A MISSION CONTROL FOR SPACESHIP EARTH

KICK OFF WORKSHOP
ESA ESRIN, ROME
25th - 29th June ‘18

RESEARCH SPRINT
30th June - 17th August
EMBED SUPERNOVA VIDEO
RE-IMAGINING DATA SCIENCE WORKFLOW
Open Source, End-to-end GPU-accelerated Workflow Built On CUDA

cuDF
Data preparation / wrangling

cuML
Optimized ML model training

Visualization
Data visualization libraries

WWW.RAPIDS.AI
DOWNLOAD AND DEPLOY

Source available on Github | Container available on NGC and Dockerhub | PIP available soon

GitHub | NGC | docker | CONDA | Python Package Index

Source code, libraries, packages

On-premises

Cloud
XAVIER

World’s First Autonomous Machines Processor

- 16 Lane CSI
  - 109 Gbps CPHY 1.1
  - 1Gb Ethernet
- DLA
  - 5.7 TFLOPS FP16
  - 11.4 TOPS INT8
- Multimedia Engines
  - 1.2 GPIX/s Encode
  - 1.8 GPIX/s Decode
  - 4 GPIX/s Video Image Compositor
  - Vision Accelerator
    - 1.7 TOPS
- Stereo & Optical Flow Engine
  - 2x 3.1 TOPS
- Volta Tensor Core GPU
  - FP32 / FP16 / INT8 Multi-Precision
  - 512 CUDA Tensor Cores
  - 2.8 CUDA TFLOPS (FP16)
  - 22.6 Tensor Core DL TOPS
- Industry Standard High-Speed IO
  - PCIe Gen4 Root and Endpoint
  - USB 3.1 gen2 Host and Device
  - UFS 2.1 Embedded Storage
- ISP
  - 2.4 GPIX/s
  - Native Full-range HDR
  - Tile-based Processing
- Carmel ARM64 CPU
  - 8 Cores
  - 10-wide Superscalar
  - 21 Specint2K6 (est.)
- 256-Bit LPDDR4X
  - 137 GB/s

Most Complex SOC Ever Made | 9 Billion Transistors, 350mm², 12FFN | ~8,000 Engineering Years

©2018 NVIDIA CORPORATION
XAVIER DLA
NOW OPEN SOURCE

Command Interface

Tensor Execution Micro-controller

Input DMA (Activations and Weights) → Unified 512KB Input Buffer → Sparse Weight Decompression → Native Winograd Input Transform → MAC Array (2048 Int8 or 1024 Int16 or 1024 FP16) → Output Accumulators → Output Postprocess or (Activation Function, Pooling etc.) → Output DMA

Memory Interface

WWW.NVDLA.ORG
JETSON TX2
SUPERCOMPUTER FOR AI AT THE EDGE

2 Core i7 PCs in <10W
256 CUDA cores
>1 TFLOPS

cuDNN, TensorRT
CUDA
Linux, ROS
COMPUTATIONAL SCALE REQUIRED

3 million labeled images
1 DGX-1 trains 300k labeled images on 1 DNN in 1 day
10 DNNs required for self-driving
10 parallel experiments at all times
100 DGX-1 per car
TRACK ML

Invaluable learnings from crowd-sourcing WW talent
DGX POD ARCHITECTURE

A single data center rack containing up to 9x NVIDIA DGX-1 servers, storage, networking & NVIDIA AI software.

- Nine DGX-1 servers
- 12 storage servers
- 10 GbE (min) storage & management switch
- Mellanox 100 Gpps intra-rack high speed network switches.