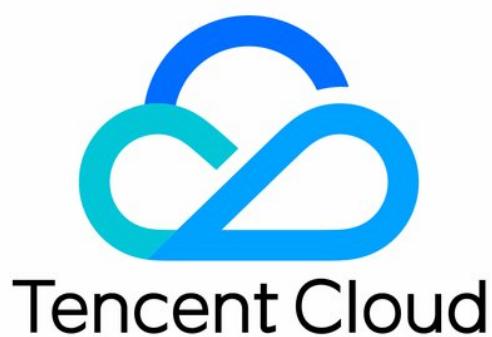


BM Cloud Physical Machine

Instance Types

Product Documentation



Copyright Notice

©2013-2023 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 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 Types

CBM Instance Specification

Instance Types

CBM Instance Specification

Last updated : 2023-01-03 15:25:22

CBM combines the elasticity of a virtual machine and the stable and strong computing performance of a physical machine. It can be integrated seamlessly with all Tencent Cloud products such as networks and databases, making it widely applicable in big data, high-performance computing, cloud gaming, and many more fields. CBM helps you quickly build dedicated and isolated physical machine clusters in the cloud, making it an ideal choice for those seeking for ultimate performance.

When you create a CBM instance, the instance type you specify determines its server hardware configuration. The computing, memory, and storage features vary by instance type. You can select an appropriate instance type based on the scale of your application to be deployed. These instances consist of different combinations of CPU, memory, storage, heterogeneous hardware, and network bandwidth, so you can flexibly select appropriate resources for your applications.

Instance Limits

- The total number of instances that can be started in a region is limited. For more information, see [Purchase Limits](#).
- Limits on system and data disks that can be mounted to an instance: To ensure premium disk I/O performance, Tencent Cloud sets limits on the size and type of data disks purchased with an instance. For more information, see the disks supplied with the corresponding instance family. You can also purchase separate cloud disks if you have higher disk requirements.
- Note that the private network bandwidth capacity of an instance specification is the private network bandwidth cap of the corresponding instance. If the private network traffic exceeds this limit, random packet loss may occur within the private network of your instances.
- The available instance specifications may vary by region. Certain configurations may be sold out, subject to the information on the purchase page.
- The numbers of sent/received packets mentioned in this document are the results of the network forwarding test. For more information on the testing method, see [Network Performance Test](#). Separate testing is needed to estimate the performance for your business.

Instances fall into different types based on the business scenario.

Standard Instance

Standard instances provide a balance of compute, memory, and network resources to accommodate most applications.

Show All

Standard BMS6

展开&收起

Standard BMS6 instances are the latest Intel standard instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

Standard BMS6 instances are applicable to the following use cases:

- Third-party hypervisor application and hybrid cloud deployment
- Enterprise applications of different types and sizes
- Medium-sized and large database systems, caches, and search clusters
- Scenarios where high numbers of massive network packets are sent and received, such as on-screen comments, live streaming, and game servers
- Video encoding/decoding, video rendering, and other applications sensitive to single-core performance

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- **CPU:** 2.7 GHz Intel® Xeon® Ice Lake processor, with a Max Turbo frequency of 3.4 GHz.
- **Memory:** Eight-channel DDR4.
- **Storage:** [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- **Network:** The private network bandwidth of up to 100 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)
BMS6	152	512	2.7/3.4	45 million	32	100	32

Standard BMSA3

展开&收起

Standard BMSA3 instances are the latest AMD standard instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

Standard BMSA3 instances are applicable to the following use cases:

- Third-party hypervisor application and hybrid cloud deployment
- Enterprise applications of different types and sizes
- Medium-sized and large database systems, caches, and search clusters
- Scenarios where high numbers of massive network packets are sent and received, such as on-screen comments, live streaming, and game servers
- Video encoding/decoding, video rendering, and other applications sensitive to single-core performance

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- **CPU:** 2.55 GHz AMD EPYC™ Milan processor, with a Max Boost frequency of 3.5 GHz.
- **Memory:** Eight-channel DDR4.
- **Storage:** [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- **Network:** The private network bandwidth of up to 100 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.

- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)
BMSA3	256	512	2.55/3.5	30 million	32	100	32

Standard BMSA3m

展开&收起

Standard BMSA3m instances are the latest AMD standard instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

Standard BMSA3m instances are applicable to the following use cases:

- Third-party hypervisor application and hybrid cloud deployment
- Enterprise applications of different types and sizes
- Medium-sized and large database systems, caches, and search clusters
- Scenarios where high numbers of massive network packets are sent and received, such as on-screen comments, live streaming, and game servers
- Video encoding/decoding, video rendering, and other applications sensitive to single-core performance

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- **CPU:** 2.55 GHz AMD EPYC™ Milan processor, with a Max Boost frequency of 3.5 GHz.
- **Memory:** Eight-channel DDR4.

- Storage:** [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- Network:** The private network bandwidth of up to 100 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)
BMSA3m	256	1024	2.55/3.5	30 million	32	100	32

Standard BMSA2

展开&收起

Standard BMSA2 instances are the latest AMD standard instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization.

Use cases

Standard BMSA2 instances are applicable to the following use cases:

- Enterprise applications of different types and sizes
- Medium-sized and large database systems, caches, and search clusters
- Scenarios that require sending and receiving massive network packets, such as on-screen video comments, live video broadcasting, and gaming
- Video encoding/decoding, video rendering, and other applications sensitive to single-core performance

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- CPU:** 2.6 GHz AMD EPYCTM ROME processor, with a Max Boost frequency of 3.3 GHz.
- Memory:** Eight-channel DDR4.
- Storage:** [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- Network:** The private network bandwidth of up to 40 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.

- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)
BMSA2	192	512	2.6/3.3	12 million	32	40	32

Standard BMS5

展开&收起

Standard BMS5 instances are the latest Intel standard instances built on the new-gen BM architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization.

Use cases

Standard BMS5 instances are applicable to the following use cases:

- Enterprise applications of different types and sizes
- Small and medium-sized database systems, caches, and search clusters
- Scenarios that require sending and receiving massive network packets, such as on-screen video comments, live video broadcasting, and gaming
- Video encoding/decoding, video rendering, and other applications sensitive to single-core performance

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- **CPU:** 2.6 GHz Intel® Xeon® Cooper Lake processor, with a Max Turbo frequency of 3.2 GHz.
- **Memory:** Six-channel DDR4.
- **Storage:** [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- **Network:** The private network bandwidth of up to 40 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)
BMS5	208	768	2.6/3.2	12 million	32	40	32

Memory Optimized Instance

Memory Optimized instances feature large memory and are suitable for applications that require extensive memory operations, searches, and computing, such as high-performance databases and distributed memory caching.

[Show All](#)

Memory Optimized BMM5c

展开&收起

Memory Optimized BMM5c instances are the latest Intel memory instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization. They guarantee quick and stable performance for workloads processing large data sets in the memory. The processor to memory ratio is 1:14, the ideal choice for large-memory computing applications.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

Memory Optimized BMM5c instances are applicable to the following use cases:

- Applications that require extensive memory operations, searches, and computing, such as high-performance databases and distributed memory caching
- External Hadoop clusters or Redis for fields such as computational genomics
- Scenarios that require sending and receiving massive network packets, such as on-screen video comments, live video broadcasting, and gaming

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- CPU:** 2.6 GHz Intel® Xeon® Cooper Lake processor, with an all-core Max Turbo frequency of 3.2 GHz.
- Memory:** Six-channel DDR4 with stable computing performance.
- Storage:** [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- Network:** The private network bandwidth of up to 40 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)
BMM5c	208	3072	2.6/3.2	12 million	32	40	32

Memory Optimized BMM5

展开&收起

Memory Optimized BMM5 instances are a new generation of Intel memory instances built on the network virtualization technology of Tencent Cloud. They have packet sending/receiving capabilities of up to 10 million PPS over the private network and support up to 25 Gbps of private network bandwidth. They guarantee quick and stable performance for workloads processing large data sets in the memory. The processor to memory ratio is 1:16, the ideal choice for large-memory computing applications.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

Memory Optimized BMM5 instances are applicable to the following use cases:

- Applications that require extensive memory operations, searches, and computing, such as high-performance databases and distributed memory caching

- External Hadoop clusters or Redis for fields such as computational genomics

Hardware specification

- CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.
- Memory:** Six-channel DDR4.
- Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks and 2 * 3,840 GB NVMe SSD high-performance local storage.
- Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Local Storage
BMM5	96	1536	2.5/3.1	10 million	16	25	2 * 480 GB SATA SSD (RAID1) and 2 * 3,840 GB NVMe SSD

Memory Optimized BMM5r

展开&收起

Memory Optimized BMM5r instances are Intel memory instances built on the network virtualization technology of Tencent Cloud. They have packet sending/receiving capabilities of up to 10 million PPS over the private network and support up to 25 Gbps of private network bandwidth. They guarantee quick and stable performance for workloads processing large data sets in the memory. The processor to memory ratio is 1:8, the ideal choice for large-memory computing applications.

Use cases

Memory Optimized BMM5r instances are applicable to the following use cases:

- Applications that require extensive memory operations, searches, and computing, such as high-performance databases and distributed memory caching
- External Hadoop clusters or Redis for fields such as computational genomics

Hardware specification

- CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.

- **Memory:** Six-channel DDR4.
- **Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks and 2 * 3,840 GB NVMe SSD high-performance local storage.
- **Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Local Storage
BMM5r	96	768	2.5/3.1	10 million	16	25	2 * 480 GB SATA SSD (RAID1) and 2 * 3,840 GB NVMe SSD

High I/O Instance

High I/O instances feature high random IOPS, high throughput and low latency and are well-suited for I/O-intensive applications that require high disk read/write performance and low latency, such as high-performance databases.

Show All

High I/O BMIA2

展开&收起

High I/O BMIA2 instances are the latest AMD high I/O instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization. Based on NVMe SSD storage, they come with storage resources featuring a low latency, an ultra high IOPS, and a high throughput, making them suitable for I/O-intensive applications such as high-performance relational databases and Elasticsearch.

Use cases

High I/O BMIA2 instances are applicable to the following use cases:

- High-performance databases, NoSQL databases (e.g. MongoDB), and clustered databases
- I/O-intensive applications that require a low latency, such as Elasticsearch
- Big data applications with storage-computing separation

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- CPU:** 2.6 GHz AMD EPYCTM ROME processor, with a Max Boost frequency of 3.3 GHz.
- Memory:** Eight-channel DDR4 with stable computing performance.
- Storage:** 4 * 3,840 GB NVMe SSD high-performance local storage. [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- Network:** The private network bandwidth of up to 40 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)	Loca Stora
BMIA2	192	512	2.6/3.3	12 million	32	40	32	4 * 3,840 GB NVMe SSD

High I/O BMIA2m

展开&收起

High I/O BMIA2m instances are the latest AMD high I/O instances built on the new-gen BM architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization. Based on NVMe SSD storage, they come with storage resources featuring a low latency, an ultra high IOPS, and a high throughput, making them suitable for I/O-intensive applications such as high-performance relational databases and Elasticsearch.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

High I/O BMIA2m instances are applicable to the following use cases:

- High-performance databases, NoSQL databases (e.g. MongoDB), and clustered databases
- I/O-intensive applications that require a low latency, such as Elasticsearch
- Big data applications with storage-computing separation

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- **CPU:** 2.6 GHz AMD EPYCTM ROME processor, with a Max Boost frequency of 3.3 GHz.
- **Memory:** Eight-channel DDR4 with stable computing performance.
- **Storage:** 4 * 3,840 GB NVMe SSD high-performance local storage. [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- **Network:** The private network bandwidth of up to 40 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Supported ENIs (Including Primary ENIs)	Loca Stora
BMIA2m	192	1024	2.6/3.3	12 million	32	40	32	4 * 3,840 GB NVMe SSD

High I/O BMI5

展开&收起

High I/O BMI5 instances, based on NVMe SSD storage, are designed for I/O-intensive workloads. They come with storage resources featuring a low latency, an ultra high IOPS, and a high throughput, making them suitable for I/O-intensive applications such as high-performance relational databases and Elasticsearch.

Use cases

High I/O BMI5 instances are applicable to the following use cases:

- High-performance databases, NoSQL databases (e.g. MongoDB), and clustered databases

- I/O-intensive applications that require a low latency, such as Elasticsearch
- Big data applications with storage-computing separation

Hardware specification

- **CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.
- **Memory:** Six-channel DDR4.
- **Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks and 2 * 3,840 GB NVMe SSD high-performance local storage.
- **Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Local Storage
BMI5	96	384	2.5/3.1	10 million	16	25	2 * 480 GB SATA SSD (RAID1) and 2 * 3,840 GB NVMe SSD

Big Data Instance

The big data family is equipped with massive storage resources, features high throughput, and is suitable for throughput-intensive applications such as Hadoop distributed computing, massive log processing, distributed file systems, and large data warehouses.

Show All

Big Data BMDA2

展开&收起

Big Data BMDA2 instances are the latest AMD instances built on the new-gen elastic bare metal architecture. They provide superior computing, network, and storage performance. They have no virtualization loss in terms of computing performance and support nested virtualization. Equipped with a high throughput and massive local storage resources, they are suitable for throughput-intensive applications such as Hadoop distributed computing and parallel data processing.

Use cases

- Distributed computing services such as Hadoop MapReduce, HDFS, Hive, and HBase
- Workloads such as Elasticsearch, log processing, and large data warehouse
- Customers in the Internet, finance, and industries that require big data computing and storage analysis, as well as workloads that require massive data storage and computing

Hardware specification

- High-performance computing services featuring high reliability, security, and stability are provided based on the Star Lake servers developed by Tencent Cloud.
- **CPU:** 2.6 GHz AMD EPYCTM ROME processor, with a Max Boost frequency of 3.3 GHz.
- **Memory:** Eight-channel DDR4 with stable computing performance.
- **Storage:** 12 * 16,000 GB SATA HDD and 1 * 3,840 GB NVMe SSD massive local storage. [Cloud disks](#) can be used as system and data disks and [expanded](#) on demand.
- **Network:** The private network bandwidth of up to 40 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed. ENI mounting is supported.
- We recommend you use the instance together with the TencentOS Server operating system to enjoy the best application performance.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Local Storage
BMDA2	192	512	2.6/3.3	12 million	32	40	12 * 16,000 GB SATA HDD and 1 * 3,840 GB NVMe SSD

Big Data BMD3

展开&收起

Big Data BMD3 instances are the latest big data instances built on the network virtualization technology of Tencent Cloud. They have packet sending/receiving capabilities of up to 10 million PPS over the private network and support up to 25 Gbps of private network bandwidth. Equipped with a high throughput and massive storage resources, they are suitable for throughput-intensive applications such as Hadoop distributed computing and parallel data processing.

Use cases

- Distributed computing services such as Hadoop MapReduce, HDFS, Hive, and HBase
- Workloads such as Elasticsearch, log processing, and large data warehouse

- Customers in the Internet, finance, and industries that require big data computing and storage analysis, as well as workloads that require massive data storage and computing

Hardware specification

- CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.
- Memory:** Six-channel DDR4 with stable computing performance.
- Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks and 12 * 12,000 GB SATA HDD massive local storage.
- Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Local Storage
BMD3	96	384	2.5/3.1	10 million	16	25	2 * 480 GB SATA SSD (RAID1) and 12 * 12,000 GB SATA HDD

GPU Compute Instance

GPU Compute instances are equipped with heterogeneous GPU to deliver real-time, fast parallel computing and floating point computing capabilities. They are suitable for high-performance applications such as deep learning, scientific computing, video encoding/decoding, and graphics workstations.

Note

- To use NVIDIA GPU instances for general computing tasks, you need to install the Tesla driver and CUDA toolkit. For more information, see [Installing NVIDIA Driver](#) and [Installing CUDA Driver](#).
- To use NVIDIA GPU instances for 3D rendering tasks such as high-performance graphics processing and video encoding/decoding, you need to install a GRID driver and configure a license server.

Show All

GPU Compute BMGNV4

展开&收起

GPU Compute BMGNV4 instances are equipped with the latest NVIDIA® Tesla® A10 GPU. They support not only general GPU computing tasks such as deep learning and scientific computing, but also graphics and image processing tasks such as 3D rendering and video encoding/decoding. They have no computing performance loss, support nested virtualization, and provide fast, stable, and elastic computing services.

Use cases

GPU Compute BMGNV4 instances are applicable to the following use cases:

- Graphic and image processing
- Video encoding and decoding
- Graph database

They are also applicable to deep learning inference and small-scale training scenarios, such as:

- AI inference for mass deployment
- Small-scale deep learning training

Hardware specification

- The high-density accelerator card is provided based on the Star Lake GPU servers developed by Tencent Cloud, which is extremely cost-effective.
- **CPU:** 3.4 GHz Intel® Xeon® Cooper Lake high-clock rate processor, with a Max Turbo frequency of 3.8 GHz.
- **GPU:** 16 * NVIDIA® Tesla® A10 GPU (31.2 TFLOPS of single-precision floating point performance, 250 TOPS for INT8, and 500 TOPS for INT4).
- **Memory:** Six-channel DDR4 with stable computing performance.
- **Storage:** 2 * 480 GB SATA SSD and 4 * 3,840 GB NVMe SSD high-performance local storage.
- **Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	GPU	GPU Video Memory	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	GPU	GPU Video Memory	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)
BMGNV4	208	768	3.4/3.8	16 * NVIDIA A10	16 * 24 GB	10 million	16	25

GPU Compute BMG5t

展开&收起

GPU Compute BMG5t instances are equipped with NVIDIA® Tesla® T4 GPU. They support not only general GPU computing tasks such as deep learning and scientific computing, but also graphics and image processing tasks such as 3D rendering and video encoding/decoding. They have no computing performance loss, support nested virtualization, and provide fast, stable, and elastic computing services.

Use cases

GPU Compute BMG5t instances are applicable to the following use cases:

- Graphic and image processing
- Video encoding and decoding
- Graph database

They are also applicable to deep learning inference and small-scale training scenarios, such as:

- AI inference for mass deployment
- Small-scale deep learning training

Hardware specification

- **CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.
- **GPU:** 4 * NVIDIA® Tesla® T4 GPU (8.1 TFLOPS of single-precision floating point performance, 130 TOPS for INT8, and 260 TOPS for INT4).
- **Memory:** Six-channel DDR4.

- Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks.
- Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	GPU	GPU Video Memory	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)
BMG5t	96	384	2.5/3.1	4 * NVIDIA T4	4 * 16 GB	10 million	16	25

GPU Compute BMG5v

展开&收起

GPU Compute BMG5v instances are equipped with NVIDIA®Tesla® V100 NVLink® 32 GB. They support not only general GPU computing tasks such as deep learning and scientific computing, but also graphics and image processing tasks such as 3D rendering and video encoding/decoding. They have no computing performance loss, support nested virtualization, and provide fast, stable, and elastic computing services.

Note

This instance type is currently made available through an allowlist. To purchase it, contact your channel manager for application.

Use cases

GPU Compute BMG5v instances are applicable to large-scale deep learning training and inference as well as scientific computing scenarios, such as:

- Deep learning
- Computational fluid dynamics
- Molecular modeling
- Genomics and others

They are also applicable to graphics and image processing scenarios, such as:

- Graphic and image processing
- Video encoding and decoding
- Graph database

Hardware specification

- **CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.
- **GPU:** 8 * NVIDIA® Tesla® V100 GPU (15.7 TFLOPS of single-precision floating point performance, 7.8 TFLOPS of double-precision floating point performance, 125 TFLOPS of deep learning accelerator performance with Tensor cores, and 300 GB/s NVLink®).
- **Memory:** Six-channel DDR4.
- **Storage:** 1 * 480 GB SATA SSD local system disk and 4 * 3,200 GB NVMe SSD high-performance local storage.
- **Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Clock Rate/Max Clock (GHz)	GPU	GPU Video Memory	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)
BMG5v	96	384	2.5/3.1	8 * NVIDIA V100	8 * 32 GB	10 million	16	25

Other Available Instances

Note

If the following instances are sold out, we recommend you select new-gen ones in the same family.

Show All

Standard BMS4

展开&收起

Standard BMS4 instances are equipped with the latest Xeon® Skylake processors and DDR4 memory. They use Tencent Cloud's proprietary network virtualization technology, offering packet sending/receiving capabilities of up to 10 million PPS over the private network and up to 25 Gbps of private network bandwidth. They have no computing performance loss and support nested virtualization.

Use cases

Standard BMS4 instances are applicable to the following use cases:

- Enterprise applications of different types and sizes
- Small and medium-sized database systems, caches, and search clusters
- Scenarios that require sending and receiving massive network packets, such as on-screen video comments, live video broadcasting, and gaming

Hardware specification

- **CPU:** 2.4 GHz Intel® Xeon® Skylake 6148 processor, with a Max Turbo frequency of 3.0 GHz.
- **Memory:** Six-channel DDR4.
- **Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks and 10 * 480 GB SATA SSD (RAID50) local data disks.
- **Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Clock Rate	Local Storage
BMS4	80	384	10 million	16	25	2.4 GHz	2 * 480 GB SATA SSD (RAID1) and 10 * 480 GB SATA SSD (RAID50)

Big Data BMD3s

展开&收起

Big Data BMD3s instances are equipped with a high throughput and massive storage resources and built on the network virtualization technology of Tencent Cloud. They have packet sending/receiving capabilities of up to 10 million

PPS over the private network and support up to 25 Gbps of private network bandwidth. They are suitable for throughput-intensive applications such as Hadoop distributed computing and parallel data processing.

Use cases

- Distributed computing services such as Hadoop MapReduce, HDFS, Hive, and HBase
- Workloads such as Elasticsearch, log processing, and large data warehouse
- Customers in the Internet, finance, and industries that require big data computing and storage analysis, as well as workloads that require massive data storage and computing

Hardware specification

- **CPU:** 2.5 GHz Intel® Xeon® Cascade Lake processor, with a Max Turbo frequency of 3.1 GHz.
- **Memory:** Six-channel DDR4.
- **Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks, 12 * 12,000 GB SATA HDD, and 1 * 3,840 GB NVMe SSD massive local storage.
- **Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Clock Rate	Local Storage
BMD3s	96	192	10 million	16	25	2.5 GHz	2 * 480 GB (RAID1) SATA SSD, 12 * 12,000 GB SATA HDD, and 1 * 3,840 GB NVMe SSD

Big Data BMD2

展开&收起

Big Data BMD2 instances are equipped with a high throughput and massive storage resources and built on the network virtualization technology of Tencent Cloud. They have packet sending/receiving capabilities of up to 10 million PPS over the private network and support up to 25 Gbps of private network bandwidth. They are suitable for throughput-intensive applications such as Hadoop distributed computing and parallel data processing.

Use cases

- Distributed computing services such as Hadoop MapReduce, HDFS, Hive, and HBase
- Workloads such as Elasticsearch, log processing, and large data warehouse

- Customers in the Internet, finance, and industries that require big data computing and storage analysis, as well as workloads that require massive data storage and computing

Hardware specification

- CPU:** 2.4 GHz Intel® Xeon® Skylake 6148 processor, with a Max Turbo frequency of 3.0 GHz.
- Memory:** Six-channel DDR4.
- Storage:** 2 * 480 GB SATA SSD (RAID1) local system disks and 12 * 12,000 GB SATA HDD massive local storage.
- Network:** The private network bandwidth of up to 25 Gbps is supported, with strong packet sending/receiving capabilities. You can purchase the [public network bandwidth](#) as needed.

Specification	vCPU	Memory (GiB)	Packets In/Out (PPS)	Number of Queues	Private Network Bandwidth Capacity (Gbps)	Clock Rate	Local Storage
BMD2	80	384	10 million	16	25	2.4 GHz	2 * 480 GB (RAID1) SATA SSD and 12 * 12,000 GB SATA HDD