

UUO3'P|JO/V\I9!'/V\/V\/V\

**uounios Guiujeai dgp ui Apeaj iw**

**OOhWONV** 的

**GRAND-C422-20D**

> Demand for Al computing is booming

**GRAND Al training server system**

The application of Al computing is absolutely not enough through the CPU computing. With the decentralized architecture, the huge data is calculated to obtain the computing result. Therefore, we have developed a water-cooled chassis system with high expansion capability by adding multiple GPUs, FPGA or VPU acceleration cards for Al deep learning and inference.

The GRAND-C422-20D is an Al training system which has maximum expansion ability to add in Al computing accelerator cards for Al model training or inference.

mum

曰+日+H+S+E

Hyper converged infrastructure (HCI) is scale-out software-defined infrastructure that converges core data services on flash-accelerated, industry-standard servers, delivering flexible and powerful building blocks under unified management.

Efficient, agile, flexible, and integrated, these systems allow for easy scale-out storage, cost-savings, and simplicity to manage your systems. To find out if hyperconverged is the best solution for your Data Center, consider the following.

**»**

**»**

**»**

**»**

**»**

**»**

**Intel® Xeon® W family processor supported**

**6 x PCIe Slot, up to 4 dual width GPU cards**

**Water cooling system on CPU**

**Support two U.2 SSD**

**Support one M.2 SSD M-key slot ( NVMe PCIe 3.0 x4)**

**Support 10GbE network**

**Virtual
Compute**

**Virtual
Memory**

**Virtual
Storage**

**Virtual
Network**

**Virtual
Management**

**IPMI remote management**

***In one easy to manage appliance***

***[www.ieiworld](http://www.ieiworld). com***



**Edge computingserverintraining**

**Edge to cloud communication**

**The advantages of edge computing:**

**Edge computing**

**GRAND-C422-20D**

**Before: Server Computing**

III

Ethernet Switch

**What is deep learning**

**Currently: Edge Computing**

**i**

I

Inference

**ARTIFICIAL INTELLIGENCE**

**MACHINE LEARNING**

**DEEP LEARNING**

**Learning from existing data**

**Forward**

Reduce data center loading, transmit less data, reduce network traffic bottlenecks. Real-time applications, the data is analyzed locally, no need long distant data center.

Lower costs, no need to implement sever grade machine to achieve non-complex applications.

Power:

Efficient, optimize power for edge workloads

Security:

High, raw data no need to be updated to server

Latency:

Low, handle the inference engine locally.

中II

Platfonn^y.. •宀;

The GRAND-C422-20D is perfectly suited for edge computing server in data training or inference. With edge computing, you can pre-process data generated within your organization or across your devices on-premise, to filter out irrelevant information and only keep valuable insights, and then further utilize them by sending or uploading to cloud platforms. You can save a great deal of cloud platform and bandwidth fees as your data to be analyzed is filtered and only relevant data will be further dealt with.

Security: Low Latency: high Power: hungry

**Deep learning break through**

**Pub he C oud**

**Pub** ic **C oud**

|  |  |
| --- | --- |
| «l: u J |  |
| 1 fim年卩同1" |  |
| \* 丄i丄,厶 |  |

**Training**

**Training Dataset**

**Trained Model**

**Inference**

**Predict new input data**

uman

**New Data**

*4*

**Forward**

With the introduction of deep learning, the most important issue is the "data blowout", and the fields of voice, image have the breakthrough due to the data-supported algorithms and computing power growed up rapidly. Artificial intelligence has ushered the new opportunities for developing new appliance. This time, it will lead the whole society to change in deeper level, this is the future we can see. The outbreak of artificial intelligence began from the Internet which brought many demands, including search, social, shopping, and so on. These demands are increasing, and it makes huge amount of data go online.

• Training

In deep learning, each level learns to transform its input data into a slightly more abstract and composite representation. In an image recognition application, the raw input may be a matrix of pixels; the first representational layer may abstract the pixels and encode edges; the second layer may compose and encode arrangements of edges; the third layer may encode a nose and eyes; and the fourth layer may recognize that the image contains a face. Importantly, a deep learning process can learn which features to optimally place in which level on its own.

In the field of Artificial Intelligence, inference engine is a component of the system that applies logical rules to the knowledge base to deduce new information. The first inference engines were components of expert systems. The typical expert system consisted of a knowledge base and an inference engine. The knowledge base stored facts about the world. The inference engine applies logical rules to the knowledge base and deduced new knowledge. This process would iterate as each new fact in the knowledge base could trigger additional rules in the inference engine.

**1000**

**800**

**600**

**400**

**200**

Data Blowout

***40% CAGR***

***2015-2020***

**171**

Computing Breakthrough

Algorithm Breakthrough

***2006 ♦ 2017***

Neural networks evolutions

CPU: dual core

multi core with accelerator card

The algorithm breakthrough make Al technology be more mature and practical

Accelerator card makes the neural netowrks training speed increased by 255 times

**AMD** 口

**2015 2016 2017 2018 2019 2020**

The data volume in global data center will increase by 40% annually in next few years

Predict new input data

**DOG**

**CAT**

**Backward**

**IDOGJ**

**Training Dataset**

**New Data**

\_ ・ ・ % Learning from existing data

Training

Trained Model

fa

Forward

Forward

***[www.ieiworld](http://www.ieiworld). com***



**> Al Trainin g System**

The Al training system GRAND-C442 is dedicated for these tasks because it offers a wide range of slots for storage expansion, acceleration cards and video capture, Thunderbolt™ or PoE add-on cards for unlimited data ac-quisition possibilities. In order to develop a useful training model, existing and widely used deep learning training frameworks such as Caffe, Tensor-Flow or cc c

Supported Software

•資***(J*** PyTorch 曲**Xnet**

TensorFlow neon

13 Keras 堂 Caffe2 Caffe

Deep Learning Models

**Image**

classification
AlexNet,VGG16,
GoogLeNet, ResNet,
MobileNet, etc.

**Object Detection**

SSD, Yolo vl/v2/v3,
R-FCN, RCNN,
Faster RCNN, etc.

**I Face Recognition**

MTCNN, Deep Face,
Face net, etc.

**Video Classification**

RNN, LSTM,etc.

Training

Intel® MKL/ NVIDIA® CUDA/ OpenCL

Container

Docker

OS

**Linux**

**Windows**

**>AI inferne cuSysts m**

C Cc c

the Mustang-FlOO-AlO is based on Intel® Arria 10GX 1150 FPGA. Both are designated for inference enhancement. The CPU acceleration card Mustang-200 combined two Intel® Core ULT CPUs and offers additional boost for training systems.

Computing accelerator models

I Microsoft

CNTK

Chainer

**Accelerator**

CPU

Mustang-200

**Accelerator**

FPGA

Mustang-F100-A10

**Accelerator**

VPU

Mustang-V100-MX8

•lntelArria10GX1150 FPGA

* PCIe Gen3 x 8
* Low profile , half size, double slot
* Intel Movidius solution
* 8 x Myriad X VPU
* PCIe Gen2 x4
* Low profile , half size, single slot

Mustang-V100

**Caffe2**

**CNTK**

**MXNet**

**Neon**

Inference

**PyTorch**

**TensorFlow**

**Image Segmentation**

SegNet, U-Net, FCN, DeepLab vl/v2, etc.

**Voice Recognition**

DeepVoice,

WaveNet, etc

CoreML (iOS)/OpenVINO/ TensorFlow Lite (Android) /TensorRT (Nvidia)

Intel Kabylake ULT

Intel FPGA

Intel VPU

**Caffe**

**XEON
inside\***

In addition, the performance of both the DL training and optimized inference models can be further enhanced by adding heterogeneous low profile computing acceleration cards such as the Mustang-FlOO-AlO with Intel® FPGA or Mustang-V100-MX8 with Intel® VPU or GPU card. The combination of GRAND-C422, TANK-870AI, the accelerator cards and a DL toolkit form lEI's Al ready solution.

Mustang-F100

FPGA

VPU

GPU

Deep Learning
Toolkit

**OpenVINO™**

TENSORRT

**IEI Al Ready**

***[www.ieiworld](http://www.ieiworld). com***



**GRAND-C422-20D**

**> Hardware Architecture**

Intel® Xeon® Processor W Family

TheXeon W family targets business and enterprise-class performance worksta­tions, situated below the scalable Xeon family and above the Xeon E3. Xeon W come with more cores, more PCIe lanes, ECC memory, generally almost all available technologies offered by the chip, volume management and various RAS featu res.

**• I/O interface**

-IPMIVGAdisplay

-lOGbE RJ45 LAN port

-IGbE GJ45 LAN port

-4xUSB3.0

・ 2xUSB2.0

Redundant power supply

-2x 1600WCRPS module

-80PLUS Platinum level

DDR4 2666 MHz ECC Memory

-Support four DDR4 ECC RDIMM/LRDIMM

-Expandable up to 256GB



, M.2 2280 form factor

-High speed PCIe3.0 by 4 signal

1~• Twenty drive Bays

-Eight 2.5" drive bays

-Twelve 3.5" drive bays

-LED for drive activity

—• - Sup port two U.2 32Gbps PCIe SSD

Water cooling system

-Gen *4* Integrated Pump and Copper Cold Plate -Standard 120mm Aluminum Heat Exchanger

-Designed for Single FanOperation



6x PCIe 3.0slots

-Four PCIe3.0x8

(dual width)

-Two PCIe 3.0x4

(single width)

-Support Al acceleartor cards, such as dual width GPU card, FPGA card, VPU card, etc

**> Expandable to suit your needs**

* Al computing requires huge computing power, so our system can support up to 4 dual-width expansion slots (PCIe x8) and 2 single­width expansio n slots (PCIe x4)formaximumexpansion ability to meet computing needs.
* All six of the backplane slots connect directly to the system host board. This is perfect for applications that require minimal latency.

**> Storage (M2 U2 SATA)**

The GRAND-C422-20D support M.2 nVMe SSD, U.2 SSD and SATA HDD/SSD. It has a built-in M.2 nVMe port and 20 bays of HDD/SSD slots inclu ding two U.2 SDD b lots. The GRAND-C422-20D su)ports M.2 sol i d-stb te disk which isthe next-ge Deration small-sized form factor introduced by Intel after mSATA. It has better performance than general SATA SSD but it is lighter and more power-saving.

2 x Single-width PCIe slots support PCIe 3.0 x4 signals

4 x Dual-width PCIe slots support PCIe 3.0 x8 signals

2 x U.2 drive bays, compatiable to SATA SSD/HDD

lx M.22280 nVMe



**> U.2 SSD**

**Water Cooling System for CPU**

U.2 uses the same concept as a general hard disk. With a connection cable, a hard disk can be installed in the case without occupying the space of the motherboard. Therefore, M.2 and U.2 interfaces can be coexistence because they have different application environment. M.2 is more suitable for laptops or microcomputers, and U.2 is more suitable on a desktop or server. The U.2 interface features high-speed, low-latency, low-power, NVMe standard protocol, and PCIe 3.0 x4 channel. The theoretical transmission speed is up to 32Gbps, while SATA is only 6Gbps, which is 5 times faster than SATA.

The U.2 interface utilizes the existing physical interface, but the bandwidth is faster. The four-channel design makes the bandwidth upgrade from PCIe x2 to PCIe 3.0 x4, which is several times more than SATA interface. The U.2 interface combines the features of SATA and SAS, and uses the signal pin to fill the connector of the SAS interface. The L-type foolproof design, except the PCIe interface, also compatible with various mainstream hard disc interface such as SATA, SAS and SATA E.

MB/s

***0***

***500***

***1000***

***1500***

***2000***

IEI uses the latest 14nm Intel Xeon Processor W family which uses the LGA2066 interface and Skylake-SP architecture with 4, 6, 8,10, 14 and 18 core versions.

High performance means higher power consumption, therefore IEI designed water cooling system for CPU with smaller size, higher efficiency cooling system makes CPU cooler and keep the high performance, and it can support up to 250W TDP.

**Cooler Size**

**Working Noise**

**Cooling Efficiency**

Small

Small

**Water Cooling**

Large

**Air Cooling**

Large

Worse

***[www.ieiworld](http://www.ieiworld). com***

10



Selection Guide

**Model**

GRAND-C422-20D

Chassis

Motherboard

|  |  |  |  |
| --- | --- | --- | --- |
| **Dimensions (H x W x D)** |  | 176.15 mm x 480.94mmx 644 mm |  |
| **System Fan** |  | 2x120mm,12VDC |  |
| **Chassis Construction** |  | 4U, Rackmount |  |
| **System Cooling** |  | 2 x Cooling Fans with Smart Fan |  |
| **CPU** |  | SupportLGA・2066lntel®Xeon®Wfamily processor |  |
| **Processor Cooling** |  | Water cooling system |  |
| **Chipest** | C422 |
|  | Total slot: 4 x DDR4 ECC RDIMM/LRDIMM |
| **Memory** |  |  |  |
| Memory expandable up to:256GB (4 x 64GB) |

Security

IPMI

**TPM** 1 x TPM 2.0 Pin header

**IPMI Solution**

**Hard Drive**

Storage

**M.2**

IPMI LAN port, IPMI VGA
12 x2.5,r/3.5u drive bay
8 x ***2.5"*** drive bay

1 x M.2 built in on SBC

**U.2**

Networking

**Ethernat IC**

2 x U.2 SSD drive bay compatible to SATA

1 GbE NIC: Intel布i210-AT with NCSI support
10 GbE NIC: Aquantia AQC107

**USB 3.0**

**USB 2.0**

I/O Interface

**Ethernet**

**Display**

1 x1GbE RJ45 combo LAN ports / IPMI
1j< 10GbE RJ45 LAN port

1 x IPMI VGA display

**Buttons**

**COM port**

Internal I/O

**USB 3.0**

2 x USB 3.0 pin header

Power button

2 x RS232 pin header

**USB 2.0**

Indicator

**LEDs**

10 GbE, Status, LAN, Storage Expansion Port Status

**LCM**

LCM, 2 buttons

1 x USB DOM header

**Expansion**

**PCIe**

4xPCIe3.0x8

2 x PCIe 3.0x4

Power

Reliability

OS

**Power Input
Power Consumption
Type/Watt
Operating Temperature
Relative Humidity
Weight
Certification**

**support OS**

110-240 AQ47-63HZ

In Operation: 285W

Redundant Power 1600W

0~40°C

5 to 95% non-condensing, wet bulb: 27°C.

23.59 kg

CE/FCC

Windows server 2016

Linux

**GRAN D-C422-20D-S1A1-R10 GRAND-C422-20D-S1B2-R10 G RAND-C422-20D-S1C3-R10 GRAND-C422-20D-S1D3-R10**

**GRAN D-C422-20D-S1E4 R10**

> Ordering information

20-bay(3.5M x12z 2.5" x 8) 4U Rackmount Intel® Xeon® W-2123 with C422 chipset, 32G DDR4w/ECC, 6 x PCIe expansion slot, and 1600W redundant PSU, RoHS

20-bay(3.5" x12, 2.5" x 8) 4U Rackmount, Intel® Xeon® W-2133 with C422 chipset, 64G DDR4 w/ECC, 6 x PCIe expansion slot, and 1600W redundant PSU, RoHS

20-bay(3.5M x12z 2.5M x 8) 4U Rackmount Intel® Xeon® W-2145 with C422 chipset, 128G DDR4 w/ECC, 6 x PCIe expansion slot, and 1600W redundant PSU, RoHS

20-bay(3.5M x12z 2.5M x 8) 4U Rackmount Intel® Xeon® W-2155 with C422 chipset, 128G DDR4 w/ECC, 6 x PCIe expansion slot, and 1600W redundant PSU, RoHS

20-bay(3.5M x12z 2.5M x 8) 4U Rackmount Intel® Xeon® W-2195 with C422 chipset, 256G DDR4 w/ECC, 6 x PCIe expansion slot, and 1600W redundant PSU, RoHS

Packing list

Flat head screws

(for 2.5" HDD)

Flat head screws

(for 3.5" HDD)

1 x Cat5e LAN cable 1 x QIG

2 x Power cord

1 x Cat6A LAN cable

Headquarters America

威強電工業電腦 IEI Integration Corp. IEI Technology USA Corp.

**No. 29, Zhongxing Rd., Xizhi Dist**・，**New laipei City 221, Taiwan 138 University Parkway, Pomona, CA 91768**

**TEL : +886-2-86916798 / +886-2-26902098 FAX : +886-2-66160028 TEL: +1 -909-595-2819 FAX : +1-909-595-2816 sales@ieiworld.com [www.ieiworld.com](http://www.ieiworld.com) sales@usa.iei world .com usa.ieiworld.com**

China

威强电工业电脑 IEI Integration (Shanghai) Corp.

上海市闵行莘庄工业区申富路**515**号

**515, Shen Fu Rd., Xin Zhuang Industrial Develop Zone, Shanghai, 201108, China TEL:+86-21-3116-7799 FAX:+86-21-3462-7797**

**sales@ieiworld.com.cn [www.ieiworld.com.cn](http://www.ieiworld.com.cn)**

2019.01

