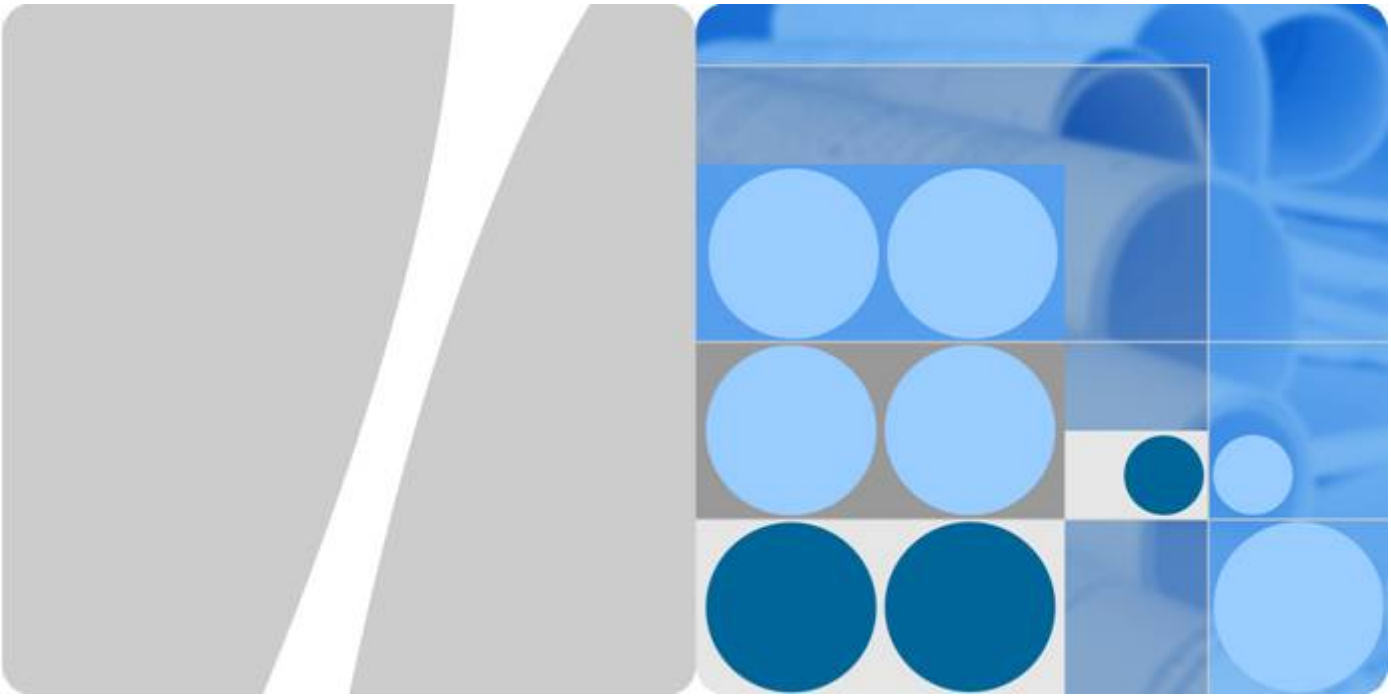


Huawei MZ612 NIC

V100R001

White Paper

Issue 06

Date 2016-11-21

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


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

Notice

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

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

Huawei Technologies Co., Ltd.

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

Website: http://e.huawei.com

About This Document

Purpose

This document describes the MZ612 in terms of its functions, appearance, features, applications, and technical specifications. You can obtain comprehensive information about the MZ612 by reading this document.

Intended Audience

This document is intended for:

- Huawei presales engineers
- Channel partner presales engineers
- Enterprise presales engineers

Symbol Conventions

The symbols that may be found in this document are defined as follows.

Symbol	Description
 DANGER	Alerts you to a high risk hazard that could, if not avoided, result in serious injury or death.
 WARNING	Alerts you to a medium or low risk hazard that could, if not avoided, result in moderate or minor injury.
 CAUTION	Alerts you to a potentially hazardous situation that could, if not avoided, result in equipment damage, data loss, performance deterioration, or unanticipated results.
 NOTE	Provides additional information to emphasize or supplement important points in the main text.

Change History

Issue 06 (2016-11-21)

This issue is the sixth official release.

Issue 05 (2016-08-17)

This issue is the fifth official release.

Issue 04 (2016-05-15)

This issue is the fourth official release.

Type	Change Description
Modify	The document content is optimized.

Issue 03 (2016-03-31)

This issue is the third official release.

Issue 02 (2015-11-09)

This issue is the second official release.

Issue 01 (2015-05-30)

This issue is the first official release.

Contents

Figures

Tables

1 Overview

About This Chapter

1.1 Functions

The MZ612 is an InfiniBand (IB) host channel adapter (HCA). It is used for E9000 compute nodes and provides two 56G IB fourteen data rate (FDR) ports for compute nodes to connect to switch modules in the chassis.

The MZ612 uses the Mellanox ConnectX-3 (CX3) chip and supports HCA applications. The 56G ports on the MZ612 support 56G/40G auto-negotiation and support 56G (FDR) and 40G (QDR) port applications. The MZ612 also supports the Remote Direct Memory Access (RDMA) feature to address low-latency network applications.

1.2 Appearance

The MZ612 can be installed in slot Mezzanine1 (Mezz1 for short) or Mezzanine2 (Mezz2 for short) on a half-width E9000 compute node.

The MZ612 provides network ports for connecting to switch modules:

- When the MZ612 is installed in slot Mezz1, its two 56GE ports connect to switch modules in slots 2X and 3X.
- When the MZ612 is installed in slot Mezz2, its two 56GE ports connect to switch modules in slots 1E and 4E.

Figure 1-1 MZ612 appearance

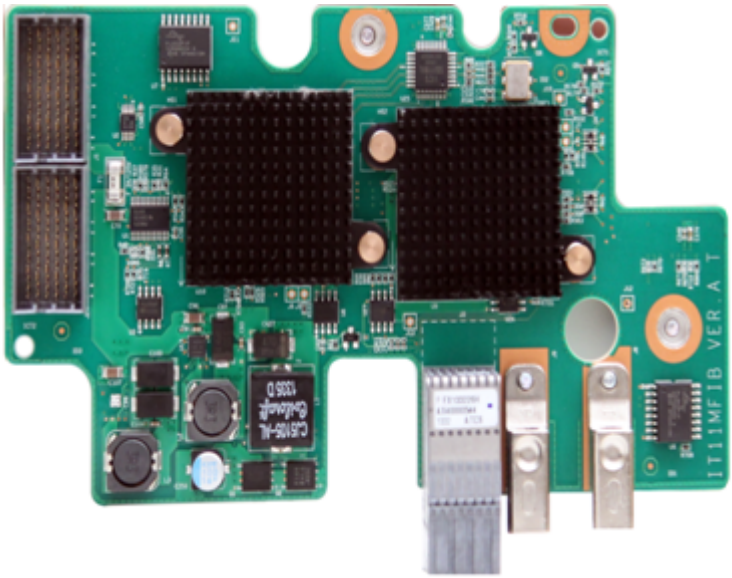
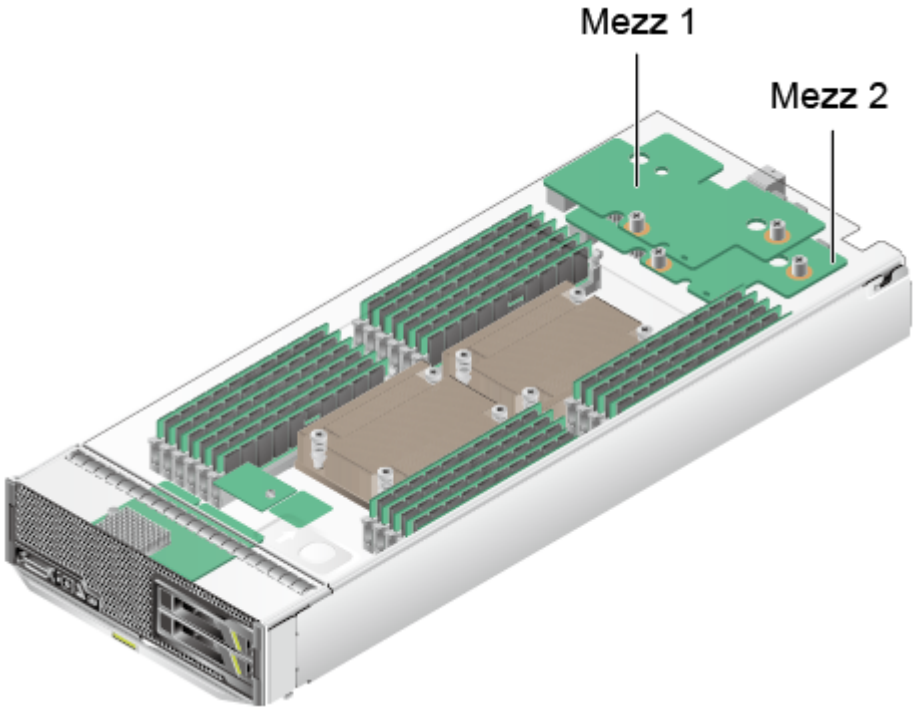


Figure 1-2 MZ612 installation positions on a half-width compute node



2 Features

About This Chapter

2.1 Feature List

The MZ612 supports the following features and performance specifications:

- InfiniBand Trade Association (IBTA) 1.2.1 specifications
- RDMA
- 32 million (2 x 16 million) I/O channels
- End-to-end QoS and 18 virtual lanes (VLs), including 16 data VLs and two control VLs
- Hardware-based congestion control
- In-band management and support for third-party Subnet Managers (SMs)

2.2 Feature Description

RDMA

The MZ612 supports the RDMA feature. This feature uses the kernel bypass technology to reduce the packet processing and forwarding latency of the HCA, to reduce the CPU usage, and to implement low-latency data transmission over the data center network. With the RDMA feature, the end-to-end read and write delay of the HCA can reach 1 us when the packet length is 128 bytes. The MZ613 supports 32 million (2 x 16 million) I/O channels (equivalent to IB QPs), and provides priority-based scheduling and flow control to support low-latency, high-bandwidth network transmission. The MZ612 supports OpenFabrics Enterprise Distribution for Linux (Linux OFED) and Mellanox OFED for Windows (WinOF).

QoS

The MZ612 supports end-to-end QoS for IB. It supports the VL mechanism defined in IB specifications, VL arbitration, and control-domain and service-domain prioritizing. The end-to-end QoS priority is represented by the service level (SL) domain of IB packets. Each SL is mapped to a VL. SMs configure VL arbitration and the SL-to-VL mapping table using network management datagrams (MADs). The MZ612 supports 18 VLs, including 16 data VLs and two control VLs.

2.3 Standards Compliance

Table 2-1 lists the standards and protocols that the MZ612 complies with.

Table 2-1 Standards compliance

Standard	Protocol
IBTA 1.2.1	InfiniBand Trade Association
ANSI INCITS 365-2002	SCSI RDMA Protocol (SRP)
IETF	iSCSI Extensions for RDMA
DAPL	User Direct Access Programming Library (uDAPL)

3 Applications

About This Chapter

3.1 Compatible Compute Nodes

The MZ612 can be installed in slot Mezz1 or Mezz2 on a half-width compute node. **Table 3-1** lists the compute nodes that support the MZ612 and its installation positions on them.

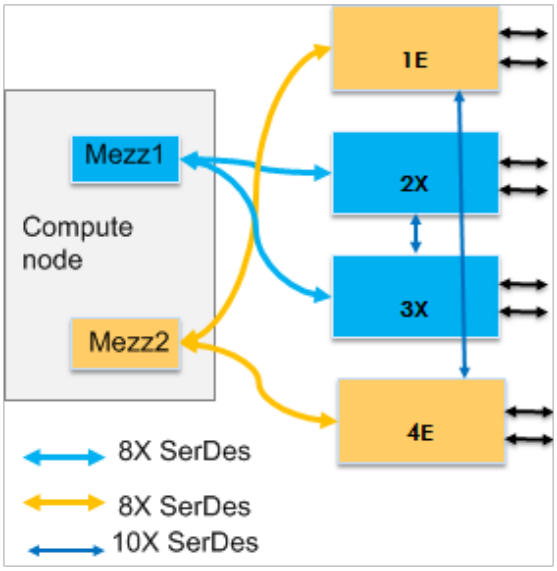
Table 3-1 Compute nodes that support the MZ612

Compute Node	Number of Mezz Module Slots	MZ612 Installation Position
CH140	2	Mezz1 and Mezz2
CH140 V3	2	Mezz1 and Mezz2

3.2 Connected I/O Modules

MZ612s can connect to I/O modules (switch modules or interface boards). **Figure 3-1** shows the connections between the MZ612s on a half-width compute node and the ports on I/O modules.

Figure 3-1 Connections between the MZ612s on a half-width compute node and the ports on I/O modules



There are two or four groups of Serializer/Deserializer (SerDes, known as high-speed interconnect line) between each compute node and I/O module slots.

- Mezz1: 8X SerDes for connecting to I/O module slots 2X and 3X
- Mezz2: 8X SerDes for connecting to I/O module slots 1E and 4E
- Mezz3 (available only on a full-width compute node): 8X SerDes for connecting to I/O module slots 2X and 3X

- Mezz4 (available only on a full-width compute node): 8X SerDes for connecting to I/O module slots 1E and 4E

NOTE

The MZ612 provides two ports, and only 4X of each 8X SerDes is used.

Table 3-2 describes the I/O modules to which the MZ612 can connect.

Table 3-2 I/O modules to which the MZ612 can connect

I/O Module	I/O Module Slot	MZ612 (Mezz1)	MZ612 (Mezz2)	Typical Configuration	Remarks
CX610	2X/3X	√	X	No	-
	1E/4E	X	√	No	-
CX611	2X/3X	√	X	No	-
	1E/4E	X	√	Yes	You are advised to install CX611s in slots 1E and 4E.

3.3 MZ612 Networking

The MZ612 can connect to I/O modules (switch modules or interface boards) to provide IB services.

The MZ612 can work with the CX610 switch module to provide 40 Gbit/s bandwidth, and connect to the external IB network through 40G QDR ports on the CX610. See Figure 3-2.

Figure 3-2 Connection between the MZ612 and the CX610

The MZ612 can work with the CX611 switch module to provide 56 Gbit/s bandwidth with its two network ports, and connect to the external IB network through 56G FDR ports on the CX611. See Figure 3-3.

Figure 3-3 Connection between the MZ612 and the CX611

3.4 Supported OSs

Table 3-3 lists the OSs supported by the MZ612.

Table 3-3 OSs supported by the MZ612

OS	Version	Remarks
Windows	Windows Server 2008 Enterprise SP2	-
	Windows Server 2008 R2 Enterprise x64	-
	Windows Server 2012 Enterprise x64	-
Red Hat	Red Hat Enterprise Linux (RHEL) 6.1 x86	-
	RHEL 6.1 x86_64	-
	RHEL 6.3 x86	-
	RHEL 6.3 x86_64	-
SUSE	SUSE Linux Enterprise Server (SLES) 11 SP1 x86	-
	SLES 11 SP1 x86_64	-
	SLES 11 SP2 x86	-
	SLES 11 SP2 x86_64	-
VMware	VMware ESXi 5.0	-
	VMware ESXi 5.1	-
	VMware ESXi 5.5	-
Oracle Linux	Oracle Linux 6.2 x86	-
	Oracle Linux 6.2 x86_64	-
	Oracle Linux 6.3 x86	-
	Oracle Linux 6.3 x86_64	-

For the latest versions of OSs, see the ***Huawei Server Compatibility Checker***.

4 Technical Specifications

About This Chapter

4.1 Technical Specifications

Table 4-1 lists the technical specifications for the MZ612.

Table 4-1 Technical specifications

Item	Specifications
Dimensions (length x width)	148 mm x 85 mm (5.83 in. x 3.35 in.)
Power supply	12 V DC
Net weight	0.5 kg (1.10 lb)
Maximum power consumption	14 W
Temperature	Operating temperature: 5°C to 40°C (41°F to 104°F) (ASHRAE Class A3 compliant)
	Storage temperature: –40°C to +65°C (–40°F to +149°F)
Temperature change rate	15°C/h (27°F/h)
Humidity	Operating humidity: 5% RH to 85% RH (non-condensing)
	Storage humidity: 5% RH to 95% RH (non-condensing)
Altitude	● 40°C (104°F) at 1800 m (5905.44 ft) ● 30°C (86°F) at 3000 m (9842.40 ft) When the MZ612 is used at an altitude between 1800 m and 3000 m, the highest operating temperature decreases by 1°C (1.8°F) as the altitude increases by 120 m (393.70 ft).
PCIe port bandwidth	2 x 64 Gbit/s (2 x PCIe 3.0 x8)
Port rate	40 Gbit/s or 56.25 Gbit/s
Number of ports	2
Port type	IB
Chip model/manufacturer	ConnectX-3 (CX3)/Mellanox

A Acronyms and Abbreviations

F	
FDR	fourteen data rate
H	
HCA	host channel adapter
I	
IB	InfiniBand
IBTA	InfiniBand Trade Association
IO	input/output
M	
MAD	Management Datagram
O	
OFED	OpenFabrics Enterprise Distribution
OS	operating system
P	
PCIe	Peripheral Component Interconnect Express
Q	
QDR	quad data rate
QoS	quality of service
QP	queue pair

R	
RDMA	Remote Direct Memory Access
S	
SL	service level
SM	Subnet Manager
V	
VL	virtual lane