Adaptable acceleration for HPC

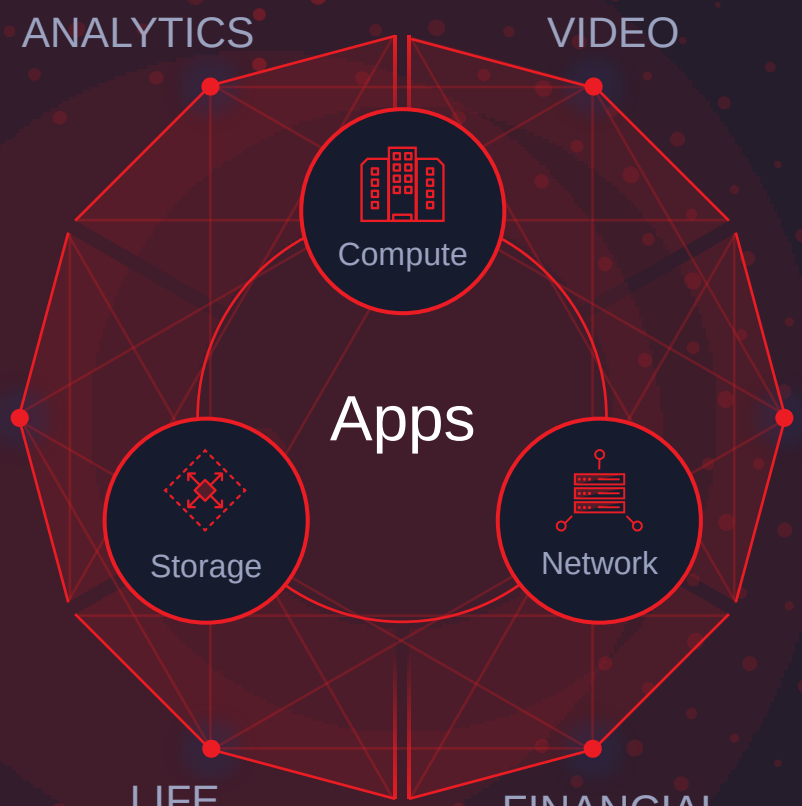
Jens Stapelfeldt – DataCenter BDM EMEA

September, 2019

JStapelfeldt@Xilinx.com
LinkedIn: www.linkedin.com/in/JensStapelfeldt

Russian Supercomputing Days
international supercomputing conference

Xilinx intro to Era of Reconfigurable Accelerators



Edge AI

+

Cloud AI + HPC

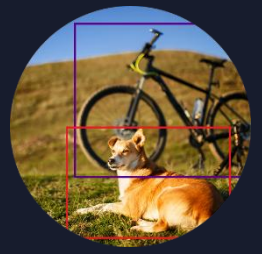
Algorithm Diversity and Fast Evolution

- **See Xilinx Model Zoo on github for free pre trained models: https://github.com/Xilinx/AI-Model-Zoo**

APPLICATIONS

Classification

Object Detection



Speech Recognition

Data Analytics

Recommendation Engine

Anomaly Detection

CNN

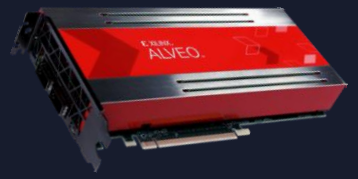
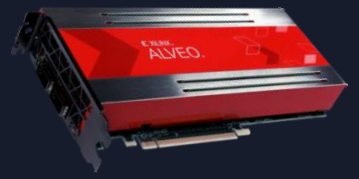
RNN, LSTM

RF, LR

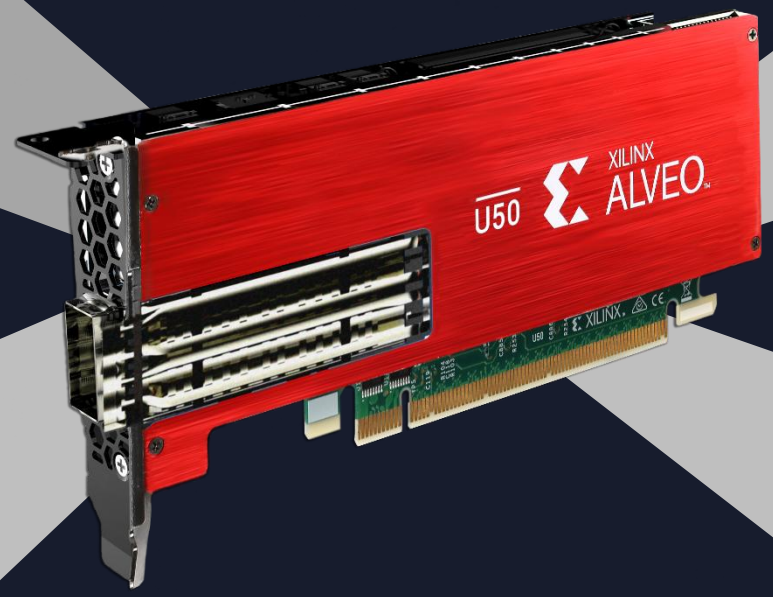
MLP

Diverse AI models and Neural Networks (NN's) for a broad range of applications

Xilinx Alveo Product Lineup

ALVEO™ U200	ALVEO™ U250	ALVEO™ U280	ALVEO™ U50 (new)
			
UltraScale+ Architecture	UltraScale+ Architecture	UltraScale+ Architecture	UltraScale+ Architecture
1,182k LUTs	1,728k LUTs	1,304k LUTs	872k LUTs
Dual slot, full height	Dual slot, full height	Dual slot, full height	Single slot, half height
64GB DDR, 77GB/sec	64GB DDR, 77GB/sec	8GB HBM2, 460GB/sec	8GB HBM2, 460GB/sec
PCIe Gen3	PCIe Gen3	PCIe Gen3, Gen4, CCIX	PCIe Gen3, Gen4, CCIX
2x QSFP 28 (100GbE)	2x QSFP 28 (100GbE)	2x QSFP 28 (100GbE)	1x QSFP 28 (100GbE)
< 225W	< 225W	< 225W	< 75W

Alveo U50 – Low Profile Acceleration for Key Applications



Fintech

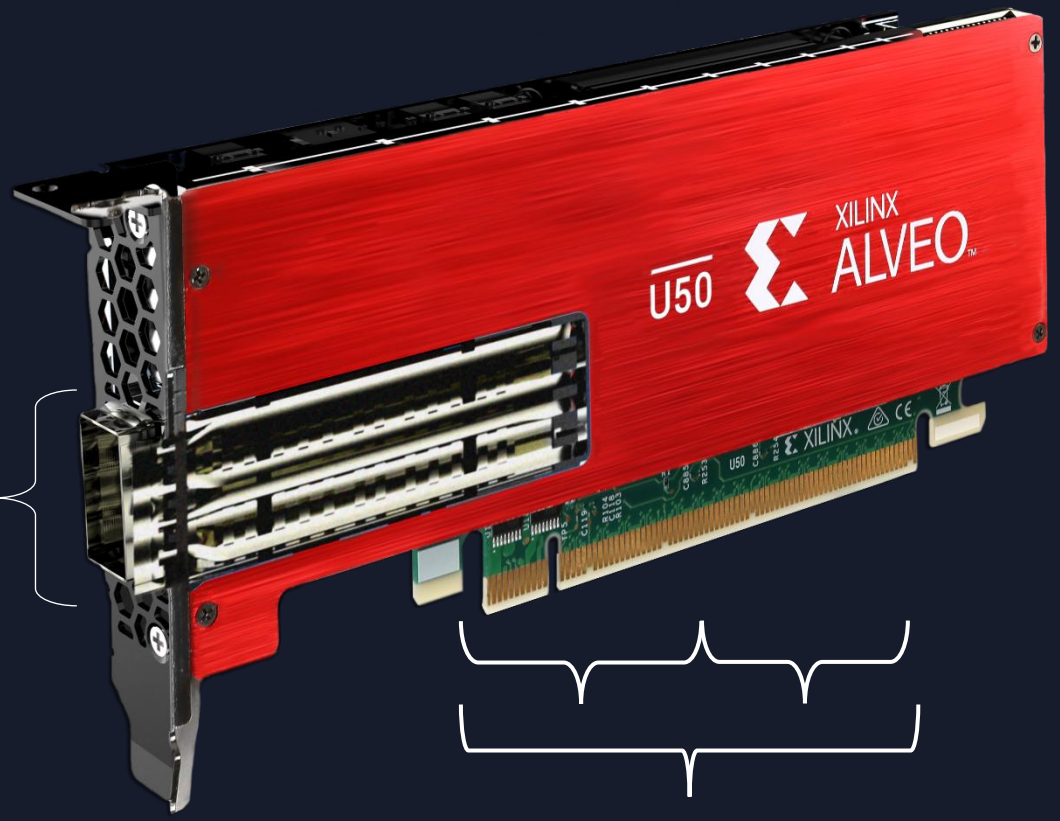
Storage

Machine Learning

HPC Acceleration

Database

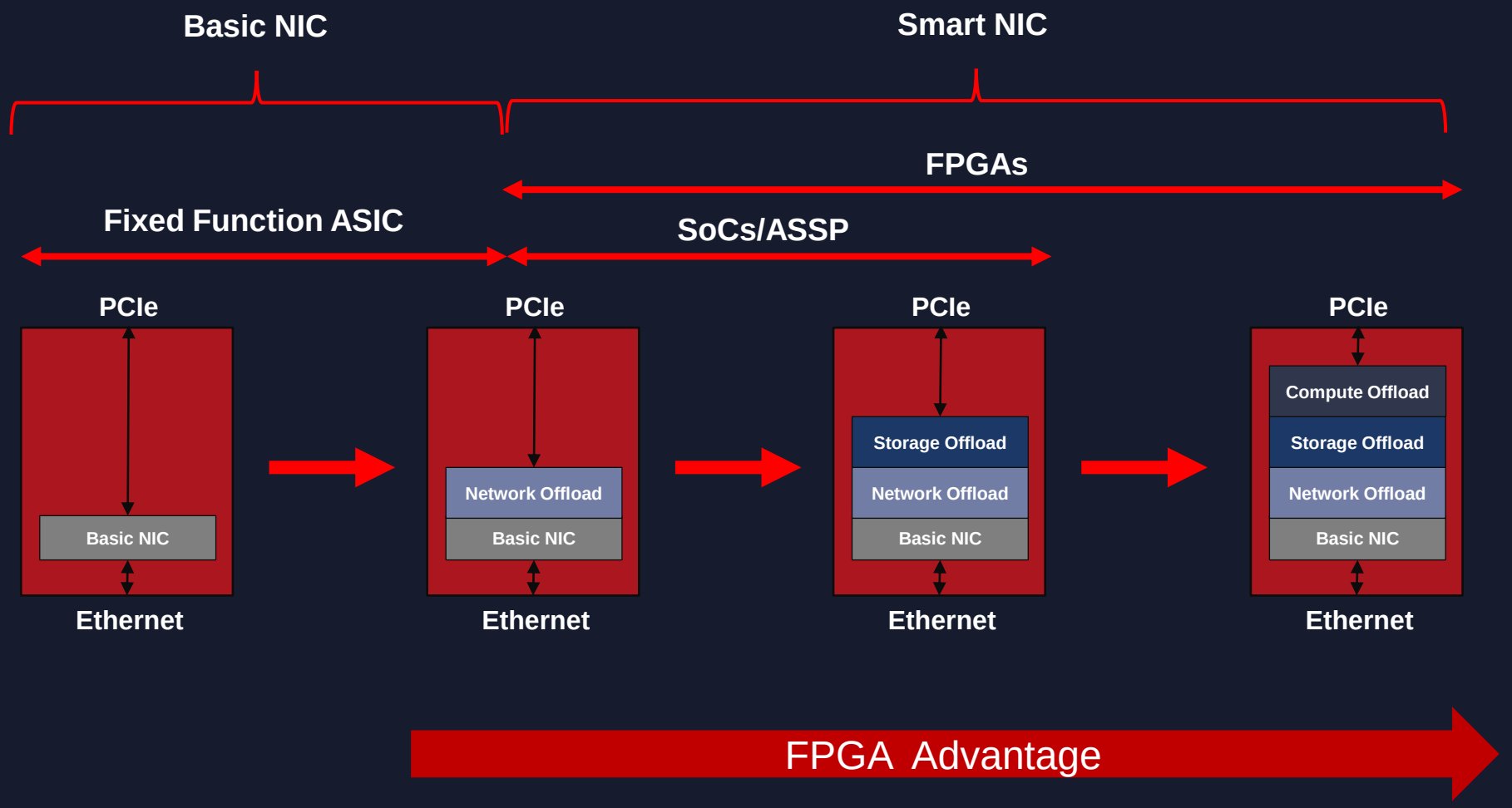
Alveo U50 – Single Slot, HHHL Scalable Acceleration Card



Board	U50
Primary Application	Fintec +Storage + Database + ML
FPGA Design	XCU50 - Dual SLR
Package	FSVH2104
CCIX	Yes
Device Speed/Voltage	2L
Width	Single slot
Form Factor (Passive)	HHHL
Memory	8 GB HBM
Memory Config	Dual Stack, 32 pseudo-ports, 460 GB/s
PCIe	2x Gen4x8, 1x Gen3x16, CCIX
Network I/F	2x SFP-DD (ES) 1X QSFP28 (Production)
Thermal	Passive
Power (Max TDP)	50 - 75W
KLuts	850

Product Schedule	
ES Shipping	**Mid-Sept 2019**
Production Shipping	**Q4 2019**

All-FPGA SmartNICs for Diverse Workloads

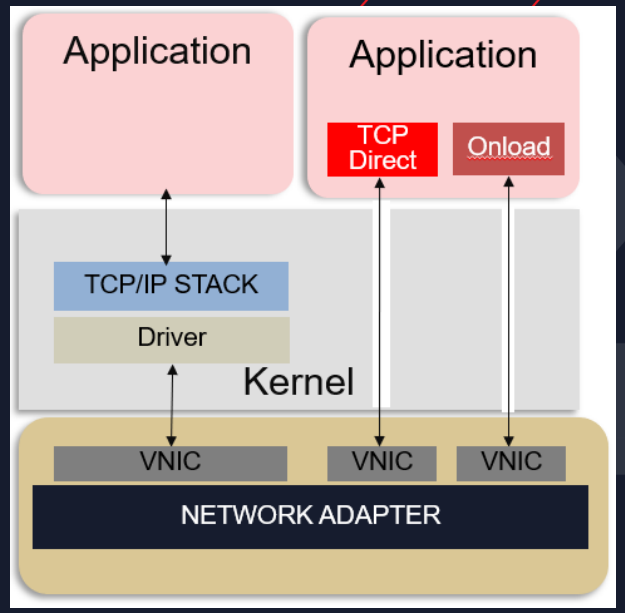


XILINX

Xilinx SmartNIC Acceleration Platform

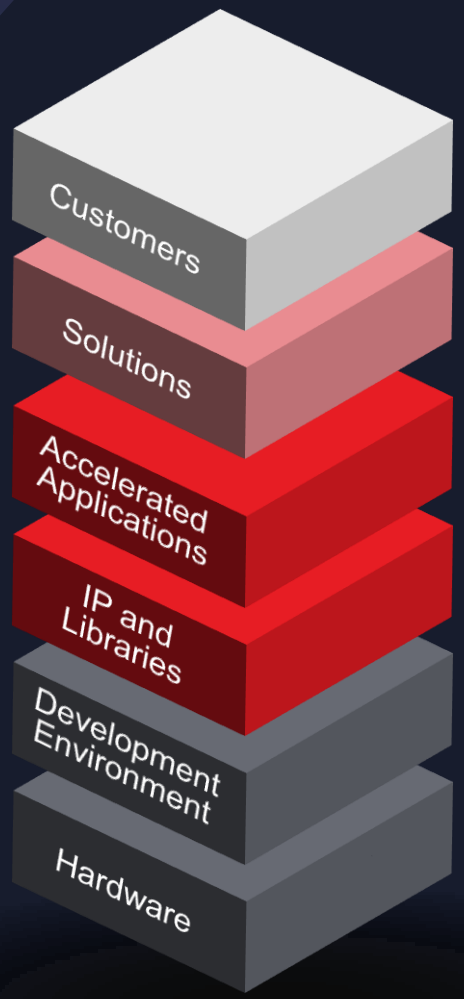
Acquisition of Solarflare on Sep. 2019 completed

- Built on top of SoftNIC
  - Standard NIC Offloads: Checksum, LSO, TSO, RSS, VMQ, VXLAN, NVGRE, etc
  - All standard host drivers provided (Linux, Windows, VMWare, DPDK, etc)
  - Built-in acceleration for common networking functions such as OVS
- Plug-in architecture which provides the ability to easily add Functions
  - NIC functions will be provided in the FPGA "shell"
  - Plug-in functions inserted into dynamic region
  - Can be customer developed or 3rd party IP
- High Performance, Low Latency
  - Target Performance is 100Mpps for Basic NIC
  - Match/Action Forwarding at 75Mpps with Xilinx Lookup IP
  - Much lower latency than SoC based solutions



XILINX.

ALVEO Solution Stack



End Customers

Solution Providers

App & IP Developers

Channel Partners

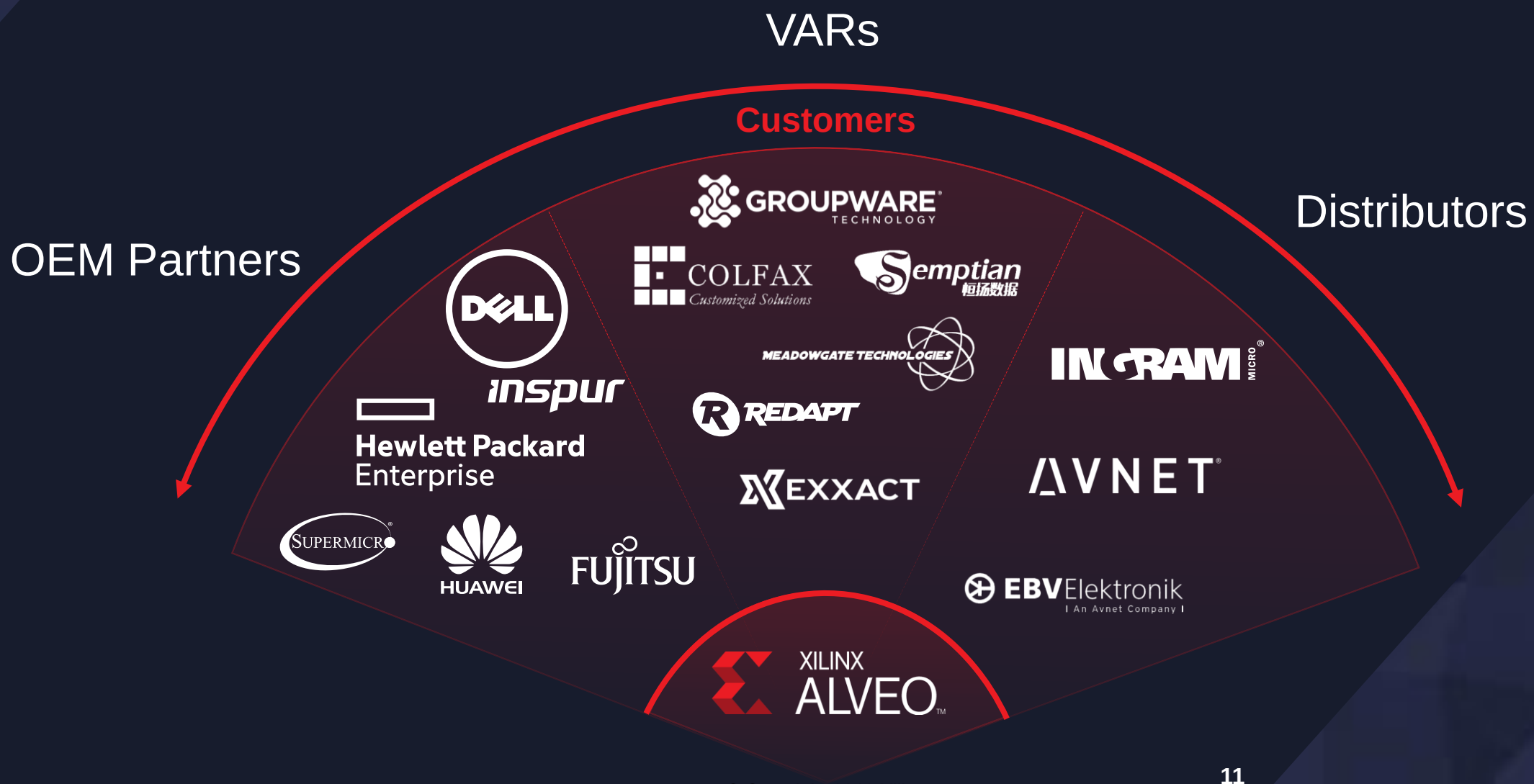
Data Analytics	Video & Image Processing	Machine Learning	Life Science & HPC	Financial Computing
Titan IC, XELERA, boon, CTACCEL, NGCODEC, V-NOVA, DEEPOLY, skreens, Nextera Video, edico genome, swarm64, bigstream, Mipsology, ALGO LOGIC, rENIAC, MAXELER Technologies, falcon COMPUTING, byte LAKE, MEGH COMPUTING, inaccel, DEEPHI, PLUNIFY, BLACKLYNX, VITESSE DATA, BigZetta Systems, LegUp, mle				
aws, Alibaba Cloud, Baidu 百度, Tencent Cloud, HUAWEI, NIMBIX, SUPERMICRO, inspur, DELL				

CLOUD

ON-PREMISE

XILINX

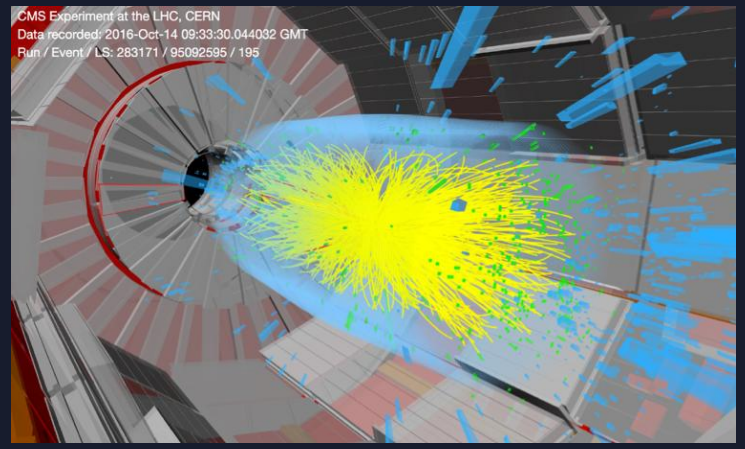
Growing Solution Partner Network



AI Accelerated Dark Matter Search (CERN)

Real-time ML Inference + Sensor pre-processing

Achieving 100ns Inference Latency on 150 Terabytes/Second Data Rates
Unachievable by CPUs & GPUs



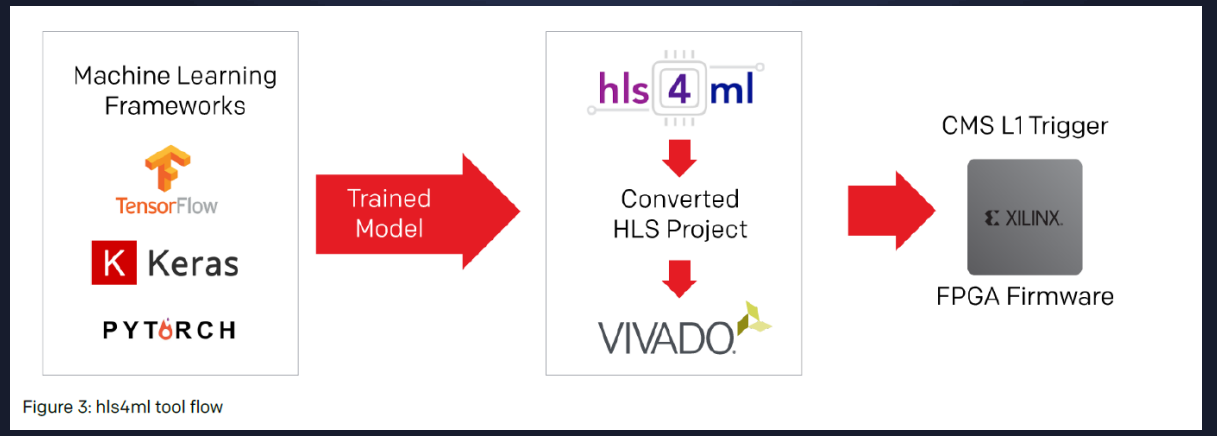


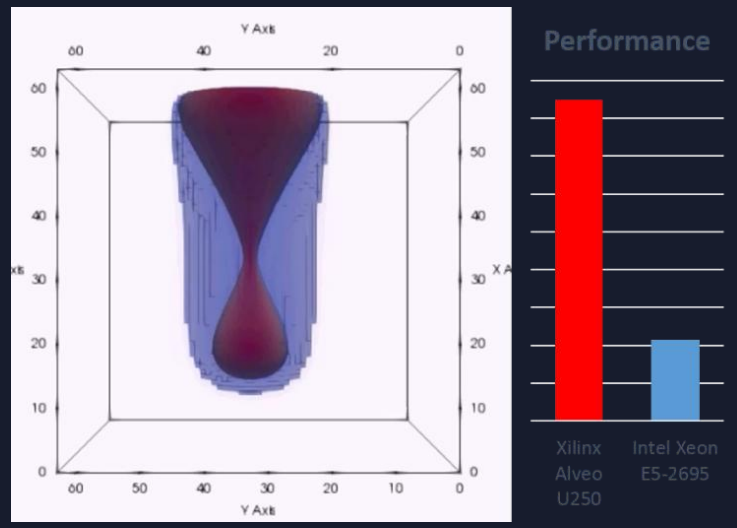
Figure 3: hls4ml tool flow

https://www.xilinx.com/content/dam/xilinx/publications/powered-by-xilinx/cern-case-study-final.pdf

XILINX

Computational Fluid Dynamics

ALVEO Accelerated CFD Kernels



byte LAKE

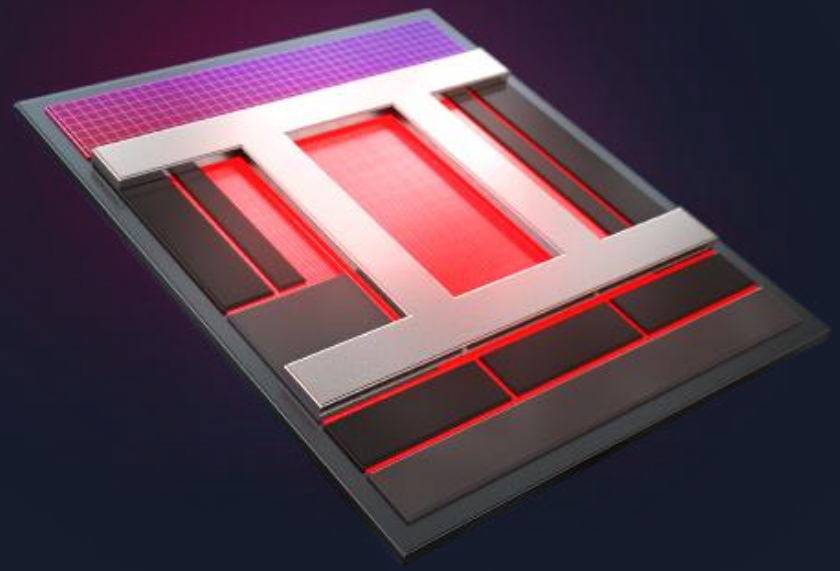
Faster Time to insight, Fewer Nodes

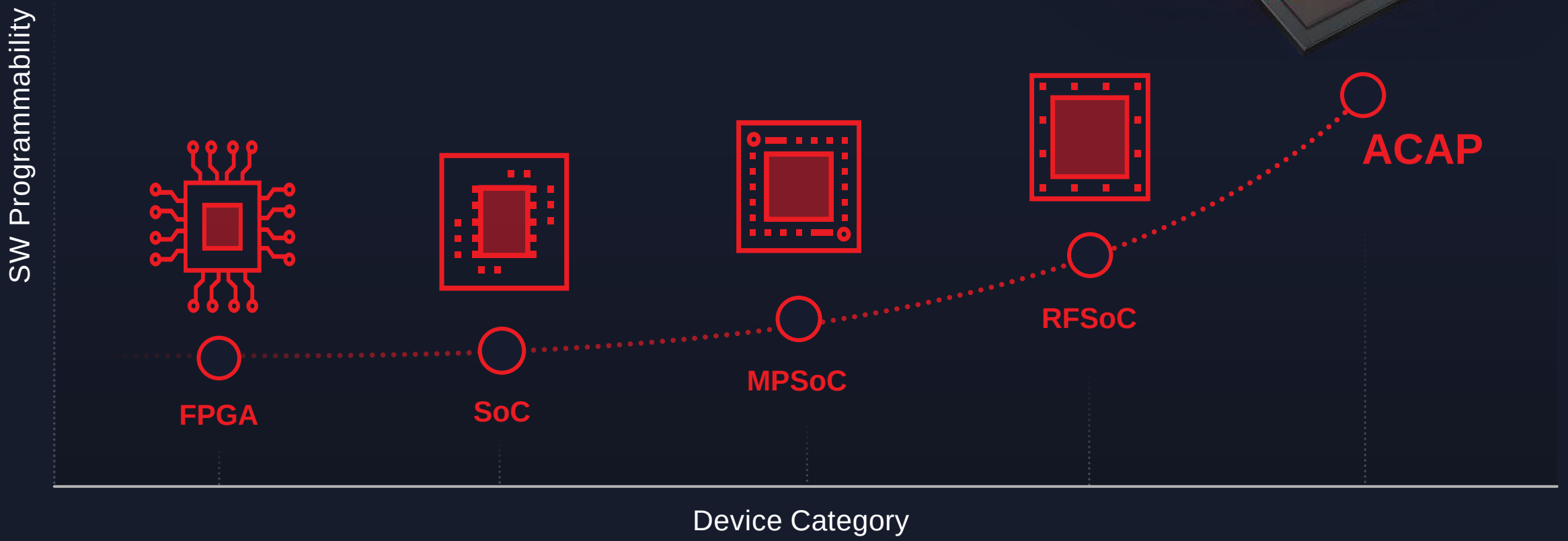
- 4x Faster simulation time
- 80% lower energy consumption
- 6x better performance per Watt

XILINX

Xilinx Transformation

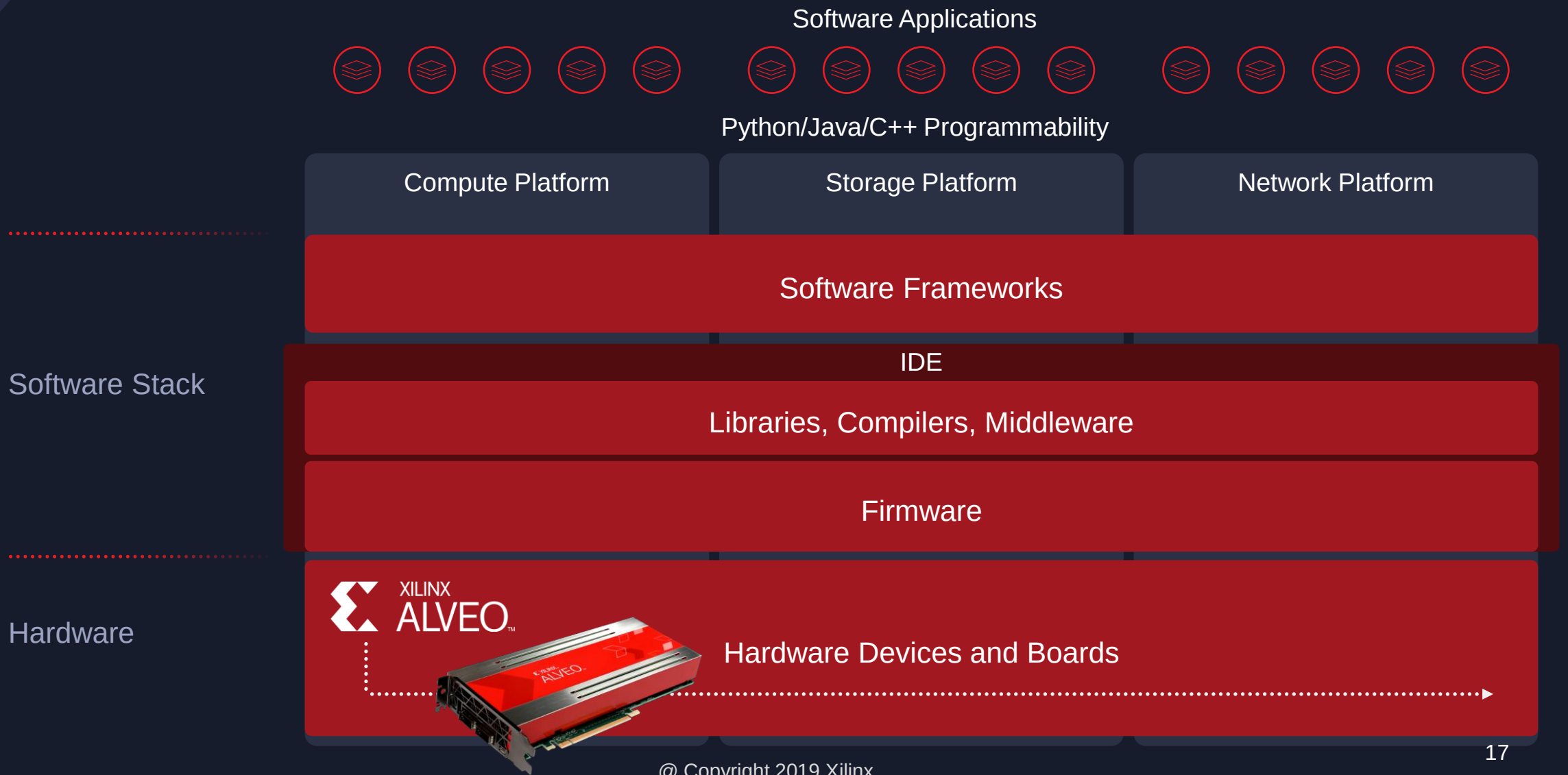
From Devices to Platforms





XILINX.

Data Center SW Stack Platforms



Conclusion

- Xilinx Alveo is a ideal and highly flexible for HPC acceleration
- Xilinx Alveo card portfolio and Eco system is growing
- Xilinx SmartNIC and Solarflare SW IP ideal for CPU off loard
- New Alveo U50 edal formfactor for HPC acceleration
- Please see us at the Avent/Silica booth now!

Xilinx Developer Forum 2019 coming soon

Check out the agenda for our Developer Forum
1. - 2. Oct. in San Jose US and
12.-13. Nov. in Den Haag Europe
– Register now and follow us #XDF2019!

- **Be the first to experience new unified software development environment**
- **Over 40 hours of hands on developer labs**
- **75+ deep-dive breakout Sessions**

Reserve your seat: https://www.xilinx.com/products/design-tools/developer-forum.html

XILINX.