Virtual Private Cloud

Best Practices

Best Practices

VPC (Virtual Private Cloud) is a network type that Alibaba Cloud recommends users to use. It is logically isolated from other VPCs in the Alibaba Cloud. More and more users choose to use VPC because of:

Isolated network environment

Based on the tunneling technology, VPC isolates the data link layer, providing an independent, isolated, and safe network for each user. Different VPCs are completely isolated with each other. The resources within a VPC can communicate with each other over the intranet, but cannot directly communicate with the resources in other VPC unless you have configured a public IP or bound an EIP to the VPC resources.

Controllable network configurations

You have full control over your own VPC network. You can select its IP address range, create subnets, and configure route tables.

For more details, see VPC overview.

You can start planning and designing your VPC by answering the following questions:

Q1: How many VPCs are need?

Q2: How many VSwitches are needed?

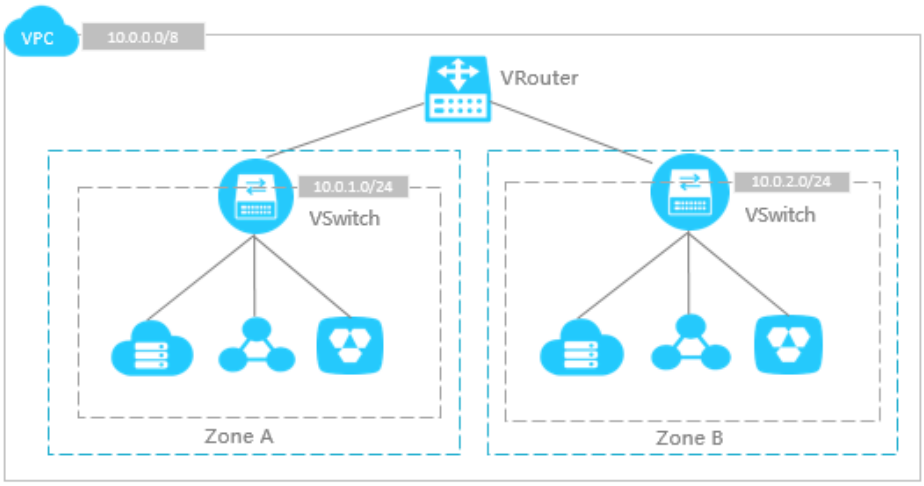
Q3: How to select CIDR blocks?

Q4: How to select CIDR blocks when planning to connect a VPC to on-premises data centers or other VPCs?

Question 1: How many VPCs are needed?

One VPC

We recommend that you create one VPC if you do not have requirements to deploy your systems in multiple regions and do not need to isolate the systems. Currently, a VPC can run as many as 10,000 cloud product instances, which is larger enough to meet your requirements.

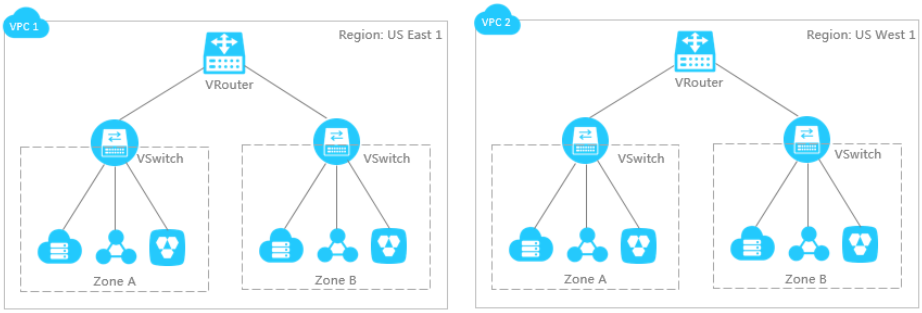


Multiple VPCs

We recommend that you create multiple VPCs if you have the following requirements:

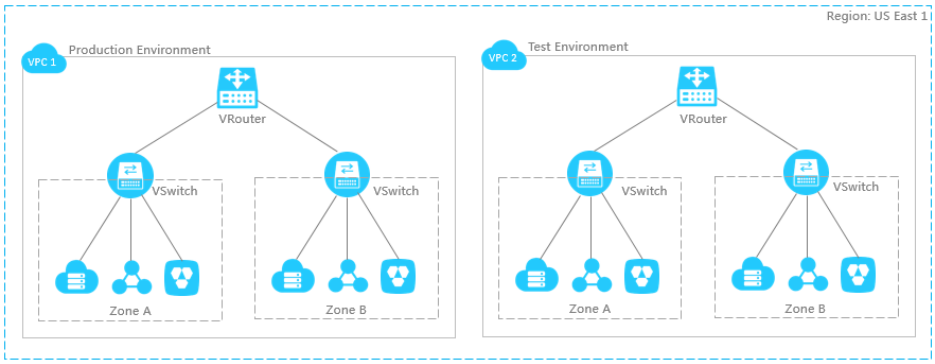
Deploy cloud product resources in different regions

VPC is a region-specific resource that cannot be deployed across regions. If you want to deploy different systems in different regions, you have to create multiple VPCs. Then use Express Connect to connect VPCs.



Isolate different systems

If you want to isolate your systems such as the production environment and the test environment, you have to create multiple VPCs. Different VPCs are completely isolated.



Question 2: How many VSwitches are needed?

In general, we recommend that you create at least two VSwitches for each VPC, and deploy these two VSwitches in two different zones so that you can protect your applications from the failure of a single location.

The network latency between different zones in the same region is very small. You need to verify the network latency in your real business system. The network latency might be larger than expected due to complicated systems calls or cross-zone calls. We recommend that you optimize and adjust the system to find a balance between high availability and low latency.

> Zones are physical areas with independent power supplies and networks in a region. Zones within the same region are interconnected over the intranet.

Additionally, the number of VSwitches is also related to the system size and system design. If the front-end system requires mutual access to the Internet, consider deploying different front-end systems under different VSwitches for better disaster tolerance, and deploying the backend system under other VSwitches.

Question 3: How to select CIDR blocks?

When creating VPCs and VSwitches, you have to specify the IP address range for the VPC and the VSwitch in the form of a Classless Inter-Domain Routing (CIDR) block.

VPC CIDR block

Apply the following rules when specifying CIDR blocks for VPCs. Note that the CIDR block cannot be modified after you create a VPC.

> Use the CIDR blocks provided in the table or their subsets as the IP address range for your VPC.

CIDR block	Number of available IP addresses	Notes
192.168.0.0/16	65,532	Exclude system reserved addresses
172.16.0.0/12	1,048,572	Exclude system reserved addresses
10.0.0.0/8	16,777,212	Exclude system reserved addresses

If you have special requirements on the IP address range, submit a ticket or contact your customer manager.

If you have multiple VPCs, or you want to build a hybrid cloud composed of one or more VPCs and on-premises data centers, we recommend that you use the subset of these standard CIDR blocks as the IP address range for your VPC and ensure that the netmask is no larger than /16.

If you just have one VPC and the VPC does not need to communicate with the on-premises data center, you are free to use any IP address ranges as listed in the previous table.

You also need to consider the use of the classic network. If you plan to connect the cloud product instances in the classic network with a VPC, we recommend that you do not use the CIDR block 10.0.0.0/8, which is also used as CIDR block by the classic network.

VSwitch CIDR block

Apply the following rules when specifying CIDR blocks for VSwitches. Note that the CIDR block cannot be modified after you create a VSwitch.

The allowed block size for a VSwitch is between a /16 netmask and /29 netmask, which can provide 8 to 65,536 IP addresses. The reason for this limitation is that the /16 netmask is larger enough to use with 65,532 IP addresses, while the /29 netmask is too small to make sense.

The CIDR block of a VSwitch can be the same as the CIDR block of the VPC that it belongs to, or a subset of the CIDR block of the VPC. For example, if the CIDR block of a VPC is 192.168.0.0/16, then the CIDR block can be 192.168.0.0/16, or any CIDR block from 192.168.0.0/17 to 192.168.0.0/29.

Note: If the CIDR block of the VSwitch is the same as that of the VPC, you can just

create one VSwitch.

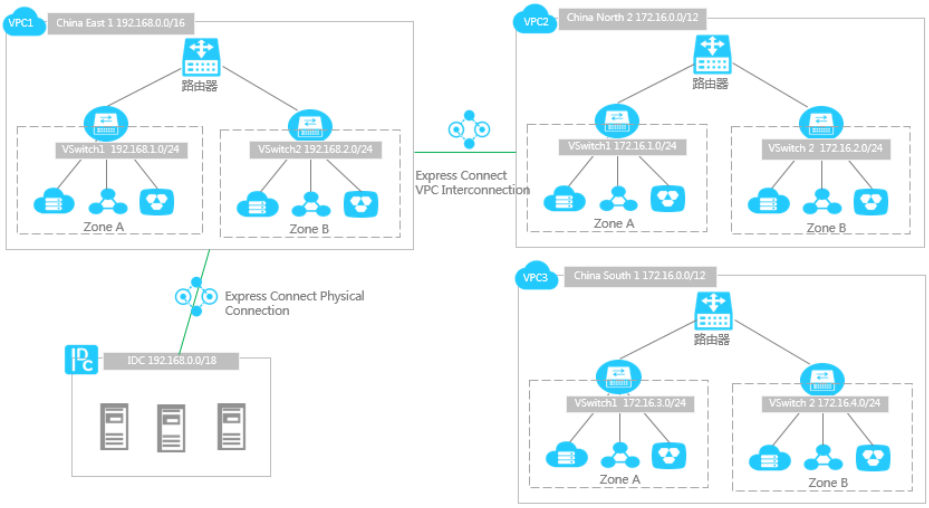
The first and the last three IP addresses are reserved by the system. Take the CIDR block 192.168.1.0/24 as an example, the IP addresses 192.168.1.0, 192.168.1.253, 192.168.1.254 and 192.168.1.255 are reserved by the system.

Alibaba Cloud is planning to provide the ClassicLink feature, which allows the ECS instances in the classic network to communicate with other cloud product instances in a VPC. If the CIDR block of the linked VPC is 10.0.0.0/8, the CIDR block of the VSwitch that requires communication to the classic ECS instance must be 10.111.0.0/16.

You also need to consider the number of cloud product instances to run in the VSwitch when specifying the CIDR block for the VSwitch.

Question 4: How to select CIDR blocks when planning to connect a VPC to on-premises data centers or other VPCs?

As shown in the following figure, assume that you have three VPCs， VPC1 is in China East 1, VPC2 is in China North 2, and VPC3 is in China South 1. VPC1 and VPC2 can communicate with each other over the intranet by using Express Connect, while VPC3 has no requirements to communicate with other VPCs for now but might need to communicate with VPC2 in the future. In addition, you have a self-built data center, which is connected to VPC1 by applying for a leased line.



When you have the requirements to connect a VPC with other VPCs or on-premises data centers, ensure that the CIDR block for each VPC is not the same. Therefore, in the previous example, the CIDR blocks of VPC1 and VPC2 are different, while the CIDR block of VPC3 is the same as the CIDR

block of VPC2, because that VPC3 does not have the requirements to communicate with VPC2. However, the VSwitches in these two VPCs use completely different CIDR blocks to meet the future demand on communication with each other.

The VPC interconnection requires that the CIDR blocks cannot conflict with each other. In this case, the CIDR block is the CIDR block of the VSwitch in each VPC, in other words, the CIDR blocks of VPCs can be the same. But for the best practices, we recommend that you use different CIDR blocks for VPCs.

Apply the following rules when specifying CIDR blocks for the VPCs that require communication with other VPCs or on-premises data centers.

Try to use different CIDR blocks for different VPCs. You can use the subsets of the standard CIDR blocks to increase the number of VPCs.

If the CIDR blocks of the VPCs are the same, then use different CIDR blocks for VSwitches in the VPCs.

If neither of the first nor second principle is met, ensure that the CIDR blocks of VSwitches requiring communication are different.

In the case where the CIDR blocks for both the on-premises data center and the VPC have been determined and cannot be modified, but they still need to communicate with each other, you can use the leased line gateway to solve this problem. Alibaba Cloud is planning to provide this function.

Currently, Alibaba Cloud does not provide a dedicated access control for VPC. The access control in a VPC depends on the cloud products themselves. This section introduces the access control features of ECS, ApsaraDB for RDS and Server Load Balancer.

ECS: security groups

A security group is a virtual firewall that provides stateful packet inspection (SPI). Security groups are used to set network access control for one or more ECS instances. As an important method of security isolation, security groups are used to divide security domains on the cloud.

Security rules

A security group is a logical group that groups instances in the same region with the same security requirements and mutual trust. Each instance belongs to at least one security group, which must be specified at the time of creation. Instances in the same security group can communicate through the network, but instances in different security groups by default cannot communicate through the intranet. However, mutual access can be authorized

between two security groups.

A VPC ECS instance can only join the security group in this VPC. You can authorize or cancel security group rules at any time. Your changes to the security group rule will be automatically applied to ECS instances associated with the security group.

You can control the inbound and outbound network traffic through the VPC ECS instances by configuring corresponding inbound and outbound rules:

Inbound: accept or deny the access of the inbound traffic from a specified IP or CIDR block to a VPC ECS instance.

Outbound: accept or deny the access of the outbound traffic from a specified IP or CIDR block to a VPC ECS instance.

If the rules conflict, the rule with the higher priority takes effect. When their priorities are the same, the Deny rule takes effect.

Configuration policies

When use a security group as whitelist, follow the principle of minimum authorization. For example, the instance used as the jump server in O&M generally has high permissions for intranet access, and is exposed to the Internet, and allows SSH login. Such instances have higher risks, we recommend that managing them separately. You can design the security group rules of the jump server instance as follows:

Deny all accesses to this instance from all addresses (0.0.0.0/0).

Allow access to this instance from the IP address of the specific O&M staff through SSH (TCP Port 22).

The ECS instances within the same security group are interconnected over the intranet, we recommend that place the ECS instances with different business requirements and access controls in the different security groups:

Place the instances that offering Internet services and intranet services in the different security groups.

Apply different security groups to different applications.

Apply different security groups to different deployment environment.

Use the following methods to allow intranet access for the ECS instances in a VPC network:

Configure mutual authorization between security groups.

Configure a security group rule to authorize the inbound and outbound access from a specific CIDR block.

When the number of security groups is small or the network plan and design are not strict, the first method is relatively simple. When the number of security groups increases with the complexity growth in business deployment, authorization on the specific IP address range can effectively simplify the security group configuration and cut down management costs with reasonable planning.

For more information about security groups, refer to Security groups.

ApsaraDB for RDS: whitelist

Powered by the whitelist feature, you can customize the IP addresses allowed to access RDS. Access from unspecified IP addresses will be denied. When you use RDS products in a VPC, you need to add the IP address of the ECS instance that requires the access to the RDS products to the whitelist of the RDS.

For more configuration on ApsaraDB for RDS whitelist, refer to Set a whitelist.

Server Load Balancer: whitelist

Powered by the whitelist feature, you can customize the IP addresses allowed to access the Server Load Balancer listener. It applies to scenarios where the application only allows access from some specific IP addresses. Server Load Balancer is a traffic distribution control service that distributes access traffic to multiple backend ECS servers based on scheduling algorithms and forwarding rules. The access to the Server Load Balancer service is usually open to Internet or intranet users. When the service is open only to specified users, or only allows intranet access, the whitelist feature can effectively control access to the service. To configure a whitelist, you just need to add the user’ s IP address that requires access to the Server Load Balancer service to the whitelist of the Server Load Balancer listener.

For more information, refer to Set whitelist access control.

Currently, Alibaba Cloud provides two Internet-facing products, Server Load Balancer (SLB) and Elastic Public IP address (EIP).

These products are applicable to different scenarios. By using these products, the VPC resources can communicate with the Internet and provide external services.

Product	Function	Scenario keywords	Pricing
SLB	Supports port mapping of the inbound traffic (DNAT). Server Load Balancer distributes the Internet traffic to backend servers based on the configured forwarding rules, and you can dynamically increase or decrease the number of the backend servers according to the business traffic changes. In addition, Server Load Balancer provides the health check function, and automatically blocks the unhealthy servers to ensure the high availability of the system. Server Load Balancer is applicable to the scenario with high availability demands and using multiple ECS servers to provide business services.	DNAT+Layer-4 load balancing/Layer-7 load balancing Health check/high availability	Pay by bandwidth
EIP	An EIP can dynamically bind to a VPC ECS instance. It provides an independent public IP address and bandwidth for each ECS instance to use.	SNAT+DNAT One EIP can bind to only one ECS instance.	Pay by bandwidth

Suggestions:

If you have more than one ECS instances and need to distribute traffic loads based on ports, then choose Server Load Balancer.

If you only need a very small number of ECS instances to provide external services, then choose EIP. You can directly bind the EIP to the VPC ECS instance without other configurations.

Typical scenarios and solutions

The following shows how to choose an Internet-facing product based on different scenarios. Assume that all ECS instances use the VPC network.

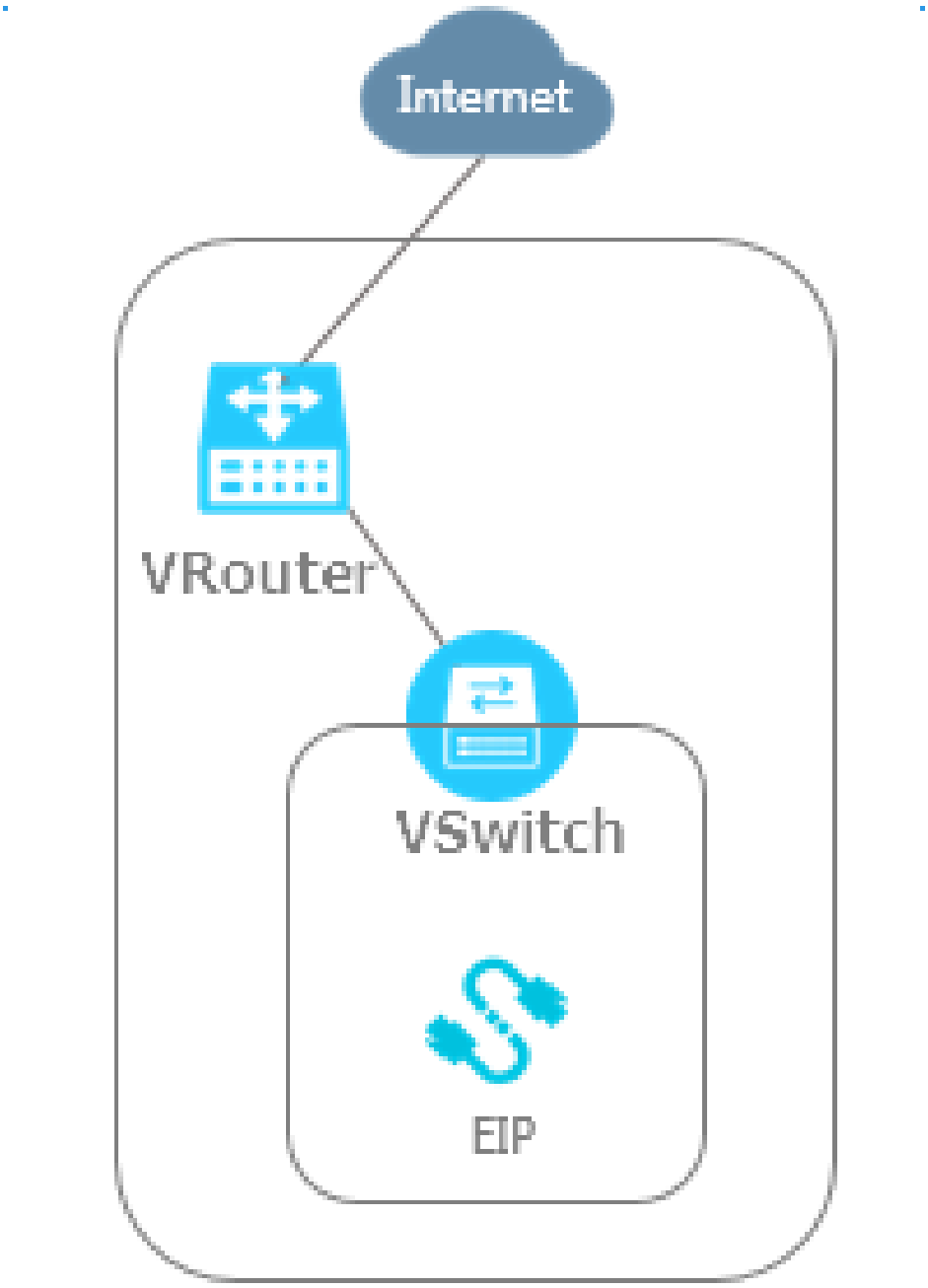
Provide external services

Access the Internet

Provide external services

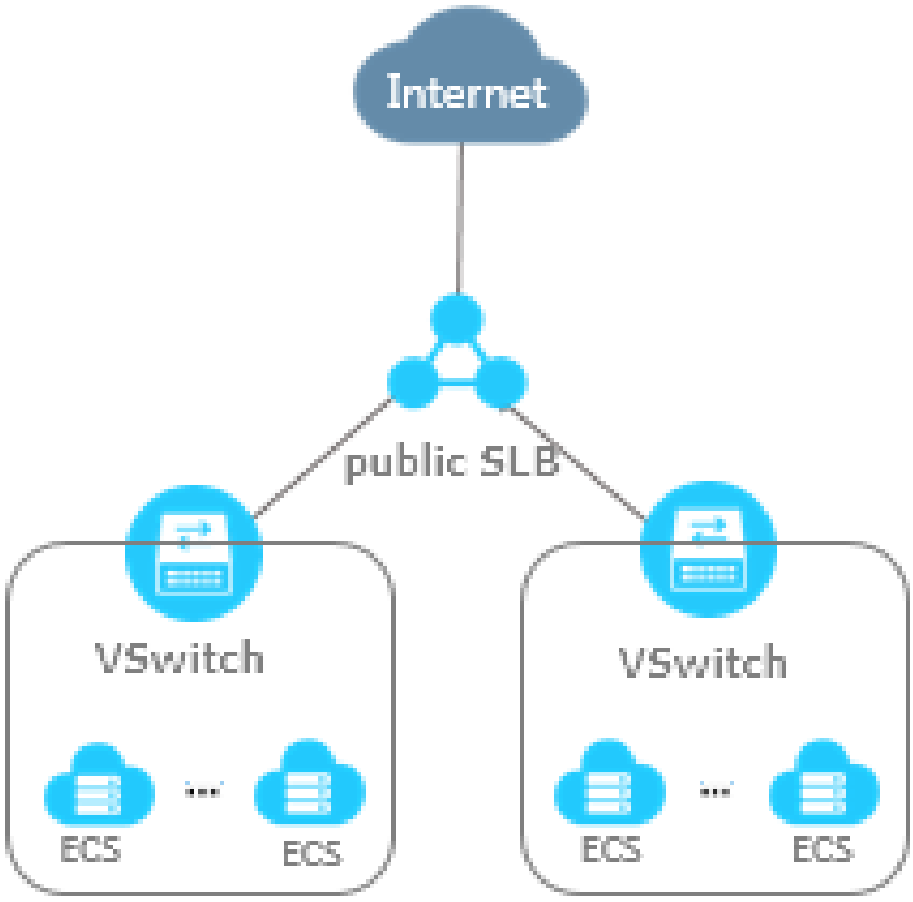
Provide external services with a single ECS instance

When you only have a single application and the access traffic is small, you can deploy the application, the database, and the files on one ECS instance. Then bind an EIP to this ECS instance to provide external services.



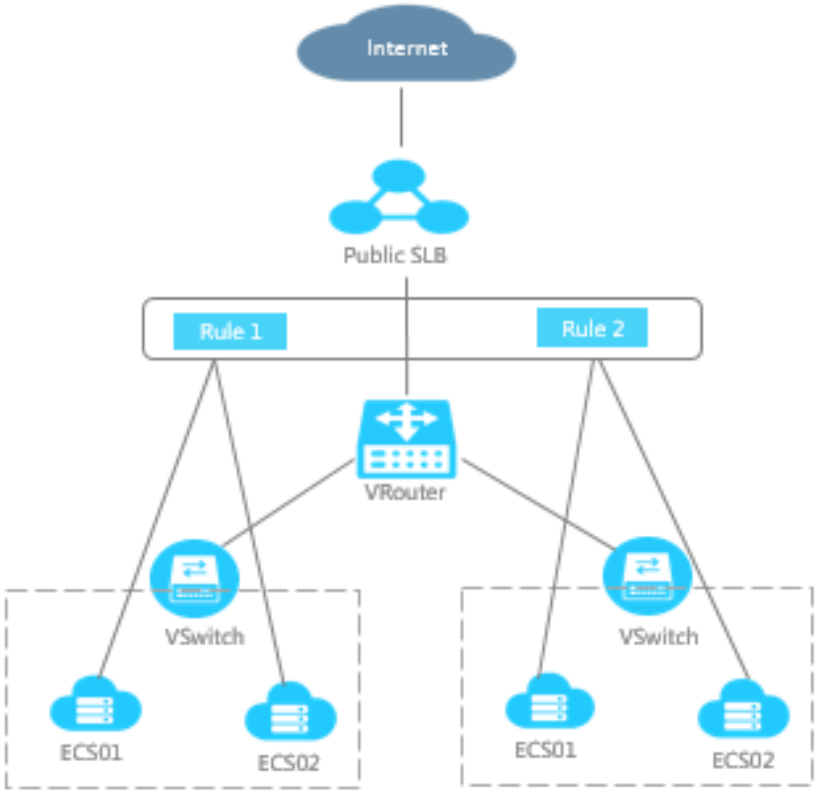
Provide a public Layer-4 load balancing service

When the ECS instance used to provide external services is unable to support the high access traffic, you need to use more ECS instances. In this situation, comparing with the more advanced Layer-7 listener, the simple Layer-4 (TCP) listener can meet your requirements. You can create a Layer-4 (TCP/UDP) listener and add multiple backend ECS instances to build the entire business architecture.



Provide a public Layer-7 load balancing service

In addition to the basic traffic distribution requirement, if you want to distribute the traffic from different business applications to different servers, choose the Layer-7 Server Load Balancer service. You can create a Layer-7 listener with multiple forwarding rules configured.



Access the Internet

Bind EIPs

If an ECS instance requires the Internet access but does not have a public IP, you can bind an EIP to this ECS. You can unbind the EIP whenever you do not need the Internet access.

How to choose to use VPC?

A VPC (Virtual Private Cloud) is an isolated private network. By default, VPC cannot communicate with each other over the intranet. ECS instances in a VPC cannot access the Internet or be accessed from Internet, and a VPC cannot access a classic network through the intranet. But most of the Alibaba Cloud products provide the capabilities of both the Internet access and the intranet access, more than 95% of the cloud products have supported VPC.

Note: The network type must be the same for the cloud product instances that require the intranet communication. For example, if a VPC ECS instance needs to access a Server Load Balancer instance and an RDS instance through the intranet, the network type of the Server Load

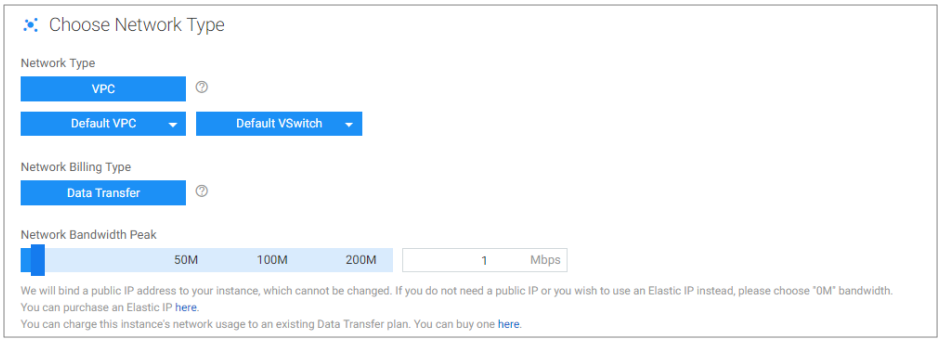
Balancer and RDS must also be VPC and belong to a same VPC, otherwise they cannot communicate with each other.

The ways to use VPC are different for different cloud products.

Select VPC on the purchase page

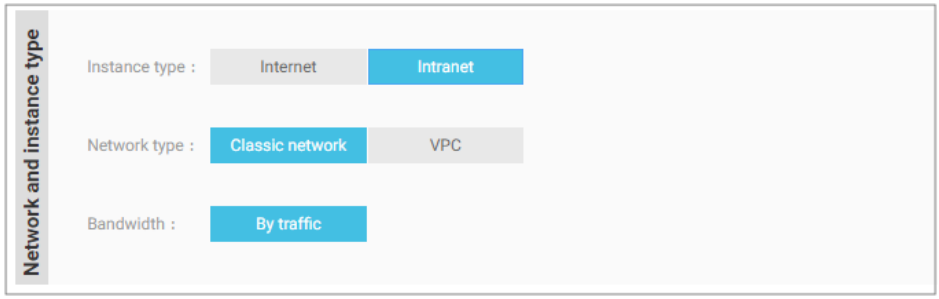
This approach mainly applies to the instance-type cloud products, such as ECS, RDS and Server Load Balancer. The purchase pages of such cloud products offer an option to select the network type.

The following figure shows the network type option on the ECS purchase page.



The following figure shows the network type option on the RDS purchase page.

The following figure shows the network type option on the Server Load Balancer purchase page.



Select the VPC access on the console

This approach mainly applies to the cloud products such as OTS, Container Service, E-MapReduce and NAS.

You can set a VPC access domain for OTS instances on the OTS console; for Container Service and E-MapReduce, you can select VPC when creating a cluster on the console; the Network Attached Storage product provides the VPC mount point.

Check the VPC access endpoint through documentation

For cloud products such as Log Service and OSS, refer to the following documents:

- Regions and endpoints
- Service endpoint

How to switch the network type?

Some instance-type cloud products have supported the network type switch. For example, you can switch the network type of an RDS instance on the console.

ECS products will also be able to switch from the classic network to VPC.

Server Load Balancer does not support switching from the classic network to VPC. As an alternative, you can purchase a VPC instance.

Note that some cloud products such as CDN and Situation Awareness do not need VPC.

For more information, refer to Migrate classic network to VPC.