AWS Data, Databases, and Analytics Online Series

Accelerating data analytics with cloud-native file storage

Luke Anderson

Storage Sales Leader, AWS

Why are customers using the cloud for data science?

Faster time to insights

Increased collaboration

Speed of innovation

Why AWS for data science workloads

Unlimited, on-demand infrastructure enabling scale and agility not attainable on-premises

The broadest portfolio of compute resources, data storage services, and transport technologies

Secure access to machine learning services, analytics services, and data science tools to simplify workflows

Better ROI

Faster time to insights

When to use a file system or a data lake for analytics

File System

Object Storage

- Applications, users or tools that require a file interface
- Strong consistency

- Global repository for large scale data analytics
- Rich metadata

Common workloads

Genomics

Machine learning

Financial modeling

Big data analytics

Traditional storage is not designed for modern data science

Administrative overhead

Lack of scalability

Lack of agility

Modern applications and data science

Traditional applications

- Shared application servers
- IT deployed & managed

Modern applications

- Developers directly deploy and manage functions and containers

Traditional data science

- Shared servers with user home directories, linked to LDAP/AD
- Approved toolsets, datasets

Modern data science

- Per-user notebook servers
- Data scientists deploy own scale-out training jobs

Using the right tool for the job

Using the right tool…for the job…for the workload

Amazon Elastic File System (Amazon EFS)

Amazon FSx for Windows File Server

Amazon FSx for Lustre

File storage for business workloads

File storage for compute-intensive workloads

Amazon EFS

Providing a more reliable, cost-effective, and cloud-native NFS service

Elastic

Highly available

Simple

High performance

Cost optimized

Access from on-premises

400% higher read operations/s

Amazon FSx for Windows File Server

Lowest-cost file storage in the cloud for Windows Workloads

- Fully managed
- Multi-AZ
- Built on Windows Server
- Fully compatible with the Windows file system
- Joins to customer AD with full Windows ACLs
- HDD or SSD storage options

Amazon FSx for Lustre

World’s most popular high-performance file system

Fully managed

Amazon Simple Storage Service (Amazon S3) datasets as POSIX file system

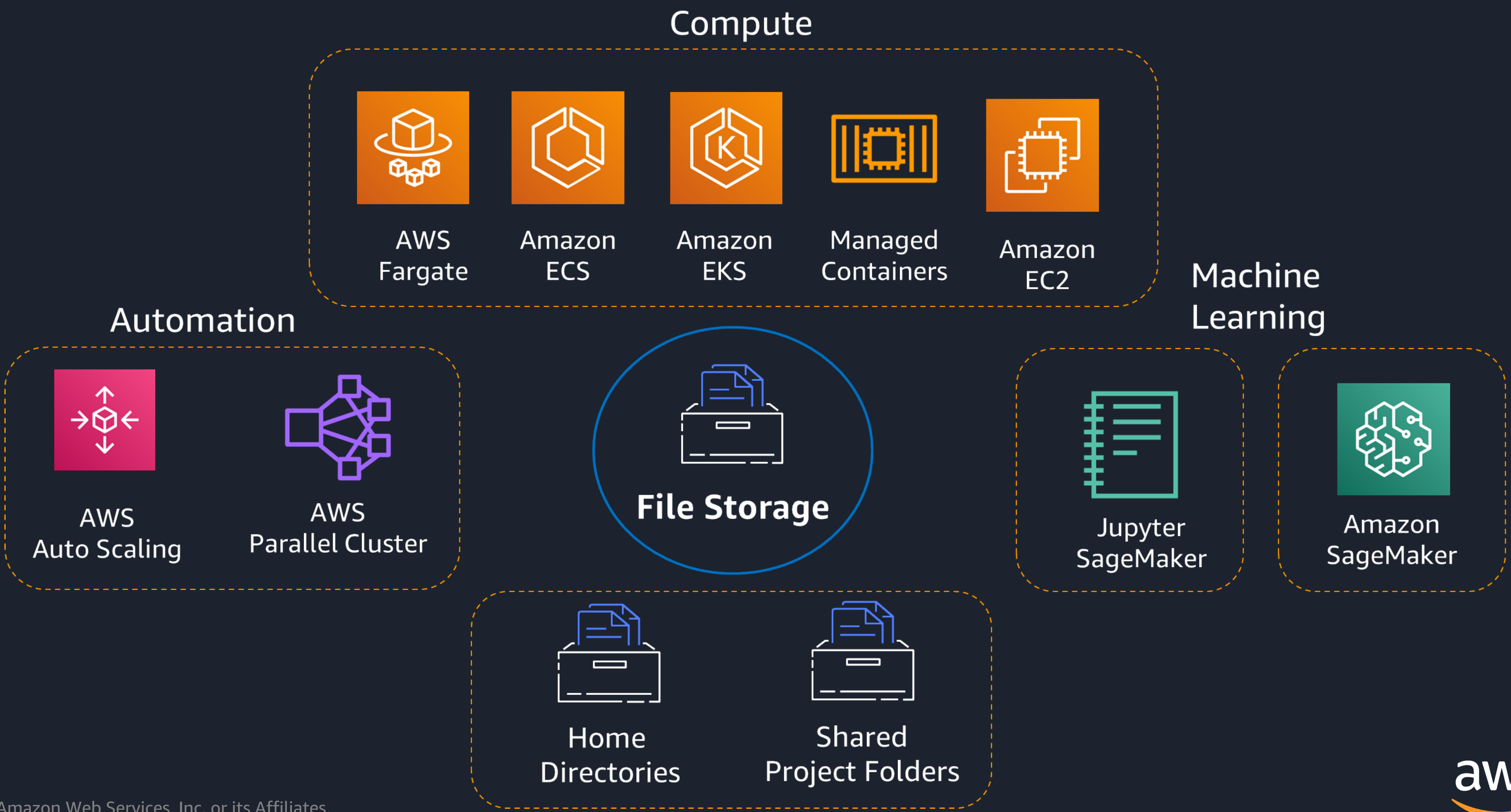
Highly performant – scratch or persistent

Designed for compute-intensive workloads

Flexible data processing options

Access from on-premises & integration with AWS services

Data science with file storage



Compute

AWS Fargate

Amazon ECS

Amazon EKS

Managed Containers

Amazon EC2

Automation

AWS Auto Scaling

AWS Parallel Cluster

File Storage

Home Directories

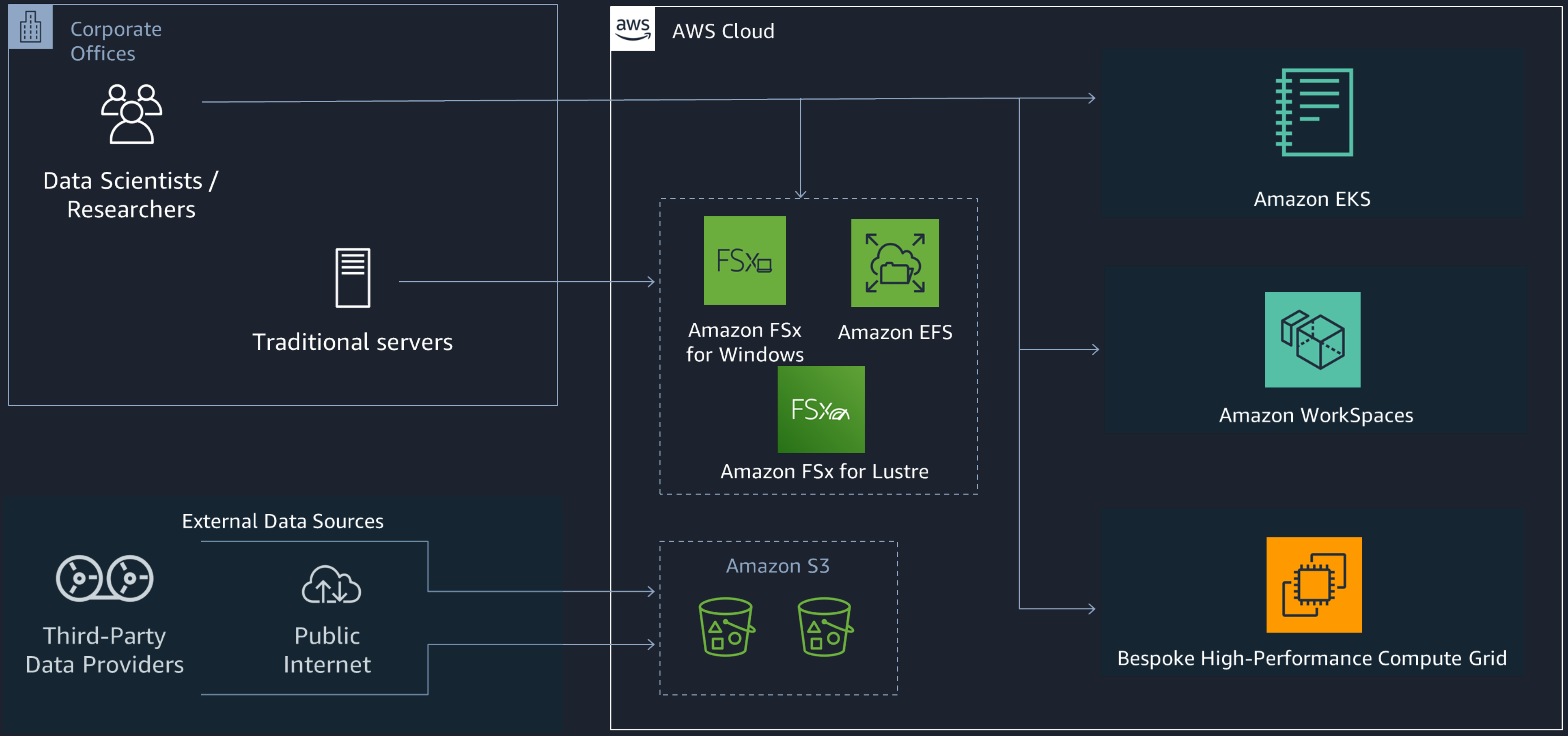
Shared Project Folders

Machine Learning

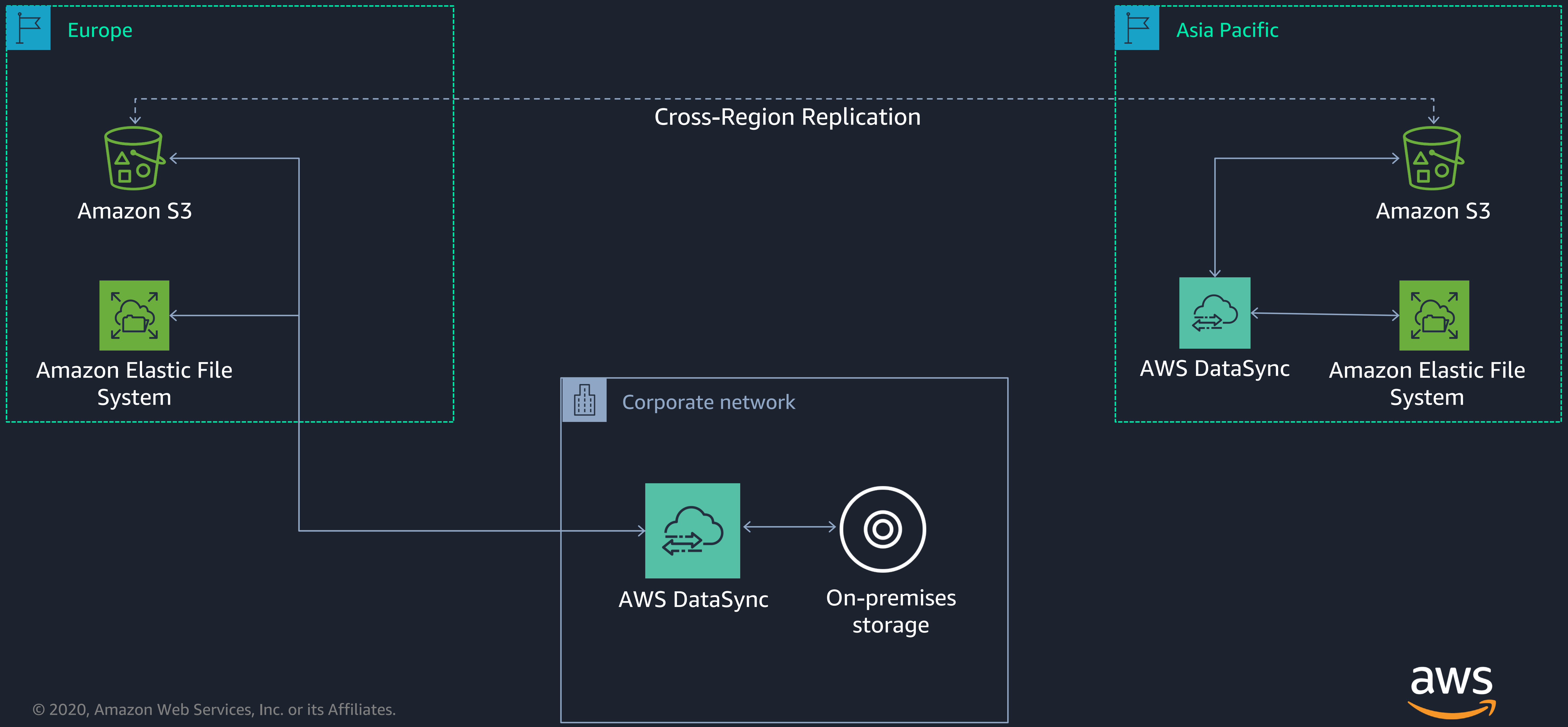
Jupyter SageMaker

Amazon SageMaker

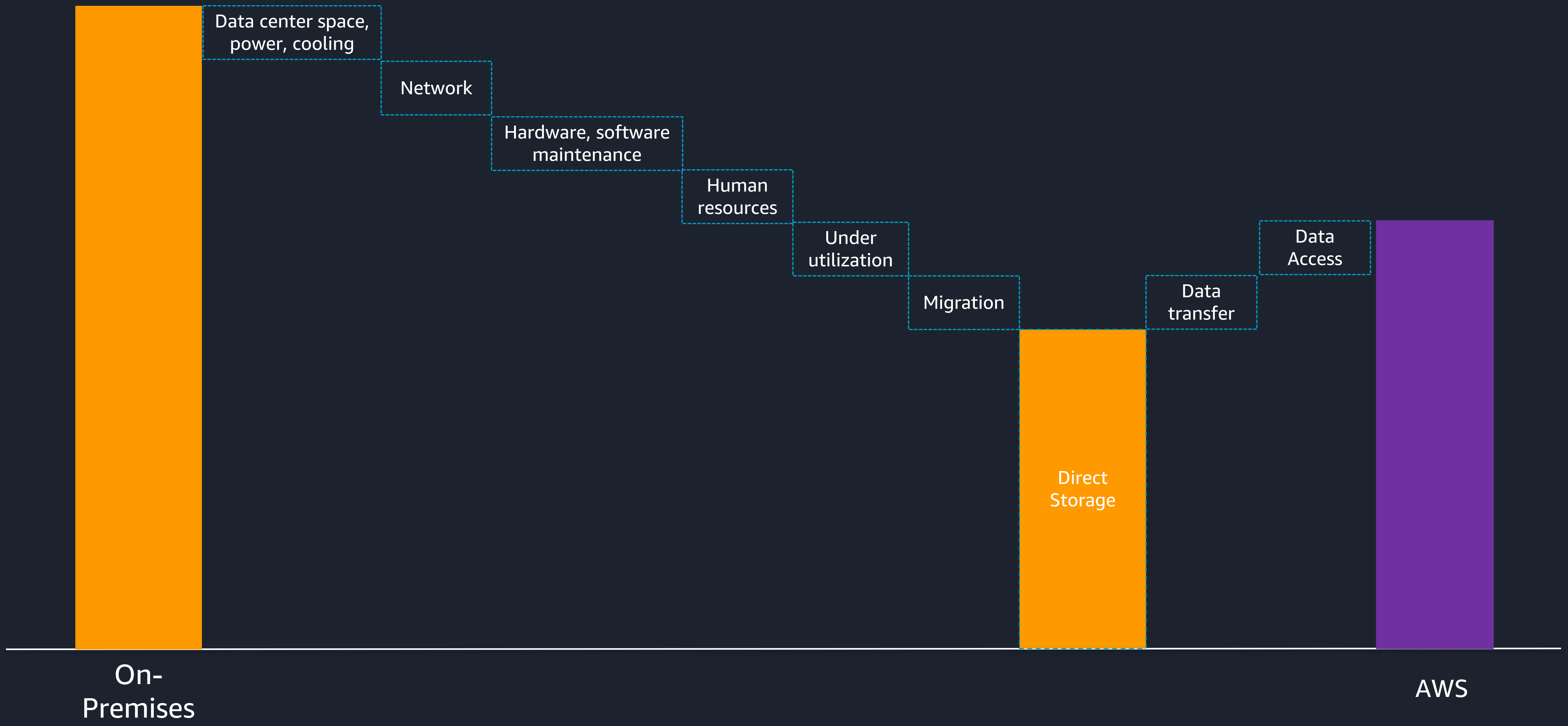
Data science architecture



Cross-region collaboration example

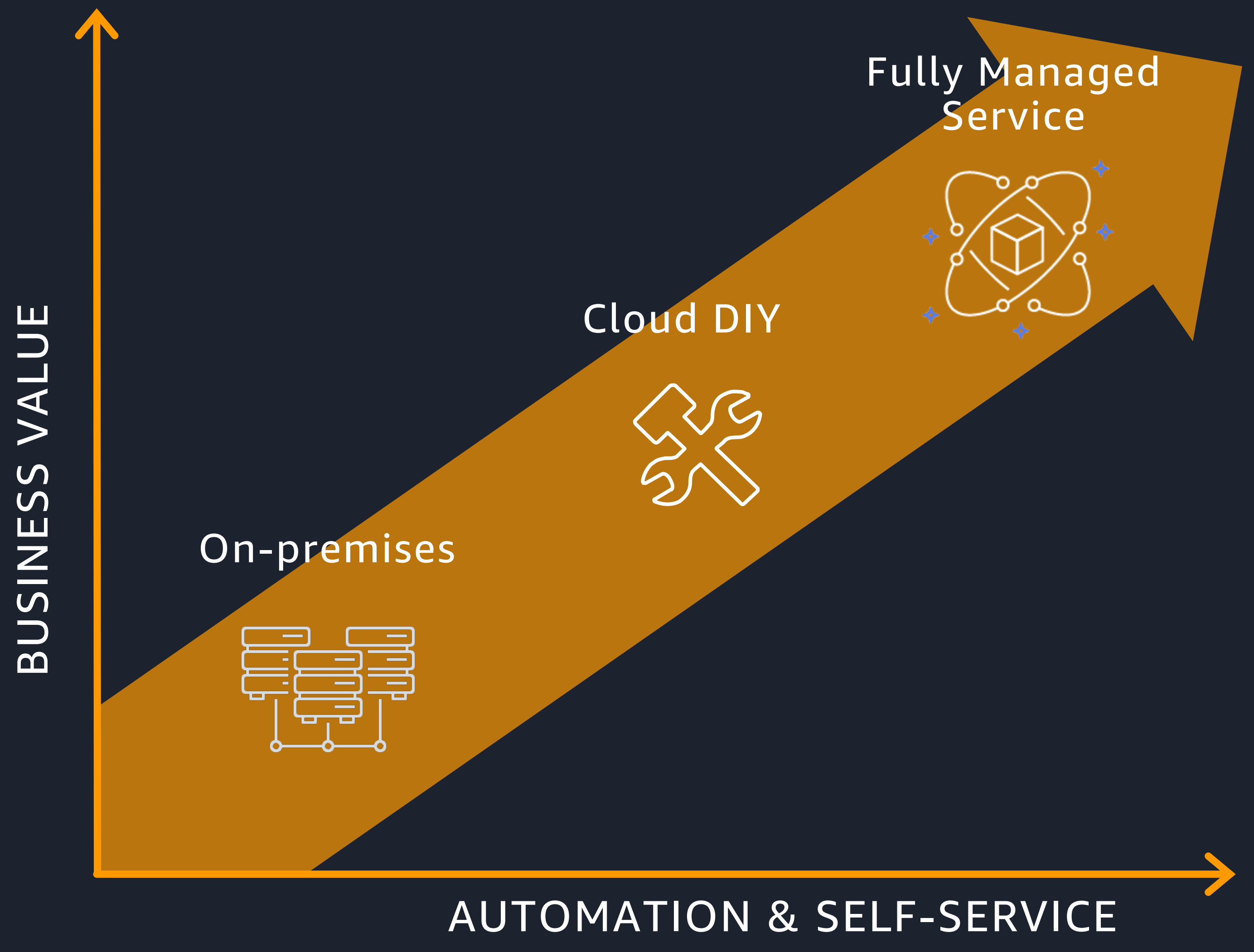


Understand your true TCO



Journey to (and in) the cloud

DISCOVER



- Moved analytics environment to AWS for agility benefits
- Built analytics environment based on a DIY file system on EC2
- Migrated to AWS managed storage service (Amazon EFS) for greater stability and ease of operations
- Moving to a fully managed storage service reduced the amount of time required to manage storage infrastructure by 90%

aws

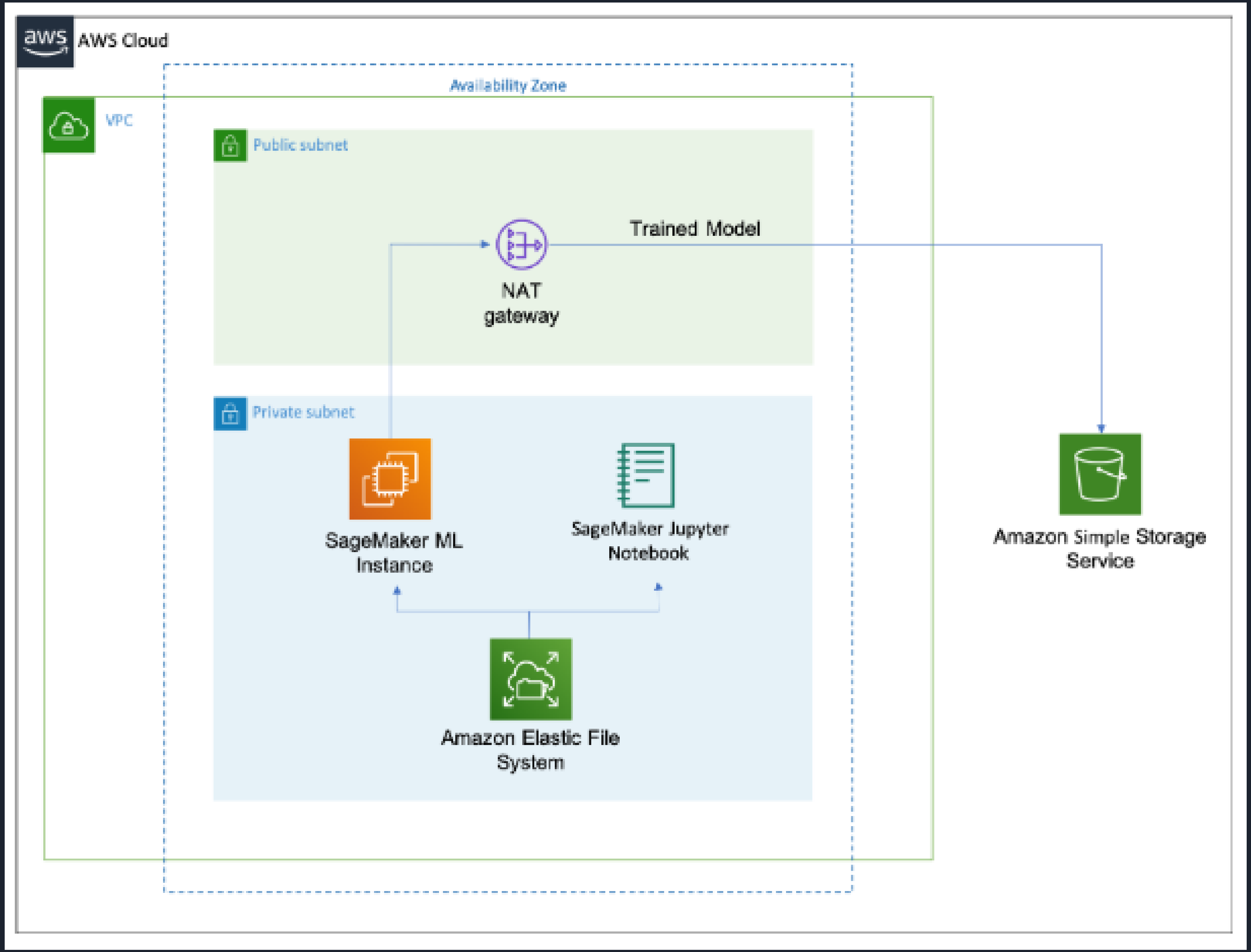
Want to learn more….then try it yourself

We have prepared a tutorial covering how to create an Amazon EFS to share with ML notebooks.

1. Go to http://www.github.com
2. Search for Amazon EFS
3. Click through to *amazon-efs-tutorial*
4. Then click through to the *data-science* folder

Data Science with Cloud-Native File Storage Workshop!

…Let’s walk through the demo…



Recap

1. Better ROI and faster time to insights
2. Broad portfolio of native file storage solutions
3. Simple and fast to get started

AWS Training and Certification

Training for the Whole Team

Explore tailored Data or Database learning paths for customers and partners

Flexibility to Learn Your Way

Build cloud skills with free digital Data training courses such as “The elements of Data Science”, or dive deep with classroom training

Validate Skills with AWS Certification

Demonstrate expertise with a Data industry-recognized credential (Data analytics and Database Specialty AWS Certifications)

https://aws.amazon.com/training/

Visit the Data, Databases, and Analytics Resource Hub for more resources

Dive deeper with these newly created whitepapers and e-books to help you uncover new insights and value from your data

- An introduction to cloud databases
- Enter the purpose-built database era
- Harness the power of data
- Creating a modern analytics architecture
- The data-driven enterprise
- … and more!

https://tinyurl.com/aws-data-databases-analytics

Visit resource hub »

Thank you for attending
AWS Data, Databases, and Analytics Online Series

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

aws

Thank you!