

GPU型gp1

• 计算：

- 处理器与内存配比为1:16，搭配768GB内存
- 处理器：采用两颗英特尔® 至强® Xeon Gold处理器，最低基频3.0GHz，最大睿频可达4.0GHz

• GPU：

- gp1-v100：配置八颗NVIDIA V100S GPU，单卡显存32GB，支持PCIe 4.0接口
- gp1-t4：配置八颗NVIDIA T4 GPU，单卡显存16GB，支持PCIe 4.0接口
- gp1-t4：配置八颗NVIDIA A100 GPU，单卡显存40GB，支持PCIe 4.0接口

• 存储：

- 搭载本地480GB SATA SSD系统盘
- 支持高性能型云硬盘，具体以订购页为准
- 提供多种类型本地数据盘
- IO性能卓越

• 网络：

- 支持IPv4+IPv6双栈网络
- 高网络收发包能力

• 适用场景：

- 深度学习推理计算加速场景
- Quadro工作站
- 人工智能学习等

规格	配置	内存	GPU	卡间互 联协议	GPU显 存	本地盘	数据盘	业务网 卡
bms.gp1-v100.12xlarge.16	2*Intel Xeon Gold 6248R (24cores, 3.0GHz)	768GB	8*NVIDIA A V100S	PCIe 4.0	8*32GB	系统盘：2*480GB SATA SSD RAID1 数据盘：2*960GB SATA SSD RAID1	读：最高15万IOPS 写：最高6万IOPS	1 x 2*25GE

bms.gp1-t4.12xlarge.16	2*Intel Xeon Gold 6248R (24cores, 3.0GHz)	768GB	8*NVIDIA T4	PCIe 4.0	8*16GB	系统盘：2*480GB SATA SSD RAID1 数据盘：2*960GB SATA SSD RAID1	读：最高15万 IOPS 写：最高6万 IOPS	1 x 2*25GE
bms.gp1-a100.12xlarge.16	2*Intel Xeon Gold 6248R(24cores, 3.0GHz)	768GB	8*NVIDIA A100	PCIe 4.0	8*40GB	系统盘：2*480GB SATA SSD 数据盘：4*1.92TB NVMe SSD	-	1 x 2*25GE

GPU型ci5

- 计算：
 - 处理器与内存配比为1:16、1：19，搭配1024GB内存
 - 处理器：采用两颗英特尔® 至强® Xeon Gold处理器
- GPU：
 - ci5-A10：配置八颗NVIDIA A10 GPU卡，单卡显存24GB，支持PCIe 5.0接口
 - ci5-KLXR200：配置八颗昆仑芯R200 GPU卡，单卡显存16GB，支持PCIe 4.0接口
- 存储：
 - 搭载本地480GB SATA SSD系统盘
 - 支持高性能型云硬盘，具体以订购页为准
 - 提供多种类型本地数据盘
 - IO性能卓越
- 网络：
 - 支持IPv4+IPv6双栈网络
 - 高网络收发包能力
- 适用场景：
 - 深度学习推理计算加速场景
 - Quadro工作站

- 人工智能学习等

规格	配置	内存	GPU	卡间互 联协议	GPU显 存	本地盘	数据盘	业务网 卡
bms.ci5- A10.16xl arge.16	CPU : 2*Intel Xeon Gold 6430(32 cores, 2.1GHz)	1024GB	8*NVIDI A A10	PCIe 5.0	8*24GB	系统 盘 : 2*480G B SATA SSD 数据 盘 : 2*1.92T NVMe	读 : 最 高35万 IOPS 写 : 最 高31万 IOPS	1 x 2*25GE
bms.ci5- KLXR20 0.14xlar ge.19	2*Intel Xeon Gold 6348(28 cores, 2.6GHz)	1024GB	8*KUNL UNIXIN R200	PCIe 4.0	8*16GB	系统 盘 : 2*480G B SATA SSD 数据 盘 : 2*1.92T NVMe	读 : 最 高36万 IOPS 写 : 最 高27万 IOPS	1 x 2*25GE

GPU型ci2a

- 计算：
 - 处理器与内存配比为1:11，搭配1024GB内存
 - 处理器：采用两颗AMD 7K62处理器，最低基频2.6GHz，最大睿频可达3.3GHz
 - 配置八颗NVIDIA A800 GPU卡，单卡显存40GB，支持NVLink接口，提供AI领域模型训练能力
- 存储：
 - 搭载本地480GB SATA SSD系统盘
 - 提供NVME数据盘
 - IO性能卓越
- 网络：
 - 支持IPv4+IPv6双栈网络
 - 高网络收发包能力
- 适用场景：
 - 深度学习推理计算加速场景
 - Quadro工作站
 - 人工智能学习等

规格	配置	内存	GPU	卡间互 联	GPU显存	本地盘	业务网 卡	参数网
----	----	----	-----	----------	-------	-----	----------	-----

				协议				
bms.ci2a.24xlarge.11	2*AMD 7K62(48core, 2.6GHz)	1024GB	8*NVIDIA A800	NVLink	8*40GB	系统盘：2*480GB SATA SSD 数据盘：4*1.92TB NVMe	1 x 2*25GE	8*200G IB

NPU型np1

- 计算：
 - 处理器与内存配比为1:14、1:10，搭配384GB内存
 - 处理器：采用两颗Intel Xeon Gold 6132处理器/Intel 5218R Gold处理器，基频2.1GHz或者2.6GHz，最大睿频可达3.7GHz
 - 配置八颗Atlas300系列推理芯片的规格，提供超强推理计算加速性能
- 存储：
 - 搭载本地480GB SATA SSD系统盘
 - 搭配多种类型本地存储数据盘
 - 支持高性能型云硬盘，具体以订购页为准
 - IO性能卓越
- 网络：
 - 支持IPv4+IPv6双栈网络
 - 高网络收发包能力
- 适用场景：
 - 深度学习推理计算加速，比如图片分类识别、语音识别等在线推理业务场景

规格	配置	内存	NPU	NPU显存	本地盘	数据盘	业务网卡
bms.np1.7xlarge.14	2*Intel Xeon Gold 6132(14cores, 2.6GHz)	384GB	8*Atlas300	8*24GB	系统盘：2*480GB SATA SSD RAID1 数据盘：4*960GB SATA SSD RAID10	读：最高16万IOPS 写：最高8万IOPS	1 x 2*10GE

bms.np1.10xlarge.10	2*Intel 5218R Gold (20cores, 2.1GHz)	384GB	8*Atlas300l	8*24GB	系统盘：2*480GB SATA SSD RAID1 数据盘：2*1.92T NVMe	读：最高32万IOPS 写：最高37万IOPS	1 x 2*25GE
---------------------	--------------------------------------	-------	-------------	--------	---	----------------------------	------------

NPU型ci5

- 计算：
 - 处理器与内存配比为1:19，搭配1024GB内存
 - 处理器：采用两颗英特尔® 至强® Xeon Gold 处理器，基频为2.6GHz，最大睿频可达3.5GHz
 - 配置八颗Atlas300V推理芯片的规格，提供超强推理计算加速性能
- 存储：
 - 搭载本地480GB SATA SSD系统盘
 - 搭配NVME数据盘
 - 支持高性能型云硬盘，具体以订购页为准
 - IO性能卓越
- 网络：
 - 支持IPv4+IPv6双栈网络
 - 高网络收发包能力
- 适用场景：
 - 深度学习推理计算加速，比如图片分类识别、语音识别等在线推理业务场景

规格	配置	内存	NPU	NPU显存	本地盘	数据盘	业务网卡
bms.ci5-Atlas300V.14xlarge.19	2*Intel Xeon Gold 6348(28cores, 2.6GHz)	1024GB	8*Atlas300V	8*24GB	系统盘：2*480GB SATA SSD 数据盘：2*1.92T NVMe	读：最高32万IOPS 写：最高37万IOPS	1 x 2*25GE

NPU型ci2k、ci2ek

采用ARM架构，搭载国产化鲲鹏920-5250处理器，并配置八颗Ascend910B芯片，在性能、安全、稳定性等方面全面升级，满足信创业务需求。

- 配置：
 - 网络：4*GE（板载）+2*10GE（管理）+2*25GE（业务+存储）+ 8*200G RoCE v2（参数）+2*100GE/2*25GE RoCE（存储PFS）
 - 单卡算力：376TFLOPS（FP16）

- 电源：4*2600W交流电源
- 计算：
 - 处理器与内存配比为1:11，搭配2048GB内存
 - 处理器：采用4颗Kunpeng920-5250处理器，基频2.6GHz
 - 配置八颗Ascend910B芯片，提供AI领域的模型训练能力
- 存储：
 - 搭载本地480GB SATA SSD系统盘
 - 搭配NVME数据盘
 - IO性能卓越
- 网络：
 - 支持IPv4+IPv6双栈网络
 - 高网络收发包能力
- 适用场景：
 - 深度学习推理计算加速，比如图片分类识别、语音识别等在线推理业务场景

规格	配置	内存	NPU	NPU显存	本地盘	板载网卡	管理网卡	业务网卡	参数网卡	样本网卡
bms.ci 2k-200G-25G-4x3.2T.48xlarge.11	4*Kunpeng920-5250(48cores, 2.6GHz)	2048GB	8*Ascend 910B	8*64GB	系统盘：2*480GB SATA SSD 数据盘：4*3.2TNVMe	4*GE	1 x 2*10GE	1 x 2*25GE	8*200G RoCE v2	1 x 2*25G RoCE v2
bms.ci 2k-200G-25G.48xlarge.11	4*Kunpeng920-5250(48cores, 2.6GHz)	2048GB	8*Ascend 910B	8*64GB	系统盘：2*480GB SATA SSD 数据盘：2*3.84TNVMe	4*GE	1 x 2*10GE	1 x 2*25GE	8*200G RoCE v2	1 x 2*25G RoCE v2

bms.ci2k-200G-100G.48xlarge.11	4*Kunpeng920-5250(48cores, 2.6GHz)	2048GB	8*Ascend910B	8*64GB	系统盘：2*960G 数据盘：2*3.2TNVMe	4*GE	1 x 2*10GE	1 x 2*25GE	8*200G RoCEv2	1 x 2*100G RoCEv2
ebms.ci2ek-200G-25G.48xlarge.11	4*Kunpeng920-5250(48cores, 2.6GHz)	2048GB	8*Ascend910B	8*64GB	系统盘：云盘 数据盘：2*480GB SATA SSD+ 2*3.84TNVMe	4*GE	1 x 2*10GE	32Gbit/s	8*200G RoCEv2	1 x 2*25G RoCEv2