目录

计费说明

计费概述

根据您使用的实例资源规格，以及每个实际运行时长按秒计费，按小时出账。计费时长从下载容器镜像（docker pull）开始至实例停止运行(进入Succeeded/Failed状态)结束。

根据您创建容器实例的方式，支持以下两种计费方式：

按vCPU和内存

根据您创建时指定的vCPU和内存进行计费。对于不支持的vCPU和内存规格，系统将自动进行规整，计费按照规整后CPU和Memory进行计费，具体规则请参考指定KCI Pod规格。

账单计算

费用 = （容器组CPU核数 x 单价 + 容器组内存大小 x 单价）x 容器组运行时间

按云服务器套餐

若在创建容器实例时指定了云服务器套餐，将根据您指定的云服务器套餐进行计费。在对实例规格有特殊需求的场景（如：网络吞吐量、网卡队列数等），您可以指定容器实例底层所使用的云服务器套餐，来获取相应机型及规格的能力。

账单计算

费用= （容器组CPU核数 x 对应机型单价 + 容器组内存大小 x 对应机型单价）x 容器组运行时间

备注：

- 按秒计费，按小时扣费。例如，10:00-11:00的账单会在12:00之前生成，具体以系统出账时间为准。
- 当汇总账单在单一计费周期内金额不足0.01元时，需按照0.01元补齐，即账单至少金额为0.01元。
- 运行时间指容器组开始创建到运行完成的时间，包括创建中和运行中状态，容器组进入成功或失败状态后即停止计费。

定价

请参考容器实例定价。

资源规格

目前通用型容器实例支持的实例规格如下（不包含指定云主机规格方式创建的容器实例）：

实例规格

CPU/核	内存/GiB
1	2，3，4
2	4，5，6，7，8
4	8-16，1GiB递增
8	16-64，1GiB递增
16	32-64，1GiB递增

地域及可用区

地域及代码	可用区
华北1（北京）cn-beijing-6	可用区A cn-beijing-6a 可用区B cn-beijing-6b
华东1（上海）ch-shanghai-2	可用区A cn-shanghai-2a 可用区B cn-shanghai-2b

支持的云服务器类型

在对实例规格或性能有特殊需求的场景（如：网络吞吐量、网卡队列数等），您可以指定容器实例底层所使用的云服务器套餐规格来创建实例。云服务器套餐定义说明如下：

- 系列：指金山云提供的不同硬件代际的服务器类型集合。随系列数增加，性能依次增强。例如，标准型S6性能优于标准型S3。
- 套餐：指云服务器的具体配置，例如，N3.2B套餐中的云服务器具体配置为2个vCPU和4G内存。
- PPS：网络收发包能力，指每秒可处理的数据包数量，数值为收发包两个方向之和。
- 内网吞吐量：内网每秒能传输的最大数据量。

容器实例提供的云服务器类型包括：

云服务器类型	子类型	描述
通用型/标准型	通用型N3 标准型S6 标准型S4 标准型S3	提供平衡的计算、内存和网络资源，适用于大多数类型和规模的企业级应用。
IO优化型	IO优化型I4 IO优化型I3	具有高随机 IOPS、高吞吐量、低访问延时等特点，适用于高负载数据库等要求高磁盘IO负载、低延迟高吞吐的场景。
计算优化型	计算优化型C5 计算优化型C4 计算优化型C3	适用于MMOGPG、MOBA游戏前端、高负载Web等要求高计算性能和高并发读写的场景。
大数据型	大数据型D4	适合 Hadoop 分布式计算、海量日志处理、分布式文件系统和大型数据仓库等吞吐密集型应用。
性能保障型	性能保障型X6 性能保障型X5	企业级机型，计算性能强劲稳定，适用于CPU消耗型业务及各种类型和规模的企业级应用等场景。

通用型/标准型云服务器

1. 通用型N3
    - 特点：
        - Intel Xeon Platinum 8168 处理器，DDR4内存
        - 支持数据盘类型：EBS 3.0和EHDD
        - 支持系统盘类型：EBS 3.0和EHDD
    - 使用场景:

      适用于各种类型和规模的企业级应用等场景。
    - 可用区域： 华北1（北京）可用区A
      华北1（北京）可用区B
      华东1（上海）可用区A
      华东1（上海）可用区B
    - 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
N3.2B	2	4	30	1	2	2	6
N3.2C	2	8	30	1	2	2	6
N3.4B	4	8	50	1.5	2	2	6
N3.4C	4	16	50	1.5	2	4	6
N3.4D	4	32	50	1.5	2	4	6
N3.8B	8	16	80	2.5	2	8	10
N3.8C	8	32	80	2.5	2	8	10
N3.8D	8	64	80	2.5	2	8	10
N3.12B	12	24	90	4	2	12	15
N3.12C	12	48	90	4	2	12	15

N3.16B	16	32	100	5	2	16	20
N3.16C	16	64	100	5	2	16	20
N3.24B	24	48	100	7	4	16	20
N3.24C	24	96	100	7	4	16	20
N3.32B	32	64	100	8	5	16	20
N3.32C	32	128	100	8	5	16	20

2. 标准型S6

- 特点：
  - Intel®Xeon®Platinum 8358P（Icelake）CPU处理器，最新八通道DDR4内存
  - 支持数据盘类型：EBS 3.0和EHDD
  - 支持系统盘类型：EBS 3.0和EHDD
- 使用场景：

  适用于各种类型和规模的企业级应用及数据库等场景。
- 可用区域： 华北1（北京）可用区A
  华北1（北京）可用区B
  华北1（北京）可用区C
  华北1（北京）可用区E
  华东1（上海）可用区A
  华东1（上海）可用区B
- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
S6.1A	1	1	25	1.5	2	1	6
S6.1B	1	2	25	1.5	2	1	6
S6.1C	1	4	25	1.5	2	1	6
S6.1D	1	8	25	1.5	2	1	6
S6.2A	2	2	30	1.5	2	2	6
S6.2B	2	4	30	1.5	2	2	6
S6.2C	2	8	30	1.5	2	2	6
S6.2D	2	16	30	1.5	2	2	6
S6.4A	4	4	50	2	2	4	6
S6.4B	4	8	50	2	2	4	6
S6.4C	4	16	50	2	2	4	6
S6.4D	4	32	50	2	2	4	6
S6.8A	8	8	80	3	2	8	10
S6.8B	8	16	80	3	2	8	10
S6.8C	8	32	80	3	2	8	10
S6.8D	8	64	80	3	2	8	10
S6.12A	12	12	110	4	3	12	15
S6.12B	12	24	110	4	3	12	15
S6.12C	12	48	110	4	3	12	15
S6.12D	12	96	110	4	3	12	15
S6.16A	16	16	150	6	4	16	20
S6.16B	16	32	150	6	4	16	20
S6.16C	16	64	150	6	4	16	20
S6.16D	16	128	150	6	4	16	20
S6.32A	32	32	300	12	4	16	20
S6.32B	32	64	300	12	4	16	20
S6.32C	32	128	300	12	4	16	20
S6.32D	32	256	300	12	4	16	20

3. 标准型S4

- 特点：
  - Intel(R) Xeon(R) Gold 6240 CPU处理器，最新六通道DDR4内存

- 支持数据盘类型：EBS 3.0和EHDD
- 支持系统盘类型：EBS 3.0和EHDD

- 使用场景：

  适用于各种类型和规模的企业级应用及数据库等场景。

- 可用区域： 华北1（北京）可用区A
  华北1（北京）可用区C
  华东1（上海）可用区A
  华东1（上海）可用区B

- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
S4.1A	1	1	25	1.5	2	1	6
S4.1B	1	2	25	1.5	2	1	6
S4.1C	1	4	25	1.5	2	1	6
S4.1D	1	8	25	1.5	2	1	6
S4.2A	2	2	30	1.5	2	2	6
S4.2B	2	4	30	1.5	2	2	6
S4.2C	2	8	30	1.5	2	2	6
S4.2D	2	16	30	1.5	2	2	6
S4.4A	4	4	50	2	2	4	6
S4.4B	4	8	50	2	2	4	6
S4.4C	4	16	50	2	2	4	6
S4.4D	4	32	50	2	2	4	6
S4.8A	8	8	85	3	2	8	10
S4.8B	8	16	85	3	2	8	10
S4.8C	8	32	85	3	2	8	10
S4.8D	8	64	85	3	2	8	10
S4.12B	12	24	85	3	2	12	15
S4.12C	12	48	85	3	2	12	15
S4.16A	16	16	100	6	2	16	20
S4.16B	16	32	100	6	2	16	20
S4.16C	16	64	100	6	2	16	20
S4.16D	16	128	100	6	2	16	20
S4.24B	24	48	100	6	3	16	20
S4.24C	24	96	100	6	3	16	20
S4.32B	32	64	150	10	5	16	20
S4.32C	32	128	150	10	5	16	20

4. 标准型S3

- 特点：

  - Intel Xeon Gold 6132处理器，DDR4内存
  - 支持数据盘类型：本地SSD、EBS 3.0、EHDD
  - 支持系统盘类型：本地SSD

- 使用场景：

  适用于各种类型和规模的企业级应用及数据库等场景。

- 可用区域： 华北1（北京）可用区A
  华北1（北京）可用区B
  华北1（北京）可用区C
  华东1（上海）可用区B

- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)

S3.1A	1	1	20	1	3	1	6
S3.1B	1	2	20	1	3	1	6
S3.1C	1	4	20	1	3	1	6
S3.2A	2	2	25	1	3	2	6
S3.2B	2	4	25	1	3	2	6
S3.2C	2	8	25	1	3	2	6
S3.4A	4	4	45	1.5	3	4	6
S3.4B	4	8	45	1.5	3	4	6
S3.4C	4	16	45	1.5	3	4	6
S3.4D	4	32	45	1.5	3	4	6
S3.8A	8	8	85	2	3	8	10
S3.8B	8	16	85	2	3	8	10
S3.8C	8	32	85	2	3	8	10
S3.8D	8	64	85	2	3	8	10
S3.12B	12	24	85	2	3	12	15
S3.12C	12	48	85	2	3	12	15
S3.16A	16	16	85	3	3	16	20
S3.16B	16	32	85	3	3	16	20
S3.16C	16	64	85	3	3	16	20
S3.24B	24	48	85	3	4	16	20
S3.24C	24	96	85	3	4	16	20
S3.32B	32	64	85	5	6	16	20
S3.32C	32	128	85	5	6	16	20

IO优化型云服务器

1. IO优化型I4
   - 特点：
     - 采用Intel(R) Xeon(R) Gold 6240 CPU处理器，最新六通道DDR4内存
     - 支持数据盘类型：本地SSD、EBS 3.0、EHDD
     - 支持系统盘类型：EBS 3.0、EHDD
     - 本地数据盘采用 NVMe SSD
       - 单盘随机读性能高达62万 IOPS（4KB块大小），顺序读吞吐能力高达3.2GB/s（128KB块大小）
       - 整机随机读性能高达200万 IOPS（4KB块大小），顺序读吞吐能力均高达12GB/s（128KB块大小）
   - 注意：该机型属于本地直连盘机型，存储在直连盘上的数据有丢失数据的风险，例如实例所在物理机发生硬件故障时，因此请勿在本地直连盘上存储长期需要保存的数据。您可以在应用层做数据冗余，保证数据可靠性；也可以通过容灾组来保证实例分布在不同物理机上，保证底层容灾能力。如果您需要高可靠性的数据存储，建议您选择其他云盘机型。
   - 使用场景：

     电商、游戏、媒体等I/O密集型应用场景，满足用户对块存储高随机IO性能以及低时延需求。
   - 可用区域：

     华北1（北京）可用区A
     华东1（上海）可用区A
   - 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
I4.16C	16	64	150	6	1	16	20
I4.32C	32	128	250	12	1	16	20
I4.64C	64	256	500	23	1	16	20

2. IO优化型I3
   - 特点：
     - Intel Xeon Platinum 8168 处理器，DDR4内存

- 支持数据盘类型：本地SSD、EBS 3.0、EHDD
- 支持系统盘类型：本地SSD

- 使用场景：

  适合于MMOGPG、MOBA游戏前端、高负载数据库、高负载Web等场景。

- 可用区域：

  华北1（北京）可用区A
  华北1（北京）可用区B
  华东1（上海）可用区A
  华东1（上海）可用区B

- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
I3.2B	2	4	30	1	2	2	6
I3.2C	2	8	30	1	2	2	6
I3.4B	4	8	50	1.5	2	4	6
I3.4C	4	16	50	1.5	2	4	6
I3.4D	4	32	50	1.5	2	4	6
I3.8B	8	16	80	2.5	2	8	10
I3.8C	8	32	80	2.5	2	8	10
I3.8D	8	64	80	2.5	2	8	10
I3.12B	12	24	90	4	2	12	15
I3.12C	12	48	90	4	2	12	15
I3.16B	16	32	100	5	2	16	20
I3.16C	16	64	100	5	2	16	20
I3.24B	24	48	100	7	4	16	20
I3.24C	24	96	100	7	4	16	20
I3.32B	32	64	100	8	5	16	20
I3.32C	32	128	100	8	5	16	20

计算优化型云服务器

1. 计算优化型C5

- 特点：

  - Intel(R) Xeon(R) Gold 6242R CPU处理器，最新六通道DDR4内存
  - 支持数据盘类型：EBS 3.0、EHDD
  - 支持系统盘类型：EBS 3.0、EHDD

- 使用场景：

  企业级机型，计算性能强劲稳定，适用于CPU消耗型业及高负载数据库、视频编码、高负载Web等各种类型和规模的企业级应用场景。

- 可用区域： 华北1（北京）可用区A
  华北1（北京）可用区B
  华东1（上海）可用区A
  华东1（上海）可用区B

- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
C5.2B	2	4	30	1.5	3	2	6
C5.2C	2	8	30	1.5	3	2	6
C5.4B	4	8	50	2	3	4	6
C5.4C	4	16	50	2	3	4	6
C5.8B	8	16	80	3	4	8	10
C5.8C	8	32	80	3	4	8	10

C5.12B	12	24	110	4	6	12	15
C5.12C	12	48	110	4	6	12	15
C5.16B	16	32	150	6	8	16	20
C5.16C	16	64	150	6	8	16	20
C5.24B	24	48	200	8	8	16	20
C5.24C	24	96	200	8	8	16	20
C5.32B	32	64	250	10	8	16	20
C5.32C	32	128	250	10	8	16	20
C5.64B	64	128	500	20	8	16	20
C5.64C	64	256	500	20	8	16	20

2. 计算优化型C4

- 特点：
  - Intel(R) Xeon(R) Gold 6254 CPU处理器，最新六通道DDR4内存
  - 支持数据盘类型：EBS 3.0、EHDD
  - 支持系统盘类型：EBS 3.0、EHDD
- 使用场景：

  适用于高负载数据库、视频编码、高负载Web等场景
- 可用区域： 华北1（北京）可用区A
  华北1（北京）可用区B
  华东1（上海）可用区A
  华东1（上海）可用区B
- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
C4.2B	2	4	30	2	2	2	6
C4.2C	2	8	30	2	2	2	6
C4.4B	4	8	60	2.5	4	4	6
C4.4C	4	16	60	2.5	4	4	6
C4.8B	8	16	100	3	4	8	10
C4.8C	8	32	100	3	4	8	10
C4.12B	12	24	120	4	4	12	15
C4.12C	12	48	120	4	4	12	15
C4.16B	16	32	150	6	4	16	20
C4.16C	16	64	150	6	4	16	20
C4.24B	24	48	150	12	6	16	20
C4.24C	24	96	150	12	6	16	20
C4.32B	32	64	250	12	8	16	20
C4.32C	32	128	250	12	8	16	20
C4.64B	64	128	250	12	8	16	20
C4.64C	64	256	250	12	8	16	20

3. 计算优化型C3

- 特点：
  - Intel Xeon Gold 6146 处理器，DDR4内存
  - 支持数据盘类型：本地SSD、EBS 3.0、EHDD
  - 支持系统盘类型：本地SSD
- 使用场景：

  适合于MMOGPG、MOBA游戏前端、高负载数据库、高负载Web等场景。
- 可用区域： 华北1（北京）可用区B
  华东1（上海）可用区A
  华东1（上海）可用区B

- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
C3.2B	2	4	30	1.5	7	2	6
C3.4B	4	8	60	2.5	7	4	6
C3.4C	4	16	60	2.5	7	4	6
C3.8B	8	16	100	3	7	8	10
C3.8C	8	32	100	3	7	8	10
C3.16B	16	32	100	6	7	16	20
C3.16C	16	64	100	6	7	16	20
C3.32B	32	64	100	10	8	16	20
C3.32C	32	128	100	10	8	16	20

大数据型云服务器

1. 大数据型D4

- 特点：
  - 采用Intel(R) Xeon(R) Gold 5220 CPU处理器，最新六通道DDR4内存
  - 支持数据盘类型：EBS 3.0、EHDD
  - 支持系统盘类型：EBS 3.0、EHDD
  - 本地数据盘采用 SATA HDD
- 使用场景：

  适合 Hadoop 分布式计算、海量日志处理、分布式文件系统和大型数据仓库等吞吐密集型应用。
- 可用区域：

  华北1（北京）可用区A
- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
D4.8B	8	16	80.0	3	2	4	10
D4.8C	8	32	80.0	3	2	4	10
D4.16B	16	32	100.0	6	2	8	20
D4.16C	16	64	100.0	6	2	8	20
D4.32B	32	64	150.0	8	2	16	20
D4.32C	32	128	150.0	8	2	16	20
D4.64B	64	128	150.0	11	2	16	20

性能保障型云服务器

1. 性能保障型X6

- 特点：
  - Intel(R) Xeon(R) Platinum 8358P（Icelake）CPU处理器，最新八通道DDR4内存
  - 支持数据盘类型：EBS 3.0和EHDD
  - 支持系统盘类型：EBS 3.0和EHDD
- 使用场景：

  企业级机型，计算性能强劲稳定，适用于CPU消耗型业务及各种类型和规模的企业级应用等场景。
- 可用区域： 华北1（北京）可用区A
  华北1（北京）可用区C
  华东1（上海）可用区A
  华东1（上海）可用区B
- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
X6.2B	2	4	30.0	1.5	2	2	6
X6.2C	2	8	30.0	1.5	2	2	6
X6.4B	4	8	50.0	2	3	4	6
X6.4C	4	16	50.0	2	3	4	6
X6.8B	8	16	80.0	3	4	8	10
X6.8C	8	32	80.0	3	4	8	10
X6.12B	12	24	110.0	4	6	12	15
X6.12C	12	48	110.0	4	6	12	15
X6.16B	16	32	150.0	6	8	16	20
X6.16C	16	64	150.0	6	8	16	20
X6.24B	24	48	200.0	8	8	16	20
X6.24C	24	96	200.0	8	8	16	20
X6.32B	32	64	250.0	10	8	16	20
X6.32C	32	128	250.0	10	8	16	20
X6.64B	64	128	500.0	20	8	16	20
X6.64C	64	256	500.0	20	8	16	20

2. 性能保障型X5

- 特点：
  - Intel(R) Xeon(R) Gold 6240 CPU处理器，最新六通道DDR4内存
  - 支持数据盘类型：EBS 3.0和EHDD
  - 支持系统盘类型：EBS 3.0和EHDD
- 使用场景：

  企业级机型，计算性能强劲稳定，适用于CPU消耗型业务及各种类型和规模的企业级应用等场景。
- 可用区域： 华北1（北京）可用区A
  华东1（上海）可用区B
- 具体套餐信息：

套餐类型名	vCPU(个)	内存容量(GB)	PPS(万)	内网吞吐量(Gbps)	弹性网卡数(个)	网卡队列数(个)	单网卡私有IP(个)
X5.2C	2	8	30.0	1.5	2	2	6
X5.4C	4	16	50.0	2	3	4	6
X5.8C	8	32	80.0	3	4	8	10
X5.12C	12	48	110.0	4	6	12	15
X5.16C	16	64	150.0	6	8	16	20
X5.24C	24	96	200.0	8	8	16	20
X5.32C	32	128	250.0	10	8	16	20
X5.64C	64	256	500.0	20	8	16	20

支持的GPU云服务器类型

GPU云服务器提供GPU加速的弹性计算服务，可以用于科学计算，AI深度学习，图形图像渲染与基于GPU的音视频编解码等诸多应用场景。容器实例已支持GPU云服务器，您可以指定容器实例底层所使用的GPU云服务器套餐规格来创建实例。

容器实例提供的GPU云服务器类型包括：

GPU云服务器	实例类型	适用场景
直通（Passthrough）	GPU推理II型GN6I GPU推理计算型P3I GPU推理计算型P3IN GPU通用计算型P4V	深度学习、语音、图形/图像学习等常见训练和推理场景
vGPU	GPU虚拟化vGN5 GPU虚拟化vGN6	云端渲染和小规模、弹性、灵活的AI应用场景

注：

1. 若要在容器实例中使用GPU云服务器，必须在Pod metadata中添加Annotation来指定GPU机型，目前不支持根据容器实例的GPU Limit值自动匹配GPU机型。指定GPU机型后，在Container配置中需添加nvidia.com/gpu字段声明GPU资源。
2. 通过Deployment等控制器创建的Pod，如果Pod申请的GPU数量超过机型的GPU数量，会出现Pod在创建失败后不断重复创建的情况。为避免此情况发生，请确保Pod在nvidia.com/gpu字段中声明的GPU数量不超过Annotation中指定机型的GPU数量。

示例如下：

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: nginx
  labels:
    app: nginx
spec:
  replicas: 2
  selector:
    matchLabels:
      app: nginx
  template:
    metadata:
      labels:
        app: nginx
      annotations:
        k8s.ksyun.com/kci-instance-type: P3I.8A1 # 根据需要指定GPU机型
        k8s.ksyun.com/kci-base-system-disk-size: "50" # vGPU类型的云服务器由于机型限制，需指定系统盘规格为50G或以上
    spec:
      containers:
      - name: nginx
        image: nginx:latest
        resources:
          limits:
            nvidia.com/gpu: 1 # 指定GPU卡数
        ports:
        - containerPort: 80
      nodeName: rbkci-virtual-kubelet # 指定nodeName将Pod调度到虚拟节点上
```

GPU推理II型GN6I

该实例适用于推理场景，以及简单的训练场景。

基于NVIDIA Tesla T4，每GPU具备16GB GDDR6显存、8.1TFLOPS的单精度（FP32）计算能力和130 TOPS的INT8计算能力。

实例特点包括：

- 处理器：2.6 GHz主频的Intel® Xeon® Gold 6240 Processor
- 支持系统盘类型：EBS3.0
- 支持数据盘类型：EBS3.0

GN6I实例包括的型号和参数规格如下表所示：

型号	GPU(Tesla T4)	GPU显存(GDDR6)	vCPU(核)	内存(GiB)	网络收发包能力(万PPS)	网络带宽能力(Gbit/s)	多队列
GN6I.4A1	1颗	16GB*1	4	16	50	4	2
GN6I.8A1	1颗	16GB*1	8	32	80	5	2
GN6I.16A1	1颗	16GB*1	16	64	120	6	4
GN6I.16B2	2颗	16GB*2	16	64	120	6	4
GN6I.32B2	2颗	16GB*2	32	128	240	8	8
GN6I.32C4	4颗	16GB*4	32	128	240	8	8

GPU推理计算型P3I

该实例适用于语音识别、语音合成、图像识别等推理预测场景。

基于NVIDIA Tesla P4，每GPU具备8GB DDR5 GPU内存、5.5TFLOPS的单精度（FP32）计算能力和22TOPS的INT8计算能力，单GPU实例在深度学习的推理预测场景下相比于CPU延时降低15倍，吞吐增加60倍。

实例特点包括：

- 处理器：2.6 GHz主频的Intel® Xeon® Processor E5-2690 v4
- 支持系统盘类型：本地SSD
- 支持数据盘类型：本地SSD、EBS3.0

P3I实例包括的型号和参数规格如下表所示：

型号	GPU(Tesla P4)	GPU显存(GDDR5)	vCPU(核)	内存(DDR4)	数据盘(本地SSD)	网络收发包能力(万PPS)	网络带宽能力(Gbit/s)
P3I.8A1	1颗	8GB*1	8	16GB	0GB	20	3
P3I.14D1	1颗	8GB*1	14	32GB	0GB	20	3
P3I.14B1	1颗	8GB*1	14	120GB	500GB	20	3
P3I.28C2	2颗	8GB*2	28	240GB	1000GB	30	6

GPU推理计算型P3IN

实例特点包括：

- 处理器：2.6 GHz主频的Intel® Xeon® Processor E5-2690 v4
- 支持系统盘类型：本地SSD
- 支持数据盘类型：本地SSD、EBS3.0

具体套餐信息：该实例的适用场景以及采用的硬件与P3I一致，包括的型号和参数规格如下表所示：

型号	GPU(Tesla P4)	GPU显存(GDDR5)	vCPU(核)	内存(DDR4)	数据盘(本地SSD)	网络收发包能力(万PPS)	网络带宽能力(Gbit/s)
P3IN.4A1	1颗	8GB*1	4	16GB	120GB	10	1.5
P3IN.8B1	1颗	8GB*1	8	32GB	180GB	20	1.5
P3IN.16C2	2颗	8GB*2	16	64GB	360GB	30	3
P3IN.32D4	4颗	8GB*4	32	128GB	720GB	40	6

GPU通用计算型P4V

该实例适用于深度学习的训练场景和推理场景。

基于NVIDIA Tesla V100，每GPU具备16GB HBM2 GPU内存、15TFLOPS的单精度 (FP32) 计算能力和125TFLOPS的混合精度计算能力。

实例特点包括：

- 处理器：2.6 GHz主频的Intel® Xeon® Processor E5-2690 v4
- 支持系统盘类型：本地SSD
- 支持数据盘类型：本地SSD、EBS3.0

P4V实例包括的型号和参数规格如下表所示：

型号	GPU(Tesla V100)	GPU显存(HBM2)	vCPU(核)	内存(DDR4)	数据盘(本地SSD)	网络收发包能力(万PPS)	网络带宽能力(Gbit/s)
P4V.8A1	1颗	16GB*1	8	32GB	240GB	20	1.5
P4V.16B2	2颗	16GB*2	16	64GB	480GB	30	3
P4V.28C4	4颗	16GB*4	28	128GB	960GB	30	6
P4V.56D8	8颗	16GB*8	56	256GB	1920GB	40	8

GPU虚拟化vGN5

该实例的适用场景包括：

- 云游戏的云端实时渲染
- AR/VR的云端实时渲染
- AI（深度学习DL/机器学习ML）

实例特点包括：

- GPU：采用NVIDIA P4 GPU

- 处理器：2.6 GHz主频的Intel ® Xeon ® E5-2690 v4 (Broadwell)
- 支持系统盘类型：本地SSD
- 支持数据盘类型：本地SSD、EBS3.0
- vGPU类别
  - vCS：专门用于深度学习，提供1/2*Tesla P4、1*Tesla P4两种实例
  - vPC：图形/图像处理场景，提供1/8*Tesla P4、1/4*Tesla P4两种实例

vGN5实例包括的型号和参数规格如下表所示：

型号	GPU(Tesla P4)	GPU显存(GDDR5)	vCPU(核)	内存(DDR4)	数据盘(本地SSD)	网络收发包能力(万PPS)	网络带宽能力(Gbit/s)
vGN5.vPC-2D8	1/8颗	2GB	2	12GB	100GB	10	1
vGN5.vPC-4C4	1/4颗	4GB	4	24GB	200GB	10	1
vGN5.vCS-8B2	1/2颗	8GB	8	48GB	400GB	20	2
vGN5.vCS-16A1	1颗	16GB	16	96GB	800GB	30	3

其中vPC适用于图形图像处理，vCS适用于CUDA计算，如AI推理等。

GPU虚拟化vGN6

该实例的适用场景包括：

- 云游戏的云端实时渲染
- AR/VR的云端实时渲染
- AI（深度学习DL/机器学习ML）

实例特点包括：

- GPU：采用NVIDIA T4 GPU
- 处理器：2.6 GHz主频的Intel® Xeon® Gold 6240 Processor
- 支持系统盘类型：EBS3.0
- 支持数据盘类型：EBS3.0
- vGPU类别
  - vCS：专门用于深度学习，提供1/4*Tesla T4、1/2*Tesla T4两种实例
  - vPC：图形/图像处理场景，提供1/8*Tesla T4、1/2*Tesla T4两种实例

vGN6实例包括的型号和参数规格如下表所示：

型号	GPU(Tesla T4)	GPU显存(GDDR6)	vCPU(核)	内存(DDR4)	网络收发包能力(万PPS)	网络带宽能力(Gbit/s)
vGN6.vPC-2D8	1/8颗	2GB	2	10GB	30	1
vGN6.vCS-4C4	1/4颗	4GB	4	20GB	50	2
vGN6.vCS-10B2	1/2颗	8GB	10	40GB	80	3
vGN6.vPC-10B2	1/2颗	8GB	10	40GB	80	3

其中vPC适用于图形图像处理，vCS适用于CUDA计算，如AI推理等。