
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

PRT315

3 innovations that redefine data protection for Amazon S3

Peter Imming

Principal Product Manager, Amazon S3
AWS

Mark Huber

Sr. Director of Engineering Enablement
Cox Automotive

Woon Ho Jung

Co-founder and CTO
Clumio

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

1. Why and how you should protect Amazon S3 data
2. How Cox Automotive simplified its Amazon S3 protection with Clumio
3. Architecture deep dive into Clumio: a highly scalable data protection platform

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3: The foundation for data-driven innovation

Data lakes and AI / ML



- Seamless data import
- Direct querying
- Ready for analytics and AI

Cloud-native apps



- Broad Container support
- Databases
- Streaming applications

IoT and Logs



- Durable storage
- Near limitless scale
- Highly available

280 TRILLION OBJECTS!

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What data in Amazon S3 needs protection?



Application data

- User-generated data
- Media
- Configs



Data lakes

- Business app data
- AI / ML training data
- Test/Dev data



Sensitive information

- Company confidential information
- Intellectual property
- PII
- Credentials
- Material nonpublic financials



Compliance data

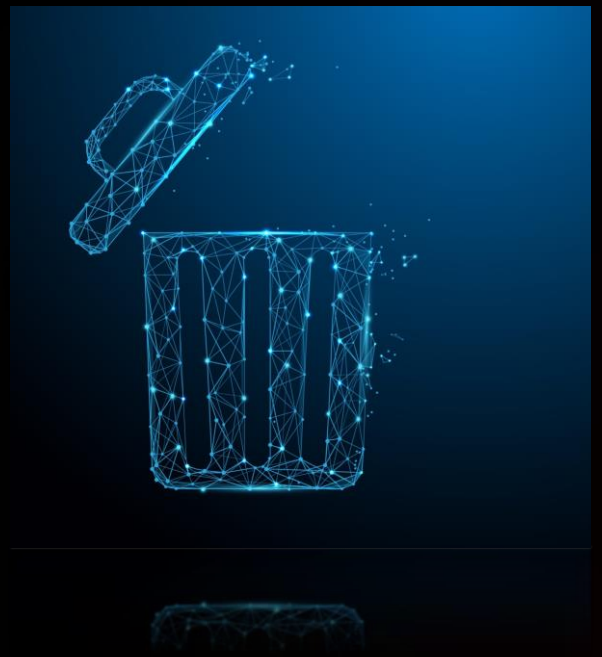
- Health records
- Employee records
- Financial records

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why should I backup my data in Amazon S3?

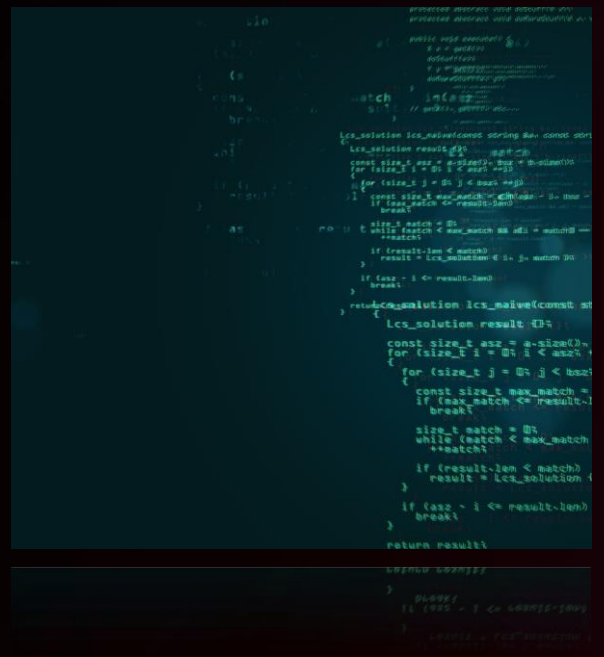
Human errors are more common than malicious incidents, and just cybersecurity won't suffice

Accidental data loss



- Human errors
- Inadvertent deletions or overwrites
- Natural disasters

Software errors



- Application deletions of data
- Erroneous deletions from misconfigured scripts
- Incorrectly configured data lifecycle rules

Malicious actors



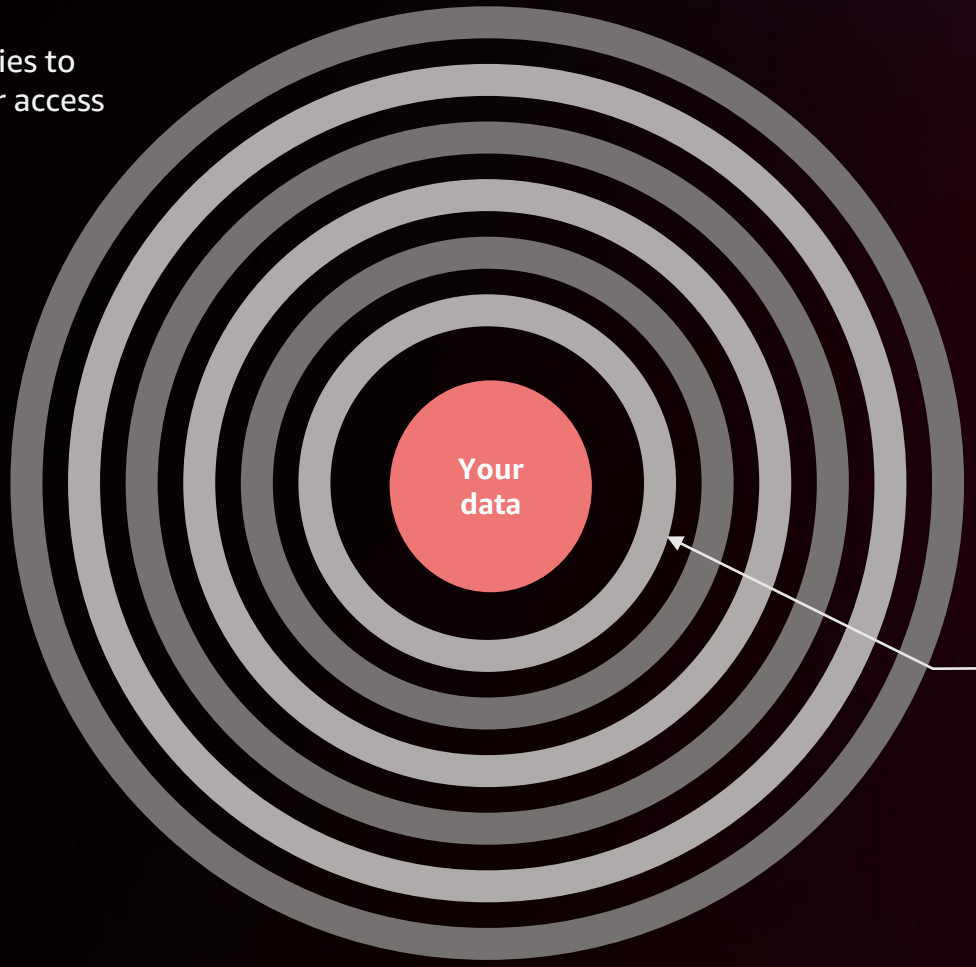
- Malicious deletions
- Ransomware mass re-encryption events
- Intentional fault injections into datasets

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 multi-layered data protection options

1. IAM policies to restrict user access
2. Enable Amazon S3 Object-Versioning on your Amazon S3 buckets
3. Use Amazon S3 Object Lock to make your data Immutable (data and backups)
4. Enable Amazon S3 Replication; use Amazon S3 Replication with Replication Time Control for SLA-backed replication performance



5. Amazon S3 Storage Lens: Enable Amazon S3 Storage Lens to get multi-account level view of data protection settings and risk levels
6. Enable AWS Storage Competency Partner solutions such as Clumio for centralized cross-account and cross-Region independent Amazon S3 backups

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cox Automotive

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cox Automotive

Transforming the Way the World Buys, Sells, Owns and Uses Vehicles

COX AUTOMOTIVE

Autotrader · Cox Automotive Mobility · Dealer.com · Dealertrack · Dickinson Fleet Services · F&I Express · Fyusion · Kelley Blue Book · Manheim · NextGear Capital · NTI Spiers New Technologies · vAuto Live Market View · VinSolutions · xtime

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cox Automotive is all in on AWS

Cloud Native


Amazon S3


Amazon CloudFront


Amazon API Gateway

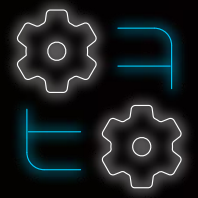

Amazon Lambda


Amazon DynamoDB


Amazon Aurora

Well-architected




Operational Excellence


Security


Reliability


Performance Efficiency


Cost Optimization


Sustainability

Scaled

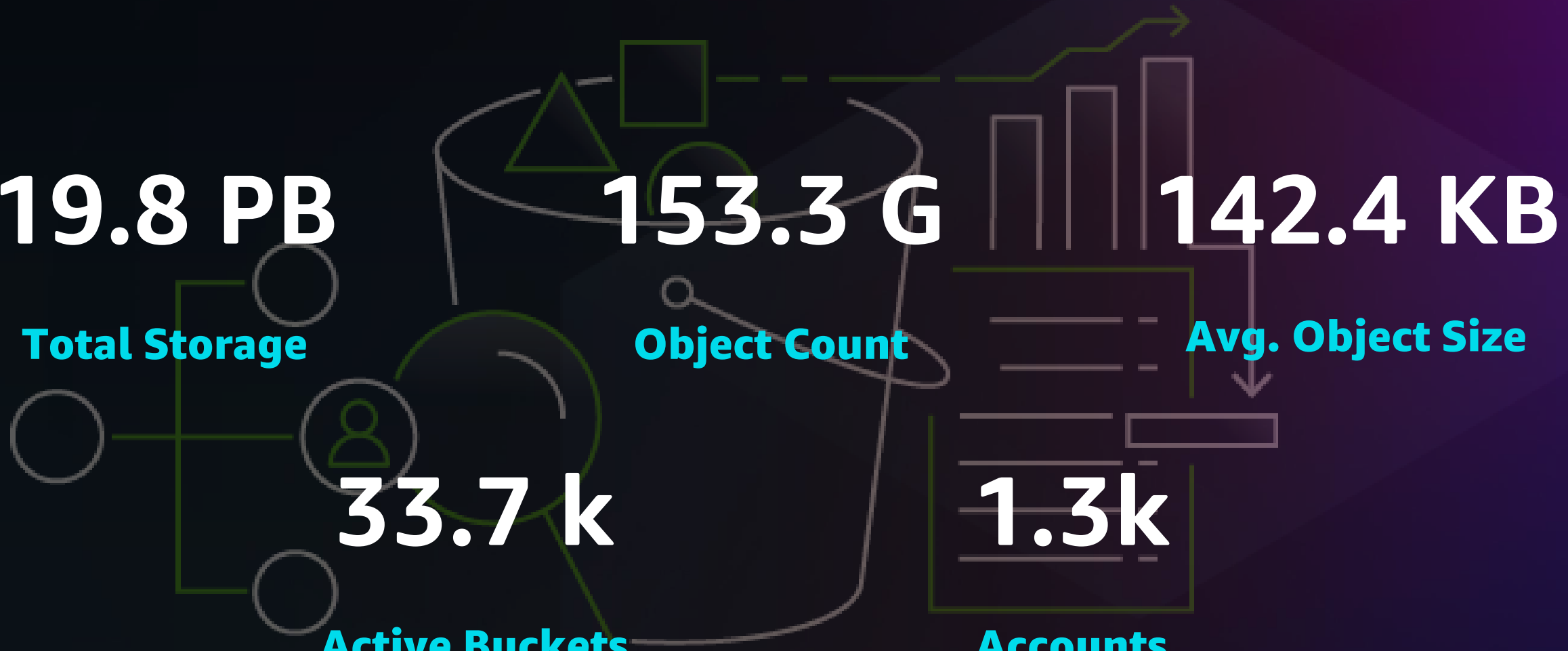
500+ Delivery Teams

1300+ AWS Accounts

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cox Automotive Amazon S3 Data By The Numbers



19.8 PB

Total Storage

153.3 G

Object Count

142.4 KB

Avg. Object Size

33.7 k

Active Buckets

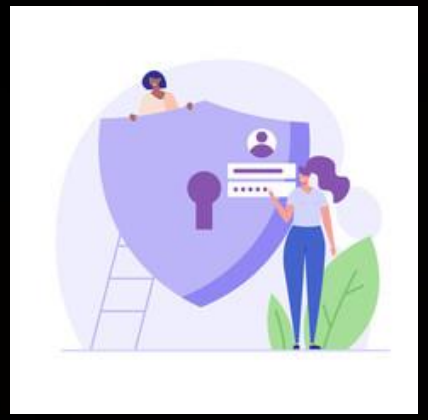
1.3k

Accounts

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Learnings from trying to build a backup strategy



If backups are not core competency, it's best to consume it as a service



Ransomware risks rising each day, and air gapped solutions are a critical component of our strategy



Can get complex to self-manage thousands of versions and replicated copies



Restoring to a specific point in time is not easy

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why we chose Clumio for AWS Data Protection



Exceptional competency

Maxing out AWS capabilities in unprecedented ways



Super simple

Protection groups for org-wide standardization enable us to scale



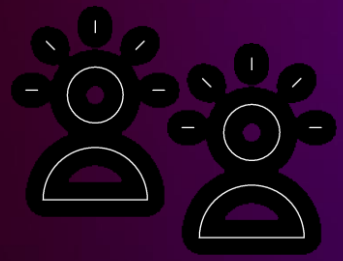
Air gapped & Native

Improves our ransomware mitigation posture without sacrificing the native power of AWS



High Performance

Delivers on speed and scale, which is critical to org of our size



Partners in design

Transparent with architecture, worked as design partners

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Results from switching to Clumio

Simpler, standardized operations

- Standardized data protection for 1300+ accounts
- Measured and defined resilience targets for 600+ workloads as part of CORE program
- Clumio Terraform provider allows automated rollout of protection policies for new accounts (7-10/week)
- Teams can get data protected in minutes of receiving a new AWS account

Better cost visibility

- Aligned with company's 'pay-as-you-use' policy
- Clear consumption reporting
- Easy charge-back
- Transparent credits system simplifies purchases

Ease of use

- Every new account comes pre-baked with Clumio
- Standardized tag strategy with sensible defaults
- Standardized protection groups makes it easy to get data protected across teams
- BYOK

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Clumio: Data protection at the scale of the internet



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

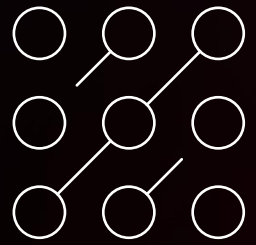
Protecting Amazon S3 is a HUGE engineering challenge

Data velocity



- New objects created every second
- Fast-changing buckets need to be inventoried continuously
- Changes, versions, metadata need to be updated constantly

Data volume



- Large enterprises have exabytes of data
- Mid-sized enterprises will get there soon
- Tens of billions of objects per bucket

Data variety



- Varying lifecycle rules
- Varying change-rates
- Disparate object types

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.
CLUMIO

Clumio: Architectural Overview



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

CLUMIO

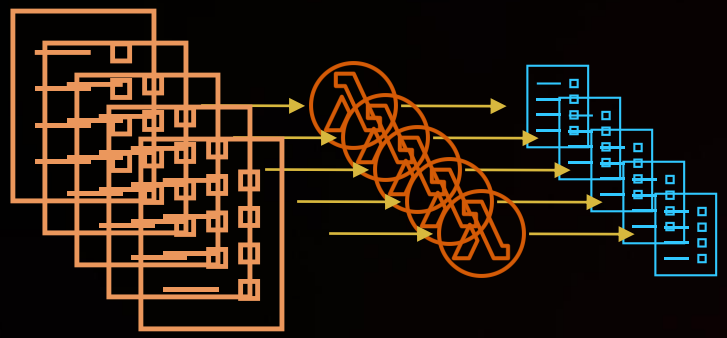
Efficient inventorying at extreme scale

Backup ONLY what is required

Filter based on—
1. Prefixes
2. Storage classes
3. Timestamps
4. …

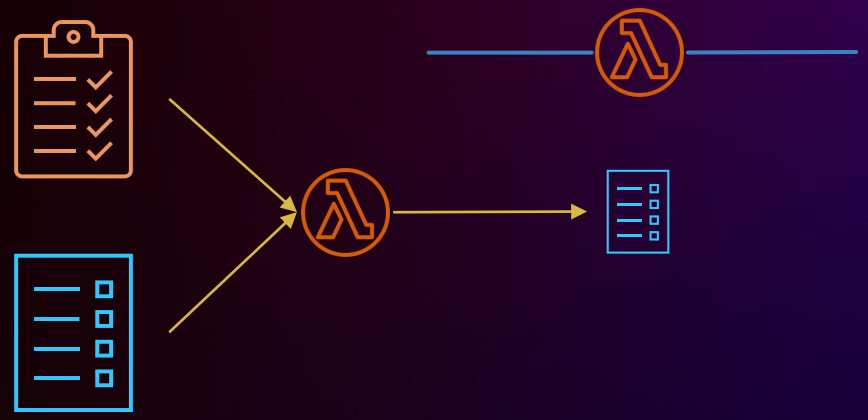
TB-sized lists

Concurrent processing



Validate Inventory periodically

Check for—
1. Missing objects
2. Dropped messages
3. 100% capture rate
4. …



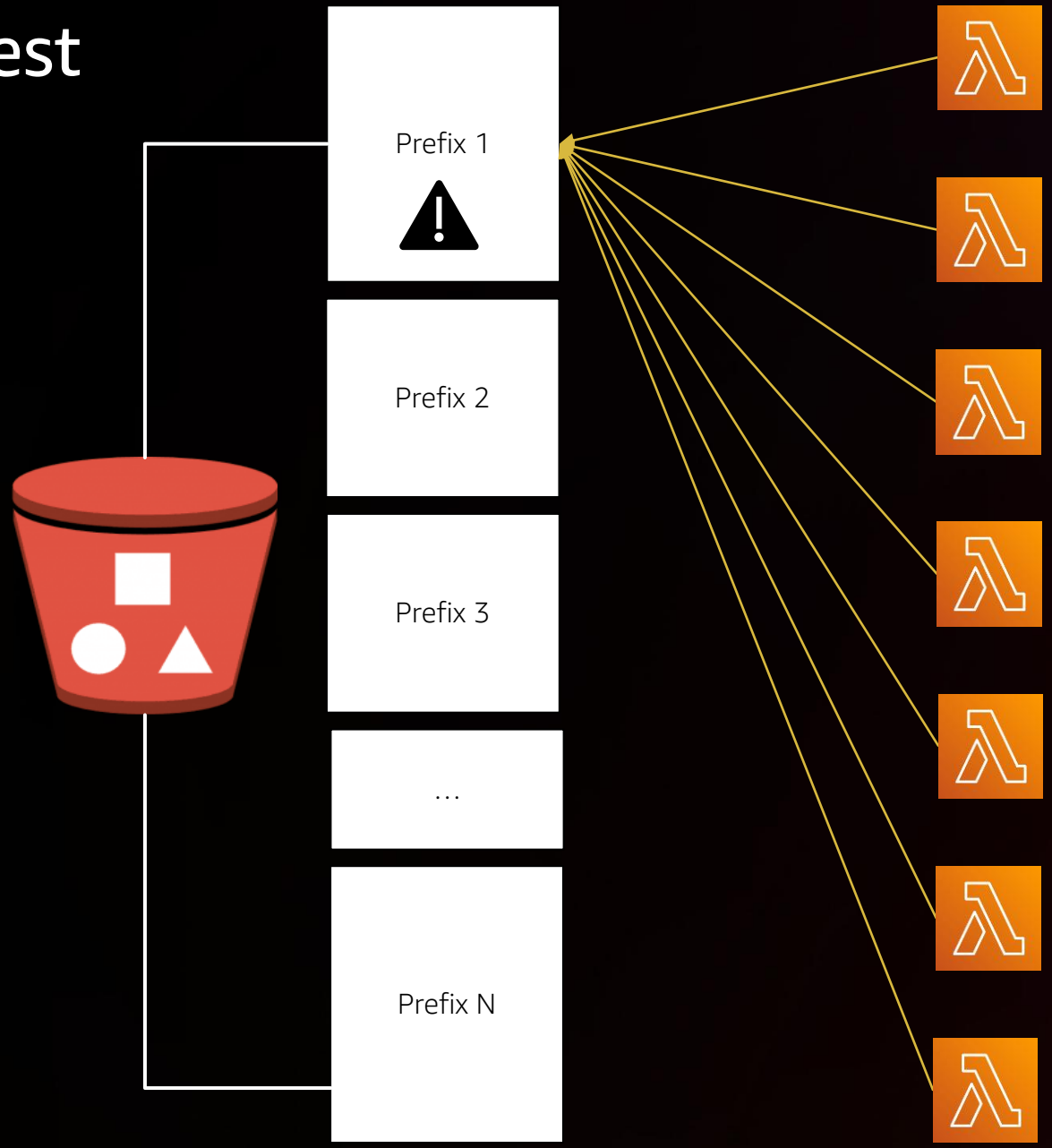
Compare billion-object inventory lists

Continuously scan, sort, merge, etc.

aws

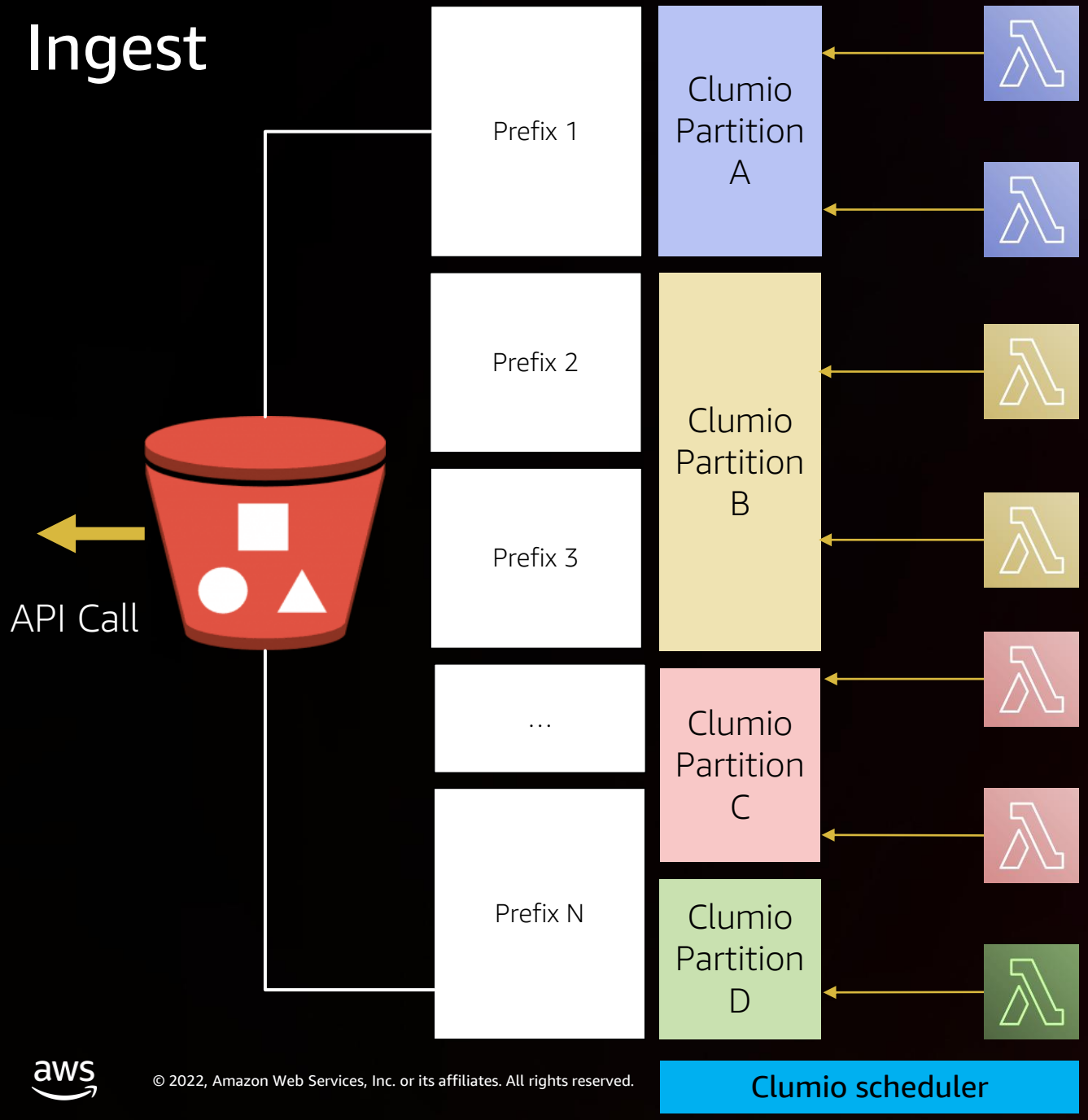
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Ingest



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

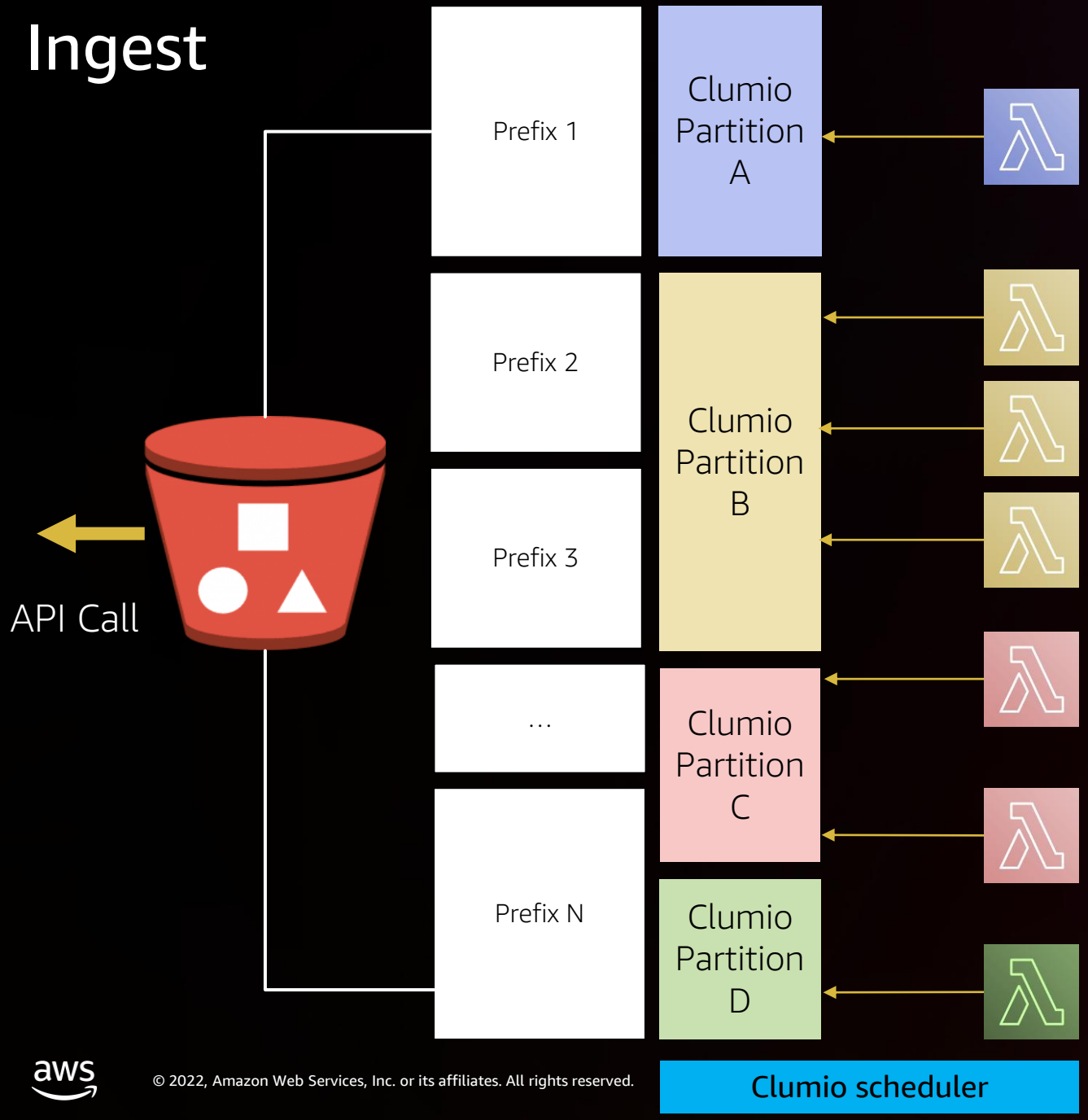
Ingest



- Use key heuristics to intelligently partition
- Allocate partitions, distribute and schedule functions

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Ingest



- Use key heuristics to intelligently partition
- Allocate partition, distribute and schedule functions
- If production bucket is being used, auto-tune scheduling
- Dynamically rate-throttle
- Auto retry on failure

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Restore / Metadata management

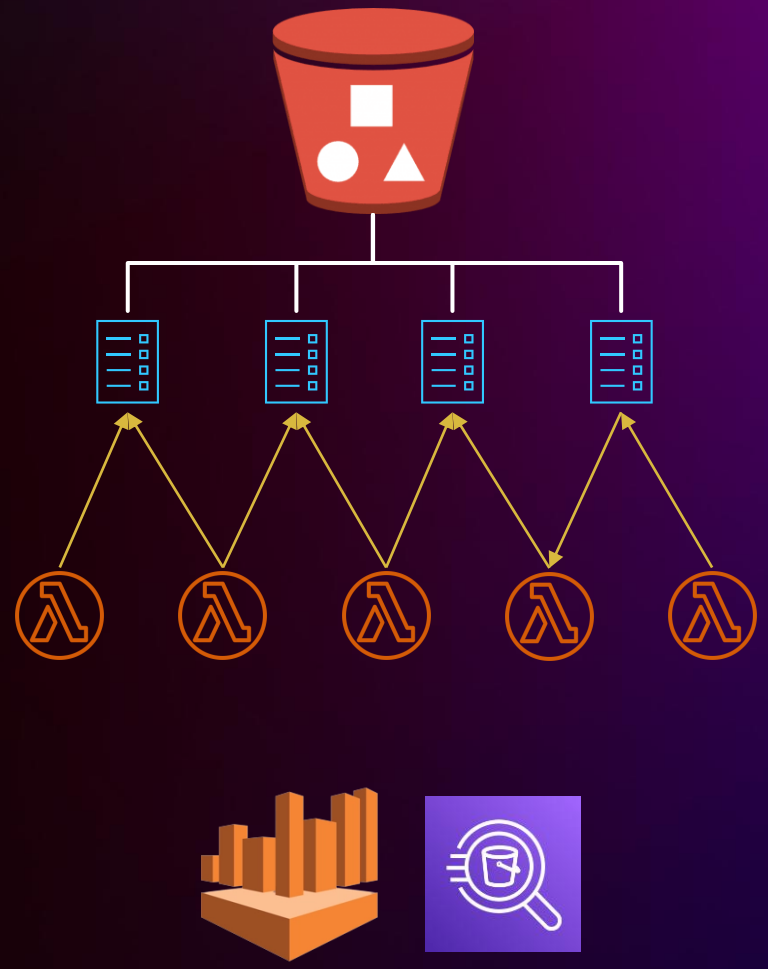
Restore ONLY what is required, FAST

1. Granular restore at the level of objects
2. Using timestamps, prefixes, etc.
3. Prioritize recovery based on criticality

Billions of metadata entries organized in Parquet files

Optimize metadata objects for querying

Partition for optimal query performance



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Observability

- Failure handing
- Monitoring
- Troubleshooting
- Upgrades

Hundreds of customers

X  X  X 

1000s of Lambda functions

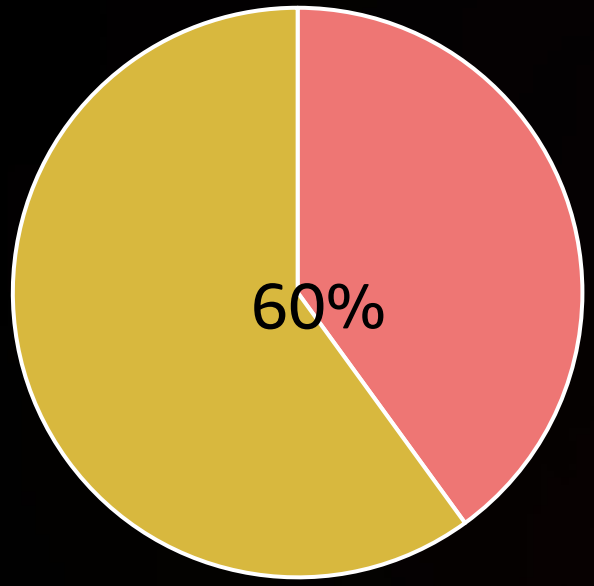
Continuous backups globally

← Clumio Workflow Engine →

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architectural efficiency leads to cost efficiency

% enterprises expected to encounter cost overruns





Your data

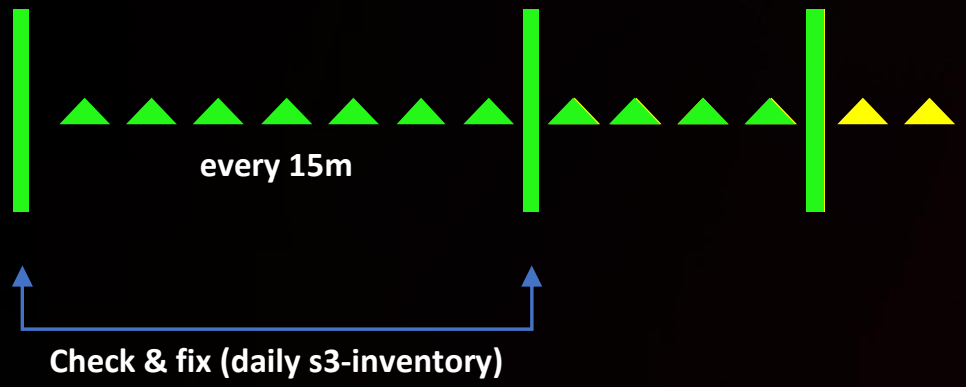


Your data in Clumio

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

New announcements!

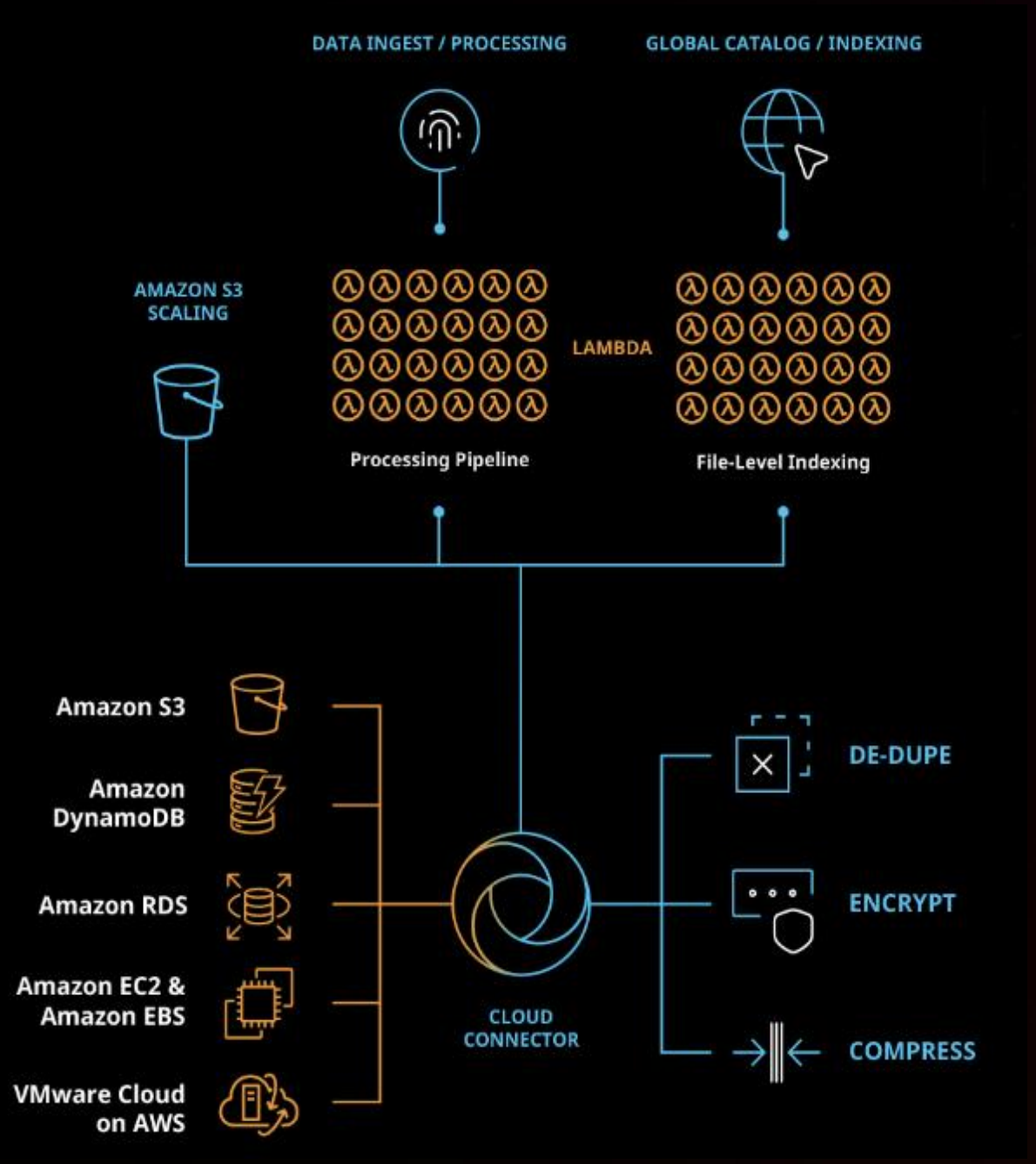
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



15-min RPO
at any scale

Continuous change tracking

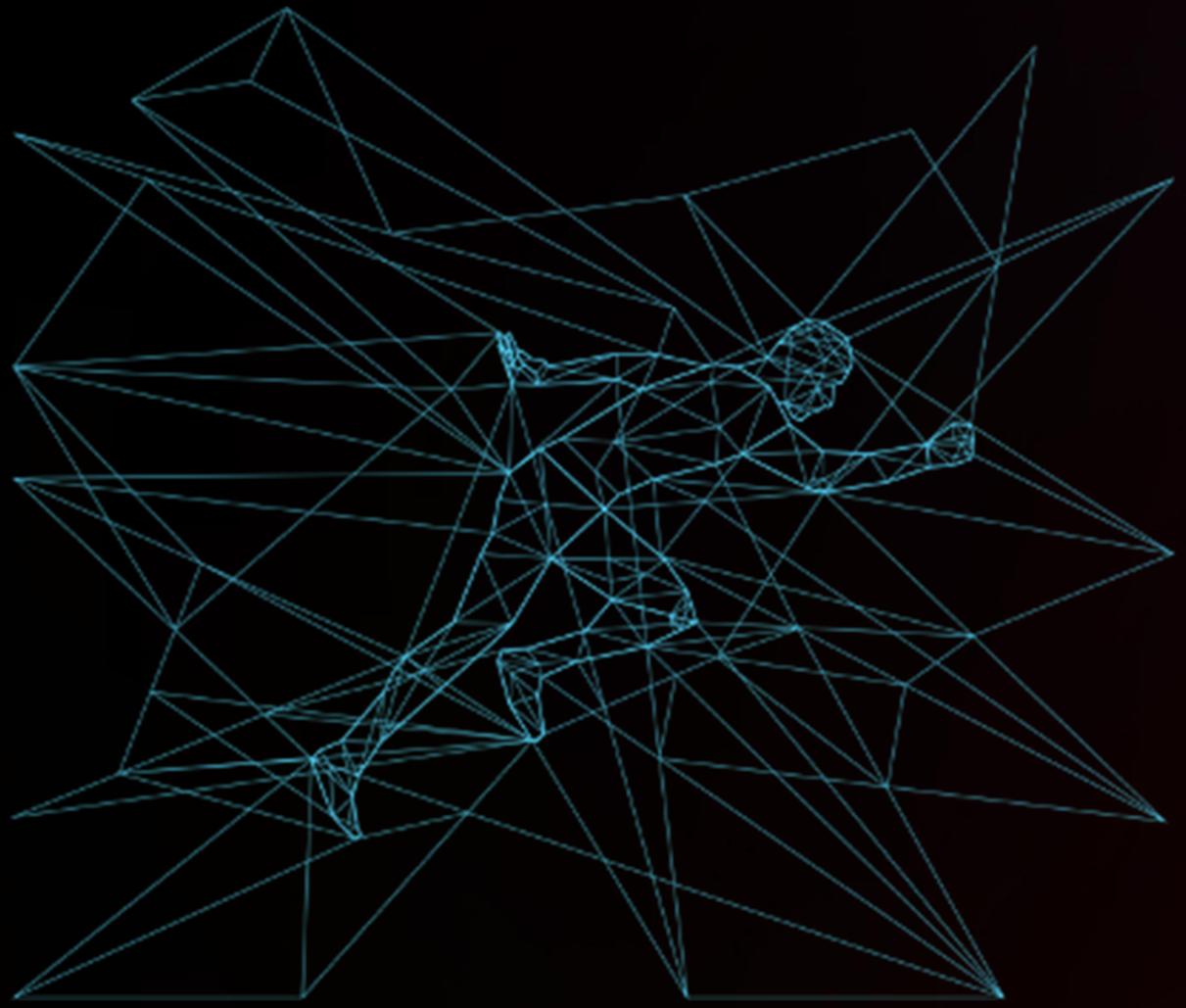
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



30 billion objects per bucket

Continuous protection for massive Amazon S3 data lakes

CLUMIO



0 RTO
instant recovery

Live, immutable, point-in-time datastore

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Visit us at Booth 1320

Across from the AWS Village

clum.io/try

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.