



# NVIDIA AUTONOMOUS DRIVING PLATFORM

Apr, 2017

Sr. Solution Architect , Marcus Oh

# Contents

Who is NVIDIA

Deep Learning in Autonomous Driving

Training Infra & Machine / DIGIT

DRIVE PX2

DRIVEWORKS

USECASE Example

Next Generation AD Platform



# NVIDIA

Founded in 1993 | CEO & Co-founder: Jen-Hsun Huang | FY16 revenue: \$5.01B | 9,500 employees  
| 7,300 patents | HQ in Santa Clara, CA |



**World's Largest  
GPU Computing Company**



**#1 in Gaming  
100M GeForce Gamers**



**#1 in Professional  
Workstations**

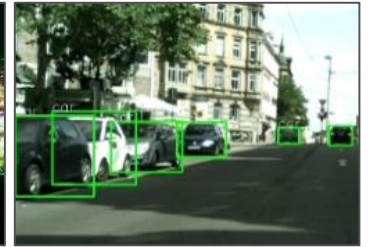
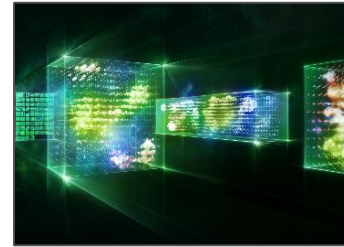
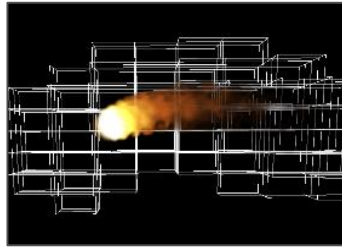
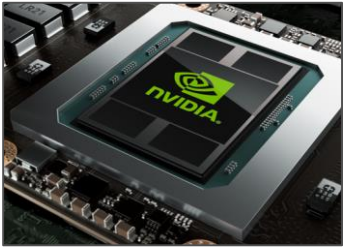
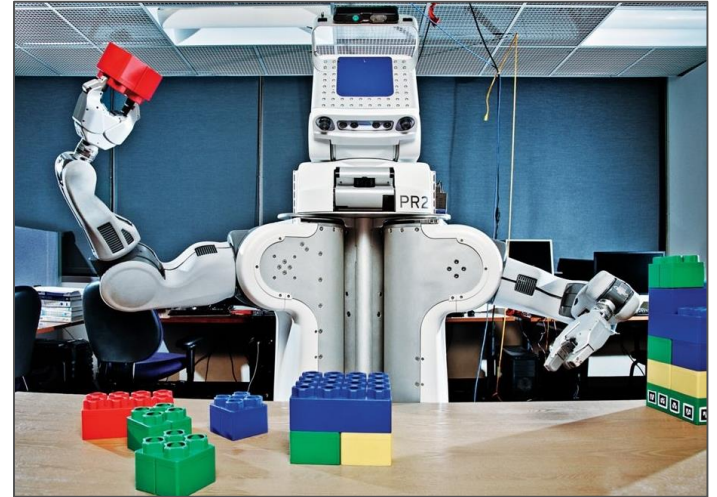


**#1 in HPC Accelerators  
70% of Top500**



**#1 Deep Learning  
Platform**

# NVIDIA – “THE AI COMPUTING COMPANY”



GPU Computing

Computer Graphics

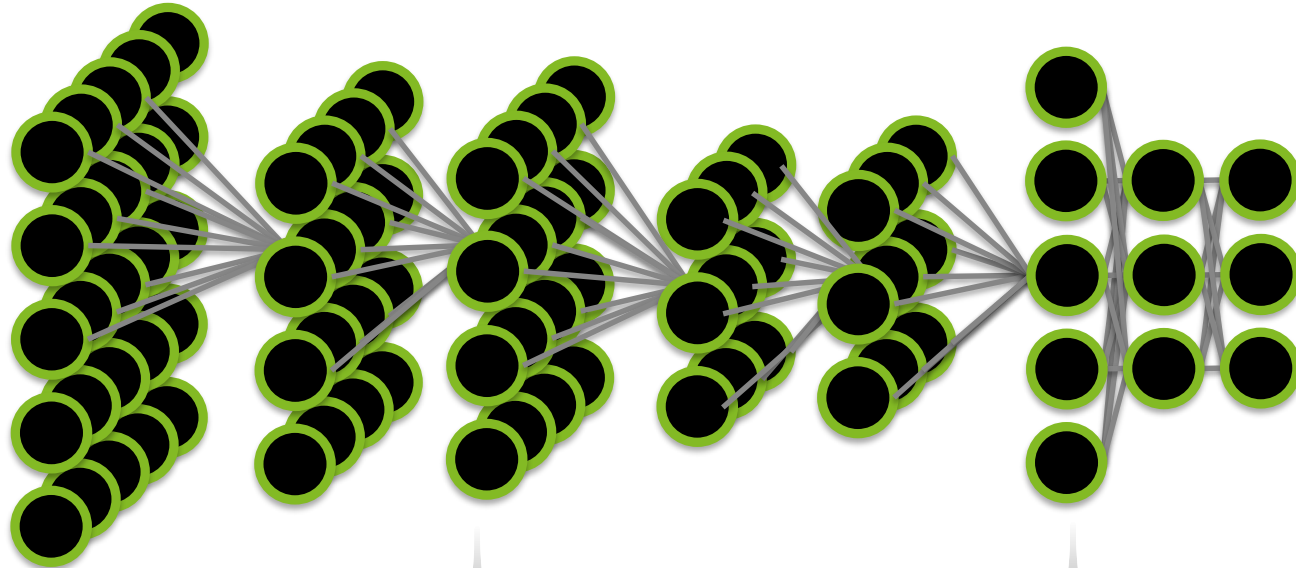
Artificial Intelligence

# DEEP LEARNING IN AUTONOMOUS DRIVING

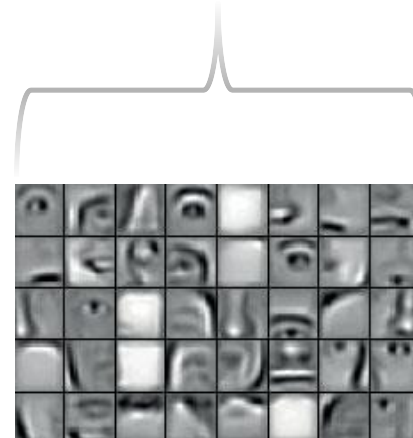
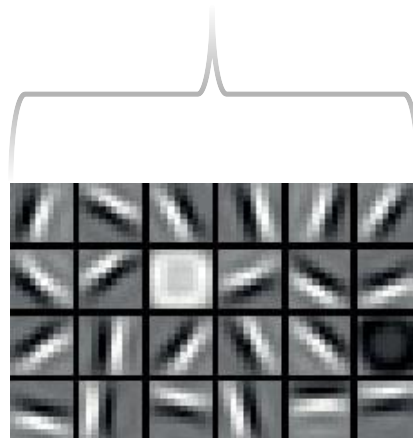
# WHAT IS DEEP LEARNING?



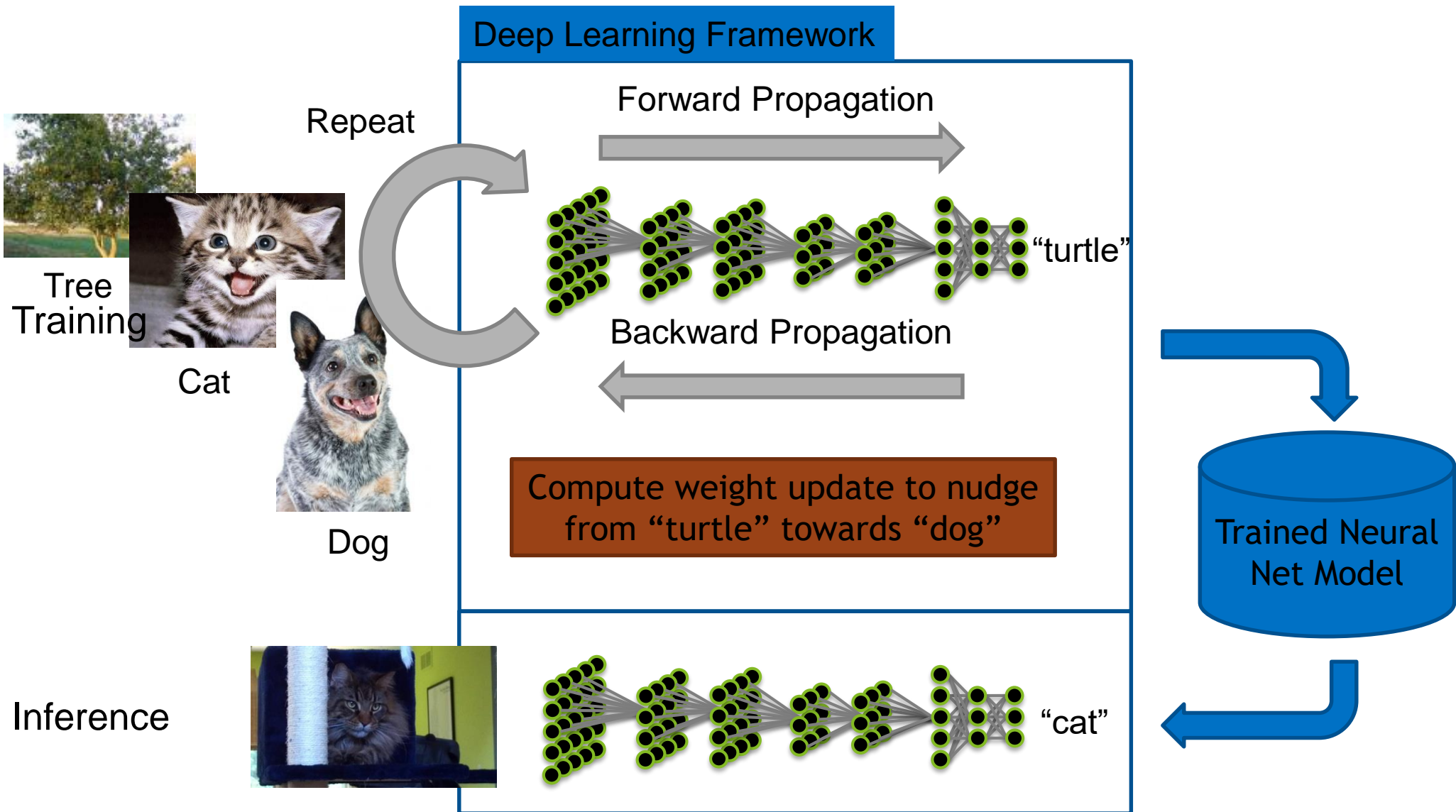
Input



Result



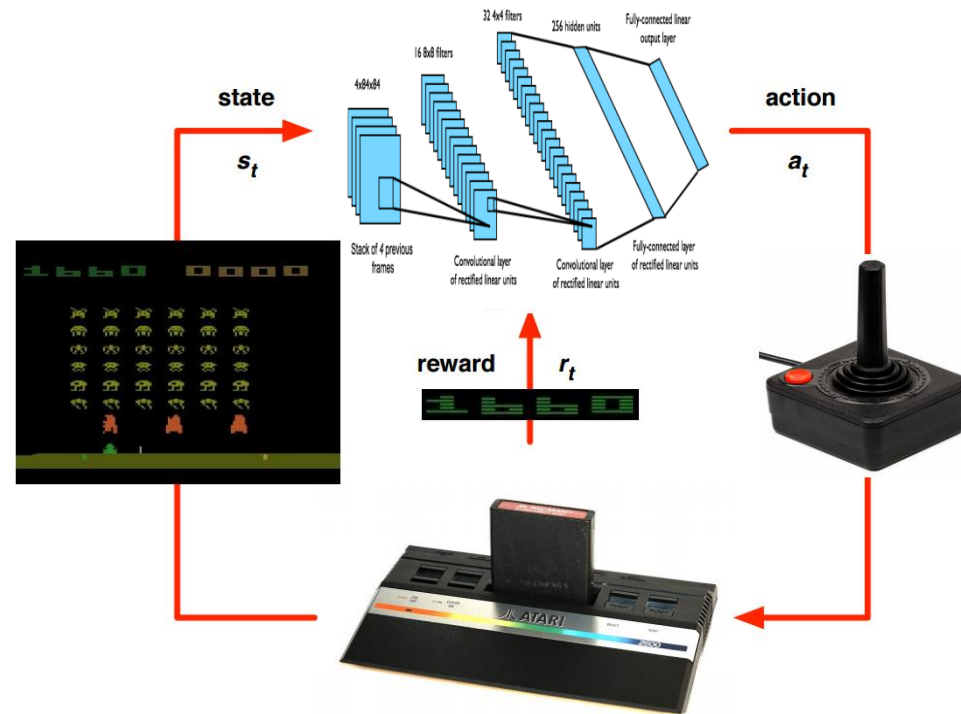




# REINFORCEMENT LEARNING

How's it work?

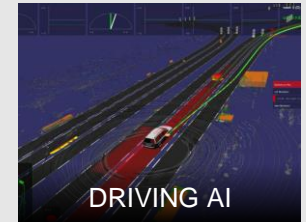
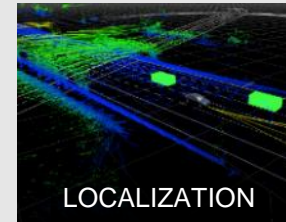
- ▶ A reinforcement learning agent includes:
  - ▶ **state** (environment)
  - ▶ **actions** (controls)
  - ▶ **reward** (feedback)
- ▶ A **value function** predicts the future reward of performing actions in the current state
  - ▶ Given the recent state, action with the maximum estimated future reward is chosen for execution
- ▶ For agents with complex state spaces, deep networks are used as **Q-value** approximator
  - ▶ Numerical solver (**gradient descent**) optimizes the network on-the-fly based on reward inputs



[github.com/dusty-nv/jetson-reinforcement](https://github.com/dusty-nv/jetson-reinforcement)

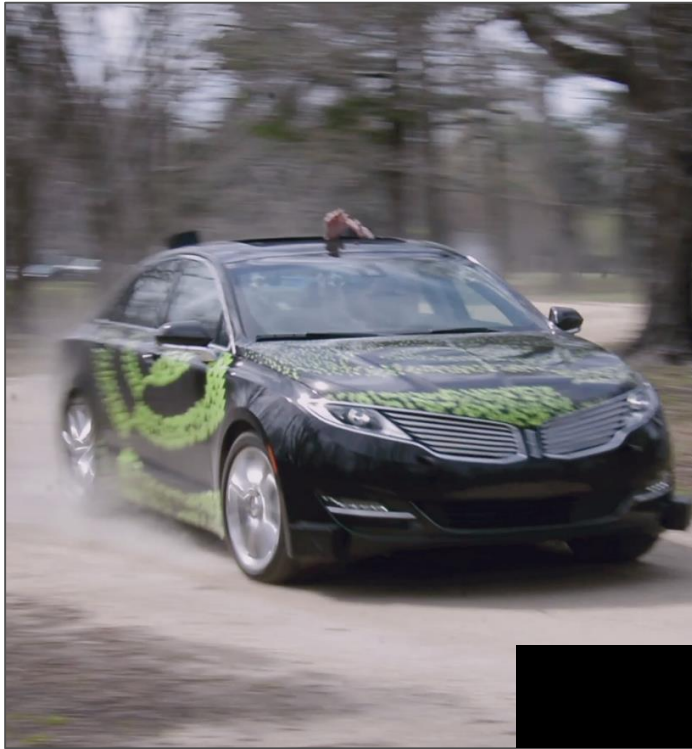


# SELF-DRIVING CARS ARE AN AI CHALLENGE



DEEP LEARNING

# NVIDIA AI SYSTEM FOR AUTONOMOUS DRIVING



Caffe

CNTK

KALDI

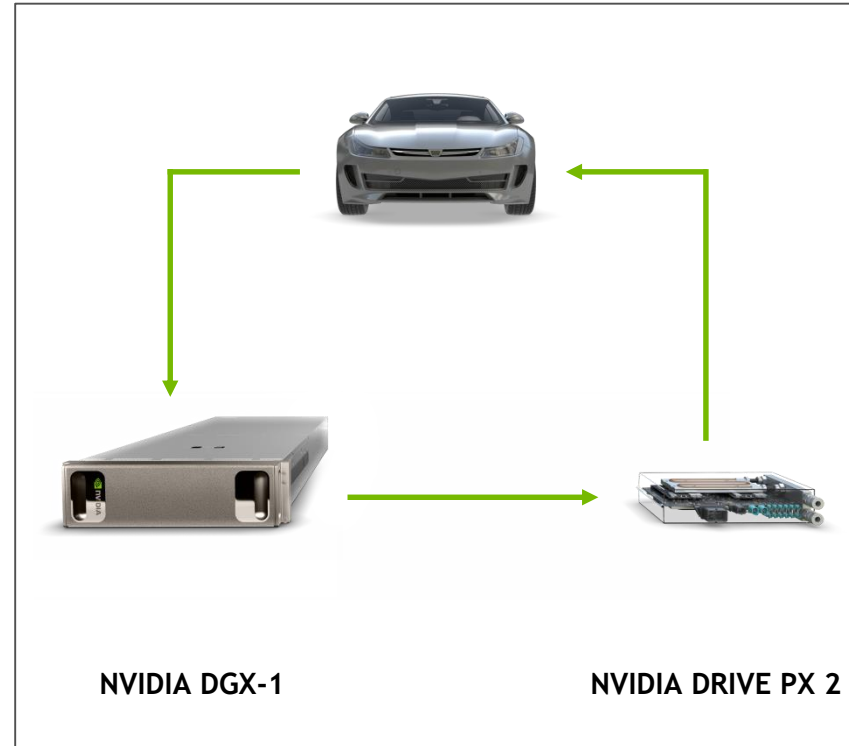


TensorFlow

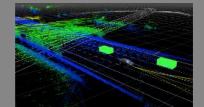
theano



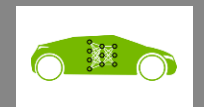
Training on  
DGX-1



MAPPING



LOCALIZATION



DRIVENET

Driving with  
DriveWorks

**TRAINING INFRA & MACHINE / DIGIT**

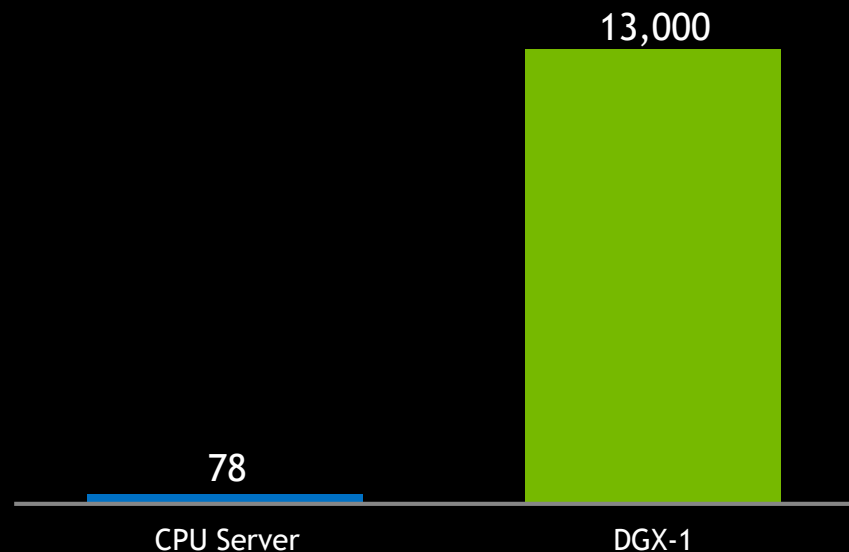
# 170X SPEED-UP OVER COTS SERVER

MICROSOFT COGNITIVE TOOLKIT SUPERCHARGED ON NVIDIA DGX-1



8x Tesla P100 | 170TF FP16 | NVLink hybrid cube mesh

**170x Faster**  
(AlexNet images/sec)

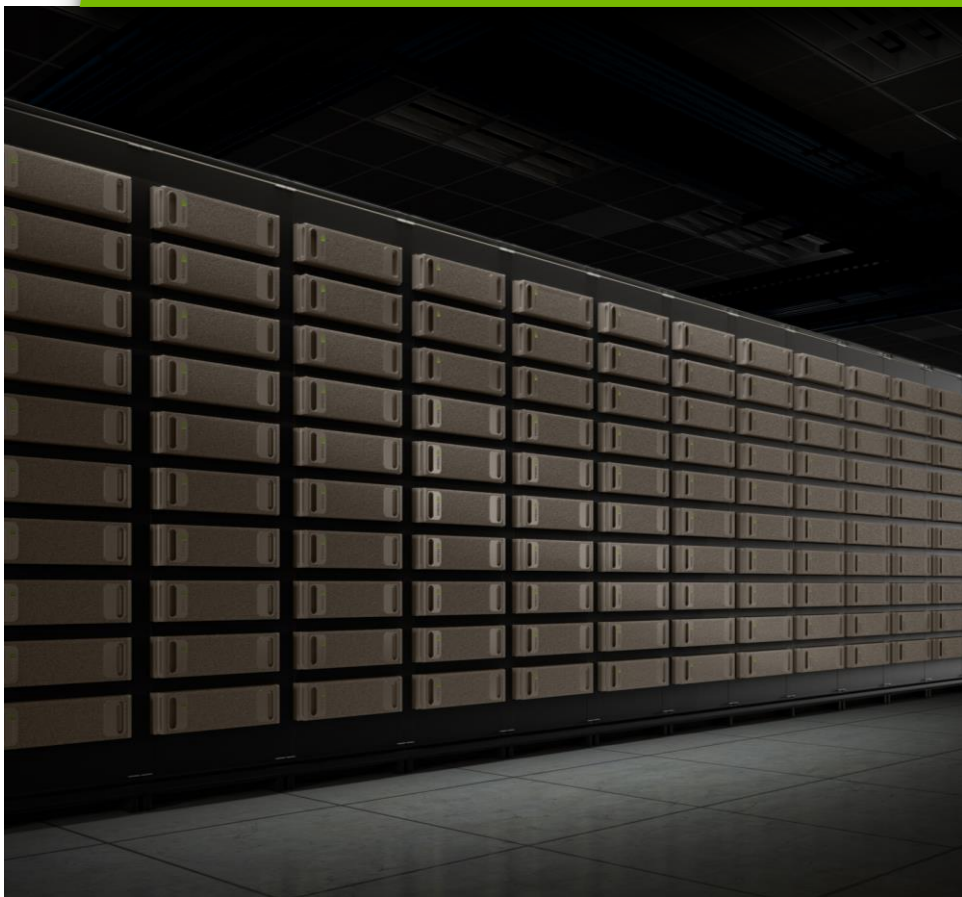


AlexNet training batch size 128, Dual Socket E5-2699v4, 44 cores CNTK 2.0b2 for CPU.  
CNTK 2.0b3 (to be released) includes cuDNN 5.1.8, NCCL 1.6.1, NVLink enabled



# INTRODUCING DGX SATURNV

124 NVIDIA DGX-1 “Rocket for Cancer Moonshot”



Fastest AI Supercomputer in TOP500

4.9 Petaflops Peak FP64

19.6 Petaflops Peak FP16



Most Energy Efficient Supercomputer

#1 Green500

9.5 GFLOPS per Watt



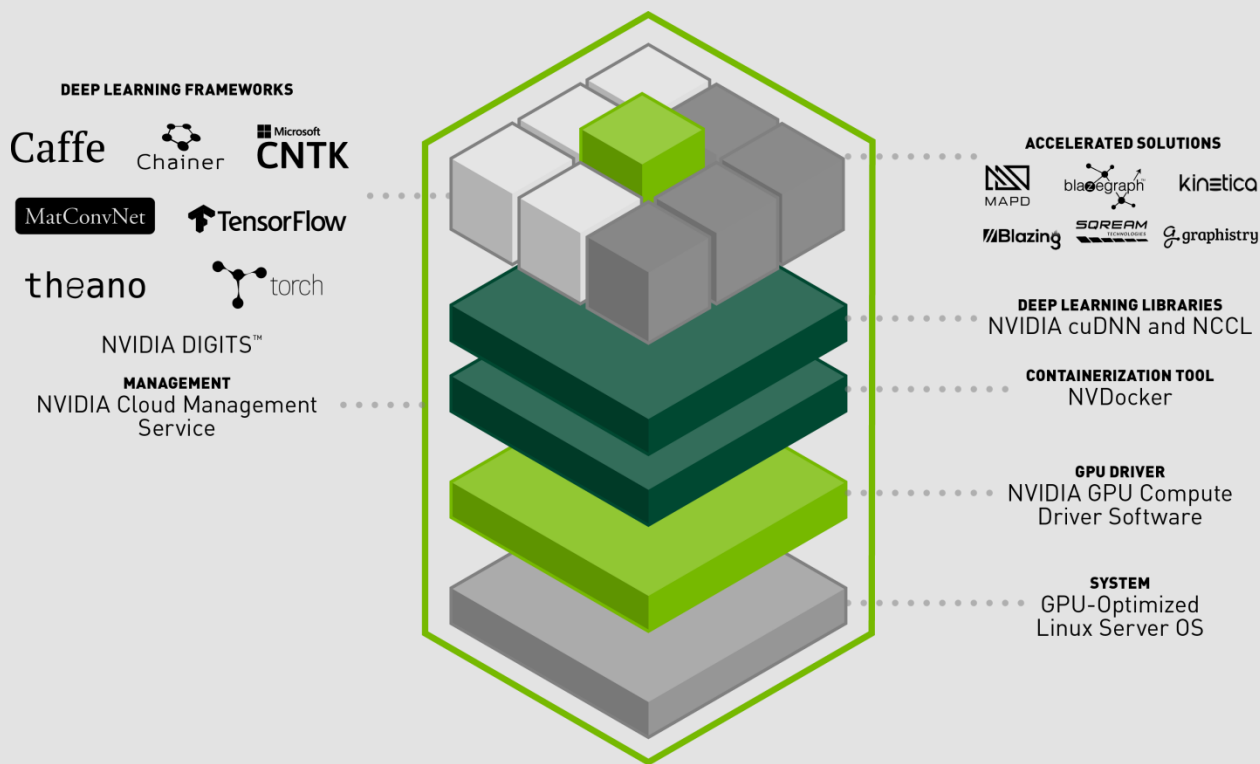
Rocket for Cancer Moonshot

CANDLE Development Platform

Common platform with DOE labs – ANL, LLNL, ORNL, LANL

# DGX STACK

Fully integrated Analytics and Deep Learning platform



Instant productivity — plug-and-play, supports every AI framework and accelerated analytics software applications

Performance optimized across the entire stack

Always up-to-date via the cloud

Mixed framework environments — baremetal and containerized

Direct access to NVIDIA experts

# DIGIT

## Deep Learning GPU Training System

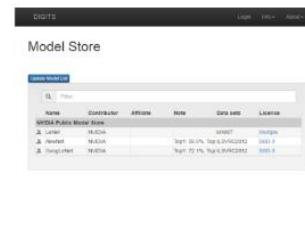
- ▶ Design, train and visualize deep neural networks for image classification, segmentation and object detection
- ▶ Download pre-trained models such as AlexNet, GoogLeNet and LeNet from the DIGITS Model Store
- ▶ Perform hyperparameter sweep of learning rate and batch size for improved model accuracy
- ▶ Schedule, monitor, and manage neural network training jobs, and analyze accuracy and loss in real time
- ▶ Import a wide variety of image formats and sources
- ▶ Scale training jobs across multiple GPUs automatically



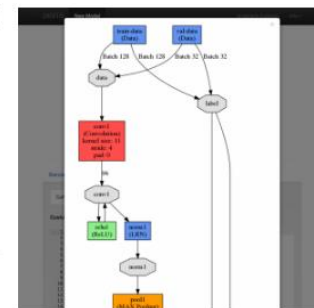
Import data for image classification and object detection neural networks



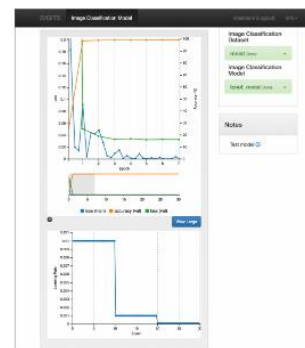
Schedule, monitor, and manage neural network training jobs



Download pre-trained models such as AlexNet, GoogLeNet and others from the DIGITS Model Store



Visualize deep neural network architectures



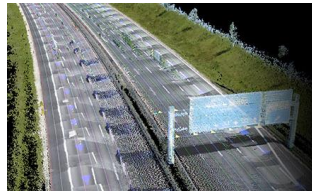
**DRIVE PX2**



# NVIDIA SELF-DRIVING BUILDING BLOCKS

Accelerate your development and deployment

## SELF DRIVING TECHNOLOGY



## SOFTWARE

PDK



cuDNN



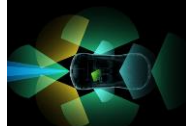
TensorRT



VPI

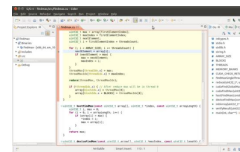


DriveWorks



## TOOLS

CUDA Development  
Nsight Eclipse Edition



Tegra System Profiler



## HARDWARE

DRIVE PX 2



DRIVE PX 2 AutoCruise



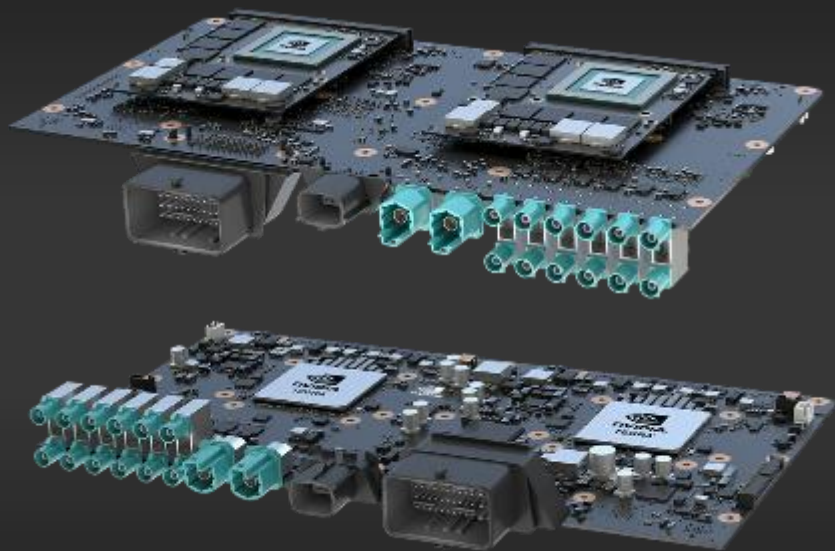
Tegra SOC, dGPU



A high-angle, close-up photograph of the NVIDIA DRIVE PX 2 Autocruise module. The module is a small, rectangular printed circuit board (PCB) populated with various electronic components. A large, square, silver-colored integrated circuit (the Tegra Parker SoC) is the central feature, with the NVIDIA logo clearly visible on its top surface. To the left of the main chip, there are several cylindrical capacitors and other smaller components. On the left edge of the module, there is a large, black, rectangular connector housing with multiple pins visible inside. Below this, two bright blue, L-shaped connectors are plugged into the board. The entire module is set against a dark, reflective background that shows a subtle reflection of the device.

# NVIDIA DRIVE PX 2 AUTOCRUISE

10W AI Car Computer | Passive Cooling | Automotive IO  
Multiple Cameras & Sensors | DriveWorks SW/SDK  
AI Highway Driving | Localization & HD Mapping  
Tegra Parker SoC – 1.3 TFLOPS, 6 CPU Cores, Integrated ISP



# NVIDIA DRIVE PX 2

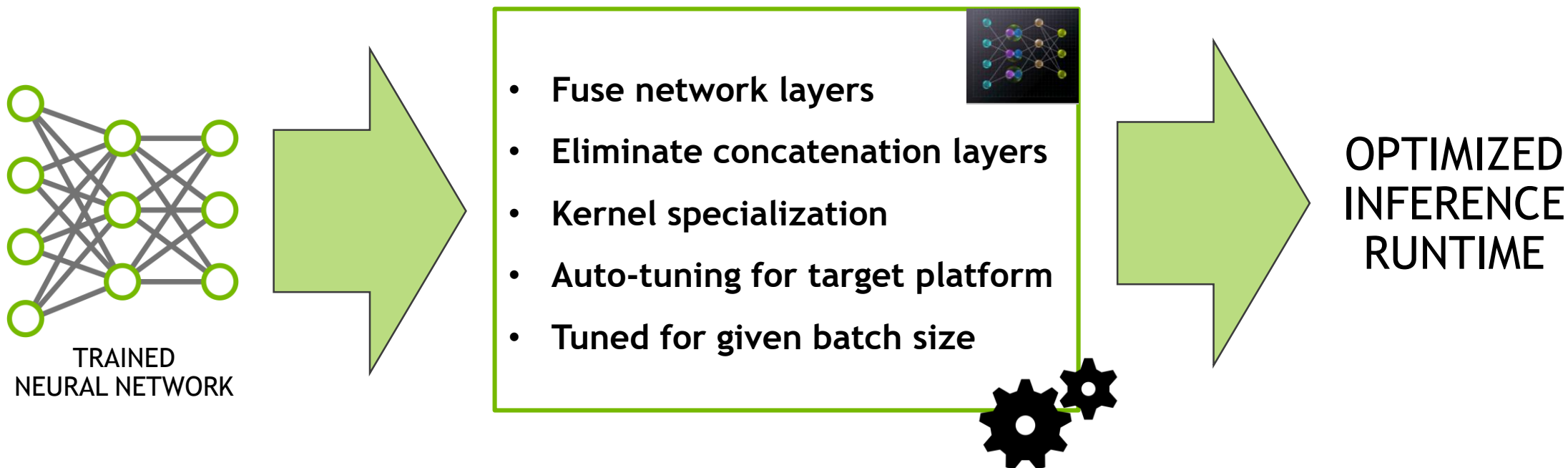
## AUTOCHAUFFEUR & FULLY AUTONOMOUS

Scalable from 1 to 4 Processors to Multiple DRIVE PX 2s  
— 2x Tegra Parker SoC, 2x Pascal dGPU, 8 TFLOPS, 24 DNN TOPs  
Up to 12 Cameras; plus LIDAR, Radar, Ultrasonic sensors  
DriveWorks SW/SDK | AI Perception | Localization & Mapping



# TensorRT: INFERENCE ENGINE

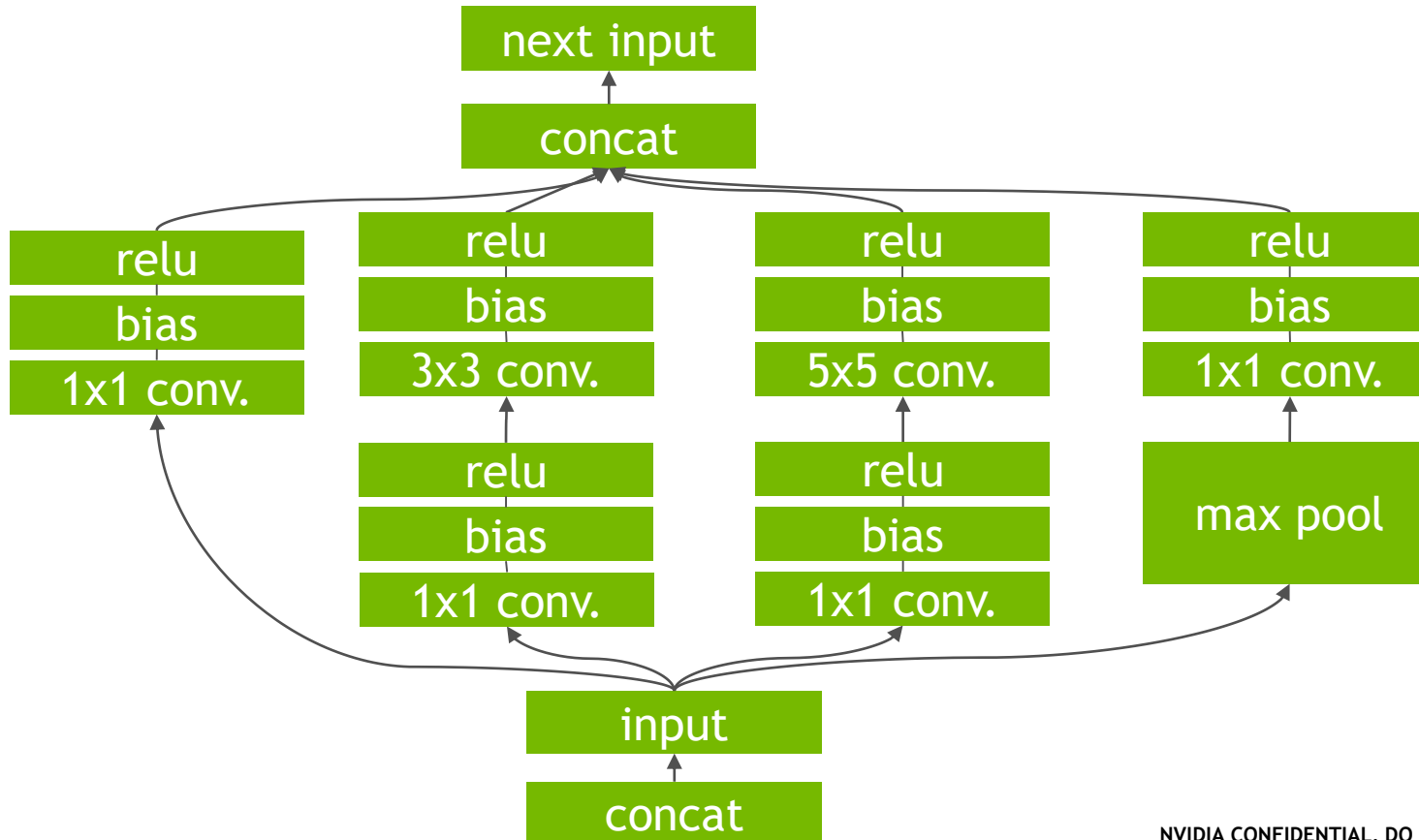
## Optimizations



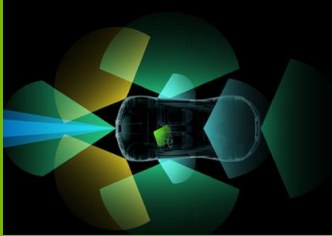


# GRAPH OPTIMIZATION

Unoptimized network



**DRIVEWORKS**



# DRIVEWORKS

Software Development Kit (SDK) for Autonomous Driving

Process sensor data through Perception, Mapping, Localization, and Path Planning steps

Provides a rich set of functionalities:

- Sensor Abstraction Layer (SAL)

- Algorithm Modules, DNNs

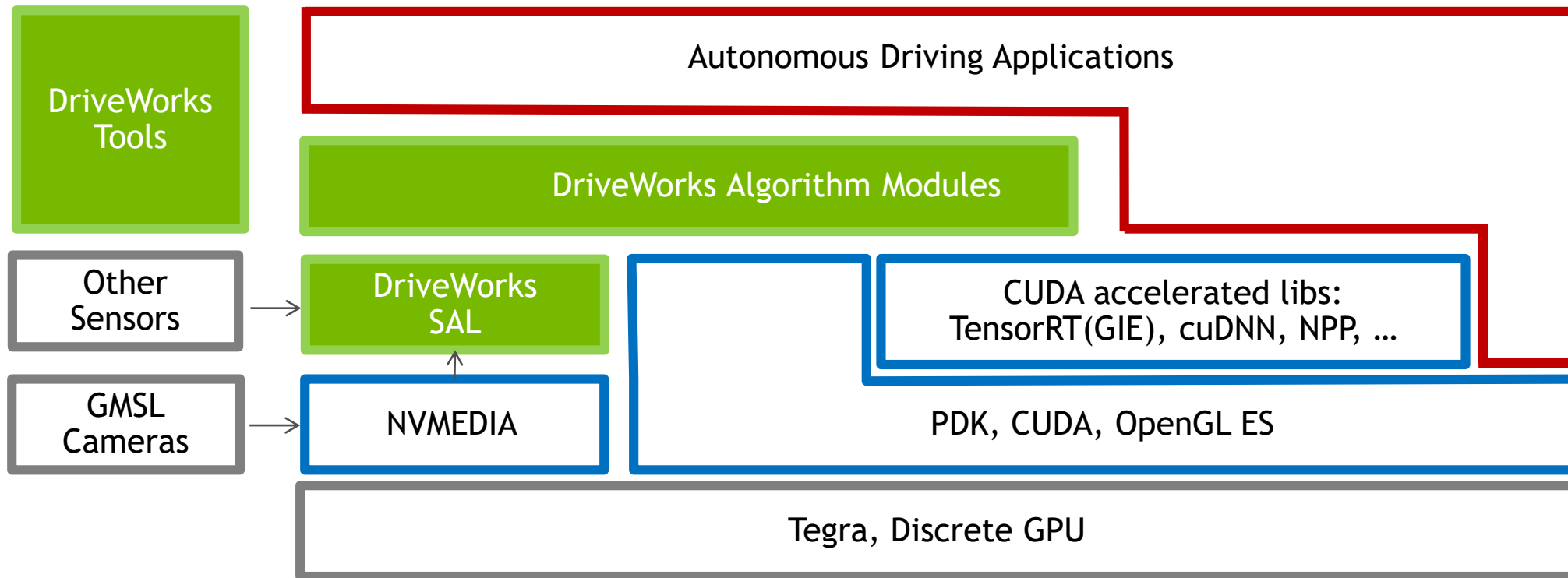
- Applications

- Tools for sensor setup and management

Flexible, modular and optimized for GPU

Runs on top of OS, CUDA/CuDNN, TensorRT, VPI

# SOFTWARE STACK WITH DRIVEWORKS



HW

Linux PDK

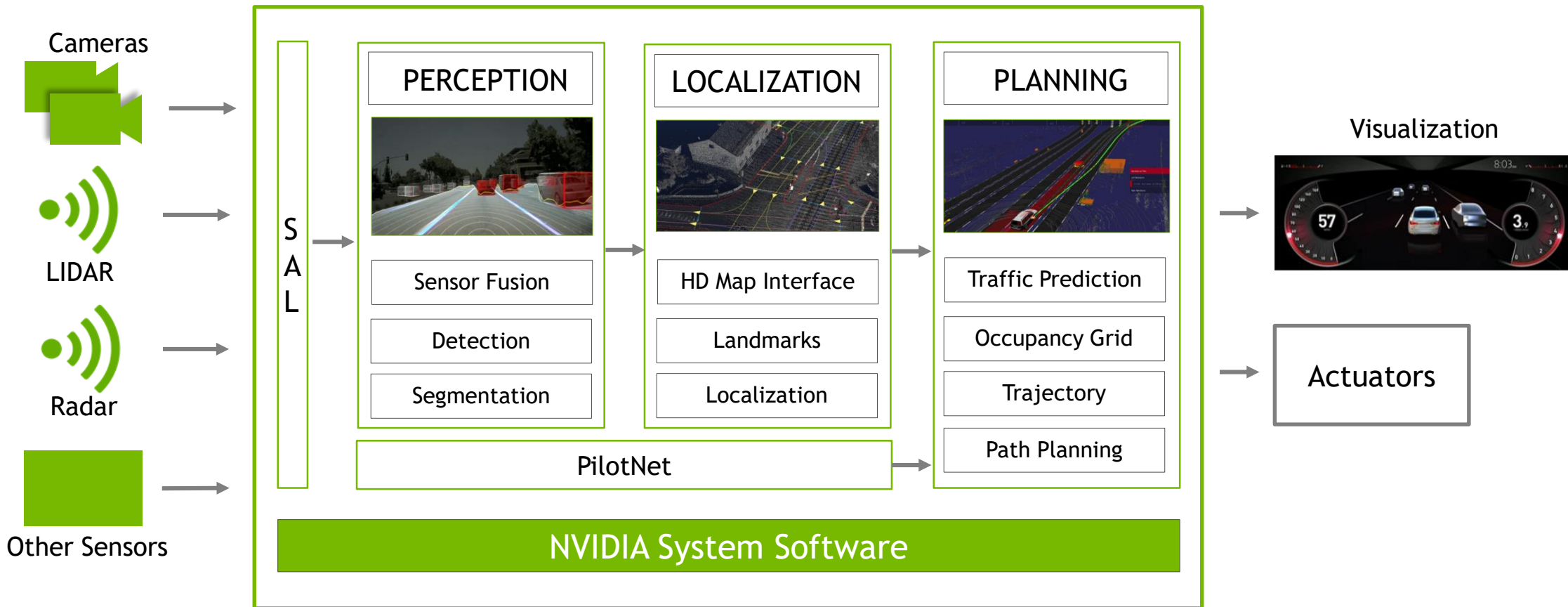
DriveWorks

OEM/Sample  
Applications

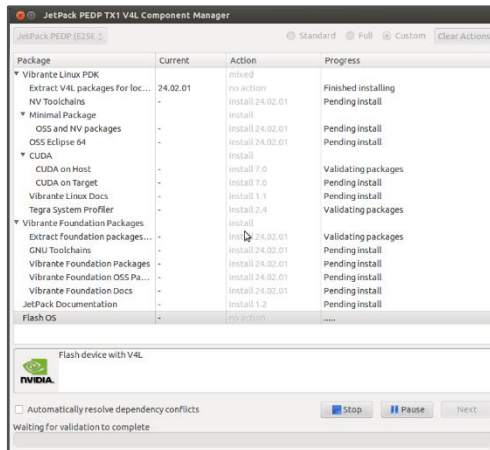


# DRIVEWORKS PROCESSING PIPELINE

End-to-End processing for Autonomous Driving



# NVIDIA DEVELOPER TOOLCHAIN

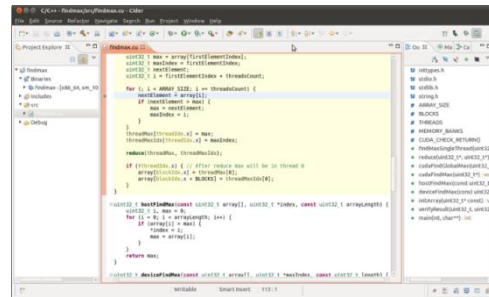


## DriveInstall

Easy installation

Devkit flashing

Sets up development environment



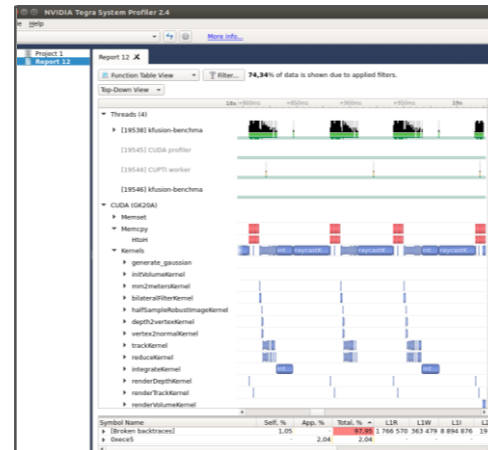
## NVIDIA® Nsight™ Eclipse Edition

CUDA build management

CUDA kernel debugging and profiling

CPU and GPU debugging

Memory checker



## Tegra System Profiler

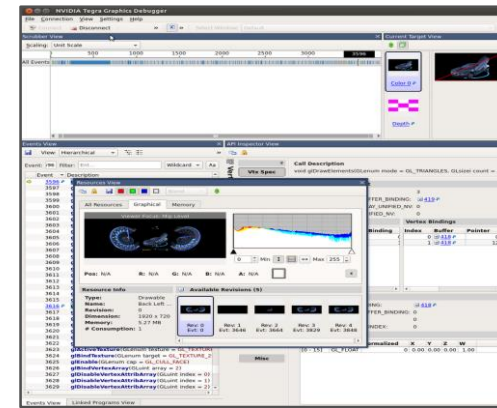
CPU sampling profiler

Application Trace

CUDA API & GPU trace

OpenGL ES API & GPU trace

Code decoration API/NVTX



## Tegra Graphics Debugger

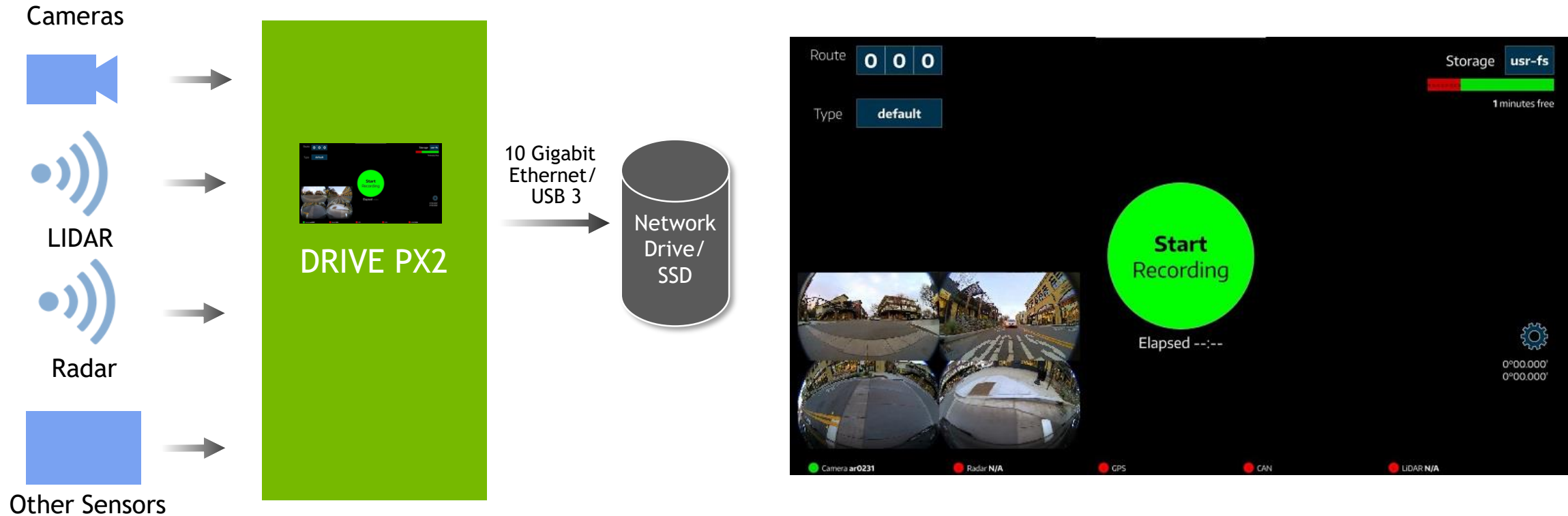
Performance monitoring

Frame debugging

Frame profiling

# USECASE EXAMPLES

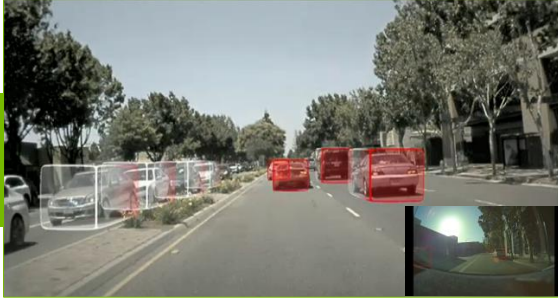
# TOOLS - DATA ACQUISITION



Intuitive UI to enable simultaneous capture data from sensors

# DEEP NEURAL NETWORKS IN DRIVEWORKS

DriveNet



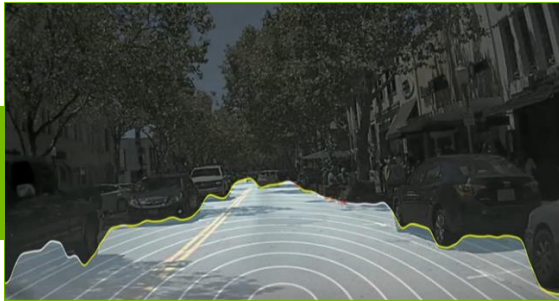
Multi-class detection: Cars, Pedestrian, Bicycles  
Upcoming: Lanes, Traffic signs

LaneNet



Lane detection

OpenRoadNet



Free space detection

PilotNet



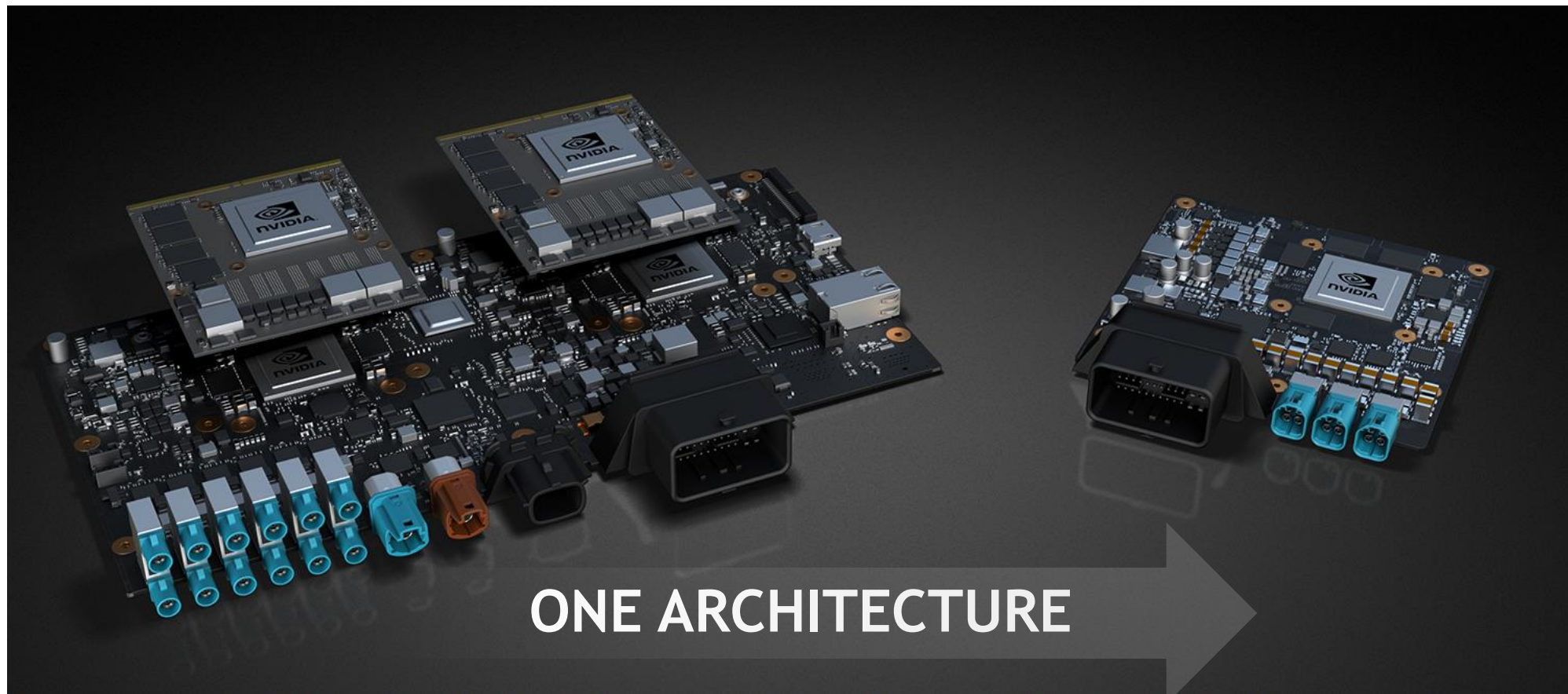
End to end in-lane driving



**NEXT GENERATION AD SOLUTION**

# INTRODUCING XAVIER

## AI SUPERCOMPUTER SOC



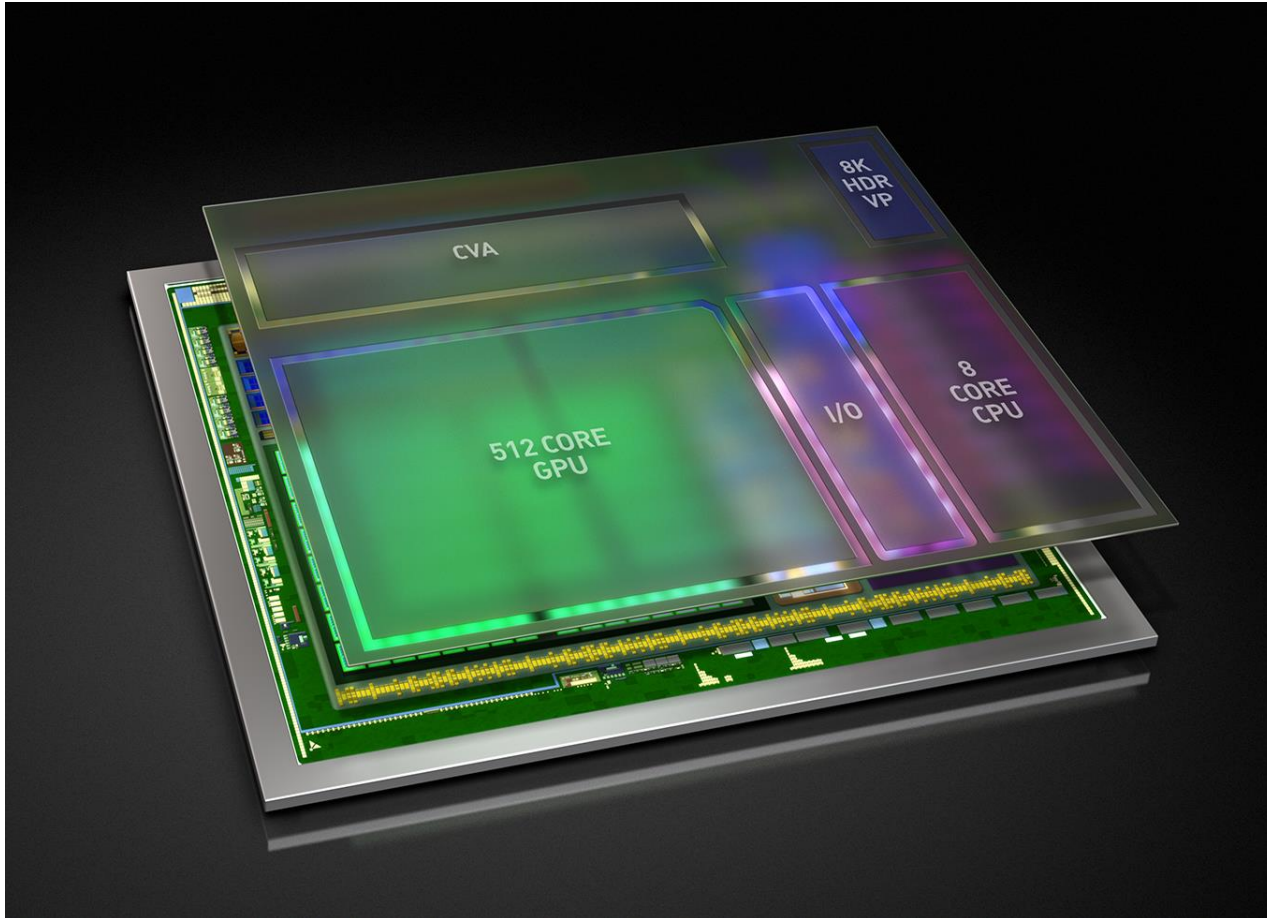
DRIVE PX 2

XAVIER

PARKER + 2 PASCAL GPU | 20 TOPS DL | 120 SPECINT | 80W 20 TOPS DL | 160 SPECINT | 20W

# INTRODUCING XAVIER

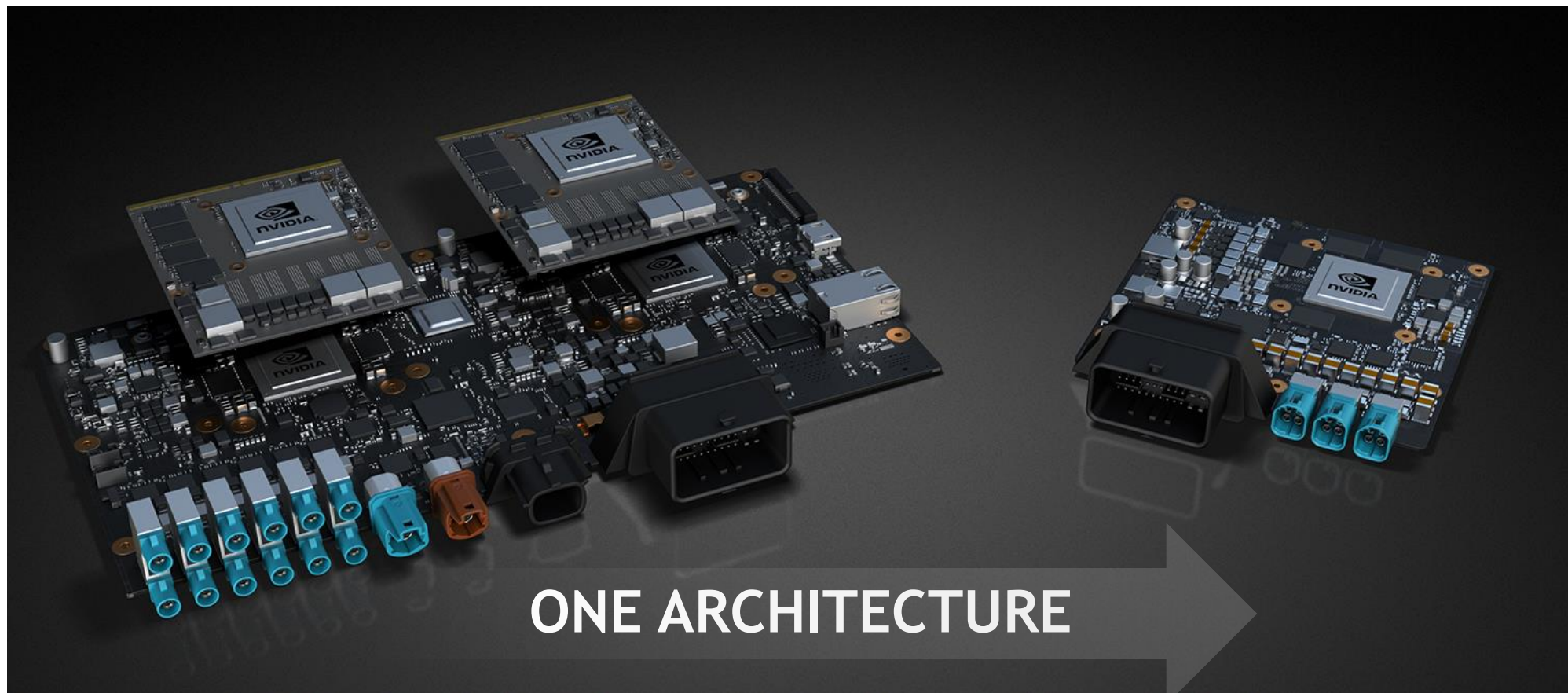
## AI SUPERCOMPUTER SOC



7 Billion Transistors 16nm FF  
8 Core Custom ARM64 CPU  
512 Core Volta GPU  
New Computer Vision Accelerator  
Dual 8K HDR Video Processors  
Designed for ASIL C Functional Safety

# INTRODUCING XAVIER

## AI SUPERCOMPUTER SOC



### DRIVE PX 2

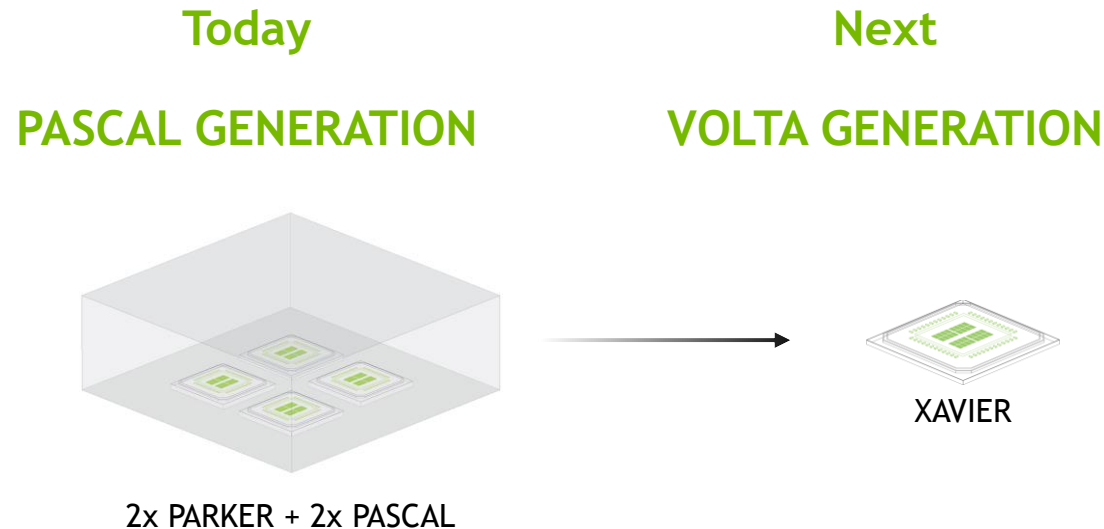
### XAVIER

2 PARKER + 2 PASCAL GPU | 20 TOPS DL | 120 SPECINT | 80W

20 TOPS DL | 160 SPECINT | 20W

# THE NEXT GENERATION

## DRIVE PX 2 IS THE TIME MACHINE TO XAVIER



# SUMMARY

End-to-end Autonomous driving SDK

Targeted for both rapid prototyping and production

Integrated with OS, TensorRT, CUDA/CuDNN and VPI to offer comprehensive feature set

Optimized for Tegra SOC and NVIDIA GPU

NVIDIA support available