1 Executive Summary

As Internet services are developing rapidly and carriers are increasing bandwidth for users, the volume of network traffic is undergoing a sharp increase. The current router capacity cannot meet network development requirements, and therefore, capacity expansion is necessary to carriers. Using conventional methods, however, may cause problems of hardware incompatibility, service interruption, and inflexible switching. In-Service Hardware Expansion (ISHE) described in this document can solve these problems.

Keywords: cluster; NE5000E; capacity expansion; ISHE
As new services such as VoIP, HDTV, and mobile services emerge and consume lots of Internet bandwidth resources, core routers are exposed to capacity, reliability, and expansibility problems. These problems have become bottlenecks in network bandwidth improvement.

Clusters are developed to adapt to rapid Internet development. A cluster is a logical single router. It is formed by multiple chassis to improve the data forwarding capability of core nodes. Clusters can be used for network expansion, meet the requirement of networks for getting flat, and maximize customers’ investment income. Clusters have become a preferred networking solution that can be applied to core nodes.

A Huawei cluster inherits and expands the architecture of a single router. Based on the Versatile Routing Platform (VRP), this architecture provides various features without affecting system stability. Routers of this architecture have been running properly on customers’ networks for a decade.

A single router uses distributed VRPs running on Main Processing Unit (MPUs) and Line Processing Units (LPUs). Communication channels are connected between these VRPs. This ensures good expansibility of the single router and allows it to be smoothly expanded to a back-to-back (BTB) system. A BTB system still uses distributed VRPs which are the same as those on a single chassis.
The BTB system of this architecture can be smoothly expanded to a larger capacity system.

Huawei clusters include BTB clusters (also known as CCC-0), CCC-1, CCC-2, and other models. A cluster can have a maximum of 16 Cluster Central Chassis (CCCs) and 64 Cluster Line-card Chassis (CLCs). As shown in the following figure, the cluster with 16 CCCs and 64 CLCs supports 100 Gbit/s switching capacity per slot, a total of 1024 100-Gbit/s interfaces, and 200 Tbit/s (unidirectional) of total interface capacity.
Sixteen 72-core optical fibers connected to 16 CCCs

Each CCC has a switching plane, with each plane being a 2048x2048 switching matrix

The software versions that have been released support the following cluster models:

CCC-0 (a BTB system): is the simplest NE5000E cluster model. CCC-0 uses the hardware architecture of a single NE5000E. CCC-0 consists of two NE5000Es (a master chassis and a slave chassis) that are connected through channels.

CCC-1: consists of one CCC and a maximum of four CLCs.

CCC-2: consists of two CCCs and a maximum of eight CLCs.

Although deploying clusters achieves capacity expansion, this still poses challenges to smooth expansion without service interruption.

Huawei ISHE is a key to smooth expansion.
3 Solution

To address the challenges posed to smooth hardware expansion, Huawei has sought out an ISHE solution by taking advantage of its advanced hardware and software design capability and its products. The ISHE solution ensures seamless hardware expansion and fully utilizes the original hardware, maximizing carriers’ investment income and greatly reducing carriers’ Operating Expense (OPEX).

3.1 Innovative Hardware Architecture

On a CCC in a cluster, the backplane is placed between Electric Cross Units (ECUs) and Switch Fabric Units (SFUs). ECUs are perpendicular to SFUs, forming an orthogonal matrix.

The orthogonal matrix ensures full connections between data planes and switching planes. The data flow in each high-speed optical fiber is evenly distributed to each switching plane on SFUs.

The orthogonal matrix also ensures full connections between hardware. This allows each shuffle that has independent bandwidth to be connected to all SFUs. The data packets from all LPUs need to be switched on the switching planes of SFUs. A newly added CLC can be connected to all SFUs without having to change its hardware structure.
This innovative hardware architecture supports intelligent sensing of cluster scales and switching modes, adjustment of switching matrix granularity, and on-demand configuration of Optical Flexible Cards (OFCs). With a flexible switching matrix, this innovative hardware architecture allows full utilization of the original hardware during hardware expansion and maximizes customers’ investment income.

### 3.2 Switching Plane Switchover

#### 3.2.1 Introduction to a Switch Fabric

A switch fabric can be called a multi-stage switch fabric if the switch chips on it can be divided into N stages ranging from Stage-1 to Stage N, and all the following conditions are met:

- All inputs connect to only Stage-1 switch chips.
- All Stage-1 switch chips connect to only inputs and Stage-2 switch chips.
- All Stage-2 switch chips connect to only Stage-1 and Stage-3 switch chips.
- All Stage N switch chips connect to only Stage N-1 switch chips and outputs.

![Figure 3-1 Schematic diagram of a single-stage switch fabric](image)

- Like a railway station
- Entrance and exit in the same station
Considering the following factors, CCC-1 and CCC-2 use a multi-stage switch fabric.

- **Complexity**
  It is very difficult and complex for a single switch chip to switch all input data.

- **Cost**
  As the switching capacity increases and data switching becomes more complex, a large-capacity switch chip of a higher cost cannot meet data switching requirements. The larger and more complex the system, the higher the cost of a switch chip and peripheral components.

- **Risk**
  On a single-stage and large capacity switch fabric, a single-chip failure may cause interruption of all services.

- **Expandability**
  A single-stage large capacity switch fabric has a fixed capacity and is not expandable, whereas a multi-stage switch fabric can be expanded as required.

In the recent scores of years, various types of switching models emerged, including shared bus switching, shared buffer switching, crossbar switching, and dynamic routing-based CLOS switching. Large capacity or super large capacity network devices usually use a crossbar switch fabric or a CLOS switch fabric. The shared buffer switch fabric is used for fully-distributed service processing and forwarding on a CLC.
A crossbar (crosspoint) switch is called a matrix switch connecting multiple inputs to multiple outputs in a matrix manner. It is generally accepted as the preferred switch fabric used for a large capacity switching system. The scheduler is a key component for crossbar switch, which is a switch that controls crossbar points. Traffic congestion due to lack of bandwidth resources does not occur on a switch fabric. Crossbar switch has problems of complex structure and higher cost.

A CLOS network is a type of multistage circuit switching network, first formalized by Charles Clos in 1953, which represents a theoretical idealization of practical multi-stage telephone switching systems. On a CLOS network, every switch chip at a stage is connected to all the switch chips at the next stage. The CLOS network is strict-sense non-blocking, rearrangeable, and scalable. Multi-stage switch matrix – CLOS network
In recent years, other vendors have launched CLOS-based and Benes-based products that can achieve multi-chassis cascading and provide 200 Tbit/s of system switching capacity. A rearrangeable non-blocking network of the CLOS network type with $m=n=2$ is generally called a Benes network.

- A multi-stage switch fabric contains fully connected switch chips.
- CLOS principle:
  $C(n,m,r), S1 \rightarrow S2 \rightarrow S3,$ \( \text{In} = \text{Out} \), synchronization, strict-sense nonblocking, multiple paths between

In recent years, other vendors have launched CLOS-based and Benes-based products that can achieve multi-chassis cascading and provide 200 Tbit/s of system switching capacity. A rearrangeable non-blocking network of the CLOS network type with $m=n=2$ is generally called a Benes network.
3.2.2 Huawei’s Switching Technologies Support Smooth Hardware Expansion

Switch chips on an NE5000E can work in three switching modes, namely, S13, S2, and Pass. The switching mode is configurable.

S13: indicates Stage-1 and Stage-3 switching of a CLOS network. During Stage-1 switching, Stage-1 switch chips receive data from TM chips and send the data to Stage-2 switch chips. During Stage-3 switching, Stage-3 switch chips receive data from Stage-2 switch chips and send the data to TM chips.

S2: indicates Stage-2 switching of a CLOS network, responsible for data switching within a single chassis or between chassis in a cluster.

Pass: Switch chips working in Pass mode only function as channels, responsible for transparent transmission of data flows.

A single chassis uses a 1-stage switch fabric, on which all switch chips work in S2 mode. A BTB system also uses a 1-stage switch fabric for data switching between two chassis. In a BTB system, some switch chips work in Pass mode, whereas others work in S2 mode.

Pass Mode: Inputs are directly connected to outputs

S2 Mode: Inputs and outputs are crossed and connected
CCC-1 and CCC-2 use a 3-stage switch fabric. In CCC-1 and CCC-2, the switch chips on all CLCs work in S13 mode, whereas the switch chips on CCCs work in S2 mode.

Working modes of switch chips vary according to product models. During the expansion from a single chassis or a BTB system to CCC-1 or CCC-2, the working modes of switch chips must be changed, and there is a possibility that 1-stage and 3-stage switching coexist. Due to the uniform cell format, packet processing will not be affected even if some switch chips work in S2 mode while others work in S13 mode. SFUs work in 3+1 load balancing mode. Therefore, when the working mode of the switch chips on one SFU is being changed, traffic on this SFU will be switched to the other three SFUs. After the working mode has been changed, the traffic can be switched back. The working mode of the switch chips on each of the other SFUs can be changed in the same manner, without interrupting services.
3.3 Control Plane Switchover

Key parts backup in a system is one of the major methods to provide fault tolerance capability for the system. Control plane switchover is based on the 1+1 backup of MPUs on an NE5000E. The hot backup function that the NE5000E supports allows reliable backup of data between the master and slave MPUs, preventing data loss during the master/slave switchover.

The following part uses the expansion from a single chassis to CCC-1 as an example. Before the expansion, the master MPU on the single chassis controls the entire system. After the expansion, the master MPU on the CCC in CCC-1 will control the entire system. The following method is used to hand over the system control power in a cluster:

As shown in the following figure, configure the master and slave MPUs on CLC1 as the system master and slave MPUs, and configure CCC1 to register with CLC1. After CCC1 has successfully registered, reset the slave MPU on CLC1. Then, select the master MPU on CCC1 as the system slave MPU. At this time, the master MPU on CLC1 is the system master MPU. After the slave MPU on CLC1 has registered, perform the master/slave switchover of system MPUs. After this inter-chassis MPU switchover is complete, the master MPU on CCC1 becomes the system master MPU. In addition, select the slave MPU on CCC1 as the system MPU. Till now the system control power has been handed over to the master MPU on CCC1. Resetting MPUs and performing the master/slave MPU switchover hand over the system control power and also perform hot backup of data.
Schematic diagram of system control power handover during the expansion from a single chassis to CCC-1
4 Typical Applications

4.1 Expansion from a Single NE5000E to a BTB System

1. Prerequisite: A carrier has already had an NE5000E. The carrier can purchase another NE5000E to expand the original single chassis to a BTB system (CCC-0) by connecting the two NE5000Es and performing necessary configurations.

2. Comparison between the data before and after expansion

<table>
<thead>
<tr>
<th>Chassis Mode</th>
<th>Number of LPUs</th>
<th>Switching Capacity (Tbit/s)</th>
<th>Interface Capacity (Tbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single chassis</td>
<td>16</td>
<td>6.4</td>
<td>3.2</td>
</tr>
<tr>
<td>CCC-0</td>
<td>32</td>
<td>12.8</td>
<td>6.4</td>
</tr>
</tbody>
</table>


4.2 Expansion from a Single NE5000E to CCC-1

1. Prerequisite: A carrier has already had an NE5000E. The carrier can purchase an NE5000E CCC and several CLCs to expand the original single chassis to CCC-1 (1+1, 1+2, 1+3, or 1+4 cluster) by connecting the NE5000E CCC and CLCs and performing necessary configurations.

2. Comparison between the data before and after expansion

<table>
<thead>
<tr>
<th>Chassis Mode</th>
<th>Number of LPUs</th>
<th>Switching Capacity (Tbit/s)</th>
<th>Interface Capacity (Tbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single chassis</td>
<td>16</td>
<td>6.4</td>
<td>3.2</td>
</tr>
<tr>
<td>CCC-1</td>
<td>16–64</td>
<td>25.6</td>
<td>3.2–12.8</td>
</tr>
</tbody>
</table>
4.3 Expansion from a Single NE5000E to CCC-2

1. Prerequisite: A carrier has already had an NE5000E. The carrier can purchase two NE5000E CCCs and several CLCs to expand the original single chassis to CCC-2 (2+1, 2+2, 2+4... or 2+8 cluster) by connecting the NE5000E CCCs and CLCs and performing necessary configurations.

2. Comparison between the data before and after expansion

<table>
<thead>
<tr>
<th>Chassis Mode</th>
<th>Number of LPUs</th>
<th>Switching Capacity (Tbit/s)</th>
<th>Interface Capacity (Tbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single chassis</td>
<td>16</td>
<td>6.4</td>
<td>3.2</td>
</tr>
<tr>
<td>CCC-2</td>
<td>16–128</td>
<td>51.2</td>
<td>3.2–25.6</td>
</tr>
</tbody>
</table>
4.4 Expansion from a BTB System to CCC-1

1. Prerequisite: A carrier has already had an NE5000E BTB system. The carrier can purchase an NE5000E CCC and several CLCs to expand the BTB system to CCC-1 (1+2, 1+3, or 1+4 cluster) by connecting the NE5000E CCC and CLCs and performing necessary configurations.

2. Comparison between the data before and after expansion

<table>
<thead>
<tr>
<th>Chassis Mode</th>
<th>Number of LPUs</th>
<th>Switching Capacity (Tbit/s)</th>
<th>Interface Capacity (Tbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC-0</td>
<td>32</td>
<td>12.8</td>
<td>6.4</td>
</tr>
<tr>
<td>CCC-1</td>
<td>32–64</td>
<td>25.6</td>
<td>6.4–12.8</td>
</tr>
</tbody>
</table>
4.5 Expansion from a BTB System to CCC-2

1. Prerequisite: A carrier has already had an NE5000E BTB system. The carrier can purchase two NE5000E CCCs and several CLCs to expand the BTB system to CCC-2 (2+2, 2+3, 2+4…or 2+8 cluster) by connecting the NE5000E CCCs and CLCs and performing necessary configurations.

2. Comparison between the data before and after expansion

<table>
<thead>
<tr>
<th>Chassis Mode</th>
<th>Number of LPUs</th>
<th>Switching Capacity (Tbit/s)</th>
<th>Interface Capacity (Tbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC-0</td>
<td>32</td>
<td>12.8</td>
<td>6.4</td>
</tr>
<tr>
<td>CCC-2</td>
<td>32–128</td>
<td>51.2</td>
<td>6.4–25.6</td>
</tr>
</tbody>
</table>
4.6 Expansion from CCC-1 to CCC-2

1. Prerequisite: A carrier has already had a CCC-1 system. The carrier can purchase an NE5000E CCC and several CLCs to expand CCC-1 to CCC-2 (2+1, 2+2, 2+3, 2+4…or 2+8 cluster) by connecting the NE5000E CCCs and CLCs and performing necessary configurations.

2. Comparison between the data before and after expansion

<table>
<thead>
<tr>
<th>Chassis Mode</th>
<th>Number of LPUs</th>
<th>Switching Capacity (Tbit/s)</th>
<th>Interface Capacity (Tbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCC-1</td>
<td>16–64</td>
<td>25.6</td>
<td>6.4–12.8</td>
</tr>
<tr>
<td>CCC-2</td>
<td>16–128</td>
<td>51.2</td>
<td>6.4–25.6</td>
</tr>
</tbody>
</table>

4.7 Expansion of CCC-1 or CCC-2 by Adding CLCs

1. Prerequisite: A carrier has already had a CCC-1 or CCC-2 system in which the number of CLCs does not reach the maximum. This means that there are less than four CLCs in CCC-1 and less than eight CLCs in CCC-2. The carrier can purchase several CLCs and add them to CCC-1 or CCC-2.

2. The switching capacity remains unchanged, whereas the number of LPUs and interface capacity increase.
5 Conclusion

The innovative ISHE solution provided by Huawei allows smooth capacity expansion without hardware replacement, ensuring high system reliability and maximizing carriers’ investment income. In this sense, ISHE promotes the evolution of core nodes from single chassis to clusters.
### A  Acronyms and Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation/Acronym</th>
<th>Full Spelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISHE</td>
<td>In-Service Hardware Expansion</td>
</tr>
<tr>
<td>CLC</td>
<td>Cluster Line-card Chassis</td>
</tr>
<tr>
<td>CCC</td>
<td>Cluster Central Chassis</td>
</tr>
<tr>
<td>ICU</td>
<td>Internal Communication Unit</td>
</tr>
<tr>
<td>SFU</td>
<td>Switch Fabric Unit</td>
</tr>
<tr>
<td>SFEA</td>
<td>Switch Fabric Enhanced unit version A</td>
</tr>
<tr>
<td>ECU</td>
<td>Electrical Cross Unit</td>
</tr>
<tr>
<td>VRP</td>
<td>Versatile Routing Platform</td>
</tr>
<tr>
<td>TM</td>
<td>Traffic Manager</td>
</tr>
</tbody>
</table>