HUAWEI

AI Fabric ,
Intelligent and Lossless
Data Center Network in the AI Era

Directory

AI Fabric, Intelligent and Lossless Data Center Network in the AI Era

1 With the Advent of the AI Era, Enterprise DCs Will Focus More on Efficient Data Processing

AlphaGo's March 2016 victory was a major milestone in artificial intelligence (AI) research. AI is profoundly changing the human social life and the world at an unprecedented speed. Huawei Global Industry Vision (GIV) predicts that the AI procurement rate will reach 86% by 2025. More and more enterprises regard AI as the subsequent strategy for digital transformation. The capability to leverage AI to make decisions, reshape business models and ecosystems, and rebuild customer experience will be a key driving force behind the success of digital transformation.

1.1 Theme of the AI Era: Mining Intelligence from Data

Digital transformation of various industries is being accelerated. The data of International Data Corporation (IDC) indicates that 64% of enterprises has become the seeker and practitioner of digital transformation. According to Gartner's data, among 2000 cross-border companies, 67% of CEOs has determined digitalization as the core of their strategies.

In the digitalization process, a large amount of data will be generated, which is becoming the core asset of companies. Huawei GIV predicts that the amount of new data will reach 180 ZB in 2025. Pure data is not so important, but intelligence extracted from the data have the eternal value. The proportion of unstructured data (such as originally collected voice and video images) keeps increasing, which will reach more than 95% in the future. However, the unstructured data cannot be handled by the big data analytics and has exceeded the processing capability of all human beings due to its large volume. The AI algorithm for deep learning based on machine calculation can resolve this problem by filtering out much invalid data and automatically reassemble useful information, thereby providing efficient decision-making suggestions and smart behavior guidance.

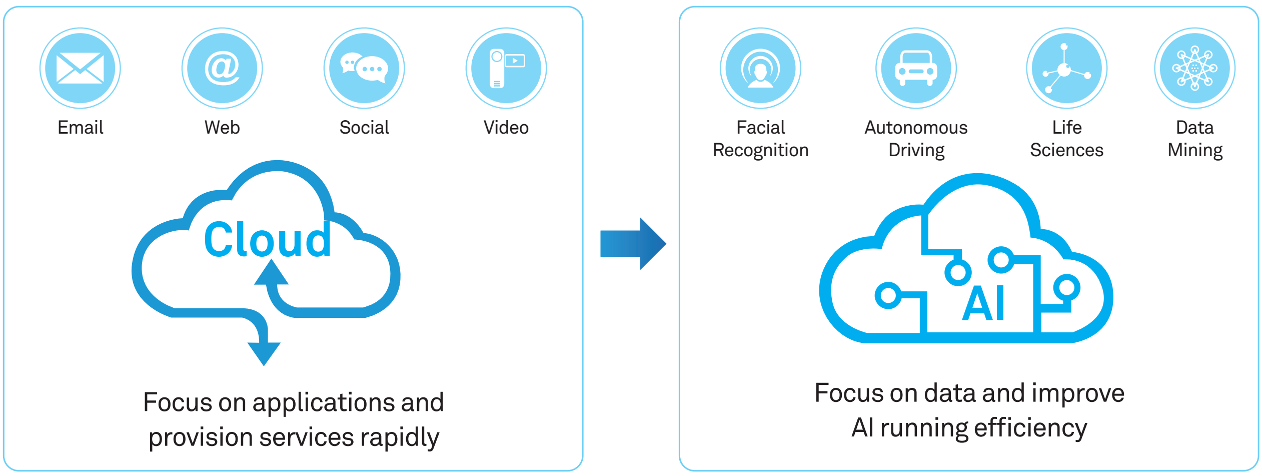


Figure 1 DCs are evolving from the cloud era to the AI era

Mining intelligence from enormous data through AI is the theme in the AI era. Data becomes an important asset for all people and enterprises. AI machine learning and real-time decision-making based on various data have become core tasks of enterprise operation. Compared with the cloud computing era, the focus of enterprise DCs in the AI era is **shifting from fast service provisioning to efficient data processing.**

1.2 Computing and Storage Transformation Improves Data Processing Efficiency, and Communication Latency Becomes a Constraint

The algorithm, computing power, and data are three key elements in AI development. The breakthrough of deep learning algorithm in 2012 has promoted the development of the AI industry. The deep learning algorithm primarily relies on enormous sample data and high-performance computing capabilities. AI training of autonomous driving is used as an example. The data collected in a day is close to the P level. If the traditional hard disk storage and common CPUs are used, the training takes at least one year, which is almost infeasible. To improve AI data processing efficiency, revolutionary changes are taking place in the storage and computing field.

Storage media are evolved from hard disk drives (HDDs) to solid state drives (SSDs) to meet real-time data access requirements. The media latency is reduced by more than 100 times. To match the needs of efficient data computing, the industry has used GPU servers or even dedicated AI chips, improving data processing capability more than 100 folds. As storage media and computing capabilities are greatly improved, the network communication latency becomes a bottleneck for performance improvement in a high-performance DC cluster. The communication latency is increased from 10% to 60% over the entire storage E2E latency; that is, the storage medium waits for idle communication for more than half of the entire storage access time. The computing bottleneck is similar. For example, for a voice recognition training, the duration of each iteration task is 650 ms to 700 ms, and the communication latency is 400 ms. That is, the expensive processor also waits for the communication synchronization of model parameters for more than half of the communication time. In addition, each training task involves millions of iterations, and the communication latency increases by millions of times accordingly.

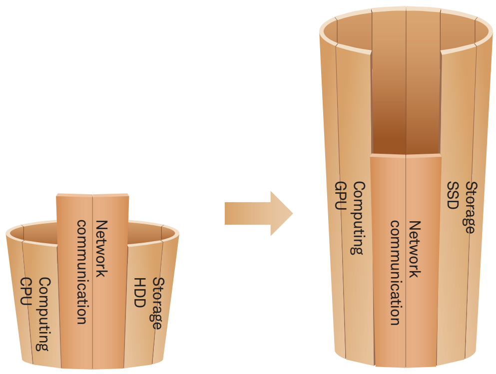


Figure 2 Network communication becomes a short plank of system performance

With the evolution of storage media and computing processors, the ratio of communication latency reaches over 50%, which hinders the improvement of computing and storage efficiency. **The communication short plank in the whole bucket is eliminated only when the communication latency is reduced to a value close to that of computing and storage. This effectively improves computing and storage performance.**

2 Transition from TCP to RDMA Drives DCN Transformation

The boom of AI, emergence of deep learning server clusters, and development of new high-performance storage media such as SSDs impose high requirements on the communication latency (microsecond level). **The traditional TCP/IP protocol stack cannot meet requirements of a high-performance system needs to be transformed first.**

2.1 Replacing TCP/IP with RDMA Become the Trend

Although the traditional TCP/IP network has been maturing over the past 30 years, the inherent technical characteristics limit its application in AI computing and distributed storage.

Restriction 1: The TCP/IP protocol stack brings the latency of tens of microseconds.

When the TCP/IP protocol stack receives or sends packets, the kernel needs to perform context switching multiple times. Each switching causes a latency of about 5 microseconds to 10 microseconds. In addition, data is copied for at least three times and protocol packets are encapsulated depending on the CPU. All this brings a fixed latency of tens of microseconds. The protocol stack latency becomes the most obvious bottleneck in microsecond-aware systems such as AI computing and SSD distributed storage.

Restriction 2: The CPU usage of the server is high due to the TCP protocol stack processing.

In addition to the fixed long latency, the TCP/IP network requires the host CPU to participate in the memory copy of the protocol stack for multiple times. During data transmission and receiving, the CPU load is proportional to the network scale and bandwidth. As a result, the CPU usage is continuously high. According to the industry's calculation result, each transmission of 1-bit data consumes 1 Hz CPU. When the network bandwidth reaches 25 Gbit/s or higher (full load), at least half of the CPU capability of most servers will be used to transmit data.

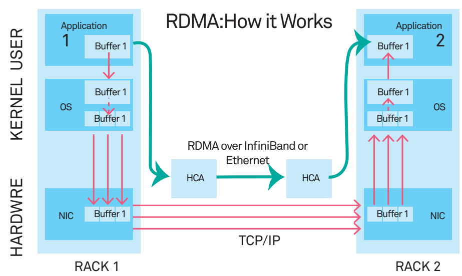


Figure 3 Working principle of RDMA and TCP/IP

RDMA mitigates the preceding limitation of TCP/IP and reduces the protocol stack latency to about 1 microsecond.

The RDMA kernel bypass mechanism allows direct data read and write between applications and NICs, reducing the data transmission latency on the server to about 1 microsecond. In addition, the RDMA memory zero copy mechanism allows the receiver to directly read data from the memory of the transmitter, which greatly reduces the CPU load and improves the CPU efficiency.

According to the test data of a well-known Internet vendor, the computing efficiency is improved 6-8 folds by using RDMA. The transmission latency of about 1 microsecond in servers makes it possible for the latency of the SSD distributed storage to be shortened from milliseconds to microseconds. In the latest NVMe interface protocol, RDMA becomes a mainstream network communication protocol stack. **Facing the network trend of AI computing and SSD distributed storage that demand ultimate performance, replacing TCP/IP with RDMA becomes an inevitable trend.**

2.2 Two Types of RDMA Network Bearer Solutions Have Disadvantages

The RDMA network is developing. The dedicated InfiniBand network and traditional Ethernet network are two network bearer solutions for RDMA. However, the two solutions have disadvantages.

2.2.1 InfiniBand Uses a Closed Architecture and Is Incompatible with the Live Network

InfiniBand uses a closed architecture, causing vendor lock-in and incompatibility with the live network.

InfiniBand switches are dedicated products provided by specific vendors and use proprietary protocols. Most live networks use IP Ethernet networks. Therefore, InfiniBand cannot meet interoperability requirements for AI computing and distributed storage systems that require many interconnections. In addition, the closed architecture also leads to vendor lock-in. If a service system that requires large-scale elastic expansion is locked by a vendor, risks cannot be controlled. InfiniBand is often used in a small-scale traditional HPC cluster.

Complex O&M and dedicated O&M personnel result in high OPEX.

InfiniBand cannot reuse the O&M accumulation and platform of users on the IP network. If InfiniBand is introduced, enterprises have to recruit dedicated O&M personnel. However, there is a lack of experienced O&M personnel in the industry as InfiniBand has only small market space (less than 1% of Ethernet). Network faults cannot be rectified in a timely manner, resulting in high OPEX.

2.2.2 Packet Loss Occurs Due to Congestion on the IP Ethernet Network, and the Throughput Is Low

The standardization of RDMA over IP networks has been completed. RDMA is carried on traditional IP Ethernet networks, which is inevitable for large-scale RDMA applications.

However, RDMA is initially carried on a lossless InfiniBand network and lacks a comprehensive packet loss protection mechanism, making it sensitive to packet loss. In Figure 5, the packet loss ratio larger than 10^{-3} causes the throughput to decrease sharply. The packet loss ratio of 2% causes the RDMA throughput to decrease to 0. To ensure that the RDMA throughput is not affected, the packet loss rate must be less than 1/100000 or 0.

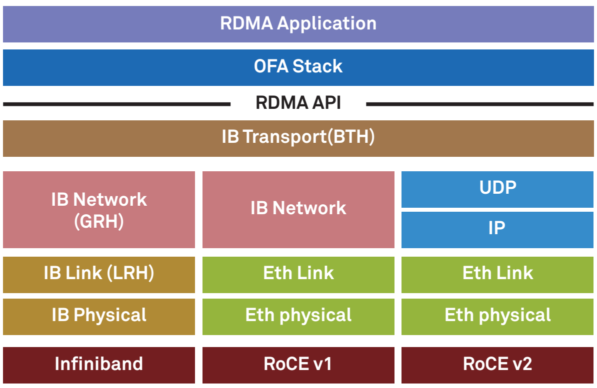


Figure 4 RDMA protocol stack

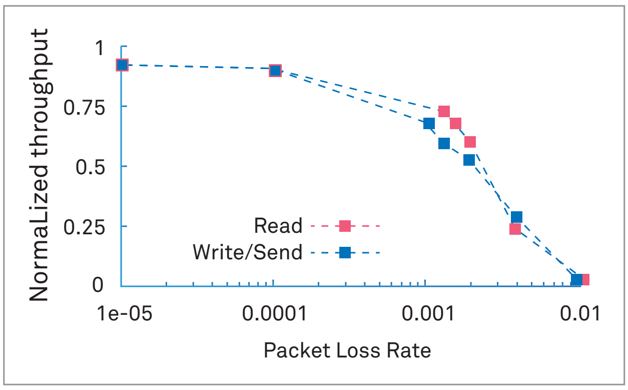


Figure 5 Impact of packet loss on RDMA throughput

Packet loss caused by congestion is the inherent defect of the traditional IP Ethernet network. Once the traffic model is complex and packet loss occurs, the throughput is very low. Many vendors use Priority-based Flow Control (PFC) and Explicit Congestion Notification (ECN) mechanisms to avoid packet loss and increase the throughput. However, the existing RDMA congestion and scheduling algorithm easily causes queue congestion on network devices and triggers PFC. If PFC is triggered continuously on the network, a network deadlock may occur, leading to system risks. In addition, PFC and ECN use backpressure signals to reduce the transmission rate, ensuring zero packet loss. Actually, the throughput is not improved.

Efficient running of RDMA must be carried over an open Ethernet with zero packet loss and high throughput.

3 Distributed Architecture Becomes the Trend, Aggravating Network Congestion and Driving Network Transformation

With popularization of cloud computing, the application architecture's evolution from centralized to distributed has become a consensus in the industry. According to statistics, more than 80% application systems in industries such as Internet, finance, and enterprise industries are migrated to the distributed systems. A large number of PCs are used to replace midrange computers, which brings benefits such as low cost, scalability, and controllability. It also brings challenges to reliability and network interconnection.

3.1 Impact of the Distributed Architecture on the Network

The distributed architecture brings much collaboration between servers.

For distributed computing, a MAP/REDUCE process is generally used for collaboration. In the MAP phase, a large computing task is divided into multiple subtasks, and each subtask is distributed to a compute node for processing. In the REDUCE phase, processing results of multiple computing nodes are collected and summarized. The two processes are repeated.

In the REDUCE phase, two changes are brought to the network to aggravate network congestion.

Change 1: incast traffic characteristics

For traditional traffic characteristics, point-to-point traffic is called unicast traffic, and point-to-multipoint traffic is called broadcast or multicast traffic. Incast traffic refers to multipoint-to-point traffic and corresponds to the REDUCE phase.

Due to incast traffic, bursts occur on the receiver and exceeds the interface capability of the receiver instantaneously. As a result, packet loss occurs due to congestion.

In addition, in a distributed architecture, the role of each server is equivalent, and a server is used as both the transmitter and receiver. That is, bursts of incast traffic cannot be resolved by increasing the interface bandwidth of the receiver.

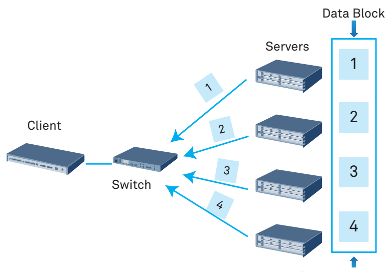


Figure 6 Traffic model of the distributed architecture

Change 2: interaction of large-sized packets

Large-sized packets mean that the length of packets exchanged between servers increases with the complexity of distributed computing. For example, in the distributed computing of image recognition, each interaction model reaches G bytes.

Similar phases exist in distributed storage systems. For example, in the write phase, data is distributed to multiple storage nodes, which is similar to the MAP process. In the read phase, data is read from multiple storage nodes, which is similar to the REDUCE process.

Bursts of incast traffic and large-sized packets caused by the distributed architecture further aggravate network congestion.

3.2 DCN Transformation Direction: Zero Packet Loss, Low Latency, and High Throughput

In terms of distributed architecture and RDMA communication efficiency, DCNs need to be transformed. Since 2000, the bandwidth of DCNs has increased from 100 Mbit/s to 100 Gbit/s, which increases 1000 folds. The Moore's Law supports the growth of bandwidth. However, due to network congestion, simply increasing bandwidth does not improve application performance. The direction of network transformation is shifting from bandwidth to latency, which is a huge leap. This is the requirement for efficient data processing in the AI era and the inevitable trend for IP network technology development.

The latency is the actual latency in the case of full load but not the test latency when the network is lightly loaded. The network latency falls into static latency and dynamic latency.

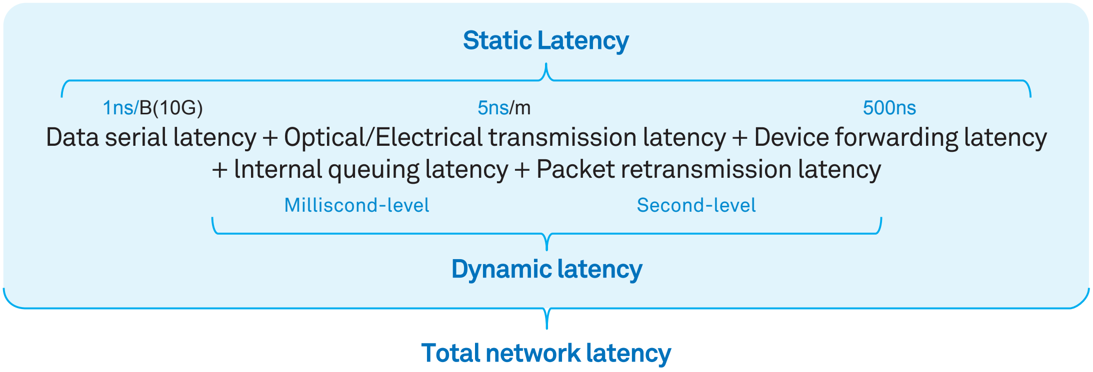


Figure 7 Network E2E latency

The static latency includes the data serial latency, device forwarding latency, and optical/electrical transmission latency. It is determined by the capability of the forwarding chip and transmission distance, and reaches nanoseconds or sub-nanoseconds in industry. The static latency ratio is less than 1%. The forwarding latency of the chip claimed by vendors reaches hundreds of nanoseconds, which is the static latency of a single packet.

The dynamic latency greatly affects the network performance. The dynamic latency ratio is greater than 99%.

The dynamic latency includes the internal queuing latency and packet retransmission latency, which are caused by network congestion and packet loss. In the AI era, traffic conflicts become more and more severe on networks. Packet queuing or packet loss often occurs, causing the latency within sub-nanoseconds. Therefore, the key of the low-latency network is the low dynamic latency.

The dynamic latency is the single-flow or multi-flow latency. That is, a flow includes multiple packets, and the flow completion time (FCT) depends on the completion time of the last packet. That is, if any packet is congested, the FCT increases.

For a distributed architecture, one job includes multiple flows, that is, the job completion time (JCT) depends on the completion time of the last flow. That is, if any flow is congested, the JCT increases.

To meet requirements for efficient data processing in the AI era, adapt to distributed architecture, and accelerate RDMA communication, zero packet loss, low latency, and high throughput become the three core requirements of next-generation DCNs.

4 AI Fabric Reconstructs DCNs and Improves Data Storage and Processing Efficiency by 25-40% in the AI Era

In the AI era, Huawei grasps the opportunity of RDMA upgrade of DCNs and innovatively builds a next-generation intelligent lossless low-latency DCN solution – AI Fabric. Based on two levels of AI chips and a unique intelligent congestion scheduling algorithm, AI Fabric implements zero packet loss, high throughput, and low latency of RDMA service flows. It accelerates computing and storage efficiency in the AI era, and finally achieves the performance comparable to the dedicated network at a low cost of the Ethernet, improving the overall ROI 45 folds. This builds a unified, converged, and highly efficient DCN for future DCs.

4.1 AI Fabric Uses a Unique Algorithm to Ensure the Highest Throughput and Lowest Latency in Addition to Zero Packet Loss

AI Fabric has three core characteristics: zero packet loss, low latency, and high throughput. Different from the universal lossless network technology in the industry, Huawei AI Fabric can achieve the optimal performance in terms of the three indicators.

These three core indicators affect each other. There are great challenges to implement the optimal performance of the three indicators:

Zero packet loss

Bandwidth is limited, resulting in low throughput and increasing the transmission latency of elephant flows.

Low latency

The waiting time of queues on a switch is reduced, resulting in low throughput.

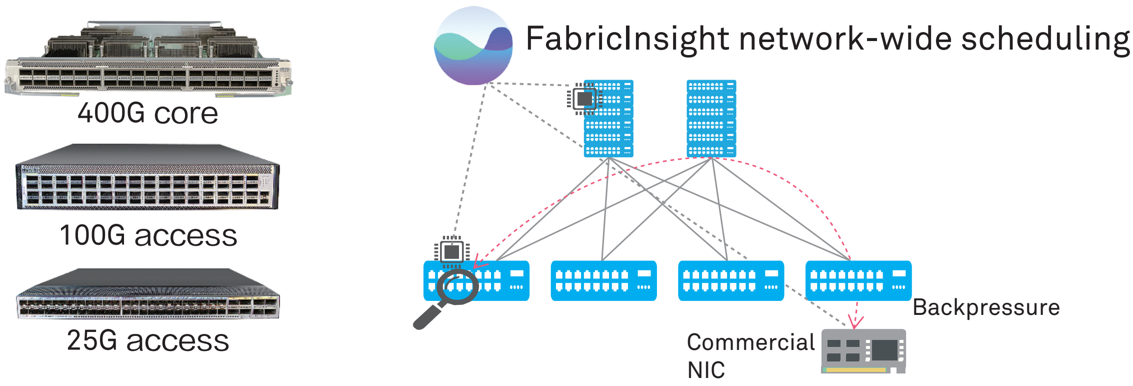
High throughput

Retaining high link utilization leads to queuing on a switch due to congestion and long latency of mice flows.

To achieve zero packet loss, low latency, and high throughput, the core technology is the congestion control algorithm.

Data Center Quantized Congestion Notification (DCQCN), which is the universal lossless network congestion control algorithm, requires collaboration between network adapters and networks. Each node needs to be configured with dozens of parameters, and hundreds of thousands of parameters are involved on the entire network. Only common configurations can be used to simplify the configuration. As a result, the three core indicators cannot be achieved in different traffic models.

25G/100G/400G Networking, Innovative iLossless Algorithm



Dynamic Threshold,Accurate Identification, and Fast Backpressure

VIQ: Virtual Input Queue
It is an internal flow control mechanism that solves the packet loss problem and controls the tail latency.
Dynamic ECN: Dynamic Congestion Waterline
It collects traffic characteristics periodically and uses the customized algorithm to apply ECN parameters matching the traffic characteristics .
Fast CNP: Fast Congestion Feedback
After packets are marked with the ECN flag,CNP packets are generated and sent to the transmitted through the original inbound interface of packets.Then the traffic sending rate of the transmitter is reduced at first time, relieving congestion in the buffer .

Figure 8 AI Fabric Unique iLossless Algorithm

Facing challenges of dynamic traffic and massive parameters, Huawei has invested in the research team to analyze various applications and extract traffic model characteristics. In addition, Huawei switches integrate the AI chip to collect traffic characteristics and network status in real time. Based on the AI algorithm, Huawei switches make decisions locally in real time and dynamically adjust network parameters so that the switch buffer can be efficiently leveraged, achieving zero packet loss on the entire network. In addition, the global intelligent analysis platform FabricInsight uses the AI algorithm to predict the future traffic model based on the traffic characteristics and network status data collected globally. The FabricInsight corrects parameter settings of network adapters and networks in real time to meet requirements of applications .

4.2 AI Fabric Accelerates Data Computing and Storage Efficiency in the AI Era, Improving ROI 45 Folds

According to the EANTC test conclusion, AI Fabric can reduce the computing latency by 44.3% in HPC scenarios, improve the IOPS by 25% in distributed storage scenarios, and ensure zero packet loss in all scenarios.
From the perspective of business value, AI Fabric improves the IOPS by 25% for storage, which is equivalent to reducing the storage investment by 25% while achieving the same performance. A distributed storage system consisting of 512 nodes is used as an example. If AI Fabric is used, 384 storage nodes can achieve IOPS performance of 512 storage nodes on a traditional network. The reduced CAPEX can bring at least 45-fold ROI improvement compared with the investment in AI Fabric.

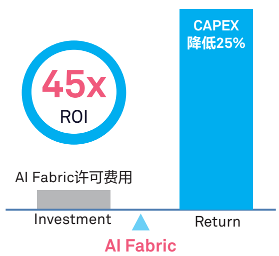


Figure 9 AI Fabric business value

Among the investment in DCs, the proportion of network devices is only 10%, while the proportion of servers and storage devices reaches 85%. AI Fabric can improve the storage performance by 25% and computing efficiency by 40%, which will bring ROI several ten folds.

4.3 AI Fabric Builds a Unified Converged Network for Future-Oriented DCs

There are three types of typical services in a DC: HPC services, storage services, and common services. Each type of service has different requirements on a network. For example, the communication between multiple nodes of the HPC service requires short latency; storage services demand high reliability and zero packet loss; the volume of common services is large and the network needs to be extended cost-effectively and easily. The current DC has three different networks:

- InfiniBand network: provides low-latency IPC communication.
- FC network: provides a storage network with high reliability and zero packet loss.
- Traditional Ethernet: carries common services.

The overall cost of DCNs is high.
High network cost: The price of the FC and InfiniBand private networks is several times higher than that of the Ethernet.
High O&M cost: The FC-based storage network and InfiniBand-based dedicated HPC network require dedicated O&M, while they do not support SDN and cannot meet requirements of cloud-network integration and automatic deployment.

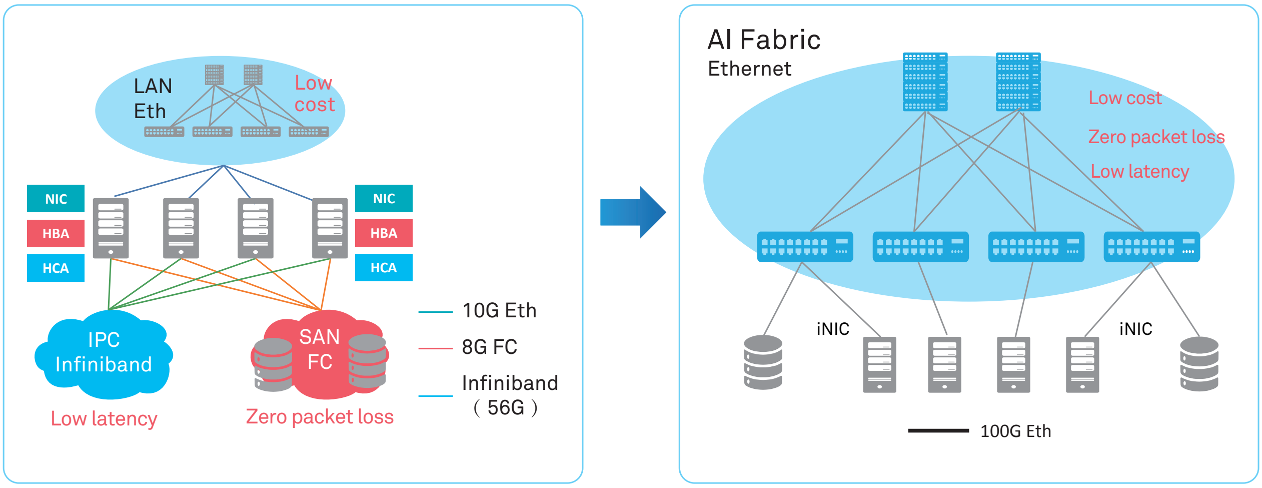


Figure 10 Transformation from independent networking to unified and converged network architecture

Based on open Ethernet, Huawei AI Fabric uses unique AI chips and algorithms to meet requirements of low cost, zero packet loss, and low latency for the Ethernet network. AI Fabric carries SAN traffic, IPC traffic, and common LAN traffic, greatly reducing CAPEX. Traditional Ethernet O&M personnel can manage AI Fabric, requiring no dedicated O&M personnel. AI Fabric supports SDN cloud-network automation, reducing OPEX by at least 60%.

AI Fabric is the optimal choice for DCs in the AI era to build a unified and converged network architecture.

4.4 AI Ready Switch Hardware Architecture, Supporting AI Fabric Long-Term Evolution

AI Fabric uses the first next-generation CloudEngine switch CloudEngine 16800 that is embedded with the AI chip to build the two-level spine-leaf intelligent architecture based on the Clos architecture. AI Fabric integrates computing and network intelligence, and global and local intelligence to build the industry's only AI Ready lossless low-latency fabric network.

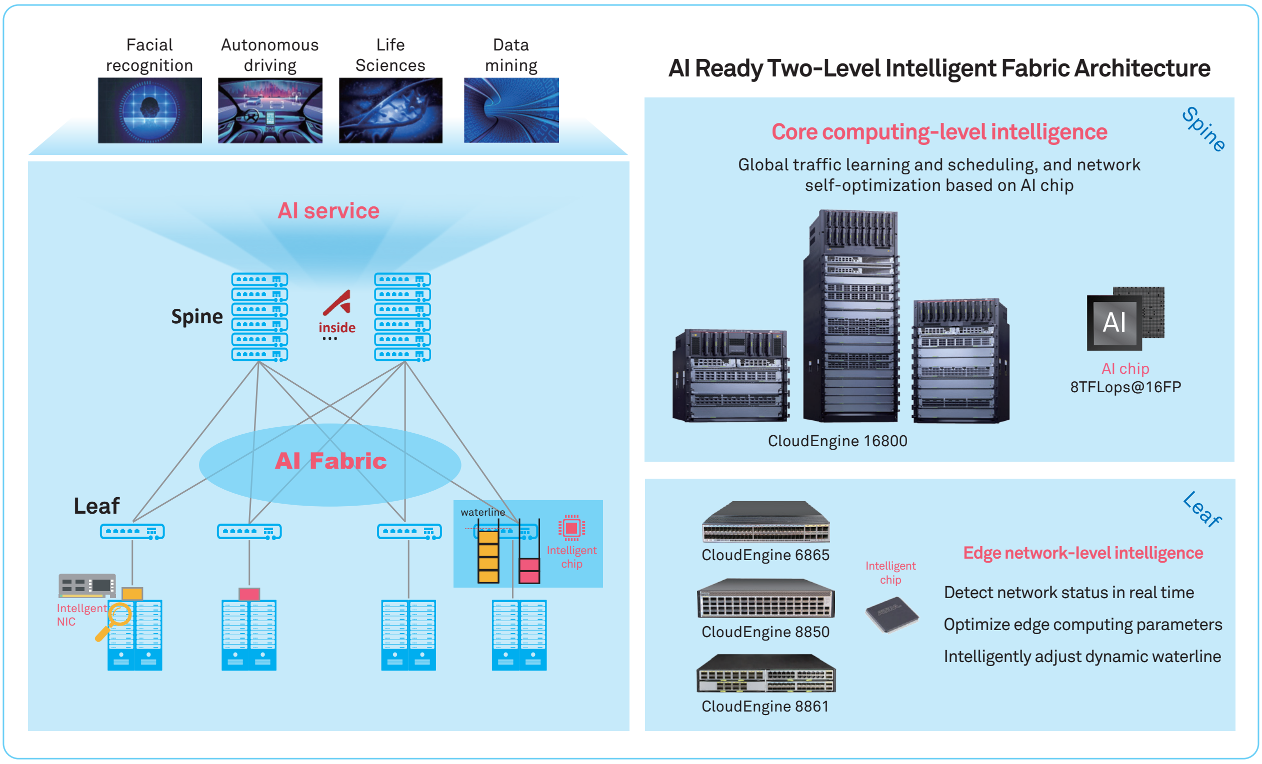


Figure 11 AI Ready two-level intelligent fabric architecture

Core computing-level intelligence

The core switches CloudEngine 16800 are embedded with the AI chips. They provide 8 TFLOPS computing capability, can learn and train network-wide traffic in real time, dynamically generate optimal network parameter settings based on characteristics of different service traffic models, and implement network optimization globally.

Edge network-level intelligence

The edge devices such as ToR switches CloudEngine 8861,CloudEngine 8850,CloudEngine 6865 are embedded with dedicated network intelligence chips to detect the network status in real time, optimize network parameters, intelligently adjust the switching queue waterline based on the local traffic status, and provide the fastest feedback to the transmit end at the best time to adjust the transmission rate, achieving high throughput in addition to zero packet loss on the network.

5 AI Fabric Success Stories

Huawei AI Fabric helps some leading digital Internet and financial enterprises to cope with requirements in the AI era. It accelerates data storage and processing and helps enterprises improve real-time and accurate decision-making.

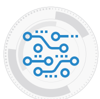
5.1 AI Fabric's Practice in the Internet Industry

AI Fabric Accelerates Training of Autonomous Driving and Shortens the Training Time by 40%

An Internet giant has deployed autonomous driving. The training of autonomous driving involves a large number of AI calculations. Data collected in one day requires hundreds of GPU servers completing training within 7 days, which severely affects the time-to-market (TTM) of autonomous driving. Huawei AI Fabric provides the ultra-high-speed lossless Ethernet with zero packet loss, low latency, and high throughput. It reduces the overall training time by 40% and accelerates the commercial use of autonomous driving.

5.2 AI Fabric's Practice in the Financial Industry

AI Fabric Accelerates the Distributed Storage of China Merchants Bank and Improves the IOPS by 20%



China Merchants Bank's Branch Cloud is an innovative pilot in the CMB cloud strategy. To provide users with the same experience as local disk access, RDMA technology is used to improve network throughput and reduce CPU consumption. Intelligent congestion scheduling of AI Fabric implements zero packet loss, low latency, and high throughput on the network, accelerating RDMA communication. Through the test, the IOPS performance of the storage cluster is improved by 20% and the performance of a single volume reaches 350,000.

As a leading ICT solution provider, Huawei has been exploring how to enable industry digitalization. AI Fabric can improve AI running efficiency and storage performance, shorten training and storage access time, and help enterprises achieve transformation and intelligent upgrade as soon as possible.

HUAWEI TECHNOLOGIES CO., LTD.

Huawei Industrial Base
Bantian Longgang
Shenzhen 518129, P.R. China
Tel: +86-755-28780808

www.huawei.com