

GPU Cloud Computing

FAQ

Product Documentation



Tencent Cloud

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

FAQ

General

GPU Usage Shows 100%

FAQ

General

Last updated : 2019-09-19 17:21:59

Q: What is NVIDIA Tesla?

A: NVIDIA Tesla is a new product line introduced by NVIDIA following the launch of professional acceleration card QUADRO and entertainment graphics card GeForce series, which is mainly used for scientific research that requires high performance computing. With NVIDIA® Tesla® GPU accelerator, it can accelerate your most demanding HPC and hyperscale data center workloads.

Q: What is computing acceleration?

A: Computing acceleration uses hardware accelerators or co-processors to perform floating point number calculations and graphics processing, which is more efficient than using software running on CPUs. Tencent Cloud offers three computing acceleration models: GPU computing GN2 and GN8 for generic computing, and GPU rendering GA2 for graphics-intensive applications.

Q: What are the advantages of GPU over CPU?

A: GPU has more arithmetic logic units (ALU) than CPU and supports large-scale multi-threaded parallel computing.

Q: When should I use GPU instances?

A: GPU instances are most suitable for highly parallel applications, such as workloads that use thousands of threads. When massive computations are required for graphics processing where each task is relatively small, a group of operations to be performed form a pipeline. The throughput of this pipeline is more important than the latency of a single operation. To build an application that makes full use of this parallelism, you need to master the expertise of GPU devices, and learn how to program for various graphical APIs (DirectX, OpenGL) or GPU computing programming models (CUDA, OpenCL).

Q: How is GCC billed?

A: Fees are calculated per second and settled per hour, and the resources are released whenever you purchase the service. This method is applicable to scenarios where the demand for devices fluctuates dramatically, such as snap-up campaign on an e-commerce site. For more information, see [Price Overview](#).

Q: Can I upgrade/degrade GPU instance specification?

A: No. Specifications of GPU instances cannot be changed.

Q: What is local SSD?

A: Local SSD is a local storage on the physical device where the CVM resides. It provides instances with block-level data access capability with a low latency, high random IOPS, and high I/O throughput.

Q: Can GPU instances access CVMs?

A: Yes. GPU instances have private IPs and public IPs, so they can communicate with other Tencent Cloud products such as CVMs.

GPU Usage Shows 100%

Last updated : 2018-11-26 18:05:50

Problem Description

When using a GPU computing instance, if you use `nvidia-smi` to view the GPU status in the system, the GPU usage may be displayed as 100% while no processes are using GPU.

```
+-----+-----+
| NVIDIA-SMI 375.51                Driver Version: 375.51           |
+-----+-----+
| GPU  Name          Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+
|   0   Tesla M40  24GB       Off       | 0000:00:06.0     Off    |          0          0 |
| N/A   53C    P0     68W / 250W |  0MiB / 22939MiB |          0%      Default |
+-----+-----+
|   1   Tesla M40  24GB       Off       | 0000:00:07.0     Off    |          0          0 |
| N/A   47C    P0     65W / 250W |  0MiB / 22939MiB |        100%      Default |
+-----+-----+
+-----+-----+
| Processes:                                 GPU Memory Usage |
| GPU       PID  Type  Process name                               |
+-----+-----+
| No running processes found                |
+-----+-----+
```

Possible Cause

This may be caused by the ECC Memory Scrubbing mechanism used when the instance loads the NVIDIA driver.

Solution

Run the `nvidia-smi -pm 1` command in the instance system to get the GPU Driver into the Persistence mode.

Procedure

1. Log in to the GPU computing instance and run the following command:

```
nvidia-smi -pm 1
```

```
[root@UM_18_107_centos data]# nvidia-smi -pm 1
Persistence mode is already Enabled for GPU 00000000:00:03.0.
Persistence mode is already Enabled for GPU 00000000:00:06.0.
All done.
[root@UM_18_107_centos data]# _
```

2. Run the following command to check GPU usage:

```
nvidia-smi
```

Here's a screen indicating that the GPU usage is normal:

```
[root@UM_18_107_centos data]# nvidia-smi
Tue Aug 29 15:31:39 2017
+-----+
| NVIDIA-SMI 384.66                Driver Version: 384.66          |
+-----+-----+
| GPU  Name      Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp      Perf    Pwr:Usage/Cap|      Memory-Usage | GPU-Util  Compute M. |
+-----+-----+-----+-----+-----+-----+
|   0   Tesla P40      On          | 00000000:00:03.0 Off  |             0         |
| N/A   22C      P8      10W / 250W |      0MiB / 22912MiB |      0%    Default  |
+-----+-----+-----+-----+-----+-----+
|   1   Tesla P40      On          | 00000000:00:06.0 Off  |             0         |
| N/A   23C      P8       9W / 250W |      0MiB / 22912MiB |      0%    Default  |
+-----+-----+-----+-----+-----+-----+
+-----+
| Processes:                               GPU Memory |
|  GPU       PID  Type  Process name                               Usage      |
+-----+-----+-----+-----+-----+
| No running processes found               |
+-----+
```