Investor Presentation Q2 FY24

August 28, 2023

Except for the historical information contained herein, certain matters in this presentation including, but not limited to, statements as to: our financial position; our markets, market opportunity, demand and growth drivers; CSPs undertaking a generational transition to update their infrastructure for the new era of accelerated computing and AI; enterprises racing to adopt generative AI; our belief that global end demand for gaming has returned to growth; the growing number of GPU cloud providers deploying in volume HGX systems based on our Hopper and Ampere GPU architecture; our expectation that data center supply will increase each quarter through next year; a large upgrade opportunity ahead; the likelihood that the seasonality of the overall gaming market will shift; RTX/DLSS ecosystem continuing to expand; the timing of a desktop workstation refresh; the expanding opportunity in Professional Visualization for RTX technology; the impact of our partnership with MediaTek; our financial outlook, expectation of sequential growth to be driven largely by Data Center, with Gaming and Professional Visualization also contributing, and our expected tax rates for the third quarter of fiscal 2024; the benefits, impact, performance, features and availability of our products and technologies; the benefits, impact, features and timing of our collaborations or partnerships; NVIDIA accelerated computing being broadly recognized as the way to advance computing as Moore's law ends; accelerated computing being needed to tackle the most impactful opportunities of our time; NVIDIA's expanding accelerated computing ecosystem; NVIDIA's value to every stakeholder in the ecosystem; AI as the greatest technology force of our time; generative AI unlocking new opportunities; data centers across industries becoming AI factories; the next wave of AI being robotics and industrial digitalization, and Omniverse being essential for robotics; NVIDIA's acceleration stacks and ecosystems helping to bring AI to the world's largest industries; NVIDIA's AI expertise and scale helping to revolutionize businesses; NVIDIA software and services enabling the world's enterprises to revolutionize industries with AI; the cost and time-to-solution savings of application speed-ups; our dividend program plan; our Automotive design win pipeline, ramp and production expectations; our aim to engage manufacturing suppliers and goal of effecting supplier adoption of science-based environmental targets by fiscal 2026; and our plan for 100% renewable electricity for our operations and data centers by fiscal 2025 and annually thereafter are forward-looking statements.

These forward-looking statements and any other forward-looking statements that go beyond historical facts that are made in this presentation are subject to risks and uncertainties that may cause actual results to differ materially. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences and demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems and other factors.

NVIDIA has based these forward-looking statements largely on its current expectations and projections about future events and trends that it believes may affect its financial condition, results of operations, business strategy, short-term and long-term business operations and objectives, and financial needs. These forward-looking statements are subject to a number of risks and uncertainties, and you should not rely upon the forward-looking statements as predictions of future events. The future events and trends discussed in this presentation may not occur and actual results could differ materially and adversely from those anticipated or implied in the forward-looking statements. Although NVIDIA believes that the expectations reflected in the forward-looking statements are reasonable, the company cannot guarantee that future results, levels of activity, performance, achievements or events and circumstances reflected in the forward-looking statements will occur. Except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances. For a complete discussion of factors that could materially affect our financial results and operations, please refer to the reports we file from time to time with the SEC, including our most recent Annual Report on Form 10-K, Quarterly Reports on Form 10-Q, and Current Reports on Form 8-K. Copies of reports we file with the SEC are posted on our website and are available from NVIDIA without charge.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements within are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

NVIDIA uses certain non-GAAP measures in this presentation including non-GAAP gross profit, non-GAAP gross margin, non-GAAP operating expenses, non-GAAP operating income, non-GAAP operating margin, non-GAAP net income, non-GAAP diluted earnings per share, and free cash flow. NVIDIA believes the presentation of its non-GAAP financial measures enhances investors' overall understanding of the company's historical financial performance. The presentation of the company's non-GAAP financial measures is not meant to be considered in isolation or as a substitute for the company's financial results prepared in accordance with GAAP, and the company's non-GAAP measures may be different from non-GAAP measures used by other companies. Further information relevant to the interpretation of non-GAAP financial measures, and reconciliations of these non-GAAP financial measures to the most comparable GAAP measures, may be found in the slide titled "Reconciliation of Non-GAAP to GAAP Financial Measures".

Content

- Q2 FY24 Earnings Summary
- Key Announcements This Quarter
- NVIDIA Overview
- Financials
- Reconciliation of Non-GAAP to GAAP Financial Measures

Q2 FY24
Earnings Summary

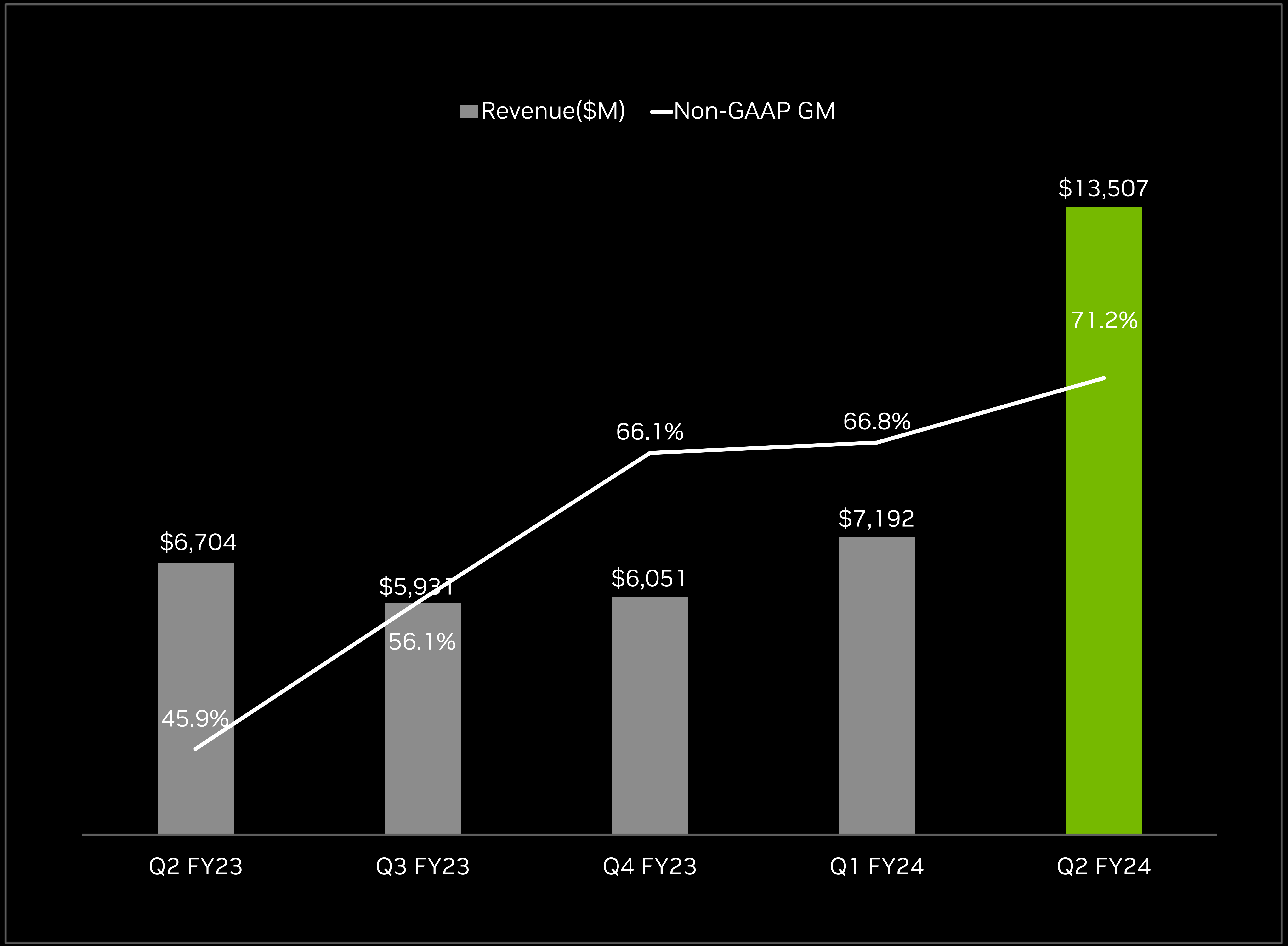
Highlights

- **Exceptional growth driven by Data Center**
  - Total revenue up 101% Y/Y to $13.51B, above outlook of $11.00B +/- 2%
  - Data Center up 171% Y/Y to $10.32B
  - Gaming up 22% Y/Y to $2.49B

- **Record Data Center revenue on tremendous demand for NVIDIA accelerated computing and AI platforms**
  - CSPs are undertaking a generational transition to upgrade their infrastructure for the new era of accelerated computing and AI
  - Consumer internet companies' investments in AI purpose-built data center infrastructure are already generating significant returns
  - Enterprises are racing to deploy generative AI, driving strong consumption of NVIDIA-powered instances in the cloud, as well as demand for on-prem infrastructure

- **Strong Gaming growth fueled by GeForce RTX 40 Series GPUs for laptops and desktops**
  - End-customer demand was solid, and consistent with seasonality; believe global end demand has returned to growth
  - Laptop GPUs posted strong growth in the key back-to-school season, led by RTX 4060 GPUs
  - Launched GeForce RTX 4060 and 4060 Ti GPUs for desktops, bringing Ada Lovelace down to price points as low as $299

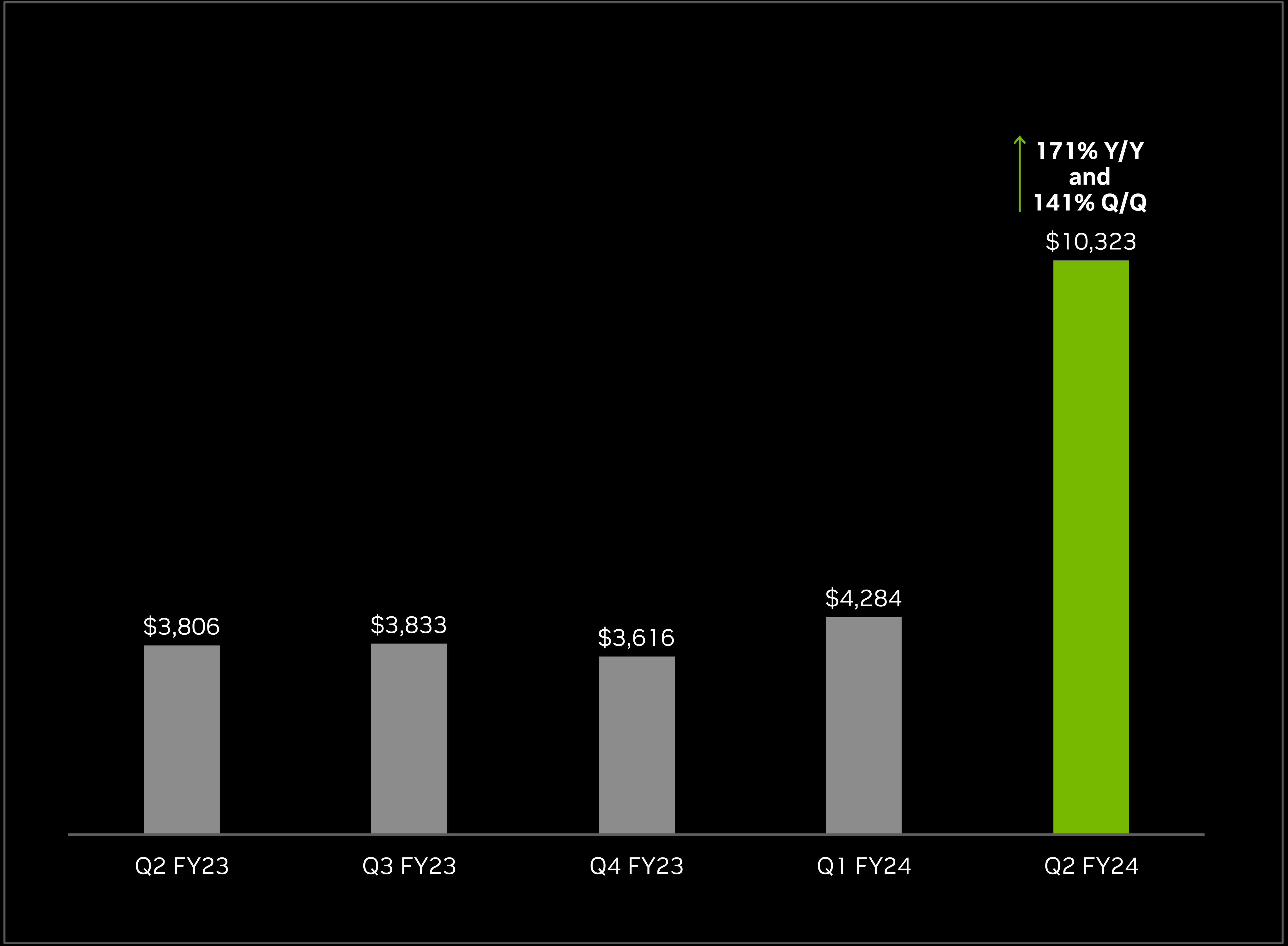
Q2 FY24 Financial Summary



	GAAP			Non-GAAP		
	Q2 FY24	**Y/Y**	**Q/Q**	**Q2 FY24**	**Y/Y**	**Q/Q**
Revenue	$13,507	+101%	+88%	$13,507	+101%	+88%
Gross Margin	70.1%	+26.6 pts	+5.5 pts	71.2%	+25.3 pts	+4.4 pts
Operating Income	$6,800	+1,263%	+218%	$7,776	+487%	+155%
Net Income	$6,188	+843%	+203%	$6,740	+422%	+148%
Diluted EPS	$2.48	+854%	+202%	$2.70	+429%	+148%
Cash Flow from Ops	$6,348	+400%	+118%	$6,348	+400%	+118%

All dollar figures are in millions other than EPS. Refer to Appendix for reconciliation of Non-GAAP measures.

Data Center

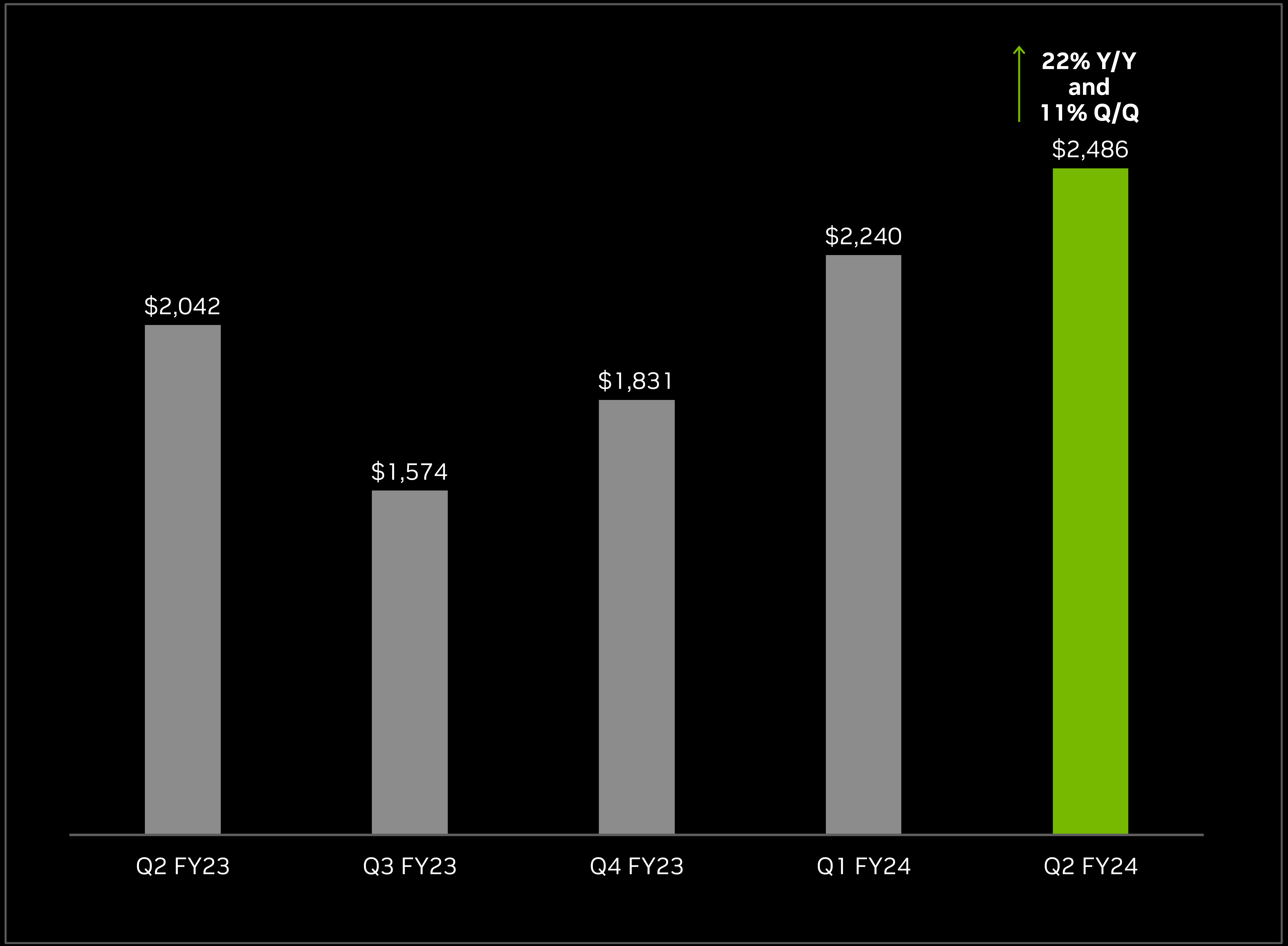


Revenue ($M)

Highlights

- Data Center compute revenue nearly tripled year-on-year, driven primarily by accelerating demand from cloud service providers and large consumer internet companies for our HGX platform
- AWS, Google Cloud, Meta, Microsoft Azure, and Oracle Cloud and a growing number of GPU cloud providers are deploying in volume HGX systems based on our Hopper and Ampere GPU architecture
- Expect supply to increase each quarter through next year
- Data Center networking revenue almost doubled year-on-year, driven by our end-to-end InfiniBand networking platform
- BlueField-3 DPU is in qualification with all major OEMs and ramping across multiple CSPs and consumer internet companies

Gaming

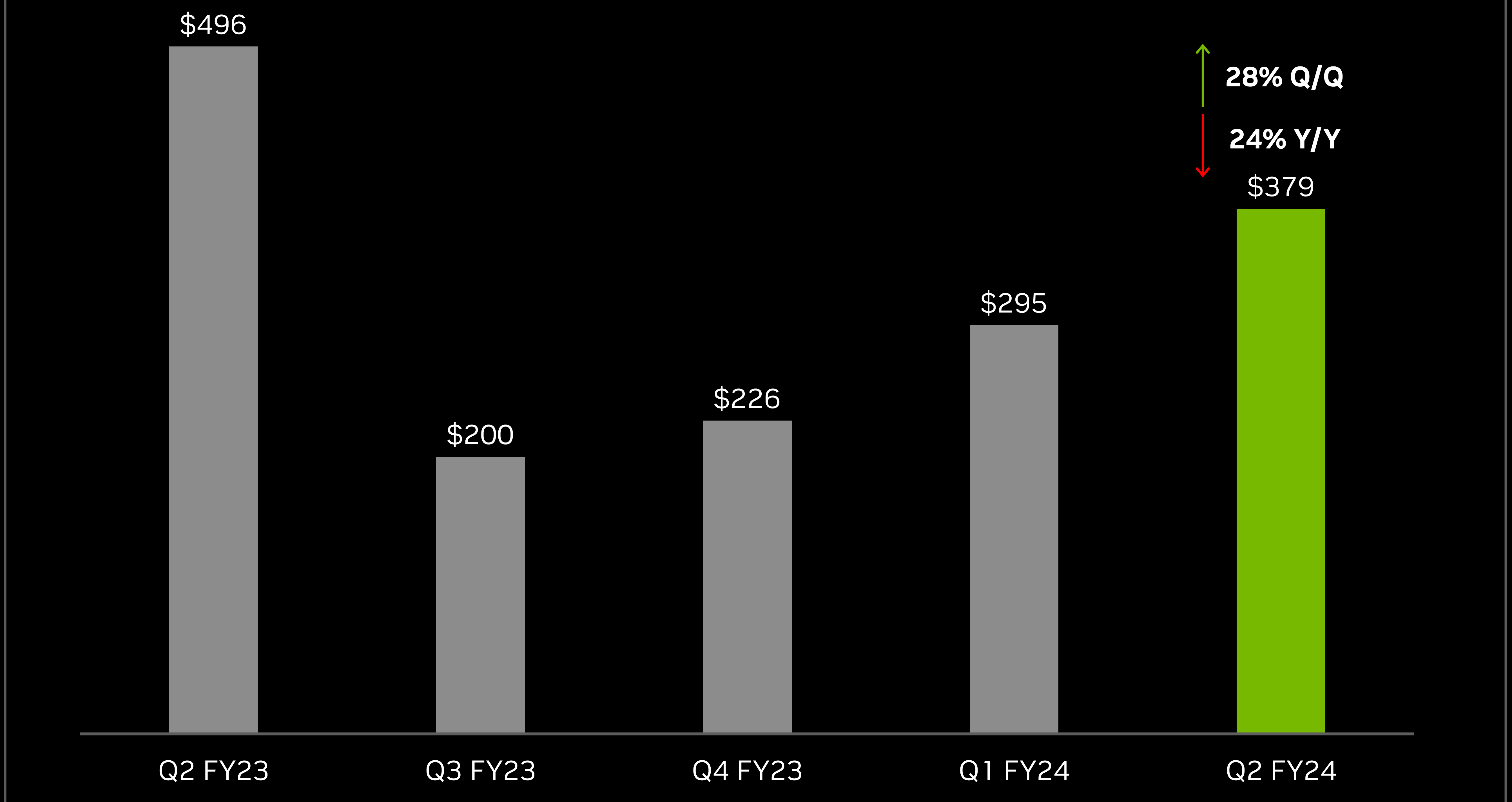


Revenue ($M)

Highlights

- Growth fueled by GeForce RTX 40 series GPUs for laptops and desktops
- Large upgrade opportunity ahead: just 47% of installed base have upgraded to RTX; ~20% have a GPU with RTX 3060 or higher
- NVIDIA GPU-powered laptops have gained in popularity; shipments now outpace desktop GPUs in several regions around the world
  - Likely to shift overall Gaming seasonality a bit, with Q2 and Q3 stronger, reflecting the back-to-school and holiday build schedules
- RTX/DLSS ecosystem continues to expand; 35 new games added DLSS support, including Diablo IV and Baldur’s Gate 3; there are now over 330 RTX accelerated games and apps
- Bringing generative AI to gaming with NVIDIA Avatar Cloud Engine (ACE) for Games

Professional Visualization

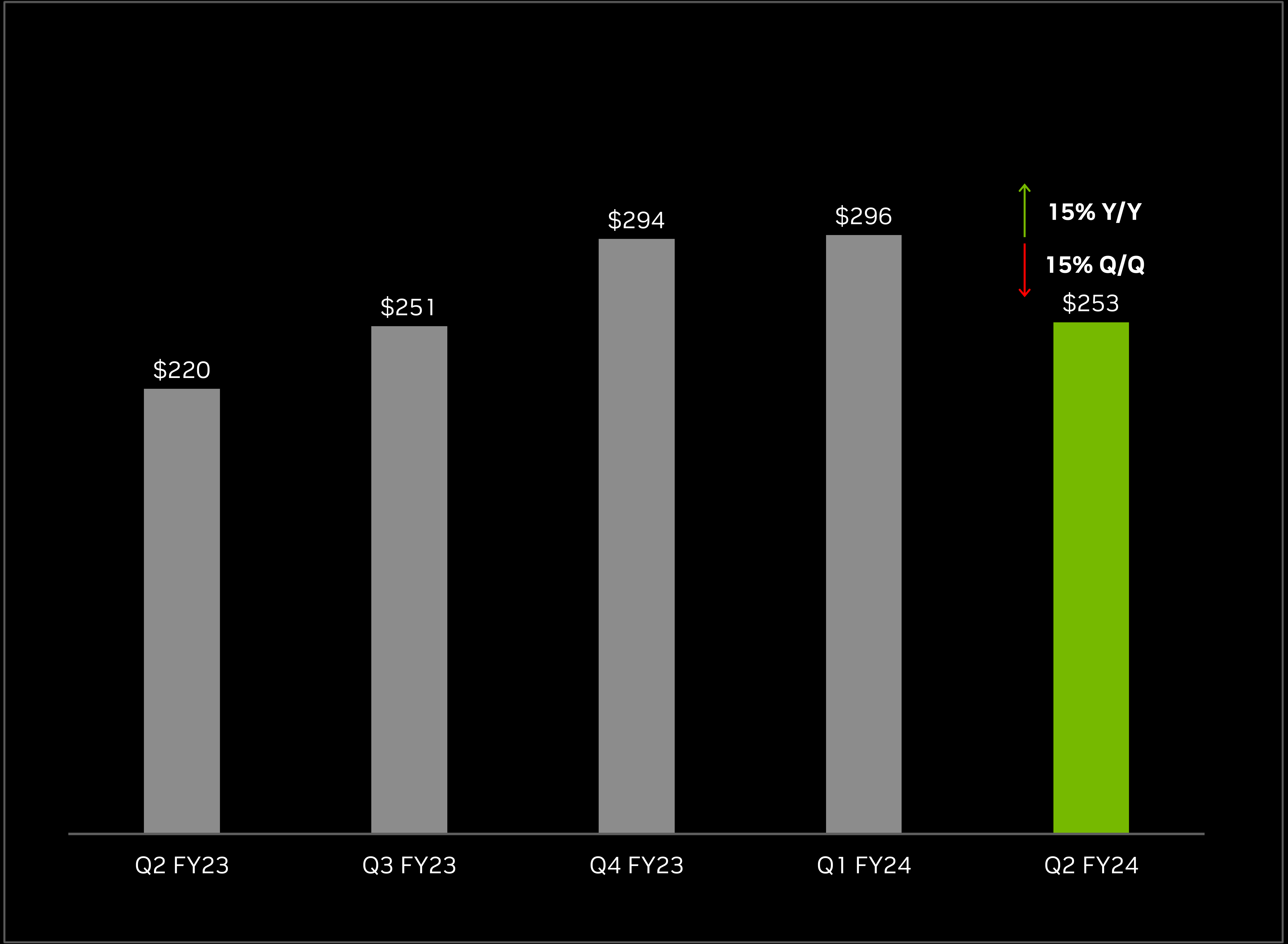


Revenue ($M)

Highlights

- Ada architecture ramp drove strong growth, rolling out initially in laptop workstations
- Desktop workstations refresh coming in Q3
  - Powerful new RTX systems with up to 4 NVIDIA RTX 6000 GPUs, configured with NVIDIA AI Enterprise or NVIDIA Omniverse Enterprise
  - Three new desktop workstation GPUs based on the Ada generation: NVIDIA RTX 5000, 4500, and 4000
- New workloads in gen AI, LLM development and data science are expanding the opportunity in Pro Viz for RTX technology

Automotive

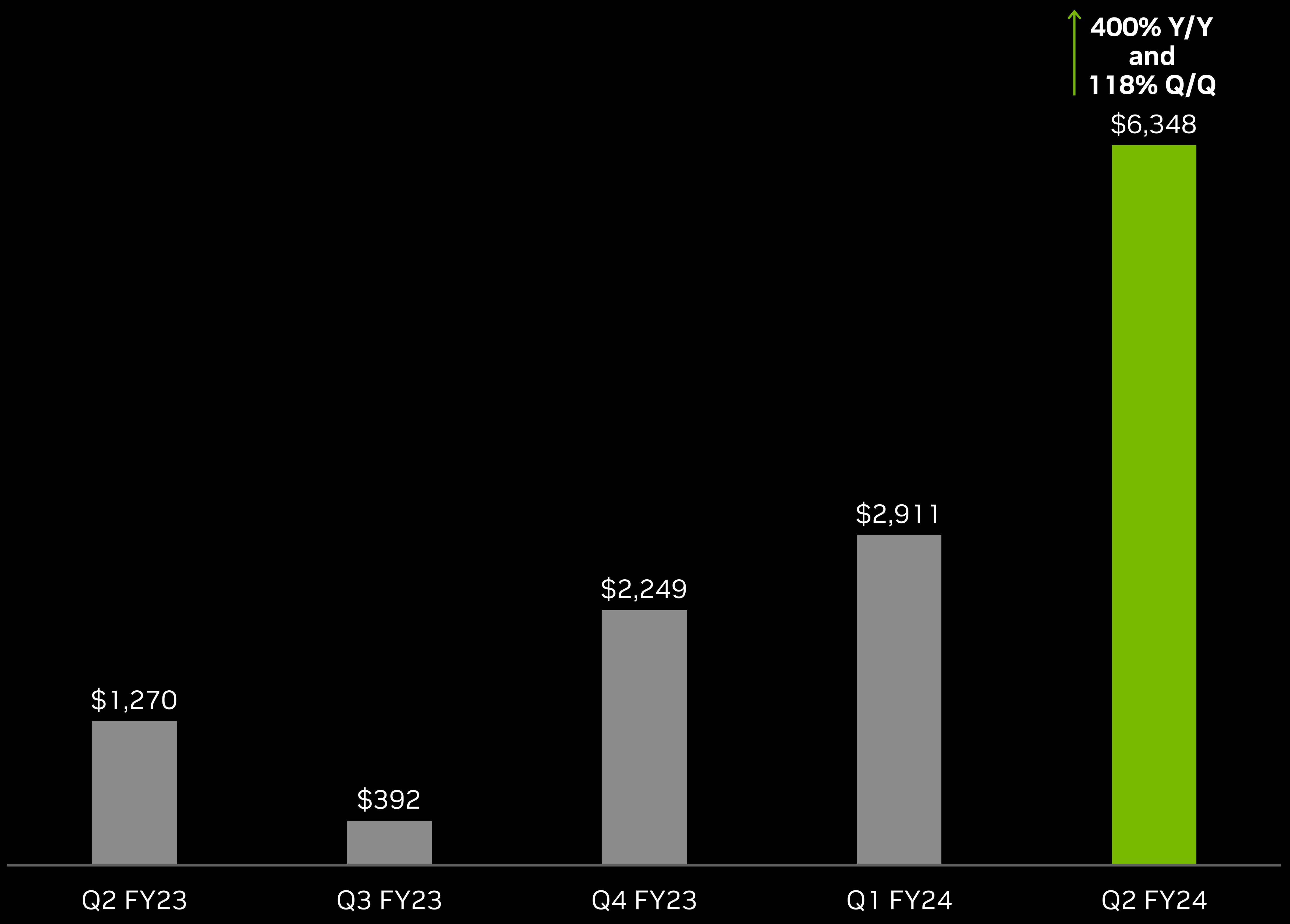


Revenue ($M)

Highlights

- Solid y/y growth driven by ramp of self-driving platforms based on NVIDIA DRIVE Orin SoC with several new energy vehicle makers
- Sequential decline reflects lower overall automotive demand, particularly in China
- Announced that NVIDIA DRIVE Orin is powering the new XPENG G6 Coupe SUV’s intelligent advanced driver assistance system
- Partnered with MediaTek, which will develop automotive SoCs and integrate a new product line of NVIDIA’s GPU chiplets

Sources & Uses of Cash



Cash Flow from Operations ($M)

Highlights

- Y/Y and Q/Q growth were both driven by higher revenue
- Returned $3.4B to shareholders in the form of shares repurchased and cash dividends
- Invested $300M in capex (includes principal payments on PP&E)
- Ended the quarter with $16.0B in gross cash and $9.8B in debt; $6.2B in net cash
- Additional $25B in stock repurchases authorized, adding to $4B that remained as of end of Q2

Gross cash is defined as cash/cash equivalents & marketable securities.
Debt is defined as principal value of debt.
Net cash is defined as gross cash less debt.

Q3 FY24 Outlook

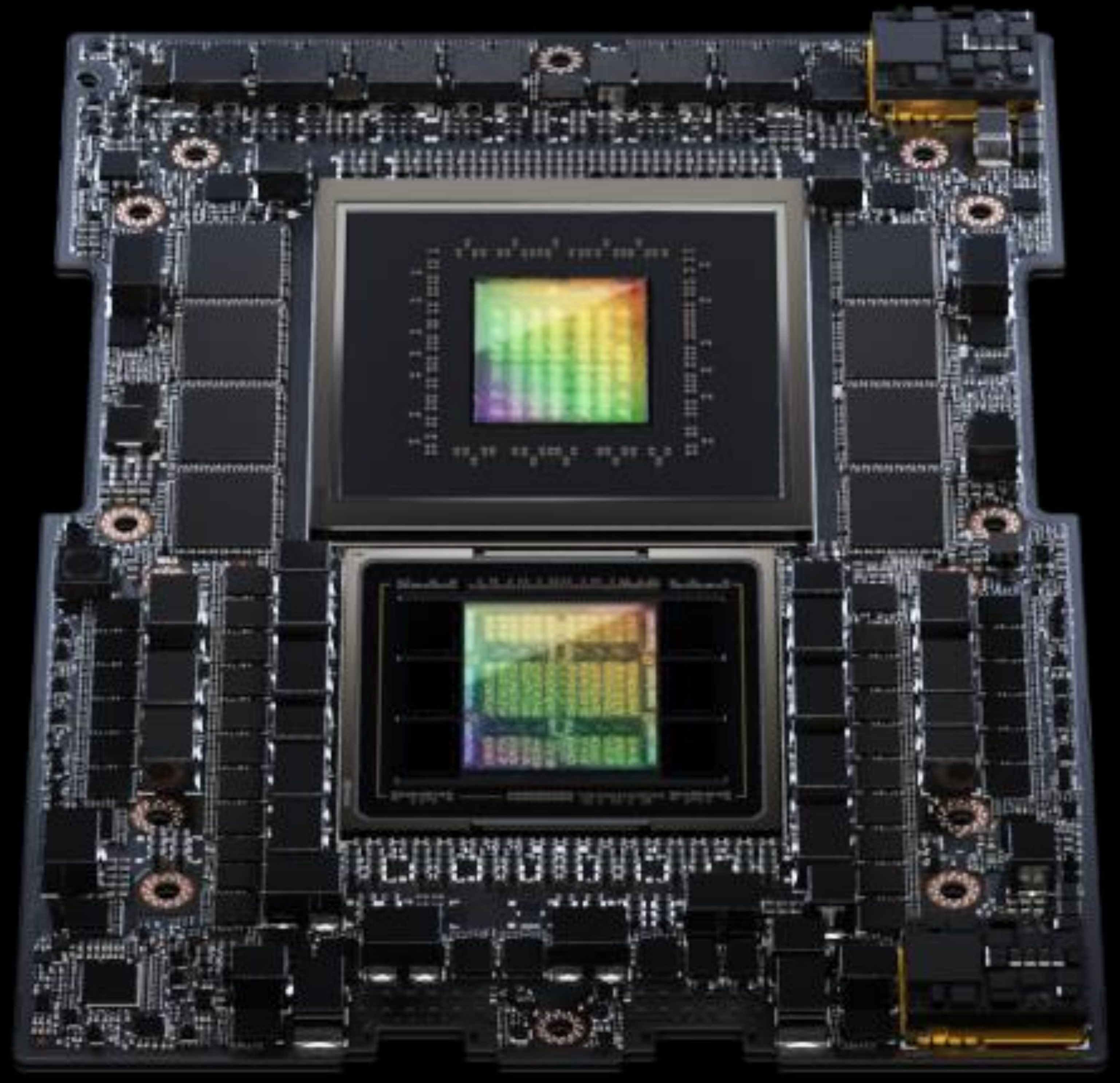
Revenue	**$16.0 billion**, plus or minus 2% Expect sequential growth to be driven largely by Data Center, with Gaming and Pro Viz also contributing
Gross Margins	**71.5%** GAAP and **72.5%** non-GAAP, plus or minus 50 basis points
Operating Expense	Approximately **$2.95 billion** GAAP and **$2.00 billion** non-GAAP
Other Income & Expense	Income of approximately **$100 million** for GAAP and non-GAAP Excluding gains and losses on non-affiliated investments
Tax Rate	**14.5%** GAAP and non-GAAP, plus or minus 1%, excluding discrete items

Refer to Appendix for reconciliation of Non-GAAP measures.

Key Announcements This Quarter

NVIDIA Grace Hopper Superchips Now Available

- GH200 Grace Hopper Superchip combines our first Arm-based Grace CPU with an H100 GPU
- Using NVIDIA NVLink-C2C interconnect at 900GB/s (7X faster than PCIe Gen5), it delivers a CPU+GPU coherent access to over 600GB of memory
- Ideal for giant AI and HPC applications like Gen AI, deep recommenders, and vector databases
- GH200 delivers order-of-magnitude speed-ups compared to x86 CPUs
  - 9X for vector databases
  - 12X for deep recommender inference
  - 284X for LLM (65B parameter) inference
- Available now from leading server manufacturers
- NVIDIA and SoftBank are collaborating on a platform based on GH200 for Gen AI and 5G/6G



DGX GH200 AI Supercomputer Extends the Frontier of LLMs

- DGX GH200 is a new class of large-memory AI supercomputer powered by 256 NVIDIA GH200 Grace Hopper Superchips and the NVIDIA NVLink Switch System, with massive, shared memory space
- 1 exaflop of performance and 144TB of shared memory for giant next-gen Gen AI language models, recommender systems and data analytics workloads
- The NVIDIA NVLink Switch System enables all 256 GPUs in a DGX GH200 to work together as one
- Google Cloud, Meta and Microsoft are among the first to gain access to the DGX GH200
- DGX GH200 supercomputers are expected to be available by the end of calendar 2023

NVIDIA Grace Hopper
Next-Gen Platform with HBM3e

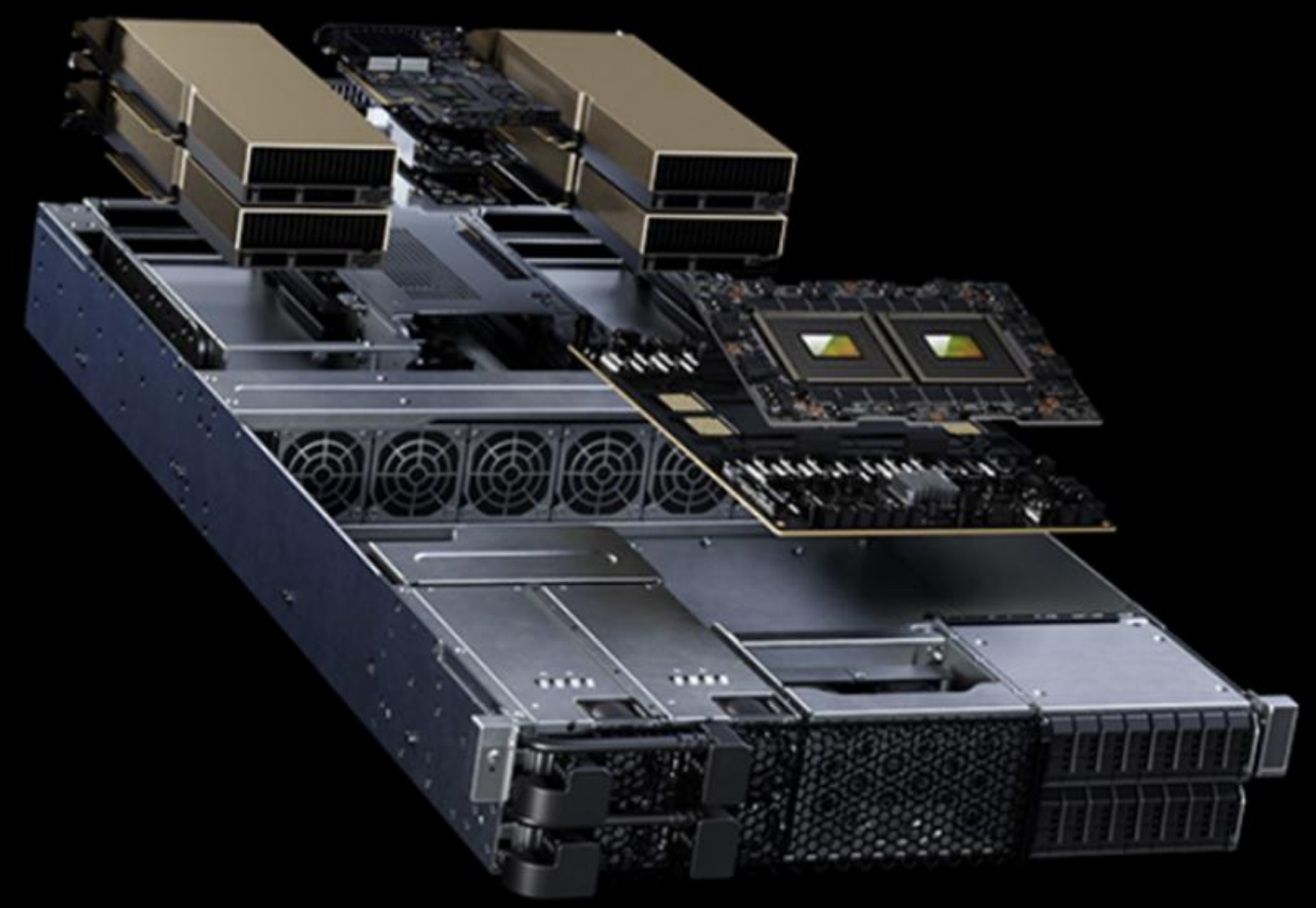
- Announced the next-gen NVIDIA GH200 Grace Hopper platform with HBM3e memory
- Based on two Grace Hopper Superchips connected by NVIDIA NVLink interconnect technology
- The new platform comprises a single server with 144 Arm Neoverse cores, eight petaflops of AI performance and 282 gigabytes of HBM3e memory
- Delivers up to 3.5X more memory capacity and 3X more bandwidth than its predecessor
- Built for the most complex Gen AI workloads, LLMs, recommender systems and vector databases
- Servers from leading manufacturers using the next generation Grace Hopper platform with HBM3e are expected to be available in Q2 of calendar 2024

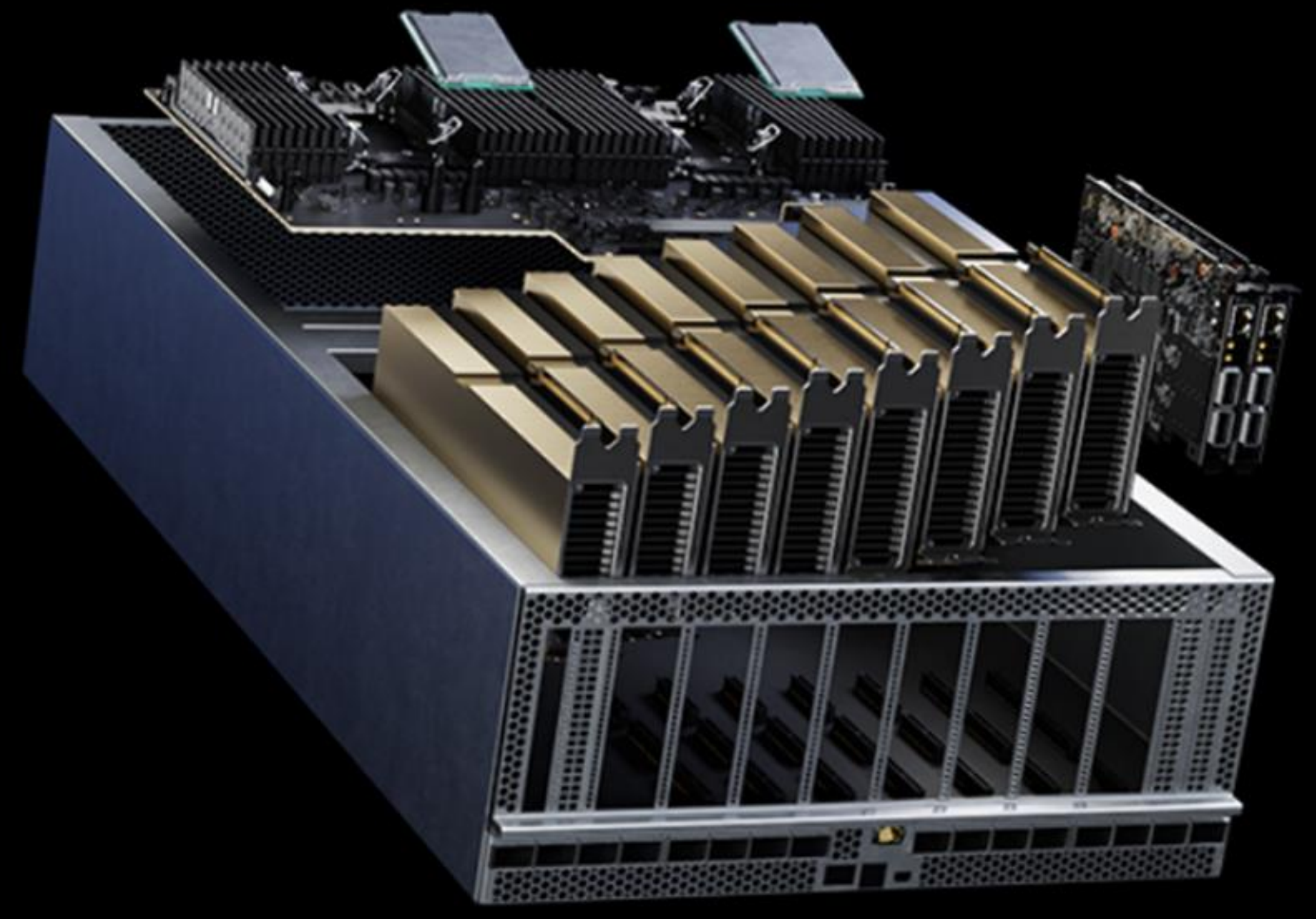
New NVIDIA L40S GPU for Gen AI and Industrial Digitalization

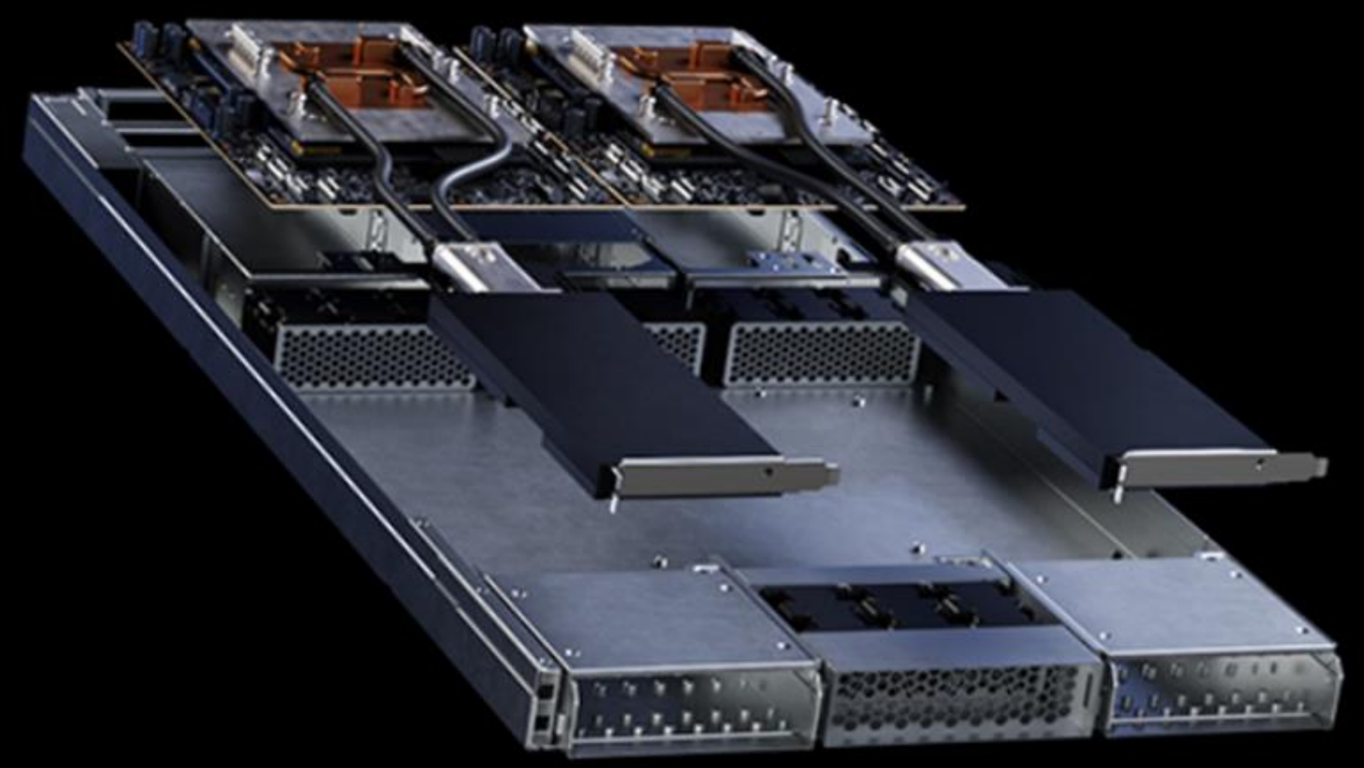
- NVIDIA L40S is a universal data center processor designed to accelerate the most compute-intensive applications, including
  - AI training and inference
  - 3D design and visualization
  - video processing
  - industrial digitalization
- Available from leading server makers in a broad range of platforms, including NVIDIA OVX and NVIDIA AI-ready servers with NVIDIA BlueField DPUs, beginning this quarter

MGX Server Reference Design for Accelerated Computing

- NVIDIA MGX server reference design is a scalable and open accelerated computing server architecture
- Helps system manufacturers quickly and cost-effectively build more than 100 server variations to address diverse AI, HPC and Omniverse applications
- Slashes development costs by up to three-quarters and reduces development time by two-thirds to just six months
- Extends NVIDIA accelerated computing into virtually every server segment of the $1T installed base of data center infrastructure
- Will be adopted by ASRock Rack, ASUS, GIGABYTE, Pegatron, QCT and Supermicro





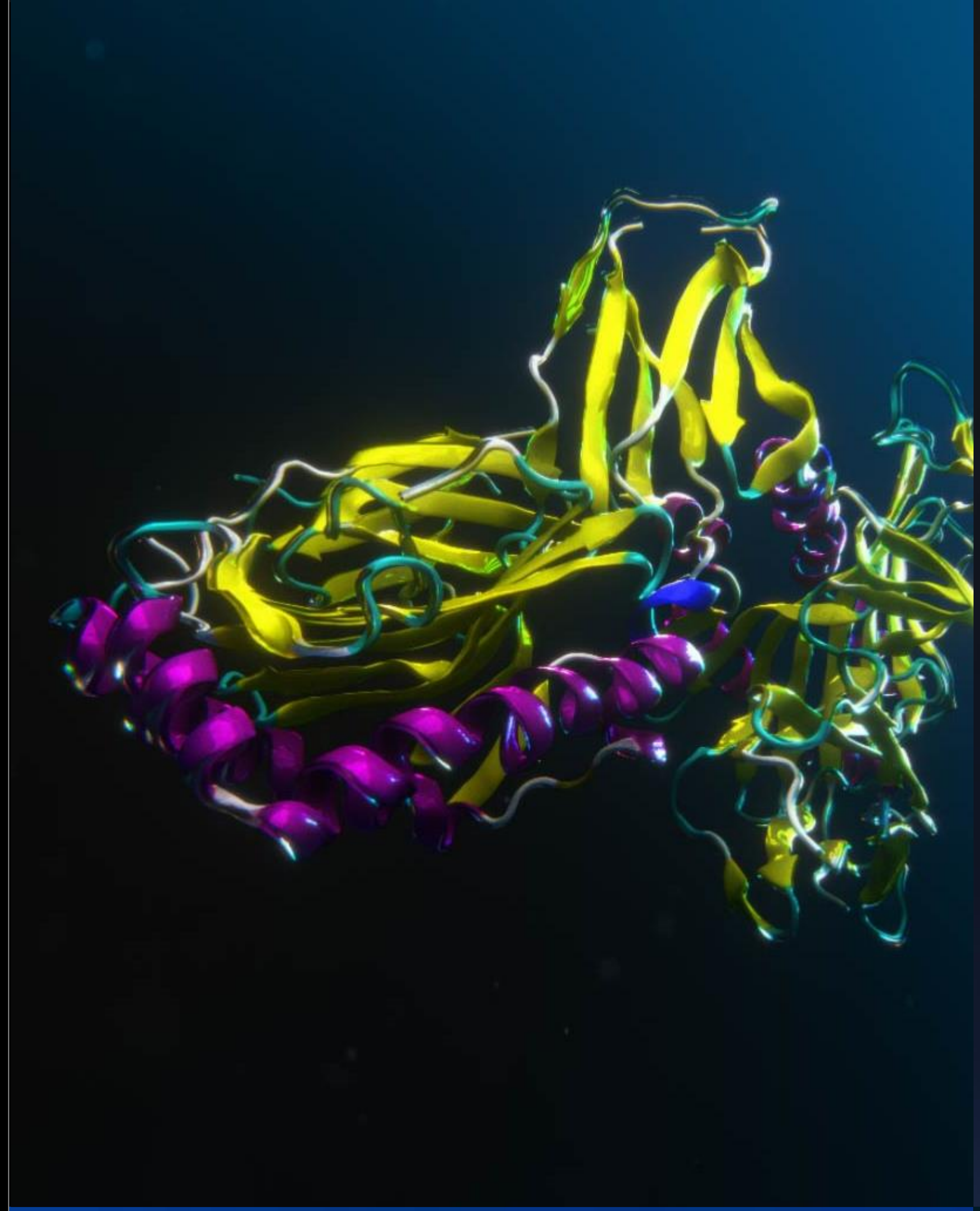


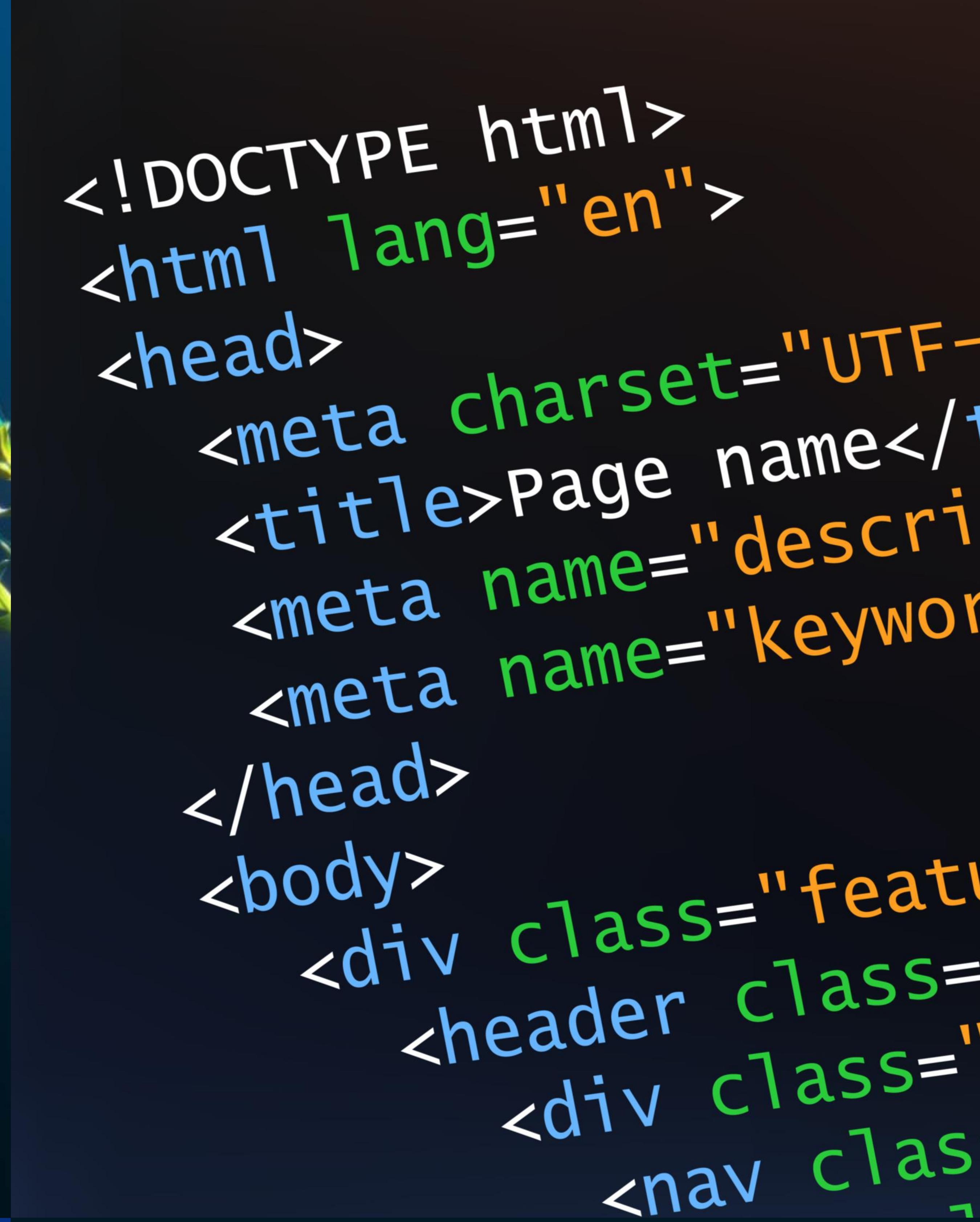
Spectrum-X End-to-End Ethernet Platform

- NVIDIA Spectrum-X is an end-to-end networking platform designed to optimize the performance and efficiency of Ethernet-based AI clouds
- Combines NVIDIA Spectrum-4 switch, with NVIDIA BlueField-3 DPU, NVIDIA LinkX, and NVIDIA software
  - The world’s first Ethernet switch platform for AI
  - NVIDIA DOCA for BlueField DPU programmability
- Spectrum-X achieves 1.7X better AI performance and power efficiency vs. traditional Ethernet
- Shipping this quarter

NVIDIA AI Announcements to Help Accelerate the Adoption of Custom Gen AI by Enterprises

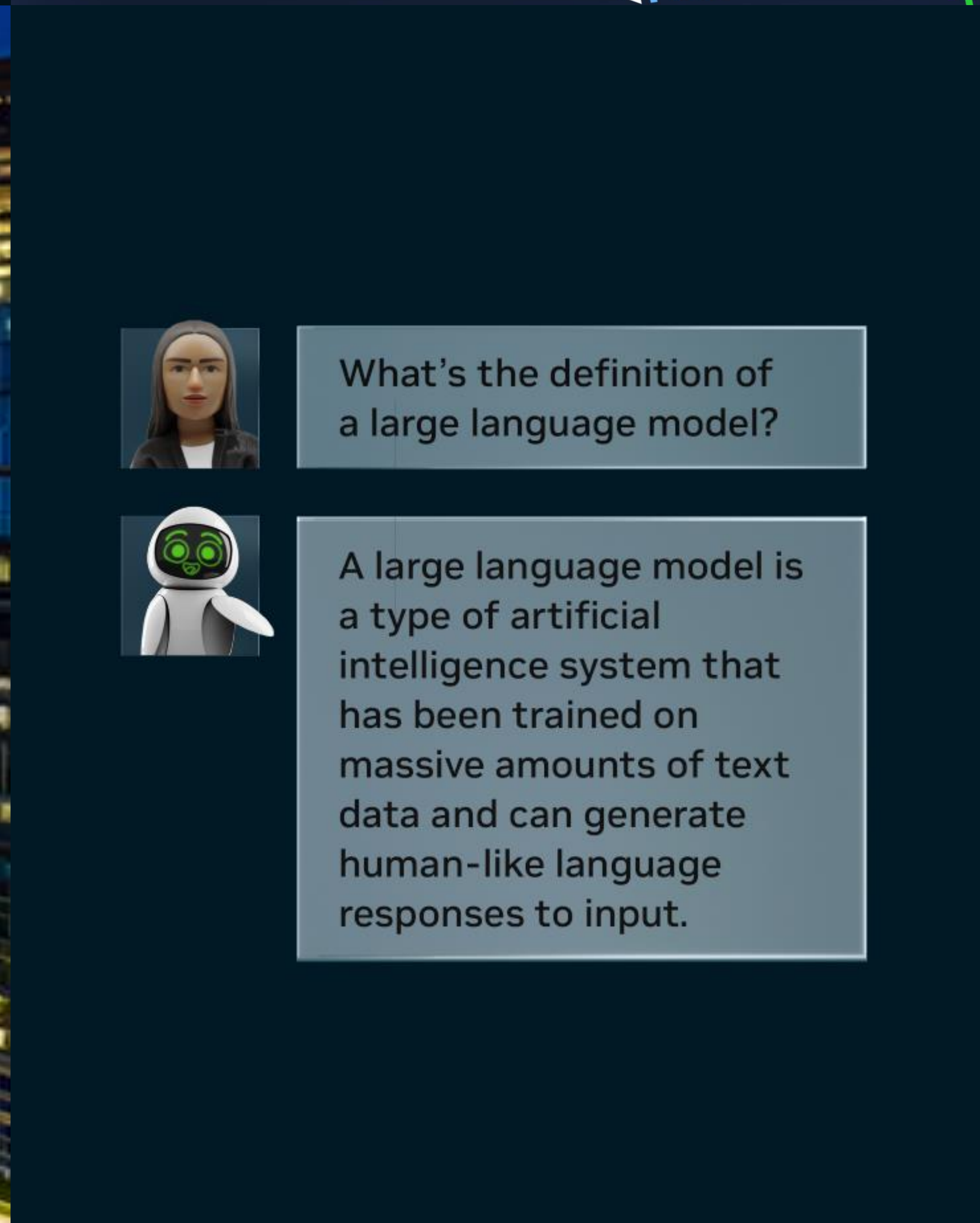
- NVIDIA AI Workbench – a unified, easy-to-use workspace that allows developers to quickly create, test and customize pretrained Gen AI models on a PC or workstation - then scale them to virtually any data center, public cloud or NVIDIA DGX Cloud
- NVIDIA AI Enterprise 4.0 – the latest version of our enterprise software platform now includes support for NVIDIA NeMo, our cloud-native framework for creating and customizing LLM applications, providing a foundation for production-ready Gen AI for customers
- NVIDIA DGX Cloud Integration in Hugging Face – partnership to give developers one-click access to NVIDIA DGX Cloud AI supercomputing within the Hugging Face platform to train and tune advanced AI models





```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-
  <title>Page name</
  <meta name="descri
  <meta name="keywor
</head>
<body>
  <div class="featu
    <header class=
      <div class="
        <nav clas
```

A bustling metropolis at sunset with intricately designed buildings and a water element.



What's the definition of a large language model?

A large language model is a type of artificial intelligence system that has been trained on massive amounts of text data and can generate human-like language responses to input.

VMware and NVIDIA Unlock Generative AI for Enterprises

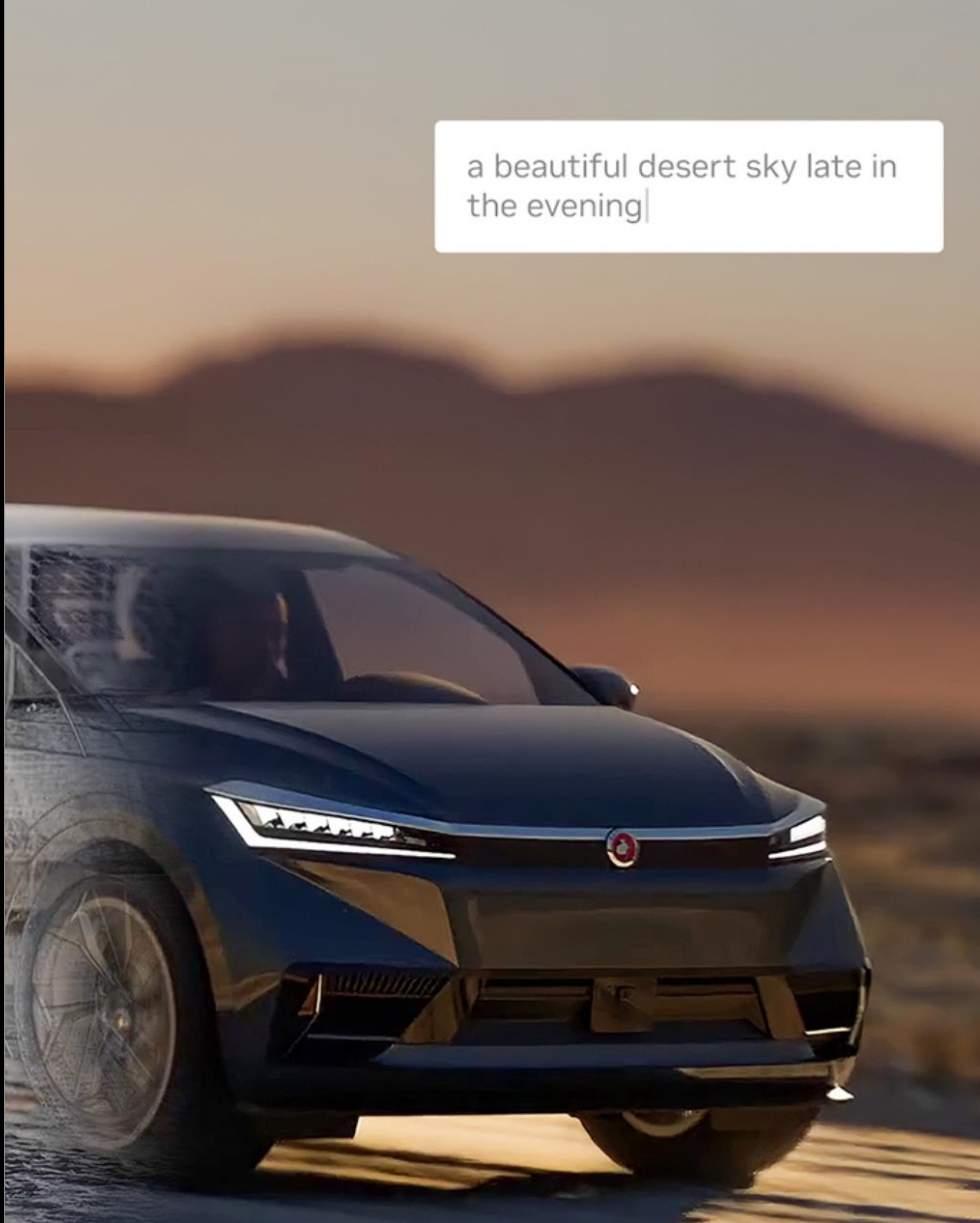
- Announced a major new enterprise offering called *VMware Private AI Foundation with NVIDIA*
- *VMware Private AI Foundation* is a fully integrated platform featuring NVIDIA AI software and accelerated computing with multi-cloud software for enterprises running VMware
- Gives enterprises access to infrastructure, AI and cloud management software needed to customize models and run Gen AI apps such as intelligent chatbots, assistants, search and summarization
- NVIDIA AI Enterprise-ready servers are fully optimized for *VMware Cloud Foundation and Private AI Foundation*
    - Nearly 100 configurations will soon be available from Dell, HPE, and Lenovo

Snowflake and NVIDIA Team to Help Businesses Harness Their Data for Gen AI in the Data Cloud

- Partnership with Snowflake to provide businesses with an accelerated path to create customized Gen AI applications using their own proprietary data, all securely within the Snowflake Data Cloud
- With the NVIDIA NeMo platform for developing LLMs and NVIDIA GPU-accelerated computing, Snowflake will enable enterprises to make custom LLMs for advanced Gen AI services, including chatbots, search and summarization

WPP Partners With NVIDIA to Build Gen AI-Enabled Content Engine for Digital Advertising

- NVIDIA and WPP are developing a content engine that harnesses NVIDIA Omniverse and AI to enable creative teams to produce high-quality commercial content faster, more efficiently and at scale
- The engine connects an ecosystem of 3D design, manufacturing and creative supply chain tools
- This includes Adobe Firefly, a family of creative Gen AI models, and exclusive visual content from Getty Images created using NVIDIA Picasso, a foundry for custom Gen AI models for visual design
- Soon available exclusively to WPP’s clients around the world



New NVIDIA RTX Ada GPUs to Power Workstations for Gen AI and LLM, Content Creation, Data Science

- Global systems builders, including BOXX, Dell Technologies, HP and Lenovo, announced powerful new workstations based on NVIDIA RTX 6000
  - Up to four GPUs in a single desktop workstation
  - Up to 5,828 TFLOPS of AI performance and 192GB of GPU memory
  - Can be configured with NVIDIA AI Enterprise or Omniverse Enterprise
- Three new desktop workstation Ada GPUs — NVIDIA RTX 5000, 4500 and 4000
  - Up to 2X the RT core throughput and up to 2X faster AI training performance versus prior gen
- NVIDIA RTX 5000 GPU available now; NVIDIA RTX 6000, 4500 and 4000 GPUs available in the fall

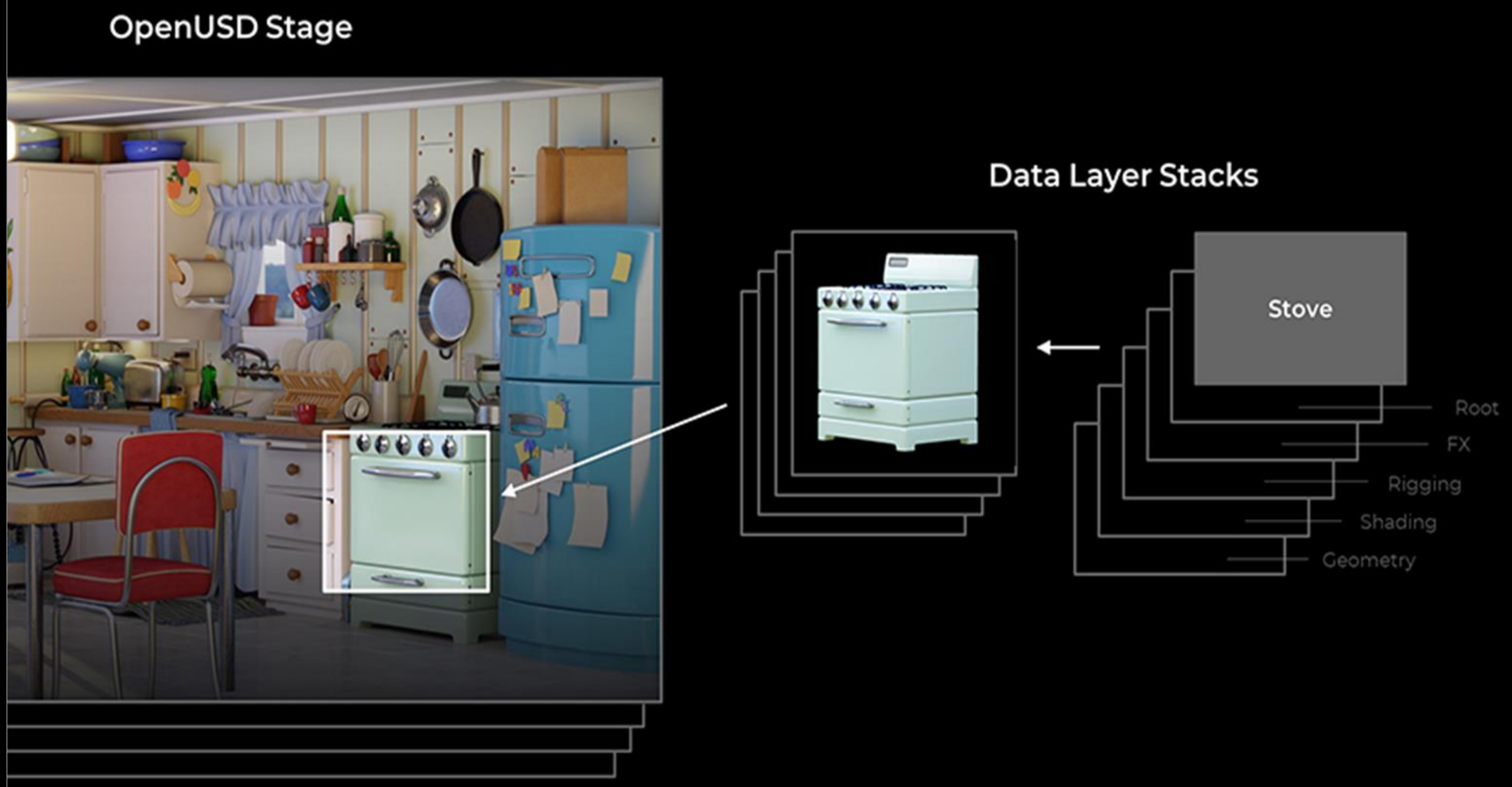
AI Art Generator
AI Image Generator
TEXT TO IMAGE
<prompt> a frog in the foreground, in the rain forest, side view, high contrast, vibrant colors, 4k resolution, photo-realistic
IMAGE TO IMAGE
Upload your image, select a style, and transform.
Drag and drop image or Browse
Select style
IMAGE DIMENSION
Width x Height of the image
512
512
Generate

NVIDIA, Pixar, Adobe, Apple and Autodesk Co-Founded the Alliance for OpenUSD (AOUSD)

- AOUSD was formed to promote the standardization, development, evolution, and growth of Pixar's Universal Scene Description technology
- Standardize the 3D ecosystem by advancing the capabilities of Open Universal Scene Description (OpenUSD)
- Enable developers and content creators to describe, compose, and simulate large-scale 3D projects and build 3D-enabled products and services with greater interoperability of 3D tools and data
- NVIDIA Omniverse is built on OpenUSD

New NVIDIA Generative AI Omniverse Cloud APIs

- Enable developers to more seamlessly implement and deploy OpenUSD pipelines and applications
  - ChatUSD – LLM copilot for USD developers that can answer USD knowledge questions or generate Python-USD code scripts
  - RunUSD – translates OpenUSD files into fully path-traced rendered images and generates renders with Omniverse Cloud
  - DeepSearch – LLM agent enabling fast semantic search through massive databases of untagged assets
  - USD-GDN Publisher – one-click service for publishing and streaming high-fidelity, OpenUSD-based experiences through NVIDIA's Graphics Delivery Network



MediaTek Partners With NVIDIA to Transform Automobiles With AI and Accelerated Computing

- MediaTek will develop mainstream automotive SoCs for global OEMs that integrate new NVIDIA GPU chiplet IP for AI and graphics
- The combination of MediaTek's SoC with NVIDIA's GPU and AI software will enable new experiences, enhanced safety and new connected services
- The chiplets are connected by an ultra-fast and coherent chiplet interconnect technology
- MediaTek will run the NVIDIA DRIVE OS, DRIVE IX, CUDA and TensorRT software on these new SoCs
- The partnership covers a wide range of vehicle segments, from luxury to entry-level

NVIDIA Overview

NVIDIA pioneered accelerated computing to help solve impactful challenges classical computers cannot. A quarter of a century in the making, NVIDIA accelerated computing is broadly recognized as the way to advance computing as Moore's law ends and AI lifts off.

NVIDIA's platform is installed in several hundred million computers, is available in every cloud and from every server maker, powers 74% of the TOP500 supercomputers, and boasts over 4 million developers.

Headquarters: Santa Clara, CA

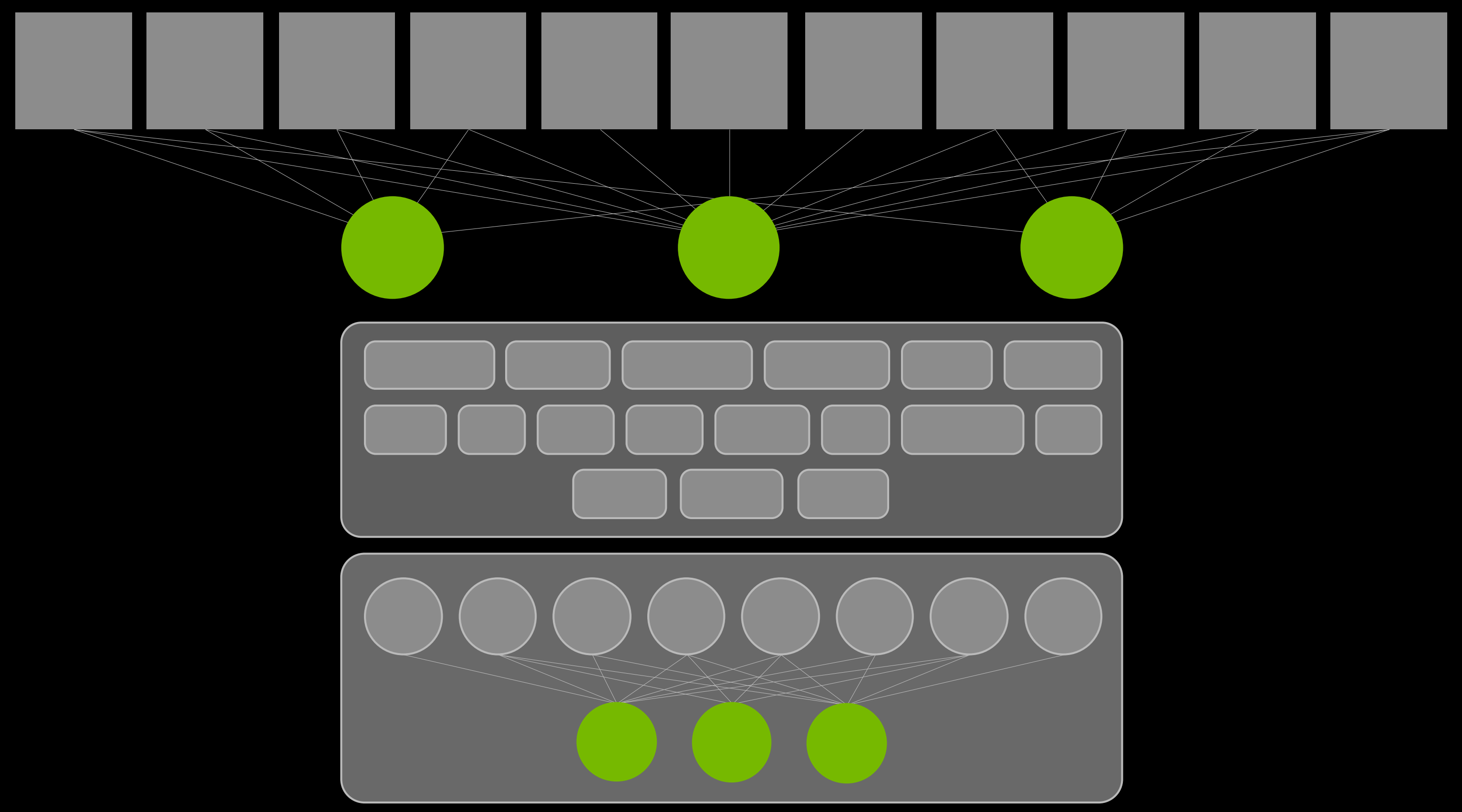
What Is Accelerated Computing?

A full-stack approach: silicon, systems, software

Not just a superfast chip – accelerated computing is a full-stack combination of:

- Chip(s) with specialized processors
- Algorithms in acceleration libraries
- Domain experts to refactor applications

To speed-up compute-intensive parts of an application.



Amdahl's law:

The overall system speed-up (S) gained by optimizing a single part of a system by a factor (s) is limited by the proportion of execution time of that part (p).

$$S = \frac{1}{(1-p) + \frac{p}{s}}$$

For example:

- If 90% of the runtime can be accelerated by 100X, the application is sped up 9X
- If 99% of the runtime can be accelerated by 100X, the application is sped up 50X
- If 80% of the runtime can be accelerated by 500X, or even 1000X, the application is sped up 5X

Why Accelerated Computing?

Advancing computing in the post-Moore's Law era

Accelerated computing is needed to tackle the most impactful opportunities of our time—like AI, climate simulation, drug discovery, ray tracing, and robotics.

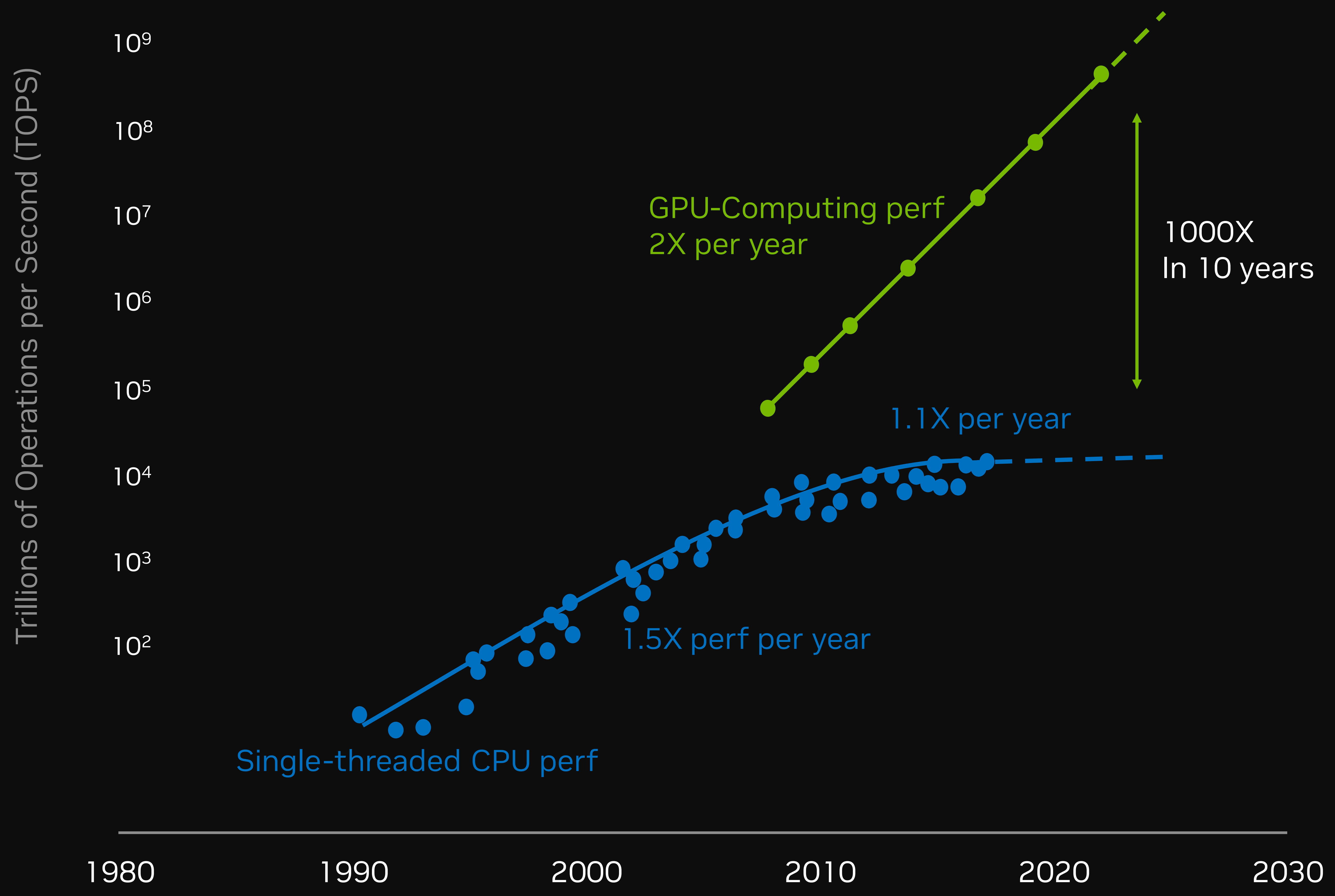
NVIDIA is uniquely dedicated to accelerated computing—working top-to-bottom—refactoring applications and creating new algorithms, and bottom-to-top—inventing new specialized processors, like RT Core and Tensor Core.

"It's the end of Moore's Law as we know it."

- John Hennessy Oct 23, 2018

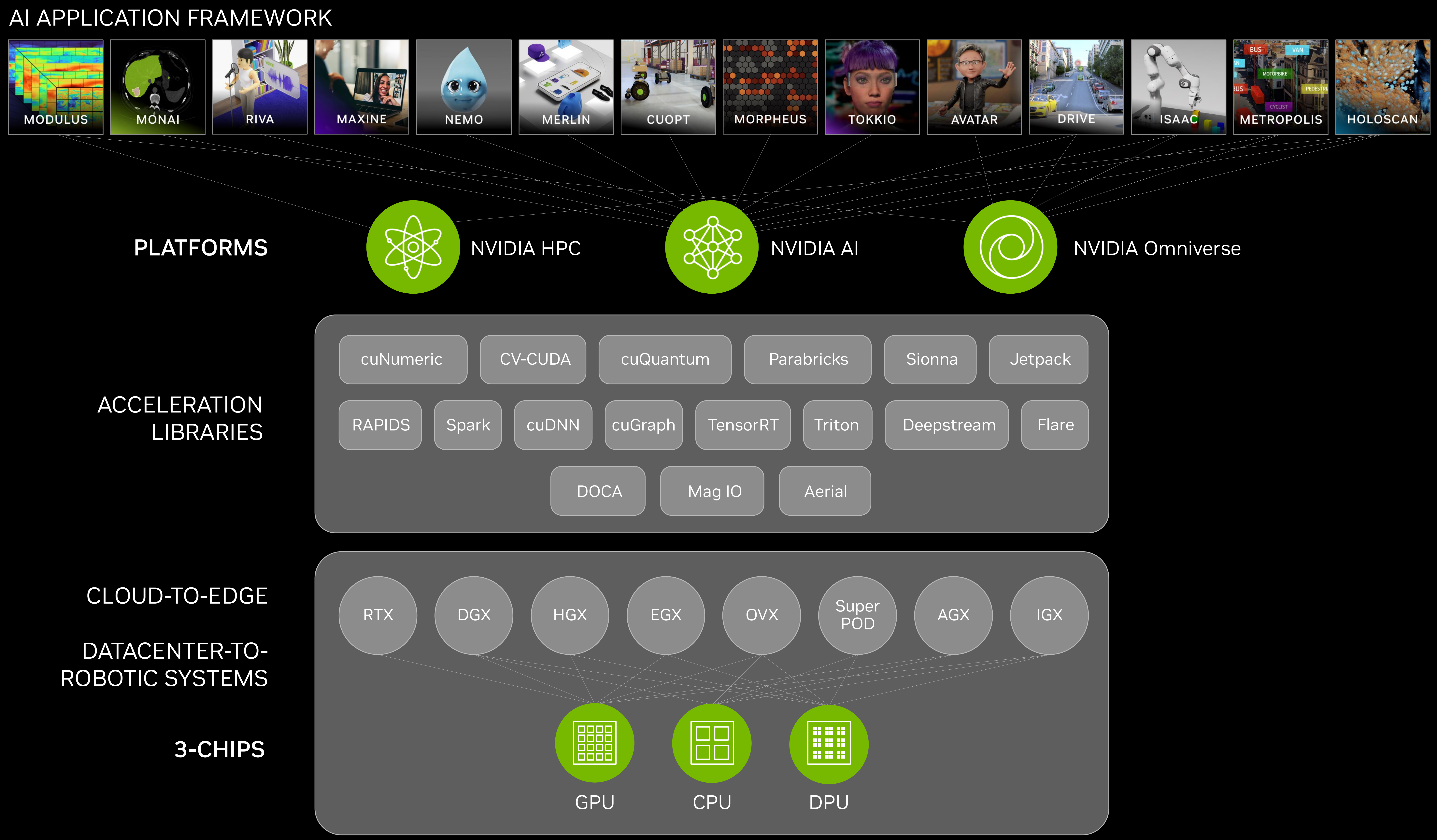
"Moore's Law is dead."

- Jensen Huang, GTC 2013



NVIDIA’s Accelerated Computing Platform

Full-stack innovation across silicon, systems and software



AI APPLICATION FRAMEWORK

MODULUS MONAI RIVA MAXINE NEMO MERLIN CUOPT MORPHEUS TOKKIO AVATAR DRIVE ISAAC METROPOLIS HOLOSCAN

PLATFORMS

NVIDIA HPC NVIDIA AI NVIDIA Omniverse

ACCELERATION LIBRARIES

cuNumeric CV-CUDA cuQuantum Parabricks Sionna Jetpack

RAPIDS Spark cuDNN cuGraph TensorRT Triton Deepstream Flare

DOCA Mag IO Aerial

CLOUD-TO-EDGE

DATACENTER-TO-ROBOTIC SYSTEMS

RTX DGX HGX EGX OVX Super POD AGX IGX

3-CHIPS

GPU CPU DPU

With nearly three decades of a singular focus, NVIDIA is expert at accelerating software and scaling compute by a Million-X, going well beyond Moore’s law.

Accelerated computing is a full-stack challenge, demanding deep understanding of the problem domain, optimizing across every layer of computing, and all three chips —GPU, CPU, and DPU.

Scaling across multi-GPUs and multi-nodes is a data center-scale challenge and requires treating the network and storage as part of the computing fabric.

Our platform extends from PCs to supercomputing centers, enterprise data centers, cloud and edge environments.

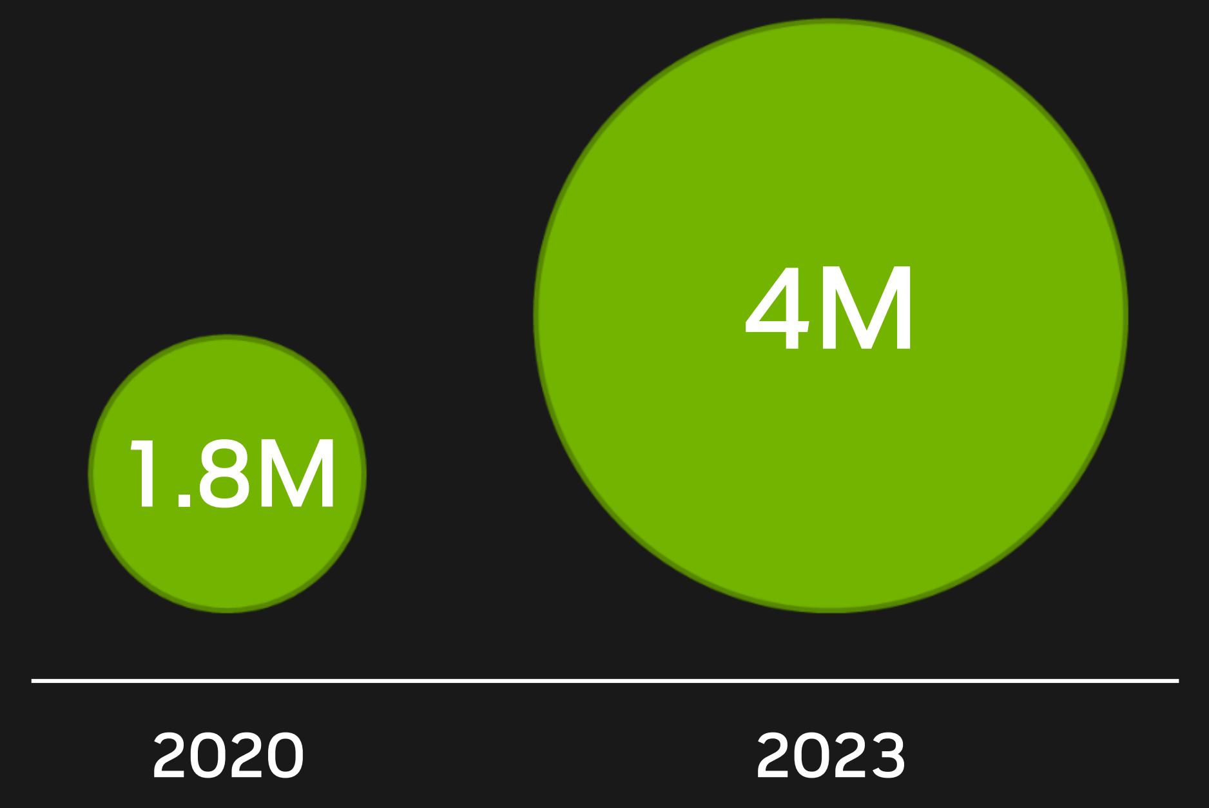
NVIDIA's Expanding Accelerated Computing Ecosystem

300 Libraries

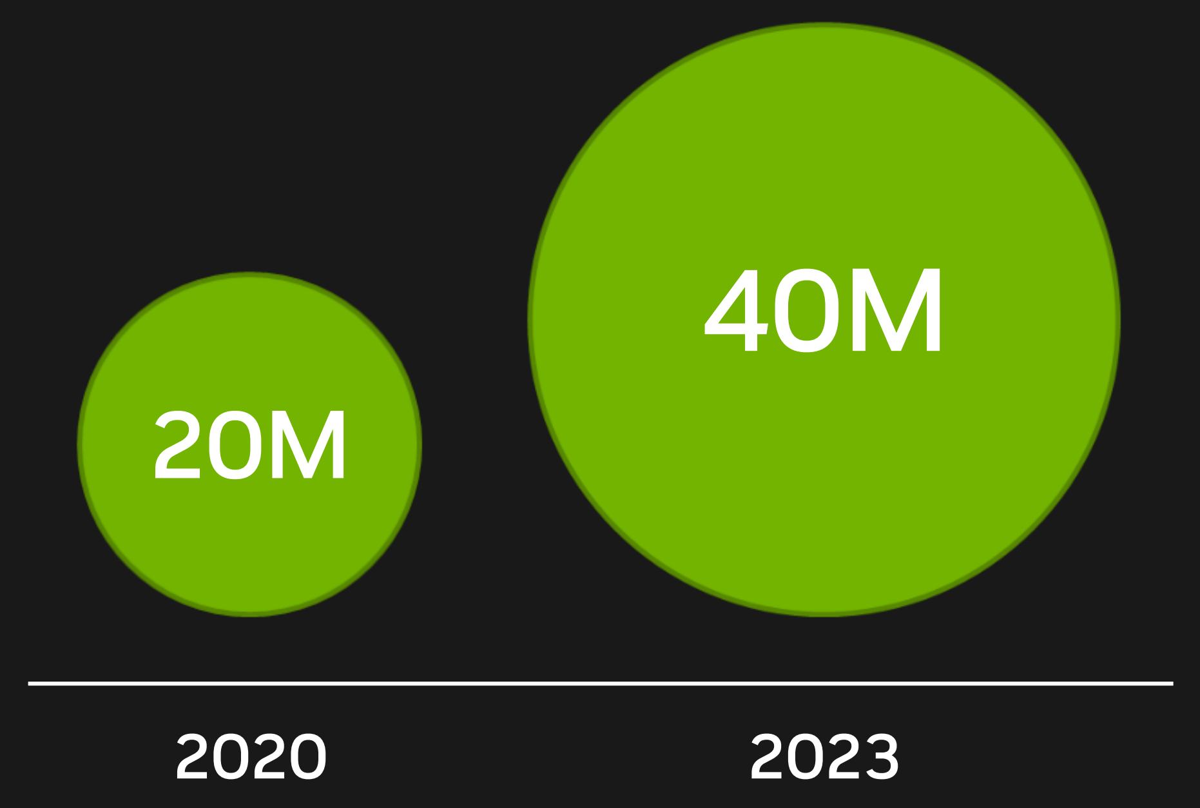
400 AI Models

100 Updated in the Last Year

Developers

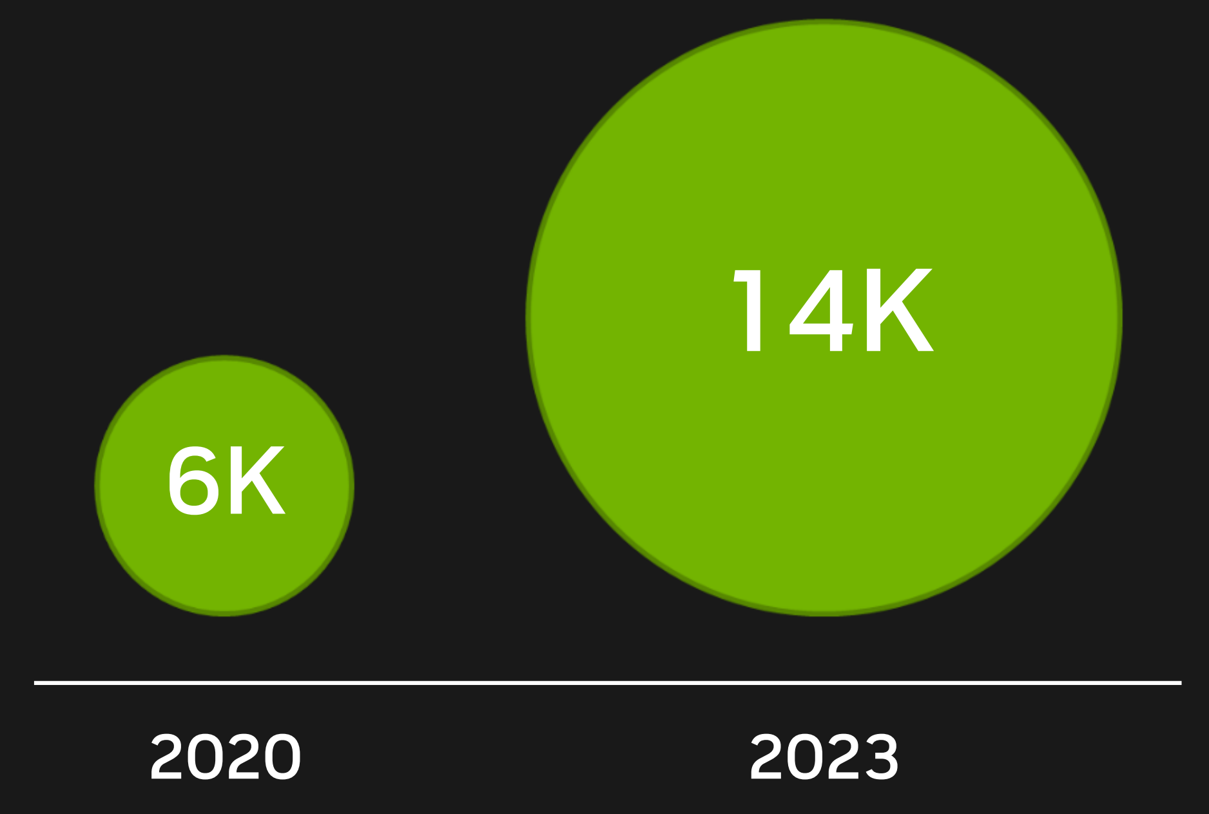


CUDA Downloads*

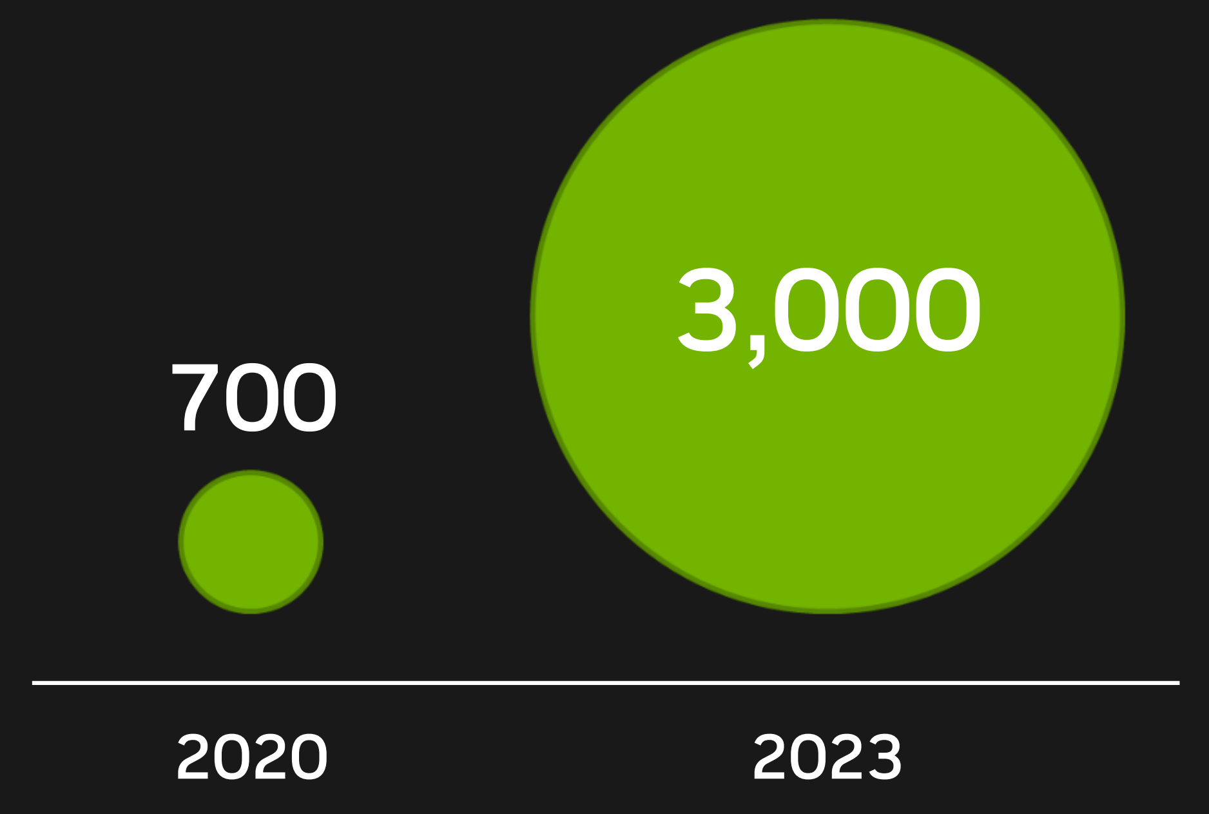


*Cumulative

AI Startups



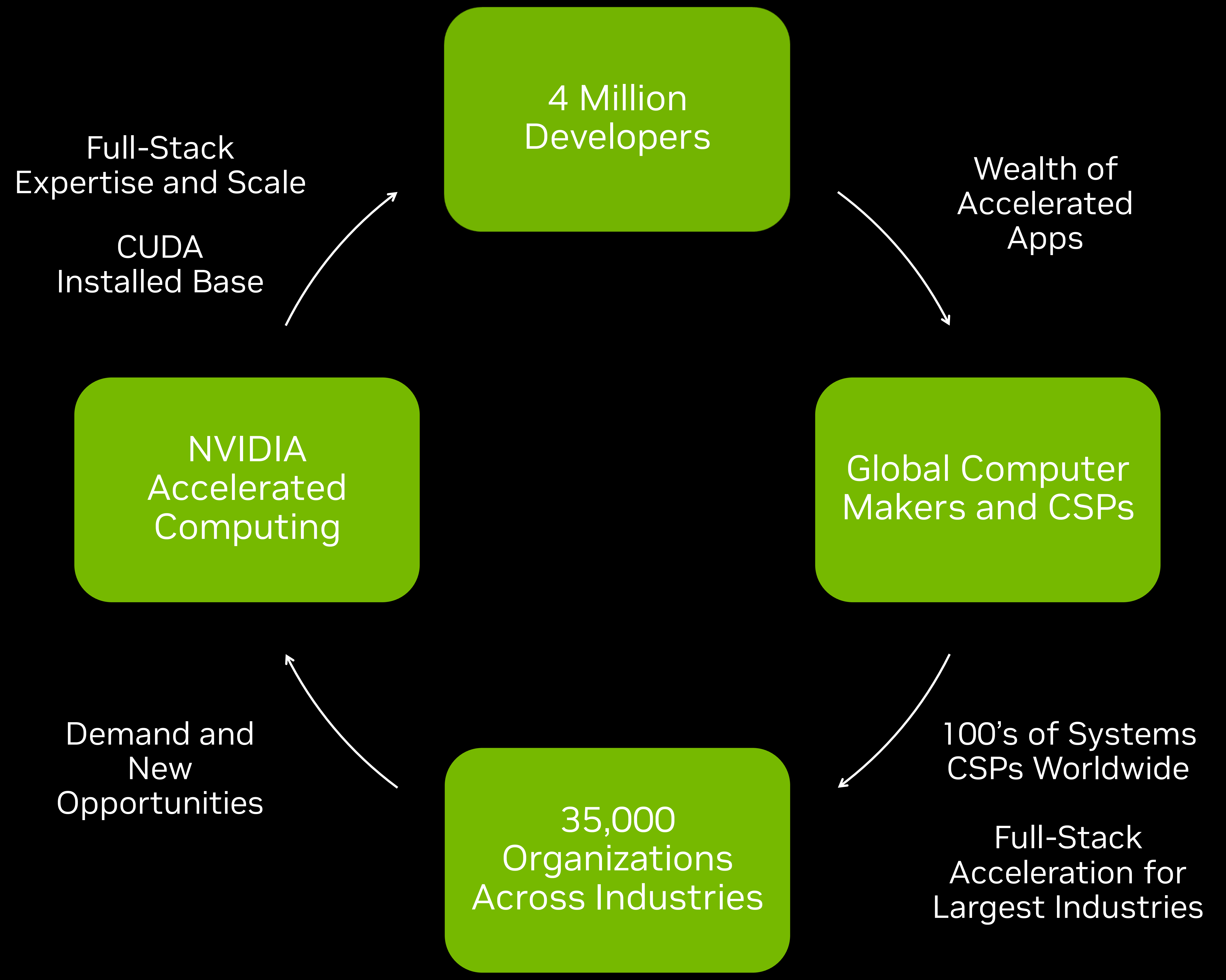
GPU-Accelerated Applications



NVIDIA's Multi-Sided Platform and Flywheel

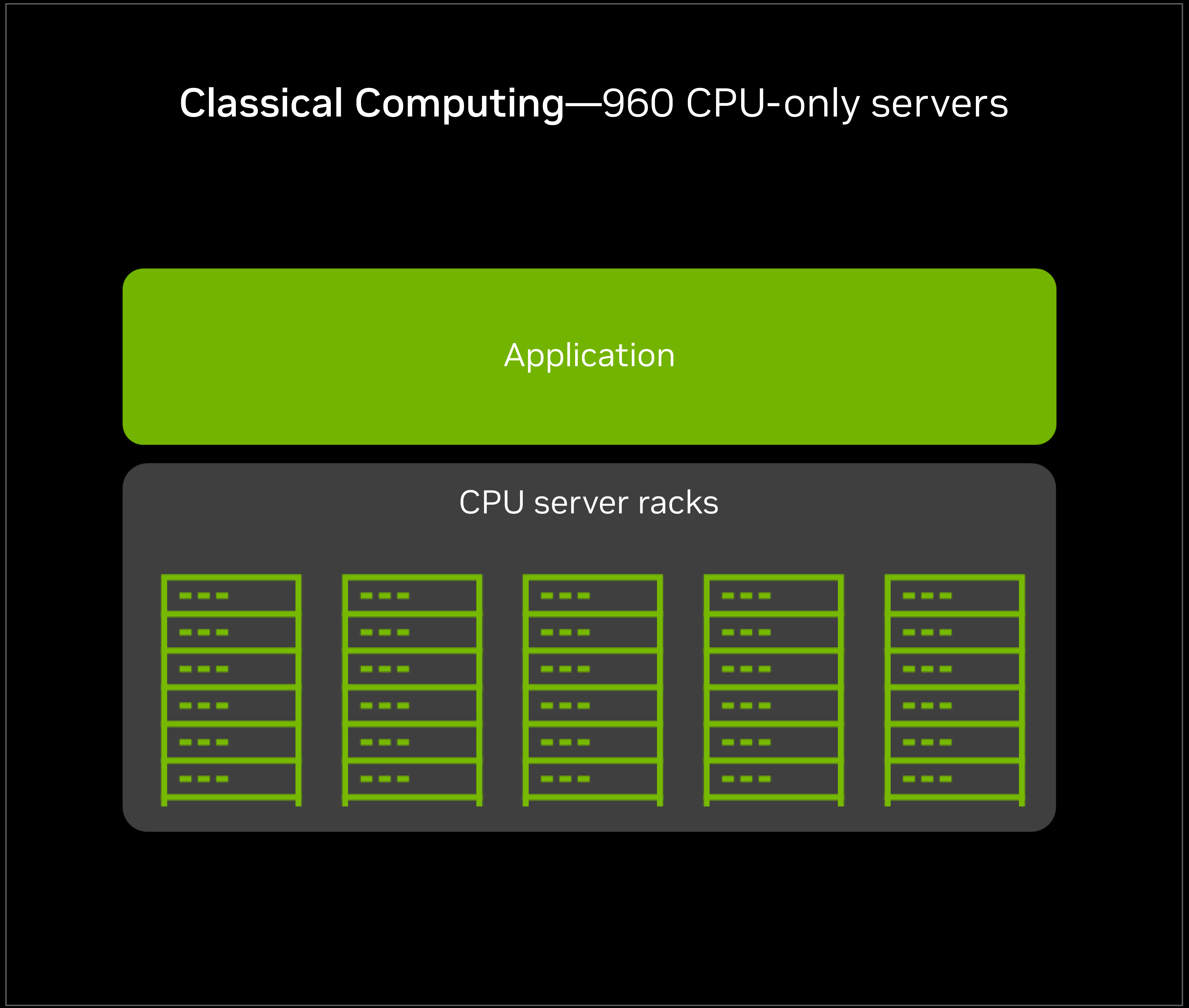
NVIDIA is valued by every stakeholder in the ecosystem:

- **For developers** – NVIDIA's one Architecture and large installed base give developer's software the best performance and greatest reach
- **For computer makers and CSPs** – NVIDIA's rich suite of Acceleration Platforms lets partners build one offering to address large markets including media & entertainment, healthcare, transportation, energy, financial services, manufacturing, retail, and more
- **For customers** – NVIDIA is offered by virtually every computing provider and accelerates the most impactful applications from cloud to edge
- **For NVIDIA** – Deep engagement with developers, computing providers, and customers in diverse industries enables unmatched expertise, scale, and speed of innovation across the entire accelerated computing stack – propelling the flywheel


4 Million Developers
Wealth of Accelerated Apps
Global Computer Makers and CSPs
100's of Systems CSPs Worldwide
Full-Stack Acceleration for Largest Industries
35,000 Organizations Across Industries
Demand and New Opportunities
NVIDIA Accelerated Computing
Full-Stack Expertise and Scale
CUDA Installed Base

Full-Stack & Data Center Scale Acceleration

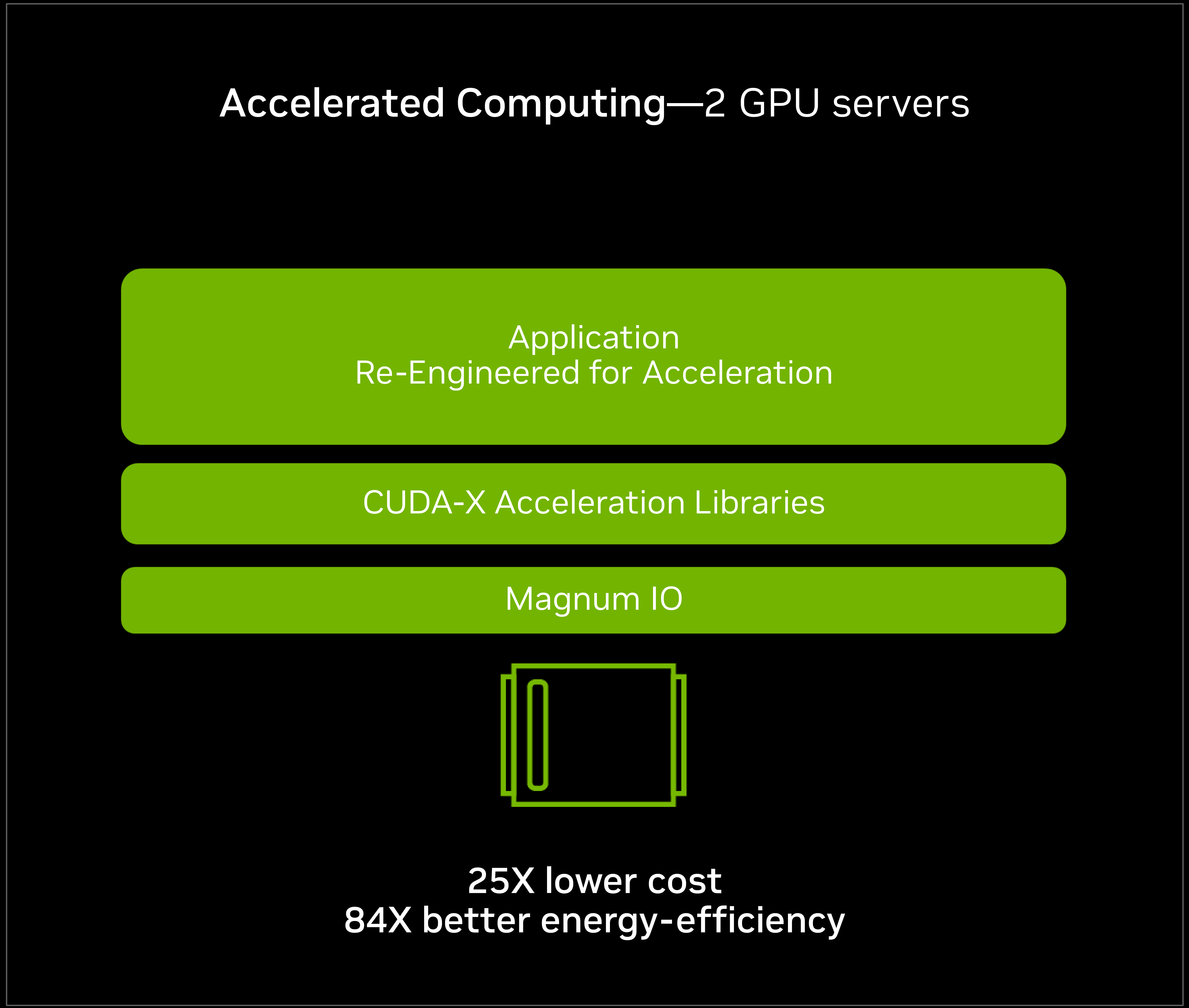
Drive Significant Cost Savings and Workload Scaling



Classical Computing—960 CPU-only servers

Application

CPU server racks



Accelerated Computing—2 GPU servers

Application
Re-Engineered for Acceleration

CUDA-X Acceleration Libraries

Magnum IO

25X lower cost
84X better energy-efficiency

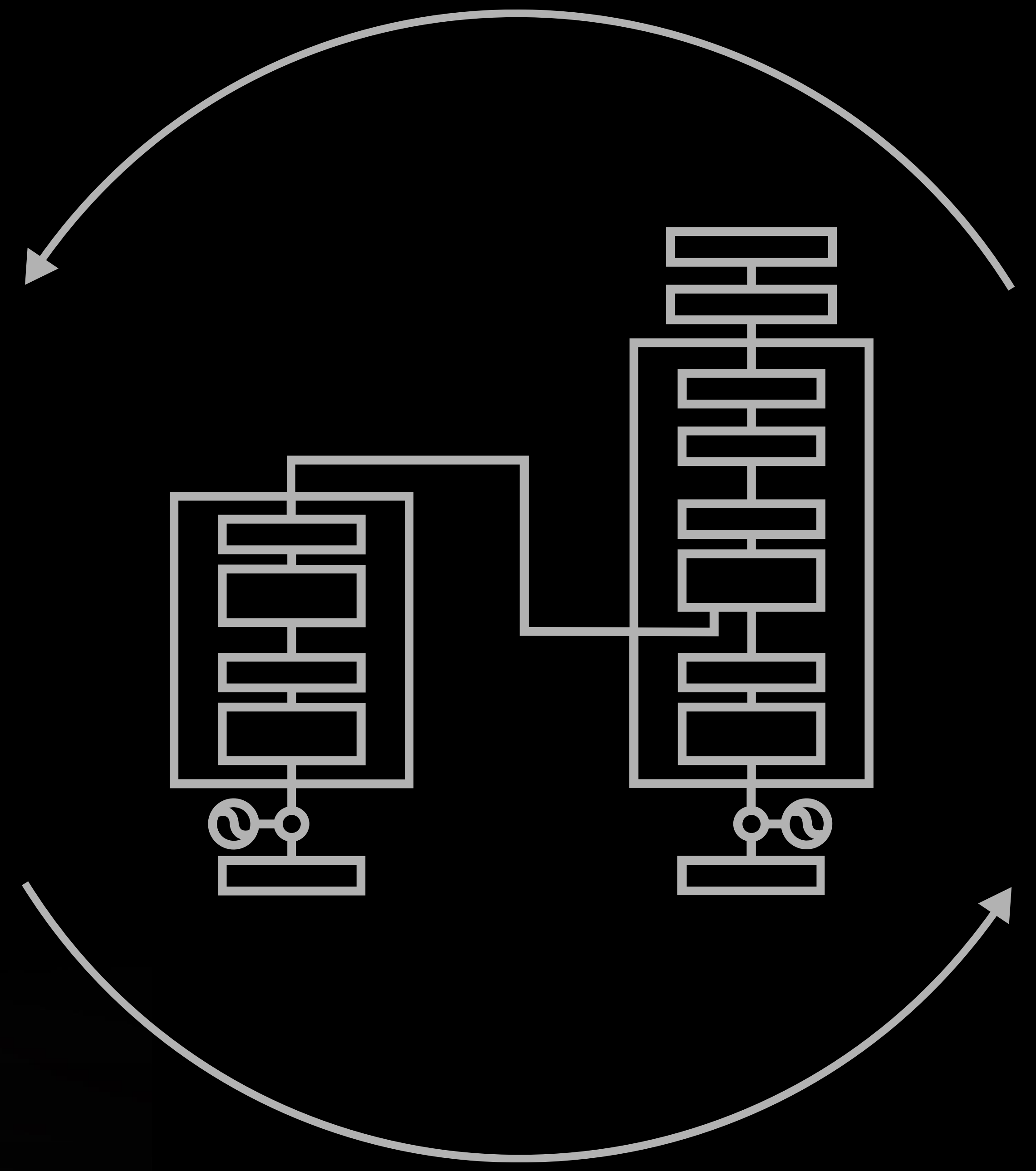
LLM Workload: Bert-Large Training and Inference | CPU Server: Dual-EYPC 7763 | GPU Server: Dual-EPYC 7763 + 8X H100 PCIe GPUs

Training & Inference — One Architecture

Cloud | On-Prem | Edge

TRAINING

AI FACTORIES



INFERENCE

IN THE DATA CENTER

NVIDIA L40
Image Generation

NVIDIA L4
AI Video

NVIDIA H100 NVL
Large Language Models

NVIDIA Grace Hopper
Recommendation Models

AT THE EDGE

IGX
Industrial-Grade System for Healthcare, Logistics, Manufacturing

AGX
Functionally-Safe System for Autonomous Vehicles

AI Is the Greatest Technology Force of Our Time

Data centers across industries will become AI factories

AI has fundamentally changed what software can make and how you make software.

Companies are processing & refining their data, making AI software—becoming intelligence manufacturers. Their data centers are AI factories.

The first wave of AI was learned perception and inference, like recognizing images, understanding speech, recommending a video, or an item to buy.

In late 2022, ChatGPT ushered in Generative AI – unlocking new opportunities for AI to generate text, images, video, code, or proteins.

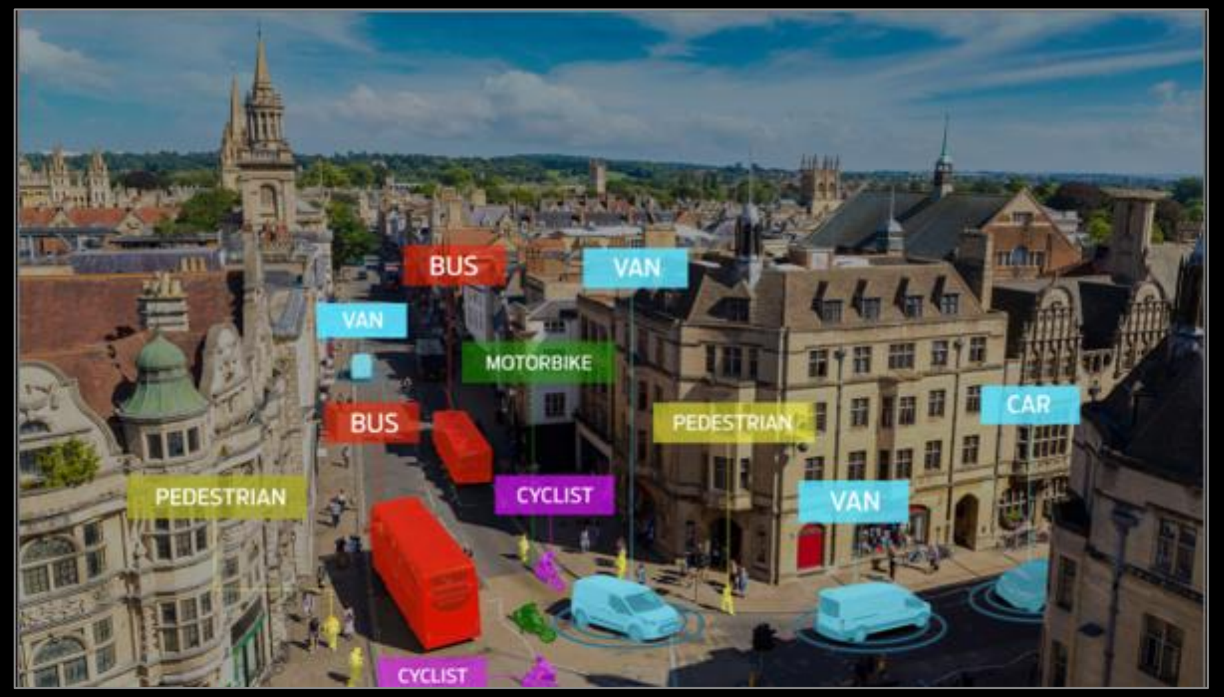
The next wave of AI is robotics and industrial digitalization — robots, avatars, and digital twins – where AI interacts with the physical world.

NVIDIA's acceleration stacks and ecosystems help bring AI to the world's largest industries.

NVIDIA's world-class AI expertise and scale can help revolutionize businesses.

Contact Center AI
500M Calls / Day

Meeting Transcription
3B Meeting Minutes / Day



Public Safety
>1B Smart City Cameras Deployed

Retail Asset Protection
$94.5B Inventory Loss / Year

Medical Imaging
10M Diagnostic Scans / Day

Industrial Inspection
$32M Vision Sensors Installed by 2025

Transportation
10T Miles / Year

Credit Card Fraud
1.28B Credit Transactions / Day

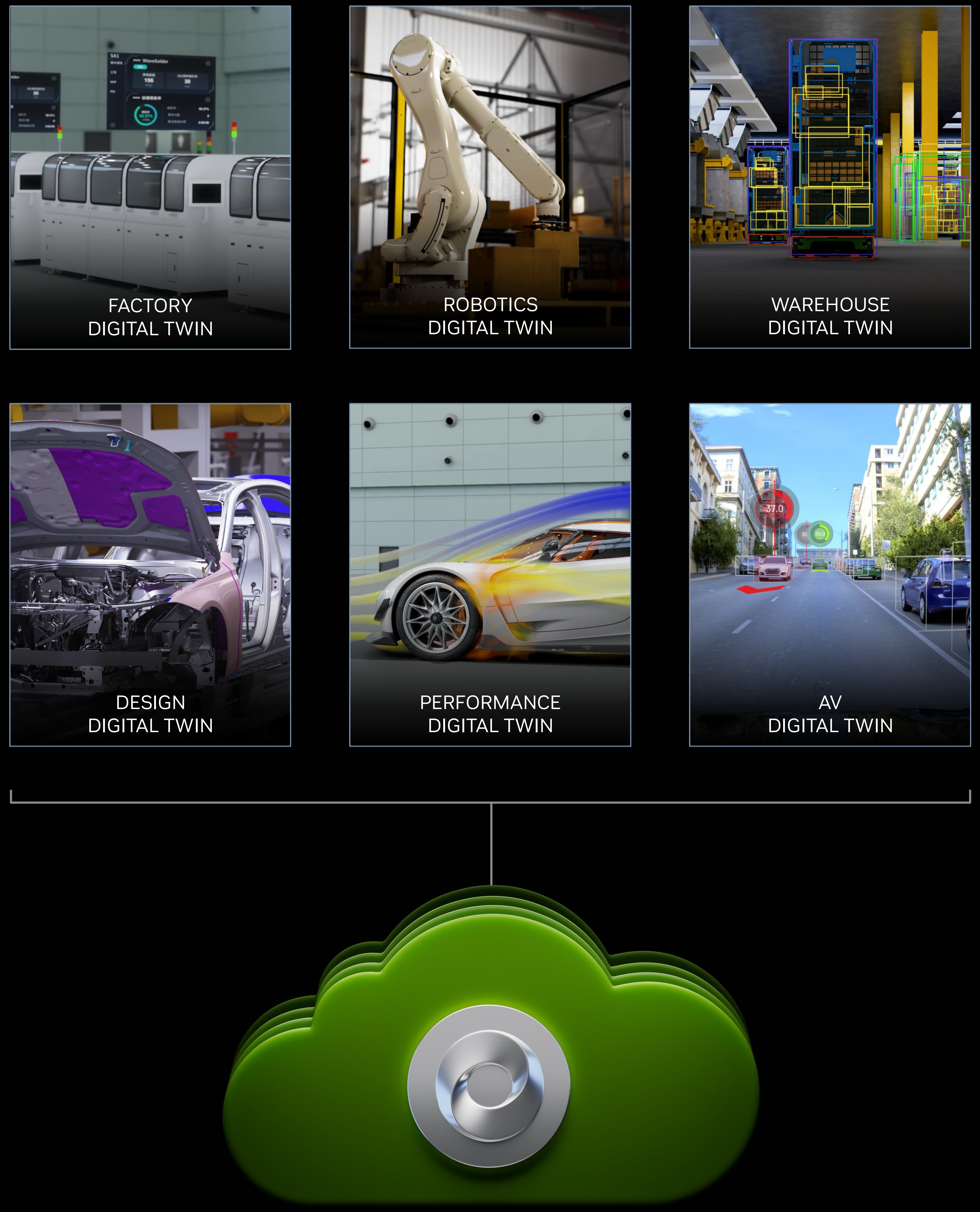
Product Recommendations
1B E-Commerce Visitors / Day

Source: Nilson Report, IHS Markit, Similar Web, NRF, WHO, ABI and NVIDIA internal analysis

Advancing Industrial Digitalization Efforts with NVIDIA Omniverse

NVIDIA OVX servers | RTX workstations
Enterprise software | Cloud services

- NVIDIA Omniverse is a software platform helping to power industrial digitalization.
- Our initial focus is on industrial digital twins used to emulate the behavior of products or factories in the physical world.
- Omniverse uses a real-time, large-scale 3D database that connects to 3D worlds via the USD (Universal Scene Descriptor) framework.
- Just as the internet connects websites over HTML, Omniverse connects 3D worlds over USD.
- Omniverse is essential for the next wave of AI—robotics—where AI interacts with the physical world.
- Applications built to run on Omniverse are like portals into the Omniverse virtual world.



FACTORY DIGITAL TWIN

ROBOTICS DIGITAL TWIN

WAREHOUSE DIGITAL TWIN

DESIGN DIGITAL TWIN

PERFORMANCE DIGITAL TWIN

AV DIGITAL TWIN

NVIDIA OMNIVERSE CLOUD

NVIDIA Software and Services

Enabling the world's enterprises to revolutionize industries with AI

NVIDIA-hosted cloud service for training Large Language Models to perform specific task s— e.g., summarize legal documents, write marketing copy, analyze market sentiment, chatbot to support customers, search documents, write and document code, paraphrase.

NeMo can help thousands of companies, train language AI's to do hundreds of tasks, in 10's of languages.

NVIDIA NeMo LLM

NVIDIA-hosted cloud service for training and deploying large biomolecular models that understand the language of chemistry, proteins, RNA, and DNA.

BioNeMo can help researchers, biotech, and pharma companies to process chemical and biological datasets to accelerate drug discovery.

NVIDIA BioNeMo

NVIDIA-hosted cloud service for building generative AI–powered visual applications.

Enterprises, software creators, and service providers can run inference on their models, train NVIDIA Edify foundation models on proprietary data, or start from pretrained models to generate image, video, and 3D content from text prompts.

Picasso service is fully optimized for GPUs and streamlines training, optimization, and inference.

NVIDIA Picasso

The operating engine of AI for end-to-end data-driven software development.

One engine license accelerates end-to-end modern AI and data science.

One engine license unlocks wealth of data processing, AI, and robotics frameworks and applications — e.g., RAPIDS, Spark, Merlin, Monai, Metropolis, cuOpt, Morpheus, Tokkio.

NVIDIA AI Enterprise

A platform for designing, building, and operating 3D and virtual world simulations.

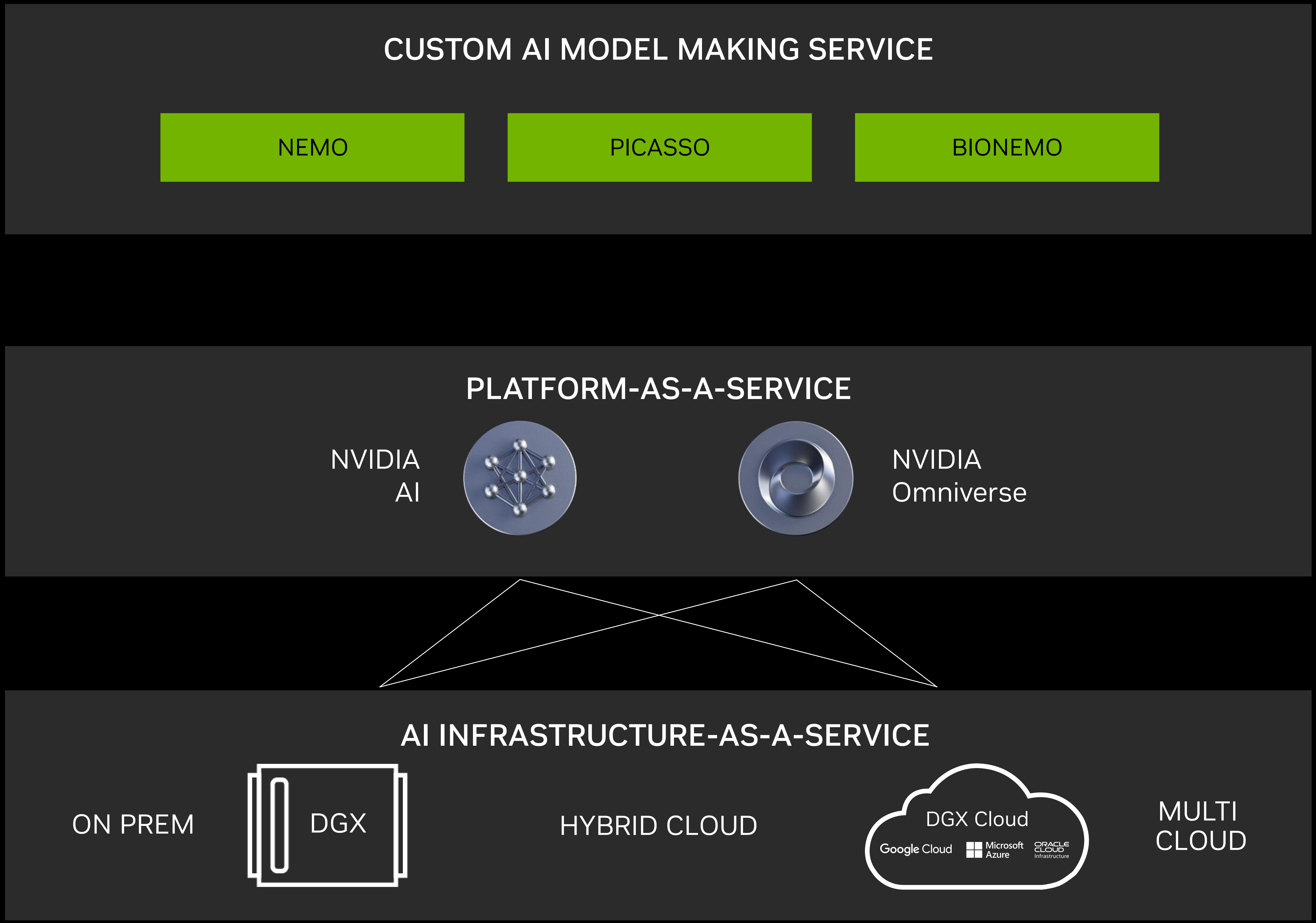
Consists of a virtual world engine, USD connectors, and portals browsing the virtual world simulation.

Omniverse is an enterprise application that connects architects, designers, hardware and software engineers, marketers, to supply-chain and factory planners.

NVIDIA Omniverse

NVIDIA Cloud Services

Engaging with customers at every layer



CUSTOM AI MODEL MAKING SERVICE

NEMO

PICASSO

BIONEMO

PLATFORM-AS-A-SERVICE

NVIDIA AI

NVIDIA Omniverse

AI INFRASTRUCTURE-AS-A-SERVICE

ON PREM

DGX

HYBRID CLOUD

DGX Cloud

MULTI CLOUD

NVIDIA Go-to-Market Across Cloud and On-Premises

Reaching customers everywhere

CLOUD

NVIDIA

DGX Cloud

Google Cloud Microsoft Azure ORACLE CLOUD Infrastructure

NeMo | Picasso | BioNeMo

PARTNERS

aws Google Cloud Microsoft Azure ORACLE CLOUD Infrastructure

HGX

INFERENCE

ON-PREM

NVIDIA

DGX

PARTNERS

DELL Technologies Hewlett Packard Enterprise Lenovo

EGX

AGX

IGX

Giant Market Opportunity

Gaming | Financial Services | Healthcare | Logistics | Manufacturing | Retail | Transportation

Gaming

Over 3B gamers and creators, a quarter of them spending over $100/year for GPUs in desktops, laptops, cloud or consoles

NVIDIA AI Enterprise Software

50M enterprise server installed base; per-server, per-year subscription price

Omniverse Enterprise Software

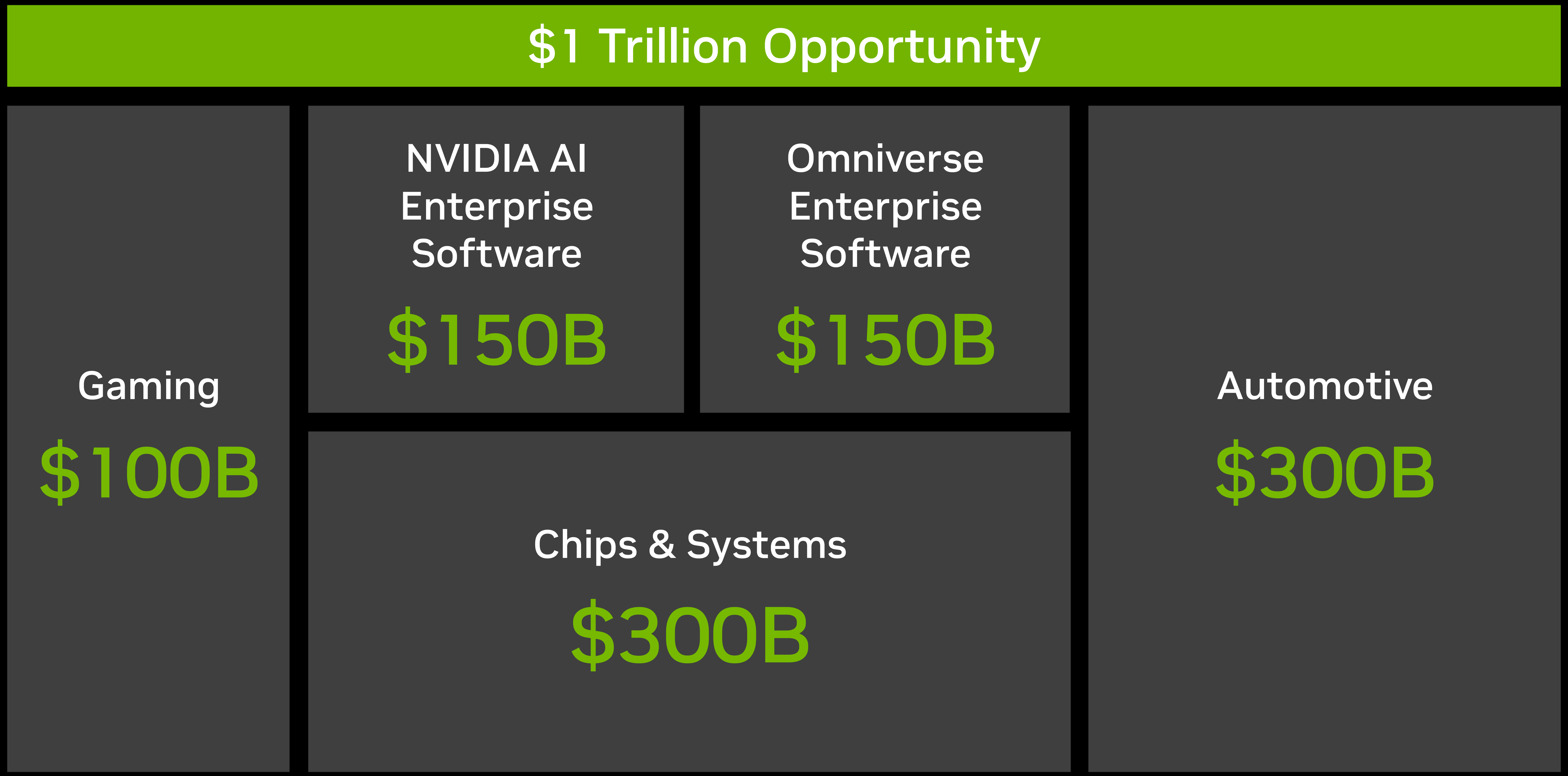
Over 45M designers and creators; 10s of millions of digital twins—per-user/digital twin, per-year subscription price

Chips and Systems

~20M servers/year—GPUs, CPUs, DPUs, NICs, switches

Automotive

100M vehicles/year hardware opportunity; 100s of millions of AV vehicles installed base software opportunity



$1 Trillion Opportunity

Gaming $100B

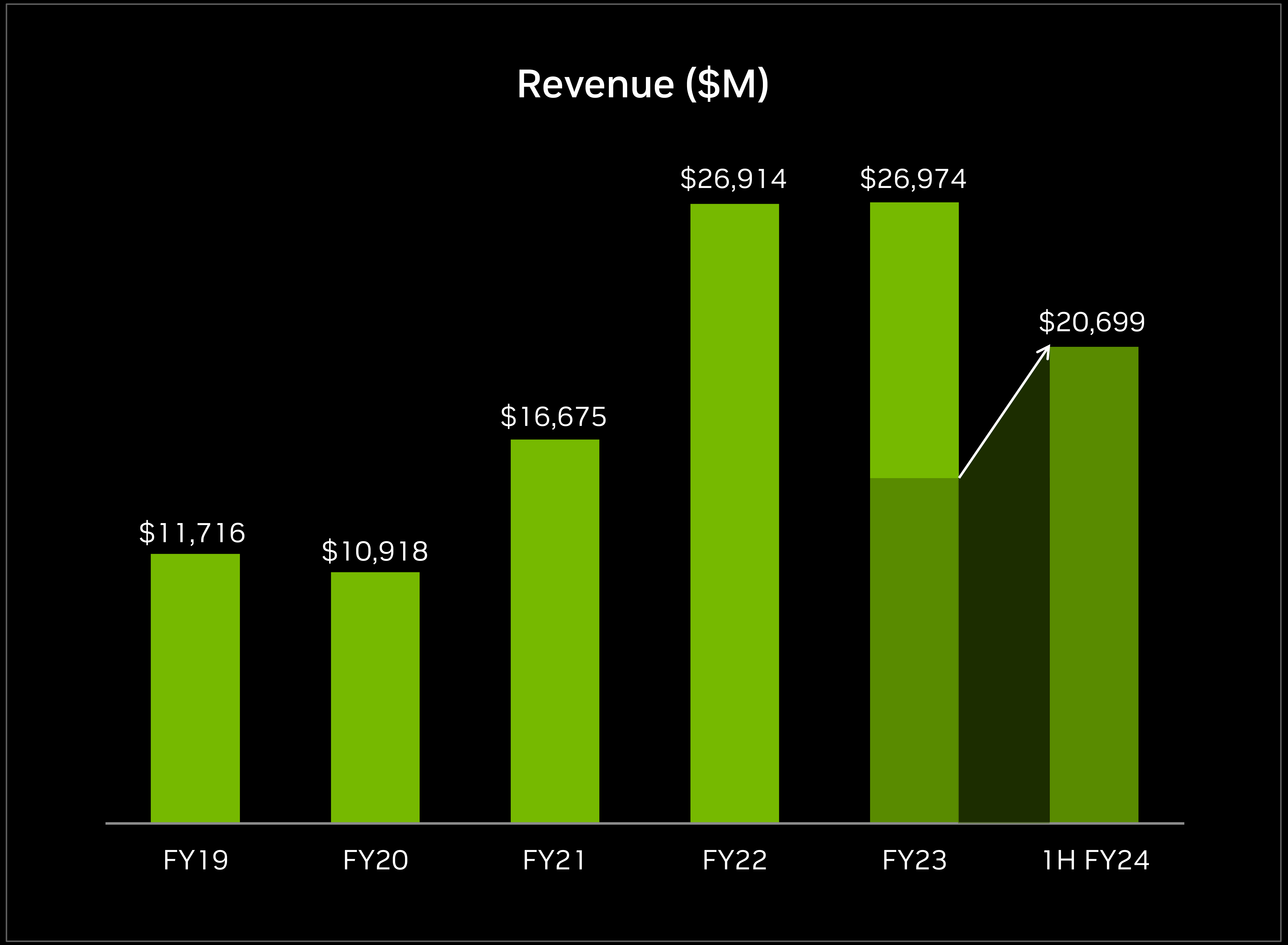
NVIDIA AI Enterprise Software $150B

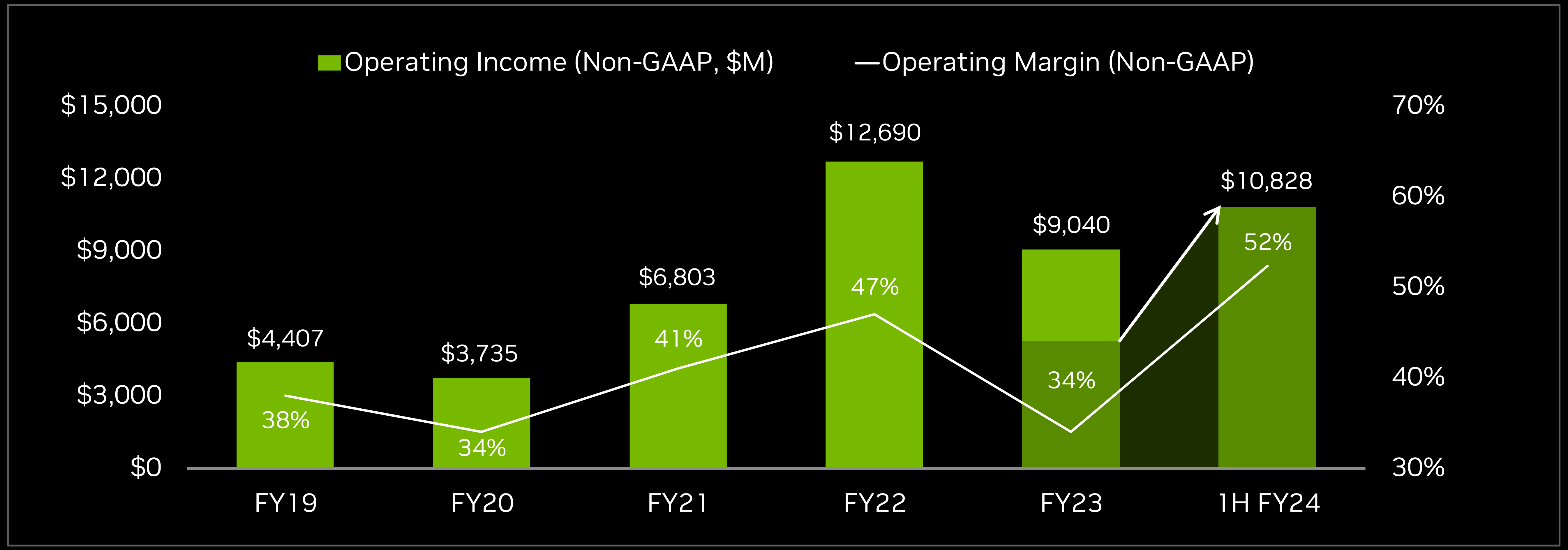
Omniverse Enterprise Software $150B

Chips & Systems $300B

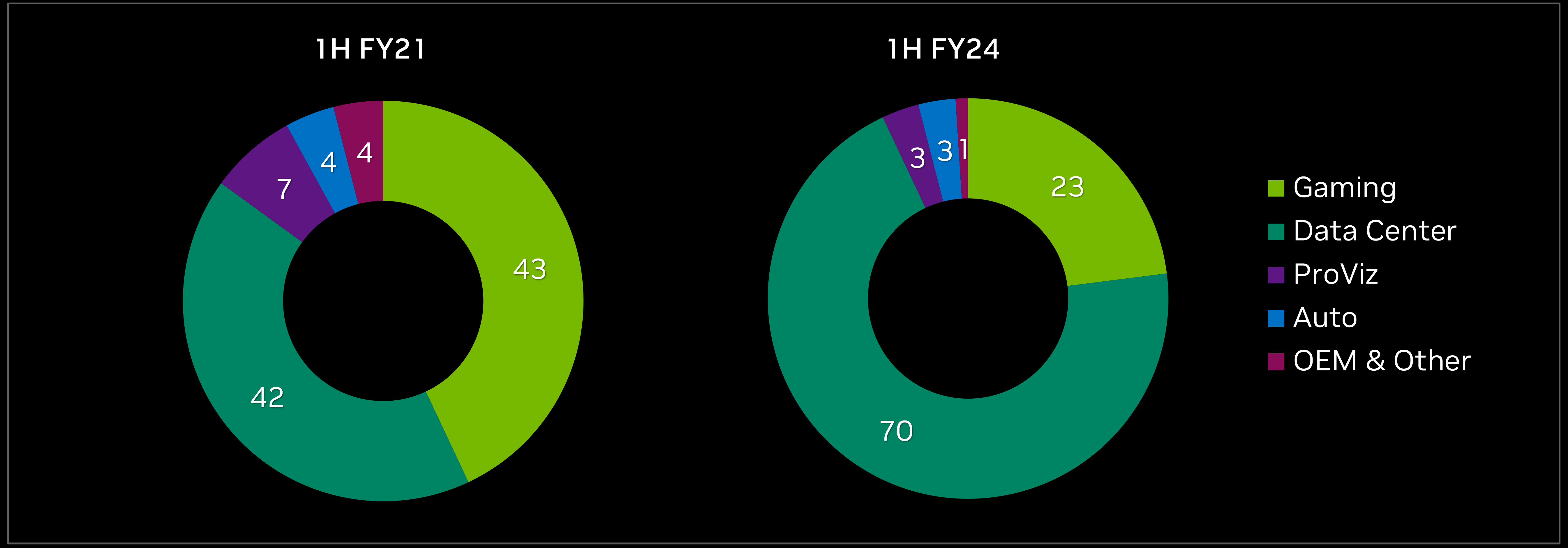
Automotive $300B

Driving Strong & Profitable Growth





Fiscal year ends in January. Refer to Appendix for reconciliation of Non-GAAP measures. Operating margins rounded to the nearest percent.



FY23 financial metrics reflect a $2.2B charge for inventory and related reserves primarily related to Data Center and Gaming.

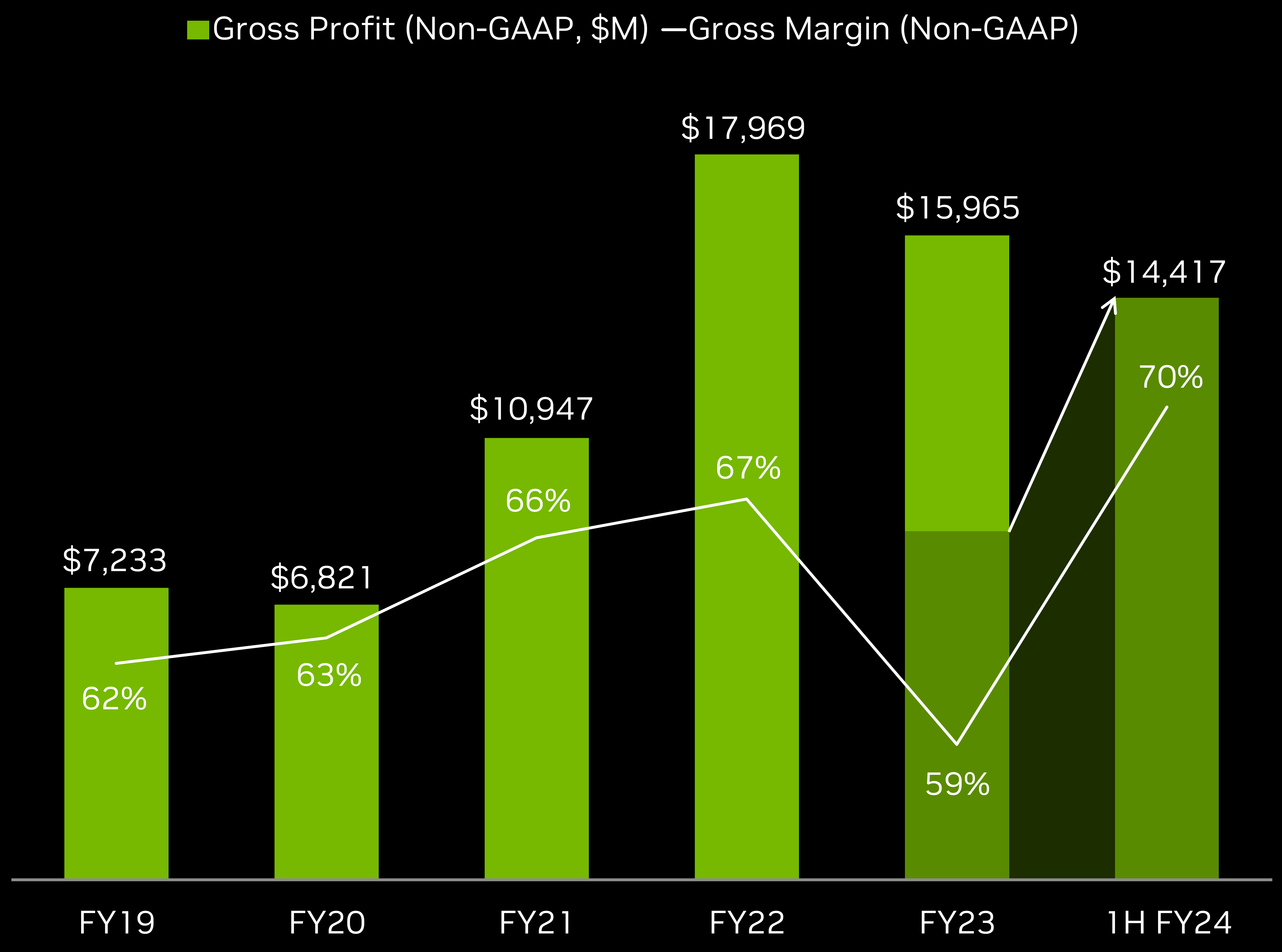
NVIDIA Gross Margins Reflect Value of Acceleration

Accelerated computing requires full-stack and data center-scale innovation across silicon, systems, algorithms and applications.

Significant expertise and effort are required, but application speed-ups can be incredible, resulting in dramatic cost and time-to-solution savings.

> For example, 10 NVIDIA HGX nodes with 80 NVIDIA A100 GPUs that cost $4M can replace 920 nodes of CPU servers that cost over $50M for AI inference.

NVIDIA chips carry the value of the full-stack, not just the chip.

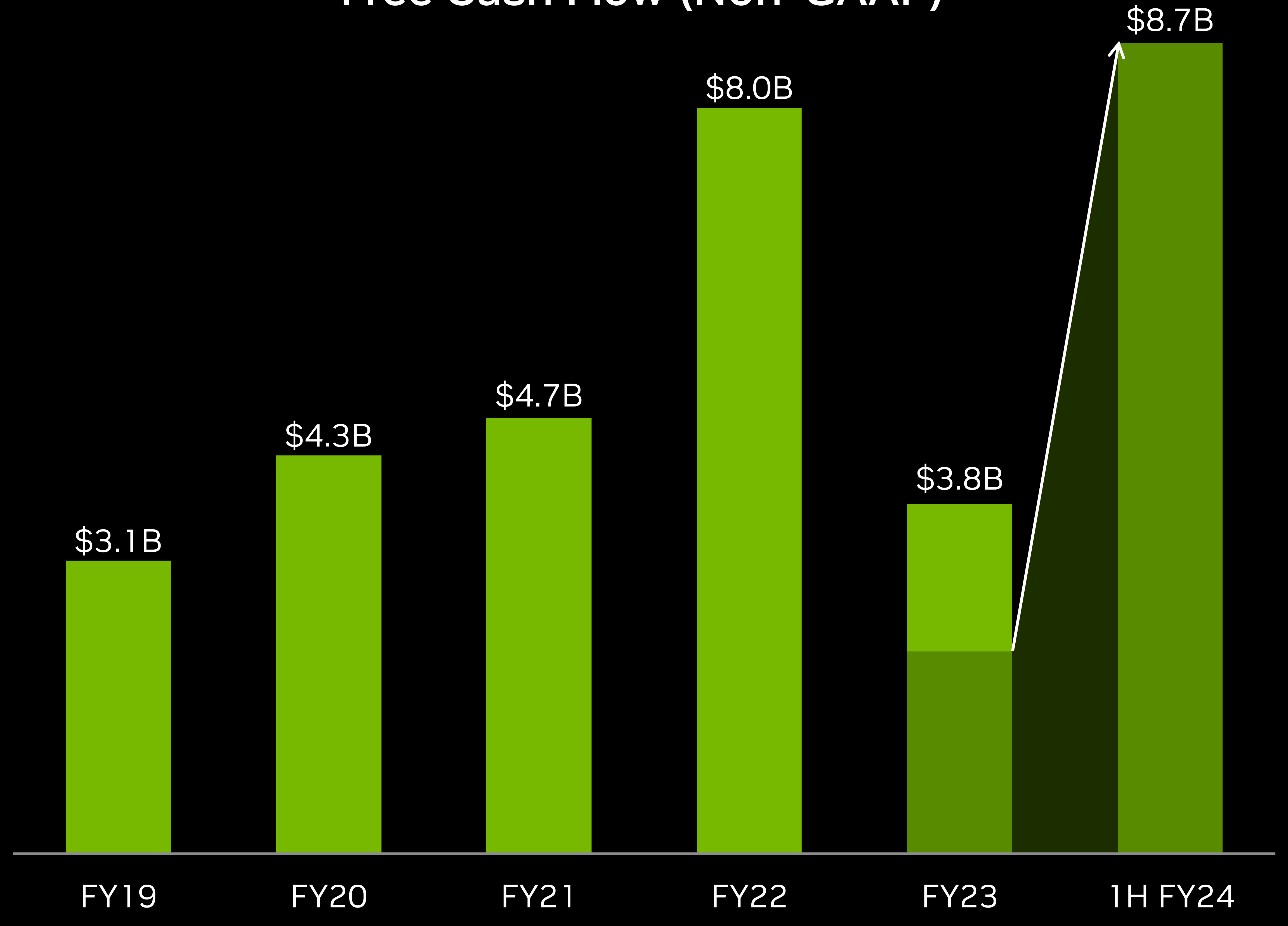


Cost comparison example based on latest available NVIDIA A100 GPU and Intel CPU inference results in the commercially available category of the MLPerf industry benchmark; includes related infrastructure costs such as networking.

FY23 financial metrics reflect a $2.2B charge for inventory and related reserves primarily related to Data Center and Gaming. Fiscal year ends in January. Refer to Appendix for reconciliation of Non-GAAP measures. Gross margins are rounded to the nearest percent.

Strong Cash Flow Generation

Free Cash Flow (Non-GAAP)



Capital Allocation

Share Repurchase

$10B repurchased in FY23

Additional $25B in stock repurchases authorized, adding to $4B which remained as of end of Q2

Dividend

$398M in FY 2023

Plan to Maintain[1]

Strategic Investments

Growing Our Talent

Platform Reach & Ecosystem

Fiscal year ends in January. Refer to Appendix for reconciliation of Non-GAAP measures.

[1] Subject to continuing determination by our Board of Directors.

Our Market Platforms at a Glance

Data Center

56% of FY23 revenue

FY23 Revenue $15.0B
5-yr CAGR 51%

DGX/HGX/EGX/IGX systems

GPU | CPU | DPU | Networking
NVIDIA AI software

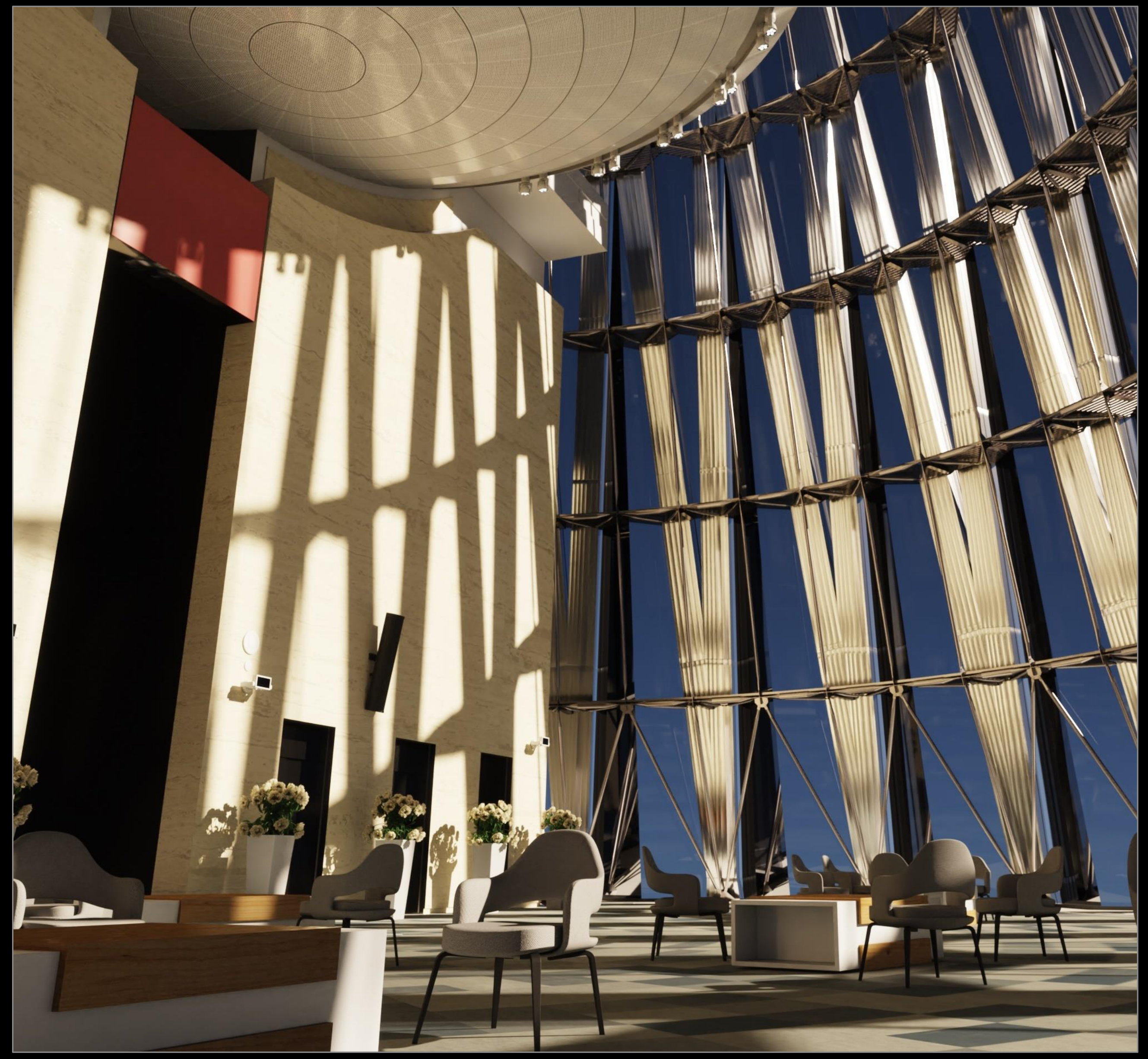
Gaming

33% of FY23 revenue

FY23 Revenue $9.1B
5-yr CAGR 10%

GeForce GPUs for PC gaming

GeForce NOW cloud gaming



Professional Visualization

6% of FY23 revenue

FY23 Revenue $1.5B
5-yr CAGR 11%

Quadro/NVIDIA RTX GPUs
for workstations

Omniverse software

Automotive

3% of FY23 revenue

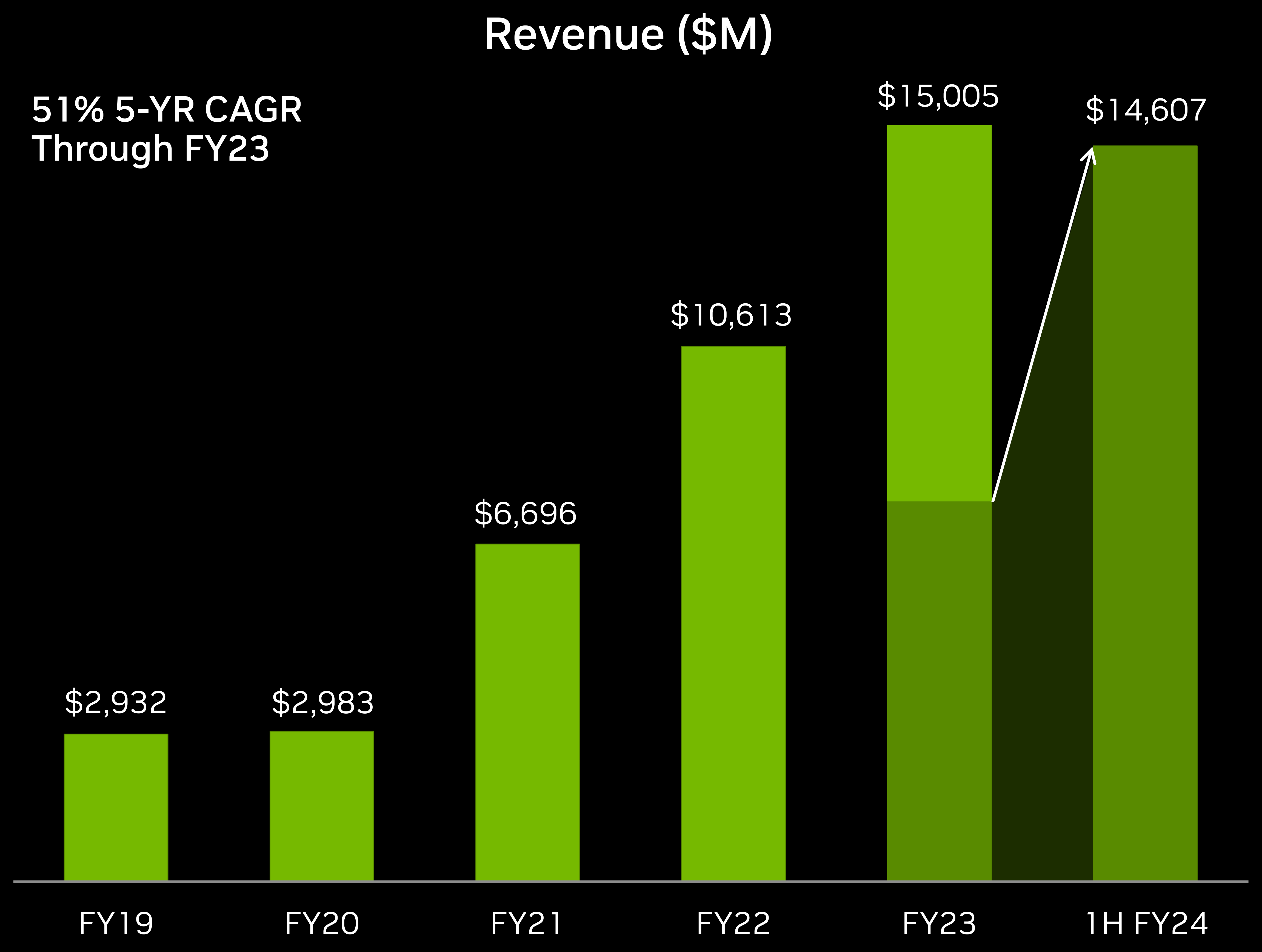
FY23 Revenue $0.9B
5-yr CAGR 10%

DRIVE Hyperion sensor architecture
with AGX compute

DRIVE AV & IX full stack software
for ADAS, AV & AI cockpit

Data Center

The leading computing platform for AI, HPC & graphics



Leader in AI & HPC

#1 in AI training and inference

Used by all hyperscale & major cloud computing providers and 35,000 organizations

Powers 74% of the TOP500 supercomputers

Growth Drivers

Rapid AI adoption across industries

Full-stack AI | Software

Three chip strategy — GPU | CPU | DPU

Rising computation requirements for modern AI

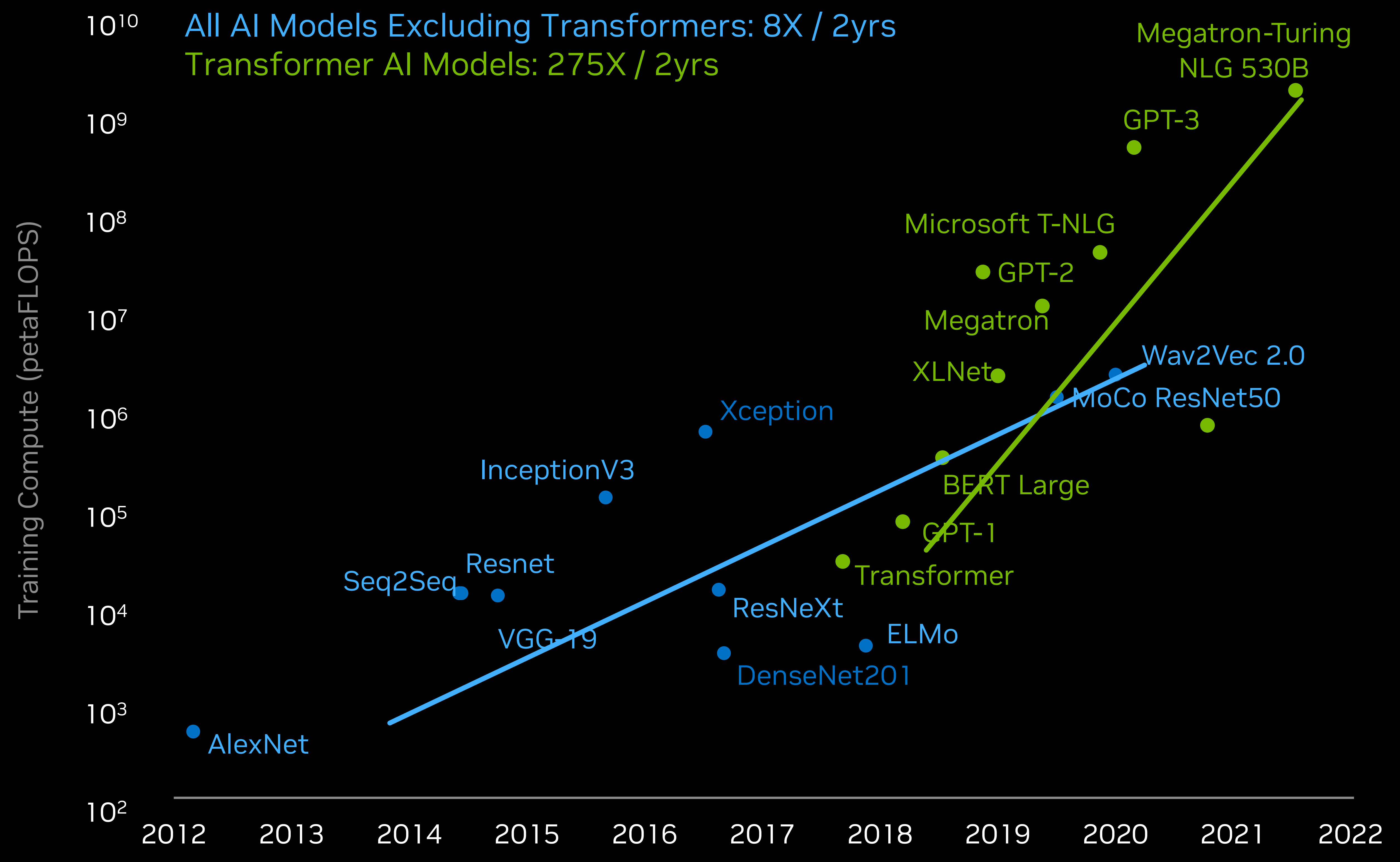
Data-center scale innovation

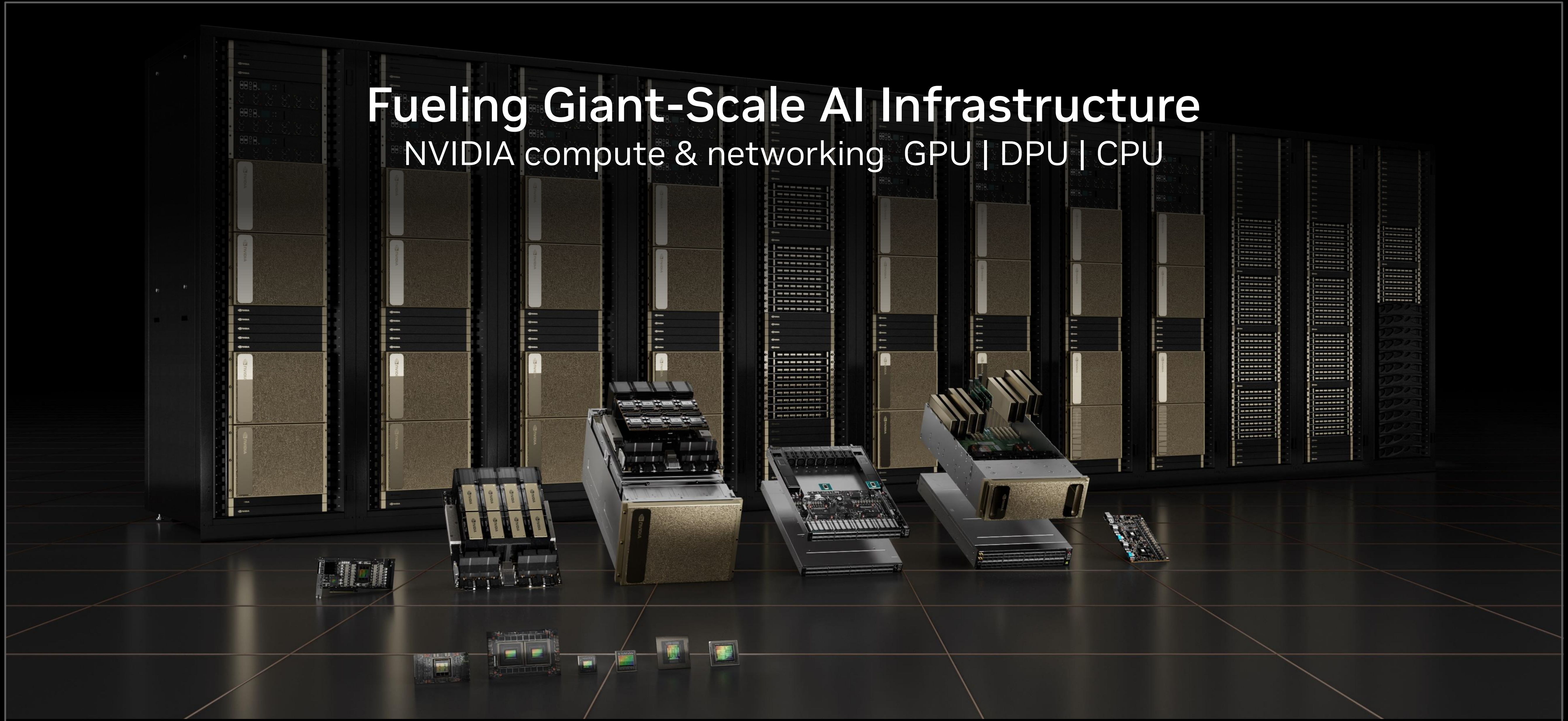
Omniverse

Modern AI is a Data Center Scale Computing Workload

Data centers are becoming AI factories: data as input, intelligence as output

AI Training Computational Requirements





Fueling Giant-Scale AI Infrastructure

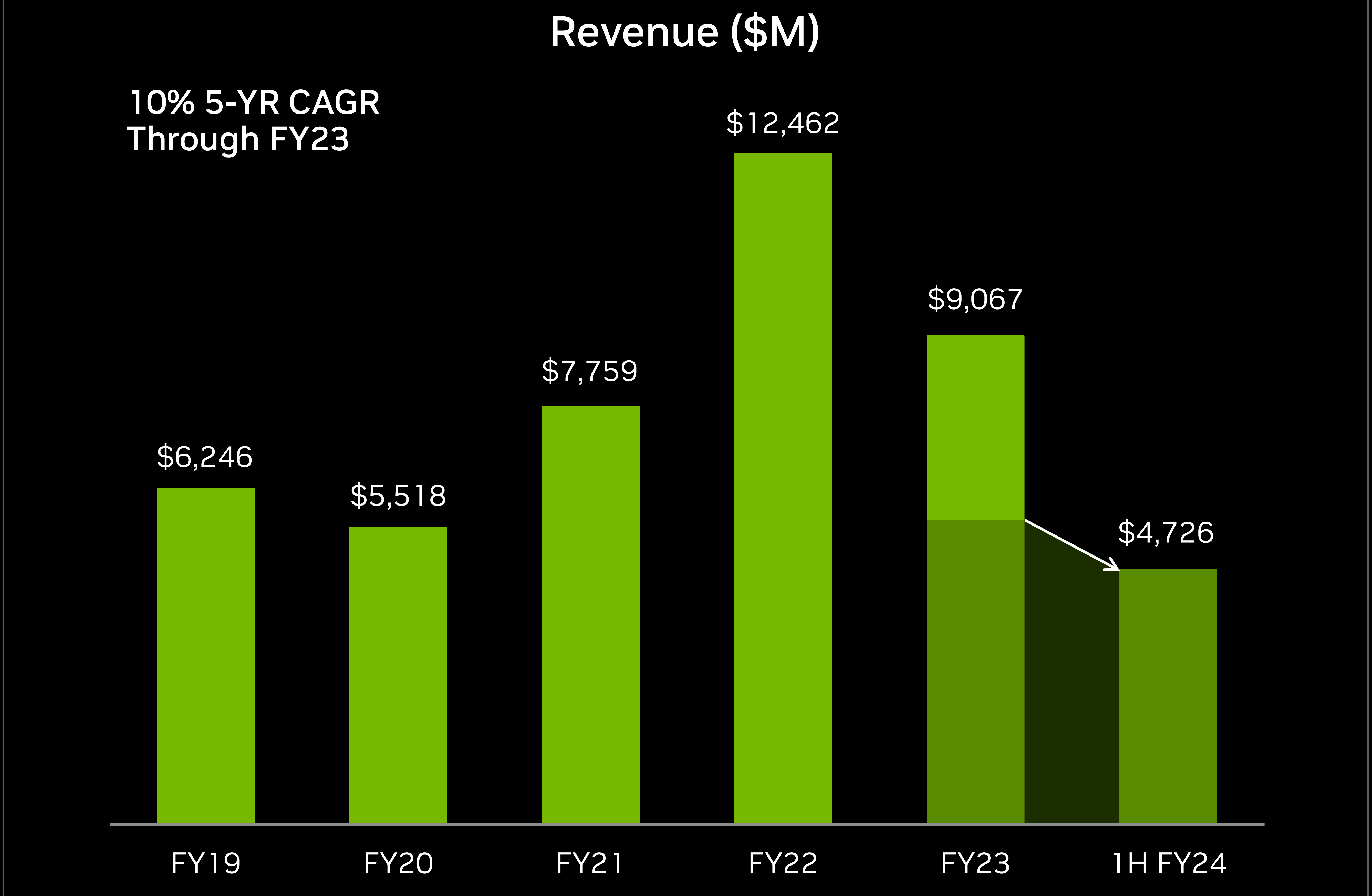
NVIDIA compute & networking GPU | DPU | CPU

Large Language Models, based on the Transformer architecture, are one of today's most important advanced AI technologies, involving up to trillions of parameters that learn from text.

Developing them is an expensive, time-consuming process that demands deep technical expertise, distributed data center-scale infrastructure, and a full-stack accelerated computing approach.

Gaming

GeForce — the world's largest gaming platform



Leader in PC Gaming

Strong #1 market position

15 of the top 15 most popular GPUs on Steam

Leading performance & innovation

200M+ gamers on GeForce

Growth Drivers

Rising adoption of NVIDIA RTX in games

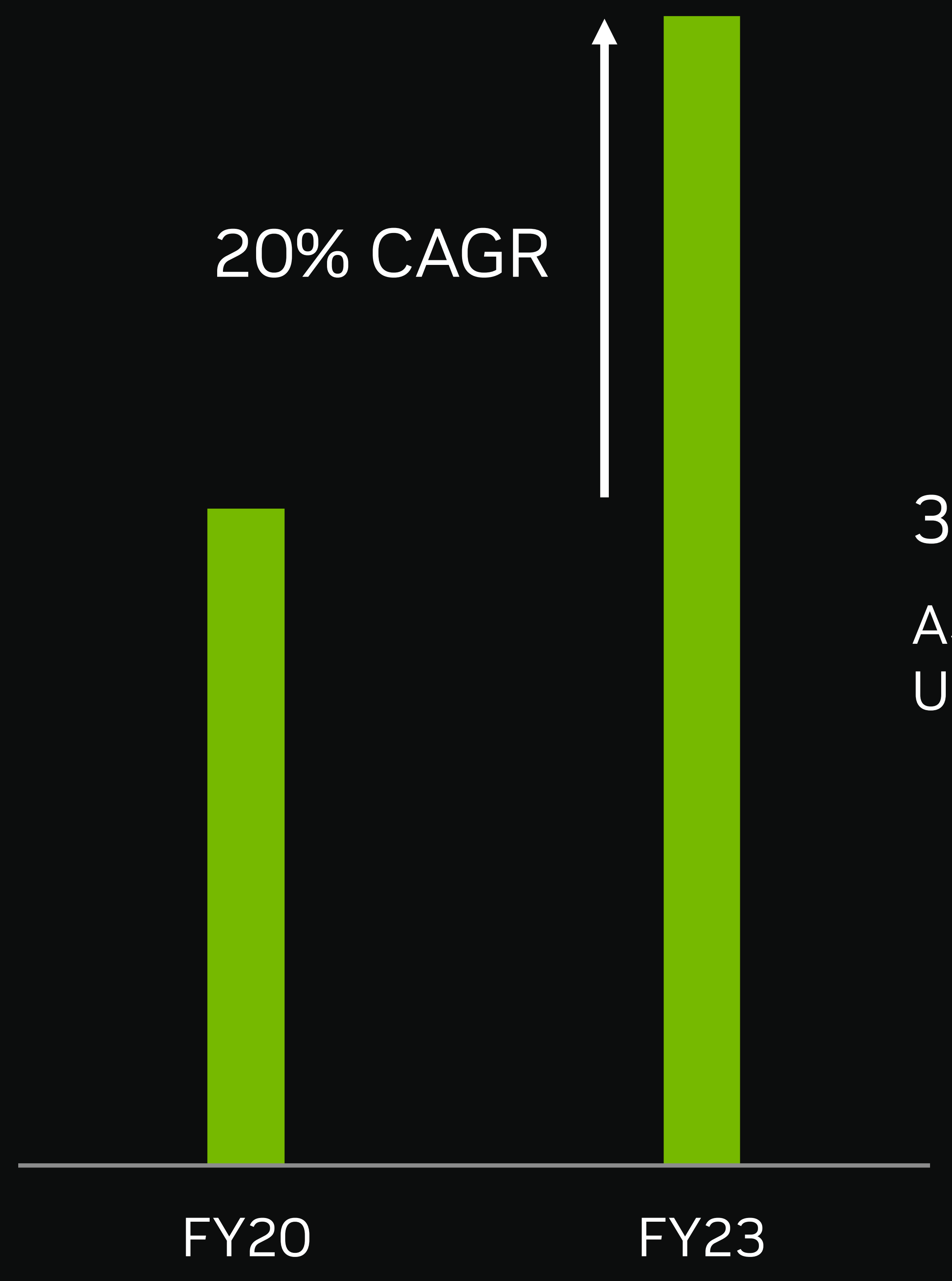
Expanding universe of gamers & creators

Gaming laptops & Gen AI on PCs

GeForce NOW Cloud gaming

GeForce Extends Growth, Large Upgrade Opportunity

GeForce Gaming Revenue

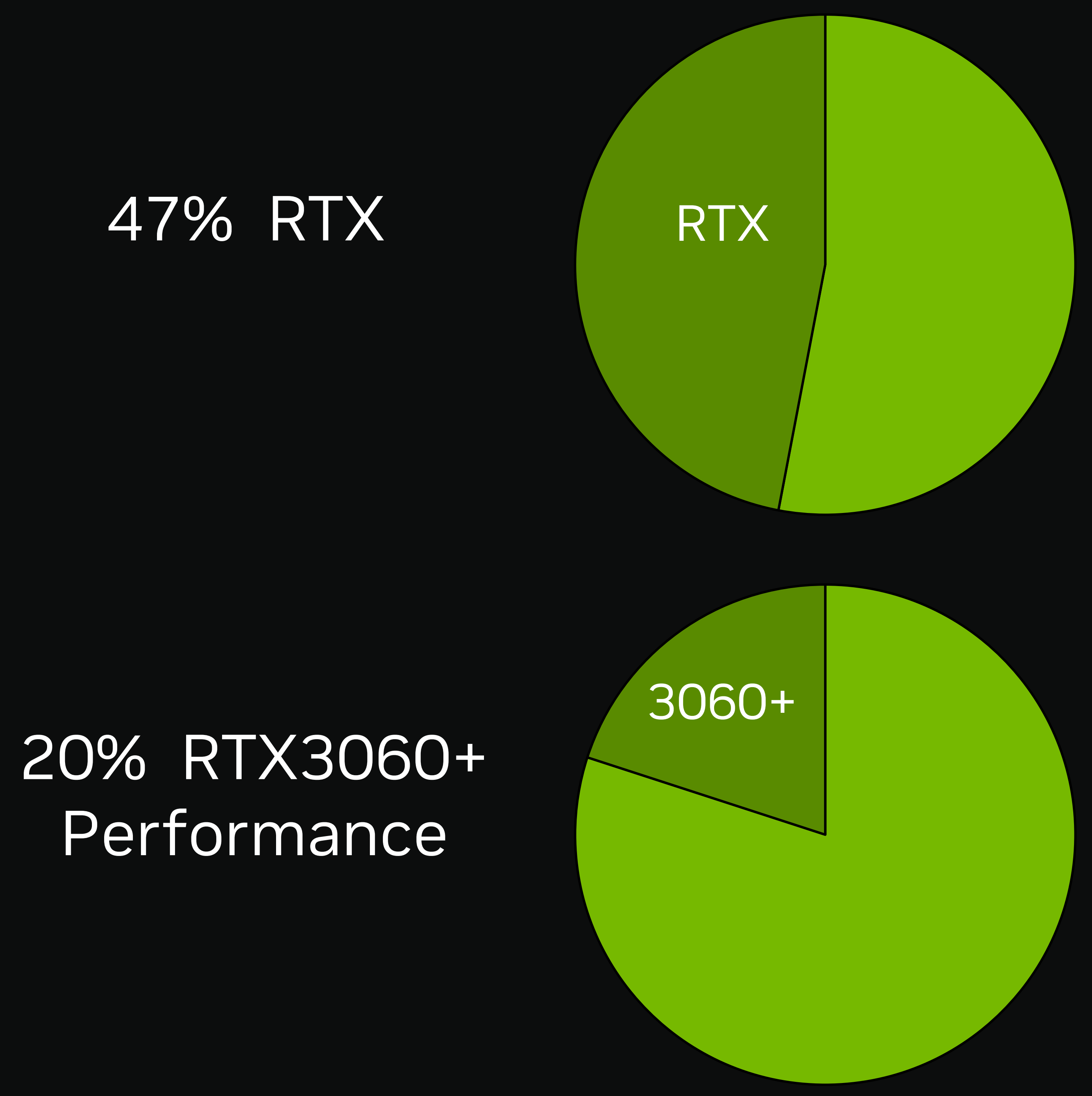


3YR CAGR

ASP 10%
Units 9%

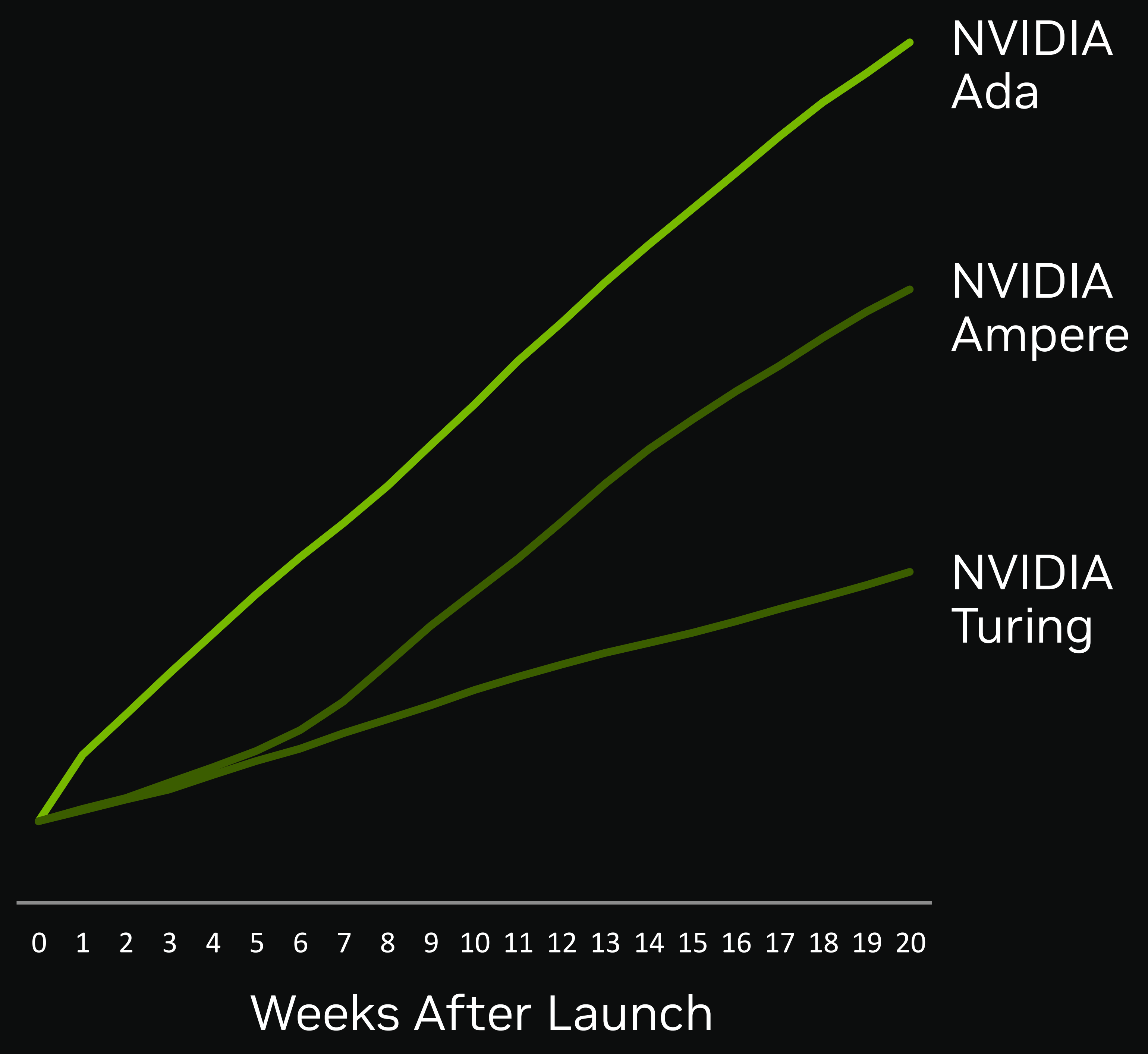
More Gamers, Richer Mix

Installed Base



Installed Base Needs Upgrade

$699+ Cumulative Sell-Through $



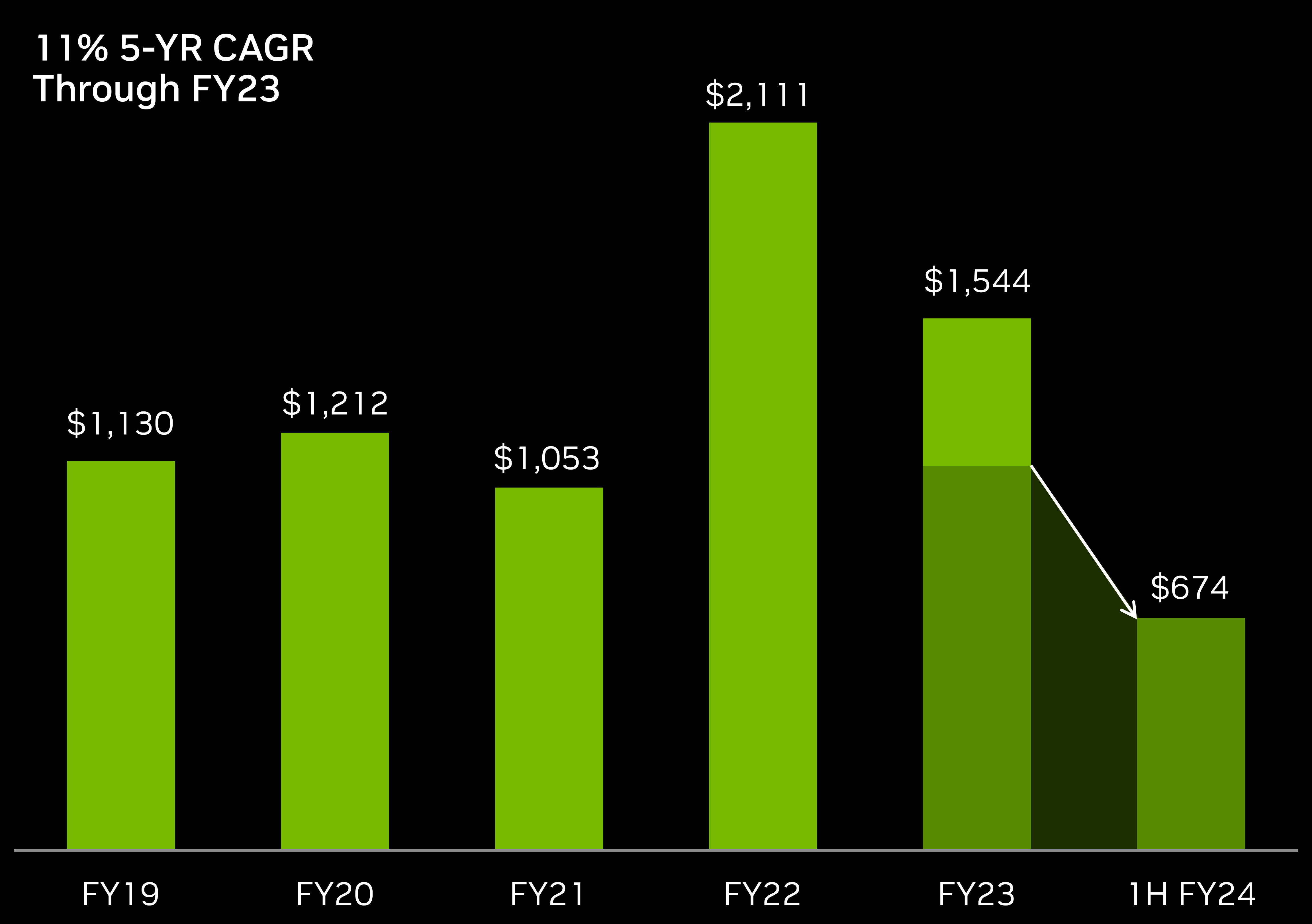
Ada: 3X Turing Ramp at $699+

Source: NVIDIA estimates

Professional Visualization

Workstation graphics

Revenue ($M)



Leader in Workstation Graphics

90%+ market share in graphics for workstations

45M Designers and Creators

Strong software ecosystem with over 100 supported applications

Growth Drivers

Ray Tracing and AI revolutionizing design

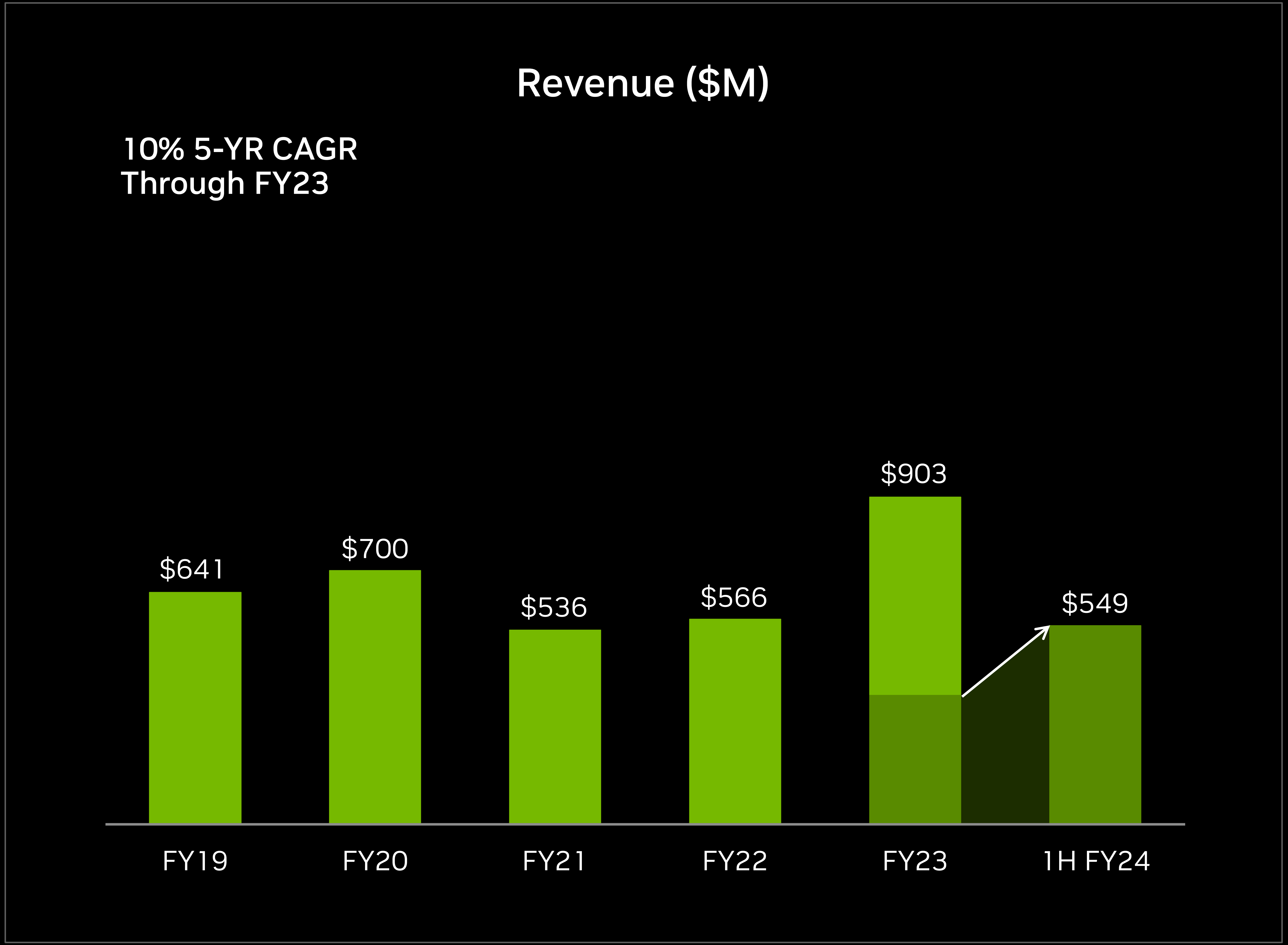
Expanding universe of designers and creators

Collaborative 3D design / Omniverse

Hybrid work environments

Automotive

Autonomous Vehicles (AV) & AI Cockpit



Leader in Autonomous Driving

Historical revenue driven largely by infotainment

Future growth primarily fueled by NVIDIA DRIVE, our AV and AI cockpit platform with full software stack

Next-generation DRIVE Thor to ramp in FY26

Growth Drivers

Adoption of centralized car computing and software-defined vehicle architectures

AV software and services:
Mercedes Benz
Jaguar Land Rover

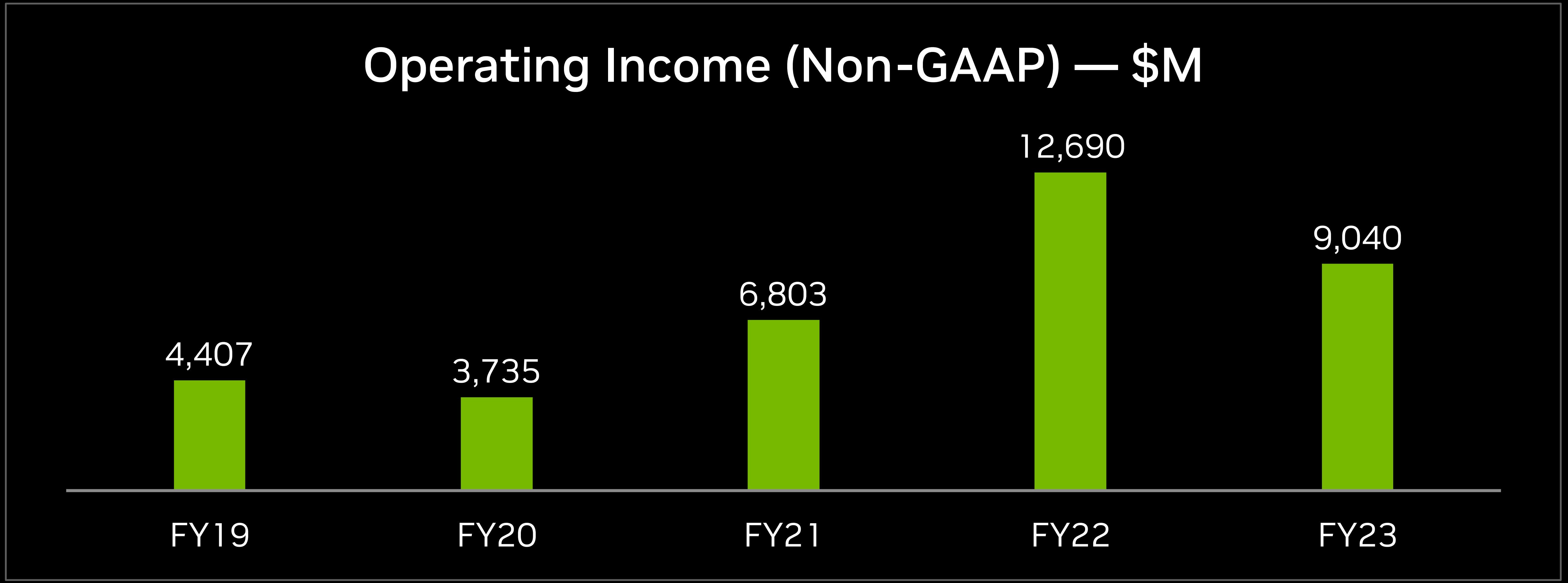
Growing NVIDIA DRIVE Pipeline

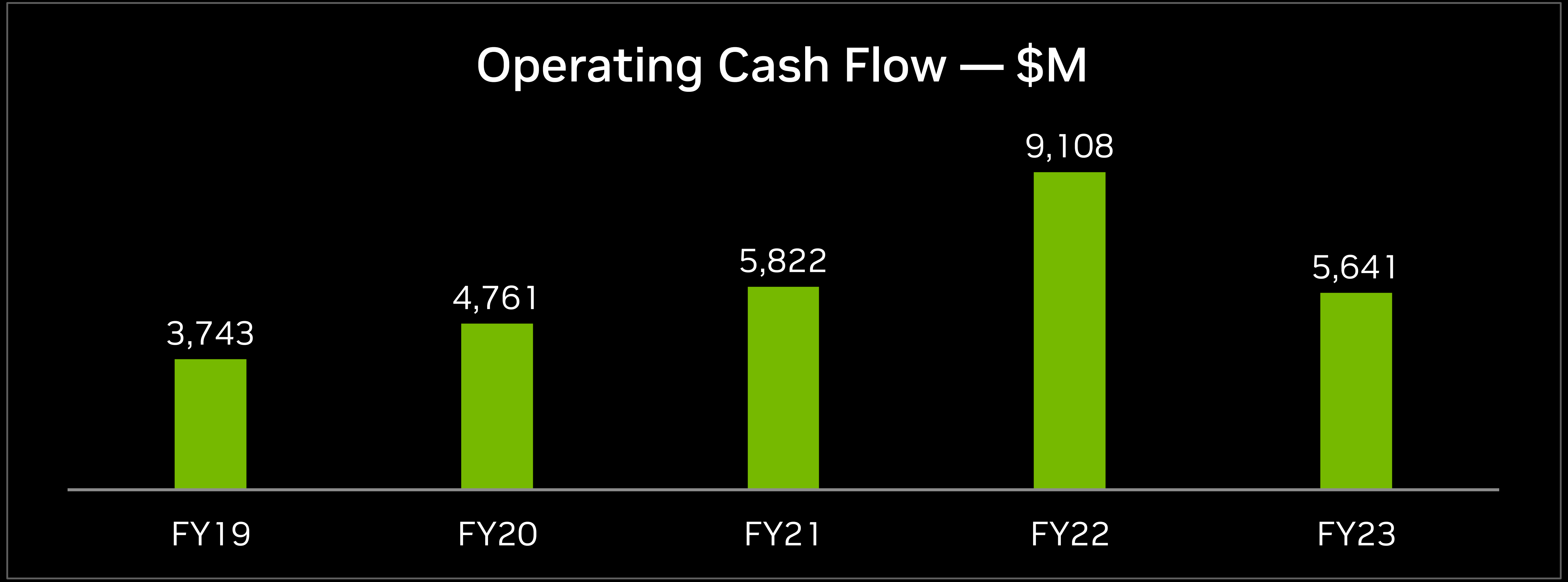
$14B Design Win Pipeline — 6 Year Horizon

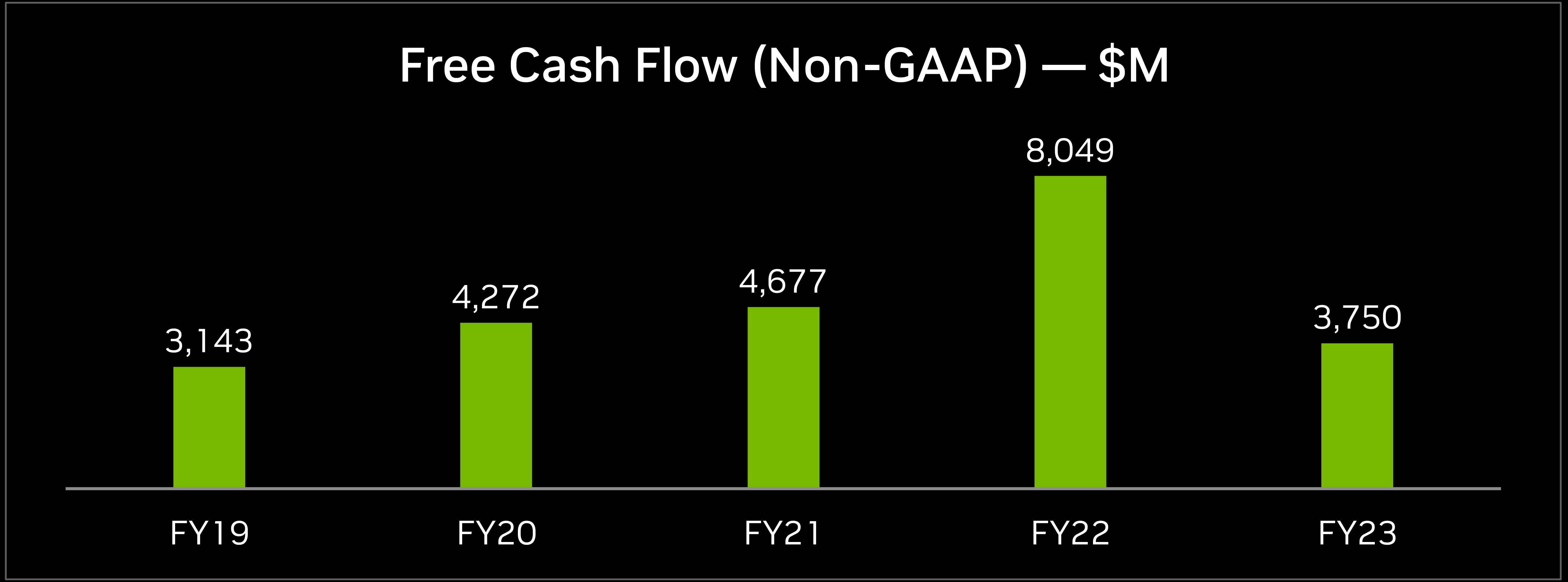
JAGUAR
LAND ROVER
BYD
LUCID
ZOOX
GENESIS
HYUNDAI
KIA
cruise
VOLVO
Polestar
NETA
LOTUS
NIO
VINFAST
Li Auto
ZEEKR
XPENG
SAIC
上汽集团
SAIC MOTOR
ROEWE
IM
智己汽车
DiDi
JIDU
2023

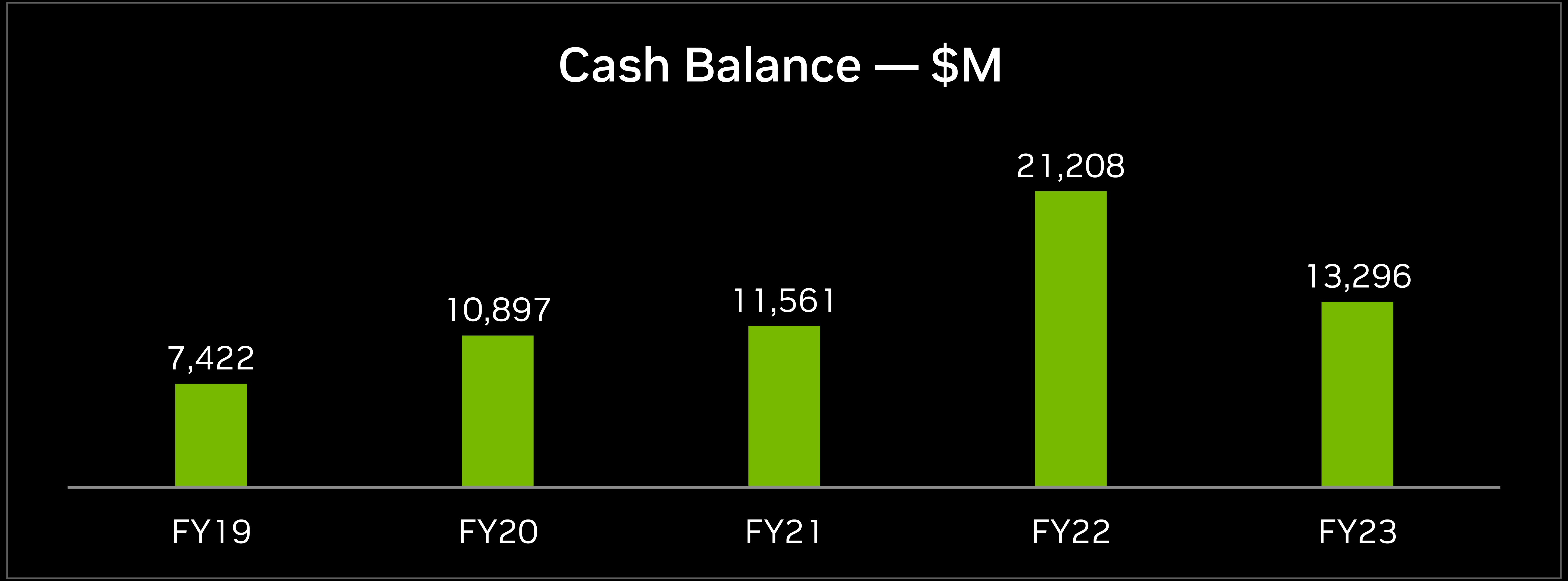
Financials

Annual Cash & Cash Flow Metrics









Cash balance is defined as cash and cash equivalents plus marketable securities
Refer to Appendix for reconciliation of non-GAAP measures

Corporate Responsibility

Environmentally Conscious

By FY26, aim to engage manufacturing suppliers comprising at least 67% of NVIDIA's scope 3 category 1 GHG emissions with goal of effecting supplier adoption of science-based targets

NVIDIA GPUs are typically 20X more energy efficient for certain AI and HPC workloads than traditional CPUs

Plan to achieve & maintain 100% renewable electricity for our operations and data centers by FY25 and annually thereafter

A Place For People To Do Their Life's Work

glassdoor BEST PLACES TO WORK 2023

"100 Best Companies to Work For"
FORTUNE

"America's Most Just Companies"
CNBC

"Most Responsible Companies"
NEWSWEEK

"Best Places to Work for LGBT Equality"
HUMAN RIGHTS CAMPAIGN

Management

Time Magazine's 100 Most Influential Companies

Fast Company's Best Workplaces for Innovators

Fortune's World's Most Admired Companies

Wall Street Journal's Management Top 250 All-Stars

Corporate Governance

43% of Board is Gender, Racially, or Ethnically Diverse

93% of Directors are independent

Reconciliation of Non-GAAP to GAAP Financial Measures

Reconciliation of Non-GAAP to GAAP Financial Measures

	Non-GAAP	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	Other (C)	Tax Impact of Adjustments	GAAP
Q2 FY24							
Gross margin ($ in million)	$9,614	(119)	(31)	(2)	—	—	$9,462
	71.2%	(0.9)	(0.2)	—	—	—	70.1%
Operating income ($ in million)	$7,776	(137)	(842)	(2)	5	—	$6,800
Net income ($ in million)	$6,740	(137)	(842)	(2)	66	363	$6,188
Shares used in diluted per share calculation (millions)	2,499	—	—	—	—	—	2,499
Diluted EPS	$2.70	—	—	—	—	—	$2.48

A. Consists of amortization of intangible assets, transaction costs, and certain compensation charges.
B. Stock-based compensation charge was allocated to cost of goods sold, research and development expense, and sales, general and administrative expense.
C. Other comprises of assets held for sale related adjustments and net gains from non-affiliated investments

Reconciliation of Non-GAAP to GAAP Financial Measures (contd.)

Gross Margin	Non-GAAP	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	GAAP
Q2 FY2023	45.9%	(1.8)	(0.6)	—	43.5%
Q3 FY2023	56.1%	(2.0)	(0.5)	—	53.6%
Q4 FY2023	66.1%	(2.0)	(0.5)	(0.3)	63.3%
Q1 FY2024	66.8%	(1.7)	(0.4)	(0.1)	64.6%

A. Consists of amortization of intangible assets
B. Stock-based compensation charge was allocated to cost of goods sold

Reconciliation of Non-GAAP to GAAP Financial Measures (contd.)

Gross Margin ($ in Millions & Margin Percentage)	Non-GAAP	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	GAAP
FY 2019	$7,233	—	(27)	(35)	$7,171
	61.7%	—	(0.2)	(0.3)	61.2%
FY 2020	$6,821	—	(39)	(14)	$6,768
	62.5%	—	(0.4)	(0.1)	62.0%
FY 2021	$10,947	(425)	(88)	(38)	$10,396
	65.6%	(2.6)	(0.5)	(0.2)	62.3%
FY 2022	$17,969	(344)	(141)	(9)	$17,475
	66.8%	(1.4)	(0.5)	—	64.9%
FY 2023	$15,965	(455)	(138)	(16)	$15,356
	59.2%	(1.7)	(0.5)	(0.1)	56.9%
1H FY 2023	$8,636	(214)	(76)	—	$8,346
	57.6%	(1.4)	(0.5)	—	55.7%
1H FY 2024	$14,417	(239)	(58)	(10)	$14,110
	69.7%	(1.2)	(0.3)	—	68.2%

A. Consists of amortization of intangible assets and inventory step-up

B. Stock-based compensation charge was allocated to cost of goods sold

Reconciliation of Non-GAAP to GAAP Financial Measures (contd.)

Operating Income and Margin ($ in Millions & Margin Percentage)	Non-GAAP	Acquisition Termination Cost	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	Other (C)	GAAP
FY 2019	$4,407	—	(2)	(557)	(35)	(9)	$3,804
	37.6%	—	—	(4.7)	(0.3)	(0.1)	32.5%
FY 2020	$3,735	—	(31)	(844)	(14)	—	$2,846
	34.2%	—	(0.3)	(7.7)	(0.1)	—	26.1%
FY 2021	$6,803	—	(836)	(1,397)	(38)	—	$4,532
	40.8%	—	(5.0)	(8.4)	(0.2)	—	27.2%
FY 2022	$12,690	—	(636)	(2,004)	(9)	—	$10,041
	47.2%	—	(2.5)	(7.4)	—	—	37.3%
FY 2023	$9,040	(1,353)	(674)	(2,710)	(16)	(63)	$4,224
	33.5%	(5.0)	(2.5)	(10.0)	(0.1)	(0.2)	15.7%
1H FY 2023	$5,280	(1,353)	(324)	(1,227)	—	(9)	$2,367
	35.2%	(9.0)	(2.2)	(8.2)	—	—	15.8%
1H FY 2024	$10,828	—	(311)	(1,576)	(10)	10	$8,941
	52.3%	—	(1.5)	(7.6)	—	—	43.2%

A. Consists of amortization of acquisition-related intangible assets, inventory step-up, transaction costs, compensation charges, and other costs

B. Stock-based compensation charge was allocated to cost of goods sold, research and development expense, and sales, general and administrative expense

C. Comprises of legal settlement costs, contributions, restructuring costs and assets held for sale related adjustments

Reconciliation of Non-GAAP to GAAP Financial Measures (contd.)

($ in Millions)	Free Cash Flow	Purchases Related to Property and Equipment and Intangible Assets	Principal Payments on Property and Equipment and Intangible Assets	Net Cash Provided by Operating Activities
FY 2019	$3,143	600	—	$3,743
FY 2020	$4,272	489	—	$4,761
FY 2021	$4,677	1,128	17	$5,822
FY 2022	$8,049	976	83	$9,108
FY 2023	$3,750	1,833	58	$5,641
1H FY 2023	$2,171	794	36	$3,001
1H FY 2024	$8,691	537	31	$9,259

Reconciliation of Non-GAAP to GAAP Financial Measures

($ in Millions)	Q3 FY24 Outlook
Non-GAAP gross margin	72.5%
Impact of stock-based compensation expense, acquisition-related costs, and other costs	(1.0%)
GAAP gross margin	71.5%
Non-GAAP operating expenses	$2,000
Impact of stock-based compensation expense, acquisition-related costs, and other costs	950
GAAP operating expenses	$2,950

NVIDIA®