Edge Computing - Infrastructure for Low-Latency AI/Media

ADLINK Technologies
Jeff Sharpe, Director Product Strategy & Business Development
## Standards Leadership & Technology Partners

Developing, Innovating, and Implementing for the Industry

<table>
<thead>
<tr>
<th>PICMG Standards</th>
<th>SGeT Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>PICMG Executive Member</td>
<td>SGeT Founding Member</td>
</tr>
<tr>
<td>Intel® IoT Solutions Alliance Premier Member</td>
<td>NXP Partner</td>
</tr>
<tr>
<td>Open Network Foundation Innovation Contributor Member</td>
<td>Windows Embedded Partner</td>
</tr>
<tr>
<td>Central Office Re-architected as a Datacenter Innovation Contributor Member</td>
<td>Telecom Infra Project Member</td>
</tr>
<tr>
<td>Edge Computing Consortium Member</td>
<td>VITA Standards Organization Member</td>
</tr>
<tr>
<td>OpenEdge Computing Consortium Member</td>
<td>Montavista Partner</td>
</tr>
<tr>
<td>AMD Premier Embedded Partner</td>
<td>Wind River Hardware/Software Gold Partner</td>
</tr>
<tr>
<td>ETSI MEC Committee Member</td>
<td>OpenFog Consortium Member</td>
</tr>
<tr>
<td>AXIe Consortium Strategic Member</td>
<td>PC/104 Embedded Consortium Affiliate Member</td>
</tr>
<tr>
<td>Future Airborne Capability Environment Consortium Member</td>
<td>PXISA Sponsor Member</td>
</tr>
<tr>
<td>Wireless Innovation Forum Member</td>
<td>Object Management Group Member</td>
</tr>
<tr>
<td>Future Airborne Capability Environment Consortium Member</td>
<td>Automated Imaging Association Member</td>
</tr>
<tr>
<td>Object Management Group Innovation Contributor Member</td>
<td>Association of Gaming Equipment Manufacturers Associate Member</td>
</tr>
</tbody>
</table>

**Leading EDGE COMPUTING**
Edge Compute for low latency

**Multi-Access**

- **Access CO**
  - Equipment Room
  - IP20, -20°C~55°C
  - Hardened Compact Server

- **Edge DC**
  - Telecom Room
  - NEBS Level 3
  - IP20, 0°C~30°C

- **Local DC**
  - DataCenter
  - ToR, X86
  - Low hardened needs

- **Regional DC**
  - 70-80 ms

**Customer Premise**

- <1ms

**Other**

- 2-5 ms

**Equipment Room**
- IP20, -20°C~55°C
- Hardened Compact Server
Distributed Edge – Inference / Virtualization

Open Eco-System for RAN deployments (Pole Mount / Container – based Appliances)
Edge Applications

- WebTV/Video Communication
- Flickr
- Text2Speech
- Speech2Text
- Video2Text
- Translation
- vCPE
- Dedicated Routing
- Enterprise Networks
- Social Events
- Local Aides
- Alerts
- GPS
- Cryptography
- Cloudlets
- VoIP
- Video
- Audio
- Resolution
- Media Transformation
- Private Networking
- Edge Broadcasting
- Compute
- Transcoding
- Content Scaling
- Local Connectivity
- Edge Content Delivery
- Aggregation
- Augmentation
- User Requests
- Proactive Caching
- Transparent Caching
- Content
- Augmented Reality
- Face Recognition
- Site Recognition
- Edge Based CDN
- Geo Information
- Preloading
- Distributed Database
- Proximity Caching
- Big Data
- Sensor Data
- Car2Car
- RAN Data
- Monitoring
- IoT
Edge Market Prerequisites – Zero Touch

Vision of Next Generation Networks at the Edge

Self Healing and Resiliency at the Edge
Edge Datacenters are remotely managed and unmanned. 5G provides low latency and high throughput delivery for services – some of which are expected to be up 99.9999% of the time.

Automation / AI / ML
The use of GPGPU technology to enhance the responsiveness of issues, policy and real-time analytics. Built in AI for container and vNF management is becoming more acute to Telecoms.

Cloud native approach and DevOps enablement
Service Providers & NFV vendors are driving a cloud native approach. Service providers acknowledge the evolvement of software in every part of the telecom network due to SDN and NFV allowing cloud-based automation for infrastructure deployment, operations and VNF management.

Role of Open Software and Hardware Architectures
Open Architectures for software and hardware infrastructure to develop, deploy and manage vNFs, containers and APIs for services – enablement of multiple, integrated partners. ONAP, ETSI, Akraino, ONF OCP TIP and others providing guidance and standards for delivery.

ADLINK & Partners to ensure low risk deployments with technology
Open Architectures – Leaders/Members
Network Edge Portfolio

**Common Architecture**

**OCP-CG Rack Core**
- Mix & Max Compute & Storage Sleds
- X2 Management Switches
- X2 Data Switches
- Balanced Solution:
  - 8x CPU (16 sleds) = 64 Sockets
  - 16x Storage sleds

**CSA-7600 Orthogonal System**
- P1: New develop Switch Card with 8x100G+40x10G;
- P2: 4xE5+RRC as new Service card for double performance/density
- Full width sleds – reusable assets

**CSA-74xx/72xx System**
- Switching up to 360Gbs
- Mix/Match Sleds
- NEBs Ready
- Support of GPU and HW acceleration
- Multiple power options
- Reusable assets
NVIDIA Partnership – AMEC Platform

Accelerated MEC – Development Kit for Telecom

Accelerated MEC is an enhanced version of Edge Computing using GPU architectures to support, high performance computing at the edge.

To provide gigabit broadband connectivity to residential properties as a last mile complement or replacement for fixed networks

To build countrywide mobile experiences that enable new use cases driven by augmented reality/virtual reality (AR/VR), ultra high-definition (UHD) video, artificial intelligence and so on

To deliver reliable, low latency networking to mission-critical businesses to boost efficiency and productivity

To develop industrial ecosystems around machine-to-machine (M2M) and the internet of things (IoT)

To deliver nationwide infrastructure as a service (IaaS)
Phased Approach to AI/ML

Tech Introduction

**ALPS 4800**
- 4U 19" Carrier grade
- 8 GPU's P4/P40/V100
- 2 Xeon Scalable Family
- High I/O

**ALPS 2200**
- 2U 19" Carrier grade
- 2 GPU's P4/P40/V100
- 2 Xeon Scalable Family
- Low-Med I/O

Market Feedback

**AMEC 4800**
- Packet Manager
- GPU Hypervisor
- CUDA toolkit/SDK
- NFV/SDN software

**Rugged Enclosure**
- Reuse of 2U system in Outdoor Enclosure
- Used for Carrier For small cell deployments

Build open architecture standard

NVIDIA Embedded Sled
- With and w/o NVLink

Hardware Development Kit -> Software Infrastructure -> Standard Product

- 2018
- 1H’19
- 2H’19
Phase I – Dev Kit w/ Infrastructure Software

NVIDIA DGX-1 Software Stack

DEEP LEARNING FRAMEWORKS
Caffe
Caffe2
Theano
TensorFlow
Torch
mxnet
PYTORCH

DEEP LEARNING USER SOFTWARE
NVIDIA DIGITS™

THIRD PARTY ACCELERATED SOLUTIONS

CONTAINERIZATION TOOL
NVIDIA Docker

GPU DRIVER
NVIDIA Driver

SYSTEM
Host OS
# MECS-7210 Specification

Reusable CPU asset, introduction of Xeon-D Type 7

<table>
<thead>
<tr>
<th>MECS-7210 Highlights</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Compact Dimension</strong></td>
<td><strong>Form Factor</strong></td>
</tr>
<tr>
<td>Height (2U) x Width (19”) x Depth (420mm)</td>
<td>2U 19” with 420mm Depth</td>
</tr>
<tr>
<td><strong>Expansion</strong></td>
<td><strong>Processor</strong></td>
</tr>
<tr>
<td>Two dual-slot full-size PCIeX16 3.0 slot; Support GTX-1080Ti &amp; Tesla P100/V100</td>
<td>Intel Xeon Processor Scalable Family (Purley, Cascade Lake compatible)</td>
</tr>
<tr>
<td><strong>Acceleration</strong></td>
<td><strong>CPU Socket</strong></td>
</tr>
<tr>
<td>Up to 100G QAT for symmetric encryption and authentication, asymmetric encryption, digital signatures, RSA, DH, and ECC, and lossless data compression</td>
<td>2x LGA3647</td>
</tr>
<tr>
<td><strong>Extreme Performance</strong></td>
<td><strong>Chipset</strong></td>
</tr>
<tr>
<td>Fully use total 12 memory channels in Skylake GEN</td>
<td>Intel C622~C627 (Support up to 100G QAT)</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td><strong>Memory</strong></td>
</tr>
<tr>
<td>On-board SSD</td>
<td>16xDDR4, Up to 512GB (RDIMM)</td>
</tr>
<tr>
<td>SATA bay</td>
<td>Up to 3x2.5” drive bays*, hot-swappable</td>
</tr>
<tr>
<td><strong>Expansion</strong></td>
<td><strong>Ethernet</strong></td>
</tr>
<tr>
<td>2x single-slot full-size PCIe 3.0 x16 slots with external panel Or 2x dual-slot full-size PCIe 3.0 x16 slots, internal</td>
<td>Up to 4x10G SFP+</td>
</tr>
<tr>
<td><strong>IO</strong></td>
<td><strong>Console Port</strong></td>
</tr>
<tr>
<td>Ethernet</td>
<td>1xRJ45</td>
</tr>
<tr>
<td>Console Port</td>
<td><strong>USB3.0</strong></td>
</tr>
<tr>
<td>USB3.0</td>
<td>2x USB on Front plane, 2x Internal USB dock on IO board</td>
</tr>
<tr>
<td>Power/Reset</td>
<td>1x Power button, 1x Reset button</td>
</tr>
</tbody>
</table>
MECS-7210 Highlights

- Configuration for expansion platform
  ✓ (Option #1) 2x single-slot full-size PCIe 3.0 x16 slots
  ✓ (Option #2) 2x dual-slot full-size PCIe 3.0 x16 slots

- 3x 2.5” SATA bay
- RJ-45 console port
- 2x RJ-45 10/100/1000BASE-T Ethernet ports
- 2x 10G SFP+ Ethernet ports (Optional 4x10G SFP+)
- 2x USB 3.0
2U MEC Edge Server appliance
## Supported Dual-slot GPU Card in MECS-7210

<table>
<thead>
<tr>
<th>Supported Dual-slot GPU card</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>LeadTek GeForce Series with cooling FAN.</td>
<td>LeadTek WinFast_GTX_1080_Ti(10780)</td>
</tr>
<tr>
<td>Requirement: Thickness &lt;=35mm</td>
<td></td>
</tr>
<tr>
<td>NVIDIA Tesla Series with passive cooling</td>
<td>Tesla V100, T4 or P40</td>
</tr>
<tr>
<td>Requirement: Thickness &lt;=40mm</td>
<td>280<em>125</em>40MM</td>
</tr>
</tbody>
</table>
Pole Mountable for 5G deployments

- Low Profile Pole Mount Enclosure
  - 4 RU: 24/48” x 11” x 19”
  - 7 RU: 24/48” x 15” x 19”
- Optional Integrated Power
  - Power and battery or power only
- Optional Battery Compartment
  - (1) String 80AH NiCD batteries
- Thermal Options:
  - 580/750W HX
  - 2K BTU HVAC (7U only)
- GR-487 & UL60950 Certified
- Verizon Approved
Call for discussion and combined input

Define OCP Edge Server use best-in-class, modular components