



# NVIDIA'S DATACENTER-LÖSUNGEN UND WIE SIE IHRE RECHENZENTRUMSTRATEGIE VERÄNDERN WERDEN

Daniela Marggraf // [dmarggraf@nvidia.com](mailto:dmarggraf@nvidia.com)

# NVIDIA IS REVOLUTIONIZING COMPUTING

An army of 8000 engineers, developing ever more complex GPUs ,  
supporting 3000 global start-ups to solve the worlds most  
challenging problems ....



A PLATFORM TO  
BRING 2D DESIGNS  
TO LIFE

**Koenigsegg**



**KPF**

AUTODESK® MAYA®

**3** AUTODESK®  
3DS MAX®

# A PLATFORM FOR AUTONOMOUS VEHICLES

VOLVO



SCANIA



TOYOTA

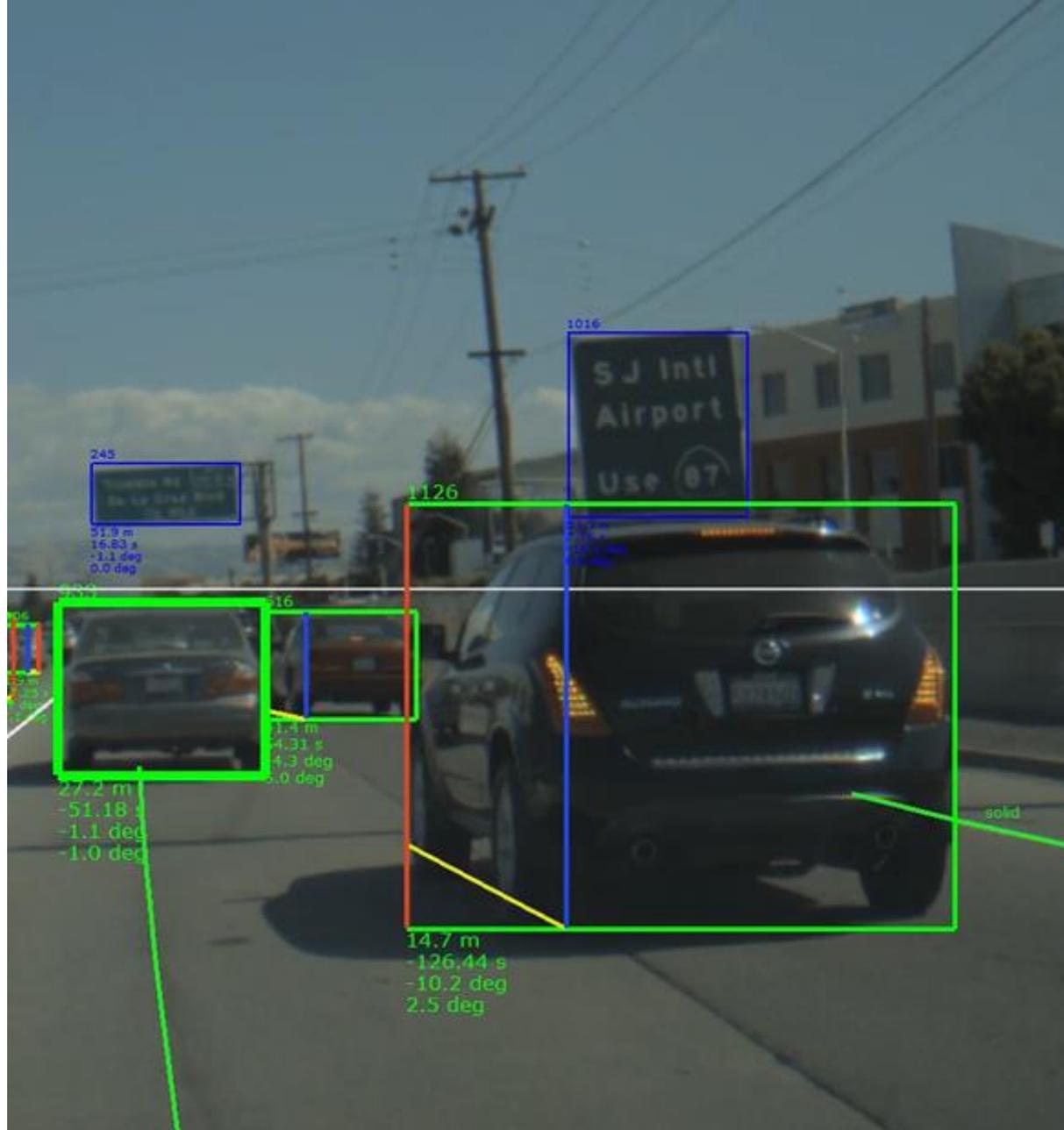


TESLA

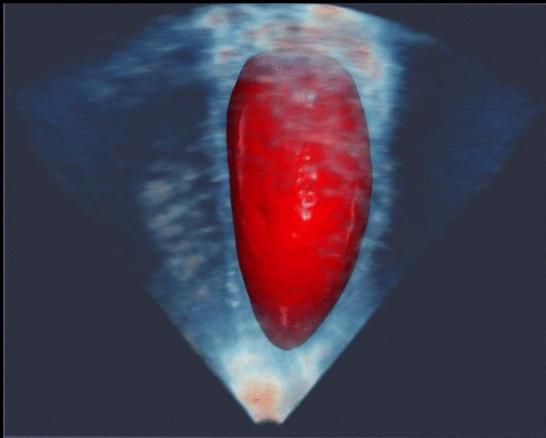
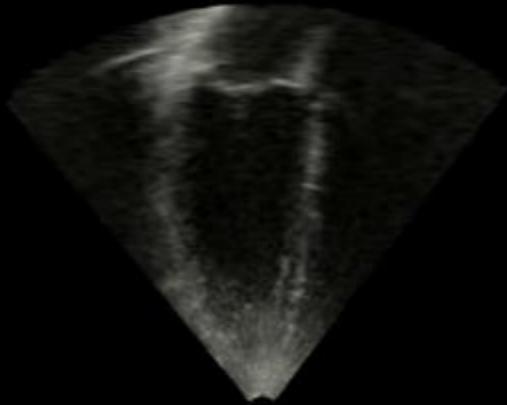
Deutsche Post DHL Group



Mercedes-Benz



200  
150  
100  
50  
0

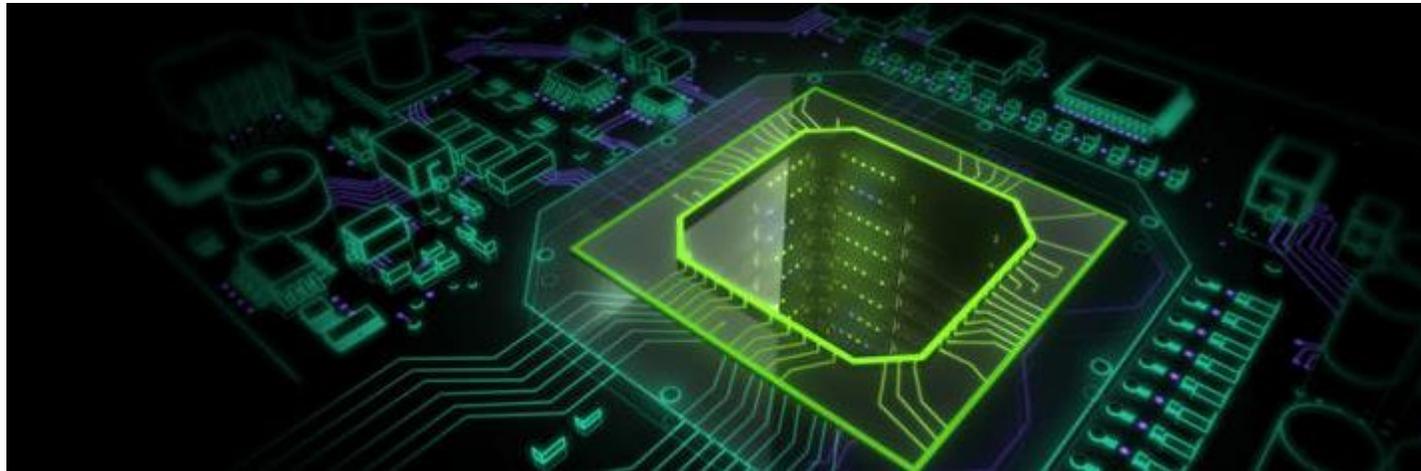


# A PLATFORM TO REDEFINE MEDICAL IMAGING



# NVIDIA IS REVOLUTIONIZING COMPUTING

*“NVIDIA didn’t innovate by changing what it does. Instead, we innovate by creating things the world needs with our technologies”*

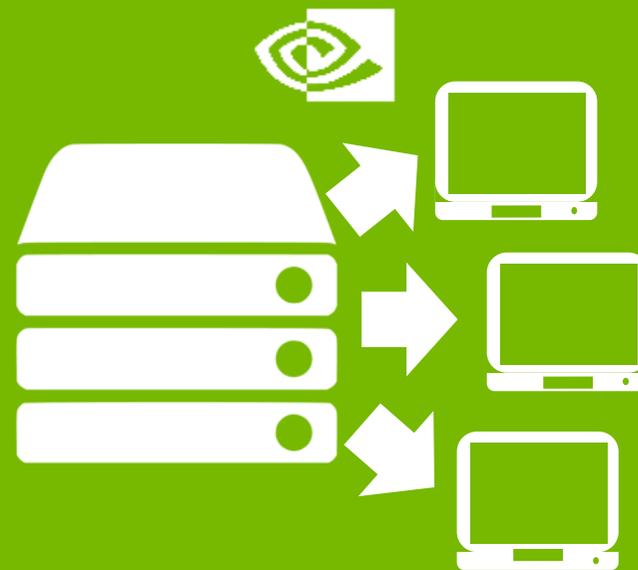
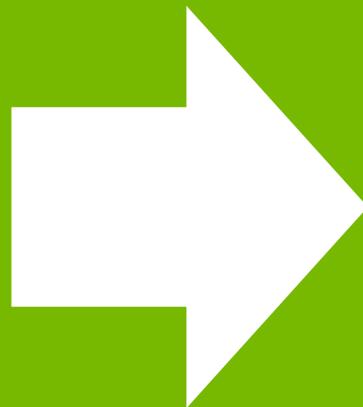


# PERFORMANCE FROM THE DATA CENTER

NVIDIA Virtual GPU technology delivers graphics accelerated virtual desktops and applications



All devices have graphics



Virtual machines also need a GPU

# ACCELERATE PRODUCTIVITY

For Every User, Any App



## Knowledge Worker

Providing business users the highest level of experience for all their apps on any device



## Creative & Technical Professional

Uncompromised experience for professional graphics users allowing them to design on the go



Scalability

Performance

# MODERN APPS IN THE DIGITAL WORKPLACE

Demands more GPU



Applications that require graphics has doubled since 2012

Over half of enterprise users access at least one graphics accelerated app

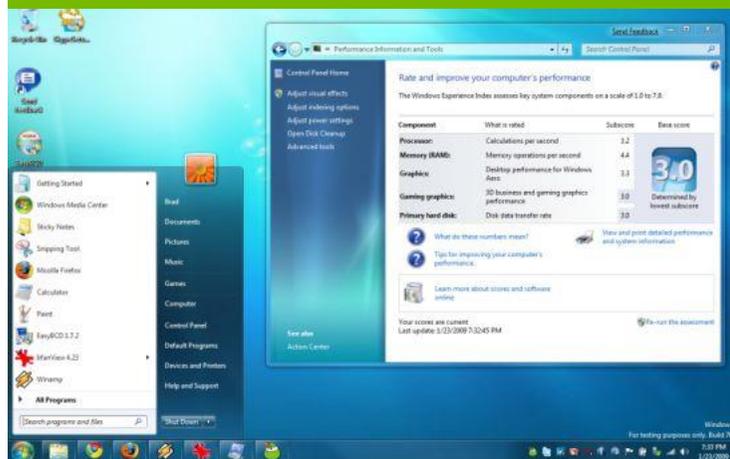
# WINDOWS 10 GRAPHICS USAGE

Highest graphics requirement from any operating system to date

## WINDOWS 95



## WINDOWS 7



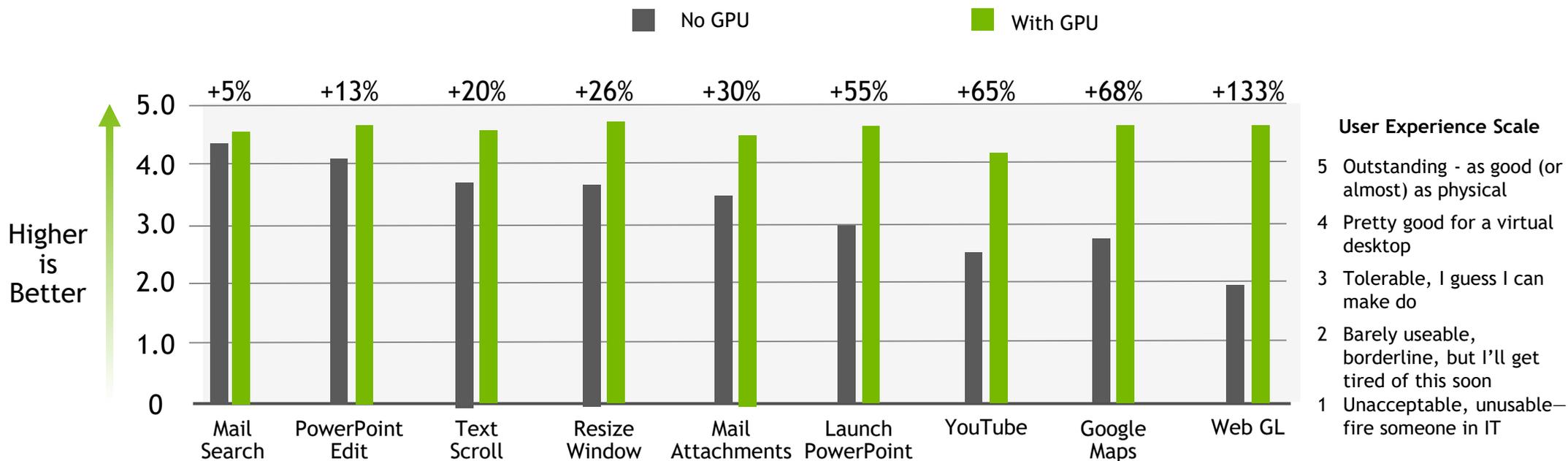
## WINDOWS 10



32% Increase in GPU Consumption, compared to Windows 7\*

# WINDOWS 10 WITH CPU vs. NVIDIA GRID

NVIDIA GRID provided an average 34% increase in Win10 UX quality



Testing ran on two identical systems, CPU system was loaded up to 60-80% utilization, the GPU system ran the same workload

# NVIDIA GRID VALUE



## Fantastic User Experience

- Immersive Windows 10 & modern apps experience
- Full multimedia experience with NVENC
- Guaranteed quality of service
- Mobility to any device, anywhere

## Cost Effective Solution

- Highest User Density with Tesla M10
- Increase User Density of Windows 10 Deployments by up to 30%

# THE WORLD'S MOST POWERFUL VIRTUAL WORKSTATION

- Ultra High-End Workflows on Quadro Virtual Data Center Workstation powered by Tesla V100
- Faster, more fluid visual interactivity throughout the design process

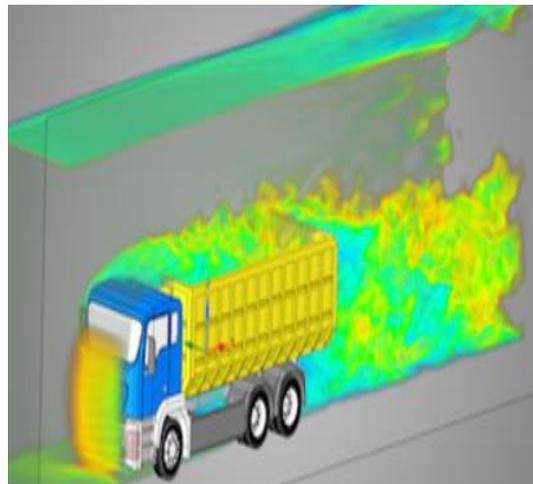
MOST ADVANCED DATA CENTER GPU



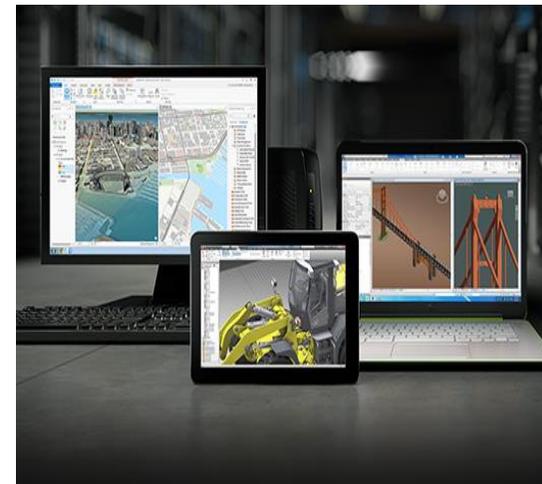
AI ENHANCED RENDERING



REAL-TIME INTERACTIVE SIMULATION



SECURE COLLABORATION



**DATACENTER READY - HARDWARE**

# NVIDIA GPUs EVERYWHERE

120+ Servers from more than 30 system vendors

## Industry Standard Servers



Hewlett Packard  
Enterprise



## Hyper-Converged Infrastructure



Hewlett Packard  
Enterprise

NUTANIX

## Blade Servers



Hewlett Packard  
Enterprise

## Public Cloud



Google Cloud Platform



Microsoft  
Azure



amazon  
web services



# NVIDIA TESLA GPUs

## Recommended for Virtualization

	V100	P100	P40	P4	M60	M10	M6	P6
<b>GPUs / Board (Architecture)</b>	1 (Volta)	1 (Pascal)	1 (Pascal)	1 (Pascal)	2 (Maxwell)	4 (Maxwell)	1 (Maxwell)	1 (Pascal)
<b>CUDA Cores</b>	5,120	3,584	3,840	2,560	4,096 (2,048 per GPU)	2,560 (640 per GPU)	1,536	2,048
<b>Memory Size</b>	32 GB/16 GB HBM2	16 GB HBM2	24 GB GDDR5	8 GB GDDR5	16 GB GDDR5 (8 GB per GPU)	32 GB GDDR5 (8 GB per GPU)	8 GB GDDR5	16 GB GDDR5
<b>H.264 1080p30 streams</b>	36	36	24	24	36	28	16	24
<b>vGPU Profiles</b>	1 GB, 2 GB, 4 GB, 8 GB, 16 GB, 32 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB	1 GB, 2 GB, 3 GB, 4 GB, 6 GB, 8 GB, 12 GB, 24 GB	1 GB, 2 GB, 4 GB, 8 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB	0.5 GB, 1 GB, 2 GB, 4 GB, 8 GB	1 GB, 2 GB, 4 GB, 8 GB, 16 GB
<b>Form Factor</b>	PCIe 3.0 Dual Slot (rack servers)	PCIe 3.0 Dual Slot (rack servers)	PCIe 3.0 Dual Slot (rack servers)	PCIe 3.0 Single Slot (rack servers)	PCIe 3.0 Dual Slot (rack servers)	PCIe 3.0 Dual Slot (rack servers)	MXM (blade servers)	MXM (blade servers)
<b>Power</b>	250W	250W	250W	75W	300W (225W opt)	225W	100W (75W opt)	90W
<b>Thermal</b>	passive	passive	passive	passive	active/passive	passive	bare board	bare board

**PERFORMANCE**  
Optimized

**DENSITY**  
Optimized

**BLADE**  
Optimized

# AI AND DEEP LEARNING

## ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



## MACHINE LEARNING

Machine learning begins to flourish.



## DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

1980's

1990's

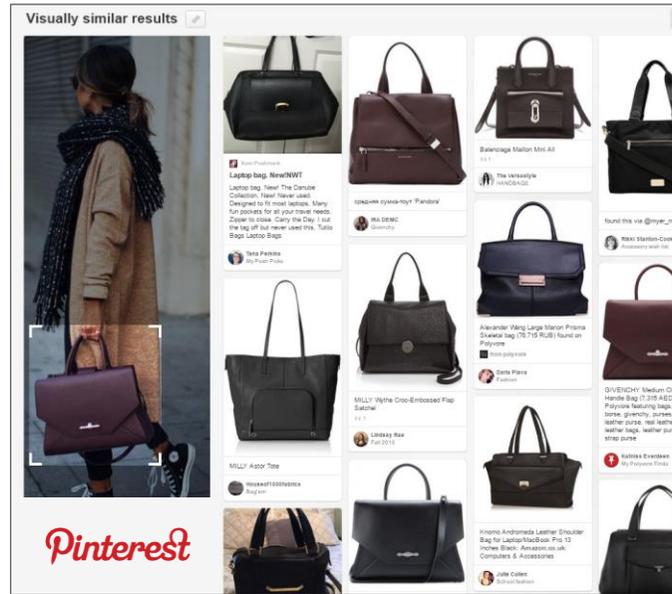
2000's

2010's

# AI IS EVERYWHERE



“Find where I parked my car”



“Find the bag I just saw in this magazine”

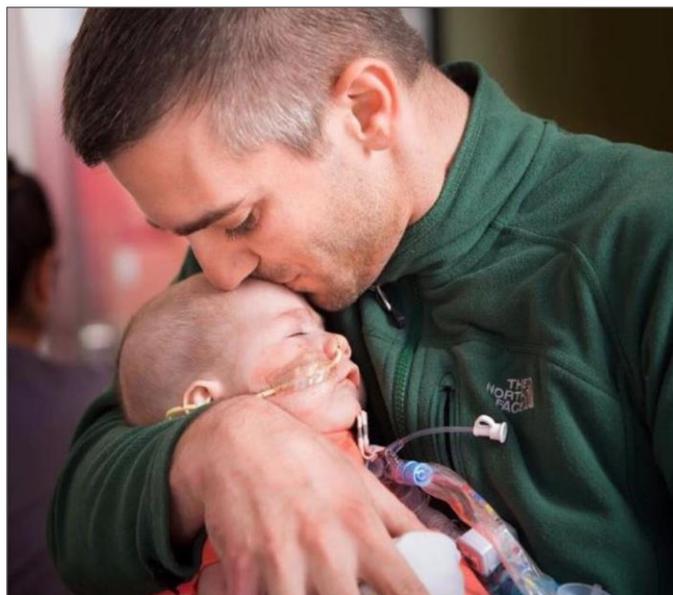


“What movie should I watch next?”

# TOUCHING OUR LIVES



Bringing grandmother closer to family by bridging language barrier



Predicting sick baby's vitals like heart rate, blood pressure, survival rate



Enabling the blind to “see” their surrounding, read emotions on faces

# AI FOR PUBLIC GOOD



Increasing public safety with smart video surveillance at airports & malls



Providing intelligent services in hotels, banks and stores



Separating weeds as it harvests, reduces chemical usage by 90%

# AI MOMENTUM

TODAY

3000

AI startups

BY 2020

85%

of all customer  
service interactions will  
be powered by  
AI bots

\$47B

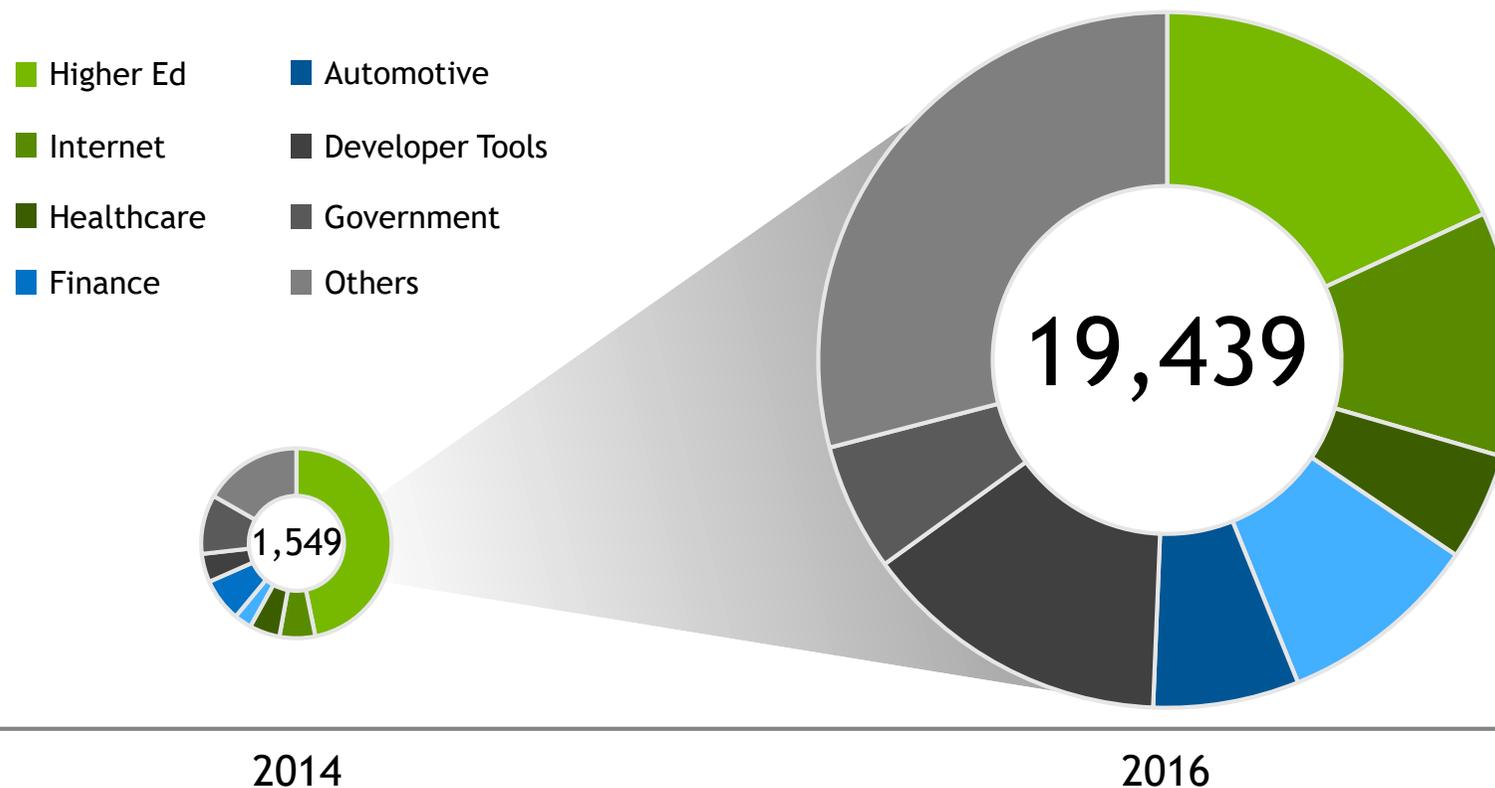
spend on AI  
technologies

20%

of companies will  
dedicate workers to  
monitor and guide  
neural networks

# EVERY INDUSTRY HAS AWOKEN TO AI

Organizations engaged with NVIDIA on Deep Learning



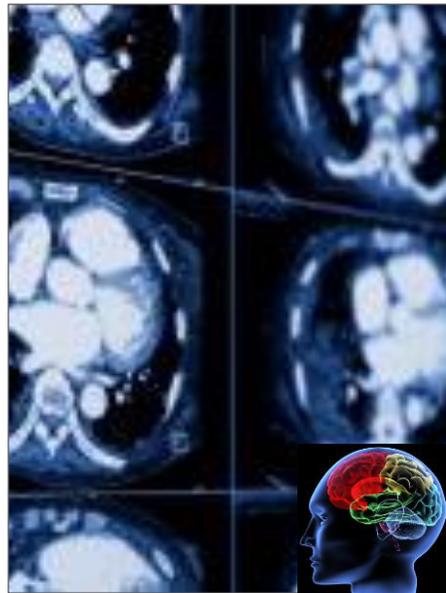
# DEEP LEARNING IS SWEEPING ACROSS INDUSTRIES

## INTERNET SERVICES



Image/Video classification  
Speech recognition  
Natural language processing

## MEDICINE



Cancer cell detection  
Diabetic grading  
Drug discovery

## MEDIA & ENTERTAINMENT



Video captioning  
Content based search  
Real time translation

## SECURITY & DEFENSE



Face recognition  
Video surveillance  
Cyber security

## AUTONOMOUS MACHINES

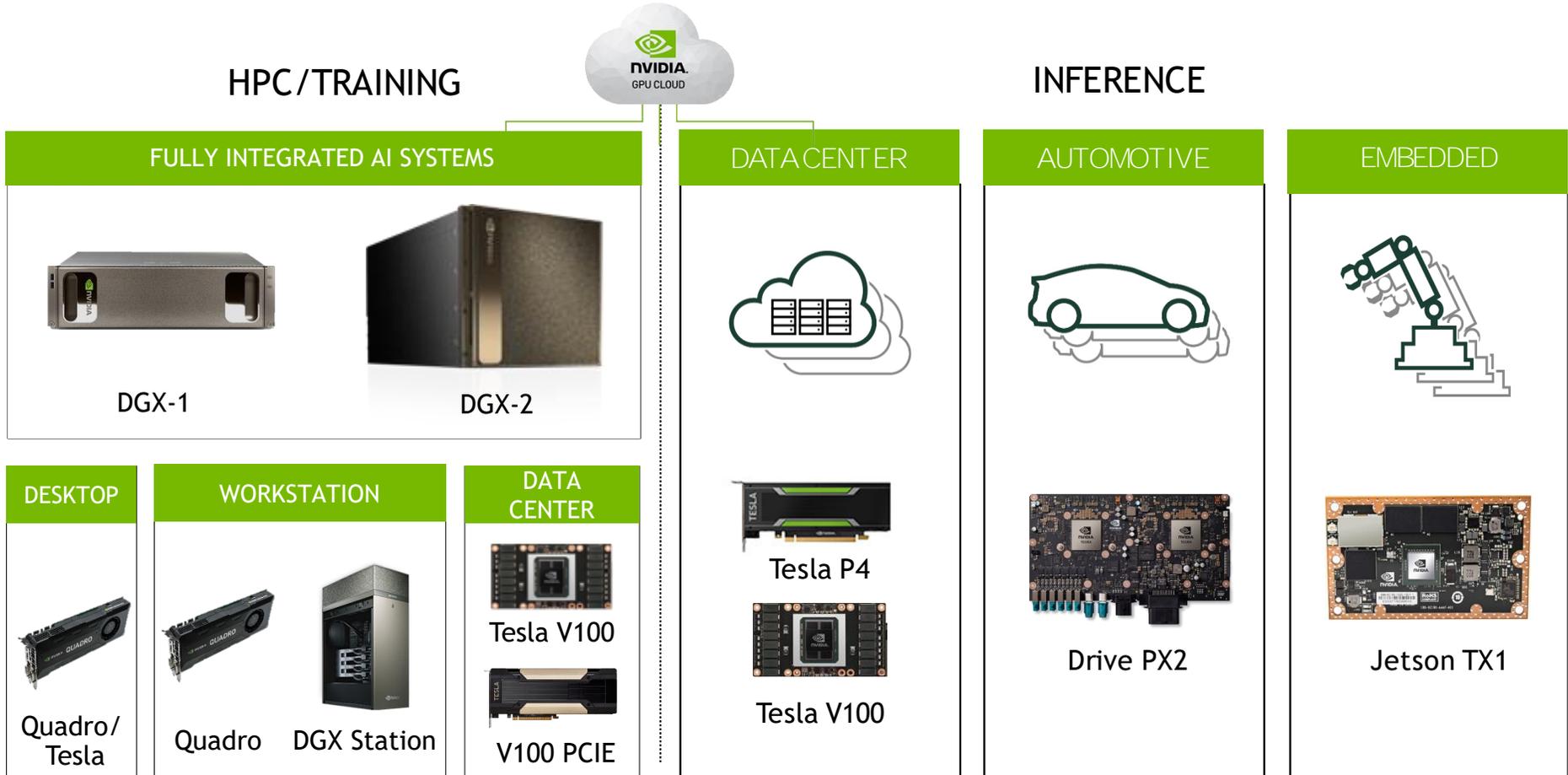


Pedestrian detection  
Lane tracking  
Recognize traffic signs

# DEEP LEARNING REQUIREMENTS

DEEP LEARNING NEEDS...	DEEP LEARNING CHALLENGES	NVIDIA DELIVERS
<b>Data Scientists</b>	Demand far exceeds supply	DIGITS, DLI Training
<b>Latest Algorithms</b>	Rapidly evolving	DL SDK, GPU-Accelerated Frameworks
<b>Fast Training</b>	Impossible -> Practical	DGX-1, P100, P40, TITAN X
<b>Deployment Platform</b>	Must be available everywhere	TensorRT, P40, P4, Jetson, Drive PX

# END-TO-END PRODUCT FAMILY



# SUMMARY

## DATACENTER STRATEGY: ALWAYS GPUS



VISUAL COMPUTING SPECTRUM

<http://www.nvidia.com/testdrive>  
[www.nvidia.com/DLI](http://www.nvidia.com/DLI)

# PLACE TO BE!

GPU Technology Conference Oct 9-11, 2018 Munich

NVIDIA's GPU Technology Conference (GTC) Europe is part of the largest global series of events focused on artificial intelligence and its applications across many important fields.

Join us in Munich and discover the latest breakthroughs in autonomous vehicles, high performance computing, healthcare, big data, Professional Visualization, virtual GPU and more.

<https://www.nvidia.com/en-eu/gtc/>

Early Bird end July 10!

VIP Code: NVDMARGGRAF to get 25% off



THANK YOU



# NVIDIA GPU RESOURCES



Test Drive

[www.nvidia.com/testdrive](http://www.nvidia.com/testdrive)

---



GRID Website

[www.nvidia.com/grid](http://www.nvidia.com/grid)

---



GRID YouTube Channel

<http://tinyurl.com/gridvideos>

---



Questions? Ask on our Forums

<https://gridforums.nvidia.com>

---



NVIDIA Virtualization on LinkedIn

<http://linkd.in/QG4A6u>

---



Follow us on Twitter

[@NVIDIAVirt](https://twitter.com/NVIDIAVirt)