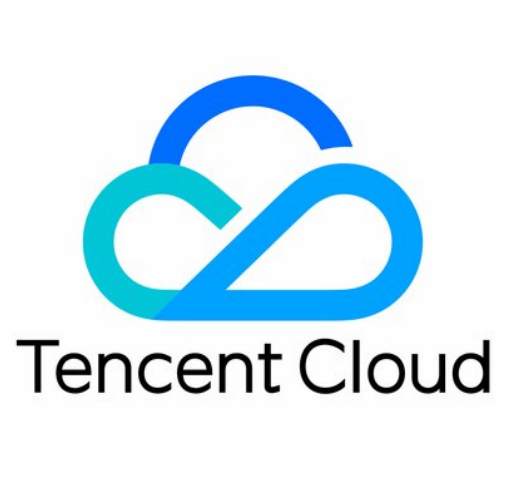
Hyper Computing Cluster

Product Introduction

Product Documentation



Contents

Product Introduction
Overview

Last updated：2024-08-20 16:50:21

Introduction

Hyper Computing Cluster uses high-performance CVM instances as nodes, and instances are interconnected via Remote Direct Memory Access (RDMA). It offers high bandwidth and ultra-low latency network services, significantly improving network performance. It can meet the parallel computing needs of large-scale high-performance computing, AI, big data recommendations, and other applications.

Basic Concepts

Before using Hyper Computing Cluster, you need to understand the following concepts:

Concept	Description
Instance	Computing resources on the public cloud, including basic computing resources such as CPU, memory, operating system, network, and disk.
Instance specification	Configurations of CPU, memory, storage, and network for high-performance CVM instances provided by Tencent Cloud.
Image	Images of high-performance CVM instances, including preset images on Windows, Linux, and other systems.
Local disk	A device on the physical server that can be used for persistent storage by an instance.
CBS disk	A distributed and persistent block storage device that can serve as the system disk or an expandable data disk of an instance.
VPC	A logically isolated virtual network space on Tencent Cloud.
IP address	Tencent Cloud offers private IP addresses and public IP addresses. Simply put, a private IP address provides LAN services for mutual access between high-performance CVM# instances. A public IP address is used when high-performance CVM instances need to access the Internet.
Elastic IP address	Static public IP addresses designed especially for dynamic networks to meet the demands for fast troubleshooting.

Security group	A security group can be understood as a virtual firewall featuring status check and data packet filtering features. It is used for network access control of one or more high-performance CVM instances. Security group is an important measure for network security isolation.
Login method	High-security SSH key pairs and passwords can be used for login.
Region and AZ	The location of instances and other resources. High-performance CVM instances in the same region but different AZs within the same VPC can communicate with each other using private IP addresses.
Tencent Cloud console	Web-based UI.

Using Hyper Computing Cluster

Tencent Cloud allows you to configure and manage Hyper Computing Cluster instances in the following ways:
Console: You can use the web-based console of CVM to configure and manage Hyper Computing Cluster instances.
API: Tencent Cloud also provides APIs to conveniently manage Hyper Computing Cluster instances. For API details, see API Overview.
TCCLI to call APIs.' style="color:#000000;font-size:12px;">**SDK**: You can use SDKs or Tencent Cloud's TCCLI to call APIs.
Note
If you have not used Hyper Computing Cluster before, see Managing Hyper Computing Cluster.

Instance Type

For detailed specifications of Hyper Computing Cluster instances, see Instance Specifications.

Billing Overview

For billing of Hyper Computing Cluster instances, see Billing Overview.

Strengths

Last updated：2024-08-06 14:54:32

Strengths of Hyper Computing Cluster

Super Computing Power

Achieve computing without virtualization loss and maintain server performance based on an elastic high-performance CVM architecture.
Support fully automated production and auto scaling, featuring flexibility and extreme performance.

High-Performance Storage

Support flexible storage capacity expansion with COS (COS) or CFS (CFS) and ensure strong consistency among three replicas.
Offer ultra-fast IO instances for selection in heavy IO load and local caching scenarios when used jointly with local NVMe SSDs.

High-Speed Network Plane

Support VPC and non-blocking RoCEv2 RDMA network for inter-node connection with transmission latency as low as 2 us, widely supporting Ethernet-based applications.
Use intelligent 25G network cards and a VPC network with a PPS of 10 million to help users build cloud-native high-performance computing applications with file storage and other services.

Comparison with Local Supercomputing Clusters

Hyper Computing Cluster has the following advantages over local supercomputing clusters:

Dimension	Feature	Hyper Computing Cluster	Local Supercomputing Cluster
Hardware resource	Computing resource supply	Flexible	Inflexible
	Auto scaling of computing resources	Flexible	Inflexible
	Hardware iteration and upgrade	Flexible	Inflexible
	Storage resource	Flexible	Inflexible

	switching		
	Virtualization	Flexibly supported	Partially supported
	Hardware resource maintenance	Simple	Complex
Software resource	Operating system type	Flexible	Inflexible
	Operating environment configuration	Simple	Complex
Network	VPC	Supported	Not supported
	RDMA network	Supported	Supported
Cost	Maintenance cost	Low	High
	Purchase cost	Pay-as-you-go	One-time large investment
Availability	Disaster recovery	High	Partially supported
	Security	High	Partially supported

Use Cases

Last updated：2024-08-20 16:52:24

Large-Scale AI Training

Scenarios such as autonomous driving, NLP, and recommendation systems are characterized by large data volumes and intensive computing. Hyper Computing Cluster instances can support high-speed, low-latency RDMA network interconnection and the next-generation CPU architecture and heterogeneous GPU components for the computing of compute-intensive workloads. In this way, it can meet business needs of high computing performance, stability, and timeliness.

Industrial Simulation

Many manufacturing enterprises in automobile, aviation, and other industries need to use simulated computing to drive design. The high-performance computing clusters built by enterprises themselves require large investments and long cycles, and it is difficult to continuously meet the demand. Hyper Computing Cluster instances can be quickly deployed and elastically scaled. Through the high-speed, low-latency RDMA network and the latest CPU architecture, Hyper Computing Cluster instances implement parallel processing to quickly meet the simulation demands of enterprises in aerospace, industrial manufacturing, and other industries and promote product R&D promptly.

Life Sciences

Hyper Computing Cluster instances can use the high-speed, low-latency RDMA network to conduct large-scale molecular dynamics simulations, predict and analyze the interactions between and changes of biological protein molecules and lipid molecules, and assist in drug research.

Scientific Research and Education

Hyper Computing Cluster instances can provide supercomputing platforms for universities and research institutions for numerical simulation, numerical computation, simulation verification, and other applications.