Data archiving and digital preservation solutions with AWS

Paul Meighan

Director, Product Management, Amazon S3
Amazon Web Services

Most of the world’s data is cold.

Why archive to AWS?

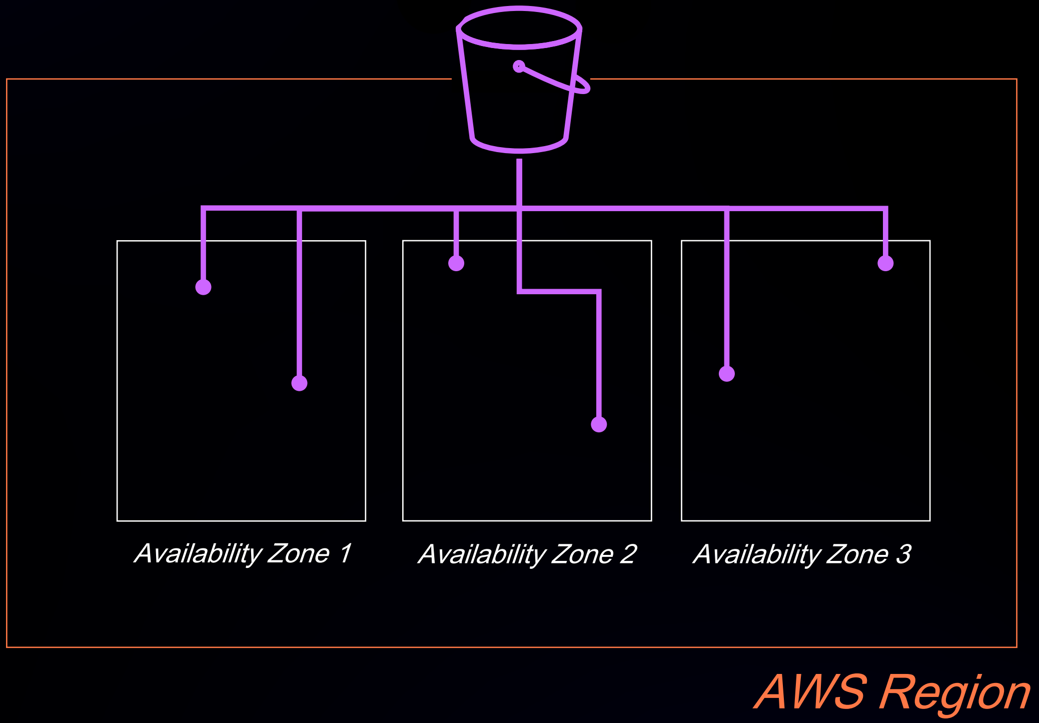
Durability & Resilience

Security & Compliance

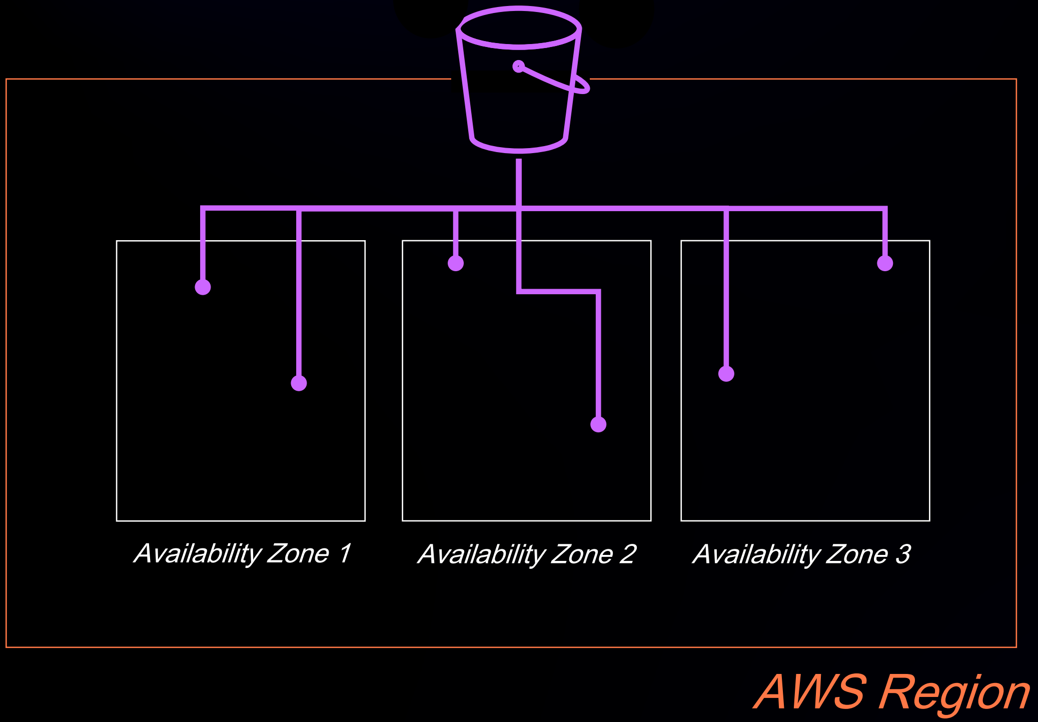
Lowest Cost

Designed to provide 99.999999999% of data durability

The unique architecture of Amazon S3

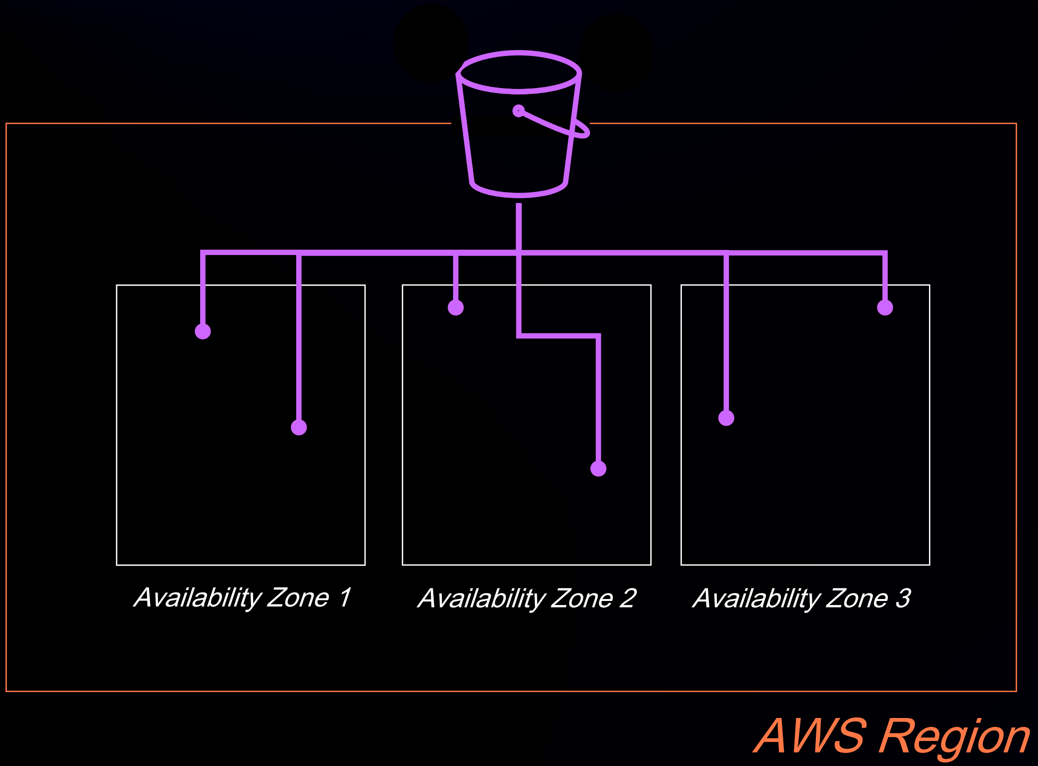


The unique architecture of Amazon S3



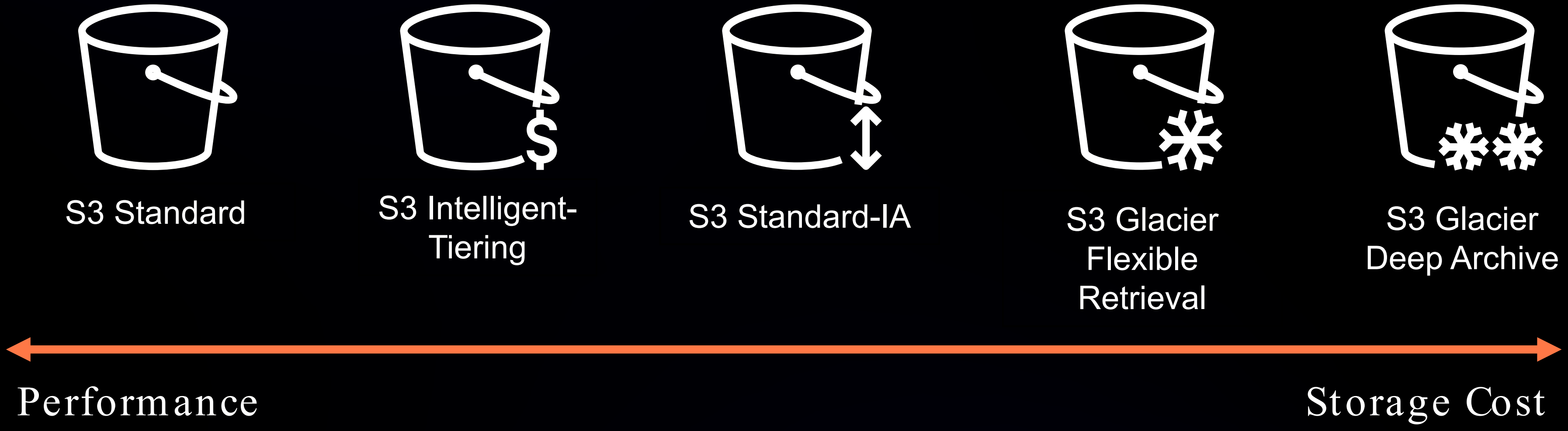
- Stored redundantly across a minimum of 3 Availability Zones
- Stored redundantly across multiple devices within an Availability Zone
- Designed to sustain concurrent device failures

A culture of durability

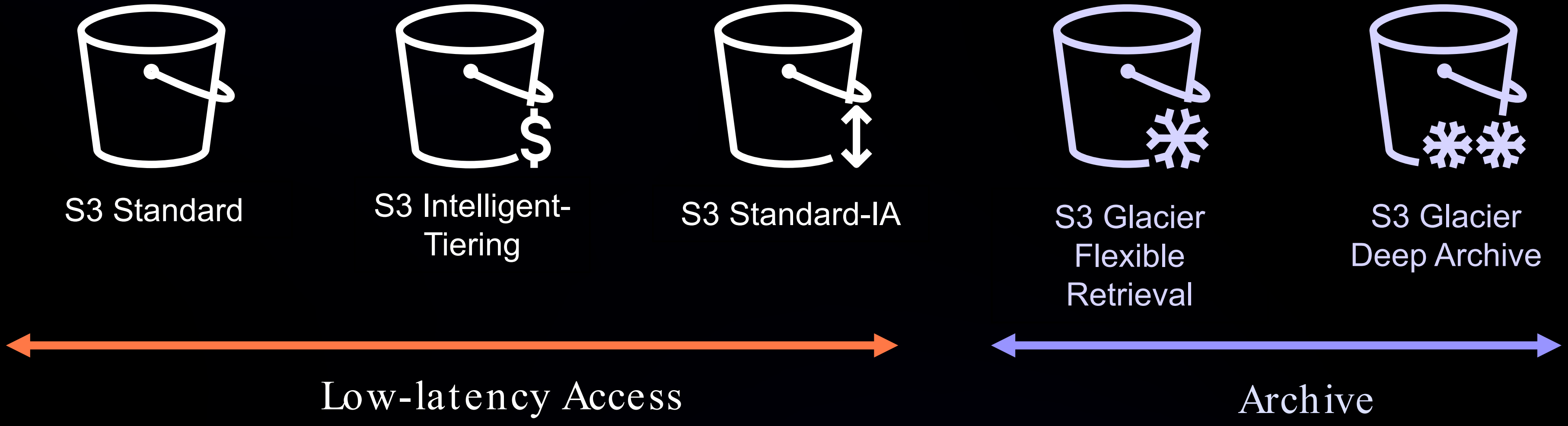


- Durability review & operational safeguards
- Integrity checking to the point of paranoia
- Auditors check and re-check data at rest

Amazon S3 Storage Classes



Amazon S3 Storage Classes



S3 Standard

S3 Intelligent-Tiering

S3 Standard-IA

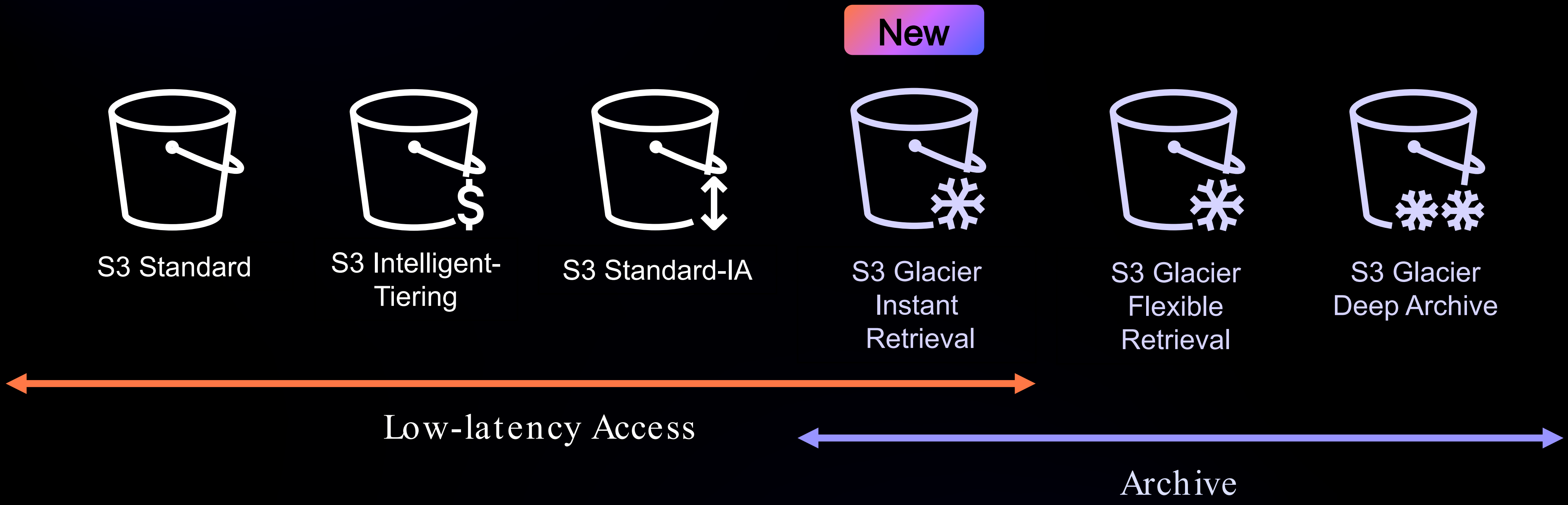
S3 Glacier Flexible Retrieval

S3 Glacier Deep Archive

Low-latency Access

Archive

Amazon S3 Storage Classes



New

S3 Standard

S3 Intelligent-Tiering

S3 Standard-IA

S3 Glacier Instant Retrieval

S3 Glacier Flexible Retrieval

S3 Glacier Deep Archive

Low-latency Access

Archive

Introducing S3 Glacier Instant Retrieval

NEW

What is it?

- For long-lived archive data that requires milliseconds retrieval
- 99.999999999% (11 9s) of durability
- Designed for 99.9% availability

What are the use cases?

- Petabytes of archive data stored for indefinite periods of time
- Only a small percentage of this archive data is accessed each year
- Archive data must be immediately accessible when requested

Amazon S3 Glacier Flexible Retrieval

Bulk Retrievals Are Now FREE !

Price Drop !

What is it?

- For long-lived archive data and long-term backup
- 99.999999999% (11 9s) of durability
- Retrievals in 3-5 hours for standard
- Free Bulk retrievals in 5-12 hours

What are the use cases?

- Petabytes of archive data stored for indefinite periods of time
- Data accessed 1-2 times per year

Amazon S3 Glacier Deep Archive

What is it?

- Archiving long-term data that which accessed infrequently
- 99.999999999% (11 9s) of durability
- Retrievals within 12 to 48 hours.

What are the use cases?

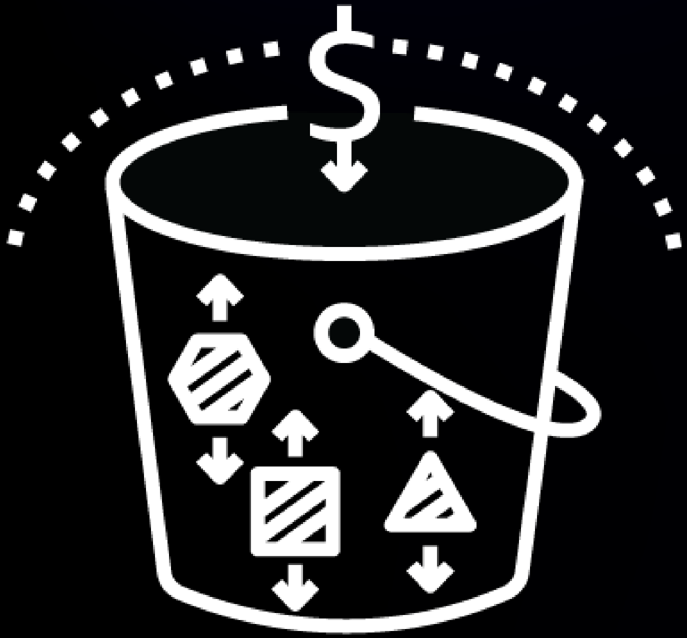
- Archive data backups that are rarely accessed
- Data that needs to be retained for the long term

Choosing between S3 Glacier archive storage

	S3 Glacier Instant Retrieval	S3 Glacier Flexible Retrieval	S3 Glacier Deep Archive
Storage cost	$0.004 per GB-month	$0.0036 per GB-month	$0.00099 per GB-month
Data retrieval	Instant	Expedited: 1-5 minutes Standard: 3-5 hours Bulk: 5-12 hours	Standard: Within 12 hours Bulk: Within 48 hours
Minimum object duration	90 Days	90 days	180 days

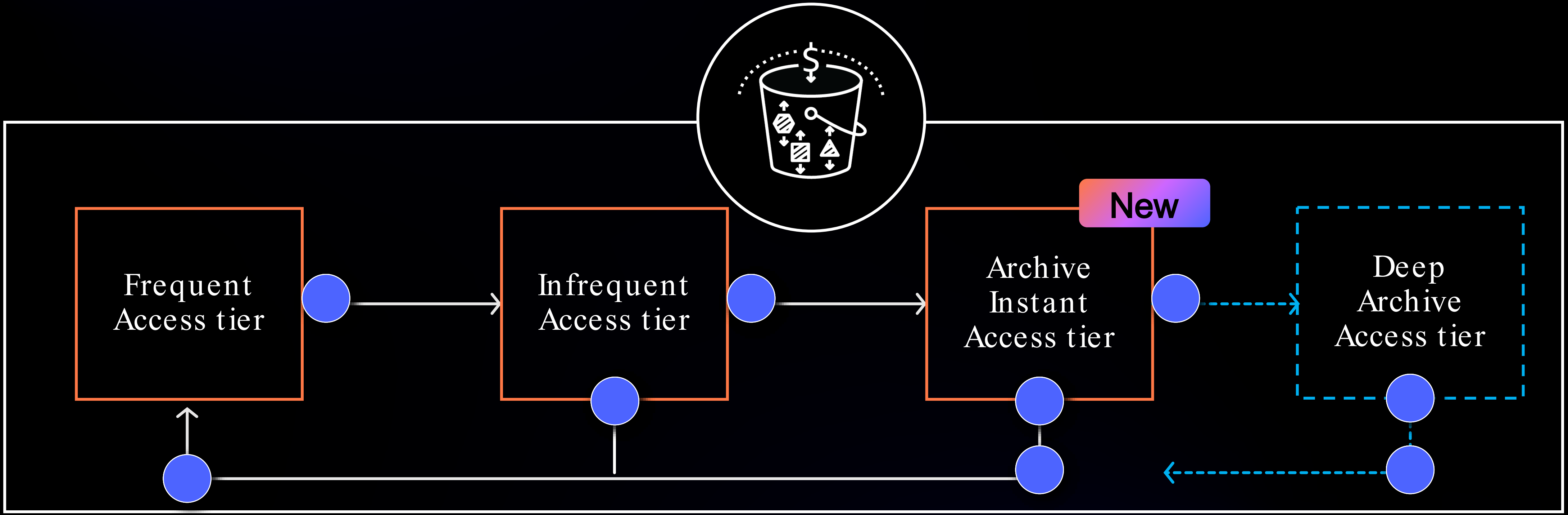
Bulk Retrievals Are Now FREE!

Amazon S3 Intelligent-Tiering



- Automatically moves objects between three access tiers
- Optional asynchronous archiving to realize lowest storage cost in the cloud
- No performance impact, operational overhead, lifecycle fees, or retrieval fees
- Designed for 99.9% availability and 99.999999999% durability

Use S3 Intelligent-Tiering by default for data with unknown or changing access patterns



Milliseconds access (automatic)

Minutes to hours (Optional)

New

Accelerate integrity checks by up to 92%

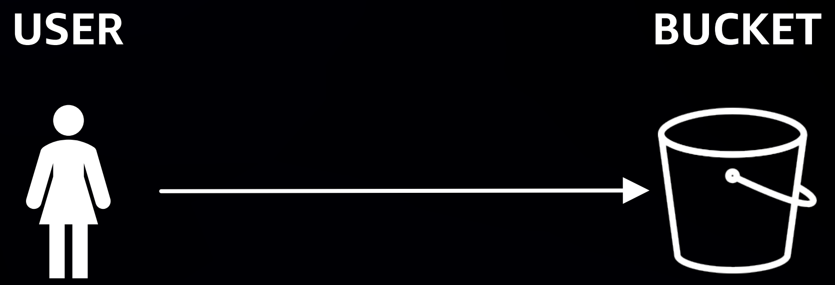
1. Trailing checksums allow you to check data while you stream it in.

2. Parallel checksums allow you to break large objects up, and leverage many cores.

aws

Checksums in S3 today

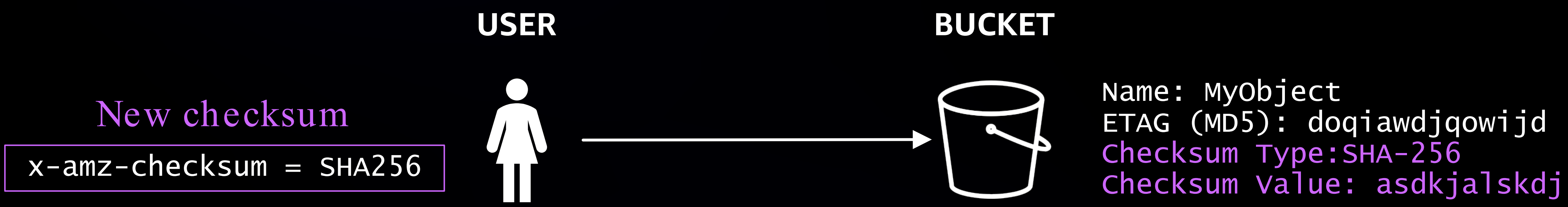
New



```
Name: MyObject
ETAG (MD5): doqiawdjqowijd
```

New

New checksum options



New checksum

```
x-amz-checksum = SHA256
```

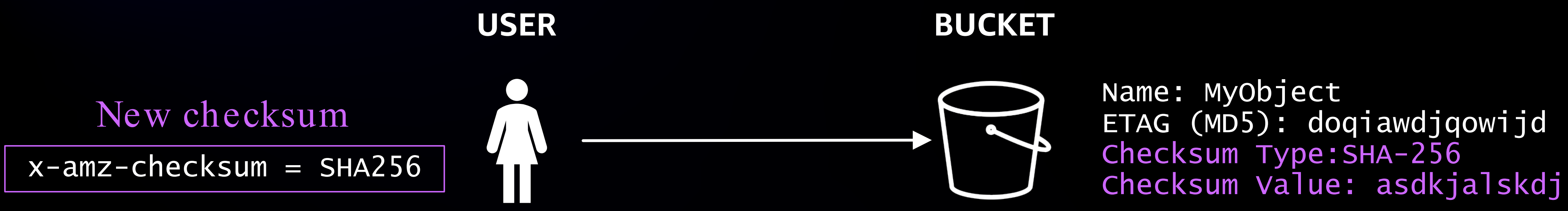
USER

BUCKET

```
Name: MyObject
ETAG (MD5): doqiawdjqowijd
Checksum Type:SHA-256
Checksum Value: asdkjalskdj
```

New checksum options

New



New checksum

```
x-amz-checksum = SHA256
```

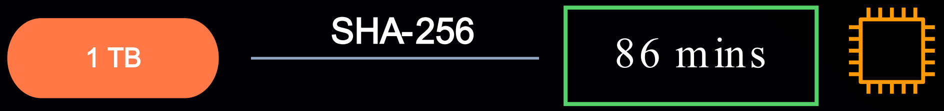
```
Name: MyObject
ETAG (MD5): doqiawdjqowijd
Checksum Type:SHA-256
Checksum Value: asdkjalskdj
```

SHA-256 | SHA-1 | CRC32 | CRC32C

New

Checksums in S3 today

Calculating Full Object Checksum



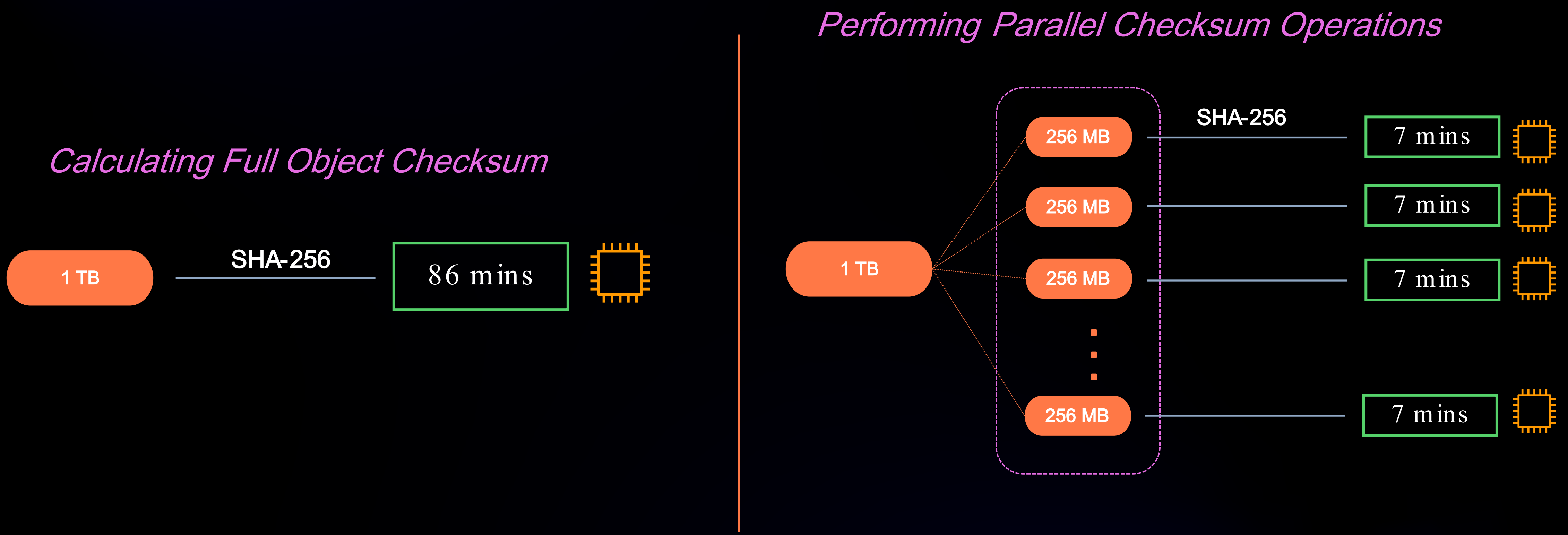
1 TB

SHA-256

86 mins

Parallelized Checksums

New



The GetObjectAttributes API

New

A new S3 API that gives you:

- Checksum Algorithm
- Checksum Value
- Number of Parts
- Part Boundaries
- Part-level Checksum Values

Thank you!