AWS re:Invent

STG307-R1

Amazon S3 + FSx for Lustre: Deep dive on high performance file storage

Edward Naim
GM, Amazon FSx
Amazon Web Services

Sayan Saha
Principal PM, Amazon FSx
Amazon Web Services

Darryl Osborne
Solution Architect, Amazon FSx
Amazon Web Services

AWS re:Invent

aws

Agenda

- Making your workloads run faster & cheaper with Amazon FSx for Lustre
- Amazon S3 + Amazon FSx
- Data processing options
- Performance
- Total cost of ownership (TCO)
- Amazon FSx for Lustre in action

Making your workloads run faster & cheaper with Amazon FSx for Lustre

AWS re:Invent

aws

Why run compute workloads on AWS?

Elasticity – **Virtually unlimited** infrastructure enabling scaling and agility not attainable on-premises

Better ROI

Functionality – Rich set of instance types, automation, orchestration, networking & visualization solutions

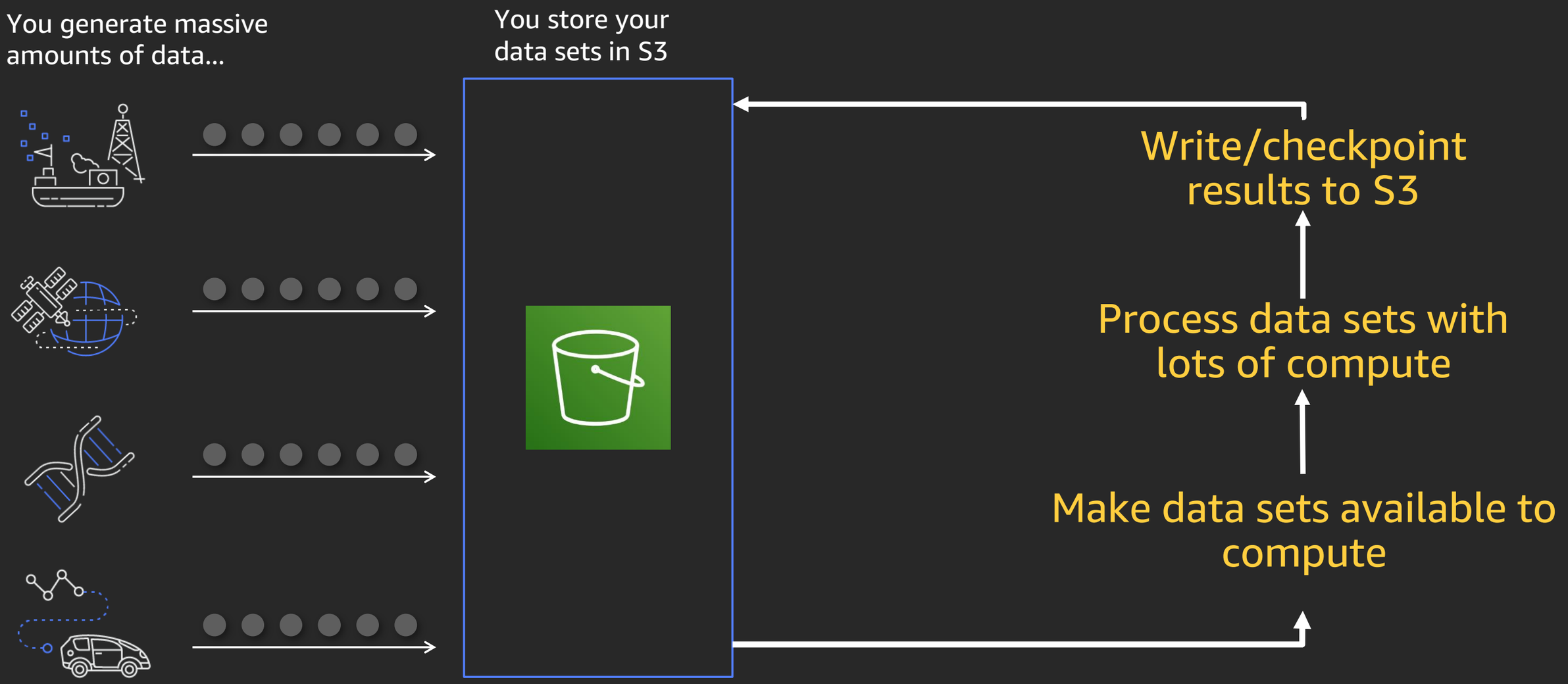
Agility – Fast fail, iterate quickly, reduce time to results

Faster time to results

Global Infrastructure – Increased collaboration with secure access to clusters around the world

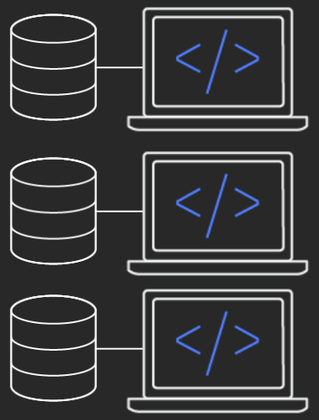
Cost Optimized – Pay for only what you use

Compute workloads on AWS typically process data sets that are stored on Amazon S3



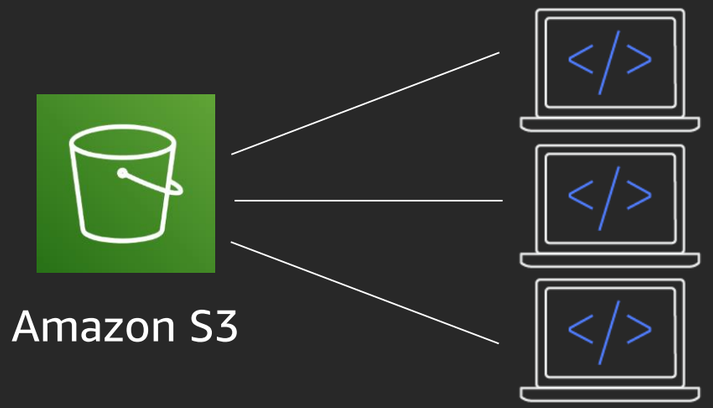
Options for cloud-native data processing

To process your data sets in Amazon S3, you either move them to temporary storage or process them directly on S3



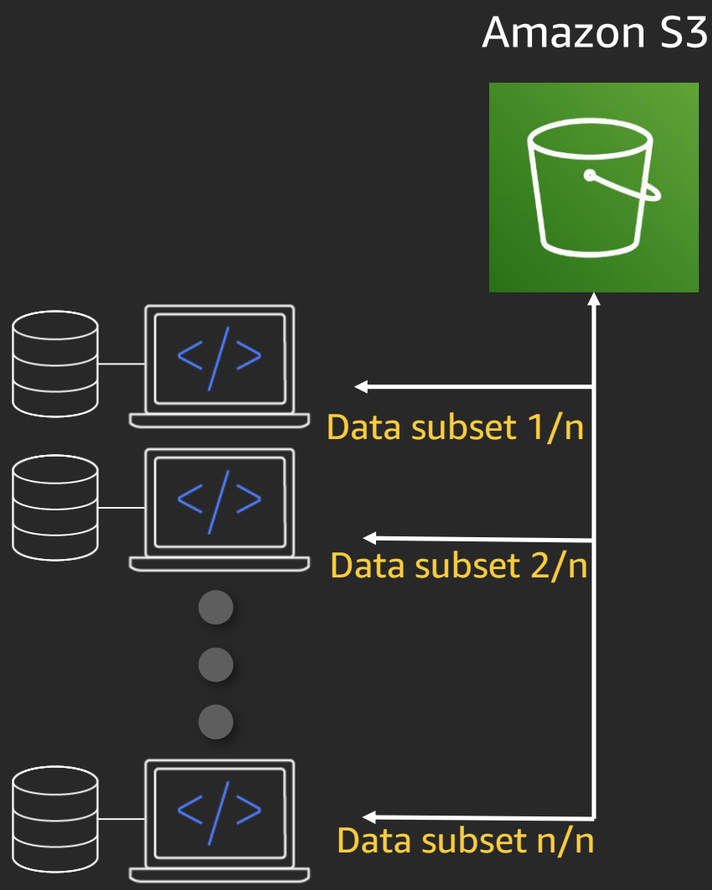
On EBS or instance storage

Self-managed file systems



Directly on S3

Data processing with EBS or instance storage

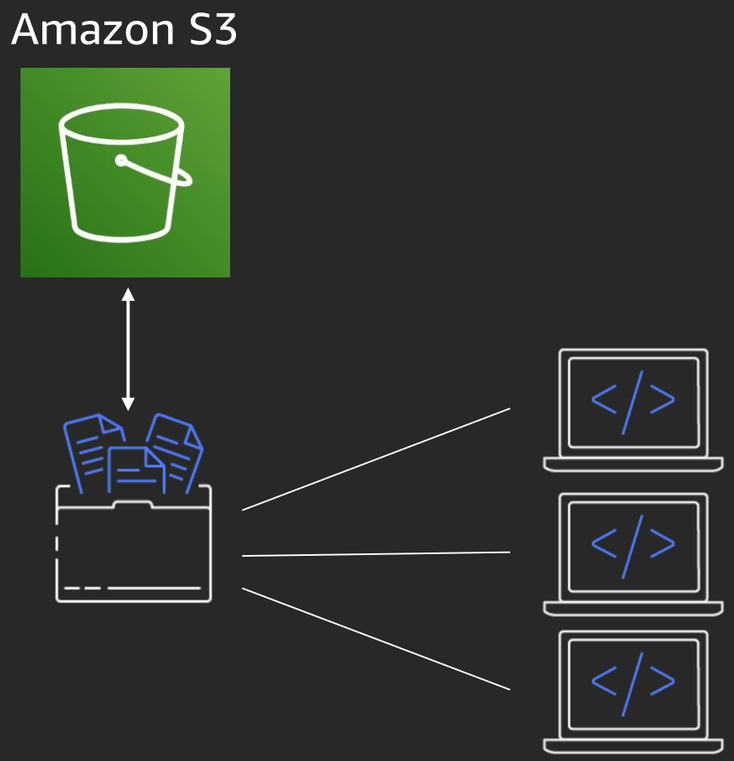


Plan active working set – You need to plan which data to move in/out S3 ahead of time, or update application logic to orchestrate the movement in real-time

Sharding your data set – You need to spread your active data set across instances or volumes, may need application rework

Data duplication – You may end up placing duplicate data across instances or volumes, depending on which instances need access to which pieces of the data

Data processing with self-managed file systems



Complex to manage and maintain

Cumbersome performance tuning

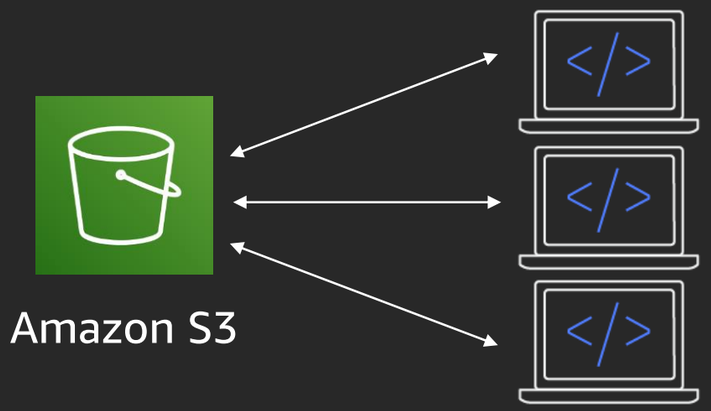
Plan active working set

Tracking changes and writing back to S3

Track what has changed

Write application or scripting logic to periodically write changes and results back to S3

Data processing without intermediate storage



Appropriate for applications that …

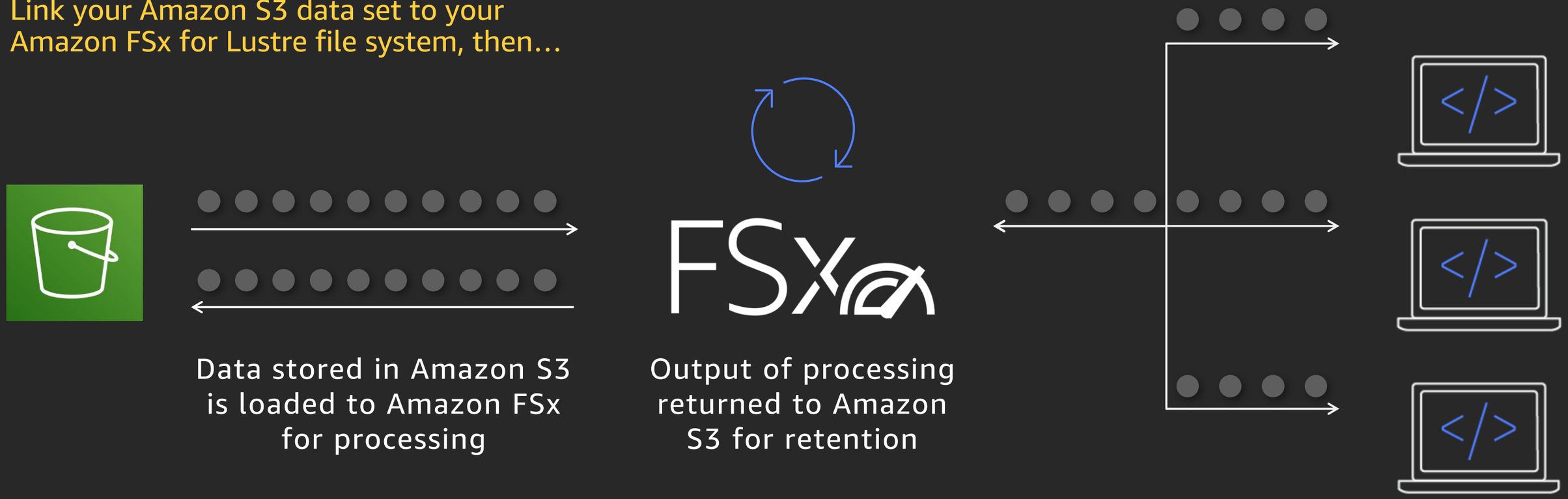
need high throughput, not latency-sensitive

use object storage interface, not POSIX

do not access the same data repeatedly

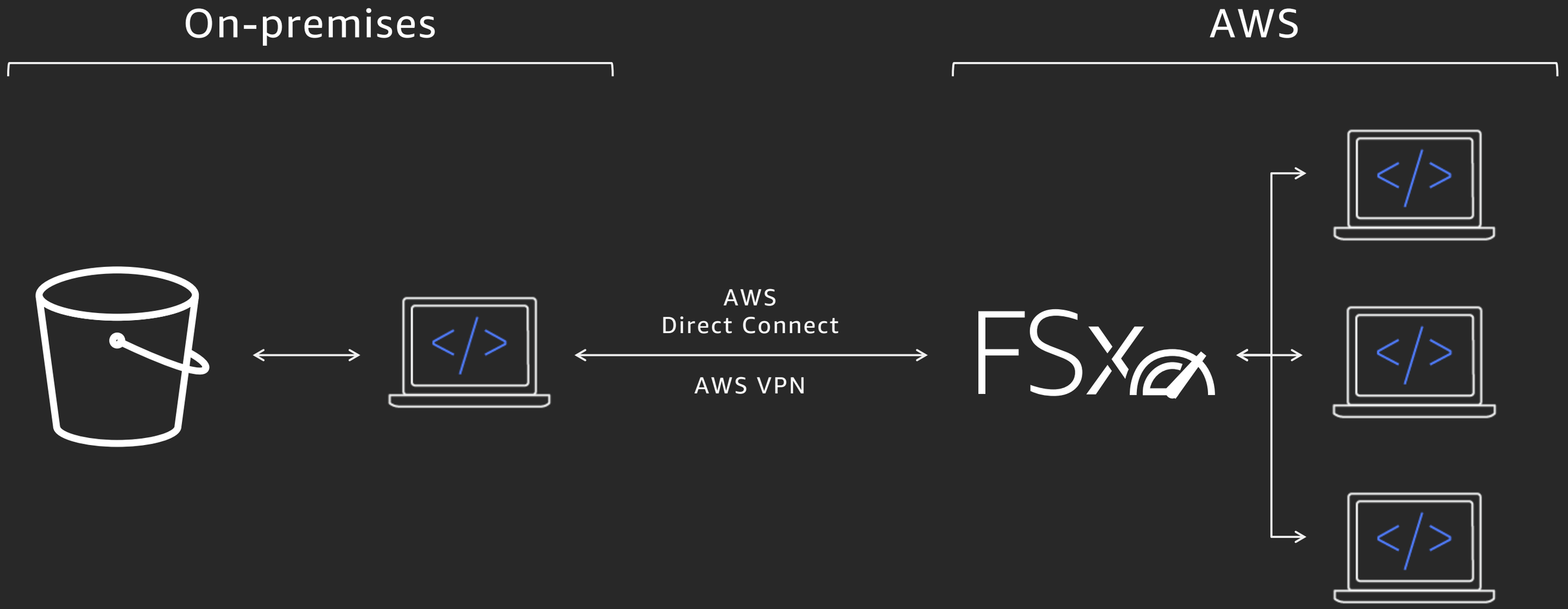
Amazon FSx for Lustre is designed for these data processing workflows in the cloud

Link your Amazon S3 data set to your Amazon FSx for Lustre file system, then…



When your workload finishes, simply delete your file system, or keep it running for long-lived workloads

Amazon FSx for Lustre also supports cloud bursting for on-prem data repos



Amazon FSx for Lustre: Making your compute workloads faster and cheaper

Fast processing with 100+ GB/s throughput & sub-millisecond latencies

Tight integration with S3

Flexible data processing options for short and longer-term

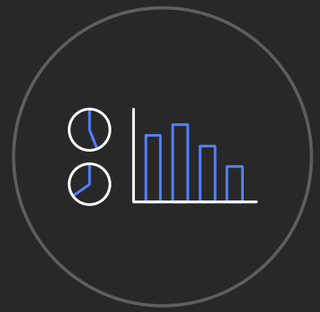
Lower TCO

Example workloads

High-performance computing

Machine learning

Media rendering and transcoding



Big data analytics

Electronic design automation

Financial modeling

Oil and gas seismic processing

Autonomous systems training

Accessible from popular Linux distributions

Amazon Linux 2

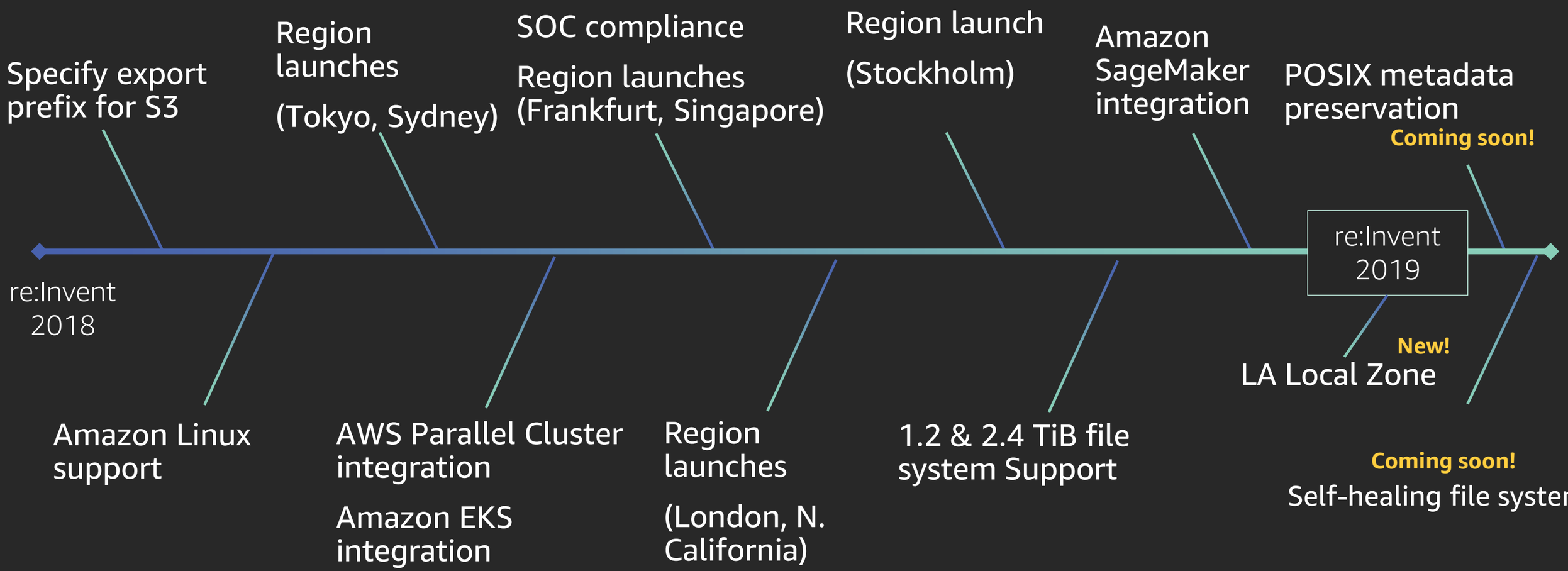
Red Hat Enterprise Linux

CentOS

Ubuntu

SuSE Linux Enterprise

Amazon FSx for Lustre pace of innovation



Amazon S3 + Amazon FSx

Amazon S3 and Amazon FSx for Lustre

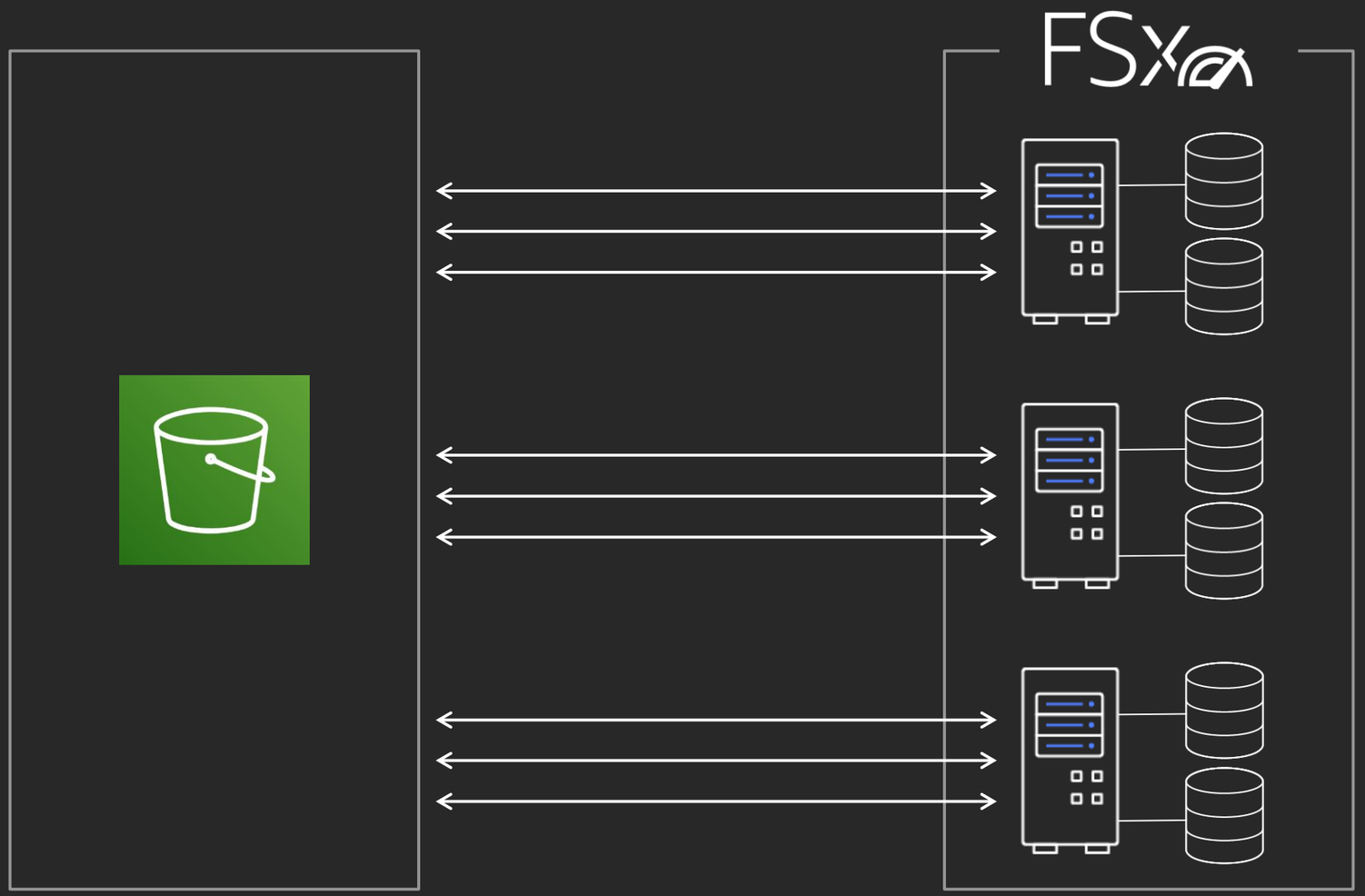
Specify an Amazon S3 bucket when you create your file system

On accessing the file system, you see all objects as files/directories

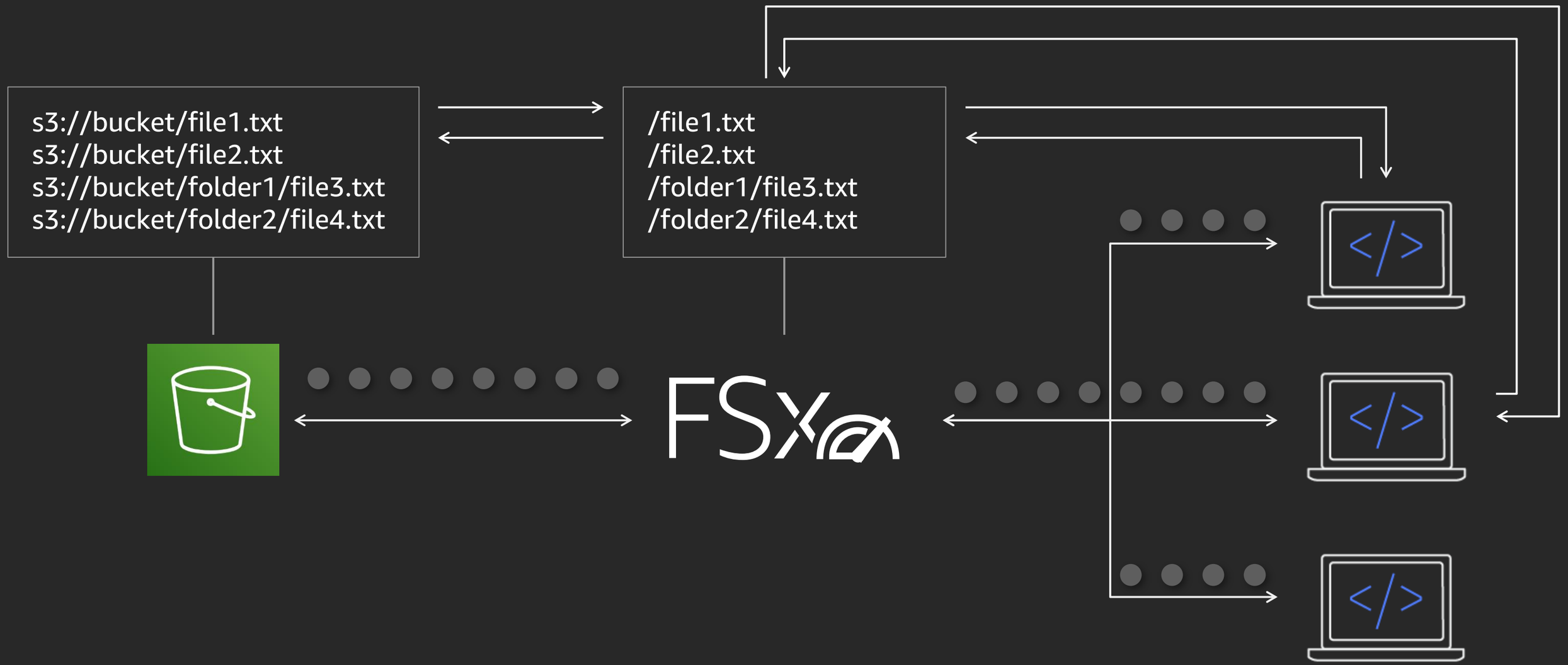
Files are moved in real time, when accessed, to file system

Use commands to write results to Amazon S3

Amazon S3 integration is performance-optimized for fast data & metadata movement



Amazon S3 lazy load example



Hierarchical Storage Management (HSM) commands for data movement

hsm_archive – Copy files to Amazon S3 from FSx for Lustre

hsm_release – Free disk space associated with files, once archived

hsm_restore – Bring back file data to FSx for Lustre from Amazon S3
(also done automatically when accessing a file for the first time)

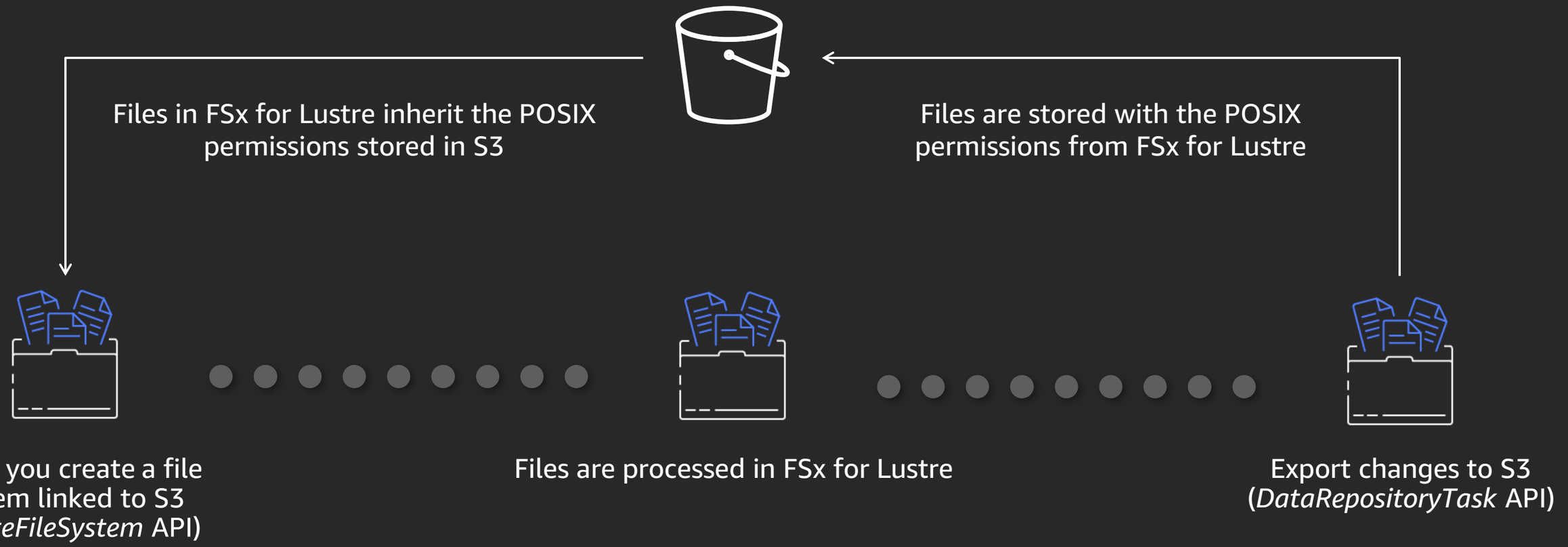
DataRepositoryTask API - Export changes to S3

Coming soon

Simplify exporting new and changed files/directories

Synchronize file/folder permissions and other metadata with S3

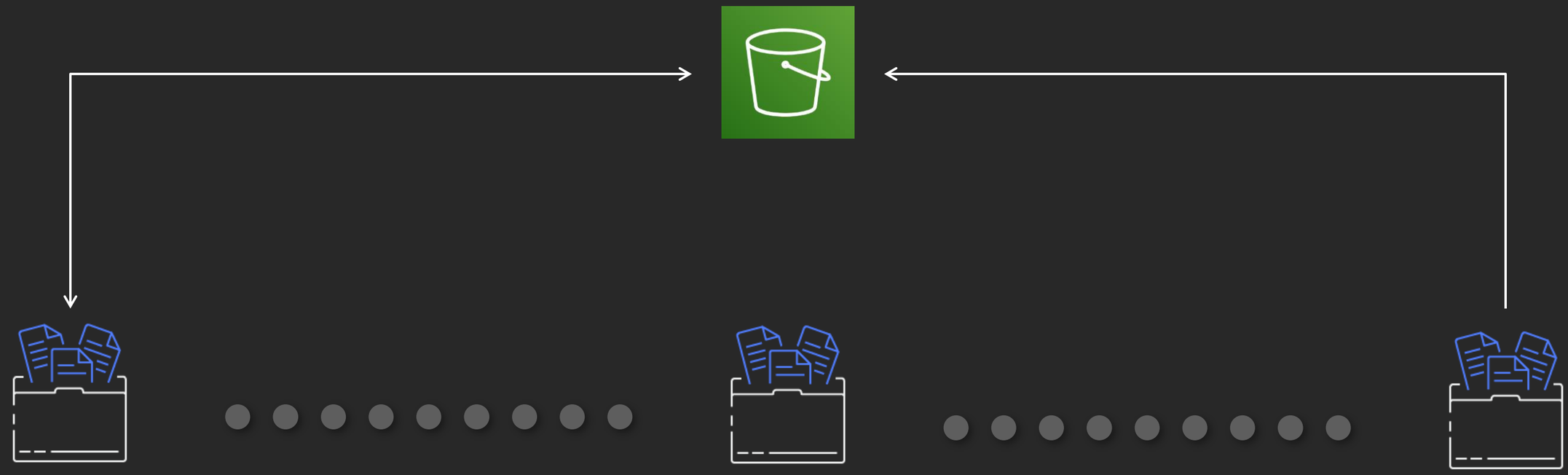
Preserve POSIX metadata across Amazon FSx and S3



POSIX metadata mapping is consistent with AWS DataSync and AWS File Gateway

POSIX Metadata	Description	Amazon S3 user-defined metadata entry
File Type	Object user-defined file permissions	x-amz-meta-file-permissions
Permissions	Object user-defined file permissions	x-amz-meta-file-permissions
User ID	Integer value of file owner uid	x-amz-meta-file-owner
Group ID	Integer value of file owner gid	x-amz-meta-file-group
Modification Time	Last modified time in nanoseconds	x-amz-meta-file-mtime
Access Time	Last accessed time in nanoseconds	x-amz-meta-file-atime
User Agent	Ignored during export, FSx sets this to “aws-fsx-lustre”	x-amz-meta-user-agent

Release inactive data sets to S3 to free up space

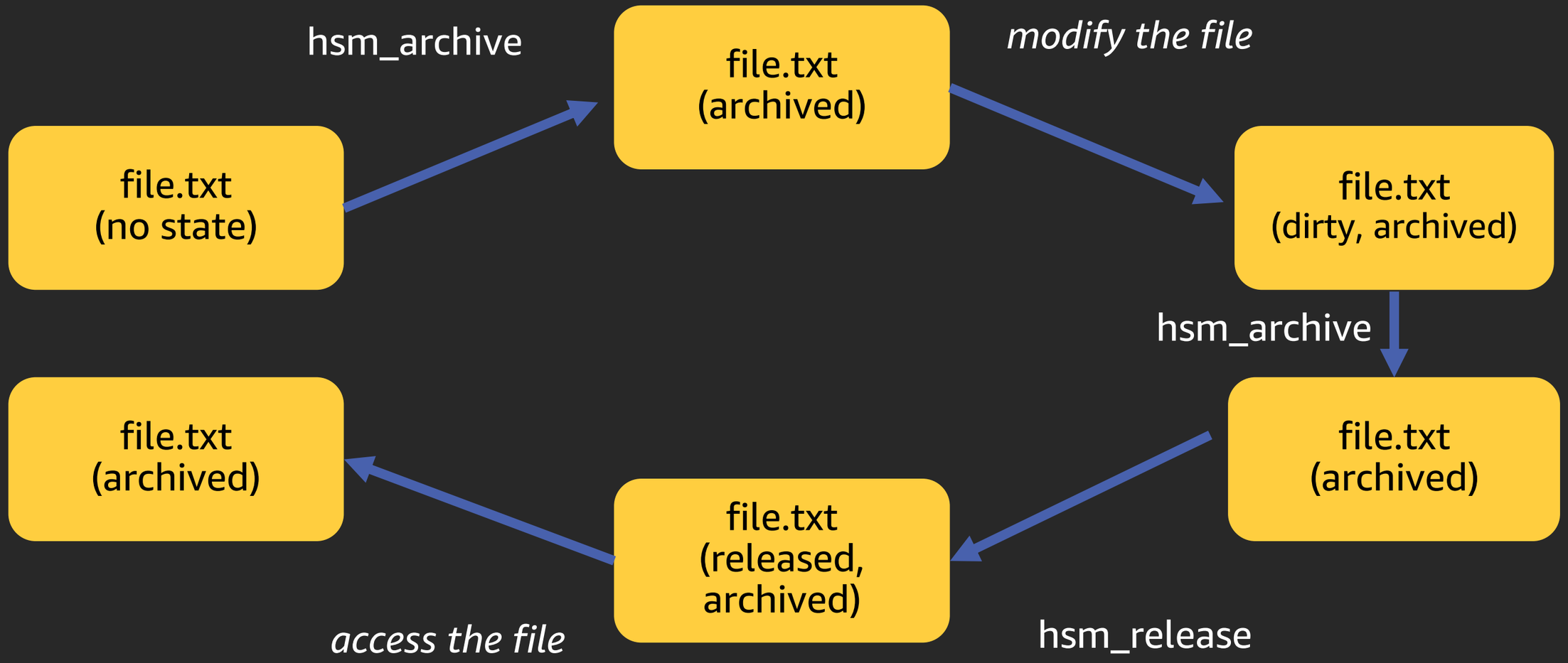


Create myfs1 linked to S3 bucket s3://mybucket and mount myfs1

Files are processed in FSx for Lustre

hsm_release files from myfs1 to S3://mybucket

The life cycle of a file



Amazon SageMaker integration

Amazon SageMaker

Amazon FSx for Lustre

Amazon S3

FSx for Lustre can be a file input data source for Amazon SageMaker

Eliminates initial S3 download step – Accelerates your training jobs

Avoid repeated download of common objects for iterative jobs on same data set

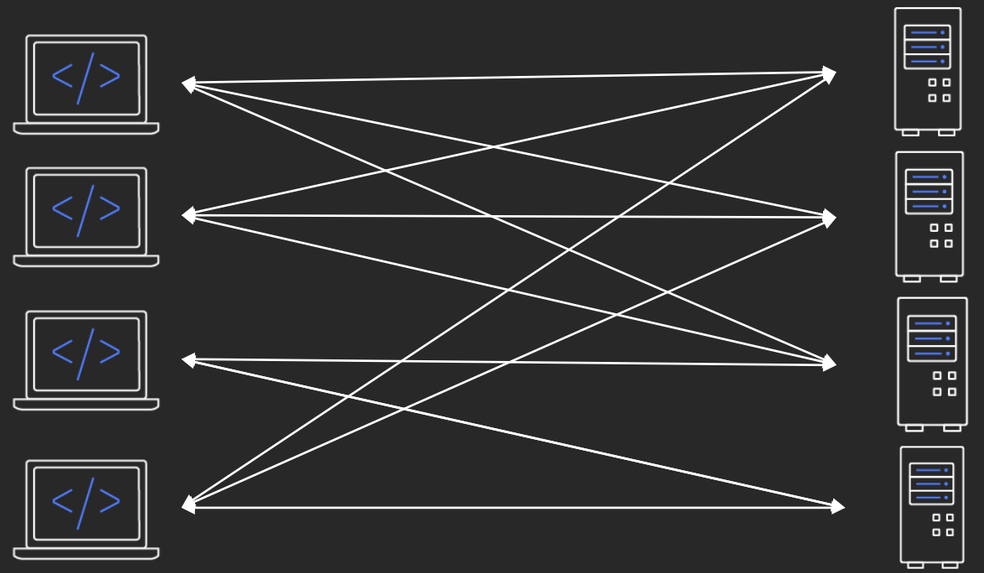
Data processing options

Reminder: How a parallel file system works

Parallel file systems store data across multiple network file servers to maximize performance and reduce bottlenecks

Clients interface directly with the servers hosting a set of data

Servers have multiple disks, data striped across disks and servers

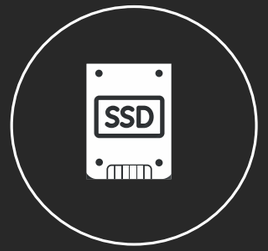


Common deployment model: S3 is long-term durable storage, FSx for Lustre is used when processing data

Use S3 as the highly durable long-term store for your data, with Amazon FSx as a high-performance file system linked to your S3 bucket

1. *Store your data set on S3*
2. *Create an FSx file system and link it to your S3 bucket*
3. *At any time, use a Lustre command to write changes back to S3*
4. *Delete your FSx file system when you are done processing*

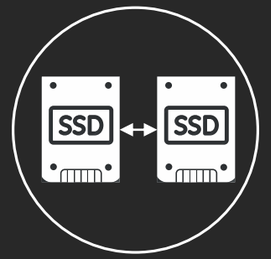
FSx for Lustre deployment options



Scratch

Short-term,
spin up, process, spin down

Coming Soon!



Self-healing

Longer-term processing, file
servers are HA, data is replicated

In both options, the Amazon FSx Control Plane (API, management layer, file system control) is designed to be highly available

What happens if a data server becomes unavailable on a scratch file system?

Workload can still continue, if designed for this scenario

Clients trying to access data on the unavailable server will get an immediate I/O error

Availability that a scratch file system is designed for

Probability of no servers permanently losing availability/durability, based on size and duration

	10 TiB	50 TiB	100 TiB
One day	99.9%	99.4%	98.8%
One week	98.9%	95.9%	92.1%

What's the behavior of a self-healing file system?

If a file server becomes unavailable on a self-healing file system, it is replaced automatically

Client requests for data on that server transparently retry, will eventually succeed

Data volumes are replicated independently from the file servers to which they are attached, with each volume designed for five 9s of durability

Performance

Conductor Technologies accelerates rendering workloads by up to 4X using Amazon FSx for Lustre

The Challenge

Conductor Technologies was faced with scaling and efficiency issues using file systems from their previous cloud provider that led to increased render times.

The Solution

Using FSx for Lustre, a fully-managed, high performance file system, the company was able to reduce their render times by up to 4X, reduce spin-up time by 30%.

The Benefits

- 30% reduction in spin-up and runtimes over traditional methods
- Improved performance by as much as 4X (30 min vs. 2 hours)

> "We chose Amazon FSx for Lustre to supercharge our VFX rendering workloads in the cloud. We reduced spin up times by 30% and accelerated run-times up to 4X, eliminating weeks and months of resources building and managing file servers."

Mac Moore, CEO, Conductor Technologies

Company: Conductor Technologies
Industry: Media & Entertainment
Country: United States
Employees: 11-50
Website: https://www.conductortech.com/

About Conductor Technologies

Conductor Technologies is a leading cloud platform for Media & Entertainment for rendering, simulation, virtual reality. Conductor enables VFX and Animation facilities to extend into the cloud, while providing data insights and controls over usage and spending.

High and scalable performance

High and scalable performance

Parallel file system

100+ GiB/s throughput
Millions of IOPS
Consistent sub-millisecond latencies

SSD-based

Supports concurrent access from hundreds of thousands of cores

Per-client throughput = EC2 instance network throughput

Scratch file system performance

Each TB of storage provides
200 MB/s of baseline throughput,
and up to 2x burst throughput

File systems can scale
to hundreds of GB/s and
millions of IOPS

Capacity	Baseline throughput	Burst throughput
1TB	200 MB/s	up to 400 MB/s
10TB	2 GB/s	up to 4 GB/s
50TB	10 GB/s	up to 20 GB/s
100TB	20 GB/s	up to 40 GB/s
1PB	200 GB/s	Up to 400 GB/s

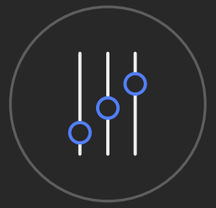
Optimizing I/O performance on FSx for Lustre

Best practices for striping file system data

Stripe files to optimize I/O performance when concurrent access is common

Average I/O size

Throughput increases with higher average I/O size



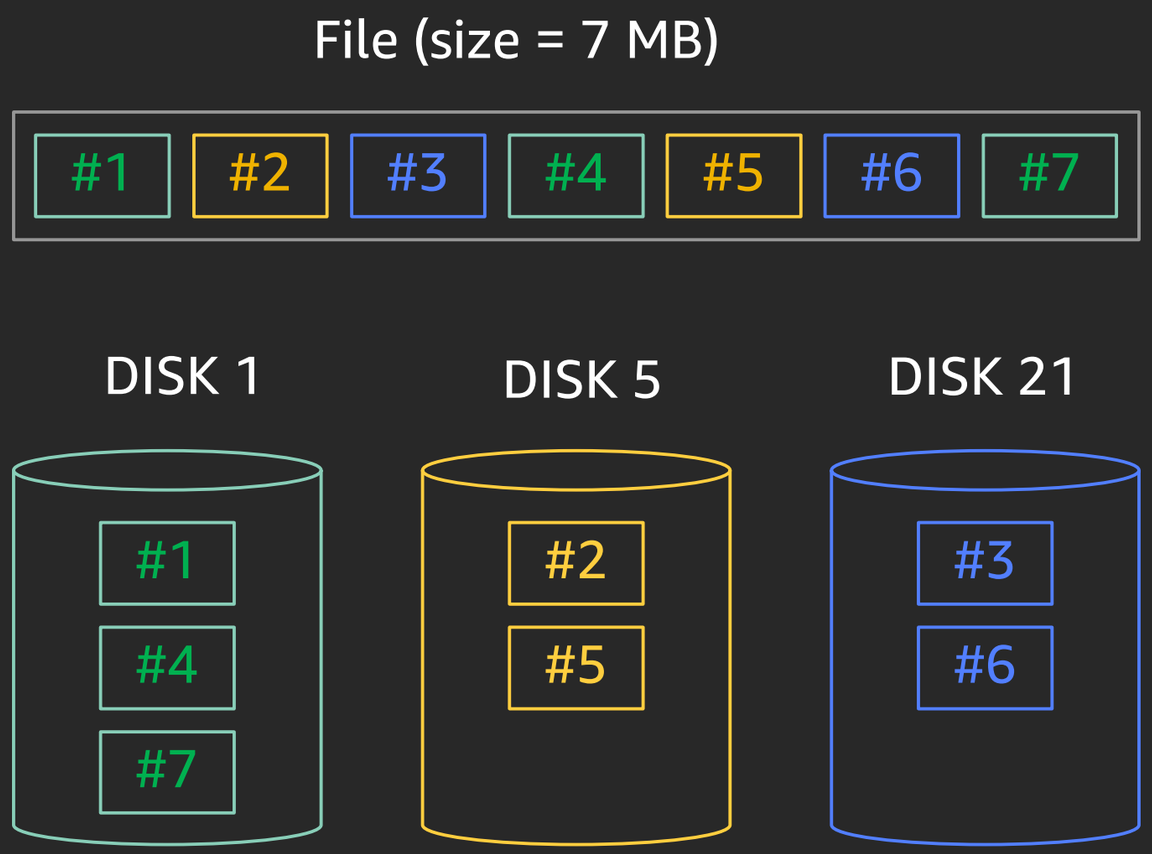
Client selection

Choose EC2 instance type with enough memory, CPU, and bandwidth

What is striping, why use it?

- Striping refers to sharding large files in to fragments and storing them across disks in multiple servers
- It allows you parallelize access to individual files, driving higher aggregate throughput
- By default each file is stored in one disk
- Striping can be set per directory or per file
- All files in a directory inherit it’s striping parameters

How striping works in FSx for Lustre



Specify stripe_count and stripe_size (lfs setstripe)

Striping can be set per directory or per file, all files in a directory inherit it's striping parameters

Stripe files across disks based on CloudWatch Max metric

Set ImportedFileChunkSize = (dominant file size / # of disks)

Total cost of ownership (TCO)

T-Mobile realizes $1.5M in annual savings and doubles the speed of SAS Grid workloads using Amazon FSx for Lustre

The challenge

T-Mobile managed their own storage for SAS Grid, which proved to be unscalable and cost prohibitive

The solution

T-Mobile deployed Amazon FSx for Lustre, a fully-managed high-performance file system, to migrate and scale their SAS Grid infrastructure

The benefits

Reduced TCO by 83% and reduced storage costs by 67%, resulting in $1.5M in annual cost savings

Cut the run time of SAS Grid analytics workloads by half

"Amazon FSx for Lustre helped us double the speed of our SAS Grid workloads, reduce our Total Cost of Ownership by 83% and completely eliminate our operational burden"

Dinesh Korde, Sr. Manager Software Development, T-Mobile

aws

Amazon FSx for Lustre is cost-optimized for data processing

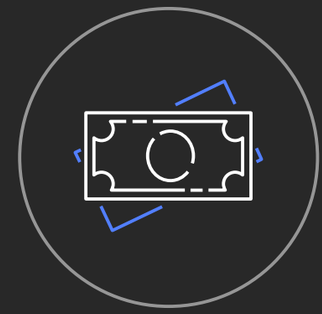
Cost-effective

Optimized for processing datasets at high performance and low cost

Process only a portion of your data set at one time

Long term data stored in low-cost Amazon S3 or on-premises

Spin up, spin down with scratch file systems or release with self-healing



No request costs for repeatedly accessed data

FSx for Lustre pricing (high-perf SSD): $0.14 per GB-month ($0.20 per TB-hour)

TCO example

Total data set: 250 terabytes

You run a daily job to process the last days-worth of data

An individual job

Processes 25 terabytes of data

Requires 5 GB/s of throughput

Runs for 10 hours each day

TCO example with direct processing on Amazon S3

S3 costs for processing 25 TB/job daily for a month. Assuming 1 MB per object, there will be 25 million GET and PUT requests per day, provided all the objects are read and written back once every day. Assuming 20 compute instances repeatedly accessing the same data per job run.

Storage on S3 (250 TB-month):	$5,550 per month
S3 request costs:	$81,000 per month
Total:	$86,550 per month ($0.346 per GB-month)

TCO example with Amazon S3 + FSx for Lustre

Amazon FSx for Lustre:

($0.20/TB-hour) * (25 TB/job) * (10 hours/job) * (30 jobs/month)= $1,460 per month

Storage on S3 (250 TB-month):	$5,550 per month
Amazon FSx for Lustre for active data:	$1,460 per month
S3 request costs	$7,800 per month
Total:	$14,810 per month ($0.059 per GB-month)

Amazon FSx availability

More coming soon!

US

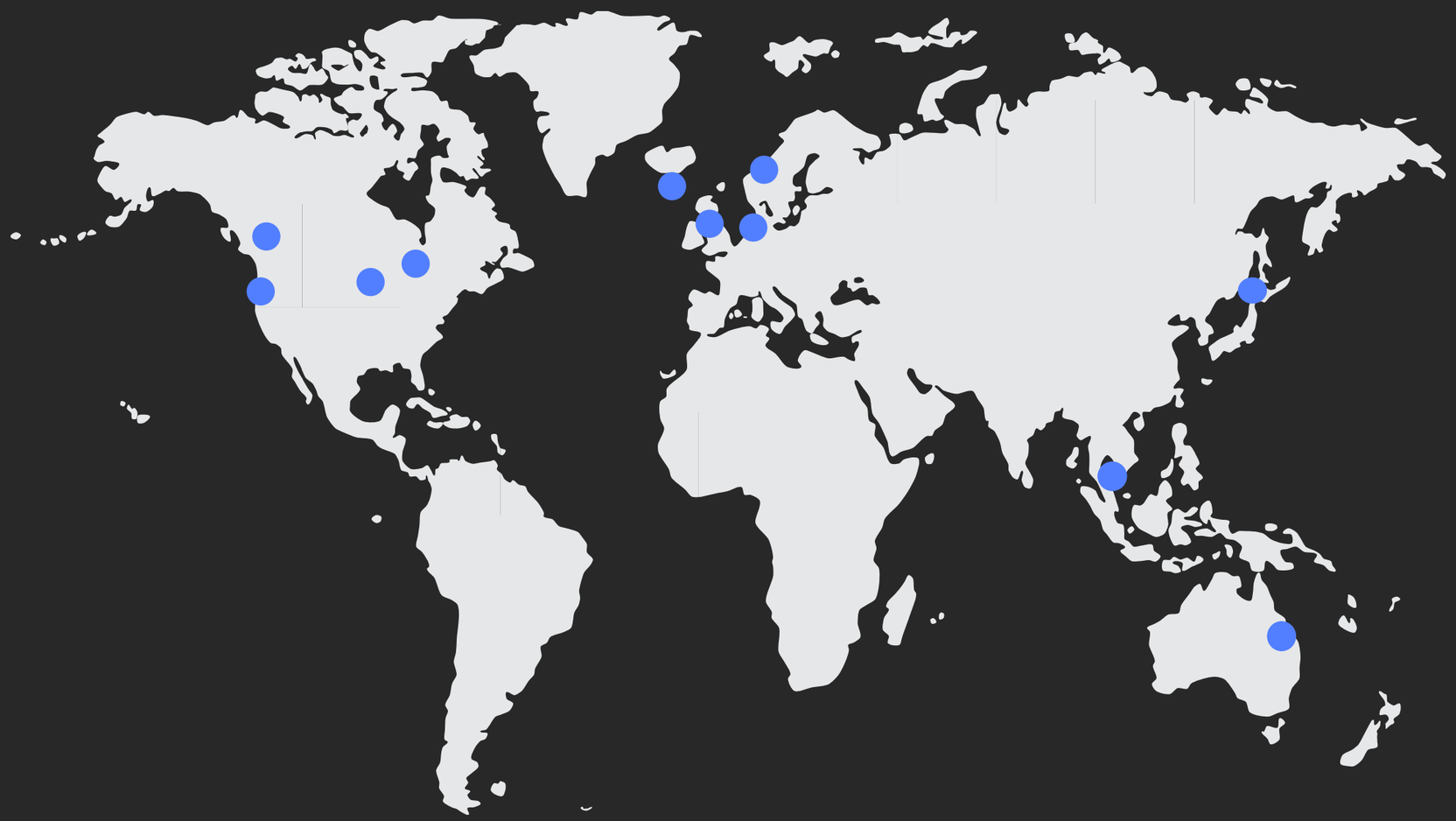
US West (Oregon)
US West (N. California)
US East (N. Virginia)
US East (Ohio)

EU

Europe (Ireland)
Europe (Frankfurt
Europe (London)
Europe (Stockholm)

APAC

Asia Pacific (Sydney)
Asia Pacific (Singapore)
Asia Pacific (Tokyo)



Related breakouts

STG 201
AWS leadership session: Storage state of the union

STG 202
What is new with the AWS file storage portfolio

STG 348
Optimize HPC workload storage using FSx for Lustre

STG 306
Deep dive on Amazon FSx for Windows File Server

STG 323R-1
Amazon FSx for Lustre—a High Performance File System integrated with S3

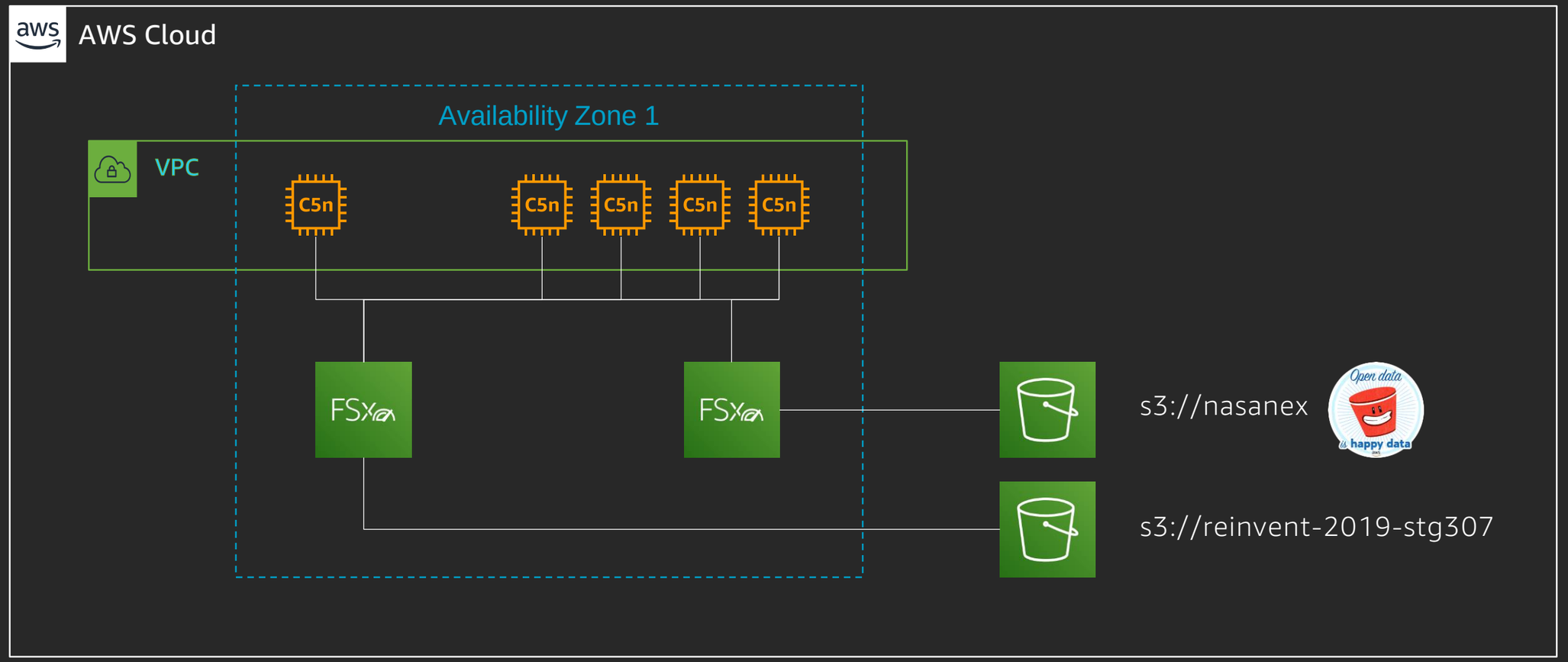
STG 347
Choosing the Right Storage for Your High-Performance Workloads

STG 349
Optimize Video Processing using FSx for Lustre (and Thinkbox)

HPC Meet-Up and Networking Reception (Dec 2, Wynn, 5 PM–7 PM)

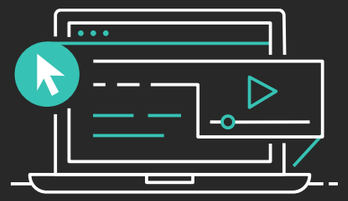
Amazon FSx for Lustre in action

Demo environment



Learn storage with AWS Training and Certification

Resources created by the experts at AWS to help you build cloud storage skills



45+ free digital courses cover topics related to cloud storage, including:

- Amazon S3
- AWS Storage Gateway
- Amazon S3 Glacier
- Amazon Elastic File System (Amazon EFS)
- Amazon Elastic Block Store (Amazon EBS)

Classroom offerings, like Architecting on AWS, feature AWS expert instructors and hands-on activities

Visit aws.amazon.com/training/path-storage/

aws training and certification

Thank you!

AWS re:Invent

aws

Please complete the session survey in the mobile app.