Orchestrating a brighter world

NEC

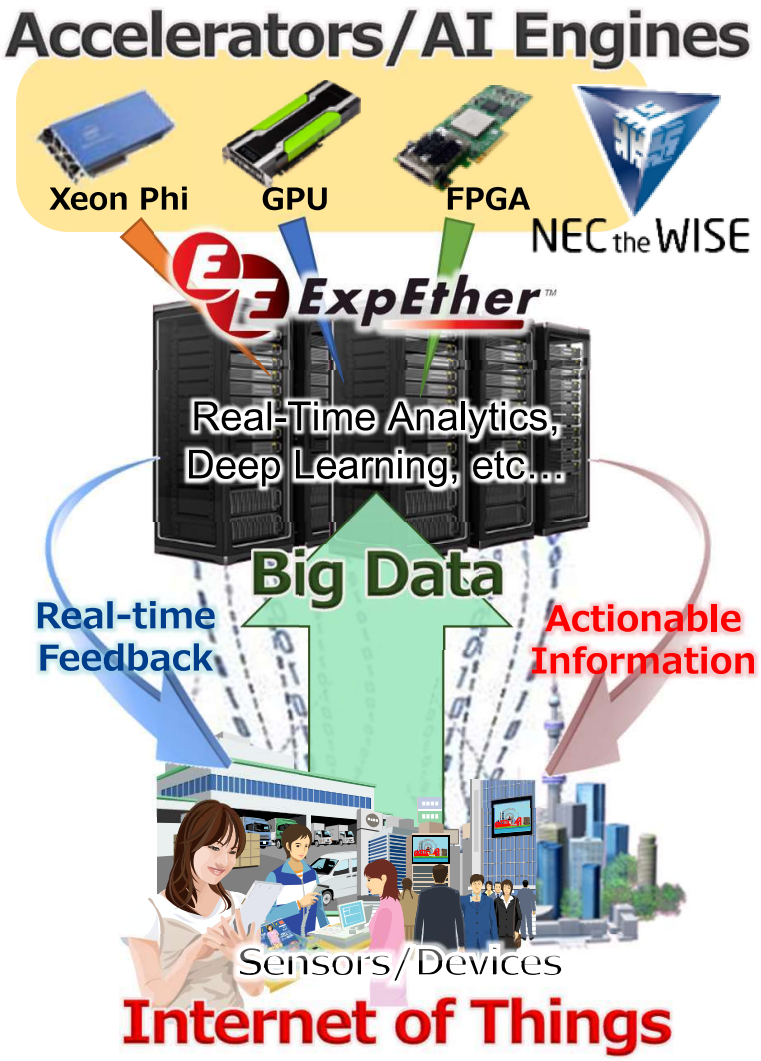
NEC Acceleration Platform

January 25th, 2017

Shinya Oda
New Platform Planning & Development Group
IoT Platform Development Division

DOC#:IoT-GE16-00112

Challenge faced in IoT Era



Situation

Big Data of varying characteristics, such as Live feeds, graphics, video, text, etc. comes into cloud computers

Demand

This data is to be processed and analyzed in real-time

Valid Solution

To accelerate such processing, a large number of accelerators such as GPUs and FPGAs, along with high speed storage are required

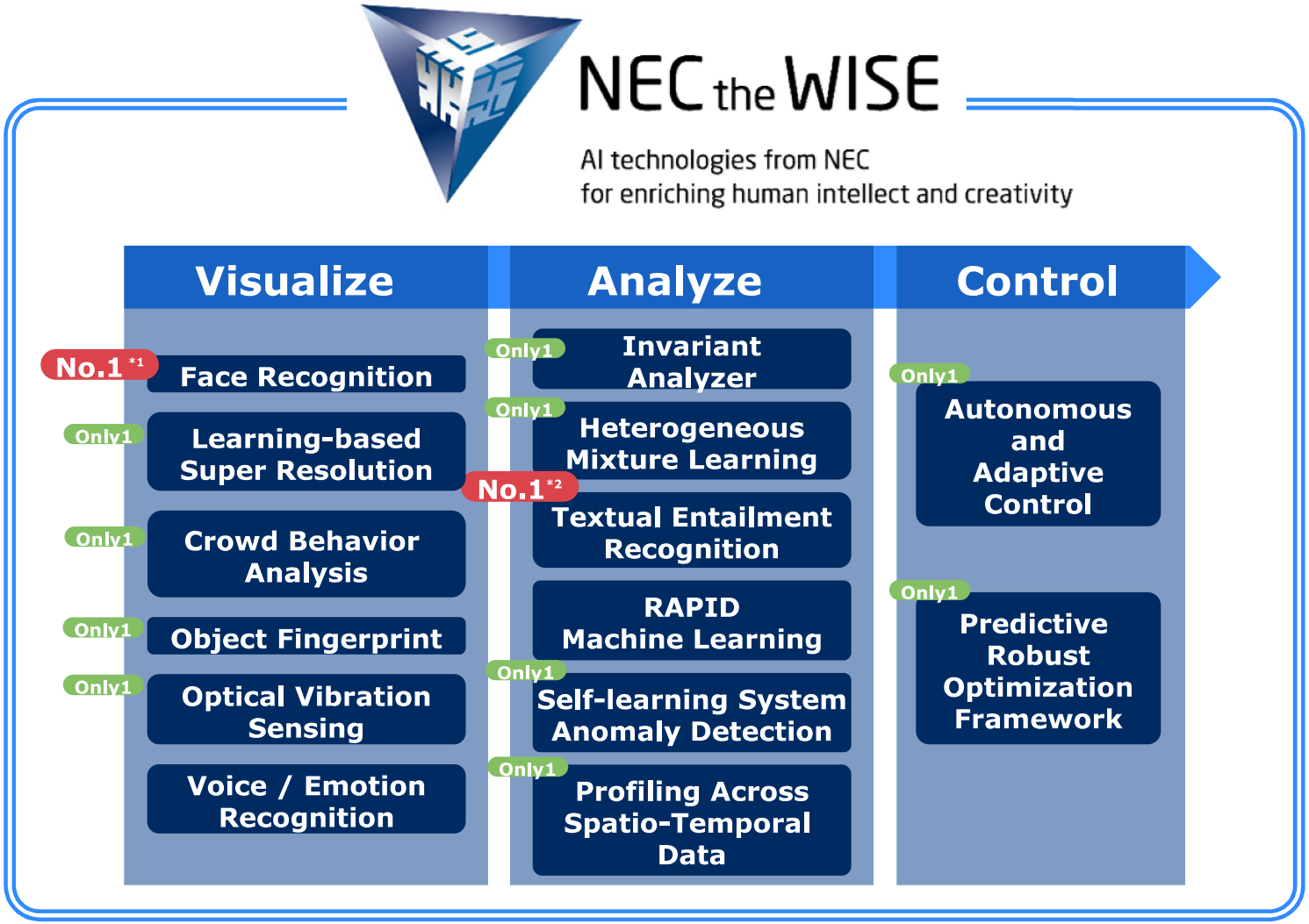
Issue

However, instead of building servers with such accelerators, Cloud vendors still prefer building homogeneous servers due to TCO and efficiency considerations

NEC's Solution

ExpEther technology that allows building dynamic accelerator deployment system.

AI technology portfolio of NEC

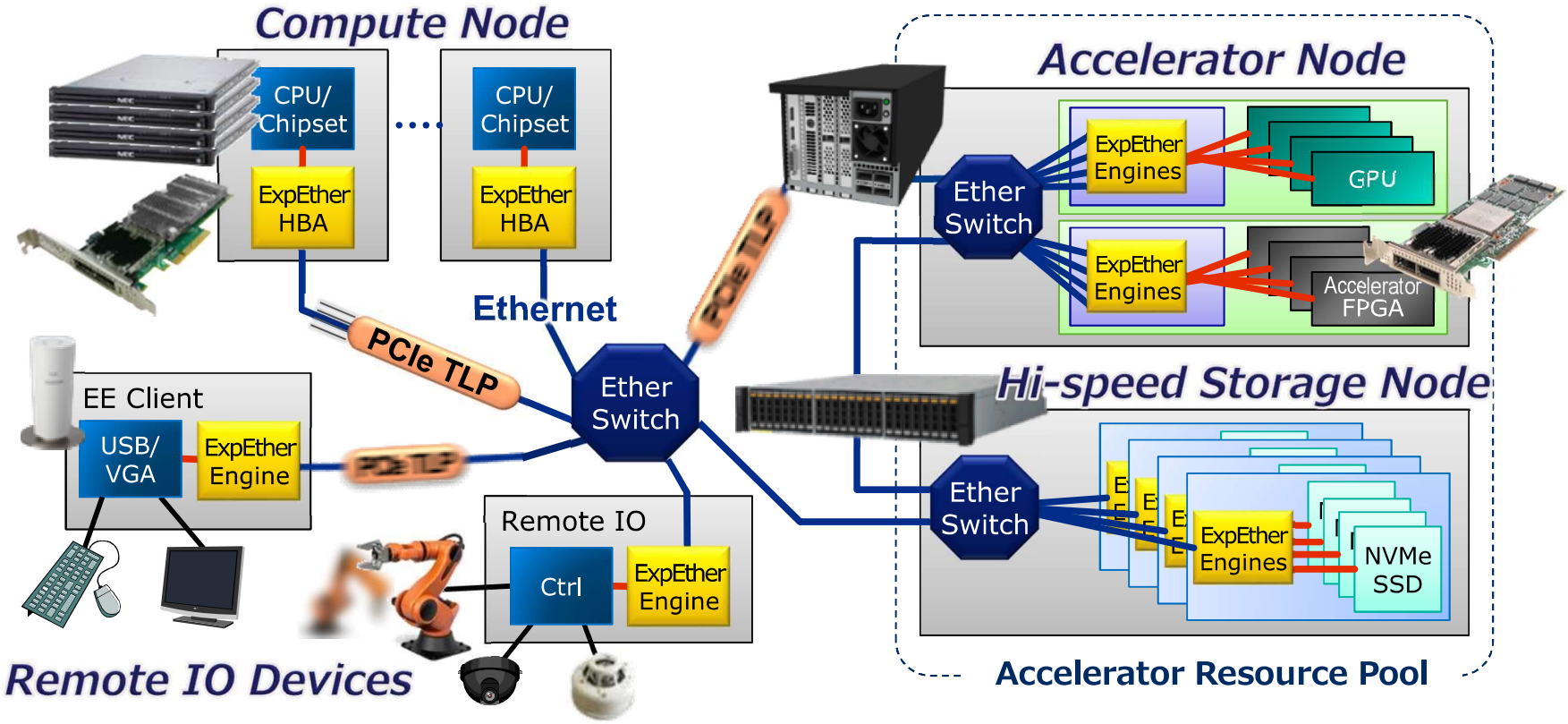


*1:Regarded as No.1 in NIST (National Institute of Standards and Technology) for 3 consecutive times
*2:No.1 in 2012 for tasks hosted by NIST

Resource Disaggregated System

Unique Selling Proposition

- IO nodes are segregated (outside) from compute node, allowing for developing disaggregated shared resource pool



- IO devices can be dynamically allocated to appropriate host according to workload
- Provides for cost optimized computing system

Acceleration Platform Products (Hardware)

80Gb ExpEther (PCI Express Switch over Ethernet)

ExpEther HBA

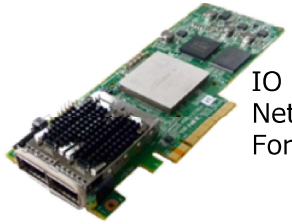
IO Interface : x8 PCI Express 3.0
Network I/F : 40G QSFP+ x 2
Form Factor : PCI Low Profile

IO Expansion Unit

IO Interface : x8 PCI Express 3.0
Slots : x16 Slot x 4
Network I/F : 40G QSFP+ x 4

NoE (NVMe over Ethernet)

NoE HBA



IO Interface : x8 PCI Express 3.0
Network I/F : 40G QSFP+ x 2
Form Factor : PCI Low Profile

NVMe SSD Storage Shelf

Slots : NVMe SSD x 24
Network I/F : 40G QSFP+ x 32

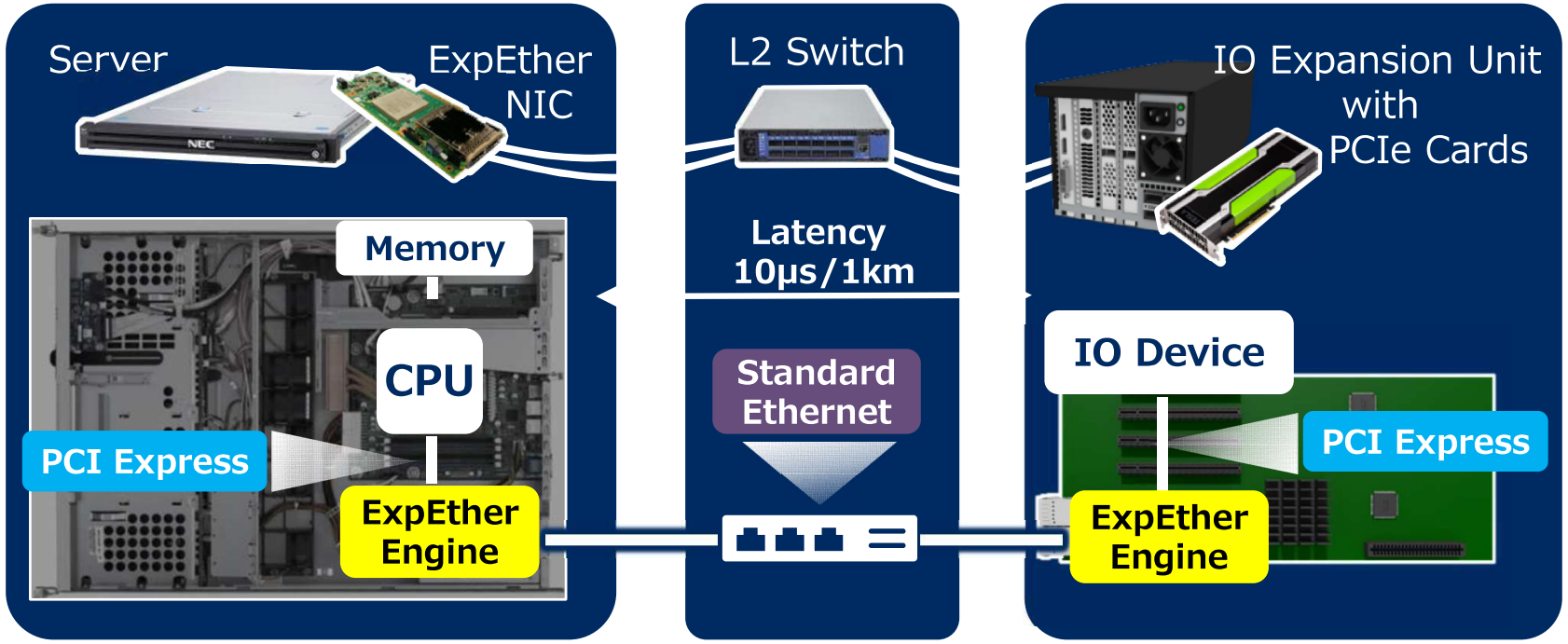
Acceleration FPGA Card

IO Interface : x8 PCI Express 3.0
Network I/F : 40G QSFP+ x 2
Form Factor : PCI Low Profile
FPGA : Altera Arria10 GX660, 1150
DRAM : DDR4 64bit+ECC x 2ch (2400MT/s, 16GB)


\Orchestrating a brighter world
NEC
ExpEther

So, what is ExpEther ?

A technology that can extend PCI Express beyond the confines of a computer chassis via Ethernet, WITHOUT any modification of existing hardware and software



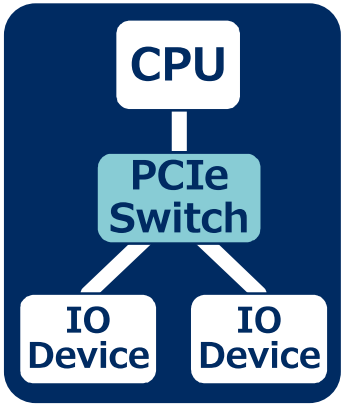
Benefit

- ✓ Extend PCIe connection over 2km *1
- ✓ Expand PCIe slot up to 128 *2
- ✓ Dynamic device allocation *3

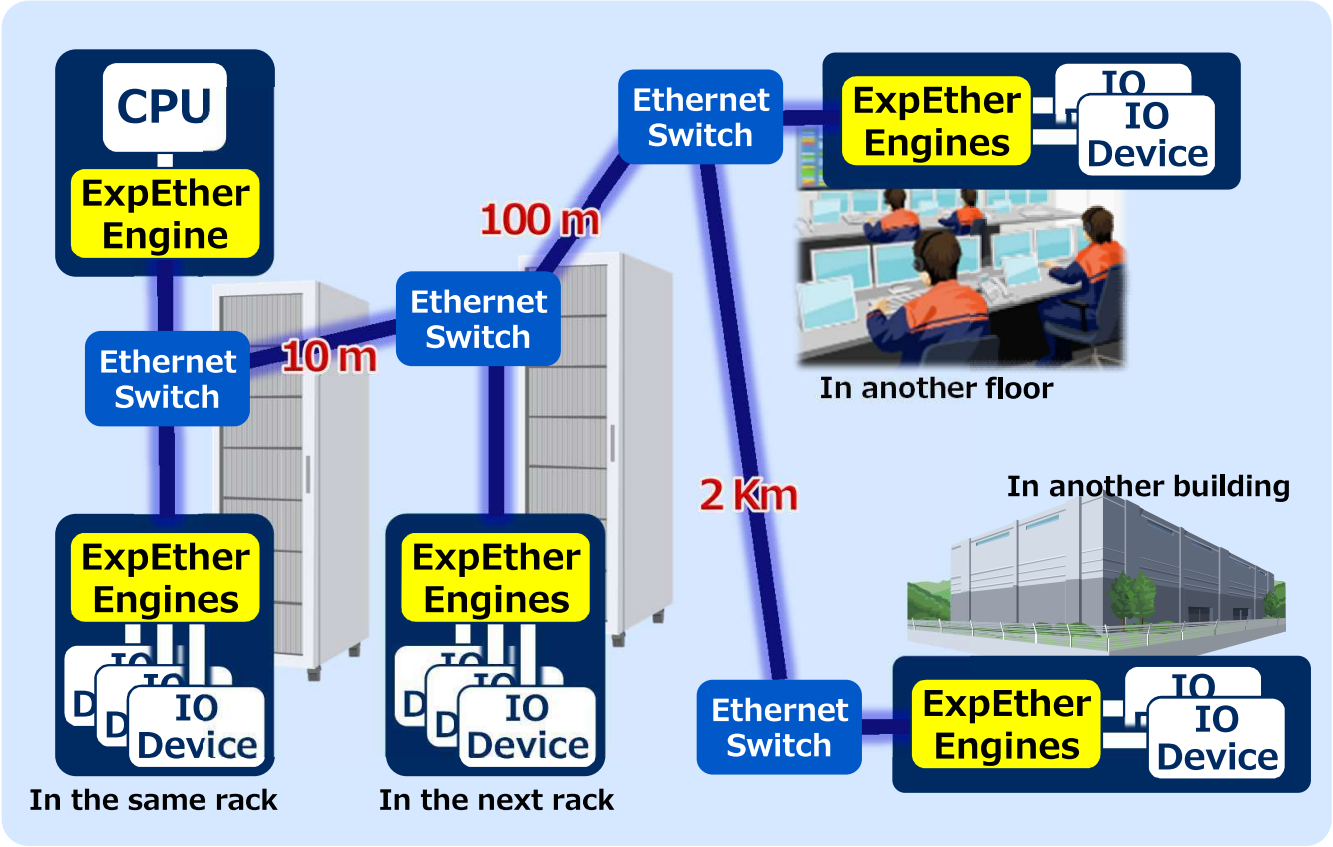
*1:Limited by latency restriction
*2:Limited by BIOS support
*3:Limited by Driver and BIOS support

Its just a ‘Broad-Scale Single Computer’ !!!

ExpEther can build new type of computing environment without physical constraints



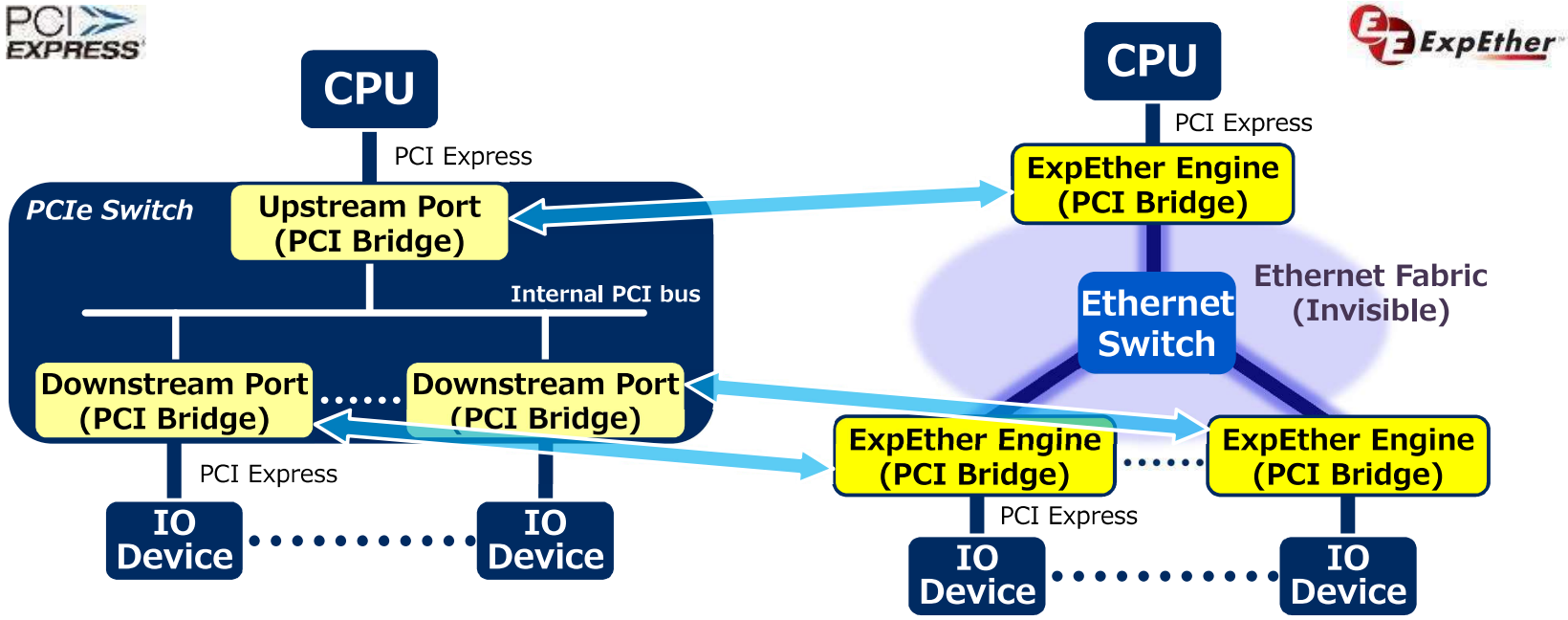
A PCI express switch is equivalent to Ethernet fabric.



Full Compatibility with PCIe Specification

ExpEther Engine is seen as PCIe Switch from CPU

- Ethernet region is invisible from the CPU

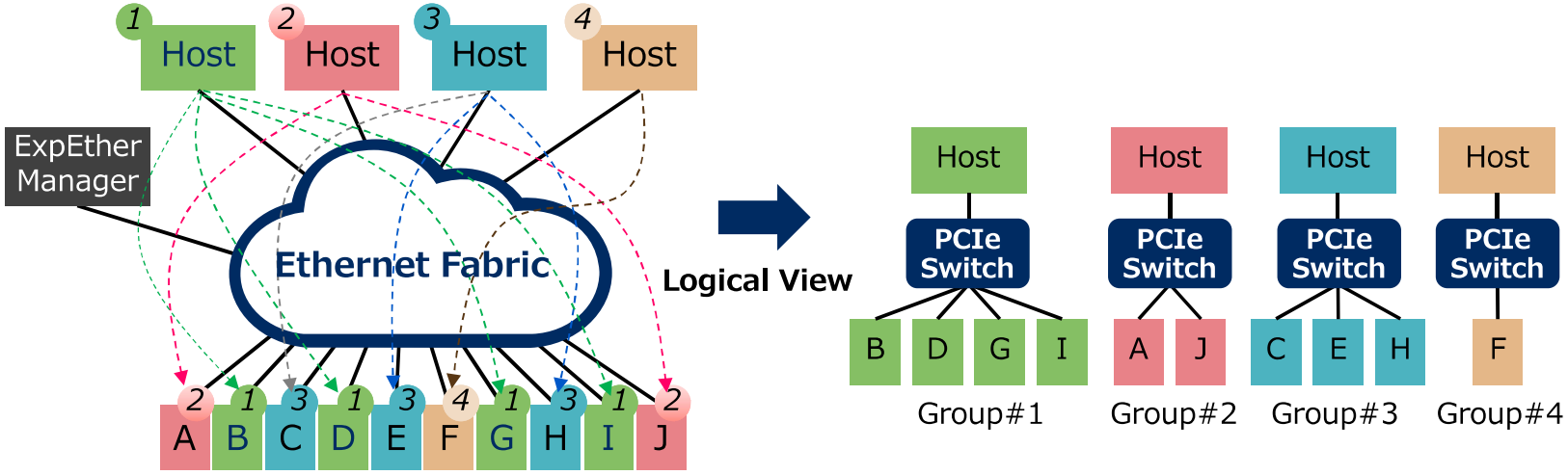


ExpEther is just another implementation of PCIe Switch

System Configuration by Grouping

Each ExpEther device has a Grouping ID to connect a Host and IO devices logically

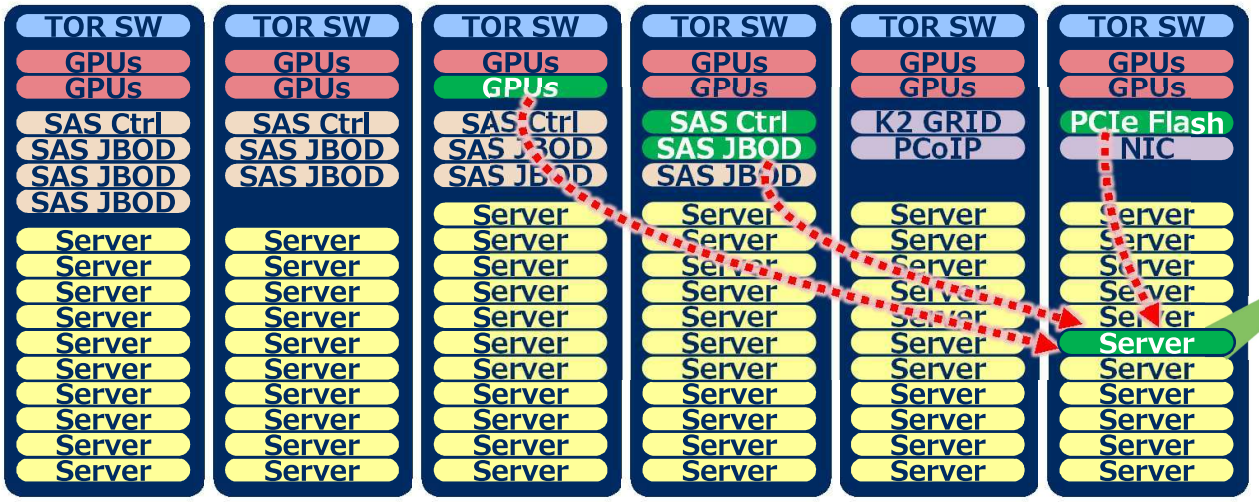
- The ID is assigned by rotary switch or Manager software
- The ID can be set from 1 to 4,000 and it is used as VLAN tag

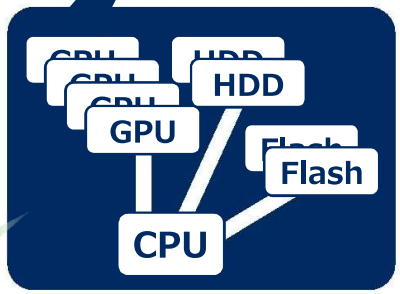


Case Study: Resource Pool System for HPC (Osaka University)

64 servers and 70 IO devices for research in Osaka University

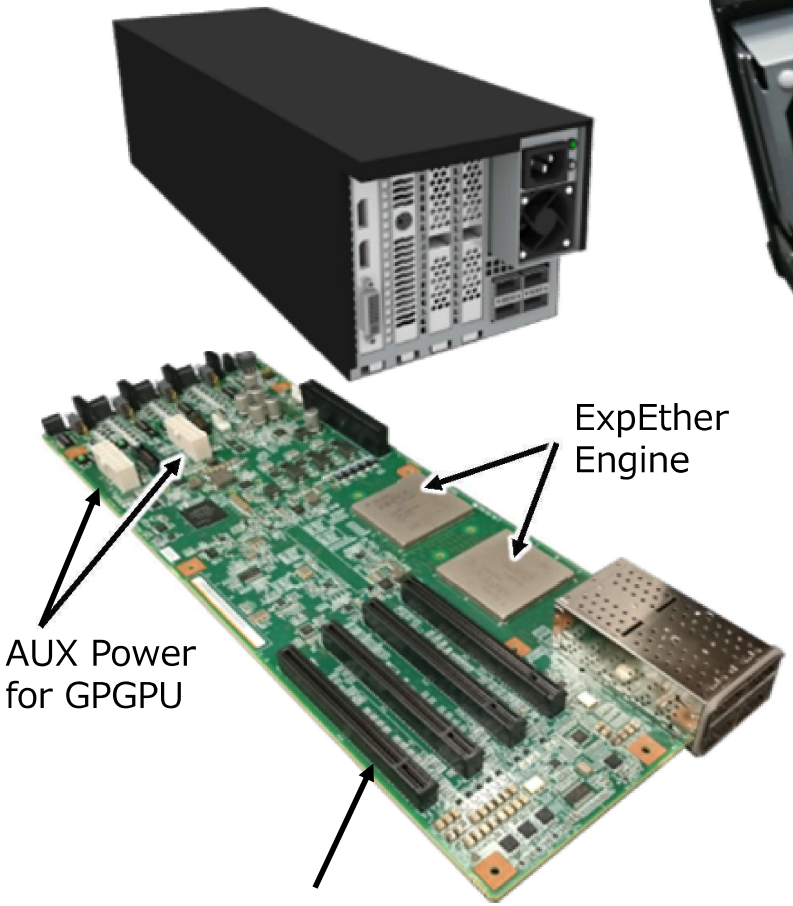
- ✓ There are GPUs, Flash storages and VDI accelerators as IO device.
- ✓ The IO devices are dynamically connected to the servers through 10G ExpEther in accordance with server's workload.

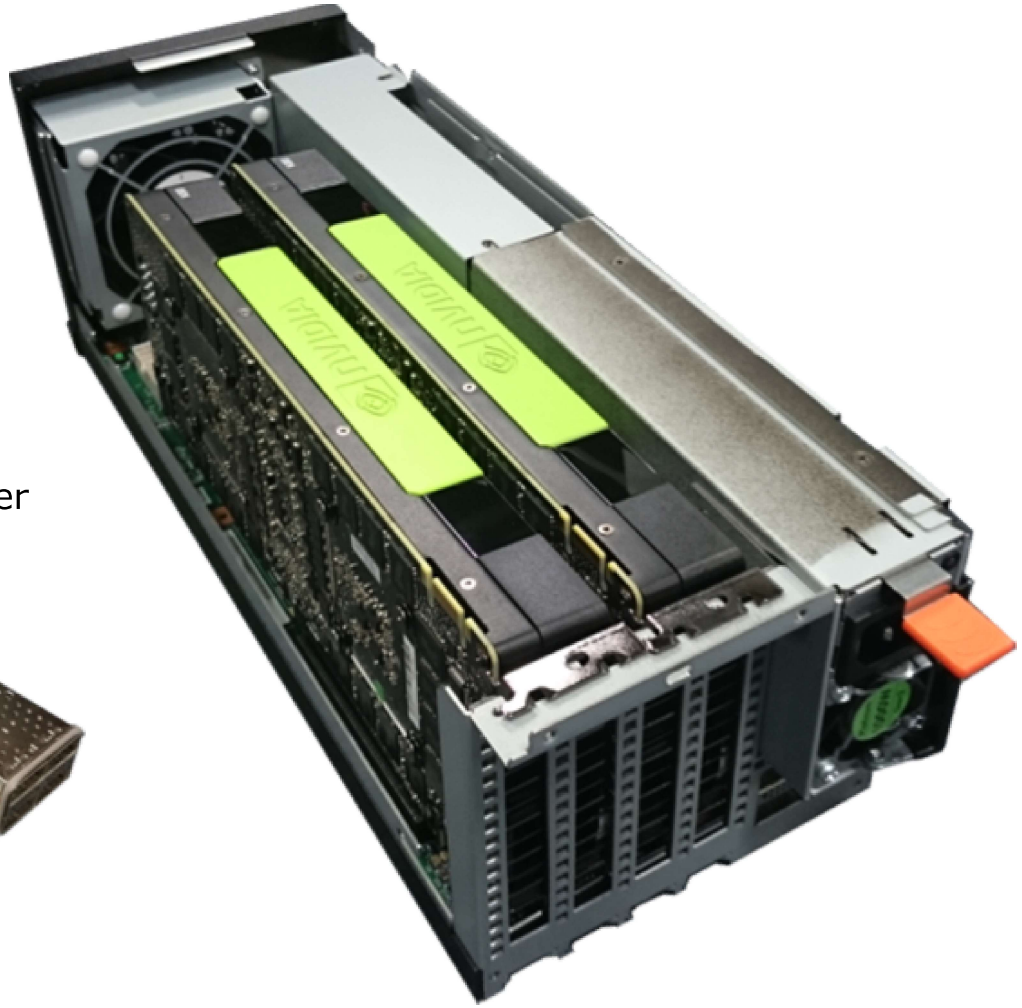




System is configured according to user requirement

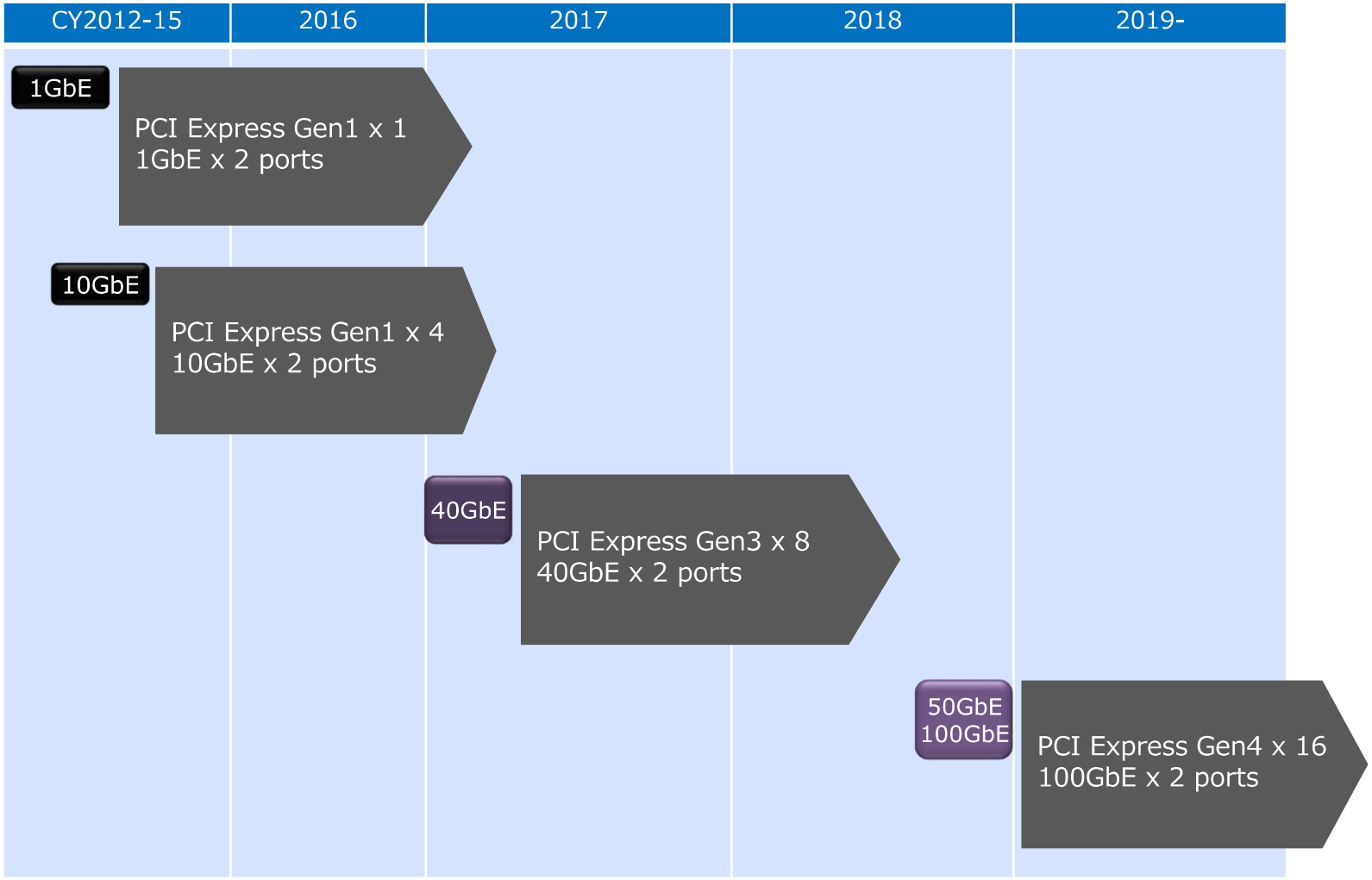
I/O Expansion Unit with NVIDIA K80





Mount two NVIDIA K80 cards

ExpEther Roadmap



\Orchestrating a brighter world NEC

NVMe over Ethernet

What is NVMe over Ethernet (NoE) ?

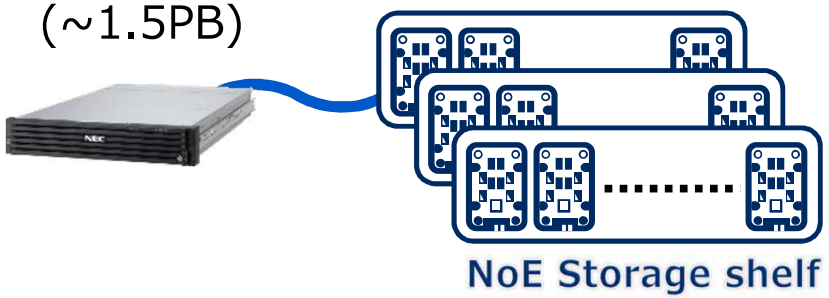
NoE extends the benefits of NVMe SSD by sharing from multiple servers through standard Ethernet Fabric.

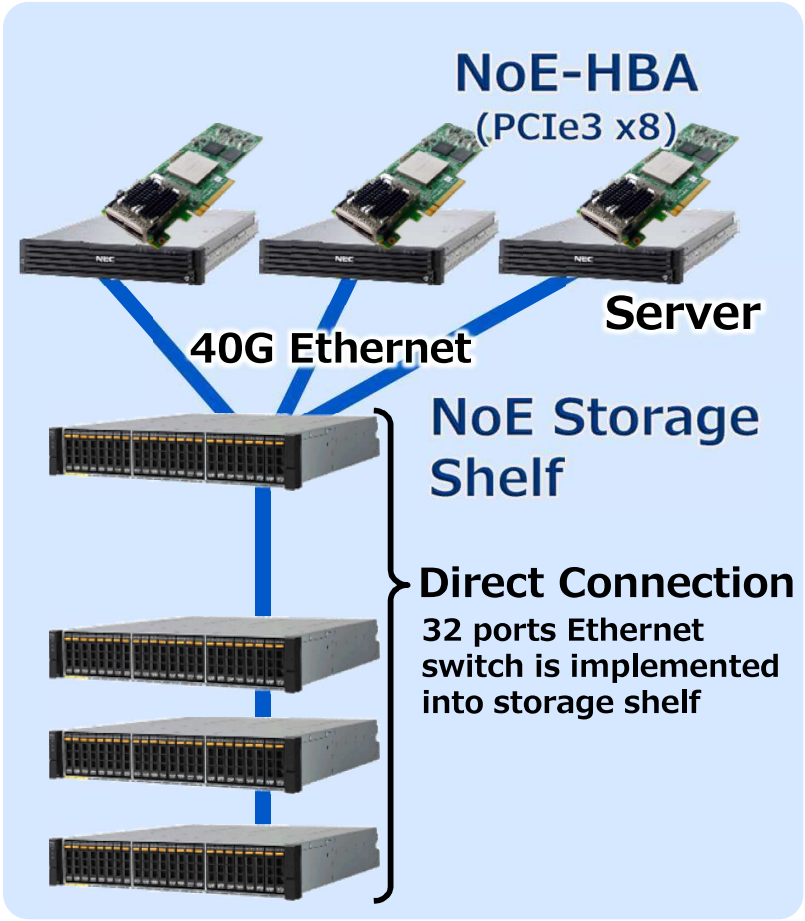
High Performance

Equivalent to internal NVMe

Easy to Scale Out

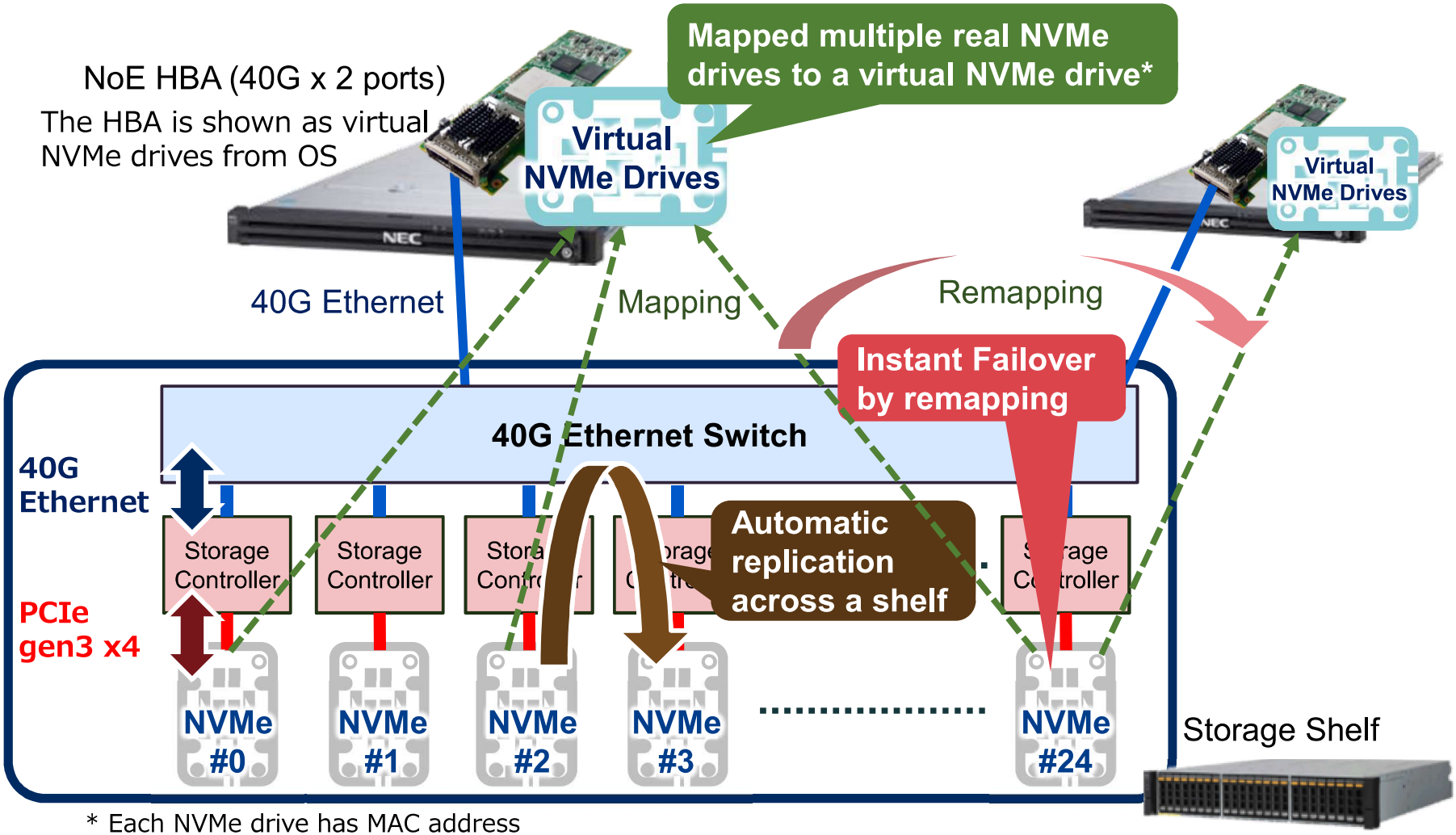
Connect over hundreds SSDs to single server beyond limitation of PCIe spec (~1.5PB)





NoE Concept

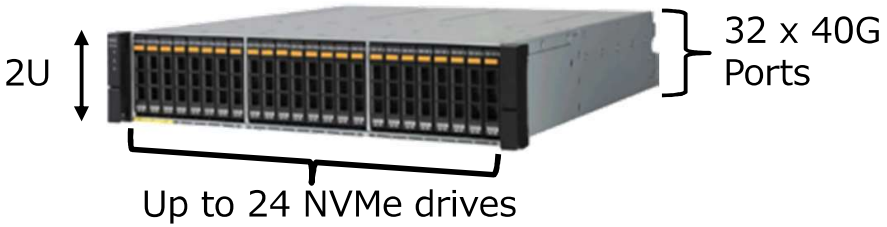
NoE provides similar usability of SAN storage to NVMe



* Each NVMe drive has MAC address

NoE Storage Shelf Product - ADS1000

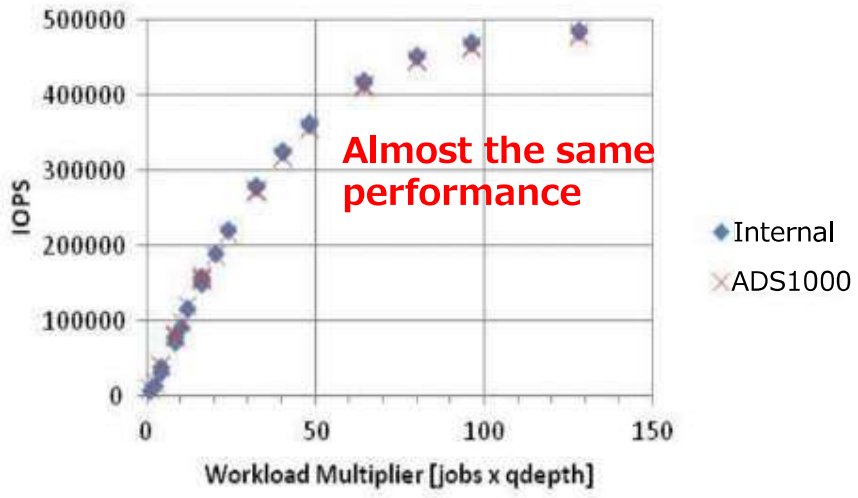
- 2U height and 19” rackmount size
- Up to 24 Standard NVMe SSDs, and Intel 3D XPoint ready
- 32 x 40G Ethernet ports are in backside



Highlighted Spec. and Performance

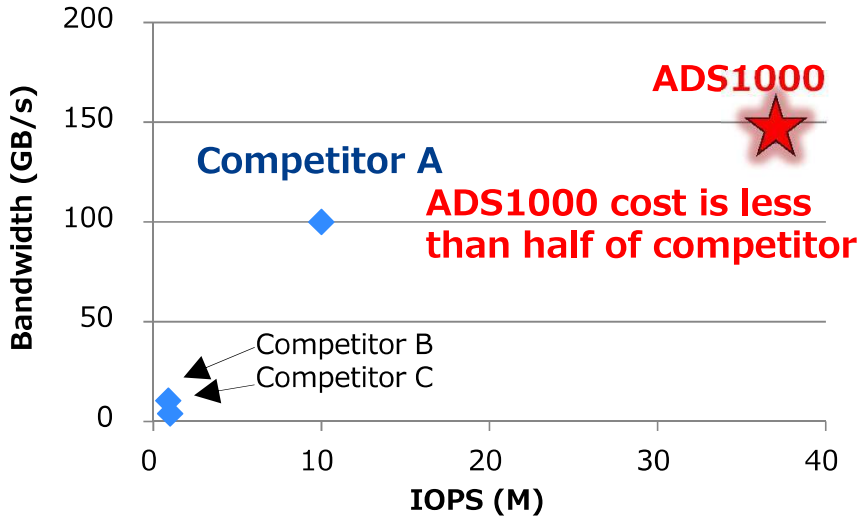
Max Capacity	192TB (8TB SSD x 24)
Latency	< 100 us (Including SSD)
Protocol Latency	< 3 us Roundtrip
Max Bandwidth	72 GB/s
IOPS	17.8M IOPS (4K Random Read)

ADS1000 drive vs Internal NVMe

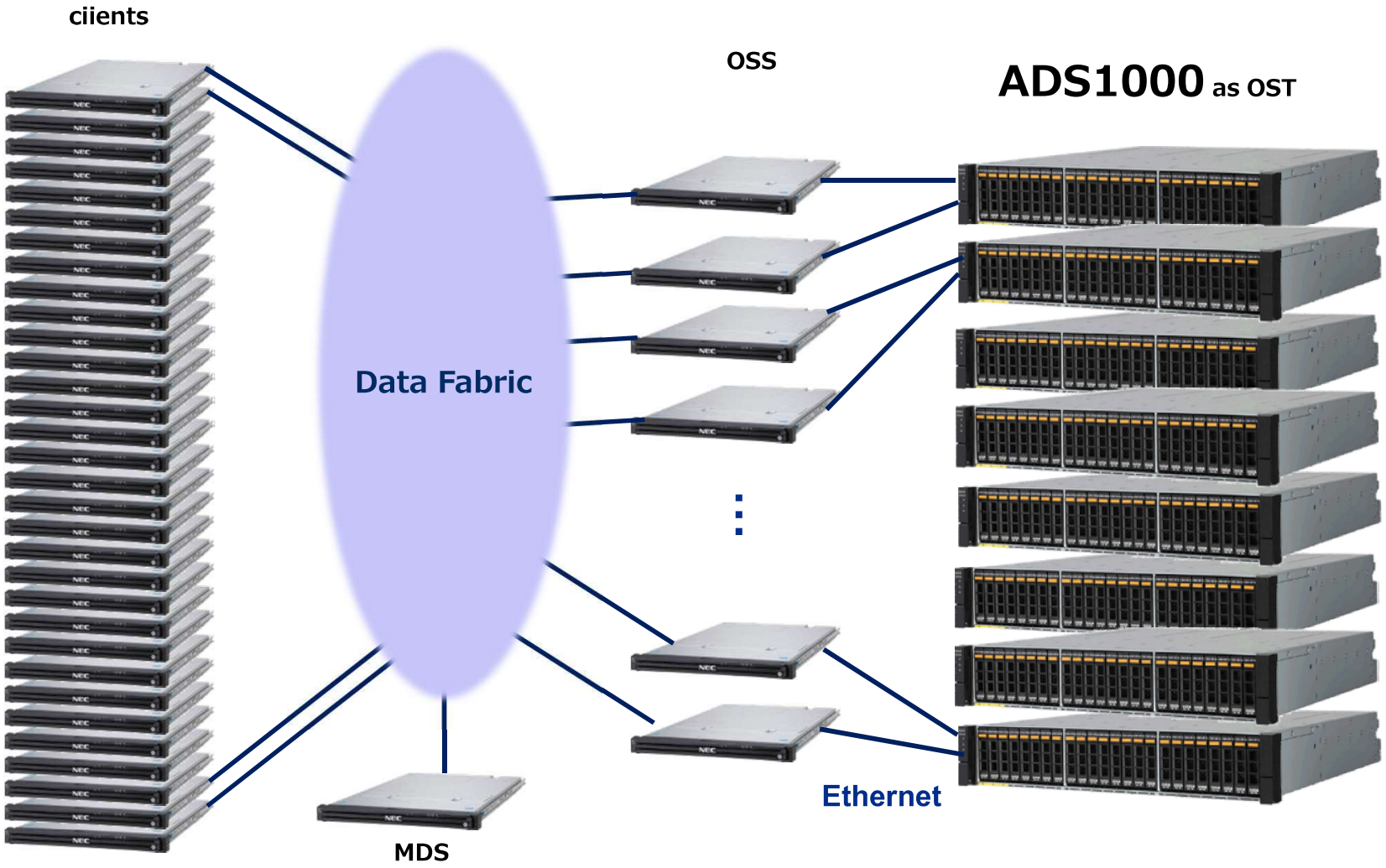


Comparison to competitors

Configuration is basically aligned to Competitor

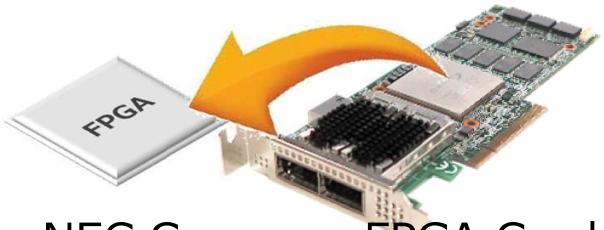


NoE Use Case (Lustre Filesystem)

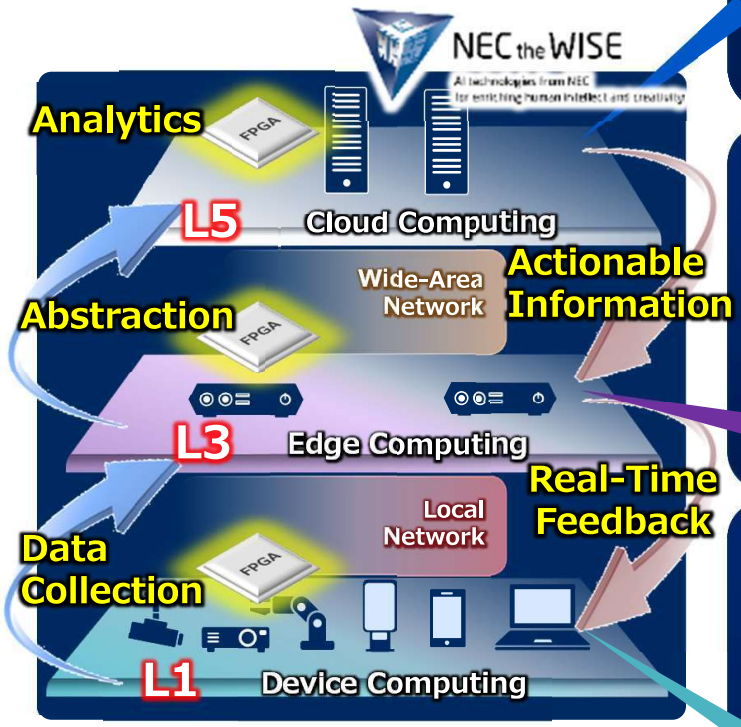



\Orchestrating a brighter world
NEC
Acceleration FPGA

Industry Trend of FPGA Usage in IoT Business



NEC Common FPGA Card



IoT Five Layers

L5 Cloud ~ Analytics/Deep Learning

GPU is being utilized for acceleration of analytics software, database by huge cloud vendors. Also they have started utilizing FPGA to reduce the power consumption. FPGA will spread to 2nd tier cloud vendors too in the near future. Intel is focusing on the deep learning solution.

L3 Edge ~ Abstraction/Real-Time Proc.

Utilization of FPGA has just started for acceleration of various processing to assist the low cost/power processor. FPGA is mainly used for data cleansing/abstraction and real-time processing. Also DNN is used as well as cloud.

L1 Device/Sensor ~ Smart Device

Many FPGAs have already been deployed in this area. As IoT, one chip solution would be developed by implementing sensor and processor into a FPGA that has intelligence.

Common FPGA Platform

Developing common FPGA platform that can also support OpenCL, Deep Learning

NEC Common Acceleration FPGA PF

HW Engines

Engine

Manager

Application (FaceDetection,…)

App's DLL

Cmn. I/F

Cmn. Driver

Download

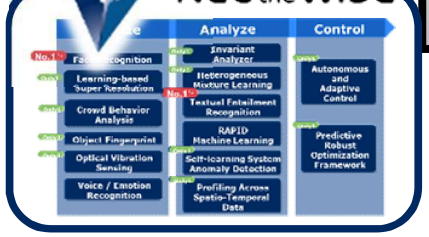
SW-to-HW Compiler (CyberWorkBench)

Cmn. Uncore

Cmn. I/F

HW Engine

NEC the WISE



OpenCL Environment

AlexNet, GoogleNet User Network

Caffe, Torch — CNN Framework

MKL-DNN — CNN Library

OpenCL

BSP

Infrastructure IP Core

CNN

NEC Common FPGA Card

Improvement of SW-to-HW Conversion Technology

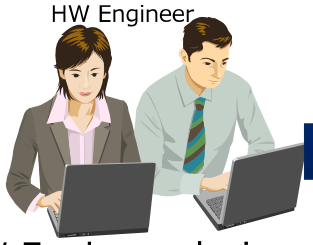
Easy to convert to hardware Engine from software

Phase-1 (Now)

Engine like facial recognition is converted by engineer

```
int main(char *str)
{
   int n, reversedInteger = 0, remainder,;
   printf("Enter an integer: ");
   scanf("%d", &n);
   originalInteger = n;
   // reversed integer is stored in variable
   while( n!=0 )
   {
      remainder = n%10;
      reversedInteger = reversedInteger*10;
      n /= 10;
   }
```

HW engineer fully understands the logic



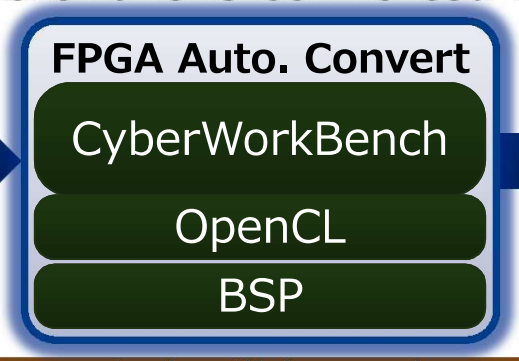
HW Engineer designs it from scratch

It takes about six months

Phase-2 (2017/4Q)

Engine base on software is converted to hardware automatically

```
int main(char *str)
{
   int n, reversedInteger = 0, remainder,;
   printf("Enter an integer: ");
   scanf("%d", &n);
   originalInteger = n;
   // reversed integer is stored in v
   while( n!=0 )
   {
      remainder = n%10;
      reversedInteger = reversedInteger*10;
      n /= 10;
   }
```



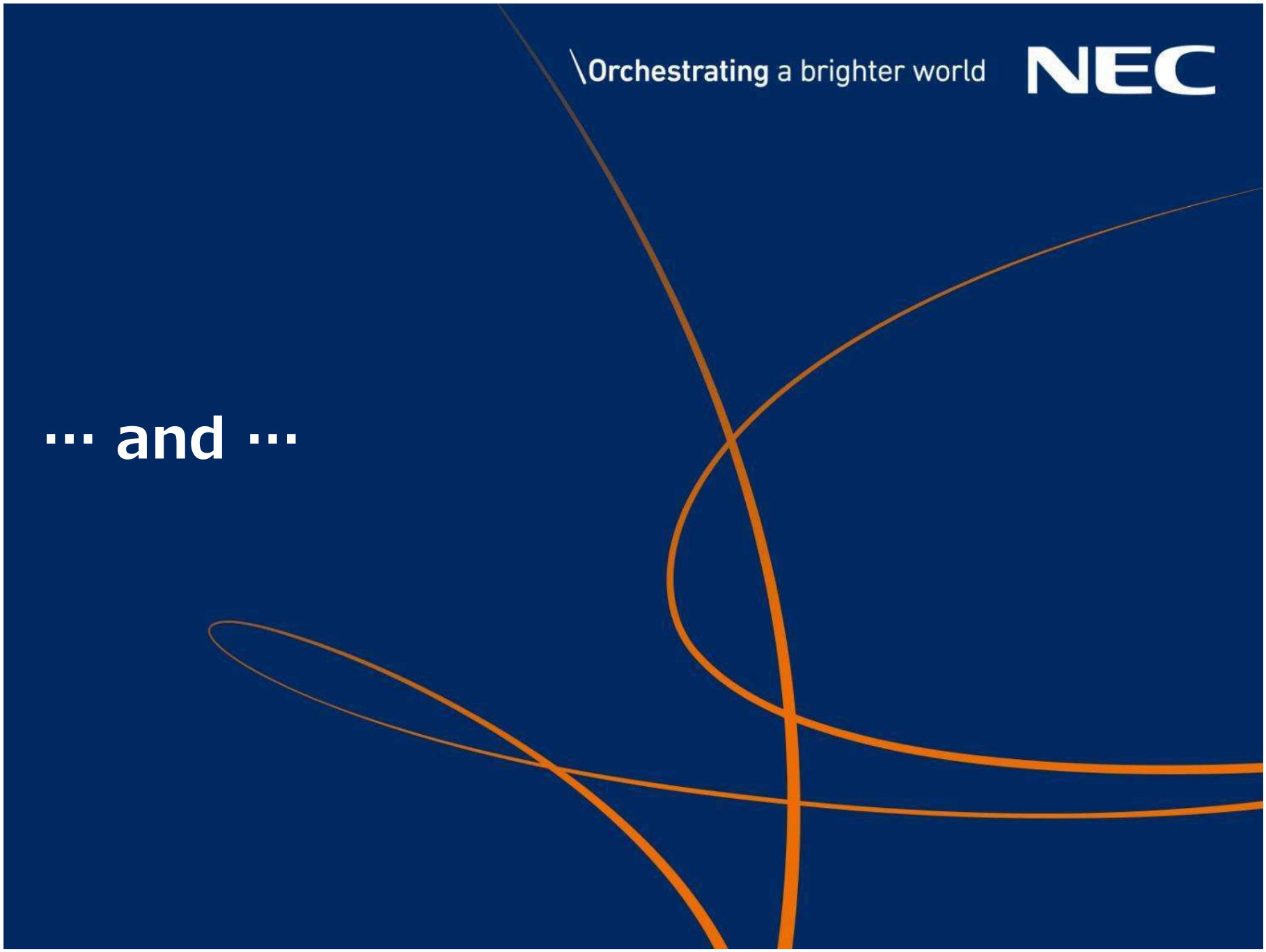
Slight Adjustment only

Development period will be reduce to about two months

NEC Common FPGA Card Overview

Acceleration card with Arria10 FPGA.
Supports 40GbE, or HDMI camera, or GPIO interface.

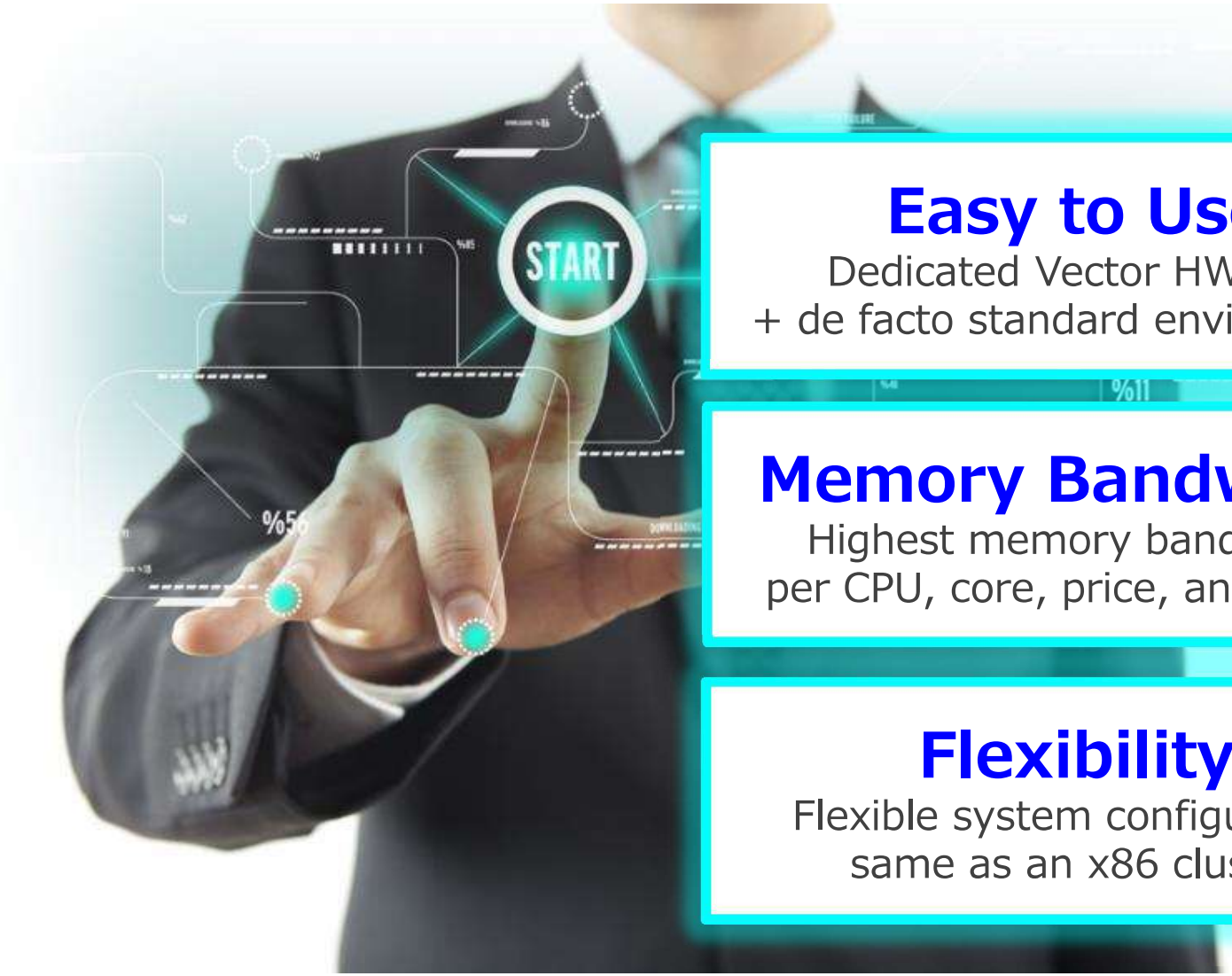
Item	Specification
Size	PCI Express Low Profile MD2 (68.9 x 167.65 mm²)
Host interface	PCIe Gen3 x 8
External interface	QSFP+ x 2, or HDMI camera x 2, or GPIO port
FPGA	ALTERA Arria10 GX 660 ~ GX 1150, ~1,150 K LE
DRAM	DDR4 x 2ch, 2400MT/s 38GB/s , 8GB~16GB
Heat sink	Active or Passive
Power	~50W (with option power cable)


\Orchestrating a brighter world
NEC
… and …


Project Aurora
Neogenesis
Vector
Supercomputer
2018

HPC Roadmap

Aurora Concept



Easy to Use

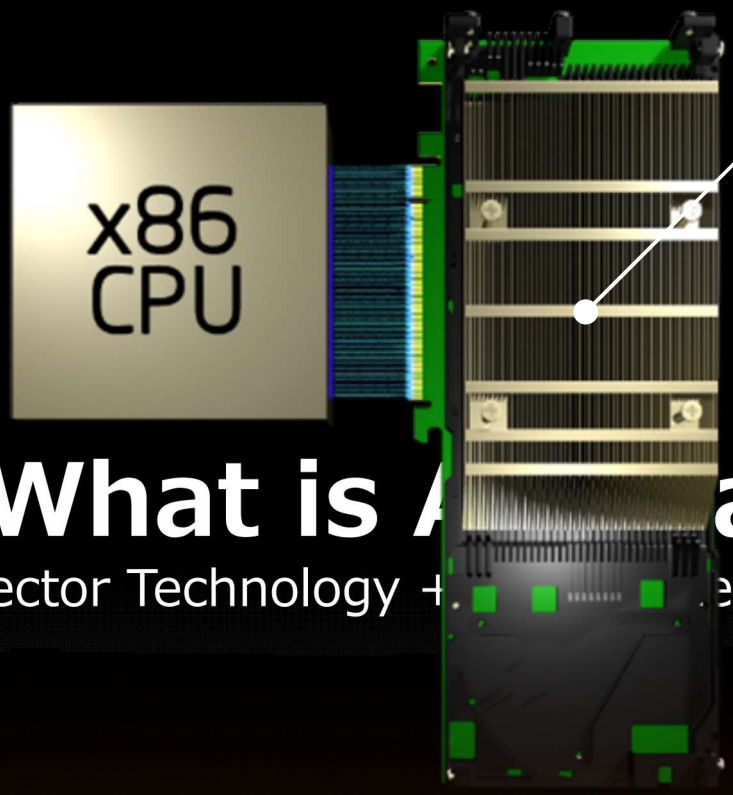
Dedicated Vector HW/SW
+ de facto standard environment

Memory Bandwidth

Highest memory bandwidth
per CPU, core, price, and power

Flexibility

Flexible system configuration
same as an x86 cluster



What is A a?

Proven Vector Technology + er Technology

Project Aurora 2018

Neogenesis Vector Supercomputer


Orchestrating a brighter world
NEC