

Intelligent EdgeFabric

Service Overview

Issue 01
Date 2023-11-23



Copyright © Huawei Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI 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.

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 Intelligent EdgeFabric? | 1 |
| 2 Functions | 3 |
| 3 Advantages | 6 |
| 4 Service Instance Editions | 9 |
| 5 Application Scenarios | 11 |
| 6 Basic Concepts | 13 |
| 7 Related Services | 15 |
| 8 Constraints | 16 |
| 9 Permissions Management | 19 |
| 10 Quotas | 20 |

1 What Is Intelligent EdgeFabric?

Why Intelligent EdgeFabric?

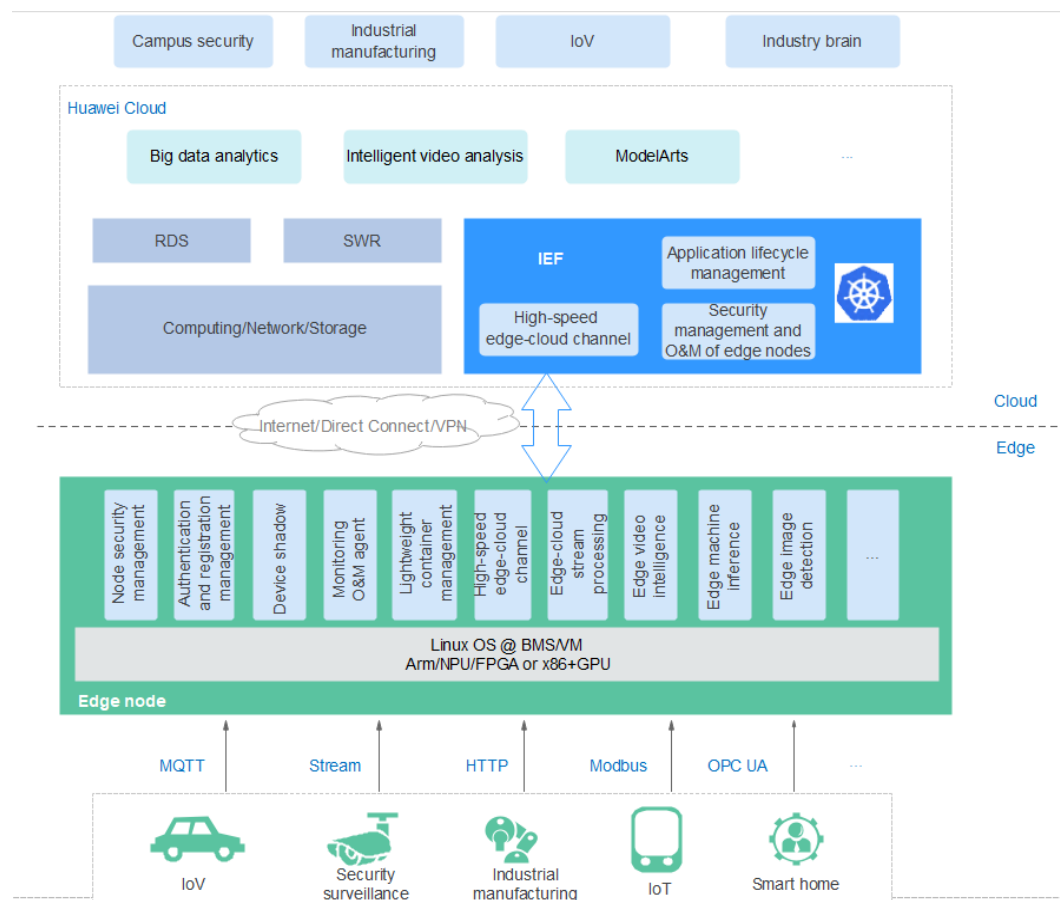
Cloud computing capabilities are centralized, which are far from devices such as cameras and sensors. It will cause long network latency, network congestion, and service quality deterioration in scenarios where high real-time computing performance is required. Furthermore, the computing capabilities of devices are insufficient and far behind those in the cloud. This is where edge computing comes in. By deploying edge nodes near devices, the computing capabilities in the cloud are extended to the edge nodes.

Intelligent EdgeFabric (IEF) provides you a complete edge computing solution, in which cloud applications are extended to the edge. By leveraging edge-cloud synergy, you can manage edge nodes and applications remotely and process data nearby, to meet your requirements for remote control, data processing, analysis, decision-making, and intelligence of edge computing resources. In addition, you can perform O&M in the cloud, including edge node monitoring, application monitoring, and log collection.

System Architecture

As shown in [Figure 1-1](#), IEF extends cloud capabilities such as AI applications to edge nodes, which are close to end devices. In this way, the edge nodes have the same capabilities as the cloud and can process device computing requirements in real time.

Figure 1-1 Edge cloud computing



Accessing IEF

IEF provides a web-based service management platform. You can access IEF through HTTPS-compliant application programming interfaces (APIs) or the management console.

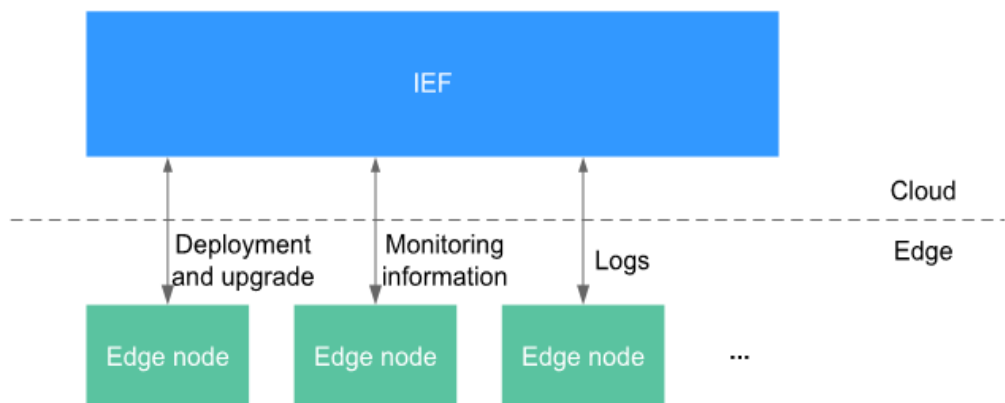
- Using APIs
Use this method if you are required to integrate IEF into a third-party system for secondary development. For detailed operations, see *Intelligent EdgeFabric API Reference*.
- Using the management console
Log in to the management console and choose **Intelligent EdgeFabric** on the homepage.

2 Functions

Edge Node Management

IEF can connect to a large number of edge nodes, automatically generate configuration information about edge nodes, and efficiently and conveniently manage edge nodes where Edge Agent is installed. In this way, all edge nodes can be managed, monitored, and maintained in the cloud.

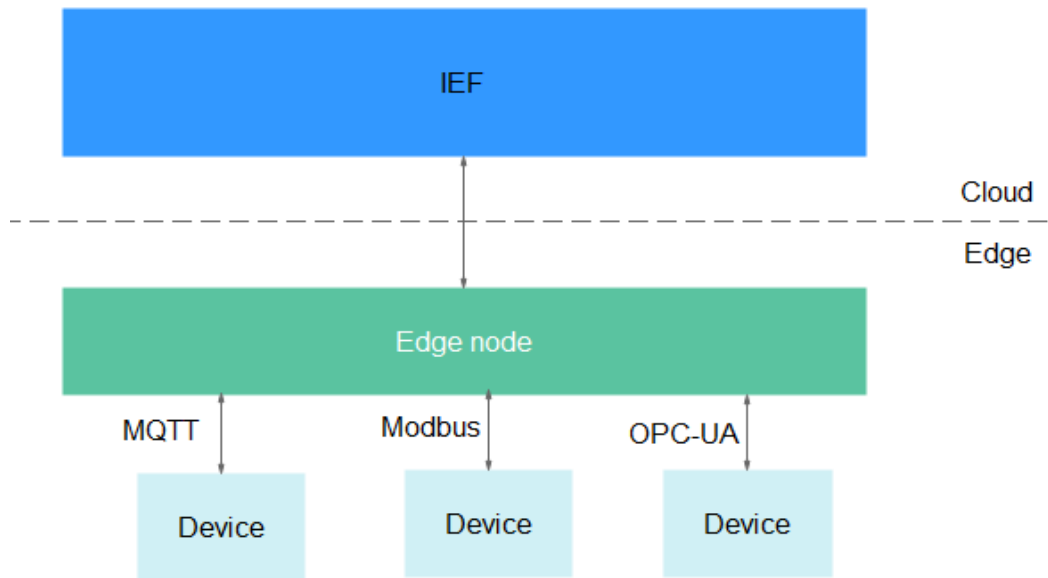
Figure 2-1 Edge node management



End Device Management

End devices can be connected to IEF through edge nodes by using the MQTT, Modbus, or OPC UA protocol. After end devices are connected to IEF, you can manage them on IEF in a unified manner.

Figure 2-2 End device management



Edge Application Management

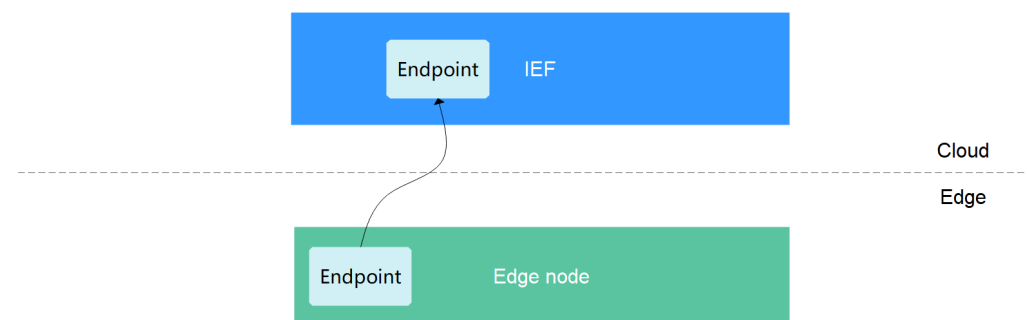
IEF allows you to deploy edge applications on edge nodes as containers. You can package your edge application into a container image, upload the image to Software Repository for Container (SWR), and use IEF to deploy the image on edge nodes. IEF also supports version upgrade, configuration change, uninstallation, monitoring, and log collection for applications.

The prosperous container ecosystem can help you seamlessly switch your containerized applications to other runtime environments and enhance their portability. In addition, containers can isolate resources better and support CPU/GPU scheduling.

Message Route Management

IEF provides the message routing function. Based on configured routes, IEF forwards edge messages to the corresponding message endpoint (sender or recipient). In this way, messages can be forwarded based on specified paths, enhancing flexibility in data routing control and improving data security.

Figure 2-3 Message forwarding paths



Batch Job Management

IEF provides the batch job management function, allowing you to register edge nodes, upgrade edge node versions, deploy applications, and upgrade applications (change container images and access configurations) in batches. For details, see [User Guide > Batch Management](#).

3 Advantages

High-Performance Intelligent Edge Hardware

IEF provides a software and hardware integrated solution, which offers users with low-cost, out-of-the-box, and centralized on-cloud O&M services. It uses Huawei general-purpose servers and AI hardware and is deeply integrated with Huawei Ascend chips to provide high-performance, low-cost edge AI inference computing power. IEF also supports TaiShan servers that use Huawei Kunpeng processors.

Figure 3-1 Edge hardware



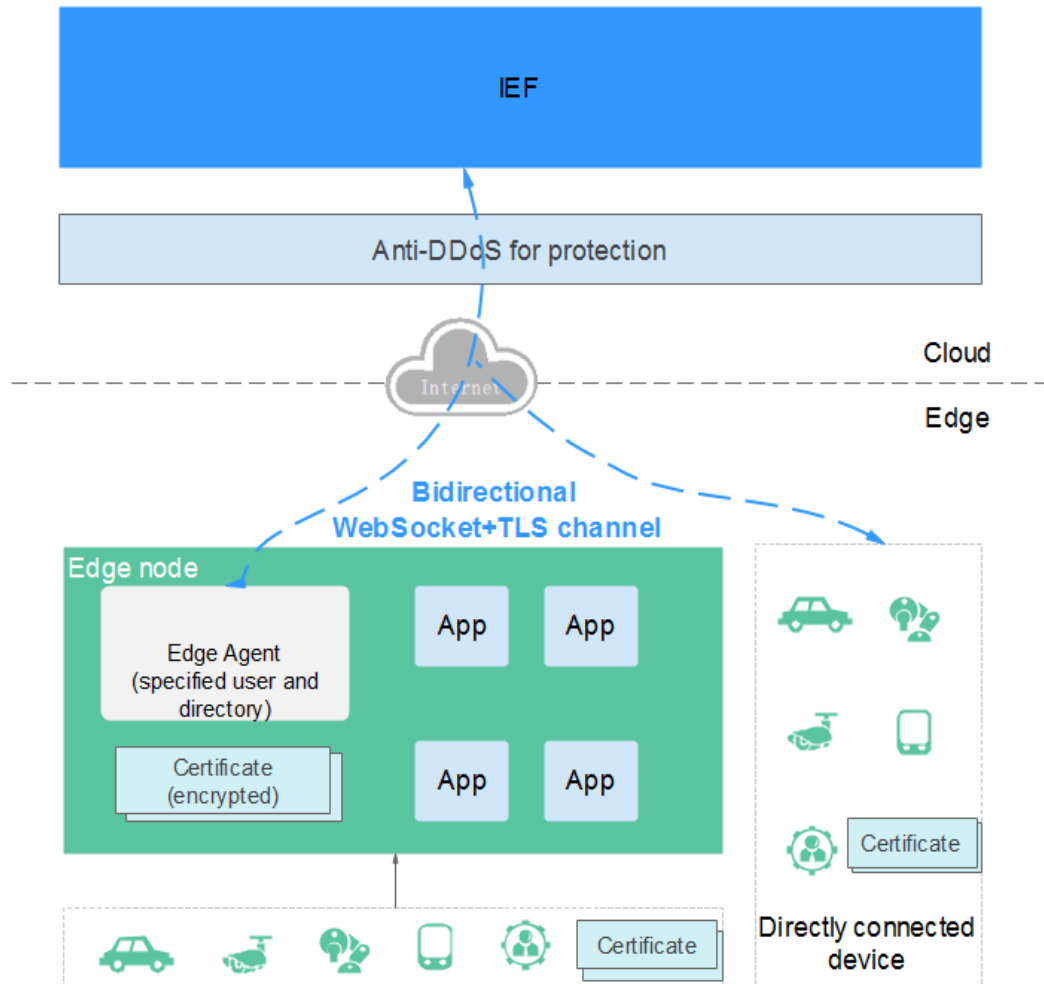
Security and Reliability

- IAM authentication
Agencies can be created in Identity and Access Management (IAM) to allow edge nodes to access resources such as Application Operations Management (AOM), Data Ingestion Service (DIS), and SoftWare Repository for Container (SWR).
- Edge node security
Edge Agent creates dedicated service users whose accessible directories and permissions are limited. Users can upload logs and monitoring information to the cloud based on their requirements.
- Edge-cloud synergy communication security
Edge Agent initiates a request to IEF for establishing a bidirectional encrypted channel. Messages exchanged between devices and IEF are authenticated and encrypted by certificates.
- Cloud security
The frontend anti-DDoS protects the cloud against malicious attacks.

A unique access certificate is issued for each edge node. Bidirectional communication is authenticated and encrypted by certificates.

- Device security
End devices use certificates for identity authentication.

Figure 3-2 IEF security solution



Open Compatibility

IEF is based on the open-source software [KubeEdge](#). Based on Kubernetes, KubeEdge provides fundamental infrastructure support for network, application deployment, and metadata synchronization between the cloud and edge.

By using KubeEdge, developers can customize and tailor the edge node runtime (Edge Agent, an edge node manager) to reduce the difficulty in using edge nodes.

Cost-effective

- The combination of cloud and edge computing implements data filtering and analysis on edge nodes, which greatly improves efficiency and reduces cloud computing costs.

- During cloud transmission, simple data processing is performed by edge nodes, so that the response time of end devices is shortened, data traffic from end devices to the cloud is decreased, and bandwidth costs are reduced.

4 Service Instance Editions

IEF offers two service instance editions for you to choose.

- Professional edition: The management plane cluster is shared by multiple users. Professional service instances allow you to manage nodes, devices, containerized applications, batch jobs, and edge-cloud messages.
- Platinum edition: Users have their own management plane clusters. Platinum service instances allow you to manage large-scale nodes and deliver higher performance. In addition to the functions provided by professional service instances, functions such as node group and application mesh are supported by platinum service instances.

Table 4-1 lists the differences between the two editions.

Table 4-1 Functions provided by the two editions

| Function | Description | Professional Edition | Platinum Edition |
|--------------------------------------|--|----------------------|------------------|
| Edge node management | Registers and manages edge nodes. | √ | √ |
| End device management | Registers end devices and binds an end device to an edge node. | √ | √ |
| Containerized application management | Delivers containerized applications to an edge node. | √ | √ |
| Edge-cloud message routing | Provides an edge-cloud message channel and supports edge-cloud message forwarding. | √ | √ |

| Function | Description | Professional Edition | Platinum Edition |
|--------------------------------|--|----------------------|------------------|
| Multi-network access | Supports access to IEF through Internet, VPN, and Direct Connect. | √ | √ |
| Monitoring and O&M | Supports monitoring and O&M. | √ | √ |
| Batch job management | Creates and updates containerized applications in batches, and registers and updates edge nodes in batches. | √ | √ |
| Edge node group | Creates edge node groups. Multiple edge nodes with the same attributes (such as the hardware architecture) can form an edge node group for unified management. | × | √ |
| Multi-instance | Supports multiple containerized application instances. | × | √ |
| Exclusive clusters | Supports exclusive management plane clusters. | × | √ |
| Application mesh | Supports service discovery and application traffic governance, such as load balancing. | × | √ |
| Add-ons | Supports add-on management. | √ | √ |
| Kubernetes native API openness | Supports operating Kubernetes clusters of a service instance through kubectl. | × | √ |

5 Application Scenarios

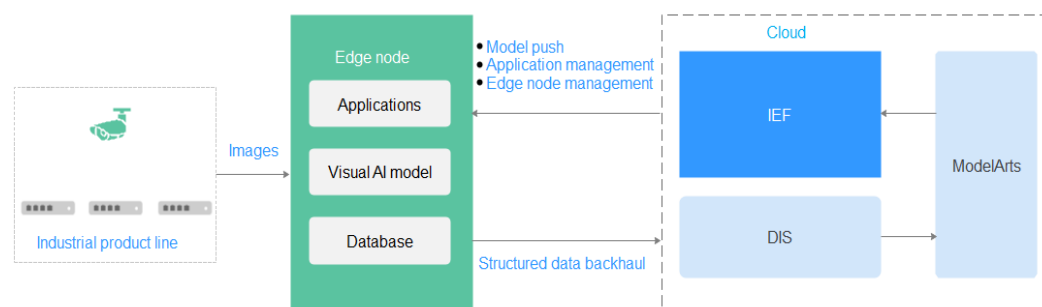
Visual Product Inspections

Traditional approaches in industrial manufacturing relied on naked eyes to detect product defects. This method was inefficient and often failed to detect flaws and even ejected products without defects from the pool of qualified products, causing losses in revenue and brand image. IEF combines cloud modeling and edge decision-making to achieve automatic visual inspection, moving away from traditional manual visual inspections.

Advantages

- **High efficiency:** Edge-side deployment of visual models trained in the cloud enables real-time product prediction, improving detection efficiency and product quality.
- **Excellent models:** Model training in the cloud, data processing at the edge, and incremental model training optimization achieve excellent models.
- **Unified control:** IEF delivers models and monitors node status in a unified manner.

Figure 5-1 Visual product inspections



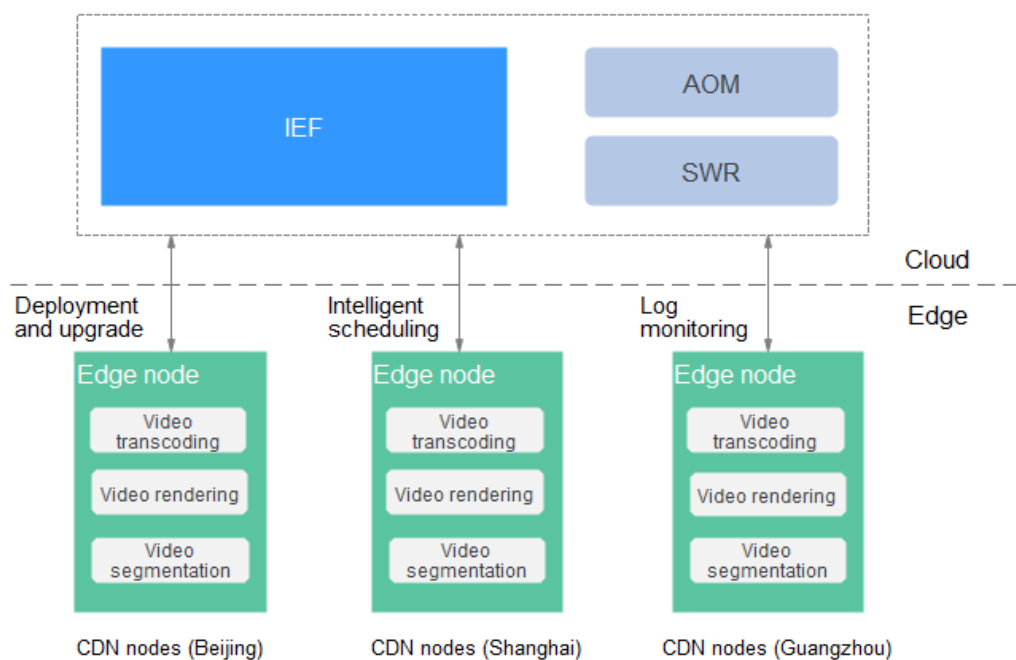
CDN Node Management

Unified management of CDN nodes deployed across the country helps users implement automatic application scheduling, auto scaling, and O&M of edge nodes and applications.

Advantages

- **Auto scaling:** IEF automatically adapts the amount of computing resources to fluctuating service load according to custom auto-scaling policies. To scale computing resources at the cluster level, IEF adds or reduces cloud servers. To scale computing resources at the workload level, IEF adds or reduces containers.
- **Intelligent scheduling:** Inter-node and inter-application affinity scheduling is supported.
- **Dimensional O&M:** Service status and edge node status are monitored in real time to ensure stable running of applications.
- **Local autonomy:** Services can run properly even when edge nodes are disconnected from cloud center networks.

Figure 5-2 CDN node management

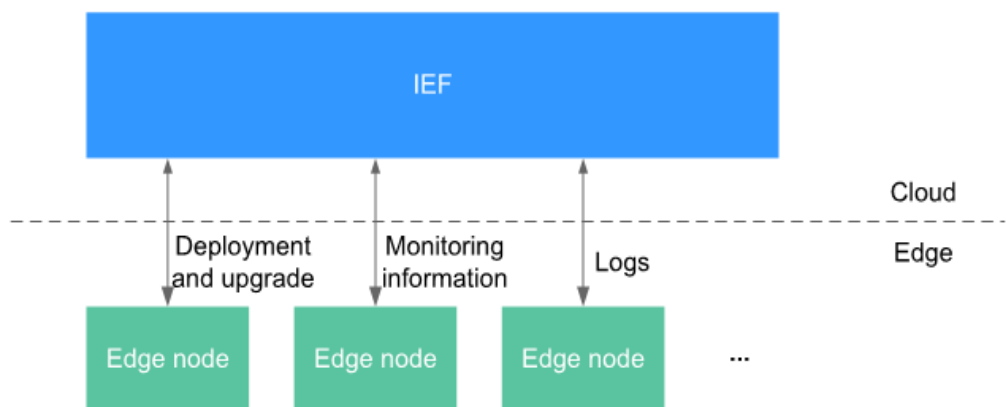


6 Basic Concepts

Edge Node

An edge computing device used to run edge applications, process data, and collaborate with cloud applications securely and conveniently.

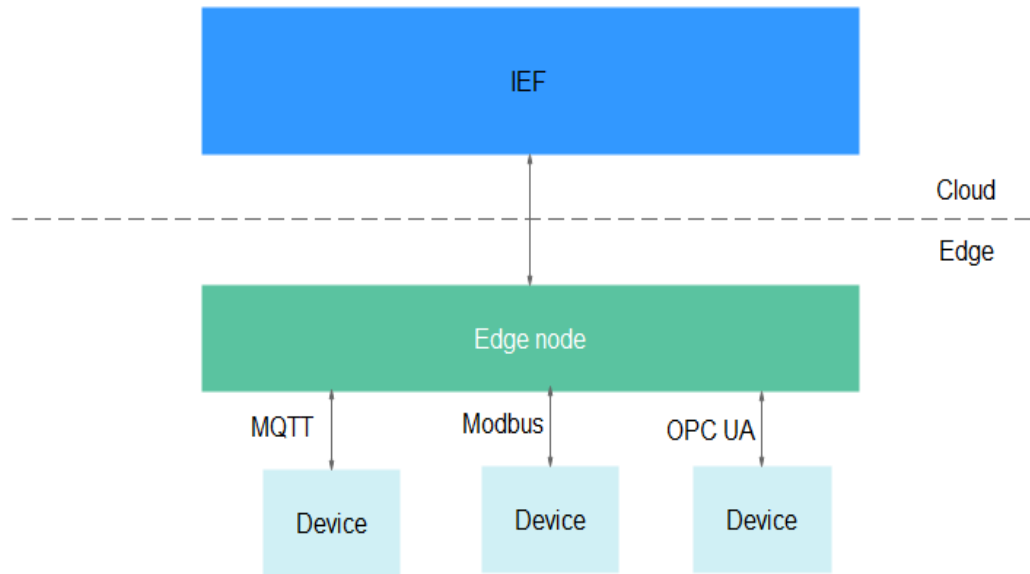
Figure 6-1 Edge node



End Device

End devices can be as small as a sensor or controller or as large as a smart camera or computer numerical control (CNC) machine tool. They can connect to IEF through edge nodes by using the MQTT, Modbus, or OPC UA protocol.

Figure 6-2 End device



Containerized Application

A functional module that runs on edge nodes. Deploying the required applications builds your own edge computing capabilities.

In the professional edition, a containerized application can have only one instance. In the platinum edition, a containerized application can have multiple instances.

Message Endpoint

Node that sends or receives data. For example, if data is sent from an end device to a cloud service, the end device is the source endpoint and the cloud service is the destination endpoint.

Message Route

A route defines source and destination endpoints and resources. The system forwards messages from the specified source endpoint resource to the specified destination endpoint resource based on the route.

Certificate

A certificate is a credential for an edge application or end device to access the MQTT broker of an edge node.

7 Related Services

Figure 7-1 shows the relationships between IEF and other services.

Figure 7-1 Relationships between IEF and other services

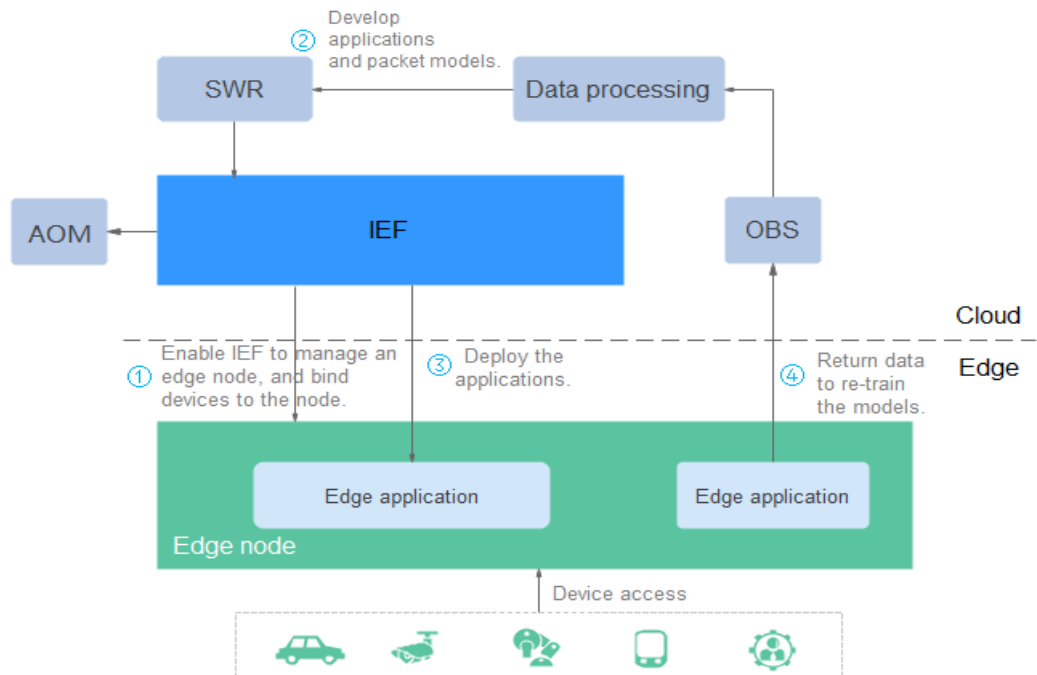


Table 7-1 Relationships between IEF and other services

| Service | Description |
|---------|---|
| SWR | Edge nodes pull container images from SWR. |
| OBS | You can upload the data generated by edge nodes to OBS for further processing on the cloud. |
| AOM | Logs, monitoring and alarms about edge nodes are reported to AOM. |

8 Constraints

Edge Node Specification Requirements

An edge node can be a physical device or a virtual machine (VM). It must meet the specifications described in [Table 8-1](#).

Table 8-1 Edge node requirements

| Item | Specifications |
|-----------|---|
| OS | <p>The language of the operating system must be English.</p> <ul style="list-style-type: none"> x86_64 architecture Ubuntu LTS (Xenial Xerus), Ubuntu LTS (Bionic Beaver), CentOS, EulerOS, RHEL, Kylin, NewStart CGS Linux, NeoKylin, openEuler, Unity Operating System (UOS), Oracle Linux (OL), Huawei Cloud Euler (HCE) Armv7i (Arm32) architecture Raspbian GNU/Linux (stretch) AArch64 (Arm64) architecture Ubuntu LTS (Bionic Beaver), CentOS, EulerOS, openEuler, Unity Operating System (UOS), Oracle Linux (OL), Huawei Cloud Euler (HCE), and openEuler 23.09 Edge <p>NOTE The openEuler 23.09 Edge operating system is recommended for edge computing scenarios.</p> |
| Memory | More than 256 MB of memory is recommended as 128 MB of memory is required to run the edge software. |
| CPU | ≥ 1 core |
| Hard disk | ≥ 1 GB |

| Item | Specifications |
|------------------|---|
| GPU (optional) | <p>The GPU models on the same edge node must be the same.</p> <p>NOTE Currently, NVIDIA Tesla GPUs such as P4, P40, and T4 are supported.</p> <p>If an edge node is equipped with GPUs, you can choose not to enable its GPUs when registering it on IEF.</p> <p>If you choose to enable GPUs of an edge node, the GPU driver has to be installed on the edge node before you can manage it on IEF.</p> <p>Currently, only x86-based GPU nodes can be managed by IEF.</p> |
| NPU (optional) | <p>Huawei Ascend AI processors</p> <p>NOTE Currently, edge nodes integrated with Huawei Ascend Processors are supported, such as Atlas 300 inference cards, and Atlas 800 inference servers. Supported NPU specifications include Ascend 310P, 310B, Ascend 310P-share, and virtualization partition NPUs..</p> <p>If you choose to enable NPUs of an edge node, ensure that the NPU driver has been installed on it. Currently, Ascend 310 supports only firmware versions 1.3.x.x and 1.32.x.x, for example, 1.3.2.B893. You can run the npu-smi info command to view your firmware version.The NPU driver version must be 22.0.4 or later. You can go to the driver path, for example, /usr/local/Ascend/driver, and run the cat version.info command to view your driver version. If the driver is not installed, contact the device manufacturer for assistance.</p> |
| Container engine | <p>The Docker version must be later than 17.06. If Docker 1.23 or later is used, set the docker cgroupfs version to 1. Docker HTTP API v2 is not supported.</p> <p>(However, Docker 18.09.0 is not recommended as it has a serious bug. For details, see https://github.com/docker/for-linux/issues/543. If this version has been installed, upgrade it at the earliest possible opportunity.)</p> <p>NOTICE After Docker is installed, configure the Docker process to start at host startup. This configuration prevents system exceptions caused by Docker startup failures after the host is restarted.</p> <p>Docker Cgroup Driver must be set to cgroupfs. For details, see How Do I Set Docker Cgroup Driver After Installing Docker on an Edge Node?.</p> |
| Glibc | <p>The Glibc version must be later than 2.17.</p> |
| Port | <p>Edge nodes require port 8883, which is the listening port of the built-in MQTT broker on edge nodes. Ensure that this port works properly.</p> |

| Item | Specifications |
|----------------------|---|
| Time synchronization | The time on an edge node must be consistent with the UTC time. Otherwise, the monitoring data and logs of the edge node may be inaccurate. You can select an NTP server for time synchronization. For details, see How Do I Synchronize Time with the NTP Server? |

MQTT Usage Constraints

Table 8-2 MQTT usage constraints

| Description | Constraint |
|---|---|
| Supported MQTT version | 3.1.1 |
| Differences from the standard MQTT protocol | <ul style="list-style-type: none"> • Quality of Service (QoS) 0 is supported. • Topic customization is supported. • QoS 1 and QoS 2 are not supported. • will and retain msg are not supported. |
| MQTTS security levels | TCP channel basic + TLS protocol (TLS v1.2) |

9 Permissions Management

IEF Permissions

IEF does not support fine-grained permission management using Identity and Access Management (IAM). You are advised to log in to IEF using an account. If you need to log in to IEF as an IAM user, assign permissions defined in the **Tenant Administrator** policy to the IAM user.

- If your Huawei Cloud account does not need individual IAM users, then you may skip over this chapter.
- IAM users need to be granted the **Tenant Administrator** policy. This policy has system administrator permissions and has the operation permissions on all cloud services except IAM.

To grant IEF permissions to a user, grant the **Tenant Administrator** permissions to a user group and add the user to the user group so that the user has the corresponding permissions. For details, see [Creating a User and Granting Permissions](#).

Helpful Links

IAM is free of charge. You pay only for the resources in your account. For more information about IAM, see [What Is IAM?](#).

10 Quotas

IEF restricts the maximum amount of resources that a user can use. For details about quotas, see [Table 10-1](#) and [Table 10-2](#).

Table 10-1 Resource quotas for the professional edition

| Resource Object | Default Quota |
|------------------------------|--|
| Node | 10 |
| End device | 500 |
| End device template | 10 |
| Containerized application | 500 |
| Application template | 10 |
| Application template version | 10 |
| Tag | A maximum of 20 tags for each resource |
| ConfigMap | 50 |
| Secret | 50 |
| Encryption data | 50 |
| Message endpoint | 20 |
| Message route | 100 |
| Node registration job | 50 |
| Node certificate | 5000 |

| Resource Object | Default Quota |
|-----------------|---------------|
| Batch job | 20 |

Table 10-2 Quotas for the platinum edition

| Resource Object | Default Quota |
|------------------------------|--|
| Service instance | 5 |
| Node | 50, 200, or 1,000, which can be selected when you create a platinum service instance |
| End device | Number of nodes x 50 |
| End device template | Same as the number of nodes |
| Containerized application | Number of nodes x 50 |
| Application template | Same as the number of nodes |
| Application template version | 10 |
| Tag | A maximum of 20 tags for each resource |
| ConfigMap | Number of nodes x 5 |
| Secret | Number of nodes x 5 |
| Encryption data | 50 |
| Message endpoint | 20 |
| Message route | 100 |
| Node registration job | 50 |
| Node certificate | Number of nodes x 500 |
| Node group | 100 |
| Node group certificate | Number of nodes x 50 |
| Service | 500 |
| Gateway | 500 |
| Virtual service | 500 |
| Batch job | 20 |