



# Scale your Studio

GPU rendering in the Cloud:

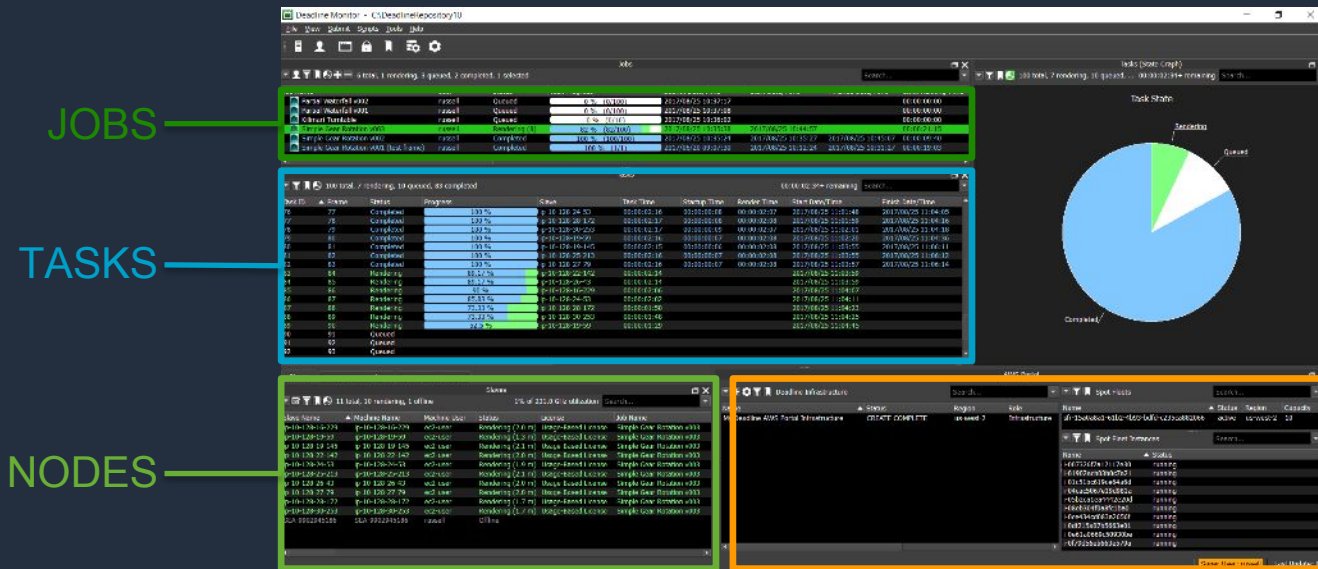
GTC 2019 / S9439

Chris Bond, Director AWS EC2, AWS Thinkbox Founder

Feb 27, 2018

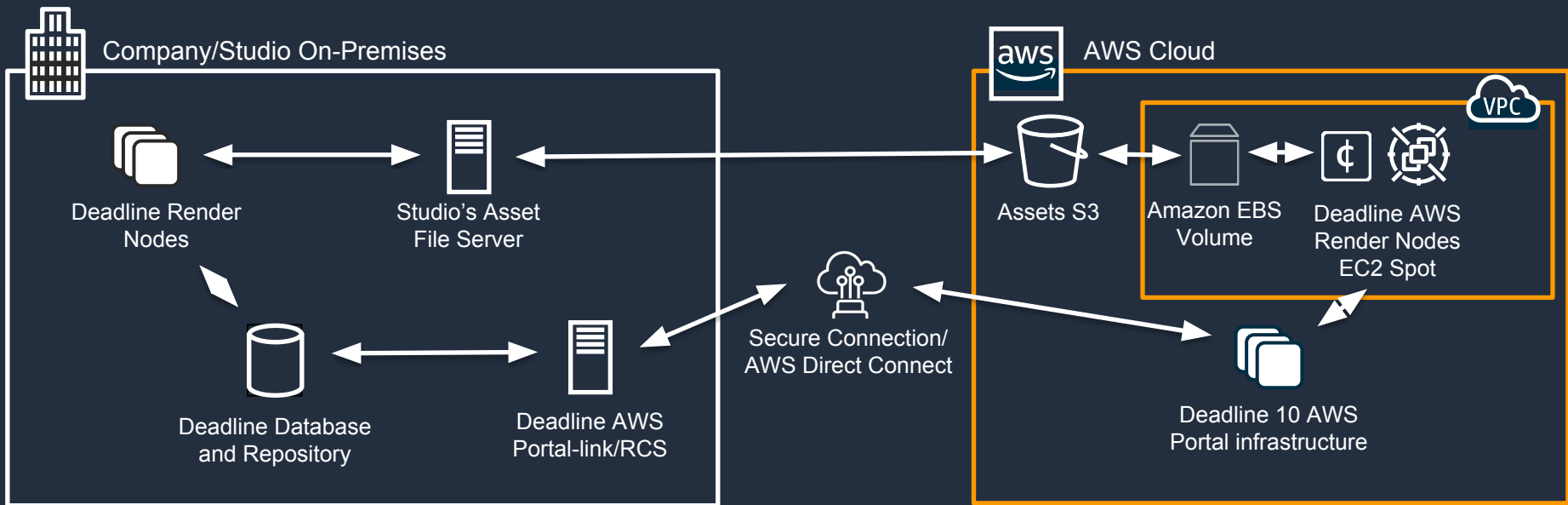
# Rendering at scale

## Deadline 10: AWS Portal



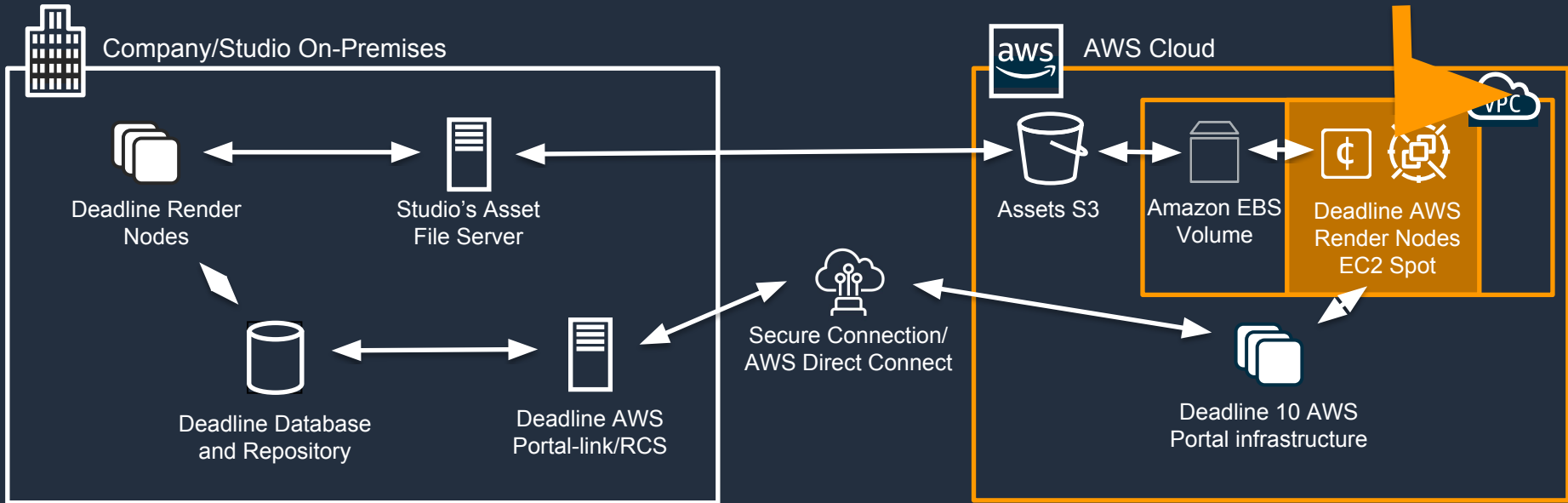
· AWS

# Customer example: Hybrid rendering pipeline



# Customer example: Hybrid rendering pipeline

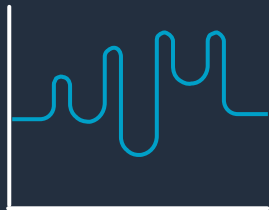
Including  
GPU!



# Amazon Elastic Cloud Compute (Amazon EC2)

## On-demand

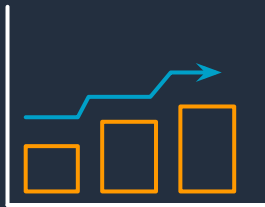
Pay for compute capacity by **the second** with no long-term commitments



Spiky workloads,  
to define needs

## Reserved instances

Make a 1- or 3-year commitment and receive a **significant discount** off on-demand prices



Committed &  
steady-state usage

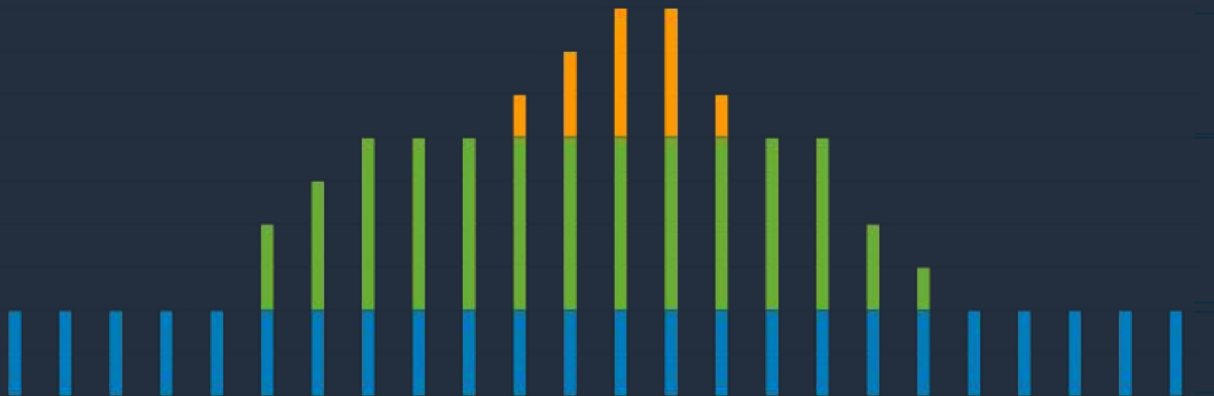
## Spot Instances

Spare Amazon EC2 capacity at **savings of up to 90%** off on-demand prices



Fault-tolerant, flexible,  
stateless workloads

# To optimize Amazon EC2, combine purchase options

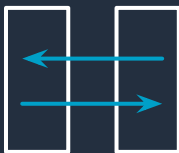


Scale using **Spot** for fault-tolerant, flexible, stateless workloads such as **Rendering**

**On-demand**, for unknown spiky workloads such as **Workstations**

Use **RIs** for known, steady-state workloads, such as **Infrastructure**

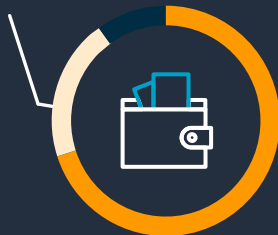
# The simple rules of Spot Instances



## Spot infrastructure

Is same as on-demand and RIs: same hardware, same capabilities, same choice of instances, same geo footprint (regions & AZs)

70–90% off



## Spot pricing

Set it and forget it pricing  
—no bidding, big savings



## Diversify

Use diverse instance fleets  
to spin up 1MM core  
(60k+ instances) clusters



# How do I use Spot in Deadline?

Deadline 10 helps you  
select GPU / CPU instances!

Spot Fleet Configuration

Spot Fleet Settings

Maximum Price (USD): 2.0000 Target Capacity: 100

Instance Type Presets: Small (8 CPU 15GB RAM) Medium (16 CPU 30GB RAM) Large (32 CPU 60GB RAM) XLarge (64 CPU 120GB RAM) GPU

Key Pair: ☒ Use Key Pair DashKey

Software Instance Types Slave Settings AWS Settings Tags

Configuration

Current Spot Price for your instance types in us-west-2: Linux: \$7.9785 (USD) Windows: \$10.288 (USD)

Spot Price is the maximum amount of money that you are willing to pay for an hour of instance time. The current price of your most expensive instance type is shown above. If the current price rises above your Spot Price, your instances will be terminated. For more information, see [Spot Instances](#).

Instance Types

Instance Type	Linux	Windows
f1.16xlarge	\$13.2	\$0.0
f1.2xlarge	\$0.495	\$0.0
f1.4xlarge	\$0.99	\$0.0
g2.2xlarge	\$0.195	\$0.312
g2.8xlarge	\$0.78	\$1.058
h1.16xlarge	\$1.32	\$4.264
h1.2xlarge	\$0.165	\$0.533
h1.4xlarge	\$0.33	\$1.066
h1.8xlarge	\$0.66	\$2.132
i2.2xlarge	\$0.5115	\$0.7525
i2.4xlarge	\$1.023	\$1.504
i2.8xlarge	\$2.046	\$3.008
i2.xlarge	\$0.2559	\$0.3759
i3.16xlarge	\$1.4976	\$4.4416
i3.2xlarge	\$0.1894	\$0.5522

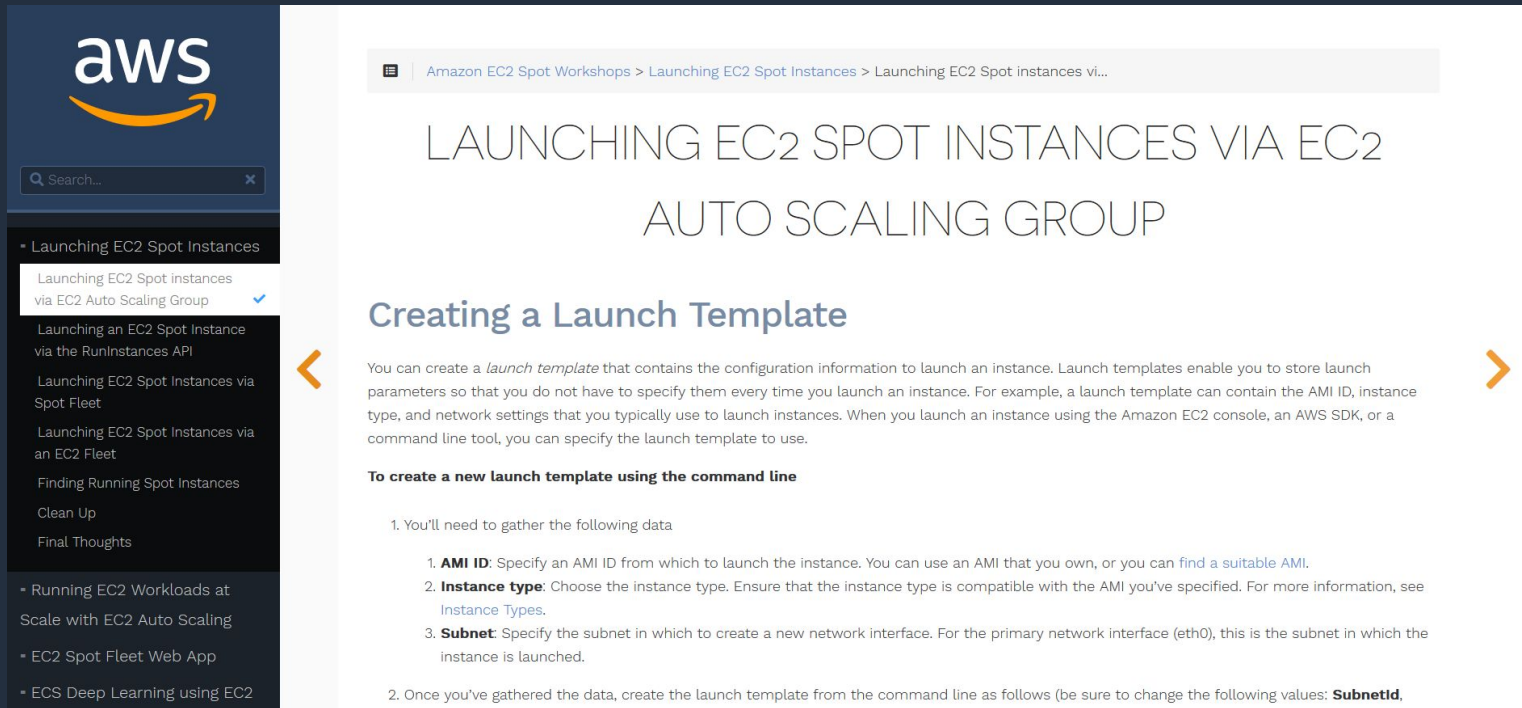
Instance Type	Linux	Windows
g3.16xlarge	\$1.368	\$4.312
g3.4xlarge	\$0.3949	\$1.078
g3.8xlarge	\$0.684	\$2.156
g3s.xlarge	\$0.225	\$0.409
p3.16xlarge	\$7.9785	\$10.288
p3.2xlarge	\$0.9765	\$1.286
p3.8xlarge	\$4.3387	\$5.144

Click here for more information on [Instance Types](#).

☒ Auto Adjust Deadline Performance Settings Preview Auto Adjust

Launch Cancel

# How do I use Spot directly?



aws

Search...

- Launching EC2 Spot Instances
  - Launching EC2 Spot instances via EC2 Auto Scaling Group ✓
  - Launching an EC2 Spot Instance via the RunInstances API
  - Launching EC2 Spot Instances via Spot Fleet
  - Launching EC2 Spot Instances via an EC2 Fleet
  - Finding Running Spot Instances
  - Clean Up
  - Final Thoughts
- Running EC2 Workloads at Scale with EC2 Auto Scaling
- EC2 Spot Fleet Web App
- ECS Deep Learning using EC2

Amazon EC2 Spot Workshops > Launching EC2 Spot Instances > Launching EC2 Spot instances vi...

## LAUNCHING EC2 SPOT INSTANCES VIA EC2 AUTO SCALING GROUP

### Creating a Launch Template

You can create a *launch template* that contains the configuration information to launch an instance. Launch templates enable you to store launch parameters so that you do not have to specify them every time you launch an instance. For example, a launch template can contain the AMI ID, instance type, and network settings that you typically use to launch instances. When you launch an instance using the Amazon EC2 console, an AWS SDK, or a command line tool, you can specify the launch template to use.

**To create a new launch template using the command line**

1. You'll need to gather the following data
  1. **AMI ID:** Specify an AMI ID from which to launch the instance. You can use an AMI that you own, or you can [find a suitable AMI](#).
  2. **Instance type:** Choose the instance type. Ensure that the instance type is compatible with the AMI you've specified. For more information, see [Instance Types](#).
  3. **Subnet:** Specify the subnet in which to create a new network interface. For the primary network interface (eth0), this is the subnet in which the instance is launched.
2. Once you've gathered the data, create the launch template from the command line as follows (be sure to change the following values: **SubnetId**,

<https://ec2spotworkshops.com/>

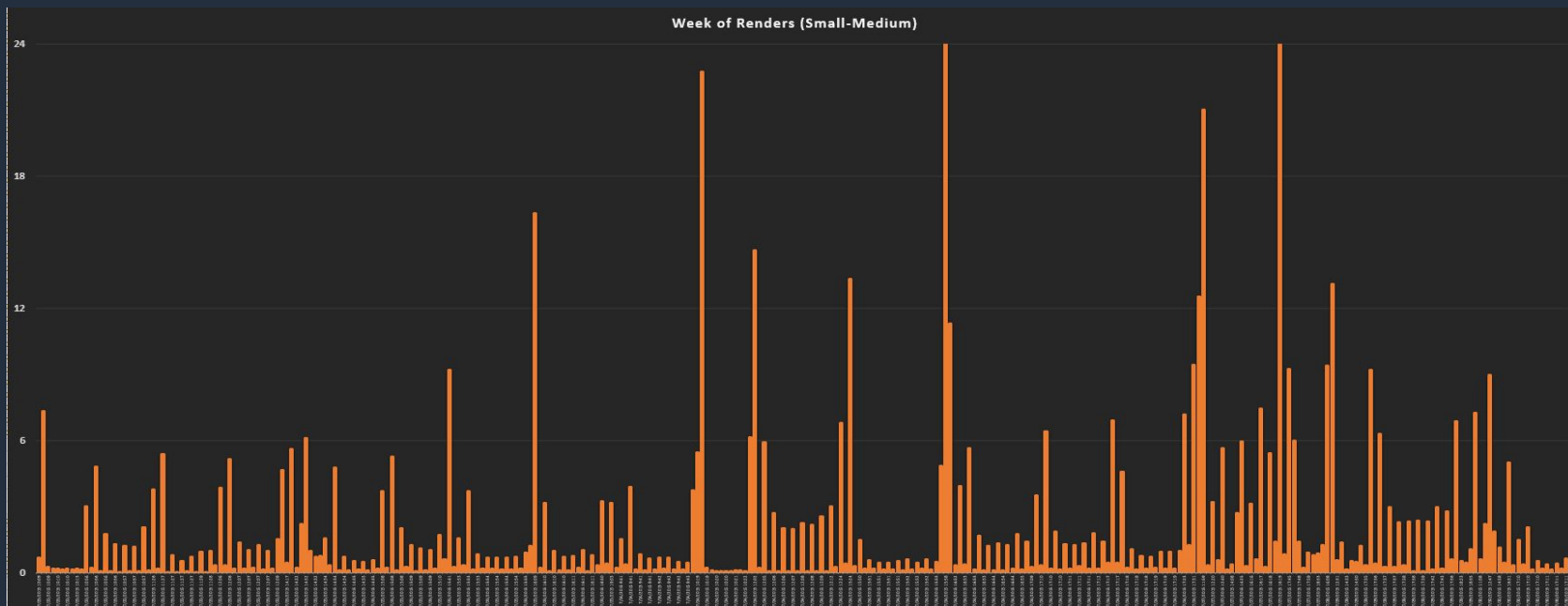
# Customer trends

# Customer trends: Early assumptions

Small-Medium studio

Peaks and valleys over a week of production

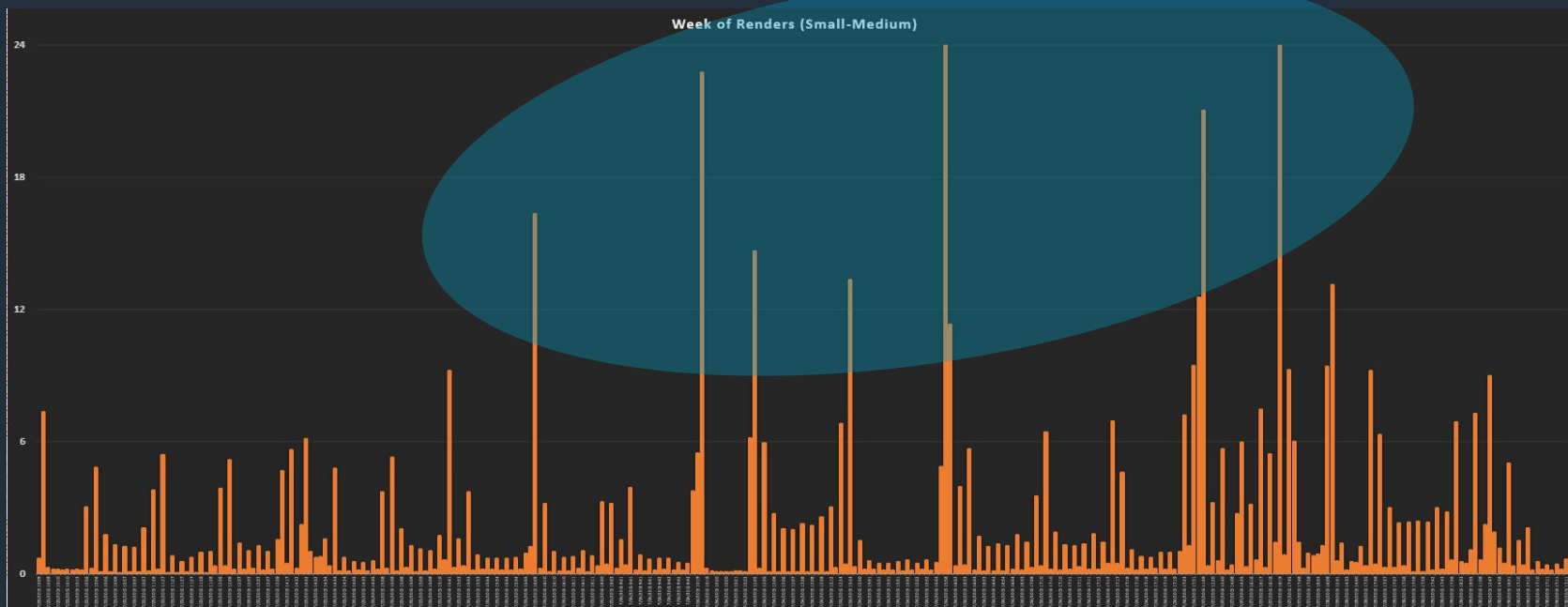
Renders peaking at just over 24 hours



# Customer trends: Early assumptions

Small-Medium studio

Customer told us they wanted to reduce these peaks with AWS Cloud rendering

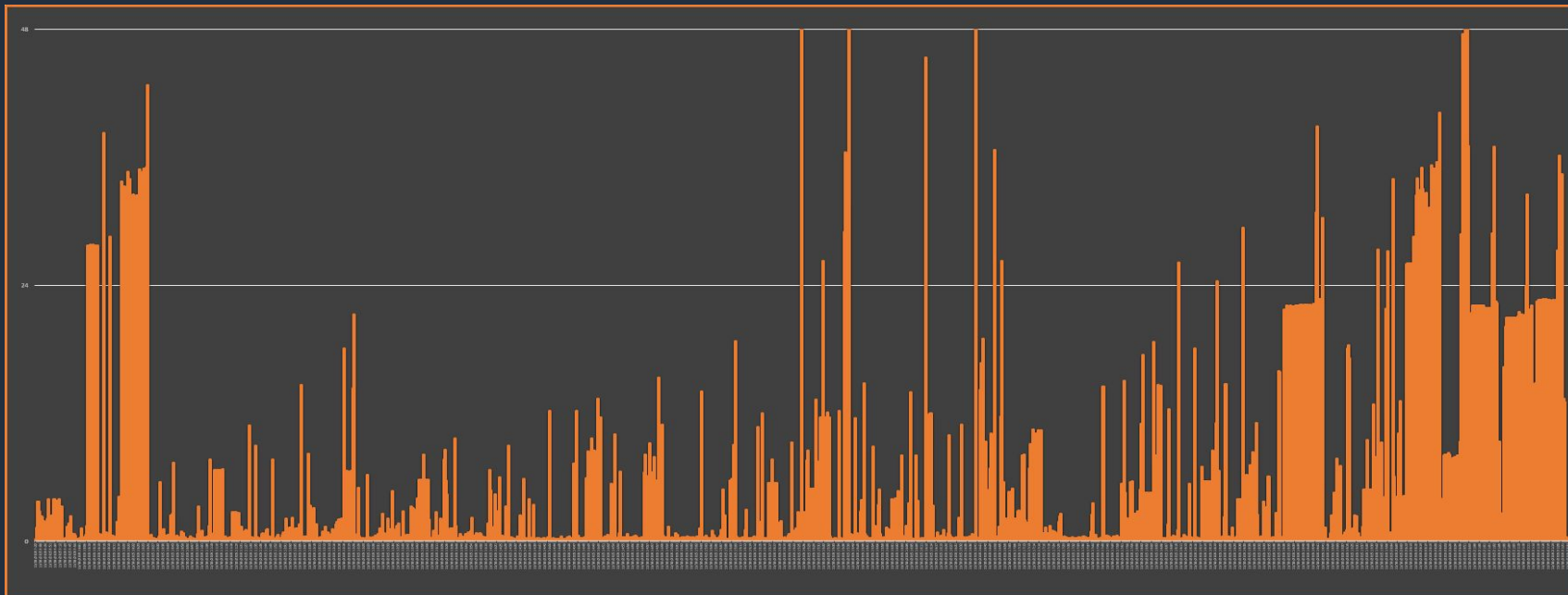


# Customer trends: Early assumptions

Large studio

Submission over a **day** of production

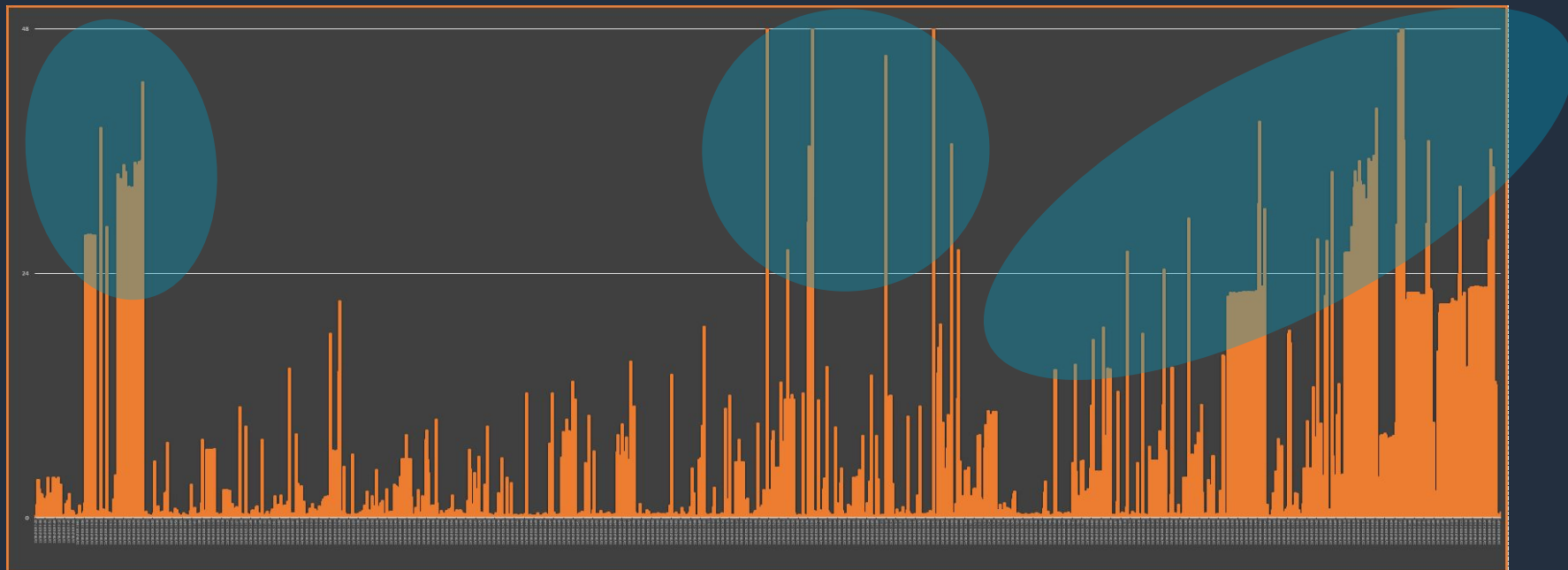
Renders peaking at **over 48 hours**





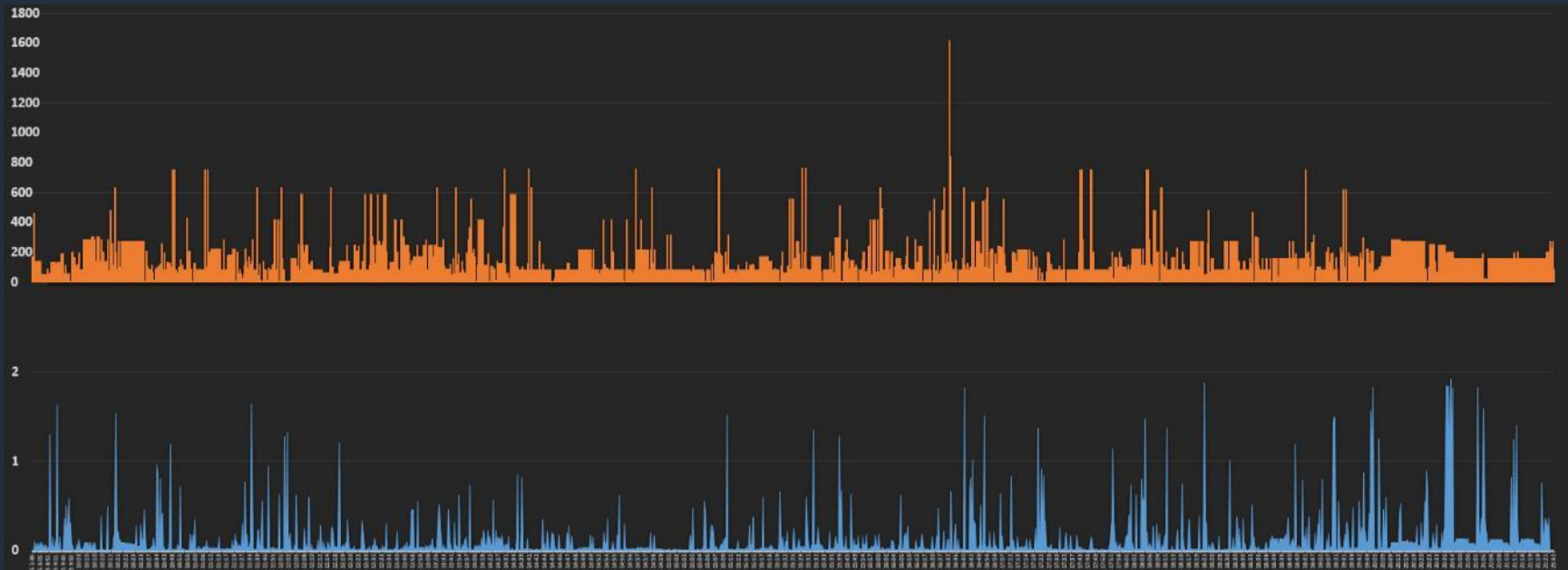
# Customer trends: Early assumptions

What Chris told his boss



# Customer trends: More capacity

With more capacity, customers can iterate more frequently  
Artists spend more time creating







Milk Visual Effects:  
Creating massive ocean and storm  
simulations for Adrift

Adrift, image courtesy of Milk Visual Effects

### Problem statement:

- Biggest creative and technical challenge Milk has undertaken
- Fluid simulations require a LOT of data (approx. 100TB per shot)
- Rendering oceans means caching/baking every frame of the shot
- Needed a rendering solution that could scale

### Use of AWS Thinkbox Deadline & Amazon EC2 Spot:

- Used Deadline to manage on-premises and EC2 Spot Instances
- Peaked at 130,000 cores, averaged 80,000 cores per day
- Deadline's flexibility allowed for custom development for their pipeline
- 10-week render job

### Business benefits:

- Scalability and elasticity of rendering with AWS enabled a small team to complete the project on time
- Milk was able to punch above their weight on a project that typically requires a much bigger team

# Customer trends

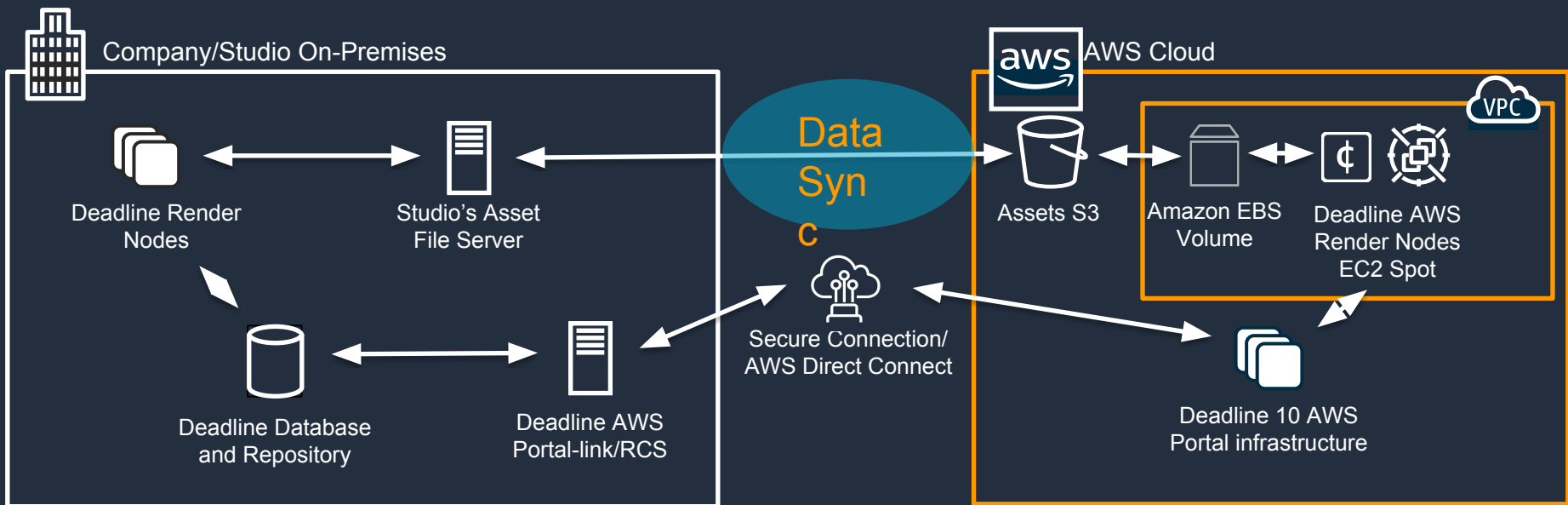
Customers are scaling **2x-10x (77x!)** their on-prem capacity

They are saying “**yes**” more frequently

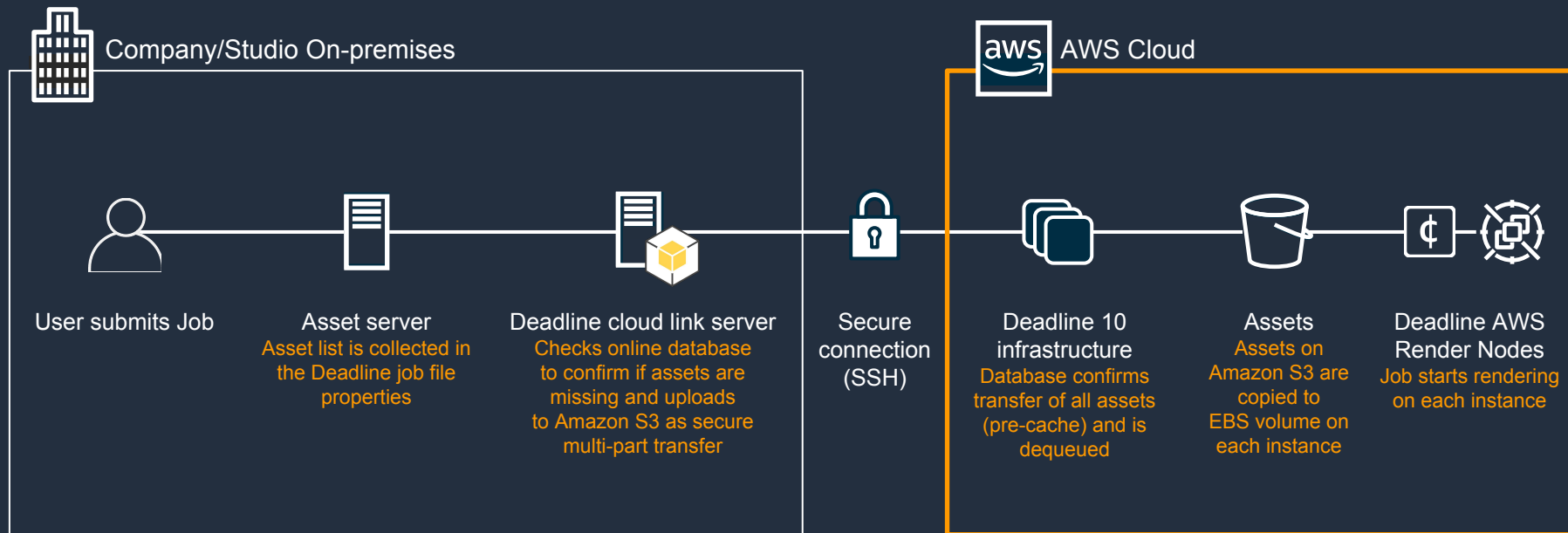
Time becomes “**elastic**”

# Infrastructure and Software

# Deadline 10 infrastructure



# Deadline 10 infrastructure



# Infrastructure flexibility

## Deadline 10

Supports hybrid  
workflows out of box  
Using Maya, 3DSMAX



Amazon S3 synchronization  
to Amazon EBS on each  
instance

## Partner

Mix of Windows/Linux clients  
Scales to billions of files

# QUMULO

High-performance  
cloud file system

## Partner

High performance for HPC  
Can leverage Amazon S3  
or Amazon Glacier

# WEKA IO

High-performance  
cloud file system

# Amazon FSx for Windows File Server

Lift and shift your Windows file storage with fully managed windows file servers



Native Windows  
compatibility



Fast and flexible  
performance



Enterprise-ready



Broad accessibility



Fully managed

# Amazon FSx for Lustre

Fully managed Lustre file system for compute-intensive workloads



Massively scalable  
performance



Seamless access to  
your data repositories



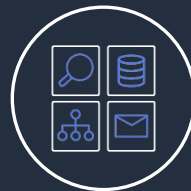
Simple  
and fully managed



Native file  
system interface



Cost-optimized for  
compute-intensive workloads



Secure  
and compliant

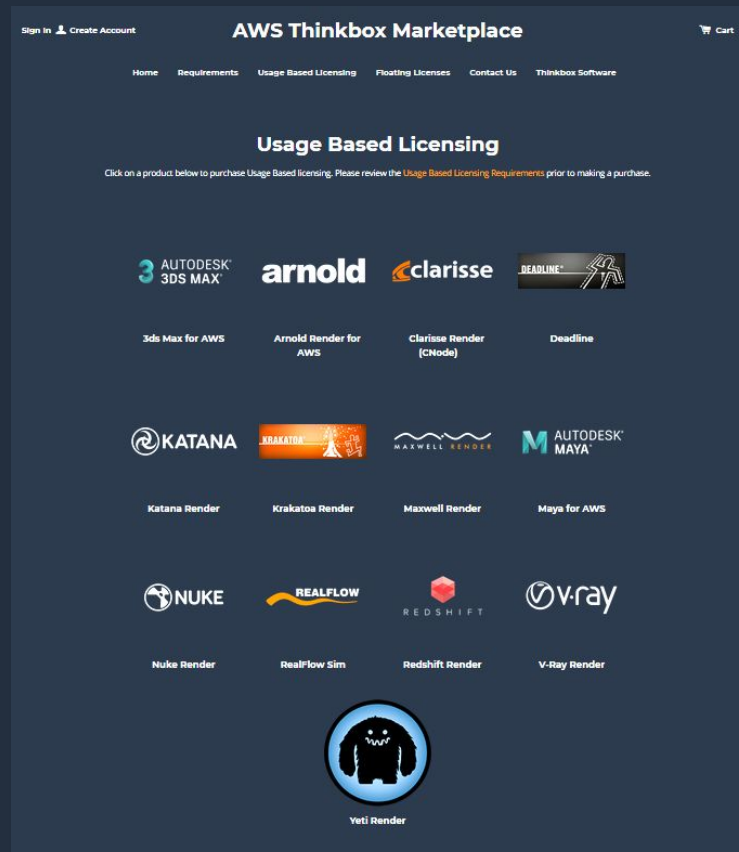


# AWS Thinkbox Marketplace

➤ Usage-based licensing (UBL)

➤ Per-minute licensing for applications

<https://marketplace.thinkboxsoftware.com>



# Workstations

# Graphics instance

## EC2 G3 instance type

MODEL	vCPU	RAM	GPU	MEMORY	CUDA	OD price/hr (Linux)	OD price/hr (Win)	SPOT price/hr (Linux)	SPOT price/hr (Win)
g3s.xlarge	4	30.5	1	8	2048	\$1.04	\$1.224	\$0.312	\$0.496
g3.4xlarge	16	122	1	8	2048	\$1.58	\$2.316	\$0.474	\$1.21
g3.8xlarge	32	244	2	16	4096	\$3.16	\$4.632	\$1.0836	\$2.42
g3.16xlarge	64	488	4	32	8192	\$6.32	\$9.264	\$6.32	\$9.264

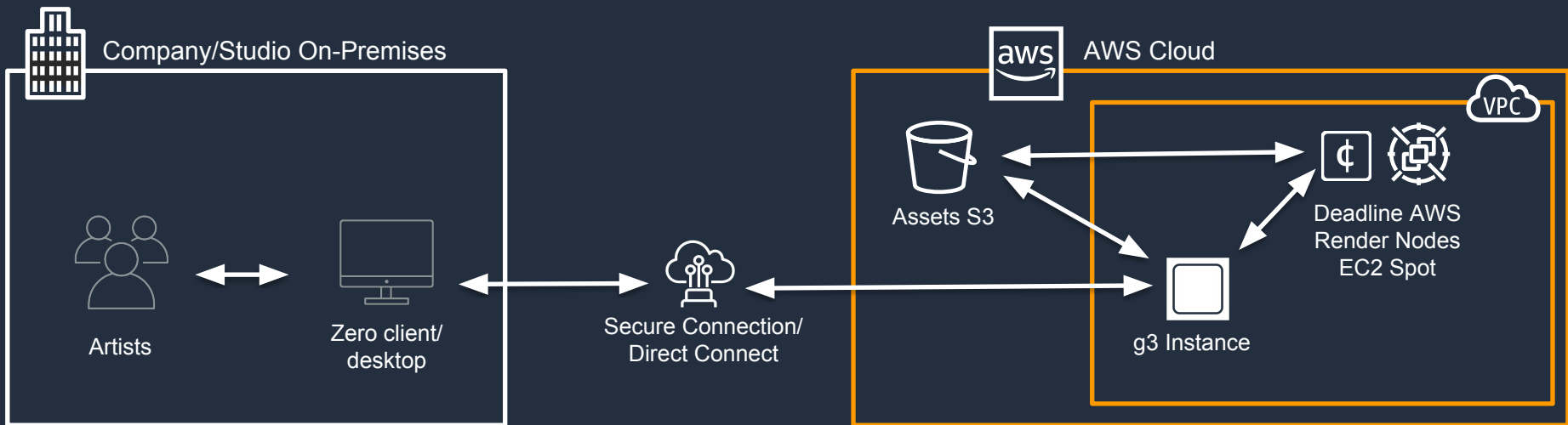
- All based on NVidia Tesla M60
- g3.4xl is half of one M60 with 2048 CUDA cores and 8GB MEM
- Multiple GPUs for CUDA processing/rendering, does not enhance viewport performance
- Supports up to 4 x monitors with max 4096 x 2160 resolution

# G4 Instances Preview!

- AWS-custom Intel CPUs (4 to 96 vCPUs)
- 1 to 8 NVIDIA T4 Tensor Core GPUs
- Up to 384 GiB of memory
- Up to 1.8 TB of fast, local NVMe storage
- Up to 100 Gbps networking

<https://pages.awscloud.com/ec2-g4-preview.html>

# Simplified studio topology



# AWS Workstation Quick Starts

AWS Quick Starts Amazon Connect Integrations FAQs Resources

## REFERENCE DEPLOYMENT

### Visual Effects Workstations on AWS

Deploy a cloud environment for VFX workstations with G3 instances and Teradici software

[View deployment guide](#)

This Quick Start deploys a highly available environment for Microsoft Windows-based, visual effects (VFX) workstations on the Amazon Web Services (AWS) Cloud.

The deployment uses G3 GPU instances, which are designed for graphics-intensive workloads. Amazon Simple Storage Service (Amazon S3) provides highly durable, secure, and scalable storage for VFX content. Teradici's PC-over-IP (PCoIP) technology and Teradici Cloud Access Software enable a powerful remote desktop experience, and you can set up AWS Direct Connect for low-latency network connections back to your studio.

This reference architecture is automated by AWS CloudFormation templates that deploy the environment on AWS in about 30 minutes. You can customize the templates to meet your specific requirements. After you deploy the Quick Start, you can install your preferred software tools.

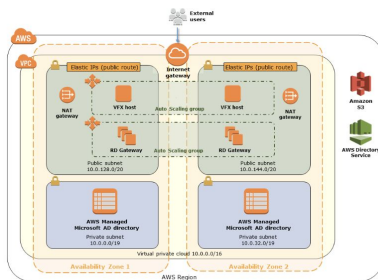
teradici.

This Quick Start was developed by Teradici in collaboration with AWS. Teradici is an APN Partner.

[What you'll build](#) [How to deploy](#) [Cost and licenses](#)

This Quick Start sets up the following environment on AWS:

- A highly available architecture that spans two Availability Zones.\*
- A virtual private cloud (VPC) configured with public and private subnets according to AWS best practices, to provide you with your own virtual networking environment, including selection of your own IP address range, creation of subnets, and configuration of route tables and network gateways.\*
- An internet gateway to allow access to the Internet.\*
- In the public subnets, managed network address translation (NAT) gateways that allow outbound internet access to resources in the private subnets, but that prevent the Internet from accessing those instances.\*
- In the public subnets, one or more Remote Desktop Gateway (RD Gateway) instances in an Auto Scaling group, to provide readily available administrative access to the environment and secure access to Microsoft Windows instances located in the private and public subnets. The RD Gateway instances use the Remote Desktop Protocol (RDP) over HTTPS to establish a secure, encrypted connection between remote users on the Internet and Windows-based EC2 instances, without needing to configure a virtual private network (VPN) connection. This helps reduce the attack surface on your Windows-based instances and provides a remote administration solution for administrators.\*



[Switch to full-screen view](#)

AWS Quick Starts Amazon Connect Integrations FAQs Resources

## REFERENCE DEPLOYMENT

### Cloud Video Editing on AWS

Deploy a cloud video editing environment using AWS services

[View deployment guide](#)

This Quick Start deploys a highly available architecture for cloud video editing on the Amazon Web Services (AWS) Cloud.

The deployment uses Amazon Simple Storage Service (Amazon S3), Amazon Elastic Compute Cloud (Amazon EC2), Amazon Virtual Private Cloud (Amazon VPC), AWS Directory Service, and Remote Desktop Gateway (RD Gateway) instances.

This reference architecture is automated by AWS CloudFormation templates that deploy the environment on AWS in about 30 minutes. You can customize the templates to meet your specific requirements.

teradici.

This Quick Start was developed by Teradici in collaboration with AWS. Teradici is an APN Partner.

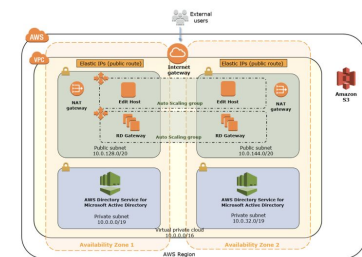
[What you'll build](#) [How to deploy](#) [Cost and licenses](#)

This Quick Start architecture includes the following infrastructure:

- A virtual private cloud (VPC) that is configured across two Availability Zones. For each Availability Zone, this Quick Start provisions one public subnet and one private subnet, according to AWS best practices.
- In the public subnets, Edit Host instances and Remote Desktop Gateway (RD Gateway) instances in an Auto Scaling group for secure access to EC2 instances in the private subnets.
- In the public subnets, managed network address translation (NAT) gateways to provide outbound internet connectivity for instances in the private subnets.
- An AWS Identity and Access Management (IAM) role to enable AWS resources created through the Quick Start to access other AWS resources when required.

The Quick Start gives you the option to build a new VPC infrastructure with these components or to use your existing VPC infrastructure. Within this infrastructure, the Quick Start deploys:

- In the private subnets, AWS Directory Service for Microsoft Active Directory, which provides most of the features that Microsoft Active Directory offers, plus integration with AWS applications.
- An Amazon Elastic Compute Cloud (Amazon EC2) G3 instance, which offers a powerful, low-cost, pay-as-you-go model for high-end workstations.
- Amazon Simple Storage Service (Amazon S3), which is an object store that provides artifacts necessary for the Quick Start. You can also use it to upload video for editing on

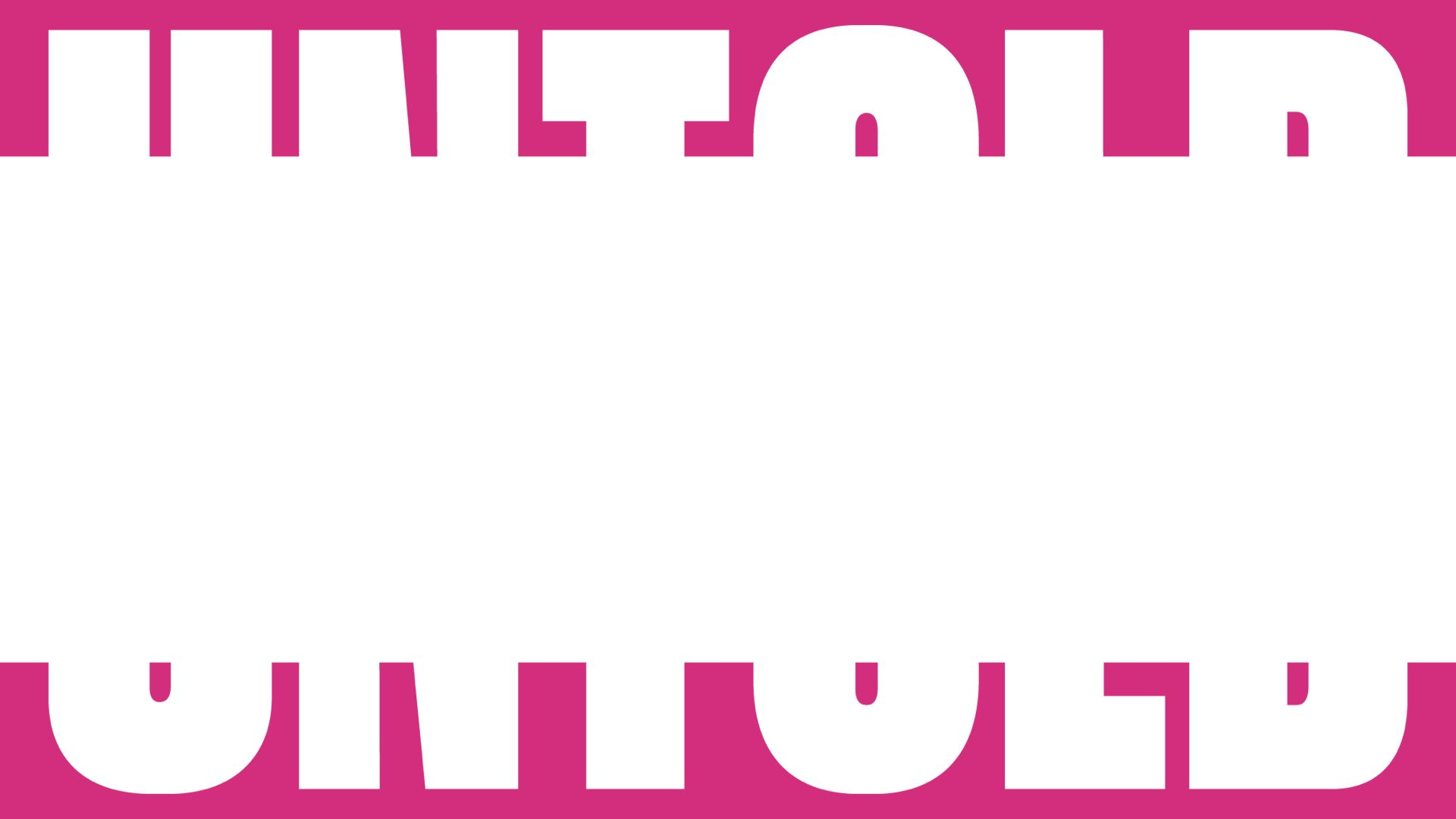


[Switch to full-screen view](#)

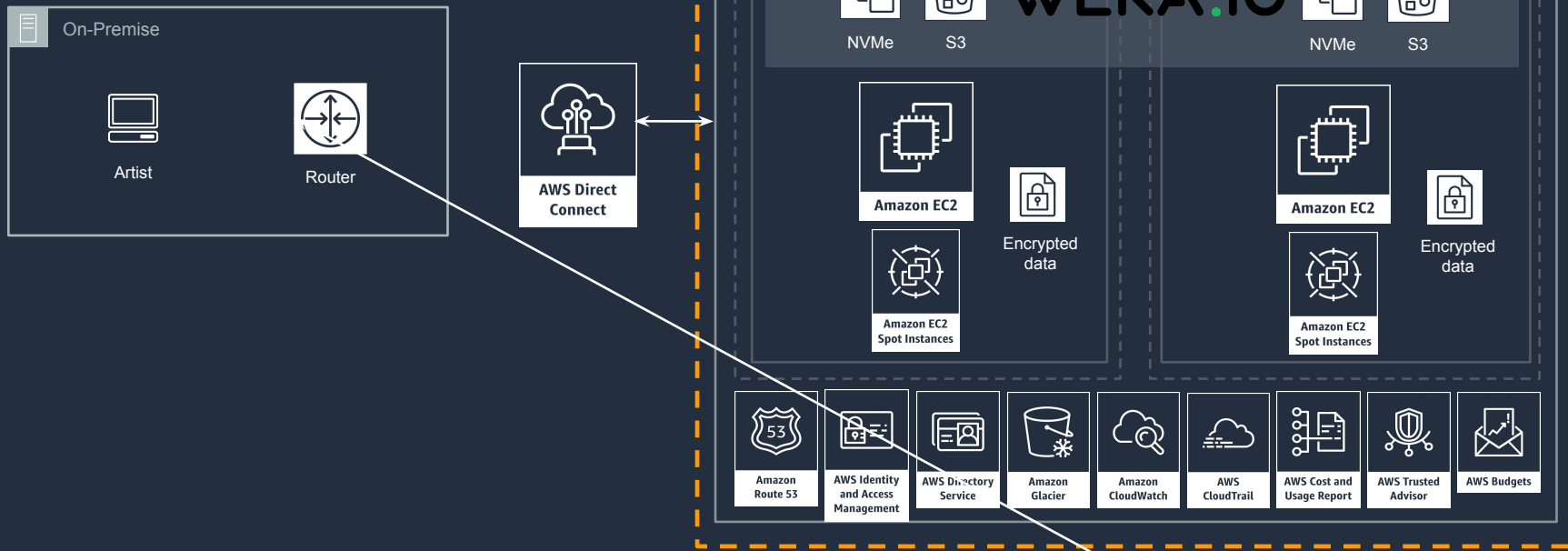
[View deployment guide for details](#)



# Customer Spotlight: Untold Studios







## "THE MACHINE ROOM"

- Diverse Fibre going over different PoP's to AWS.
- 2 x 10Gbps Circuits.
- Redundant CPE's, Firewalls & Routers.







# Thank you!!

[www.thinkboxsoftware.com](http://www.thinkboxsoftware.com)

<https://ec2spotworkshops.com/>

<https://aws.amazon.com/quickstart/architecture/vfx-workstations-with-teradici/>

Chris Bond, Director AWS EC2, AWS Thinkbox Founder