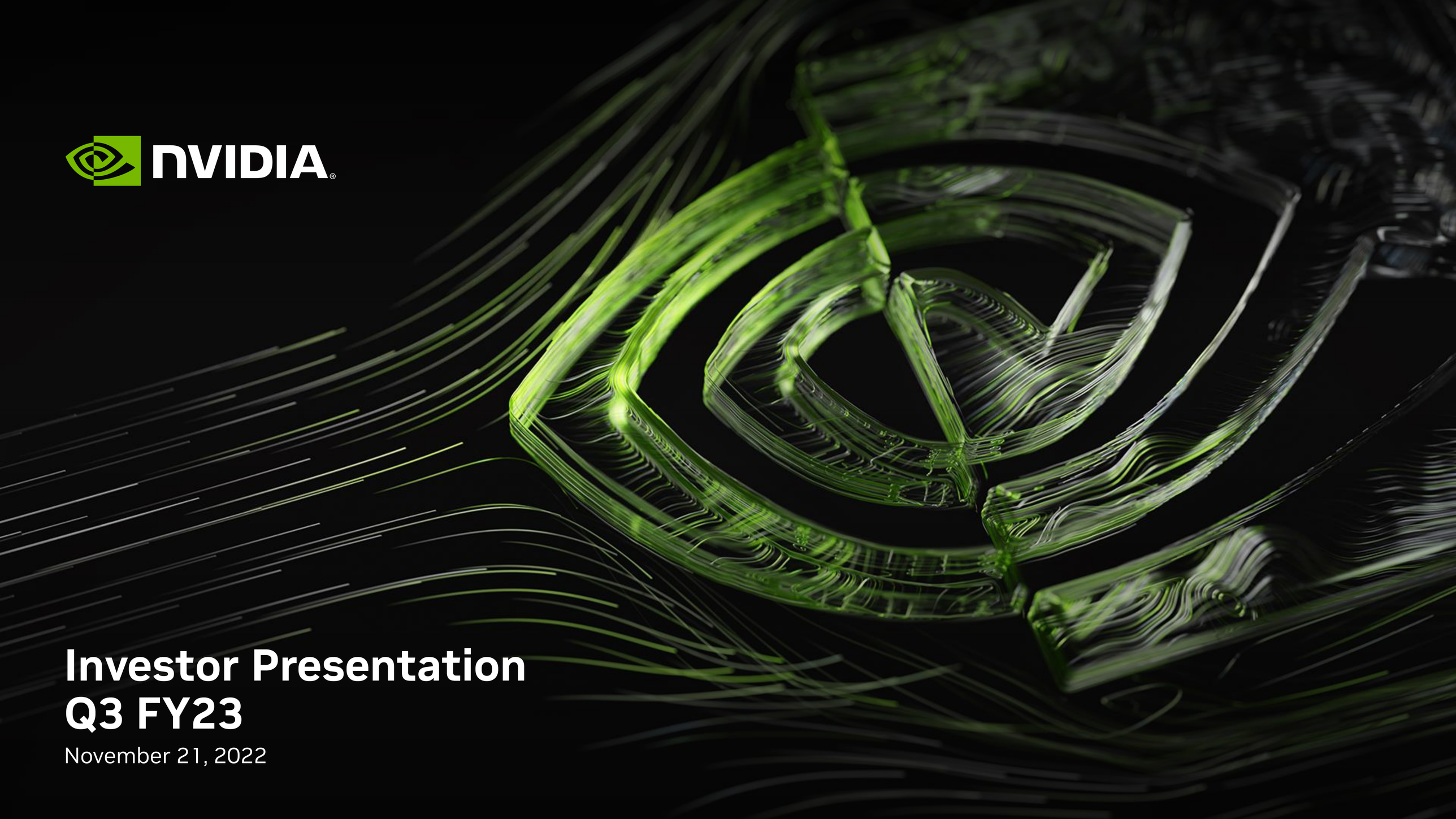


Investor Presentation
Q3 FY23

November 21, 2022

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 and growth drivers; Gaming channel inventories being on track to approach normal levels exiting Q4; our expectation that Data Center demand in China will broadly remain soft into the current quarter; Automotive having great momentum and being our next multi-billion dollar platform; our long-term Professional Visualization opportunity fueled by AI, simulation, and computationally intensive design and engineering workloads; the continued scaling of our customers' DRIVE Orin-based production ramps; our financial outlook, our expected tax rates and our expected capital expenditures for the fourth quarter of fiscal 2023; the benefits, impact, performance and availability of our products and technologies, including NVIDIA Ada Lovelace architecture, GeForce RTX 40 Series GPUs, NVIDIA H100 GPUs, NVIDIA NeMo LLM Service and NVIDIA BioNeMo LLM Service, NVIDIA Omniverse Computing Systems (OVX), NVIDIA Omniverse Cloud services, NVIDIA DRIVE Thor, Jetson Orin Nano, NVIDIA IGX platform, NVIDIA's acceleration stacks and ecosystems, NVIDIA's AI expertise and scale, NVIDIA Omniverse, NVIDIA DGX A100, NVIDIA AI Enterprise, Bluefield-3 DPU, and Grace CPU Superchip; NVIDIA's partnership with Oracle to add tens of thousands more NVIDIA GPUs to Oracle Cloud Infrastructure (OCI); Meta's next-gen AI platform, Grand Teton, using NVIDIA H100 GPUs; Rescale adopting NVIDIA AI Enterprise and other software; NVIDIA's collaboration with Microsoft to build a cloud AI supercomputer; accelerated computing being needed to tackle the most impactful opportunities of our time; expected TOPS processing; AI as the greatest technology force of our time; data centers across industries becoming AI factories; digital robots, avatars and physical robots perceiving, planning and acting; NVIDIA's value to every stakeholder in the ecosystem; the cost and time-to-solution savings of application speed-ups; our remaining repurchase authorization and dividend program plan; upcoming launches of our Data Center products; our Automotive design win pipeline, ramp and production expectations; LLMs being widely viewed as the most important AI models today; NVIDIA accelerated computing being broadly recognized as the way to advance computing as Moore's law ends; the next wave of AI being robotics; building and operating Metaverse applications being the next wave; and our plan for 100% of our global electricity usage for our offices and data centers to be renewable by 2025 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.

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

- Q3 FY23 Earnings Summary
- Key Announcements This Quarter
- NVIDIA Overview
- Financials
- Reconciliation of Non-GAAP to GAAP Financial Measures

Q3 FY23
Earnings Summary

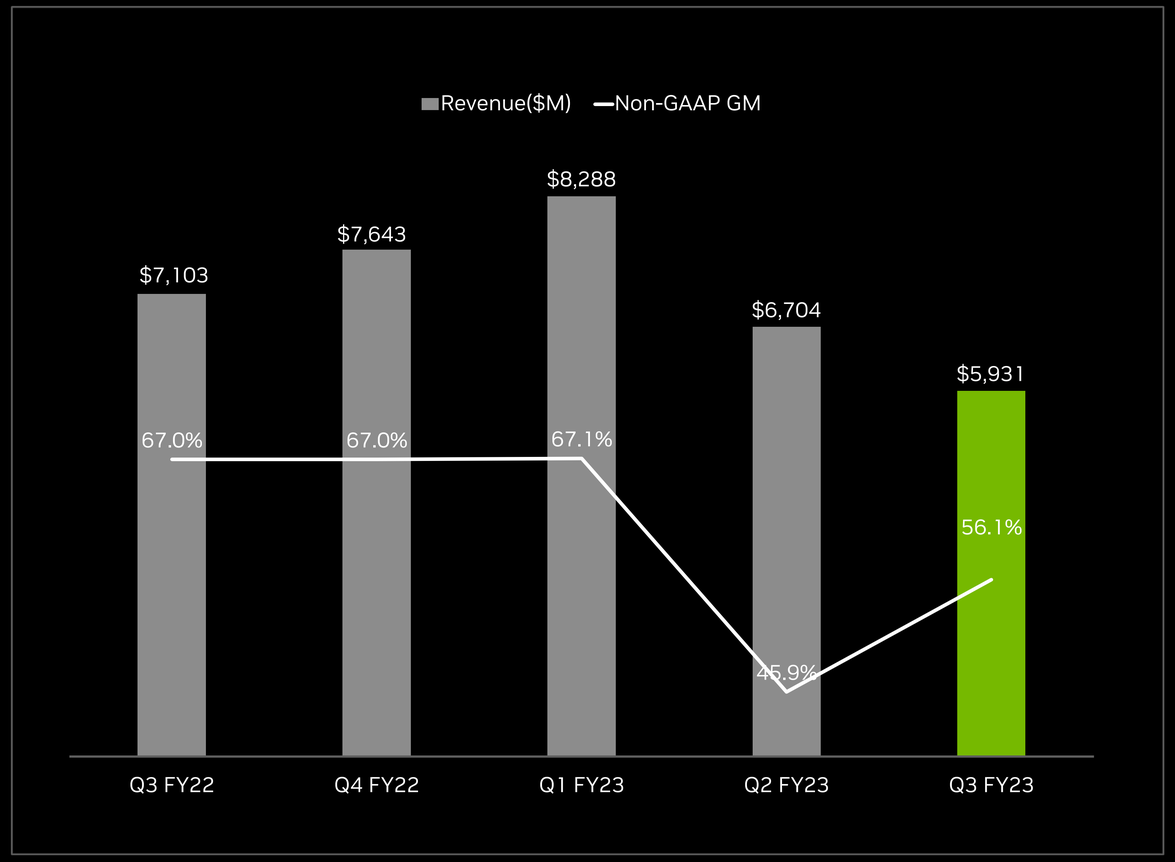
Highlights

- **Record Data Center and Automotive revenue; Gaming and Pro Viz revenue declines on channel inventory corrections and challenging external conditions**
  - Total revenue down 17% Y/Y to $5.93B, inline with outlook of $5.90B +/- 2%
  - Data Center up 31% Y/Y to $3.83B
  - Gaming down 51% Y/Y to $1.57B

- **Data Center posted very solid performance in the face of macroeconomic challenges, export controls and lingering supply chain disruptions**
  - Y/Y growth was driven primarily by leading U.S. cloud providers and a broadening set of consumer internet companies
  - Networking posted strong growth driven by hyperscale customers and easing supply constraints
  - Sequential growth impacted by China regulatory and macro headwinds

- **Gaming decline on continued inventory correction; believe channel inventory on track to approach normal levels exiting Q4**
  - Sell-through for gaming products was relatively solid in the Americas and EMEA, but softer in Asia-Pac
  - Macroeconomic conditions and COVID lockdowns in China continued to weigh on consumer demand
  - New Ada Lovelace GPU architecture had an exceptional launch; sold out quickly in many locations

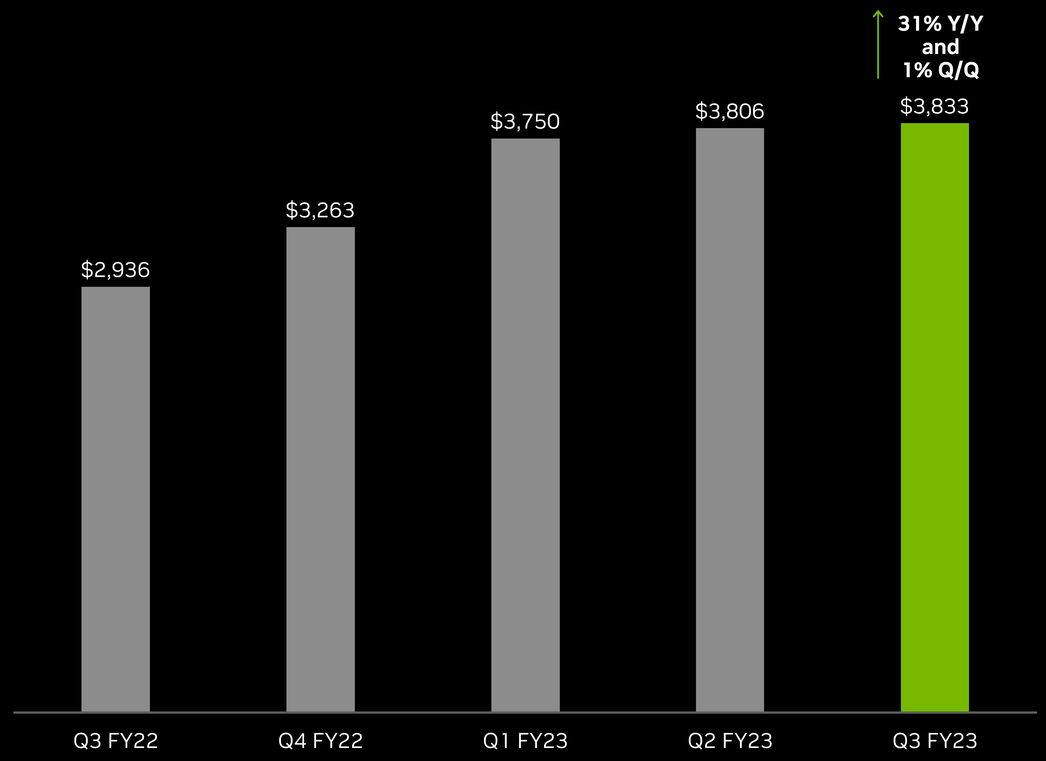
Q3 FY23 Financial Summary



	GAAP			Non-GAAP		
	Q3 FY23	Y/Y	Q/Q	Q3 FY23	Y/Y	Q/Q
Revenue	$5,931	-17%	-12%	$5,931	-17%	-12%
Gross Margin	53.6%	-11.6 pts	+10.1 pts	56.1%	-10.9 pts	+10.2 pts
Operating Income	$601	-77%	+20%	$1,536	-55%	+16%
Net Income	$680	-72%	+4%	$1,456	-51%	+13%
Diluted EPS	$0.27	-72%	+4%	$0.58	-50%	+14%
Cash Flow from Ops	$392	-74%	-69%	$392	-74%	-69%

All dollar figures are in millions other than EPS.

Data Center



Revenue ($M)

Highlights

- Very solid performance in the face of macroeconomic challenges, new export controls, and lingering supply chain disruptions
- Year-on-year, growth was driven primarily by leading US cloud providers and a broadening set of consumer internet companies for workloads such as large language models, recommendation systems and generative AI drove growth
- Other vertical industries, such as automotive and energy, also contributed to growth with key workloads relating to autonomous driving, high performance computing,, simulations and analytics
- Demand in China remains soft, and we expect that to continue in the current quarter
- Started shipping flagship H100 data center GPU based on the new Hopper architecture

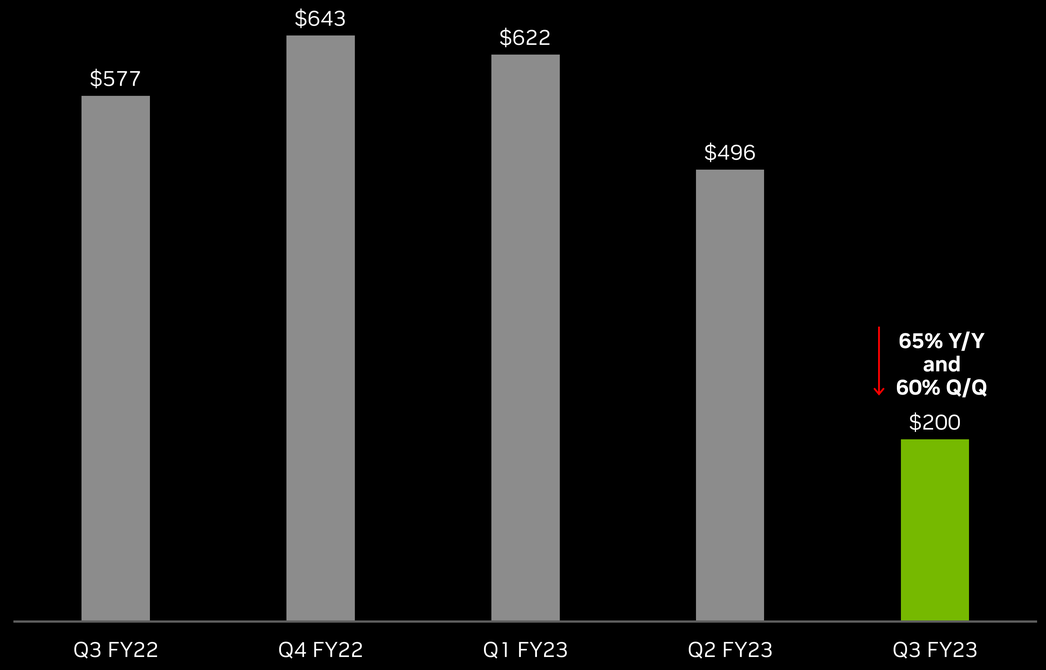
Gaming

Revenue ($M)

Highlights

- Decline reflects lower sell-in to partners to help align channel inventory levels with current demand expectations
- We believe channel inventories are on track to approach normal levels as we exit Q4
- Sell-through was relatively solid in the Americas and EMEA, but softer in Asia Pac as macroeconomic conditions and COVID lockdowns in China continued to weigh on consumer demand.
- The Ada Lovelace GPU architecture launched to tremendous demand and positive feedback
  - GeForce RTX 4090 sold out quickly in many locations
  - GeForce RTX 4080 now available
- Total number of RTX games and applications now exceeds 350
- Added over 85 games to the GeForce NOW library, bringing the total to over 1,400

Professional Visualization

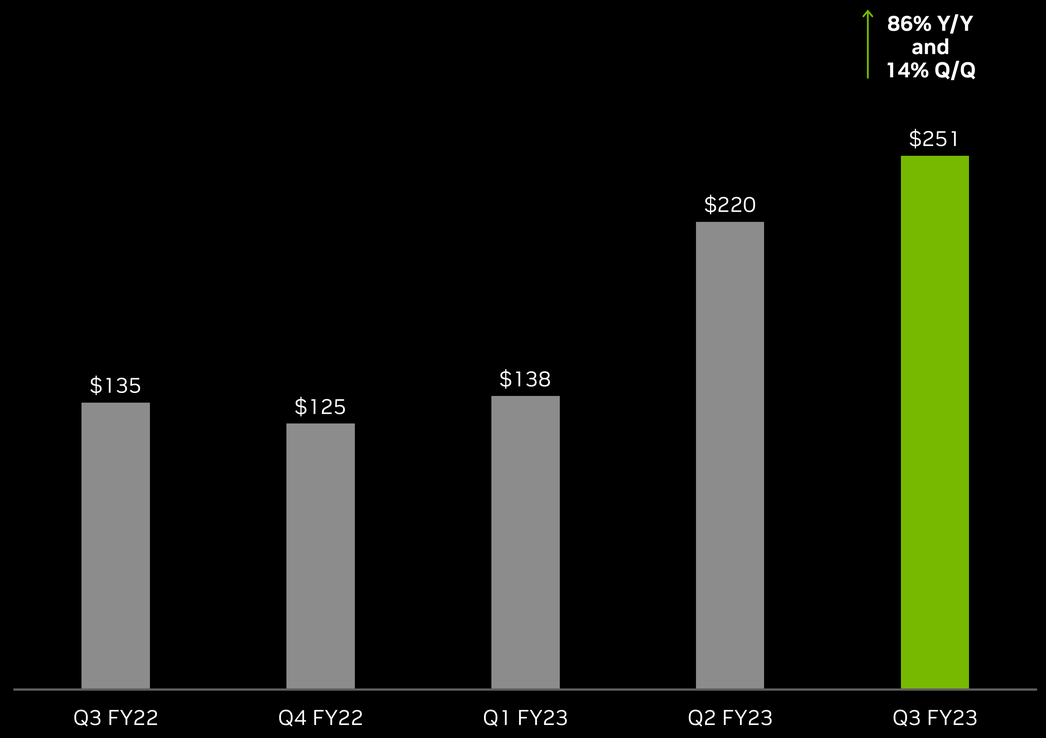


Revenue ($M)

Highlights

- Decline reflects lower sell-in to partners to help align channel inventory levels with current demand expectations
- Despite near-term challenges, long-term opportunity remains intact, fueled by AI, simulation, and computationally intensive design and engineering workloads
- Leaders in some of the world's largest industries continue to adopt Omniverse, including Lowe's, Charter Communications with HEAVY.AI, and Deutsche Bahn

Automotive

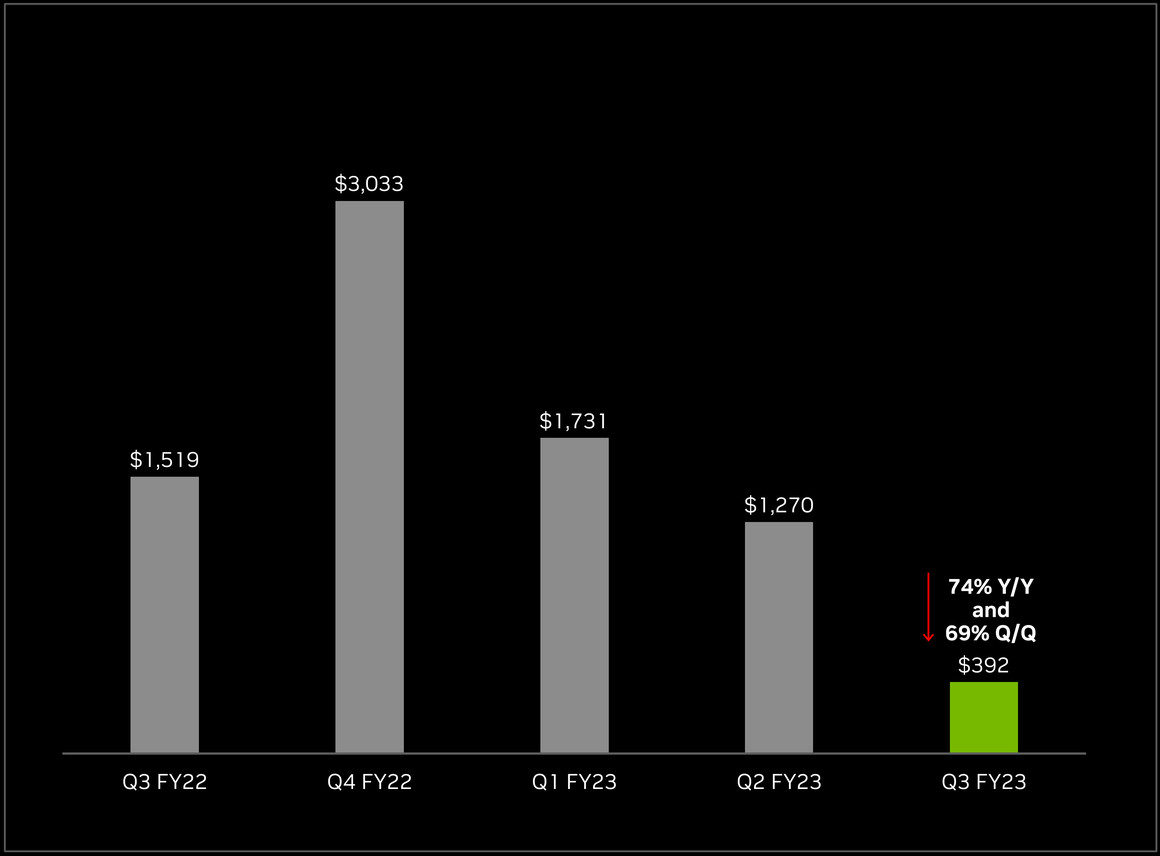


Revenue ($M)

Highlights

- Growth was driven by an increase in AI Automotive Solutions, as our customers' DRIVE Orin-based production continues to ramp
- Volvo Cars unveiled the all-new flagship Volvo EX90 SUV powered by the NVIDIA DRIVE platform
- Other recently announced design wins and new model introductions include Hozon Auto, NIO, Polestar, and XPENG

Sources & Uses of Cash



Cash Flow from Operations ($M)

Highlights

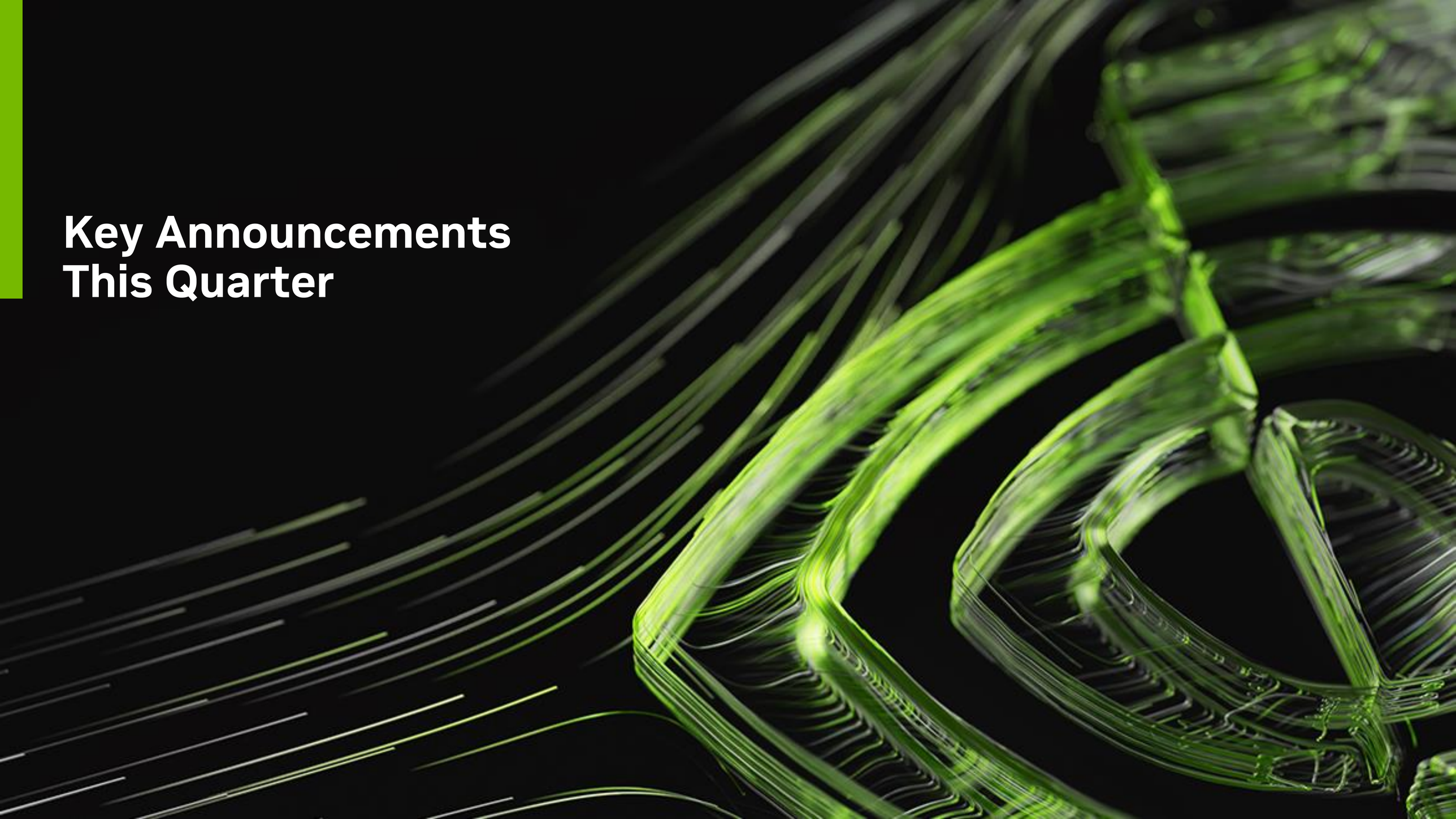
- Y/Y decrease reflects lower operating income, timing of supplier payments and inventory deliveries, partially offset by lower supplier prepayments
- Q/Q decrease reflects timing of supplier and other payments as well as inventory deliveries, partially offset by lower cash tax payments
- Returned $3.75 billion to shareholders in the form of share repurchases and cash dividends
- Invested $548M in capex (includes principal payments on PP&E)
- Ended the quarter with $13.1B in gross cash and $11.0B in debt; $2.1B in net cash

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.

Q4 FY23 Outlook

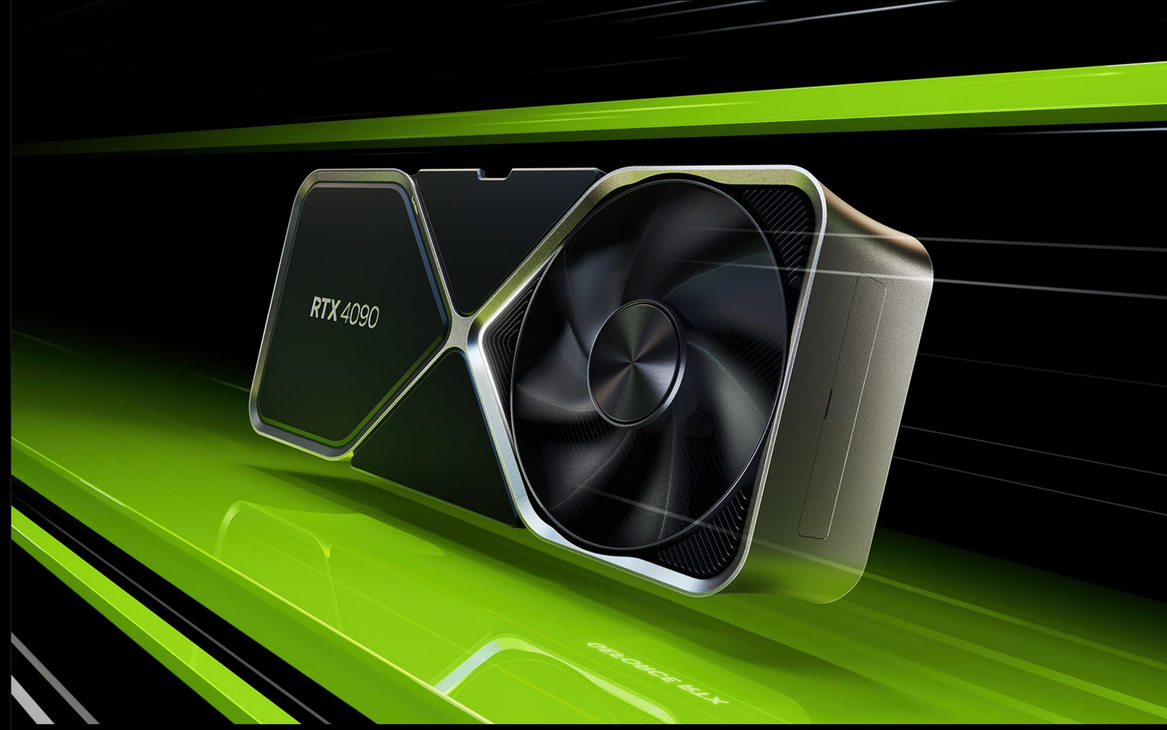
Revenue	**$6.0 billion**, plus or minus 2% Expect modest sequential growth, driven by Automotive, Gaming and Data Center
Gross Margins	**63.2%** GAAP and **66.0%** non-GAAP, plus or minus 50 basis points
Operating Expense	Approximately **$2.56 billion** GAAP and **$1.78 billion** non-GAAP
Other Income & Expense	Net income of approximately **$40 million** for GAAP and non-GAAP Excluding gains and losses on non-affiliated investments
Tax Rate	**9.0%** GAAP and non-GAAP, plus or minus 1%, excluding discrete items
Capital Expenditures	Approximately **$500 million** to **$550 million**

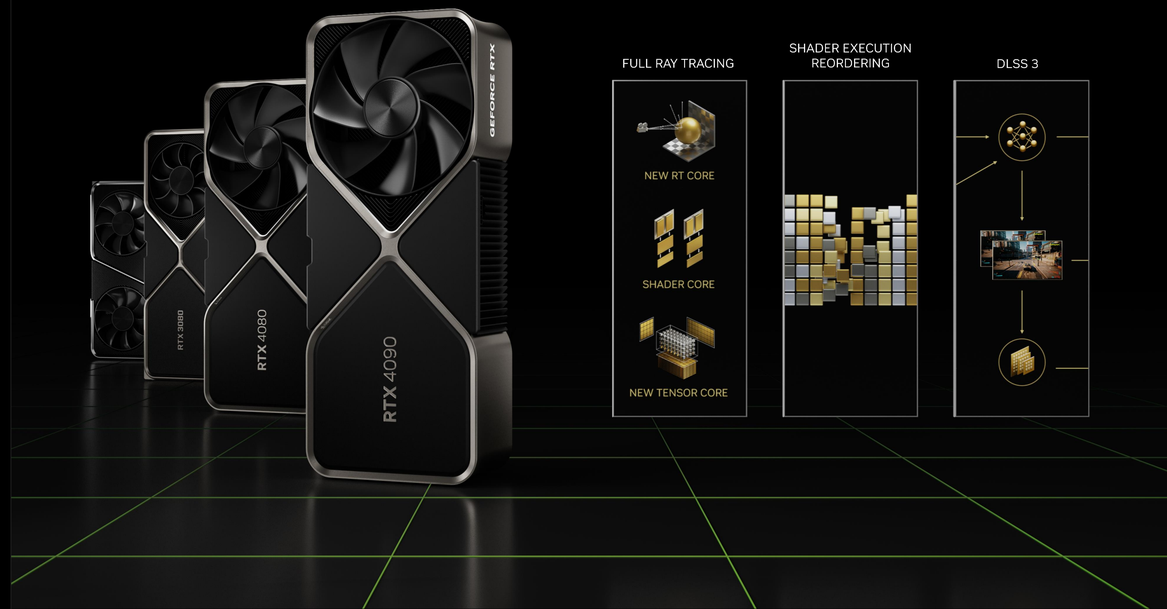
Key Announcements This Quarter



NVIDIA GeForce RTX 40 Series GPU Launch

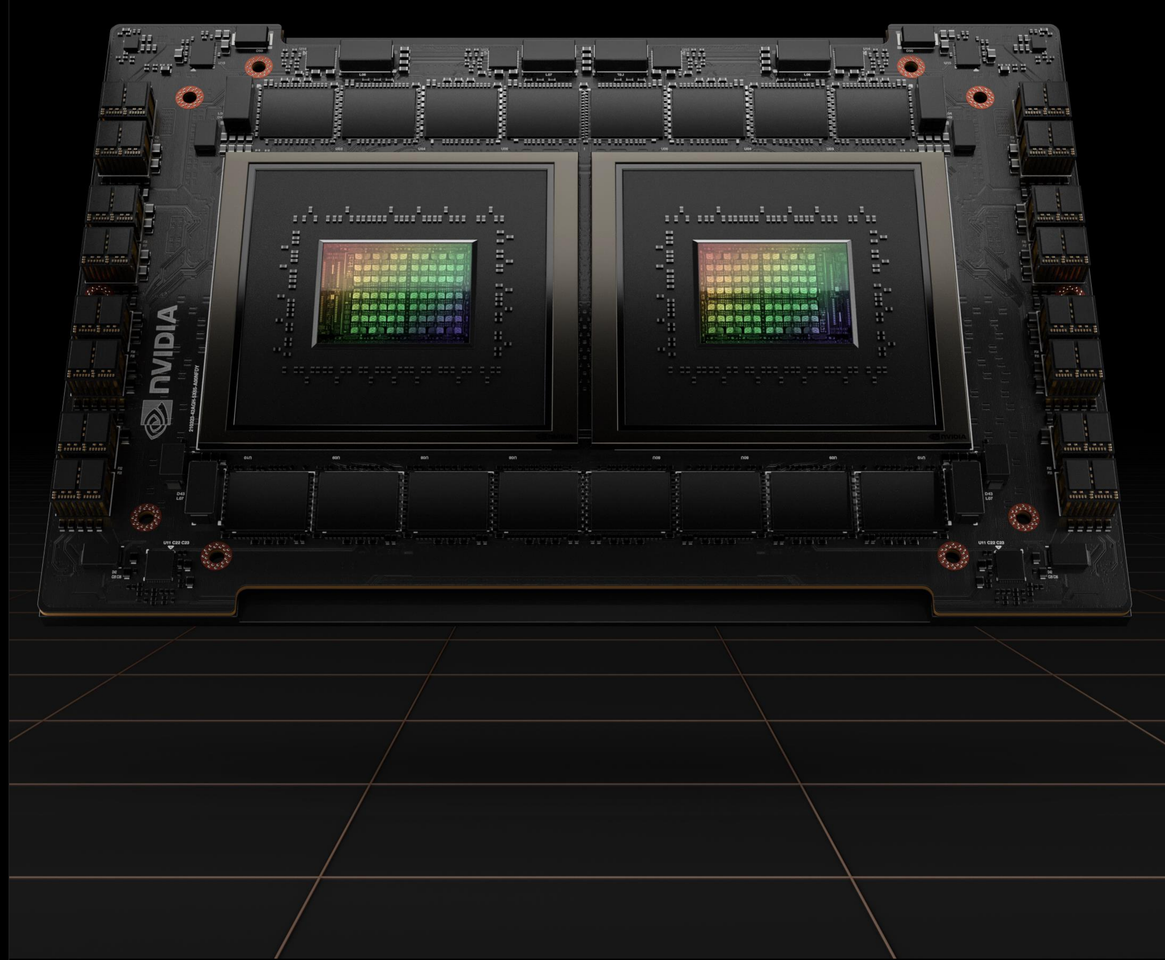
- Based on the new NVIDIA Ada Lovelace architecture, GeForce 40 Series GPUs deliver up to 4x the performance of the previous generation
- Features DLSS 3, the neural rendering technology that uses AI to generate entire frames for faster gameplay
- The initial line-up includes two high-end GPUs:
  - RTX 4090 24GB is up to 4x faster compared to RTX 3090 Ti; began selling last quarter, starting at $1,599
  - RTX 4080 16GB is 2x as fast as the RTX 3080 Ti and more performance than RTX 3090 Ti at lower power; available now, starting at $1,199





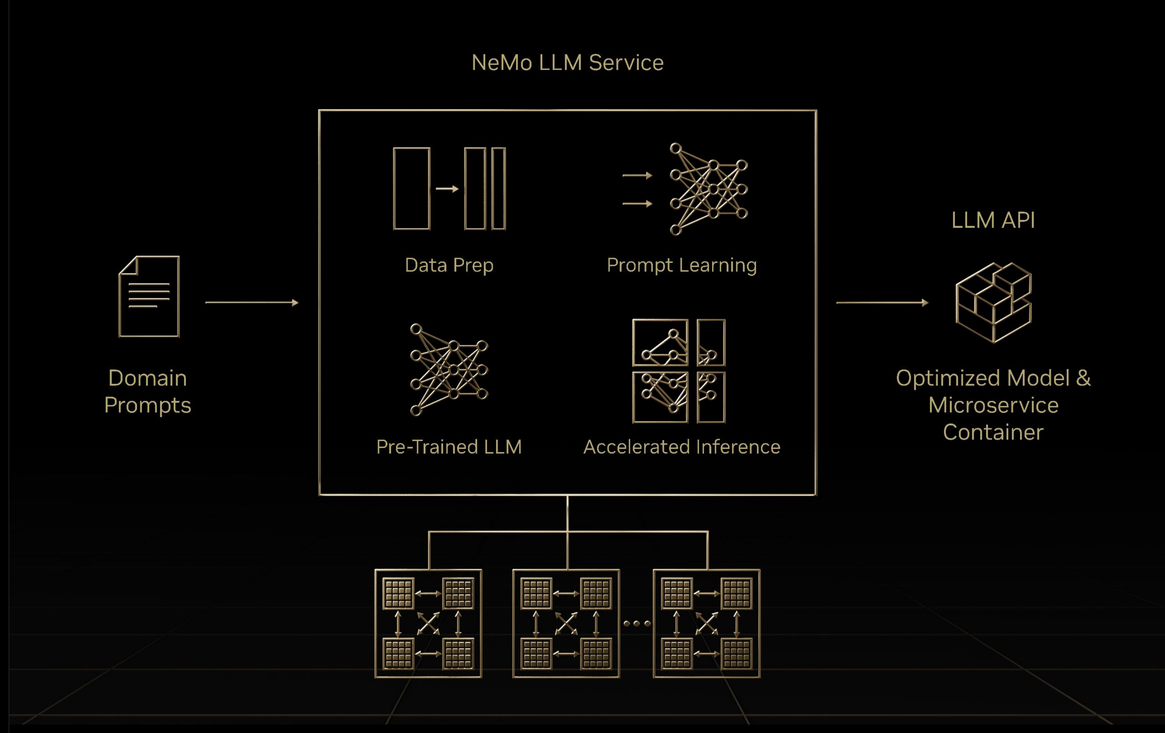
NVIDIA H100 in Full Production

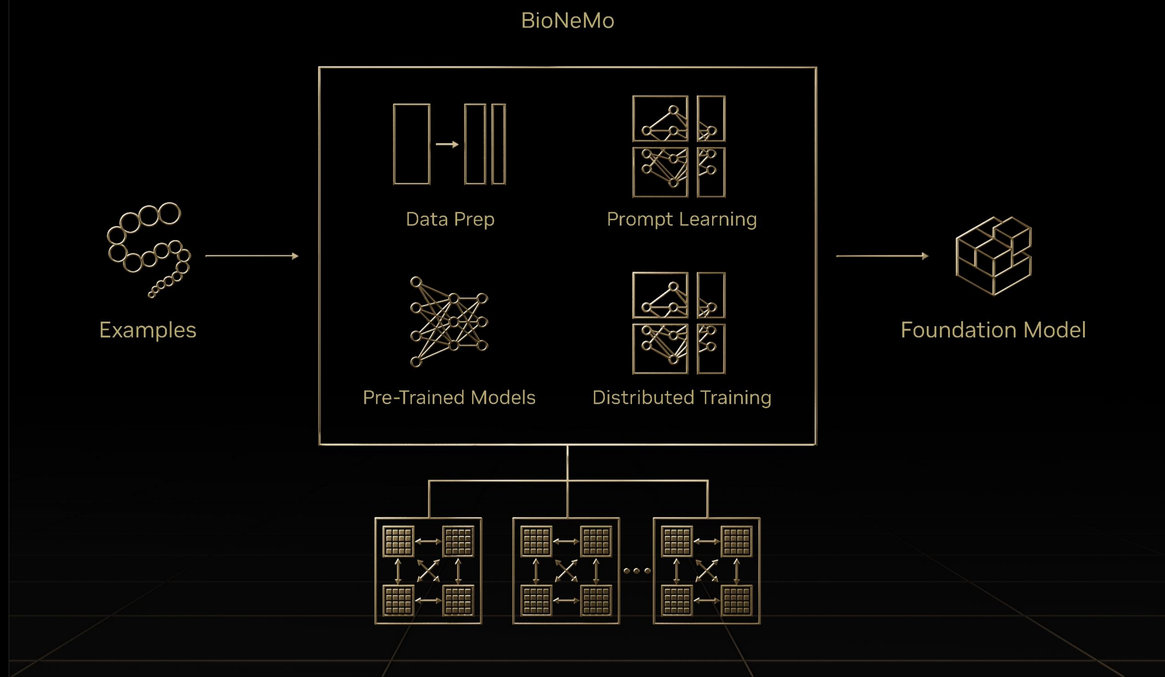
- NVIDIA H100 is in full production, with global tech partners planning to roll out the first wave of products and services in Q4
- H100 delivers the same AI performance with 3.5x better energy efficiency and 3x lower total cost of ownership, while using 5x fewer server nodes, over the previous generation
- The NVIDIA AI Enterprise software suite is now included with H100 enterprise servers
  - Optimized for development and deployment of AI and give enterprises access to NVIDIA's AI application frameworks and tools needed to build customer support chatbots, recommender systems, IoT edge AI and more
- H100 systems will be available from Atos, Cisco, Dell Technologies, Fujitsu, GIGABYTE, Hewlett Packard Enterprise, Lenovo and Supermicro
- The first H100-based cloud systems will be deployed on Amazon Web Services, Google Cloud, Microsoft Azure and Oracle Cloud Infrastructure early next year



Large Language Model (LLM) Cloud Services to Advance AI and Digital Biology

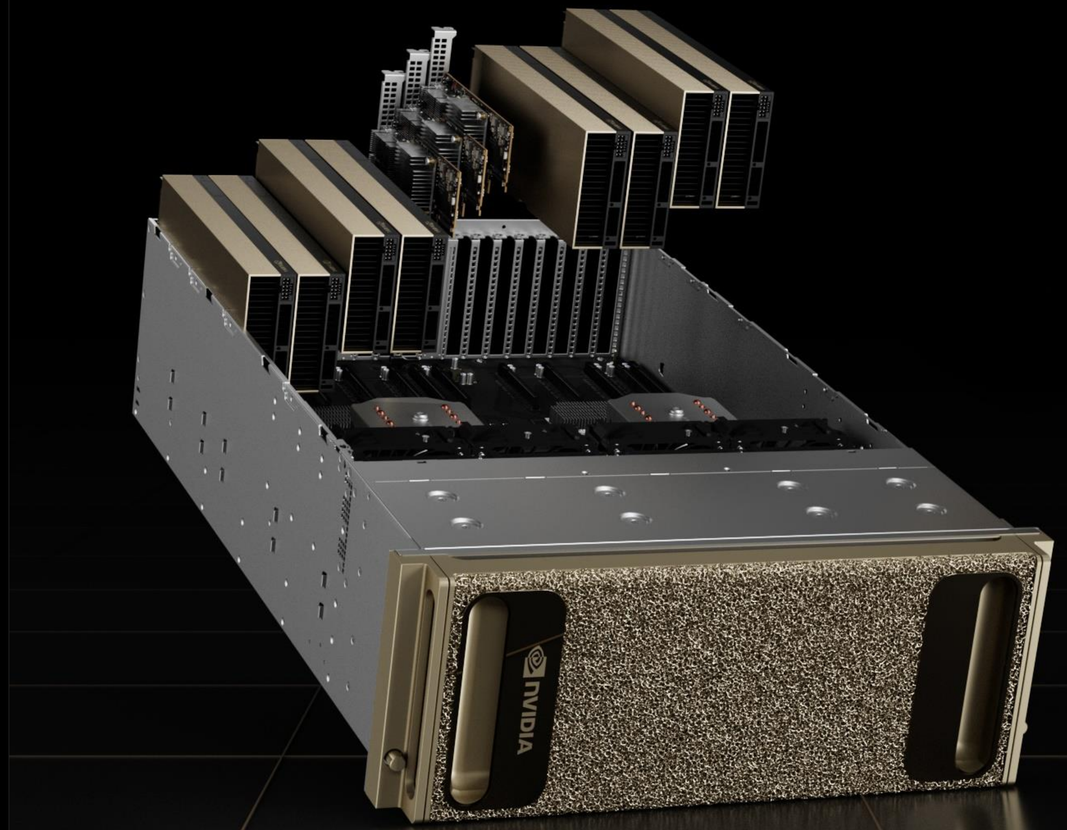
- The NVIDIA NeMo LLM Service and the NVIDIA BioNeMo LLM Service enable developers to easily adapt LLMs and deploy customized AI applications
  - Applications include content generation, text summarization, chatbots, code development, protein structure and biomolecular property predictions
- Based on the Transformer architecture, LLMs can learn to represent the language of humans, chemistry, or biology, without supervision or labeled datasets
- LLMs are widely viewed as the most important AI models today
- Early Access availability soon





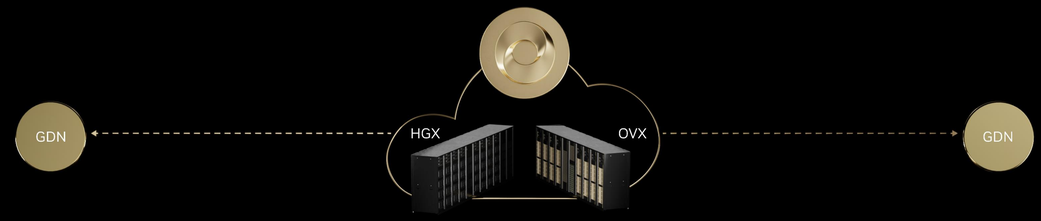
NVIDIA Omniverse Computing Systems (OVX)

- NVIDIA OVX systems—the graphics and simulation foundation for the metaverse—are designed to build 3D virtual worlds and to operate immersive digital twin simulations in NVIDIA Omniverse Enterprise
- NVIDIA OVX is powered by NVIDIA L40 GPU based on Ada Lovelace and NVIDIA CX7 400Gb/s NIC
  - OVX servers can be deployed in NVIDIA OVX POD and SuperPOD configurations with the NVIDIA Spectrum-3 Ethernet platform



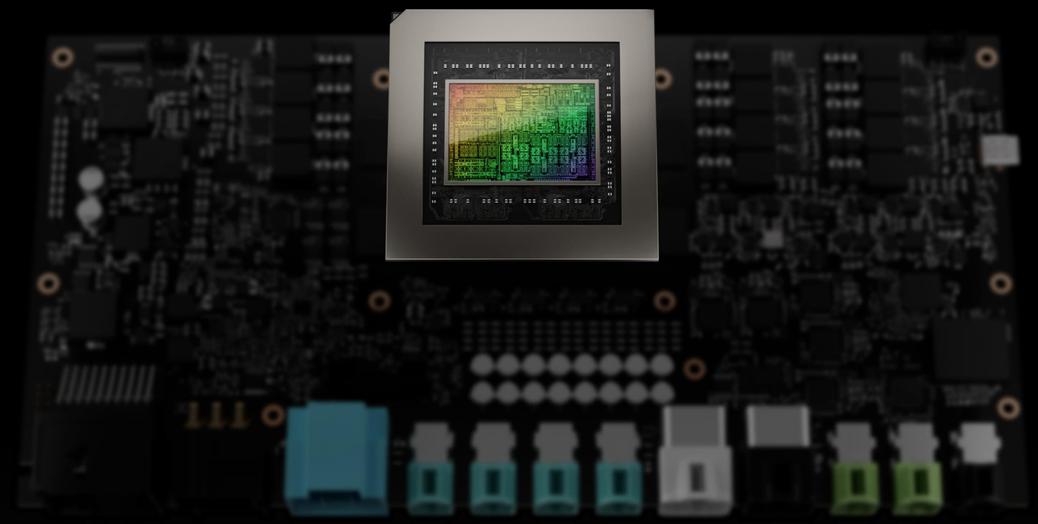
NVIDIA Omniverse Cloud Services

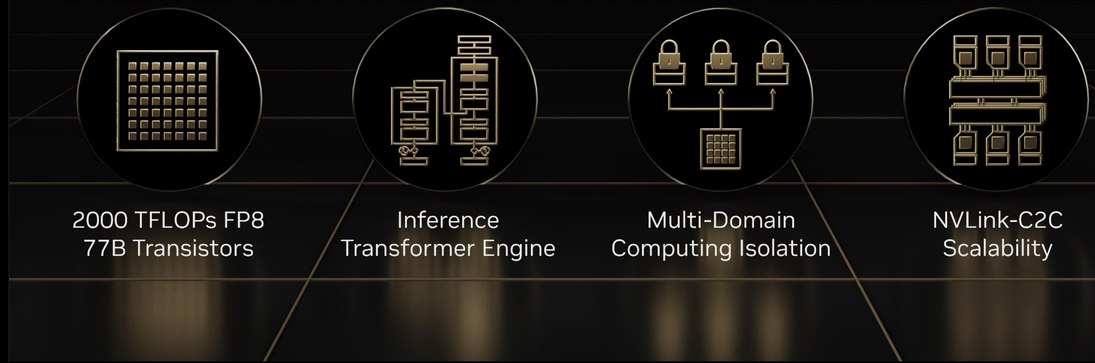
- NVIDIA's first software- and infrastructure-as-a-service offering
- A suite of cloud services for artists, developers and enterprise teams to design, publish, and operate metaverse applications anywhere
- Omniverse Cloud services run on the Omniverse OVX for graphics and physics simulation, NVIDIA HGX for AI workloads and the NVIDIA Graphics Delivery Network
- Services in Omniverse Cloud – including Omniverse Farm, Replicator and Isaac Sim – are available as containers on NVIDIA NGC for self-service deployment on AWS using Amazon EC2 G5 instances featuring NVIDIA A10G GPUs
- NVIDIA Omniverse Cloud managed services are available via early access by application
- Early customers of Omniverse Cloud include RIMAC Group, WPP and Siemens



NVIDIA DRIVE Thor Superchip

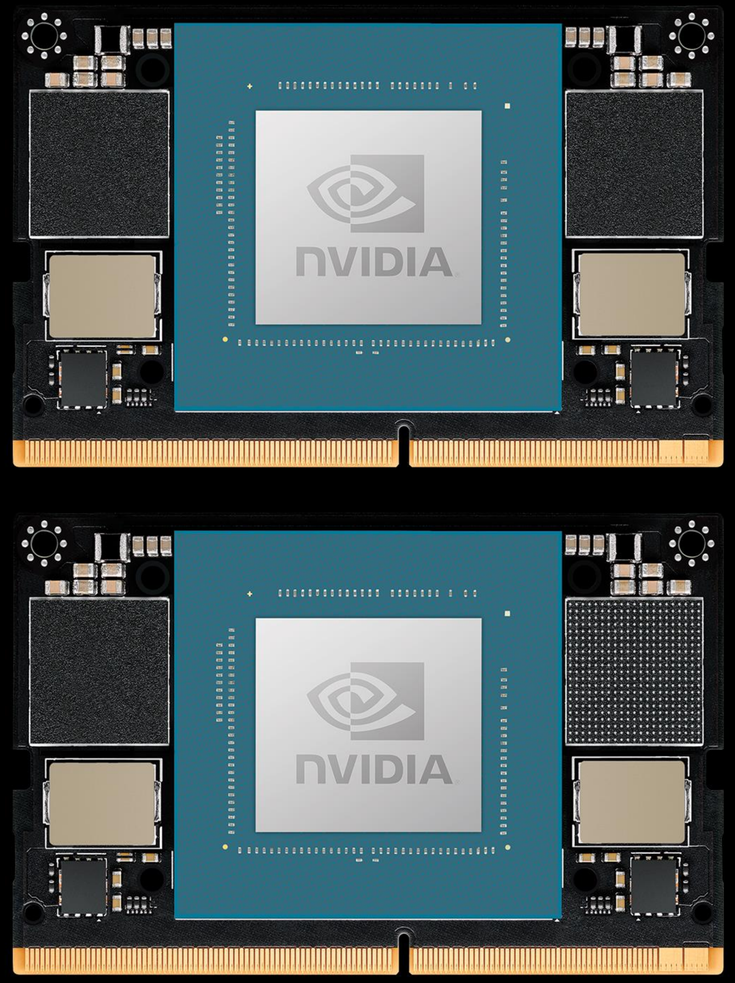
- DRIVE Thor, with 2,000 teraflops of performance, unifies separate functions into a single architecture for greater efficiency and lower overall system cost. Functions include:
  - Automated and assisted driving
  - Parking
  - Driver and occupant monitoring
  - Digital instrument cluster
  - In-vehicle infotainment (IVI) and rear-seat entertainment
- Includes AI capabilities introduced in the NVIDIA Hopper Multi-Instance GPU architecture, along with the NVIDIA Grace CPU and Ada Lovelace GPU
- Chosen by Geely-owned automaker, ZEEKR, for its next-generation intelligent EVs for early 2025
- Available for automakers' 2025 models





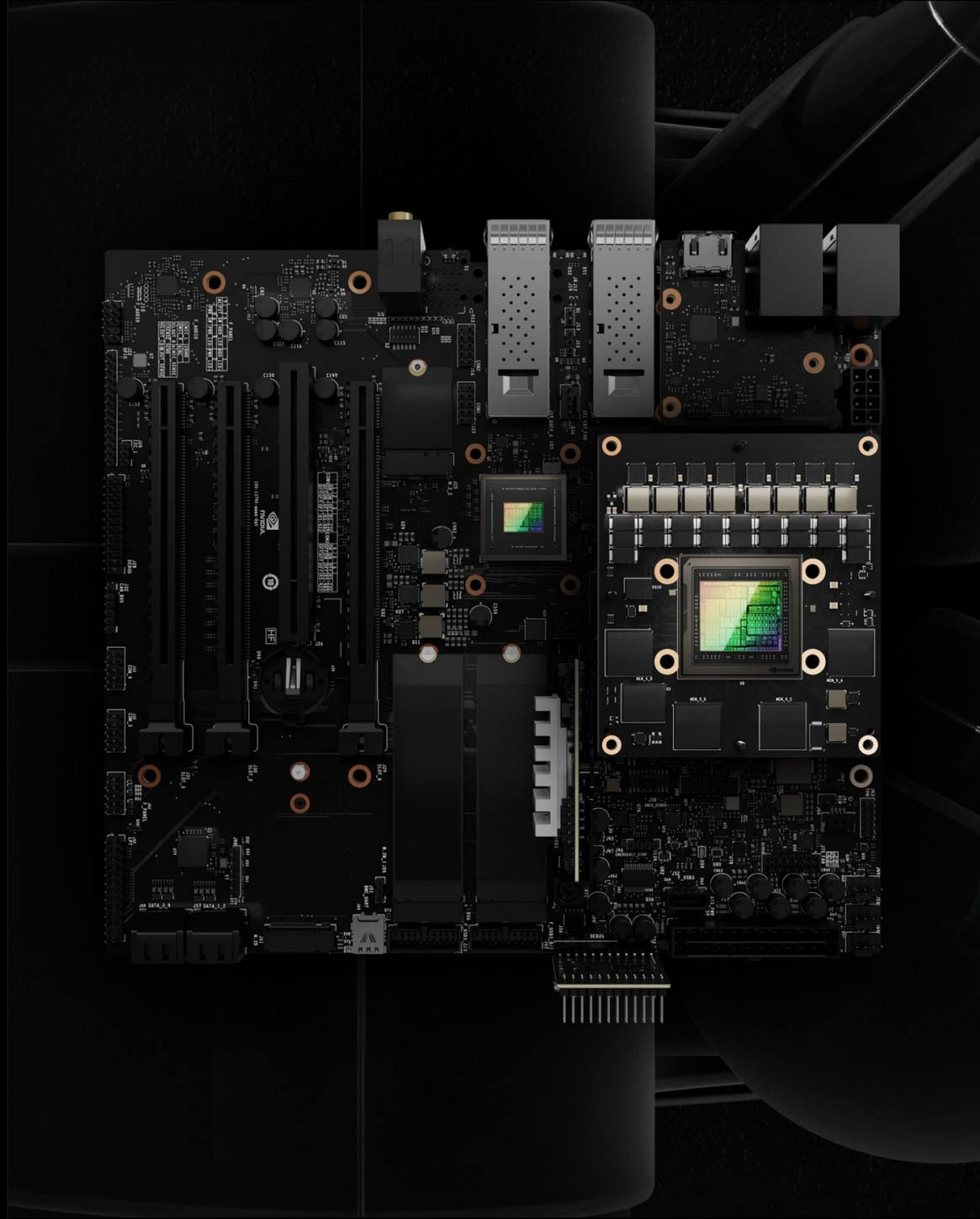
Jetson Orin Nano for Robotics

- The new Jetson Orin Nano system-on-module delivers up to 40 TOPS, or up to 80x the performance of the prior generation Jetson Nano, setting a new standard for entry-level edge AI and robotics
- NVIDIA's Jetson ecosystem includes 1 million developers, 6,000 customers and 150 partners
- Canon, John Deere, Microsoft and Teradyne among the many companies building robots with Jetson
- Available in January 2023 starting at $199



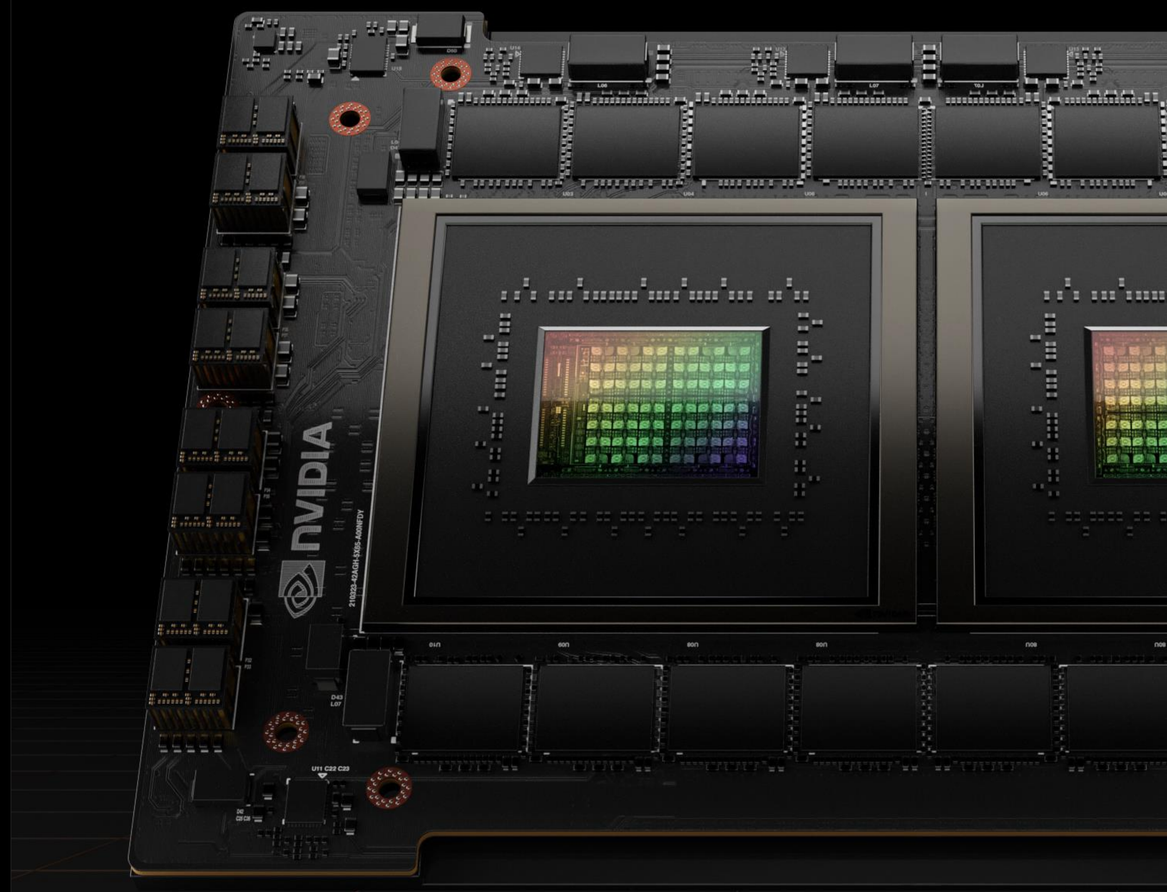
IGX Edge AI Platform for Safe, Secure Autonomous Systems

- The NVIDIA IGX platform brings advanced security and proactive safety to sensitive industries such as manufacturing, logistics and healthcare
- Includes NVIDIA IGX Orin, a compact and energy-efficient AI system, and can run NVIDIA AI Enterprise and NVIDIA Fleet Command software
  - NVIDIA Fleet Command allows organizations to deploy secure, over-the-air software and system updates from a central cloud console
- Embedded-computing manufacturers to create products based on IGX include ADLINK, Advantech, Dedicated Computing, Kontron, Leadtek, MBX, Onyx, Portwell, Prodrive Technologies and YUAN
- Siemens will use IGX at the industrial edge



Oracle Adding Tens of Thousands More NVIDIA GPUs in OCI, Partnering to Speed AI Adoption

- Oracle Cloud Infrastructure (OCI) is adding tens of thousands more NVIDIA GPUs, including the A100 and upcoming H100
  - Provides enterprises a broad, easily accessible portfolio of options for AI training and inference at scale
- The partnership aims to bring the full NVIDIA accelerated computing stack—from GPUs to systems to software—to OCI
- Puts NVIDIA AI within easy reach for a broad range of industries and thousands of companies
  - NVIDIA AI Enterprise for secure and scalable end-to-end AI development and deployment
  - NVIDIA RAPIDS acceleration for Apache Spark on the OCI Data Flow managed service
  - NVIDIA Clara for medical imaging, genomics, natural language processing and drug discovery



Meta’s Next-Gen AI Platform, Grand Teton, Brings NVIDIA Hopper to Its Data Centers

- Grand Teton uses NVIDIA H100 GPUs and offers 2x the network bandwidth and 4x the bandwidth between host processors and GPU accelerators compared to Meta’s prior Zion system
- The added network bandwidth enables Meta to create larger clusters of systems for training AI models
- NVIDIA H100 GPUs, when connected with NVIDIA networking across thousands of servers in hyperscale data centers, can be 300X more energy efficient than CPU-only servers
- With Meta sharing Grand Teton, system builders will have access to an open design for H100-powered hyperscale data center compute infrastructure

Reinforcing NVIDIA's Leadership in Latest MLPerf Training Benchmark

What is MLPerf?

- The industry's first objective AI benchmark for measuring machine learning performance
- Consortium of over 70 universities and companies, including Google, Intel, Baidu, and NVIDIA, founded in 2018
- NVIDIA has consistently delivered leading results and record performances in both MLPerf Training and Inference benchmarks

MLPerf November 2022—AI Training

- Seventh consecutive top showing in training tests NVIDIA H100 GPUs set records across all enterprise AI workloads in the training benchmarks
- NVIDIA AI was not only the winning platform across the board, but also the only platform to run all eight tests
- NVIDIA H100 GPUs delivered up to 6.7X faster performance than prior generation A100 GPUs when they were first submitted
- Today's A100 GPUs are 2.5X more performant, thanks to advances in software and full-stack optimization
- 11 companies, including the Microsoft Azure cloud service, made submissions using NVIDIA A100, A30 and A40 GPUs, demonstrating the reach of the NVIDIA ecosystem

NVIDIA AI Adopted by Rescale HPC-as-a-Service Platform

- Rescale is adopting NVIDIA AI Enterprise and other software to address the industrial scientific community's rising demand for AI in the cloud
- Brings new capabilities to Rescale's HPC-as-a-service offerings, which include simulation and engineering software used by hundreds of customers across industries
- NVIDIA is also accelerating the Rescale Compute Recommendation Engine, which enables customers to identify the right infrastructure options to optimize cost and speed objectives

NVIDIA Teams with Microsoft to Build Massive Cloud AI Computer

- Multi-year collaboration with Microsoft to build one of the most powerful AI supercomputers to help enterprises train, deploy and scale AI, including large, state-of-the-art models
- Microsoft Azure will incorporate the NVIDIA AI stack, adding tens of thousands of A100 and H100 GPUs, Quantum-2 400Gb/s InfiniBand networking and the NVIDIA AI Enterprise software suite to its platform
- NVIDIA will utilize Azure's scalable virtual machine instances to research and further accelerate advances in generative AI, a rapidly emerging area of AI

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 361 of the TOP500 supercomputers, and boasts over 3.5 million developers.

Headquarters: Santa Clara, CA

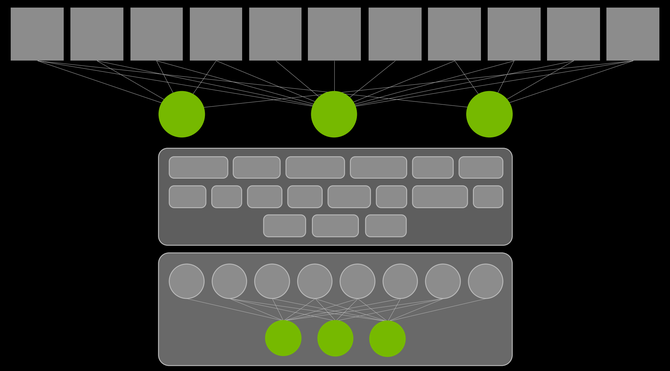
What Is Accelerated Computing?

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.

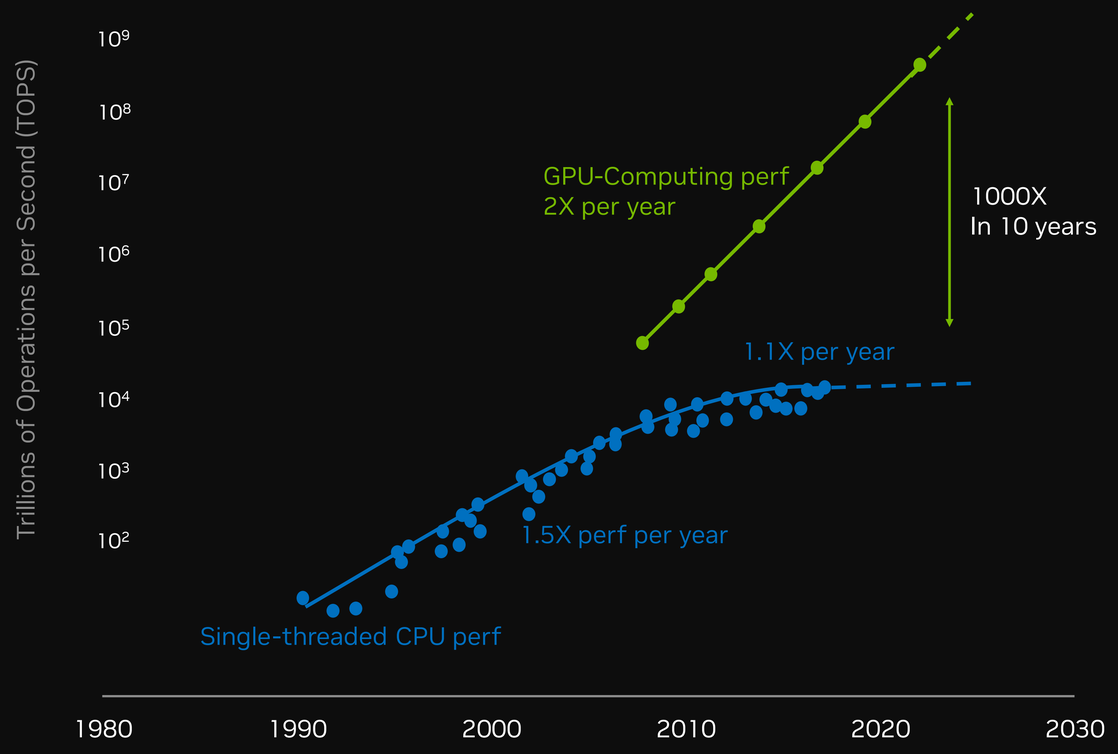
NVIDIA is uniquely dedicated to accelerated computing —working top-to-bottom—refactoring applications and creating new algorithms, and bottom-to-top—inventing new specialized processors, like RT Core and Tensor Core.

"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



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 is learned perception and inference, like recognizing images, understanding speech, recommending a video, or an item to buy.

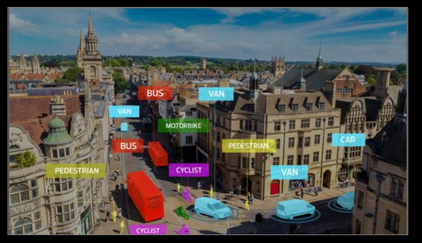
The next wave of AI is robotics—AI planning actions. Digital robots, avatars, and physical robots will perceive, plan and act.

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

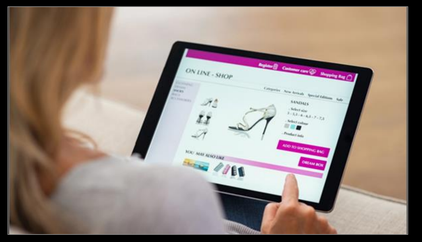
Retail Asset Protection
$94.5B Inventory Loss / Year

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

NVIDIA

Building and Operating Metaverse Applications Is the Next Wave

NVIDIA Omniverse—Runs on NVIDIA OVX servers | RTX workstations | Enterprise software | Cloud services

NVIDIA Omniverse is a software platform for building and operating metaverse applications.

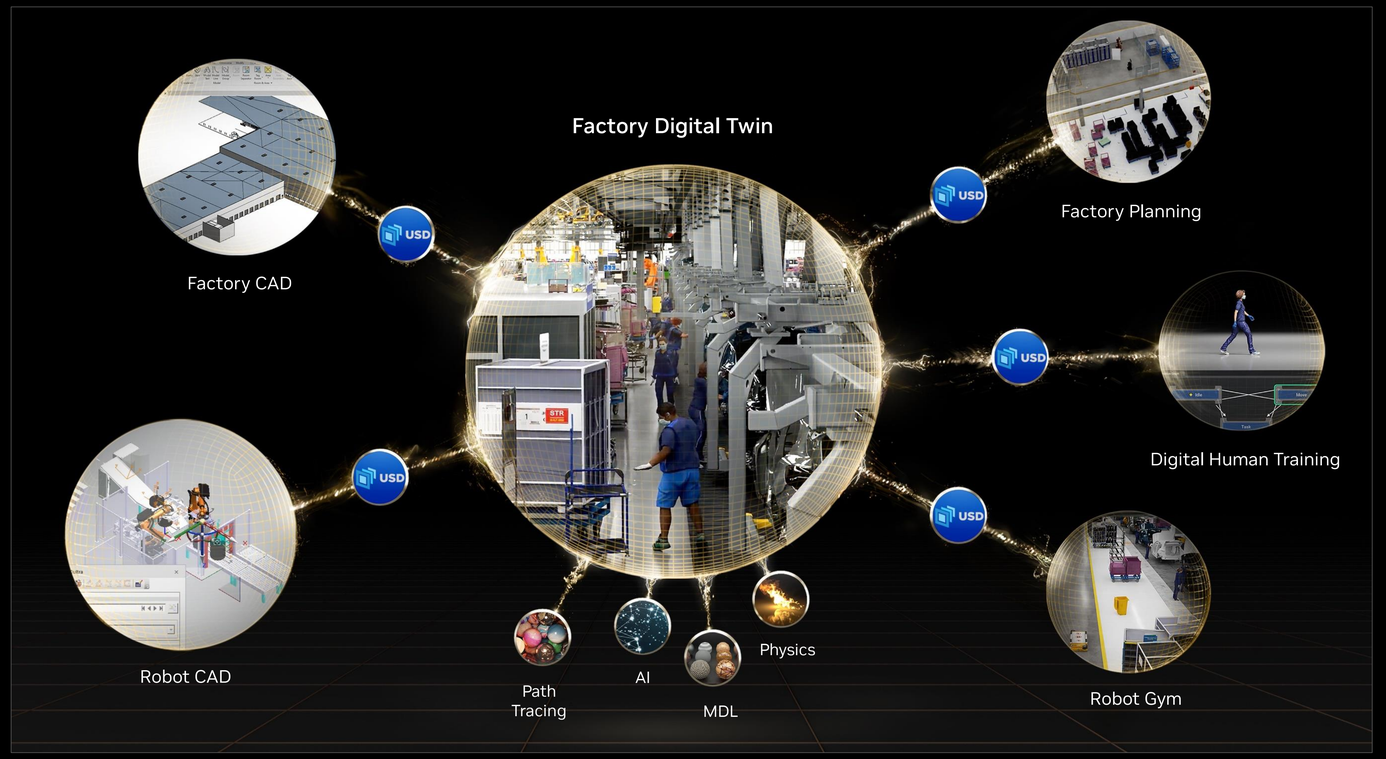
Our initial focus is on industrial metaverses, such as digital twins used to emulate the behavior of products or factories in the physical world.

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.



Product Development

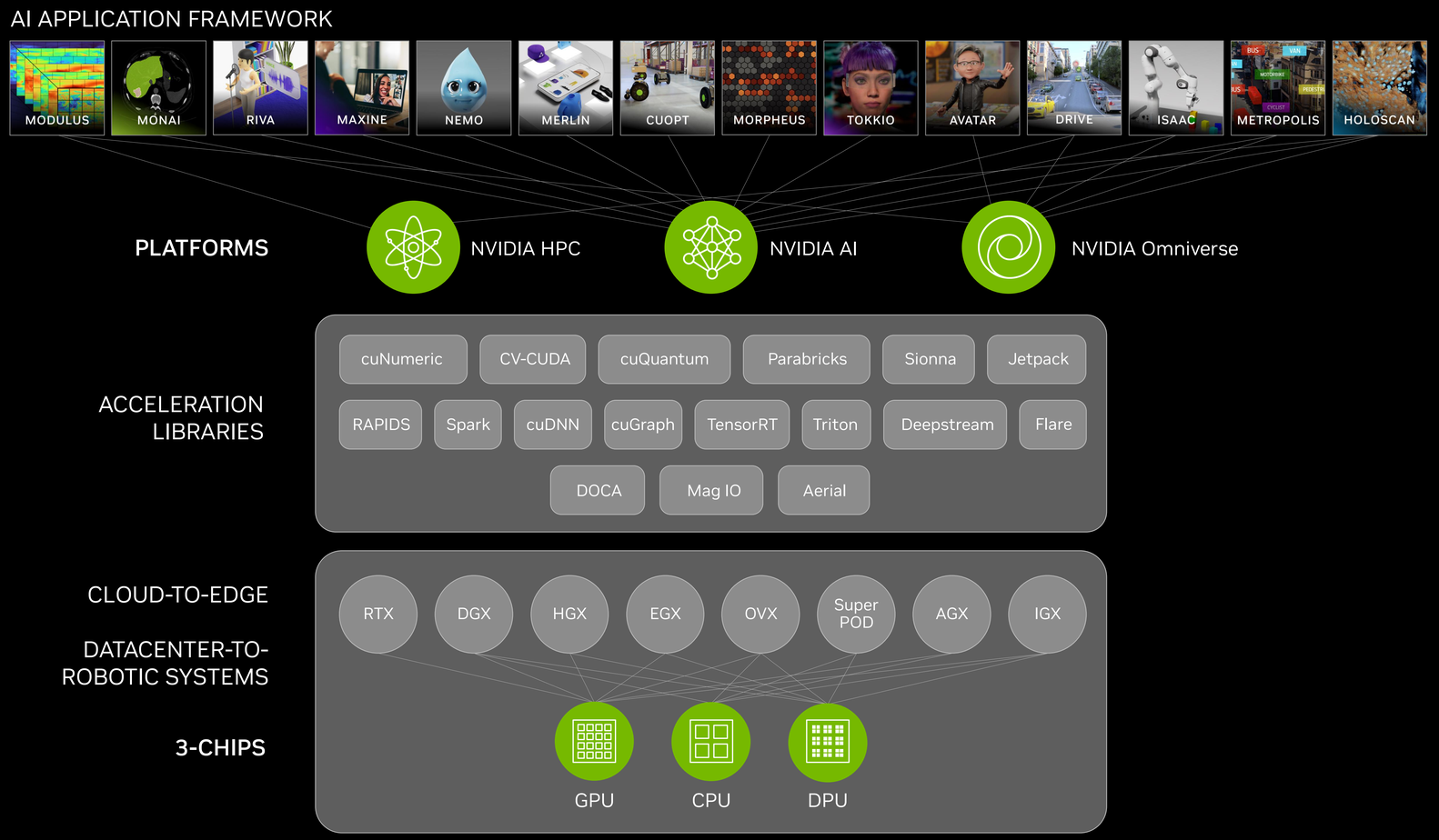
Connecting 3D Creators & AI Assistants in Virtual Worlds

Operate Digital Twin

Connecting Robots in a Virtual World

NVIDIA's Accelerated Computing Platform

Full-stack innovation across silicon, systems and software



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.

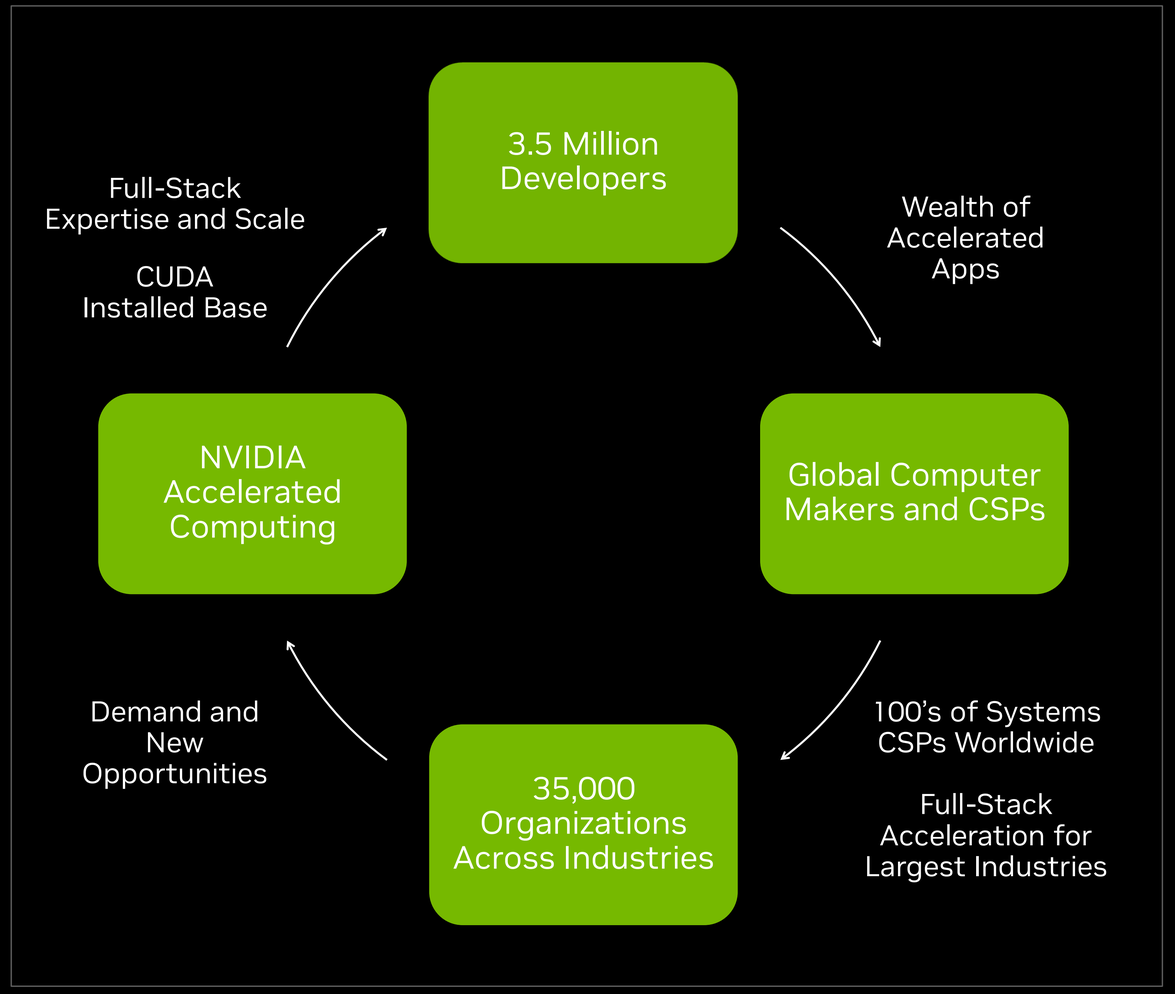
Scaling across multi-GPUs and multi-nodes is a data center-scale challenge and requires treating the network and storage as part of the computing fabric.

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

NVIDIA's 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



Full-Stack & Data Center Scale Acceleration

Drive Significant Cost Savings and Workload Scaling

Classical Computing—92 CPU-only servers

$3.3M (including switches, cables, racks)

Application

CPU server racks

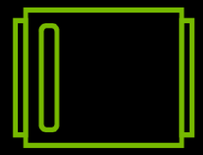
Accelerated Computing—1 NVIDIA DGX A100

$220,000 DGX and $100,000 NVIDIA AI software

Application
Re-Engineered for Acceleration

CUDA-X Acceleration Libraries

Magnum IO



DGX

10X lower cost
14X better energy-efficiency

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.

New NVIDIA Software and Services

Enabling the World's Enterprises to Revolutionize Industries with AI

NVIDIA AI Enterprise

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.

Per GPU On-Prem Subscription

Per GPU-HR Cloud Consumption

NVIDIA Omniverse

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.

Per Connection On-Prem Subscription

Per GPU-HR Cloud Consumption

NVIDIA Nemo LLM

NVIDIA-hosted cloud service for training Large Language Models to perform specific tasks—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.

Per GPU On-Prem Subscription

Per GPU-HR Cloud Consumption

NVIDIA BioNemo

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.

Per GPU On-Prem Subscription

Per GPU-HR Cloud Consumption

Giant Market Opportunity

Gaming & Metaverse

Financial Services

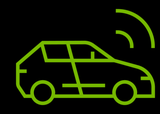
Healthcare

Logistics

Manufacturing

Retail

Transportation



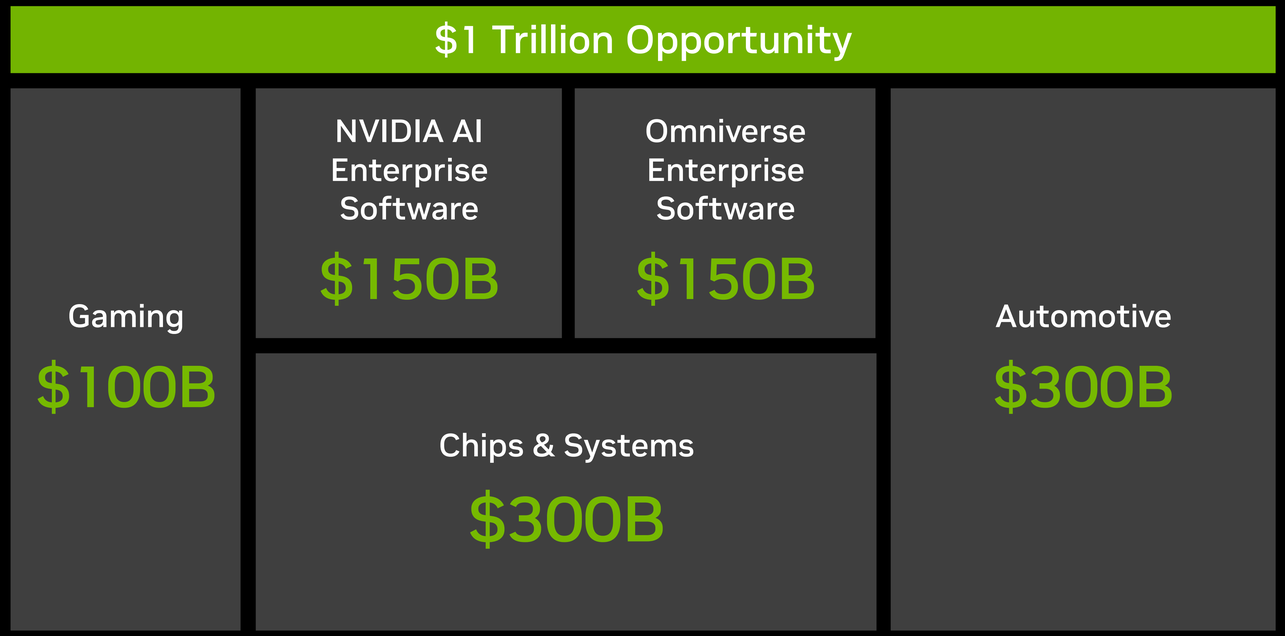
Gaming
Over 3B gamers and creators, a quarter of them spending over $100/year for GPUs in desktops, laptops, cloud or consoles

NVIDIA AI Enterprise Software
50M enterprise server installed base; per-server, per-year subscription price

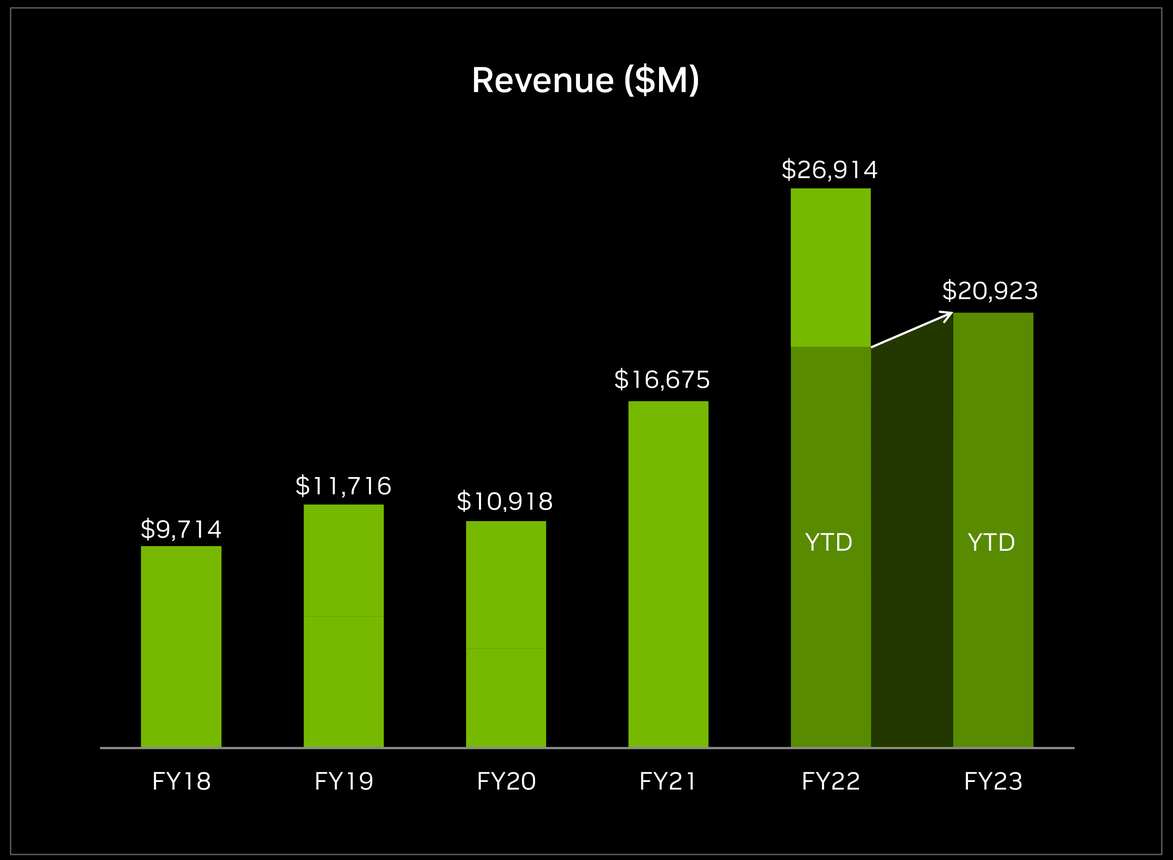
Omniverse Enterprise Software
Over 45M designers and creators; 10s of millions of digital twins —per-user/digital twin, per-year subscription price

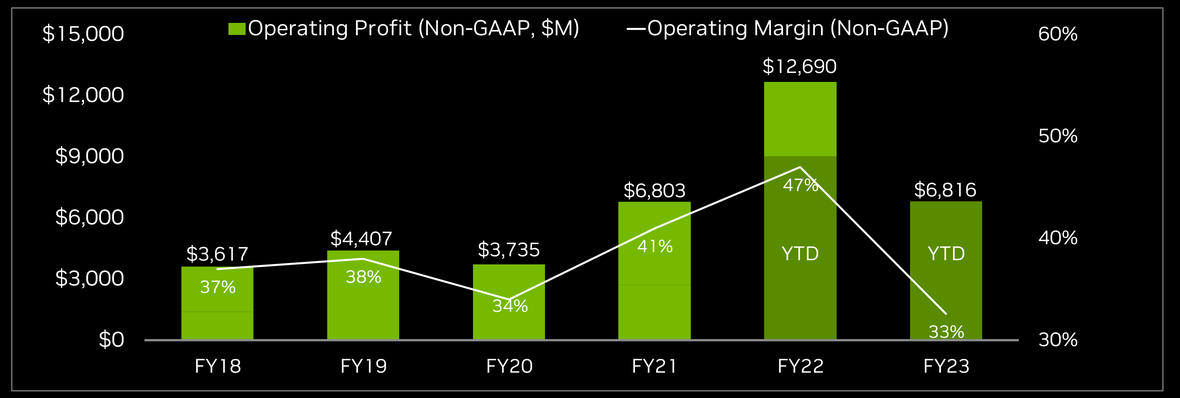
Chips and Systems
~20M servers/year—GPUs, CPUs, DPUs, NICs, switches

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

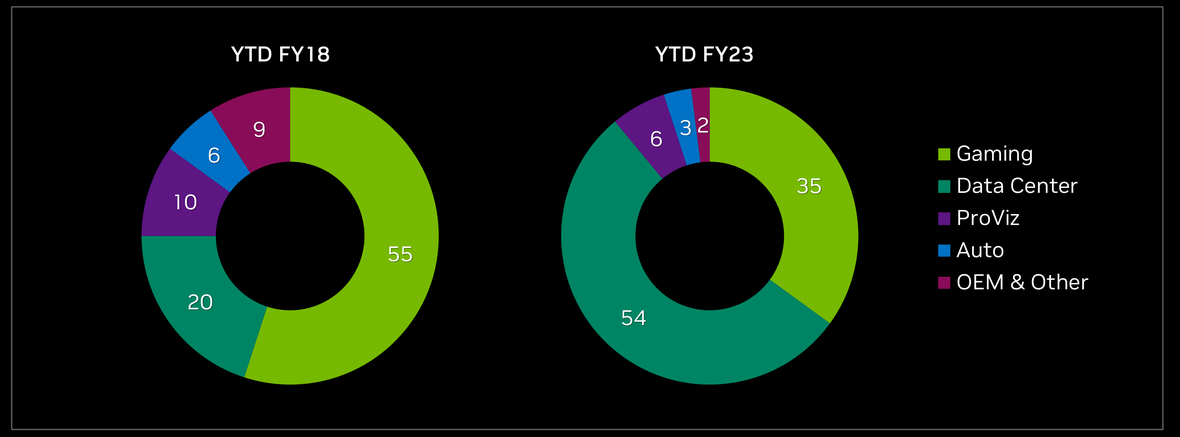


Driving Strong & Profitable Growth





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



YTD FY23 financial metrics reflect a $1.9B charge for inventory and related reserves primarily related to Data Center and Gaming.

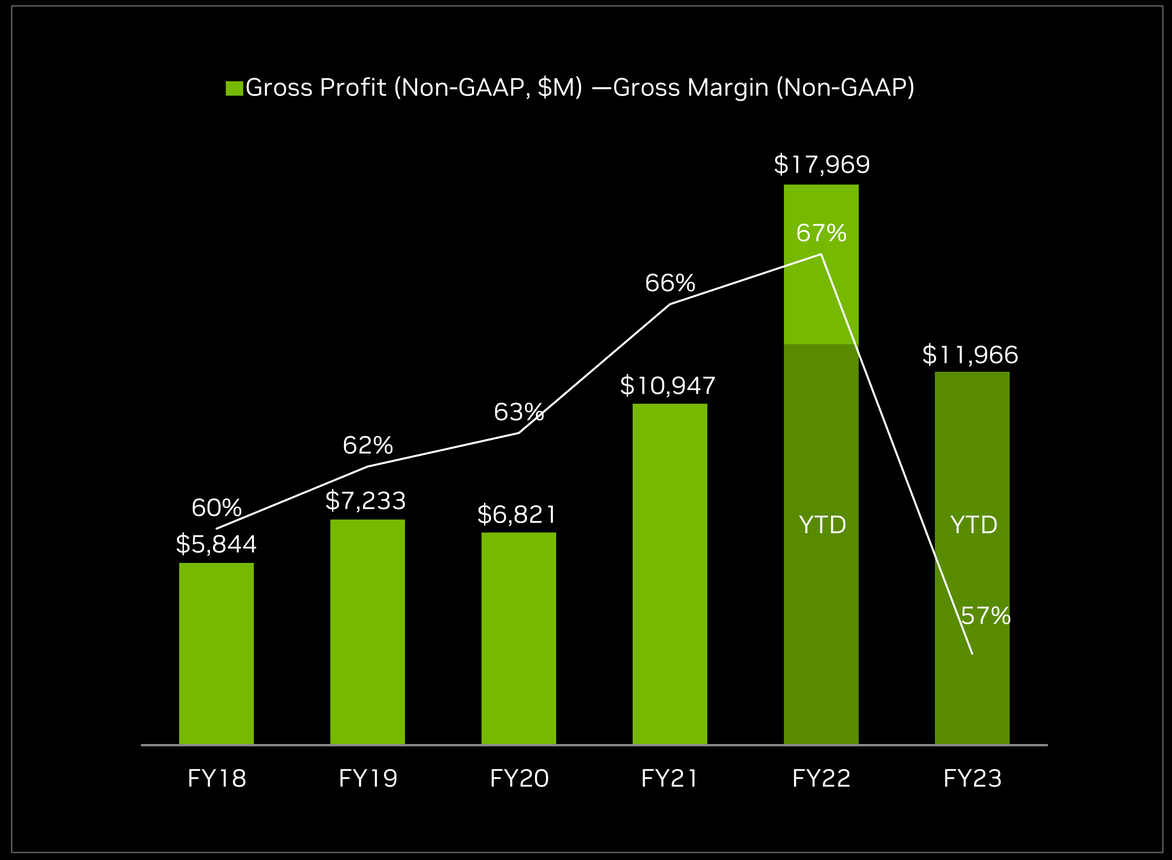
NVIDIA Gross Margins Reflect Value of Acceleration

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.

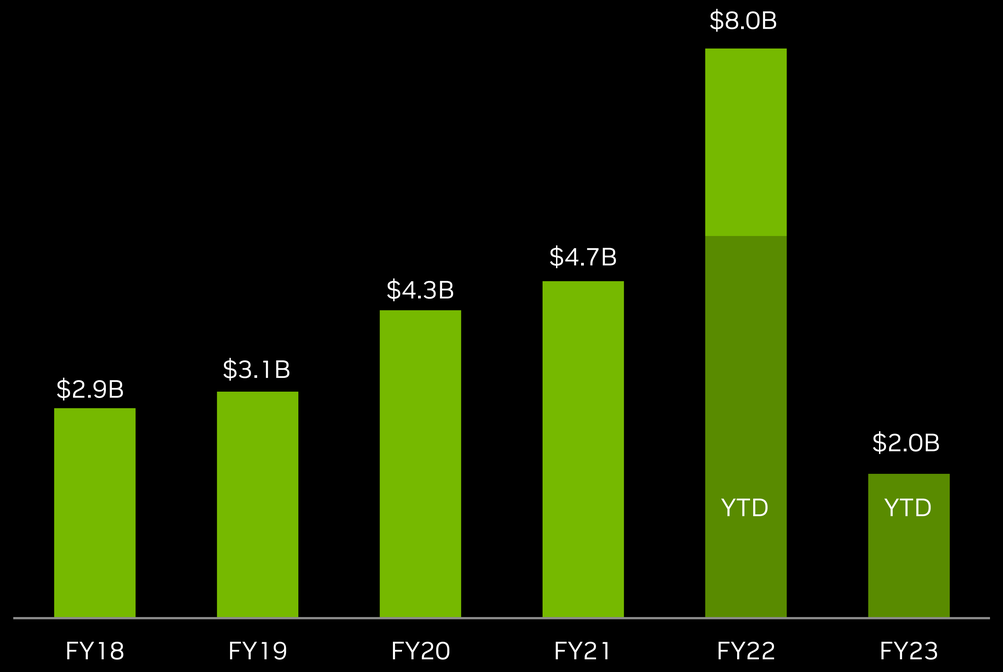


Cost comparison example based on latest available NVIDIA A100 GPU and Intel CPU inference results in the commercially available category of the MLPerf industry benchmark; includes related infrastructure costs such as networking.

YTD FY23 financial metrics reflect a $1.9B 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

Resumed Buybacks in Q1 FY 2023

$9.0B repurchased YTD FY23; $8.3B Remaining Authorization Through Dec 2023 as of Oct 30, 2022

Dividend

$400M in FY 2022

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

Gaming

46% of FY22 revenue

FY22 Revenue $12.5B
5-yr CAGR 25%

GeForce GPUs for PC gaming

GeForce NOW cloud gaming

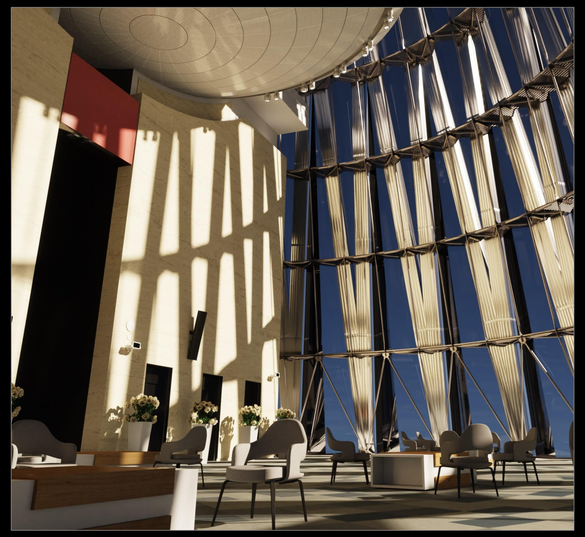
Data Center

40% of FY22 revenue

FY22 Revenue $10.6B
5-yr CAGR 66%

DGX/HGX/EGX/IGX systems

GPU | CPU | DPU | Networking
NVIDIA AI software



Professional Visualization

8% of FY22 revenue

FY22 Revenue $2.1B
5-yr CAGR 20%

Quadro/NVIDIA RTX GPUs
for workstations

Omniverse software

Automotive

2% of FY22 revenue

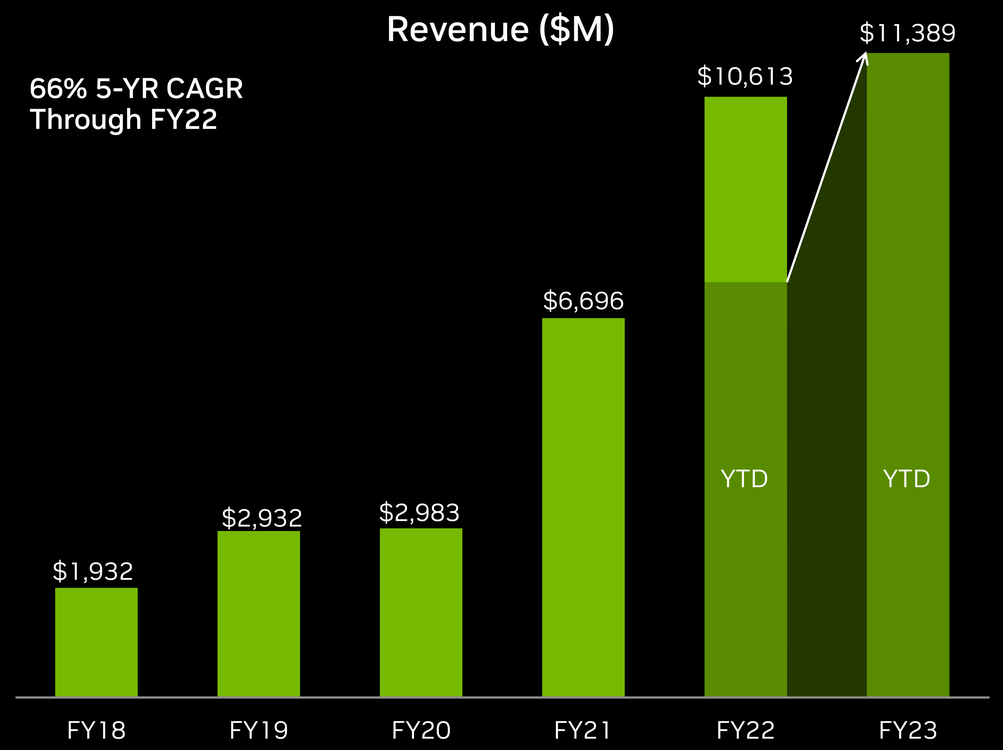
FY22 Revenue $0.6B
5-yr CAGR 3%

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 361 of the TOP500 supercomputers

Growth Drivers

Rapid AI adoption across industries

Full-stack AI | Software

Three chip strategy—GPU | CPU | DPU

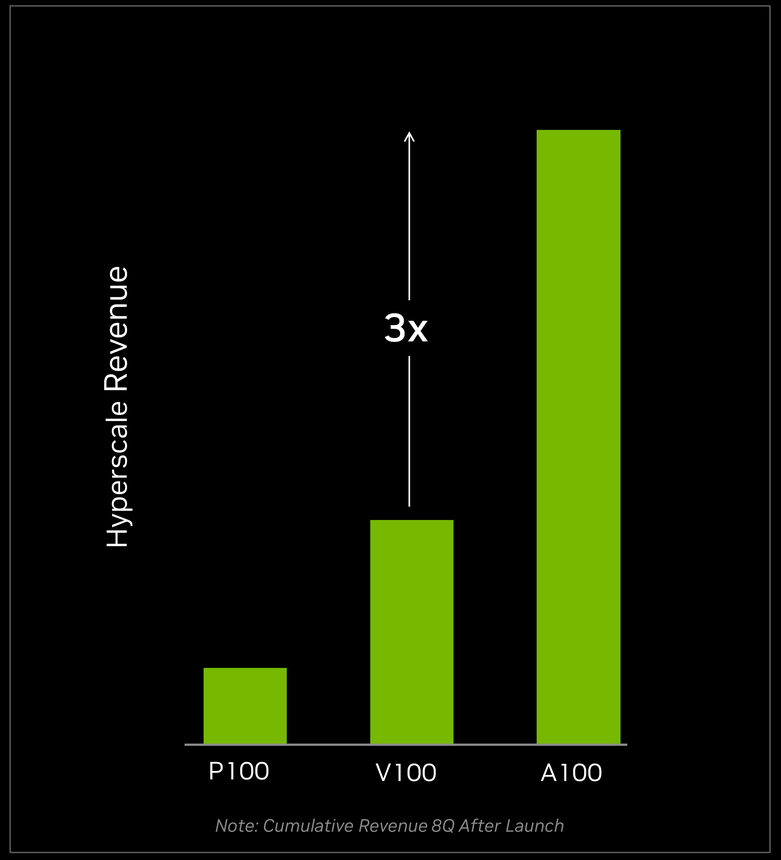
Rising computation requirements for modern AI

Data-center scale innovation

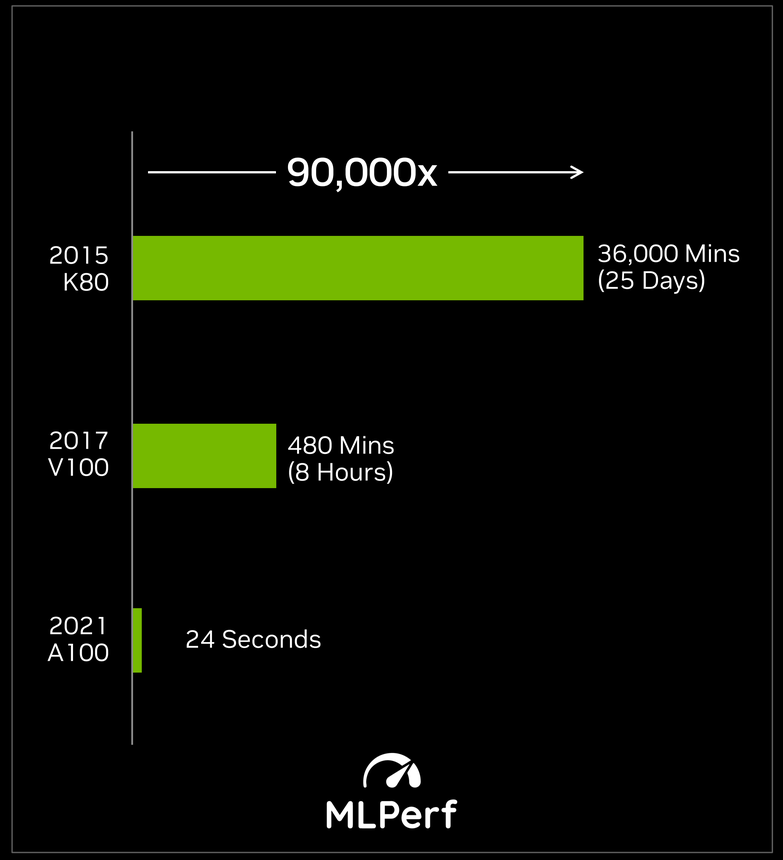
Omniverse

Data Center

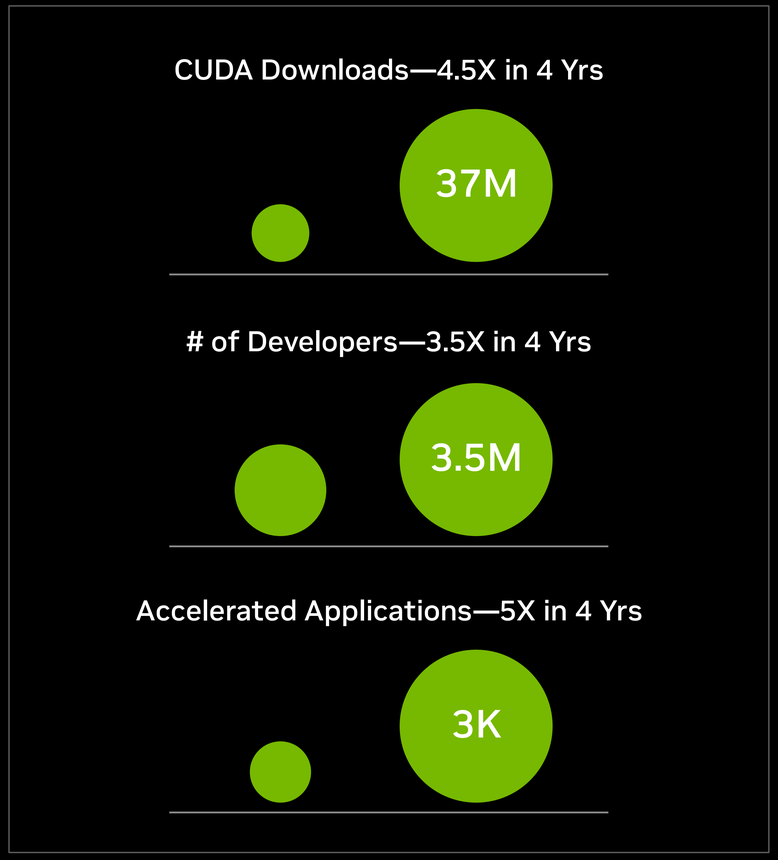
Strong growth fueled by AI performance leadership and huge developer ecosystem



Accelerating Adoption with Every Architecture



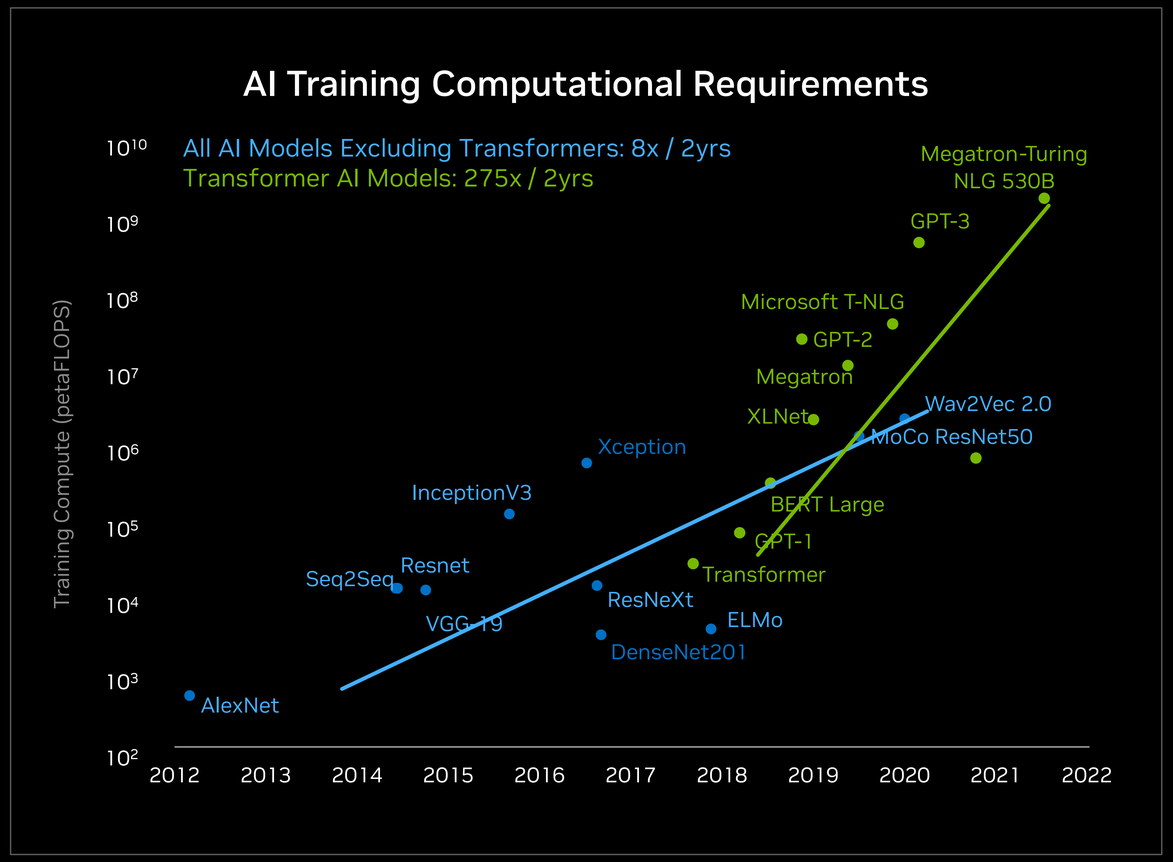
#1 in AI Training & Inference Performance

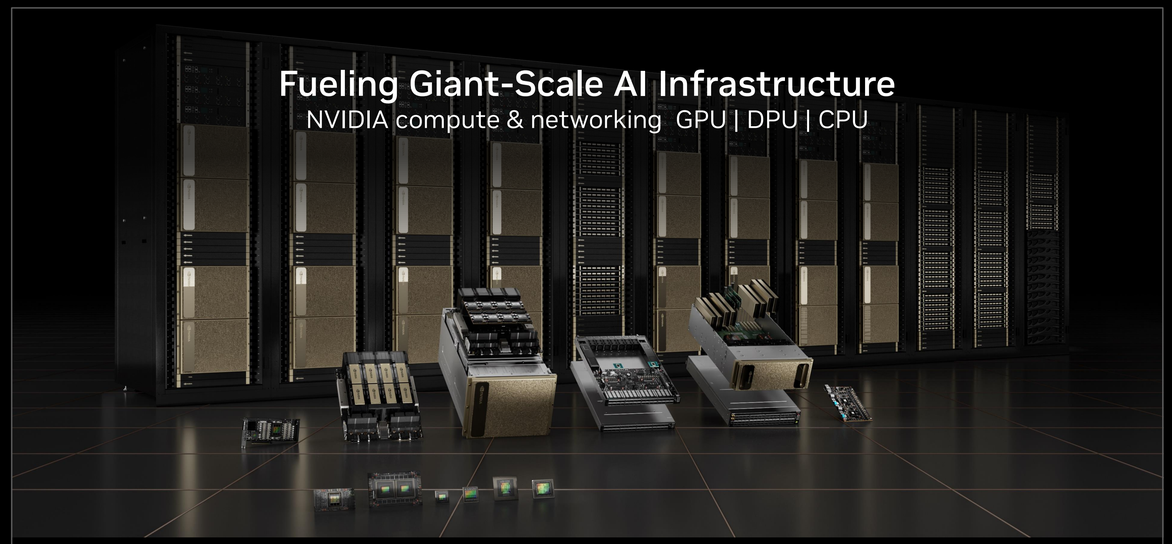


The Largest Accelerated Computing Ecosystem

Modern AI is a Data Center Scale Computing Workload

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





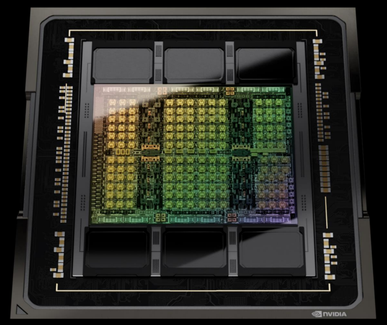
Large Language Models, based on the Transformer architecture, are one of today's most important advanced AI technologies, involving up to trillions of parameters that learn from text.

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.

Wave of New Data Center Products

Ramping new architectures for GPU, CPU and DPU

H100 GPU



World's Most Advanced Chip

80B Transistors

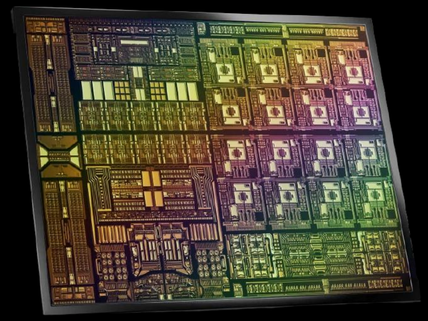
Transformer Engine – 6X Perf

Confidential Computing

4th Gen NVLink—7X PCIe Gen5

2H FY23

Bluefield-3 DPU



First 400 Gb/s DPU

Line-rate processing of software-defined networking, storage, and cybersecurity

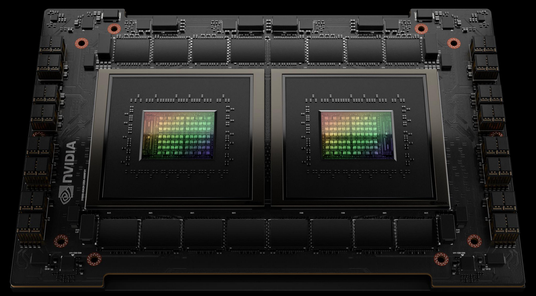
VMware vSphere 8 integration

Zero-trust security

~600 infrastructure software partners

1H FY23

Grace CPU Superchip



High Performance CPU for HPC and AI

144 Cores | 740 SPECrate®2017_int_base est.

1TB/s Memory Bandwidth

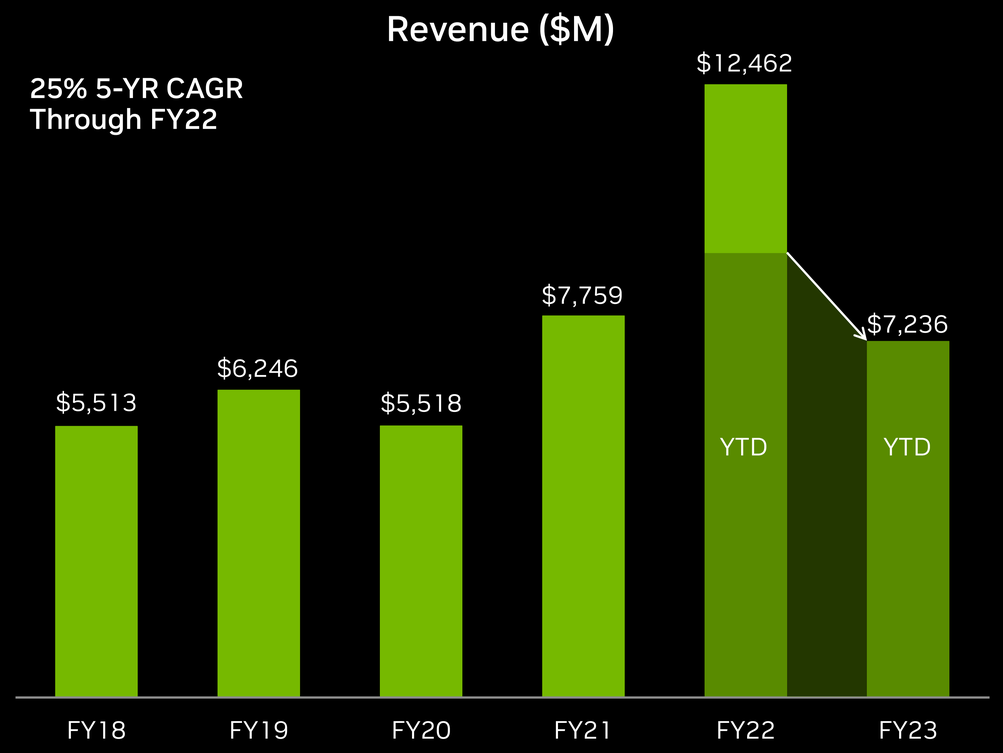
2X Perf/Watt Over Traditional Servers

Runs NVIDIA Computing Stacks

1H FY24

Gaming

GeForce—the world's largest gaming platform



Leader in PC Gaming

Strong #1 market position with over 80% share

15 of the Top 15 most popular GPUs on Steam

Leading performance & innovation

200M+ gamers on GeForce

Growth Drivers

Rising adoption of NVIDIA RTX

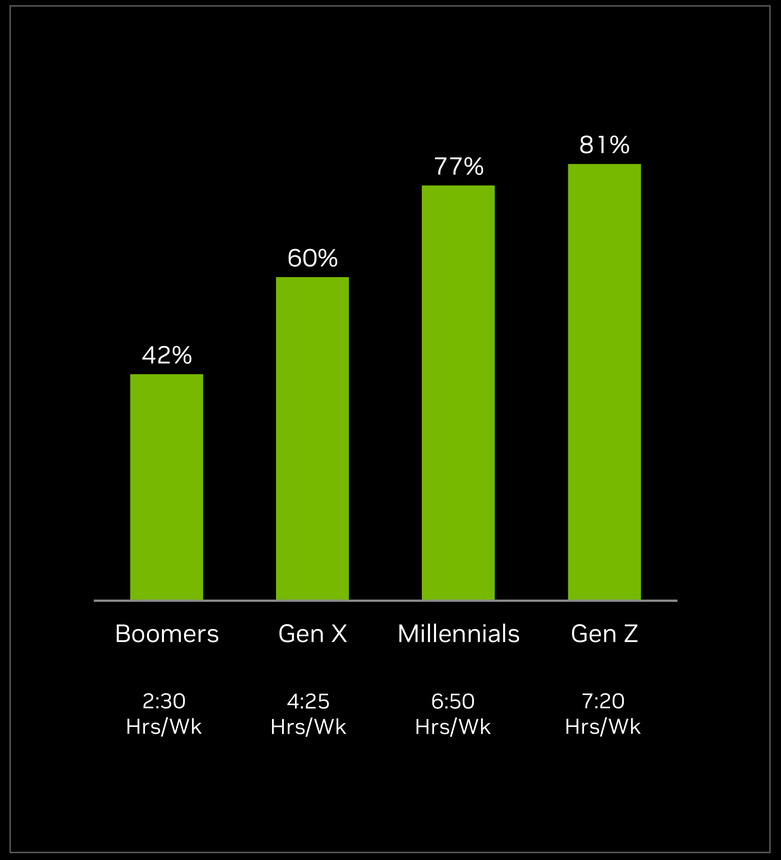
Expanding universe of gamers & creators

Gaming laptops & game consoles

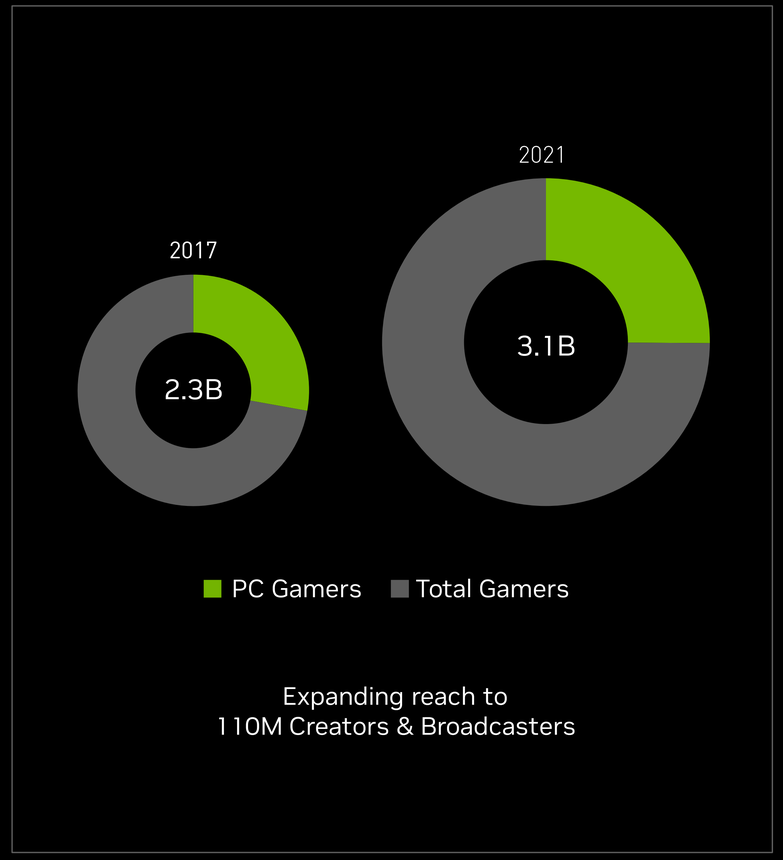
GeForce NOW Cloud gaming

Strong Gaming Fundamentals

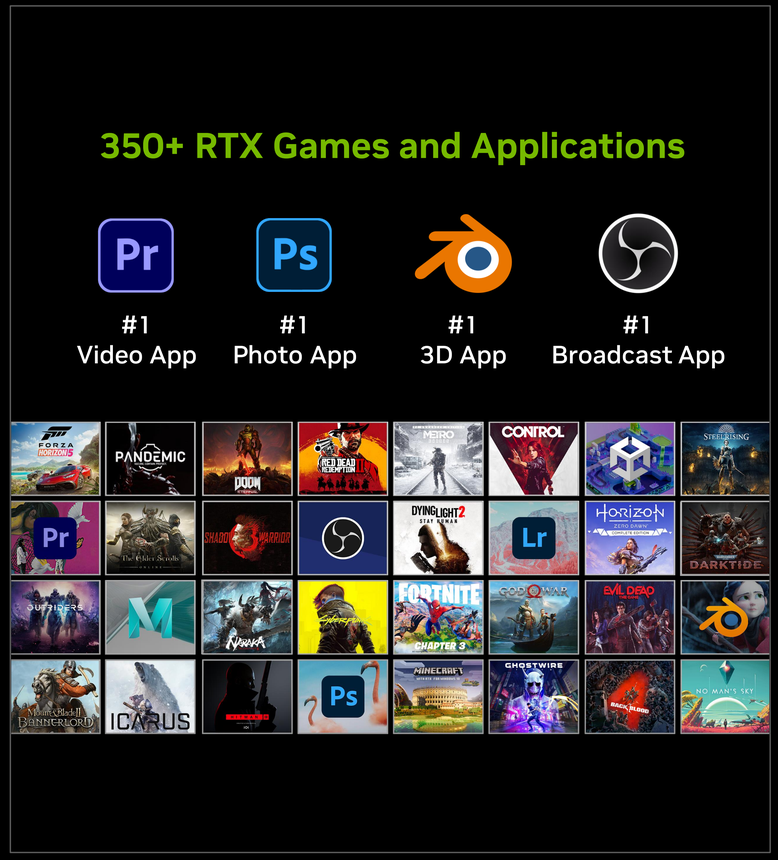
New generation, more gamers



New generation, more gamers



Expanding universe of gamers and creators

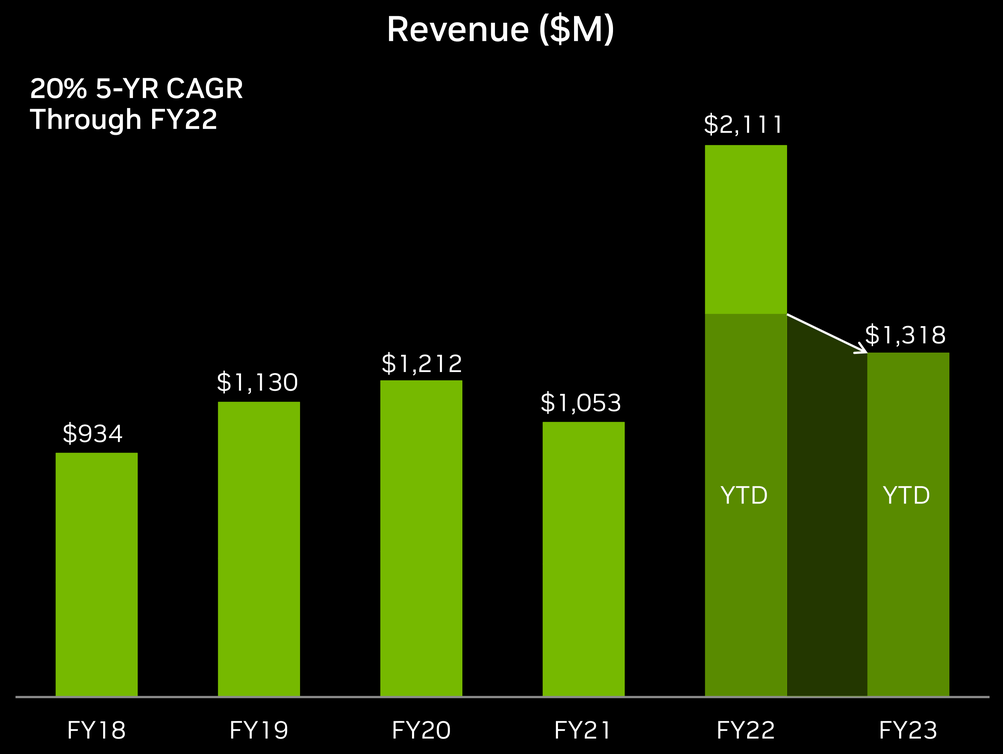


Robust NVIDIA ecosystem

Source: NewZoo and NVIDIA internal analysis

Professional Visualization

Workstation graphics



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

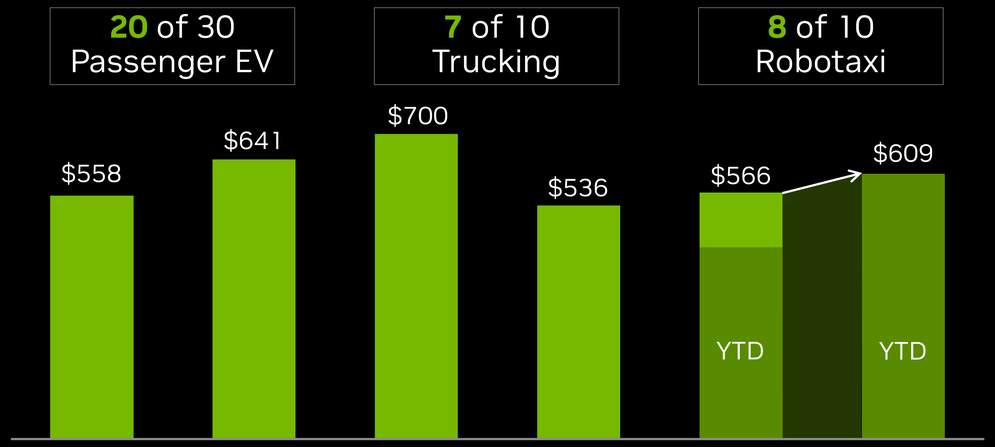
Autonomous Vehicles (AV) & AI Cockpit

Revenue ($M)

Our Next Billion-Dollar Business

Over $11B design win pipeline across 40 customers

Leadership Position in All Segments



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

Over $11B design win pipeline through FY28 based on DRIVE Orin, which started ramp in FY23

Next-generation DRIVE Thor to ramp in FY25

Growth Drivers

Adoption of centralized car computing and software-defined vehicle architectures

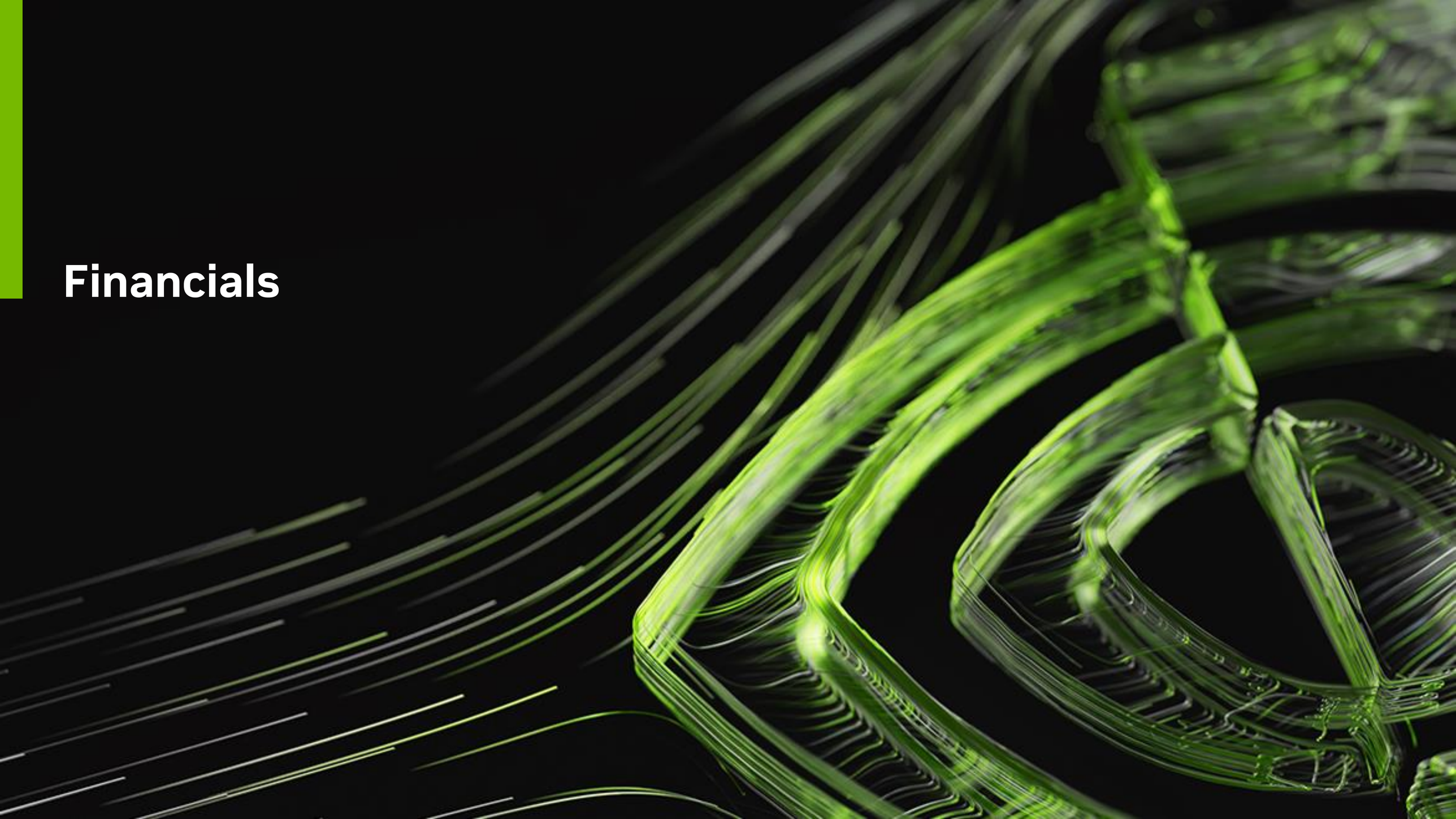
AV software and services:

Mercedes Benz FY25 SOP*

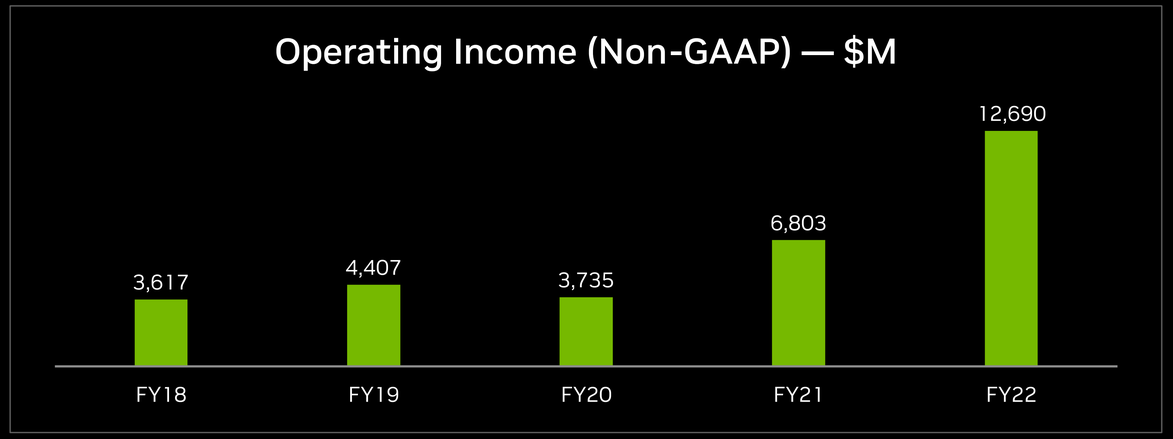
Jaguar Land Rover FY26 SOP

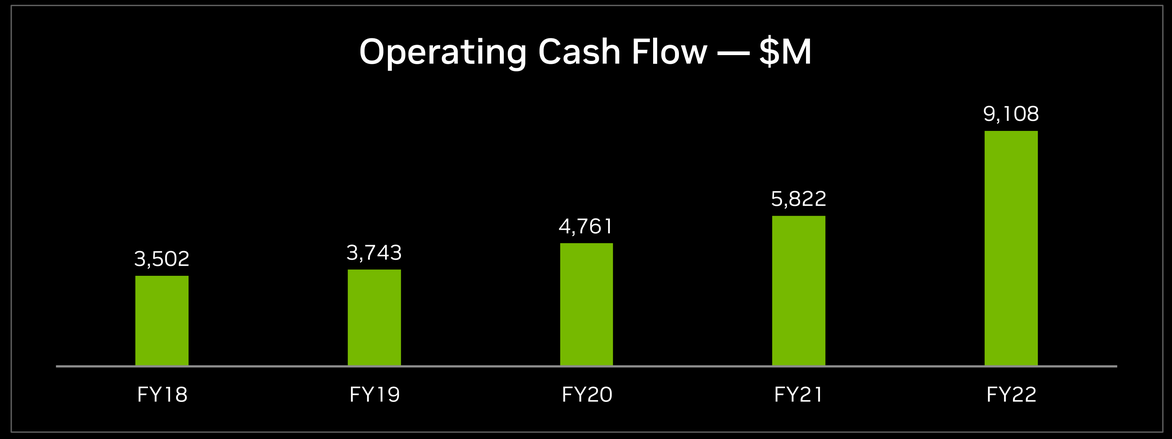
**SOP = Start of Production*

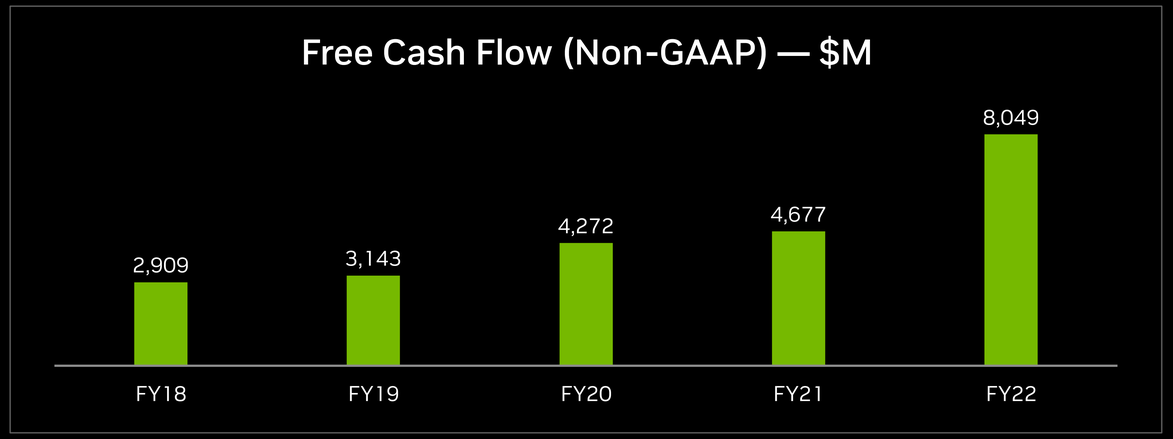
Financials

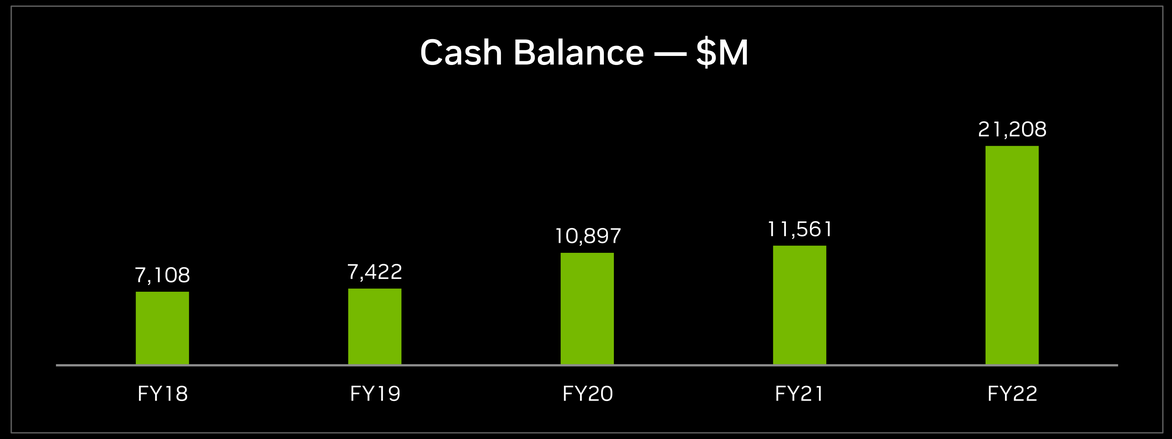


Annual Cash & Cash Flow Metrics









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

NVIDIA's two HQ campuses have received **LEED Gold status**

23 of Top 30 Supercomputers on the Nov 2022 Green500 powered by NVIDIA including the #1 system, Henri

We Plan For **100% of Our Global Electricity Usage For Our Offices and Data Centers** to Be Renewable by 2025

A Place For People To Do Their Life's Work

#1

"100 Best Companies to Work For"
FORTUNE

"America's Most Just Companies"
FORBES

"Most Responsible Companies"
NEWSWEEK

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

Management

Time Magazine's 100 Most Influential Companies

CEO Magazine's 10 Best CEOs

Fortune's World's Most Admired Companies

Wall Street Journal's Management Top 250 All-Stars

Corporate Governance

38% Of Board is Gender, Racially, or Ethnically Diverse

92% of Directors are Independent

Reconciliation of Non-GAAP to GAAP Financial Measures

Reconciliation of Non-GAAP to GAAP Financial Measures

	Non-GAAP	Acquisition Termination Cost	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	Other (C)	Tax Impact of Adjustments	GAAP
Q3 FY23							
Gross margin ($ in million)	$3,329	—	(120)	(32)	—	—	$3,177
	56.1%	—	(2.0)	(0.5)	—	—	53.6%
Operating income ($ in million)	$1,536	—	(174)	(745)	(16)	—	$601
Net income ($ in million)	$1,456	—	(174)	(745)	(28)	171	$680
Shares used in diluted per share calculation (millions)	2,499	—	—	—	—	—	2,499
Diluted EPS	$0.58	—	—	—	—	—	$0.27

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 restructuring, contributions and net losses 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
Q3 FY2022	67.0%	(1.2)	(0.6)	—	65.2%
Q4 FY2022	67.0%	(1.1)	(0.5)	—	65.4%
Q1 FY2023	67.1%	(1.1)	(0.5)	—	65.5%
Q2 FY2023	45.9%	(1.8)	(0.6)	—	43.5%

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 2018	$5,844	—	(21)	(1)	$5,822
	60.2%	—	(0.3)	—	59.9%
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%

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

Gross Margin ($ in Millions & Margin Percentage)	Non-GAAP	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	GAAP
YTD Q3 FY2022	$12,844	(258)	(102)	(8)	$12.476
	66.6%	(1.4)	(0.5)	—	64.7%
YTD Q3 FY2023	$11,966	(335)	(108)	—	$11,523
	57.2%	(1.6)	(0.5)	—	55.1%

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 Margin ($ in Millions & Margin Percentage)	Non-GAAP	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	Other	GAAP
FY 2018	$3,617	(13)	(391)	(3)	$3,210
	37.2%	(0.2)	(4.0)	—	33.0%
FY 2019	$4,407	(2)	(557)	(44)	$3,804
	37.6%	—	(4.7)	(0.4)	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%

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 IP-related costs, legal settlement costs, contributions, and restructuring and other charges

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

Operating Margin ($ in Millions & Margin Percentage)	Non-GAAP	Acquisition Termination Cost	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	Other	GAAP
YTD Q3 FY2022	$9,014	—	(482)	(1,453)	(8)	$7,071
	46.8%	—	(2.6)	(7.5)	—	36.7%
YTD Q3 FY2023	$6,816	(1,353)	(499)	(1,971)	(25)	$2,968
	32.6%	(6.5)	(2.4)	(9.4)	(0.1)	14.2%

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 IP-related costs, legal settlement costs, contributions, and restructuring and other charges

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	Net Cash Provided by Operating Activities
FY 2018	$2,909	593	—	$3,502
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
YTD Q3 FY 2022	$5,310	703	62	$6,075
YTD Q3 FY 2023	$2,015	1,324	54	$3,393

Reconciliation of Non-GAAP to GAAP Financial Measures

($ in Millions)	Q4 FY23 Outlook
Non-GAAP gross margin	66.0%
Impact of stock-based compensation expense, acquisition-related costs, and other costs	(2.8%)
GAAP gross margin	63.2%
Non-GAAP operating expenses	$1,780
Impact of stock-based compensation expense and acquisition-related costs	780
GAAP operating expenses	$2,560

NVIDIA