
AWS
re:Invent
DECEMBER 2 – 6, 2024 | LAS VEGAS, NV

STG354-NEW

New default Amazon S3 data integrity protections

Raghu Balivada
Principal Engineer, Amazon S3
AWS

Akshat Sandh
Sr. Product Manager, Tech, Amazon S3
AWS

Raghu

Akshat

Geospatial or lunar imagery

Internet of Things (IoT) sensor data

Medical images and records

Analytics

Customer call-center records

Digital record preservation

Data lakes

Mobile sync and storage

Compliance records

Media master files

Home video recordings

Pharmaceutical study data

Model checkpoints

Seismic and reservoir simulation data

DNA sequences

Backups

Website hosting

Surveillance video/closed-circuit television

Machine learning training data

Log files

Media assets

Financial records

Autonomous vehicle data

User-generated content

Meteorological and environmental research

Oil and gas topography

AWS enables you to protect the durability of your data

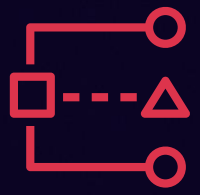
MEET YOUR NEEDS WITH NATIVE AWS OFFERINGS

AWS Backup

MFA token

S3 Object Lock

Permissions



AWS Identity and Access Management Access Analyzer

AWS Key Management Service (AWS KMS)

S3 Versioning

S3 replication

Find the difference between these 2 images

Find the difference between these 2 images

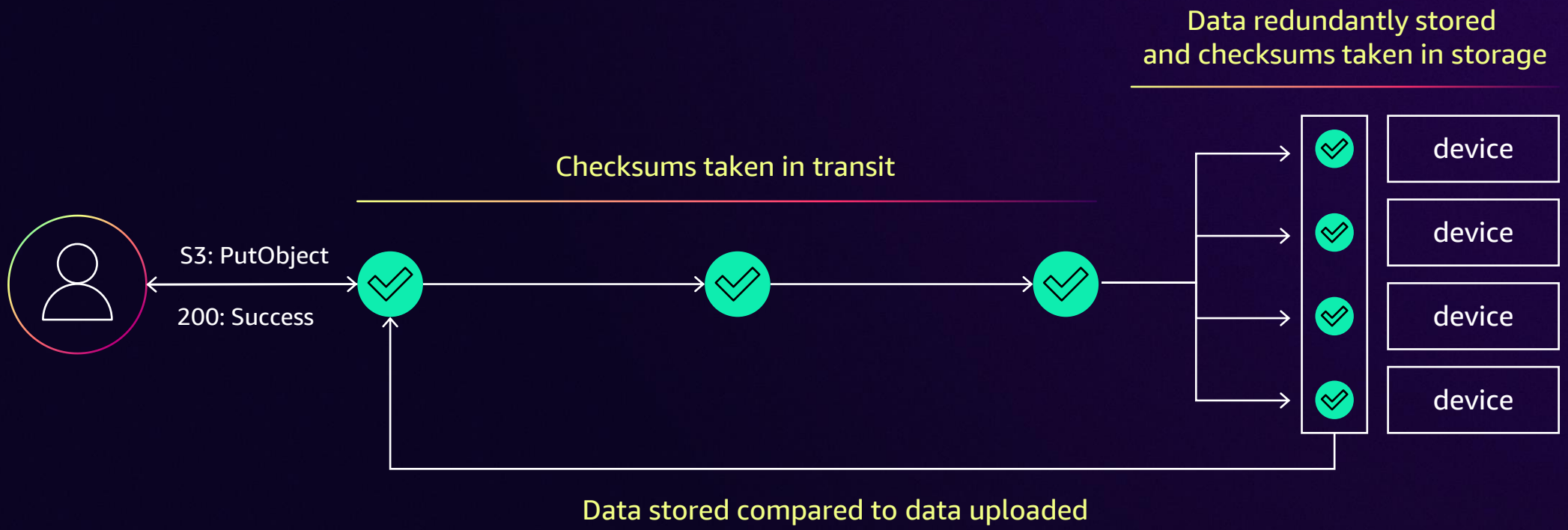
CHECKSUMS HELP CONFIRM THE INTEGRITY OF YOUR DATA

SHA-1: 8r0d5ezWizLzeFUi1Ko/vUHbQ5w=

SHA-1: AZLZVuL/xtHGXG2yhs4ZNU7Zrgk=

Amazon S3 data integrity basics

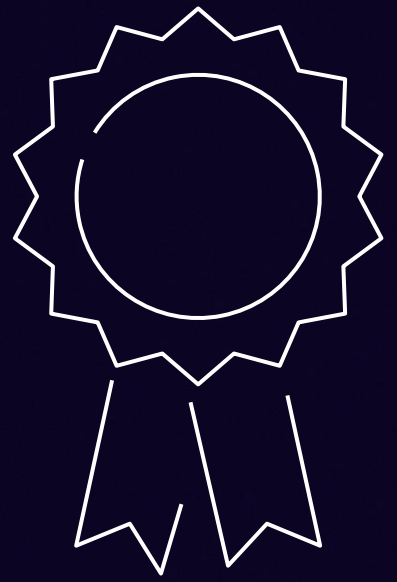
END-TO-END CHECKSUMS TO VALIDATE DATA INTEGRITY



Pattern

“Chain of custody”

Maintaining a chain of custody for your data is key as it moves across systems



Congratulations!

You are now a data integrity expert!
(relative to the global population)

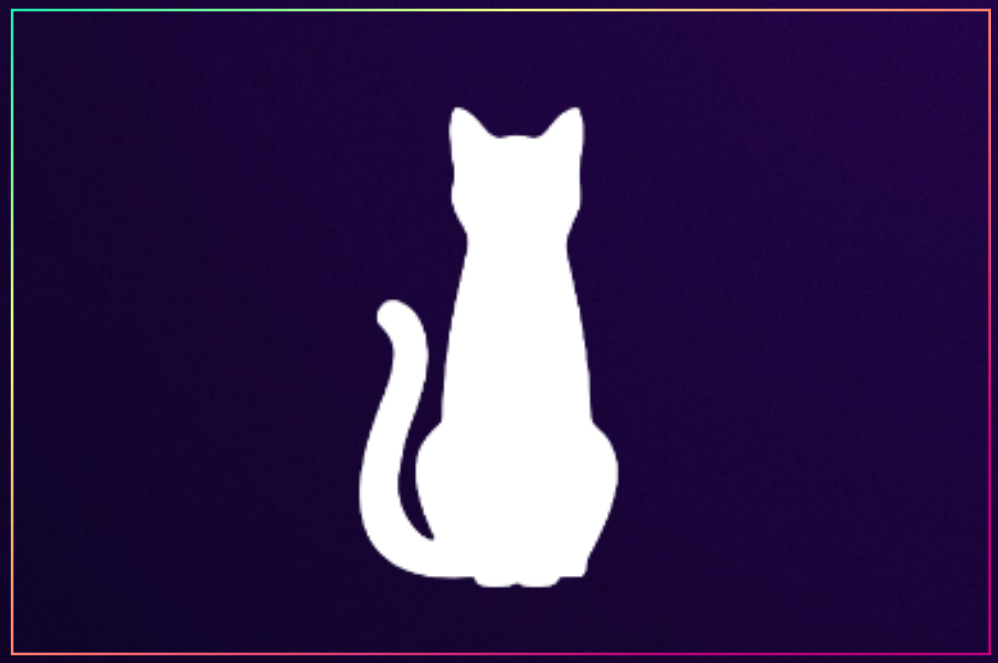
01 Content-MD5 and the Amazon S3 ETag

02 Amazon S3 checksum enhancements

03 AmazonS3 default checksums

Session legend

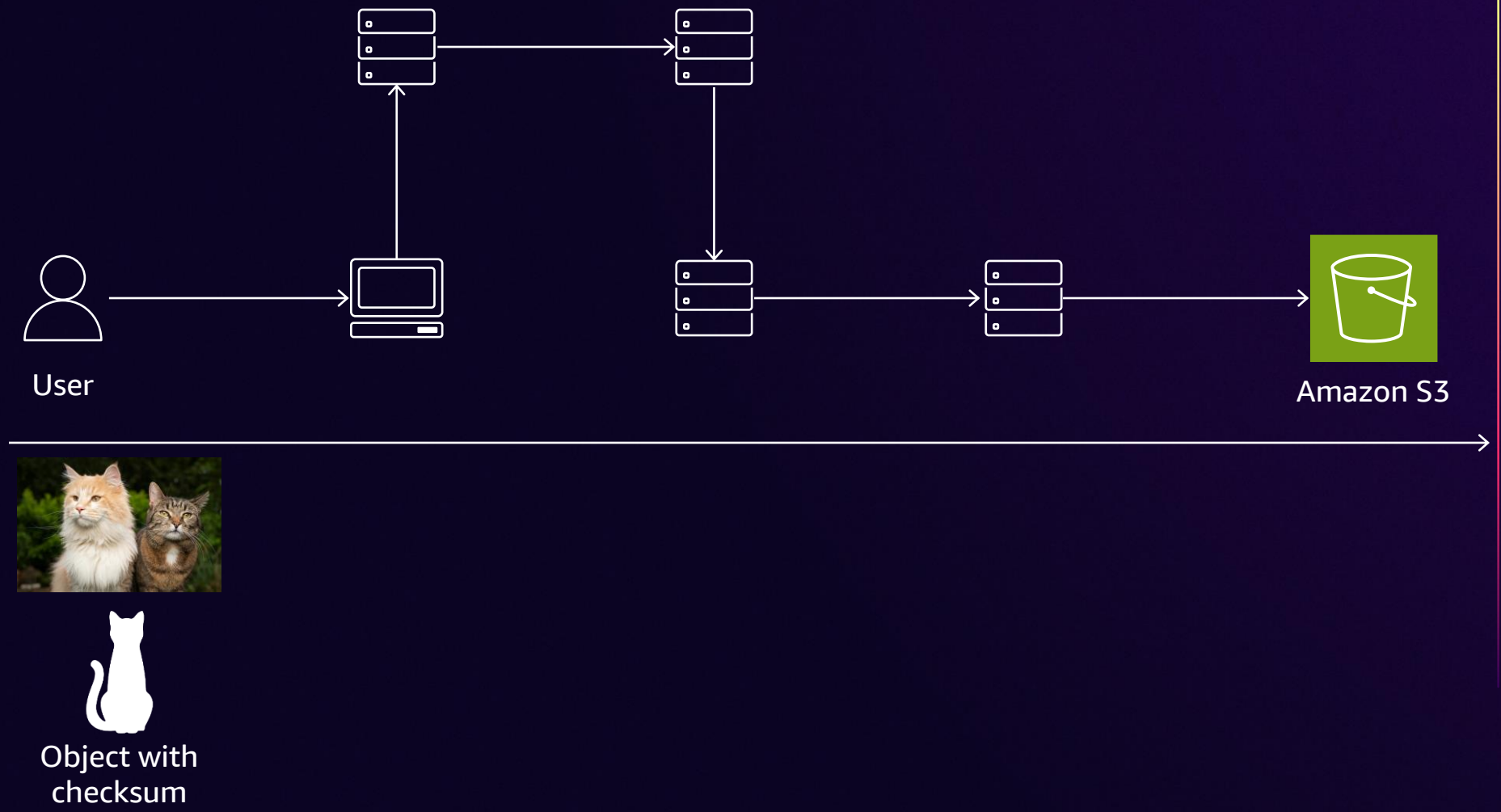
Object



Checksum

1 Content-MD5 and the Amazon S3 Etag

Using Content-MD5 to validate integrity on upload



MD5 checksum of the object generated by the user traverses the public internet with the object

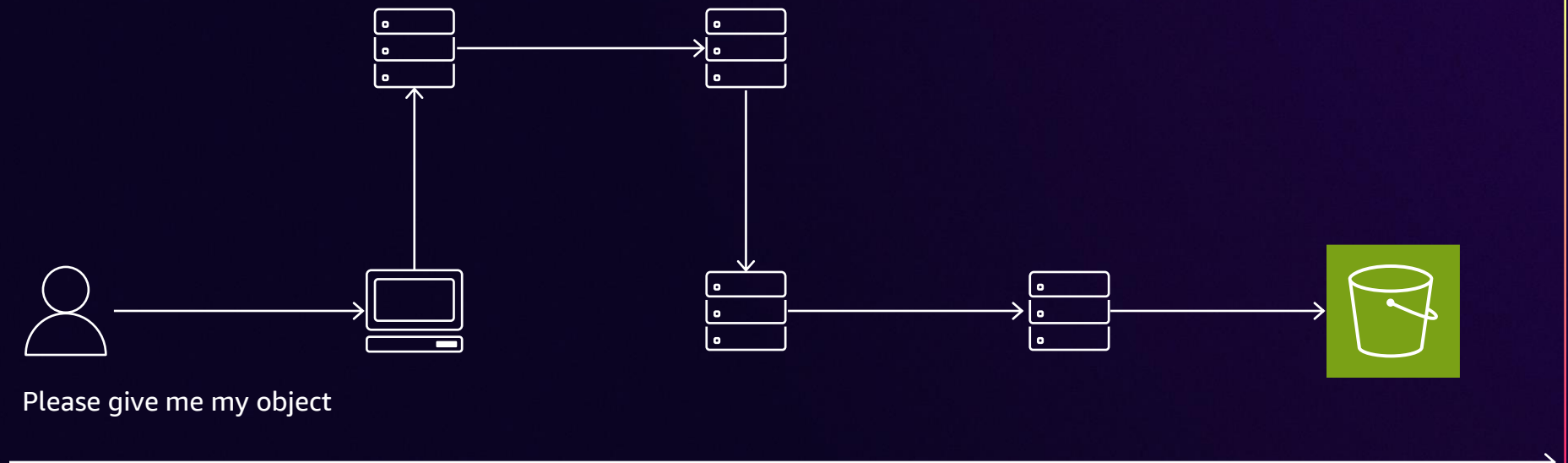
Amazon S3 validates the checksum before storing the object

Content-MD5 is the header in the HTTP spec

But

How do I check object integrity when I download the object?

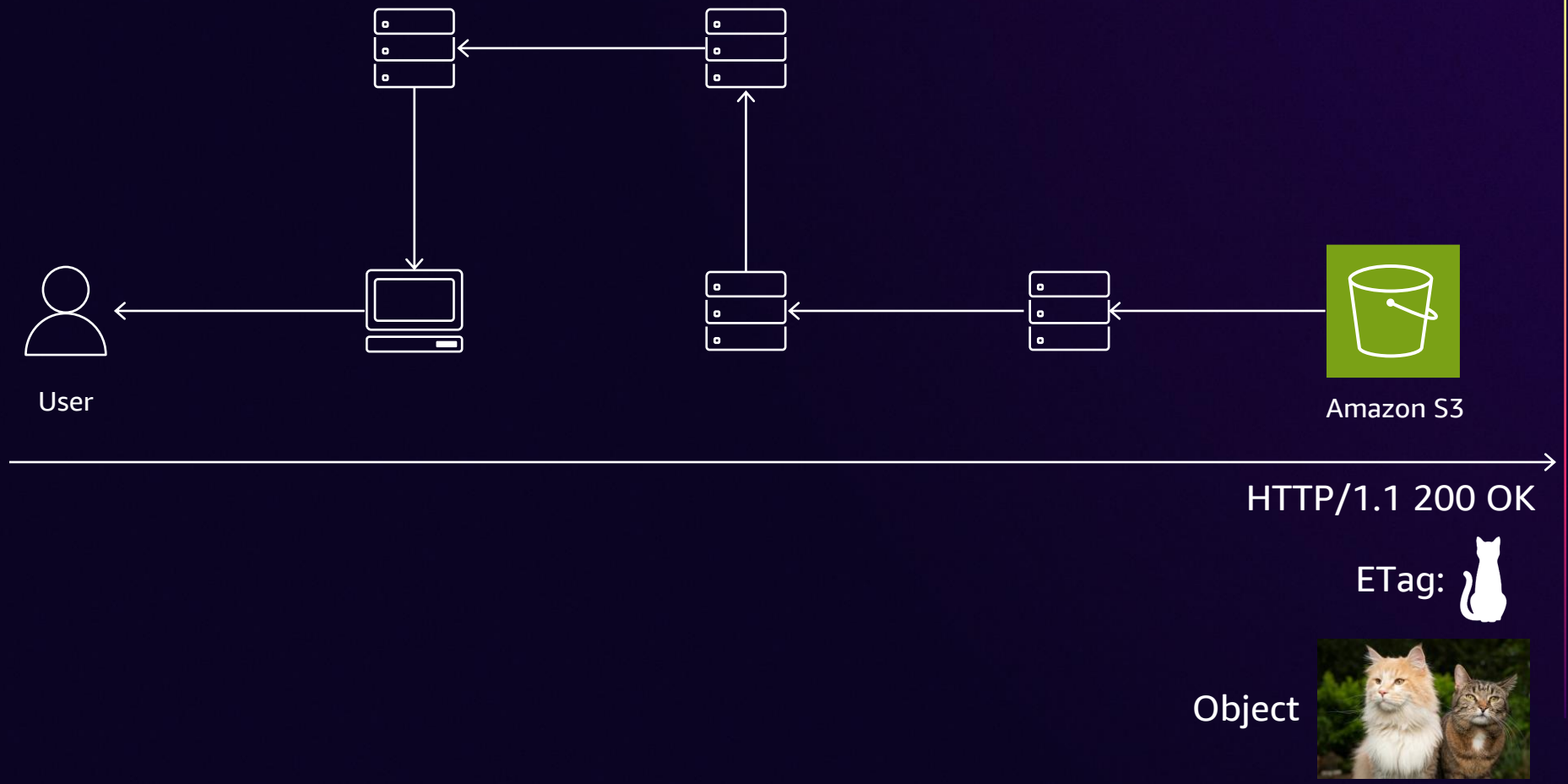
Using the Amazon S3 ETag to validate integrity on download



You request the object

Amazon S3 receives the request and validates if you have access to the object

Using the Amazon S3 ETag to validate integrity on download



Amazon S3 returned the object with the Entity Tag (ETag)

The ETag is a checksum of your object

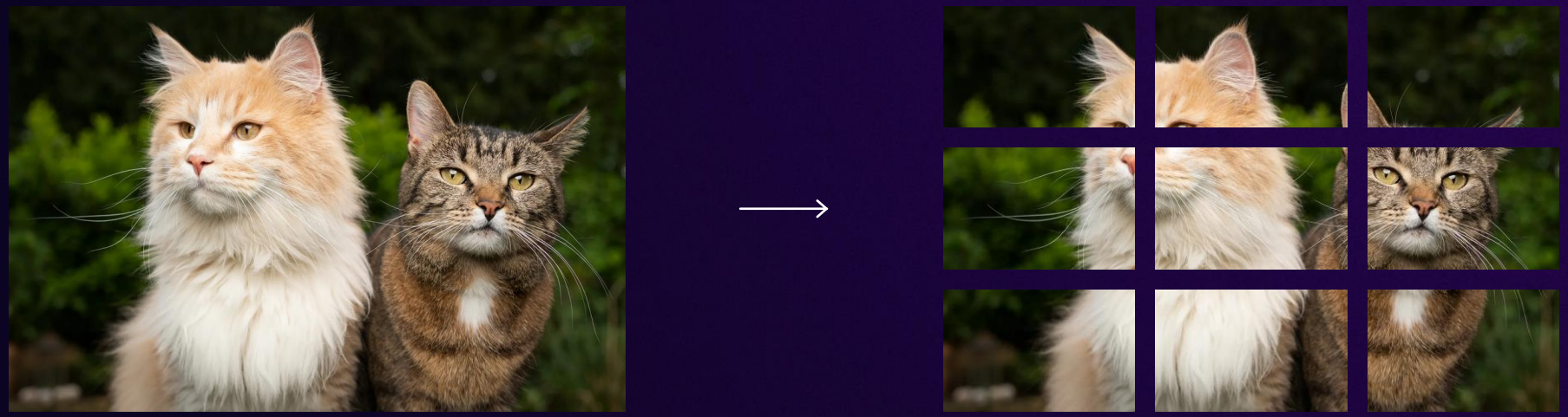
Simple, right ?

Why is the session 1 hour long?

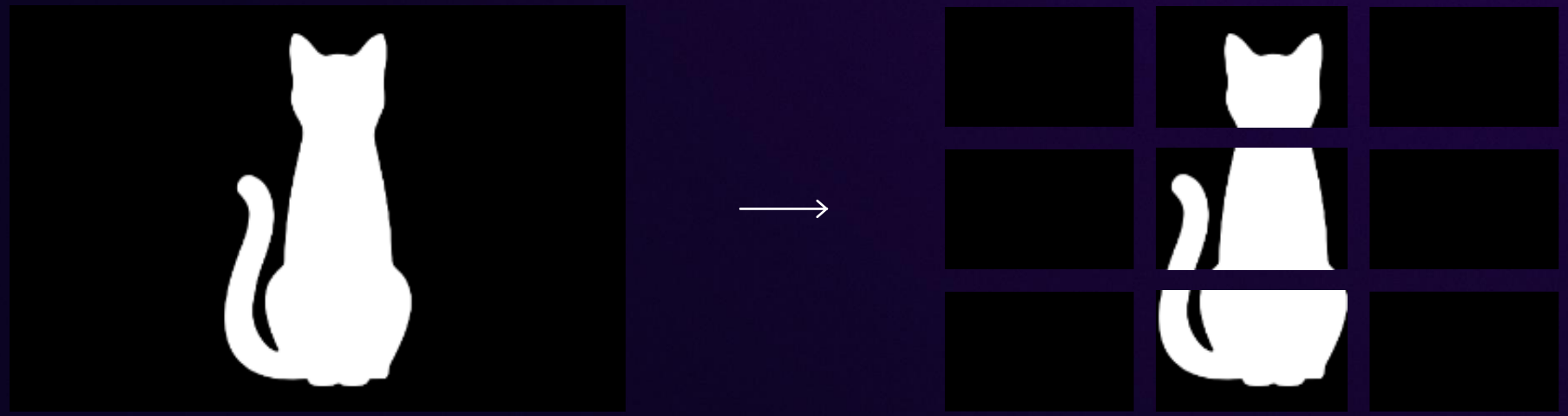
Amazon S3 multipart uploads (MPU)

THE OBJECT IS UPLOADED IN PARALLEL CHUNKS

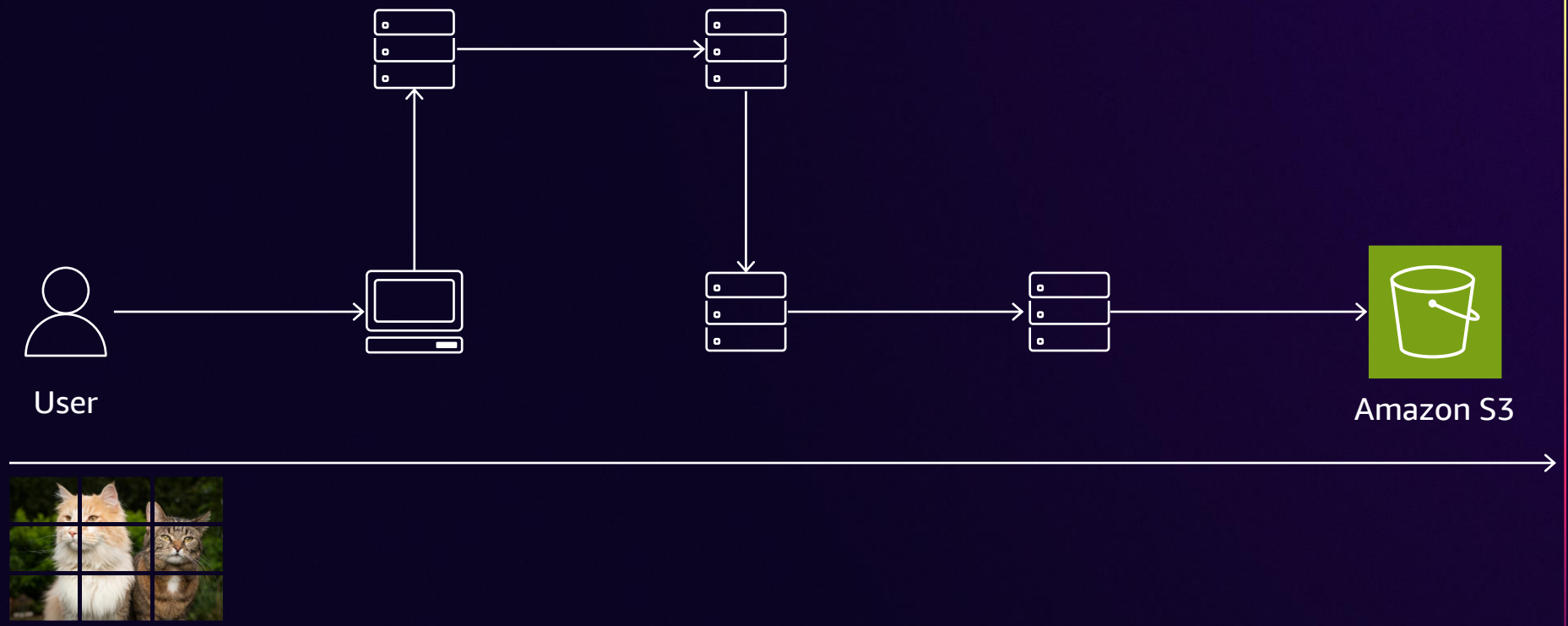
Step 1: Object is split into multiple parts



Step 2: The checksum of each part is computed as it’s uploaded

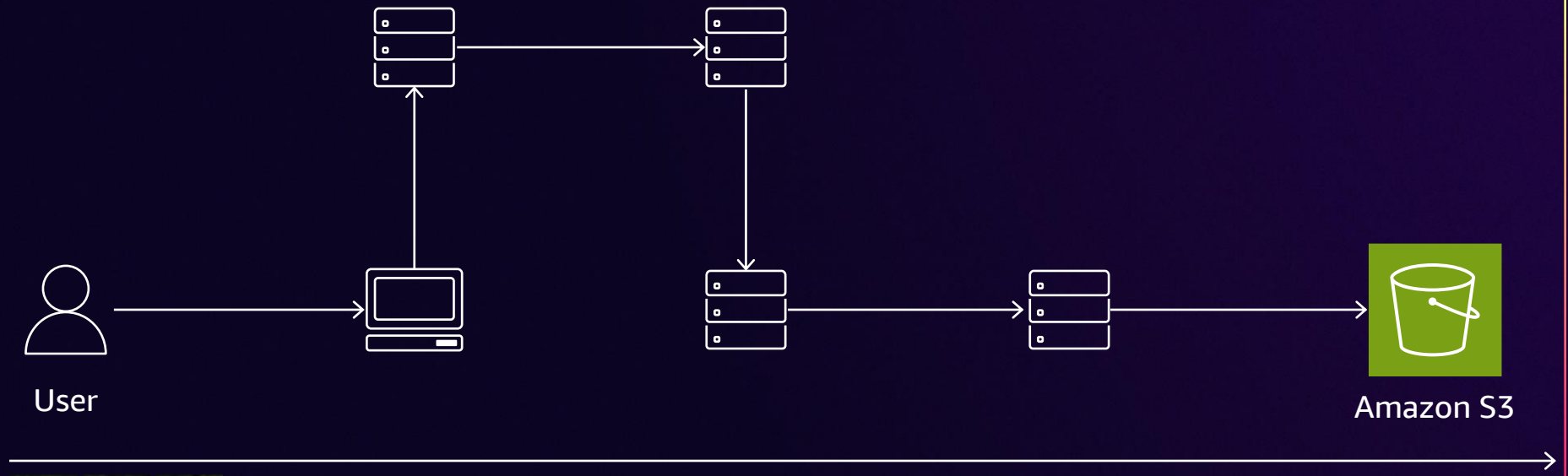


Using Content-MD5 to validate integrity on MPU

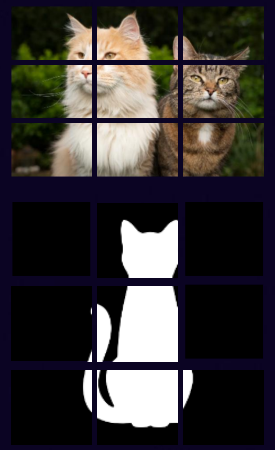


The object is split into parts

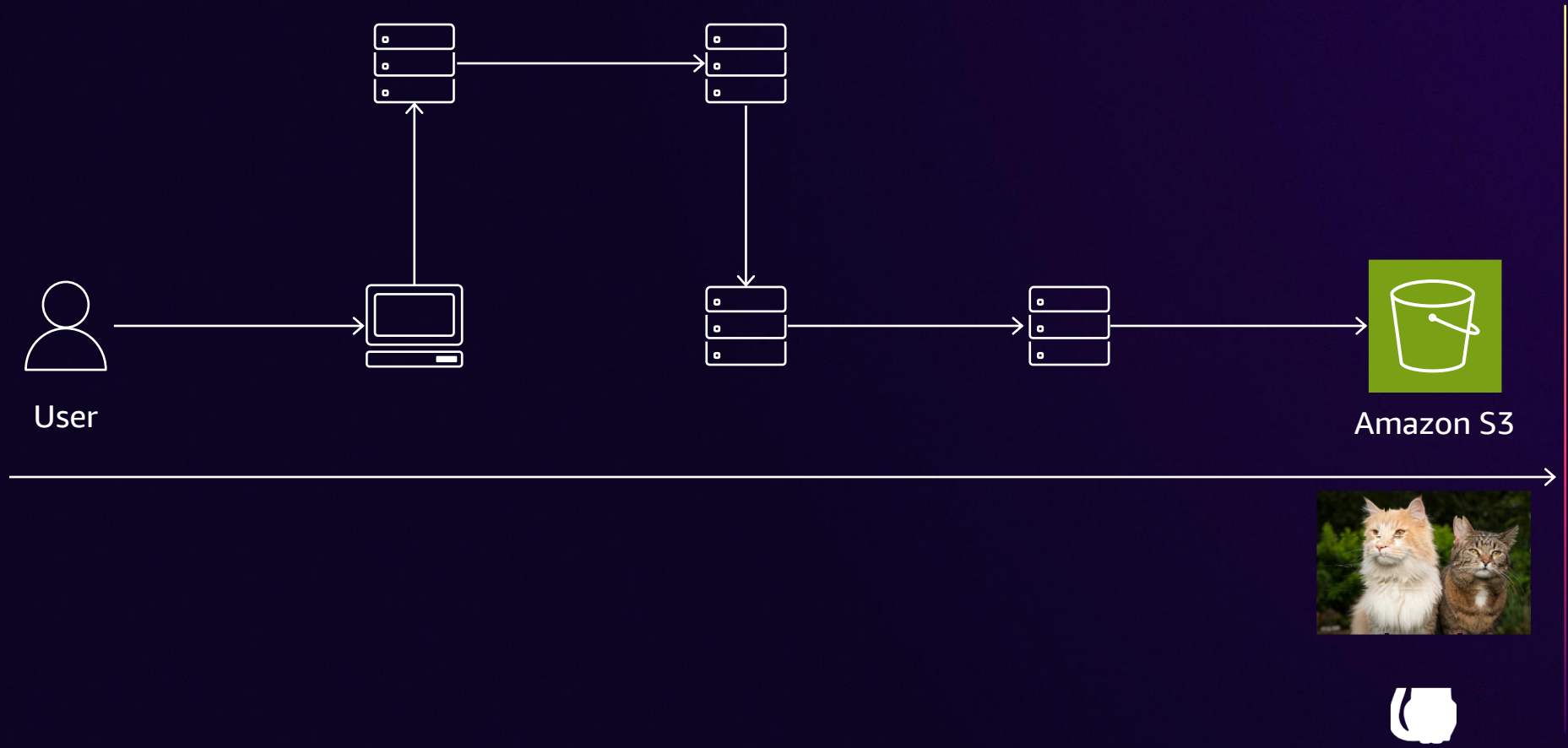
Using Content-MD5 to validate integrity on MPU



Each part is uploaded with its own checksum



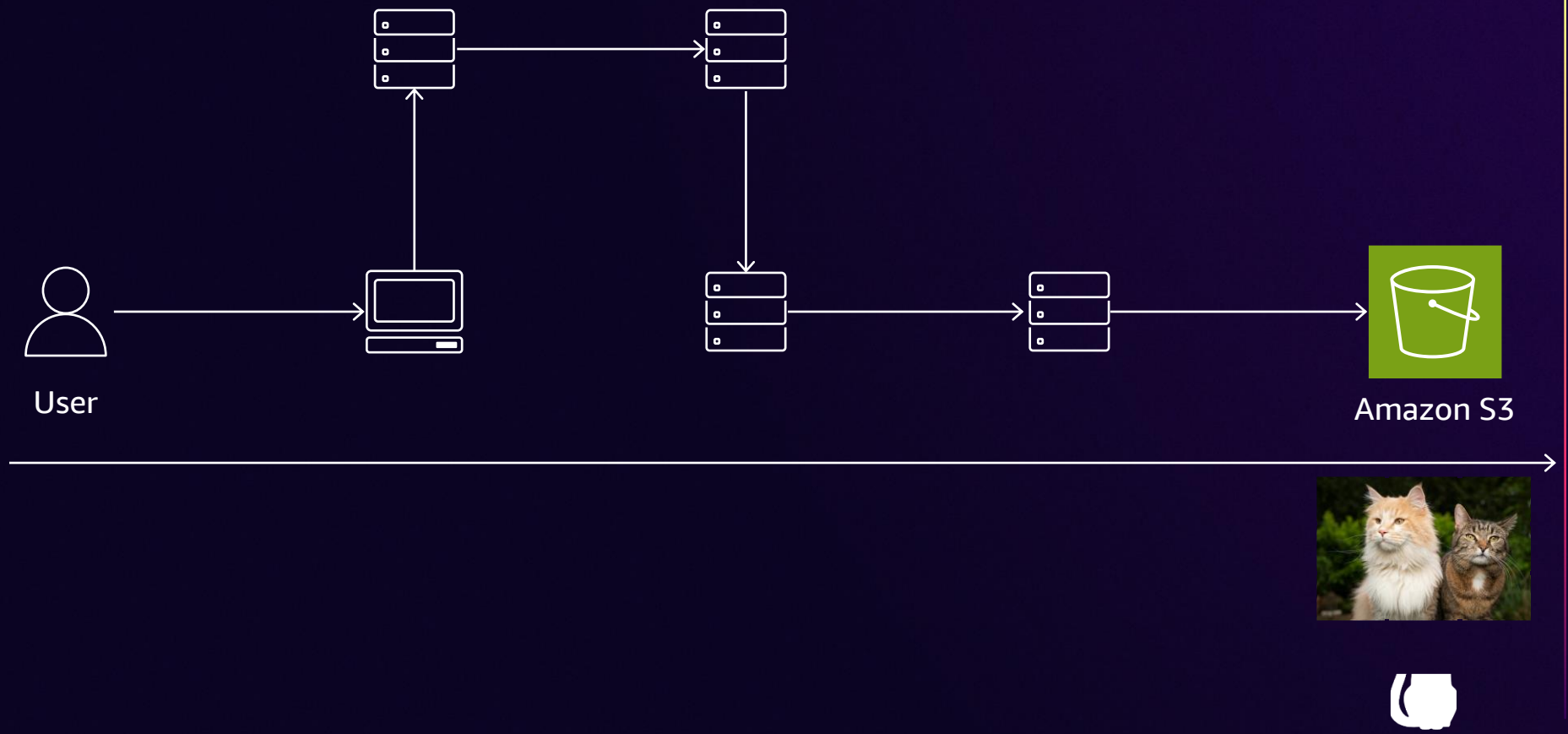
Using Content-MD5 to validate integrity on upload



Amazon S3 reassembles the object

The checksum of the reassembled object is not the same as that of the whole object

Using Content-MD5 to validate integrity on upload



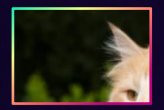
Amazon S3 reassembles the object

The checksum of the reassembled object is not the same as that of the whole object

ETag of the checksums of MPU objects

DETAILS ON HOW THE CHECKSUM IS CALCULATED

fa0bf5fdcf9894c0d3c1a0
02bba08ea0

 = 1391d901cbfb3c538bcc96aab5fdd798

 = ec7e213b880cd469593feb7a0e52884c

 = 037d519b58da2c9d64255cc43ff0256d

 = 09b3e616e86102d107b27a6e336c4ff9

 = d13b56e3f4afcfade1c128637ee32786

 = ac6a98310da49c7716c421642f93342a

 = 22ef28b61617484231 65a810f19973f3

 = ca54f7358d128443283633f90fe3fd57

 = 34a456303d1f77680e3ffc6647ad495e

F242af177779f5
d0d4e45cd3acaa
12e5-9

Pattern

"Checksums are a function of object data"

Object checksums should only be available to users who can get the object

MD5 and the Amazon S3 ETag

MPU AND ENCRYPTION

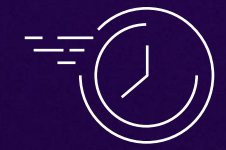
Upload type	MD5
PUT	fa0bf5fdcf9894c0d3c1a002bba08ea0
MPU	f242af177779f5d0d4e45cd3acaa12e5-9

Encryption type	MD5
SSE-S3	Can be used to validate integrity
KMS/DSSE-KMS/SSE-C	Cannot be used to validate integrity

2 Amazon S3 checksum enhancements

Choosing the right checksum algorithm

ALGORITHMS HAVE TRADEOFFS, AND CHOOSING THE RIGHT ONE IS KEY



SHA-256 **SHA-1**

SHA (secure hash algorithms) have to be calculated in serial

They are slow but offer a much lower collision rate (2 distinct objects having the same checksums)

MD5

CRC32/32C

CRCs (cyclical redundancy check) algorithms – very fast but not cryptographically secure

Best used for over-the-network integrity validation

Demo

Providing information on part boundaries

Consistency across encryption modes

Demo

Providing information on part boundaries

Consistency across encryption modes

Integrity validation using Amazon S3 additional checksum options

WORKS FOR MPU OBJECTS, CONSISTENT ACROSS ENCRYPTION TYPES

Upload type	SHA-1
PUT	bb1223d01b88f8cf9a6c3273de825344afd805ab
MPU	694ba911fdd6e0e4538ed879f5c49340d101090b-9

Encryption type	SHA-1
SSE-S3	Can be used to validate integrity
KMS/DSSE-KMS/SSE-C	Can be used to validate integrity

3

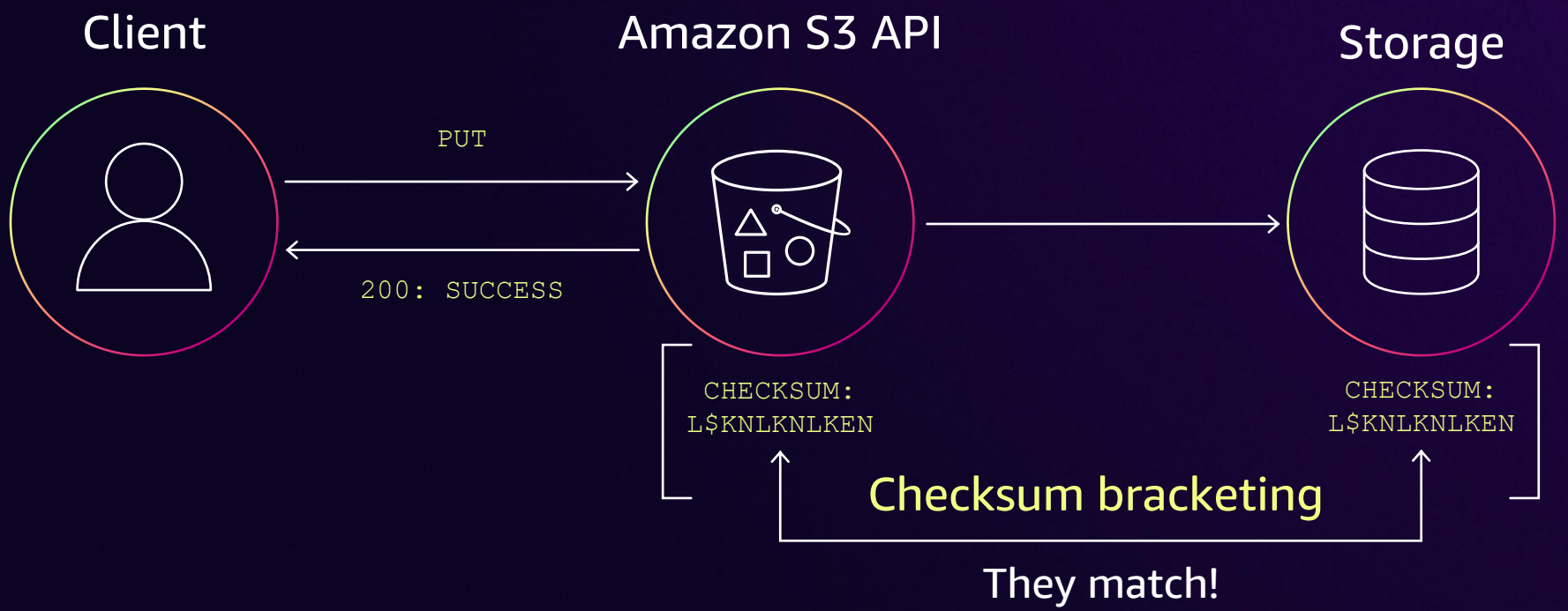
New

Amazon S3 default checksums

Customer needs we set out to solve

- Full object checksums for objects uploaded in parts
- Default checksums for objects uploaded without checksums
- Work consistently across encryption modes
- Remove the need to opt in/make client-side changes

Durability in flight



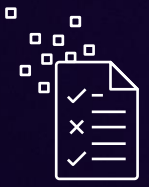
Durability in flight

GA 12/1/2024

New data integrity options

DATA VALIDATION OVER THE WIRE, NEW CHECKSUM INFO IN METADATA

Amazon S3's checksum over-the-wire techniques extended out to customer applications



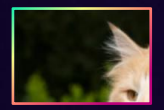
Whole object checksums taken using CRC64, stored in object metadata

New data validation applied by default, with no changes required to customer applications

CRC-based checksums are combinable

YOU CAN COMBINE PART-LEVEL CHECKSUMS TO GET THE CHECKSUM OF THE OBJECT

LQJtNrLOxbo=

 = 9TUStcoOD4W7

 = 4kcnJFxGIZpI

 = Hk4hm9UFFMKz

 = FqtE8vvgoaLl

 = mRsw018ojKdX

 = 1vLiPoMzlu0ra

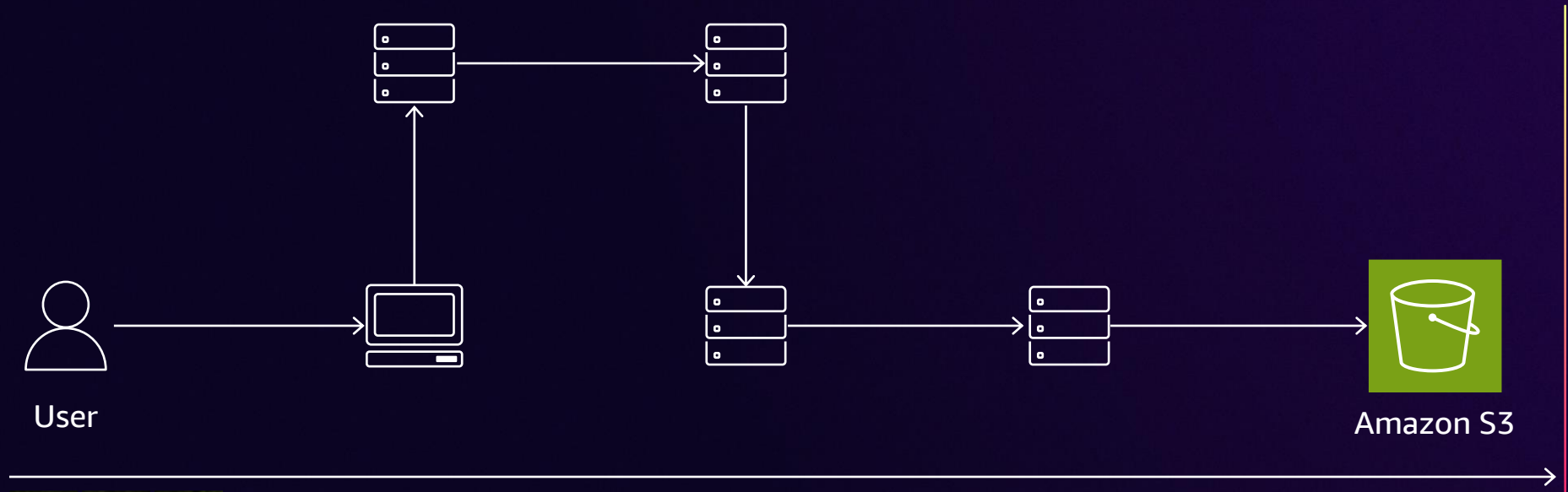
 = ahKi0L3zvmPb

 = 0RRZAQVvh2R6

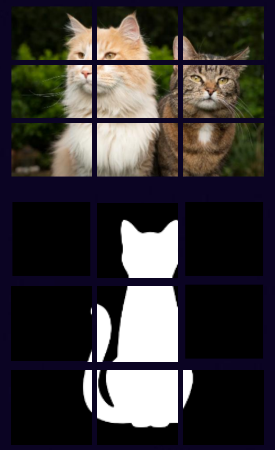
 = MXncUAeXWaiI

→ LQJtNrLOxbo=

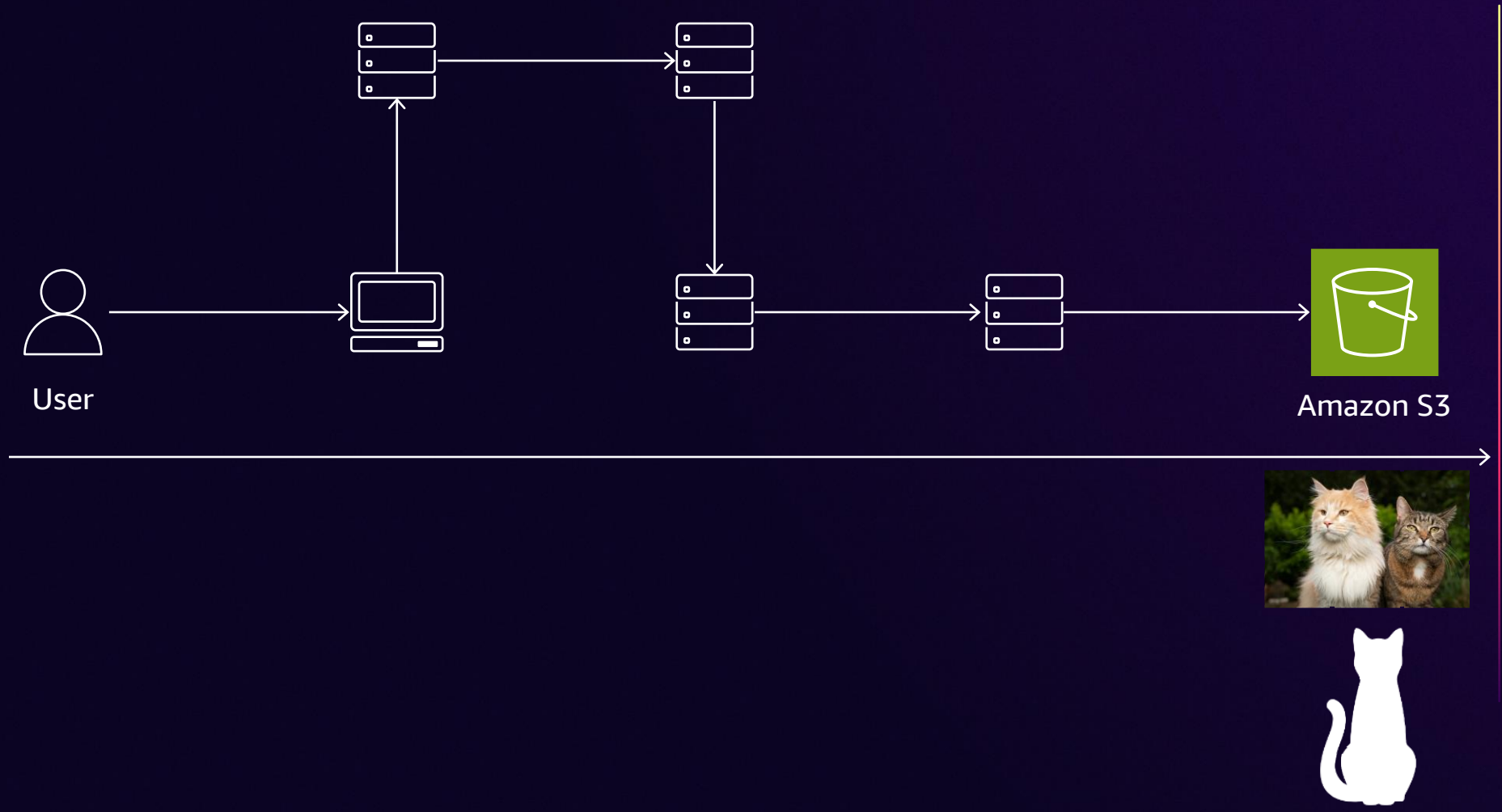
Default checksums: Full object for MPUs



Each part is uploaded with its own checksum



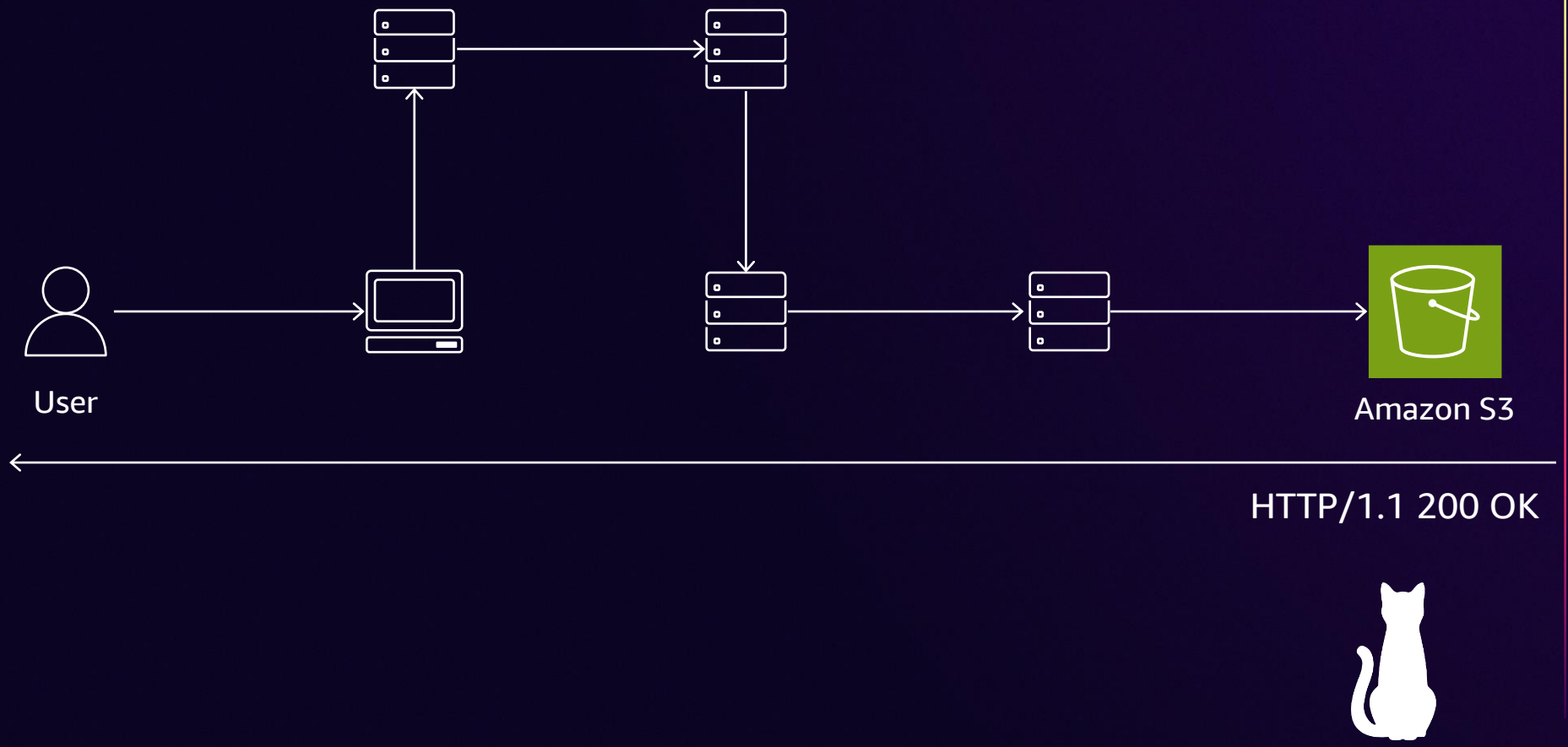
Default checksums: Full object for MPUs



Each part is uploaded with its own checksum

The checksum is combined server-side to build a checksum of the whole object

Default checksums: Full object for MPUs



Full object checksum for objects uploaded in parts, returned to you

Demo: Amazon S3 default checksums

Objects uploaded with no checksum

3 supported full-object algorithms

Consistency across encryption types

Updated SDKs

Recap of integrity checking on Amazon S3

01 Checksums and their use in validating integrity

02 Content-MD5 and the Amazon S3 ETag

03 Amazon S3 checksum enhancements to meet compliance needs

04 Amazon S3 default checksums to simplify your workloads

Thank you!

Please complete the session survey in the mobile app

Raghu Balivada
balivada@amazon.com

Akshat Sandh
akssandh@amazon.com