UCS Service Overview

Issue 01
Date 2023-09-12

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process.* For details about this process, visit the following web page:
https://www.huawei.com/en/psirt/vul-response-process
For vulnerability information, enterprise customers can visit the following web page:
https://securitybulletin.huawei.com/enterprise/en/security-advisory

Contents

1 What Is Huawei Cloud UCS?.....1
2 Advantages.....3
3 Application Scenarios.....6
3.1 Live Commerce.....6
3.2 Finance.....7
3.3 Automobile.....8
4 Billing.....10
5 Permissions.....11
6 Constraints.....17
7 Related Services.....19

1 What Is Huawei Cloud UCS?

Huawei Cloud Ubiquitous Cloud Native Service (UCS) is the first distributed cloud native product in the industry. It provides consistent experience in the cloud native application deployment, management, and ecosystem. Cloud native applications can freely run across regions and clouds with intelligent traffic distribution.

Running on Karmada, CNCF's first multi-cloud container orchestration project, Huawei Cloud UCS enables you to run cloud native applications across clouds or regions, no matter whether they are running on Huawei Cloud (CCE and CCE Turbo clusters), partner clouds (CCE clusters), other clouds (other cloud vendors' Kubernetes clusters), or on-premises infrastructure (clusters provided by Huawei Cloud and clusters deployed by yourself). UCS extends cloud native to central regions, hotspot areas, customer premises, and business locations.

Huawei Cloud UCS innovates in three ways:

- A new way for application-data collaboration

  Your data can migrate to where your applications run. Integrated migration, scaling, and disaster recovery remove geographical restrictions on your application running.
- A new way to provision compute

  With distributed scheduling, millions of nodes collaborate to provision compute to applications across clouds at any time.
- A new way to manage application traffic

  Service requests can be intelligently distributed in real time, across regions, and on demand.

Functions

- **Unified cluster management**

  You can connect Huawei Cloud clusters, on-premises clusters, attached clusters, partner cloud clusters, and multi-cloud clusters across clouds and regions to UCS and manage them in a unified manner.
- **Central delivery of cluster configurations**

  You can manage the configurations of your multi-cloud clusters all in one place by **controlling permissions** of tenants and users in enterprise projects, and audit cluster compliance through a unified **policy center**.

- **Visualized monitoring and O&M**

  You can obtain insights on your containers and service meshes from multiple dimensions. UCS is compatible with open source Prometheus and OpenTelemetry, supports custom dashboards, and checks the health status of your running services.

- **Collaborative compute supply and optimal deployment**

  Running on Karmada, UCS can connect to thousands of distributed Kubernetes clusters, coordinate compute resources on millions of nodes, and respond in just seconds. Supporting multiple types of distributed deployment policies, UCS can find the best-fit location to deploy your application based on global resource distribution, service characteristics, geographical locations, network QoS, and resource balancing.

- **Unified traffic management**

  UCS distributes requests globally according to user locations and service policies across clouds and clusters. Application traffic can be split based on weight and content. Advanced functions such as grayscale release, failover, circuit breaking, and rate limiting are also available.

- **Application-data collaboration**

  UCS integrates data and services and automates migration, cloning, data replication, and cross-cloud scaling for your applications. Data at the storage, container, and middleware layers is associated to support application DR, auto scaling, and migration.

- **One ecosystem with globally available applications**

  With an in-house deployment engine, UCS provides ready-to-use components with unified specifications, which can be deployed globally with just a few clicks and managed throughout their lifecycle.

2 Advantages

Huawei Cloud UCS Advantages

UCS helps you manage cloud native services across clouds and regions while ensuring consistent experience. Extending cloud native to wherever your services run, Huawei Cloud UCS eases your journey to digital upgrade.

- **Unified multi-cloud experience**

  You can connect distributed Kubernetes clusters to UCS, including those running on Huawei Cloud (central region, IEC, and CloudPond), on-premises infrastructure, and third-party clouds. You can manage the configurations of your multi-cloud clusters all in one place by controlling permissions of tenants in enterprise projects, and perform fine-grained management on IAM users' permissions on Kubernetes resources. You can also audit the service compliance of your clouds and clusters, as UCS manages security policies and resource access restrictions of each cluster in a unified manner.

- **Collaborative computing**

  Huawei Cloud UCS is built on Karmada, a multi-cluster orchestration project contributed by Huawei Cloud to CNCF. With multi-cloud capabilities, UCS can connect to thousands of Kubernetes clusters across clouds and regions, and schedule applications by coordinating millions of nodes. Your applications can scale across clouds and clusters, migrate upon failures, and run in the best condition based on global resource distribution, geographical location, network QoS, affinity, and resource balancing. With UCS, compute is at your fingertips anytime, anywhere.

- **Intelligent traffic distribution**

  Unified container network orchestration and service discovery implement a flattened network across clouds and clusters. This network allows consistent service experience and makes communications secure and reliable. Also UCS can send service requests to the best-fit backend cluster. It does so with less access latency and by the policies you set based on factors such as visitor CIDR blocks, regions, and carriers. UCS works with service meshes for unified service governance. Scheduling can be based on network QoS priorities. Geographical affinity can be implemented. Automated grayscale release, visualized service topology, and service tracing are now all available for you to manage access traffic globally in real time and on demand.

- **Data migration with applications**

UCS automates data replication across clouds for the storage infrastructure layer, container cluster layer, and middleware layer. Data goes wherever your applications run. You can scale your apps on the distributed infrastructure with ease. During scaling, data scanning and rebuild are automated and application-centric. Integrated migration, scaling, and disaster recovery are completed for the entire service.

Huawei Cloud UCS vs Traditional Cloud Native

Table 2-1 Differences between Huawei Cloud UCS and traditional cloud native

Item	Traditional Cloud Native	Huawei Cloud UCS
Experience	Vendor lock-in exists due to customizations on cloud native technologies. Therefore, users may have inconsistent experience when managing their clusters in different regions, and the learning curves could be steep.	**Unified multi-cloud experience** Huawei Cloud UCS connects your clusters running on different clouds across central areas, hotspot areas, on-premises data centers, and business locations. All in one place with unified experience.
Scalability	Compute resources cannot be scheduled across clouds.	**Collaborative compute** Running on Karmada, Huawei Cloud UCS schedules multi-cloud resources in a unified manner and bursts on-premises applications to the public cloud. Supporting multiple types of distributed deployment policies, UCS can find the best-fit location to deploy your application based on global resource distribution, service characteristics, geographical locations, network QoS, and resource balancing.
Application management	In most cases, traditional cloud native manages applications in a single region, demanding little on application migration. When scaling apps across clouds, O&M personnel need to clone and migrate app data manually. Low efficiency and heavy workload.	**Data migration with applications** Huawei Cloud UCS supports synchronous data replication across clouds for you to scale your applications on the distributed infrastructure. Application DR, scaling, and migration become much easier.

Item	Traditional Cloud Native	Huawei Cloud UCS
Traffic management	Traffic management is decoupled from services. Requests are not distributed on demand. Access latency is high for cross-region/carrier requests.	**Unified traffic management** Huawei Cloud UCS distributes requests to the nearest backend cluster to reduce the access latency based on different policies such as CIDR blocks, regions, and carriers.
Efficiency	Applications need to be manually deployed in each cluster across clouds, a labor-intensive process.	**Ready-to-use services** Huawei Cloud UCS allows you to batch deliver application settings to each cluster in different regions through edge-cloud synergy. Much faster than before without repetitive configuration.
O&M	Services scattered in the central region, IDC, and edge nodes need to be monitored separately, a heavy burden for O&M.	**Multi-dimensional O&M** Huawei Cloud UCS supports multi-dimensional monitoring and O&M on your resources in all regions, and is compatible with open source Prometheus and OpenTelemetry ecosystems.

3 Application Scenarios

3.1 Live Commerce

Scenario

E-commerce apps or platforms may fail or crash due to the sharp increase of live streaming viewers or buyers for promotion and flash sales, running out of server resources and resulting in high service latency.

UCS balances service traffic and edge-cloud resource allocation to ensure smooth services and user experience during peak hours.

Advantages

- **Nearby access**

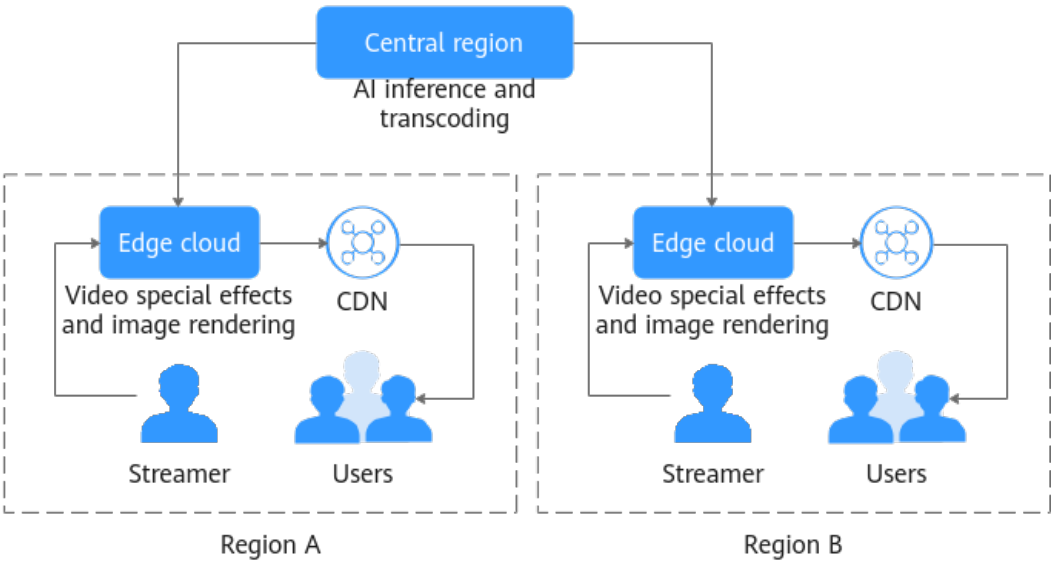
  Intelligent routing and nearby access based on user locations, reducing end-to-end service latency

- **Unified compute supply**

  Flexible scheduling of cross-region edge and cloud compute resources based on the number of live streaming viewers and application requirements, improving resource utilization

Suggested Solution

Figure 3-1 Solution for live commerce



3.2 Finance

Scenario

Customers in this industry may find it hard to balance mobile services and data privacy. The existing hybrid cloud architecture could be a solution, but there are still some pain points unsolved.

- Pain point 1: Services are scattered and cannot be quickly expanded or managed on a large scale. Traffic bursts impact system running.
- Pain point 2: Service instances are difficult to deploy in different environments due to no unified cloud ecosystem specifications. Financial cloud native SaaS applications are in great demand.
- Pain point 3: Available traffic governance capabilities cannot satisfy data-sensitive and delay-sensitive services.
- Pain point 4: Smart devices challenge management, operations, and supervision.
- Pain point 5: Without cross-DC service monitoring and governance, service instances cannot be migrated across clouds.

Advantages

UCS centrally manages resources and data spanning on-premises infrastructure, edge cloud, and central cloud, and supports one-stop distribution and scheduling.

- **Collaborative compute**

  For mobile financial services, UCS supports fast scaling and large-scale governance. On-premises, edge, and cloud resources are collaboratively scheduled to cope with traffic bursts.

- **Unified ecosystem**

  UCS builds a standard financial application ecosystem. Your applications can be easily distributed, deployed, and migrated across regions and clouds.

- **Edge-cloud synergy**

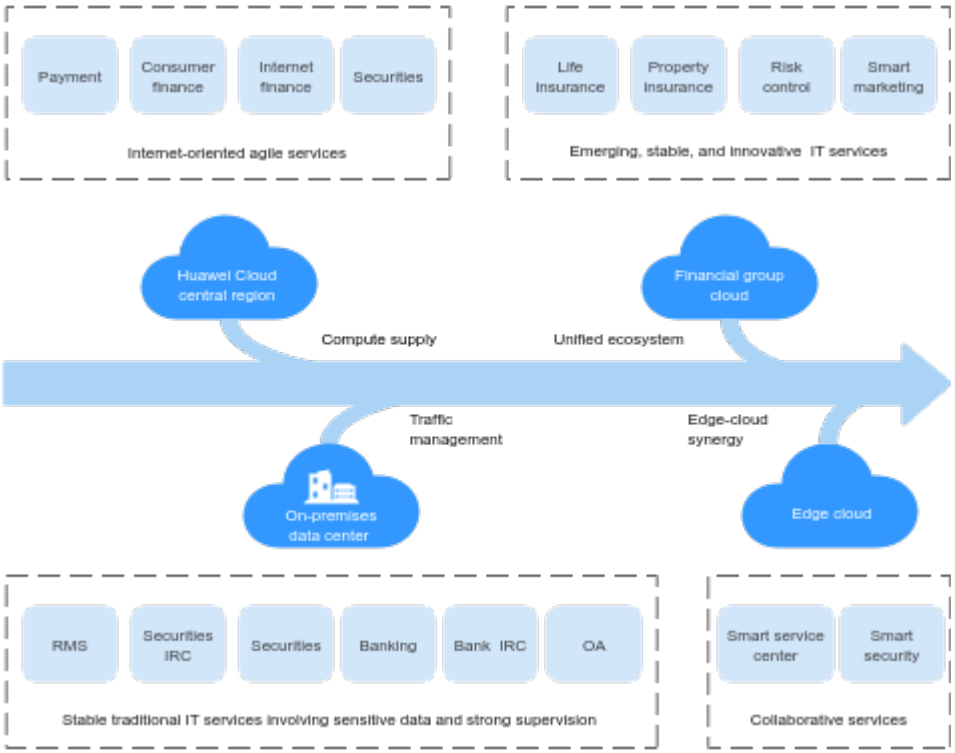
  UCS can collaboratively manage a large number of terminals, edge devices, and applications to build intelligent security and smart branches.

- **Multi-cloud synergy**

  UCS helps you build a multi-region, multi-center digital architecture for unified governance across clouds and DCs.

Suggested Solution

Figure 3-2 Solution for finance



3.3 Automobile

Scenario

New service scenarios such as Internet of Vehicles (IoV) are demanding innovations in digital marketing, smart production, and smart stores. The industry is diving into digital transformation, but there are many obstacles under the surface.

- Challenge 1: Traditional stable services do not maximize resource utilization. Infrastructure resources cannot collaborate.

- Challenge 2: Poor scaling cannot support a large number of concurrent users. The network latency is high.
- Challenge 3: Services vary in types and deployment locations, causing difficult O&M.

Advantages

UCS integrates resources on the edge cloud, on-premises infrastructure, and Huawei Cloud to speed up digital transformation for automobile companies.

- **Collaborative compute**

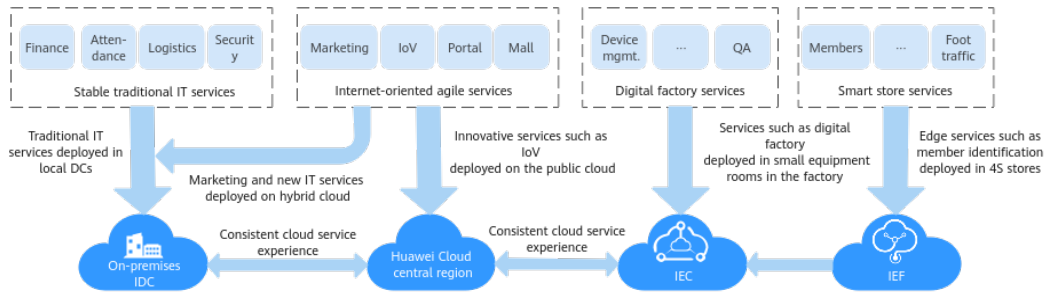
  Agile and stable services run on the same platform that integrates all infrastructure resources, improving resource utilization.
- **Unified traffic governance**

  IoV and Internet services can connect to the most efficient networks to reduce latency.
- **Unified management**

  O&M and operations are performed on one platform for network-wide distributed applications, improving O&M efficiency.

Suggested Solution

Figure 3-3 Solution for automobile



4 Billing

Billing Modes

There are yearly/monthly and pay-per-use billing modes to meet your requirements.

- Yearly/Monthly is a prepaid billing mode. You pay in advance for a subscription term. Before purchasing yearly/monthly resources, ensure that your account has sufficient balance.
- Pay-per-use is a postpaid billing mode. You pay as you go and just pay for what you use.

After purchasing clusters or cluster resources, you can change their billing modes if the current billing mode cannot meet your service requirements.

Billed Items

You will be billed for clusters managed by UCS. The UCS price depends on the cluster type, number of vCPUs of a cluster, and required duration. To view the number of vCPUs (included in the UCS price) of each cluster, run the following command:

kubectl get nodes -o jsonpath='{range .items[*]}{.metadata.name}{"\t"} {.status.conditions[?(@.type=="Ready")].status}{"\t"} {.status.capacity.cpu} {"\n"}' | grep True

5 Permissions

If you need to assign different permissions to employees in your enterprise to access your UCS resources, IAM is a good choice for fine-grained permission management. IAM provides identity authentication, permission management, and access control, helping you secure access to your resources.

With IAM, you can use your account to create IAM users, and assign permissions to the users to control their access to specific resources. For example, some software developers in your enterprise need to use UCS cluster resources but should not perform high-risk operations such as unregistering clusters. In this case, you can create IAM users for the software developers and grant them only the permissions for using UCS cluster resources.

UCS Permission Types

UCS provides refined permission management based on the role access control (RBAC) capability of IAM and Kubernetes. Permission control can be implemented by UCS service resource and Kubernetes resource in a cluster. The two permission types apply to different resource types and are granted using different methods.

- **UCS resource permissions** are granted based on the system policies of IAM. UCS resources include container fleets, clusters, and federation instances. Administrators can grant different permissions to different user roles (such as development and O&M) to control their use of UCS resources.
- **Kubernetes resource permissions in a cluster** are granted based on the Kubernetes RBAC capability. Refined permissions can be granted to Kubernetes resource objects in a cluster. With permission setting, the permissions for performing operations on different Kubernetes resource objects (such as workloads, jobs, and services) will vary with users.

If your team mainly uses UCS resources, IAM system policies can meet your requirements. If you need refined permissions on Kubernetes resource objects in the cluster, use IAM system policies together with Kubernetes RBAC.

UCS Resource Permissions

New IAM users do not have any permissions assigned by default. You need to first add them to one or more groups and attach policies or roles to these groups. The users then inherit permissions from the groups and can perform specified operations on cloud services.

UCS is a global service deployed in all physical regions. You can grant UCS permissions to users in the global service project. In this way, users do not need to switch regions when they access UCS.

You can grant users permissions by using roles and policies.

- Roles: A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. IAM provides a limited number of roles for permission management. When using roles to grant permissions, you also need to assign other roles on which the permissions depend to take effect. However, roles are not an ideal choice for fine-grained authorization and secure access control.
- Policies: A fine-grained authorization tool that defines permissions for performing operations on specific cloud resources under certain conditions. This type of authorization is more flexible. It is ideal for keeping permissions minimized. For example, you can grant IAM users only the permissions for managing a certain type of fleets and clusters.

Table 5-1 describes all system permissions of UCS.

Table 5-1 UCS system permissions

System Role/ Policy Name	Description	Type
UCS FullAccess	UCS administrator with full permissions, including creating permissions policies and security policies	System policy
UCS CommonOperations	Common UCS user with permissions for creating workloads, distributing traffic, and other operations	System policy
UCS CIAOperations	UCS Container Intelligent Analysis administrator with full permissions	System policy
UCS ReadOnlyAccess	Read-only permissions on UCS (except for Container Intelligent Analysis)	System policy

Services on Huawei Cloud are interdependent, and UCS depends on other cloud services to implement some functions, such as image repository and domain name resolution. Therefore, the preceding four system policies are often used together with roles or policies of other cloud services for refined permission granting. When granting permissions to IAM users, the administrator must comply with the principle of least privilege. **Table 5-2** lists the minimum permissions required by the administrator, operation, and read-only permissions of each UCS function.

Table 5-2 Minimum permissions required by UCS

Description	Permission Type	Permission	Minimum Permission
Fleet	Admin	• Creating and deleting a fleet • Registering a Huawei Cloud cluster (CCE cluster and CCE Turbo cluster), on-premises cluster, or attached cluster • Unregistering a cluster • Adding a cluster to or removing a cluster from a fleet • Associating permission policies with a cluster or fleet • Enabling cluster federation and performing federation management operations (such as creating a federated workload and creating domain name access)	UCS FullAccess
	Viewer	Querying clusters and fleets or their details	UCS ReadOnlyAccess
Huawei Cloud clusters	Admin	Read-write permissions on Huawei Cloud clusters and all Kubernetes resource objects (including nodes, workloads, jobs, and services)	UCS FullAccess + CCE Administrator
	Operation	Read-write permissions on Huawei Cloud clusters and most Kubernetes resource objects and read-only permissions on Kubernetes resource objects such as namespaces and resource quotas	UCS CommonOperations + CCE Administrator
	Viewer	Read-only permissions on Huawei Cloud clusters and all Kubernetes resource objects (including nodes, workloads, jobs, and services)	UCS ReadOnlyAccess + CCE Administrator
On-premises/ Attached/ Multi-cloud clusters	Admin	Read-write permissions on on-premises/attached/multi-cloud clusters and all Kubernetes resource objects (including nodes, workloads, jobs, and services)	UCS FullAccess

Description	Permission Type	Permission	Minimum Permission
	Operation	Read-write permissions on on-premises/attached/multi-cloud clusters and most Kubernetes resource objects and read-only permissions on Kubernetes resource objects such as namespaces and resource quotas	UCS CommonOperations + UCS RBAC (The list permission for namespaces is required.)
	Viewer	Read-only permissions on on-premises/attached/multi-cloud clusters and all Kubernetes resource objects (including nodes, workloads, jobs, and services)	UCS ReadOnlyAccess + UCS RBAC (The list permission for namespaces is required.)
Image repository	Admin	All permissions on SoftWare Repository for Container (SWR), including creating organizations, uploading images, viewing images or details, and downloading images	SWR Administrator
Permissions	Admin	• Creating and deleting a permission policy • Viewing permissions or details **NOTE** When creating a permission policy, you need to grant the **IAM ReadOnlyAccess** permission (read-only permissions on IAM) to IAM users to obtain the IAM user list.	UCS FullAccess + IAM ReadOnlyAccess
	Viewer	Viewing permissions or details	UCS ReadOnlyAccess + IAM ReadOnlyAccess
Policy Center	Admin	• Enabling the Policy Center • Creating and disabling a policy • Querying policies • Viewing policy implementation details	UCS FullAccess
	Viewer	For fleets and clusters with Policy Center enabled, users with this permission can view policies and policy implementation details.	UCS CommonOperations or UCS ReadOnlyAccess

Description	Permission Type	Permission	Minimum Permission
Traffic distribution	Admin	Operations such as creating a traffic policy, suspending and deleting a scheduling policy	(Recommended) UCS CommonOperations + DNS Administrator or UCS FullAccess + DNS Administrator
	Viewer	Viewing traffic policies or details	UCS ReadOnlyAccess + DNS Administrator
Container Intelligent Analysis	Admin	• Connecting clusters to a fleet or canceling cluster connection • Viewing monitoring data in multiple aspects, such as infrastructure and application workload	UCS CIAOperations

Kubernetes Resource Permissions in a Cluster

Kubernetes resource permissions in a cluster are granted according to Kubernetes RBAC. The administrator can grant users refined permissions on specific Kubernetes resource objects in the cluster. These resources are cluster-level and namespace-level. Refined operation permissions include **get**, **list**, **watch**, **create**, **update**, **patch**, and **delete**. The permissions take effect on the namespaces of a fleet or on clusters that are not added to the fleet. The operation permissions are described as follows:

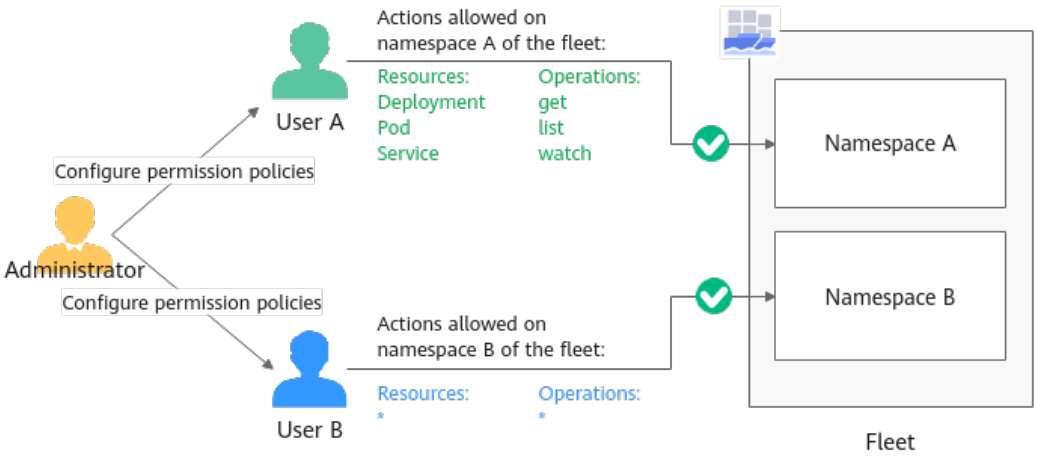
- **get**: Retrieves a specific resource object by name.
- **list**: Retrieves all resource objects of a specific type in the namespace.
- **watch**: Responds to resource changes.
- **create**: Creates a resource.
- **update**: Updates a resource.
- **patch**: Updates resources partially.
- **delete**: Deletes a resource.

NOTE

For details about cluster-level and namespace-level resources, see **Kubernetes Resource Objects**.

For example, after permission policies are configured according to the scheme shown in **Figure 5-1**, user A can perform **get**, **list**, and **watch** (read-only) operations only on Deployments, pods, and Services in namespace A of the fleet, and user B can perform all operations on all resources in namespace B of the fleet.

Figure 5-1 Granting permissions on Kubernetes resources



Three common permission types are available on the UCS console: admin, developer, and read-only. You can grant these permission types to users. If these permission types cannot meet your requirements, you can customize permissions by specifying the operation type and resource object.

Table 5-3 Permission types

Permission Type	Description
Admin	Read-write permissions on all Kubernetes resource objects
Developer	Read-write permissions on most Kubernetes resource objects and read-only permissions on Kubernetes resource objects such as namespaces and resource quotas
Viewer	Read-only permissions on all Kubernetes resource objects

6 Constraints

This section describes the constraints on using Huawei Cloud UCS.

Kubernetes Versions

The version of each Kubernetes cluster connected to UCS must be between 1.19 and 1.26.

Regions

When a cluster connects to UCS through a private network, you need to use Direct Connect (DC) or Virtual Private Network (VPN) to connect the on-premises network to the cloud VPC, and use VPC Endpoint (VPCEP) to connect to UCS through the private network.

In this scenario, when creating a DC, VPN, VPC, or VPCEP, you can create it only in AP-Singapore. No region restriction is involved if you are not using a private network to connect your cluster to UCS.

Functions

UCS is in open beta test (OBT). Functions including service mesh, container intelligent analysis (CIA), and Operator Service Center (OSC) are not yet available.

Quota Limits

Quotas put limits on the quantity or capacity of resources available to users. UCS has quota limits on clusters, fleets, permissions, and cluster federations, as shown in **Table 6-1**. If the default quota provided by UCS cannot meet your service requirements, you can submit a service ticket to increase your quota.

- Cluster quota: specifies the maximum number of clusters connected to UCS. This item applies to Huawei Cloud clusters, on-premises clusters, attached clusters, multi-cloud clusters, and partner cloud clusters.
- Fleet quota: specifies the maximum number of fleets owned by a user.
- Permission quota: specifies the maximum number of permission policies that a user can create on the **Permissions** page.
- Cluster federation quota: specifies the maximum number of cluster federations that a user can enable. You cannot request a quota increase.

Table 6-1 UCS quota items

Quota Item	Default
Cluster	50
Fleet	50
Permission	50
Cluster federation	1

For other cloud services you may also use when running UCS, such as Elastic Cloud Server (ECS), Elastic Volume Service (EVS), Virtual Private Cloud (VPC), Elastic Load Balance (ELB), SoftWare Repository for Container (SWR), and Domain Name Service (DNS), their quotas are independent of those of UCS and are managed by themselves. For details, see **Quotas**.

7 Related Services

Huawei Cloud is a unified cluster management platform. **Figure 7-1** shows the relationships between UCS and other services.

Figure 7-1 Relationships between UCS and other services

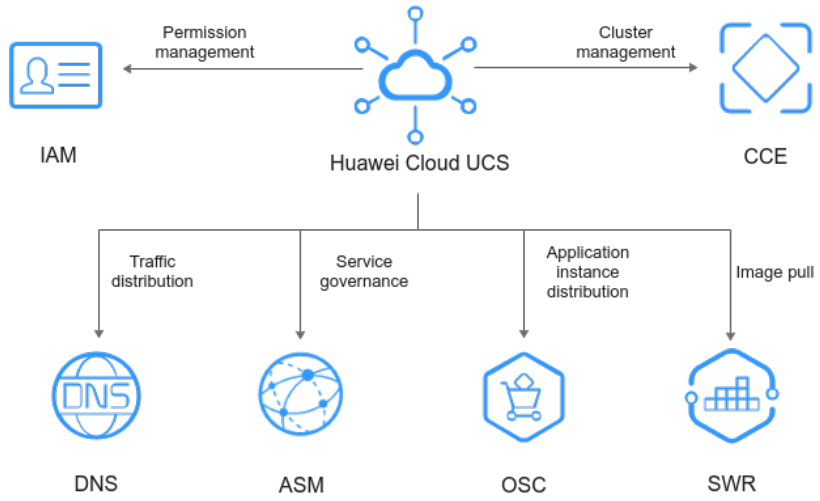


Table 7-1 Relationships between UCS and other services

Service	Relationship	Related Feature
Cloud Container Engine (CCE)	UCS automatically takes over CCE and CCE Turbo clusters and provides clusters with functions such as application distribution, traffic/data management, and cluster monitoring.	Registering a Huawei Cloud Cluster
Identity and Access Management (IAM)	UCS provides fine-grained permission management based on IAM.	Permissions

Service	Relationship	Related Feature
Domain Name Service (DNS)	UCS integrates with DNS to resolve domain names for large-scale traffic governance.	Traffic Distribution
Application Service Mesh (ASM)	UCS is backed by ASM to provide non-intrusive service governance.	Service Meshes
Operator Service Center (OSC)	UCS builds a unified cloud native ecosystem through OSC to support unified distribution and deployment of cloud native applications.	OSC
SoftWare Repository for Container (SWR)	UCS is interconnected with SWR. You can use the images stored in SWR to create workloads on UCS.	Image Repositories

