
AWS
re:Invent

STG221

SAS Grid on AWS: Optimize price, performance, and agility

Laura Shepard

Storage Specialist, Amazon FSx

AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Overview of SAS Grid on AWS

Defining storage and compute requirements for your application

References and migration resources



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why SAS Grid on AWS



The same high-performance analytics available on premises



Access to spin-up compute and storage resources on demand



Lower costs by paying only for IT resources consumed

SAS Grid using FSx for Lustre: Business impact

T-Mobile



Amazon FSx for Lustre

83% reduction in total cost of ownership

2x increase in speed of SAS Grid workloads

$1.5M in annual savings

2x improvement in storage utilization

2 hrs to zero reduced backup admin process time

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

T-Mobile

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

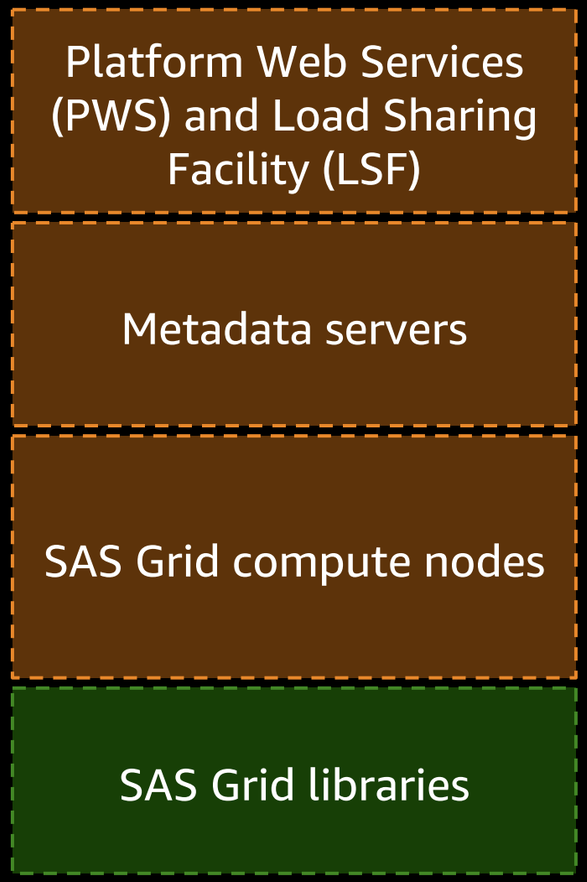
Cut the run time of SAS Grid analytics workloads by half

> "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

SAS Grid on AWS



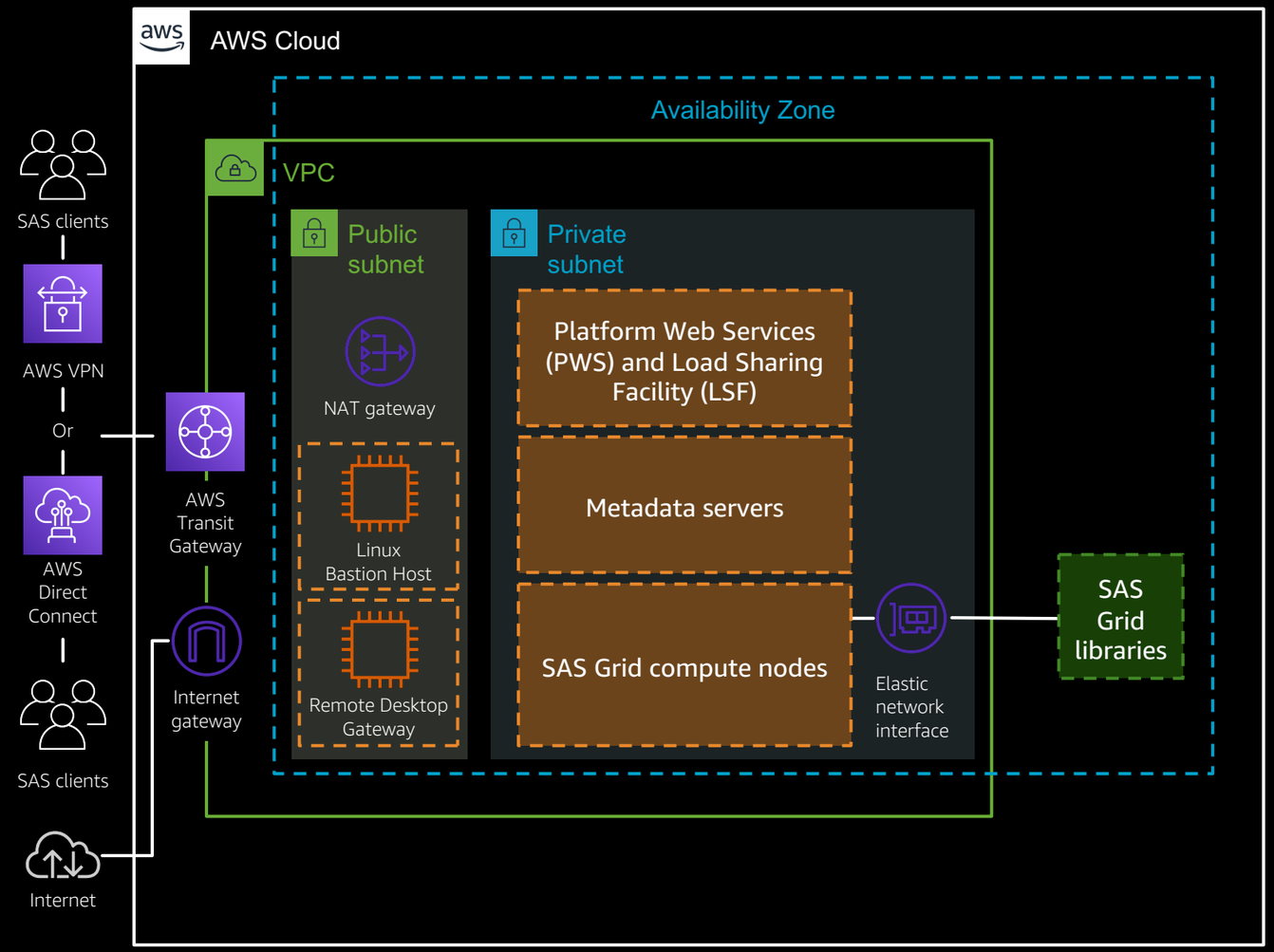
Dispatch jobs submitted and return the status of each job

Contain the metadata repository that stores the metadata definitions of all SAS Grid manager products

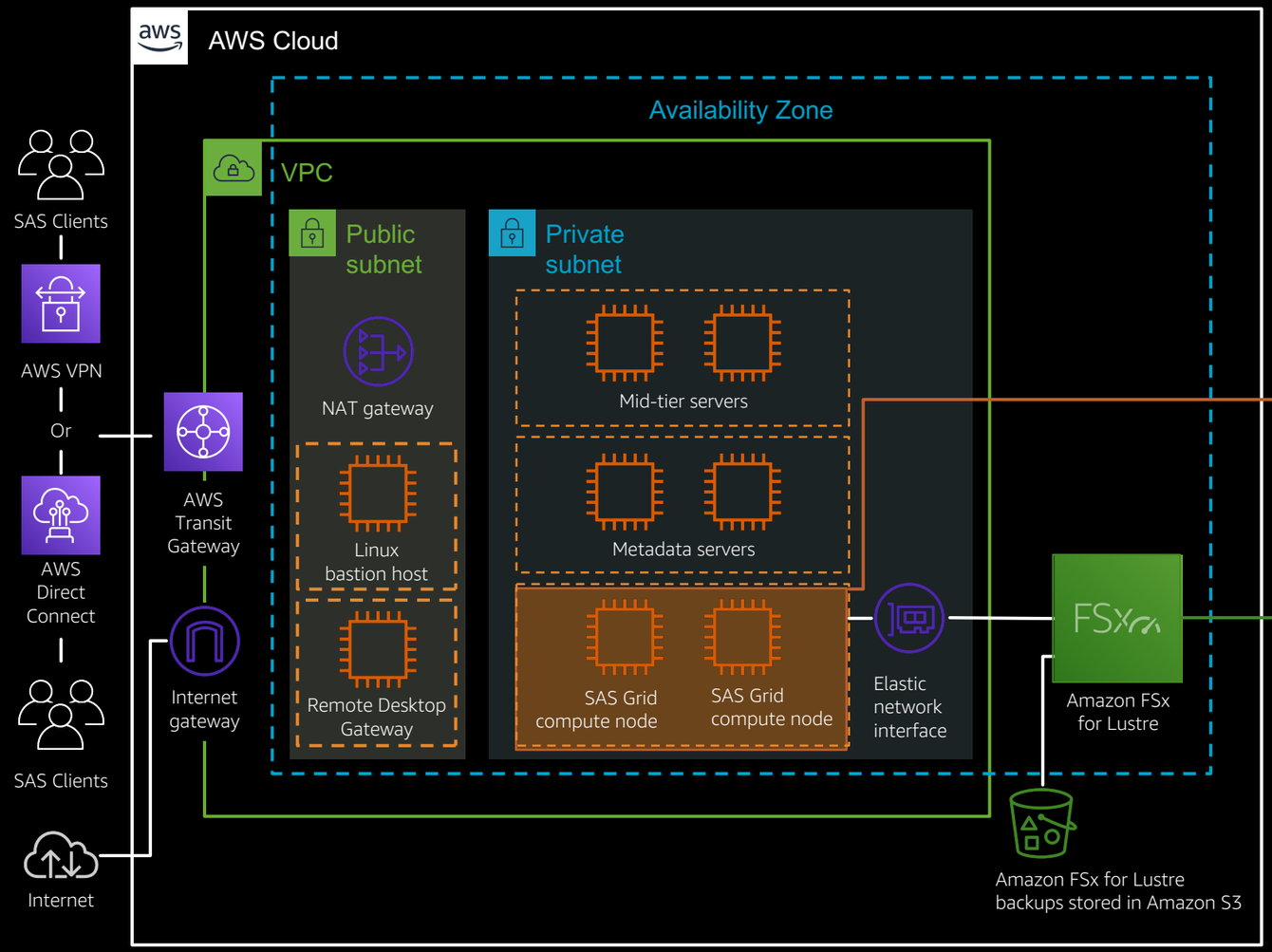
Run a portion of the work allocated to the grid

Store your enterprise data

SAS Grid on AWS



SAS Grid on AWS



SAS Grid costs and performance are driven by:

SAS Grid compute nodes

Storage
- /SASDATA
- /SASWORK
- /UTILLOC

Why Lustre	The world’s most popular high-performance file system
Why Lustre on AWS	Increase agility and reduce costs
Why Amazon FSx for Lustre	Reduce storage administration and accelerate innovation

Amazon FSx for Lustre

Example use cases



Financial modeling and analytics



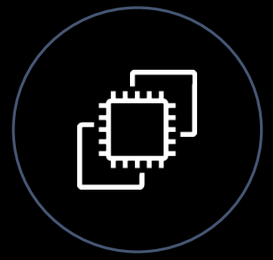
Genome analysis



Media rendering and transcoding



ECU simulations and object detection

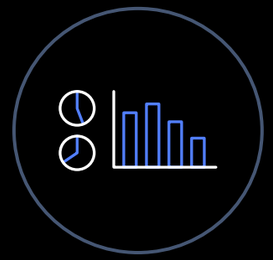


Semiconductor EDA



Seismic data processing

. . . and application areas



Big data analytics

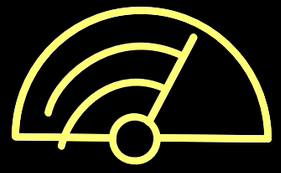


Machine learning



High-performance computing

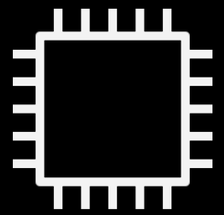
Steps to selecting storage and compute



Determine throughput requirements



Define /SASDATA performance requirements



Identify optimal Amazon EC2 instances



Select storage location for /SASWORK and /UTILLOC

/SASDATA throughput requirements



50
cores

x

100
MB/s per core

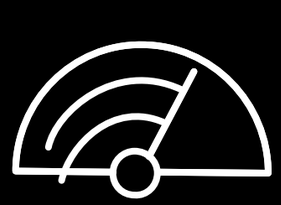
=

5,000
MB/s



Amazon FSx
for Lustre

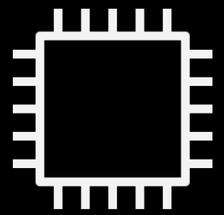
Steps to selecting storage and compute



Determine throughput requirements



Define /SASDATA performance requirements



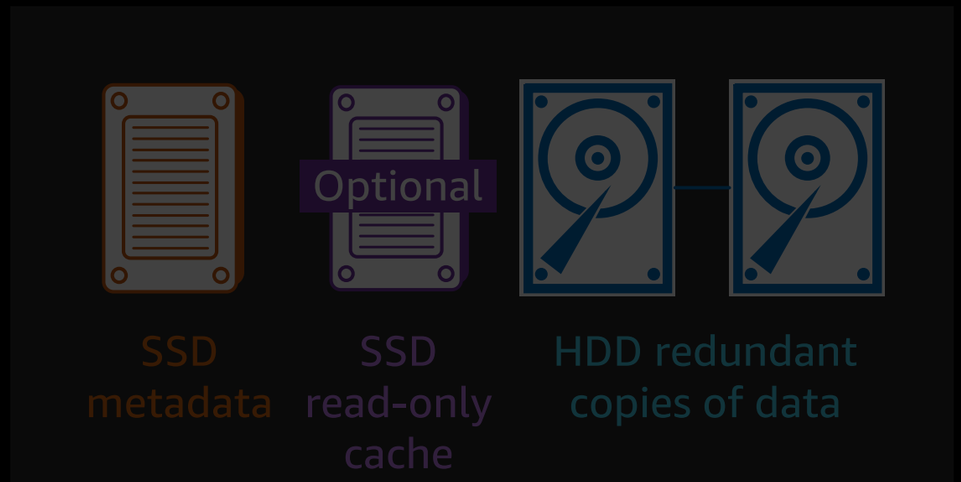
Identify optimal Amazon EC2 instances

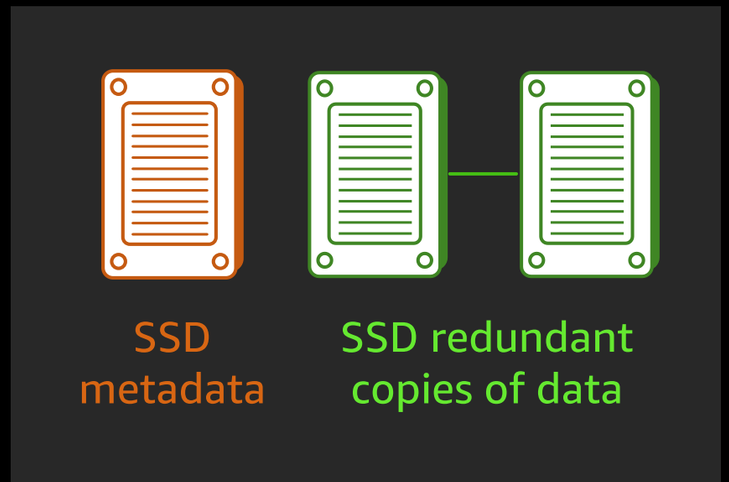


Select storage location for /SASWORK and /UTILLOC

FSx for Lustre: Multiple file system options



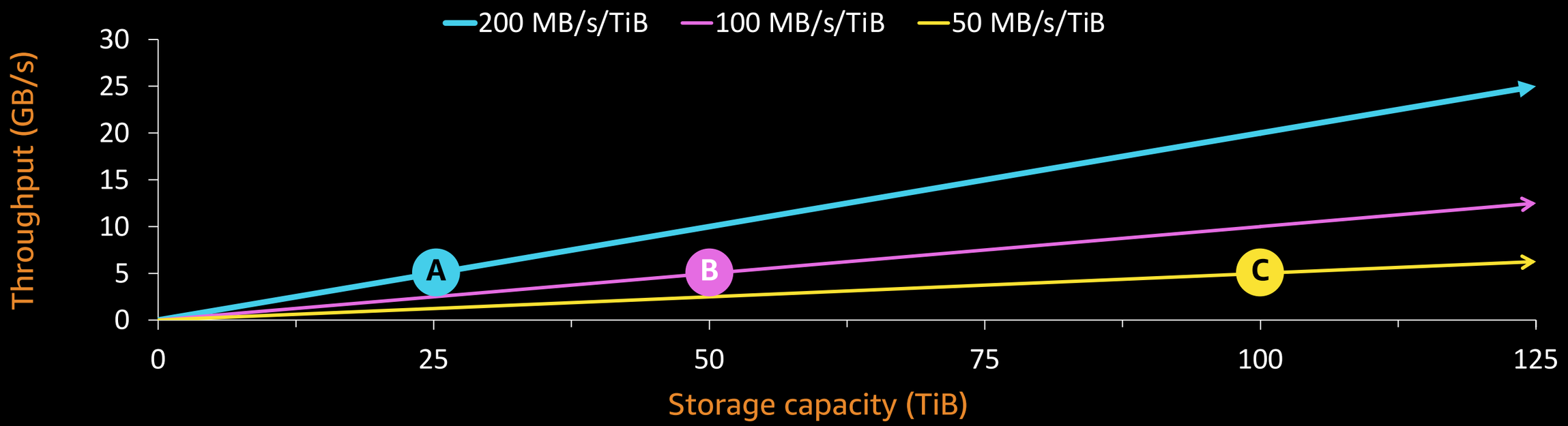






Optimal selection for SAS Grid libraries

Powered by Lustre's scalable performance



	Throughput	FSx Throughput Tier	Storage capacity provisioned	Price, $ per GB-month[1]
A	5,000 MB/s	200 MB/s per TiB	25 TiB	$0.29
B	5,000 MB/s	100 MB/s per TiB	50 TiB	$0.19
C	5,000 MB/s	50 GB/s per TiB	100 TiB	$0.14

[1] Sample pricing for US East (N. Virginia). *Prices are as of Aug 30, 2020, and subject to change without notice. Pricing varies by AWS Region.*

Burst up to 26x baseline throughput



Deployment type		Persistent-50	Persistent-100	Persistent-200
Disk throughput (MB/s per TiB)	Baseline	50	100	200
	Burst	240	240	240
Disk latencies per file operation (P50)		Sub-ms		
Cache storage (GiB per TiB)		2.2 RAM	4.4 RAM	8.8 RAM
Network IOPS (IOPS)		Tens of thousands baseline Hundreds of thousands burst		
Network throughput MB/s per TiB	Baseline	250	500	750
	Burst	1,300	1,300	1,300
Network throughput burst multiple of baseline disk throughput		26x	13x	6x

Steps of a DIY on-premises Lustre deployment

- Employ Lustre experts (difficult to acquire and keep)
- Budget and pay for power and cooling (expensive)
- Invest in capital expenditure (expensive)
- Procure and purchase hardware (weeks to months)
- Rack and stack storage servers (days to weeks)
- Build ZFS packages to work with supported OS (hours to days)
- Build Lustre server and client packages on supported OS (hours to days)
- Deploy metadata server (MDS) and required storage (hours to days)
- Deploy object storage servers (OSS) and object storage targets (OST) (hours to days)
- Install ZFS packages and Lustre server packages (hours to days)
- Set up ZFS pools and configure file system on MDS and OSS (hours to days)
- Implement tunings for Lustre (days to weeks)

FSx for Lustre file system creation demo



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Storage

Amazon FSx

Launch and run feature-rich and highly performant file systems with just a few clicks

Get started

Create file system

Amazon FSx for Windows File Server

Amazon FSx for Lustre

With Amazon FSx, you only pay for the resources you use.

How it works

Amazon FSx makes it easy and cost effective to launch and run popular file systems. With Amazon FSx, you can leverage the rich feature sets and fast performance of widely-used open source and commercially-licensed file systems, while avoiding time-consuming administrative tasks like hardware provisioning, software configuration, patching, and backups. It provides cost-efficient capacity and high levels of reliability, and it integrates with other AWS services so that you can manage and use the file systems in cloud-native ways.

Amazon FSx provides you with two file systems to choose from: Amazon FSx for Windows File Server for enterprise workloads and Amazon FSx for Lustre for high-performance workloads.

FSx Amazon FSx for Windows File Server

FSx Amazon FSx for Lustre

Amazon FSx for Windows File Server Getting started and documentation

Getting started

Documentation

Community Forum

Amazon FSx for Lustre Getting started and documentation

Getting started

Documentation

Community Forum

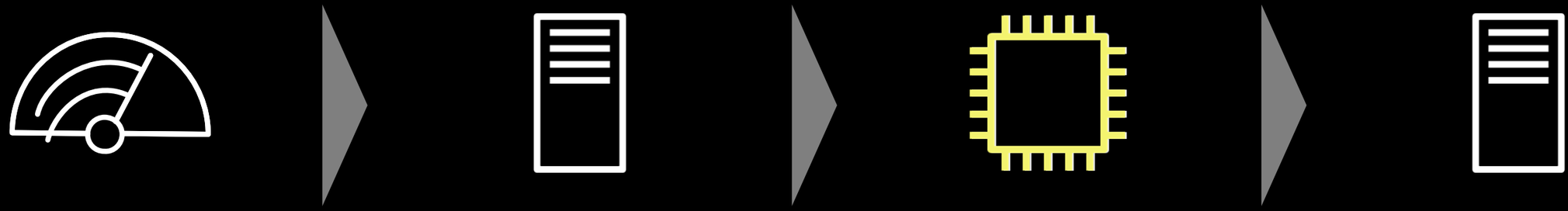
Benefits

Simple and fully managed

Fast

© 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Steps to selecting storage and compute



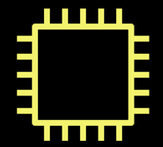
Determine throughput requirements

Define /SASDATA performance requirements

Identify optimal Amazon EC2 instances

Select storage location for /SASWORK and /UTILLOC

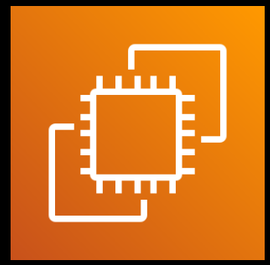
Compute node requirements



8 GB

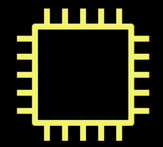
minimum memory per physical core

SAS Library	Throughput
/SASDATA	75-100 MB/s
/SASWORK	125 MB/s
/UTILLOC	125 MB/s



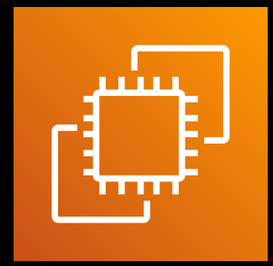
Amazon EC2

Compute node requirements



8 GB
minimum memory
per physical core

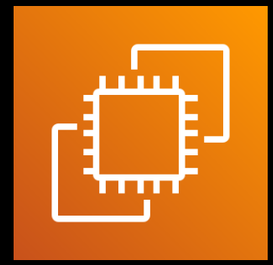
125 MB/s
minimum throughput
per physical core



Amazon EC2

Instance types vs. compute node requirements

EC2 instance type	Minimum storage (8GB)	Minimum throughput (100 MB/s)
i3en	16 GB	>100 MB/s
i3	15 GB	<100 MB/s
r5n / r5dn	16 GB	>100 MB/s
r5a / r5ad	16 GB	<100 MB/s
r6g / r6gd	16 GB	<100 MB/s
r5	16 GB	<100 MB/s
m5n / m5dn	8 GB	>100 MB/s
m5a / m5ad	8 GB	<100 MB/s
m6g / m6gd	8 GB	<100 MB/s
m5	8 GB	<100 MB/s
c5n	5.25 GB	>100 MB/s
c5a / c5ad	4 GB	<100 MB/s
c6g / c6gd	4 GB	<100 MB/s
c5	4 GB	<100 MB/s

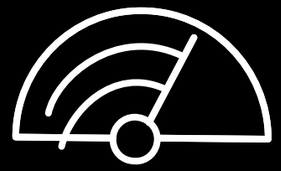


Amazon EC2

Compute node recommendations for SAS Grid

	Physical cores	-Per physical cores specification			Instance store volumes	
	(# of physical cores)	Memory (GiB)	Variable network performance (MB/s)	Consistent network performance (MB/s)	NVMe storage capacity (GiB)	Throughput performance (MB/s)
m5n	1, 2, 4, 8, 16, 24, 32, 48	8	396.18 - 1774.61	258.89	Not available	Not available
m5dn	1, 2, 4, 8, 16, 24, 32, 48	8	396.13 - 1915.05	259.17	75 – 3,600	62.48
r5n	1, 2, 4, 8, 16, 24, 32, 48	16	396.12 - 1789.06	259.05	Not available	Not available
r5dn	1, 2, 4, 8, 16, 24, 32, 48	16	396.14 - 1844.39	258.84	75 – 3,600	62.22
i3en	1, 2, 4, 6, 12, 24, 48	16	528.18 - 1745.51	259.81	1250 – 60,000	154.26

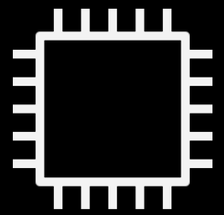
Steps to selecting storage and compute



Determine throughput requirements



Define /SASDATA performance requirements



Identify optimal Amazon EC2 instances



Select storage location for /SASWORK and /UTILLOC

Example: Storage for /SASWORK & /UTILLOC



Compute node	Cores	Memory	Storage for additional SAS libraries	Cost per instance[1]	Cost for 10 instances and FSx storage[1]
Minimum	2	64 GB	600 GB		
m5dn.2xlarge	2	32 GB			
m5n.2xlarge	2	32 GB			
r5dn.2xlarge	2	64 GB	300 GB NVMe		
i3en.2xlarge	2	64 GB	5,000 GB NVMe	$660	$6,600
r5n.2xlarge	2	64 GB	600 GB of FSX	$220	$3,945

[1] N. Virginia (us-east-1) pricing

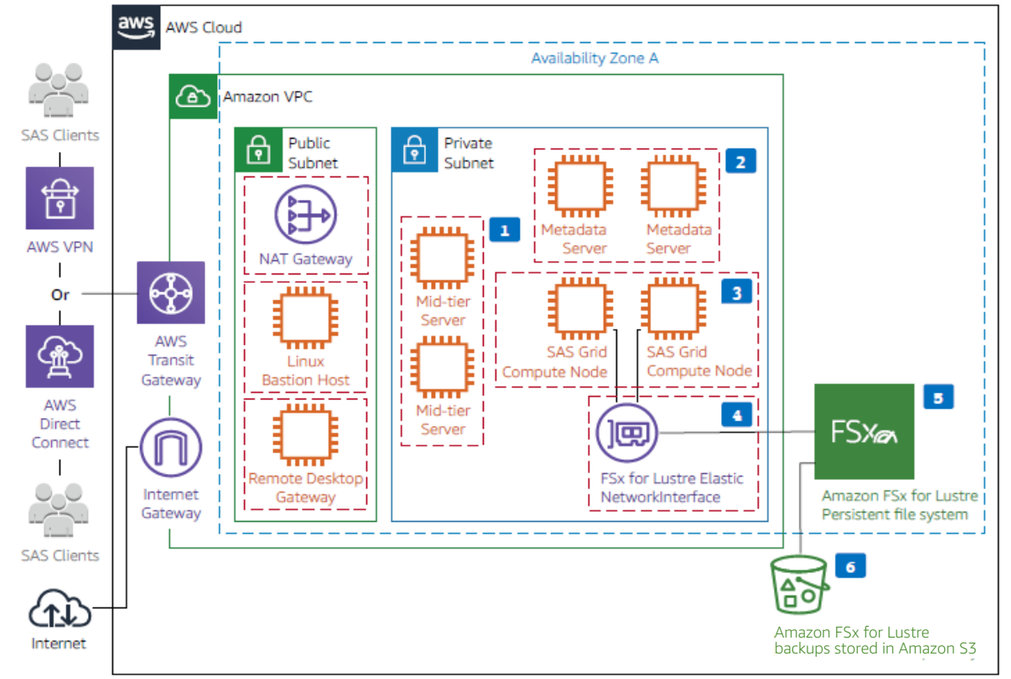
40% ***savings using FSx for /SASWORK and /UTILLOC***

Resources

Reference architecture

Running SAS® Grid on AWS

Reference architecture for deploying high-performing Amazon FSx for Lustre file system storage and guidance for the types of Amazon EC2 instances best suited for the SAS Grid compute tier.



1. For mid-tier servers, select **Amazon EC2 r5 instance types** to run Platform Web Services (PWS) and the Load Sharing Facility (LSF) client. Using two or more instances is not a SAS requirement unless you require High Availability (HA).
2. For metadata servers, select **Amazon EC2 r5 instance types** that meet or exceed minimum recommendations from SAS. Memory: the larger of 8 GB per physical core or 24 GB. Using three or more is not a SAS requirement unless you require HA.
3. For SAS Grid compute nodes running the LSF platform, select **Amazon EC2 instance types** that meet or exceed minimum recommendations from SAS. Memory: 8 GB per physical core. File system performance (network for **Amazon FSx for Lustre**): 100-125 MB/s per physical core. We recommend the **Amazon EC2 m5n and r5n instance types** if hosting /SASDATA, /SASWORK, and /UTILLOC on **Amazon FSx for Lustre**. Use **i3en instance types** if offloading /SASWORK to local instance store volumes.
4. The **Amazon FSx for Lustre** file system is accessible through an elastic network interface that resides in your virtual private cloud (VPC). Use standard **Amazon VPC** security groups to control network access to your **Amazon FSx for Lustre** file system.
5. Use **Amazon FSx for Lustre** persistent file systems for all SAS Grid libraries, including /SASDATA, /SASWORK, and /UTILLOC. **Amazon FSx for Lustre** is a fully managed open-source Lustre file system capable of hundreds of gigabytes per second throughput, millions of IOPS, and submillisecond latencies.
6. For additional data protection, Amazon FSx for Lustre persistent file systems offer support for point-in-time backups stored on Amazon S3

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Reference Architecture

Migration guide

AWS Prescriptive Guidance

Migrating SAS Grid to the AWS Cloud

Recap

- Many customers are migrating SAS Grid applications to AWS to take advantage of the agility and cost savings of running these workloads in the cloud
- Application-specific compute and storage requirements can be easily mapped to AWS storage and compute resources
  - Amazon FSx for Lustre provides fully managed high-performance file systems that scale to meet your SAS Grid throughput needs; persistent SSD storage is the recommended FSx storage offering for all SAS Grid libraries
  - Amazon EC2 provides several types of compute instances to meet SAS Grid requirements (r5n, r5dn, m5n, m5dn, i3en)
- AWS provides comprehensive resources to support SAS Grid migrations to the cloud

Go to aws.amazon.com/fsx for details and next steps

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.