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

STG401

Optimize analytics workloads with AWS file solutions

Shrinath Kurdekar

Principal Solutions Architect
Amazon File Services
AWS

Laura Shepard

Principal GTM Specialist
Amazon FSx for Lustre
AWS

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

Organizations want more value from their data



Make better decisions



Improve efficiencies



Respond faster



Uncover opportunities

aws

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

Challenges

Traditional on-premises data-analytics infrastructure does not scale well enough and is too expensive.

To run analytics at scale, we need . . .

- Agile and flexible compute options
- Scalable and performant storage
- Easy lift-and-shift for existing workflows
- Modernize monolithic applications using microservices
- The ability to burst workloads to the cloud
- The option to consolidate data from disparate data sources

aws

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

Running your data analytics at scale

Microservices

Containers and serverless compute

Traditional compute

Serverless



AWS Lambda

Containers



AWS Fargate



Red Hat OpenShift

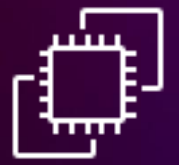


Amazon EKS



Amazon ECS

Compute



Amazon EC2

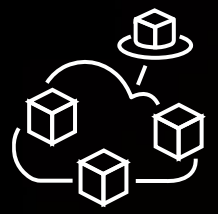


Performant, scalable, persistent shared storage

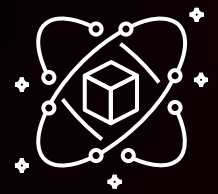
aws

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

We meet you where you are
in your cloud journey



Burst to the cloud



Move to the cloud



Build in the cloud

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

We’re razor-focused on providing the best file solutions to serve your usage patterns



Amazon FSx

File systems purpose-built for NAS or scale-out workloads, cloudified

1. Run NAS workloads in the cloud
2. Leverage the cloud’s virtually unlimited compute scale



Amazon EFS

Serverless, fully elastic file storage designed for builders

3. Easily share data in cloud-native apps and workflows



Amazon File Cache

High-speed cache for on-premises or AWS datasets

4. Easily burst your workloads to cloud

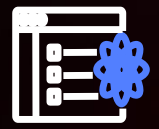
aws

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

Put data to work



AI/ML and data science



High-performance analytics



Financial analytics



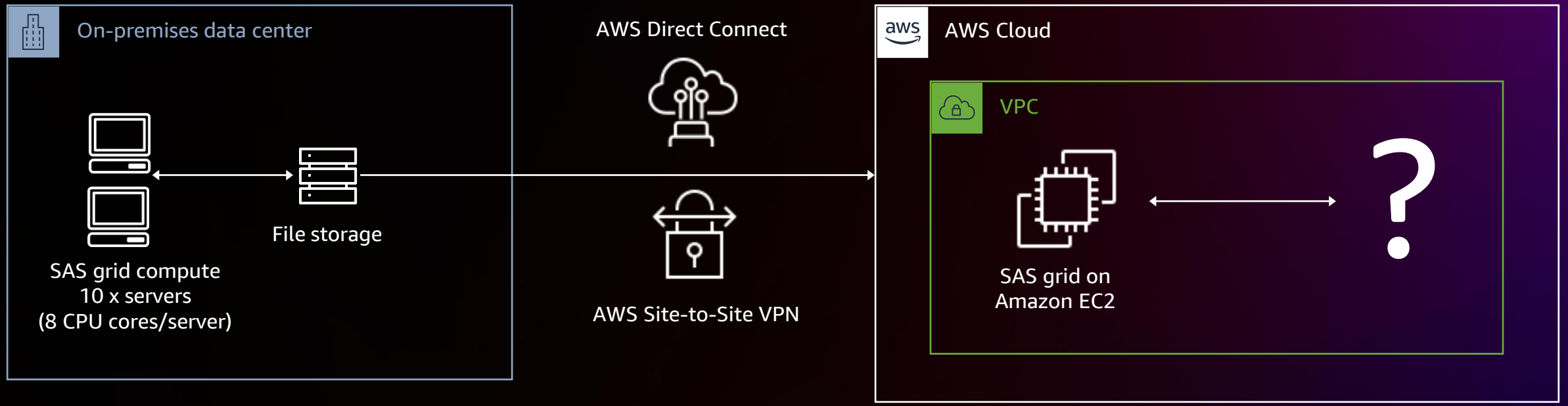
Genomics research

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

Customer scenario 1

Lift and shift SAS analytics workflows to cloud

- Minimum 125 MB/s per physical core
- Minimum of 8 GiB of memory per core



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

Let’s chalk talk

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

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, and cut the run time of SAS grid analytics workloads by half

---

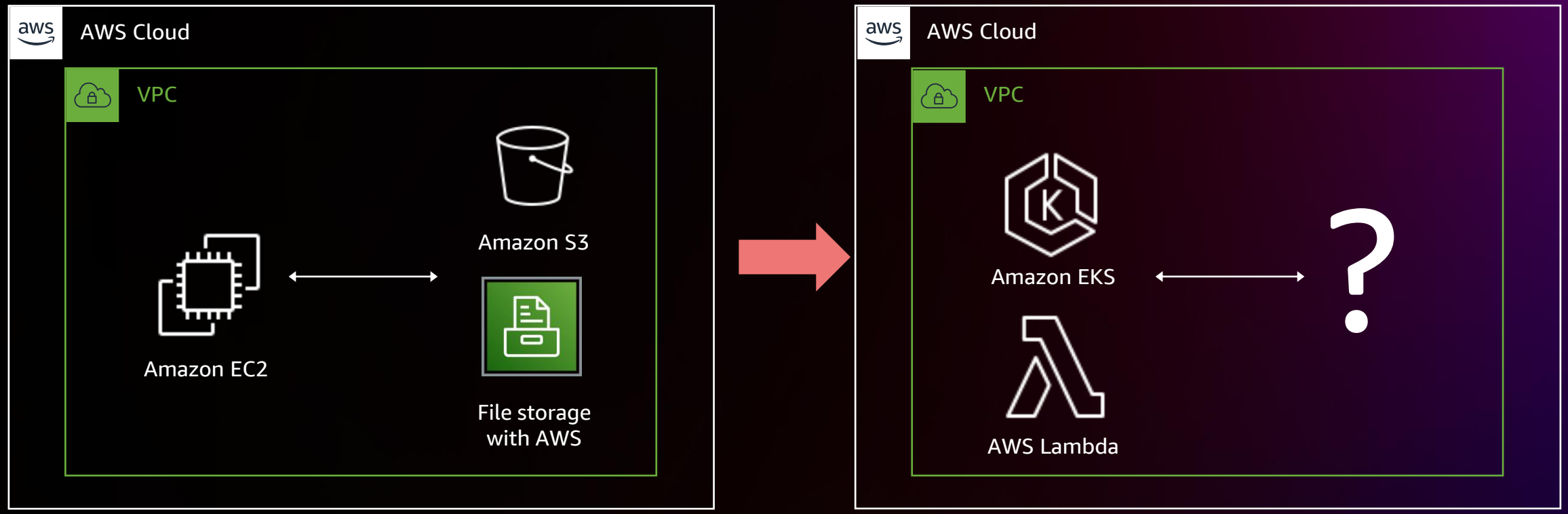
“Amazon FSx for Lustre helped us double the speed of our SAS grid workloads, reduce our total cost of ownership by 83%, and completely eliminate our operational burden.”

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

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

Customer scenario 2

Modernize using microservices architectures with containers and serverless



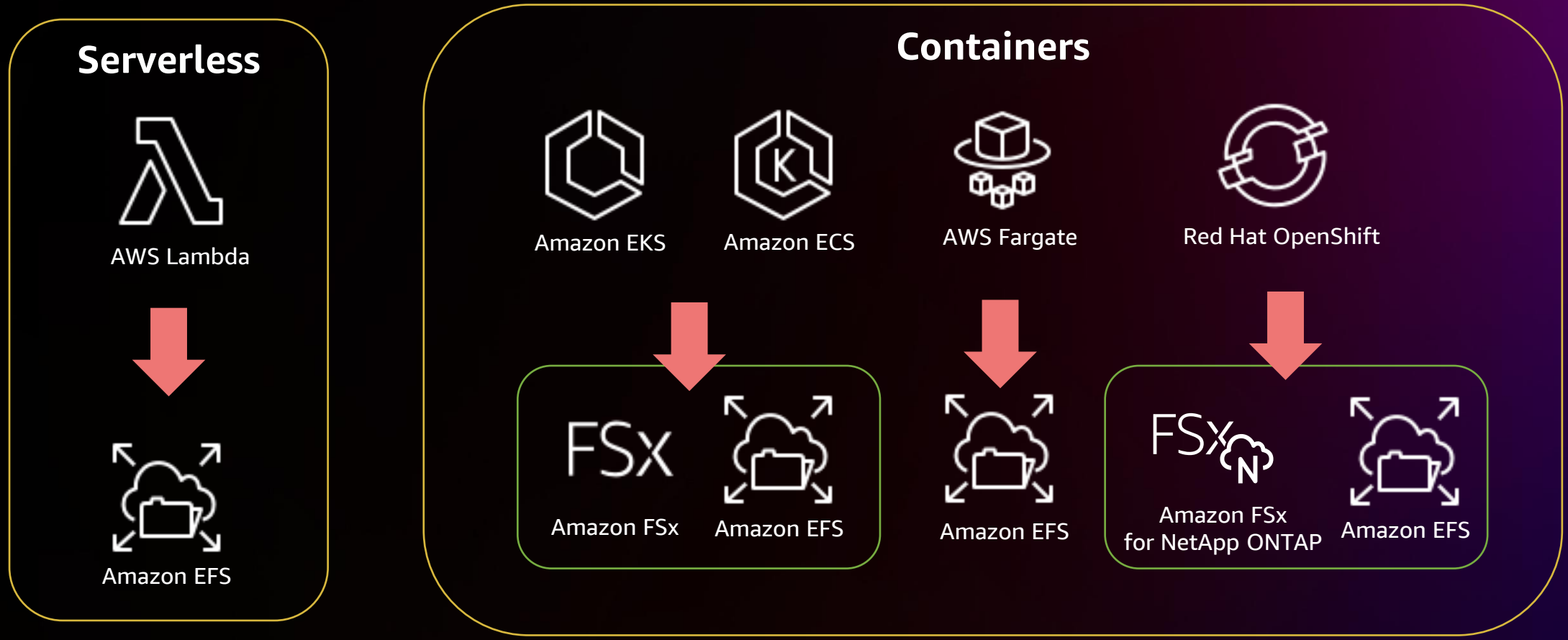
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Let’s chalk talk

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

File storage for modern or new applications

Modern applications built on serverless or containers can benefit from **persistent**, **durable**, and **scalable shared** storage



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

J&J builds data science platform using Amazon EKS and Amazon EFS

The challenge

J&J needed storage to share and perform analytics on their data-science workbench for genomics neuroscience, R&D, and drug discovery; a highly scalable solution was required to share and analyze hundreds of TiB of data from external and internal studies

The solution

Amazon EFS provides performant analytics storage with shared file access to data scientists using open-source genomics and Shiny, as well as Domino Data Lab, running on an Amazon EKS cluster

The benefits

- Faster time to insights: reduced analysis time by 35%
- Scales elastically to 500 TiB and growing
- Reduced costs by 37% using Amazon EFS lifecycle management via EFS One Zone and lower cost storage classes

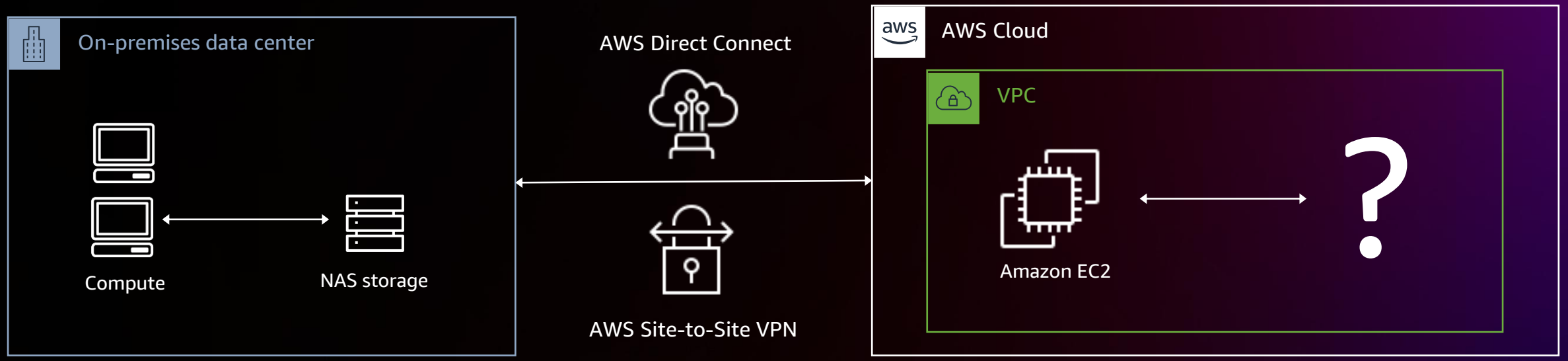
“Data scientists need access to massive data from different sources for genomics, R&D, and drug discovery to perform advanced analytics. With Amazon EFS, we were able to share data on a massive scale, increase innovation by accelerating analytics, and save 37% in costs.”

Greg Rusin, IT Manager, Johnson & Johnson

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

Customer scenario 3

Cloud-bursting your analytics workflows



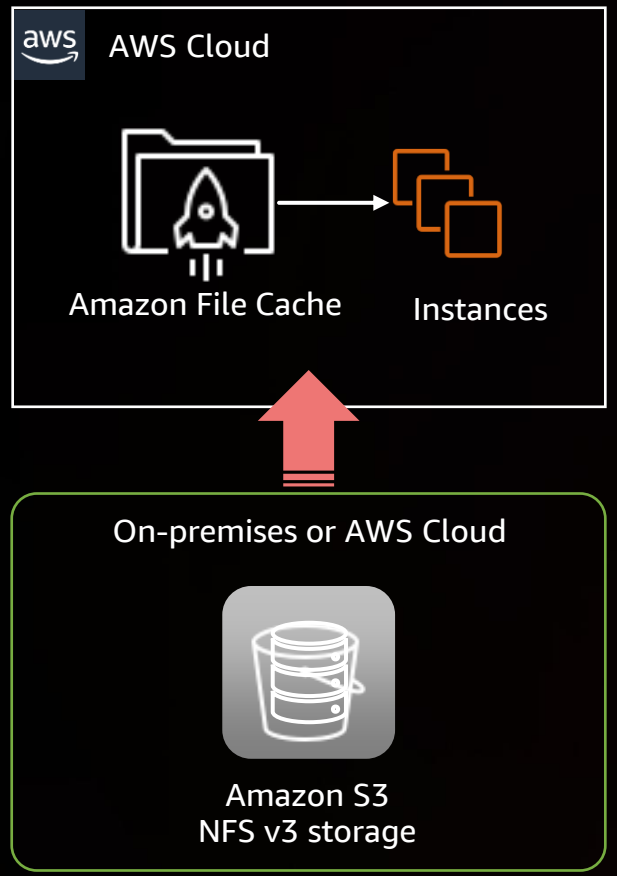
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Let’s chalk talk

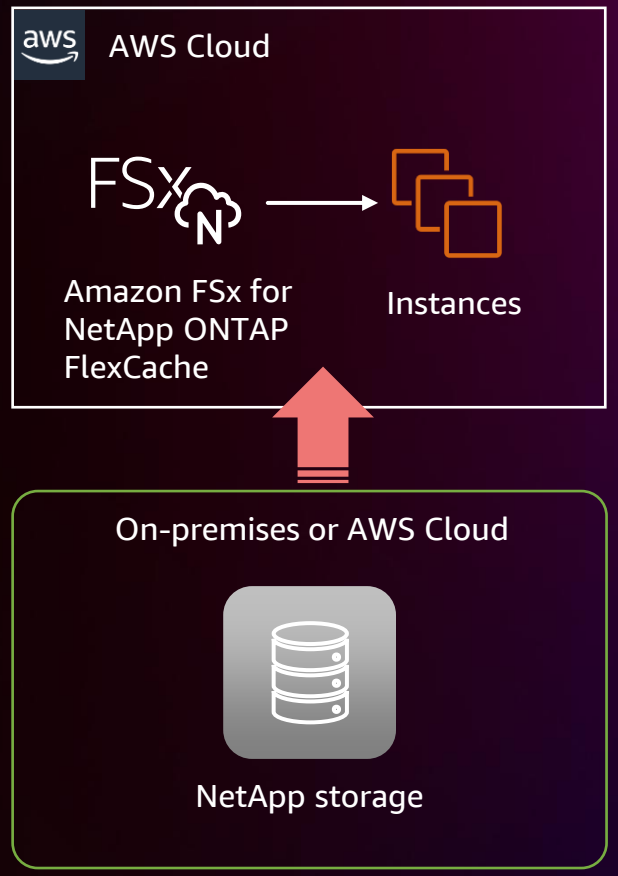
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cloud-bursting your analytics workloads

Amazon File Cache



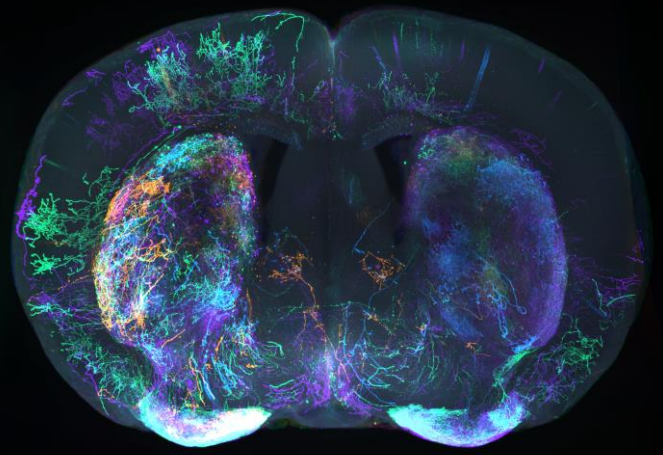
Amazon FSx for NetApp ONTAP



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

Cajal Neuroscience accelerates time to insights with Amazon File Cache

Cajal Neuroscience



Cajal Neuroscience is a drug discovery company focused on neurodegenerative disease

Cajal Neuroscience processes hundreds of terabytes of raw brain images to research how the brain degenerates in diseases like Parkinson's and Alzheimer's. This information is used to develop new medicines. On-premises data growth makes it slow and expensive to upload these large image files to the cloud for processing.

Amazon File Cache will enable them to bring the power and scale of the AWS Cloud to their on-premises data for a faster, easier image processing workflow.

"Tenfold growth in lab imaging is driving the need for more on-premises integration - we look forward to using Amazon File Cache to bring the compute to our data wherever it resides."

Avi Levitan, Principal Software Engineer
Cajal Neuroscience

aws

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

Related re:Invent sessions

STG218 **Accelerate compute workloads with Amazon File Cache**
Wed 11/30 4:45 PM in Upper Level, Cristal 7, Wynn

STG403 **Take your Amazon EFS container deployments to the next level**
Wed 11/30 5:30 PM in Level 1, Academy 417, Caesars Forum

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

Thank you!

Shrinath Kurdekar
(linkedin)

Laura Shepard
(linkedin)



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

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