Next-Generation Cloud Infrastructure in the Age of AI

Wiwynn OCP, Project Olympus, and 19” Solutions

Ethan Yang, Deputy Manager
Wiwynn Corporation
Mainly owned by Wistron Corp.
Established on April 2, 2012

A Value Partner for Customers in Cloud Service Business

Cloud Service
Mission

Customer Value = \frac{Workload Performance}{Total Cost of Ownership}
Actively Proposing Workload Optimum Platform

Computing Intensive

Power Saving

Minimum CAPEX

Minimum Power Consumption

Storage Intensive

Storage Capacity (HDDs)

I/O Intensive

IO Performance (IOPS)

I/O Intensive

Storage Intensive

Computing Intensive
Cloud Infrastructure Grows into an Age of AI

DEMANDS
Fundamental Demands For Various Applications

APPLICATIONS
Various Evolving Applications

INFRASTRUCTURES
Building blocks and IT Gears on Advanced Technologies
3 Next Generation Cloud Infrastructures

INFRASSTRUCTURES
Buildin blocks and IT Gears on Advanced Technologies

OCP Infrastructure
Project Olympus Infrastructure
19" Infrastructure
Open Compute Project Infrastructure

**21” Open Rack**  
(Open Compute Project)

- **OCP ACCEPTED**
- **PDU**
- **AC-DC 5.5% Loss**
- **PSU**
- **BBU**
- **Server**
- **2% Loss**

**19” EIA Rack**  
(Traditional Infrastructure)

- **UPS**
- **ATS/PDU**
- **Hot-swap 3% Loss**
- **Server PS**
- **AC-DC:AC 6~12% Loss**
- **AC-DC 10% Loss**

**7.5% Total Power Loss**

**22~27% Total Power Loss**

**Wiwynn OCP Solution Reduces Power Loss by 20%**
Project Olympus Infrastructure

**NEXT-GEN CLOUD HARDWARE**
Open sourced cutting-edge Hyperscale cloud hardware developed at Microsoft

**OPEN SOURCE DEVELOPMENT MODEL**
New collaboration model with OCP community – co-develop open hardware at cloud speed

**INDUSTRY ECOSYSTEM**
Bootstrap a vibrant ecosystem in OCP for the next generation of datacenter hardware
Project Olympus Product Family

SV5100G3
- DDR4 memory
- 24 DIMM slots
- GbE Redfish Management
- 3 PCI-E x16 FHHL slots

ST5100G3
- Hot-Plug NVMe storage
- Flexible design for M.2 and Ruler NVMe SSD
- Front I/O
Best HW Platform for Compute Accelerator

4U16x Compute Accelerator
Connecting 1 to 4 Hosts

16 Intel® FPGA Cards

(Wiwynn® XC200)

Optimized Rack with FPGAs for Inference

128 FPGAs*
36 GPGPUs*

- 100% Rack Utilization
- Modularized Design
- Disaggregated Solution
- Easy Scale-out Accelerator

*Based on 9kw Power Consumption
Software-Defined Data Center

BIG DATA ANALYTICS

Scale Up Dynamically

CLOUD SERVICE

Optimize for Workload

Disaggregated and Composable Rack Solution

Physical Infrastructure

Logical Infrastructure

AI

Logical Infrastructure

Computing Power

I/O Throughput

Storage Capacity

Logical Infrastructure

Computing Power

I/O Throughput

Storage Capacity

Logical Infrastructure

Computing Power

I/O Throughput

Storage Capacity

Logical Infrastructure

Computing Power

I/O Throughput

Storage Capacity
Cloud Infrastructure Grows into an Age of AI

**Applications**
Various Evolving Applications

**Demands**
Fundamental Demands For Various Applications

**Infrastructures**
Building blocks and IT Gears on Advanced Technologies