

RMA Field Diagnostics Test using Repair Rack

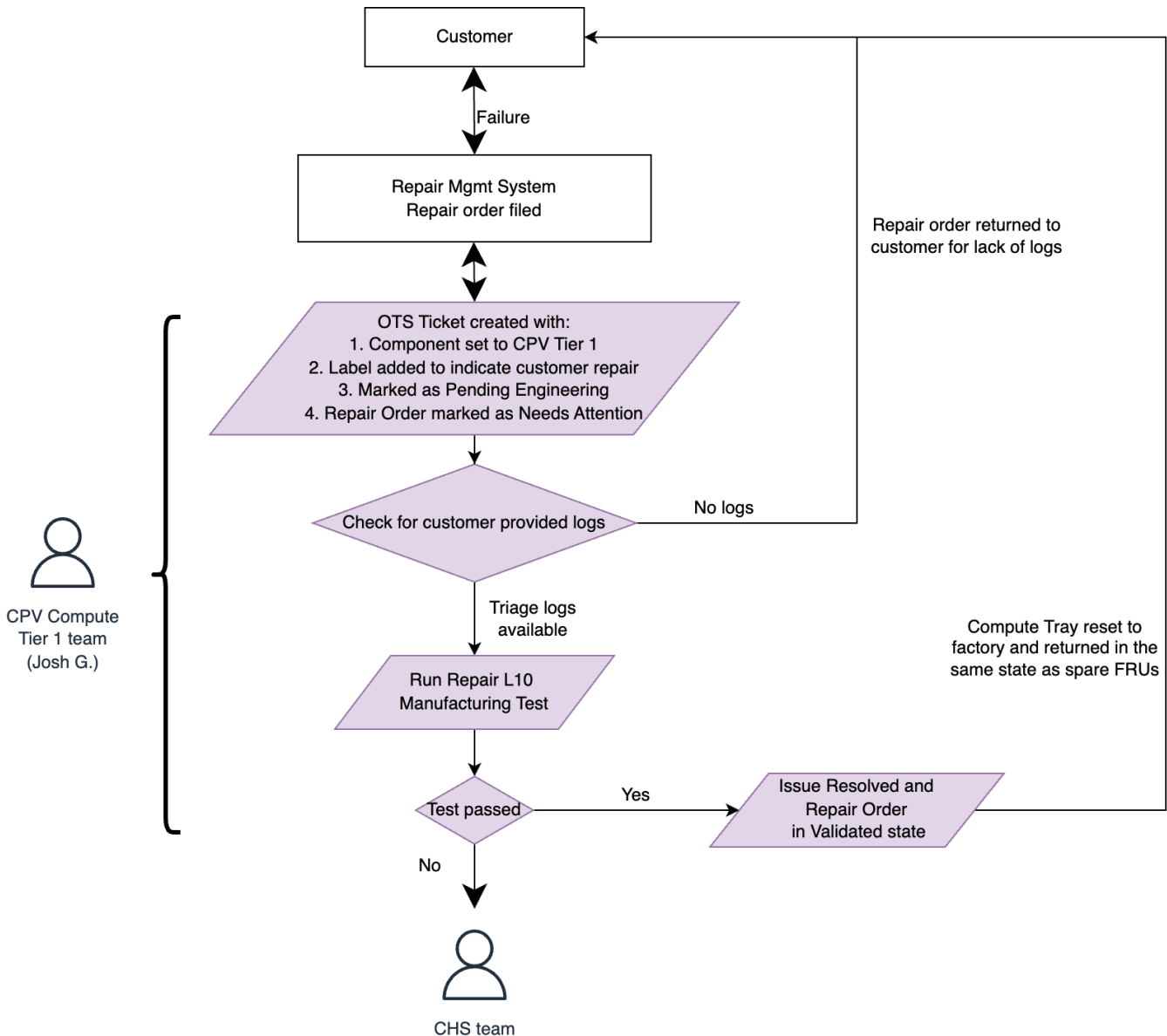
We are currently executing on a proof-of-concept that will repair and test GB200 compute trays inside the ABL data center. As part of this initiative, we have setup a repair rack with a Diagnostics Server that is used to execute full L10 manufacturing test on a compute tray in the rack (see: [Repair Validation Rack](#)). With this current setup, we have also validated the ability to run those same tests against compute trays in the customer rack, without impacting the customer that may have instances running in the same rack. As a short-term stop-gap solution until Warminator is ready, we plan to use this repair setup to test compute trays with repair orders from the customer.

RMA Short-Term Workflow



After we finish running L10 Manufacturing tests on a compute tray, the compute tray will be factory reset and running the latest manufacturing version of firmware. The compute tray must be handed back to the customer like we would a spare replacement. The compute tray will be configured for PXE booting as a spare, but will need to be configured and have the correct customer firmware flashed.

This full Warminator flow is defined in [Warminator flow](#), with the following flow as a short-term process until Warminator is available.



L10 Manufacturing Test (FDT)

The Field Diagnostic Test using the L10 Manufacturing test procedure will execute the following:

Test	Details/Command
Check, Update, and Validate ILOM firmware	Check ILOM is at the expected version, flash ILOM firmware, and check version
Check, Update, and Validate PLDM firmware	Check PLDM is at the expected version, flash PLDM firmware, and check version
Run hardware diagnostic i2c test	Run hwdiag i2c test and only validate result
Boot live host image to update CX8, SWB, IOB, Nvme firmware	Except firmware update, this step checks nvme count and cx8 count, we can move these checks when booting to Ubuntu 2404 before running partnerdiag
Check open problems	Fail for any ILOM Open Problems
hwdiag i2c test all	<code>hwdiag i2c test all</code>
hwdiag fan test	<code>hwdiag fan test -o</code>
cx8 count validation	check cx8 count is as expected <code>lspci grep Mellanox grep ConnectX-8</code>
Loopback cable type check	make sure using optical transceiver <code>mlxlink -d mlx5_{ } -m grep -i "Cable Type"</code>
Validate all PCI bus	<code>mst start && mst status -v</code>
Configure CX8 PCI bus using 2*400Gb/s	<code>mlxconfig -y -d {pci_id} set MODULE_SPLIT_M0[0..3]=1 "</code> <code>f"MODULE_SPLIT_M0[4..7]=2 MODULE_SPLIT_M0[8..15]=FF</code> <code>mlxconfig -y -d {pci_id} set NUM_OF_PF=2</code>
Get mlxlink full output to check if transceiver is reachable	<code>mlxlink -d { } -m -c -e -j</code>
osfp_internal_prbs test	
ultrapass cable flip check	
RDMA serial cable check	<code>mlxlink -d mlx5_{ } -m</code>
Partnerdiag test	Run partnerdiag tests (see below)
Disable internal loopback	
Reset to Factory	

NVIDIA Partnerdiag Tests

The following table contains all tests executed as part of L10 manufacturing partnerdiag. This table contains information from the NVIDIA document (DU-11965-001_18), customized based on what we execute and what failures we ignore in L10 manufacturing.

Actions	Test Duration	Test Description	Pass/Fail Criteria
Inventory	1 minute	System-level check of components against expected versions.	Fails if the firmware version does not match.
inforom	40 seconds	Checks the GPU inforom	
tegra_cpu (TegraCpu)	10 minutes (configurable)	Performs CPU diagnostics testing.	Fails if the CPU operation is not stable.

tegra_memory (TegraMemory)	45 seconds	Saturates the system memory bus with the concurrent data traffic that was generated by High Speed Scrubbing and the CPU. It requires a MODS secure partition.	Fails if unable to perform read/write/allocate transactions on all memory channels.
tegra_memory (CpuMemorySweep)	22 minutes	Perform reads, writes, correctness checks for CPU memory. This requires a MODS secure partition.	Fails if the test cannot allocate read/write/allocate transactions on all memory channels.
tegra_clink (TegraClink)	40 seconds	CPU-CPU NVIDIA® NVLink™ Test.	Fails if the link quality thresholds are not met.
ssd	2 minutes (configurable)	SSD Stress test.	Skipped in Repair Test
pcieproperties (CxpCieProperties, BfPcieProperties, BfMgmtPcieProperties)	1 second	Verifies the PCIe connection properties.	Fail if the detected PCIe properties do not match the spec JSON file. Skipped BfPcieProperties and BfMgmtPcieProperties for ABL SKU
gpustress (Gpustress)	4.5 minutes	GPU stress tests.	Fails if there is a CRC miscompare during computation.
gpumem (Gpumem)	One minute	GPU memory and interface (FBIO) tests.	Fails if the GPU memory is unstable.
pcie	12.5 minutes	GPU PCIe bandwidth, speed switching, and eye diagram tests.	Fails if the GPU PCIe connection is unstable or cannot achieve required functions.
thermal (ThermalSteadyState)	11.5 minutes (configurable)	Thermal test.	Fails if the temperature exceeds the limitation.
connectivity	12 minutes	Validates that the electrical quality of NVLinks and PCIE link speeds/width match the POR.	Skipped in L10 Test
nvlink (NvlBwStress, NvlBwStressBg610)	12 minutes	NVLink bandwidth and eye diagram tests.	Skipped in L10 Test
ibstress (IbStressCables, IbStressBf3PhyLoopback, IbStressBf3Loopout, IbStressCx7PhyLoopback, IbConfigureCx7Cables400G_8X, IbStressLoopout400G_8X, IbConfigureCx7Cables400G_4X, IbStressLoopout400G_4X, Cx8GpuDirectLoopback, Cx8GpuDirectCrossGpu)	2 minutes (configurable)	Performs infiniband stress read and write operations and verifies the performance.	Fails if the expected bandwidth is not met. Skipped IbStressCables and IbStressCx7PhyLoopback in Repair Test Skipped IbStressBf3PhyLoopback, IbStressBf3Loopout, IbStressLoopout400G_8X, IbStressLoopout400G_4X, Cx8GpuDirectLoopback, and Cx8GpuDirectCrossGpu for ABL SKU
dpudiag (CxeyegradeStart, CxeyegradeStop, Bf3PcieInterfaceTraffic)	5 minutes (configurable)	Performs infiniband traffic stress, eye diagram, and thermal stress of ConnectX Devices	Fails if the ConnectX-8 does not meet performance standards. Skipped CxeyegradeStop and Bf3PcieInterfaceTraffic for ABL SKU

powersync (CpuGpuSyncPulsePower1Hz50duty, CpuGpuSyncPulsePower3Hz50duty, CpuGpuSyncPulsePower10Hz50duty, CpuGpuSyncPulsePower25Hz50duty, CpuGpuSyncPulsePower50Hz50duty, CpuGpuSyncPulsePower100Hz50duty, CpuGpuSyncPulsePower200Hz50duty, CpuGpuSyncPulsePower300Hz50duty, CpuGpuSyncPulsePower400Hz50duty, CpuGpuSyncPulsePower500Hz50duty, CpuGpuSyncPulsePower600Hz50duty, CpuGpuSyncPulsePower700Hz50duty, CpuGpuSyncPulsePower800Hz50duty, CpuGpuSyncPulsePower900Hz50duty, CpuGpuSyncPulsePower1kHz50duty, CpuGpuSyncPulsePower1_5kHz50duty, CpuGpuSyncPulsePower2kHz50duty, CpuGpuSyncPulsePower3kHz50duty, CpuGpuSyncPulsePower4kHz50duty)	2.5 minutes per frequency (configurable)	Performs a synchronous CPU and GPU power stress.	Fails if the temperature exceeds the limitation.
chkoccurrences (SyslogErrorCheck, KernLogErrorCheck, DmesgLogErrorCheck, SyslogAERCheck, KernLogAERCheck)	1.5 minutes	Verifies that the required permissions and tools are installed for all actions selected to be run.	The required dependencies and/or permissions for the selected actions are not met. <i>Ignored on Failure for Repair Test</i>