

AWS  
re:Invent

C M P 2 0 3

# Studio in the Cloud: Content production on AWS

**Jason Schleifer**

Creative Director

Amazon Web Services

**Rick Grandy**

Sr. Solutions Architect

NVIDIA



# Studio in the Cloud

Rendering

Software

Virtual Workstations

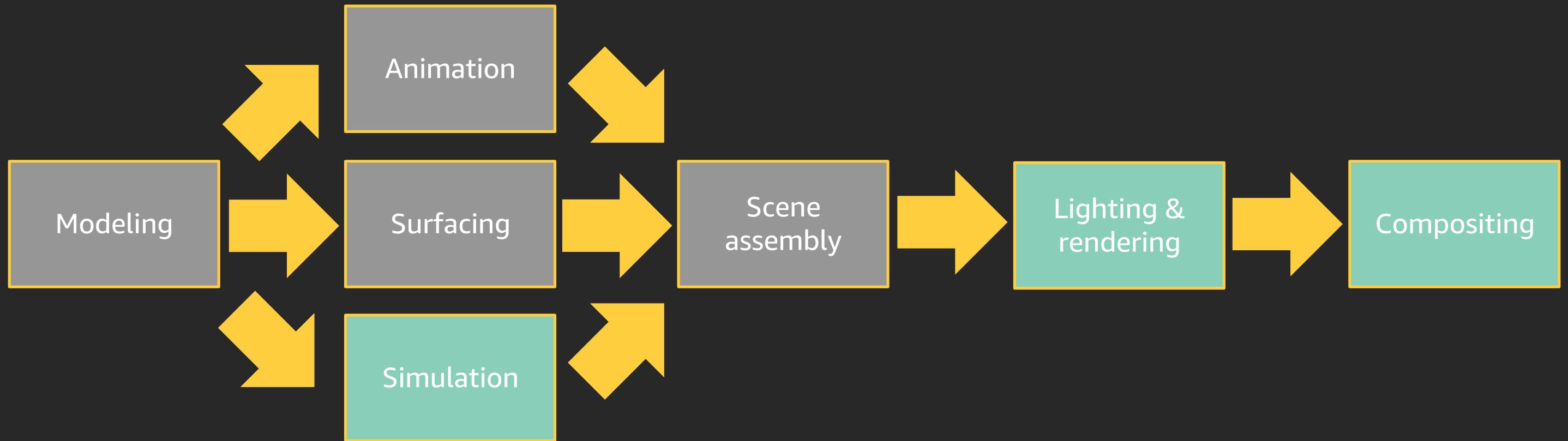
Infrastructure

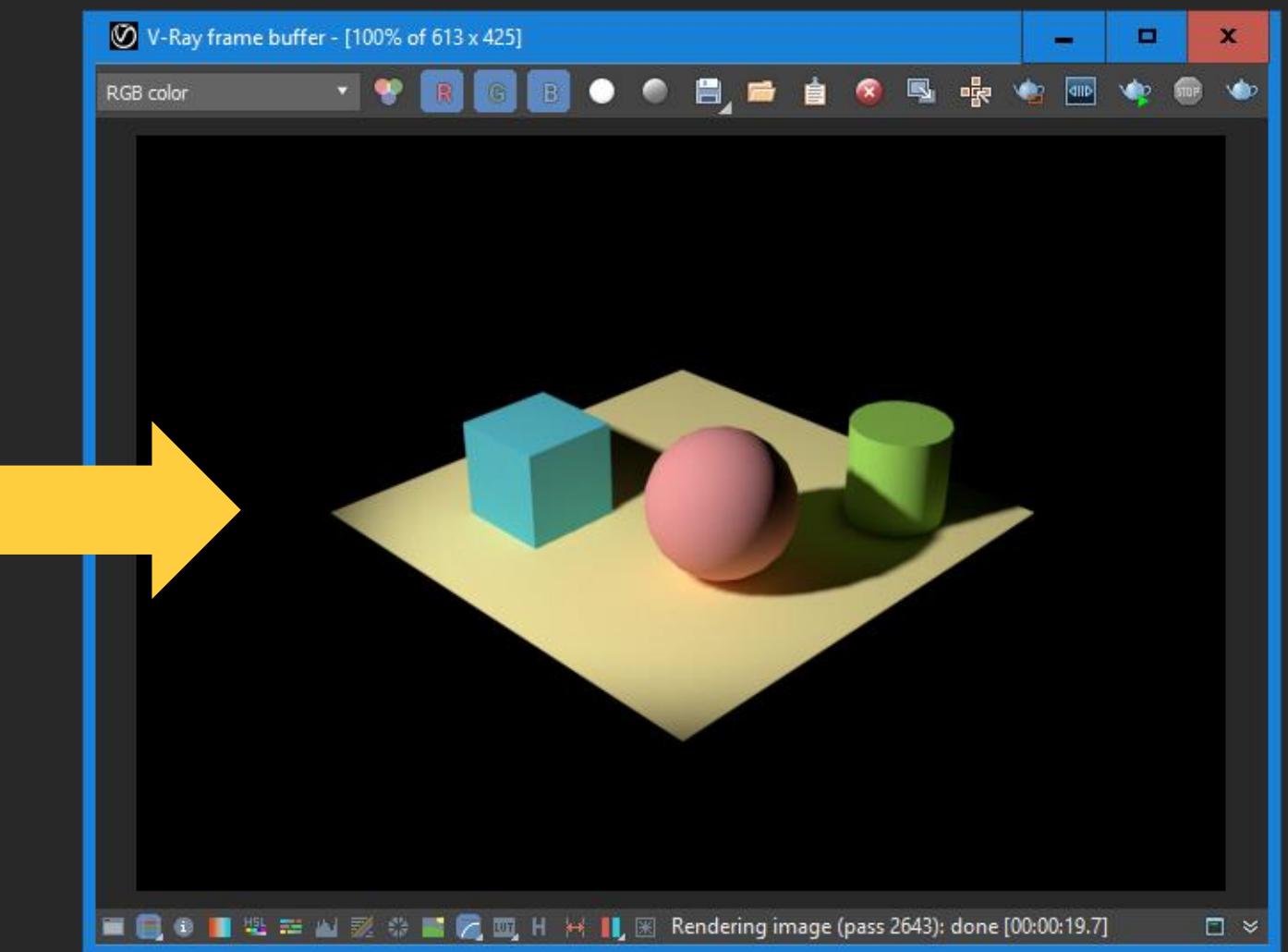
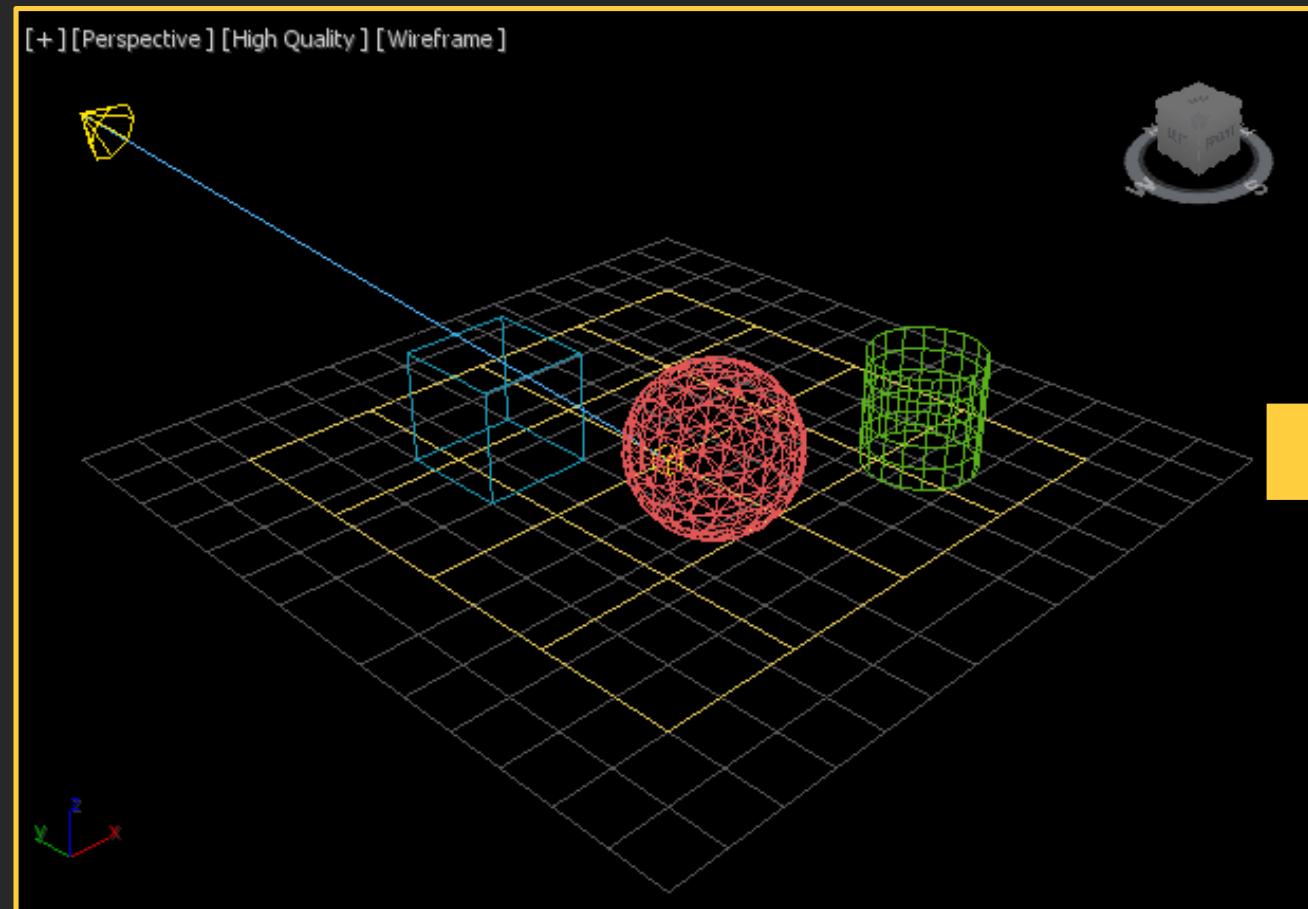
# Animation & VFX



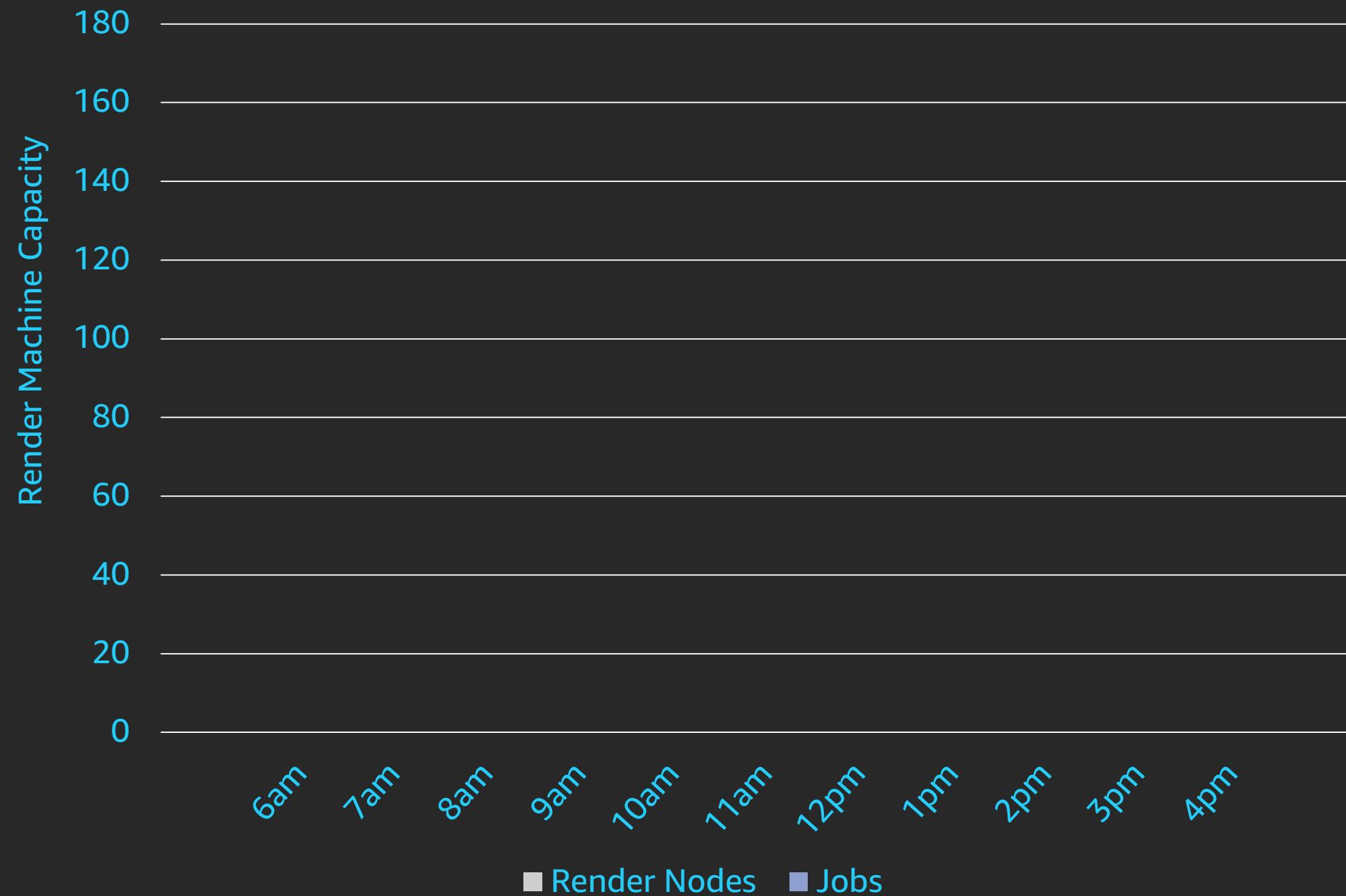
milk

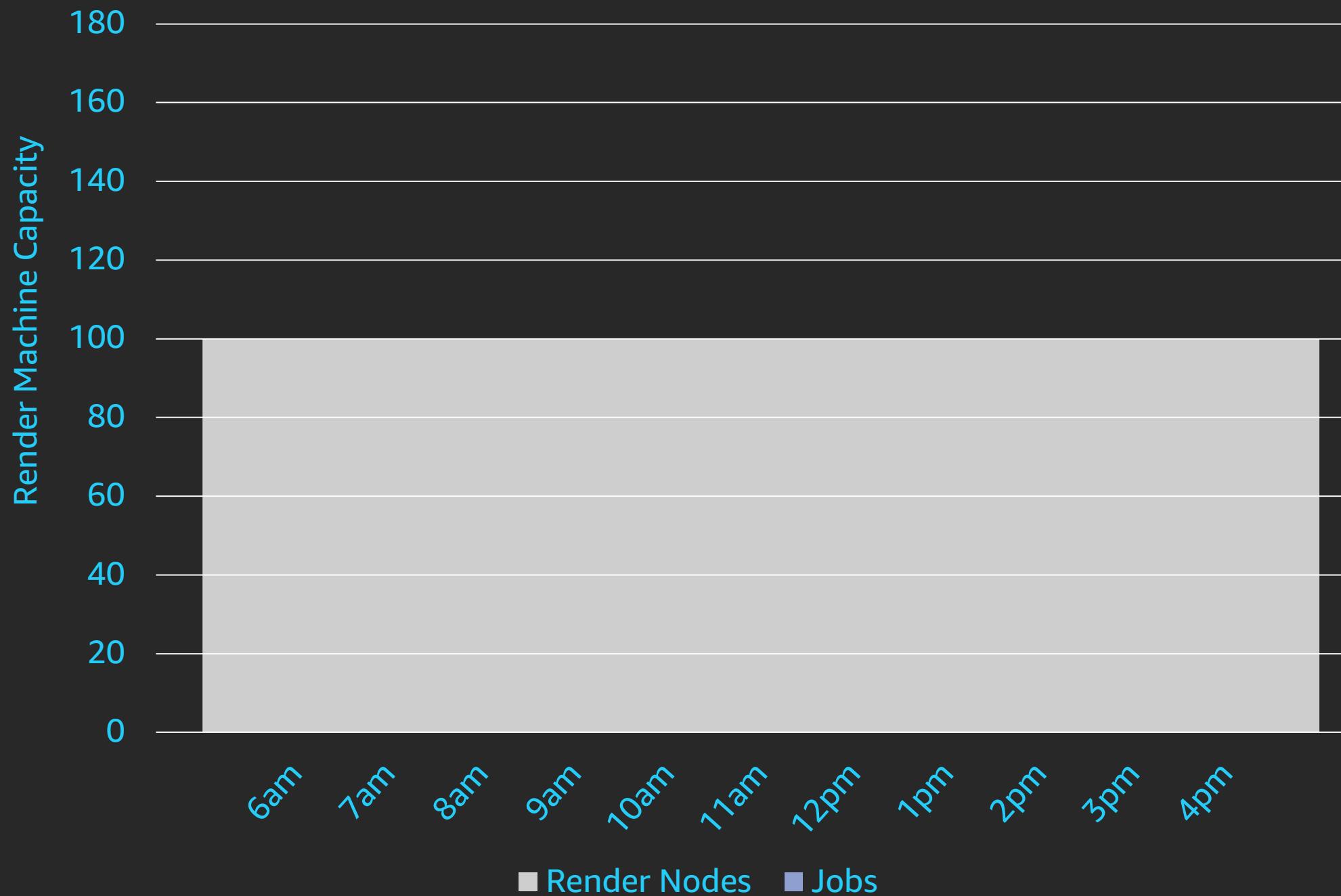
# Rendering

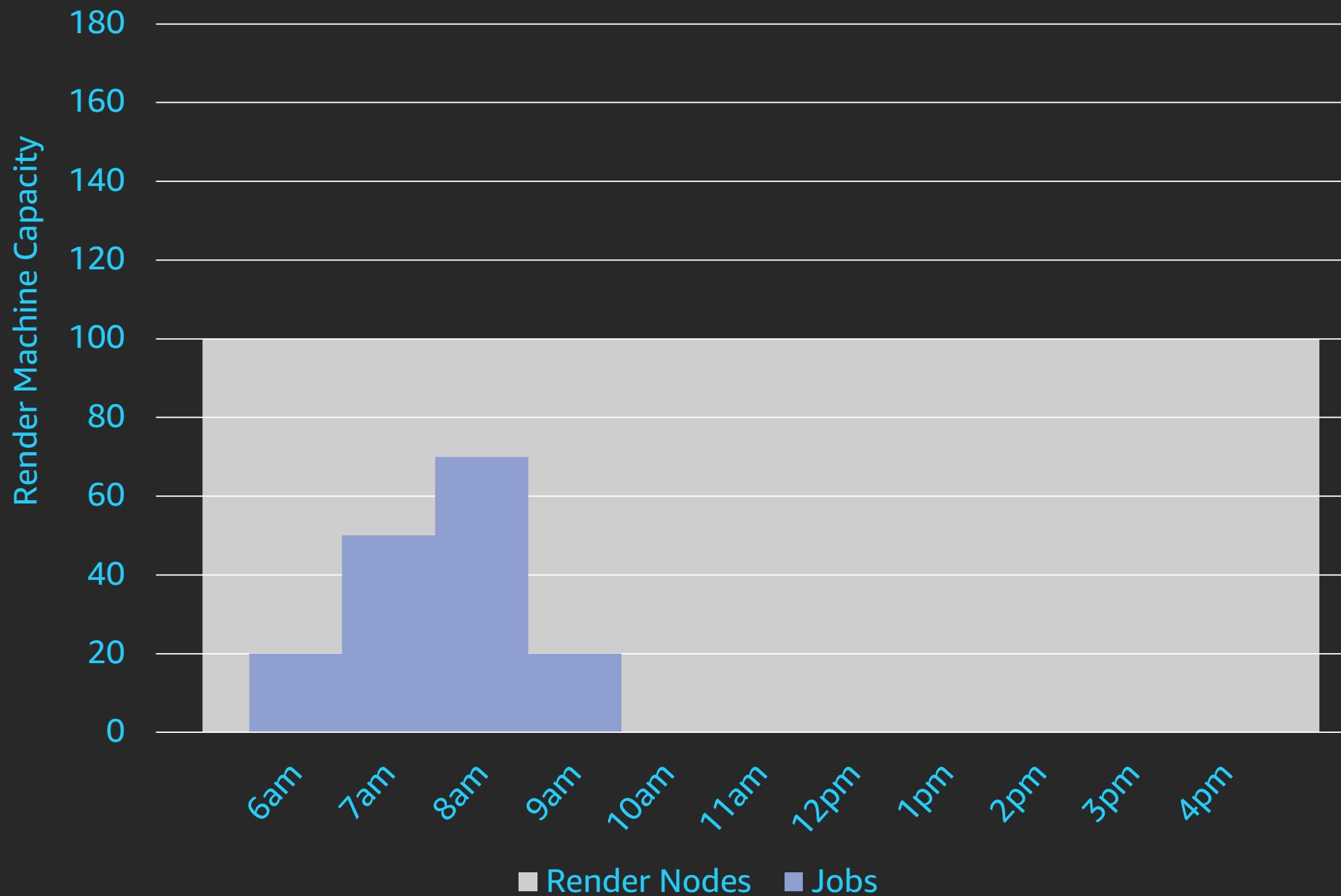


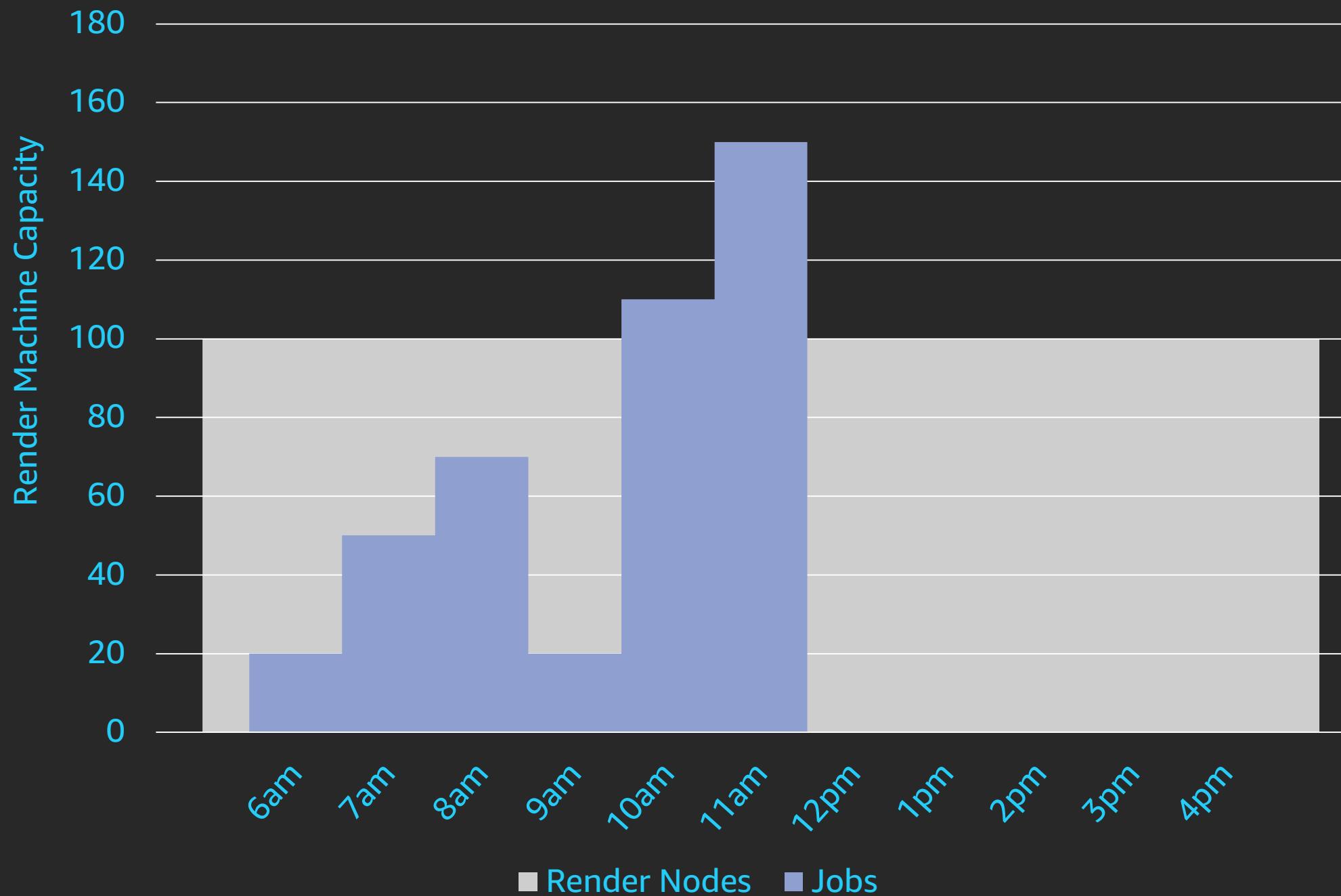


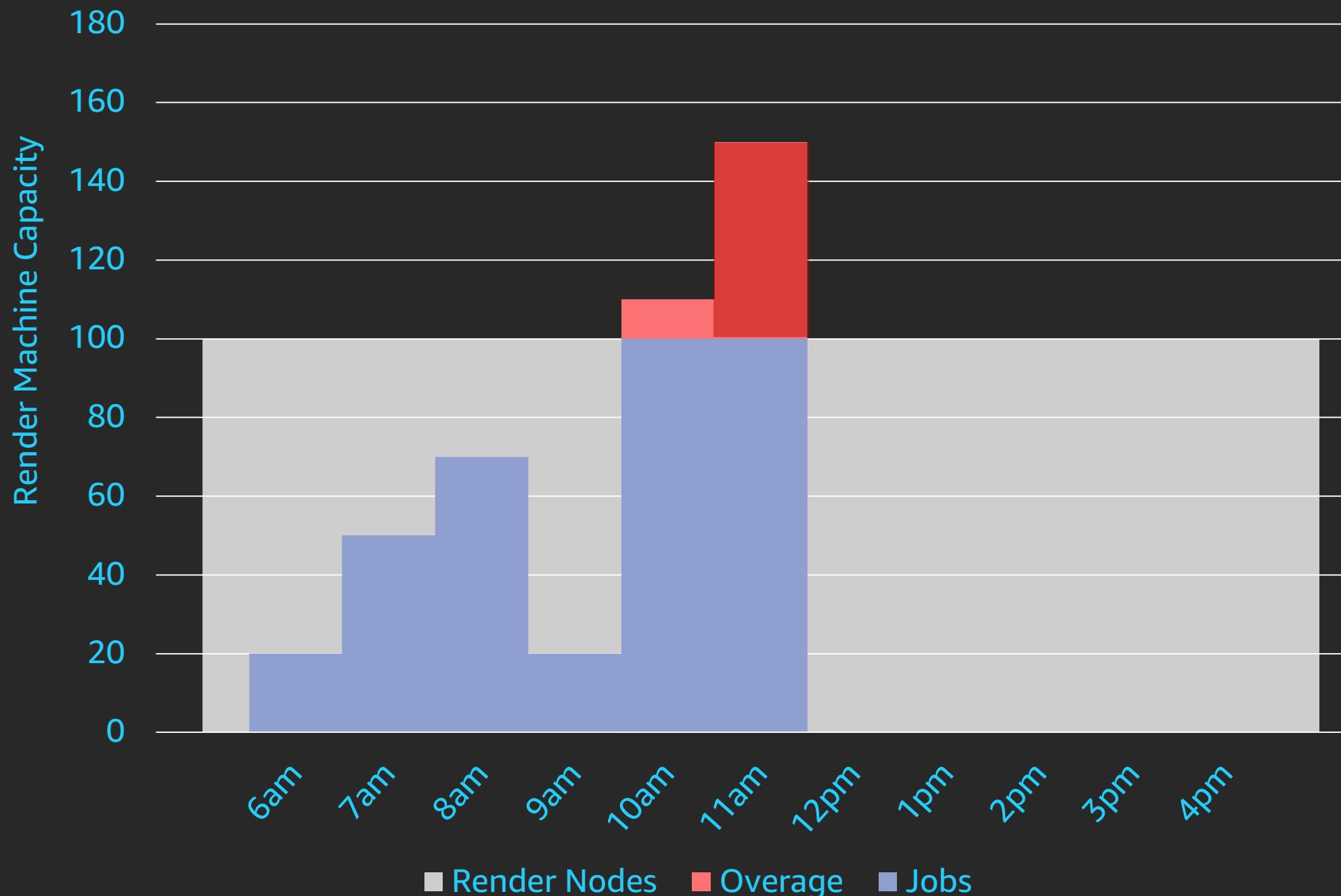
# Common rendering patterns

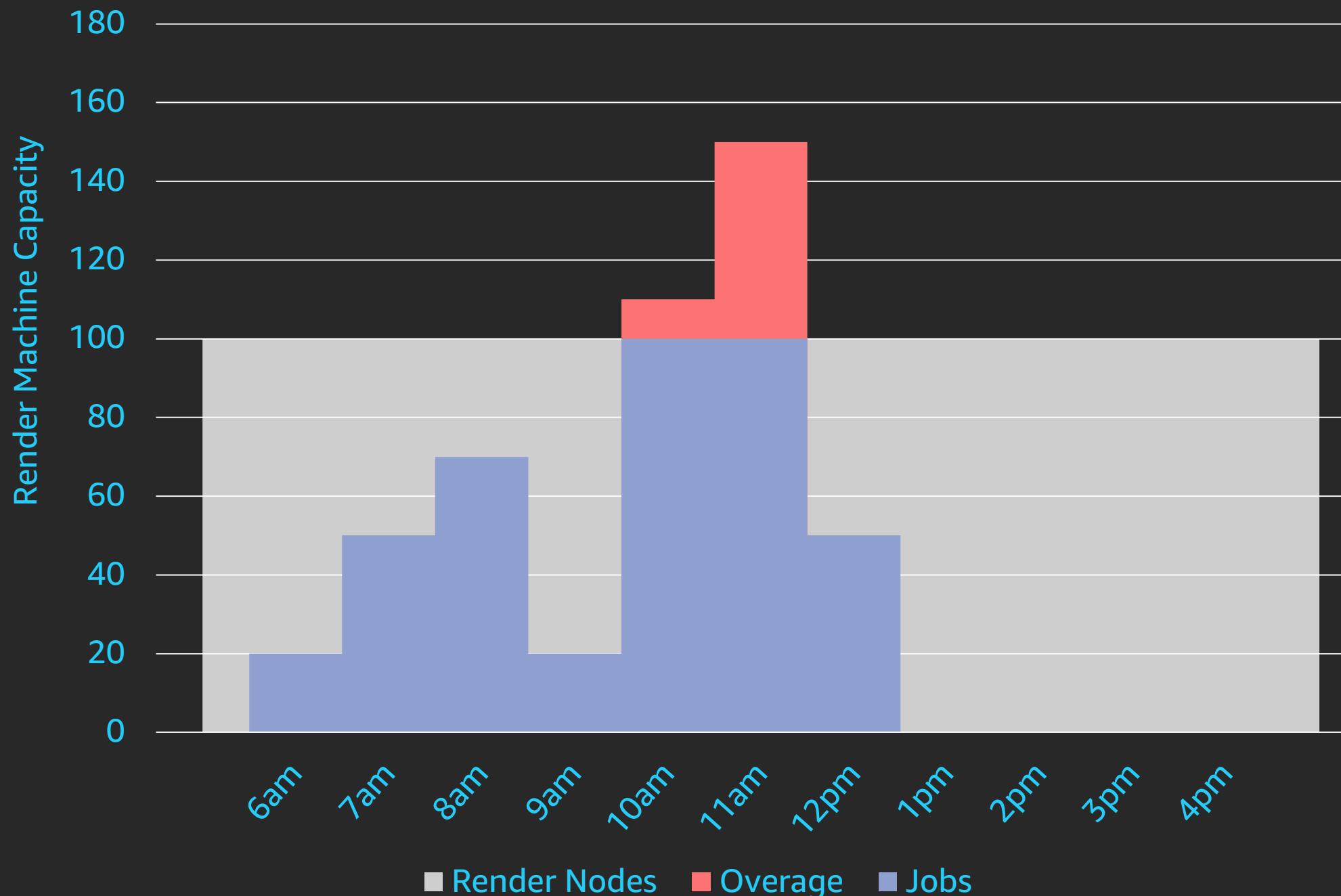


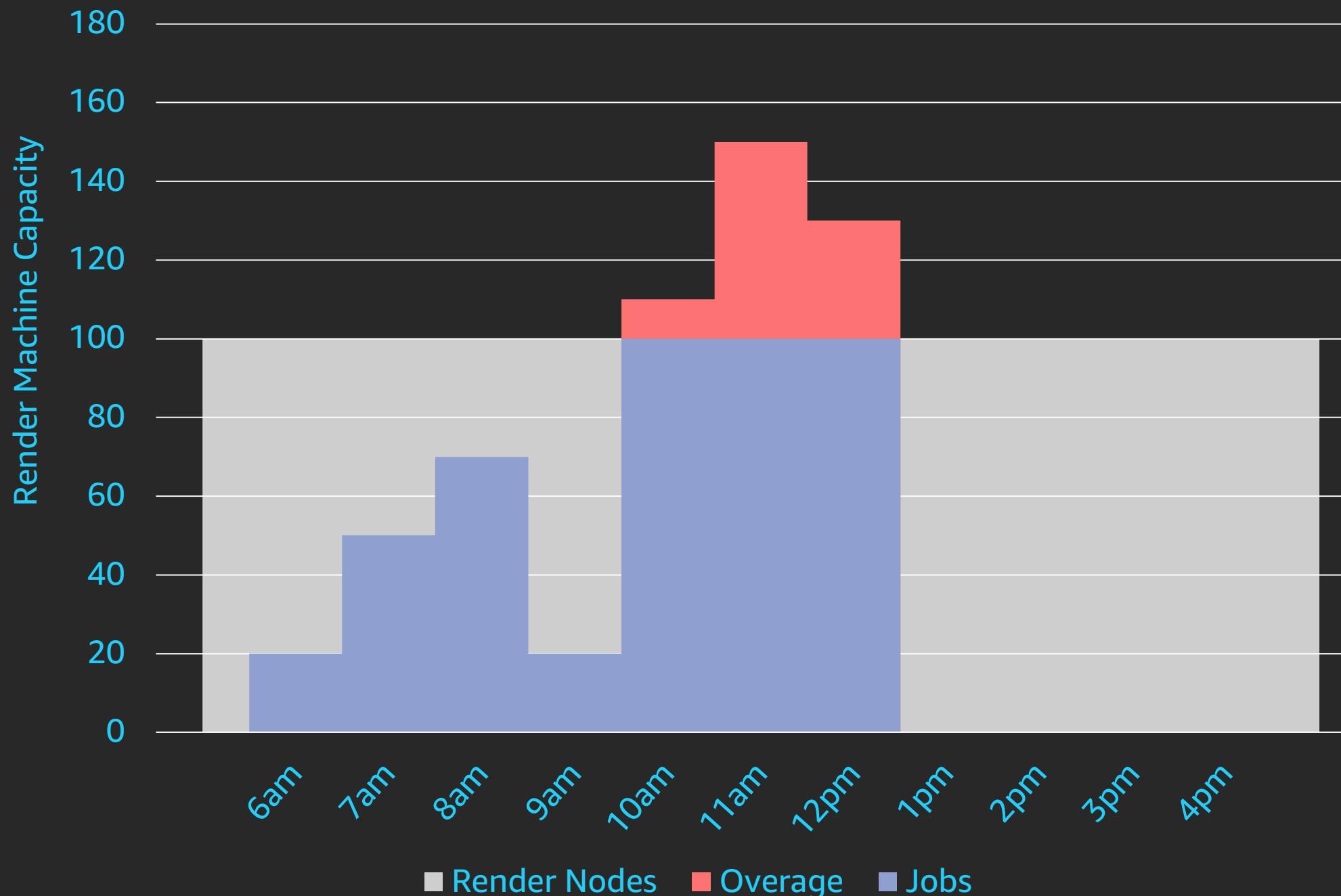


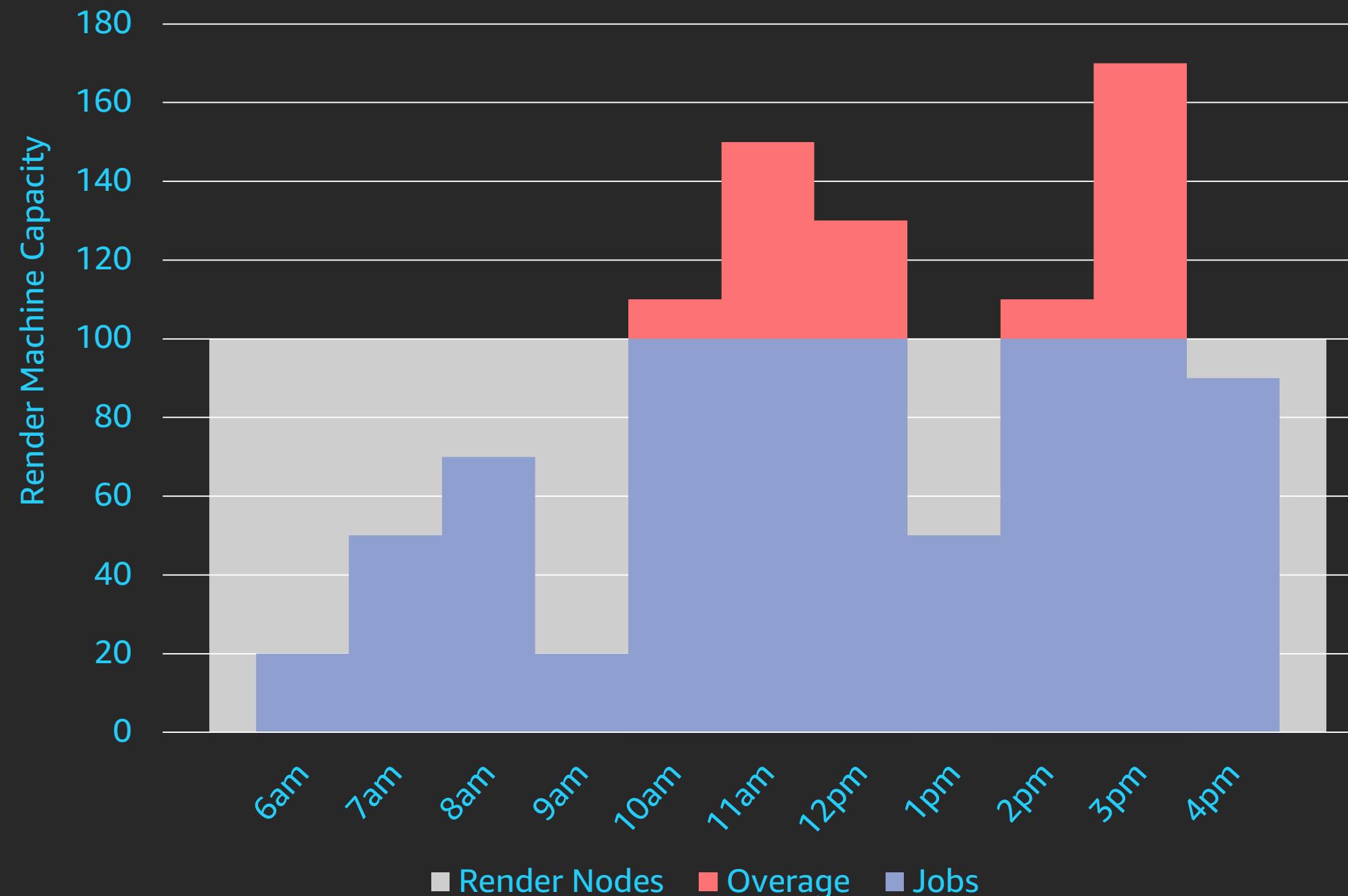


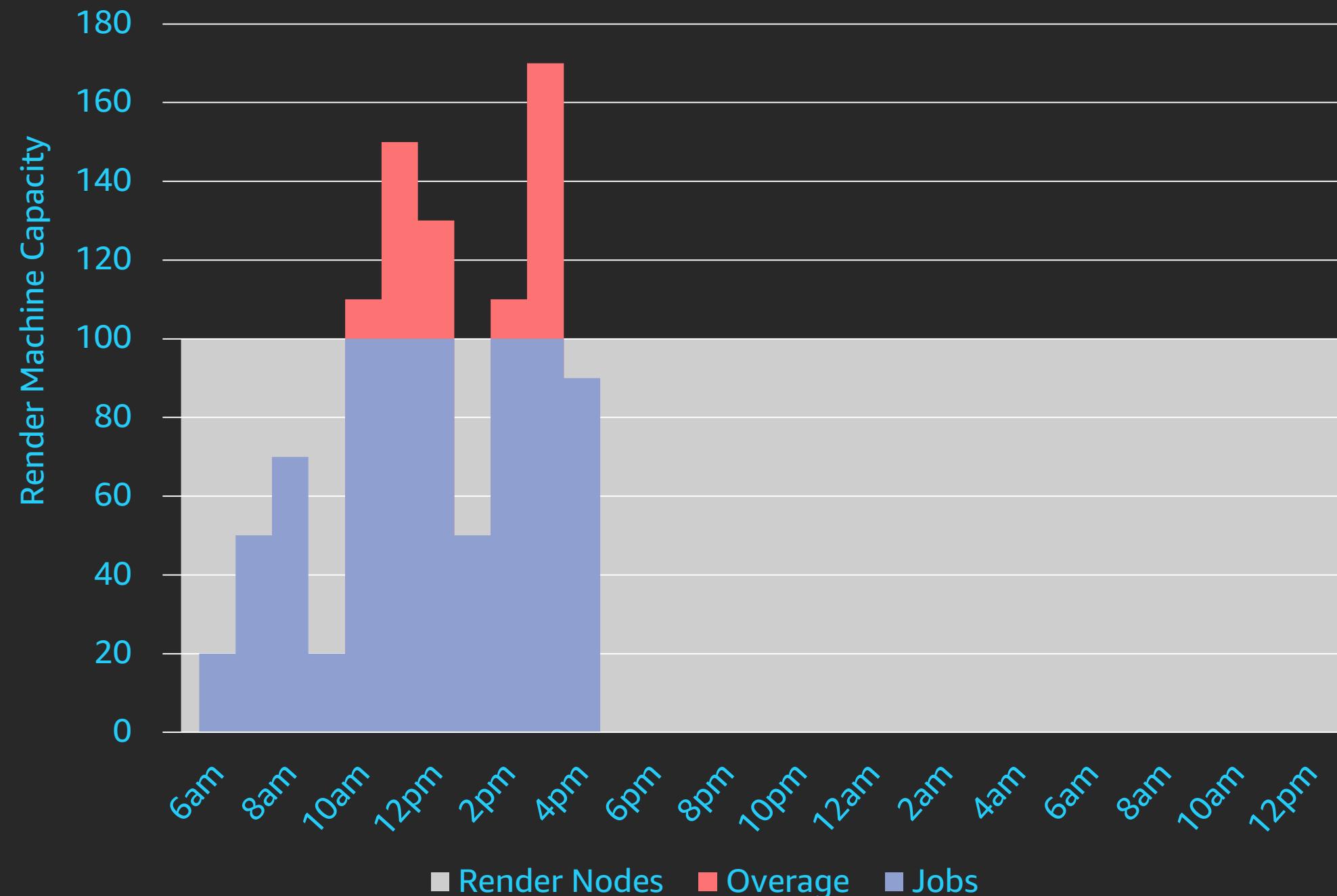


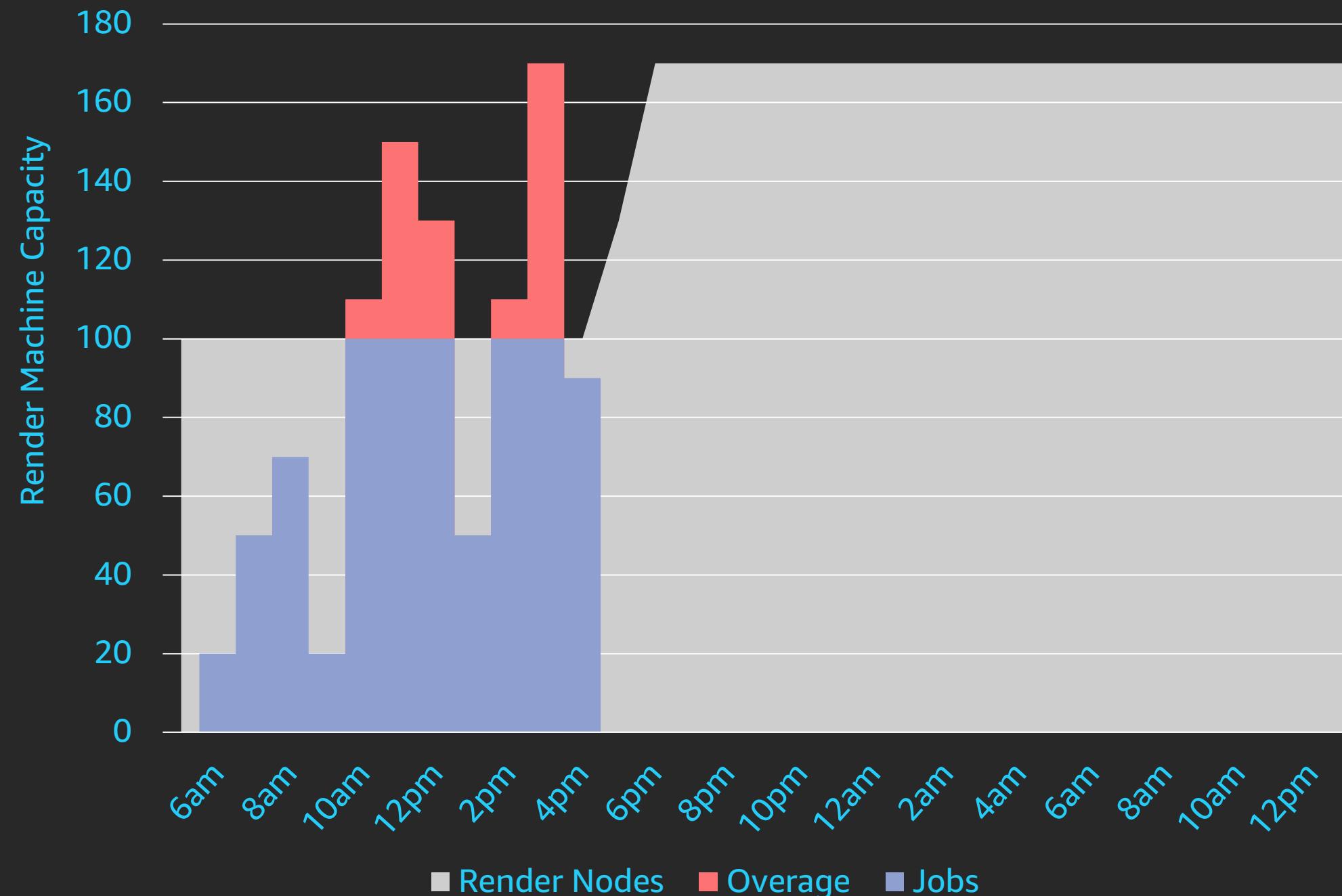


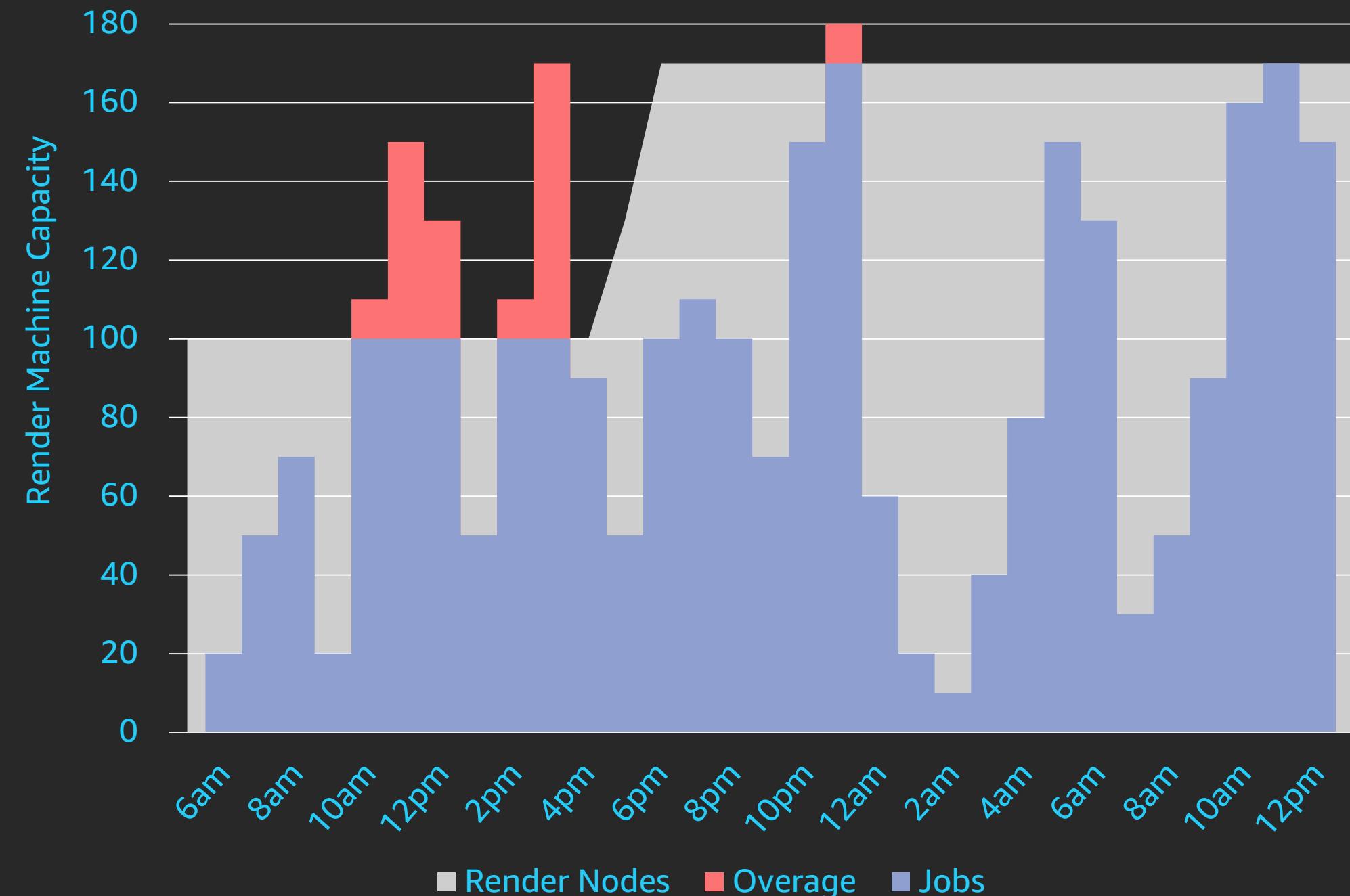


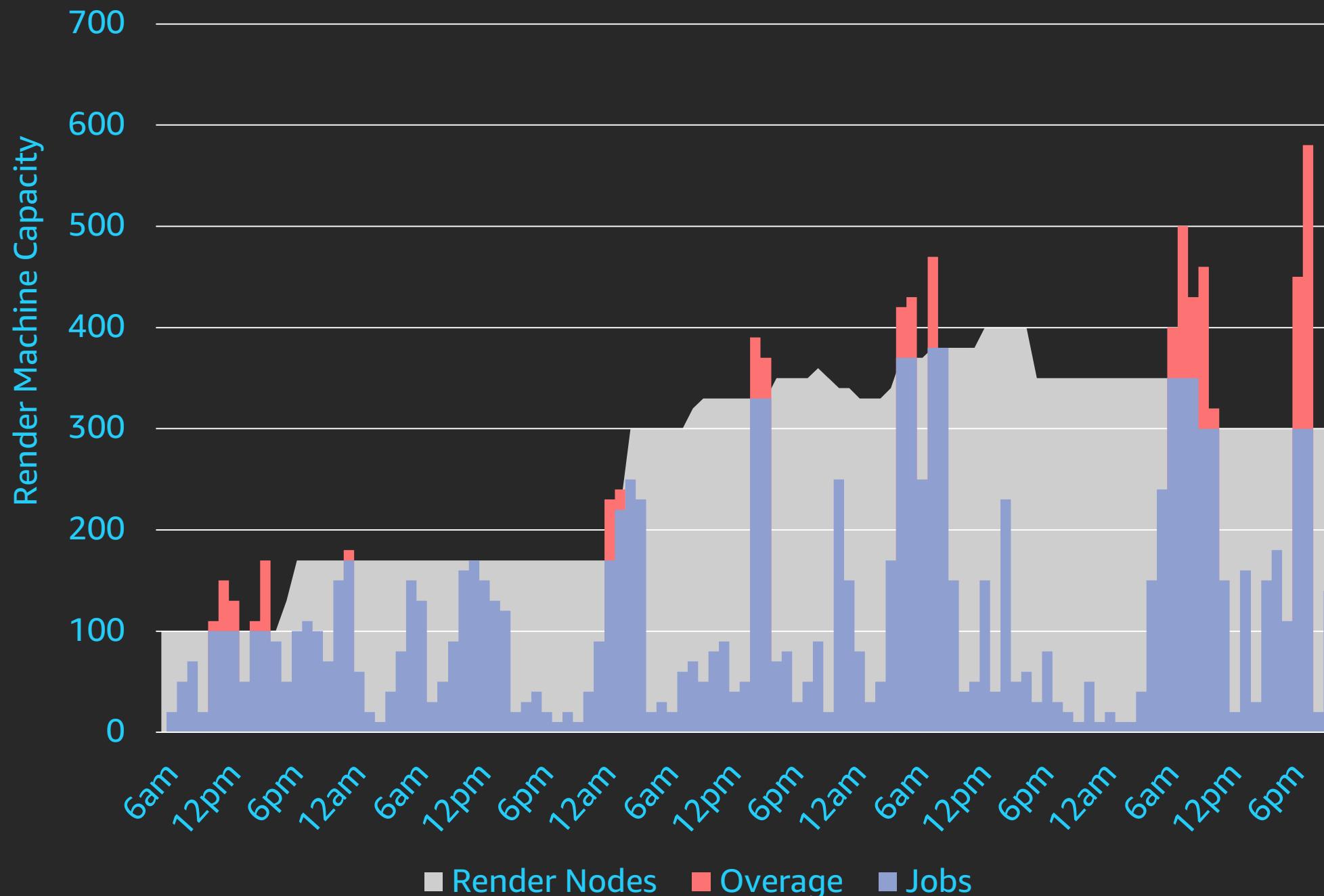


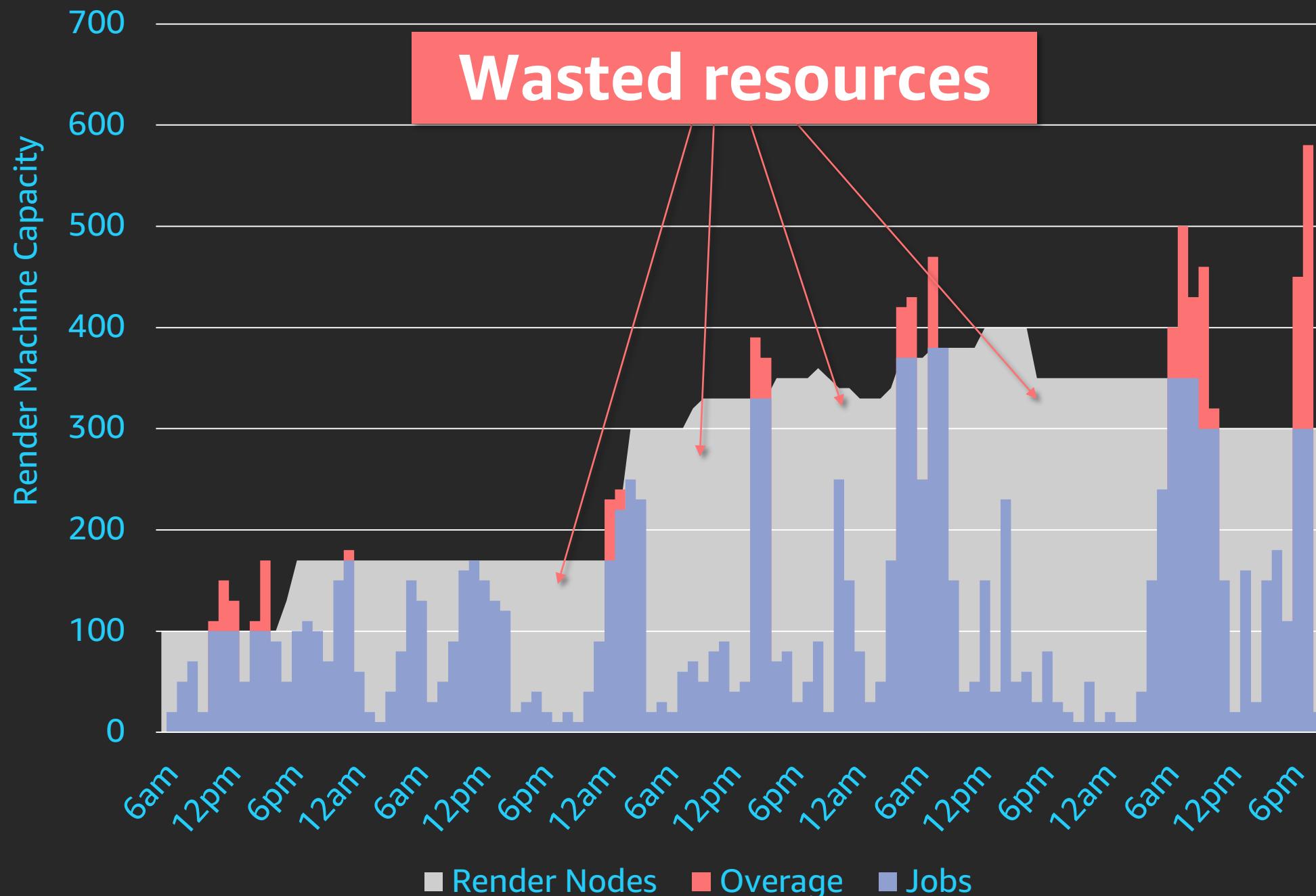


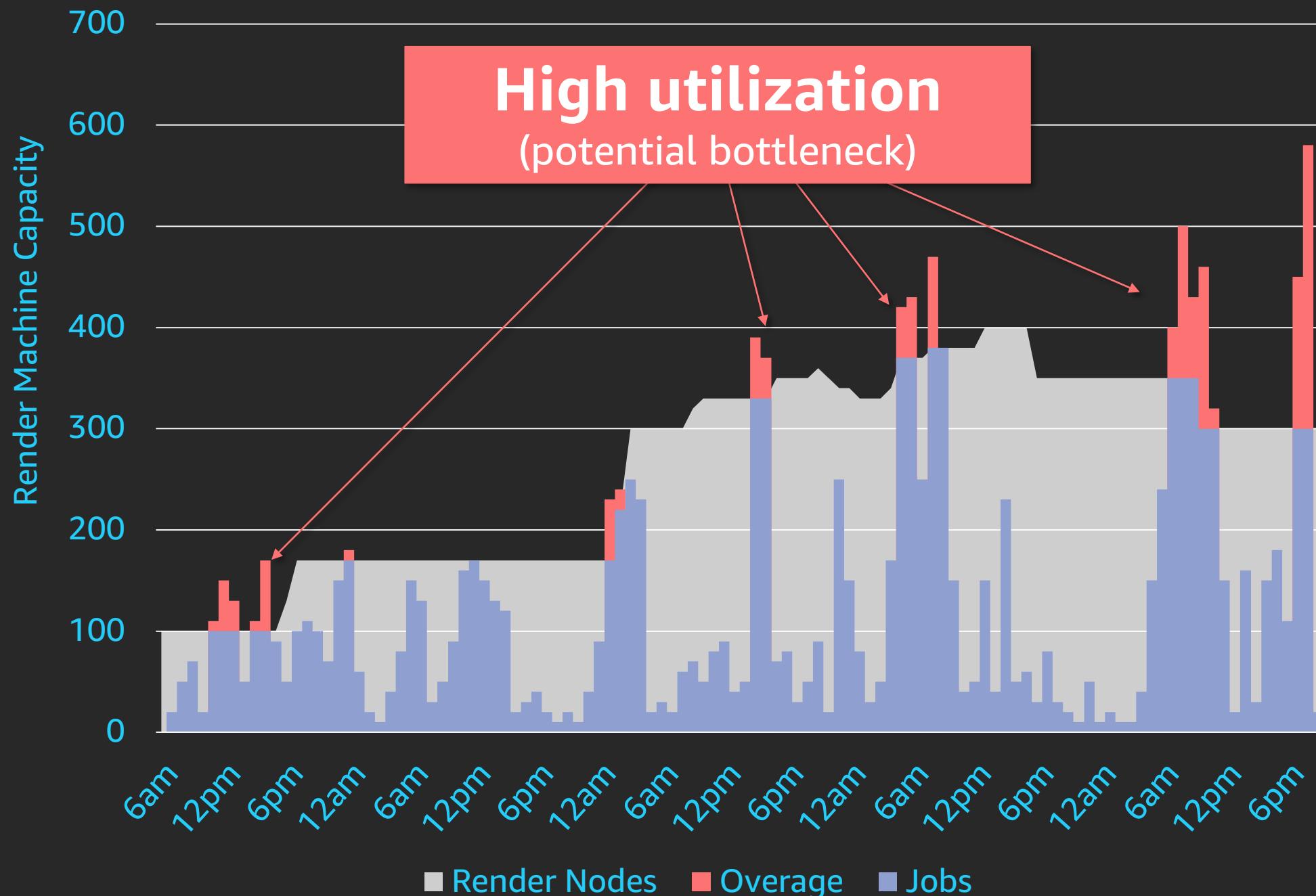


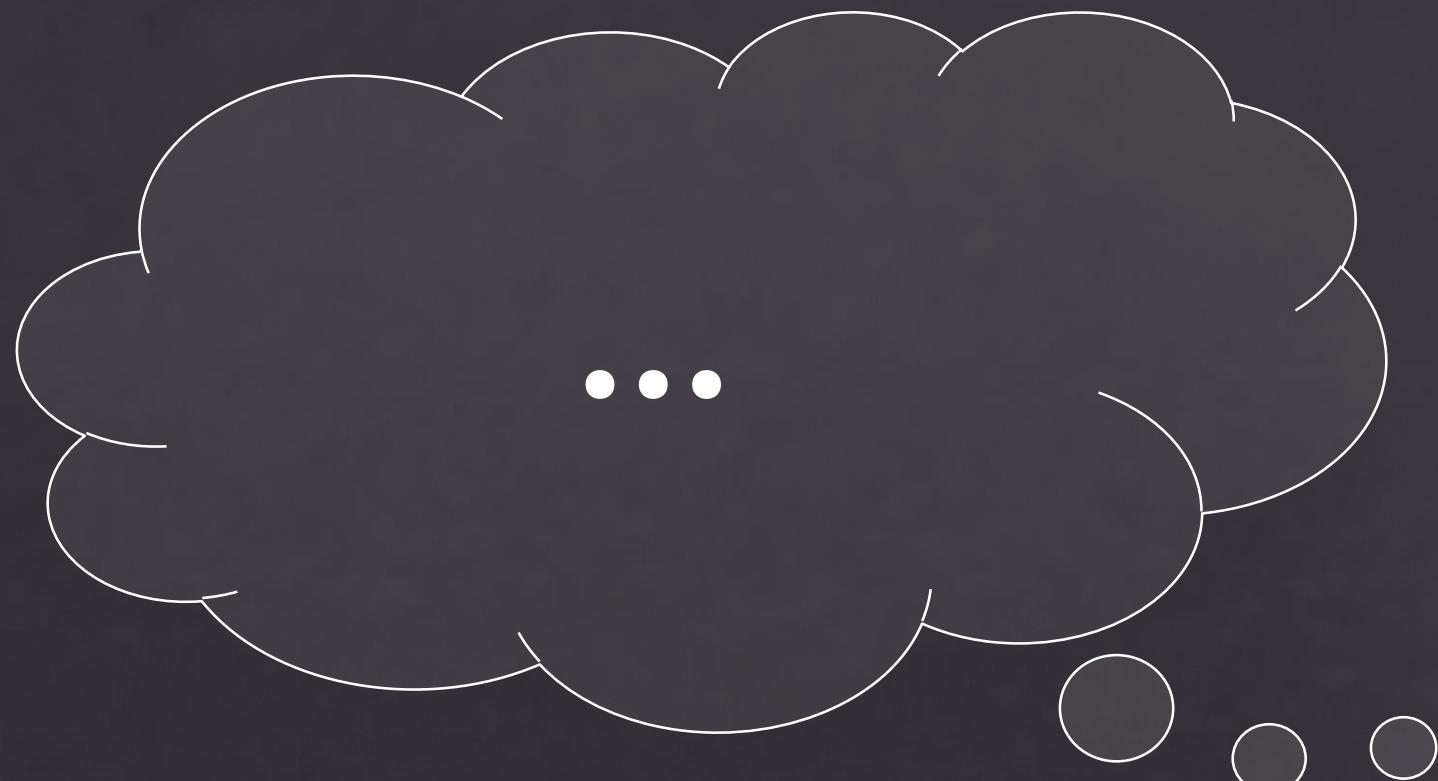


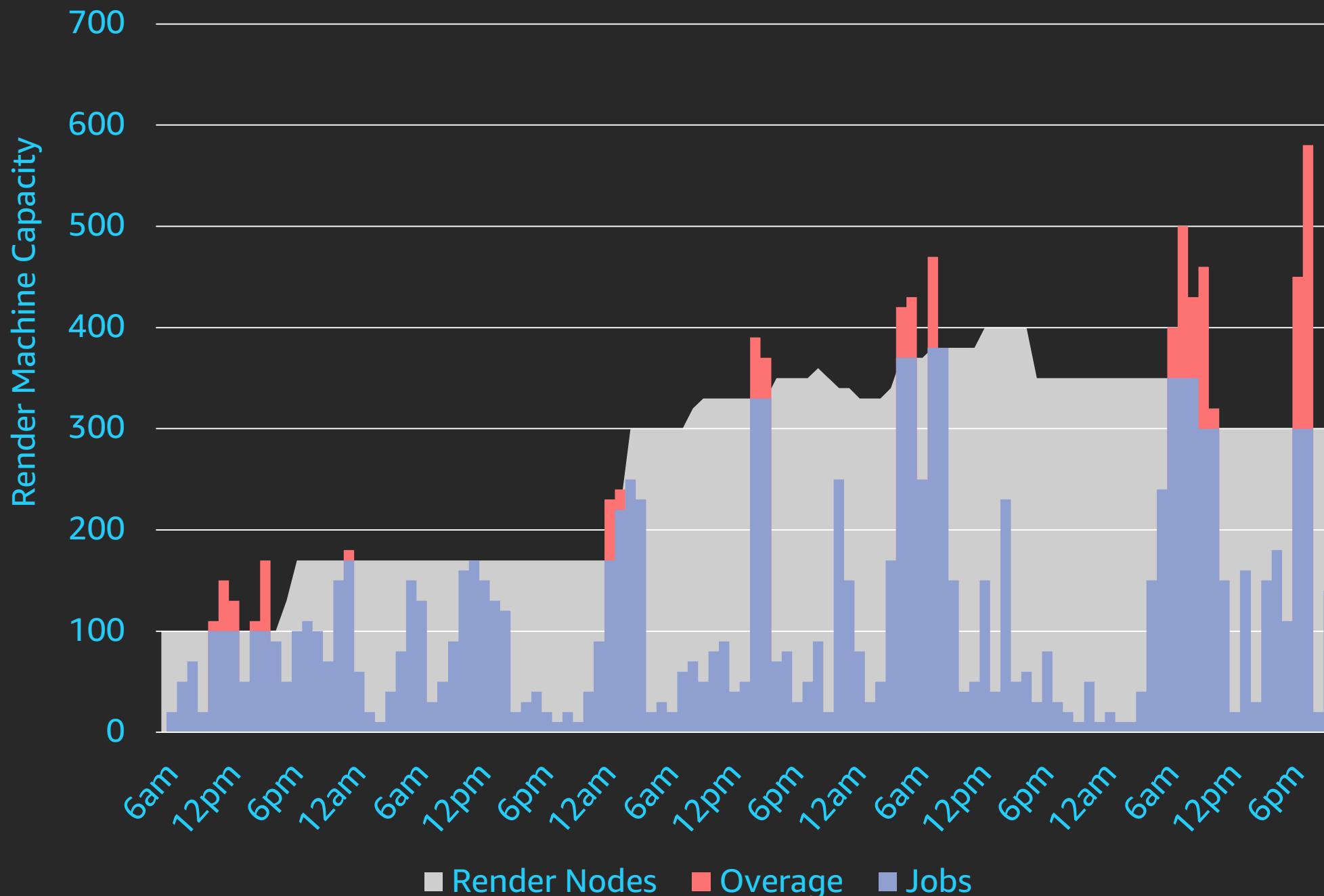


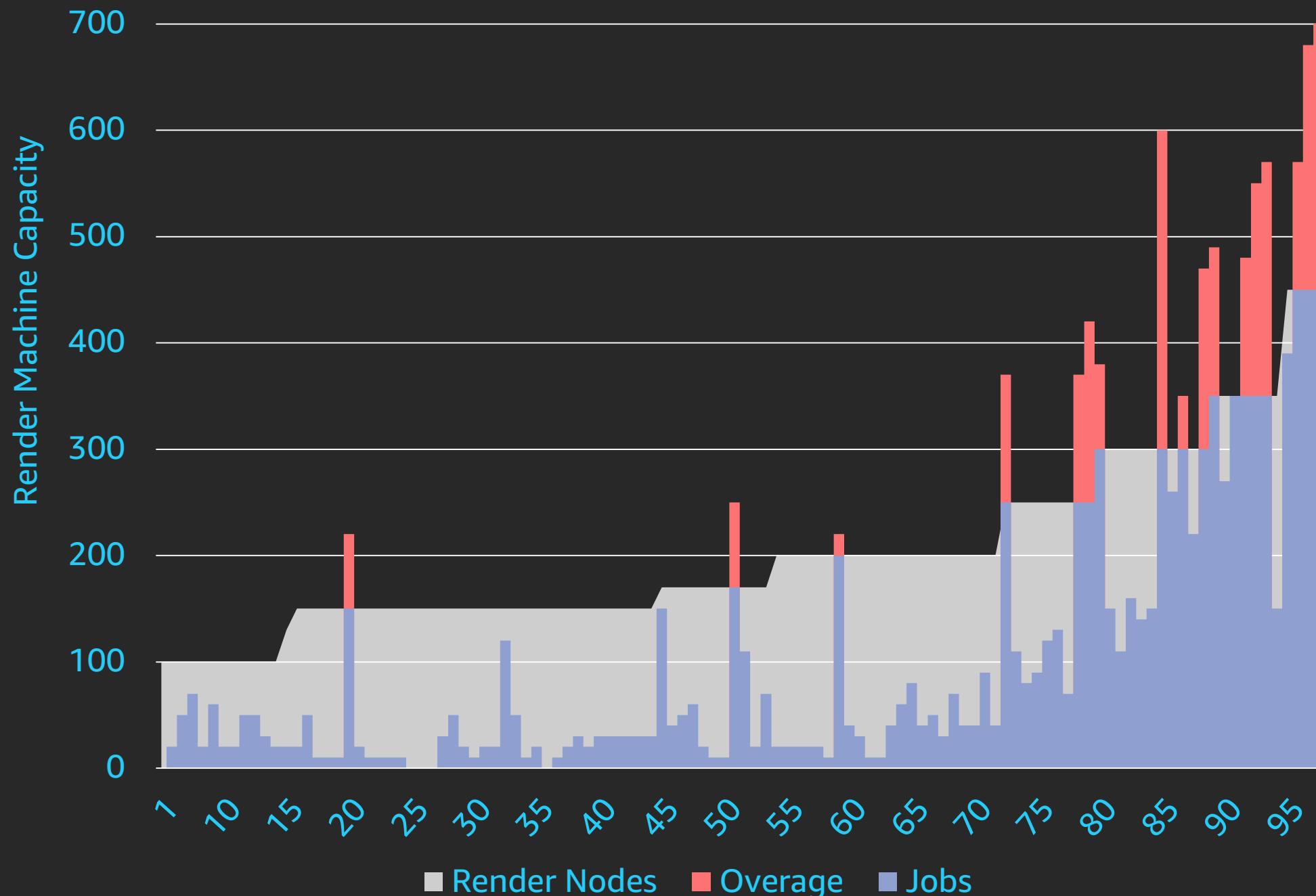




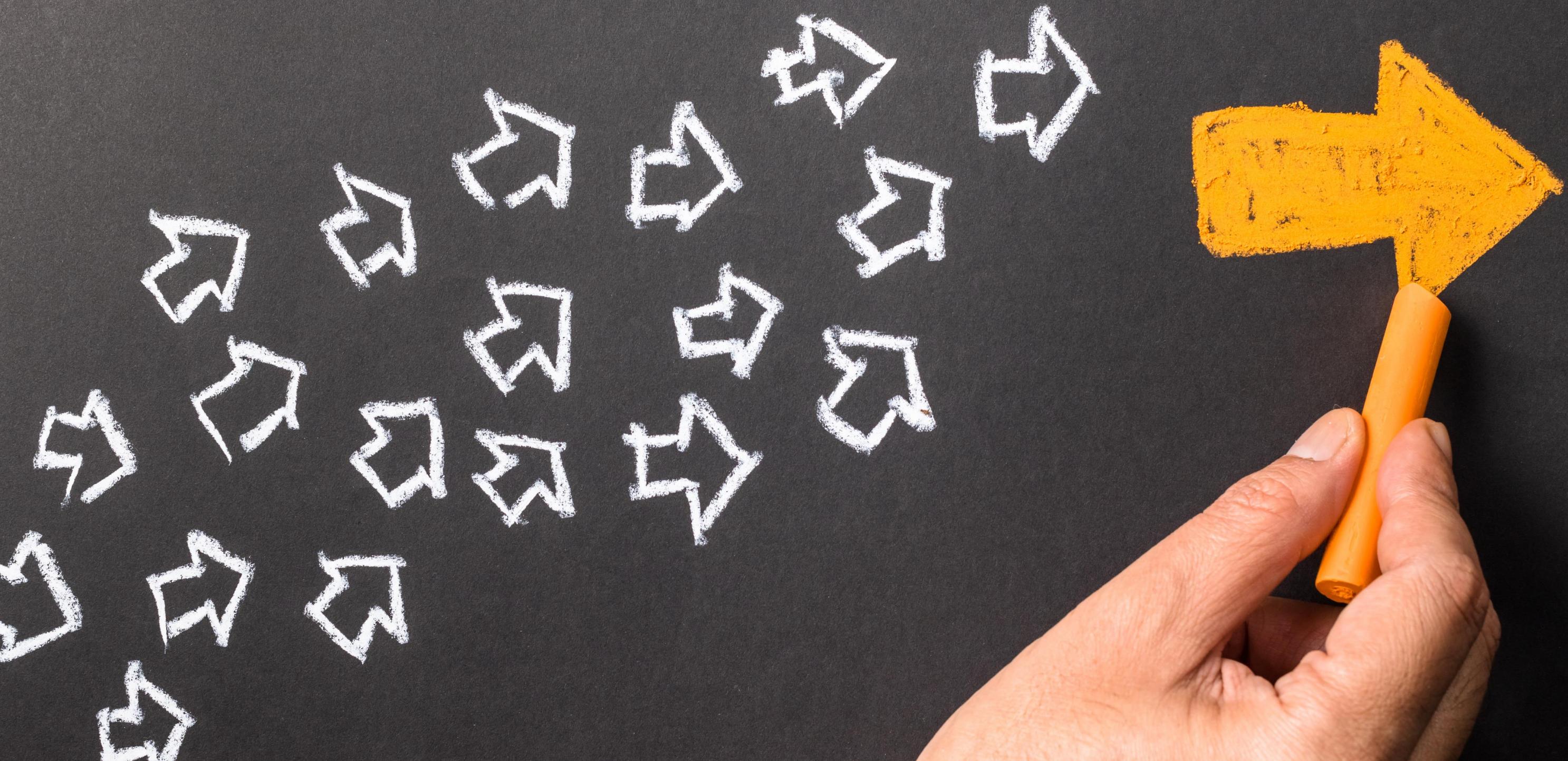


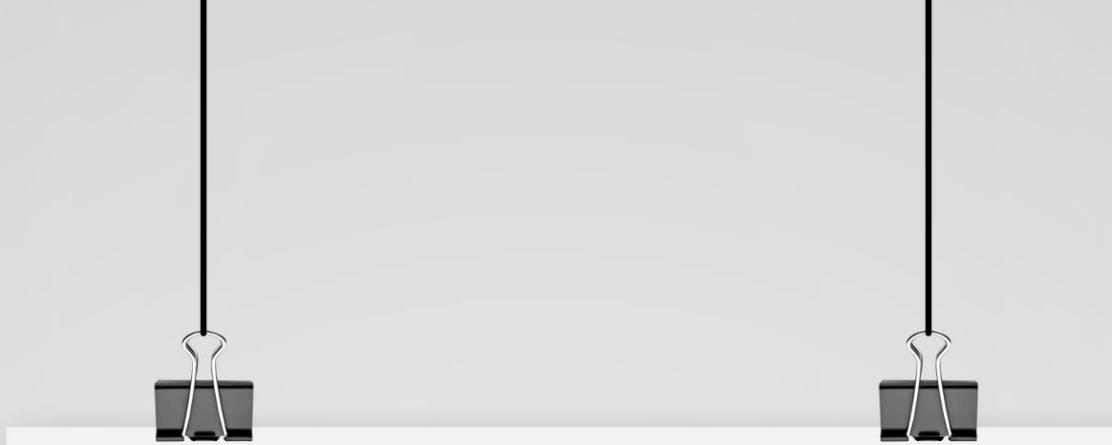






# Opportunity





# How?

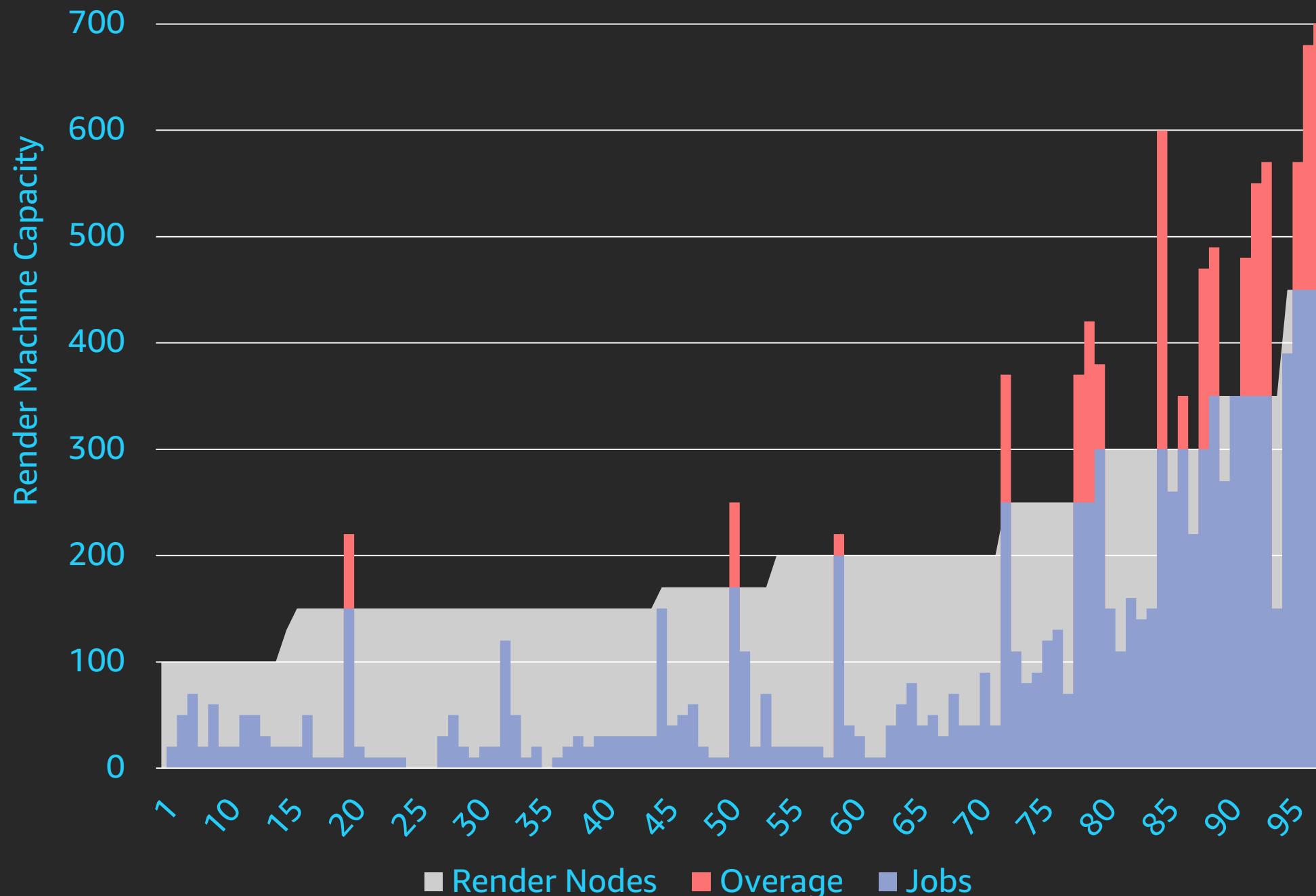
Buy more hardware ...

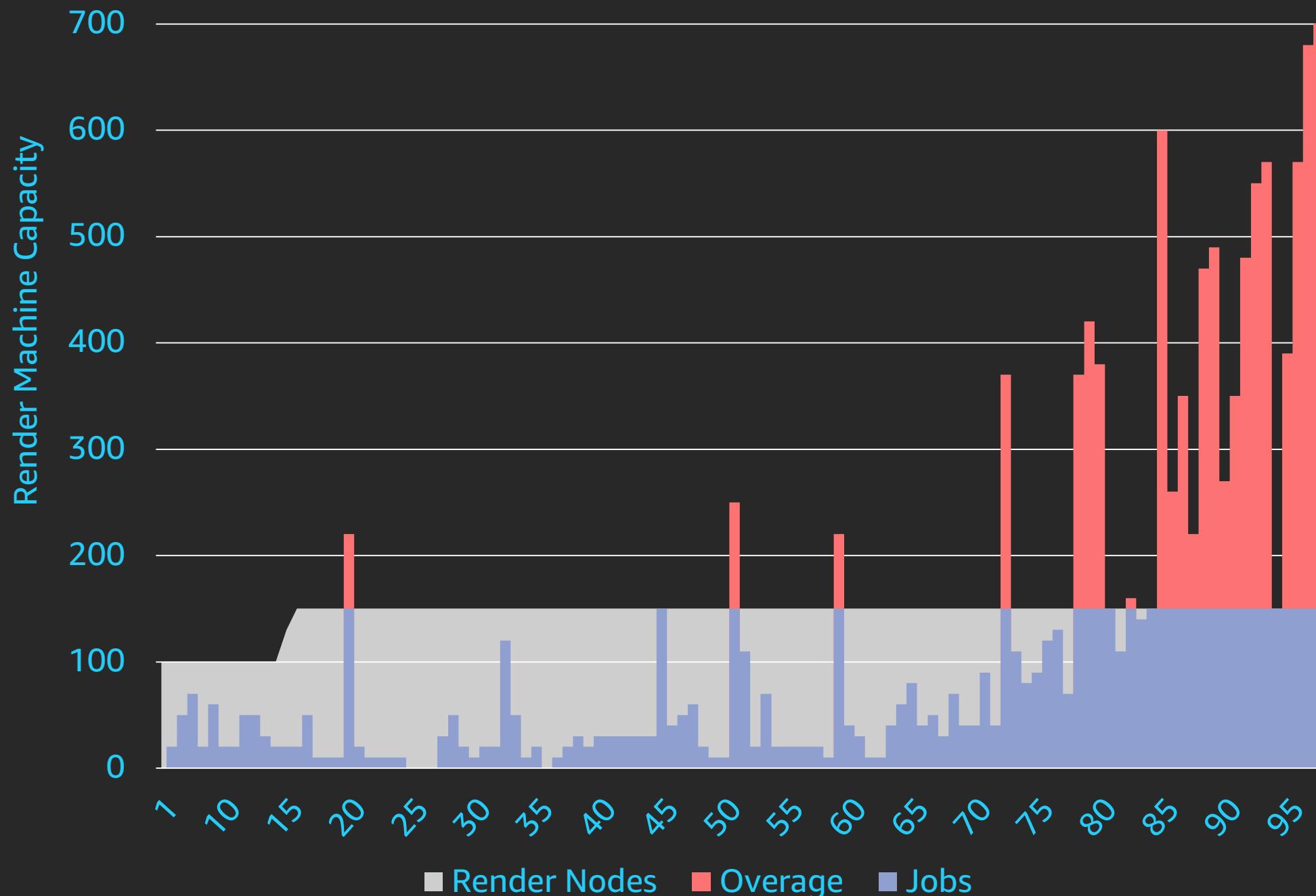
Sacrifice quality ...

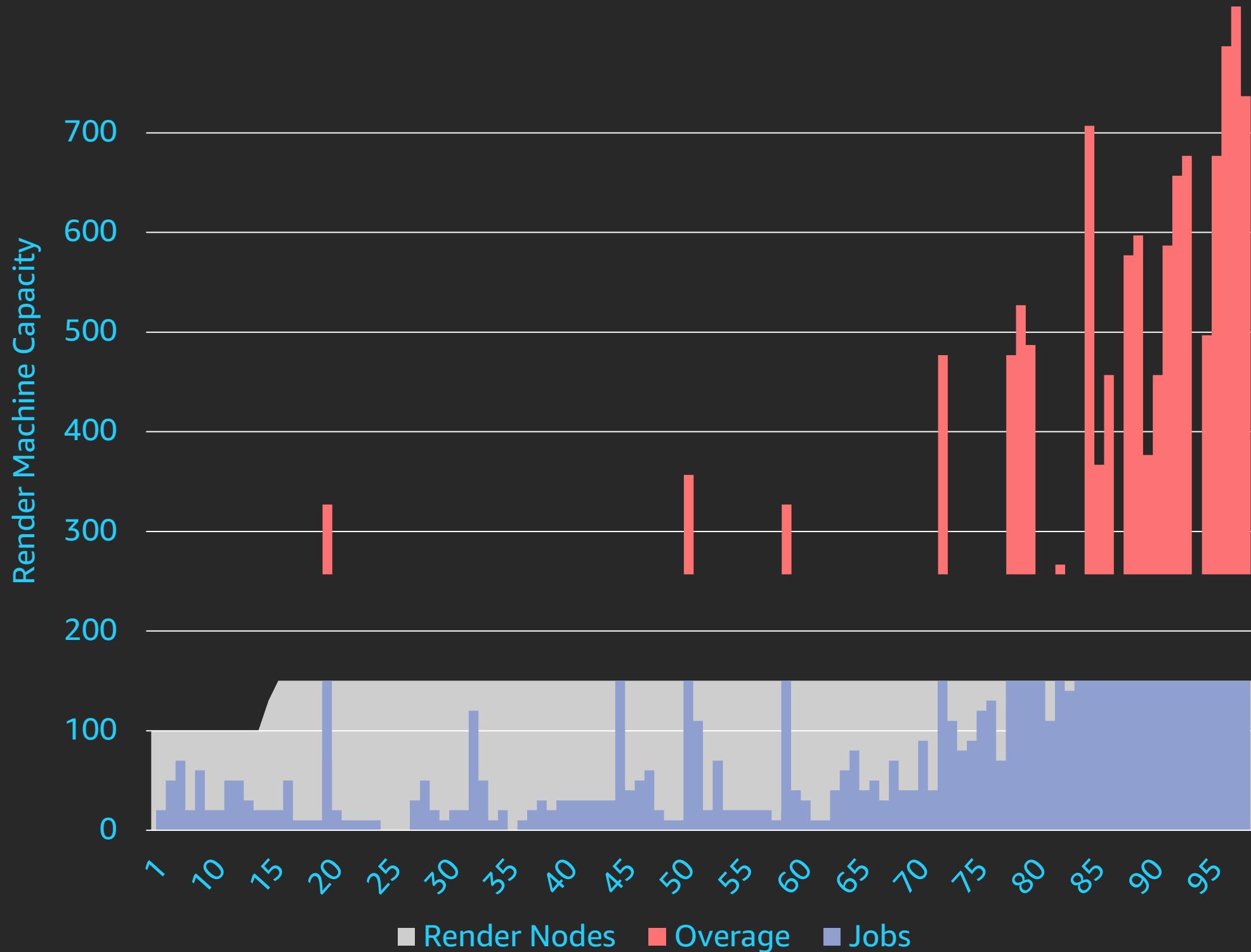
Turn down new work ...

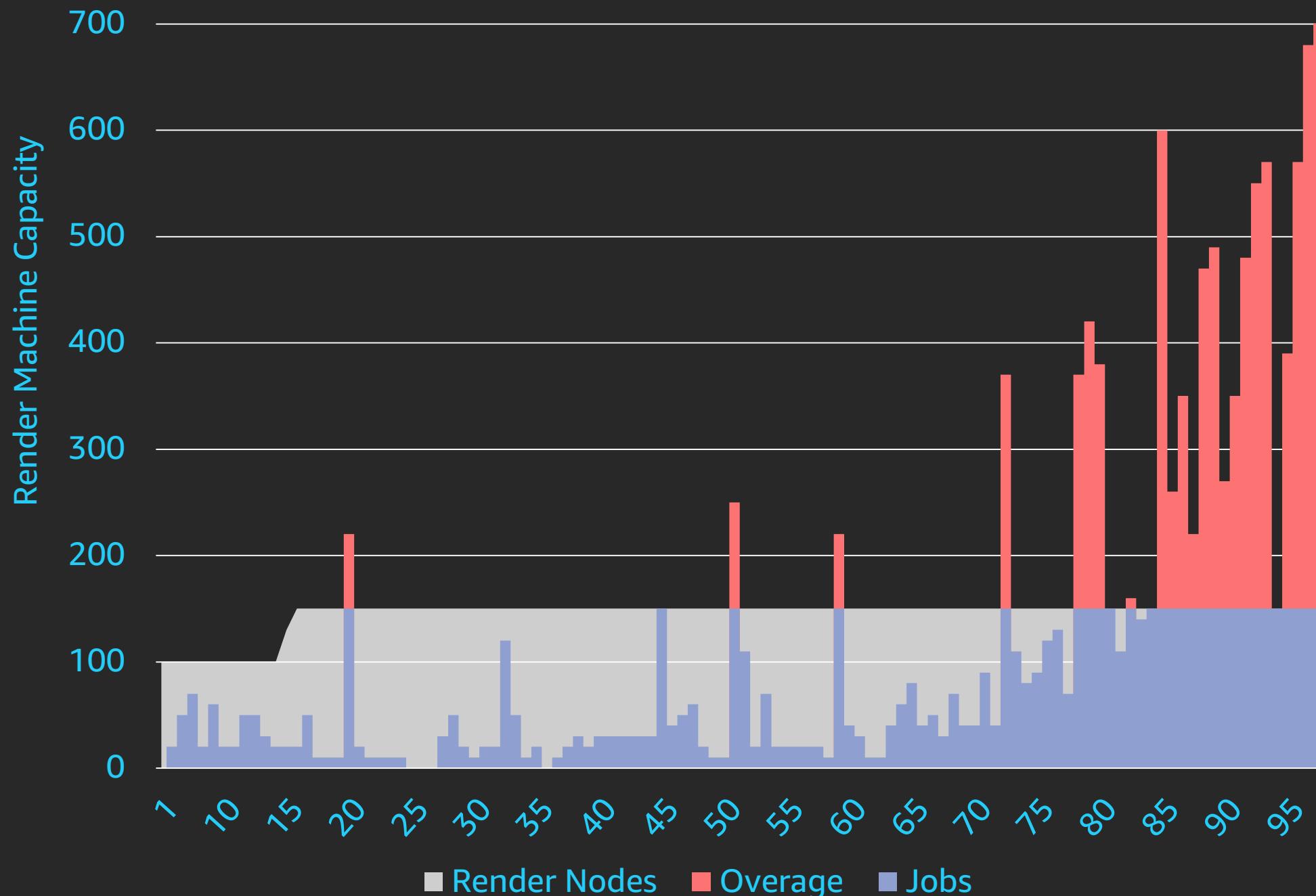




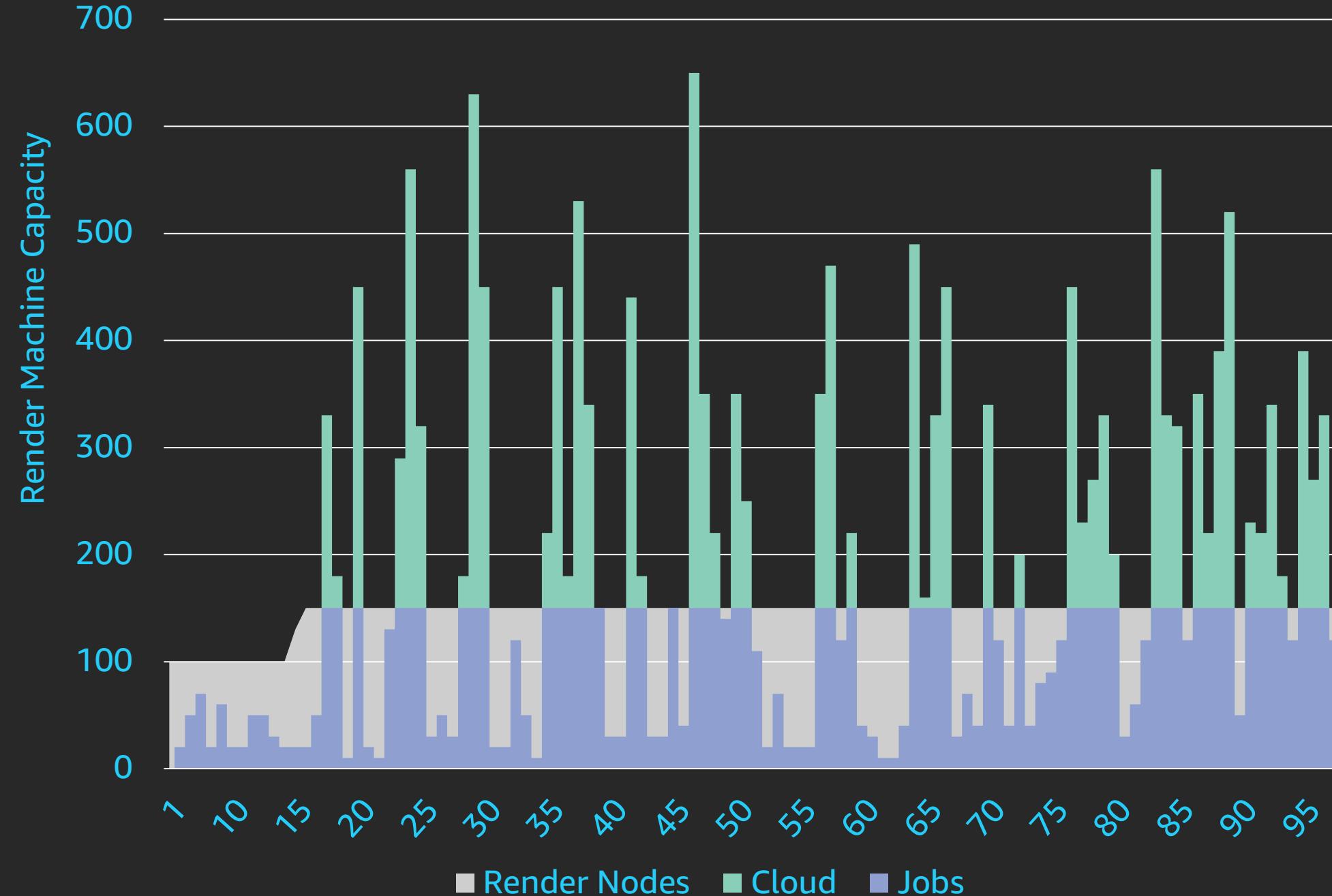








# More capacity ... more iterations!



Customer trends

Scaling

2-10x

on-premise capacity

Customer trends

Time becomes

elastic

# Customer trends

Saying

YES

# Good Omens



milk

# Challenge

5 months to deliver 650 shots

Wide variety of work, from characters  
to environments to FX

Needed a rendering solution that  
could scale

milk

# AWS Thinkbox Deadline & Amazon EC2 Spot

Managed on-premise and EC2 Spot

Peak 83k cores, average of 15k cores per day

Deadline's flexibility allowed for custom pipeline

milk

# Benefits

Scalability and elasticity enabled small team to complete on time

Milk punched above their weight on a project typically requiring a much larger team

The logo for Milk Visual Effects, featuring the word "milk" in a lowercase, sans-serif font. The letter "m" is stylized with a vertical bar on the left and a horizontal bar on the right, creating a shape reminiscent of wings or a liquid flow. The "i" has a small vertical bar extending from its top. The "l" is a simple vertical line, and the "k" has a vertical bar on the left and a horizontal bar on the right, mirroring the "m". The "s" is a standard lowercase "s". The entire logo is rendered in a light blue-grey color.



# Orville Space Battle

## Episode 209





# Challenge

Nearly 8 minutes of space battle – rivaling the likes of Star Wars – for TV

228 shots; over 1,000 person hours in 8 weeks



# AWS Thinkbox Deadline & AWS EC2 Spot

Used Deadline to manage on-premise and EC2 Spot Instances

Integrated solution with in-house production platform Nucleus

100,000+ hours of rendering

200,000+ frames rendered on EC2 Spot with a peak over 100,000 cores

500,000+ total frames rendered

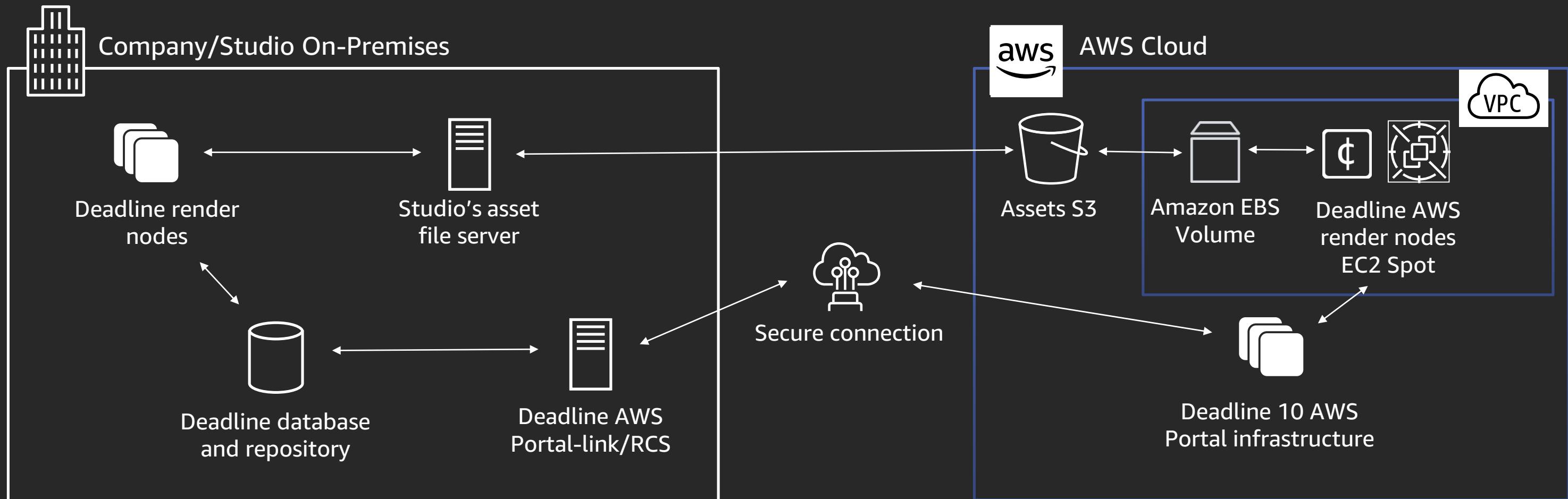
100,000+ independent versions of shots rendered



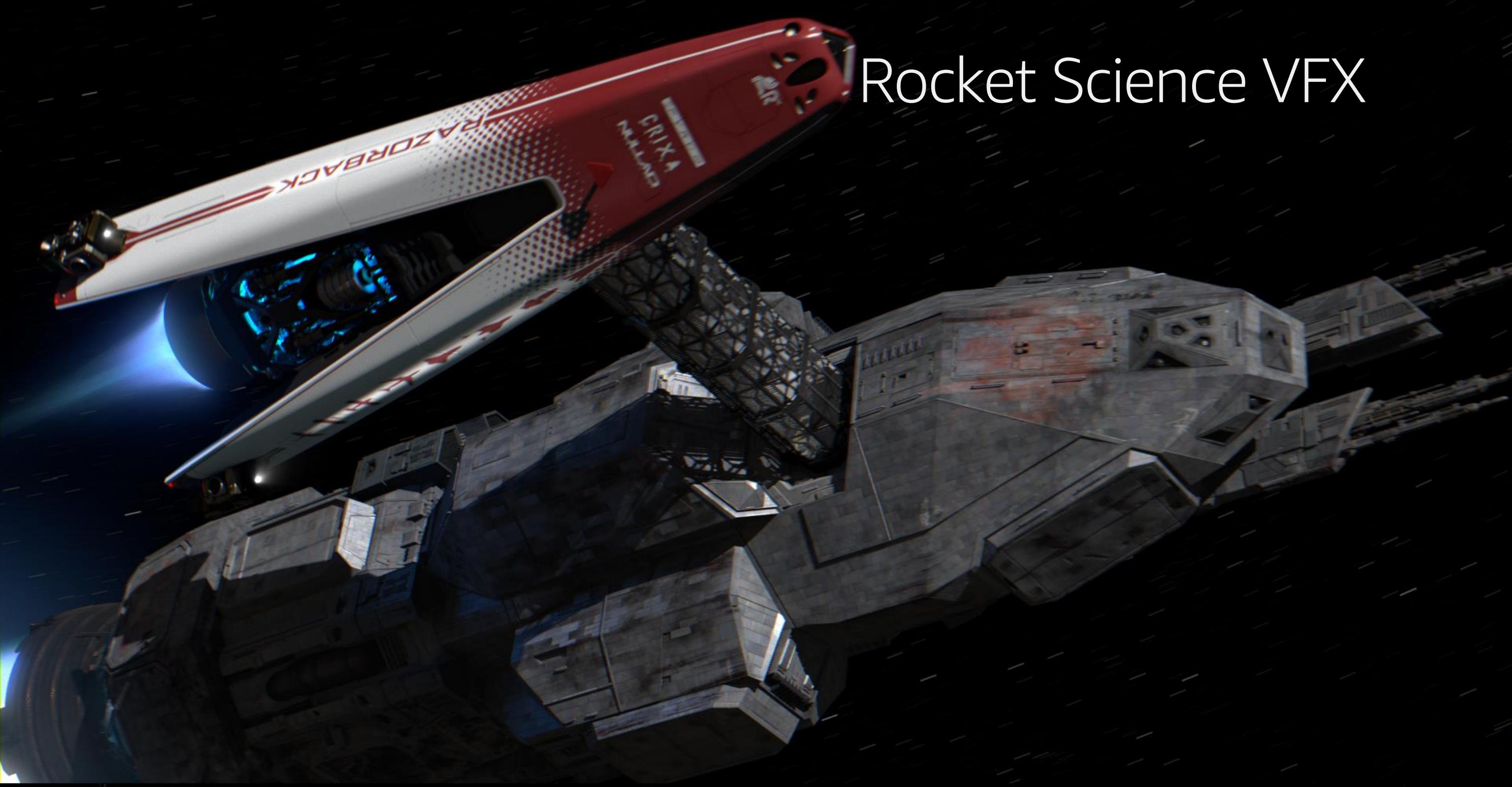
# Benefits

Impossible task without cloud capacity  
Elastic capacity of cloud rendering allows production teams to hit delivery targets with confidence

# Example hybrid rendering architecture



# Rocket Science VFX



THE  
**EXPANSE** SEASON 3

Rocket Science



# Challenge

Needed to expand GPU resources  
Several projects running in parallel  
with different pipelines



# AWS Thinkbox Deadline & AWS EC2 Spot

Scaled up NVIDIA GPU instances using EC2 Spot for rendering with Redshift

Used CPU rendering with Autodesk Arnold for parallel project

Thinkbox Deadline allowed render wranglers to view and manage local and cloud-based resources in a single interface



# Benefits

More time to iterate and refine shots  
Flexible pipeline that allows for adoption of different workflows

# Software

# AWS Thinkbox Marketplace



AUTODESK®  
3DS MAX®



arnold



clarisse



- Usage-based licensing (UBL)
- Per-minute licensing for applications



Houdini  
ENGINE™



KATANA



KeyShot®  
by Luxion



Krakatoa Render

Houdini Engine

Katana Render

Keyshot Render



AUTODESK®  
MAYA®



NUKE



RED SHIFT



V-Ray Render

Maya for AWS

Nuke Render

Redshift Render



Yeti Render

<https://marketplace.thinkboxsoftware.com>



# New open source foundation

The Academy Software Foundation provides a neutral forum for open source software developers in the motion picture and broader media industries to share resources and collaborate on technologies for image creation, visual effects, animation and sound.





## PREMIER MEMBERS



DREAMWORKS



Google Cloud



## GENERAL MEMBERS



## ASSOCIATE MEMBERS



# Virtual workstations

A photograph of a person sitting at a white circular table by a swimming pool, working on a white laptop. A small white cup of coffee sits next to the laptop. The background is the dark blue water of the pool.

Reduce overhead



Talent  
is global



Burst workforce  
and compute



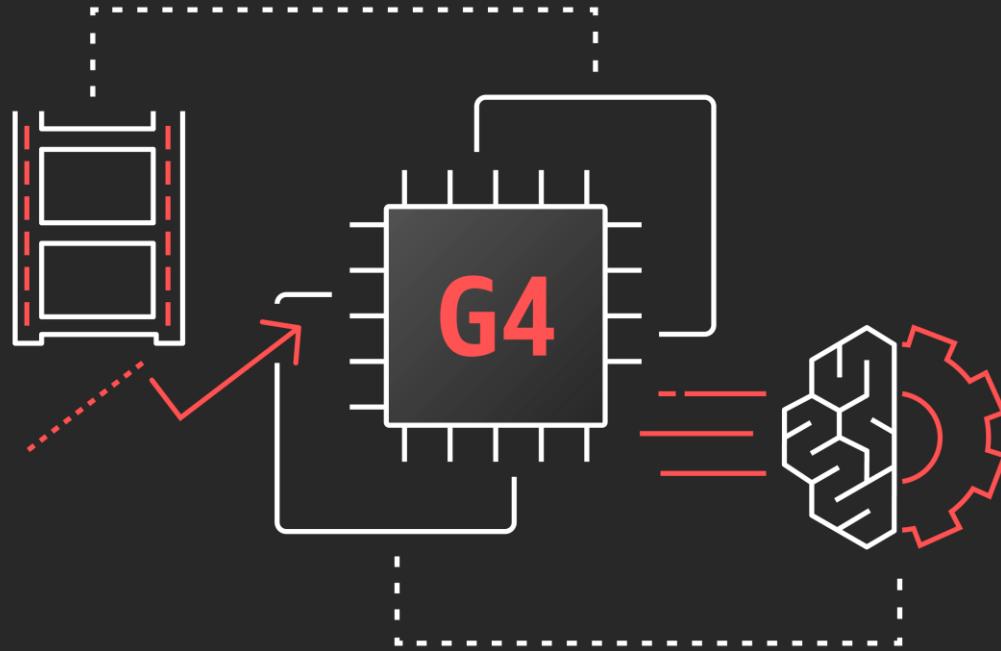
Secure (MPAA/TPN)

# Virtual workstations for content production



High-spec graphics workstations  
Linux and Windows  
Dual HD/2K monitors  
Color/Pixel accuracy  
Video/Audio sync  
Wacom/Cintiq support

# Introducing Amazon EC2 G4 Instances



Machine learning inference  
Graphics intensive applications  
RTX-Enabled Applications  
Video transcoding

NVIDIA T4 Tensor Core GPUs  
Up to 1.8 TB of Local NVMe storage  
Up to 100 Gbps of networking throughput

**The most cost-effective compute in the cloud for running machine learning inference and graphics-intensive applications**

# 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.

**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 network on AWS. You have complete control over your 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.

**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](#)

# Infrastructure

# AWS global infrastructure

22 Geographical Regions, 1 Local Region, 69 Availability Zones, 200 POPs

## Region and number of Availability Zones (AZs)

### GovCloud (US)

US-East (3), US-West (3)

### US West

Oregon (4)

Northern California (3)

### US East

N. Virginia (6), Ohio (3)

### Canada

Central (2)

### South America

São Paulo (3)

### Europe

Frankfurt (3)

Ireland (3)

London (3)

Paris (3)

Stockholm (3)

### Middle East

Bahrain (3)

### Asia Pacific

Singapore (3), Sydney (3),

Tokyo (4), Osaka-Local (1)\*

Seoul (3), Mumbai (3),

Hong Kong (3)

### China

Beijing (2), Ningxia (3)

AWS now spans 69 Availability Zones within 22 geographic regions around the world, and has announced plans for thirteen more Availability Zones and four more AWS Regions in Indonesia, Italy, South Africa, and Spain.



## Announced Regions

Four Regions and 13 AZs in Indonesia, Italy, South Africa, and Spain

\* Available to select AWS customers who request access. Customers wishing to use the Asia Pacific (Osaka) Local Region should speak with their sales representative.

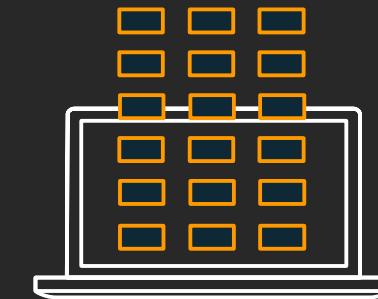
# AWS Local Zones

New type of AWS infrastructure deployment

Places compute, storage, database, and other services closer to customers  
For demanding applications that require single-digit latencies



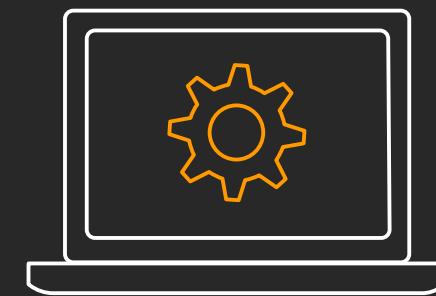
AWS  
infrastructure at  
the edge



Local compute,  
storage, database,  
and other services



Connect to  
services in  
AWS Regions

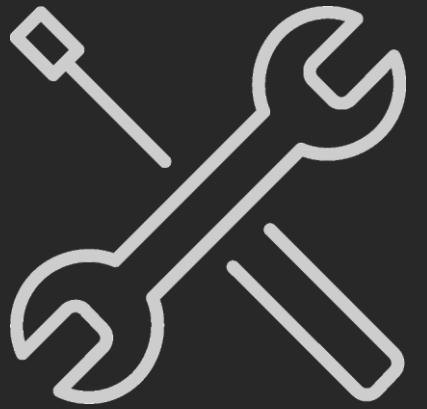


Deliver new  
low-latency  
apps

# Benefits of AWS Local Zones



Low latency



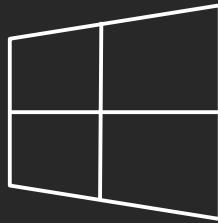
Consistent AWS  
experience



Flexible and  
scalable

# Amazon FSx for Windows File Server

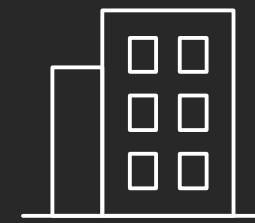
Lift and shift your Windows file storage with fully managed Windows file servers



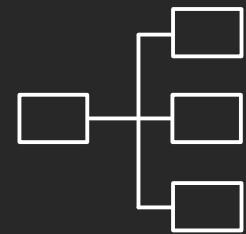
Native Windows  
compatibility



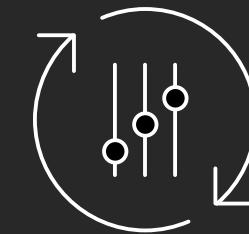
Fast and flexible  
performance



Enterprise-ready



Broad accessibility



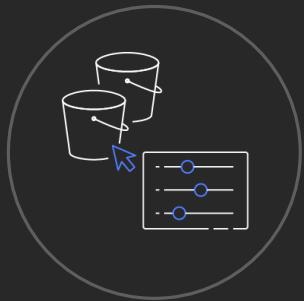
Fully managed

# 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

# 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



High-performance cloud file system

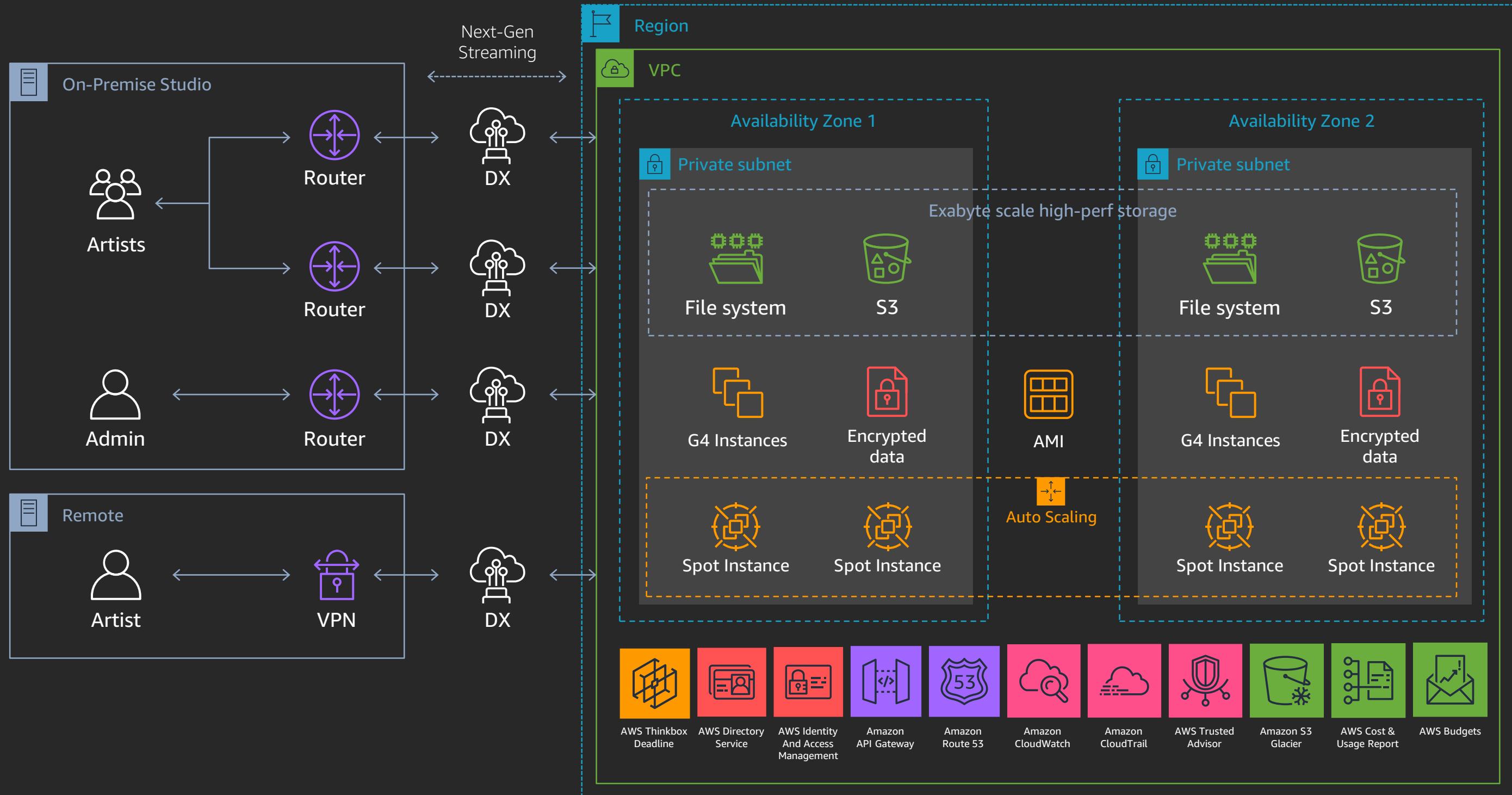
## Partner

High performance for HPC  
Can leverage Amazon S3 or Amazon Glacier



High-performance cloud file system

# Studio in the Cloud reference architecture



The background of the image is a wide, open landscape, possibly a field or a coastal area, with a horizon line in the distance. The sky above is filled with large, billowing clouds, ranging from white to dark grey, suggesting an overcast day or a approaching storm. The overall atmosphere is one of vastness and openness.

# Studio in the Cloud

Rendering

Software

Virtual Workstations

Infrastructure

# NVIDIA



# NVIDIA GPU TECHNOLOGY FOR MEDIA & ENTERTAINMENT

GPU Accelerated Performance for the Enterprise



# QUADRO VIRTUAL WORKSTATION ON AMAZON CLOUD FUELS AN INDUSTRY THAT'S INCREASINGLY MOBILE AND GRAPHICS-INTENSIVE



# TRENDS IN M&E

Companies must create quality content in less time and for less money



## Greater demands

Discerning audiences are rejecting lower quality visual effects and animation

## Dynamic landscape

Mergers, acquisitions, and new startups require scaling of compute resources and quick set up of users

## Production acceleration

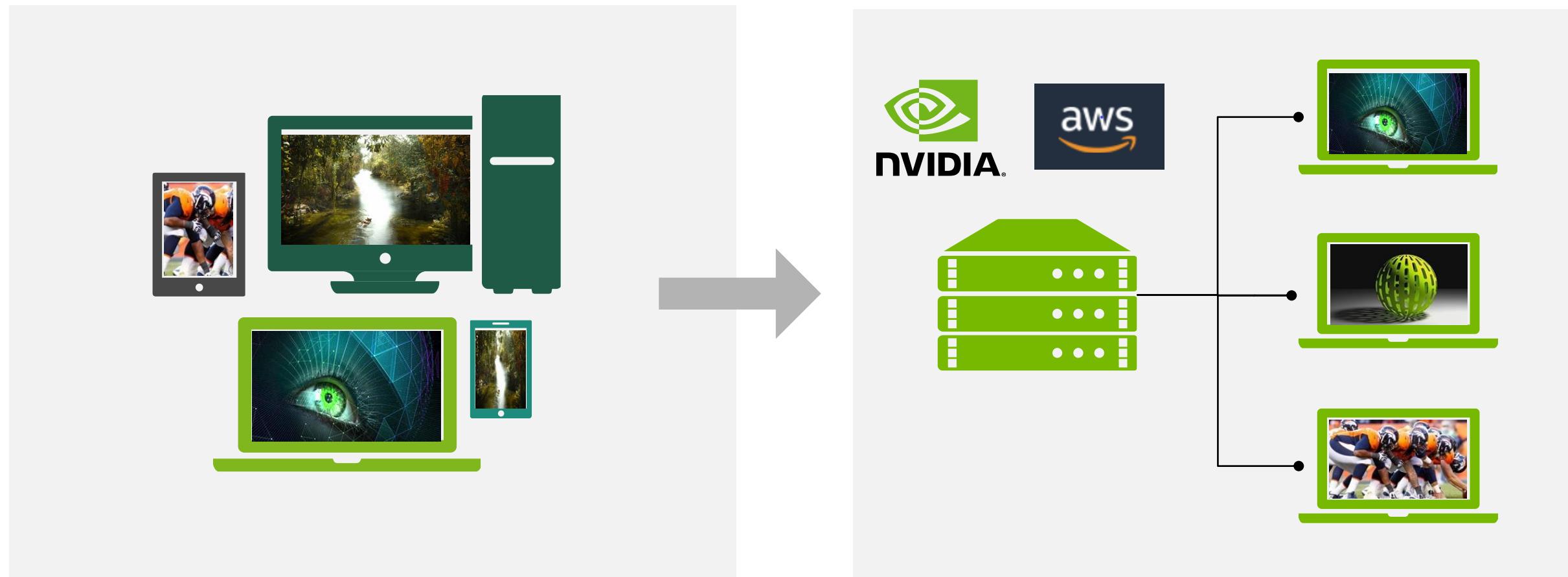
Creators are challenged to maximize iterations in shorter timeframes; firms seek to hire globally to get the best talent and keep projects moving forward

## Enhanced security

Centralize assets in the data center or cloud, to prevent costly leaks and data disruption

# QUADRO PERFORMANCE FROM AWS

NVIDIA Virtual GPU technology delivers graphics accelerated virtual desktops and applications

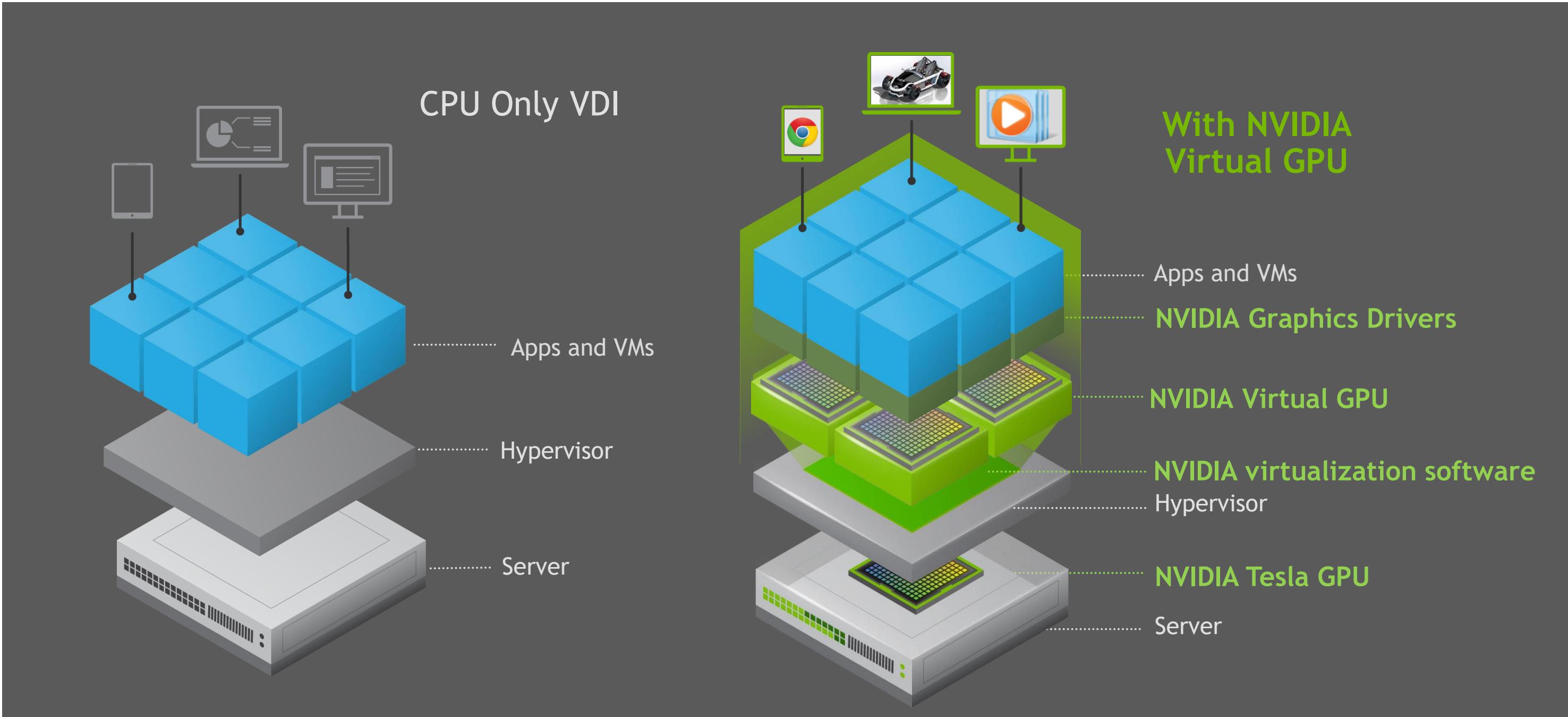


All devices have graphics

Virtual machines also need a GPU

# HOW IT WORKS

NVIDIA delivers a GPU experience to every virtual desktop



# M&E KEY USER GROUPS

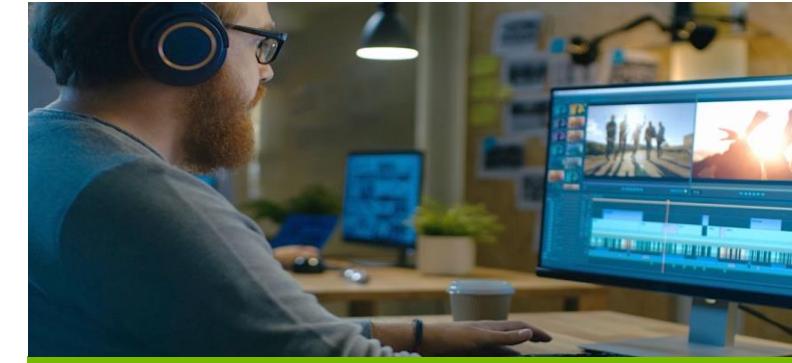


Animators, production artists, VFX producers

Rendering and making changes to graphics-intensive scenes

Recommend

AWS g4dn.4xlarge  
AWS g4dn.8xlarge  
AWS g4dn.12xlarge  
AWS g4dn.16xlarge



Video editors

Remotely viewing and editing film footage, including real-time on-air production, highlight reels, and rotoscoping

Recommend

AWS g4dn.2xlarge  
AWS g4dn.4xlarge  
AWS g4dn.8xlarge



Marketing, creative, design, illustrators

General purpose VDI using Windows 10 and graphics-rich productivity apps, as well as virtualized design and creative apps such as Adobe Creative Cloud

Recommend

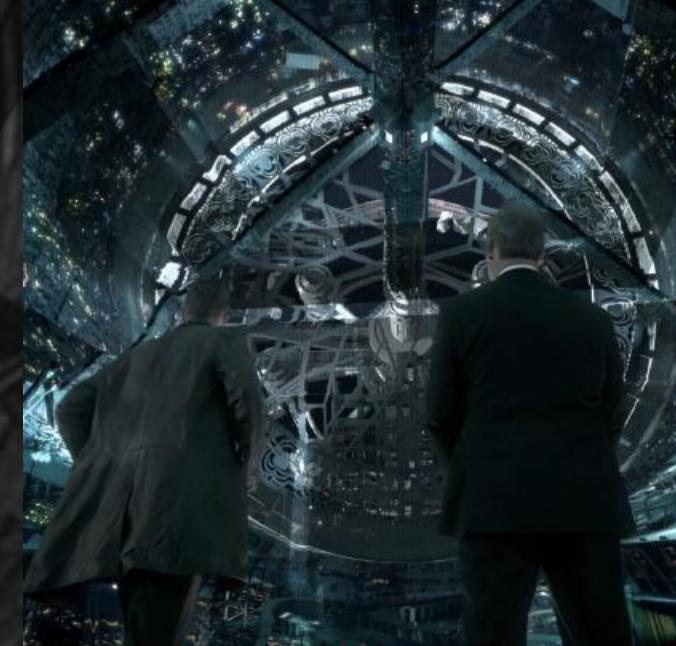
AWS g4dn.xlarge  
AWS g4dn.2xlarge

# ROCKET SCIENCE VFX

*Hybrid Studio workflow  
combining local resources  
with cloud rendering  
on AWS GPU instances*

Project: The Expanse  
Renderer: Redshift

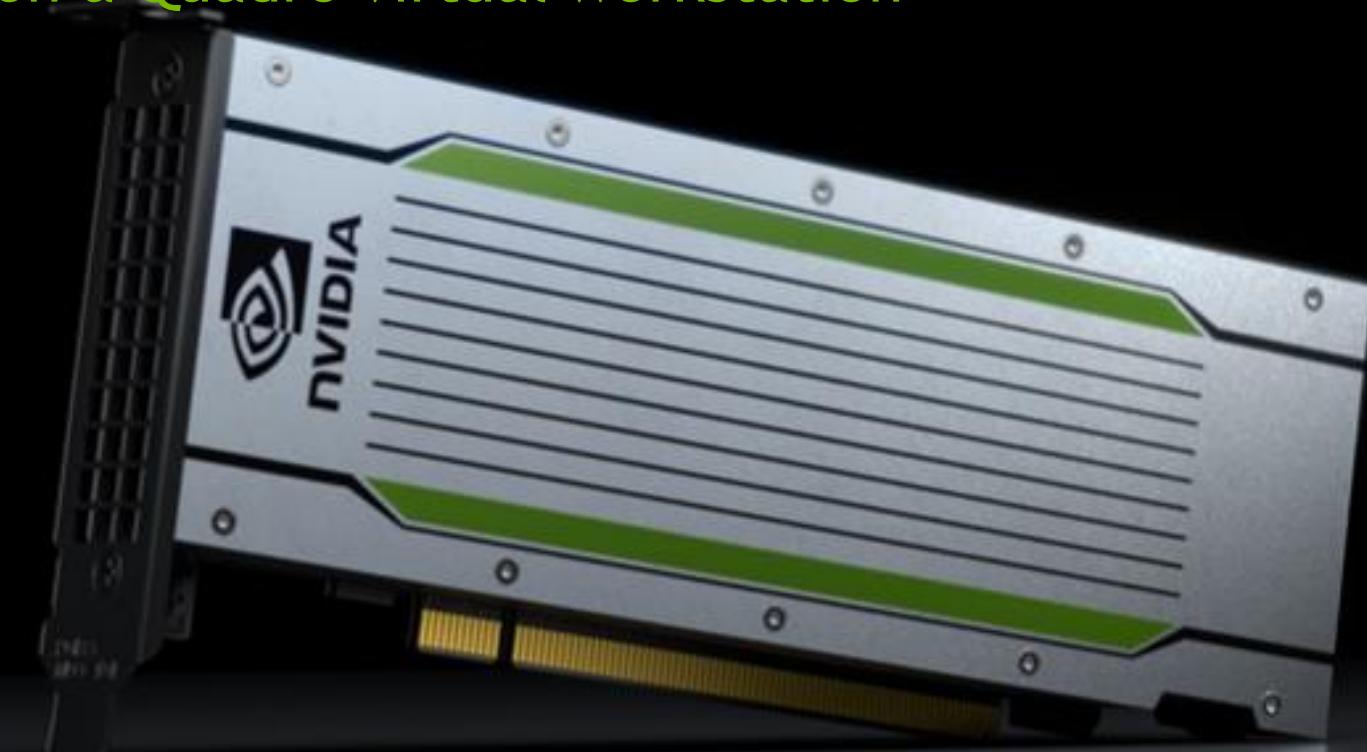
*Images Courtesy:  
Rocket Science VFX and Amazon*



# Amazon EC2 G4: NVIDIA T4 with Quadro Virtual Workstation

NVIDIA Turing generation of computer graphics on a Quadro virtual workstation

- Up to 2X graphics performance versus M60
- 5 Giga Rays per second for real-time, interactive ray tracing
- 2,560 NVIDIA CUDA® Cores
- 320 NVIDIA Turing Tensor Cores
- 40 RT Cores for RTX-enabled applications
- 16 GB GDDR6 Memory



# Amazon EC2 G4: NVIDIA T4 with Quadro Virtual Workstation

NVIDIA Turing generation of computer graphics on a Quadro virtual workstation

- Up to 2X graphics performance versus M60
- 5 Giga Rays per second for real-time, interactive ray tracing
- 2,560 NVIDIA CUDA® Cores
- 320 NVIDIA Turing Tensor Cores
- 40 RT Cores for RTX-enabled applications
- 16 GB GDDR6 Memory

## WHY SHOULD YOU CARE?



# Amazon EC2 G4: NVIDIA T4 with Quadro Virtual Workstation

NVIDIA Turing generation of computer graphics on a Quadro virtual workstation

- Up to 2X graphics performance versus M60
  - 5 Giga Rays per second for real-time, interactive ray tracing
  - 2,560 NVIDIA CUDA® Cores
  - 320 NVIDIA Turing Tensor Cores
  - 40 RT Cores for RTX-enhanced applications
  - 16 GB GDDR6 Memory
- 



# RTX RAY TRACING



# RTX RAY TRACING

*Turing Architecture GPUs  
to optimize your  
production workflow*

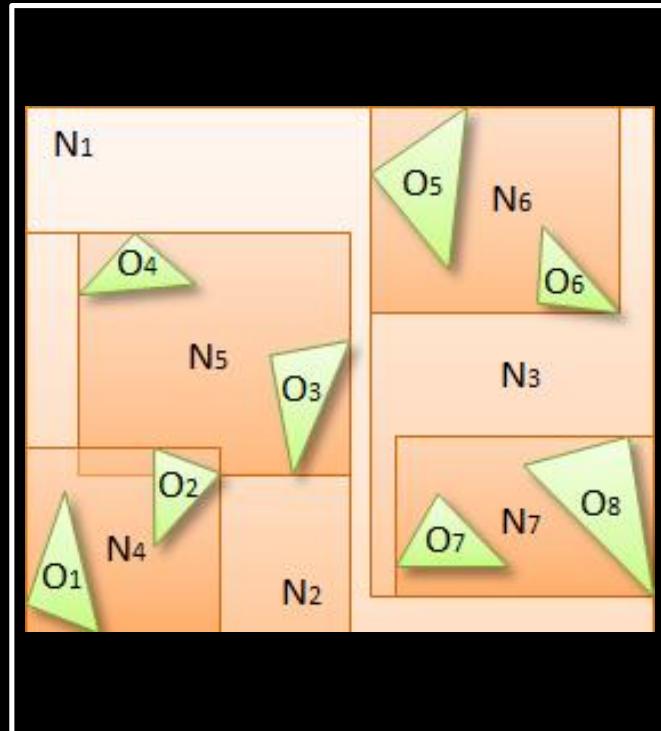
Experience faster content creation workflows and GPU-accelerated rendering with NVIDIA RTX GPUs, Maya Batch Rendering and Arnold

*Asset Courtesy:  
Image Engine and Netflix*



# A FULL PICTURE

## Breakdown of a ray tracer



Acceleration structure  
management



Shading



Denoising



Ray tracing

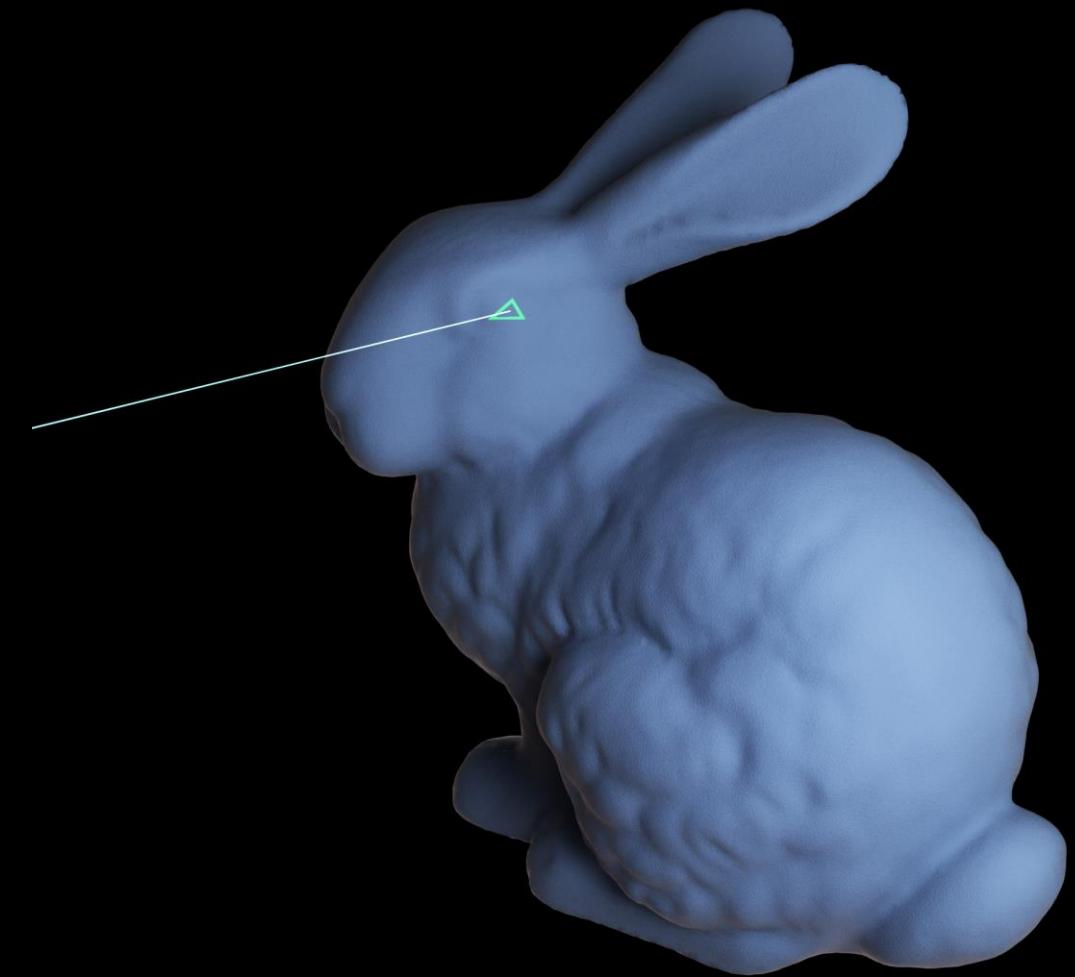
# OptiX: A General Purpose Ray Tracing Engine

Steven G. Parker<sup>1\*</sup> James Bigler<sup>1</sup> Andreas Dietrich<sup>1</sup> Heiko Friedrich<sup>1</sup> Jared Hoberock<sup>1</sup> David Luebke<sup>1</sup>  
David McAllister<sup>1</sup> Morgan McGuire<sup>1,2</sup> Keith Morley<sup>1</sup> Austin Robison<sup>1</sup> Martin Stich<sup>1</sup>  
NVIDIA<sup>1</sup> Williams College<sup>2</sup>



**Figure 1:** Images from various applications built with OptiX. Top: Physically based light transport through path tracing. Bottom: Ray tracing of a procedural Julia set, photon mapping, large-scale line of sight and collision detection, Whitted-style ray tracing of dynamic geometry, and ray traced ambient occlusion. All applications are interactive.

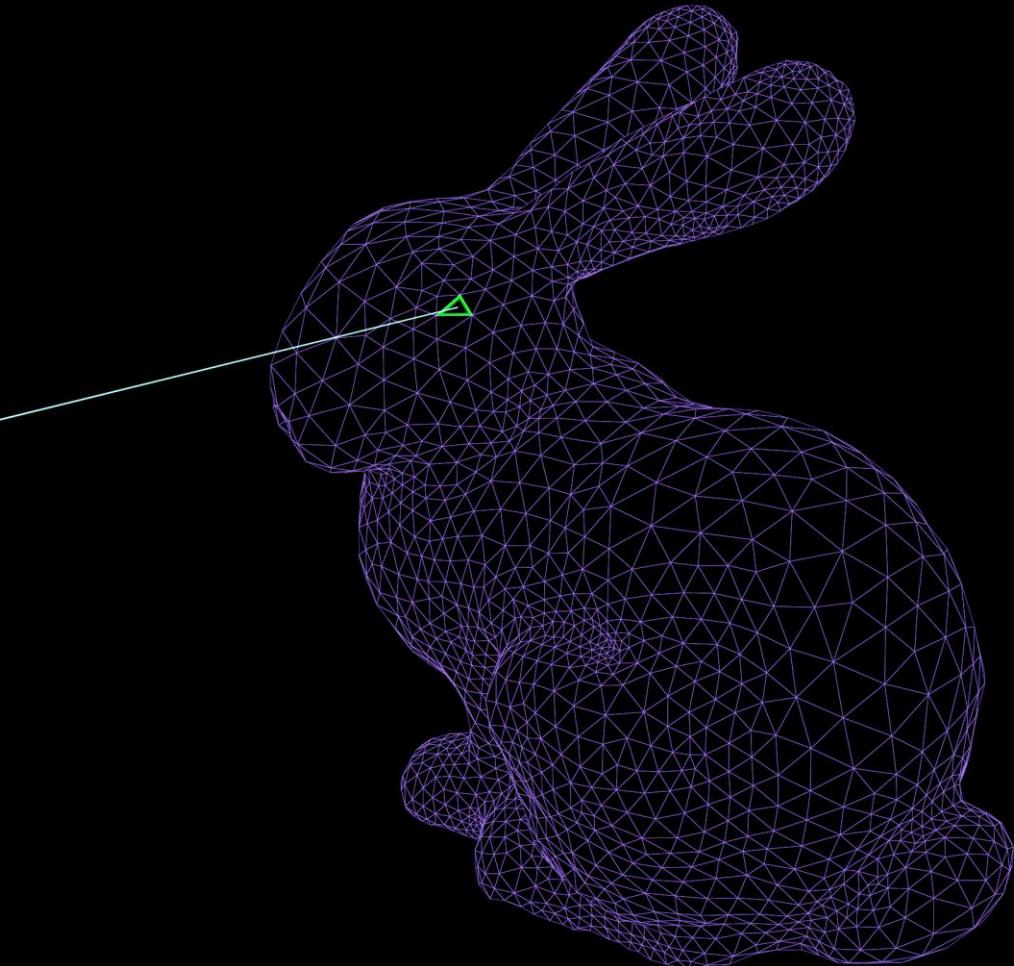
# THE RAY TRACING CHALLENGE



# THE RAY TRACING CHALLENGE

How to find the “needle” in the triangle data “haystack”?

|        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 548.09 | 26.60  | 977.30 | 309.07 | 942.51 | 943.84 | 954.14 | 546.50 | 495.62 |
| 954.35 | 457.14 | 614.35 | 682.99 | 714.16 | 985.30 | 412.09 | 583.21 | 245.93 |
| 23.22  | 157.44 | 644.73 | 867.76 | 820.81 | 273.56 | 297.39 | 896.97 | 811.89 |
| 873.92 | 188.68 | 664.87 | 216.65 | 514.58 | 362.04 | 457.05 | 796.50 | 786.00 |
| 292.98 | 320.87 | 662.20 | 89.48  | 793.76 | 696.15 | 171.60 | 400.56 | 833.53 |
| 67.24  | 956.67 | 322.79 | 623.92 | 81.26  | 320.69 | 504.97 | 900.10 | 705.55 |
| 0.39   | 828.89 | 614.22 | 830.05 | 14.85  | 895.96 | 351.53 | 577.34 | 508.73 |
| 182.26 | 970.25 | 48.49  | 291.07 | 354.99 | 776.41 | 1.16   | 25.56  | 581.09 |
| 539.16 | 273.37 | 656.67 | 407.03 | 578.86 | 151.16 | 142.14 | 494.58 | 534.29 |
| 495.54 | 159.08 | 763.44 | 395.82 | 343.53 | 149.02 | 945.97 | 161.74 | 74.26  |
| 476.37 | 569.43 | 238.29 | 817.46 | 648.77 | 393.45 | 84.25  | 249.51 | 663.99 |
| 733.85 | 250.94 | 669.23 | 850.30 | 890.15 | 833.83 | 790.60 | 168.37 | 10.38  |
| 595.39 | 611.15 | 846.96 | 15.34  | 893.95 | 311.69 | 258.05 | 174.31 | 88.11  |
| 109.68 | 712.32 | 436.50 | 876.75 | 657.30 | 205.79 | 881.99 | 279.31 | 255.43 |
| 15.18  | 22.55  | 196.99 | 157.67 | 382.02 | 287.38 | 121.32 | 592.23 | 91.53  |
| 64.05  | 444.40 | 395.57 | 15.49  | 531.77 | 801.03 | 11.21  | 189.37 | 342.66 |
| 419.16 | 172.03 | 3.11   | 784.85 | 21.57  | 630.50 | 639.69 | 128.88 | 688.59 |
| 345.05 | 512.51 | 429.58 | 306.48 | 908.12 | 879.09 | 760.31 | 243.03 | 173.92 |
| 285.29 | 289.68 | 345.85 | 841.70 | 924.04 | 899.56 | 519.15 | 706.46 | 599.07 |
| 77.45  | 518.36 | 105.82 | 364.65 | 134.91 | 69.86  | 724.23 | 531.37 | 57.28  |
| 943.10 | 161.83 | 70.13  | 922.84 | 30.88  | 913.59 | 635.33 | 288.13 | 409.63 |
| 396.59 | 81.56  | 812.85 | 624.20 | 718.88 | 746.88 | 410.21 | 771.24 | 898.46 |
| 903.82 | 596.27 | 289.02 | 317.16 | 267.82 | 762.04 | 454.07 | 5.85   | 673.96 |
| 501.70 | 802.38 | 46.61  | 883.84 | 41.63  | 947.00 | 812.63 | 909.83 | 884.7  |

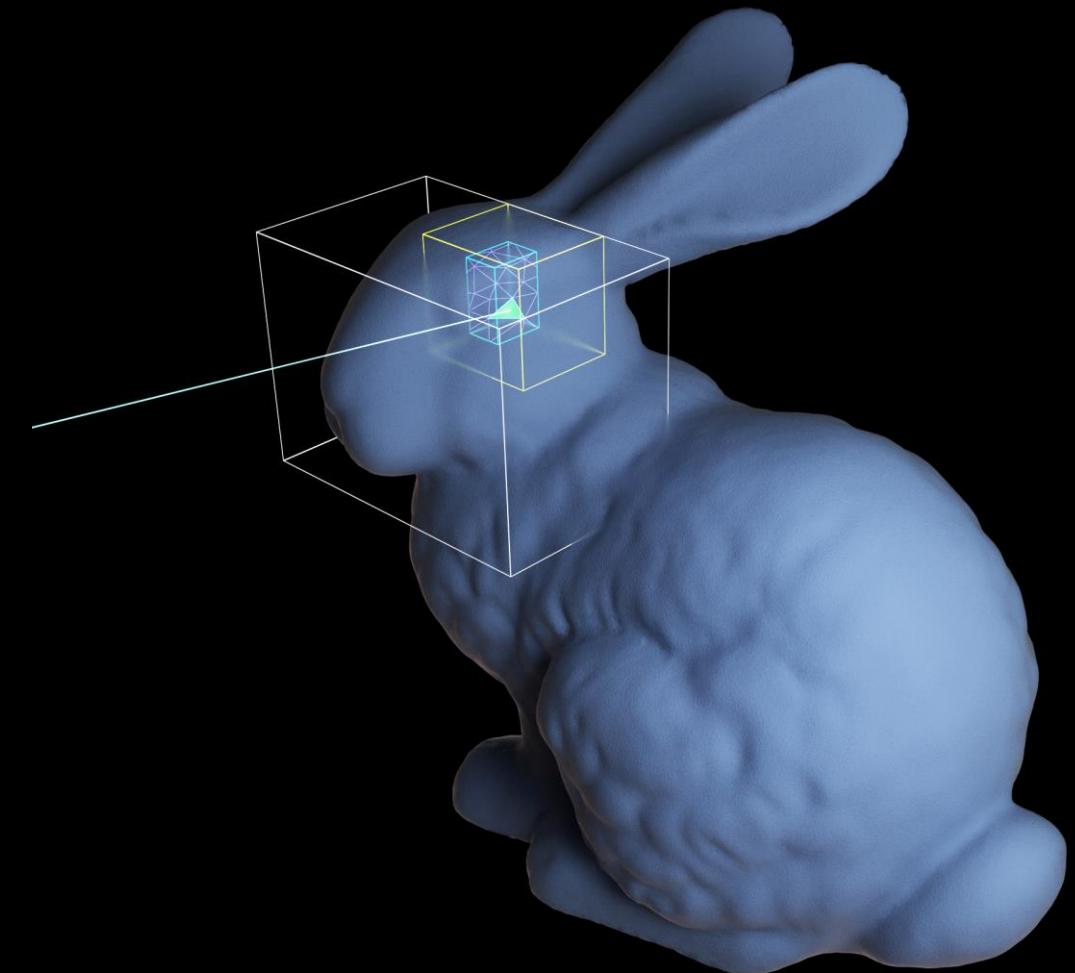


# THE RAY TRACING CHALLENGE

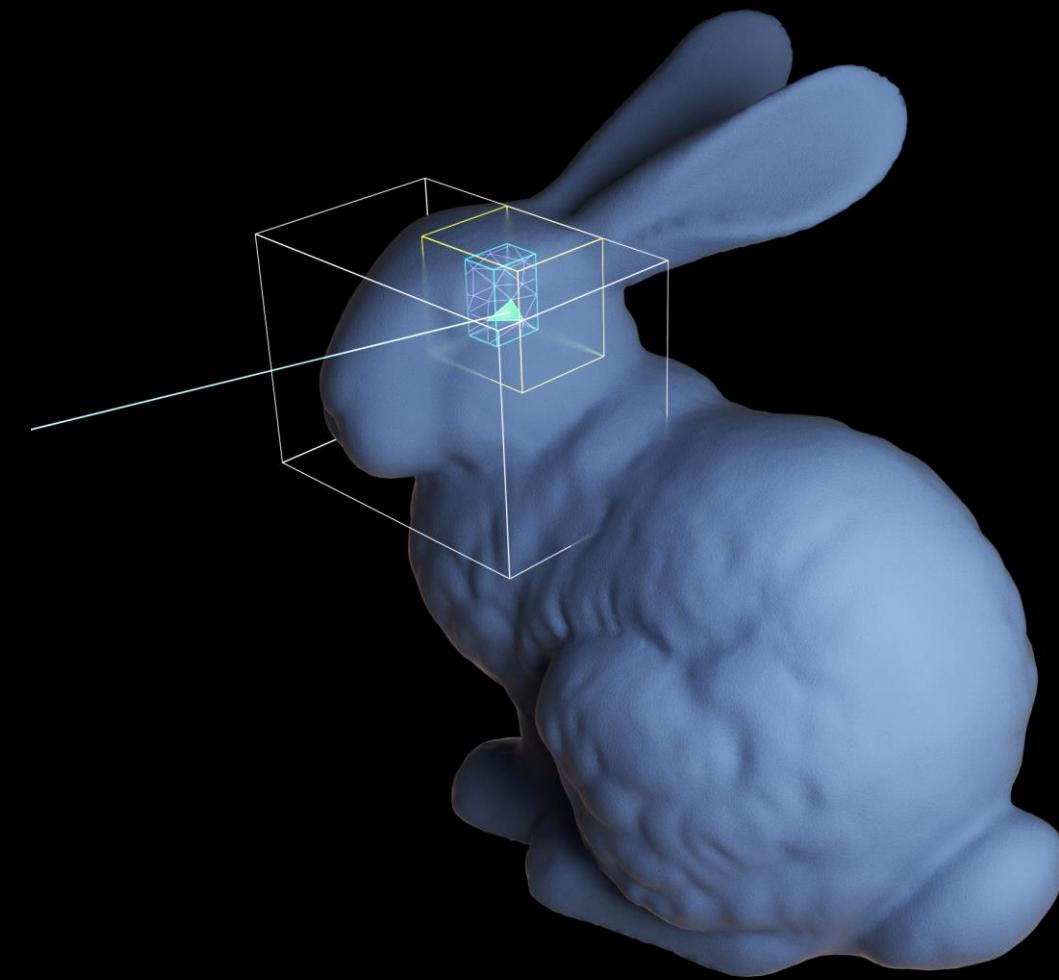
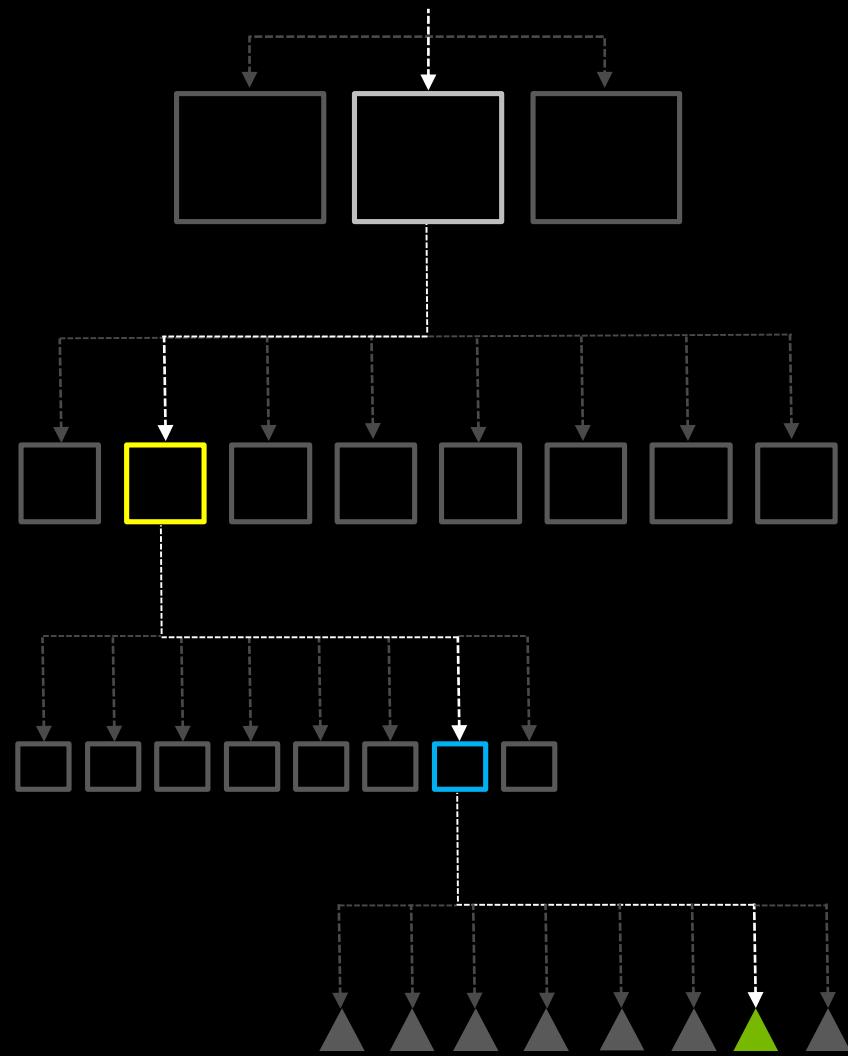
How to find the “needle” in the triangle data “haystack”?

BVH = Bounding Volume Hierarchy

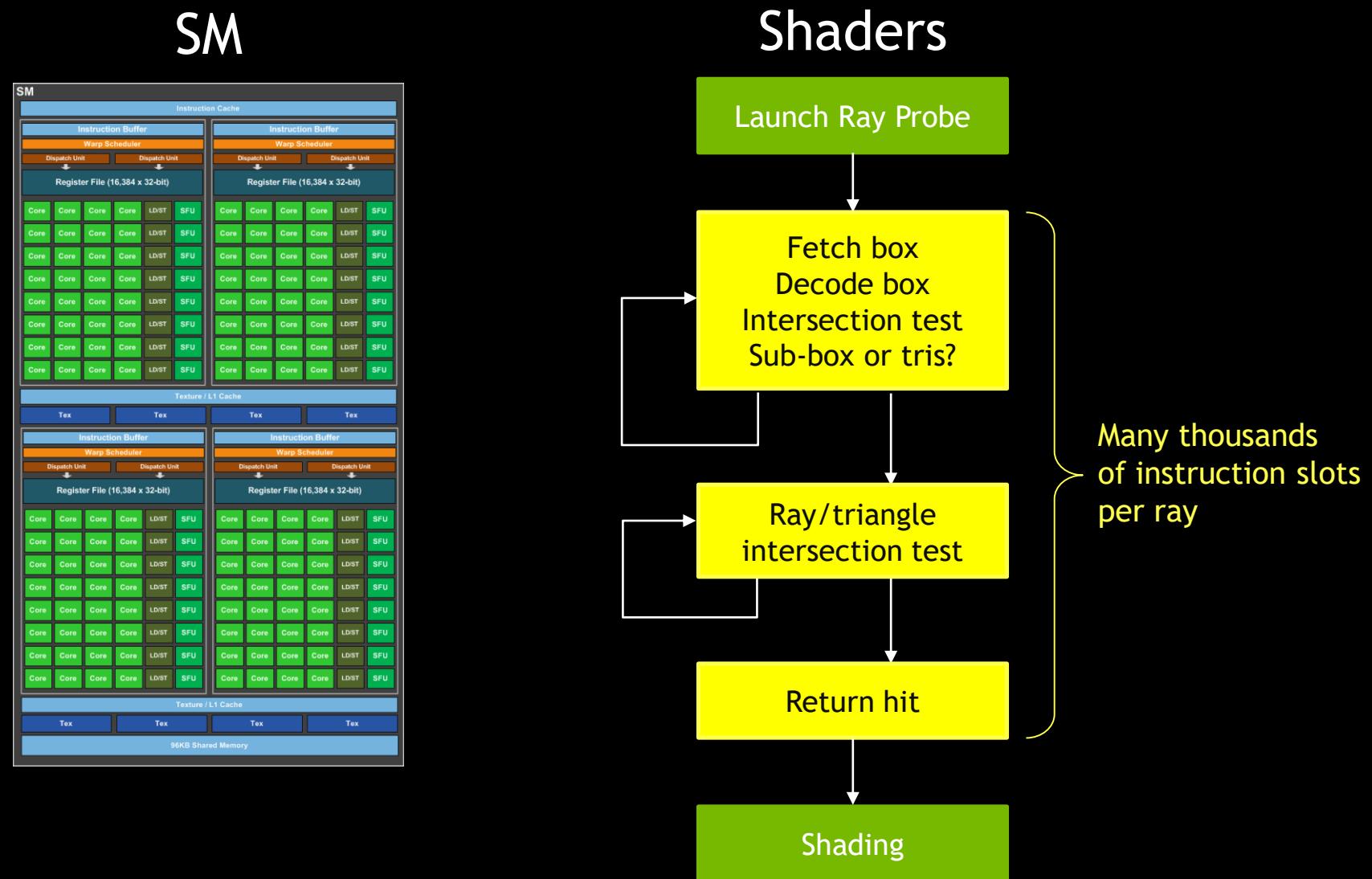
AS = Acceleration Structure



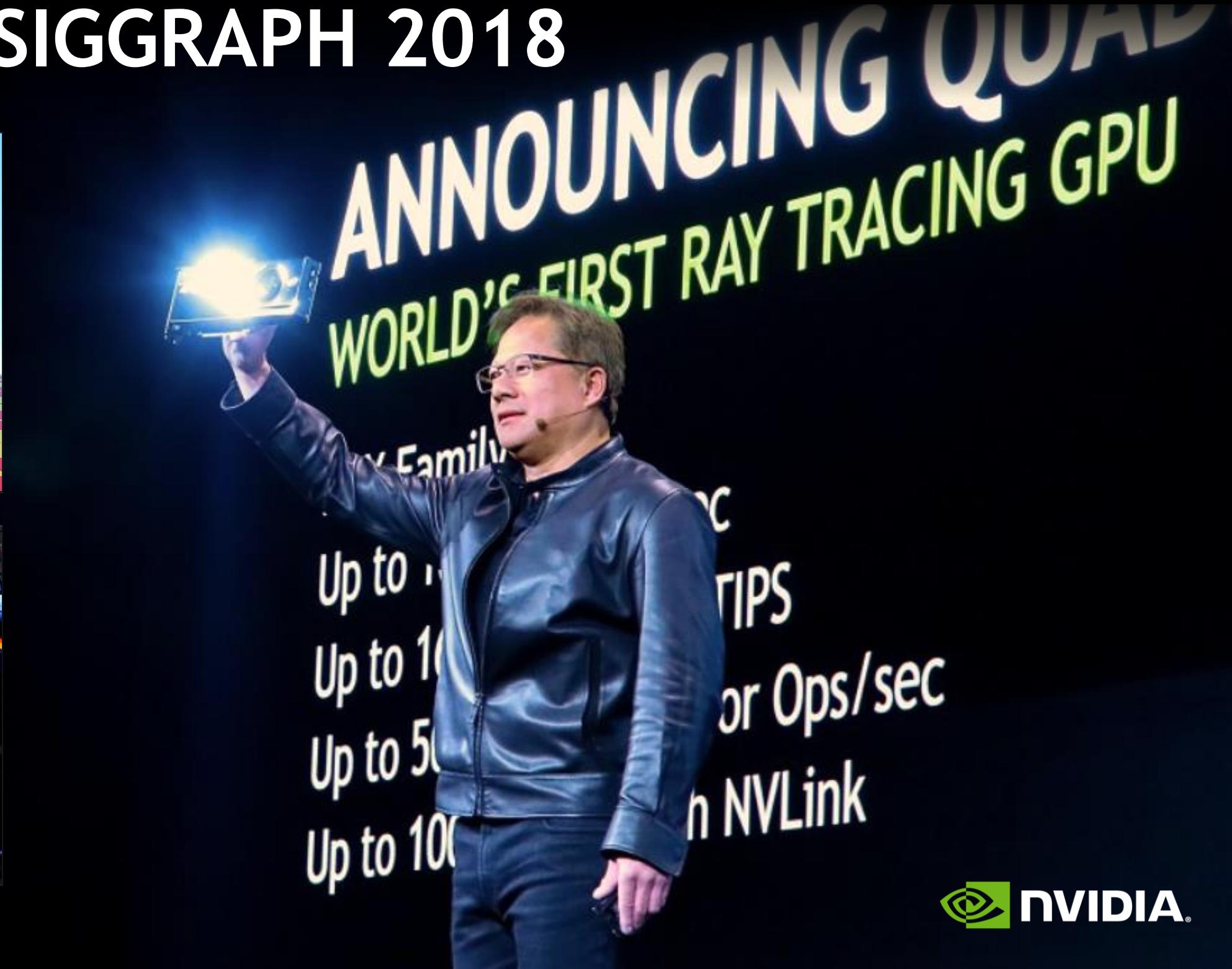
# BVH ALGORITHM



# RENDERING WITHOUT RTX

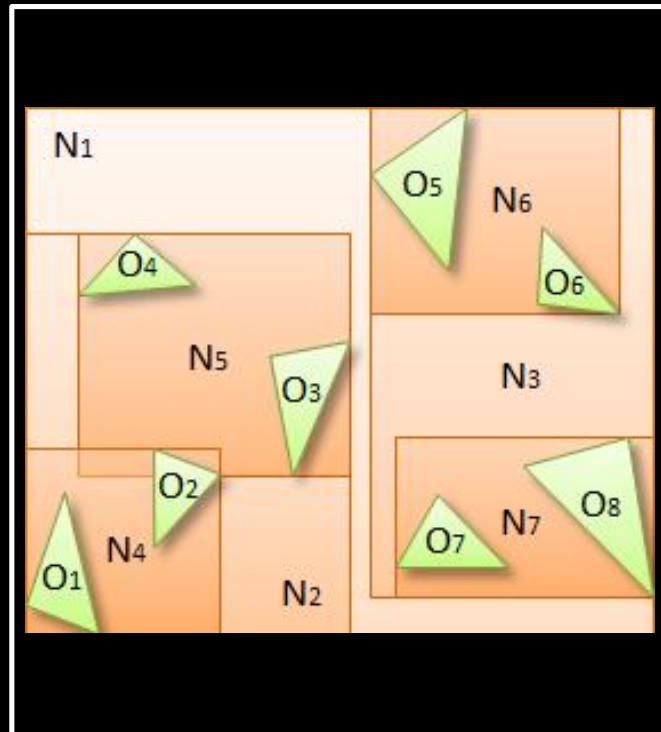


# SIGGRAPH 2018



# A FULL PICTURE

## Breakdown of a ray tracer



Acceleration structure  
management



Shading



Denoising



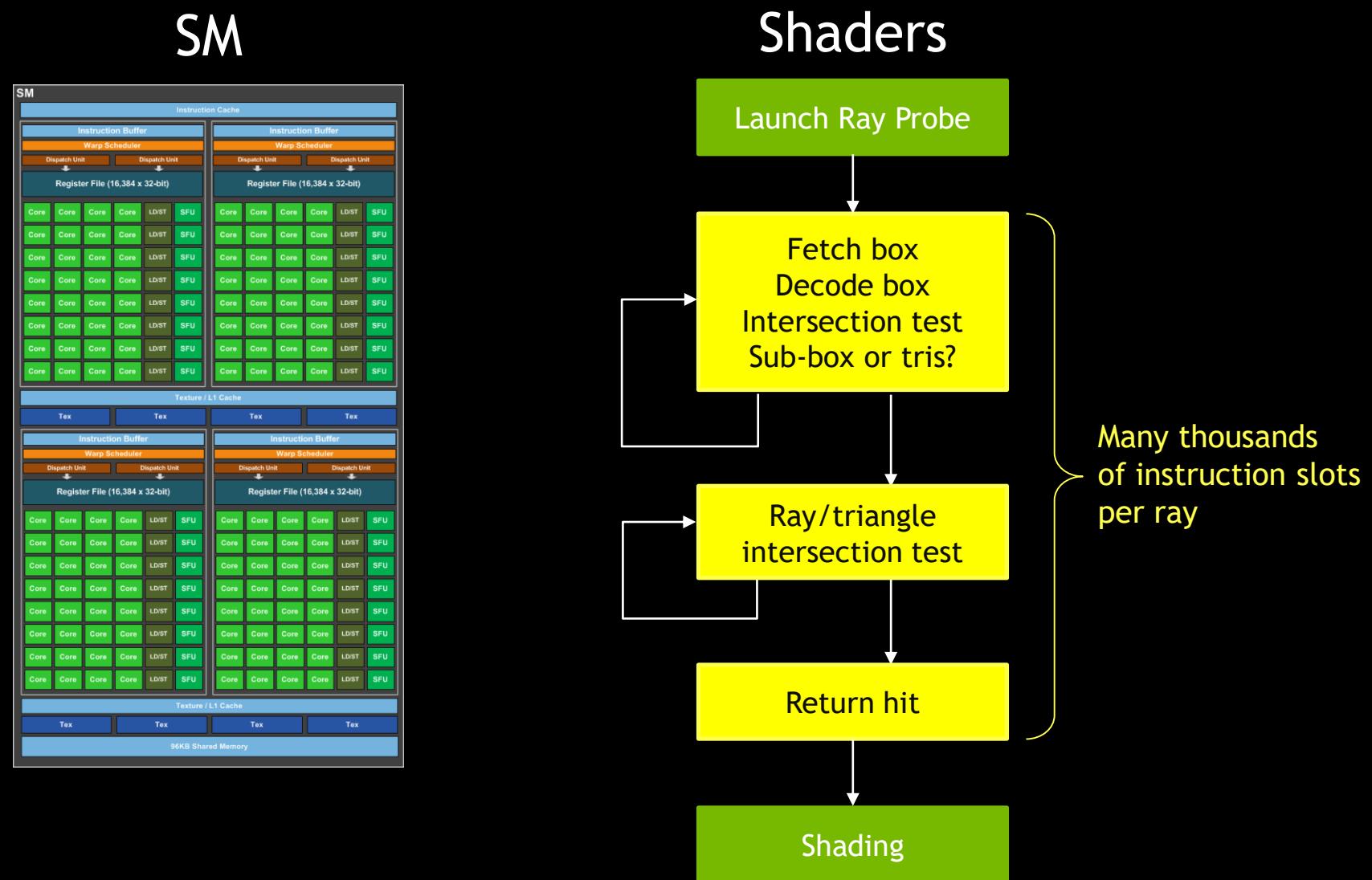
Ray tracing

CUDA cores

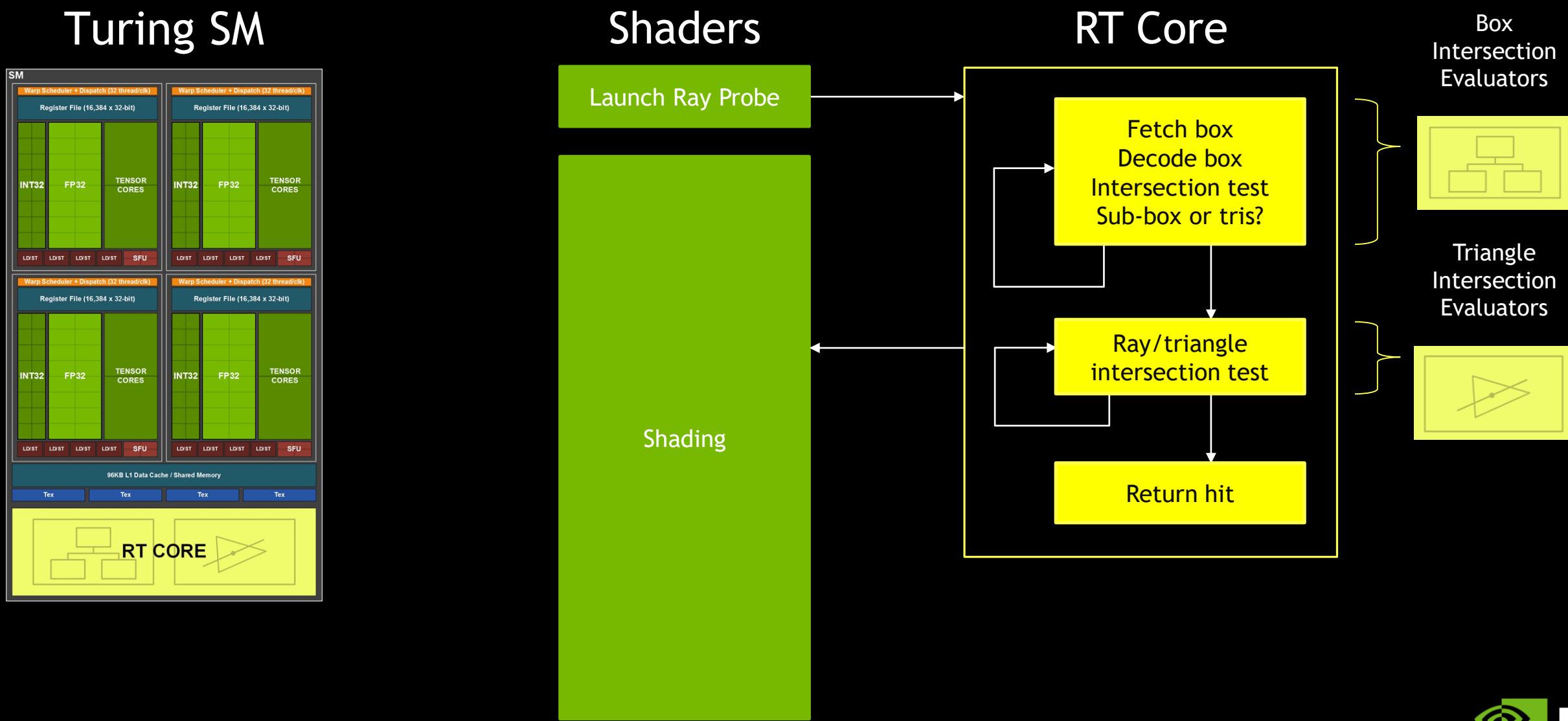
Tensor cores

RT cores

# RENDERING WITHOUT RTX



# RENDERING WITH RTX





CPU render



GPU render







CUDA-X AI



# OPTIX RECURRENT DENOISING AUTOENCODER

Interactive Reconstruction of Monte Carlo Image Sequences using a Recurrent Denoising Autoencoder

CHAKRAVARTY R. ALLA CHAITANYA (1,2,4)

ANTON S. KAPLANYAN (1)

CHRISTOPH SCHIED (1,3)

MARCO SALVI (1)

AARON LEFOHN (1)

DEREK NOWROUZEZAHRAI (4)

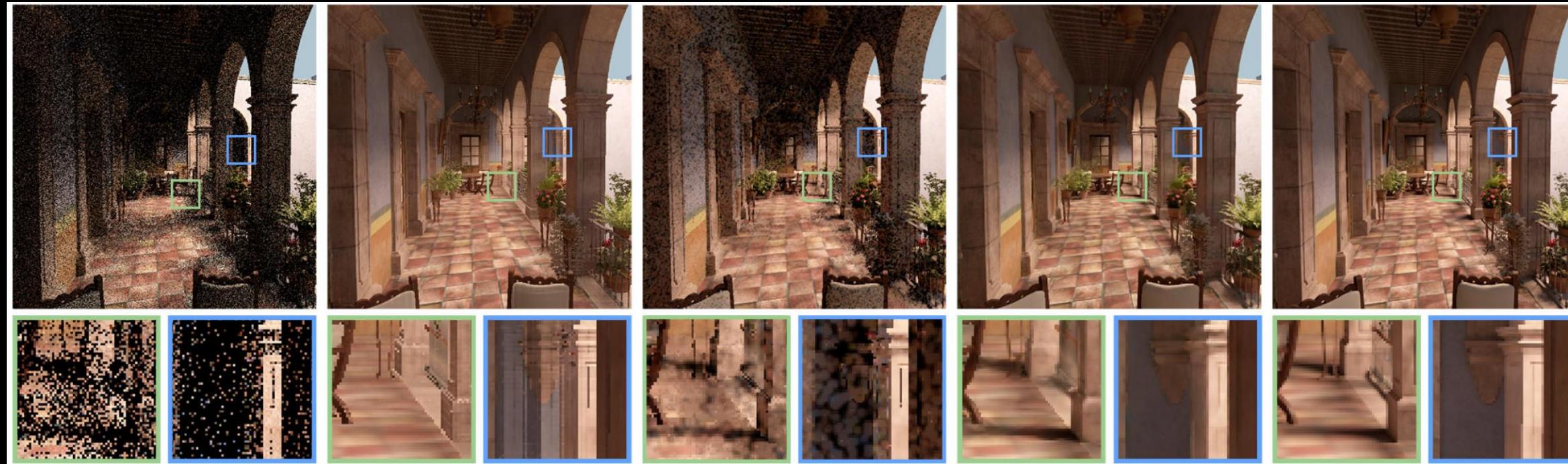
TIMO AILA (1)

(1)Nvidia

(2)University of Montreal

(3) Karlsruhe Institute of Technology

(4) McGill University



<http://www.developer.nvidia.com/optix-denoiser>





# SUPER RESOLUTION

Smart upscaling using Deep Learning



Upscale



Available at: <https://gwmt.nvidia.com>



# BILINEAR



# BICUBIC



# DEEP LEARNING



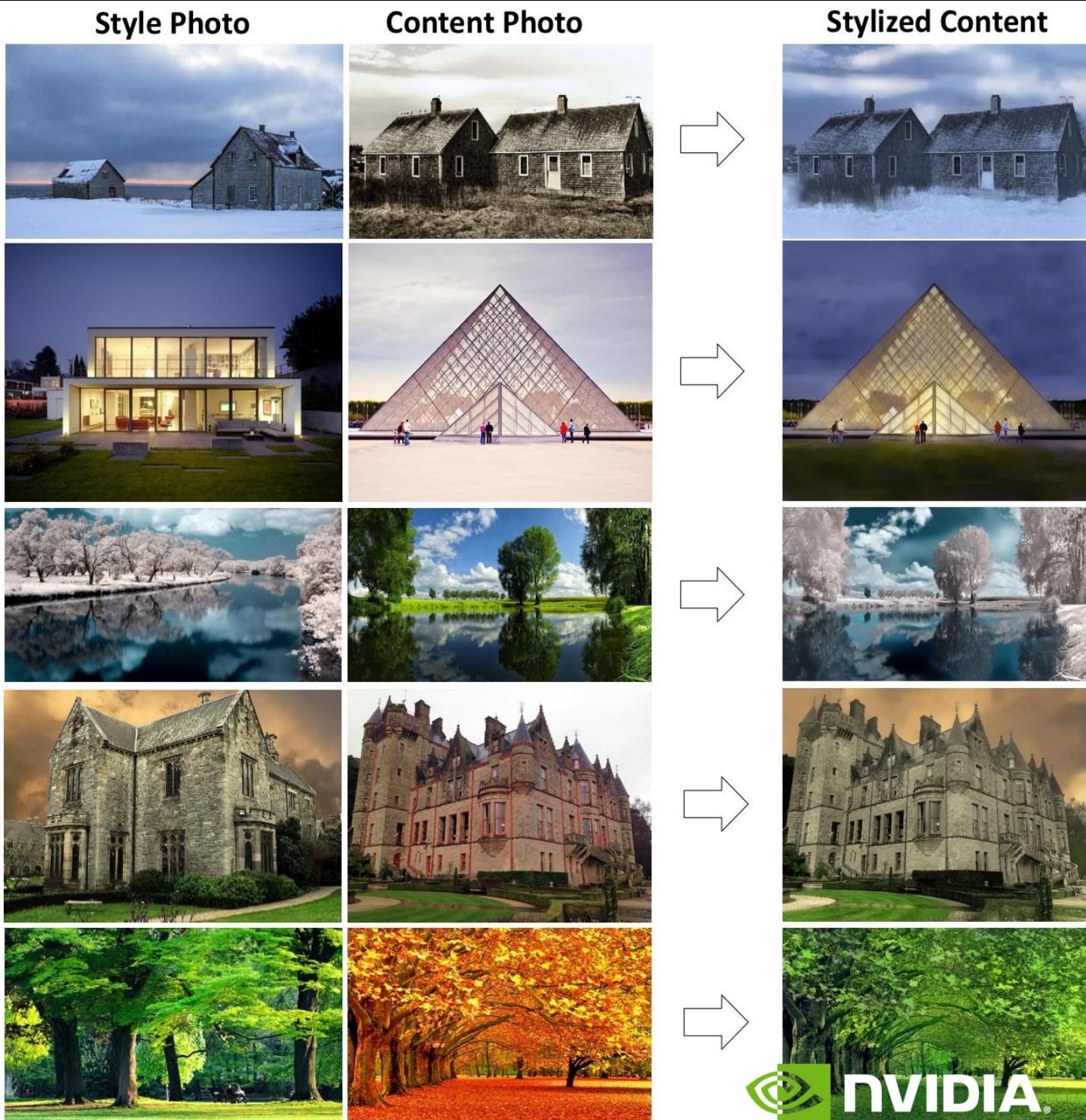
# FAST PHOTO STYLE

## NVIDIA's Style Transfer Research

Applies “features” of one image to another

NVIDIA approach applies style and smoothing for enhanced result

<https://github.com/NVIDIA/FastPhotoStyle>









## Deep Learning for locomotion

Realistic character motion created in real-time through artificial intelligence

Replicate the nuances of an animator or performer

Adapt to changing environments

Scale easily from single characters to crowds

Give time to the artists for the “Hero” moments

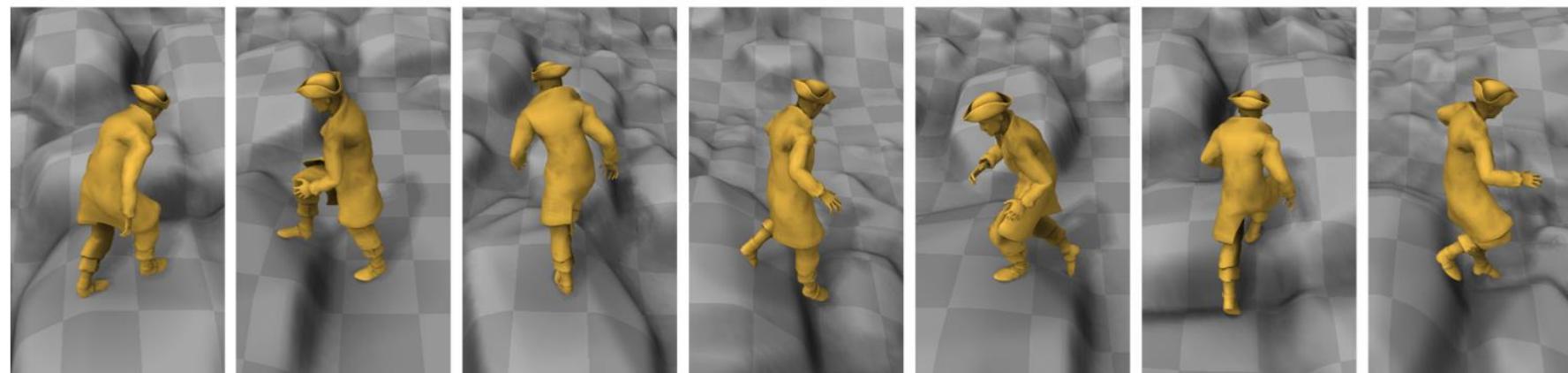
# DEEP LEARNING ANIMATION: PFNN

## Phase-Functioned Neural Networks for Character Control

DANIEL HOLDEN, University of Edinburgh

TAKU KOMURA, University of Edinburgh

JUN SAITO, Method Studios



<http://theorangeduck.com/page/phase-functioned-neural-networks-character-control>



# HOW DOES IT WORK?

## Get the data

- 3D skeletal data
- Animated or motion captured
- All possible actions required

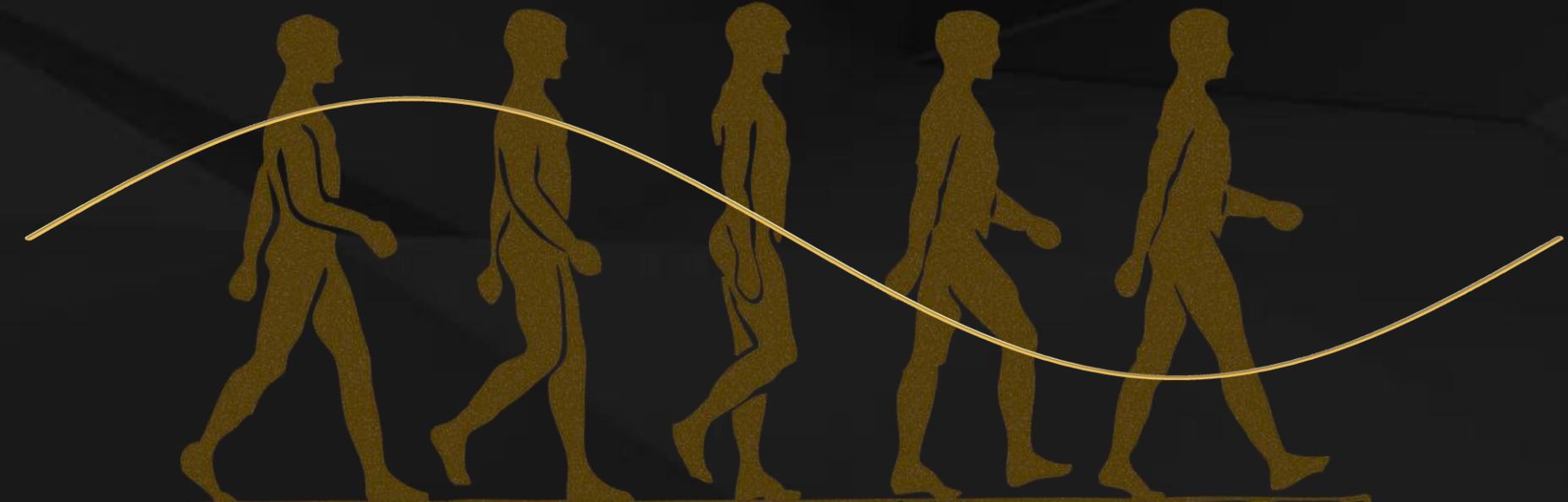


Many thanks to the Fox VFX Lab  
for time on their motion capture stage

# HOW DOES IT WORK?

## Label the data

- Action (running, walking, crouching, etc.)
- Phase
- Footstep positions







THANK YOU

# Thank you!

**Jason Schleifer & Rick Grandy**

[aws-thinkbox-bd@amazon.com](mailto:aws-thinkbox-bd@amazon.com)  
[rgrandy@nvidia.com](mailto:rgrandy@nvidia.com)



Please complete the session  
survey in the mobile app.