02 Introduction to Cloud Virtual Machine (CVM)

Course Objectives

- At the end of this course, you will have a better understanding of:
  - The basics of Cloud Virtual Machine (CVM)
  - The features, advantages, and use cases of CVM
  - The billing plans for CVM

Contents

Chapter I CVM Basics

Chapter II CVM Products

Chapter III Billing Plans for CVM

Chapter I CVM Basics

1.1 Regions and Availability Zones

1.2 CVM Instances

1.3 CVM Images

1.4 CVM Storage

1.5 CVM Evolution

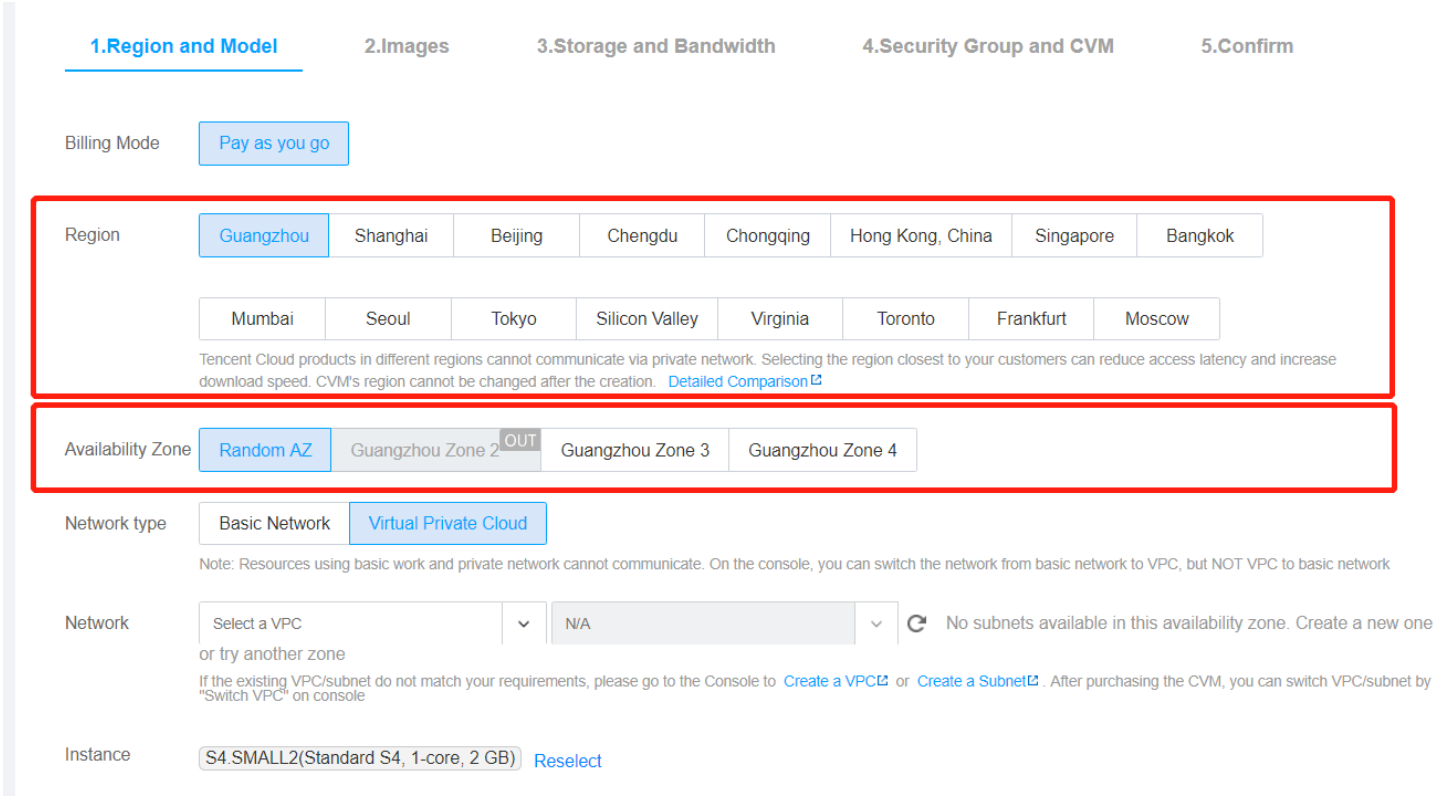
1.1 Regions and Availability Zones

- **Regions:** These are independent geographical areas. Naming convention: [coverage area + name of the city where the data center is located].

- **Availability zones**: These are physical data centers with independent electrical and network systems in the same Region. Naming convention: [city + number].

- Tencent Cloud currently has **25** regions and **53** availability zones worldwide.

CVM purchase step 1: Select a region and availability zone



1.1 Deployment of Multiple Regions and Availability Zones

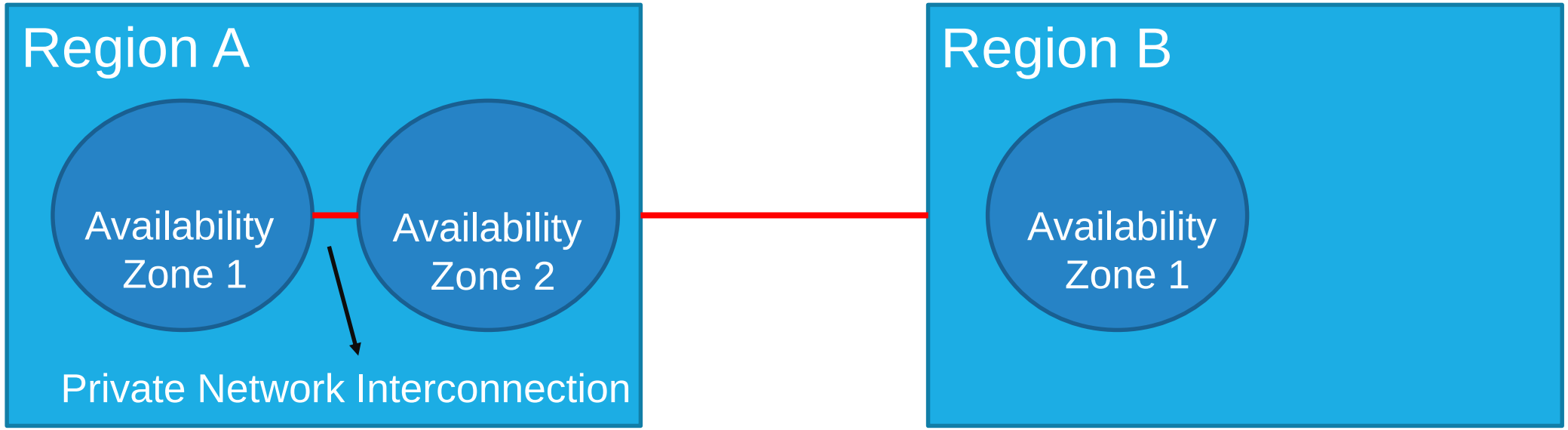
Nearby access

Fault Isolation

Load Balancing

High availability

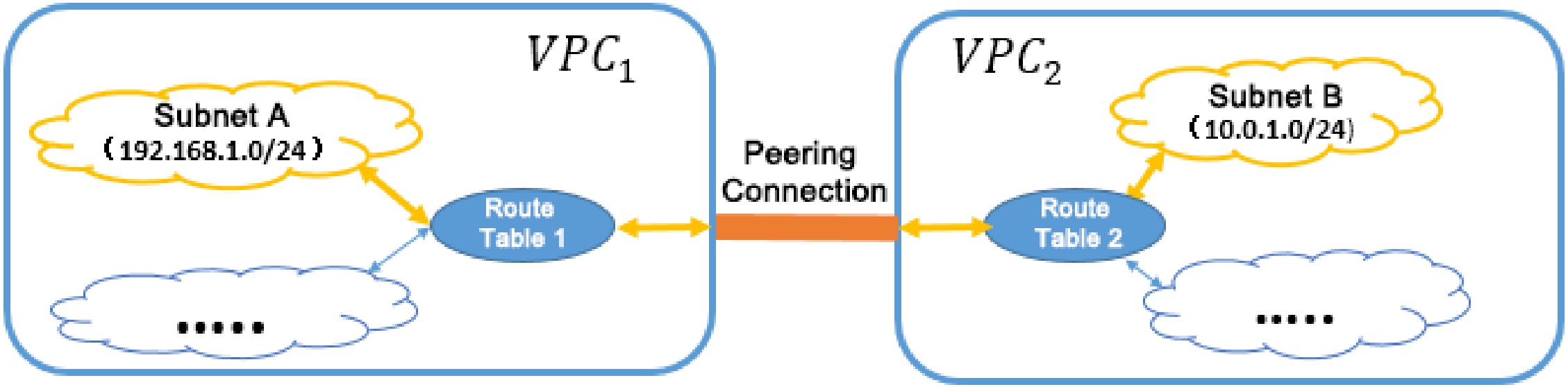
1.1.1 Cross-Region Access



- Cloud services in private networks can communicate via the high-speed connections provided by Tencent Cloud.
- CLB does not support traffic forwarding across regions.
- The private network interconnection illustrated above applies to the sharing of resources under the same account.

1.1.1 Cross-Region Access (continued)

- VPC peering enables data synchronization between VPC networks. It routes traffic between different private IP addresses in peer VPC networks as if they were in the same network.



1.2 CVM Instances

- **Instances:** A CVM instance includes components such as CPU, memory, disk, network, and operating system.
- **Naming convention:** Series.Model (Specification, Number of vCPUs, Memory size)

  S4.SMALL2(Standard S4, 1-core, 2 GB)

- **Management:** Instances are managed by the console or APIs. You can adjust the instance configuration online and restart the instance to apply the changes.
- **Security:** CVMs support security policies, security groups, and login control.

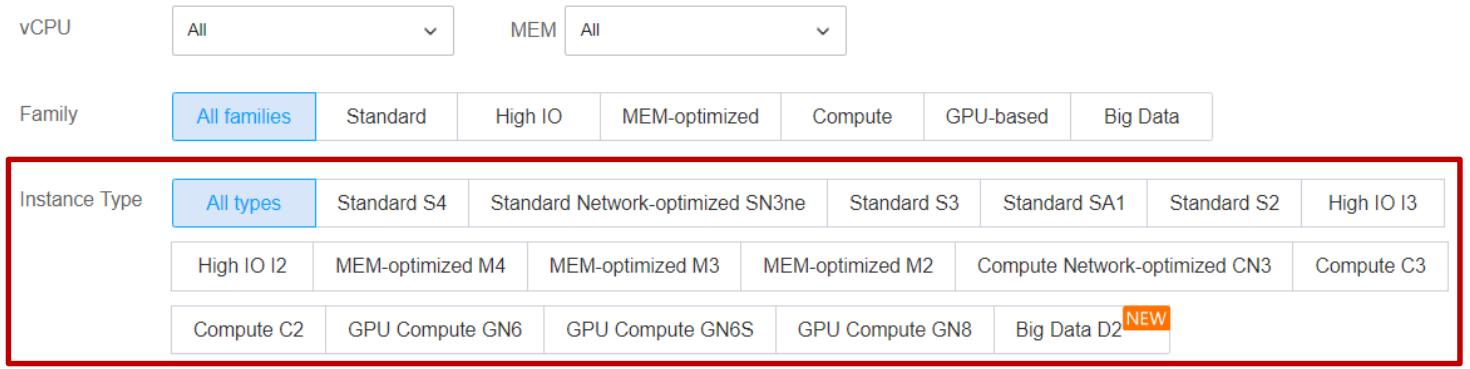
1.2 CVM Instance Models

- Instances of different models have different **hardware and configurations** (CPU, memory, bandwidth, and maximum configurations).

	Standard S1	Standard S2	Standard S3	Standard S4
CPU	Intel Xeon CPU	Xeon Broadwell (v4)	Skylake 6133	Skylake 6148
Memory	DDR3	DDR4	Latest DDR4	Latest DDR4
Private network bandwidth	10 Gbps	10 Gbps	10 Gbps	25 Gbps
Maximum configuration	48-core, 96 GB	56-core, 224 GB	80-core, 320 GB	72-core, 228 GB

1.2 CVM Instance Families and Types

- An instance family is a collection instance types with similar purposes and configurations. An instance type is a collection of instance configuration options. These options include the model, specifications, number of vCPUs, memory size, CPU model, private network bandwidth, throughput, and supported availability zones.



CVM Instance Types

Selected: S4.SMALL2 (Standard S4, 1-core 2 GB)

Only display Guangzhou Zone 4 Supported models

Model	Specifications	vCPU	MEM	CPU model (clock rate)	Private network bandwidth	Send/Receive packets	Supported availability zones	Note	Fee
Standard S4	S4.SMALL2	1-core	2GB	Intel Xeon Skylake 6148(2.4 GHz)	1.5Gbps	25 (in 10k) PPS	11	N/A	0.05USD/hr
Standard S4	S4.MEDIUM4	2-core	4GB	Intel Xeon Skylake 6148(2.4 GHz)	1.5Gbps	30 (in 10k) PPS	11	N/A	0.1USD/hr
Standard S4	S4.MEDIUM8	2-core	8GB	Intel Xeon Skylake 6148(2.4 GHz)	1.5Gbps	30 (in 10k) PPS	11	N/A	0.15USD/hr
Standard S4	S4.LARGE8	4-core	8GB	Intel Xeon Skylake 6148(2.4 GHz)	1.5Gbps	50 (in 10k) PPS	11	N/A	0.2USD/hr

1.2 Features and Use Cases of Different Instance Types

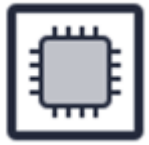
腾讯云

Type	Features	Use Cases
Standard	Balanced computing power, memory, and network capabilities	Small and medium-sized web applications and databases, and official company websites
Memory	Large memory size	Memory-intensive operations, such as searching, computing, and distributed in-memory caching
High I/O	High I/O, high throughput, and low latency	NoSQL databases, clustered databases, and OLTP systems
Computational	CPU with a high clock speed and max price/performance ratio	High-traffic web servers, MMO game servers, and HPC

1.2 How to Choose a Suitable CVM

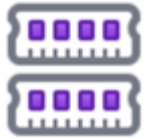
Standard

Small- and medium-sized web applications and databases



Computational

Large game servers, ad service engines, and high-performance computing



Memory

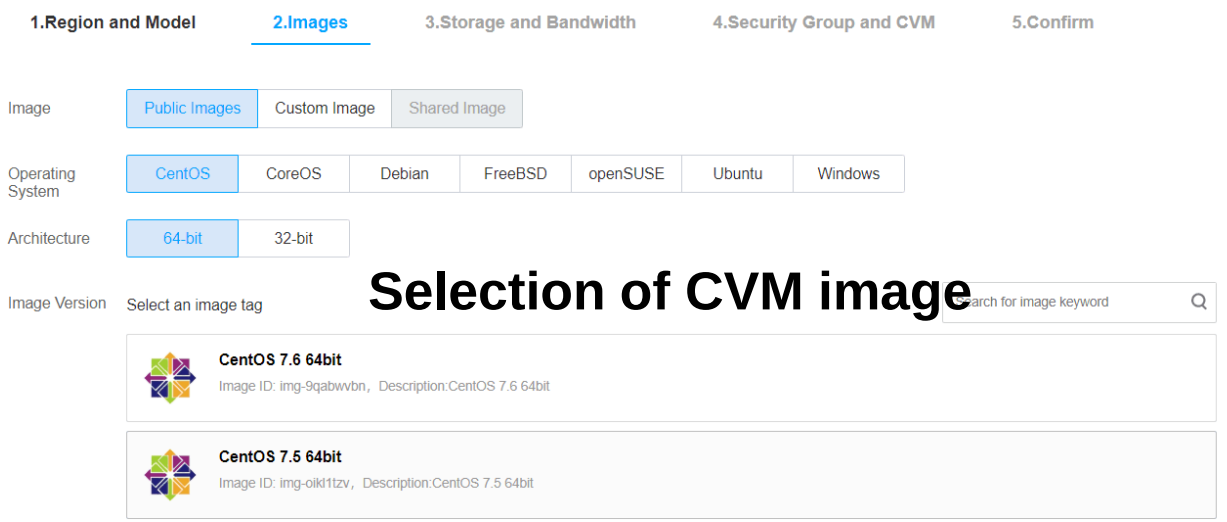
Memory-intensive operations, such as searching and computing

High I/O

Low-latency and I/O-intensive applications

1.3 CVM Images

- **Definition:**
  - Images are templates (installation disks) for creating CVMs. They contain operating systems and relevant configuration information.
- **Functions:**
  - Batch deployment
  - Specific software deployment
  - Runtime environment backup



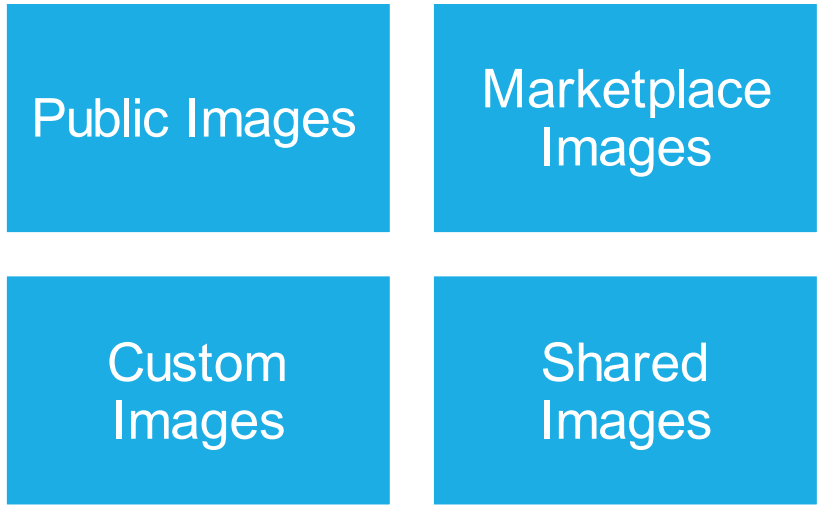
Selection of CVM image

1.3 CVM Images (continued)

- Image deployment vs. Manual installation:

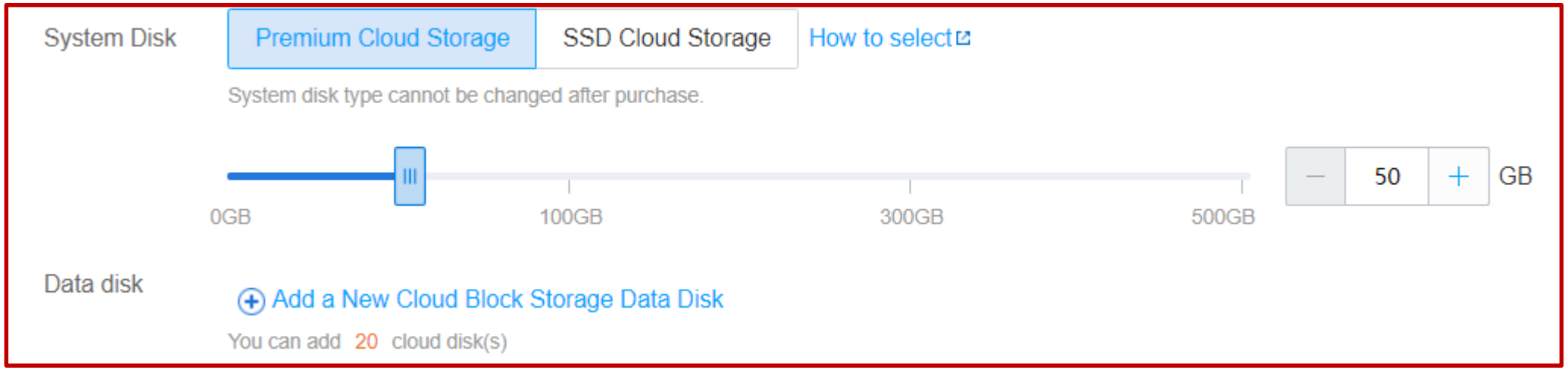
	Image Deployment	Manual Installation
Deployment time	3-5 minutes	1-2 days
Deployment process	Select an image and deploy it with one click	Manually select and install the software and perform debugging
Security	Tested and approved by Tencent Cloud	Relies on the expertise of development and deployment staff
Source	Public images, marketplace images, custom images, and shared images	Manual deployment from scratch

- Types of CVM images:



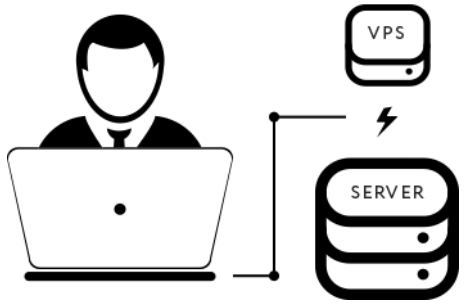
1.4 CVM Storage

- By storage media: HDD Cloud Storage, Premium Cloud Storage, and SSD Cloud Storage
- By partition type: System disk and data disk
- By architecture: Local disk, Cloud Block Storage (CBS), and Cloud Object Storage (COS)



CVM storage settings

1.5 CVM Evolution



Virtual Server

A portion of the server disk space is partitioned for the user.

Dedicated Server

The customer rents a dedicated server to host a website.

VPS

A physical server hosts several virtual servers.

CVM

Virtual servers are hosted on clusters.

Chapter II CVM Products

2.1 What Is CVM

2.2 GPU Cloud Computing

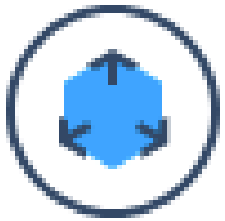
2.3 FPGA Cloud Computing

2.4 CVM Dedicated Hosts

2.5 Cloud Physical Machine (CPM)

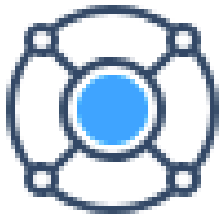
2.1 What Is CVM

- Cloud Virtual Machine (CVM) is an elastically scalable computing service.



Elastic Computing

Scaling
in minutes

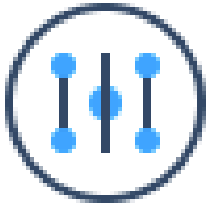


Flexible Configuration

CPU, memory, and bandwidth

Stability and reliability

99.95% availability
and three replicas



Simple management

APIs, CLI, or
console

Secure network

VPC, ACL, security groups

Comprehensive protection

Trojan and vulnerability detection and
protection against brute force attacks

2.1 CVM: E-commerce Promotion Service

- **Challenges:**
  - Traffic peaks
  - Service stability
  - Cost control

Configuration Example

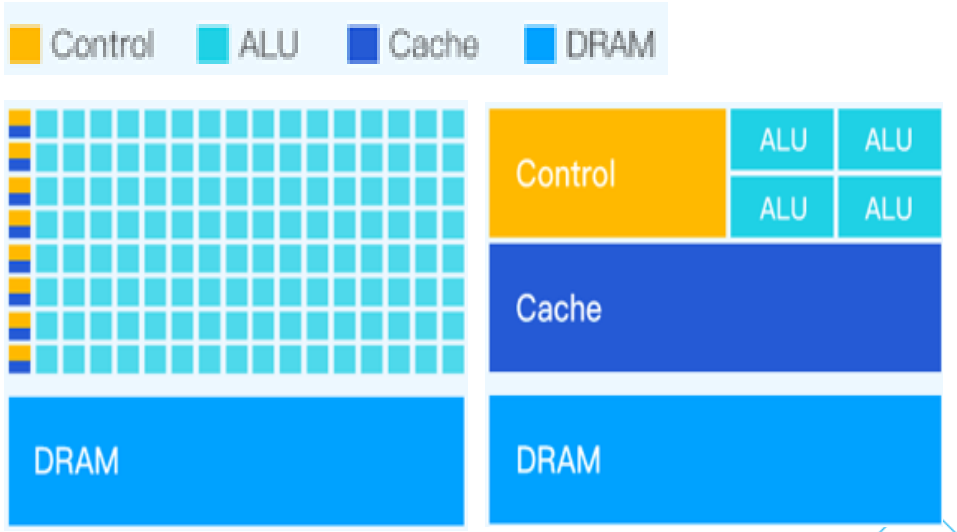
Product	Quantity	Specifications
CVM fixed service group	20 instances	Standard S2, 2-core, 4 GB
CVM auto scaling group	Dynamically adjusted	Standard S2, 1-core, 1 GB

- **Solution:**
  - Standard CVM instances and image deployment
  - Auto scaling: reduces the number of instances during slack periods and increases the number during peak hours

2.2 GPU Cloud Computing

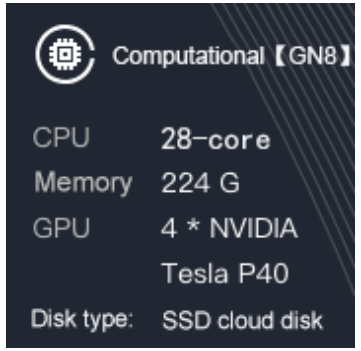
- This is a GPU-based computing service for real-time high-speed parallel computing and floating-point computing.
- It breaks through the limitations of traditional CPUs and features high performance parallel computing, high throughput, and low latency. Its scientific computing power is 50 times greater than that of the traditional architecture.

GPU	CPU
Thousands of accelerator cores	Dozens of cores
Highly efficient ALUs, multi-threading, and high parallel throughput	Complex logic control unit and powerful ALU
Compute-intensive and parallel computing	Logic control and serial arithmetic



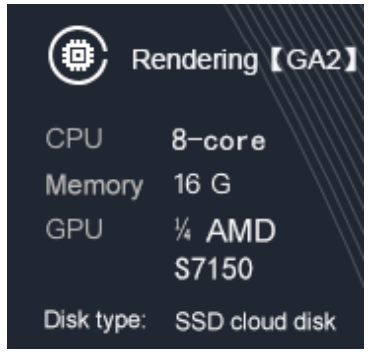
GPU CPU

2.2 GPU Cloud Computing



Computational Instance

- **Deep learning:** Deep learning training and inference, image recognition, and speech recognition
- **Scientific computing:** Computational finance, seismic analysis, molecular modeling, and genomics
- **Video coding:** HD video transcoding, security video surveillance, and large-scale video conference.



Rendering Instance

- **Graphic workstations:** 3D design and rendering, audio, video, and animation production, engineering modeling and simulation (CAD/CAE), medical imaging, and game testing
- **Nonlinear editing (NLE):** Movie editing and TV post-production, including editing, subtitling, special effects, color correction, and packaging

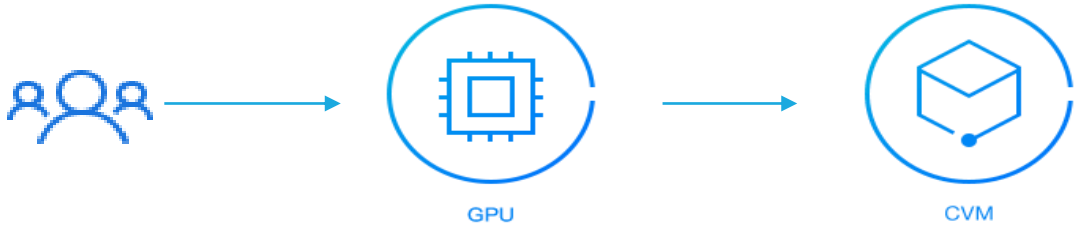
2.2 GPU Cloud Computing: Big Data Analysis

- **Challenges:**
  - AI-based modeling
  - High computing power

GPU (Tesla M40)	vCPU (Xeon E5 v4)	Memory (DDR4)	GPU Memory (GDDR5)
1	28-core	60 GB	24GB

- **Solution:**
  - GPU instances
  - Reduced training time for deep learning
  - No up-front hardware costs

GPU Instance Configuration



Deep Learning

2.3 FPGA Cloud Computing

- This is a computing service based on Field Programmable Gate Array (FPGA). You can use an FPGA image to deploy FPGA instances in a few minutes and enable hardware acceleration.

- **Three types of FPGA instances:**

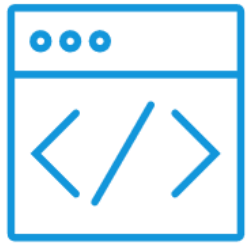
Instance Specifications	FPGA	vCPU	Memory (GB)	Data Disk	Network
FX2.7xlarge60	1	14-core	60	SSD cloud disk	Gigabit network
FX2.14xlarge120	2	28-core	120	SSD cloud disk	Gigabit network
FX2.28xlarge240	4	56-core	240	SSD cloud disk	Gigabit network

2.3 FPGA Cloud Computing

- **Advantages:**

Hardware Acceleration

FPGA hardware acceleration can reach speeds 20 times faster than CPU servers.



Programmable Hardware

FPGA instances can be programed for customized hardware acceleration without the need to change the hardware.

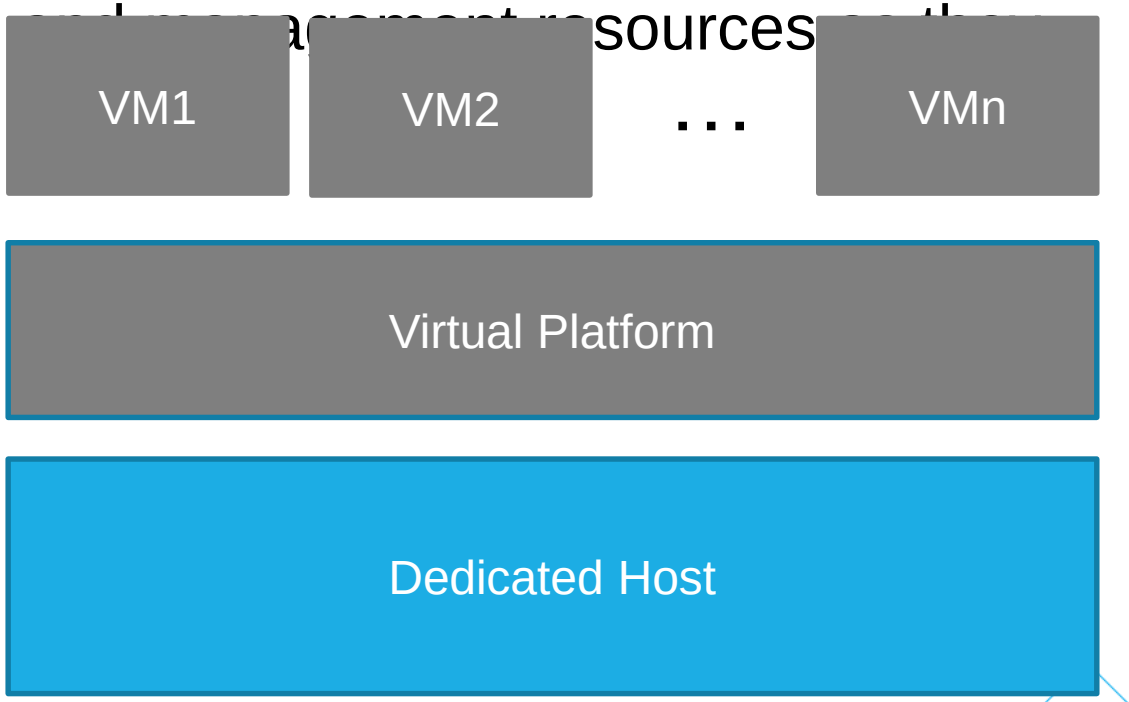
Property Rights Marketplace

We provide a unified, standardized, secure, and reliable FPGA hardware platform and FPGA service marketplace.

- **Use Cases:** Deep learning modeling and real-time image compression

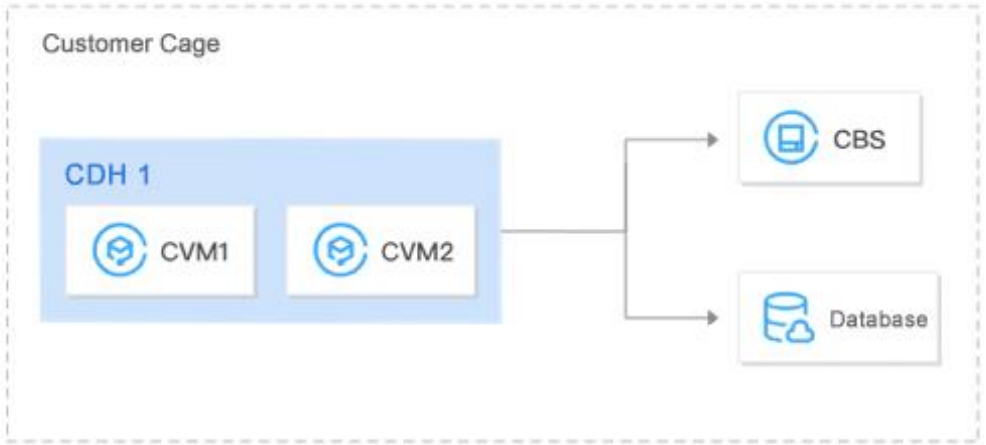
2.4 CVM Dedicated Hosts

- CVM Dedicated Hosts (CDH) enables the purchase of dedicated servers which come installed with the Tencent virtualization system. This allows our customers to create custom CVMs and management resources as they see fit.
- **Product advantages:**
  - Physical isolation
  - Dedicated resources
  - Secure and compliant
  - Flexible configuration

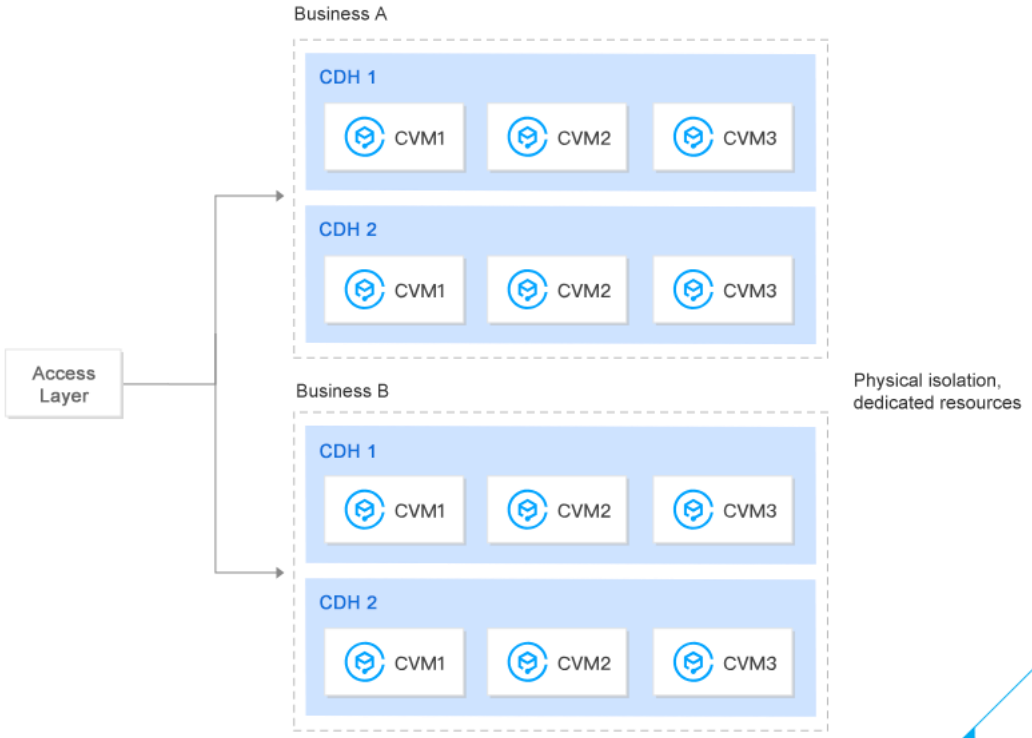


2.4 CDH Use Cases

- Financial services: Secure and compliant



- Businesses with high performance needs: Dedicated resources

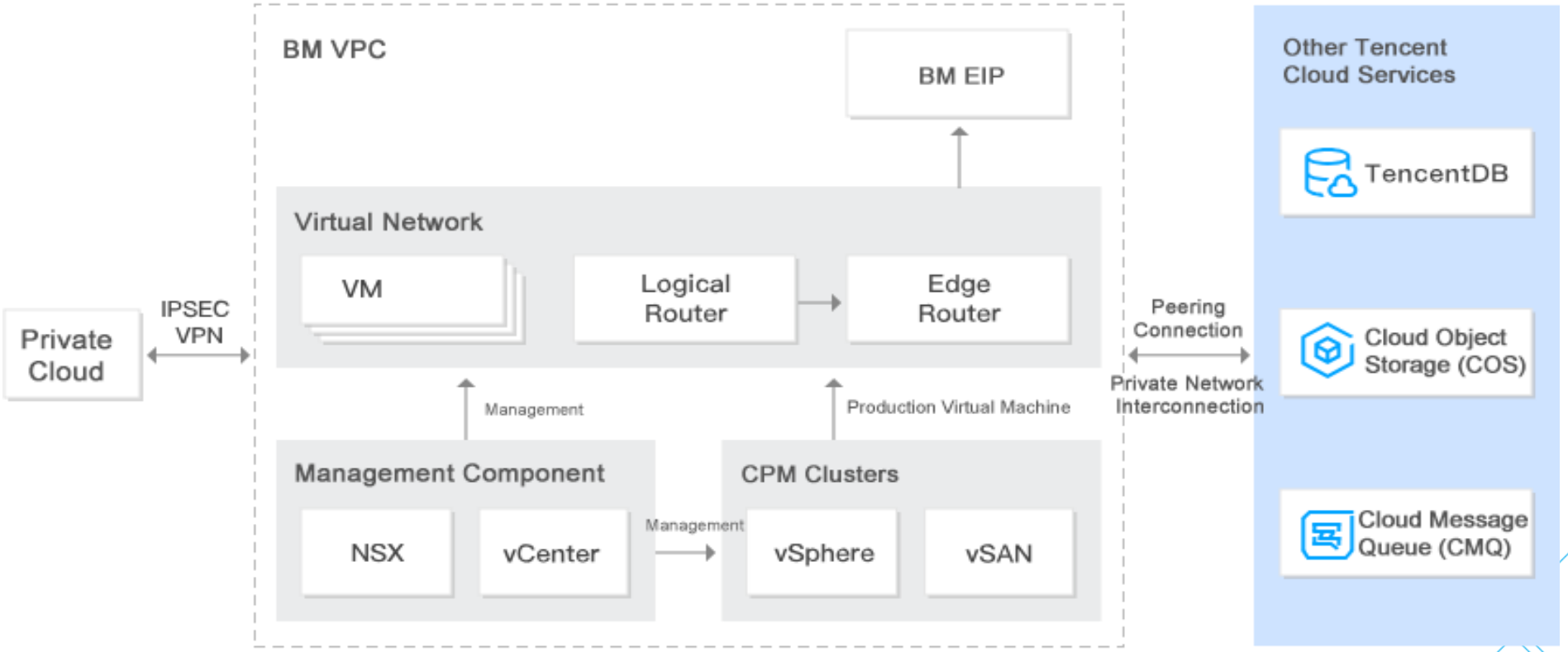


2.5 Cloud Physical Machine (CPM)

- Cloud Physical Machine (CPM) is a physical server rental service that allows you to purchase servers as needed and pay as you go. It provides efficient, safe, and isolated physical clusters designed for cloud computing.
  - CPM Stack-V (deep VMware integration)
  - CPM for ARM
  - CPM for OpenPOWER

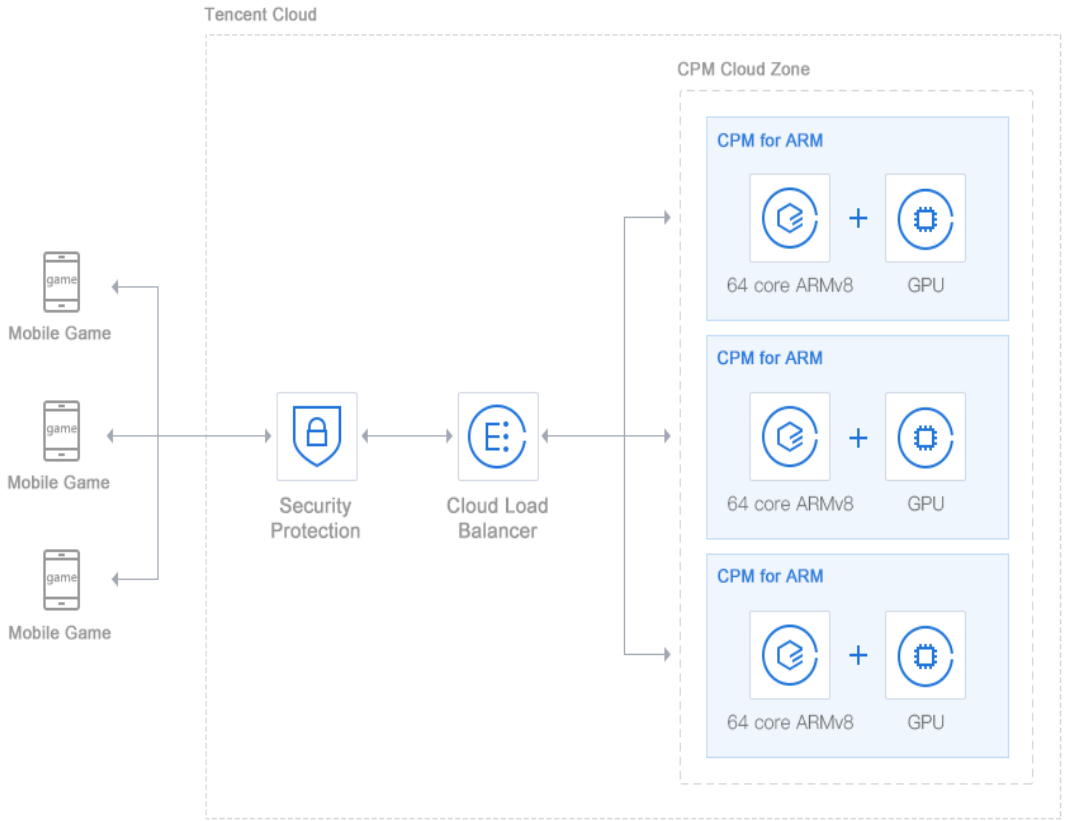
2.5 CPM Stack-V: Hybrid Cloud

- Stack-V clusters can be used with a suite of VMware solutions, and comes with licenses for vSphere 6.5, vSAN 6.6, NSX 6.3, and other products. Through private network interconnection, it can also access mulitple Tencent Cloud services.



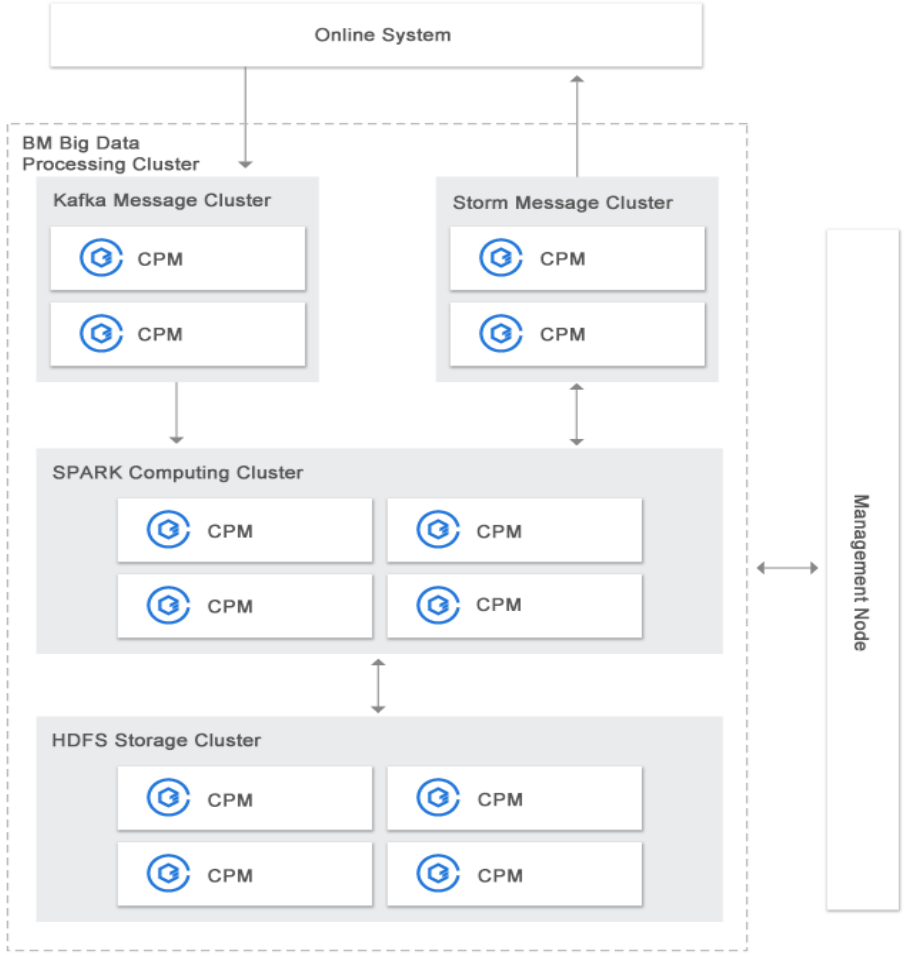
2.5 CPM for ARM: Mobile Gaming

- **Challenges:**
  - Mobile game development requires testing on real devices, which are upgraded frequently.
  - Installation packages are often very large.
  - The cost of external operations is high.
- **Solution:**
  - CPM for ARM instances use the same instruction set as mobile devices with ARM CPUs. They are ideal for running Android emulators for testing and operation.



2.5 CPM for OpenPOWER: Big Data

- **Challenges:**
  - Big data processing requires high performance, high concurrence, high I/O, large memory size, and high-speed transmission.
- **Solution:**
  - CPM for OpenPOWER features a high clock speed, high concurrency, multiple threads, high memory, high bandwidth, and huge cache capacity.



Chapter III Billing Plans for CVM Services

3.1 CVM Pricing

3.2 CVM Billing Mode

3.1 CVM Pricing

CVM Total Price = Configuration Fee* + Bandwidth Fee**

Fee	Configuration Fee	Bandwidth Fee
	0.05USD /hr	0.08USD /GB

* Configuration Fee = CVM Model Fee + Cloud Disk Fee

  For pricing information of CVM models, see here.

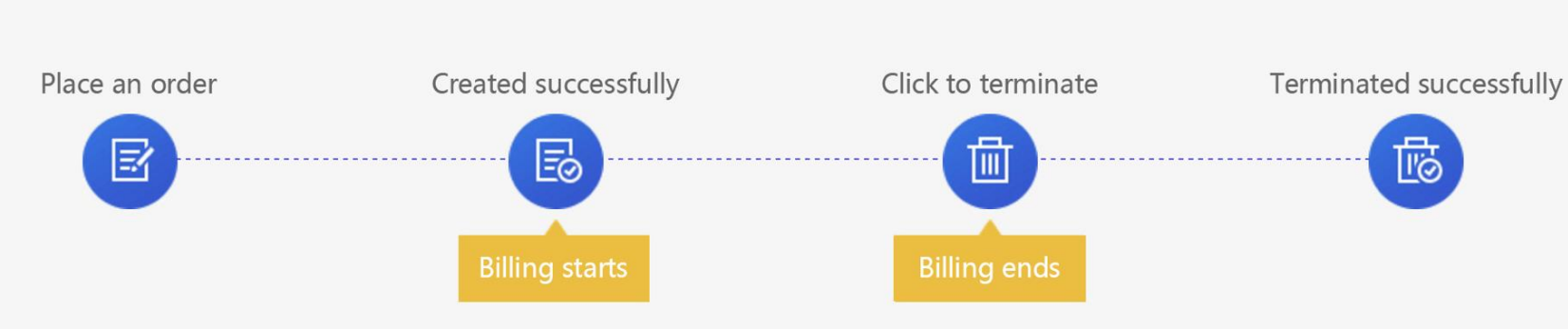
  For pricing information of Cloud Disk, see here.

** Bandwidth fee is billed by traffic and depend on only outbound traffic per unit. Detail pricing here.

3.1 CVM Billing Mode

- Pay-as-you-go

Tencent Cloud CVM is pay-as-you-go, so you only pay for what you use. This flexible billing method is designed to meet your demands in multiple business scenarios. You can activate or terminate CVMs at any time and be billed only for the actual usage. Billable time is **accurate to the second**. It is best suited for big online sales or other scenarios with vastly fluctuating demands.



Billing starts after successful creation of CVM instance and ends when termination is initiated.

Course Summary:

This course covered the following topics:

- Basic knowledge about CVM: Regions, availability zones, instances, images, and storage
- CVM services: CVM, GPU Cloud Computing, FPGA Cloud Computing, CDH, and CPM
- Billing plans for CVM services

Thank you