

The background features a dark blue gradient with large, overlapping, semi-transparent shapes in shades of purple and magenta. Thin, light blue lines crisscross the scene, creating a sense of depth and movement.

AWS re:Invent

DECEMBER 2 - 6, 2024 | LAS VEGAS, NV

STG212

What's new with Amazon S3

Paul Meighan

Director, Product Management
Amazon S3 & Glacier
AWS

Mallory Gershenfeld

Senior Product Manager
Amazon S3
AWS



*30 launches in
60 minutes*

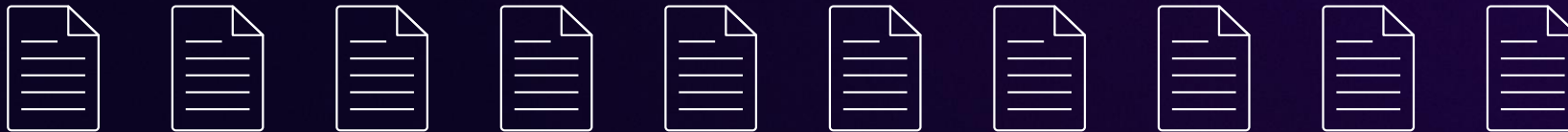
Structured Data
Unstructured Data
Storage Fundamentals
S3 Clients



Structured data in Amazon S3

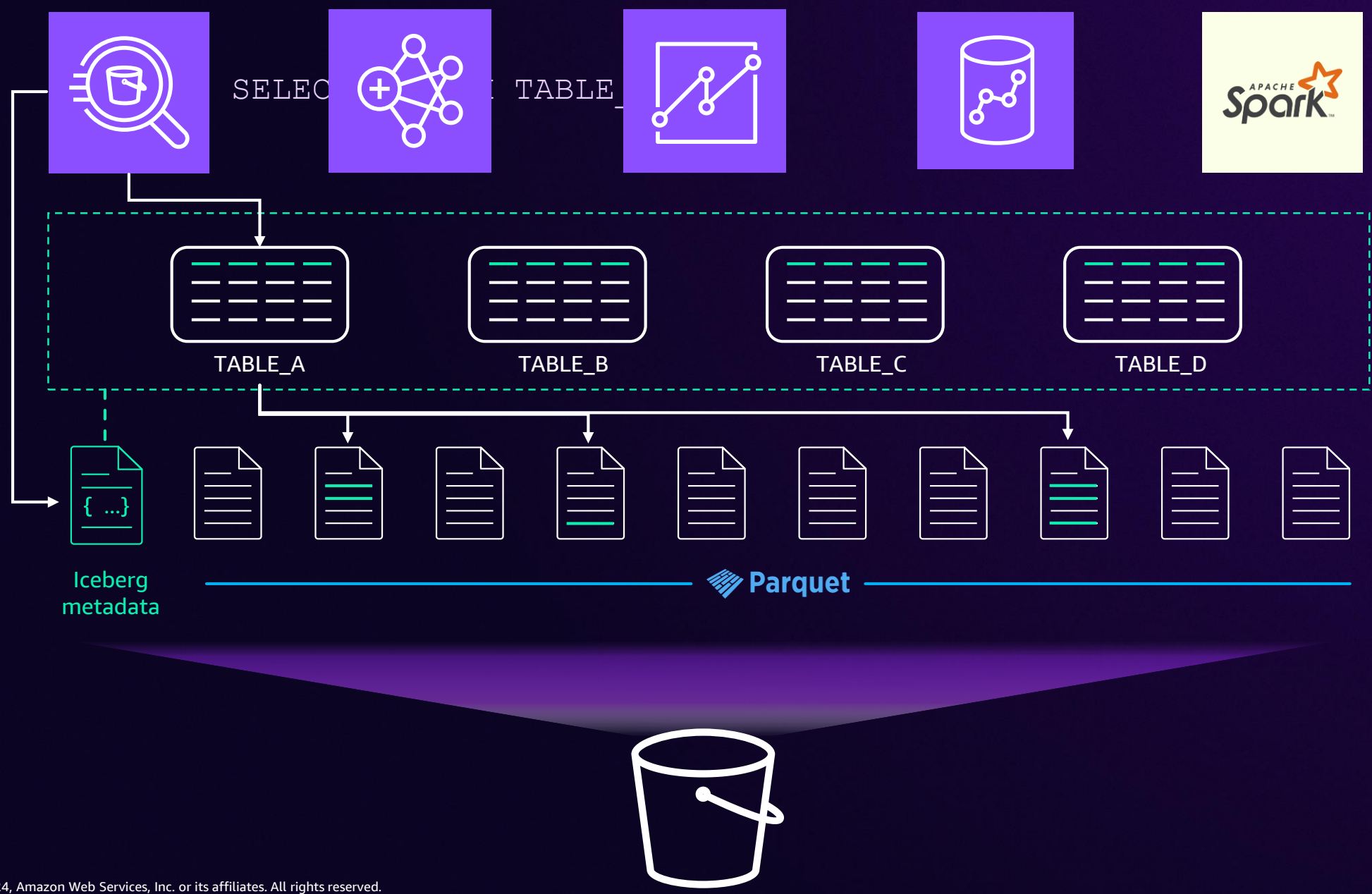
Today, Amazon S3 is also a *tabular* data store

Exabytes of Parquet data stored
15 million+ requests per second
Serve hundreds of petabytes every day

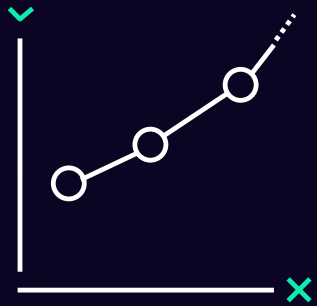


 Parquet





Customer problems to solve



Growing scale requires
more and more
performance



Enforcing table-level
security and integrity is
complex



Optimizing storage cost
drives unexpected
operational burden

GA Dec. 3, 2024

Amazon S3 Tables

Fully managed Apache Iceberg tables
in Amazon S3

Introducing S3 Tables

NEW APIS



Table
Management

Metadata
Updates

```
S3tables:CreateTable  
S3tables:ListTables  
S3tables:RenameTable  
S3tables:GetTable  
S3tables:DropTable
```

```
S3tables:GetTableMetadataLocation  
S3tables:UpdateTableMetadataLocation
```

Introducing S3 Tables

NEW APIS



The table is a first class **AWS resource!**

- Has an ARN
- Has a dedicated endpoint
- Can take an S3 resource policy

Introducing S3 Tables

NEW APIS



S3 applies Iceberg-specific optimization behind the scenes

Applications read and write data via the Apache Iceberg standard



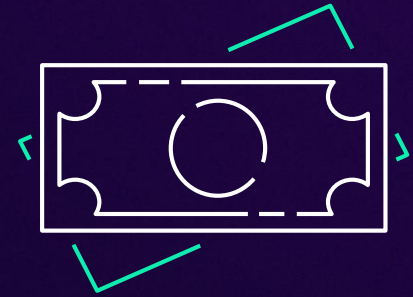
Benefits



Improved
Performance



Simplified
Security



Automated Cost
Optimization

S3 Table Performance

10x

Transactions per Second
(TPS)

3x

Improvement to
Query Performance

S3 Table Performance

10X TPS BASED ON OPTIMIZED KEY NAMING & INDEX TUNING

10x

Transactions per Second
(TPS)

Optimized **key naming** & layout

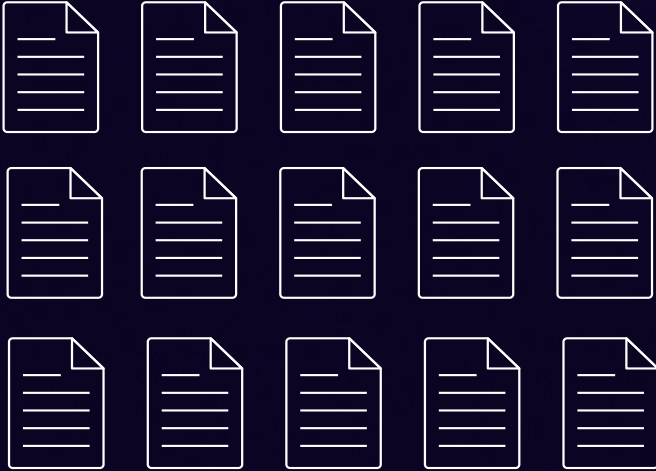
S3 index subsystem **tuned specifically for Iceberg** workloads

Enables a **higher starting point** for S3 TPS scaling



S3 Table Performance

3X QUERY PERFORMANCE BASED ON AUTOMATIC DATA COMPACTION



3x

Improvement to
Query Performance

S3 Table Performance

3X QUERY PERFORMANCE BASED ON AUTOMATIC DATA COMPACTION

Automatic **Compaction**



Fewer requests to S3
Better tail latencies
Better throughput

3x

Improvement to
Query Performance



S3 Table Performance

10x

Transactions per Second
(TPS)

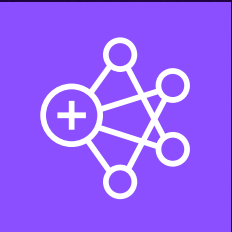
3x

Improvement to
Query Performance

Simplifying Security with S3 Tables



```
{  
  "Effect": "Allow",  
  "Action": [  
    "s3tables:GetTableData",  
    "s3tables:PutTableData"  
  ]  
}
```



Introducing S3 Table Buckets

SECURITY



```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowTablebucketActions",
      "Effect": "Allow",
      "Principal": {
        "AWS": "arn:aws:iam::acct_id:role/datasteward"
      },
      "Action": [
        "s3tables:ListTables",
        "s3tables:GetTableMetadataLocation",
        "s3tables:GetTableData",
        "s3tables:GetTableMaintenanceConfig",
        "s3tables:GetTableBucketMaintenanceConfig"
      ],
      "Resource": "arn:aws:s3tables:region:acct_id:bucket/table-bucket-1/table/*"
    }
  ]
}
```

Introducing S3 Table Buckets

STORAGE COST OPTIMIZATION



Introducing S3 Table Buckets

STORAGE COST OPTIMIZATION



Nightly Maintenance Runs

- Snapshot Expiration
- Garbage Collection

Introducing S3 Table Buckets

STORAGE COST OPTIMIZATION



100% Policy-driven Table Maintenance:

```
aws s3tables put-table-maintenance-configuration \  
--table-bucket-arn="arn:aws:s3tables:us-east-2:accountID:bucket/my- test-table-bucket" \  
--type="icebergSnapshotManagement" \  
--namespace="my_namespace" \  
--name="my_table_name" \  
--value="{ 'status': 'enabled', \ settings: { 'minimumSnapshots': '10', 'maximumSnapshotAge': '2500' } \  
}"
```



High performance out of the box

Easy to secure

Automatic storage cost optimization



Unstructured data in Amazon S3

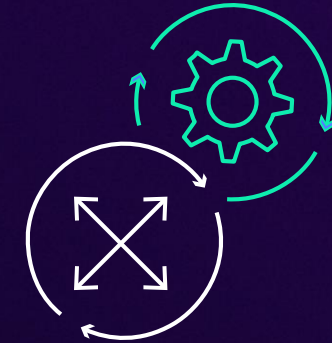
Customer problems to solve



Finding the right data to complete a task takes longer and longer



Metadata stores are difficult to build, live outside of storage



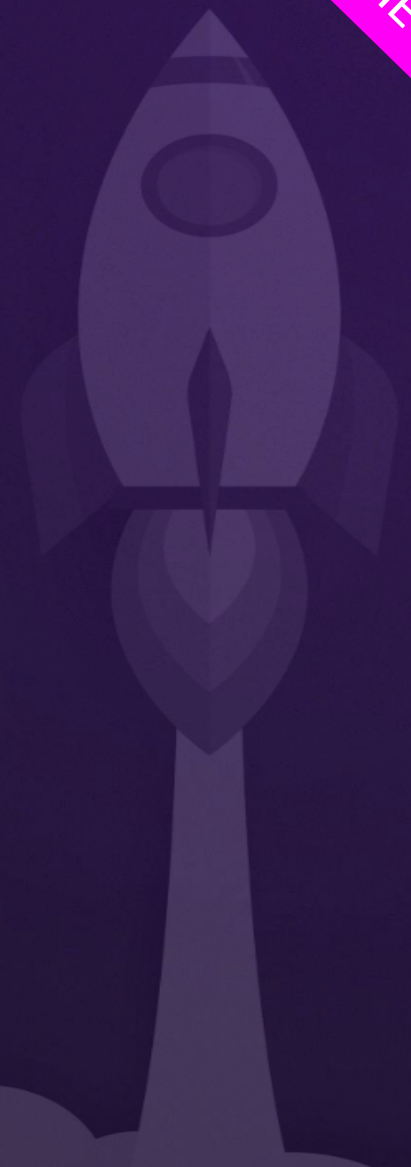
Metadata is only as useful as it is current

PREVIEW

PREVIEW Dec. 3, 2024

Amazon S3 Metadata

Automatic metadata generation,
accessible with simple SQL semantics

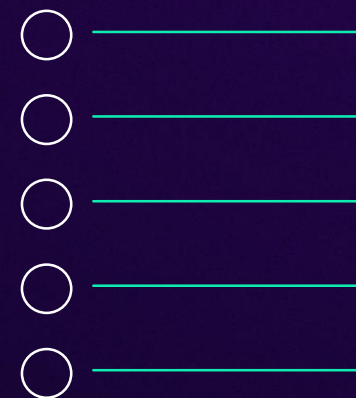
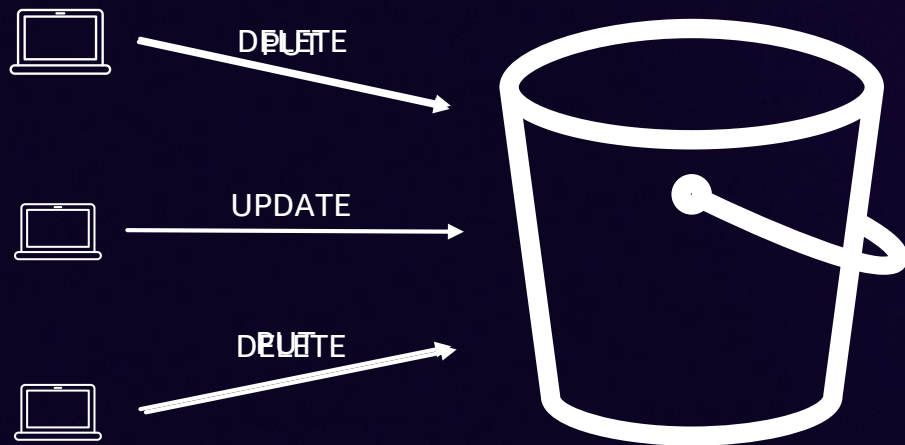


Clients

Data

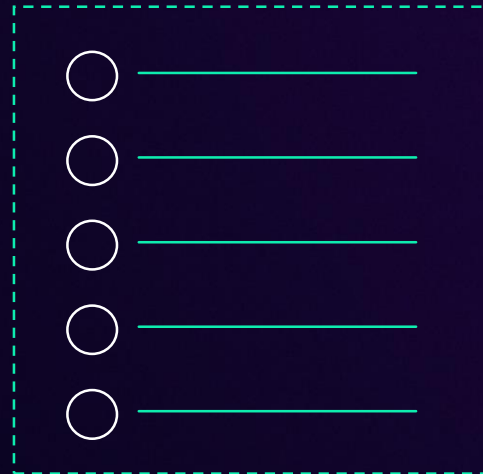
Metadata

Journal table





Journal table



AWS-managed table

Read-only for non-AWS principals

It's a "system table" for your bucket

| Metadata | Data Type |
|-----------------------|----------------------|
| bucket | String |
| key | String |
| sequence_number | String |
| record_type | String |
| record_timestamp | Timestamp |
| version_id | String |
| is_delete_marker | Boolean |
| size | Long |
| last_modified_date | Timestamp |
| e_tag | String |
| storage_class | String |
| is_multipart | Boolean |
| encryption_status | String |
| is_bucket_key_enabled | Boolean |
| kms_key_arn | String |
| checksum_algorithm | String |
| object_tags | Map <String, String> |
| user_metadata | Map <String, String> |
| requester | String |
| source_ip_address | String |
| request_id | String |

21 system metadata fields automatically recorded for all new objects in your bucket

Object size

Storage class

Create date

Client information

Encryption information

...

Use cases

Find objects to process

```
// List the objects in my bucket (without using the LIST API)
SELECT bucket, key, storage_class, user_metadata
FROM aws_s3_metadata.my_metadata_table
WHERE key LIKE '{SOME_PREFIX}%';
```

Understand and track data lineage

```
// Who is deleting data? What are they deleting?
SELECT DISTINCT bucket, key, record_type, record_timestamp, requester, source_ip_address
FROM aws_s3_metadata.my_metadata_table
WHERE record_type = 'DELETE';
```

Learn about storage usage

```
// How much storage are my tagged objects using?
SELECT object_tags['MyTag'] as mytag, storage_class, count(*) as count,
sum(size) / 1024 / 1024 as usage
FROM aws_s3_metadata.my_metadata_table
GROUP BY tags['Module'], storage_class;
```

Custom metadata

Object tags



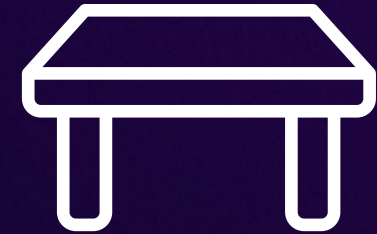
Mutable
Dedicated APIs
Dedicated permissions
Paid Amazon S3 feature

User-defined



Immutable
Only added on PUT
No cost

Joined tables



Mutable
Fully customized
Detached from objects
Cost of table storage

Video generated via Amazon Bedrock is automatically annotated with custom metadata

Amazon
Bedrock



In Amazon Bedrock's PUT request:

```
x-amz-meta-content-source: AmazonBedrock
x-amz-meta-content-model-id: arn:aws::model/xyz-v1
```

Journal
table



Query from Amazon S3 Metadata:

```
SELECT bucket, key, user_defined_metadata, row_type
FROM aws_s3_metadata.my_metadata_table
WHERE user_defined_metadata['content-source'] = AmazonBedrock
```


Storage Fundamentals

S3 Scale



When limited by bucket counts...



```
/customer-1/application-blue  
/customer-1/application-green  
/customer-2/application-red  
/customer-3/application-yellow
```

AWS account 987654321



```
/customer-4/application-purple  
/customer-5/application-pink  
/customer-6/application-grey  
/customer-7/application-red2
```

AWS account 123456789

Customer problems to solve



Limits are yet another operational concern



Prefix-level policies add complexity



Cross-account permissions add complexity

GA Nov. 14, 2024

Higher S3 Bucket Limits

Amazon S3 now supports up to
1 million buckets per AWS account



Service Quota Support

The screenshot shows the AWS Service Quota console for 'General purpose buckets'. It includes a 'Details' section with a description, quota code (L-DC2B2D3D), and quota ARN. Below this is a table with columns for Utilization (1.01), Applied account-level quota value (10,000), AWS default quota value (10,000), and Adjustability (Account level). There are tabs for Monitoring, Request history, Alarms, and Tags. The Monitoring tab is active, showing a line graph titled 'General purpose buckets - (Utilization %)' with a y-axis from 0 to 1.01 and an x-axis from 06/01 to 11/01. The graph shows a step increase in utilization starting on 10/01.

Launched: Oct. 1, 2024

ListBucket API Updates

Paginate your requests:

```
GET / max-buckets=1000  
host:s3.amazonaws.com
```

Filter by AWS Region:

```
GET / bucket-region=us-east-2  
max-buckets=1000  
host:s3.amazonaws.com
```

Filter by prefix:

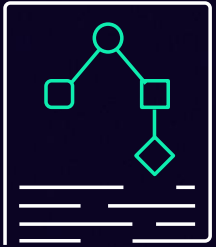
```
GET / prefix=reinvent-bucket  
max-buckets=1000  
host:s3.amazonaws.com
```

Launched: Aug. 15, 2024 & Oct. 16, 2024

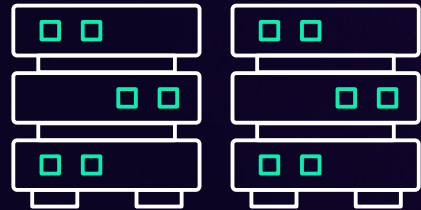
Storage Fundamentals

S3 Durability

99.9999999999% Data Durability



End-to-end integrity checking of requests



Data always stored on redundant devices



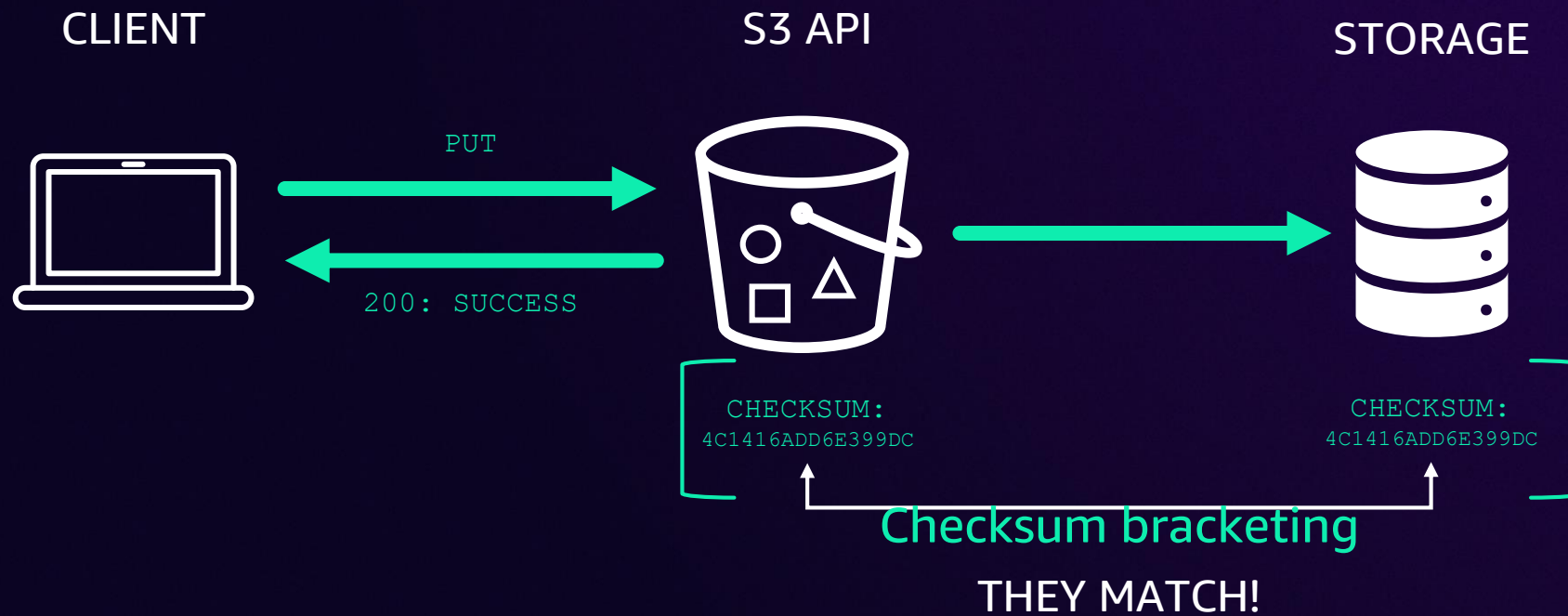
Periodic durability auditing for data at rest



Data stored across multiple Availability Zones



Durability in flight



Durability in flight



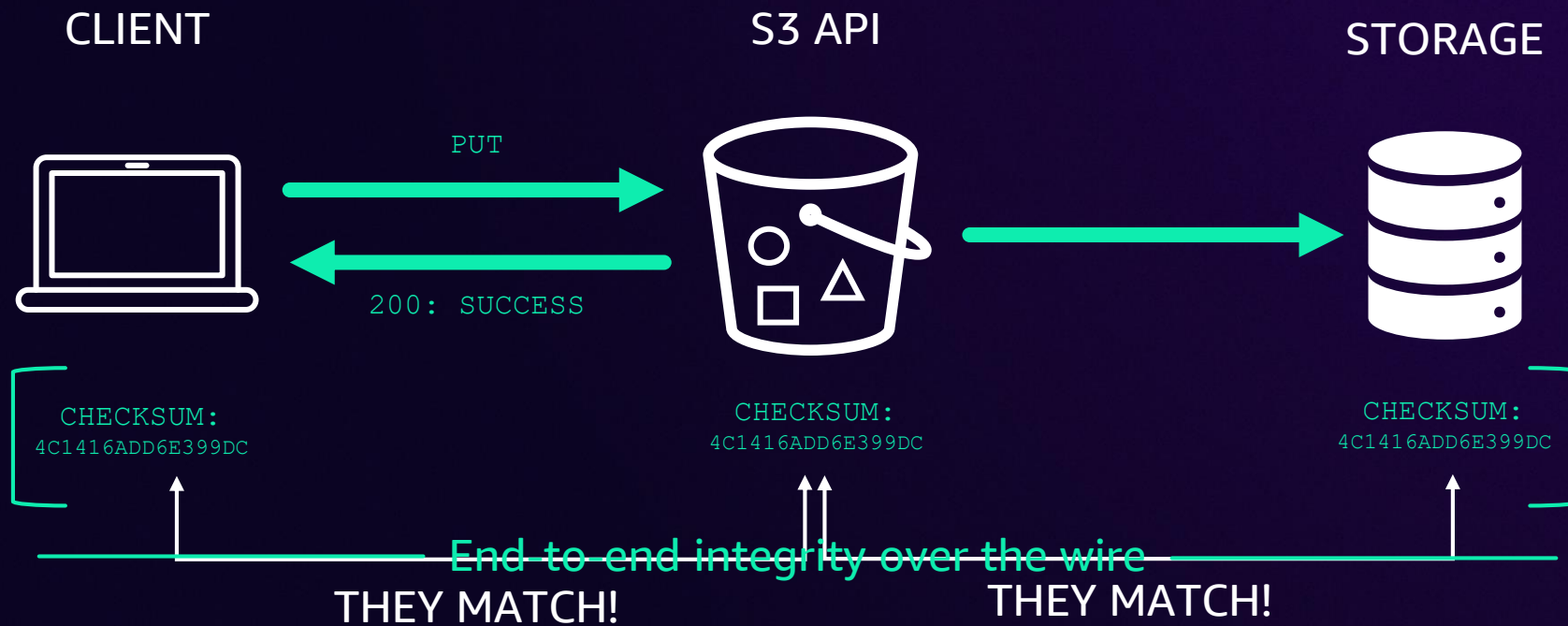
GA Dec. 1, 2024

New data integrity defaults

Data validation over the wire,
new checksum info in metadata



Durability in flight



Whole object checksums in metadata



```
{  
  "AcceptRanges": "bytes",  
  "LastModified": "2024-10-24T20:41:17+00:00",  
  "ContentLength": 531132416,  
  "ETag": "\"61911f803a7c5e2533c6b6d6703996bc\"",  
  "ContentType": "binary/octet-stream",  
  "ServerSideEncryption": "AES256",  
  "Metadata": {},  
  "checksum-crc64nvme": "OtNfLhPRTuQ=",  
  "checksum-type": "FULL_OBJECT"  
}
```

On by default for all PUT requests

CLIENT



S3 API



STORAGE





Data Durability for Distributed Applications

Two clients writing to the same key



Start Multi-part Upload

Upload Parts

Complete Upload

200: SUCCESS



PutObject

200: SUCCESS

Successful request, but no data stored (ultimately)



Time

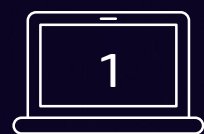
GA 2024

Conditional S3 Requests

Put-if-absent and Put-if-match
conditions simplify distributed apps



Two clients writing to the same key, with **if-no-match** condition



Start Multi-part
Upload

Upload Parts

Complete Upload
(if-no-match)

412:
Precondition
Failed



PutObject

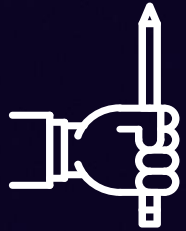
200: SUCCESS

Data remains in S3

Time

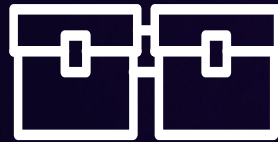
4 new conditional request primitives

If-None-Match



Never overwrite an existing object

If-Match
(PUT)



Only overwrite an object with a given Etag

If-Match
(DELETE)



Only delete an object with a given Etag

Condition
Keys



Only allow conditional requests

Storage Fundamentals

S3 Security

Troubleshooting 403: Access Denied Errors

GA Aug. 21, 2024

New Context in 403 Responses

More information for requests made
within an AWS Account



Includes **type** of policy, **reason** for denial, and **AWS IAM user or role** that requested access

Resource Policy Denial

User: `arn:aws:iam::123456789012:user/MaryMajor` is not authorized to perform: `s3:GetObject` on resource: `"arn:aws:s3:::amzn-s3-demo-bucket1/object-name"` with an **explicit deny** in a **resource-based policy**

Block Public Access Denial

User: `arn:aws:iam::123456789012:user/MaryMajor` is not authorized to perform: `s3:GetObject` on resource: `"arn:aws:s3:::amzn-s3-demo-bucket1/object-name"` because **public access control lists (ACLs) are blocked** by the **BlockPublicAcls block public access setting**.

Service Control Policy Denial

User: `arn:aws:iam::777788889999:user/MaryMajor` is not authorized to perform: `s3:GetObject` with an **explicit deny** in a **service control policy**



GA re:Invent 2023

Amazon S3 Access Grants

Simplified permissions for data lakes,
driven by identity directories



S3 Access Grants integrations

Security



AWS IAM
Identity Center



AWS
STS



AWS
IAM



AWS
KMS

Machine Learning



Amazon
SageMaker

Analytics



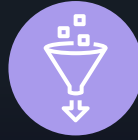
Amazon EMR on
EC2, EKS, and
Serverless



Amazon
Athena



Amazon
Redshift



AWS Glue
ETL

Developer and Open-Source Tools



AWS SDKs



AWS CLI



Boto3
(Python)

GA Sep. 5, 2024

S3 Access Grants

ListCallerAccessGrants API

Return and display datasets that an end-user has permissions to



Storage Fundamentals

S3 Performance

Performance demands



GA re:Invent 2023

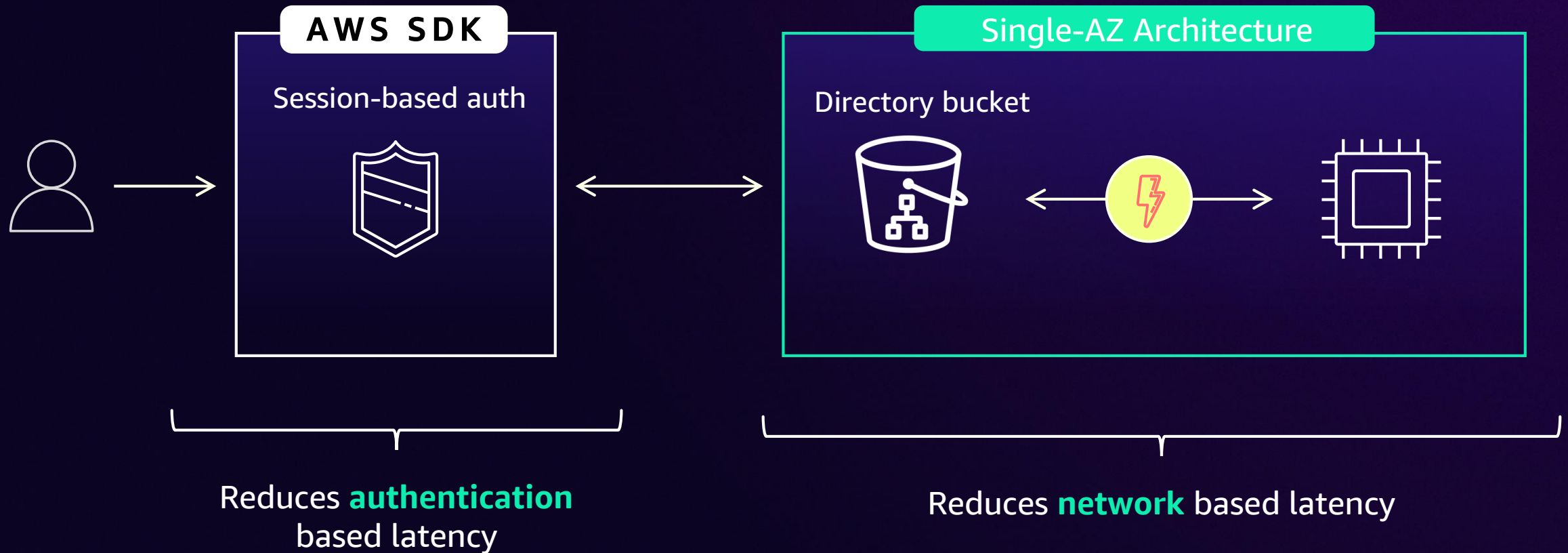
Amazon S3

Express One Zone
Storage Class

For compute-intensive workloads



Designed for performance





S3 Express One Zone integrations

New

Monitoring and Automation



Amazon CloudWatch



AWS CloudTrail



AWS CloudFormation



Amazon VPC



AWS IAM



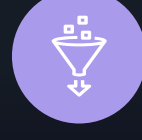
AWS KMS



Amazon EMR on EC2, EKS, and Serverless



Amazon Athena



AWS Glue

Security

Analytics

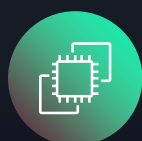
Compute

Machine Learning

Developer and Open-Source Tools



Amazon Elastic Kubernetes



Amazon EC2



AWS Lambda



Amazon SageMaker



S3 Lifecycle Expiration



Mountpoint for Amazon S3



AWS SDKs



S3A Connector

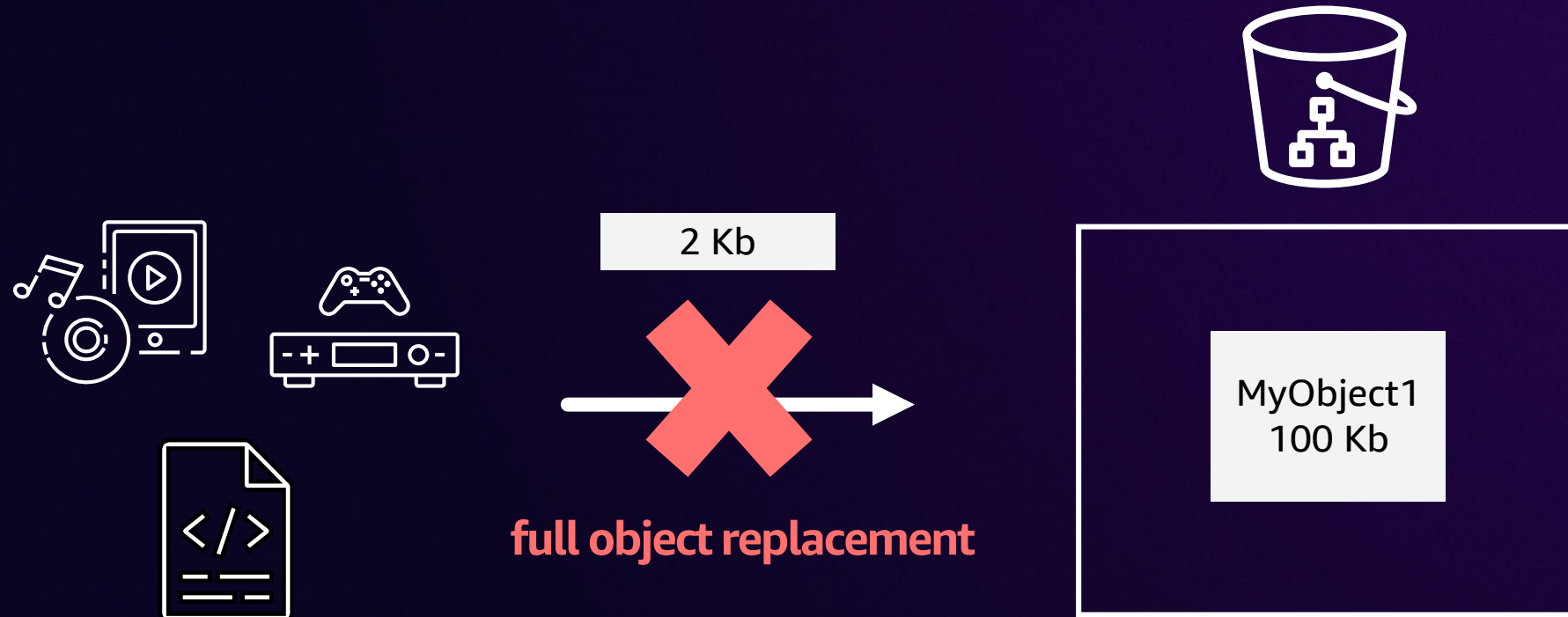


AWS CLI



S3 Connector for PyTorch

Continuous data workloads



add data to the same object

GA Nov. 21, 2024

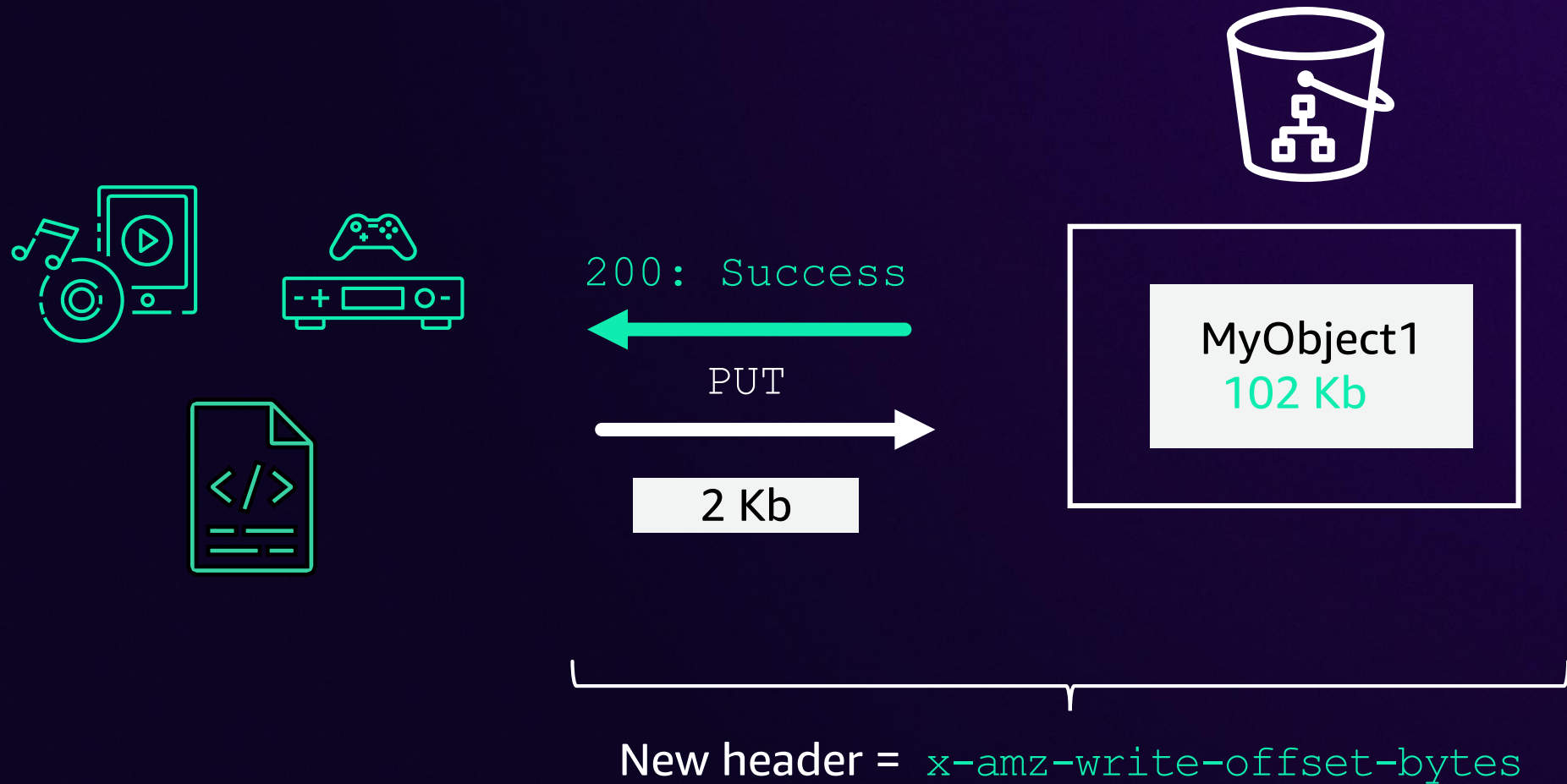
Append data

with S3 Express One Zone

Add data to the end of an existing object



Append data for S3 Express One Zone



Amazon S3 Express One Zone locations

UNITED STATES

Northern Virginia

Oregon

Ohio

EUROPE

Stockholm

Ireland

ASIA PACIFIC

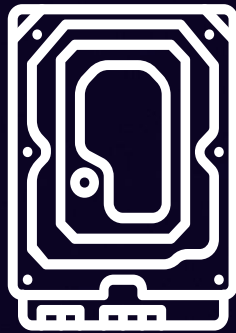
Tokyo

Mumbai

GA Nov 21, 2024

● Available location ● New

Data sovereignty problems to solve



Host applications on
isolated infrastructure



Stringent regulatory
standards

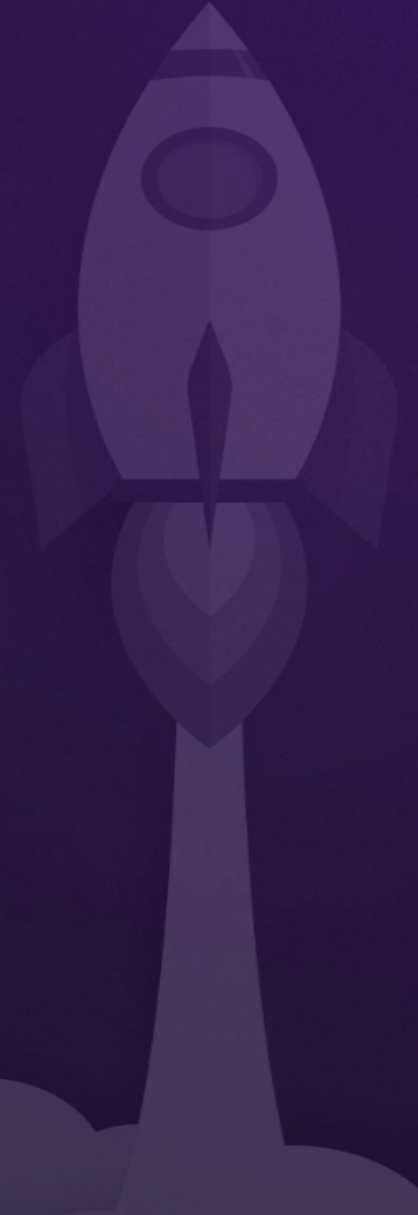


Sovereign location
requirements

GA Dec. 1, 2024

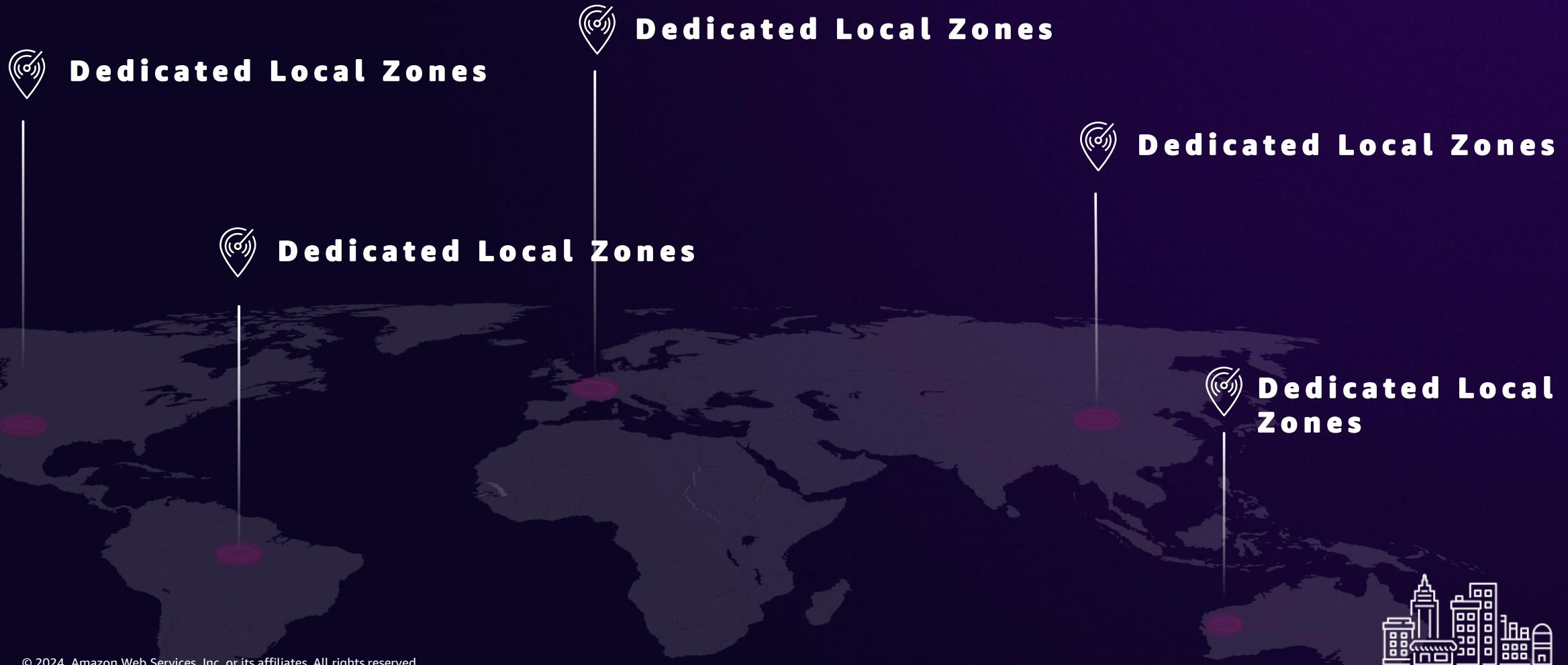
Amazon S3 storage classes for AWS Dedicated Local Zones

Supported in directory buckets

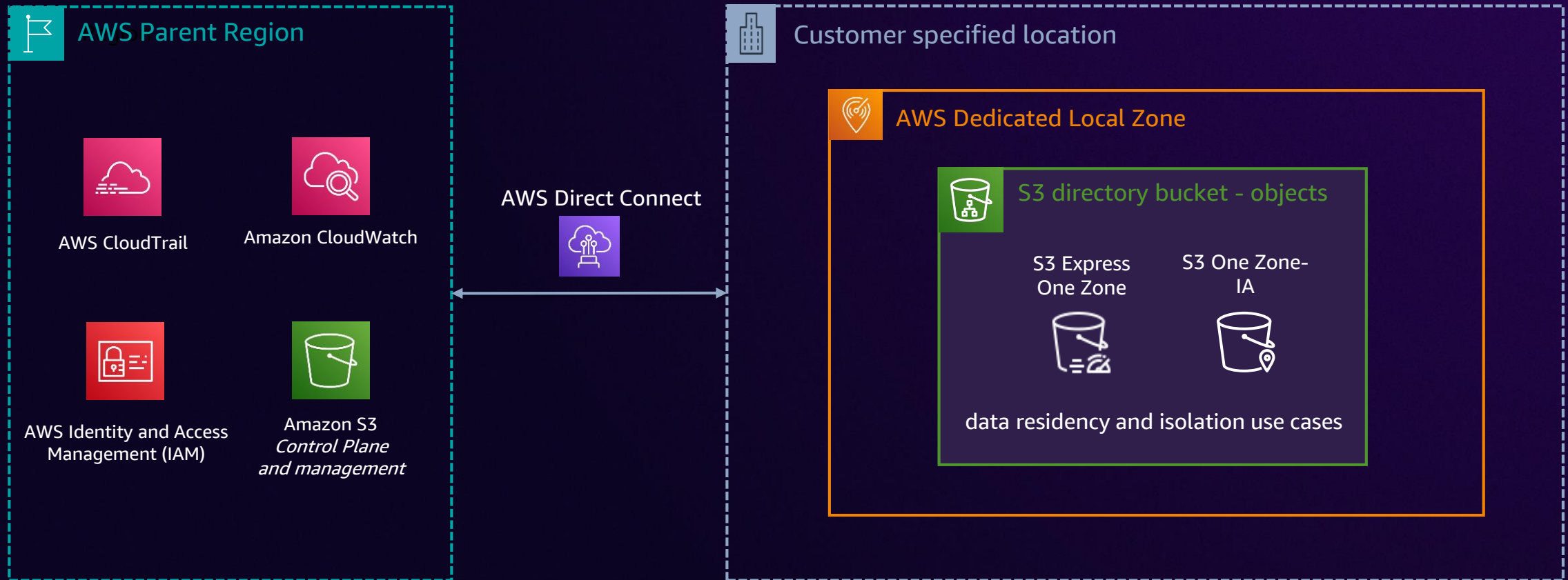


AWS Dedicated Local Zones

AWS infrastructure built exclusively for you



AWS Dedicated Local Zones



Bucket Location =  **AWS Local Zones**



At AWS, we have always believed that for the cloud to realize its full potential it is essential that you have control over your data and choices for how you secure and manage your data in the cloud. We give you the flexibility to choose how and where you want to run your workloads.

AWS

Digital Sovereignty Pledge





Open Source Clients for Amazon S3



Extending the S3 experience



On ramp to S3



The client family for Amazon S3

Clients



Mountpoint for
Amazon S3



Mountpoint for
Amazon S3
for containers



Amazon S3
Connector for
PyTorch

Features



Mountpoint for
Amazon S3 with
local cache



Mountpoint for
Amazon S3
Express One Zone

GA re:Invent 2023

Mountpoint for Amazon S3

Open source file client that translates local file system API calls to S3 object API calls



Mountpoint for append data with S3 Express One Zone



```
mkdir ~/mnt  
mount-s3 amzn-s3-demo-bucket--azid--x-s3 ~/mnt  
--incremental-upload
```

Append data flag

Mountpoint support for AWS Dedicated Local Zones



```
mkdir ~/mnt  
mount-s3 amzn-s3-demo-bucket--lzid--x-s3 ~/mnt
```

ZoneID

S3 directory bucket name



GA Nov. 21, 2024

Mountpoint high performance shared cache

With Amazon S3 Express One Zone

How it works



MOUNTPOINT HIGH PERFORMANCE SHARED CACHE

Source bucket

```
mkdir ~/mnt  
mount-s3 amzn-s3-demo-bucket ~/mnt  
--cache-xz amzn-s3-demo-bucket--usw2-az1--x-s3
```

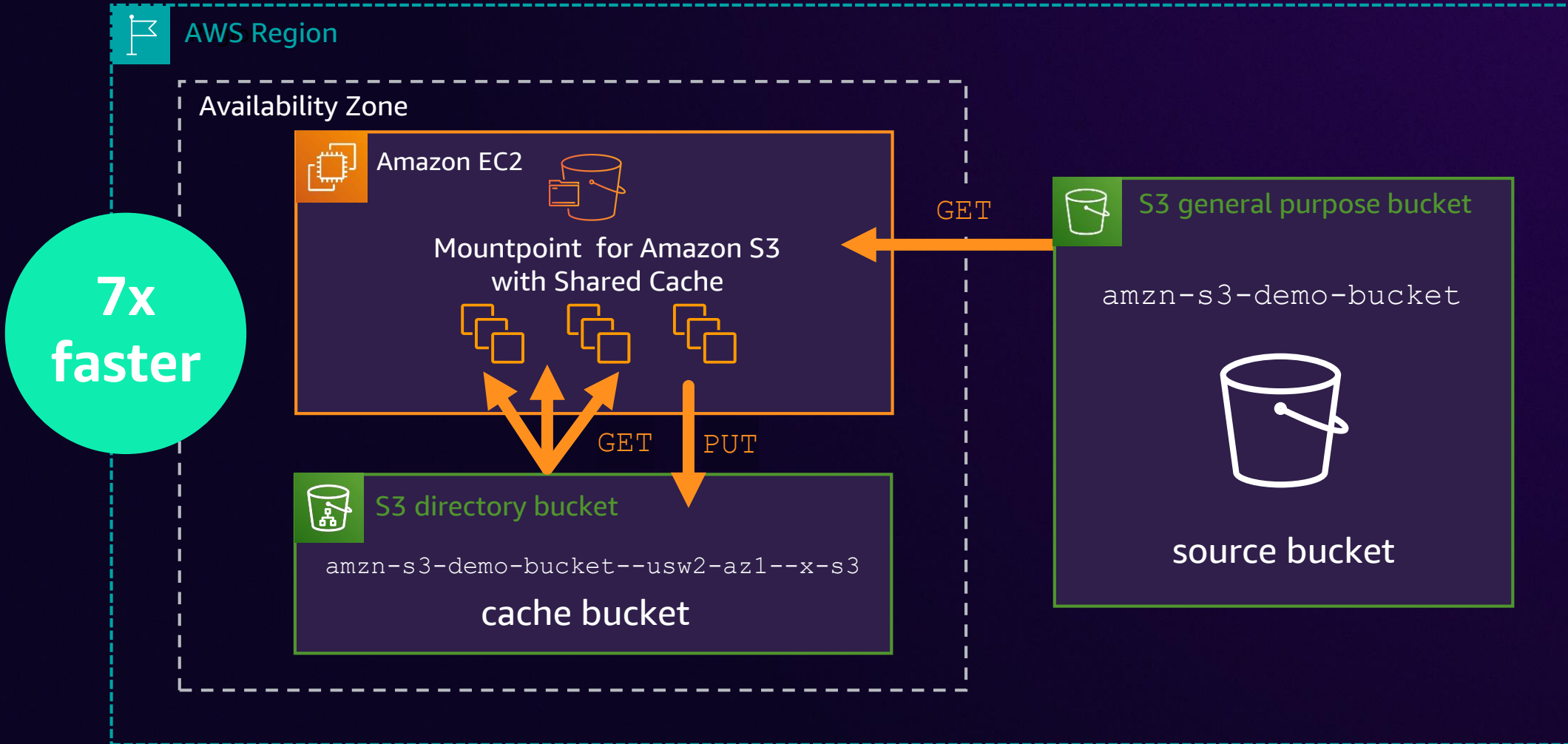
Cache type

S3 directory bucket cache

How it works



MOUNTPOINT HIGH PERFORMANCE SHARED CACHE



GA Nov. 22, 2023

Amazon S3 Connector for PyTorch

High throughput for PyTorch training jobs
that access or store data in Amazon S3



GA Mar. 13, 2024

Amazon S3 Connector for PyTorch checkpoints with PyTorch Lightning

Improving time to write checkpoints by
up to 40%



GA Nov. 22, 2024

Amazon S3 Connector for PyTorch Distributed Checkpoint

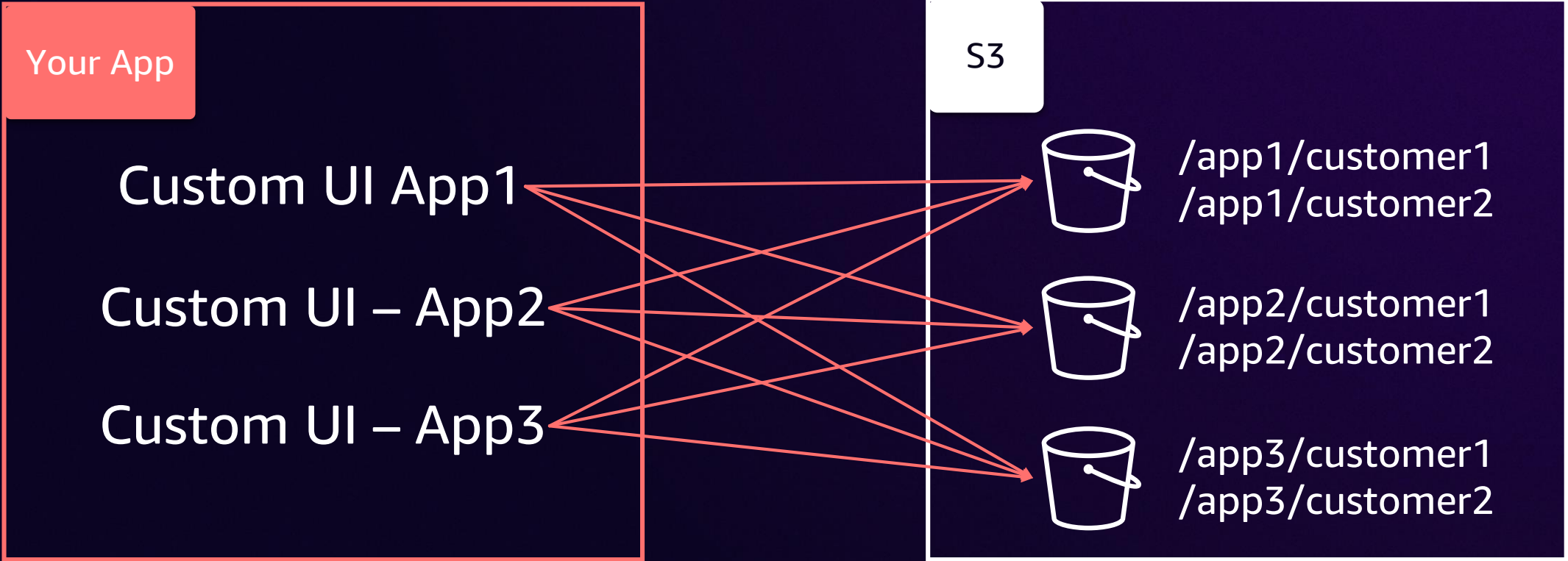
Improving time to write checkpoints by
up to 40%



Custom UX/UI problems to solve

Your application...

Application connections



GA Dec. 1, 2024

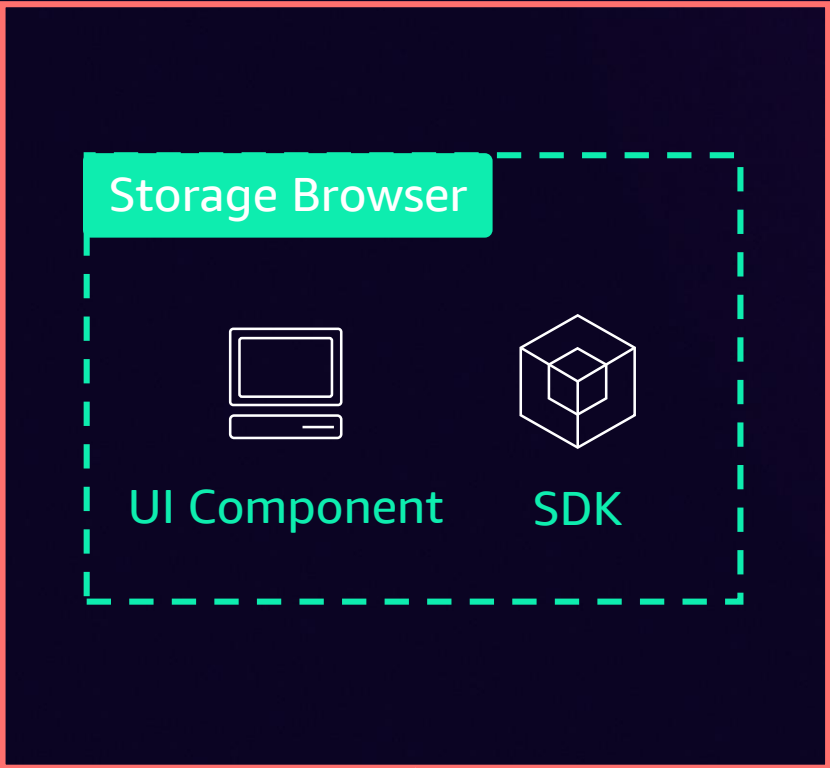
Storage Browser for Amazon S3

Provide your end users with a simple
interface for data stored in S3

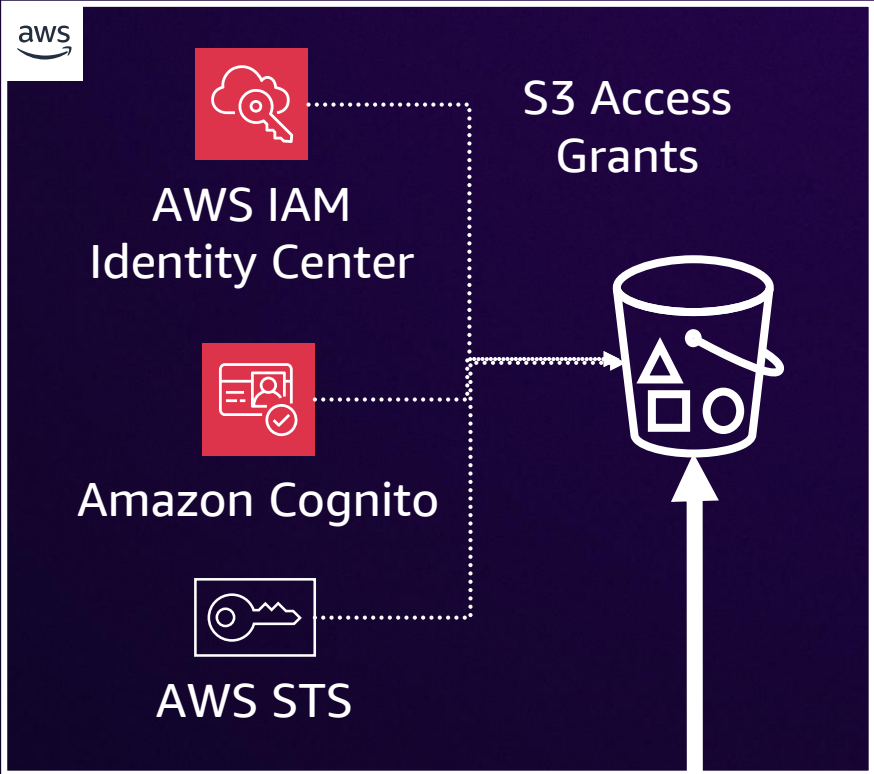


Storage Browser for Amazon S3

Your application...



AWS Cloud



Authorization
.....>

S3 APIs

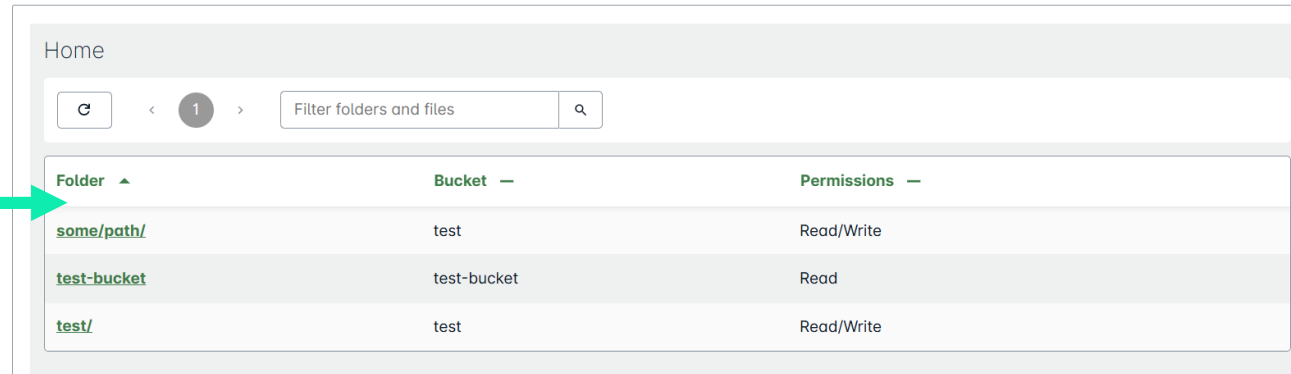


Storage Browser for Amazon S3

Customize

Theme object

You can use the `createTheme()` function to theme the `StorageBrowser` component



```
import * as React from 'react';
import {
  ThemeStyle,
  createTheme,
  defineComponentTheme,
} from '@aws-amplify/ui-react/server';
import { View } from '@aws-amplify/ui-react';

const storageBrowserTheme = defineComponentTheme({
  name: 'storage-browser',
  theme: (tokens) => {
    return {
      _element: {
        controls: {
          flexDirection: 'row-reverse',
          backgroundColor: tokens.colors.background.primary,
          padding: tokens.space.small,
          borderRadius: tokens.radii.small,
        },
      },
    };
  },
});
```

Copy

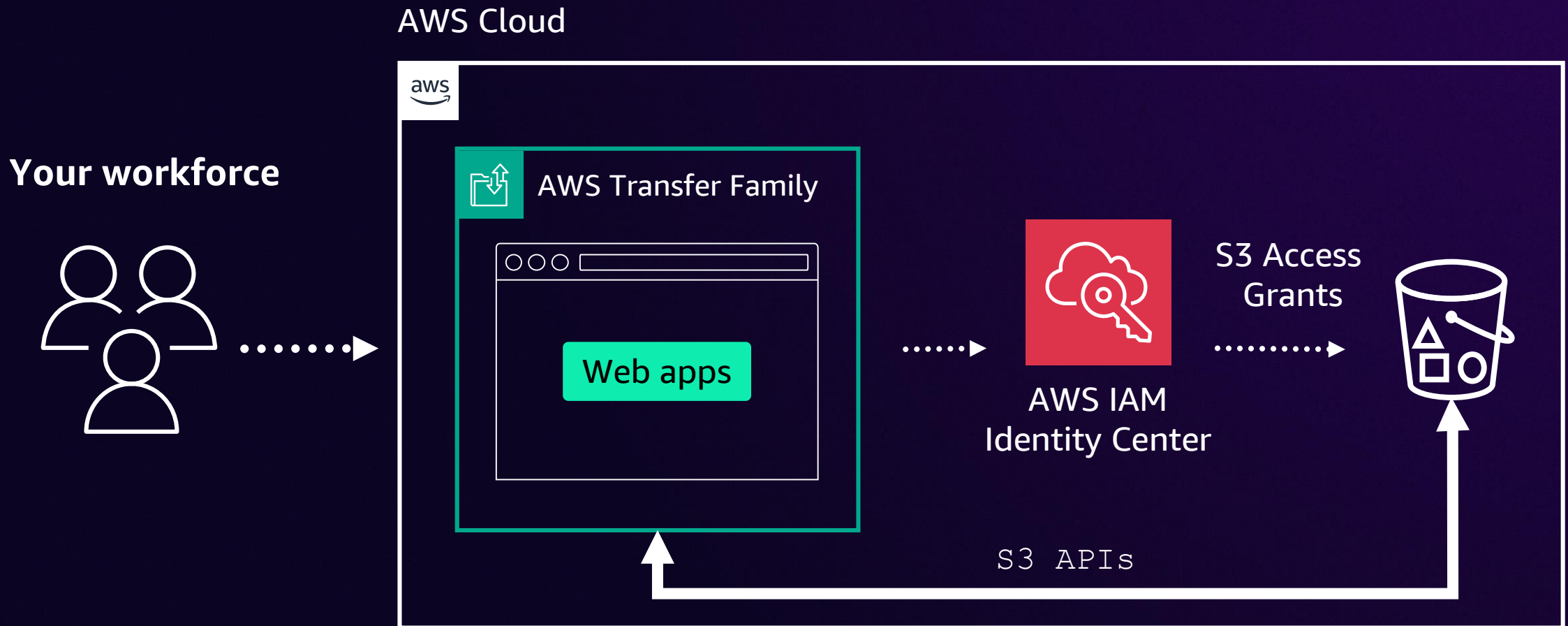
GA Dec. 1, 2024

AWS Transfer Family web apps

Fully managed interface for data
stored in Amazon S3



AWS Transfer Family web apps



AWS Transfer Family web apps

Customize Title & Logo

Design web app - optional

Web app design [Info](#)

Add customization so your web app aligns with your company branding and is visible to your end users.

Page title - optional

Specify the title of your web app that your end user would see displayed in their browser tab.

Title tag between 0 and 60 characters

Logo - optional

Upload a logo for your web app that will be displayed to your end users in the user interface. Maximum recommended

[Choose file](#)

Maximum file size: 50KB

Favicon - optional [Info](#)

Upload a favicon for your web app that your end user would see displayed in their browser tab.

[Choose file](#)

Maximum file size: 20KB

Shareable URL for your web app

amazon

demo user

Home / transferwebapps

transferwebapps

Search current folder Submit Include subfolders

| <input type="checkbox"/> | Name ^ | Type | Last Modified | Size |
|--------------------------|--------------------|--------|-----------------------|--------|
| <input type="checkbox"/> | FY 2019/ | Folder | | |
| <input type="checkbox"/> | FY 2020/ | Folder | | |
| <input type="checkbox"/> | FY 2021/ | Folder | | |
| <input type="checkbox"/> | FY 2022/ | Folder | | |
| <input type="checkbox"/> | FY 2023/ | Folder | | |
| <input type="checkbox"/> | FY 2024/ | Folder | | |
| <input type="checkbox"/> | Meeting Notes.docx | docx | 11/8/2024, 8:30:31 AM | 3.9 MB |

Static website hosting problems to solve

Your Website



Tasks to do

- DNS configuration
- Domain management
- TLS certificate
- CDN integration
- Logging

GA Oct. 30, 2024

Amazon S3 static website hosting with AWS Amplify

Seamlessly host static website content
stored on S3 with just a few clicks



Amazon S3 static website hosting with AWS Amplify

BUCKET PROPERTIES > CREATE AMPLIFY APP

Amazon S3 [Close]

Buckets

- Access Grants
- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- IAM Access Analyzer for S3

Block Public Access settings for this account

▼ **Storage Lens**

- Dashboards
- Storage Lens groups
- AWS Organizations settings

Object Lock
Disabled

Requester pays [Edit]

When enabled, the requester pays for requests and data transfer costs, and anonymous access to this bucket is disabled. [Learn more](#)

Requester pays
Disabled

Static website hosting [Edit]

Use this bucket to host a website or redirect requests. [Learn more](#)

We recommend using AWS Amplify Hosting for static website hosting
Deploy a fast, secure, and reliable website quickly with AWS Amplify Hosting. Learn more about [Amplify Hosting](#) or [View your existing Amplify apps](#)

Create Amplify app

S3 static website hosting
Disabled

30 launches in 60 minutes

1. S3 Tables
2. S3 Metadata (preview)
3. Higher Bucket Counts
4. Service Quota Support
5. ListBucket API Improvements
6. If-match condition for PUT & DELETE
7. If-none-match condition for PUT
8. IAM condition keys for conditional requests
9. New data integrity defaults
10. New context in 403 responses
11. S3 Access Grants support for SageMaker
12. S3 Access Grants support for Redshift
13. S3 Access Grants support for Glue ETL
14. S3 Access Grants support for boto3
15. S3 Access Grants ListCallerAccessGrants API
16. S3 Append for directory buckets
17. S3 Express support for CloudTrail
18. S3 Express support for SageMaker
19. S3 Express support for Amazon EMR
20. S3 Express support for PyTorch
21. S3 Express support for SSE-KMS
22. S3 Lifecycle support for S3 Express
23. S3 Express: 4 new AWS Regions
24. S3 storage classes for AWS Dedicated Local Zones
25. S3 Mountpoint Distributed caching
26. S3 Connector for PyTorch support for PyTorch Lightning checkpoints
27. S3 Connector for PyTorch support for distributed checkpoints
28. Storage Browser for Amazon S3
29. AWS Transfer Family for web apps
30. Amazon S3 static websites with AWS Amplify

Thank you!

Paul Meighan

Director, Product Management
Amazon S3 & Glacier
[@paulmeighan.bsky.social](https://bsky.app/profile/paulmeighan.bsky.social)

Mallory Gershenfeld

Senior Product Manager
Amazon S3



Please complete the session survey in the mobile app

