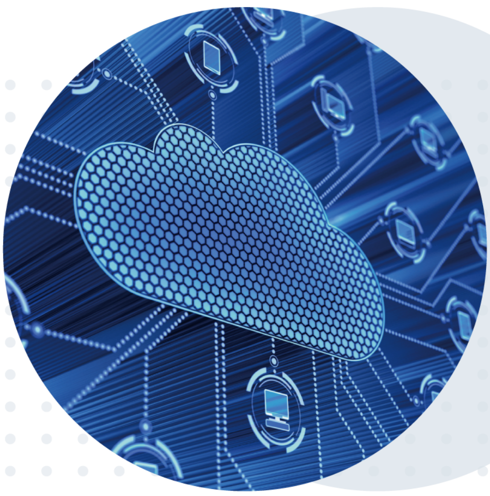
北京 2022 年冬奥会官方合作伙伴
Official Partner of the Olympic Winter Games Beijing 2022

White Paper for China Unicom's Edge-Cloud

Service Platform Architecture and Industrial Eco-System

Issued by China Unicom



February 2018

Contents

White Paper for China Unicom's Edge-Cloud Platform Architecture and Industrial Eco-System

1 Overview

1.1 Vision of the White Paper

As the information technology deeply changes our daily life, digital reshaping will penetrate into all industries. The communication industry has undergone several leaps and reforms during the past decades. From dial-up Internet access to FTTH and from GSM/TD-SCDMA/Wimax to LTE, information pipes have undergone radical changes, and the IT cloudification technology is reshaping CT and OT industries. The architecture of both the wireless network and wired network has been evolving into C/U separation and Fixed Mobile Convergence (FMC) to cope with all challenges brought by industrial digital transformation and industrial structure upgrade. In the predictable future, the integration at the infrastructure layer is imperative.

With the combination of increasingly mature SDN/NFV, big data, and artificial intelligence, 5G networks will become key infrastructure in the digital transformation of all industries. In 5G-oriented IoE scenarios, new types of services have the characteristics of a lower latency, larger bandwidth, and more intelligence. The traditional vertical type network architecture has many deficiencies in resource sharing, agile innovation, flexible expansion, and simple O&M. To effectively meet the service requrements of eMBB, mMTC, and uRLLC and enhance industrial competitiveness, global telecom operators successfively begin network reconstruction and transformation, aiming to construct a DC-centered full cloudification network.

The multi-access edge computing technology is the result of ICT integration, and is also a key technology that allows telecom operators to pursue 5G-oriented network transformation. This meets the development requirements of HD video, VR/AR, industrial IoT, and IoV in the future. Operators also regard tens of thousands of edge DCs as the best high-quality resources compared with OTT, gaining a wide application space for edge computing. As the official communication service partner for the 2022 Winter Olympics, China Unicom has already established an Edge-Cloud strategic partnership with BAT, and has begun the planning and construction of thousands of edge DCs. China Unicom is committed to building an open and open-source Edge-Cloud Service PaaS Platform and providing rich platform service capabilities and unified APIs for application developers, aiming to accelerate the incubation and commercial use of innovative services related to edge applications.

As the pilot project that is constructed based on the edge ecology incubation and edge service platform, China Unicom works with ZTE, Intel, and Tencent video to build an Edge DC testbed in Jingjin University Town, Tianjin. The deployment tests of edge vCDNs, edge transcoding, and edge intelligent analysis are already completed. For the edge vCDN, the deployment of applications to the network edge not only reduces the bandwidth pressure and latencies (compare to the traditional CDN mode, the average RTT latency is reduced by 50 percent, and the HTTP download rate is increased by 43 percent) but also significantly enhances user experience and helps content providers reduce costs. This pilot project blazes a trail in the edge computing collaboration between China Unicom and OTT, and is of great guidance value for the reconstruction of the edge DCs and the commercial incubation of 5G innovative services.

This white paper describes the architecture and evolution roadmaps of China Unicom's Edge-Cloud platform in accordance with 5G FMC requirements and network cloudification evolution. It also detailedly describes China Unicom's key service scenarios and business cases that are based on the Edge-Cloud platform, with a combination of the standardization progress of the edge computing technology and the status of the industrial chain.

1.2 Version of the White Paper

In June 2017, China Unicom issued the *White Paper for China Unicom's Edge Computing Technology* at Shanghai MWC. This white paper further extends the Edge-Cloud platform architecture, evolution planning, the advancement of commercial use, and ecosystem construction, and hopes to be helpful for the development of the whole industrial chain. With the pre-commercial deployment of 5G networks, freezing of ETSI and 3GPP standards, more contents will be added to the future version, and any comment or suggestion is welcomed.

2 Service Demand and Network Evolution Trend

Throughout the history of telecommunications, every industry reform, from wired to wireless, from voice to video, is driven by service demand. In addition, the reform in an industry may be driven by the technological development of other industries. For example, the photography industry began with film but now is reshaped by digital images. The development of edge computing is driven by both of the following factors: service demand and cloud-enabled evolution.

2.1 Service Demand

With the popularization of 4G and optical fiber communication networks, people have become accustomed to the convenience brought by broadband services and abundant applications, and desire more realistic and interactive experience. At the same time, the rapid development of the Internet of Things (IoT) makes the originally isolated equipment interconnected with each other and improves efficiency. Therefore, the future information society will be bound to present diverse demand. Following the trend of the time, the 3GPP defines three scenarios based on the research on 5G mobile networks, namely Enhance Mobile Broadband (eMBB), Massive Machine Type of Communication (mMTC), and Ultra Reliable Low Latency Communication (uRLLC).

- 5G eMBB provides users with ultra HD videos and immersive service experience (such as VR and AR) through a transmission rate of 10 Gbps.

- The mMTC technology supports the access and interconnection of massive devices typically in smart cities and smart buildings by connecting millions of devices per square kilometer.

- By virtue of the ultra-low latency and ultra-high reliability, uRLLC greatly improves the operation efficiency of the applications in vertical industries, such as the Internet of Vehicles and industrial Internet.

Meanwhile, the development of fixed-network services also shows a similar 5G trend. For example, the 8K video, 3D video, industrial control, and government and enterprise private cloud services are evolving toward high bandwidth, low latency, and massive connection. According to the development trends of wireless and fixed networks, operators will face the following service demand challenges:

- The eMBB services, such as 8K videos, 3D videos, and VR/AR, will create a demand for ultra-high network bandwidth (hundreds of Gbps), thus causing great transmission load on backhaul networks. The investment in capacity expansion of convergence networks and Metropolitan Area Networks (MANs) can significantly improve the transmission costs per media stream, but cannot generate investment income.

- The URLLC services, typically V2X and industrial Internet, require an end-to-end ultra-low latency of 1 ms. This cannot be fulfilled just through the technology progress in the physical layer and transport layer of the wireless and fixed networks. Therefore, it is necessary to introduce innovative deployment of network and industry applications according to the characteristics of vertical industries.

- The mMTC services, such as smart city and smart building, will produce a large amount of data, leading to huge challenges to operation management. Practices show that only the centralized cloud cannot afford such a complex system. Instead, localized intelligent control and management must be introduced.

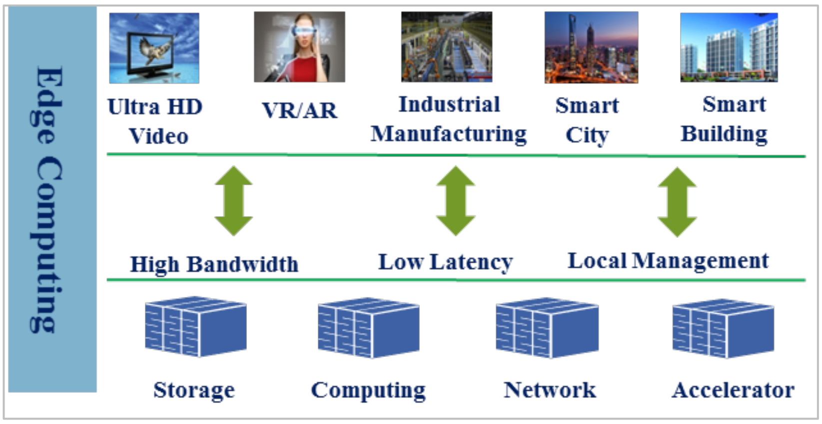


Figure 2.1 Edge Computing Supports Fixed-Mobile Service Development

From the perspective of operators, it is a great challenge to fully support the service scenarios above without decreasing investment income. As shown in Figure 2.1, we believe that effectively use of storage, computing, network, and acceleration resources provided by the edge service platform can satisfy the performance requirements for the development of future 5G and fixed broadband services, which sinks the edge of some key service applications to the edge of access networks. This can reduce thebandwidth and latency loss caused by network transmission and multi-level forwarding, and decrease the CAPEX and OPEX through the localized management and operation on the unified edge computing platform.

2.2 Network Cloudification Evolution

In order to satisfy the development requirement of future services in the high bandwidth, low latency, and local management, the control plane is separated from the forwarding plane in the 3GPP 4G CUPS and 5G New Core, thus implementing a flat network architecture. The forwarding plane gateway is migrated to the wireless side, distributed as required, and scheduled by the control plane in a centralized manner. Combined with edge computing, gateway anchors can support end-to-end massive services with low latency, high bandwidth, and balanced load, fundamentally resolving the overload and latency bottlenecks in transmission and core networks caused by unidirectional service flow in traditional mobile networks.

Coincidentally, the key NEs, such as HGU, OLT and BRAS, in fixed networks also evolve with the control plane separated from the forwarding plane to provide flexible on-demand network services for governments, enterprises, and households. The control plane is deployed in a centralized manner, while the forward plane is simplified and standardized, and managed in a unified way by SDN controllers through NC/Yang, realizing flexible and efficient service orchestration. The CORD project, led by AT&T and the Linux foundation, is a classic example. It is a major challenge in the evolution of fixed networks how to further improve and optimize the system architecture based on the existing achievements in the CORD project. And it has become a big topic of fixed network evolution that puts forward a more practical, efficient, fast, and intensive network evolution solution with a view to the service demand for edge communication cloud and edge service cloud of wired and wireless networks.

As shown in figure 2.2, with the evolution of the network architecture that the control plane separated from the forwarding plane, traditional closed proprietary equipment systems are gradually disintegrating, and transforming into cloud-based open systems based on universal hardware and SDN/NFV open system. They carry telecommunication network functions based on virtual machines and containers, orchestrate services and resources in a unified way through MANO and cloud management, and construct the integrated DevOps capability to significantly reduce the time to come into market of the new services, implementing in-depth convergence of IT and CT and enhancing the competitiveness of operators in services. Therefore, it is an inevitable choice to reconstruct telecommunication infrastructure based on cloud technologies from central cloud to edge cloud.

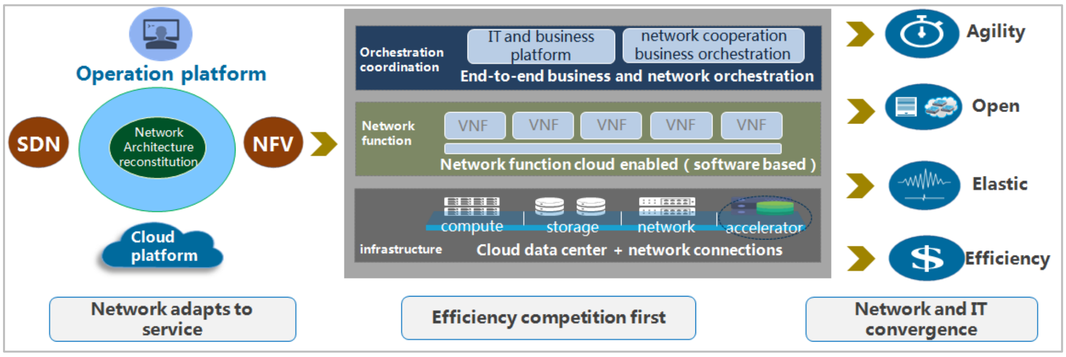


Figure 2.2 Trend of Network Cloudification Evolution of China Unicom

As a product of ICT convergence, the edge computing–based platform can support network function virtualization deployment, such as Central Unit, UPF, vSGW-U/vPGW-U, vCPE, vBRAS, and vOLT. In addition, operators can make the storage, computing, network, and security capabilities of the platform open up to third-party application developers and content providers, and provide OTT applications with unified edge deployment and management. Further, the edge computing platform can abstract telecom network capability information into a variety of services (for example, RNIS, location services, bandwidth management, and TCP optimization), and make them open up to third-party applications and vertical industries to improve their service performance and competitiveness among their peers.

The cloud-based edge computing platform allows operators to flexibly deploy network functions and support infrastructure for fixed-mobile network convergence, and the hosting and infrastructure service capabilities of the platform itself can help third-party applications improve user experience, maximizing the value of both applications and networks. This ensures that operators realize the digital transformation from pipeline builders to eco-builders.

3 Edge-Cloud Platform Architecture and Evolution

3.1 Overall Layout of Communication Cloud

Communication cloud is the network infrastructure that is constructed in a unified way based on the existing network deployment and operation experience of operators and uses SDN/NFV/cloud computing as the core technology to support the cloud-based evolution of networks and match network transformation deployment. In the future, all the NEs in 5G networks will be deployed based on cloud and NFV. To follow the trend, China Unicom plans to transform traditional end offices into data centers based on the Central Office Re-architected as a Data Center (CORD). The network architecture in the future uses DC-centered 3-level communication cloud–based DC layout, and deploys and builds edge, local and regional DCs at different levels to plan the communication cloud resource pool in a unified way, achieving unified access bearers and services for fixed networks, mobile networks, IoT, and enterprise dedicated lines. In a cloud-based network, the lateral architecture still consists of the access networks in traditional communication systems, MANs, and backbone networks, while the vertical architecture is hierarchical, composed of transmission bearing, IP, NFV, and network management and orchestration.

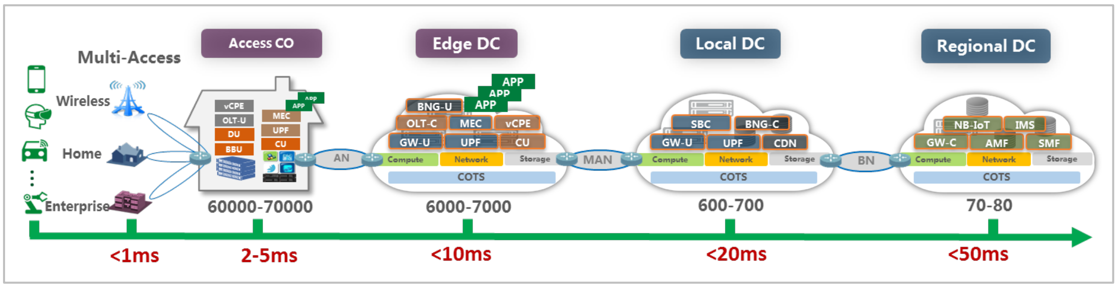


Figure 3.1 Cloud-Based Network Architecture of China Unicom Communication Cloud

As shown in Figure 3.1, the communication cloud–based network of China Unicom is generally deployed at four layers, including three layers of DCs and one layer of access CO. The features are as follows:

1) **Regional DC:** There are about 70 to 80 regional DCs across China with the end-to-end delay less than 50 ms, which are used for the services and control plane NEs (such as group OSS/NFVOs, provincial cloud pipe platforms, NFVOs, and VNFMs) of

the whole country, large regions, and provinces. Regional DCs mainly serve for the control NEs at the provincial level and the group level, including the IMS, GW-C, AMF, SMF, MME, and NB-IoT core network NEs. Regional DCs are an important part of communication cloud management. They mainly manage NEs, networks, and infrastructure in a unified way. Regional DCs also realize the collaboration between NFVOs and OSSs. An NFVO provides orchestration, resource management, and fault management for virtual networks. An OSS coordinates services and resources, and manages traditional networks. NFVOs work with OSSs for unified management of traditional networks and virtual networks.

2) **Local DC:** There are about 600 to 700 local DCs across China with the end-to-end delay less than 20 ms, which are deployed in prefecture-level cities and key county-level cities of a province. Local DCs mainly serve for the control plane NEs in MANs, centralized media plane NEs, services of local networks, control plane NEs, and some user plane NEs, including the CDN, SBC, BGN-C, UPF, and GW-U NEs.

3) **Edge DC:** There are about 6,000 to 7,000 edge DCs across China with the end-to-end delay less than 10 ms. The target of edge DC–based transformation is converging equipment rooms. Edge DCs serve for access layer NEs and edge computing NEs. They terminate and forward media streams. With the characteristics of a low latency and high bandwidth, the future 5G RAN-CU, MEC, BNG-U, OLT-U, and UPF NEs can be flexibly deployed at edge DCs, and provide location perception and wireless network information for network edge users. Edge DCs allow cloud service environment, computing, storage, network, and acceleration resources to be deployed on network edges, realizing closer combination of applications and networks and providing more network resources and services for users.

4) **Access CO:** There are about 60,000 to 70,000 access COs across China with the end-to-end delay between 2 ms and 5 ms. Access COs achieve unified access and bearing for the public, governments and enterprises, and users of fixed and mobile networks to make resources more intensive and provide ultimate experience for users. Access COs are widely distributed in villages and towns. The scenarios with a high requirement for the delay and bandwidth, such as 5G CUs, DUs, MECs, and vCPEs, can be deployed in access COs as required or directly deployed in the existing equipment rooms. It is difficult to transform an access CO because an access CO mainly accommodates access and traffic forwarding equipment and the condition in an access

CO is worse than that in an edge DC. Therefore, the DC-based transformation of the unified infrastructure for access COs are not considered at present (province-level branch China Unicom can transform the access COs according to the actual service needs).

3.2 Edge-Cloud Platform Architecture

The construction of the Edge-Cloud platform involves the infrastructure layer, virtualization layer, edge computing platform layer, and upper-layer APPs. China Unicom is committed to building an open and open-source edge services Edge-Cloud platform, aiming to provide rich network capabilities, services and unified APIs for developers, see Figure 3.2.

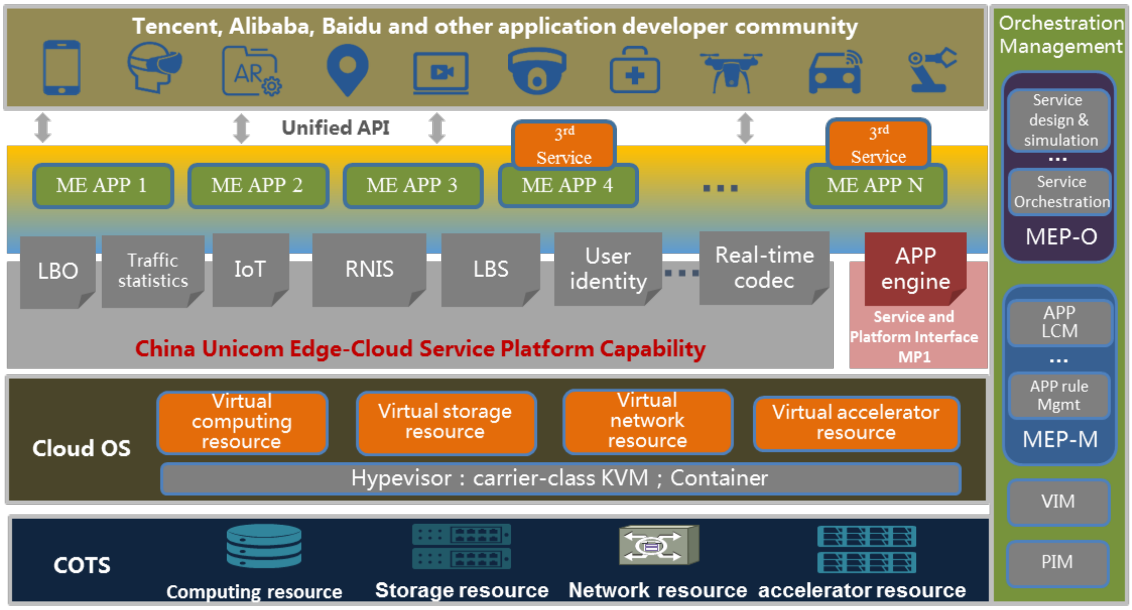


Figure 3.2 China Unicom's Edge- Cloud Platform Architecture

3.2.1 Infrastructure Architecture

The business feature of the Edge-Cloud edge service platform determines that the platform must be close to terminals and users in the network to provide the best experience. Therefore, in edge service platform deployment, the positions of the access CO and edge DC are key points requiring attention in the communication cloud architecture.

- The edge DC implements traffic convergence for the access network and leads traffic to the local DC. The IDC can be a reference in guiding the DC reconstruction of the infrastructure level, and similar environmental conditions of the IDC should be implemented. The DC should have specified operating space, bearing capacity, power supply capacity, and heat dissipation capability, and should support deployment of general COTS devices. It is acceptable to use customized servers provided by device manufacturers to satisfy requirements for many factors including the limited deployment space and power supply.

- The access CO is relatively closer to terminals and users, aiming to increase resource intensity and achieve perfect user experience. In most cases, access devices and traffic forwarding devices are deployed in the access CO, and there is not an urgent demand for DC reconstruction. The edge-service platform can be directly settled based on the actual equipment room condition. In this case, general COTS devices are not suitable for deployment due to the size, power consumption, and weight.

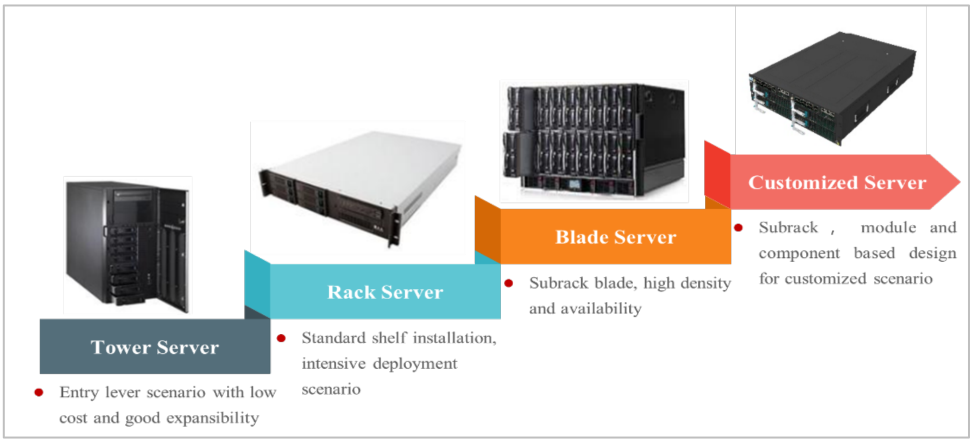


Figure 3.3 Server Evolution

In the IT field, the COTS server has evolved through several stages, including tower server, rack server, blade server, and customized server, see Figure 3.3. The tower server is mainly oriented towards scenarios with entry-level low-cost applications scenarios. Based on the tower server, the rack server has higher computing density and good extensibility through industrial standardization. The blade server not only has higher computing density but also owns the advantages in costs, deployment, and serviceability through sharing fans, power supplies, and networks among subracks and providing built-in network solutions. With the development of cloud computing and large-scale DCs, the above server types can no longer satisfy the operators' demands in costs, performance, and energy efficiency. Therefore, server customization, including the multi-node form and integrated cabinet form, has become the focus of Internet companies, aiming to reduce costs and share resources to the maximum extent.

In the CT field, with the development of the NFV technology, the traditional private hardware platform of device manufacturers has also evolved towards the open general COTS server. However, the general COTS server still has problems in several aspects, including the supports of telecom feature, installation environment restrictions, and capability of system architecture evolution, mainly reflected in:

- The design philosophy of rack servers may cause a waste of edge DC equipment room resources. For example, for computing-intensive services (such as 5G CU), the design of using more local hard disks and IO extensions may cause a waste of space resources and costs.

- The blade server requires a great installation height and power consumption over 5 kW, posing a considerable challenge in power supply, heat dissipation, and bearing capacity. It may cause that the edge DC equipment rooms are difficult to reconstructed or require huge costs.

- The depths of existing cabinets differ from the required cabinet depth. There are a lot of 600–800 mm cabinets in the existing edge DC equipment rooms. However, the general COTS devices should be at least 1 m deep.

- The bearing capacity requirement for equipment rooms is high. In most cases, the bearing capacity of the IDC reaches 600 kg/m2, but the bearing capacity of edge DC equipment rooms is far less than the value. Thus an additional constraint is added on the deployment.

- I/O and resource acceleration are not designed in accordance with the NUMA Balance technology, causing difference in performance and delay among cores.

- The management interfaces, protocols, and data formats on servers of different vendors differ from each other. Remote fault locating, diagnosis, and troubleshooting capabilities are not adequate, unable to satisfy the management requirements for equipment rooms in unattended mode.

- The BIOS, FW, and OS configurations of servers of different vendors differ from each other, causing the VNF unable to be migrated seamlessly and deployed dynamically.

- The specific –48 V power supply for the communication industry and the high-voltage DC power supply possibly used in the future are not compatible with the AC mainly used for general COTS devices.

Therefore, the edge-service platform has specific restrictions in consideration of the hardware architecture, including the position, space, power supply, and service scenario. It requires customized development and configurations. For example, the following capabilities should be considered: data plane implementation and extension (virtualized, unified and open API), rapid traffic offloading (intelligent network cards), real-time codec capability (GPU), rapid image recognition (GPU/FPGA), and even artificial intelligence. Both are important drivers of personalized openness and configuration. Local computing, storage, I/O balancing, hardware acceleration, high integration, and device Energy Efficiency Ratio (EER) are key factors that should be considered in the future. Servers are customized designed in accordance with edge application environments and edge service requirements to achieve the best solution. The following factors should be considered in customization of general servers:

- General X86 architecture.
- Modular design to achieve flexible function extension on demand, for example, hardware acceleration to satisfy service performance requirements.
- Flexible configurations in accordance with the requirements for computing, storage, network, and hardware acceleration.
- Built-in network to provide better O&M experience through simplified network management and topology architecture.
- Support for telecom-class features such as electromagnetic compatibility and environmental adaptability. Considering various deployment positions, the edge server must possess adaptability to extreme environments, for example, the server must be able to operate in the 45°C environment for a certain period of time.
- Matching existing condition of equipment rooms and minimize equipment room reconstruction.
- Proper physical specifications, including subrack height (3–4U. A great height may cause heavy weight and large costs while a small height may reduce the gain brought by common part sharing), subrack depth (800 mm cabinet installation supported), and front maintenance mode (better O&M experience).

Behind the most general concept of COTS device, the real core requirement of customers is decoupling of software from hardware. It means that a customer can flexibly deploy services without being restricted by hardware device manufacturers. From this perspective, the customized general servers based on the X86 architecture can be classified to the range of COTS devices. In addition to the features of decoupling software from hardware, virtualization, and flexible service deployment, customization of general servers also involves scenario-based device appearance and device architecture, aiming to achieve higher energy efficiency and better cost-effectiveness. Therefore, the customized general server is comparatively the best hardware choice for the edge DC and access CO.

3.2.2 Virtualized Layer Architecture

VIM refers to the virtual resource management platform which manages the NFVIs in a domain. In the hierarchical DC architecture, each edge DC, local DC, and regional DC are all respectively corresponds to an NFVI, and all the NFVIs are managed by the VIM. China Unicom recommends to use that the OpenStack, enhanced VIM based on open-source Openstack, or VIM based on the Kubernetes container technology.

Compared with regional DCs and local DCs, edge DCs have different scales and carry various types of services, and thus have flexible deployment modes. Reconstruction of an edge DC should be implemented in stages. In the initial stage, virtualized reconstruction can be implemented at some hot spots with good deployment conditions. For large-scale edge DCs, each edge DC is deployed with Openstack. For small-scale edge DCs, deploying Openstack for each edge DC may cause a huge waste of resources. Thus, it is recommended that compact Openstack deployment or distributed OpenStack deployment can be used locally. In local compact Openstack deployment mode, resources can be saved by deploying the control node and computing node as one node and decouples some components to the actual requirements. In distributed Openstack deployment mode, only computing modes are deployed in edge DCs with fewer nodes, control nodes are deployed in regional DCs or edge DCs with more nodes, and all edge DCs are managed in a unified manner.

The container technology provides better elasticity response speed, system capacity flexibility, and computing resource utilization. Considering the development of new services and technology evolution trend, the virtualization layer should support the container technology to construct distributed container clouds. The container technology supports two deployment modes: container deploys in a physical machine and container deploys in a virtual machine. In container deployment scenarios, the Cloud Container Engine should be deployed on the VIM to provide highly reliable and high-performance management services for enterprise container applications, support native applications and tools of the Kubernetes community, and simplify the construction of automation container operating environment on the cloud.

The low delay and high-speed forwarding features of the edge computing service require that the virtualization layer should have the hardware acceleration capability. The VIM provides unified interfaces to adapt to various accelerator types. Accelerators are abstracted and managed in a virtualized manner as computing devices, storage devices, and networks. Accelerators are presented as acceleration resources and provide comprehensive acceleration services. With the development of the NFV standard, hardware acceleration has become increasingly important. The standardization of functional interfaces of accelerator resources is being gradually conducted. Openstack has started the Cyborg project, aiming to provide a general hardware acceleration management framework. The acceleration hardware includes the encryption card, intelligent network adapter, GPU, and FPGA. China Unicom is actively promoting relevant research.

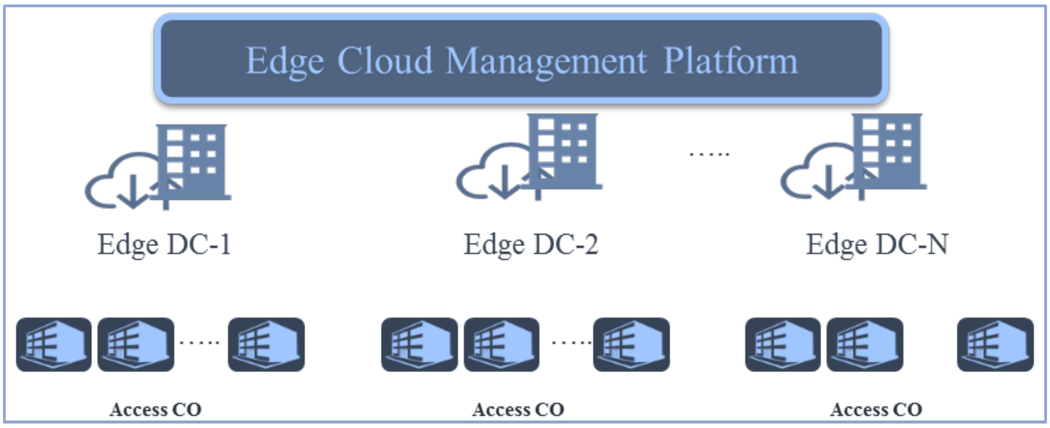


Figure 3.4 Edge Cloud Management Platform

In ETSI NFV MANO specification, the VIM executes the virtual resource management only. However, the edge cloud requires an integrated platform having the VIM, PIM, and service management capabilities, which is known as the Cloud Management Platform (CMP). For edge clouds, the CMP should support hierarchical architecture, for example, the edge cloud management platform distributed in provinces and cities and the unified cloud management platform of the group. As shown in Figure 3.4, an edge DC and its access COs comprise an edge cloud, and an edge cloud management platform manages one or multiple edge clouds. The edge cloud management platform can be deployed either in the edge DC or in the regional DC in accordance with the actual condition. It is logically interconnected with multiple VIMs, and provides access point convergence, unified management of heterogeneous resources, and position-based resource scheduling based on VIM virtualized management.

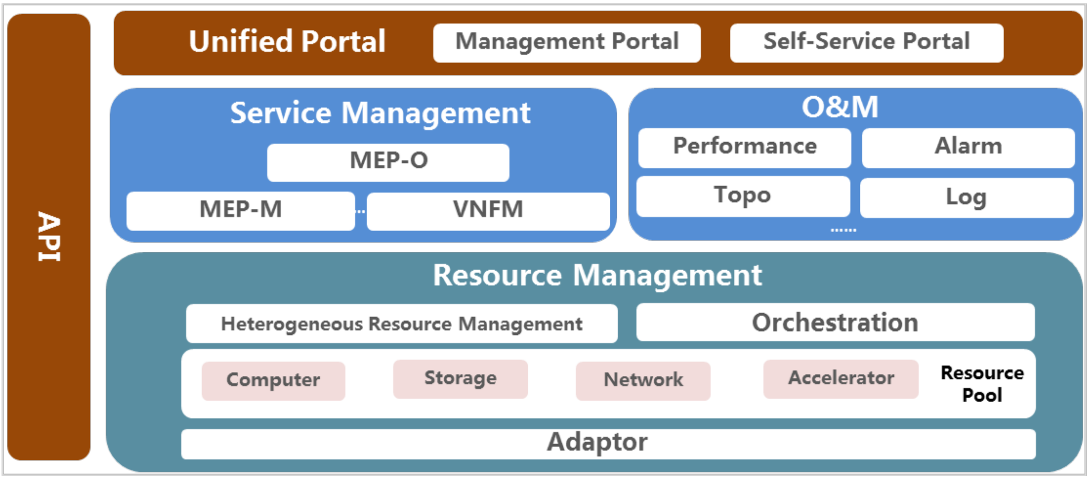


Figure 3.5 Edge Cloud Management Platform Functional Diagram

As shown in Figure 3.5, the edge cloud management platform provides functions including cloud resource management, O&M management, edge cloud service management, and capability openness. The cloud resource management connector (adaptation layer) accesses the VIM (OpenStack, K8S) of the edge DC through an plug-in or agent to form a geographically distributed but logically unified cloud resource pool (computing resources, network resources, storage resources, and acceleration resources) to implement unified management and scheduling of heterogeneous resources. The O&M module implements unified monitoring, alarm management, performance management, and log management of resources. At the early stage of edge cloud computing construction, the ecological chain has not been formed. To implement rapid validation of new services and put the services into operation quickly, the NEs (for example, MEPO, MEPM, and VNFM) in the MANO framework coordinating with NFVO to implement VNF and third-party vApp lifecycle management can be packaged and delivered in a unified manner as the components of the edge cloud management platform. After the edge cloud ecological chain becomes mature, these components can be split off from the edge cloud management platform and be integrated into the MANO framework. The edge cloud management platform is interconnected with the unified cloud management platform and NFVO through the unified capability openness layer. The unified portal provides a unified GUI for cloud resource presentation and operation. It is used for unified management of edge cloud resources and provisioning and maintenance of local edge cloud services.

3.2.3 Service Orchestration and Management Architecture

In the overall communication cloud layout of China Unicom, regional MANO and the unified cloud management platform are deployed for regional DCs to implement communication cloud infrastructure management. MANO management includes:

- NFVO: Based on the authorization of the unified cloud management platform, this module manages and orchestrates authorized resources in a unified manner, including definitions, collaborative scheduling, and lifecycle management of resources and services, to put the services into operation quickly. Cross-regional services are orchestrated by the group NFVO.

- VNFM: This module is responsible for lifecycle management for VNF NEs, including creation, modification, deletion, and elasticity expansion of VNF NEs.

The traditional MANO can be used to manage ETSI-NFV VNFs only. On the Edge-Cloud service platform, not only VNFs but also third-party edge applications (Edge-APPs) are deployed. Thus, Edge-APPs should also be orchestrated and managed. The Edge-APP requirements fall into two categories:

1) Application package requirement

- The following information should be included: software image (or image link), image format (for example, virtual machine, container), application description (requirements and rules which applications depend on or prefer to), provider signature (including the abstract and public key of the signature and relevant certificates).

- The files in an application package can be signed individually. For each signed file, the corresponding public key, algorithm, and certificate are stored in a position known to all in the application package.

2) Application description requirement

- Includes the minimum computing resources (for example, quantity, features, and functions of virtual machines), virtual storage resources, and virtual network resources required by the application.

- Includes the (additional) service list and (additional) feature list required for the application operation.

- Supports descriptions of traffic rules, DNS rules, and delays descriptions for Edge-APPs.

The requirements for computing, storage, and network resources are consistent with the VNF requirements, and can be parsed and processed through traditional MANO. However, the requirements for description of delays, traffic rules, and DNS rules cannot be parsed and processed through traditional MANO. Considering the special requirements for third-party Edge-APPs, China Unicom extends the traditional MANO management domain, finally realizes the management of the edge service platforms and Edge-APPs.

As shown in Figure 3.6, based on NFVO and VNFM, two logical function modules are added: MEP-O and MEP-M.

- MEP-O: This module collaborates with NFVO to implement version package management (including loading, querying, enabling, disabling, and deleting version packages) and application lifecycle management. MEP-O supports parsing and directly processing descriptions related to Edge-APPs, for example, it can select a proper edge service platform to configure features for the MEP-M. For common descriptions, MEP-O interacts with NFVO and uses NFVO to orchestrate resources, and orchestrates NFV network services for a group of Edge-APPs.

- MEP-M: This module manages NEs on the edge service platform and Edge-APP rules and requirements (for example, traffic rules and DNS rules), and implements lifecycle management for the edge service platform and Edge-APPs.

The unified Cloud Management Platform (CMP) manages infrastructure layers of multiple DCs within its management areas, divides the resource pool and manages authorities, and monitors various types of information of virtualized resources and non-virtualized resources, including topology, alarms, performance, and capacity. It also provides reports related to whole resources, and helps O&M personnel evaluate potential risks based on AI algorithms to avoid or eliminate problems in time. Considering that Edge-Cloud is mainly oriented to vertical industries, CMP can provide service portals oriented to third parties to support MEP capability openness , Edge-APP lifecycle management and Devops, promoting the construction of edge content ecology.

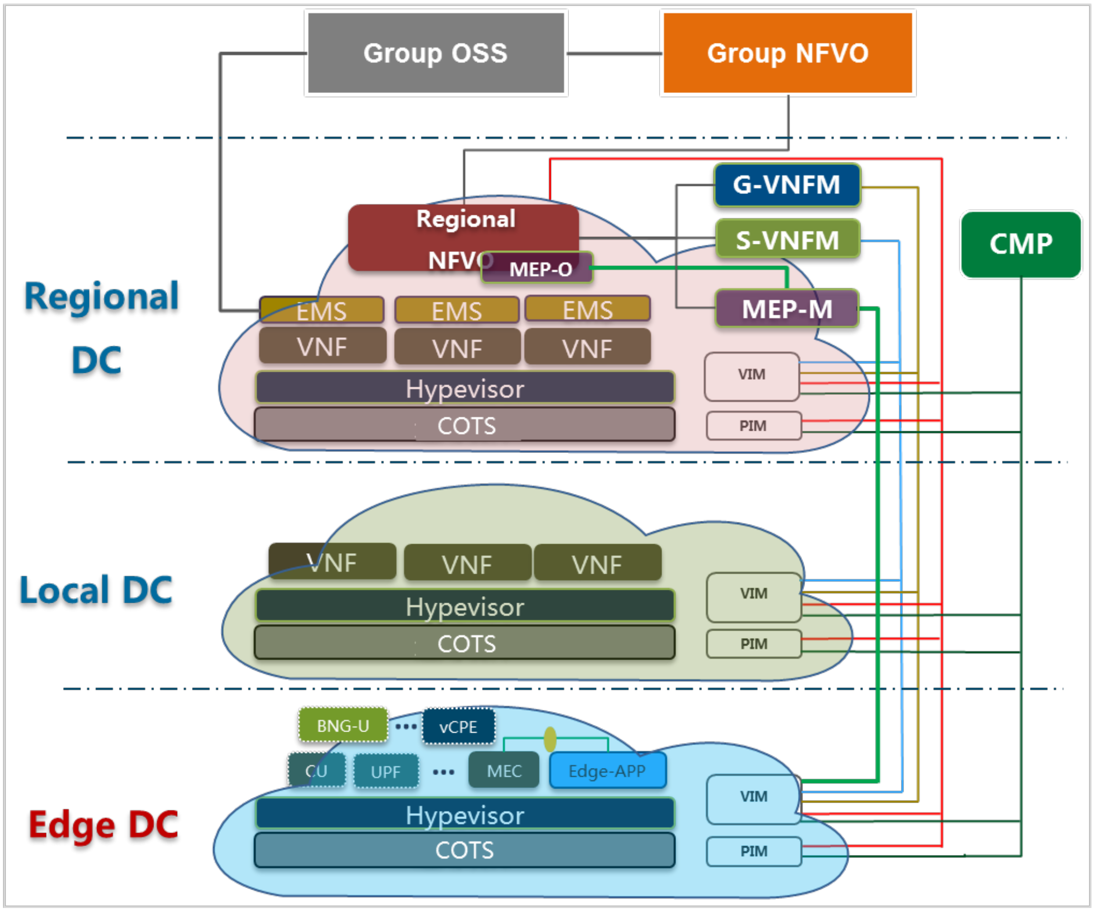


Figure 3.6 Edge-Cloud Platform and MANO Architecture of China Unicom

As defined in the ETSI GS IFA 009, NFVO has two functional modules. The RO module is responsible for cross-VIM NFVI resource orchestration while the NSO module is responsible for NS LCM. From the perspective of long-term ICT evolution, CMP, RO of NFVO, and VNFM can be integrated into a functional entity to provide a unified Automation & Orchestration platform for IT and CT applications.

3.2.4 Edge-Cloud Platform Capabilities

China Unicom's Edge-Cloud platform not only provides various basic services but also provides a unified capability openness framework by formulating MEC-Enabled API specifications on the basis of ETSI standards. This platform serves and manages third-party applications. Figure 3.7 shows the capability openness framework of China Unicom Edge-Cloud platform. The service capabilities of basic networks can be provided to third-party applications in a safe and efficient manner, and can be reliably shared between third-party applications. This promotes rich edge application ecology and meets various edge application scenarios.

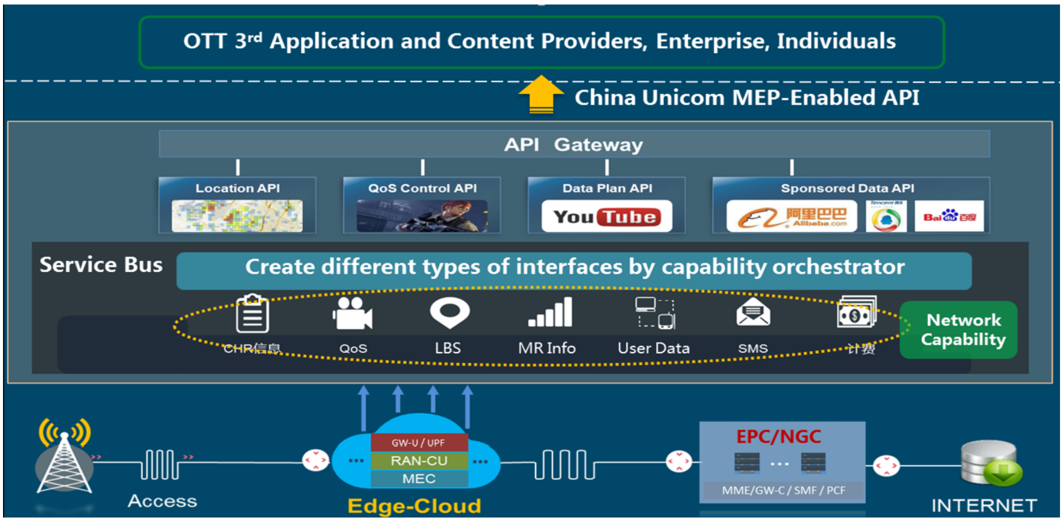


Figure 3.7 Capability Openness Framework of the Edge-Cloud platform

3.2.4.1 Edge-APP Management Capability

The APPs deployed on the Edge-Cloud platform can obtain necessary services, and can also provide services for other users. To achieve the above purposes, the Edge-Cloud platform needs to support APP management. The ESTI-MEC specification defines the interface between the Edge-Cloud platform and APPs as the Mp1 interface. Based on the Mp1 interface specified by ETSI-MEC, China Unicom defines a unified APP management framework, which provides the following functions for third-party applications:

- Service registration and discovery
- Service termination
- Subscription and notification of service-related events
- Service API and access control
- Authentication and authorization of service APIs
- Statistics and logs of service APIs
- Flow control rule configuration

- DNS rule configuration

A third-party application can obtain the capabilities provided by the platform and provide the capabilities for other applications through the platform. The descriptions of specific capabilities and related service APIs should be provided by the third-party application developer.

3.2.4.2 Location Service Capability

User location information is a key factor to associate online services with real scenarios, especially in large-sized super markets, railway stations, and airports with dense population and high traffic, where the requirements of accurate sales, online payment, mobile advertising, indoor location service, operation innovation, and service innovation are growing fast. The Edge-Cloud based indoor location solution uses the fingerprint database matching location algorithm based on uplink SRS measurement results. The location procedure is divided into two phases:

- Offline fingerprint database collection: collects the SRS fingerprint data in the location area and stores it in the Edge-Cloud. The data includes the coordinates of each indoor position and the signal strength array, which are one-to-one matched.

- Online location: Queries the fingerprint database to find K similar samples in accordance with the real-time RRU measurement data of each UE to be located, and calculates the real-time position coordinates of the UE in accordance with the k-Nearest Neighbor (KNN) algorithm.

An indoor eNodeB reports the UE position coordinates and UE ID (for example, UE IP) through the API defined by the Edge-Cloud platform. This platform can either provide the user location service directly or provide a customized interactive service by using the real-time IDs, services, and behaviors of authorized UEs in the mobile network.

3.2.4.3 Traffic Breakout Capability

The breakout capability based on the Edge-Cloud platform can shunt some services to the local network, for example, a campus network, or to the applications deployed on the edge service platform to reduce the pressure on transmission bandwidth, decrease delay, and improve user experience of the low-latency and high-bandwidth services. For the entire architecture of local breakout, see Figure 3.8. The procedure is described as follows:

- In the uplink, the Edge-Cloud platform transparently transports the data that does not need to be processed in the local network to the mobile core network, and decapsulates the GTP headers of the data that needs to be processed and shunt packets to the local network.

- In the downlink, the Edge-Cloud platform transparently transports the data that is not generated by a non-local network, for example, generated by Internet, to eNodeBs. Similarly encapsulates the data that is generated by the local network into S1-U packets, and sends them to eNodeBs.

Local breakout rules include breakout by domain name, address, and user ID. Meanwhile, the network will be a converged network containing 4G, 5G, and WiFi, and the Edge-Cloud platform needs to support network convergence breakout.

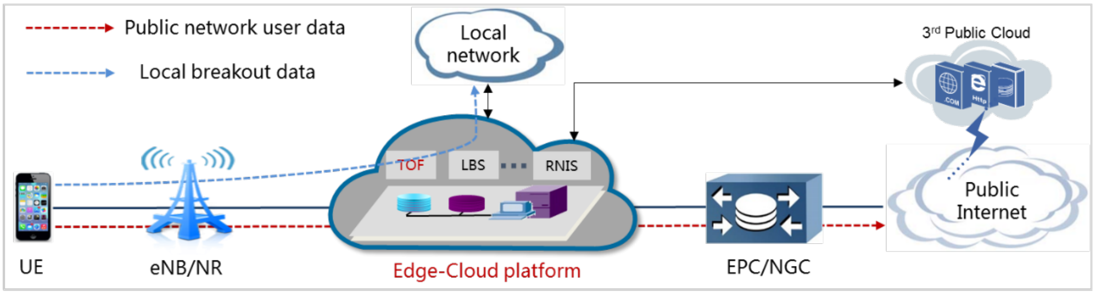


Figure 3.8 Diagram of Service Breakout Based on the Edge-Cloud Platform

3.2.4.4 Real-Time Codec Service Capability

To support media-level services, for example, local live show, local video surveillance, and video cloud games, the edge service platform needs to support the real-time codec service capability.

1) **Live show:** The camera collects audio and video signals and pushes them to the media server deployed on the edge service platform. To adapt to more video playing terminals, the media server needs to support real-time transcoding and video frame re-encapsulation.

2) **VR games:** To provide perfect and shocked experience, VR games need CPUs and GPUs with a high computing rate and large computing cost, which prevent VR games from being popular. If VR cloud games are deployed on network edge through

the Edge-Cloud platform, the strong computing capability on pictures and real-time media of the Edge-Cloud platform can accelerate VR game popularization.

3) **Video communication:** In a video call, especially a multi-party call, each data stream needs to be coded, decoded, and synthesized on the UE and cloud. Deploying the codec capability on the edge service platform can solve both the codec synchronization problem caused by inconsistent UE capabilities and the bottleneck problem caused by centralized coding, decoding, and synthesis. The distributed video communication coding and decoding capability based on the edge service platform, combined with the UE capability and cloud capability, can allocate network resources in a reasonable manner and provide better user experience.

3.2.4.5 Intelligent Analysis Service Capability

Now the video surveillance service becomes one of the major services of Smart City. Video surveillance generates lots of backhaul traffic, while most pictures are still or worthless. Now the video surveillance service uploads all video streams to the back-end server for processing, with high cost and low efficiency. Deploying edge intelligence capability on the edge service platform to implement intelligent video analysis can either return the changed pictures only or send structural data processed by intelligent analysis to the DC, saving the transmission resources of the backhaul network. To avoid high latency, bad experience, and high backhaul bandwidth cost due to lots of alternative service routes, the low-value surveillance contents can be saved on the local server. In addition, distributed video storage based on the edge service platform can reduce the bottleneck problem caused by centralized storage.

3.2.4.6 Radio Network Information Service Capability

Deployed on network edge, the Edge-Cloud platform facilitates the awareness of real-time radio network information. The Radio Network Information Service (RNIS) includes the real-time radio network conditions, measurement and statistics information on the user plane, for example, CQI, SINR, and BLER, and UE bearer information, for example, UE context and RAB.

The RNIS provides open interfaces in the format of API to third-party applications through the Edge-Cloud platform, helping the applications optimize service procedures, improving user experience, and implementing deep convergence of the network and services. Based on the radio network information awareness capability provided by the Edge-Cloud platform, third-party applications can obtain differentiated network services in accordance with their requirements, at last increasing the customer satisfaction level.

3.2.4.7 User Identification Service Capability

The edge service platform authenticates third-party applications, and accepts the API requests from only the authorized third-party applications and forwards the requests to the internal services of the Edge-Cloud platform. Authorized Edge-Cloud application instances can activate or deactivate the associated configuration rules by user ID. By means of using the identification service, third-party applications can map external application IDs to the internal IDs in the mobile network to implement data operations for specific users.

3.2.4.8 Bandwidth Management Service Capability

After being deployed on the edge computation platform, various applications require different network resources, for example, the bandwidth. The bandwidth management service can set bandwidth parameters on different applications, for example, bandwidth, priority, and allocation policy of static or dynamic bandwidth, and manage the parameters in accordance with the service policies of the third-party applications in a unified management.

3.2.4.9 QoS Service Capability

With the increase of video-type services, QoS on the radio access network side will be increasingly focused. And more and more users, including individual users and enterprise users, want to obtain different user experience. Research on how to schedule air-interface resources for HD video services becomes a new direction. In the high-user-density spots such as airports, railway stations, and campuses, the Edge-Cloud platform can provide different QoS services for users acquiring the same type of services by detecting real-time data, for example, the user requirement class, service type, and network situation. For the video QoS optimization diagram, see Figure 3.9.

The Edge-Cloud platform can identify the service type and user type in accordance with user data packets, and provides QoS information for eNodeBs in accordance with the analysis results. eNodeBs provide different radio bandwidth and latency guarantee for the data packets of different users. By integrating OTT service information and radio access network information, the Edge-Cloud platform can use the advantages of intelligent channels to guarantee QoS of key users and services.

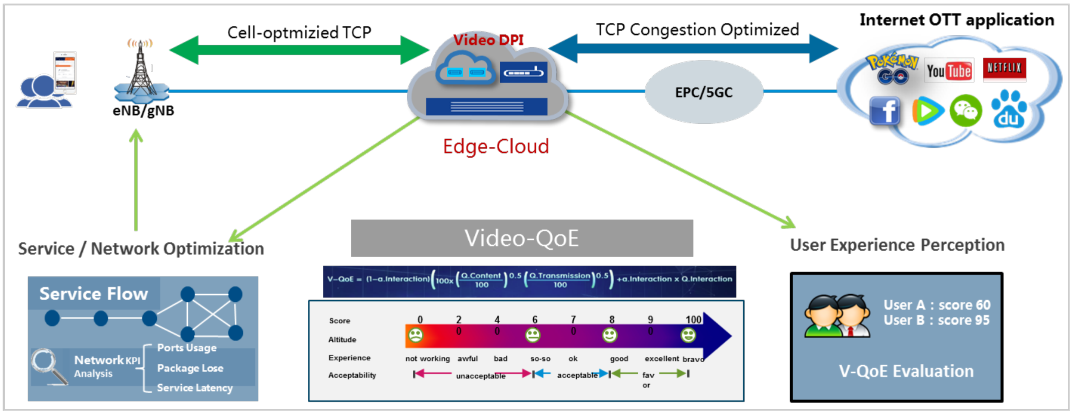


Figure 3.9 Video QoS Optimization Diagram

3.2.4.10 Procedure Collection and Charging Service Capability

In traditional networks, the core network NEs implement charging function. Due to the location change of the Edge-Cloud platform and deployment of third-party applications, the charging service has new requirements.

- Charging requirements related to APP providers: charging on the operation procedure of third-party applications in accordance with the factors related to third-party applications, for example, the number of times that an APP service is used and service duration. Third-party applications reports the generated charging information to the third-party charging center to which they belong.

- Charging requirements related to operators: charging on third-party applications in accordance with statistics information, for example, the number of times that an APP is on boarding, number of times that an APP is instantiated, resource consumption data, number of times that an APP is migrated, number of times that a service is used, and service duration. The charging data generated by the Edge-Cloud should be sent to the charging center of the operator.

On account of 5G specifications are not frozen yet, the charging capability will be developed in multiple ways, for example, the specification, technology, and engineering. The detailed solution is described as follows:

The UPF implements charging on users. The UPF forwards 5G user data to the Edge-Cloud platform, implements charging based on the charging policy, guarantees security, and reports charging information to the charging center. Accurate charging on users is supported either before or after the Edge-Cloud is deployed.

The Edge-Cloud platform implements charging on APPs/APIs. The Edge-Cloud needs to monitor the events that can be charged, generates charging records in accordance with the events, and reports the records to the charging gateway. The events that can be charged are classified into three types:

- Resource occupation events: provide the number of occupied virtualized resources and/or the usage duration.

- Management and orchestration events: The Edge-Cloud management elements invoke the API to implement APP management functions, for example, creating or deleting APPs, and migrating APPs.

- Capability openness API invoking events: APPs invoke the API and use the services provided by the Edge-Cloud platform or network.

In addition, the Edge-Cloud platform needs to support online charging and offline charging.

- Offline charging: The Edge-Cloud periodically collects and reports the resource occupation data and management/orchestration data to the offline charging function module of the network. The charging system generates charging data of consumers when a bill period ends.

- Online charging: The Edge-Cloud monitors resource occupation. When the quota is being used up, the Edge-Cloud reports resource occupation to the online charging function module and requests a new quota. If your account balance is insufficient, the Edge-Cloud rejects the quota request, and suspends the resources being used.

3.2.4.11 IoT Service Capability

If all IoT data is processed on the cloud, network traffic and application instantaneity become new challenges. For the applications that require extra-high instantaneity, for example, detection, control, and execution applications for industry systems, and the multimedia applications that generate massive data, the analysis and decision functions of some data can be deployed on network edge. For the applications that require high privacy, the IoT capability can be moved to the network edge or campus to satisfy the security requirements of specific users.

Performing data analysis and pre-processing on the Edge-Cloud can efficiently deal with data explosion and reduce network traffic. The Edge-Cloud can decrease the device response time and reduce data from devices to cloud data centers to efficiently allocate network resources. The Edge-Cloud can be deployed inside a campus to ensure data privacy. The Edge-Cloud platform supports the following IoT capabilities:

- Multi-access capability: allows UEs and gateways to access the network through multiple protocols, including TCP, UDP, MQTT, and COAP, and supports multiple types of UEs, for example, LORA, ZIGGEE, and NB-IoT.
- Device management: supports device status management, map and location management, and storage, display, and distribution of the data reported by devices.
- Data analysis: supports analysis and pre-processing of the data reported by devices.
- Rule management: supports operations on the data reported by devices in accordance with the user-defined rules, for example, generating alarms, and operating the device or other related devices.
- Coordination management: supports coordination with the higher-layer cloud platform, and implements convergence, storage, and long-term analysis of important data on high-layer cloud platform nodes.

With the service development and network evolution, the Edge-Cloud platform, based on the capabilities mentioned above, will bear more capabilities and applications to accelerate fixed-mobile convergence.

3.3 Evolution Roadmap of the Edge-Cloud Platform

Figure 3.10 shows the evolution planning of China Unicom's Edge-Cloud platform.

1) Large-scale pilot projects (September 2017-June 2018)

China Unicom began large-scale pilot projects from September 2017. By the end of June 2018, such projects will be constructed in China's 15 provinces: Beijing, Tianjin, Shanghai, Sichuan, Zhejiang, Chongqing, Guangdong, Henan, Liaoning, Fujian, Jiangsu, Hubei, Hebei, Hunan, and Shandong. The application scenarios include smart campuses, smart venues, smart parks, industrial Internet, and IoV.

During this stage, the Edge-Cloud platform can implement software&hardware decoupling (COTS&Cloud OS decoupling). The Cloud OS and MEP manufactured by the same vendor are deployed. It is recommended that the Edge-APP and MEP can be deployed onto the same Cloud OS and the MEP Mp1 rule be followed. The ETSI MEC ISG has not strictly defined API standards, and there are great differences among platforms manufactured by different vendors. It is not needed to unify the standards of northbound API interfaces between the Edge-APP (manufactured by the same vendor as the MEP) and MEP. China Unicom has begun the formulation of unified northbound API standards that will be completed in June 2018.

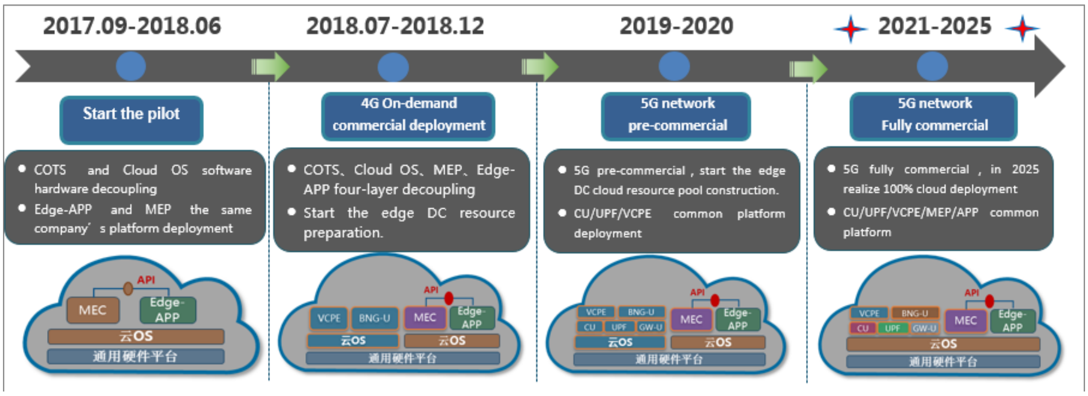


Figure 3.10 Evolution Planning of China Unicom's Edge-Cloud Platform

2) On-demand commercial deployment of 4G (from July 2018 to December 2018)

In accordance with the early-stage Edge-Cloud pilot network construction and service demonstration effect, China Unicom will replicate and promote the success stories of the incubated edge services and implement commercial construction step by step. During this stage, the Edge-Cloud platform will implement 4-level decoupling between the COTS, Cloud OS, MEP, and Edge-APP, and embed the VCPE and BNG-U functions by using the same COTS and different Cloud OSs.

On the Edge-Cloud platform, third-party application developers develop the Edge-APP in accordance with China Unicom's unified API specifications, and are not restricted by the platforms manufactured by different vendors, thus greatly shortening the service launching low. In addition, China Unicom will begin the preparation of the equipment rooms of edge DCs in multiple key provinces (including 15 provinces where pilot projects are located).

3) Pre-commercial construction of 5G networks (2019-2020)

China is actively promoting 5G research and industrial development, and raises the goal of “Become a Pioneer in 5G”. With the rich Edge-Cloud pilot project and commercial construction experience accumulated in early stages, China Unicom will start the large-scale construction of edge-DC cloud resource pools in 2019 to boost the pre-commercial use of 5G networks.

During a short period, the 5G CU, UPF, VCPE, and BNG-U manufactured by the same vendors are deployed in the same Cloud OS, smoothly implementing VNF&Cloud OS decoupling. The northbound API capabilities between the MEP and Edge-APP will become richer (4-level decoupling is still kept). Considering the differences between the CT domain and IT domain, the deployment across different Cloud OSs is allowable. Considering the scenario where 4G and 5G networks will coexist for a long time, the Edge-Cloud platform supports the simultaneous connection of 4G and 5G base stations and traffic unloading.

4) Large-scale commercial use of 5G networks (2021-2025)

During this stage, 5G networks will be rapidly constructed and put into large-scale commercial use. At the same time, the cloudification architecture of China Unicom’s communication networks will be shaped. It is estimated that full cloudification deployment will be achieved in 2025. China Unicom's Edge-Cloud platform supports the deployment of the CU, UPF, GW-U, VCPE, BNG-U, MEC, and Edge-APP manufactured by different vendors, completely implementing 4-level decoupling. The Edge Cloud-oriented various services are more prosperous, and business modes are more mature. China Unicom will build more sophisticated edge ecology and implement service layout with the characteristics of a standard leader, full commercial use, self-control, openness and open source.

China Unicom will gradually implement the seamless integration of the business/enterprise private Ethernet, FTTP access network for family users, and FTTC access network for mobile users in accordance with actual service requirements. The vision is to provide a unified Edge-Cloud platform for FMC users.

4 Edge-Cloud Standardization and Industrial Chain

4.1 Standardization Progress

In 2014, the ETSI first initiated the MEC standard project that aimed to build a service platform with cloud-based computing and an IT-based environment at the edge of mobile networks for application developers and content providers, and to open wireless network information through the platform, supporting and locally managing high-bandwidth and low-latency services. Some leading vendors and operators, such as Vodafone, ZTE, Intel, Huawei, Nokia, and HP are members of the alliance. Phase I of the project is completed, which is deployed based on traditional 4G network architecture and defines the service scenarios, MEC system reference architecture, MEC platform application enablement APIs, framework of the system management and operation, application lifecycle and management and wireless service capability APIs (e.g. RNIS, location, and bandwidth management). To satisfy the requirements for commercial MEC deployment and fixed-mobile convergence, ongoing Phase II of the project focuses on multi-access edge computing systems, including 5G, WiFi, and fixed networks, mainly involving the MEC in NFV reference architecture, end-to-end MEC mobility, network slicing support, legal interception, container-based application deployment, V2X support, and WiFi and fixed-network capability exposure.

With the extension of ETSI MEC influence, the 3GPP is also engaged in the support research on edge computing. The 4G CUPS and 5G New Core introduce the separation of the control plane from the forwarding plane. The forwarding plane is deployed on the edge of wireless networks in a distributed way, and the control plane deploys and controls the forwarding plane in a centralized manner, ensuring the on-demand local distribution of services. The SA2 5G system architecture fully supports edge computing in terms of local routing and service manipulation, session and service continuity, network capability exposure, QoS and Policy Control. In addition, the research on SA5 network functions management of next generation network architecture and features, SA6 Common API Framework for 3GPP Northbound APIs will further consider the demand for edge computing. The 3GPP is accelerating the process of commercial edge computing as an effective supplement to ETSI MEC.

Following the fixed-mobile convergence trend, the Broadband Forum is converging wired and wireless networks by using the unified core network and network slicing technology. The definition of the unified interface between an access network and a core network enables 5G New Core to support the convergence of wired and wireless services. This further extends the enabling capability of the multi-access edge computing platform and support edge computing services for fixed-mobile convergence.

In addition to the efforts of the international organizations, China also makes contributions. In 2017, the China Communications Standards Association (CCSA) launched an edge computing research project. As a part of the project, China Unicom initiated and led the research on the capability openness technology of the 5G edge computing platform, which studies on capability exposure for edge computing in 5G. The solutions will be based on the architecture of the edge computing platform and the capability of mobile networks to standardize the framework and content for network capability exposure. In January 2018, an international standard project, "IoT Requirements for Edge Computing", led by China Unicom, was successfully initiated at the plenary meeting of ITU-T SG20 WP1, which was the first time that an edge computing project is initiated by the ITU-T in the field of industrial Internet. China Unicom will lead the international standardization efforts of ITU-T IoT edge computing to promote health and sustainable development of the edge computing industry.

However, edge computing is an ecosystem that includes applications, platforms, and networks, so standard organizations cannot promote the rapid development of the industry alone. Therefore, it is necessary for an edge computing industry alliance, for example, OpenFog Consortium, to fully explore industry service scenarios, integrate industry resources through operators, and combine application requirements with edge computing platform standards, to promote the development of the edge computing industry together.

4.2 Situation of the Industrial Chain

As shown in Figure 4.1, the Edge-Cloud industry is an ecosystem built by telecom operators, telecom vendors, IT vendors, third-party APP developers, and content providers. Telecom operators are the core of the whole industrial ecological chain, and are also the foundation of industrial collaboration.

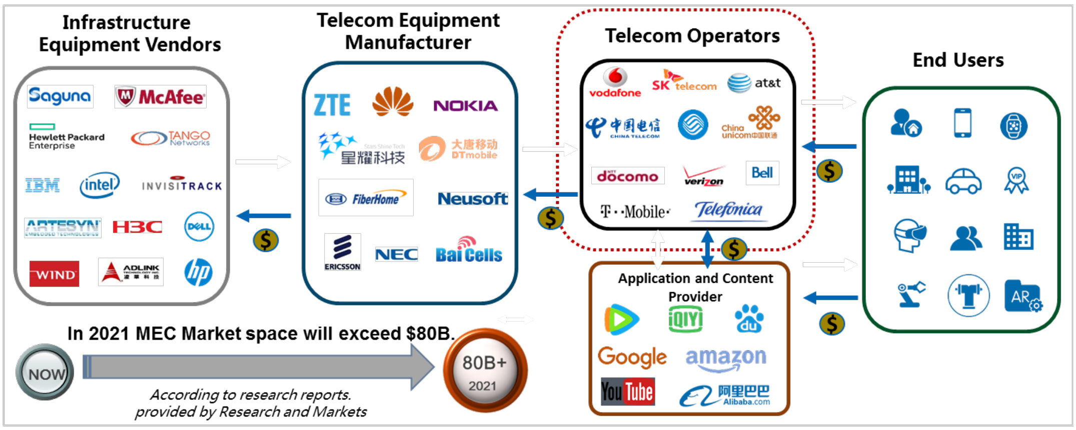


Figure 4.1 Landscape of the Edge-Cloud Industrial Chain

It is predicted that the number of connected IoT devices will reach 50 billion in 2020, and 50 percent of these devices should have the edge computing capability. For the global 5G market, it is estimated that the scale of Edge-Cloud devices and system markets will reach $80 billion in 2021. In the whole Edge-Cloud industrial chain, the ratio of pipeline connection value is only 10 to 15 percent, and the ratio of application services is 45 to 65 percent. Telecom operators are successively beginning network reconstruction and transformation and changing from traditional pipeline connectors to industrial integrators, aiming to become service providers. By relying on the establishment of edge cloud platforms, telecom operators can provide its platform capabilities to third-party APP developers, rapidly launch user-oriented innovative services, and shorten the Time to Market (TTM).

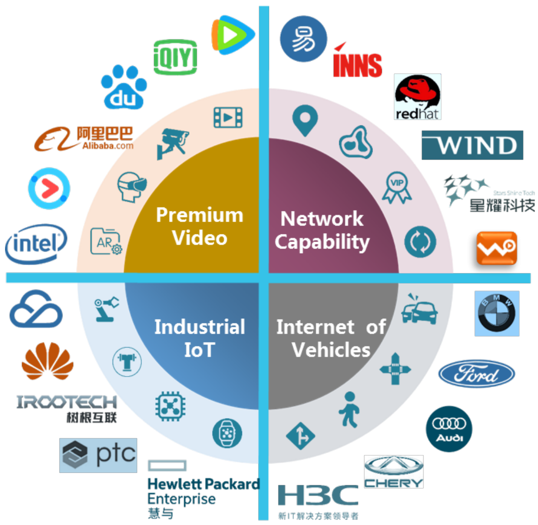


Figure 4.2 China Unicom's Edge Cloud Ecosystem

China Unicom is committed to creating a brand-new value chain and an edge ecosystem full of vitality by collaborating with all industries. The subjects of all processes are motivated to work closely together to promote the integration among IT, CT, and OT and deeply excavate the potential value of the edge network. As shown in Figure 4.2, China Unicom focuses on big video, VR/AR, industrial IoT, IoV, and other high-bandwidth and low-latency services by adhering to the “Openness and Cooperation” philosophy. It provides location, charging, QoS, and other platform capabilities to third parties, aiming to construct an Edge-Cloud content ecosystem with industrial partners. Currently, China Unicom has over 100 Edge-Cloud partners, including Tencent, Alibaba, Baidu, Jingdong, iQIYI, Intel, ZTE, Star Technology, H3C, HPE and so on.

4.3 Orientation of China Unicom's Edge Cloud

Global operators are encountered with the dilemma in which voice services create less value. At the same time, the rapid growth of data traffic does not bring proportional revenue increase due to the coming of innovative enterprises. How to resolve the dilemma and expand value-added services are the problems to be solved. During the ICT integration tide, operators regard thousands of edge DCs as the best high-quality resources compared with OTT. China Unicom will deeply excavate the Edge-Cloud service platform capability, the same time, still enhancing pipeline capabilities, and open a new window for OTT collaboration.

- A unified Edge-Cloud resource pool will be rapidly constructed during mixed-ownership reforms, on the basis of 5G cloudification evolution, aiming at digital transformation.
- An open, open-source, and friendly edge service PaaS platform is built to provide rich platform capabilities and unified APIs for third-party APP developers.
- 4G networks are 5G-oriented and important service scenarios are deployed to provide business support and incubation.
- The ecological collaboration depth and width of the Edge-Cloud industry will be expanded, and all industrial partners are involved to explore business models and co-build the edge ecology of 5G networks.

5 Key Service Scenarios and Commerial Applications

5.1 Scenario 1: Private Customization of Campus Area Network

Campus Area Network generally refers to the private intranet of university or enterprise, which covers key requirements, such as regular mobile office work and internal enterprise communication, and needs to be isolated from public networks for security. With the development of IoT, in addition to the demand for interpersonal communications, that the demand for management of things is increasingly urgent. WiFi-based dedicated mobile networks show the disadvantages of a long construction period, high costs, high maintenance costs, and cannot ensure sufficient, stable, or secure coverage or provide continuous services. For user experience, it is inconvenient for end users that dedicated networks can be used only for special purposes. Public services and private services are separated from each other is also inconvenient.

Figure 5.1 Customized Campus Area Network Solution

As shown in Figure 5.1, through the local breakout function, the Edge-Cloud Service Platform can create a mobile virtual private network to shunt enterprise or campus users' local data services into edge servers, which forms a local 4G virtual private network, ensuring the security and stability of mobile office work, avoiding traffic alternation, and reducing the delay in service access. It can also help the enterprise or campus create a private IoT network and deploy related services.

China Unicom is committed to building Edge-Cloud based "Smart Park Solutions" in several provinces in China to shunt service access by enterprise or campus user and provide carrier-class security identity authentication. Besides, the Edge-Cloud based IoT management platform can manage wireless sensors such as for fire sprinklers and fire alarms jointly, perform indoor positioning–based rescue, and organize real-time fire information–based evacuation to build a safe and reliable private network in a Smart Park.

5.2 Scenario 2: Mobile vCDN

The rapid growth of video services significantly affects the bearer capacity of mobile operators. OTT vendors have deployed many CDN nodes on a large scale. However, the CDN nodes are mainly deployed inside fixed networks, and mobile users can access video services only on the back end of the core network, which poses a huge challenge to operators' network resources, especially the transmission bandwidth on the S1 interface. When CDN nodes are deployed inside a mobile network, for example, vCDN nodes are migrated downward to the edge DCs of operators through the deployment on the Edge-Cloud platform, the traffic load on traditional networks can be greatly reduced and mobile user experience of video services can be improved.

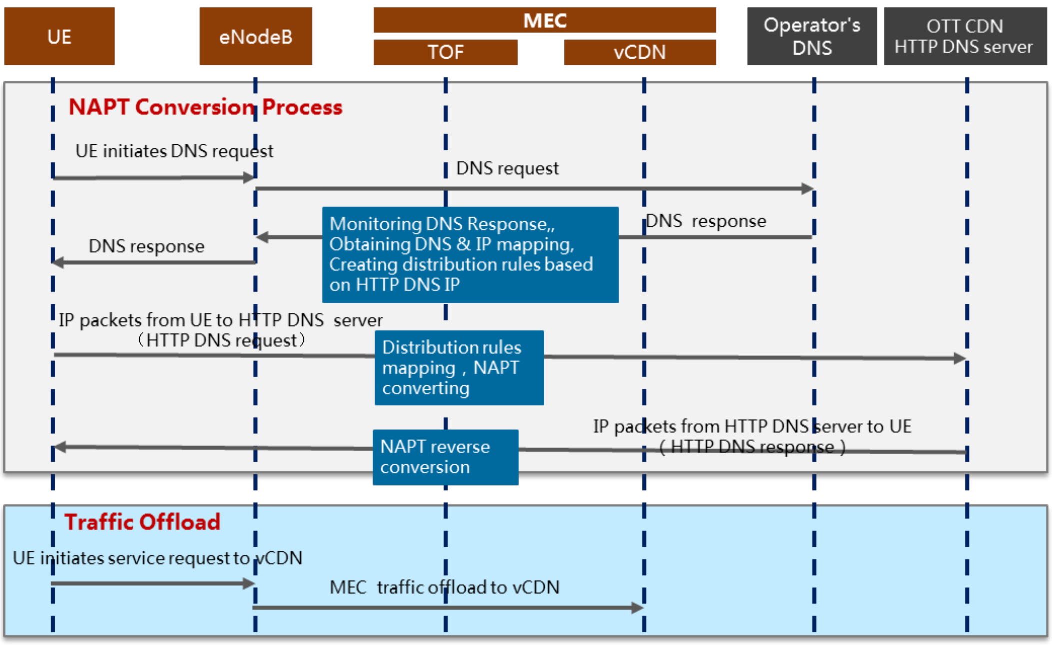


Figure 5.2 Domain Name–Based Breakout and NAPT Function Procedure

Working with ZTE and Intel, China Unicom verified edge vCDN of Tencent videos and Wo videos. Edge-Cloud shunts video services by domain name through the NAPT function. As shown in Figure 5.2, the CDN HTTP DNS servers of OTT vendors establish server mapping by IP address to ensure that the requests converted by Edge-

Cloud NAPT can be mapped to local edge servers (edge vCDNs) according to the load balancing strategy.

- The Edge-Cloud platform configures the specific domain name, listens to the DNS response sent to the UE, obtains the IP address corresponding to the domain name from the DNS response, and dynamically generates a distribution rule for the IP address.
- The Edge-Cloud platform matches the distribution rule, and performs NAPT (the source IP address is replaced with a specific IP address) for the packet sent to the HTTP DNS.
- The Edge-Cloud platform makes a reverse NAPT conversion on the packet returned to the UE by the HTTP DNS and then sends it to the UE.
- The Edge-Cloud server establishes a breakout mechanism, and locally shunts the packets whose destination address is the edge server.

As shown in figure 5.3, compared to provincial node server, the RTT latency reduces 50% and HTTP download speed increases 43%. At the same time, China Unicom together with ZTE have verified video optimization based on wireless information in the existing networks. In the preliminary stage, the indicators about video quality and wireless network quality are reported through the Wo video app. Based on the reported indicators, the Edge-Cloud platform determines the adaptation of the video bit rate and informs the UE to adjust the bit rate. The UE app reports a video bit rate switchover instruction to the Wo video server and the server switches to the corresponding video source.

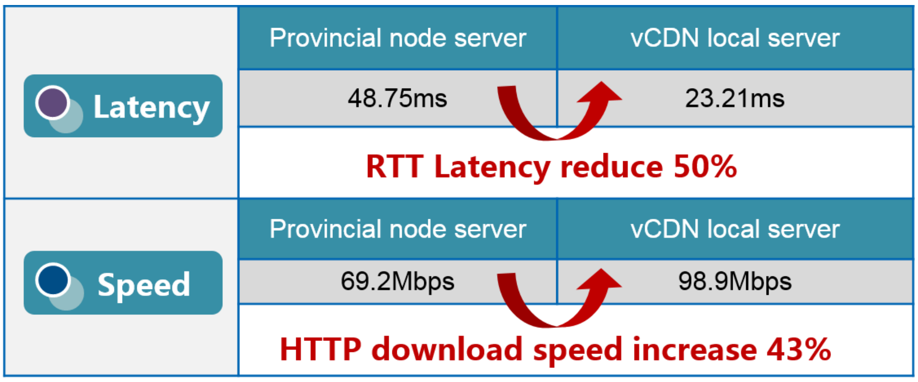


Figure 5.3 Test result of vCDN deployed in Edge DC

5.3 Scenario 3: Video Monitoring and AI Analysis

Traditional monitoring are deployed by using cables and all the cameras must be connected to the cable network. So a large number of cables are used, resulting in high costs but low efficiency. The monitoring videos are transmitted to the cloud or servers for storage and handling through the bearer networks and core networks. This method increases network burden and the end-to-end delay of services. To solve the problems, monitoring data can be shunted to the Edge-Cloud service platform, effectively reducing the network transmission load and the end-to-end service delay. Video monitoring can also be combined with AI. The AI video analysis module is deployed on edge cloud and used in intelligent security, video monitoring, and face recognition scenarios. Edge cloud can realize local analysis, rapid processing, and real-time response due to its low latency, high bandwidth, and fast response, which effectively make up for the disadvantages of AI-based video analysis, such as a high latency and poor user experience.

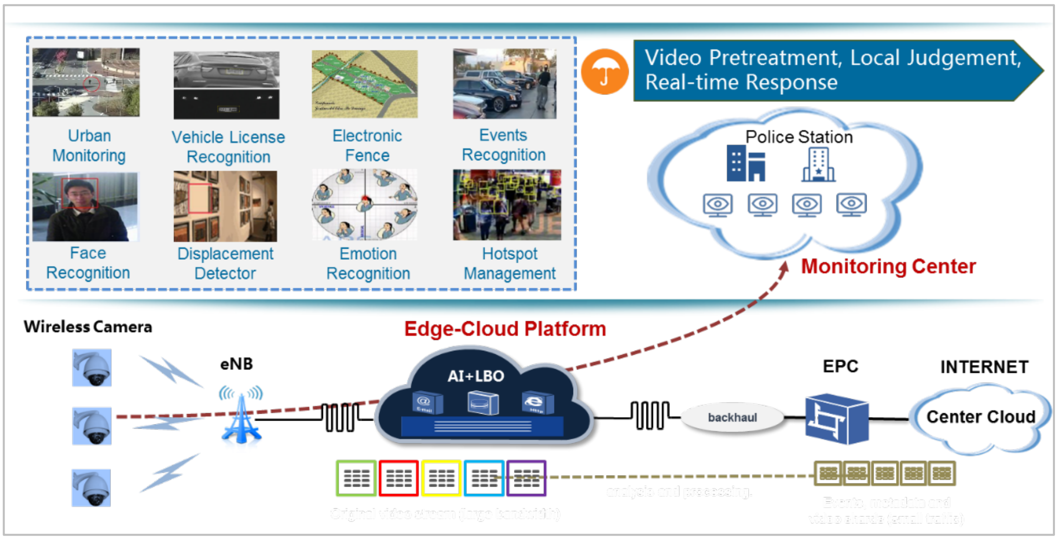


Figure 5.4 Edge-Cloud–Based Intelligent Monitoring Solution

Figure 5.4 shows the AI video analysis system architecture based on the Edge-Cloud service platform. The Edge-Cloud platform locally shunts the videos collected by 4G cameras, which reduces the use of transmission bandwidth resources of the core network and the backbone network. This shortens the end-to-end delay. The cloud computing center performs AI training tasks, and the Edge-Cloud platform executes AI inference. Local decisions and real-time responses make the system applicable to many local AI applications, such as expression recognition, behavior detection, path tracking, hot-spot management, and physical attribute recognition. In addition, cable cameras can also access the intelligent Edge-Cloud analysis platform.

Based on this solution, China Unicom has commercially deployed Edge-Cloud Intelligent Security systems in Zhejiang and Hubei provinces of China in the Xueliang (comprehensive security) projects. The system will be commercially used in more provinces on a large scale.

5.4 Scenario 4: Augmented Reality (AR)

The biggest challenge of AR is the delay and bandwidth. To overcome the challenge, AR information (such as user locations and camera angles) needs to be locally processed to minimize the AR delay and improve the accuracy of data processing. The existing AR solutions are generally based on the interaction between mobile phones and the AR servers on the cloud. However, it always takes a long time for image uploading, image recognition on the cloud, and information feedback due to a large amount of image information. The long time will affect the user experience of AR.

As shown in Figure 5.5, the AR server is deployed on the edge of network, the Edge-Cloud platform determines the user's location through network data calculation. The AR server provides real-time content aggregation and pushes the content, shortening the delay of transmitting a large amount of image data, and effectively reducing the image recognition and calculation load on the cloud. The content on the edge server can be dynamically updated based on the characteristics of the region in any scenario, so the user experience can be enhanced.

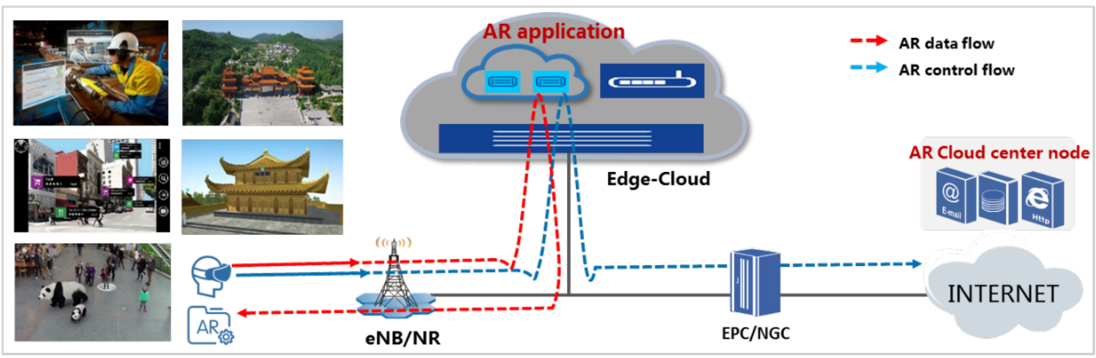


Figure 5.5 Edge-Cloud–Based AR Solution

Based on the above solution, China Unicom is carrying out pilot project of the current network in scenic spots and museums. In a repair scenario, for example, the AR glasses acquire and upload device images for the on-site engineer who wear the glasses.

The repair app on the edge service platform identifies the images, and shows device information and repair instructions on the AR glasses. Experts can remotely provide written or audio guides on the back end. In a scenic spot scenario, the AR glasses take pictures for a tourist and upload them. The value-added scenic app deployed on the edge cloud platform identifies the images, displays enhanced scenic information on the AR glasses, and navigates the tourist based on AR pictures. AR can be widely used in numerous application scenarios, such as museums, art galleries, city memorial halls, scenic spots, and factories, and greatly enhance user experience.

5.5 Scenario 5: Live VR Broadcast

In 2017, China further strengthened the strategic positioning of the VR industry, and explicitly specifies VR as a development key in *"The 13th Five-Year Plan for National Science and Technology Innovation"* and *"The 13th Five-Year Plan for* National *Strategic Emerging Industry Development"*. Live VR broadcast differs from traditional live broadcast in its high resolution ratio, high-code-rate 360-degree vision, 3D image and interaction. Considering the high-bandwidth and low-latency requirements, live VR broadcast is greatly restricted in current networks.

The Edge-Cloud platform allows an ultra-low latency in live VR broadcast. Traditional image collection cameras and VR cameras are wired. Using wireless cell network backhaul instead of the wired solution is more helpful in mobile scenarios and outdoor live broadcast. In addition, the distribution, assembling, and computation of VR video and HD live-broadcast video can be unloaded to the network edge. Cameras only collect image information, thus reducing costs. They can also be manufactured in small size and easy to carry. The near-end distribution of live broadcast contents may result in lower latency and better experience.

As shown in Figure 5.6, China Unicom, Intel, and Tecent Cloud built an Edge-Cloud platform in the "Mercedes-Benz Cloud Center" situated in Shanghai, and deployed the Edge-Cloud virtualized network function on the platform. The open API and EVO is also provided. In this venue, the interaction between fans and web celebrities can be interacted through the live VR broadcast or multi-angle HD video, with an ultra-large bandwidth and a near-zero latency. Compared with the traditional Internet (Center-Cloud) live broadcast with a latency of 30 to 40 seconds, the real-time live broadcast greatly enhances user experience, gaining a good reputation among users.

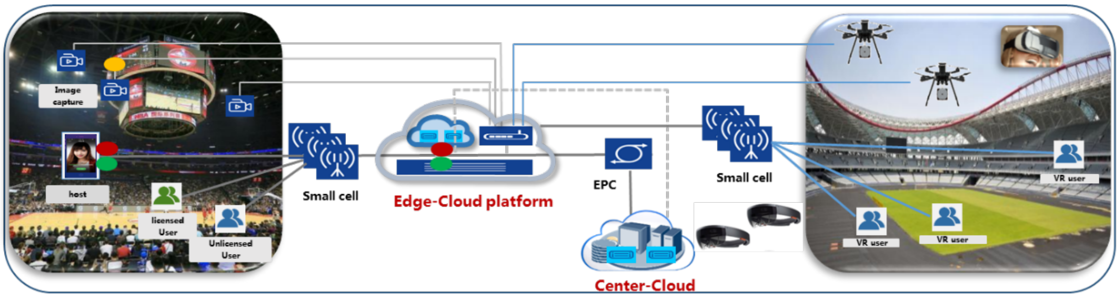


Figure 5.6 Live VR Broadcast Solution (Edge Cloud-Based)

Currently, as the official communication service partner of the 2022 Winter Olympics, China Unicom has already begun the preparation and planning of over 20 Olympics venues located in Beijing, Yanqing, and Zhangjiakou. With the commercial use of 5G networks, mobile live VR broadcast applications (Edge Cloud-based) will surely reach a new vertex.

5.6 Scenario 6: IoV C-V2X

IoV refers to all-round network connections (inside vehicles, between vehicles and men, between vehicles, between vehicles and roads, and between vehicles and service platforms) by using a new generation of information and communication technologies to enhance the intelligent level and automatic driving capabilities of vehicles. IoV will construct new patterns of vehicles and transport services. This improves transportation efficiency and enhances driving and riding experience, providing intelligent, comfortable, safe, energy-saving, and efficient comprehensive services. As a new type of application that develops very fast, IoV raises new requirements for service quality:

- Information and entertainment applications (environmental information sharing and in-vehicle entertainment): a bandwidth of 10 to 100 Mbps and a latency of 500 ms.
- Traffic safety driving applications (platooning and anti-collision): a bandwidth smaller than 1 Mbps, and a latency of 20 to 100 ms.
- Automatic driving: a bandwidth of 100 Mbps and a latency of 1 to 10 ms.

To meet the extreme requirements of IoV for latencies and reduce the load of the bearer network and core network, the Edge Cloud-based IoV network architecture is composed of an Edge-Cloud platform, RSU for drive testing, mobile terminals, vehicle-mounted terminal (OBU), and traffic lights, see Figure 5.7. The RSU collects data uploaded by various terminals and distributes instructions to various terminals. The Edge-Cloud platform can execute regional algorithms to implement safety avoidance, speed guidance, and regional traffic analysis.The same time it can support handovers and dynamic data synchronization for moving vehicles.

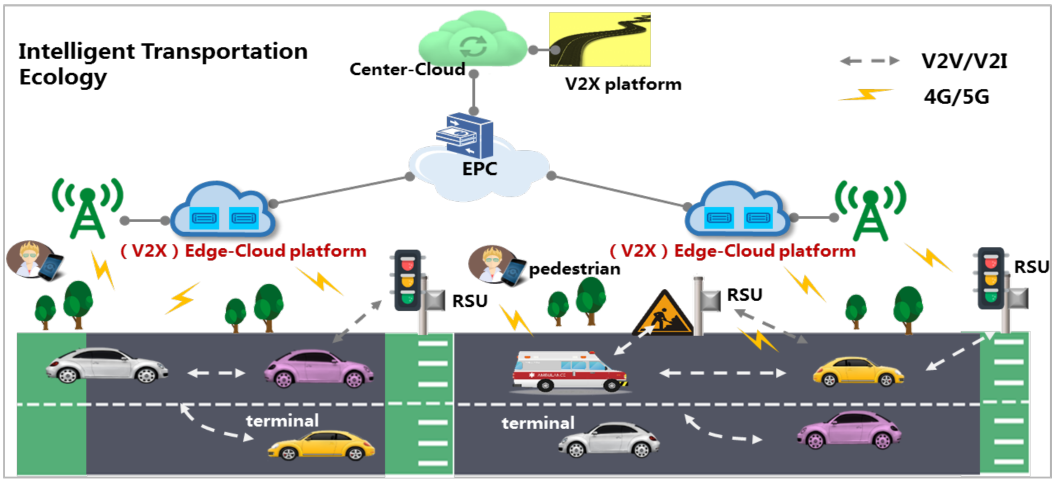


Figure 5.7 Edge Cloud-Based V2X Solution

Based on the above solution, China Unicom already starts the current-network pilot site of V2X in Chongqing. The verified applications include the pushing of traffic information based on relational models, APP development of driving vehicle queues based on the Edge-Cloud Platform, intelligent traffic scheduling, and big data analysis.

- The real-time traffic information collected by the RSU for drive testing and vehicle-mounted terminal (OBU) is sent back to the local Edge-Cloud server through the LTE-V base station. The local Edge-Cloud server carries out the filtering, comparison, association, and fitting operations based on relational models in accordance with the sensing information in the specified region or area. And then generates a real-time traffic information transfer chain. At last, it also pushes and displays the traffic information about special scenarios (such as congestion, traffic accident, and road information, as well as prealarm information) to in-vehicle terminals or other terminals in accordance with the relational message model.

- By using location, route, planned arriving time, climate, and road data, the Edge-Cloud platform carries out the filtering, computation, and combination operations in accordance with the pre-configurations, dynamic requests, and route merging information on the Edge-Cloud. It generates a proper vehicle list, route plan, driving speed, and relevant platooning conditions, sends them to the corresponding vehicles, and also provides decision information for vehicle-related servo organizations.

- The intelligent traffic scheduling of the Edge-Cloud platform collects statistics on vehicles and passengers in different directions in accordance with the V2X RSU-collected traffic object information that is set at crossroads. The V2X-based traffic object statistics and predication model is also designed, and analysis computation is carried out on the Edge-Cloud platform. Traffic lights are dynamically adjusted in accordance with the model to increase the traffic efficiency at crossroads.

- The big data and data analysis platform is used to construct a unified smart core transportation system. The real-time status information, traffic and map data of the IoV is analyzed, and data streams are sorted and classified. A model is then constructed, so that a big data and data analysis platform can be constructed for IoV applications. Finally, a unified smart core transportation system is built on the basis of this platform.

Chongqing is a city with complicated road condition in China. At the pilot site of C-V2X, China Unicom selected the roads in Liangjiang New Area for experiment purposes. 12 different types of typical roads (such as open and densely-populated roads as well as roads with complicated traffic status, 9.6 km long) were covered, meeting the complexity and typical requirements from the perspective of wireless coverage, optimization, in-vehicle information collection, and Edge-Cloud data analysis. On the experimental road, the prealarms about running red lights at signal crossroads, assistant driving of signal-control crossroads, anti-collision reminding at crossroads, and vehicle speed guidance at signal-control crossroads, emergency vehicle prealarms, optimization of signal-light control parameters, road-crossing prealarms of passengers, and simulation of platooning driving were tested. A great amount of rich and precious experimental data was obtained, laying a foundation for the future IoV expansion experiments and industrial applications.

5.7 Scenario 7: Indoor Positioning

Location information is the online identifies of real objects, and can truly reflect the interaction between real worlds and virtual spaces in the three-dimensional coordinate system. It is also the basis of IoE. However, the location problems of men and things in IoT scenarios need to be solved first. However, isolated user location data has very limited value. The business value of information can be fully expanded only after the men, things, and data are connected and placed in scenarios through user location data.

The Edge-Cloud platform provides high-precision indoor positioning through various access modes, and provides high-precision positioning services for smart inventory and logistics, smart manufacturing, emergency rescue, personnel asset management and service robots by using indoor high-precision positioning. Grid-level low-precision positioning is provided for the retail industry. Location information records the movement tracks and areas of each customer. If other big data (such as consumption information, staying time, and behavior habits) is combined, disruptive changes will be caused due to more precise marketing, more efficient delivery, more precise pre-judgment analysis, and more efficient transaction.

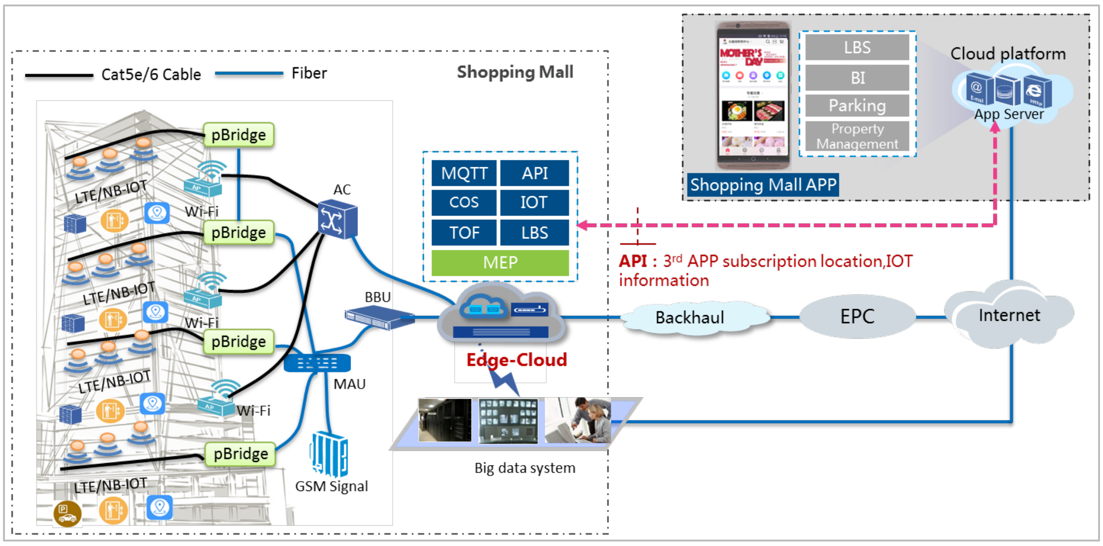


Figure 5.8 Mixed Indoor Positioning Solution (Edge-Cloud Platform-Based)

As shown in Figure 5.8, China Unicom is committed to launching a smart shopping mall solution based on the Edge-Cloud platform. This solution involves multiple positioning technologies such as indoor base stations, WiFi, and NB-IoT, allowing five-meters high-precision positioning. At the same time, the Edge-Cloud-based IoT management platform can manage wireless sensors jointly such as geomagnetic field sensors, fire sprinklers, and fire alarms. In addition, the Edge-Cloud platform is used to provide the capabilities of indoor location services for third-party applications and the big data platform of shopping malls. Indoor navigation and smart parking are provided for users, and property management is also provided for owners. Indoor locations and big data are combined to excavate valid data for further analysis and calculation, aiming to get the highest valuable information.

5.8 Scenario 8: Industrial IoT

Currently, smart manufacturing, electricity, and public utilities have been digitized, intelligized, and network-oriented, and face a series of key challenges.The challenges includes: OT and ICT cross-border collaboration, interconnection and integration of physical worlds and digital worlds, valid information flowing and integration caused by the co-existence of multiple heterogeneous technologies, integration of industrial Know-How and data driving in the process of knowledge modelization, and end-to-end collaboration and integration resulting from the extension of the industrial chain.

Based on the edge computing model and intelligent distributed architecture, an intelligent Edge-Cloud Platform can be constructed on the industrial site to achieve unified network connections, unified intelligent distributed architecture, unified information models, unified data services, unified control models, and unified service orchestration. The industrial Edge-Cloud integrates local network, computation, storage, and application core capabilities inside factories, and provides a nearby local cloud platform for production management services. It complies with the mode of "Moving Factory Management to the Edge End and Moving Perceiving End Upwards", and conditionally integrates OT, IT, and CT. In this way, the massive, heterogeneous, and real-time connections between terminals and devices are achieved, and networks are automatically deployed and maintained to ensure the safety of connections. The hardware infrastructure of industrial Edge-Cloud uses a universal X86 architecture server, and the industrial characteristics are stronger than those of the commercial server, providing more stability and security. Based on the Openstack or Kubernets technology, a VM-based or Container-based virtualized platform is built on the basis of hardware infrastructure, aiming to placing computation, storage, and network resources in a pool. This allows application developers in vertical industries to design various applications.

Figure 5.9 shows the function architecture of the industrial Edge-Cloud Platform. This architecture provides a development and deployment operation service framework, and can achieve intelligent collaboration between development and deployment. In this way, the unification of software development interfaces and the automation of deployment operation are implemented. The southbound interfaces can interconnect industrial sensors and devices. The northbound interfaces can interconnect industrial APPs and implement various typical applications through OT&ICT integration, including the predictive maintenance of industrial devices, power consumption analysis of industrial sites, service orchestration of the industrial production process, and network connection of devices on industrial sites. Additionally, the industrial Edge-Cloud Platform allows multiple types of vertical applications customized for different production enterprises, including the online real-time status detection of production devices, online maintenance of production devices, network connection of the edge cloud platform, and remote cloud-end management.

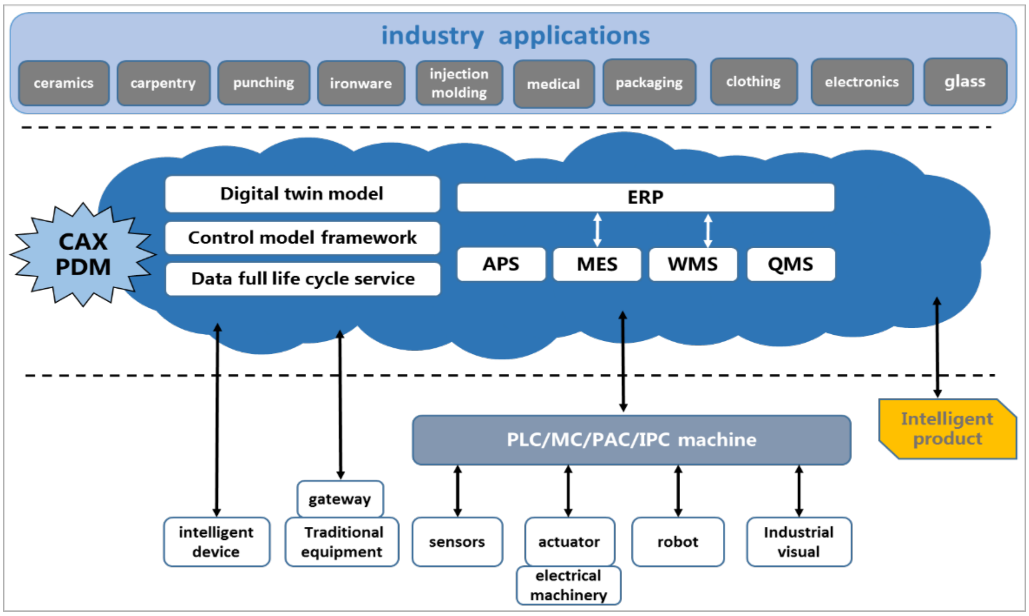


Figure 5.9 Function Architecture of the Industrial Edge-Cloud Platform

Currently, China Unicom positively responds to the national "*Internet + Intelligent Manufacturing*" strategy, and works along with the Guangdong Intelligent Manufacturing Institute and South China Intelligent Robot Innovative Institute for intelligent manufacturing Edge-Cloud pilot engineering held in Foshan, Guangdong province. Besides, China Unicom has launched the R&D and service demonstration project towards the 5G network-oriented industrial manufacturing Edge-Cloud Platform in accordance with the key national technological project *"Industrial Manufacturing-Oriented 5G Service R&D and Experiments"*.

6 Summary and Forecasting

When facing the coming 5G era and successive tides of ICT integration, we are ready to brace the future through a series of reforms. Different from the role of pipeline provider in the 4G era, telecom operators will gain more opportunies to expand their value-added services in the coming 5G era, and become comprehensive end-to-end service providers. As a new technology of ICT integration, edge computing deploys high-bandwidth, low-latency, and localized services onto the network edge, and provides unified telecom infrastructure support for FMC, which is of vital importance for telecom operator’s digital transformation and industrial structure upgrade. China Unicom's Edge-Cloud platform will provide differentiated network capabilities and services for application parties in vertical industries, and provide unified operation management support, thus significantly enhancing industrial competitiveness.

Although both standards organizations and industry alliances are actively engaged in the research of the technological development of edge computing, constructing the ecology of the edge industry needs more open collaborative research and bold industrial practice. By working with Baidu, Alibaba, Tecent, ZTE, and Intel, China Unicom has begun large-scale Edge-Cloud pilot projects and pre-commerical network construction in 15 provinces and cities of China. It is committed to building an open and open-source Edge-Cloud Service PaaS Platform to flexible allocate computation, storage, network, and accelerator resources, carry out orchestration amangement for varied edge services, and provide rich platform service capabilities and unified APIs for application developers, aiming to accelerate the incubation and promotion of edge services. This successful industrial practice will be of great guidance for the construction of the ecology of edge computing. Next, China Unicom will pursue the exploration of nationwide smart campuses, smart venues, IoV V2X, and industrial IoT.

All industries have unlimited anticipation and expectation for various application scenarios of edge service platforms in the future. However, the fulfillment of this dream needs the joint effort of the whole industrial chain. China Unicom is expecting to work with more industrial partners during the next large-scale Edge-Cloud pilot projects and the advancement of commercial use, and jointly discuss the collaboration modes of edge service platforms, aiming to co-build a 5G-oriented edge ecosystem and fully promote the vigorous development of edge services.

This White Paper is Co-Written by

- ZTE INTEL

China Unicom's Edge-Cloud Ecological Partners

Abbreviations

Abbreviation	Full Name
OT	Operation Technology
IT	Information Technology
CT	Communication Techonology
ICT	Information and Communication Technology
MEC	Multi-access Edge Computing
OTT	Over The Top
CDN	Content Delivery Network
DC	Data Center
eMBB	enhanced Mobile Broadband
mMTC	massive Machine Type of Communication
uRLLC	ultra Reliable & Low Latency Communication
CAPEX	Capital Expenditure
OPEX	Operating Expense
CUPS	Control and User Plane Separation of EPC nodes
HGU	Home Gateway Unit
OLT	Optical Line Terminal
BRAS	Broadband Remote Access Server
SDN	Software Defined Network

CORD	Central Office Re-architected as a Data Center
NFV	Network Function Virtualization
NFVI	Network Functions Virtualisation Infrastructure
NFVO	Network Functions Virtualisation Orchestrator
VNF	Virtual network function
VNFM	VNF Manager
MANO	MANagement and Orchestration
DevOps	Development and Operations
UPF	User Plane Function
RNIS	Radio Network Information Service
CPE	Customer Premise Equipment
OSS	Operation support system
AMF	Access and Mobility Management Function
SMF	Session Management Function
MME	Mobility Management Entity
IoT	Internet of Things
NB-IoT	Narrow Band Internet of Things
CU	Central Unit
DU	Distributed Unit
SBC	Session Border Controller
BGN	Broadband Network Gateway

APP	Application
API	Application Programming Interface
IDC	Internet Data Center
COTS	Commercial Off The Shelf
BIOS	Basic Input Output System
FW	FirmWare
OS	Operation System
GPU	Graphics Processing Unit
FPGA	Field Programmable Gate Array
VIM	Virtualised Infrastructure Manager
PIM	Physical Infrastructure Manager
CMP	Cloud management platforms
MEPO	Multi-access Edge Platform Operation
MEPM	Multi-access Edge Platform Manager
RO	Resource Orchestrator
NS	Network Service
NSO	Network Service Orchestrator
LCM	Life Cycle Management
SRS	Sounding Reference Signal
RRU	Radio Remote Unit
GTP	GPRS Tunnelling Protocol

AR	Augmented Reality
VR	Virtual Reality
CQI	Channel Quality Indicator
SINR	Signal to Interference plus Noise Ratio
BLER	Block Error Ratio
QoS	Quality of Service
MQTT	Message Queuing Telemetry Transport
COAP	Constrained Application Protocol
FTTP	Fiber To The Premise
FTTC	Fibre To The Cabinet
V2X	Vehicle to Everything
NAPT	Network Address Port Translation
AI	Artificial Intelligence
EVO	EDGE Video Orchestration
RSU	Road Side Unit
OBU	On Board Unit

White Paper for China Unicom's Edge-Cloud

Service Platform Architecture and Industrial Eco-System

Issued by China Unicom