

Cnuas Product Roadmap

Last updated 2026-08-22 14:25 UTC, from the GitHub project board

Available In Progress Planned Blocked

Cnuas Timing: Calibrated Virtual-time and Performance Model 7 tasks	Deterministic integer-picosecond virtual clock and event schedule...	Versioned cross-component profiles for NIC switch GPU link...	Target-specific calibration fitting and held-out error...	QEMU icount deterministic TCG experiment mode [VT4]	Physical hardware calibration datasets and validated target...
	Runtime telemetry adapters and automatic trace ingestion [VT6]	Dedicated calibrated virtual-time research paper [VT7]			

CnuasNIC: RDMA NIC (RoCEv2 + native InfiniBand) 26 tasks	RoCEv2 TX path [H1]	RoCEv2 RX path [H2]	RC reliability + RDMA READ	RC atomics (CMP_SWP_FETCH_ADD)	Multi-packet WRITE/READ honouring path MTU
	Real ICRC compute + validate on RX	Shared Receive Queue (SRQ)	UD QPs + IB multicast	Per-port IB link layer + dual port_immutable [I1]	LRH + IB opcodes; parallel IB skb builder [I2-13]
	Wire IB skb through QEMU vNIC + switch pipeline [I4]	QP0 MAD plumbing; SMA GET/SET; directed-route SMPs [I5a-I5c]	SM wire-format + kernel SMI groundwork [I5f]	LRH-aware RX dispatcher [I6]	End-to-end test gate [I7]
	NCCL / UCX build+link gates	DMA-BUF MR import and I/O-safe CnuasGPU BAR access [PM2]	Userspace verbs fast path [H5]	Native IB LRH-only TX path for RC QPs [H1.5]	SA path-record responder on QP1 [I5f-SA]
	PMA + perquery support [I5g]	opensm-in-VM compatibility lab [I5h]	RoCE/IB HowTo rewrite [I8]	Debug lb_umad umadN creation	UCCL-P2P RDMA transfer engine smoke
	UCCL collective (AllReduce) host-mem smoke				

CnuasGPU: Virtual GPU + Runtime 33 tasks	GPU driver scaffold + /dev/cnuasgpuN char device [Phas...	ioctl set (INFO/ALLOC/FREE/MEMCPY/LINK)	libcnuasdev low-level driver shim	libcnuasrt CUDA-style runtime	Vector add compute primitive [5b]
	Tile-based SGEMM (AVX2/FMA)	AVX-512 SGEMM [6.3]	Multi-GPU verification	cnuassmi (cnuas-smi) CLI	Compute-backend refactor and dispatch loader [6.1]
	Split backends into .so + dlopen loader [6.1c]	Device-backend refactor [6.2]	CnuasIR object format and reference parser [6.4a]	CnuasIR packer disassembler and validator tools [6.4b]	CnuasIR interpreter and kernel launch API [6.4c]
	Host device backend for standalone Soft-GPU	Vectorised activations for AVX2 and AVX-512	ggml backend libggml-cnuas.so for llama.cpp and Ollama	Lifetime-safe allocation and subrange DMA-BUF export [PM1]	libcnuaspeermem export and standard verbs registration API...
	Tile/accumulate primitive exposing the TTU	TVP (Tile Vector Processor) lanes	CCP (Cluster Control Processor) command processor	LSU (Load Store Unit) operand movement	TCP (Tile Control Processor) as an addressable block
	SFU (Special Function Unit) as an addressable block	TF32 matrix multiply with FP32 accumulation	BF16 matrix multiply entry point and conversion	2:4 structured sparsity metadata layout and sparse entry point	Quantised dequantise-to-SGEMM compute paths
	Host character device for the ioctl ABI	vLLM integration	Documentation pass [6.6]		

Cnuas Peer Memory 1 task	Guest CnuasGPU to CnuasNIC RDMA transfer validation [PM4]				
--	---	--	--	--	--

CnuasGPU FPGA IP: PolarFire SoC Icicle Kit 10 tasks	Board preparation and Linux on the U54 cores [6.5.0]	Minimum PCIe bitstream with BAR0 readback [6.5.1]	Firmware daemon on the U54 cores serving CCP and TCP [6.5.2]	Vector add on fabric as the first TVP lane in RTL [6.5.3]	SAXPY scale dot and SGEMM on fabric TVP lanes [6.5.4]
	CnuasIR execution on the U54 cores [6.5.5]	CnuasIR in fabric [6.5.6]	TTU in fabric on the MACC array	FPGA IP implementation and legal release review [6.5]	Bitstream distribution as tagged releases

CnuasSwitch: Virtual TOR Switch 10 tasks	10-port switch daemon	RoCEv2 + native IB forwarding	DCB PFC/ETS/ECN	Ethernet FDB + IB LFT	Switch-side SMA [I5d]
	In-switch Subnet Manager [I5e]	SM SMP wire-format fix [I5f]	OpenTelemetry export	JSON management API	Switch-side SA/PMA alignment [I5g]

CnuasLink: GPU Fabric Interconnect 8 tasks	GPU fabric switch daemon [P1]	8-port frame-based protocol	FDB learning + discovery + keepalive	CnuasGPU to switch wiring [P2]	Link state in sysfs + cnuassmi [P2.1]
	cnuaslink-cli management CLI [P4]	Collective offload end-to-end	Bandwidth / latency modelling		

Platform Tooling & VM Lifecycle 11 tasks	VM lifecycle CLI [6.2e]	Package + image build subcommands [6.2c-6.2d]	Kernel port to v6.19 + binary packaging [6.2f]	Deployment tooling	Self-hosted apt repository and offline install
	Repo split into submodules	QEMU fork tagged for consumers	CI strategy (Azure DevOps) [6.0e]	Build system (Bazel + Make)	PCI vendor/device ID allocation request
	Upstreaming tracks 1 to 5				

Documentation 7 tasks	Design docs [6.0]	Component + platform datasheets	Live version wiring from git tags	Version pinning reference	Convert ASCII diagrams to mermaid
	Product roadmap	Vector redraw of the finished logo mark for poster scale			

Deferred design decisions 23 tasks	Math mode selection API	Host matrix engine (AMX/SME) exposure	Mixed-precision GEMM via cnuasSgemmEx	NUMA-aware backend selection	CnuasIR JIT for AVX2 / AVX-512 / SVE / RVV
	Asynchronous compute queues and streams	Growable host arena	Shared device memory IPC across processes	pcie_fpga backend versus a capability bit	Heterogeneous multi-device backends
	Fabric compute scope for the first FPGA target	Bitstream distribution as tagged releases	PolarFire SoC Discovery Kit as a second target	Zephyr or bare metal on the U54	Container CI jobs versus raw VM agents
	Outbound allow-list for CI agents	CI parallel job capacity	Cross-architecture CI for ARM64 and RISC-V64	DKMS support for non-Cnuas kernels	Multi-machine image distribution
	Golden image signing	--prefix flag for libvirt domain names	libvirt-python bindings instead of subprocess		

Facility Digital Twin 12 tasks	Campus buildings and room programme generator	Detailed data hall with the emulated racks	Plant yards sized from the load	Power roll-up to a grid connection	OpenUSD emission in Z-up metres
	Custom JSON campus specification spatial validation and...	Live rack power from the ORV3 shelf	Runtime load input to the PSU model	Workload profiles driving the shelf	Cooling plant that responds to measured heat
	Workloads on emulated blades driving the draw	Grid frequency and generator start sequencing			

Accelerator Software Stack 19 tasks	CnuasML NVML-style management library [8.8]	Cnuasdebug low-level inspection tool [8.8]	cnuas_gpu_tools.py device inspection script [8.8]	CnuasBLAS v0.1 over the shipped SGEMM and vector kernels [8.1]	Userspace CnuasLink transport wrapper in CnuasDev [8.2a]
	CnuasCCL rank bootstrap protocol [8.2b]	CnuasCCL v0.1 broadcast AllReduce and AllGather [8.2c]	CnuasIR scalar subset completion as the compiler target [8.3a]	cnuascc front end device C subset parser and type checker [8.3b]	cnuascc code generator emitting CnuasIR objects scalar subset...
	Activation set in the compute backend ABI 0.2 the TVP layer...	Host device backend needing no kernel module or guest [8.3e]	All six activations vectorised for AVX2 and AVX-512 [8.3f]	ggml backend so unpatched llama.cpp or Ollama runs on...	Transposed matrix multiply through the stack compute ABI 0....
	CnuasSHMEM one sided put get and atomics [8.4]	CnuasBLAS v0.2 kernels generated through cnuascc [8.5]	CnuasDNN CnuasFFT CnuasSPARSE and CnuasSOLVER [8.6]	cnuas-prof profiler and cnuas-dogmi telemetry [8.7]	

CnuasBMC: OCP Rack Management 12 tasks	PCIe Link Capabilities on CnuasNIC and CnuasGPU (Gen5 x16...	Management plane design (fidelity tiers topology protocol surface)	meta-cnuas OpenBMC layer (cnuas-sled/tor/rmc machines)	Renode satellite skeletons (BMS PSU fan optics sequencer front...	UART front-panel satellite firmware and protocol
	Isolated management bridge and one real OpenBMC on QEMU ast2600	Redfish power wired to VM lifecycle and Serial over LAN	cnuas bmc/power/rack command groups and Redfish aggregation	Cnuas machine device trees and branded firmware image	BMS over RS-485 with real firmware in Renode
	Thermal Simulation model for fans and airflow (research)	DC-SCM root of trust and SPDMM attestation bench (optional)			