
AWS
re:Invent

STG208

Best practices for archiving large datasets with AWS

Matt Jordan

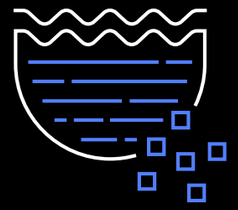
Sr. Software Dev Manager, Amazon S3 Glacier

Amazon Web Services, Inc.

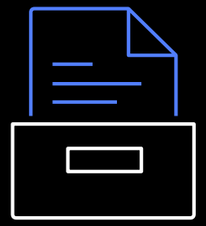
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Large data archival use cases



Data retention

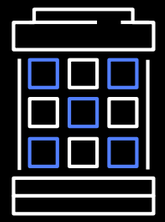


Compliance

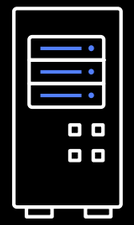


Disaster recovery

On-premises archival challenges



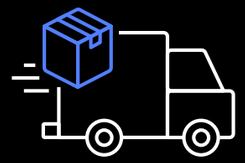
Hosting – space, power, cooling, network



Capital – servers, libraries, tapes



Administration and operations



Off-site storage and transportation

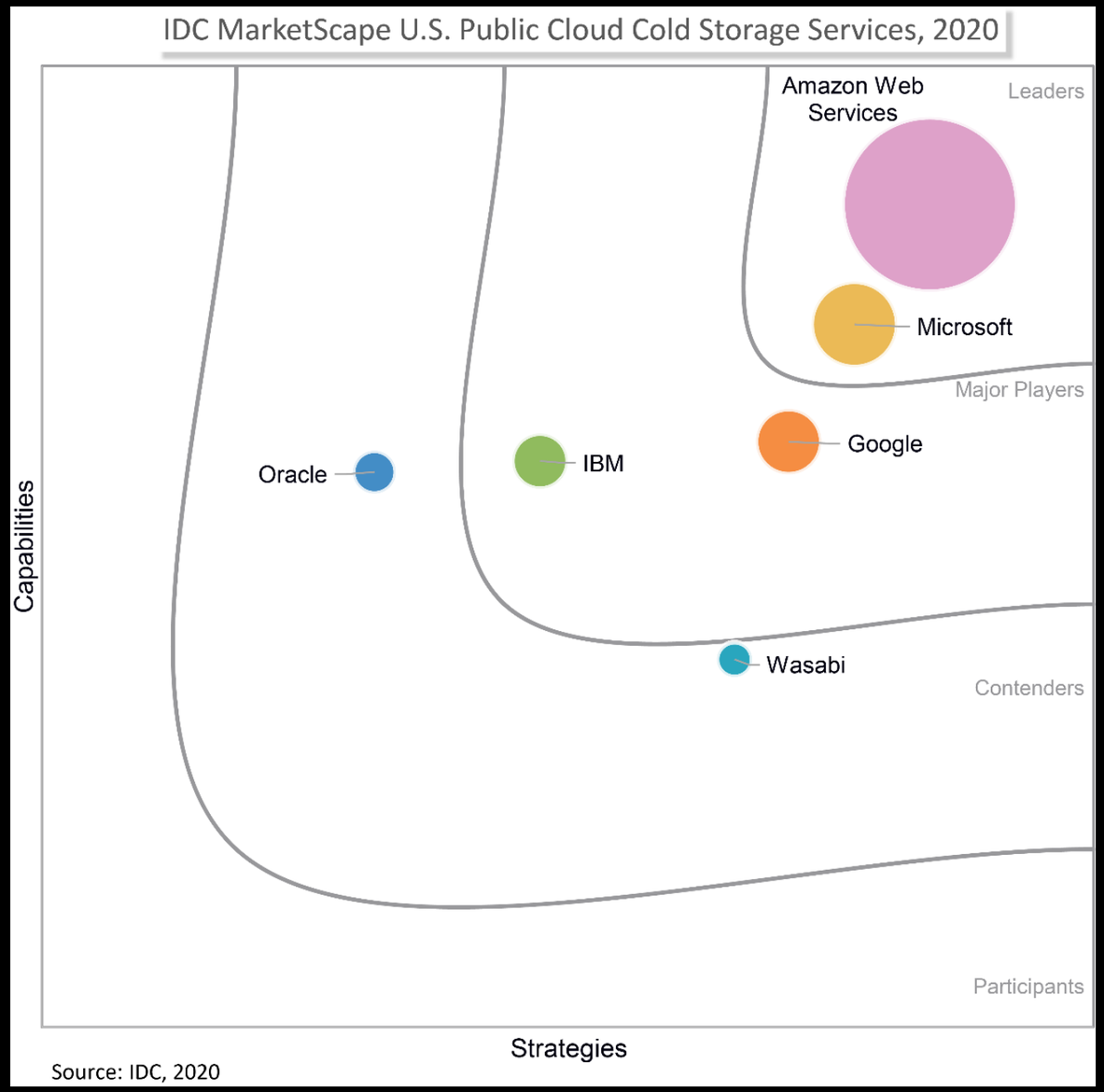


Opportunity cost

AWS recognized as a 2020 Cloud Leader in IDC Public Cloud Cold Storage MarketScape

"AWS was one of the first vendors to offer a branded solution in the public cloud cold storage market with the release of Amazon Glacier in 2012."

Source: IDC MarketScape: U.S. Public Cloud Cold Storage Services 2020 Vendor Assessment, Doc # US46855520, September 2020.
https://aws.amazon.com/resources/analyst-reports/idc-public-cloud-cold-storage-2020/



Your choice of Amazon Simple Storage Service (Amazon S3) storage classes



S3 Standard

- Active, frequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0210/GB



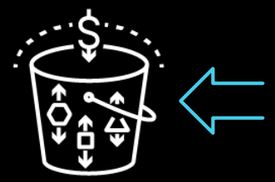
S3 Standard-IA

- Infrequently accessed data
- Milliseconds access
- ≥ 3 AZ
- $0.0125/GB



S3 One Zone-IA

- Re-creatable, less accessed data
- Milliseconds access
- 1 AZ
- $0.0100/GB



S3 Intelligent-Tiering

- Data with changing access patterns
- Opt-in for automatic archiving
- Millisecond to hours access
- ≥ 3 AZ
- $0.0210 to $0.00099/GB



S3 Glacier

- Archive data
- Minutes or hours access
- ≥ 3 AZ
- $0.0040/GB



S3 Glacier Deep Archive

- Archive data
- Hours to access
- ≥ 3 AZ
- $0.00099/GB

← Frequent — Access frequency — Infrequent →

Amazon S3 Intelligent-Tiering storage class



Only cloud storage that delivers **automated cost savings**

- **Up to 40% cost savings** using the Frequent and Infrequent Access tiers
- **Up to 95% cost savings** using the Archive and Deep Archive tiers

Monitors and optimizes costs at a granular object level

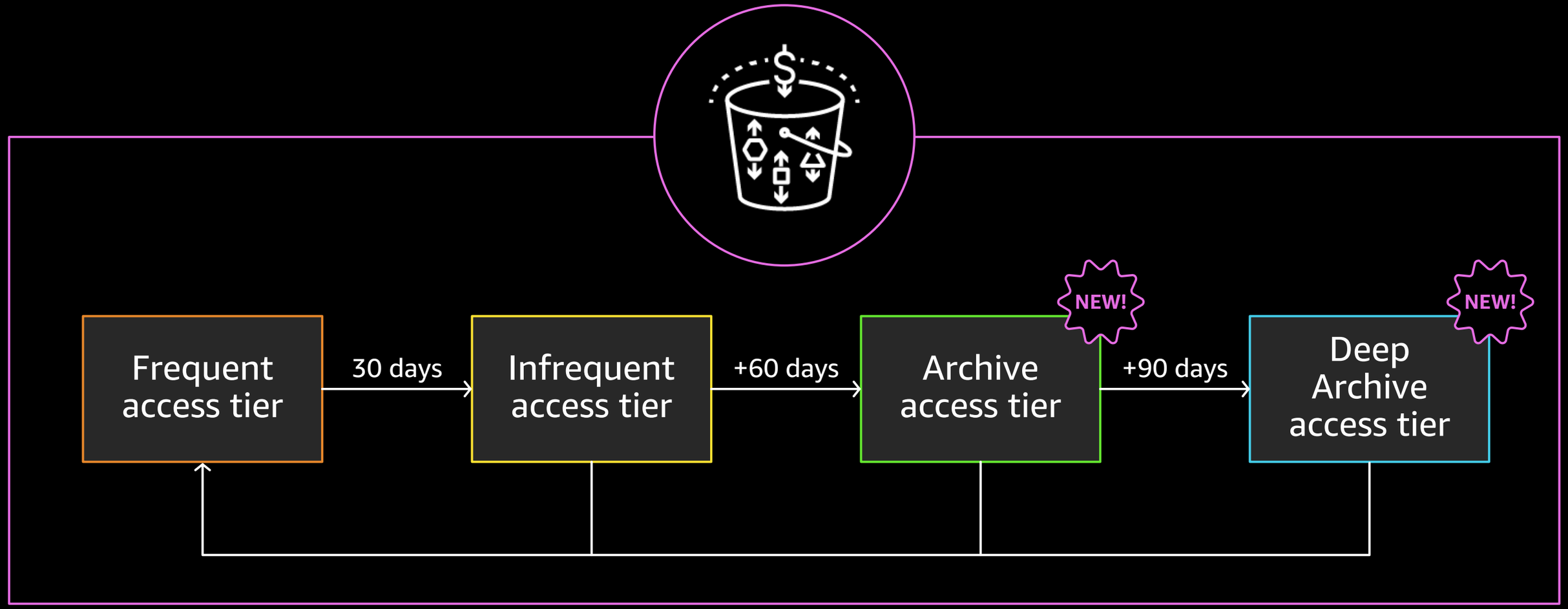
Moves objects between **4 access tiers** for a small monthly monitoring and automation fee

Two, low-latency access tiers for frequent and infrequent access and two **new optional archive access tiers** designed for access in minutes and hours

No operational overhead, no lifecycle fees, and no retrieval fees

Designed for 99.9% availability and 99.999999999% (11 9s) of durability

Amazon S3 Intelligent-Tiering in action



Storage classes for large data use cases



Data retention



Disaster recovery



Compliance



S3 Intelligent-Tiering



S3 Glacier



S3 Glacier Deep Archive

← Unknown — Access pattern — Known →

← Frequent — Access frequency — Infrequent →

Accessing data





	S3 Glacier	S3 Glacier Deep Archive
Expedited	Typically 1–5 minutes	Not available
Standard	Typically 3–5 hours	Typically within 12 hours
Bulk	Typically 5–12 hours	Typically within 48 hours

Archival considerations



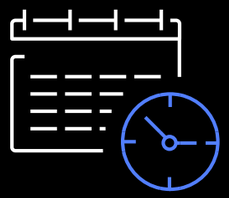
Object lifetime



Number of objects



Object size



Retrieval frequency



Retrieval ordering

Cost Comparison: 1 PB

IMPACTS OF OBJECT SIZE ON COSTS

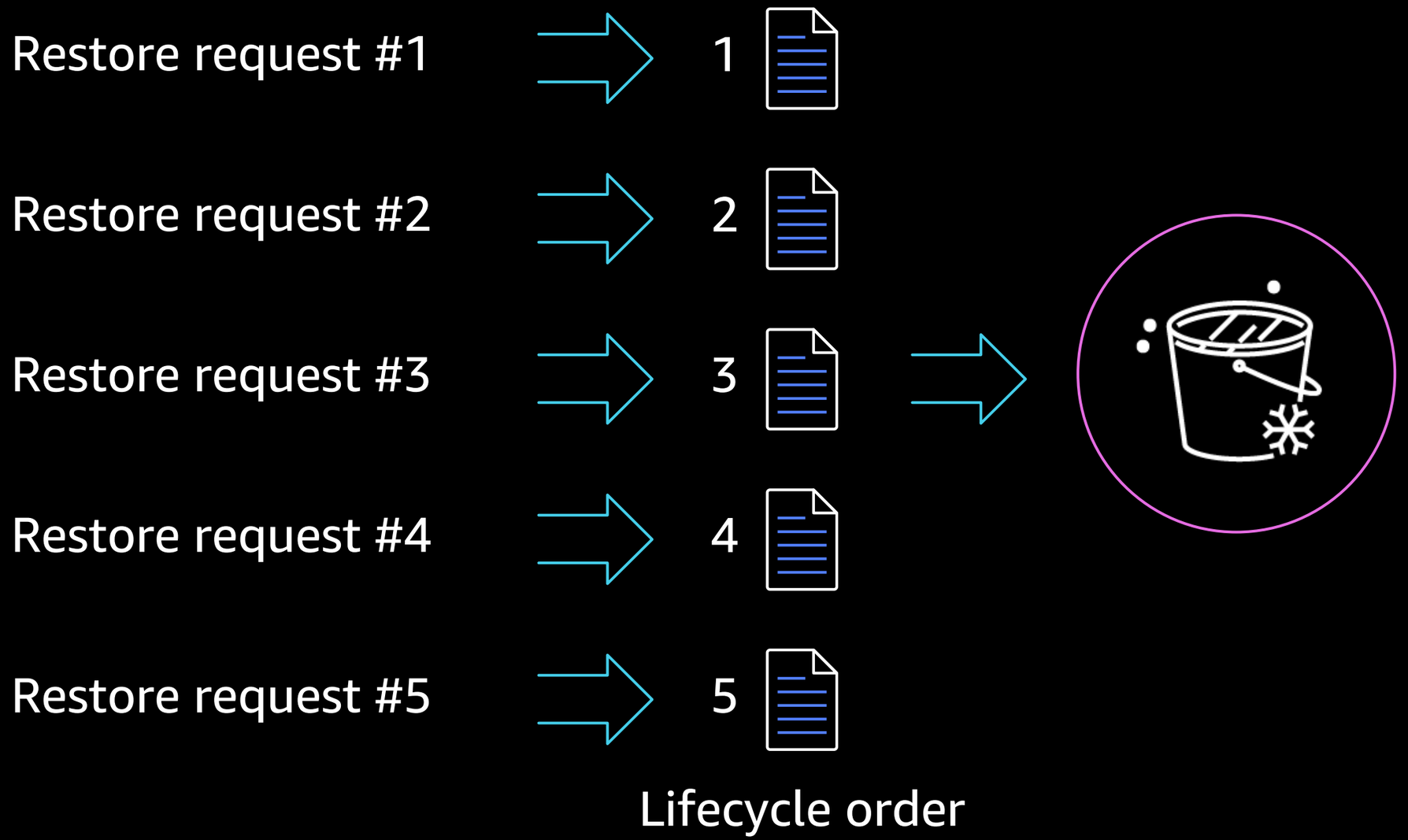
	1 MB	1 GB	1 TB
Restore requests	1,000,000,000	1,000,000	1,000
S3 Glacier Storage	$4,160	$4,000.16	$4,000
S3 Glacier Deep Archive Storage	$1,029.60	$990.04	$990
Lifecycle transition costs	$50,000	$50	$.05
S3 Glacier Standard retrievals	$60,400	$10,050.40	$10,000.05
S3 Glacier Bulk retrievals	$27,600	$2,525.10	$2,500.03
S3 Glacier Deep Archive Standard retrievals	$120,800	$20,100.80	$20,000.10
S3 Glacier Deep Archive Bulk retrievals	$27,600	$2,525.10	$2,500.03

Cost Comparison: 1 PB

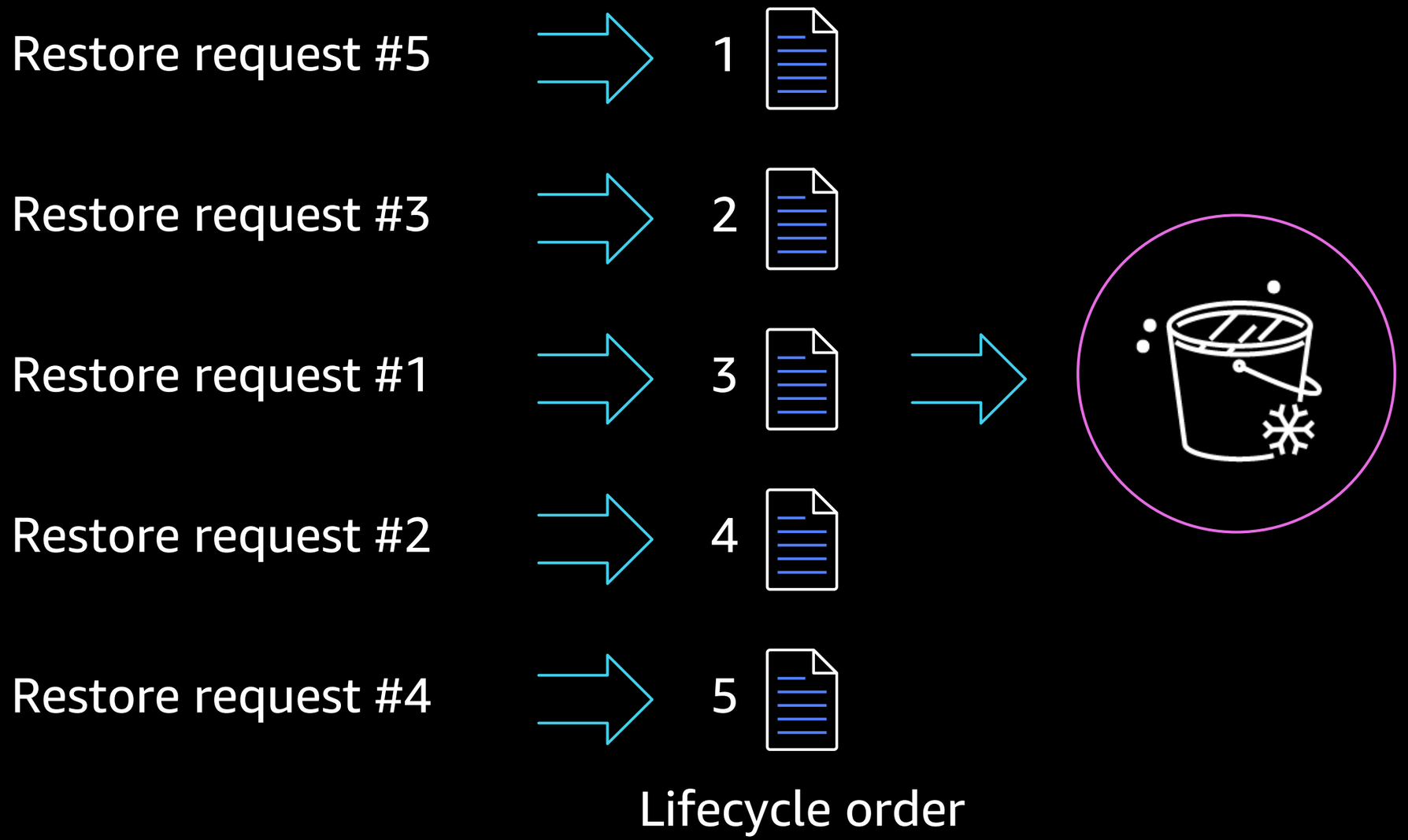
IMPACTS OF OBJECT SIZE ON COSTS

	1 MB	1 GB	1 TB
Restore requests	1,000,000,000	1,000,000	1,000
S3 Glacier Storage	$4,160	$4,000.16	$4,000
S3 Glacier Deep Archive Storage	$1,029.60	$990.04	$990
Lifecycle transition costs	$50,000	$50	$.05
S3 Glacier Standard retrievals	$60,400	$10,050.40	$10,000.05
S3 Glacier Bulk retrievals	$27,600	$2,525.10	$2,500.03
S3 Glacier Deep Archive Standard retrievals	$120,800	$20,100.80	$20,000.10
S3 Glacier Deep Archive Bulk retrievals	$27,600	$2,525.10	$2,500.03

Restore performance: sequential ordering



Restore performance: random ordering



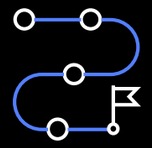
Best practices



1. Have an archival plan



2. Know your data access



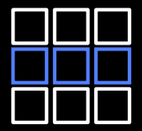
3. Plan ahead for asynchronous interactions



4. Have a cost plan for the entire lifecycle

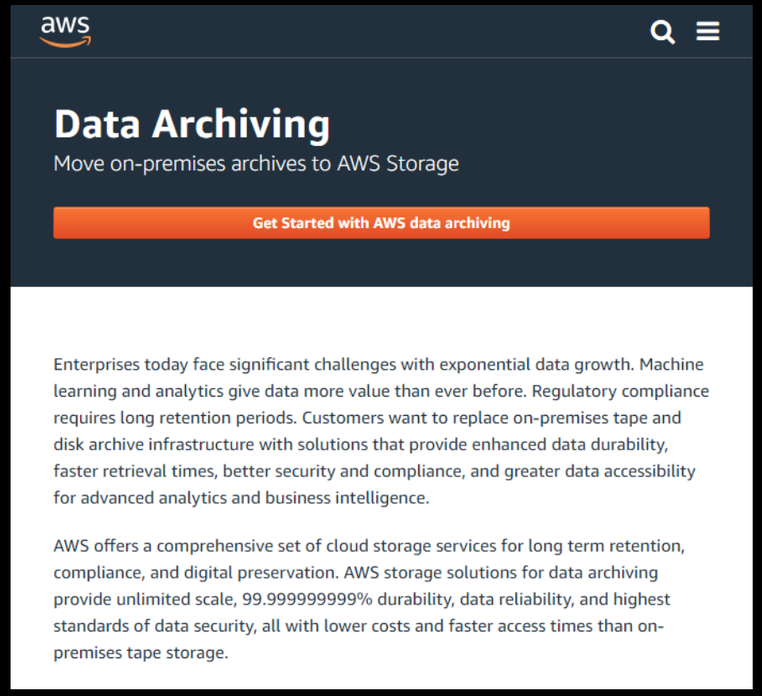


5. Optimize your objects for archival plan



6. Minimize randomness when restoring

Resources



Data Archiving:
https://aws.amazon.com/archive



Optimizing storage costs using Amazon S3:
https://aws.amazon.com/s3/cost-optimization



Amazon S3 Glacier & S3 Glacier Deep Archive:
https://aws.amazon.com/glacier

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.