IxNetwork RoCEv2 Test Solution

Validate Data Center Fabric delivering the best performance for AI training workload

Introduction

Training large AI model has driven the growth of cluster size and training workload. This involves many compute nodes of Servers with GPUs doing parallel computing with collective communication operations among these devices. The network connecting these devices needs to provide high bandwidth throughput, low latency, and lossless traffic.

The AI training network design can be 2-tiers or 3-tiers depends on required scale and design choice. Care needs to be taken for ECMP hashing, PFC deadlock and end-to-end communication latency. To validate and benchmark the AI network fabric performance, switch fabric needs to exercise RoCE Congestion Control and Priority Flow Control (PFC) to optimize buffer management for AI/ML workload.

Keysight RoCEv2 Lossless Ethernet Test solution includes high-density cost-effect test platform and IxNetwork test application. It emulates Queue-Pair (QP) connections and flows, generates congestion notification, performs DCQCN based dynamic rate control, as well as provides needed flexibility to test throughput, buffer management and ECMP hashing for optimizing the fabric performance delivering AI training workload. It provides scalable and cost-effective solution to validate the effectiveness of congestion control and benchmark fabric performance.

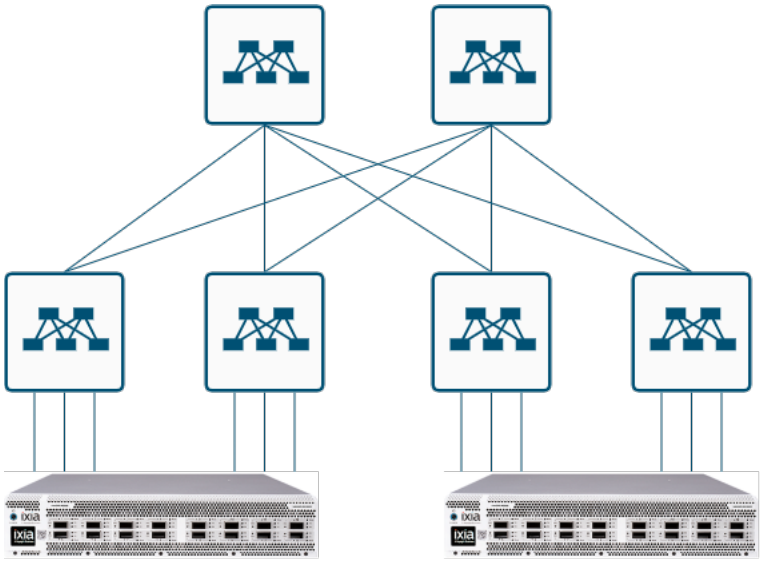


Figure 1. Keysight RoCEv2 Lossless Ethernet Test Solution

Highlights

- Future proof high-density and cost-effective 400GE Test Platform
- Hardware-based ECN/CNP congestion notification and DCQCN rate control per Q-Pair
- Stress AI network fabric with realistic RoCEv2 traffic emulating AI workload
- Comprehensive statistics to help troubleshooting and accelerate time-to-market

Key features

- Introduces new FPGA mode supporting RoCEv2 Queue-Pairs (QPs) flow traffic engine
- Supports 4x100GE NRZ and 8x100GE/4x200GE PAM4(56G) RoCEv2 speed modes per RG
- Emulates up to 4K Q-Pairs performing RDMA WRITE with Reliable Connection (RC) service type
- Assigns DSCP code per Q-Pair
- Auto-generates Q-Pair number or allows user configurable Q-Pair number
- Transfers up to 252MB buffer size at 4K MTU
- Detects ECN congestion signaling and generates CNP congestion notification
- Performs DCQCN rate control algorithm per Q-Pair with user tunable parameters
- Handles PFC Pause frame for traffic pause and resume
- Supports in-cast (N:1), M:N, and all-to-all traffic patterns with fixed or continuous transmission
- Control traffic rate in % of line rate or inter-batch-gap
- Provides per port statistic and per QP RoCEv2 statistic
- Support 1500 bytes to 14K bytes MTU
- Comprehensive TCL, Python/REST API support for automation

High-density, Cost-effective test platform

AresONE-S 400GE QSFP-DD 16-port fixed chassis system is the industry’s highest density 400GE test platform. It supports 16 x 400GE/ 32 x 200GE/ 64 x 100GE PAM4 56G speeds, as well as 100GE NRZ speed, is an ideal platform for AI Fabric validation.

The RoCEv2 FPGA is a selectable mode to enable RoCEv2 flow engine per Resource Group (RG). Each RG supports 2 x 400GE, 4 x 200GE and 8x 100GE PAM4 speeds, and 4 x 100GE NRZ speed. It detects ECN-CE, generating CNP congestion notification, and performing DCQCN rate control.

The test platform support multi-user with up to 8 users per system. It supported both regular L23 control and data plane test, and RoCEv2 test concurrently across different RGs, enable efficient usage and reduce cost of ownership.

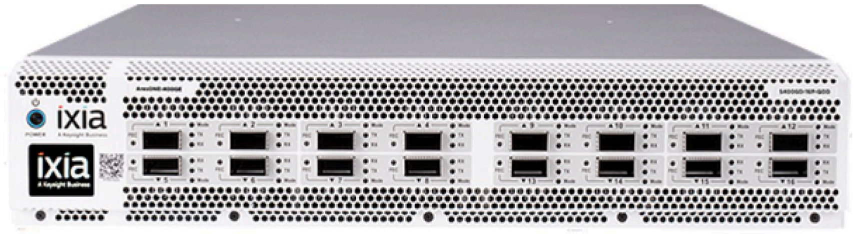


Figure 2. Keysight RoCEv2 Lossless Ethernet Test Solution - AresONE-S 400GE QSFP-DD

RDMA Endpoint Emulation

IxNetwork emulates RDMA endpoints establishing Q-Pair connections, performing RDMA WRITE operation in Reliable Connection (RC) mode, generating in-cast (N:1), M:N, and all-to-all traffic patterns with fixed or continuous transmission, and providing per Q-Pair RoCEv2 statistics.

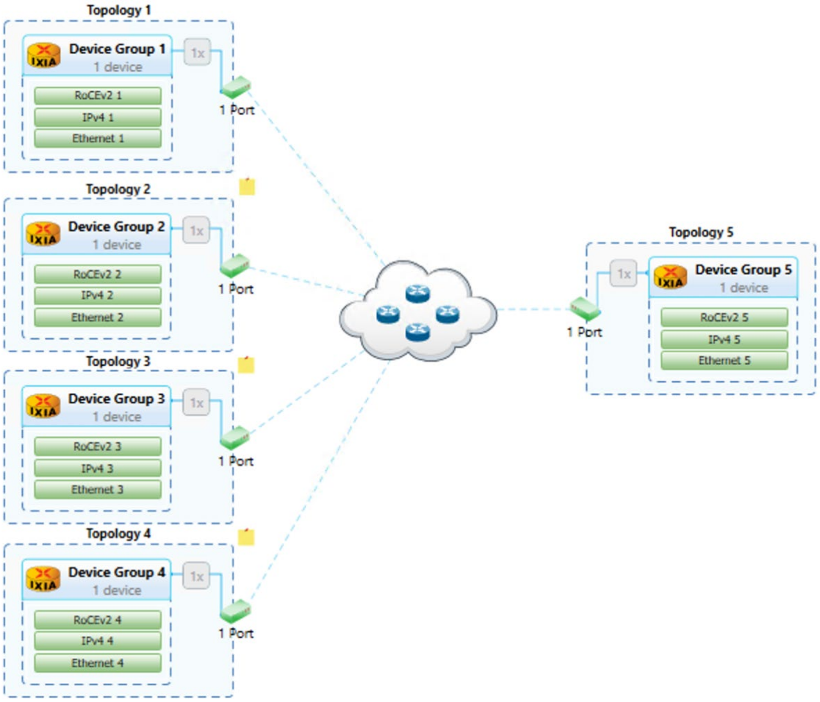
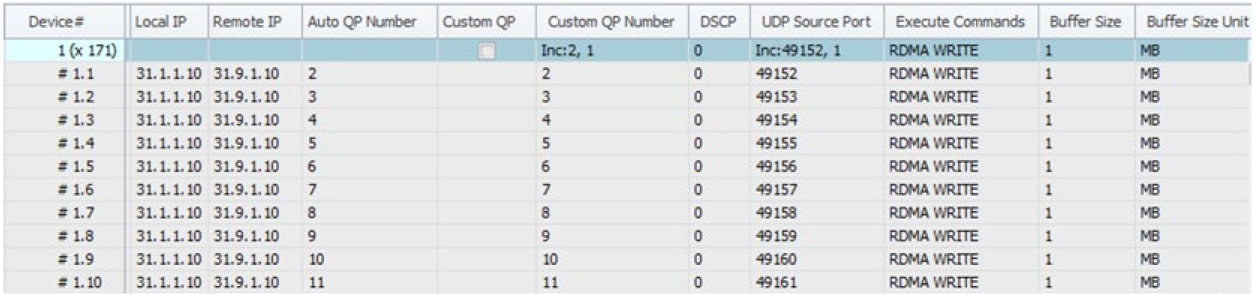


Figure 3. IxNetwork RoCEv2 Endpoints Emulation – 4:1 In-cast

Q-Pair configuration auto-generates Q-Pair number or allows user configurable Q-Pair number. DSCP can be mapped at per QP level and buffer size supports up to 256MB at 4K MTUs.

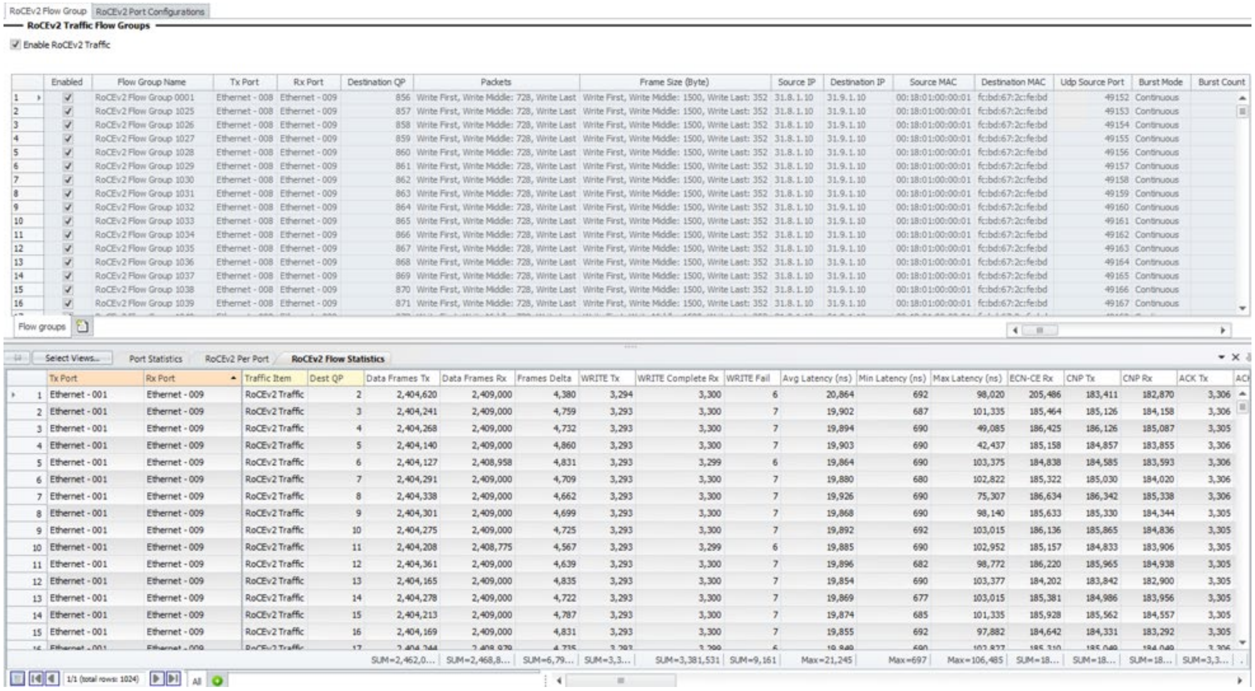


Device#	Local IP	Remote IP	Auto QP Number	Custom QP	Custom QP Number	DSCP	UDP Source Port	Execute Commands	Buffer Size	Buffer Size Unit
1 (x 171)				☐	Inc:2, 1	0	Inc:49152, 1	RDMA WRITE	1	MB
# 1.1	31.1.1.10	31.9.1.10	2		2	0	49152	RDMA WRITE	1	MB
# 1.2	31.1.1.10	31.9.1.10	3		3	0	49153	RDMA WRITE	1	MB
# 1.3	31.1.1.10	31.9.1.10	4		4	0	49154	RDMA WRITE	1	MB
# 1.4	31.1.1.10	31.9.1.10	5		5	0	49155	RDMA WRITE	1	MB
# 1.5	31.1.1.10	31.9.1.10	6		6	0	49156	RDMA WRITE	1	MB
# 1.6	31.1.1.10	31.9.1.10	7		7	0	49157	RDMA WRITE	1	MB
# 1.7	31.1.1.10	31.9.1.10	8		8	0	49158	RDMA WRITE	1	MB
# 1.8	31.1.1.10	31.9.1.10	9		9	0	49159	RDMA WRITE	1	MB
# 1.9	31.1.1.10	31.9.1.10	10		10	0	49160	RDMA WRITE	1	MB
# 1.10	31.1.1.10	31.9.1.10	11		11	0	49161	RDMA WRITE	1	MB

Figure 4. IxNetwork RoCEv2 Q-Pair Configuration

One-click Q-Pair flow generation provides QP flow details. User can control traffic rate either using % of line rate or inter batch period.

RoCEv2 per QP stats provide RDMA WRITE operation count with successful and fail operation, packet count and latency, ECN/CNP/ACK/NAK counters to help verify congestion and troubleshoot failures.



RoCEv2 Flow Group | RoCEv2 Port Configurations

RoCEv2 Traffic Flow Groups

☑ Enable RoCEv2 Traffic

	Enabled	Flow Group Name	Tx Port	Rx Port	Destination QP	Packets	Frame Size (Byte)	Source IP	Destination IP	Source MAC	Destination MAC	Udp Source Port	Burst Mode	Burst Count
1	☑	RoCEv2 Flow Group 0001	Ethernet - 008	Ethernet - 009	856	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49152	Continuous	
2	☑	RoCEv2 Flow Group 1025	Ethernet - 008	Ethernet - 009	857	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49153	Continuous	
3	☑	RoCEv2 Flow Group 1026	Ethernet - 008	Ethernet - 009	858	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49154	Continuous	
4	☑	RoCEv2 Flow Group 1027	Ethernet - 008	Ethernet - 009	859	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49155	Continuous	
5	☑	RoCEv2 Flow Group 1028	Ethernet - 008	Ethernet - 009	860	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49156	Continuous	
6	☑	RoCEv2 Flow Group 1029	Ethernet - 008	Ethernet - 009	861	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49157	Continuous	
7	☑	RoCEv2 Flow Group 1030	Ethernet - 008	Ethernet - 009	862	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49158	Continuous	
8	☑	RoCEv2 Flow Group 1031	Ethernet - 008	Ethernet - 009	863	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49159	Continuous	
9	☑	RoCEv2 Flow Group 1032	Ethernet - 008	Ethernet - 009	864	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49160	Continuous	
10	☑	RoCEv2 Flow Group 1033	Ethernet - 008	Ethernet - 009	865	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49161	Continuous	
11	☑	RoCEv2 Flow Group 1034	Ethernet - 008	Ethernet - 009	866	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49162	Continuous	
12	☑	RoCEv2 Flow Group 1035	Ethernet - 008	Ethernet - 009	867	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49163	Continuous	
13	☑	RoCEv2 Flow Group 1036	Ethernet - 008	Ethernet - 009	868	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49164	Continuous	
14	☑	RoCEv2 Flow Group 1037	Ethernet - 008	Ethernet - 009	869	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49165	Continuous	
15	☑	RoCEv2 Flow Group 1038	Ethernet - 008	Ethernet - 009	870	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49166	Continuous	
16	☑	RoCEv2 Flow Group 1039	Ethernet - 008	Ethernet - 009	871	Write First, Write Middle: 728, Write Last	Write First, Write Middle: 1500, Write Last: 352	31.8.1.10	31.9.1.10	00:18:01:00:00:01	fc:bd:67:2c:fe:bd	49167	Continuous	

Flow groups

	Tx Port	Rx Port	Traffic Item	Dest QP	Data Frames Tx	Data Frames Rx	Frames Delta	WRITE Tx	WRITE Complete Rx	WRITE Fail	Avg Latency (ns)	Min Latency (ns)	Max Latency (ns)	ECN-CE Rx	CNP Tx	CNP Rx	ACK Tx
1	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	2	2,404,620	2,409,000	4,380	3,294	3,300	6	20,864	692	98,020	205,486	183,411	182,870	3,306
2	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	3	2,404,241	2,409,000	4,759	3,293	3,300	7	19,902	687	101,335	185,464	185,126	184,158	3,306
3	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	4	2,404,268	2,409,000	4,732	3,293	3,300	7	19,894	690	49,085	186,425	186,126	185,087	3,305
4	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	5	2,404,140	2,409,000	4,860	3,293	3,300	7	19,903	690	42,437	185,158	184,857	183,855	3,306
5	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	6	2,404,127	2,408,958	4,831	3,293	3,299	6	19,864	690	103,375	184,838	184,585	183,593	3,306
6	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	7	2,404,291	2,409,000	4,709	3,293	3,300	7	19,880	680	102,822	185,322	185,030	184,020	3,306
7	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	8	2,404,338	2,409,000	4,662	3,293	3,300	7	19,926	690	75,307	186,634	186,342	185,338	3,306
8	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	9	2,404,301	2,409,000	4,699	3,293	3,300	7	19,868	690	98,140	185,633	185,330	184,344	3,305
9	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	10	2,404,275	2,409,000	4,725	3,293	3,300	7	19,892	692	103,015	186,136	185,865	184,836	3,305
10	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	11	2,404,208	2,408,775	4,567	3,293	3,299	6	19,885	690	102,952	185,157	184,833	183,906	3,305
11	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	12	2,404,361	2,409,000	4,639	3,293	3,300	7	19,896	682	98,772	186,220	185,965	184,938	3,305
12	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	13	2,404,165	2,409,000	4,835	3,293	3,300	7	19,854	690	103,377	184,202	183,842	182,900	3,305
13	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	14	2,404,278	2,409,000	4,722	3,293	3,300	7	19,869	677	103,015	185,381	184,986	183,956	3,305
14	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	15	2,404,213	2,409,000	4,787	3,293	3,300	7	19,874	685	101,335	185,928	185,562	184,557	3,305
15	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	16	2,404,169	2,409,000	4,831	3,293	3,300	7	19,855	692	97,882	184,642	184,331	183,292	3,305
16	Ethernet - 001	Ethernet - 009	RoCEv2 Traffic	17	2,404,244	2,408,979	4,735	3,293	3,299	6	19,849	690	102,827	185,310	185,049	184,049	3,306
					SUM=2,462,0...	SUM=2,468,8...	SUM=6,79...	SUM=3,3...	SUM=3,381,531	SUM=9,161	Max=21,245	Max=697	Max=106,485	SUM=18...	SUM=18...	SUM=18...	SUM=3,3...

1/1 (total rows: 1024)

Specifications

ISIS segment routing

Hardware platform	• AresONE-S 400GE QSFP-DD 16-port fixed chassis model • AresONE-S 400GE QSFP-DD 8-port fixed chassis model • AresONE-M 800GE QSFP-DD800 8-port fixed chassis model • AresONE-M 800GE QSFP-DD800 4-port fixed chassis model
Ethernet speeds	• 100GE NRZ • 100GE PAM4 56G • 200GE PAM4 56G
Q-Pairs configuration	• Local and remote IP • Auto QP Number or custom QP Number • DSCP mapping • Execute Command: RDMA WRITE • Buffer Size and unit • Connection: Connect Request, Connect Reply, ReadyToUse
Congestion control	• ECN-CE detection • CNP generation and DSCP priority • DCQCN Rate Control Parameters • CNP Delay Timer
Traffic flow configuration	• Q-Pair Mesh: In-cast (N:1), All-to-all, Partial mesh (M:N) • Burst mode: Fixed, Continuous • Rate: Target % Line Rate, Inter batch Period • DCQCN Setting
Statistics	• Packet count and Packet latency • RDMA WRITE Count: Complete or Fail • ECN Rx, CNP Tx/Rx, ACK Tx/Rx, NAK Tx/Rx • Sequency error

Ordering Information

RoCEv2 part numbers

Part number	Description
905-1092	Keysight RoCEv2 Lossless Ethernet Enablement FACTORY INSTALLED Option for AresONE-S 400GE and AresONE-M 800GE fixed chassis models (905-1092)
905-1093	Keysight RoCEv2 Lossless Ethernet Enablement FIELD UPGRADE Option for AresONE-S and AresONE-M fixed chassis models (905-1093)
930-2208	Keysight IxNetwork RoCEv2 Lossless Ethernet Test Package for AresONE-S 400GE and AresONE-M 800GE fixed chassis models (930-2208)

RoCEv2 bundles

Part number	Description
947-4071	Keysight RoCEv2 Lossless Ethernet Test Bundle for AresONE-S 400GE QSFP-DD 16-port fixed chassis model (947-4071)
947-4072	Keysight RoCEv2 Lossless Ethernet Test Bundle for AresONE-S 400GE QSFP-DD 8-port fixed chassis model (947-4072)
947-4073	Keysight RoCEv2 Lossless Ethernet Test Bundle for AresONE-M 800GE QSFP-DD800 8-port fixed chassis model (947-4073)
947-4074	Keysight RoCEv2 Lossless Ethernet Test Bundle for AresONE-M 800GE QSFP-DD800 4-port fixed chassis model (947-4074)

KEYSIGHT

Keysight enables innovators to push the boundaries of engineering by quickly solving design, emulation, and test challenges to create the best product experiences. Start your innovation journey at www.keysight.com.

This information is subject to change without notice. © Keysight Technologies, 2024, Published in USA, February 21, 2024, 3124-1115.EN