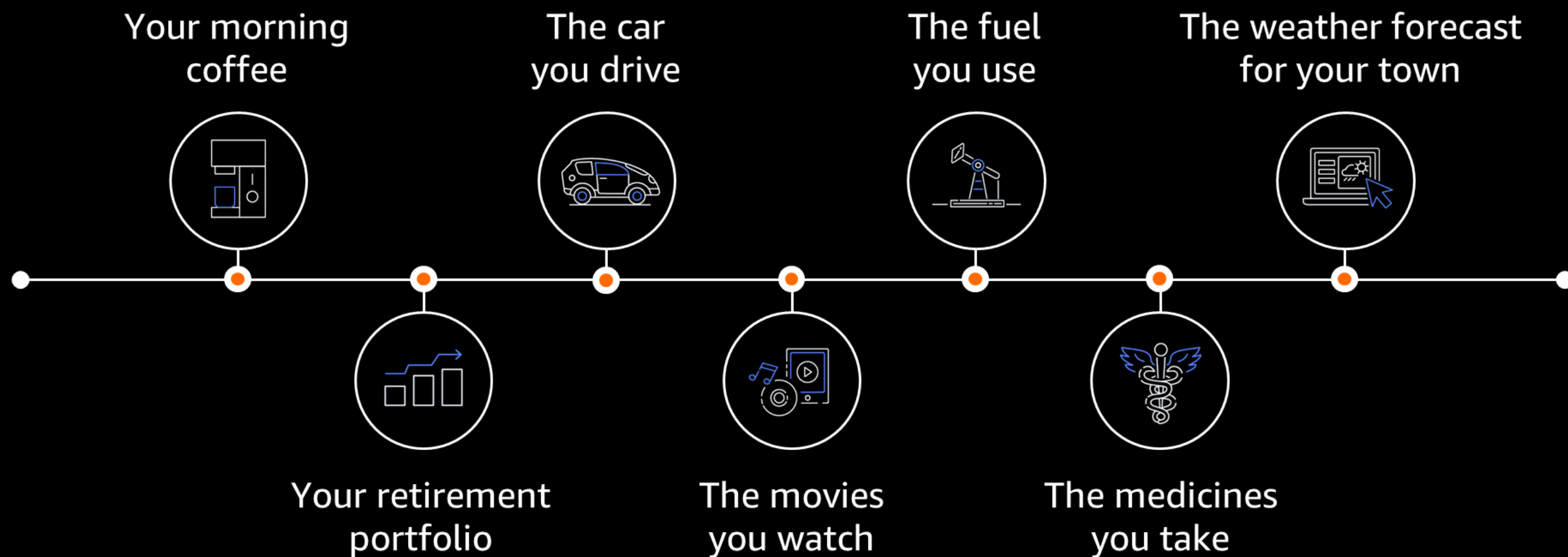


Accelerate innovation and ML workloads using your data with Amazon FSx for Lustre and Amazon S3

Wali Akbari
Senior APJ Storage Solutions Architect
Amazon Web Services

ML and Analytics impacts your life every day



Machine Learning

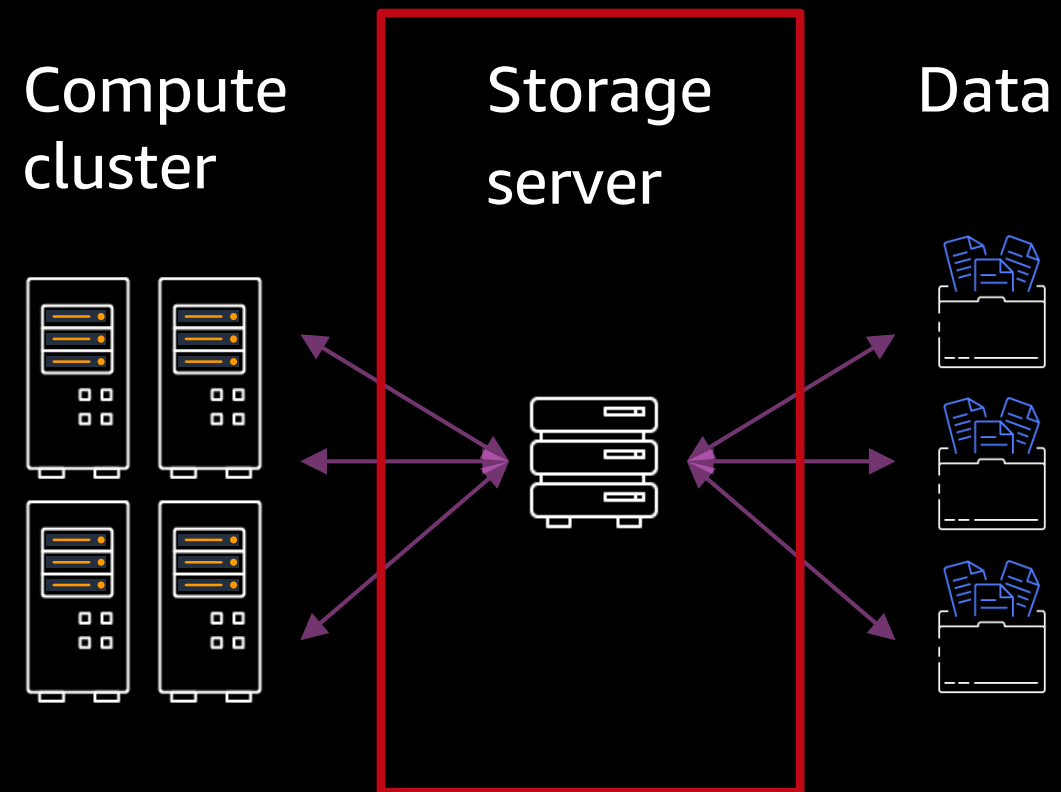
- Collect and prepare training data
- Choose and optimize ML algorithms
- Setup and manage resources
- Train and tune model
- Deploy model in production
- Scale and manage

Agenda

- Overview of Amazon FSx for Lustre
- Benefits for Machine Learning and analytics
- Integration with Amazon SageMaker, Amazon EC2 and Amazon Simple Storage Service (Amazon S3)
- Demo
- Recap

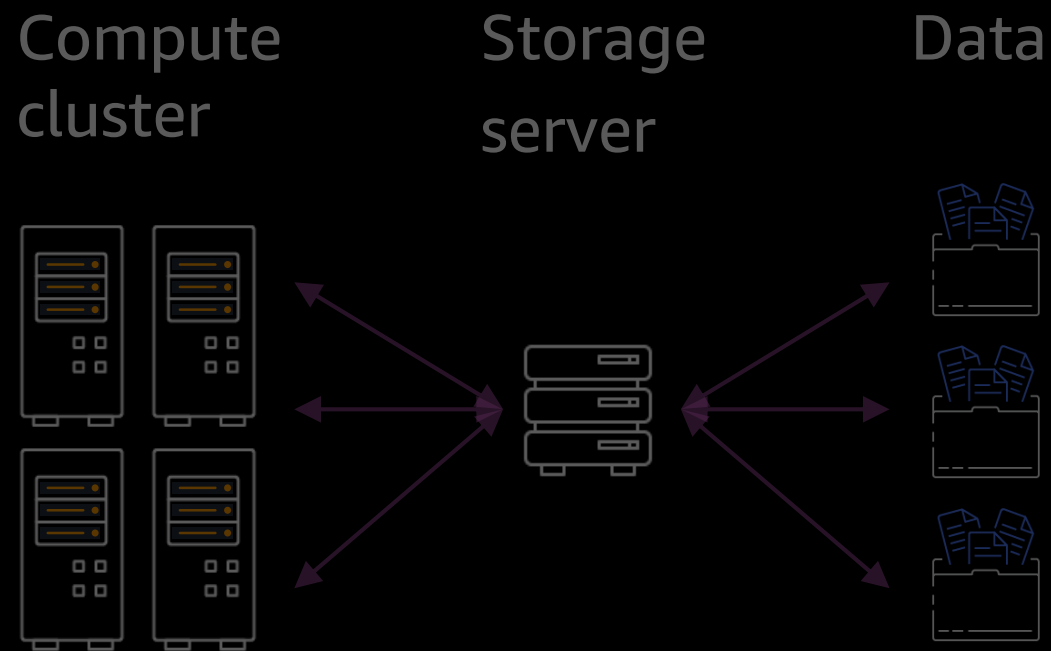
Why high performance storage matters

Compute cluster with storage bottleneck

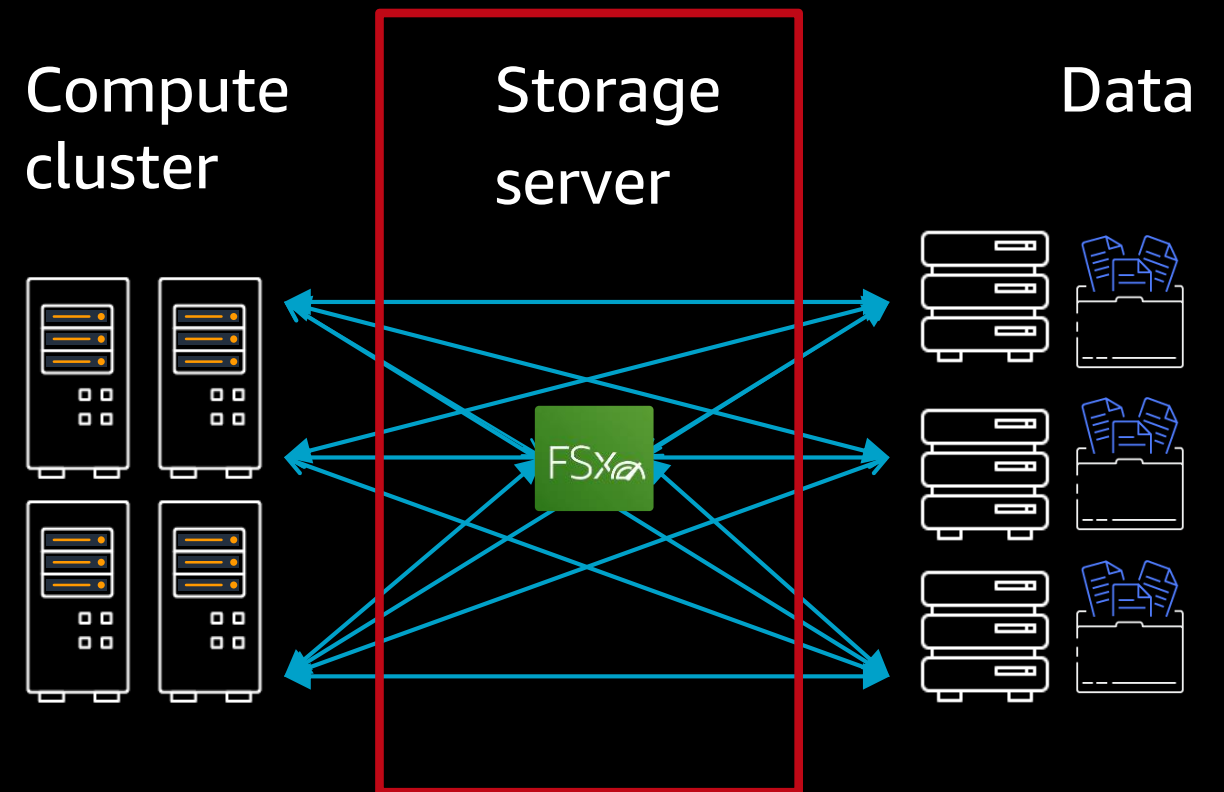


Why high performance storage matters

Compute cluster with storage bottleneck



Compute cluster with **FSx parallel file system**



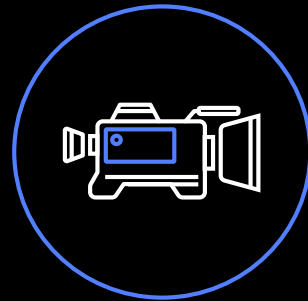
Organizations have lots of data



Financial services:
Modeling and
analytics



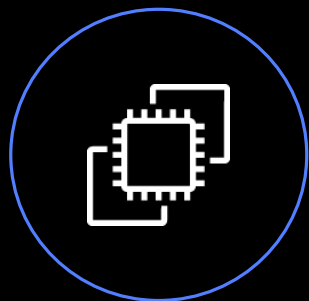
Life sciences:
Genome analysis



Media and
entertainment:
Rendering and
transcoding



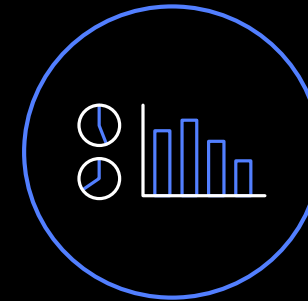
Automotive:
ECU simulations
and object detection



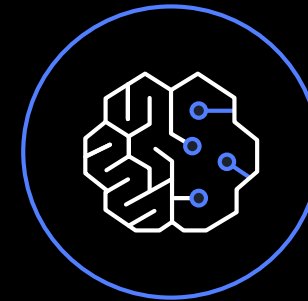
Semiconductor:
Electronic design
automation



Oil and gas:
Seismic data
processing



Big data
analytics



Machine
learning



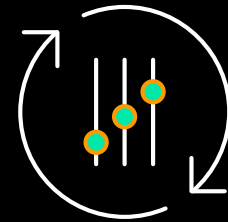
High-performance
computing

Amazon FSx for Lustre

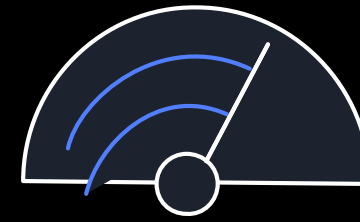


Amazon FSx
for Lustre

High-performance
Workloads



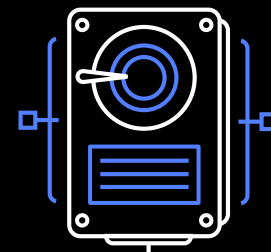
Fully
managed



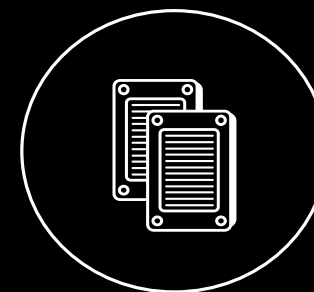
Very high
performance



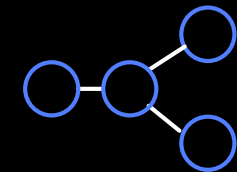
Amazon S3 data
sets as POSIX
file system



Cost
optimized



Scratch or
Persistent
mode



Access from
on-premises &
AWS services



Amazon FSx for Lustre



Simple



Fast



Cost effective

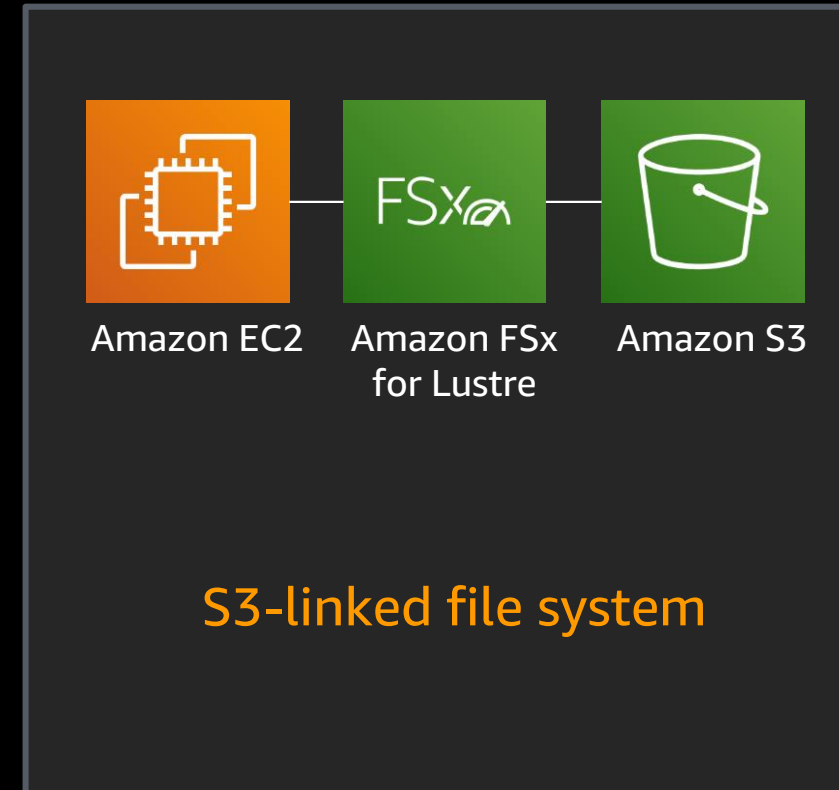
Provides shared storage to power your workloads

Simple



- 1. Deploy** Amazon FSx for Lustre file system:
Standalone file system or link it with your Amazon S3 bucket
- 2. Integrate** Amazon FSx for Lustre with your workloads:
Amazon SageMaker – select it as file system data-source
Mount it as a file share on your compute fleet nodes, such as Amazon EC2 instances
- 3. Export** your new and changed data:
To your Amazon S3 bucket – using the native Amazon FSx for Lustre DRT export integration
To your desired location – using your own application and process

Amazon FSx for Lustre : Standalone or Amazon S3-linked file system



Amazon SageMaker with Amazon FSx for Lustre



1. Navigate to training jobs

Amazon SageMaker > Training jobs

Training jobs

1

Name	Creation time	Duration	Status
There are currently no resources.			

2. Select file system as data source

Data source

S3

File system

File system ID

Type

Access mode

Directory path



Compute fleet nodes with Amazon FSx for Lustre

1. Mount Amazon FSx for Lustre file system on compute node

```
[root@ip-10-10-10-10 ~]# mount -t lustre -o noatime,flock fs-0245f411c5691760e.fsx.ap-southeast-2.amazonaws.com@tcp:/sdtwlbmv /fsx
```

2. View Amazon FSx for Lustre file system

```
[root@ip-10-10-10-10 ~]# df -h
```

Filesystem	Size	Used	Avail	Use%	Mounted on
devtmpfs	7.6G	0	7.6G	0%	/dev
tmpfs	7.7G	0	7.7G	0%	/dev/shm
tmpfs	7.7G	380K	7.7G	1%	/run
tmpfs	7.7G	0	7.7G	0%	/sys/fs/cgroup
/dev/nvme0n1p1	30G	1.8G	29G	6%	/
tmpfs	1.6G	0	1.6G	0%	/run/user/1000
172.31.10.10@tcp:/sdtwlbmv	6.6T	28M	6.6T	1%	/fsx



Toyota Research Institute chooses Amazon FSx for Lustre to **reduce object recognition machine learning training times**

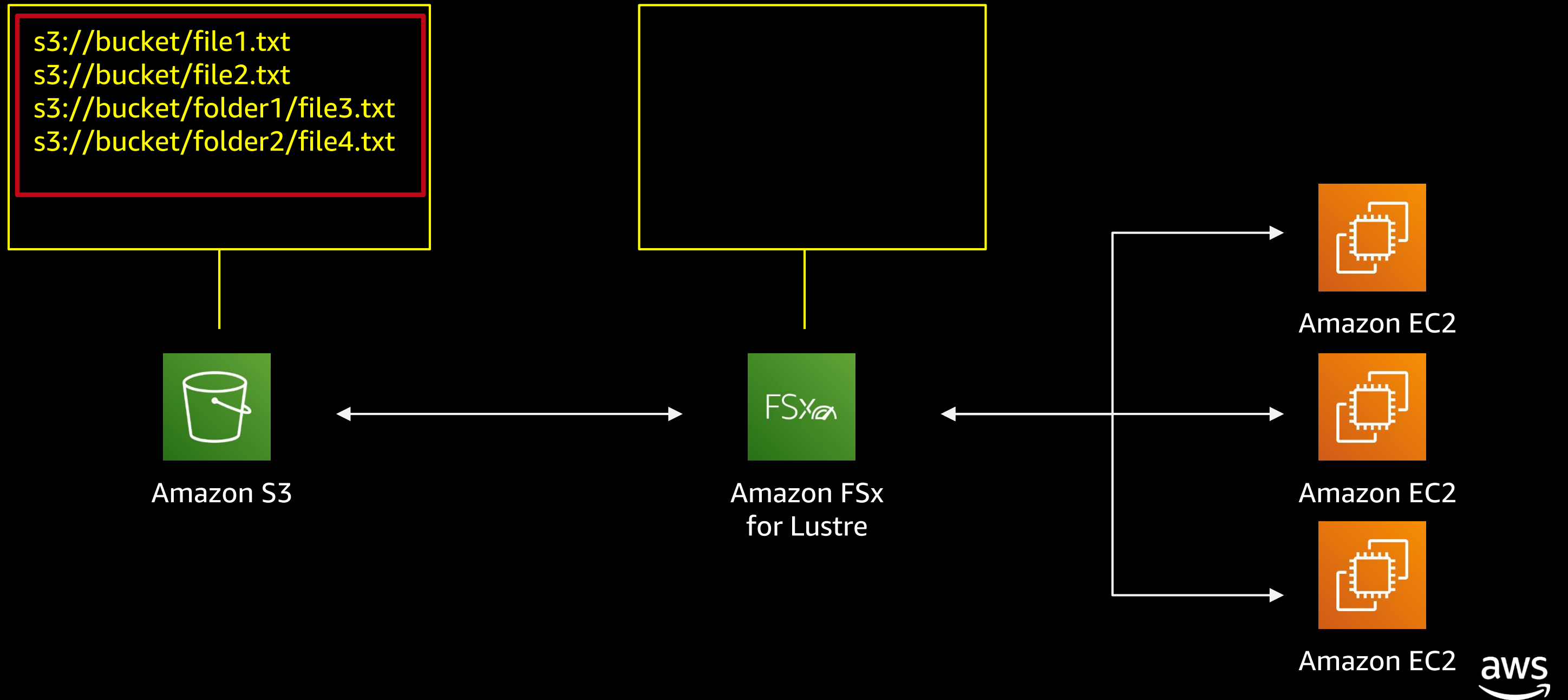
“We needed a parallel file system for our ML training data sets and chose Amazon FSx for Lustre . . . The **integration with AWS services, including S3**, also made it the preferred option for our high performance file storage.”

David Fluck
Software Engineer
Toyota Research Institute



Amazon S3-linked file systems

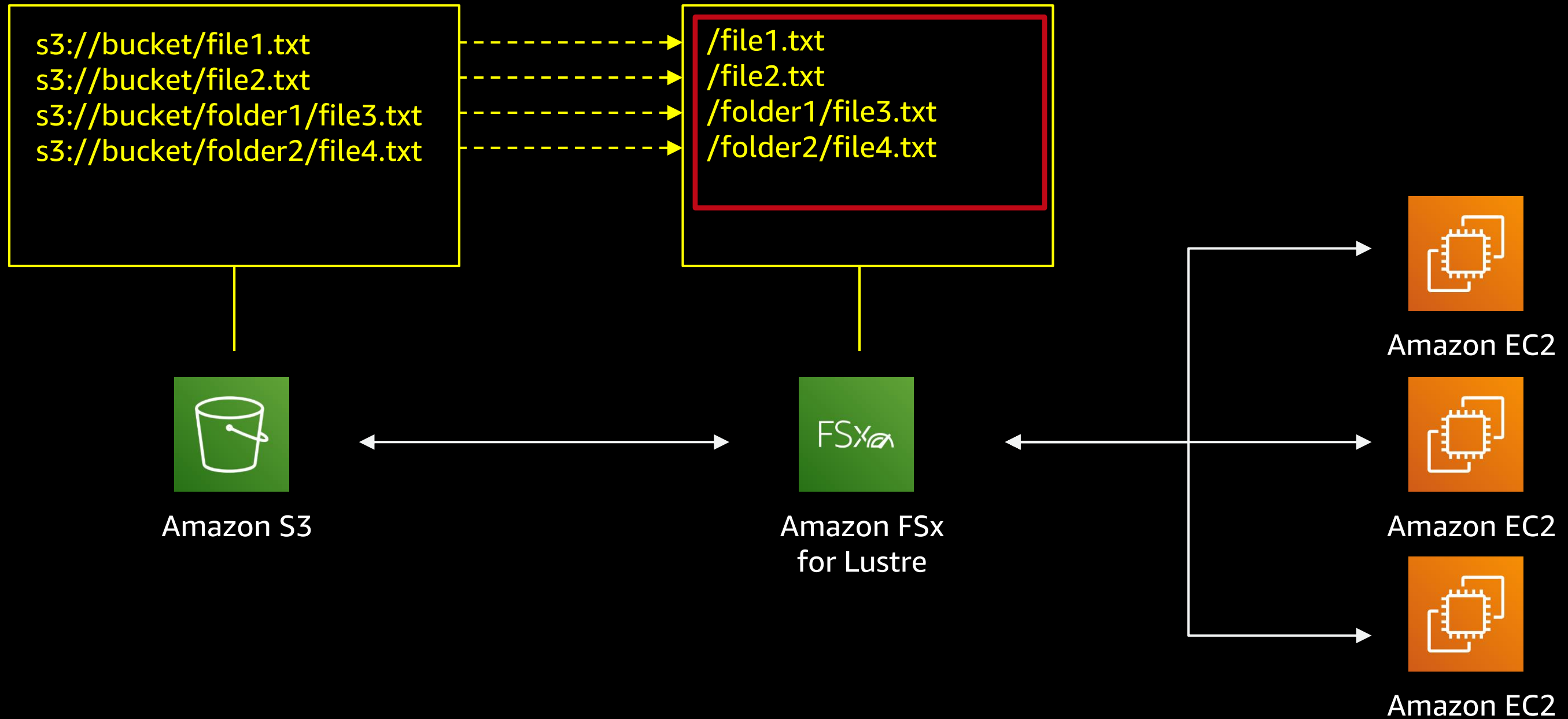
Objects stored in Amazon S3 appear as files on Amazon FSx file systems





Amazon S3-linked file systems

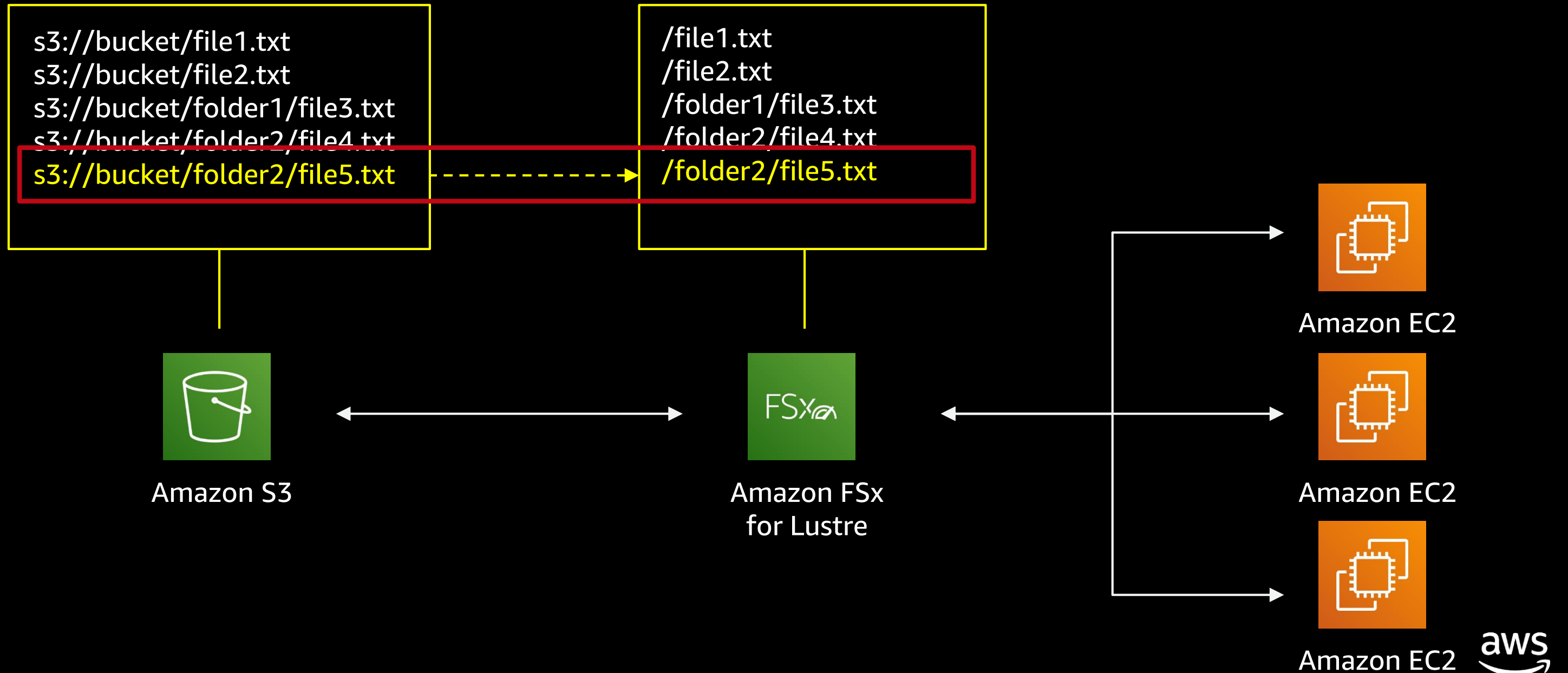
Objects stored in Amazon S3 appear as files on Amazon FSx file systems





Amazon S3-linked file systems

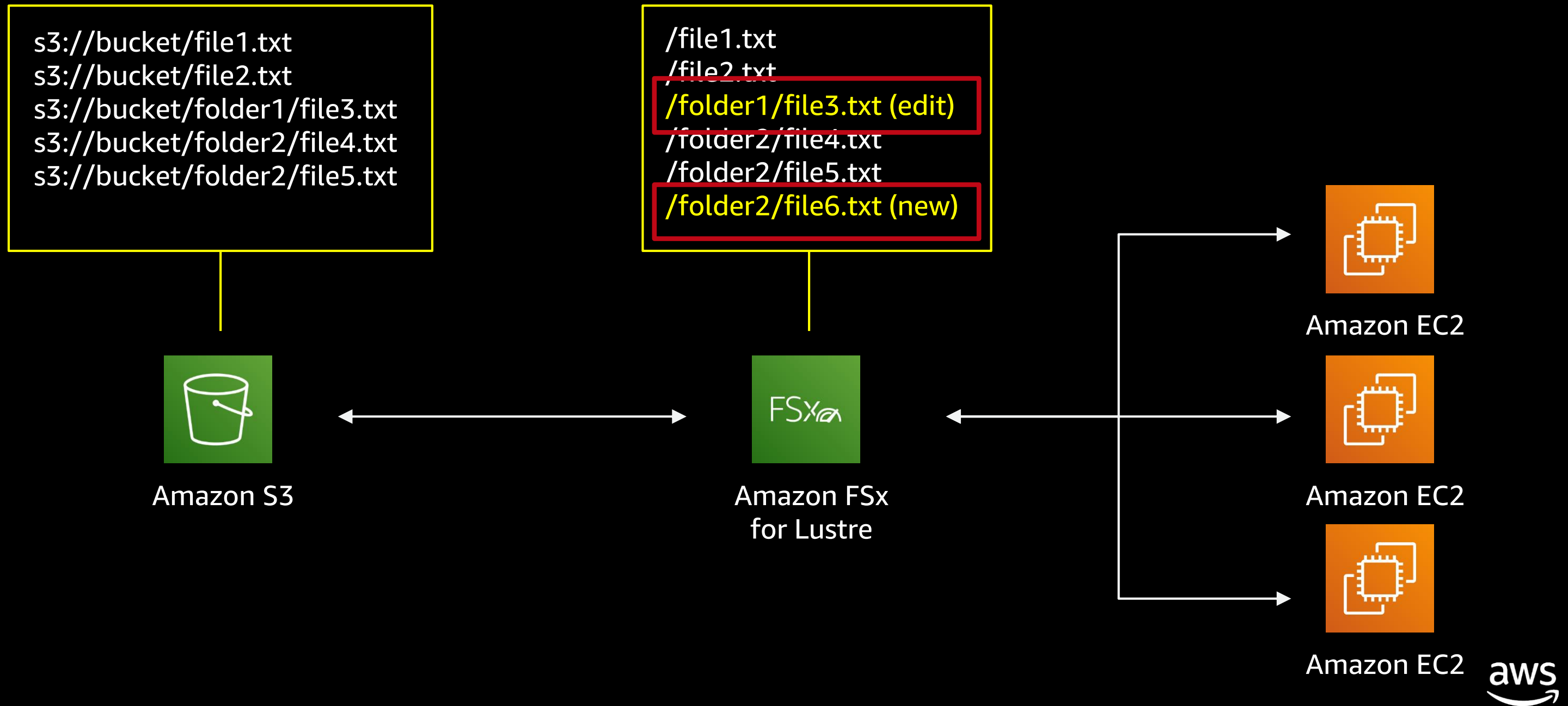
Objects stored in Amazon S3 appear as files on Amazon FSx file systems

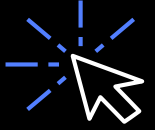




Amazon S3-linked file systems

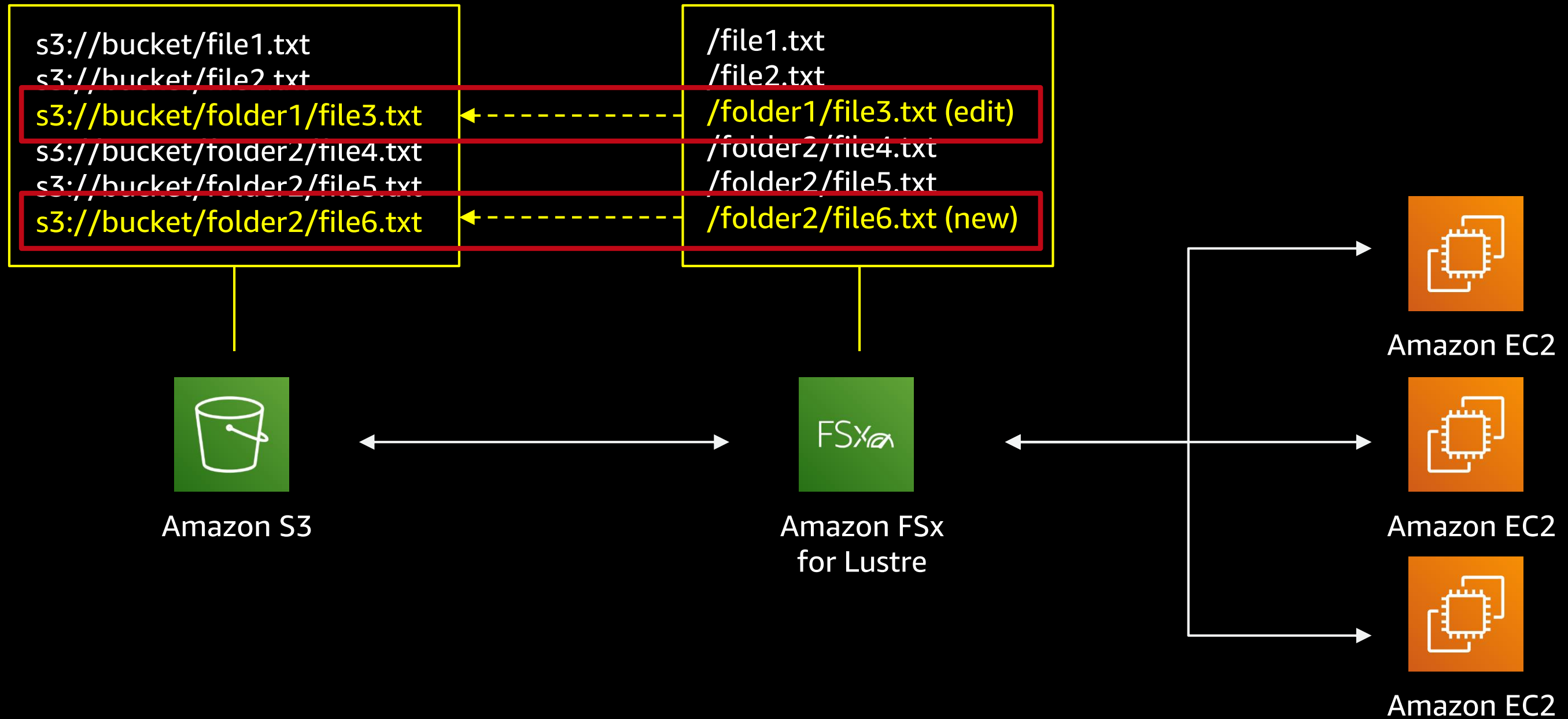
New or modified files can be exported from Amazon FSx to Amazon S3





Amazon S3-linked file system

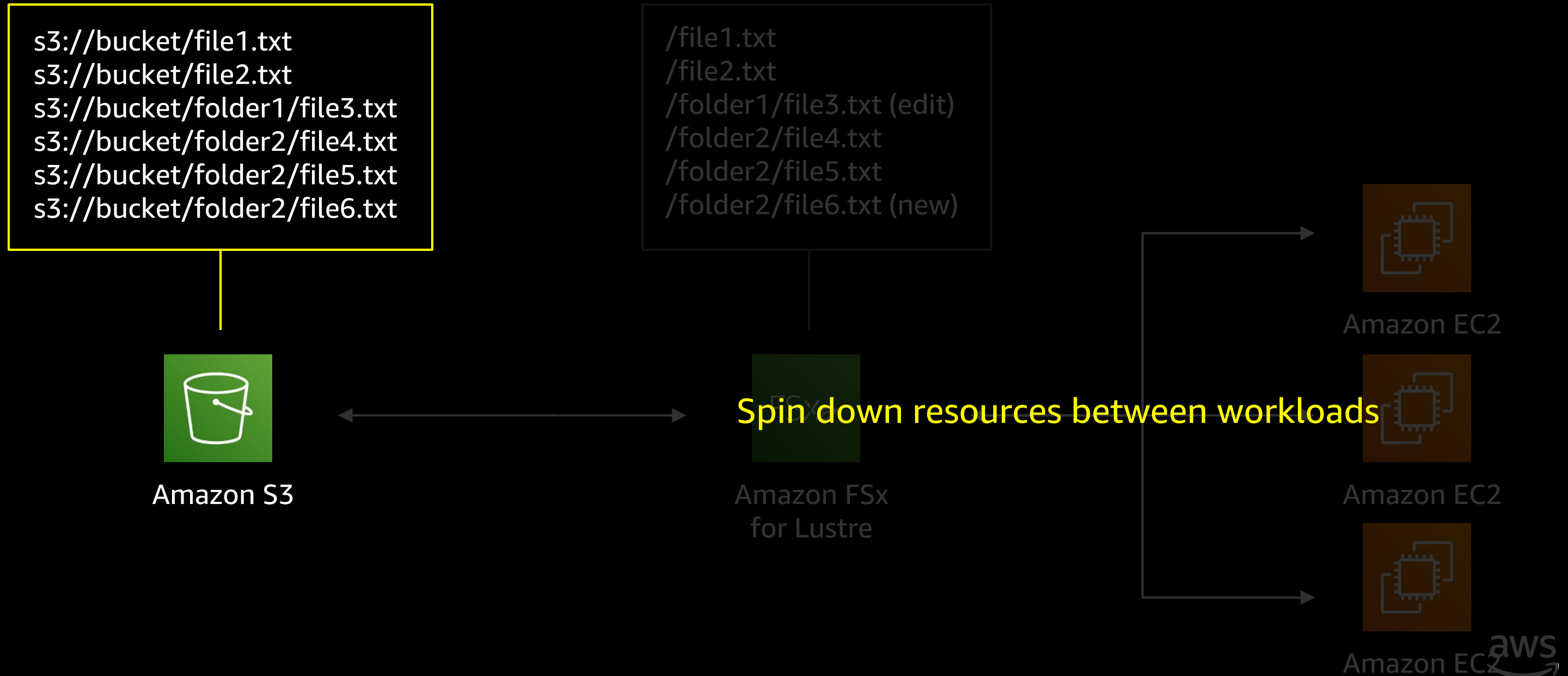
New or modified files can be exported from Amazon FSx to Amazon S3





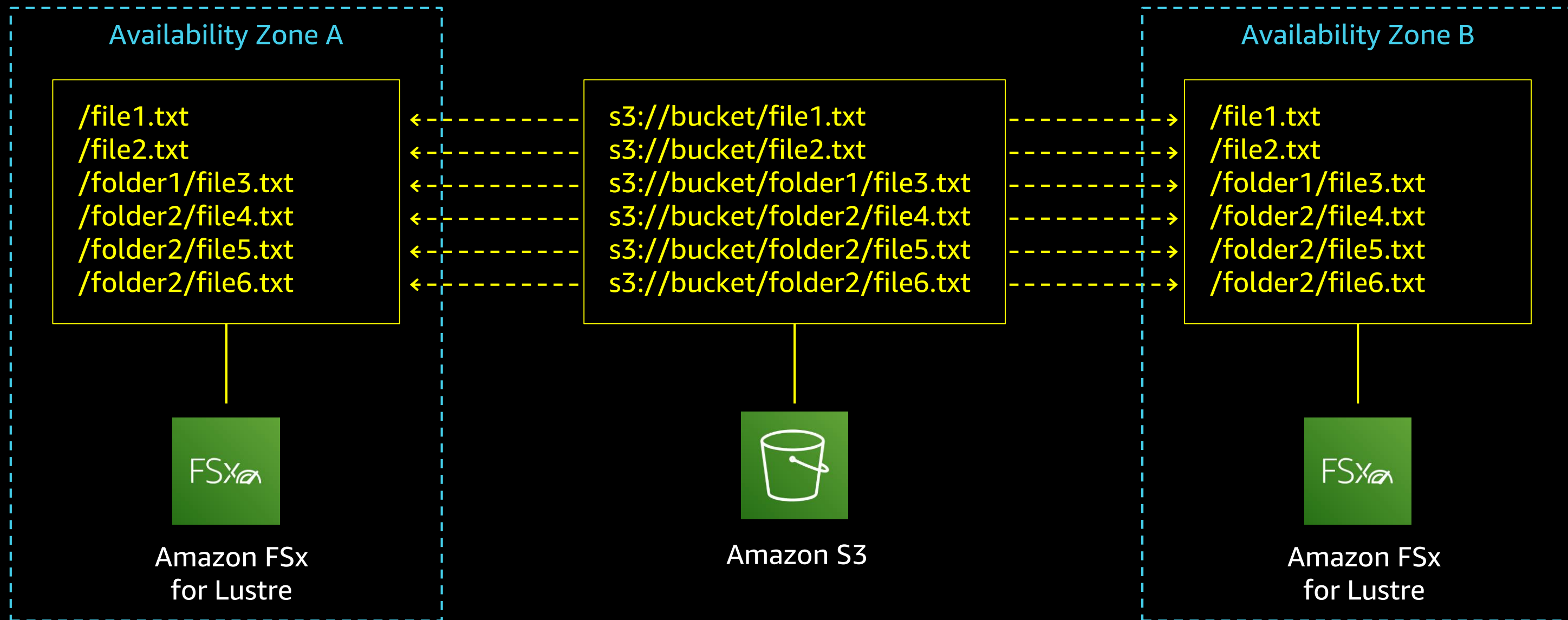
Amazon S3-linked file systems:

Spin up and spin down with compute resources



Amazon S3-linked file systems

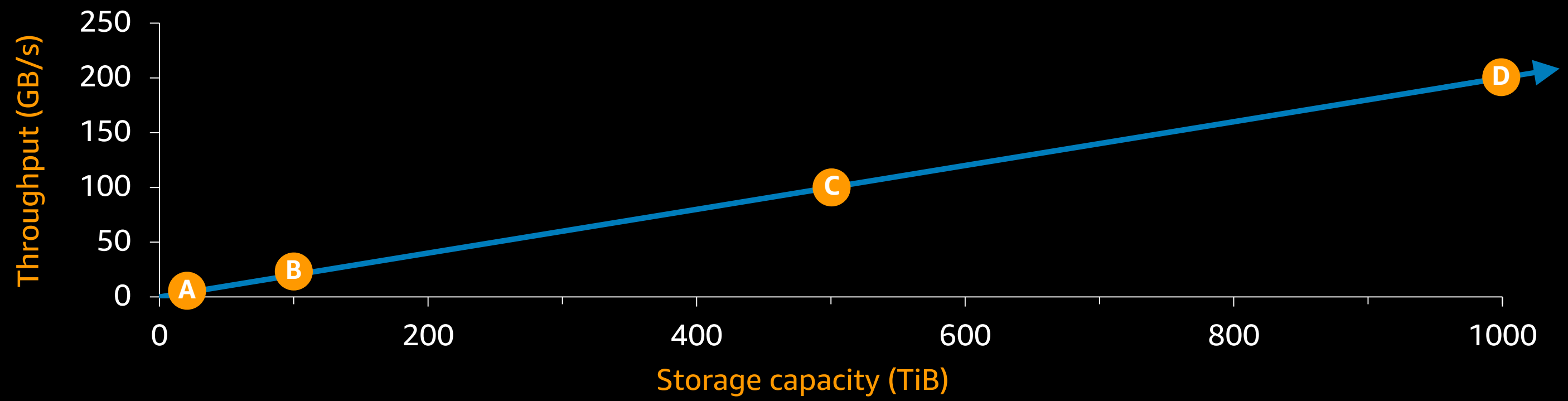
Multiple availability zones



Amazon FSx for Lustre: Amazon S3-linked file system demo



Fast - performance scales with storage capacity



	Storage capacity	Baseline throughput	Burst throughput	IOPS	Latencies
A	10 TiB	2 GB/s	13 GB/s	Tens of thousands	Sub-millisecond
B	100 TiB	20 GB/s	130 GB/s	Hundreds of thousands	Sub-millisecond
C	500 TiB	100 GB/s	650 GB/s	>1 Million	Sub-millisecond
D	1,000 TiB	200 GB/s	1,300 GB/s	Millions	Sub-millisecond

Source: <https://docs.aws.amazon.com/fsx/latest/LustreGuide/performance.html>



Amazon FSx for Lustre: Performance demo

Cost effective - Multiple deployment options



SSD
metadata



SSD single
copy of data

SSD scratch storage

For short-term,
latency-sensitive,
IOPS-focused
workloads



Cost effective - Multiple deployment options



SSD
metadata



SSD single
copy of data

SSD scratch storage

For short-term,
latency-sensitive,
IOPS-focused
workloads



SSD
metadata



SSD redundant
copies of data

SSD persistent storage

For longer-term,
latency-sensitive,
IOPS-focused
workloads

Cost effective - Multiple deployment options



SSD metadata



SSD single copy of data

SSD scratch storage

For short-term, latency-sensitive, IOPS-focused workloads



SSD metadata



SSD redundant copies of data

SSD persistent storage

For longer-term, latency-sensitive, IOPS-focused workloads



SSD metadata



Optional
SSD read-only cache



HDD redundant copies of data



HDD persistent storage

For longer-term, throughput-focused workloads



Optimize cost and performance for your workload

Storage type	Baseline throughput	Price (\$ per GB-month) ¹	Example applications
HDD	12 MB/s per TiB	\$0.025	Genome analysis
		\$0.041 (w/ SSD cache)	
	40 MB/s per TiB	\$0.083	Financial simulations
		\$0.099 (w/ SSD cache)	
SSD	50 MB/s per TiB	\$0.14	Machine learning for imaging
	100 MB/s per TiB	\$0.19	SAS grid
	200 MB/s per TiB	\$0.29	EDA simulations
		\$0.14 (scratch)	Video rendering

Pay only for storage provisioned; no additional throughput or request costs

¹Sample pricing for US East (N. Virginia); prices are as of August 30, 2020 and subject to change without notice; pricing varies by AWS Region

Source: <https://aws.amazon.com/fsx/lustre/pricing/>



Optimize cost and performance for your workload

Storage type	Baseline throughput	Price (\$ per GB-month) ¹	Example applications
HDD	12 MB/s per TiB	\$0.025	Genome analysis
		\$0.041 (w/ SSD cache)	
	40 MB/s per TiB	\$0.083	Financial simulations
		\$0.099 (w/ SSD cache)	
SSD	50 MB/s per TiB	\$0.14	Machine learning for imaging
	100 MB/s per TiB	\$0.19	SAS grid
	200 MB/s per TiB	\$0.29	EDA simulations
		\$0.14 (scratch)	Video rendering

Pay only for storage provisioned; no additional throughput or request costs

¹Sample pricing for US East (N. Virginia); prices are as of August 30, 2020 and subject to change without notice; pricing varies by AWS Region

Source: <https://aws.amazon.com/fsx/lustre/pricing/>



Amazon FSx for Lustre | AWS integrations

Security



Amazon VPC

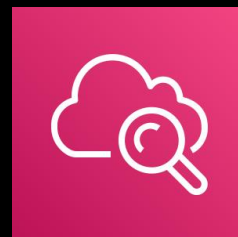


AWS Identity and Access Management (IAM)

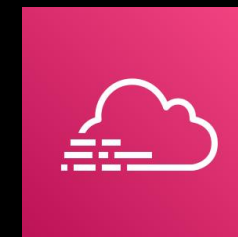


AWS Key Management Service (KMS)

Monitoring and automation



Amazon CloudWatch



AWS CloudTrail



AWS CloudFormation



Amazon Elastic Kubernetes Service (EKS)

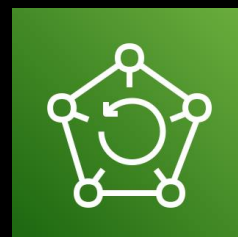


AWS ParallelCluster



AWS Batch

Compute clusters and orchestration



AWS Backup

Data protection



Amazon S3

Data repository



Amazon SageMaker

Machine learning



Amazon FSx availability

AMER

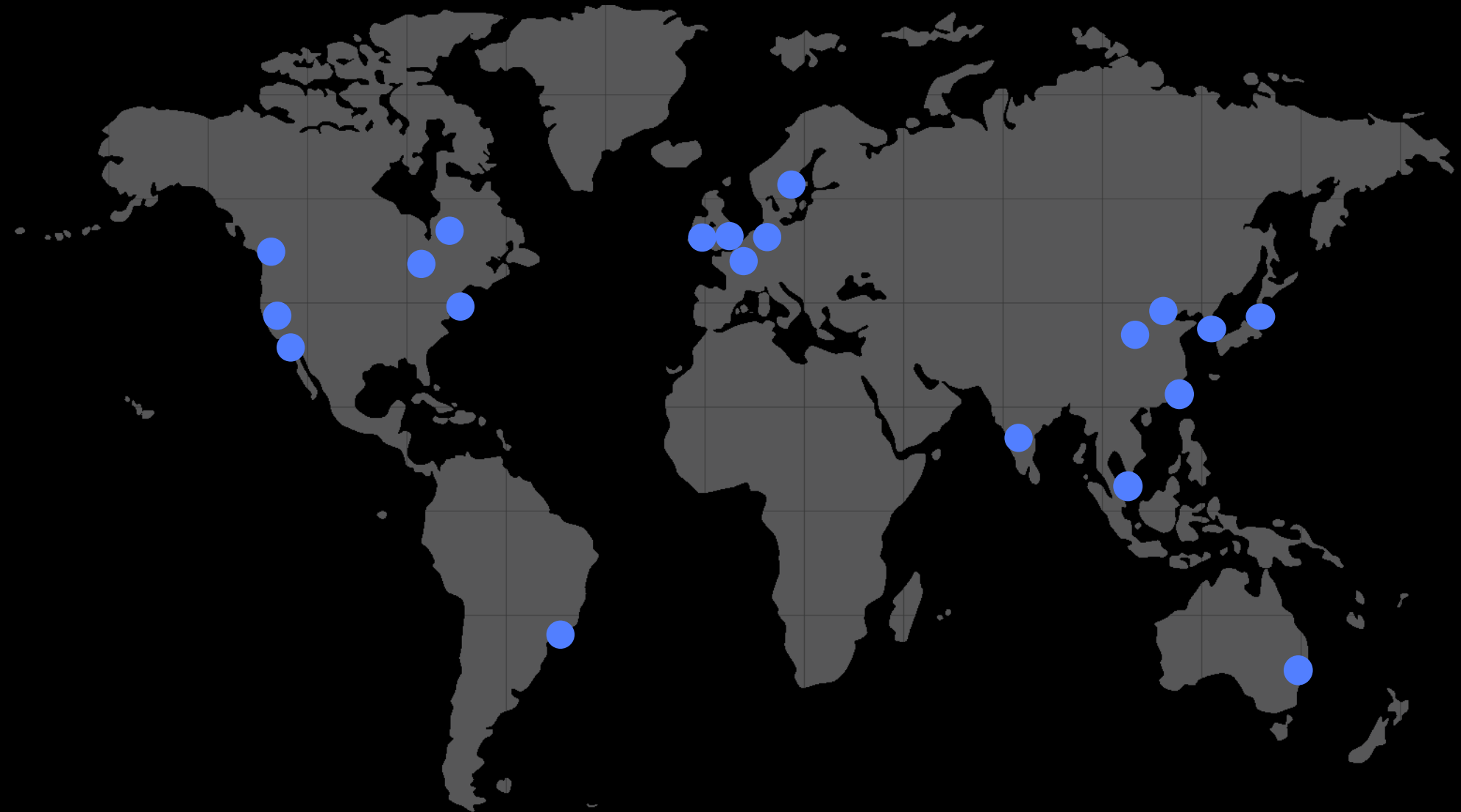
US West (Oregon)
US West (N. California)
US West (Los Angeles)
US East (N. Virginia)
US East (Ohio)
Canada (Central)
South America (Sao Paulo)

EU

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

APAC

Asia Pacific (Hong Kong)
Asia Pacific (Mumbai)
Asia Pacific (Seoul)
Asia Pacific (Sydney)
Asia Pacific (Singapore)
Asia Pacific (Tokyo)
China (Beijing)
China (Ningxia)



Recap

Amazon FSx for Lustre is a fully managed service that provides simple, cost-effective, high-performance storage for compute workloads, with multiple deployment options

Accelerate innovation - Simplicity in architecture using FSx for Lustre

- Seamless integration with Amazon SageMaker for Machine Learning
- Easily integrate with your data on Amazon S3, which can serve a primary data store for your spin-up / spin-down workloads

Accelerate results - Supercharge your workloads using FSx for Lustre

- Improves workload performance by reducing data-access latencies
- Provides file system consistency across multiple Availability Zones



Get up and running with these resources

Learn about Amazon FSx for Lustre:

<https://aws.amazon.com/fsx/lustre>

Get hands-on experience with the Amazon FSx for Lustre:

<https://github.com/aws-samples/amazon-fsx-workshop/tree/master/lustre>

Visit the AI and Machine Learning Resource Hub for more resources

Dive deeper with these resources, get inspired and learn how you can use machine learning to accelerate business outcomes.

- The machine learning journey e-book
- Machine learning enterprise guide
- 7 leading machine learning use cases e-book
- A strategic playbook for data, analytics, and machine learning
- Accelerating ML innovation through security e-book
- ... and more!

[Visit resource hub »](#)



<https://tinyurl.com/aiml-aws>

AWS Machine Learning (ML) Training and Certification

Learn like an Amazonian, based on the curriculum we've used to train our own developers and data scientists



AWS is how you build machine learning skills

Courses built on the curriculum leveraged by Amazon's own teams. Learn from the experts at AWS.



Flexibility to learn your way

Learn online with 65+ on-demand digital courses or live with virtual instructor-led training, plus hands-on labs and opportunities for practical application.



Validate your expertise

Demonstrate expertise in building, training, tuning, and deploying machine learning models with an industry-recognized credential.

aws.training/machinelearning

Thank You for Attending AWS Innovate

We hope you found it interesting! A kind reminder to **complete the survey**.
Let us know what you thought of today's event and how we can improve
the event experience for you in the future.



aws-apac-marketing@amazon.com



twitter.com/AWSCloud



facebook.com/AmazonWebServices



youtube.com/user/AmazonWebServices



slideshare.net/AmazonWebServices



twitch.tv/aws



Thank you!



INNOVATE

AI/ML EDITION