

GPU Cloud Computing

Instance Type

Product Documentation



Copyright Notice

©2013-2019 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Instance Type
NVIDIA Instances

Instance Type

NVIDIA Instances

Last updated : 2019-08-06 17:03:33

Instance Overview

NVIDIA Series GPU Instance GN? is not only suitable for GPU general computing scenarios such as deep learning, scientific computing, but also for graphic/image processing (3D rendering, video encoding/decoding) scenarios. Tencent Cloud provides fast, stable and elastic computing services **managed in the same way as with standard CVM.**

Note:

Using the GN? Series instance for 3D graphics rendering (not supported by GN2) requires installing the GRID driver and configuring the license server.

Use Cases

It is suitable for working scenarios where data throughput is large and calculation speed is demanding.

- Deep learning;
- Graphic/image processing;
- Video encoding/decoding;
- Graphic database;
- High-performance database;
- Computational fluid dynamics;
- Computational finance;
- Earthquake analysis;
- Molecular modeling;
- Genomics and others;

Hardware Specification

The basic hardware specifications are as follows

Instance Types	GPU	GPU Mem	vCPU	Mem(DDR4)	Performance	Storage/Net	Region
GN8							
GN8.LARGE56	1 x Tesla P40	24 GB (GDDR5)	6 (Xeon E5-2680v4)	56 GB	12 TF SP · 47T INT8	· Premium CBS · SSD CBS · SSD Local · 10G Network	Beijing 1, Beijing 2, Beijing 4, Shanghai 3, Guangzhou 3, Chengdu 1, Chongqing 1, Silicon Valley 1, Hongkong 1
GN8.3XLARGE112	2 x Tesla P40	48 GB (GDDR5)	14 (Xeon E5-2680v4)	112 GB	24 TF SP · 94T INT8		
GN8.7XLARGE224	4 x Tesla P40	96 GB (GDDR5)	28 (Xeon E5-2680v4)	224 GB	48 TF SP · 188T INT8		
GN8.14XLARGE448	8 x Tesla P40	192 GB (GDDR5)	56 (Xeon E5-2680v4)	448 GB	96 TF SP · 376T INT8		
GN6S							
GN6S.LARGE20	1 x Tesla P4	8 GB (GDDR5)	4 (Xeon Silver 4110)	20 GB	5.5 TF SP · 22T INT8	· Premium CBS · SSD CBS · ESSD CBS · 25G Network	Beijing 4, Shanghai 3, Shanghai 4, Guangzhou 3
GN6S.2XLARGE40	2 x Tesla P4	16 GB (GDDR5)	8 (Xeon Silver 4110)	40 GB	11 TF SP · 44T INT8		
GN10X							
GN10X.2XLARGE40	1 x Tesla V100-NVLINK-32G	32 GB (HBM2)	8 (Xeon Gold 6133)	40 GB	7.8 TF DP · 15.7 TF SP · 125 TF DL	· Premium CBS · SSD CBS · ESSD CBS · 25G Network	Beijing 4, Beijing 5, Shanghai 2, Shanghai 3, Guangzhou 3
GN10X.9XLARGE160	4 x Tesla V100-NVLINK-32G	128 GB (HBM2)	36 (Xeon Gold 6133)	160 GB	31.2 TF DP · 62.8 TF SP · 500 TF DL		
GN10X.18XLARGE320	8 x Tesla V100-NVLINK-32G	256 GB (HBM2)	72 (Xeon Gold 6133)	320 GB	62.4 TF DP · 125.6 TF SP · 1000 TF DL		

Specifications:

- GPU performance: The main indicator is GPU's floating-point computing performance. TF stands for TFlops, SP for single-precision floating-point computing, DP for double-precision floating-point computing, INT8 for INT8 integer computing, and DL for Deep Learning Tensor Core computing (V100 only).
- Storage/network: The storage list shows the storage types supported by the current instance; the network bandwidth refers to the network bandwidth of the physical server where an instance of this type is located. See the purchase page for the network bandwidth assigned by an instance of a certain type.
- Availability zone: Beijing 2 represents Beijing Zone 2, Shanghai 1 represents Shanghai Zone 1, and Guangzhou 3 represents Guangzhou Zone 3, and so on.

Note:

GN2 and GN8 provide SSD-based local storage. When instances are stored locally, their system and data disks only exist within the life cycle of the instance. When these instances expire or are terminated by you, the applications and data in the instance storage will be wiped out. We suggest that you back up or copy the data in the instance storage regularly.

Service Options

- It can be launched in basic network and [VPC](#).
- It can be interfaced with [Cloud Load Balance](#) and other Tencent Cloud products, without additional management and OPS costs. Private network traffic is free of charge.