Juniper® Validated Design

JVD Test Report Brief: AI Data Center Network with Juniper Apstra, NVIDIA GPUs, and WEKA Storage

test-report-brief-jvd-aiclusterdc-aiml-02-09

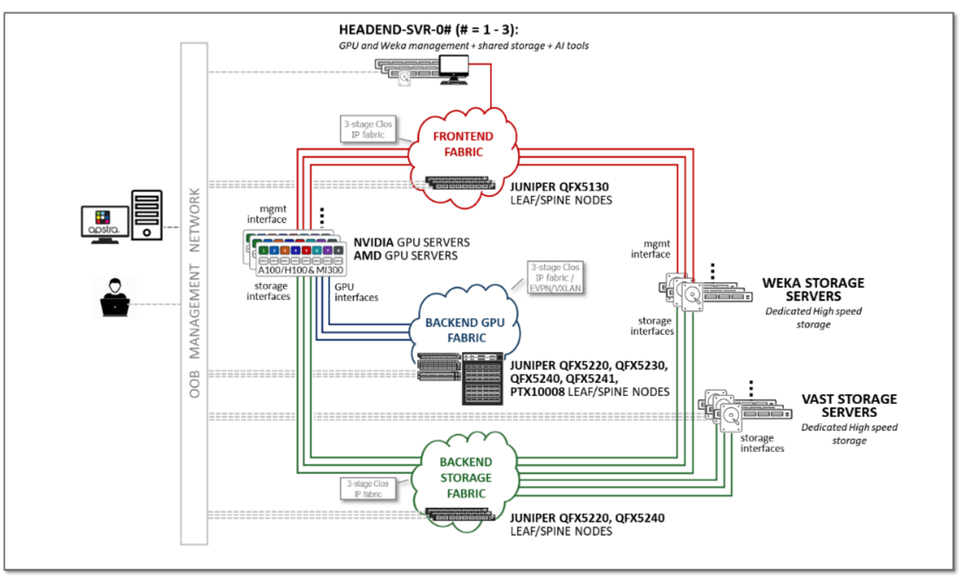
Introduction

This test report brief contains qualification test results for the AI Data Center Network with Juniper Apstra, NVIDIA GPUs, and WEKA Storage Juniper Validated Design (JVD). This qualification evaluated the AI ML Cluster solution using IP Clos deployment for effective congestion control, with QFX5220-32CD, QFX5230-64CD, QFX5240, and QFX5241 as Devices Under Test (DUT) and Apstra as the control point.

The AI ML cluster includes (as depicted in Figures 1 and 2):

1. Front-end fabric with QFX5220 switches as spine and leaf nodes connecting the headend servers with the NVIDIA GPU servers for job management and WEKA storage devices
2. Dedicated storage backend fabric with QFX5220s or QFX5240s connecting the WEKA storage servers and NVIDIA GPU servers.
3. Backend GPU (compute) fabric as described in the Test Topology section.
4. The Cluster solution is orchestrated by Juniper Apstra and hence is dependent on a Juniper Apstra server and Apstra flow deployment for configurations and telemetry.

Figure 1 - AI ML cluster



Test Topology

The Backend GPU (compute) 3-clos IP fabric includes the following components, as depicted in Figure 2):

The Backend GPU (compute) 3-clos IP fabric includes the following components, as depicted in Figures 2 and 3):

	CLUSTER 1	
DUT	LEAF NODES	Juniper QFX5220-32CD switches as Leaf nodes (stripe 1) Juniper QFX5230-64CD as Leaf nodes (stripe 2)
	SPINE NODES	QFX5230-64CD PTX10008 with JNP10K-LC1201
GPU Servers		Nvidia A100 GPU servers (compute nodes). GPUs with ConnectX7 NICs
	CLUSTER 2	
DUT	LEAF NODES	QFX5240-64OD/QFX5241-64OD
	SPINE NODES	QFX5240-64OD/QFX5241-64OD
GPU Servers		Nvidia H100 GPU servers (compute nodes). GPUs with ConnectX7 NICs

Figure 2: Reference Topology Cluster 1

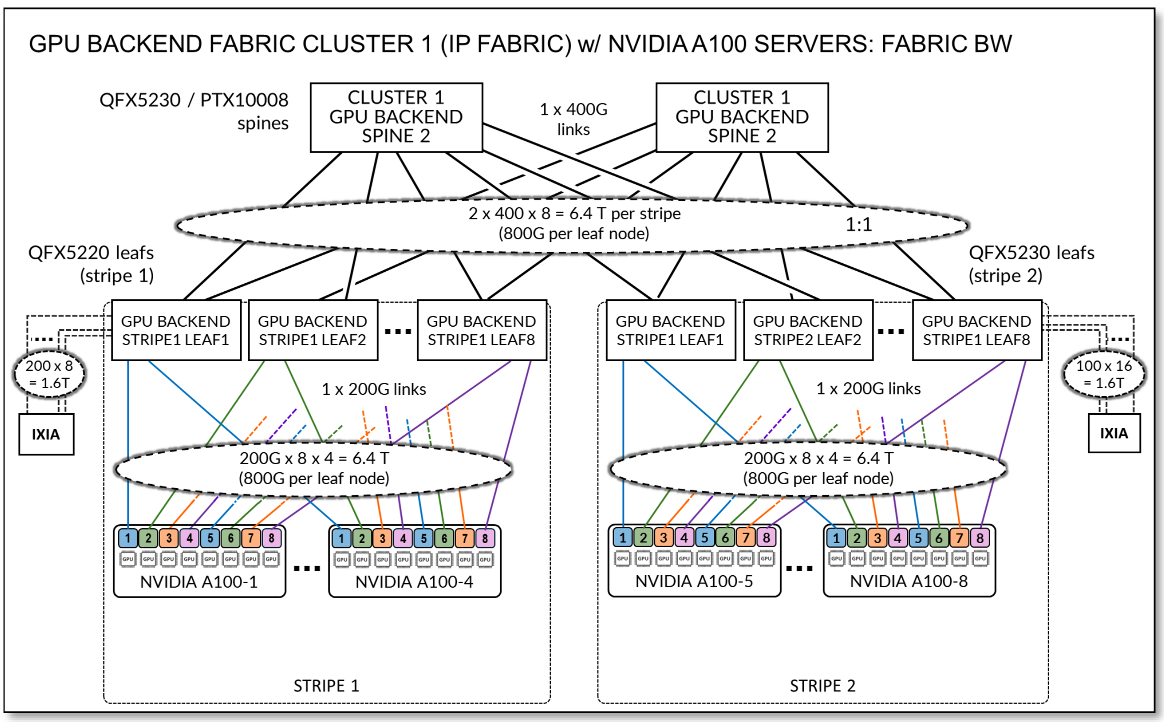
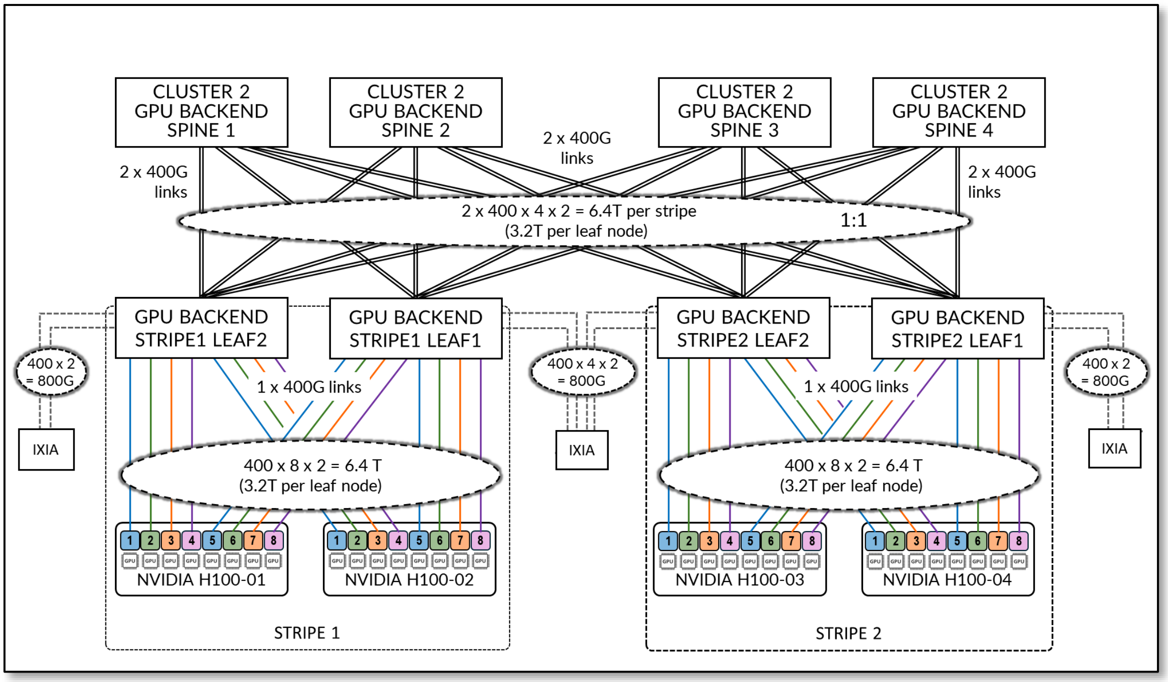


Figure 2 - Reference Topology Cluster 2



If you have questions about this Juniper Validated Design, contact your Juniper representative.

Test Approach

The testing plan focused on validating a set of congestion control scenarios with Junos 23.4X100-D20, implementing congestion control and ECMP DLB flow-let mode for lossless RoCEv2 traffic forwarding in the AI Data Center Network with Juniper Apstra, NVIDIA GPUs, and WEKA Storage Juniper Validated Design (JVD) solution, when running RoCEv2 traffic flows.

These congestion scenarios include fine-tuning dependent parameters (shared buffer allocation, drop-profiles, DLB for a lossless fabric, and more). The overall goal is to establish a lossless RoCEv2 network with ECMP DLB flow-let mode enabled.

Job Completion Time (JCT) values are compared against the MLCommons benchmarks.

Test Goals and Non-Goals

The goal of this phase is to arrive at a better controlled performance of an AI ML storage cluster compared to previously validated values of congestion threshold parameters with the fine-tuning of additional knobs supported in the Junos OS Evolved 23.4X100-D20 release for improved throughput, latency and JCT values with Nvidia A100/H100 GPU servers.

Additional goals include:

1. Determine an optimal Alpha-per-Queue setting for a profile experiencing varying congestion levels with dynamic alpha setting possible per queue level.

2. Establish a PFC XON limit for the ingress ports PG shared buffer threshold setting at which a peer resumes transmitting the packets after a brief PAUSE because of the PFC sent by this node.

3. Monitor ECN-marked packets per congested queue at the CLI level, rather than at the interface level, to enable congestion control specific to each congested queue.

4. Determine optimal values for PFC Watchdog parameters to detect and mitigate PFC pause storms for a recovery from the congested scenario. While this feature itself avoids PFC propagating through the network due to back-pressure halting the traffic, this also ensures the PFC deadlock does not happen in case of link/device failures.

Non-goals for this validation include:

1. Validation of base protocols like EBGP and BFD; this validation focused on congestion threshold parameters.

Platforms Tested

Table 1: Platforms

ROLE	PLATFORM	VERSION
LEAF	QFX5230-64CD	Junos OS Evolved Release 23.4X100-D20
	QFX5240-64OD	Junos OS Evolved Release 23.4X100-D20
	QFX5241-64OD	Junos OS Evolved Release 23.4X100-D42
SPINE	QFX5230-64CD	Junos OS Evolved Release 23.4X100-D20
	QFX5240-64OD	Junos OS Evolved Release 23.4X100-D20
	QFX5241-64OD	Junos OS Evolved Release 23.4X100-D42
CLUSTER DEPLOYMENT/MONITORING	Apstra	5.0

Version Qualification History

This JVD has been qualified in Junos OS Evolved Release 23.4X100-D20.

Performance Data for Cluster-1 (with A100 GPUs)

Test Environment:

Buffer management values for QFX (default):

- Shared buffer – lossless 80%, headroom 10%, lossy 10%
- Dynamic threshold – 7 (default)
- ECN fill level – 55%

Buffer management values for PTX (default):

- ECN fill level – 49%

DLB values for QFX (default):

- flowlet-table-size 256

- flowlet inactivity-timer 256us
- flowlet sampling-rate 62500/s
- flowlet egress-quantization min 20
- flowlet egress-quantization max 50
- flowlet egress-quantization rate-weightage 50
- flowlet reassignment disabled

Traffic Flows:

- Ixia RoCEv2 Tx traffic: 50% (400G) & 75% (600G) traffic load sent to stripe-1 leaf2 & 4, stripe-2 leaf2 & 4.
- 16 QPs per port from IXIA + Model to the leaf devices as ingress traffic
- Only v4 flows with single DSCP code-point marking

NOTE: Apstra pushes default “<ifd>.0 family inet” for unused interfaces. When COS config is enabled, headroom is allocated for all interfaces configured in system. With 10% headroom allocation config, only some interfaces will be allocated headroom, and the rest will have headroom allocation errors as 10% is not enough for all interfaces. We have deleted all un-used interfaces in system and performed the testing to get the performance numbers.

Test Results

Table 2: DLRM model Job Completion time (JCT) test results

Traffic Profile	Tuned parameters in leaf/spine (all other parameters as per defaults above)	JCT [sec]
Only Model	flowset-table-size 2048	3.1
Model + Ixia (50% traffic)	flowset-table-size 2048	3.53
Model + Ixia (75% traffic)	flowset-table-size 2048	4.2

Table 3: BERT model Job Completion time (JCT) test results

Traffic Profile	Tuned parameters in leaf/spine (all other parameters as per defaults above)	JCT [sec]
Only Model	flowset-table-size 2048	2.66
Model + Ixia (50% traffic)	flowset-table-size 2048	2.88
Model + Ixia (75% traffic)	flowset-table-size 2048	2.9

Table 4: NCCL Bandwidth

Traffic Profile	Tuned parameters in leaf/spine (all other parameters as per defaults above)	Bandwidth (GB/s)
NCCL All-reduce QP's – 16	flowset-table-size 2048	115.98
NCCL All-to-All QP's – 4	flowset-table-size 2048	13.92

Table 5: Customer-A model

Traffic Profile	Tuned parameters in leaf/spine (all other parameters as per defaults above)	JCT [msec]
Only Model	flowset-table-size 2048	401.2
Model + Ixia (50% traffic)	flowset-table-size 2048	442.8
Model + Ixia (75% traffic)	flowset-table-size 2048	632

Table 6: LLAMA2 model

Traffic Profile	Tuned parameters in leaf/spine (all other parameters as per defaults above)	JCT [sec]
Only Model	flowset-table-size 2048	70.82
Model + Ixia (50% traffic)	flowset-table-size 2048	70.86
Model + Ixia (75% traffic)	flowset-table-size 2048	70.98

High Level Features Tested

The following table shows the features that were tested with this JVD.

Table 7: Features Tested

Feature	Description
LLDP	Apstra default link discovery mechanism
EBGP sessions	Default routing profile for the Apstra IP Clos deployment
Type 5 routing	Default Apstra ip-prefix-routes announcement type
BGP multi-path multi-AS with load balancing	Enables load balancing across nodes in different AS

Feature	Description
IP ECMP with fast-reroute	Equal traffic distribution at all multipath points (e.g., left-to-spine1/spine2), tested in steady state and with link/protocol/device/process flapping
BFD	BFD with BGP timers validated in steady state and with link/protocol/device/process flapping
COS support	COS features validation on QFX5k for L2 and L3 interfaces
Congestion control with ECN	Congestion control with ECN for lossless queues supporting DCQCN
Congestion avoidance with DLB	Dynamic Load Balancing (DLB) flow mode and packet mode on QFX5k
RoCEv2 support for v4 and v6	RoCEv2-based congestion control for lossless traffic with DSCP-based PFC for both v4 and v6 flows
Accounting features	Account ECN packets per queue
PFC watchdog	PFC watchdog as a solution feature during PFC storm or cyclic buffer dependency states
PFC-DSCP x-on and x-off support	Tunable X-on to enable more robust congestion control
DLB features	1. Tunable hash bucket size for improved congestion avoidance based on the traffic pattern 2. Reactive path rebalancing to enable improved load balancing based on the quality of the ECMP links 3. Ability to selectively enable DLB via user-defined firewall filters (UDF)
Telemetry/SNMP support	Telemetry/SNMP support ECN/PFC counters and ingress buffer usage and drop accounting

Event Testing

The following table shows the events that were tested with this JVD.

Table 8: Events Tested

Test	Details
Create Agent profiles for managed devices of Backend GPU Network	Apstra Deployment
Create pristine configurations and acknowledge managed devices	Apstra Deployment
Create a logical device, interface map, and device profile for leaf devices	Apstra Deployment
Create a logical device, interface map, device profile for spine devices	Apstra Deployment
Create a logical device, interface map, device profile for external router and emulated servers	Apstra Deployment
Create a rack type for leaf and spines	Apstra Deployment
Create a blueprint with the template	Apstra Deployment
Assign all resource pools for the blueprint	Apstra Deployment
Assign interface maps to managed devices	Apstra Deployment
Check cabling map	Apstra Deployment

Test	Details
Commit the blueprint for backend GPU fabric	Apstra Deployment
Configure overlay network routing zones	Apstra Deployment
Assign IP loopbacks	Apstra Deployment
Create virtual networks	Apstra Deployment
Commit overlay updates	Apstra Deployment
Create a blueprint with similar steps for backend storage fabric	Apstra Deployment
Create a blueprint with similar steps for frontend fabric	Apstra Deployment
Create configlets for DCQCN and DLB	Apstra-based Config
Verify baseline configuration	Validate establishment of control and forwarding planes.
Verify overlay connectivity	Validate forwarding plane establishment with emulated host test traffic.
Validate congestion control scenarios	Validate Congestion Control with PFC/ ECN for lossless queues supporting DCQCN with RoCEv2 Traffic
Validate the lossless feature of the fabric under congestion scenarios	Ensure no traffic loss for RoCEv2 Traffic during Congestion, rate-control to happen until congestion is resolved
Reboot devices	Ensure minimal traffic loss when a redundant node is rebooted, and that control and forwarding plane are restored with ECMP when the node reboot is completed.
Leaf to spine link failure	Interface down/up at leaf-spine layer. Minimal traffic loss as all leaves are dual-homed.
Process restart	Ensure minimal traffic loss and full recovery when various JUNOS processes are killed/restarted
Deactivate BGP on leaf	Ensure minimal traffic loss when various leaf and spine BGP sessions are deactivated. Full traffic restoration when sessions are restored.
Fabric device upgrade from Apstra	Junos image changes performed via Apstra on all managed devices. Verify control and forward planes are functional after system upgrade

Traffic Profiles

The following table shows the traffic profiles that were tested with this JVD.

Table 9: Tested Traffic Profiles

Traffic Type	Fabric Traffic Mode	RDMA Flow Count	Packet Size (IXIA)
RoCEv2 from IXIA	Intra-Stripe	16 QPs per port	9000 bytes
RoCEv2 from IXIA	Inter-Stripe	16 QPs per port	9000 bytes
RoCEv2 from IXIA plus GPU-based DLRM Model	Intra-Stripe	16 QPs per port	9000 bytes
RoCEv2 from IXIA plus GPU-based DLRM Model	Inter-Stripe	16 QPs per port	9000 bytes
RoCEv2 from IXIA plus GPU-based Jane St. Model	Intra-Stripe	16 QPs per port	9000 bytes
RoCEv2 from IXIA plus GPU-based Jane St. Model	Inter-Stripe	16 QPs per port	9000 bytes

Known Limitations

Below are some of the caveats which will affect job completion time (JCT):

- DLB does not use the port BW for calculation of quality of link. Instead, the quality of the link is based on the amount of traffic recently transmitted over each ECMP link, and the amount of traffic enqueued on each ECMP link for transmission. This could lead to flows being assigned to a lower BW link instead of a higher BW link, which could cause congestion. Also, if a link quality degrades a flow that has been assigned to a link will not be reassigned, unless it pauses for a period of time longer than the inactivity interval. The port quality metrics and inactivity interval can be adjusted to overcome this condition; see Customize Egress Port Link Quality Metrics for DLB. Also, consider implementing Reactive Path Rebalancing.

- Out-of-order packets can be seen when implementing DLB flowlet mode if the inactivity interval is too small (e.g. 16 usec). The recommended value for the inactivity interval is 256 usec (default).

Corporate and Sales Headquarters
Juniper Networks, Inc.
1133 Innovation Way
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER (888.586.4737)
or +1.408.745.2000
Fax: +1.408.745.2100
www.juniper.net

APAC and EMEA Headquarters
Juniper Networks International B.V.
Boeing Avenue 240
1119 PZ Schiphol-Rijk
Amsterdam, The Netherlands
Phone: +31.207.125.700
Fax: +31.207.125.701