

Ericsson Unicycle SR-IOV Validation HW, Networking and IP plan

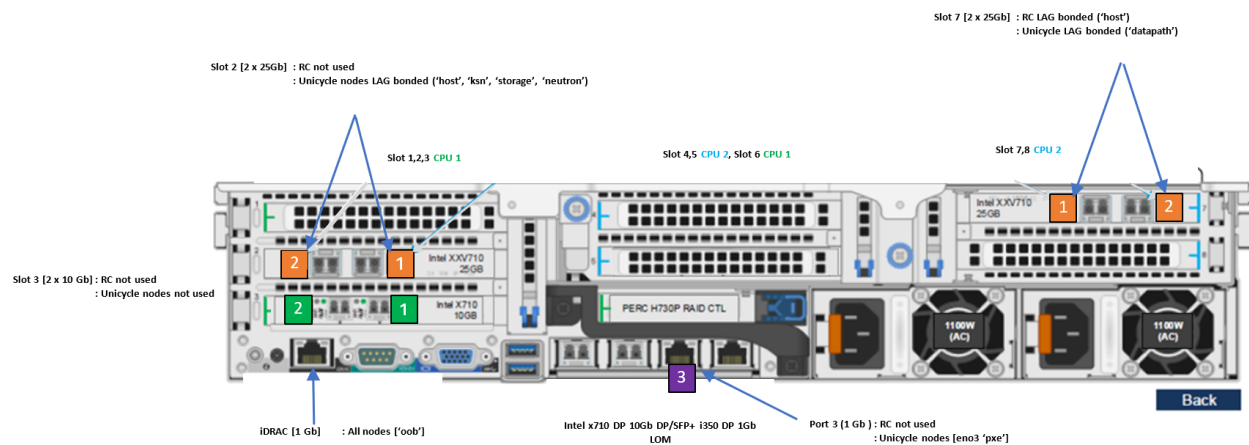
- [Unicycle SR-IOV Validation Servers](#)
- [Unicycle SR-IOV Validation Networking](#)
- [Unicycle SR-IOV Validation IP/VLAN Plan](#)
- [Unicycle SR-IOV BGP Plan](#)
- [Unicycle SR-IOV LAG Details](#)

Unicycle SR-IOV Validation Servers

Verification was based on a Build Server VM and four identical Dell 740XD servers for the Regional Controller and three Unicycle nodes.

Dell Purley 740XD Server

L2 networks 'oob', 'host', 'pxe', 'ksn', 'storage', 'neutron', 'vxlan' – Refer to IP plan spreadsheet for values and further details



PowerEdge R740XD Server



Components

- 1 PowerEdge R740/R740XD Motherboard
- 1 Intel Xeon Gold 6152 2.1G, 22C/44T, 10.4GT/s , 30M Cache, Turbo, HT (140W) DDR4-2666
- 1 iDRAC Group Manager, Enabled
- 1 iDRAC,Legacy Password
- 1 Chassis with Up to 24 x 2.5 Hard Drives for 2CPU, GPU Capable Configuration
- 1 Riser Config 6, 5 x8, 3 x16 slots
- 1 PowerEdge R740 Shipping Material
- 1 No Quick Sync
- 1 Performance Optimized
- 1 2666MT/s RDIMMs
- 12 32GB RDIMM 2666MT/s Dual Rank
- 1 Intel Xeon Gold 6152 2.1G, 22C/44T, 10.4GT/s , 30M Cache, Turbo, HT (140W) DDR4-2666
- 1 iDRAC9,Enterprise
- 4 480GB SSD SATA Read Intensive 6Gbps 512 2.5in Hot-plug AG Drive, 1 DWPD, 876 TBW
- 6 2.4TB 10K RPM SAS 12Gbps 512e 2.5in Hot-plug Hard Drive
- 1 PERC H730P RAID Controller, 2GB NV Cache, Adapter, Low Profile
- 2 C13 to C14, PDU Style, 10 AMP, 6.5 Feet (2m), Power Cord
- 1 Dual, Hot-plug, Redundant Power Supply (1+1), 1100W
- 1 No Trusted Platform Module
- 1 Order Configuration Shipbox Label (Ship Date, Model, Processor Speed, HDD Size, RAM)
- 1 GPU Ready Configuration Cable Install Kit
- 1 PE R740XD Luggage Tag
- 1 Intel X710 Dual Port 10Gb Direct Attach, SFP+, Converged Network Adapter
- 2 Intel XXV710 Dual Port 25GbE SFP28 PCIe Adapter, Full Height
- 1 Intel X710 DP 10Gb DA/SFP+, + I350 DP 1Gb Ethernet, Network Daughter Card
- 1 HS Install Kit,GPU Config.No cable
- 1 ReadyRails Sliding Rails Without Cable Management Arm
- 1 Unconfigured RAID
- 1 OME Server Configuration Management

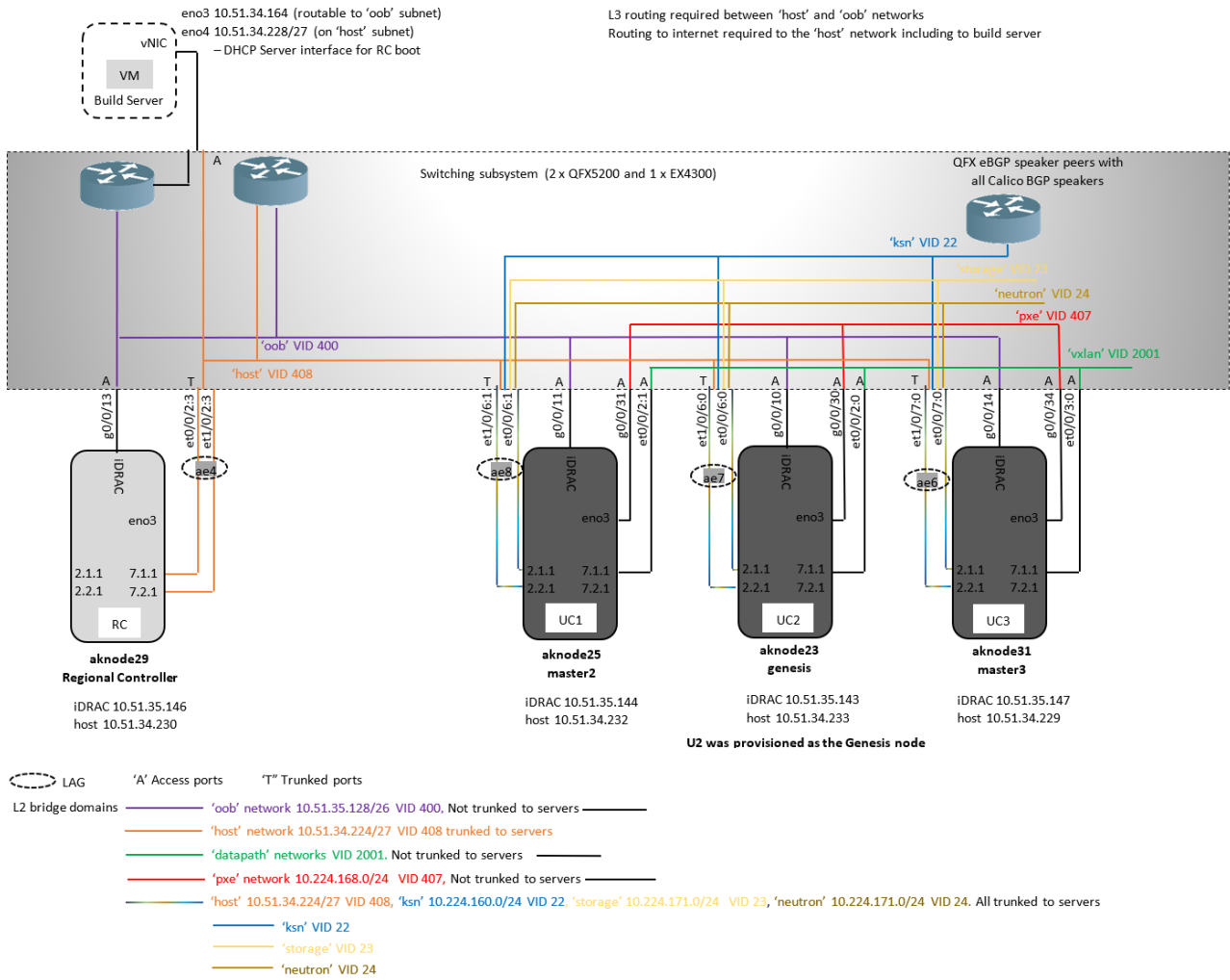
Software

- 1 6 Performance Fans forR740/740XD
- 1 Performance BIOS Settings
- 1 No Operating System
- 1 No Systems Documentation, No OpenManage DVD Kit

Unicycle SR-IOV Validation Networking

The diagram below shows an example physical and L2 connectivity, the IP subnet plan, host and iDRAC addressing scheme similar to that used in the validation.

The Unicycle deployment does not configure the networking subsystem thus the choice of switch subsystem components is not restricted to those shown and they can be replaced with devices offering equivalent functionality.



Unicycle SR-IOV Validation IP/VLAN Plan

Below is an example of a detailed network and IP address plan similar to that used during validation.

NOTE: LAG groups always created on 2 ports of the same NIC card on compute. LAGs not split across NIC card sets.
2 x 25G NIC card not used
NOTE: Build Server is only used to create the Regional Controller. Once the RC is built the build server has no further role in the deployment of the flow or Unicycle pods and can be removed
Note: **The RC uses the LAG bonded interfaces on 7.1.1 and 7.1.2 to send Redfish API calls to the flow and Unicycle nodes connected to the 'oob' network thus routing between the 'host' and 'oob' networks is required.

Network name	Server	Network name	Compute HW	Present	NIC slot/port	Port rate	LAG	VID	Subnet	Host address	Network name	Note
'oob'	Build Server*	oob	(vm)	vnic		Any	No	native	Any	10.51.34.164	'oob'	Build server uses Redfish API calls to the Regional Controller to provision BM. Regional Controller then uses Redfish API calls to flow/UC1, UC2 and UC3 over 'oob' network
	Regional controller	oob	aknode29	Yes	BMC	10	No	native		10.51.35.146		Initially build server issues Redfish calls to setup RC. Subsequently RC issues Redfish commands to flow and Unicycle nodes. Build server VMs and interfaces assigned 10.51.34.164/27 and eno4 10.51.34.228/27 for DHCP server
	Unicycle (master2)	oob	aknode25	Yes	BMC	10	No	native		10.51.35.144		Switch port configured as access port bridged to VLAN 400 in switch network
	Unicycle (genesis3)	oob	aknode23	Yes	BMC	10	No	native		10.51.35.143		Switch port configured as access port bridged to VLAN 400 in switch network
	Unicycle (master3)	oob	aknode31	Yes	BMC	10	No	native		10.51.35.147		Switch port configured as access port bridged to VLAN 400 in switch network
'host'	Build Server*	host	(vm)	vnic		Any	No	native	Any	10.51.34.230	'host' (single seen as 'host')	When Regional Controller flows based on DHCP Request is served by Build Server's DHCP server on eno4
	Regional controller	host	aknode29	Yes	2.1.1.8.2.1.1	2x25G	Yes	408	10.51.34.224/27	10.51.34.230		Flow/Unicycle's 'genesis node' DHCP Requests are served by Regional Controller's DHCP servers. The RC also uses this phy interface for iDRAC management of the flow and Unicycle nodes
	Unicycle (master2)	host	aknode25	Yes	2.1.1.8.2.1.1	2x25G	Yes	408		10.51.34.232		The genesis node sends its initial DHCP Req via this interface to the RC
	Unicycle (genesis3)	host	aknode23	Yes	2.1.1.8.2.1.1	2x25G	Yes	408		10.51.34.233		
	Unicycle (master3)	host	aknode31	Yes	2.1.1.8.2.1.1	2x25G	Yes	408		10.51.34.229		
'pxe'	Build Server*	pxe	(vm)	vnic		Any	No	native	Any	10.224.168.11	'pxe'	No pxe network on RC. RC node pxe boots over 'host' network from Build Server
	Regional controller	pxe	aknode29	Yes	eno4	First 10 ports on integrated card	No	native		10.224.168.0/24		Unicycle (genesis3) boots from RC over 'host' network over 'pxe' L2 network (over WAN). After setup Unicycle's node support a MAAS server to boot other local Unicycle nodes
	Unicycle (master2)	pxe	aknode25	Yes	eno4	First 10 ports on integrated card	No	native		10.224.168.12		Unicycle (master2) boots from local MAAS server on Unicycle's node over 'pxe' L2 physical site local network
	Unicycle (genesis3)	pxe	aknode23	Yes	eno4	First 10 ports on integrated card	No	native		10.224.168.13		Unicycle (genesis3) boots from local MAAS server on Unicycle's node over 'pxe' L2 physical site local network
	Unicycle (master3)	pxe	aknode31	Yes	eno4	First 10 ports on integrated card	No	native		10.224.168.14		
'storage'	Build Server*	storage	(vm)	vnic		Any	No	native	Any	10.224.171.0/24	'storage'	Calico is used as the MDS CN.
	Regional controller	storage	aknode29	Yes	2.1.1.8.2.1.1	2x25G	Yes	23	10.224.171.0/24	10.224.171.11		Calico's BGP peer pairing to the TOR router based on eBGP pairing.
	Unicycle (master2)	storage	aknode25	Yes	2.1.1.8.2.1.1	2x25G	Yes	23		10.224.171.12		TOR router requires dynamic BGP pairing configuration to accept each Calico BGP speaker's BGP pairing request.
	Unicycle (genesis3)	storage	aknode23	Yes	2.1.1.8.2.1.1	2x25G	Yes	23		10.224.171.13		This is the OpenStack storage network
	Unicycle (master3)	storage	aknode31	Yes	2.1.1.8.2.1.1	2x25G	Yes	23		10.224.171.14		Note
'neutron'	Build Server*	neutron	(vm)	vnic		Any	No	native	Any	10.224.171.0/24	'neutron'	This is the OpenStack neutron network
	Regional controller	neutron	aknode29	Yes	2.1.1.8.2.1.1	2x25G	Yes	24	10.224.171.0/24	10.224.171.11		
	Unicycle (master2)	neutron	aknode25	Yes	2.1.1.8.2.1.1	2x25G	Yes	24		10.224.171.12		
	Unicycle (genesis3)	neutron	aknode23	Yes	2.1.1.8.2.1.1	2x25G	Yes	24		10.224.171.13		
	Unicycle (master3)	neutron	aknode31	Yes	2.1.1.8.2.1.1	2x25G	Yes	24		10.224.171.14		
'datapath'	Build Server*	datapath	(vm)	vnic		Any	No	native	Any	10.224.171.0/24	'datapath'	No server traffic on the RC.
	Regional controller	datapath	aknode29	Yes	7.1.1.8.2.1.1	2x25G	Yes	2001-3000		10.224.171.11		If SR-IOV Unicycle BP is deployed these three lines apply. LAG in Unicycle SR-IOV BP is used.
	Unicycle (master2)	datapath	aknode25	Yes	7.1.1.8.2.1.1	2x25G	Yes	2001-3000		10.224.171.12		These VLANs support the Openstack network. This
	Unicycle (genesis3)	datapath	aknode23	Yes	7.1.1.8.2.1.1	2x25G	Yes	2001-3000		10.224.171.13		Initially separate VMs are deployed on the master nodes.
	Unicycle (master3)	datapath	aknode31	Yes	7.1.1.8.2.1.1	2x25G	Yes	2001-3000		10.224.171.14		

Unicycle SR-IOV BGP Plan

The three unicycle nodes automatically peer with an external fabric BGP speaker using eBGP. The calico nodes do not peer using an internal iBGP mesh. Below is an example similar to that used in the validation.

BGP	ASN	IP address/subnet	Start IP	End IP			
Calico ('k8s' network)	65531	10.224.160.0/24	10.224.160.134	10.224.160.254			
external router (maybe QFX or other HW/SW router TBD)	65001	10.224.160.129	NA	NA			Any private ASN other than the Calico ASN can be used. The external router must be on the same L2 domain as the 'k8s' network but can be a physical router or a SW based router Dynamic BGP peering to the Calico subnet must be configured on the external peer router to accept BGP peering requests from any k8s Calico BGP speaker The external BGP router must be setup to advertise routes to BGP speakers of the same ASN that they are received from.

Unicycle SR-IOV LAG Details

The RC and Unicycle genesis nodes boot via VLAN tagged 'host' interface which is pre-provisioned on the QFX switches with LAG bonding. Since booting occurs before the linux kernel can bring up its LAC-P signalling the QFX switches must be configured to pass traffic on their primary (first) link before the LAG bundle is up.

Note the Unicycle [master2] and unicycle [master3] nodes do not boot over the 'host' network but rather over the edge site 'pxe' network which is a single link to each unicycle server and thus not lag bonded.

LAG							
LAG config on QFX5200	FORCE UP	DHCP and http pxe boot occurs over interfaces configured for LAG on JPR switches					