Keysight AI Fabric RoCEv2 Test Solution

Validate data center fabrics performance for AI training workload

Introduction

Training large AI models requires thousands of AI-accelerated compute nodes processing in parallel and using collective communication operations to share data and results. In many cases, the time and cost to complete the training task is determined by the network performance supporting these collective communications. Keysight's AI Fabric Test Solution recreates these collective communication patterns so engineers can optimize network performance under AI training workloads.

To ensure the most timely and cost-effective platform for collective communications, the underlying network must provide high bandwidth throughput, low latency, and lossless traffic. Care needs to be taken for ECMP hashing, PFC deadlock, and end-to-end communication latency. To validate and benchmark the AI network fabric's performance, the fabric needs to exercise RoCE Congestion Control and Priority Flow Control (PFC) to optimize buffer management.

The Keysight AI Fabric Test Solution includes a high-density cost-effective test hardware platform (AresONE-M 800GE or AresONE-S 400GE) and the IxNetwork test application. IxNetwork models the AI training workload running on the tester's target topology. The system creates traffic that results from collective communications emanating from simulated end points. This includes emulating Queue-Pair (QP) connections and flows, generating congestion notifications, performing DCQCN-based dynamic rate control, and providing flexibility to test throughput, buffer management and ECMP hashing. The combination allows engineers to optimize the fabric's performance under the stress of the target AI workload and resulting collective communication patterns.

KEYSIGHT

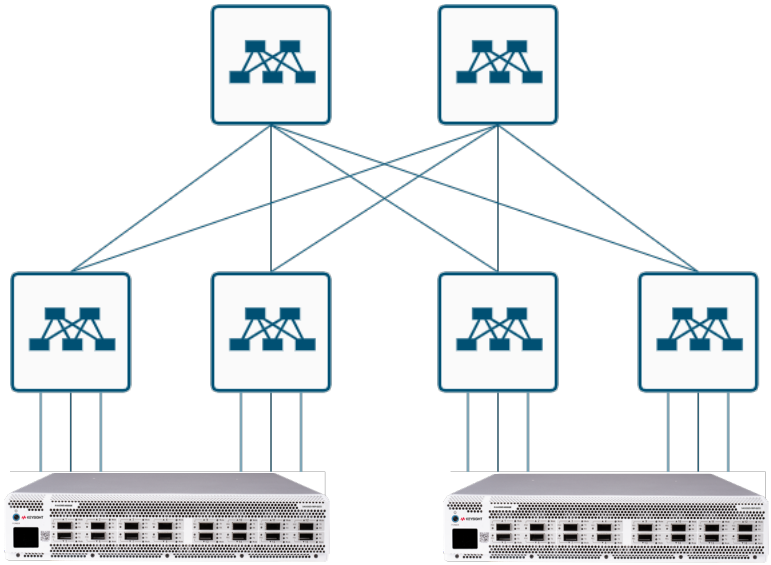

Figure 1. Keysight AI Fabric Test Solution

Highlights

- Stress AI fabric with realistic RoCEv2 traffic emulating AI workload
- Comprehensive statistics to help troubleshooting and accelerate time-to-market
- Hardware-based ECN/CNP congestion notification and DCQCN rate control per Q-Pair
- Future proof high-density and cost-effective 400GE and 800GE test platform

Key features

- Supports RoCEv2 over 800/400/200/100GE (106G, 53G PAM4) and100G NRZ links
- Emulates RoCEv2 endpoints over IPv4 and IPv6 transport
- Performs RDMA WRITE with Reliable Connection (RC) service type
- Generate in-cast (N:1), M:N, and All-to-all traffic patterns with fixed or continuous transmission
- Provides user configurable Q-Pair number and DSCP code
- Detects ECN congestion signaling and generates CNP congestion notification
- Performs DCQCN rate control algorithm per Q-Pair with user tunable parameters
- Responds to PFC Pause frame for to pause or resume traffic
- Auto-generates PFC pause with simulated buffer size and threshold to test PFC backpressure
- Provides per test, per port and per QP RoCEv2 statistics
- Allows user configurable Ethernet MTU and IB MTU
- Provides comprehensive TCL, Python/REST API support for automation

High-density, cost-effective test platform

Keysight supports the AI Fabric RoCEv2 Test Solution on both 800GE and 400GE hardware platforms. The solution installs an additional selectable mode on the hardware enabling the RoCEv2 flow engine. The engine not only generates traffic emanating from created Q-pairs but also detects ECN-CE, generates CNP congestion notifications, and performs DCQCN rate control. Both platforms support regular L23 control and data plane test, and RoCEv2 test, concurrently across different resource groups, enabling efficient usage and reducing the cost of ownership.

AresONE 800GE-M 8-port system is the industry's highest density 800GE platform and supports both QSFP-DD800 and OSFP800 interfaces. The Keysight AI Fabric RoCEv2 Test Solution is supported in 800/400/200/100GE speed mode (both 106G and 53G PAM4 electrical lanes), and 100GE NRZ mode (25G electrical lanes).

Figure 2. Keysight AI Fabric Test Solution - AresONE-M QSFP-DD800 and OSFP800

AresONE-S 400GE QSFP-DD 16-port fixed chassis system supports the AI Fabric RoCEv2 Test solution in 400/200/100GE speed mode (both 106G and 53G PAM4 electrical lanes), and 100GE NRZ mode (25G electrical lanes).

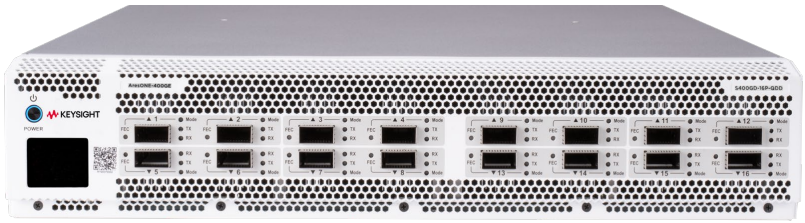

Figure 3. Keysight AI Fabric Test Solution - AresONE-S 400GE QSFP-DD

RDMA Endpoint Emulation

The Keysight AI Fabric RoCEv2 Test Solution recreates the traffic resulting from the collective communications during AI training jobs. It emulates RDMA endpoints establishing Q-Pair connections, performs RDMA WRITE operations in Reliable Connection (RC) mode, with fixed or continuous transmission, and provides per test, per port and per Q-Pair RoCEv2 statistics.

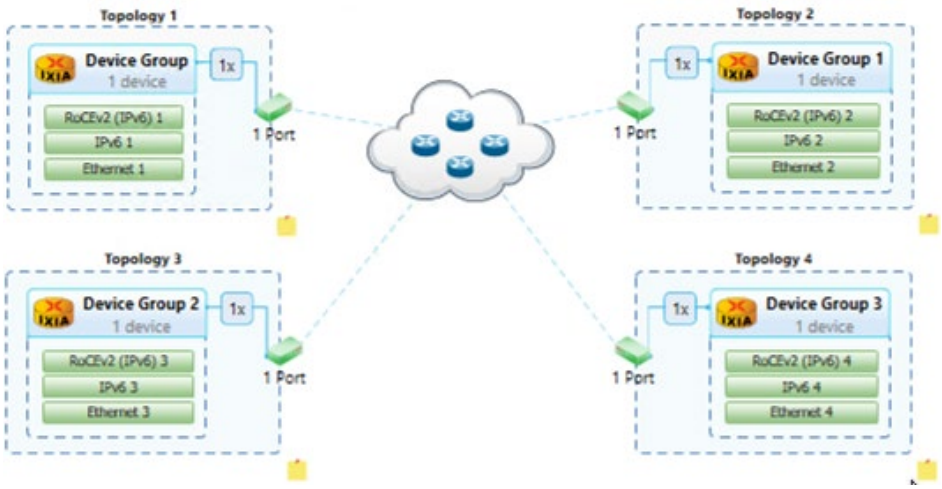

Figure 4. Keysight AI Fabric RoCEv2 Test Solution Endpoints Emulation – 4 ports all-to-all

Q-Pair configuration auto-generates Q-Pair number or allows user configurable Q-Pair number. DSCP code can be mapped at per QP level and Message size can be configured.

Device#	Local Peer	Remote Peer	Local IP	Remote IP	Auto QP Number	Custom QP	Custom QP Number	DSCP	ECN	UDP Source Port	Execute Commands	Message Size	Message Size Unit
1 (x 3)	RoCEv2 (IPv6) 1						Inc:2, 1	24	1	Inc:49152, 1	RDMA WRITE	1	MB
# 1.1		RoCEv2 (IPv6) 2	201::1:2	202::1:2	2		2	24	1	49152	RDMA WRITE	1	MB
# 1.2		RoCEv2 (IPv6) 3	201::1:2	203::1:2	3		3	24	1	49153	RDMA WRITE	1	MB
# 1.3		RoCEv2 (IPv6) 4	201::1:2	204::1:2	4		4	24	1	49154	RDMA WRITE	1	MB
1 (x 3)	RoCEv2 (IPv6) 2						Inc:2, 1	24	1	Inc:49152, 1	RDMA WRITE	1	MB
# 1.1		RoCEv2 (IPv6) 1	202::1:2	201::1:2	5		2	24	1	49152	RDMA WRITE	1	MB
# 1.2		RoCEv2 (IPv6) 3	202::1:2	203::1:2	6		3	24	1	49153	RDMA WRITE	1	MB
# 1.3		RoCEv2 (IPv6) 4	202::1:2	204::1:2	7		4	24	1	49154	RDMA WRITE	1	MB
1 (x 3)	RoCEv2 (IPv6) 3						Inc:2, 1	24	1	Inc:49152, 1	RDMA WRITE	1	MB
# 1.1		RoCEv2 (IPv6) 1	203::1:2	201::1:2	8		2	24	1	49152	RDMA WRITE	1	MB
# 1.2		RoCEv2 (IPv6) 2	203::1:2	202::1:2	9		3	24	1	49153	RDMA WRITE	1	MB
# 1.3		RoCEv2 (IPv6) 4	203::1:2	204::1:2	10		4	24	1	49154	RDMA WRITE	1	MB
1 (x 3)	RoCEv2 (IPv6) 4						Inc:2, 1	24	1	Inc:49152, 1	RDMA WRITE	1	MB
# 1.1		RoCEv2 (IPv6) 1	204::1:2	201::1:2	11		2	24	1	49152	RDMA WRITE	1	MB
# 1.2		RoCEv2 (IPv6) 2	204::1:2	202::1:2	12		3	24	1	49153	RDMA WRITE	1	MB
# 1.3		RoCEv2 (IPv6) 3	204::1:2	203::1:2	13		4	24	1	49154	RDMA WRITE	1	MB

Figure 5. Keysight AI Fabric RoCEv2 Test Solution 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.

Tx Port	Rx Port	Destination QP	Packets	Frame Size (Byte)	Source IP	Destination IP	Source MAC	Destination MAC	Udp Source Port	DSCP	Burst Mode	Burst Count
Ethernet - 001	Ethernet - 002	2	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	201::1:2	202::1:2	00:11:01:00:00:01	fc:bd:67:2c:fe:bd	49152	24	Continuous	1
Ethernet - 001	Ethernet - 003	3	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	201::1:2	203::1:2	00:11:01:00:00:01	fc:bd:67:2c:fe:bd	49153	24	Continuous	1
Ethernet - 001	Ethernet - 004	4	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	201::1:2	204::1:2	00:11:01:00:00:01	fc:bd:67:2c:fe:bd	49154	24	Continuous	1
Ethernet - 002	Ethernet - 001	5	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	202::1:2	201::1:2	00:12:01:00:00:01	fc:bd:67:2c:fe:bd	49152	24	Continuous	1
Ethernet - 002	Ethernet - 003	6	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	202::1:2	203::1:2	00:12:01:00:00:01	fc:bd:67:2c:fe:bd	49153	24	Continuous	1
Ethernet - 002	Ethernet - 004	7	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	202::1:2	204::1:2	00:12:01:00:00:01	fc:bd:67:2c:fe:bd	49154	24	Continuous	1
Ethernet - 003	Ethernet - 001	8	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	203::1:2	201::1:2	00:13:01:00:00:01	fc:bd:67:2c:fe:bd	49152	24	Continuous	1
Ethernet - 003	Ethernet - 002	9	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	203::1:2	202::1:2	00:13:01:00:00:01	fc:bd:67:2c:fe:bd	49153	24	Continuous	1
Ethernet - 003	Ethernet - 004	10	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	203::1:2	204::1:2	00:13:01:00:00:01	fc:bd:67:2c:fe:bd	49154	24	Continuous	1
Ethernet - 004	Ethernet - 001	11	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	204::1:2	201::1:2	00:14:01:00:00:01	fc:bd:67:2c:fe:bd	49152	24	Continuous	1
Ethernet - 004	Ethernet - 002	12	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	204::1:2	202::1:2	00:14:01:00:00:01	fc:bd:67:2c:fe:bd	49153	24	Continuous	1
Ethernet - 004	Ethernet - 003	13	Write First, Write Middle: 1022, Write Last	Write First: 1122, Write Middle: 1106, Write Last: 1106	204::1:2	203::1:2	00:14:01:00:00:01	fc:bd:67:2c:fe:bd	49154	24	Continuous	1

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 Statistics | RoCEv2 Data Plane Port Statistics | RoCEv2 Test Summary Statistics | L2-L3 Test Summary Statistics

Traffic Item	Dest QP	Data Frames Tx	Data Frames Rx	Frames Delta	WRITE Tx	WRITE Complete Rx	WRITE Fail	Flow Completion Time (ms)	Avg Latency (ns)	Min Latency (ns)	Max Latency (ns)	ECN-CE Rx
RoCEv2 Traffic	2	23,818,579,410	23,814,491,982	4,087,428	23,260,331	23,256,339	3,992	2,109,795.690	904	587	38,137	13,122,188
RoCEv2 Traffic	3	22,961,478,727	22,966,951,367	5,472,640	22,423,319	22,428,663	5,344	2,110,606.112	885	590	30,552	4,557,405
RoCEv2 Traffic	4	23,683,865,312	23,679,559,464	4,305,848	23,128,774	23,124,569	4,205	2,109,711.046	894	585	31,037	6,595,129
RoCEv2 Traffic	5	26,756,953,551	26,769,222,897	12,269,346	26,129,837	26,141,818	11,981	2,110,517.141	880	587	78,412	6,587,971
RoCEv2 Traffic	6	24,587,640,266	24,600,674,028	13,033,762	24,011,367	24,024,095	12,728	2,110,606.111	879	587	30,572	6,123,799
RoCEv2 Traffic	7	27,063,189,171	27,067,253,406	4,064,235	26,428,895	26,432,864	3,969	2,109,711.046	888	580	31,060	8,984,172
RoCEv2 Traffic	8	26,858,162,548	26,860,350,143	2,187,595	26,228,674	26,230,809	2,135	2,110,517.140	878	590	78,417	5,933,066
RoCEv2 Traffic	9	25,537,757,431	25,531,396,575	6,360,856	24,939,216	24,933,004	6,212	2,109,795.690	863	587	38,122	6,621,607
RoCEv2 Traffic	10	27,195,601,839	27,188,309,410	7,292,429	26,558,204	26,551,083	7,121	2,109,711.045	889	585	31,075	8,680,772
RoCEv2 Traffic	11	23,413,746,973	23,423,321,359	9,574,386	22,864,987	22,874,337	9,350	2,110,517.140	867	582	36,152	3,731,755
RoCEv2 Traffic	12	23,402,572,737	23,404,078,171	1,505,434	22,854,074	22,855,545	1,471	2,109,795.689	857	580	38,110	4,631,056
RoCEv2 Traffic	13	26,308,885,965	26,323,073,995	14,188,030	25,692,271	25,706,126	13,855	2,110,606.111	925	582	30,532	13,270,769

Figure 6. Keysight AI Fabric RoCEv2 Test Solution Statistics

Specifications

Keysight AI Fabric Test Solution	
Hardware Platform	• AresONE-M 800GE OSFP 8-port fixed chassis model • AresONE-M 800GE OSFP 4-port fixed chassis model • AresONE-M 800GE QSFP-DD800 8-port fixed chassis model • AresONE-M 800GE QSFP-DD800 4-port fixed chassis model • AresONE-S 400GE QSFP-DD 16-port fixed chassis model • AresONE-S 400GE QSFP-DD 8-port fixed chassis model
Ethernet Speeds	• 800GE PAM4 (106G, 53G) • 400GE PAM4 (106G, 53G) • 200GE PAM4 (106G, 53G) • 100GE PAM4 (53G) • 100GE NRZ (25G)
Q-Pairs Configuration	• Local and remote IP • Auto QP Number or custom QP Number • DSCP code • ECN setting • UDP Source Port • Execute Command: RDMA WRITE • Message Size and unit
Q-Pairs Establishment	• Connect Request • Connect Reply • ReadyToUse
Congestion Control	• CNP Priority Type, CNP Priority Value, CNP ECN value • ACK Priority Type, ACK Priority Value, ACK ECN value • NAK Priority Type, NAK Priority Value, NAK ECN value • 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 Rate Control Parameters
Statistics	• Data Frame Tx/Rx count • Tx/Rx L1/L2 Rate • Completion time, Flow completion time • RDMA WRITE Count: Complete or Fail • Min/Max/Avg Packet latency • ECN Rx, CNP Tx/Rx, ACK Tx/Rx, NAK Tx/Rx • Sequency error

Ordering Information

Keysight AI Fabric Test Solution 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)

Keysight AI Fabric Test Solution 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)
947-4075	Keysight RoCEv2 Lossless Ethernet Test Bundle for AresONE-M 800GE OSFP 8-port fixed chassis model (947-4075)
947-4076	Keysight RoCEv2 Lossless Ethernet Test Bundle for AresONE-M 800GE OSFP 4-port fixed chassis model (947-4076)