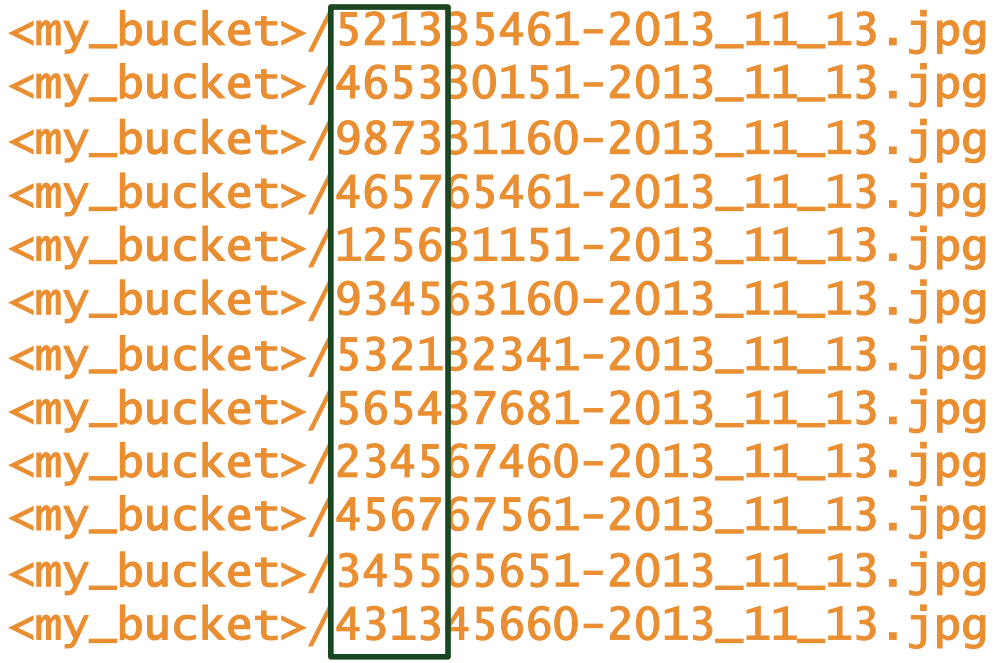
- Use a key-naming scheme with randomness at the beginning for high TPS
  - Most important if you regularly exceed 100 TPS on a bucket
  - Avoid starting with a date
  - Consider adding a hash or reversed timestamp (ssmmhhddmmyy)

- Don’t do this…

```
<my_bucket>/2013_11_13-164533125.jpg
<my_bucket>/2013_11_13-164533126.jpg
<my_bucket>/2013_11_13-164533127.jpg
<my_bucket>/2013_11_13-164533128.jpg
<my_bucket>/2013_11_12-164533129.jpg
<my_bucket>/2013_11_12-164533130.jpg
<my_bucket>/2013_11_12-164533131.jpg
<my_bucket>/2013_11_12-164533132.jpg
<my_bucket>/2013_11_11-164533133.jpg
<my_bucket>/2013_11_11-164533134.jpg
<my_bucket>/2013_11_11-164533135.jpg
<my_bucket>/2013_11_11-164533136.jpg
```

Distributing key names

- Add randomness to the beginning of the key name…



Optimize storage spending by tiering on AWS

Price Blending – Optimizing Storage Costs

Example of leveraging storage class and lifecycle management lower their costs

- 1/3 * $0.03
- + 1/3 * $0.0125
- + 1/3 * $0.007

$0.0165

AWS Storage Services

Amazon **S3**

Durable object storage for all types of data

Economics

Pay as you go

No upfront investment
No commitment

No risky capacity planning

Amazon **Glacier**

Archival storage for infrequently accessed data

Easy to Use

Self service administration

SDKs for simple integration

Amazon **EBS**

Block storage for use with Amazon EC2

Durable, Secure

Avoid risks of physical media handling

High performance

Amazon **EFS**

File storage for use with Amazon EC2

Agility, Scale

Reduce time to market

Focus on your business, not your infrastructure

Summary: the Amazon storage options

S3

- Object storage: data presented as buckets of objects
- Data access by using APIs over the Internet

Glacier

- Archival storage: data presented as vaults/archives of objects
- Lowest-cost storage, infrequent access by using APIs over the Internet

EBS

- Block storage (analogous to SAN): data presented as disk volumes
- Lowest-latency access from single EC2 instances

EFS

- File storage (analogous to NAS): data presented as a file system
- Shared low-latency access from multiple EC2 instances

Storage Gateway

- Back up and archive data into S3 and Amazon Glacier

Hands-on Lab #2: S3

Hands-on Lab#2: S3

- The lab will cover following topics:
  - Create a bucket
  - Add an object to your bucket
  - Manage access permissions on an object
  - Create a bucket policy
  - Use bucket versioning

- If you do not have existing qwiklabs account, please create a free one at https://qwiklabs.com/
- Once you have the account created, complete following lab: Click Here.

Content Delivery Network

AWS Cloud Global Infrastructure

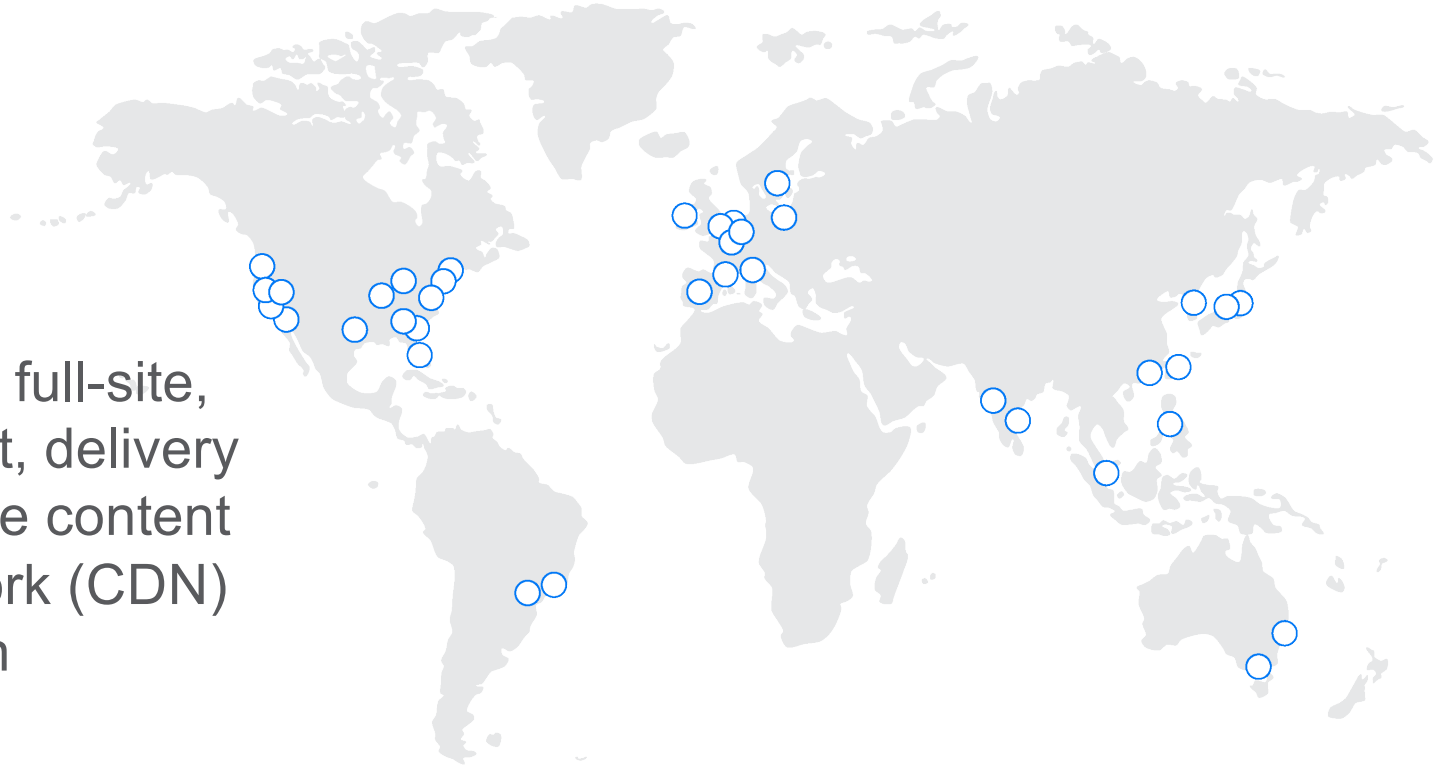
The AWS Cloud spans 69 Availability Zones within 22 geographic Regions.

Source: https://aws.amazon.com/about-aws/global-infrastructure/

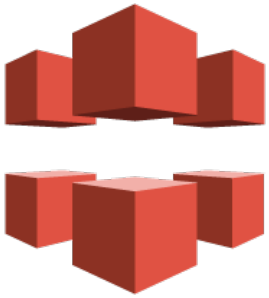
Benefits of CDN?

- Web Service that speeds up distribution of static web content
- Delivers content through a worldwide network of data centers called edge locations
- Reduces Latency
- Puts content closer to users
- Offload traffic from the origin by caching content and pooling connections

AWS Content delivery

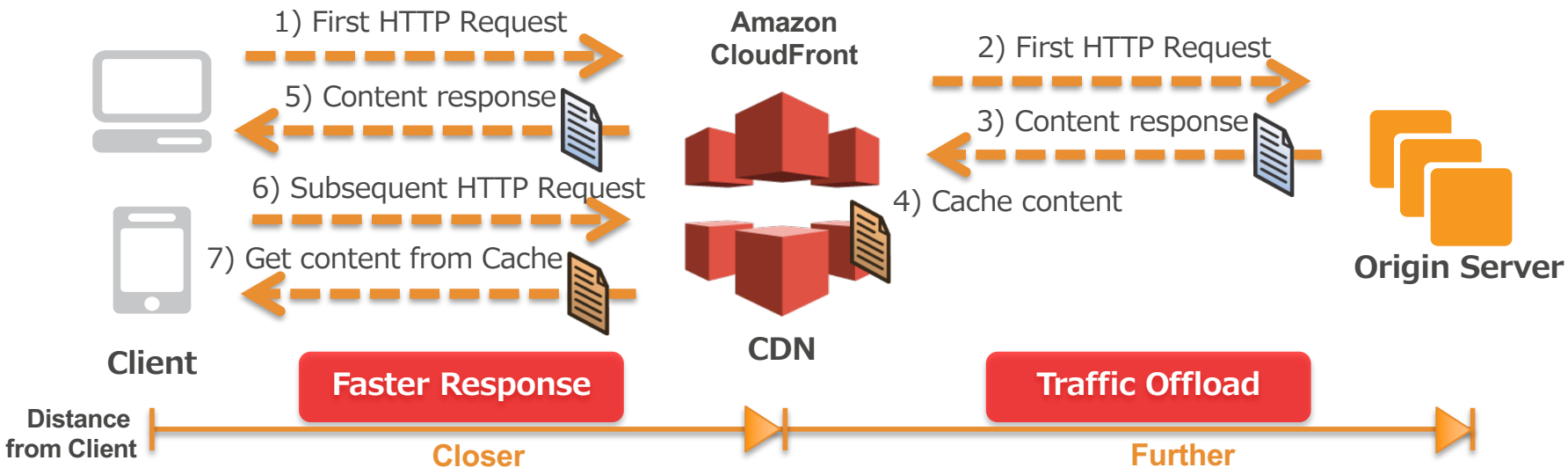


- AWS provides full-site, or media asset, delivery via a worldwide content delivery network (CDN) called Amazon **CloudFront**.



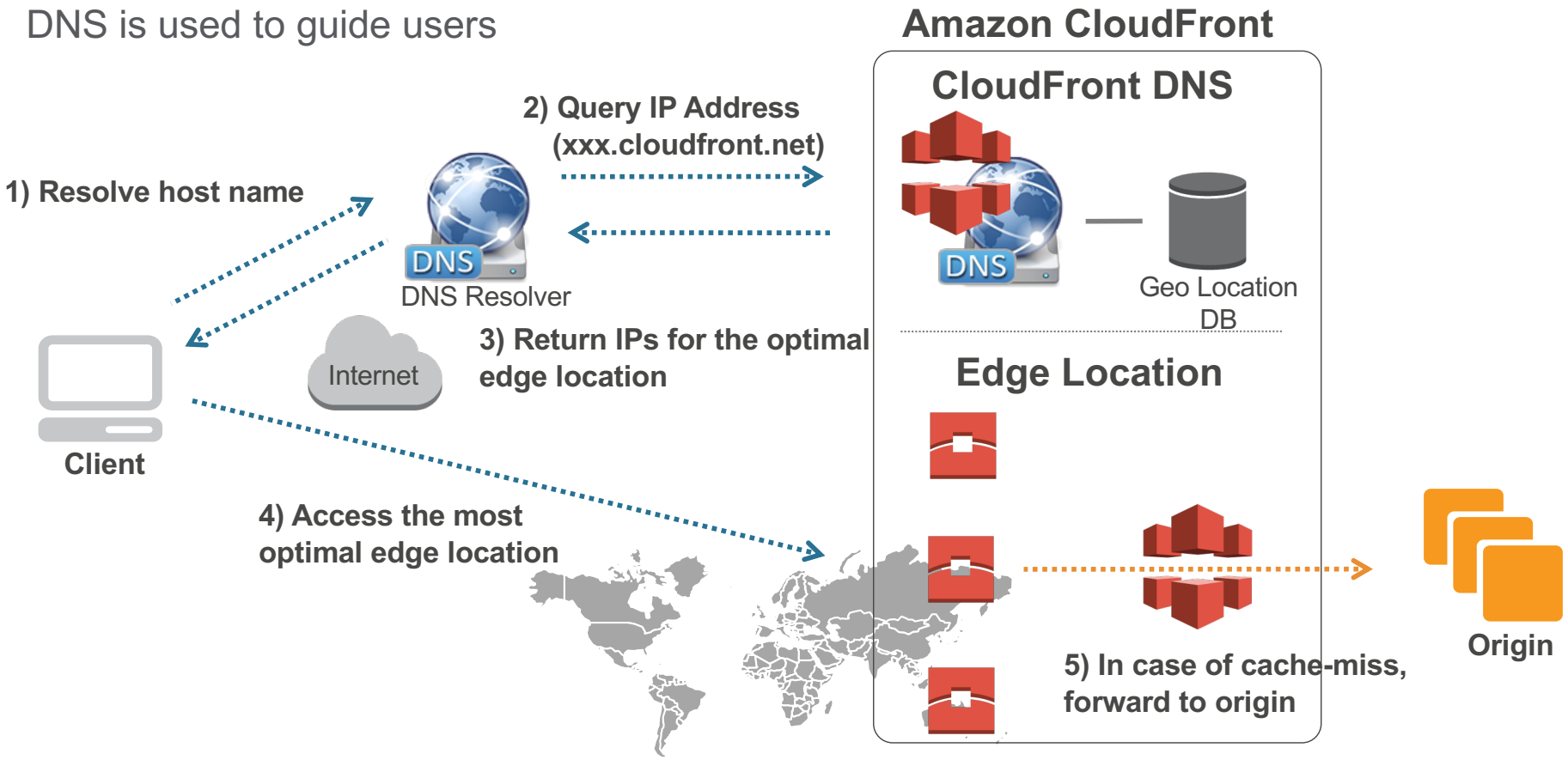
What is Amazon CloudFront?

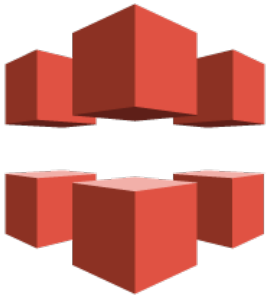
CloudFront is a CDN



Guiding users to the Optimal Edge Location

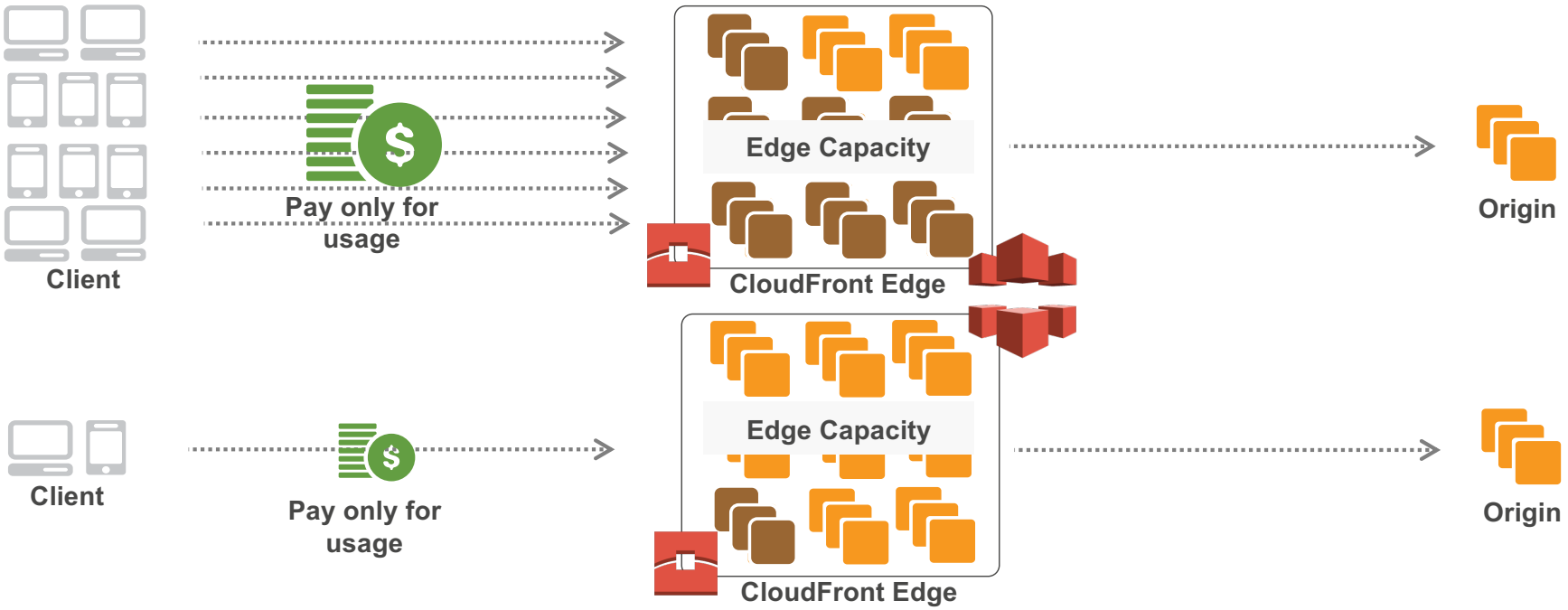
DNS is used to guide users





Scalability, Availability, Offload

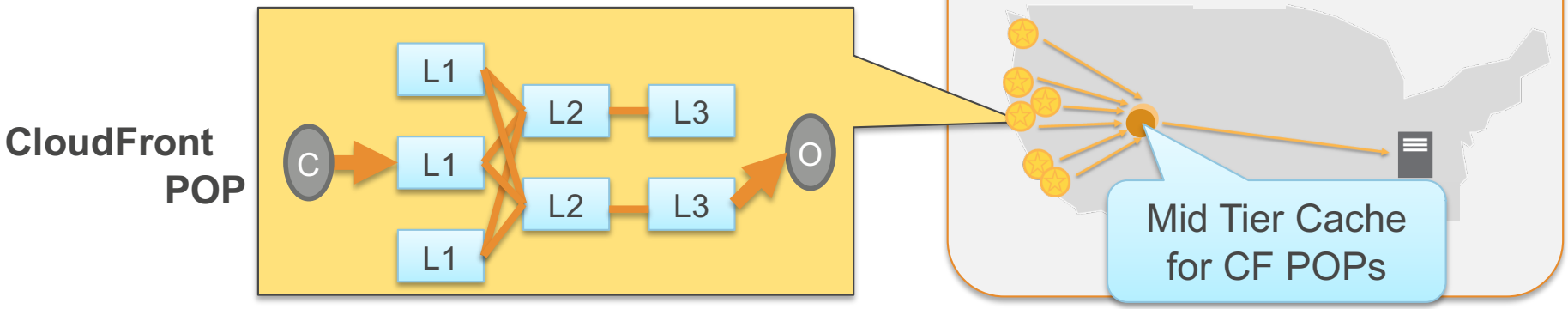
Scalability and Pay as You Go Pricing

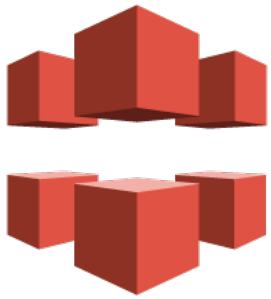


CloudFront scales automatically, eliminating the need for difficult capacity planning and loss of opportunity

Offload

- Each CloudFront Edge Location (POP) has multiple caching layers to maximize offload and performance.
- Regional Edge Caches provide a mid-tier cache for CloudFront POPs within a region
- Current Regional Edge Caches are available in Tokyo, Singapore, Sydney, Seoul, N. Virginia, Oregon, Sao Paolo, Frankfurt, Mumbai



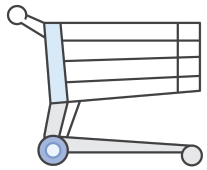


Ease of Use

CF Use Cases

Amazon CloudFront can be used for different purposes. There is no need to sign up for different services with different rates depending on the CDN use case

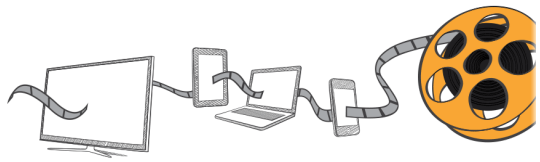
- **Static / Dynamic Web Site Delivery**
  (Cacheable and non cacheable content including web sites and APIs)



- **Static File Delivery**
  (Static objects embedded in web sites and apps)

- **Large File Downloads**
  (Games, Video, Software Update, etc)

- **Streaming Video (Live/Ondemand)**
  (Smooth Streaming、HLS、HDS delivery of video)



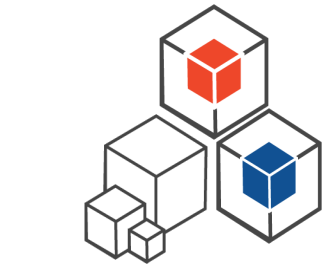
Cache Invalidations

- AWS Management Console or API

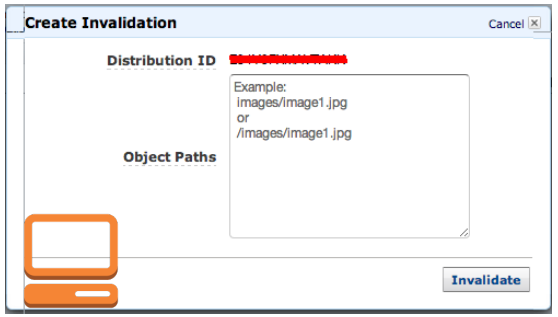
- Invalidation by content paths
  - Up to 3,000 concurrent path requests

- Invalidation by wildcard(*) paths
  - Up to 15 concurrent wildcard path requests
  - No limit on number of objects

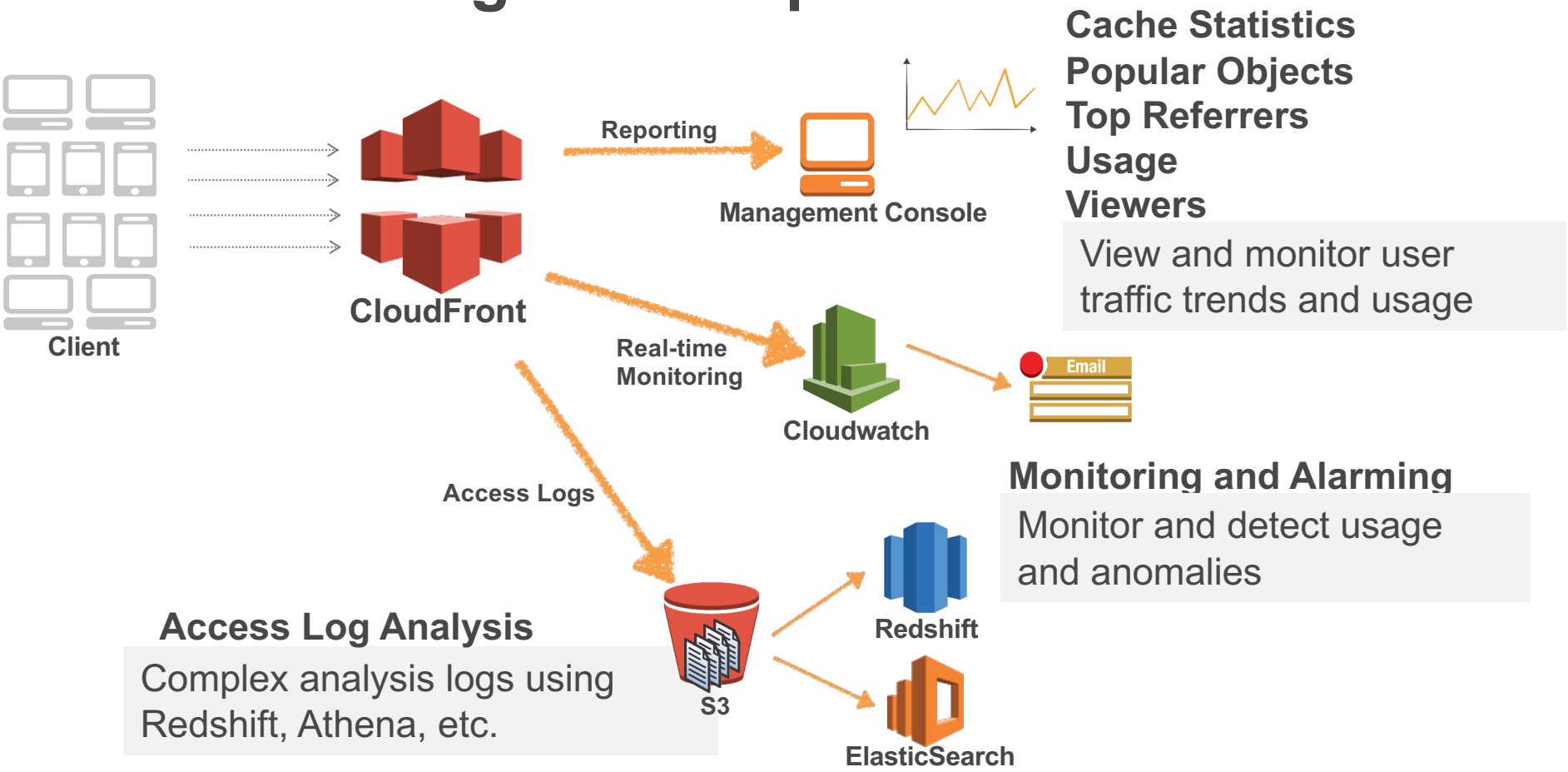
- Faster Invalidations
  - 90% of edge servers completed within 5 seconds
  - 100% Completed within 1 minute



AWS SDK / CLI / API



CloudFront Logs and Reports



Restricting Access

- CloudFront provides the capability to restrict access to users based on Geographic Location
- Restrictions can be performed using either a whitelist or blacklist
- Restrictions are specified on the Restrictions tab of the distribution

Accessing your Website

- CloudFront will generate a unique URL similar to:
  - http://d210owefw5zol3z.cloudfront.net

- In order to use your Domain name, a CNAME DNS record must be created to point your existing domain (www.mydomain.com) to your CloudFront URL

- Wildcards are support (*.myexample.com) can all map to the CloudFront URL

Hands-on Lab #3: CloudFront

Hands-on Lab# 3: CloudFront

- The lab will cover following topics:
  - Create a new Amazon CloudFront Distribution
  - Use your Amazon CloudFront Distribution to serve an image file
  - Delete your Amazon CloudFront Distribution when it is no longer required

- If you do not have existing qwiklabs account, please create a free one at https://qwiklabs.com/
- Once you have the account created, complete following lab: Click here.

Next Steps

Next Steps

- **Read**:
  - Review the lecture #3 slides
  - Complete extra reading
  - Complete lab#2 and Lab#3
  - Read S3 and CloudFront FAQ from AWS site
  - Review S3 and CloudFront user guides from AWS

- **More Labs:**
  - http://docs.aws.amazon.com/cli/latest/userguide/cli-chap-welcome.html
  - http://docs.aws.amazon.com/cli/latest/userguide/using-s3-commands.html