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 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:

<table>
<thead>
<tr>
<th>Instance Types</th>
<th>GPU</th>
<th>GPU Mem</th>
<th>vCPU</th>
<th>Mem (DDR4)</th>
<th>Performance</th>
<th>Storage/Net</th>
<th>Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>GN8.LARGE56</td>
<td>1 x Tesla P40</td>
<td>24 GB [R00RG]</td>
<td>6 [Bronze 2600v4]</td>
<td>56 GB</td>
<td>12 TF SP · 47 TF INT8</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td>Beijing 1, Beijing 2, Beijing 4, Shanghai 3, Guangzhou 3, Chengdu 1, Chongqing 1, Silicon Valley 1, Hongkong 1</td>
</tr>
<tr>
<td>GN8.3XLARGE112</td>
<td>2 x Tesla P40</td>
<td>48 GB [R00RG]</td>
<td>14 [Bronze 2600v4]</td>
<td>112 GB</td>
<td>24 TF SP · 94 TF INT8</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td></td>
</tr>
<tr>
<td>GN8.7XLARGE224</td>
<td>4 x Tesla P40</td>
<td>96 GB [R00RG]</td>
<td>28 [Bronze 2600v4]</td>
<td>224 GB</td>
<td>48 TF SP · 188 TF INT8</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td></td>
</tr>
<tr>
<td>GN8.14XLARGE448</td>
<td>8 x Tesla P40</td>
<td>192 GB [R00RG]</td>
<td>56 [Bronze 2600v4]</td>
<td>448 GB</td>
<td>96 TF SP · 376 TF INT8</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td></td>
</tr>
<tr>
<td>GN65</td>
<td>1 x Tesla P40</td>
<td>8 GB [R00RG]</td>
<td>4 [Bronze Silver 4110]</td>
<td>20 GB</td>
<td>5.5 TF SP · 22 TF INT8</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td>Beijing 4, Shanghai 3, Shanghai 4, Guangzhou 3</td>
</tr>
<tr>
<td>GN65.2XLARGE40</td>
<td>2 x Tesla P40</td>
<td>16 GB [R00RG]</td>
<td>8 [Bronze Silver 4110]</td>
<td>40 GB</td>
<td>11 TF SP · 44 TF INT8</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td></td>
</tr>
<tr>
<td>GN10X</td>
<td>1 x Tesla V100-NVLink-32G</td>
<td>32 GB [H8M2]</td>
<td>8 [Bronze Gold 6143]</td>
<td>40 GB</td>
<td>7.8 TF DP · 15.7 TF SP · 125 TF DL</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td>Beijing 4, Beijing 5, Shanghai 2, Shanghai 3, Guangzhou 3</td>
</tr>
<tr>
<td>GN10X.9XLARGE160</td>
<td>4 x Tesla V100-NVLink-32G</td>
<td>128 GB [H8M2]</td>
<td>36 [Bronze Gold 6143]</td>
<td>160 GB</td>
<td>31.2 TF DP · 62.8 TF SP · 500 TF DL</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td></td>
</tr>
<tr>
<td>GN10X.18XLARGE320</td>
<td>8 x Tesla V100-NVLink-32G</td>
<td>256 GB [H8M2]</td>
<td>72 [Bronze Gold 6143]</td>
<td>320 GB</td>
<td>62.4 TF DP · 125.6 TF SP · 1000 TF DL</td>
<td>-Premium CBS SSD CBS SSD Local 10G Network-</td>
<td></td>
</tr>
</tbody>
</table>

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.